Mathématiques
Algèbre - Probabilités

Lycée du Diadème

Cours de préparation à l'agrégation interne

Vidal AGNIEL

Table des matières

Avant-propos

Vous trouverez au fil de ce cours différents symboles :

- Le symbole "  ", situé dans la marge, signifie que le point correspondant est un point délicat (il s'agit d'un *virage dangereux*).
- Le symbole "□" est un marqueur signifiant la fin d'une démonstration.
-  Ce cours peut comporter des fautes de frappe, des coquilles, voire des erreurs d'argumentation. Ainsi, il faut toujours être vigilant lorsque vous le suivez et que vous le travaillez . Vérifier que les exemples sont justes et que les preuves n'ont pas de fautes est un exercice très utile (et indispensable) pour comprendre les notions et comprendre leurs utilisations.

Vous trouverez aussi des notations mathématiques :

$\forall$	"pour tout"
$\exists$	"il existe"
$\in$	"appartenant à"
$\subset$	"inclus dans"
$\Rightarrow$	"implique"
$\Leftrightarrow$	"équivalent à"
$\mathcal{R}$	une relation binaire
$\equiv$	pour certaines relations d'équivalence
$\mathbb{N}$	l'ensemble des entiers naturels
$\mathbb{Z}$	l'ensemble des entiers relatifs
$\mathbb{Q}$	l'ensemble des nombres rationnels
$\mathbb{R}$	l'ensemble des entiers relatifs
$\mathbb{C}$	l'ensemble des nombres complexes
$(G, \times)$	un groupe
e_G	l'élément neutre d'un groupe G
g^{-1}	l'inverse dans un groupe d'un élément g
$\langle g \rangle$	le sous-groupe de G engendré par g
$(H, +)$	un groupe commutatif (parfois)
$a\mathbb{Z}$	l'ensemble des multiples de a
$\text{pgcd}(a, b)$	le plus grand diviseur commun d'entiers a et b
$\text{ppcm}(a, b)$	le plus petit multiple commun d'entiers a et b
$a \equiv b \bmod n$	a est congru à b modulo n (n divise $b - a$)
$\mathbb{K}$	un corps (en général $\mathbb{R}, \mathbb{C}$ ou $\mathbb{Q}$, parfois $\mathbb{Z}/p\mathbb{Z}$)
$\mathscr{M}_{n,p}(\mathbb{K})$	l'ensemble des matrices à n lignes et p colonnes à coefficients dans $\mathbb{K}$
$\mathscr{M}_n(\mathbb{K})$	l'ensemble des matrices carrées $n \times n$ à coefficients dans $\mathbb{K}$
$E_{i,j}$	les matrices de la base canonique de $\mathscr{M}_{n,p}(\mathbb{K})$
M^{-1}	l'inverse d'une matrice carrée M
${}^tM$	la transposée d'une matrice M
$rg(M)$	le rang d'une matrice M
$Tr(M)$	la trace d'une matrice carrée M
$\mathbb{K}[X]$	l'ensemble des polynômes à une indéterminée à coefficients dans $\mathbb{K}$
$\mathbb{K}_n[X]$	l'ensemble des polynômes à une indéterminée à coefficients dans $\mathbb{K}$, de degré au plus n
$\deg(P)$	le degré d'un polynôme P
$a \equiv b \bmod n$	la relation de congruence modulo n (n divise $b - a$)
$\mathbb{Z}/n\mathbb{Z}$	l'ensemble des classes d'équivalence pour la relation de congruence modulo n

$(A, +, \times)$	un anneau
0_A	l'élément nul d'un anneau A pour l'addition $+$
1_A	l'élément unitaire d'un anneau A pour la multiplication $\times$
a^{-1}	l'inverse dans A d'un élément a inversible
I	un idéal de l'anneau A
$\langle a \rangle$	l'idéal de A engendré par a
$(\mathbb{Z}/n\mathbb{Z}, +, \times)$	l'anneau des classes d'équivalence modulo n
$\mathbb{Z}/p\mathbb{Z}$	le corps à p éléments, avec p premier
$\mathbb{K}(X)$	le corps des fractions rationnelles à coefficients dans $\mathbb{K}$
$\sigma(E)$	l'ensemble des bijections d'un ensemble E dans lui-même
$\circ$	la composition des applications, donc $(\sigma(E), \circ)$ est un groupe.
$\mathcal{S}_n$	l'ensemble des permutations de $[\![1, n]\!]$ (le groupe symétrique d'ordre n)
(i, j)	une transposition
$(a_1, a_2, \ldots, a_p)$	un p-cycle
$\varepsilon(\sigma)$	la signature de σ
$\mathcal{A}_n$	groupe alterné d'ordre n
E	un ensemble ou un espace vectoriel
F	un sous-espace vectoriel de l'espace vectoriel E
u	un endomorphisme de l'espace vectoriel E
$\det_{\mathcal{B}}(u_1, \ldots, u_n)$	le déterminant de la famille de vecteurs $(u_1, \ldots, u_n)$ par rapport à la base $\mathcal{B}$
$\det(f)$	le déterminant de l'endomorphisme f
$\det(A)$	le déterminant d'une matrice carrée A
$\mathrm{com}\,(A)$	la comatrice
$\mathrm{Tr}\,(u)$	la trace d'un endomorphisme u
$\mathrm{Tr}\,(M)$	la trace d'une matrice carrée M
$E_\lambda(u)$	le sous-espace vectoriel propre à une valeur propre λ de l'endomorphisme u
$F_\lambda(u)$	le sous-espace vectoriel caractéristique à une valeur propre λ de l'endomorphisme u
$\mathrm{Spec}\,(u)$	le spectre d'un endomorphisme u
$\chi_A(X)$	le polynôme caractéristique d'une matrice carrée A
$\chi_u(X)$	le polynôme caractéristique d'un endomorphisme u
$C_{P(X)}$	la matrice compagnon d'un polynôme P
$m(\lambda)$	l'ordre de multiplicité d'une valeur propre λ (dans χ_u)
$r_u(\lambda)$	l'indice d'une valeur propre λ (ordre de multiplicité dans μ_u)
e_u	le morphisme d'évaluation des polynômes en un endomorphisme u
e_A	le morphisme d'évaluation des polynômes en une matrice carré A
$\mathbb{K}[u]$	la sous-algèbre engendrée par un endomorphisme u
$\mathcal{I}_u$	l'idéal annulateur de u
μ_u	le polynôme minimal d'un endomorphisme u
μ_A	le polynôme minimal d'une matrice carrée A
$< x \mid y >$	le produit scalaire de deux vecteurs(éléments) x et y
$N(x), \|\|x\|\|$	la norme de x
$d(x, y)$	la distance entre deux vecteurs(éléments) x et y
$x \perp y$	deux vecteurs(éléments) x et y sont orthogonaux
$A^\perp$	l'orthogonal ou le supplémentaire orthogonal de A
$E = F \overset{\perp}{\oplus} F^\perp$	F et $F^\perp$ sont supplémentaires de E
ϕ_a	la forme linéaire $\phi_a(x) =< a \mid x >$
P_F	la projection orthogonale sur F parallèlement à $F^\perp$

Chapitre 1 Vocabulaire, logique et raisonnements

Table des matières du chapitre

1.1 Variables et quantificateurs

Définition 1
Une **proposition** est une phrase mathématique qui est soit **vraie**, soit **fausse**.

En mathématiques, on utilise des **variables**. Il s'agit de lettres (x, y, a, n, α, β, ...) qui ont parfois des indices (x_1, x_2, ...). Une variable ne désigne pas un objet particulier mais des objets appartenant à un certain ensemble.

Les objets mathématiques que nous utilisons se conçoivent dans trois grandes familles : les **ensembles** ($\mathbb{N}, \mathbb{Z}, \mathbb{Q}, \mathbb{R}, \mathbb{R}^2, \mathbb{R}[X], M_r$ les **nombres et éléments** d'un ensemble (entiers, rationnels, réels, vecteurs, matrices, ...), les **fonctions** (dérivées, suites, polynômes, séries, ...).

Souvent, une proposition dépend d'une ou plusieurs variables. La **valeur de vérité** de la proposition (vraie ou fausse) peut alors être donnée lorsque l'on précise les **valeurs** des variables. En général, on note $P(x)$ une proposition qui dépend de la variable est x.

1.1.1 Quelques règles de calcul pour les propositions

Proposition 2
Soient P, Q et R des propositions. On a :

- $non(non\, P) \equiv P$,
- $non(P\ et\ Q) \equiv (non\, P)\ ou\ (non\, Q)$,
- $non(P\ ou\ Q) \equiv (non\, P)\ et\ (non\, Q)$,
- $P\ et\ (Q\ ou\ R) \equiv (P\ et\ Q)\ ou\ (P\ et\ R)$,
- $P\ ou\ (Q\ et\ R) \equiv (P\ ou\ Q)\ et\ (P\ ou\ R)$,
- La proposition « $P\ et\ (non\, P)$ » est toujours fausse,
- La proposition « $P\ ou\ (non\, P)$ » est toujours vraie : soit P est vraie, soit $non(P)$ est vraie.

Preuve — On peut démontrer ces propriétés avec des tables de vérité. Donnons un exemple.

P	Q	$P\ et\ Q$	$non(P\ et\ Q)$	$non(P)$	$non(Q)$	$non(P)\ ou\ non(Q)$
V	V	V	F	F	F	F
V	F	F	V	F	V	V
F	V	F	V	V	F	V
F	F	F	V	V	V	V

Donc $non(P\ et\ Q) \equiv non(P)\ ou\ non(Q)$. □

1.1.2 Implication

Soient P et Q des propositions. La proposition « $P \Rightarrow Q$ » est la proposition qui est

- fausse lorsque P est vraie et Q est fausse,
- vraie dans les autres cas.

REMARQUE 3 — *Lorsque « $P \Rightarrow Q$ » est vraie, on dit que*

- *P est une **condition suffisante** pour Q,*
- *Q est une **condition nécessaire** pour P.*

DÉFINITION 4

- *La proposition « $Q \Rightarrow P$ » s'appelle la **réciproque** de l'implication « $P \Rightarrow Q$ ».*
- *La proposition « $non(Q) \Rightarrow non(P)$ » s'appelle la **contraposée** de l'implication « $P \Rightarrow Q$ ».*

PROPOSITION 5 (**Implication et contraposée**)
On a $P \Rightarrow Q \ \equiv \ (non\, Q) \Rightarrow (nonP)$.

1.1.3 Équivalence

Soient P et Q des propositions. La proposition « $P \Leftrightarrow Q$ » (se lit « P équivalent à Q » ou « P si et seulement si Q ») est la proposition qui est

- vraie lorsque les propositions P et Q sont toutes les deux vraies ou toutes les deux fausses,
- fausse dans les autres cas.

PROPOSITION 6 (**Equivalence et double-implication**)
On a $P \Leftrightarrow Q \ \equiv \ (P \Rightarrow Q)\ et\ (Q \Rightarrow P)$.

1.1.4 Quantificateurs

Soit E un ensemble. Soit $P(x)$ une proposition dépendant de la variable x, avec $x \in E$.

DÉFINITION 7 (**Quantificateur "pour tout"**)
Le quantificateur $\forall$ (se lit « pour tout » ou « quel que soit ») permet de définir la proposition « $\forall\, x \in E,\ P(x)$ » qui est :

- vraie lorsque pour tous les éléments x appartenant à E, $P(x)$ est vraie,
- fausse sinon (c'est-à-dire si $P(x)$ est fausse pour **au moins** un élément x de E).

DÉFINITION 8 (**Quantificateur "il existe"**)
Le quantificateur $\exists$ (se lit « il existe ») permet de définir la proposition « $\exists\, x \in E,\ P(x)$ » (« il existe ... tel que ») qui est :

- vraie lorsque $P(x)$ est vraie pour au moins un élément x de E,
- fausse lorsque $P(x)$ est fausse pour tous les éléments x de E.

Après ces deux quantificateurs, il faut toujours préciser l'ensemble où on prend notre variable : « $\forall x,\, x(1-x) \geq 0$ » n'a pas de sens (pour tout x dans quoi ?).
« $\forall x \in \mathbb{R},\, x(1-x) \geq 0$ » est fausse, mais « $\forall x \in [0,1],\, x(1-x) \geq 0$ » est vraie.

Attention ! Les symboles « $\forall$ » et « $\exists$ » ne sont pas des **abréviations**, ils ne doivent pas être utilisés dans une phrase en français.

PROPOSITION 9 (**Négation et quantificateurs**)
On a :

- $non(\forall\, x \in E,\ P(x)) \ \equiv \ \exists\, x \in E,\ non(P(x))$,
- $non(\exists\, x \in E,\ P(x)) \ \equiv \ \forall\, x \in E,\ non(P(x))$.

Attention ! Dans une proposition, on ne peut pas échanger les positions de $\forall$ et $\exists$.

Par exemple, "$\forall x \in \mathbb{R}, \exists y \in \mathbb{R}$ tel que $3y - 4 = x$" est une proposition qui est vraie. Mais, "$\exists y \in \mathbb{R}$ t.q. $\forall x \in \mathbb{R}$ on a $3y - 4 = x$" est une proposition qui est fausse.

1.2 Méthodes de démonstration - Exemples, raisonnements classiques

Un **raisonnement mathématique** est un processus permettant d'établir, à partir de propositions vraies, de nouvelles propositions, de nouveaux résultats, en utilisant des principes logiques. Dans cette partie, nous étudions différents types de raisonnement.

Lorsque l'on écrit une proposition mathématique, il est sous-entendu qu'elle est vraie.
Sinon, on ne l'écrit pas.

1.2.1 Vocabulaire

Une proposition s'énonce souvent sous la forme « Si A alors B » ($A \Rightarrow B$).

- La proposition A regroupe les **hypothèses** .
- La proposition B regroupe les **conclusions** .

1.2.2 Quelques exemples de rédaction

Quand on dit « **Supposons** P », on part de l'hypothèse que P est vraie.

Pour bien rédiger en mathématiques, on doit respecter certaines règles.

- On doit introduire les nouveaux objets.
  - Pour introduire une variable x qui représente un **élément quelconque** d'un ensemble E, on peut écrire :
    « *Soit* $x \in E$ » ou « *Soit* x *un élément de* E ».
  - Pour donner un nom, par exemple M, à une quantité connue ou à un objet que l'on va souvent utiliser, on peut écrire :
    « ***Posons*** $M = \ldots$ » ou « ***Notons*** $M = \ldots$ » (ou « On pose $M = \ldots$, On note $M = \ldots$ »).
    Par exemple, « *Posons* $M = \dfrac{\sqrt{2}+3}{4}$ ».
- On doit mettre des liens logiques entre les arguments, comme par exemple :
  - « **Donc** » , « **D'où** », « **Ainsi** » ,
  - « **On en déduit que** »,
  - « **Or** » (permet d'ajouter un argument),
  - « **Finalement** » (pour une conlusion à la fin du raisonnement), ...
- On peut annoncer ce que l'on veut faire. Cela aide a bien clarifier l'objectif.
  « ***Montrons que*** ... », « On veut montrer que ... ».

Exemples 10

- *On veut montrer la proposition » $\forall x \in E$, $P(x)$ ».*
  On pose x un élément quelconque de E et on montre que $P(x)$ est vraie.
  Exemple de rédaction :
  Soit $x \in E$. Montrons que $P(x)$ est vraie.
  $\vdots$
  Donc $P(x)$.
  Ainsi, pour tout $x \in E$, on a $P(x)$.
- *On veut montrer la proposition « $\exists x \in E$, $P(x)$ ». En général, on trouve explicitement un élément $x_0 \in E$ tel que $P(x_0)$ est vraie. Le raisonnement peut se faire par **analyse - synthèse**.*
- *On veut montrer qu'un $x \in E$ qui vérifie $P(x)$ doit être unique.*
  On suppose qu'il en existe deux et on montre qu'ils sont égaux.
- *On veut montrer une inclusion $A \subset B$. On prend un élément quelconque x dans A, on montre que x est dans B.*
- *On veut montrer une égalité d'ensembles $A = B$. Souvent, on procède par **double inclusion** : on montre $A \subset B$ puis on montre $B \subset A$.*

- *On veut montrer une implication « $P \Longrightarrow Q$ ». On suppose que P est vraie et on montre que Q est vraie. S'il n'y a pas besoin de faire d'autres hypothèses/méthode de démonstration, on parle alors de* ***raisonnement direct****.*
- *On veut montrer une équivalence « $P \Longleftrightarrow Q$ ». Souvent, on procède par* ***double implication*** *: on montre d'abord « $P \Longrightarrow Q$ » puis on montre « $Q \Longrightarrow P$ ».*
- *On veut montrer la proposition « P ou Q ». On peut montrer l'implication « $non(P) \Longrightarrow Q$ ».*

Rappelons que pour prouver qu'une proposition P est fausse, on peut montrer que sa négation $non(P)$ est vraie. Par exemple, pour montrer que la proposition « $\forall x \in E,\ P(x)$ » est fausse, on peut montrer que sa négation « $\exists\, x \in E,\ non(P(x))$ »
est vraie. Donner un élément x_0 de E tel que $non(P(x_0))$ est vraie s'appelle un **contre-exemple** .

Attention ! Lorsque l'on utilise la flèche « $\Leftrightarrow$ », il faut être sûr que le sens direct ($\Rightarrow$) et le sens réciproque ($\Leftarrow$) soient vrais.

Remarque 11 (**Utiliser une implication**) — *Dans un exercice, pour appliquer un théorème de la forme $A \Rightarrow B$ (« Si A alors B »), on commence donc par vérifier que A (les hypothèses) est vraie. On écrit par exemple*

« On a A. D'après le théorème ..., on sait que A implique B. Donc on a B »

Exemples 12 (Exemples de rédaction de preuves)

- ***Démontrons*** *que pour tout $n \in \mathbb{Z}$, si n est pair alors n^2 est pair.*
  Il s'agit de la proposition « $\forall n \in \mathbb{Z},\ (n$ est pair $\Rightarrow n^2$ est pair$)$ ».

  Preuve *: Soit $n \in \mathbb{Z}$. Montrons que si n est pair alors n^2 est pair.*
  Supposons n pair. Nous allons montrer que n^2 est pair par ***raisonnement direct***
  Par hypothèse, n est pair, donc il existe $k \in \mathbb{Z}$ tel que $n = 2k$. On a donc $n^2 = (2k)^2 = 4k^2 = 2(2k^2)$ et $2k^2 \in \mathbb{Z}$.
  Donc n est pair.
  Donc, pour tout $n \in \mathbb{Z}$, si n est pair alors n^2 est pair.

- *Démontrons que, pour tout $(a, b) \in \mathbb{R}^2$, $(a+b)^2 = a^2 + b^2$ si et seulement si $a = 0$ ou $b = 0$.*
  Il s'agit de la proposition « $\forall (a, b) \in \mathbb{R}^2,\ \big((a+b)^2 = a^2 + b^2 \Leftrightarrow (a = 0$ ou $b = 0)\big)$. »

  Preuve *: Soit $(a, b) \in \mathbb{R}^2$. Montrons, par* ***double implication****, que $(a+b)^2 = a^2 + b^2$ si et seulement si $a = 0$ ou $b = 0$.*

  ▷ *Supposons que $(a+b)^2 = a^2 + b^2$. Montrons que $a = 0$ ou $b = 0$.*
  On sait que $(a+b)^2 = a^2 + 2ab + b^2$ et, par hypothèse, $(a+b)^2 = a^2 + b^2$. Donc

  $$a^2 + 2ab + b^2 = a^2 + b^2.$$

  Donc, $2ab = 0$, c'est-à-dire $ab = 0$. Donc $a = 0$ ou $b = 0$.
  Donc si $(a+b)^2 = a^2 + b^2$ alors $a = 0$ ou $b = 0$.

  ◁ *Réciproquement, supposons $a = 0$ ou $b = 0$. Montrons que $(a+b)^2 = a^2 + b^2$.*
  $a = 0$ ou $b = 0$, dans les deux cas, on a $ab = 0$ et donc $(a+b)^2 = a^2 + 2ab + b^2 = a^2 + b^2$.
  Donc si $a = 0$ ou $b = 0$ alors $(a+b)^2 = a^2 + b^2$.

- *Démontrons que la proposition « $\forall n \in \mathbb{N},\ 2^n > n^2$ » est fausse.*

  Preuve *: Donnons un* ***contre-exemple****.*
  Pour $n = 3$, on a $8 = 2^3 < 3^2 = 9$. Il existe donc un entier naturel n tel que $2^n \leq n^2$. Donc la proposition « $\forall n \in \mathbb{N},\ 2^n > n^2$ » est fausse.

- *Démontrons l'égalité d'ensembles $\big\{x \in \mathbb{R} \mid x^2 + x + 2 \leq 2\big\} = [-1, 0]$.*

  Preuve *: Procédons par* ***double inclusion****.*
  • *Montrons que $\big\{x \in \mathbb{R} \mid x^2 + x + 2 \leq 2\big\} \subset [-1, 0]$. Soit $x \in \mathbb{R}$ tel que $x^2 + x + 2 \leq 2$, alors $x^2 + x \leq 0$, donc $x \leq -x^2 \leq 0$. On a donc $x \leq 0$ et $x(x+1) \leq 0$ donc $x + 1 \geq 0$, c'est-à-dire $x \geq -1$. Finalement, $-1 \leq x \leq 0$.*
  • *Montrons que $[-1, 0] \subset \big\{x \in \mathbb{R} \mid x^2 + x + 2 \leq 2\big\}$. Soit $x \in [-1, 0]$. Alors $x^2 + x + 2 = x(x+1) + 2 \leq 2$ car $x \leq 0$ et $x + 1 \geq 0$.*
  Par double inclusion, on a donc l'égalité.

1.2.3 Raisonnements classiques

Nous avons déjà mentionné et fait des exemples pour :

- Le **raisonnement direct** : Pour montrer qu'on a « $P \Rightarrow Q$ », on suppose que P est vraie, et on utilise directement des calculs/théorèmes pour obtenir que Q est vraie.
- Le **contre-exemple** : Pour montrer que « $P \Rightarrow Q$ » est fausse, on cherche un cas de figure (un nombre entier n, un nombre réel x, une fonction f) dans lequel P est vraie et Q est fausse.
- La **double-implication** : Pour montrer qu'on a « $P \Leftrightarrow Q$ », on montre d'une part que « $P \Rightarrow Q$ » est vraie, et réciproquement que « $P \Leftarrow Q$ » est vraie. En général, l'une des implications est bien plus facile que l'autre.
- La **double-inclusion** : Pour montrer qu'on a $E = F$ (pour E, F des ensembles), on montre d'une part que $E \subset F$, et réciproquement que $F \subset E$. En général, l'une des inclusions est bien plus facile que l'autre.

Raisonnement par contraposée (ou par contraposition)

Pour montrer que la proposition « $P \Rightarrow Q$ » est vraie, on peut montrer que sa contraposée, qui est la proposition « $non(Q) \Rightarrow non(P)$ », est vraie. On parle de **raisonnement par contraposée** .

EXEMPLE 13 — *Démontrons que pour tout $n \in \mathbb{Z}$, si n^2 est pair alors n est pair.*

Cette proposition s'écrit « $\forall n \in \mathbb{Z}$, (n^2 pair $\Rightarrow$ n pair) ».

Preuve *: Soit $n \in \mathbb{Z}$.*
Plutôt que de montrer « n^2 pair $\Rightarrow$ n pair », on montre la contraposée « n impair $\Rightarrow$ n^2 impair », plus facile à démontrer.
Supposons que n est impair. Nous allons montrer que n^2 est impair.
Par hypothèse, n est impair, donc il existe $k \in \mathbb{Z}$ tel que $n = 2k + 1$. On a donc

$$n^2 = (2k+1)^2 = 4k^2 + 4k + 1 = 2(2k^2 + 2k) + 1,$$

et $2k^2 + 2k \in \mathbb{Z}$. Donc n^2 est impair.
Donc si n est impair alors n^2 est impair. On en déduit par contraposée que si n^2 est pair alors n est pair.

Raisonnement par l'absurde

On souhaite démontrer qu'une proposition P est vraie. Le **raisonnement par l'absurde** consiste à supposer que P est fausse, c'est-à-dire à supposer que $non(P)$ est vraie et montrer que cela conduit à une contradiction. On en déduit alors que P est vraie.

EXEMPLE 14 — *Démontrons que $\sqrt{2}$ est un **nombre irrationnel**[1] .*

Preuve *: Supposons, par l'absurde, que $\sqrt{2}$ est un nombre rationnel. Alors il existe deux entiers $p \in \mathbb{Z}$ et $q \in \mathbb{N}^*$, **premiers entre eux**[2], tels que $\sqrt{2} = \frac{p}{q}$.*
*On a donc $2q^2 = p^2$. On en déduit que p^2 est pair. Or, nous avons vu à l'exemple 13 que pour tout $n \in \mathbb{Z}$, si n^2 est pair alors n est pair. Donc p est pair. Il existe donc $k \in \mathbb{Z}$ tel que $p = 2k$. On a donc $p^2 = 4k^2$, puis $2q^2 = 4k^2$. Donc finalement, $q^2 = 2k^2$. On en déduit que q^2 est pair. Donc, comme précédemment, q est pair. On en déduit que 2 **divise** p et 2 divise q. Ceci est **absurde** car cela contredit le fait que p et q sont premiers entre eux !*
Donc $\sqrt{2}$ est un nombre irrationnel.

Raisonnement par disjonction de cas

Le **raisonnement par disjonction de cas** permet de simplifier un raisonnement en distinguant toutes les situations possibles. Cela est notamment utilisé lorsque la proposition dépend d'une variable x.

EXEMPLE 15 — *Démontrons que, pour tout $n \in \mathbb{N}$, $\dfrac{n(n+1)(2n+1)}{6}$ est un entier naturel.*

1. L'ensemble des nombres irrationnels est $\mathbb{R} \setminus \mathbb{Q}$: ce sont les nombres réels qui ne sont pas des nombres rationnels
2. Si un entier naturel d divise p et divise q alors $d = 1$

Preuve *: Soit $n \in \mathbb{N}$. Montrons que $n(n+1)$ est divisible par 2.*
• 1^{er} *cas : n est pair. Alors $n(n+1)$ est pair.*
• 2^{nd} *cas : n est impair. Alors $n+1$ est pair et donc $n(n+1)$ est pair.*
Donc, dans tous les cas, $n(n+1)$ est divisible par 2.
Montrons maintenant que $n(n+1)(2n+1)$ est divisible par 3.
• 1^{er} *cas : n est un multiple de 3. Alors $n(n+1)(2n+1)$ est un multiple de 3.*
• $2^{ème}$ *cas : $n+1$ est un multiple de 3. Alors $n(n+1)(2n+1)$ est un multiple de 3.*
• $3^{ème}$ *cas : $n-1$ est un multiple de 3. Alors $2n+1 = 2(n-1)+3$ est un multiple de 3, et donc $n(n+1)(2n+1)$ est un multiple de 3.*
Donc, dans tous les cas, $n(n+1)(2n+1)$ est divisible par 3.
Ainsi $n(n+1)(2n+1)$ est toujours divisible par 3 et par 2. Or 2 et 3 sont premiers entre eux donc $n(n+1)(2n+1)$ est divisible par 6.

Raisonnement par récurrence

Soit $P(n)$ une proposition dépendant d'une variable $n \in \mathbb{N}$. Soit $n_0 \in \mathbb{N}$.

Démontrer **par récurrence** que la proposition « $\forall n \geq n_0,\ P(n)$ » est vraie repose sur le principe suivant :

Si $P(n_0)$ est vraie (**initialisation**) ET pour tout $n \geq n_0$, « $P(n) \Rightarrow P(n+1)$ » est vraie (**hérédité**), alors, pour tout $n \geq n_0$, $P(n)$ est vraie.

Remarque 16 — *En général, n_0 vaut 0, 1 ou 2.*

On peut donc, par exemple, rédiger un raisonnement par récurrence comme suit :

« *Démontrons le résultat par récurrence. Notons, pour tout entier $n \geq n_0$, $P(n)$ la propriété "..... "* »

- Initialisation : *Vérifions $P(n_0)$ (ce n_0 est à déterminer en fonction de l'énoncé et la vérification est souvent facile.)*

  $\vdots$

  D'où $P(n_0)$.
- Hérédité : *Soit $n \geq n_0$. Supposons $P(n)$, montrons $P(n+1)$. Dans cette étape, on va utiliser la propriété $P(n)$, qui est l'**hypothèse de récurrence** .*

  $\vdots$

  D'où $P(n+1)$.

Par récurrence, on en déduit donc que, pour tout $n \geq n_0$, on a $P(n)$.

Remarque 17 — ***Attention :** Il est très important de démarrer l'hérédité au même entier n_0 que l'initialisation. Si on montre $P(0)$ et pour tout $n \geq 1$, $P(n) \Longrightarrow P(n+1)$, cela ne montre pas que $P(n)$ est vraie pour tout $n \geq 0$.*

Exemple 18 — *Démontrons que, pour tout $n \in \mathbb{N}^*$, $\displaystyle\sum_{k=1}^{n} k = \frac{n(n+1)}{2}$.*

Preuve : *Démontrons le résultat par récurrence sur n. Notons, pour tout $n \in \mathbb{N}^*$, $P(n)$ la propriété :*

$$\text{« } \sum_{k=1}^{n} k = \frac{n(n+1)}{2} \text{ ».}$$

- *Initialisation : Pour $n = 1$, on a $1 = \dfrac{1(1+1)}{2}$. D'où $P(1)$.*
- *Hérédité : Soit $n \geq 1$. Supposons $P(n)$, montrons $P(n+1)$.*
  On a
  $$\sum_{k=1}^{n+1} k = \left(\sum_{k=1}^{n} k\right) + (n+1).$$
  Donc d'après l'hypothèse de récurrence $P(n)$,
  $$\sum_{k=1}^{n+1} k = \frac{n(n+1)}{2} + n + 1 = \frac{n(n+1) + 2(n+1)}{2} = \frac{(n+1)(n+2)}{2}.$$
  D'où $P(n+1)$.

Par récurrence, on a donc démontré que, pour tout $n \in \mathbb{N}^*$, $\displaystyle\sum_{k=1}^{n} k = \frac{n(n+1)}{2}$.

Le principe que l'on vient de détailler est appelé une **récurrence simple** : on déduit $P(n+1)$ directement de $P(n)$. Parfois, on ne peut déduire $P(n+2)$ que de $P(n+1)$ et $P(n)$. On parle alors de **récurrence double** . Le principe est le suivant :

Si $P(n_0)$ et $P(n_0+1)$ sont vraies (*initialisation*) ET pour tout $n \geq n_0$, la proposition « $\Big(P(n)$ et $P(n+1)\Big) \Rightarrow P(n+2)$ » est vraie (*hérédité*), alors, pour tout $n \geq n_0$, $P(n)$ est vraie.

Exemple 19 — *Soit $(u_n)_{n\in\mathbb{N}}$ la **suite** définie par*
$$\begin{cases} u_0 = 2, \\ u_1 = 3, \\ \text{pour tout } n \in \mathbb{N}, u_{n+2} = 3u_{n+1} - 2u_n. \end{cases}$$

Démontrons que, pour tout $n \in \mathbb{N}$, $u_n = 1 + 2^n$.

Preuve : *Démontrons le résultat par récurrence sur n. Notons, pour tout $n \in \mathbb{N}$, $P(n)$ la propriété :*
$$\text{« } u_n = 1 + 2^n \text{ ».}$$

- *Initialisation : On a $u_0 = 2 = 1 + 2^0$ et $u_1 = 3 = 1 + 2^1$ donc $P(0)$ et $P(1)$ sont vraies.*
- *Hérédité : Soit $n \in \mathbb{N}$. Supposons $P(n)$ et $P(n+1)$, montrons $P(n+2)$.*
  On a $u_{n+2} = 3u_{n+1} - 2u_n$, donc par hypothèses de récurrence,
  $$u_{n+2} = 3 \times (1 + 2^{n+1}) - 2 \times (1 + 2^n) = 3 + 3 \times 2^{n+1} - 2 - 2^{n+1} = 1 + 2 \times 2^{n+1} = 1 + 2^{n+2}.$$
  D'où $P(n+2)$.

Par récurrence, on a donc démontré que, pour tout $n \in \mathbb{N}$, $u_n = 1 + 2^n$.

Enfin, il arrive que $P(n+1)$ ne puisse se déduire que de $P(n_0)$, $P(n_0+1)$, …, $P(n)$. On parle alors de **récurrence forte** . Le principe est le suivant :

Si $P(n_0)$ est vraie (*initialisation*) ET pour tout $n \geq n_0$, la proposition « $\Big(P(n_0)$ et $P(n_0+1)$ et … et $P(n)\Big) \Rightarrow P(n+1)$ » est vraie (*hérédité*), alors, pour tout $n \geq n_0$, $P(n)$ est vraie.

Rappelons que pour tout $n \in \mathbb{N}$ et pour tout $x \in \mathbb{C}$, $\displaystyle\sum_{k=0}^{n} x^k = \begin{cases} \dfrac{x^{n+1}-1}{x-1} & \text{si } x \neq 1 \\ n+1 & \text{si } x = 1 \end{cases}$. Cette somme s'appelle une **somme géométrique**.

Raisonnement par double-inégalité

Dans certaines situations, pour montrer que deux nombres réels a, b sont égaux, on montre à la place que $a \leq b$ et que $b \leq a$.

Cette méthode est analoque de la double-implication et à la double-inclusion.

Ce raisonnement est en général utile lorsque a et/ou b sont obtenus comme des majorants/minorants/limites. Dans ces situations, il faut souvent revenir aux propriétés de sup, inf lim pour obtenir les inégalités voulues, et ainsi obtenir l'égalité.

Exemple : Le théorème des gendarmes contient une double-inégalité. Avec $u_n \leq v_n$ et $v_n \leq w_n$ on s'assure que $\lim_n u_n \leq \lim_n v_n$ puis que $\lim_n v_n \leq \lim_n w_n$. L'hypothèse $\lim_n u_n = \lim_n w_n$ apporte le résultat.

Exemple 20 — *Soit E un espace vectoriel normé, et soit $A \in \mathcal{L}_C(E)$ une application linéaire continue sur E. Par continuité de A,* $\sup(\{\frac{\|AX\|}{\|X\|}, X \in E^*\})$ *existe et est fini. On note ce nombre $|||A|||$.*

Alors, on a $\sup(\{\frac{\|AX\|}{\|X\|}, X \in E^*\}) = |||A||| = \inf(\{C > 0 \text{ t.q. } \|AX\| \leq C\|X\|, \forall X \in E\})$, *et* $|||A||| = \sup(\{\frac{\|AX\|}{\|X\|}, X \in E, \|X\| = 1\})$.
Pour montrer que ces deux sup et que cet inf sont égaux, le plus pratique est de raisonner par double-inégalité. On note a, b, c ces trois quantités, puis on vérifie (Exercice) que $a \leq b$, $b \leq a$, $a \leq c$, $c \leq a$.

Lorsque E est de dimension finie, la sphère unité de E est un compact. Une fonction continue sur un compact atteint ses bornes, donc le deuxième sup est atteint. Il existe X de norme 1 tel que $\|AX\| = |||A|||.\|X\| = |||A|||$. Si l'on a une idée de la valeur de $|||A|||$ (disons 2), on peut alors dans un premier temps montrer que $\|AX\| \leq 2\|X\|$ pour tout $X \in E$ ($|||A||| \leq 2$), puis trouver un X non-nul tel que $\|AX\| = 2\|X\|$ ($2 \leq |||A|||$).

Mais, trouver un vecteur X en lequel A atteint sa norme n'est pas toujours facile. En dimension infinie il n'existe en général pas. Cependant, on peut généraliser la méthode en utilisant la définition séquentielle du sup *: On montre dans un premier temps que $\|AX\| \leq 2\|X\|$ pour tout $X \in E$ ($|||A||| \leq 2$). Puis, on trouve une suite $(X_n)_n$ de vecteurs de norme 1 tels que $\|AX_n\| \to_n 2$ ($2 \leq |||A|||$).*

Raisonnement par analyse-synthèse

Lorsque l'on veut chercher les solutions d'un problème et montrer que celles que l'on a trouvées sont les seules, on utilise le raisonnement par **analyse-synthèse**.

Ce raisonnement s'effectue en deux étapes :

1. *Analyse* : On suppose que l'on a une solution du problème et on cherche des propriétés vérifiées par cette solution.
2. *Synthèse :* Parmi les éléments vérifiant les propriétés obtenues dans l'analyse, on détermine ceux qui sont bien solutions du problème (il n'y en a pas d'autres).

On obtient ainsi l'ensemble des solutions du problème.

Ce raisonnement est particulièrement utile pour démontrer l'**existence** et l'**unicité** d'une solution à un problème. Cependant, c'est le raisonnement le plus difficile à utiliser car c'est le plus élaboré (il faut trouver les bonnes propriétés lors de la phase d'analyse pour pouvoir bien conclure avec la synthèse).
Pour un raisonnement par l'absurde, par contraposée, par récurrence, on sait ce que l'on doit obtenir (on connaît le "résultat" auquel on doit aboutir), alors que pour une analyse-synthèse on ne sait pas ce que l'on doit obtenir (et il faudra trouver le "résultat", en plus d'y aboutir).

Exemple 21 — *Déterminons l'ensemble des fonctions $f : \mathbb{R} \longrightarrow \mathbb{R}$ telles que, pour tout $(x, y) \in \mathbb{R}^2$,*

$$f(y - f(x)) = 2 - x - y.$$

Preuve : *Raisonnons par analyse-synthèse.*
• *Analyse : Soit $f : \mathbb{R} \longrightarrow \mathbb{R}$ une fonction telle que, pour tout $(x, y) \in \mathbb{R}^2$, $f(y - f(x)) = 2 - x - y$.*
Soit $x \in \mathbb{R}$. Prenons $y = f(x) \in \mathbb{R}$. Alors $f(0) = 2 - x - f(x)$. Donc $f(x) = 2 - f(0) - x$.
Donc f est de la forme $f(x) = a - x$ où $a \in \mathbb{R}$.

• *Synthèse : Déterminons parmi les fonctions de la forme* $x \longmapsto a - x$ *celles qui vérifient la condition de l'énoncé. Soit* $a \in \mathbb{R}$. *Soit* $f : x \longmapsto a - x$.
On a

$$f(y - f(x)) = f(y - (a - x)) = f(y + x - a) = a - (y + x - a) = 2a - x - y.$$

Donc f *vérifie la condition de l'énoncé si et seulement si* $2a = 2$, *soit encore si et seulement si* $a = 1$.
• *Conclusion : Il existe donc une unique fonction vérifiant la condition de l'énoncé, c'est la fonction* $x \longmapsto 1 - x$.

Chapitre 2 Fonctions

Table des matières du chapitre

2.1 Définitions

Définition 1
Soient E et F deux ensembles. On appelle **fonction de E dans F** un triplet $f = (E, F, \mathcal{G})$ où $\mathcal{G}$ est un ensemble vérifiant :

1. $\mathcal{G} \subset E \times F$,
2. pour tout $x \in E$, il existe un unique $y \in F$ tel que $(x, y) \in \mathcal{G}$.

Avec les notations précédentes,

- E est appelé l'**ensemble de définition** ou ensemble de départ de f,
- F est appelé l'**ensemble d'arrivée** de f,
- $\mathcal{G}$ est appelé le **graphe** de f.
- Pour tout $x \in E$, l'unique $y \in F$ tel que $(x, y) \in \mathcal{G}$ est noté $f(x)$. On l'appelle l'**image** de x par f.
- Si $y \in F$ et si $x \in E$ est tel que $f(x) = y$, on dit que x est un **antécédent** de y par f.
- L'ensemble des fonctions de E dans F est noté F^E ou $\mathcal{F}(E, F)$.

Notation : En général, le triplet $f = (E, F, \mathcal{G})$ est noté de la façon suivante :

$$\begin{array}{rccc} f : & E & \longrightarrow & F \\ & x & \longmapsto & f(x) \end{array},$$

avec $f(x)$ remplacé par son expression.

On utilise aussi la notation $f : E \longrightarrow F$ pour dire que f est une fonction de E dans F.

Exemples 2

- *Soit E un ensemble. On appelle **fonction identité** de E, notée id_E, la fonction définie par*

$$\begin{array}{rccc} \mathrm{id}_E : & E & \longrightarrow & E \\ & x & \longmapsto & x \end{array}.$$

- *Soient E et F deux ensembles. Soit a un élément de F. On appelle **fonction constante** à a la fonction définie par*

$$\begin{array}{rccc} f : & E & \longrightarrow & F \\ & x & \longmapsto & a \end{array}.$$

- *Soient E un ensemble non vide et A une partie de E. On appelle **fonction indicatrice de** A, notée $\mathbb{1}_A$ la fonction définie par*

$$\begin{array}{rccc} \mathbb{1}_A : & E & \longrightarrow & \{0,1\} \\ & x & \longmapsto & \begin{cases} 1 \text{ si } x \in A \\ 0 \text{ si } x \notin A \end{cases} \end{array}.$$

Image directe, image réciproque

Définition 3
Soit $f : E \longrightarrow F$ une fonction. Soit $A \subset E$.

- L'**image directe** $f(A)$ de A par f est l'ensemble des images par f des éléments de A :

$$f(A) = \{y \in F \mid \exists\, x \in A, y = f(x)\} = \{f(x) \mid x \in A\}.$$

- L'image de E par f est simplement appelée **image de** f. Elle est souvent notée $\mathrm{Im}(f)$ plutôt que $f(E)$.

$$y \in f(A) \Leftrightarrow \exists\, x \in A, y = f(x).$$

Si $x \in A$ alors $f(x) \in f(A)$ mais la réciproque est fausse : $f(x) \in f(A)$ n'implique pas $x \in A$ (voir le schéma ci-dessus).

Définition 4
Soit $f : E \longrightarrow F$ une fonction. Soit $B \subset F$. L'**image réciproque** de B par f est l'ensemble

$$f^{-1}(B) = \{x \in E \mid f(x) \in B\}.$$

$f^{-1}(B)$ correspond à l'ensemble des éléments de E qui sont envoyés dans B par la fonction f. C'est aussi l'ensemble des antécédents des éléments de B par f.

$$x \in f^{-1}(B) \Leftrightarrow f(x) \in B.$$

Proposition 5
Soit $f : E \longrightarrow F$ une fonction. Soient A et B des parties de F.

1. Si $A \subset B$ alors $f^{-1}(A) \subset f^{-1}(B)$.
2. $f^{-1}(A \cup B) = f^{-1}(A) \cup f^{-1}(B)$.
3. $f^{-1}(A \cap B) = f^{-1}(A) \cap f^{-1}(B)$.
4. $f^{-1}\left(\complement_E A\right) = \complement_E f^{-1}(A)$.

Proposition 6
Soit $f : E \longrightarrow F$ une fonction. Soient $A \subset E$ et $B \subset F$.

1. $A \subset f^{-1}\left(f(A)\right)$.
2. $f\left(f^{-1}(B)\right) \subset B$.

2.2 Injectivité, surjectivité et bijectivité

Injectivité

Définition 7
On dit qu'une fonction $f : E \longrightarrow F$ est **injective** si tout élément de l'ensemble d'arrivée F a au plus un antécédent dans E par f.
Dit autrement, $f : E \longrightarrow F$ est injective si pour tous x_1, $x_2 \in E$, on a

$$f(x_1) = f(x_2) \Longrightarrow x_1 = x_2.$$

Ou encore, $f : E \longrightarrow F$ est injective si pour tous $x_1 \neq x_2 \in E$, on a $f(x_1) \neq f(x_2)$.

Méthode 8 — *Pour montrer qu'une fonction f n'est* pas *injective, on montre qu'il existe deux éléments x_1 et x_2 de E* distincts *tels que $f(x_1) = f(x_2)$, en donnant explicitement les éléments x_1 et x_2.*

PROPOSITION 9
Soient $f : E \longrightarrow F$ et $g : F \longrightarrow G$ deux fonctions.

- Si f et g sont injectives alors $g \circ f$ est injective.
- Si $g \circ f$ est injective alors f est injective.

Surjectivité

DÉFINITION 10
On dit qu'une fonction $f : E \longrightarrow F$ est **surjective** si tout élément de l'ensemble d'arrivée F a au moins un antécédent dans E par f.
Dit autrement, $f : E \longrightarrow F$ est surjective si pour tout $y \in F$, il existe $x \in E$ tel que $f(x) = y$.
Ou encore, $f : E \longrightarrow F$ est surjective si $\mathrm{Im}(f) = F$.

REMARQUE 11 — *Soit $f : E \longrightarrow F$ une fonction. La fonction* $\widetilde{f} : \begin{array}{ccc} E & \longrightarrow & \mathrm{Im}(f) \\ x & \longmapsto & f(x) \end{array}$ *est surjective.*

MÉTHODE 12 — *Pour montrer qu'une fonction $f : E \longrightarrow F$ est surjective, on peut montrer que pour tout $y \in F$, l'équation $y = f(x)$ admet au moins une solution $x \in E$.*

PROPOSITION 13
Soient $f : E \longrightarrow F$ et $g : F \longrightarrow G$ deux fonctions.

- Si f et g sont surjectives alors $g \circ f$ est surjective.
- Si $g \circ f$ est surjective alors g est surjective.

Si $g \circ f$ est surjective, la fonction f n'est pas forcément surjective.

Bijectivité

DÉFINITION 14
On dit que $f : E \longrightarrow F$ est **bijective** si f est injective et surjective.
Ainsi, $f : E \longrightarrow F$ est bijective si tout élément de l'ensemble d'arrivée F a un unique antécédent dans E par f.
Dit autrement, $f : E \longrightarrow F$ est bijective si pour tout $y \in F$, il existe un unique $x \in E$ tel que $f(x) = y$.
Ou encore, $f : E \longrightarrow F$ est bijective si pour tout $y \in F$, l'équation $y = f(x)$ a une unique solution $x \in E$.

EXEMPLES 15

- *La fonction* $\mathrm{id}_E : E \longrightarrow E$ *est bijective.*
- *La fonction* $f : \begin{array}{ccc} \mathbb{R}_+ & \longrightarrow & \mathbb{R}_+ \\ x & \longmapsto & x^2 \end{array}$ *est injective et surjective, donc bijective.*

PROPOSITION-DÉFINITION 16
La fonction $f : E \longrightarrow F$ est bijective si et seulement s'il existe une fonction $g : F \longrightarrow E$ telle que $g \circ f = \mathrm{id}_E$ et $f \circ g = \mathrm{id}_F$.

Dans ce cas, la fonction g est unique. Elle est appelée **bijection réciproque** de f et est notée f^{-1}.

Il ne suffit pas que $g \circ f = \mathrm{id}_E$ ou que $f \circ g = \mathrm{id}_F$ pour que f soit bijective.

Si f est bijective, la bijection réciproque est la fonction qui à un élément y de F associe l'unique antécédent de y par f dans E, noté x :

$$f^{-1} : \begin{array}{ccc} F & \longrightarrow & E \\ y & \longmapsto & x \text{ tel que } y = f(x) \end{array} .$$

Si f est bijective : $\quad y = f(x) \Leftrightarrow x = f^{-1}(y)$

MÉTHODE 17 — *Pour déterminer si une fonction est bijective,*

- *Soit on donne directement l'expression d'une fonction g telle que $g \circ f = \mathrm{id}_E$ et $f \circ g = \mathrm{id}_F$, et dans ce cas, f est bijective, de bijection réciproque $f^{-1} = g$,*
- *Soit on résout, pour tout $y \in F$, l'équation $y = f(x)$ d'inconnue x. Si, pour tout $y \in F$, cette équation admet une unique solution x, alors f est bijective et on a également obtenu l'expression de f^{-1},*
- *Soit on montre que f est injective et surjective, mais dans ce cas on n'a pas l'expression de f^{-1}.*

Exemple 18 — *Montrons que la fonction* $\mathrm{sh} : \mathbb{R} \longrightarrow \mathbb{R} \; ; \; x \longmapsto \dfrac{\mathrm{e}^x - \mathrm{e}^{-x}}{2}$ *est bijective et déterminons sa bijection réciproque.*

Soit $y \in \mathbb{R}$. Résolvons l'équation $y = \mathrm{sh}(x)$ d'inconnue $x \in \mathbb{R}$. Pour tout $x \in \mathbb{R}$,

$y = \mathrm{sh}(x) = \dfrac{\mathrm{e}^x - \mathrm{e}^{-x}}{2}$ *si et seulement si* $(\mathrm{e}^x)^2 - 2y\mathrm{e}^x - 1 = 0$, *soit encore si et seulement si* e^x *est une racine strictement positive de* $X^2 - 2yX - 1$, *soit encore si et seulement si* $\mathrm{e}^x = y + \sqrt{y^2+1}$.

Donc, pour tout $x \in \mathbb{R}$, $y = \mathrm{sh}(x)$ si et seulement si $x = \ln(y + \sqrt{y^2+1})$.

Ainsi, pour tout $y \in \mathbb{R}$, l'équation $y = \mathrm{sh}(x)$ admet une unique solution x sur $\mathbb{R}$, $x = \ln(y + \sqrt{y^2+1})$.

La fonction sh *est donc bijective, de bijection réciproque* $\mathrm{sh}^{-1} : \mathbb{R} \longrightarrow \mathbb{R} \; ; \; y \longmapsto \ln\left(y + \sqrt{y^2+1}\right)$.

Proposition 19
Soient $f : E \longrightarrow F$ et $g : F \longrightarrow G$ deux fonctions.

- Si f est bijective alors f^{-1} est bijective et $\left(f^{-1}\right)^{-1} = f$.
- Si f et g sont bijectives alors $g \circ f$ est bijective et $(g \circ f)^{-1} = f^{-1} \circ g^{-1}$.

Exemples 20

- *La fonction* $\begin{array}{rccc} f : & \mathbb{R}^* & \longrightarrow & \mathbb{R}^* \\ & x & \longmapsto & \dfrac{1}{x} \end{array}$ *est une involution et est donc bijective, de bijection réciproque elle-même.*
- *La fonction* $\begin{array}{rccc} f : & \mathcal{P}(E) & \longrightarrow & \mathcal{P}(E) \\ & A & \longmapsto & \complement_E A \end{array}$ *est une involution et est donc bijective, de bijection réciproque elle-même.*

Chapitre 3 Relations binaires - Relations d'équivalences

Table des matières du chapitre

3.1 Relations binaires

Définition 1
Soient E et F deux ensembles. Soit G une patrie de $E \times F$.
On appelle **relation binaire** le triplet $\mathcal{R} = (E, F, G)$.
Si $(x, y) \in G$, on dit que x **est en relation avec** y, et on le note $x\mathcal{R}y$.
On parle de la relation $\mathcal{R}$.

 L'ordre dans un couple est important. On peut avoir $x\mathcal{R}y$ mais pas $y\mathcal{R}x$.

Dans la suite du chapitre, on s'intéresse aux relations sur E et E. Ce sont les relations les plus utiles en mathématiques.
Lorsque $E = F$, la relation $\mathcal{R}$ est appelée **relation binaire sur** E.

On note souvent une relation $\mathcal{R}$ avec un symbole $\equiv$, $\leq$, $\sim$, $\subset$...

Exemples 2 *Donnons quelques exemples de relations binaires sur un ensemble :*

- *la relation d'égalité $=$ sur E,*
- *les relations d'inégalité $\leq$, $<$ sur $\mathbb{R}$, $\mathbb{Q}$, $\mathbb{Z}$ ou $\mathbb{N}$,*
- *la relation d'inclusion $\subset$ sur $\mathcal{P}(E)$ (ensemble des parties de E),*
- *la relation de comparaison $\leq$ sur $Fonct(E, \mathbb{R})$ (fonctions de E dans $\mathbb{R}$), définie par $f \leq g$ si $f(x) \leq g(x)$ $\forall x \in E$,*
- *la relation de divisibilité $|$ sur $\mathbb{Z}$, définie par $m \mid n$ s'il existe $k \in \mathbb{Z}$ tel que $n = mk$.*
- *pour tout $n \in \mathbb{N}$, la relation de congruence modulo n sur $\mathbb{Z}$, notée $\equiv \mod n$, définie par $a \equiv b \mod n$ s'il existe $k \in \mathbb{Z}$ tel que $a = b + kn$.*
- *pour tout $\alpha \in \mathbb{R}$, la relation de congruence modulo α sur $\mathbb{R}$, notée $\equiv \mod \alpha$, définie par $a \equiv b \mod \alpha$ s'il existe $k \in \mathbb{Z}$ tel que $a = b + k\alpha$.*
- *la relation « avoir le même signe » sur $\mathbb{R}^*$.*

On définit des propriétés intéressantes pour les relations.

Définition 3
Soit E un ensemble. Soit $\mathcal{R}$ une relation binaire sur E.

- $\mathcal{R}$ est dite **réflexive** si : pour tout $x \in E$, on a $x\mathcal{R}x$,
- $\mathcal{R}$ est dite **symétrique** si : pour tout $(x, y) \in E^2$, si $x\mathcal{R}y$ alors $y\mathcal{R}x$,
- $\mathcal{R}$ est dite **antisymétrique** si : pour tout $(x, y) \in E^2$, si $x\mathcal{R}y$ et $y\mathcal{R}x$ alors $x = y$,
- $\mathcal{R}$ est dite **transitive** si : pour tout $(x, y, z) \in E^3$, si $x\mathcal{R}y$ et $y\mathcal{R}z$ alors $x\mathcal{R}z$.

Exemples 4

- *La relation d'égalité $=$ sur E est réflexive, symétrique, antisymétrique et transitive.*
  On remarque qu'une relation peut donc être symétrique et antisymétrique.
- *La relation $\leq$ sur $\mathbb{R}$ est réflexive, antisymétrique et transitive.*
  Elle n'est pas symétrique car par exemple $2 \leq 3$ mais $3 \not\leq 2$.
- *La relation $<$ sur $\mathbb{R}$ est symétrique et transitive.*
  Elle n'est ni réflexive, ni antisymétrique. Par exemple, $1 \not< 1$.

- *La relation de divisibilité sur $\mathbb{Z}$ est réflexive et transitive.*
  Elle n'est ni symétrique, ni antisymétrique. En effet, on a $1|2$ mais $2 \nmid 1$, et $1|-1$ et $-1|1$ mais $1 \neq -1$.
- *La relation d'inclusion $\subset$ sur $\mathcal{P}(E)$ est réflexive, antisymétrique et transitive.*
  Elle n'est pas symétrique car par exemple $\{1\} \subset \{1,2\}$ mais $\{1,2\} \not\subset \{1\}$.
- *La relation « avoir le même signe » sur $\mathbb{R}^*$ est réflexive, symétrique et transitive.*
  Elle n'est pas antisymétrique car par exemple, $1\mathcal{R}2$ et $2\mathcal{R}1$ mais $1 \neq 2$.

Nous allons étudier deux grandes familles de relations.

3.2 Relations d'équivalence, classes d'équivalence

La première famille est celle des relations d'équivalence.
Une telle relation permet d'identifier des éléments de E qui sont en relation muuelle, pour regarder les ensembles d'éléments qui sont en relation. Par exemple, la nationalité.

Définition et exemples

Définition 5
Soit E un ensemble et $\mathcal{R}$ une relation binaire sur E.
On dit que $\mathcal{R}$ est une **relation d'équivalence** sur E si $\mathcal{R}$ est réflexive, symétrique et transitive.

Une relation d'équivalence est souvent notée $\equiv$ ou $\sim$. On parle d'éléments de E "équivalents", "semblables".

Exemples 6

- *La relation d'égalité $=$ sur E est une relation d'équivalence.*
- *La relation « avoir le même signe » sur $\mathbb{R}^*$ est une relation d'équivalence.*
- *Pour tout $n \in \mathbb{N}$, la relation de congruence modulo n sur $\mathbb{Z}$ est une relation d'équivalence.*
- *Pour tout $\alpha \in \mathbb{R}$, la relation de congruence modulo α sur $\mathbb{R}$ est une relation d'équivalence.*
- *Si $(A_i)_{i\in I}$ est une partition de E ($A_i \cap A_j = \emptyset$ si $i \neq j$, et $\cup_{i\in I} A_i = E$), la relation d'appartenance à un sous-ensemble A_i est une relation d'équivalence. ($x\mathcal{R}y$ si $\exists i \in I$ tel que $x, y \in A_i$)*

Classes d'équivalence et ensemble quotient

Définition 7
Soient E un ensemble et $\mathcal{R}$ une relation d'équivalence sur E. Soit x un élément de E.
On appelle **classe d'équivalence** de x pour la relation $\mathcal{R}$ (ou plus simplement classe de x) l'ensemble des éléments y de E qui sont en relation avec x :

$$\mathrm{Cl}(x) = \{y \in E \text{ tels que } x\mathcal{R}y\}.$$

On la note $\mathrm{Cl}(x)$ ou $\overline{x}$.

Exemples 8

- *La classe d'équivalence de 1 pour la relation d'équivalence « avoir le même signe » sur $\mathbb{R}^*$ est l'ensemble des nombres réels non nuls de même signe que 1, c'est-à-dire l'ensemble des nombres réels strictement positifs :* $\mathrm{Cl}(1) = \mathbb{R}_+^*$.
- *Soit $n \in \mathbb{N}$. Soit $r \in \mathbb{Z}$. La classe d'équivalence de r pour la relation de congruence modulo n dans $\mathbb{Z}$ est*

$$\begin{aligned}\mathrm{Cl}(r) &= \{p \in \mathbb{Z} \mid p \equiv r \bmod n\}\\ &= \{p \in \mathbb{Z} \mid \exists k \in \mathbb{Z}, p = r + kn\}\\ &= \{r + kn \mid k \in \mathbb{Z}\}\\ &= n\mathbb{Z} + r\end{aligned}$$

Proposition 9
Soient E un ensemble et $\mathcal{R}$ une relation d'équivalence sur E.
Pour tout $(x, y) \in E^2$, on a $\mathrm{Cl}(x) = \mathrm{Cl}(y)$ si et seulement si $x\mathcal{R}y$.

Notons donc que si $y \in \mathrm{Cl}(x)$ alors $\mathrm{Cl}(y) = \mathrm{Cl}(x)$.

Définition 10
Soient E un ensemble et $\mathcal{R}$ une relation d'équivalence sur E. Soit C une classe d'équivalence pour $\mathcal{R}$.
On appelle **représentant** de la classe d'équivalence C tout élément x de C.
On a alors $C = Cl(x)$.

Exemple 11 — *Nous avons vu que $\mathbb{R}_+^*$ est une classe d'équivalence (celle de 1) pour la relation d'équivalence « avoir le même signe » sur $\mathbb{R}^*$.*
Tout élément de $\mathbb{R}_+^*$ est un représentant de cette classe. Des représentants de cette classe sont donc par exemple 1, ou π, ou $\sqrt{2}$... Ainsi, $\mathbb{R}_+^* = \mathrm{Cl}(1) = \mathrm{Cl}(\pi) = \mathrm{Cl}(\sqrt{2})$...

Proposition 12
Soient E un ensemble et $\mathcal{R}$ une relation d'équivalence sur E.
L'ensemble des classes d'équivalence de E forme une partition de E, c'est-à-dire :

- Elles sont non vides,
- Elle sont deux à deux disjointes,
- Leur réunion est égale à l'ensemble E.

Remarque 13 — *Ainsi, toute relation d'équivalence sur E donne une partition de E, et toute partition de E donne une relation d'équivalence sur E.*
Chaque relation d'équivalence sur E correspond à une partition de E. (penser par exemple au pays de naissance, cela partionne l'ensemble des êtres humains)

Exemples 14

- *La relation « avoir le même signe » sur $\mathbb{R}^*$ a exactement deux classes d'équivalence : $\mathbb{R}_+^*$ et $\mathbb{R}_-^*$. Ces deux classes d'équivalence forment bien une partition de $\mathbb{R}^*$.*
- *Pour tout $n \in \mathbb{N}$, la relation de congruence modulo n sur $\mathbb{Z}$ possède exactement n classes d'équivalence : les ensembles $n\mathbb{Z} + r = \{nk + r \mid k \in \mathbb{Z}\}$ avec $r \in \{0, \dots, n-1\}$. On les note souvent $\overline{0}$, $\overline{1}$, ..., $\overline{n-1}$.*
On a choisi comme représentant des différentes classes les entiers 0, 1, ..., $n-1$.

Définition 15
Soient E un ensemble et $\mathcal{R}$ une relation d'équivalence sur E.
L'ensemble des classes d'équivalence de E pour la relation $\mathcal{R}$ s'appelle **l'ensemble quotient de E par $\mathcal{R}$**.
On le note $E/\mathcal{R}$. C'est un sous-ensemble de $\mathcal{P}(E)$ (ensemble des parties de E).

Exemples 16

- *L'ensemble quotient de $\mathbb{R}^*$ par la relation « avoir le même signe » est l'ensemble $\{\mathbb{R}_-^*, \mathbb{R}_+^*\}$.*
- *L'ensemble quotient de E par la relation d'égalité $=$ est l'ensemble $\{\{x\}, x \in E\}$.*
- *Soit $n \in \mathbb{Z}$. L'ensemble quotient de $\mathbb{Z}$ par la relation de congruence modulo n est l'ensemble $\{n\mathbb{Z}, n\mathbb{Z} + 1, \dots, n\mathbb{Z} + n - 1\} = \{\overline{0}, \overline{1}, \dots, \overline{n-1}\}$. On note cet ensemble $\mathbb{Z}/n\mathbb{Z}$.*
- *Pour $f : E \to F$, la relation « avoir la même image par f » est une relation d'équivalence. L'ensemble quotient de E par cette relation d'équivalence est $\{f^{-1}(\{y\}), y \in Im(f)\}$. (A vérifier.)*

Les relations d'équivalence servent à étudier les éléments d'un ensemble E de façon simplifiée.
Par exemple, la congruence modulo n sert à étudier les entiers de $\mathbb{Z}$ seulement selon le reste de leur division euclidienne par n. La relation « signe » sur $\mathbb{R}$ étudie les nombres réels selon leur signe.
La relation « avoir le même rang » chez les matrices $n \times p$ étudie les matrices selon leur rang (ou les familles de p vecteurs de $\mathbb{K}^n$ selon leur rang).

Chapitre 4 Structures algébriques : Groupes, Anneaux, Corps

Table des matières du chapitre

4.1 Groupes

Dans tout ce chapitre, si cela n'est pas précisé, E désigne un ensemble.

Définition 1
Soit E un ensemble.
On appelle **loi de composition interne** (ou opération) sur E toute fonction $\varphi : E \times E \to E$. L'image $\varphi(a, b)$ (écriture préfixe) sera souvent notée $a * b$ (écriture infixe), et la loi de composition sera appelée $*$.

Exemples 2— *Les ensembles et opérations suivants :* $(\mathbb{Z}, +)$, $(\mathcal{P}(E), \cap)$, $(\mathbb{R}^3, \wedge)$, *où* $\wedge$ *est le produit vectoriel, sont des ensembles munis d'une loi de composition interne.*

Définition 3
Soit $(E, *)$ un ensemble muni d'une loi de composition interne $*$.
Une partie $F \subset E$ de E est dite **stable par** $*$ si l'on a $x * y \in F$ pour tous $x, y \in F$.

La restriction de $*$ à $F \times F$ est alors appelée loi induite sur F.

Définition 4
Soit $(G, \star)$ un ensemble muni d'une loi de composition interne.
On dit que $(G, \star)$ est un **groupe** si la loi $\star$ vérifie les propriétés suivantes :

1. La loi $\star$ est **associative** :
$$\forall x, y, z \in G, \text{ on a } x * (y * z) = (x * y) * z\,;$$
2. L'ensemble G possède un **élément neutre** pour la loi $\star$:
$$\exists e \in G, \text{ tel que } \forall x \in G,\ e * x = x * e = x\,;$$
3. tout élément est inversible :
$$\forall x \in G,\ \exists y \in G \text{ tel que } x * y = y * x = e.$$

On appelle y le **symétrique** de x que l'on notera parfois y^{-1} (l'inverse) ou $-y$ (l'opposé).
De plus, le groupe $(G, \star)$ est **commutatif** (ou **abélien**) si la loi est commutative : $\forall x, y \in G,\ x * y = y * x$.

Exemple 5 — *Nous allons d'abord donner quelques exemples que nous avons déjà rencontrés.*

1. *Les ensembles $\mathbb{Z}$, $\mathbb{Q}$, $\mathbb{R}$ et $\mathbb{C}$ munis de la loi d'addition $+$ sont des groupes, qui sont abéliens.*
2. *L'ensemble $\mathbb{N}$ muni de l'addition n'est pas un groupe. (Pourquoi ?)*
3. *Les ensembles de matrices $\mathscr{M}_{n,p}(\mathbb{K})$ munis de la loi d'addition sont des groupes, qui sont abéliens.*
4. *Les ensembles $\mathbb{K}[X]$ munis de la loi d'addition sont des groupes, qui sont abéliens.*
5. *Les espaces vectoriels E munis de leur loi d'addition de vecteurs sont des groupes, qui sont commutatifs.*
6. *Les ensembles $\mathbb{Q}$, $\mathbb{R}$ et $\mathbb{C}$ munis de la multiplication ne sont pas des groupes. (Pourquoi ?)*
   Mais $\mathbb{Q}^*$, $\mathbb{R}^*$ et $\mathbb{C}^*$ munis de la multiplication le sont, et sont commutatifs.
7. *Les ensembles de matrices inversibles $GL_n(\mathbb{K})$ munis de la multiplication matricielle sont des groupes. Si $n \geq 2$ ces groupes ne sont pas abéliens.*
   Par exemple $A = \begin{pmatrix} 1 & 1 \\ 0 & 1 \end{pmatrix}$ et $B = \begin{pmatrix} 1 & 0 \\ 1 & 1 \end{pmatrix}$ sont des matrices inversibles, mais $AB \neq BA$.

Remarque 6 — *Soit $(G, \star)$ un groupe. Alors G possède un unique élément neutre.*

En effet, pour e, e' deux éléments neutres de G, on a $e \star e' = e'$ et $e \star e' = e$, donc $e = e'$.

Définition 7
Soit $(G, \star)$ un groupe. Pour tout élément $x \in G$ et tout entier $n \in \mathbb{Z}^*$, on notera :

- x^{-1} le **symétrique** de x (ou **l'inverse** de x) ;
- $x^0 := e$;
- $x^n := \underbrace{x \star \ldots \star x}_{n \text{ fois}}$ si n est strictement positif ;
- $x^n := \underbrace{x^{-1} \star \ldots \star x^{-1}}_{-n \text{ fois}}$ si n est strictement négatif.

Remarque 8 — *Avec cette notation, pour tout $x \in G$ et tous $n, m \in \mathbb{Z}$ et $k \in \mathbb{N}^*$, on a :*

$$x^n \star x^m = x^{n+m} \quad \text{et} \quad (x^n)^k = \underbrace{x^n \star \ldots \star x^n}_{k \text{ fois}} = x^{kn}.$$

Proposition 9
Soit $(G, \star)$ un groupe. Soient $a, b \in G$.
Alors, on a $(a \star b)^{-1} = b^{-1} \star a^{-1}$.

Preuve — On a $(a \star b) \star (b^{-1} \star a^{-1}) = e_G$, et $(b^{-1} \star a^{-1}) \star (a \star b) = e_G$. □

Si les éléments a et b ne commutent pas, alors $(a \star b)^{-1} = b^{-1} \star a^{-1} \neq a^{-1} \star b^{-1}$.

Proposition 10
Soient $(G_1, \star_1)$ et $(G_2, \star_2)$ deux groupes.

L'ensemble $G = G_1 \times G_2$ muni de la loi produit :

$$\begin{array}{rccc} \star & G \times G & \to & G \\ & ((g_1, g_2), (h_1, h_2)) & \mapsto & (g_1 \star_1 h_1, g_2 \star_2 h_2) \end{array}$$

est lui aussi un groupe.
On l'appelle **produit direct** des groupes G_1 et G_2.

4.1.1 Sous-groupes et ordre d'un élément

Définition 11
Soit $(G, \star)$ un groupe. Soit H un sous-ensemble de G.
On dit que H est un **sous-groupe** de $(G, \star)$ s'il vérifie les propriétés suivantes :

1. L'élément neutre e appartient à H ;
2. Pour tous x, y appartenant à H, le produit $x \star y$ appartient à H (autrement dit H est stable par la loi $\star$) ;
3. Pour tout x appartenant à H, l'inverse de x appartient à H.

Si H est un sous-groupe de G, on notera $H < G$.

Proposition 12
Soient $(G, \star)$ un groupe et H un sous-groupe de G.
Alors l'ensemble H muni de la loi $\star$ restreinte à H est un groupe.

La proposition suivante donne un critère simple pour montrer qu'un sous-ensemble est un sous-groupe.

PROPOSITION 13
Soient $(G, \star)$ un groupe et H un sous-ensemble de G.
Les propriétés suivantes sont équivalentes :

1. H est un sous-groupe ;
2. H est non vide, et pour tous x, y appartenant à H, $x \star y^{-1}$ appartient à H.

EXEMPLE 14 —

1. *Soit $(G, \star)$ un groupe et e_G son élément neutre. L'ensemble $\{e_G\}$ est un sous-groupe de G. On l'appelle le sous-groupe trivial de G.*
2. *Soit $n \in \mathbb{Z}$ un entier. L'ensemble $n\mathbb{Z} = \{nk, k \in \mathbb{Z}\}$ des multiples de n est un sous-groupe de $(\mathbb{Z}, +)$.*
3. *L'ensemble $\mathbb{U} = \{z \in \mathbb{C},\ |z| = 1\}$ est un sous-groupe de $(\mathbb{C}^*, \times)$.*
   Pour $n > 0$, l'ensemble $\mathbb{U}_n = \{z \in \mathbb{C},\ z^n = 1\}$ est un sous-groupe de $\mathbb{U}$ (Pourquoi ?).
4. *Soit E un ensemble et A une partie de E. L'ensemble des bijections $f : E \to E$ telles que $f(x) = x$, $\forall x \in A$, est un sous-groupe de $(Bij(E), \circ)$.*

REMARQUE 15 —
• *Soit $(G, \star)$ un groupe et H un sous-groupe de G. Soit $x \in H$. Alors H contient tous les x^n, $n \in \mathbb{Z}$.*
• *Pout G un groupe, et X, Y des sous-groupes de G, la relation : $X < Y$ si X est un sous-groupe de Y, est une relation d'ordre. Cette relation d'ordre n'est pas totale en général.*

Il faut bien penser à vérifier que le sous-ensemble H est non-vide pour montrer que c'est un sous-groupe. L'ensemble vide $\emptyset$ est un sous-ensemble de G, mais pas un sous-groupe.

PROPOSITION 16 (Intersection de sous-groupes)
Soient $(G, \star)$ un groupe et $(H_i)_{i \in I}$ une famille de sous-groupes de G.
Alors le sous-ensemble $\cap_{i \in I} H_i$ est un sous-groupe de G.

PROPOSITION 17
Les sous-groupes de $(\mathbb{Z}, +)$ sont de la forme $n\mathbb{Z}$, avec $n \in \mathbb{N}$ uniquement déterminé.

Sous-groupe engendré par une partie

PROPOSITION-DÉFINITION 18
Soient $(G, \star)$ un groupe et S une partie de G.
On appelle sous-groupe **engendré** par S le plus petit sous-groupe contenant S.
On note $\langle S \rangle$ ce sous-groupe. Il est caractérisé par la relation :

$$\langle S \rangle = \bigcap_{S \subset H < G} H.$$

Lorsque S est un singleton $\{x\}$, on note $\langle x \rangle$ le groupe engendré par $\{x\}$.
On l'appelle aussi le sous-groupe engendré par x.

REMARQUE 19 — *Soit $(G, \star)$ un groupe et S une partie de G.*
L'ensemble A de tous les produits $a_1 \star \ldots \star a_n$, avec $a_i \in S$ ou $a_i^{-1} \in S$ pour tout $i \in [\![1, n]\!]$, est un sous-groupe de G .

En particulier, pour $x \in G$ le sous-groupe engendré par x est

$$\langle x \rangle = \{x^n | n \in \mathbb{Z}\}.$$

DÉFINITION 20
Soit $(G, \star)$ un groupe.
Le groupe G est **monogène** s'il existe $a \in G$ tel que $G =< a >$.
L'élément a est appelé un **générateur** de G.

Si G est monogène et fini, on dit que G est un groupe **cyclique** .

Exemple 21 —

1. *Tout sous-groupe de $\mathbb{Z}$ est de la forme $n\mathbb{Z}$ et est donc monogène.*
2. *Pour $n \geq 1$, l'ensemble $\mathcal{U}_n$ des racines n-ièmes de l'unité muni de la multiplication est un groupe.*
   Il est cyclique, et $\chi = \exp\left(\frac{2i\pi}{n}\right)$ est un générateur du groupe.
   Le groupe $(\mathcal{U}_n, \times)$ est un groupe à n éléments. Il existe ainsi des groupes finis de toutes les tailles possibles.

Ordre d'un élément

Définition 22
Soit $(G, \star)$ un groupe fini.
On appelle **ordre** de G le nombre d'éléments de G (c'est le cardinal de G).
On le note $|G|$ (ou $\text{Card}(G)$).

Définition 23
Soient $(G, \star)$ un groupe et $a \in G$.

Si le sous-groupe $\langle a \rangle$ est fini, alors l'**ordre** de a, noté $\text{ord}(a)$, est le cardinal de ce sous-groupe.
Sinon, on dit que a est d'ordre infini.

Exemple 24 — *Dans $(\mathbb{C}^*, \times)$, i est un élément d'ordre 4. En effet, on a :*

$$i^0 = 1, \quad i^1 = i, \quad i^2 = -1, \quad i^3 = -i, \quad i^4 = 1 = i^0,$$

donc $\langle i \rangle = \{1, i, -1, -i\}$.

Proposition 25
Soient $(G, \star)$ est un groupe et $x \in G$.
L'élement x est d'ordre fini si et seulement s'il existe $k \in \mathbb{N}^*$ tel que $x^k = e$.
Dans ce cas, $\text{ord}(x)$ est le plus petit entier n strictement positif tel que $x^n = e$.

Remarque 26 — *En utilisant l'idée de la division euclidienne, on montre que si $x \in G$ est d'ordre fini n, alors on a $x^k = e$ si et seulement si n divise k.*

De plus, tout groupe $(G, \star)$ possède un unique élément d'ordre 1 : son élément neutre e_G.

Remarque 27 — *Les groupes pouvant se définir de façon très formelle, comment classifier les groupes, comment les identifier ?*

Morphismes de groupes, Isomorphismes

Avec des ensembles, on regarde des fonctions définies sur ces ensembles.
Pour les espaces vectoriels on regarde les applications linéaires, les fonctions sur des e.v. qui préservent la structure d'espace vectoriel. Pour les groupes, nous allons regarder les fonctions qui préservent la structure de groupe.

Définition 28
Soit $(G, *)$ et $(H, \triangle)$ deux groupes.
Une fonction $\varphi : G \to H$ est un **morphisme de groupes** si l'on a :

$$\forall x, y \in G, \ \varphi(x * y) = \varphi(x) \triangle \varphi(y).$$

De plus, on dit que φ est :

1. un endomorphisme de groupes si $G = H$.
2. un **isomorphisme** de groupes si φ est bijective.
3. un automorphisme de groupes si φ est un endomorphisme bijectif.

Exemple 29 —

1. *Pour $(G, \star_1)$ et $(H, \star_2)$ deux groupes, la fonction $\varphi : x \in G \mapsto e_H \in H$ est toujours un morphisme de groupes.*
   On l'appelle ***morphisme trivial*** *de G vers H.*
2. *L'application $n \mapsto 2n$ est-elle un morphisme de groupes sur $(\mathbb{Z}, +)$?*
3. *Soit $(G, \star)$ un groupe et $h \in G$ fixé.*
   La fonction $\varphi : g \in G \mapsto hg \in G$ est-elle un morphisme de groupes ? Est-elle bijective ?
4. *Soit $(G, \star)$ un groupe et $h \in G$ fixé.*
   La fonction $\varphi : g \in G \mapsto hgh^{-1} \in G$ est-elle un morphisme de groupes ? Est-elle bijective ?
5. *La fonction $\theta \to e^{i\theta}$ est un morphisme de groupes de $(\mathbb{R}, +)$ dans $(\mathbb{C}^*, \times)$. Est-elle bijective ?*

Les morphismes de groupes sont les fonctions $f : G \to H$ qui préservent la structure de groupe de G vers H (autrement dit, les fonctions qui sont compatibles avec les lois de ces groupes).

Proposition 30
Soient $(G, *)$, $(H, \triangle)$ des groupes. Soit $\varphi : G \to H$ un morphisme de groupes. Alors :

1. On a $\varphi(e_G) = e_H$ et $\varphi(x^{-1}) = \varphi(x)^{-1}$ pour tout $x \in G$;
2. Si G' est un sous-groupe de G, alors $\varphi(G') = \{\varphi(x),\ x \in G'\}$ est un sous-groupe de H ;
3. Si H' est un sous-groupe de H, alors $\varphi^{-1}(H') = \{x \in G$,tels que $\varphi(x) \in H'\}$ est un sous-groupe de G.

Définition 31
Soient $(G, *)$, $(H, \triangle)$ des groupes. Soit $\varphi : G \to H$ un morphisme de groupes.
On appelle **noyau** de φ l'image réciproque de e_H.
On le note $\mathrm{Ker}(\varphi) = \varphi^{-1}(\{e_H\})$ (de l'allemand "*Ker*nel" = "noyau").

Le noyau de φ, $\mathrm{Ker}(\varphi)$, est un sous-groupe de G.

Proposition 32
Soient $(G, *)$, $(H, \triangle)$ des groupes et $\varphi : G \to H$ un morphisme de groupes.
Le morphisme φ est injectif si et seulement si $\ker \varphi = \{e_G\}$.

Preuve — Si la fonction φ est injective on a bien $\mathrm{Ker}\varphi = \{e_G\}$ puisque $\varphi(e_G) = e_H$.

Réciproquement, supposons que $\mathrm{Ker}\varphi = \{e_G\}$. Soient $x, x' \in G$ tels que $\varphi(x) = \varphi(x')$. On a alors alors $\varphi(x)\triangle\varphi(x')^{-1} = e_H$, donc $\varphi(x * x'^{-1}) = e_H$. Cela donne $x * x'^{-1} = e_G$, c'est-à-dire $x = x'$. Donc φ est injectif. □

Exemple 33 — *La fonction $\varphi : z \in (\mathbb{C}^*, \times) \mapsto |z| \in (\mathbb{R}_+^*, \times)$ est un morphisme de groupes, et* $\mathrm{Ker}(\varphi) = \varphi^{-1}(\{1\}) = \mathbb{U}$. *Ainsi, $(\mathbb{U}, \times)$ est un sous-groupe de $(\mathbb{C}, \times)$ et φ n'est pas injectif.*

Isomorphismes de groupes

Définition 34
Soient $(G, *)$, $(H, \triangle)$ deux groupes.
On dit que G et H sont **isomorphes** s'il existe un isomorphisme de groupes φ entre G et H.
On le note $G \simeq H$.

Proposition 35
Soient $(G, *)$, $(G', \triangle)$ deux groupes et $\varphi : G \to G'$ un isomorphisme de groupes.
Alors $\varphi^{-1} : G' \to G$ est encore un isomorphisme de groupes.

Remarque 36 — *Soient G, H deux groupes isomorphes (via $\varphi : G \to H$).*

- *Alors G et H ont même cardinal.*
- *Tout élément $x \in G$ d'ordre fini n est envoyé sur un élément $\varphi(x) \in H$ d'ordre fini n.*
- *On a une bijection entre les solutions de l'équation $x^n = e_G$ et celles de $y^n = e_H$.*
  Plus généralement, on a une bijection entre les solutions de $x^n = g$ et celles de $y^n = \varphi(g)$.
- *De même, G est abélien si et seulement si H est abélien.*

Ces résultats sont utiles pour montrer que deux groupes ne sont pas isomorphes.
Cela permet aussi de voir quels éléments de la "structure" d'un groupe sont totalement identiques entre G et H quand $G \simeq H$.

4.1.2 Le groupe $\mathbb{Z}/n\mathbb{Z}$

Relation de congruence

Définition 37
Soit $\varepsilon \in \mathbb{R}$. On définit sur $\mathbb{R}$ la relation de **congruence modulo** ε par :

$$x \equiv y \bmod \varepsilon \iff x - y \in \varepsilon\mathbb{Z} = \{\varepsilon k,\, k \in \mathbb{Z}\}.$$

Proposition 38
Pour $x \in \mathbb{R}$, l'ensemble des éléments y tels que $x \equiv y \bmod \varepsilon$ est :

$$\bar{x} = x + \varepsilon\mathbb{Z} = \{x + \varepsilon k,\, k \in \mathbb{Z}\}.$$

Si $\varepsilon \neq 0$, alors pour tout $x \in \mathbb{R}$ il existe un unique $y \in [0, \varepsilon[$ tel que $x \equiv y \bmod \varepsilon$. C'est-à-dire :

$$\exists!(y,k) \in [0,\varepsilon[\times\mathbb{Z},\ x = y + k\varepsilon.$$

Proposition 39
Tout sous-groupe de $(\mathbb{R}, +)$ est soit dense dans $\mathbb{R}$, soit de la forme $a\mathbb{Z}$, $a \in \mathbb{R}^+$.

Définition 40
Soit $n \in \mathbb{N}$. Pour $p \in \mathbb{Z}$, on note $\overline{p} := p + n\mathbb{Z} = \{p + nk,\, k \in \mathbb{Z}\}$ l'ensemble des entiers congrus à p modulo n.
Ce sont les **classes d'équivalences** pour la relation de congruence modulo n.
L'entier p est ainsi un **représentant** de la classe d'équivalence $\overline{p}$.

On définit $\mathbb{Z}/n\mathbb{Z} = \{\overline{0}, \overline{1}, \cdots, \overline{n-1}\}$ l'ensemble des classes d'équivalence modulo n.
$\mathbb{Z}/n\mathbb{Z}$ est l'ensemble quotient de la relation d'équivalence de congruence modulo n. (voir chapitre Relations)

Proposition-Définition 41 (Opérations sur $\mathbb{Z}/n\mathbb{Z}$)
Soit $n \in \mathbb{N}$. Pour tous $p, q \in \mathbb{Z}$, on a :

$$\overline{p} + \overline{q} = \{a + b,\, a \in \overline{p}, b \in \overline{p}\} = \overline{p+q},$$
$$\overline{p} \times \overline{q} = \{ab,\, a \in \overline{p}, b \in \overline{p}\} = \overline{pq},$$

les opérations d'addition et de multiplication étant ici celles définies pour des sous-ensembles de $\mathbb{Z}$.
Ces opérations définissent ainsi des lois de composition internes sur $\mathbb{Z}/n\mathbb{Z}$, notées $\overline{+}$ et $\overline{\times}$.
Ces lois de composition sont associatives et commutatives.
La loi $\overline{+}$ admet pour élément neutre $\overline{0}$ et la loi $\overline{\times}$ admet pour élément neutre $\overline{1}$.
Enfin, $\left(\mathbb{Z}/n\mathbb{Z}, \overline{+}\right)$ est un groupe.

Remarque 42 —

1. *Soit $\overline{a} \in \mathbb{Z}/n\mathbb{Z}$ et $k \in \mathbb{N}$, on note $k\overline{a} = \overline{a} \overline{+} \cdots \overline{+} \overline{a}$, la somme de k copies de a. On a $k\overline{a} = \overline{ka}$.*
2. *Le symétrique pour la loi $\overline{+}$ de $\overline{p}$ est $\overline{-p}$.*
3. *Le groupe $(\mathbb{Z}/n\mathbb{Z}, \overline{+})$ est un groupe de cardinal n, qui est commutatif. ($\overline{p} + \overline{q} = \overline{q} + \overline{p}$)*
4. *La fonction $k \in \mathbb{Z} \mapsto \overline{k} \in \mathbb{Z}/n\mathbb{Z}$, est un morphisme de groupes surjectif, de noyau $n\mathbb{Z}$.*
5. *Pour $n = 1$, on a $\mathbb{Z}/1\mathbb{Z} = \{\overline{0}\}$. Ce groupe ne contient donc que son élément neutre.*
   *On appelle un tel groupe le **groupe trivial**. (le groupe à 1 élément)*

Proposition 43
Soit $n \in \mathbb{N}^*$. Le groupe $\left(\mathbb{Z}/n\mathbb{Z}, \overline{+}\right)$ possède des générateurs.
Ce sont exactement les classes $\overline{m}$, avec $\text{pgcd}(m, n) = 1$. (on a $\mathbb{Z}/n\mathbb{Z} = \langle\overline{m}\rangle$)

Définition 44
Soit $n \geq 1$. On définit $\varphi(n)$ le nombre de générateurs du groupe $\mathbb{Z}/n\mathbb{Z}$.
L'entier $\varphi(n)$ est appelé **l'indicatrice d'Euler** de n.
On a donc $\varphi(n) = Card(\{1 \leq m \leq n-1 \text{ tels que } \text{pgcd}(m,n) = 1\})$.

Théorème 45 (Théorème des restes chinois)
Soient $m_1, m_2 \in \mathbb{N}^*$ premiers entre eux. Soient $a, b \in \mathbb{Z}$.
Alors le système de congruences d'inconnue $x \in \mathbb{Z}$:

$$\begin{cases} x \equiv a \bmod m_1 \\ x \equiv b \bmod m_2 \end{cases}$$

admet une solution dans $\mathbb{Z}$.
De plus, si $x_0 \in \mathbb{Z}$ est une solution, alors l'ensemble des solutions dans $\mathbb{Z}$ est $x_0 + m_1 m_2 \mathbb{Z} = \{x_0 + m_1 m_2 k \mid k \in \mathbb{Z}\}$.
Autrement dit, pour x_0 une solution du système d'équations, on a :

$$\begin{cases} x \equiv a \bmod m_1 \\ x \equiv b \bmod m_2 \end{cases} \Leftrightarrow \Big\{ x \equiv x_0 \bmod m_1 m_2$$

THÉORÈME 46 (Théorème d'isomorphisme chinois)
Soient $m_1, m_2 \in \mathbb{N}^*$ premiers entre eux.
Soit $\phi : (\mathbb{Z}/(m_1 m_2)\mathbb{Z}) \to (\mathbb{Z}/m_1\mathbb{Z}) \times (\mathbb{Z}/m_2\mathbb{Z})$ définie par $\phi(\overset{-(m_1 m_2)}{c}) = (\overset{-(m_1)}{c}, \overset{-(m_2)}{c})$.
Alors, ϕ est un isomorphisme de groupes.
Soit $um_1 + vm_2 = 1$ une relation de Bézout pour m_1 et m_2. On a alors :

$$\phi^{-1}(\overset{-(m_1)}{a}, \overset{-(m_2)}{b}) = \overset{-(m_1 m_2)}{avm_2 + bum_1}.$$

Ainsi, on a

$$(\mathbb{Z}/(m_1 m_2)\mathbb{Z}, +) \simeq ((\mathbb{Z}/m_1\mathbb{Z}) \times (\mathbb{Z}/m_2\mathbb{Z}), +).$$

THÉORÈME 47 (Théorème d'isomorphisme chinois généralisé)
Soit $r \geq 2$. Soient $m_1, \ldots, m_2 \in \mathbb{N}^*$ premiers entre eux deux à deux.
Soit $\phi : (\mathbb{Z}/(m_1 \ldots m_r)\mathbb{Z}) \to (\mathbb{Z}/m_1\mathbb{Z}) \times \ldots \times (\mathbb{Z}/m_r\mathbb{Z})$ définie par $\phi(\overset{-(m_1 \ldots m_r)}{c}) = (\overset{-(m_1)}{c}, \ldots, \overset{-(m_r)}{c})$.
Alors, ϕ est un isomorphisme de groupes.
Pour tout $1 \leq i \leq r$, soient $u_i m_i + v_i \Pi_{j \neq i} m_j = 1$ des relations de Bézout pour m_i et $\Pi_{j \neq i} m_j$. On a alors :

$$\phi^{-1}(\overset{-(m_1)}{a_1}, \ldots, \overset{-(m_r)}{a_r}) = \overset{-(m_1 \ldots m_r)}{\sum_{i=1}^{r} a_i v_i \Pi_{j \neq i} m_j}.$$

Ainsi, on a

$$(\mathbb{Z}/(m_1 \ldots m_r)\mathbb{Z}, +) \simeq ((\mathbb{Z}/m_1\mathbb{Z}) \times \ldots \times (\mathbb{Z}/m_r\mathbb{Z}), +).$$

4.1.3 Groupes monogènes, Théorème de Lagrange

THÉORÈME 48 (Caractérisation des groupes monogènes)
Soit $(G, \star)$ un groupe monogène engendré par a.

1. Si G est d'ordre infini, alors $\varphi : k \in \mathbb{Z} \mapsto a^k \in G$ est un isomorphisme de groupes.
   On a donc $G \simeq \mathbb{Z}$.
2. Si G est cyclique d'ordre n, alors $\overline{\varphi} : \overline{k} \in \mathbb{Z}/n\mathbb{Z} \mapsto a^k \in G$ est un isomorphisme de groupes.
   On a donc $G \simeq \mathbb{Z}/n\mathbb{Z}$.

THÉORÈME 49
Soit G un groupe fini de cardinal n. Alors :

$$\forall a \in G,\ a^n = e.$$

THÉORÈME 50 (Théorème de Lagrange)
Soit G un groupe fini et H un sous-groupe de G.
Alors, l'ordre de H divise l'ordre de G : $\mathrm{Card}(H) \mid \mathrm{Card}(G)$.

COROLLAIRE 51
Soit G un groupe fini.
Alors pour tout $a \in G$, l'ordre de a divise l'ordre de G : $\mathrm{ord}(a) \mid \mathrm{Card}(G)$.

REMARQUE 52 — *Le théorème de Lagrange et ses conséquences sont des résultats très importants en théorie des groupes.*
Ils apportent des informations sur des groupes finis qui peuvent être très gros et compliqués (par exemple Bij$(\{1, \ldots, n\})$, groupe à $n!$ éléments, ou l'un de ses sous-groupes).

Les groupes finis commutatifs ne sont qu'un cas particulier des groupes finis, et les groupes cycliques (les $\mathbb{Z}/n\mathbb{Z}$ à isomorphisme près) ne sont qu'un cas particulier des groupes finis commutatifs. (par exemple, $\mathbb{Z}/2\mathbb{Z} \times \mathbb{Z}/2\mathbb{Z} \times \mathbb{Z}/2\mathbb{Z}$ n'a rien à voir avec $\mathbb{Z}/8\mathbb{Z}$)
Un groupe fini non-commutatif se comporte très différemment d'un groupe fini commutatif. On a besoin des théorèmes comme le théorème de Lagrange pour étudier, obtenir des informations sur ces groupes.

Exemple 53 — *Soit G un groupe d'ordre 105. On a $105 = 3.5.7$.*
Pour H un sous-groupe de G, le cardinal de H vaut alors $1, 3, 5, 7, 15, 21, 35$ ou 105.
Pour $a \in G$, l'ordre de a vaut alors $1, 3, 5, 7, 15, 21, 35$ ou 105.
Si on a un élément a d'ordre 105, alors G est cyclique et $G \simeq \mathbb{Z}/105\mathbb{Z}$.

Pour a un élément d'ordre 5 et b un élément d'ordre 7, le sous-groupe $\langle\{a, b\}\rangle$ est alors de cardinal 35 ou 105, car son cardinal est un multiple de 5 et de 7, et est un diviseur de 105.

Proposition 54
Soit G un groupe fini d'ordre p, avec p premier.
Alors G est isomorphe à $\mathbb{Z}/p\mathbb{Z}$.
Autrement dit, $\mathbb{Z}/p\mathbb{Z}$ est l'unique groupe (à isomorphisme près) d'ordre p.

Remarque 55 (Bilan sur les groupes) — *Les familles de groupes que l'on utilise ou aborde souvent sont :*

1. *$\mathbb{Z}, \mathbb{Q}, \mathbb{R}, \mathbb{C}$, pour la loi additive $+$;*
2. *$\mathbb{Q}^*, \mathbb{R}^*, \mathbb{C}^*$, pour la loi multiplicative $\times$.*
   Le groupe $\mathbb{U}$ des complexes de module 1, pour la lo $\times$;
3. *Les groupes quotients $\mathbb{Z}/n\mathbb{Z}$, $n \geq 1$, pour la loi additive $+$;*
4. *Les groupes de bijections $Bij(E)$, pour la loi de composition $\circ$.*
   Les groupes de permutation S_n, pour $n \geq 1$, pour la loi de composition $\circ$ (voir Géom. 2).
   Les groupes d'isométrices du plan ou de l'espace, $Isom(E)$, pour la loi $\circ$ (voir Géom. 2).
   Les groupes d'isométries qui préservent un polygone régulier, D_n, pour la loi $\circ$ (voir Géom. 2) ;
5. *Les groupes de matrices inversibles $Gl_n(\mathbb{K})$, pour la loi multiplicative $\times$;*
6. *Les anneaux A, pour la loi $+$ (voir chapitre Anneaux) ;*
7. *Les groupes des inversibles d'un anneau, $A^\times$, pour la loi multiplicative $\times$ (voir chap. Anneaux) ;*

Remarque 56 (Propriétés des groupes) —
Les types de groupes que nous avons sont :

1. *Groupes monogènes : $\mathbb{Z}/n\mathbb{Z}, \mathbb{Z}$, sous-groupes de ces groupes.*
2. *Groupes commutatifs finis : Produits d'un nombre fini de $\mathbb{Z}/n_i\mathbb{Z}$ avec les n_i non-tous premiers entre eux. Sous-groupes de ces groupes.*
3. *Groupes commutatifs : $\mathbb{Q}, \mathbb{R}, \mathbb{C}$, $\mathbb{Z}^n$, $(\mathbb{Q}^*, \times)$, $(\mathbb{R}^*, \times)$, $(\mathbb{C}^*, \times)$, $(\mathbb{U}, \times)$. Produits d'un nombre fini (ou infini) de groupes commutatifs. Sous-groupes de groupes commutatifs.*
4. *Groupes finis (non commutatifs) : $(S_n, \circ)$. Groupes d'isométries du plan/de l'espace. Produits d'un nombre fini de ces groupes.*
5. *Groupes généraux : $(Bij(E), \circ)$ pour E infini, $(Gl_n(\mathbb{K}), \times)$. Produits et sous-groupes de ces groupes.*

4.2 Anneaux

Définition 57
Soit $(A, +, \times)$ un ensemble muni de deux lois de composition internes.
On dit que $(A, +, \times)$ est un **anneau** si :

1. $(A, +)$ est un groupe commutatif, d'élément neutre noté 0_A ;
2. La loi $\times$ est associative et admet un élément neutre noté 1_A, avec $1_A \neq 0_A$;
3. La loi $\times$ est distributive à gauche et à droite par rapport à $+$:

$$\forall x, y, z \in A,\ x \times (y + z) = (x \times y) + (x \times z) \text{ et } (x + y) \times z = (x \times z) + (y \times z).$$

De plus, on dit que A est un anneau **commutatif** si la loi $\times$ est commutative ($x \times y = y \times x$).

Remarque 58 — *Dans le cas d'un anneau A, le symétrique d'un élément a pour la loi $+$ est noté $-a$ (**opposé** de a) et le symétrique de a pour la loi $\times$, s'il existe, est noté a^{-1} (**inverse** de a).*
Les notations 0_A et 1_A sont très générales, et on les abrège parfois en 0 et 1 (élément neutre pour l'addition, élément neutre pour la multiplication).
On notera souvent ab pour $a \times b$.
On écrira souvent "Soit A un anneau", les lois $+$ et $\times$ étant sous-entendues.

Exemples 59

1. $(\mathbb{Z},+,\times)$, $(\mathbb{Q},+,\times)$, $(\mathbb{R},+,\times)$, $(\mathbb{R},+,\times)$ *sont des anneaux commutatifs.*
2. *Pour E un ensemble non-vide et $\mathcal{F}(E,\mathbb{C})$ l'ensemble des fonctions de E dans $\mathbb{C}$, $(\mathcal{F}(E,\mathbb{C}),+,\times)$ est un anneau commutatif.*
   Cela reste vrai si l'on remplace $\mathbb{C}$ par un anneau A. (Le vérifier)
3. *En particulier, $(\mathbb{K}^{\mathbb{N}},+,\times)$ est un anneau. C'est l'ensemble des suites à coefficients dans un corps $\mathbb{K}$ muni de l'addition et de la multiplication termes à termes.*
4. *$(\mathbb{K}[X],+,\times)$ est un anneau. C'est l'ensemble des polynômes à coefficients dans un corps $\mathbb{K}$, pour l'addition et la multiplication de polynômes.*

Exemples 60 1. *Pour $\mathbb{K}=\mathbb{Q},\mathbb{R},\mathbb{C}$, $(\mathscr{M}_n(\mathbb{K}),+,\times)$ est un anneau.*
   Si $n\geq 2$, cet anneau n'est pas commutatif. (Le vérifier)
2. *Pour E un $\mathbb{K}$-espace vectoriel, $(\mathcal{L}(E),+,\circ)$ est un anneau.*
   Si $\dim(E)\geq 2$, cet anneau n'est pas commutatif.
3. *Pour A un anneau, $(\mathcal{F}(A,A),+,\circ)$ n'est pas un anneau. En effet, on a bien $(f+g)\circ g = f\circ g + f\circ h$, mais en général on a $f\circ(g+h)\neq f\circ g + f\circ h$.*

Remarque 61 — *Dans la littérature mathématique, on peut aussi définir un anneau sans demander l'existence de l'élément unitaire 1_A. Un anneau possédant un élément unitaire 1_A est alors appelé anneau unitaire.*

Les exemples classiques d'anneaux commutatifs sont $\mathbb{Z},\mathbb{Q},\mathbb{R},\mathbb{K}[X]$, voire un produit de certains d'entre eux.
Il ne faut pas hésiter à les utiliser pour tester si un résultat sur les anneaux commutatifs est vrai, ou voir s'il est faux (chercher un contre-exemple).
Les exemples classiques d'anneaux non-commutatifs sont les $\mathscr{M}_n(\mathbb{K})$ $(n\geq 2)$. Il ne faut pas hésiter à les utiliser (pour $n=2,3,4$) afin de tester si un résultat sur les anneaux commutatifs est vrai, ou voir s'il est faux (chercher un contre-exemple).

4.2.1 Groupe des éléments inversibles

Définition 62
Soit A un anneau et $a\in A$.
On dit que a **inversible** s'il admet un symétrique pour la loi $\times$: $b\in A$ tel que $ab=ba=1_A$.
Ce symétrique est appelé inverse de a, et est noté a^{-1}.
On note $A^\times$ l'ensemble des éléments inversibles de l'anneau A. C'est le **groupe des inversibles** de A.

Proposition 63
L'ensemble des éléments inversible d'un anneau A, $A^\times$, est un groupe pour la multiplication $\times$.
Si A est commutatif, alors ce groupe est abélien.

Remarque 64 — *Attention à ne pas confondre $A^*=A\setminus\{0_A\}$ et $A^\times$ l'ensemble des inversibles de A. Dans cerains cas (par ex $\mathbb{R},\mathbb{C},\mathbb{Q}$), ces deux ensembles sont égaux.*

On a par contre $\mathbb{Z}^\times=\{-1,1\}$. De même, $\mathscr{M}_n(\mathbb{K})^\times = GL_n(\mathbb{K})\neq\mathscr{M}_n(\mathbb{K})^*$.

Définition 65
Soit $(A,+,\times)$. Soit $a\in A$ non-nul. Alors :

1. Pour $n\geq 0$, on définit $0_A^n=0_A$ (donc $0_A^0=0_A$) ;
2. Pour $n\geq 1$, on définit $a^n=a\times\ldots\times a$ (n fois) ;
3. Pour $n=0$, on définit $a^0=1_A$;
4. Si a est inversible, pour $n<0$, on définit $a^n=a^{-1}\times\ldots\times a^{-1}$ ($-n$ fois).

Remarque 66 — *On retrouve alors les propriétés habituelles.*
Pour $k,m,n\in\mathbb{N}$ et $a\in A$, on a :

$$a^m\times a^n=a^{m+n}\quad\text{et}\quad (a^n)^k=a^{kn}.$$

Si $a\in A$ est inversible, cela reste vrai pour a^n, pour tout $n\in\mathbb{Z}$.

4.2.2 L'anneau $(\mathbb{Z}/n\mathbb{Z}, +, \times)$

Proposition 67
Soit $n \geq 2$. $(\mathbb{Z}/n\mathbb{Z}, \overline{+}, \overline{\times})$ est un anneau, qui est commutatif.
Son élément neutre pour l'addition est $\overline{0}$, l'élément nul de $\mathbb{Z}/n\mathbb{Z}$.
Son élément neutre pour la multiplication est $\overline{1}$, l'élément unitaire de $\mathbb{Z}/n\mathbb{Z}$.

Remarque 68 —
• *Les anneaux de la forme $\mathbb{Z}/n\mathbb{Z}$ sont des anneaux avec un nombre fini d'éléments.*
On va alors trouver des phénomènes très différents de $\mathbb{Z}$, $\mathbb{Q}$, $\mathbb{R}$, $\mathbb{C}$.
• *Par exemple, dans $\mathbb{Z}/n\mathbb{Z}$, on a $\sum_{k=1}^{n} 1 = 1 + \ldots + 1 (n \text{ fois}) = 0$.*
• *De même, pour n qui n'est pas permier (par ex. $n = 2.3$), en écrivant $n = a.b$ avec $a, b \in \{1, \ldots, n-1\}$, on va avoir $\bar{a} \neq 0$, $\bar{b} \neq 0$ dans $\mathbb{Z}/n\mathbb{Z}$, mais $\bar{a}\bar{b} = \overline{ab} = \bar{n} = 0$. (le produit de deux nombres non-nuls peut être nul)*
• *Autre exemple de phénomène. Dans $\mathbb{Z}/8\mathbb{Z}$ on a vu avec les congruences que $\bar{1}^2 = \bar{3}^2 = \bar{5}^2 = \bar{7}^2 = \bar{1}$. Ainsi, l'équation $x^2 = \bar{1}$ possède plus de 2 solutions.*

Remarque 69 — *Dans la suite du cours, quand les lois d'addition et de multiplication seront claires (quand on sait sur quel anneau on travaille), on notera $+$ à la place de $\overline{+}$ et $\times$ à la place de $\overline{\times}$ pour les lois de $\mathbb{Z}/n\mathbb{Z}$.*

Le groupe des inversibles de $\mathbb{Z}/n\mathbb{Z}$

Proposition-Définition 70
Soit $n \geq 2$. Pour $(\mathbb{Z}/n\mathbb{Z})^\times$ l'ensemble des éléments de $\mathbb{Z}/n\mathbb{Z}$ inversibles, on a :

$$(\mathbb{Z}/n\mathbb{Z})^\times = \{\overline{m},\ m \in \mathbb{Z} \text{ avec } \operatorname{pgcd}(m, n) = 1\}.$$

C'est un groupe fini et abélien.
On a aussi $\operatorname{Card}(\mathbb{Z}/n\mathbb{Z}^\times) = \varphi(n)$, où $\varphi(n)$ est **l'indicatrice d'Euler** de n.

Proposition 71
Soit p un nombre premier.

1. Alors on a $(\mathbb{Z}/p\mathbb{Z})^\times = \mathbb{Z}/p\mathbb{Z}^*$.
   Ainsi, on obtient $\varphi(p) = p - 1$.
2. Pour tout $k \geq 1$, on a $\varphi(p^k) = p^k - p^{k-1}$.

Théorème 72 (Petit théorème de Fermat)
Soient $a \in \mathbb{Z}$ et $n \in \mathbb{N}^*$ avec $\operatorname{pgcd}(a, n) = 1$. Alors, on a :

$$a^{\varphi(n)} \equiv 1 \bmod n$$

$$\text{c'est-à-dire } \overline{a}^{\varphi(n)} = \overline{1} \text{ dans } \mathbb{Z}/n\mathbb{Z}.$$

Si n est un nombre premier, alors on obtient $a^{n-1} \equiv 1 \bmod n$.

Preuve — L'élément $\overline{a}$ est alors contenu dans le groupe $(\mathbb{Z}/n\mathbb{Z})^\times$, qui est de cardinal $\varphi(n)$. Le théorème de Lagrange permet alors de conclure. □

4.2.3 Anneaux intègres

Définition 73
Soit $a \in A$, avec $a \neq 0$. On dit que a est un **diviseur de zéro** s'il existe $b \in A$ non-nul tel que $ab = 0$.

Remarque 74 — *Si $a \in A$ est un diviseur de 0, alors a n'est pas inversible.*
En effet, si a était inversible on aurait $b = a^{-1}ab = 0_A$, ce qui est impossible.

Définition 75
Soit A un anneau. On dit que A est **intègre** s'il n'a pas de diviseurs de zéro.

Proposition 76
Soit A un anneau intègre.

1. $\forall a, b \in A$, si $ab = 0$ alors $a = 0_A$ ou $b = 0_A$.
2. $\forall a, b, c \in A$ avec $a \neq 0_A$, si $ab = ac$ alors $b = c$.

3. $\forall a, b, c \in A$ avec $a \neq 0_A$, si $ba = ca$ alors $b = c$.

Remarque 77 — *Dans un anneau intègre, tous les éléments non-nuls ne sont pas forcément inversibles, mais on peut simplifier une équation en la factorisant.*
Cela est par exemple le cas dans l'anneau $\mathbb{Z}$.

Proposition 78
Soit A un anneau intègre. Soient $b_1, \ldots, b_n \in A$ (pas forcément distincts).
Alors, on a $(x - b_1) \ldots (x - b_n) = 0$ si et seulement si $x = b_i$ pour un $1 \leq i \leq n$.
L'équation $(x - b_1) \ldots (x - b_n) = 0$ possède donc au plus n solutions.

Remarque 79 —
• *Si A est un anneau intègre, alors l'équation $x^2 = 1$ a pour unique solution 1 et -1 car l'équation s'écrit $(x - 1_A)(x + 1_A) = 0_A$.*
• *Si l'on trouve une équation polynômiale dans A de la forme $\Pi_i(x - b_i) = 0$ qui possède plus de n solutions, on sait donc immédiatement que l'anneau A n'est pas intègre.*
• *Dans un anneau produit $A \times B$ tel que $1_A \neq -1_A$ et $1_B \neq -1_B$, il y a au moins 4 solutions à cette équation et donc $A \times B$ n'est jamais intègre.*
• *Dans $\mathcal{M}_2(\mathbb{R})$, pour tout $\theta \in \mathbb{R}$, $S(\theta) = \begin{pmatrix} \cos\theta & \sin\theta \\ \sin\theta & -\cos\theta \end{pmatrix}$ vérifie $x^2 = 1$.*
Les anneaux $\mathscr{M}_n(\mathbb{K})$ ne sont jamais intègres pour $n \geq 2$.

Exemple 80 —

- *Les anneaux $\mathbb{Z}, \mathbb{Q}, \mathbb{R}, \mathbb{C}$ sont intègres.*
- *Les anneaux $\mathbb{Z}/p\mathbb{Z}$ pour p premier sont intègres.*
- *Les anneaux $\mathbb{K}[X]$ sont intègres.*
- *Les anneaux $\mathbb{Z}/n\mathbb{Z}$, pour $n \geq 2$ non premier, ne sont pas intègres.*
- *Les anneaux $\mathscr{M}_n(\mathbb{K})$, pour $n \geq 2$, ne sont pas intègres. (Le vérifier.)*
- *Les anneaux produits $A \times B$ ne sont pas intègres.*
- *Les anneaux $\mathcal{F}(E, A)$ ne sont pas intègres si $Card(E) \geq 2$.*

4.2.4 Calcul dans les anneaux

Proposition 81
Soit $(A, +, \times)$ un anneau. Soient $a, b \in A$ qui commutent ($a \times b = b \times a$). Alors, on a :

1. $(a + b)^n = \sum_{k=0}^{n} \binom{n}{k} a^k b^{n-k}$, $\forall n \geq 1$ **(Formule du binôme)** ;
2. $a^n - b^n = (a - b) \sum_{k=0}^{n-1} a^{n-k-1} b^k$, $\forall n \geq 2$ **(Formule de Bernoulli)** ;
3. $a^{2n+1} + b^{2n+1} = (a + b) \sum_{k=0}^{2n} (-1)^k a^{2n-k} b^k$, $\forall n \geq 1$.

Remarque 82 — *Vous devez connaître ces formules et leurs cas particuliers quand n est petit, comme par exemple*
1. $a^3 - b^3 = (a - b)(a^2 + ab + b^2)$;
2. $a^3 + b^3 = (a + b)(a^2 - ab + b^2)$.

Corollaire 83
On retrouve la formule de la **somme géométrique** : $1 - a^n = (1 - a) \sum_{k=0}^{n-1} a^k$.

Si $1 - a$ est inversible, on a alors : $\sum_{k=0}^{n-1} a^k = (1 - a)^{-1}(1 - a^n)$.

Exercice 1 — *On dit qu'un élément $a \in A$ est **nilpotent** s'il existe $n \in \mathbb{N}^*$ tel que $a^n = 0$. ("nil" : "nul", "potent" : "puissance")*
*On appelle **indice de nilpotence** de a le plus petit $n \geq 1$ tel que $a^n = 0$.*

1. *Pour $A = \mathscr{M}_n(\mathbb{R})$ avec $n = 2, 3$ ou 4, trouver des exemples de matrices nilpotentes. (On pourra chercher des matrices triangulaires supérieures/inférieures).*
2. *Soient $a, b \in A$ qui commutent et sont nilpotents d'indices p et q. Montrer que $a+b$ est nilpotent et majorer son indice de nilpotence.*
3. *Si a et b ne commutent plus, peut-on encore dire quelque chose sur $a + b$? (Si non, on cherchera un contre-exemple)*
4. *Si a est nilpotent d'indice p, que peut-on dire de $1 - a$?*

4.2.5 Sous-anneaux, Idéaux

Définition 84
Soit $(A, +, \times)$ un anneau et $B \subset A$.
On dit que B est un **sous-anneau** de A si :

1. $(B, +)$ est un sous-groupe de $(A, +)$;
2. $1_A \in B$;
3. la loi $\times$ est stable sur B : $\forall x, y \in B$, on a $x \times y \in B$.

Remarque 85 —

- *Pour B un sous-anneau de A, l'ensemble $(B, +, \times)$ muni de la restriction de $+$ et $\times$ à B est un anneau.*
- *Z est un sous-anneau de $\mathbb{Q}$, qui est un sous-anneau de $\mathbb{R}$, qui est un sous-anneau de $\mathbb{C}$.*
- *$\mathbb{K}$ est un sous-anneau de $\mathbb{K}[X]$.*
- *$\mathbb{K}_n[X]$ n'est pas un sous-anneau de $\mathbb{K}[X]$. $X\mathbb{K}[X]$ n'est pas un sous-anneau non plus.*
- *Il est important de vérifier que B contient l'élément 1_A :*
  Pour $n \geq 2$, $n\mathbb{Z}$ n'est pas un sous-anneau de $\mathbb{Z}$. Il vérifie toutes les propriétés nécessaires sauf celle de l'existence d'un élément neutre pour la multiplication.

Proposition 86
Soit $(A, +, \times)$ un anneau et $B \subset A$. B est un sous-anneau de A si et seulement si :

1. $1_A \in B$;
2. $\forall (x, y) \in B^2$, $x - y \in B$;
3. $\forall (x, y) \in B^2$, $xy \in B$.

Exemple 87 —
• *On pose $\mathbb{Z}[\sqrt{2}] = \{a + b\sqrt{2},\ (a, b) \in \mathbb{Z}^2\}$. C'est un sous-anneau de $(\mathbb{R}, +, \times)$.*
• *On pose $\mathbb{Z}[X] = \{P \in \mathbb{Q}[X],$ avec P à coefficients dans $\mathbb{Z}\}$.*
C'est un sous-anneau de $(\mathbb{Q}[X], +, \times)$.

Proposition-Définition 88
Soient $n \geq 1$ et $\mathbb{K}$ un corps. Soient $A \in \mathscr{M}_n(\mathbb{K})$ et $P \in \mathbb{K}[X]$.
On définit $\mathbb{K}[A] = Vect(A^k,\ k \geq 0)$ et $\mathbb{K}[P] = Vect(P^k,\ k \geq 0)$.
Alors, $\mathbb{K}[A]$ est un sous-anneau de $\mathscr{M}_n(\mathbb{K})$ et $\mathbb{K}[P]$ est un sous-anneau de $\mathbb{K}[X]$.
$\mathbb{K}[P]$ est un sous-anneau commutatif et intègre (car $\mathbb{K}[X]$ est commutatif et intègre).
$\mathbb{K}[A]$ est un sous-anneau commutatif.

Remarque 89 —
• *Les sous-anneaux $\mathbb{K}[A]$ et $\mathbb{K}[P]$ sont construits en utilisant la structure de $\mathbb{K}$-algèbre de $\mathscr{M}_n(\mathbb{K})$ et $\mathbb{K}[X]$, que l'on étudiera plus tard. Cela fournit quelques sous-anneaux que l'on peut étudier.*
• *La commutativité de $\mathbb{K}[A]$ est très utile pour faire des calculs. (déjà vu au chapitre Matrices)*
• *Le sous-anneau $\mathbb{K}[A]$ n'est pas intègre en général.*
Par exemple, si $A^2 = I_n$ (symétrie), $A^2 = A$ (projection), $A^m = 0$ (nilpotent), ou $A = Diag(\lambda_1, \ldots, \lambda_n)$ (diagonal), avec $A \neq \lambda I_n$, alors le sous-anneau $\mathbb{K}[A]$ n'est pas intègre.
On peut trouver $B_1, B_2 \in \mathbb{K}[A]$ non-nulles telles que $B_1 B_2 = 0$.

Par rapport aux sous-groupes qui sont seulement stables pour la loi $+$, les sous-anneaux doivent être stables pour les lois $+$ et $\times$ et en plus contenir l'élément 1.
Pour tout $a \in A$ non-inversible, l'ensemble des multiples de a, aA, n'est donc pas un sous-anneau.
On va définir un objet mathématique différent des sous-anneaux pour étudier cela, les idéaux.

Définition 90
Soient A un anneau commutatif et $I \subset A$.
On dit que I est un **idéal** de A si :

1. $(I, +)$ est un sous-groupe de $(A, +)$;
2. $\forall x \in I$, $\forall a \in A$, on a $ax \in I$ (on dit que I est **absorbant** pour la loi $\times$).

Exemple 91 —

1. *Le singleton $\{0_A\}$ et A sont des idéaux de A.*
2. *Pour $a \in A$, l'ensemble aA des multiples de a est un idéal de A.*
3. *La définition d'idéal ne concerne que les anneaux commutatifs, afin d'avoir $ax = xa$.*
   On ne s'intéresse pas aux anneaux non-commutatifs (ex : matrices) dans ce cas.
4. *$\mathbb{Z}, \mathbb{Q}, \mathbb{R}$ ne sont pas des idéaux de $\mathbb{C}$ car ils ne sont pas absorbants pour la multiplication. Pourtant, ce sont des sous-anneaux de $\mathbb{C}$.*
5. *Si un idéal I contient un élément inversible, alors $I = A$. En particulier, les idéaux de $\mathbb{Q}, \mathbb{R}, \mathbb{C}$ sont $\{0\}$ et l'anneau tout entier.*
6. *Tout sous-anneau B strict d'un anneau A n'est pas un idéal de A, car B contient 1_A mais ne contient pas A tout entier.*
7. *Pour tout $n \in \mathbb{Z}$, $n\mathbb{Z}$ est un idéal de $\mathbb{Z}$. Pourtant, le seul sous-anneau de $\mathbb{Z}$ est $\mathbb{Z}$ lui-même.*
8. *Pour $n \in \mathbb{Z}^*$, $n\mathbb{Z}$ est un idéal de $\mathbb{Z}$ mais ce n'est pas un idéal de $\mathbb{Q}, \mathbb{R}$ ou $\mathbb{C}$.*

Remarque 92 — *Un idéal I d'un anneau A est un sous-ensemble qui est en partie similaire à un sous-anneau (identique pour l'addition, mêmes règles de multiplication).*
Par contre, il ne contient en général pas d'élément neutre pour la multiplication (pas de 1).

La définition de sous-anneau de $(A, +, \times)$ ne dépend que des opérations sur A, et non de tous les éléments de A (par exemple, Z est un sous-anneau de $\mathbb{Q}$, de $\mathbb{R}$ et de $\mathbb{C}$, mais les nombres réels et complexes n'interviennent nulle part).
Pour un idéal, la condition d'absorbance implique un lien entre les éléments de I et ceux de A.

Par exemple, $2\mathbb{Z}$ est un idéal de $\mathbb{Z}$ mais n'est pas un idéal de $\mathbb{Q}$,$\mathbb{R}$ ou $\mathbb{C}$. La condition d'absorbance fait entrer les nombres rationnels/réels/complexes dans les multiplications possibles, et $2\mathbb{Z}$ ne contient pas tout cela.

Proposition 93
Soit $(A, +, \times)$ un anneau commutatif et $(I_k)_{k \in E}$ une famille d'idéaux de A indexée par un ensemble E. Alors

1. $\bigcap_{k \in E} I_k$ est un idéal de A.
2. Si E est de cardinal fini, alors $\sum_{k \in E} I_k$ est un idéal de A.
3. $\sum_{k \in E} I_k = \left\{ \sum_{k=1}^{n} x_i, \quad \text{avec } x_i \in \bigcup_{k \in E} I_k,\ \forall i \in [\![1, n]\!],\ \forall n \in \mathbb{N}^* \right\}$ est un idéal de A.

4.2.6 Anneaux principaux

Définition 94
Soit $(A, +, \times)$ un anneau commutatif, et I un idéal de A.

1. On dit que l'idéal I est **principal** s'il existe $a \in A$ tel que $I = aA$.
2. On dit que l'anneau A est **principal** si A est intègre et si tout idéal de A est principal.

Exemple 95 —

1. *L'anneau $(\mathbb{Z}, +, \times)$ est un anneau principal.*
2. *Les corps $(\mathbb{Q}, \mathbb{R}, \mathbb{C}, \mathbb{Z}/p\mathbb{Z})$ sont des anneaux principaux.*
3. *Les anneaux $\mathbb{Z}/n\mathbb{Z}$ avec n non-premier ne sont pas des anneaux principaux car ils ne sont pas intègres.*

Exemple 96 —

1. *Tous les anneaux qui ne sont pas intègres ne sont pas principaux ($\mathbb{Z}/n\mathbb{Z}$ avec n non-premier,...).*
2. *L'anneau $(\mathbb{Z}[X], +, \times)$ des polynômes à coefficients entiers n'est pas un anneau principal.*
   En effet, l'idéal $\langle 2, X \rangle$ n'est pas un idéal principal de $\mathbb{Z}[X]$. (Le vérifier.)

Propriétés de l'anneau $(\mathbb{K}[X], +, \times)$

Théorème 97
L'anneau $(\mathbb{K}[X], +, \times)$ est principal.
Soit I un idéal de $\mathbb{K}[X]$ qui n'est pas réduit à $\{0\}$. Alors il existe un unique polynôme unitaire P tel que $I = P\mathbb{K}[X]$.

Remarque 98 — *La preuve de ce théorème utilise la division euclidienne de polynômes dans* $\mathbb{K}[X]$.
Les anneaux $\mathbb{K}[X]$ *et* $\mathbb{Z}$ *sont appelés* ***anneaux euclidiens*** *(des anneaux possédant une division euclidienne).*
La preuve du théorème permet en fait de montrer qu'un anneau qui possède une division euclidienne (un anneau euclidien) est un anneau intègre.

PGCD et PPCM, Théorèmes de Bézout et de Gauss dans les anneaux principaux

Proposition 99
Soit A un anneau intègre. Soient $z, z' \in A$.
Alors, on a $zA = z'A$ si et seulement si z et z' sont associés.
Ainsi, l'idéal principal $I = zA = \langle z \rangle$ possède un unique élément générateur à association près.

Remarque 100 — *Dans l'étude des propriétés des anneaux principaux (divisiblité, factorisations), les éléments inversibles ne sont pas très intéressants (ce sont des éléments qui divisent tout le monde, comme* -1 *dans* $\mathbb{Z}$ *ou comme* $\lambda \neq 0$ *dans* $\mathbb{K}[X]$*).*
Ainsi, la majorité des résultats sera à association près, c'est-à-dire à un représentant de classe d'équivalence près pour la relation d'association.
Dans $\mathbb{Z}$*, ce représentant sera un nombre positif. Dans* $\mathbb{K}[X]$*, ce sera un polynôme unitaire.*
Par contre, dans le cas général, il n'y a pas de méthode pour choisir un représentant particulier dans la relation "être associé". On pensera donc bien à préciser que l'unicité d'un élément ou d'une factorisation sera "à association près".

Définition 101
Soit A un anneau principal. Soient x,$y \in A$.
On définit le **plus grand diviseur commun** de x et y, noté $\text{pgcd}(x, y)$ ou $x \wedge y$, comme l'unique élément de A (à association près) tel que :

$$\langle x, y \rangle = xA + yA = \text{pgcd}(x, y)A.$$

On définit le plus petit diviseur commun de x et y, noté $\text{ppcm}(x, y)$ ou $x \vee y$, comme l'unique élément de A (à association près) tel que :

$$xA \cap yA = \text{ppcm}(x, y)A.$$

Définition 102
Soit A un anneau principal. Soient x,$y \in A$.
On dit que x et y sont **premiers entre eux** si $\text{pgcd}(x, y) = 1$.

Proposition 103 (Théorème de Bézout)
Soit A un anneau principal. Soient x,$y \in A$.
Les élémnts x et y sont premiers entre eux si et seulement s'il existe $u, u \in A$ tels que $ux + vy = 1$.

Preuve — On a $\text{pgcd}(x, y) = 1$ ssi $xA + yA = A$. Cela implique qu'il existe $u, v \in A$ tels que $1 = ux + vy$.

Réciproquement, si l'on a $1 = xu + vy$, alors $xA + yA$ contient $1.A = A$, d'où $xA + yA = A$. □

Remarque 104 — *On peut alors montrer de l'exacte même façon qu'avec les entiers que* $\text{pgcd}(x, y)$ *et* $\text{ppcm}(x, y)$ *sont bien les plus grands commun diviseur et plus petit commun multiple pour le relation de divisibilité, à association près.*

Proposition 105
Soit A un anneau principal. Soient x,$y \in A$ et $s \in A$.
Alors s est égal à $\text{pgcd}(x, y)$ à association près si et seulement si :

1. $s \mid x$ et $s \mid y$,
2. Pour tout $t \in A$ tel que $t \mid x$ et $t \mid y$, on a $t \mid s$.

Autrement dit, s est le plus grand diviseur de x et de y.

Proposition 106
Soit A un anneau principal. Soient x,$y \in A$ et $s \in A$.
Alors s est égal à ppcm(x, y) à association près si et seulement si :
1. $x \mid s$ et $y \mid s$,
2. Pour tout $t \in A$ tel que $x \mid t$ et $y \mid t$, on a $s \mid t$.
Autrement dit, s est le plus petit multiple de x et de y.

Proposition 107 (Théorème de Gauss)
Soit A un anneau principal. Soient $x, y, z \in A$.
Si x divise yz et pgcd$(x, y) = 1$, alors $x|z$.

Preuve — la preuve est identique au cas des entiers. D'après e théorème de Bézout, il existe $u, v \in A$ tels que $ux + vy = 1$. Cela donne $xuz + yvz = z$. Comme $x|yz$, on a $x|(xuz + yvz) = z$. □

Remarque 108 — *Si l'anneau A est principal mais ne possède pas de division euclidienne (s'il n'est pas euclidien), on sait que pour $x, y \in A$ il existe $u, v \in A$ tels que $xu + yv = pgcd(x, y)$, mais on n'a aucune méthode pour calculer u et v.*
La division euclidienne, dont découle l'algorithme d'Euclide, est une propriété qui permet de calculer $pgcd(x, y)$ et de calculer les nombres u, v.

Conséquences de ces théorèmes

Proposition 109
Soit A un anneau principal. Soient $a, b, c \in A$.
Si a et b premiers entre eux et si $a \mid c$ et $b \mid c$, alors $ab \mid c$.

Preuve — Supposons que $a \mid c$ et $b \mid c$ avec pgcd$(a, b) = 1$. Le théorème de Bézout nous dit qu'il existe alors des entiers $u, v \in A$ tels que $au + bv = 1$. On a donc $c = auc + bvc$. Comme $a \mid c$, on a on a $ab \mid bc$. Comme $b \mid c$, on a on a $ab \mid ac$. Donc, ab divise $acu + bcv = c$. □

Corollaire 110
Soit A un anneau principal. Soient a_1,...,a_n,$c \in A$.
Si les a_i sont premiers entre eux deux à deux et si $a_i \mid c$ pour tout $1 \leq i \leq n$, alors $a_1 \times a_2 \times \ldots a_n \mid c$.

Preuve — Cette généralisation de la proposition précédente se démontre par récurrence sur $n \geq 2$. □

Si a et b ne sont pas premiers entre eux, on ne peut rien dire ! Par exemple, dans $\mathbb{Z}$, $4 \mid 4$ et $2 \mid 4$ mais $4 \times 2 = 8$ ne divise pas 4.

Exemple 111 — *Si $4 \mid n$ et $3 \mid n$ alors $12 \mid n$ car 4 et 3 sont premiers entre eux.*

Proposition 112
Soit A un anneau principal. Soient $a, b, c \in A$.
Si a est premier avec b et si a est premier avec c, alors a est premier avec bc.

Preuve — Supposons a premier avec b et avec c. D'après le théorème de Bézout, il existe u_1, u_2,v_1,$v_2 \in A$ tels que $1 = au_1 + bv_1$ et $1 = au_2 + cv_2$. Par multiplication, on obtient :

$$1 = a(au_1u_2 + u_1cv_2 + bv_1u_2) + bc(v_1v_2).$$

Ainsi, d'après le théorème de Bézout, a et bc sont premiers entre eux. □

Corollaire 113
Soit A un anneau principal. Soient a,b_1,...,$b_n \in \mathbb{Z}$.
Si a est premier avec b_i pour tout $i \in \{1, \ldots, n\}$, alors a est premier avec $b_1 \times b_2 \times \ldots \times b_n$.

Preuve — Cette généralisation de la proposition précédente se démontre par récurrence sur $n \geq 2$. □

Proposition 114
Soit A un anneau principal. Soient $a, b \in A$.
Si a est premier avec b alors a^m est premier avec b^n pour tous $m, n \in \mathbb{N}$.

Preuve — Soient $m, n \in \mathbb{N}$. Si $m = 0$ ou $n = 0$ on a $a^m = 1$ ou $b^n = 1$, et dans ce cas le résultat est vrai.

Supposons $m, n \neq 0$. Comme a est premier avec b, le corollaire précédent nous dit que a est premier avec b^n. Comme b^n est premier avec a, le corollaire précédent nous dit que b^n est premier avec a^m. □

Exemple 115 — *Soit A un anneau principal. Soit $a \in \mathbb{N}^*$. Comme a est premier avec $a-1$ et avec $a+1$, a est premier avec $(a-1)(a+1) = a^2 - 1$.*

Proposition 116
Soit A un anneau principal. Soient $a, b \in A$. Posons $d = \text{pgcd}(a, b)$.
Alors il existe des éléments a' et b' de A tels que

$$a = da',\ b = db', \quad \text{et} \quad \text{pgcd}(a', b') = 1.$$

Preuve — Si $(a, b) = (0, 0)$, alors $a' = b' = 1$ conviennent.

Supposons que $(a, b) \neq (0, 0)$. Comme $d = \text{pgcd}(a, b)$, on sait que $d \mid a$ et $d \mid b$. Les nombres $a' = \frac{a}{\text{pgcd}(a,b)}, b' = \frac{b}{\text{pgcd}(a,b)}$ sont donc bien définis, et tels que $a = da'$ et $b = db'$. On a alors $d = \text{pgcd}(a, b) = \text{pgcd}(da', db') = d\,\text{pgcd}(a', b')$. Donc, comme d est non nul, on a $\text{pgcd}(a', b') = 1$. □

Remarque 117 (Bilan sur les anneaux étudiés) — *Voici les principales propriétés des anneaux que nous avons définies et étudiées. Elles sont triées par ordre décroissant, et accompagnées d'exemples.*

1. *Les anneaux : $\mathscr{M}_n(\mathbb{K})$, $(Fonct(G, G), +, \circ)$ (avec G un groupe commutatif), produits d'anneaux non-tous commutatifs,...*
2. *Les anneaux commutatifs : $(Fonct(E, A), +, \times)$, $\mathbb{Z}/n\mathbb{Z}$, produits d'anneaux commutatifs,...*
3. *Les anneaux intègres commutatifs : $\mathbb{Z}[X], \mathbb{K}[a]$(pour $a \in A$),...*
4. *Les anneaux principaux : $\mathbb{Z}, \mathbb{Z}[i], \mathbb{Z}[sqrt2], \mathbb{K}[X], \ldots$*
5. *Les corps : $\mathbb{Q}, \mathbb{R}, \mathbb{C}, \mathbb{Z}/p\mathbb{Z}, \mathbb{Q}[i], \mathbb{Q}[sqrt2], \ldots$*

Chaque propriété supplémentaire permet de faciliter les calculs ou les études sur l'anneau A.
• Si A est commutatif, on peut alors effectuer des développements et factorisations, comme la formule du binôme, la somme géométrique, les identités remarquables.
• Si A est intègre, on peut alors résoudre des équations produit-nul ($ab = 0$) et donc simplifier beaucoup d'équations où un terme non-nul est en facteur.
• Si A est principal, il existe des éléments qui ne se factorisent pas, et tous les éléments se factorisent comme produit (à association près) d'éléments irréductibles. Cela aide aussi dans beaucoup de factorisations, à déterminer des divisibilités, et à résoudre certaines équations, en regardant ce qui se passe pour chaque facteur irréductible.
• Si A possède une division euclidienne, tous les éléments que l'on peut définir quand A est principal peuvent se calculer à l'aide de divisions euclidienne et de l'algorithme d'Euclide : diviseurs, pgcd,ppcm,...
• Si A est un corps (tous les éléments non-nuls sont inversibles), alors beaucoup de questions de factorisation et de divisibilité se résolvent instantanément. (nous étudierons les corps par la suite)
Les corps ne sont pas intéressants à étudier comme anneaux principaux (ils n'ont pas d'éléments irréductibles), mais ils sont très intéressants pour définir d'autres objets (anneaux $\mathbb{K}[X]$, $\mathbb{K}$-espaces vectoriels, $\mathbb{K}$-algèbres,...).

Remarque 118 — *Il existe aussi des anneaux qui ne vérifient pas toutes ces propriétés. Il n'y a pas forcément d'exemple très simple pour ces anneaux, mais leur existence montre que l'étude des anneaux en général est très riche.*
Cela montre aussi qu'il existe des anneaux possédant certains propriétés pratiques, mais qui peuvent être très différents des exemples "classiques" que nous avons étudiés dans ce cours.

1. *Anneaux intègres non-commutatifs : $\mathbb{H}$ (anneau des quaternions),...*
2. *Anneaux commutatifs, non-intègres, avec tous les idéaux principaux : $\mathbb{Z}/n\mathbb{Z}$ pour n non premier, produit d'anneaux principaux,...*
3. *Anneaux principaux sans division euclidienne : $\mathbb{Z}[\frac{1+i\sqrt{19}}{2}]$, ... (pas facile du tout !)*

4.2.7 Morphismes d'anneaux, Isomorphismes

Définition 119
Soient $(A, +, \times)$ et $(B, \triangle, .)$ deux anneaux.
Une fonction $f : A \to B$ est un **morphisme d'anneaux** si :

1. f est un morphisme de groupes de $(A, +)$ dans $(B, \triangle)$: $\forall x, y \in A,\ f(x + y) = f(x) \triangle f(y)$;
2. $f(1_A) = 1_B$;

3. $\forall x, y \in A,\ f(x \times y) = f(x).f(y)$.

Un morphisme d'anneaux est ainsi une fonction qui est compatible avec les opérations d'addition ($+$ et $\triangle$), avec les opérations de multiplication ($\times$ et $.$), et telle que $f(1) = 1$.

Exemple 120 —

1. *Soit $(A, +, \times)$ un anneau. La fonction identité $Id : A \to A$ est un morphisme d'anneaux.*
2. *Si B est un sous-anneau de $(A, +, \times)$, alors l'injection $i : x \in B \mapsto x \in A$ est un morphisme d'anneaux. Ainsi, $x \in \mathbb{R} \mapsto x \in \mathbb{C}$ est un morphisme d'anneaux.*
3. *Soient E un ensemble et $x_0 \in E$. La fonction*

$$\varphi_{x_0} : \begin{array}{ccc} \mathcal{F}(E, \mathbb{R}) & \to & \mathbb{R} \\ g & \mapsto & g(x_0) \end{array}$$

   est un morphisme d'anneaux.
   *On l'appelle **morphisme d'évaluation** en x_0.*
4. *La conjugaison complexe $z \in \mathbb{C} \mapsto \overline{z} \in \mathbb{C}$ est un morphisme d'anneaux.*
5. *Soit $n \in \mathbb{N}^*$. La fonction*

$$r : \begin{array}{ccc} \mathbb{Z} & \to & \mathbb{Z}/n\mathbb{Z} \\ x & \mapsto & \overline{x} \end{array}$$

   est un morphisme d'anneaux.
6. *Soit I un intervalle de $\mathbb{R}$. La fonction*

$$D : \begin{array}{ccc} \mathcal{C}^1(I) & \to & \mathcal{C}^0(I) \\ f & \mapsto & f' \end{array}$$

   n'est pas un morphisme d'anneaux.

Définition 121
Soient $(A, +, \times)$, $(B, \triangle, .)$ deux anneaux, er $\varphi : A \to B$ un morphisme d'anneaux. On dit que :

- φ est un endomorphisme d'anneaux si $(B, \triangle, .) = (A, +, \times)$;
- φ est un **isomorphisme d'anneaux** si φ est bijective ;
- φ est un automorphisme d'anneaux si φ est un endomorphisme bijectif.

On dit que les anneaux A et B sont **isomorphes** s'il existe un isomorphisme d'anneaux entre A et B.

Exemple 122 —

1. *Dans $\mathscr{M}_n(\mathbb{C})$, $Vect(I_n) = I_n.\mathbb{C}$ est un sous-anneau isomorphe à $\mathbb{C}$.*
2. *Dans $\mathscr{M}_n(\mathbb{K})$, l'ensemble des matrices diagonales est un sous-anneau, isomorphe à l'anneau $\mathbb{K}^n$.*
3. *Dans $\mathscr{M}_2(\mathbb{R})$, pour $A = \begin{pmatrix} 0 & -1 \\ 1 & 0 \end{pmatrix}$, $Vect(I_2, A)$ est un sous-anneau qui est isomorphe à l'anneau $\mathbb{C}$. (Prendre $f(xI_2 + yA) = x + iy$.)*
4. *Dans $\mathbb{K}[X]$, $Vect(1) = 1.\mathbb{K}$ est un sous-anneau isomorphe à $\mathbb{K}$.*
5. *L'anneau $\mathbb{Q}[i]$ n'est pas isomorphe comme anneau à $\mathbb{Q}^2$.*

Proposition 123
Soient A, B, C des anneaux.

- Pour $\varphi_1 : A \to B$, $\varphi_2 : B \to C$ des morphismes d'anneaux, alors $\varphi_2 \circ \varphi_1 : A \to C$ est un morphisme d'anneaux.
- Pour $\varphi : A \to B$ un isomorphisme d'anneaux, $\varphi^{-1} : B \to A$ est un isomorphisme d'anneaux.

Exemple 124 — *La fonction $\varphi : a + ib \in C \mapsto \begin{pmatrix} a & -b \\ b & a \end{pmatrix} \in \mathscr{M}_2(\mathbb{R})$ est un morphisme d'anneaux, qui est injectif. Ainsi, $\mathscr{M}_2(\mathbb{R})$ contient un sous-anneau commutatif qui est isomorphe (comme anneau) à $\mathbb{C}$.*

Proposition 125
Soient A, B des anneaux et $\varphi : A \to B$ un morphisme d'anneaux. Alors on a :

1. $\varphi(0_A) = 0_B$;
2. $\varphi(na) = n\varphi(a)$, $\forall a \in A$, $\forall n \in \mathbb{Z}$;
3. $\varphi(a^n) = \varphi(a)^n$, $\forall a \in A$, $\forall n \in \mathbb{N}$.

Proposition 126
Soient $(A, +, \times)$, $(B, \triangle, .)$ deux anneaux commutatifs et $\varphi : A \to B$ un morphisme d'anneaux.

1. Soit I un idéal de A. Alors $\varphi(I)$ est un idéal de $\varphi(A)$.
2. En particulier, pour $S \subset A$ on a $f(\langle S \rangle) = \langle f(S) \rangle$.
3. Soit J un idéal de B. Alors $\varphi^{-1}(J)$ est un idéal de A.
4. En particulier, $\mathrm{Ker}(\varphi)$ est un idéal de A.

Exemple 127 —

- *Pour $n \in \mathbb{N}$ et $\varphi_n : a \in \mathbb{Z} \mapsto \overline{a} \in \mathbb{Z}/n\mathbb{Z}$, on a* $\mathrm{Ker}(\varphi_n) = n\mathbb{Z}$.
  On retrouve le fait que $n\mathbb{Z}$ est un idéal de $\mathbb{Z}$.
  D'un autre côté, pour B un anneau commutatif et pour $\varphi : \mathbb{Z} \to B$ un morphisme d'anneaux, on a $\mathrm{Ker}(\varphi) = m\mathbb{Z}$ *pour un $m \in \mathbb{N}$, car ces idéaux sont les seuls idéaux de $\mathbb{Z}$. ($\mathbb{Z}$ est un anneau principal)*
- *Soit B un anneau quelconque et pour $\varphi : \mathbb{R} \to B$ un morphisme d'anneaux.*
  Comme $\varphi(\mathbb{R})$ est un sous-anneau commutatif de B, la restriction $\varphi : \mathbb{R} \to \varphi(\mathbb{R})$ est un morphisme entre anneaux commutatifs.
  Donc, $\mathrm{Ker}(\varphi)$ *est un idéal de $\mathbb{R}$. Les idéaux de $\mathbb{R}$ sont $\{0\}$ et $\mathbb{R}$.*
  Comme $\varphi(1) = 1_B \neq 0_B$, on a donc $\mathrm{Ker}(\varphi) = \{0\}$*, donc un tel morphisme est toujours injectif.*

Théorème d'isomorphisme chinois

Théorème 128 (Théorème d'isomorphisme chinois)
Soit $r \geq 2$. Soient $m_1, \ldots, m_2 \in \mathbb{N}^*$ premiers entre eux deux à deux.
Soit $\phi : (\mathbb{Z}/(m_1 \ldots m_r)\mathbb{Z}) \to (\mathbb{Z}/m_1\mathbb{Z}) \times \ldots \times (\mathbb{Z}/m_r\mathbb{Z})$ définie par $\phi(\overset{-(m_1 \ldots m_r)}{c}) = (\overset{-(m_1)}{c}, \ldots, \overset{-(m_r)}{c})$.
Alors, ϕ est un isomorphisme d'anneaux.
Pour tout $1 \leq i \leq r$, soient $u_i m_i + v_i \Pi_{j \neq i} m_j = 1$ des relations de Bézout pour m_i et $\Pi_{j \neq i} m_j$. On a alors :

$$\phi^{-1}(\overset{-(m_1)}{a_1}, \ldots, \overset{-(m_r)}{a_r}) = \sum_{i=1}^{r} \overset{-(m_1 \ldots m_r)}{a_i v_i \Pi_{j \neq i} m_j}.$$

Ainsi, les deux anneaux suivants sont isomorphes :

$$(\mathbb{Z}/(m_1 \ldots m_r)\mathbb{Z}, +, \times) \simeq ((\mathbb{Z}/m_1\mathbb{Z}) \times \ldots \times (\mathbb{Z}/m_r\mathbb{Z}), +, \times).$$

4.3 Corps

Définition 129
Soit $(\mathbb{K}, +, \times)$ un anneau.
On dit que l'anneau $\mathbb{K}$ est un **corps** s'il est commutatif et si tout élément non nul est inversible pour $\times$.

Remarque 130 —

- *Si $(K, +, \times)$ est un corps, alors $K^\times = \mathbb{K}^* := K \setminus \{0\}$.*
- *Ainsi, $(K^*, \times)$ est un groupe abélien.*

Exemple 131 — $\mathbb{R}$,$\mathbb{C}$*, et $\mathbb{Q}$ sont des corps pour les lois $+$ et $\times$ habituelles.*
$\mathbb{Q}[\sqrt{2}]$ *et $\mathbb{Q}[i]$ sont aussi des corps.*

Proposition 132
Soit $\mathbb{K}$ un corps.
Alors $\mathbb{K}$ est un anneau intègre.

Proposition 133
Soit $\mathbb{K}$ un corps.
Alors $\mathbb{K}$ est un anneau principal. Ses seuls idéaux sont $\{0\}$ et $\mathbb{K}$.

Remarque 134 —

- *La réciproque est fausse : $\mathbb{Z}$ est un anneau intègre mais ce n'est pas un corps.*
- *Un anneau commutatif A est un corps si et seulement si ses idéaux sont exactement $\{0\}$ et A, les idéaux triviaux. (Le vérifier.)*
- *Un corps est donc un cas très particulier d'anneau commutatif, intègre, principal. Ces particularités peuvent être utilisées pour construire des structures très utiles ($\mathbb{K}$-espaces vectoriels, matrices, polynômes, $\mathbb{K}$-algèbres,...).*
  Un corps est un anneau, mais il est tellement important qu'on lui consacre un chapitre à part entière.

Exemple 135 — *Pour p est nombre premier, $(\mathbb{Z}/p\mathbb{Z}, +, \times)$ est un corps.*
En effet, c'est un anneau commutatif, intègre, et fini.

Ainsi, la liste des corps usuels en mathématiques est : $\mathbb{Q}, \mathbb{R}, \mathbb{C}$, $\mathbb{Z}/p\mathbb{Z}$ pour p premier.
On peut construire beaucoup d'autres corps de plusieurs façons différentes (corps des fractions, extensions de corps), mais les corps les plus courants sont ceux-là.

4.4 Structure de $\mathbb{K}$-algèbre, Sous-algèbres

Après avoir vu groupes, anneaux, corps, et espaces vectoriels, regardons des ensembles qui possèdent à la fois une structure d'anneau et à la fois une structure de $\mathbb{K}$-espace vectoriel.

Définition 136
Soit $\mathbb{K}$ un corps.
Une $\mathbb{K}$-algèbre est un quadruplet $(A, +, \times, .)$ tel que

1. $(A, +, \times)$ est un anneau.
2. $(A, +, .)$ est un espace vectoriel.
3. $\forall \lambda \in \mathbb{K}$, $\forall x, y \in A$, $(\lambda.x)y = \lambda.(xy) = x(\lambda.y)$.

De plus, si la loi $\times$ est commutative, on dit que l'algèbre est commutative.

Donnons maintenant les exemples fondamentaux qui motivent l'étude des $\mathbb{K}$-algèbres.

Exemple 137 — *Soit $\mathbb{K}$ un corps.*

1. *L'ensemble $(\mathbb{K}[X], +, \times, .)$ (polynômes à coefficients dans $\mathbb{K}$) est une $\mathbb{K}$-algèbre commutative.*
2. *L'ensemble $(\mathcal{M}_n(\mathbb{K}), +, \times, .)$ (matrices carrées) est une $\mathbb{K}$-algèbre.*
   Elle est non commutative si $n \geq 2$.
   De même $(\mathcal{L}(E), +, \circ, .)$ (endomorphismes sur E) est une $\mathbb{K}$-algèbre.
3. *$(\mathcal{F}(X, \mathbb{K}), +, \times, .)$ (fonctions de X dans $\mathbb{K}$) est une $\mathbb{K}$-algèbre.*
4. *$\mathbb{C}$ est une $\mathbb{R}$-algèbre et une $\mathbb{Q}$-algèbre.*
5. *$\mathbb{R}$ est une $\mathbb{Q}$-algèbre.*
6. *$\mathbb{Q}[\sqrt{2}]$ est une $\mathbb{Q}$-algèbre.*

Définition 138
Soit $(A, +, \times, .)$ une $\mathbb{K}$-algèbre.
Une **sous-algèbre** B est une partie $B \subset A$ telle que

1. $(B, +, \times)$ est un sous-anneau de $(A, +, \times)$;
2. $(B, +, .)$ est un sous-espace vectoriel de $(A, +, .)$.

Proposition-Définition 139
Soient $\mathbb{K}$ un corps, A une $\mathbb{K}$-algèbre, et $a \in A$.
On définit $\mathbb{K}[a] := Vect(a^k, k \geq 0)$.
L'ensemble $\mathbb{K}[a]$ est une sous-algèbre de A, qui est commutative.
C'est la sous-algèbre engendrée par a.

4.4.1 Morphismes de $\mathbb{K}$-algèbres

Comme pour les autres objets, une fois leurs propriétés établies, on s'intéresse aux fonctions qui préservent la structure choisie.

Définition 140
Soit $(A, +, \times, .)$ et $(B, +, \times, .)$ deux $\mathbb{K}$-algèbres.
Un **morphisme d'algèbres** $\varphi : A \to B$ est une application vérifiant

1. $\varphi(1_A) = 1_B$;
2. $\forall x, y \in A$, $\varphi(xy) = \varphi(x)\varphi(y)$
3. $\forall x, y \in A$, $\forall \lambda, \mu \in \mathbb{K}$, $\varphi(\lambda x + \mu y) = \lambda\varphi(x) + \mu\varphi(y)$.

Exemple 141 —

- *La conjugaison complexe $z \mapsto \bar{z}$ est un morphisme de $\mathbb{R}$-algèbres.*
- *Sur $\mathbb{K}[X]$, $f : P \mapsto P \circ Q$ est un morphisme de $\mathbb{K}$-algèbres.*
- *Sur $\mathscr{M}_n(\mathbb{K})$, pour M inversible, $f : A \mapsto M^{-1}AM$ est un morphisme de $\mathbb{K}$-algèbres.*
- *Pour $a \in \mathbb{K}$, $f : P \in \mathbb{K}[X] \mapsto P(a) \in \mathbb{K}$ est un morphisme de $\mathbb{K}$-algèbres.*
- *Pour $x \in E$ et A une $\mathbb{K}$-algèbre, $\phi : g \in F(E, A) \mapsto g(x) \in A$ est un morphisme de $\mathbb{K}$-algèbres.*
  On l'appelle le morphisme d'évaluation en x.
- *Pour B une base de $\mathbb{K}^n$, la fonction $f \in \mathcal{L}(\mathbb{K}^n) \mapsto Mat_B(f) \in \mathscr{M}_n(\mathbb{K})$ est un morphisme de $\mathbb{K}$-algèbres.*
- *Sur $\mathbb{Q}[\sqrt{2}]$, $f : a + \sqrt{2}b \mapsto a - \sqrt{2}b$ est un morphisme de $\mathbb{K}$-algèbres.*
- *$f : P \in \mathbb{K}[X] \mapsto (x \mapsto P(x)) \in F(\mathbb{K}, \mathbb{K})$ est un morphisme de $\mathbb{K}$-algèbres.*
  Il est injectif si $\mathbb{K}$ est infini, et surjectif si $\mathbb{K}$ est fini. (Le vérifier.)

Proposition 142
Soient $\mathbb{K}$ un corps, A, B des $\mathbb{K}$-algèbres, et $f : A \to B$ un morphisme de $\mathbb{K}$-algèbres.
Soient $a \in A$ et $P(X) = a_nX^n + \ldots + a_0 \in \mathbb{K}[X]$.
• Alors, on a $f(P(a)) = P(f(a))$.
• Si $P(a) = 0$ alors on a $P(f(a)) = 0$.

Remarque 143 —
• *Ce résultat permet de déterminer l'existence ou non de morphismes de $\mathbb{K}$-algèbres entre deux algèbres A et B.*
• *Par exemple, dans la $\mathbb{R}$-algèbre $\mathbb{C}$, on a i qui est racine de $X^2 + 1$.*
Ainsi, un morphisme de $\mathbb{R}$-algèbres sur $\mathbb{C}$ doit envoyer 1 sur 1 et i sur $\pm i$. On a ainsi deux choix possibles de morphismes ($Id_{\mathbb{C}}$ et $z \mapsto \bar{z}$).
• *La $\mathbb{K}$-algèbre $\mathbb{K}[X]$ est engendrée par l'élément X en tant que $\mathbb{K}$-algèbre. (tous les polynômes sont des combinaisons linéaires de puissances de X) Ainsi, un morphisme d'algèbres $f : \mathbb{K}[X] \to B$ est entièrement déterminé par le choix de $f(X)$.*
Cela permet de montrer qu'un morphisme d'algèbres $f : \mathbb{K}[X] \to \mathbb{K}[X]$ est forcément de la forme $P \mapsto P \circ Q$. (Prendre $Q = f(X)$.)

Chapitre 5 Permutations, groupe symétrique

Définition 1
Soit $n \in \mathbb{N}^*$. L'ensemble $\mathrm{Bij}(\llbracket 1, n \rrbracket)$ des fonctions bijectives de $\{1, \dots, n\}$ dans $\{1, \dots, n\}$ est un groupe pour la composition de fonctions $\circ$. On le note $\mathcal{S}_n$.
On appelle le groupe $(\mathcal{S}_n, \circ)$ le **groupe symétrique d'ordre** n .
Un élément de $\mathcal{S}_n$ est appelé une **permutation** .

Décrire une permutation $\sigma : \{1, 2, \dots, n\} \longrightarrow \{1, 2, \dots, n\}$ revient à donner les images de chaque i par σ, pour $1 \leq i \leq n$. On peut ainsi noter σ par :

$$\begin{bmatrix} 1 & 2 & \cdots & n \\ \sigma(1) & \sigma(2) & \cdots & \sigma(n) \end{bmatrix}.$$

Par exemple, pour :

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

on a $\sigma(1) = 3$, $\sigma(2) = 7$, $\sigma(3) = 5$, $\sigma(4) = 4$, $\sigma(5) = 6$, $\sigma(6) = 1$, $\sigma(7) = 2$.

Définition 2 (Transposition)
Soit $n \geq 2$. Soient $i, j \in \llbracket 1, n \rrbracket$, avec $i \neq j$. On définit la fonction $\tau : \llbracket 1, n \rrbracket \to \llbracket 1, n \rrbracket$ par $\tau(i) = j$, $\tau(j) = i$, et $\tau(k) = k$ pour tout $k \neq i, j$.
La fonction τ est alors une permutation de $\mathcal{S}_n$. On la note $\tau = (i, j)$.
On appelle cette permutation une **transposition** .

La transposition (i, j) intervertit i et j et ne change pas les autres élément de $\llbracket 1, n \rrbracket$.

Définition 3
Soit $n \geq 2$. Soit $2 \leq p \leq n$ un entier. Soient $a_1, \dots, a_p \in \llbracket 1, n \rrbracket$ des éléments distincts. On définit la fonction $\sigma : \llbracket 1, n \rrbracket \to \llbracket 1, n \rrbracket$ par :

$$\begin{aligned} \sigma(x) &= x, \ \forall x \notin \{a_1, a_2, \dots, a_p\}\,; \\ \sigma(a_i) &= a_{i+1}, \ \forall i \in \llbracket 1, p-1 \rrbracket\,; \\ \sigma(a_p) &= a_1. \end{aligned}$$

La fonction σ est alors une permutation de $\mathcal{S}_n$. On la note $\sigma = (a_1, a_2, \dots, a_p)$.
On appelle cette permutation un **p-cycle** ou **cycle d'ordre** p .

Théorème 4
Soit $n \geq 2$.
Alors toute permutation de $\mathcal{S}_n$ peut s'écrire comme un produit de transpositions.

Exemple 5 — *Pour $a_1, \dots, a_p \in \llbracket 1, n \rrbracket$ distincts, le p-cycle $\gamma = (a_1, a_2, \dots, a_p)$ se décompose en :*

$$\gamma = (a_1, a_2)(a_2, a_3) \dots (a_{p-1}, a_p).$$

Remarque 6 — *Attention : Comme l'opération entre permutations est une composée de fonctions, pour calculer l'image de $k \in \llbracket 1, n \rrbracket$ par une tel produit de transpositions, il faut commencer par la permutation de droite. $(\sigma\tau(k) = \sigma(\tau(k)))$*

Définition 7
Soient $n \geq 2$ et $\sigma \in \mathcal{S}_n$. On définit le **support** de σ comme l'ensemble des entiers k tels que $\sigma(k) \neq k$.

Remarque 8 —

1. *Le support d'un p-cycle $(a_1, \dots, a_p)$ est l'ensemble $\{a_1, \dots, a_p\}$.*
2. *Soient σ, τ deux permutations avec des supports disjoints. Alors σ et τ commutent.* ◀

Théorème 9 (Décomposition en produit de cycles à support disjoint)
Soit $n \geq 2$. Soit $\sigma \in \mathcal{S}_n$.
Alors il existe $\sigma_1, \sigma_2, \ldots, \sigma_p$ des cycles dont les supports sont disjoints, tels que :

$$\sigma = \sigma_1 \circ \ldots \circ \sigma_p.$$

On dit que toute permutation de $\mathcal{S}_n$ se décompose en produit de cycles à supports disjoints.
De plus cette décomposition est unique à l'ordre près.

Exemple 10 — *La décomposition de* $\sigma = \left[\begin{smallmatrix} 1 & 2 & 3 & 4 & 5 & 6 & 7 & 8 \\ 5 & 2 & 1 & 8 & 3 & 7 & 6 & 4 \end{smallmatrix}\right]$ *en produit de cycles à supports disjoints est* $\sigma = (1\ 5\ 3) \circ (4\ 8) \circ (6\ 7)$.
Pour l'unicité il faut comprendre : unique à l'écriture de chaque cycle près (exemple : $(1\ 5\ 3) = (5\ 3\ 1)$*) et à l'ordre près (exemple :* $(4\ 8) \circ (1\ 5\ 3) = (1\ 5\ 3) \circ (4\ 8)$*).*

Définition 11 (Signature d'une permutation)
Soient $n \geq 2$ et $\sigma \in \mathcal{S}_n$. On définit la **signature** de σ comme le nombre

$$\varepsilon(\sigma) = \Pi_{1 \leq i < j \neq 1} \frac{\sigma(j) - \sigma(i)}{j - i}.$$

Proposition 12
La signature d'une transposition est égale à -1.

Théorème 13
Soit $n \geq 2$. Soient $\sigma, \tau \in \mathcal{S}_n$.Alors, on a :

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

La fonction $\varepsilon : \mathcal{S}_n \to \{-1, 1\}$ est un morphisme de groupes de $\mathcal{S}_n$ vers $\big(\{-1, 1\}, \times\big)$.

Définition 14
Soient $n \geq 2$ et $\sigma \in \mathcal{S}_n$.
On dit que la permutation σ est **paire** si sa signature vaut 1.
On dit que la permutation σ est **impaire** si sa signature vaut -1.

Exemple 15 — *Soit* $p \geq 2$*. La décomposition* $(a_1, a_2, \ldots, a_p) = (a_1, a_2)(a_2, a_3) \ldots (a_{p-1}, a_p)$ *prouve qu'un p-cycle a pour signature* $(-1)^{p-1}$.

Corollaire 16 (Calcul de $\varepsilon(\sigma)$)
Soit $n \geq 2$. Soit $\sigma \in \mathcal{S}_n$. Soit $\sigma = \sigma_1\,\sigma_2 \ldots \sigma_p$ la décomposition de σ en produit de cycles à support disjoint.
Soient $\alpha_1, \ldots, \alpha_p$ les longueurs de ces cycles.
Alors, on a $\varepsilon(\sigma) = (-1)^{\alpha_1 + \ldots + \alpha_p - p}$.

Définition 17 (Groupe alterné)
Soit $n \geq 1$. On pose $\mathcal{A}_n$ l'ensemble des permutations de $\mathcal{S}_n$ de signature 1 (permutations paires).
On appelle $\mathcal{A}_n$ le **groupe alterné d'ordre** n .

Remarque 18 — $\mathcal{A}_n$ *est un sous-groupe de* $\mathcal{S}_n$*, car cet ensemble est le noyau de la signature (qui est un morphisme de groupes).*

Proposition 19
Soit $n \geq 2$. Soit $\tau \in \mathcal{S}_n$ une permutation impaire. Alors, l'ensemble des permutations impaires est égal à $\mathcal{A}_n\,\tau = \{\sigma\,\tau, \sigma \in \mathcal{A}_n\}$.

Chapitre 6 Arithmétique dans $\mathbb{Z}$

Table des matières du chapitre

6.1 Divisibilité dans $\mathbb{Z}$

6.1.1 Définitions et premières propriétés

Définition 1
Soient $a, b \in \mathbb{Z}$ des entiers. On dit que a **divise** b s'il existe un entier $k \in \mathbb{Z}$ tel que $b = ka$.
Cette relation est notée $a \mid b$.

On dit aussi que a est un **diviseur** de b, ou que b est **divisible** par a, ou que b est un **multiple** de a.

Exemples 2

- 2 *divise* 6 *mais* 2 *ne divise pas* 7.
- *Soit* $a \in \mathbb{Z}$. *Alors* 1, -1, a *et* $-a$ *divisent* a.
- *Pour tout* $a \in \mathbb{Z}$ *on a* $a \mid 0$.
- *Le seul multiple de* 0 *est* 0 *: Si* $0 \mid a$ *alors* $a = 0$.

Remarque 3 — *Soit* $a \in \mathbb{Z}$. *L'ensemble des multiples de* a *est l'ensemble*

$$a\mathbb{Z} = \{ak,\ k \in \mathbb{Z}\} = \{\ldots, -3a, -2a, -a, 0, a, 2a, 3a, \ldots\}.$$

Si $a = 0$, *on a* $a\mathbb{Z} = \{0\}$.
Sinon, on peut remarquer que $|a|$ *est le plus petit entier strictement positif contenu dans* $a\mathbb{Z}$, *c'est-à-dire :* $|a| = \inf(\{k \in a\mathbb{Z},\ k > 0\})$.

Proposition 4
Soient $a, b \in \mathbb{Z}$. L'entier a divise b si et seulement si l'ensemble des multiples de b est inclus dans l'ensemble des multiples de a :

$$a \mid b \quad \Leftrightarrow \quad b\mathbb{Z} \subset a\mathbb{Z}.$$

Preuve — Supposons que $a \mid b$. Soit $m \in b\mathbb{Z}$. Alors il existe $p \in \mathbb{Z}$ tel que $m = pb$. Comme $a \mid b$, il existe $k \in \mathbb{Z}$ tel que $b = ka$. Donc $m = pka \in a\mathbb{Z}$, d'où $b\mathbb{Z} \subset a\mathbb{Z}$.

Réciproquement, supposons que $b\mathbb{Z} \subset a\mathbb{Z}$. Comme $b = 1 \times b \in b\mathbb{Z}$ on a $b \in a\mathbb{Z}$. Donc, il existe $k \in \mathbb{Z}$ tel que $b = ka$. Ainsi, a divise b, ce qui donne le résultat. □

Proposition 5
Soient $a, b \in \mathbb{Z}$ avec $b \neq 0$. Si a divise b alors $|a| \leq |b|$.

Preuve — Supposons que a divise b. Alors il existe $k \in \mathbb{Z}$ tel que $b = ka$. Comme b est non nul, k l'est également. Le nombre k étant un entier, on a ainsi $|k| \geq 1$, d'où $|b| = |ka| \geq |a|$. □

PROPOSITION 6 (Propriétés de la relation de divisibilité)
Soient $a, b, c, d \in \mathbb{Z}$.

- Si $a \mid b$ et $b \mid c$, alors $a \mid c$.
- On a les équivalences :

$$a \mid b \text{ et } b \mid a \quad \Leftrightarrow \quad a\mathbb{Z} = b\mathbb{Z} \quad \Leftrightarrow \quad |a| = |b| \quad \Leftrightarrow \quad a = b \text{ ou } a = -b.$$

- Si $a \mid b$ et $c \mid d$, alors $ac \mid bd$.
  En particulier, si $a \mid b$, alors $a^n \mid b^n$pour tout $n \in \mathbb{N}$
- Si $ab \mid c$ alors $a \mid c$ et $b \mid c$.
- Si $d \mid a$ et $d \mid b$, alors pour tous $u, v \in \mathbb{Z}$ on a $d \mid (au + bv)$.

Preuve —

- On a $k_1, k_2 \in \mathbb{Z}$ tels que $b = k_1 a$ et $c = k_2 b$. Donc $c = k_1 k_2 a$ avec $k_1 k_2 \in \mathbb{Z}$, donc a divise c.
- Si $a \mid b$ et $b \mid a$, la Proposition 4 nous donne $a\mathbb{Z} \subset b\mathbb{Z}$ et $b\mathbb{Z} \subset a\mathbb{Z}$, donc $a\mathbb{Z} = b\mathbb{Z}$.
  Supposons que $a\mathbb{Z} = b\mathbb{Z}$. Si $a = 0$ ou $b = 0$ on a alors $a\mathbb{Z} = b\mathbb{Z} = \{0\}$, donc $a = b = 0$, d'où $|a| = |b|$. Si $a \neq 0$ et $b \neq 0$, la Proposition 5 nous donne $|a| \leq |b|$ et $|b| \leq |a|$, donc $|a| = |b|$.
  Supposons que $|a| = |b|$. On a alors $a = \pm|b|$, donc $a = \pm b$, c'est-à-dire $a = b$ ou $a = -b$.
  Supposons que $a = b$ ou $a = -b$. On a alors $a \mid b$ et $b \mid a$. Cela démontre l'équivalence entre toutes ces conditions.
- Supposons que $a \mid b$ et $c \mid d$. Alors il existe $k_1 \in \mathbb{Z}$ tel que $b = ak_1$ et il existe $k_2 \in \mathbb{Z}$ tel que $d = ck_2$. Donc $bd = ack_1k_2$ et $k_1k_2 \in \mathbb{Z}$. Donc $ac \mid bd$.
- Si $ab \mid c$, alors il existe $k \in \mathbb{Z}$ tel que $c = kab = a(kb) = b(ka)$. On a donc $a \mid c$ et $b \mid c$.
- Soient $u, v \in \mathbb{Z}$. Supposons que $d \mid a$ et $d \mid b$. Alors il existe $k_1, k_2 \in \mathbb{Z}$ tels que $a = k_1 d$ et $b = k_2 d$. On a ainsi $au + bv = d(uk_1 + vk_2)$ avec $uk_1 + vk_2 \in \mathbb{Z}$, donc $d \mid (au + bv)$.

□

REMARQUE 7 — *La réciproque de l'avant-dernière proposition est fausse :* $4 \mid 12$ *et* $6 \mid 12$ *mais* $4 \times 6 = 24$ *ne divise pas* 12.

EXEMPLE 8 — *Déterminons les entiers naturels* n *tels que* $2n + 3$ *divise* $3n + 7$.

Soit $n \in \mathbb{N}$. *Supposons que* $2 + 3n \mid 3n + 7$. *Comme* $2n + 3 \mid 2n + 3$, *on a ainsi*

$$2 + 3n \mid 2(3n + 7) - 3(2n + 3) = 5.$$

Les diviseurs de 5 *sont* $1, -1, 5$ *et* -5. *Vu que* $n \in \mathbb{N}$, *on a* $2n + 3 > 0$. *On obtient donc* $2n + 3 = 1$ *ou* $2n + 3 = 5$, *soit* $n = -1$ *ou* $n = 1$. *Comme* n *est positif, on en déduit que* $n = 1$.

Réciproquement, si $n = 1$ *alors* $2n + 3 = 5$ *et* $3n + 7 = 10$, *et donc* $2n + 3 \mid 3n + 7$.

Il existe donc un unique entier naturel, $n = 1$, *tel que* $2n + 3$ *divise* $3n + 7$.

EXERCICE 2 — *Déterminer les entiers* $n \in \mathbb{Z}$ *tels que* $n + 3 \mid n^2$.

6.1.2 Division euclidienne

On rappelle que la **partie entière** (ou plancher) d'un nombre réel x, notée $\lfloor x \rfloor$ ou E, est le plus grand entier n tel que $n \leq x$, c'est-à-dire : $\lfloor x \rfloor = \sup(\{n \in \mathbb{Z} \text{ tel que } n \leq x\})$.

THÉORÈME 9 (Division euclidienne d'entiers)
Soient $a \in \mathbb{Z}$ et $b \in \mathbb{N}^*$. Alors existe un unique couple d'entiers $(q, r) \in \mathbb{Z} \times \mathbb{Z}$ tel que

$$a = bq + r \quad \text{et} \quad 0 \leq r < b.$$

L'entier q est appelé le **quotient** et l'entier r est appelé le **reste** de la division euclidienne de a par b .

Preuve —

- Existence : Posons $q = \left\lfloor \frac{a}{b} \right\rfloor$ et $r = a - qb$. Alors $(q, r) \in \mathbb{Z} \times \mathbb{N}$ et $a = bq + r$. Comme $q = \left\lfloor \frac{a}{b} \right\rfloor$, on a $q \leq \frac{a}{b} < q + 1$. Comme b est strictement positif, on obtient ainsi $bq \leq a < bq + b$. Donc, on a $0 \leq r = a - bq < b$. Ainsi, le couple (q, r) convient.
- Unicité : Soient (q_1, r_1) et (q_2, r_2) deux couples vérifiant l'énoncé. Alors on a $a = bq_1 + r_1$ et $a = bq_2 + r_2$. Cela donne $bq_1 + r_1 = bq_2 + r_2$, d'où $b(q_1 - q_2) = r_2 - r_1$.
  Comme r_1 et r_2 sont positifs, on a $|r_2 - r_1| \leq \max(r_1, r_2) < b$. Ainsi, on a $b|q_1 - q_2| < b$, donc $|q_1 - q_2| < 1$. Comme $q_1 - q_2 \in \mathbb{Z}$ on obtient $q_1 - q_2 = 0$, soit $q_1 = q2$, et donc $r_1 = r_2$, ce qui prouve l'unicité.

□

Remarque 10 — *On a montré en particulier que* $q = \left\lfloor \frac{a}{b} \right\rfloor$

Exemples 11

- *On a* $22 = 3 \times 6 + 4$ *et* $0 \leq 4 < 6$*, donc le quotient de la division euclidienne de* 22 *par* 6 *est* 3 *et le reste est* 4*.*
  Les expressions $22 = 2 \times 6 + 10$ *ou* $22 = 4 \times 6 - 2$ *ne vérifient pas la condition imposée sur le reste* r*.*
- *On a* $-12 = -3 \times 5 + 3$ *et* $0 \leq 3 < 5$*, donc le quotient de la division euclidienne de* -12 *par* 5 *est* -3 *et le reste est* 3*.*
  L'expressions $-12 = -2 \times 5 - 2$ *ne vérifie pas la condition imposée sur le reste* r*.*

Proposition 12
Soit $(a, b) \in \mathbb{Z} \times \mathbb{N}^*$. Alors b divise a si et seulement si le reste de la division euclidienne de a par b est nul.

Preuve —

- Supposons que b divise a. Alors il existe $k \in \mathbb{Z}$ tel que $a = kb$. On a donc $a = kb + 0$, et par unicité de la division euclidienne le reste de la division euclidienne de a par b vaut 0.
- Réciproquement, supposons que le reste de la division euclidienne de a par b soit nul. Alors il existe $q \in \mathbb{Z}$ tel que $a = qb + 0 = qb$. Donc b divise a.

□

6.1.3 Relation de congruence modulo un entier

Définition 13
Soient a, b et $n \in \mathbb{Z}$. On dit que a **est congru à** b **modulo** n si n divise $b - a$, ou encore, s'il existe $k \in \mathbb{Z}$ tel que $b = a + kn$. On note alors $a \equiv b \bmod n$.

Exemple 14 — *On a :* $11 \equiv 1 \bmod 5$, $\quad -1 \equiv 2 \bmod 3$, $\quad 0 \equiv 100 \bmod 2$.

Remarque 15 — *Soient* $a, n \in \mathbb{Z}$*. On a* $n \mid a \Leftrightarrow a \equiv 0 \bmod n$*.*

Proposition 16
Soient $a, b, c, n \in \mathbb{Z}$ On a :

- $a \equiv a \bmod n$ (la congruence est symétrique) ;
- Si $a \equiv b \bmod n$, alors $b \equiv a \bmod n$ (la congruence est réflexive) ;
- Si $a \equiv b \bmod n$ et $b \equiv c \bmod n$, alors $a \equiv c \bmod n$ (la congruence est transitive).

Preuve —

- On a $a - a = 0$ et n divise 0.
- Si $a \equiv b \bmod n$ alors n divise $b - a$. Donc n divise $a - b = (-1)(b - a)$, donc $b \equiv a \bmod n$.
- Il existe des entiers $k_1, k_2 \in \mathbb{Z}$ tels que $b = a + k_1 n$ et $c = b + k_2 n$. On a donc $c = a + (k_1 + k_2)n$, donc $a \equiv c \bmod n$.

□

Proposition 17 (Opérations sur les congruences)
Soient $a, b, c, d, m, n \in \mathbb{Z}$. On a :

1. $a \equiv b \bmod n$ si et seulement si $a + c \equiv b + c \bmod n$.
2. Si $a \equiv b \bmod n$ et $c \equiv d \bmod n$ alors $a + c \equiv b + d \bmod n$.
   La congruence modulo n est compatible avec la somme d'entiers.
3. Si $a \equiv b \bmod n$ alors $ac \equiv bc \bmod n$.
4. Si $a \equiv b \bmod n$ et $c \equiv d \bmod n$ alors $ac \equiv bd \bmod n$.
   La congruence modulo n est compatible avec la multiplication d'entiers.
   En particulier, si $a \equiv b \bmod n$ alors pour tout $k \in \mathbb{N}$, $a^k \equiv b^k \bmod n$.

5. Si m est non nul, alors on a $a \equiv b \bmod n$ si et seulement si $ma \equiv mb \bmod mn$.

Preuve —

1. On a $a \equiv b \bmod n$ si et seulement si n divise $b - a = (b + c) - (a + c)$, soit si et seulement si $a + c \equiv b + c \bmod n$.
2. Supposons $a \equiv b \bmod n$ et $c \equiv d \bmod n$. D'après le point précédent on a $a + c \equiv b + c \bmod n$ et $b + c \equiv b + d \bmod n$. Donc, par transitivité de la congruence modulo n, on a $a + c \equiv b + d \bmod n$.
3. Supposons $a \equiv b \bmod n$. Alors n divise $b - a$, donc n divise $c(b - a) = bc - ac$. Donc $ac \equiv bc \bmod n$.
4. Supposons $a \equiv b \bmod n$ et $c \equiv d \bmod n$. D'après le point précédent on a $ac \equiv bc \bmod n$ et $bc \equiv bd \bmod n$. Donc, par transitivité de la congruence modulo n, on a $ac \equiv bd \bmod n$.
5. Supposons que $a \equiv b \bmod n$. Alors il existe $k \in \mathbb{Z}$ tel que $b = a + kn$. Donc $mb = ma + k(mn)$ et $ma \equiv mb \bmod mn$.
   Réciproquement, supposons que $ma \equiv mb \bmod mn$. Alors il existe $k \in \mathbb{Z}$ tel que $mb = ma + kmn$. Comme m est non nul, la division par m donnc $b = a + kn$, donc $a \equiv b \bmod n$.

□

Exemples 18

- $2^{518} + 8^{211}$ *est divisible par* 3.
  Preuve — *On a* $2 \equiv -1 \bmod 3$. *Donc* $2^{518} \equiv (-1)^{518} \equiv 1 \bmod 3$. *De même,* $8 \equiv -1 \bmod 3$ *donc* $8^{211} \equiv (-1)^{211} \equiv -1 \bmod 3$. *Ainsi,*
  $$2^{518} + 8^{211} \equiv 1 - 1 \equiv 0 \bmod 3.$$
  Donc 3 *divise* $2^{518} + 8^{211}$. □
- *Déterminer les entiers* n *tels que* $3n + 5 \equiv 4 \bmod 7$.
  Soit $n \in \mathbb{Z}$. *Alors on a :*
  $$3n + 5 \equiv 4 \bmod 7 \Leftrightarrow 3n \equiv -1 \bmod 7 \Rightarrow 5 \times 3n \equiv (-1 \times 5) \bmod 7 \Leftrightarrow n \equiv 2 \bmod 7.$$
  On vérifie alors réciproquement que tous les entiers de l'ensemble $\{2 + 7k \mid k \in \mathbb{Z}\} = 2 + 7\mathbb{Z}$ *sont des solutions de* $3n + 5 \equiv 4 \bmod 7$.
  Une autre façon de prouver la réciproque est la suivante :
  $$n \equiv 2 \bmod 7 \Leftrightarrow 15n \equiv -5 \bmod 7 \Rightarrow 45n \equiv -15 \bmod 7 \Leftrightarrow 3n \equiv -1 \bmod 7 \Leftrightarrow 3n + 5 \equiv 4 \bmod 7.$$
  Donc $3n + 5 \equiv 4 \bmod 7 \Leftrightarrow n = 2 + 7k$, $k \in \mathbb{Z}$.
- *Pour tout entier* $n \in \mathbb{Z}$ *impair,* 8 *divise* $n^2 - 1$.
  Preuve — *En effet, soit* n *un entier impair. Il existe donc* $k \in \mathbb{Z}$ *tel que* $n = 2k + 1$. *Alors* $n^2 - 1 = 4k^2 + 4k = 4k(k + 1)$. *Or* k *et* $k + 1$ *étant deux entiers successifs, l'un d'entre eux est pair et donc* $k(k + 1)$ *est pair. Donc* $k(k + 1) \equiv 0 \bmod 2$. *Donc* $4k(k + 1) \equiv 0 \bmod 8$. *Donc* $n^2 - 1 \equiv 0 \bmod 8$. *D'où le résultat.* □

6.2 PGCD, PPCM

6.2.1 Plus grand diviseur commun

Définition 19
Soient $a_1, \ldots, a_n$ des éléments de $\mathbb{Z}$. On appelle **diviseur commun** de $a_1, \ldots, a_n$ tout élément $d \in \mathbb{Z}$ tel que $d \mid a_i$ pour tout $i \in \{1, \ldots, n\}$.

Exemples 20

- 6 *est un diviseur commun de* 12 *et* 18.
- 3 *est un diviseur commun de* 9, 12 *et* 21.

Lemme 21
Soient a et b deux éléments de $\mathbb{Z}$. L'ensemble $a\mathbb{Z} + b\mathbb{Z} = \{ak_1 + bk_2 \mid (k_1, k_2) \in \mathbb{Z}^2\}$ vérifie les propriétés suivantes :

- $0 \in a\mathbb{Z} + b\mathbb{Z}$;
- Pour $x \in a\mathbb{Z} + b\mathbb{Z}$, on a $-x \in a\mathbb{Z} + b\mathbb{Z}$;
- Pour $x, y \in a\mathbb{Z} + b\mathbb{Z}$, on a $x + y \in a\mathbb{Z} + b\mathbb{Z}$.

L'ensemble $a\mathbb{Z} + b\mathbb{Z}$ est ainsi un sous-groupe de $(\mathbb{Z}, +)$ (voir chapitre Structures algébriques).

Preuve — D'après sa définition, $a\mathbb{Z} + b\mathbb{Z}$ est un sous-ensemble de $\mathbb{Z}$.

- On a $0 = 0 \times a + 0 \times b$, donc $0 \in a\mathbb{Z} + b\mathbb{Z}$.

- Soit $x \in a\mathbb{Z} + b\mathbb{Z}$. Il existe k_1,k_2 dans $\mathbb{Z}$ tels que $x = ak_1 + bk_2$.
  On a alors $-x = -ak_1 - bk_2 = a(-k_1) + b(-k_2)$, donc $-x \in a\mathbb{Z} + b\mathbb{Z}$.
- Soit $(x, y) \in (a\mathbb{Z} + b\mathbb{Z})^2$. Il existe k_1,k_2,k_3,k_4 dans $\mathbb{Z}$ tels que $x = ak_1 + bk_2$ et $y = ak_3 + bk_4$.
  On a alors $x + y = ak_1 + bk_2 + (ak_3 + bk_4) = a(k_1 + k_3) + b(k_2 + k_4)$, donc $x + y \in a\mathbb{Z} + b\mathbb{Z}$.

□

PROPOSITION 22
Soient $a, b \in \mathbb{Z}$. Alors il existe un unique entier naturel $d \in \mathbb{N}$ tel que $a\mathbb{Z} + b\mathbb{Z} = d\mathbb{Z}$.

Preuve — • Si $a = 0$ et $b = 0$, on a $a\mathbb{Z} + b\mathbb{Z} = \{0\} + \{0\} = \{0\}$. Pour $d = 0$ on a ainsi $d\mathbb{Z} = 0\mathbb{Z} = \{0\} = a\mathbb{Z} + b\mathbb{Z}$. Pour tout autre entier $c \neq 0$, l'ensemble $c\mathbb{Z}$ contient une infinité d'éléments. Ainsi $d = 0$ est l'unique entier vérifiant $a\mathbb{Z} + b\mathbb{Z} = d\mathbb{Z}$.

• Supposons maintenant que $a \neq 0$ ou $b \neq 0$.

Existence : L'ensemble $a\mathbb{Z} + b\mathbb{Z}$ contient $|a| = \pm a + 0.b$ et $|b| = 0.a + \pm b$, donc il contient au moins un entier strictement positif. L'ensemble $(a\mathbb{Z} + b\mathbb{Z}) \cap \mathbb{N}^* = \{n \in a\mathbb{Z} + b\mathbb{Z},\ n > 0\}$ est donc une sous-partie de $\mathbb{N}$ qui est non-vide. Cet ensemble admet donc un plus petit élément, que l'on note d. Montrons par double inclusion que $a\mathbb{Z} + b\mathbb{Z} = d\mathbb{Z}$.

◁ Comme $d \in a\mathbb{Z} + b\mathbb{Z}$, il existe $u, v \in \mathbb{Z}$ tels que $d = au + bv$. Pour tout $k \in \mathbb{Z}$, on a donc $kd = a(ku) + b(kv) \in a\mathbb{Z} + b\mathbb{Z}$. Donc, $d\mathbb{Z} \subset a\mathbb{Z} + b\mathbb{Z}$.

▷ Pour l'inclusion réciproque, prenons $c \in a\mathbb{Z} + b\mathbb{Z}$. On effectue la division euclidienne de c par d : $c = dq + r$, avec $0 \leq r < d$. Alors $r = c - dq$ appartient à $a\mathbb{Z} + b\mathbb{Z}$ d'après la proposition précédente. On doit alors avoir $r = 0$ par minimalité de d.

Ainsi, on a $c = dq + 0$, donc c est un multiple de d, donc $c \in d\mathbb{Z}$. Cela donne $a\mathbb{Z} + b\mathbb{Z} \subset d\mathbb{Z}$, et donc $a\mathbb{Z} + b\mathbb{Z} = d\mathbb{Z}$.

Unicité : Supposons avoir $d, d' \in \mathbb{N}$ tels que $d\mathbb{Z} = a\mathbb{Z} + b\mathbb{Z} = d'\mathbb{Z}$. On a donc $d\mathbb{Z} = d'\mathbb{Z}$. La Proposition 6 nous donne alors $d = d'$, ce qui conclut la preuve. □

DÉFINITION 23
Soient $a, b \in \mathbb{Z}$. L'unique entier $d \in \mathbb{N}$ tel que $a\mathbb{Z} + b\mathbb{Z} = d\mathbb{Z}$ est appelé le **plus grand diviseur commun** de a et b (en abrégé pgcd). On le note $d = \text{pgcd}(a, b)$ ou encore $d = a \wedge b$.

REMARQUE 24 — *Si $a\mathbb{Z} + b\mathbb{Z} = d\mathbb{Z}$, alors tout élément de $a\mathbb{Z} + b\mathbb{Z}$ est un multiple de d. Comme $a\mathbb{Z} + b\mathbb{Z}$ contient $a = 1.a + 0.b$ et $b = 0.a + 1.b$, on remarque en particulier que a et b dont des multiples de d, c'est-à-dire que d est un diviseur commun à a et à b.*

PROPOSITION 25
Soient $a, b \in \mathbb{Z}$. Alors $d = \text{pgcd}(a, b)$ si et seulement si :

1. $d \mid a$ et $d \mid b$,
   Autrement dit, d est un diviseur commun à a et b.
2. Pour tout $d' \in \mathbb{Z}$ tel que $d' \mid a$ et $d' \mid b$, on a $d' \mid d$.
   Autrement dit, tout diviseur commun de a et b divise d.

Preuve —

• Supposons que $d = \text{pgcd}(a, b)$. On a $a\mathbb{Z} + b\mathbb{Z} = d\mathbb{Z}$. Le premier point vient de la remarque précédente.

Soit $d' \in \mathbb{Z}$ tel que $d' \mid a$ et $d' \mid b$. La Proposition 4 donne $a\mathbb{Z} \subset d'\mathbb{Z}$ et $b\mathbb{Z} \subset d'\mathbb{Z}$ donc $a\mathbb{Z} + b\mathbb{Z} \subset d'\mathbb{Z}$. Donc $d\mathbb{Z} \subset d'\mathbb{Z}$, et $d' \mid d$. D'où le second point.

• Réciproquement, supposons 1) et 2). Montrons par double inclusion que $a\mathbb{Z} + b\mathbb{Z} = d\mathbb{Z}$.

▷ Comme $d \mid a$ et $d \mid b$, la Proposition 4 donne $a\mathbb{Z} \subset d\mathbb{Z}$ et $b\mathbb{Z} \subset d\mathbb{Z}$, donc $a\mathbb{Z} + b\mathbb{Z} \subset d\mathbb{Z}$.

◁ Soit $d' = pgcd(a, b)$. Alors on a $d' \mid a$ et $d' \mid b$. L'hypothèse 2 donne $d' \mid d$, donc $d\mathbb{Z} \subset d'\mathbb{Z} = a\mathbb{Z} + b\mathbb{Z}$.

Finalement, on a montré que $d\mathbb{Z} = a\mathbb{Z} + b\mathbb{Z}$, ce qui conclut. □

REMARQUE 26 — *Le pgcd de a et b est donc bien le plus grand diviseur commun pour la relation $\leq$ sur $\mathbb{N}$.*

REMARQUE 27 — *Pour une famille d'entiers a_1,...,$a_n \in \mathbb{Z}$ on peut démontrer par récurrence qu'il existe un unique $d \in \mathbb{N}$ tel que $a_1\mathbb{Z} + \ldots + a_n\mathbb{Z} = d\mathbb{Z}$, et définir ainsi le pgcd de la famille $a_1, \ldots, a_n$.*

On peut alors démontrer de même que ce pgcd est un diviseur commun de a_1, ..., a_n qui est un multiple de tout autre diviseur commun.

EXEMPLES 28— *On a :* $\text{pgcd}(12, 18) = 6$, $\text{pgcd}(10, 12, 18) = 2$, $\text{pgcd}(2, 3) = 1$, $\text{pgcd}(8, 6) = 2$.

PROPOSITION 29
Soient $a, b, c, k \in \mathbb{Z}$. On a :

- $\text{pgcd}(a,b) = \text{pgcd}(|a|,|b|)$,
- $\text{pgcd}(a,0) = |a|$,
- $\text{pgcd}(a,1) = 1$,
- $\text{pgcd}(a,b) = \text{pgcd}(b,a)$,
- $\text{pgcd}(a,\text{pgcd}(b,c)) = \text{pgcd}(\text{pgcd}(a,b),c)$,
- $\text{pgcd}(ka,kb) = |k|\text{pgcd}(a,b)$.

Preuve — Ces propriétés découlent immédiatement de la définition du pgcd comme l'unique entier positif d tel que $d\mathbb{Z} = a\mathbb{Z} + b\mathbb{Z}$. En effet, on a :

1. $a\mathbb{Z} + b\mathbb{Z} = |a|\mathbb{Z} + |b|\mathbb{Z}$,
2. $a\mathbb{Z} + 0\mathbb{Z} = a\mathbb{Z}$,
3. $a\mathbb{Z} + \mathbb{Z} = \mathbb{Z}$,
4. $d\mathbb{Z} = a\mathbb{Z} + b\mathbb{Z} = b\mathbb{Z} + a\mathbb{Z}$,
5. $a\mathbb{Z} + (b\mathbb{Z} + c\mathbb{Z}) = (a\mathbb{Z} + b\mathbb{Z}) + c\mathbb{Z}$,
6. $ak\mathbb{Z} + bk\mathbb{Z} = |k|(a\mathbb{Z} + b\mathbb{Z})$.

□

Proposition 30
Soient $a, b \in \mathbb{Z}$, et $d = \text{pgcd}(a,b)$.
Alors il existe deux entiers $u_0, v_0 \in \mathbb{Z}$ tels que

$$au_0 + bv_0 = d.$$

Preuve — Par définition on a $d\mathbb{Z} = a\mathbb{Z} + b\mathbb{Z}$. Donc $d \in a\mathbb{Z} + b\mathbb{Z}$. Il existe donc $(u_0, v_0) \in \mathbb{Z}^2$ tel que $d = au_0 + bv_0$. □

Exemple 31 — *On a* $\text{pgcd}(4,6) = 2$ *et* $4 \times (-1) + 6 \times 1 = 2$. *On a aussi* $4 \times 2 + 6 \times (-1) = 2$.

On peut donc remarquer que les entiers u et v ne sont pas uniques.

6.2.2 Calcul du PGCD avec l'algorithme d'Euclide

Le pgcd de deux entiers se calcule facilement de manière algorithmique. Ce calcul est basé sur le résultat suivant.

Lemme 32
Soient $a\mathbb{Z}$, $b \in \mathbb{N}^*$. Notons r le reste de la division euclidienne de a par b. Alors on a :

$$\text{pgcd}(a,b) = \text{pgcd}(b,r).$$

Preuve — Par division euclidienne, il existe $q \in \mathbb{Z}$ tel que $a = bq + r$. On vérifie alors que $a\mathbb{Z} + b\mathbb{Z} = r\mathbb{Z} + b\mathbb{Z}$. Par définition du pgcd, on a donc $\text{pgcd}(a,b) = \text{pgcd}(b,r)$. □

L'algorithme d'Euclide permet de calculer le pgcd de deux entiers, il est basé sur des divisions euclidiennes successives.

Principe de l'algorithme d'Euclide

Soient a et b deux entiers tels que $0 \leq b \leq a$.

Si $b = 0$ alors $\text{pgcd}(a,b) = a$ *et c'est terminé. On suppose donc b non nul.*

- Étape 1 : *On effectue la division euclidienne de a par b : $a = bq_0 + r_0$ avec $0 \leq r_0 < b$.*
  D'après le lemme, $\text{pgcd}(a,b) = \text{pgcd}(b,r_0)$.
  Si $r_0 = 0$ alors $\text{pgcd}(a,b) = b$ *et c'est terminé.*
  Sinon, on passe à l'étape suivante.
- Étape 2 : *On effectue la division euclidienne de b par r_0 : $b = r_0q_1 + r_1$ avec $0 \leq r_1 < r_0$.*
  D'après le lemme, $\text{pgcd}(a,b) = \text{pgcd}(b,r_0) = \text{pgcd}(r_0,r_1)$.
  Si $r_1 = 0$ alors $\text{pgcd}(r_0,r_1) = r_0$ *et donc* $\text{pgcd}(a,b) = r_0$ *et c'est terminé.*
  Sinon, on passe à l'étape suivante.
- Étape 3 : *On effectue la division euclidienne de r_0 par r_1 : $r_0 = r_1q_2 + r_2$ avec $0 \leq r_2 < r_1$.*
  D'après le lemme, $\text{pgcd}(a,b) = \text{pgcd}(r_0,r_1) = \text{pgcd}(r_1,r_2)$.
  Si $r_2 = 0$ alors $\text{pgcd}(r_1,r_2) = r_1$ *et donc* $\text{pgcd}(a,b) = r_1$ *et c'est terminé.*
  Sinon on passe à l'étape suite, etc.
- ...

La suite des restes obtenus est une suite strictement décroissante d'entiers positifs, il existe donc un entier $n_0 \in \mathbb{N}$ tel que $r_{n_0} = 0$. D'après le lemme précédent, on a : $\text{pgcd}(a,b) = \text{pgcd}(b,r_0) = \ldots = \text{pgcd}(r_{n_0-1},r_{n_0}) = r_{n_0-1}$.

Remarque 33 — *On peut toujours se ramener au cas où* $0 \leq b \leq a$ *en utilisant le fait que* $\text{pgcd}(a,b) = \text{pgcd}(b,a)$ *et que* $\text{pgcd}(a,b) = \text{pgcd}(|a|,|b|)$.

Exemple 34 — *Calculons le pgcd de* 721 *et* 658 *à l'aide de l'algorithme d'Euclide.*

1. *Division euclidienne de* 721 *par* 658 *:* $721 = 658 \times 1 + 63$. *Le reste vaut* 63.
2. *Division euclidienne de* 658 *par* 63 *:* $658 = 63 \times 10 + 28$. *Le reste vaut* 28.
3. *Division euclidienne de* 63 *par* 28 *:* $63 = 28 \times 2 + 7$. *Le reste vaut* 7.
4. *Division euclidienne de* 28 *par* 7 *:* $28 = 7 \times 4 + 0$. *Le reste est nul !*

Le dernier reste non nul dans la suite des divisions euclidienne est donc 7. *Ainsi,* $\text{pgcd}(721, 658) = 7$.

6.2.3 Plus petit multiple commun

Définition 35
Soient $a_1,\ldots,a_n \in \mathbb{Z}$. On appelle **multiple commun** de $a_1,\ldots,a_n$ tout élément m de $\mathbb{Z}$ tel que m est un multiple de a_i (ou encore, $a_i \mid m$) pour tout $i \in \{1,\ldots,n\}$.

Exemples 36

- 12 *est un multiple commun de* 4 *et* 6.
- 36 *est un multiple commun de* 2, 3 *et* 9.
- 105 *est un multiple commun de* 3, 5 *et* 7.

Proposition 37
Soient $a, b \in \mathbb{Z}$. Alors il existe un unique entier positif $m \in \mathbb{N}$ tel que $a\mathbb{Z} \cap b\mathbb{Z} = m\mathbb{Z}$.

Preuve — • Si $a = 0$ et $b = 0$, on a $a\mathbb{Z} \cap b\mathbb{Z} = \{0\} \cap \{0\} = \{0\}$. Pour $m = 0$ on a ainsi $m\mathbb{Z} = 0\mathbb{Z} = \{0\} = a\mathbb{Z} \cap b\mathbb{Z}$. Pour tout autre entier $c \neq 0$, l'ensemble $c\mathbb{Z}$ contient une infinité d'éléments. Ainsi $m = 0$ est l'unique entier vérifiant $a\mathbb{Z} \cap b\mathbb{Z} = m\mathbb{Z}$.

• Supposons maintenant que $a \neq 0$ ou $b \neq 0$.

Existence : L'entier positif $|ab|$ est dans $a\mathbb{Z}$ et dans $b\mathbb{Z}$, donc $|ab| \in a\mathbb{Z} \cap b\mathbb{Z}$. Cet ensemble contient donc au moins un entier strictement positif. L'ensemble $(a\mathbb{Z} \cap b\mathbb{Z}) \cap \mathbb{N}^* = \{n \in a\mathbb{Z} \cap b\mathbb{Z},\ n > 0\}$ est donc une sous-partie de $\mathbb{N}$ qui est non-vide. Cet ensemble admet donc un plus petit élément, que l'on note m. Montrons par double inclusion que $a\mathbb{Z} \cap b\mathbb{Z} = m\mathbb{Z}$.

◁ Comme $m \in a\mathbb{Z} \cap b\mathbb{Z}$, il existe $u, v \in \mathbb{Z}$ tels que $m = au$ et $m = bv$. Pour tout $k \in \mathbb{Z}$, on a donc $km = a(ku) \in a\mathbb{Z}$ et $km = b(kv) \in b\mathbb{Z}$. Donc, $m\mathbb{Z} \subset a\mathbb{Z} \cap b\mathbb{Z}$.

▷ Pour l'inclusion réciproque, prenons $c \in a\mathbb{Z} \cap b\mathbb{Z}$. On effectue la division euclidienne de c par m : $c = mq + r$, avec $0 \leq r < m$. Si l'on avait $r > 0$, alors $r = c - mq$ appartiendrait à $a\mathbb{Z}$ et à $b\mathbb{Z}$, donc à $a\mathbb{Z} \cap b\mathbb{Z}$. Mais comme $0 < r < m$, cela contredirait le fait que m est le plus petit entier strictement positif contenu dans $a\mathbb{Z} \cap b\mathbb{Z}$.

Ainsi, on a $c = mq + 0$, donc c est un multiple de m, donc $c \in m\mathbb{Z}$. Cela donne $a\mathbb{Z} \cap b\mathbb{Z} \subset m\mathbb{Z}$, et donc $a\mathbb{Z} \cap b\mathbb{Z} = m\mathbb{Z}$.

Unicité : Supposons avoir $m, m' \in \mathbb{N}$ tels que $m\mathbb{Z} = a\mathbb{Z} \cap b\mathbb{Z} = m'\mathbb{Z}$. On a donc $m\mathbb{Z} = m'\mathbb{Z}$. La Proposition 6 nous donne alors $m = m'$, ce qui conclut la preuve. □

Définition 38
Soient a, b des éléments de $\mathbb{Z}$. L'unique entier naturel $m \in \mathbb{N}$ tel que $a\mathbb{Z} \cap b\mathbb{Z} = m\mathbb{Z}$ est appelé le **plus petit multiple commun** de a et de b (en abrégé ppcm).
On le note $m = \text{ppcm}(a,b)$ ou encore $m = a \vee b$.

Proposition 39
Soient $a, b \in \mathbb{Z}$ et $m \in \mathbb{N}$. On a $m = \text{ppcm}(a,b)$ si et seulement si :

1. $a \mid m$ et $b \mid m$,
   Autrement dit, m est un multiple commun de a et de b.
2. Pour tout $m' \in \mathbb{Z}$ tel que $a \mid m'$ et $b \mid m'$, on a $m \mid m'$.
   Autrement dit, tout multiple commun de a et de b est un multiple de m.

Preuve —

• Supposons que $m = \text{ppcm}(a,b)$, c'est-à-dire $a\mathbb{Z} \cap b\mathbb{Z} = m\mathbb{Z}$. Comme $m \in m\mathbb{Z} \subset a\mathbb{Z}$, on a $a \mid m$. On obtient de même $b \mid m$. D'où le premier point.

Soit $m' \in \mathbb{Z}$ tel que $a \mid m'$ et $b \mid m'$. Alors $m' \in a\mathbb{Z}$ et $m' \in b\mathbb{Z}$, donc $m' \in a\mathbb{Z} \cap b\mathbb{Z} = m\mathbb{Z}$. Donc $m \mid m'$.

• Réciproquement, supposons 1) et 2). Montrons par double inclusion que $mZ = a\mathbb{Z} \cap b\mathbb{Z}$.

▷ Le premier point donne $m \in a\mathbb{Z}$ et $m \in b\mathbb{Z}$, donc $m \in a\mathbb{Z} \cap b\mathbb{Z}$. Ainsi, on a $m\mathbb{Z} \subset a\mathbb{Z} \cap b\mathbb{Z}$.

◁ Soit $m' = \text{ppcm}(a,b)$. On a alors $a \mid m'$ et $b \mid m'$. Le point 2) nous donne $m \mid m'$. Ainsi, on a $m'\mathbb{Z} \subset m\mathbb{Z}$, soit $a\mathbb{Z} \cap b\mathbb{Z} = m'\mathbb{Z} subset m\mathbb{Z}$.

Finalement, on obtient $a\mathbb{Z} \cap b\mathbb{Z} = m\mathbb{Z}$, ce qui conclut la preuve. □

REMARQUE 40 — *Le ppcm de a et b est donc bien le plus petit multiple commun pour la relation $\leq$ sur $\mathbb{N}$.*

REMARQUE 41 — *Pour une famille d'entiers $a_1,\ldots,a_n \in \mathbb{Z}$ on peut démontrer par récurrence qu'il existe un unique $m \in \mathbb{N}$ tel que $a_1\mathbb{Z} \cap \ldots \cap a_n\mathbb{Z} = m\mathbb{Z}$, et définir ainsi le ppcm de la famille $a_1, \ldots, a_n$.*

On peut alors démontrer de même que ce ppcm est un multiple commun de a_1, …, a_n qui est un diviseur de tout autre multiple commun.

EXEMPLE 42 — $\text{ppcm}(3,6) = 6$, $\text{ppcm}(4,6) = 12$, $\text{ppcm}(2,3) = 6$.

PROPOSITION 43
Soient $a, b, k \in \mathbb{Z}$. On a :

1. $\text{ppcm}(a,b) = \text{ppcm}(|a|,|b|)$,
2. $\text{ppcm}(a,0) = 0$,
3. $\text{ppcm}(1,a) = |a|$,
4. $\text{ppcm}(a,b) = \text{ppcm}(b,a)$,
5. $\text{ppcm}(a,\text{ppcm}(b,c)) = \text{ppcm}(\text{ppcm}(a,b),c)$,
6. $\text{ppcm}(ka,kb) = |k|\text{ppcm}(a,b)$.

Preuve — Ces propriétés découlent de la définition du ppcm. □

Preuve — Ces propriétés découlent immédiatement de la définition du ppcm comme l'unique entier positif m tel que $m\mathbb{Z} = a\mathbb{Z} \cap b\mathbb{Z}$. En effet, on a :

1. $a\mathbb{Z} \cap b\mathbb{Z} = |a|\mathbb{Z} \cap |b|\mathbb{Z}$,
2. $a\mathbb{Z} \cap 0\mathbb{Z} = 0\mathbb{Z}$,
3. $a\mathbb{Z} \cap \mathbb{Z} = a\mathbb{Z}$,
4. $m\mathbb{Z} = a\mathbb{Z} \cap b\mathbb{Z} = b\mathbb{Z} \cap a\mathbb{Z}$,
5. $a\mathbb{Z} \cap (b\mathbb{Z} \cap c\mathbb{Z}) = (a\mathbb{Z} \cap b\mathbb{Z}) \cap c\mathbb{Z}$,
6. $ak\mathbb{Z} \cap bk\mathbb{Z} = |k|(a\mathbb{Z} \cap b\mathbb{Z})$.

□

6.3 Théorème de Bézout et théorème de Gauss

6.3.1 Nombres entiers premiers entre eux

DÉFINITION 44
Soient $a, b \in \mathbb{Z}$. On dit que a et b sont **premiers entre eux** si $\text{pgcd}(a,b) = 1$.

EXEMPLE 45 — *2 et 3 sont premiers entre eux. 9 et 16 sont premiers entre eux. 6 et 4 ne sont pas premiers entre eux.*

Si a ne divise pas b et b ne divise pas a, on ne peut pas dire que a et b sont premiers entre eux !
Par exemple, 6 ne divise pas 15 et 15 ne divise pas 6 mais $\text{pgcd}(6,15) = 3$, donc 6 et 15 ne sont pas premiers entre eux.

DÉFINITION 46
Soient $a_1,\ldots,a_n \in \mathbb{Z}$. On dit que $a_1,\ldots,a_n$ sont **premiers entre eux deux à deux** si $\text{pgcd}(a_i,a_j) = 1$ pour tous i, j dans $\{1,\ldots,n\}$ avec $i \neq j$.

6.3.2 Théorème de Bézout et théorème de Gauss

THÉORÈME 47 (Théorème de Bézout)
Soient $a, b \in \mathbb{Z}$. Les entiers a et b sont premiers entre eux si et seulement s'il existe deux entiers $u, v \in Z$ tels que

$$au + bv = 1.$$

Preuve — Supposons que a et b sont premiers entre eux. Alors on a $\text{pgcd}(a,b) = 1$. Cela veut dire que $a\mathbb{Z} + b\mathbb{Z} = \mathbb{Z}$. Ainsi, il existe $u, v \in \mathbb{Z}$ tels que $au + bv = 1$.

Réciproquement, supposons qu'il existe $u, v \in \mathbb{Z}$ tels que $au + bv = 1$. On a alors $1 \in a\mathbb{Z} + b\mathbb{Z}$, donc $1\mathbb{Z} = \mathbb{Z} \subset a\mathbb{Z} + b\mathbb{Z}$. Comme $a\mathbb{Z} + b\mathbb{Z} \subset \mathbb{Z}$, on en déduit que $a\mathbb{Z} + b\mathbb{Z} = 1\mathbb{Z}$. Donc $\text{pgcd}(a,b) = 1$, et a et b sont premiers entre eux. □

EXEMPLE 48 — *n et $n+1$ sont premiers entre eux car $(n+1)\times 1 + n\times(-1) = 1$.*

L'algorithme d'Euclide étendu permet d'obtenir les coefficients u et v, appelés **coefficients de Bézout**. Alors que l'algorithme d'Euclide s'intéresse uniquement aux restes de divisions euclidiennes successives, l'algorithme d'Euclide étendu considère également les quotients de ces divisions euclidiennes.

PRINCIPE : À l'aide de l'algorithme d'Euclide, on construit de proche en proche des éléments u_k et v_k de $\mathbb{Z}$ tels que l'ont ait à chaque étape :

$$r_k = au_k + bv_k,$$

où les r_k sont les restes des divisions euclidiennes successives de l'algorithme d'Euclide.

EXEMPLE 49 — *Expliquons sur un exemple, avec $a = 1795$ et $b = 343$.*

$$\begin{array}{rcccccc|l}
343 & = & 0 & \times 1795 & + & 1 & \times 343 & \\
80 & = & 1 & \times 1795 & + & (-5) & \times 343 & 1795 - 5\times 343 = 80 \\
23 & = & 0+(-4).1 & \times 1795 & + & 1+(-4)(-5) & \times 343 & 343 - 4\times 80 = 23 \\
23 & = & -4 & \times 1795 & + & 21 & \times 343 & \\
11 & = & 1+(-3)(-4) & \times 1795 & + & (-5)+(-3)(21) & \times 343 & 80 - 3\times 23 = 11 \\
11 & = & 13 & \times 1795 & + & (-68) & \times 343 & \\
1 & = & (-4)+(-2)(13) & \times 1795 & + & 21+(-2)(-68) & \times 343 & 23 - 2\times 11 = 1 \\
1 & = & -30 & \times 1795 & + & 157 & \times 343 &
\end{array}$$

L'algorithme d'Euclide (à droite) nous dit que $\text{pgcd}(1795, 343) = 1$. *Et on a obtenu en même temps (à gauche) que $1 = 1795\times u + 343\times v$ avec $u = -30$ et $v = 157$.*

THÉORÈME 50 (Théorème de Gauss)
Soient $a, b, c \in \mathbb{Z}$ des éléments de $\mathbb{Z}$.
Si a et b sont premiers entre eux et si $a \mid bc$, alors $a \mid c$.

Preuve — On suppose que a et b sont premiers entre eux et que $a \mid bc$. D'après le théorème de Bézout, il existe des entiers $u, v \in \mathbb{Z}$ tels que $au + bv = 1$. Comme $a \mid bc$, il existe $k \in \mathbb{Z}$ tel que $bc = ak$. On a donc $c = auc + bvc = auc + avk = a(uc + vk)$. Donc $a \mid c$. □

Quand a et b ne sont pas premiers entre eux, si $a \mid bc$ et même si a ne divise pas b, on ne peut pas dire que $a \mid c$!
Par exemple, 8 divise 4×6 mais 8 ne divise ni 4 ni 6.

EXEMPLE 51 — *Si $4 \mid 3n$ alors $4 \mid n$, car 4 et 3 sont premiers entre eux.*

Conséquences de ces théorèmes

PROPOSITION 52
Soient $a, b, c \in \mathbb{Z}$.
Si a et b premiers entre eux et si $a \mid c$ et $b \mid c$, alors $ab \mid c$.

Preuve — Supposons que $a \mid c$ et $b \mid c$ avec $\text{pgcd}(a, b) = 1$. Le théorème de Bézout nous dit qu'il existe alors des entiers $u, v \in \mathbb{Z}$ tels que $au + bv = 1$. On a donc $c = auc + bvc$. Comme $a \mid c$, on a on a $ab \mid bc$. Comme $b \mid c$, on a on a $ab \mid ac$. Donc, ab divise $acu + bcv = c$. □

COROLLAIRE 53
Soient $a_1, \ldots, a_n, c \in \mathbb{Z}$.
Si les a_i sont premiers entre eux deux à deux et si $a_i \mid c$ pour tout $1 \leq i \leq n$, alors $a_1 \times a_2 \times \ldots a_n \mid c$.

Preuve — Cette généralisation de la proposition précédente se démontre par récurrence sur n. □

Si a et b ne sont pas premiers entre eux, on ne peut rien dire ! Par exemple $4 \mid 4$ et $2 \mid 4$ mais $4 \times 2 = 8$ ne divise pas 4.

EXEMPLE 54 — *Si $4 \mid n$ et $3 \mid n$ alors $12 \mid n$ car 4 et 3 sont premiers entre eux.*

PROPOSITION 55
Soient $a, b, c \in \mathbb{Z}$.
Si a est premier avec b et si a est premier avec c, alors a est premier avec bc.

Preuve — Supposons a premier avec b et avec c. D'après le théorème de Bézout, il existe $u_1, u_2, v_1, v_2 \in \mathbb{Z}$ tels que $1 = au_1 + bv_1$ et $1 = au_2 + cv_2$. Par multiplication, on obtient :

$$1 = a(au_1u_2 + u_1cv_2 + bv_1u_2) + bc(v_1v_2).$$

Ainsi, d'après le théorème de Bézout, a et bc sont premiers entre eux. □

COROLLAIRE 56
Soient $a, b_1, \ldots, b_n \in \mathbb{Z}$. Si a est premier avec b_i pour tout $i \in \{1, \ldots, n\}$, alors a est premier avec $b_1 \times b_2 \times \ldots \times b_n$.

Preuve — Cette généralisation de la proposition précédente se démontre par récurrence sur n. □

PROPOSITION 57
Soient $a, b \in \mathbb{Z}$. Si a est premier avec b alors a^m est premier avec b^n pour tous $m, n \in \mathbb{N}$.

Preuve — Soient $m, n \in \mathbb{N}$. Si $m = 0$ ou $n = 0$ on a $a^m = 1$ ou $b^n = 1$, et dans ce cas le résultat est vrai.

Supposons $m, n \neq 0$. Comme a est premier avec b, le corollaire précédent nous dit que a est premier avec b^n. Comme b^n est premier avec a, le corollaire précédent nous dit que b^n est premier avec a^m. □

EXEMPLE 58 — *Soit $n \in \mathbb{N}^*$. Comme n est premier avec $n-1$ et avec $n+1$, n est premier avec $(n-1)(n+1) = n^2-1$.*

PROPOSITION 59
Soient $a, b \in \mathbb{Z}$. Posons $d = \text{pgcd}(a, b)$. Alors il existe des éléments a' et b' de $\mathbb{Z}$ tels que

$$a = da',\ b = db', \quad \text{et} \quad \text{pgcd}(a', b') = 1.$$

Preuve — Si $(a, b) = (0, 0)$, alors $a' = b' = 1$ conviennent.

Supposons que $(a, b) \neq (0, 0)$. Comme $d = \text{pgcd}(a, b)$, on sait que $d \mid a$ et $d \mid b$. Les nombres $a' = \frac{a}{\text{pgcd}(a,b)}, b' = \frac{b}{\text{pgcd}(a,b)}$ sont donc des entiers, tels que $a = da'$ et $b = db'$. On a alors $d = \text{pgcd}(a, b) = \text{pgcd}(da', db') = d\,\text{pgcd}(a', b')$. Donc, comme d est non nul, on a $\text{pgcd}(a', b') = 1$. □

PROPOSITION 60
Soit $r \in \mathbb{Q}$ un nombre rationnel. Alors il existe un unique couple d'entiers $(p, q) \in \mathbb{Z} \times \mathbb{N}^*$ tel que $r = \dfrac{p}{q}$, avec p et q premiers entre eux.
L'écriture d'un rationnel sous cette forme est appelée **forme irréductible**.

Preuve —
• Existence : Comme $r \in \mathbb{Q}$, il existe $(a, b) \in \mathbb{Z} \times \mathbb{N}^*$ tel que $r = \dfrac{a}{b}$. Posons $d = \text{pgcd}(a, b)$. D'après la proposition précédente, il existe $p \in \mathbb{Z}$ et $q \in \mathbb{N}^*$ tels que $a = pd$, $b = qd$ et $\text{pgcd}(p, q) = 1$. On a donc $r = \dfrac{a}{b} = \dfrac{pd}{qd} = \dfrac{p}{q}$, avec p et q premiers entre eux.

• Unicité : Soient $(p_1, q_1) \in \mathbb{Z} \times \mathbb{N}^*$ et $(p_2, q_2) \in \mathbb{Z} \times \mathbb{N}^*$ tels que $r = \dfrac{p_1}{q_1} = \dfrac{p_2}{q_2}$ et $\text{pgcd}(p_1, q_1) = \text{pgcd}(p_2, q_2) = 1$.

On a alors $p_1q_2 = p_2q_1$, donc $q_2 \mid p_2q_1$. Comme q_2 et p_2 sont premiers entre eux, le théorème de Gauss nous di que $q_2 \mid q_1$. Par symétrie des rôles de q_1 et q_2, on en déduit que $q_1 \mid q_2$. Ainsi on a $|q_1| = |q_2|$, et par positivité de q_1 et q_2 on obtient $q_1 = q_2$. Comme $p_1q_2 = p_2q_1$, on obtient également $p_1 = p_2$, ce qui conclut la preuve. □

6.4 NOMBRES PREMIERS

6.4.1 L'ensemble des nombres premiers

DÉFINITION 61
Soit $p \in \mathbb{N}$. On dit que p est un nombre **premier** si p est différent de 1 et si ses seuls diviseurs positifs sont 1 et p.
On note $\mathcal{P}$ l'ensemble des nombres premiers.

EXEMPLE 62 — *2, 3, 5, 7, 11, 13, 17, 23, ..., sont les plus petits nombres premiers.*

THÉORÈME 63
Tout nombre entier $n \geq 2$ a au moins un diviseur premier.

Preuve — Soit $n \geq 2$. L'ensemble des diviseurs de n qui sont positifs et différents de 1 est une partie non vide de $\mathbb{N} \setminus \{0,1\}$. Il admet donc un minimum p.

Si p n'était pas premier, alors il admettrait lui-même un diviseur q tel que $2 \leq q < p$. Comme $q \mid p$ et $p \mid n$, on aurait $q \mid n$ q. Cela est une contradiction avec la minimalité de p.

Ainsi, p est un nombre premier, et p divise n. □

PROPOSITION 64
Tout nombre entier $n \geq 2$ qui n'est pas premier a au moins un diviseur premier p tel que $2 \leq p \leq \sqrt{n}$.

Preuve — Soit n un entier supérieur ou égal à 2 et non premier. En reprenant la démonstration précédente, le minimum p de l'ensemble des diviseurs supérieurs ou égaux à 2 de n est un nombre premier. Comme p divise n, il existe $q \in \mathbb{N}$ tel que $n = pq$. q est alors un diviseur de n, donc $p \leq q$ par minimalité de p. On a donc $p^2 \leq pq = n$, d'où $p \leq \sqrt{n}$. □

REMARQUE 65 — *Le résultat précédent fournit une méthode pour déterminer si un nombre n est premier ou non :*
on effectue successivement la division euclidienne de n par tous les entiers inférieurs à $\sqrt{n}$, et si l'une des divisions donne un reste nul alors n n'est pas premier. Sinon, n est premier.

On peut améliorer cette méthode et dresser la liste des nombres premiers $p \leq n$ de manière algorithmique en utilisant le crible d'Eratosthène.
Pour cela, on écrit tous les nombres compris entre 2 à n, puis on procède comme suit :

1. *Le plus petit nombre est 2 qui est premier, et tous les multiples stricts de 2 ne sont pas premiers, on élimine alors tous ces multiples,*
2. *Le premier nombre restant est 3, qui est donc premier, et tous les multiples stricts de 3 ne sont pas premiers, on élimine alors tous ces multiples,*
3. *Le premier nombre restant est 5, qui est donc premier, et tous les multiples stricts de 5 ne sont pas premiers, on élimine alors tous ces multiples,*
4. *On poursuit ainsi jusqu'à tomber sur un nombre supérieur à $\sqrt{n}$.*

Les entiers non élimés sont alors exactement les nombres premiers inférieurs à n, puisque les entiers non premiers inférieurs à n possèdent un diviseur premier inférieur à $\sqrt{n}$ et ont donc été éliminés.

PROPOSITION 66
Soient p un nombre premier et $a \in \mathbb{Z}$.
Alors soit p divise a, soit p et a sont premiers entre eux.

Preuve — Comme le pgcd de p et a divise p et que p est premier, on a $\text{pgcd}(p,a) = 1$ ou $\text{pgcd}(p,a) = p$. Supposons que p ne divise pas a. On a alors $\text{pgcd}(p,a) \neq p$ car $\text{pgcd}(a,p) \mid a$, donc $\text{pgcd}(p,a) = 1$ et a et p sont premiers entre eux. □

Ce résultat, comme un certain nombre en arithmétique, n'est vrai que si p est un nombre premier. Par exemple, 6 ne divise pas 15 et 6 et 15 ne sont pas premiers entre eux.

6.4.2 Théorème d'Euclide et petit théorème de Fermat

PROPOSITION 67
Soient $p, q \in \mathbb{N}$ deux nombres premiers distincts. Alors p et q sont premiers entre eux.

Preuve — Soient p, q deux nombres premiers. Supposons que p et q ne sont pas premiers entre eux. La proposition précédente nous dit alors que p divise q et que q divise p. On obtient donc que $p = q$. Par contraposition, on obtient le résultat. □

THÉORÈME 68 (Théorème d'Euclide)
Soient p un nombre premier et $a, b \in \mathbb{Z}$.
Si $p \mid ab$, alors $p \mid a$ ou $p \mid b$.

Preuve — Supposons que $p \mid ab$.

• Si p divise a, c'est bon.

• Sinon, p ne divise pas a. Comme p est premier, p et a sont alors premiers entre eux. Le théorème de Gauss nous dit alors que p divise b, ce qui conclut la preuve. □

PROPOSITION 69
Soient p un nombre premier et $a_1, \ldots, a_n \in \mathbb{Z}$.
Si p divise le produit $a_1 \times \ldots \times a_n = \Pi_{i=1}^n a_i$, alors p divise l'un des a_i.

Preuve — Cette généralisation du théorème précédent se démontre par récurrence sur n. □

EXEMPLE 70 — *Soit $(a, b) \in \mathbb{Z}^2$. Si $2 \mid ab$, alors $2 \mid a$ ou $2 \mid b$, car 2 est un nombre premier.*

THÉORÈME 71 (Petit théorème de Fermat)
Soit p est un nombre premier et $a \in \mathbb{Z}$. On a :

$$a^p \equiv a \bmod p.$$

Si p ne divise pas a, alors :

$$a^{p-1} \equiv 1 \bmod p.$$

Preuve — Démontrons dans un premier lieu le résultat pour $a \geq 0$. Nous allons procéder par récurrence sur a.
• Initialisation : Pour $a = 0$ on a $0^p \equiv 0 \bmod p$.
• Hérédité : Supposons que $a^p \equiv a \bmod p$ pour un $a \geq 0$. La formule du binôme nous donne : $(a+1)^p = \sum_{k=0}^p a^k \binom{k}{p}$. On rappelle que $\binom{k}{p}$ est égal à $\binom{k}{p} = \frac{p!}{k!(p-k)!}$, où $n! = 1 \times 2 \times \ldots \times n$, et que ce nombre est un entier.
Soit $1 \leq k \leq p-1$. Comme p est premier, p ne divise donc pas $k!$ ni $(p-k)!$, alors que p divise $p!$. Ainsi, le théorème d'Euclide appliqué à $(k!(p-k)!)\binom{k}{p} = p!$ nous dit que p divise $\binom{k}{p}$. On obtient donc :

$$(a+1)^p = \sum_{k=0}^{p} a^k \binom{k}{p} \equiv 1 + 0 + \ldots + 0 + a^p \bmod p \equiv a + 1 \bmod p,$$

ce qui prouve que le résultat est vrai pour $a+1$.

Le résultat est ainsi vrai pour tout $a \geq 0$.

Soit maintenant $a \neq 0$. Si $p = 2$, on a $(-1)^2 = 1 \equiv -1 \bmod 2$. Si $p \neq 2$ alors p est impair et $(-1)^p = -1 \equiv -1 \bmod p$. Ainsi, on obtient :

$$a^p = (-|a|)^p \equiv (-1)^p |a| \bmod p.$$

Maintenant, lorsque p ne divise pas a, alors a est premier avec p. D'après le théorème de Bézout il existe $u, v \in \mathbb{Z}$ tels que $au + bp = 1$. Cela donne :

$$ua^p \equiv a^{p-1} \bmod p \equiv ua \bmod p \equiv 1 \bmod p.$$

□

REMARQUE 72 — *Le petit théorème de Fermat est très utile pour calculer/simplifier les puissances d'un nombre entier a modulo p. Nous reverrons ce théorème, qui est très important, en étudiant les groupes (voir chapitre Structures algébriques).*

EXEMPLE 73 — *Calculer $2021^{2021} \bmod 13$ (déterminer le reste de 2021^{2021} dans la division euclidienne par 13). Le nombre 13 est premier. D'après le petit théorème de Fermat, on a donc $2021^{12} \equiv 1 \bmod 13$.*

La division euclidienne de 2021 par 12 donne : $2021 = 12 \times 168 + 5$.

D'autre part, on a $2021 \equiv 6 \bmod 13$. On a donc :

$$2021^{2021} \equiv 6^{12 \times 168 + 5} \bmod 13 \equiv (6^{12})^{168} \times 6^5 \bmod 13 \equiv 6^5 \bmod 13.$$

Cela permet de terminer le calcul :

$$6^2 = 36 \equiv -3 \bmod 13, \quad 6^4 = (6^2)^2 \equiv 9 \bmod 13, \quad 6^5 \equiv 9 \times 6 \bmod 13 \equiv 2 \bmod 13.$$

Donc, $2021^{2021} \equiv 2 \bmod 13$.

PROPOSITION 74
L'ensemble $\mathcal{P}$ est infini : il existe une infinité de nombres premiers.

Preuve — Supposons par l'absurde qu'il existe un nombre fini N de nombres premiers, notés $p_1, p_2, \ldots, p_N$.

On pose alors $p = p_1 \times p_2 \times \ldots \times p_N + 1$. Pour tout $1 \leq j \leq N$, $\Pi_{i=1}^N p_i$ est un multiple de p_j. Ainsi, pour tout $1 \leq j \leq N$, p_j ne divise pas p. En effet, sinon p_j diviserait $p - \Pi_{i=1}^N p_i = 1$, ce qui est impossible puisque $p_j > 1$.

On remarque que p est un nombre entier supérieur ou égal à 2. Il admet donc un diviseur premier. Par hypothèse, ce diviseur est forcément de la forme p_{i_0} pour un $i_0 \in \{1, \ldots, N\}$. Cela implique que $p_{i_0} \mid p$, ce qui est absurde.

Ainsi, le nombre de nombres premiers est infini, ce qui conclut la preuve. □

6.4.3 Décomposition en produit de facteurs premiers

THÉORÈME 75 (Théorème fondamental de l'arithmétique)
Soit $n \in \mathbb{N}$ un entier naturel, avec $n \geq 2$.
Alors n se décompose, de manière unique à l'ordre près des termes, en produit de facteurs premiers :

$$n = p_1^{\alpha_1} \times p_2^{\alpha_2} \times \ldots \times p_N^{\alpha_N},$$

où les p_i sont des nombres premiers deux à deux distincts et les α_i sont des entiers naturels non nuls.

Preuve —

- *Existence :* Démontrons ce résultat par récurrence sur n. Pour tout $n \geq 2$, on note (H_n) la propriété : « n se décompose en produit de facteurs premiers. »
  • *Initialisation :* $n = 2$ est un nombre premier, donc n s'écrit comme le produit de nombres premiers. Donc (H_2) est vraie.
  • *Hérédité :* Soit $n \geq 3$. Supposons (H_k) vraie pour tout $2 \leq k \leq n-1$.
  Si n est premier, alors n se décompose en produit de facteurs premiers.
  Sinon, il existe des entiers naturels a, b, avec $a, b > 1$, tels que $n = ab$. On a ainsi $a < n$ et $b < n$, donc l'hypothèse de récurrence s'applique à a et à b, qui se décomposent en un produit de facteurs premiers. Comme n est le produit de a et b, il est un produit de facteurs premiers, donc (H_n) est vraie.
- *Unicité :* Supposons que n se décompose en deux produits :

  $$n = p_1^{\alpha_1} \times \ldots \times p_N^{\alpha_N} \quad \text{et} \quad n = q_1^{\beta_1} \times \ldots \times q_R^{\beta_R},$$

  où les p_i et q_j sont des nombres premiers, avec les p_i distincts deux à deux, les q_j également, et où les α_i, β_j sont des entiers naturels non nuls.
  Pour tout $i \in \{1, \ldots, N\}$ on a $p_i \mid n$, donc p_i divise l'un des q_j. Comme p_i et q_j sont des nombres premiers, on a $p_i = q_j$. Donc, $\{p_1, \ldots, p_N\} \subset \{q_1, \ldots, q_R\}$.
  Par symétrie des rôles, on en déduit que $\{q_1, \ldots, q_R\} \subset \{p_1, \ldots, p_N\}$. Ces deux ensembles sont donc égaux et on a $N = R$. Quitte à permuter les indices, on peut supposer que $p_i = q_i$ pour tout $i \in \{1, \ldots, N\}$.
  Soit $\leq i \leq N$. On a $p_i^{\alpha_i} \mid n$, avec $n = q_1^{\beta_1} \times \ldots q_N^{\beta_N} = q_i^{\beta_i} \times k$. Comme $q_i = p_i$, on a $\text{pgcd}(p_i^{\alpha_i}, k) = 1$. Ainsi, le théorème de Gauss nous dit que $p_i^{\alpha_i} \mid p_i^{\beta_i}$. Donc $\alpha_i \leq \beta_i$. Par symétrie des rôles, on a de même $\beta_i \leq \alpha_i$, donc finalement $\alpha_i = \beta_i$.
  Cette décomposition en produit de facteurs premiers est donc unique à l'ordre près.

□

DÉFINITION 76
Soient $n \geq 2$ un entier naturel et p un nombre premier.
On définit $\nu_p(n)$ l'exposant de p dans la décomposition de n en facteurs premiers (avec $\nu_p(n) = 0$ si p ne divise pas n). L'entier $\nu_p(n)$ est appelé la **valuation p-adique** de n.

REMARQUE 77 — *On a* $\nu_p(n) = \max\{k \in \mathbb{N} \mid p^k \text{ divise } n\}$.
On peut aussi écrire l'entier n comme :

$$n = \Pi_{p \in \mathcal{P},\, p \leq n} p^{\nu_p(n)}.$$

MÉTHODE 78 — *Pour décomposer un nombre entier $n \geq 2$, on peut procéder de la façon suivante :*

1. *On cherche la plus grande puissance $\alpha_1 \geq 0$ de 2 divisant n, on obtient $n = 2^{\alpha_1} n_1$ où $n_1 \in \mathbb{N}$ et n_1 n'est plus divisible par 2. Si $n_1 = 1$, on a terminé, sinon on passe à l'étape suivante.*
2. *On cherche la plus grande puissance $\alpha_2 \geq 0$ de 3 divisant n_1, on obtient alors $n = 2^{\alpha_1} \times 3^{\alpha_2} n_2$ où $n_2 \in \mathbb{N}$ et n_2 n'est plus divisible par 3 (ni 2 par la première étape). Si $n_2 = 1$, on a terminé, sinon on passe à l'étape suivante.*
3. *On cherche la plus grande puissance $\alpha_3 \geq 0$ de 5 divisant n_2, etc.*

EXEMPLE 79 — $360 = 2^3 \times 3^2 \times 5, \quad 147 = 3 \times 7^2, \quad 1575 = 3^2 \times 5^2 \times 7.$

PROPOSITION 80
Soit un entier $n \geq 2$. Soit $n = p_1^{\alpha_1} \times \ldots \times p_N^{\alpha_N}$ la décomposition de n en facteurs premiers, avec p_i des nombres premiers distincts deux à deux et α_i des entiers naturels non nuls.

Alors les diviseurs positifs de n sont exactement les entiers de la forme $p_1^{\beta_1} \ldots p_N^{\beta_N}$ avec $0 \leq \beta_i \leq \alpha_i$ pour tout $1 \leq i \leq N$.

L'entier n possède ainsi $\Pi_{i=1}^{N}(\alpha_i + 1)$ diviseurs.

Preuve — Soit m un diviseur de n. Alors tous les facteurs premiers de m sont des diviseurs de n. Ainsi, m est de la forme $m = p_1^{\beta_1} \ldots p_N^{\beta_N}$.

Comme $m \mid n$ on a $p_i^{\beta_i} \mid n$ pour tout $1 \leq i \leq N$, donc $\beta_i \leq \alpha_i$ d'après la preuve du théorème.

Réciproquement, tout entier m de la forme $m = p_1^{\beta_1} \dots p_N^{\beta_N}$ avec $\beta_i \leq \alpha_i \ \forall 1 \leq i \leq N$ est un diviseur de n.

Par unicité de la décomposition en facteurs premiers (à l'ordre près des termes), le nombre de diviseurs de n est égal au nombre de choix possibles des N entiers $(\beta_1, \dots, \beta_N)$. Comme on a $\beta_i \in \{0, \dots, \alpha_i\}$, on a $\alpha_i + 1$ choix pour β_i, ce qui donne $\Pi_{i=1}^{N}(\alpha_i + 1)$ diviseurs de n. □

EXEMPLE 81 — *Les diviseurs positifs de* $45 = 3^2 \times 5$ *sont les suivants :* $3^0 \times 5^0 = 1$, $3^0 \times 5 = 5$, $3 \times 5^0 = 3$, $3 \times 5 = 15$, $3^2 \times 5^0 = 9$, $3^2 \times 5 = 45$.

On dispose du résultat suivant pour calculer le pgcd et le ppcm de deux entiers à partir de leur décomposition en produit de nombres premiers.

PROPOSITION 82
Soient $a, b \in \mathbb{N}$ supérieurs ou égaux à 2. On suppose que $a = p_1^{\alpha_1} \times \dots \times p_N^{\alpha_N}$ et $b = p_1^{\beta_1} \times \dots \times p_N^{\beta_N}$, où les p_i sont des nombres premiers distincts deux à deux et les α_i, β_i sont des entiers naturels (éventuellement nuls). Alors on a :

- $\text{pgcd}(a, b) = p_1^{\min(\alpha_1, \beta_1)} \times \dots \times p_N^{\min(\alpha_N, \beta_N)}$,
- $\text{ppcm}(a, b) = p_1^{\max(\alpha_1, \beta_1)} \times \dots \times p_N^{\max(\alpha_N, \beta_N)}$.

Preuve — On montre que $p_1^{\min(\alpha_1, \beta_1)} \times \dots \times p_N^{\min(\alpha_N, \beta_N)}$ est un diviseur commun de a et b, et le plus grand de leurs diviseurs communs.

On montre que $p_1^{\max(\alpha_1, \beta_1)} \times \dots \times p_N^{\max(\alpha_N, \beta_N)}$ est un multiple commun de a et b, et le plus petit de leurs multiples communs.

Les Propositions 25 et 39 nous disent alors que ces quantités sont $\text{pgcd}(a, b)$ et $\text{ppcm}(a, b)$. □

EXEMPLES 83

- $\text{pgcd}(147, 1575) = 3 \times 7 = 21$,
- $\text{ppcm}(147, 1575) = 3^2 \times 5^2 \times 7^2 = 11025$.

On établit alors la relation suivante qui lie pgcd et ppcm.

PROPOSITION 84
Soient $a, b \in \mathbb{Z}$. Alors on a

$$\text{pgcd}(a, b) \times \text{ppcm}(a, b) = |a| \times |b|.$$

En particulier, si a et b sont premiers entre eux, on a $\text{ppcm}(a, b) = |a| \times |b|$.

Preuve — On peut supposer a et b positifs. Si $a = 0$ ou $b = 0$ on a $\text{ppcm}(a, b) = 0$, et le résultat est vrai. Si $a = 1$ ou $b = 1$ on a $\text{pgcd}(a, b) = 1$, et le résultat est vrai.

Supposons que $a \geq 2$ et $b \geq 2$. Soient $p_1, \dots, p_N$ les nombres premiers divisant a ou b. D'après le théorème fondamental de l'arithmétique, on a alors $a = p_1^{\alpha_1} \times \dots \times p_N^{\alpha_N}$ et $b = p_1^{\beta_1} \times \dots \times p_N^{\beta_N}$, où les α_i, β_i sont des entiers naturels (éventuellement nuls).

La proposition précédents nous fournit alors les valeurs de $\text{pgcd}(a, b)$ et $\text{ppcm}(a, b)$ en fonction des p_i, α_i et β_i. On a alors :

$$\begin{aligned} \text{pgcd}(a, b) \times \text{ppcm}(a, b) &= p_1^{\min(\alpha_1, \beta_1) + \max(\alpha_1, \beta_1)} \times \dots \times p_N^{\min(\alpha_N, \beta_N) + \max(\alpha_N, \beta_N)} \\ &= p_1^{\alpha_1 + \beta_1} \times \dots \times p_N^{\alpha_N + \beta_N} = ab \end{aligned},$$

ce qui conclut. □

EXEMPLE 85 — *Les multiples communs à* 12 *et* 18 *sont les multiples de* 36.
En effet, $\text{ppcm}(12, 18) = \dfrac{12 \times 18}{\text{pgcd}(12, 18)} = \dfrac{12 \times 18}{6} = 36$.

PROPOSITION 86
Soient $a, b \in \mathbb{Z}$. Alors a et b sont premiers entre eux si et seulement s'ils n'ont pas de facteurs premiers en commun dans leur décomposition en produit de facteurs premiers.

Preuve — Si a et b sont premiers entre eux, alors leur seul diviseur commun positif est 1 et ils n'ont donc pas de facteur premier en commun.

Réciproquement, supposons que a et b n'ont pas de facteurs premiers en commun. Notons $d = \text{pgcd}(a, b)$. Si l'on avait $d \geq 2$, alors d admettrait un diviseur premier p. Comme d divise a et b, p diviserait également a et b et il serait donc un facteur premier commun à a et à b, ce qui est impossible. On a donc $d = 1$, donc a et b sont premiers entre eux. □

EXEMPLE 87 — $825 = 3 \times 5^2 \times 11$ *et* $56 = 2^3 \times 7$ *sont premiers entre eux.*

Chapitre 7 Polynômes à une indéterminée

Table des matières du chapitre

Un polynôme s'écrit de la forme

$$P(X) = a_n X^n + a_{n-1} X^{n-1} + \ldots + a_1 X + a_0$$

où les a_i s'appellent les coefficients de P et X où est l'indéterminée.
On retrouvera les polynômes tant en analyse (par ex. les développements limités) qu'en algèbre (par ex. polynôme caractéristique d'une application linéaire).
Une bonne maîtrise des produits, divisions et factorisations de polynômes ainsi que de la caractérisation des racines est indispensable.

7.1 Polynômes, opérations, degré, fonctions polynômiales

Polynômes à une indéterminée

Définition 1
Soit $\mathbb{K}$ un corps. On appelle **polynôme à une indéterminée** à coefficients dans $\mathbb{K}$ toute suite d'éléments $(a_n)_{n\in\mathbb{N}}$ d'éléments de $\mathbb{K}$ qui est nulle à partir d'un certain rang :

$$\exists d \in \mathbb{N}, \text{ tel que } (a_n)_{n\in\mathbb{N}} = (a_0, a_1, \ldots, a_d, 0, \ldots, 0, \ldots).$$

L'ensemble des polynômes est noté $\mathbb{K}[X]$.

Définition 2
Soit $\mathbb{K}$ un corps. On définit sur $\mathbb{K}[X]$ deux lois internes $(+, \times)$ et une loi externe $(.)$:

1. L'addition, $+$, est définie par :

$$(a_0, \ldots, a_d, 0, \ldots) + (b_0, \ldots, b_{d'}, 0, \ldots) = (a_0 + b_0, a_1 + b_1, \ldots, a_k + b_k, \ldots) = (a_n + b_n)_{n\in\mathbb{N}}.$$

   Une telle suite est bien dans $\mathbb{K}[X]$ car $c_k = 0$ pour $k > sup(d, d')$.
   La suite nulle, notée $0_{\mathbb{K}[X]} = (0, \ldots, 0, \ldots)$ ou 0, est l'élément neutre pour l'addition $+$.

2. La multiplication, $\times$, est définie par :

$$(a_0, \ldots, a_d, 0, \ldots) \times (b_0, \ldots, b_{d'}, 0, \ldots) = (c_n)_{n\in\mathbb{N}}, \text{ avec } c_n = \sum_{k=0}^{n} a_k b_{n-k}.$$

   Une telle suite est bien dans $\mathbb{K}[X]$ car $c_k = 0$ pour $k > m + n$.
   On note $1_{\mathbb{K}[X]} = (1, 0, \ldots, 0, \ldots)$, ou 1, l'élément neutre pour la multiplication $\times$.

3. La multiplication par un scalaire de $\mathbb{K}$, ., définie par :

$$\begin{array}{rcl} \mathbb{K} \times \mathbb{K}[X] & \to & \mathbb{K}[X] \\ (\lambda, (a_0, a_1, \ldots, a_d, 0 \ldots)) & \mapsto & \lambda.(a_n)_{n\in\mathbb{N}} = (\lambda.a_0, \lambda a_1, \ldots, \lambda.a_d, 0, \ldots). \end{array}$$

Proposition 3
Soit $\mathbb{K}$ un corps. Soient $P, Q, R \in \mathbb{K}[X]$. Soit $\lambda \in \mathbb{K}$. On a :

1. $P + (Q + R) = (P + Q) + R$ ($+$ est associative) ;
2. $P + Q = Q + P$ ($+$ est commutative) ;

3. $P + 0 = 0 + P = P$ (0 est le neutre de $+$) ;
4. $\lambda.(P + Q) = \lambda.P + \lambda.Q$ ($.$ est distributive sur $+$) ;
5. $P \times (Q \times R) = (P \times Q) \times R$ ($\times$ est associatve) ;
6. $(P \times Q) = (Q \times P)$ ($\times$ est commutative) ;
7. $(P \times 1) = (1 \times P) = P$ (1 est le neutre de $\times$) ;
8. $P \times (Q + R) = P \times Q + P \times R = (Q + R) \times P$ ($\times$ est distributive sur $+$) ;
9. $P \times (\lambda.Q) = \lambda.P \times Q$ ($\times$ et $.$ commutent).

L'ensemble $(\mathbb{K}[X], +, .)$ est donc un $\mathbb{K}$-espace vectoriel.
L'ensemble $(\mathbb{K}[X], +, \times)$ est donc un anneau commutatif.
L'ensemble $(\mathbb{K}[X], +, \times, .)$ est donc une $\mathbb{K}$-algèbre.

Proposition 4
Soit $\mathbb{K}[X]$ un corps. Soient $P, Q \in \mathbb{K}[X]$.
On a $P \times Q = 0$ si et seulement si $P = 0$ ou $Q = 0$.
L'anneau $\mathbb{K}[X]$ est intègre.

Écriture d'un polynôme

Définition 5
Soit $\mathbb{K}$ un corps. On définit **l'indéterminée** de $\mathbb{K}[X]$ comme la suite $X = (0, 1, 0, \ldots, 0, \ldots)$.

Proposition 6
Dans $\mathbb{K}[X]$, on pose $X^0 = 1$. Pour tout $k \in \mathbb{N}$, on a alors :

$$X^k = \underbrace{X \times \ldots \times X}_{k \text{ fois}} = (\underbrace{0, \ldots, 0}_{k}, 1, 0, \ldots, 0, \ldots).$$

Tout polynôme $P \in \mathbb{K}[X]$ non nul s'écrit de manière unique de la forme :

$$P = a_n X^n + a_{n-1} X^{n-1} + \ldots + a_0,$$

avec $a_0, \ldots, a_n \in \mathbb{K}$ et $a_n \neq 0$.

Remarque 7 — *L'indéterminée X est un élément très important pour travailler dans $\mathbb{K}[X]$.*
On écrit souvent $P(X)$ à la place de P. Cette écriture est parfois très utile (par exemple pour différencier un polynôme $P(X)$ de sa fonction polynômiale associée $x \mapsto P(x)$).
Un polynôme quelconque de $\mathbb{K}[X]$ est ainsi de la forme $P(X) = \sum_{k=0}^{n} a_k X^k$, pour un $n \geq 0$ et des $a_0, \ldots, a_n \in \mathbb{K}$.

Corollaire 8
Soit $P(X) = \sum_{k=0}^{n} a_i X^i \in \mathbb{K}[X]$ un polynôme. Alors $P(X)$ est le polynôme nul si et seulement si l'on a $a_k = 0$ pour tout $k \in \{0, \ldots, n\}$.

Remarque 9 — *La somme de polynômes sous la nouvelle écriture ne pose pas de problèmes. Pour le produit, on a :*

$$\left(\sum_{k=0}^{n} a_k X^k\right) \times \left(\sum_{k=0}^{m} b_k X^k\right) = \sum_{k=0}^{m+n} c_k X^k, \text{ avec } c_k = \sum_{i=0}^{k} a_i b_{k-i}.$$

Exemple 10 — *Pour multiplier rapidement deux polynômes, on utilise la distributivité du produit sur la somme et on regroupe les termes de même degré :*

$$(X+1)(X^3+X+2) = X^4(1.1) + X^3(1.1) + X^2(1.1) + X(1.1+1.2) + (1.2) = X^4 + X^3 + X^2 + 3X + 2,$$
$$X^2+X+1)(X^2-4X+3) = X^4(1.1) + X^3(1.(-4)+1.1) + X^2(1.1+1.(-4)+1.3) + X(1.3+1.(-4)) + (1.3) \quad = X^4 -$$

Degré d'un polynôme

DÉFINITION 11
Soit $P(X) = \sum_{k=0}^{n} a_k X^k$ un polynôme non nul. On appelle degré de P, noté $\deg(P)$, le plus grand entier k tel que $a_k \neq 0$.
Pour $d = \deg P$, le terme $a_d X^d$ est appelé terme dominant du polynôme P, a_d le coefficient dominant de P.
a_0 est appelé le coefficient constant de P.
On dit que P est un polynôme **unitaire** si son coefficient dominant vaut 1.
Enfin, le degré du polynôme nul est par convention $\deg(0) = -\infty$.

EXEMPLES 12 *Le polynôme $2X^2 + X + 1$ n'est pas unitaire, mais $X^7 + X^3 + 2$ l'est. On a $\deg(X^7 + X^3 + 2) = 7$.*
Pour $\lambda \in \mathbb{K}^*$ on a $\deg(\lambda) = 0$, tandis que $\deg(0) = -\infty$.
Pour tout $n \geq 0$, on a $\deg(X^n) = n$.

PROPOSITION 13
Soient P,$Q \in \mathbb{K}[X]$. Il résulte des définitions de $+$ et $\times$ que

1. $\deg(P + Q) \leq \max(\deg(P), \deg(Q))$;
   Si $\deg P \neq \deg Q$, alors $\deg P + Q = \max(\deg P, \deg Q)$.
2. $\deg(P \times Q) = \deg P + \deg Q$;
3. $\forall \lambda \in \mathbb{K}^*$, $\deg \lambda.P = \deg P$.

EXEMPLES 14

1. $\deg((X^3 + X + 3) + (X^2 + 2)) = 3$;
2. $\deg((X^3 + X + 3) + (-X^3 + 3X + 7)) = 1$;
3. $\deg((X^3 + X + 2)(X^5 + 3X^4 + 2)) = 8$.

REMARQUE 15 — *L'écriture $P(X) = \sum_{k=0}^{n} a_k X^k$ avec $a_0, \ldots, a_n \in \mathbb{K}$ nous dit seulement que $\deg(P) \leq n$.*

Il faut rajouter la condition $a_n \neq 0$ pour avoir $\deg(P) = n$.

DÉFINITION 16
Soient $\mathbb{K}$ un corps et $n \in \mathbb{N}$. On définit $\mathbb{K}_n[X]$ l'ensemble des polynômes sur $\mathbb{K}$ de degré inférieur ou égal à n :

$$\mathbb{K}_n[X] = \{P \in \mathbb{K}[X],\ \deg(P) \leq n\}.$$

L'ensemble $\mathbb{K}_n[X]$ est un sous-espace vectoriel de $(\mathbb{K}[X], +, .)$.

Fonctions polynomiales

DÉFINITION 17
Soit $P(X) \in \mathbb{K}[X]$, avec $P(X) = \sum_{k=0}^{n} a_k X^k$. On appelle fonction polynomiale associée au polynôme $P(X)$, la fonction notée P ou f_P ou $(x \mapsto P(x))$, définie par :

$$P : \begin{array}{ccc} \mathbb{K} & \to & \mathbb{K} \\ x & \mapsto & P(x) := \sum_{k=0}^{n} a_k x^k \end{array} .$$

REMARQUE 18 — *La fonction $\psi : P(X) \in \mathbb{K}[X] \mapsto (x \mapsto P(x)) \mathcal{F}(\mathbb{K}, \mathbb{K})$ vérifie les propriétés :*

- $\psi(P + Q) = \psi(P) + \psi(Q)$;
- $\psi(\lambda.P) = \lambda\psi(P)$;
- $\psi(P \times Q) = \psi(P) \times \psi(Q)$.

En particulier, ψ est une application linéaire de $\mathbb{K}[X]$ vers $\mathcal{F}(\mathbb{K}, \mathbb{K})$, ainsi qu'un morphisme d'anneaux, et un morphisme de $\mathbb{K}$-algèbres.

7.2 L'espace vectoriel $\mathbb{K}[X]$

Familles échelonnées en degré

Proposition 19
Soient $\mathbb{K}$ un corps et $n \in \mathbb{N}$. L'espace vectoriel $\big(\mathbb{K}[X], +, .\big)$ est un $\mathbb{K}$-espace vectoriel de dimension infinie.
La famille $\{1, X, \dots, X^n, \dots\} = \{X^k,\ k \geq 0\}$ est une base de cet espace vectoriel, appelée base canonique de $\mathbb{K}[X]$.
Le sous-espace vectoriel $\mathbb{K}_n[X]$ des polynômes de degré au plus n est un sous-espace vectoriel de $\mathbb{K}[X]$, de dimension $n+1$.
La famille $\{1, X, \dots, X^n\}$ est une base de cet espace, appelée base canonique de $\mathbb{K}_n[X]$.

Définition 20
Soit $\{P_0, \dots, P_n\}$ une famille de polynômes de $\mathbb{K}[X]$. On dit que cette famille est échelonnée en degré si $\deg P_i = i$ pour tout $i \in [\![0, n]\!]$.

Proposition 21
Soit $\{P_0, \dots, P_n\}$ une famille de polynômes de $\mathbb{K}[X]$ échelonnée en degré. Alors cette famille est une base de $\mathbb{K}_n[X]$.

Remarque 22 — *Si vous pouvez montrer qu'une famille de $n+1$ polynômes est échelonnée en degré, vous aurez montré que c'est une base de $\mathbb{K}_n[X]$. La famille $(1, 1+X, 1+X+X^2, \dots, 1+X+\dots+X^n)$ est une base de $\mathbb{K}_n[X]$ car elle est échelonnée en degré.*
Plus généralement, une famille $\{P_1, \dots, P_n\}$ de polynômes qui sont de degrés tous distincts est libre. En réordonnant ces polynômes selon leur degré, on peut voir cet ensemble comme une sous-famille d'une famille échelonnée en degré.

Nous allons voir une famille de polynômes assez classique et très utile qui elle n'est pas écéhelonnée en degré, mais qui forme une base de $\mathbb{K}_n[X]$.

Polynômes interpolateurs de Lagrange

Définition 23
Soit $n \geq 1$. Soient $a_0, \dots, a_n \in \mathbb{K}$ des éléments deux à deux distincts.
On définit la famille de polynômes $\{L_0, \dots, L_n\}$, appelés **polynômes interpolateurs de Lagrange**, par :

$$L_i(X) == \frac{\prod\limits_{k=0, k\neq i}^{n} (X - a_k)}{\prod\limits_{k=0, k\neq i}^{n} (a_i - a_k)} = \frac{(X-a_0)\dots(X-a_{i-1})(X-a_{i+1})\dots(X-a_n)}{(a_i-a_0)\dots(a_i-a_{i-1})(a_i-a_{i+1})\dots(a_i-a_n)},\ \forall i \in [\![0, n]\!].$$

Proposition 24
Soit $n \geq 1$. Soient $a_0, \dots, a_n \in \mathbb{K}$ deux à deux distincts. On a alors :

1. $L_i(a_j) = \delta_{i,j}$, $\forall i \in [\![0, n]\!]$. ($\delta_{i,j} = 1$ si $i = j$ et $\delta_{i,j} = 0$ si $i \neq j$)
2. La famille $\{L_0, \dots, L_n\}$ est une famille de polynômes de degré n qui forme une base de $\mathbb{K}_n[X]$.
3. Soit $P \in \mathbb{K}_n[X]$. Alors $P(X) = \sum\limits_{i=0}^{n} P(a_i) L_i(X)$.
4. Soient b_0, ..., $b_n \in \mathbb{K}$.
   Il existe un unique polynôme $Q \in \mathbb{K}_n[X]$ tel que $Q(a_i) = b_i$ pour tout $0 \leq i \leq n$. On a :

$$Q(X) = \sum_{i=0}^{n} b_i L_i(X).$$

Exemple 25 (Matrice de Vandermonde) — *Soient a_0, ..., $a_n \in \mathbb{K}$ des éléments deux à deux distincts. On cherche à montrer que la matrice suivante, appelée matrice de Vandermonde, est inversible et à calculer son inverse.*

$$V(a_0, \dots, a_n) = \begin{pmatrix} 1 & a_0 & \dots & a_0^n \\ \vdots & \vdots & \dots & \vdots \\ 1 & a_{n-1} & \dots & a_{n-1}^n \\ 1 & a_n & \dots & a_n^n \end{pmatrix}.$$

Soient $y_0, \dots, y_n \in \mathbb{K}$. On cherche à résoudre :

$$V(a_0,\dots,a_n)\begin{pmatrix} x_0 \\ \vdots \\ x_n \end{pmatrix} = \begin{pmatrix} x_0.1 + x_1a_0 + \dots + x_na_0^n \\ \vdots \\ x_0.1 + x_1a_{n-1} + \dots + x_na_{n-1}^n \\ x_0.1 + x_1a_n + \dots + x_na_n^n \end{pmatrix} = \begin{pmatrix} y_0 \\ \vdots \\ y_n \end{pmatrix}.$$

En posant $P(X) = x_0 + x_1X + \dots + x_nX^n$, le polynôme P serait alors un polynôme de degré au plus n tel que $P(a_i) = y_i$ pour tout $i \in [\![0,n]\!]$. D'après ce qui précède, un tel polynôme existe et vaut :

$$P(X) = \sum_{i=0}^{n} y_iL_i(X) = x_0 + x_1X + \dots + x_nX^n.$$

Ainsi, le système linéaire initial possède toujours une solution. Cela implique que la matrice $V(a_0,\dots,a_n)$ est inversible. Pour trouver la valeur de $(x_0,\dots,x_n)$ en fonction de $(y_0,\dots,y_n)$, il suffit de développer chaque polynôme interpolateur de Lagrange. Pour $L_j = \sum_{i=0}^{n} a_{i,j}X^i$, on a alors $x_i = \sum_{j=0}^{n} a_{i,j}y_j$. Ainsi, l'inverse de la matrice de Vandermonde est $V(a_0,\dots,a_n)^{-1} = (a_{i,j})_{0\le i,j\le n}$.

REMARQUE 26 — *Nous venons de montrer qu'en prenant $n+1$ points deux à deux distincts, un polynôme de degré au plus n, était uniquement déterminé par sa valeur en ces $n+1$ points.*
Si $\mathbb{K}$ a une infinité d'éléments, alors il existe des familles de polynômes interpolateurs de Lagrange pour toute famille de points aussi grande que l'on veut.

Il existe cependant des corps $\mathbb{K}$ avec un nombre fini d'éléments. Sur un tel corps, on ne peut interpoler que pour $n+1 \le$ Card($\mathbb{K}$) points distincts.

EXEMPLE 27 — *Trouver le polynôme $P \in \mathbb{R}[X]$ de degré au plus 2 tel que $P(0) = 1$, $P(1) = 2$ et $P(3) = 3$.*

7.3 L'anneau $\mathbb{K}[X]$, division euclidienne de polynômes

THÉORÈME 28 (Division euclidienne de polynômes)
Soient A et $B \in \mathbb{K}[X]$ avec $B \neq 0$. Il existe un unique couple $(Q,R) \in \mathbb{K}[X]^2$ tel que :

$$A = QB + R \qquad \text{avec} \qquad \deg R < \deg B.$$

Preuve — L'unicité se montre comme pour la division euclidienne d'entiers : on suppose qu'il existe deux couples possibles, et on montre qu'ils sont égaux.
Existence : Le cas $B = \lambda \in \mathbb{K}^*$ ($\deg B = 0$) est immédiat avec $(Q,R) = (\lambda^{-1}A, 0)$. Supposons B non constant.
On procède par récurrence sur $\deg(A)$. On remarque d'une part que si $\deg(A) < \deg(B)$, alors $(Q,R) = (0,A)$.
D'autre part, si $\deg A \ge \deg B$, en écrivant :

$$A = a_nX^n + \dots + a_0,\ B = b_mX^m + \dots + b_0,\ \text{avec } a_nb_m \neq 0,$$

on remarque que le polynôme $A - \dfrac{a_n}{b_m}X^{n-m}B$ est de degré strictement inférieur àdeg(A), ce qui permet d'appliquer l'hypothèse de récurrence à ce dernier. □

REMARQUE 29 — *Soient $A, B \in \mathbb{K}[X]$ avec B non-nul. On a $B|A$ si et seulement si le reste de la division euclidienne de A par B est nul.*

EXEMPLE 30 (Algorithme de la division euclidienne) —
On effectue une division euclidienne de polynômes en faisant descendre le degré du polynôme à diviser. Voici en exemple la division euclidienne de $A = X^5 + 4X^4 + 2X^3 + X^2 - X - 1$ par $B = X^3 - 2X + 3$:

$$\begin{array}{ccccccc|c} X^5 & +4X^4 & +2X^3 & +X^2 & -X & -1 & X^3-2X+3 \\ \cline{7-7} & 4X^4 & +4X^3 & -2X^2 & -X & -1 & X^2+4X+4 \\ & & 4X^3 & +6X^2 & -13X & -1 & \\ & & & 6X^2 & -5X & -13 & \end{array}$$

On trouve finalement $X^5 + 4X^4 + 2X^3 + X^2 - X + 1 = (X^3 - 2X + 3)(X^2 + 4X + 4) + (6X^2 - 5X - 13)$.

REMARQUE 31 — *Soit $\mathbb{K}$ un corps.*
Alors, l'anneau des polynômes $\mathbb{K}[X]$ est un anneau euclidien.
En particulier, c'est un anneau intègre et un anneau principal.
Tous les résultats sur les anneaux principaux (pgcd et ppcm, Bézout,Gauss, dcomposition en facteurs irréductibles) s'appliquent donc à $\mathbb{K}[X]$.

7.4 Racines d'un polynôme, dérivation, factorisation

Définition 32
Soient $P \in \mathbb{K}[X]$ et $\alpha \in \mathbb{K}$. On dit que α est une **racine** du polynôme P si l'on a $P(\alpha) = 0$, où $P(\alpha)$ désigne l'image de α par la fonction polynômiale associée à P.

Proposition 33
Soient $P \in \mathbb{K}[X]$ et $a \in \mathbb{K}$.
Alors, a est une racine de P si et seulement si $(X - a)|P(X)$.

Preuve — On écrit la division euclidienne de $P(X)$ par $(X - a)$: $P(X) = (X - a)Q(X) + R(X)$ avec $\deg(R) < \deg(X - A) = 1$. $R(X)$ est donc un polynôme constant : $R(X) = \lambda$. L'évaluation en a donne $P(a) = 0.Q(a) + R(a) = \lambda$. Ainsi, a est une racine de P si et seulement si $R(X) = 0$, si et seulement si $(X - a)$ divise $P(X)$. □

Définition 34
Soient $P \in \mathbb{K}[X]$, $a \in \mathbb{K}$, et $k \geq 1$. On dit que a est une racine de **multiplicité** k de P si l'on a $(X - \alpha)^k|P$ et $(X - a)^{k+1} \not| P$.
Une racine de multiplicité 1 est appelée racine simple de P.

Proposition 35
Soient $P \in \mathbb{K}[X]$ et $a_1, \ldots, a_r \in \mathbb{K}$, tels que $a_1, \ldots, a_r$ sont des racines de P de multiplicités respectives α_1, ...,α_r.
Alors il existe $Q \in \mathbb{K}[X]$ tel que :

$$P = (X - a_1)^{\alpha_1} \ldots (X - a_r)^{\alpha_r} Q \qquad \text{et} \qquad Q(a_i) \neq 0, \forall 1 \leq i \leq r.$$

Corollaire 36
Soit $\mathbb{K}$ un corps et $P \in \mathbb{K}[X]$ de degré $n \geq 0$.
Alors P possède au plus n racines, comptées avec leur multiplicité.
• Pour tout $n \geq 0$, le seul polynôme Q de $\mathbb{K}_n[X]$ qui possède $n + 1$ racines ou plus est le polynôme nul.

Définition 37
Soit $P \in \mathbb{K}[X]$ non-nul.
On dit que P est **scindé** s'il admet autant de racines (comptées avec multiplicité) que son degré.
Il est équivalent de dire que $P(X) = a_n \prod_{i=1}^{r} (X - z_i)^{\alpha_i}$, pour des $z_1, \ldots, z_r \in \mathbb{K}$.
On dit que P est **scindé à racines simples** si le polynôme P est scindé et si toutes ses racines sont distinctes.
Il est équivalent de dire que $P(X) = a_n \prod_{i=1}^{n} (X - z_i)$, pour des $z_1, \ldots, z_n \in \mathbb{K}$ distincts.

Exemple 38 — *Le polynôme $X^n - 1$ admet donc n racines dans $\mathbb{C}$. On a vu qu'il ne possède aucune racine double, ce qui montre qu'il existe exactement n racines n-ièmes de l'unité dans $\mathbb{C}$.*

Remarque 39 — *Si le corps $\mathbb{K}$ est infini, deux fonctions polynomiales sont égales sur $\mathbb{K}$ si et seulement si leurs polynômes associés ont les mêmes coefficients.*
Ainsi, par exemple, $P : x \in \mathbb{R} \mapsto x^7 + 5x^4 + 1 \in \mathbb{R}$ n'est pas du tout la même fonction que $Q : x \in \mathbb{R} \mapsto x^7 + 5x^2 + 1 \in \mathbb{R}$.
De telles fonctions ne sont égales qu'en au plus 4 points (on a $P(x) = Q(x)$ ssi $(P-Q)(x) = 0$, et $\deg(P-Q) = 4$).

Dérivation dans $\mathbb{K}[X]$

Dérivée d'un polynôme

Définition 40
Soit $P \in \mathbb{K}[X]$ avec $P = a_n X^n + \ldots + a_0$. On appelle **polynôme dérivé** de P, noté P', le polynôme :

$$P'(X) = n a_n X^{n-1} + (n-1) a_{n-1} X^{n-2} + \ldots + a_1.$$

Proposition 41
Soit $\mathbb{K}$ un corps tel que $\mathbb{K}$ contient $\mathbb{Z}$ ($Car(\mathbb{K}) = 0$). Soit $P \in \mathbb{K}[X]$.
On a $\deg(P') = \deg(P) - 1$ si $\deg(P) \geq 1$, et $P'(X) = 0$ sinon.

Proposition 42
Soit $\mathbb{K}$ un corps. La fonction $D : P \in \mathbb{K}[X] \mapsto P' \in \mathbb{K}[X]$ est une application linéaire.
Si $\mathbb{K}$ contient $\mathbb{Z}$, alors $\mathrm{Ker}(D) = \{\lambda,\ \lambda \in \mathbb{K}\}$, l'ensemble des polynômes constants.

PROPOSITION 43 (Formules de dérivation)
Soient $P, Q \in \mathbb{K}[X]$, $\lambda \in \mathbb{K}$, $m \geq 1$. On a :

1. $(\lambda P)'(X) = \lambda P'(X)$;
2. $(P+Q)'(X) = P'(X) + Q'(X)$ (dérivée d'une somme) ;
3. $(PQ)'(X) = P'(X)Q(X) + P(X)Q'(X)$ (dérivée d'un produit) ;
4. $(P^m)'(X) = mP'(X)P(X)^{m-1}$ (dérivée d'une puissance) ;
5. $(P \circ Q)'(X) = Q'(X).(P' \circ Q)(X)$ (dérivée d'une composée).

Formule de Taylor

PROPOSITION 44
Soient $\alpha \in \mathbb{K}$, $n \geq 1$, $k \geq 0$. On pose $P(X) = (X-\alpha)^n$. En notant $P^{(k)}$ le polynôme dérivé k-ième de P, on a :

$$P^{(k)}(X) = \frac{n!}{(n-k)!}(X-\alpha)^{n-k} \text{ si } 0 \leq k \leq n, \text{ et } P^{(k)}(X) = 0 \text{ si } k > n.$$

On en déduit que $P^{(n)}(X) = n!$, et que $P^{(k)}(\alpha) = 0$ si $k \neq n$.

THÉORÈME 45 (Formule de Taylor pour les polynômes)
Soit $\mathbb{K}$ un corps contenant $\mathbb{Z}$. Soient $a \in \mathbb{K}$ et $P \in \mathbb{K}[X]$ de degré n. On a l'égalité suivante :

$$P = \sum_{k=0}^{n} \frac{P^{(k)}(a)}{k!}(X-a)^k = P(a) + \frac{P'(a)}{1!}(X-a) + \ldots + \frac{P^{(n)}(a)}{n!}(X-a)^n.$$

REMARQUE 46 — *Nous avons montré que $\{Q_0, Q_1, \ldots, Q_n\}$ est une base de $\mathbb{K}_n[X]$. Pour P un polynôme de degré au plus n, la formule de Taylor nous dit alors que les coordonnées de P dans cette base $(P(a), P'(a), \ldots, P^{(n)}(a))$.*

EXEMPLE 47 — *On a $X^2 - 10X + 1 = 1 + \frac{10}{1}(X-10) + \frac{2}{2}(X-10)^2 = 1 + 10(X-10) + (X-10)^2$.*
Appliquer la fomule de Taylor à

1. *$X^3 + X^2 + X + 1$ et $\alpha = 1$;*
2. *$2X^4 + 2X + 1$ et $\alpha = -1$.*

7.4.1 Caractérisation des racines multiples

PROPOSITION 48 (Caractérisation des racines simples)
Soient $\mathbb{K}$ un corps, $a \in \mathbb{K}$, et $P \in \mathbb{K}[X]$.
L'élément a est une racine simple du polynôme P si et seulement si $P(a) = 0$ et $P'(a) \neq 0$.

PROPOSITION 49 (Caractérisation des racines multiples)
Soit $\mathbb{K}$ un corps contenant $\mathbb{Z}$ ($Car(\mathbb{K}) = 0$). Soient $a \in \mathbb{K}$, $k \geq 1$, et $P \in \mathbb{K}[X]$.
Alors a est une racine de P de multiplicité k si et seulement si $P(a)$, $P'(a)$,...,$P^{(k-1)}(a) = 0$ et $P^{(k)}(a) \neq 0$.

REMARQUE 50 — *Le critère de caractérisation des racines simples est vrai dans tout corps $\mathbb{K}$, tandis que celui des racines multiples n'est vrai que dans un corps $\mathbb{K}$ contenant $\mathbb{Z}$ (de caractéristique nulle).*

EXEMPLE 51 — *Dans $\mathbb{C}[X]$, le polynôme $P = X^n - 1$ n'admet que des racines simples, puisque $P' = nX^{n-1}$ n'admet pas de racine commune avec P.*

PROPOSITION 52 (Caractérisation des facteurs multiples 2)
Soient $\mathbb{K}$ un corps contenant $\mathbb{Z}$ et $P \in \mathbb{K}[X]$. Soit $P(X) = a_n P_1(X)^{\alpha_1} \times P_2(X)^{\alpha_2} \ldots P_N(X)^{\alpha_N}$ sa décomposition en produit de facteurs irréductibles, avec P_i irréductibles distincts.
Alors, on a $\text{pgcd}(P, P') = P_1(X)^{\alpha_1 - 1} \times P_2(X)^{\alpha_2 - 1} \ldots P_N(X)^{\alpha_N - 1}$.
Ainsi, P est à facteurs irréductibles simples ($\alpha_1 = \ldots = \alpha_N = 1$) si et seulement si $\text{pgcd}(P, P') = 1$.

EXEMPLE 53 — *Pour $P(X) = X^4 + 3X^2 + 1 \in \mathbb{Q}[X]$, on a $P'(X) = 4X^3 + 6X$. Le calcul donne* $\text{pgcd}(P, P') = 1$. *Ainsi, P est un polynôme à facteurs premiers dans $\mathbb{Q}[X]$. (idem pour $\mathbb{R}$ et $\mathbb{C}$)*

Pour $P \in \mathbb{Q}[X]$

REMARQUE 54 — *Pour $P \in \mathbb{K}[X]$ avec $\mathbb{K} = \mathbb{Q}, \mathbb{R}, \mathbb{C}$, en calculant P' puis* $\text{pgcd}(P, P')$, *ce qui se fait très bien avec l'algorithme d'Euclide, on obtient ainsi un diviseur de P. Si P a un facteur irréductible de multiplicité au moins 2, on a* $\text{pgcd}(P, P') \neq 1$, *donc ce diviseur est non-trivial.*

On peut ensuite effectuer la division euclidienne de P par $\text{pgcd}(P, P')$ *pour obtenir le polynôme $Q = P_1 \ldots P_N$.*

Théorème de Rolle pour les polynômes réels

Proposition 55 (Théorème des valeurs intermédiaires)
Soit $I = [a, b]$ un intervalle de $\mathbb{R}$. Soit $f : [a, b] \to \mathbb{R}$ une fonction continue sur $[a, b]$.
• Si $f(a) \neq f(b)$, pour tout $d \in]f(a), f(b)[$, il existe $c \in]a, b[$ tel que $f(c) = d$.
• Si $f(a) < 0$ et $f(b) > 0$ (ou $f(a) > 0$ et $f(b) < 0$), alors il existe $c \in]a, b[$ tel que $f(c) = 0$.

Proposition 56 (Théorème de Rolle)
Soit $I = [a, b]$ un intervalle de $\mathbb{R}$. Soit $f : [a, b] \to \mathbb{R}$ une fonction continue sur $[a, b]$ et dérivable sur $]a, b[$.
Si $f(a) = f(b)$, alors il existe $c \in]a, b[$ tel que $f'(c) = 0$.

Corollaire 57
Soit $P \in \mathbb{R}[X]$.
• Soient a, b deux racines de P distinctes.
Alors il existe $c \in]a, b[$ tel que $P'(c) = 0$ (c est une racine de P).
• Si P possède r racines distinctes $a_1 < a_2 < \ldots < a_r$, alors le polynôme P' possède au moins $r - 1$ racines $b_1, \ldots, b_{r-1}$ telles que $b_i \in]a_i, a_{i+1}[$.
P' possède donc au moins $r - 1$ racines distinctes qui ne sont pas des racines de P.

Factorisation dans $\mathbb{C}[X]$

Théorème 58 (Théorème de D'alembert-Gauss)
Tout polynôme non constant de $\mathbb{C}[X]$ admet au moins une racine.

Corollaire 59
Soit $P \in \mathbb{C}[X]$ de degré $n \geq 1$. Alors P se décompose en :

$$P = a_n \prod_{i=1}^{r} (X - z_i)^{\alpha_i},$$

où $\alpha_1, \ldots, \alpha_r$ sont des entiers non nuls et $z_1, \ldots, z_r$ sont des nombres complexes deux à deux distincts. Cette décomposition est unique à l'ordre des z_i près.

Remarque 60 — *On peut formuler le corollaire en disant que les polynômes irréductibles dans $\mathbb{C}$ sont exactement les polynômes de degré 1, ou encore en disant que tout polynôme $P \in \mathbb{C}[X]$ se décompose en un produit de polynômes de degré 1.*

Factorisation dans $\mathbb{R}[X]$

La situation dans $\mathbb{R}$ est relativement différente.

Lemme 61
Soit $P \in \mathbb{R}[X]$. Soit $\alpha \in \mathbb{C} \setminus \mathbb{R}$ une racine complexe de P. Alors, $\bar{\alpha}$ est aussi une racine de P.

Preuve — On écrit $P = a_n X^n + \ldots + a_0$ avec $a_i \in \mathbb{R}$. Alors, on a :

$$P(\bar{\alpha}) = a_n \bar{\alpha}^n + \ldots + a_1 \bar{\alpha} + a_0 = \overline{a_n \alpha^n + \ldots + a_1 \alpha + a_0} = \overline{P(\alpha)} = 0.$$

Donc $\bar{\alpha}$ est bien une racine de P. □

Proposition 62
Les polynômes irréductibles de $\mathbb{R}[X]$ sont :

1. Les polynômes de degré 1, $\lambda(X - \beta)$, avec $\lambda \neq 0$;
2. Les polynômes de degré 2, $aX^2 + bX + c$, avec $b^2 - 4ac < 0$.

Exemple 63 —

1. *Le polynôme $X^3 + 1$ n'est pas irréductile dans $\mathbb{R}[X]$ car -1 est une racine. Il se décompose en $X^3 + 1 = (X + 1)(X^2 - X + 1)$.*
2. *$X^4 + 1$ n'a pas de racines sur $\mathbb{R}$ mais n'est pas irréductible. Sa décomposition est :*

$$X^4 + 1 = (X^2 - \sqrt{2}X + 1)(X^2 + \sqrt{2}X + 1).$$

3. Tout polynôme réel P de degré impair admet au moins une racine réelle. (Pourquoi ?)

COROLLAIRE 64
Soit $P \in \mathbb{R}[X]$ de degré $n \geq 1$. Alors P se décompose en :

$$P = a_n \prod_{i=1}^{r} (X - c_i)^{\alpha_i} \times \prod_{j=1}^{m} (X^2 + c_j X + d_j)^{\beta_j},$$

où $\alpha_1, \ldots, \alpha_r, \beta_i, \ldots, \beta_m$ sont des entiers non nuls, les b_i sont distincts, les (c_j, d_j) sont distincts, avec $c_j^2 - 4d_j < 0$. Cette décomposition est unique à l'ordre l'ordre des b_i et des (c_j, d_j) près.

Chapitre 8 Espaces vectoriels - Dimension finie

Table des matières du chapitre

8.1 Définitions - Premières propriétés

Définitions - Exemples

Définition 1
Soit $\mathbb{K}$ un corps. Un **$\mathbb{K}$-espace vectoriel** E est un ensemble muni :

1. D'une addition interne

$$\begin{aligned} E \times E &\to E \\ (x, y) &\mapsto x + y \end{aligned}$$

telle que :
   (a) Pour tous $x, y, z \in E$: $x + (y + z) = (x + y) + z$;
   (b) Il existe un élément de E noté 0, appelé vecteur nul, tel que pour tout $x \in E$: $x + 0 = x$;
   (c) Pour tout $x \in E$, il existe un élément de E noté $-x$, appelé symétrique de x, tel que : $x + (-x) = 0$;
   (d) Pour tous $x, y \in E$: $x + y = y + x$;
2. D'une multiplication externe

$$\begin{aligned} \mathbb{K} \times E &\to E \\ (\lambda, x) &\mapsto \lambda.x \end{aligned}$$

telle que pour tous x, $y \in E$ et tous λ, $\mu \in \mathbb{K}$,
   (a) $(\lambda + \mu).x = \lambda.x + \mu.x$.
   (b) $\lambda.(x + y) = \lambda.x + \lambda.y$.
   (c) $\lambda.(\mu.x) = (\lambda\mu).x$.
   (d) $1.x = x$.

Les éléments de E s'appellent des vecteurs , les éléments de $\mathbb{K}$ s'appellent des scalaires .
On notera également $(E, +, .)$ lorsque l'on veut préciser quelles sont les opérations d'addition et de multiplication.

Remarque 2 — *On notera toujours les lois $+$ et $.$ et le plus souvent, on oubliera même de mettre le point :* $2.x = 2x$.

Exemple 3 —

1. *Un espace vectoriel n'est jamais vide, car il contient au moins le vecteur nul* 0.
   L'ensemble $E = \{0\}$ est un espace vectoriel sur n'importe quel corps de scalaires $\mathbb{K}$. L'addition et la multiplication par un scalaire sont uniquement déterminées.
2. *L'espace $\mathbb{R}^2$ est un $\mathbb{R}$-espace vectoriel. Un vecteur u de $\mathbb{R}^2$ est défini par ses coordonnées $\begin{pmatrix} x \\ y \end{pmatrix}$. La somme et le produit par un scalaire sont définis de la façon suivante :*

$$\forall u = \begin{pmatrix} x \\ y \end{pmatrix}, v = \begin{pmatrix} x' \\ y' \end{pmatrix} \in \mathbb{R}^2,\ \forall \lambda \in \mathbb{R},\ u + v = \begin{pmatrix} x + x' \\ y + y' \end{pmatrix},\ \lambda.u = \begin{pmatrix} \lambda x \\ \lambda y \end{pmatrix}.$$

3. *De la même manière, l'espace* $\mathbb{K}^2$ *est un* $\mathbb{K}$*-espace vectoriel. Un vecteur* u *est défini par ses coordonnées* $\begin{pmatrix} x \\ y \end{pmatrix}$*. La somme et le produit par un scalaire sont définis par :*

$$\forall u = \begin{pmatrix} x \\ y \end{pmatrix}, v = \begin{pmatrix} x' \\ y' \end{pmatrix} \in \mathbb{K}^2,\ \forall \lambda \in \mathbb{K},\ u + v = \begin{pmatrix} x + x' \\ y + y' \end{pmatrix},\ \lambda.u = \begin{pmatrix} \lambda x \\ \lambda y \end{pmatrix}.$$

Exemples fondamentaux

Proposition 4
$\mathbb{R}$ est un $\mathbb{Q}$-espace vectoriel.
$\mathbb{C}$ est un $\mathbb{R}$-espace vectoriel et un $\mathbb{Q}$-espace vectoriel.

Proposition 5
Si E et F sont deux $\mathbb{K}$-espaces vectoriels alors $E \times F$ muni de l'addition

$$\begin{aligned} (E \times F) \times (E \times F) &\to E \times F \\ ((x, y), (x', y')) &\mapsto (x + x', y + y') \end{aligned}$$

et de la multiplication externe

$$\begin{aligned} \mathbb{K} \times (E \times F) &\to E \times F \\ (\lambda, (x, y)) &\mapsto (\lambda x, \lambda y) \end{aligned}$$

est un $\mathbb{K}$-espace vectoriel. $E \times F$ est appelé espace vectoriel produit de E et F.

Exemple 6 — *Si* $n \in \mathbb{N}^*$ *alors* $\mathbb{K}^n$ *est un* $\mathbb{K}$*-espace vectoriel, avec*

$$\begin{cases} \forall u = \begin{pmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{pmatrix}, v = \begin{pmatrix} y_1 \\ y_2 \\ \vdots \\ y_n \end{pmatrix} \in \mathbb{K}^n, & u + v = \begin{pmatrix} x_1 + y_1 \\ x_2 + y_2 \\ \vdots \\ x_n + y_n \end{pmatrix} \\ \forall \lambda \in \mathbb{K}, & \lambda.u = \begin{pmatrix} \lambda x_1 \\ \lambda x_2 \\ \vdots \\ \lambda x_n \end{pmatrix} \end{cases}$$

Le vecteur nul est $(0, 0, \ldots, 0)$.

Proposition 7
Soit E un $\mathbb{K}$-espace vectoriel et X un ensemble quelconque.
Alors l'ensemble $\mathcal{F}(X, E)$ des fonctions $f : X \to E$ est un $\mathbb{K}$-espace vectoriel, les lois étant définies par :

$$\forall f, g \in \mathcal{F}(X, E),\ \forall \lambda \in \mathbb{K},\ f + g : t \mapsto f(t) + g(t),\ \lambda.f : t \mapsto \lambda f(t).$$

Le vecteur nul est alors la fonction constante à 0.

Exemple 8 — 1. $\mathcal{F}([-1, 1], \mathbb{R})$ *est un* $\mathbb{R}$*-espace vectoriel.*

2. $\mathbb{K}^{\mathbb{N}} = \mathcal{F}(\mathbb{N}, \mathbb{K})$*, l'ensemble des suites* $(u_n)_{n \in \mathbb{N}}$ *à valeurs dans* $\mathbb{K}$*, est un* $\mathbb{K}$*-espace vectoriel. On définit l'addition et la multiplication par un scalaire :*

$$\forall (u_n)_{n \in \mathbb{N}}, (v_n) \in \mathbb{K}^{\mathbb{N}},\ \forall \lambda \in \mathbb{K},\ \begin{cases} (u_n)_{n \in \mathbb{N}} + (v_n)_{n \in \mathbb{N}} = (u_n + v_n)_{n \in \mathbb{N}} \\ \lambda.(u_n)_{n \in \mathbb{N}} = (\lambda u_n)_{n \in \mathbb{N}} \end{cases}$$

8.2 Sous-espaces vectoriels

Définition 9
Soit E un $\mathbb{K}$-espace vectoriel.
On dit que $F \subset E$ est un **sous-espace vectoriel** de E si $F \neq \emptyset$ et si

$$\forall \lambda, \mu \in \mathbb{K},\ \forall x, y \in F, \lambda x + \mu y \in F.$$

Remarque 10 — *Si F est un sous-espace vectoriel de E alors le vecteur nul 0 de E appartient à F.*

Remarque 11 — *Les ensembles $\{0\}$ et E sont des sous-espaces vectoriels de E.*

Proposition 12
Si F est un sous-espace vectoriel du $\mathbb{K}$-espace vectoriel $(E, +, .)$, alors $(F, +, .)$ est un $\mathbb{K}$-espace vectoriel.

Preuve — Vous pouvez faire la vérification sans problème. □

Remarque 13 — *Pour montrer qu'un espace est un espace vectoriel, on montrera le plus souvent que c'est un sous-espace vectoriel d'un espace vectoriel plus grand.*

Méthode 14 — *Soit E un $\mathbb{K}$-espace vectoriel et $F \subset E$. F est un sous-espace vectoriel de E si et seulement si*

1. *$0_E \in F$,*
2. *F est stable par addition : pour tous $x, y \in F$, $x + y \in F$,*
3. *F est stable par multiplication par un scalaire : pour tout $\lambda \in \mathbb{K}$ et tout $x \in F$, $\lambda x \in F$.*

Exemple 15 — *On sait que $\mathcal{F}(\mathbb{R}, \mathbb{R})$ est un $\mathbb{R}$-espace vectoriel. On veut montrer que l'espace $\mathcal{C}^0(\mathbb{R}, \mathbb{R})$ des fonctions continues de $\mathbb{R}$ dans $\mathbb{R}$ est un sous-espace vectoriel de $\mathcal{F}(\mathbb{R}, \mathbb{R})$. On a bien $\mathcal{C}^0(\mathbb{R}, \mathbb{R}) \subset \mathcal{F}(\mathbb{R}, \mathbb{R})$ et :*

1. *La fonction nulle est continue ;*
2. *Si $f : \mathbb{R} \to \mathbb{R}$ et $g : \mathbb{R} \to \mathbb{R}$ sont continues alors $f + g : \mathbb{R} \to \mathbb{R}$ est continue, donc $\mathcal{C}^0(\mathbb{R}, \mathbb{R})$ est stable par addition ;*
3. *Si $\lambda \in \mathbb{R}$ et $f : \mathbb{R} \to \mathbb{R}$ est continue alors $\lambda f : \mathbb{R} \to \mathbb{R}$ est continue, donc $\mathcal{C}^0(\mathbb{R}, \mathbb{R})$ est stable par multiplication par un scalaire.*

On en déduit que $\mathcal{C}^0(\mathbb{R}, \mathbb{R})$ est un sous-espace vectoriel de $\mathcal{F}(\mathbb{R}, \mathbb{R})$.

Proposition 16
Toute intersection de sous-espaces vectoriels de E est un sous-espace vectoriel de E.

Preuve — L'intersection est non vide : elle contient le vecteur nul 0_E car tous les sous-espaces vectoriels de E contiennent 0_E. Soit $x, y \in \bigcap_{i \in I} F_i$ où les F_i sont des sous-espaces vectoriels de E, soit $\lambda, \mu \in \mathbb{K}$. Alors $\lambda x + \mu y \in F_i$ pour tout $i \in I$, car F_i est un sous-espace vectoriel, et donc $\lambda x + \mu y \in \bigcap_{i \in I} F_i$. □

Remarque 17 — ***Attention*** *: Si F et G sont des sous-espaces vectoriels de E, en général $F \cup G$ n'est pas un sous-espace vectoriel.*
Contre-exemple : $F = \mathbb{R}(1, 0)$ et $G = \mathbb{R}(0, 1)$ sont des sous-espaces vectoriels de $\mathbb{R}^2$, or $(1, 0) + (0, 1) = (1, 1) \notin F \cup G$.

8.3 Somme de sous-espaces vectoriels, combinaisons linéaires

Combinaisons linéaires

Définition 18
On dit que x est une combinaison linéaire de x_1, ..., x_n s'il existe λ_1, ..., $\lambda_n \in \mathbb{K}$ tels que

$$x = \lambda_1 x_1 + \cdots + \lambda_n x_n.$$

Proposition 19
Une partie $F \subset E$ d'un espace vectoriel E est un sous-espace vectoriel si et seulement si F est non vide et F est stable par combinaison linéaire : pour tous x_1, ..., $x_n \in F$ et tous λ_1, ..., $\lambda_n \in \mathbb{K}$, on a

$$\lambda_1 x_1 + \cdots + \lambda_n x_n \in F.$$

Définition 20
Soit $A \subset E$ une partie de E, on définit le **sous-espace vectoriel engendré par** A comme l'ensemble des combinaisons linéaires d'éléments de A :

$$\mathrm{Vect}(A) := \left\{ \sum_{i=1}^{n} \lambda_i x_i \mid n \in \mathbb{N}, \lambda_1, \ldots, \lambda_n \in \mathbb{K}, x_1, \ldots, x_n \in A \right\}.$$

C'est un sous-espace vectoriel de E.
En particulier, si $x_1, \dots, x_n \in E$, on note l'ensemble des combinaisons linéaires de $x_1, \dots, x_n$:

$$\text{Vect}(x_1, \cdots, x_n) = \{\lambda_1 x_1 + \cdots + \lambda_n x_n \mid \lambda_1, \cdots, \lambda_n \in \mathbb{K}\}.$$

PROPOSITION 21
Vect(A) est le plus petit sous-espace vectoriel contenant A.
De même, Vect$(x_1, \dots, x_n)$ est le plus petit sous-espace vectoriel contenant tous les vecteurs $x_1, \dots, x_n$.

Preuve — Soit H un sous-espace vectoriel de E tel que $A \subset H$.
H est stable par combinaison linéaire donc Vect$(A) \subset H$. □

EXEMPLE 22 —

1. *Dans $\mathbb{R}^3$ muni de (i, j, k), soit $e_1 = (0, 1, 2)$ et $e_2 = (0, 2, 3)$. Alors* Vect(e_1, e_2) *est le plan vectoriel engendré par e_1 et e_2. On a $k = 2e_1 - e_2 \in \text{Vect}\,(e_1, e_2)$ et $j = e_1 - 2k = -3e_1 + 2e_2 \in \text{Vect}\,(e_1, e_2)$, donc* Vect $(j, k) \subset$ Vect (e_1, e_2). *Et réciproquement, $e_1, e_2 \in \text{Vect}\,(j, k)$, d'où* Vect $(e_1, e_2) =$ Vect (j, k).
2. *Soit* $P = \text{Vect}\left(\begin{pmatrix}1\\1\\-1\end{pmatrix}, \begin{pmatrix}1\\2\\3\end{pmatrix}\right)$. *On a*
$$P = \left\{ x\begin{pmatrix}1\\1\\-1\end{pmatrix} + y\begin{pmatrix}1\\2\\3\end{pmatrix} = \begin{pmatrix}x+y\\x+2y\\-x+3y\end{pmatrix} \mid x, y \in \mathbb{K}^2 \right\}.$$
3. *Soit* $F = \left\{ \begin{pmatrix}2x-y\\x+2y\\3x-2y\end{pmatrix},\ x, y \in \mathbb{K}^2 \right\} \subset \mathbb{K}^3$. *Alors F est un sous-espace vectoriel car*
$$F = \left\{ x\begin{pmatrix}2\\1\\3\end{pmatrix} + y\begin{pmatrix}-1\\2\\-2\end{pmatrix},\ x, y \in \mathbb{K} \right\} = \text{Vect}\left(\begin{pmatrix}2\\1\\3\end{pmatrix}, \begin{pmatrix}-1\\2\\-2\end{pmatrix}\right).$$

DÉFINITION 23
Soit E un $\mathbb{K}$-espace vectoriel et soient F et G deux sous-espaces vectoriels de E. On définit la **somme** de F et G de la façon suivante :

$$F + G := \{u + v \mid u \in F, v \in G\}.$$

PROPOSITION 24
Si F et G sont deux sous-espaces vectoriels de E alors $F + G$ est un sous-espace vectoriel de E.

Preuve — $F + G \subset E$ et on a $0 \in F$ et $0 \in G$ donc $0 = 0 + 0 \in F + G$. Soient $\lambda \in \mathbb{K}$ et $x, y \in F + G$, c'est-à-dire qu'il existe $x_F, y_F \in F$ et $x_G, y_G \in G$ tels que $x = x_F + x_G$ et $y = y_F + y_G$. On a donc $x + y = (x_F + y_F) + (x_G + y_G) \in F + G$ et $\lambda x = \lambda x_F + \lambda x_G \in F + G$. □

EXEMPLE 25 — *Si $F = \mathbb{R}x_1$ et $G = \mathbb{R}x_2$, alors $F + G = \mathbb{R}x_1 + \mathbb{R}x_2 = \text{Vect}(x_1, x_2)$.*
$F + G$ est l'ensemble des combinaisons linéaires de x_1 et x_2.

EXEMPLE 26 —

1. $\mathbb{C} = \mathbb{R}1 + \mathbb{R}i$.
2. $\mathbb{R}^3 = \{(0, y, z) \mid y, z \in \mathbb{R}\} + \{(x, y, 0) \mid x, y \in \mathbb{R}\}$.
3. $\mathbb{R}^{\mathbb{N}} = \{(u_n) \in \mathbb{R}^{\mathbb{N}} \mid u_0 = 0\} + \{(u_n) \in \mathbb{R}^{\mathbb{N}} \mid u_1 = 0\}$.
   En effet, soit $u \in \mathbb{R}^{\mathbb{N}}$, $u = v + w$ avec $v_0 = 0$, $v_n = u_n$ si $n > 0$ et $w_0 = u_0$, $w_n = 0$, si $n > 0$.

DÉFINITION 27
Soit $F_1, \dots, F_n$ des sous-espaces vectoriels de E. La somme des F_i est l'ensemble

$$\sum_{i=1}^{n} F_i := \left\{ \sum_{i=1}^{n} x_i \mid \forall i\ x_i \in F_i \right\}.$$

C'est le plus petit sous-espace vectoriel contenant les F_i. C'est aussi l'ensemble des combinaisons linéaires d'éléments des F_i.

EXEMPLE 28 — $\sum_{i=1}^{n} \mathbb{K}x_i = \text{Vect}(x_1, \dots, x_n)$ *est l'ensemble des combinaisons linéaires des x_i.*

Somme directe

Définition 29
Soient F, G deux sous-espaces vectoriels de E.
On dit que F et G sont en **somme directe** si pour tout $x \in F+G$, il existe un **unique** couple $(x_F, x_G) \in F \times G$ tel que $x = x_F + x_G$.
On note alors la somme $F \oplus G$.

Théorème 30
Deux sous-espaces vectoriels F et G sont en somme directe si et seulement si $F \cap G = \{0\}$.

Exemple 31 — *$F = \mathbb{R}\begin{pmatrix}1\\0\end{pmatrix}$ et $G = \mathbb{R}\begin{pmatrix}0\\1\end{pmatrix}$ sont des sous-espaces vectoriels de $\mathbb{R}^2$. Soit $u = \begin{pmatrix}x\\y\end{pmatrix} \in F \cap G$. $u \in F$ donc $y = 0$, $u \in G$ donc $x = 0$. Finalement, $u = 0$ et donc F et G sont en somme directe.*

Définition 32
Deux sous-espaces vectoriels F et G de E sont dits **supplémentaires** s'ils sont en somme directe et si $F + G = E$.
On note alors $E = F \oplus G$. On a donc

$$E = F \oplus G \iff (E = F + G \text{ et } F \cap G = \{0\}).$$

Définition 33
Soit F_1, ..., F_n des sous espaces vectoriels de E, on dit qu'ils sont en **somme directe** si pour tout $x \in \sum_{i=1}^{n} F_i$, il existe un **unique** n-uplet $(x_1, \ldots, x_n) \in F_1 \times \ldots \times F_n$ tel que $x = \sum_{i=1}^{n} x_i$.
On dit qu'ils sont **supplémentaires** s'ils sont en somme directe et si $\sum_{i=1}^{n} F_i = E$.

Proposition 34
Soit $F_1, \ldots, F_n$ des sous espaces vectoriels de E. Les conditions suivantes sont équivalentes :

1. $F_1, \ldots, F_n$ sont en somme directe.
2. $\sum_{i=1}^{n} x_i = 0$ avec $x_i \in F_i$ entraîne $x_i = 0$ pour tout i.
3. Pour tout i, $F_i \cap \sum_{j \neq i} F_j = \{0\}$.

Remarque 35 —

1. *On a $\mathbb{K}^3 = \mathbb{K}.\begin{pmatrix}1\\0\\0\end{pmatrix} \oplus \mathbb{K}\begin{pmatrix}0\\1\\0\end{pmatrix} \oplus \mathbb{K}\begin{pmatrix}0\\0\\1\end{pmatrix}$, car tout vecteur $u = \begin{pmatrix}x\\y\\z\end{pmatrix} \in \mathbb{K}^3$ s'écrit de manière unique comme une combinaison linéaire de ces trois vecteurs :*
$$u = x\begin{pmatrix}1\\0\\0\end{pmatrix} + y\begin{pmatrix}0\\1\\0\end{pmatrix} + z\begin{pmatrix}0\\0\\1\end{pmatrix}.$$
2. *D'après la définition 33, $F = F_1 \oplus F_2 \cdots \oplus F_n$ si et seulement si tout vecteur $x \in F$ se décompose de manière unique dans $F_1, \ldots, F_n$: $x = x_1 + \ldots + x_n$.*
3. ***Attention** : pour montrer que F_1 et F_2 sont en somme directe, il suffit de montrer que $F_1 \cap F_2 = \{0\}$, mais ce n'est plus vrai pour $n \geq 3$:*
$$\forall i \neq j \in \{1, \cdots, n\},\ F_i \cap F_j = \{0\} \not\Rightarrow F_1 \oplus \cdots \oplus F_n$$
comme le montre l'exemple dans $\mathbb{R}^2$: $F_1 = \mathbb{R}\begin{pmatrix}1\\0\end{pmatrix}$, $F_2 = \mathbb{R}\begin{pmatrix}0\\1\end{pmatrix}$ et $F_3 = \mathbb{R}\begin{pmatrix}1\\1\end{pmatrix}$.

8.4 Familles libres, familles génératrices, bases

Familles libres

Définition 36

1. Une famille de vecteurs $(x_1, \dots, x_n)$ est dite **libre** si elle vérifie pour tous $\lambda_1, \dots, \lambda_n \in \mathbb{K}$:
$$\sum_{i=1}^{n} \lambda_i x_i = 0 \Longrightarrow \lambda_i = 0 \text{ pour tout } i.$$
On dit aussi que les vecteurs $x_1, \dots, x_n$ sont linéairement indépendants .
2. Une famille qui n'est pas libre est dite **liée** : $(x_1, \dots, x_n)$ est liée si et seulement si il existe $\lambda_1, \dots, \lambda_n \in \mathbb{K}$ tels que
$$\sum_{i=1}^{n} \lambda_i x_i = 0,$$
avec au moins l'un des λ_i non nuls. On dit aussi que les vecteurs $x_1, \dots, x_n$ sont linéairement dépendants .
3. Une famille de vecteurs $(x_i)_{i\in I}$ non nécessairement finie est dite **libre** si toute sous-famille finie $(x_{i_1}, \dots, x_{i_n})$ est libre, pour tout $n \in \mathbb{N}^*$.
Dans le cas contraire, la famille est dite **liée**.

Exemple 37 —

1. *Une famille de deux vecteurs (x_1, x_2) est liée si et seulement si x_1 et x_2 sont colinéaires , c'est-à-dire s'il existe $\lambda \in \mathbb{K}$ tel que $x_1 = \lambda x_2$ ou $x_2 = \lambda x_1$.*
2. *Une famille de trois vecteurs (x_1, x_2, x_3) dans $\mathbb{R}^3$ est libre si et seulement si ils ne sont pas coplanaires (leur déterminant est non nul).*
3. ***Attention :** $(1, i)$ est libre dans $\mathbb{R}$ mais est liée dans $\mathbb{C}$.*

Caractérisation

Exemple 38 — *Si une famille $(u_1, \dots, u_n)$ contient deux vecteurs colinéaires alors elle est liée.*

Proposition 39
Soit $(x_1, \dots, x_n)$ une famille de vecteurs telle que $x_1 \neq 0$ et pour tout $2 \leq i \leq n$, $x_i \notin \text{Vect}\,(x_1, \dots, x_{i-1})$.
Alors la famille est libre.

Preuve — On procède par récurrence sur n :

- **Initialisation :** Pour $n = 1$, la propriété est évidente.
- **Hérédité :** Soit $n \geq 2$, supposons que la proposition est vraie pour toute famille de $n-1$ vecteurs. Soit $(x_1, \dots, x_n)$ tels que $x_1 \neq 0$ et pour tout $2 \leq i \leq n$, $x_i \notin \text{Vect}\,(x_1, \dots, x_{i-1})$. Si on a relation de dépendance $\lambda_1 x_1 + \dots + \lambda_{n-1} x_{n-1} + \lambda_n x_n = 0$, alors $\lambda_n = 0$ car sinon
$$x_n = \frac{1}{\lambda_n}(\lambda_1 x_1 + \dots \lambda_{n-1} x_{n-1}) \in \text{Vect}\,(x_1, \cdots, x_{n-1}),$$
ce qui contredit les hypothèses. On a donc une relation de dépendance entre $x_1, \dots, x_{n-1}$, et comme par récurrence cette famille est libre, on en déduit que les λ_i sont nuls pour tout i. La famille est donc libre.

□

Proposition 40
Soit F et G deux sous-espaces vectoriels de E en somme directe, $(x_1, \dots, x_n)$ une famille libre de F et $(y_1, \dots, y_m)$ une famille libre de G.
Alors $(x_1, \dots, x_n, y_1, \dots, y_m)$ est une famille libre de $F \oplus G$.

Preuve — Supposons $\alpha_1 x_1 + \cdots + \alpha_n x_n + \beta_1 y_1 + \cdots + \beta_m y_m = 0$, alors
$$\alpha_1 x_1 + \cdots + \alpha_n x_n = -(\beta_1 y_1 + \cdots + \beta_m y_m)$$
et ce vecteur appartient à $F \cap G$. F et G sont en somme directe, donc $\alpha_1 x_1 + \dots + \alpha_n x_n = 0$ et $\beta_1 y_1 + \cdots + \beta_m y_m = 0$. Les familles (x_i) et (y_i) sont libres donc tous les α_i et les β_i sont nuls, ce qui permet de conclure. □

Familles génératrices

Définition 41
Une partie $A \subset E$ est dite **génératrice** si $\text{Vect}(A) = E$.
Dit autrement, une famille de vecteurs $(x_1, \dots, x_n)$ est génératrice si tout vecteur de E est combinaison linéaire des $x_1, \dots, x_n$: pour tout $y \in E$, il existe $\lambda_1, \dots, \lambda_n \in \mathbb{K}$ tels que $y = \lambda_1 x_1 + \dots + \lambda_n x_n$.

Exemple 42 —

1. *La famille* $\big(e_1 = (1, 0, \dots, 0), \dots, e_n = (0, \dots, 0, 1)\big)$ *est une famille génératrice de* $\mathbb{K}^n$. *En effet, soit* $x = (x_1, \dots, x_n) \in \mathbb{K}^n$, *alors* $x = \sum_{i=1}^{n} x_i e_i$.
2. *Soit*
$$\mathbb{K}_n[X] := \{(u_k)_{k\in\mathbb{N}} \in \mathbb{K}[X] \mid \forall k \geq n+1,\ u_k = 0\}$$
l'espace vectoriel des suites dont les termes sont tous nuls à partir du rang $n+1$. *La famille* $(X^0, \dots, X^n)$ *engendre* $K_n[X]$. *En effet, si* u *est une suite dont tous les termes sont nuls à partir du rang* $n+1$, *on a* $u = \sum_{i=0}^{n} u_i X^i$.
3. *La famille* $(g_n)_{n\in\mathbb{N}}$, *avec* $g_k(x) = x^k$, *est génératrice de l'espace* $\mathbb{K}[x]$ *des fonctions polynomiales de* $\mathbb{R}$ *dans* $\mathbb{R}$.
4. *Si on rajoute des vecteurs à une famille génératrice, elle reste génératrice.*

Proposition 43
Soit F et G deux sous-espaces vectoriels de E, $(x_1, \dots, x_n)$ une famille génératrice de F et $(y_1, \dots, y_m)$ une famille génératrice de G. Alors $(x_1, \dots, x_n, y_1, \dots, y_m)$ est une famille génératrice de $F + G$.

Preuve — Tout élément $z \in F + G$ s'écrit $z = f + g$ avec $f \in F$ et $g \in G$. Or par hypothèse, $f = \lambda_1 x_1 + \dots + \lambda_n x_n$ et $g = \mu_1 y_1 + \dots + \mu_m y_m$. Donc
$$z = \lambda_1 x_1 + \dots + \lambda_n x_n + \mu_1 y_1 + \dots + \mu_m y_m \in \text{Vect}\,(x_1, \cdots, x_n, y_1, \cdots, y_m),$$
et la famille est bien génératrice. □

Bases

Définitions

Définition 44
Une **base** d'un espace vectoriel E est une famille de vecteurs de E qui est à la fois libre et génératrice.

Exemple 45 —

1. $(1, i)$ *est une base de* $\mathbb{C}$ *comme* $\mathbb{R}$*-espace vectoriel.* (1) *est une base de* $\mathbb{C}$ *comme* $\mathbb{C}$*-espace vectoriel.*
2. *La famille* $(E_{ij})_{1\leq i\leq n,\, 1\leq j\leq p}$ *est une base de l'ensemble* $\mathscr{M}_{n,p}(\mathbb{K})$ *des matrices à* n *lignes et* p *colonnes (voir cours d'Algèbre 1).*
3. *Pour* $1 \leq i \leq n$, *soit* $e_i := (0, \dots, 0, 1, 0, \dots, 0)$ *où le* 1 *est en* i*-ème position. On a vu dans l'exemple* **??** *que la famille* $(e_1, \dots, e_n)$ *est libre dans* $\mathbb{K}^n$ *et dans l'exemple 42 qu'elle génère* $\mathbb{K}^n$. *C'est donc une base, que l'on appelle base canonique de* $\mathbb{K}^n$.
4. *On a également vu dans l'exemple* **??** *que la famille* $(X^0, \dots, X^n)$ *était libre et dans l'exemple 42 qu'elle génère* $\mathbb{K}_n[X]$, *c'est donc une base de* $\mathbb{K}_n[X]$, *que l'on appelle aussi base canonique de* $\mathbb{K}_n[X]$.
5. *De la même façon, la famille* $(X^0, X^1, X^2, \dots)$ *est une base (infinie) de* $\mathbb{K}[X]$, *on l'appelle la base canonique de* $\mathbb{K}[X]$.

Définition 46
Soit E un espace vectoriel admettant une base finie $(e_1, \dots, e_n)$.
On appelle coordonnées du vecteur $u \in E$ dans cette base l'unique n-uplet $(x_1, \dots, x_n)$ tel que
$$u = x_1 e_1 + \dots + x_n e_n.$$

Remarque 47 — *L'existence vient du fait que la famille est génératrice, l'unicité du fait qu'elle est libre : supposons qu'il existe un autre* n*-uplet* $(y_1, \dots, y_n)$ *tel que* $u = y_1 e_1 + \dots + y_n e_n$.
En prenant la différence, on obtient $(x_1 - y_1)e_1 + \dots + (x_n - y_n)e_n = 0$, *ce qui implique que* $x_i = y_i$ *pour tout* i *car la famille est libre.*

Caractérisation

PROPOSITION 48
Si E est engendré par une famille à n éléments $(e_1, \dots, e_n)$, alors toute famille libre $(f_1, \dots, f_m)$ a au plus n éléments (c'est-à-dire $m \leq n$).

Preuve — On procède par récurrence sur n. Le cas $n = 1$ est évident.
Soit $n \geq 1$. Supposons $\mathcal{P}(n)$: "dans un espace vectoriel engendré par n éléments, toute famille libre a au plus n éléments" est vrai.
Montrons $\mathcal{P}(n+1)$: supposons que E est engendré par $n+1$ éléments et soit $(f_1, \dots, f_m)$ une famille libre, on veut montrer que $m \leq n+1$.
Si pour tout i, $f_i \in \text{Vect}(e_1, \dots, e_n)$, alors $(f_1, \dots, f_m)$ est une famille libre d'un espace engendré par n éléments et donc $m \leq n$ par hypothèse de récurrence.
Sinon, il existe i_0 tel que

$$f_{i_0} = \lambda_{i_0} e_{n+1} + \text{ une combinaison linéaire des } e_1, \dots, e_n \text{ avec } \lambda_{i_0} \neq 0.$$

Mais pour tout $i \in \{1, \dots, m\}$,

$$f_i = \lambda_i e_{n+1} + \text{ une combinaison linéaire des } e_1, \text{ ..., } e_n.$$

donc pour $i \neq i_0$, $f_i - \frac{\lambda_i}{\lambda_{i_0}} f_{i_0} \in \text{Vect}(e_1, \dots, e_n)$. Ces vecteurs forment donc une famille libre de $m-1$ vecteurs dans un espace engendré par n vecteurs d'où, par hypothèse de récurrence, $m - 1 \leq n$ et donc $m \leq n+1$. □

COROLLAIRE 49 (Fondamental)
Soit $n \geq 1$.
Toute base d'un espace vectoriel engendré par n éléments a au plus n éléments.

COROLLAIRE 50 (Très pratique)
Soit $n \geq 1$.
Toute famille libre à n éléments dans un espace E engendré par n éléments est une base.

Preuve — Soit une famille $(e_1, \dots, e_n)$ libre. Montrons que la famille est génératrice : soit $x \in E$, alors la famille $(e_1, \dots, e_n, x)$ est liée car elle contient $n+1$ éléments et E est engendré par n éléments. On a donc $\lambda_1 e_1 + \dots + \lambda_n e_n + \lambda_{n+1} x = 0$ avec les λ_i non tous nuls. Si $\lambda_{n+1} = 0$, alors les autres λ_i sont nuls puisque la famille $(e_1, \dots, e_n)$ est libre, donc $\lambda_{n+1} \neq 0$. On a alors

$$x = -\frac{\lambda_1}{\lambda_{n+1}} e_1 + \dots + \frac{\lambda_n}{\lambda_{n+1}} e_n \in \text{Vect}(e_1, \dots, e_n).$$

La famille $(e_1, \dots, e_n)$ est donc génératrice et forme une base. □

EXEMPLE 51 —

1. *Définissons les vecteurs $f_1 = (1, 2, 0)$, $f_2 = (1, 0, 3)$ et $f_3 = (0, 1, 0)$ de $\mathbb{R}^3$. Ils forment une famille libre de $\mathbb{R}^3$ et donc une base (car $\mathbb{R}^3$ est engendré par les 3 vecteurs de sa base canonique).*
2. *Deux vecteurs non colinéaires dans $\mathbb{R}^2$ ou trois vecteurs non coplanaires dans $\mathbb{R}^3$ forment une base.*

8.5 DIMENSION, ESPACES VECTORIELS DE DIMENSION FINIE

Dimension d'un espace vectoriel

DÉFINITION 52
Un $\mathbb{K}$-espace vectoriel E est dit de **dimension finie** si E admet une famille génératrice finie.
Dans le cas contraire, on dit que E est de **dimension infinie**.

EXEMPLE 53 —

1. *$\mathbb{K}^n$ est de dimension finie. On a vu en effet dans l'exemple 45 que la famille $(e_1, \dots, e_n)$ génère $\mathbb{K}^n$. C'est même une base.*
2. *$\mathbb{K}_n[X]$ est de dimension finie.*
3. *$\mathscr{M}_{n,p}(\mathbb{K})$ est de dimension finie.*
4. *$\mathbb{K}[X]$ est de dimension infinie. En effet, d'après la Proposition **??**, s'il existait une famille génératrice finie de $\mathbb{K}[X]$, toute famille libre aurait un cardinal plus petit. Donc toute famille libre serait finie, ce qui contredit l'exemple **??**, où on a montré que la famille infinie $(X^0, X^1, \dots)$ est libre dans $\mathbb{K}[X]$.*

5. *On prouve de la même façon que l'ensemble $\mathcal{F}(\mathbb{R},\mathbb{R})$ est de dimension infinie. En effet, on a montré dans l'exemple* **??** *que pour tout $n \in \mathbb{N}^*$ et tous réels distincts $\alpha_1 < \cdots < \alpha_n$, la famille de fonctions $(x \mapsto e^{\alpha_1 x}, \ldots, x \mapsto e^{\alpha_n x})$ est libre.*

THÉORÈME 54
Soit $E \neq \{0\}$ un $\mathbb{K}$-espace vectoriel de dimension finie. Alors

1. E admet une base finie.
2. Toute base de E a le même nombre déléments n.

On dit alors que E est de dimension n sur $\mathbb{K}$, ce que l'on note $\dim_{\mathbb{K}} E = n$ ou plus simplement $\dim E = n$ si le corps $\mathbb{K}$ est clair. Par convention si $E = \{0\}$, on pose $\dim E = 0$.

Preuve — 1. E est de dimension finie donc il existe une famille génératrice finie $(f_1, \ldots, f_m)$: on a $E = \text{vect}\,(f_1, \ldots, f_m)$. On va en extraire une base de E.
On choisit $e_1 \in \{f_1, \ldots, f_m\}$ non nul (existe car $E \neq \{0\}$).
Puis on choisit $e_2 \in \{f_1, \ldots, f_m\} \setminus \{e_1\}$ tel que $e_2 \notin \text{vect}(e_1)$.
Puis on choisit $e_3 \in \{f_1, \ldots, f_m\} \setminus \{e_1, e_2\}$ tel que $e_3 \notin \text{vect}(e_1, e_2)$, etc.
On continue tant qu'il existe $1 \leq r \leq m$ et $e_r \in \{f_1, \ldots, f_m\} \setminus \{e_1, \ldots, e_{r-1}\}$ tels que $e_r \notin \text{vect}\,(e_1, \ldots, e_{r-1})$.
On s'arrête quand pour tout i, $f_i \in \text{vect}(e_1, \ldots, e_r\}$. La famille $(f_1, \ldots, f_m)$ étant génératrice, on en déduit que $(e_1, \ldots, e_r)$ l'est aussi. De plus, comme pour tout $2 \leq i \leq r$, $e_i \notin \text{vect}(e_1, \ldots, e_{i-1})$, la Proposition **??** nous dit que la famille $(e_1, \ldots, e_r)$ est libre. C'est donc une base finie de E.
2/ Si $(e_1, \ldots, e_r)$ et $(e'_1, \ldots, e'_n)$ sont deux bases de E, alors $(e_1, \cdots, e_r)$ est une famille génératrice et $(e'_1, \cdots, e'_n)$ est une famille libre, la proposition **??** nous dit donc que $n \leq r$. De même, on montre que $n \leq r$ et donc $n = r$. □

REMARQUE 55 — *On a en fait montré que l'on peut extraire une base de toute famille génératrice.*

Exemples

EXEMPLE 56 —

1. $\mathbb{K}^n$ *est de dimension n car* $\left(e_1 = \begin{pmatrix} 1 \\ 0 \\ \vdots \\ 0 \end{pmatrix}, \ldots, e_n = \begin{pmatrix} 0 \\ \vdots \\ 0 \\ 1 \end{pmatrix} \right)$ *en est une base.*
2. ***Attention*** *: $\mathbb{C}$ est de dimension 1 en tant que $\mathbb{C}$-espace vectoriel mais est de dimension 2 en tant que $\mathbb{R}$-espace vectoriel (en effet, la famille $(1, i)$ est une base de $\mathbb{C}$). Il est possible de montrer que $\mathbb{R}$ est de dimension infinie en tant que $\mathbb{Q}$-espace vectoriel.*
3. $\mathbb{K}_n[X]$ *est de dimension $n+1$ car $(X^0, X^1, \ldots, X^n)$ en est une base.*
4. *L'espace de matrices $\mathscr{M}_{n,p}(\mathbb{K})$ est de dimension np, car la famille $(E_{ij})_{1 \leq i \leq n,\, 1 \leq j \leq p}$ est une base.*
5. *Dans $\mathbb{R}^3$, on pose $e_1 = (1,1,0)$, $e_2 = (0,1,2)$, $e_3 = (-1,0,2)$ et $e_4 = (0,1,1)$. La famille (e_1, e_2, e_3, e_4) est génératrice. On a $e_2 \notin \text{vect}(e_1)$, $e_3 \in \text{vect}(e_1, e_2)$ et $e_4 \notin \text{vect}(e_1, e_2)$. Donc la famille (e_1, e_2, e_4) est une base de $\mathbb{R}^3$.*

8.5.1 Caractérisation des bases en dimension finie

PROPOSITION 57
Soit E un espace vectoriel de dimension $n \geq 1$ et $(f_1, \cdots, f_n)$ une famille de vecteurs de E. Alors

$$(f_1, \cdots, f_n) \text{ est une base} \iff (f_1, \cdots, f_n) \text{ est libre} \iff (f_1, \cdots, f_n) \text{ est génératrice.}$$

Preuve —

- Si $(f_1, \ldots, f_n)$ est une base, c'est une famille libre.
- Soit $(f_1, \ldots, f_n)$ une famille libre. Puisque E est engendré par n vecteurs, le Corollaire **??** s'applique.
- Soit $(f_1, \ldots, f_n)$ une famille génératrice. D'après la Remarque 55, on peut en extraire une base $(e_1, \ldots, e_m)$. Or le Théorème 54 nous dit que $m = n$, donc que $(e_1, \ldots, e_m)$ est la famille $(f_1, \ldots, f_n)$ toute entière. C'est donc une base de E.

□

EXEMPLE 58 — *Définissons la famille (f_1, f_2, f_3) avec $f_1 = (1,0,2)$, $f_2 = (1,1,2)$ et $f_3 = (2,0,2)$. Soit (e_1, e_2, e_3) la base canonique de $\mathbb{R}^3$. On a $e_1 = f_3 - f_1$, $e_2 = f_2 - f_1$ et $e_3 = f_1 - \frac{1}{2} f_3$, donc la famille (f_1, f_2, f_3) est génératrice dans $\mathbb{R}^3$. C'est donc une base puisque $\mathbb{R}^3$ est de dimension 3.*

8.5.2 Théorème de la base incomplète

On a vu que d'une famille génératrice, on peut extraire une base. On étudie ici le processus inverse.

THÉORÈME 59 (Théorème de la base incomplète)
Soit E un espace vectoriel non réduit à $\{0\}$, soit $(e_1, \ldots, e_n)$ une famille génératrice de E et soit $(f_1, \ldots, f_m)$ une famille libre de E.
Alors, on peut complèter la famille $(f_1, \ldots, f_m)$ en une base $(f_1, \cdots, f_{m+p})$.
De plus, il est possible de choisir les vecteurs f_{m+1}, ..., f_{m+p} parmi les e_i.

Preuve — On applique l'algorithme suivant :

1. Si pour tout $1 \leq j \leq n$, $e_j \in \text{vect}(f_1, \ldots, f_m)$, alors $(f_1, \ldots, f_m)$ est génératrice, c'est donc une base et on a fini.
2. Sinon, soit j tel que $e_j \notin \text{vect}(f_1, \ldots, f_m)$, on pose $f_{m+1} = e_j$.
3. On recommence en remplaçant m par $m + 1$.

L'algorithme s'arrête après au plus n étapes. On obtient ainsi une famille $(f_1, \cdots, f_{m+p})$ libre et telle que tout vecteur de $(e_1, \ldots, e_n)$ est combinaison linéaire de $f_1, \ldots, f_{m+p}$. Puisque la famille $(e_1, \ldots, e_n)$ est génératrice, on obtient que la famille $(f_1, \cdots, f_{m+p})$ est génératrice aussi. C'est donc une base de E. □

EXEMPLE 60 — *Soit $\mathbb{R}^3$ muni de sa base canonique (e_1, e_2, e_3) et soit $f_1 = (1, 1, 1)$. On a : $e_1 \notin \text{vect}(f_1)$ et $e_2 \notin \text{vect}(f_1, e_1)$ donc (f_1, e_1, e_2) est une base de $\mathbb{R}^3$.*

8.5.3 Sous-espaces vectoriels et dimension

PROPOSITION 61
Soit E un espace vectoriel de dimension n, et $F \subset E$ un sous-espace vectoriel de E. Alors :

1. F est de dimension finie, et $\dim F \leq \dim E$.
2. Si $\dim F = \dim E$ alors $F = E$.

Preuve — Si $F = \{0\}$ alors F est de dimension finie 0 et $\dim F \leq \dim E$. On suppose maintenant que $F \neq \{0\}$.

1. On applique l'algorithme suivant :
   (a) Soit $f_1 \in F$ tel que $f_1 \neq 0$.
   (b) S'il existe $g \in F$ tel que $g \notin \text{vect}(f_1)$, on pose $f_2 = g$.
   (c) S'il existe $g \in F$ tel que $g \notin \text{vect}(f_1, f_2)$, on pose $f_3 = g$.
   (d) On continue jusqu'à ce que pour tout $g \in F$, $g \in \text{vect}(f_1, \ldots, f_p)$.

   On a alors une famille libre $(f_1, \ldots, f_p)$ dans F donc dans E, et donc $p \leq n$. De plus, $(f_1, \ldots, f_p)$ est génératrice de F donc F est de dimension finie et $\dim F = p \leq n = \dim E$.
2. Si $\dim F = \dim E$ dans la preuve précédente, alors $p = n$ et la famille libre $(f_1, \ldots, f_p)$ est une base de E. Donc $E = \text{vect}(f_1, \ldots, f_p) = F$.

□

EXEMPLE 62 —

1. *Une droite vectorielle $\mathbb{K}x_0$ avec $x_0 \neq 0$ est de dimension 1. Tout sous-espace vectoriel de dimension 1 est une droite vectorielle.*
2. *Un plan vectoriel $\mathbb{K}u + \mathbb{K}v$, avec u et v non colinéaires, est de dimension 2. Tout sous-espace vectoriel de dimension 2 est un plan vectoriel.*
3. *Dans $E = \mathbb{R}^4$, on pose $F := \{(x, y, z, t) \in \mathbb{R}^4,\ x + y = 0 \text{ et } z + t = 0\}$. Alors F contient les vecteurs $f_1 = (0, 0, 1, -1)$ et $f_2 = (1, -1, 0, 0)$ donc $\text{vect}(f_1, f_2) \subset F$. La famille (f_1, f_2) est libre donc $\dim F \geq 2$. De plus, si $(x, y, z, t) \in F$, alors $(x, y, z, t) = (x, -x, z, -z)$ puisque $x + y = 0$ et $z + t = 0$, donc $(x, y, z, t) = xf_1 + zf_2 \in \text{vect}(f_1, f_2)$. On en déduit que $F = \text{vect}(f_1, f_2)$ et $\dim F = 2$.*

DÉFINITION 63
Soit E un $\mathbb{K}$-espace vectoriel de dimension n, un sous-espace vectoriel de dimension $n - 1$ est appelé **hyperplan** de E.

EXEMPLE 64 — *Dans $E = \mathbb{R}^4$, soit $G := \{(x, y, z, t) \in \mathbb{R}^4,\ x + y + z + t = 0\}$, alors les vecteurs $(1, -1, 0, 0)$, $(1, 0, -1, 0)$ et $(1, 0, 0, -1)$ forment une famille libre de G, donc $\dim G \geq 3$. Or $G \neq \mathbb{R}^4$ car $(1, 0, 0, 0) \notin G$, donc $\dim G \leq 3$. On en déduit que G est de dimension 3, c'est un hyperplan de $\mathbb{R}^4$. Une base de G est donnée par les vecteurs $(1, -1, 0, 0)$, $(1, 0, -1, 0)$ et $(1, 0, 0, -1)$.*

PROPOSITION 65 (Dimension d'un produit)
Soient E et F deux $\mathbb{K}$-espaces vectoriels de dimension finie. Alors

$$\dim(E \times F) = \dim E + \dim F.$$

Preuve — En effet, si $(e_1, \dots, e_n)$ est une base de E, et $(f_1, \dots, f_m)$ est une base de F, on vérifie facilement que la famille

$$\big((e_1, 0), \cdots, (e_n, 0), (0, f_1), \cdots, (0, f_m)\big)$$

est libre et génératrice, donc une base de $E \times F$. □

THÉORÈME 66 (Formule de Grassmann)
Soit E un espace vectoriel de dimension finie. Soient F et G deux sous-espaces vectoriels de E. Alors

$$\dim(F + G) = \dim F + \dim G - \dim F \cap G.$$

Preuve — Soit $(e_1, \dots, e_p)$ une base de $F \cap G$ que l'on complète (Théorème 59) en deux familles :

$$\begin{cases} (e_1, \dots, e_p, e_{p+1}, \dots, e_n) \text{ base de } F; \\ (e_1, \dots, e_p, e'_{p+1}, \dots, e'_m) \text{ base de } G. \end{cases}$$

On va montrer que $(e_1, \dots, e_p, e_{p+1}, \dots, e_n, e'_{p+1}, \dots, e'_m)$ est une base de $F + G$.

- Soient $\lambda_1, \dots, \lambda_n, \mu_{p+1}, \dots, \mu_m \in \mathbb{K}$ tels que $\lambda_1 e_1 + \ldots + \lambda_n e_n + \mu_{p+1} e'_{p+1} + \ldots + \mu_m e'_m = 0$. Alors

$$\underbrace{\lambda_1 e_1 + \ldots + \lambda_n e_n}_{\in F} = \underbrace{-\mu_{p+1} e'_{p+1} - \ldots - \mu_m e'_m}_{\in G} \in F \cap G.$$

  Or pout tout $p+1 \leq i \leq n$, $e_i \notin F \cap G$ et pour tout $p+1 \leq j \leq m$, $e'_j \notin F \cap G$, donc $\lambda_{p+1} = \cdots = \lambda_n = \mu_{p+1} = \cdots = \mu_m = 0$. On obtient donc $\lambda_1 e_1 + \ldots + \lambda_p e_p = 0$, ce qui implique que $\lambda_1 = \cdots = \lambda_p = 0$ puisque la famille $(e_1, \dots, e_p)$ est libre.
- Soit $u = f + g \in F + G$ avec $f \in F$ et $g \in G$. Alors f est combinaison linéaire des $e_1, \dots, e_n$ et g est combinaison linéaire des $e_1, \dots, e_p, e'_{p+1}, \dots, e'_m$, donc u est combinaison linéaire des $e_1, \dots, e_p, e_{p+1}, \dots, e_n, e'_{p+1}, \dots, e'_m$.

Ainsi $\dim F + G = n + m - p$ avec $\dim F = n$, $\dim G = m$ et $\dim F \cap G = p$. □

COROLLAIRE 67
(Très utile) Soit F et G des sous-espaces vectoriels d'un espace vectoriel E de dimension finie. Alors

$$F \oplus G \iff \dim(F + G) = \dim F + \dim G.$$

REMARQUE 68 —

1. *De même $F_1 \oplus \cdots \oplus F_l$ si et seulement si $\dim(F_1 + \cdots + F_l) = \dim F_1 + \cdots + \dim F_l$.*
2. *L'intersection de deux plans vectoriels distincts dans $\mathbb{R}^3$ est de dimension 1 (la situation change complètement du point de vue affine : les plans peuvent être parallèles).*
3. *Soit E un espace vectoriel de dimension n et F un sous-espace vectoriel de dimension p, alors tout supplémentaire est de dimension $n - p$.*

PROPOSITION 69
Si E est de dimension finie alors tout sous-espace vectoriel possède (au moins) un supplémentaire.

Preuve — Soit F un sous-espace vectoriel de base $(f_1, \dots, f_m)$, on complète cette famille libre en une base $(f_1, \dots, f_{m+p})$ de E. Soit $G := \text{vect}(f_{m+1}, \dots, f_{m+p})$ alors F et G sont supplémentaires. □

PROPOSITION 70
Les hyperplans de $\mathbb{K}^n$ sont exactement les ensembles de la forme

$$\{(x_1, \dots, x_n) \mid a_1 x_1 + \ldots + a_n x_n = 0\}$$

pour une certaine famille $a_1, \dots, a_n$ de scalaires qui ne sont pas tous nuls (c'est-à-dire : $\exists i,\ a_i \neq 0$).

Preuve —

1. Soit $(a_1, \dots, a_n) \neq (0, \dots, 0) \in \mathbb{K}^n$ et soit

   $$H := \{(x_1, \dots, x_n) \mid a_1 x_1 + \ldots + a_n x_n = 0\}.$$

   Quitte à permuter les indices, on suppose que $a_1 \neq 0$. Alors $(x_1, \dots, x_n) \in H$ si et seulement si $x_1 = -\frac{a_2}{a_1} x_2 - \ldots - \frac{a_n}{a_1} x_n$. On définit, pour $2 \leq i \leq n$, $u_i := \left(-\frac{a_i}{a_1}, 0 \dots, 0, 1, 0 \dots, 0\right)$ où 1 est à la i-ème position. On a alors $H = \text{vect}(u_2, \dots, u_n)$ et $(u_2, \dots, u_n)$ est libre. Donc $\dim H = n - 1$ et H est un hyperplan.
2. Admis.

□

Rang d'une famille

Définition 71
On appelle **rang** d'une famille $(e_1, \cdots, e_l) \in E^n$ la dimension de l'espace vectoriel engendré par ces vecteurs :

$$\mathrm{rg}\,(e_1, \ldots, e_l) = \dim \mathrm{vect}(e_1, \ldots, e_l).$$

Proposition 72
On a $\mathrm{rg}(e_1, \ldots, e_l) \leq l$.
De plus, il y a égalité si et seulement si la famille est libre.

Preuve — La famille $(e_1, \ldots, e_l)$ est génératrice de $\mathrm{vect}(e_1, \ldots, e_l)$. Donc $\dim \mathrm{vect}(e_1, \ldots, e_l) = l$ si et seulement si $(e_1, \ldots, e_l)$ est une base, c'est-à-dire si et seulement si cette famille est libre. □

Corollaire 73
Soit E un espace vectoriel de dimension n.
Alors $(e_1, \ldots, e_n)$ est de rang n si et seulement si c'est une base de E.

Chapitre 9 Matrices

Table des matières du chapitre

Dans ce chapitre, après avoir défini les matrices et les opérations sur ces matrices :

- On les utilise pour représenter un système linéaire et pour le résoudre ;
- On étudie les structures de l'ensemble des matrices (structure d'espace vectoriel et d'anneau, sous-ensembles particuliers).

9.1 Définitions

Définition 1
Soit $\mathbb{K}$ un corps. Soient n et p deux entiers naturels non nuls.
Soit A un tableau, avec n lignes, p colonnes, dont les nombres sont dans $\mathbb{K}$. C'est-à-dire :

$$A = \begin{pmatrix} a_{1,1} & a_{1,2} & \dots & a_{1,j} & \dots & a_{1,p} \\ a_{2,1} & a_{2,2} & \dots & a_{2,j} & \dots & a_{2,p} \\ \vdots & \vdots & & \vdots & & \vdots \\ a_{i,1} & a_{i,2} & \dots & a_{i,j} & \dots & a_{i,p} \\ \vdots & \vdots & & \vdots & & \vdots \\ a_{n,1} & a_{n,2} & \dots & a_{n,j} & \dots & a_{n,p} \end{pmatrix}.$$

Les nombres $a_{1,1}, \dots, a_{n,p}$ sont appelés **coefficients** de A.
On dit alors que A est une **matrice** à n lignes et p colonnes à coefficients dans $\mathbb{K}$, ou matrice $n \times p$.
On note $\mathscr{M}_{n,p}(\mathbb{K})$ l'ensemble des matrices $n \times p$.

Pour A une matrice à n lignes et p colonnes, A est définie par ses $n \times p$ coefficients. On écrit aussi cette matrice comme :

$$A = (a_{i,j})_{(i,j)\in[\![1,n]\!]\times[\![1,p]\!]} \quad \text{ou} \quad A = (a_{i,j})_{1\leqslant i\leqslant n, 1\leqslant j\leqslant p}.$$

Les entiers i, j sont appelés **indices** des coefficients de la matrice A.
Le premier indice d'un coefficient est le numéro de sa ligne, et le second indice est le numéro de sa colonne.

Exemple 2 — $A = \begin{pmatrix} 1 & 2 & 3 \\ -1 & 10 & 5 \end{pmatrix}$ *est une matrice à* 2 *lignes et* 3 *colonnes, donc* $A \in \mathscr{M}_{2,3}(\mathbb{Q})$.
$B = \begin{pmatrix} \frac{1}{5} \\ 10 \\ \sqrt{2} \end{pmatrix}$ *est une matrice à* 3 *lignes et* 1 *colonne, donc* $B \in \mathscr{M}_{3,1}(\mathbb{R})$. $C = \begin{pmatrix} 1+i & \exp(\frac{i\Pi}{6}) \\ -1 & \sqrt{5} \end{pmatrix}$ *est une matrice à* 2 *lignes et* 2 *colonnes, donc* $C \in \mathscr{M}_{2,2}(\mathbb{C})$. *Pour* $C = (c_{i,j})_{1\leqslant i\leqslant n, 1\leqslant j\leqslant p}$, *on a* $c_{1,1} = 1+i$, $c_{1,2} = \exp(\frac{i\Pi}{6})$, $c_{2,1} = -1$, $c_{2,2} = \sqrt{5}$.

Définition 3
Soient $\mathbb{K}$ un corps et $n, p \in \mathbb{N}^*$. Soit A une matrice à n lignes et p colonnes à coefficients dans $\mathbb{K}$.

- On dit que la matrice A est **nulle** si tous les coefficients $a_{i,j}$ sont nuls.
  On la note alors $A = 0_{\mathscr{M}_{n,p}(\mathbb{K})}$ ou $0_{n,p}$ ou 0 ;
- Si $p = 1$, on dit alors que A est une **matrice ligne**. Si $q = 1$, on dit alors que A est une **matrice colonne**.
  L'usage est de lire une matrice en suivant ses colonnes ;

- Si $n = p$, on dit alors que $A = (a_{i,j})_{1 \leqslant i,j \leqslant n}$ est une **matrice carrée** .
  On note $\mathscr{M}_n(\mathbb{K})$ l'ensemble $\mathscr{M}_{n,n}(\mathbb{K})$ des matrices de taille $n \times n$.

L'espace vectoriel $\mathscr{M}_{n,p}(\mathbb{K})$

Définition 4
Soit $\mathbb{K}$ un corps. Soient $n, p \in \mathbb{N}^*$.
Pour chaque $(i,j) \in [\![1,n]\!] \times [\![1,p]\!]$, on définit $E_{i,j}$ la matrice de $\mathscr{M}_{n,p}(\mathbb{K})$ dont tous les coefficients sont nuls, sauf le coefficient d'indice (i,j) qui vaut 1. C'est-à-dire :

$$E_{ij} = \begin{pmatrix} a_{1,1}=0 & \dots & a_{1,j}=0 & \dots & a_{1,p}=0 \\ \vdots & & \vdots & & \vdots \\ a_{i,1}=0 & \dots & a_{i,j}=1 & \dots & a_{i,p}=0 \\ \vdots & & \vdots & & \vdots \\ a_{n,1}=0 & \dots & a_{n,j}=0 & \dots & a_{n,p}=0 \end{pmatrix} = (\delta_{i,k}\delta_{j,l})_{k,l}.$$

Exemple 5 — *Dans $\mathscr{M}_2(\mathbb{K})$, on a* $E_{1,2} = \begin{pmatrix} 0 & 1 \\ 0 & 0 \end{pmatrix}$, $E_{2,2} = \begin{pmatrix} 0 & 0 \\ 0 & 1 \end{pmatrix}$, $E_{2,1} = \begin{pmatrix} 0 & 0 \\ 1 & 0 \end{pmatrix}$.

Définition 6 (Multiplication par un scalaire, Somme de matrices)
Soit $\mathbb{K}$ un corps. Soient $n, p \in \mathbb{N}^*$. On définit les opérations suivantes :

1. Le **produit d'un scalaire $\lambda \in \mathbb{K}$ et d'une matrice** $A = (a_{i,j})_{1 \leqslant i \leqslant n, 1 \leqslant j \leqslant p} \in \mathscr{M}_{n,p}(\mathbb{K})$ est la matrice notée $\lambda \cdot A$ ou λA obtenue en multipliant tous les coefficients par λ :

$$\lambda \cdot A = (\lambda a_{i,j})_{1 \leqslant i \leqslant n, 1 \leqslant j \leqslant p} \in \mathscr{M}_{n,p}(\mathbb{K}).$$

2. La **somme de deux matrices** $A = (a_{i,j}) \in \mathscr{M}_{n,p}(\mathbb{K})$ et $B = (b_{i,j}) \in \mathscr{M}_{n,p}(\mathbb{K})$ est la matrice

$$A + B = (a_{i,j} + b_{i,j}) \in \mathscr{M}_{n,p}(\mathbb{K}).$$

Exemple 7 — $2.\begin{pmatrix} 1 & 0 \\ 5 & 1 \end{pmatrix} = \begin{pmatrix} 5 & 0 \\ 10 & 5 \end{pmatrix}$.
$\begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix} + \begin{pmatrix} 1 & 1 \\ 0 & -1 \end{pmatrix} = \begin{pmatrix} 2 & 1 \\ 0 & 0 \end{pmatrix}$
$\begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix} + \begin{pmatrix} 1 & 0 & 2 \\ 1 & 1 & 3 \end{pmatrix}$ *n'a pas de sens.*

Proposition 8
L'ensemble $(\mathscr{M}_{n,p}(\mathbb{K}), +, .)$ des matrices $n \times p$, muni des opérations d'addition et de mutliplication par un scalaire, est un $\mathbb{K}-$espace vectoriel.
Il est de dimension $n \cdot p$. La famille $(E_{i,j})_{(i,j) \in [\![1,n]\!] \times [\![1,p]\!]}$ est une base de $\mathscr{M}_{n,p}(\mathbb{K})$.
On l'appelle **la base canonique de $\mathscr{M}_{n,p}(\mathbb{K})$** .

Définition 9
Soient $\mathbb{K}$ un corps et $n \in \mathbb{N}^*$. On dit qu'une matrice carrée $A \in \mathscr{M}_n(\mathbb{K})$ est

- **triangulaire supérieure** si $\forall (i,j)$ avec $i > j$, on a $a_{i,j} = 0$. C'est-à-dire si A est de la forme :

$$\begin{pmatrix} * & \cdots & & * \\ 0 & & & \\ \vdots & \ddots & \ddots & \vdots \\ 0 & \cdots & 0 & * \end{pmatrix}$$

  où chaque $*$ est un scalaire de $\mathbb{K}$ quelconque.
  On note $\mathscr{T}_n(\mathbb{K})$ l'ensemble des matrices $n \times n$ triangulaires supérieures.
  (De même, A est une matrice **triangulaire inférieure** si $\forall i < j$ on a $a_{i,j} = 0$.)

- **diagonale** si $\forall (i,j)$ avec $i \neq j$ on a $a_{i,j} = 0$, c'est-à-dire si A est de la forme :

$$\begin{pmatrix} \lambda_1 & 0 & \cdots & 0 \\ 0 & \lambda_2 & \ddots & \vdots \\ \vdots & \ddots & \ddots & 0 \\ 0 & \cdots & 0 & \lambda_n \end{pmatrix}.$$

  Cette matrice se note $\mathrm{Diag}(\lambda_1, \lambda_2, \ldots, \lambda_n)$.
  On note $\mathscr{D}_n(\mathbb{K})$ l'ensemble des matrices diagonales de taille $n \times n$.

Exemple 10 — $\begin{pmatrix} 0 & 1 & 5 \\ 0 & 2 & 1 \\ 0 & 0 & 1 \end{pmatrix}$, $\begin{pmatrix} 0 & 0 & 5 \\ 0 & 0 & 1 \\ 0 & 0 & 0 \end{pmatrix}$ *sont des matrices triangulaires supérieures.*
$\begin{pmatrix} 1 & 0 \\ 0 & 2 \end{pmatrix}$ *est une matrice diagonale.*
La matrice nulle $0_{\mathscr{M}_n(\mathbb{K})}$ *est une matrice diagonale.*
Les matrices diagonales sont des matrices triangulaires supérieures et des matrices triangulaires inférieures.

Remarque 11 — *L'ensemble* $\mathscr{T}_n(\mathbb{K})$ *est un sous-espace vectoriel de* $\mathscr{M}_n(\mathbb{K})$*. La famille* $(E_{i,j})_{1 \leqslant i \leqslant j \leqslant n}$ *est une famille génératrice de cet espace vectoriel. Ainsi, on a* $\dim \mathscr{T}_n(\mathbb{K}) = \dfrac{n(n+1)}{2}$.
L'ensemble $\mathscr{D}_n(\mathbb{K})$ *est un sous-espace vectoriel de* $\mathscr{M}_n(\mathbb{K})$*. La famille* $(E_{i,i})_{1 \leqslant i \leqslant n}$ *est une famille génératrice de cet espace vectoriel. Ainsi, on a* $\dim \mathscr{D}_n(\mathbb{K}) = n$.

Produit de matrices, propriétés

Définition 12
Soient $\mathbb{K}$ un corps et $p, q, r \in \mathbb{N}^*$. On définit le **produit** de deux matrices

$$A = (a_{i,j})_{(i,j) \in [\![1,p]\!] \times [\![1,q]\!]} \in \mathscr{M}_{p,q}(\mathbb{K}) \quad \text{et} \quad B = (b_{j,k})_{(j,k) \in [\![1,q]\!] \times [\![1,r]\!]} \in \mathscr{M}_{q,r}(\mathbb{K}),$$

noté $A \times B$ ou AB, comme la matrice

$$C = (c_{i,k})_{(i,k) \in [\![1,p]\!] \times [\![1,r]\!]} \in \mathscr{M}_{p,r}(\mathbb{K}) \text{ avec } c_{i,k} = \sum_{j=1}^{q} a_{i,j}\, b_{j,k}.$$

Remarque 13 —

1. *Le produit* AB *n'a de sens que si le nombre de colonnes de la matrice* A *soit égal au nombre de lignes de la matrice* B*.*
2. *Pour* $n \in \mathbb{N}^*$*, si* A *et* B *appartiennent à* $\mathscr{M}_n(\mathbb{K})$*, alors le produit* $A \times B$ *est bien défini et est aussi un élément de* $\mathscr{M}_n(\mathbb{K})$*.*
3. *Dans le calcul de* $c_{i,k}$ *interviennent les coefficients de la* $i^{\text{ème}}$ *ligne de* B *et les coefficients de la* $k^{\text{ème}}$ *colonne de* A *:*

$$\begin{pmatrix} a_{1,1} & \cdots & a_{1,j} & \cdots & a_{1,q} \\ \vdots & & & & \vdots \\ \hline a_{i,1} & \cdots & a_{i,j} & \cdots & a_{i,q} \\ \hline \vdots & & & & \vdots \\ a_{p,1} & \cdots & a_{p,j} & \cdots & a_{p,q} \end{pmatrix} \quad \to \quad \begin{pmatrix} b_{1,1} & \cdots & b_{1,k} & \cdots & b_{1,r} \\ \vdots & & \vdots & & \vdots \\ b_{j,1} & & b_{j,k} & & b_{j,r} \\ \vdots & & \vdots & & \vdots \\ b_{q,1} & \cdots & b_{q,k} & \cdots & b_{q,r} \end{pmatrix}$$

$$\downarrow$$

$$\begin{pmatrix} c_{1,1} & \cdots & & \cdots & c_{1,r} \\ \vdots & & & & \vdots \\ & & \boxed{c_{i,k}} & & \\ \vdots & & & & \vdots \\ c_{p,1} & \cdots & & \cdots & c_{p,r} \end{pmatrix}$$

Exemples 14

1. $\begin{pmatrix} 2 & -1 & 3 \\ -2 & 2 & -1 \end{pmatrix} \begin{pmatrix} -1 & 2 & 1 \\ 4 & -2 & 3 \\ -2 & 1 & -1 \end{pmatrix} = \begin{pmatrix} -12 & 9 & -4 \\ 12 & -9 & 5 \end{pmatrix}$.
2. $\begin{pmatrix} -1 & 2 & 1 \\ 4 & -2 & 3 \\ -2 & 1 & -1 \end{pmatrix} \begin{pmatrix} 2 & -1 & 3 \\ -2 & 2 & -1 \end{pmatrix}$ *n'a pas de sens.*
3. *Le produit d'une matrice carrée et d'une matrice colonne est une matrice colonne. Par exemple :*

$$\begin{pmatrix} 1 & 2 & 3 \\ 0 & 0 & 0 \\ 1 & 0 & 0 \end{pmatrix} \begin{pmatrix} 1 \\ 2 \\ 3 \end{pmatrix} = \begin{pmatrix} 14 \\ 0 \\ 1 \end{pmatrix} \quad \textit{et} \quad \begin{pmatrix} 1 & 2 & 3 \\ 0 & 0 & 0 \\ 1 & 0 & 0 \end{pmatrix} \begin{pmatrix} 0 \\ -3 \\ 2 \end{pmatrix} = \begin{pmatrix} 0 \\ 0 \\ 0 \end{pmatrix}.$$

Cet exemple permet de remarquer que le produit de deux matrices non-nulles peut être une matrice nulle. Ainsi :

$$A\,X = 0 \implies A = 0 \quad \textit{ou} \quad X = 0.$$

Proposition 15
Soient $\mathbb{K}$ un corps et $p, q \in \mathbb{N}^*$. Soient A et B dans $\mathscr{M}_{p,q}(\mathbb{K})$. On a :

(i) Si $AX = BX$ pour toute matrice colonne $X \in \mathscr{M}_{q,1}(\mathbb{K})$, alors $A = B$;

(ii) En particulier, si l'on a $AX = 0$ pour tout $X \in \mathscr{M}_{q,1}(\mathbb{K})$, alors la matrice A est la matrice nulle.

Preuve —

(i) Soit X_j la matrice colonne dont tous les coefficients sont nuls sauf le j-ième qui vaut 1. Le produit $A\,X_j$ est alors la j-ième colonne de la matrice A. De nême, $B\,X_j$ est la j-ième colonne de B.
Comme pour chaque $j \in [\![1, q]\!]$ on a $A\,X_j = B\,X_j$, les matrices A et B ont ainsi les mêmes colonnes. Donc $A = B$.

(ii) On est dans le cas particulier où la matrice B est nulle. Le point (i) donne alors $A = B = 0$.

□

Remarque 16 —

1. *Le produit d'une matrice ligne et d'une matrice colonne de même longueur est une matrice* 1×1 *qu'on identifie à un scalaire. Par exemple :*

$$\begin{pmatrix} 1 & 2 & 3 \end{pmatrix} \begin{pmatrix} 1 \\ 2 \\ 3 \end{pmatrix} = (14) = 14.$$

2. *Le produit d'une matrice colonne et d'une matrice ligne de même longueur est une matrice carrée. Par exemple :*

$$\begin{pmatrix} 1 \\ 2 \\ 3 \end{pmatrix} \begin{pmatrix} 1 & 2 & 3 \end{pmatrix} = \begin{pmatrix} 1 & 2 & 3 \\ 2 & 4 & 6 \\ 3 & 6 & 9 \end{pmatrix}.$$

Proposition 17
Soient $\mathbb{K}$ un corps, $n \geq 1$, et $1 \leq i, j \leq n$. On a :

$$E_{i,j}\,E_{k,l} = \delta_{j,k}\,E_{i,l}.$$

Les matrices $E_{i,j}$ formant la base canonique de l'espace vectoriel $\mathscr{M}_n(\mathbb{K})$, connaître le produit de deux de ces matrices est parfois très utile.

Proposition 18 (Propriétés du produit matriciel)
Soient $\mathbb{K}$ un corps et $p, q, r, s \in \mathbb{N}^*$. Soient $A, A' \in \mathscr{M}_{p,q}(\mathbb{K})$, $B, B' \in \mathscr{M}_{q,r}(\mathbb{K})$, $C \in \mathscr{M}_{r,s}(\mathbb{K})$, et $\lambda \in \mathbb{K}$. On a :

1. $A\,(\lambda\,B) = \lambda\,(A\,B)$.
   Le produit matriciel et la multiplication par un scalaire commutent.
2. $A\,(B + B') = (A\,B) + (A\,B') \qquad$ et $\qquad (A + A')\,B = (A\,B) + (A'\,B)$.
   Le produit matriciel est distributif à gauche et à droite par rapport à l'addition de matrices.
3. $A\,(B\,C) = (A\,B)\,C$.
   On dit que le produit matriciel a la propriété d'**associativité** . Le résultat d'une chaîne de produits matriciels ne dépend pas de l'ordre dans lequel on effectue les produits.

L'anneau $\mathscr{M}_n(\mathbb{K})$

Définition 19
Soient $\mathbb{K}$ un corps et $n \in \mathbb{N}^*$. On définit la **la matrice identité** $n \times n$, notée I_n, comme la matrice :

$$I_n = \text{Diag}\,(1, 1, \ldots, 1) = \begin{pmatrix} 1 & 0 & \cdots & 0 \\ 0 & 1 & \ddots & \vdots \\ \vdots & \ddots & \ddots & 0 \\ 0 & \cdots & 0 & 1 \end{pmatrix}.$$

Définition 20
Soient $\mathbb{K}$ un corps et $n \in \mathbb{N}^*$. Soit $A \in \mathscr{M}_n(\mathbb{K})$.
Pour $k \in \mathbb{N}$ on définit la **puissance k-ième** de A, notée A^k, par :

$$\begin{array}{rcl} A^0 & = & I_n \\ A^k & = & A \times A \times \ldots \times A \text{ (k fois), si } k > 0 \end{array}$$

Cette définition a bien un sens car il a été montré à la Prop. 18 que le produit matriciel est associatif $(A(BC) = (AB)C =_{def} ABC)$.

Proposition 21
Soient $\mathbb{K}$ un corps et $n \in \mathbb{N}^*$. L'ensemble $(\mathscr{M}_n(\mathbb{K}), +, \times)$ des matrices carrées $n \times n$ muni de l'addition de matrices et de la multiplication matricielle est un anneau.
De plus, cet anneau n'est pas intègre :

$$AB = 0 \not\Longrightarrow \quad A = 0 \text{ ou } B = 0,$$

C'est-à-dire qu'il existe des matrices A, B non-nulles telles que $AB = 0$.
Et cet anneau n'est pas commutatif :

$$AB \text{ n'est pas toujours égal à } BA.$$

C'est-à-dire qu'il existe des matrices A, B telles que $AB \neq BA$.

Preuve —Les notions de groupe et d'anneau seront détaillées dans le cours Algèbre 2. Commençons par montrer que $(\mathscr{M}_n(\mathbb{K}), +\times)$ est un anneau :

- L'ensemble $(\mathscr{M}_n(\mathbb{K}), +)$ est un groupe commutatif dont l'élément neutre est la matrice nulle (voir cours Algèbre 2) ;
- D'après la proposition 18, la multiplication matricielle $\times$ est une loi de composition interne associative ;
- D'après la proposition 18, la multiplication matricielle $\times$ est distributive à gauche et à droite par rapport à la loi d'addition matricielle $+$. La matrice identité I_n est un élément neutre pour $\times$ car $I_n A = AI_n = A$ pour tout $A \in \mathscr{M}_n(\mathbb{K})$ (voir **??**).

Cet anneau n'est pas intègre car $E_{1,2}E_{1,1} = 0$ (voir **??**), et n'est pas commutatif car

$$E_{1,2}E_{1,1} = 0 \neq E_{1,2} = E_{1,1}E_{1,2}.$$

□

Remarque 22 (Difficultés dans les anneaux non commutatifs) — *Pour A et B deux matrices de $\mathscr{M}_n(\mathbb{K})$, on ne peut en général pas appliquer les formules du binôme pour développer $(A+B)^2$, $(A+B)^m$ ou pour factoriser $A^2 - B^2$,$A^m - B^m$.*
On a par exemple $(A+B)^2 = (A+B)(A+B) = A^2 + AB + BA + B^2$, mais on ne peut pas simplifier plus cette expression car A et B ne commutent pas forcément (on ne sait rien entre AB et BA).
Lorsque les matrices A et B commutent on peut appliquer les formules de développement ou de factorisation, ce qui en fait un cas très particulier.

Exemple 23 (Difficultés dans les anneaux non intègres) —
Prenons $a \in [0, 1]$ et $b = \sqrt{1-a^2}$. Posons $A = \begin{pmatrix} a & b \\ b & -a \end{pmatrix}$. On a alors :

$$A^2 = \begin{pmatrix} a & b \\ b & -a \end{pmatrix}\begin{pmatrix} a & b \\ b & -a \end{pmatrix} = \begin{pmatrix} a^2+b^2 & ab-ba \\ ab-ba & b^2+a^2 \end{pmatrix} = \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix} = I_2$$

Ainsi, dans $\mathscr{M}_2(\mathbb{K})$ (pour $\mathbb{K} = \mathbb{Q}$,$\mathbb{R}$ ou $\mathbb{C}$), il existe une infinité de matrices A telles que $A^2 = I_2$.
Cela bien que l'équation $x^2 = 1$ ne possède que deux solutions dans $\mathbb{Q}$,$\mathbb{R}$ ou $\mathbb{C}$.
Cela est lié au fait que l'anneau $\mathscr{M}_2(\mathbb{K})$ n'est pas intègre. En effet on a :

$$\begin{aligned} A^2 = I_2 &\iff A^2 - I_2 = 0 \iff A^2 - I_2^2 = 0 \\ &\iff (A - I_2)(A + I_2) = 0, \text{ car } A \text{ et } I_2 \text{ commutent,} \end{aligned}$$

mais on ne peut pas avancer plus loin car les résultats que l'on voudrait utiliser ne sont pas vrais en général dans $\mathscr{M}_2(\mathbb{K})$.

9.2 Matrices inversibles, groupe linéaire

Définition 24
Soient $\mathbb{K}$ un corps et $n \in \mathbb{N}^*$. Une matrice $A \in \mathscr{M}_n(\mathbb{K})$ est **inversible** s'il existe une matrice B vérifiant :

$$A\,B = B\,A = I_n.$$

Cela revient à dire que A possède un inverse pour la loi de multiplication matricielle sur $\mathscr{M}_n(\mathbb{K})$.
L'ensemble des matrices inversibles de $\mathscr{M}_n(\mathbb{K})$ se note $\mathrm{GL}_n(\mathbb{K})$. On l'appelle le **groupe linéaire de $\mathscr{M}_n(\mathbb{K})$** .

Exemple 25 — *Pour $\lambda_1, \cdots, \lambda_n \in \mathbb{K}$ non-nuls, la matrice diagonale $A = \mathrm{Diag}\,(\lambda_1, \lambda_2, \cdots, \lambda_n)$ est inversible.*
En effet, pour $B = \mathrm{Diag}\left(\frac{1}{\lambda_1}, \frac{1}{\lambda_2}, \cdots, \frac{1}{\lambda_n}\right)$, on a

$$A\,B = B\,A = \mathrm{Diag}\,(1, 1, \cdots, 1) = I_n$$

En particulier, la matrice identité I_n est elle-même inversible.
Par contre, la matrice nulle 0 n'est pas inversible car pour tout matrice B on a $0.B = 0 \neq I_n$.

Proposition 26
Soient $\mathbb{K}$ un corps et $n \in \mathbb{N}^*$. Soient $A, B \in \mathscr{M}_n(\mathbb{K})$ inversibles. Alors :

- **L' inverse** de A est unique. On le note A^{-1} ;
- La matrice AB est inversible, et
$$(AB)^{-1} = B^{-1}A^{-1} \text{ ;}$$
(Le passage à l'inverse renverse le produit matriciel.)
- La matrice A^{-1} est inversible, et $(A^{-1})^{-1} = A$;

L'ensemble $(\mathrm{GL}_n(\mathbb{K}), \times)$ muni de la loi de multiplication matricielle est un groupe.

Preuve —

- Soient $C, D \in \mathscr{M}_n(\mathbb{K})$ tels que $AC = CA = I_n = AD = DA$.
On a alors $CAD = C(AD) = C\,I_n = C$ et $CAD = (CA)D = I_n\,D = D$, donc $C = D$.
- On a :
$$(B^{-1}A^{-1})(AB) = B^{-1}(A^{-1}A)B = B^{-1}I_nB = B^{-1}B = I_n$$
$$(AB)(B^{-1}A^{-1}) = A(BB^{-1})A^{-1} = AI_nA^{-1} = AA^{-1} = I_n.$$
Donc $(AB)^{-1} = B^{-1}A^{-1}$.
- On a : $AA^{-1} = A^{-1}A = I_n$, donc A^{-1} est inversible d'inverse A.

Les propriétés qui ont été montrées assurent que $(\mathrm{GL}_n(\mathbb{K}), \circ)$ est un groupe. □

Proposition 27
Soient $\mathbb{K}$ un corps et $n \in \mathbb{N}^*$. Soit $A \in \mathscr{M}_n(\mathbb{K})$.
La matrice A est inversible si et seulement s'il existe $B \in \mathscr{M}_n(\mathbb{K})$ telle que $AB = I_n$ ou $BA = I_n$.
Il suffit d'avoir une seule de ces relations pour montrer que A est inversible.

Proposition 28
Soient $\mathbb{K}$ un corps et $n \in \mathbb{N}^*$. Soit $A \in \mathscr{M}_n(\mathbb{K})$. S'il existe $B \in \mathscr{M}_n(\mathbb{K})$ non-nulle telle que $AB = 0$, alors la matrice A n'est pas inversible.

Preuve — Si A était inversible, on aurait

$$A^{-1}(AB) = A^{-1}.0 = 0 \text{ et } A^{-1}(AB) = (A^{-1}A)B = I_n.B = B,$$

donc $B = 0$, ce qui est impossible car B est non-nulle. Donc A n'est pas inversible. □

Corollaire 29
Soient $\mathbb{K}$ un corps et $n \in \mathbb{N}^*$. Soient $\lambda_1, \cdots, \lambda_n \in \mathbb{K}$. La matrice diagonale $A = \text{Diag}(\lambda_1, \cdots, \lambda_n)$ est inversible si et seulement si $\lambda_i \neq 0$, $\forall 1 \leq i \leq n$.

Preuve — Si tous les λ_i sont non-nuls, on a montré que A est inversible d'inverse $A^{-1} = \text{Diag}\left(\frac{1}{\lambda_1}, \frac{1}{\lambda_2}, \cdots, \frac{1}{\lambda_n}\right)$.
Supposons qu'il existe un indice j tel que $\lambda_j = 0$. Posons $B = \text{Diag}(\gamma_1 \cdots, \gamma_n)$ avec $\gamma_i = 0$ si $i \neq j$ et $\gamma_j = 1$. On a alors

$$A\,B = \text{Diag}(\lambda_1\gamma_1, \cdots, \lambda_n\gamma_n) = \text{Diag}(0, 0, \cdots, 0) = 0,$$

donc la matrice A n'est pas inversible d'après la proposition précédente. □

Proposition 30
Soient $\mathbb{K}$ un corps et $n \in \mathbb{N}^*$. Soit $A \in \mathscr{M}_n(\mathbb{K})$ une matrice dont une ligne est nulle ou une colonne est nulle. Alors A n'est pas inversible.

Matrices de taille 2×2 *inversibles*

Proposition 31
Soit $A = \begin{pmatrix} a & b \\ c & d \end{pmatrix} \in \mathscr{M}_2(\mathbb{K})$.
La matrice A est inversible si et seulement si $ad - bc \neq 0$.
Si $ad - bc \neq 0$, on a :

$$A^{-1} = \frac{1}{ad - bc} \begin{pmatrix} d & -b \\ -c & a \end{pmatrix}$$

9.3 Système linéaire, matrice d'un système linéaire

Système linéaire

Soient n et p deux entiers naturels non nuls et $\mathbb{K}$ un corps. Un **système linéaire** de n équations à p inconnues s'écrit

$$(\mathscr{S}) \; : \; \left\{ \begin{array}{ccccccccccccc} a_{1,1}\,x_1 & + & a_{1,2}\,x_2 & + & \ldots & + & a_{1,j}\,x_j & + & \ldots & + & a_{1,p}\,x_p & = & b_1 \\ \vdots & & \vdots & + & \vdots & & \vdots & & \vdots & & \vdots & & \vdots \\ a_{i,1}\,x_1 & + & a_{i,2}\,x_2 & + & \ldots & + & a_{i,j}\,x_j & + & \ldots & + & a_{i,p}\,x_p & = & b_i \\ \vdots & & \vdots & + & \vdots & & \vdots & & \vdots & & \vdots & & \vdots \\ a_{n,1}\,x_1 & + & a_{n,2}\,x_2 & + & \ldots & + & a_{n,j}\,x_j & + & \ldots & + & a_{n,p}\,x_p & = & b_n \end{array} \right.$$

- Les p inconnues sont x_1, x_2,..., x_p. On appelle **une solution du système** toute p-liste $(x_1, x_2, \ldots, x_p) \in \mathbb{K}^p$ vérifiant les n équations de $(\mathscr{S})$.
- Les np scalaires $a_{i,j} \in \mathbb{K}$ sont les **coefficients** du système.
- La $n-$liste $(b_1, b_2, \ldots, b_n) \in \mathbb{K}^n$ est le **second membre** du système.
- Si $b_1 = b_2 = \cdots = b_n = 0$, alors on dit que le système est **homogène** , ou que le système est **sans second membre**.

La matrice $A = (a_{i,j})_{\substack{1 \leqslant i \leqslant n \\ 1 \leqslant j \leqslant p}}$ s'appelle la **matrice du système linéaire** . Elle contient les coefficients du système linéaire. Comme le système linéaire s'écrit aussi :

$$(\mathscr{S}) \; : \; \sum_{j=1}^{p} a_{i,j}\,x_j = b_i \,, \forall i \in [\![1, n]\!]$$

on peut utiliser la matrice A pour réécrire le système linéaire sous la forme :

$$(\mathscr{S}) \; : \; A\,X = B$$

où

$$X = \begin{pmatrix} x_1 \\ x_2 \\ \vdots \\ x_p \end{pmatrix} \quad \text{et} \quad B = \begin{pmatrix} b_1 \\ b_2 \\ \vdots \\ b_n \end{pmatrix}$$

sont les matrices colonnes des inconnues et du second membre.

Nous allons présenter une méthode pour résoudre n'importe quel système linéaire $AX = Y$.
Nous commencerons par résoudre des systèmes linéaires simples (les systèmes échelonnés), puis nous verrons une méthode pour se ramener à un système linéaire échelonné (la méthode du Pivot).

Matrices échelonnées

Définition 32
Soient $\mathbb{K}$ un corps et $n, p \in \mathbb{N}^*$. Soit $A \in \mathscr{M}_{n,p}(\mathbb{K})$, avec $A = (a_{i,j})_{i,j}$.
Soit $1 \leq i \leq n$. Si la i-ème ligne de A est non-nulle, on définit $\alpha_i = \inf(\{1 \leq k \leq p \text{ tels que } a_{i,k} \neq 0\})$.

Pour L_i la i-ème ligne de A, si L_i est non-nulle on a : $L_i = (0, 0, \ldots, 0, a_{i,\alpha_i}, *, *)$.

Si la k-ème ligne de A est nulle, on pose alors $\alpha_i = p + i$.

On dit alors que la matrice A est **échelonnée** si l'on a :

$$\alpha_1 < \alpha_2 < \ldots < \alpha_n.$$

Exemple 33 — $A = \begin{pmatrix} 1 & 2 & 3 \\ 0 & 0 & 2 \\ 0 & 0 & 0 \end{pmatrix}$ *est échelonnée.* $B = \begin{pmatrix} 0 & 2 & 3 \\ 1 & 0 & 2 \\ 0 & 2 & 0 \end{pmatrix}$ *n'est pas échelonnée.*

$C = \begin{pmatrix} 1 & 2 & 3 & 1 \\ 0 & 2 & 2 & -1 \\ 0 & 2 & 0 & 0 \\ 0 & 0 & 1 & 1 \end{pmatrix}$ *n'est pas échelonnée.* $D = \begin{pmatrix} 0 & 0 & 3 & 1 \\ 0 & 0 & 0 & -1 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \end{pmatrix}$*, on a* $\alpha_1 = 3$*,* $\alpha_2 = 4$*,* $\alpha_3 = 7$*,* $\alpha_4 = 8$ *est échelonnée.*

Pour $E = 0_{\mathscr{M}_{n,p}(\mathbb{K})}$ *on a* $\alpha_1 = p + 1$,$\alpha_2 = p + 2$,$\ldots$, $\alpha_n = p + n$*. La matrice nulle est échelonnée.*

Pour $F \in \mathscr{M}_n(\mathbb{K})$ *une matrice diagonale dont les coefficients diagonaux sont non-nuls,* F *est échelonnée.*

Remarque 34 — *On dit que le système linéaire* $(\mathcal{S})$ *:* $AX = Y$*, où* $X, Y \in \mathscr{M}_{p,1}(\mathbb{K})$ *et* $A \in \mathscr{M}_{n,p}(\mathbb{K})$ *est échelonné si la matrice* A *est échelonnée.*

Remarque 35 — *Soit* $A \in \mathscr{M}_n(\mathbb{K})$ *une matrice échelonnée. Comme on a*

$$1 \leq \alpha_1 < \alpha_2 < \ldots < \alpha_n,$$

on remarque en particulier que A *est une matrice triangulaire supérieure.*

Les matrices carrées échelonnées sont un cas particulier de matrices triangulaires supérieures.

Résolution d'un système échelonné

Proposition 36 (Résolution d'un système linéaire échelonné)
Soient $\mathbb{K}$ un corps, $n, p \in \mathbb{N}^*$. Soit $A \in \mathscr{M}_{n,p}(\mathbb{K})$ une matrice échelonnée. Soient $(y_1, \ldots, y_p) \in \mathbb{K}^p$.
Alors, on peut toujours résoudre le système $(\mathcal{S})$: $AX = Y$.

Remarque 37 — *On détermine les solutions* $(x_1, \ldots, x_p)$ *du système échelonné* $(\mathcal{S})$ *:* $AX = Y$ *en remontant ligne par ligne :*

Soit r *le numéro de la dernière ligne non-nulle de* A*.*
Comme A *est échelonnée, les lignes* $r + 1, \ldots, n$ *sont nulles, et les lignes* $1, \ldots, r$ *sont non-nulles.*

• *Toutes les lignes nulles de la matrice* A *donnent des équations de la forme :* $0 = y_j$*, pour* $r + 1 \leq j \leq n$*.*
- Si l'un de ces y_j *est non-nul, alors le système* $(\mathcal{S})$ *n'a pas de solutions, et la résolution est terminée.*
- Si tous les y_j *sont nuls pour* $r + 1 \leq j \leq n$*, alors ces équations sont de la forme :* $0 = 0$*.*
On retire ces équations du système (équations toujours vraies), et on continue la résolution.

• *Pour tout* $1 \leq i \leq r$*, on a* $a_{i,\alpha_i} \neq 0$*.*
Pour chaque ligne L_i*, on exprime* x_{α_i} *en fonction de* b_i*, et* $x_{\alpha_i+1}, \ldots, x_p$*.*

• *Ligne* r *: Le coefficient* x_{α_r} *est alors déterminé.*

• *Ligne $r-1$: On remplace x_{α_r} par l'expression à la ligne r. Le coefficient $x_{\alpha_{r-1}}$ est alors déterminé.*

• *Ligne $r-2$: On remplace $x_{\alpha_r}, x_{\alpha_{r-1}}$ par leurs expressions. Le coefficient $x_{\alpha_{r-2}}$ est alors déterminé.*

• ...

• *Ligne 1 : On remplace $x_{\alpha_r}, \dots, x_{\alpha_2}$ par leurs expressions. Le coefficient x_{α_1} est alors déterminé.*

On obtient alors les valeurs de $x_{\alpha_1}, \dots, x_{\alpha_r}$ en fonction de $y_1, \dots, y_r$ ainsi que des x_j pour $j \neq \alpha_1, \dots, \alpha_r$.

EXEMPLE 38 — *Résolution dans $\mathbb{R}$ du système linéaire :*

$$(\mathcal{S}) : \left\{ \begin{array}{rcrcrcl} x_1 & + & x_2 & + & 2x_3 & = & 3 \\ 0 & + & 2x_2 & + & 2x_3 & = & 4 \\ 0 & + & 0 & - & x_3 & = & 3 \\ 0 & + & 0 & + & 0 & = & 0 \end{array} \right.$$

La matrice A associée à ce système linéaire est échelonnée. Elle possède une ligne nulle.

On a ainsi :

$$(\mathcal{S}) \Longleftrightarrow \left\{ \begin{array}{rl} x_1 & = 3 - x_2 - 2x_3 \\ x_2 & = 2 - x_3 \\ x_3 & = -3 \\ 0 & = 0 \end{array} \right. \Longleftrightarrow \left\{ \begin{array}{rl} x_1 & = 3 - 5 - 2(-3) = 4 \\ x_2 & = 5 \\ x_3 & = -3 \end{array} \right.$$

L'ensemble des solutions de $(\mathcal{S})$ est donc $\{(4, 5, 3)\}$. Ce système linéaire possède une unique solution.

EXEMPLE 39 — *Résolution dans $\mathbb{R}$ du système linéaire d'équations :*

$$(\mathcal{S}) : \left\{ \begin{array}{rcrcrcrcl} x_1 & + & 3x_2 & - & 2x_3 & + & 5x_4 & = & -1 \\ 0 & + & 2x_2 & + & 2x_3 & - & 2x_4 & = & 4 \end{array} \right.$$

On remarque que la matrice A associée à ce système linéaire est échelonnée. On a :

$$(\mathcal{S}) \Leftrightarrow \left\{ \begin{array}{rl} x_1 & = -1 - 3x_2 + 2x_3 - 5x_4 \\ x_2 & = 2 - x_3 + x_4 \end{array} \right. \Leftrightarrow \left\{ \begin{array}{rl} x_1 & = -1 - 3(2 - x_3 + x_4) + 2x_3 - 5x_4 = -7 + 5x_3 - 8x_4 \\ x_2 & = 2 - x_3 + x_4 \end{array} \right.$$

L'ensemble des solutions de $(\mathcal{S})$ est donc :

$$\begin{aligned} S &= \{(-7 + 5x_3 - 8x_4, 2 - x_3 + x_4, x_3, x_4),\ x_3, x_4 \in \mathbb{R}\} \\ &= \{(-7, 2, 0, 0) + (5x_3, -x_3, x_3, 0) + (-8x_4, x_4, 0, x_4),\ x_3, x_4 \in \mathbb{R}\} \\ &= (-7, 2, 0, 0) + Vect((5, -1, 1, 0), (-8, 1, 0, 1)). \end{aligned}$$

Ce système linéaire possède une infinité de solutions.

EXEMPLE 40 — *Résoudre dans, selon $(y_1, y_2, y_3, y_4) \in \mathbb{R}^4$, le système linéaire d'équations :*

$$(\mathcal{S}) : \left\{ \begin{array}{rcrcrcrcl} x_1 & + & 2x_2 & + & x_3 & - & x_4 & = & y_1 \\ 0 & - & x_2 & + & x_3 & + & 0 & = & y_2 \\ 0 & + & 0 & + & 2x_3 & + & 6x_4 & = & y_3 \\ 0 & + & 0 & + & 0 & + & 0 & = & y_4 \end{array} \right.$$

On remarque que la matrice A associée à ce système linéaire est échelonnée.
• *La quatrième ligne du système est $0 = y_4$.*
- *Si $y_4 \neq 0$, ce système n'a pas de solutions ($S = \emptyset$).*

- *Si $y_4 = 0$, on poursuit la résolution. On a :*

$$(\mathcal{S}) \Longleftrightarrow \left\{ \begin{array}{rl} x_1 & = y_1 - 2x_2 - x_3 + x_4 \\ x_2 & = -y_2 + x_3 \\ x_3 & = \frac{1}{2}y_3 - 3x_4 \\ 0 & = 0 \end{array} \right. \Longleftrightarrow \left\{ \begin{array}{rl} x_1 & = y_1 - 2x_2 - x_3 + x_4 \\ x_2 & = -y_2 + \frac{1}{2}y_3 - 3x_4 \\ x_3 & = \frac{1}{2}y_3 - 3x_4 \end{array} \right.$$

$$(\mathcal{S}) \iff \begin{cases} x_1 &= y_1 - 2(-y_2 + \frac{1}{2}y_3 - 3x_4) - (\frac{1}{2}y_3 - 3x_4) + x_4 = y_1 + 2y_2 - \frac{3}{2}y_3 + 10x_4 \\ x_2 &= -y_2 + \frac{1}{2}y_3 - 3x_4 \\ x_3 &= \frac{1}{2}y_3 - 3x_4 \end{cases}$$

L'ensemble des solutions de $(\mathcal{S})$ est donc :

$$\begin{aligned} S &= \{(y_1 + 2y_2 - \frac{3}{2}y_3 + 10x_4, -y_2 + \frac{1}{2}y_3 - 3x_4, \frac{1}{2}y_3 - 3x_4, x_4),\ x_4 \in \mathbb{R}\} \\ &= (y_1 + 2y_2 - \frac{3}{2}y_3, -y_2 + \frac{1}{2}y_3, \frac{1}{2}y_3, 0) + Vect((10, -3, -3, 1)). \end{aligned}$$

En conclusion, si $y_4 \neq 0$ ce système linéaire n'admet pas de solutions. Si $y_4 = 0$, alors ce système linéaire possède une infinité de solutions, décrites au-dessus.

9.4 Méthode du Pivot

La méthode du Pivot permet de transformer toute matrice en une matrice échelonnée, en la multipliant par des matrices inversibles.

Matrices élémentaires

Définition 41
Soient $\mathbb{K}$ un corps, $n \in \mathbb{N}^*$. Soient $1 \leq i, j \leq n$, $i \neq j$ et $\lambda \in \mathbb{K}$.
On définit les matrices suivantes, appelées **matrices élémentaires** :

- $E(i, j, \lambda) = I_n + \lambda.E_{i,j}$;
- $M(i, \lambda) = I_n - E_{i,i} + \lambda.E_{i,i}$;
- $S(i, j) = I_n - E_{i,i} - E_{j,j} + E_{i,j} + E_{j,i}$.

Exemple 42 — *Dans $\mathscr{M}_3(\mathbb{K})$, on a :*
$E(1, 3, \lambda) = \begin{pmatrix} 1 & 0 & \lambda \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{pmatrix}$, $M(2, \lambda) = \begin{pmatrix} 1 & 0 & 0 \\ 0 & \lambda & 0 \\ 0 & 0 & 1 \end{pmatrix}$, *et* $S(1, 3) = \begin{pmatrix} 0 & 0 & 1 \\ 0 & 1 & 0 \\ 1 & 0 & 0 \end{pmatrix}$.

La matrice $E(i, j, \lambda)$ est la matrice identité pour laquelle on a ajouté λ en (i, j).

La matrice $M(i, \lambda)$ est une matrice diagonale, qui vaut λ en (i, i), et 1 sur le reste la diagonale.

La matrice $S(i, j)$ est la matrice identité pour laquelle on a déplacé les coefficients (i, i) et (j, j) (en (i, j) et (j, i)).

Proposition 43
Soient $\mathbb{K}$ un corps, $n, p \in \mathbb{N}^*$. Soient $1 \leq i, j \leq n$, $i \neq j$ et $\lambda \in \mathbb{K}$. Soit $A \in \mathscr{M}_{n,p}(\mathbb{K})$. Alors :

- $E(i, j, \lambda)$ est inversible.
  L'opération $A \mapsto E(i, j, \lambda)A$ revient à effectuer : Ligne i devient (Ligne i+λ.Ligne j).
  On note $L_i \leftarrow L_i + \lambda.L_j$ cette opération.
- $M(i, \lambda)$ est inversible si $\lambda \neq 0$.
  L'opération $A \mapsto M(i, \lambda)A$ revient à effectuer : Ligne i devient λ.Ligne i.
  On note $L_i \leftarrow \lambda.L_i$ cette opération.
- $S(i, j)$ est inversible.
  L'opération $A \mapsto S(i, j)A$ revient à effectuer : Ligne i devient Ligne j, et Ligne j devient Ligne i.
  On note $L_i \leftrightarrow L_j$ cette opération.

Théorème 44
Soient $\mathbb{K}$ un corps, $n, p \in \mathbb{N}^*$. Soit $A \in \mathscr{M}_{n,p}(\mathbb{K})$.
Alors il existe $B \in \mathscr{M}_{n,p}(\mathbb{K})$ une matrice échelonnée, et $M_1, \ldots, M_r$ des matrices élémentaires de $\mathscr{M}_n(\mathbb{K})$ telles que :

$$M_1 \times \ldots \times M_r \times A = B.$$

Autrement dit, il est possible de transformer la matrice A en une matrice échelonnée B en un nombre fini d'opérations élémentaires.

Remarque 45 — *Soient $X, Y \in \mathscr{M}_{p,1}(\mathbb{K})$ des vecteurs colonne. Comme les matrices élémentaires sont inversibles, on a :*

$$AX = Y \iff (M_1 \dots M_r A)X = M_1 \dots M_r Y \iff BX = Y'.$$

En possédant une méthode pour trouver de telles matrices élémentaires $M_1, \dots, M_r$, on peut alors résoudre tout système linéaire $AX = Y$ en se ramenant à un système linéaire échelonné $BX = Y'$.

Méthode du Pivot

Remarque 46 (Méthode du Pivot) — *Soit $A \in \mathscr{M}_{n,p}(\mathbb{K})$.*
On échelonne la matrice A en utilisant les opérations élémentaires.
On procède colonne par colonne, de la gauche vers la droite.
On rappelle que pour $1 \leq i \leq n$, l'entier α_i désigne la position du coefficient non-nul de la ligne L_i le plus à gauche (et $\alpha_i = n + i$ si la ligne L_i est nulle).

- *Pour L_i et L_j deux lignes telles que $\alpha_i < \alpha_j$ mais $i > j$, l'opération $L_i \leftrightarrow L_j$ permute les lignes L_i et L_j :*

$$\begin{cases} L_j : (0,0,\dots,0,0,\dots,0,a_{i,\alpha_i},*,\dots,*) \\ L_i : (0,0,\dots,0,a_{i,\alpha_j},*,\dots,*,*,\dots,*) \end{cases} \xrightarrow[L_i \leftrightarrow L_j]{} \begin{cases} L_j : (0,0,\dots,0,a_{i,\alpha_j},*,\dots,*,*,\dots,*) \\ L_i : (0,0,\dots,0,0,\dots,0,a_{i,\alpha_i},*,\dots,*) \end{cases}$$

- *Pour a_{i,α_i} le coefficient non-nul le plus à gauche de L_i, l'opération $L_i \leftarrow \frac{1}{a_{i,\alpha_i}} L_i$ change ce coefficient en un 1 (utile pour certains calculs).*

$$L_i : (0,0,\dots,0,a_{i,\alpha_i},*,\dots,*) \xrightarrow[L_i \leftarrow \frac{1}{a_{i,\alpha_i}} L_i]{} L_i : (0,0,\dots,0,1,*,\dots,*).$$

- *Pour L_i et L_j deux lignes telles que $\alpha_i = \alpha_j$, avec $j < i$, l'opération $L_i \leftarrow L_i + \frac{-a_{i,\alpha_i}}{a_{j,\alpha_j}} L_j$ annule le coefficient a_{i,α_i} et préserve les 0 situés avant :*

$$\begin{cases} L_j : (0,0,\dots,0,a_{j,\alpha_j},*,\dots,*) \\ L_i : (0,0,\dots,0,a_{i,\alpha_j},*,\dots,*) \end{cases} \xrightarrow[L_i \leftarrow L_i + \frac{-a_{i,\alpha_i}}{a_{j,\alpha_j}} L_j]{} \begin{cases} L_j : (0,0,\dots,0,a_{j,\alpha_j},*,\dots,*) \\ L_i : (0,0,\dots,0,0,*,\dots,*) \end{cases}$$

Voyons cela sur des exemples (échelonnage de matrice, résolution d'un système linéaire).

Exemple 47 — *Appliquer la méthode du Pivot sur* $A = \begin{pmatrix} 1 & 2 & 0 & 3 \\ -1 & 1 & 1 & 0 \\ 2 & 1 & 7 & 1 \end{pmatrix}$.

On a : $\begin{pmatrix} 1 & 2 & 0 & 3 \\ -1 & 1 & 1 & 0 \\ 2 & 1 & 7 & 1 \end{pmatrix} \xrightarrow[L_3 \leftarrow L_3 - 2L_1]{L_2 \leftarrow L_2 + L_1} \begin{pmatrix} 1 & 2 & 0 & 3 \\ 0 & 3 & 1 & 3 \\ 0 & -3 & 7 & -5 \end{pmatrix} \xrightarrow[L_3 \leftarrow L_3 + 3L_2]{L_2 \leftarrow \frac{1}{3}L_2} \begin{pmatrix} 1 & 2 & 0 & 3 \\ 0 & 1 & \frac{1}{3} & 1 \\ 0 & 0 & 8 & -2 \end{pmatrix}$.

On a bien obtenu une matrice échelonnée.
Pour B cette matrice échelonnée, les opérations effectuées donnent :

$$E(3,2,3)M(2,\frac{1}{3})E(3,1,-2)E(1,2,1)A = B.$$

Exemple 48 — *Appliquer la méthode du Pivot sur* $A = \begin{pmatrix} 0 & 2 & 1 \\ 2 & 2 & 1 \\ 1 & 4 & 3 \end{pmatrix}$.

On a : $\begin{pmatrix} 0 & 2 & 1 \\ 2 & 2 & 1 \\ 1 & 4 & 3 \end{pmatrix} \xrightarrow[L_1 \leftrightarrow L_3]{} \begin{pmatrix} 1 & 4 & 3 \\ 2 & 2 & 1 \\ 0 & 2 & 1 \end{pmatrix} \xrightarrow[L_2 \leftarrow L_2 - 2L_1]{} \begin{pmatrix} 1 & 4 & 3 \\ 0 & -6 & -5 \\ 0 & 2 & 1 \end{pmatrix}$.

$$\xrightarrow[L_3 \leftarrow L_3 - 2L_2]{L_2 \leftarrow \frac{1}{-6}L_2} \begin{pmatrix} 1 & 4 & 3 \\ 0 & 1 & \frac{5}{6} \\ 0 & 0 & \frac{-2}{3} \end{pmatrix}.$$

On a bien obtenu une matrice échelonnée. Pour B cette matrice échelonnée, les opérations effectuées donnent : $E(3,2,-2)M(2,\frac{-1}{6})E(2,1,-2)S(1,3)A = B$.

Si l'on avait effectué $L_1 \leftrightarrow L_2$ dans l'exemple précédent, on aurait obtenu une matrice échelonnée différente. Cela ne dérange pas.

Exemple 49 (Résolution d'un système linéaire avec la méthode du Pivot) —
Résoudre le système linéaire : $(\mathcal{S}) : \left\{ \begin{array}{rcrcrcl} x_1 & + & 4x_2 & - & 5x_3 & = & 1 \\ 2x_1 & - & 2x_2 & + & x_3 & = & 0 \\ 3x_1 & - & x_2 & - & x_3 & = & 1 \end{array} \right.$.

On utilise la méthode du Pivot :

$$(\mathcal{S}) \begin{array}{c} \iff \\ L_2 \leftarrow L_2 - 2L_1 \\ L_3 \leftarrow L_3 - 3L_1 \end{array} \left\{ \begin{array}{rcrcrcl} x_1 & + & 4x_2 & - & 5x_3 & = & 1 \\ 0 & - & 10x_2 & + & 11x_3 & = & -2 \\ 0 & - & 9x_2 & + & 14x_3 & = & -2 \end{array} \right.$$

$$(\mathcal{S}) \begin{array}{c} \iff \\ L_2 \leftarrow \frac{1}{-10} L_2 \\ L_3 \leftarrow L_3 + 9L_2 \end{array} \left\{ \begin{array}{rcrcrcl} x_1 & + & 4x_2 & - & 5x_3 & = & 1 \\ 0 & + & x_2 & + & \frac{-11}{10}x_3 & = & \frac{1}{5} \\ 0 & + & 0 & + & (14 + \frac{-99}{10})x_3 & = & -2 + \frac{9}{5} \end{array} \right. \quad (\textit{système échelonné})$$

$$(\mathcal{S}) \iff \left\{ \begin{array}{rcrcrcl} x_1 & + & 4x_2 & - & 5x_3 & = & 1 \\ 0 & + & x_2 & + & \frac{-11}{10}x_3 & = & \frac{1}{5} \\ 0 & + & 0 & + & \frac{41}{10}x_3 & = & \frac{-1}{5} \end{array} \right. \iff \left\{ \begin{array}{rl} x_1 & = 1 - 4x_2 + 5x_3 \\ x_2 & = \frac{1}{5} + \frac{11}{10}x_3 \\ x_3 & = \frac{-10}{5.41} = \frac{-2}{41} \end{array} \right.$$

$$(\mathcal{S}) \iff \left\{ \begin{array}{rl} x_1 & = 1 - 4x_2 + 5x_3 \\ x_2 & = \frac{1}{5} + \frac{11}{10}\frac{-2}{41} = \frac{60}{410} = \frac{6}{41} \\ x_3 & = \frac{-2}{41} \end{array} \right. \iff \left\{ \begin{array}{rl} x_1 & = 1 - 4\frac{6}{41} + 5\frac{-2}{41} = \frac{7}{41} \\ x_2 & = \frac{6}{41} \\ x_3 & = \frac{-2}{41} \end{array} \right.$$

L'ensemble des solutions de $(\mathcal{S})$ *est* $\{(\frac{7}{41}, \frac{6}{41}, \frac{-2}{41})\}$.

Calcul de l'inverse d'une matrice avec la méthode du Pivot

Pour $A \in \mathscr{M}_n(\mathbb{K})$ une matrice carrée, la méthode du Pivot peut être utilisée pour déterminer si A est inversible ou non, et pour calculer A^{-1}.

Proposition 50
Soient $\mathbb{K}$ un corps, $n \in \mathbb{N}^*$. Soit $A \in \mathscr{M}_n(\mathbb{K})$.
Si, en appliquant la méthode du Pivot à A, on obtient à une étape une matrice dont une ligne est nulle, alors A n'est pas inversible.
Si, en appliquant la méthode du Pivot à A, on obtient une matrice échelonnée sans ligne nulle, alors A est inversible.

Exemple 51 — *La matrice* $A = \begin{pmatrix} 1 & 2 & 1 \\ -4 & 3 & 2 \\ -3 & 5 & 3 \end{pmatrix}$ *est-elle inversible ?*

On applique la méthode du Pivot à A *:*

$$\begin{pmatrix} 1 & 2 & 1 \\ -4 & 3 & 2 \\ -3 & 5 & 3 \end{pmatrix} \begin{array}{c} \longrightarrow \\ L_2 \leftarrow L_2 + 4L_1 \\ L_3 \leftarrow L_3 + 3L_1 \end{array} \begin{pmatrix} 1 & 2 & 1 \\ 0 & 11 & 6 \\ 0 & 11 & 6 \end{pmatrix} \begin{array}{c} \longrightarrow \\ L_3 \leftarrow L_3 - L_2 \end{array} \begin{pmatrix} 1 & 2 & 1 \\ 0 & 11 & 6 \\ 0 & 0 & 0 \end{pmatrix}$$

On a obtenu une matrice avec la ligne nulle pendant la méthode du Pivot. Donc A n'est pas inversible.

Remarque 52 — *Soit* $A \in \mathscr{M}_n(\mathbb{K})$. *S'il existe des matrices élémentaires* $M_1, \dots, M_r$ *telles que* $M_1 \dots M_r A = I_n$, *alors on a* $A = (M_1 \dots M_r)^{-1}$. *Donc A est inversible et* $M_1 \dots M_r = A^{-1}$.

Proposition 53
Soient $\mathbb{K}$ un corps, $n \in \mathbb{N}^*$. Soit $A \in \mathscr{M}_n(\mathbb{K})$.
On pose $B = (A \mid I_n) \in \mathscr{M}_{n,2n}(\mathbb{K})$, la matrice obtenue en "collant" les matrices A et I_n. C'est-à-dire :

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

Si, en appliquant la méthode du Pivot à B, on obtient une matrice de la forme $(I_n \mid M)$, alors on a $M = A^{-1}$.

EXEMPLE 54 — *Montrer que la matrice* $A = \begin{pmatrix} 1 & 2 & 1 \\ -4 & 3 & 2 \\ 1 & 5 & 3 \end{pmatrix}$ *est inversible, et calculer son inverse.*

On applique la méthode du Pivot à la matrice $B = (A \mid I_3)$ *:*

$$\begin{pmatrix} 1 & 2 & 1 & 1 & 0 & 0 \\ -4 & 3 & 2 & 0 & 1 & 0 \\ 1 & 5 & 3 & 0 & 0 & 1 \end{pmatrix} \begin{array}{c} \longrightarrow \\ L_2 \leftarrow L_2 + 4L_1 \\ L_3 \leftarrow L_3 - L_1 \end{array} \begin{pmatrix} 1 & 2 & 1 & 1 & 0 & 0 \\ 0 & 11 & 6 & 4 & 1 & 0 \\ 0 & 3 & 2 & -1 & 0 & 1 \end{pmatrix}$$

$$\begin{array}{c} \longrightarrow \\ L_2 \leftarrow \frac{1}{11} L_2 \\ L_3 \leftarrow L_3 - 3L_2 \end{array} \begin{pmatrix} 1 & 2 & 1 & 1 & 0 & 0 \\ 0 & 1 & \frac{6}{11} & \frac{4}{11} & \frac{1}{11} & 0 \\ 0 & 0 & \frac{4}{11} & \frac{-23}{11} & \frac{-3}{11} & 1 \end{pmatrix} \begin{array}{c} \longrightarrow \\ L_3 \leftarrow \frac{11}{4} L_3 \\ L_2 \leftarrow L_2 - \frac{6}{11} L_2 \\ L_1 \leftarrow L_1 - L_3 \end{array} \begin{pmatrix} 1 & 2 & 0 & \frac{27}{4} & \frac{3}{4} & \frac{-11}{4} \\ 0 & 1 & 0 & \frac{4}{11} + \frac{23.6}{4.11} & \frac{1}{11} + \frac{3.6}{4.11} & \frac{-6}{4} \\ 0 & 0 & 1 & \frac{-23}{4} & \frac{-3}{4} & \frac{11}{4} \end{pmatrix}$$

$$= \begin{pmatrix} 1 & 2 & 0 & \frac{27}{4} & \frac{3}{4} & \frac{-11}{4} \\ 0 & 1 & 0 & \frac{7}{2} & \frac{1}{2} & \frac{-3}{2} \\ 0 & 0 & 1 & \frac{-23}{4} & \frac{-3}{4} & \frac{11}{4} \end{pmatrix} \begin{array}{c} \longrightarrow \\ L_1 \leftarrow L_1 - 2L_2 \end{array} \begin{pmatrix} 1 & 0 & 0 & \frac{-1}{4} & \frac{-1}{4} & \frac{1}{4} \\ 0 & 1 & 0 & \frac{7}{2} & \frac{1}{2} & \frac{-3}{2} \\ 0 & 0 & 1 & \frac{-23}{4} & \frac{-3}{4} & \frac{11}{4} \end{pmatrix}$$

On a obtenu une matrice de la forme $(I_3 \mid M)$*. Donc* A *est inversible et* $A^{-1} = M$*, avec :*

$$A^{-1} = \frac{1}{4} \begin{pmatrix} -1 & -1 & 1 \\ 14 & 2 & -6 \\ -23 & -3 & 11 \end{pmatrix}.$$

COROLLAIRE 55
Soient $\mathbb{K}$ un corps, $n \in \mathbb{N}^*$.
Alors, le groupe linéaire $Gl_n(\mathbb{K})$ est engendré comme groupe par les matrices élémentaires (dilatations, transvections, réflexions).

9.5 TRANSPOSÉE D'UNE MATRICE

DÉFINITION 56
Soient $\mathbb{K}$ un corps et $p, q \in \mathbb{N}^*$. Soit $A \in \mathscr{M}_{p,q}(\mathbb{K})$.
On définit la **transposée** de A, notée ${}^tA$, par $b_{i,j} = a_{j,i}$, $\forall 1 \leq i \leq p, 1 \leq j \leq q$.

EXEMPLE 57 — *Pour* $A = \begin{pmatrix} 1 & 2 \\ 0 & -1 \end{pmatrix}$*, on a* ${}^tA = \begin{pmatrix} 1 & 0 \\ 2 & -1 \end{pmatrix}$.

Pour $B = \begin{pmatrix} a_1 & a_2 & \dots & a_p \\ 0 & 0 & \dots & 0 \\ 0 & 0 & \dots & 0 \end{pmatrix} \in \mathscr{M}_{3,p}(\mathbb{K})$*, on a* ${}^tB = \begin{pmatrix} a_1 & 0 & 0 \\ a_2 & 0 & 0 \\ \vdots & & \vdots \\ a_p & 0 & 0 \end{pmatrix} \in \mathscr{M}_{p,3}(\mathbb{K})$.

REMARQUES 58

1. *Transposer une matrice transforme ses lignes en colonnes (et ses colonnes en lignes).*
2. *La transposée d'une matrice ligne est une matrice colonne.*
   La transposée d'une matrice carrée est une matrice carrée.
   La transposée d'une matrice triangulaire supérieure est une matrice triangulaire inférieure.
3. *Soit* $A \in \mathscr{M}_n(\mathbb{K})$ *une matrice carrée. Alors les matrices carrées* A *et* ${}^tA$ *:*
   (a) *ont la même diagonale ;*
   (b) *sont les symétriques l'une de l'autre par rapport à la diagonale.*

PROPOSITION 59
Pour $A, B \in \mathscr{M}_{p,q}(\mathbb{K})$ et $\lambda \in \mathbb{K}$ un scalaire, on a :

$${}^t({}^tA) = A \qquad \text{et} \qquad {}^t(\lambda A) = \lambda {}^tA \qquad \text{et} \qquad {}^t(A + B) = {}^tA + {}^tB.$$

PROPOSITION 60
Soient $\mathbb{K}$ un corps et $p, q, r \in \mathbb{N}^*$. Soient $A \in \mathscr{M}_{p,q}(\mathbb{K})$ et $B \in \mathscr{M}_{q,r}(\mathbb{K})$. On a :

$$^t(A\,B) = {}^tB\,{}^tA.$$

La transposition renverse l'ordre du produit matriciel.

Preuve — Soient $A = (a_{i,j})_{(i,j)}$ et $B = (b_{j,k})_{(j,k)}$.

La matrice $C = A\,B$ possède p lignes et r colonnes, donc $C' = {}^t(A\,B)$ possède r lignes et p colonnes. On a :

$$c'_{k,i} = c_{i,k} = \sum_{j=1}^{q} a_{i,j}\,b_{j,k}.$$

Comme $A' = {}^tA \in \mathscr{M}_{q,p}(\mathbb{K})$ et $B' = {}^tB \in \mathscr{M}_{r,q}(\mathbb{K})$, le produit $D = {}^tB\,{}^tA$ existe et possède lui aussi r lignes et p colonnes. La relation :

$$d_{k,i} = \sum_{j=1}^{q} b'_{k,j}\,a'_{j,i} = \sum_{j=1}^{q} b_{j,k}\,a_{i,j} = c'_{k,i}$$

prouve que les matrices ${}^tB{}^tA$ et ${}^t(A\,B)$ ont tous leurs coefficients égaux, donc qu'elles sont égales. □

EXERCICE 3 —

1. *Soit A une matrice inversible.*
   Montrer que sa transposée est inversible et que ${}^t\left(A^{-1}\right) = ({}^tA)^{-1}$.
2. *Soit B une matrice dont la colonne C_j est combinaison linéaire des C_i, $1 \leq i \leq n$, $i \neq j$: $C_j = \sum_{i=0,\,i\neq j}^{n} \lambda_i C_i$.*
   Montrer que B n'est pas inversible.

DÉFINITION 61
Soient $\mathbb{K}$ un corps et $n \in \mathbb{N}^*$. Soit $A \in \mathscr{M}_n(\mathbb{K})$. On dit que la matrice A est **symétrique** si ${}^tA = A$. On note $\mathscr{S}_n(\mathbb{K})$ l'ensemble des matrices $n \times n$ symétriques.
On dit que la matrice A est **antisymétrique** si ${}^tA = -A$. On note $\mathscr{A}_n(\mathbb{K})$ l'ensemble des matrices $n \times n$ antisymétriques.

PROPOSITION 62
Soient $\mathbb{K}$ un corps et $n \in \mathbb{N}^*$. Alors :

(i) Toute matrice $M \in \mathscr{M}_n(\mathbb{K})$ s'écrit d'une unique façon comme la somme d'une matrice symétrique et d'une matrice antisymétrique.

(ii) Les ensembles $\mathscr{S}_n(\mathbb{K})$ et $\mathscr{A}_n(\mathbb{K})$ sont des sous-espaces vectoriels de $\mathscr{M}_n(\mathbb{K})$ qui sont supplémentaires dans $\mathscr{M}_n(\mathbb{K})$ (voir Géométrie 1) :

$$\mathscr{S}_n(\mathbb{K}) \oplus \mathscr{A}_n(\mathbb{K}) = \mathscr{M}_n(\mathbb{K}).$$

Preuve — (i) Soit $M \in \mathscr{M}_n(\mathbb{K})$. On peut alors écrire :

$$M = S + A, \qquad \text{avec} \qquad S = \frac{M + {}^tM}{2} \quad \text{et} \quad A = \frac{M - {}^tM}{2}.$$

D'après la Proposition 59, S est symétrique et A est antisymétrique.

Montrons que cette écriture est unique. Supposons que $M = S + A = S' + A'$.

Alors on a $S - S' = A' - A$ avec $S - S'$ symétrique et $A' - A$ antisymétrique.

Pour $N = S - S'$, N est à la fois symétrique et antisymétrique, donc $N = {}^tN = -N$. Cela donne $2N = 0$, donc N est nulle. Ainsi, on a $S = S'$ et $A = A'$.

(ii) Montrons que $\mathscr{S}_n(\mathbb{K})$ et $\mathscr{A}_n(\mathbb{K})$ sont des sous-espaces vectoriels de $\mathscr{M}(n)\mathbb{K}$.

La matrice nulle est symétrique. L'ensemble $\mathscr{S}_n(\mathbb{K})$ est donc non vide.

Pour A,B deux matrices symétriques, et λ,μ deux scalaires, on a ${}^t(\lambda A + \mu B) = \lambda {}^tA + \mu {}^tB = \lambda A + \mu B$. Donc la matrice $\lambda A + \mu B$ est symétrique. Ainsi, $\mathscr{S}_n(\mathbb{K})$ est bien un sous-espace vectoriel de $\mathscr{M}_n(\mathbb{K})$.

On montre la même chose pour $\mathscr{A}_n(\mathbb{K})$.

De plus, la preuve de (i) a montré que $\mathscr{S}_n(\mathbb{K}) \cap \mathscr{A}_n(\mathbb{K}) = \{0\}$. Ces sous-espaces vectoriels sont donc supplémentaires dans $\mathscr{M}_n(\mathbb{K})$. □

EXERCICE 4 — *Montrer que* $\dim \mathscr{S}_n(\mathbb{K}) = \dfrac{n(n+1)}{2}$ *et* $\dim \mathscr{A}_n(\mathbb{K}) = \dfrac{n(n-1)}{2}$.

9.6 Matrice d'une famille de vecteurs, rang d'une matrice

Définition 63
Soit E un $\mathbb{K}$-ev de dimension finie. Soit $\mathcal{B} = (e_1, \dots, e_n)$ une base de E. Soient $x_1, \dots, x_p$ des vecteurs de E. Pour $1 \leq j \neq p$, soit $x_j = \sum_{i=1}^{n} a_{i,j} e_i$ la décomposition du vecteur x_j dans la base E.
On définit la matrice $\mathrm{Mat}_{\mathcal{B}}(x_1, \dots, x_p) = (a_{i,j})_{i,j} \in \mathscr{M}_{n,p}(\mathbb{K})$.
Cette matrice est appelée **la matrice de la famille** $(x_1, \dots, x_p)$ **dans la base** $\mathcal{B}$.

Remarque 64 — *La j-ème colonne C_j de la matrice $Mat_{\mathcal{B}}(x_1, \dots, x_p)$ contient les coefficients de la décomposition du vecteur x_j dans la base $\mathcal{B}$.*

Exemple 65 — *Pour $E = \mathbb{K}^2$, $\mathcal{B} = ((1,-1),(0,1))$, et $x_1 = (1,0), x_2 = (1,1), x_3 = (2,1)$, on a :*

$$Mat_{\mathcal{B}}(x_1, x_2, x_3) = \begin{pmatrix} 1 & 1 & 2 \\ 1 & 2 & 3 \end{pmatrix}.$$

Remarque 66 — *Soit $A \in \mathscr{M}_{n,p}(\mathbb{K})$. Notons $C_1, \dots, C_p$ les colonnes de A, vues comme vecteurs colonne de $\mathbb{K}^n$. Alors, la matrice A est exactement la matrice de la famille de vecteurs $(C_1, \dots, C_p)$ dans la base banonique de $\mathbb{K}^n$.*

Définition 67
Soient $\mathbb{K}$ un corps et $n, p \geq 1$. Soit $A \in \mathscr{M}_{n,p}(\mathbb{K})$. Soient $C_1, \dots, C_p$ les colonnes de A, vues comme vecteurs colonnes de $\mathbb{K}^n$.
On définit le **rang de** A, noté $\mathrm{rang}(A)$ ou $\mathrm{rg}(A)$, comme le rang de la famille de vecteurs $(C_1, \dots, C_p)$:

$$\mathrm{rang}(A) = \mathrm{rang}(C_1, \dots, C_p).$$

Remarque 68 — *On a ainsi $rang(A) \leq \max(n,p)$, puisque cela correspond au rang d'une famille de p vecteurs dans un e.v. de dimension n.*

Exemple 69 —

- *On a $rg(I_n) = n$*
- *On a $rg(0) = 0$*
- *Soient $a_1, \dots, a_p \in \mathbb{K}$. On pose :*

$$A = \begin{pmatrix} a_1 & \dots & a_p \\ 0 & \dots & 0 \\ \vdots & & \vdots \\ 0 & \dots & 0 \end{pmatrix} \in \mathscr{M}_{n,p}(\mathbb{K}).$$

Alors $rang(A) = 0$ si tous les a_i sont nuls, et $rang(A) = 1$ sinon.

Proposition 70
Soit E un $\mathbb{K}$-ev de dimension n. Soit $\mathcal{B}$ une base de E. Soient $x_1, \dots, x_p$ des vecteurs de E. Alors, on a :

$$\mathrm{rg}(\mathrm{Mat}_{\mathcal{B}}(x_1, \dots, x_p)) = \mathrm{rg}(x_1, \dots, x_p).$$

Autrement dit, le rang de la famille $x_1, \dots, x_p$ est égal au rang de sa matrice associée dans la base $\mathcal{B}$.

Remarque 71 — *Ainsi, on peut déterminer le rang d'une famille de vecteurs en calculant celui de sa matrice associée dans une base $\mathcal{B}$ bien choisie.*

Corollaire 72
Soit E un $\mathbb{K}$-ev de dimension finie. Soient $\mathcal{B}, \mathcal{B}'$ deux bases de E. Soient $x_1, \dots, x_p$ des vecteurs de E. Alors, on a :

$$\mathrm{rg}(\mathrm{Mat}_{\mathcal{B}}(x_1, \dots, x_p)) = \mathrm{rg}(\mathrm{Mat}_{\mathcal{B}'}(x_1, \dots, x_p)).$$

Autrement dit le rang de la matrice associée à la famille de vecteurs $(x_1, \dots, x_p)$ ne dépend pas de la base choisie.

PROPOSITION 73
Soient $\mathbb{K}$ un corps et $n, p \geq 1$. Soit $A \in \mathscr{M}_{n,p}(\mathbb{K})$ une matrice échelonnée.
Alors, rang(A) est égal au nombre de lignes non-nulles de A.

EXEMPLE 74 — *Soit* $A = \begin{pmatrix} -1 & 2 & 0 & 3 \\ 0 & 0 & 2 & 5 \\ 0 & 0 & 0 & 7 \\ 0 & 0 & 0 & 0 \end{pmatrix}$. *Alors, on a rang*$(A) = 3$.
Les vecteurs colonne C_1, C_3, C_4 *de* A *forment une famille échelonnée dans le sous-ev* Vect e_1, e_2, e_3.

PROPOSITION 75
Soient $\mathbb{K}$ un corps et $n, p \geq 1$. Soit $A \in \mathscr{M}_{n,p}(\mathbb{K})$. Soit $M \in \mathscr{M}_n(\mathbb{K})$ une matrice inversible. Alors, on a :

$$\text{rang}(MA) = \text{rang}(A) = \text{rang}(AM).$$

Autrement dit, multiplier la matrice A par une matrice inversible ne change pas son rang.

COROLLAIRE 76 (Calcul du rang par la méthode du Pivot)
Soient $\mathbb{K}$ un corps et $n \geq 1$. Soit $A \in \mathscr{M}_n(\mathbb{K})$.
Soient $M_1, \ldots, M_r$ des matrices élémentaires et B une matrice échelonnée telles que $M_r \ldots M_1 A = B$.
Alors, le rang de A est égal au nombre de lignes non-nulles de B.

Preuve — Les matrices élémentaires sont inversibles, donc on a $\text{rg}(A) = \text{rg}(B)$, et B est une matrice échelonnée. □

En appliquant la méthode du Pivot à la matrice A pour se ramener à une matrice échelonnée, on obtient ainsi un calcul du rang de A.

THÉORÈME 77
Soient $\mathbb{K}$ un corps et $n \geq 1$. Soit $A \in \mathscr{M}_n(\mathbb{K})$.
Alors A est inversible si et seulement si $\text{rang}(A) = n$.

9.7 Trace d'une matrice

DÉFINITION 78
Soient $\mathbb{K}$ un corps et $n \in \mathbb{N}^*$. Soit $A = (a_{i,j})_{(i,j)} \in \mathscr{M}_n(\mathbb{K})$. On définit la **trace** de A, notée $\text{Tr}(A)$ ou $\text{tr}(A)$, comme la somme des éléments de la diagonale de A :

$$\text{Tr}(A) = a_{1,1} + \cdots + a_{n,n} = \sum_{i=1}^{n} a_{i,i} \in \mathbb{K}.$$

REMARQUE 79 — *La trace d'une matrice n'est définie que pour les matrices carrées.*

PROPOSITION 80
Soient $A, B \in \mathscr{M}_n(\mathbb{K})$. On a :

$$\text{Tr}(AB) = \text{Tr}(BA).$$

Preuve — Pour $A = (a_{i,j})_{(i,j)}$ et $B = (b_{i,j})_{(i,j)}$, on a :

$$\begin{array}{rcl} \text{Tr}(AB) & = & \text{Tr}\left(\left(\sum_{k=1}^{n} a_{i,k} b_{k,j}\right)_{(i,j)}\right) \\ & = & \sum_{i=1}^{n} \left(\sum_{k=1}^{n} a_{i,k} b_{k,i}\right) \\ & = & \sum_{k=1}^{n} \left(\sum_{i=1}^{n} b_{k,i} a_{i,k}\right) \\ & = & \text{Tr}\left(\left(\sum_{i=1}^{n} b_{k,i} a_{i,j}\right)_{(k,j)}\right) \\ & = & \text{Tr}(BA). \end{array}$$

□

REMARQUE 81 — *Soient A et B deux matrices carrées. La trace de AB n'est en général pas égale à la trace de A multipliée par la trace de B. Par exemple :*

$$\text{Tr}(I_2 I_2) = \text{Tr}(I_2) = 2 \neq 4 = \text{Tr}(I_2)\text{Tr}(I_2).$$

EXERCICE 5 — *Soit* $A \in \mathscr{M}_n(\mathbb{K})$.

1. *On prend* $\mathbb{K} = \mathbb{R}$. *Montrer alors que* $\text{Tr}({}^tAA)$ *est un nombre positif. Montrer que* $\text{Tr}({}^tAA)$ *est nul si et seulement si* A *est la matrice nulle.*

2. *Montrer qu'il existe une matrice non nulle A dans $\mathscr{M}_2(\mathbb{C})$ telle que $\mathrm{Tr}({}^tAA)$ est nul.*

Exercice 6 —
Montrer que l'ensemble des matrices dont la trace est nulle est un sous-espace vectoriel de $\mathscr{M}_n(\mathbb{K})$. Déterminer une base de ce sous-espace vectoriel. Quelle est la dimension de ce sous-espace vectoriel ?

Chapitre 10 Applications linéaires

Table des matières du chapitre

10.1 Définitions

Définition 1
Soient E et F deux $\mathbb{K}-$espaces vectoriels. On dit qu'une application $u : E \longrightarrow F$ est une **application linéaire** si

$$\forall (x,y) \in E^2\,,\ \forall (\alpha,\beta) \in \mathbb{K}^2\,,\ u(\alpha x + \beta y) = \alpha u(x) + \beta u(y).$$

On dit aussi que u est un **morphisme d'espaces vectoriels** et on parle :

- d'**endomorphisme** si $E = F$;
- d'**isomorphisme** si u est bijective ;
- d'**automorphisme** si u est bijective et $E = F$.

On note :

- $\mathcal{L}(E,F)$ l'ensemble des applications linéaires de E vers F ;
- $\mathcal{L}(E)$ l'ensemble $\mathcal{L}(E,E)$ des endomorphismes de E.
- $\mathrm{GL}(E)$ (**Groupe Linéaire** de E) l'ensemble des automorphismes de l'espace vectoriel E.

Exemple 2 —

1. *L'application* $u : \begin{array}{ccc} \mathbb{K} & \longrightarrow & \mathbb{K}^2 \\ x & \longmapsto & (2x,x) \end{array}$ *est linéaire.*
2. *Soit* $k \in \mathbb{K}$. ***L'homothétie de rapport*** k *est l'application* $\begin{array}{ccc} E & \longrightarrow & E \\ x & \longmapsto & kx \end{array}$.
   C'est un endomorphisme de E. *C'est un automorphisme si et seulement si* $k \neq 0$.
3. *La dérivation est une application linéaire de* $\mathcal{C}^1(\mathbb{R})$ *vers* $\mathcal{C}^0(\mathbb{R})$ *:*

$$D : \begin{array}{ccc} \mathcal{C}^1(\mathbb{R}) & \longrightarrow & \mathcal{C}^0(\mathbb{R}) \\ f & \longmapsto & f' \end{array}$$

   Sa restriction à $\mathcal{C}^\infty(\mathbb{R})$ *est un endomorphisme de* $\mathcal{C}^\infty(\mathbb{R})$.

Remarque 3 — *Si* $u : E \to F$ *est une application linéaire, alors* $u(0_E) = 0_F$.

Preuve — $u(0_E) = u(0_\mathbb{K} 0_E) = 0_\mathbb{K}\, u(0_E) = 0_F$. □

Soient $(a,b) \in \mathbb{K}^2$ et l'application $f : \begin{array}{ccc} \mathbb{K} & \longrightarrow & \mathbb{K} \\ x & \longmapsto & ax+b \end{array}$.
Si $b \neq 0$, alors f n'est pas linéaire car $f(0) = b \neq 0$. (On dit que f est une application affine.)

Remarque 4 (Image d'une combinaison linéaire) — *Si* $u : E \longrightarrow F$ *est linéaire alors pour tout* $n \in \mathbb{N}^*$, *pour tous* $x_1, \ldots, x_n \in E$ *et tous* $\lambda_1, \ldots, \lambda_n \in \mathbb{K}$, *on a*

$$u(\lambda_1 x_1 + \ldots + \lambda_n x_n) = \lambda_1 u(x_1) + \ldots + \lambda_n u(x_n).$$

Proposition 5
Soit $u \in \mathcal{L}(E,F)$.

1. Si A est un sous-espace vectoriel de E, alors $u(A)$ est un sous-espace vectoriel de F.
2. Si B est un sous-espace vectoriel de F, alors $u^{-1}(B)$ est un sous-espace vectoriel de E.

Preuve —

1. $0_E \in A$ donc $0_F = u(0_E) \in u(A)$. De plus, si x', $y' \in u(A)$, il existe x, $y \in A$ tels que $x' = u(x)$ et $y' = u(y)$. On a alors, pour tous λ, $\mu \in \mathbb{K}$,
$$\lambda x' + \mu y' = \lambda u(x) + \mu u(y) = u(\lambda x + \mu y) \in u(A).$$
2. $0_E \in u^{-1}(B)$ car $u(0_E) = 0_F$. De plus, si x, $y \in u^{-1}(B)$, on a $u(x) \in B$ et $u(y) \in B$. Donc pour tous λ, $\mu \in \mathbb{K}$, $u(\lambda x + \mu y) = \lambda u(x) + \mu u(y) \in B$, c'est-à-dire $\lambda x + \mu y \in u^{-1}(B)$.

□

Définition 6
On appelle **forme linéaire** de E une application linéaire de E vers $\mathbb{K}$.

Exemple 7 —

1. *Soit $(a,b) \in \mathbb{K}^2$. L'application* $f : \begin{array}{ccc} \mathbb{K}^2 & \longrightarrow & \mathbb{K} \\ (x,y) & \longmapsto & ax+by \end{array}$ *est une forme linéaire de $\mathbb{K}^2$.*
2. *Pour tout $n \in \mathbb{N}^*$, et pour tout $k \in [\![1,n]\!]$, l'application* $\begin{array}{ccc} \mathbb{K}^n & \longrightarrow & \mathbb{K} \\ (x_1,\cdots,x_n) & \longmapsto & x_k \end{array}$ *associe à un vecteur x sa $k-$ième coordonnée x_k. C'est une forme linéaire de $\mathbb{K}^n$.*
3. *Soit un segment $I=[a,b]$. L'application* $\begin{array}{ccc} \mathcal{C}^0(I) & \longrightarrow & \mathbb{R} \\ f & \longmapsto & \int_a^b f(t)dt \end{array}$ *est une forme linéaire de $\mathcal{C}^0(I)$.*
4. *Soit $a \in \mathbb{K}$. L'évaluation* $\begin{array}{ccc} \mathbb{K}[X] & \longrightarrow & \mathbb{K} \\ P & \longmapsto & P(a) \end{array}$ *d'un polynôme en a est une forme linéaire de $\mathbb{K}[X]$.*
5. *Soit E l'espace vectoriel des suites $(u_n)_{n\in\mathbb{N}} \in \mathbb{K}^{\mathbb{N}}$ qui convergent. L'application* $\begin{array}{ccc} E & \longrightarrow & \mathbb{K} \\ (u_n)_{n\in\mathbb{N}} & \longmapsto & \lim_{n\to+\infty} u_n \end{array}$ *est une forme linéaire sur E.*

10.2 Noyau et image d'une application linéaire

Définition 8
Soient E et F deux espaces vectoriels. Soit u une application linéaire de E vers F.
Le noyau de u, noté $\mathrm{Ker}\,(u)$, est l'ensemble des vecteurs x de E tels que $u(x) = 0_F$:
$$\mathrm{Ker}\,(u) = \{x \in E \mid u(x) = 0_F\} = u^{-1}(\{0_F\}).$$
C'est un sous-espace vectoriel de E.

Exemple 9 — *Déterminons le noyau de l'application linéaire*
$$\begin{array}{ccc} f : \mathbb{R}^3 & \longrightarrow & \mathbb{R}^2 \\ (x,y,z) & \longmapsto & \begin{pmatrix} x+2y+3z \\ x-2y \end{pmatrix} \end{array}$$
.
Soit $(x,y,z) \in \mathbb{R}^3$,
$$\begin{aligned} f(x,y,z) = 0 &\iff \left\{ \begin{array}{lll} x+2y+3z & = 0 \\ x-2y & = 0 \end{array} \right. \\ &\underset{L_2' = L_2 - L_1}{\iff} \left\{ \begin{array}{lll} x+2y+3z & = 0 \\ -4y-3z & = 0 \end{array} \right. \\ &\iff \left\{ \begin{array}{lll} x & = -2y-3z = -\frac{3}{2}z \\ y & = -\frac{3}{4}z \end{array} \right. . \end{aligned}$$

Donc $\mathrm{Ker}\,(u) = \mathbb{K}\begin{pmatrix} -6 \\ -3 \\ 4 \end{pmatrix}$.

THÉORÈME 10
Soit u une application linéaire de E vers F. u est injective si et seulement si $\mathrm{Ker}\,(u) = \{0\}$.

Preuve —

- Si u est injective, alors pour tout $x \in \mathrm{Ker}\,(u)$, on a $u(x) = 0 = u(0)$ et donc $x = 0$.
- Si $\mathrm{Ker}\,(u) = \{0\}$ alors pour tous x, $y \in E$,

$$u(x) = u(y) \Longrightarrow u(x-y) = 0 \Longrightarrow x - y \in \mathrm{Ker}\,(u) \Longrightarrow x - y = 0.$$

□

EXEMPLE 11 — *La dérivation*

$$\begin{array}{rccc} D : & \mathcal{C}^1(\mathbb{R},\mathbb{R}) & \longrightarrow & \mathcal{C}^0(\mathbb{R},\mathbb{R}) \\ & f & \longmapsto & f' \end{array}$$

est une application linéaire. Elle n'est pas injective car son noyau est l'ensemble Vect*(1) des fonctions constantes.*

DÉFINITION 12
Soit u une application linéaire de E vers F.
On définit **l'image de** u par

$$\mathrm{Im}(u) = \{y \in F \mid \exists x \in E,\ y = u(x)\} = u(E).$$

C'est un sous-espace vectoriel de F.

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

EXEMPLE 13 — *La dérivation* $D : \mathcal{C}^1(\mathbb{R},\mathbb{R}) \longrightarrow \mathcal{C}^0(\mathbb{R},\mathbb{R})$ *est surjective, car toute fonction continue est la dérivée d'une fonction de classe* $\mathcal{C}^1$.

Image d'une famille de vecteurs

LEMME 14
Soit $u \in \mathcal{L}(E,F)$, soit $(f_1,\ldots,f_n)$ une famille génératrice de E. Alors $(u(f_1),\ldots,u(f_n))$ est génératrice de $\mathrm{Im}(u)$. C'est-à-dire :

$$\mathrm{Im}(u) = \mathrm{Vect}(u(f_1),\ldots,u(f_n)).$$

Preuve — Soit $y \in \mathrm{Im}(u)$, il existe $x \in E$ tel que $u(x)y$. Or $x = x_1f_1 + \ldots + x_nf_n$ donc $u(x) = x_1u(f_1) + \ldots + x_nu(f_n)$. □

PROPOSITION 15
Soit $\mathcal{B} = (e_1, \cdots, e_p)$ une base d'un espace vectoriel E. Soit $u \in \mathcal{L}(E,F)$.

1. Alors, u est surjective si et seulement si $(u(e_1),\ldots,u(e_n))$ est une famille génératrice de F.
2. Alors, u est injective si et seulement si $(u(e_1),\ldots,u(e_n))$ est libre.
3. Alors, u est un isomorphisme si et seulement si $(u(e_1),\ldots,u(e_n))$ est une base de F.

THÉORÈME 16
Soit $(e_1,\ldots,e_n)$ une base de E et soit $(f_1,\ldots,f_n)$ une famille (quelconque) de vecteurs de F. Alors il existe une unique application linéaire de E vers F telle que pour tout $i \in [\![1,n]\!]$, $u(e_i) = f_i$.

Preuve —

- **Unicité :** Soient u, $v \in \mathcal{L}(E,F)$ telles que pour tout $i \in [\![1,n]\!]$, $u(e_i) = f_i$. Soit $x = x_1e_1 + \ldots + x_ne_n \in E$, on a

$$u(x) = x_1u(e_1) + \ldots + x_nu(e_n) = x_1f_1 + \ldots + x_nf_n = x_1v(e_1) + \ldots + x_nv(e_n) = v(x).$$

- **Existence :** Si $x \in E$, on note $(x_1,\ldots,x_n)$ ses coordonnées dans la base $(e_1,\ldots,e_n)$. Soit

$$\begin{array}{rccc} u : & E & \longrightarrow & F \\ & x & \longmapsto & x_1f_1 + \ldots + x_nf_n. \end{array}$$

f est bien définie par unicité des coordonnées dans une base. Pour tout $i \in [\![1,n]\!]$, le vecteur des coordonnées de e_i dans $(e_1,\ldots,e_n)$ est $(0,\ldots,0,1,0,\ldots,0)$ avec 1 en position i. Donc $u(e_i) = f_i$.
De plus, si x, $y \in E$ et λ, $\mu \in \mathbb{K}$, on a

$$u(\lambda x + \mu y) = (\lambda x + \mu y)_1 f_1 + \ldots + (\lambda x + \mu y)_n f_n = (\lambda x_1 + \mu y_1)f_1 + \ldots + (\lambda x_n + \mu y_n)f_n = \lambda u(x) + \mu u(y).$$

Donc u est bien une application linéaire de E vers F.

□

Corollaire 17
Si u, $v \in \mathcal{L}(E,F)$ et si $\mathcal{B} = (e_1, \dots, e_p)$ est une base de E, alors :

- On a $u = v \iff \forall j \in [\![1,p]\!],\ u(e_j) = v(e_j)$;
- En particulier, on a $u = 0 \iff \forall j \in [\![1,p]\!],\ u(e_j) = 0$.

Définition 18
On dit que deux espaces vectoriels E et F sont **isomorphes** s'il existe un isomorphisme de E vers F.

Corollaire 19
Deux espaces vectoriels de dimension finie sont isomorphes si et seulement si ils ont la même dimension.

Exemple 20 — *Tout $\mathbb{K}$-espace vectoriel de dimension n est isomorphe à $\mathbb{K}^n$.*

10.3 Matrice d'une application linéaire

Représenter une application linéaire par une matrice

Si E est de dimension finie, le Théorème 16 montre que, pour définir une application linéaire u de E vers F, il suffit de :

- **choisir** une base $\mathcal{B}_1 = (e_1, \dots, e_p)$ de E ;
- **choisir** l'image $u(\mathcal{B}_1) = (u(e_1), \dots, u(e_p))$ de cette base.

Si l'espace vectoriel F est aussi de dimension finie, alors on peut ensuite :

- **choisir** une base $\mathcal{B}_2 = (f_1, \dots, f_n)$ de F ;
- **écrire** dans cette base les p vecteurs $u(e_j)$ sous la forme $u(e_j) = \sum_{i=1}^{n} a_{i,j}\, f_i$.

On dit alors que l'application linéaire u est représentée par la matrice $A = (a_{i,j})_{(i,j)\in[\![1,n]\!]\times[\![1,p]\!]}$ dans les bases $\mathcal{B}_1$ et $\mathcal{B}_2$:

Définition 21
Soient E et F deux $\mathbb{K}$-espaces vectoriels de **dimensions finies** : $\dim E = p$ et $\dim F = n$. Soit u une application linéaire de E vers F.

On choisit une base $\mathcal{B}_1 = (e_1, \dots, e_p)$ de E et une base $\mathcal{B}_2 = (f_1, \dots, f_n)$ de F. La **matrice de u dans les bases $\mathcal{B}_1$ et $\mathcal{B}_2$**, notée $\mathrm{Mat}_{\mathcal{B}_1,\mathcal{B}_2}(u)$, est la matrice

$$A = \begin{pmatrix} a_{1,1} & a_{1,2} & \dots & a_{1,j} & \dots & a_{1,p} \\ a_{2,1} & a_{2,2} & \dots & a_{2,j} & \dots & a_{2,p} \\ \vdots & \vdots & & \vdots & & \vdots \\ a_{i,1} & a_{i,2} & \dots & a_{i,j} & \dots & a_{i,p} \\ \vdots & \vdots & & \vdots & & \vdots \\ a_{n,1} & a_{n,2} & \dots & a_{n,j} & \dots & a_{n,p} \end{pmatrix}$$

telle que :

$$\forall j \in [\![1,p]\!],\ u(e_j) = \sum_{i=1}^{n} a_{i,j}\, f_i.$$

C'est une matrice à n lignes et p colonnes. Elle se lit en colonnes : pour chaque $j \in [\![1,p]\!]$, la $j-$ième colonne contient les coordonnées, dans la base $\mathcal{B}_2$, du vecteur $u(e_j)$. Autrement dit,

$$\mathrm{Mat}_{\mathcal{B}_1,\mathcal{B}_2}(u) = \mathrm{Mat}_{\mathcal{B}_2}(u(\mathcal{B}_1)).$$

Exemple 22 —

1. *Déterminons la matrice de l'application* $u : \begin{array}{ccc} \mathbb{K} & \longrightarrow & \mathbb{K}^2 \\ x & \longmapsto & (x, 2x) \end{array}$ *dans les bases* $\mathcal{B}_1 = (2)$ *et* $\mathcal{B}_2 = ((1,0),(1,1))$.
On a $u(2) = (2,4) = -2(1,0) + 4(1,1)$. *Donc*
$$\mathrm{Mat}_{\mathcal{B}_1,\mathcal{B}_2}(u) = \begin{pmatrix} -2 \\ 4 \end{pmatrix}.$$

2. *Soit* $\varphi : \begin{array}{ccc} \mathbb{R}_3[X] & \longrightarrow & \mathbb{R}_2[X] \\ u = (u_0, u_1, u_2, u_3, 0, 0, \ldots) & \longmapsto & (u_1, 2u_2, 3u_3, 0, 0, \ldots) \end{array}$, *et soit* $\mathcal{B}_1 = (X^0, X^1, X^2, X^3)$ *la base canonique de* $\mathbb{R}_3[X]$ *et* $\mathcal{B}_1 = (X^0, X^1, X^2)$ *la base canonique de* $\mathbb{R}_2[X]$. *On a :*
$$\begin{aligned} \varphi(X^0) &= 0, \\ \varphi(X^1) &= X^0, \\ \varphi(X^2) &= 2X^1, \\ \varphi(X^3) &= 3X^2. \end{aligned}$$
On obtient donc
$$\mathrm{Mat}_{\mathcal{B}_1,\mathcal{B}_2}(\varphi) = \begin{pmatrix} 0 & 1 & 0 & 0 \\ 0 & 0 & 2 & 0 \\ 0 & 0 & 0 & 3 \end{pmatrix}.$$

Proposition 23
Soit $\mathcal{B}_1 = (e_1, \ldots, e_p)$ une base d'un espace vectoriel E et $\mathcal{B}_2 = (f_1, \ldots, f_n)$ une base d'un espace vectoriel F. Soient deux vecteurs $x = x_1e_1 + \ldots + x_pe_p \in E$ et $y = y_1f_1 + \ldots + y_nf_n \in F$. Soit $u \in \mathcal{L}(E,F)$. On a :

$$\boxed{y = u(x) \iff Y = A \cdot X}$$

où $X = \begin{pmatrix} x_1 \\ \vdots \\ x_p \end{pmatrix}$ est le vecteur des coordonnées de x dans $\mathcal{B}_1$, $Y = \begin{pmatrix} y_1 \\ \vdots \\ y_n \end{pmatrix}$ est le vecteur des coordonnées de y dans $\mathcal{B}_2$ et $A = \mathrm{Mat}_{\mathcal{B}_1,\mathcal{B}_2}(u)$.

Remarque 24 —

1. *Dans le cas où u est un endomorphisme de E, on peut choisir la même base $\mathcal{B}$ pour écrire un vecteur $x \in E$ et son image $u(x) \in E$. La matrice $A = \mathrm{Mat}_{\mathcal{B},\mathcal{B}}(u)$ est alors carrée. On l'appelle la matrice de u dans la base $\mathcal{B}$ et elle se note plus simplement $\mathrm{Mat}_{\mathcal{B}}(u)$.*

2. *Dans deux bases différentes $\mathcal{B}$ et $\mathcal{C}$ de E, un même endomorphisme $u \in \mathcal{L}(E)$ peut être représenté par deux matrices $\mathrm{Mat}_{\mathcal{B}}(u)$ et $\mathrm{Mat}_{\mathcal{C}}(u)$ différentes.*

3. *Soient $(\ell_1, \ell_2, \cdots, \ell_p) \in \mathbb{K}^p$ et $\mathcal{B}_1 = (e_1, e_2, \cdots, e_p) \in E^p$ une base d'un espace vectoriel E. Soit $u \in \mathcal{L}(E, \mathbb{K})$ la forme linéaire définie par : pour chaque vecteur $x = x_1e_1 + x_2e_2 + \cdots + x_pe_p \in E$,*
$$u(x) = \ell_1x_1 + \ell_2x_2 + \cdots + \ell_px_p = \sum_{j=1}^{p} \ell_j\, x_j = \begin{pmatrix} \ell_1 & \ell_2 & \cdots & \ell_p \end{pmatrix} \begin{pmatrix} x_1 \\ x_2 \\ \vdots \\ x_p \end{pmatrix}.$$
Si on choisit le scalaire 1 comme base $\mathcal{B}_2$ de $\mathbb{K}$, alors $L = \begin{pmatrix} \ell_1 & \ell_2 & \cdots & \ell_p \end{pmatrix}$ est la matrice $\mathrm{Mat}_{\mathcal{B}_1,\mathcal{B}_2}(u)$ de u dans les bases $\mathcal{B}_1$ et $\mathcal{B}_2$.

4. *Si A est une matrice à n lignes et p colonnes, alors on peut définir l'application*
$$\mathbb{K}^p \to \mathbb{K}^n, \quad \begin{pmatrix} x_1 \\ x_2 \\ \vdots \\ x_p \end{pmatrix} \mapsto A \cdot \begin{pmatrix} x_1 \\ x_2 \\ \vdots \\ x_p \end{pmatrix}.$$
*C'est une application linéaire u de $\mathbb{K}^p$ vers $\mathbb{K}^n$ et A est la matrice de u dans les bases canoniques de $\mathbb{K}^p$ et $\mathbb{K}^n$. Cette application linéaire u est appelée l'**application linéaire canoniquement associée** à la matrice A.*

10.4 Changement de bases, matrices de passage

Matrice de passage

Soit E un espace vectoriel de dimension finie et soient $\mathcal{B} = (e_1, \dots, e_n)$ et $\mathcal{B}' = (e'_1, \dots, e'_n)$ deux bases de E.

Définition 25
La **matrice de passage** de la base $\mathcal{B}$ à la base $\mathcal{B}'$ est la matrice

$$\mathrm{P}_{\mathcal{B}\to\mathcal{B}'} = \mathrm{Mat}_{\mathcal{B}}(e'_1, \dots, e'_n) = \mathrm{Mat}_{\mathcal{B}}\left(\mathcal{B}'\right).$$

Dit autrement, $\mathrm{P}_{\mathcal{B}\to\mathcal{B}'}$ est la matrice dont la j-ième colonne contient les coordonnées du vecteur e'_j dans la base $\mathcal{B}$ *(voir le cours d'Algèbre 1 pour la définition de la matrice d'une famille de vecteurs).*

C'est donc la matrice carrée à n lignes et n colonnes :

$$\begin{array}{c} \\ e_1 \\ e_2 \\ \\ e_i \\ \\ e_n \end{array} \begin{array}{c} \begin{array}{ccccccc} e'_1 & & e'_2 & & & e'_j & & & e'_n \end{array} \\ \begin{pmatrix} a_{1,1} & & a_{1,2} & \dots & & a_{1,j} & \dots & & a_{1,n} \\ a_{2,1} & & a_{2,2} & \dots & & a_{2,j} & \dots & & a_{2,n} \\ \vdots & & \vdots & & & \vdots & & & \vdots \\ a_{i,1} & & a_{i,2} & \dots & & a_{i,j} & \dots & & a_{i,n} \\ \vdots & & \vdots & & & \vdots & & & \vdots \\ a_{n,1} & & a_{n,2} & \dots & & a_{n,j} & \dots & & a_{n,n} \end{pmatrix} \end{array}$$

telle que :

$$\forall j \in [\![1, n]\!], \quad e'_j = \sum_{i=1}^{n} a_{i,j}\, e_i.$$

Remarque 26 — *La matrice de passage de $\mathcal{B}$ à $\mathcal{B}'$ est aussi à la matrice de l'identité dans les bases $\mathcal{B}'$ et $\mathcal{B}$:*

$$\mathrm{P}_{\mathcal{B}\to\mathcal{B}'} = \mathrm{Mat}_{\mathcal{B}',\mathcal{B}}\left(\mathrm{Id}_E\right).$$

En effet, pour tout $j \in [\![1, n]\!]$, on a $\mathrm{Id}_E(e'_j) = e'_j$ et on range les coordonnées de $\mathrm{Id}_E(e'_j)$ dans la base $\mathcal{B}$.

Soit x un vecteur de E dont les coordonnées sont $(x_1, \dots, x_n)$ dans $\mathcal{B}$ et $(x'_1, \dots, x'_n)$ dans $\mathcal{B}'$. On peut écrire :

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

D'après l'unicité de la décomposition dans la base $\mathcal{B}$:

$$\forall i \in [\![1, n]\!], \quad x_i = \sum_{j=1}^{n} a_{i,j}\, x'_j.$$

On a donc prouvé la le résultat suivant :

Théorème 27 (Formule de changement de base pour un vecteur)
Soient deux bases $\mathcal{B}$ et $\mathcal{B}'$ d'un espace vectoriel E de dimension finie. Soit P la matrice de passage de $\mathcal{B}$ à $\mathcal{B}'$. Si Xest la matrice colonne des coordonnées d'un vecteur x dans la base $\mathcal{B}$ et X' la matrice colonne des coordonnées du même vecteur x dans la base $\mathcal{B}'$, alors

$$X = P\, X'.$$

Proposition 28
La matrice de passage de $\mathcal{B}$ à $\mathcal{B}'$ est inversible et son inverse est la matrice de passage de $\mathcal{B}'$ à $\mathcal{B}$:

$$\left(\mathrm{P}_{\mathcal{B}\to\mathcal{B}'}\right)^{-1} = \mathrm{P}_{\mathcal{B}'\to\mathcal{B}}.$$

Remarque 29 — *La matrice de passage donne :*

- *(en lisant les colonnes) les vecteurs de la **nouvelle** base $\mathcal{B}'$ en fonction des vecteurs de l'**ancienne** base $\mathcal{B}$;*

- *(en calculant $X = P\,X'$) les* ***anciennes*** *coordonnées X d'un vecteur quelconque en fonction de ses* ***nouvelles*** *coordonnées X'.*

Exemple 30 — *Dans l'espace vectoriel $\mathbb{R}^2$, soient une première base (e_1, e_2) et la seconde base (f_1, f_2) définie par :*

$$f_1 = 2e_1 + 3e_2 \quad \text{et} \quad f_2 = 4e_1 + 5e_2.$$

La famille (f_1, f_2) est bien une nouvelle base de $\mathbb{R}^2$ car les deux vecteurs f_1 et f_2 ne sont pas colinéaires. La matrice de passage de la première base (e_1, e_2) à la seconde base (f_1, f_2) est $\begin{pmatrix} 2 & 4 \\ 3 & 5 \end{pmatrix}$.

Les formules de changement de bases

$$\begin{cases} x = 2x' + 4y' \\ y = 3x' + 5y' \end{cases}$$

expriment les anciennes coordonnées (x, y) en fonction des nouvelles coordonnées (x', y').

Changement de base pour une application linéaire

Proposition 31
Soient E et F deux espaces vectoriels de dimension finie et u une application linéaire de E vers F. Si :

- $\mathcal{B}_1$ et $\mathcal{B}'_1$ sont deux bases de E, et P la matrice de passage de $\mathcal{B}_1$ à $\mathcal{B}'_1$,
- $\mathcal{B}_2$ et $\mathcal{B}'_2$ sont deux bases de F, et Q la matrice de passage de $\mathcal{B}_2$ à $\mathcal{B}'_2$,
- $A = \mathrm{Mat}_{\mathcal{B}_1,\mathcal{B}_2}(u)$ et $A' = \mathrm{Mat}_{\mathcal{B}'_1,\mathcal{B}'_2}(u)$,

alors :

$$A' = Q^{-1}\,A\,P.$$

Corollaire 32
Soient u un endomorphisme d'un espace vectoriel E de dimension finie. Si :

- $\mathcal{B}$ et $\mathcal{B}'$ sont deux bases de E,
- P est la matrice de passage de $\mathcal{B}$ à $\mathcal{B}'$,
- $A = \mathrm{Mat}_{\mathcal{B}}(u)$ et $A' = \mathrm{Mat}_{\mathcal{B}'}(u)$,

alors

$$\boxed{A' = P^{-1}\,A\,P.}$$

Remarque 33 — *S'il existe une matrice inversible P telle que $A' = P^{-1}AP$, on dit que les matrices A et A' sont* ***semblables****.*

Remarque 34 — *La matrice d'un endomorphisme u change quand on change de base (A devient $A' = P^{-1}\,A\,P$) mais la trace de cette matrice ne change pas : $\mathrm{Tr}(A) = \mathrm{Tr}(A')$. On peut donc parler de la trace d'un endomorphisme, sans préciser dans quelle base :*

$$\mathrm{Tr}(u) = \mathrm{Tr}(\mathrm{Mat}_{\mathcal{B}}(u))$$

ne dépend pas de la base $\mathcal{B}$.
On dit que la trace est un invariant de similitude.

10.5 Opérations sur les applications linéaires

Combinaison linéaire

Proposition 35
Soient u et v deux applications linéaires de E vers F, soient λ, $\mu \in \mathbb{K}$.

1. L'application $\lambda u + \mu v$ est aussi linéaire.

2. Si E et F sont de dimensions finies, alors

$$\mathrm{Mat}_{\mathcal{B}_1,\mathcal{B}_2}(\lambda u+\mu v)=\lambda\,\mathrm{Mat}_{\mathcal{B}_1,\mathcal{B}_2}(u)+\mu\,\mathrm{Mat}_{\mathcal{B}_1,\mathcal{B}_2}(v),$$

où $\mathcal{B}_1$ est une base de E et $\mathcal{B}_2$ une base de F.

Exemple 36 — *Si u est un endomorphisme d'un espace vectoriel E, alors tout vecteur $x\in E$ tel que $u(x)=x$ est appelé un **vecteur invariant** par u. L'ensemble des vecteurs invariants par u est un sous-espace vectoriel de E : C'est le noyau de $f-\mathrm{id}_E$. En effet :*

$$\forall x\in E,\quad f(x)=x\iff f(x)-x=0_E\iff (f-\mathrm{id}_E)(x)=0_E\iff x\in\mathrm{Ker}\,(f-\mathrm{id}_E)\,.$$

Corollaire 37

1. $\mathcal{L}(E,F)$ est un $\mathbb{K}$-espace vectoriel.
2. Si E et F sont de dimensions finies, alors :
   - l'espace vectoriel $\mathcal{L}(E,F)$ est isomorphe à $\mathscr{M}_{n,p}(\mathbb{K})$, avec $\dim E=p$ et $\dim F=n$;
   - $\mathcal{L}(E,F)$ est de dimension finie et $\dim(\mathcal{L}(E,F))=\dim(E)\times\dim(F)$.

Composition

Proposition 38
Soient E, F et G trois espaces vectoriels.

1. Si $u\in\mathcal{L}(E,F)$ et $v\in\mathcal{L}(F,G)$, alors $v\circ u\in\mathcal{L}(E,G)$.
   Autrement dit : la composée de deux applications linéaires est linéaire.
2. Si E, F et G sont de dimensions finies, alors

$$\mathrm{Mat}_{\mathcal{B}_1,\mathcal{B}_3}(v\circ u)=\mathrm{Mat}_{\mathcal{B}_2,\mathcal{B}_3}(v)\cdot\mathrm{Mat}_{\mathcal{B}_1,\mathcal{B}_2}(u)$$

où $\mathcal{B}_1$ est une base de E, $\mathcal{B}_2$ une base de F et $\mathcal{B}_3$ une base de G.

Exemple 39 — *Les deux applications linéaires*

$$\begin{array}{rcl} f:\mathbb{K}^2 & \longrightarrow & \mathbb{K}^2 \\ (x,y) & \longmapsto & (y,x)\end{array}$$

et

$$\begin{array}{rcl} g:\mathbb{K}^2 & \longrightarrow & \mathbb{K}^2 \\ (x,y) & \longmapsto & (x+y,x).\end{array}$$

ne commutent pas car

$$(f\circ g)(1,2)=f(3,1)=(1,3)\quad et\quad (g\circ f)(1,2)=g(2,1)=(3,2).$$

Donc la composition n'est pas commutative dans $\mathcal{L}(\mathbb{K}^2)$.

Proposition 40
Soit E un $\mathbb{K}$-espace vectoriel.
Alors, $(\mathcal{L}(E),+,\circ)$ est un anneau.
Et $(\mathcal{L}(E),+,\circ,.)$ est une $\mathbb{K}$-algèbre.
Si E est de dimension finie, $\mathcal{L}(E)$ est un $\mathbb{K}$-ev de dimension $\dim(E)^2$.

Définition 41
Si $u\in\mathcal{L}(E)$ est un endomorphisme de E, alors on peut définir **la puissance** n de u, par $u^n=\underbrace{u\circ u\circ\cdots\circ u}_{n\text{ fois}}$.
Plus précisément, on note $u^1=u$, $u^2=u\circ u$ et pour tout $n\geq 1$, $u^{n+1}=u^n\circ u=u\circ u^n$.
Par convention, on a $u^0=\mathrm{Id}_E$.

Pour tout $n\in\mathbb{N}$, la puissance n de u est un endomorphisme de E : $u^n\in\mathcal{L}(E)$.

Applications linéaires inversibles

Proposition 42
Soient E et F deux espaces vectoriels. Soit u un isomorphisme de E vers F.

1. Alors, la bijection réciproque u^{-1} est linéaire et bijective de F vers E.
2. Si E et F sont de dimension finie, alors $\mathrm{Mat}_{\mathcal{B}_1,\mathcal{B}_2}(u)$ est inversible et

$$\big(\mathrm{Mat}_{\mathcal{B}_1,\mathcal{B}_2}(u)\big)^{-1} = \mathrm{Mat}_{\mathcal{B}_2,\mathcal{B}_1}(u^{-1}),$$

où $\mathcal{B}_1$ est une base de E et $\mathcal{B}_2$ une base de F.

Corollaire-Définition 43
On note $\mathrm{GL}(E)$ l'ensemble des automorphismes d'un espace vectoriel E.
Il s'agit du groupe des éléments inversibles de l'anneau $\mathcal{L}(E)$. On l'appelle le **Groupe Linéaire**.
Si E est de dimension finie n, alors $(\mathrm{GL}(E), \circ)$ est isomorphe au groupe $(\mathrm{GL}_n(\mathbb{K}), \times)$.

Proposition 44
Soit E un espace vectoriel de dimension finie n muni d'une base $\mathcal{B}$. Une famille $(x_1, x_2, \ldots, x_n)$ de vecteurs de E est une base de E si, et seulement si, la matrice $\mathrm{Mat}_{\mathcal{B}}(x_1, x_2, \ldots, x_n)$ est inversible.

10.6 Rang d'une application linéaire, théorème du rang

Définition 45
Soit une application linéaire $u : E \to F$ d'un espace vectoriel E vers un espace vectoriel F.
Si $\mathrm{Im}\,(u)$ est de dimension finie, alors **le rang** de u est

$$\mathrm{rg}(u) = \dim \mathrm{Im}\,(u)\,.$$

Remarque 46 — *Soient E et F deux espaces vectoriels de dimensions finies, $\mathcal{B}$ une base de E et $\mathcal{C}$ une base de F. Soient $u \in \mathcal{L}(E,F)$ et $A = \mathrm{Mat}_{\mathcal{B},\mathcal{C}}(u)$. Alors*

$$\mathrm{rg}(u) = \mathrm{rg}(u(\mathcal{B})) = \mathrm{rg}(A).$$

Le théorème du rang

Lemme 47
Soit u une application linéaire d'un espace vectoriel E vers un espace vectoriel F. Soit E_0 un supplémentaire de $\mathrm{Ker}\,(u)$: $E = E_0 \oplus \mathrm{Ker}\,(u)$. La restriction de u à E_0 est un isomorphisme de E_0 vers $\mathrm{Im}\,(u)$.

Preuve — L'application

$$\begin{array}{rccc} v\,:\, E_0 & \longrightarrow & \mathrm{Im}\,(u) \\ x & \longmapsto & u(x) \end{array}$$

est bien définie car l'image de tout vecteur de E (et donc de E_0) appartient à $\mathrm{Im}\,(u)$. L'application v est :

- linéaire car u est linéaire ;
- injective car
$$\mathrm{Ker}\,(v) = \big\{x \in E_0,\ u(x) = 0\big\} = E_0 \cap \mathrm{Ker}\,(u)$$
est égal $\{0_E\}$ car E_0 et $\mathrm{Ker}\,(u)$ sont en somme directe ;
- surjective car $\forall y \in \mathrm{Im}\,(u)\,,\ \exists x \in E,\ y = u(x)$. Or E_0 et $\mathrm{Ker}\,(u)$ sont supplémentaires, d'où
$$x = x_1 + x_2 \quad \text{avec} \quad x_1 \in E_0 \quad \text{et} \quad x_2 \in \mathrm{Ker}\,(u)\,.$$
D'où
$$y = u(x) = u(x_1) + u(x_2) = u(x_1) = v(x_1).$$

□

Théorème 48 (Théorème du rang)
Soient E et F deux espaces vectoriels, u une application linéaire de E vers F. Si E est de dimension finie, alors

$$\dim(E) = \dim(\mathrm{Ker}\,(u)) + \underbrace{\dim(\mathrm{Im}\,(u))}_{\mathrm{rg}(u)}.$$

Preuve — Soit E_0 un supplémentaire de $\mathrm{Ker}\,(u)$. D'après le lemme précédent, les sous-espaces vectoriels E_0 et $\mathrm{Im}\,(u)$ sont isomorphes. Ils ont donc même dimension. D'où

$$\dim(E) = \dim(E_0) + \dim(\mathrm{Ker}\,(u)) = \dim(\mathrm{Im}\,(u)) + \dim(\mathrm{Ker}\,(u)).$$

□

PROPOSITION 49
Soient E un espace vectoriel de dimension finie et u un endomorphisme de E :

$$u \text{ injectif} \iff u \text{ surjectif} \iff u \text{ bijectif.}$$

Preuve — D'après le théorème du rang,

$$\dim(E) = \dim(\mathrm{Im}\,(u)) + \dim(\mathrm{Ker}\,(u)). \qquad (*)$$

L'application u est injective si, et seulement si, $\mathrm{Ker}\,(u) = \{0\}$, ce qui équivaut d'après $(*)$, à $\dim(E) = \dim(\mathrm{Im}\,(u))$, c'est-à-dire à $E = \mathrm{Im}\,(u)$ et finalement à la surjectivité de u. □

En dimension finie, pour montrer qu'un endomorphisme est bijectif, il suffit donc de montrer qu'il est injectif (ou surjectif, mais l'injectivité est souvent plus simple à démontrer que la surjectivité).

REMARQUE 50 — *Si l'espace vectoriel E est de dimension infinie, alors un endomorphisme de E peut être :*

- *injectif sans être surjectif, comme le prouve l'exemple de l'endomorphisme*

$$\mathbb{R}[X] \to \mathbb{R}[X],\ P(X) \mapsto X\,P(X)\ ;$$

- *surjectif sans être injectif, comme le prouve l'exemple de la dérivation*

$$\mathbb{R}[X] \to \mathbb{R}[X],\ P(X) \mapsto P'(X).$$

Rang et transposée d'une matrice

Soient un entier $r \in [\![0, \min(n,p)]\!]$ et la matrice

$$J_r = \begin{pmatrix} I_r & 0_{r,p-r} \\ 0_{n-r,r} & 0_{n-r,p-r} \end{pmatrix}.$$

Cette matrice appartient à $\mathscr{M}_{n,p}(\mathbb{K})$ et son rang est r.

PROPOSITION 51
Une matrice M de $\mathscr{M}_{n,p}(\mathbb{K})$ est de rang r si, et seulement si, il existe $Q \in \mathrm{GL}_n(\mathbb{K})$ et $P \in \mathrm{GL}_p(\mathbb{K})$ telles que $M = Q^{-1}\,J_r\,P$.

Preuve — Supposons que $\mathrm{rg}(M) = r$ et notons u l'application linéaire de $E = \mathbb{K}^p$ vers $F = \mathbb{K}^n$ canoniquement associée à M. Le rang de u est r, d'où (théorème du rang) $\dim \mathrm{Ker}\,(u) = p - r$. Soient $(e_{r+1}, e_{r+2}, \cdots, e_p)$ une base de $\mathrm{Ker}\,(u)$ et $(e_1, e_2, \cdots, e_r)$ une base d'un supplémentaire E_0 de $\mathrm{Ker}\,(u)$. On obtient alors une base $\mathcal{B}_1 = (e_1, \cdots, e_p)$ de E.

D'après le lemme 47, la restriction de u à E_0 est un isomorphisme de E_0 vers $\mathrm{Im}\,(u)$. D'où la famille $(f_1, f_2, \cdots, f_r)$ définie par $f_1 = u(e_1)$, $f_2 = u(e_2)$,..., $f_r = u(e_r)$ est une base de $\mathrm{Im}\,(u)$. On peut la compléter pour obtenir une base $\mathcal{B}_2 = (f_1, f_2, \ldots, f_r, f_{r+1}, \ldots, f_n)$ de F.

$J_r = \mathrm{Mat}_{\mathcal{B}_1,\mathcal{B}_2}(u)$, d'où (proposition 31) il existe des matrices inversibles Q et P telles que $M = Q^{-1}\,J_r\,P$.

Réciproquement : si $M = Q^{-1}\,J_r\,P$, alors M et J_r représentent un même endomorphisme u. Or $\mathrm{rg}(u) = \mathrm{rg}(J_r) = r$. Donc $\mathrm{rg}(M) = r$.
□

COROLLAIRE 52
Le rang d'une matrice M est égal au rang de sa transposée : $\mathrm{rg}(M) = \mathrm{rg}({}^tM)$.

Preuve — Soient $M \in \mathscr{M}_{n,p}(\mathbb{K})$ et r le rang de M. D'après la proposition précédente, il existe $P \in \mathrm{GL}_n(\mathbb{K})$ et $Q \in \mathrm{GL}_p(\mathbb{K})$ tels que $M = Q^{-1}\,J_r\,P$. D'où

$${}^tM = {}^tP\,{}^tJ_r\,{}^t(Q^{-1}).$$

Les matrices ${}^tP$ et ${}^t(Q^{-1})$ sont inversibles et la matrice ${}^tJ_r$ est du même type que J_r mais avec p lignes et n colonnes. Donc, d'après la proposition précédente, le rang de ${}^tM$ est r. □

La transposition échange les lignes et les colonnes, par suite le rang d'une matrice $M \in \mathscr{M}_{n,p}(\mathbb{K})$ est égal au rang :

- de la famille de ses vecteurs colonnes (d'où $\mathrm{rg}(M) \leq p$) ;
- de la famille de ses vecteurs lignes (d'où $\mathrm{rg}(M) \leq n$).

10.7 Formes linéaires et hyperplans

Proposition-Définition 53
Soit H un sous-espace vectoriel d'un $\mathbb{K}-$espace vectoriel E de dimension finie n. Les propriétés suivantes sont équivalentes :

(*i*) $\dim(H) = n-1$;
(*ii*) il existe une droite vectorielle D telle que $E = H \oplus D$;
(*iii*) il existe une forme linéaire non nulle $f : E \to \mathbb{K}$ telle que $H = \mathrm{Ker}(f)$.

On appelle **hyperplan** de E tout sous-espace vectoriel H de E vérifiant l'une de ces conditions.

Proposition 54
Soient E un $\mathbb{K}-$espace vectoriel de dimension finie n, et$\mathcal{B} = (e_1, \cdots, e_n)$ une base de E. Soit $x = x_1e_1 + \cdots + x_ne_n$ la décomposition dans la base $\mathcal{B}$.
Alors, toute forme linéaire $f : E \to \mathbb{K}$ s'écrit $f(x) = a_1x_1 + \cdots a_nx_n$.

L'équation d'un hyperplan $H = \mathrm{Ker}(f)$ est donc

$$a_1x_1 + \cdots + a_nx_n = 0, \quad \text{avec} \quad (a_1, \cdots, a_n) \in \mathbb{K}^n \setminus \{(0, \cdots, 0)\},$$

Proposition 55
Deux formes linéaires f et g non nulles ont le même noyau si, et seulement si, il existe $\lambda \in \mathbb{K}^*$ tel que $f = \lambda g$.
Autrement dit : deux hyperplans d'équations

$$\sum_{i=1}^{n} a_i\, x_i = 0 \quad \text{et} \quad \sum_{i=1}^{n} b_i\, x_i = 0$$

sont égaux si, et seulement si, il existe $\lambda \in \mathbb{K}^*$ tel que $(b_1, \ldots, b_n) = \lambda\,(a_1, \ldots, a_n)$.

10.8 Projecteurs & symétries

Projecteurs

Définition 56
Soient F et G deux sous-espaces vectoriels **supplémentaires** de E : $E = F \oplus G$. Tout vecteur $x \in E$ s'écrit alors de manière unique $x = x_F + x_G$, où $x_F \in F$ et $x_G \in G$.

1. L'application $p : E \to E$ définie par $p(x) = x_F$ est le **projecteur sur F parallèlement à G**. Le vecteur $p(x)$ est le **projeté de x sur F parallèlement à G**.
2. L'application $q : E \to E$ définie par $q(x) = x_G$ est le **projecteur sur G parallèlement à F**. Le vecteur $q(x)$ est le **projeté de x sur G parallèlement à F**.

Proposition 57
Soient p et q comme dans la définition précédente. Alors :

1. Les projecteurs p et q sont des endomorphismes de E, et $p + q = \mathrm{id}_E$.
2. On a $\mathrm{Im}(p) = F$, $\mathrm{Im}(q) = G$, $\mathrm{Ker}(p) = G$, $\mathrm{Ker}(q) = F$
3. On a $p \circ p = p$, $q \circ q = q$, $p \circ q = q \circ p = 0$

Remarque 58 —

1. *On retiendra donc que : un projecteur projette sur son image parallèlement à son noyau.*
2. *Le projecteur p projette sur F mais p n'est pas une application linéaire de E vers F : c'est une application linéaire de E vers E, c'est-à-dire un endomorphisme de E. Ne pas confondre l'ensemble d'arrivée E et l'image F.*
3. *Si $p \in \mathcal{L}(E)$ est un projecteur, alors l'image de p est aussi l'ensemble des vecteurs invariants (voir exemple 36) par p :*
$$\mathrm{Im}\,(p) = \mathrm{Ker}\,(p - \mathrm{id}_E)\,.$$

Preuve — *D'après la proposition, $p - \mathrm{id}_E = -q$ et $\mathrm{Ker}\,(q) = \mathrm{Im}\,(p)$.* □

Proposition 59
Soit $p \in \mathcal{L}(E)$.
Alors, p est un projecteur si, et seulement si, $p \circ p = p$.

Symétries

Définition 60
Soient F et G deux sous-espaces vectoriels **supplémentaires** de E : $E = F \oplus G$. Tout vecteur $x \in E$ s'écrit alors de manière unique $x = x_F + x_G$, où $x_F \in F$ et $x_G \in G$.

L'application $s : E \to E$ définie par $s(x) = x_F - x_G$ est **la symétrie par rapport à F parallèlement à G**. Le vecteur $s(x)$ est le **symétrique de x par rapport à F parallèlement à G**.

Exemple 61 —

1. *La conjugaison $\mathbb{C} \to \mathbb{C},\ z \mapsto \bar{z}$ est la symétrie par rapport à $\mathbb{R}$ parallèlement à $i\mathbb{R}$.*
2. *L'application $\mathbb{K}[X] \to \mathbb{K}[X],\ P(X) \mapsto P(-X)$ est la symétrie par rapport au sous-espace vectoriel des polynômes pairs parallèlement au sous-espace vectoriel des polynômes impairs.*

Proposition 62
Soient p et s deux endomorphismes d'un espace vectoriel E tels que

$$s + \mathrm{id}_E = 2p.$$

Soient F et G deux sous-espaces vectoriels supplémentaires de E : p est le projecteur sur F parallèlement à G si, et seulement si, s est la symétrie par rapport à F parallèlement à G.

Cette proposition montre qu'une symétrie s est linéaire (car $s = 2p - \mathrm{id}_E \in \mathcal{L}(E)$).

Proposition 63
Soit s un endomorphisme de E.
Alors, s est une symétrie si, et seulement si, $s \circ s = \mathrm{id}_E$.

Remarque 64 —

1. *Toute symétrie est bijective et égale à sa réciproque : $s = s^{-1}$*
2. *Toute symétrie s est une symétrie par rapport à $\mathrm{Ker}\,(s - \mathrm{id}_E)$ et parallèlement à $\mathrm{Ker}\,(s + \mathrm{id}_E)$.*

Chapitre 11 Notion de déterminant

Ce chapitre est constitué de deux étapes :

1. Avec le groupe symétrique, nous définissons les *formes multi-linéaires alternées*.
2. Avec les formes multi-linéaires alternées, nous définissons le *déterminant*. Nous étudions alors les nombreuses propriétés de cette fonction.

Table des matières du chapitre

11.1 Formes p-linéaires

Définition 1 (**forme p-linéaire**)
Soit E un $\mathbb{K}$-espace vectoriel. Soit $p \geq 1$ un entier. Soit $f : E^p \to \mathbb{K}$ une fonction. On dit que la fonction f est une **forme p-linéaire** si, pour toute famille $(a_1, a_2, \ldots, a_p) \in E^p$ et pour tout $i \in [\![1, p]\!]$, la i-ème fonction :

$$\begin{array}{rccc} f_i : & E & \longrightarrow & \mathbb{K} \\ & x & \longmapsto & f(a_1, a_2, \ldots, a_{i-1}, x, a_{i+1}, \ldots, a_p) \end{array}$$

est une application linéaire.

Exemples 2

1. *Les formes 1-linéaires sont les formes linéaires.*
2. *Les formes bilinéaires sont les formes 2-linéaires.*
3. *Dans un espace euclidien, le produit scalaire est une forme bilinéaire.*
4. *La fonction :*
$$\begin{array}{rccc} \phi : & \mathcal{C}^0\big([0,1],\mathbb{R}\big)^2 & \longrightarrow & \mathbb{R} \\ & (u, v) & \longmapsto & \int_0^1 u(t)\, v(t)\, dt \end{array}$$
est ainsi une forme bilinéaire sur $E = \mathcal{C}^0\big([0,1],\mathbb{R}\big)$.
5. *La fonction :*
$$\begin{array}{rccc} \psi : & \mathbb{C}^p & \longrightarrow & \mathbb{C} \\ & (z_1, z_2, \ldots, z_p) & \longmapsto & z_1\, z_2 \ldots z_p \end{array}$$
est une forme p-linéaire sur $\mathbb{C}$.

Autrement dit, une forme p-linéaire est une fonction $f : (u_1, \ldots, u_p) \in E^p \to \mathbb{K}$ qui est *linéaire par rapport à chaque variable* $u_1, \ldots, u_p$ (linéaire en u_1, linéaire en u_2, ..., linéaire en u_p).

11.1.1 Formes p-linéaires alternées, antisymétriques

Définition 3
Soient E un $\mathbb{K}$-e.v. et $p \geq 2$. Soit $f : E^p \to \mathbb{K}$ une forme p-linéaire sur E. On dit que f est une forme p-linéaire **alternée** si, pour tout $(u_1, u_2, \ldots, u_p) \in E^p$ et pour tout $i \neq j$, on a :

$$u_i = u_j \Longrightarrow f(u_1, u_2, \ldots, u_p) = 0.$$

Définition 4
Soient E un $\mathbb{K}$-e.v. et $p \geq 2$. Soit $f : E^p \to \mathbb{K}$ une forme p-linéaire sur E. On dit que f est une forme p-linéaire **antisymétrique** si, pour tout $(u_1, u_2, \ldots, u_p) \in E^p$ et pour tout $i < j$, on a :

$$f(u_1, \ldots, \underset{\substack{\uparrow \\ i\text{-ième}}}{u_i}, \ldots \underset{\substack{\uparrow \\ j\text{-ième}}}{u_j}, \ldots, u_p) = -f(u_1, \ldots, \underset{\substack{\uparrow \\ i\text{-ième}}}{u_j}, \ldots \underset{\substack{\uparrow \\ j\text{-ième}}}{u_i}, \ldots, u_p).$$

Proposition 5
Soient E un $\mathbb{K}$-e.v., $p \geq 2$, et f une forme p-linéaire sur E.
Si f est alternée, alors elle est antisymétrique.
Si est $\mathbb{K}$ un corps tel que $Car(\mathbb{K}) \neq 2$ $(1 + 1 \neq 0)$ et si f est antisymétrique, alors f est alternée.

Preuve —

- "$\Rightarrow$" Soient $u_1, \ldots, u_p \in E$. Soient i, j tels que $i < j$. Alors, la fonction :

$$\begin{array}{rcl} g : E^2 & \longrightarrow & F \\ (x, y) & \longmapsto & f(u_1, \ldots, \underset{\substack{\uparrow \\ i\text{-ième}}}{x}, \ldots \underset{\substack{\uparrow \\ j\text{-ième}}}{y}, \ldots, u_p) \end{array}$$

  est une forme bilinéaire sur E, avec :

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

  Soient $(x, y) \in E^2$. Comme g est bilinéaire, on a :

$$\begin{aligned} 0 = g(x + y, x + y) &= g(x, x + y) + g(y, x + y) = g(x, x) + g(x, y) + g(y, x) + g(y, y) \\ &= 0 + g(x, y) + g(y, x) + 0. \end{aligned}$$

  On a donc :

$$g(x, y) = -g(y, x), \ \forall (x, y) \in E^2.$$

- "$\Leftarrow$" Soit f une forme bilinéaire antisymétrique. Soient i, j avec $i < j$. On a alors par antisymétrie :

$$f(u_1, \ldots, \underset{\substack{\uparrow \\ u_i}}{x}, \ldots \underset{\substack{\uparrow \\ u_j}}{x}, \ldots, u_p) = -f(u_1, \ldots, \underset{\substack{\uparrow \\ u_j}}{x}, \ldots \underset{\substack{\uparrow \\ u_j}}{x}, \ldots, u_p),$$

  cela donne : $2.f(u_1, \ldots, x, \ldots, x, \ldots, u_p) = 0$.
  Comme $1 + 1 = 2 \neq 0$ dans $\mathbb{K}$, on obtient alors $f(u_1, \ldots, x, \ldots, x, \ldots, u_p) = 0$, ce qui conclut.

□

Exemples 6

1. *La forme nulle $f : (u_1, \ldots, u_p) \in E^p \mapsto 0 \in \mathbb{K}$ est une forme p-linéaire alternée.*
2. *La fonction*

$$\begin{array}{rcl} f : (\mathbb{K}^2)^2 & \longrightarrow & \mathbb{K} \\ ((x_1, x_2), (y_1, y_2)) & \longmapsto & x_1 y_2 - x_2 y_1 \end{array}$$

   est une forme bilinéaire alternée. Par contre, $(x, y) \mapsto x_1 y_2 + x_2 y_1$ n'est pas alternée. ◄
3. *Pour E un $\mathbb{R}$-e.v. euclidien, le produit scalaire sur E est une forme bilinéaire qui n'est pas alternée.*

Proposition 7
Soient E un $\mathbb{K}$-e.v., $p \geq 2$, et f une forme p-linéaire alternée sur E. Soit $(u_1, u_2, \ldots, u_p)$ une famille de vecteurs de E qui est liée. Alors, on a $f(u_1, u_2, \ldots, u_p) = 0$.

Corollaire 8
Soient une forme p-linéaire alternée f sur E et $(u_1, u_2, \ldots, u_p) \in E^p$. Alors le nombre $f(u_1, u_2, \ldots, u_p) \in \mathbb{K}$ est inchangé si l'on ajoute à l'un des u_j une combinaison linaire des vecteurs u_i, $i \neq j$.

Formes p-linéaires alternées et permutation

Proposition 9
Soient E un $\mathbb{K}$-e.v. et $f : E^n \to \mathbb{K}$ une forme n-linéaire alternée sur E. Soit $\sigma \in \mathcal{S}_n$ une permutation. Alors, on a :

$$f(u_{\sigma(1)}, u_{\sigma(2)}, \ldots, u_{\sigma(n)}) = \varepsilon(\sigma)\, f(u_1, u_2, \ldots, u_n), \ \forall (u_1, \ldots, u_n) \in E^n.$$

11.1.2 Formes n-linéaires alternées en dimension n

THÉORÈME 10
Soient E un $\mathbb{K}$-e.v. de dimension n et $\mathcal{B} = (e_1, e_2, \dots, e_n)$ une base de E. Alors :

1. Il existe une unique forme n-linéaire alternée ϕ_0 sur E telle que :
$$\phi_0(e_1, e_2, \dots, e_n) = 1.$$
2. Toute forme n-linéaire alternée sur E est proportionnelle à ϕ_0.

COROLLAIRE 11
Soit E un $\mathbb{K}$-e.v. de dimension n. Alors l'ensemble des forme n-linéaires alternées sur E est un e.v. de dimension 1.

11.2 DÉTERMINANT

11.2.1 Diverses notions de déterminants

Déterminant d'une famille de vecteurs dans une base

DÉFINITION 12
Soient E un $\mathbb{K}$-e.v. de dimension n, $\mathcal{B}$ une base de E, et ϕ_0 l'unique forme n-linéaire alternée telle que $\phi_0(\mathcal{B}) = 1$.
Soient $u_1, \dots, u_n \in E$.
On dit que le nombre $\phi_0(u_1, u_2, \dots, u_n)$ s'appelle **déterminant de la famille** $(u_1, u_2, \dots, u_n)$ **dans la base** $\mathcal{B}$. On le note $\det_{\mathcal{B}}(u_1, u_2, \dots, u_n)$.

On note $\det_{\mathcal{B}}$ la fonction $\det_{\mathcal{B}} : (u_1, u_2, \dots, u_n) \in E^n \mapsto \det_{\mathcal{B}}(u_1, u_2, \dots, u_n) \in \mathbb{K}$.

Déterminant d'un endomorphisme

PROPOSITION-DÉFINITION 13
Soient E un $\mathbb{K}$-e.v. de dimension n et $f : E \to E$ un endomorphisme. Alors il existe un unique $\lambda \in \mathbb{K}$ tel que pour toute base $\mathcal{B}$ de E et pour tout $(u_1, u_2, \dots, u_n) \in E^n$, on ait :
$$\det_{\mathcal{B}}\big(f(u_1), f(u_2), \dots, f(u_n)\big) = \lambda \det_{\mathcal{B}}(u_1, u_2, \dots, u_n).$$

Ce nombre est appelé **déterminant** de f . On le note $\det(f)$ ou $\det f$.
Pour $\mathcal{B} = (e_1, \dots, e_n)$ une base de E, on a :
$$\lambda = \det f = \det_{\mathcal{B}}\big(f(e_1), \dots, f(e_n)\big) = \det_{\mathcal{B}}\big(f(\mathcal{B})\big).$$

REMARQUE 14 — *Pour $B = (e_1, \dots, e_n)$ une base de E, et $u_1, \dots, u_n \in E$ avec $u_j = \sum_{i=1}^{n} a_{i,j} e_i$, on a :*
$$det_B(u_1, \dots, u_n) = \sum_{\sigma \in \mathcal{S}_n} \varepsilon(\sigma) a_{\sigma(1),1}\, a_{\sigma(2),2} \dots a_{\sigma(n),n}.$$

Cette expression est la formule générique du déterminant d'une famille de n vecteurs dans une base.
Pour $f : E \to E$ une application linéaire, $det(f)$ est égal au déterminant de la famille $f(e_1), \dots, f(e_n)$ dans la base B.
Dans le cas où $E = \mathbb{K}^n$ et où B est la base canonique de $\mathbb{K}^n$, on notera parfois det à la place de det_B.

EXEMPLES 15

1. *Pour $\mathcal{B}$ une base de E, on a* $\det \mathrm{Id}_E = \det_{\mathcal{B}}(\mathrm{Id}_E(\mathcal{B})) = \det_{\mathcal{B}}(\mathcal{B}) = 1$.
2. *Soient F et G deux sous-espaces supplémentaires de E. Soit s la symétrie par rapport à F parallèlement à G. Soien $(e_1, e_2, \dots, e_p)$ et $(e_{p+1}, e_{p+2}, \dots, e_n)$ des bases de F et de F. Alors $\mathcal{B} = (e_1, e_2, \dots, e_n)$ est une base de E , et on a :*
$$\begin{aligned}\det s &= \det_{\mathcal{B}}\big(s(e_1), \dots, s(e_n)\big) = \det_{\mathcal{B}}(e_1, \dots, e_p, -e_{p+1}, \dots, -e_n) \\ &= (-1)^{n-p} = (-1)^{\dim G}.\end{aligned}$$

3. *Soient $\sigma \in \mathcal{S}_n$ et $\mathcal{B} = (e_1, e_2, \dots, e_n)$ une base de E. On définit l'application linéaire $f : E \to E$ par :*

$$f(e_i) = e_{\sigma(i)}, \forall i \in [\![1, n]\!].$$

Alors le déterminant de f vaut :

$$\det f = \det_{\mathcal{B}}(e_{\sigma(1)}, e_{\sigma(2)}, \dots, e_{\sigma(n)}) = \varepsilon(\sigma) \det_{\mathcal{B}}(\mathcal{B}) = \varepsilon(\sigma).$$

Déterminant d'une matrice carrée

Définition 16
Soit $n \geq 1$. Soit $A \in \mathscr{M}_n(\mathbb{K})$. Soient $C_1, \dots, C_n$ les colonnes de A, vues comme vecteurs colonnes de $\mathbb{K}^n$. Soit B la base canonique de $\mathbb{K}^n$.
On définit le **déterminant de la matrice** A, noté $\det(A)$ ou $\det A$, par $det(A) = det_B(C_1, \dots, C_n)$.

Le déterminant d'une matrice carrée A est donc le déterminant de ses vecteurs colonnes, par rapport à la base canonique de $\mathbb{K}^n$.

Remarque 17 — *Pour $A = (a_{i,j})_{\substack{1 \leqslant i \leqslant n \\ 1 \leqslant j \leqslant n}}$, le déterminant de A se note aussi :*

$$\det A = \begin{vmatrix} a_{1,1} & \cdots & a_{1,j} & \cdots & a_{1,n} \\ \vdots & & \vdots & & \vdots \\ a_{i,1} & \cdots & a_{i,j} & \cdots & a_{i,n} \\ \vdots & & \vdots & & \vdots \\ a_{n,1} & \cdots & a_{n,j} & \cdots & a_{n,n}. \end{vmatrix}.$$

On a :

$$det(A) = \sum_{\sigma \in \mathcal{S}_n} \varepsilon(\sigma) a_{\sigma(1),1}\, a_{\sigma(2),2} \dots a_{\sigma(n),n}.$$

Ceci est l'expression générique du déterminant d'une matrice carrée.

Exemples 18

1. *Pour $n = 2$, on a :*

$$\begin{vmatrix} a & b \\ c & d \end{vmatrix} = ad - bc.$$

2. *Pour $n = 3$, on a :*

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

Maintenant que le déterminant d'une matrice carrée a été défini, nous allons pouvoir énoncer des propriétés vérifiées par cette fonction. Ces propriétés sont nombreuses, ce qui montre l'importance de cette fonction.

Proposition 19
Soient $\mathbb{K}$ un corps et $n \geq 1$. Alors la fonction $\det : A \in \mathscr{M}_n(\mathbb{K}) \mapsto \det(A) \in \mathbb{K}$ est une fonction polynômiale en les coefficients de la matrice d'entrée A.

Preuve — Cela découle de l'expression de $\det(A)$. □

Proposition 20 (Déterminant d'une matrice et déterminant d'une famille de vecteurs)
Soit $\mathcal{B} = (e_1, \dots, e_n)$ une base de E. Soit $(u_1, u_2, \dots, u_n)$ une famille de n vecteurs de E. Soit A la matrice des coefficients de $(u_1, \dots, u_n)$ dans la base $\mathcal{B}$.
Alors, on a $\det_{\mathcal{B}}(u_1, u_2, \dots, u_n) = \det(A)$.

Proposition 21 (Déterminant d'une matrice et déterminant d'une application linéaire)
Soient $f : E \to E$ une application linéaire et $\mathcal{B} = (e_1, \dots, e_n)$ une base de E. Soit $A = \mathrm{Mat}_{\mathcal{B}}(f)$ la matrice de f dans la base $\mathcal{B}$.
Alors, on a $\det f = \det A$.

Exemple 22 — *Si D est une matrice diagonale d'éléments diagonaux $\lambda_1, \lambda_2, \ldots, \lambda_n$, on a :*

$$\det D = \prod_{i=1}^{n} \lambda_i.$$

La matrice D est la matrice de la famille $(\lambda_1 e_1, \lambda_2 e_2, \ldots, \lambda_n e_n)$ dans la base $\mathcal{B} = (e_1, e_2, \ldots, e_n)$. Par n-linéarité du déterminant, on a donc :

$$\det D = \det_{\mathcal{B}}(\lambda_1 e_1, \lambda_2 e_2, \ldots, \lambda_n e_n) = \left(\prod_{i=1}^{n} \lambda_i\right) \det_{\mathcal{B}}(\mathcal{B}) = \prod_{i=1}^{n} \lambda_i.$$

Proposition 23 (Déterminant et géométrie du plan)
Soit $n = 2$. Soit $\mathcal{B} = (\vec{\imath}, \vec{\jmath})$ une base orthonormée du plan $\mathbb{R}^2$. Alors, $|\det_{\mathcal{B}}(\vec{u}, \vec{v})|$ est égal à la surface du parallélogramme de côtés $\vec{u} = \begin{pmatrix} a \\ b \end{pmatrix}$ et $\vec{v} = \begin{pmatrix} c \\ d \end{pmatrix}$ non nuls.

Remarque 24 (Déterminant et géométrie de l'espace) — *Soit $n = 3$. Soit $\mathcal{B} = (\vec{\imath}, \vec{\jmath}, \vec{k})$ une base orthonormée directe de l'espace $\mathbb{R}^3$. Alors,*

$$\det_{\mathcal{B}}(\vec{u}, \vec{v}, \vec{w}) = \vec{u} \cdot (\vec{v} \wedge \vec{w}) \qquad \textit{et} \qquad V = |\det_{\mathcal{B}}(\vec{u}, \vec{v}, \vec{w})|$$

est le volume du parallépipède[1] *de côtés $\vec{u}$, $\vec{v}$ et $\vec{w}$.*

11.2.2 Propriétés du déterminant

Théorème 25
Soient E un $\mathbb{K}$-e.v. de dimension n et $\mathcal{B}$ une base de E. Soit $(u_1, u_2, \ldots, u_n)$ est une famille de n vecteurs de E. Alors, les propriétés suivantes sont équivalentes :

i) $(u_i)_{1 \leqslant i \leqslant n}$ est une base de E,
ii) $\det_{\mathcal{B}}(u_1, u_2, \ldots, u_n) \neq 0$.

Proposition 26
Soient E un $\mathbb{K}$-e.v. de dimension n, f, g deux endomorphismes de E, et $\lambda \in \mathbb{K}$. Alors, on a :

- $\det(\lambda f) = \lambda^n \det(f)$;
- $\det(f \circ g) = \det(f) \det(g)$.

Proposition 27
Soit $n \geq 1$. Soient $A, B \in \mathscr{M}_n(\mathbb{K})$. Soit $\lambda \in \mathbb{K}$. Alors on a :

- $\det(\lambda A) = \lambda^n \det(A)$;
- $\det(A B) = \det(A) \det(B)$.

Remarque 28 — *Pour $(A, B) \in \mathscr{M}_n(\mathbb{K})^2$, on a donc* $\det(A B) = \det(A) \det(B) = \det(B A)$.

Proposition 29
Soit $n \geq 1$. Soit E un $\mathbb{K}$-e.v. de dimension n.

1. Soit $f : E \to E$ une application linéaire. Alors, f est bijective si, et seulement si, $\det f \neq 0$.
   Dans ce cas, on obtient :
   $$\det\left(f^{-1}\right) = (\det f)^{-1}.$$
2. Soit $A \in \mathscr{M}_n(\mathbb{K})$. Alors, A est inversible si, et seulement si, $\det A \neq 0$.
   Dans ce cas, on obtient :
   $$\det\left(A^{-1}\right) = (\det A)^{-1}.$$

Proposition 30
Soit $n \geq 2$. Soit $A \in \mathscr{M}_n(\mathbb{K})$. Alors, on a :
$$\det A = \det\left({}^tA\right).$$

Notons $L_1, \ldots, L_n$ les lignes de A, vues comme vecteurs lignes de $\mathbb{K}^n$. Alors, on a $\det(A) = \det({}^tL_1, \ldots, {}^tL_1)$. La fonction $A \mapsto \det(A)$ est une forme n-linéaire alternée en les lignes de la matrice A.

1.

Remarque 31 — *Pour $A \in \mathscr{M}_n(\mathbb{K})$, la quantité* $\det(A)$ *reste la même que l'on considère A comme une matrice composée de n vecteurs colonnes $C_1, \dots, C_n$ ou une matrice de n vecteurs lignes $L_1, \dots, L_n$.*
On peut ainsi calculer le déterminant d'une matrice A en effectuant des opérations sur lignes ou sur les colonnes de A (que des opérations sur les lignes, ou que des opérations sur les colonnes, ou un mélange d'opérations sur les lignes et d'opérations sur les colonnes).

11.2.3 Opérations sur les lignes ou les colonnes d'un déterminant

Comme on l'a précédemment vu , le déterminant d'une famille de vecteurs par rapport à une base, ou du déterminant d'un endomorphisme, est le déterminant d'une certaine matrice carrée. Nous nous intéresserons donc aux méthodes permettant de calculer le déterminant d'une matrice A.

Le déterminant d'une matrice étant une forme n-linéaire alternée des colonnes ou des lignes de cette matrice, les propriétés des formes n-linéaires alternées permettent d'énoncer les règles suivantes.

Proposition 32 (Déterminant et opérations sur les lignes/colonnes)
Soit $n \geq 2$. Soit $A \in \mathscr{M}_n(\mathbb{K})$.

- Si A a deux colonnes (resp. deux lignes) identiques, alors $\det(A) = 0$.
- L'échange de deux colonnes de A (resp. deux lignes) multiplie son déterminant par -1.
- Si une colonne (resp. une ligne) de A est combinaison linaire des **autres** colonnes (resp. des **autres** lignes), alors $\det(A) = 0$.
- Si une colonne (resp. une ligne) de A est formée de 0, alors $\det(A) = 0$.
- La valeur de $\det(A)$ est inchangée si l'on ajoute à une colonne (resp. à une ligne) de A une combinaison linaire des **autres** colonnes (resp. des **autres** lignes).
- Si l'on multiplie une colonne de A (resp. une ligne) par λ, alors son déterminant est multiplié par λ.
  Donc, si l'on multiplie la matrice A par λ, son déterminant est multiplié par λ^n.

Remarque 33 — *Ces règles de transformation d'un déterminant permettent :*

- *Soit de prouver qu'il est nul,*
- *Soit d'introduire dans une colonne (resp. une ligne) un maximum de 0 afin de pouvoir utiliser les résultats qui vont suivre.*

Cette proposition décrit le comportement du déterminant lorsque l'on applique à une matrice A des opérations élémentaires sur ses lignes ou sur ses colonnes ($L_i \leftarrow L_i + \lambda L_j$, $L_i \leftarrow \lambda L_i$, $L_i \leftrightarrow L_j$).
On peut ainsi utiliser la ***méthode du Pivot*** *(ou Pivot de Gauss) pour calculer* $\det(A)$.
La méthode du Pivot permet de se ramener à une matrice échelonnée, et nous allons voir des façons de calculer le déterminant d'une matrice échelonnée.

Exemples 34

1. *On a :*
$$\begin{vmatrix} 1 & 2 & 3 \\ 1 & 2 & 4 \\ 1 & 2 & 5 \end{vmatrix} = 0$$
puisque la matrice a deux colonnes proportionnelles.
2. *On a :*
$$\begin{vmatrix} 1 & 2 & 3 \\ 2 & 3 & 4 \\ 3 & 4 & 5 \end{vmatrix} = 0$$
puisque la deuxième ligne est la demi-somme des deux autres. ($L_3 = \frac{L_1+L_2}{2}$)

11.2.4 Développement d'un déterminant selon une colonne ou une ligne

Proposition 35
Soit $n \geq 2$. Soit $A \in \mathscr{M}_n(\mathbb{K})$ de la forme :
$$A = \begin{pmatrix} & & & 0 \\ & A' & & \vdots \\ & & & 0 \\ * & \dots & * & a_{n,n} \end{pmatrix}.$$

Alors, on a $\det A = a_{n,n} \det A'$.

COROLLAIRE 36
Soit $A = (a_{i,j})_{i,j} \in \mathscr{M}_n(\mathbb{K})$. Si A est une matrice triangulaire, alors on a : $\det A = \prod_{i=1}^{n} a_{i,i}$. [2]

Preuve

- Comme $\det(A) = \det({}^tA)$, il suffit de démontrer le résultat pour les matrices triangulaires inférieures.
- Si A est triangulaire inférieure, la proposition précédente nous donne $\det A = a_{n,n} \det A'$ où A' est la matrice A privée de sa dernière ligne et de sa dernière colonne. Une récurrence nous donne alors le résultat.

□

EXEMPLE 37 — *Soient $a, b, c \in \mathbb{K}$. On veut calculer le déterminant :*

$$\Delta = \begin{vmatrix} 1 & a & a^2 \\ 1 & b & b^2 \\ 1 & c & c^2 \end{vmatrix}.$$

On a successivement :

$$\begin{aligned}
\Delta &= \begin{vmatrix} 1 & a & a^2 \\ 0 & b-a & b^2-a^2 \\ 0 & c-a & c^2-a^2 \end{vmatrix} && L_2 \leftarrow L_2 - L_1,\ L_3 \leftarrow L_3 - L_1 \\
&= (b-a)(c-a) \begin{vmatrix} 1 & a & a^2 \\ 0 & 1 & b+a \\ 0 & 1 & c+a \end{vmatrix} && \text{mise en facteur dans } L_2 \text{ et } L_3 \\
&= (b-a)(c-a) \begin{vmatrix} 1 & a & a^2 \\ 0 & 1 & b+a \\ 0 & 0 & c-b \end{vmatrix} && L_3 \leftarrow L_3 - L_2 \\
&= (b-a)(c-a)(c-b) && \text{d'après le corollaire 36.}
\end{aligned}$$

DÉFINITION 38
Soit $n \geq 2$. Soit $A = (a_{i,j}) \in \mathscr{M}_n(\mathbb{K})$. Soient $i, j \in [\![1, n]\!]$.
On note $\Delta_{i,j}$ le déterminant de la matrice extraite de A obtenue en supprimant la i-ième ligne et la j-ième colonne de A.
Ce déterminant est appelé **mineur** de A. [3]
Le nombre $(-1)^{i+j}\Delta_{i,j}$ est appelé **cofacteur** de A.

THÉORÈME 39 (Développement du déterminant selon une colonne)
Soit $n \geq 2$. Soit $A = (a_{i,j})_{i,j} \in \mathscr{M}_n(\mathbb{K})$. On a :

$$\det A = \sum_{i=1}^{n} a_{i,j} \, (-1)^{i+j} \Delta_{i,j}, \ \forall j \in [\![1, n]\!].$$ [4] (développement selon la j-ème colonne de A)

En appliquant ce résultat à la transposée de A, on obtient :

THÉORÈME 40 (Développement du déterminant selon une ligne)
Soit $n \geq 2$. Soit $A = (a_{i,j})_{i,j} \in \mathscr{M}_n(\mathbb{K})$. On a :

$$\det A = \sum_{j=1}^{n} a_{i,j} (-1)^{i+j} \Delta_{i,j}, \ \forall i \in [\![1, n]\!].$$ (développement selon la i-ème ligne de A)

EXEMPLES 41

1. *Pour un déterminant 3×3, on a donc : (développement selon la 1-ère ligne)*

$$\begin{vmatrix} a_{1,1} & a_{1,2} & a_{1,3} \\ a_{2,1} & a_{2,2} & a_{2,3} \\ a_{3,1} & a_{3,2} & a_{3,3} \end{vmatrix} = a_{1,1} \begin{vmatrix} a_{2,2} & a_{2,3} \\ a_{3,2} & a_{3,3} \end{vmatrix} - a_{1,2} \begin{vmatrix} a_{2,1} & a_{2,3} \\ a_{3,1} & a_{3,3} \end{vmatrix} + a_{1,3} \begin{vmatrix} a_{2,1} & a_{2,2} \\ a_{3,1} & a_{3,2} \end{vmatrix}.$$

2.
3.
4.

Mais aussi : (développement selon la 2-ème colonne)

$$\begin{vmatrix} a_{1,1} & a_{1,2} & a_{1,3} \\ a_{2,1} & a_{2,2} & a_{2,3} \\ a_{3,1} & a_{3,2} & a_{3,3} \end{vmatrix} = -a_{1,2} \begin{vmatrix} a_{2,1} & a_{2,3} \\ a_{3,1} & a_{3,3} \end{vmatrix} + a_{2,2} \begin{vmatrix} a_{1,1} & a_{1,3} \\ a_{3,1} & a_{3,3} \end{vmatrix} - a_{3,2} \begin{vmatrix} a_{1,1} & a_{1,3} \\ a_{2,1} & a_{2,3} \end{vmatrix}.$$

2. *Le déterminant de Vandermonde .*
Soient $x_0, x_1, \ldots, x_n \in \mathbb{K}$. On définit $V(x_0, x_1, \ldots, x_n)$ le déterminant de Vandermonde par :

$$V(x_0, x_1, \ldots, x_n) = \begin{vmatrix} 1 & 1 & \cdots & 1 \\ x_0 & x_1 & \cdots & x_n \\ x_0^2 & x_1^2 & \cdots & x_n^2 \\ \vdots & \vdots & & \vdots \\ x_0^n & x_1^n & \cdots & x_n^n \end{vmatrix}$$

On a alors :

$$V(x_0, x_1, \ldots, x_n) = \prod_{0 \leqslant j < i \leqslant n} (x_i - x_j).$$

Cela montre que :

$$V(x_0, x_1, \ldots, x_n) = 0 \iff \exists i, j \in [\![0, n]\!] \text{ tels que } i \neq j \text{ et } x_i = x_j.$$

Montrons cela par récurrence sur n. On définit la propriété H_n : Soient $x_0, x_1, \ldots, x_n \in \mathbb{K}$. On a :

$$V(x_0, x_1, \ldots, x_n) = \prod_{0 \leqslant j < i \leqslant n} (x_i - x_j).$$

- *H_1 est vraie, car pour $x_0, x_1 \in \mathbb{K}$, on a :*

$$V(x_0, x_1) = \begin{vmatrix} 1 & 1 \\ x_0 & x_1 \end{vmatrix} = x_1 - x_0.$$

- *Soit $n \geq 2$. Supposons que H_{n-1} est vraie. Soient $x_0, x_1, \ldots, x_n \in \mathbb{K}$.*
  - *Si les scalaires $x_0, x_1, \ldots, x_n$ ne sont pas distincts deux à deux, le déterminant $V(x_0, x_1, \ldots, x_n)$ a deux colonnes identiques et est donc nul. Dans ce cas $V(x_0, x_1, \ldots, x_n) = \prod_{0 \leqslant j < i \leqslant n} (x_i - x_j) = 0$.*
  - *Si les scalaires $x_0, x_1, \ldots, x_n$ sont distincts deux à deux, on développe ce déterminant par rapport à la dernière colonne. Cela montre que la fonction $f : x \mapsto V(x_0, x_1, \ldots, x_{n-1}, x)$ est une fonction polynomiale en x, de degré inférieur ou égal à n, et dont le terme de degré n est :*

$$V(x_0, x_1, \ldots, x_{n-1})\, x^n.$$

D'après H_{n-1}, on a :

$$V(x_0, x_1, \ldots, x_{n-1}) = \prod_{0 \leqslant j < i \leqslant n-1} (x_i - x_j)$$

Cette quantité est donc non nulle car $x_0, x_1, \ldots, x_{n-1}$ sont deux à deux distincts. Ainsi, f est une fonction polynomiale de degré n.
Or, f admet comme racines $x_0, x_1, \ldots, x_{n-1}$ car pour tout $i \in [\![0, n-1]\!]$, la quantité $f(x_i)$ est un déterminant admettant deux colonnes identiques. On connaît donc toutes les racines de f. Cela donne :

$$V(x_0, x_1, \ldots, x_{n-1}, x) = V(x_0, x_1, \ldots, x_{n-1}) \prod_{j=0}^{n-1} (x - x_j), \ \forall x \in \mathbb{K}.$$

Ceci implique :

$$V(x_0, x_1, \ldots, x_n) = \prod_{0 \leqslant j < i \leqslant n-1} (x_i - x_j) \prod_{j=0}^{n-1} (x_n - x_j) = \prod_{0 \leqslant j < i \leqslant n} (x_i - x_j),$$

ce qui montre que H_n est vraie, ce qui termine la récurrence.

11.2.5 Déterminant d'une matrice par blocs

PROPOSITION 42
Soit $n \geq 2$. Soit $A \in \mathscr{M}_n(\mathbb{K})$ possédant un bloc de zéros en bas à gauche. On l'écrit sous la forme :

$$A = \begin{pmatrix} C & D \\ 0 & E \end{pmatrix}$$

où C est une matrice $p \times p$, D une matrice $p \times (n-p)$, 0 la matrice nulle $(n-p) \times p$, et E une matrice $(n-p) \times (n-p)$. Alors on a :

$$\det A = (\det C) \cdot (\det E).$$

Par itération de ce théorème, on peut calculer facilement les déterminants de matrices "triangulaires par blocs".

EXEMPLES 43

1. $$\det \begin{bmatrix} 1 & -1 & 4 & 3 & 7 & -2 \\ 1 & 2 & -2 & -3 & 5 & -6 \\ 0 & 0 & 4 & 2 & -3 & 7 \\ 0 & 0 & -1 & -3 & 5 & 2 \\ 0 & 0 & 0 & 0 & -2 & 3 \\ 0 & 0 & 0 & 0 & 3 & -2 \end{bmatrix} = \det \begin{bmatrix} 1 & -1 \\ 1 & 2 \end{bmatrix} \cdot \det \begin{bmatrix} 4 & 2 \\ -1 & -3 \end{bmatrix} \cdot \det \begin{bmatrix} -2 & 3 \\ 3 & -2 \end{bmatrix}.$$

2. *Comme le déterminant est invariant par transposition, on peut aussi calculer des déterminants "triangulaires inférieurs par blocs" :*

$$\det \begin{bmatrix} 1 & 1 & 0 & 0 \\ 1 & 2 & 0 & 0 \\ 1 & 2 & 3 & 4 \\ 5 & 6 & 7 & 8 \end{bmatrix} = \det \begin{bmatrix} 1 & 1 \\ 1 & 2 \end{bmatrix} \cdot \det \begin{bmatrix} 3 & 4 \\ 7 & 8 \end{bmatrix}.$$

Le déterminant d'une matrice "diagonale par blocs" est facile à calculer : c'est le produit des déterminants de tous les blocs.

COROLLAIRE 44
Soit $n \geq 2$. Soit $1 \leq k \geq n$ et $n_1, \ldots, n_k \in \mathbb{N}^*$ tels que $n_1 + \ldots + n_k = n$. Soit $A \in \mathscr{M}_n(\mathbb{K})$

1. Si A une matrice diagonale par blocs :

$$A = \begin{pmatrix} A_1 & 0 & \cdots & 0 \\ 0 & A_2 & \cdots & 0 \\ \vdots & \vdots & \ddots & 0 \\ 0 & 0 & \cdots & A_k \end{pmatrix}, \text{ avec } A_i \in \mathscr{M}_{n_i}(\mathbb{K}),$$

   alors on a $\det A = \prod_{i=1}^{k} \det A_i$.

2. Soit E un $\mathbb{K}$-e.v. de dimension n.Soient $E_1, \ldots, E_k$ des sous-e.v. de E en somme directe : $E = \bigoplus_{i=1}^{k} E_i$. Soit $u : E \to E$ un endomorphisme, tel que chaque sous-e.v. E_i est stable par u. On note $u_i : E_i \to E_i$ l'endomorphisme induit par u sur le sous-espace E_i.
   Alors, on a $\det u = \prod_{i=1}^{k} \det u_i$.

Preuve — La seconde partie est seulement la traduction de la première partie au niveau des endomorphismes.
La première partie se prouve par récurrence sur k, en se ramenant à la formule $\det \begin{pmatrix} A_1 & 0 \\ 0 & A_2 \end{pmatrix} = (\det A_1) \cdot (\det A_2)$. □

11.2.6 Comatrice

DÉFINITION 45
Soit $n \geq 2$. Soit $A = (a_{i,j})_{i,j} \in \mathscr{M}_n(\mathbb{K})$. On définit la matrice $\operatorname{com}(A) = (b_{i,j})_{i,j}$, avec $b_{i,j} = (-1)^{i+j}\Delta_{i,j}$.
Cette matrice $\operatorname{com}(A)$ est appelée **comatrice** .
C'est la matrice des cofacteurs de A ($b_{i,j}$ est le cofacteur (i,j) de A).

PROPOSITION 46
Soient $n \geq 2$ et $A = (a_{i,j})_{i,j} \in \mathscr{M}_n(\mathbb{K})$. Alors, on a :

$$A\, {}^t\operatorname{com}(A) = {}^t\operatorname{com}(A)\ A = (\det A)\, I_n.$$

Corollaire 47
Soient $n \geq 2$ et $A = (a_{i,j})_{i,j} \in \mathscr{M}_n(\mathbb{K})$. Si A est inversible, on a alors :

$$A^{-1} = \frac{1}{\det A}\, {}^t\mathrm{com}\,(A)\,.$$

Exemple 48 — *Si* $A = \begin{pmatrix} a & b \\ c & d \end{pmatrix}$, *on a* $\mathrm{com}\,(A) = \begin{pmatrix} d & -c \\ -b & a \end{pmatrix}$.
Si A est inversible, on a :

$$A^{-1} = \frac{1}{ad-bc}\begin{pmatrix} d & -b \\ -c & a. \end{pmatrix}$$

Remarques 49

- *À l'exception de ce cas des matrices* 2×2 *on utilise rarement la formule précédente pour inverser une matrice. En effet,* $\mathrm{com}\,(A)$ *est une matrice dont chacun des coefficients est un déterminant de taille* $(n-1) \times (n-1)$.
- *Par contre, la matrice* $\mathrm{com}\,(A)$ *est une matrice dont tous les coefficients sont des polynômes en les coefficients de A. Ainsi,* A^{-1} *est une matrice dont tous les coefficients sont un quotient de deux polynômes en les coefficients de A.*
  Cette formule peut être très utile lorsque la matrice A dépend d'un paramètre, pour étudier les propriétés de continuité, de dérivabilité... des coefficients de A^{-1}.
  Cette formule peut aussi être très utile en algèbre pour caractériser les coefficients de A^{-1} *par rapport à ceux de A. Comme les coefficients de* A^{-1} *sont obtenus à partir de polynômes à coefficients entiers en les coefficients de A, cela montre que l'expression de* A^{-1} *ne dépend pas du corps* $\mathbb{K}$ *choisi. Pour une matrice A à coefficients rationnels, A est dans* $\mathscr{M}_n(\mathbb{Q})$,$\mathscr{M}_n(\mathbb{R})$,$\mathscr{M}_n(\mathbb{C})$. *Peu importe le corps choisi, la valeur de* $\det(A)$ *reste la même, et si A est inversible alors* A^{-1} *sera à coefficients rationnels.*

11.3 Rappels sur la trace

Trace d'une matrice

Définition 50
Soit $M = (m_{i,j})_{(i,j)\in[\![1,n]\!]^2} \in \mathscr{M}_n(\mathbb{K})$. On appelle *trace* de la matrice M le nombre :

$$\mathrm{Tr}\,(M) = \sum_{i=1}^{n} m_{i,i}.$$

Proposition 51
Soient $(A, B) \in \mathscr{M}_n(\mathbb{K})^2$. On a :

$$\mathrm{Tr}\,(AB) = \mathrm{Tr}\,(BA)\,.$$

Preuve — Posons $A = (a_{i,j})$ et $B = (b_{i,j})$. Alors :

$$\mathrm{Tr}\,(AB) = \sum_{i=1}^{n}\sum_{k=1}^{n} a_{i,k}b_{k,i} = \sum_{k=1}^{n}\sum_{i=1}^{n} b_{k,i}a_{i,k} = \mathrm{Tr}\,(BA)\,.$$

□

On remarque en particulier le fait que, pour toute matrice $M \in \mathscr{M}_n(\mathbb{K})$ et pour toute matrice inversible $P \in \mathrm{GL}_n(\mathbb{K})$, on a :

$$\mathrm{Tr}\left(P^{-1}MP\right) = \mathrm{Tr}\left(MPP^{-1}\right) = \mathrm{Tr}\,(M)\,.$$

Corollaire 52
Soit $n \geq 1$. Soit $\phi \in \mathcal{L}(\mathscr{M}_n(\mathbb{K}), \mathbb{K})$ une forme linéaire. Alors, il existe une unique matrice $A \in \mathscr{M}_n(\mathbb{K})$ telle que :

$$\phi : M \longmapsto \mathrm{Tr}\,(AM)\,.$$

De plus, toute forme linéaire $\phi \in \mathcal{L}(\mathscr{M}_n(\mathbb{K}), \mathbb{K})$ vérifiant

$$\phi(MN) = \phi(NM),\ \forall (M, N) \in \mathscr{M}_n(\mathbb{K})^2,$$

est de la forme :

$$\phi : M \longmapsto \alpha\,\mathrm{Tr}\,(M)$$

pour un unique $\alpha \in \mathbb{K}$.

Preuve — La famille $(E_{i,j})_{(i,j)\in[\![1,n]\!]^2}$ est la base canonique de $\mathscr{M}_n(\mathbb{K})$.

- Soit $\phi \in \mathcal{L}(\mathscr{M}_n(\mathbb{K}), \mathbb{K})$. Soit $M = (m_{i,j}) \in \mathscr{M}_n(\mathbb{K})$. On a :
$$\phi(M) = \sum_{(i,j)\in[\![1,n]\!]^2} m_{i,j}\phi(E_{i,j}).$$
La matrice $A = (a_{i,j})$ avec $a_{i,j} = \phi(E_{j,i})$ est donc l'unique matrice vérifiant $\phi(M) = \operatorname{Tr}(AM)$ pour tout M.
Réciproquement, la fonction $M \mapsto \operatorname{Tr}(AM) \in \mathbb{K}$ est bien une forme linéaire.
- Soit $\phi \in \mathcal{L}(\mathscr{M}_n(\mathbb{K}), \mathbb{K})$ vérifiant $\phi(MN) = \phi(NM)$ pour tout $(M, N) \in \mathscr{M}_n(\mathbb{K})^2$.
Pour tous $i, j, k, l \in \{1, \dots, n\}$, on a :
$$E_{i,j}E_{k,l} = \delta_{j,k}E_{i,l}$$
Ainsi pour tout (i, j) avec $i = j$, on obtient :
$$\phi(E_{i,i}) = \phi(E_{i,j}E_{j,i}) = \phi(E_{j,i}E_{i,j}) = \phi(E_{j,j}) = \phi(E_{1,1}).$$
Pour tout (i, j) avec $i \neq j$, on a :
$$\phi(E_{i,j}) = \phi(E_{i,i}E_{i,j}) = \phi(E_{i,j}E_{i,i}) = \phi(0) = 0.$$
On obtient donc :
$$\phi(M) = \sum_{i=1}^{n} m_{i,i}\phi(E_{1,1}) = \phi(E_{1,1}) \sum_{i=1}^{n} m_{i,i} = \phi(E_{1,1}) \operatorname{Tr}(M).$$

□

Trace d'un endomorphisme

Proposition-Définition 53
Soit E un $\mathbb{K}$-e.v. de dimension n. Soit $B = (e_1, \dots, e_n)$ une base de E. Soit $f : E \to E$ un endomorphisme.
On définit la **trace** de f, notée $\operatorname{Tr}(f)$, comme :
$$\operatorname{Tr}(f) = \operatorname{Tr}(\operatorname{Mat}_B(f)).$$

Cette définition ne dépend pas de la base B choisie.

Exemple 54 — *Soit E un $\mathbb{K}$-e.v. de dimension n. Soit p un projecteur sur E. Alors :*

$$\boxed{\operatorname{rg}(p) = \operatorname{Tr}(p).}$$

En effet, en prenant une base $(e_1, \dots, e_n)$ adaptée à la décomposition en somme directe
$$E = \operatorname{Im}(u) \oplus \operatorname{Ker}(u),$$
alors la matrice de p dans cette base est
$$\begin{pmatrix} I_r & O_{r,n-r} \\ O_{n-r,r} & O_{n-r} \end{pmatrix}.$$
D'où le résultat.

Proposition 55
Soient $\mathbb{K}$ un corps et $n \geq 2$.
Alors le déterminant $\det(.)$ et la trace $\operatorname{Tr}(.)$ sont des **invariants de similitude** sur $\mathscr{M}_n(\mathbb{K})$:
Pour tout $A \in \mathscr{M}_n(\mathbb{K})$, pour tout $M \in \mathscr{M}_n(\mathbb{K})$ inversible, on a :
$$\det(MAM^{-1}) = \det(A)$$
$$\operatorname{Tr}(MAM^{-1}) = \operatorname{Tr}(A).$$

Autrement dit, pour A et B deux matrices qui sont semblables ($B = MAM^{-1}$ pour un M inversible), alors A et B ont même trace et même déterminant.

Chapitre 12 Eléments propres d'un endomorphisme - Diagonalisation

Les notions de cette section concernent les endomorphismes. **Elles concernent aussi les matrices en considérant les endomorphismes canoniquement associés $X \mapsto AX$.** Nous allons prinipalement apprendre à chercher des sous-espaces stables à un endomorphisme afin de mieux le comprendre.

L'existence de sous-espaces vectoriels stables est très importante dans l'étude géométrique et algébrique d'un endomorphisme. On s'intéresse particulièrement aux sous-espaces vectoriels stables qui sont différents de $\{0\}$ et minimaux, ou aux noyaux des endomorphismes qui s'écrivent comme un "polynôme" en l'endomorphisme u.

Tous les objets que nous verrons seront définis sur des $\mathbb{K}$-espaces vectoriels E. Par contre, certains résultats ne seront vrais que pour des e.v. E de dimension finie.

Table des matières du chapitre

12.1 Valeurs propres, vecteurs propres & sous-espaces vectoriels propres

Définition 1
Soit E un $\mathbb{K}$-e.v. Soit $u : E \to E$ un endomorphisme. Soient $\lambda \in \mathbb{K}$ et $x \in E$, $x \neq 0$.

1. S'il existe $y \in E$, $y \neq 0$ tel que $u(y) = \lambda y$, on dit alors que λ est une **valeur propre** de u.
2. S'il existe $\gamma \in \mathbb{K}$ tel que $u(x) = \gamma x$, on dit que x est un **vecteur propre** de u.
3. On définit $\mathrm{Spec}_{\mathbb{K}}(u)$ l'ensemble des valeurs propres de u sur $\mathbb{K}$.
   Cet ensemble est appelé le **spectre** de u sur $\mathbb{K}$.
4. Pour $\lambda \in \mathrm{Spec}_{\mathbb{K}}(u)$, on définit $E_\lambda(u) = \mathrm{Ker}\,(u - \lambda Id_E)$ l'ensemble des vecteurs propres de u associés à la valeur propre λ (auquel on rajoute 0).
   Cet ensemble est appelé **le sous-espace vectoriel propre** de u pour la valeur propre λ.

Soient $n \in \mathbb{N}^*$, $A \in \mathscr{M}_n(\mathbb{K})$ une matrice carrée, et $X \in \mathscr{M}_{n,1}(\mathbb{K})$ un vecteur colonne.

1. S'il existe un vecteur colonne $Y \in \mathscr{M}_{n,1}(\mathbb{K})$, $Y \neq 0$ tel que $AY = \lambda Y$, on dit alors que λ est une **valeur propre** de A.
2. S'il existe $\gamma \in \mathbb{K}$ tel que $AX = \gamma X$, on dit que X est un **vecteur propre** de A.
3. On définit $\mathrm{Spec}_{\mathbb{K}}(A)$ l'ensemble des valeurs propres de A sur $\mathbb{K}$.
   Cet ensemble est appelé le **spectre** de A sur $\mathbb{K}$.
4. Pour $\lambda \in \mathrm{Spec}_{\mathbb{K}}(A)$, on définit $E_\lambda(A) = \mathrm{Ker}\,(A - \lambda I_n)$ l'ensemble des vecteurs propres de A associés à la valeur propre λ (auquel on rajoute 0).
   Cet ensemble est appelé **le sous-espace vectoriel propre** de A pour la valeur propre λ.

Remarque 2 — *Si E est de dimension n, en prenant B une base de E et en représentant le vecteur $x \in E$ par une vecteur colonne $X \in \mathscr{M}_{n,1}(\mathbb{K})$, on a*

$$u(x) = \lambda x \iff Mat_B(u)X = \lambda X.$$

Ainsi, x est un vecteur propre de u si, et seulement si, X est un vecteur propre de $A = Mat_B(u)$, pour la même valeur propre λ.

Si x est un vecteur propre de u, alors x est associé à une unique valeur propre λ telle que $u(x) = \lambda x$.
Si λ est une valeur propre de u, l'ensemble des vecteurs propres associés est l'ensemble des vecteurs non nuls de $\mathrm{Ker}\,(u - \lambda\,\mathrm{Id}_E)$. ◀

Exemples 3
Soit E un $\mathbb{K}$-e.v. et $u : E \to E$ un endomorphisme.

1. *$E_0(u) = \mathrm{Ker}\,(u)$ est le noyau de u.*
2. *$E_1(u) = \mathrm{Ker}\,(u - \mathrm{Id}_E) = \{x \in E, u(x) = x\}$ est le sous-espace des vecteurs de E invariants par u.*
3. *Pour $u = \lambda Id_E$, on a $Spec_{\mathbb{K}}(u) = \{\lambda\}$. Le sous-espace propre associé à λ est $E_\lambda(u) = E$.*
4. *Pour $E = C^\infty(\mathbb{R}, \mathbb{R})$ et $D : f \mapsto f'$ l'endomorphisme de dérivation, on étudie les valeurs propres de D en résolvant les équations différentielles :*

$$f' = \lambda f, \text{ pour un } \lambda \in \mathbb{R}.$$

La théorie des équations différentielles linéaires du premier ordre montre que tout $\lambda \in \mathbb{R}$ est une valeur propre de D et que le sous-espace vectoriel propre associé à λ est $E_\lambda(D) = Vect(x \mapsto e^{\lambda x})$. Cet endomorphisme a donc une infinité de valeurs propres, et chaque sous-espace propre est de dimension 1.

5. *Soient $E = \mathbb{R}^2$, $t \in \mathbb{R}$, et $A = \begin{pmatrix} \cos(t) & -\sin(t) \\ \sin(t) & \cos(t) \end{pmatrix}$. Alors :*

$$Spec_{\mathbb{R}}(A) = \begin{cases} \{1\} & si\ t = 2k\pi,\ k \in \mathbb{Z}, \\ \{-1\} & si\ t = \pi + 2k\pi,\ k \in \mathbb{Z}, \\ \emptyset & sinon. \end{cases}$$

En effet, on a $\det(A - \lambda I_2) = (\cos(t) - \lambda)^2 + \sin(t)^2$. Or, on a $\mathrm{Ker}\,(A - \lambda I_2) \neq \{0\}$ si et seulement si $A - \lambda I_2$ est non-inversible, ssi $\det(A - \lambda I_2) = 0$. C'est-à-dire si et seulement si $\sin(t) = 0$ et $\lambda = \cos(t)$. Lorsque $t = 2k\pi$ on a $A = I_2$, donc $E_1(A) = \mathbb{R}^2$. Lorsque $t = \pi + 2k\pi$ on a $A = -I_2$, donc $E_{-1}(A) = \mathbb{R}^2$. Par contre, pour $\mathbb{K} = \mathbb{C}$ et $E' = \mathbb{C}^2$, le spectre de la matrice A vaut :

$$Spec_{\mathbb{C}}(A) = \{e^{it}, e^{-it}\}.$$

En effet, sur $\mathbb{C}$ le polynôme $(\cos(t) - X)^2 + \sin(t)^2$ admet pour racines e^{it} et e^{-it}.
Le calcul montre que $(1, -i)$ est un vecteur propre de A pour e^{it}, et que $(1, i)$ est un vecteur propre de A pour e^{-it}.

Remarques 4

1. *Un endomorphisme d'un espace vectoriel non réduit à $\{0\}$ n'admet pas nécessairement de valeur propre ; c'est ce que nous avons vu en exemple pour les rotation d'angle $\theta \notin \pi\mathbb{Z}$ sur $\mathbb{R}^2$.*
2. *Pour A une matrice à coefficients réels, on peut avoir $Spec_{\mathbb{R}}(A) \neq Spec_{\mathbb{C}}(A)$. Les valeurs propres de A dépendent du corps $\mathbb{K}$ avec lequel on considère tous nos objets.*
3. *Si au contraire il n'y a pas d'ambiguité sur le corps $\mathbb{K}$, on notera parfois $\mathrm{Spec}\,(u)$ au lieu de $Spec_{\mathbb{K}}(u)$.*

Exemple 5 — *Voici les valeurs propres de quelques endomorphismes :*

1. *Pour une homothétie $h = \lambda\,\mathrm{Id}_E$, on a $\mathrm{Spec}\,(h) = \{\lambda\}$ et $E_\lambda(h) = E$;*
2. *Pour un projecteur p non-trivial ($p \neq 0$, $p \neq Id_E$), on a $\mathrm{Spec}\,(p) = \{0, 1\}$, $E_0(p) = \mathrm{Ker}\,(p)$ et $E_1(p) = \mathrm{Im}\,(p)$;*
3. *Pour une symétrie s non-triviale ($s \neq Id_E$,$s \neq -Id_E$), on a $\mathrm{Spec}\,(s) = \{-1, 1\}$, $E_{-1}(s) = \mathrm{Ker}\,(s + \mathrm{Id}_E)$ et $E_1(s) = \mathrm{Ker}\,(s - \mathrm{Id}_E)$.*

Proposition 6
Soit E un $\mathbb{K}$-e.v. de dimension finie. Soient $u : E \to E$ un endomorphisme et $\lambda \in \mathbb{K}$. Alors, on a :

$$\begin{aligned} \lambda \in \mathrm{Spec}\,(u) &\iff \mathrm{Ker}\,(u - \lambda\,\mathrm{Id}_E) \neq \{0_E\} \iff u - \lambda\,\mathrm{Id}_E \text{ non injective} \\ &\iff u - \lambda\,\mathrm{Id}_E \text{ non bijective} \iff \det(u - \lambda\,\mathrm{Id}_E) = 0 \end{aligned}$$

Preuve — En dimension finie, un endomorphisme est injectif ssi il est surjectif ssi il est bijectif. □

Remarque 7 — *Pour E un $\mathbb{K}$-e.v. de dimensioon finie, on a donc que u est bijectif si et seulement si $0 \notin \mathrm{Sp}(u) \iff \det(u) = 0$.*
De même, $A \in \mathscr{M}_n(\mathbb{K})$ est inversible si et seulement si $0 \notin \mathrm{Sp}(A) \iff \det A \neq 0$.

Remarque 8 — *Pour A une matrice à coefficients réels, on peut considérer que A appartient à $\mathscr{M}_n(\mathbb{R})$ ou à $\mathscr{M}_n(\mathbb{C})$. Or, pour $\lambda \in \mathbb{R}$ $\det(A-\lambda I_n)$ est un nombre qui ne dépend pas du fait que l'on ait pris $\mathbb{K}=\mathbb{R}$ ou $\mathbb{K}=\mathbb{C}$. C'est-à-dire, $A-\lambda I_n$ est inversible dans $\mathscr{M}_n(\mathbb{R})$ ssi $A-\lambda I_n$ est inversible dans $\mathscr{M}_n(\mathbb{C})$. On a donc :*

$$\mathrm{Spec}_{\mathbb{R}}(A) = \mathrm{Spec}_{\mathbb{C}}(A) \cap \mathbb{R}.$$

De même, pour $B \in \mathscr{M}_n(\mathbb{Q})$, on a :

$$Spec_{\mathbb{Q}}(B) = \mathrm{Spec}_{\mathbb{R}}(B) \cap \mathbb{Q} = \mathrm{Spec}_{\mathbb{C}}(B) \cap \mathbb{Q}.$$

Exemple 9 —

- *Pour $A = \left(\begin{smallmatrix} 0 & -1 \\ 1 & 0 \end{smallmatrix}\right)$ on a vu que $\mathrm{Spec}_{\mathbb{R}}(A) = \emptyset$ mais que $\mathrm{Spec}_{\mathbb{C}}(A) = \{-i, i\}$.*
- *Pour $A = \left(\begin{smallmatrix} 0 & 2 \\ 1 & 0 \end{smallmatrix}\right)$ on a $\det(A-\lambda I_2) = \lambda^2 - 2$. Ainsi, on a $Spec_{\mathbb{Q}}(A) = \emptyset$ et $\mathrm{Spec}_{\mathbb{R}}(A) = \mathrm{Spec}_{\mathbb{C}}(A) = \{-\sqrt{2}, \sqrt{2}\}$.*

Proposition 10
Soient E un $\mathbb{K}$-e.v., $u : E \to E$ un endomorphisme, $\phi : E \to E$ un isomorphisme.
Alors, on a $\mathrm{Spec}(u) = \mathrm{Spec}\left(\phi \circ u \circ \phi^{-1}\right)$.
De plus, pour tout $\lambda \in \mathrm{Spec}(u)$, on a $\dim(E_\lambda(u)) = \dim(E_\lambda(\phi u \phi^{-1}))$.
Soient $n \geq 1$, $A, P \in \mathscr{M}_n(\mathbb{K})$ avec P inversible.
Alors, on a $\mathrm{Spec}(A) = \mathrm{Spec}\left(PAP^{-1}\right)$.
De plus, pour tout $\lambda \in \mathrm{Spec}(A)$, on a $\dim(E_\lambda(A)) = \dim(E_\lambda(PAP^{-1}))$.
Le spectre d'un endomorphisme et la dimension des sous-espaces propres sont des **invariants de similitude**.

Preuve — Soit $\phi : E \to E$ un isomorphisme. Soient $x \in E$ non-nul et $\lambda \in \mathbb{K}$. On a :

$$\begin{aligned} u(x) = \lambda x &\iff u \circ \phi^{-1} \circ \phi(x) = \lambda x \\ &\iff (\phi \circ u \circ \phi^{-1})(\phi(x)) = \phi(\lambda x) = \lambda \phi(x). \end{aligned}$$

Comme ϕ est un isomorphisme, on a $\phi(x) \neq 0$ si et seulement si $x \neq 0$.
Donc, x est un vecteur propre pour u associé à λ si et seulement si $\phi(x)$ est un vecteur propre pour $\phi u \phi^{-1}$ associé à λ.

Ces endomorphismes ont donc les mêmes valeurs propres, et l'on a :

$$E_\lambda\left(\phi \circ u \circ \phi^{-1}\right) = \phi\left(E_\lambda(u)\right).$$

Leurs sous-espaces propres associés à λ sont donc isomorphes, ce qui implique qu'ils ont la même dimension. □

Remarque 11 — *Soit E un $\mathbb{K}$-e.v. de dimension n. Soit $u : E \to E$ un endomorphisme. La proposition précédente nous dit que pour n'importe quelle base B de E, la matrice $Mat_B(u)$ possède le même spectre que u, et que leurs sous-espaces propres associés à une valeur propre donnée sont de même dimension.*
On peut ainsi étudier le spectre et les espaces propres de u, ou ceux de $Mat_B(u)$. Ce résultat est très utile s'il existe une base B telle que $Mat_B(u)$ a une expression qui permet de calculer plus facilement $\det(Mat_B(u) - \lambda I_n)$.

12.2 Polynôme caractéristique

On rappelle que pour $\mathbb{K}$ un corps, l'ensemble $\mathbb{K}(X)$ est le corps des fractions rationnelles à coefficients dans $\mathbb{K}$. Cet ensemble est un corps, qui contient $\mathbb{K}[X]$.

Définition du polynôme caractéristique

Proposition-Définition 12
Soient $n \geq 1$ et $A \in \mathscr{M}_n(\mathbb{K})$.
Alors, le déterminant de la matrice $XI_n - A \in \mathscr{M}_n(\mathbb{K}(X))$, $\det(XI_n - A)$, est un polynôme. On le note $\chi_A(X)$.
Le polynôme $\chi_A(X)$ est appelé **polynôme caractéristique de A**.
$\chi_A(X)$ est un polynôme unitaire, de degré n, avec :

$$\chi_A(X) = X^n - (\mathrm{Tr}(A))X^{n-1} + \cdots + (-1)^n \det(A).$$

Proposition 13
Soient $n \geq 1$ et $A, P \in \mathscr{M}_n(\mathbb{K})$ avec P inversible.
Alors, on a $\chi_{PAP^{-1}}(X) = \chi_A(X)$.
Le polynôme caractéristique est un **invariant de similitude**.

Preuve — On a :

$$\det\left(XI_n - PAP^{-1}\right) = \det\left(XPI_nP^{-1} - PAP^{-1}\right) = \det\left(P\left(XI_n - A\right)P^{-1}\right) = \det(P)\det\left(XI_n - A\right)\det\left(P^{-1}\right) = \det\left(XI_n - A\right).$$

□

Proposition-Définition 14
Soit E un $\mathbb{K}$-e.v. de dimension n. Soit $u :\in \mathcal{L}(E)$ un endomorphisme.
Alors il existe un unique polynôme, noté $\chi_u(X)$, tel que pour toute base B de E on ait $\chi_u(X) = \chi_{\mathrm{Mat}_B(u)}(X)$.
Ce polynôme est appelé le **polynôme caractéristique de** u.
Le polynôme χ_u est unitaire, de degré n, avec :

$$\chi_u(X) = X^n - \left(\mathrm{Tr}\,(u)\right)X^{n-1} + \cdots + (-1)^n \det(u).$$

Pour tout $\lambda \in \mathbb{K}$, on a :

$$\chi_u(\lambda) = \det\left(\lambda\,\mathrm{Id}_E - u\right).$$

Exemples 15

1. *Le polynôme caractéristique de* $(\alpha) \in \mathscr{M}_1(\mathbb{K})$ *est* $X - \alpha$.
2. *Le polynôme caractéristique de* $A = \begin{pmatrix} \alpha & \beta \\ \gamma & \delta \end{pmatrix} \in \mathscr{M}_2(\mathbb{K})$ *est :*

$$\chi_A(X) = X^2 - (\alpha+\delta)X + (\alpha\delta - \beta\gamma) = X^2 - \mathrm{Tr}\,(A)\,X + \det(A).$$

3. *Le polynôme caractéristique de la matrice :*

$$A = \begin{pmatrix} \alpha & \beta & \gamma \\ \alpha' & \beta' & \gamma' \\ \alpha'' & \beta'' & \gamma'' \end{pmatrix}$$

de $\mathscr{M}_3(\mathbb{K})$ *est :* $X^3 - \mathrm{Tr}\,(A)\,X^2 + \left((\alpha\beta' - \alpha'\beta) + (\alpha\gamma'' - \alpha''\gamma) + (\beta'\gamma'' - \beta''\gamma')\right)X - \det(A)$.

4. *Pour* $A \in \mathscr{M}_n(\mathbb{C})$, *on a* $\chi_A = \overline{\chi_{\overline{A}}}$.
5. *Soit* E *un* $\mathbb{K}$*-e.v. de dimension* n. *Soit* $u : E \to E$ *un endomorphisme de rang* 1.
Dans une base B *adaptée au sous-espace vectoriel* $\mathrm{Im}\,(u)$, *la matrice de* u *est de la forme :*

$$Mat_B(u) = \begin{pmatrix} \alpha_{1,1} & \dots & \dots & \alpha_{1,n} \\ 0 & \dots & \dots & 0 \\ \vdots & & \vdots & \\ 0 & \dots & \dots & 0 \end{pmatrix}.$$

Son polynôme caractéristique est donc $X^{n-1}(X - \alpha_{1,1})$. *Comme* $\alpha_{1,1} = \mathrm{Tr}\,(Mat_B(u)) = \mathrm{Tr}\,(u)$, *on obtient :*

$$\chi_u(X) = X^{n-1}\left(X - \mathrm{Tr}\,(u)\right).$$

Remarque 16 — *Soit* $A \in \mathscr{M}_n(\mathbb{K})$.
Alors on a $\det(XI_n - {}^tA) = \det({}^tXI_n - A) = \det(XI_n - A)$. *Donc,* $\chi_{{}^tA}(X) = \chi_A(X)$.

Exemples 17 (Matrices triangulaires)

1. *Soit* A *une matrice triangulaire supérieure, de diagonale* $(\alpha_1, \dots, \alpha_n)$. *On a alors :*

$$\chi_A(X) = \begin{vmatrix} X - \alpha_1 & \dots & \dots & * \\ 0 & \ddots & & \vdots \\ \vdots & \ddots & \ddots & \vdots \\ 0 & \dots & 0 & X - \alpha_n \end{vmatrix} = \prod_{k=1}^{n}\left(X - \alpha_k\right).$$

2. *Pour* M *une matrice triangulaire supérieure par blocs de la forme :*

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

on a :

$$\chi_M(X) = \begin{vmatrix} XI_p - A & -C \\ 0 & XI_q - D \end{vmatrix} = \chi_A(X)\chi_D(X).$$

Ces résultats sont aussi vrais pour des matrices triangulaires inférieures.

PROPOSITION 18
Soit $n \geq 1$. Soient $n_1, \ldots, n_r \in \{1, \ldots, n\}$ tels que $n_1 + \ldots, n_r = n$. Soit $M \in \mathscr{M}_n(\mathbb{K})$ une matrice triangulaire supérieure par blocs :

$$M = \begin{pmatrix} A_1 & * & \ldots & * \\ 0 & A_2 & * & \vdots \\ \vdots & \ddots & \ddots & * \\ 0 & \ldots & 0 & A_r \end{pmatrix}, \text{ pour } A_i \in \mathscr{M}_{n_i}(\mathbb{K}).$$

Alors, on a $\chi_M(X) = \chi_{A_1}(X) \ldots \chi_{A_r}(X) = \Pi_{i=1}^r \chi_{A_i}(X)$.
Le résultat est aussi vrai si M est une matrice triangulaire inférieure par blocs.

Preuve — On démontre le résultat par récurrence sur $r \geq 2$, en utilisant le fait que la matrice M est aussi de la forme $M = \begin{pmatrix} A_1 & B \\ 0 & A' \end{pmatrix}$, où A' est une matrice triangulaire supérieure par blocs avec $r-1$ blocs diagonaux, et que $XI_n - M$ est aussi une matrice triangulaire supérieure par blocs, avec r blocs diagonaux. □

Polynôme caractéristique et valeurs propres

PROPOSITION 19
Soit E un $\mathbb{K}$-e.v. de dimension finie. Soit $u : E \to E$ un endomorphisme.
Alors, le spectre de u est égal à l'ensemble des racines du polynôme caractéristique $\chi_u(X)$.
Soit $n \geq 1$. Soit $A \in \mathscr{M}_n(\mathbb{K})$.
Alors, on a $\text{Spec}\,(A) = \{\lambda \in \mathbb{K},\ \chi_A(\lambda) = 0\}$.

Preuve — On a montré qu'en dimension finie, on a $\lambda \in \text{Spec}\,(u)$ si et seulement si $\det(u - \lambda Id_E) = 0$. Or, on a $\chi_u(\lambda) = \det(\lambda Id_E - u) = (-1)^n \det(u - \lambda Id_E)$.
La preuve est identique pour une matrice A. □

REMARQUE 20 — *Ainsi, un seul calcul de déterminant peut suffire pour déterminer toutes les valeurs propres d'une matrice A (ou d'un endomorphisme u), du moment que l'on arrive à factoriser le polynôme caractéristique $\chi_A(X)$.*

COROLLAIRE 21
Soit E un $\mathbb{K}$-e.v. de dimension finie. Soient $u : E \to E$ un endomorphisme et $A \in \mathscr{M}_n(\mathbb{K})$.
Alors u (ou A) possède au plus n valeurs propres distinctes.

- Si $\mathbb{K} = \mathbb{C}$, alors u (ou A) a au moins une valeur propre.
- Si $\mathbb{K} = \mathbb{R}$ et si n est impair, alors u (ou A) a au moins une valeur propre.

Preuve — Les valeurs propres de u sont les racines de $\chi_u(X)$. Comme ce polynôme est de degré n, il possède au plus n racines distinctes. □

Le corollaire précédent montre d'obtenir le polynôme caractéristique sous forme factorisée est très important en pratique. On factorise en général $\chi_A(X)$ en calculant calculant le déterminant $\det(XI_n - A)$ par opérations élémentaires afin de faire apparaître des facteurs communs dans les lignes ou les colonnes.

EXEMPLE 22 — *Soit*

$$A = \begin{pmatrix} 2 & 5 & -6 \\ 4 & 6 & -9 \\ 3 & 6 & -8 \end{pmatrix}.$$

Le polynôme caractéristique de A est :

$$\chi_A(X) = \begin{vmatrix} X-2 & -5 & 6 \\ -4 & X-6 & 9 \\ -3 & -6 & X+8 \end{vmatrix}.$$

La somme des coefficients des lignes du déterminant ci-dessus étant $X - 1$, l'opération $C_1 \leftarrow C_1 + C_2 + C_3$ donne :

$$\begin{vmatrix} X-2 & -5 & 6 \\ -4 & X-6 & 9 \\ -3 & -6 & X+8 \end{vmatrix} \underset{C_1 \leftarrow C_1 + C_2 + C_3}{=} (X-1) \begin{vmatrix} 1 & -5 & 6 \\ 1 & X-6 & 9 \\ 1 & -6 & X+8 \end{vmatrix} \underset{\substack{L2 \leftarrow L_2 - L_1 \\ L_3 \leftarrow L_3 - L_1}}{=} (X-1) \begin{vmatrix} 1 & -5 & 6 \\ 0 & X-1 & 3 \\ 0 & -1 & X+2 \end{vmatrix}.$$

On obtient donc :

$$\chi_A(X) = (X-1)((X-1)(X+2)+3) = (X-1)(X^2+X+1).$$

Le spectre de A est donc $\{1, j, j^2\}$ dans $\mathbb{C}$ et seulement $\{1\}$ dans $\mathbb{R}$.

Définition 23
Soit E un $\mathbb{K}$-e.v. de dimension finie. Soit $u : E \to E$ un endomorphisme. Soit $\lambda \in \mathbb{K}$ une valeur propre de u.
On définit $m_u(\lambda)$ la multiplicité du facteur $(X-\lambda)$ pour le polynôme caractéristique $\chi_u(X)$.
L'entier $m(\lambda)$ est appelé **multiplicité de la valeur propre** λ, pour u.
Pour $n \geq 1$, $A \in \mathscr{M}_n(\mathbb{K})$, et $\lambda \in \mathrm{Spec}\,(A)$, on définit de même $m_A(\lambda)$ la multiplicité du facteur $(X-\lambda)$ pour le polynôme caractéristique $\chi_A(X)$.

Exemple 24 —

- *Soit $A \in \mathscr{M}_5(\mathbb{K})$ une matrice triangulaire supérieure dont les coefficients diagonaux sont $(1,1,2,2,3)$.*
  On a alors $\chi_A(X) = (X-1)(X-1)(X-2)(X-2)(X-3)$.
  On a donc $\mathrm{Spec}\,(A) = \{1,2,3\}$, avec $m_A(1) = 2$,$m_A(2) = 2$,$m_A(3) = 1$.
- *Pour $C = I_2$, on a $\chi_C(X) = (X-1)^2$, donc $\mathrm{Spec}\,(C) = \{1\}$, avec $m_C(1) = 2$. On a aussi $E_1(B) = \mathrm{Ker}\,(B - I_2) = \mathbb{K}^2$, donc $\dim(E_1(C)) = 2$.*
- *Soit $B = \begin{pmatrix} 1 & 1 \\ 0 & 1 \end{pmatrix}$. On a $\chi_B(X) = (X-1)^2$, donc $\mathrm{Spec}\,(B) = \{1\}$, avec $m_B(1) = 2$. Par contre, on a $E_1(B) = \mathrm{Ker}\,(B - I_2) = Vect((1,0))$, donc $\dim(E_1(B)) = 1$.*
  On a ici $\dim(E_1(B)) \neq m(1)$. On peut de plus remarquer que B n'est pas semblable à I_2 car $\dim(E_1(B)) \neq \dim(E_1(I_2))$.

Proposition 25
Soient E un $\mathbb{K}$-e.v., $u : E \to E$ un endomorphisme, $\phi : E \to E$ un isomorphisme. Soit $\lambda \in \mathrm{Spec}\,(u)$.
Alors, on a $m_u(\lambda) = m_{\phi^{-1}u\phi}(\lambda)$. Soient $n \geq 1$ et $A, P \in \mathscr{M}_n(\mathbb{K})$ avec P inversible. Soit $\lambda \in \mathrm{Spec}\,(A)$.
Alors, on a $m_A(\lambda) = m_{P^{-1}AP}(\lambda)$.
La multiplicité des valeurs propres est un **invariant de similitude**.

Preuve — On a vu que u et $\phi^{-1}u\phi$ ont le même spectre et le même polynôme caractéristique, ce qui conclut. Même chose pour A et $P^{-1}AP$. □

Remarque 26 — *Soient $A, B \in \mathscr{M}_n(\mathbb{K})$ deux matrices carrées (respectivement $u, v : E \to E$ deux endomorphismes).*
Si A et B (resp. u et v) n'ont pas le même spectre ou pas le même polynôme caractéristique ou pas les mêmes multiplicités de valeurs propres ou pas les mêmes dimensions d'espaces propres, alors A et B (resp. u et v) ne sont pas semblables.
En effet, tous ces objets sont des invariants de similitudes.
Par contre, si deux matrices A et B qui ont le même spectre/poly. caractéristique/multiplicités des valeurs propres/dimensions des sous-espaces propres, on ne sait pas si A et B sont semblables ou non.

Exemple 27 (Exemple avec des matrices nilpotentes) — *Soient $A = \begin{pmatrix} 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 \\ 0 & 0 & 0 & 0 \end{pmatrix}$ et $B = \begin{pmatrix} 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \end{pmatrix}$.*
Les matrices A et B sont triangulaires supérieures par blocs, avec deux blocs diagonaux de taille 2×2. Le calcul donne alors :

$$\chi_A(X) = X^2 \times X^2 = X^4 \text{ et } \chi_B(X) = X^2 \times X^2 = X^4.$$

On a donc $\chi_A(X) = \chi_B(X)$. De plus, on a $\mathrm{Spec}\,(A) = \mathrm{Spec}\,(B) = \{0\}$, et $m_A(0) = 4 = m_B(0)$.
Le calcul donne : $E_0(A) = Vect(e_1, e_3)$ et $E_0(B) = Vect(e_1, e_2)$. On a donc $\dim(E_0(A)) = 2 = \dim(E_0(B))$.
Par contre, on a $A^2 = 0$ et $B^2 = E_{1,4} \neq 0$. Les matrices A et B ne sont donc pas semblables, car on aurait sinon $B^2 = (P^{-1}A)^2 = P^{-1}A^2P = 0$, ce qui n'est pas le cas.

Cet exemple montre qu'il existe des matrices qui ont le même spectre/poly. caractéristique/multiplicités des valeurs propres/dimensions des sous-espaces propres, mais qui ne sont pas semblables.

Remarque 28 — *Pour u un endomorphisme et $\lambda \in \mathrm{Spec}\,(u)$, attention à ne pas confondre $m_u(\lambda)$ la multiplicité de $(X-\lambda)$ dans $\chi_u(X)$ et $\dim(E_\lambda(u))$ la dimension du sous-espace propre associé à λ.*
Ces deux quantités ne sont en général pas égales.

12.3 Sous-espaces vectoriels stables par un endomorphisme

Définition

Définition 29
Soient E un $\mathbb{K}$-e.v. et $u : E \to E$ un endomorphisme. Soit F un sous-espace vectoriel de E.
Si l'on a $u(F) \subset F$, on dit alors que le sous-e.v. F est **stable par** u, ou u-stable.

Exemples 30

1. *Pour tout $u : E \to E$ endomorphisme, les sous-espaces vectoriels $\{0\}$ et E sont stables par u.*
2. *Il existe des endomorphismes u dont les seuls sous-e.v. stables sont $\{0\}$ et E. Par exemple, $R_{\frac{\pi}{2}} : \mathbb{R}^2 \to \mathbb{R}^2$ la rotation d'angle $\frac{\pi}{2}$ dans $\mathbb{R}^2$ ne possède aucun sous-espace stable non-trivial (différent de $\{0\}$ et $\mathbb{R}^2$).*
3. *Pour $u : x \mapsto \lambda x$ une homothétie de rapport λ, on vérifie que tout sous-ev F de E est stable par u.*
   On peut montrer que les seuls endomorphismes v de E tels que tout sous-e.v. de E est stable par v sont les homothéties. ◀
4. *Soit $u : E \to E$ un endomorphisme. Pour $F \subset \mathrm{Ker}\,(u)$, alors F est stable par u (car $u(F) = \{0\}$). Pour $G \supset Im(u)$, alors G est stable par u (car $u(G) \subset Im(u)$).*
5. *Soit $u : E \to E$ un endomorphisme. Pour F, G deux sous-ev stables par u, alors $F \cap G$ et $F + G$ sont des sous-ev stables par u.* ◀

Endomorphisme induit par stabilité

Définition 31
Soient E un $\mathbb{K}$-e.v., $u : E \to E$ un endomorphisme, et F un sous-e.v. de E.
Si F est stable par u, on peut alors définir la fonction $u_F : x \in F \mapsto u(x) \in F$.
Cete fonction est un endomorphisme, appelé **l'endomorphisme induit par** u **sur** F.

Remarque 32 — *Le fait que $u(F) \subset F$ permet de remplacer l'espace d'arrivée E par son sous-espace vectoriel F.*
Attention à ne pas confondre la restriction $u_{|F} : F \to E$ de u à F, qui est une application linéaire de F vers E que l'on peut définir pour tout u, et l'endomorphisme $u_F : F \to F$ induit par u sur F, qui est un endomorphisme sur F que l'on ne peut définir que lorsque F est stable par u.
On notera que l'image de u_F est $Im(u_F) = u(F)$ et que le noyau de u_F est $\mathrm{Ker}\,(u_F) = \mathrm{Ker}\,(u) \cap F$. ◀

Proposition 33
Soient E un $\mathbb{K}$-e.v., $u : E \to E$ un endomorphisme, et F un sous-e.v. de E. Si F est stable par u, alors les valeurs propres de l'endomorphisme u_F induit par u sur F sont les valeurs propres de u telles que $E_\lambda(u) \cap F \neq \{0\}$.
On a alors :

$$E_\lambda(u_F) = E_\lambda(u) \cap F.$$

Preuve — Soit F est stable par u. Alors pour tout $\lambda \in \mathbb{K}$, F est stable par $u - \lambda Id_E$. On a $(u - \lambda Id_E)_F = u_F - \lambda Id_F$.
Avec la remarque précédente, on obtient : $\mathrm{Ker}\,(u_F - \lambda Id_F) = \mathrm{Ker}\,(u - \lambda Id_F) \cap F$. Ce qui permet de conclure. □

À l'aide des vecteurs propres d'un endomorphisme u, il est facile de construire certains sous-espaces stables par u.

Proposition 34 (Sous-espaces stables engendrés par des vecteurs propres)
Soient E un $\mathbb{K}$-e.v. et $u : E \to E$ un endomorphisme. Soient $v_1, \ldots, v_k$ des vecteurs propres de u (pas forcément de même valeur propre).
Alors $F = \mathrm{Vect}\,(v_1, \ldots, v_k)$ est stable par u.

Preuve — Soit $x \in F$. Alors on peut écrire $x = \sum_{j=1}^{k} \alpha_j v_j$. Notons λ_j la valeur propre associée à v_j. On a alors :

$$u(x) = u\left(\sum_{j=1}^{k} \alpha_j v_j\right) = \sum_{j=1}^{k} \alpha_j u(v_j) = \sum_{j=1}^{k} (\alpha_j \lambda_j) v_j \in F.$$

□

Endomorphisme induit et polynôme caractéristique

Proposition 35 (Sous-espaces stables et représentation matricielle)
Soient E un $\mathbb{K}$-e.v. de dimension n et $u : E \to E$ un endomorphisme. Soit F un sous-espace de E. Soit

$\mathcal{B} = (e_1, \dots, e_n)$ une base de E telle que $F = \text{Vect}\,(e_1, \dots, e_p)$ pour un certain $p \in [\![1, n]\!]$.
Alors, le sous-e.v. F est stable par u si et seulement si la matrice $\text{Mat}_{\mathcal{B}}(u)$ est de la forme :

$$\text{Mat}_{\mathcal{B}}(u) = \begin{pmatrix} a_{1,1} & \cdots & * & * & \cdots & * \\ \vdots & & \vdots & \vdots & & \vdots \\ a_{p,1} & \cdots & a_{p,p} & a_{p,p+1} & \cdots & * \\ 0 & \cdots & 0 & a_{p+1,p+1} & \cdots & * \\ \vdots & & \vdots & \vdots & & \vdots \\ 0 & \cdots & 0 & a_{n,p+1} & \cdots & a_{n,n} \end{pmatrix}$$

Proposition 36 (Endomorphisme induit et polynôme caractéristique)
Soient E un $\mathbb{K}$-e.v. de dimension n et $u : E \to E$ un endomorphisme. Soit F un sous-espace de E stable par u. Alors, $\chi_{u_F}(X)$ divise $\chi_u(X)$, le polynôme caractéristique de u.

Preuve — Soit $\mathcal{B} = (e_1, \dots, e_n)$ une base de E telle que $F = \text{Vect}(e_1, \dots, e_p)$. La matrice de u dans $\mathcal{B}$ est aors de la forme :

$$\text{Mat}_{\mathcal{B}}(u) = \begin{pmatrix} A & B \\ 0_{n-p,p} & D \end{pmatrix}.$$

On a alors $A = \text{Mat}_{(e_1,\dots,e_p)}(u_F)$, par définition de l'endomorphisme induit u_F. Le polynôme caractéristique de u vérifie donc $\chi_u(X) = \chi_{\text{Mat}_{\mathcal{B}}(u)}(X) = \chi_A(X)\chi_D(X)$. Ce polynôme est donc divisible par $\chi_A(X)$, qui est le polynôme caractéristique de u_F, ce qui conclut. □

Ainsi, trouver des sous-espaces F stables par un endomorphisme u permet de trouver des diviseurs du polynôme caractéristque de u $\chi_u(X)$. Cela aide à factoriser $\chi_u(X)$.

Proposition 37
Soient E un $\mathbb{K}$-e.v. de dimension n, et $u : E \to E$ un endomorphisme. Soit $\lambda \in \text{Spec}\,(u)$. Alors, on a :

$$1 \leq \dim(E_\lambda(u)) \leq m_u(\lambda).$$

Preuve — Soit $\lambda \in \text{Spec}\,(u)$. Le sous-espace propre $F = E_\lambda(u)$ étant-non réduit à $\{0\}$, il est de dimension au moins 1. Ce sous-espace vectoriel est aussi stable par u.
On remarque que l'endomorphisme induit u_F vérifie : $u_F = \lambda Id_F$.

Ainsi, le polynôme caractéristique de u_F est $\chi_{u_F}(X) = (X - \lambda)^{\dim(E_\lambda(u))}$.
Comme ce polynôme divise $\chi_u(X)$, on obtient $\dim(E_\lambda(u)) \leqslant m(\lambda)$. □

Exemples 38

1. *Si* $\text{rg}(u) = r$ *on a* $\dim E_0(u) = \text{Ker}\,(u) = n - r$ *, donc* $\chi_u(X)$ *est divisible par* X^{n-r}. ◀
2. *La matrice :*

$$A = \begin{pmatrix} 0 & \dots & 0 & \alpha_1 \\ \vdots & & \vdots & \vdots \\ 0 & \dots & 0 & \alpha_{n-1} \\ \alpha_1 & \dots & \alpha_{n-1} & 0 \end{pmatrix}$$

est de rang égal à 2 *(ou* 0*). Donc, son polynôme charactéristique est de la forme :*

$$\chi_A(X) = X^{n-2}\left(X^2 + uX + v\right) = X^n + uX^{n-1} + vX^{n-2}.$$

On a de plus $u = -\,\text{Tr}\,(A) = 0$*. En développant le déterminant :*

$$\det(XiI_n - A) = \begin{vmatrix} X & \dots & 0 & -\alpha_1 \\ \vdots & & \vdots & \vdots \\ 0 & \dots & X & -\alpha_{n-1} \\ -\alpha_1 & \dots & -\alpha_{n-1} & X \end{vmatrix}$$

les termes non nuls de degré $n-2$ *sont obtenus pour les transpositions* $\tau = (i, n)$ *avec* $i < n$*. On obtient donc :*

$$v = -\alpha_1^2 - \cdots - \alpha_{n-1}^2.$$

Ainsi, on a $\chi_A(X) = X^{n-2}(X^2 - (\alpha_1^2 + \ldots + \alpha_n^2))$.

Sous-espaces propres et commutativité

Proposition 39
Soient E un $\mathbb{K}$-e.v. et $u, v : E \to E$ des endomorphismes.
Si u et v commutent ($u \circ v = v \circ u$), alors tout sous-espace propre de v, $E_\lambda(v)$, est stable par u.

Preuve — Soit $\lambda \in \text{Spec}\,(v)$. Soit $x \in E_\lambda(v)$. On a $v(x) = \lambda x$.
Cela donne :

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

donc $u(x) \in E_\lambda(v)$, ce qui conclut. □

Remarque 40 — *Attention ! Ce résultat est faux pour un sous-espace stable F de v quelconque. Par exemple, pour $E = \mathbb{K}^2$, $u(x, y) = (y, x)$, $v = Id_E$, le sous-ev $F = Vect((1, 0))$ est stable par v, mais il n'est pas du tout stable par u.*
Il faut bien remarquer dans la preuve de la proposition précédente que $F = E_\lambda(v)$ et que $v(x) = \lambda x$ sont des informations nécessaires.

12.3.1 Matrice compagnon

Proposition 41
Soient E un $\mathbb{K}$-e.v. de dimension n, $u : E \to E$ un endomorphisme. Soit $x \in E$ non-nul.
Soit p le plus grand entier tel que la famille $\mathcal{B} = (x, u(x), \cdots, u^p(x))$ est libre.
Alors, on a $\dim(\text{Vect}\,(\mathcal{B})) = p + 1$.
De plus, pour $a_0, \ldots, a_p \in \mathbb{K}$ tels que $u^{p+1}(x) = -a_0 x - a_1 u(x) - \ldots - a_p u^p(x)$, on a :

$$\text{Mat}_{\mathcal{B}}(u_{\text{Vect}\,(\mathcal{B})}) = \begin{pmatrix} 0 & \cdots & \cdots & \cdots & 0 & -a_0 \\ 1 & \ddots & & & \vdots & -a_1 \\ 0 & \ddots & \ddots & & \vdots & \vdots \\ \vdots & \ddots & \ddots & \ddots & \vdots & \vdots \\ 0 & \cdots & 0 & 1 & 0 & -a_{p-1} \\ 0 & \cdots & 0 & 0 & 1 & -a_p \end{pmatrix}.$$

Proposition-Définition 42
Soit $n \geq 1$. Soit $P(X) = X^n + \alpha_{n-1}X^{n-1} + \cdots \alpha_1 X + \alpha_0 \in \mathbb{K}[X]$ un polynôme unitaire.
On appelle **matrice compagnon de P** [1], la matrice :

$$C_{P(X)} = \begin{pmatrix} 0 & 0 & \ldots & \ldots & 0 & -\alpha_0 \\ 1 & 0 & \ldots & \ldots & 0 & -\alpha_1 \\ 0 & 1 & \ddots & & & \\ \vdots & \vdots & \ddots & \ddots & \vdots & \vdots \\ 0 & 0 & \ldots & 1 & 0 & -\alpha_{n-2} \\ 0 & 0 & \ldots & 0 & 1 & -\alpha_{n-1} \end{pmatrix} \in \mathscr{M}_n(\mathbb{K}).$$

Alors, on a :

$$\chi_{C_{P(X)}}(X) = P(X).$$

Le polynôme caractéristique de la matrice compagnon de P est le polynôme $P(X)$.

Preuve — Notons $L_1, \ldots, L_{n+1}$ les lignes de la matrice $C_{P(X)}$. Son polynôme caractéristique est :

$$\chi_{C_{P(X)}} = \begin{vmatrix} X & 0 & \ldots & \ldots & 0 & \alpha_0 \\ -1 & X & \ldots & \ldots & 0 & \alpha_1 \\ 0 & -1 & \ddots & & & \\ \vdots & \vdots & \ddots & \ddots & \vdots & \vdots \\ 0 & 0 & \ldots & -1 & X & \alpha_{n-2} \\ 0 & 0 & \ldots & 0 & -1 & X + \alpha_{n-1} \end{vmatrix}$$

1.

Avec l'opération $L_1 \leftarrow L_1 + XL_2 + \cdots + X^{n-1}L_n$, ce déterminant devient :

$$\chi_{C_{P(X)}} = \begin{vmatrix} 0 & 0 & \dots & \dots & 0 & P(X) \\ -1 & X & \dots & \dots & 0 & \alpha_1 \\ 0 & -1 & \ddots & & & \\ \vdots & \vdots & \ddots & \ddots & \vdots & \vdots \\ 0 & 0 & \dots & -1 & X & \alpha_{n-2} \\ 0 & 0 & \dots & 0 & -1 & X+\alpha_{n-1} \end{vmatrix}.$$

Le développement selon la première ligne donne : $\chi_{C_{P(X)}} = (-1)^{n+1}P(X)(-1)^{n-1} = P(X)$. □

12.3.2 Polynômes caractéristiques scindés

Rappelons que l'on dit qu'un polynôme $P(X) \in \mathbb{K}[X]$ est **scindé** sur $\mathbb{K}$ s'il peut s'écrire comme produit de facteurs du premier degré de $\mathbb{K}[X]$ et qu'il est scindé à racines simples sur $\mathbb{K}$ si de plus ses racines sont de multiplicité 1.

Exemples 43

1. *Soit E un $\mathbb{C}$-e.v. de dimension finie. Pour tout endomorphisme $u : E \to E$, le polynôme caractéristique $\chi_u(X)$ est scindé, puisque tout polynôme de $\mathbb{C}[X]$ est scindé.*
2. *Pour $E = \mathbb{R}^2$ et $R(\theta) = \begin{pmatrix} cos\theta & -sin\theta \\ sin\theta & cos\theta \end{pmatrix}$, avec $\theta \in \mathbb{R}$, son polynôme caractéristique est $\chi_{R(\theta)}(X) = X^2 - 2Xcos\theta + 1$. Si $\theta \not\equiv 0 \bmod \pi$, ce polynôme est irréductible dans $\mathbb{R}[X]$. Il n'est donc pas scindé. Par contre, $\chi_{R(\theta)}(X)$ est scindé à racines simples dans $\mathbb{C}[X]$.*

Proposition 44
Soit E un $\mathbb{K}$-e.v. de dimension n. Soit $u : E \to E$ un endomorphisme.
Si χ_u est scindé, alors pour $(\mu_1, \dots, \mu_n)$ les racines de χ_u comptées avec multiplicité, on a :

$$\mathrm{Tr}\,(u) = \mu_1 + \cdots + \mu_n \quad \text{et} \quad \det(u) = \mu_1 \dots \mu_n.$$

Preuve — On a $\chi_u(X) = X^n - \mathrm{Tr}\,(u)\,X^{n-1} + \ldots + (-1)^n \det(u) = (X-\mu_1)\dots(X-\mu_n) = X^n - (\mu_1 + \ldots + \mu_n)X^{n-1} + \ldots + (-1)^n\mu_1 \dots \mu_n$. □

Remarque 45 — *Pour $A = (\alpha_{i,j})_n$ une matrice triangulaire supérieure, son polynôme caractéristique est scindé et ses racines (comptées avec multiplicité) sont $(\alpha_{1,1}, \dots, \alpha_{n,n})$.*

Proposition 46
Soient u et v deux endomorphismes de E qui commutent i.e. $u \circ v = v \circ u$. Alors l'image et le noyau de u sont stables par v, et l'image et le noyau de v sont stables par u.

Preuve — Il suffit par symétrie de montrer que l'image et le noyau de u sont stables par v. Montrons que l'image de u est stable par v : soit $x \in \mathrm{Im}\,(u)$. Alors il existe $y \in E$ avec $x = u(y)$. Dès lors

$$v(x) = v(u(y)) = u(v(y)) \in \mathrm{Im}\,(u).$$

La stabilité de $\mathrm{Im}\,(u)$ par v est prouvée.

Montrons que le noyau de u est stable par v : soit $x \in \mathrm{Ker}\,(u)$. Alors

$$u(v(x)) = v(u(x)) = v(0) = 0.$$

Donc $v(x) \in \mathrm{Ker}\,(u)$, et la stabilité de $\mathrm{Ker}\,(u)$ par v est prouvée. □

En particulier, si u et v commutent, alors tout sous-espace propre de u est stable par v. En effet, comme u, v commutent, on a aussi, pour tout $\lambda \in K$, la relation

$$(u - \lambda\,\mathrm{Id}_E) \circ v = u \circ v - \lambda v = v \circ u - \lambda v = v \circ (u - \lambda\,\mathrm{Id}_E).$$

Donc le noyau de $u - \lambda\,\mathrm{Id}_E$ (à savoir l'espace propre de u correspondant à la valeur propre λ) est stable par v, grâce au théorème que nous venons de prouver.

Corollaire 47
Les sous-espaces vectoriels propres d'un endomorphisme u de E sont stables par tout endomorphisme v commutant avec u.

12.3.3 Sous-espaces propres et sommes directes

THÉORÈME 48
Soient E un $\mathbb{K}$-e.v. et $u : E \to E$ un endomorphisme. Soient $\lambda_1, \ldots, \lambda_p \in \mathrm{Spec}\,(u)$ des valeurs propres de u distinctes. Soient $x_1, \ldots, x_p \in E$ des vecteurs propres associés à $\lambda_1, \ldots, \lambda_p$.
Alors, la famille $(x_1, \ldots, x_p)$ est une famille libre.

Preuve — Démonstration par récurrence sur $p \geq 1$. □

Autrement dit, toute famille finie de vecteurs propres associés à des valeurs propres deux à deux distinctes est libre. On notera que cette propriété est valable en dimension infinie.

THÉORÈME 49
Soient E un $\mathbb{K}$-e.v. et $u : E \to E$ un endomorphisme. Soient $\lambda_1, \ldots, \lambda_p \in \mathrm{Spec}\,(u)$ des valeurs propres de u distinctes.
Alors, la famille des sous-espaces propes $(E_{\lambda_1}(u), \ldots, E_{\lambda_p}(u))$ est en somme directe.

Preuve — Soit $(x_1, \ldots, x_p) \in E_{\lambda_1}(u) \times \ldots \times E_{\lambda_p}(u)$ une famille de vecteurs telle que

$$\sum_{i=1}^{p} x_i = 0.$$

Supposons par l'absurde que les vecteurs $x_1, \ldots, x_p$ ne sont pas tous nuls. Quitte à réordonner les vecteurs, on peut supposer que les vecteurs $x_1, \ldots, x_r$ sont non-nuls, pour un $1 \leq r \leq p$, et que les vecteurs $x_{r+1}, \ldots, x_p$ sont nuls.
Les vecteurs $x_1, \ldots, x_r$ sont des vecteurs propres non-nuls, associés à des valeurs propres distinctes, pour lesquels on aurait $x_1 + \ldots + x_r = 0$.

D'après le théorème précédent, la famille $(x_1, \ldots, x_r)$ est libre, ce qui contredit la relation $x_1 + \ldots + x_r = 0$.

Donc tous les vecteurs $x_1, \ldots, x_p$ sont nuls. Ainsi, les espaces propres $E_{\lambda_1}(u), \ldots, E_{\lambda_p}(u)$ sont en somme directe. □

EXEMPLE 50 — *Soient $E = \mathcal{C}^\infty(\mathbb{R}, \mathbb{C})$ et $D : f \in E \mapsto f' \in E$ l'application linéaire de dérivation. Pour tout $\lambda \in \mathbb{C}$, la fonction $x \in \mathbb{R} \longmapsto e^{\lambda x}$ est un vecteur propre de D pour la valeur propre λ. La famille $\left(x \in \mathbb{R} \longmapsto e^{\lambda x} \in \mathbb{C}\right)_{\lambda \in \mathbb{C}}$ est donc libre, d'après les résultats précédents (toute combinaison linéaire d'un nombre fini de ces fonctions qui est nulle est forcément la combinaison linéaire nulle). Comme cette famille est infinie, on en déduit que l'espace vectoriel $\mathcal{C}^\infty(\mathbb{R}, \mathbb{C})$ est de dimension infinie (ce résultat a déjà été démontré de plusieurs manières dans les cours précédents).*

12.4 DIAGONALISABILITÉ

Définition et caractérisations élémentaires

DÉFINITION 51
Soit E un $\mathbb{K}$-e.v. de dimension n. Soit $u : E \to E$ un endomorphisme.
On dit que l'endomorphisme u est **diagonalisable** s'il existe une base B de E telle que $\mathrm{Mat}_B(E)$ est une matrice diagonale.
Une base B dans laquelle la matrice $\mathrm{Mat}_B(E)$ est diagonale est appelée une **base de diagonalisation** de u.
Soient $n \geq 1$ et $A \in \mathscr{M}_n(\mathbb{K})$.
On dit que la matrice A est **diagonalisable** s'il existe une matrice inversible P telle que $P^{-1}AP$ est une matrice diagonale.

REMARQUES 52

- *Pour $\mathcal{B} = (e_1, \ldots, e_n)$ une base de diagonalisation de u, les vecteurs $e_1, \ldots, e_n$ sont alors des vecteurs propres de u.*
- *Une matrice $A \in \mathscr{M}_n(\mathbb{K})$, est diagonalisable si et seulement si l'application linéaire $X \in \mathbb{K}^n \mapsto AX$ est diagonalisable.*
  En effet, un changement de base vers la base B correspond à la matrice PAP^{-1}, où P est la matrice de passage de la base canonique vers la base B

THÉORÈME 53
Soient E un $\mathbb{K}$-e.v. de dimension n et $u : E \to E$ un endomorphisme.
On a les équivalences suivantes :

i) u est diagonalisable ;
ii) Il existe une base $\mathcal{B}$ de E constituée de vecteurs propres de u ;
iii) E est la somme directe de sous-espaces propres de u :

$$E = \bigoplus_{\lambda \in \mathrm{Spec}(u)} E_\lambda(u) ;$$

iv) Pour $\mathrm{Spec}\,(u) = \{\lambda_1, \dots, \lambda_r\}$, on a :

$$\sum_{k=1}^{r} \dim\left(E_{\lambda_k}(u)\right) = n ;$$

v) Le polynôme caractéristique χ_u est scindé, et pour $\mathrm{Spec}\,(u) = \{\lambda_1, \dots, \lambda_r\}$, on a :

$$\forall k \in [\![1, r]\!], \dim\left(E_{\lambda_k}(u)\right) = m(\lambda_k).$$

Corollaire 54
Soient E un $\mathbb{K}$-e.v. de dimension n et $u : E \to E$ un endomorphisme.
Si le polynôme caractéristique χ_u est scindé à racines simples, alors l'endomorphisme u possède n valeurs propres distinctes et est diagonalisable.
Pour $\mathrm{Spec}\,(u) = \{\lambda_1, \dots, \lambda_n\}$ et pour $e_1, \dots, e_n$ des vecteurs propres de u associés aux valeurs propres $\lambda_1, \dots, \lambda_n$, la famille $\mathcal{B} = (e_1, \dots, e_n)$ est une base de E, et $\mathrm{Mat}_{\mathcal{B}}(u) = \mathrm{Diag}(\lambda_1, \dots, \lambda_n)$.

Exemple 55 — *Soit $k \in \mathbb{C}$. On définit la matrice de $\mathscr{M}_4(\mathbb{C})$:*

$$A_k = \begin{pmatrix} 0 & 1 & 0 & 0 \\ 1 & k & 1 & 1 \\ 0 & 1 & 0 & 0 \\ 0 & 1 & 0 & 0 \end{pmatrix}.$$

Cette matrice est de rang 2, et de trace k. Un calcul de déterminant nous donne $\chi_A(X) = X^2(X^2 - kX - 3)$. Le sous espace-propre associé à 0, qui est $\mathrm{Ker}\,(A)$, s'obtient en résolvant le système :

$$\begin{cases} y = 0 \\ x + ky + z + t = 0 \end{cases}$$

Il est de dimension 2, engendré par les vecteurs :

$$\begin{pmatrix} 1 \\ 0 \\ -1 \\ 0 \end{pmatrix} \quad \text{et} \quad \begin{pmatrix} 1 \\ 0 \\ 0 \\ -1 \end{pmatrix}.$$

On retrouve d'ailleurs que A est de rang 2.

Ainsi :

- *Si k vérifie $k^2 + 12 \neq 0$, le polnôme $X^2 - kX - 3$ a deux racines λ_1 et λ_2 distinctes non nulles. Les sous-espaces vectoriels propres associés étant de dimension non-nulle, la matrice A est donc diagonalisable.*
- *Si k est égal à $\pm 2i\sqrt{3}$, le polynôme $X^2 - kX - 3$ a une racine $\lambda = \frac{k}{2}$ de multiplicité 2. On détermine le sous-espace propre associé en résolvant :*

$$\begin{cases} -\lambda x + y = 0 \\ x + (k-\lambda)y + z + t = 0 \\ y - \lambda z = 0 \\ y - \lambda t = 0 \end{cases}.$$

Une résolution avec la méthode du Pivot montre que ce sous-espace propre est de dimension 1, engendré par

$$\begin{pmatrix} 1 \\ \lambda \\ 1 \\ 1 \end{pmatrix}.$$

La dimension de la somme des sous-espaces propres de A vaut 3, donc la matrice A n'est pas diagonalisable dans ce cas.

Remarque 56 — *Soient E un e.v. de dimension et n $u : E \to E$ un endomorphisme. Si $\chi_u(X) = (X-\lambda)^n$, u n'a qu'une seule valeur propre λ.*
Alors u est diagonalisable si et seulement si $Ker(u - \lambda Id_E) = E$, si et seulement si $u = \lambda \operatorname{Id}_E$.
Par exemple, la matrice $A = \begin{pmatrix} 2 & a & b \\ 0 & 2 & c \\ 0 & 0 & 2 \end{pmatrix}$ est diagonalisable si et seulement si $a = b = c = 0$.

Méthode 57 (**Etude de la diagonalisabilité**) —
Pour déterminer si un endomorphisme u est diagonalisable, on procède en général de la façon suivante :

1. *On détermine χ_u le polynôme caractéristique de u ;*
2. *On factorise ce polynôme pour obtenir le spectre de u ;*
   Cela passe par un calcul de $\det(XI_n - Mat_B(u))$ via la méthode du Pivot pour trouver des facteurs de χ_u, par une recherche de vecteurs propres évidents pour trouver des facteurs $(X-\lambda)^k$ de χ_u, ou par une recherche de sous-espaces stables par u pour trouver des diviseurs de χ_u.
   Si le polynôme caractéristique χ_u n'est pas scindé, alors u n'est pas diagonalisable.
3. *On détermine les sous-espaces propres $E_\lambda(u)$ de u. (une base $\mathcal{B}_\lambda$ et leur dimension)*
4. *Conclure : Si la somme des dimensions des sous-espaces propres $E_\lambda(u)$ de u ne vaut pas n, alors u n'est pas diagonalisable.*
   Sinon, u est diagonalisable et $\mathcal{B} = \bigcup_{\lambda \in \mathrm{Spec}(u)} B_\lambda$ est une base de E qui diagonalise u.

Exemple 58 — *On veut étudier la matrice :*

$$A = \begin{pmatrix} 0 & 3 & 2 \\ -2 & 5 & 2 \\ 2 & -3 & 0 \end{pmatrix} \in \mathscr{M}_3(\mathbb{K}).$$

Le polynôme caractéristique de A vaut : (factorisation via la méthode du Pivot ou via des racines évidentes)

$$\chi_A(X) = X^3 - 5X^2 + 8X - 4 = (X-2)^2(X-1).$$

Les racines de χ_A comptées avec multiplicité sont $1, 2, 2$. Le sous-espace propre $E_1(A) = Ker(A - I_3)$ est alors de dimension 1. Trouvons-en une base. On résout l'équation $AX = X$, soit :

$$\begin{cases} -x + 3y + 2z &= 0 \\ -2x + 4y + 2z &= 0 \\ 2x - 3y - z &= 0 \end{cases}.$$

Ce système linéaire est équivalent à $x = y = -z$. Ainsi on a $E_1(A) = \mathbb{K}v_1$, pour

$$v_1 = \begin{pmatrix} 1 \\ 1 \\ -1 \end{pmatrix}.$$

Le sous-espace propre $E_2(A) = Ker(A2I_3)$ est de dimension 1 ou 2. Trouvons-en une base. On résout l'équation $AX - 2X = 0$, soit :

$$\begin{cases} -2x + 3y + 2z &= 0 \\ -2x + 3y + 2z &= 0 \\ 2x - 3y - 2z &= 0 \end{cases}.$$

Ce système linéaire est équivalent à $2x - 3y - 2z = 0$. Ainsi on a $E_2(A) = \mathbb{K}v_2 \oplus \mathbb{K}v_3$ où

$$v_2 = \begin{pmatrix} 3 \\ 2 \\ 0 \end{pmatrix} \quad et \quad v_3 = \begin{pmatrix} 1 \\ 0 \\ 1 \end{pmatrix}.$$

La matrice A est donc diagonalisable car la somme des dimensions de ses sous-espaces propres vaut 3. Une base de diagonalisation de A est $\mathcal{C} = (v_1, v_2, v_3)$. La matrice de passage de la base canonique à $\mathcal{C}$ est :

$$P = \begin{pmatrix} 1 & 3 & 1 \\ 1 & 2 & 0 \\ -1 & 0 & 1 \end{pmatrix}.$$

La matrice de l'endomorphisme $X \mapsto AX$ dans la base $\mathcal{C}$ est $P^{-1}AP = \operatorname{Diag}(1,2,2) = D$. On obtient alors $A = PDP^{-1}$, soit après calcul de P^{-1} :

$$A = \begin{pmatrix} 1 & 3 & 1 \\ 1 & 2 & 0 \\ -1 & 0 & 1 \end{pmatrix} \cdot \begin{pmatrix} 1 & 0 & 0 \\ 0 & 2 & 0 \\ 0 & 0 & 2 \end{pmatrix} \cdot \begin{pmatrix} 2 & -3 & -2 \\ -1 & 2 & 1 \\ 2 & -3 & -1 \end{pmatrix}.$$

12.4.1 Réduction des endomorphismes diagonalisables

Remarque 59 — *Soit E un $\mathbb{K}$-e.v. de dimension n. Soit $u : E \to E$ un endomorphisme. Supposons que u est diagonalisable. Soit* $\mathrm{Spec}\,(u) = \{\lambda_1, \ldots, \lambda_r\}$. *On a montré dans cette section que l'on a alors :*

- $\sum_{k=1}^{r} \dim(E_{\lambda_k}(u)) = n$;
- $E = E_{\lambda_1}(u) \oplus \ldots \oplus E_{\lambda_r}(u)$;
- $\dim(E_{\lambda_k}(u)) = m_u(\lambda_k)$, $\forall 1 \leq k \leq r$;
- $\chi_u(X) = \Pi_{i=1}^{r}(X - \lambda_k)^{\dim(E_{\lambda_k}(u))}$.
- *Pour $\mathcal{B}_1, \ldots, \mathcal{B}_r$ des bases de $E_{\lambda_1}(u), \ldots, E_{\lambda_r}(u)$, la famille $\mathcal{B} = \mathcal{B}_1 \cup \ldots \cup \mathcal{B}_r$ est une base de E.*
- *La matrice $Mat_{\mathcal{B}}(u)$ est une matrice diagonale.*

Pour tout $1 \leq k \leq r$, on a $u_{E_{\lambda_k}(u)} = \lambda_k Id_{E_{\lambda_1}(u)}$.
Ainsi, en notant $D_k = \lambda_k I_{\dim(E_{\lambda_k}(u))}$, la matrice $Mat_{\mathcal{B}}(u)$ est diagonale par blocs, avec

$$Mat_{\mathcal{B}}(u) = Diag(D_1, \ldots, D_r).$$

Pour $1 \leq k \leq r$, soit $p_k : E \to E$ la projection sur $E_{\lambda_k}(u)$ parallèlement à $\bigoplus_{j \neq k} E_{\lambda_j}(u)$. On a alors :

$$u = \lambda_1 p_i + \ldots + \lambda_r p_r,$$

car ces deux endomorphismes sont égaux sur la base $\mathcal{B}$.

Remarque 60 (Puissances d'un endomorphisme diagonalisable) — *Soit E un $\mathbb{K}$-e.v. de dimension n. Soit $u : E \to E$ un endomorphisme. Supposons que u est diagonalisable, et reprenons les notations de la remarque précédente. Soit $m \geq 0$.*
On a alors $u^m_{E_{\lambda_k}(u)} = \lambda_k^m Id_{E_{\lambda_1}(u)}$.
Ainsi, on a :

$$Mat_{\mathcal{B}}(u^m) = Diag(D_1^m, \ldots, D_r^m).$$

En utilisant les projections p_k, cela donne :

$$u^m = \lambda_1^m p_1 + \ldots + \lambda_r^m p_r,$$

car ces deux endomorphismes sont égaux sur la base $\mathcal{B}$.
Ainsi, l'endomorphisme u^m est totalement décrit à l'aide du spectre et de la base de diagonalisation de u.
En particulier, u^m est diagonalisable, et les racines de son polynôme caractéristique χ_{u^m} sont $\lambda_1^m, \ldots, \lambda_r^m$ (certaines racines pouvant être égales). C'est-à-dire :

$$\chi_{u^m}(X) = \Pi_{i=1}^{k}(X - \lambda_k^m)^{\dim(E_{\lambda_k}(u))}.$$

D'un point de vue matriciel, si $A \in \mathscr{M}_n(\mathbb{K})$ est diagonalisable, on a P une matrice inversible et $D = Diag(a_1, \ldots, a_n)$ une matrice diagonale telles que $A = PDP^{-1}$.
Pour tout $m \geq 0$, on a alors :

$$A^m = (PDP^{-1})^m = PD^mP^{-1} = PDiag(a_1^m, \ldots, a_n^m)P^{-1}.$$

On peut alors calculer facilement la matrice A^m à l'aide de deux produits matriciels.

Exemple 61 — *Reprenons le premier exemple de la page 127. Nous avons vu que la matrice*

$$A = \begin{pmatrix} 0 & 3 & 2 \\ -2 & 5 & 2 \\ 2 & -3 & 0 \end{pmatrix}$$

est diagonalisable. Elle s'écrit $A = P \cdot \mathrm{Diag}\,(1, 2, 2) \cdot P^{-1}$ avec :

$$P = \begin{pmatrix} 1 & 3 & 1 \\ 1 & 2 & 0 \\ -1 & 0 & 1 \end{pmatrix} \quad et \quad P^{-1} = \begin{pmatrix} 2 & -3 & -2 \\ -1 & 2 & 1 \\ 2 & -3 & -1 \end{pmatrix}.$$

Pour $k \in \mathbb{N}$, obtient alors les puissances de A par la relation $A^k = P \cdot \mathrm{Diag}\left(1^k, 2^k, 2^k\right) \cdot P^{-1}$. Cela donne :

$$A^k = 1^k P \cdot \mathrm{Diag}\,(1, 0, 0) \cdot P^{-1} + 2^k P \cdot \mathrm{Diag}\,(0, 1, 1) \cdot P^{-1} = \begin{pmatrix} 2 & -3 & -2 \\ 2 & -3 & -2 \\ -2 & 3 & 2 \end{pmatrix} + 2^k \begin{pmatrix} -1 & 3 & 2 \\ -2 & 4 & 2 \\ 2 & -3 & -1 \end{pmatrix}.$$

On remarque que les matrices devant 1^k et 2^k dans l'écriture ci-dessus correspondent aux matrices dans la base canonique des projections p_1 et p_2 associées à la somme directe de sous-espaces propres $\mathbb{K}^3 = E_1(A) \oplus E_2(A)$.

REMARQUE 62 — *Dans l'exemple précédent, pour M_1 et M_2 les matrices associées aux projections p_1 et p_2, on a :*

$$M_1 = \begin{pmatrix} 2 & -3 & -2 \\ 2 & -3 & -2 \\ -2 & 3 & 2 \end{pmatrix} = (2I_3 - A), \quad et\ M_2 = \begin{pmatrix} -1 & 3 & 2 \\ -2 & 4 & 2 \\ 2 & -3 & -1 \end{pmatrix} = (A - I_3).$$

Les polynômes $L_1(X) = (2 - X)$ et $L_2(X) = (X - 1)$ sont les polynômes d'interpolation associés à l'ensemble $\{1, 2\}$ ($L_1(1) = 1$, $L_1(2) = 0$, $L_2(2) = 2$, $L_2(1) = 0$).
Nous verrons dans le chapitre suivant que lorsque A est une matrice diagonalisable, les projections p_k associées aux sous-espaces propres de A peuvent être calculées avec des polynômes d'interpolation de Lagrange.
Ce calcul ne nécessite pas de déterminer une base de vecteurs propres (donc la matrice de passage P), ni de faire un inverse de matrice (calculer P^{-1}).

Chapitre 13 Polynômes d'endomorphismes

Table des matières du chapitre

13.1 Morphisme d'évaluation

Définition 1

Soient E un $\mathbb{K}$-e.v., $u \in \mathcal{L}(E)$ et $P = \sum_{k=0}^{n} a_k X^k \in \mathbb{K}[X]$.

- On appelle **évaluation** du polynôme P en u l'endomorphisme de E suivant :

$$P(u) = \sum_{k=0}^{n} a_k u^k = a_0 \operatorname{Id}_E + a_1 u + \cdots + a_n u^n,$$

où $u^k = \underbrace{u \circ u \circ \cdots \circ u}_{k \text{ fois}}$ est la composée k-ème de u, et $u^0 = \operatorname{Id}_E$.

La fonction $P \mapsto P(u)$ est appelée **morphisme d'évaluation** en u.

- Soit $A \in \mathscr{M}_n(\mathbb{K})$. On définit **l'évaluation** du polynôme P en A par la matrice :

$$P(A) = \sum_{k=0}^{n} a_k A^k = a_0 I_n + a_1 A + \cdots + a_n A^n.$$

La fonction $P \mapsto P(A)$ est appelée **morphisme d'évaluation** en A.

Remarque 2 — *Soit E un $\mathbb{K}$-e.v. de dimension n. Soient $u \in \mathcal{L}(E)$, et $\mathcal{B}$ une base de E. On a déjà vu que* $\operatorname{Mat}_{\mathcal{B}}(u^k) = \operatorname{Mat}_{\mathcal{B}}(u)^k$, *pour tout $k \geq 0$.*
Ainsi, pour tout polynôme $P \in \mathbb{K}[X]$, on a :

$$\operatorname{Mat}_{\mathcal{B}}(P(u)) = P\left(\operatorname{Mat}_{\mathcal{B}}(u)\right).$$

◀

Exemples 3

1. *Soit E un $\mathbb{K}$-e.v. et $u : E \to E$ un endomorphisme. Alors p est une projection si, et seulement si, $P(p) = 0$ pour $P = X^2 - X$.*
2. *Soit $\lambda \in \mathbb{K}$. Pour $u = \lambda \cdot id_E$ et $P \in \mathbb{K}[X]$, on trouve alors $P(u) = P(\lambda) \cdot id_E$.*
3. *Pour $A = \begin{pmatrix} 1 & 0 \\ 0 & 2 \end{pmatrix}$ et $P(X) = (X-1)(X-2) = X^2 - 3X + 2$, on a :*

$$P(A) = A^2 - 3A + 2 = \begin{pmatrix} 0 & 0 \\ 0 & 0 \end{pmatrix} = 0.$$

4. *Pour $E = \mathcal{C}^\infty(\mathbb{R}, \mathbb{C})$, on pose $u : f \in E \mapsto f' \in E$ l'endomorphisme de dérivation. Soit $P = \sum_{k=0}^{n} a_k X^k \in \mathbb{C}[X]$. Alors, l'endomorphisme $P(u)$ est l'opérateur différentiel sur $\mathcal{C}^\infty(\mathbb{R}, \mathbb{C})$ suivant :*

$$\begin{array}{rccc} P(u) : & \mathcal{C}^\infty(\mathbb{R}, \mathbb{C}) & \longrightarrow & \mathcal{C}^\infty(\mathbb{R}, \mathbb{C}) \\ & f & \longmapsto & a_0 f + a_1 f' + \cdots + a_n f^{(n)}. \end{array}$$

5. *L'évaluation de $P(X) = 3 - 2X + X^2$ en $A = \begin{pmatrix} 1 & -1 \\ 1 & 2 \end{pmatrix}$ est :*

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

Proposition 4
Soient E un $\mathbb{K}$-e.v. et $u : E \to E$ un endomorphisme. Soient $P, Q \in \mathbb{K}[X]$, $\lambda \in \mathbb{K}$. Alors on a :

1. $(\lambda P)(u) = \lambda P(u)$;
2. $(P + Q)(u) = P(u) + Q(u)$;
   Le morphisme d'évaluation en u, $P \in \mathbb{K}[X] \mapsto P(u) \in \mathcal{L}(E)$, est une application linéaire.
3. $(PQ)(u) = P(u) \circ Q(u)$;
   Le morphisme d'évaluation en u, $P \in \mathbb{K}[X] \mapsto P(u) \in \mathcal{L}(E)$, est un morphisme d'anneaux.

Soient $n \geq 1$ et $A \in \mathscr{M}_n(\mathbb{K})$.
Alors le morphisme d'évaluation en A, $P \in \mathbb{K}[X] \mapsto P(A) \in \mathscr{M}_n(\mathbb{K})$, est un morphisme de $\mathbb{K}$-algèbres.

Remarque 5 — *Pour $u : E \to E$ un endomorphisme, l'image du morphisme d'évaluation $P \mapsto P(u)$ est donc un sous-anneau de $\mathcal{L}(E)$ contenant u, et même une sous-algèbre de la $\mathbb{K}$-algèbre $(\mathcal{L}(E), +, \circ, .)$.*

Proposition 6
Soient E un $\mathbb{K}$-e.v. et $u : E \to E$ un endomorphisme.
On note $\mathbb{K}[u]$ l'ensemble $\{P(u),\ P \in \mathbb{K}[X]\}$ des polynômes en u. Alors :

- $\mathbb{K}[u]$ est une sous-algèbre de $\mathcal{L}(E)$, qui est commutative ;
- $\mathbb{K}[u]$ est la plus petite sous-algèbre de $\mathcal{L}(E)$ qui contient u.

Soient $n \geq 1$ et $A \in \mathscr{M}_n(\mathbb{K})$.
On note $\mathbb{K}[A]$ l'ensemble $\{P(A),\ P \in \mathbb{K}[X]\}$ des polynômes en A.
Alors, $\mathbb{K}[A]$ est une sous-algèbre commutative de $\mathscr{M}_n(\mathbb{K})$. On a $\mathbb{K}[A] = Vect(A^k,\ k \geq 0)$.
C'est la plus petite sous-algèbre de $\mathscr{M}_n(\mathbb{K})$ qui contient A.

Remarque 7 — *Nous avons vu à plusieurs reprises que les anneaux $\mathcal{L}(E)$ et $\mathscr{M}_n(\mathbb{K})$ ne sont pas commutatifs, ce qui empêche d'effectuer certaines opérations et rend faux certains résultats.*
Le fait que les sous-anneaux $\mathbb{K}[u]$ et $\mathbb{K}[A]$ soient commutatifs est très important.

Remarque 8 — *Soient E un $\mathbb{K}$-ev avec $\dim E \geq 2$ et $u : E \to E$ un endomorphisme. Alors le morphisme d'évaluation n'est pas surjectif car l'anneau $\mathcal{L}(E)$ n'est pas commutatif alors que $\mathbb{K}[u]$ oui. (pareil pour $\mathscr{M}_n(\mathbb{K})$ avec $n \geq 2$)*

Exemple 9 — *Soit $A = \begin{pmatrix} 1 & 1 \\ 0 & 1 \end{pmatrix} \in \mathscr{M}_2(\mathbb{R})$. La matrice $B = \begin{pmatrix} 0 & 1 \\ 0 & 0 \end{pmatrix}$ appartient à $\mathbb{R}[A]$ car on a :*

$$B = P(A) \text{ où } P = (X - 1).$$

Mais on a aussi :

$$B = Q(A) \text{ où } Q = (X - 1)^2 + (X - 1).$$

Proposition 10
Soient E un $\mathbb{K}$-e.v., et u un endomorphisme sur E. Soit ϕ un endomorphisme inversible sur E. Alors, on a :

$$P(\phi u \phi^{-1}) = \phi P(u) \phi^{-1},\ \forall P \in \mathbb{K}[X].$$

Soient $n \geq 1$ et $A, P \in \mathscr{M}_n(\mathbb{K})$ avec P inversible. Alors, on a :

$$P(PAP^{-1}) = PP(A)P^{-1},\ \forall P \in \mathbb{K}[X].$$

Si deux endomorphismes u, v (ou matrices A, B) sont semblables, alors $P(u)$ et $P(v)$ (ou $P(A)$ et $P(B)$) sont semblables.

Preuve — Soit $k \geq 1$. On a
$$(\phi u \phi^{-1})^k = (\phi u \phi^{-1})(\phi u \phi^{-1}) \dots (\phi u \phi^{-1}) = \phi u^k \phi^{-1}.$$
Ainsi, pour $P(X) = a_0 + \ldots + a_n X^n$, on a :
$$P(\phi u \phi^{-1}) = a_0 Id_E + a_1 \phi u \phi^{-1} + \ldots + a_n (\phi u \phi^{-1})^n = \phi(a_0 Id_E + a_1 u + \ldots + a_n u^n)\phi^{-1} = \phi P(u) \phi^{-1}.$$
La preuve est identique pour les matrices. □

Proposition 11
Soient $n \geq 1$ et $A \in \mathscr{M}_n(\mathbb{K})$. Alors :
$$P({}^tA) = {}^tP(A), \forall P \in \mathbb{K}[X].$$

Proposition 12 (Polynômes de matrices triangulaires)
Soit $n \geq 1$. Soit $A \in \mathscr{M}_n(\mathbb{K})$ une matrice triangulaire supérieure par blocs :
$$A = \begin{pmatrix} A_1 & * & \dots & & * \\ 0 & A_2 & * & & \vdots \\ 0 & \ddots & \ddots & & \vdots \\ 0 & \dots & 0 & A_{r-1} & * \\ 0 & \dots & & 0 & A_r \end{pmatrix}, \ A_i \in \mathscr{M}_{n_i}(\mathbb{K}), \ n_1 + \ldots + n_r = n.$$

Alors, pour tout $P \in \mathbb{K}[X]$, $P(A)$ est une matrice triangulaire supérieure par blocs :
$$P(A) = \begin{pmatrix} P(A_1) & * & \dots & & * \\ 0 & P(A_2) & * & & \vdots \\ 0 & \ddots & \ddots & & \vdots \\ 0 & \dots & 0 & P(A_{r-1}) & * \\ 0 & \dots & & 0 & P(A_r) \end{pmatrix}$$

Corollaire 13
Soient E un $\mathbb{K}$-e.v. de dimension n et $u : E \to E$ un endomorphisme.
Si u est diagonalisable, alors $P(u)$ est diagonalisable pour tout $P \in \mathbb{K}[X]$.

Preuve — Si u est diagonalisable, il existe une base $\mathcal{B}$ de E telle que $\mathrm{Mat}_{\mathcal{B}}(u)$ soit diagonale.

Soit $P \in \mathbb{K}[X]$. D'après la proposition précédente, la matrice $\mathrm{Mat}_{\mathcal{B}}(P(u)) = P(\mathrm{Mat}_{\mathcal{B}}(u))$ est elle aussi diagonale, donc $P(u)$ est diagonalisable. □

Les deux propositions suivantes vont se révéler par la suite particulièrement utiles.

Proposition 14
Soient E un $\mathbb{K}$-e.v. et $u \in \mathcal{L}(E)$. Soit F un sous-espace stable par u.
Alors, pour tout $P \in \mathbb{K}[X]$, le sous-espace F est stable par $P(u)$. On a de plus :
$$P(u)_F = P(u_F).$$

Proposition 15
Soient E un $\mathbb{K}$-e.v., $u \in \mathcal{L}(E)$, et $P \in \mathbb{K}[X]$.
Alors, les sous-espaces vectoriels $\mathrm{Im}\,(P(u))$ et $\mathrm{Ker}\,(P(u))$ sont stables par u.

Preuve — Les sous-ev $\mathrm{Im}\,(P(u))$ et $\mathrm{Ker}\,(P(u))$ sont stables par $P(u)$. Comme u et $P(u)$ commutent, ces sous-ev sont stables par u. □

13.2 Idéal annulateur et Polynôme minimal

Polynômes annulateurs

Définition 16
Soient E un $\mathbb{K}$-e.v., $u \in \mathcal{L}(E)$, et $P \in \mathbb{K}[X]$.
On dit que P est un **polynôme annulateur de** u si $P(u) = 0$.
Soient $n \geq 1$ et $A \in \mathscr{M}_n(\mathbb{K})$.
On dit que P est un **polynôme annulateur de** A si $P(A) = 0$.

Remarque 17 — *Pour $e_u : P \in \mathbb{K}[X] \mapsto P(u) \in \mathcal{L}(X)$, l'ensemble des polynômes annulateurs de u est exactement $Ker(e_u)$.*
Comme e_u est une application linéaire et un morphisme d'anneaux, ensemble est ainsi un sous-ev de $\mathbb{K}[X]$ et un idéal de $\mathbb{K}[X]$.

On rappelle que les idéaux de $\mathbb{K}[X]$ sont de la forme $M.\mathbb{K}[X]$, pour un $M \in \mathbb{K}[X]$ (ces idéaux sont principaux). De plus :

- *Si $M = 0$ on a $M\mathbb{K}[X] = \{0\}$;*
- *Si $M \neq 0$, alors il existe un unique polynôme unitaire N tel que $N\mathbb{K}[X] = M\mathbb{K}[X]$.*

Exemples 18

1. *Pour tout endomorphisme u, le polynôme nul $P = 0$ est un polynôme annulateur de u.*
2. *Le polynôme constant $P = 1$ n'est le polynôme annulateur d'aucun endomorphisme u.*
3. *Pour p une projection, p est annulée par le polynôme $X^2 - X$ (car $p^2 = p$).*
4. *Pour s une symétrie, s est annulée par le polynôme $X^2 - 1$ (car $s^2 = Id_E$).*
5. *Pour $u = \lambda Id_E$, $P(X) = (X - \lambda)$ est un polynôme annulateur de u.*
6. *Soit $A \in \mathscr{M}_n(\mathbb{K})$ une matrice diagonalisable. On a $P \in \mathscr{M}_n(\mathbb{K})$ inversible et $\lambda_1, \ldots, \lambda_n \in \mathbb{K}$ tels que $A = PDiag(\lambda_1, \ldots, \lambda_n)P^{-1}$.*
   Posons $Q(X) = (X - \lambda_1) \ldots (X - \lambda_n)$. On a alors :

$$\begin{aligned} Q(A) &= Q(PDiag(\lambda_1, \ldots, \lambda_n)P^{-1}) = PQ(Diag(\lambda_1, \ldots, \lambda_n))P^{-1} \\ &= PDiag(Q(\lambda_1), \ldots, Q(\lambda_n))P^{-1} = PDiag(0, \ldots, 0)P^{-1} \\ &= 0, \end{aligned}$$

   donc Q est un polynôme annulateur de A.
   D'après les résultats du chapitre précédent sur les matrices diagonalsables, on remarque que $Q(X) = \pi_{i=1}^{n}(X - \lambda_i) = \chi_A(X)$. Donc χ_A est un polynôme annulateur de A. Nous reviendrons sur ce résultat dans ce chapitre (Théorème de Cayley-Hamilton).
7. *Dans $E = \mathcal{C}^\infty(\mathbb{R}, \mathbb{C})$, l'endomorphisme de dérivation $D : f \mapsto f'$ ne possède pas de polynôme annulateur non nul. Soit $P \in \mathbb{C}[X]$. Pour tout $\lambda \in \mathbb{C}$, posons $f_\lambda : x \mapsto e^{\lambda x}$.*
   Alors on a $D(f_\lambda) = \lambda f_\lambda$, donc $P(D)(f_\lambda) = P(\lambda)f_\lambda$.
   Si l'on a $P(D) = 0$, alors on a $P(\lambda) = 0$ pour tout $\lambda \in \mathbb{C}$, ce qui implique que $P(X) = 0$.

13.2.1 Polynôme minimal, cas de la dimension finie

Proposition 19
Soit E un $\mathbb{K}$-e.v. de dimension de n. Soit $u \in \mathcal{L}(E)$.
Alors il existe $P \in \mathbb{K}[X]$ non-nul tel que $P(u) = 0$.
Soit $A \in \mathscr{M}_n(\mathbb{K})$.
Alors la matrice A possède un polynôme annulateur non-nul.

Définition 20
Soit E un $\mathbb{K}$-e.v. Soit $u : E \to E$ un endomorphisme qui possède un polynôme annulateur non-nul. Pour $e_u : P \in \mathbb{K}[X] \mapsto P(u) \in \mathcal{L}(E)$, le noyau de ce morphisme n'est pas réduit à $\{0\}$.
Il existe donc un unique polynôme $M \in \mathbb{K}[X]$, unitaire, tel que $\mathrm{Ker}(e_u) = M\mathbb{K}[X]$.
Ce polynôme est appelé **polynôme minimal** de u . On le note μ_u, ou M_u.

Remarque 21 — *Si E est de dimension finie, alors tout endomorphisme u sur E possède un polynôme minimal. Pour u un endomorphisme possédant un polynôme minimal, on a :*

$$P(u) = 0 \text{ si et seulement si } P \mid \mu_u.$$

Le polynôme minimal de u, μ_u, est le polynôme unitaire annulant de u de plus petit degré.
*Pour $A \in \mathscr{M}_n(\mathbb{K})$, on définit de même le **polynôme minimal** de A que l'on note μ_A.*

La question qui se pose alors est : comment calculer le polynôme annulateur d'un endomorphisme ou d'une matrice en dimension finie ?
Ce problème est très souvent difficile. Il existe cependant une méthode générale qui fonctionne, mais elle est trop longue en général pour être utilisée d'un point de vue algorithmique :

1. On commence par calculer un polynôme annulateur de u ou de A. Nous verrons dans le chapitre suivant comment trouver un tel polynôme dans le cas de la dimension finie (théorème de Cayley-Hamilton).
2. Pour P un polynôme annulateur de u (ou A), on factorise P dans $\mathbb{K}[X]$.
3. Parmi tous les diviseurs de P, on cherche ceux qui annulent u (ou A) et qui sont de plus petit degré possible.

PROPOSITION 22
Soit E un $\mathbb{K}$-e.v. de dimension n. Soient $u, v \in \mathcal{L}(E)$ avec v inversible. Soit $\mathcal{B}$ une base de E. On a :

1. $$\mu_{\mathrm{Mat}_{\mathcal{B}}(u)}(X) = \mu_u(X)\,;$$
2. $$\mu_{vuv^{-1}}(X) = \mu_u(X)\,;$$
3. Soient $A, P \in \mathscr{M}_n(\mathbb{K})$, avec P inversible. Alors :
$$\mu_{PAP^{-1}}(X) = \mu_A(X).$$
Le polynôme minimal est un **invariant de similitude**.
4. On a de plus :
$$\mu_{{}^tA}(X) = \mu_A(X).$$

EXEMPLES 23

1. *Pour $u = 0$ l'endomorphisme nul, on a $\mu_0(X) = X$.*
2. *On a $\mu_{Id_E}(X) = X - 1$.*
3. *Le polynôme minimal d'un endomorphisme est un polynôme de degré supérieur ou égal à 1.*
4. *Pour u un endomorphisme, on a $u = \lambda Id_E$ si et seulement si $\mu_u(X) = X - \lambda$.* ◀
5. *Soit p une projection. Si $p = 0$ on a $\mu_p(X) = X$. Si $p = Id_E$ on a $\mu_p(X) = X - 1$. Sinon, on a $\mu_p(X) = X^2 - X$.*
En effet, $X^2 - X$ est un polynôme annulateur de p, donc son polynôme minimal μ_p est un diviseur unitaire de $X^2 - X$, de degré au moins 1. Ce polynôme vaut donc $X^2 - X$ ou X ou $X - 1$. Comme on a supposé $p \neq 0$ et $p \neq Id_E$, on en déduit que $\mu_p(X) = X^2 - X$.
6. *De la même façon, si s est une symétrie différente de* id *et de* $-$id*, alors $\mu_s(X) = X^2 - 1$.*
7. *Pour u un endomorphisme nilpotent ($\exists k \geq 1$ tel que $u^k = 0$), on a $\mu_u(X) = X^r$, où r est l'indice de nilpotence de u. Réciproquement, si $\mu_u(X) = X^r$, alors u est nilpotent d'indice r.*

Nous terminons cette section par une propriété très utile que nous avons déjà vue pour le polynôme caractéristique χ_u.

PROPOSITION 24
Soient E un $\mathbb{K}$-e.v. et u un endomorphisme sur E. Soit F un sous-ev de E stable par u.
Si u admet un polynôme minimal, alors l'endomorphisme induit u_F possède un polynôme minimal, et l'on a :

$$\mu_{u_F} \text{ divise } \mu_u.$$

Preuve — Soit $P \in \mathbb{K}[X]$. On a vu que $P(u_F) = P(u)_F$. Ainsi, μ_u est un polynôme annulateur de u_F. Donc u_F possède un polynôme minimal, et celui-ci divise μ_u. □

EXEMPLE 25 — *Soit A la matrice diagonale par blocs suivante :*

$$A = \begin{pmatrix} 1 & 0 & 0 \\ 0 & 2 & 3 \\ 0 & 0 & 2 \end{pmatrix} \in \mathscr{M}_3(\mathbb{R}).$$

On associe A à l'endomorphisme $u : X \in \mathbb{R}^3 \mapsto AX \in \mathbb{R}^3$. Comme A est diagonale par blocs, on sait alors que les sous-ev $F_1 = \mathbb{R}e_1$ et $F_2 = \mathrm{Vect}(e_2, e_3)$ sont stables par u.

- *On a $u_{F_1} = Id_{F_1}$, donc $\mu_{u_{F_1}} = X - 1$.*

- *Sur F_2, on a :*

$$\mathrm{Mat}_{(e_2,e_3)}(u_{F_2}) = \begin{pmatrix} 2 & 3 \\ 0 & 2 \end{pmatrix}.$$

Comme

$$\left(\begin{pmatrix} 2 & 3 \\ 0 & 2 \end{pmatrix} - 2I_2\right)^2 = 0,$$

on en déduit que $(X-2)^2$ est un polynôme annulateur de u_{F_2}. Comme $(X-2)$ n'est pas un polynôme annulateur de u_{F_2}, on a alors $\mu_{u_{F_2}} = (X-2)^2$.

Ainsi le polynôme minimal de A est un multiple de $(X-1)$ et de $(X-2)^2$. Comme $(X-1)(X-2)^2$ est un polynôme annulateur de A, on en déduit que $\mu_A = (X-1)(X-2)^2$.

13.2.2 Calculs de polynômes d'endomorphismes ou de matrices

Proposition 26
Soient E un $\mathbb{K}$-e.v. et u un endomorphisme sur E.
Si u possède un polynôme annulateur P tel que $P(0) \neq 0$, alors u est inversible.
Pour $P(X) = a_0 + \ldots + a_n X^n$, on a $u^{-1} = \sum_{k=1}^{n} \frac{-a_k}{a_0} u^{k-1}$.

Preuve — On a :

$$0 = P(u) = a_0 Id_E + a_1 u + \ldots + a_n u^n = a_0 Id_E + \sum_{k=1}^{n} a_k u^k = a_0 Id_E + u(\sum_{k=1}^{n} a_k u^{k-1}).$$

Ainsi, on obtient :

$$Id_E = u(\sum_{k=1}^{n} \frac{-a_k}{a_0} u^{k-1}) = (\sum_{k=1}^{n} \frac{-a_k}{a_0} u^{k-1})u,$$

donc u est inversible, d'inverse $u^{-1} = \sum_{k=1}^{n} \frac{-a_k}{a_0} u^{k-1}$. □

Proposition 27
Soient E un $\mathbb{K}$-e.v. et u un endomorphisme sur E qui a un polynôme annulateur non-nul P. Soit $M \in \mathbb{K}[X]$.
Soit $R \in \mathbb{K}[X]$ le reste de la division euclidienne de M par P. On a alors :

$$M(u) = R(u).$$

Ainsi, tout polynôme en u est égal à un polynôme de degré au plus $\deg(\mu_u) - 1$ en u, que l'on peut déterminer à l'aide d'une division euclidienne.

Preuve — Soit $M = PQ + R$, $\deg(R) < \deg(P)$ la division euclidienne de M par P. On a alors :

$$M(u) = (PQ)(u) + R(u) = Q(u) \circ P(u) + R(u) = 0 + R(u),$$

ce qui conclut. □

Exemple 28 — *Soit $n \geqslant 2$. On considère la matrice :*

$$J = \begin{pmatrix} 0 & 1 & \ldots & 1 \\ 1 & 0 & (1) & \vdots \\ \vdots & (1) & \ddots & 1 \\ 1 & \ldots & 1 & 0 \end{pmatrix}.$$

- *La relation $(I_n + J)^2 = n(I_n + J)$ montre que l'on a :*

$$J^2 = (n-1)I_n + (n-2)J.$$

Ainsi le polynôme $P(X) = X^2 - (n-2)X - (n-1) = (X+1)(X-(n-1))$ est un polynôme annulateur de J. Vu que J n'est pas une matrice diagonale, elle ne peut être annulée par un polynôme de degré 1. On a donc $\mu_J(X) = X^2 - (n-2)X - (n-1) = (X+1)(X-(n-1))$.
Comme $\mu_J(0) \neq 0$, J est inversible, d'inverse :

$$J^{-1} = \frac{1}{n-1}\left(J - (n-2)I_n\right).$$

- *On calcule alors J^k, pour $k \in \mathbb{N}$, de la manière suivante :*

— *On effectue la division euclidienne de X^k par μ_J :*

$$X^k = \mu_J Q + \alpha_k X + \beta_k.$$

On obtient α_k et β_k en évaluant X^k en -1 et $n-1$. Cela donne :

$$\alpha_k = \frac{1}{n}\left((n-1)^k - (-1)^k\right) \quad \text{et} \quad \beta_k = (-1)^k + \frac{1}{n}\left((n-1)^k - (-1)^k\right).$$

— *On obtient ainsi :*

$$J^k = \alpha_k J + \beta_k I_n = (n-1)^k \frac{1}{n}(I_n + J) + (-1)^k \left(\frac{n-1}{n} I_n - \frac{1}{n} J\right) \quad \forall k \in \mathbb{N}^*.$$

Nous terminons notre étude sur le polynôme minimal par un exemple que l'on retrouve fréquemment, et qui se révèle très utile : celui de la *matrice compagnon*[1].

Proposition 29
Soit $n \geq 1$. Soit $P(X) = a_0 + a_1 X + \ldots + X^n$ un polynôme unitaire de degré n. Soit $C_P \in \mathscr{M}_{n+1}(\mathbb{K})$ la matrice compagnon de P.
Alors, on a $\mu_{C_P}(X) = P(X)$.

13.3 Polynômes d'endomorphismes et éléments propres

Les notions développées dans cette section concernent les endomorphismes. Elles s'appliquent aux matrices de $\mathscr{M}_n(\mathbb{K})$ en considérant les endomorphismes canoniquement associés sur $\mathbb{K}^n$.

Valeurs propres

Proposition 30
Soient E un $\mathbb{K}$-e.v., $u \in \mathcal{L}(E)$, $\lambda \in \mathrm{Spec}\,(u)$, et $P \in \mathbb{K}[X]$.
Pour tout $x \in E_\lambda(u)$, on a $P(u)(x) = P(\lambda)x$.
Donc, $P(\lambda) \in \mathrm{Spec}\,(P)\,(u)$.

Preuve — Comme $x \in E_\lambda(u)$, on a $u^k(x) = \lambda^k x$ pour tout k, d'où :

$$P(u)(x) = \left(\sum_{k=0}^{p} \alpha^k u^k\right)(x) = \sum_{k=0}^{p} \alpha^k u^k(x) = \left(\sum_{k=0}^{p} \alpha^k \lambda^k\right) x = P(\lambda)x.$$

En prenant $x \neq 0$, on obtient donc $P(u)(x) = P(\lambda)x$ avec $x \neq 0$, ce qui conclut. □

Corollaire 31
Soient E un $\mathbb{K}$-e.v. et $u \in \mathcal{L}(E)$. Soit P un polynôme annulateur de u.
Alors, les valeurs propres de u sont des racines de P.

Preuve — Les relations $P(u) = 0$ et $u(x) = \lambda x$ avec $x \neq 0$ entraînent $P(\lambda)x = 0$ et donc $P(\lambda) = 0$. □

Théorème 32
Soient E un $\mathbb{K}$-e.v. et $u \in \mathcal{L}(E)$ possédant un polynôme annulateur non-nul.
Alors, les valeurs propres de u sont exactement les racines de $\mu_u(X)$ dans $\mathbb{K}$.

Preuve —

- Comme $\mu_u(u) = 0$, les valeurs propres de u sont des racines de μ_u.
- Soit α une racine de $M\mu_u$. Ona donc $M_u(X) = (X-\alpha)N(X)$. Par minimalité de μ_u, l'endomorphisme $N(u)$ est non nul. Donc $\mathrm{Im}\,(N(u)) \neq 0$. Soit $x = N(u)(z) \in \mathrm{Im}\,(N(u))$ avec $x \neq 0$. On a alors :

$$(u - \alpha\,\mathrm{Id}_E)(x) = (u - \alpha\,\mathrm{Id}_E) \circ N(u)(z) = m u_u(u)(z) = 0,$$

donc α est une valeur propre de u.

□

1. Voir la proposition-définition 42.

13.3.1 Théorème de Cayley-Hamilton

Pour le reste de ce chapitre, les espaces E seront de dimension finie.

THÉORÈME 33 (Théorème de Cayley-Hamilton)
Soient E un $\mathbb{K}$-e.v. de dimension n et $u : E \to E$ un endomorphisme.
Alors, le polynôme caractéristique de u annule u :

$$\chi_u(u) = 0.$$

COROLLAIRE 34
Soient E un e.v. de dimension n et $u \in \mathcal{L}(E)$.
Alors $\mu_u(X)$ divise $\chi_u(X)$.
Le polynôme minimal de u divise son polynôme caractéristique.

13.3.2 Endomorphismes nilpotent

PROPOSITION 35
Soient E un $\mathbb{K}$-e.v. de dimension n et $f \in \mathcal{L}(E)$ un endomorphisme nilpotent.
Alors, on a : $\chi_f(X) = X^n$.

Preuve — Soit $q \in \mathbb{N}^*$ tel que $f^q = 0$. Nous donnons deux méthodes de démonstration de ce résultat.
Méthode No 1 : Passage aux complexes. Cette preuve est valable si $\mathbb{K} = \mathbb{C}$ ou si $\mathbb{K}$ est un sous-corps de $\mathbb{C}$.
Méthode No 2 : Récurrence sur la dimension de l'espace.

□

REMARQUE 36 — *Cette proposition permet de caractériser les endomorphismes nilpotents : f est nilpotentesi et seulement si son polynôme caractéristique vaut $X^{\dim(E)}$.*

PROPOSITION 37
Soient E un $\mathbb{K}$-e.v. de dimension finie, et $f \in \mathcal{L}(E)$ un endomorphisme nilpotent.
Alors, les nombres entiers suivants sont égaux :

- L'indice de nilpotence de f (le plus petit entier $k \geq 1$ tel que $f^k = 0$) ;
- Le degré du polynôme minimal de f.

13.3.3 Lemme des noyaux

THÉORÈME 38 (Lemme des noyaux)
Soient E un $\mathbb{K}$-e.v. et $u : E \to E$ un endomorphisme. Soient $P_1, ..., P_r\mathbb{K}[X]$ des polynômes premiers entre eux deux à deux, et soit $P = P_1 \dots P_r$. Alors, on a la décomposition en somme directe :

$$\operatorname{Ker}(P(u)) = \bigoplus_{k=1}^{r} \operatorname{Ker}(P_k(u)).$$

De plus, la projection p_k de $\operatorname{Ker}(P(u))$ sur $\operatorname{Ker}(P_k(u))$ parallèlement à $\bigoplus_{i\neq k} \operatorname{Ker}(P_i(u))$ est de la forme $U_k(u)_{\operatorname{Ker}(P(u))}$ pour un polynôme $U_k \in \mathbb{K}[X]$.

Le lemme des noyaux est un résultat très important en algèbre linéaire.

COROLLAIRE 39
Soient E un $\mathbb{K}$-e.v. et $u \in \mathcal{L}(E)$. Soit P un polynôme annulateur non-nul de u. Soient $P_1 \dots, P_r$ des polynômes premiers entre eux deux à deux tels que $P = P_1 \cdots P_r$.
Alors, on a :

$$E = \bigoplus_{k=1}^{r} \operatorname{Ker}(P_k(u)).$$

EXEMPLE 40 — *Pour* $A = \begin{pmatrix} 1 & 0 & 0 \\ 0 & 2 & 1 \\ 0 & 0 & 2 \end{pmatrix} \in \mathscr{M}_3(\mathbb{R})$, *on a vu que* $(X-1)(X-2)^2$ *est un polynôme annulateur de* A. *Ainsi, on a* $\mathbb{R}^3 = \operatorname{Ker}(A - I_3) \oplus \operatorname{Ker}\left((A - 2I_3)^2\right)$.

Remarque 41 — *Pour u un endomorphisme et $\lambda_1, \ldots, \lambda_r$ des valeurs propres de u distinctes, les polynômes $X-\lambda_1, \ldots, X-\lambda_r$ sont premiers entre eux deux à deux.*
On retrouve avec le lemme des noyaux le fait que $\mathrm{Ker}\,(u-\lambda_1 Id_E), \ldots, \mathrm{Ker}\,(u-\lambda_r Id_E)$ sont en somme directe.

Remarque 42 — *Soient u un endomorphisme et $P_1, \ldots, P_r$ des polynômes premiers entre eux deux à deux. Notons $Q_k = \prod_{i\neq k} P_i$. Alors les polynômes $Q_1, \ldots, Q_r$ sont premiers entre eux dans leur ensemble. D'après le théorème de Bezout généralisé, il existe donc $A_1, \ldots, A_r \in \mathbb{K}[X]$ tels que $Q_1A_1 + \ldots + Q_rA_r = 1$, de sorte que*

$$\mathrm{Id}_E = Q_1(u)\circ A_1(u) + \ldots + Q_r(u)\circ A_r(u).$$

En reprenant les éléments de la preuve du lemme des noyaux, pour $F = \mathrm{Ker}\,((P_1 \ldots P_r)(u))$ la projection p_k dans F sur $\mathrm{Ker}\,(P_k(u))$ parallèlement à $\bigoplus_{i\neq k}^r \mathrm{Ker}\,(P_i(u))$ est alors égale à : $p_k = Q_kA_k(u)_F$.

13.3.4 Synthèse sur la réduction

Faisons un bilan de certains résultats obtenus dans ce chapitre et dans le chapitre précédent.

Soit E un $\mathbb{K}$-e.v. E de dimension n. Soit $u \in \mathcal{L}(E)$ un endomorphisme.

§ *1. Polynôme caractéristique* — Le premier élément d'information que nous avons pour u est son *polynôme caractéristique* $\chi_u(X)$. On a :

$$\chi_u(X) = \prod_{k=1}^{r} (X-\lambda_k)^{m_u(\lambda_k)} \pi_{l=1}^{s} P_i^{\alpha_i},$$

avec

$$\mathrm{Spec}\,(u) = \{\lambda_1, \ldots, \lambda_r\}, \text{ et } P_1 \ldots, P_s \text{ irréductibles et de degré } > 1.$$

Pour pour toute valeur propre λ_k, l'entier $m_u(\lambda_k)$ est appelé la *multiplicité* de λ_k pour u.
Comme χ_u est unitaire et de degré n, on a $m_u(\lambda_1) + \ldots + m_u(\lambda_r) \leq n$.

§ *2. Polynôme minimal* — D'après le théorème de Caley-Hamilton, le polynôme minimal de u, μ_u, divise son polynôme caractéristique. Ainsi, on a $\deg(\mu_u) \leq \deg(\chi_u) \leq n$.
De plus, les valeurs propres de u sont les racines de μ_u dans $\mathbb{K}$.
On obtient donc :

$$\mu_u(X) = \prod_{k=1}^{r} (X-\lambda_k)^{r_u(\lambda_k)} \pi_{l=1}^{s} P_i^{\beta_i},$$

avec $1 \leqslant r_u(\lambda_k) \leqslant m_u(\lambda_k)$ pour tout $1 \leq k \leq r$ et $\beta_i \leq \alpha_i$ pour tout $1 \leq i \leq s$.
L'entier $r_u(\lambda_k)$ est appelé **l'indice de la valeur propre** λ_k. (c'est la multiplicité de $X-\lambda_k$ dans $\mu_u(X)$) λ_k pour u.

§ *3. Sous-espace propres* — Nous avons vu que les sous-espaces propres de u,

$$E_{\lambda_k}(u) = \mathrm{Ker}\,(u-\lambda_k \mathrm{Id}_E),\ 1 \leq k \leq r,$$

sont en somme directe. Leur dimension vérifie :

$$1 \leqslant \dim\,(E_{\lambda_k}(u)) \leqslant m_u(\lambda_k),\ \forall 1 \leq k \leq r.$$

Proposition 43
Soient E un $\mathbb{K}$-e.v. de dimension n, et $u : E \to E$ un endomorphisme avec $\mathrm{Spec}\,(u) = \{\lambda_1, \ldots, \lambda_r\}$.
Avec les notations précédentes, on a alors :

$$\dim\left(\mathrm{Ker}\left((u-\lambda_k \mathrm{Id}_E)^{r_u(\lambda_k)}\right)\right) = m_u(\lambda_k),\ \forall 1 \leq k \leq r$$

13.4 Diagnalisation et polynôme minimal

Les notions développées dans cette section concernent les endomorphismes. Elles s'appliquent aux matrices carrées de $\mathscr{M}_n(\mathbb{K})$ en considérant les endomorphismes associés sur $\mathbb{K}^n$.

Diagonalisation

Soient E un ev de dimension finie et $u \in \mathcal{L}(E)$ un endomorphisme. On a vu que u est diagonalisable si et seulement si $\chi_u(X)$ est scindé, et si la dimension du sous-espace propre $E_{\lambda_k}(u)$ vaut $m(\lambda_k)$ pour tout $k = 1, \dots, r$. Cela donne le résultat :

THÉORÈME 44
Soient E un ev de dimension finie et $u \in \mathcal{L}(E)$ un endomorphisme.
Alors, u est diagonalisable si et seulement si son polynôme minimal est scindé à racines simples.

Preuve — On pose $\mathrm{Spec}\,(u) = \{\lambda_1, \dots, \lambda_r\}$.

- Si u est diagonalisable, on a $E = \bigoplus_{i=1}^{r} \mathrm{Ker}\,(u - \lambda_i Id_E)$.
  Pour $P(X) = (X - \lambda_1) \dots (X - \lambda_r)$, le lemme des noyaux donne :
  $$E = \mathrm{Ker}\,(P(u))\,.$$
  C'est-à-dire, $P(u) = 0$. Donc le polynôme minimal de u, χ_u, divise P. Comme P est un polynôme scindé à racines simples, μ_u est alors un polynôme scindé à racines simples.
  Comme $\lambda_1, \dots, \lambda_r$ sont les racines de μ_u, on a même $\mu_u(X) = P(X)$.
- Réciproquement, on suppose que μ_u est scindé à racines simples. Comme les racines de μ_u sont les valeurs propres de u, on a donc $\mu_u(X) = (X - \lambda_1) \dots (X - \lambda_r)$. Le lemme des noyaux nous donne alors :
  $$E = \mathrm{Ker}\,(\mu_u(u)) = \bigoplus_{k=1}^{s} \mathrm{Ker}\,((u - \alpha_k \,\mathrm{Id}_E))\,.$$
  Donc E est la somme directe des sous-espaces propres de E. Cela veut dire que u est diagonalisable.

□

COROLLAIRE 45
Soient E un ev de dimension finie et $u \in \mathcal{L}(E)$ un endomorphisme.
Alors, u est diagonalisable si et seulement s'il est annulé par un polynôme scindé à racines simples.

COROLLAIRE 46
Soient E un ev de dimension finie et $u \in \mathcal{L}(E)$ un endomorphisme.
Soit F un sous-ev de E stable par u. Si u est diagonalisable, alors u_F est diagonalisable.

Preuve — On sait que u_F est annulé par μ_u, ce qui conclut. □

PROPOSITION 47 (Rappel)
Soient E un ev de dimension finie et $u \in \mathcal{L}(E)$ un endomorphisme diagonalisable avec $\mathrm{Spec}\,(u) = \{\lambda_1, \dots, \lambda_r\}$.
Soit $p_k : E \to E$ la projection sur $E_{\lambda_k}(u)$ parallèlement à $\bigoplus_{i \neq k} E_{\lambda_i}(u)$. Alors, on a :

$$u = \lambda_1 p_1 + \ldots + \lambda_r p_r$$
$$u^m = \lambda_1^m p_1 + \ldots + \lambda_r^m p_r,\ \forall m \geq 0$$

PROPOSITION 48
Soient E un ev de dimension finie et $u \in \mathcal{L}(E)$ un endomorphisme diagonalisable avec $\mathrm{Spec}\,(u) = \{\lambda_1, \dots, \lambda_r\}$.
On a $u = \lambda_1 p_1 + \ldots + \lambda_r p_r$, où $p_k : E \to E$ est la projection sur $E_{\lambda_k}(u)$ parallèlement à $\bigoplus_{i \neq k} E_{\lambda_i}(u)$.
Soient $L_1, \dots, L_r \in \mathbb{K}[X]$ les polynômes d'interpolation de Lagrange associés à l'ensemble $\{\lambda_1, \dots, \lambda_r\}$. Alors, on a :
$$p_k = L_k(u) = \Pi_{i \neq k} \frac{u - \lambda_i}{\lambda_k - \lambda_i},\ \forall 1 \leq k \leq r.$$

COROLLAIRE 49
Soit $u \in \mathcal{L}(E)$ un endomorphisme diagonalisable, avec $\mathrm{Spec}\,(u) = \{\lambda_1, \dots, \lambda_r\}$, dont on connaît $A = \mathrm{Mat}_{\mathcal{B}}(u)$.
On peut alors calculer u^m de deux façons :

- *Méthode 1 :* On détermine une base $\mathcal{B}'$ de vecteurs propres de u.
  On détermine P la matrice de passage de $\mathcal{B}$ vers $\mathcal{B}'$. On calcule P^{-1} l'inverse de P.
  Alors, on a $\mathrm{Mat}_{\mathcal{B}}(u) = A = PDP^{-1}$, où D est une matrice diagonale. Cela donne :
  $$\mathrm{Mat}_{\mathcal{B}}(u^m) = \mathrm{Mat}_{\mathcal{B}}(u)^m = A^m = PD^mP^{-1},\ \forall m \geq 0.$$
  On détermine donc u^m en calculant D^m (facile), puis en effectuant le produit PD^mP^{-1}.
- *Méthode 2 :* On détermine les polynômes interpolateurs de Lagrange $L_1, \dots, L_r$ associés à $\{\lambda_1, \dots, \lambda_r\}$ en développant l'expression $L_k(X) = \Pi_{i \neq k} \frac{X - \lambda_i}{\lambda_k - \lambda_i}$.
  Alors, on a
  $$u^m = \lambda_1^m L_1(u) + \ldots + \lambda_r^m L_r(u),\ \forall m \geq 0.$$

Exemple 50 — *On pose* $A = \begin{pmatrix} -1 & 0 & 1 \\ 2 & 0 & 2 \\ 2 & 0 & 0 \end{pmatrix}$. *Montrer que* A *est diagonalisable, et* A^m *pour tout* $m \geq 0$.

On commence par caluler le polynôme caractéristique de A. *Un calcul de déterminant donne :* $\chi_A(X) = X(X-1)(X+2)$.
Le polynôme caractéristique de A *est scindé à racines simples, donc* A *est diagonalisable. De plus, on obtient que* $\mathrm{Spec}\,(A) = \{-2, 0, 1\}$.
Soient L_1, L_2, L_3 *les polynômes d'interpolation de Lagrange associés à* $\{-2, 0, 1\}$. *On a :*

$$L_1(X) = \frac{(X-0)(X-1)}{(-2-0)(-2-1)} = \frac{X^2 - X}{6}$$
$$L_2(X) = \frac{(X+2)(X-1)}{(0+2)(0-1)} = \frac{X^2 + X - 2}{-2}$$
$$L_3(X) = \frac{(X+2)(X-0)}{(1+2)(1-0)} = \frac{X^2 + 2X}{3}$$

Pour tout $m \geq 0$, *on a donc :*

$$A^m = (-2)^m L_1(A) + 0^m L_2(A) + 1^m L_3(A).$$

Pour $m \geq 1$, *cela donne* $A^m = (-2)^m \frac{1}{6}(A^2 - A) + 0 + \frac{1}{3}(A^2 + 2A)$.
On calcule :

$$A^2 = \begin{pmatrix} 3 & 0 & -1 \\ 2 & 0 & 2 \\ -2 & 0 & 2 \end{pmatrix}.$$

On obtient donc :

$$A^m = \frac{(-2)^m}{6} \begin{pmatrix} 4 & 0 & -2 \\ 0 & 0 & 0 \\ -4 & 0 & 2 \end{pmatrix} + \frac{1}{3} \begin{pmatrix} 1 & 0 & 1 \\ 6 & 0 & 6 \\ 2 & 0 & 2 \end{pmatrix}.$$

Trigonalisation (HP)

Définition 51
Soient E un ev de dimension finie et $u \in \mathcal{L}(E)$ un endomorphisme.
On dit que u est **trigonalisable** s'il existe une base $\mathcal{B}$ de E dans laquelle la matrice $\mathrm{Mat}_{\mathcal{B}}(u)$ est triangulaire supérieure.
La base $\mathcal{B}$ est appelée une **base de trigonalisation** de u.[2].
Soit $A \in \mathscr{M}_n(\mathbb{K})$. On dit que A est **trigonalisable** s'il existe une matrice inversible P telle que $P^{-1}AP$ soit triangulaire supérieure.

Théorème 52
Soient E un ev de dimension finie et $u \in \mathcal{L}(E)$ un endomorphisme.
Alors, u est trigonalisable si et seulement s'il existe $P \in \mathbb{K}[X]$ scindé tel que $P(u) = 0$.

Corollaire 53
Soient E un $\mathbb{K}$-ev de dimension finie et $u \in \mathcal{L}(E)$ un endomorphisme.
On a les équivalences :

1. u est trigonalisable ;
2. χ_u est scindé ;
3. μ_u est scindé ;
4. $\mathrm{Spec}\,(u)$ est non-vide, et pour $\mathrm{Spec}\,(u) = \{\lambda_1, \ldots, \lambda_r\}$, on a
$$m_u(\lambda_1) + \ldots + m_u(\lambda_r) = \dim(E).$$

Remarque 54 — *Tout endomorphisme sur un* $\mathbb{C}$*-espace vectoriel de dimension finie est donc trigonalisable. De même, toute matrice complexe* A *est trigonalisable et s'écrit :*

$$A = PTP^{-1}$$

avec P *inversible et* T *triangulaire supérieure.*
Ce n'est pas le cas sur $\mathbb{R}$ *où un endomorphisme (une matrice) est trigonalisable si, et seulement si, sn polynôme caractéristique est scindé. Par exemple une matrice de rotation rotation* $R(\theta)$ *(*$\theta \not\equiv 0 \bmod \pi$*) de* $\mathbb{R}^2$ *n'est pas trigonalisable.*

2. Nous avons vu que c'est le cas si tous les sous-espaces vectoriels $\mathrm{Vect}\,(e_1, \ldots, e_i)$ sont stables par u. Voir le corollaire **??**.

13.4.1 Décomposition de Jordan-Chevalley (dite de Dunford)

Nous allons étudier une nouvelle forme de réduction plus poussée que la diagonalisation/trigonalisation.

Proposition 55 (Décomposition de Jordan-Chevalley (dite de Dunford))
Soient E un $\mathbb{K}$-ev de dimension finie et $u \in \mathcal{L}(E)$ un endomorphisme.
Si χ_u est scindé, alors il existe un unique couple d'endomorphismes $(d, n) \in (\mathcal{L}(E))^2$ avec d diagonalisable et n nilpotent tels que

$$u = d + n \quad \text{et} \quad d \circ n = n \circ d.$$

De plus, ces endomorphismes sont des polynômes en u.

Preuve — On écrit $\chi_u(X) = \prod_{k=1}^{r}(X - \lambda_k)^{m_k}$ et $F_k = \operatorname{Ker}\left((u - \lambda_k \operatorname{Id}_E)^{m_k}\right)$ le sous-espace caractéristique de λ_k.

Existence. D'après le lemme des noyaux on a $E = \bigoplus_{k=1}^{r} F_k$. D'après la remarque 42, la projection p_k sur F_k parallèlement à $\bigoplus_{i\neq k}^{r} F_i$ est un polynôme en u. On le note $p_k = P_k(u)$. On pose alors :

$$\boldsymbol{d = \sum_{k=1}^{r} \lambda_k p_k} \qquad \text{et} \qquad \boldsymbol{n = u - d = \sum_{k=1}^{r} (u - \lambda_k \operatorname{Id}_E) \circ p_k}.$$

On rappelle que $p_i \circ p_j = \delta_{i,j} p_i$, pour tous $1 \leq i, j \leq r$. Ainsi, d laisse stable chaque F_k, et l'on a $d_{F_k} = \lambda_k Id_{F_k}$. Donc d est diagonalisable.
De même, on montre par récurrence sur $q \geq 0$ que

$$n^q = \sum_{k=1}^{r} (u - \lambda_k \operatorname{Id}_E)^q \circ p_k, \ \forall \forall q \in \mathbb{N}^*.$$

Pour $q = \max\{m_k, k = 1, 2, \ldots, r\}$, on a $(u - \lambda_k \operatorname{Id}_E)^q_{F_k} = 0$ car $F_k = \operatorname{Ker}\left((u - \lambda_k \operatorname{Id}_E)^{m_k}\right)$.
Cela donne $(u - \lambda_k \operatorname{Id}_E)^q \circ p_k = 0$. Ainsi, on a $n^q = 0$, donc n est nilpotent.
Enfin, d et n sont des polynômes en u, donc ils commutent.

Unicité. Soit (d', n') un autre couple vérifiant les conditions. Comme d' et n' commutent, ils commuent avec $u = d' + n'$, donc ils commuent aussi avec d et n qui sont des polynômes en u. Ainsi, d et d' sont simultanément diagonalisables dans une même base, ce qui implique que $d - d'$ est diagonalisable.
D'autre part, n et n' commutent et sont trigonalisables, donc ils sont trigonalisables dans une même base $\mathcal{B}$. Alors, $\operatorname{Mat}_{\mathcal{B}}(n' - n) = \operatorname{Mat}_{\mathcal{B}}(n') - \operatorname{Mat}_{\mathcal{B}}(n)$ est une matrice triangulaire supérieure de diagonale nulle. On en déduit que $\chi_{n'-n}(X) = X^{\dim(E)}$ et donc que $n' - n$ est nilpotent.
Comme on a $d - d' = n' - n$, $d - d'$ est donc nilpotent. Le seul endomorphisme diagonalisable et nilpotent étant l'endomorphisme nul, on obtien $d - d' = n' - n = 0$, ce qui donne l'unicité.

□

Remarques 56

1. *Ainsi, tout endomorphisme trigonalisable u possède une décomposition de Dunford.*
2. *En fait, comme u laisse ses sous-espaces caractéristiques stables, on a défini d et n sur chaque sous-espace caractéristique F_k par :*

   $$d_{F_k} = \lambda_k Id_{F_k}, \ \text{et } n_{F_k} = u_{F_k} - d_{F_k} = u_{F_k} - \lambda_k Id_{F_k}.$$

   On remarque alors que d_{F_k} et n_{F_k} sont des polynôme en u_{F_k}, et que ces endomorphismes commutent.
   Le lemme des noyaux nous dit que les espaces F_k sont en somme directe, avec $E = \bigoplus_{k=1}^{r} F_k$, ce qui permet de faire remonter le comportement sur chaque F_k à un comportement sur E.
3. *L'écriture $u = d + n$ donnée par la décomposition de Dunford s'utilise pour calculer u^p :*

   $$u^p = (d + n)^p = \sum_{k=0}^{p} \binom{p}{k} d^k \circ n^{p-k}.$$

   Dans l'expression ci-dessus, on peut retirer les termes de la somme pour lesquels $p - k$ est plus grand que l'indice de nilpotence de n.

13.5 Applications aux EDL et aux suites récurrentes linéaires

Applications aux équations différentielles

Soit l'équation différentielle linéaire à coefficients constants d'ordre p :

$$f^{(p)} + \alpha_{p-1} f^{(p-1)} + \cdots + \alpha_1 f' + \alpha_0 f = 0 \qquad (ED)$$

avec $(\alpha_0, \ldots, \alpha_{p-1}) \in \mathbb{C}^p$, d'inconnue $f \in \mathcal{C}^\infty(\mathbb{R}, \mathbb{C})$.

On appelle **polynôme caractéristique** de (ED) le polynôme :

$$\chi(X) = X^p + \sum_{k=0}^{p-1} \alpha_k X^k.$$

On note sa factorisation dans $\mathbb{C}[X]$:

$$\chi(X) = \prod_{k=1}^{r} (X - \lambda_k)^{p_k},$$

où les λ_k sont deux à deux distincts et les p_k sont non-nuls.
Soit D l'endomorphisme de dérivation sur $\mathcal{C}^\infty(\mathbb{R}, \mathbb{C})$. Alors, l'ensemble $\mathcal{S}$ des solutions de (ED) est le noyau de l'endomorphisme $\chi(D)$.
Le lemme des noyaux nous fournit alors la décomposition en somme directe :

$$\mathcal{S} = \bigoplus_{k=1}^{r} \operatorname{Ker}\left((D - \lambda_k \operatorname{Id}_E)^{p_k}\right).$$

Résoudre l'équation (ED) se ramène donc à résoudre r équations différentielles linéaires plus simples.
Soient $1 \leq k \neq r$ et $f \in \mathcal{C}^\infty(\mathbb{R}, \mathbb{C})$. On démontre par récurrence sur $p_k \geq 1$ la relation suivante :

$$(D - \lambda_k \operatorname{Id}_E)^{p_k} \left(f(x) e^{\lambda_k x}\right) = e^{\lambda_k x} D^{p_k}(f)(x).$$

On en déduit alors que :

$$\operatorname{Ker}\left((D - \lambda_k \operatorname{Id}_E)^{p_k}\right) = \left\{x \in \mathbb{R} \longmapsto P(x) e^{\lambda_k x},\ P \in \mathbb{C}_{p_k - 1}[X]\right\}.$$

Ainsi, tout élément f de $\mathcal{S}$ s'écrit de façon unique comme :

$$x \in \mathbb{R} \longmapsto f(x) = \sum_{k=1}^{r} P_k(x) e^{\lambda_k x},\ P_1, \ldots, P_r \in \mathbb{C}[X],\ \deg(P_k) \leq p_k - 1.$$

L'espace vectoriel des solutions $\mathcal{S}$ est donc de dimension $p_1 + \ldots + p_k = p$.
On retrouve ce dernier point comme conséquence du théorème de Cauchy sur les équations linéaires à coefficients constants d'ordre 1 ou 2.

Exemple 57 — *Résoudre sur $\mathbb{R}$ l'équation différentielle :*

$$y^{(5)} - 13y^{(4)} + 67y^{(3)} - 171y" + 216y' = 108y. \ (E)$$

L'équation (E) est une équa. diff. linéaire à coefficients constants, d'ordre 5.
Pour $D : y \mapsto y'$ l'endomorphisme de dérivation sur $C^\infty(\mathbb{R}, \mathbb{C})$, les solutions de (E) sont exactement les fonctions dans $Ker(P(D))$, avec $P(X) = X^5 - 13X^4 + 67X^3 - 171X^2 + 216X - 108$. (les fonctions y telles que $P(D)(y) = 0$)
On remarque que 2 est une racine de P. En calculant P', on remarque que $P'(2) = 0$, donc 2 est une racine double de P.
On obtient alors en factorisant : $P(X) = (X^2 - 4X + 4)(X^3 - 9X^2 + 27X - 27)$.
On reconnaît le polynôme $(X - 3)^3$. On a donc $P(X) = (X - 2)^2 (X - 3)^3$.
D'après le Lemme des noyaux, on a donc :

$$Ker(P(D)) = Ker((D - 2Id)^2) \oplus Ker((D - 3Id)^3).$$

D'après les résultats de la section, on a ainsi :

$$Ker(P(D)) = Vect(x \mapsto \exp(2x), x \mapsto x\exp(2x), x \mapsto \exp(3x), x \mapsto x\exp(3x), x \mapsto x^2 \exp(3x)).$$

Une fonction y est solution de (E) si et seulement s'il existe $a_1, \ldots, a_5 \in \mathbb{C}$ tels que

$$y(x) = (a_1 x + a_2)\exp(2x) + (a_3 x^2 + a_4 x + a_5)\exp(3x),\ \forall x \in \mathbb{R}.$$

Applications aux suites récurrentes linéaires à coefficients constants

On prend $\mathbb{K} = \mathbb{Q}, \mathbb{R}$, ou $\mathbb{C}$. (ou un sous-corps de $\mathbb{C}$)
On considère l'*équation récurrente linéaire* à coefficients constants d'ordre p :

$$\forall n \in \mathbb{N}, u_{n+p} + \alpha_{p-1}u_{n+p-1} + \cdots + \alpha_1 u_{n+1} + \alpha_0 u_n = 0 \qquad (ER)$$

avec $\alpha_0, \ldots, \alpha_{p-1} \in \mathbb{K}$. L'inconnue est une suite $u = (u_n)_n \in \mathbb{K}^{\mathbb{N}}$.

On appelle **polynôme caractéristique** de (ER) le polynôme :

$$\chi(X) = X^p + \sum_{k=0}^{p-1} \alpha_k X^k.$$

On suppose que le polynôme $\chi(X)$ est scindé sur $\mathbb{K}$, de factorisation :

$$\chi(X) = \prod_{k=1}^{r} (X - \lambda_k)^{p_k},$$

avec les racines λ_k deux à deux distinctes et $p_k > 0$.
On note T l'endomorphisme de translation sur $\mathcal{L}(\mathbb{K}^{\mathbb{N}})$, défini par :

$$\begin{array}{rcl} T : \mathbb{K}^{\mathbb{N}} & \longrightarrow & \mathbb{K}^{\mathbb{N}} \\ (u_n)_{n\in\mathbb{N}} & \longmapsto & (u_{n+1})_{n\in\mathbb{N}}. \end{array}$$

L'ensemble $\mathcal{S}$ des solutions de l'équation (ER) est le noyau de l'endomorphisme $\chi(T)$ sur $\mathbb{K}^{\mathbb{N}}$.
Comme une suite dans $\mathcal{S}$ est déterminée par ses p premiers termes, l'application

$$\begin{array}{rcl} u : \mathcal{S} & \longrightarrow & \mathbb{K}^p \\ (u_n)_{n\in\mathbb{N}} & \longmapsto & (u_0, \ldots, u_{p-1}) \end{array}$$

est un isomorphisme, donc $\dim(\mathcal{S}) = p$.
De plus, le lemme des noyaux nous fournit la décomposition en somme directe :

$$\mathcal{S} = \bigoplus_{k=1}^{r} \operatorname{Ker}\left((T - \lambda_k \operatorname{Id}_{\mathbb{K}^{\mathbb{N}}})^{p_k}\right).$$

Il faut alors déterminer, pour tout k, le noyau de l'endomorphisme $(T - \lambda_k \operatorname{Id}_{\mathbb{K}^{\mathbb{N}}})^{p_k}$.
Ce sous-espace est de dimension p_k.

- Si $\lambda_k = 0$, il s'agit du noyau de T^{p_k}. C'est évidemment l'espace vectoriel des suites (u_n) telle que $u_n = 0$ pour tout $n \geqslant p_k$. On retrouve qu'il est de dimension p_k.
- Supposons $\lambda_k \neq 0$. Soit $U \in \mathbb{K}[X]$ un polynôme. Un calcul montre que

  $$(T - \lambda_k \operatorname{Id}_E)((\lambda_k^n U(n))_{n\in\mathbb{N}}) = \left(\lambda_k^{n+1} V(n)\right)_{n\in\mathbb{N}},$$

  avec $V(n)$ le polynôme $U(n+1) - U(n)$.
  On montre alors par récurrence sur p_k que $\operatorname{Ker}\left((T - \lambda_k \operatorname{Id}_E)^{p_k}\right)$ contient l'ensemble :

  $$\left\{(\lambda_k^n U(n))_{n\in\mathbb{N}},\ U \in \mathbb{K}[X],\ \deg(U) \leq p_k - 1\right\}.$$

  Cet ensemble est lui aussi un sous-espace vectoriel de dimension p_k, puisqu'il est l'image de $\mathbb{K}_{p_k-1}[X]$ par l'application linéaire injective

  $$\begin{array}{rcl} \Phi : \mathbb{K}_{p_k-1}[X] & \longrightarrow & \operatorname{Ker}\left((T - \lambda_k \operatorname{Id}_E)^{p_k}\right) \\ U & \longmapsto & (\lambda_k^n U(n))_{n\in\mathbb{N}}. \end{array}$$

  Notons qu'on utilise ici le fait que $\mathbb{K}$ contienne $\mathbb{N}$ (c'est faux si $\mathbb{K} = \mathbb{Z}/p\mathbb{Z}$ par ex.)
  Par égalité des dimensions, on obtient alors :

  $$\operatorname{Ker}\left((T - \lambda_k \operatorname{Id}_E)^{p_k}\right) = \left\{(\lambda_k^n U(n))_{n\in\mathbb{N}}, U \in \mathbb{K}_{p_k-1}[X]\right\}.$$

En supposant que $\alpha_0 \neq 0$, on trouve finalement que toute solution u de (ER) est de la forme :

$$\forall n \in \mathbb{N}, u_n = \sum_{k=1}^{r} \lambda_k^n U_k(n)$$

pour une unique famille $U_1, \ldots, U_r \in \mathbb{K}[X]$ avec $\deg(U_k) \leq p_k - 1$.
On retrouve aussi que l'espace $\mathcal{S}$ est de dimension p.

EXEMPLE 58 — *Trouver toutes les suites $(u_n)_n$ à coefficients complexes vérifiant*

$$u_{n+4} = 4u_{n+3} - 3u_{n+2} - 4u_{n+1} + 4u_n, \forall n \geq 0 \, (E)$$

Pour $T : (u_n)_n \mapsto (u_{n+1})_n$ l'application linéaire de décalage à gauche sur $(\mathbb{C})^{\mathbb{N}}$, les suites $u = (u_n)_n$ vérifiant (E) sont exactement les suites telles que $P(T)(u) = 0$, pour P le polynôme $P(X) = X^4 - 4X^3 + 3X^2 + 4X - 4$. On trouve que $1, -1, 2$ sont des racines de P. On obtient $P(X) = (X-1)(X+1)(X-2)^2$.
Ainsi, on a $Ker(P(T)) = Vect((1^n)_n, ((-1)^n)_n, (2^n)_n, (n2^n)_n)$.
Donc, une suite u vérifie (E) si et seulement s'il existe $a, b, c, d \in \mathbb{C}$ tels que

$$u_n = a + b(-1)^n + (cn + d)2^n, \forall n \geq 0.$$

Chapitre 14 Algèbre bilinéaire

Table des matières du chapitre

14.1 Formes bilinéaires, formes bilinéaires symétriques

Pour certains objets (ex : produit scalaire) nous nous restreindrons aux corps $\mathbb{K} = \mathbb{Q}, \mathbb{R}, \mathbb{C}$, voire seulement $\mathbb{R}$.

Définition 1
Soit E un $\mathbb{K}$-espace vectoriel E. Soit $f : E \times E \longrightarrow \mathbb{K}$.
On dit que f est une **forme bilinéaire** , si pour tous $x, y \in E$, les fonctions

$$f(.,y) : \begin{array}{ccc} E & \longrightarrow & \mathbb{K} \\ x' & \longmapsto & f(x',y) \end{array} \quad \text{et} \quad f(x,.) : \begin{array}{ccc} E & \longrightarrow & \mathbb{K} \\ y' & \longmapsto & f(x,y') \end{array}$$

sont des applications linéaires.

Exemples 2

1. *Pour $E = \mathbb{R}^2$, la fonction $f : (u,v) \in \mathbb{R}^2 \times \mathbb{R}^2 \mapsto xx' + yy' \in \mathbb{R}$ est une forme bilinéaire.*
2. *Pour $E = \mathbb{C}$, la fonction $f : (z,z') \in \mathbb{C} \times \mathbb{C} \mapsto zz' \in \mathbb{C}$ est une forme bilinéaire.*
3. *Pour $E = \mathbb{R}[X]$, la fonction $f : (P,Q) \in E^2 \mapsto P(0)Q(0) \in \mathbb{R}$ est une forme bilinéaire.*
4. *Pour $E = \mathcal{C}^0([a,b],\mathbb{R})$, avec $a < b$, la fonction $S : (f,g) \in E^2 \mapsto \int_a^b f(t)g(t)dt \in \mathbb{R}$ est une forme bilinéaire.*

Remarque 3 — *Une forme bilinéaire $f : E \times E \to \mathbb{K}$ est très différente d'une forme linéaire $g : E \times E \to \mathbb{K}$. Pour $(x,y),(x',y') \in E \times E$, on a $g(x+x', y+y') = g(x,y) + g(x',y')$ par linéarité, tandis que*

$$f(x+x', y+y') = f(x,y) + f(x,y') + f(x',y) + f(x',y'), \text{ par bilinéarité .}$$

De même, pour $\lambda \in \mathbb{K}$ on a $g(\lambda x, \lambda y) = \lambda g(x,y)$ par linéarité, tandis que

$$f(\lambda x, \lambda y) = \lambda f(x, \lambda y) = \lambda^2 f(x,y), \text{ par bilinéarité .}$$

Remarque 4 — *Pour $f, g : E \times E \to \mathbb{K}$ des formes bilinéaires et $\lambda \in \mathbb{K}$, la fonction $f + \lambda g$ est encore une forme bilinéaire.*

En effet, pour tous $x, y \in E$, les fonctions $f(x,.) + \lambda g(x,.)$ et $f(.,y) + \lambda g(.,y)$ sont des applications linéaires (toute combinaison linéaire d'applications linéaires est une application linéaire).

Donc, l'ensemble des formes bilinéaires sur E est un $\mathbb{K}$-espace vectoriel.

14.2 Matrice d'une forme bilinéaire

Proposition-Définition 5
Soit E un $\mathbb{K}$-espace vectoriel de dimension finie n.
Soient $\mathcal{B} = (e_1, \cdots, e_n)$ une base de E et $f : E^2 \to \mathbb{K}$ une forme bilinéaire.

Soit $A = (a_{ij})_{i,j} \in \mathscr{M}_n(\mathbb{K})$ avec $a_{i,j} = f(e_i, e_j)$.
On dit que A est **la matrice de la forme bilinéaire** f dans la base $\mathcal{B}$, notée $Mat_{\mathcal{B}}(f)$.

Pour $x, y \in E$ avec $x = x_1e_1 + \cdots + x_ne_n$ et $y = y_1e_1 + \cdots + y_ne_n$, on a :

$$f(x,y) = \sum_{i=1}^{n}\sum_{j=1}^{n} a_{ij}x_iy_i = \underbrace{\begin{pmatrix} x_1 & x_2 & \cdots & x_n \end{pmatrix}}_{{}^tX} \underbrace{\begin{pmatrix} a_{1,1} & a_{1,2} & \dots & a_{1,p} \\ a_{2,1} & a_{2,2} & \dots & a_{2,p} \\ \vdots & \vdots & & \vdots \\ a_{n,1} & a_{n,2} & \dots & a_{n,p} \end{pmatrix}}_{A} \underbrace{\begin{pmatrix} y_1 \\ y_2 \\ \vdots \\ y_n \end{pmatrix}}_{Y} = {}^tX \cdot A \cdot Y.$$

Preuve —

$$\begin{aligned} f(x,y) &= f(\sum_{i=1}^{n} x_ie_i, \sum_{j=1}^{n} y_je_j) = \sum_{i=1}^{n} x_i f(e_i, \sum_{j=1}^{n} y_je_j) = \sum_{i=1}^{n}\sum_{j=1}^{n} x_iy_jf(e_i,e_j) \text{ par linéarité de } f(.,y) \text{ et } f(e_i,.). \\ &= \sum_{i=1}^{n}\sum_{j=1}^{n} a_{ij}x_iy_j = \sum_{i=1}^{n} x_i \left(\sum_{j=1}^{n} a_{ij}y_j \right) = {}^tX \cdot (A \cdot Y)\,. \end{aligned}$$

□

Définition 6
Soient E un $\mathbb{K}$-espace vectoriel (abrégé e.v.) et $f : E \times E \to \mathbb{K}$ une forme bilinéaire. On dit que f est une forme bilinéaire **symétrique** si l'on a $f(x,y) = f(y,x)$ pour tous $x, y \in E$.

Exemple 7 —

- *Pour $E = \mathbb{K}[X]$ et $\varphi : (P,Q) \in E^2 \mapsto P(1)Q(1) \in \mathbb{K}$, φ est une forme bilinéaire symétrique.*
- *Pour $E = \mathbb{R}^2$, $\varphi : (x,y) \in E^2 \mapsto x_1y_1 + y_2y_2 \in \mathbb{R}$ est une forme bilinéaire symétrique.*
- *Pour $E = \mathbb{R}^2$, $\varphi : (x,y) \in E^2 \mapsto x_1y_2 - x_2y_1 \in \mathbb{R}$ et $\psi : (x,y) \in E^2 \mapsto x_1y_2 \in \mathbb{R}$ sont des formes bilinéaires qui ne sont pas symétriques ($\varphi((1,0),(0,1)) = 1 \neq -1 = \varphi((0,1),(1,0))$).*

Proposition 8
Soient E un $\mathbb{K}$-e.v. de dimension finie n, et $\mathcal{B} = (e_1, \dots, e_n)$ une base de E. Soit $f : E \times E \to \mathbb{K}$ une forme bilinéaire.
Alors f est une forme bilinéaire symétrique si et seulement si $Mat_{\mathcal{B}}(f)$ est une matrice symétrique.

Preuve —

- Les coefficients de la matrice $A = Mat_{\mathcal{B}}(f)$ sont les $a_{i,j} = f(e_i,e_j)$, $1 \leq i,j \leq n$.
  Si f est symétrique, alors on a $f(e_i,e_j) = f(e_j,e_i)$, d'où $a_{ij} = a_{ji}$, donc A est une matrice symétrique ($A = {}^tA$).
- Réciproquement, supposons A est symétrique. On a $f(x,y) = {}^tX \cdot A \cdot Y$. La matrice ${}^tX \cdot A \cdot Y$ est une matrice 1×1. Elle est donc égale à sa transposée. D'où :
  $$f(x,y) = {}^tX \cdot A \cdot Y = {}^t({}^tX \cdot A \cdot Y) = {}^tY \cdot {}^tA \cdot {}^t({}^tX) = {}^tY \cdot A \cdot X = f(y,x).$$
  Donc f est une forme bilinéaire symétrique.

□

Il ne faut pas confondre la matrice d'une forme bilinéaire $\varphi : E \times E \to \mathbb{K}$ et la matrice d'une application linéaire $f : E \to E$. Pour $\mathcal{B} = (e_1, e_2, \cdots, e_n)$ de E, une même matrice $A \in \mathcal{M}_n(\mathbb{K})$ peut représenter un endomorphisme $f(X \mapsto AX)$ et une forme bilinéaire φ $((X,Y) \mapsto {}^tXAY)$.

Si l'on change de base de E, ces matrices ne changent pas de la même manière.

Proposition 9
Soit E un $\mathbb{K}$-e.v. de dimension finie n. Soient $\mathcal{B}$, $\mathcal{B}'$ deux bases de E. Soient $f : E \to E$ une application linéaire et $\varphi : E^2 \to \mathbb{K}$ une forme bilinéaire. Soit P la matrice de passage de la base $\mathcal{B}$ vers $\mathcal{B}'$. Alors on a :

- $Mat_{\mathcal{B}'}(f) = P^{-1}Mat_{\mathcal{B}}(f)P$;
- $Mat_{\mathcal{B}'}(\varphi) = {}^tPMat_{\mathcal{B}}(\varphi)P$.

Preuve —

endomorphisme f		forme bilinéaire φ
$y = f(x) \Leftrightarrow Y = Mat_{\mathcal{B}}(f)X$		$\varphi(x,y) = {}^tXMat_{\mathcal{B}}(\varphi)Y$
$\Leftrightarrow Y' = BX'$		$= {}^tX'CY'$
	$X = PX'$	
	$Y = PY'$	
D'où $B = P^{-1}Mat_{\mathcal{B}}(f)P$.		D'où $C = {}^tPMat_{\mathcal{B}}(\varphi)P$.

□

14.3 Produit scalaire

Passons maintenant à des espaces vectoriels réels.

Définition 10
Soit E un $\mathbb{R}$-espace vectoriel. Soit $S : E^2 \to \mathbb{R}$ une forme bilinéaire.
On dit que S est un **produit scalaire** si celle-ci vérifie :

- $\forall x, y \in E$, $S(x,y) = S(y,x)$ (S est symétrique) ;
- $\forall x \in E$, $S(x,x) \geq 0$, avec $S(x,x) = 0 \Longleftrightarrow x = 0$ (S est **définie positive**).

Un produit scalaire (ou forme bilinéaire symétrique définie positive) est noté $\langle x|y\rangle$, ou $\langle x,y\rangle$ ou $x \cdot y$.

Remarque 11 — *En géométrie, on utilise souvent la notation $\overrightarrow{u} \cdot \overrightarrow{v}$ pour désigner le produit scalaire des vecteurs $\overrightarrow{u}$ et $\overrightarrow{v}$.*

La notion de produit scalairé nécessite que le corps $\mathbb{K}$ possède des éléments "positifs". Nous n'étudierons le produit scalaire que pour $\mathbb{K} = \mathbb{R}$.

Théorème 12 (Inégalité de Cauchy-Schwarz)
Soit E un $\mathbb{R}$-e.v. muni d'un produit scalaire $< .|. >$.

1. Pour x et y dans E, on a : $\langle x|y\rangle^2 \leqslant \langle x|x\rangle\langle y|y\rangle$.
2. Cette inégalité est une égalité si, et seulement si, x et y sont colinéaires ($x = \lambda y$ ou $y = \lambda x$, pour un $\lambda \in \mathbb{R}$).

Preuve

1.
   - Si $y = 0$, l'inégalité est évidente (c'est une égalité).
   - Sinon, posons $P(\lambda) =< x + \lambda\, y|x + \lambda\, y >= \lambda^2 \ < y|y > +2\lambda \ < x|y > + < x|x >$.
     Alors P est une fonction polynomiale de degré 2 (puisque $\forall y \neq 0, \langle y|y\rangle > 0$) telle que $P(\lambda) \geqslant 0$, $\forall \lambda \in \mathbb{R}$. Son discriminant :
     $$\Delta = 4 < x|y >^2 -4 < x|x >< y|y >$$
     est donc négatif ou nul, ce qui donne l'inégalité annoncée.
2.
   - Si x et y sont proportionnels, il existe un scalaire λ tel que $y = \lambda\, x$ ou $x = \lambda\, y$.
     Supposons par exemple $y = \lambda\, x$. Alors on a : $< x|y >^2 =< x|\lambda\, x >^2 = \lambda^2 \ < x|x >^2 =< x|x >< y|y >$.
   - Réciproquement, supposons que $< x|y >^2 =< x|x >< y|y >$.
     - Si $y = 0$, alors x et y sont proportionnels.
     - Sinon, le polynôme P est de degré 2 avec un discriminant nul. Il existe donc $\lambda \in \mathbb{R}$ tel que $P(\lambda) = 0$.
       Cela donne $< x + \lambda\, y|x + \lambda\, y >= 0$. Par définition du produit scalaire, on en déduit que $x + \lambda\, y = 0$, et donc que x est proportionnel à y.

□

Exemples 13

1. *Pour $E = \mathbb{R}^n$, $n \geq 1$, la forme bilinéaire symétrique $S : \mathbb{R}^n \times \mathbb{R}^n \to \mathbb{R}$ définie par :*
   $$S\big((x_1, x_2, \ldots, x_n), (y_1, y_2, \ldots, y_n)\big) = \sum_{i=1}^{n} x_i\, y_i$$
   est un produit scalaire.
   *On l'appelle le **produit scalaire canonique** sur $\mathbb{R}^n$.*
2. *Pour $E = \mathcal{C}^0([a,b], \mathbb{R})$ et $S : (f,g) \in E^2 \mapsto \int_0^1 f(t)g(t)dt \in \mathbb{R}$. S est une forme bilinéaire symétrique ainsi qu'un produit scalaire.*

Remarque 14 — *Pour $S : E \times E \to \mathbb{R}$ une forme bilinéaire symétrique telle que $S(x,x) \geq 0$, S vérifie l'inégalité de Cauchy-Schwarz.*
Le cas d'égalité est par contre faux si S n'est pas un produit scalaire (si $S(x,x) = 0$ n'implique pas $x = 0$).

Remarque 15 — *Pour vérifier qu'une forme bilinéaire symétrique S est un produit scalaire, il faut regarder si $S(x,x)$ est toujours positif ou nul, et si $S(x,x) = 0 \Leftrightarrow x = 0$.*

Remarque 16 — *Soient $n \geq 1$ et $A \in \mathscr{M}_n(\mathbb{R})$. Alors la forme bilinéaire $(X,Y) \mapsto {}^tXAY$ est un produit scalaire si et seulement si :*

- *A est une matrice symétrique ;*
- *${}^tXAX \geq 0$ pour tout vecteur colonne $X \in \mathbb{R}^n$;*
- *${}^tXAX = 0$ si et seulement si $X = 0$.*

Exemple 17 — *Soit $E = \mathbb{R}^n$. Soit $A \in \mathscr{M}_n(\mathbb{R})$ inversible. On pose $B = {}^tAA$.*
Alors la fonction $S : (X, Y) \in \mathbb{R}^n \times \mathbb{R}^n \mapsto {}^tXBY \in \mathbb{R}$ est un produit scalaire.
Cette fonction est bien une forme bilinéaire, comme vu précédemment.
La matrice B est symétrique, donc cette forme bilinéaire est symétrique.

Soit $X \in \mathbb{R}^n$. En posant $Z = AX$, on a $S(X, X) = {}^tX{}^tAAY = {}^tZZ$. Pour $\langle .|. \rangle$ le produit scalaire canonique sur $\mathbb{R}^n$, on a donc :

$$S(X, X) = \langle Z|Z\rangle = \langle AX|AX\rangle \geq 0.$$

La forme bilin. sym. S est donc positive.

Enfin, si l'on a $S(X, X) = 0$, alors on a $\langle AX|AX\rangle = 0$, donc $AX = 0$. Comme la matrice A est inversible, on obtient $X = 0$. Donc S est une forme bilin. sym. définie positive, c'est-à-dire un produit scalaire.

14.4 Norme euclidienne

Définition 18
Soit E un $\mathbb{R}$-e.v. Soit $N : E \to \mathbb{R}_+$.
On dit que N est une **norme** sur E si cette fonction vérifie les axiomes suivants :

- $N(x) = 0 \iff x = 0$ (axiome de séparation) ;
- $N(\lambda \cdot x) = |\lambda|\, N(x)$, $\forall (\lambda, x) \in \mathbb{R} \times E$ (homogénéité) ;
- $N(x + y) \leqslant N(x) + N(y)$, $\forall (x, y) \in E^2$ (inégalité triangulaire).

Proposition-Définition 19
Soit E un $\mathbb{R}$-e.v. muni d'un produit scalaire $< .|. >$. Alors, la fonction :

$$\begin{array}{rccl} ||\cdot|| : & E & \longrightarrow & \mathbb{R}_+ \\ & x & \longmapsto & ||x|| = \sqrt{< x|x >} \end{array}$$

est une norme sur E.

Cette norme est appelée **norme euclidienne** associée au produit scalaire $< .|. >$.
La distance associée $d : (x, y) \mapsto \|x - y\|$ est appelée **distance euclidienne** .
En utilisant cette norme, l'inégalité de Cauchy-Schwarz s'écrit :

$$\big|< x|y >\big| \leqslant ||x||\, ||y||\,.$$

Exemples 20

1. *Pour $n \geq 1$, soit $\phi : (x, y) \in \mathbb{R}^n \times \mathbb{R}^n \mapsto \sum_{i=1}^n x_i\, y_i \in \mathbb{R}$ le produit scalaire canonique de $\mathbb{R}^n$. Alors la norme euclidienne associée est :*

$$||(x_1, x_2, \ldots, x_n)|| = \sqrt{\sum_{i=1}^n x_i^2}.$$

L'inégalité de Cauchy-Schwarz pour ce produit scalaire s'écrit : ◀

$$\left|\sum_{i=1}^n x_i y_i\right| \leqslant \left(\sum_{i=1}^n x_i^2\right)^{1/2} \left(\sum_{i=1}^n y_i^2\right)^{1/2}.$$

2. *Dans un espace vectoriel de dimension n muni d'une base $\mathcal{B}$, on peut définir un produit scalaire en posant :*

$$\langle x|y\rangle = \sum_{i=1}^n x_i\, y_i,$$

où x_1, x_2,..., x_n (resp. y_1, y_2,..., y_n) sont les composantes dans la base $\mathcal{B}$ du vecteur x (resp. y).

3. *Pour $E = \mathbb{C}^0\big([a, b]\big)$, la fonction $(f, g) \mapsto \int_a^b f(x)\, g(x)\, dx$ est un produit scalaire sur cet e.v..* ◀
L'inégalité de Cauchy-Schwarz correspondante est :

$$\left|\int_a^b f(x)\, g(x)\, dx\right| \leqslant \left(\int_a^b f^2(x)\, dx\right)^{1/2} \left(\int_a^b g^2(x)\, dx\right)^{1/2}.$$

4. *Soit $E = C^0(\mathbb{R}/2\pi\mathbb{Z})$ l'espace vectoriel des fonctions continues et 2π-périodiques sur $\mathbb{R}$. La fonction $(f,g) \mapsto \frac{1}{\pi}\int_0^{2\pi} f(x)\,g(x)\,dx$ est un produit scalaire sur cet e.v..*
5. *Pour $E = \mathbb{R}^3$, la forme bilinéaire symétrique S définie par :*

$$S\big((x,y,z),(x',y',z')\big) = x\,x' + y\,y' + z\,z' + \frac{1}{2}(x\,y' + x'\,y + x\,z' + x'\,z + y\,z' + y'\,z)$$

est un produit scalaire. En effet, on a : ◀

$$\begin{aligned} S\big((x,y,z),(x,y,z)\big) &= x^2 + y^2 + z^2 + x\,y + x\,z + y\,z \\ &= \frac{1}{2}\left((x+y)^2 + (y+z)^2 + (z+x)^2\right). \end{aligned}$$

Donc $S\big((x,y,z),(x,y,z)\big)$ est positif et ne peut être nul que si $x = y = z = 0$.

Proposition 21
Soit E un $\mathbb{R}$-e.v. muni d'un produit scalaire. Soient $x, y \in E$.
Alors, on a $\|x+y\| = \|x\| + \|y\|$ si et seulement si $x = \lambda y$ ou $y = \lambda x$ pour un $\lambda \in \mathbb{R}_+$ (ssi x et y sont *positivement liés*).

Preuve — D'après la preuve de la Proposition-Définition 19, x et y vérifient le cas d'égalité dans l'inégalité triangulaire si et seulement si $2\langle x|y\rangle = 2\|x\|\|y\|$.
Ainsi, x et y vérifient le cas d'égalité dans l'inégalité de Cauchy-Schwarz. On a donc $x = \lambda y$ ou $y = \lambda x$ pour un $\lambda \in \mathbb{R}$.
Comme $\langle x|y\rangle = \|x\|\|y\| \geq 0$, on doit avoir $\lambda \geq 0$. □

Remarque 22 — *Pour E un $\mathbb{R}$-e.v. et $N : E \to \mathbb{R}_+$ une fonction. Si l'on peut trouver un produit scalaire $\langle .|. \rangle$ tel que $N(x)^2 = \langle x|x\rangle$, alors on aura montré que N est une norme.*

Théorème 23 (**Equivalence des normes en dimension finie**)
Soit E un $\mathbb{R}$-e.v. de dimension finie.
Alors, toutes les normes sur E sont équivalentes.
Autrement dit, pour $\|.\|_1, \|.\|_2$ deux normes sur E, il existe $a, b > 0$ tels que :

$$a\|x\|_1 \leq \|x\|_2 \leq b\|x\|_1,\ \forall x \in E.$$

Preuve — *Idées de preuve* : On choisit pour E $\mathbb{R}^n$, et pour $\|.\|_1$ la norme infinie.
La sphère unité S pour $\|.\|_1$ est un compact, car c'est un fermé borné de $\mathbb{R}^n$.
On a $\|x\|_2 = \|x_1e_1 + \ldots + x_ne_n\|_2 \leq |x_1|\|e_1\|_2 + \ldots + |x_n|\|e_n\|_2 \leq \max_i(\|e_i\|_2)(\sum_{k=1}^n |x_i|) = \max_i(\|e_i\|_2)\|x\|_1$.
Donc, la fonction $\|.\|_2$ est continue sur $(E, \|.\|_1)$ (car $\max_i(\|e_i\|_2)$-Lipschitzienne).
Elle admet alors un maximum et un minimum sur S (fonction continue sur un compact). Notons a, b ces extrema.
Pour $x \in E$ non-nul on a $\frac{x}{\|x\|_1} \in S$.
Cela donne $a \leq \|\frac{x}{\|x\|_1}\|_2 \leq b$, ce qui permet de conclure. □

14.5 Espaces vectoriels euclidiens

Définition 24
Soit E un $\mathbb{R}$-e.v. muni d'un produit scalaire. On dit alors que E est un **espace vectoriel euclidien** .

Si $(E, <.|.>)$ est un espace euclidien, il est donc naturellement muni de la norme euclidienne associée à son produit scalaire.

Exemple 25 — *Tous les produits scalaires considérés précédemment munissent leur espace vectoriel associé d'une structure d'espace euclidien.*

Remarque 26 — *La norme euclidienne $\|\cdot\|$ d'un espace vectoriel euclidien E est définie à partir de son produit scalaire $\langle .|. \rangle$.*

Nous allons montrer que la réciproque est vraie : Si l'on connaît toutes les valeurs de la norme euclidienne $\|\cdot\|$, alors on peut retrouver les valeurs du produit scalaire $\langle .|. \rangle$. Pour cela, nous aurons besoin des égalités suivantes.

PROPOSITION 27
Soit E un espace vectoriel euclidien. Soit $(x, y) \in E^2$. On a :

- **Identités de polarisation** :
  1. $||x+y||^2 = ||x||^2 + ||y||^2 + 2 < x|y >$;
  2. $||x-y||^2 = ||x||^2 + ||y||^2 - 2 < x|y >$;
  3. $||x+y||^2 - ||x-y||^2 = 4 < x|y >$.
- **Identité du parallélogramme** :
  $||x+y||^2 + ||x-y||^2 = 2\left(||x||^2 + ||y||^2\right)$ [1].

Preuve —

- On a : $||x+y||^2 =< x+y|x+y >=< x|x > + < x|y > + < y|x > + < y|y >=< x|x > +2 < x|y > + < y|y >$.
- Appliquer l'égalité précédente à x et $-y$ pour développer $||x+y||^2$.
- Les deux dernières égalités se déduisent des précédentes par somme et différence.

□

REMARQUE 28 — *Pour $\|.\|$ une norme issue d'un produit scalaire, les identités de polarisation nous disent alors que*

$$\langle x|y\rangle = \frac{||x+y||^2 - ||x-y||^2}{4}.$$

On peut donc bien déterminer les valeurs de $\langle .|.\rangle$ en fonction des valeurs de $\|.\|$.
L'identité du parallélogramme est une identité que vérifient toutes les normes euclidiennes. Mais certaines normes ne sont pas euclidiennes. Une façon qui permet de le montrer est de trouver x et y qui ne vérifient pas l'identité du parallélogramme.

EXEMPLES 29

1. *Sur $\mathbb{R}^2$, on définit $N((x,y)) = \sqrt{x^2+2xy+3y^2}$. Pour montrer que définit une norme euclidienne sur $\mathbb{R}^2$, on commence par vérifier que :*
$$x^2+2xy+3y^2 = (x+y)^2 + 2y^2$$
est positif et ne peut être nul que si $(x,y) = (0,0)$.
Il faut alors exhiber le produit scalaire dont N provient. D'après la dernière identité de polarisation, il doit être égal à :
$$\begin{aligned} S\big((x,y),(x',y')\big) &= \frac{N((x+x',y+y'))^2 - N((x-x',y-y'))^2}{4} \\ &= x\,x' + x\,y' + y\,x' + 3y\,y'. \end{aligned}$$
On remarque que la fonction S est bien une forme bilinéaire symétrique S, et que l'on a :
$$\forall (x,y) \in \mathbb{R}^2\,,\ S\big((x,y),(x,y)\big) = N((x,y))^2,$$
Ainsi, S est bien un produit scalaire, donc N est bien une norme euclidienne (et donc une norme).
2. *Sur $\mathbb{R}^2$, on définit la norme « infinie » par :*
$$||(x,y)||_\infty = \max\big(|x|,|y|\big)$$
Cette fonction est bien une norme (voir Analyse 4) mais ce n'est pas une norme euclidienne.
En effet, pour $u = (2,1)$ et $v = (1,2)$, on a :
$$||u+v|| = 3 \qquad ||u-v|| = 1 \qquad ||u|| = ||v|| = 2$$
et donc : $||u+v||^2 + ||u-v||^2 = 10 \neq 8 = 2\left(||u||^2 + ||v||^2\right)$.
3. *Pour $E = \mathbb{R}^n$, on définit la norme « ℓ^1 » par :*
$$\|x\|_{\ell^1} = \sum_{i=1}^{n} |x_i|.$$
Cette fonction est bien une norme (voir Analyse 4), mais ce n'est pas une norme euclidienne.

1. La somme des carrés des quatre côtés est égale à la somme des carrés des deux diagonales.

Le produit scalaire permet par exemple de définir la notion d'angle entre deux vecteurs :

COROLLAIRE 30 (Lien entre produit scalaire et angles en géométrie)
Soit E un e.v. euclidien. Soient $x, y \in E$ non-nuls. Alors :

$$\exists! \theta \in [0, \pi], \text{ tel que } \langle x|y\rangle = \|x\| \cdot \|y\| \cdot \cos\theta.$$

Preuve — D'après l'inégalité de Cauchy-Schwarz, on a $|\langle x|y\rangle| \leqslant \|x\| \cdot \|y\|$. Si les vecteurs x et y ne sont pas nuls, on en déduit que $\dfrac{|\langle x|y\rangle|}{\|x\| \cdot \|y\|} \leqslant 1$. D'où $-1 \leqslant \dfrac{\langle x|y\rangle}{\|x\| \cdot \|y\|} \leqslant +1$, donc $\exists! \theta \in [0, \pi]$, tel que $\dfrac{\langle x|y\rangle}{\|x\| \cdot \|y\|} = \cos\theta$. □

Chapitre 15 Orthogonalité

Table des matières du chapitre

15.1 Bases orthonormées

Nous travaillerons jusqu'à la fin de ce chapitre sur un $\mathbb{R}$-e.v. euclidien E désigne un espace vectoriel euclidien, muni de sa norme euclidienne $||\cdot||$ et de la distance d associée.

Familles orthonormées

Définition 1
Soit E un $\mathbb{R}$-e.v. euclidien. Soit $x \in E$.
On dit que x est **unitaire** , ou **normé** , si $\|x\| = 1$.

Exemples 2

1. *Dans* $\mathbb{R}^2$, $x = (\frac{1}{\sqrt{2}}, \frac{-1}{\sqrt{2}})$ *est unitaire.*
2. *Dans* $\mathbb{R}^n$ *muni du produit scalaire canonique, les vecteurs de la base canonique sont unitaires.*
3. *Pour* $E = C^0(\mathbb{R}/2\pi Z)$ *l'e.v. des fonctions continues et* 2π*-périodiques sur* $\mathbb{R}$*, muni du produit scalaire :*

$$(f, g) \longmapsto \frac{1}{\pi} \int_0^{2\pi} f(x)\, g(x)\, dx,$$

les fonctions sin *et* cos *sont unitaires.*

Définition 3
Soit E un $\mathbb{R}$-e.v. euclidien. Soient $x, y \in E$.
On dit que x et y sont **orthogonaux** si l'on a $< x|y >= 0$. On note alors $x \perp y$.

Remarque 4 — *Comme* $\langle .|. \rangle$ *est symétrique, on a* $< x|y >= 0$ *si et seulement si* $< y|x >= 0$*. Donc la relation précédente est symétrique.*

Exemples 5

1. *Dans* $\mathbb{R}^2$, *on a* $x(\frac{1}{\sqrt{2}}, \frac{-1}{\sqrt{2}}) \perp (\frac{1}{\sqrt{2}}, \frac{1}{\sqrt{2}})$.
2. *Pour* $\mathbb{R}^n$ *muni du produit scalaire canonique, les vecteurs* $e_1, \ldots, e_n$ *de la base canonique sont orthogonaux deux à deux.*
3. *Pour* $E = C^0(\mathbb{R}/2\pi\mathbb{Z})$ *l'e.v. des fonctions continues et* 2π*-périodiques sur* $\mathbb{R}$*, muni du produit scalaire :*

$$(f, g) \longmapsto \frac{1}{\pi} \int_0^{2\pi} f(x)\, g(x)\, dx,$$

les fonctions sin *et* cos *sont orthogonales.*

Définition 6
Soit E un $\mathbb{R}$-e.v. euclidien. Soit A une partie de E.
On définit **l'orthogonal de** A, noté $A^\perp$, par :

$$A^\perp = \{x \in E \text{ tels que } x \perp y,\ \forall y \in A\}.$$

PROPOSITION 7
Soient E un $\mathbb{R}$-e.v. euclidien, et A une partie de E.
Alors $A^\perp$ est un sous-espace vectoriel de E.

Preuve — Soit A une partie de E.

- $A^\perp$ contient le vecteur nul puisque celui-ci est orthogonal à tout élément de A.
- Soient $x, y \in A^\perp$ et $\lambda \in \mathbb{R}$. Pour tout $y \in A$, on a :
$$\langle t|x+\lambda y\rangle = \langle t|x\rangle + \lambda\langle t|y\rangle = 0.$$
Ainsi, $x+\lambda y \in A^\perp$, donc $A^\perp$ est un sous-espace vectoriel de E.

□

EXEMPLES 8

1. *On a* $\{0\}^\perp = E$.
2. *On a* $E^\perp = \{0\}$. *En effet :*
   - $E^\perp$ *est un sous-e.v. de* E, *donc il contient* 0.
   - *Pour* $x \in E^\perp$, *alors on a* $x \perp x$, *donc* $||x||^2 =< x|x >= 0$. *Cela donne* $x = 0$, *donc* $E^\perp = \{0\}$.
3. *Soient* A, B *des sous-parties de* E. *Si* $A \subset B$, *alors* $B^\perp \subset A^\perp$. ◀
4. *Soit* $a \in E$ *non-nul. Alors le sous-e.v.* $\{a\}^\perp$ *est exactement le noyau de la forme linéaire non nulle* $\phi_a : x \mapsto < a|x >$. *Ainsi,* $\{a\}^\perp$ *est un hyperplan de* E. ◀

DÉFINITION 9
Soit E un $\mathbb{R}$-e.v. euclidien. Soient F, G deux sous-e.v. de E.
On dit que F et G sont **orthogonaux** si l'on a $x \perp y$, $\forall (x,y) \in F \times G$.

EXEMPLES 10

1. *Pour* F *un sous-e.v. de* E, *les sous-e.v.* F *et* $F^\perp$ *sont orthogonaux, puisque par définition les éléments de* $F^\perp$ *sont orthogonaux à tous les éléments de* F.
2. *Les sous-espace vectoriels* F *et* G *sont orthogonaux si, et seulement si,* $F \subset G^\perp$. ◀

THÉORÈME 11 (Théorème de Pythagore)
Soient E un $\mathbb{R}$-e.v. euclidien et $x, y \in E$.
Alors x et y sont orthogonaux si et seulement si $||x+y||^2 = ||x||^2 + ||y||^2$.

Preuve — Conséquence de la Proposition 27 et de la définition de l'orthogonalité. □

DÉFINITION 12
Soit E un $\mathbb{R}$-e.v. euclidien. Soit $(e_i)_{i\in I}$ une famille de vecteurs de E.
On dit que la famille $(e_i)_{i\in I}$ est une **famille orthogonale** si tous ses vecteurs sont deux à deux orthogonaux.
On dit que la famille $(e_i)_{i\in I}$ est une **famille orthonormée** (ou **orthonormale**) si tous ses vecteurs sont unitaires et deux à deux orthogonaux.

EXEMPLES 13

1. *Dans* $\mathbb{R}^2$ *muni du produit scalaire canonique, la famille* $((\frac{1}{\sqrt{2}}, \frac{-1}{\sqrt{2}}), (\frac{1}{\sqrt{2}}, \frac{1}{\sqrt{2}}))$ *est orthonormée.*
2. *Dans* $\mathbb{R}^n$ *muni de son produit scalaire canonique, la base canonique* $(e_1, \ldots, e_n)$ *est une famille orthonormée.*
3. *Pour* $E = C^0(\mathbb{R}/2\pi Z)$ *l'e.v. des fonctions continues et* 2π*-périodiques sur* $\mathbb{R}$, *muni du produit scalaire :* ◀
$$(f,g) \longmapsto \frac{1}{\pi}\int_0^{2\pi} f(x)\,g(x)\,dx,$$
la famille $\{x \mapsto \frac{1}{\sqrt{2}}, \quad x \mapsto \cos(x), \quad x \mapsto \cos(2x), \quad x \mapsto \cos(nx), \forall n \in \mathbb{N}^*,$
$x \longmapsto \sin(x), \quad x \longmapsto \sin(2x), \quad x \mapsto \sin(nx), \forall n \in \mathbb{N}^*\}$ *est orthonormée.*

PROPOSITION 14
Soit E un $\mathbb{R}$-e.v. euclidien. Soit $(x_1, x_2, \ldots, x_n)$ une famille de vecteurs de E.
Si cette famille est orthogonale, alors on a :
$$\left|\left|\sum_{i=1}^n x_i\right|\right|^2 = \sum_{i=1}^n ||x_i||^2.$$

Preuve — Par bilinéarité du produit scalaire, on a :

$$< \sum_{i=1}^{n} x_i | \sum_{i=1}^{n} x_i >= \sum_{1 \leqslant i,j \leqslant n} < x_i | x_j >= \sum_{i=1}^{n} < x_i | x_i >$$

puisque $< x_i | x_j >= 0$ si $i \neq j$. □

PROPOSITION 15
Soit E un $\mathbb{R}$-e.v. euclidien. Soit $(e_1, e_2, \ldots, e_n)$ une famille de vecteurs de E.
Si cette famille est orthonormée, alors on a :

1. Pour $x = \sum_{i=1}^{n} \lambda_i \, e_i \in \text{Vect}(e_1, \ldots, e_n)$, on a $\lambda_i =< e_i | x >$, $\forall 1 \leq i \leq n$.
2. La famille $(e_1, e_2, \ldots, e_n)$ est libre.

Preuve

1. Par linéarité à droite du produit scalaire, on a :

$$< e_i | x >= \sum_{j=1}^{n} \lambda_j \ < e_i | e_j >= \lambda_i.$$

2. Soit $(\lambda_i)_{1 \leqslant i \leqslant n} \in \mathbb{R}^n$ tel que $\sum_{k=1}^{n} \lambda_k \, e_k = 0$. Alors on :

$$0 = \quad < e_i | 0 >=< e_i | \sum_{i=k}^{n} \lambda_k \, e_k >= \sum_{i=k}^{n} \lambda_k < e_i | e_k >= \lambda_i, \ \forall i \in [\![1, n]\!].$$

Donc cette famille est bien libre.

□

Bases orthonormées

DÉFINITION 16
Soit E un $\mathbb{R}$-e.v. euclidien de dimension finie n. Soit $\mathcal{B} = (e_1, e_2, \ldots, e_n)$ une base de E.
On dit que la base $\mathcal{B}$ est une **base orthonormée** de E (ou b.o.n.) si la famille $(e_1, e_2, \ldots, e_n)$ est orthonormée.

EXEMPLE 17 — *Dans $\mathbb{R}^2$ muni du produit scalaire canonique, la famille $((\frac{1}{\sqrt{2}}, \frac{-1}{\sqrt{2}}), (\frac{1}{\sqrt{2}}, \frac{1}{\sqrt{2}}))$ est une base orthonormée.*

Dans $\mathbb{R}^n$ muni du produit scalaire canonique, la base canonique $(e_1, \ldots, e_n)$ est une base orthonormée.

PROPOSITION 18
Soit E un $\mathbb{R}$-e.v. euclidien de dimension finie n.
Alors E possède au moins une base orthonormée.

Preuve — Récurrence sur n la dimension de E. □

PROPOSITION 19
Soient E un $\mathbb{R}$-e.v. euclidien de dimension n, et $\mathcal{B} = (e_1, e_2, \ldots, e_n)$ une base orthonormée de E. Soient $x, y \in E$.
On a :

1. $x = \sum_{i=1}^{n} < x | e_i > \ e_i$.
2. Pour $x = \sum_{i=1}^{n} x_i \, e_i$ et $y = \sum_{i=1}^{n} y_i \, e_i$, on a

$$< x | y >= \sum_{i=1}^{n} x_i \, y_i \quad \text{et} \quad ||x||^2 = \sum_{i=1}^{n} x_i^2.$$

Preuve — L'égalité $x = \sum_{i=1}^{n} < x | e_i > \ e_i$ est une conséquence de la Proposition 15.

Pour $x = \sum_{i=1}^{n} x_i \, e_i$ et $y = \sum_{i=1}^{n} y_i \, e_i$, la bilinéarité du produit scalaire donne :

$$\langle x | y \rangle = \langle \sum_{i=1}^{n} x_i \, e_i | \sum_{j=1}^{n} y_j \, e_j \rangle = \sum_{i=1}^{n} \sum_{j=1}^{n} x_i y_j \langle e_i | e_j \rangle = \sum_{i=1}^{n} x_i \, y_i.$$

On en déduit alors que $||x||^2 = \langle x | x \rangle = \sum_{i=1}^{n} x_i^2$ □

COROLLAIRE 20
Soit E un $\mathbb{R}$-e.v. euclidien de dimension n, et $\mathcal{B} = (e_1, e_2, \ldots, e_n)$ une base orthonormée de E. Alors, on a :

- La fonction $f : x = x_1 e_1 + \ldots + x_n e_n \in E \mapsto (x_1, \ldots, x_n) \in \mathbb{R}^n$ isomorphisme d'espaces vectoriels de E sur $\mathbb{R}^n$.
- Si l'on munit $\mathbb{R}^n$ de sa structure euclidienne canonique, cet isomorphisme f conserve la norme et le produit scalaire.
- Pour $x, y \in E$ de coordonnées $(x_1, x_2, \ldots, x_n)$ et $(y_1, y_2, \ldots, y_n)$, on a :

$$d(x,y) = \|x - y\| = \left(\sum_{i=1}^{n} (x_i - y_i)^2 \right)^{1/2}.$$

Ainsi, un e.v. euclidien de dimension n s'étudie et se manipule de la même façon que $(\mathbb{R}^n, \langle . | . \rangle)$.

15.1.1 Procédé d'orthonormalisation de Schmidt

THÉORÈME 21
Soit E un $\mathbb{R}$-e.v. euclidien de dimension n. Soit $\mathcal{B} = (e_1, e_2, \ldots, e_n)$ une base de E.
Alors il existe une base orthonormée $(f_1, f_2, \ldots, f_n)$ de E telle que :

$$\text{Vect}(e_1, e_2, \ldots, e_p) = \text{Vect}(f_1, f_2, \ldots, f_p),\ \forall p \in [\![1, n]\!].$$

On peut construire cette base à l'aide d'un algorithme.

Preuve — Nous allons démontrer cela par récurrence sur p. Cette démonstration donne en fait un algorithme efficace qui permet de construire cette base. On l'appelle le **procédé d'orthonormalisation de Schmidt** .

- Le vecteur f_1 doit être un vecteur unitaire et colinéaire à e_1. On prend alors $f_1 = \frac{e_1}{||e_1||}$.
- Soit $1 \leq p \leq n$. Supposons avoir une famille $(f_1, f_2, \ldots, f_p)$ orthonormée telle que :

  $$\text{Vect}(e_1, e_2, \ldots, e_k) = \text{Vect}(f_1, f_2, \ldots, f_k),\ \forall k \in [\![1, p]\!].$$

  Comme on a $\text{Vect}(e_1, e_2, \ldots, e_p) = \text{Vect}(f_1, f_2, \ldots, f_p)$, tout vecteur de $\text{Vect}(e_1, e_2, \ldots, e_{p+1})$ peut s'écrire comme combinaison linéaire de $f_1, f_2, \ldots, f_p, e_{p+1}$.
  On cherche donc g_{p+1} orthogonal à $f_1, f_2, \ldots, f_p$, de la forme :

  $$g_{p+1} = e_{p+1} - \sum_{i=1}^{p} \lambda_i f_i.$$

  L'orthogonalité avec $f_1, \ldots, f_p$ donne :

  $$0 = < f_i | g_{p+1} > = < f_i | e_{p+1} > - \lambda_i,\ \forall 1 \leq i \leqslant p$$

  En prenant $\lambda_i = < f_i | e_{p+1} >$, le vecteur g_{p+1} est donc orthogonal à $f_1, \ldots, f_p$. Le vecteur g_{p+1} est aussi non nul puisque :

  $$e_{p+1} \notin \text{Vect}(e_1, e_2, \ldots, e_p) = \text{Vect}(f_1, f_2, \ldots, f_p).$$

  On pose alors $f_{p+1} = \frac{g_{p+1}}{||g_{p+1}||}$.
  La famille $(f_1, f_2, \ldots, f_{p+1})$ est alors une famille orthonormée, donc libre, dans $\text{Vect}(e_1, e_2, \ldots, e_{p+1})$. Comme elle possède $p+1$ vecteurs, c'est donc une base de ce sous-e.v. :

  $$\text{Vect}(f_1, f_2, \ldots, f_{p+1}) = \text{Vect}(e_1, e_2, \ldots, e_{p+1}).$$

□

EXEMPLE 22 — *On prend $E = \mathbb{R}^3$, muni du produit scalaire (voir Exemple 5) :*

$$< (x,y,z) | (x',y',z') > = x\,x' + y\,y' + z\,z' + \frac{1}{2}(x\,y' + x'\,y + x\,z' + x'\,z + y\,z' + y'\,z)$$

dont la norme associée est :

$$||(x,y,z)|| = \sqrt{x^2 + y^2 + z^2 + x\,y + x\,z + y\,z}.$$

A partir de la base canonique (e_1, e_2, e_3), construisons une base orthonormée (f_1, f_2, f_3) de $\mathbb{R}^3$ avec le procédé de Schmidt.

- *Le vecteur $e_1 = (1,0,0)$ est unitaire, donc on peut prendre $f_1 = e_1$.*
- *Cherchons g_2 orthogonal à f_1 de la forme : $g_2 = e_2 - \lambda\, f_1$.*
  On a $< f_1 | g_2 > = < f_1 | e_2 > - \lambda$, donc il suffit de prendre $\lambda = < f_1 | e_2 > = \frac{1}{2}$. Cela donne :

$$g_2 = \left(-\frac{1}{2}, 1, 0 \right) \quad \text{et} \quad f_2 = \frac{1}{\sqrt{3}}(-1, 2, 0).$$

- *Cherchons g_3 orthogonal à f_1 et f_2 de la forme : $g_3 = e_3 - \lambda f_1 - \mu f_2$. Il suffit de prendre :*

$$\lambda =< f_1|e_3 >= \frac{1}{2} \quad et \quad \mu =< f_2|e_3 >= \frac{1}{2\sqrt{3}}.$$

Cela donne :

$$g_3 = \left(-\frac{1}{3}, -\frac{1}{3}, 1\right) \quad et \quad f_3 = \frac{1}{\sqrt{6}}(-1, -1, 3).$$

REMARQUE 23 — *Soit E un $\mathbb{R}$-e.v. euclidien de dimension n, $\mathcal{B} = (e_1, e_2, \cdots, e_n)$ une base de E,et $\mathcal{C} = (f_1, f_2, \cdots, f_n)$ une base orthonormée obtenue avec le procédé de Schmidt.*
Soit P La matrice de passage de $\mathcal{B}$ vers $\mathcal{C}$. Alors la matrice P est triangulaire supérieure et ses éléments diagonaux sont non-nuls (f_i est une combinaison linéaire de $e_1, \ldots, e_i$).

15.1.2 Supplémentaire orthogonal

PROPOSITION-DÉFINITION 24
Soit E un $\mathbb{R}$-e.v. euclidien de dimension n. Soit F un sous-e.v. de E. Alors, on a :

1. F et $F^\perp$ sont supplémentaires dans E ;
2. $\dim\left(F^\perp\right) + \dim(F) = \dim(E)$;
3. $(F^\perp)^\perp = F$.

Le sous-ev $F^\perp$ est appelé le **supplémentaire orthogonal** de F. On écrit alors $E = F \overset{\perp}{\oplus} F^\perp$, pour signifier que la somme directe entre F et $F^\perp$ est orthogonale.

Preuve —

1. - On a $F \cap F^\perp = \{0\}$ puisqu'un vecteur x dans F et $F^\perp$ est orthogonal à lui-même, ce qui implique que $x = 0$.
   - D'après le théorème de Scmhidt, il existe $\mathcal{B} = (e_1, e_2, \ldots, e_p)$ une base orthonormée de F. Soit $x \in E$. On pose $y = \sum_{i=1}^{p} < e_i|x > e_i$. On a donc :
   $$< e_i|y >=< e_i|x >, \forall i \in [\![1, p]\!].$$
   En posant $z = x - y$, le vecteur z est ainsi orthogonal à $e_1, \ldots, e_p$, donc à $\text{Vect}(e_1, \ldots, e_p) = F$. Ainsi, on a $x = y + z$, avec $y \in F$ et $z \in F^\perp$. Donc F et $F^\perp$ sont supplémentaires dans E (le vecteur y est appelé projeté orthogonal de x sur F).
2. Conséquence du fait que F et $F^\perp$ sont supplémentaires.
3. Comme tout élément de F est orthogonal à tout élément de $F^\perp$ on a $F \subset (F^\perp)^\perp$. Comme on a :
$$\dim\left(\left(F^\perp\right)^\perp\right) = n - \dim\left(F^\perp\right) = \dim(F),$$
on en déduit que $F = (F^\perp)^\perp$.

□

PROPOSITION 25
Soit E un $\mathbb{R}$-e.v. euclidien de dimension n.
Alors, toute famille orthonormée de E peut être complétée en une base orthonormée de E.

Preuve — Soit $(e_1, e_2, \ldots, e_p)$ une famille orthonormée de E. On pose $F = \text{Vect}(e_1, \ldots, e_p)$. Alors $F^\perp$ est le supplémentaire orthogonal de F dans E. D'après le procédé de Schmidt, $F^\perp$ possède une base orthonormée. Notons-la $(e_{p+1}, \ldots, e_n)$. On obtient alors que $(e_1, e_2, \ldots, e_n)$ est une famille orthonormée de E à $n = \dim(E)$ vecteurs, donc c'est une b.o.n. de E. □

REMARQUES 26

- *Comme nous l'avons vu dans la démonstration précédente, si $E = F \overset{\perp}{\oplus} G$, la réunion d'une base orthonormée de F et d'une base orthonormée de G est une base orthonormée de E.*
- *Réciproquement, si $\mathcal{B} = (e_1, e_2, \ldots, e_n)$ est une base orthonormée de E et si $p \leqslant n$, alors les sous-espaces vectoriels :*
$$F = \text{Vect}\{e_1, \ldots, e_p\} \quad et \quad G = \text{Vect}\{e_{p+1}, \ldots, e_n\}$$
sont deux sous-espaces vectoriels orthogonaux et supplémentaires dans E.

15.1.3 Équations d'un hyperplan

Définition 27
Soit E un $\mathbb{R}$-e.v. euclidien. Soient H un hyperplan de E et $u \in E$. On dit que le vecteur u est un **vecteur normal** l'hyperplan H si u est non-nul et si u est orthogonal à H.

Proposition 28
Soit E un $\mathbb{R}$-e.v. euclidien de dimension n. Soient $\mathcal{B} = (e_1, \ldots, e_n)$ une b.o.n de E, H un hyperplan de E et $a \in E$ non-nul, avec $a = a_1 e_1 + \ldots + a_n e_n$.
Alors, les propriétés suivantes sont équivalentes :

(i) $H = \mathrm{Ker}\,(\phi_a)$;

(ii) $H = \{x = x_1 e_1 + \ldots x_n e_n, \text{ tels que } \sum_{i=1}^{n} a_i\, x_i = 0.\}$

(iii) a est un vecteur normal à H.

Théorème 29 (**Théorème de représentation de Riesz**)
Soit E un $\mathbb{R}$-e.v. euclidien de dimension n. Soit $f : E \to \mathbb{R}$ une forme linéaire sur E.
Alors il existe un unique $a \in E$ tel que $f = \phi_a : x \mapsto \langle x, a\rangle$.
Pour B une b.o.n. de E et $(a_1, \ldots, a_n)$ les coordonnées de a dans la base B, on a :

$$f(x) = \langle x, a\rangle = \sum_{i=1}^{n} a_i x_i.$$

Preuve

Unicité. Si $a, b \in E$ sont tels que $\phi_a = \phi_b$, alors on a :

$$\forall x \in E\,,\ < a|x > = < b|x > \ \text{ c'est-à-dire } \forall x \in E\,,\ < a - b|x > = 0.$$

Le vecteur $a - b$ est donc orthogonal à tous les éléments de E, et par conséquent il est nul.

Existence.

- Si $f = 0$, alors le vecteur nul convient.
- Sinon, le noyau de f est un hyperplan H. Comme $\dim(H) = n - 1$, l'orthogonal $H^{\perp}$ de H est de dimension 1. Pour $a \in H^{\perp}$ non-nul, on a alors $H = \mathrm{Ker}\,(\phi_a)$. Ces deux formes linéaires non nulles ont le même noyau. Elles sont donc proportionnelles : il existe $\lambda \in \mathbb{R}$ tel que $f = \lambda\, \phi_a = \phi_{\lambda\, a}$.

□

Remarque 30 — *On peut aussi démontrer cette proposition en utilisant le théorème du rang. La fonction $f : a \in E \mapsto \phi_a \in \mathcal{L}(E, \mathbb{R})$ est une application linéaire de E dans $\mathcal{L}(E, \mathbb{R})$, l'espace vectoriel des formes linéaires sur E.*

- *Son noyau est réduit à 0, car si $\phi_a = 0$, alors tous les vecteurs de E sont orthogonaux au vecteur a et donc $a = 0$. Donc f est injective.*
- *Comme $\dim(E) = n = \dim(\mathcal{L}(E, \mathbb{R}))$, f est donc isomorphisme. Ainsi, toute forme linéaire s'écrit de façon unique sous la forme ϕ_a.*

15.2 Projections orthogonales

Projections vectorielles

Définition 31
Soit E un $\mathbb{R}$-e.v. euclidien de dimension finie. Soit F un sous-e.v. de E. Soit P la projection sur F parallèlement à $F^{\perp}$.
La projection P est appelée **projection orthogonale** sur F. On la note $P = p_F$.

Proposition 32
Soient E un $\mathbb{R}$-e.v. euclidien de dimension finie, et F un sous-e.v. de E. Soit $\mathcal{B} = (e_1, e_2, \ldots, e_p)$ une base orthonormée de F. Alors, pour tout $x \in E$, on a :

$$p_F(x) = \sum_{i=1}^{p} < x|e_i > \ e_i.$$

Preuve — On a $x = p_F(x) + (x - p_F(x))$. Comme $x - p_F(x) \in F^\perp$, on a $\langle x - p_F(x)|e_i\rangle = 0$, c'est-à-dire $< x|e_i >=< p_F(x)|e_i >$. Comme $p_F(x) \in F$, on obtient le résultat. □

Remarque 33 — *Ainsi, il suffit d'avoir une base orthonormée du sous-espace F, par exemple avec l'algorithme de Gram-Schmidt, pour pouvoir calculer très facilement la projection orthogonale sur F.*

15.2.1 Distance à un sous-espace

Définition 34
Soit E un $\mathbb{R}$-e.v. normé. Soient A une partie non-vide de E, et $x \in E$. On définit la **distance** de x à A comme :

$$d(x, A) = \inf_{y\in A} d(x, y) = \inf_{y\in A} \|x - y\|.$$

L'existence de cette quantité $d(x, A)$ vient du fait que $\big\{d(x, y),\ y \in A\big\}$ est une partie non vide de $\mathbb{R}_+$.

Proposition 35
Soit E un $\mathbb{R}$-e.v. euclidien de dimension finie. Soient F un sous-ev de E et $x \in E$.
Alors, la distance de x à F vérifie $d(x, F) = \|x - p_F(x)\|$, où $p_F(x)$ est le projeté orthogonal de x sur F.
De plus, $p_F(x)$ et l'unique vecteur $y \in F$ tel que $d(x, F) = \|x - y\|$.

Preuve — Soit $y \in F$. D'après le théorème de Pythagore, on a :

$$d(x, y)^2 = \|x - y\|^2 = \|(x - p_F(x)) + (p_F(x) - y)\|^2 = \|x - p_F(x)\|^2 + \|p_F(x) - y\|^2 \geq \|x - p_F(x)\|^2 = d(x, p_F(x))^2.$$

On a donc $d(x, F) \geqslant d(y, p_F(x))$. Et comme $p_F(y) \in \mathbb{F}$, on obtient $d(x, F) = d(x, p_F(x))$.

De plus, on a $d(x, y) = d(x, F) = d(x, p_F(x))$ si et seulement si $\|p_F(x) - y\|^2 = 0$, ce qui est équivalent à $y = p_F(x)$. □

Remarque 36 —
• *Pour chercher à minimiser $\|x - y\|$ avec $y \in F$, on peut chercher à minimiser $\|x - y\|^2$, comme on le fait dans la preuve de la proposition.*
Comme $\|.\|$ est une norme euclidienne, on a $\|x - y\|^2 = \langle x - y, x - y\rangle$, et on peut utiliser les propriétés du produit scalaire dans les calculs.
• *Si $(e_1, \dots, e_r)$ est une base orthonormée de F, on a $p_F(x) = \sum_{k=1}^r \langle x, e_k\rangle e_k$.*
Ainsi, on a

$$d(x, F)^2 = \|x - p_F(x)\|^2 = \|x - \sum_{k=1}^{r} \langle x, e_k\rangle e_k\|^2.$$

Si on complète la famille orthonormée $(e_1, \dots, e_r)$ en une base orthonormée $(e_1, \dots, e_n)$ de E, on a alors $x = \sum_{i=1}^n \langle x, e_i\rangle e_i$.
Cela donne donc

$$d(x, F)^2 = \|x - \sum_{k=1}^{r} \langle x, e_k\rangle e_k\|^2 = \|\sum_{i=r+1}^{n} \langle x, e_i\rangle e_i\|^2 = \sum_{i=r+1}^{n} \langle x, e_i\rangle^2.$$

Avec une base orthonormée de F complétée en une base orthonormée de E, on obtient une expression très simple de la distance d'un vecteur x à F.

Chapitre 16 Dénombrement, sommabilité

Table des matières du chapitre

16.1 L'ensemble $\mathcal{P}(\Omega)$

On revoit rapidement dans ce chapitre les éléments de théorie des ensembles et de dénombrement qui sont nécessaires en probabilités.
On suppose dans toute la suite que Ω est un ensemble.

Un **sous-ensemble** ou une **partie** de Ω est un ensemble dont tous les éléments sont dans Ω.

Axiome *Soit Ω un ensemble et P une propriété sur les éléments de Ω. Alors la collection des éléments de Ω qui vérifient la propriété P forme un ensemble. On note cet ensemble :*

$$\{x \in \Omega \text{ tq } P(x)\} \quad \text{ou} \quad \{x \in \Omega \mid P(x)\} \quad \text{ou} \quad \{x \in \Omega\ ,\ P(x)\}.$$

Remarque 1 — *Si $A \subset \Omega$, on définit $\overline{A}$ (ou A^C), le **complémentaire de A dans Ω** comme*

$$\overline{A} = \{\omega \in \Omega |\ \omega \notin A\}.$$

*Ou encore, pour $w \in \Omega$, le **singleton** $\{\omega\}$ est $\omega = \{\alpha \in \Omega \mid \alpha = \omega\}$.*

Axiome *La collection des parties d'un ensemble Ω est encore un ensemble noté $\mathcal{P}(\Omega)$:*

$$A \in \mathcal{P}(\Omega) \iff A \subset \Omega.$$

$\mathcal{P}(\Omega)$ est l'ensemble de toutes les parties de Ω.

Exemple 2 —

1. *Si $\Omega = \emptyset$, alors $\mathcal{P}(\emptyset) = \{\emptyset\}$ et donc a un élément. Et $\mathcal{P}(\{\emptyset\}) = ?$*
   On peut voir les ensembles comme des boîtes.
   L'ensemble vide est une boîte vide. Et $\{\emptyset\}$ est une boîte qui contient une boîte vide.
2. *Si $\Omega = \{1, 2, 3\}$, alors*

$$\mathcal{P}(\Omega) = \{\emptyset, \{1\}, \{2\}, \{3\}, \{1, 2\}, \{1, 3\}, \{2, 3\}, \{1, 2, 3\}\}.$$

Remarque 3 — *Pour tout ensemble Ω, on a toujours $\emptyset \in \mathcal{P}(\Omega)$ et $\Omega \in \mathcal{P}(\Omega)$.*

Définition 4
Soit I un ensemble et $(\Omega_i)_{i\in I}$ une famille d'ensembles.
Le **produit cartésien** de la famille $(\Omega_i)_{i\in I}$ est l'ensemble $\prod_{i\in I} \Omega_i$ des familles $(\omega_i)_{i\in I}$, avec $\omega_i \in \Omega_i$.

Exemple 5 — *Par exemple, $(v_n)_{n\in\mathbb{N}}$ est un élément de $\Omega^{\mathbb{N}}$.*

Remarque 6 — *Sur $\mathcal{P}(E)$ on définit le complémentaire d'une partie, ainsi que l'union et l'intersection d'une famille quelconque de parties. On a par exemple*

$$\Omega = \bigcup_{\omega\in\Omega} \{w\}.$$

On peut montrer que les unions et les intersections sont associatives et commutatives pour une famille $(A_i)_{i\in I}$ de parties de Ω.
La distributivité entre l'union et l'intersection est valable pour des familles quelconques de parties. En particulier, on a

$$A \cap \left(\bigcup_{i\in I} A_i\right) = \bigcup_{i\in I}(A \cap A_i) \text{ et } A \cup \left(\bigcap_{i\in I} A_i\right) = \bigcap_{i\in I}(A \cup A_i).$$

On a aussi

$$\bigcup_{i\in I}\left(\bigcap_{j\in J} A_{i,j}\right) \subset \bigcap_{j\in J}\left(\bigcup_{i\in I} A_{i,j}\right).$$

En effet, si $x \in \bigcup_{i\in I}\left(\bigcap_{j\in J} A_{i,j}\right)$, alors il existe $i_0 \in I$ tel que $\forall j \in J$, $x \in A_{i_0,j}$.
Donc $\forall j \in J$ on a $x \in \bigcup_{i\in I} A_{i,j}$ (l'indice i_0 convient).
Finalement, on a prouvé $x \in \bigcap_{j\in J}\left(\bigcup_{i\in I} A_{i,j}\right)$, ce qui termine la preuve.
En général il n'y a pas égalité.

Exemple 7 — *Soit $A_{i,j} = [j + i - 1, i + j[$. On a*

$$\bigcup_{i\in\mathbb{Z}}\left(\bigcap_{j\in\mathbb{N}} A_{i,j}\right) = \bigcup_{i\in\mathbb{Z}}\left(\bigcap_{j\in\mathbb{N}} [j + i - 1, i + j[\right) = \emptyset$$

et

$$\bigcap_{j\in\mathbb{N}}\left(\bigcup_{i\in\mathbb{Z}} A_{i,j}\right) = \bigcap_{j\in\mathbb{N}}\left(\bigcup_{i\in\mathbb{Z}} [j + i - 1, i + j[\right) = \mathbb{R}$$

Nous allons définir sur un ensemble particulier (un espace probabilisé) des fonctions particulières (les variables aléatoires). Avant de définir ces objets, les fonctions indicatrices en sont un exemple fondamental. Les fonctions indicatrices permettent de définir des opérations ensemblistes (union, intersection,...) en terme algébriques (produit, somme...).

Définition 8 (**Fonction indicatrice**)
Soit $A \subset \Omega$. On appelle **fonction indicatrice** de A, notée $\mathbb{1}_A$ (ou χ_A), la fonction $\mathbb{1}_A : \Omega \to \mathbb{R}$ définie par

$$\mathbb{1}_A(x) = \begin{cases} 1 \text{ si } x \in A \\ \\ 0 \text{ sinon.} \end{cases}$$

Proposition 9
Soient A, $B \in \mathcal{P}(\Omega)$. On a

1. $A \subset B$ si et seulement si $\mathbb{1}_A \leq \mathbb{1}_B$.
2. $\mathbb{1}_{\overline{A}} = 1 - \mathbb{1}_A$.
3. $\mathbb{1}_{A\cap B} = \mathbb{1}_A.\mathbb{1}_B$.
4. $\mathbb{1}_{A\cup B} = \mathbb{1}_A + \mathbb{1}_B - \mathbb{1}_A.\mathbb{1}_B$

Exemple 10 —

1. *Écrire les fonctions indicatrices de $A \setminus B$ et de $A\Delta B$ en fonction de celles de A et B.*
2. *Soit une famille $(A_i)_{i\in[\![1,n]\!]}$ de parties non vides de Ω. On a :*

$$\mathbb{1}_{\bigcap\limits_{i\in[\![1,n]\!]} A_i} = \prod_{i\in[\![1,n]\!]} \mathbb{1}_{A_i} \text{ et } \mathbb{1}_{\bigcup\limits_{i\in[\![1,n]\!]} A_i} = 1 - \prod_{i\in I}(1 - \mathbb{1}_{A_i})$$

Proposition 11
Soit $n \in \mathbb{N}^*$. Une famille $(A_i)_{i\in[\![1,n]\!]}$ de parties non vides de Ω est une partition de Ω si et seulement si

$$\mathbb{1}_\Omega = \sum_{i=1}^{n} \mathbb{1}_{A_i}.$$

Preuve — On a les équivalences

1. Soit $i \neq j$. Alors $x \in A_i \cap A_j$ ssi
$$\sum_{i\in I} \mathbb{1}_{A_i}(x) \geq 2.$$
2. De plus, $x \in \bigcup_{i\in j} A_i$ ssi il existe $i \in J$ tel que $x \in A_i$ ssi $\sum_{i\in I} \mathbb{1}_{A_i}(x) \geq 1$.

Ceci montre que Ω est l'union disjointe de la famille $(A_i)_{i\in I}$ ssi $\mathbb{1}_\Omega = \sum_{i\in I} \mathbb{1}_{A_i}$.

□

Corollaire 12
Une famille $(A_i)_{i\in[\![1,n]\!]}$ de parties non vides de Ω est une partition de Ω si et seulement si

$$\forall A \in \mathcal{P}(\Omega),\ \mathbb{1}_A = \sum_{i\in I} \mathbb{1}_{A_i\cap A}.$$

16.2 Cardinal d'un ensemble

Définition 13
On dit que deux ensembles Ω et Ω' ont **même cardinal**, et on note $|\Omega| = |\Omega'|$, s'il existe une bijection de Ω dans Ω'.
Si Ω est en bijection avec $\{1, \ldots, n\}$, on dit que Ω est un ensemble fini de **cardinal** n et l'on note $|\Omega| = n$ ou $\mathrm{card}\,(\Omega) = n$.
Par convention, $\mathrm{card}\,(\emptyset) = 0$.
Un ensemble de cardinal infini est un ensemble qui n'est pas de cardinal fini. On écrira $|\Omega| = \infty$.

Remarque 14 — *Cela nous donne une première technique majeure pour calculer le cardinal Ω d'un ensemble. On définit une bijection g entre Ω et un ensemble Ω' dont on connaît le cardinal.*

Exemple 15 — *On cherche le cardinal de Ω l'ensemble des suites de 4 chiffres (u_0, u_1, u_2, u_3) à valeurs dans $[\![0, 9]\!]$ telles que $u_1 - u_0 = 2$, $u_2 - u_1 = 1$ et $u_3 - u_2 = 3$.*
On définit l'application

$$\varphi : [\![0, 3]\!] \to \Omega,\ k \mapsto (k, k + 2, k + 3, k + 6).$$

On a bien $\varphi(k) \in \Omega$. φ est injective car $\varphi(k) = \varphi(k')$ implique sur le premier terme $k = k'$. Et φ est surjective car si $(u_0, u_1, u_2, u_3) \in \Omega$, alors $u_3 = u_0 + 6 \leq 9$ implique $u_0 \leq 3$ et $\varphi(u_0) = (u_0, u_1, u_2, u_3)$.
On en déduit que $|\Omega| = 4$.

Proposition 16
Soit Ω un ensemble fini de cardinal n.
Si $F \subset \Omega$, alors F est fini et $\mathrm{card}\,(F) \leq \mathrm{Card}\,(\Omega)$.
De plus, $\mathrm{Card}\,(\Omega) = \mathrm{Card}\,(F)$ si et seulement si $F = \Omega$.

Preuve — Si Ω est de cardinal n, on pose $\Omega = \{\omega_1, \cdots, \omega_n\}$. On construit par récurrence une suite (f_n)

1. Si $F \neq \emptyset$, alors $f_0 = \omega_{\inf\{k\ |w_k\in F\}}$
2. Si f_{k-1} est construit et si $F \setminus \{f_0, \cdots, f_{k-1}\} \neq \emptyset$, alors $f_k = \omega_{\inf\{k\ |w_k\in F\setminus\{f_0,\cdots,f_{k-1}\}\}}$
3. Si f_{k-1} est construit et si $F \setminus \{f_0, \cdots, f_{k-1}\} = \emptyset$, alors $F = \{f_0, \cdots, f_{k-1}\}$ et $|F| = k$. Et on s'arrête.

Tous les éléments de Ω sont parcourus ssi $n = k$ et $F = \Omega$. □

Remarque 17 — *On en déduit alors que $\mathbb{N}$ n'est pas un ensemble fini. Par exemple, $\mathbb{N}$ est en bijection avec les nombres pairs en prenant $k \mapsto 2k$, donc $\mathbb{N}$ a le même cardinal qu'un sous-ensemble strict, ce qui n'est pas possible pour un ensemble fini.*

Définition 18
On dit qu'un ensemble Ω est

- **dénombrable** s'il existe une bijection de Ω dans $\mathbb{N}$.
- au plus dénombrable si Ω est fini ou dénombrable.

Proposition 19
Soient Ω et Ω' deux ensembles.

1. S'il existe une surjection de Ω dans Ω', alors on a $|\Omega| \geq |\Omega'|$.
2. S'il existe une injection de Ω dans Ω', alors on a $|\Omega| \leq |\Omega'|$.

Proposition 20
Soit Ω un ensemble. Il n'existe pas d'injection de $\mathcal{P}(\Omega)$ dans Ω.

Preuve — On raisonne par l'absurde : supposons qu'il existe une injection $f : \mathcal{P}(\Omega) \to \Omega$, $A \mapsto f(A)$.
Alors $\mathcal{A} = \{A \in \mathcal{P}(\Omega) \mid f(A) \notin A\}$ est un sous-ensemble de $\mathcal{P}(\Omega)$ et $B = f(\mathcal{A})$ est un sous-ensemble de Ω, c'est-à-dire un élément de $\mathcal{P}(\Omega)$. On peut donc calculer l'image de B par f. On a deux possibilités

1. Soit $f(B) \notin B$, donc $B \in \mathcal{A}$. On en déduit $f(B) \in f(\mathcal{A}) = B$, ce qui est contradictoire.
2. Soit $f(B) \in B$, donc $B \notin \mathcal{A}$. De plus, comme $f(\mathcal{A}) = B$, il existe $C \in \mathcal{A}$ tel que $f(B) = f(C)$, mais cela contredit l'injectivité de f.

On en déduit qu'il n'existe pas d'injection de $\mathcal{P}(E)$ dans E. □

Remarque 21 — *La proposition ici est pour Ω de cardinal quelconque. On peut donc l'appliquer à $\mathbb{N}$: il n'existe pas de bijection de $\mathbb{N}$ dans $\mathcal{P}(\mathbb{N})$.*
Tous les ensembles de cardinal infini ne sont donc pas nécessairement dénombrables.
Nous allons souvent étudier les probabilités sur des ensembles Ω dénombrables et nous devrons étudier $\mathcal{P}(\Omega)$, qui n'est plus dénombrable.
Par contre les opérations d'unions, produits d'ensembles dénombrables restent dénombrables, comme nous le rappelons ci-dessous.

Proposition 22
Soit Ω un ensemble et A une partie finie de Ω. Alors

$$\operatorname{card} A = \sum_{x \in \Omega} \mathbb{1}_A(x)$$

Proposition 23
Si A et B sont deux sous-ensembles finis de Ω, alors $A \cup B$ est fini et

$$\operatorname{card}(A \cup B) = \operatorname{Card}(A) + \operatorname{Card}(B) - \operatorname{Card}(A \cap B).$$

Corollaire 24
Soit Ω un ensemble fini et $(A_i)_{i \in I}$ une partition de Ω. Alors on a

$$|\Omega| = \sum_{i \in I} |A_i|.$$

Preuve — Les A_i sont deux à deux distincts non vides, donc $|I| \leq |\Omega|$: pour tout i, on choisit un élément $a_i \in A_i$ et l'application $i \mapsto a_i$ est une injection.
Puis on procède par récurrence sur $|I|$, l'étape $|I| = 1$ implique $A_1 = \Omega$ et l'hérédité résulte immédiatement de la proposition précédente. □

Remarque 25 —
• *Cela nous donne une autre méthode fondamentale pour calculer le cardinal d'un ensemble Ω.*
On cherche une partition de $(A_i)_{i \in I}$ de Ω telle que tous les A_i sont de cardinal connu. La proposition ci-dessus permet ainsi de calculer le cardinal de Ω.
En particulier, pour $f : \Omega \to \Omega'$ une fonction surjective, alors $\left(f^{-1}(\omega')\right)_{\omega' \in \Omega'}$ est une partition de Ω.
Ainsi, si Ω est un ensemble fini, on a $|\Omega| = \sum_{\omega' \in \Omega'} \left|f^{-1}(\omega')\right|$.

Proposition 26
Soient Ω et Ω' deux ensembles finis. Alors $\Omega \times \Omega'$ est fini et

$$\operatorname{card}(\Omega \times \Omega') = \operatorname{Card}(\Omega)\operatorname{Card}(\Omega').$$

Preuve — Il suffit d'écrire $\Omega \times \Omega'$ comme l'union disjointe $\cup_{y \in \Omega'} \Omega \times \{y\}$. □

PROPOSITION 27
Soient Ω et Ω' deux ensembles finis de cardinaux n et p.
Alors l'ensemble des applications de Ω dans Ω', Ω'^{Ω}, est un ensemble fini de cardinal p^n.

Preuve — L'application qui à $f : \Omega \to \Omega'$ associe le n-uplet $(f(x_1), \ldots, f(x_n))$ est une bijection. Or Ω'^n est de cardinal p^n d'après la proposition 26, d'où le résultat. □

PROPOSITION 28
Soit Ω un ensemble fini de cardinal n.
Alors $\mathcal{P}(\Omega)$ est fini, de cardinal 2^n.

Preuve — L'application $\psi : \mathcal{P}(\Omega) \to \{0,1\}^{\Omega}$, $A \mapsto \mathbb{1}_A$ est bijective. □

16.3 ENSEMBLES DÉNOMBRABLES

REMARQUE 29 — *Soit Ω un ensemble dénombrable et $\varphi : \mathbb{N} \to \Omega$ une bijection.*
On pose $\varphi(n) = \omega_n$. Alors, les éléments de Ω peuvent être indexés par $\mathbb{N}$: $\Omega = \{\omega_n\}_{n\in\mathbb{N}}$.
*On dit que φ est une **énumération** de Ω.*

PROPOSITION 30
Soit Ω un ensemble dénombrable.
Alors toute partie A de Ω est finie ou dénombrable.

COROLLAIRE 31
Toute partie de $\mathbb{N}$ est soit de cardinal fini, soit dénombrable.

PROPOSITION 32
Soient Ω et Ω' deux ensembles dénombrables. Alors on a

1. $\Omega \cup \Omega'$ est dénombrable.
2. $\Omega \times \Omega'$ est dénombrable.

Preuve — On écrit $\Omega = (\omega_n)_{n\in\mathbb{N}}$.

1. On pose $\Omega'' = \Omega' \setminus \Omega$ et $\Omega \cup \Omega' = \Omega \cup \Omega''$.
   Si $|\Omega''| = p$, on numérote les éléments de Ω' par ω''_0, ..., ω''_{p-1} et ainsi $\varphi : \mathbb{N} \to \Omega \cup \Omega'$ définie par
   $$\varphi(k) = \begin{cases} \omega''_k \text{ si } k \in [\![0, p-1]\!] \\ \\ \omega_{k-p} \text{ si } k \geq p \end{cases}$$
   est bijective.
   Si Ω'' est dénombrable, on écrit $\Omega'' = (\omega''_n)_{n\in\mathbb{N}}$ et l'application $\varphi : \mathbb{N} \to \Omega \cup \Omega'$ définie par
   $$\varphi(k) = \begin{cases} \omega_{k/2} \text{ si } k \text{ pair} \\ \\ \omega''_{(k-1)/2} \text{ sinon} \end{cases}$$
   est bijective.
2. Si $\varphi : \Omega \to \mathbb{N}$ et $\psi : \Omega' \to \mathbb{N}$, alors $\varphi \times \psi : \Omega \times \Omega' \to \mathbb{N} \times \mathbb{N}$ est bijective.
   L'application $f : \mathbb{N} \times \mathbb{N} \to \mathbb{N}$ définie come suit est bijective :
   $$f : \begin{cases} \mathbb{N}^2 & \to & \mathbb{N} \\ (p,q) & \mapsto & 2^p(2q+1) - 1 \end{cases}.$$
   L'application φ est bijective : d'après la décomposition en facteur premier, tout nombre $n \in \mathbb{N}^*$ se décompose de manière unique $2^p m$ avec m impair et donc $q = \dfrac{m-1}{2}$. Donc il existe un unique couple (p,q) tel que $\varphi(p,q) = n-1$.

□

REMARQUE 33 —
• *La première partie de la preuve du 1) correspond à l'histoire suivante (hôtel de Hilbert) : Un hôtel possède une infinité dénombrable de chambres, et il est complet. Un car arrive avec 60 passagers.*
La personne de la réception répond qu'il pas de problème. On décale tout le monde de soixante chambres, et les 60 premières seront libres.
• *La seconde partie de la preuve correspond au cas où chacun client se décale de sorte que seules les chambres paires soient occupées, ce qui laisse donc un nombre infini de chambres libres.*

PROPOSITION 34
Soit Ω un ensemble. Pour tout $n \geq 0$ soit A_n une partie de Ω dénombrable.
Alors, $\cup_{n\geq 0} A_n$ est dénombrable.
Une union dénombrable d'ensembles dénombrables est dénombrable.

REMARQUE 35 — *On déduit des résultats de ce chapitre que $\mathbb{Z}$, $\mathbb{Q}$ sont dénombrables.*
Ce n'est pas le cas de $\{0,1\}^{\mathbb{N}}$ (en bijection avec $\mathcal{P}(\mathbb{N})$), de $\mathbb{R}$, de $\mathbb{C}$, ni des intervalles $[a,b]$ avec $a<b$.

EXEMPLE 36 —

1. *Soit $f : I \to \mathbb{R}$ une fonction monotone (croissante ou décroissante). Alors l'ensemble des points de discontinuité de f est au plus dénombrable.*
   En effet, on suppose que f est croissante. Alors f est discontinue en a ssi $]\lim_{a^-} f, \lim_{a^+} f[$ est un intervalle non-vide. Un intervalle non-vide de $\mathbb{R}$ contient un rationnel. Comme f est croissante, les points de discontinuité a sont donc en bijection avec une partie $\mathbb{Q}$. D'où le résultat.
2. *L'ensemble des racines des polynômes à coefficients entiers est dénombrable.*
   L'ensemble $\mathbb{Z}_n[X]$ des polynômes à coefficients entiers de degré $\leq n$ est en bijection avec $\mathbb{Z}^{n+1}$, qui est dénombrable (produit fini d'ensembles dénombrables).
   Chaque polynôme de $\mathbb{Z}_n[X]$ a au plus n racines, donc l'ensemble des racines des polynômes de $\mathbb{Z}_n[X]$ est dénombrable (réunion finie d'ensembles dénombrables).
   Et on a $\mathbb{Z}[X] = \cup_{n\geq 0}\mathbb{Z}_n[X]$ (réunion dénombrable). Donc l'ensemble des racines des polynômes de $\mathbb{Z}[X]$ est une réunion dénombrable d'ensembles dénombrables, ce qui est dénombrable.
   Un nombre réel qui n'est pas racine d'un polynôme à coefficients entiers est appelé un nombre transcendant. Ainsi, la majorité des nombres réels sont transcendants, même on en connait très peu dans les faits.

16.4 COEFFICIENTS BINOMIAUX, NOMBRES D'ARRANGEMENTS

DÉFINITION 37
Soit Ω un ensemble de cardinal n. Soit $p \geq 0$.
On appelle p-combinaison de Ω toute partie de Ω de cardinal p.

REMARQUE 38 — *On peut montrer que le nombre de p-combinaisons d'un ensemble de cardinal n ne dépend que de p et de n.*

DÉFINITION 39
Soient $n \geq 0$, $p \in \mathbb{Z}$.
On note $\binom{n}{p}$ (ou C_n^p) le nombre de p-combinaisons d'un ensemble à n éléments.
Ce nombre est appelé **coefficient binomial** (ou p **parmi** n).
On convient que si $p<0$ et si $p>n$, alors $\binom{n}{p} = 0$.

PROPOSITION 40 (**Propriété du triangle de Pascal**)
Soient $n, p \in \mathbb{N}$. On a

$$\binom{n+1}{p} = \binom{n}{p-1} + \binom{n}{p}.$$

Preuve — Soit Ω de cardinal $n+1$ et $a \in \Omega$. Alors l'ensemble des p-combinaisons de Ω est égal à l'union disjointe des p combinaisons qui contiennent a et de celles qui ne le contiennent pas. Le premier ensemble a pour cardinal $\binom{n}{p-1}$ et le second de cardinal $\binom{n}{p}$, d'où la formule. □

PROPOSITION 41
Soient $n, p \in \mathbb{N}$. Alors on a

$$\binom{n}{p} = \binom{n}{n-p}.$$

Preuve — Si $p>n$, l'égalité se résume à $0=0$. Sinon, soit Ω de cardinal n et $\varphi : \mathcal{P}(\Omega) \to \mathcal{P}(\Omega)$, $A \mapsto \bar{A}$. Alors φ est involutive, donc bijective et induit donc une bijection des p-combinaisons avec les $n-p$-combinaisons, l'égalité est démontrée. □

PROPOSITION 42
Soient $n, p \in \mathbb{N}$ avec $0 \leq p \leq n$. Alors on a

$$\binom{n}{p} = \frac{n!}{p!(n-p)!}$$

REMARQUE 43 — *On a ainsi* $\binom{n}{0} = \binom{n}{n} = 1$, $\binom{n}{1} = \binom{n}{n-1} = 1$, $\binom{n}{2} = \frac{n(n-1)}{2}$.

EXEMPLE 44 — *Soit* $n \in \mathbb{N}^*$ *fixé, on cherche* $p \in \mathbb{N}$ *tel que* $\binom{n}{p}$ *soit maximal. pour cela, on calcule pour* $p \neq 0$

$$\frac{\binom{n}{p}}{\binom{n}{p-1}} = \frac{n!}{(p)!(n-p)!}\frac{(p-1)!(n-p+1)!}{n!} = \frac{n+1-p}{p} = \frac{n+1}{p} - 1$$

Le quotient est décroissant en p, *vaut* n *en* 1, 0 *en* $n+1$ *et vaut* 1 *ssi* $n+1 = 2p$.
On en déduit que $\binom{n}{p}$ *est maximal pour* $p = \left[\frac{n}{2}\right]$ *(si le rapport vaut* 1 *le maximum est atteint pour deux valeurs de* p*).*

DÉFINITION 45 (Arrangements)
Soient $n, p \in \mathbb{N}^*$.
On appelle **arrangement** à p éléments de $\{1, \dots, n\}$ tout p-uplet $(a_1, \cdots, a_p)$ de $\{1, \dots, n\}^p$ tel que les a_i soient deux à deux distincts.

Au lieu de prendre seulement p éléments parmi n, un arrangement tient aussi compte de l'ordre des éléments choisis. (par ex $(2,3)$ et $(3,2)$ sont deux arrangements différents à 2 él. de $\{1,2,3\}$)

PROPOSITION 46
Soient $n, p \geq 1$.
Si $1 \leq p \leq n$, le nombre d'arrangements à p éléments de $\{1, \cdots, n\}$ vaut $A_p^n = \dfrac{n!}{(n-p)!} = p!\binom{n}{p}$.
Si $p > n$, le nombre d'arrangements à p éléments de $\{1, \cdots, n\}$ vaut $A_p^n = 0 = p!\binom{n}{p}$.

Preuve — Le nombre d'arrangements de p éléments dans I_n revient à la donnée d'un élément a_1 dans I_n, puis d'un élément $a_2 \in I_n \setminus \{a_1\}$, etc puis $a_p \in I_n \setminus \{a_1, \cdots, a_{p-1}\}$... et donc $A_p^n = n(n-1)\cdots(n-p+1)$, ce que nous voulions. Mais cette écriture, bien que convaincante, n'est pas entièrement rigoureuse. Reprenons le raisonnement. On note $\mathcal{A}$ l'ensemble des p-arrangements de $[\![1, n]\!]$.

1. Pour tout $i \in [\![0, n]\!]$, on pose A_i l'ensemble des arrangements qui commencent par i. Les A_i forment une partition de $\mathcal{A}$, donc $|\mathcal{A}| = \sum_{i=1}^{n} |A_i|$. Si σ_i est la permutation $(1\ i)$, $\varphi_i : A_1 \to A_i$, $(1, a_2, \cdots, a_p) \mapsto (\sigma(1), \sigma(a_2), \cdots, \sigma(a_p))$ est une bijection (donnez la réciproque !). Donc $|A_i| = |A_1|$ et $|\mathcal{A}| = n|A_1|$ car il y a n ensembles A_i.
2. Puis on pose, pour $i \neq j$, $A_{(i,j)}$ l'ensemble des permutations qui commencent par (i, j) : $(i, j, a_3, \cdots, a_n)$. On montre que tous les $A_{(i,j)}$ ont même cardinal et les $A_{(i,j)}$ pour $j \in [\![1, n]\!] \setminus \{i\}$ forment une partition de A_i. On en déduit que $|A_i| = (n-1)|A_{(1,2)}|$
3. Avec les notations précédentes, $|A_{(a_i, \cdots, a_r)}| = (n-r)|A_{(a_i, \cdots, a_{r+1})}|$ tant que $r \leq p-1$.

⋮

4. Si $p = r$, il est clair que $|A_{(a_1, \cdots, a_p)}| = 1$ et on en déduit le résultat.

En probabilité, la formule des probabilités composées nous permettra de clarifier ce type de raisonnements. Mais il est préférable ici de faire, par exemple, une simple récurrence sur p pour n fixé en reprenant l'étape 1 ci-dessus.
Une autre méthode est de considérer l'application φ surjective de $\mathcal{A}_n^p$, l'ensemble des arrangements de p éléments de I_n, dans $\mathcal{C}_n^p$, l'ensemble des p-combinaison de I_n, définie par

$$\varphi : \mathcal{A}_n^p \to \mathcal{C}_n^p,\ (a_1, \cdots, a_p) \mapsto \{a_1, \cdots, a_p\}.$$

Pour toute p combinaison $\{a_1, \cdots, a_p\}$,

$$\varphi^{-1}(\{a_1, \cdots, a_p\}) = \left\{(a_{\sigma(1)}, \cdots, a_{\sigma(p)}),\ \sigma \in S_p\right\}$$

On en déduit que $\left|\varphi^{-1}(\{a_1, \cdots, a_p\})\right| = p!$. Or

$$|\mathcal{A}_n^p| = \sum_{\{a_1, \cdots, a_p\} \in \mathcal{C}_n^p} \left|\varphi^{-1}(\{a_1, \cdots, a_p\})\right| = p!\,|\mathcal{C}_n^p| = \frac{n!}{(n-p)!}.$$

□

16.5 EXEMPLES DE DÉNOMBREMENT

EXEMPLE 47 — *Soit* M *un mot de* n *lettres, composé des lettres* $a_1, \dots, a_l$, *qui apparaissant respectivement* $\alpha_1, \dots, \alpha_l$ *fois. On a donc* $n = \alpha_1 + \cdots + \alpha_l$.
Alors, le nombre d'anagrammes (mots obtenus par permutation des lettres) de M, *noté* $N(n, \alpha_1, \cdots, \alpha_l)$, *est égal à*

$$N(n, \alpha_1, \cdots, \alpha_l) = \frac{n!}{\alpha_1! \cdots \alpha_l!}.$$

On démontre cela par récurrence sur $l \geq 1$.
Initialisation : *Pour $l = 1$, il n'y a qu'un seul anagramme du mot.*
Hérédité : *Supposons la proposition vraie pour $l - 1 \geq 1$.*
La lettre a_l doit apparaître α_l fois dans l'anagramme. Soit $\mathcal{C}_n^{\alpha_l}$, l'ensemble des α_l-combinaisons dans I_n. Soit $s \in \mathcal{C}_n^{\alpha_l}$. On note Ω_s l'ensemble des anagrammes tels que la position de la lettre a_l corresponde au choix s. Les ensembles Ω_s forment une partition de Ω.
Or, chaque Ω_s correspond aux anagrammes du mot d'origine privé de la lettre a_l et donc est de cardinal $N(n - \alpha_l, \alpha_1, \cdots, \alpha_{l-1})$. Comme $|\mathcal{C}_n^p| = \binom{n}{\alpha_l}$, on obtient

$$N(n, \alpha_1, \cdots, \alpha_l) = \binom{n}{\alpha_l} N(n - \alpha_l, \alpha_1, \cdots, \alpha_{l-1}) = \frac{n!}{\alpha_l!(n-\alpha_l)!} \times \frac{(n-\alpha_l)!}{\alpha_1! \cdots \alpha_{l-1}!} = \frac{n!}{\alpha_1! \cdots \alpha_l!}$$

ce qui termine la preuve.
• *Le nombre $N(n, \alpha_1, \cdots, \alpha_l)$ est appelé* ***coefficient multinomial****.*
Quand $n = 2$, on a $N(2, \alpha_1, \alpha_2) = \binom{\alpha_1+\alpha_2}{\alpha_1}$, on retrouve les coefficients binomiaux.

Les coefficients binomiaux apparaissent dans le développement de $(a + b)^n$ (pour a, b des éléments d'un anneau commutatif).
On peut généraliser ce résultat avec les coefficients multinomiaux.

Proposition 48
Soient $p \geq 2$ et $a_1, \ldots, a_p$ des éléments d'un anneau commutatif A. Alors, on a

$$(a_1 + \cdots + a_p)^n = \sum_{\alpha_1 + \cdots \alpha_p = n} \frac{n!}{\alpha_1! \cdots \alpha_p!} a_1^{\alpha_1} \cdots a_p^{\alpha_p}, \forall n \geq 0.$$

Preuve — On calcule le coefficient de $a_1^{\alpha_1} \cdots a_p^{\alpha_p}$ sachant que la puissance totale est n car on a n facteurs. Cela correspond à écrire un mot de n lettres avec l'alphabet $\{a_1, \cdots, a_p\}$. La proposition ci-dessus permet de conclure.

Par exemple

$$(a_1 + \cdots + a_p)^2 = a_1^2 + \cdots + a_p^2 + 2a_1a_2 + 2a_1a_3 + \cdots + 2a_{n-1}a_n.$$
$$(a_1 + a_2 + a_3)^3 = a_1^3 + a_2^3 + a_3^3 + 3a_1^2a_2 + 3a_1^2a_3 + 3a_2^2a_1 + 3a_2^2a_3 + 3a_3^2a_1 + 3a_3^2a_2 + 6a_1a_2a_3.$$

□

Exemple 49 — *Soit E un ensemble fini à n éléments, et soit $A \subset E$ de cardinal p $(0 \leq p \leq n)$.*
Dénombrer l'ensemble

$$\Omega = \{(X, Y) \in \mathcal{P}(E)^2 \mid A \subset X \cap Y \text{ et } X \cup Y = E\}$$

Méthode 1 : *On simplifie d'abord le problème. Par hypothèse, A est inclus dans X et Y, en posant $X' = X \setminus A$ et $Y' = Y \setminus A$, on cherche tous les couples $(X', Y') \in \mathcal{P}(E \setminus A)^2$ tels que $X' \cup Y' = E' = E \setminus A$.*

$$\Omega' = \{(X', Y') \in \mathcal{P}(E')^2 \mid X' \cup Y' = E\}$$

On a une application

$$\varphi : \Omega \to \Omega', \ (X, Y) \mapsto (X \setminus A, Y \setminus A).$$

On vérifie facilement que φ est surjective et que $\varphi^{-1}(X', Y') = (X' \cup A, Y' \cup A)$. Donc $|\Omega| = |\Omega'|$. Pour calculer $|\Omega'|$, on pose Ω_k l'ensemble des couples $(X', Y') \in \Omega'$ tels que X' a k éléments dans E'. Les Ω_k forment une partition de Ω'.
Calculons $|\Omega_k|$: pour X' fixé, soit $\Omega_{k,X'}$ l'ensemble des couples $(X', Y') \in \Omega'$, c'est-à-dire ssi Y' contient le complémentaire de X' dans E' : $Y' = \overline{X'} \cup Z$, $Z \in \mathcal{P}(X')$ qui est de cardinal 2^k.
Or les $\Omega_{k,X'}$ avec X' de cardinal k forment une partition de Ω_k et il y en a $\binom{n-p}{k}$.
On en déduit que

$$|\Omega_k| = \sum_{X' \in \mathcal{P}(E'),\ |X'| = k} |\Omega_{k,X'}| = \binom{n-p}{k} 2^k.$$

Enfin k varie entre 0 et $n - p$ donc

$$|\Omega| = |\Omega'| = \sum_{k=0}^{n-p} |\Omega_k| = \sum_{k=0}^{n-p} \binom{n-p}{k} 2^k = 3^{n-p}.$$

Méthode 2 : *La simplicité du résultat nous suggère que l'on a peut-être une méthode plus efficace qui traduit la remarque suivante : pour tout $a \in E \setminus A$, on a trois possibiltés*

1. *a appartient à X et pas à Y*
2. *a appartient à Y et pas à X*
3. *a appartient à X et Y*

On peut le rédiger de la manière suivante : posons $\psi : \Omega \to B$, $(X, Y) \mapsto (\mathbb{1}_X, \mathbb{1}_Y)$ avec

$$B = \left\{ f : E \to \{0,1\}^2 \mid \forall a \in A,\ f(a) = (1,1) \text{ et } \forall b \in E \setminus A, f(b) \in \{(0,1),(1,0),(1,1)\} \right.$$

On montre que ψ est bijective. Enfin, B est en bijection avec $\{(1,0),(0,1),(1,1)\}^{E\setminus A}$. Donc $|\Omega| = |B| = 3^{n-p}$ car $|E \setminus A| = n - p$.

EXEMPLE 50 — *Ce dernier exemple est plus complexe, mais on peut le résoudre simplement si on "modélise" correctement l'ensemble Ω.*
Soient $1 \leq p \leq n$ et soit Ω l'ensemble des répartitions de n boules dans p trous, numérotés T_1, ..., T_p.
Faites un dessin pour bien comprendre.
On considère que

1. *si T_1 a k_1 boules blanches, on écrit $T_1 = \underbrace{1, \cdots, 1}_{k_1 \text{ fois}}$.*

$\vdots$

p. si T_p a k_p boules blanches, on écrit $T_p = \underbrace{1, \cdots, 1}_{k_p \text{ fois}}$.

Puis on définit l'application

$$\begin{array}{rccc} \varphi : & \Omega & \to & [\![0,1]\!]^{n+p-1} \\ & (T_1, \cdots, T_p) & \mapsto & (T_1, 0, T_2, 0, \cdots, 0, T_p) \end{array}$$

L'application φ est clairement injective.
Déterminons l'image de φ : nous avons introduit $p-1$ zéros.
De plus, soit $(a_1, \cdots, a_{n+p-1}) \in [\![0,1]\!]^{n+p-1}$ tel qu'il existe $1 \leq j_1 < j_2 < \cdots < j_{p-1} \leq n$ tels que $a_{j_1} = \cdots = a_{j_{p-1}} = 0$ et pour tout $i \in [\![0, n-1]\!]$, i différent j_1, ..., j_{p-1}, alors $a_i = 1$:

$$(a_1, \cdots, a_{n+p-1}) = (1, \cdots, 1, \underset{j_1}{0}, 1, \cdots, 1, \underset{j_2}{0}, 1 \cdots, 1, \underset{j_{p-1}}{0}, 1 \cdots, 1).$$

On pose

1. *$T_1 = \underbrace{1, \cdots, 1}_{j_1 - 1}$.*

$\vdots$

2. *$T_k = \underbrace{1, \cdots, 1}_{j_k - j_{k-1}}$ si $k \in [\![2, p-1]\!]$.*

$\vdots$

p. si $T_p = \underbrace{1, \cdots, 1}_{n - j_{p-1}}$.

Par construction, $\varphi(T_1, \cdots, T_P) = (a_1, \cdots, a_{n+p-1})$.
On en déduit que $\varphi(\Omega)$ est en bijection avec l'ensemble des $p-1$ combinaisons parmi n.
On obtient donc, $|\Omega| = \binom{n+p-1}{n} = \binom{n+p-1}{p-1}$.
***Application :** Si x_1, ..., x_p sont des nombres complexes, combien existe-t-il de monômes $x_1^{\alpha_1} \cdots x_p^{\alpha_p}$ tels que $\alpha_1 + \cdots + \alpha_p = n$?*

16.6 TIRAGES

En probabilités, lorsque l'on tire des éléments (des cartes, des boules dans une urne,...) on rencontre 4 méthodes principales pour faire le tirage.
Le tirage avec ou sans remise, et avec ou sans ordre.
Dans un tirage avec remise, on peut tirer plusieurs fois le même élément.
Dans un tirage sans remise, tous les éléments tirés doivent être différents.
Dans un tirage avec ordre, l'ordre des éléments tirés compte (on regardera une liste).
Dans un tirage sans ordre, l'ordre des éléments tirés ne compte pas (on regardera un ensemble). Prenons $E = \{1, \ldots, n\}$ pour ensemble de départ, et tirons k éléments ($0 \leq k$) parmi les n éléments de E.

- **Tirage sans remise, sans ordre**
  L'élément tiré est $\{a_1, \dots, a_k\}$, avec $a_i \in E$ t.q. $a_i \neq a_j \, \forall i \neq j$.
  Il y a $\binom{n}{k}$ tirages possibles.
- **Tirage sans remise, avec ordre**
  L'élément tiré est $(a_1, \dots, a_k)$, avec $a_i \in E$ t.q. $a_i \neq a_j \, \forall i \neq j$.
  Il y a $A_k^n = k!\binom{n}{k}$ tirages possibles.
- **Tirage avec remise, avec ordre**
  L'élément tiré est $(a_1, \dots, a_k, \ a_i)$, avec $a_i \in E$.
  Il y a n^k tirages possibles.
- **Tirage avec remise, sans ordre**
  L'élément tiré est $\{\{a_1, \dots, a_k\}$, avec $a_i \in E$ (un ensemble où on autorise les répétitions).
  Il y a $\binom{n+k-1}{n-1}$ tirages possibles.

On retrouve alors les coefficients binomiaux, les nombres d'arrangements, le nombre de fontions d'un ensemble à k éléments vers un ensemble à n éléments, et le nombre de dispositions de k éléments dans n boîtes.

16.7 Familles sommables

Les sommes $\sum_{k=0}^{n} u_k$ se retrouvent en probabilités dans le cas fini (calcul de probabilité, espérance, variance,...).
Dans le cas infini, on retrouvera des séries $\sum_{k\geq 0} u_k$ et leur somme associée.
Chez les séries, permuter les termes peut changer le fait d'être sommable ou non, ainsi que la valeur de la somme.
En général, une série est sensible à l'ordre de la sommation.
En probabilités, nous voulons éviter ces phénomènes afin de pouvoir sommer dans n'importe quel ordre.

Parmi les séries convergentes, celles qui posent problème sont les séries semi-convergentes.
Une série $\sum u_n$ **semi-convergente** est une série qui est convergente mais qui n'est pas absolument convergente ($\sum |u_n|$ diverge).

Exemple 51 — *La série harmonique alternée* $H_n^- = \sum \dfrac{(-1)^n}{n}$ *est une série semi-convergente.*

Proposition 52
Soit $\sum u_n$ une série semi-convergente.
Pour tout $l \in \mathbb{R}$, il existe une bijection $f : \mathbb{N} \to \mathbb{N}$ telle que $\sum u_{f(n)}$ converge vers l.

A contrario, définissons la notion de famille sommable, qui est la notion dont on a besoin en probabilités.

Définition 53
Soit I un ensemble non vide et $(\alpha_i)_{i\in I}$ une famille de nombres réels positifs. On pose

$$\sum_{i\in I} \alpha_i = \sup_{J\subset I, \text{ t.q. } Card(J)<+\infty} \sum_{i\in J} \alpha_i$$

avec la convention que la borne supérieure vaut $+\infty$ si l'ensemble n'est pas majoré.
Si la borne supérieure est finie, on dit que la famille $(\alpha_i)_i$ est **sommable**.

Une famille sommable de nombres positifs est une famille de nombres pour laquelle toutes les sommes finies sont majorées par une constante commune.

Définition 54
Soient I un ensemble non-vide et $(\alpha_i)_{i\in I}$ une famille de réels.
On dit que $(\alpha_i)_{i\in I}$ est **sommable** si les familles $(\alpha_i^+)_{i\in I}$ et $(\alpha_i^-)_{i\in I}$ sont sommables en tant que familles de réels positifs.
On définit alors la somme

$$\sum_{i\in I} \alpha_i = \sum_{i\in I} \alpha_i^+ - \sum_{i\in I} \alpha_i^- .$$

On dit qu'une famille de nombres complexes $(\beta_i)_{i\in I}$ est **sommable** si la famille des parties réelles et la famille des parties imaginaires sont sommables. On définit alors la somme

$$\sum_{i\in I}\beta_i = \sum_{i\in I} Re(\beta_i) + i\sum_{i\in I} Im(\beta_i).$$

REMARQUE 55 — *Pour tout nombre complexe $z\in\mathbb{C}$, on a*

$$\max(|\mathrm{Re}\, z|, |\mathrm{Im}\, z|) \leq |z| \leq |\mathrm{Re}\, z| + |\mathrm{Im}\, z|$$

On en déduit qu'une famille de nombres complexes $(z_i)_{i\in I}$ est sommable si et seulement si la famille $(|z_i|)_{i\in I}$ est sommable.

PROPOSITION 56
Soit $(\alpha_i)_{i\in I}$ une famille de réels. Soit J l'ensemble des indices j tels que $\alpha_j \neq 0$.
• Si $(\alpha_i)_{i\in I}$ est sommable, alors l'ensemble J est au plus dénombrable.
• Si J est infini dénombrable, pour $(j_n)_{n\in\mathbb{N}}$ une énumération de J, on a alors que la famille $(\alpha_i)_{i\in I}$ est sommable ssi la série $\sum_{k=0}^{n}\alpha_{j_k}$ est absolument convergente.
Dans ce cas, on obtient :

$$\sum_{i\in I}\alpha_i = \sum_{j\in J}\alpha_j = \sum_{k=0}^{\infty}\alpha_{j_k}.$$

Ainsi, les familles sommables correspondent exactement aux suites de nombres qui fournissent des séries absolument convergentes.
La différence se situe au niveau des définitions : Pour une famille sommable les éléments ne sont en général pas ordonnés (par ex la famille $(\frac{1}{q^p})_{r=\frac{p}{q}\in\mathbb{Q}}$ qui est indexée sur $\mathbb{Q}$, ou la famille $(\frac{1}{2^i}\frac{1}{3^j})_{(i,j)\in\mathbb{N}^2}$ qui est indexée sur $\mathbb{N}^2$), alors que dans une série les éléments sont ordonnés.
Pour vérifier qu'une famille $(a_i)_{i\in I}$ est sommable, on ordonne ses éléments et on montre que $\sum|a_i| < +\infty$ (par comparaison avec une série de référence, par équivalent avec le terme d'une série CV, par comparaison série-intégrale, etc).

16.7.1 Théorème de sommation par paquets, théorème de Fubini

THÉORÈME 57 (**Théorème de sommation par paquets**)
Soit $(\alpha_i)_{i\in I}$ une famille de nombres complexes et soit $(A_j)_{j\in J}$ une partition de I.
La famille $(\alpha_i)_{i\in I}$ est sommable si et seulement si chacune des familles $(\alpha_i)_{i\in A_j}$ est sommable, et si la famille $\left(\sum_{i\in A_j}|\alpha_i|\right)_{j\in J}$ est elle aussi sommable.
Autrement dit, la famille $(\alpha_i)_{i\in I}$ est sommable si et seulement si $\sum_{j\in J}\sum_{i\in A_j}|\alpha_i| < +\infty$. Dans ce cas, on a alors

$$\sum_{i\in I}\alpha_i = \sum_{j\in J}\left(\sum_{i\in A_j}\alpha_i\right).$$

REMARQUE 58 — • *Pour que la famille $(\alpha_i)_{i\in I}$ soit sommable, il ne suffit pas que chacune des familles $(\alpha_i)_{i\in A_j}$ soit sommable et que la famille $\left(\sum_{i\in A_j}\alpha_i\right)_{j\in J}$ soit sommable, à cause de problèmes de signes.*

Contre-exemple : *si $A_j = \{2j, 2j+1\}$ pour tout $j\in\mathbb{N}$ et $\alpha_j = (-1)^j$, la famille $(1,-1,1,-1,...)$ indexée par $\mathbb{N}$ n'est pas sommable.*
Il faut bien vérifier que la quantité $\sum_j\sum_{i\in A_j}|\alpha_i|$ est finie.

• *La condition à bien retenir pour les sommes doubles, qui est aussi valable pour des sommes d'intégrales ou pour des intégrales doubles, est que "la somme de la somme du module est CV".*

Dans le cas de sommes doubles, on peut avoir l'existence des sommes $s_j = \sum_{i=0}^{+\infty}u_{i,j}$ et de la somme $\sum_{j=0}^{+\infty}s_j = \sum_{j=0}^{+\infty}\left(\sum_{i=0}^{+\infty}u_{i,j}\right)$ sans pour autant pouvoir permuter les indices i et j.

Tout ce qui touche à la permutation dans des sommes infinies (ou des intégrales) a besoin en général de convergence absolue pour fonctionner.

THÉORÈME 59 (**Théorème de Fubini**)
Soit $(u_{i,j})_{(i,j)\in\mathbb{N}^2}$ une famille de nombres complexes.

1. La famille $(u_{i,j})_{i,j}$ est sommable si et seulement si $\sum_{i=0}^{+\infty}\left(\sum_{j=0}^{+\infty} u_{i,j}\right) < +\infty$.
2. On a alors
$$\sum_{(i,j)\in\mathbb{N}} u_{i,j} = \sum_{i=0}^{+\infty}\left(\sum_{j=0}^{+\infty} u_{i,j}\right) = \sum_{j=0}^{+\infty}\left(\sum_{i=0}^{+\infty} u_{i,j}\right).$$

Preuve — Le théorème de Fubini est un cas particulier du théorème de sommation par paquets, appliqué à $A_j = \{(i,j),\ i\in\mathbb{N}\}$ pour $j\in\mathbb{N}$.

□

REMARQUE 60 — *Pour montrer qu'une famille $(u_{i,j})_{(i,j)\in\mathbb{N}}$ est sommable, on va ainsi montrer que la série $\sum_i |u_{i,j}|$ est convergente pour tout j, puis que la série $\sum_j\left(\sum_i |u_{i,j}|\right)$ est convergente.*
On pourra alors appliquer le théorème de Fubini pour intervertir les deux sommes de séries.
Le théorème de Fubini est le théorème principal qui permet d'intervertir les sommes de séries (ou série et intégrale en analyse).

COROLLAIRE 61 (Produit de séries absolument convergentes)
Soient $\sum_{i\geq 0} a_i$ et $\sum_{i\geq 0} b_i$ deux séries absolument convergentes.
Alors la famille $(u_{i,j} = a_i b_j)_{(i,j)\in\mathbb{N}^2}$ est sommable, et on a
$$\sum_{(i,j)\in\mathbb{N}} a_i b_j = \sum_{i\in\mathbb{N}} a_i \sum_{i\in\mathbb{N}} b_i.$$

REMARQUE 62 — *Si l'on pose $c_n = \sum_{k=0}^{n} a_k b_{n-k}$ et que l'on utilise le théorème de sommation par paquets, on retrouve le fait que le produit de Cauchy de deux séries absolument convergentes est absolument convergent, et que la somme du produit est le produit des sommes.*
Avec des séries entières il est courant de faire des produit $\sum a_n x^n \sum b_n x^n = \sum c_n x^n$.
Tant que les séries sont absolument convergentes, on pourra parler aussi bien de produit de familles sommables que de produit de Cauchy.

EXEMPLE 63 — *Montrer que la famille $(\alpha_{m,n}) = \left(\frac{(-1)^{mn}}{m^2n^2}\right)_{m,n\in\mathbb{N}^*}$ est sommable et calculer sa somme S.*
On a
$$\sum_{(m,n)\in(N^*)^2} |\alpha_{m,n}| = \sum_{(m,n)\in(N^*)^2} \frac{1}{m^2n^2} = \left(\sum_{n>0}\frac{1}{n^2}\right)^2 < +\infty,$$
donc la famille est bien sommable.
De plus le produit $m.n$ est impair ssi m et n sont impairs.
On sait que $A = \sum_{n>0}\frac{1}{n^2} = \frac{\pi^2}{6}$ et que si P est l'ensemble des entiers pairs non nuls et I les entiers impairs, alors
$$U = \sum_{p\in P}\frac{1}{p^2} = \sum_{n\in\mathbb{N}^*}\frac{1}{(2n)^2} = \frac{1}{4}A$$
et
$$V = \sum_{i\in I}\frac{1}{i^2} = \sum_{n\in\mathbb{N}^*}\frac{1}{n^2} - \sum_{p\in P}\frac{1}{p^2} = \frac{3}{4}A.$$
On en déduit par découpage que
$$S = U^2 + 2UV - V^2 = -\frac{\pi^4}{288}.$$

EXEMPLE 64 — *Démontrer que la famille* $(\alpha_{n,p}) = \left(\frac{1}{n^p}\right)_{n,p\geq 2}$ *est sommable et calculer sa somme* S.
En effet, pour $n \geq 2$ *fixé, la série géométrique* $\sum_{p\geq 2} \frac{1}{n^p}$ *converge et vaut* $\frac{1}{n^2} \times \frac{1}{1-\frac{1}{n}} = \frac{1}{n-1} - \frac{1}{n}$ *qui est le terme général d'une série télescopique convergente vers* 1.
La famille est bien sommable de somme 1 *et on a ainsi montré que*

$$\sum_{p=2}^{+\infty}\left(\sum_{n=2}^{+\infty}\frac{1}{n^p}\right) = 1 = \sum_{p=2}(\zeta(p)-1) = 1$$

où la fonction zeta de Riemann vaut $\zeta(z) = \sum_{k=1}^{+\infty}\frac{1}{n^z}$ *pour* $|z| > 1$.

EXEMPLE 65 —

1. *Calculer la somme* $\sum_{n=0}^{+\infty}\left(\sum_{k\geq n}\frac{1}{k!}\right)$.
2. *On pose* $a_{n,p} = \frac{1}{n^2-p^2}$ *si* $n \neq p$ *et* 0 *sinon.*
   (a) *Expliquer pourquoi la famille n'est pas sommable.*
   (b) $\sum_n\sum_p a_{n,p}$ *et* $\sum_p\sum_n a_{n,p}$.

1. *La somme inversée vaut* $\sum_{k=0}^{+\infty}\sum_{n=0}^{k}\frac{1}{k!} = \sum_{k\geq 0}\frac{k+1}{k!} = \sum_{k=0}^{+\infty}\frac{1}{k!} + \sum_{k=1}^{+\infty}\frac{1}{(k-1)!} = 2e$. *Comme la famille est à termes positifs, on en déduit que la famille est sommable et que la somme demandée vaut encore* $2e$.
2. (a) *La somme des* $a_{n,n-1} \sim \frac{1}{2n}$ *tend vers* $+\infty$
   (b) *Il faut calculer pour tout* $n \in \mathbb{N}$, *la somme* $\sum_{p=0}^{+\infty}\frac{1}{n^2-p^2}$.

   Pour $n = 0$, *la somme* $-\sum_{p=0}^{+\infty}\frac{1}{p^2} = -\frac{\pi^2}{6}$.

   Si $n \neq 0$, *on remarque que* $\frac{1}{n^2-p^2} = \frac{1}{2n}\left(\frac{1}{n+p} + \frac{1}{n-p}\right)$, *donc pour* N *grand*

$$\begin{aligned}\sum_{p=0,p\neq n}^{N}\left(\frac{1}{n+p}+\frac{1}{n-p}\right) &= \sum_{k\geq n,k\neq 2n}^{N+n}\frac{1}{k} + \sum_{k=1}^{n}\frac{1}{k} - \sum_{k=1}^{N-n}\frac{1}{k}\\ &= \frac{1}{n} - \frac{1}{2n} + \sum_{k=1}^{N+n}\frac{1}{k} - \sum_{k=1}^{N-n}\frac{1}{k}\\ &= \frac{1}{2n} + \sum_{k=N-n+1}^{N+n}\frac{1}{k}\end{aligned}$$

   et on obtient

$$\sum_{p=0}^{+\infty}\frac{1}{n+p}+\frac{1}{n-p} = \lim_{N\to+\infty}\sum_{p=0}^{N}\frac{1}{n+p}+\frac{1}{n-p} = \frac{1}{2n}.$$

   En remplaçant dans l'expression

$$\sum_n\sum_p a_{n,p} = \sum_{n\neq 0}\frac{1}{4n^2} - \frac{\pi^2}{6} = -\frac{\pi^2}{8} = -\sum_p\sum_n a_{n,p}.$$

Chapitre 17 Espaces probabilisés

Table des matières du chapitre

17.1 Le langage des probabilités

Définition 1
On appelle expérience aléatoire une expérience $\mathcal{E}$ qui, reproduite dans des conditions identiques, peut conduire à plusieurs résultats possibles, et dont on ne peut prévoir le résultat par avance.

Définition 2
L'ensemble de tous les résultats possibles d'une expérience aléatoire est appelé ensemble d'états ou **univers**.
Il est noté Ω.
Un élément de Ω est noté généralement ω ($\omega \in \Omega$).
On dit que ω est un résultat possible de l'expérience aléatoire.

Exemple 3 —

1. *On lance une pièce :* $\Omega = \{P, F\}$*, assimilé à* $\Omega = \{0, 1\}$.
2. *On lance un dé :* $\Omega = \{1, 2, 3, 4, 5, 6\}$.
3. *Génotype d'un individu :* $\Omega = \{AA, Aa, aa\}$.
4. *On étudie n individus :* $\Omega = \{AA, Aa, aa\}^n$.
5. *On étudie la durée de vie d'une bactérie :* $\Omega = [0, +\infty[$.
6. *On étudie la durée d'une communication téléphonique :* $\Omega = [0, +\infty[$.
7. *On envoie une fléchette sur une cible circulaire de 30 cm de diamètre :* $\Omega = \{(x, y), x^2 + y^2 \leq 15\}$.
8. *Cours d'un actif financier sur un intervalle de temps* $[t_1, t_2]$ *:* $\Omega = \mathcal{C}^0([t_1, t_2], \mathbb{R}_+^*)$.

Remarque 4 — *Cette longue liste d'exemples montre que l'espace* Ω *peut varier énormément dans sa structure, d'une expérience à l'autre. Cela permet de réaliser la richesse de la théorie qu'il faut mettre en place pour créer un modèle qui englobe tous les cas.*

Remarque 5 — *Quelle information pouvons-nous tirer de l'expérience ? Dans le jeu de fléchettes, on s'intéresse à la chance de tomber dans une des couronnes ou un des secteurs de la cible.*
Les résultats du jeu peuvent se décrire à l'aide de parties du disque. Pas la température de la pièce...

Définition 6
Soit Ω un ensemble associé à une expérience aléatoire.
On appelle **événement aléatoire** (associé à l'expérience $\mathcal{E}$) un sous-ensemble de $A \subset \Omega$ dont on peut dire au vu de l'expérience s'il est réalisé ou non.

Exemple 7 —

1. $\Omega = \{0, 1\}$. *"La pièce tombe sur Pile" :* $A = \{0\}$.
2. $\Omega = \{0, 1\}^n$, $\omega = (\omega_1, \cdots, \omega_n)$. *"Le nombre de Faces est supérieur au nombre de Piles" :*

$$A = \{\omega \in \Omega, \sum_{i=1}^{n} \omega_i \geq \frac{n}{2}\}.$$

3. Dans un lancer de deux dés

$$A = \{\textit{La somme des deux dés est inférieure à } 4\}$$

est un évènement aléatoire mais

$$B = \{\textit{Le résultat du premier dé lancé est inférieur à } 4\}$$

n'en est pas si Ω ne contient que les résultats non ordonnés des tirages.

Un événement aléatoire A est un sous-ensemble, donc il est caractérisé par l'ensemble des ω qu'il contient (l'ensemble des résultats tels que l'événement se réalise).

17.1.1 Opérations ensemblistes et description des événements aléatoires

Comme les événements sont des sous-ensembles, on peut effectuer des opérations ensemblistes dessus, et donner des interprétations.
Soient A et B sont deux événements d'une expérience aléatoire. (deux parties d'un ensemble Ω) On a :

1. A n'est pas réalisé : $\overline{A}$ (le complémentaire de A, noté aussi A^C)
2. A **et** B sont réalisés : $A \cap B$
3. A **ou** B sont réalisés : $A \cup B$
4. A réalisé $\Rightarrow$ B réalisé : $A \subset B$.
5. A et B sont incompatibles : $A \cap B = \emptyset$.
6. Toujours vrai : Ω est l'événement certain (il arrive toujours).
7. Jamais vrai : $\emptyset$ est l'événement impossible (il n'arrive jamais).

On note $\mathcal{A}$ l'ensemble de tous les événements possibles dans l'expérience aléatoire.
C'est un ensemble de parties de Ω.
Cet ensemble modélise l'information que l'on peut obtenir à partir des résultats de l'expérience.
Il est important de comprendre que l'on n'a pas toujours $\mathcal{A} = P(\Omega)$, ensemble de toutes les parties de Ω : dans le jeu de fléchettes, on ne considère pas que pour gagner la flêche doit avoir des coordonnées rationnelles dans un repère centré au milieu du disque.

Remarque 8 — *Pour que la modélisation soit cohérente avec l'intuition, l'ensemble $\mathcal{A}$ doit être stable par les opérations ensemblistes : si A, $B \in \mathcal{A}$, alors on doit avoir $A \cap B \in \mathcal{A}$, $A \cup B \in \mathcal{A}$, $\overline{A} \in \mathcal{A}$, mais aussi $\Omega \in \mathcal{A}$, $\emptyset \in \mathcal{A}$.*

Remarque 9 — *Comment savoir si la pièce est truquée dans un jeu de Pile ou Face ?*

Approche intuitive Considérons une expérience aléatoire donnée $\mathcal{E}$ et un événement A pour cette expérience.
Le but : associer à chaque événement A un nombre $\mathbb{P}(A)$ compris entre 0 et 1, qui représente la chance a priori que cet événement soit réalisé.
Ce nombre réel est appelé **probabilité de l'événement** A.
Supposons que l'on répète n fois l'expérience $\mathcal{E}$. On note n_A le nombre de fois où l'événement A s'est réalisé. Alors,

$$f_n(A) = \frac{n_A}{n}$$

donne la fréquence des réalisations de A sur ces n essais.
On remarque alors que

1. $f_n(A) \in [0, 1]$;
2. $f_n(\Omega) = 1$ et $f_n(\emptyset) = 0$;
3. Si $A \cap B = \emptyset$, on a
$$f_n(A \cup B) = f_n(A) + f_n(B).$$
4. Si $A \subset B$ on a $f_n(A) \leq f_n(B)$;
5. Intuitivement, on imagine avoir
$$\mathbb{P}(A) = \lim_{n \to +\infty} f_n(A).$$

Cette conception de la probabilité d'un événement A comme fréquence d'apparition de l'événement est l'approche intuitive que l'on a de la notion.
Pour une expérience où on lance un dé à 6 face (dé équilibré, non truqué), on a l'intuition que la probabilité de l'événement $\{$Obtenir 4$\}$ est égale à $\frac{1}{6}$, car cette quantité représente la fréquence à laquelle les lancers de dé vont donner un 6.

Exemple 10 (Expérience aléatoire avec une infinité de résultats possibles) — *On considère un jeu de Pile ou Face infini, avec pièce équilibrée. L'ensemble des résultats est ainsi* $\Omega = \{0,1\}^{\mathbb{N}}$.
On veut regarder l'événement

$$A = \{\textit{on ne tire jamais Pile}\}.$$

Que pourrait valoir la probabilité de l'événement A, $\mathbb{P}(A)$ *?*
Si l'on regrde les n *premiers tirages, on est sur l'ensemble fini* $\Omega_n = \{0,1\}^n$ *et, on regarde l'événement* $A_n = \{$*on ne tire jamais Pile en* n *lancers*$\}$.
Alors, on a $\mathbb{P}(A_n) = \dfrac{1}{2^n}$.
Il semble alors évident d'écrire que

$$\mathbb{P}(A) = \mathbb{P}\left(\bigcap_n An\right) = \lim_{n\to+\infty} \mathbb{P}(A_n) = 0.$$

Cependant, pour qu'un tel résultat soit vrai, il faut que la probabilité $\mathbb{P}$ *possèdent une propriété de passage à la limite. Cela implique aussi que l'ensemble des événements* $\mathcal{A}$ *posède une condition de stabilité pour des intersections dénombrables (on a* $A = \cap_{n\geq 1} A_n$*).*

Avec cette section, nous avons donné quelques exemples d'expériences aléatoires, nous avons vu un peu de vocabulaire du monde des probabilités (expérience, événement, réalisation, probabilité), et nous avons vu que deux objets semblaient importants (l'ensemble des événements, la probabilité de chaque événement).
Il est temps de définir mathématiquement ces objets.

17.2 Sigma-algèbre, mesure de probabilités

Pour avoir une structure mathématique qui permet de modéliser facilement et fidèlement les expériences aléatoires que l'on veut étudier dans le monde réel, nous allons avoir besoin de trois objets fondamentaux : les **sigma-algèbres** (ou tribus), les **mesures de probabilités**, et les **variables aléatoires**.
Dans ce chapitre, nous allons définir les sigma-algèbres et les mesures de probabilités, sur un ensemble Ω donné. Les variables aléatoires seront définies au prochain chapitre.
Ces définitions sont courtes, mais extrêmement importantes. De ces définitions découlent toutes les propriétés que nous utiliserons dans nos raisonnemens et nos calculs.
Nous verrons des exemples de mesures de probabilités sur des ensembles finis ou dénombrables, et avec les propriétés des mesures de probabilités nous utiliserons les techniques de dénombrement pour calculer des probabilités.

Commençons par l'objet de base, la sigma-algèbre (ou tribu).

Définition 11
Soit Ω un ensemble. Soit $\mathcal{A} \subset \mathcal{P}(\Omega)$ un ensemble de parties de Ω.
On dit que $\mathcal{A}$ est une **σ-algèbre** (ou une **tribu**) si elle vérifie les propriétés suivantes :

1. On a $\emptyset \in \mathcal{A}$ et $\Omega \in \mathcal{A}$.
2. $\mathcal{A}$ est stable par passage au complémentaire :
   Pour tout $A \in \mathcal{A}$ on a $\overline{A} \in \mathcal{A}$.
3. $\mathcal{A}$ est stable par réunion et intersection dénombrables :
   Pour $(A_n)_{n\in\mathbb{N}}$ une famille d'éléments de $\mathcal{A}$, on a
   $\bigcup_{n\in\mathbb{N}} A_n \in \mathcal{A}$ et $\bigcap_{n\in\mathbb{N}} A_n \in \mathcal{A}$.

Remarque 12 —

1. *Une* σ*-algèbre* $\mathcal{A}$ *toujours non vide.*
2. *Pour* A, $B \in \mathcal{A}$, *alors* $A \cup B \in \mathcal{A}$ *et* $A \cap B \in \mathcal{A}$.
   Une σ*-algèbre est stable par intersection et par réunion (finies).*

3. *L'ensemble $\mathcal{A}$ est un ensemble de parties de Ω, et pas un sous-ensemble de Ω.*
   Cet ensemble est un sous-anneau de $(\mathcal{P}(\Omega), \cup, \cap)$, et une $\mathbb{Z}/2\mathbb{Z}$-algèbre, avec une propriété de "dénombrabilité" en plus.
   Le nom de "sigma-algèbre" fait référence à cette structure algébrique, et à cette propriété en plus.
4. *Notons également que la propriété 3. n'implique pas que $\mathcal{A}$ soit stable par réunion ou intersection infinie non dénombrable.*
5. *La σ-algèbre représente l'ensemble des parties de Ω que l'on va chercher à mesurer, dont on va chercher à donner une probabilité (pour tout $A \in \mathcal{A}$, on donnera un sens à $\mathbb{P}(A)$, la probabilité de la partie A). C'est le premier élément fondamental pour modéliser les expériences aléatoires.*

Exemple 13 —

1. *$\mathcal{A} = \{\emptyset, \Omega\}$ est une σ-algèbre.*
   On l'appelle la tribu grossière, ou triviale. C'est la plus petite tribu de Ω.
2. *L'ensemble $\mathcal{P}(\Omega)$ de toutes les parties de Ω est une σ-algèbre sur Ω.*

Proposition-Définition 14
Soit Ω un ensemble. Soit $C \subset P(\Omega)$.
On appelle tribu engendrée par C la plus petite tribu contenant C.
Cette tribu existe toujours, car d'une part $\mathcal{P}(\Omega)$ est une tribu contenant C, et d'autre part l'intersection d'une famille quelconque de tribus est une tribu.
Ainsi, la tribu engendrée par C est l'intersection de toutes les tribus contenant C.

Exemple 15 —

1. *La tribu engendrée par l'ensemble $\{A\}$ est $\{\emptyset, A, \overline{A}, \Omega\}$.*
2. *Si $(A_i)_{i\in I}$ est une partition finie ou dénombrable de Ω, la tribu engendrée par $\{A_i, i \in I\}$ est l'ensemble des réunions $B_J = \bigcup_{i\in J} A_i$, pour tout $J \subset I$.*

Proposition 16
Soit Ω un ensemble fini ou dénombrable.
Soit $\mathcal{A}$ une σ-algèbre sur Ω qui contienne tous les singletons : $\forall \omega \in \Omega$, on a $\{\omega\} \in \mathcal{A}$.
Alors, on a $\mathcal{A} = \mathcal{P}(\Omega)$.

Preuve — Soit B une partie de Ω. On a $B = \cup_{\omega\in B}\{\omega\}$.
Comme Ω est dénombrable, l'ensemble B est fini ou dénombrable.
Comme tous les singletons $\{\omega\}$ sont dans $\mathcal{A}$, on en déduit que leur réunion finie ou dénombrable est dans $\mathcal{A}$; donc B est dans $\mathcal{A}$.
On a donc bien que $\mathcal{A} = \mathcal{P}(\Omega)$. □

Remarque 17 — *A chaque fois que l'ensemble Ω sera fini ou dénombrable, on choisira comme σ-algèbre $\mathcal{A}$ l'ensemble $\mathcal{P}(\Omega)$.*
En effet, c'est la seule σ-algèbre de Ω qui contienne tous les singletons.

Remarque 18 — *Quand Ω est infini non dénombrable (par exemple $\mathbb{R}$), la σ-algèbre engendrée par les singletons est différente de $\mathcal{P}(\Omega)$.*
De plus, la tribu $\mathcal{P}(\Omega)$ sera trop grande pour que l'on puisse définir la probabilité de tous ses éléments de façon simple.

Définition 19
Soit $\Omega = \mathbb{R}$.
On appelle **tribu borélienne** la tribu de $\mathbb{R}$ engendrée par l'ensemble de tous les intervalles.
On la note $\mathcal{B}(\mathbb{R})$.

Proposition 20
La tribu borélienne de $\mathbb{R}$ est la tribu engendrée par les intervalles de la forme $]-\infty, a]$ pour $a \in \mathbb{Q}$.
De plus, cette tribu contient tous les singletons.

Preuve — • Rappelons que toute tribu est stable par passage au complémentaire, par réunion ou intersection dénombrable. Puisque $]-\infty, a]$ est le complémentaire de l'intervalle ouvert $]a, +\infty[$, il appartient à la tribu borélienne, et donc la tribu C engendrée par ces intervalles est incluse dans la tribu borélienne.
Réciproquement, soit $]x, y[$ un intervalle ouvert de $\mathbb{R}$. Soit $(x_n)_{n\geq 0}$ une suite de rationnels décroissant vers x et $(y_n)_{n\geq 0}$ une suite de rationnels croissant strictement vers y.
On a :

$$]x, y[= \bigcup_{n\geq 0} \left( ]-\infty, y_n] \cap \overline{]-\infty, x_n]} \right).$$

Nous en déduisons que tout intervalle ouvert appartient à C, d'où le résultat.
• Soit $x \in \mathbb{R}$. Alors le complémentaire de $\{x\}$ est $\mathbb{R} \setminus \{x\}$.
Or, les intervalles $]-\infty, x[$ et $]x, +\infty[$ sont dans $\mathcal{B}(\mathbb{R})$. Donc, par réunion et passage au complémentaire, on en déduit que $\{x\}$ est dans $\mathcal{B}(\mathbb{R})$. □

Remarque 21 —

1. *La tribu borélienne de $\mathbb{R}$ contient tous les singletons $\{x\}$.*
2. *Cette σ-algèbre $\mathcal{B}(R)$ contient strictement la σ-algèbre engendrée par tous les singletons. Elle est aussi strictement incluse dans $\mathcal{P}(\mathbb{R})$.*
   C'est-à-dire qu'il existe des éléments dans $\mathcal{B}(\mathbb{R})$ qu'on ne peut pas engendrer simplement à partir de singletons, et qu'il existe des parties E de $\mathbb{R}$ qui ne sont pas dans la tribu borélienne.
3. *Dans la suite du cours, les ensembles Ω sur lesquels nous prendrons des tribus seront en général soit finis, soit dénombrables, soit égaux à $\mathbb{R}$.*
   • *Si Ω est fini ou dénombrable, on aura comme tribu $\mathcal{A} = \mathcal{P}(\Omega)$.*
   • *Si $\Omega = \mathbb{R}$, on aura comme tribu $\mathcal{B}(\mathbb{R})$.*
   Ces cas paraissent simples, mais ils permettent déjà de construire énormément de variables aléatoires, et d'obtenir beaucoup de phénomènes différents. La majorité des exemples d'expériences aléatoires que vous connaissez peuvent se modéliser avec ces ensembles et ces tribus.

Maintenant que nous avons les sigma-algèbres, nous pouvons donner la définition d'une mesure de probabilité, et celle d'une probabilité.

Définition 22
Soient Ω un ensemble, et $\mathcal{A}$ une tribu sur Ω. Soit $\mathbb{P} : \mathcal{A} \to [0,1]$ une fonction.
On dit que $\mathbb{P}$ est une **mesure de probabilité** sur le couple $(\Omega, \mathcal{A})$ si elle vérifie les propriétés suivantes :

1. $$\mathbb{P}(\Omega) = 1. \quad \text{(mesure totale de l'ensemble)}(*)$$
2. Pour toute famille dénombrable $(A_n)_{n\in\mathbb{N}}$ d'éléments de $\mathcal{A}$ deux-à-deux disjoints, on a
$$\mathbb{P}\left(\bigcup_{n\in\mathbb{N}} A_n\right) = \sum_{n=0}^{+\infty} \mathbb{P}(A_n). \quad \text{(sigma-additivité)}(**)$$

Le nombre réel $\mathbb{P}(A)$ est appelé **probabilité** de la partie A.

Remarque 23 —

1. *La propriété $(**)$ est appelée σ-additivité.*
   Cette propriété implique que la série de terme général $\mathbb{P}(A_n)$ est convergente.
   En effet, c'est une série à termes positifs et majorée par 1.
   De plus, sa somme est égale à $\mathbb{P}\left(\bigcup_{n\in\mathbb{N}} A_n\right)$.
2. *Tout comme on distingue un polynôme $P(X)$ de son évaluation en x, $P(X)$, et la fonction dérivée f' du nombre dérivé en x, $f'(x)$, on distingue bien la **mesure de probabilité** $\mathbb{P}$ de la **probabilité** d'une partie A, $\mathbb{P}(A)$.*
   Le premier est une fonction, le second un nombre réel entre 0 et 1.
3. *Pour que la propriété 1) ait un sens, il était nécessaire que Ω soit un élément de $\mathcal{A}$.*
   Pour que la propriété 2) ait un sens, il était nécessaire que $\mathcal{A}$ soit stable par réunion dénombrable.
4. *La mesure de probabilité $\mathbb{P}$ n'est définie que sur $\mathcal{A}$.*
   Ainsi, la probabilité $\mathbb{P}(A)$ a un sens uniquement pour les éléments de la sigma-algèbre $\mathcal{A}$.
   *Si une partie $E \subset A$ n'est pas dans la sigma-algèbre $\mathcal{A}$, parler de "probabilité de E" **n'a pas de sens**.*
   Par exemple, pour $\Omega = \mathbb{R}$ et $\mathcal{A}$ la sigma-algèbre des boréliens, il existe des parties E de $\mathbb{R}$ qui ne sont pas dans $\mathcal{A}$, et pour lesquelles on ne pourra pas parler de "probabilité" (pas avec les mesures de probabilités les plus classiques sur $\mathbb{R}$).
5. *La mesure de probabilité est le deuxième objet mathématique essentiel pour modéliser les expériences aléatoires.*

Exemple 24 — *Soit Ω un ensemble, et $\mathcal{A}$ une sigma-algèbre sur Ω.*
Soit $\omega_0 \in \Omega$.
On définit la fonction $\delta_{\omega_0} : A \in \mathcal{A} \mapsto \begin{cases} 1 \text{ si } \omega_0 \in A \\ 0 \text{ sinon} \end{cases}$.
Alors, δ_{ω_0} est une mesure de probabilité. (Le montrer)

Cette mesure est appelée ***mesure de Dirac*** *en* ω_0.
C'est un des exemples les plus simples de mesure de probabilité que l'on peut construire. On le reverra par la suite, car il est en fait très utile.

Proposition-Définition 25
Soit Ω un ensemble fini.
On appelle **mesure de probabilité uniforme** sur Ω la mesure de probabilité telle que

$$\mathbb{P}(\{\omega\}) = \frac{1}{\text{card}\,(\Omega)}, \forall \omega \in \Omega.$$

Pour toute partie $A \in \mathcal{P}(\Omega)$, on a alors

$$\mathbb{P}(A) = \frac{\text{card}\,(A)}{\text{card}(\Omega)}.$$

Preuve — On pose la fonction $\mathbb{P} : \mathcal{P}(\Omega) \to [0,1]$ définie par $\mathbb{P}(A) = \frac{\text{card}\,(A)}{\text{card}(\Omega)}$.
Il faut montrer que cette fonction est une mesure de probabilité, afin de prouver le résultat.
• Tout d'abord, on peut remarquer que $\mathbb{P}$ est bien définie : $\mathbb{P}(A)$ est compris entre 0 et 1.
• Ensuite, on remarque que $\mathbb{P}(\Omega) = 1$. La fonction vérifie donc la propriété $(*)$.
• Maintenant, soit $(A_n)_{n\geq 0}$ une famille d'éléments de la sigma-algèbre $\mathcal{P}(\Omega)$ qui soient deux à deux disjoints.
Comme l'ensemble Ω est fini, cette famille de parties de Ω a seulement un nombre fini de parties qui sont non-vides.
Quitte à réordonner, on peut supposer que $A_0, \ldots, A_m$ sont non-vides, et que $A_n = \emptyset$, pour tout $n \geq m+1$.
On a alors :

$$\mathbb{P}(\cup_{n\geq 0} A_n) = \mathbb{P}(\cup_{n=0}^{m} A_n) = \frac{\text{card}\,(\cup_{n=0}^{m} A_n)}{\text{card}(\Omega)}$$

$$\mathbb{P}(\cup_{n\geq 0} A_n) = \frac{\sum_{n=0}^{m} \text{card}\,(A_n)}{\text{card}(\Omega)} = \sum_{n=0}^{m} \mathbb{P}(A_n).$$

Cela montre que $\mathbb{P}$ vérifie la propriété $(**)$, donc que $\mathbb{P}$ est une mesure de probabilité.
On a bien construit la mesure de probabilité uniforme ! □

Remarque 26 —

1. *Cette probabilité décrit mathématiquement l'expression intuitive de "au hasard" (tirage au hasard d'une carte, lancer au hasard d'un dé, etc).*
   C'est-à-dire, que l'on a plusieurs résultats possibles pour une expérience aléatoire (les 6 *faces d'un dé, les* 52 *cartes d'un jeu,...), et qu'aucun résultat n'est avantagé par rapport aux autres.*
   Autrement dit, tous les résultats de l'expérience ont une probabilité identique d'arriver. Cette probabilité est donc : $\frac{1}{\text{nombre de résultats possibles}}$.
2. *Pour calculer la probabilité d'un événement A avec la mesure de probabilité uniforme, il faut calculer Card(A), c'est-à-dire dénombrer A (compter le nombre d'éléments de A).*
3. *Ainsi, le calcul des probabilités (avec la mesure uniforme) se ramène à du calcul combinatoire (au contenu du premier chapitre).*
   La difficulté est de bien décrire et dénombrer l'ensemble total Ω *et a partie A qui nous intéresse.*

Il existe beaucoup d'autres mesures de probabilités sur un ensemble fini. Mais avant de continuer, intéressons-nous aux propriétés de ces fonctions.

La définition d'une mesure de probabilité est courte, mais la propriété de sigma-additivité $(**)$ engendre beaucoup d'autres propriétés qui sont extrêmement utiles pour le calcul de probabilités.
Ces propriétés font intervenir toutes les propriétés des ensembles classiques (union, intersection, complémentaire, inclusion), et toutes les propriétés d'une sigma-algèbre (union dénombrable, intersection dénombrable).

Proposition 27 (**Propriétés élémentaires**)
Soient Ω un ensemble, et $\mathcal{A}$ une tribu sur Ω.
Soit $\mathbb{P}$ une mesure de probabilité. Soient $A, B \in \mathcal{A}$. Alors, on a les résultats suivants :

1. $\mathbb{P}(\emptyset) = 0$;
2. $\mathbb{P}(\bar{A}) = 1 - \mathbb{P}(A)$;
3. Si $A \subset B$, alors $\mathbb{P}(A) \leq \mathbb{P}(B)$.
   La fonction $\mathbb{P}$ est croissante pour l'inclusion.
4. $\mathbb{P}(A) + \mathbb{P}(B) = \mathbb{P}(A \cap B) + \mathbb{P}(A \cup B)$

Preuve — • Commençons par prouver le point 3).
On découpe B en deux sous-ensembles : $B \cap A = A$ et $B \cap \bar{A}$.

Ces deux sous-ensembles sont inclus dans $\mathcal{A}$, et sont disjoints.
On a donc par sigma-additivité de $\mathbb{P}$ que $\mathbb{P}(B) = \mathbb{P}(A) + \mathbb{P}(B \cap \bar{A}) \geq \mathbb{P}(A)$.
• Prouvons ensuite le point 4).
On sépare l'ensemble $A \cup B$ en trois morceaux : $A_1 = A \setminus (A \cap B)$, $A_2 = A \cap B$, $A_3 = B \setminus (A \cap B)$.
Ces sous-ensembles appartiennent tous à la tribu $\mathcal{A}$ comme réunions/intersections/complémentaires d'éléments de $\mathcal{A}$. (par exemple, $A \setminus (A \cap B) = A \cap \overline{(A \cap B)}$)
De plus, ces sous-ensembles sont deux à deux disjoints.
Ainsi, la propriété de sigma-addiivité de la mesure de probabilité $\mathbb{P}$ donne :

$$\mathbb{P}(A) + \mathbb{P}(B) = \mathbb{P}(A_1 \cup A_2) + \mathbb{P}(A_2 \cup A_3) = \mathbb{P}(A_1) + \mathbb{P}(A_2) + \mathbb{P}(A_2) + \mathbb{P}(A_3)$$
$$\mathbb{P}(A) + \mathbb{P}(B) = \mathbb{P}(A_1 \cup A_2 \cup A_3) + \mathbb{P}(A_2) = \mathbb{P}(A \cap B) + \mathbb{P}(A \cup B).$$

• On obtient maintenant le point 2) en utilisant 4) avec $B = \bar{A}$.
• On obtient maintenant le point 1) en utilisant 2) avec $A = \Omega$. □

Remarque 28 —

1. *On voit à nouveau dans cette Proposition que les propriétés de la sigma-algèbre A sont utilisées.*
   Le point 1) *n'aurait pas de sens si $\mathcal{A}$ ne contenait pas l'ensemble vide $\emptyset$.*
   Le point 2) *n'aurait pas de sens si $\mathcal{A}$ n'était pas stable par passage au complémentaire.*

Le résultat suivant est lui aussi extrêmement utile dans la pratique.

Proposition 29 (**Probabilités et suites croissantes/décroissantes**)
Soient Ω un ensemble, et $\mathcal{A}$ une tribu sur Ω.
Soit $\mathbb{P}$ une mesure de probabilité. Alors, on a les propriétés suivantes :

(i) Pour toute famille $(A_n)_{n\in\mathbb{N}} \in \mathcal{A}^{\mathbb{N}}$ croissante pour l'inclusion, on a

$$\mathbb{P}(\bigcup_{n\in\mathbb{N}} A_n) = \lim_{n\to+\infty} \mathbb{P}(A_n).$$

(ii) Pour toute famille $(A_n)_{n\in\mathbb{N}} \in \mathcal{A}^{\mathbb{N}}$ décroissante pour l'inclusion, on a

$$\mathbb{P}(\bigcap_{n\in\mathbb{N}} A_n) = \lim_{n\to+\infty} \mathbb{P}(A_n).$$

Preuve —
(*i*) Soit $(A_n)_{n\in\mathbb{N}}$ une suite d'éléments de $\mathcal{A}$ croissante pour l'inclusion. Notons $A = \bigcup_{n\geq 0} A_n$.
Posons $B_0 = A_0$, et définissons par récurrence $B_n = A_n \setminus B_{n-1}$, pour $n \geq 1$.
Comme $A_n = \bigcup_{p\leq n} B_p$, $\bigcup_{n\in\mathbb{N}} B_n = A$ et comme les B_n sont deux-à-deux disjoints, nous avons

$$\mathbb{P}(A) = \sum_{n\geq 0} \mathbb{P}(B_n) = \lim_{n\to+\infty} \sum_{p=0}^{n} \mathbb{P}(B_p) = \lim_{n\to+\infty} \mathbb{P}(A_n).$$

(*ii*) Si $(A_n)_n$ est une suite décroissante pour l'inclusion, alors la suite des complémentaires $C_n = \bar{A_n}$ est croissante pour l'inclusion.
En appliquant le point (*i*) et en utilisant la propriété $\mathbb{P}(\bar{A}) = 1 - \mathbb{P}(A)$, on en déduit le résultat. □

Remarque 30 —

1. *Ce résultat entraîne en particulier que si $(A_n)_{n\in\mathbb{N}}$ est une suite croissante ou décroissante d'événements, la suite des probabilités $(\mathbb{P}(A_n))_{n\geq 0}$ admet une limite quand n tend vers l'infini.*
2. *Pour que ce résultat ait un sens, il fallait que la sigma-algèbre $\mathcal{A}$ soit stable par intersection dénombrable. Et voilà, toutes les propriétés d'une sigma-algèbre ont été nécessaires pour obtenir des propriétés sur les mesures de probabilités.*
3. *Par définition d'une mesure de probabilité, on peut calculer la probabilité d'une réunion dénombrable de parties disjointes.*
   Avec la propriété précédente, on peut calculer la probabilité d'une réunion d'ensembles formant une suite croissante (ou intersection d'ensembles formant une suite décroissante).
   On remarque dans la preuve de la proposition que la condition de suite croissante/décroissante était absolument nécessaire pour obtenir le résultat.

Si les parties A_n ne sont pas deux-à-deux disjointes, et ne forment pas de suite croissante/décroissante, on ne peut pas appliquer les propriétés précédentes.
Par contre, nous avons la majoration suivante, très utile dans la pratique.

PROPOSITION 31 (**Probabilité d'une réunion quelconque**)
Soient Ω un ensemble, $\mathcal{A}$ une tribu sur Ω, et $\mathbb{P}$ une mesure de probabilité sur $(\Omega, \mathcal{A})$.
Soit $(A_n)_{n\in I}$ une famille finie ou dénombrable d'élements de $\mathcal{A}$ (une famille d'événements). On a alors

$$\mathbb{P}\left(\bigcup_{n\in I} A_n\right) \leq \sum_{n\in I} \mathbb{P}(A_n).$$

Preuve — • Supposons d'abord l'ensemble I fini. Il s'agit de montrer que pour tout entier k,

$$\mathbb{P}(A_1 \cup \cdots \cup A_k) \leq \mathbb{P}(A_1) + \cdots + P(Ak).$$

Nous montrons cette propriété par récurrence sur k :
Initialisation : Elle est évidente pour $k = 1$.
Hérédité : Supposons la propriété vraie pour $k-1$, avec $k \geq 2$.
Posons $B = A_1 \cup \cdots \cup A_{k-1}$ et $C = B \cup A_k$.
Nous savons que

$$\mathbb{P}(C) + \mathbb{P}(B \cap A_k) = \mathbb{P}(B) + \mathbb{P}(A_k),$$

donc $\mathbb{P}(C) \leq \mathbb{P}(B) + \mathbb{P}(A_k)$, et nous en déduisons immédiatement la proposition au rang k.
• Considérons maintenant le cas où I est dénombrable.
Nous pouvons supposer sans restriction que $I = \mathbb{N}$, d'après les résultats du premier chapitre.
Posons $B_n = \bigcup_{i=0}^{n} A_i$, qui est une suite croissante, qui converge en croissant vers l'ensemble $C = \bigcup_{n\in\mathbb{N}} A_n$.
D'après la première partie de la preuve, nous avons

$$\mathbb{P}(B_n) \leq \sum_{i=0}^{n} \mathbb{P}(A_i).$$

Mais le membre de gauche ci-dessus croît vers $\mathbb{P}(C)$ en vertu de la proposition précédente, tandis que le membre de droite croît vers $\sum_{n\in\mathbb{N}} \mathbb{P}(A_n)$.
En passant à la limite, nous obtenons le résultat. □

Et voilà, nous avons énoncé ici toutes les propriétés les plus fondamentales d'une mesure de probabilité.
Avec les deux objets que sont la sigma-algèbre (ou tribu) et la mesure de probabilité, on regarde en général le triplet suivant :

DÉFINITION 32
Soient Ω un ensemble, $\mathcal{A}$ une tribu sur Ω, et $\mathbb{P}$ une mesure de probabilité sur $(\Omega, \mathcal{A})$.
Le triplet $(\Omega, \mathcal{A}, \mathbb{P})$ est appelé un **espace probabilisé** (ou **espace de probabilité**).

Un espace de probabilité est tout simplement un ensemble que l'on a muni d'une mesure de probabilité. Et pour définir une mesure de probabilité, il faut aussi définir une sigma-algèbre.

La définition suivante est fondamentale en théorie des probabilités.
Elle introduit une notion de "vrai ou faux", qui dépend de la probabilité choisie sur la sigma-algèbre $\mathcal{A}$ (sur l'ensemble de tous les événements que l'on va considérer).

DÉFINITION 33
Soit $(\Omega, \mathcal{A}, \mathbb{P})$ un espace probabilisé.
Soient $A, B \in \mathcal{A}$.

- Si on a $\mathbb{P}(A) = 0$, on dit alors que A est un **événement négligeable** (un événement de probabilité nulle).
- Si on a $\mathbb{P}(B) = 1$, on dit alors que B est un événement **vrai $\mathbb{P}$-presque-sûrement** (un événement de probabilité 1).
  Cela veut dire que l'ensemble des $\omega \in \Omega$ tels que $\omega \notin B$ est un ensemble de probabilité nulle. (un événement négligeable).

REMARQUE 34 — *Une chose qui est importante est que pour une mesure de probabilité $\mathbb{P}$ donnée, on peut avoir plusieurs parties A telles que $\mathbb{P}(A) = 0$ (pas seulement $A = \emptyset$).*
Une partie de probabilité nulle modélise un événement qui n'arrivera jamais. Ainsi, il est important de savoir identifier les événements de probabilité nulle, afin de les écarter dans les calculs.

Pour bien comprendre ce qu'est une mesure de probabilité, et comment s'en servir, comment utiliser ses propriétés, nous allons plonger dans des exemples.
Commençons par le cas où l'ensemble Ω est fini.

17.3 Probabilités sur un ensemble fini - Calcul combinatoire

Dans ce paragraphe, nous supposons que l'ensemble Ω est fini.
Nous rappelons que dans ce cas, nous choisirons toujours $\mathcal{A} = \mathcal{P}(\Omega)$.
Comme l'ensemble Ω est fini, nous rappelons aussi qu'en posant $n = Card(\Omega)$, on pourra écrire $\Omega = \{\omega_1, \ldots, \omega_n\}$.
L'ensemble Ω est alors en bijection avec $\{1, \ldots, n\}$, donc on pourra aussi se ramener à considérer des mesures de probabilités sur $\{1, \ldots, n\}$.

Proposition 35 (**Caractérisation des mesures de probabilités**)*Soit $\Omega = \{\omega_1, \cdots, \omega_n\}$ un ensemble fini.*

(i) Soit $\mathbb{P}$ une mesure de probabilité sur Ω, et soit $A \in \mathcal{P}(\Omega)$. Nous avons alors

$$\mathbb{P}(A) = \sum_{\omega \in A} \mathbb{P}(\{\omega\}).$$

(ii) Soit $\mathbb{P}$ une mesure de probabilité sur Ω.
La fonction $\mathbb{P}$ est entièrement caractérisée par ses valeurs sur les singletons, c'est-à-dire par la liste

$$(\mathbb{P}(\{\omega_1\}), \ldots, \mathbb{P}(\{\omega_n\})).$$

(iii) Soit $(p_i)_{1 \leq i \leq n}$ une famille de nombres réels incluse dans $[0,1]$ et telle que $\sum_{i=1}^n p_i = 1$.
Alors, il existe une unique mesure de probabilité $\mathbb{P}$ telle que l'on ait $p_i = P(\{\omega_i\})$ pour tout $\omega_i \in \Omega$.
On a : $\mathbb{P}(A) = \sum_{i=1}^n p_i . \chi_A(\omega_i) = \sum_{i=1}^n p_i \delta_{\omega_i}(A)$.

Preuve —

1. Comme l'ensemble Ω est fini, l'esemble A est fini.
   Donc, dans l'écriture $A = \cup_{\omega \in A}\{\omega\}$, on a une réunion finie.
   Par propriété de la mesure de probabilité $\mathbb{P}$, on a donc :
   $$\mathbb{P}(A) = \sum_{\omega \in A} \mathbb{P}(\{\omega\}).$$
2. Soient $\mathbb{P}$ et Q deux mesures de probabilités sur Ω telles que $\mathbb{P}(\{\omega_i\}) = Q(\{\omega_i\})$, pour tout $1 \leq i \leq n$.
   Alors, pour toute partie A dans $\mathcal{P}(\Omega)$, on a
   $$\mathbb{P}(A) = \sum_{\omega \in A} \mathbb{P}(\{\omega\}) = \sum_{\omega \in A} Q(\{\omega\}) = Q(A).$$
   On en déduit donc que l'on a $\mathbb{P} = Q$.
3. Soit $(p_i)_{1 \leq i \leq n}$ avec $p_i \in [0,1]$ et $\sum_{i=1}^n p_i = 1$.
   On définit la fonction $\mathbb{P} : \mathcal{P}(\Omega) \to [0,1]$ par $\mathbb{P}(A) = \sum_{i=1}^n p_i \chi_A(\omega_i)$.
   • On remarque en premier lieu que $\mathbb{P}$ est bien définie : on a toujours $\mathbb{P}(A) \in [0,1]$.
   • On a bien $\mathbb{P}(\{\omega_i\}) = p_i$ par construction de $\mathbb{P}$.
   • On a bien $\mathbb{P}(\Omega) = \sum_i p_i = 1$, donc $\mathbb{P}$ vérifie la condition $(*)$.
   • Pour la preuve de la propriété $(**)$, cela est identique à la preuve du même point pour la mesure de proba. uniforme.
   • La preuve de l'unicité vient du point *(ii)*.

□

Remarque 36 —

- *Ainsi, pour définir une mesure de probabilité sur $\{1, \ldots, n\}$, il faut et suffit simplement de choisir des nombres réels $p_1, \ldots, p_n$ dans $[0,1]$ et dont la somme vaut 1.*
  Avec cette proposition, nous pouvons donc construire très facilement toutes les mesures de probabilités que l'on veut sur un ensemble de mesure finie.
  La partie difficile est ensuite le calcul de la probabilité $\mathbb{P}(A)$ pour une partie A donnée.
- ***Remarque de notation :** Pour une mesure de probabilité $\mathbb{P}$ et pour $\omega \in \Omega$, on calcule la probabilité du singleton $\{\omega\}$.*
  La fonction $\mathbb{P}$ est définie sur $\mathcal{P}(\Omega)$ et pas sur Ω, donc parler de "$P(\omega)$" n'a pas de sens !

Exemple 37 — *Loi de Bernoulli de paramètre $p \in [0,1]$: $\mathcal{B}(p)$*
On prend un ensemble Ω à deux éléments, et un nombre réel $p \in [0,1]$. On pose

$$\Omega = \{\omega_1, \omega_2\} \text{ et } p_{\omega_1} = p,\ p_{\omega_2} = 1 - p.$$

Le singleton $\{p_1\}$ sera ainsi de probabilité p, et le singleton $\{p_2\}$ de probabilité $1 - p$.
La mesure de probabilité associée modélise en particulier la chance pour une pièce de tomber sur Pile (ou Face) dans un jeu de pile ou face.
Dans ce cas $\Omega = \{P, F\}$ peut être assimilé à $\{0,1\}$.

• *Pour un lancer de pièce avec une pièce "équilibrée", le nombre réel p sera égal à $\frac{1}{2}$.*
On retrouve alors une mesure de probabilité uniforme.

DÉFINITION 38 *Soit $\Omega = \{\omega_1, \ldots, \omega_n\}$ un ensemble fini, et $\mathbb{P}$ une mesure de probabilité sur Ω.*
Le n-uplet $(\mathbb{P}(\{\omega_1\}), \ldots, \mathbb{P}(\{\omega_n\}))$, $\omega_i \in \Omega\}$ est appelé ***loi de probabilité*** *de $\mathbb{P}$.*

Nous avons vu que la loi de probabilité de $\mathbb{P}$ caractérise la fonction $\mathbb{P}$.
Cela est une façon bien plus pratique de définir et de manipuler une mesure de probabilités.

Différentes méthodes de tirages

Un grand exemple de mesure de probabilité est celui de l'urne contenant des boules de couleur.
Le modèle général est le suivant : Une urne contient N boules de k couleurs différentes, réparties en N_1 boules de couleur 1, N_2 boules de couleur 2, ..., N_k boules de couleur k.
Nous appelons

$$p_i = \frac{N_i}{N}$$

la proportion de boules de couleur i.
Tirons au hasard uniforme n boules de cette urne, $n \leq N$, et intéressons-nous à la répartition des couleurs dans l'échantillon obtenu.
Nous notons par $P_{n_1 n_2 \cdots n_k}$ la probabilité d'obtenir n_1 boules de couleur 1, n_2 boules de couleur 2,... , n_k boules de couleur k, avec bien sûr

$$n_1 + n_2 + ... + n_k = n.$$

Vous connaissez trois grandes façons de tirer les boules au hasard : tirage avec remise, tirage sans remise, tirage simultané. Pour chaque tirage, l'ensemble Ω des résultats est différent.
En fonction de la situation, il faudra choisir le tirage qui correspond, sinon les calculs de probabilités ne seront pas bons.

Tirage simultané Nous tirons n boules en même temps.
L'ensemble Ω est alors l'ensemble de toutes les parties possibles de n éléments distincts, et le nombre de résultats possibles est $\binom{N}{n}$.
La probabilité recherchée est (après calculs) :

$$\widehat{p}_{n_1 n_2 \cdots n_k} = \frac{\binom{N_1}{n_1} \cdots \binom{N_k}{n_k}}{\binom{N}{n}}$$

Dans le cas de deux couleurs, on a :

$$\widehat{p}_{n_1, n-n_1} = \frac{\binom{N_1}{n_1}\binom{N-N_1}{n-n_1}}{\binom{N}{n}}.$$

EXEMPLE 39 — *Si dans une usine de fabrication de pièces, nous savons que parmi N pièces usinées il y en a M qui sont à mettre au rebut, et si nous choisissons au hasard uniforme et simultanément un échantillon de n pièces, alors la probabilité pour que cet échantillon contienne k pièces défectueuses est*

$$\frac{\binom{M}{k}\binom{N-M}{n-k}}{\binom{N}{M}}.$$

Tirage sans remise Nous tirons maintenant successivement les boules de l'urne, mais sans les replacer dans l'urne après tirage.
L'ensemble Ω de tous les tirages possibles est alors l'ensemble des listes de n éléments distincts parmi N, et le nombre de cas possibles sera égal au nombre d'arrangements :

$$N(N-1)\cdots(N-n+1) = A_N^n.$$

Cependant, les événements que l'on regarde ne tiennent pas compte de l'ordre (on veut uniquement un certain nombre de boules de chaque couleur). Le calcul des probabilités donne :

$$\widehat{p}_{n_1 n_2 \cdots n_k} = \frac{\binom{N_1}{n_1} \cdots \binom{N_k}{n_k}}{\binom{N}{n}}.$$

On obtient ainsi la même mesure de probabilité que celle du cas de tirage simultané.
Proposition : Le tirage sans remise et le tirage simultané sont deux façons de "tirer au sort" qui sont équivalentes, si l'on ne s'intéresse pas à l'ordre des éléments tirés.

Tirage avec remise Les tirages sont successifs. Nous replaçons la boule tirée dans l'urne avant le tirage suivant. Nous pouvons donc tirer plusieurs fois la même boule.
L'ensemble Ω est alors l'ensemble de tous les n-uplets d'éléments de l'urne.
Comme on a N boules au total, on a donc $Card(\Omega) = N^n$.
On obtient après calculs une probabilité de :

$$P_{n_1 n_2 \cdots n_k} = \frac{n!}{n_1! n_2! ... n_k!} \frac{N_1^{n_1} \cdots N_k^{n_k}}{N^n}$$

Dans le cas particulier où $k = 2$, on pose $p = \dfrac{N_1}{N} = p_1$. La probabilité vaut alors :

$$p_{n_1, n-n_1} = \binom{n}{n_1} p^{n_1} (1-p)^{n-n_1}.$$

Exemples

Exemple 40 — *Les yeux bandés, vous manipulez au hasard 7 fiches où sont écrites les lettres E, E, T, B, R, L, I. Quelle est la probabilité que vous écriviez le mot LIBERTE ?*
Solution : $\frac{nb\ de\ cas\ favorables}{nb\ de\ cas\ possibles} = \dfrac{2}{7!} = \dfrac{1}{2520}$
Le fait de dire "au hasard", et de dire que l'on manipule toutes les fiches de la même façon indique que la mesure de probabilité que l'on choisit pour modéliser l'expérience est la mesure uniforme.

Exemple 41 — *On tire au hasard quatre cartes d'un jeu de cinquante-deux cartes.*
Quelle est la probabilité pour que, parmi ces quatre cartes, il y ait exactement deux rois ?
Solution : L'hypothèse au hasard amène à modéliser cette expérience comme un tirage uniforme dans un certain ensemble Ω qu'il faut préciser.
Ici, on prend pour Ω l'ensemble des parties à 4 éléments de l'ensemble de 52 cartes. Le cardinal de Ω est donc $\binom{52}{4}$, et $\mathbb{P}$ est la probabilité uniforme sur Ω.
Les résultats favorables sont les tirages qui contiennent exactement 2 rois, à savoir 2 rois et 2 cartes parmi les 48 cartes autres que des rois. Ainsi, la probabilité cherchée vaut $\frac{nb\ de\ cas\ favorables}{nb\ de\ cas\ possibles} = \dfrac{\binom{4}{2}\binom{48}{2}}{\binom{52}{4}}$.

Exemple 42 — *On lance trois dés parfaitement équilibrés.*
Montrer que la probabilité que la somme des points dépasse strictement dix est égale à la probabilité que cette somme ne dépasse pas dix. (Cela permettra de construire un jeu parfaitement équitable.)
Solution : L'ensemble Ω est ici l'ensemble des familles (a_1, a_2, a_3) de 3 nombres compris entre 1 et 6, $\{1, .., 6\}^3$, muni de la probabilité $\mathbb{P}$ uniforme.
On remarque que

$$a_1 + a_2 + a_3 > 10 \iff (7 - a_1) + (7 - a_2) + (7 - a_3) \leq 10.$$

Ainsi, si A désigne l'événement "la somme des points est strictement supérieure à 10", nous remarquons que l'application $(a_1, a_2, a_3) \to (7 - a_1, 7 - a_2, 7 - a_3)$ est une bijection de A sur $\overline{A}$.
Les événements A et $\overline{A}$ ont donc même cardinal, et donc même probabilité de réalisation (qui vaut donc $\frac{1}{2}$).

Remarque 43 — *Une difficulté majeure dans ce genre de calculs combinatoires est de bien préciser le modèle probabiliste (l'ensemble et la mesure de probabilité que l'on choisit).*
De célèbres paradoxes sont nés de cette difficulté.

Exemple 44 — *Rappelons le problème du chevalier de Méré. Ce personnage marquant de la cour de Louis XIV qui "avait très bon esprit, mais n'était pas très bon géomètre" (cf. lettre de Pascal à Fermat du 29 juillet 1654) était un joueur impénitent, toujours à la recherche de règles cachées lui permettant d'avoir un avantage sur ses adversaires. Voici deux de ses règles.*

1. *Il est avantageux de parier sur l'apparition d'au moins un 6 en lançant un dé 4 fois de suite.*

   Cette règle est bonne puisque la probabilité de l'événement qui nous intéresse vaut

   $$1 - \left(\frac{5}{6}\right)^4 \simeq 0.5177 > \frac{1}{2}.$$

   La différence avec $\dfrac{1}{2}$ est faible, mais apte à fournir à long terme des gains assurés : le chevalier devait jouer souvent...

2. *Il est avantageux de parier sur l'apparition d'au moins un double-six en lançant deux dés 24 fois de suite.*

   Cette règle est mauvaise, puisque la probabilité de l'événement cherché vaut :

$$1 - \left(\frac{35}{36}\right)^{24} = 0.4914 < \frac{1}{2}.$$

   Le Chevalier était donc moins heureux avec cette règle qu'avec la précédente. En fait, il s'était laissé abuser par un soi-disant argument d'homothétie : en lançant un dé, il y a 6 résultats possibles, en lançant deux dés, il y en a $6^2 = 36$*, soit 6 fois plus.*
   Comme il est avantageux de parier sur l'apparition d'au moins un 6 en lançant un dé 4 fois de suite, il doit être avantageux de parier sur l'apparition d'un double-six en lançant deux dés $4 \times 6 = 24$ *fois de suite.*

17.4 Probabilités sur un ensemble dénombrable

On suppose dans cette section Ω est dénombrable.
L'ensemble est en bijection avec $\mathbb{N}$, donc nous pouvons numéroter ses éléments : $\Omega = \{\omega_i,\ i \in \mathbb{N}\}$.
La proposition suivante généralise au cas dénombrable la proposition vue dans le cas fini.
On rappelle que la sigma-algèbre considérée sur Ω est l'ensemble $\mathcal{P}(\Omega)$.

Proposition 45
Soit Ω un ensemble dénombrable.

1. Soit $(p_n)_{n\geq 0}$ une suite de nombres réels tels que $0 \leq p_n \leq 1$ et $\sum_{n=0}^{+\infty} p_n = 1$.
   Alors il existe une unique mesure de probabilité $\mathbb{P}$ telle que pour tout $A \subset \Omega$, on ait

$$\mathbb{P}(A) = \sum_{\omega_n \in A} p_n = \sum_{\omega_n \in A} \mathbb{P}(\{\omega_n\}).$$

2. Soit $\mathbb{P}$ une probabiité sur $(\Omega, \mathcal{P}(\Omega))$.
   La fonction $\mathbb{P}$ est entièrement caractérisée par ses valeurs sur les singletons, c'est-à-dire par la famille des

$$(\mathbb{P}(\{\omega_i\})_{i\geq 0}.$$

Preuve —

1. **Existence :** On définit la fonction $\mathbb{P}$ sur $\mathcal{P}(\Omega)$ par $\mathbb{P}(A) = \sum_{\omega_n \in A} p_n$.

   Rappelons que cette somme infinie vaut :

$$\sum_{\omega_n \in A} p_n = \sup_{B \subset A,\ |B|<\infty} \sum_{\omega_n \in B} p_n$$

   Cette somme de termes positifs est majorée par $\sum_{n=0}^{+\infty} p_n = 1$, donc elle est finie.

   • Ainsi, la fonction $\mathbb{P}$ est bien définie.

   • On remarque que l'on a $\mathbb{P}(\Omega) = \sum_{n=0}^{+\infty} p_n = 1$. Donc la propriété $(*)$ est vérifiée.

   Si A est un ensemble fini, on en déduit par additivité de $\mathbb{P}$ que

$$\mathbb{P}(A) = \sum_{\omega_n \in A} \mathbb{P}(\{\omega_n\}).$$

   Enfin, si $A \subset \Omega$ est dénombrable, alors A correspond à une suite extraite $(\omega_{\rho(n)})_{n\geq 0}$ de $(\omega_n)_{n\geq 0}$.
   Par σ-additivité

$$\mathbb{P}(A) = \mathbb{P}\left(\bigcup_{n=0}^{+\infty} \{\omega_{\rho(n)}\}\right) = \sum_{n=0}^{+\infty} p_{\rho(n)}.$$

   Ceci montre que si $\mathbb{P}$ existe, elle est uniquement déterminée. Il reste à montrer que $\mathbb{P}$ vérifie bien les axiomes d'une probabilité :
   • Soit $(A_n)_{n\geq 0}$ une famille dénombrable d'éléments de $\mathcal{A}$ deux à deux disjoints. On note A leur union.

$$\mathbb{P}(A) = \sum_{\omega_n \in A} p_n = \sum_{k=0}^{+\infty} \left(\sum_{\omega_n \in A_k} p_n \right) = \sum_{k\geq 0} \mathbb{P}(A_k).$$

   Cela découle du théorème de sommation par paquets (voir chapitre Dénombrement). Ainsi la fonction $\mathbb{P}$ vérifie la propriété $(**)$, c'est donc une mesure de probabilité. **Unicité :** Si on avait $\mathbb{P}$ et Q deux mesures de probabilité telles que

$$\mathbb{P}(A) = \sum_{\omega_n \in A} p_n = Q(A),$$

   alors on a $\mathbb{P} = Q$ car ces fonctions sont égales sur $\mathcal{P}(\Omega)$.

2. On pose $p_n = \mathbb{P}(\{\omega_n\})$.
Alors la famille $(p_n)_{n\geq 0}$ vérifie les conditions du point 1).
Par unicité de la mesure de probabilité associée aux $(p_n)_n$, on en déduit que pour toute partie $A \in \mathcal{P}(\Omega)$, on a
$$\mathbb{P}(A) = \sum_{\omega_n \in A} \mathbb{P}(\{\omega_n\}),$$
ce qui montre que $\mathbb{P}$ est entièrement déterminée par sa valeur en les singletons.

□

Exemple 46 — *Soit $\theta > 0$ et $p_n = e^{-\theta}\dfrac{\theta^n}{n!}$.*
Il est facile de vérifier que $0 \leq p_n \leq 1$ et que
$$\sum_{n=0}^{+\infty} p_n = e^{-\theta}\sum_{n=0}^{+\infty}\frac{\theta^n}{n!} = 1$$
La suite $(p_n)_{n\in\mathbb{N}}$ définit une probabilité sur $\mathbb{N}$, appelée ***loi de Poisson de paramètre*** *θ : $\mathcal{P}(\theta)$.*

Exemple 47 — *Soit $\Omega = \Omega = \{\omega_0, \omega_1, \cdots, \omega_n, \cdots\}$ un ensemble dénombrable $\mathbb{P}$ une mesure probabilité sur Ω, et $p_n = \mathbb{P}(\{\omega_n\})$. Alors pour tout $A \in \mathcal{A}$, on a*
$$\mathbb{P}(A) = \sum_{n\in\mathbb{N}} p_n \delta_{\omega_n}(A).$$
La mesure de probabilité $\mathbb{P}$ peut donc s'écrire comme somme de la série de fonctions :
$$\mathbb{P} = \sum_{n=0}^{+\infty} p_n \delta_{\omega_n}.$$
Ainsi, toute mesure de probabilité sur un ensemble fini ou dénombrable peut s'écrire comme une "combinaison convexe" de mesures de Dirac. (une combinaison convexe est une combinaison linéaire avec tous les coefficients positifs ou nuls, et dont la somme vaut 1)

17.5 Conditionnement et indépendance

La notion de conditionnement est l'une des plus fructueuses de la théorie des probabilités (de regarder des résultats "à condition que", "sachant que").
L'idée de base qui permet de comprendre de cette notion est la suivante : une information supplémentaire sur l'expérience modifie la vraisemblance que l'on accorde à l'événement étudié.

Exemple 48 — *Cherchons, pour un lancer de deux dés équilibrés, la probabilité de l'événement "la somme est supérieure ou égale à 10". Elle vaut*
- $\dfrac{1}{6}$ *sans information supplémentaire.*
- $\dfrac{1}{2}$ *si l'on sait que le résultat du second dé est* 6
- 0 *si l'on sait a priori que le résultat d'un des dés est* 2.

Pour obtenir ces résultats, nous avons dans chaque cas calculé le rapport du nombre de résultats favorables sur le nombre de cas possibles.
Il est à chaque fois indispensable de bien définir l'espace probabilisé associé à l'expérience en tenant compte des informations a priori.
On remarque que l'information a priori change la valeur de la probabilité de l'événement aléatoire. (c'est le même événement, mais regardé ici sur des espaces probabilisés un peu différents)

L'approche pour donner un sens mathématique à cette notion se base à nouveau sur la notion de fréquence d'apparition.
Cela donne la définition suivante.

Définition 49
Soit $(\Omega, \mathcal{A}, \mathbb{P})$ un espace probabilisé.
Soit $A, B \in \mathcal{A}$ deux événements, avec $\mathbb{P}(B) > 0$.
On appelle la **probabilité conditionnelle** de A sachant B le nombre
$$\mathbb{P}(A|B) = \mathbb{P}_B(A) = \frac{\mathbb{P}(A \cap B)}{\mathbb{P}(B)}.$$

Pour B donné, la fonction $\mathbb{P}(.|B)$ définit encore une mesure de probabilité.

PROPOSITION 50
Soit $(\Omega, \mathcal{A}, \mathbb{P})$ un espace probabilisé.
Soit $B \in \mathcal{A}$ une partie telle que $\mathbb{P}(B) > 0$. Alors,

1. La fonction $\mathbb{P}(.|B) : A \in \mathcal{A} \mapsto \mathbb{P}(A|B) = \frac{\mathbb{P}(A)}{\mathbb{P}(B)} \in [0,1]$ est une mesure de probabilité sur Ω. On l'appelle **mesure de probabilité conditionnelle** sachant B.
2. Soit $A \in \mathcal{A}$. Si $\mathbb{P}(A) > 0$ et $\mathbb{P}(B) > 0$, on a
$$\mathbb{P}(A|B)\mathbb{P}(B) = \mathbb{P}(A \cap B) = \mathbb{P}(B|A)\mathbb{P}(A).$$

Preuve — • Comme $\mathbb{P}(B) > 0$, il est clair que $0 \leq \mathbb{P}(A|B) \leq 1$, donc $\mathbb{P}(.|B)$ est bien définie.
• De même, on a $\mathbb{P}(\Omega|B) = 1$.
• Soit $(A_n)_{n\geq 0}$ une famille d'élements de $\mathcal{A}$ qui sont deux à deux disjoints.
Alors $(A_n \cap B)_{n\geq 0}$ est encore une famille d'éléments de $\mathcal{A}$ deux à deux disjoints.
Par σ additivité de $\mathbb{P}$, on obtient
$$\mathbb{P}(\bigcup_{n\geq 0} A_n|B) = \frac{\mathbb{P}\left(\bigcup_{n\geq 0}(A_n \cap B)\right)}{\mathbb{P}(B)} = \frac{\sum_{n\geq 0}\mathbb{P}(A_n \cap B)}{\mathbb{P}(B)} = \sum_{n\geq 0}\mathbb{P}(A_n|B).$$
Cela montre bien que $\mathbb{P}(.|B)$ est une mesure de probabilités.
Le point 2) découle de la définition de la probabilité conditionnelle. □

PROPOSITION 51 (Formule des probabilités composées)
Soit $(\Omega, \mathcal{A}, \mathbb{P})$ un espace probabilisé.
Soient $A_1, \ldots, A_n \in A$ tels que $\mathbb{P}(A_1 \cap A_2 \cap \cdots \cap A_{n-1}) > 0$.
Alors, on a
$$\mathbb{P}(A_1 \cap A_2 \cap \cdots \cap A_n) = \mathbb{P}(A_1)\mathbb{P}(A_2|A_1)\mathbb{P}(A_3|A_1 \cap A_2)\cdots\mathbb{P}(A_n|A_1 \cap A_2 \cap \cdots \cap A_{n-1}).$$

Preuve — Par hypothèse, on a $\mathbb{P}(A_1 \cap \ldots \cap A_k) > 0$ pour tout $1 \leq k \leq n$.
On peut alors écrire le télescopage suivant :
$$\mathbb{P}(A_1 \cap A_2 \cap \ldots \cap A_n) = \frac{\Pi_{i=1}^n \mathbb{P}(A_1 \cap \ldots \cap A_i)}{\Pi_{j=1}^{n-1}\mathbb{P}(A_1 \cap \ldots \cap A_j)} = \mathbb{P}(A_1)\Pi_{k=2}^n \frac{\mathbb{P}(A_1 \cap \ldots \cap A_k}{\mathbb{P}(A_1 \cap \ldots A_{k-1})}.$$
Cela donne le résultat. □

Pour $A, B \in \mathcal{A}$ deux événéments sur Ω, regarder la probabilité conditionnelle de A sachant B ne donne qu'une information partielle sur la probabilité de A.
Avec suffisamment d'événements B bien choisis (en utilisant des partitions), on peut arriver à retrouver $\mathbb{P}(A)$.

DÉFINITION 52 (**Partition**)*Soient Ω un ensemble, $n \in \mathbb{N}^*$, et $A_1, \ldots, A_n$ des parties de Ω.*
*On dit que la famille $(A_1, \ldots, A_n)$ est une **partition** de Ω si :*

1. *Les A_i sont deux à deux disjoints : $A_i \cap A_j = \emptyset$, $\forall i, j \in \{1, \ldots, n\}$ t.q. $i \neq j$*
2. *La réunion des A_i vaut Ω : $\cup_{i=1}^n A_i = \Omega$.*

DÉFINITION 53 (**Système complet d'événements**)*Soit $(\Omega, \mathbb{P})$ un espace probabilisé.*
Soit $(B_1, \ldots, B_n)$ une partition de Ω, telle que $\mathbb{P}(B_i) > 0$ pour chaque i.
*Une telle partition est appelée **système complet d'événements**.*

PROPOSITION 54 (**Formule des probabilités totales**)
Soit $(\Omega, \mathcal{A}, \mathbb{P})$ un espace probabilisé, et soit $(B_i)_{i\in I}$ un système complet d'événements de Ω.
Alors, pour tout $A \in \mathcal{A}$, on a
$$\mathbb{P}(A) = \sum_{i\in I}\mathbb{P}(A \cap B_i) = \sum_{i\in I}\mathbb{P}(A|B_i)\mathbb{P}(B_i).$$

Preuve — Nous avons $A = \bigcup_{i\in I}(A \cap B_i)$. Par hypothèse, les ensembles $(A \cap B_i)$ sont deux-à-deux disjoints, et par ailleurs $\mathbb{P}(A \cap B_i) = \mathbb{P}(A|B_i)\mathbb{P}(B_i)$. La σ-additivité donne le résultat. □

THÉORÈME 55 (**Formule de Bayes**)
Soit $(\Omega, \mathcal{A}, \mathbb{P})$ un espace probabilisé, et soit $(B_i)_{i\in I}$ un système complet d'événements de Ω.
Soit $A \in \mathcal{A}$ tel que $\mathbb{P}(A) > 0$. Alors, on a
$$\mathbb{P}(B_i|A) = \frac{P(A|B_i)P(B_i)}{\sum_{j\geq 0}\mathbb{P}(A|B_j)P(B_j)}, \ \forall i \in I.$$

Preuve — On sait d'après la proposition précédente que $\sum_{j\geq 0}\mathbb{P}(A|B_j)P(B_j)=\mathbb{P}(A)$, et par définition d'une probabilité conditionnelle on a

$$\mathbb{P}(B_i|A)=\frac{\mathbb{P}(A\cap B_i)}{\mathbb{P}(A)}=\frac{\mathbb{P}(A|B_i)\mathbb{P}(B_i)}{\mathbb{P}(A)}.$$

□

Remarque : Notre intuition habituelle est très mauvaise quand il s'agit d'estimer certaines probabilités conditionnelle !
La formule de Bayes est le résultat qui permet de mettre cela en évidence.

Exemple 56 — *Un individu est tiré au hasard dans une population où l'on trouve une proportion* 10^{-4} *de séropositifs. On lui fait passer un test de détection de la séropositivité.*
Par ailleurs, des essais antérieurs ont permis de savoir que la probabilité d'avoir un résultat positif lors du test si l'individu est séropositif est $0,99$ *(c'est la sensibilité du test, la proba de vrais positifs), et que celle d'avoir un résultat positif si l'individu n'est pas séropositif est de* $0,001$ *(*$0,999=1-0,001$ *est la spécificité du test, la proba de vrais négatifs).*
Sachant que le test donne un résultat positif, quelle est la probabilité pour que l'individu soit vraiment séropositif ?
Solution : Considérons les événements A "l'individu est séropositif", et B "le test de détection donne un résultat positif".
Les données de l'énoncé fournissent $\mathbb{P}(A)=10^{-4}$*, donc* $\mathbb{P}(\overline{A})=0,9999$*, ainsi que* $\mathbb{P}(B|A)=0,99$ *(vrai positif) et* $\mathbb{P}(B|\overline{A})=0,001$ *(faux positif).*
Nous trouvons alors

$$\begin{aligned}\mathbb{P}(A|B) &= \frac{\mathbb{P}(A\cap B)}{\mathbb{P}(B)}\\ &= \frac{\mathbb{P}(B|A)\mathbb{P}(A)}{\mathbb{P}(B|A)\mathbb{P}(A)+\mathbb{P}(B|\overline{A})\mathbb{P}(\overline{A})}\\ &= \frac{0,99\times 10^{-4}}{0,99\times 10^{-4}+0,001\times 0,9999}\simeq 0,09\end{aligned}$$

Contrairement à l'intuition, cette probabilité est plutôt faible ($0.09=\frac{9}{100}=9\%$*).*
La proportion de gens séropositifs est très faible, ce qui fait que même si le test détecte très bien la maladie, le volume de faux positifs est finalement bien plus important que le volume de vrais positifs.
Dans cette population, ce test, bien que très efficace, n'est pas extrêmement fiable (le test seul n'est pas suffisant pour vraiment savoir si on est séropositif ou pas).

Exemple 57 — *On classe les gérants de portefeuilles en deux catégories, les bien informés et les autres.*
Lorsqu'un gérant bien informé achète une valeur boursière pour son client, on peut montrer par une étude préalable que la probabilité que le cours de cette valeur monte est de $0,8$*.*
Si le gérant est mal informé, la probabilité que le cours descende est de $0,6$*.*
On sait par ailleurs que si l'on choisit au hasard un gérant de portefeuille, il y a une chance sur 10 *que celui-ci soit un gérant bien informé.*
Un client choisit au hasard un gérant dans l'annuaire, et lui demande d'acheter une valeur. Sachant que le cours de cette valeur est monté, cherchons la probabilité pour que le gérant soit mal informé.
Solution : Notons M l'événement "la valeur monte" et I l'événement "le gérant est bien informé".
Par la formule des probabilités totales, la probabilité que la valeur monte vaut

$$\mathbb{P}(M)=\mathbb{P}(M|I)\mathbb{P}(I)+\mathbb{P}(M|\overline{I})\mathbb{P}(\overline{I})=0,8\times 0,1+0,4\times 0,9=0,44.$$

La formule de Bayes donne alors

$$\mathbb{P}(\overline{I}|M)=\frac{\mathbb{P}(M|\overline{I})\mathbb{P}(\overline{I})}{\mathbb{P}(M)}=\frac{0,4\times 0,9}{0,44}\simeq 0,818.$$

17.5.1 Evénements indépendants

La notion d'indépendance est un outil absolument fondamental en probabilités.

Intuitivement, deux événements A et B sont indépendants si le fait de savoir que A est réalisé ne donne aucune information sur la réalisation de B, et réciproquement.
L'indépendance est modélisée mathématiquement par cette définition.

Définition 58
Soient $(\Omega,\mathcal{A},\mathbb{P})$ un espace probabilisé, et $A,B\in\mathcal{A}$ deux événements.
On dit que les événements A et B **sont indépendants** si $\mathbb{P}(A\cap B)=\mathbb{P}(A)\mathbb{P}(B)$.

Remarque 59 —

1. *Si* $\mathbb{P}(A) > 0$ *et* $\mathbb{P}(B) > 0$*, alors*

$$\mathbb{P}(A \cap B) = \mathbb{P}(A)\mathbb{P}(B) \Leftrightarrow \mathbb{P}(A|B) = \mathbb{P}(A) \Leftrightarrow \mathbb{P}(B|A) = \mathbb{P}(B).$$

Ainsi, si A *est indépendant de* B*, la probabilité de voir* A *réalisé ne dépend pas de la réalisation de* B*, et réciproquement.*

2. *La notion d'indépendance est une notion qui dépend totalement de la mesure de probabilité* $\mathbb{P}$*.*
Cette notion n'a rien à voir avec les opérations ensemblistes dans $\mathcal{P}(\Omega)$*.*
Par exemple, cela n'a rien à voir avec le fait que A *et* B *soient disjoints ou non. (Cf. Exemple ci-dessous).*

Exemple 60 —

1. *On lance* 3 *fois de suite un dé équilibré.*
Si A_i *est un événement qui ne dépend que du* i*-ème lancer, alors* A_1*,* A_2*,* A_3 *sont indépendants (pour la mesure uniforme).*

2. *Si deux événements* A *et* B *sont disjoints mais pas de probabilité nulle, on a alors* $\mathbb{P}(A \cap B) = \mathbb{P}(\emptyset) = 0$ *et* $\mathbb{P}(A)\mathbb{P}(B) > 0$*.*
Donc A *et* B *ne sont pas indépendants pour la mesure de probas* $\mathbb{P}$*. (par exemple* $A = \{$*faire Pile*$\}$ *et* $B = \{$*faire Face*$\}$ *dans un jeu de Pile ou Face équilibré)*

3. *On tire une carte au hasard uniforme dans un jeu de* 52 *cartes.*

$$A = \{\textit{la carte est une dame}\}\,; \quad B = \{\textit{la carte est un coeur}\}.$$

Il est facile de voir que $\mathbb{P}(A) = \dfrac{4}{52}$*,* $\mathbb{P}(B) = \dfrac{13}{52}$*, et*

$$\mathbb{P}(A \cap B) = \mathbb{P}(\{\textit{la carte est la dame de coeur}\}) = \frac{1}{52} = \mathbb{P}(A)\mathbb{P}(B).$$

Ainsi, les événements A *et* B *sont indépendants pour la mesure uniforme* $\mathbb{P}$*.*

4. *Supposons maintenant que le jeu de cartes soit trafiqué.*
Soit $\mathbb{Q}$ *la nouvelle mesure de probabilité correspondant au tirage de cartes. Supposons que*

$$\mathbb{Q}(\{\textit{dame de coeur}\}) = \frac{1}{2},\ \mathbb{Q}(\{\textit{autre carte}\}) = \frac{1}{2} \times \frac{1}{51} = \frac{1}{102}.$$

Alors

$$\mathbb{Q}(A \cap B) = \frac{1}{2} \neq \mathbb{Q}(A)\mathbb{Q}(B) = (\frac{1}{2} + \frac{3}{102}) \times (\frac{1}{2} + \frac{12}{102}).$$

Les événements A *et* B *ne sont pas indépendants pour la mesure de probas* $\mathbb{Q}$*.*

Proposition 61
Soient $(\Omega, \mathcal{A}, \mathbb{P})$ un espace probabilisé, et $A, B \in \mathcal{A}$ deux événements.
Si A et B sont indépendants, alors il en est de même de A et $\overline{B}$, $\overline{A}$ et B, $\overline{A}$ et $\overline{B}$.

Preuve — Supposons A et B indépendants.

$$\mathbb{P}(A) = \mathbb{P}(A \cap B) + \mathbb{P}(A \cap \overline{B}) \Rightarrow \mathbb{P}(A \cap \overline{B}) = \mathbb{P}(A) - \mathbb{P}(A)\mathbb{P}(B) = \mathbb{P}(A)\mathbb{P}(\overline{B})$$

Donc A et $\overline{B}$ sont indépendants.

$$\mathbb{P}(\overline{A}) = \mathbb{P}(\overline{A} \cap \overline{B}) + \mathbb{P}(\overline{A} \cap B) \Rightarrow \mathbb{P}(\overline{A} \cap \overline{B}) = \mathbb{P}(\overline{A}) - \mathbb{P}(\overline{A})\mathbb{P}(B) = \mathbb{P}(\overline{A})\mathbb{P}(\overline{B})$$

où l'on a appliqué la première partie de la preuve à $\overline{A}$ et B.
Les deux derniers cas s'en déduisent immédiatement. □

La notion d'indépendance se généralise à une famille finie ou dénombrable d'événements de la manière suivante.

Définition 62
Soient $(\Omega, \mathcal{A}, \mathbb{P})$ un espace probabilisé, et $(A_n)_{n\geq 0} \in \mathcal{A}^{\mathbb{N}}$ une famille d'événements.
On dit que cete famille est **indépendante** si

$$\mathbb{P}(A_{i_1} \cap \cdots \cap A_{i_k}) = \mathbb{P}(A_{i_1}) \cdots \mathbb{P}(A_{i_k})$$

pour toute famille finie finie $(i_1, \cdots, i_k)$ d'entiers, avec $i_1 < i_2 < \ldots < i_k$.

Remarque 63 — *Il faut faire très attention avec cette définition.*

1. *Pour que la suite (A, B, C) soit indépendante, la propriété doit être vérifiée pour toutes les intersections de deux ensembles et l'intersection des 3 ensembles. Il ne suffit pas d'avoir*
$$\mathbb{P}(A \cap B \cap C) = P(A)P(B)P(C).$$
Par exemple, prenons un lancer de 1 dé avec $A = \{1, 2, 3\}$, $B = \{2, 4, 6\}$ et $C = \{1, 2, 4, 5\}$.
Nous avons $\mathbb{P}(A) = \frac{1}{2}$, $\mathbb{P}(B) = \frac{1}{2}$, $\mathbb{P}(C) = \frac{2}{3}$.
Ainsi, nous avons bien $\mathbb{P}(A \cap B \cap C) = \mathbb{P}(A)\mathbb{P}(B)\mathbb{P}(C)$, mais $\mathbb{P}(A \cap B) \neq \mathbb{P}(A)\mathbb{P}(B)$, donc la famille (A, B, C) n'est pas indépendante.
2. *Il ne suffit pas non plus que les événements soient indépendants deux à deux.*
Par exemple, on joue 2 fois à Pile ou Face (avec pièce équilibrée) et on considère les événements $A = \{$ Face au premier lancer $\}$, $B = \{$ Face au deuxième lancer $\}$ et $C = \{$ les deux tirages donnent le même résultat $\}$.
On vérifie que ces événements sont deux à deux indépendants, mais par contre on a $\mathbb{P}(A \cap B \cap C) \neq \mathbb{P}(A)\mathbb{P}(B)\mathbb{P}(C)$, donc la famille (A, B, C) n'est pas indépendante.

17.5.2 Lemme de Borel-Cantelli

On termine ce chapitre par un résultat aux conséquences vraiment surprenantes et qui permet de mieux comprendre la notion d'indépendance.
Nous commençons par définir la limite supérieure et inférieure d'une suite de parties d'un ensemble.

Définition 64
Soient Ω un ensemble, $\mathcal{A}$ une σ-algèbre, et $(A_n)_{n\in\mathbb{N}}$ une suite d'éléments de $\mathcal{A}$.
On définit la **limite supérieure** de la famille $(A_n)_n$ comme l'ensemble
$$\limsup_n A_n = \bigcap_p \left(\bigcup_{n\geq p} A_n \right) \in \mathcal{A},$$
et la **limite inférieure** de la famille $(A_n)_n$ comme l'ensemble
$$\liminf_n A_n = \bigcup_p \left(\bigcap_{n\geq p} A_n \right) \in \mathcal{A}.$$

Remarque 65 —

1. *L'ensemble $\limsup_n A_n$ est l'ensemble des ω qui apparaissent une infinité de fois parmi les A_n.*
Inversement, on a $\omega \notin \limsup_n A_n$ ssi ω appartient à au plus un nombre fini de A_n.
On peut remarquer que la suite des $(\bigcup_{n\geq p} A_n)_p$ est une suite décroissante.
L'ensemble $\limsup_n(A_n)$ est donc une limite de suite décroissante pour l'inclusion.
2. *L'ensemble $\liminf_n A_n$ est l'ensemble des ω qui apparaissent dans tous les A_n à partir d'un certain rang p (pour tout $n \geq p$).*
On peut remarquer que la suite des $(\bigcap_{n\geq p} A_n)_p$ est croissante.
L'ensemble $\liminf_n(A_n)$ est donc une limite de suite croissante pour l'inclusion.
3. *On peut aussi remarquer que $\liminf_n(A_n) \subset \limsup_n(A_n)$.*
Ces ensembles ne sont en général pas égaux.

4. *Comme le passage au complémentaire change les unions en intersections, et les intersections en unions, on montre facilement que $\liminf_n(A_n) = \limsup_n(\overline{A_n})$.*

Le lemme de Borel-Cantelli nous permet de déterminer facilement la probabilité de $\limsup_n(A_n)$ dans certaines situations.

Théorème 66 (**Lemme de Borel-Cantelli**)
Soient Ω un ensemble, $\mathcal{A}$ une σ-algèbre, et $(A_n)_{n\in\mathbb{N}}$ une suite d'éléments de $\mathcal{A}$.

1. Si on a $\sum_{n\geq 0} \mathbb{P}(A_n) < +\infty$, alors $\mathbb{P}(\limsup_n A_n) = 0$.
2. Si la famille $(A_n)_{n\geq 0}$ est indépendante, alors on a
$$\sum_{n\geq 0} \mathbb{P}(A_n) = +\infty \text{ implique } \mathbb{P}(\limsup_n A_n) = 1.$$

Preuve —

1. Comme la suite $\left(\bigcup_{n\geq p} A_n\right)_p$ est décroissante, par propriétés de la mesure $\mathbb{P}$ on a

$$\mathbb{P}(\limsup_n A_n) = \lim_{p\to+\infty} \mathbb{P}\left(\bigcup_{n\geq p} A_n\right) \leq \lim_{p\to+\infty} \sum_{n\geq p} \mathbb{P}(A_n).$$

Si la série $\sum_{n\geq 0} \mathbb{P}(A_n)$ est convergente, le reste de cette série tend vers 0. Donc, on a $\mathbb{P}(\limsup A_n) = 0$.

2. Supposons maintenant que les A_n soient indépendants et que la série $\sum_{n\geq 0} \mathbb{P}(A_n)$ diverge.

Soit m un nombre entier. Nous avons

$$\mathbb{P}\left(\bigcup_{i=p}^{m} A_i\right) = 1 - \mathbb{P}\left(\bigcap_{i=p}^{m} \overline{A}_i\right) = 1 - \prod_{i=p}^{m} \mathbb{P}(\overline{A}_i) = 1 - \prod_{i=p}^{m}(1-\mathbb{P}(A_i)) \geq 1 - e^{-\sum_{i=p}^{m} \mathbb{P}(A_i)}$$

grâce à l'inégalité $1 - x \leq e^{-x}$ pour $x \geq 0$.
Ainsi, en passant à la limite on obtient

$$\mathbb{P}\left(\bigcup_{i=p}^{+\infty} A_i\right) \geq 1 - e^{-\sum_{i=p}^{+\infty} \mathbb{P}(A_i)} = 1.$$

On a donc $\mathbb{P}(\bigcup_{i=p}^{+\infty} A_i) = 1$ pour tout $p \geq 1$. Comme la suite des $(\bigcup_{i=p}^{+\infty} A_i)_p$ est décroissante et converge vers $\limsup_n(A_n)$, les propriétés de $\mathbb{P}$ nous donnent $\mathbb{P}(\limsup_n A_n) = \lim_{p\to+\infty} \mathbb{P}\left(\bigcup_{n\geq p} A_n\right) = 1$.

□

Remarque 67 —

1. *Il est clair que le point* 2) *est totalement faux dans le cas où la famille n'est pas indépendante.*
   On peut prendre, par exemple, tous les A_n égaux à un même événement A de probabilité $\mathbb{P}(A) \in]0,1[$.
2. *Le théorème montre que si la suite $(A_n)_{n\geq 0}$ est indépendante, alors $\limsup_n A_n$ est de probabilité 0 ou 1 suivant que la série $\sum_{n\geq 0} \mathbb{P}(A_n)$ converge ou diverge.*
3. *Le lemme de Borel-Cantelli porte sur $\limsup_n(A_n)$. Attention à ne pas confondre $\limsup$ et $\liminf$!*
4. *On peut parfois calculer $\mathbb{P}(\liminf_n(A_n))$ avec Borel-Cantelli en utilisant le fait que $\liminf_n(A_n) = \limsup_n(\overline{A_n})$.*

Exemple 68 — *Supposons que vous vous installez les yeux bandés devant votre clavier et que vous tapez indéfiniment (de façon dénombrable) sur les touches au hasard (probabilité uniforme).*
Prenons M un mot de longueur l, et pour $k \geq 1$ définissons l'évènement A_k : "les lettres lk à $k(l+1)-1$ forment le mot M"
En prenant $\mathbb{P}$ la mesure de probabilité uniforme sur notre ensemble dénombrable, cette famille d'événements est alors indépendante.
On a de plus $\mathbb{P}(A_1) \geq \frac{1}{l!} > 0$, et $\mathbb{P}(A_k) = \mathbb{P}(A_1)$ (la proba de A_k ne dépend pas de k).
Cela donne donc $\sum_{k=0}^{+\infty} \mathbb{P}(A_k) = +\infty$.
On peut donc appliquer le lemme de Borel-Cantelli pour obtenir que la probabilité que le mot M apparaisse une infinité de fois est de 1. (c'est un événement $\mathbb{P}$-presque sûr, et l'événement contraire est de probabilité nulle).

Chapitre 18 Variables aléatoires

Table des matières du chapitre

18.1 Variables aléatoires

Maintenant que nous avons défini les espaces probabilisés $(\Omega, \mathcal{A}, \mathbb{P})$, nous allons pouvoir définir les variables aléatoires.

Définition 1
Soit $(\Omega, \mathcal{A}, \mathbb{P})$ un espace probabilisé. Soit $(F, \mathcal{F})$ où un ensemble muni d'une σ-algèbre.
Soit $X : \Omega \to F$ une fonction.
On dit que X est une **variable aléatoire** de $(\Omega, \mathcal{A}, \mathbb{P})$ dans $(F, \mathcal{F})$ si on a $X^{-1}(B) \in \mathcal{A}$, $\forall B \in \mathcal{F}$.

Remarque 2 — *Le nom donné, qui est utilisé maintenant couramment, n'est pas le mieux choisi : une variable aléatoire, malgré son nom, n'est pas une variable, mais une fonction (une fonction en la variable $\omega \in \Omega$).*
Une variable aléatoire est une fonction !
On peut abréger le nom "variable aléatoire" en ***v.a.****.*

Faisons un exemple.

Exemple 3 — *Étudions un lancer de deux dés équilibrés.*
Dans ce cas, l'ensemble des états est $\Omega = \{(i, j) : 1 \leq i \leq 6; 1 \leq j \leq 6\}$.
On a alors aussi $\mathcal{A} = \mathcal{P}(\Omega)$. Puisque les dés sont équilibrés, on prend pour $\mathbb{P}$ la mesure de probabilité uniforme. Pour $A \subset \Omega$ un événement, on a donc

$$\mathbb{P}(A) = \frac{\operatorname{card} A}{36}.$$

L'application $X : \Omega \to \{1, 2, \cdots, 12\}$ définie par

$$X(i, j) = i + j$$

est la variable aléatoire "somme des résultats des deux dés". Elle a pour loi

$$\mathbb{P}_X(B) = \frac{\textit{nombre de couples } (i, j) \textit{ tels que } i + j \in B}{36}.$$

On peut associer à une variable aléatoire X une mesure de probabilité, de la façon suivante.

PROPOSITION-DÉFINITION 4
Soient $(\Omega, \mathcal{A}, \mathbb{P})$ un espace probabilisé et $(F, \mathcal{F})$ un ensemble avec une σ-algèbre. Soit $X : \Omega \to F$ une variable aléatoire.
On définit la fonction $\mathbb{P}_X : \mathcal{F} \to [0,1]$ par

$$\mathbb{P}_X(B) = \mathbb{P}(X^{-1}(B)) =_{def} \mathbb{P}(X \in B).$$

Alors, la fonction $\mathbb{P}_X$ est une mesure de probabilité.
Cette mesure de probabilité est appelée **loi de la variable aléatoire** X.
Dans le langage probabiliste, on notera aussi $\mathbb{P}_X(B) =_{def} \mathbb{P}(X \in B)$.

Preuve — Comme X est une variable aléatoire, tous les $X^{-1}(B)$ sont des éléments de $\mathcal{A}$, donc la onction $\mathbb{P}_X$ est bien définie.
On rappelle que

$$X^{-1}(\emptyset) = \emptyset,\ X^{-1}(F) = \Omega,\ X^{-1}(\overline{B}) = \overline{X^{-1}(B)},$$
$$X^{-1}(\bigcap_i A_i) = \bigcap_i X^{-1}(A_i),\ X^{-1}(\bigcup_i A_i) = \bigcup_i X^{-1}(A_i).$$

Comme $\mathcal{F}$ est une σ-algèbre sur F, on montre alors facilement que $\mathbb{P}_X$ est une mesure de probabilité. (elle vérifie $(*)$ et $(**)$) □

La loi d'une variable aléatoire X donne énormément d'informations sur la variable aléatoire X, tout comme la loi d'une mesure de probabilité $\mathbb{P}$ (si Ω fini ou dénombrable) donne énormément d'informations sur $\mathbb{P}$.

EXEMPLE 5 — *Dans l'exemple précédent du lancer de deux dés équilibrés, et pour X la variable aléatoire "somme des deux faces obtenues", la loi de probabilité de X, $\mathbb{P}_X$, est une mesure de probas sur l'ensemble $\{2, \dots, 12\}$.*
On a par exemple

$$\mathbb{P}_X(\{2\}) = \mathbb{P}_X(\{12\}) = \frac{1}{36},\ \mathbb{P}_X(\{3\}) = \frac{2}{36},\ \mathbb{P}_X(\{5\}) = \frac{4}{36}.$$

REMARQUE 6 — *Pour une expérience aléatoire donnée, en faire une modélisation mathématique implique de trouver une σ-algèbre $\mathcal{F}$ sur l'ensemble d'arrivée F telle que $X^{-1}(B) \in \mathcal{A}$, pour tout $B \in \mathcal{F}$.*
Cette σ-algèbre $\mathcal{F}$ peut être a priori diffcile à décrire.

Donnons quelques exemples d'expériences aléatoires classiques que l'on va chercher à modéliser en mathématiques avec des espaces probabilisés et des variables aléatoires.

EXEMPLE 7 —

1. *Le nombre de 6 obtenus dans un lancer de 3 dés équilibrés.*
   Les espaces sont $\Omega = \{1, \dots, 6\}^3$, $\mathcal{A} = \mathcal{P}(\Omega)$, $\mathbb{P}$ la mesure uniforme sur Ω, $F = \{0,1,2,3\}$, $\mathcal{F} = \mathcal{P}(F)$.
   La variable aléatoire est $X : (a_1, a_2, a_3) \in \Omega \mapsto \chi_6(a_1) + \chi_6(a_2) + \chi_6(a_3) \in F$.
2. *Le nombre d'appels dans un central téléphonique pendant une heure $F = \mathbb{N}$.*
3. *La distance du point atteint par une flèche flèche par rapport centre de la cible (cible de 15 cm de rayon) : $F = [0, 15]$.*
4. *La valeur maximale du prix d'un actif sur un intervalle de temps donné : $F = \mathbb{R}_+$.*

En général, l'ensemble F sera un ensemble fini ou dénombrable, ou $\mathbb{R}$ ou $\mathbb{R}^d$, ou un ensemble un peu plus particulier.

REMARQUE 8 —
• *Si l'ensemble Ω est fini ou dénombrable, on utilise $\mathcal{A} = \mathcal{P}(\Omega)$.*
Ainsi, pour toute fonction $X : \Omega \to F$ et pour tout $B \in \mathcal{F}$, on a $X^{-1}(B) \in \mathcal{P}(\Omega)$.
On a donc montré que toute fonction sur un ensemble fini ou dénombrable est une variable aléatoire !
• *Si cette fois l'ensemble $X(\Omega)$ est fini ou dénombrable (ce qui sera souvent le cas dans le cours), alors la σ-algèbre $\mathcal{F}$ contient $\mathcal{P}(X(\Omega))$.*
De plus, la condition $X^{-1}(B) \in \mathcal{A}$, $\forall B \in \mathcal{F}$, se réduit à

$$\forall x \in F,\ X^{-1}(\{x\}) \in \mathcal{A}.$$

Ceci nous conduit à la notion de variables aléatoires discrètes

18.2 VARIABLES ALÉATOIRES DISCRÈTES

DÉFINITION 9
Soient $(\Omega, \mathcal{A}, \mathbb{P})$ un espace probabilisé et $(F, \mathcal{F})$ un ensemble avec une σ-algèbre.
Soit $X : \Omega \to F$ une variable aléatoire.
Si $X(\Omega)$ est un ensemble fini ou dénombrable, on dit que X est une **variable aléatoire discrète**.

Remarque 10 —
• *Une variable aléatoire discrète est une variable aléatoire qui prend un nombre fini ou dénombrable de valeurs.*
• *Pour $X : \Omega \to F$ une fonction telle que $X(\Omega)$ est fini ou dénombrable, X est une variable aléatoire si et seulement si $X^{-1}(\{x\}) \in \mathcal{A}$, pour tout $x \in X(\Omega)$.*
• *Si Ω est fini ou dénombrable, on sait alors automatiquement que toute fonction $X : \Omega \to F$ est une variable aléatoire discrète.*

Exemple 11 (**Fonction indicatrice**) —
Soit $(\Omega, \mathcal{A}, \mathbb{P})$ un espace probabilisé. Soit $A \in \mathcal{A}$.
*On définit $\mathbb{1}_A$ (ou χ_A) la **fonction indicatrice** de A, par $\mathbb{1}_A(\omega) = 1$ si $\omega \in A$ et 0 sinon.*
Alors la fonction $\mathbb{1}_A$ est une variable aléatoire discrète sur $(\Omega, \mathcal{A}, \mathbb{P})$.
Ces variables aléatoires sont les v.a. les plus simples que l'on puisse construire (avec les v.a. constantes).
Elles sont extrêmement utiles dans les calculs. (pour des sommes, produits, découpages en partition)
On a par exemple que $\mathbb{P}_{\mathbb{1}_A}(\{1\}) = \mathbb{P}(\mathbb{1}_A^{-1}(\{1\}) = \mathbb{P}(A)$.

On rappelle que quand Ω est fini ou dénombrable, une mesure de probabilité $\mathbb{P}$ sur Ω sera caractérisée par les $p_\omega = \mathbb{P}(\{\omega\})$ (par sa loi).
Cela n'est pas vrai quand Ω est infini non dénombrable. Mais, si X est une v.a. discrète sur Ω, on peut quand même déterminer la mesure $\mathbb{P}_X$ avec la probabilité de singletions.
C'est ce que nous donne le résultat suivant.

Proposition 12
Soient $(\Omega, \mathcal{A}, \mathbb{P})$ un espace probabilisé. Soit $X : \Omega \to F$ une variable aléatoire.

- Si Ω est fini ou dénombrable, alors pour tout $y \in F$, on a

$$\mathbb{P}_X(\{y\}) = \mathbb{P}(X^{-1}(\{y\}))\mathbb{P}(\{\omega \text{ t.q. } X(\omega) = y\}) = \sum_{\omega,\ X(\omega)=y} \mathbb{P}(\{\omega\}).$$

  Dans le langage probabiliste, on notera aussi $\mathbb{P}_X(\{y\}) =_{def} \mathbb{P}(X = y)$.
  Quand Ω est fini ou dénombrable, on peut calculer toutes les probabilités de la forme $\mathbb{P}(X = A)$ en utilisant la probabilité de tous les singletons $\{\omega\}$.
- Si F est dénombrable, alors la loi de la v.a. X est caractérisée par la famille des $(\mathbb{P}_X(\{y\}))_{y_i \in F}$.

Exemple 13 — *Une variable aléatoire X de loi uniforme sur $\{1, \cdots, n\}$ a pour loi la famille $(\frac{1}{n})_{1 \le k \le n}$.*

18.3 Espérance des v.a. discrètes réelles

Définition 14
Soient $(\Omega, \mathcal{A}, \mathbb{P})$ un espace probabilisé et $(F, \mathcal{F})$ un ensemble avec une σ-algèbre.
Soit $X : \Omega \to F$ une variable aléatoire.
Si F est inclus dans $\mathbb{R}$ on dit que X est une **variable aléatoire réelle**.
On pourra abréger ce nom en **v.a.r.**.

Dans la majorité des exemples que nous avons vus, les variables aléatoires étaient réelles.
Les variables aléatoires étaient aussi discrètes.
Nous allons donc nous intéresser aux v.a.r. qui sont discrètes.

Motivation : Considérons X une variable aléatoire réelle, définie sur un ensemble Ω fini ou dénombrable.
On peut en général répéter l'expérience aléatoire associée à X autant de fois que l'on veut. Pour n répétitions de l'expérience X_1, ... , X_n les valeurs successives prises par X.
Pour avoir une idée du comportement de la variable X, il est naturel de considérer leur moyenne arithmétique

$$M_n = \frac{1}{n}(X_1 + ... + X_n).$$

En regroupant suivant les différents résultats y de l'expérience, nous obtenons

$$M_n = \sum_{y \in X(\Omega)} f_n(\{y\})y,$$

où $f_n(\{y\})$ est la fréquence de réalisation du résultat $\{y\}$ au cours des n expériences, c'est-à-dire de la fréquence de réalisation de l'événement $X^{-1}(\{y\})$ (dans l'ensemble de départ Ω).
D'après le précédent raisonnement intuitif sur $\mathbb{P}(X^{-1}(\{y\}))$, vue comme une mesure de la fréquence de réalisation de cet événement, on peut supposer que $f_n(\{y\})$ converge vers $\mathbb{P}(X^{-1}(\{y\})) = \mathbb{P}_X(\{y\})$ que l'on note aussi $\mathbb{P}(X = y)$.
Et si on peut de plus intervertir la somme et la limite dans l'expression ci-dessus (par exemple vrai si X prend un nombre fini de valeurs), alors la suite $(M_n)_{n\in\mathbb{N}^*}$ converge vers

$$\sum_{y\in X(\Omega)} f_n(\{y\})y = \sum_{y\in F} f_n(\{y\})y.$$

L'espérance d'une variable aléatoire, ou moyenne, est à percevoir comme la limite de ses moyennes arithmétiques, lorsque le nombre d'expériences tend vers l'infini.

Définition 15
Soient $(\Omega, \mathcal{A}, \mathbb{P})$ un espace probabilisé et $X : \Omega \to F$ une variable aléatoire réelle discrète.
Si la somme $\sum_{y\in F} |y|\mathbb{P}(X = y)$ est finie, on dit que X est **d'espérance finie** (pour la mesure $\mathbb{P}$).
On appelle **espérance de la v.a.** X le nombre $\mathbb{E}(X) = \sum_{y\in F} y\mathbb{P}(X = y)$.

Remarque 16 —

1. *On a besoin de supposer que X est une v.a. discrète pour que la famille des $(y\mathbb{P}(X = y))_{y\in F}$ possède au plus un nombre dénombrable de termes non-nuls.*
2. *On peut remarquer que le nombre réel $\mathbb{E}(X)$ ne dépend que de la loi de X (de la famille $(\mathbb{P}_X(\{y\})_{y\in F}$.*
3. *L'hypothèse de convergence absolue de $\sum_{y\in F} y\mathbb{P}(X = y)$, permet de s'assurer que la somme est indépendante de l'ordre de sommation.*
4. *Le terme d'espérance (introduit par Pascal) fait référence aux problèmes de jeux et d'espérance de gain. (au fait d'espérer gagner de l'argent en jouant longtemps à un jeu de hasard)*
5. *Si la variable aléatoire X est d'espérance finie, alors la fonction $|X|$, qui est aussi une v.a.r., est d'espérance finie.*
   En effet, on a $\mathbb{E}(|X|) = \sum_{y\in F} |y|\mathbb{P}(X = y) < +\infty$.

Les v.a. réelles sont des fonctions à valeurs dans $\mathbb{R}$. On peut donc les additionner $(X + Y)$, les multiplier par une constante (aX), mais aussi les multiplier entre elles (XY), ou les composer par une fonction réelle $(f(X)$, pour $f : \mathbb{R} \to \mathbb{R})$.
Dans le cas où X, Y sont discrètes, toutes ces opérations définissent encore des v.a.r. discrètes. (A vérifier.)
On peut alors étudier ce qui se passe par rapport à l'espérance, et par rapport au fait d'être intégrable.

Les sommes finies ou dénombrables qui apparaissent dans l'espérance sont liées aux intégrales. La définition suivante va permettre de relier ces notions.

Définition 17
Soit $(\Omega, \mathcal{A}, \mathbb{P})$ un espace probabilisé. Soit X une v.a. réelle discrète sur Ω.
Si X est d'espérance finie, on dit aussi que X est **intégrable**.
On note $L^1(\Omega, \mathcal{A}, \mathbb{P})$ l'ensemble de toutes les v.a. réelles discrètes intégrables (sur $(\Omega, \mathcal{A}, \mathbb{P})$).

Avec ce point de vue, nous allons continuer à étudier les ensembles de v.a.r. discrètes comme des ensembles de fonctions.
Quand l'ensemble Ω et la mesure de probabilité $\mathbb{P}$ sont clairs, on utilisera parfois l'abréviation L^1 pour $L^1(\Omega, \mathcal{A}, \mathbb{P})$.
Attention, dans ce chapitre toutes les v.a. que l'on considèrera intégrables/de carré intégrable/etc seront **discrètes**.

Proposition 18
Soit $(\Omega, \mathcal{A}, \mathbb{P})$ un espace probabilisé. Alors

1. $L^1(\Omega, \mathcal{A}, \mathbb{P})$ est un $\mathbb{R}$-espace vectoriel.
2. L'espérance $\mathbb{E} : L^1(\Omega, \mathcal{A}, \mathbb{P}) \to \mathbb{R}$ est une forme linéaire. On a $\mathbb{E}(aX + bY) = a\mathbb{E}(X) + b\mathbb{E}(Y)$, $\forall X, Y \in L^1,\ \forall a, b \in \mathbb{R}$.
3. On a $X \in L^1(\Omega, \mathcal{A}, \mathbb{P})$ ssi $|X| \in L^1(\Omega, \mathcal{A}, \mathbb{P})$. De plus, on a $|\mathbb{E}(X)| \leq \mathbb{E}(|X|)$.

4. L'espérance est positive : Si $X \geq 0$ (i.e. $X(\omega) \geq 0\ \forall \omega$) et $X \in L^1$, alors on a $\mathbb{E}(X) \geq 0$.
5. Soient $X, Y \in L^1$ avec $X \leq Y$. Alors on a $\mathbb{E}(X) \leq \mathbb{E}(Y)$.
6. $L^1(\Omega, \mathcal{A}, \mathbb{P})$ contient toutes les variables aléatoires réelles bornées. (les fonctions X telles que $|X| \leq b$ pour un $b \in \mathbb{R}$).
7. Si X est une v.a.r. constante ($X = a$ pour un $a \in \mathbb{R}$), alors $\mathbb{E}(X) = a$.
8. Si Ω est fini, alors $L^1(\Omega, \mathcal{A}, \mathbb{P})$ contient toutes les v.a. réelles, ce qui est aussi égal à l'ensemble de toutes les fonctions de Ω vers $\mathbb{R}$.

Preuve — On démontre chaque point en utilisant la définition de l'espérance et la définition des v.a. discrètes.
Aucun de ces résultat n'est difficile à obtenir. □

Remarque 19 —

- *Ces propriétés font fortement penser à celles des espaces vectoriels normés (voir Analyse 4), en utilisant la fonction $X \mapsto \mathbb{E}(|X|)$.*
  Mais, en général, cette fonction n'est pas une norme.
  On montre facilement que l'on a $\mathbb{E}(|X|) = 0$ si et seulement si ($X(\omega) = 0$ ou $\mathbb{P}(\{\omega\}) = 0$, pour tout $\omega \in \Omega$), si et seulement si $\mathbb{P}(X = 0) = 1$.(A vérifier.)
  Ainsi, si la probabilité de certains singletons vaut 0, il existe des v.a. X dont l'espérance vaut 0.
  Par exemple, sur $\{1, 2, 3\}$ si on prend $\mathbb{P}$ de loi $(0, \frac{1}{2}, \frac{1}{2})$, alors la fonction indicatrice $X = \delta_1(.)$ est une fonction qui est non-nulle mais telle que $\mathbb{E}(|X|) = 0$.
- *La fonction $\mathbb{E}(|.|)$ est appelée une **semi-norme**. Elle vérifie toutes les propriétés d'une norme, sauf celle pour le cas nul.*
- *Si Ω n'est pas fini, on peut avoir des v.a.r. discrètes qui ne sont pas intégrables.*
  Pour $\Omega = \mathbb{N}$ et $\mathbb{P}$ de loi $(\frac{1}{2}, \frac{1}{4}, \frac{1}{8}, \ldots)$, la fonction $X : \omega \mapsto 2^\omega$ est bien définie et est une v.a.r. discrète (car Ω est dénombrable).
  Par contre, son espérance est infinie car $\mathbb{P}(\{n\}).X(n) = \frac{2^n}{2^{n+1}} = \frac{1}{2}$ (et la famille $(\frac{1}{2})_{n\geq 0}$ n'est pas sommable).
  Donc X n'est pas intégrable.

Exemple 20 — *Soit $(\Omega, \mathcal{A}, \mathbb{P})$ un espace probabilisé. Soit $A \in \mathcal{P}(\Omega)$. Pour $\mathbb{1}_A$ la fonction indicatrice de A, cette v.a. réelle est bornée, donc intégrable, et on a*

$$\mathbb{E}(\mathbb{1}_A) = \mathbb{P}(A).$$

Cela donne donne un lien très utile entre la probabilité d'un événement et l'espérance d'une variable aléatoire.

La définition d'espérance utilise une somme sur l'espace d'arrivée F, somme qui est dénombrable car la v.a. X est discrète.
Mais si l'ensemble Ω est fini ou dénombrable, ne peut-on pas décomposer chaque terme $\mathbb{P}(X = y)$ en une somme sur des parties de Ω, et exprimer l'espérance comme une somme sur chaque $\omega \in \Omega$?
C'est ce que nous allons démontrer.

18.3.1 Lemme de transfert, théorème de transfert

Théorème 21 (**Lemme de transfert**)
Soit $(\Omega, \mathcal{P}(\Omega), \mathbb{P})$ un espace probabilisé fini ou dénombrable. Soit $X : \Omega \to F$ une v.a. réelle sur Ω, qui est d'espérance finie.
On a alors la formule fondamentale suivante :

$$\mathbb{E}(X) = \sum_{y \in F} y\mathbb{P}(X = y) = \sum_{\omega \in \Omega} p_\omega X(\omega).$$

Preuve — Par hypothèse d'espérance finie, la famille des $y\mathbb{P}(X = y)$ est sommable. Comme Ω est dénombrable, la famille des $p_\omega X(\omega)$ est dénombrable.
On pose alors les ensembles $A_y = X^{-1}(y)$. Ces ensembles (pour $y \in X(\Omega)$) forment une partition de Ω (les autres A_y sont vides), et le théorème de sommation par paquets nous permet d'obtenir le résultat. □

Exemple 22 — *Un nombre m est choisi au hasard uniforme entre 1 et 10, et nous devons deviner ce nombre en posant des questions auxquelles il ne sera répondu que par oui ou par non.*
Calculons l'espérance du nombre N de questions nécessaires dans les cas suivants :

1. *Premier cas : la question numéro i "Est-ce que $m = i$?".*
   Avec ce choix de questions, on obtient

$$\mathbb{P}(N = k) = \mathbb{P}(\text{ le nombre } k \text{ a été choisi}) = \frac{1}{10}.$$

   Ainsi, l'espérance de N vaut :

$$\mathbb{E}(N) = \sum_{k=1}^{10} k\mathbb{P}(N = k) = \frac{10(10+1)}{2} \times \frac{1}{10} = \frac{11}{2}$$

2. *Deuxième cas : Avec chaque question, nous essayons d'éliminer à peu près la moitié des réponses possibles, avec le protocole suivant : Est-ce que $m \leq 5$? $m \leq 2$? (resp. $m \leq 7$?), $m \leq 4$? (resp. $m \leq 9$?).*
   Alors, il faut 3 questions pour trouver 1, 2, 5, 6, 7 et 10. Et il faut 4 questions pour trouver 3, 4, 8 et 9.
   L'espérance de N dans ce cas vaut donc

$$\mathbb{E}(N) = 3 \times \frac{6}{10} + 4 \times \frac{4}{10} = \frac{17}{5}$$

   L'espérance dans le second cas est strictement inférieure. L'interprétation est que la seconde stratégie va "en moyenne" permettre de trouver le nombre m en moins de questions qu'avec la première stratégie.
   L'espérance donne le nombre "moyen" de questions qu'il faudra poser pour trouver m. Elle ne dit par contre rien sur le nombre minimal ni le nombre maximal de questions que l'on peut avoir à poser pour trouver m.

Remarque 23 —
• *Dans le lemme de transfert, il est absolument nécessaire que Ω soit fini ou dénombrable.*
• *La somme des $p_\omega X(\omega)$ peut avoir un sens si Ω n'est pas dénombrable (par exemple quand $\Omega = \mathbb{R}$ et $\mathcal{A} = \mathcal{B}(\mathbb{R})$), mais elle ne sera en général pas égale à $\mathbb{E}(X)$.*
• *La raison : Quand est Ω non-dénombrable, il existe des mesures de probabilité $\mathbb{P}$ telles que $\mathbb{P}(\{\omega\}) = 0$ pour tout $\omega \in \Omega$ (la probabilité de chaque singleton est nulle).*
Mais comme la mesure doit vérifier $\mathbb{P}(\Omega) = 1$, on se retrouve avec $1 = \mathbb{P}(\Omega) \neq \sum_{\omega\in\Omega} \mathbb{P}(\{\omega\}) = 0$.
• *Dans la théorie générale des probabilités discrètes, on remplace la deuxième somme du théorème de transfert par une intégrale.*
Cependant il faut d'abord avoir construit cette intégrale, et construit beaucoup d'outils en plus, et ce n'est pas du tout l'objectif de ce cours.

Soit maintenant $f : F \to \mathbb{R}$ une fonction. Ainsi $Y = f(X)$ est encore une v.a.r. discrète. Cela permet de généraliser le lemme de transfert.

Théorème 24 (**Théorème de transfert**)
Soit $(\Omega, \mathcal{A}, \mathbb{P})$ un espace probabilisé fini ou dénombrable.
Soient $X : \Omega \to F$ une v.a. intégrable, et $f : F \to \mathbb{R}$.
Si la v.a. discrète $f(X)$ et intégrable, on a alors

$$\mathbb{E}(f(X)) = \sum_{x_i \in F} f(x_i)\mathbb{P}(X = x_i) = \sum_{\omega\in\Omega} f(X(\omega))p_\omega.$$

Preuve — Ceci est encore une conséquence du théorème de sommation par paquets. □

Définition 25
Soit $(\Omega, \mathcal{A}, \mathbb{P})$ un espace probabilisé.
On définit l'ensemble $L^2(\Omega, \mathcal{A}, \mathbb{P})$ comme l'ensemble des v.a. réelles discrètes X telles que X^2 est intégrable.
Si $X \in L^2(\Omega, \mathcal{A}, \mathbb{P})$, on dit que X est **de carré sommable**.

Si l'espace probabilisé $(\Omega, \mathcal{A}, \mathbb{P})$ est clair, on pourra noter L^2 à la place de $L^2(\Omega, \mathcal{A}, \mathbb{P})$.
Attention à bien voir que l'ensemble $L^2(\Omega, \mathcal{A}, \mathbb{P})$ est un ensemble de v.a. **réelles** et **discrètes**.

Proposition 26
Soit $(\Omega, \mathcal{A}, \mathbb{P})$ un espace probabilisé.
L'ensemble $L^2(\Omega, \mathcal{A}, \mathbb{P})$ est un sous-espace vectoriel de $L^1(\Omega, \mathcal{A}, \mathbb{P})$.
Pour tout $X \in L^2(\Omega, \mathcal{A}, \mathbb{P})$, on a

$$|\mathbb{E}(X)| \leq \mathbb{E}(|X|) \leq \sqrt{\mathbb{E}(X^2)}.$$

Preuve — Soient X et Y deux variables aléatoires réelles et $a \in \mathbb{R}$. Si X et Y sont dans L^2, l'inégalité

$$(aX+Y)^2 \leq 2a^2X^2 + 2Y^2$$

montre que que $aX + Y \in L^2$: L^2 est bien un espace vectoriel.
L'inclusion $L^2 \subset L^1$ découle de $|X| \leq 1 + X^2$.
La première inégalité a déjà été vue.
Pour la seconde, on peut supposer X est positive.
Soit alors $a = \mathbb{E}(X)$ et $Y = X - a$. Par linéarité

$$\mathbb{E}(Y^2) = \mathbb{E}(X^2) - 2a\mathbb{E}(X) + a^2 = E(X^2) - a^2,$$

et $\mathbb{E}(Y^2) \geq 0$. Donc $a^2 \leq \mathbb{E}(X^2)$, ce qui est le résultat cherché. □

Définition 27
Soit $(\Omega, \mathcal{A}, \mathbb{P})$ un espace probabilisé. Soit $X \in L^2(\Omega, \mathcal{A}, \mathbb{P})$.
On définit la **variance** de X par :

$$\mathrm{Var}\,(X) = \mathbb{E}((X - \mathbb{E}(X))^2) = \sum_{x_i \in F} (x_i - \mathbb{E}(X))^2 p_i^X.$$

On note aussi $\sigma_X = \sqrt{Var(X)}$, **l'écart-type** de X.

Remarque 28 —
• *En développant le carré* $(X - \mathbb{E}(X))^2$ *on obtient*

$$Var(X) = \sigma_X^2 == \mathbb{E}(X^2) - \mathbb{E}(X)^2,$$

ce qui permet de montrer que ce nombre réel est bien défini quand X *est de carré intégrable.*
• *Par définition, on constate que* $Var(X) \geq 0$. *Cela montre donc que* $\mathbb{E}(X^2) - \mathbb{E}(X)^2 \geq 0$.
• *L'écart-type est une grandeur qui mesure une distance de la v.a* X *par rapport à son espérance* $\mathbb{E}(X)$. *(penser à* $d(x,y) = \sqrt{\langle x-y, x-y\rangle}$ *pour les espaces euclidiens)*
Elle mesure, dans un sens, à quel point la v.a. X *s'écarte en moyenne de* $\mathbb{E}(X)$.

Exemple 29 (Un jeu de loto) —
Le joueur coche 6 *numéros sur une grille qui en comporte* 49. *Les* 6 *numéros gagnants sont déterminés par tirage au sort. Soit* n *le nombre de numéros gagnants d'une grille.*
Pour une mise de 2 *Euros, on reçoit le gain* $G = g(n)$ *suivant :*

n numéros gagnants	*gain* $g(n)$	*probabilité*
6	2 132 885 E	$7,2\ 10^{-8}$
5	3 575 E	$7,8\ 10^{-5}$
4	94 E	$9,7\ 10^{-4}$
3	11 E	$7,8\ 10^{-2}$

Le gain moyen est donc de

$$\begin{aligned}\mathbb{E}(G) &= \sum_n g(n)\mathbb{P}(N=n)\\ &= 11 \times 7.8\ 10^{-2} + 94 \times 9.7\ 10^{-4} + 3575 \times 1.8\ 10^{-5} + 2132885 \times 7.2\ 10^{-8}\\ &= 1,16\ E.\end{aligned}$$

Ainsi le bénéfice moyen du joueur, qui vaut $\mathbb{E}(G) - 2 = -0.84$, *est négatif, et le jeu est défavorable au joueur. On peut calculer aussi que l'écart-type de ce jeu vaut* 572. *La grande valeur de l'écart-type vient du fait que ce jeu peut rapporter énormément d'argent (même si cela est très très rare), alors qu'en moyenne chaque joueur perd un peu d'argent à chaque partie.*
Beaucoup de jeux de hasard sont basés sur ce ptincipe : gros gains avec très faibles probabilités (grande variance), et gains moyens légèrement négatifs (espérance légèrement négative).

Remarque 30 — *Pour* X *une v.a.r. discrète, on sait que* X^2 *intégrable implique* X *intégrable.*
La réciproque est par contre fausse en général.
Contre-exemple : *On prend* $\Omega = \mathbb{N}$ *et* $\mathbb{P}$ *la mesure dont la loi est* $(\frac{1}{2^{n+1}})_{n\geq 0}$. *On pose* $X : \mathbb{N} \to \mathbb{R}$ *avec* $X(n) = \sqrt{2}^{n+1}$.

Alors X est une v.a. discrète, positive, et $\mathbb{E}(X) = \sum_{n\geq 0} \frac{1}{\sqrt{2}^{n+1}} < +\infty$. (la famille $(X(n)\mathbb{P}(X=n))_n$ est sommable, et on applique la formule de transfert)
Par contre, on a $X^2(n) = 2^{n+1}$, ce qui donne $\mathbb{E}(X^2) = \sum_{n\geq 0} 1 = +\infty$. (la famille $(X^2(n)\mathbb{P}(X=n))_n$ n'est pas sommable)

On peut généraliser la définition d'intégrabilité pour toutes les puissances de X.

Définition 31
Soient $(\Omega, \mathcal{A}, \mathbb{P})$ un espace probabilisé, X une v.a.r. discrète, et $k \geq 1$.
Si la v.a. X^k est intégrable (si $\mathbb{E}(|X^k|) < +\infty$), on dit que X possède un **moment d'ordre** k.
Le moment d'ordre k de X est la quantité $\mathbb{E}(X^k)$.

18.3.2 Fonction génératrice d'une v.a. à valeurs dans $\mathbb{N}$

On va montrer que pour X une v.a. discrète, sa loi $\mathbb{P}_X$ peut être caractérisée par une fonction, appelée fonction génératrice, définie sur $[0, 1]$ et indéfiniment dérivable sur $[0, 1[$.
Comme X est une v.a. réelle discrète, à bijection près on peut considérer que X est à valeurs dans $\mathbb{N}$. Soit une variable aléatoire X à valeurs dans $\mathbb{N}$, dont la loi est caractérisée par les nombres $p_n = p_n^X = \mathbb{P}(X = n)$.

Définition 32
Soit $(\Omega, \mathcal{A}, \mathbb{P})$ un espace probabilisé. Soit X une v.a.r. à valeurs dans $\mathbb{N}$.
On appelle **fonction génératrice** de X, la fonction $G_X : [0,1] \to \mathbb{R}_+$ définie par

$$G_X(s) = \sum_{n=0}^{\infty} s^n \mathbb{P}(X=n) = \sum_{n=0}^{\infty} s^n p_n, \ \forall s \in [0,1].$$

Remarque 33 —

1. *Cette quantité est la somme d'une série entière à termes positifs, dont tous les termes sont majorés par 1. Son rayon de convergence est donc d'au moins 1.*
   On sait aussi que pour $s = 1$ la série est convergente, de somme 1. La fonction G_X est donc bien définie sur $[0,1]$.
2. *Pour tout $s \in [0,1]$, la fonction $s^X : \omega \in \Omega \mapsto s^{X(\omega)} \in \mathbb{R}_+$ est bien définie et est une v.a. discrète (c'est la composée d'une v.a. discrète par une fonction). La fonction génératrice G_X s'écrit alors*

$$G_X(s) = \mathbb{E}(s^X).$$

3. *La fonction génératrice ne dépend que de la famille de probabilités $(\mathbb{P}(X=n))_{n\geq 0}$, c'est-à-dire de la loi de X.*

Proposition 34
Soit $(\Omega, \mathcal{A}, \mathbb{P})$ un espace probabilisé. Soit X une v.a.r. à valeurs dans $\mathbb{N}$.
La fonction génératrice G_X est continue sur $[0,1]$ et infiniment dérivable (de classe C^∞) sur $[0,1[$.
De plus, on peut retrouver la loi de X à partir de G_X.

Preuve — La fonction G_X est la somme d'une série entière à termes positifs qui converge normalement sur $[0,1]$, puisque $\sum_{n=0} \mathbb{P}(X=n) = 1$.
Les propriétés de continuité et de dérivabilité en découlent. (voir Analyse 4)
De plus, on sait que

$$\mathbb{P}(X=n) = p_n = \frac{G_X^{(n)}(0)}{n!}.$$

Ainsi, on peut retrouver la famille $(\mathbb{P}(X=n))_n$ (la loi de X) à partir de la fonction G_X.
On dit aussi que la fonction G_X **caractérise la loi** de X. □

La fonction génératrice G_X ne donne pas seulement toutes les informations sur la loi de X, elle permet aussi de dire si X est intégrable, et de calculer son espérance.

Proposition 35
Soit $(\Omega, \mathcal{A}, \mathbb{P})$ un espace probabilisé. Soit X une v.a.r. à valeurs dans $\mathbb{N}$.
La v.a. X est intégrable si et seulement si G_X est dérivable à gache en $s = 1$.
Dans ce cas, on a $\mathbb{E}(X) = G'_X(1)$.

Preuve — La fonction G_X est dérivable sur $[0,1[$ et

$$G'_X(s) = \sum_{n\geq 0} np_n s^{n-1}.$$

Si la variable aléatoire X est intégrable alors $\mathbb{E}(X) = \sum_{n\geq 0} np_n$ est convergente et la série entière définissant G'_X est normalement convergente sur $[0,1]$, donc G'_X admet une limite en 1^-.
0n en déduit que G_X est de classe $\mathcal{C}^1$ sur $[0,1]$ et que

$$G'_X(1) = \mathbb{E}(X).$$

Si X n'est pas intégrable, alors $\left(\sum_{k=0}^{n} kp_k\right)_{n\geq 0}$ tend vers $+\infty$.
Supposons que $G'_X(s)$ admet une limite A en 1^-. La série étant à termes positifs, la fonction est croissante en la variable s et

$$\forall n \in \mathbb{N},\ \sum_{k=0}^{n} kp_k s^{k-1} \leq \sum_{k=0}^{+\infty} kp_k s^{k-1} \leq A.$$

Par passage à la limite en 1, on obtient $\sum_{k=0}^{n} kp_k \leq A$, puis $\sum_{k=0}^{+\infty} kp_k \leq A$, ce qui est absurde. Donc $G'_X(s)$ n'admet pas de limite 1^- et comme la fonction est croissante,

$$\lim_{s\to 1^-} G'_X(s) = +\infty.$$

On en déduit que G_X n'est pas dérivable en 1. □

Plus généralement, la même démonstration prouve que

Proposition 36
Soient $(\Omega, \mathcal{A}, \mathbb{P})$ un espace probabilisé, X une v.a.r. à valeurs dans $\mathbb{N}$,et $p \geq 1$.
La v.a. $X(X-1)...(X-p)$ est intégrable si et seulement si G_X est $p+1$ fois dérivable à gauche en $s=1$.
Dans ce cas, on a $\mathbb{E}(X(X-1)...(X-p)) = G_X^{(p+1)}(1)$.
En particulier on a $\mathbb{E}(X(X-1)) = G''_X(1)$, d'où $\mathrm{Var}(X) = G''_X(1) - (G'_X(1))^2 + G'_X(1)$.

Preuve — On procède par récurrence sur p, le cas $p=0$ étant déjà traité.

$$\forall s \in [0,1[,\ G_X^{(p+1)}(s) = \sum_{n\geq 0} n(n-1)\cdots(n-p)\, p_n s^{n-p-1}$$

et

$$\mathbb{E}(X(X-1)...(X-p)) = \sum_{n\geq 0} n(n-1)\cdots(n-p)\, p_n.$$

Un raisonnement similaire au cas $p=0$, montre que $G_X^{(p+1)}$ admet une limite en 1^- ssi $\mathbb{E}(X(X-1)...(X-p))$ existe et alors on a l'égalité attendue. □

Remarque 37 —
• *On peut en fait montrer avec la dernière proposition que X^p est intégrable ssi G_X est $(p+1)$ fois dérivable à gauche en $s=1$. (On utilise le fait que X^p est une combinaison linéaire des $X(X-1)\dots(X-k)$, pour $0 \leq k \leq p$.)*
Ainsi, en étudiant la fonction génératrice G_X, on peut dire si la v.a. X possède des moments d'ordre p (si $X^p \in L^1$).
• *Si X a un moment d'ordre p, comme X est à valeurs dans $\mathbb{N}$ on a $\mathbb{E}(X^p) = \sum_{n\geq 0} n^p \mathbb{P}(X=n)$.*
On peut alors calculer la somme de cette série à l'aide des dérivées de la fonction G_X.
Même lorsque $k = 1, 2$ (pour calculer l'espérance ou la variance), il peut être beaucoup plus rapide et simple d'utiliser les dérivées de la fonction génératrice plutôt qu'un calcul direct.

18.4 Variables aléatoires discrètes usuelles

Dans cette section, nous présentons des v.a.r. discrètes usuelles. Ces v.a. sont à chaque fois à valeurs dans $\mathbb{N}$. $(X : \Omega \to \mathbb{N})$
Pour une v.a. discrète, on l'a dit, la v.a. X est totalement déterminée par sa mesure de probabilité associée $\mathbb{P}_X$. Ainsi, pour décrire une v.a. discrète, il n'est pas nécessaire de vraiment décrire l'espace probabilité $(\Omega, \mathcal{A}, \mathbb{P})$.

Définition 38
Soit $(\Omega, \mathcal{A}, \mathbb{P})$ un espace probabilisé. Soit $X : \Omega \to \mathbb{N}$ une v.a.
Soit $p \in [0,1]$.
Si on a $\mathbb{P}(X=1) = p$ et $\mathbb{P}(X=0) = 1-p$, on dit que X est une **variable aléatoire de Bernouilli**, de paramètre p.

PROPOSITION 39
Soient $p \in [0,1]$ et X une v.a. de Bernouilli de paramètre p.
Alors, on a

$$\begin{aligned} \mathbb{E}(X) &= p \\ \mathrm{Var}\,(X) &= p(1-p) \\ G_X(s) &= (1-p+ps). \end{aligned}$$

Preuve — On calcule

$$\mathbb{E}(X) = p \times 1 + 0 \times (1-p) = p$$

et

$$\mathrm{Var}\,(X) = p - p^2 = p(1-p).$$

Enfin, on a $G_X(s) = (1-p)s^0 + ps^1$. □

REMARQUE 40 — *Quand on modélise le jeu du pile ou pace avec une pièce, en supposant que face (1) apparaît avec la probabilité p et pile (0) avec la probabilité $1-p$, on obtient une v.a. de Bernoulli de paramètre p.*

DÉFINITION 41
Soit $(\Omega, \mathcal{A}, \mathbb{P})$ un espace probabilisé. Soit $X : \Omega \to \mathbb{N}$ une v.a.
Soient $p \in [0,1]$ et $n \geq 1$.
Si on a $X(\Omega) = \{1, \ldots, n\}$ avec $\mathbb{P}(X = k) = \binom{n}{k} p^k (1-p)^{n-k}$, pour tout $0 \leq k \leq n$, on dit que X est une **variable aléatoire binomiale**, de paramètres n et p.
On note sa loi de probabilités $\mathbb{P}_X = B(n,p)$.

REMARQUE 42 —

1. *On retrouve cette v.a. (et sa mesure de probas associée) dans le modèle des urnes : on tire n boules parmi des boules de 2 couleurs (blanc ou noir), sachant que la probabilité de choisir une boule noire est p. Si X donne le nombre de boules noires, alors X est une v.a. binomiale.*
2. *On peut aussi considérer n lancers de Pile ou Face, sachant que la probabilité d'obtenir Face est p, et X la v.a. compte le nombre de Faces au bout de n lancers.*
3. *Pour X une v.a. binomiale, on dit aussi que sa loi de probabilité est une **loi binomiale**.*
4. *La loi de probas $B(1,p)$ est égale à la loi de probas de Bernoulli de paramètre p.*

PROPOSITION 43
Soient $n \geq 1$, $p \in [0,1]$, et X une variable binomiale de loi $B(n,p)$.
Alors, on a

$$\begin{aligned} G_X(s) &= (1-p+ps)^n \\ \mathbb{E}(X) &= np \\ \mathrm{Var}\,(X) &= np(1-p). \end{aligned}$$

Preuve — Ici, on va calculer l'espérance et la variance de X grâce à la fonction génératrice G_X.
On calcule :

$$G_X(s) = \sum_{k=0}^{n} \binom{n}{k} p^k s^k (1-p)^{n-k} = (1-p+ps)^n.$$

En dérivant G_X et en calculant $G'_X(1)$, sachant que $k\binom{n}{k} = n\binom{n-1}{k-1}$ pour $k \neq 0$, on obtient

$$\mathbb{E}(X) = \sum_{k=0}^{n} k \binom{n}{k} p^k (1-p)^{n-k} = np(1-p+p\times 1)^{n-1} = np$$

De même,

$$G''_X(1) = n(n-1)p^2(1-p+p\times 1)^{n-2} = n(n-1)p^2.$$

et donc

$$\begin{aligned} \mathrm{Var}\,(X) &= \mathbb{E}(X(X-1)) + \mathbb{E}(X) - \mathbb{E}(X)^2 \\ &= n(n-1)p^2 + np - (np)^2 \\ &= np(1-p) \end{aligned}$$

□

EXEMPLE 44 — *Aux jeux olympiques de Vancouver (2010), 86 médailles d'or ont été mises en jeu.*
Nous faisons l'hypothèse que la probabilité qu'un pays remporte une médaille est proportionnelle à sa population.
Soit X le nombre de médailles prévues pour la France. X va suivre une loi binomiale $B(86,p)$, où

$$p = \frac{\textit{population France}}{\textit{population monde}} = \frac{60 \times 10^6}{6000 \times 10^6} = 0,01.$$

Ainsi l'espérance de X sera égale à $86 \times 0,01 = 0,86$.
Cherchons la probabilité pour que le nombre de médailles soit inférieur à 3. Elle vaut

$$\mathbb{P}(X \leq 3) = \mathbb{P}(X = 0) + \mathbb{P}(X = 1) + \mathbb{P}(X = 2) + \mathbb{P}(X = 3),$$

avec pour tout $k \in \{0, \cdots, 86\}$

$$\mathbb{P}(X = k) = \binom{86}{k} (0,01)^k (0,99)^{86-k}.$$

Tous calculs faits, nous trouvons

$$\mathbb{P}(X \leq 3) = 0,9889.$$

La France a en fait remporté 2 médailles d'or (la France en a obtenu 4 sur 99 en 2015 et 5 sur 103 en 2018).

Dans un jeu de Pile ou Face (on lance autant de fois que l'on veut, avec $\mathbb{P}(Face) = p$), on considère la variable X qui donne le numéro du premier lancer donnant Face (les précédents étant Pile).
Comme on considère les lancers indépendants, on obtient

$$\mathbb{P}(X = k) = (1-p)^{k-1} p.$$

Définition 45
Soit $(\Omega, \mathcal{A}, \mathbb{P})$ un espace probabilisé. Soit $X : \Omega \to \mathbb{N}$ une v.a.
Soit $p \in]0, 1[$.
Si on a $\mathbb{P}(X = k) = p(1-p)^{k-1}$ pour tout $k \geq 1$, on dit que X est une **variable aléatoire géométrique** de paramètre p.

Proposition 46
Soient $p \in]0, 1[$ et X une v.a. géométrique de paramètre p.
Alors, on a

$$\begin{aligned} \mathbb{E}(X) &= \frac{1}{p} \\ \mathrm{Var}\,(X) &= \frac{1-p}{p^2} \\ G_X(s) &= \frac{ps}{1-(1-p)s} \end{aligned}$$

Preuve — Le critère de D'Alembert montre que

$$\mathbb{E}(X) = \sum_{k=1}^{\infty} kp(1-p)^{k-1} < +\infty.$$

A nouveau, calculons la fonction génératrice, puis déterminons espérance et variance avec G_X.

$$G_X(s) = \sum_{k \geq 1} p(1-p)^{k-1} s^k = \frac{ps}{1-(1-p)s}.$$

Comme $0 < p < 1$, il n'y a pas de problèmes pour les quotients. On obtient alors directement

$$G'_X(s) = \frac{p}{(1-(1-p)s)^2}, \; G''_X(s) = \frac{2p(1-p)}{(1-(1-p)s)^3},$$

et donc $\mathbb{E}(X) = G'_X(1) = \dfrac{1}{p}$.
On a aussi $G''_X(1) = \dfrac{2(1-p)}{p^2}$ et donc

$$\mathrm{Var}\,(X) = \frac{2(1-p)}{p^2} + \frac{1}{p} - \frac{1}{p^2} = \frac{1-p}{p^2}.$$

□

Remarque 47 — *En d'autres termes, si l'on regarde une expérience de Bernouilli de paramètre p, et qu'on la répète de façon indépendante jusqu'à obtenir un succès (un Pile par exemple), alors le nombre moyen de répétitions à faire est $\frac{1}{p}$.*
faut donc, en moyenne, s'attendre à lancer 6 fois un dé équilibré avant d'obtenir le premier 1.
On retrouve ici un résultat intuitif qui dit que pour A un événement de probabilité p ($0 < p < 1$), il faudra faire en moyenne $\frac{1}{p}$ tentatives pour que l'événement A se réalise. (en moyenne 36 lancers pour obtenir un double-6 avec deux dés équilibrés, en moyenne 52 tirages avec remise pour piocher l'as de coeur, etc)

Définition 48
Soit $(\Omega, \mathcal{A}, \mathbb{P})$ un espace probabilisé. Soit $X : \Omega \to \mathbb{N}$ une v.a.
Soit $\theta > 0$.
Si on a $\mathbb{P}(X = k) = e^{-\theta}\dfrac{\theta^k}{k!}$ pour tout $k \in \mathbb{N}$, on dit que X est une **variable aléatoire de Poisson** de paramètre $\theta > 0$.
On dit aussi que sa loi de probas $\mathbb{P}_X$ est une **loi de Poisson** de paramètre $\theta > 0$.

Proposition 49
Soient $\theta > 0$ et X une v.a. de Poisson de paramètre θ.
Alors, on a

$$\begin{aligned} \mathbb{E}(X) &= \theta \\ \mathrm{Var}\,(X) &= \theta \\ G_X(s) &= e^{\theta(s-1)}. \end{aligned}$$

Preuve — On peut calculer facilement l'espérance de X ainsi que la fonction génératrice G_X.

$$\mathbb{E}(X) = e^{-\theta}\sum_{k=0}^{\infty} k\frac{\theta^k}{k!} = \theta,$$

$$G_X(s) = e^{-\theta}\sum_{k=0}^{\infty} \frac{\theta^k s^k}{k!} = e^{\theta(s-1)}.$$

On calcule facilement $G''_X(1) = \theta^2$ et donc on en déduit la variance de X vaut

$$\mathrm{Var}\,(X) = \theta^2 + \theta - \theta^2 = \theta.$$

□

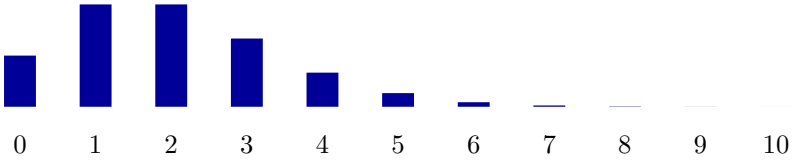
Remarque 50 — *La loi de Poisson est une loi de probabilité discrète. Elle décrit le nombre d'évènements se produisant dans un laps de temps fixé, dans le cas où ces évènements se produisent avec une fréquence moyenne connue, et indépendamment du temps écoulé depuis l'évènement précédent.*

Exemple 51 — *Une société constate en moyenne trois accidents du travail par an. L'effectif total est relativement élevé, aussi considère-t-on que le nombre d'accidents suit une loi de Poisson. Quelle est la probabilité que plus de quatre accidents surviennent dans l'année ?*
On comprend ici que $\theta = 3$.
On calcule alors :

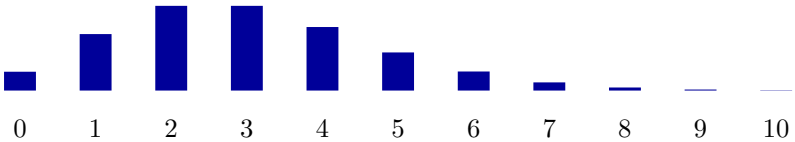
$$\mathbb{P}(X \geq 4) = 1 - e^{-3} \times \sum_{k=0}^{3} \frac{3^k}{k!} \simeq 0.26$$

Diagrammes

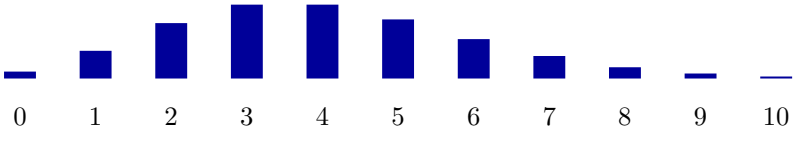
Loi de poisson de paramètre $\theta = 2$: $\mathbb{P}(X = 2) = 0,27$



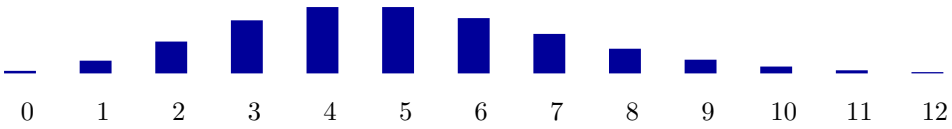
Loi de poisson de paramètre $\theta = 3$: $\mathbb{P}(X = 3) = 0,22$



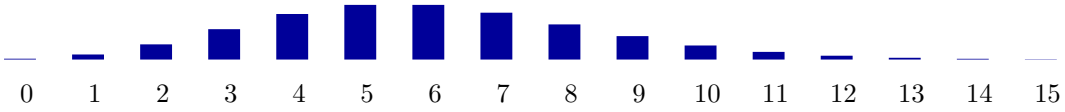
Loi de poisson de paramètre $\theta = 4$: $\mathbb{P}(X = 4) = 0,20$



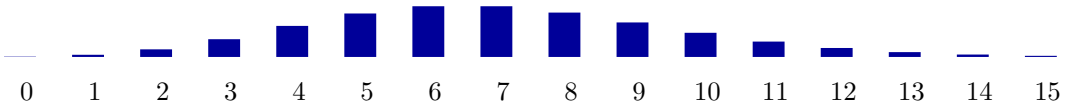
Loi de poisson de paramètre $\theta = 5$: $\mathbb{P}(X = 5) = 0,18$



Loi de poisson de paramètre $\theta = 6$: $\mathbb{P}(X = 6) = 0,16$



Loi de poisson de paramètre $\theta = 7$: $\mathbb{P}(X = 7) = 0,15$



18.5 Variables aléatoires indépendantes

Nous avons vu la notion de probabilités indépendantes. Cette notion, totalement dépendante de la mesure de probas $\mathbb{P}$, donne des informations très utiles sur la réalisation d'événements A et B.
Nous allons généraliser cette notion aux variables aléatoires.

Définition 52
Soit $(\Omega, \mathcal{A}, \mathbb{P})$ un espace probabilisé. Soit $X : \Omega \to F$ et $Y : \Omega \to G$ deux v.a. discrètes.
On dit X et Y sont des **variables aléatoires indépendantes** si on a

$$\mathbb{P}(X = x, Y = y) = \mathbb{P}(X = x)\,\mathbb{P}(Y = y),\ \forall (x, y) \in F \times G.$$

Proposition 53
Soit $(\Omega, \mathcal{A}, \mathbb{P})$ un espace probabilisé. Soit $X : \Omega \to F$ et $Y : \Omega \to G$ deux v.a. discrètes.
Alors X et Y sont indépendantes si et seulement si on a

$$\mathbb{P}(X \in A, Y \in B) = \mathbb{P}(X \in A)\,\mathbb{P}(Y \in B),\ \forall A \subset F,\ \forall B \subset G.$$

Preuve — L'implication $\Leftarrow$ s'obtient en prenant $A = \{x\}$ et $B = \{y\}$.
Pour l'implication réciproque, on écrit que $A \times B$ est l'union disjointe des singletons $\{x, y\}$, $x \in A$, $y \in B$. De plus, comme X et Y sont des v.a. discrètes, il n'y a qu'un nombre fini ou dénombrables d'éléments de A (resp. B) qui sont atteints par X (resp. Y). et

$$\begin{aligned}
\mathbb{P}(X \in A, Y \in B) &= \sum_{(x,y)\in A\times B} \mathbb{P}(X = x, Y = y) \\
&= \sum_{(x,y)\in A\times B} \mathbb{P}(X = x)\mathbb{P}(Y = y) \\
&= \left(\sum_{x\in A} \mathbb{P}(X = x)\right) \times \left(\sum_{y\in B} \mathbb{P}(Y = y)\right) \\
&= \mathbb{P}(A) \times \mathbb{P}(B)
\end{aligned}$$

Notons que les regroupements de somme sont licites puisque nous sommons des réels positifs et que ces sommes sont majorées par 1. □

Remarque 54 — *Pour X, Y deux v.a.d. indépendantes, si posant $Z = (X, Y)$, alors Z est une v.a. discrète et sa loi de probas est donnée par la famille $(\mathbb{P}(X = x)\mathbb{P}(Y = y))_{(x,y)}$.*

On peut aussi généraliser la notion d'indépendances à n variables aléatoires :

Définition 55
Soit $(\Omega, \mathcal{A}, \mathbb{P})$ un espace probabilisé. Soient $X_i : \Omega \to F_i$, $1 \leq i \leq n$, n v.a. discrètes.
On dit que $X_1, \ldots, X_n$ sont des **variables aléatoires indépendantes** si

$$\forall (x_1, \cdots, x_n) \in F_1 \times \cdots \times F_n,\ \mathbb{P}(X_1 = x_1, \cdots, X_n = x_n) = \mathbb{P}(X_1 = x_1) \cdots \mathbb{P}(X_n = x_n).$$

Exemple 56 — *Considérons X_1, ..., X_n n variables aléatoires de Bernoulli de paramètre $p \in]0,1[$ qui sont indépendantes.*
Soit $x_i \in \{0,1\}$, pour $i \in \{1,...,n\}$.
La probabilité que la suite $(X_1, \cdots, X_n)$ soit égale à $(x_1, \cdots, x_n)$, vaut alors

$$\mathbb{P}(X_i = x_i,\ 1 \leq i \leq n) = \prod_{i=1}^{n} p^{x_i}(1-p)^{1-x_i} = p^{\sum_i x_i}(1-p)^{n-\sum_i x_i}.$$

On retouve le modèle du tirage dans une urne sans remise (loi binomiale).

Proposition 57
Soit $(\Omega, \mathcal{A}, \mathbb{P})$ un espace probabilisé.
Soient X, Y deux v.a. discrètes indépendantes. Soient $f : F \to \mathbb{R}$ et $g : G \to \mathbb{R}$, telles que les v.a. $f(X)$ et $g(X)$ sont intégrables.
Alors le produit $f(X)g(Y)$ est aussi intégrable et vérifie

$$\mathbb{E}\big(f(X)g(Y)\big) = \mathbb{E}(f(X))\mathbb{E}(g(Y)).$$

Preuve — On écrit

$$\begin{aligned}\sum_{(x,y)\in F\times G} |f(x)g(y)|\mathbb{P}((X,Y)=(x,y)) &= \sum_{x\in F, y\in G} |f(x_i)|\,|g(y)|\,\mathbb{P}(X=x)\mathbb{P}(Y=y)\\ &= \left(\sum_{x\in F} |f(x)|\mathbb{P}(X=x)\right) \times \left(\sum_{y\in G} |g(y)|\mathbb{P}(Y=y)\right)\end{aligned}$$

Le terme de droite étant fini par hypothèse, on en déduit que la variable aléatoire $f(X)g(Y)$ est intégrable.
Les égalités sont alors valables sans les valeurs absolues, ce qui montre exactement que $f(X)$ et $g(Y)$ sont indépendantes. □

Corollaire 58
Soit $(\Omega, \mathcal{A}, \mathbb{P})$ un espace probabilisé. Soient X, Y deux v.a.d. réelles et indépendantes.
Alors, on a $\mathbb{E}(XY) = \mathbb{E}(X)\mathbb{E}(Y)$ et $\sigma^2_{X+Y} = \sigma^2_X + \sigma^2_Y$.

Preuve — On utilise la proposition précédente. □

Corollaire 59
Soit $(\Omega, \mathcal{A}, \mathbb{P})$ un espace probabilisé. Soient $X_i : \Omega \to F_i$, $1 \leq i \leq n$, n v.a. discrètes. Soient $f_1 : F_1 \times \cdots \times F_k \to G_1$ et $f_2 : F_{k+1} \times \cdots \times E_n \to G_2$ des fonctions.
Si les v.a. X_1, ...,X_n sont indépendantes, alors les v.a. $f_1(X_1, \cdots, X_k)$ et $f_2(X_{k+1}, \cdots, X_n)$ sont indépendantes.

Preuve — On utilise la proposition précédente. □

Corollaire 60
Soit $(\Omega, \mathcal{A}, \mathbb{P})$ un espace probabilisé. Soient $X_1, \ldots, X_n$ des v.a. réelles discrètes.
Si les v.a. $X_1, \ldots, X_n$ sont indépendantes et ont la même loi de probas, alors on a $\sigma_{X_1+\cdots+X_n} = \sqrt{n}\sigma_{X_1}$.

Preuve — On applique un corollaire précédent. □

Attention, la notion d'indépendance pour les variables aléatoires a quelques bonnes propriétés, mais certaines manipulations ne préservent pas cette indépendance.
L'intérêt d'avoir des v.a. indépendantes est d'étudier des fonctions en ces v.a.

Proposition 61
Soit $(\Omega, \mathcal{A}, \mathbb{P})$ un espace probabilisé. Soient X, Y deux v.a. à valeurs dans $\mathbb{Z}$.
On pose $Z = X + Y$. Alors, on a

$$\mathbb{P}(Z=i) = \sum_{j\in\mathbb{Z}} \mathbb{P}((X,Y)=(j,i-j)) = \sum_{j\in\mathbb{Z}} \mathbb{P}((X,Y)=(i-j,j)).$$

En particulier, si X et Y sont indépendantes, on a

$$\mathbb{P}(Z=i) = \sum_{j\in\mathbb{Z}} \mathbb{P}(X=j)\mathbb{P}(Y=i-j) = \sum_{j\in\mathbb{Z}} \mathbb{P}(X=i-j)\mathbb{P}(Y=j).$$

REMARQUE 62 — *La loi de probabilité définie par* $\mathbb{P}_{X+Y}(\{i\}) = \sum_{j\in\mathbb{Z}} \mathbb{P}(X=i-j)\mathbb{P}(Y=j)$ *s'appelle le* ***produit de convolution*** *des deux mesures de probabilité* $\mathbb{P}_X$ *et* $\mathbb{P}_Y$*. On le note* $\mathbb{P}_X * \mathbb{P}_Y$*.*
Nous n'allons pas plus étudier la convolution dans ce cours.

EXEMPLE 63 — *Soit* $(\Omega, \mathcal{A}, \mathbb{P})$ *un espace probabilisé. Soit* $n \geq 1$ *un entier et soient* $X, Y : \Omega \to \{0, \ldots, n\}$ *deux v.a. qui sont de loi de probas uniforme, et qui sont indépendantes.*
On étudie la loi de la v.a. $Z = X + Y$*. On a :*

$$\mathbb{P}(Z=k) = \sum_{j\in\mathbb{N}} \mathbb{P}(X=j)\mathbb{P}(Y=k-j).$$

Or

$$\mathbb{P}(X=j) = \mathbb{P}(Y=j) = \begin{cases} \dfrac{1}{n+1} \text{ si } j \in \{0,1,\cdots,n\} \\ \\ 0 \text{ sinon.} \end{cases}$$

On doit donc séparer deux cas :

1. *si* $0 \leq k \leq n$*, alors*

$$\mathbb{P}(Z=k) = \sum_{j=0}^{k} \left(\frac{1}{n+1}\right)^2 = \frac{k+1}{(n+1)^2}$$

2. *si* $n \leq k \leq 2n$*, alors*

$$\mathbb{P}(Z=k) = \sum_{j=k-n}^{n} \left(\frac{1}{n+1}\right)^2 = \frac{2n-k+1}{(n+1)^2}$$

On vérifie que $\mathbb{P}(Z=2n-k) = \mathbb{P}(Z=k)$ *pour* $0 \leq k \leq n$*.*
Le diagramme en bâton de la loi de probas Z *est en fait un triangle. On l'appelle loi triangulaire.*

EXEMPLE 64 — *Soit* $(\Omega, \mathcal{A}, \mathbb{P})$ *un espace probabilisé. Soient* $\lambda_1, \lambda_2 > 0$ *, et* $X, Y : \Omega \to \mathbb{N}$ *deux v.a. de Poisson de paramètres* λ_1 *et* λ_2*, et qui sont indépendantes.*
On étudie la loi de la v.a. $Z = X + Y$*. On a :*

$$\begin{aligned} \mathbb{P}(Z=i) &= \sum_{j\in\mathbb{N}} \mathbb{P}(X=j)\,\mathbb{P}(Y=i-j) \\ &= \sum_{j=0}^{i} e^{-(\lambda_1+\lambda_2)} \frac{\lambda_1^j \lambda_2^{i-j}}{j!(i-j)!} \\ &= \frac{e^{-(\lambda_1+\lambda_2)}}{i!} \sum_{j=0}^{i} \binom{i}{j} \lambda_1^j \lambda_2^{i-j} \\ &= \frac{e^{-(\lambda_1+\lambda_2)}}{i!} (\lambda_1+\lambda_2)^i. \end{aligned}$$

Cela montre que Z *suit encore une loi de Poisson, de paramètre* $\lambda_1 + \lambda_2$*.*

REMARQUE 65 — *On peut souvent généraliser les études de sommes de 2 v.a. indépendantes à des sommes de* n *v.a. indépendantes.*

18.5.1 Fonction génératrice et indépendance

PROPOSITION 66
Soit $(\Omega, \mathcal{A}, \mathbb{P})$ un espace probabilisé. Soient $X, Y : \Omega \to \mathbb{N}$ deux v.a. indépendantes.
On pose $Z = X + Y$. Alors, les fonctions génératrices de X, Y, Z vérifient

$$G_Z(s) = G_X(s)G_Y(s),\ \forall s \in [0,1].$$

Preuve — Il suffit de remarquer que pour $s \in [0,1]$,

$$G_Z(s) = \mathbb{E}(s^Z) = E(s^{X+Y})$$

et $G_X(s) = \mathbb{E}(s^X)$ et $G_Y(s) = \mathbb{E}(s^Y)$.
La proposition 57 permet de conclure. □

Remarque 67 — *Comme la fonction génératrice de la v.a. Z décrit la loi de probas de Z (ainsi que son espérance, sa variance,...), on peut utiliser cette proposition pour calculer dans certains cas très facilement la loi d'une somme de variables aléatoires.*

Exemple 68 —

1. *Soient $n \geq 1$, $p \in [0,1]$, et $X_1, \dots, X_n$ des variables de Bernoulli de paramètre p qui sont indépendantes. Alors, la v.a. $X = X_1 + \cdots + X_n$ a pour fonction génératrice $(1-p+ps)^n$. Ainsi, X est une v.a. binomiale de paramètres n et p.*
2. *Soient X et Y sont des variables aléatoires indépendantes de lois binomiales $B(n,p)$ et $B(m,p)$ respectivement (avec le même $p \in [0,1]$).*
   Alors, leur somme $Z = X + Y$ vérifie
   $$G_Z(s) = (1-p+ps)^n(1-p+ps)^m = (1-p+ps)^{n+m}.$$
   On en déduit que $X+Y$ est une loi binomiale de paramètres $n+m$ et p.
3. *Soient X, Y des v.a. indépendantes de loi de Poisson de paramètres λ_1 et λ_2 respectivement.*
   Alors, leur somme $Z = X + Y$ vérifie
   $$G_Z(s) = e^{\lambda_1(s-1)}e^{\lambda_2(s-1)} = e^{(\lambda_1+\lambda_2)(s-1)}.$$
   Cela démontre à nouveau (et plus facilement) que $X+Y$ est une v.a. de Poisson de paramètre $\lambda_1 + \lambda_2$.

Exemple 69 — *Soit $n > 1$ un entier fixé. On choisit de manière équiprobable un entier x dans $\{1, \dots, n\}$. Pour tout entier $m \leq n$, on note A_m l'événement "m divise x". On note également B l'événement "x est premier avec n". Enfin, on note $p_1, \dots, p_r$ les diviseurs premiers de n.*

1. *Exprimer B en fonction des A_{p_k}.*
2. *Pour tout $m \leq n$ qui divise n, calculer la probabilité de A_m.*
3. *Montrer que les événements $A_{p_1}, \dots, A_{p_r}$ sont mutuellement indépendants.*
4. *En déduire la probabilité de B.*
5. *En déduire que $\phi(n)$ le nombre d'éléments premiers avec n (inversibles de $\mathbb{Z}/n\mathbb{Z}$). Démontrer que*
   $$\phi(n) = n \prod_{k=1}^{r} \left(1 - \frac{1}{p_k}\right).$$

1. *On sait que x est premier avec n si et seulement si aucun des diviseurs premiers de n ne divise x. On a donc :*
   $$B = A_{p_1}^c \cap \cdots \cap A_{p_r}^c.$$
2. *Il suffit de calculer le cardinal de A_m. Mais si $n = km$, alors les multiples de m qui sont inférieurs ou égaux à n sont $m, 2m, \dots, km$. On a donc*
   $$\mathbb{P}(A_m) = \frac{k}{n} = \frac{1}{m}.$$
3. *Soit $i_1 < \cdots < i_m$ des entiers distincts choisis dans $\{1, \dots, r\}$. On doit prouver que*
   $$\mathbb{P}(A_{p_{i_1}}) \dots \mathbb{P}(A_{p_{i_m}}) = \mathbb{P}(A_{p_{i_1}} \cap \cdots \cap A_{p_{i_m}}).$$
   Mais,
   $$\mathbb{P}(A_{p_{i_1}}) \dots \mathbb{P}(A_{p_{i_m}}) = \prod_{j=1}^{m} \frac{1}{p_{i_j}}.$$
   D'autre part, puisque $p_{i_1}, \dots, p_{i_m}$ sont premiers entre eux deux à deux, un entier est multiple de $p_{i_1} \dots p_{i_m}$ si et seulement s'il est multiple de chaque p_{i_j}, $j = 1, \dots, m$. On en déduit que
   $$A_{p_{i_1}} \cap \cdots \cap A_{p_{i_m}} = A_{p_{i_1} \dots p_{i_m}},$$
   soit
   $$\mathbb{P}(A_{p_{i_1}} \cap \cdots \cap A_{p_{i_m}}) = \frac{1}{p_{i_1} \dots p_{i_m}},$$
   ce qui prouve le résultat voulu.

4. Les événements $A^c_{p_1}, \dots, A^c_{p_r}$ sont également indépendants. On en déduit que

$$\mathbb{P}(B) = \prod_{j=1}^{r} \mathbb{P}(A^c_{p_j}) = \prod_{j=1}^{r} \left(1 - \frac{1}{p_j}\right).$$

5. On a donc

$$\mathbb{P}(B) = \frac{\phi(n)}{n}$$

ce qui, grâce à la question précédente, donne le résultat voulu.

Quand on modélise l'expérience de deux jets de dés équilibrés consécutifs, on considère comme ensemble d'états le produit $\Omega = \{1, \cdots, 6\} \times \{1, .., 6\}$ et on prend pour probabilité la probabilité uniforme sur $(\Omega, \mathcal{P}(\Omega))$.
On vérifie alors facilement que les variables aléatoires X_1 et X_2 qui donnent le résultat du premier jet (resp. second jet) sont des variables aléatoires de loi uniforme sur $\{1, \cdots, 6\}$ qui sont indépendantes.
On se pose la question inverse : On prend $(E_i, \mathcal{A}_i, \mathbb{P}_i)_{i\in I}$ une famille d'espaces probabilisés dont les mesures $\mathbb{P}_i$ sont discrètes. Peut-on construire un espace probabilisé $(\Omega, \mathcal{A}, \mathbb{P})$ et une famille de v.a. discrètes et indépendantes $(X_i)_{i\in I}$, avec X_i à valeurs dans $(E_i, \mathcal{A}_i)$, tels que la loi de X_i soit $\mathbb{P}_i$?
Ce problème est fondamental pour pour la construction des modèles probabilistes, et on dépasse très vite le cadre dénombrable.
Par exemple un lancer de pièces infini (dénombrable) conduit à un espace $\Omega = \{0, 1\}^{\mathbb{N}}$ non dénombrable.
On se limite donc au cas où I est fini.

Proposition 70
Soit $n \geq 1$. Soient $(\Omega_1, \mathcal{P}(\Omega_i), \mathbb{P}_i)$, $1 \leq i \leq n$ des espaces probabilités, tels que Ω_i est dénombrable et $\mathbb{P}_i$ une mesure de probabilité de loi (x^i_j, p^i_j), $j \in \mathbb{N}$.
On pose $\Omega = \prod_{i=1}^n E_i$ et $X_i : \Omega \to \Omega_i$ la projection de Ω sur Ω_i.
Alors, la famille

$$\left((x^1_{i_1}, \cdots, x^n_{i_n}), p^1_{i_1} \cdots p^n_{i_n}\right)_{(i_1,\cdots,i_n)\in\mathbb{N}^n}$$

définit une mesure de probabilité $\mathbb{P}$ sur Ω telle que les variables X_i soient indépendantes.
Cette mesure de probas est appelée **mesure produit** de $\mathbb{P}_1, \dots, \mathbb{P}_n$.

Preuve — On vérifie que

$$\sum_{(x^1_{i_1},\cdots,x^n_{i_n})\in\Omega} p^1_{i_1} \cdots p^n_{i_n} = \prod_{i=1}^{n} \left(\sum_{j\geq 0} p^i_j\right) = 1$$

donc on a bien définit une probabilité.
De plus, de la même manière

$$(X_{i_0} = x^{i_0}_j) = \bigcup_{(a_1,\cdots,a_n)\in\Omega,\ a_{i_0}=x^{i_0}_j} \{(a_1, \cdots, a_n)\}$$

L'union étant disjointe, en sommant, on trouve

$$\mathbb{P}(X_{i_0} = x^{i_0}_j) = p^{i_0}_j.$$

Enfin, pour tout $(i_1, \cdots, i_n) \in \mathbb{N}$, on a par définition de $\mathbb{P}$

$$\mathbb{P}\left(\bigcap_{j=1}^{n} (X_j = x^j_{i_j})\right) = \prod_{j=1}^{n} \mathbb{P}(X_j = x^j_{i_0}).$$

Les variables X_i sont bien indépendantes pour $\mathbb{P}$. □

Théorème 71 (**Théorème de Kolmogorov**)
Soient $(\Omega_1, \mathcal{P}(\Omega_i), \mathbb{P}_i)$, $i \in \mathbb{N}$ des espaces probabilités, tels que Ω_i est dénombrable et $\mathbb{P}_i$ une mesure de probabilité.
Soient $\Omega = \prod_{i=1}^{\infty} \Omega_i$ et $X_i : \Omega \to \Omega_i$ la projection de Ω sur E_i.
Soit $\mathcal{T}$ la sigma-algèbre sur Ω engendrée par

$$(A_n)_{n\in\mathbb{N}} \in (\mathcal{P}(\Omega_i))^{\mathbb{N}} \text{ t.q. } \exists n_0 \text{ avec } A_n = \Omega_n, \forall n \geq n_0.$$

Alors il existe une mesure de probabilité $\mathbb{P}$ sur $(\Omega, \mathcal{T})$ telle que

$$\mathbb{P}\left(\prod_{i=1}^{+\infty} A_n\right) = \prod_{i=1}^{+\infty} \mathbb{P}_i\left(A_n\right).$$

Les v.a. X_i sont alors indépendantes.

Remarque 72 — *Ce théorème nous permet par exemple de modéliser le jeu de pile ou face infini avec la probabilité p d'obtenir face. Ainsi, la probabilité d'obtenir face au i-ème lancer reste p.*
Remarquons que, même dans ce cas élémentaire, la tribu $\mathcal{T}$ n'est pas dénombrable.
La probabilité d'obtenir une suite lancers donnés, par exemple $(P, F, P, F, P \cdots)$, est nulle.

Exemple 73 — *Soit $p \in]0,1[$. On considère une suite de parties de pile ou face de paramètre p, où les parties sont indépendantes.*
Pour $n \geq 1$, on définit la fonction T_n qui donne le numéro de la partie à laquelle on obtient exactement n fois pile (et $T_n(\omega) - n$ fois face).
Enfin, on pose les fonctions $A_1 = T_1$ et $A_n = T_n - T_{n-1}$.

1. *Quelle est la loi de la v.a. T_1 ?*
   Donner la valeur de son espérance.
2. *Soit $n \geq 2$.*
   Montrer que les v.a. $A_1, \ldots, A_n$ sont indépendantes et ont la même loi de probas.

1. *La variable aléatoire T_1 est le temps d'attente du premier pile ; elle suit la loi géométrique de paramètre p, donc d'espérance $1/p$.*
2. *Notons X_n la variable aléatoire égale à 1 si la partie numéro n amène pile et 0 sinon. Les variables X_n sont des variables aléatoires de Bernoulli indépendantes de même paramètre p. Soit $(i_1, \ldots, i_n) \in \mathbb{N}^n$. L'événement $(A_1 = i_1, \ldots, A_n = i_n)$ s'écrit aussi :*

$$X_1 = \cdots = X_{i_1-1} = 0, X_{i_1} = 1, X_{i_1+1} = \cdots = X_{i_1+i_2-1} = 0, X_{i_1+i_2} = 1, \ldots, X_{i_1+\cdots+i_n} = 1.$$

Donc, en posant $q = 1 - p$, on a :

$$P(A_1 = i_1, \ldots, A_n = i_n) = q^{i_1-1}pq^{i_2-1}p \ldots q^{i_n-1}p.$$

En sommant pour $(i_1, \ldots, i_{n-1})$ parcourant $(\mathbb{N}^*)^{n-1}$, on a :

$$P(A_n = i_n) = q^{i_n-1}p.$$

(A_n) suit bien une loi géométrique de paramètre p. De plus l'expression ci-dessus prouve que :

$$P(A_1 = i_1, \ldots, A_n = i_n) = P(A_1 = i_1) \ldots P(A_n = i_n),$$

ce qui montre que les variables $A_1, \ldots, A_n$ sont indépendantes.

18.6 Fonction de répartition

Nous avons défini la fonction génératrice G_X d'une variable aléatoire discrète à valeurs dans $\mathbb{N}$. Celle-ci caractérise la loi de X et permet de calculer l'espérance et plus généralement les moments d'ordre p pour tout p.

Nous considérons ici des variables aléatoires réelles :

Définition 74
Soit X une variable aléatoire et $\mathbb{P}_X$ sa loi. On appelle fonction de répartition de X la fonction suivante :

$$\forall x \in \mathbb{R},\ F_X(x) = \mathbb{P}_X(]-\infty, x]) = \mathbb{P}(X \leq x)$$

Exemple 75 — *Si $\mathbb{P}_X$ est la mesure de Dirac en 0, c'est-à-dire $\mathbb{P}(X = 0) = 1$, la fonction de répartition de X est la fonction de Heaviside en 0 : $H(x) = 0$ si $x < 0$ et $H(x) = 1$ si $x \geq 0$.*

Remarque 76 — *L'exemple ci-dessus montre qu'en général, une fonction de répartition n'est pas une fonction continue. L'étude générale des fonctions de répartition et en particulier des points de continuité est hors programme. Cependant quelques propriétés élémentaires sont à connaître.*

Proposition 77
Avec les notations précédentes, on a les propriétés suivantes

1. F_X est croissante
2. F_X est continue à droite : $\lim\limits_{y \to x^+} F(y) = F(x)$
3. $\lim\limits_{x \to -\infty} F_X(x) = 0$, $\lim\limits_{x \to +\infty} F_X(x) = 1$.

Preuve — La croissance de F est immédiate. Pour montrer 2/, on considère une suite $(x_n)_{n\geq 0}$ décroissant strictement vers x. Alors la suite $(]-\infty, x_n])_{n\geq 0}$ décroît vers $]-\infty, x]$. Par limite décroissante, $F_X(x_n)$ décroît vers $F_X(x)$. Comme F est monotone, on en déduit que F admet une limite à droite.
De même, si $]-\infty, x]$ décroît vers $\emptyset$ (resp. croît vers $\mathbb{R}$) lorsque x décroît vers $-\infty$ (resp. croît vers $+\infty$) et la monotonie de F permet de conclure. □

Remarque 78 — *Comme F est croissante, elle admet une limite à gauche en chaque point notée $F(x-)$. En remarquant que $]-\infty, y[= \bigcup_{n\in\mathbb{N}}]-\infty, y_n]$ si (y_n) croît strictement vers y, on obtient facilement que pour $x < y$,*

1. $\mathbb{P}_X(]x, y]) = \mathbb{P}(x < X \leq y) = F(y) - F(x)$,
2. $\mathbb{P}_X(]x, y[) = \mathbb{P}(x < X < y) = F(y-) - F(x)$,
3. $\mathbb{P}_X([x, y]) = \mathbb{P}(x \leq X \leq y) = F(y) - F(x-)$,
4. $\mathbb{P}_X([x, y[) = P(x \leq X < y) = F(y-) - F(x-)$.

En particulier,

$$\mathbb{P}_X(\{x\}) = F(x) - F(x-)$$

est le saut de la fonction F au point x. Nous avons donc

Proposition 79

$$\mathbb{P}_X(\{x\}) = \mathbb{P}(X = x) = 0 \Leftrightarrow F \text{est continue en } x.$$

Si X est identiquement égale au réel $a \in \mathbb{R}$, alors sa loi est la mesure de Dirac en a et sa fonction de répartition est $F(x) = \mathbb{1}_{[a,\infty[}(x)$.

§ *1. Variable aléatoire à valeurs dans* $\mathbb{N}$ Si X prend ses valeurs dans $\mathbb{N}$, alors sa loi $\mathbb{P}_X$ est caractérisée par la suite $p_n = \mathbb{P}_X(\{n\}) = \mathbb{P}(X = n)$. La fonction de répartition F de X vaut alors

$$\begin{array}{rcl} F : \mathbb{R} & \to & [0,1] \\ x & \mapsto & \begin{cases} 0 \text{ si } x < 0 \\ p_0 + ... + p_n \text{ si } n \leq x < n+1, n \in \mathbb{N} \end{cases} \end{array} .$$

La fonction F est une fonction en escalier qui saute de l'amplitude pn au point n. Puisque $F(x) \in [0,1]$, F admet au plus k sauts de taille supérieure à $\frac{1}{k}$, pour tout $k \in \mathbb{N}^*$.

Exemple 80 — *Soit X une variable géométrique et F sa fonction de répartition. Alors*

$$\forall n \in \mathbb{N}^*,\ F(n) = \sum_{k=1}^{n} p(1-p)^{k-1} = p \times \frac{1-(1-p)^n}{1-(1-p)} = 1-(1-p)^n$$

et donc $F(x) = 1-(1-p)^{E(x)}$ si $x \geq 0$ et 0 sinon.

Plus généralement, si X prend ses valeurs dans une partie finie ou dénombrable E de $\mathbb{R}$, la loi $\mathbb{P}_X$ de X est caractérisée pour tout $x_i \in E$ par $p_i = \mathbb{P}_X(\{x_i\}) = P(X = x_i)$. La fonction de répartition F de X est alors

$$\begin{array}{rcl} F : \mathbb{R} & \to & [0,1] \\ x & \mapsto & \sum_{x_i \leq x} p_i \end{array} .$$

avec la convention qu'une somme "vide" vaut 0.

Proposition 81
La loi d'une variable aléatoire discrète est uniquement déterminée par sa fonction de répartition.

Preuve — La loi d'une variable discrète est par définition la distribution $(x_i, p_i)_{i\in\mathbb{N}}$. D'après la remarque 78, $p_i = \mathbb{P}(X = x_i) = F(x_i) - F(x_i-)$, ce qui montre bien que F détermine uniquement la loi de X. La réciproque étant par définition. □

Remarque 82 — *Notons aussi que l'ensemble E, quoiqu'au plus dénombrable, peut être dense dans $\mathbb{R}$, par exemple il peut être égal à l'ensemble des rationnels $\mathbb{Q}$. Dans ce cas, si $q_i > 0$ pour tout $i \in \mathbb{Q}$, la fonction F nous donne un exemple de fonction discontinue en tout nombre rationnel, et continue partout ailleurs.*

Définition 83
On dit qu'une variable aléatoire discrète $X : \Omega \to \mathbb{N}^*$ est sans mémoire si

$$\forall n, m \in \mathbb{N},\ \mathbb{P}(X > m + n | X > n) = \mathbb{P}(X > m).$$

Remarque 84 — *La variable X est sans mémoire ou sans viellissement car si $\mathbb{P}$ est la probabilité qu'un processus dure n unités de temps, alors si le processus a déjà duré n unités de temps, sa probabilité qu'il dure encore m unités de temps est la même que s'il partait de l'instant* 0.

Proposition 85
Si X est une variable aléatoire discrète sans mémoire $X : \Omega \to \mathbb{N}^*$, alors X est une variable géométrique.

Preuve — Posons $p = \mathbb{P}(X = 1)$. D'après la proposition 81, il suffit de montrer que la fonction de répartition F de X coïncide avec celle d'une variable géométrique : on pose $p = \mathbb{P}(X = 1)$ et on veut montrer que pour tout $n \in \mathbb{N}$, $F(n) = \mathbb{P}(X \leq n) = 1 - (1-p)^n$ ce qui est équivalent à $\mathbb{P}(X > n) = (1-p)^n$.
On a bien :

$$\mathbb{P}(X > 1) = 1 - \mathbb{P}(X = 1) = 1 - p$$

puisque X à valeurs dans $\mathbb{N}^*$.
Enfin,

$$\mathbb{P}(X > k+1 | X > k) = \mathbb{P}(X > 1) \iff \mathbb{P}(X > k+1) = (1-p) \times \mathbb{P}(X > k).$$

Ce qui détermine bien une suite géométrique de raison $(1-p)$. □

On termine ce paragraphe avec une propriété sur l'espérance des variables aléatoires à valeurs dans $\mathbb{N}$ qui peut être fort utile

Proposition 86
Si X est une variable aléatoire discrète $X : \Omega \to \mathbb{N}^*$ intégrable, alors

$$\mathbb{E}(X) = \sum_{n \geq 1} \mathbb{P}(X \geq n).$$

Preuve — Soit (k, p_k) la distribution de X. Alors

$$\mathbb{P}(X \geq n) = \sum_{k \geq n} p_k$$

et

$$\sum_{n \geq 1} \mathbb{P}(X \geq n) = \sum_{n \geq 1} \left(\sum_{k \geq n} p_k \right) = \sum_{k \geq 1} \left(\sum_{n=1}^{k} p_k \right) = \sum_{k \geq 1} k p_k$$

la dernière égalité résulant du théorème de sommation par paquets :

$$I = \bigcup_{n \in \mathbb{N}} \{(n, k)\ ,\ k \geq n\} = \bigcup_{k \in \mathbb{N}} \{(n, k)\ ,\ n \in [\![1, k]\!]\}$$

la somme indexée par la dernière partition étant sommable par hypothèse. □

18.7 L'ensemble $L^2(\Omega, \mathcal{A}, \mathbb{P})$

Nous revenons à l'étude de l'espace vectoriel $L^2(\Omega, \mathcal{A}, \mathbb{P})$.
Le but est de progresser dans la comparaison de variables aléatoires en s'aidant de quantités supplémentaires (covariance, approximation).

Remarque 87 — *Nous avons déjà vu que $L^2(\Omega, \mathcal{A}, \mathbb{P})$, l'ensemble des v.a. réelles et discrètes de carré intégrable sur $(\Omega, \mathcal{A}, \mathbb{P})$, est un espace vectoriel.*

18.7.1 Covariance, approximation linéaire

Proposition 88
Soit $(\Omega, \mathcal{A}, \mathbb{P})$ un espace probabilisé.
• La fonction $(X, Y) \mapsto \mathbb{E}(XY)$ définit une forme bilinéaire symétrique positive sur $L^2(\Omega, \mathcal{A}, \mathbb{P})$.
• L'inégalité de Cauchy-Schwarz donne

$$\mathbb{E}(XY)^2 \leq \mathbb{E}(X^2)\mathbb{E}(Y^2).$$

Preuve — On vérifie avec les propriétés de l'espérance toutes les propriétés de l'énoncé. □

Corollaire 89
En prenant $Y = 1$, on retrouve l'inégalité

$$\mathbb{E}(X)^2 \leq \mathbb{E}(X^2).$$

Remarque 90 — *La forme bilin. $(X, Y) \mapsto \mathbb{E}(XY)$ n'est en général pas un produit scalaire. pas définie positive. Pour X une v.a.r. discrète, de loi de probas $(x_i, \mathbb{P}(X = x_i))$, on a*

$$\mathbb{E}(X^2) = \sum_{i \geq 0} x_i^2 \, \mathbb{P}(X = x_i).$$

Selon la mesure de probas $\mathbb{P}$, on peut avoir $\mathbb{P}(X = x_i) = 0$ même si $x_i \neq 0$.
On a en général seulement :

$$\mathbb{E}(X^2) = 0 \Leftrightarrow \mathbb{P}(X = 0) = 1.$$

C'est-à-dire, $\mathbb{E}(X^2) = 0$ si et seulement si la v.a. X est $\mathbb{P}$-presque sûrement égale à 0.

Proposition-Définition 91
Soit $(\Omega, \mathcal{A}, \mathbb{P})$ un espace probabilisé.
• La fonction **covariance** $Cov : (X, Y) \mapsto \mathbb{E}\big((X - \mathbb{E}(X))(Y - \mathbb{E}(Y))\big)$ définit une forme bilinéaire symétrique positive sur $L^2(\Omega, \mathcal{A}, \mathbb{P})$.
• L'inégalité de Cauchy-Schwarz donne

$$\mathrm{Cov}(X, Y)^2 \leq \mathrm{Var}(X) \, \mathrm{Var}(Y).$$

On peut remarquer que pour $X \in L^2$, on a $Cov(X, X) = \mathrm{Var}(X)$.

Preuve — L'application est clairement bilinéaire symétrique positive puisque $\mathbb{E}$ est linéaire, positive.
Il faut cependant justifier que la covariance est bien définie : développons

$$(X - \mathbb{E}(X))(Y - \mathbb{E}(Y) = XY - \mathbb{E}(X)Y - \mathbb{E}(Y)X + \mathbb{E}(X)\mathbb{E}(Y)$$

qui est somme de variables intégrables puisque $L^2 \subset L^1$ et la variable constante est intégrable. □

Corollaire 92
Soit $(\Omega, \mathcal{A}, \mathbb{P})$ un espace probabilisé. Soient $X, Y \in L^2(\Omega, \mathcal{A}, \mathbb{P})$. On a

$$\mathrm{Cov}(X, Y) = \mathbb{E}(XY) - \mathbb{E}(X)\mathbb{E}(Y).$$

Remarque 93 —
• *La covariance n'est pas un produit scalaire non plus.*
En effet, on a $Cov(X, X) = 0$ ssi $Var(X) = 0$ ssi $\mathbb{E}((X - \mathbb{E}(X))^2) = 0$, ssi $\mathbb{P}(X - \mathbb{E}(X) = 0) = 1$, ssi $\mathbb{P}(X = \mathbb{E}(X)) = 1$.
Une v.a.r. discrète X vérifie $Cov(X, X) = 0$ si et seulement si X est $\mathbb{P}$-presque sûrement égale à son espérance.
• *La covariance est un outil très important dans l'étude des v.a. (de carré intégrable). Elle permet de quantifier à quel point des v.a. X_1, X_2 ne sont pas indépendantes.*

Proposition 94 (Variance d'une somme de v.a.)
Soit $(\Omega, \mathcal{A}, \mathbb{P})$ un espace probabilisé. Soient $X_1, \ldots, X_n$ des v.a. de carré intégrable. On a alors

$$\mathrm{Var}\left(\sum_{i=1}^{n} X_i\right) = \sum_{i=1}^{n} \mathrm{Var}(X_i) + 2 \sum_{1 \leq i < j \leq n} \mathrm{Cov}(X_i, X_j).$$

Preuve — Cela résulte immédiatement du fait que cov est une forme bilinéaire symétrique et que $\mathrm{Var}(X) = \mathrm{cov}(X, X)$. □

Définition 95
Soit $(\Omega, \mathcal{A}, \mathbb{P})$ un espace probabilisé. Soient $X, Y \in L^2(\Omega, \mathcal{A}, \mathbb{P})$ de variance non-nulle.
On définit le **coefficient de corrélation** des v.a. X et Y come

$$\rho(X, Y) = \frac{\mathrm{cov}(X, Y)}{\sqrt{\mathrm{Var}(X)\mathrm{Var}(Y)}} = \frac{\mathrm{cov}(X, Y)}{\sigma_X \sigma_Y}$$

Pour des produits scalaires, cela revient à écrire $\frac{\langle X, Y \rangle}{\|X\| \|Y\|}$.

PROPOSITION 96
Soient $X, Y \in L^2(\Omega, \mathcal{A}, \mathbb{P})$ de variance non-nulle. On a alors

$$-1 \leq \rho(X,Y) \leq 1$$

De plus, on a

$$|\rho(X,Y)| = 1 \iff \exists a, b, c \in \mathbb{R} \text{ tels que } \mathbb{P}(aX + bY + c = 0) = 1.$$

Preuve — On applique l'inégalité de Cauchy-Schwarz au produit bilinéaire symétrique positif cov. Le cas d'égalité est équivalent à $\text{cov}(Y,Y) = 0$ ou $\text{cov}(X+bY, X+bY) = 0$.
D'après la remarque 90

$$\mathbb{E}\Big((X + bY - \mathbb{E}(X+bY))^2\Big) = 0 \Leftrightarrow \mathbb{P}\Big(X + bY - \mathbb{E}(X+bY) = 0\Big) = 1$$

et de même si $\text{cov}(Y,Y) = 0$, alors $\mathbb{P}(Y - \mathbb{E}(Y) = 0) = 1$. □

PROPOSITION 97
Soient $X, Y \in L^2(\Omega, \mathcal{A}, \mathbb{P})$, et $a, b, a', b' \in \mathbb{R}$. Alors on a

$$\text{cov}(aX + b, a'Y + b') = aa'\text{cov}(X,Y).$$

En particulier, cela donne

$$\text{var}(aX + b) = a^2 \text{Var}(X).$$

Preuve — Comme la covariance est une forme bilinéaire symétrique, il suffit en fait de développer l'expression par bilinéarité, et de montrer que $Cov(X, b') = Cov(b, Y) = 0$ (la covariance entre une v.a. et une v.a. constante vaut 0).
Cela donne :

$$\begin{aligned} \mathbb{E}((aX+b)(a'Y+b')) - \mathbb{E}(aX+b)\mathbb{E}(a'Y+b')) &= aa'\mathbb{E}(XY) + ab'\mathbb{E}(X) + a'b\mathbb{E}(Y) + bb' \\ &- \left[aa'\mathbb{E}(X)\mathbb{E}(Y) + ab'\mathbb{E}(X) + a'b\mathbb{E}(Y) + bb'\right] \\ &= aa'\left[\mathbb{E}(XY) - \mathbb{E}(X)\mathbb{E}(Y)\right] \end{aligned}$$

Ce qui montre l'égalité désirée. □

COROLLAIRE 98
Soient $X, Y \in L^2(\Omega, \mathcal{A}, \mathbb{P})$ de variance non-nulle et $a, a', b, b' \in \mathbb{R}$.
Si $aa' > 0$, alors on a $\rho(X,Y) = \rho(aX + b, a'Y + b')$.
Les v.a X et Y, et $aX + b$ et $a'Y + b'$ ont le même coefficient de corrélation linéaire.

Preuve — D'après la proposition ci-dessus, on a

$$\rho(aX + b, a'Y + b') = \frac{aa'\text{cov}(X,Y)}{a\sqrt{\text{cov}(X,X)}a'\sqrt{\text{cov}(Y,Y)}} = \rho(X,Y)$$

□

REMARQUE 99 — *Soit $X \in L^2(\Omega, \mathcal{A}, \mathbb{P})$ une v.a. de carré intégrable, dont l'écart-type est strictement positif ($\sigma_X > 0$).*
Alors la variable aléatoire $Y = \dfrac{X - \mathbb{E}(X)}{\sigma_X}$ est une v.a. de L^2, qui est d'espérance nulle et d'écart-type 1.
*On dit que Y est une v.a. **centrée** ($\mathbb{E}(Y) = 0$) et **réduite** ($\sigma_Y = 1$).*

REMARQUE 100 — *L'inégalité de Minkowski pour la forme bilin $Var(.,.)$ (ou inégalité triangulaire) montre que si X et Y sont de carré intégrable, alors on a*

$$\sigma_{X+Y} \leq \sigma_X + \sigma_Y.$$

Dans l'étude des variables aléatoires, il est fréquent de connaître certaines informations sur la v.a. X (son espérance, sa variance, un échantillon de valeurs $X(\omega)$,...) mais pas toutes les informations.
On peut alors chercher à approximer la v.a. X avec les informations que l'on connaît.
Quelles sont les v.a. (les fonctions) les plus si

Nous considérons X, Y deux v.a. de carré intégrable dont on connaît que les variances et la covariance.
On suppose que la v.a. X est entièrement connue, mais qu Y ne l'est pas vraiment.
On veut ainsi chercher à approximer Y par une v.a. qui dépende de X, avec les informations que l'on connaît.
L'approximation la plus fondamentale que l'on peut faire en mathématiques est l'approximation linéaire, l'approximation par des fonctions affines.
On va ainsi chercher une v.a. de la forme $aX + b$ qui soit la plus "proche" de notre v.a. Y.

Plus proche au quel sens ? On choisira ici au sens **des moindres carrés**, c'est à dire qui minimise la quantité $\mathbb{E}((Y-(aX+b))^2)$.
On retrouve ici des question de minimisation de "distance" sur des espaces vectoriels.
Comment trouver ces coefficients a et b ? En utilisant les mêmes idées que celles du projeté orthogonal !

On a $\{aX+b,\, a,b\in\mathbb{R}\}$ qui est un sous-ev de v.a. de dimension 1 ou 2.
Rappel : Dans un e.v. E, si on connaît une base orthonormée $(f_1,\dots,f_n)$ d'un sous-ev F, alors pour un vecteur $x\in E$ le vecteur $y\in F$ qui est le plus proche de x correspond au projeté orthogonal de x sur F, et on a $y=\sum_{i=1}^n\langle x,f_i\rangle f_i$.
Nous allons ainsi chercher une base "orthonormée" de $Vect(1,X)$ pour la forme bilinéaire symétrique positive $\mathbb{E}(Z.W)$ (qui n'est pas toujours un produit scalaire).

On suppose ici que $Var(X)\neq 0$, donc en particulier que X n'est pas une v.a. constante (sinon l'étude se simplifie trop).
On va chercher à "orthonormaliser" la famille $(1,X)$.
• La v.a. 1 vérifie $\mathbb{E}(1^2)=1$, c'est bon.
• On cherche une v.a. de la forme $X-\lambda.1$ telle que $\mathbb{E}(1\cdot(X-\lambda.1))=0$.
On obtient alors $\lambda=\mathbb{E}(X)$, c'est-à-dire $X'=X-\mathbb{E}(X)$.
• On veut alors "renormaliser" X'. On a $\mathbb{E}((X')^2)=Var(X)=\sigma_X^2\neq 0$.
Donc, la v.a. $X''=\frac{X'}{\sigma_X}=\dfrac{X-\mathbb{E}(X)}{\sigma_X}$ convient.
Posons $\beta=\mathbb{E}(Y.1)$ et $\alpha=\mathbb{E}(YX'')$.
Avec ce choix de coefficients, la bilinéarité de $(Z_1,Z_2)\mapsto\mathbb{E}(Z_1.Z_2)$ nous donne :

$$\mathbb{E}((Y-(\alpha X''+\beta)).(aX+b))=a.0+b.0=0,\ \forall a,b\in\mathbb{R}.$$

Ainsi, d'après le théorème de Pythagore appliqué à $(Z_1,Z_2)\mapsto\mathbb{E}(Z_1.Z_2)$, on a donc :

$$\mathbb{E}((Y-(aX+b))(Y-(aX+b)))=\mathbb{E}([Y-(\alpha X''+\beta)+(\alpha X''+\beta-aX-b)][Y-(\alpha X''+\beta)+(\alpha X''+\beta-aX-b)]),$$
$$\mathbb{E}((Y-(aX+b))(Y-(aX+b)))=\mathbb{E}((Y-(\alpha X''+\beta))^2)+\mathbb{E}((\alpha X''+\beta-aX-b)^2),$$
$$\mathbb{E}((Y-(aX+b))(Y-(aX+b)))\geq\mathbb{E}((Y-(\alpha X''+\beta))^2).$$

On en déduit donc que la meilleure approximation de Y par une v.a. de la forme $aX+b$, au sens des moindres carrés, est $\alpha X''+\beta$.
Reste à calculer α et β.
On a $\beta=\mathbb{E}(Y)$, et $\alpha=\frac{\mathbb{E}(YX-Y\mathbb{E}(X))}{\sigma_X}$.
En ré-exprimant $\alpha X''+\beta$ en fonction de X,σ_Y, $Cov(X,Y)$, $Var(Y)$, on obtient :

Proposition 101
Soient $X,Y\in L^2(\Omega,\mathcal{A},\mathbb{P})$, avec $\sigma_X\neq 0$.
La meilleure approximation de Y par une fonction de la forme $aX+b$, au sens des moindres carrés, est la v.a. :

$$\frac{Cov(X,Y)}{Var(X)}(X-\mathbb{E}(X))+\mathbb{E}(Y)=\rho(X,Y)\frac{\sigma_Y}{\sigma_X}\big(X-\mathbb{E}(X)\big)+\mathbb{E}(Y).$$

Définition 102
Avec les notations précédenes, on appelle **droite de régression linéaire** (ou la droite des moindres carrés) de Y en fonction de X la droite d'équation

$$\big(y-\mathbb{E}(Y)\big)-\rho(X,Y)\frac{\sigma_Y}{\sigma_X}\big(x-\mathbb{E}(X)\big)=0,\ (x,y)\in\mathbb{R}^2.$$

Remarque 103 — *Avec les résultats sur l'orthogonalité, on en déduit aussi que l'écart entre Y et son approximation linéaire en X, au sens des moindres carrés, vaut :*

$$\begin{aligned}\mathbb{E}\left[\Big(\big(Y-\mathbb{E}(Y)\big)-a\big(X-\mathbb{E}(X)\big)\Big)^2\right] &= \mathbb{E}[(Y-\alpha X''-\beta)^2]\\ &= \mathbb{E}(Y^2)-\alpha^2-\beta^2=\mathbb{E}(Y^2)-\frac{Cov(X,Y)^2}{\sigma_X^2}-\mathbb{E}(Y)^2\\ &= Var(Y)-\frac{Cov(X,Y)^2}{Var(X)}=Var(Y)(1-\rho(X,Y)^2)\end{aligned}$$

Ceci montre que plus $|\rho(X,Y)|$ est proche de 1 (plus $|Cov(X,Y)|$ est proche de $Var(X)Var(Y)$), plus l'approximation est bonne.
A l'inverse, si $Cov(X,Y)=0$, alors la distance est maximale et vaut $Var(Y)$.

18.8 Lois conditionnelles

Soit $(\Omega, \mathcal{A}, \mathbb{P})$ un espace probabilisé, et $X : \Omega \to F$, $Y : \Omega \to G$, deux v.a. discrètes.
Nous avons vu dans le paragraphe précédent que si X et Y sont réelles, alors la covariance permet de trouver des v.a. de la forme $f(X)$ qui approximent Y de façon optimale (qui minimise une certaine quantité).
On suppose ici que F et G sont dénombrables.
Pour comprendre comment X et Y sont liées l'une par rapport à l'autre, nous allons utiliser les probabilités conditionnelles.
On considère le couple $Z = (X, Y) : \Omega \to F \times G$. C'est encore une v.a. discrète à valeurs dans un ensemble dénombrable.
Alors, la loi de probabilité de $\mathbb{Z}$ est caractérisée par la famille $(\mathbb{P}(Z = (x, y)))_{(x,y)\in F\times G}$.
On rappelle que l'on a $\mathbb{P}(Z = (x, y)) = \mathbb{P}(X = x \text{ et } Y = y)$.

Définition 104
Soit $(\Omega, \mathcal{A}, \mathbb{P})$ un espace probabilisé, et $X : \Omega \to F$, $Y : \Omega \to G$, deux v.a. discrètes.
Les mesures $\mathbb{P}_X$ et $\mathbb{P}_Y$ sont appellées les **lois marginales** de la v.a. $Z = (X, Y)$.

Exemple 105 — *Un joueur lance en même temps un dé rouge et un dé bleu.*
Soient X le résultat du dé rouge et Y le résultat de la somme des deux dés.
Il est clair que la connaissance de la valeur de X va influer sur les valeurs possibles que peut prendre Y et sur sa loi.
Par exemple, si $X = 3$, alors Y ne pourra prendre que des valeurs supérieures ou égales à 4, ce qui n'est pas le cas si $X = 1$. Il est donc naturel de s'intéresser, pour chaque valeur fixée x_i de X, à la loi de Y avec l'information a priori que $X = x_i$.

Remarque 106 — *Plus généralement, quand on étudie un phénomène aléatoire, on obtient une série de mesures qui chacune donne une information partielle sur le résultat.*
Chacune de ces mesures correspond à une variable aléatoire.
Une bonne compréhension du phénomène correspond à étudier les liens entre ces valeurs.

Définition 107
Soient $X : \Omega \to F$, $Y : \Omega \to G$, deux v.a. discrètes, avec F, G dénombrables.
Soit $x \in F$ tel que $\mathbb{P}(X = x) > 0$.
On appelle **loi de probabilité conditionnelle de Y sachant $X = x$** la mesure de probabilité sur G associée à la famille $(\mathbb{P}(Y = y|X = x))_{y\in G}$.
On la note $\mathbb{P}_{Y|X=x}$.

Remarque 108 — *Ces lois conditionnelles de Y sachant $X = x$ sont a priori différentes pour chaque valeur de x, et différentes de la mesure de probas $\mathbb{P}_Y$.*
Cela vient des propriétés des probabilités conditionnelles que nous avons vues dans le chapitre précédent.

Nous avons en fait les relations suivantes.

Proposition 109
Soient $X : \Omega \to F$, $Y : \Omega \to G$, deux v.a. discrètes, avec F, G dénombrables. On a alors

1. $\mathbb{P}(X = x) = \sum_{y\in G} \mathbb{P}\big(Z = (x, y)\big)$, $\forall x \in F$;
2. $\mathbb{P}_{Y|X=x}(\{y\}) = \mathbb{P}(Y = y|X = x) = \dfrac{\mathbb{P}(Z = (x, y))}{\mathbb{P}(X = x)}$, si $\mathbb{P}(X = x) > 0$.
3. $\mathbb{P}\big(Z = (x, y)\big) = \begin{cases} \mathbb{P}(Y = y|X = x)\mathbb{P}(X = x) \text{ si } \mathbb{P}(X = x) > 0, \\ 0 \text{ sinon} \end{cases}$

Ainsi, la mesure de probas $\mathbb{P}_Z$ est caractérisée par $\mathbb{P}_X$ et par les $\mathbb{P}Y|X = x$ $(x \in F)$, et la réciproque est vraie.

Preuve — Pour montrer le 1/ : l'ensemble $\{X = x_i\}$ est la réunion (finie ou dénombrable) des ensembles deux-à- deux disjoints $\{X = x_i, Y = y_j\} = \{Z = (x_i, y_j)\}$ pour $y_j \in G$. Par σ-additivité, on en déduit le résultat.
L'égalité 2/ est en fait la définition de la probabilité conditionnelle.
Enfin, si $p_i^X = P(X = x_i) > 0$, l'égalité vient de 2/ et sinon $\mathbb{P}(X = x_i, Y = y_j) = 0$ d'après 1/. □

Proposition 110

Avec les notations précédentes et en supposant Y intégrable, on a

$$\mathbb{E}(Y) = \sum_{i\geq 0} \mathbb{E}(Y|X = x_i)\mathbb{P}(X = x_i).$$

Preuve — Par définition $\mathbb{E}(Y) = \sum_{y_j \in G} y_j p_j^Y$. L'égalité 1/ de la proposition 109 nous dit que

$$p_j^Y = \sum_{x_i \in F} \mathbb{P}(Y = y_j, X = x_i).$$

On a donc

$$\mathbb{E}(Y) = \sum_{j\geq 0} \left(\sum_{i\geq 0} y_j \, \mathbb{P}(Y = y_j, X = x_i) \right)$$

La série étant absolument convergente, on peut intervertir les sommes

$$\begin{aligned}
\mathbb{E}(Y) &= \sum_{i\geq 0} \left(\sum_{j\geq 0} y_j \, \mathbb{P}(Y = y_j, X = x_i) \right) \\
&= \sum_{i\geq 0} \left(\sum_{j\geq 0} y_j \, \mathbb{P}(Y = y_j | X = x_i)\mathbb{P}(X = x_i) \right) \text{ Prop 109 3/} \\
&= \sum_{i\geq 0} \left(\mathbb{P}(X = x_i) \sum_{j\geq 0} y_j \, \mathbb{P}(Y = y_j | X = x_i) \right) \\
&= \sum_{i\geq 0} \mathbb{E}(Y|X = x_i)\mathbb{P}(X = x_i)
\end{aligned}$$

On fait abstraction des indices i tels que $\mathbb{P}(X = x_i) = 0$, mais alors, $\mathbb{P}(X = x_i, Y = y_j) = 0$. □

Remarque 111 — *Ce résultat permet de calculer $\mathbb{E}(Y)$ en conditionnant par une variable auxiliaire X. Il généralise la formule des probabilités totales, qui correspond ici à $Y = \mathbb{1}_A$, et $B_i = \{X = x_i\}$:*

$$\mathbb{P}(A) = \sum_{i\in I} \mathbb{P}(A \cap B_i) = \sum_{i\in I} \mathbb{P}(A|B_i)\mathbb{P}(B_i)$$

où les B_i forment un système complet.

Exemple 112 — *Le nombre N de voitures passant devant une station d'essence en un jour suit la loi de Poisson de paramètre $\lambda > 0$. Chaque voiture décide de s'arrêter à la station avec probabilité p indépendamment des autres. On note K le nombre de voitures qui s'arrêtent à la station. Cherchons $\mathbb{E}(K)$.*
Par hypothèse,

$$p_n^N = e^{-\lambda}\frac{\lambda^n}{n!}, \; p_k^{K|N=n} = \binom{n}{k} p^k (1-p)^{n-k}$$

D'où

$$\mathbb{E}(K|N = n) = \sum_{k=0}^{n} k p_k^{K|N=n} = \sum_{k=0}^{n} k \binom{n}{k} p^k (1-p)^{n-k} = np,$$

espérance d'une loi binomiale de paramètre n et p.
D'après la proposition précédente

$$\mathbb{E}(K) = \sum_{n\geq 0} \mathbb{E}(K|N = n)\mathbb{P}(N = n) = p \sum_{n\geq 0} n \, \mathbb{P}(N = n) = p\lambda$$

puisque $\lambda = \mathbb{E}(N)$.

A l'opposé des variables aléatoires discrètes, une autre famille de variables aléatoires possède une caractérisation qui rend très pratique le calcul d'espérance et de variance (ainsi que les résultats associés) : ce sont les variables aléatoires à densité.

18.9 Variables aléatoires à densité

Définition 113 (Fonction de densité)
Soit $f : \mathbb{R} \to \mathbb{R}$ une fonction continue par morceaux.
On dit que f est une **fonction de densité** sur $\mathbb{R}$ si :

1. $f \geq 0$ (f est positive)
2. $f \in L^1(\mathbb{R})$ (f est intégrable sur $\mathbb{R}$)
3. $\int_{\mathbb{R}} f(x)dx = 1$.

Plus généralement, une fonction de densité est une fonction mesurable sur $\mathbb{R}$, pour la σ-algèbre des boréliens de $\mathbb{R}$ (pas forcément continue par morceaux). Cependant nous n'avons pas abordé clairement cette notion, et tous les exemples et manipulations que vous aurez utiliseront des fonctions continues par morceaux (indicatrices, fonctions puissance, exponentielles,...).

Exemple 114 —

1. *Pour $a < b \in \mathbb{R}$, la fonction $f = \frac{1}{b-a}\chi_{[a,b]}$ est une fonction de densité.*
   Elle est bien continue par morceaux, positive partout, intégrable, et son intégrale vaut $\frac{b-a}{b-a} = 1$.
2. *Pour $c > 0$, la fonction $g(x) = \exp(-cx)c\chi_{\mathbb{R}_+}(x)$ est une fonction de densité.*
   Elle est bien continue par morceaux, positive partout, et intégrable.
   Son intégrale sur $\mathbb{R}$ vaut $\int_{\mathbb{R}} g(x)dx = \int_0^{+\infty} c\exp(-cx)dx = [-\exp(-cx)]_0^{+\infty} = 0 - (-1) = 1$.
3. *La fonction $g(x) = \frac{1}{\sqrt{2\pi}}\exp(-\frac{x^2}{2})$ est une fonction de densité.*
   Elle est bien continue par morceaux, positive partout, et intégrable.
   Comme on a $\int_{\mathbb{R}} \exp(-\frac{x^2}{2})dx = \sqrt{2\pi}$, on a $\int_{\mathbb{R}} h(x)dx = 1$.

Plus généralement, si f est une fonction continue par morceaux, positive, et intégrable sur $\mathbb{R}$, alors $g = \frac{1}{\int_{\mathbb{R}} f(x)dx}.f$ est une fonction à densité. Ce sont plus généralement les fonctions de $L^1(\mathbb{R})$ qui sont positives et de norme 1 ($\|f\|_{L^1(\mathbb{R})} = 1$). On obtient donc assez facilement des fonctions de densité à partir des fonctions intégrables sur $\mathbb{R}$ en les renormalisant.

Définition 115 (Variable aléatoire à densité)
Soient $(\Omega, \mathcal{A}, \mathbb{P})$ un espace probabilisé et $X : \Omega \to \mathbb{R}$ une variable aléatoire.
On dit que X est une **variable aléatoire à densité** s'il existe une fonction de densité f telle que :
pour tout $A \in \mathcal{B}(\mathbb{R})$, on a $\mathbb{P}(X \in A) = \int_A f(x)dx$.

Lorsque l'on a une v.a. X à densité, tous les calculs de probabilités en fonction de X se ramènent à des calculs d'intégrale en fonction de f. On peut ainsi appliquer tous les outils du calcul intégral déjà connus pour les fonctions de $\mathbb{R}$ dans $\mathbb{R}$.

Proposition 116
Soient $(\Omega, \mathcal{A}, \mathbb{P})$ un espace probabilisé et $X : \Omega \to \mathbb{R}$ une variable aléatoire à densité, de fonction de densité f.
Alors, la loi de la v.a. X est caractérisée par la fonction f :
Pour tout $A \in \mathcal{B}(\mathbb{R})$, on a $\mathbb{P}_X(A) = \mathbb{P}(X \in A) = \int_A f(x)dx$.

Pour X une v.a. à densité, les propriétés de X découleront des propriétés de sa fonction de densité f.

Remarque 117 — *Pour X une v.a à densité, l'image de X est reliée au **support** de sa fonction de densité f : $Supp(f) = \{x \in \mathbb{R} \ t.q. \ f(x) \neq 0\}$.*
Pour avoir une fonction f continue par morceaux dont l'intégrale vaut 1, il faut que le support de f contienne un intervalle (plus généralement, qu'il soit de mesure de Lebesgue non-nulle). Cet ensemble sera donc forcément infini non-dénombrable : la fonction f ne peut pas prendre un nombre fini ou dénombrable de valeurs, et de même pour la v.a. X.
Ainsi, les v.a. discrètes ne sont pas des v.a. à densité, et réciproquement.

Remarque 118 — *Pour X une v.a. à densité de densité f, et pour $\lambda \neq 0$, la v.a. λX est encore à densité, de fonction de densité $g(x) = \frac{1}{\lambda}f(\frac{x}{\lambda})$.*
Par contre, la somme de deux v.a. à densité n'est pas forcément à densité. Si X est à densité, $-X$ aussi, mais $X + (-X) = 0$ (v.a. constante) n'est pas une v.a. à densité (c'est une v.a. discrète !).
L'ensemble des v.a. à densité n'est donc pas un espace vectoriel, il n'est pas stable par addition.

Théorème 119 (Construction de v.a. à densité)
Soit $f : \mathbb{R} \to \mathbb{R}$ une fonction de densité.
Alors il existe $(\Omega, \mathcal{A}, \mathbb{P})$ un espace probabilisé et $X : \Omega \to \mathbb{R}$ une v.a. telle que X est à densité, de densité f.

Pour chercher des exemples de v.a. à densité ou des contre-exemples à certains résultats, on se contentera ainsi

de chercher des fonctions f qui sont intégrables mais telles que $\sqrt{f}$ ou $x \mapsto xf(x)$ (ou $x \mapsto x^2 f(x)$) ne sont pas intégrables.

18.10 Espérance des v.a. à densité

Contrairement aux v.a. discrètes, où l'espérance se définit sans problèmes grâce à des sommes finies, pour les v.a. à densité cela fait intervenir des intégrales.
La définition formelle de v.a. intégrables fait intervenir une intégrale sur l'ensemble de départ Ω, ce qui sort du cadre de la théorie de l'intégration sur $\mathbb{R}$. Dans le cas des v.a. à densité, l'intégrabilité peut être vérifiée plus facilement. Nous nous contenterons de ce critère.

Théorème 120 (**Lemme de transfert**)
Soit $(\Omega, \mathcal{P}(\Omega), \mathbb{P})$ un espace probabilisé. Soit $X : \Omega \to \mathbb{R}$ une v.a. à densité, de densité f.
Alors, X est intégrable si et seulement si $\int_{-\infty}^{+\infty} |x| f(x)dx < +\infty$.
Dans ce cas, on a alors :

$$\mathbb{E}(X) = \int_{-\infty}^{+\infty} xf(x)dx.$$

Toutes les propriétés générale de l'espérance vues avec les v.a. discrètes restent vraies pour les v.a. à densité (linéarité, positivité, comparaison, variance, covariance, Cauchy-Schwarz,...). Le principal changement est la différence d'écriture de $\mathbb{E}[X]$ (d'un côté une somme/somme de série, de l'autre une intégrale sur $\mathbb{R}$).

18.10.1 Théorème de transfert

Théorème 121 (**Théorème de transfert**)
Soit $(\Omega, \mathcal{A}, \mathbb{P})$ un espace probabilisé.
Soient $X : \Omega \to \mathbb{R}$ une v.a. à densité intégrable, de densité f, et $g : \mathbb{R} \to \mathbb{R}$.
Alors $g(X)$ est intégrable ssi $\int_{-\infty}^{+\infty} |g(x)| f(x)dx < +\infty$.
Si la v.a. à densité $g(X)$ est intégrable, on a alors :

$$\mathbb{E}(g(X)) = \int_{-\infty}^{+\infty} g(x)f(x)dx.$$

Pour X une v.a. à densité, on vérifie si $X^2, X^3, \ldots$ est intégrable en regardant si les intégrales sur $\mathbb{R}$ de $x^2 f(x), |x^3| f(x), \ldots$ sont finies ou non. Cela se ramène à de l'intégrabilité pour des fonctions de $\mathbb{R}$ dans $\mathbb{R}$.

Remarque 122 — *La v.a. $g(X)$ n'est pas toujours à densité. Par exemple pour $g = 0$ (fonction nulle), on a $g(X) = 0$ (v.a. constante) qui n'est pas une v.a. à densité.*
Si la fonction g est bijective et dérivable ($x \mapsto \lambda.x$ avec $\lambda \neq 0$, $\exp$, $\ln$ si X est positive), alors $g(X)$ sera une v.a. à densité, de densité $h(x) = \frac{f(g^{-1}(x))}{g'(g^{-1}(x))}$.
Dans certains cas plus généraux ($x \mapsto x^2$, $x \mapsto exp(-|x|)$), $g(X)$ sera aussi une v.a. à densité, mais la fonction de densité est plus compliquée à écrire.
Tout comme pour les v.a. discrètes, on ne cherche en général pas à détailler précisément la loi de $g(X)$ (sa fonction de densité si elle est à densité), on utilise plutôt le théorème de transfert pour calculer de façon pratique l'espérance de $g(X)$.

Lemme 123
Soient $(\Omega, \mathcal{A}, \mathbb{P})$ un espace probabilisé, et $X : \Omega \to \mathbb{R}$ une v.a. à densité, de densité f.
On a $X \in L^1(\Omega, \mathcal{A}, \mathbb{P})$ si et seulement si $\int_{-\infty}^{+\infty} |x| f(x)dx = 0$.
On a $X \in L^2(\Omega, \mathcal{A}, \mathbb{P})$ si et seulement si $\int_{-\infty}^{+\infty} x^2 f(x)dx = 0$.
Si X et X^2 sont intégrables, on a alors $Var(X) = \mathbb{E}((X - \mathbb{E}(X))^2] = \int_{\mathbb{R}} (x - \mathbb{E}(X))^2 f(x)dx = \mathbb{E}(X^2) - \mathbb{E}(X)^2$.

Proposition 124
Soit $(\Omega, \mathcal{A}, \mathbb{P})$ un espace probabilisé.
L'ensemble $L^2(\Omega, \mathcal{A}, \mathbb{P})$ est un sous-espace vectoriel de $L^1(\Omega, \mathcal{A}, \mathbb{P})$.
Pour tout $X \in L^2(\Omega, \mathcal{A}, \mathbb{P})$, on a

$$|\mathbb{E}(X)| \leq \mathbb{E}(|X|) \leq \sqrt{\mathbb{E}(X^2)}.$$

Preuve — La preuve est identique à celle effectuée pour les v.a. discrètes. □

Remarque 125 — *On retrouve le fait que X^2 intégrable implique X intégrable.*
La réciproque est par contre fausse en général.
Contre-exemple : *On pose $f(x) = \frac{1}{1+|x^3|}\frac{1}{\int_{\mathbb{R}}\frac{1}{1+|x|^3}dx}$.*
La fonction f est continue, positive, intégrable, et $\int_{\mathbb{R}} f(x)dx = 1$. C'est bien une fonction de densité.
Soit X une v.a. à densité, de densité f. On a $|x|f(x) = \frac{|x|}{1+|x|^3}$. Cette fonction est continue sur $\mathbb{R}$, et on a $|x|f(x) \sim \frac{1}{|x|^2}$ quand $x \to +\infty$ et quand $x \to -\infty$. Ainsi la fonction $|x|f(x)$ est intégrable sur $\mathbb{R}$.
Donc, la v.a. X est intégrable, elle est dans $L^1(\Omega, \mathcal{A}, \mathbb{P})$.
Par contre, pour $x^2 f(x) = \frac{x^2}{1+|x|^3}$, on a $x^2 f(x) \sim \frac{1}{|x|}$ quand $x \to +\infty$ et quand $x \to -\infty$. Cette fonction n'est donc pas intégrable sur $\mathbb{R}$.
Donc on a $X \notin L^2(\Omega, \mathcal{A}, \mathbb{P})$.

On généralise aussi la définition d'intégrabilité pour toutes les puissances de X.

Définition 126
Soient $(\Omega, \mathcal{A}, \mathbb{P})$ un espace probabilisé, X une v.a. à densité, et $k \geq 1$.
Si la v.a. X^k est intégrable (si $\mathbb{E}(|X^k|) < +\infty$), on dit que X possède un **moment d'ordre** k.
Le moment d'ordre k de X est la quantité $\mathbb{E}(X^k)$.

18.11 Variables aléatoires à densité usuelles

Définition 127 (Loi uniforme)
Soit $(\Omega, \mathcal{A}, \mathbb{P})$ un espace probabilisé. Soit $X : \Omega \to \mathbb{R}$ une v.a à densité, de densité f.
Si on a $f = \frac{1}{b-a}\chi_{[a,b]}$, on dit que X est une **variable aléatoire uniforme** sur $[a, b]$.

Proposition 128
Soit X une v.a. uniforme sur $[a, b]$.
Alors, on a $\mathbb{E}(X) = \frac{a+b}{2}$ et $Var(X) = \frac{(b-a)^2}{12}$

Preuve — La v.a. X est bornée ($|X| \leq \max(|a|, |b|)$), donc elle est intégrable. X^2 est elle aussi intégrable car bornée par $\max(a^2, b^2)$.
On a $\mathbb{E}(X) = \int_{\mathbb{R}} x.f(x)dx = \int_a^b x\frac{1}{b-a}dx = \frac{1}{b-a}[\frac{x^2}{2}]_a^b = \frac{b^2-a^2}{2(b-a)} = \frac{b+a}{2}$.
On a $\mathbb{E}(X^2) = \int_{\mathbb{R}} x^2.f(x)dx = \int_a^b x^2\frac{1}{b-a}dx = \frac{1}{b-a}[\frac{x^3}{3}]_a^b = \frac{b^3-a^3}{3(b-a)} = \frac{b^2+ab+a^2}{3}$
Cela donne $Var(X) = \mathbb{E}(X^2) - \mathbb{E}(X)^2 = \frac{b^2+ab+a^2}{3} - (\frac{b+a}{2})^2 = \frac{b^2-2ab+a^2}{12} = \frac{(b-a)^2}{12}$. □

Définition 129 (Loi normale)
Soit $(\Omega, \mathcal{A}, \mathbb{P})$ un espace probabilisé. Soit $X : \Omega \to \mathbb{R}$ une v.a. à densité, de densité f.
Soient $m \in \mathbb{R}$ et $\sigma > 0$. Si on a $f(x) = \frac{1}{\sqrt{2\pi\sigma^2}}\exp(-\frac{(x-m)^2}{2\sigma^2})$, on dit que X est une **variable aléatoire de loi normale**, de paramètres m, σ. On le note aussi $X \sim \mathcal{N}(m, \sigma^2)$.
Si $m = 0$ et $\sigma = 1$, on dit que X est une v.a. de **loi normale centrée réduite** $\mathcal{N}(0, 1)$.

Proposition 130
Soient $m \in \mathbb{R}$, $\sigma > 0$ et X une v.a. de loi normale $\mathcal{N}(m, \sigma^2)$.
Alors, on a $\mathbb{E}(X) = m$ et $Var(X) = \sigma^2$.

Preuve — On utilise le fait que $\int_{-\infty}^{+\infty}\exp(-\frac{1}{2}x^2)dx = \sqrt{2\pi}$.
Avec ce résultat, nous avons montré dans un exemple précédent que la fonction f est une densité de probabilité (continue, positive, intégrable, dont l'intégrale sur $\mathbb{R}$ vaut 1).
On a $|x|f(x) = \frac{|x|}{\sqrt{2\pi\sigma^2}}\exp(-\frac{(x-m)^2}{2\sigma^2})$. Cette fonction est continue sur $\mathbb{R}$, donc ses éventuels problèmes d'intégrabilité sont en $\pm\infty$.
Quand $x \to +\infty$, on a $\frac{(x-m)^2}{2\sigma^2} \geq x$ à partir d'un certain réel m_0.
Donc, pour $x > m_0$, on a $|x|f(x) \leq \frac{|x|}{\sqrt{2\pi\sigma^2}}\exp(-x)$.
D'après les croissances comparées, on a $|x^3|\exp(-x) \leq 1$ à partir d'un certain réel m_1.
Cela s'écrit aussi $|x|\exp(-x) \leq \frac{1}{x^2}$.
Donc, pour $x > \max(m_0, m_1)$, on a $|x|f(x) \leq \frac{1}{\sqrt{2\pi\sigma^2}}\frac{1}{x^2}$, qui est intégrable en $+\infty$.
Donc $|x|f(x)$ est intégrable en $+\infty$.
Comme la fonction $|x|f(x)$ est paire, elle est de même intégrable en $-\infty$. Donc, elle est intégrable sur $\mathbb{R}$.
Donc, X est intégrable : $X \in L^1(\Omega, \mathcal{A}, \mathbb{P})$.

Et, on a $\mathbb{E}(X) = \int_{-\infty}^{+\infty} \frac{x}{\sqrt{2\pi\sigma^2}} \exp(-\frac{(x-m)^2}{2\sigma^2})dx = \int_{-\infty}^{+\infty} \frac{x-m+m}{\sqrt{2\pi\sigma^2}} \exp(-\frac{(x-m)^2}{2\sigma^2})dx$
$\mathbb{E}(X) = \int_{-\infty}^{+\infty} \frac{x-m}{\sqrt{2\pi\sigma^2}} \exp(-\frac{(x-m)^2}{2\sigma^2})dx + \frac{m}{\sqrt{2\pi\sigma^2}} \int_{-\infty}^{+\infty} \exp(-\frac{(x-m)^2}{2\sigma^2})dx$
$\mathbb{E}(X) = [-\frac{\sigma^2}{\sqrt{2\pi\sigma^2}} \exp(-\frac{(x-m)^2}{2\sigma^2})]_{-\infty}^{+\infty} + \frac{m}{\sqrt{2\pi\sigma^2}} \int_{-\infty}^{+\infty} \exp(-\frac{y^2}{2})(dy\sqrt{\sigma^2})$
$\mathbb{E}(X) = 0 - 0 + \frac{m}{\sqrt{2\pi\sigma^2}} \sqrt{\sigma^2}\sqrt{2\pi} = m.$
Pour la v.a. X^2, on peut réutiliser les mêmes arguments. La fonction $x \mapsto x^2 f(x)$ est continue sur $\mathbb{R}$ donc n'a des problèmes d'intégrabilité qu'en $+\infty$ ou $-\infty$.
On peut montrer que $x^2 f(x) \leq \frac{x^2}{\sqrt{2\pi\sigma^2}} \exp(-x)$ à partir d'un certain réel m_0 et que $x^2 \exp(-x) \leq \frac{1}{x^2}$ à partir d'un certain réel m_1, et donc que $x^2 f(x) \leq \frac{1}{\sqrt{2\pi\sigma^2}} \frac{1}{x^2}$ quand $x \geq \max(m_0, m_1)$. Cela prouve que $x^2 f(x)$ est intégrable en $+\infty$.
Comme $x^2 f(x)$ est symétrique, elle est donc aussi intégrable en $-\infty$.
On obtient donc que X^2 est intégrable.
On a $Var(X) = \int_{-\infty}^{+\infty} \frac{(x-m)^2}{\sqrt{2\pi\sigma^2}} \exp(-\frac{(x-m)^2}{2\sigma^2})dx = \int_{-\infty}^{+\infty} \frac{\sigma^2 y^2}{\sqrt{2\pi\sigma^2}} \exp(-\frac{y^2}{2})(dy\sqrt{\sigma^2})$ $Var(X) = \frac{\sigma^2\sqrt{\sigma^2}}{\sqrt{2\pi\sigma^2}} \int_{-\infty}^{+\infty} y^2 \exp(-\frac{y^2}{2})dy =$
$\frac{\sigma^2}{\sqrt{2\pi}} \int_{-\infty}^{+\infty} y^2 \exp(-\frac{y^2}{2})dy$
$Var(X) = \frac{\sigma^2}{\sqrt{2\pi}}([y(-\exp(-\frac{y^2}{2}))]_{-\infty}^{+\infty} - \int_{-\infty}^{+\infty} -\exp(-\frac{y^2}{2})dy) = \frac{\sigma^2}{\sqrt{2\pi}}(0 - 0 + \sqrt{2\pi}) = \sigma^2.$ □

Remarque 131 — *Il est plus léger de calculer l'espérance et la variance d'une v.a. X de loi normale centrée réduite $\mathcal{N}(0,1)$ (espérance nulle, variance qui vaut 1). On peut ensuite se ramener au cas général en disant que $\sigma X + m$ est une v.a. de loi normale $\mathcal{N}(m, \sigma^2)$ (composée de X avec la bijection $g(x) = \sigma x + m$).*
Et on sait alors que $\mathbb{E}(\sigma X + m) = \sigma\mathbb{E}(X) + m = m$ et que $Var(\sigma X + m) = Var(\sigma X) = \sigma^2 Var(X) = \sigma^2$.

Définition 132 (Loi exponentielle)
Soit $(\Omega, \mathcal{A}, \mathbb{P})$ un espace probabilisé. Soient $X : \Omega \to \mathbb{R}$ une v.a à densité, de densité f, et $\lambda > 0$.
Si on a $f(x) = \lambda \exp(-\lambda x)\chi_{\mathbb{R}_+}(x)$, on dit que X est une **variable aléatoire de loi exponentielle**, de paramètre λ.

Proposition 133
Soit X une v.a. de loi exponentielle de paramètre λ.
Alors, on a $\mathbb{E}(X) = \frac{1}{\lambda}$ et $Var(X) = \frac{1}{\lambda^2}$

Preuve — On montre que les fonctions $|x|f(x)$ et $x^2 f(x)$ sont intégrables sur $\mathbb{R}$ (soit car on les majore par $\frac{1}{x^2}$ quand $x \to +\infty$ via les croissances comparées, soit car on calcule l'intégrale de $-N$ à N de la fonction et que l'on montre que cette intégrale converge quand $N \to +\infty$).
Soit $N > 0$.
On a : $\int_{-N}^{N} |x|f(x)dx = \int_{-N}^{N} |x|\lambda \exp(-\lambda x)\chi_{\mathbb{R}_+}(x)dx = \int_0^N x\lambda \exp(-\lambda x)dx$.
$\int_{-N}^{N} |x|f(x)dx = [-x\exp(-\lambda x)]_0^N - \int_0^N 1.(-\exp(-\lambda x))dx = -N\exp(-\lambda N) - 0 + [\frac{1}{-\lambda}\exp(-\lambda x)]_0^N$
$\int_{-N}^{N} |x|f(x)dx = -N\exp(-\lambda N) - \frac{1}{\lambda}\exp(-\lambda N) + \frac{1}{\lambda}$.
Or, on a $N\exp(-\lambda N) \to_{N\to+\infty} 0$ et $\exp(-\lambda N) \to_{N\to+\infty} 0$ car $\lambda > 0$.
Donc, l'intégrale $\int_{-N}^{N} |x|f(x)dx$ est convergente quand $N \to +\infty$, et sa limite vaut $\frac{1}{\lambda}$.
On a donc montré que X est intégrable. Et on a $\mathbb{E}[X] = \int_{\mathbb{R}} xf(x)dx = \int_0^{+\infty} xf(x)dx = \frac{1}{\lambda}$.
Le calcul est similaire pour $x^2 f(x)$, on peut calculer $\int_0^N x^2\lambda \exp(-\lambda x)dx$ à l'aide de deux intégrations par partie. Le résultat obtenu converge vers $\frac{2}{\lambda^2}$ quand $N \to +\infty$, ce qui montre que X^2 est intégrable et que $\mathbb{E}[X^2] = \frac{2}{\lambda^2}$.
On a alors $Var(X) = \mathbb{E}[X^2] - \mathbb{E}[X]^2 = \frac{2}{\lambda^2} - \frac{1}{\lambda^2} = \frac{1}{\lambda^2}$. □

Définition 134 (Loi de Cauchy)
Soit $(\Omega, \mathcal{A}, \mathbb{P})$ un espace probabilisé. Soient $X : \Omega \to \mathbb{R}$ une v.a à densité, de densité f, $x_0 \in \mathbb{R}$ et $a > 0$.
Si on a $f = \frac{1}{a\pi(1+(\frac{x-x_0}{a})^2)}$, on dit que X est une **variable aléatoire de loi de Cauchy**, centrée en x_0, de paramètre d'échelle a.

Proposition 135
Soit X une v.a. de loi de Cauchy.
Alors X n'est ni intégrable ni de carré intégrable.

Preuve — Quand $x \to +\infty$, on a $|x|f(x) = \frac{|x|}{a\pi(1+(\frac{x-x_0}{a})^2)} \sim \frac{a}{\pi x}$, ce qui n'est pas intégrable en $+\infty$.
Comme on a $L^2(\Omega, \mathcal{A}, \mathbb{P}) \subset L^1(\Omega, \mathcal{A}, \mathbb{P})$, si X n'est pas intégrable alors elle n'est pas non plus de carré intégrable. □

Définition 136 (Loi Gamma à un paramètre)
Soit $(\Omega, \mathcal{A}, \mathbb{P})$ un espace probabilisé. Soient $X : \Omega \to \mathbb{R}$ une v.a à densité, de densité f, $a > 0$.
Si on a $f = \frac{x^{a-1}e^{-x}}{\Gamma(a)}\chi_{\mathbb{R}_+^*}(x)$, on dit que X est une **variable aléatoire de loi Gamma à un paramètre**, de paramètre a.

Proposition 137
Soit X une v.a. de loi Gamma à un paramètre.

Alors on a $\mathbb{E}[X] = a$ et $Var(X) = a$.

Preuve — Soit $N > 0$. On a $\int_{-N}^{N} |x| f(x) = \int_{-N}^{N} |x| \frac{x^{a-1}e^{-x}}{\Gamma(a)} \chi_{\mathbb{R}_+^*}(x)dx = \int_0^N \frac{x^a e^{-x}}{\Gamma(a)} dx = \frac{1}{\Gamma(a)} \int_0^N x^{(a+1)-1} e^{-x} dx$.
Comme on a $a + 1 > 0$, la fonction $x \mapsto x^{(a+1)-1}e^{-x}$ est intégrable sur $]0, +\infty[$, et son intégrale sur $]0, +\infty[$ vaut $\Gamma(a + 1)$.
Donc, on a $\frac{1}{\Gamma(a)} \int_0^N x^{(a+1)-1} e^{-x} dx \to_{N\to+\infty} \frac{Gamma(a+1)}{Gamma(a)} = a$.
Ainsi, la v.a. X est intégrable. Et on a $\mathbb{E}[X] = \int_{\mathbb{R}} xf(x)dx = \int_0^{+\infty} xf(x) = a$.
Soit $N > 0$. On a $\int_{-N}^{N} x^2 f(x) = \int_{-N}^{N} x^2 \frac{x^{a-1}e^{-x}}{\Gamma(a)} \chi_{\mathbb{R}_+^*}(x)dx = \int_0^N \frac{x^{a+1}e^{-x}}{\Gamma(a)} dx = \frac{1}{\Gamma(a)} \int_0^N x^{(a+2)-1} e^{-x} dx$.
Comme on a $a + 2 > 0$, la fonction $x \mapsto x^{(a+2)-1}e^{-x}$ est intégrable sur $]0, +\infty[$, et son intégrale sur $]0, +\infty[$ vaut $\Gamma(a + 2)$.
Donc, on a $\frac{1}{\Gamma(a)} \int_0^N x^{(a+1)-1} e^{-x} dx \to_{N\to+\infty} \frac{Gamma(a+2)}{Gamma(a)} = a(a + 1)$.
Ainsi, la v.a. X est de carré intégrable. Et on a $\mathbb{E}[X^2] = \int_{\mathbb{R}} x^2 f(x)dx = a(a + 1)$.
On a donc $Var(X) = \mathbb{E}[X^2] - \mathbb{E}[X]^2 = a(a + 1) - a^2 = a$. □

Chapitre 19 Résultats asymptotiques

Table des matières du chapitre

Les variables aléatoires sur un espace probabilisé $(\Omega, \mathcal{A}, \mathbb{P})$ sont des fonctions. On est ainsi amené à étudier des suites de variables aléatoires (des suites de fonctions), par exemple une suite de lancers de Pile ou Face (ou une suite de tirage de cartes, une suite de jets de dés,...).
Avec une suite de fonctions, on cherche à étudier le comportement de la suite (bornée, non-bornée, normes, convergences, limite,...).

19.1 Convergences de v.a.

Nous allons définir plusieurs notions de convergence pour les suites de v.a. réelles.

Définition 1
Soit $(\Omega, \mathcal{A}, \mathbb{P})$ un espace probabilisé. Soient X et $(X_n)_{n\geq 0}$ des v.a. réelles.
On dit que

1. $(X_n)_n$ converge **uniformément presque partout** vers X si X_n et X sont bornées presque partout et si
$$\forall \epsilon > 0,\ \exists n_0 \geq 0 \text{ t.q. } \forall n \geq n_0 \text{ on a } P(|X_n - X| < \epsilon) = 1.$$
On notera parfois $\mathbb{P}(\lim_{n\to+\infty} X_n = X) = 1$.
2. $(X_n)_n$ converge **dans** L^2 vers X si ces v.a. sont de carré intégrable et si $(\mathbb{E}\left[(X - X_n)^2\right])_n$ converge vers 0.
3. $(X_n)_n$ converge **dans** L^1 vers X si ces v.a. sont intégrables et si $(\mathbb{E}\left[|X - X_n|\right])_n$ converge vers 0.
4. $(X_n)_n$ converge **en probabilité** vers X si
$$\lim_{n\to+\infty} \mathbb{P}(|X - X_n| \geq \varepsilon) = 0,\ \forall \varepsilon > 0$$
5. $(X_n)_n$ converge **presque sûrement** vers X si
$$\mathbb{P}(\{\omega \in \Omega \text{ tq } \lim_{n\to+\infty} X_n(\omega) \text{ existe et vaut } X(\omega)\}) = 1.$$
On notera parfois $\mathbb{P}(\lim_{n\to+\infty} X_n = X) = 1$.

Remarque 2 —

- *La convergence uniforme presque partout est la généralisation probabiliste de la convergence uniforme. La suite de fonctions $(X_n)_n$ est uniformément proche de la fonction X, sauf sur une partie de mesure nulle.*
- *La convergence presque sûre est la généralisation probabiliste de la convergence ponctuelle. La suite fonctions $(X_n)_n$ converge "point par point" vers la fonction X sur un emsemble de probabilité 1. On s'autorise à avoir des valeurs ω pour lesquelles on n'a pas la convergence voulue tant que l'ensemble des valeurs problématiques est de probabilité nulle.*
En probabilités, tout ce qui peut se passer dans un ensemble de probabilité nulle ne change rien aux phénomènes que l'on observe (calculs de $\mathbb{P}, \mathbb{E}, Var()$,...). De même pour la convergence, on l'observe "presque partout" (à un ensemble de mesure nulle près).
- *La convergence en probabilité est un phénomène propre aux probabilités (grâce à la mesure de probas $\mathbb{P}$). On veut que X_n soit à distance de X partout sauf sur un ensemble dont la mesure tend vers 0.*
- *Les convergences L^1 et L^2 sont identiques à celles pour les fonctions d'une variable réelle, étant donné que l'espérance $\mathbb{E}$ est une intégrale (une intégrale sur Ω, mais une intégrale).*
- *Si la suite $(X_n)_n$ converge, sa **limite X est unique presque partout**. Si X et Y sont deux limites alors l'ensemble des ω tels que $X(\omega) \neq Y(\omega)$ est de probabilité 0.*

- *Ainsi, si $(X_n)_n$ converge vers X pour une notion (ex presque sûrement) et que l'on s'intéresse à une autre notion de convergence (ex L^2), on sait que $(X_n)_n$ ne pourra converger que vers X pour cette seconde notion, ce qui rend la vérification plus pratique.*
- *Il n'y a aucune équivalence entre ces modes de convergence. Certaines implications sont vraies, mais pour toutes les réciproques on peut construire des contre-exemples.*
- *La convergence uniforme presque partout (convergence dans L^∞) implique la convergence dans L^2 et dans L^1.*
  Si les X_n et X sont bornées presque partout, alors elles sont bien dans L^2 et dans L^1, et on a $\mathbb{E}[|X_n - X|^p] \leq \mathbb{E}[\sup(|X_n - X|)^p] = \sup(|X_n - X|)^p \to_n 0$ (pour $p = 1, 2$).
- *La convergence L^2 implique la convergence L^1.*
  Si X_n et X sont dans L^2 alors elles sont dans L^1, et on peut alors utiliser l'inégalité de Cauchy-Schwarz ($|\mathbb{E}(X_n - X)| \leq \sqrt{\mathbb{E}((X_n - X)^2)}\sqrt{\mathbb{E}(1)}) \to_n 0$.
- *La convergence dans L^1 implique la convergence en probabilité.*
  En effet, l'inégalité de Markov (voir plus bas) donne : $\mathbb{P}(|X_n - X| > \epsilon) \leq \frac{\mathbb{E}[|X_n - X|]}{\epsilon} \to_n 0$.
- *La convergence uniforme presque partout implique la convergence presque sûre.*
  En effet, l'ensemble des ω tels que $X_n(\omega) \to_n X(\omega)$ est alors de mesure 1.
- *La convergence presque sûre implique la convergence en probabilité.*
- *Toutes les autres implications sont fausses en général.*

Exemple 3 — *La convergence L^1 n'implique pas la convergence L^2.*
Pour X_n telle que $\mathbb{P}(X_n = n) = \frac{1}{n\sqrt{n}}$ et $\mathbb{P}(X_n = 0) = 1 - \frac{1}{n\sqrt{n}}$, la v.a. X_n converge presque sûrement vers 0.
On a $\mathbb{E}[X_n - 0] = \frac{n}{n\sqrt{n}} = \frac{1}{\sqrt{n}} \to_n 0$, donc $(X_n)_n$ converge dans L^1 vers 0, mais $\mathbb{E}[X_n^2 - 0^2] = \frac{n^2}{n\sqrt{n}} = \sqrt{n} \to_n +\infty$ donc $(X_n)_n$ ne converge pas dans L^2 vers 0 (et donc ne converge pas tout court dans L^2).
On peut aussi remarquer que $(X_n)_n$ converge en probabilité vers 0 ($\mathbb{P}(|X_n - 0| \geq \epsilon) = \frac{1}{n\sqrt{n}}$ quand n est grand), mais ne converge pas uniformément vers 0.
Donc, la convergence L^1 n'implique pas la convergence uniforme, et la convergence en probabilité n'implique pas la convergence uniforme.

Exemple 4 — *Soit $(\Omega, \mathcal{A}, \mathbb{P})$ un espace probabilisé et $(A_n)_{n\in\mathbb{N}}$ des éléments de $\mathcal{A}$.*

1. *La suite de v.a. $(\mathbb{1}_{A_n})_{n\geq 0}$ converge en probabilités vers 0 si et seulement si $\lim_{n\to+\infty} \mathbb{P}(A_n) = 0$.*
   En effet, pour tout $\varepsilon > 0$ avec $\varepsilon < 1$, on a
   $$\mathbb{P}(|\mathbb{1}_{A_n} - 0| > \varepsilon) = \mathbb{P}(A_n).$$
   Ainsi, le terme de gauche tend vers 0 ssi celui de droite tend vers 0.
2. *La suite de v.a. $(\mathbb{1}_{A_n})_{n\geq 0}$ converge dans L^2 vers 0 si et seulement si $\mathbb{P}(A_n)$ tend vers 0.*
   En effet, on a $\mathbb{E}((\mathbb{1}_{A_n} - 0)^2) = \mathbb{E}(\mathbb{1}_{A_n}) = \mathbb{P}(A_n)$.
   Dans cet exemple, la cv en probabilités est équivalente à la cv L^2.
3. *On suppose que la suite $(\mathbb{1}_{A_n})_{n\geq 0}$ converge presque sûrement vers 0.*
   Soit $\omega \in \Omega$ tel que $\mathbb{1}_{A_n}(\omega)$ tend vers 0. Alors la suite $(\mathbb{1}_{A_n}(\omega))_n$ est stationnaire à partir d'un certain rang et vaut 0.
   Donc ω n'appartient pas à un nombre infini de A_n. Par hypothèse de CV presque sûre, on en déduit que
   $$\mathbb{P}(\limsup A_n) = \mathbb{P}\left(\bigcap_{n\in\mathbb{N}} \bigcup_{p\geq n} Ap\right) = 0.$$
   Par propriété de limite décroissante, on en déduit que $\lim_{n\to+\infty} \mathbb{P}\left(\cup_{p\geq n} A_p\right) = 0$. Comme $A_n \subset \cup_{p\geq n} A_p$, on obtient alors $\lim_{n\to+\infty} \mathbb{P}(A_n) = 0$. Dans cet exemple, la convergence presque sûre implique la convergence en probabilités (ou la cv L^2).
4. *On prend maintenant*
   $$\begin{array}{llll} A_1 = [0,1] & & & \\ A_2 = [0,1/2] & A_3 =]1/2,1] & & \\ A_4 = [0,1/3] & A_5 =]1/3,2/3] & A_6 =]2/3,1] & \\ A_7 = [0,1/4] & A_8 =]1/4,1/2] & A_9 =]1/2,3/4] & A_{10} =]3/4,1] \\ \quad \ldots & & & \end{array}$$

On a alors $\lim \mathbb{P}(A_n) = 0$ *et* $\limsup A_n = [0,1]$. *Donc* $\mathbb{P}\big(\limsup A_n\big) = 1$.
D'après les points précédents, la suite $(\mathbb{1}_{A_n})_{n\geq 0}$ *cv en probabilités (et dans* L^2 *et* L^1*), mais elle ne converge pas presque sûrement.*
Donc, la convergence en probabilité n'implique pas la convergence presque sûre, de même pour la convergence dans L^2 *et pour celle dans* L^1.

Sous certaines conditions (domination, croissance, extraction), on peut récupérer certaines implications. C'est ce qu'énonce par exemple le théorème de Riesz-Fischer.

Théorème 5 (Théorème de Riesz-Fisher)
Soient X_n, X des v.a. dans L^1 (ou L^2) telles que $(X_n)_n$ converge vers X dans L^1 (ou L^2).
Alors, il existe une sous-suite de $(X_n)_n$ qui converge presque sûrement vers X.

Il reste un dernier mode de convergence à définir :

Définition 6 (Convergence en loi)
Soit $(\Omega, \mathcal{A}, \mathbb{P})$ un espace probabilisé. Soient X et $(X_n)_{n\geq 0}$ des v.a. réelles.
On dit que $(X_n)_n$ converge **en loi** vers X si

$$\forall \phi \in C_b^0(\mathbb{R}), \text{ on a } \mathbb{E}[\phi(X_n)] \to_n \mathbb{E}[\phi(X)],$$

où $C_b^0(\mathbb{R})$ est l'ensemble des fonctions continues et bornées sur $\mathbb{R}$.

Remarque 7 —

1. *La convergence en loi est la notion de convergence la plus "faible" pour les v.a., au sens où pratiquement toutes les autres notions impliquent la convergence en loi (alors que la réciproque est fausse).*
2. *Si* $(X_n)_n$ *converge en probabilité vers* X*, alors* $(X_n)_n$ *converge en loi vers* X.
3. *De par sa définition, cette notion de convergence est plus difficile à manipuler que les autres, car on ne peut ni utiliser les* $X_n(\omega)$*, ni étudier* $|X_n - X|$.

4. *Il existe plusieurs caractérisations équivalentes de la convergence en loi, qui permettent de la vérifier plus facilement, mais cela sort du cadre de ce cours.*

L'utilisation la plus classique de la convergence en loi est la suivante :

Proposition 8
Soit $(\Omega, \mathcal{A}, \mathbb{P})$ un espace probabilisé. Soient X et $(X_n)_{n\geq 0}$ des v.a. réelles telles que $(X_n)_n$ converge en loi vers X.
Alors, pour tous $a, b \in \mathbb{R}$ tels que $\mathbb{P}(X = a) = 0 = \mathbb{P}(X = b)$, on a $\mathbb{P}(a \leq X_n \leq b) \to_n \mathbb{P}(a \leq X \leq b)$.
Et, pour tout $A \in \mathcal{B}(\mathbb{R})$ borélien dont la frontière vérifie $\mathbb{P}(X \in Fr(A)) = 0$, on a $\mathbb{P}(X_n \in A) \to_n \mathbb{P}(X \in A)$.

Autrement dit, pour des v.a. réelles, si $(X_n)_n$ converge en loi vers X, si l'on prend A une partie de $\mathbb{R}$ qui est un borélien (disons un intervalle ou une réunion d'intervalles), et si X n'a pas de "masse" sur le bord de A, alors la suite de probabilités $\mathbb{P}(X_n \in A)$ converge vers $\mathbb{P}(X \in A)$.
C'est ce résultat que l'on utilise souvent en pratique avec le TCL (voir plus bas) afin d'estimer une valeur pour des probabilités de la forme $\mathbb{P}(a \leq X_n \leq b)$.

Dans les manipulations liées aux convergences en probabilité et presque sûres (résultats, cas particuliers), une famille d'ensembles revient souvent : les lim inf et lim sup.
Nous rappelons donc ce que sont ces ensembles, et le Lemme qui les concerne.

Définition 9
Soient Ω un ensemble, $\mathcal{A}$ une σ-algèbre, et $(A_n)_{n\in\mathbb{N}}$ une suite d'éléments de $\mathcal{A}$.
On définit la **limite supérieure** de la famille $(A_n)_n$ comme l'ensemble

$$\limsup_n A_n = \bigcap_p \left(\bigcup_{n\geq p} A_n \right) \in \mathcal{A},$$

et la **limite inférieure** de la famille $(A_n)_n$ comme l'ensemble

$$\liminf_n A_n = \bigcup_p \left(\bigcap_{n\geq p} A_n \right) \in \mathcal{A}.$$

La $\limsup_n(A_n)$ contient tous les ω qui appartiennent à une infinité de A_n. La $\liminf_n(A_n)$ contient tous les ω qui sont dans tous les A_n à partir d'un certain rang.

THÉORÈME 10 (**Lemme de Borel-Cantelli**)
Soient Ω un ensemble, $\mathcal{A}$ une σ-algèbre, et $(A_n)_{n\in\mathbb{N}}$ une suite d'éléments de $\mathcal{A}$.

1. Si on a $\sum_{n\geq 0}\mathbb{P}(A_n) < +\infty$, alors $\mathbb{P}(\limsup_n A_n) = 0$.
2. Si la famille $(A_n)_{n\geq 0}$ est indépendante, alors on a

$$\sum_{n\geq 0}\mathbb{P}(A_n) = +\infty \text{ implique } \mathbb{P}(\limsup_n A_n) = 1.$$

19.2 Théorèmes limites

En probabilités, on s'intéresse énormément à des suites de v.a. qui sont de même loi (même espérance, même variance, mêmes probabilités,...) et qui sont indépendantes.
Pour ces suites de v.a., il existe trois grands théorèmes qui fournissent des convergences.

Inégalité de Markov, Loi faible des grands nombres

Nous commençons par une inégalité "simple" mais très souvent utile.

PROPOSITION 11 (**Inégalité de Markov**)
Soient $(\Omega, \mathcal{A}, \mathbb{P})$ un espace probabilisé, et $X \in L^1(\Omega, \mathcal{A}, \mathbb{P})$.
Pour tout $\varepsilon > 0$, on a

$$\mathbb{P}(|X| \geq \varepsilon) \leq \frac{\mathbb{E}(|X|)}{\varepsilon}.$$

Preuve — On rappelle que comme X est intégrable, alors par définition $|X|$ l'es aussi et $\mathbb{E}(|X|)$ existe.
Soit $D = \{\omega \in \Omega \text{ tq } |X(\omega)| \geq \varepsilon\}$.
On a alors l'inégalité de fonctions : $|X| \geq \epsilon \mathbb{1}_D(.)$. (Cela est vrai sur D et sur $\bar{D}$)
Les propriétés de l'espérance (qui est une forme d'intégrale) donnent :

$$\mathbb{E}(|X|) \geq \mathbb{E}(\epsilon\mathbb{1}_D) = \epsilon\mathbb{E}(\mathbb{1}_D) = \epsilon\mathbb{P}(D) = \epsilon\mathbb{P}(|X| \geq \varepsilon).$$

□

REMARQUE 12 — *On peut montrer de la même façon que pour tout $p \geq 1$, si $X^p \in L^1$, alors pour tout $\varepsilon > 0$ on a*

$$\mathbb{P}(|X| \geq \varepsilon) \leq \frac{\mathbb{E}(|X^p|)}{\varepsilon^p}.$$

Cela vient de l'inégalité de fonctions :

$$|X|^p \geq \varepsilon^p \mathbb{1}_{[\varepsilon;+\infty[}(|X|),$$

que l'on combine aux propriétés de l'espérance.
En particulier, en prenant $p = 2$ et $Y = X - \mathbb{E}(X)$, on obtient la proposition suivante

PROPOSITION 13 (**Inégalité de Bienaymé-Tchebychev**)
Soit $X \in L^2(\Omega, \mathcal{A}, \mathbb{P})$.
Alors, pour tout $\varepsilon > 0$ on a

$$\mathbb{P}(|X - \mathbb{E}(X)| \geq \varepsilon) \leq \frac{Var(X)}{\varepsilon^2} = \frac{\sigma_X^2}{\varepsilon^2}.$$

REMARQUE 14 — *L'inégalité de Bienaymé-Tchebychev majore la probabilité que la v.a. X s'éloigne d'au moins ε de sa moyenne.*
Cette inégalité va nous servir à démontrer le prochain théorème, la loi faible des grands nombres.

THÉORÈME 15 (**Loi faible des grands nombres**)
Soit $(\Omega, \mathcal{A}, \mathbb{P})$ un espace probabilisé. Soit $(X_n)_{n\geq 1}$ une suite de v.a. réelles indépendantes, et de même loi de probabilité.

Soit $S_n = \sum_{k=1}^{n} X_k$.
Alors pour tout $\varepsilon > 0$, on a

$$\lim_{n\to+\infty} \mathbb{P}\left(|\frac{S_n}{n} - \mathbb{E}(X_1)| \geq \varepsilon\right) = 0.$$

Preuve —[Cas où X_n est de carré intégrable] On va utiliser les résultats précédents sur la v.a. $\frac{X_1+\ldots+X_n}{n}$.
Comme les v.a. X_n ont la même loi de probabilité, on a en particulier que $\mathbb{E}(X_n) = \mathbb{E}(X_1)$ et $Var(X_n) = Var(X_1)$, $\forall n \geq 1$.
Avec la linéarité de l'espérance, on obtient donc

$$\mathbb{E}\left(\frac{S_n}{n}\right) = \frac{1}{n}\sum_{k=1}^{n} \mathbb{E}(X_k) = \mathbb{E}(X_1).$$

Avec l'indépendance des X_k, le calcul de variance donne

$$Var(\frac{S_n}{n}) = \frac{1}{n^2}\sum_{k=1}^{n} Var(X_k) = \frac{Var(X_1)}{n}.$$

On applique alors l'inégalité de Bienaymé-Tchebychev à la v.a. $\frac{S_n}{n}$:

$$0 \leq \mathbb{P}\left(\left|\frac{S_n}{n} - m\right| \geq \varepsilon\right) \leq \frac{Var(X_1)}{n\varepsilon^2}.$$

Le terme de droite tend vers 0 quand $n \to +\infty$, ce qui conclut. □

REMARQUE 16 — *La loi faible des grands nombrs nous dit que la suite de v.a. $(\frac{1}{n}S_n)_{n\geq 1}$ converge en probabilité vers la v.a. constante $\mathbb{E}(X_1)$.*
• *Cette suite de v.a. $(X_n)_n$ représente une suite de réalisations de la même expérience (on relance la même pièce autant de fois que l'on veut, le même dé autant de fois que l'on veut) et cela de façon indépendante (chaque relance ne tient pas compte des résultats précédents). Ce théorème nous dit alors que le résultat moyen de n réalisations de la même expérience ($\frac{S_n}{n}$) a une probabilité 1 de converger vers l'espérance $\mathbb{E}(X_1)$.*
• *Elle permet de mieux comprendre l'approche des probabilités basée sur la notion de fréquence, celle utilisée au début du chapitre.*
• *Attention, l'hypothèse d'indépendance est indispensable : Si on considère une variable aléatoire X non constante et $(X_n)_{n\geq 1}$ avec $X_n = X$. Alors on a $\frac{1}{n}S_n = X$, et $\mathbb{P}\left(\left|\frac{1}{n}S_n - \mathbb{E}(X)\right| \geq \varepsilon\right))\mathbb{P}(|X - \mathbb{E}(X)| \geq \varepsilon) \neq 0$.*

Loi forte des grands nombres

THÉORÈME 17 (Loi forte des grands nombres)
Soit $(\Omega, \mathcal{A}, \mathbb{P})$ un espace probabilisé. Soit $(X_n)_{n\geq 1}$ une suite de v.a. réelles intégrables, indépendantes, et de même loi de probabilité.
Soit $S_n = \sum_{k=1}^{n} X_k$.
Alors, la suite $(S_n)_n$ converge presque sûrement vers $\mathbb{E}[X_1]$.

REMARQUE 18 — *On parle de loi "faible" et de loi "forte" des grands nombres par rapport au type de convergence obtenu. La convergence presque sûre est plus "forte" que la convergence en probabilité.*

Voici une utilisation de la loi des grands nombres en analyse.

EXEMPLE 19 — *Preuve du théorème d'approximation de Weierstrass par les probabilités.*
Soit $f : [0,1] \to \mathbb{R}$ une fonction continue. Alors il existe une suite de polynômes $(B_n)_{n\geq 1}$ qui converge uniformément vers f sur $[0,1]$.
Preuve : *On pose*

$$B_n(x) = \sum_{k=0}^{n} f\left(\frac{k}{n}\right)\binom{n}{k}x^k(1-x)^{n-k},$$

avec la convention $0^0 = 1$.
Le polynôme B_n s'appelle polynôme de Bernstein de degré au plus n associé à f.
Soit $(\Omega, \mathcal{A}, \mathbb{P})$ un espace probabilisé. Pour $x \in]0,1[$ fixé, on considère une suite $(X_n)_{n\geq 1}$ de v.a. de loi de probas de Bernouilli de paramètre x, et indépendantes.
On pose $S_n = \sum_{i=1}^{n} X_i$. On va pouvoir appliquer des résultats précédents à la v.a. S_n.

1. *La loi de S_n est binomiale : pour tout $k \in \{0, \cdots, n\}$, $\mathbb{P}(S_n = k) = \binom{n}{k} x^k (1-x)^{n-k}$.*

2. *Le théorème de transfert nous dit que*

$$\mathbb{E}\left[f\left(\frac{S_n}{n}\right)\right] = \sum_{k=0}^{n} f\left(\frac{k}{n}\right)\binom{n}{k} x^k (1-x)^{n-k}.$$

3. *Pour tout $\varepsilon > 0$, on pose*

$$\delta(\varepsilon) = \sup\left\{|f(x) - f(y)|,\ x, y \in [0,1] \text{ et } |x-y| \leq \varepsilon\right\}.$$

Le théorème de Heine nous dit que f est uniformément continue et donc $\delta(\varepsilon)$ tend vers 0 quand ε tend vers 0.
Alors, pour tout $x \in]0,1[$, on a

$$|B_n(x) - f(x)| = \left|\mathbb{E}\left[f\left(\frac{S_n}{n}\right)\right] - f(x)\right|,$$

et donc

$$\begin{aligned} |B_n(x) - f(x)| &\leq \mathbb{E}\left[\mathbb{1}_{|\frac{S_n}{n}-x|<\varepsilon}\left|f\left(\frac{S_n}{n}\right) - f(x)\right|\right] \\ &+ \mathbb{E}\left[\mathbb{1}_{|\frac{S_n}{n}-x|\geq\varepsilon}\left|f\left(\frac{S_n}{n}\right) - f(x)\right|\right] \\ &\leq \delta(\varepsilon) + 2\|f\|_\infty \mathbb{P}\left(\left|\frac{S_n}{n} - x\right| \geq \varepsilon\right) \end{aligned}$$

Mais d'après l'inégalité de Bienaymé-Tchebychev, on a

$$\mathbb{P}\left(\left|\frac{S_n}{n} - \mathbb{E}\left(\frac{S_n}{n}\right)\right| \geq \varepsilon\right) \leq \frac{1}{\varepsilon^2}\sigma^2_{\frac{S_n}{n}}.$$

Or

$$\mathbb{E}\left(\frac{S_n}{n}\right) = x$$

et les variables X_n étant indépendantes, on a

$$\sigma^2_{\frac{S_n}{n}} = \frac{1}{n^2}\sigma^2_{S_n} = \frac{1}{n}\sigma^2_{X_1}.$$

On sait enfin que $\sigma^2_{X_1} = x(1-x) \leq 1$ et donc

$$|B_n(x) - f(x)| \leq \delta(\varepsilon) + 2\|f\|_\infty \frac{1}{n\varepsilon^2}.$$

Comme $\lim_{\varepsilon\to 0} \delta(\varepsilon) = 0$, pour tout $\varepsilon' > 0$, il existe $\delta' > 0$, tel que $\varepsilon < \delta'$ implique $\delta(\varepsilon) \leq \frac{\varepsilon'}{2}$.
Et il existe $N \in \mathbb{N}$, tel que $n \geq N$ implique $2\|f\|_\infty \frac{1}{n\varepsilon^2} \leq \frac{\varepsilon'}{2}$. Et finalement, on a montré que pour $n \geq N$,

$$|B_n(x) - f(x)| \leq \delta(\varepsilon) + 2\|f\|_\infty \frac{1}{n\varepsilon^2} \leq \varepsilon',$$

la majoration ne dépendant pas de x, la convergence uniforme est donc bien prouvée.

Théorème central de la limite

THÉORÈME 20 (Théorème Central de la Limite (TCL))
Soit $(\Omega, \mathcal{A}, \mathbb{P})$ un espace probabilisé. Soit $(X_n)_{n\geq 1}$ une suite de v.a.réelles, de carré intégrable, indépendantes, et de même loi de probabilité.
Soit $S_n = \sum_{k=1}^{n} X_k$. On pose $m = \mathbb{E}(X_1)$ et $\sigma^2 = Var(X_1)$
Alors, la suite $(\frac{S_n - m}{\sqrt{n}\sigma})_n$ converge en loi vers une v.a. X de loi normale $\mathcal{N}(0,1)$.

REMARQUE 21 — *Comme une v.a. de loi normale est une v.a. à densité, elle n'a pas de masse concentrée en un seul point.*
Ainsi, la convergence en loi implique que pour tout intervalle (ou réunion d'intervalles) I, on a : $\mathbb{P}(\frac{X_1+\ldots+X_n-m}{\sqrt{n}\sigma} \in I) \to_n \mathbb{P}(X \in I) = \int_I \frac{1}{\sqrt{2\pi}} \exp(-\frac{x^2}{2})dx$.

Approximation d'une loi de Poisson

Théorème 22 (**Théorème de Poisson**)
Soit $\lambda > 0$. Soit $(p_n)_{n\in\mathbb{N}}$ une suite de réels dans $]0,1[$ telle que $\lim_{n\to+\infty} np_n = \lambda$.
Soit $(X_n)_{n\geq 1}$ une suite de v.a., telle que X_n est de loi binomiale $B(n,p_n)$.
Alors, pour tout entier $k \geq 0$, la suite de probabilités $(\mathbb{P}(X_n = k))_n$ converge. On a

$$\lim_{n\to+\infty} \mathbb{P}(X_n = k) = \exp(-\lambda)\frac{\lambda^k}{k!}.$$

Preuve — Soit $k \geq 0$ fixé.
Comme X_n est une v.a. de loi binomiale de paramètres n et p_n, on a

$$\mathbb{P}(X_n = k) = \binom{n}{k} p_n^k (1-p_n)^{n-k}.$$

Par hypothèse, on a

$$p_n = \frac{\lambda}{n} + o\left(\frac{1}{n}\right).$$

On obtient alors :

$$\mathbb{P}(X_n = k) = \binom{n}{k}\left[\frac{\lambda}{n} + o\left(\frac{1}{n}\right)\right]^k \left[1 - \frac{\lambda}{n} + o\left(\frac{1}{n}\right)\right]^{n-k}.$$

Quand n tend vers $+\infty$, le de gauche est équivalent à

$$\binom{n}{k}\left[\frac{\lambda}{n} + o\left(\frac{1}{n}\right)\right]^k = \frac{n(n-1)\cdots(n-k+1)}{k!n^k}\left[\lambda + o\left(1\right)\right]^k \sim \frac{\lambda^k}{k!}.$$

Le terme de droite est équivalent à

$$\left[1 - \frac{\lambda}{n} + o\left(\frac{1}{n}\right)\right]^{n-k} \sim \exp(-\lambda).$$

Avec les propriétés des équivalents, on obtient donc

$$\mathbb{P}(X_n = k) \sim_{n\to+\infty} \exp(-\lambda)\frac{\lambda^k}{k!}.$$

□

Remarque 23 —

1. *Pour approximer correctement une loi de probabilités avec une suite de v.a., nous avons déjà remarqué que l'espérance et l'écart-type doivent coïncider.*
   L'espérance d'une variable aléatoire de loi $B(n,p_n)$ est np_n qui tend bien vers λ, l'espérance de la loi de Poisson correspondante. L'écart-type est $np_n(1-p_n)$ qui tend bien vers λ puisque (p_n) tend vers 0.
2. *On peut améliorer le résultat en obtenant une information sur la vitesse de convergence des $(\mathbb{P}(X_n = k))_n$:*

$$\sum_{k=0}^{+\infty}\left|\mathbb{P}(X_n = k) - \exp(-\lambda)\frac{\lambda^k}{k!}\right| \leq \frac{2\lambda}{n}\min(2,\lambda).$$

   La preuve est cependant plus technique et sort du cadre de ce cours.