

Chapitre 1

Espaces vectoriels normés

Etant donnés deux points A et B dans le plan, on a l'habitude de dire que le chemin le plus court entre A et B est la ligne droite. Qu'entend-on exactement par *le plus court* ? Comment mesure-t-on cette distance ? Identifions le plan muni d'un repère à \mathbb{R}^2 . Dans ce repère, les deux points A et B ont respectivement pour coordonnées (x_A, y_A) et (x_B, y_B) . La distance entre ces deux points est généralement définie comme la norme du vecteur $\overrightarrow{AB} = (x_B - x_A, y_B - y_A)$, à savoir

$$\text{dist}(A, B) = \|\overrightarrow{AB}\| = \sqrt{\langle \overrightarrow{AB}, \overrightarrow{AB} \rangle} = \sqrt{(x_B - x_A)^2 + (y_B - y_A)^2}.$$

Mais il existe en fait bien d'autres manières de mesurer la norme d'un vecteur (et donc aussi la distance entre deux points). La norme décrite précédemment est dite *euclidienne*, parce qu'elle est définie à partir d'un produit scalaire. On la notera $\|\cdot\|_2$ dans la suite.

Imaginons maintenant que pour rejoindre les points A et B , on ne puisse suivre que des lignes horizontales ou verticales. Alors partant de A , on suivra par exemple la ligne horizontale passant par A jusqu'au point intermédiaire C de coordonnées (x_B, y_A) puis on remontera la verticale jusqu'au point B . On aura parcouru la distance $|x_A - x_B| + |y_A - y_B|$. Tout autre chemin entre A et B et suivant uniquement des horizontales ou verticales aurait au moins la même longueur. On obtient de cette manière une autre norme du vecteur \overrightarrow{AB} , appelée la norme $\|\cdot\|_1$, et définie par

$$\|\overrightarrow{AB}\|_1 = |x_A - x_B| + |y_A - y_B|.$$

Il y a en fait une infinité de manières de mesurer la norme d'un vecteur, selon les critères qu'on veut privilégier (nombre limité de directions possibles, comme les horizontales et verticales par exemple). Pourtant, toutes ces normes sont comparables. Ainsi, pour tous points A et B ,

$$\|\overrightarrow{AB}\|_2 \leq \|\overrightarrow{AB}\|_1 \leq \sqrt{2} \|\overrightarrow{AB}\|_2.$$

On dit que ces normes sont équivalentes.

Dans ce chapitre, on introduit précisément la notion de norme, dont on vient de voir deux exemples. On ne se limite pas à \mathbb{R}^2 , puisqu'on peut vouloir mesurer des distances dans \mathbb{R}^3 , ou même dans des espaces de dimension infinie. Dans toute la suite, \mathbb{K} désigne l'ensemble des nombres réels \mathbb{R} ou l'ensemble des nombres complexes \mathbb{C} .

1.1 Norme

1.1.1 Rappels sur les espaces vectoriels

Espaces vectoriels

Définition 1 Soit E un ensemble sur lequel on a défini deux opérations :

- une addition, notée $+$, qui à deux éléments x et y de E , associe un troisième élément de E : $x + y$,
- une multiplication, notée \cdot , qui à un élément x de E et un nombre $\lambda \in \mathbb{K}$, associe un autre élément de E : $\lambda \cdot x$.

On dit alors que E est un espace vectoriel sur \mathbb{K} lorsque l'addition et la multiplication vérifient les propriétés suivantes :

1. Pour tout $x, y, z \in E$,

$$x + (y + z) = (x + y) + z \quad , \quad x + y = y + x.$$

2. Il existe un élément de E , noté 0_E , tel que pour tout $x \in E$,

$$x + 0_E = 0_E + x = x.$$

3. Pour tout $x \in E$, il existe un élément de E , noté $-x$, tel que

$$x + (-x) = (-x) + x = 0_E.$$

4. Pour tout $x, y \in E$, pour tout $\lambda \in \mathbb{K}$,

$$\lambda \cdot (x + y) = (\lambda \cdot x) + (\lambda \cdot y) \quad , \quad (\lambda + \mu) \cdot x = (\lambda \cdot x) + (\mu \cdot x),$$

$$\lambda \cdot (\mu \cdot x) = (\lambda\mu) \cdot x \quad , \quad 1 \cdot x = x.$$

L'exemple le plus simple d'espace vectoriel sur \mathbb{R} est \mathbb{R} lui-même, muni de l'addition et de la multiplication usuelles. L'ensemble \mathbb{R}^2 des couples de réels, muni de l'addition et de la multiplication usuelles définies coordonnées par coordonnées :

$$\forall (x_1, x_2), (y_1, y_2) \in \mathbb{R}^2, (x_1, x_2) + (y_1, y_2) = (x_1 + y_1, x_2 + y_2) \quad , \quad \lambda \cdot (x_1, y_1) = (\lambda x_1, \lambda y_1)$$

est aussi un espace vectoriel. Plus généralement, l'ensemble des n -uplets \mathbb{R}^n est un espace vectoriel lorsqu'on le munit de l'addition et de la multiplication coordonnées par coordonnées. En fait, le produit de deux espaces vectoriels quelconques est encore un espace vectoriel, pour l'addition et la multiplication définies coordonnées par coordonnées.

On note $\mathcal{M}_{n,p}(\mathbb{K})$ l'ensemble des matrices à n lignes et p colonnes à coefficients dans \mathbb{K} . C'est donc l'ensemble des *tableaux* à n lignes et p colonnes dont chaque case est remplie par un nombre (i.e. un élément de \mathbb{K}). On considère deux matrices $A = (a_{ij})_{\substack{1 \leq i \leq n, \\ 1 \leq j \leq p}}$, et $B = (b_{ij})_{\substack{1 \leq i \leq n, \\ 1 \leq j \leq p}}$. On peut encore écrire

$$A = \begin{pmatrix} a_{11} & \cdots & a_{1p} \\ \vdots & \ddots & \vdots \\ a_{n1} & \cdots & a_{np} \end{pmatrix} \quad , \quad B = \begin{pmatrix} b_{11} & \cdots & b_{1p} \\ \vdots & \ddots & \vdots \\ b_{n1} & \cdots & b_{np} \end{pmatrix}.$$

Alors l'addition de A et B est définie coefficients par coefficients : $A + B$ est la matrice $C = (c_{ij})_{\substack{1 \leq i \leq n, \\ 1 \leq j \leq p}}$, où pour tout $(i, j) \in \llbracket 1, n \rrbracket \times \llbracket 1, p \rrbracket$, $c_{ij} = a_{ij} + b_{ij}$. Autrement dit,

$$A + B = \begin{pmatrix} a_{11} + b_{11} & \cdots & a_{1p} + b_{1p} \\ \vdots & \ddots & \vdots \\ a_{n1} + b_{n1} & \cdots & a_{np} + b_{np} \end{pmatrix}.$$

La multiplication de la matrice A par un nombre $\lambda \in \mathbb{K}$ est la matrice $D = (d_{ij})_{\substack{1 \leq i \leq n, \\ 1 \leq j \leq p}}$, où $d_{ij} = \lambda a_{ij}$.

Autrement dit,

$$\lambda A = \begin{pmatrix} \lambda a_{11} & \cdots & \lambda a_{1p} \\ \vdots & \ddots & \vdots \\ \lambda a_{n1} & \cdots & \lambda a_{np} \end{pmatrix}.$$

Muni de ces deux opérations, l'ensemble des matrices $\mathcal{M}_{n,p}(\mathbb{K})$ est un espace vectoriel.

Donnons un dernier exemple d'espace vectoriel. Soit I un intervalle de \mathbb{R} . On peut munir l'ensemble $\mathcal{F}(I; \mathbb{R})$ des fonctions définies sur I et à valeurs dans \mathbb{R} , de deux opérations naturelles : pour tout $f, g \in \mathcal{F}(I; \mathbb{R})$ et $\lambda \in \mathbb{R}$, on définit les fonctions $f + g$ et $\lambda \cdot f$ par

$$\forall x \in I, (f + g)(x) = f(x) + g(x) \quad , \quad (\lambda \cdot f)(x) = \lambda f(x).$$

Muni de ces deux opérations, l'ensemble $\mathcal{F}(I; \mathbb{R})$ devient un espace vectoriel sur \mathbb{R} . Plus généralement, l'ensemble $\mathcal{F}(A; E)$ des fonctions d'un ensemble quelconque A dans un espace vectoriel E sur \mathbb{K} , muni des deux opérations naturelles d'addition et de multiplication par un nombre *point par point*, est un espace vectoriel sur \mathbb{K} .

Sous-espaces vectoriels

Soit E un espace vectoriel sur \mathbb{K} muni des opérations $+$ et \cdot . Soit F une partie non vide de E . Supposons que F soit stable pour l'addition et la multiplication, autrement dit, pour tout $x, y \in F$, pour tout $\lambda \in \mathbb{K}$,

$$x + y \in F \quad , \quad \lambda \cdot x \in F.$$

Alors on peut définir la *restriction* de l'addition et de la multiplication à F , c'est-à-dire considérer $+$ comme une application qui à deux éléments de F associe un élément de F , et considérer \cdot comme une application qui à un élément de F et un nombre de \mathbb{K} associe un élément de F . L'ensemble F est alors muni de deux opérations.

Proposition 1 Une partie non vide F d'un espace vectoriel E qui est stable pour l'addition et la multiplication par les nombres est un espace vectoriel.

Définition 2 On dit alors que F est un sous-espace vectoriel de E .

On en déduit la méthode la plus utilisée pour montrer qu'un ensemble est un espace vectoriel :

Méthode 1 Etant donné un ensemble F , pour montrer que F est un espace vectoriel sur \mathbb{K} associé à deux opérations $+$ et \cdot ,

1. on identifie un espace vectoriel E pour ces deux lois qui contient F ,
2. on montre que F est non vide,
3. on montre que F est stable pour l'addition et la multiplication, par exemple en montrant que pour tout $x, y \in F$, pour tout $\lambda \in \mathbb{K}$, $\lambda x + y \in F$.

Exemple 1 L'ensemble des fonctions continues $C^0([a, b]; \mathbb{R})$ du segment $[a, b]$ à valeurs réelles est un espace vectoriel lorsqu'on le munit de l'addition point par point et la multiplication par des nombres point par point.

On sait que $\mathcal{F}([a, b]; \mathbb{R})$ muni de l'addition et multiplication point par point est un espace vectoriel sur \mathbb{R} qui contient $C^0([a, b]; \mathbb{R})$. L'ensemble $C^0([a, b]; \mathbb{R})$ est non vide, par exemple parce qu'il contient les fonctions constantes. Pour tout $f, g \in C^0([a, b]; \mathbb{R})$ et pour tout $\lambda \in \mathbb{R}$, λf est continu (comme produit d'une fonction continue par un nombre) puis $\lambda f + g$ est continue (comme somme de deux fonctions continues). Ainsi, $C^0([a, b]; \mathbb{R})$ est stable pour l'addition et la multiplication par les nombres. En conclusion, $C^0([a, b]; \mathbb{R})$ est un espace vectoriel (comme sous-espace vectoriel de $\mathcal{F}([a, b]; \mathbb{R})$).

Dimension

Soit E un espace vectoriel.

Définition 3 Soient (x_1, \dots, x_n) une famille d'éléments de E .

1. On dit que cette famille est *génératrice* si tout élément $x \in E$ peut s'écrire comme une combinaison linéaire d'éléments de E , autrement dit, il existe $\lambda_1, \dots, \lambda_n \in \mathbb{K}$ tels que

$$x = \lambda_1 x_1 + \dots + \lambda_n x_n.$$

2. On dit que cette famille est *libre* quand pour tout $\lambda_1, \dots, \lambda_n \in \mathbb{K}$, si

$$\lambda_1 x_1 + \dots + \lambda_n x_n = 0,$$

alors $\lambda_1 = \dots = \lambda_n = 0$.

3. On dit que cette famille est une *base* de E si elle est à la fois libre et génératrice.

Définition 4 On dit que E est de dimension finie lorsqu'il existe $n \in \mathbb{N}^*$ et une famille (x_1, \dots, x_n) d'éléments de E qui est une famille génératrice de E .

Il faut retenir les résultats suivants :

Théorème 1 1. Tout espace de dimension finie admet une base.

2. De toute famille génératrice de E , on peut extraire une base de E .
3. Toute famille libre de E peut être complétée en une base.

Théorème 2 Soit E un espace vectoriel de dimension finie. Toutes les bases de E ont même cardinal n . L'entier n s'appelle la dimension de E (par convention, si $E = \{0\}$, on dit que E est de dimension 0).

Théorème 3 Soit E un espace vectoriel de dimension $n \in \mathbb{N}^*$. Alors

1. Toute famille libre a au plus n éléments et une famille libre qui a exactement n éléments est une base.
2. Toute famille génératrice a au moins n éléments et une famille qui a exactement n éléments est une base.

Exemple 2 L'ensemble $\mathbb{R}[X]$ des polynômes à coefficients dans \mathbb{R} est un espace vectoriel sur \mathbb{R} (on vérifie que c'est un sous-espace vectoriel de $\mathcal{F}(\mathbb{R}; \mathbb{R})$). Pour tout $k \in \mathbb{N}$, notons P_k le polynôme $P_k : x \mapsto x^k$. Alors

1. pour tout $n \in \mathbb{N}$, la famille (P_0, P_1, \dots, P_n) est libre,
2. l'espace $\mathbb{R}[X]$ n'est pas de dimension finie,
3. pour tout $n \in \mathbb{N}$, l'espace $\mathbb{R}_n[X]$ des polynômes de degré $\leq n$ est un espace vectoriel de dimension finie égale à $n + 1$.

Montrons que la famille (P_0, P_1, \dots, P_n) est libre. Soient $\lambda_0, \dots, \lambda_n \in \mathbb{R}$ et supposons que $\lambda_0 P_0 + \lambda_1 P_1 + \dots + \lambda_n P_n = 0$. Alors pour tout $x \in \mathbb{R}$,

$$\lambda_0 P_0(x) + \lambda_1 P_1(x) + \dots + \lambda_n P_n(x) = 0.$$

Ainsi,

$$\lambda_0 + \lambda_1 x + \dots + \lambda_n x^n = 0.$$

Posons $f(x) := \lambda_0 + \lambda_1 x + \dots + \lambda_n x^n, x \in \mathbb{R}$. Alors pour tout $k \in \llbracket 0, n \rrbracket$, $f^{(k)}(0) = k! \lambda_k$. Comme par ailleurs f est la fonction nulle, toutes ses dérivées sont nulles, en particulier en 0. Ainsi pour tout $k \in \llbracket 0, n \rrbracket$, $k! \lambda_k = 0$ et donc $\lambda_k = 0$. Conclusion : la famille (P_0, P_1, \dots, P_n) est libre.

Cela permet d'en déduire que $\mathbb{R}[X]$ n'est pas de dimension finie. Pour le voir, procédons par l'absurde. Si $\mathbb{R}[X]$ était de dimension finie, il existerait un entier $n \in \mathbb{N}$ tel que $\mathbb{R}[X]$ serait de dimension n . Alors toute famille libre aurait au plus n éléments. Or, d'après le premier point, la famille (P_0, \dots, P_n) qui possède $n + 1$ éléments, est libre : contradiction. Conclusion : $\mathbb{R}[X]$ n'est pas de dimension finie.

Pour tout $n \in \mathbb{N}$, l'espace $\mathbb{R}_n[X]$ des polynômes de degré $\leq n$ est un espace vectoriel. Pour le voir, montrons que c'est un sous-espace vectoriel de l'espace vectoriel $\mathbb{R}[X]$. D'abord, $\mathbb{R}_n[X]$ est non vide, car il contient le polynôme nul. Ensuite, si $P, Q \in \mathbb{R}_n[X]$ et $\lambda \in \mathbb{R}$, alors λP a le même degré que P si $\lambda \neq 0$ et est le polynôme nul si $\lambda = 0$. Dans les deux cas, $\lambda P \in \mathbb{R}_n[X]$. De plus, $\lambda P + Q$ est de degré $\leq \max(\deg \lambda P, \deg Q)$. Donc $\lambda P + Q \in \mathbb{R}_n[X]$. Ainsi, $\mathbb{R}_n[X]$ est stable par addition et multiplication par les nombres. C'est donc un sous-espace vectoriel de $\mathbb{R}[X]$ et donc un espace vectoriel.

Montrons que $\mathbb{R}_n[X]$ est de dimension $n+1$. Pour cela, il suffit de vérifier que la famille (P_0, \dots, P_n) , qui est bien contenue dans $\mathbb{R}_n[X]$, est une base de $\mathbb{R}_n[X]$. On sait déjà qu'elle est libre. Soit $P \in \mathbb{R}_n[X]$. Il existe $a_0, \dots, a_n \in \mathbb{R}$ tels que

$$P(x) = a_0 + a_1 x + \dots + a_n x^n, \quad \forall x \in \mathbb{R}.$$

On en déduit que $P = a_0 P_0 + \dots + a_n P_n$. Autrement dit P s'écrit comme combinaison linéaire de P_0, \dots, P_n . Cela montre que (P_0, \dots, P_n) est une famille génératrice de $\mathbb{R}_n[X]$. Finalement, (P_0, \dots, P_n) est une base de $\mathbb{R}_n[X]$, ce qui implique que $\mathbb{R}_n[X]$ est de dimension $n + 1$.

1.1.2 Définition d'une norme

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

Définition 5 On appelle norme sur E toute application $N : E \rightarrow \mathbb{R}^+$ vérifiant

1. (séparation) $\forall x \in E, N(x) = 0$ si et seulement si $x = 0$,
2. (positive homogénéité) $\forall x \in E, \forall \lambda \in \mathbb{K}, N(\lambda x) = |\lambda| N(x)$,
3. (inégalité triangulaire) $\forall x, y \in E, N(x + y) \leq N(x) + N(y)$.

Exemple 3 Sur \mathbb{K}^n , les applications

$$\|x\|_1 = \sum_{i=1}^n |x_i|, \quad \|x\|_2 = \sqrt{\sum_{i=1}^n |x_i|^2}, \quad \|x\|_\infty = \max_{1 \leq i \leq n} |x_i|$$

sont des normes. En revanche, l'application $x \in \mathbb{R} \mapsto x \in \mathbb{R}$ n'est pas une norme.

Exemple 4 Sur $\mathcal{M}_n(\mathbb{K})$, on définit les normes suivantes : pour tout $A = (a_{ij})_{\substack{1 \leq i \leq n \\ 1 \leq j \leq n}}$,

$$N_1(A) = \sum_{1 \leq i, j \leq n} |a_{ij}|, \quad N_2(A) = \sqrt{\sum_{1 \leq i, j \leq n} |a_{ij}|^2}, \quad N_\infty(A) = \max_{1 \leq i, j \leq n} |a_{ij}|.$$

Proposition 2 Soit $N : E \rightarrow \mathbb{R}^+$ une norme.

1. $N(-x) = N(x)$,
2. $N(x_1 + \dots + x_n) \leq N(x_1) + \dots + N(x_n)$,
3. $|N(x) - N(y)| \leq N(x - y)$.

Preuve : La première propriété est une conséquence de la positive homogénéité. La deuxième propriété se démontre par récurrence sur n à l'aide de l'inégalité triangulaire. Montrons la dernière propriété. Soient $x, y \in E$. Comme $x = y + (x - y)$, l'inégalité triangulaire donne

$$N(x) \leq N(y) + N(x - y),$$

d'où l'on déduit

$$N(x) - N(y) \leq N(x - y).$$

En écrivant maintenant $y = x + (y - x)$, on obtient

$$N(y) - N(x) \leq N(y - x) = N(x - y).$$

Ainsi,

$$|N(x) - N(y)| \leq N(x - y).$$

□

Définition 6 A toute norme $N : E \rightarrow \mathbb{R}^+$ on associe une distance $d(x, y) = N(x - y)$.

Définition 7 Un espace vectoriel normé (E, N) est un espace vectoriel E muni d'une norme N . On dit alors qu'une suite $(x_k)_{k \in \mathbb{N}} \subset E$ converge vers x si la suite de réels $(N(x_k - x))_{k \in \mathbb{N}}$ converge vers 0 :

$$\lim_{k \rightarrow +\infty} N(x_k - x) = 0.$$

Remarque 1 Comme pour les suites réelles, la limite d'une suite, lorsqu'elle existe, est unique.

En effet, si x et x' sont deux limites d'une suite $(x_k)_{k \in \mathbb{N}}$, on a

$$N(x - x') \leq N(x - x_k) + N(x' - x_k).$$

En prenant la limite quand $k \rightarrow +\infty$ de l'inégalité précédente, on obtient

$$N(x - x') = 0,$$

et donc $x = x'$. Cela prouve l'unicité de la limite.

Ainsi, la convergence d'une suite est directement liée à la norme qu'on considère sur E . D'où la question : si on se donne une autre norme N' sur E , est-ce qu'une suite qui convergeait au sens de la norme N va encore converger pour la norme N' ? La notion d'équivalence des normes permet justement de donner une condition suffisante pour que ce soit le cas :

Définition 8 On dit que deux normes N et N' sont équivalentes quand il existe $C > 0$ et $C' > 0$ tels que pour tout $x \in E$,

$$N(x) \leq CN'(x) \quad , \quad N'(x) \leq C'N(x).$$

Insistons : les constantes C et C' sont indépendantes de x !

La propriété d'équivalence des normes permet de montrer que la convergence des suites ne dépend pas de la norme équivalente choisie :

Proposition 3 Si deux normes sont équivalentes alors toute suite qui converge pour l'une converge vers la même limite pour l'autre.

Preuve : Soient N et N' deux normes équivalentes et $(x_k)_{k \in \mathbb{N}} \subset E$ une suite convergente pour N . Appelons x sa limite : $\lim_{k \rightarrow +\infty} N(x_k - x) = 0$. On sait qu'il existe $C > 0$ tel que $N'(y) \leq CN(y)$ pour tout $y \in E$. Alors pour tout $k \in \mathbb{N}$,

$$0 \leq N'(x_k - x) \leq CN(x_k - x).$$

On en déduit par le théorème des gendarmes que $\lim_{k \rightarrow +\infty} N'(x_k - x) = 0$. Donc $(x_k)_{k \in \mathbb{N}}$ converge vers x pour N' . □

Remarque 2 *La réciproque de la Proposition 3 est vraie.*

On se donne deux normes N et N' sur E et on suppose que toute suite qui converge pour l'une converge vers la même limite pour l'autre. On veut montrer que N et N' sont équivalentes. On procède par l'absurde : pour tout $C > 0$, il existe $x_C \in E$ tel que

$$N'(x_C) > CN(x_C).$$

On applique cette assertion pour chaque valeur de C entière. Ainsi, pour tout $k \in \mathbb{N}$, il existe $x_k \in E$ tel que $N'(x_k) > kN(x_k)$. En particulier, $N'(x_k) \neq 0$. On pose $y_k = x_k / N'(x_k)$. Alors

$$N'(y_k) = 1, \quad N(y_k) = \frac{N(x_k)}{N'(x_k)} < \frac{1}{k}.$$

Donc $N(y_k)$ tend vers 0, ou de manière équivalente, $(y_k)_{k \in \mathbb{N}}$ tend vers 0 pour N . Ce n'est pas le cas pour N' : cela contredit notre hypothèse que toute suite qui converge pour N converge pour N' vers la même limite ! Conclusion : N et N' sont équivalentes.

1.1.3 Normes en dimension finie

Dans un espace vectoriel de dimension finie, la situation est beaucoup plus simple. En effet,

Théorème 4 *Dans un espace vectoriel de dimension finie, toutes les normes sont équivalentes.*

La preuve de ce théorème est difficile. Elle repose sur la notion de compacité qu'on verra à la fin de ce chapitre. Très souvent, il est possible de montrer que deux normes sont équivalentes en exhibant des constantes explicites.

Exemple 5

$$\forall x \in \mathbb{R}^n, \|x\|_\infty \leq \|x\|_2 \leq \|x\|_1 \leq n\|x\|_\infty.$$

Selon le problème, il sera commode d'utiliser plutôt une norme qu'une autre. Par exemple,

Exemple 6 *Soit $(u_n)_{n \in \mathbb{N}}$ une suite de \mathbb{R}^2 . Pour tout $n \in \mathbb{N}$, on note v_n la première coordonnée de u_n et w_n la seconde coordonnée de u_n : $u_n = (v_n, w_n)$. Alors $(u_n)_{n \in \mathbb{N}}$ converge si et seulement si les suites $(v_n)_{n \in \mathbb{N}}$ et $(w_n)_{n \in \mathbb{N}}$ convergent.*

Comme la convergence d'une suite ne dépend pas de la norme équivalente choisie et que toutes les normes sont équivalentes dans \mathbb{R}^2 , on peut démontrer chaque implication avec la norme qui nous convient le mieux.

Supposons d'abord que $(u_n)_{n \in \mathbb{N}}$ converge et montrons que $(v_n)_{n \in \mathbb{N}}$ et $(w_n)_{n \in \mathbb{N}}$ convergent. Notons $\ell = (\ell_1, \ell_2) \in \mathbb{R}^2$ la limite de la suite $(u_n)_{n \in \mathbb{N}}$. On choisit sur \mathbb{R}^2 la norme $\|\cdot\|_\infty$. Comme pour tout $n \in \mathbb{N}$, $0 \leq |v_n - \ell_1| \leq \|u_n - \ell\|_\infty$, on en déduit $\lim_{n \rightarrow +\infty} |v_n - \ell_1| = 0$. Ainsi, $(v_n)_{n \in \mathbb{N}}$ converge vers ℓ_1 . De même, $(w_n)_{n \in \mathbb{N}}$ converge vers ℓ_2 . Cela prouve l'implication directe.

Réciproquement, supposons $(v_n)_{n \in \mathbb{N}}$ et $(w_n)_{n \in \mathbb{N}}$ convergentes, vers des limites qu'on note a et b respectivement. Montrons alors que $(u_n)_{n \in \mathbb{N}}$ converge vers (a, b) . On choisit sur \mathbb{R}^2 la norme $\|\cdot\|_1$. Comme pour tout $n \in \mathbb{N}$, $\|u_n - (a, b)\|_1 = |v_n - a| + |w_n - b|$ et que les deux termes du membre de droite tendent vers 0, on en déduit

$$\lim_{n \rightarrow +\infty} \|u_n - (a, b)\|_1 = 0.$$

Cela prouve l'implication réciproque.

Exemple 7 Soit $A \in \mathcal{M}_2(\mathbb{R})$ une matrice qui a deux valeurs propres distinctes de module < 1 . Alors la suite $(A^n)_{n \in \mathbb{N}}$ tend vers 0.

Comme A a deux valeurs propres distinctes (a priori complexes), elle est diagonalisable sur \mathbb{C} : il existe $P \in GL_2(\mathbb{C})$ et une matrice diagonale $D \in \mathcal{M}_2(\mathbb{R})$ tels que $A = PDP^{-1}$. Les éléments diagonaux de D sont les valeurs propres λ_1 et λ_2 de A . Alors pour tout $n \in \mathbb{N}$, $A^n = PD^nP^{-1}$.

On introduit sur $\mathcal{M}_2(\mathbb{R})$ une norme bien adaptée à notre problème. On considère d'abord la norme N_1 définie pour tout $C = (c_{ij})_{1 \leq i, j \leq 2}$ par $N_1(C) = \sum_{1 \leq i, j \leq 2} |c_{ij}|$. On définit maintenant la norme $N(C) = N_1(P^{-1}CP)$ (on admet ici qu'il s'agit bien d'une norme).

Il reste à observer que

$$N(A^n) = N(PD^nP^{-1}) = N_1(P^{-1}(PD^nP^{-1})P) = N_1(D^n) = |\lambda_1|^n + |\lambda_2|^n.$$

Comme $|\lambda_1| < 1$ et $|\lambda_2| < 1$, on en déduit $\lim_{n \rightarrow +\infty} N(A^n) = 0$, q.e.d.

Définition 9 Etant donné une norme sur $\|\cdot\|$ sur \mathbb{K}^n , on appelle norme d'opérateur associée la norme sur $\mathcal{M}_n(\mathbb{K})$ définie par

$$\|A\| = \sup_{\substack{x \in \mathbb{K}^n \\ x \neq 0}} \frac{\|Ax\|}{\|x\|}.$$

Remarque 3 On a aussi

$$\|A\| = \sup_{\substack{x \in \mathbb{K}^n \\ \|x\|=1}} \|Ax\|. \quad (1.1)$$

En effet, comme $\{x \in \mathbb{K}^n \mid \|x\| = 1\} \subset \{x \in \mathbb{K}^n \mid x \neq 0\}$ et que $\frac{\|Ax\|}{\|x\|} = \|Ax\|$ lorsque $\|x\| = 1$, on a

$$\sup_{\substack{x \in \mathbb{K}^n \\ \|x\|=1}} \|Ax\| \leq \|A\|.$$

Ensuite, pour tout $y \in \mathbb{K}^n$, $y \neq 0$, posons $x = \frac{y}{\|y\|}$. Alors $\|x\| = 1$ et $\frac{\|Ay\|}{\|y\|} = \|A(y/\|y\|)\| = \|Ax\|$. On en déduit

$$\frac{\|Ay\|}{\|y\|} \leq \sup_{\substack{x \in \mathbb{K}^n \\ \|x\|=1}} \|Ax\|.$$

Finalement, on a bien l'égalité (1.1).

Proposition 4 La norme d'opérateur est bien une norme !

Preuve : Observons d'abord que l'application $A \in \mathcal{M}_n(\mathbb{K}) \mapsto \|A\|$ est bien une application à valeurs positives. Elle vaut 0 en 0 et si $A \in \mathcal{M}_n(\mathbb{K})$ vérifie $\|A\| = 0$, alors pour tout $x \in \mathbb{K}^n$, $Ax = 0$. En prenant pour x les vecteurs de la base canonique $e_i = (0, \dots, 0, 1, 0, \dots, 0)$ (où le 1 est en $i^{\text{ème}}$ position), on voit que chaque colonne de A est nulle. Ainsi $A = 0$, ce qui achève la preuve de la séparation. Par ailleurs, pour tout $\lambda \in \mathbb{R}$, pour tout $x \in \mathbb{K}^n$,

$$\|(\lambda A)x\| = |\lambda| \|Ax\|.$$

On en déduit que

$$\sup_{\substack{x \in \mathbb{K}^n \\ \|x\|=1}} \|(\lambda A)x\| = \sup_{\substack{x \in \mathbb{K}^n \\ \|x\|=1}} |\lambda| \|Ax\| = |\lambda| \sup_{\substack{x \in \mathbb{K}^n \\ \|x\|=1}} \|Ax\|,$$

et donc $\|\lambda A\| = |\lambda| \|A\|$, ce qui prouve la positive homogénéité. Enfin, pour tout $A, B \in \mathcal{M}_n(\mathbb{K})$, pour tout $x \in \mathbb{K}^n$ tel que $\|x\| = 1$,

$$\|(A+B)x\| = \|Ax+Bx\| \leq \|Ax\| + \|Bx\| \leq \|A\| + \|B\|.$$

On en déduit $\|A+B\| \leq \|A\| + \|B\|$, c'est-à-dire l'inégalité triangulaire. Conclusion : la norme opérateur est bien une norme. □

Proposition 5 Soient $A, B \in \mathcal{M}_n(\mathbb{K})$ et $x \in \mathbb{K}^n$.

1. $\|Ax\| \leq \|A\| \|x\|$,
2. $\|AB\| \leq \|A\| \|B\|$.

Preuve : Pour tout $x \in \mathbb{K}^n \setminus \{0\}$, $\|A\| \geq \frac{\|Ax\|}{\|x\|}$. Donc

$$\|Ax\| \leq \|A\| \|x\|.$$

Ensuite, pour tout $x \in \mathbb{K}^n$,

$$\|ABx\| = \|A(Bx)\| \leq \|A\| \|Bx\| \leq \|A\| \|B\| \|x\|.$$

Donc $\|AB\| \leq \|A\| \|B\|$. □

Exercice 1 On considère la matrice

$$A = \begin{pmatrix} 1 & 2 \\ 3 & 4 \end{pmatrix}.$$

Calculer la norme opérateur de A lorsqu'on prend sur K^2 la norme $\|\cdot\|_1$, puis la norme $\|\cdot\|_2$, et enfin la norme $\|\cdot\|_\infty$.

Exercice 2 Sur $\mathcal{M}_n(\mathbb{K})$, on définit l'application N par : pour tout $A = (a_{ij})_{\substack{1 \leq i \leq n, \\ 1 \leq j \leq n}}$,

$$N(A) = \max_{\substack{1 \leq i \leq n, \\ 1 \leq j \leq n}} |a_{ij}|.$$

1. Montrer que N est une norme sur $\mathcal{M}_n(\mathbb{K})$.
2. Montrer qu'il n'existe pas de norme $\|\cdot\|$ sur K^n telle que N soit la norme opérateur associée à $\|\cdot\|$.

1.1.4 Normes en dimension infinie

En dimension infinie, il n'est plus vrai que les normes soient toutes équivalentes. On se sert souvent de la contraposée de la Proposition 3 pour montrer que deux normes ne sont pas équivalentes :

Méthode 2 Pour montrer que deux normes N et N' ne sont pas équivalentes, il suffit de trouver une suite $(x_n)_{n \in \mathbb{N}}$ dans E qui converge vers 0 pour N mais pas pour N' .

Exemple 8 Sur l'espace vectoriel $\mathbb{R}[X]$ des polynômes à coefficients réels, on définit les deux applications N et N' suivantes : étant donné un polynôme $P(x) = a_0 + a_1x + \dots + a_nx^n$, $x \in \mathbb{R}$,

$$N(P) := \max_{0 \leq i \leq n} |a_i|, \quad N'(P) := \sum_{0 \leq i \leq n} |a_i|.$$

Alors N et N' sont des normes sur $\mathbb{R}[X]$ qui ne sont pas équivalentes.

Admettons ici que N et N' sont des normes sur $\mathbb{R}[X]$ et montrons qu'elles ne sont pas équivalentes. Pour tout $n \in \mathbb{N}$, on définit le polynôme

$$Q_n(x) = \frac{1}{n+1}(1 + x + \dots + x^n).$$

Alors $N(Q_n) = \frac{1}{n+1}$ tandis que $N'(Q_n) = 1$. On en déduit que la suite $(Q_n)_{n \in \mathbb{N}}$ converge vers 0 pour la norme N mais pas pour la norme N' .

Exemple 9 Soit $[a, b]$ un segment de \mathbb{R} . On considère sur $\mathcal{C}^0([a, b]; \mathbb{K})$ les deux normes suivantes :

$$\|f\|_1 = \int_a^b |f(t)| dt, \quad \|f\|_2 = \sqrt{\int_a^b |f(t)|^2 dt}.$$

Ces deux normes ne sont pas équivalentes.

On admet ici qu'il s'agit bien de normes. Pour montrer qu'elles ne sont pas équivalentes, il suffit de trouver une suite $(f_n)_{n \in \mathbb{N}^*} \subset \mathcal{C}^0([a, b]; \mathbb{K})$ telle que $\|f_n\|_1 \rightarrow 0$ alors que $\|f_n\|_2 \not\rightarrow 0$. On prend

$$f_n(x) = \begin{cases} -n^3(x-a) + n & \text{si } a \leq x \leq a + \frac{b-a}{n^2}, \\ 0 & \text{sinon.} \end{cases}$$

Alors $\|f_n\|_1 = \frac{b-a}{2n}$ (il s'agit de l'aire sous la courbe) et donc $(f_n)_{n \in \mathbb{N}^*}$ tend vers 0 pour la norme $\|\cdot\|_1$. Pour estimer $\|f_n\|_2$, on observe que

$$\begin{aligned} \|f_n\|_2^2 &= \int_a^{a+\frac{b-a}{n^2}} (-n^3(t-a) + n)^2 dt = n^2 \int_0^{\frac{b-a}{n^2}} (-n^2t + 1)^2 dt \\ &= \int_0^{b-a} (-t + 1)^2 dt \neq 0. \end{aligned}$$

En particulier, $(\|f_n\|_2)_{n \in \mathbb{N}^*}$ ne tend pas vers 0.

Conclusion : ces deux normes ne sont pas équivalentes sur $\mathcal{C}^0([a, b]; \mathbb{K})$.

Sur l'espace $\mathcal{C}^0([a, b]; \mathbb{K})$, une autre norme aura beaucoup d'importance dans la suite. Pour la définir, on a besoin de rappeler que

Théorème 5 Toute fonction $f \in \mathcal{C}^0([a, b]; \mathbb{K})$ est bornée : il existe $M > 0$ tel que pour tout $x \in [a, b]$, $|f(x)| \leq M$. De plus, f atteint ses bornes : il existe $c, d \in [a, b]$ tels que pour tout $x \in [a, b]$,

$$f(c) \leq f(x) \leq f(d).$$

On a donc $f(c) = \min_{x \in [a, b]} f(x)$, $f(d) = \max_{x \in [a, b]} f(x)$.

On définit alors

$$\|f\|_\infty = \max_{x \in [a, b]} |f(x)|.$$

Observer qu'avec les notations précédentes $\|f\|_\infty = \max(|f(c)|, |f(d)|)$.

Proposition 6 L'application $f \in \mathcal{C}^0([a, b]; \mathbb{K})$ est une norme.

Preuve : D'abord, il s'agit bien d'une application à valeurs positives. La fonction nulle (qui est constamment égale à 0) vérifie $\|0\|_\infty = 0$. Réciproquement, si $f \in \mathcal{C}^0([a, b]; \mathbb{K})$ vérifie $\|f\|_\infty = 0$, alors pour tout $x \in [a, b]$, $|f(x)| \leq \max_{x' \in [a, b]} |f(x')| = 0$ donc f est la fonction nulle, ce qui prouve la séparation.

Pour tout $\lambda \in \mathbb{K}$, $f \in \mathcal{C}^0([a, b]; \mathbb{K})$, on a

$$|\lambda f(x)| = |\lambda| |f(x)|, \quad x \in [a, b].$$

On en déduit

$$\|\lambda f\|_\infty = \max_{x \in [a, b]} |\lambda| |f(x)|$$

On observe ensuite que $\max_{x \in [a, b]} |\lambda| |f(x)| = |\lambda| \max_{x \in [a, b]} |f(x)|$. En effet, pour tout $x \in [a, b]$, $|\lambda| |f(x)| \leq |\lambda| \max_{x \in [a, b]} |f(x)|$ d'où l'on déduit que $\max_{x \in [a, b]} |\lambda| |f(x)| \leq |\lambda| \max_{x \in [a, b]} |f(x)|$. Puis, pour tout $x \in [a, b]$ et tout $\lambda \neq 0$, $|f(x)| = \frac{1}{|\lambda|} |\lambda f(x)|$, et donc

$$|f(x)| \leq \frac{1}{|\lambda|} \max_{x \in [a, b]} |\lambda f(x)|.$$

Cela implique $\max_{x \in [a, b]} |f(x)| \leq \frac{1}{|\lambda|} \max_{x \in [a, b]} |\lambda f(x)|$ et finalement

$$|\lambda| \max_{x \in [a, b]} |f(x)| \leq \max_{x \in [a, b]} |\lambda f(x)|.$$

Cette inégalité reste vraie pour $\lambda = 0$. On a donc bien $\max_{x \in [a, b]} |\lambda| |f(x)| = |\lambda| \max_{x \in [a, b]} |f(x)|$. Alors

$$\|\lambda f\|_\infty = |\lambda| \max_{x \in [a, b]} |f(x)| = |\lambda| \|f\|_\infty.$$

La positive homogénéité est démontrée.

Enfin, soient $f, g \in \mathcal{C}^0([a, b]; \mathbb{K})$. Alors pour tout $x \in [a, b]$,

$$|f(x) + g(x)| \leq |f(x)| + |g(x)| \leq \|f\|_\infty + \|g\|_\infty.$$

Donc

$$\max_{x \in [a, b]} |f(x) + g(x)| \leq \|f\|_\infty + \|g\|_\infty,$$

ce qui prouve bien l'inégalité triangulaire.

□

On dispose de certaines inégalités entre les normes $\|\cdot\|_1$, $\|\cdot\|_2$ et $\|\cdot\|_\infty$. Par exemple,

$$\|f\|_1 \leq (b-a)\|f\|_\infty.$$

En effet, il suffit d'intégrer sur $[a, b]$ l'inégalité $|f(x)| \leq \|f\|_\infty$. Une autre inégalité plus subtile est l'inégalité de Cauchy-Schwarz :

Proposition 7

$$\|f\|_1 \leq \sqrt{b-a} \sqrt{\int_a^b |f(t)|^2 dt}.$$

Pour montrer cette proposition, il est commode d'introduire le produit scalaire naturellement associé à la norme $\|\cdot\|_2$. C'est l'un des objets de la section suivante.

1.1.5 Norme et produit scalaire

Définition 10 Une application $\langle \cdot, \cdot \rangle : V \times V \rightarrow \mathbb{R}$ est un produit scalaire si elle vérifie les propriétés suivantes :

1. (linéarité par rapport à la première variable) pour tout $w \in V$, la fonction $v \in V \mapsto \langle v, w \rangle \in \mathbb{R}$ est linéaire,
2. (linéarité par rapport à la deuxième variable) pour tout $v \in V$, la fonction $w \in V \mapsto \langle v, w \rangle \in \mathbb{R}$ est linéaire,
3. (symétrie) pour tout $v, w \in V$, $\langle v, w \rangle = \langle w, v \rangle$,
4. (défini positif) pour tout $v \in V$, $\langle v, v \rangle \geq 0$ avec égalité si et seulement si $v = 0$.

Exemple 10 Sur \mathbb{R}^n , on définit $\langle x, y \rangle = \sum_{j=1}^n x_j y_j$ en notant $x = (x_1, \dots, x_n)$, $y = (y_1, \dots, y_n)$. On vérifie qu'il s'agit d'un produit scalaire.

Exemple 11 Dans $C^0([a, b]; \mathbb{R})$, $(f, g) \mapsto \int_a^b f(t)g(t) dt$ est un produit scalaire.

En effet, par linéarité de l'intégrale, on a bien linéarité par rapport à la première variable f et linéarité par rapport à la seconde variable g . Pour montrer la symétrie, on observe que

$$\int_a^b f(t)g(t) dt = \int_a^b g(t)f(t) dt$$

Enfin,

$$\int_a^b f(t)f(t) dt = \int_a^b f(t)^2 dt \geq 0$$

avec égalité si et seulement si $f = 0$ (on utilise ici que l'intégrale d'une fonction continue positive est nulle ssi la fonction est partout nulle).

Proposition 8 (Inégalité de Cauchy-Schwarz) Pour tout $v, w \in V$

$$|\langle v, w \rangle| \leq \sqrt{\langle v, v \rangle} \sqrt{\langle w, w \rangle}.$$

Preuve : Soit $t \in \mathbb{R}$. Alors

$$0 \leq \langle v + tw, v + tw \rangle = t^2 \langle w, w \rangle + 2t \langle v, w \rangle + \langle v, v \rangle.$$

Comme le trinôme du second degré du membre de droite ne change pas de signe, c'est que son discriminant est ≤ 0 :

$$\langle v, w \rangle^2 \leq \langle v, v \rangle \langle w, w \rangle.$$

On en déduit l'inégalité désirée en prenant les racines carrées de chaque membre. □

Exemple 12 Pour tout $f \in C^0([a, b]; \mathbb{R})$, $\|f\|_1 \leq \sqrt{b-a} \|f\|_2$, autrement dit

$$\int_a^b |f(t)| dt \leq \sqrt{b-a} \sqrt{\int_a^b |f(t)|^2 dt}.$$

Pour le voir, on introduit sur $C^0([a, b]; \mathbb{R})$ le produit scalaire

$$(f, g) \in C^0([a, b]; \mathbb{R}) \times C^0([a, b]; \mathbb{R}) \mapsto \int_a^b f(t)g(t) dt.$$

On applique l'inégalité de Cauchy-Schwarz aux fonctions $t \mapsto f(t)$ et $t \mapsto 1$.

Proposition 9 Etant donné un produit scalaire $\langle \cdot, \cdot \rangle$, on lui associe l'application $\| \cdot \| : V \rightarrow \mathbb{R}$

$$\|v\| = \sqrt{\langle v, v \rangle}, \quad v \in V$$

qui est une norme sur V .

Preuve : D'abord, pour tout $v \in V$, $\|v\| \geq 0$. De plus, $\|v\| = 0$ si et seulement si $\langle v, v \rangle = 0$, ce qui est équivalent à $v = 0$ (car le produit scalaire est défini positif).

Pour tout $\lambda \in \mathbb{R}$, $v \in V$,

$$\langle \lambda v, \lambda v \rangle = \lambda^2 \|v\|^2.$$

Donc

$$\|\lambda v\| = |\lambda| \|v\|.$$

Enfin, pour tout $v, w \in V$, en utilisant l'inégalité de Cauchy-Schwarz,

$$\begin{aligned} \langle v+w, v+w \rangle &= \langle v, v+w \rangle + \langle w, v+w \rangle \\ &= \langle v, v \rangle + \langle v, w \rangle + \langle w, v \rangle + \langle w, w \rangle \\ &= \|v\|^2 + \|w\|^2 + 2\langle v, w \rangle \\ &\leq \|v\|^2 + \|w\|^2 + 2\|v\|\|w\| \\ &= (\|v\| + \|w\|)^2. \end{aligned}$$

Ainsi,

$$\|v+w\|^2 \leq (\|v\| + \|w\|)^2,$$

d'où l'on déduit que $\| \cdot \|$ vérifie l'inégalité triangulaire. □

1.2 Ouverts et fermés

Dans cette section, on donne la définition des ouverts et des fermés. Ce sont des mots qui sont déjà apparus pour qualifier des intervalles. Un intervalle ouvert est un intervalle qui ne contient pas ses extrémités alors qu'un intervalle fermé est un intervalle qui contient ses extrémités. Plus généralement, un ouvert est une partie d'un espace qui ne contient pas son bord alors qu'un fermé est une partie de l'espace qui contient son bord (on dit aussi sa frontière). Il s'agit là d'une définition informelle, qui sert à guider l'intuition, mais qui doit être précisée (par exemple, la notion de bord reste floue).

Dans la suite, on se place dans un espace vectoriel normé $(E, \| \cdot \|)$.

1.2.1 Boules et sphères

Définition 11 On définit

1. la boule ouverte de centre a et de rayon $r > 0$

$$B(a, r) = \{x \in E : \|x - a\| < r\},$$

2. la boule fermée de centre a et de rayon $r > 0$

$$\overline{B(a, r)} = \{x \in E : \|x - a\| \leq r\},$$

3. la sphère de centre a et de rayon $r > 0$

$$S(a, r) = \{x \in E : \|x - a\| = r\}.$$

Exemple 13 Dessiner les boules centrées en 0 et de rayon 1 pour les normes $\|\cdot\|_1$, $\|\cdot\|_2$ et $\|\cdot\|_\infty$ dans \mathbb{R}^2 .

Par définition

$$B_{\|\cdot\|_2}(0, 1) = \{x = (x_1, x_2) \in \mathbb{R}^2 : \sqrt{x_1^2 + x_2^2} < 1\} = \{x = (x_1, x_2) \in \mathbb{R}^2 : x_1^2 + x_2^2 < 1\}.$$

On reconnaît le disque (ouvert) centré en 0 de rayon 1.

Ensuite, pour dessiner

$$B_{\|\cdot\|_1}(0, 1) = \{x = (x_1, x_2) \in \mathbb{R}^2 : |x_1| + |x_2| < 1\},$$

on commence par considérer les points (x_1, x_2) dans cette boule tels que $x_1 > 0, x_2 > 0$. On trace la droite $x_1 + x_2 = 1$ et on garde donc tous les points au-dessous de cette droite dans le quadrangle supérieur droit. On répète la même opération pour les autres quadrangles, ou bien on procède par symétrie. On obtient le carré de sommets $(\pm 1, 0)$ et $(0, \pm 1)$.

Pour la boule

$$B_{\|\cdot\|_\infty}(0, 1) = \{x = (x_1, x_2) \in \mathbb{R}^2 : \max(|x_1|, |x_2|) < 1\},$$

on suit la même stratégie, et on obtient le carré de sommets $(\pm 1, \pm 1)$.

Une boule, ça peut donc être un carré !

Remarque 4 L'équivalence des normes se traduit géométriquement en termes d'inclusions de boules.

Si N_1 et N_2 sont deux normes équivalentes, i.e. il existe $C_1, C_2 > 0$ telles que

$$N_1 \leq C_1 N_2 \quad , \quad N_2 \leq C_2 N_1$$

alors en notant

$$B_{N_1}(a, r) = \{x \in E : N_1(x - a) < r\} \quad , \quad B_{N_2}(a, r) = \{x \in E : N_2(x - a) < r\},$$

on a

$$B_{N_1}(a, r) \subset B_{N_2}(a, C_2 r) \quad , \quad B_{N_2}(a, r) \subset B_{N_1}(a, C_1 r).$$

Définition 12 1. Une partie $A \subset E$ est dite bornée si elle est contenue dans une boule (ouverte ou fermée) : il existe $a \in E$ et $r > 0$ tel que $A \subset B(a, r)$.

2. Une suite d'éléments de E est dite bornée si tous ses termes sont contenus dans une partie bornée de E .

Proposition 10 Une suite de E convergente est bornée.

Preuve : Soit $(x_n)_{n \in \mathbb{N}}$ une suite convergente. Notons $\ell \in E$ sa limite. Alors il existe $n_0 \in \mathbb{N}$ tel que pour tout $n \geq n_0$, $\|x_n - \ell\| \leq 1$. On a alors $\|x_n\| \leq \|x_n - \ell\| + \|\ell\| \leq 1 + \|\ell\|$. Notons $M := 1 + \|\ell\| + \max_{n < n_0} \|x_n\|$. En distinguant les cas $n < n_0$ et $n \geq n_0$, on vérifie que pour tout $n \in \mathbb{N}$,

$$\|x_n\| \leq M.$$

Ainsi, la suite $(x_n)_{n \in \mathbb{N}}$ est contenue dans la boule $\overline{B(0, M)}$. Elle est donc bornée. □

1.2.2 Ouverts et fermés

Mathématiquement, pour traduire rigoureusement le fait qu'un ensemble ne contient pas son bord, on procède ainsi :

Définition 13 On dit qu'un ensemble $\Omega \subset E$ est un ouvert si Ω est une réunion de boules ouvertes.

En particulier, une boule ouverte est un ouvert !

Remarque 5 La réunion de deux ouverts (ou même d'une infinité d'ouverts) est un ouvert.

Exercice 3 Dans l'espace vectoriel \mathbb{R} muni de la norme donnée par la valeur absolue, montrer qu'un intervalle ouvert $]a, b[$ est un ouvert.

Proposition 11 Soit $\Omega \subset E$. Alors Ω est un ouvert si et seulement si pour tout $x \in \Omega$, il existe $r_x > 0$ tel que $B(x, r_x) \subset \Omega$.

Preuve : Supposons que Ω soit un ouvert. Soit $x \in \Omega$. Comme Ω est une réunion de boules, il en existe une, de centre y et de rayon r , qui contient x : $x \in B(y, r)$. Posons alors $r_x := r - |x - y|$. On a bien $r_x > 0$ car $x \in B(y, r)$. De plus, si $z \in E$ vérifie $|x - z| < r_x$, alors par l'inégalité triangulaire,

$$|z - y| \leq |z - x| + |x - y| < r_x + |x - y| = r.$$

On en déduit $z \in B(y, r)$ et donc $z \in \Omega$. Ainsi, $B(x, r_x) \subset \Omega$.

Réciproquement, si pour tout $x \in \Omega$, il existe $r_x > 0$ tel que $B(x, r_x) \subset \Omega$, montrons qu' Ω est un ouvert. Il suffit d'observer que

$$\Omega = \cup_{x \in \Omega} B(x, r_x).$$

En effet, l'inclusion \subset provient du fait que chaque $x \in \Omega$ appartient à $B(x, r_x)$ et donc appartient à la réunion $\cup_{x \in \Omega} B(x, r_x)$. L'inclusion \supset est conséquence du fait que $B(x, r_x) \subset \Omega$ pour chaque $x \in \Omega$. L'égalité est démontrée et Ω est donc bien une réunion de boules ouvertes. □

La notion d'ouvert ne dépend pas de la norme équivalente qu'on met sur l'espace :

Proposition 12 On considère deux normes N_1 et N_2 sur un espace vectoriel E et on suppose qu'elles sont équivalentes. Soit $\Omega \subset E$. Si Ω est un ouvert dans (E, N_1) , alors Ω est un ouvert de (E, N_2) .

Preuve : Soit Ω un ouvert de (E, N_1) . Soit $x \in \Omega$. Comme Ω est un ouvert pour N_1 , il existe $r_x > 0$ tel que

$$B_{N_1}(x, r_x) = \{y \in E : N_1(y - x) < r_x\} \subset \Omega.$$

Comme N_1 et N_2 sont équivalentes, il existe $C > 0$ tel que $N_1 \leq CN_2$. Par la remarque 4, on en déduit que

$$B_{N_2}(x, \frac{r_x}{C}) \subset B_{N_1}(x, r_x) \subset \Omega.$$

Donc Ω est un ouvert de (E, N_2) . □

Exemple 14 *Un demi-plan ouvert dans \mathbb{R}^2 (et plus généralement un demi-espace dans \mathbb{R}^n) est un ouvert.*

On considère un demi-plan ouvert de \mathbb{R}^2 : une droite de \mathbb{R}^2 sépare un plan en deux parties ne contenant pas cette droite. C'est ce qu'on appelle les demi-plans ouverts associés à cette droite. On n'a pas besoin de norme pour définir cette notion.

Montrons qu'un tel demi-plan est ouvert (pour n'importe quelle norme, puisqu'on a vu que la notion d'ouvert ne dépendait pas de la norme équivalente qu'on choisissait et que toutes les normes sont équivalentes dans \mathbb{R}^2). On va prendre dans la suite la norme $\|\cdot\|_2$. On se donne donc une droite Δ et on note Π l'un des demi-plans définis par cette droite. Soit $x \in \Pi$. On note y la projection orthogonale de x sur Δ . Comme $x \notin \Delta$, $y \neq x$ et donc $\|x - y\|_2 > 0$. Montrons que

$$B_{\|\cdot\|_2}(x, \|x - y\|_2) \subset \Pi.$$

Soit $z \notin \Pi$. La demi-droite de sommet x passant par z coupe Δ en z' . Alors

$$\|x - z\|_2^2 \geq \|x - z'\|_2^2 = \|x - y\|_2^2 + \|y - z'\|_2^2.$$

La dernière égalité résulte du théorème de Pythagore. On en déduit

$$\|x - z\|_2 \geq \|x - y\|_2,$$

et donc $z \notin B_{\|\cdot\|_2}(x, \|x - y\|_2)$. Par contraposée, on a montré que pour tout $z \in B_{\|\cdot\|_2}(x, \|x - y\|_2)$, on a $z \in \Pi$. Conclusion : $B_{\|\cdot\|_2}(x, \|x - y\|_2) \subset \Pi$. On en déduit que Π est un ouvert.

Définition 14 *Soit $A \subset E$. L'intérieur de A est la réunion des boules ouvertes contenues dans A .*

L'intérieur de A est donc aussi le plus grand (au sens de l'inclusion) ouvert contenu dans A .

Définition 15 *Soit $V \subset E$ et $x \in V$. On dit que V est un voisinage de x si x est dans l'intérieur de V .*

Autrement dit, il existe une boule ouverte contenant x et contenue dans V .

Définition 16 *Une partie $F \subset E$ est dite fermée si son complémentaire $F^c = E \setminus F$ est ouvert.*

Comme pour les ouverts, cette notion est invariante par changement de norme équivalente.

Exemple 15 *1. une boule fermée dans un e.v.n (E, N) ,
2. une droite dans \mathbb{R}^2 .*

1. Soit $\overline{B(a, r)}$ une boule fermée dans un e.v.n. (E, N) . Montrons que c'est un fermé, c'est-à-dire que son complémentaire est un ouvert. Soit $x \notin \overline{B(a, r)}$. Donc $N(x - a) > r$. Soit $r_x = N(x - a) - r$ et montrons que $B(x, r_x) \subset E \setminus \overline{B(a, r)}$. Soit $y \in B(x, r_x)$. Alors

$$N(y - a) \geq N(x - a) - N(y - x) > N(x - a) - r_x = r.$$

Donc $y \notin \overline{B(a, r)}$. Donc $B(x, r_x) \subset E \setminus \overline{B(a, r)}$. Donc $E \setminus \overline{B(a, r)}$ est un ouvert, ce qui montre que $\overline{B(a, r)}$ est un fermé.

2. Soit Δ une droite de \mathbb{R}^2 . Le complémentaire de cette droite est constitué de deux demi-plans ouverts, dont on a montré qu'ils étaient ouverts. Comme la réunion de deux ouverts est un ouvert, on en déduit que le complémentaire d'une droite est un ouvert. Donc une droite est un fermé.

Définition 17 Soit $A \subset E$.

1. La frontière de A , notée ∂A , est l'ensemble des points x de E tels que pour tout $r > 0$,

$$B(x, r) \cap A \neq \emptyset \quad , \quad B(x, r) \setminus A \neq \emptyset.$$

2. L'adhérence de A , notée \overline{A} , est la réunion de A et de sa frontière : $\overline{A} = \partial A \cup A$.

Comme d'habitude, ces notions ne dépendent pas de la norme équivalente choisie. Tous les points de A sont adhérents à A mais un point adhérent à A n'est pas forcément dans A . Par exemple, dans \mathbb{R} , les bornes d'un intervalle ouvert sont adhérentes à cet intervalle et pourtant n'appartiennent pas à cet intervalle.

Proposition 13 Un point $a \in A$ est dans l'adhérence de A si et seulement si pour tout $r > 0$, $B(a, r) \cap A \neq \emptyset$.

Preuve : Soit $a \in \overline{A}$. Soit $r > 0$. Montrons que $B(a, r) \cap A \neq \emptyset$. Par définition, $\overline{A} = \partial A \cup A$. On distingue alors deux cas. Si $a \in A$, alors $a \in B(a, r) \cap A$ et le résultat est vrai. Sinon, $a \in \partial A$ et par définition de ∂A , le résultat est encore vrai.

Réciproquement, supposons que pour tout $r > 0$, $B(a, r) \cap A \neq \emptyset$. Montrons que $a \in \overline{A}$. Si $a \in A$, la conclusion attendue est vraie, puisque $A \subset \overline{A}$. Sinon, $a \notin A$ et donc pour tout $r > 0$, $a \in B(a, r) \setminus A$. Ainsi, $a \in \partial A \subset \overline{A}$. Dans les deux cas, on a bien $a \in \overline{A}$.

□

Exercice 4 Soit $x \in E$ et $r > 0$. Alors

1. la sphère $S(x, r)$ est la frontière de la boule ouverte $B(x, r)$ et aussi la frontière de la boule fermée $\overline{B(x, r)}$.
2. la boule fermée $\overline{B(x, r)}$ est l'adhérence de la boule ouverte $B(x, r)$,
3. la boule ouverte $B(x, r)$ est l'intérieur de la boule fermée $\overline{B(x, r)}$.

1.3 Continuité

Soient (E, N) et (F, N') deux espaces vectoriels normés. Ainsi, N est une norme sur E et N' est une norme sur F . On s'intéresse dans cette section à des fonctions définies sur E (ou plus généralement sur une partie A de E) à valeurs dans F .

Soit donc $A \subset E$ et $f : A \rightarrow F$. Ainsi, à tout vecteur $x \in A$, on associe un (unique) vecteur noté $f(x)$ dans F .

1.3.1 Limite

Définition 18 Soit $a \in \overline{A}$. On dit que f tend vers $\ell \in F$ en a si pour tout $\varepsilon > 0$, il existe $\eta > 0$ tel que pour tout $y \in A$,

$$N(y - x) < \eta \implies N'(f(y) - \ell) < \varepsilon.$$

Cette définition prolonge la définition de la limite pour les fonctions définies sur un intervalle de \mathbb{R} et à valeurs dans \mathbb{R} (simplement, les valeurs absolues sont remplacées par des normes). Lorsque l'espace d'arrivée est \mathbb{R} , on définit comme avant la convergence vers $+\infty$ ou $-\infty$.

Remarquer que comme x est adhérent à A , il existe toujours des vecteurs $y \in A \cap B(x, \eta)$. Par ailleurs, ℓ n'est pas forcément égal à $f(x)$, puisque x n'est pas nécessairement dans A et que f n'est définie que sur A .

Toutes les propriétés vues sur les limites pour les fonctions de \mathbb{R} dans \mathbb{R} restent vraies ici, notamment l'unicité de la limite, et les limites d'une somme, d'un produit par un scalaire, d'une composée de fonctions.

1.3.2 Continuité

Définition 19 On dit qu'une fonction $f : A \rightarrow F$ est continue en un point $x \in A$ si pour tout $\varepsilon > 0$, il existe $\eta > 0$ tel que pour tout $y \in A$,

$$N(y - x) < \eta \implies N'(f(y) - f(x)) < \varepsilon.$$

Avec des symboles mathématiques, cela devient : f est continue en x si $\forall \varepsilon > 0, \exists \eta > 0$ tel que

$$f^{-1}(B_{N'}(f(x), \varepsilon)) \supset B_N(x, \eta) \cap A.$$

Ici, $B_N(x, \eta)$ désigne la boule ouverte de centre x et de rayon η dans E :

$$B_N(x, \eta) = \{y \in E : N(x - y) < \eta\}.$$

De même, $B_{N'}(f(x), \varepsilon)$ désigne la boule ouverte de centre $f(x)$ et de rayon ε dans F .

Intuitivement, cela signifie que l'on peut rendre $f(y)$ aussi proche que l'on veut de $f(x)$ pourvu que l'on prenne y suffisamment proche de x . Pour signifier que f est continue en x , on dit aussi que f tend vers $f(x)$ en x et on note $\lim_{y \rightarrow x} f(y) = f(x)$.

Exercice 5 Soit $f : A \subset E \rightarrow \mathbb{R}$ une fonction continue (ici, comme d'habitude, E désigne un espace vectoriel muni d'une norme et \mathbb{R} est muni de la valeur absolue). Alors

1. l'ensemble $\{x \in A : f(x) > 0\}$ est un ouvert,
2. les ensembles $\{x \in A : f(x) = 0\}$ et $\{x \in A : f(x) \geq 0\}$ sont des fermés.

La définition 19 généralise la définition de continuité pour les fonctions définies sur un intervalle de \mathbb{R} et à valeurs dans \mathbb{R} .

Proposition 14 La continuité d'une fonction en un point est invariante par changement de norme équivalente.

Cela signifie que si on prend une autre norme N_1 sur E , équivalente à N , et une autre norme N'_1 sur F , équivalente à N' , alors toute fonction continue lorsqu'on prend les normes N et N' sera continue lorsqu'on prend les normes N_1 et N'_1 .

Preuve : Soit une fonction $f : A \subset E \rightarrow F$ continue en $x \in A$ pour les normes N sur E et N' sur F : pour tout $\varepsilon > 0$, il existe $\eta > 0$ tel que pour tout $y \in A$,

$$N(y - x) < \eta \implies N'(f(y) - f(x)) < \varepsilon.$$

On prétend qu'alors f est continue en x pour les normes N_1 et N'_1 . En effet, soit $\varepsilon > 0$. Comme N' et N'_1 sont équivalentes dans F , il existe $C > 0$ tel que pour tout $z \in F$, $N'_1(z) \leq CN'(z)$. On applique alors la définition de continuité pour les normes N et N' avec ε/C : on en déduit l'existence de $\eta > 0$ tel que pour tout $y \in A$,

$$N(y - x) < \eta \implies N'(f(y) - f(x)) < \varepsilon/C. \quad (1.2)$$

Maintenant, on utilise que les normes N et N_1 sont équivalentes dans E : en particulier, il existe $C' > 0$ tel que pour tout $y \in E$, $N(y) \leq C'N_1(y)$. Ainsi, si $y \in A$ vérifie $N_1(y - x) < \eta/C'$, alors $N(y - x) < \eta$. Donc par (1.2), $N'(f(y) - f(x)) < \varepsilon/C$. On en déduit

$$N'_1(f(y) - f(x)) \leq CN'(f(y) - f(x)) < \varepsilon.$$

On a bien montré que pour tout $\varepsilon > 0$, il existe $\eta' > 0$ (dans ce qui précède, c'est η/C') tel que pour tout $y \in A$ vérifiant $N_1(y - x) < \eta'$, on a $N'_1(f(y) - f(x)) < \varepsilon$. Cela démontre la continuité de f pour les normes N_1 et N'_1 . □

Comme pour les fonctions d'une seule variable, la continuité est préservée par les opérations usuelles :

- Proposition 15**
- 1) Soient $f, g : A \rightarrow F$ deux fonctions continues en un vecteur $x \in A$. Alors $f + g$ est continue en x .
 - 2) Soient $f : A \rightarrow F$ une fonction continue en un vecteur $x \in A$ et $\lambda \in \mathbb{R}$. Alors λf est continue en x .
 - 3) Soit $f : A \rightarrow F$ une fonction continue en un point $x \in A$. Soient (G, N'') un autre espace vectoriel normé, B une partie de F contenant $f(A)$ et $g : B \rightarrow G$ une fonction continue en $f(x)$. Alors $g \circ f$ est continue en x .

Preuve : Montrons d'abord le point 1). Soit $\varepsilon > 0$. Comme f est continue en x , il existe $\eta > 0$ tel que pour tout $y \in A \cap B_N(x, \eta)$, $N'(f(y) - f(x)) < \varepsilon/2$. De même, comme g est continue en x , il existe $\eta' > 0$ tel que pour tout $y \in A \cap B_N(x, \eta')$, $N'(g(y) - g(x)) < \varepsilon/2$. On en déduit en posant $\eta'' = \min(\eta, \eta')$ que pour tout $y \in A \cap B_N(x, \eta'')$,

$$\begin{aligned} N'((f + g)(y) - (f + g)(x)) &= N'((f(y) - f(x)) + (g(y) - g(x))) \\ &\leq N'(f(y) - f(x)) + N'(g(y) - g(x)) \\ &< \frac{\varepsilon}{2} + \frac{\varepsilon}{2} = \varepsilon, \end{aligned}$$

ce qui montre la continuité en x .

On laisse le point 2) en exercice et montrons à présent le point 3). Soit $\varepsilon > 0$. Par continuité de g , il existe $\eta > 0$ tel que

$$g^{-1}(B_{N''}(g(f(x)), \varepsilon)) \supset (B_{N'}(f(x), \eta) \cap B).$$

Par continuité de f , il existe $\eta' > 0$ tel que

$$f^{-1}(B_{N'}(f(x), \eta)) \supset (B_N(x, \eta') \cap A).$$

On en déduit que

$$(g \circ f)^{-1}(B_{N''}(g \circ f(x), \varepsilon)) = f^{-1}(g^{-1}(B_{N''}(g(f(x)), \varepsilon))) \supset f^{-1}(B_{N'}(f(x), \eta) \cap B).$$

Comme $f(A) \subset B$, $f^{-1}(B_{N'}(f(x), \eta) \cap B) = f^{-1}(B_{N'}(f(x), \eta))$. Donc

$$f^{-1}(B_{N'}(f(x), \eta) \cap B) \supset (B_N(x, \eta') \cap A).$$

On a donc montré que

$$(g \circ f)^{-1}(B_{N''}(g \circ f(x), \varepsilon)) \supset (B_N(x, \eta') \cap A).$$

C'est bien la définition de la continuité de $g \circ f$ en x . □

Dans l'énoncé suivant, on suppose que les fonctions sont à valeurs réelles, autrement dit $F = \mathbb{R}$ et N' est la valeur absolue sur \mathbb{R} .

Proposition 16 1) *Le produit de deux fonctions $f, g : A \subset E \rightarrow \mathbb{R}$ continues en un vecteur $x \in A$ est continue en x .*
 2) *Soit $f : A \rightarrow \mathbb{R}$ une fonction qui ne s'annule pas sur A et qui est continue en un vecteur $x \in A$. Alors $1/f$ est continue en x .*

Définition 20 (Continuité sur un ensemble) *On dit que $f : A \subset E \rightarrow F$ est continue sur A quand f est continue en tout point $x \in A$.*

Exemple 16 *L'application suivante*

$$f : (x, y) \in \mathbb{R}^2 \mapsto e^{x^2+4y^5} \in \mathbb{R}$$

est continue sur \mathbb{R}^2 .

Théorème 6 *Toute fonction réelle et continue sur une partie fermée et bornée est bornée et atteint ses bornes.*

Cela signifie que si $f : A \subset E \rightarrow \mathbb{R}$ est continue sur la partie A de E supposée fermée et bornée, alors il existe $M > 0$ tel que pour tout $x \in A$, on a $|f(x)| \leq M$. De plus, il existe $x_- \in A$ et $x_+ \in A$ tels que

$$\forall x \in A, \quad f(x_-) \leq f(x) \leq f(x_+).$$

La preuve de ce théorème repose sur la notion de compacité évoquée à la fin de ce chapitre.

1.3.3 Continuité en dimension finie

On s'intéresse dans la suite à des fonctions définies sur une partie D de \mathbb{R}^n et à valeurs dans \mathbb{R}^p . Soit $x \in D$. On note (x_1, \dots, x_n) ses coordonnées. L'image $f(x)$ est un élément de \mathbb{R}^p . Ses coordonnées dépendent donc de x . On note

$$f(x_1, \dots, x_n) = (f_1(x_1, \dots, x_n), \dots, f_p(x_1, \dots, x_n)).$$

On adopte aussi une notation matricielle :

$$f(x_1, \dots, x_n) = \begin{pmatrix} f_1(x_1, \dots, x_n) \\ \vdots \\ f_p(x_1, \dots, x_n) \end{pmatrix}.$$

Les f_i sont les fonctions composantes de f . Elles sont définies sur $D \subset \mathbb{R}^n$ et à valeurs dans \mathbb{R} .

Remarque 6 *Lorsque $n = 1$ et $p = 1$, on a l'habitude de représenter une fonction par son graphe. On peut faire de même lorsque $n = 2$ et $p = 1$ dans un cadre tridimensionnel. Lorsque $n = 1$ et $p = 2$ ou $p = 3$, on peut encore dessiner des courbes paramétrées dans \mathbb{R}^2 et \mathbb{R}^3 (dans ce cas, on ne dessine pas la variable x).*

Dans toute la suite de cette section, D désigne une partie de \mathbb{R}^n et f une fonction définie sur D à valeurs dans \mathbb{R}^p .

Proposition 17 Une fonction $f : D \subset \mathbb{R}^n \rightarrow \mathbb{R}^p$ est continue en un point $x \in D$ si et seulement si toutes ses composantes $f_j : D \subset \mathbb{R}^n \rightarrow \mathbb{R}$ sont continues en x .

Preuve : Comme la proposition ne dépend pas des normes équivalentes choisies, on peut faire intervenir celles qui nous conviennent le mieux. Ici, on ne précisera pas la norme sur \mathbb{R}^n mais on choisit la norme $\|\cdot\|_\infty$ sur \mathbb{R}^p . Supposons que f soit continue en x : pour tout $\varepsilon > 0$, il existe $\eta > 0$ tel que pour tout $y \in D$, si $\|y - x\| < \eta$, alors $\max_{1 \leq j \leq p} |f_j(y) - f_j(x)| < \varepsilon$, d'où l'on tire : pour tout $1 \leq j \leq p$,

$$|f_j(y) - f_j(x)| < \varepsilon.$$

Ainsi pour tout $1 \leq j \leq p$, f_j est continue en x . Réciproquement, si pour tout $1 \leq j \leq p$, f_j est continue en x , alors pour tout $\varepsilon > 0$, pour tout $1 \leq j \leq p$, il existe $\eta_j > 0$ tel que pour tout $y \in D$, si $\|y - x\| < \eta_j$, alors $|f_j(y) - f_j(x)| < \varepsilon$. Posons $\eta = \min_{1 \leq j \leq p} \eta_j$. Ainsi pour tout x , si $\|y - x\| < \eta$, alors pour tout $1 \leq j \leq p$, $|f_j(y) - f_j(x)| < \varepsilon$, et donc $\|f(y) - f(x)\|_\infty < \varepsilon$. Ceci montre la seconde implication. □

Proposition 18 Soit $\Phi : \mathbb{R}^n \rightarrow \mathbb{R}^p$ une application linéaire. Alors Φ est continue.

Preuve : Pour tout $x, y \in \mathbb{R}^n$,

$$\Phi(x - y) = \Phi \left(\sum_{i=1}^n (x_i - y_i) e_i \right) = \sum_{i=1}^n (x_i - y_i) \Phi(e_i).$$

Donc

$$\|\Phi(x - y)\| \leq \sum_{i=1}^n |x_i - y_i| \|\Phi(e_i)\| \leq M \|x - y\|_1$$

en posant $M = \max \|\Phi(e_i)\|$. On en déduit que pour tout $\varepsilon > 0$, si $\|x - y\|_1 \leq \varepsilon/M$, alors

$$\|\Phi(x) - \Phi(y)\| = \|\Phi(x - y)\| \leq \varepsilon.$$

Donc Φ est continue. □

Exercice 6 Montrer que toute application linéaire $f : \mathbb{R}^n \rightarrow \mathbb{R}^p$ est lipschitzienne, c'est-à-dire qu'il existe $C > 0$ tel que pour tout $x, y \in \mathbb{R}^n$, $\|f(x) - f(y)\| \leq C \|x - y\|$ (la constante C dépend de f et du choix des normes sur \mathbb{R}^n et \mathbb{R}^p).

Exemple 17 On note $p_j : x = (x_1, \dots, x_n) \in \mathbb{R}^n \mapsto x_j \in \mathbb{R}$ la projection sur la $j^{\text{ème}}$ coordonnée de x . Alors p_j est une application continue sur \mathbb{R}^n .

Preuve : L'application p_j est linéaire. Par la proposition précédente, on en déduit que p_j est continue. □

Proposition 19 Soit $f : D \subset \mathbb{R}^n \rightarrow \mathbb{R}^p$ une fonction continue en $a = (a_1, \dots, a_n)$. Alors toutes les restrictions de f par rapport à 1, 2, ..., ou $n - 1$ variables sont continues.

Preuve : On montre que la restriction par rapport à la première variable $x_1 \in \mathbb{R} \mapsto f(x_1, a_2, \dots, a_n)$ est continue (les autres assertions se démontrent de manière analogue). Il s'agit de la composée de $x_1 \mapsto (x_1, a_2, \dots, a_n)$ qui est continue en a_1 et de f qui est continue en a . On conclut grâce à la proposition sur la continuité des composées. □

Remarque 7 Attention : la réciproque est fautive. Par exemple pour $f : \mathbb{R}^2 \rightarrow \mathbb{R}$, ce n'est pas parce que $x \mapsto f(x, a_2)$ et $y \mapsto f(a_1, y)$ sont continues que f est continue en x .

Exemple 18

$$f(x, y) = \begin{cases} \frac{xy}{x^2+y^2} & \text{si } (x, y) \neq (0, 0), \\ 0 & \text{si } (x, y) = (0, 0). \end{cases}$$

Les fonctions $f(0, \cdot)$ et $f(\cdot, 0)$ sont constamment égales à 0 et en particulier continues. Pourtant, pour tout $y \neq 0$, $f(y, y) = \frac{1}{2}$. En particulier, $|f(0, 0) - f(y, y)| \geq \frac{1}{2}$ et donc f n'est pas continu en $(0, 0)$.

Exemple 19

$$g(x, y) = \begin{cases} \frac{x^2 e^x + y^2}{x^2 + y^2} & \text{si } (x, y) \neq (0, 0), \\ 1 & \text{si } (x, y) = (0, 0). \end{cases}$$

Sur $\mathbb{R}^2 \setminus \{(0, 0)\}$, la fonction g est continue comme quotient de fonctions continues, le dénominateur ne s'annulant pas. Montrons la continuité de g en $(0, 0)$. Pour tout $(x, y) \neq (0, 0)$,

$$|g(x, y) - 1| = \frac{|e^x - 1|x^2}{x^2 + y^2} \leq |e^x - 1|.$$

Soit $\varepsilon > 0$. Par continuité de $x \mapsto e^x$ en 0, il existe $\eta > 0$ tel que pour tout $|x| < \eta$, on a $|e^x - 1| < \varepsilon$. Donc pour tout $\|(x, y)\|_\infty < \eta$, $|g(x, y) - 1| < \varepsilon$. Ainsi, g est bien continue en $(0, 0)$.

1.4 Compacité

Soit $n \geq 1$. On munit \mathbb{R}^n d'une norme (elles sont toutes équivalentes).

1.4.1 Critères séquentiels

Une partie de \mathbb{R}^n est dite fermée lorsque son complémentaire est ouvert (intuitivement, un fermé est donc un ensemble qui contient son bord). On dispose d'une caractérisation séquentielle (i.e. faisant intervenir les suites) des fermés :

Proposition 20 Une partie $F \subset \mathbb{R}^n$ est fermée si toute suite d'éléments de F qui converge dans \mathbb{R}^n a sa limite dans F .

Il existe aussi une caractérisation séquentielle des fonctions continues :

Proposition 21 Soit $D \subset \mathbb{R}^n$, $f : D \rightarrow \mathbb{R}^p$ et $x_0 \in D$. Alors f est continue en x_0 si et seulement si pour toute suite $(x_i)_{i \in \mathbb{N}}$ contenue dans D , la suite $(f(x_i))_{i \in \mathbb{N}}$ converge vers $f(x_0)$.

1.4.2 Compacité

On en vient alors à la seule définition nouvelle de cette section :

Définition 21 Soit $K \subset \mathbb{R}^n$. On dit que K est compact si et seulement si de toute suite d'éléments de $K : (x_i)_{i \in \mathbb{N}} \subset K$, on peut extraire une sous-suite qui converge dans K : il existe une fonction strictement croissante $\varphi : \mathbb{N} \rightarrow \mathbb{N}$ telle que la sous-suite $(x_{\varphi(i)})_{i \in \mathbb{N}}$ converge, et sa limite appartient à K .

On a en fait une caractérisation plus simple pour les compacts de \mathbb{R}^n :

Proposition 22 Une partie K de \mathbb{R}^n est compacte si et seulement si elle est fermée et bornée.

Tout ce précède reste vrai lorsqu'on remplace \mathbb{R}^n par un espace vectoriel normé plus général, à l'exception (notable !) de la proposition 22 qui n'est plus vraie lorsque l'espace vectoriel est de dimension infinie. Dans ce cas, seule une implication est vraie : une partie compacte est toujours fermée et bornée mais la réciproque n'a pas systématiquement lieu.

Exercice 7 ((Extrait du concours CCP 2010))

1. Rappeler la définition (par les suites) d'une partie compacte d'un espace vectoriel normé.
2. Soit E et F deux espaces vectoriels normés, et f une application continue de E dans F . Si A est une partie compacte de E , démontrer que $f(A)$ est une partie compacte de F . L'image réciproque par f d'une partie compacte de F est-elle nécessairement une partie compacte de E ?

Solution :

1. Une partie d'un espace vectoriel normé est compacte si de toute suite de cette partie, on peut extraire une sous-suite qui converge dans cette partie.
2. Soit $(y_i)_{i \in \mathbb{N}}$ une suite d'éléments de $f(A)$. Alors pour tout $i \in \mathbb{N}$, il existe $x_i \in A$ tel que $f(x_i) = y_i$. Comme A est compact, on peut extraire de la suite $(x_i)_{i \in \mathbb{N}}$ une sous-suite convergente dans A : il existe une fonction strictement croissante $\varphi : \mathbb{N} \rightarrow \mathbb{N}$ telle que la sous-suite $(x_{\varphi(i)})_{i \in \mathbb{N}}$ converge vers un vecteur $x \in A$. Par continuité de f , la suite $(f(x_{\varphi(i)}))_{i \in \mathbb{N}}$ converge vers $f(x)$. Autrement dit, la suite $(y_{\varphi(i)})_{i \in \mathbb{N}}$ converge vers $f(x) \in f(A)$. Conclusion : $f(A)$ est compact.

Considérons la fonction nulle $f : x \in \mathbb{R} \mapsto 0 \in \mathbb{R}$. On a $f(\mathbb{R}) = \{0\}$, $\{0\}$ est un compact de \mathbb{R} et pourtant $f^{-1}(\{0\}) = \mathbb{R}$ n'est pas un compact de \mathbb{R} , par exemple parce qu'il n'est pas borné. Conclusion : en général, l'image réciproque d'un compact par une application continue n'est pas compacte.

Chapitre 2

Séries de Fourier

Dans ce chapitre, toutes les fonctions considérées sont à valeurs dans \mathbb{C} (ce qui n'exclut pas les fonctions qui ne prennent que des valeurs réelles).

2.1 Introduction

2.1.1 Un détour par les séries entières

Lorsqu'une fonction $f : \mathbb{R} \rightarrow \mathbb{R}$ est développable en série entière, cela signifie qu'on peut l'écrire sous la forme

$$f(x) = \sum_{n=0}^{+\infty} a_n x^n,$$

où $(a_n)_{n \in \mathbb{N}}$ est une suite de réels. En particulier, pour tout $x \in \mathbb{R}$, la série $\sum_n a_n x^n$ converge. En posant pour tout $n \in \mathbb{N}$, $f_n : x \in \mathbb{R} \mapsto a_n x^n$, on dit que la série de fonctions $\sum_n f_n$ converge simplement vers f , autrement dit

$$\forall x \in \mathbb{R}, \quad f(x) = \lim_{N \rightarrow +\infty} \sum_{n=0}^N f_n(x).$$

En fait, la série converge normalement sur tout segment de \mathbb{R} , c'est-à-dire

$$\forall M > 0, \quad \sum_{n=0}^{+\infty} \sup_{x \in [-M, M]} |f_n(x)| < \infty.$$

Cela implique que la série de fonctions converge uniformément sur tout segment vers f :

$$\forall M > 0, \quad \lim_{N \rightarrow +\infty} \sup_{x \in [-M, M]} \left| \sum_{n=0}^N f_n(x) - f(x) \right| = 0.$$

On sait de plus que f est \mathcal{C}^∞ sur \mathbb{R} , qu'on peut dériver termes à termes (attention, cela n'est pas possible pour toutes les séries de fonctions) et

$$\forall n \in \mathbb{N}, \quad a_n = \frac{1}{n!} f^{(n)}(0).$$

Dans le cas où f n'est pas développable en série entière sur \mathbb{R} tout entier mais seulement sur un intervalle (symétrique par rapport à 0), alors tout ce qui précède reste vrai sur cet intervalle seulement.

Quoi qu'il en soit, le développement en série entière permet d'approcher une fonction f par une suite de polynômes $(\sum_{n=0}^N a_n x^n)_{N \in \mathbb{N}}$, uniformément sur tout segment de l'intervalle de convergence. Cela n'est possible que pour des fonctions f qui sont nécessairement \mathcal{C}^∞ (mais toutes les fonctions \mathcal{C}^∞ ne sont pas développables en séries entières).

2.1.2 Polynômes trigonométriques à coefficients réels

Considérons à présent une fonction *périodique* $f : \mathbb{R} \rightarrow \mathbb{R}$: on suppose donc qu'il existe $T > 0$ (appelée la *période*) telle que

$$\forall x \in \mathbb{R}, \quad f(x + T) = f(x).$$

Alors plutôt que de chercher à approcher f par des polynômes (qui ne sont pas eux-mêmes périodiques), il est naturel de construire une approximation à l'aide de fonctions périodiques. Pour fixer les idées, supposons que la période T de f vaut 2π . Les fonctions usuelles \cos et \sin étant elles-mêmes 2π périodiques, on peut tenter d'approcher f par des fonctions élémentaires construites à partir de \cos et \sin . On appelle *polynôme trigonométrique à coefficients réels* toute fonction de la forme

$$P : x \in \mathbb{R} \mapsto \sum_{n=0}^N \alpha_n \cos(nx) + \beta_n \sin(nx),$$

où $N \in \mathbb{N}$ et les α_n, β_n sont des réels. On observe que les fonctions $x \mapsto \cos(nx)$ et $x \mapsto \sin(nx)$ sont toutes 2π périodiques. Il en est donc de même de P .

Il est possible de retrouver les coefficients α_n et β_n lorsqu'on connaît la fonction P . Ainsi, par linéarité de l'intégrale,

$$\int_0^{2\pi} P(x) dx = \sum_{n=0}^N \alpha_n \int_0^{2\pi} \cos(nx) dx + \beta_n \int_0^{2\pi} \sin(nx) dx.$$

Comme $x \mapsto \cos(nx)$ et $x \mapsto \sin(nx)$ ont des primitives explicites, il est facile de voir que leur intégrale sur $[0, 2\pi]$ est nulle, sauf pour \cos lorsque $n = 0$:

$$\int_0^{2\pi} \cos(0x) dx = \int_0^{2\pi} 1 dx = 2\pi.$$

On trouve ainsi

$$\alpha_0 = \frac{1}{2\pi} \int_0^{2\pi} P(x) dx.$$

Pour obtenir les coefficients α_k et β_k , on multiplie d'abord P par $\cos(kx)$ ou $\sin(kx)$. Par exemple, lorsque $k \in \mathbb{N}^*$,

$$\int_0^{2\pi} P(x) \cos(kx) dx = \sum_{n=0}^N \alpha_n \int_0^{2\pi} \cos(nx) \cos(kx) dx + \beta_n \int_0^{2\pi} \sin(nx) \cos(kx) dx.$$

Pour calculer l'intégrale de la fonction $\cos(nx) \cos(kx)$, on cherche d'abord à la linéariser, c'est-à-dire à transformer le produit en une somme. Pour cela, il est commode de passer par l'exponentielle complexe :

$$\begin{aligned} \cos(nx) \cos(kx) &= \frac{e^{inx} + e^{-inx}}{2} \frac{e^{ikx} + e^{-ikx}}{2} = \frac{e^{i(n+k)x} + e^{-i(n+k)x} + e^{i(n-k)x} + e^{-i(n-k)x}}{4} \\ &= \frac{\cos(n+k)x + \cos(n-k)x}{2}. \end{aligned}$$

On a de même

$$\begin{aligned} \sin(nx) \cos(kx) &= \frac{e^{inx} - e^{-inx}}{2i} \frac{e^{ikx} + e^{-ikx}}{2} = \frac{e^{i(n+k)x} - e^{-i(n+k)x} + e^{i(n-k)x} - e^{-i(n-k)x}}{4i} \\ &= \frac{\sin(n+k)x + \sin(n-k)x}{2}. \end{aligned}$$

En intégrant sur $[0, 2\pi]$, il vient lorsque $k \neq n$

$$\int_0^{2\pi} \cos(nx) \cos(kx) dx = \frac{1}{2} \int_0^{2\pi} \cos(n+k)x + \cos(n-k)x dx = 0,$$

tandis que pour $k = n$,

$$\int_0^{2\pi} \cos(nx) \cos(kx) dx = \frac{1}{2} \int_0^{2\pi} \cos(n+k)x + \cos(n-k)x dx = \pi.$$

Par ailleurs, pour tout $n \in \mathbb{N}$,

$$\int_0^{2\pi} \sin(nx) \cos(kx) dx = \frac{1}{2} \int_0^{2\pi} \sin(n+k)x + \sin(n-k)x dx = 0.$$

En conclusion, lorsque $k \neq 0$

$$\int_0^{2\pi} P(x) \cos(kx) dx = \pi \alpha_k,$$

soit encore

$$\alpha_k = \frac{1}{\pi} \int_0^{2\pi} P(x) \cos(kx) dx.$$

On calcule de même pour tout $k \in \mathbb{N}$,

$$\beta_k = \frac{1}{\pi} \int_0^{2\pi} P(x) \sin(kx) dx.$$

Si on peut écrire une fonction 2π périodique continue (ou continue par morceaux) f sous la forme

$$f(x) = \sum_{n=0}^{+\infty} \alpha_n \cos(nx) + \beta_n \sin(nx)$$

(on suppose donc ici que la série converge pour tout x vers $f(x)$), on s'attend à ce que

$$\alpha_0 = \frac{1}{2\pi} \int_0^{2\pi} f(x) dx \quad \text{et } \forall n \in \mathbb{N}^*, \quad \alpha_n = \frac{1}{\pi} \int_0^{2\pi} f(x) \cos(nx) dx, \quad \beta_n = \frac{1}{\pi} \int_0^{2\pi} f(x) \sin(nx) dx.$$

On appelle *coefficients de Fourier de f* les nombres

$$\forall n \in \mathbb{N}, \quad a_n(f) = \frac{1}{\pi} \int_0^{2\pi} f(x) \cos(nx) dx, \quad b_n(f) = \frac{1}{\pi} \int_0^{2\pi} f(x) \sin(nx) dx.$$

Exercice 8 Soit $f : \mathbb{R} \rightarrow \mathbb{R}$ continue et 2π périodique.

1. Si f est paire, alors pour tout $n \in \mathbb{N}$, $b_n(f) = 0$.
2. Si f est impaire, alors pour tout $n \in \mathbb{N}$, $a_n(f) = 0$.

2.1.3 Polynômes trigonométriques à coefficients complexes

Comme on l'a vu dans la section précédente, il est souvent plus facile pour faire des calculs de passer par les nombres complexes. Aussi, dans tout ce chapitre, on va considérer des fonctions $f : \mathbb{R} \rightarrow \mathbb{C}$ à valeurs complexes. Lorsqu'une telle fonction est continue (ou continue par morceaux), on peut définir son intégrale en décomposant f en partie réelle et partie imaginaire : $f = \operatorname{Re} f + i \operatorname{Im} f$.

$$\int_0^{2\pi} f(x) dx = \int_0^{2\pi} \operatorname{Re} f(x) dx + i \int_0^{2\pi} \operatorname{Im} f(x) dx.$$

L'intégrale des fonctions complexes hérite de toutes les propriétés de l'intégrale des fonctions réelles (linéarité, relation de Chasles).

Considérons à nouveau un polynôme trigonométrique, cette fois à coefficients complexes :

$$P : x \in \mathbb{R} \mapsto \sum_{n=0}^N \alpha_n \cos(nx) + \beta_n \sin(nx),$$

où $N \in \mathbb{N}$ et les α_n, β_n sont maintenant des nombres complexes. En écrivant $\cos(nx) = (e^{inx} + e^{-inx})/2$ et $\sin(nx) = (e^{inx} - e^{-inx})/(2i)$, on obtient

$$\begin{aligned} P(x) &= \alpha_0 + \sum_{n>0} \alpha_n \frac{e^{inx} + e^{-inx}}{2} + \sum_{n>0} \beta_n \frac{e^{inx} - e^{-inx}}{2i} \\ &= \alpha_0 + \sum_{n>0} \left(\frac{\alpha_n}{2} + \frac{\beta_n}{2i} \right) e^{inx} + \sum_{n>0} \left(\frac{\alpha_n}{2} - \frac{\beta_n}{2i} \right) e^{-inx} \\ &= \alpha_0 + \sum_{n>0} \frac{\alpha_n - i\beta_n}{2} e^{inx} + \sum_{n<0} \frac{\alpha_{-n} + i\beta_{-n}}{2} e^{inx}. \end{aligned}$$

En utilisant l'expression de α_n et β_n en fonction de P obtenue dans la section précédente, il vient si $n > 0$,

$$\frac{\alpha_n - i\beta_n}{2} = \frac{1}{2\pi} \left(\int_0^{2\pi} P(x) \cos(nx) dx - i \int_0^{2\pi} P(x) \sin(nx) dx \right) = \frac{1}{2\pi} \int_0^{2\pi} P(x) e^{-inx} dx,$$

tandis que si $n < 0$,

$$\frac{\alpha_{-n} + i\beta_{-n}}{2} = \frac{1}{2\pi} \left(\int_0^{2\pi} P(x) \cos(-nx) dx + i \int_0^{2\pi} P(x) \sin(-nx) dx \right) = \frac{1}{2\pi} \int_0^{2\pi} P(x) e^{-inx} dx.$$

Ainsi,

$$P(x) = \sum_{n \in \mathbb{Z}} \gamma_n e^{inx}$$

en posant

$$\gamma_n = \frac{1}{2\pi} \int_0^{2\pi} f(x) e^{-inx} dx.$$

Plus généralement, pour une fonction $f : \mathbb{R} \rightarrow \mathbb{C}$ continue (ou continue par morceaux) et 2π périodique, on définit ses *coefficients de Fourier*

$$c_n(f) = \frac{1}{2\pi} \int_0^{2\pi} f(x) e^{-inx} dx.$$

Les calculs précédents montrent que

$$c_n(f) = \begin{cases} (a_n(f) - ib_n(f))/2 & \text{si } n > 0, \\ a_0(f)/2 & \text{si } n = 0, \\ (a_{-n}(f) + ib_{-n}(f))/2 & \text{si } n < 0. \end{cases}$$

On peut aussi exprimer les coefficients $(a_n(f))$ et $(b_n(f))$ avec les formules :

$$\forall n \in \mathbb{N}, \quad a_n(f) = c_n(f) + c_{-n}(f) \quad , \quad b_n(f) = i(c_n(f) - c_{-n}(f)).$$

Enfin, on définit la *série de Fourier* de f par

$$\forall N \in \mathbb{N}, \quad S_N(f)(x) = \frac{a_0(f)}{2} + \sum_{n=1}^N a_n(f) \cos(nx) + b_n(f) \sin(nx) = \sum_{n=-N}^N c_n(f) e^{inx}.$$

Remarque 8 La série $\sum_n c_n(f)e^{inx}$ est indexée sur \mathbb{Z} et non pas sur \mathbb{N} .

La théorie pour de telles séries se ramène à la théorie classique en disant qu'une série de fonctions $\sum_{k \in \mathbb{Z}} f_k$ converge si et seulement si les deux séries $\sum_{k \in \mathbb{N}} f_k$ et $\sum_{k \in \mathbb{N}} f_{-k}$ (qui sont deux séries indexées sur \mathbb{N}) convergent. Dans ce cas, on définit

$$\sum_{k=-\infty}^{+\infty} f_k = \sum_{k=0}^{+\infty} f_k + \sum_{k=1}^{+\infty} f_{-k} = \lim_{N \rightarrow +\infty} \sum_{k=0}^N f_k + \lim_{N \rightarrow +\infty} \sum_{k=1}^N f_{-k}.$$

Exercice 9 Montrer que la série de Fourier d'un polynôme trigonométrique P de la forme

$$P(x) = \sum_{n=0}^N \alpha_n \cos(nx) + \beta_n \sin(nx)$$

est donnée par

$$S_M(P)(x) = \begin{cases} \sum_{n=0}^M \alpha_n \cos(nx) + \beta_n \sin(nx) & \text{si } M \leq N, \\ \sum_{n=0}^N \alpha_n \cos(nx) + \beta_n \sin(nx) = P(x) & \text{si } M \geq N. \end{cases}$$

L'objet de ce chapitre est l'étude des propriétés de convergence des séries de Fourier. On note $C_{2\pi}^0(\mathbb{R}; \mathbb{C})$ l'ensemble des fonctions continues et 2π périodiques à valeurs complexes.

2.2 Convergence uniforme et convergence ponctuelle

2.2.1 Injectivité des coefficients de Fourier

Cette section est fondée sur le théorème suivant dont on donnera la preuve ultérieurement :

Théorème 7 Soit $f : \mathbb{R} \rightarrow \mathbb{R}$ une fonction continue 2π périodique. Si tous les coefficients de Fourier de f sont nuls, alors $f = 0$.

Corollaire 1 Si deux fonctions continues et 2π périodiques f et g ont les mêmes coefficients de Fourier, alors $f = g$.

Preuve : La fonction $h = f - g$ est continue et 2π périodique. De plus, pour tout $n \in \mathbb{Z}$,

$$\begin{aligned} c_n(h) &= \frac{1}{2\pi} \int_0^{2\pi} (f - g)(x)e^{-inx} dx = \frac{1}{2\pi} \int_0^{2\pi} f(x)e^{-inx} dx - \frac{1}{2\pi} \int_0^{2\pi} g(x)e^{-inx} dx \\ &= c_n(f) - c_n(g) = 0. \end{aligned}$$

On en déduit par le théorème précédent que $h = 0$, d'où la conclusion. □

2.2.2 La norme uniforme

On note $C_{2\pi}^0(\mathbb{R}; \mathbb{C})$ l'ensemble des fonctions continues 2π périodiques à valeurs dans \mathbb{C} . On peut introduire sur $C_{2\pi}^0(\mathbb{R}; \mathbb{C})$ la norme uniforme $\|\cdot\|_\infty$, où

$$\forall f \in C_{2\pi}^0(\mathbb{R}; \mathbb{C}), \quad \|f\|_\infty = \sup_{x \in \mathbb{R}} |f(x)|.$$

En effet, si une fonction f est continue 2π périodique, alors l'ensemble des valeurs qu'elle prend coïncide avec l'ensemble des valeurs qu'elle prend sur $[0, 2\pi]$ (on utilise ici la 2π périodicité) :

$$f(\mathbb{R}) = \{f(x); x \in \mathbb{R}\} = f([0, 2\pi]).$$

Sur le segment $[0, 2\pi]$, la fonction f est continue, donc est bornée et atteint ses bornes : il existe $c, d \in [0, 2\pi]$ tels que pour tout $x \in [0, 2\pi]$,

$$f(c) \leq f(x) \leq f(d).$$

Cette double inégalité reste vraie pour tout $x \in \mathbb{R}$. On peut donc définir

$$\|f\|_\infty = \max_{x \in [0, 2\pi]} |f(x)| = \max_{x \in \mathbb{R}} |f(x)|.$$

Rappelons quelques propriétés de la convergence uniforme :

1. Si une suite $(f_n)_{n \in \mathbb{N}}$ de fonctions continues sur \mathbb{R} converge uniformément sur \mathbb{R} vers une fonction $f : \mathbb{R} \rightarrow \mathbb{R}$, alors f est aussi continue. De plus, si chaque f_n est 2π périodique, alors f est aussi périodique (par passage à la limite dans l'égalité $f_n(x + 2\pi) = f_n(x)$, $x \in \mathbb{R}$).
2. La convergence normale d'une série implique la convergence uniforme.
3. Si une série converge uniformément, alors on peut intervertir somme et intégrale sur un segment.

2.2.3 Coefficients de Fourier sommables

Théorème 8 Soit $f \in C_{2\pi}^0(\mathbb{R}; \mathbb{C})$. On suppose que la famille des coefficients de Fourier est sommable :

$$\sup_{n \in \mathbb{N}} \sum_{\substack{k \in \mathbb{Z}, \\ |k| \leq n}} |c_k(f)| < +\infty.$$

Alors $(S_n(f))_{n \in \mathbb{N}}$ converge normalement vers f sur \mathbb{R} .

Preuve : On a pour tout $t \in \mathbb{R}$

$$S_n(f)(t) = \sum_{k=-n}^n c_k(f) e^{ikt}.$$

La convergence normale des sommes partielles signifie par définition que la série numérique

$$\sum_{k=-n}^n \|c_k(f) e^{ikt}\|_\infty$$

converge. Or

$$\|c_k(f) e^{ikt}\|_\infty = |c_k|.$$

Par hypothèse, on en déduit la convergence normale de $S_n(f)$, et donc la convergence uniforme de $S_n(f)$ vers une fonction qu'on note g .

$$\lim_{n \rightarrow +\infty} \|S_n(f) - g\|_\infty = 0.$$

Il reste à montrer que $f = g$. Comme chaque terme de la somme $c_k e_k$ est continue, on en déduit que g est également continue sur \mathbb{R} . Par convergence simple, la 2π périodicité de $S_n(f)$ entraîne celle de g .

Ainsi $g \in \mathcal{C}_{2\pi}^0(\mathbb{R}; \mathbb{C})$. Calculons les coefficients de Fourier de g . Par convergence uniforme de $S_n(f)$ vers g sur $[0, 2\pi]$, on a

$$c_k(g) = \frac{1}{2\pi} \int_0^{2\pi} e^{-ikt} \lim_{n \rightarrow +\infty} S_n(f)(t) dt = \frac{1}{2\pi} \lim_{n \rightarrow +\infty} \int_0^{2\pi} e^{-ikt} S_n(f)(t) dt.$$

Or,

$$\frac{1}{2\pi} \int_0^{2\pi} e^{-ikt} S_n(f)(t) dt = \sum_{\ell=-n}^n c_\ell(f) \frac{1}{2\pi} \int_0^{2\pi} e^{-ikt} e^{i\ell t} dt$$

et

$$\frac{1}{2\pi} \int_0^{2\pi} e^{-ikt} e^{i\ell t} dt = \begin{cases} 0 & \text{si } \ell \neq k, \\ 1 & \text{si } \ell = k. \end{cases}$$

Donc pour tout $n \geq k$,

$$\frac{1}{2\pi} \int_0^{2\pi} e^{-ikt} S_n(f)(t) dt = c_k(f),$$

ce qui implique que pour tout $k \in \mathbb{N}$, $c_k(g) = c_k(f)$. On en déduit que $f = g$. □

Corollaire 2 Soit $f \in \mathcal{C}_{2\pi}^2(\mathbb{R}; \mathbb{C})$. Alors $(S_n(f))_{n \in \mathbb{N}}$ converge normalement vers f sur \mathbb{R} .

Preuve : En utilisant la définition de f' comme limite d'un taux d'accroissement, on vérifie que f' est 2π périodique :

$$\forall x \in \mathbb{R}, \forall h \in \mathbb{R}^*, \quad \frac{f(x + 2\pi + h) - f(x + 2\pi)}{h} = \frac{f(x + h) - f(x)}{h}$$

d'où en passant à la limite en $h \rightarrow 0$, $f'(x + 2\pi) = f'(x)$.

On calcule par intégration par parties :

$$c_k(f') = \frac{1}{2\pi} \int_0^{2\pi} f'(t) e^{-ikt} dt = \frac{1}{2\pi} \left[f(t) e^{-ikt} \right]_0^{2\pi} - \frac{1}{2\pi} \int_0^{2\pi} (-ik) f(t) e^{-ikt} dt.$$

Par 2π périodicité de $t \mapsto f(t) e^{-ikt}$, on obtient $c_k(f') = ik c_k(f)$. Le même calcul avec f' au lieu de f donne $c_k(f'') = ik c_k(f') = -k^2 c_k(f)$. Ainsi, pour tout $k \in \mathbb{Z}^*$,

$$c_k(f) = \frac{-1}{k^2} c_k(f'').$$

Par ailleurs,

$$|c_k(f'')| = \frac{1}{2\pi} \left| \int_0^{2\pi} f''(x) e^{-ikx} dx \right| \leq \frac{1}{2\pi} \int_0^{2\pi} |f''(x) e^{-ikx}| dx \leq \|f''\|_\infty.$$

On en déduit que

$$|c_k(f)| \leq \frac{\|f''\|_\infty}{k^2}$$

et donc la série $\sum_k |c_k(f)|$ converge. Par le théorème précédent, la suite des sommes partielles $S_n(f)$ converge normalement vers f . □

2.2.4 L'espace $\mathcal{C}_{m,2\pi}^0(\mathbb{R}; \mathbb{C})$

Définition 22 Soit $[a, b]$ un segment de \mathbb{R} et $f : [a, b] \rightarrow \mathbb{C}$ une fonction.

1. On dit que f est continue par morceaux sur $[a, b]$ s'il existe une subdivision $a_0 = a < a_1 < \dots < a_m = b$ telle que pour tout $i = 0, \dots, m-1$, $f|_{]a_i, a_{i+1}[}$ se prolonge en une fonction continue à $[a_i, a_{i+1}]$, autrement dit : $f|_{]a_i, a_{i+1}[}$ est continue et les limites $\lim_{x \rightarrow a_i^+} f(x)$, $\lim_{x \rightarrow a_{i+1}^-} f(x)$ existent et sont finies.
2. On dit que f est \mathcal{C}^1 par morceaux sur $[a, b]$ s'il existe une subdivision $a_0 = a < a_1 < \dots < a_m = b$ telle que pour tout $i = 0, \dots, m-1$, $f|_{]a_i, a_{i+1}[}$ se prolonge en une fonction \mathcal{C}^1 à $[a_i, a_{i+1}]$, autrement dit : $f|_{]a_i, a_{i+1}[}$ est \mathcal{C}^1 et les limites

$$\lim_{x \rightarrow a_i^+} f(x), \quad \lim_{x \rightarrow a_{i+1}^-} f(x), \quad \lim_{x \rightarrow a_i^+} f'(x), \quad \lim_{x \rightarrow a_{i+1}^-} f'(x)$$

existent et sont finies.

Définition 23 Soit I un intervalle quelconque de \mathbb{R} et $f : I \rightarrow \mathbb{R}$ une fonction. On dit que f est continue (resp. \mathcal{C}^1) par morceaux sur I si la restriction de f à tout segment $[a, b] \subset I$ est continue (resp. \mathcal{C}^1) par morceaux sur $[a, b]$.

On note $\mathcal{C}_m^0(I; \mathbb{C})$ l'ensemble des fonctions continues par morceaux sur I à valeurs complexes, et $\mathcal{C}_m^1(I; \mathbb{C})$ l'ensemble des fonctions \mathcal{C}^1 par morceaux.

2.2.5 Le théorème de Dirichlet

Lorsque f est seulement continue par morceaux mais pas continue, $S_n(f)$ ne peut pas converger uniformément vers f , car une limite uniforme de fonctions continues est continue. Par contre, on dispose du théorème suivant (dont on admet la preuve) :

Théorème 9 (Théorème de Dirichlet) Soit $f \in \mathcal{C}_{m,2\pi}^1(\mathbb{R})$. Alors pour tout $x \in \mathbb{R}$,

$$\lim_{n \rightarrow +\infty} S_n(f)(x) = \frac{1}{2}(f(x^-) + f(x^+)),$$

où $f(x^-)$ (resp. $f(x^+)$) est la limite à gauche (resp. à droite) de f en x .

En particulier, si f est continue en $x \in \mathbb{R}$, alors $\lim_{n \rightarrow +\infty} S_n(f)(x) = f(x)$.

2.3 Convergence quadratique

2.3.1 La norme $\|\cdot\|_2$

Sur l'ensemble $\mathcal{C}_{2\pi}^0(\mathbb{R}; \mathbb{C})$ des fonctions continues 2π périodiques à valeurs dans \mathbb{C} , on introduit la norme

$$\|f\|_2 = \sqrt{\frac{1}{2\pi} \int_0^{2\pi} |f(t)|^2 dt}, \quad \forall f \in \mathcal{C}_{2\pi}^0(\mathbb{R}; \mathbb{C}).$$

L'une des raisons provient du calcul suivant : si P est un polynôme trigonométrique, de la forme

$$P : x \in \mathbb{R} \mapsto \sum_{n=-N}^N \gamma_n e^{inx},$$

où $N \in \mathbb{N}$ et les γ_n sont dans \mathbb{C} , on a

$$\begin{aligned}
\|P\|_2^2 &= \frac{1}{2\pi} \int_0^{2\pi} |P(t)|^2 dt = \frac{1}{2\pi} \int_0^{2\pi} \overline{P(t)} P(t) dt \\
&= \frac{1}{2\pi} \int_0^{2\pi} \overline{\sum_{n=-N}^N \gamma_n e^{int} P(t)} dt = \frac{1}{2\pi} \int_0^{2\pi} \sum_{n=-N}^N \overline{\gamma_n} e^{-int} P(t) dt \\
&= \sum_{n=-N}^N \overline{\gamma_n} \frac{1}{2\pi} \int_0^{2\pi} e^{-int} P(t) dt = \sum_{n=-N}^N \overline{\gamma_n} \frac{1}{2\pi} \int_0^{2\pi} e^{-int} \sum_{m=-N}^N \gamma_m e^{imt} dt \\
&= \sum_{n=-N}^N \overline{\gamma_n} \sum_{m=-N}^N \gamma_m \frac{1}{2\pi} \int_0^{2\pi} e^{-int} e^{imt} dt \\
&= \sum_{n=-N}^N \overline{\gamma_n} \gamma_n = \sum_{n=-N}^N |\gamma_n|^2.
\end{aligned}$$

Autrement dit, on peut calculer explicitement la norme 2 d'un polynôme trigonométrique en fonction de ses coefficients de Fourier. Le théorème principal de cette section, que nous admettrons en partie, affirme que ce calcul reste vrai pour toute fonction continue par morceaux et 2π périodique. C'est l'objet du paragraphe suivant.

2.3.2 Théorème de Parseval

Théorème 10 Soit $f \in \mathcal{C}_{m,2\pi}^0(\mathbb{R}; \mathbb{C})$. Alors

1. la série $\sum_{n \in \mathbb{Z}} |c_n(f)|^2$ converge,
2. et sa limite est $\|f\|_2^2$.

Preuve : on montre seulement la première assertion. On observe d'abord que

$$\begin{aligned}
\frac{1}{2\pi} \int_0^{2\pi} \overline{S_N(f)(t)} f(t) dt &= \sum_{n=-N}^N \overline{c_n(f)} \frac{1}{2\pi} \int_0^{2\pi} f(t) e^{-int} dt \\
&= \sum_{n=-N}^N |c_n(f)|^2 = \|S_N(f)\|_2^2.
\end{aligned} \tag{2.1}$$

On en déduit

$$\begin{aligned}
\|f\|_2^2 &= \frac{1}{2\pi} \int_0^{2\pi} |(f(t) - S_N(f)(t)) + S_N(f)(t)|^2 dt \\
&= \frac{1}{2\pi} \int_0^{2\pi} |f(t) - S_N(f)(t)|^2 dt + \frac{1}{2\pi} \int_0^{2\pi} |S_N(f)(t)|^2 dt \\
&\quad - \frac{1}{\pi} \int_0^{2\pi} \operatorname{Re} \left(\overline{S_N(f)(t)} (f(t) - S_N(f)(t)) \right) dt.
\end{aligned}$$

Dans la deuxième égalité, on a utilisé l'identité $|z - z'|^2 = |z|^2 + |z'|^2 - 2 \operatorname{Re} \bar{z} z'$. Or,

$$\int_0^{2\pi} \operatorname{Re} \left(\overline{S_N(f)(t)} (f(t) - S_N(f)(t)) \right) dt = \operatorname{Re} \int_0^{2\pi} \overline{S_N(f)(t)} (f(t) - S_N(f)(t)) dt$$

et

$$\begin{aligned} \frac{1}{2\pi} \int_0^{2\pi} \overline{S_N(f)(t)}(f(t) - S_N(f)(t)) dt &= \frac{1}{2\pi} \int_0^{2\pi} \overline{S_N(f)(t)}f(t) dt - \frac{1}{2\pi} \int_0^{2\pi} \overline{S_N(f)(t)}S_N(f)(t) dt \\ &= \|S_N(f)\|_2^2 - \|S_N(f)\|_2^2 = 0. \end{aligned}$$

Dans la dernière ligne, on a utilisé (2.1). Il vient donc

$$\|f\|_2^2 = \frac{1}{2\pi} \int_0^{2\pi} |(f(t) - S_N(f)(t))|^2 dt + \frac{1}{2\pi} \int_0^{2\pi} |S_N(f)(t)|^2 dt.$$

Cette dernière identité peut se réécrire

$$\|f\|_2^2 = \|f - S_N(f)\|_2^2 + \|S_N(f)\|_2^2. \quad (2.2)$$

En particulier,

$$\|S_N(f)\|_2^2 \leq \|f\|_2^2.$$

Or, comme $S_N(f)$ est un polynôme trigonométrique, $\|S_N(f)\|_2^2 = \sum_{n=-N}^N |c_n(f)|^2$. La série à termes positifs $\sum_n |c_n(f)|^2$ est majorée donc converge, ce qui montre la première partie de l'énoncé. \square

Corollaire 3 Soit $f \in \mathcal{C}_{m,2\pi}^0(\mathbb{R}; \mathbb{C})$. Alors

$$\lim_{N \rightarrow +\infty} \|f - S_N(f)\|_2 = 0.$$

Preuve : D'après (2.2),

$$\|f - S_N(f)\|_2^2 = \|f\|_2^2 - \|S_N(f)\|_2^2.$$

Par le théorème précédent, le membre de droite tend vers 0, ce qui montre le corollaire. \square

Remarque 9 On peut observer que le théorème précédent est vrai pour toute fonction continue 2π périodique, et pas seulement pour les fonctions \mathcal{C}^2 qui sont 2π périodiques, comme dans le Corollaire 2. En revanche, la conclusion du Corollaire 2 est plus forte que celle du Corollaire 3.

En effet, les normes $\|\cdot\|_\infty$ et $\|\cdot\|_2$ ne sont pas équivalentes sur $\mathcal{C}_{2\pi}^0(\mathbb{R}; \mathbb{C})$. Certes, on a pour tout $f \in \mathcal{C}_{2\pi}^0(\mathbb{R}; \mathbb{C})$,

$$\|f\|_2^2 = \frac{1}{2\pi} \int_0^{2\pi} |f(t)|^2 dt \leq \frac{1}{2\pi} \int_0^{2\pi} \|f\|_\infty^2 dt \leq \|f\|_\infty^2$$

et donc $\|\cdot\|_2 \leq \|\cdot\|_\infty$. Cela implique que la convergence uniforme implique la convergence quadratique. En revanche, la réciproque est fautive.

Remarque 10 Le théorème 10 implique aussi le résultat qu'on avait admis au début de la section précédente : si une fonction f continue et 2π périodique a tous ses coefficients de Fourier nuls, alors $f = 0$. En effet, une telle fonction vérifie :

$$\|f\|_2^2 = \sum_{n \in \mathbb{Z}} |c_n(f)|^2 = 0,$$

donc $f = 0$.

Le théorème 10 permet aussi d'avoir une amélioration du Corollaire 2 :

Corollaire 4 Soit $f \in \mathcal{C}_{2\pi}^1(\mathbb{R}; \mathbb{C})$. Alors la série de Fourier de f converge normalement vers f .

Preuve : Il suffit de prouver que la série $\sum_n |c_n(f)|$ converge. Comme f' est continue et 2π périodique, on sait que la série $\sum_n |c_n(f')|^2$ converge.

Par intégration par parties, on a déjà vu que $c_n(f) = \frac{1}{in} c_n(f')$. Donc

$$|c_n(f)| = \frac{1}{n} |c_n(f')| \leq \frac{1}{2} \left(\frac{1}{n^2} + |c_n(f')|^2 \right).$$

Comme les séries de terme général $1/n^2$ et $|c_n(f')|^2$ converge, la série de terme général $|c_n(f)|$ converge également. □

On peut vérifier que ce résultat reste vrai si f est seulement continu et C^1 par morceaux.

2.3.3 Interprétation en termes de produit scalaire complexe

Toute cette section sur la convergence quadratique peut se reformuler élégamment en termes de produit scalaire complexe. La définition diffère légèrement du produit scalaire réel.

Définition 24 Soit V un espace vectoriel sur \mathbb{C} . Un produit scalaire complexe (ou hermitien) est une application $\langle \cdot, \cdot \rangle : V \times V \rightarrow \mathbb{C}$ telle que

1. (antilinearité par rapport à la première variable) pour tout $w \in V$, la fonction $v \in V \mapsto \langle v, w \rangle \in \mathbb{R}$ est antilinéaire,
2. (linéarité par rapport à la deuxième variable) pour tout $v \in V$, la fonction $w \in V \mapsto \langle v, w \rangle \in \mathbb{R}$ est linéaire,
3. (antisymétrie) pour tout $v, w \in V$, $\langle v, w \rangle = \overline{\langle w, v \rangle}$,
4. (défini positif) pour tout $v \in V$, $\langle v, v \rangle \geq 0$ avec égalité si et seulement si $v = 0$.

On peut montrer que les produits scalaires complexes vérifient les principales propriétés des produits scalaires réels :

1. pour tout $v, w \in V$,

$$\langle v + w, v + w \rangle = \|v\|^2 + \|w\|^2 + 2 \operatorname{Re} \langle v, w \rangle.$$

2. l'inégalité de Cauchy-Schwarz : pour tout $v, w \in V$

$$|\langle v, w \rangle| \leq \sqrt{\langle v, v \rangle} \sqrt{\langle w, w \rangle}.$$

3. l'application $\| \cdot \| : V \rightarrow \mathbb{R}$

$$\|v\| = \sqrt{\langle v, v \rangle} \quad , v \in V$$

est une norme sur V .

4. Pour deux vecteurs orthogonaux v et w , c'est-à-dire $\langle v, w \rangle = 0$, on a la relation de Pythagore :

$$\|v + w\|^2 = \|v\|^2 + \|w\|^2.$$

Sur l'espace vectoriel $\mathcal{C}_{2\pi}^0(\mathbb{R}; \mathbb{C})$, on introduit l'application

$$(f, g) \in \mathcal{C}_{m, 2\pi}^0(\mathbb{R}; \mathbb{C}) \mapsto \langle f, g \rangle = \frac{1}{2\pi} \int_0^{2\pi} \overline{f(t)} g(t) dt.$$

Proposition 23 L'application $\langle \cdot, \cdot \rangle$ est un produit scalaire sur $\mathcal{C}_{2\pi}^0(\mathbb{R}; \mathbb{C})$.

Preuve : Par linéarité de l'intégrale, on a bien antilinéarité par rapport à la première variable f , linéarité par rapport à la seconde variable g . Pour vérifier l'antisymétrie, on doit se souvenir comment on définit l'intégrale d'une fonction continue h à valeurs complexes. On décompose h en partie réelle et partie imaginaire $h = h_1 + ih_2$ et on pose

$$\int_a^b h(t) dt = \int_a^b h_1(t) dt + i \int_a^b h_2(t) dt.$$

On en déduit

$$\overline{\int_a^b h(t) dt} = \int_a^b h_1(t) dt - i \int_a^b h_2(t) dt.$$

Or $\bar{h} = h_1 - ih_2$ et donc

$$\int_a^b \bar{h}(t) dt = \int_a^b h_1(t) dt - i \int_a^b h_2(t) dt.$$

On a donc montré

$$\overline{\int_a^b h(t) dt} = \int_a^b \bar{h}(t) dt.$$

Appliquons ce résultat à $h = \bar{f}g$:

$$\overline{\langle f, g \rangle} = \overline{\int_a^b \bar{f}g dt} = \int_a^b \overline{\bar{f}g}(t) dt = \int_a^b \bar{g}f(t) dt = \langle g, f \rangle,$$

d'où l'on déduit l'antisymétrie. Enfin,

$$\langle f, f \rangle = \int_a^b \overline{f(t)}f(t) dt = \int_a^b |f(t)|^2 dt \geq 0$$

avec égalité si et seulement si $f = 0$ (on utilise ici que l'intégrale d'une fonction continue positive est nulle ssi la fonction est partout nulle). □

La norme associée à ce produit scalaire est

$$\sqrt{\frac{1}{2\pi} \int_0^{2\pi} |f(t)|^2 dt}.$$

On retrouve la norme $\|\cdot\|_2$ déjà introduite.

Définition 25 On dit qu'une famille $\{e_j\}_{j \in J}$ (J est un ensemble d'indices fini ou non) de vecteurs de V est

1. orthogonale si pour tout $j, j' \in J, j \neq j', \langle e_j, e_{j'} \rangle = 0$,
2. orthonormale (ou orthonormée) si pour tout $j, j' \in J, \langle e_j, e_{j'} \rangle = \delta_{jj'}$.

Exercice 10 Montrer qu'une famille orthonormée est libre.

On définit pour tout $k \in \mathbb{Z}, e_k : t \mapsto e^{ikt}$.

Proposition 24 La famille $\{e_k\}$ est orthonormale dans $C_{2\pi}^0(\mathbb{R}; \mathbb{C})$.

Preuve : En effet, pour tout $k \in \mathbb{Z}$,

$$\frac{1}{2\pi} \int_0^{2\pi} |e^{ikt}|^2 dt = 1.$$

Pour tout $k \neq \ell$,

$$\frac{1}{2\pi} \int_0^{2\pi} e^{-ikt} e^{i\ell t} dt = \frac{1}{2\pi} \left[\frac{1}{i(\ell - k)} e^{i(\ell - k)t} \right]_0^{2\pi} = 0.$$

Ainsi, la famille $\{e_k\}$ est orthonormée. □

On note $\mathcal{P}_{2\pi}^N(\mathbb{R}; \mathbb{C})$ l'ensemble des polynômes trigonométriques de degré $\leq N$, c'est-à-dire le sous-espace vectoriel engendré par la famille $\{e_{-N}, \dots, e_N\}$ dans $\mathcal{C}_{2\pi}^0(\mathbb{R}; \mathbb{C})$. Alors, la série de Fourier $S_N(f)$ de f est la *projection orthogonale* de f sur $\mathcal{P}_{2\pi}^N(\mathbb{R}; \mathbb{C})$. Cela signifie que $f - S_N(f)$ est orthogonale à tous les éléments de $\mathcal{P}_{2\pi}^N(\mathbb{R}; \mathbb{C})$. Pour le voir, il suffit de vérifier que $f - S_N(f)$ est orthogonal à chacun des e_n , $-N \leq n \leq N$:

$$\langle e_n, f \rangle = \frac{1}{2\pi} \int_0^{2\pi} e^{-int} f(t) dt = c_n(f)$$

et

$$\langle e_n, S_N(f) \rangle = \frac{1}{2\pi} \int_0^{2\pi} e^{-int} S_N(f)(t) dt = \sum_{\ell=-N}^N c_\ell(f) \frac{1}{2\pi} \int_0^{2\pi} e^{-int} e^{i\ell t} dt = c_n(f).$$

On en déduit

$$\langle e_n, f - S_N(f) \rangle = c_n(f) - c_n(f) = 0,$$

qui est le résultat attendu.

En particulier, la fonction $P \in \mathcal{P}_{2\pi}^N(\mathbb{R}; \mathbb{C}) \mapsto \|f - P\|_2^2$ atteint son minimum en $P = S_N(f)$. En effet, c'est une conséquence du *Théorème de Pythagore* : comme $f - S_N(f)$ et $S_N(f) - P$ sont orthogonaux,

$$\|f - P\|_2^2 = \|f - S_N(f)\|_2^2 + \|S_N(f) - P\|_2^2 \geq \|f - S_N(f)\|_2^2,$$

avec égalité si et seulement si $P = S_N(f)$.

Méthode 3 Pour développer une fonction 2π périodique $f : \mathbb{R} \rightarrow \mathbb{R}$ en série de Fourier, on pourra procéder comme suit :

1. Tracer le graphe de f sur plusieurs périodes (au moins 3).
2. Déterminer la classe de f : continue par morceaux ou continue, C^1 par morceaux ou C^1 .
3. Calculer les coefficients de Fourier de f : a_k et b_k , ou c_k . On utilisera ici la parité éventuelle de f .
4. Appliquer
 - (a) le théorème de Parseval : convergence en $\|\cdot\|_2$ pour les fonctions continues par morceaux,
 - (b) le théorème de Dirichlet : convergence en tout point de continuité de f pour les fonctions C^1 par morceaux,
 - (c) le théorème de convergence normale pour les fonctions continues C^1 par morceaux.

Chapitre 3

Fonctions de plusieurs variables

Soient $n, p \geq 1$. L'objet de ce chapitre est de définir (une généralisation de) la dérivabilité pour les fonctions définies sur une partie de \mathbb{R}^n et prenant leurs valeurs dans \mathbb{R}^p . Les applications linéaires constituent l'exemple le plus important de telles fonctions mais toutes les fonctions de \mathbb{R}^n dans \mathbb{R}^p ne sont pas linéaires.

Pour définir rigoureusement cette notion de dérivabilité, nous ferons intervenir des normes sur \mathbb{R}^n et \mathbb{R}^p . Comme il s'agit d'espaces de dimension finie, toutes les normes sont équivalentes. On rappelle que deux normes N_1 et N_2 sont équivalentes s'il existe deux constantes $C_1 > 0$ et $C_2 > 0$ telles que pour tout vecteur x , $N_1(x) \leq C_1 N_2(x)$ et $N_2(x) \leq C_2 N_1(x)$.

Lorsque deux normes sont équivalentes, si l'une vérifie une inégalité, alors l'autre vérifie la même inégalité à une constante multiplicative près. Aussi, beaucoup des notions qui vont suivre, faisant intervenir des inégalités, ne dépendront pas de la norme choisie. C'est la raison pour laquelle, en général, nous ne précisons pas la norme que nous utilisons, et nous la noterons $\|\cdot\|$. Les boules $B(0, r)$, $B(0, \eta)$, ... sont définies à partir de cette norme.

3.1 Dérivées partielles des fonctions de \mathbb{R}^n dans \mathbb{R}

Dans cette section, on considère une fonction $g : D \subset \mathbb{R}^n \rightarrow \mathbb{R}$ où D est un ouvert de \mathbb{R}^n . On fixe aussi un point $a = (a_1, \dots, a_n) \in D$. Pour tout $1 \leq j \leq n$, on va considérer la fonction d'une variable

$$x_j \mapsto g(a_1, \dots, a_{j-1}, x_j, a_{j+1}, \dots, a_n).$$

Cette fonction est bien sûr définie sur

$$D_j = \{x_j \in \mathbb{R} : (a_1, \dots, a_{j-1}, x_j, a_{j+1}, \dots, a_n) \in D\}.$$

Comme D est ouvert, il existe $\varepsilon > 0$ tel que pour tout $x \in \mathbb{R}^n$, si $\|x - a\|_\infty < \varepsilon$, alors $x \in D$. En particulier, si $|x_j - a_j| < \varepsilon$, alors

$$\|(a_1, \dots, a_{j-1}, x_j, a_{j+1}, \dots, a_n) - (a_1, \dots, a_n)\|_\infty < \varepsilon$$

et donc $(a_1, \dots, a_{j-1}, x_j, a_{j+1}, \dots, a_n) \in D$. Ainsi, la fonction

$$x_j \mapsto g(a_1, \dots, a_{j-1}, x_j, a_{j+1}, \dots, a_n)$$

est définie sur $]a_j - \varepsilon, a_j + \varepsilon[$.

Définition 26 On dit que g admet une dérivée partielle par rapport à x_j en a quand la fonction

$$x_j \mapsto g(a_1, \dots, a_{j-1}, x_j, a_{j+1}, \dots, a_n)$$

est dérivable en a_j . On note

$$\partial_j g(a) = \lim_{x_j \xrightarrow{\neq} a_j} \frac{g(a_1, \dots, a_{j-1}, x_j, a_{j+1}, \dots, a_n) - g(a_1, \dots, a_n)}{x_j - a_j}.$$

Même si pour calculer $\partial_j g(a)$, on gèle les variables $x_1 = a_1, \dots, x_{j-1} = a_{j-1}, x_{j+1} = a_{j+1}, \dots, x_n = a_n$, et on dérive par rapport à x_j , le résultat $\partial_j g(a)$ dépend de toutes les coordonnées de a et pas seulement de a_j .

Définition 27 On dit que g est de classe C^1 sur D quand

1. $\forall a \in D, \forall 1 \leq j \leq n, g$ admet une dérivée partielle $\partial_j g(a)$,
2. $\forall 1 \leq j \leq n, a \mapsto \partial_j g(a)$ est continue sur D .

Définition 28 Soit g une fonction de classe C^1 sur D .

1. On appelle différentielle de g en a l'application linéaire

$$h = (h_1, \dots, h_n) \in \mathbb{R}^n \mapsto h_1 \partial_1 g(a) + \dots + h_n \partial_n g(a).$$

On la note $Dg(a)$.

2. On appelle gradient de g en a le vecteur de \mathbb{R}^n

$$\nabla g(a) = \begin{pmatrix} \partial_1 g(a) \\ \vdots \\ \partial_n g(a) \end{pmatrix}.$$

On observe que pour tout $h \in \mathbb{R}^n$,

$$Dg(a)[h] = \langle \nabla g(a), h \rangle.$$

En particulier, si (e_1, \dots, e_n) désigne la base canonique de \mathbb{R}^n ,

$$Dg(a)[e_i] = \langle \nabla g(a), e_i \rangle = \partial_i g(a).$$

La matrice de l'application linéaire $Dg(a)$ dans les bases canoniques est

$$\begin{pmatrix} \partial_1 g(a) & \partial_2 g(a) & \dots & \partial_n g(a) \end{pmatrix}.$$

Il s'agit de la transposée du gradient.

Les fonctions dérivables d'une seule variable admettent un développement limité d'ordre 1 : si $f : \mathbb{R} \rightarrow \mathbb{R}$ est dérivable en 0, alors pour tout $t \in \mathbb{R}$,

$$f(t) = f(0) + tf'(0) + o(t),$$

où $o(t)$ désigne une fonction de la forme $t \mapsto t\delta(t)$, avec δ une fonction qui tend vers 0 en 0.

Le théorème fondamental qui suit affirme qu'une fonction de plusieurs variables qui est C^1 admet un développement limité d'ordre 1.

Théorème 11 Soit $g : D \rightarrow \mathbb{R}$ une fonction de classe C^1 . Alors pour tout $a \in D$, il existe $r > 0$ et une fonction $\delta : D \rightarrow \mathbb{R}$ vérifiant $\lim_{x \rightarrow a} \delta(x) = 0$ et telle que pour tout $x \in D$:

$$g(x) = g(a) + Dg(a)[x - a] + \|x - a\|\delta(x) = g(a) + \sum_{i=1}^n \partial_i g(a)(x_i - a_i) + \|x - a\|\delta(x).$$

Rappelons que la propriété $\lim_{x \rightarrow a} \delta(x) = 0$ signifie que pour tout $\varepsilon > 0$, il existe $\eta > 0$ tel que pour tout $x \in B(a, \eta)$,

$$|\delta(x)| \leq \varepsilon.$$

Sur $D \setminus \{a\}$, la fonction δ vaut nécessairement :

$$\delta(x) = \frac{g(x) - g(a) - Dg(a)[x - a]}{\|x - a\|}, \quad x \neq a.$$

L'intérêt du théorème est donc d'affirmer que cette quantité tend vers 0 quand x tend vers a .

Preuve du théorème : On prouve le résultat pour $n = 2$ (le cas général est analogue mais moins lisible). On fixe $a = (a_1, a_2) \in \mathbb{R}^2$. Soit $\varepsilon > 0$. Comme $\partial_1 g$ et $\partial_2 g$ sont continues en a , il existe $\eta > 0$ tel que pour tout $\|(x_1, x_2) - (a_1, a_2)\|_\infty < \eta$,

$$|\partial_1 g(x_1, x_2) - \partial_1 g(a_1, a_2)| \leq \varepsilon, \quad |\partial_2 g(x_1, x_2) - \partial_2 g(a_1, a_2)| \leq \varepsilon. \quad (3.1)$$

Pour tout $(x_1, x_2) \in B((a_1, a_2), \eta)$, on écrit

$$g(x_1, x_2) - g(a_1, a_2) = (g(x_1, x_2) - g(x_1, a_2)) + (g(x_1, a_2) - g(a_1, a_2)).$$

Par le théorème fondamental du calcul différentiel appliqué à la fonction $x_2 \mapsto g(x_1, x_2)$,

$$g(x_1, x_2) - g(x_1, a_2) = \int_0^1 \frac{d}{dt} (g(x_1, a_2 + t(x_2 - a_2))) dt = \int_0^1 \partial_2 g(x_1, a_2 + t(x_2 - a_2))(x_2 - a_2) dt.$$

On en déduit

$$|g(x_1, x_2) - g(x_1, a_2) - \partial_2 g(a_1, a_2)(x_2 - a_2)| \leq |x_2 - a_2| \int_0^1 |\partial_2 g(x_1, a_2 + t(x_2 - a_2)) - \partial_2 g(a_1, a_2)| dt.$$

Pour tout $t \in [0, 1]$, on a $\|(x_1, a_2 + t(x_2 - a_2)) - (a_1, a_2)\|_\infty < \eta$. Donc par (3.1),

$$|\partial_2 g(x_1, a_2 + t(x_2 - a_2)) - \partial_2 g(a_1, a_2)| \leq \varepsilon.$$

Ainsi,

$$|g(x_1, x_2) - g(x_1, a_2) - \partial_2 g(a_1, a_2)(x_2 - a_2)| \leq \varepsilon |x_2 - a_2|.$$

On montre de même

$$|g(x_1, a_2) - g(a_1, a_2) - \partial_1 g(a_1, a_2)(x_1 - a_1)| \leq \varepsilon |x_1 - a_1|.$$

Il suit que

$$|g(x_1, x_2) - g(a_1, a_2) - (x_1 - a_1)\partial_1 g(a_1, a_2) - (x_2 - a_2)\partial_2 g(a_1, a_2)| \leq \varepsilon(|x_1 - a_1| + |x_2 - a_2|) = 2\varepsilon \|(x_1 - a_1, x_2 - a_2)\|_\infty.$$

Pour résumer, on a montré que si $\|(x_1 - a_1, x_2 - a_2)\|_\infty < \eta$, alors $\delta(x_1, x_2) < 2\varepsilon \|(x_1 - a_1, x_2 - a_2)\|_\infty$, en notant

$$\delta(x_1, x_2) = \begin{cases} \frac{g(x_1, x_2) - g(a_1, a_2) - Dg(a_1, a_2)[x_1 - a_1, x_2 - a_2]}{\|(x_1 - a_1, x_2 - a_2)\|} & \text{si } (x_1, x_2) \neq (a_1, a_2), \\ 0 & \text{si } (x_1, x_2) = (a_1, a_2). \end{cases}$$

Autrement dit, $\lim_{(x_1, x_2) \rightarrow 0} \delta(x_1, x_2) = 0$. Cela conclut la preuve du théorème. \square

On dit parfois que le gradient d'une fonction f indique la *direction de plus grande pente*. Ecrivons le développement limité d'une fonction f de classe C^1 en un point $a \in D$ avec $x = a + te$, où $t > 0$ et $e \in \mathbb{R}^n$, $\|e\| = 1$ (comme D est ouvert, $a + te \in D$ si t est assez petit) :

$$f(a + te) = f(a) + \langle \nabla f(a), te \rangle + |t|\delta(a + te) = f(a) + t\langle \nabla f(a), e \rangle + |t|\delta(a + te),$$

où δ est une fonction qui tend vers 0 en a . Supposons que $\nabla f(a) \neq 0$. Pour un t fixé, la direction e pour laquelle f augmente le plus, c'est-à-dire pour laquelle la différence $f(a + te) - f(a)$ est maximale, est obtenue lorsque la quantité $\langle \nabla f(a), e \rangle$ est maximale, si du moins on néglige le reste $|t|\delta(a + te)$. Or d'après l'inégalité de Cauchy-Schwarz,

$$\langle \nabla f(a), e \rangle \leq \|\nabla f(a)\| \|e\|,$$

avec égalité si et seulement si e est positivement colinéaire à $\nabla f(a)$:

$$e = \frac{\nabla f(a)}{\|\nabla f(a)\|}.$$

C'est donc la direction qui permet d'augmenter le plus $f(a + te)$.

Comme pour les fonctions dérivables, on peut ajouter, multiplier, composer les fonctions C^1 :

Proposition 25 Soient g_1, g_2 deux fonctions C^1 sur D et $\lambda \in \mathbb{R}$. Alors $g_1 + g_2, g_1g_2, \lambda g_1$ sont C^1 sur D . De plus, si g_1 ne s'annule pas sur D , alors $\frac{1}{g_1}$ est C^1 sur D . On a le formulaire :

$$\partial_j(g_1 + g_2)(a) = \partial_j g_1(a) + \partial_j g_2(a) \quad , \quad \partial_j(g_1 g_2)(a) = \partial_j g_1(a) g_2(a) + \partial_j g_2(a) g_1(a),$$

$$\partial_j(\lambda g_1)(a) = \lambda \partial_j g_1(a) \quad , \quad \partial_j \left(\frac{1}{g_1} \right) = -\partial_j g_1(a) \frac{1}{g_1(a)^2}.$$

Proposition 26 Une forme linéaire est C^1 sur \mathbb{R}^n .

Preuve : Une forme linéaire est une fonction de la forme $\ell : x \mapsto a_1 x_1 + \dots + a_n x_n$. Alors pour tout $1 \leq i \leq n$,

$$\partial_i \ell(x) = a_i$$

et les fonctions (constantes) $x \mapsto a_i$ sont continues ! Donc ℓ est de classe C^1 .

Exemple 20

$$g : (x, y) \in \mathbb{R}^2 \mapsto \frac{\sin(x^3 - y^5)}{1 + x^2 + y^2}.$$

La fonction g est C^1 comme quotient de fonctions C^1 , le dénominateur ne s'annulant pas.

Exercice 11 Soit $f : \mathbb{R}^2 \rightarrow \mathbb{R}$ définie par

$$f(x, y) = \begin{cases} \frac{(\sin x)^4}{x^2 + y^2} & \text{si } (x, y) \neq (0, 0), \\ 0 & \text{sinon} \end{cases}.$$

Montrer que f est de classe C^1 sur \mathbb{R}^2 .

3.2 Dérivées partielles des fonctions de \mathbb{R}^n dans \mathbb{R}^p

Définition 29 Soit $D \subset \mathbb{R}^n$ ouvert et $f : \mathbb{R}^n \rightarrow \mathbb{R}^p$. On note (f_1, \dots, f_p) les fonctions coordonnées de f :

$$x = (x_1, \dots, x_n) \mapsto \begin{pmatrix} f_1(x_1, \dots, x_n) \\ \vdots \\ f_p(x_1, \dots, x_n) \end{pmatrix}.$$

On dit que f admet une dérivée partielle par rapport à x_j en $a = (a_1, \dots, a_n) \in D$ si pour tout $1 \leq i \leq p$, f_i admet une dérivée partielle par rapport à x_j en a . On note

$$\partial_j f(a) = \begin{pmatrix} \partial_j f_1(a) \\ \vdots \\ \partial_j f_p(a) \end{pmatrix}.$$

Exemple 21

$$f : (x, y) \mapsto (e^{x^2 + y^3}, xy)$$

$$\partial_1 f(x, y) = (2xe^{x^2 + y^3}, y) \quad , \quad \partial_1 f(1, 2) = (2e^{1+8}, 2).$$

Définition 30 Soit $f : D \rightarrow \mathbb{R}^p$. On dit que f est de classe C^1 sur D si pour tout $1 \leq i \leq p$, $f_i : D \rightarrow \mathbb{R}$ est de classe C^1 et on définit alors la différentielle de f en un point $a \in D$ par :

$$Df(a) : h \in \mathbb{R}^n \mapsto (Df_1(a)[h], \dots, Df_p(a)[h]) \in \mathbb{R}^p \quad , \quad h \in \mathbb{R}^n.$$

Comme chaque $Df_i(a)$ est une application linéaire de \mathbb{R}^n vers \mathbb{R} , $Df(a)$ est une application linéaire de \mathbb{R}^n vers \mathbb{R}^p . Pour calculer sa matrice dans les bases canoniques, notées $(e_i)_{1 \leq i \leq n}$ et $(\varepsilon_j)_{1 \leq j \leq p}$ respectivement, on calcule pour $1 \leq i \leq n$:

$$Df(a)[e_i] = (Df_1(a)[e_i], \dots, Df_p(a)[e_i]) = (\partial_i f_1(a), \dots, \partial_i f_p(a)).$$

Il s'agit du $i^{\text{ème}}$ vecteur colonne de la matrice de $Df(a)$ dans les bases canoniques. Cette matrice s'écrit donc :

$$\begin{pmatrix} \partial_1 f_1 & \dots & \partial_n f_1 \\ \vdots & \ddots & \vdots \\ \partial_1 f_p & \dots & \partial_n f_p \end{pmatrix}.$$

On vient d'écrire la matrice jacobienne de f en a , qu'on notera $M_{Df(a)}$ dans la suite.

Exemple 22 Pour l'exemple précédent,

$$M_{Df(x,y)} = \begin{pmatrix} 2xe^{x^2+y^3} & 3y^2e^{x^2+y^3} \\ y & x \end{pmatrix}.$$

Proposition 27 Soient $f, h : D \rightarrow \mathbb{R}^p$ deux applications C^1 , et $\lambda \in \mathbb{R}$.

1. Alors $f + h, \langle f, h \rangle, \lambda f$ sont C^1 sur D .
2. On a le formulaire : $\forall a \in D$,

$$\partial_j(f + h)(a) = \partial_j f(a) + \partial_j h(a) \quad , \quad \partial_j(\lambda f)(a) = \lambda \partial_j f(a),$$

$$\partial_j \langle f, h \rangle(a) = \langle \partial_j f(a), h(a) \rangle + \langle f(a), \partial_j h(a) \rangle.$$

Preuve : Les deux premières identités se vérifient en dérivant chaque fonction coordonnée de f . Pour la troisième, on note que

$$\langle f, h \rangle = f_1 h_1 + \dots + f_p h_p.$$

Donc

$$\partial_j \langle f, h \rangle = (\partial_j f_1) h_1 + f_1 (\partial_j h_1) + \dots + (\partial_j f_p) h_p + f_p (\partial_j h_p) = \langle \partial_j f, h \rangle + \langle f, \partial_j h \rangle.$$

□

Remarque 11 Si $p = 3$, on montre par une preuve analogue que

$$\partial_j(f \wedge g) = (\partial_j f) \wedge g + f \wedge (\partial_j g).$$

On a toujours le résultat fondamental :

Théorème 12 Soit $f : D \subset \mathbb{R}^n \rightarrow \mathbb{R}^p$ une fonction de classe C^1 . Pour tout $a \in D$, il existe $\delta : D \rightarrow \mathbb{R}^p$ vérifiant $\lim_{x \rightarrow a} \delta(x) = 0$ et tel que pour tout $x \in D$,

$$f(x) = f(a) + Df(a)[x - a] + \|x - a\| \delta(x).$$

Preuve : ce résultat s'obtient comme conséquence du cas $p = 1$ en travaillant coordonnée par coordonnée.

□

La propriété $\lim_{x \rightarrow a} \delta(x) = 0$ signifie que pour tout $\varepsilon > 0$, il existe $\eta > 0$ tel que pour tout $x \in B(a, \eta)$,

$$\|\delta(x)\| \leq \varepsilon.$$

Le théorème précédent admet la réciproque suivante :

Proposition 28 Soit $f : D \subset \mathbb{R}^n \rightarrow \mathbb{R}^p$ une fonction continue. Supposons qu'il existe une fonction continue $L : D \rightarrow \mathcal{L}(\mathbb{R}^n; \mathbb{R}^p)$ telle que pour tout $a \in D$, il existe $\delta : D \rightarrow \mathbb{R}^p$ vérifiant $\lim_{x \rightarrow a} \delta(x) = 0$ et

$$\forall x \in D, \quad f(x) = f(a) + L(a)[x - a] + \|x - a\|\delta(x). \quad (3.2)$$

Alors f est C^1 sur D et pour tout $a \in D$, $Df(a) = L(a)$.

Preuve : Observer que pour tout $a \in D$, $L(a)$ est une application linéaire de \mathbb{R}^n vers \mathbb{R}^p . La continuité de L se réfère à une norme quelconque sur l'espace de dimension finie $\mathcal{L}(\mathbb{R}^n; \mathbb{R}^p)$ (où toutes les normes sont équivalentes).

Fixons $a \in D$. Notons e_1 le premier vecteur de la base canonique et prenons $x = a + te_1$, $t \in \mathbb{R}^*$, dans (3.2). Alors

$$\frac{f(a + te_1) - f(a)}{t} = L(a)[e_1] + \|e_1\|\delta(a + te_1).$$

On en déduit

$$\lim_{t \rightarrow 0} \frac{f(a + te_1) - f(a)}{t} = L(a)[e_1].$$

Comme $f(a + te_1) = f(a_1 + t, a_2, \dots, a_n)$, en notant $a = (a_1, \dots, a_n)$, la limite dans le membre de gauche de l'égalité précédente est justement $\partial_1 f(a)$. Ainsi, $\partial_1 f(a)$ existe et vaut $L(a)[e_1]$.

L'application $\varphi \in \mathcal{L}(\mathbb{R}^n; \mathbb{R}^p) \mapsto \varphi[e_1] \in \mathbb{R}^p$ est linéaire sur un espace de dimension finie, donc continue. Par composition avec l'application L qui est aussi continue, on déduit que l'application $a \in D \mapsto L(a)[e_1]$ est continue. Ainsi, $a \mapsto \partial_1 f(a)$ est continue.

Le résultat reste vrai pour tout $1 \leq i \leq n$. Ainsi, les dérivées partielles de f existent et sont continues. Donc f est C^1 . De plus, pour tout $a \in D$ et pour tout $x = (x_1, \dots, x_n) \in D$,

$$Df(a)[x - a] = \sum_{i=1}^n (x_i - a_i) \partial_i f(a) = \sum_{i=1}^n (x_i - a_i) L(a)[e_i] = L(a) \left[\sum_{i=1}^n (x_i - a_i) e_i \right] = L(a)[x - a].$$

On en déduit que $Df(a) = L(a)$, ce qui achève la preuve. □

Remarque 12 Lorsque $n = 1$ et que f est C^1 sur un intervalle ouvert I de \mathbb{R} , pour tout $a \in I$, on peut écrire $f(x) = f(a) + f'(a)(x - a) + |x - a|\delta(x)$, pour tout $x \in I$. Ici, δ est une fonction qui tend vers 0 en a . Les arguments de la preuve précédente montrent que $Df(a)[x - a] = f'(a)(x - a)$. Autrement dit, pour $n = 1$, la différentielle $Df(a)$ de f en a est simplement la multiplication par $f'(a)$.

Composition des fonctions C^1

On rappelle que si $f, g : \mathbb{R} \rightarrow \mathbb{R}$ sont deux fonctions dérivables, alors $g \circ f$ est dérivable et pour tout $a \in \mathbb{R}$,

$$(g \circ f)'(a) = g'(f(a)) \cdot f'(a).$$

On dispose d'une formule analogue pour les fonctions de plusieurs variables. Soient $f : D \subset \mathbb{R}^n \rightarrow \mathbb{R}^p$ et $g : E \subset \mathbb{R}^p \rightarrow \mathbb{R}^q$ deux fonctions C^1 sur les ouverts D et E respectivement. On suppose que $f(D) \subset E$, de sorte que la composée $g \circ f$ est bien définie.

Théorème 13 La fonction $g \circ f$ est C^1 sur D et

$$D(g \circ f)(a) = Dg(f(a)) \circ Df(a) \quad , \quad a \in D.$$

On se souvient que $Df(a)$ est une application linéaire de \mathbb{R}^n vers \mathbb{R}^p et que $Dg(f(a))$ est une application linéaire de \mathbb{R}^p vers \mathbb{R}^q . La composée est donc une application linéaire de \mathbb{R}^n vers \mathbb{R}^q , ce qui est aussi le cas de $D(g \circ f)(a)$.

Preuve : On écrit

$$f(x) = f(a) + Df(a)[x - a] + \|x - a\|\delta(x) \quad , \quad g(y) = g(f(a)) + Dg(f(a))[y - f(a)] + \|y - f(a)\|\delta'(y),$$

où $\lim_{x \rightarrow a} \delta(x) = 0$ et $\lim_{y \rightarrow f(a)} \delta'(y) = 0$. Lorsqu'on prend $y = f(x)$, il vient

$$g(f(x)) = g(f(a)) + Dg(f(a))[f(x) - f(a)] + \|f(x) - f(a)\|\delta'(f(x)).$$

Comme $Dg(f(a))$ est une application linéaire et que $f(x) - f(a) = Df(a)[x - a] + \|x - a\|\delta(x)$,

$$\begin{aligned} Dg(f(a))[f(x) - f(a)] &= Dg(f(a)) [Df(a)[x - a] + \|x - a\|\delta(x)] \\ &= Dg(f(a))[Df(a)[x - a]] + \|x - a\|Dg(f(a))[\delta(x)]. \end{aligned}$$

On en déduit

$$\begin{aligned} g(f(x)) &= g(f(a)) + Dg(f(a))[Df(a)[x - a]] + \|x - a\|Dg(f(a))[\delta(x)] + \|f(x) - f(a)\|\delta'(f(x)) \\ &= g(f(a)) + (Dg(f(a)) \circ Df(a))[x - a] + \|x - a\|Dg(f(a))[\delta(x)] + \|f(x) - f(a)\|\delta'(f(x)) \\ &= g(f(a)) + (Dg(f(a)) \circ Df(a))[x - a] + \|x - a\|\delta''(x) \end{aligned}$$

où on a introduit la fonction

$$\delta''(x) = Dg(f(a))[\delta(x)] + \frac{\|f(x) - f(a)\|}{\|x - a\|} \delta'(f(x)) \quad \forall x \neq a,$$

et $\delta''(a) = 0$. Montrons que $\lim_{x \rightarrow a} \delta''(x) = 0$. Comme $\lim_{x \rightarrow a} \delta(x) = 0$ et $Dg(f(a))$ est une application linéaire en dimension finie donc continue, $\lim_{x \rightarrow a} Dg(f(a))[\delta(x)] = 0$. On sait aussi que $\lim_{x \rightarrow a} f(x) = f(a)$ et $\lim_{y \rightarrow f(a)} \delta'(y) = 0$ d'où l'on déduit $\lim_{x \rightarrow a} \delta'(f(x)) = 0$. Enfin,

$$\|f(x) - f(a)\| = \|Df(a)[x - a] + \|x - a\|\delta(x)\| \leq \|Df(a)\| \|x - a\| + \|x - a\| \|\delta(x)\| \leq (\|Df(a)\| + 1) \|x - a\| \|\delta(x)\|.$$

Ainsi,

$$\frac{\|f(x) - f(a)\|}{\|x - a\|} \leq (\|Df(a)\| + 1) \|\delta(x)\|.$$

On en déduit que

$$\lim_{x \rightarrow a} \left(\frac{\|f(x) - f(a)\|}{\|x - a\|} \delta'(f(x)) \right) = 0.$$

On a donc

$$(g \circ f)(x) = (g \circ f)(a) + Dg(f(a)) \circ Df(a)[x - a] + \|x - a\| \delta''(x),$$

avec $\lim_{x \rightarrow a} \delta''(x) = 0$. On conclut la preuve du théorème en utilisant la proposition 28 appliquée à $g \circ f$. □

En termes de matrices jacobiniennes, le théorème de composition se réécrit :

$$M_{D(g \circ f)}(a) = M_{Dg(f(a))} M_{Df(a)}.$$

En effet, la matrice de la composée de deux applications linéaires s'obtient comme le produit des matrices de ces deux applications linéaires.

En lisant les coefficients de $M_{D(g \circ f)}(a)$, on obtient une version équivalente du théorème de composition des différentielles en termes de dérivées partielles :

Proposition 29 Soient $f : D \subset \mathbb{R}^n \rightarrow \mathbb{R}^p$ et $g : E \subset \mathbb{R}^p \rightarrow \mathbb{R}^q$ deux applications C^1 telles que $f(D) \subset E$:

$$f(x_1, \dots, x_n) = \begin{pmatrix} f_1(x_1, \dots, x_n) \\ \vdots \\ f_p(x_1, \dots, x_n) \end{pmatrix}, \quad g(u_1, \dots, u_p) = \begin{pmatrix} g_1(u_1, \dots, u_p) \\ \vdots \\ g_q(u_1, \dots, u_p) \end{pmatrix}.$$

Alors $g \circ f : D \rightarrow \mathbb{R}^q$ est C^1 et de plus, pour tout $1 \leq j \leq n$,

$$\partial_j(g \circ f)(a) = \partial_1 g(f(a)) \partial_j f_1(a) + \dots + \partial_p g(f(a)) \partial_j f_p(a).$$

Un cas particulier important du théorème de composition est donné dans la proposition suivante :

Proposition 30 Soit $I \subset \mathbb{R}$ un intervalle ouvert et $f : I \rightarrow \mathbb{R}^n$ une fonction C^1 : en notant $f(t) = (f_1(t), \dots, f_n(t))$ les coordonnées de f , cela signifie que chaque f_i est C^1 . On suppose que $f(I) \subset D$. Soit $g : D \rightarrow \mathbb{R}^q$ une fonction C^1 . Alors la fonction $G = g \circ f$ est C^1 sur I et on a

$$G'(t) = \partial_1 g(f(t)) f_1'(t) + \dots + \partial_n g(f(t)) f_n'(t) = Dg(f(t))[f'(t)].$$

Exemple 23

$$h : (r, \theta) \mapsto (r \cos \theta, r \sin \theta) \quad , \quad f : (x, y) \mapsto (f_1(x, y), f_2(x, y), f_3(x, y)).$$

On pose $G = f \circ h$.

Alors

$$\partial_r G = \begin{pmatrix} \partial_x f_1(r \cos \theta, r \sin \theta) \cos \theta \\ \partial_x f_2(r \cos \theta, r \sin \theta) \cos \theta \\ \partial_x f_3(r \cos \theta, r \sin \theta) \cos \theta \end{pmatrix} + \begin{pmatrix} \partial_y f_1(r \cos \theta, r \sin \theta) \sin \theta \\ \partial_y f_2(r \cos \theta, r \sin \theta) \sin \theta \\ \partial_y f_3(r \cos \theta, r \sin \theta) \sin \theta \end{pmatrix},$$

ce qui correspond bien à

$$\partial_r G = \cos \theta \partial_x f(r \cos \theta, r \sin \theta) + \sin \theta \partial_y f(r \cos \theta, r \sin \theta).$$

De même,

$$\partial_\theta G = -r \sin \theta \partial_x f(r \cos \theta, r \sin \theta) + r \cos \theta \partial_y f(r \cos \theta, r \sin \theta).$$

On vérifie que

$$M_{DG(r,\theta)} = M_{Df(h(r,\theta))} M_{Dh(r,\theta)}.$$

3.3 Changement de variables

Dans cette section, D et Ω désignent deux ouverts de \mathbb{R}^n .

Définition 31 On dit que $f : D \rightarrow \mathbb{R}^n$ est un C^1 difféomorphisme de D sur Ω si

1. f est C^1 ,
2. f est une bijection de D sur Ω ,
3. f^{-1} est de classe C^1 dans Ω .

Proposition 31 Soit $f : D \rightarrow \mathbb{R}^n$ un C^1 difféomorphisme sur Ω . Notons $h = f^{-1}$ son inverse. Alors pour tout $a \in D$, $Df(a)$ est inversible et

$$Dh(f(a)) = [Df(a)]^{-1}$$

Preuve : Pour tout $x \in D$, on a $h \circ f(x) = x$. En différentiant, on obtient

$$Dh(f(x)) \circ Df(x) = Id.$$

On en déduit que $Df(x)$ est une application linéaire inversible, d'inverse $Dh(f(x))$. □

Définition 32 On note $J_f(x) = \det M_{Df(x)}$ le jacobien de f en x .

Le théorème suivant, appelé théorème d'inversion globale, donne une condition nécessaire et suffisante pour que f soit un difféomorphisme.

Théorème 14 Soit $f : D \rightarrow \mathbb{R}^n$ une application C^1 . Alors f est un C^1 difféomorphisme de D sur Ω si et seulement si

1. f est une bijection de D sur Ω ,
2. $\forall x \in D, J_f(x) = \det Df(x) \neq 0$.

Exemple 24 (Coordonnées polaires dans le plan)

$$\mathcal{C} : (r, \theta) \in D =]0, +\infty[\times]-\pi, \pi[\mapsto (r \cos \theta, r \sin \theta) \in \Omega = \mathbb{R}^2 \setminus \{(x, 0), x \leq 0\}$$

Alors \mathcal{C} est C^1 (produit de fonction C^1). Pour tout $(x, y) \in \Omega$, posons $r = \sqrt{x^2 + y^2}$ et $\theta \in]-\pi, \pi[$ défini par

$$\cos \theta = \frac{x}{\sqrt{x^2 + y^2}}, \quad \sin \theta = \frac{y}{\sqrt{x^2 + y^2}}.$$

Comme $(x/r, y/r) \neq (-1, 0)$, on peut bien exclure $\theta = \pm\pi$. Alors on a bien $\mathcal{C}(r, \theta) = (x, y)$ donc \mathcal{C} est surjectif de D sur Ω . Par ailleurs, \mathcal{C} est injectif : si $(r \cos \theta, r \sin \theta) = (r' \cos \theta', r' \sin \theta')$, alors en prenant la $\|\cdot\|_2$ de ces vecteurs, on trouve $r = r'$, puis par injectivité de $\theta \mapsto (\cos \theta, \sin \theta)$ sur $]-\pi, \pi[$, on en déduit $\theta = \theta'$.

Enfin, on calcule $J_{\mathcal{C}}(x, y) = r \neq 0$. Le théorème s'applique : \mathcal{C} est un C^1 difféomorphisme de D sur Ω .

Exemple 25 *En revanche,*

$$\tilde{C} : (r, \theta) \in [0, +\infty[\times]-\pi, \pi[\mapsto (r \cos \theta, r \sin \theta) \in \mathbb{R}^2 \setminus \{(x, 0), x < 0\}$$

n'est pas injectif de $[0, +\infty[\times]-\pi, \pi[$ sur son image, et n'est donc pas un difféomorphisme.

Exemple 26 *La fonction*

$$\psi : (x, y) \in D = \mathbb{R} \times \mathbb{R}^{*+} \mapsto \left(\frac{x}{y}, x^2 + y^2\right) \in \Omega = \mathbb{R} \times \mathbb{R}^{*+}.$$

est un C^1 difféomorphisme.

3.4 Dérivées partielles d'ordre supérieur

Dans cette section, D est un ouvert de \mathbb{R}^n et $f : D \rightarrow \mathbb{R}^p$.

Définition 33 *On dit que f est de classe C^2 si f est C^1 et toutes les dérivées partielles de f , i.e. les fonctions $\frac{\partial f}{\partial x_j} : D \rightarrow \mathbb{R}^p$ sont aussi C^1 .*

On note

$$\partial_i(\partial_j f) = \partial_{ij}^2 f.$$

Théorème 15 (Théorème de Schwarz) *Si f est de classe C^2 sur D , alors*

$$\partial_{ij}^2 f = \partial_{ji}^2 f.$$

Définition 34 *On définit par récurrence les fonctions de classe C^n : f est de classe C^n si f est C^1 et toutes les dérivées partielles de f sont C^{n-1} . De manière équivalente, f est C^n si f est C^{n-1} et les dérivées partielles de f d'ordre $n-1$ sont C^1 .*

Théorème 16 (Formule de Taylor-Young) *Soient $f : D \rightarrow \mathbb{R}$ de classe C^2 , $a \in D$. Alors il existe $\delta : D \rightarrow \mathbb{R}$ tel que $\lim_{x \rightarrow a} \delta(x) = 0$ et tel que pour tout $x \in D$,*

$$f(x) = f(a) + Df(a)[x - a] + \frac{1}{2} \sum_{1 \leq i, j \leq n} \partial_{ij}^2 f(a)(x_i - a_i)(x_j - a_j) + \|x - a\|^2 \delta(x).$$

Preuve : On travaille avec la norme $\|\cdot\|_1$ sur \mathbb{R}^n (changer de norme revient à changer de fonction δ , mais pas le fait que cette fonction tende vers 0 en a). On introduit la fonction

$$\varphi(x) := f(x) - \sum_{i=1}^n \partial_i f(a)(x_i - a_i) - \frac{1}{2} \sum_{1 \leq i, j \leq n} \partial_{ij}^2 f(a)(x_i - a_i)(x_j - a_j).$$

Alors pour tout $k \in \llbracket 1, n \rrbracket$,

$$\begin{aligned} \partial_k \varphi(x) &= \partial_k f(x) - \partial_k f(a) - \frac{1}{2} \sum_{i=1}^n \partial_{ik}^2 f(a)(x_i - a_i) - \frac{1}{2} \sum_{j=1}^n \partial_{kj}^2 f(a)(x_j - a_j) \\ &= \partial_k f(x) - \partial_k f(a) - \sum_{i=1}^n \partial_i(\partial_k f)(a)(x_i - a_i). \end{aligned}$$

Soit $\varepsilon > 0$. Comme $\partial_k f$ est C^1 pour tout $k \in \llbracket 1, n \rrbracket$, il existe $\eta > 0$ tel que pour tout $x \in B(a, \eta)$, pour tout $k \in \llbracket 1, n \rrbracket$, $|\partial_k \varphi(x)| \leq \varepsilon \|x - a\|_1$.
En utilisant le lemme ci-dessous, on en déduit que pour tout $x \in B(a, \eta)$,

$$|\varphi(x) - \varphi(a)| \leq \varepsilon \|x - a\|_1^2,$$

ce qui se réécrit

$$|f(x) - f(a) - Df(a)[x - a] - \frac{1}{2} \sum_{1 \leq i, j \leq n} \partial_{ij}^2 f(a)(x_i - a_i)(x_j - a_j)| \leq \varepsilon \|x - a\|_1^2.$$

Cela montre bien que la fonction

$$\delta(x) = \frac{1}{\|x - a\|_1^2} \left(f(x) - f(a) - Df(a)[x - a] - \frac{1}{2} \sum_{1 \leq i, j \leq n} \partial_{ij}^2 f(a)(x_i - a_i)(x_j - a_j) \right), \quad x \neq a,$$

(prolongée par 0 en a) tend vers 0 en a .

□

Lemme 1 Soit $\varphi : B(a, \eta) \subset \mathbb{R}^n \rightarrow \mathbb{R}$ une fonction de classe C^1 . On suppose qu'il existe $\varepsilon > 0$ tel que pour tout $x \in B(a, \eta)$, pour tout $k \in \llbracket 1, n \rrbracket$, $|\partial_k \varphi(x)| \leq \varepsilon \|x - a\|_1$. Alors pour tout $x \in B(a, \eta)$,

$$|\varphi(x) - \varphi(a)| \leq \varepsilon \|x - a\|_1^2.$$

Preuve : Soit $x \in B(a, \eta)$. On considère la fonction $\theta(t) = \varphi(a + t(x - a))$. Alors θ est C^1 par composition, de dérivée $\theta'(t) = D\varphi(a + t(x - a))[x - a]$. On a

$$\varphi(x) - \varphi(a) = \theta(1) - \theta(0) = \int_0^1 \theta'(t) dt = \int_0^1 D\varphi(a + t(x - a))[x - a] dt.$$

On utilise maintenant que $a + t(x - a) \in B(a, \eta)$ pour écrire

$$|D\varphi(a + t(x - a))[x - a]| = \left| \sum_{k=1}^n \partial_k \varphi(a + t(x - a))(x_k - a_k) \right| \leq \sum_{k=1}^n |\partial_k \varphi(a + t(x - a))| |x_k - a_k| \leq \varepsilon \|t(x - a)\|_1 \sum_{k=1}^n |x_k - a_k| \leq \varepsilon t \|x - a\|_1^2 \leq \varepsilon \|x - a\|_1^2.$$

On en déduit

$$|\varphi(x) - \varphi(a)| \leq \int_0^1 |D\varphi(a + t(x - a))[x - a]| dt \leq \varepsilon \|x - a\|_1^2,$$

ce qui achève la preuve. □

3.5 Extrema

Dans toute cette section, D est un ouvert de \mathbb{R}^2 et $g : D \rightarrow \mathbb{R}$ une fonction de classe C^1 .

Définition 35 1. On dit que g admet un maximum local en (x_0, y_0) s'il existe $r > 0$ tel que

$$\forall (x, y) \in B((x_0, y_0), r), \quad g(x, y) \leq g(x_0, y_0),$$

2. minimum local : \geq

3. maximum local strict : $<$

4. minimum local strict : $>$

Définition 36 1. g admet un maximum global en (x_0, y_0) si pour tout $(x, y) \in D$,

$$g(x, y) \leq g(x_0, y_0),$$

2. g admet un minimum global en (x_0, y_0) si pour tout $(x, y) \in D$,

$$g(x, y) \geq g(x_0, y_0).$$

Ces définitions généralisent les notions de maximum et minimum pour les fonctions d'une seule variable. Quelques rappels dans ce cadre : soit $f : \mathbb{R} \rightarrow \mathbb{R}$ une fonction au moins C^2 . Si x_0 est un minimum local de f , alors $f'(x_0) = 0$. On peut le voir par exemple en considérant les taux d'accroissements : pour tout $h > 0$ assez petit,

$$\frac{f(x_0 + h) - f(x_0)}{h} \geq 0.$$

On en déduit en passant à la limite pour $h \rightarrow 0^+$ que $f'(x_0) \geq 0$. Maintenant, pour tout $h < 0$,

$$\frac{f(x_0 + h) - f(x_0)}{h} \leq 0.$$

En passant à la limite pour $h \rightarrow 0^-$, on trouve cette fois $f'(x_0) \leq 0$. Finalement, $f'(x_0) = 0$.

Attention, ce n'est pas parce que f' s'annule en un point qu'il s'agit automatiquement d'un minimum ou d'un maximum local. Par exemple, la fonction $f(x) = x^3$ vérifie $f'(0) = 0$ et pourtant 0 n'est ni un minimum ni un maximum local de f . Autrement dit, la condition $f'(x_0) = 0$ est nécessaire mais non suffisante pour être un extremum local.

Une condition suffisante est donnée par la dérivée seconde. Plus précisément, si $x_0 \in \mathbb{R}$ est tel que $f'(x_0) = 0$ et $f''(x_0) > 0$, alors x_0 est un minimum local (on pourra aisément deviner un énoncé analogue pour les maxima locaux). La preuve repose sur la formule de Taylor-Young : il existe $r > 0$ et $\delta :]-r, r[\rightarrow \mathbb{R}$ une fonction qui tend vers 0 en x_0 tels que

$$f(x) = f(x_0) + (x - x_0)f'(x_0) + \frac{1}{2}(x - x_0)^2 f''(x_0) + (x - x_0)^2 \delta(x) \quad , \quad x \in]x_0 - r, x_0 + r[.$$

Comme $f'(x_0) = 0$, on en déduit

$$f(x) = f(x_0) + \frac{1}{2}(x - x_0)^2 f''(x_0) + (x - x_0)^2 \delta(x) \quad , \quad x \in]x_0 - r, x_0 + r[.$$

Quitte à réduire $r > 0$, on peut supposer que $|\delta(x)| \leq \frac{1}{4}f''(x_0)$ lorsque $x \in]x_0 - r, x_0 + r[$ (ici, on utilise l'hypothèse sur δ). Alors pour tout $x \in]x_0 - r, x_0 + r[$,

$$f(x) \geq f(x_0) + \frac{1}{2}(x - x_0)^2 f''(x_0) - \frac{1}{4}f''(x_0)(x - x_0)^2 = f(x_0) + \frac{1}{4}(x - x_0)^2 f''(x_0) \geq f(x_0)$$

et donc x_0 est un minimum local.

Ces conditions nécessaires ou suffisantes d'extrémalité peuvent se généraliser aux fonctions de plusieurs variables. On se limite dans la suite à deux variables.

Proposition 32 *Si g présente un minimum ou un maximum local en (x_0, y_0) et si g est C^1 , alors*

$$\partial_1 g(x_0, y_0) = 0 = \partial_2 g(x_0, y_0).$$

Définition 37 *Si $Dg(x_0, y_0) = 0$, on dit que (x_0, y_0) est un point critique.*

Preuve : Les fonctions $x \mapsto g(x, y_0)$ et $y \mapsto g(x_0, y)$ admettent toutes les deux un maximum ou un minimum local en x_0 et y_0 respectivement. Donc

$$\partial_1 g(x_0, y_0) = 0 = \partial_2 g(x_0, y_0).$$

□

Remarque 13 *Les extremums sont à chercher parmi les points critiques, mais tous les points critiques ne sont pas forcément des extremums.*

Pour pouvoir déterminer si un point critique est un extremum ou pas, on a besoin des dérivées partielles d'ordre 2, et plus précisément de la formule de Taylor-Young d'ordre 2. Pour simplifier les notations, on introduit les notations :

$$r = \partial_{11}^2 g(x_0, y_0), \quad s = \partial_{12}^2 g(x_0, y_0), \quad t = \partial_{22}^2 g(x_0, y_0).$$

Au voisinage d'un point critique (x_0, y_0) (où les dérivés partielles s'annulent), la formule de Taylor-Young s'écrit :

$$g(x_0 + h_1, y_0 + h_2) = g(x_0, y_0) + \frac{1}{2} (r^2 h_1^2 + t^2 h_2^2 + 2s^2 h_1 h_2) + \|(h_1, h_2)\|_2^2 \delta(h_1, h_2),$$

avec $\lim_{(h_1, h_2) \rightarrow 0} \delta(h_1, h_2) = 0$. En utilisant la notation matricielle,

$$g(x_0 + h_1, y_0 + h_2) - g(x_0, y_0) = \frac{1}{2} \begin{pmatrix} h_1 & h_2 \end{pmatrix} \begin{pmatrix} r & s \\ s & t \end{pmatrix} \begin{pmatrix} h_1 \\ h_2 \end{pmatrix} + \|(h_1, h_2)\|_2^2 \delta(h_1, h_2). \quad (3.3)$$

Théorème 17 Soient $g : D \subset \mathbb{R}^2 \rightarrow \mathbb{R}$ une fonction de classe C^2 , $(x_0, y_0) \in D$ un point critique de g . On note $r = \partial_{11}^2 g(x_0, y_0)$, $s = \partial_{12}^2 g(x_0, y_0)$, $t = \partial_{22}^2 g(x_0, y_0)$.

1. (x_0, y_0) est un minimum local strict si et seulement si $rt - s^2 > 0$ et $r > 0$,
2. (x_0, y_0) est un maximum local strict si et seulement si $rt - s^2 > 0$ et $r < 0$.

Avant de donner la preuve du théorème, on aura besoin du résultat suivant en algèbre linéaire.

Lemme 2 Si $A \in \mathcal{M}_2(\mathbb{R})$ est une matrice symétrique définie positive, alors il existe $\alpha > 0$ tel que pour tout $h \in \mathbb{R}^2$,

$$h^T Ah \geq \alpha \|h\|_2^2.$$

Preuve du lemme : Noter que Ah est un vecteur (colonne) de \mathbb{R}^n et $h^T Ah$ est un vecteur colonne de \mathbb{R} , autrement dit un nombre réel !

Comme A est symétrique, il existe une matrice orthogonale P et une matrice diagonale D telle que $A = P^T DP$. Les éléments diagonaux de D sont les valeurs propres λ_1 et λ_2 de A : $D = \begin{pmatrix} \lambda_1 & 0 \\ 0 & \lambda_2 \end{pmatrix}$.

Alors

$$h^T Ah = h^T P^T DP h = k^T D k$$

en posant $k = Ph$. En notant $k = \begin{pmatrix} k_1 \\ k_2 \end{pmatrix}$, il vient

$$h^T Ah = \lambda_1 k_1^2 + \lambda_2 k_2^2 \geq \min(\lambda_1, \lambda_2)(k_1^2 + k_2^2) = \min(\lambda_1, \lambda_2) \|k\|_2^2.$$

Comme P est orthogonale, $\|k\|_2 = \|h\|_2$. Ainsi,

$$h^T Ah \geq \min(\lambda_1, \lambda_2) \|h\|_2^2.$$

On pose $\alpha = \min(\lambda_1, \lambda_2)$. Comme A est définie positive, $\lambda_1 > 0$ et $\lambda_2 > 0$, donc $\alpha > 0$. Cela permet de conclure la preuve du lemme. □

On peut alors se lancer dans la preuve du théorème :

Preuve : On prouve le théorème dans le cas $r > 0$. Le cas $r < 0$ se déduit du précédent en considérant $-f$ au lieu de f .

La matrice

$$A = \begin{pmatrix} r & s \\ s & t \end{pmatrix}$$

est symétrique, donc diagonalisable dans une base orthonormée. Notons λ_1 et λ_2 ses valeurs propres. Le déterminant de A qui vaut $rt - s^2 > 0$ est le produit des valeurs propres $\lambda_1 \lambda_2$. On en déduit que λ_1 et λ_2 sont de même signe. Comme $r > 0$ et $rt - s^2 > 0$, alors $t > s^2/r > 0$. La trace de A , qui est la somme des valeurs propres $\lambda_1 + \lambda_2$, vaut $r + t > 0$. On en déduit que λ_1 et λ_2 sont toutes les deux > 0 . Autrement dit, A est une matrice symétrique définie positive.

D'après le lemme, il existe $\alpha > 0$ tel que pour tout $h \in \mathbb{R}^2$,

$$h^T Ah \geq \alpha \|h\|_2^2.$$

Par la formule (3.3), on obtient :

$$\begin{aligned} g(x_0 + h_1, y_0 + h_2) &= g(x_0, y_0) + \frac{1}{2} h^T Ah + \|(h_1, h_2)\|_2^2 \delta(h_1, h_2) \\ &\geq g(x_0, y_0) + \frac{\alpha}{2} \|(h_1, h_2)\|_2^2 + \|(h_1, h_2)\|_2^2 \delta(h_1, h_2) \\ &\geq g(x_0, y_0) + \left(\frac{\alpha}{2} + \delta(h_1, h_2)\right) \|(h_1, h_2)\|_2^2. \end{aligned}$$

Comme $\lim_{(h_1, h_2)} \delta(h_1, h_2) = 0$, il existe $r > 0$ tel que pour tout $(h_1, h_2) \in B(0, r)$, $|\delta(h_1, h_2)| \leq \frac{\alpha}{2}$. Alors pour de tels vecteurs (h_1, h_2) ,

$$g(x_0 + h_1, y_0 + h_2) \geq g(x_0, y_0),$$

ce qui montre bien que (x_0, y_0) est un minimum local de g . □

Dans le cas où $rt - s^2 < 0$, on dit que (x_0, y_0) est un point selle. En notant comme dans la preuve ci-dessus λ_1, λ_2 les valeurs propres de $\nabla^2 g(x_0, y_0)$, alors le signe du déterminant $rt - s^2 = \lambda_1 \lambda_2$ montre que λ_1 et λ_2 sont de signe opposé. Supposons par exemple $\lambda_1 < 0 < \lambda_2$. Notons ε_1 un vecteur propre associé à λ_1 et ε_2 un vecteur propre associé à λ_2 . Alors on peut montrer que 0 est un minimum local de $t \mapsto f((x_0, y_0) + t\varepsilon_1)$ et un maximum local de $t \mapsto f((x_0, y_0) + t\varepsilon_2)$.

Méthode 4 Pour étudier les extrema locaux d'une fonction $f : U \rightarrow \mathbb{R}$ de classe C^2 sur un ouvert $U \subset \mathbb{R}^n$,

1. on calcule les points critiques de f , c'est-à-dire les $x \in U$ tels que $\nabla f(x) \neq 0$,
2. on calcule la hessienne de f en chaque point critique,
3. on détermine s'il s'agit d'extrema (minima ou maxima locaux) en calculant les valeurs propres de ces matrices hessiennes.

Exercice 12 Chercher les extrema locaux des fonctions suivantes sur leur domaine de définition

1. $(x, y) \mapsto \cos x \sin y$,
2. $(x, y) \mapsto xy$,
3. $(x, y) \mapsto x^2 - 2y^2 - 2x + 1$.
4. $(x, y) \mapsto x^3 + \frac{4}{3}y^3 + 3y^2 - 3x + 1$.

Chapitre 4

Equations différentielles

4.1 Introduction

Les équations différentielles constituent une véritable zoologie : les scalaires et les vectorielles, les linéaires et les non linéaires, les homogènes et les non homogènes. En voici quelques exemples :

$$(1) y' = t$$

$$(2) y' = y^2 + \frac{1}{\sqrt{t}}$$

$$(3) y'' = y$$

$$(4) y''' = y'^2 + y^2$$

$$(5) ty' + y = \cos t$$

$$(6) \begin{cases} y'_1 = ty_1(t) + y_2(t), \\ y'_2(t) = y_1(t) + ty_2(t). \end{cases}$$

Ici, y désigne une fonction inconnue dérivable sur un intervalle de \mathbb{R} et à valeurs dans \mathbb{R} , au moins 1 fois dérivable pour (1), (2) et (5), 2 fois dérivable pour (3) et 3 fois dérivable pour (4). Dans le cas (6), on parle de système différentiel : il y a deux fonctions inconnues y_1 et y_2 .

Ces fonctions inconnues, on les cherche sur les intervalles où les fonctions qui apparaissent dans l'équation sont continues. Dans les cas (1), (3), (4) et (5) et (6), ces fonctions sont définies sur \mathbb{R} . On cherchera donc les solutions parmi les fonctions définies et dérivables (autant de fois que nécessaire) sur \mathbb{R} (on devra pourtant se contenter parfois de solutions définies seulement sur des intervalles de \mathbb{R}). En revanche, dans le cas (2), la fonction $t \mapsto \frac{1}{\sqrt{t}}$ n'étant définie que sur $]0, +\infty[$, les solutions seront cherchées uniquement parmi les fonctions dérivables sur $]0, +\infty[$ (ou sur des sous-intervalles de $]0, +\infty[$).

Le cas (5) est un peu particulier. En effet, pour résoudre une telle équation, on la met sous forme résolue :

$$y' + \frac{1}{t}y = \frac{\cos t}{t}.$$

Cela n'est possible que si $t \neq 0$. On résout cette nouvelle équation sur $] -\infty, 0[$ puis sur $]0, +\infty[$, avant de chercher à raccorder les solutions obtenues en une solution globale, c'est-à-dire définie sur \mathbb{R} tout entier. On détaillera ce point dans la section suivante.

Il arrive qu'une équation différentielle n'ait pas de solution, mais le plus souvent, elle en a une infinité. Pourtant, on ne sait pas en général résoudre *explicitement* une équation différentielle. Il existe deux cas importants où c'est néanmoins possible : les équations différentielles linéaires d'ordre 1 et les équations différentielles linéaires à coefficients constants.

4.2 Equations différentielles linéaires scalaires d'ordre n

4.2.1 Solution

Soit $n \in \mathbb{N}^*$. Une équation différentielle linéaire scalaire d'ordre n est de la forme

$$(E) \quad a_0(t)y^{(n)}(t) + a_1(t)y^{(n-1)}(t) + \dots + a_n(t)y(t) = f(t)$$

où les fonctions a_j et f sont définies et continues sur un intervalle I de \mathbb{R} . Il s'agit des données du problème. Une solution de cette équation est une fonction $y : I \rightarrow \mathbb{R}$ qui est n fois dérivable sur I et qui vérifie la relation (E).

Insistons : on ne cherche à résoudre une équation différentielle que sur des intervalles où toutes les fonctions a_j et f sont *continues* !

Dans la suite, on considèrera surtout les cas $n = 1$ et $n = 2$.

Exemple 27

$$ty'(t) + y(t) = \cos t.$$

4.2.2 Homogène et résolue

Lorsque $a_0 \equiv 1$, on dit que (E) est sous forme résolue. Pour passer de l'équation générale (E) à une équation sous forme résolue, il suffit de diviser toute l'équation par a_0 , sur tout sous-intervalle de I où a_0 ne s'annule pas. On obtient alors l'équation sous forme résolue

$$(\tilde{E}) \quad y^{(n)}(t) + \frac{a_1(t)}{a_0(t)}y^{(n-1)}(t) + \dots + \frac{a_n(t)}{a_0(t)}y(t) = \frac{f(t)}{a_0(t)}.$$

Il ne faudra pas oublier ensuite de chercher à *raccorder* entre elles les solutions obtenues sur chacun de ces sous-intervalles.

Exemple 28 L'équation résolue associée à notre équation différentielle linéaire scalaire préférée est

$$y'(t) + \frac{1}{t}y(t) = \frac{\cos t}{t}$$

qu'on résout sur $] -\infty, 0[$, sur $]0, +\infty[$ puis on cherche si on peut raccorder les solutions obtenues en une fonction dérivable sur \mathbb{R} .

Raccorder les solutions, cela signifie trouver les solutions $y_- :] -\infty, 0[\rightarrow \mathbb{R}$ et $y_+ :]0, +\infty[\rightarrow \mathbb{R}$ telles qu'il existe $\ell \in \mathbb{R}$ de sorte que

1. $\lim_{t \rightarrow 0^-} y_-(t) = \ell = \lim_{t \rightarrow 0^+} y_+(t)$,
2. la fonction

$$y : t \in \mathbb{R} \mapsto \begin{cases} y_-(t) & \text{si } t < 0, \\ \ell & \text{si } t = 0, \\ y_+(t) & \text{si } t > 0, \end{cases}$$

est dérivable sur \mathbb{R} (en fait, on sait déjà qu'elle est dérivable hors de 0 donc il suffit de vérifier la dérivabilité en 0),

3. la fonction y est solution de $ty'(t) + y(t) = \cos t$ pour $t \in \mathbb{R}$ (en fait, on sait que c'est déjà le cas pour $t \neq 0$, donc il suffit de le vérifier pour $t = 0$).

On appelle équation différentielle linéaire homogène associée à (E) l'équation (E_0) où on remplace le second membre f par 0.

Exemple 29 L'équation homogène associée à la précédente est :

$$y'(t) + \frac{1}{t}y(t) = 0.$$

4.2.3 Linéaire

Le mot *linéaire* fait référence à des espaces vectoriels que l'on va expliciter. Pour cela, introduisons l'espace $\mathcal{F}(I; \mathbb{R})$ des fonctions définies sur l'intervalle I et à valeurs dans \mathbb{R} . C'est un espace vectoriel pour la loi $+$ et la loi \cdot définies ponctuellement : pour tout $f, g \in \mathcal{F}(I; \mathbb{R})$, pour tout $\lambda \in \mathbb{R}$, la somme $f + g$ et le produit $\lambda \cdot f$ sont les fonctions définies par

$$\forall x \in \mathbb{R} \quad , \quad (f + g)(x) = f(x) + g(x) \quad , \quad (\lambda \cdot f)(x) = \lambda f(x).$$

On peut de même définir le produit de deux fonctions.

L'ensemble des fonctions n fois dérivables sur I , noté $\mathcal{D}^n(I; \mathbb{R})$, est un sous-espace vectoriel de $\mathcal{F}(I; \mathbb{R})$. En effet, c'est un ensemble non vide (il contient la fonction constante égale à 0). Prenons deux éléments f et g dans $\mathcal{D}^n(I; \mathbb{R})$ et $\lambda \in \mathbb{R}$. Alors $f + g \in \mathcal{D}^n(I; \mathbb{R})$ (autrement dit, la somme de deux fonctions n fois dérivables est encore n fois dérivable) et $\lambda \cdot f \in \mathcal{D}^n(I; \mathbb{R})$ (car le produit d'une fonction n fois dérivable par une constante reste n fois dérivable).

Considérons une équation homogène (E_0) sous forme résolue

$$(E_0) \quad y^{(n)} + a_1 y^{(n-1)} + \dots + a_n y = 0,$$

Proposition 33 *L'ensemble des solutions \mathcal{S}_0 de (E_0) est un sous-espace vectoriel de $\mathcal{D}^n(I; \mathbb{R})$.*

Preuve : Observons d'abord que la fonction constante égale à 0 est toujours solution de (E_0) . Soient $y_1, y_2 \in \mathcal{D}^n(I; \mathbb{R})$ deux solutions de (E_0) et $\lambda \in \mathbb{R}$. Ainsi y_1 et y_2 sont n fois dérivables et donc $y_1 + \lambda y_2$ l'est aussi, avec $(y_1 + \lambda y_2)^{(j)} = y_1^{(j)} + \lambda y_2^{(j)}$, $1 \leq j \leq n$. De plus,

$$y_1^{(n)} + a_1 y_1^{(n-1)} + \dots + a_n y_1 = 0 \quad , \quad y_2^{(n)} + a_1 y_2^{(n-1)} + \dots + a_n y_2 = 0.$$

En multipliant la deuxième équation par λ et en l'ajoutant à la première, on obtient

$$(y_1 + \lambda y_2)^{(n)} + a_1 (y_1 + \lambda y_2)^{(n-1)} + \dots + a_n (y_1 + \lambda y_2) = 0.$$

Ainsi $y_1 + \lambda y_2$ est bien solution de (E_0) .

Conclusion : l'ensemble des solutions de (E_0) est un sous-espace vectoriel de $\mathcal{D}^n(I; \mathbb{R})$. □

Proposition 34 *L'ensemble des solutions \mathcal{S} de (E) est un sous-espace affine de $\mathcal{D}^n(I; \mathbb{R})$: si y_p est une solution particulière quelconque de (E) ,*

$$\mathcal{S} = \{y : I \rightarrow \mathbb{R} : \text{il existe } y_0 \in \mathcal{S}_0 \text{ tel que } y = y_p + y_0\}.$$

Montrons l'égalité des deux ensembles par double inclusion. On fixe une solution particulière y_p de (E) . Soit $y_0 \in \mathcal{S}_0$. On a donc

$$y_p^{(n)} + a_1 y_p^{(n-1)} + \dots + a_n y_p = f \quad , \quad y_0^{(n)} + a_1 y_0^{(n-1)} + \dots + a_n y_0 = 0.$$

Par somme, on en déduit

$$(y_p + y_0)^{(n)} + a_1 (y_p + y_0)^{(n-1)} + \dots + a_n (y_p + y_0) = f$$

et donc $y_p + y_0 \in \mathcal{S}$. Réciproquement, soit $y \in \mathcal{S}$. On définit $y_0 = y - y_p$. Comme

$$y_p^{(n)} + a_1 y_p^{(n-1)} + \dots + a_n y_p = f \quad , \quad y^{(n)} + a_1 y^{(n-1)} + \dots + a_n y = f,$$

on a

$$y_0^{(n)} + a_1 y_0^{(n-1)} + \dots + a_n y_0 = 0$$

et donc $y_0 \in \mathcal{S}_0$. Ainsi, y est bien de la forme attendue $y_p + y_0$ avec $y_0 \in \mathcal{S}_0$. □

Les deux propositions précédentes indiquent la stratégie de résolution d'une équation différentielle. Dans une première étape, on résout l'équation homogène (mise au préalable sous forme résolue). Dans une deuxième étape, on s'efforce de trouver une solution particulière. Enfin, on conclut en donnant l'ensemble des solutions sous la forme de la somme de cette solution particulière et d'une solution générale de l'équation homogène.

4.2.4 Scalaire et vectoriel

Le mot *scalaire* signifie que toutes les fonctions qu'on considère ici sont à valeurs dans \mathbb{R} . Ce ne sera pas toujours le cas. On étudiera dans la dernière section de ce chapitre des équations différentielles vectorielles, où les inconnues seront des fonctions à valeurs dans un espace vectoriel (de dimension finie). Plus précisément, considérons un système différentiel du premier ordre de forme résolue

$$(V) \begin{cases} y_1'(t) = a_{11}(t)y_1(t) + \dots + a_{1n}(t)y_n(t) + f_1(t) \\ \vdots \\ y_n'(t) = a_{n1}(t)y_1(t) + \dots + a_{nn}(t)y_n(t) + f_n(t) \end{cases}$$

Les données sont les fonctions a_{ij} et f_i qui sont définies et continues sur un intervalle I de \mathbb{R} . Les inconnues sont les fonctions y_i , $1 \leq i \leq n$ qu'on cherche dans l'ensemble des fonctions dérivables sur I .

Comme pour les systèmes linéaires, il est très commode d'adopter une notation matricielle pour ce système différentiel. Introduisons pour cela l'espace de fonctions $\mathcal{D}(I; \mathbb{R}^n)$ des fonctions de I dans \mathbb{R}^n qui sont dérivables une fois. Un élément Y de $\mathcal{D}(I; \mathbb{R}^n)$ est donc une fonction qui à un réel t de I associe un vecteur de \mathbb{R}^n qu'on va noter sous forme matricielle, c'est-à-dire à l'aide d'un vecteur colonne :

$$Y(t) = \begin{pmatrix} Y_1(t) \\ Y_2(t) \\ \vdots \\ Y_n(t) \end{pmatrix} \quad \forall t \in I.$$

Cela permet de définir les fonctions coordonnées de Y , qui sont les Y_i , $1 \leq i \leq n$. Demander que Y soit dérivable signifie par définition que chaque Y_i est dérivable. On définit la dérivée de Y comme la fonction

$$Y'(t) = \begin{pmatrix} Y_1'(t) \\ Y_2'(t) \\ \vdots \\ Y_n'(t) \end{pmatrix} \quad \forall t \in I.$$

Alors le système différentiel (V) peut se réécrire sous forme matricielle

$$Y'(t) = A(t)Y(t) + F(t)$$

avec

$$A(t) = \begin{pmatrix} a_{11}(t) & \dots & a_{1n}(t) \\ \vdots & \ddots & \vdots \\ a_{n1}(t) & \dots & a_{nn}(t) \end{pmatrix}, \quad F(t) = \begin{pmatrix} f_1(t) \\ f_2(t) \\ \vdots \\ f_n(t) \end{pmatrix}.$$

Lorsqu'on étudie une équation différentielle scalaire d'ordre 2 ou 3 (et a fortiori d'ordre supérieur), il est utile de se ramener à un système différentiel d'ordre 1. On procède comme suit. Donnons-nous une équation scalaire d'ordre 3 sous forme résolue :

$$(E_3) \quad y^{(3)}(t) + a_1(t)y''(t) + a_2(t)y'(t) + a_3(t)y(t) = f(t).$$

On construit la matrice

$$A(t) = \begin{pmatrix} 0 & 1 & 0 \\ 0 & 0 & 1 \\ -a_3(t) & -a_2(t) & -a_1(t) \end{pmatrix},$$

et le vecteur

$$F(t) = \begin{pmatrix} 0 \\ 0 \\ f(t) \end{pmatrix}.$$

On considère enfin le système

$$(V_3) \quad Y'(t) = A(t)Y(t) + F(t).$$

Résoudre (E_3) est équivalent à résoudre (V_3) . En effet, si y est solution de (E_3) , alors définissons

$$Y(t) = \begin{pmatrix} y(t) \\ y'(t) \\ y''(t) \end{pmatrix}$$

de sorte que

$$Y'(t) = \begin{pmatrix} y'(t) \\ y''(t) \\ y^{(3)}(t) \end{pmatrix}.$$

Comme y est solution de (E_3) , on en déduit

$$Y'(t) = \begin{pmatrix} y'(t) \\ y''(t) \\ -a_3(t)y(t) - a_2(t)y'(t) - a_1(t)y''(t) + f(t) \end{pmatrix}.$$

Le membre de droite est justement $A(t)Y(t) + F(t)$. Ainsi Y est solution de (V_3) .

Réciproquement, si Y est solution de (V_3) , alors en notant Y_1, Y_2 et Y_3 ses coordonnées, on obtient que

$$Y_1' = Y_2, \quad Y_2' = Y_3, \quad Y_3' = -a_3Y_1 - a_2Y_2 - a_1Y_3 + f(t).$$

En notant $y = Y_1$, on a donc $y' = Y_2, y'' = Y_3$ et

$$y^{(3)}(t) = -a_3(t)y(t) - a_2(t)y'(t) - a_1(t)y''(t) + f(t).$$

Ainsi y est solution de (E_3) .

On vient de mettre en évidence une bijection entre les solutions de (E_3) et les solutions de (V_3) .

4.3 Equations différentielles linéaires scalaires du premier ordre

4.3.1 Résolution explicite

Partons de l'équation sous forme résolue :

$$(E) \quad y' + a(t)y = f,$$

où a et f sont deux fonctions continues sur un intervalle I .

Equation homogène

On considère dans un premier temps l'équation homogène :

$$(E_0) \quad y' + a(t)y = 0.$$

Fixons pour toute la suite $t_0 \in I$. Notons $A : I \rightarrow \mathbb{R}$ la primitive de a qui s'annule en t_0 :

$$A(t) = \int_{t_0}^t a(s) ds.$$

On remarque que

1. $(\exp(A(t)))' = \exp(A(t))A'(t) = \exp(A(t))a(t)$,
2. $\exp(A(t)) > 0$ pour tout $t \in I$.

Une fonction dérivable $y : I \rightarrow \mathbb{R}$ est solution de (E_0) si et seulement si

$$\exp(A(t))(y' + a(t)y) = 0 \iff (y \exp(A(t)))' = 0,$$

donc si et seulement si il existe une constante $C \in \mathbb{R}$ telle que $y(t) \exp A(t) = C$ pour tout $t \in \mathbb{R}$, soit encore $y(t) = C \exp(-A(t))$. Conclusion :

Proposition 35 *L'ensemble des solutions \mathcal{S}_0 de (E_0) est*

$$\mathcal{S}_0 = \left\{ C \exp \left(- \int_{t_0}^t a(s) ds \right), C \in \mathbb{R} \right\},$$

où t_0 est un point fixé de I .

Ainsi, \mathcal{S}_0 est un espace vectoriel de dimension 1 engendré par la fonction

$$t \mapsto \exp \left(- \int_{t_0}^t a(s) ds \right).$$

En remplaçant t_0 par un autre $t_1 \in I$, on obtient un autre générateur

$$t \mapsto \exp \left(- \int_{t_1}^t a(s) ds \right),$$

qui est colinéaire au précédent. En particulier, les solutions de (E_0) sont de signe constant.

Solution particulière

Maintenant, on cherche une solution particulière de (E) sous la forme $y_p(t) = C(t)y_0(t)$ où y_0 est une solution quelconque mais non nulle de (E_0) (et donc y_0 ne s'annule pas). La fonction $t \mapsto C(t)$ est donc à déterminer. Un tel y_p est solution de (E) si et seulement si

$$y_p' + ay_p = f \iff C'y_0 + Cy_0' + aCy_0 = f \iff C'y_0 + C(y_0' + ay_0) = f \iff C'y_0 = f.$$

Ici, on a utilisé que y_0 est solution de (E_0) . De manière équivalente,

$$C'(t) = \frac{f(t)}{y_0(t)}, \quad t \in I.$$

La fonction

$$C(t) = \int_{t_0}^t \frac{f(s)}{y_0(s)} ds$$

vérifie cette condition. On en déduit qu'une solution particulière est donnée par

$$y_p(t) = y_0(t)C(t) = y_0(t) \int_{t_0}^t \frac{f(s)}{y_0(s)} ds.$$

Proposition 36 L'ensemble des solutions \mathcal{S} de (E) est donné par l'ensemble des fonctions

$$\mathcal{S} = \left\{ t \mapsto y_0(t) \int_{t_0}^t \frac{f(s)}{y_0(s)} ds + \alpha y_0(t), \alpha \in \mathbb{R} \right\}$$

où t_0 est un point fixé de I . L'ensemble \mathcal{S} est un espace affine de dimension 1.

Il n'est pas nécessaire de retenir par coeur cette formule. Il est plus sûr de la retrouver à chaque fois, en faisant les calculs sous forme symbolique le plus longtemps possible, et en ne remplaçant par l'expression explicite des fonctions qu'au moment du calcul des primitives.

Exemple 30 Résoudre sur \mathbb{R} l'équation $y' + (\sin t)y = 2 \sin t \cos t$.

Réponse : $\mathcal{S} = \{2 \cos t + 2 + \alpha \exp \cos t, \alpha \in \mathbb{R}\}$.

Exemple 31 Résoudre sur $] -\frac{\pi}{2}, \frac{\pi}{2}[$ l'équation $y' + (\tan t)y = \cos t$.

Réponse : $\mathcal{S} = \{t \cos t + \alpha \cos t, \alpha \in \mathbb{R}\}$.

Exemple 32 Résoudre sur \mathbb{R} l'équation $t(t-1)^2 y' + (t^2-1)y = (-1+t)$.

Réponse : sur chaque intervalle $] -\infty, 0[,]0, 1[,]1, +\infty[$, l'ensemble des solutions est de la forme $\mathcal{S} = \left\{ \frac{t \ln |t| + 1}{(t-1)^2} + \alpha \frac{t}{(t-1)^2}, \alpha \in \mathbb{R} \right\}$. On peut raccorder continûment en 0 mais pas de raccord dérivable en 0. Raccord en 1 : il existe une unique solution dérivable sur $]0, +\infty[$ donnée par

$$y(t) = \frac{t \ln t + 1 - t}{(t-1)^2}, y(1) = \frac{1}{2}, y'(1) = -\frac{1}{6}.$$

4.3.2 Théorème de Cauchy-Lipschitz pour les équations linéaires du premier ordre

Comme conséquence des résolutions explicites du paragraphe précédent, on obtient :

Théorème 18 (Cauchy-Lipschitz) Soient I un intervalle de \mathbb{R} et $t_0 \in I$. Soient a et f des fonctions continues de I dans \mathbb{R} . On considère l'équation sous forme résolue

$$(E) \quad y' + ay = f.$$

Alors pour tout $z_0 \in \mathbb{R}$, il existe une unique solution y de (E) telle $y(t_0) = z_0$.

Preuve : On sait que l'ensemble des solutions \mathcal{S} de (E) est donné par l'ensemble des fonctions de la forme

$$t \mapsto y_0(t) \int_{t_0}^t \frac{f(s)}{y_0(s)} ds + \alpha y_0(t),$$

où α est un réel quelconque et y_0 une solution de l'équation homogène non nulle (et donc ne s'annulant pas). En prenant en compte la condition $y(t_0) = z_0$, on voit que la seule valeur possible pour α est

$$\alpha = \frac{z_0}{y_0(t_0)}.$$

D'où l'existence et unicité de la solution. □

4.4 Equations différentielles linéaires scalaires du deuxième ordre

4.4.1 Théorème de Cauchy-Lipschitz

Théorème 19 (Cauchy-Lipschitz) Soient I un intervalle de \mathbb{R} et $t_0 \in I$. Soient a_1, a_2 et f des fonctions continues de I dans \mathbb{R} . On considère l'équation sous forme résolue

$$(E) \quad y'' + a_1 y' + a_2 y = f.$$

Alors pour tout $(z_0, z_1) \in \mathbb{R}^2$, il existe une unique solution y de (E) telle $y(t_0) = z_0$ et $y'(t_0) = z_1$.

La solution est donc entièrement déterminée par sa valeur et celle de sa dérivée en t_0 .

En notant (E_0) l'équation homogène associée $y'' + a_1 y' + a_2 y = 0$, on obtient

Corollaire 5 L'ensemble des solutions \mathcal{S}_0 de (E_0) est un sous-espace vectoriel de dimension 2 de $\mathcal{D}^2(I; \mathbb{R})$.

Preuve : Soit $t_0 \in I$. Par le théorème de Cauchy-Lipschitz, il existe y_0 solution de (E_0) telle que $y_0(t_0) = 1, y_0'(t_0) = 0$ et aussi \widetilde{y}_0 solution de (E_0) telle que $\widetilde{y}_0(t_0) = 0$ et $\widetilde{y}_0'(t_0) = 1$. Montrons que la famille constituée des deux fonctions y_0 et \widetilde{y}_0 est libre. Soient $\lambda, \mu \in \mathbb{R}$ tels que $\lambda y_0 + \mu \widetilde{y}_0 = 0$. Cela signifie que pour tout $t \in I$, $\lambda y_0(t) + \mu \widetilde{y}_0(t) = 0$. On exploite cette relation de deux manières : d'abord en l'évaluant en $t = t_0$, ensuite en la dérivant par rapport à t et en évaluant l'égalité obtenue en $t = t_0$. On trouve ainsi

$$\lambda = 0, \quad \mu = 0.$$

Ceci démontre que la famille $\{y_0, \widetilde{y}_0\}$ est libre. Montrons à présent qu'elle engendre \mathcal{S}_0 . Soit $y \in \mathcal{S}$. Notons $\lambda = y(t_0)$ et $\mu = y'(t_0)$. Alors y et $\lambda y_0 + \mu \widetilde{y}_0$ sont solutions de (E_0) et vérifient les mêmes conditions en t_0 :

$$y(t_0) = (\lambda y_0 + \mu \widetilde{y}_0)(t_0) = \lambda, \quad y'(t_0) = (\lambda y_0 + \mu \widetilde{y}_0)'(t_0) = \mu.$$

Par unicité dans le théorème de Cauchy-Lipschitz, on en déduit $y = \lambda y_0 + \mu \widetilde{y}_0$. Donc la famille $\{y_0, \widetilde{y}_0\}$ est génératrice de \mathcal{S}_0 .

Conclusion : \mathcal{S}_0 est de dimension 2. □

Corollaire 6 Si on connaît 2 solutions indépendantes de (E_0) notées y_{01} et y_{02} , alors

$$\mathcal{S}_0 := \{y \in \mathcal{D}(I; \mathbb{R}) : \text{il existe } \alpha_1, \alpha_2 \in \mathbb{R} \text{ tels que } y = \alpha_1 y_{01} + \alpha_2 y_{02}\}.$$

Preuve : Notons provisoirement \mathcal{V} l'ensemble figurant dans le membre de droite de l'égalité précédente. C'est l'espace vectoriel engendré dans $\mathcal{D}(I; \mathbb{R})$ par les deux fonctions y_{01} et y_{02} . Comme ces deux fonctions sont linéairement indépendantes, \mathcal{V} est de dimension 2. Comme y_{01} et y_{02} appartiennent à l'espace vectoriel \mathcal{S}_0 , \mathcal{V} est contenu dans \mathcal{S}_0 . Or on sait que \mathcal{S}_0 est de dimension 2, d'où l'égalité des deux ensembles. □

Exemple 33 Résoudre l'équation sur $I =]0; +\infty[$:

$$t^2 y'' + t y' - y = -3.$$

On se ramène pour commencer à l'équation sous forme résolue : comme $1/t^2$ ne s'annule pas sur I , l'équation précédente est équivalente à

$$y'' + \frac{1}{t} y' - \frac{1}{t^2} y = \frac{-3}{t^2}.$$

On résout dans un premier temps l'équation homogène :

$$y'' + \frac{1}{t} y' - \frac{1}{t^2} y = 0.$$

On tente la recherche de solutions de la forme $t \mapsto t^\alpha$ avec α à choisir :

$$\alpha(\alpha - 1)t^{\alpha-2} + \alpha t^{\alpha-2} - t^{\alpha-2} = 0 \iff (\alpha^2 - 1)t^{\alpha-2} = 0.$$

On trouve ainsi deux solutions de (E_0) : $y_{01} = t$ et $y_{02} = 1/t$. On en déduit que l'ensemble des solutions de l'équation homogène est

$$\mathcal{S}_0 = \left\{ y : I \rightarrow \mathbb{R} : \text{il existe } \alpha_1, \alpha_2 \in \mathbb{R} \text{ tels que } y = \alpha_1 t + \frac{\alpha_2}{t} \right\}.$$

On passe à l'équation initiale. Une solution particulière évidente est $y_p = 3$. On en déduit que l'ensemble des solutions de l'équation est

$$\mathcal{S} = \left\{ y : I \rightarrow \mathbb{R} : \text{il existe } \alpha_1, \alpha_2 \in \mathbb{R} \text{ tels que } y = 3 + \alpha_1 t + \frac{\alpha_2}{t} \right\}$$

4.4.2 Wronskien

Définition 38 Soient y_{01}, y_{02} deux solutions de (E_0)

$$y'' + a_1 y' + a_2 y = 0$$

où a_1 et a_2 sont des fonctions continues sur un intervalle I . Le Wronskien de (y_{01}, y_{02}) est la fonction

$$W_{y_{01}, y_{02}} = y_{01} y'_{02} - y'_{01} y_{02} = \det \begin{pmatrix} y_{01} & y_{02} \\ y'_{01} & y'_{02} \end{pmatrix}.$$

Proposition 37 Le Wronskien $W_{y_{01}, y_{02}}$ est solution de l'équation linéaire du premier ordre

$$w' + a_1 w = 0.$$

On en déduit que $W_{y_{01}, y_{02}}$ est de signe constant. En particulier, il s'annule en un point si et seulement si il s'annule partout.

L'intérêt fondamental du Wronskien réside dans la proposition suivante :

Proposition 38 Les solutions y_{01}, y_{02} sont linéairement indépendantes si et seulement si il existe $t_0 \in I$ tel que $W_{y_{01}, y_{02}}(t_0) \neq 0$.

Preuve : Si y_{01} et y_{02} sont linéairement dépendantes, alors il existe $\lambda, \mu \in \mathbb{R}$ tels que

$$\lambda y_{01} + \mu y_{02} = 0.$$

Cela signifie que pour tout $t \in I$, $\lambda y_{01}(t) + \mu y_{02}(t) = 0$. En particulier, on peut dériver cette identité :

$$\lambda y'_{01}(t) + \mu y'_{02}(t) = 0.$$

Ainsi,

$$\lambda \begin{pmatrix} y_{01} \\ y'_{01} \end{pmatrix}(t) + \begin{pmatrix} y_{02} \\ y'_{02} \end{pmatrix}(t) = 0.$$

Ceci montre que pour tout $t \in I$, la matrice

$$\begin{pmatrix} y_{01}(t) & y_{02}(t) \\ y'_{01}(t) & y'_{02}(t) \end{pmatrix}$$

est de rang au plus 1, et donc son déterminant, qui est justement $W_{y_{01}, y_{02}}$, est nul sur I .

Réciproquement, si $W_{y_{01}, y_{02}}(t) = 0$ pour tout $t \in I$, fixons $t_0 \in I$. La matrice

$$\begin{pmatrix} y_{01}(t_0) & y_{02}(t_0) \\ y'_{01}(t_0) & y'_{02}(t_0) \end{pmatrix}$$

est de déterminant nul, donc ses colonnes sont liées. On peut supposer par exemple que la première est un multiple de la seconde : il existe $\lambda \in \mathbb{R}$ tel que

$$\begin{pmatrix} y_{01}(t_0) \\ y'_{01}(t_0) \end{pmatrix} = \lambda \begin{pmatrix} y_{02}(t_0) \\ y'_{02}(t_0) \end{pmatrix},$$

autrement dit

$$y_{01}(t_0) = \lambda y_{02}(t_0) \quad , \quad y'_{01}(t_0) = \lambda y'_{02}(t_0).$$

Ainsi y_{01} et λy_{02} sont deux fonctions solutions de (E_0) qui coïncident ainsi que leur dérivée en t_0 . Par l'unicité dans le théorème de Cauchy-Lipschitz, on en déduit qu'elles sont égales : pour tout $t \in I$,

$$y_{01}(t) = \lambda y_{02}(t)$$

Ainsi y_{01} et y_{02} sont linéairement dépendantes. □

Le Wronskien est donc un outil très utile pour montrer que deux solutions de l'équation homogène sont linéairement indépendantes.

4.4.3 Méthode de variations de la constante

On considère l'équation

$$(E) \quad y'' + a(t)y' + b(t)y = f(t)$$

où a, b et f sont trois fonctions continues sur un intervalle I . On suppose que l'on connaît une base $\{y_{01}, y_{02}\}$ des solutions \mathcal{S}_0 de l'équation homogène (E_0) . Pour achever la résolution de (E) , on cherche donc une solution particulière y_p de (E) .

Étape 1 On cherche y_p sous la forme $y_p = \alpha(t)y_{01}(t) + \beta(t)y_{02}(t)$ en imposant que $0 = \alpha'y_{01} + \beta'y_{02}$. On a alors

$$\begin{aligned} y_p' &= \alpha y_{01}' + \beta y_{02}' + \alpha' y_{01} + \beta' y_{02} = \alpha y_{01}' + \beta y_{02}', \\ y_p'' &= \alpha' y_{01}' + \beta' y_{02}' + \alpha y_{01}'' + \beta y_{02}'' . \end{aligned}$$

Alors

$$\begin{aligned} y_p'' + a(t)y_p' + b(t)y_p &= \alpha' y_{01}' + \beta' y_{02}' + \alpha y_{01}'' + \beta y_{02}'' + a(t)(\alpha y_{01}' + \beta y_{02}') \\ &\quad + b(t)(\alpha y_{01}(t) + \beta y_{02}(t)) \\ &= \alpha(y_{01}'' + a(t)y_{01}' + b(t)y_{01}) + \beta(y_{02}'' + a(t)y_{02}' + b(t)y_{02}) + \alpha' y_{01}' + \beta' y_{02}' . \end{aligned}$$

En utilisant que y_{01} et y_{02} sont solutions de l'équation, on en déduit :

$$y_p'' + a(t)y_p' + b(t)y_p = \alpha' y_{01}' + \beta' y_{02}' .$$

Ainsi, y_p est une solution particulière si et seulement si α et β vérifient

$$\begin{cases} \alpha' y_{01}' + \beta' y_{02}' = f \\ \alpha y_{01}' + \beta y_{02}' = 0 \end{cases} .$$

Étape 2 On résout le système précédent de deux équations à deux inconnues, les inconnues étant les fonctions α' et β' . Observez que le déterminant de ce système est précisément le Wronskien $W_{y_{01}, y_{02}}(t)$. Comme on est parti de deux solutions y_{01} et y_{02} linéairement indépendantes, on sait que le Wronskien ne s'annule jamais. Ceci garantit que pour tout $t \in I$, le système précédent a toujours une unique solution $(\alpha'(t), \beta'(t))$.

Étape 3 En prenant les primitives des fonctions obtenues à l'étape précédente, on en déduit des fonctions α et β et on peut alors affirmer que $y_p = \alpha y_{01} + \beta y_{02}$ est une solution particulière de (E) .

Exemple 34 Résoudre $y'' + \frac{1}{t}y' - \frac{1}{t^2}y = \cos t$ sur $]0; +\infty[$.

Réponse : $y_{01}(t) = t, y_{02}(t) = \frac{1}{t}$ et $y_p = -\cos t + \frac{\sin t}{t}$.

4.4.4 Cas particulier des équations à coefficients constants

En général, on ne sait pas résoudre les équations du second ordre homogène. Il y a pourtant un cas particulier très important où c'est toujours possible : lorsque les coefficients sont constants.

On s'intéresse ici aux équations de la forme

$$(E) \quad y'' + ay' + by = f$$

où a, b sont deux réels et $f : I \rightarrow \mathbb{R}$ est une fonction continue. Comme d'habitude, on considère l'équation homogène

$$(E_0) \quad y'' + ay' + by = 0.$$

On introduit l'équation caractéristique : $r^2 + ar + b = 0$.

Théorème 20 1. *S'il existe deux racines réelles distinctes r_1, r_2 alors*

$$\mathcal{S}_0 = \{y : \mathbb{R} \rightarrow \mathbb{R} : \text{il existe } \alpha_1, \alpha_2 \in \mathbb{R} \text{ tels que } y = \alpha_1 e^{r_1 t} + \alpha_2 e^{r_2 t}\}.$$

2. *S'il existe une racine double r alors*

$$\mathcal{S}_0 = \{y : \mathbb{R} \rightarrow \mathbb{R} : \text{il existe } \alpha, \beta \in \mathbb{R} \text{ tels que } y = \alpha e^{rt} + \beta t e^{rt}\}.$$

3. *S'il existe deux racines complexes conjuguées $r = \sigma \pm i\omega$, alors*

$$\mathcal{S}_0 = \{y : \mathbb{R} \rightarrow \mathbb{R} : \text{il existe } \alpha, \beta \in \mathbb{R} \text{ tels que } y = \alpha e^{\sigma t} \cos \omega t + \beta e^{\sigma t} \sin \omega t\}.$$

Preuve : On sait que \mathcal{S}_0 est un espace vectoriel de dimension 2. Dans chaque cas, il suffit donc de montrer qu'on dispose de deux solutions linéairement indépendantes.

1. Cas 1 : deux racines réelles distinctes r_1 et r_2 . On pose $y_{0j}(t) = e^{r_j t}$, $j = 1, 2$. On vérifie que

$$y''_{0j}(t) + ay'_{0j}(t) + by_{0j}(t) = e^{r_j t}(r_j^2 + ar_j + b) = 0.$$

De plus, montrons que y_{01} et y_{02} sont linéairement indépendantes. Soient $\lambda, \mu \in \mathbb{R}$ tels que $\lambda y_{01} + \mu y_{02} = 0$. On évalue cette relation en 0 et en 1. On en déduit que λ et μ sont solutions du système :

$$\begin{cases} \lambda + \mu & = & 0 \\ \lambda e^{r_1} + \mu e^{r_2} & = & 0. \end{cases}$$

Le déterminant de ce système $e^{r_2} - e^{r_1}$ est non nul (car $r_1 \neq r_2$) et donc $\lambda = \mu = 0$. Ainsi les fonctions y_{01} et y_{02} sont linéairement indépendantes¹. Comme \mathcal{S}_0 est de dimension 2, on en déduit que $\{y_{01}, y_{02}\}$ est une base de solutions.

2. Cas 2 : une racine double $r = -a/2$. Comme précédemment, $y_{01}(t) = e^{rt}$ est solution de l'équation. Montrons que $y_{02}(t) = t e^{rt}$ l'est aussi. On a

$$\begin{aligned} y''_{02}(t) + ay'_{02}(t) + by_{02}(t) &= e^{rt}(tr^2 + 2r + atr + a + bt) \\ &= e^{rt}(t(r^2 + ar + b) + 2r + a) \\ &= 0. \end{aligned}$$

Dans la dernière ligne, on a utilisé le fait que r est racine et que $r = -a/2$. On vérifie comme dans le premier cas que y_{01} et y_{02} sont linéairement indépendantes. Une base de solutions est donc $\{y_{01}, y_{02}\}$.

3. Cas 3 : deux racines complexes conjuguées $r_{\pm} = \sigma \pm i\omega$, avec $\omega \neq 0$. On introduit ici une démarche nouvelle. On va chercher dans un premier temps des solutions complexes avant d'en déduire deux solutions réelles linéairement indépendantes. En fait, exactement comme dans le premier cas (les calculs sont les mêmes), on vérifie que $z_{\pm}(t) = e^{r_{\pm} t}$ sont solutions de l'équation. On pose alors

$$y_{01}(t) = \frac{z_+ + z_-}{2} = e^{\sigma t} \cos \omega t, \quad y_{02}(t) = \frac{z_+ - z_-}{2i} = e^{\sigma t} \sin \omega t.$$

Les fonctions y_{01} et y_{02} sont solutions de (E_0) comme combinaison linéaire de solutions de (E_0) . De plus, ce sont des fonctions à valeurs réelles. En évaluant une relation linéaire liant y_{01} et y_{02} en 0 et $\frac{\pi}{2\omega}$ comme dans le premier cas, on en déduit que $\{y_{01}, y_{02}\}$ est une famille libre, qui engendre donc l'espace de dimension 2 des solutions \mathcal{S}_0 . □

Une fois qu'on a obtenu deux solutions y_{01} et y_{02} linéairement indépendantes de l'équation homogène, on peut chercher une solution particulière y_p de l'équation complète (E) grâce à la méthode de variation des constantes. Cependant, il y a un cas où on peut procéder rapidement : lorsque le second membre est en exponentielle-polynôme. On se place donc dans le cas particulier :

$$(E) \quad y'' + ay' + by = P(t)e^{\lambda t}$$

avec $a, b, \lambda \in \mathbb{R}$ et P est un polynôme.

Proposition 39 *L'équation (E) admet une solution particulière de la forme*

$$y_p(t) = Q(t)e^{\lambda t}$$

où Q est un polynôme. De plus,

1. si λ n'est pas racine de l'équation caractéristique $r^2 + ar + b = 0$, $\deg Q = \deg P$,
2. si λ est racine simple, $\deg Q = \deg P + 1$,
3. si λ est racine double, $\deg Q = \deg P + 2$.

1. Alternativement, on aurait pu calculer le Wronskien de y_{01} et y_{02} , qui vaut $(r_2 - r_1)(e^{(r_1+r_2)t})$, et donc ne s'annule jamais.

Preuve : Insérons $y_p(t) = Q(t)e^{\lambda t}$ dans l'équation :

$$y_p'' + ay_p' + by_p = ((\lambda^2 + a\lambda + b)Q + (2\lambda + a)Q' + Q'')e^{\lambda t}.$$

Donc y_p est une solution particulière si et seulement si

$$((\lambda^2 + a\lambda + b)Q + (2\lambda + a)Q' + Q'') = P$$

Posons $\alpha = \lambda^2 + a\lambda + b$ et $\beta = 2\lambda + a$. Il s'agit donc de montrer que pour tout polynôme P , il existe un polynôme Q solution de l'équation : $\alpha Q + \beta Q' + Q'' = P$. Notons $d \in \mathbb{N}$ le degré de P et $\mathbb{R}_d[X]$ l'ensemble des polynômes de degré $\leq d$.

1. Cas 1 : $\alpha \neq 0$. On veut montrer que l'application $\Phi : Q \in \mathbb{R}_d[X] \mapsto \alpha Q + \beta Q' + Q'' \in \mathbb{R}_d[X]$ est surjective. Comme c'est une application linéaire en dimension finie, il est équivalent de montrer qu'elle est injective. Soit $Q \in \mathbb{R}_d[X]$ tel que $\Phi(Q) = 0$. Ainsi $Q = \frac{-\beta}{\alpha}Q' - \frac{1}{\alpha}Q''$. Comme $\deg Q' = \deg Q - 1$, $\deg Q'' = \deg Q - 2$, l'égalité précédente n'est possible que si $Q = 0$ (auquel cas $\deg Q = -\infty$). Donc Φ est injective, donc surjective, d'où l'on déduit l'existence de Q .
2. Cas 2 : $\alpha = 0$ mais $\beta \neq 0$. On veut montrer que l'application $\Psi : Q \in \mathbb{R}_{d+1}[X] \mapsto \beta Q' + Q'' \in \mathbb{R}_d[X]$ est surjective. On montre comme dans le premier cas que l'application $T \in \mathbb{R}_d[X] \mapsto \beta T + T' \in \mathbb{R}_d[X]$ est surjective. Par ailleurs, l'application de dérivation $Q \in \mathbb{R}_{d+1}[X] \mapsto Q' \in \mathbb{R}_d[X]$ est surjective. Par composition, on en déduit que Ψ est surjective.
3. Cas 3 : $\alpha = \beta = 0$. L'application $Q \in \mathbb{R}_{d+2}[X] \mapsto Q'' \in \mathbb{R}_d[X]$ est surjective (primitiver deux fois).

□

Exemple 35 $y'' - 5y' + 6y = te^t$

Réponse : $\mathcal{S} = \{y : \mathbb{R} \rightarrow \mathbb{R} : \text{il existe } \alpha, \beta \in \mathbb{R} \text{ tels que } y(t) = (\frac{t}{2} + \frac{3}{4})e^t + \alpha e^{2t} + \beta e^{3t}\}$.

4.4.5 D'autres techniques de résolution

Le principe de superposition

Proposition 40 Soient a_1, \dots, a_n des fonctions continues sur un intervalle I , et f_1, f_2 deux fonctions continues sur I . On note $f = f_1 + f_2$. Notons (E_f) l'équation : $y^{(n)} + a_1 y^{(n-1)} + \dots + a_n y = f$, et de même $(E_{f_1}), (E_{f_2})$ les équations lorsqu'on remplace le second membre f par f_1 et f_2 respectivement. Alors l'ensemble \mathcal{S}_f des solutions de (E) est égal à

$$\mathcal{S}_f := \{y : I \rightarrow \mathbb{R} : \text{il existe } y_1 \in \mathcal{S}_{f_1}, y_2 \in \mathcal{S}_{f_2} \text{ tels que } y = y_1 + y_2\},$$

où $\mathcal{S}_{f_1}, \mathcal{S}_{f_2}$ désignent l'ensemble des solutions de E_{f_1}, E_{f_2} respectivement.

La méthode de Lagrange

Lorsqu'on connaît une solution y_{01} de l'équation homogène ne s'annulant pas, alors on peut trouver l'autre en la cherchant sous la forme $t \mapsto \mu(t)y_{01}(t)$.

Application : $y'' - y'/t + y/t^2 = 0$ sur $]0, +\infty[$. On obtient deux solutions linéairement indépendantes : $y_{01}(t) = t$ et $y_{02}(t) = t \ln t$.

Recherche de solutions développables en série entière

Lorsque l'équation (mise sous forme résolue) a ses coefficients développables en série entière, alors il existe des solutions de l'équation homogène qui sont développables en série entière. On peut donc chercher une telle solution y_0 sous la forme $\sum_n a_n (t - t_0)^n$ (où t_0 est un point de I ; il est d'ailleurs souvent commode de se ramener par translation au cas $t_0 = 0$). On insère cette fonction dans l'équation, ce qui donne des conditions (nécessaires et suffisantes) sur les a_n . On en déduit l'expression des a_n et on vérifie que la série $\sum_n a_n t^n$ a un rayon de convergence non nul. Avec cette méthode, on peut obtenir une solution ou deux solutions linéairement indépendantes de l'équation homogène.

Application : $y'' + ty' + y = 0$, $ty'' + 2y' + ty = 0$.

4.5 Systèmes différentiels linéaires

4.5.1 Théorème de Cauchy-Lipschitz

Théorème 21 (Cauchy-Lipschitz) Soit $A = [a_{ij}]_{1 \leq i, j \leq n}$ une matrice dont les n^2 coefficients sont des fonctions continues de I vers \mathbb{R} et $F : I \rightarrow \mathbb{R}^n$ une fonction continue. On considère le système

$$(V) \quad Y'(t) = A(t)Y(t) + F(t).$$

Alors pour tout $Y_0 \in \mathbb{R}^n$, il existe une unique solution de (V) telle que $Y(t_0) = Y_0$.

La solution est donc entièrement déterminée par sa valeur en t_0 .

Corollaire 7 L'ensemble des solutions du système homogène $(V_0) : Y'(t) = A(t)Y(t)$ est un sous-espace vectoriel de $\mathcal{D}(I; \mathbb{R}^n)$ de dimension n . De plus, si on connaît n solutions indépendantes de $(V_0) : Y_{01}, \dots, Y_{0n}$, alors l'ensemble des solutions de (V_0) est donné par

$$\{Y : I \rightarrow \mathbb{R}^n : \text{il existe } \alpha_1, \dots, \alpha_n \text{ tels que } Y = \alpha_1 Y_{01} + \dots + \alpha_n Y_{0n}\}.$$

Enfin, si Y_p est une solution particulière quelconque de (V), alors l'ensemble des solutions de (V) est un espace affine de dimension n donné par

$$\{Y : I \rightarrow \mathbb{R}^n : \text{il existe } Y_0 \in \mathcal{S}_0 \text{ tel que } Y = Y_p + Y_0\}.$$

Autrement dit, "solution générale de (E) = solution particulière + solution générale de (E_0) ".

4.5.2 Résolution des systèmes à coefficients constants diagonalisables

On considère le système d'équations

$$(V) \quad Y' = AY + F$$

où $A \in \mathcal{M}_n(\mathbb{R})$ est une matrice et F est une fonction continue sur un intervalle I à valeurs dans \mathbb{R}^n . On note (V_0) le système homogène associé.

On se limite au cas où A est diagonalisable. On note $\lambda_1, \dots, \lambda_n$ les valeurs propres de A comptées avec multiplicité; autrement dit, si λ_i est une racine du polynôme caractéristique de multiplicité $p \in \{1, \dots, n\}$, alors λ_i apparaît p fois dans l'énumération $\lambda_1, \dots, \lambda_n$, avec p indices différents. On note (e_1, \dots, e_n) une base de vecteurs propres associée (une telle base existe puisque A a été supposée diagonalisable).

Théorème 22 Une base de l'ensemble des solutions \mathcal{S}_0 de (E_0) est donnée par les n fonctions

$$Y_{0i}(t) = e^{\lambda_i t} e_i.$$

Preuve : Pour chaque $1 \leq i \leq n$, on a

$$Y'_{0i}(t) = e^{\lambda_i t} \lambda_i e_i = e^{\lambda_i t} A e_i = A Y_{0i}.$$

Donc chaque Y_{0i} est solution du système. Il reste à montrer que la famille $\{Y_{0i}\}$ est une base de \mathcal{S}_0 . On sait qu'il s'agit d'un espace de dimension n . Il suffit donc de montrer que la famille est libre. Soit $\{\mu_i\}_i \subset \mathbb{R}$ tel que

$$\sum_{i=1}^n \mu_i Y_{0i} = 0.$$

On évalue cette relation en $t = 0$. On obtient

$$\sum_{i=1}^n \mu_i e_i = 0.$$

Comme $\{e_i\}_i$ est une base, on en déduit que tous les μ_i sont nuls. Donc la famille $\{Y_{0i}\}$ est libre, donc une base de \mathcal{S}_0 . □

On présente à présent une autre preuve qui ne nécessite pas l'utilisation du théorème de Cauchy-Lipschitz. En fait, quand la matrice A est diagonalisable, un système d'ordre n se ramène à n équations d'ordre 1, pour la résolution desquelles on n'a pas eu à s'appuyer sur le théorème de Cauchy-Lipschitz.

Pour cela, on introduit la matrice de passage P de la base canonique de \mathbb{R}^n à la base de vecteurs propres $\{e_i\}_{1 \leq i \leq n}$, ainsi que la matrice diagonale

$$D = \begin{pmatrix} \lambda_1 & 0 & 0 \\ 0 & \ddots & 0 \\ 0 & \dots & \lambda_n \end{pmatrix}.$$

On sait que $A = PDP^{-1}$. L'équation $Y' = AY$ se réécrit $Z' = DZ$ en posant $Z = P^{-1}Y$ ou de manière plus explicite :

$$z'_i(t) = \lambda_i z_i(t).$$

Ceci est équivalent à : pour tout $1 \leq i \leq n$, il existe $\alpha_i \in \mathbb{R}$ tel que

$$z_i = \alpha_i e^{\lambda_i t}.$$

Ainsi,

$$Z(t) = \begin{pmatrix} \alpha_1 e^{\lambda_1 t} \\ \vdots \\ \alpha_n e^{\lambda_n t} \end{pmatrix}.$$

Comme $Y(t) = PZ(t)$ et que les colonnes de la matrice P donnent les coordonnées des vecteurs de la nouvelle base $\{e_i\}_{1 \leq i \leq n}$ dans l'ancienne (la base canonique), on a donc par définition de la multiplication matricielle (et en identifiant un vecteur de \mathbb{R}^n avec le vecteur colonne de ses coordonnées dans la base canonique)

$$Y(t) = \sum_{i=1}^n \alpha_i e^{\lambda_i t} e_i = \sum_{i=1}^n \alpha_i Y_{0i}(t)$$

en posant comme ci-dessus $Y_{0i}(t) = e^{\lambda_i t} e_i$. Ceci montre donc que l'ensemble des solutions du système homogène est engendré par les vecteurs Y_{0i} (sans utiliser que cet ensemble est un espace vectoriel de dimension n).

Exemple 36

$$A = \begin{pmatrix} 2 & 0 & 0 \\ 2 & -1 & 3 \\ -2 & 3 & -1 \end{pmatrix}$$

Réponse : $Y_{011} = (3, 2, 0)e^{2t}$, $Y_{012} = (-3, 0, 2)e^{2t}$, $Y_{021} = (0, 1, -1)e^{-4t}$.

4.5.3 Méthode de variation des constantes

On considère ici le système

$$Y'(t) = AY(t) + F(t).$$

où F est une fonction continue sur un intervalle I . On suppose connue une base de solutions du système homogène $\{Y_{01}, \dots, Y_{0n}\}$. On cherche alors une solution particulière sous la forme

$$Y_p(t) = \sum_{i=1}^n \alpha_i Y_{0i}.$$

Une telle fonction est solution de $Y_p' = AY_p + F$ si et seulement si

$$\sum_{i=1}^n (\alpha_i'(t) Y_{0i}(t) + \alpha_i(t) Y_{0i}'(t)) = A \left(\sum_{i=1}^n \alpha_i(t) Y_{0i}(t) \right) + F(t).$$

Comme $A \left(\sum_{i=1}^n \alpha_i(t) Y_{0i}(t) \right) = \sum_{i=1}^n \alpha_i(t) AY_{0i}(t)$, on en déduit

$$\sum_{i=1}^n (\alpha_i'(t) Y_{0i}(t) + \alpha_i(t) Y_{0i}'(t)) = \sum_{i=1}^n \alpha_i(t) AY_{0i}(t) + F(t).$$

Comme chaque Y_{0i} est solution de $Y_{0i}' = AY_{0i}$, il vient

$$\sum_{i=1}^n \alpha_i'(t) Y_{0i}(t) = F(t).$$

Il s'agit d'un système de n équations à n inconnues $\alpha_1'(t), \dots, \alpha_n'(t)$. On peut montrer qu'il a une unique solution (pourvu que les Y_{0i} soient bien linéairement indépendants). Une fois résolu, on trouve une primitive aux solutions de ce système pour obtenir des fonctions $\alpha_1, \dots, \alpha_n$ convenables.

On peut alors affirmer que $Y_p(t) = \sum_{i=1}^n \alpha_i Y_{0i}$ est une solution particulière du système $Y'(t) = AY(t) + F(t)$.

Exemple 37 Résoudre $Y' = AY + F$ avec

$$A = \begin{pmatrix} 4 & 2 \\ -3 & -1 \end{pmatrix}, \quad F(t) = \begin{pmatrix} e^t \cos t \\ e^t \end{pmatrix}.$$

Solution :

$$\mathcal{S} = \left\{ \alpha e^{2t} \begin{pmatrix} 1 \\ -1 \end{pmatrix} + \beta e^t \begin{pmatrix} 2 \\ -3 \end{pmatrix} + (\sin t - t) e^{2t} \begin{pmatrix} 1 \\ -1 \end{pmatrix} + \left(-2 + \frac{3}{2} \cos t - \frac{3}{2} \sin t\right) \begin{pmatrix} 2 \\ -3 \end{pmatrix} \right\}.$$

La méthode de variation de la constante pour les équations linéaires d'ordre 2 apparaît comme un cas particulier de la méthode de variation de la constante pour les systèmes. En effet, une équation de la forme $y'' + ay + b = f$, où a, b, f sont des fonctions continues sur un intervalle I peut se réécrire sous la forme d'un système

$$Y' = AY + F$$

en posant

$$Y = \begin{pmatrix} y \\ y' \end{pmatrix}, \quad A = \begin{pmatrix} 0 & 1 \\ -b & -a \end{pmatrix}, \quad F(t) = \begin{pmatrix} 0 \\ f \end{pmatrix}.$$

Etant donné deux solutions linéairement indépendantes Y_{01} et Y_{02} de $Y' = AY$, la méthode de variation de la constante pour $Y' = AY + F$ revient à résoudre le système

$$\alpha'Y_{01} + \beta'Y_{02} = F,$$

soit encore

$$\begin{cases} \alpha'y_{01} + \beta'y_{02} = 0 \\ \alpha'y'_{01} + \beta'y'_{02} = f \end{cases} .$$

On reconnaît précisément le système qu'on est amené à résoudre pour les équations d'ordre 2 lorsqu'on connaît deux solutions y_{01} et y_{02} linéairement indépendantes de l'équation homogène $y'' + ay + b = 0$.

4.5.4 D'autres techniques de résolution

Principe de superposition pour les systèmes linéaires

C'est l'exact analogue des équations linéaires.

Chapitre 5

Intégrales à paramètres et intégrales multiples

Ici, \mathbb{K} désigne \mathbb{R} ou \mathbb{C} .

5.1 Régularité des intégrales à paramètres

Soient I et J deux intervalles de \mathbb{R} . Ces intervalles ne sont pas nécessairement des segments, c'est-à-dire des intervalles fermés bornés, de la forme $[a, b]$, où $a < b$ sont deux réels. Lorsqu'on considère des intégrales sur J dans le cas où J n'est pas un segment, on fait référence à la théorie des intégrales sur un intervalle quelconque étudiée au semestre précédent.

5.1.1 Rappels sur l'intégrale sur un intervalle quelconque

On suppose ici connue la définition de l'intégrale d'une fonction continue (ou continue par morceaux) sur un segment.

On se donne un intervalle quelconque J dont on note $-\infty \leq \alpha < \beta \leq +\infty$ les extrémités. On a donc $J = [\alpha, \beta]$ (dans ce cas J est un segment), ou $J =]\alpha, \beta]$, ou $J = [\alpha, \beta[$ ou $J =]\alpha, \beta[$.

1. On dit qu'une fonction continue et positive $\varphi : J \rightarrow \mathbb{R}^+$ est intégrable sur J s'il existe $M > 0$ tel que pour tout segment $[a, b] \subset J$,

$$\int_{[a,b]} \varphi(t) dt \leq M.$$

Si J est un segment, l'intégrabilité est automatique.

2. On dit qu'une fonction continue $\varphi : J \rightarrow \mathbb{R}$ est intégrable sur J si $|f|$ (qui est une fonction continue positive) est intégrable sur J .
3. Dans ce cas, pour tout $x_0 \in J$, les limites suivantes existent :

$$\lim_{y \rightarrow \beta} \int_{x_0}^y \varphi(t) dt \quad , \quad \lim_{x \rightarrow \alpha} \int_x^{x_0} \varphi(t) dt$$

et on note

$$\int_{\alpha}^{\beta} \varphi(t) dt = \lim_{x \rightarrow \alpha} \int_x^{x_0} \varphi(t) dt + \lim_{y \rightarrow \beta} \int_{x_0}^y \varphi(t) dt.$$

4. Il se peut que les limites précédentes existent sans que φ ne soit intégrable. Dans ce cas, on note encore

$$\int_{\alpha}^{\beta} f(t) dt = \lim_{x \rightarrow \alpha} \int_x^{x_0} \varphi(t) dt + \lim_{y \rightarrow \beta} \int_{x_0}^y \varphi(t) dt$$

et on dit que l'intégrale est semi-convergente.

En pratique, étant donné une fonction φ définie sur un intervalle $J \subset \mathbb{R}$, on procède selon la démarche suivante :

1. On s'assure que la fonction est continue par morceaux sur l'intervalle J .
2. Si l'extrémité gauche de J n'est pas $-\infty$ et n'appartient pas à J (autrement dit, J est de la forme $] \alpha, \beta[$ ou $] \alpha, \beta]$ avec $\alpha \in \mathbb{R}$), on étudie la possibilité de prolonger φ par continuité en α . Si φ est continue en α ou peut être prolongée par continuité en α , alors φ est automatiquement intégrable sur tout segment de la forme $] \alpha, x_0]$, pour tout $x_0 \in J$.
3. On fait le même travail en β .
4. Si $\alpha = -\infty$ ou bien s'il n'a pas été possible de prolonger φ par continuité en α , on fixe un $x_0 \in] \alpha, \beta[$ et on se donne un $a \in] \alpha, x_0[$. Comme φ est continue par morceaux sur le segment $[a, x_0]$, on peut considérer les intégrales $\int_a^{x_0} |\varphi|$ et $\int_a^{x_0} \varphi$.
 - (a) S'il existe un majorant de $\int_a^{x_0} |\varphi|$ qui soit indépendant de a , alors φ sera intégrable sur $] \alpha, x_0]$. C'est en fait le cas si et seulement si la fonction $a \mapsto \int_a^{x_0} |\varphi|$ a une limite quand a tend vers α . Dans ce cas, on peut affirmer que φ est intégrable sur $] \alpha, x_0]$ et

$$\int_{\alpha}^{x_0} \varphi = \lim_{a \rightarrow \alpha} \int_a^{x_0} \varphi.$$

- (b) Si φ n'est pas intégrable sur $] \alpha, x_0]$, on regarde si la fonction $a \mapsto \int_a^{x_0} \varphi$ a une limite quand a tend vers α . Si c'est le cas, l'intégrale $\int_{\alpha}^{x_0} \varphi$ est semi-convergente.
5. On fait le même travail en β .
6. Si φ est intégrable sur $] \alpha, x_0]$ et sur $[x_0, \beta[$, alors φ est intégrable sur $] \alpha, \beta[$ et

$$\int_{\alpha}^{\beta} \varphi = \int_{\alpha}^{x_0} \varphi + \int_{x_0}^{\beta} \varphi.$$

Si $\int_{\alpha}^{x_0} \varphi$ et $\int_{x_0}^{\beta} \varphi$ sont semi-convergentes, alors $\int_{\alpha}^{\beta} \varphi$ est semi-convergente et

$$\int_{\alpha}^{\beta} \varphi = \int_{\alpha}^{x_0} \varphi + \int_{x_0}^{\beta} \varphi.$$

Exemple 38 La fonction $t \mapsto \frac{1}{t^2}$ est absolument intégrable sur l'intervalle $[1, +\infty[$ mais ne l'est pas sur l'intervalle $]0, +\infty[$.

En effet, la fonction est continue et positive sur $]0, +\infty[$. De plus, pour tout $b > 1$,

$$\int_1^b \frac{dt}{t^2} = 1 - \frac{1}{b}.$$

Cette quantité est majorée par 1 (qui ne dépend pas de b). Donc la fonction $t \mapsto \frac{1}{t^2}$ est intégrable sur $[1, +\infty[$, d'intégrale :

$$\int_1^{+\infty} \frac{dt}{t^2} = \lim_{b \rightarrow +\infty} \left(1 - \frac{1}{b}\right) = 1.$$

Maintenant, pour tout $\varepsilon \in]0, 1[$,

$$\int_{\varepsilon}^1 \frac{dt}{t^2} = \frac{1}{\varepsilon} - 1.$$

Cette quantité tend vers $+\infty$ quand ε tend vers 0. En particulier, elle ne peut être majorée indépendamment de ε . On en déduit que $t \mapsto \frac{1}{t^2}$ n'est pas intégrable sur $]0, +\infty[$.

Exemple 39 La fonction $t \mapsto \frac{\cos t}{t^2}$ est intégrable sur $[1, +\infty[$ mais pas sur $]0, +\infty[$.

La fonction $\varphi : t \mapsto \frac{\cos t}{t^2}$ est continue sur $]0, +\infty[$. Soit $b > 1$. Comme pour tout $t \in [1, +\infty[$, $|\varphi(t)| \leq 1/t^2$,

$$\int_1^b |\varphi(t)| dt \leq \int_1^b \frac{dt}{t^2}.$$

La fonction $t \mapsto \frac{1}{t^2}$ est intégrable sur $[1, +\infty[$ d'après l'exemple précédent. Donc $\int_1^b \frac{dt}{t^2}$ peut être majoré indépendamment de b , par exemple (l'intégrande $t \mapsto 1/t^2$ étant à valeurs positives) par $\int_1^{+\infty} \frac{dt}{t^2}$.

On en déduit que $|\varphi|$ est intégrable sur $[1, +\infty[$ et donc φ aussi. Montrons à présent que φ n'est pas intégrable sur $]0, 1]$. Comme la fonction \cos est décroissante sur $[0, 1]$, pour tout $t \in [0, 1]$,

$$\frac{|\cos t|}{t^2} = \frac{\cos t}{t^2} \geq \frac{\cos 1}{t^2}.$$

On en déduit que pour tout $\varepsilon \in]0, 1[$,

$$\int_{\varepsilon}^1 \frac{|\cos t|}{t^2} dt \geq \int_{\varepsilon}^1 \frac{\cos 1}{t^2} dt = \left(\frac{1}{\varepsilon} - 1\right) \cos 1.$$

Donc $\lim_{\varepsilon \rightarrow 0} \int_{\varepsilon}^1 |\varphi| = +\infty$, ce qui montre que φ n'est pas intégrable sur $]0, 1]$. En fait, l'intégrale de φ n'est même pas semi-convergente sur $]0, 1]$. En effet, par intégration par parties,

$$\begin{aligned} \int_{\varepsilon}^1 \frac{\cos t}{t^2} dt &= \left[\frac{-\cos t}{t} \right]_{\varepsilon}^1 - \int_{\varepsilon}^1 \frac{\sin t}{t} dt \\ &= \frac{\cos \varepsilon}{\varepsilon} - \cos 1 - \int_{\varepsilon}^1 \frac{\sin t}{t} dt. \end{aligned}$$

D'une part,

$$\lim_{\varepsilon \rightarrow 0} \frac{\cos \varepsilon}{\varepsilon} = +\infty.$$

D'autre part, la fonction $t \mapsto \frac{\sin t}{t}$ est une fonction continue sur $]0, 1]$ qui se prolonge par continuité en 0. Donc c'est une fonction intégrable sur $[0, 1]$ et

$$\lim_{\varepsilon \rightarrow 0} \int_{\varepsilon}^1 \frac{\sin t}{t} dt = \int_0^1 \frac{\sin t}{t} dt.$$

On en déduit que

$$\lim_{\varepsilon \rightarrow 0} \int_{\varepsilon}^1 \frac{\cos t}{t^2} dt = +\infty.$$

Conclusion : l'intégrale de φ n'est pas semi-convergente sur $]0, 1]$.

Exemple 40 La fonction $\psi : t \mapsto \frac{\sin t}{t}$ est d'intégrale convergente sur l'intervalle $]0, +\infty[$ mais elle n'est pas absolument convergente sur ce même intervalle.

La fonction ψ est continue sur $]0, +\infty[$ et se prolonge par continuité en 0. Il suffit donc d'étudier son intégrabilité en $+\infty$. Soit $k \in \mathbb{N}^*$. On écrit

$$\int_0^{2k\pi} \frac{|\sin t|}{t} dt = \sum_{\ell=0}^{k-1} \int_{2\ell\pi}^{2(\ell+1)\pi} \frac{|\sin t|}{t} dt.$$

Pour tout $\ell \in \{0, \dots, k-1\}$, par 2π périodicité de \sin ,

$$\int_{2\ell\pi}^{2(\ell+1)\pi} \frac{|\sin t|}{t} dt = \int_0^{2\pi} \frac{|\sin t|}{t + 2\ell\pi} dt.$$

L'intégrande étant positif,

$$\int_{2\ell\pi}^{2(\ell+1)\pi} \frac{|\sin t|}{t} dt \geq \int_{\pi/4}^{3\pi/4} \frac{|\sin t|}{t + 2\ell\pi} dt = \int_{\pi/4}^{3\pi/4} \frac{\sin t}{t + 2\ell\pi} dt.$$

Sur l'intervalle $[\pi/4, 3\pi/4]$, la fonction \sin est minorée par $\frac{\sqrt{2}}{2}$. Donc

$$\begin{aligned} \int_{\pi/4}^{3\pi/4} \frac{\sin t}{t + 2\ell\pi} dt &\geq \frac{\sqrt{2}}{2} \int_{\pi/4}^{3\pi/4} \frac{dt}{t + 2\ell\pi} \\ &= \frac{\sqrt{2}}{2} \ln \left(\frac{\frac{3\pi}{4} + 2\ell\pi}{\frac{\pi}{4} + 2\ell\pi} \right). \end{aligned}$$

Introduisons

$$u_\ell = \ln \left(\frac{\frac{3\pi}{4} + 2\ell\pi}{\frac{\pi}{4} + 2\ell\pi} \right), \quad \ell \geq 0.$$

Ainsi

$$\int_{2\ell\pi}^{2(\ell+1)\pi} \frac{|\sin t|}{t} dt \geq \frac{\sqrt{2}}{2} u_\ell.$$

On a donc montré que

$$\int_0^{2k\pi} \frac{|\sin t|}{t} dt \geq \frac{\sqrt{2}}{2} \sum_{\ell=0}^{k-1} u_\ell. \quad (5.1)$$

Montrons que la série $\sum_{\ell} u_\ell$ diverge. D'abord, $u_\ell \geq 0$. Ensuite,

$$\begin{aligned} u_\ell &= \ln \left(\frac{1 + \frac{3}{8\ell}}{1 + \frac{1}{8\ell}} \right) \\ &= \ln \left(1 + \frac{3}{8\ell} \right) - \ln \left(1 + \frac{1}{8\ell} \right) \\ &= \left(\frac{3}{8\ell} + O\left(\frac{1}{\ell^2}\right) \right) - \left(\frac{1}{8\ell} + O\left(\frac{1}{\ell^2}\right) \right) \\ &= \frac{1}{4\ell} + O\left(\frac{1}{\ell^2}\right). \end{aligned}$$

On en déduit que $u_\ell \sim \frac{1}{4\ell}$ pour $\ell \rightarrow +\infty$. Par comparaison de séries à termes positifs, la série $\sum_{\ell} u_\ell$ diverge et on a

$$\lim_{k \rightarrow +\infty} \sum_{\ell=0}^{k-1} u_\ell = +\infty.$$

Revenant à (5.1), on obtient

$$\lim_{k \rightarrow +\infty} \int_0^{2k\pi} \frac{|\sin t|}{t} dt = +\infty.$$

Il est donc impossible de majorer $\int_0^b \frac{|\sin t|}{t} dt$ indépendamment de b . Conclusion : La fonction ψ n'est pas intégrable sur $[0, +\infty[$.

On montre à présent que ψ a une intégrale semi-convergente sur $[1, +\infty[$. Par intégration par parties,

$$\begin{aligned} \int_1^b \frac{\sin t}{t} dt &= \left[\frac{-\cos t}{t} \right]_1^b - \int_1^b \frac{\cos t}{t^2} dt \\ &= \cos 1 - \frac{\cos b}{b} - \int_1^b \frac{\cos t}{t^2} dt. \end{aligned}$$

Par l'exemple précédent, la fonction $t \mapsto \frac{\cos t}{t^2}$ est intégrable sur $[1, +\infty[$ donc $b \mapsto \int_1^b \frac{\cos t}{t^2} dt$ a une limite en $+\infty$ et

$$\lim_{b \rightarrow +\infty} \int_1^b \frac{\cos t}{t^2} dt = \int_1^{+\infty} \frac{\cos t}{t^2} dt.$$

Comme $|\cos b| \leq 1$, $\lim_{b \rightarrow +\infty} \frac{\cos b}{b} = 0$. On en déduit que ψ a une intégrale semi-convergente sur $[1, +\infty[$ et

$$\lim_{b \rightarrow +\infty} \int_1^b \frac{\sin t}{t} dt = \cos 1 - \int_1^{+\infty} \frac{\cos t}{t^2} dt,$$

ce qu'il fallait démontrer.

5.1.2 Continuité sous le signe intégral

Dans la suite, on considère une fonction de deux variables $f : (x, t) \in I \times J \mapsto \mathbb{K}$ et on s'intéresse à la fonction

$$F : x \mapsto \int_J f(x, t) dt.$$

Théorème 23 (Théorème de continuité sous le signe intégral) *On suppose que*

1. (continuité) la fonction f est continue sur $I \times J$,
2. (domination) il existe une fonction $\varphi : J \rightarrow \mathbb{R}$ continue et absolument intégrable sur J telle que

$$\forall (x, t) \in I \times J, |f(x, t)| \leq \varphi(t).$$

Alors la fonction F est bien définie et continue sur I .

Dans le théorème précédent, si J est un segment $[a, b]$, alors l'hypothèse de domination est superflue (en fait, elle est automatiquement vérifiée sur tout ensemble de la forme $I' \times J$, pour chaque segment $I' \subset I$, ce qui est suffisant pour obtenir la conclusion).

Exercice 13 Soit $f : [a, b] \times [c, d] \rightarrow \mathbb{K}$ une fonction continue. Montrer que la fonction

$$F : (x, y) \in [a, b] \times [c, d] \mapsto \int_a^y f(x, t) dt$$

est continue sur $[a, b] \times [c, d]$.

Notons $\| \cdot \|$ une norme quelconque sur \mathbb{R}^2 (elles sont toutes équivalentes). Soit $(x_0, y_0) \in [a, b] \times [c, d]$ et montrons que F est continue en (x_0, y_0) . Soit $\varepsilon > 0$. Pour tout $(x, y) \in [a, b] \times [c, d]$,

$$F(x, y) - F(x_0, y_0) = F(x, y) - F(x, y_0) + F(x, y_0) - F(x_0, y_0).$$

Par continuité de f en (x_0, y_0) , il existe $\eta_1 > 0$ tel que pour tout $(x, y) \in [a, b] \times [c, d]$, si $\|(x, y) - (x_0, y_0)\| \leq \eta_1$, alors

$$|f(x, y) - f(x_0, y_0)| \leq 1$$

et donc

$$|f(x, y)| \leq 1 + |f(x_0, y_0)|.$$

Pour un tel couple (x, y) , on a donc

$$\begin{aligned} |F(x, y) - F(x, y_0)| &= \left| \int_a^y f(x, t) dt - \int_a^{y_0} f(x, t) dt \right| \\ &\leq \left| \int_{y_0}^y |f(x, t)| dt \right| \\ &\leq (1 + |f(x_0, y_0)|) |y - y_0|. \end{aligned}$$

Si on exige de plus $|y - y_0| \leq \frac{\varepsilon}{2(1+|f(x_0, y_0)|)}$, alors

$$|F(x, y) - F(x, y_0)| \leq \frac{\varepsilon}{2}.$$

Comme la fonction $(x, t) \mapsto f(x, t)$ est continue sur $[a, b] \times [c, d]$ et qu'on intègre sur le segment $[a, y_0]$, on peut appliquer le théorème de continuité sous le signe intégral à la fonction $x \mapsto F(x, y_0) = \int_a^{y_0} f(x, t) dt$. Il existe donc $\eta_2 > 0$ tel que pour tout $x \in [a, b]$, si $|x - x_0| \leq \eta_2$, alors

$$|F(x, y_0) - F(x_0, y_0)| \leq \frac{\varepsilon}{2}.$$

Récapitulons : notant $\eta = \min\left(\eta_1, \eta_2, \frac{\varepsilon}{2(1+|f(x_0, y_0)|)}\right)$, si $\|(x, y) - (x_0, y_0)\| \leq \eta$, alors $|y - y_0| \leq \frac{\varepsilon}{2(1+|f(x_0, y_0)|)}$ et $|x - x_0| \leq \eta_2$. Donc

$$|F(x, y) - F(x_0, y_0)| \leq |F(x, y) - F(x, y_0)| + |F(x, y_0) - F(x_0, y_0)| \leq \frac{\varepsilon}{2} + \frac{\varepsilon}{2} = \varepsilon,$$

ce qui achève la preuve de la continuité de F en (x_0, y_0) .

5.1.3 Dérivabilité sous le signe intégral

Théorème 24 *On suppose que*

1. (continuité de la fonction et de sa dérivée partielle) la fonction f est continue sur $I \times J$ et admet une dérivée partielle $\frac{\partial f}{\partial x}$ par rapport à la première variable qui est aussi continue sur $I \times J$,
2. (domination) il existe une fonction $\varphi : J \rightarrow \mathbb{R}$ continue et absolument intégrable sur J telle que

$$\forall (x, t) \in I \times J, |f(x, t)| + \left| \frac{\partial f}{\partial x}(x, t) \right| \leq \varphi(t).$$

Alors la fonction F est bien définie et C^1 sur I , de dérivée

$$F'(x) = \int_J \frac{\partial f}{\partial x}(x, t) dt.$$

Dans le théorème précédent, si J est un segment $[a, b]$, alors l'hypothèse de domination est superflue (en fait, elle est automatiquement vérifiée sur tout ensemble de la forme $I' \times J$, pour chaque segment $I' \subset I$, ce qui est suffisant pour obtenir la conclusion).

Exercice 14 1. Montrer que pour tout $t \in \mathbb{R}$, l'intégrale de la fonction $x \mapsto e^{-x^2} \cos(tx)$ converge sur \mathbb{R} .

2. Montrer que la fonction

$$f : t \mapsto \int_{\mathbb{R}} e^{-x^2} \cos(tx) dx$$

est continue sur \mathbb{R} .

3. Montrer que f est C^1 sur \mathbb{R} et que

$$f'(t) = - \int_{\mathbb{R}} x e^{-x^2} \sin(tx) dx.$$

4. Montrer que f est solution de l'équation $y'(t) + \frac{t}{2}y(t) = 0$.

5. On admet que $\int_{\mathbb{R}} e^{-x^2} dx = \sqrt{\pi}$. En déduire une expression explicite de f .

Exercice 15 Soit I un intervalle ouvert, $a, b \in \mathbb{R}$ et $f : I \times [a, b] \rightarrow \mathbb{R}$ une fonction continue admettant une dérivée partielle $\frac{\partial f}{\partial x}$ continue sur $I \times [a, b]$. Soient $g_1, g_2 : I \rightarrow \mathbb{R}$ deux fonctions C^1 sur I telles que pour tout $x \in I$, $g_1(x), g_2(x) \in]a, b[$. Alors

$$F : x \in I \mapsto \int_{g_1(x)}^{g_2(x)} f(x, t) dt$$

est C^1 et

$$F'(x) = g_2'(x)f(x, g_2(x)) - g_1'(x)f(x, g_1(x)) + \int_{g_1(x)}^{g_2(x)} \frac{\partial f}{\partial x}(x, t) dt.$$

Introduisons la fonction

$$G(x, y, z) = \int_y^z f(x, t) dt$$

définie pour tout $y, z \in [a, b]$ et $x \in I$. Le nombre $G(x, y, z)$ est bien défini comme intégrale de la fonction continue $t \mapsto f(x, t)$ sur le segment d'extrémités y et z . De plus,

$$F(x) = G(x, g_1(x), g_2(x)).$$

Pour montrer que F est C^1 sur I , il suffit donc de montrer que G est une fonction C^1 sur $I \times]a, b[\times]a, b[$. Fixons $t_0 \in [a, b]$. Alors par la relation de Chasles, pour tout $x \in I$, $y, z \in [a, b]$,

$$G(x, y, z) = G(x, y, t_0) + G(x, t_0, z).$$

Par l'exercice 13, la fonction

$$(x, z) \in I \times [a, b] \mapsto \int_{t_0}^z f(x, t) dt = G(x, t_0, z)$$

est continue en ses deux variables. Il en est de même de $(x, y) \mapsto G(x, y, t_0)$. On en déduit que la fonction G est continue.

Fixons $z \in [a, b]$. La fonction $(x, t) \mapsto f(x, t)$ est continue et admet une dérivée partielle continue sur $I \times [a, b]$. Par intégration sur le segment d'extrémités t_0 et z , le théorème de dérivation sous le signe intégral implique que

$$x \mapsto \int_{t_0}^z f(x, t) dt$$

est C^1 de dérivée

$$x \mapsto \int_{t_0}^z \frac{\partial f}{\partial x}(x, t) dt.$$

De même, pour y fixé, on montre que la fonction

$$x \mapsto \int_y^{t_0} f(x, t) dt$$

est C^1 de dérivée

$$x \mapsto \int_y^{t_0} \frac{\partial f}{\partial x}(x, t) dt.$$

Cela prouve que G admet une dérivée partielle par rapport à x qui est continue sur $I \times [a, b] \times [a, b]$ et

$$\frac{\partial G}{\partial x}(x, y, z) = \int_y^z \frac{\partial f}{\partial x}(x, t) dt.$$

Par ailleurs, à x fixé, la fonction

$$z \mapsto \int_{t_0}^z f(x, t) dt$$

est une primitive de la fonction continue $t \mapsto f(x, t)$. Elle est donc C^1 sur $]a, b[$ de dérivée

$$z \mapsto f(x, z).$$

De même, la fonction

$$y \mapsto \int_y^{t_0} f(x, t) dt$$

est C^1 sur $]a, b[$ de dérivée

$$y \mapsto -f(x, y).$$

On en déduit que les dérivées partielles de G par rapport à y et z existent et sont continues. Récapitulons :

$$\frac{\partial G}{\partial x}(x, y, z) = \int_y^z \frac{\partial f}{\partial x}(x, t) dt, \quad \frac{\partial G}{\partial y}(x, y, z) = -f(x, y), \quad \frac{\partial G}{\partial z}(x, y, z) = f(x, z).$$

Par la règle de différentiation des fonctions composées, on en déduit que F est C^1 et de dérivée

$$\begin{aligned} F'(x) &= \frac{\partial G}{\partial x}(x, g_1(x), g_2(x)) + \frac{\partial G}{\partial y}(x, g_1(x), g_2(x))g_1'(x) + \frac{\partial G}{\partial z}(x, g_1(x), g_2(x))g_2'(x) \\ &= \int_{g_1(x)}^{g_2(x)} \frac{\partial f}{\partial x}(x, t) dt - f(x, g_1(x))g_1'(x) + f(x, g_2(x))g_2'(x), \end{aligned}$$

ce qu'il fallait démontrer.

5.1.4 Différentielle totale exacte

On sait que toute fonction continue sur un intervalle de \mathbb{R} admet une primitive sur ce même intervalle. Est-ce que ce résultat admet une généralisation pour les fonctions de 2 variables ?

Plus précisément, on considère le problème suivant : soient $f, g : \mathbb{R}^2 \rightarrow \mathbb{R}$ deux fonctions de classe C^1 . A quelles conditions sur f et g existe-t-il une fonction H de classe C^1 sur \mathbb{R}^2 telle que

$$\frac{\partial H}{\partial x} = f, \quad \frac{\partial H}{\partial y} = g \quad ?$$

Autrement dit, quand peut-on trouver une primitive de la fonction

$$(x, y) \in \mathbb{R}^2 \mapsto (f(x, y), g(x, y)) \quad ?$$

Pour répondre à cette question, on procède par analyse et synthèse. Autrement dit, on suppose que H existe, et on déploie toutes les conséquences possibles sur f et g en espérant mettre la main sur des conditions suffisantes pour garantir l'existence de H .

Etape d'analyse Si H existe, alors ses dérivées partielles existent et sont C^1 et donc H est C^2 . Par le théorème de Schwarz, ses dérivées partielles secondes commutent, et donc

$$\frac{\partial f}{\partial y} = \frac{\partial^2 H}{\partial y \partial x} = \frac{\partial^2 H}{\partial x \partial y} = \frac{\partial g}{\partial x}$$

et on obtient donc une première condition nécessaire, qui porte sur f et g :

$$\frac{\partial f}{\partial y} = \frac{\partial g}{\partial x}.$$

Fixons $(x, y) \in \mathbb{R}^2$. Introduisons la fonction

$$h : t \mapsto H(tx, ty).$$

La fonction h est C^1 par composition de fonctions C^1 et de plus

$$h'(t) = x \frac{\partial H}{\partial x}(tx, ty) + y \frac{\partial H}{\partial y}(tx, ty) = xf(tx, ty) + yg(tx, ty).$$

En écrivant

$$h(t) - h(0) = \int_0^1 h'(s) ds,$$

on obtient donc

$$H(x, y) - H(0, 0) = \int_0^1 (xf(sx, sy) + yg(sx, sy)) ds,$$

ce qui constitue une expression explicite de la fonction H cherchée (à une constante additive près).

Synthèse Supposons que

$$\frac{\partial f}{\partial y} = \frac{\partial g}{\partial x} \tag{5.2}$$

et posons pour tout $(x, y) \in \mathbb{R}^2$,

$$H(x, y) = \int_0^1 (xf(sx, sy) + yg(sx, sy)) ds.$$

Pour (x, y) fixé, la fonction $s \mapsto xf(sx, sy) + yg(sx, sy)$ est continue sur \mathbb{R} donc le nombre $H(x, y)$ (obtenu comme l'intégrale d'une fonction continue sur un segment) est bien défini.

Pour x fixé, la fonction $\rho : (y, s) \mapsto xf(sx, sy) + yg(sx, sy)$ est continue sur \mathbb{R}^2 , admet une dérivée partielle par rapport à y :

$$\frac{\partial \rho}{\partial y}(y, s) = sx \frac{\partial f}{\partial y}(sx, sy) + sy \frac{\partial g}{\partial y}(sx, sy) + g(sx, sy)$$

qui est une fonction continue sur \mathbb{R}^2 . Par intégration sur le segment $[0, 1]$, le théorème de dérivation sous le signe intégral implique que H admet une dérivée partielle par rapport à y qui est continue par rapport à (x, y) et

$$\begin{aligned} \frac{\partial H}{\partial y}(x, y) &= \int_0^1 \frac{\partial \rho}{\partial y}(y, s) ds \\ &= \int_0^1 (sx \frac{\partial f}{\partial y}(sx, sy) + sy \frac{\partial g}{\partial y}(sx, sy) + g(sx, sy)) ds. \end{aligned}$$

En utilisant l'hypothèse (5.2), il vient

$$\frac{\partial H}{\partial y}(x, y) = \int_0^1 (sx \frac{\partial g}{\partial x}(sx, sy) + sy \frac{\partial g}{\partial y}(sx, sy) + g(sx, sy)) ds.$$

Or

$$x \frac{\partial g}{\partial x}(sx, sy) + y \frac{\partial g}{\partial y}(sx, sy) = \frac{\partial}{\partial s}(g(sx, sy)).$$

On en déduit

$$\frac{\partial H}{\partial y}(x, y) = \int_0^1 s \frac{\partial}{\partial s}(g(sx, sy)) + g(sx, sy) ds.$$

Par intégration par parties

$$\frac{\partial H}{\partial y}(x, y) = \left[sg(sx, sy) \right]_0^1 - \int_0^1 g(sx, sy) ds + \int_0^1 g(sx, sy) ds = g(x, y).$$

On montre de même que la dérivée partielle de H par rapport à x existe et vaut

$$\frac{\partial H}{\partial x}(x, y) = f(x, y).$$

Ainsi la fonction H a bien pour dérivées partielles f et g . Comme il s'agit de fonctions continues, H est C^1 sur \mathbb{R}^2 . On peut écrire

$$DH(x, y) = f(x, y)dx + g(x, y)dy, \quad \forall (x, y) \in \mathbb{R}^2,$$

en notant $dx : \mathbb{R}^2 \rightarrow \mathbb{R}$ la projection sur la première coordonnée, et $dy : \mathbb{R}^2 \rightarrow \mathbb{R}$ la projection sur la deuxième coordonnée.

5.1.5 Théorème de Fubini

Soient $[a, b]$ et $[c, d]$ deux segments de \mathbb{R} . On considère une fonction $f : [a, b] \times [c, d] \rightarrow \mathbb{K}$.

Théorème 25 *On suppose que f est continue sur $[a, b] \times [c, d]$. Alors*

$$\int_a^b \left(\int_c^d f(x, t) dt \right) dx = \int_c^d \left(\int_a^b f(x, t) dx \right) dt.$$

Preuve : Par le théorème de continuité sous le signe intégral, la fonction $x \mapsto \int_c^d f(x, t) dt$ est continue sur $[a, b]$. On peut donc considérer son intégrale sur $[a, b]$ et le membre de gauche de l'identité à démontrer est donc bien définie. Il en est de même du membre de droite.

Introduisons la fonction $F(x, y) = \int_c^y f(x, t) dt$. Alors F est continue sur $[a, b] \times [c, d]$.

De plus, par le théorème fondamental du calcul différentiel, F admet une dérivée partielle par rapport $x : \frac{\partial F}{\partial x}(x, y) = f(x, y)$, qui est continue sur $[a, b] \times [c, d]$.

Par le théorème de dérivation sous le signe intégral (cas où le domaine d'intégration est un segment, on n'a donc pas besoin de vérifier l'hypothèse de domination), on en déduit que la fonction

$$G : y \in [c, d] \mapsto \int_a^b F(x, y) dx$$

est bien définie et C^1 sur $[c, d]$, de dérivée

$$G'(y) = \int_a^b \frac{\partial F}{\partial y}(x, y) dx = \int_a^b f(x, y) dx.$$

On peut à présent conclure. D'une part, par définition de G ,

$$G(d) - G(c) = G(d) = \int_a^b \int_c^d f(x, y) dy dx.$$

D'autre part,

$$G(d) - G(c) = \int_c^d G'(y) dy = \int_c^d \int_a^b f(x, y) dx dy.$$

Ainsi on a bien

$$\int_a^b \int_c^d f(x, y) dy dx = \int_c^d \int_a^b f(x, y) dx dy.$$

□

5.2 Intégrales multiples

Dans cette section, on définit l'intégrale d'une fonction continue sur des domaines bornés de \mathbb{R}^2 . On se contentera ici d'une définition intuitive.

Soit $D \subset \mathbb{R}^2$ un borné. On découpe \mathbb{R}^2 en petits rectangles R_{ij} de la forme $]a_i, a_{i+1}[\times]b_j, b_{j+1}[$ et on ne retient que ceux qui intersectent D . On se donne pour chaque rectangle R_{ij} un point à l'intérieur (x_i, y_j) et on forme les sommes :

$$\sum_i \sum_j f(x_i, y_j) (a_{i+1} - a_i) (b_{j+1} - b_j).$$

Si quand on raffine le maillage, c'est-à-dire qu'on diminue les côtés des petits rectangles, la somme double ci-dessus tend vers une limite indépendante du découpage en petits rectangles et des points (x_i, y_j) , alors cette limite est par définition l'intégrale de f sur D . On la note

$$\iint_D f.$$

Il s'agit de l'analogie en dimension 2 de la construction de l'intégrale de Riemann en dimension 1. En effet, pour définir l'intégrale d'une fonction continue φ sur un segment $[a, b] \subset \mathbb{R}$, on découpe $[a, b]$ en petits segments $]a_i, a_{i+1}[$, on se donne un point x_i dans $]a_i, a_{i+1}[$ et on forme la somme

$$\sum_i (a_{i+1} - a_i) \varphi(x_i),$$

qui est une approximation d'autant meilleure de $\int_a^b \varphi$ que les intervalles $]a_i, a_{i+1}[$ sont petits. Les rectangles $]a_i, a_{i+1}[\times]0, \varphi(x_i)[$ sont remplacés en dimension 2 par les pavés

$$]a_i, a_{i+1}[\times]b_j, b_{j+1}[\times]0, f(x_i, y_j)[.$$

De la même manière que l'intégrale $\int_a^b \varphi(t) dt$ est égale à l'aire algébrique de la surface située entre la courbe représentant φ et l'axe des abscisses, l'intégrale $\int \int_D f$ est égal au volume compris entre la surface représentant f au-dessus de D et le plan d'équation $z = 0$.

L'intégrale multiple vérifie les propriétés suivantes :

Proposition 41 Soient f, g deux fonctions continues sur un ouvert Ω contenant la partie D bornée.

1. Pour tout $\lambda \in \mathbb{R}$,

$$\iint_D (\lambda f + g) dx dy = \lambda \iint_D f dx dy + \iint_D g dx dy.$$

2. Si D est une réunion finie d'arcs de courbes réguliers,

$$\iint_D f(x, y) dx dy = 0.$$

3. Si $D = D_1 \cup D_2 \cup \Gamma$ avec Γ réunion finie d'arcs de courbes réguliers et $D_1 \cap D_2 = \emptyset$, alors

$$\iint_D f(x, y) dx dy = \iint_{D_1} f(x, y) dx dy + \iint_{D_2} f(x, y) dx dy.$$

En pratique, pour justifier l'existence de l'intégrale d'une fonction f sur un borné D , il suffira d'arguer de la continuité de f sur un ouvert contenant D . De plus, il ne sera demandé de calculer une intégrale multiple que sur des domaines très particuliers. On en donne quelques exemples dans ce qui suit.

Soient deux fonctions continues $\phi, \psi : [a, b] \rightarrow \mathbb{R}$ telles que pour tout $x \in [a, b]$, $\phi(x) \leq \psi(x)$. On note D le sous-ensemble de \mathbb{R}^2 défini par

$$D = \{(x, y) \in \mathbb{R}^2 : a \leq x \leq b, \phi(x) \leq y \leq \psi(x)\}.$$

Soit $f : D \rightarrow \mathbb{R}$ une fonction continue. On a alors

$$\iint_D f := \int_a^b \left(\int_{\phi(x)}^{\psi(x)} f(x, y) dy \right) dx.$$

Soit maintenant deux fonctions continues $\zeta, \xi : [c, d] \rightarrow \mathbb{R}$ telles que pour tout $y \in [c, d]$, $\zeta(y) \leq \xi(y)$. On note

$$\Delta = \{(x, y) \in \mathbb{R}^2 : a \leq y \leq b, \zeta(y) \leq x \leq \xi(y)\}.$$

On a alors pour toute fonction continue $g : \Delta \rightarrow \mathbb{R}$

$$\iint_{\Delta} g := \int_a^b \left(\int_{\zeta(y)}^{\xi(y)} g(x, y) dx \right) dy.$$

Considérons le cas particulier où D et Δ sont égaux, et donc égaux à un rectangle de la forme $[a, b] \times [c, d]$, avec $\phi \equiv c$, $\psi \equiv d$, $\zeta \equiv a$ et $\xi \equiv b$. Alors par le théorème de Fubini

$$\begin{aligned} \iint_D f &= \int_a^b \left(\int_c^d f(x, y) dy \right) dx \\ &= \int_c^d \int_a^b f(x, y) dx dy \\ &= \iint_{\Delta} f. \end{aligned}$$

Comme $D = \Delta$, il est rassurant que les deux objets $\iint_D f$ et $\iint_{\Delta} f$ coïncident !

Une autre manière de calculer des intégrales multiples est de se ramener par un changement de variables à un domaine de la forme précédente. On a besoin pour cela du théorème de changement de variables.

Théorème 26 (Théorème de changement de variables) Soit $\Phi : \Omega \rightarrow D$ un C^1 difféomorphisme entre deux bornés Ω et D de \mathbb{R}^2 . Alors pour toute fonction continue $f : D \rightarrow \mathbb{R}$,

$$\iint_D f = \iint_{\Omega} f \circ \Phi(x) |Jac \Phi(x)| dx.$$

Le changement de variable le plus répandu est le passage en coordonnées polaires. On rappelle au besoin que

$$\Phi_{pol} : (r, \theta) \in]0, +\infty[\times]-\pi, \pi[\mapsto (r \cos \theta, r \sin \theta) \in \mathbb{R}^2 \setminus \{(x, 0) : x \leq 0\}$$

est un C^1 difféomorphisme, de jacobien égal à r .

Exercice 16 Soit D le disque unité : $D = \{(x, y) \in \mathbb{R}^2 : x^2 + y^2 < 1\}$. Calculer $\iint_D x dx dy$.

Notons $f(x, y) = x$, qui constitue bien une fonction continue sur \mathbb{R}^2 . Pour utiliser le passage en coordonnées polaires, on se ramène d'abord à un domaine d'intégration contenu dans $\Phi_{pol}(\Omega)$ où $\Omega =]0, +\infty[\times]-\pi, \pi[$. Pour cela, on écrit

$$\iint_D f = \iint_{D \setminus \Delta} f$$

en notant $\Delta = \{(x, 0) : x \leq 0\}$. Alors en utilisant le théorème de changement de variables,

$$\begin{aligned} \iint_{D \setminus \Delta} f &= \int \int_{\Phi_{pol}^{-1}(D \setminus \Delta)} f \circ \Phi_{pol}(r, \theta) |Jac \Phi(r, \theta)| dr d\theta \\ &= \int_{r=0}^1 \int_{\theta=-\pi}^{\pi} f(r \cos \theta, r \sin \theta) r dr d\theta \\ &= \int_{r=0}^1 \int_{\theta=-\pi}^{\pi} r \cos \theta dr d\theta \\ &= \int_{r=0}^1 r dr \int_{\theta=-\pi}^{\pi} \cos \theta d\theta = 0. \end{aligned}$$

On aurait pu s'attendre à ce résultat en observant l'imparité de la fonction f par rapport à l'axe des ordonnées. Autrement dit, pour tout $(x, y) \in D$, $f(-x, y) = -f(x, y)$. En appliquant le théorème de changement de variables avec le C^1 difféomorphisme $\Phi : (x, y) \mapsto (-x, y)$ qui est de jacobien -1 , on obtient

$$\iint_D f = \iint_{\Phi^{-1}(D)} f \circ \Phi(x, y) |Jac \Phi(x, y)| dx dy = - \iint_D f(x, y) dx dy = - \iint_D f$$

et finalement

$$\iint_D f = 0.$$

Pour calculer certaines intégrales pour des fonctions d'une seule variable, il est parfois commode de se ramener à l'intégrale de fonctions de plusieurs variables.

Exercice 17 Calculer $\int_{\mathbb{R}} e^{-x^2} dx$.

En toute rigueur, on devrait commencer à calculer l'intégrale sur un segment de la forme $[0, A]$ puis passer à la limite quand $A \rightarrow +\infty$. Le calcul qui suit n'est donc pas parfaitement justifié (en particulier, on calcule des intégrales multiples sur des domaines non bornés).

L'idée est de calculer d'abord le carré de l'intégrale dont on cherche la valeur :

$$\begin{aligned} \left(\int_{\mathbb{R}} e^{-x^2} dx \right)^2 &= \left(\int_{\mathbb{R}} e^{-x^2} dx \right) \left(\int_{\mathbb{R}} e^{-y^2} dy \right) \\ &= \iint_{\mathbb{R}^2} e^{-x^2} e^{-y^2} dx dy = \iint_{\mathbb{R}^2} e^{-x^2-y^2} dx dy. \end{aligned}$$

Par passage en coordonnées polaires (après avoir enlevé la demi-droite $\{(x, 0) : x \leq 0\}$), il vient

$$\begin{aligned} \left(\int_{\mathbb{R}} e^{-x^2} dx \right)^2 &= \int_{r=0}^{+\infty} \int_{\theta=-\pi}^{\pi} e^{-r^2} r dr d\theta \\ &= 2\pi \left[\frac{-1}{2} e^{-r^2} \right]_{r=0}^{+\infty} = \pi. \end{aligned}$$

On en déduit

$$\int_{\mathbb{R}} e^{-x^2} dx = \sqrt{\pi}.$$

5.3 Cas particuliers de la formule de Stokes

Dans cette section, on présente deux cas particuliers de la formule de Stokes, souvent utilisée en électromagnétisme. Il ne s'agit rien d'autre que d'une intégration par parties pour des fonctions de plusieurs variables.

On commence par le cas où le domaine de départ est un rectangle

$$D = [a, b] \times [c, d] \quad , \quad -\infty < a < b < +\infty, -\infty < c < d < +\infty.$$

On considère aussi une fonction $X : \mathbb{R}^2 \rightarrow \mathbb{R}^2$ de classe C^1 . On rappelle que la divergence de X en un point $(x, y) \in \mathbb{R}^2$ est définie comme la trace de la jacobienne de X (la matrice de $DX(x, y)$ dans la base canonique), autrement dit :

$$\operatorname{div} X = \frac{\partial X_1}{\partial x}(x, y) + \frac{\partial X_2}{\partial y}(x, y).$$

On a donc

$$\begin{aligned} \iint_D \operatorname{div} X(x, y) dx dy &= \int_a^b \int_c^d \frac{\partial X_1}{\partial x}(x, y) + \frac{\partial X_2}{\partial y}(x, y) dx dy \\ &= \int_c^d dy \int_a^b \frac{\partial X_1}{\partial x}(x, y) dx + \int_a^b dx \int_c^d \frac{\partial X_2}{\partial y}(x, y) dy \\ &= \int_c^d (X_1(b, y) - X_1(a, y)) dy + \int_a^b (X_2(x, d) - X_2(x, c)) dx \end{aligned}$$

On introduit la normale sortante ν au domaine D , c'est-à-dire la fonction définie sur le bord de D (privé des sommets du rectangle) et à valeurs dans \mathbb{R}^2

$$\nu = (1, 0) \text{ sur } \{b\} \times [c, d] \quad , \quad \nu = (0, 1) \text{ sur } [a, b] \times \{d\}$$

$$\nu = (-1, 0) \text{ sur } \{a\} \times [c, d] \quad , \quad \nu = (0, -1) \text{ sur } [a, b] \times \{c\}.$$

Alors on peut reformuler l'égalité précédente

$$\begin{aligned} \iint_D \operatorname{div} X(x, y) dx dy &= \int_{\{b\} \times [c, d]} \langle X(b, y), \nu(b, y) \rangle dy + \int_{[a, b] \times \{d\}} \langle X(x, d), \nu(x, d) \rangle dx \\ &\quad + \int_{\{a\} \times [c, d]} \langle X(a, y), \nu(a, y) \rangle dy + \int_{[a, b] \times \{c\}} \langle X(x, c), \nu(x, c) \rangle dx. \end{aligned}$$

On réécrit souvent cette dernière quantité sous la forme :

$$\iint_D \operatorname{div} X(x, y) dx dy = \int_{\partial D} \langle X, \nu \rangle d\sigma.$$

On considère maintenant un second domaine. Soit $(a_1, a_2) \in \mathbb{R}^2$ et $R > 0$. On note D le disque (euclidien) de centre (a_1, a_2) et de rayon $R > 0$:

$$D = \{(x, y) \in \mathbb{R}^2 : \|(x, y) - (a_1, a_2)\| < R\}.$$

Soit $X = (X_1, X_2) : \mathbb{R}^2 \rightarrow \mathbb{R}^2$ une fonction de classe C^1 .

Alors en utilisant le théorème de changement de variables avec le C^1 difféomorphisme :

$$\Phi : (r, \theta) \in]0, R[\times]-\pi, \pi[\mapsto (a_1 + r \cos \theta, a_2 + r \sin \theta) \in D \setminus \Delta$$

où $\Delta = \{(x, a_2) : a_1 - R < x \leq a_1\}$, il vient

$$\begin{aligned} \iint_D \operatorname{div} X(x, y) \, dx \, dy &= \int_{D \setminus \Delta} \operatorname{div} X(x, y) \, dx \, dy \\ &= \int_{r=0}^R \int_{\theta=-\pi}^{\pi} \operatorname{div} X(a_1 + r \cos \theta, a_2 + r \sin \theta) r \, dr \, d\theta. \end{aligned}$$

Posons

$$Y(r, \theta) = (Y_1(r, \theta), Y_2(r, \theta)) = (X_1(a_1 + r \cos \theta, a_2 + r \sin \theta), X_2(a_2 + r \cos \theta, a_2 + r \sin \theta)).$$

Alors

$$\begin{aligned} \frac{\partial Y}{\partial r}(r, \theta) &= \frac{\partial X}{\partial x}(a_1 + r \cos \theta, a_2 + r \sin \theta) \cos \theta + \frac{\partial X}{\partial y}(a_1 + r \cos \theta, a_2 + r \sin \theta) \sin \theta \\ \frac{\partial Y}{\partial \theta}(r, \theta) &= \frac{\partial X}{\partial x}(a_1 + r \cos \theta, a_2 + r \sin \theta)(-r \sin \theta) + \frac{\partial X}{\partial y}(a_1 + r \cos \theta, a_2 + r \sin \theta)(r \cos \theta). \end{aligned}$$

On en déduit

$$\begin{aligned} \frac{\partial X_1}{\partial x}(a_1 + r \cos \theta, a_2 + r \sin \theta) &= \cos \theta \frac{\partial Y_1}{\partial r}(r, \theta) - \frac{\sin \theta}{r} \frac{\partial Y_1}{\partial \theta}(r, \theta), \\ \frac{\partial X_2}{\partial y}(a_1 + r \cos \theta, a_2 + r \sin \theta) &= \sin \theta \frac{\partial Y_2}{\partial r}(r, \theta) + \frac{\cos \theta}{r} \frac{\partial Y_2}{\partial \theta}(r, \theta). \end{aligned}$$

Alors

$$\begin{aligned} (\operatorname{div} X)(a_1 + r \cos \theta, a_2 + r \sin \theta) &= \frac{\partial X_1}{\partial x}(a_1 + r \cos \theta, a_2 + r \sin \theta) + \frac{\partial X_2}{\partial y}(a_1 + r \cos \theta, a_2 + r \sin \theta) \\ &= \cos \theta \frac{\partial Y_1}{\partial r}(r, \theta) - \frac{\sin \theta}{r} \frac{\partial Y_1}{\partial \theta}(r, \theta) + \sin \theta \frac{\partial Y_2}{\partial r}(r, \theta) + \frac{\cos \theta}{r} \frac{\partial Y_2}{\partial \theta}(r, \theta) \\ &= \left\langle \frac{\partial Y}{\partial r}(r, \theta), u_r(\theta) \right\rangle + \frac{1}{r} \left\langle \frac{\partial Y}{\partial \theta}(r, \theta), u_\theta(\theta) \right\rangle \end{aligned}$$

en notant $u_r(\theta) = (\cos \theta, \sin \theta)$ le vecteur normal et $u_\theta(\theta) = (-\sin \theta, \cos \theta)$ le vecteur tangent. On obtient donc

$$\iint_D \operatorname{div} X(x, y) \, dx \, dy = \int_{r=0}^R \int_{\theta=-\pi}^{\pi} \left(\left\langle \frac{\partial Y}{\partial r}(r, \theta), u_r(\theta) \right\rangle + \frac{1}{r} \left\langle \frac{\partial Y}{\partial \theta}(r, \theta), u_\theta(\theta) \right\rangle \right) r \, dr \, d\theta. \quad (5.3)$$

On fait une intégration par parties pour chaque terme de la somme précédente : d'une part,

$$\begin{aligned} \int_0^R \left\langle \frac{\partial Y}{\partial r}(r, \theta), u_r(\theta) \right\rangle r \, dr &= \int_0^R \frac{\partial}{\partial r} \langle Y(r, \theta), u_r(\theta) \rangle r \, dr \\ &= \left[\langle Y(r, \theta), u_r(\theta) \rangle r \right]_0^R - \int_0^R \langle Y(r, \theta), u_r(\theta) \rangle \, dr \\ &= R \langle Y(R, \theta), u_r(\theta) \rangle - \int_0^R \langle Y(r, \theta), u_r(\theta) \rangle \, dr. \end{aligned}$$

D'autre part, en remarquant que

$$\frac{\partial}{\partial \theta} \langle Y, u_\theta(\theta) \rangle = \left\langle \frac{\partial Y}{\partial \theta}, u_\theta(\theta) \right\rangle + \left\langle Y, \frac{\partial u_\theta}{\partial \theta}(\theta) \right\rangle,$$

on obtient

$$\int_{-\pi}^{\pi} \left\langle \frac{\partial Y}{\partial \theta}(r, \theta), u_\theta(\theta) \right\rangle d\theta = \left[\langle Y, u_\theta \rangle \right]_{-\pi}^{\pi} - \int_{-\pi}^{\pi} \langle Y(r, \theta), \frac{\partial u_\theta}{\partial \theta}(\theta) \rangle d\theta.$$

On calcule $\frac{\partial u_\theta}{\partial \theta} = -u_r$. En utilisant aussi la 2π périodicité de la fonction $\theta \mapsto \langle Y, u_\theta \rangle$, il vient

$$\int_{-\pi}^{\pi} \left\langle \frac{\partial Y}{\partial \theta}(r, \theta), u_\theta(\theta) \right\rangle d\theta = \int_{-\pi}^{\pi} \langle Y(r, \theta), u_r(\theta) \rangle d\theta.$$

En revenant à (5.3), on a donc établi

$$\begin{aligned} \iint_D \operatorname{div} X(x, y) \, dx \, dy &= \int_{-\pi}^{\pi} R \langle Y(R, \theta), u_r(\theta) \rangle d\theta - \int_{-\pi}^{\pi} \int_0^R \langle Y(r, \theta), u_r(\theta) \rangle \, dr \, d\theta \\ &\quad + \int_0^R \int_{-\pi}^{\pi} \langle Y(r, \theta), u_r(\theta) \rangle \, d\theta \, dr = \int_{-\pi}^{\pi} R \langle Y(R, \theta), u_r(\theta) \rangle d\theta. \end{aligned}$$

On note cette dernière intégrale

$$\int_{\partial D} \langle X, \nu \rangle d\sigma$$

où ν est la normale sortante à D et coïncide ici avec u_r .

Plus généralement, la formule de Stokes s'écrit sur un domaine borné (régulier) $D \subset \mathbb{R}^2$:

Théorème 27 Soit $X : \mathbb{R}^2 \rightarrow \mathbb{R}^2$ une fonction C^1 . Alors

$$\iint_D \operatorname{div} X(x, y) \, dx \, dy = \int_{\partial D} \langle X, \nu \rangle d\sigma$$

où ν est la normale sortante extérieure au domaine D .

On n'a défini jusqu'à présent la quantité $\int_{\partial D} \dots$ uniquement dans le cas où D est un carré ou un disque. Plus généralement, si on peut paramétrer ∂D avec une fonction C^1 (ou continue et C^1 par morceaux)

$$\gamma : [a, b] \rightarrow \mathbb{R}^2$$

telle que γ est injective, $\gamma([a, b]) = \partial D$ et γ' ne s'annule pas (i.e. n'est jamais égale au vecteur nul), alors pour toute fonction continue $g : \partial D \rightarrow \mathbb{R}$,

$$\int_{\partial D} g \, d\sigma = \int_a^b g(\gamma(t)) |\gamma'(t)| \, dt.$$

On peut vérifier facilement que cette quantité ne dépend pas de la paramétrisation γ de ∂D .

Chapitre 6

Probabilités discrètes

6.1 Événements aléatoires

6.1.1 Issues et événements

L'ensemble des issues possibles d'une expérience aléatoire est appelé l'univers, et souvent noté Ω . Les issues, c'est-à-dire les éléments de Ω , seront souvent désignées par la lettre ω .

Exemple 41 *On peut modéliser*

1. un lancer de dé en prenant $\Omega = \{1, 2, \dots, 6\}$,
2. deux lancers successifs d'un même dé par $\Omega = \{1, 2, \dots, 6\}^2$. Une issue est donc un couple (i, j) où i désigne le résultat du premier lancer, et j le résultat du deuxième lancer,
3. un lancer de deux dés, l'un rouge, l'autre vert, avec $\Omega = \{1, 2, \dots, 6\}^2$. Une issue est un couple de nombres, le premier indiquant le résultat du dé rouge, le second indiquant le résultat du dé vert,
4. un lancer de deux dés identiques par $\Omega := \{\{i, j\} : 1 \leq i, j \leq 6\}$. Les issues ne sont plus des couples, mais des ensembles qui sont soit des singletons : $\{1\}, \dots, \{6\}$ (c'est le cas si les deux dés indiquent le même résultat), soit des ensembles à deux éléments : $\{i, j\}$ avec $i \neq j$.
5. On lance un dé une infinité de fois. Une issue sera alors la suite des résultats obtenus, c'est-à-dire une suite de nombres compris entre 1 et 6. L'univers est alors $\{1, \dots, 6\}^{\mathbb{N}}$.
6. On lance une pièce de monnaie et on s'arrête dès qu'on obtient un pile. Chaque issue correspondra à un nombre de lancers supérieur ou égal à 1. Dans ce cas l'univers s'identifie à $\mathbb{N}^* \cup \{\infty\}$ (on rajoute $\{\infty\}$ pour l'issue où on n'obtient jamais pile).

On note $\mathcal{P}(\Omega)$ l'ensemble des parties de Ω . Les éléments de $\mathcal{P}(\Omega)$ sont donc des parties de Ω . On considère un sous-ensemble \mathcal{E} de $\mathcal{P}(\Omega)$ qu'on appelle l'ensemble des événements de l'expérience aléatoire. Un événement est donc une partie de Ω , autrement dit un ensemble d'issues de l'expérience aléatoire.

- Exemple 42**
1. Lors du lancer d'un dé, on pourra considérer l'événement le nombre obtenu est 1, ou bien le nombre obtenu est pair.
 2. Lors du lancer répété d'une pièce de monnaie qu'on arrête dès qu'on obtient pile, on pourra considérer l'événement on obtient pile avant le cinquième lancer.

On choisira toujours l'ensemble des événements \mathcal{E} de sorte que si A et B sont deux événements, alors l'ensemble

- $A \cup B = \{\omega \in \Omega : \omega \in A \text{ ou } \omega \in B\}$ est un événement, appelé événement A ou B ,

- $A \cap B = \{\omega \in \Omega : \omega \in A \text{ et } \omega \in B\}$ est un événement, appelé *événement A et B*,
- $\Omega \setminus A = \{\omega \in \Omega : \omega \notin A\}$, noté encore A^c , est un événement, appelé *événement contraire de A*.

On supposera aussi que l'ensemble vide, noté \emptyset , qui est la partie de Ω qui ne contient aucun élément, est un événement. On l'appelle l'*événement impossible*.

On peut considérer des suites d'événements : pour chaque entier $n \in \mathbb{N}$, on se donne un événement A_n . On note $(A_n)_{n \in \mathbb{N}}$ cette suite d'événements.

Définition 39 La réunion d'une suite d'événements $(A_n)_{n \in \mathbb{N}}$ est définie par

$$\cup_{n \in \mathbb{N}} A_n = \{\omega \in \Omega : \text{il existe } n \in \mathbb{N} \text{ tel que } \omega \in A_n\}.$$

On dit qu'une suite d'événements est complète lorsque

1. les événements sont deux à deux disjoints :

$$\forall n, m \in \mathbb{N}, \quad A_n \cap A_m = \emptyset,$$

2. de réunion Ω :

$$\cup_{n \in \mathbb{N}} A_n = \Omega.$$

On supposera toujours que la réunion d'une suite d'événements est encore un événement.

Enfin, on dit que deux événements A et B sont *incompatibles* si $A \cap B = \emptyset$.

Exemple 43 1. Dans le cas du dé, l'événement le nombre obtenu est divisible par 7 est impossible. Les événements le nombre obtenu est pair et le nombre obtenu est impair sont incompatibles.

2. Dans le cas du lancer de pièce, pour chaque $n \in \mathbb{N}^*$, on peut définir l'événement A_n comme on obtient pile avant le lancer n . Alors les événements A_n ne sont pas deux à deux disjoints. Si on considère l'événement B_n défini par on obtient pile au lancer n , alors les B_n sont deux à deux disjoints.

3. Toujours dans le cas du lancer de pièce, on note C_n l'événement on obtient pile après le lancer $2^{n-1} - 1$ et avant le lancer 2^n , pour tout $n \in \mathbb{N}^*$, et C_0 l'événement on n'obtient jamais pile, alors la suite d'événements $(C_n)_{n \in \mathbb{N}}$ est complète.

6.1.2 Probabilités

Une mesure de probabilité vise à donner une idée quantitative de l'importance de chaque événement. Elle traduit par un nombre la chance que l'événement a de se réaliser.

Définition 40 Une mesure de probabilité \mathbb{P} sur Ω est une application de l'ensemble des événements \mathcal{E} à valeurs dans $[0, 1]$ telle que

1. $\mathbb{P}(\Omega) = 1, \mathbb{P}(\emptyset) = 0$,
2. si A et B sont deux événements tels que $A \subset B$, alors $\mathbb{P}(A) \leq \mathbb{P}(B)$,
3. si A et B sont deux événements tels que $A \cap B = \emptyset$, alors $\mathbb{P}(A \cup B) = \mathbb{P}(A) + \mathbb{P}(B)$,
4. si $(A_n)_{n \in \mathbb{N}}$ est une suite d'événements deux à deux disjoints, alors

$$\mathbb{P}(\cup_{n \in \mathbb{N}} A_n) = \sum_{n=0}^{+\infty} \mathbb{P}(A_n).$$

Le dernier axiome signifie que la série à termes positifs $\sum_n \mathbb{P}(A_n)$ converge vers le nombre $\mathbb{P}(\cup_{n \in \mathbb{N}} A_n)$.

Remarque 14 En fait, on peut montrer que les axiomes 1 et 4 dans la définition 40 impliquent les axiomes 2 et 3.

Proposition 42 Si \mathbb{P} est une mesure de probabilité, et $A, B \subset \Omega$,

1. $\mathbb{P}(A^c) = 1 - \mathbb{P}(A)$,
2. $\mathbb{P}(B) = \mathbb{P}(A) - \mathbb{P}(A \setminus B)$ si $B \subset A$,
3. $\mathbb{P}(A \cup B) = \mathbb{P}(A) + \mathbb{P}(B) - \mathbb{P}(A \cap B)$.

Preuve : En écrivant $\Omega = A \cup A^c$, on obtient

$$1 = \mathbb{P}(\Omega) = \mathbb{P}(A \cup A^c) = \mathbb{P}(A) + \mathbb{P}(A^c)$$

d'où la première relation. En écrivant $A = B \cup (A \setminus B)$, on obtient la deuxième égalité. Enfin, en observant que $A \cup B = A \cup (B \setminus (A \cap B))$, la troisième égalité découle de la deuxième. □

Exemple 44 1. Si Ω est un univers fini (comme pour le lancer de dé), on définit la mesure de probabilité uniforme sur Ω par

$$\forall A \subset \Omega, \quad \mathbb{P}(A) = \frac{|A|}{|\Omega|}.$$

Ici, $|A|$ désigne le cardinal de A

2. Si $\Omega = \{0, 1\}$, on définit la mesure de Bernoulli de paramètre $p \in [0, 1]$ par

$$\mathbb{P}(\{1\}) = p, \quad \mathbb{P}(\{0\}) = 1 - p.$$

3. Si Ω est $\{0, \dots, N\}$, la mesure binomiale est définie par :

$$\forall k \in \{0, \dots, N\}, \quad \mathbb{P}(\{k\}) = C_N^k p^k (1-p)^{N-k}.$$

Définition 41 Soient A et B deux événements. On suppose que $\mathbb{P}(B) > 0$. La probabilité conditionnelle de A sachant B est

$$P(A|B) = \frac{P(A \cap B)}{P(B)}.$$

Lemme 3 Soit $\{A_n\}_{n \in \mathbb{N}}$ une suite d'événements disjoints deux à deux. Alors pour tout événement $B \subset \Omega$, la série $\sum_n \mathbb{P}(A_n \cap B)$ converge et

$$\sum_{n=0}^{+\infty} \mathbb{P}(A_n \cap B) = \mathbb{P}((\cup_{n \in \mathbb{N}} A_n) \cap B).$$

Preuve : Observer que

$$(\cup_{n \in \mathbb{N}} A_n) \cap B = \cup_{n \in \mathbb{N}} (A_n \cap B)$$

(appartenir à l'ensemble de gauche signifie qu'on appartient à B et à l'un des A_n , tandis qu'appartenir à l'ensemble de droite signifie qu'il existe n pour lequel on appartient à B et à A_n : il s'agit donc du même ensemble !). De plus, les A_n étant deux à deux disjoints, il en est de même des $A_n \cap B$:

$$(A_n \cap B) \cap (A_m \cap B) = A_n \cap A_m \cap B = (A_n \cap A_m) \cap B = \emptyset \cap B = \emptyset.$$

On en déduit que la série $\sum_n \mathbb{P}(A_n \cap B)$ converge et sa limite est $\mathbb{P}(\cup_{n \in \mathbb{N}} (A_n \cap B))$. Finalement,

$$\mathbb{P}((\cup_{n \in \mathbb{N}} A_n) \cap B) = \sum_{n=0}^{+\infty} \mathbb{P}(A_n \cap B).$$

□

Ce lemme a deux conséquences. La première est le fait que les probabilités conditionnelles sont des mesures de probabilité :

Proposition 43 On suppose que $\mathbb{P}(B) > 0$. L'application $P(\cdot|B) : \mathcal{E} \rightarrow \mathbb{R}$ est une mesure de probabilités sur Ω .

Preuve : Par définition de la probabilité conditionnelle,

$$\mathbb{P}(\emptyset|B) = \frac{\mathbb{P}(\emptyset \cap B)}{\mathbb{P}(B)} = \frac{\mathbb{P}(\emptyset)}{\mathbb{P}(B)} = 0, \mathbb{P}(\Omega|B) = \frac{\mathbb{P}(\Omega \cap B)}{\mathbb{P}(B)} = \frac{\mathbb{P}(B)}{\mathbb{P}(B)} = 1.$$

De plus, si $(A_n)_{n \in \mathbb{N}}$ est une suite d'événements deux à deux disjoints,

$$\mathbb{P}(\cup_{n \in \mathbb{N}} A_n | B) = \frac{\mathbb{P}((\cup_{n \in \mathbb{N}} A_n) \cap B)}{\mathbb{P}(B)}.$$

D'après le lemme,

$$\mathbb{P}((\cup_{n \in \mathbb{N}} A_n) \cap B) = \sum_{n=0}^{\infty} \mathbb{P}(A_n \cap B)$$

d'où

$$\mathbb{P}(\cup_{n \in \mathbb{N}} A_n | B) = \frac{\sum_{n=0}^{\infty} \mathbb{P}(A_n \cap B)}{\mathbb{P}(B)} = \sum_{n=0}^{\infty} \frac{\mathbb{P}(A_n \cap B)}{\mathbb{P}(B)} = \sum_{n=0}^{\infty} P(A_n|B).$$

Conclusion : $P(\cdot|B)$ est bien une probabilité. □

La deuxième conséquence du lemme est la formule des probabilités totales.

Proposition 44 (Formule des probabilités totales) Si $\{A_n\}_n$ est une suite complète d'événements, alors pour tout événement $B \subset \Omega$:

$$\mathbb{P}(B) = \sum_n \mathbb{P}(B|A_n)\mathbb{P}(A_n)$$

(avec la convention $\mathbb{P}(B|A_n) = 0$ si $\mathbb{P}(A_n) = 0$).

Preuve : D'après le lemme, $\sum_n \mathbb{P}(B \cap A_n)$ converge vers

$$\mathbb{P}(B \cap (\cup_n A_n)) = \mathbb{P}(B \cap \Omega) = \mathbb{P}(B).$$

On conclut en utilisant que $\mathbb{P}(B|A_n)\mathbb{P}(A_n) = \mathbb{P}(B \cap A_n)$. □

On en déduit aussitôt :

Remarque 15 (Formule de Bayes) Sous l'hypothèse de la proposition précédente et si $\mathbb{P}(B) > 0$,

$$\mathbb{P}(A_1|B) = \frac{\mathbb{P}(A_1 \cap B)}{\mathbb{P}(B)} = \frac{\mathbb{P}(B|A_1)\mathbb{P}(A_1)}{\sum_n \mathbb{P}(B|A_n)\mathbb{P}(A_n)}.$$

Pour calculer des probabilités conditionnelles, on utilise souvent un arbre dont les branches portent les probabilités conditionnelles tandis que les noeuds portent les probabilités d'événements :

Exemple 45 Le dépistage d'une maladie conduit à un taux de détection de 90 pour 100. Le taux de faux positif est de 5 pour 100. Le taux de détection de 90 pour 100 signifie que si 100 individus malades passent ce test, en moyenne 90 seront détectés comme malade et 10 ne le seront pas. Parallèlement, le taux de faux positif de 5 pour 100 signifie que si 100 individus sains passent ce test, en moyenne 5 seront identifiés à tort comme malades. Sachant que la probabilité qu'un individu soit malade est environ de $1/650$, calculer la probabilité qu'un individu soit sain sachant qu'il a été déclaré malade.

Modélisons : l'ensemble des issues est constitué par l'ensemble des individus. Notons A l'événement : l'individu est sain, et B l'événement : le test est positif. On traduit l'énoncé : $\mathbb{P}(B|A^c) = 0.9$, $\mathbb{P}(B|A) = 0.05$, $\mathbb{P}(A^c) = \frac{1}{650}$. Par la formule de Bayes, on en déduit

$$\begin{aligned} \mathbb{P}(A|B) &= \frac{\mathbb{P}(B|A)\mathbb{P}(A)}{\mathbb{P}(B|A)\mathbb{P}(A) + \mathbb{P}(B|A^c)\mathbb{P}(A^c)} \\ &= \frac{0.05 \times (1 - \frac{1}{650})}{0.05 \times (1 - \frac{1}{650}) + 0.9 \times \frac{1}{650}} = \frac{0.05 \times 649}{0.05 \times 649 + 0.9} = 0,97. \end{aligned}$$

6.2 Variables aléatoires discrètes

Dans toute cette section, on se donne un univers Ω et un ensemble d'événements \mathcal{E} sur Ω . On se donne de plus un ensemble $\mathcal{X} \subset \mathbb{R}$ de la forme

$$\mathcal{X} = \{x_i, i \in I\}$$

où I est soit \mathbb{N} soit un ensemble fini. Dans cette écriture, on suppose implicitement que les x_i sont distincts (autrement dit, l'application $i \in I \mapsto x_i \in \mathcal{X}$ est injective).

Définition 42 Une variable aléatoire sur \mathcal{X} est une fonction $X : \Omega \rightarrow \mathcal{X}$ telle que pour tout $x \in \mathcal{X}$, l'ensemble

$$X^{-1}(\{x\}) = \{\omega \in \Omega : X(\omega) = x\}$$

est un événement.

Pour $x \in \mathcal{X}$, on notera souvent $[X = x]$ l'ensemble $X^{-1}(\{x\})$. Si $A \subset \mathcal{X}$, $[X \in A]$ désignera l'événement $X^{-1}(A) = \{\omega \in \Omega : X(\omega) \in A\}$.

Définition 43 La loi d'une variable aléatoire est la fonction $x \in \mathcal{X} \mapsto p(x) := \mathbb{P}([X = x])$.

Exemple 46 1. Si $\mathcal{X} = \{0, 1\}$ et $p \in [0, 1]$, on dit que X suit une loi de Bernoulli lorsque $p(1) = p$ et $p(0) = 1 - p$.

2. Si $\mathcal{X} = \{1, \dots, N\}$, on dit que X suit une loi uniforme si $p(i) = \frac{1}{N}$ pour tout $i \in \{1, \dots, N\}$.

Remarque 16 Si $A \subset \Omega$, on note 1_A la fonction indicatrice de A définie par

$$1_A(\omega) = \begin{cases} 1 & \text{si } \omega \in A, \\ 0 & \text{si } \omega \in A^c. \end{cases}$$

L'indicatrice de l'événement A suit une loi de Bernoulli de paramètre $\mathbb{P}(A)$.

Proposition 45 Pour toute partie $\mathcal{X}' \subset \mathcal{X}$,

$$\mathbb{P}([X \in \mathcal{X}']) = \sum_{x \in \mathcal{X}'} p(x).$$

En particulier,

$$\sum_{x \in \mathcal{X}} p(x) = 1.$$

Preuve : L'ensemble \mathcal{X}' est de la forme $\{x_j : j \in J\}$ où J est une partie de I . Pour tout $j \in J$, notons A_j l'événement $[X = x_j]$. Alors les A_j sont deux à deux disjoints et de réunion $[X \in \mathcal{X}']$. On en déduit

$$\sum_{x \in \mathcal{X}'} p(x) = \sum_{j \in J} \mathbb{P}(A_j) = \mathbb{P}(\cup_{j \in J} A_j) = \mathbb{P}([X \in \mathcal{X}']),$$

ce qui montre la première assertion. Dans le cas où $\mathcal{X}' = \mathcal{X}$, on obtient

$$\sum_{x \in \mathcal{X}} p(x) = \mathbb{P}([X \in \mathcal{X}]) = \mathbb{P}(\Omega) = 1,$$

ce qui prouve la seconde assertion. □

Définition 44 La fonction de répartition associée à une variable aléatoire $X : \Omega \rightarrow \mathcal{X}$ est la fonction $F_X : \mathbb{R} \rightarrow [0, 1]$ définie par

$$F_X(x) = \mathbb{P}([X \leq x]).$$

Une issue $\omega \in [X \leq x]$ si et seulement si $X(\omega) \leq x$. De manière équivalente, il existe $k \in I$ tel que $X(\omega) = x_k$ et $x_k \leq x$. Ainsi

$$[X \leq x] = \cup_{k: x_k \leq x} [X = x_k].$$

Dans le membre de droite de l'égalité ci-dessus, on prend la réunion sur tous les indices k tels que $x_k \leq x$. On en déduit :

$$F_X(x) = \mathbb{P}(\cup_{k: x_k \leq x} [X = x_k]) = \sum_{k: x_k \leq x} p(x_k). \quad (6.1)$$

Proposition 46 *La fonction de répartition F_X est une fonction croissante. De plus,*

$$\lim_{x \rightarrow -\infty} F_X(x) = 0 \quad , \quad \lim_{x \rightarrow +\infty} F_X(x) = 1.$$

Preuve : La monotonie de F_X provient de la monotonie de la mesure \mathbb{P} par rapport à l'inclusion (cela correspond à l'axiome 2 de la définition d'une probabilité). Alternativement, on peut utiliser (6.1). Montrons maintenant la dernière propriété (la limite en $-\infty$ se démontre de manière analogue). Comme F_X est croissante et bornée par 1, sa limite en $+\infty$ existe et est ≤ 1 . Montrons que cette limite vaut 1. Soit $\varepsilon > 0$. Alors il existe $k_0 \in \mathbb{N}$

$$\sum_{k \leq k_0} p(x_k) \geq 1 - \varepsilon.$$

Ici, on a utilisé que la série dans le membre de gauche converge vers 1, par la proposition 45. Alors d'après la formule (6.1), en posant $y = \max\{x_k : k \leq k_0\}$,

$$F(y) = \sum_{k: x_k \leq y} p(x_k) \geq \sum_{k \leq k_0} p(x_k) \geq 1 - \varepsilon.$$

Comme F_X est croissante, pour tout $x \geq y$, $F_X(x) \geq F_X(y) \geq 1 - \varepsilon$. Cela prouve que $\lim_{x \rightarrow +\infty} F_X(x) = 1$ comme attendu.

6.3 Moments des variables aléatoires discrètes réelles

La notion d'espérance est la traduction mathématique de l'idée de moyenne des résultats obtenus en répétant une même expérience dans des conditions identiques et indépendantes.

6.3.1 Espérance

Soit X une variable aléatoire sur Ω à valeurs dans $\mathcal{X} = \{x_i, i \in I\} \subset \mathbb{R}$, où I est soit \mathbb{N} , soit un ensemble fini. Chaque x_i est un nombre réel.

Définition 45 *Si $\sum_{i \in I} |x_i| \mathbb{P}(X = x_i) < \infty$, on dit que X est intégrable et on définit l'espérance de X par*

$$\mathbb{E}(X) = \sum_{i \in I} x_i \mathbb{P}([X = x_i]).$$

Noter que si I est fini, alors X est automatiquement intégrable.

Remarque 17 *Soit A un événement. L'indicatrice de A prend exactement deux valeurs 0 et 1. On en déduit*

$$\mathbb{E}(1_A) = 1 \cdot \mathbb{P}([1_A = 1]) = \mathbb{P}(A).$$

Pour une variable aléatoire X , on peut donc exprimer son espérance par la formule suivante :

$$\mathbb{E}(X) = \sum_{i \in I} x_i \mathbb{P}([X = x_i]) = \sum_{i \in I} x_i \mathbb{E}(1_{[X=x_i]}).$$

6.3.2 Composition et linéarité

Proposition 47 [Théorème de transfert] Soient $X : \Omega \rightarrow \mathcal{X}$ est une variable aléatoire et $f : \mathcal{X} \rightarrow \mathbb{R}$ une fonction. Si $\sum_{i \in I} |f(x_i)| \mathbb{P}([X = x_i])$ converge, alors $f(X)$ est intégrable et :

$$\mathbb{E}(f(X)) = \sum_{i \in I} f(x_i) \mathbb{P}([X = x_i]).$$

Preuve : On prouve ce résultat dans deux cas particuliers. On commence par faire l'hypothèse supplémentaire que f est injective. Dans ce cas, $f(X)$ prend les valeurs distinctes $f(x_i)$, $i \in I$. On en déduit par définition de l'espérance

$$\mathbb{E}(f(X)) = \sum_{i \in I} f(x_i) \mathbb{P}([f(X) = f(x_i)]).$$

Comme f est injective, $[f(X) = f(x_i)] = [X = x_i]$ et donc $\mathbb{P}([f(X) = f(x_i)]) = \mathbb{P}([X = x_i])$. On en déduit

$$\mathbb{E}(f(X)) = \sum_{i \in I} f(x_i) \mathbb{P}([X = x_i]),$$

ce qui conclut la preuve dans ce cas là.

On ne suppose plus que f est injective, mais que \mathcal{X} est fini. On a toujours $f(\mathcal{X}) = \{f(x_i) : i \in I\}$ mais comme f n'est plus supposée injective, il se peut que les $f(x_i)$ ne soit pas distincts. Introduisons alors $K \subset I$ tel que $\{f(x_k) : k \in K\}$ soit l'ensemble des valeurs distinctes prises par f . Alors par définition de l'espérance,

$$\mathbb{E}(f(X)) = \sum_{k \in K} f(x_k) \mathbb{P}([f(X) = f(x_k)]).$$

Pour tout $k \in K$, notons $I_k := \{i \in I : f(x_i) = f(x_k)\}$. Alors

$$\mathbb{P}([f(X) = f(x_k)]) = \mathbb{P}(\cup_{i \in I_k} [X = x_i]) = \sum_{i \in I_k} \mathbb{P}([X = x_i]).$$

En reportant dans l'inégalité précédente, il vient

$$\mathbb{E}(f(X)) = \sum_{k \in K} f(x_k) \sum_{i \in I_k} \mathbb{P}([X = x_i]) = \sum_{k \in K} \sum_{i \in I_k} f(x_k) \mathbb{P}([X = x_i]) = \sum_{k \in K} \sum_{i \in I_k} f(x_i) \mathbb{P}([X = x_i]).$$

Pour obtenir la dernière égalité, on a utilisé que $f(x_i) = f(x_k)$ lorsque $i \in I_k$. Comme $I = \cup_{k \in K} I_k$ et qu'il s'agit d'ensembles finis, on obtient finalement

$$\mathbb{E}(f(X)) = \sum_{i \in I} f(x_i) \mathbb{P}([X = x_i])$$

ce qui achève la preuve. □

En particulier, si X est une variable aléatoire intégrable, alors il en est de même de λX pour tout $\lambda \in \mathbb{R}$ et de plus

$$\mathbb{E}(\lambda X) = \lambda \mathbb{E}(X).$$

Un autre cas particulier implique l'inégalité triangulaire :

Proposition 48 (Inégalité triangulaire) Soit X une variable aléatoire intégrable. Alors $|X|$ est intégrable et

$$|\mathbb{E}(X)| \leq \mathbb{E}(|X|).$$

Preuve : En utilisant l'inégalité triangulaire pour les séries numériques et en appliquant la proposition 47 à $f = |\cdot|$, on obtient :

$$|\mathbb{E}(X)| = \left| \sum_{i \in I} x_i \mathbb{P}([X = x_i]) \right| \leq \sum_{i \in I} |x_i| \mathbb{P}([X = x_i]) = \mathbb{E}(|X|),$$

ce qui achève la preuve. □

Proposition 49 Si X et Y sont deux variables aléatoires intégrables à valeurs dans \mathcal{X} et \mathcal{Y} respectivement, alors on peut définir l'espérance de $X + Y$ et

$$\mathbb{E}(X + Y) = \mathbb{E}(X) + \mathbb{E}(Y).$$

On admettra cette proposition.

Il suit de la définition que si X est une variable aléatoire positive (i.e. ne prenant que des valeurs positives) et intégrable, alors $\mathbb{E}(X) \geq 0$. En utilisant la linéarité de l'espérance, on obtient :

Proposition 50 (Monotonie) Soient X, Y deux variables aléatoires telles que pour tout $\omega \in \Omega$, $X(\omega) \leq Y(\omega)$. Si X et Y sont intégrables, alors

$$\mathbb{E}(X) \leq \mathbb{E}(Y).$$

6.3.3 Variance

Définition 46 (Moments d'ordre m) Soit X une variable aléatoire. Si $\sum_i |x_i|^m \mathbb{P}([X = x_i])$ converge, on définit le moment d'ordre m de X par $\mathbb{E}(X^m)$.

Une variable prenant un nombre fini de valeurs a tous ses moments finis.

Comme conséquence de la proposition 47, on obtient :

$$\mathbb{E}(X^m) = \sum_{k \in I} x_k^m p(x_k).$$

La proposition suivante peut être vue comme un cas particulier de l'inégalité de Cauchy-Schwarz :

Proposition 51 Soit X une variable aléatoire de second moment fini. Alors X est intégrable et

$$\mathbb{E}(|X|) \leq [\mathbb{E}(|X|^2)]^{1/2}.$$

Une variable de second moment fini est donc de premier moment fini. L'écart-type permet de mesurer la distance à la moyenne :

Définition 47 Soit X une variable aléatoire de second moment fini. Alors on appelle variance de X la quantité

$$\mathbb{V}(X) = \mathbb{E}[(X - \mathbb{E}(X))^2] = \mathbb{E}(X^2) - (\mathbb{E}(X))^2.$$

La racine carrée de la variance s'appelle l'écart-type.

Pour justifier l'égalité dans l'énoncé précédent, observons que

$$\mathbb{E}[(X - \mathbb{E}(X))^2] = \mathbb{E}(X^2 + (\mathbb{E}(X))^2 - 2\mathbb{E}(X)X).$$

Lorsqu'on écrit dans la ligne ci-dessus $(\mathbb{E}(X))^2$, il s'agit ici de la fonction constante égale au nombre $(\mathbb{E}(X))^2$. Alors par linéarité,

$$\begin{aligned} \mathbb{E}[(X - \mathbb{E}(X))^2] &= \mathbb{E}(X^2) + \mathbb{E}((\mathbb{E}(X))^2) - 2\mathbb{E}(\mathbb{E}(X)X) \\ &= \mathbb{E}(X^2) + (\mathbb{E}(X))^2 \mathbb{E}(1) - 2\mathbb{E}(X)\mathbb{E}(X) = \mathbb{E}(X^2) - (\mathbb{E}(X))^2. \end{aligned}$$

Dans la ligne ci-dessus, on a utilisé le fait que l'espérance de la fonction constante égale à 1 est, par définition de l'espérance, $\mathbb{E}(1) = 1 \times \mathbb{P}(\Omega) = 1$.

Remarque 18 La variance n'est pas linéaire : si $\lambda \in \mathbb{R}$,

$$\mathbb{V}(\lambda X + \mu) = \lambda^2 \mathbb{V}(X).$$

Preuve :

$$\begin{aligned}\mathbb{V}(\lambda X + \mu) &= \mathbb{E}((\lambda X + \mu)^2) - (\mathbb{E}(\lambda X + \mu))^2 = \mathbb{E}(\lambda^2 X^2 + \mu^2 + 2\lambda\mu X) - (\lambda\mathbb{E}(X) + \mu)^2 \\ &= \lambda^2\mathbb{E}(X^2) + \mu^2 + 2\lambda\mu\mathbb{E}(X) - \lambda^2(\mathbb{E}(X))^2 - \mu^2 - 2\lambda\mu\mathbb{E}(X) \\ &= \lambda^2\mathbb{V}(X).\end{aligned}$$

□

Remarque 19 Si $\mathbb{V}(X) = 0$, la variable X est constante en dehors d'un événement de probabilité nulle.

Preuve : Par hypothèse et en utilisant la proposition 47, $0 = \mathbb{V}(X) = \sum_{i \in I} (x_i - \mathbb{E}(X))^2 p(x_i)$. Comme il s'agit d'une série de nombres positifs, cela implique que pour tout $i \in I$, $x_i = \mathbb{E}(X)$ ou $p(x_i) = 0$. On en déduit

$$\mathbb{P}(\cup_{i \in I: x_i \neq \mathbb{E}(X)} [X = x_i]) = \sum_{i \in I: x_i \neq \mathbb{E}(X)} p(x_i) = 0.$$

De plus, $X = \mathbb{E}(X)$ sur $\Omega \setminus \cup_{i \in I: x_i \neq \mathbb{E}(X)} [X = x_i]$, ce qui conclut la preuve.

□

Exemple 47 Si X est une variable de Bernoulli de paramètre p , alors

$$\mathbb{E}(X) = 1 \times p = p, \quad \mathbb{V}(X) = \mathbb{E}(X^2) - (\mathbb{E}(X))^2 = 1^2 \times p - p^2 = p(1 - p).$$

Exemple 48 Si X suit une loi binomiale de paramètre p et $n > 1$, alors

$$\mathbb{E}(X) = np.$$

En effet,

$$\mathbb{E}(X) = \sum_{k=0}^n k C_n^k p^k (1-p)^{n-k} = \sum_{k=1}^n k C_n^k p^k (1-p)^{n-k}.$$

On remarque ensuite

$$k C_n^k = k \frac{n!}{k!(n-k)!} = n \frac{(n-1)!}{(k-1)!((n-1)-(k-1))!} = n C_{n-1}^{k-1}.$$

Il vient alors

$$\begin{aligned}\mathbb{E}(X) &= n \sum_{k=1}^n C_{n-1}^{k-1} p^k (1-p)^{n-k} = np \sum_{k=1}^n C_{n-1}^{k-1} p^{k-1} (1-p)^{(n-1)-(k-1)} \\ &= np \sum_{k=0}^{n-1} C_{n-1}^k p^k (1-p)^{(n-1)-k} = np(p + (1-p))^{n-1} = np.\end{aligned}$$

Exercice 18 Calculer la variance d'une variable aléatoire qui suit une loi binomiale de paramètres n et p .

Les deux inégalités suivantes sont très souvent utilisées dans la théorie des probabilités.

Proposition 52 1. **Inégalité de Markov** Si X est une variable aléatoire intégrable à valeurs positives et $a > 0$, alors

$$\mathbb{P}([X \geq a]) \leq \frac{\mathbb{E}(X)}{a}.$$

2. Inégalité de Bienaymé-Tchebychev Si X est une variable aléatoire de second moment fini et $a > 0$, alors

$$\mathbb{P}(|X - \mathbb{E}(X)| \geq a) \leq \frac{\mathbb{V}(X)}{a^2}.$$

Preuve : Pour tout $\omega \in \Omega$,

$$1 \geq 1_{[X \geq a]}(\omega),$$

ce qui implique

$$X(\omega) \geq X(\omega)1_{[X \geq a]}(\omega) \geq a1_{[X \geq a]}(\omega). \quad (6.2)$$

La dernière inégalité résulte du fait que si $X(\omega) < a$, alors $1_{[X \geq a]}(\omega) = 0$. Comme (6.2) est vraie pour tout ω , on a donc $X \geq a1_{[X \geq a]}$. Alors par monotonie de l'espérance,

$$\mathbb{E}(X) \geq \mathbb{E}(a1_{[X \geq a]}) = a\mathbb{E}(1_{[X \geq a]}) = a\mathbb{P}([X \geq a]),$$

ce qui prouve l'inégalité de Markov. En appliquant l'inégalité de Markov à $(X - \mathbb{E}(X))^2$ et a^2 , il vient

$$\mathbb{P}((X - \mathbb{E}(X))^2 \geq a^2) \leq \frac{\mathbb{E}((X - \mathbb{E}(X))^2)}{a^2}.$$

Pour conclure, il suffit d'observer que $[(X - \mathbb{E}(X))^2 \geq a^2] = [|X - \mathbb{E}(X)| \geq a]$ et donc $\mathbb{P}((X - \mathbb{E}(X))^2 \geq a^2) = \mathbb{P}(|X - \mathbb{E}(X)| \geq a)$.

Définition 48 Si X et Y sont deux variables aléatoires de second moment fini, la covariance de X et Y est

$$\text{Cov}(X, Y) = \mathbb{E}((X - \mathbb{E}(X))(Y - \mathbb{E}(Y))) = \mathbb{E}(XY) - \mathbb{E}(X)\mathbb{E}(Y).$$

Lorsque $\mathbb{V}(X) > 0$ et $\mathbb{V}(Y) > 0$, le coefficient de corrélation est donné par

$$\text{Cor}(X, Y) = \frac{\text{Cov}(X, Y)}{\sqrt{\mathbb{V}(X)\mathbb{V}(Y)}}.$$

Exercice 19 Montrer la propriété d'invariance d'échelle :

$$\forall a > 0, c > 0, b \in \mathbb{R}, d \in \mathbb{R}, \quad \text{Cor}(X, Y) = \text{Cor}(aX + b, cY + d).$$

6.4 Indépendance

6.4.1 Indépendance d'événements

Définition 49 (Indépendance de deux événements) Deux événements $A, B \subset \Omega$ sont dit indépendants si $\mathbb{P}(A \cap B) = \mathbb{P}(A)\mathbb{P}(B)$.

Remarque 20 Si $\mathbb{P}(B) > 0$, alors A et B sont indépendants ssi $\mathbb{P}(A|B) = \mathbb{P}(A)$.

Exemple 49 Lorsqu'on lance deux dés simultanément, les deux résultats obtenus sont indépendants. Mathématiquement, l'univers est $\Omega = \{1, \dots, 6\}^2$ muni de la probabilité uniforme et pour tout $A, B \subset \{1, \dots, 6\}$, les événements $A \times \{1, \dots, 6\}$ et $\{1, \dots, 6\} \times B$ sont indépendants :

$$\begin{aligned} \mathbb{P}((A \times \{1, \dots, 6\}) \cap (\{1, \dots, 6\} \times B)) &= \mathbb{P}(A \times B) = \frac{1}{36}(\text{card } A \times B) \\ &= \frac{1}{36}(\text{card } A)(\text{card } B) = \mathbb{P}(A \times \{1, \dots, 6\})\mathbb{P}(\{1, \dots, 6\} \times B). \end{aligned}$$

6.4.2 Indépendance de variables aléatoires

Définition 50 On dit que n variables aléatoires X_1, \dots, X_n à valeurs dans $\mathcal{X}_1, \dots, \mathcal{X}_n$ sont indépendantes si pour tout $A_1 \subset \mathcal{X}_1, \dots, A_n \subset \mathcal{X}_n$,

$$\mathbb{P}([X_1 \in A_1] \cap \dots \cap [X_n \in A_n]) = \mathbb{P}([X_1 \in A_1]) \dots \mathbb{P}([X_n \in A_n]).$$

On en déduit facilement :

Remarque 21 Si X et Y sont deux variables aléatoires indépendantes, alors pour toutes fonctions f et g de \mathbb{R} dans \mathbb{R} , les variables $f(X)$ et $g(Y)$ sont indépendantes.

Proposition 53 Les variables aléatoires discrètes X_1, \dots, X_n sont indépendantes si et seulement si pour tout $x_1 \in \mathcal{X}_1, \dots, x_n \in \mathcal{X}_n$,

$$\mathbb{P}([X_1 = x_1, X_2 = x_2, \dots, X_n = x_n]) = \mathbb{P}([X_1 = x_1])\mathbb{P}([X_2 = x_2]) \dots \mathbb{P}([X_n = x_n]).$$

Preuve (dans le cas où chaque \mathcal{X}_j est fini) : Si les X_1, \dots, X_n sont indépendantes, alors pour tout $x_1 \in \mathcal{X}_1, \dots, x_n \in \mathcal{X}_n$, notons A_j l'événement $[X_j = x_j], 1 \leq j \leq n$. Alors

$$\begin{aligned} \mathbb{P}([X_1 = x_1, X_2 = x_2, \dots, X_n = x_n]) &= \mathbb{P}([X_1 \in A_1] \cap \dots \cap [X_n \in A_n]) \\ &= \mathbb{P}([X_1 \in A_1]) \dots \mathbb{P}([X_n \in A_n]) \\ &= \mathbb{P}([X_1 = x_1]) \dots \mathbb{P}([X_n = x_n]). \end{aligned}$$

ce qui montre l'implication directe.

Pour l'implication réciproque, soient $A_1 \subset \mathcal{X}_1, \dots, A_n \subset \mathcal{X}_n$. Pour chaque $1 \leq j \leq n$, on note x_{ij} les éléments de A_j , où i parcourt l'ensemble fini d'indices I_j . On a donc

$$A_j = \{x_{ij} : i \in I_j\}.$$

Alors

$$\begin{aligned} [X_1 \in A_1] \cap \dots \cap [X_n \in A_n] &= (\cup_{i_1 \in I_1} [X_1 = x_{i_1 1}]) \cap \dots \cap (\cup_{i_n \in I_n} [X_n = x_{i_n n}]) \\ &= \cup_{(i_1, \dots, i_n) \in I_1 \times \dots \times I_n} [X_1 = x_{i_1 1}] \cap \dots \cap [X_n = x_{i_n n}] \\ &= \cup_{(i_1, \dots, i_n) \in I_1 \times \dots \times I_n} [X_1 = x_{i_1 1}, \dots, X_n = x_{i_n n}]. \end{aligned}$$

On en déduit

$$\mathbb{P}([X_1 \in A_1] \cap \dots \cap [X_n \in A_n]) = \sum_{(i_1, \dots, i_n) \in I_1 \times \dots \times I_n} \mathbb{P}([X_1 = x_{i_1 1}, \dots, X_n = x_{i_n n}]).$$

En utilisant l'hypothèse de l'énoncé, on obtient

$$\begin{aligned} \mathbb{P}([X_1 \in A_1] \cap \dots \cap [X_n \in A_n]) &= \sum_{(i_1, \dots, i_n) \in I_1 \times \dots \times I_n} \mathbb{P}([X_1 = x_{i_1 1}]) \dots \mathbb{P}([X_n = x_{i_n n}]) \\ &= \sum_{i_1 \in I_1} \mathbb{P}([X_1 = x_{i_1 1}]) \dots \sum_{i_n \in I_n} \mathbb{P}([X_n = x_{i_n n}]) \\ &= \mathbb{P}([X_1 \in A_1]) \dots \mathbb{P}([X_n \in A_n]), \end{aligned}$$

ce qui montre que les variables aléatoires X_1, \dots, X_n sont indépendantes et achève la preuve. \square

Exemple 50 Deux variables de Bernoulli B_1 et B_2 sont indépendantes si et seulement si les événements $[B_1 = 1]$ et $[B_2 = 1]$ sont indépendants.

En effet, d'après la proposition précédente, il suffit de montrer que pour tout $\varepsilon_1, \varepsilon_2 \in \{0, 1\}$, on a

$$\mathbb{P}([B_1 = \varepsilon_1] \cap [B_2 = \varepsilon_2]) = \mathbb{P}([B_1 = \varepsilon_1])\mathbb{P}([B_2 = \varepsilon_2]).$$

Considérons le cas $\varepsilon_1 = \varepsilon_2 = 1$. Alors par hypothèse,

$$\mathbb{P}([B_1 = \varepsilon_1] \cap [B_2 = \varepsilon_2]) = \mathbb{P}([B_1 = 1] \cap [B_2 = 1]) = \mathbb{P}([B_1 = 1])\mathbb{P}([B_2 = 1]).$$

Considérons à présent le cas $\varepsilon_1 = 0, \varepsilon_2 = 1$. Alors on écrit

$$\begin{aligned} [B_1 = \varepsilon_1] \cap [B_2 = \varepsilon_2] &= [B_1 = 0] \cap [B_2 = 1] = (\Omega \setminus [B_1 = 1]) \cap [B_2 = 1] \\ &= (\Omega \cap [B_2 = 1]) \setminus ([B_1 = 1] \cap [B_2 = 1]) = [B_2 = 1] \setminus ([B_1 = 1] \cap [B_2 = 1]). \end{aligned}$$

Alors

$$\begin{aligned}
\mathbb{P}([B_1 = 0] \cap [B_2 = 1]) &= \mathbb{P}([B_2 = 1]) - \mathbb{P}([B_1 = 1] \cap [B_2 = 1]) \\
&= \mathbb{P}([B_2 = 1]) - \mathbb{P}([B_1 = 1])\mathbb{P}([B_2 = 1]) = (1 - \mathbb{P}([B_1 = 1]))\mathbb{P}([B_2 = 1]) \\
&= (\mathbb{P}(\Omega) - \mathbb{P}([B_1 = 1]))\mathbb{P}([B_2 = 1]) = \mathbb{P}(\Omega \setminus [B_1 = 1])\mathbb{P}([B_2 = 1]) \\
&= \mathbb{P}([B_1 = 0])\mathbb{P}([B_2 = 1]).
\end{aligned}$$

Les autres cas se montrent de manière tout à fait similaire.

Exercice 20 *Le lancer de deux dés est modélisé par l'univers $\Omega = \{1, \dots, 6\}^2$ muni de la probabilité uniforme, chaque lancer correspondant à une variable aléatoire : $X_i : (\omega_1, \omega_2) \in \Omega \mapsto \omega_i, i = 1, 2$. Alors X_1 et X_2 sont indépendantes.*

On admet la proposition suivante :

Proposition 54 *Si $X_1, \dots, X_k, Y_1, \dots, Y_\ell$ sont des variables indépendantes et $f : \mathbb{R}^k \rightarrow \mathbb{R}$ est une fonction continue, alors $f(X_1, \dots, X_k), Y_1, \dots, Y_\ell$ sont indépendantes.*

6.4.3 Espérance et indépendance

Soient $A, B \subset \Omega$ deux événements indépendants. En termes de fonctions indicatrices, l'indépendance de A et B s'écrit

$$\mathbb{E}(1_{A \cap B}) = \mathbb{E}(1_A)\mathbb{E}(1_B).$$

Comme $1_{A \cap B} = 1_A 1_B$, on peut aussi écrire

$$\mathbb{E}(1_A 1_B) = \mathbb{E}(1_A)\mathbb{E}(1_B).$$

Autrement, l'espérance du produit est égale au produit des espérances. Ce phénomène est tout à fait général sous l'hypothèse d'indépendance convenable :

Proposition 55 *Soient X et Y deux variables aléatoires à valeurs dans \mathcal{X} et \mathcal{Y} respectivement. Soient $f : \mathcal{X} \rightarrow \mathbb{R}$ et $g : \mathcal{Y} \rightarrow \mathbb{R}$ deux fonctions telles que $f(X)$ et $g(Y)$ sont intégrables. Si X et Y sont indépendantes, alors le produit $f(X)g(Y)$ est intégrable et*

$$\mathbb{E}(f(X)g(Y)) = \mathbb{E}(f(X))\mathbb{E}(g(Y))$$

Preuve (dans le cas où \mathcal{X} et \mathcal{Y} sont finis) : Notons $\mathcal{X} = \{x_i : i \in I\}$ et $\mathcal{Y} = \{y_j : j \in J\}$. Alors la fonction $\omega \in \Omega \mapsto f(X(\omega))g(Y(\omega))$ prend la valeur $f(x_i)g(y_j)$ sur l'ensemble $[X = x_i, Y = y_j]$. On en déduit en suivant les arguments de la preuve de la proposition 47 que

$$\mathbb{E}(f(X)g(Y)) = \sum_{(i,j) \in I \times J} f(x_i)g(y_j)\mathbb{P}([X = x_i, Y = y_j]).$$

Comme X et Y sont indépendantes, il suit que

$$\begin{aligned}
\mathbb{E}(f(X)g(Y)) &= \sum_{(i,j) \in I \times J} f(x_i)g(y_j)\mathbb{P}([X = x_i])\mathbb{P}([Y = y_j]) \\
&= \sum_{i \in I} f(x_i)\mathbb{P}([X = x_i]) \sum_{j \in J} g(y_j)\mathbb{P}([Y = y_j]) = \mathbb{E}(f(X))\mathbb{E}(g(Y)).
\end{aligned}$$

□

Comme cas particulier de la proposition précédente, on a

Corollaire 8 *Si les variables X_1, X_2 sont indépendantes et intégrables, alors*

$$\mathbb{E}(X_1 X_2) = \mathbb{E}(X_1)\mathbb{E}(X_2)$$

Si on suppose de plus que X_1 et X_2 sont de second moment fini, alors

$$\mathbb{V}(X_1 + X_2) = \mathbb{V}(X_1) + \mathbb{V}(X_2).$$

En effet, en utilisant la proposition 8, on a

$$\mathbb{E}((X_1 + X_2)^2) = \mathbb{E}((X_1)^2) + \mathbb{E}((X_2)^2) + 2\mathbb{E}(X_1X_2) = \mathbb{E}((X_1)^2) + \mathbb{E}((X_2)^2) + 2\mathbb{E}(X_1)\mathbb{E}(X_2).$$

Par ailleurs,

$$(\mathbb{E}(X_1 + X_2))^2 = (\mathbb{E}(X_1) + \mathbb{E}(X_2))^2 = (\mathbb{E}(X_1))^2 + (\mathbb{E}(X_2))^2 + 2\mathbb{E}(X_1)\mathbb{E}(X_2).$$

On en déduit

$$\mathbb{V}(X_1 + X_2) = \mathbb{E}((X_1 + X_2)^2) - (\mathbb{E}(X_1 + X_2))^2 = \mathbb{V}(X_1) + \mathbb{V}(X_2),$$

comme attendu. Comme conséquence de la définition de la covariance, on a également :

Proposition 56 *La covariance de deux variables indépendantes de second moment fini est nulle.*

Remarque 22 *Cette condition n'est pas suffisante.*

6.5 Processus de Bernoulli

On considère dans cette section une suite $X_i, i \in \mathbb{N}^*$ de variables indépendantes suivant une loi de Bernoulli de paramètre p . L'indépendance signifie ici que, pour tout $n \geq 1$, les variables $X_i, 1 \leq i \leq n$, sont indépendantes.

Pour toute issue $\omega \in \Omega$, on note

$$T(\omega) = \min\{i \in \mathbb{N}^* : X_i(\omega) = 1\}.$$

Autrement dit, T est la variable aléatoire qui donne l'indice du premier 1. Si l'ensemble du membre de droite est vide, alors on convient de poser $T(\omega) = \infty$. En fait, ce-dernier cas advient si pour tout $i \in \mathbb{N}^*$, $X_i(\omega) = 0$. Cet événement est de probabilité nulle : pour tout $N \in \mathbb{N}^*$,

$$P(\cap_{i \in \mathbb{N}^*} [X_i = 0]) \leq P(\cap_{1 \leq i \leq N} [X_i = 0]) \leq \prod_{i=1}^N P([X_i = 0]),$$

où la dernière inégalité résulte de l'indépendance des X_i . Ainsi,

$$P(\cap_{i \in \mathbb{N}^*} [X_i = 0]) \leq \prod_{i=1}^N (1-p) = (1-p)^N.$$

Comme ceci doit être vrai pour tout $N \in \mathbb{N}^*$ et que $\lim_{N \rightarrow +\infty} (1-p)^N = 0$, on en déduit

$$P(\cap_{i \in \mathbb{N}^*} [X_i = 0]) = 0.$$

Conclusion : hors d'un ensemble de probabilité nulle, T prend ses valeurs dans \mathbb{N}^* .

Fixons $n \in \mathbb{N}^*$. Alors une issue $\omega \in \Omega$ vérifie $T(\omega) = n$ (le premier 1 apparaît en numéro n) si et seulement si $X_i(\omega) = 0$ pour tout $1 \leq i \leq n-1$ (les $n-1$ premiers numéros ont donné 0) et $X_n(\omega) = 1$. Ainsi,

$$[T = n] = [X_n = 1] \cap \cap_{1 \leq i \leq n-1} [X_i = 0].$$

Par indépendance, il suit que

$$P([T = n]) = P([X_n = 1]) \prod_{i=1}^{n-1} P([X_i = 0]) = p(1-p)^{n-1}.$$

De même,

$$[T > n] = \cap_{1 \leq i \leq n} [X_i = 0]$$

et donc

$$P([T > n]) = (1-p)^n.$$

Définition 51 On dit qu'une variable aléatoire $T : \Omega \rightarrow \mathbb{N}^*$ telle que $\mathbb{P}([T = n]) = p(1 - p)^{n-1}$ suit une loi géométrique de paramètre p .

On note $\mathcal{G}(p)$ la loi géométrique de paramètre p .

Proposition 57 Si $T : \Omega \rightarrow \mathbb{N}^*$ suit une loi géométrique de paramètre p , alors pour tout $k, j \in \mathbb{N}^*$,

$$\mathbb{P}([T > j + k] | [T > j]) = \mathbb{P}([T > k]).$$

Preuve : On revient à la définition de la probabilité conditionnelle :

$$\begin{aligned} \mathbb{P}([T > j + k] | [T > j]) &= \frac{\mathbb{P}([T > j + k] \cap [T > j])}{\mathbb{P}([T > j])} = \frac{\mathbb{P}([T > j + k])}{\mathbb{P}([T > j])} \\ &= \frac{(1 - p)^{j+k}}{(1 - p)^j} = (1 - p)^k \\ &= \mathbb{P}([T > k]), \end{aligned}$$

comme attendu. □

L'espérance d'une variable aléatoire suivant une loi géométrique est

$$\mathbb{E}(T) = \sum_{n \geq 1} n \mathbb{P}([T = n]) = \sum_{n \geq 1} np(1 - p)^{n-1} = p \sum_{n \geq 1} n(1 - p)^{n-1}.$$

La série entière $\sum_{n \geq 0} x^n$ converge sur $] - 1, 1[$ vers $\frac{1}{1-x}$. Elle est donc dérivable terme à terme sur cet intervalle et on a

$$\frac{1}{(1 - x)^2} = \sum_{n \geq 0} nx^{n-1} = \sum_{n \geq 1} nx^{n-1}.$$

En appliquant cette égalité à $x = 1 - p$, on obtient

$$\mathbb{E}(T) = p \frac{1}{p^2} = \frac{1}{p}.$$

Exercice 21 Montrer que la variance d'une variable aléatoire T suivant une loi géométrique de paramètre p est

$$\mathbb{V}(T) = \frac{1 - p}{p^2}.$$

On introduit à présent pour tout $n \geq 1$, la variable aléatoire :

$$S_n = X_1 + \dots + X_n.$$

Pour tout $\omega \in \Omega$, $0 \leq S_n(\omega) \leq n$. Pour tout $k \in \{0, \dots, n\}$, $S_n(\omega) = k$ si exactement k des variables aléatoires X_1, \dots, X_n prennent la valeur 1 en ω :

$$[S_n = k] = \bigcup_{\substack{I \subset \{1, 2, \dots, n\}, \\ |I|=k}} (\cap_{i \in I} [X_i = 1]) \cap (\cap_{i \notin I} [X_i = 0]).$$

Dans le membre de droite, on prend l'union sur tous les sous-ensembles I de $\{1, 2, \dots, n\}$ qui sont de cardinal k . Ainsi, comme il s'agit d'une union disjointe et en utilisant l'indépendance des X_j ,

$$\mathbb{P}([S_n = k]) = \sum_{\substack{I \subset \{1, 2, \dots, n\}, \\ |I|=k}} \prod_{i \in I} \mathbb{P}([X_i = 1]) \prod_{i \notin I} \mathbb{P}([X_i = 0]) = \sum_{\substack{I \subset \{1, 2, \dots, n\}, \\ |I|=k}} p^k (1 - p)^{n-k}.$$

Comme il y a C_n^k façons de choisir k éléments parmi n , c'est-à-dire C_n^k sous-ensembles I possibles, on en déduit

$$\mathbb{P}([S_n = k]) = C_n^k p^k (1-p)^{n-k}.$$

Ainsi, la loi de S_n est la loi binomiale de paramètres n et p .

Proposition 58 *La somme de deux variables binomiales indépendantes de paramètres (k, p) et (ℓ, p) respectivement est une variable binomiale de paramètres $(k + \ell, p)$.*

Preuve : Notons $S : \Omega \rightarrow \{0, \dots, k\}$ une variable binomiale de paramètres (k, p) et $T : \Omega \rightarrow \{0, \dots, \ell\}$ une variable binomiale de paramètres (ℓ, p) . Supposons que S et T soient indépendantes. Alors $S + T : \Omega \rightarrow \{0, \dots, k + \ell\}$ vérifie

$$\forall i \in \{0, \dots, k + \ell\}, \quad [S + T = i] = \cup_{j=0}^i [S = j] \cap [T = i - j].$$

Comme les événements $[S = j] \cap [T = i - j]$ sont disjoints pour deux valeurs distinctes de j ,

$$\mathbb{P}([S + T = i]) = \sum_{j=0}^i \mathbb{P}([S = j] \cap [T = i - j]).$$

Comme S et T sont indépendantes,

$$\mathbb{P}([S + T = i]) = \sum_{j=0}^i \mathbb{P}([S = j])\mathbb{P}([T = i - j]).$$

On utilise maintenant que S et T sont binomiales :

$$\mathbb{P}([S + T = i]) = \sum_{j=0}^i C_k^j p^j (1-p)^{k-j} C_\ell^{i-j} p^{i-j} (1-p)^{\ell+i-j} = p^i (1-p)^{\ell+k-i} \sum_{j=0}^i C_k^j C_\ell^{i-j}.$$

Pour calculer $\sum_{j=0}^i C_k^j C_\ell^{i-j}$, on calcule de deux manières différentes la quantité $(1+x)^k (1+x)^\ell$. D'une part,

$$(1+x)^k (1+x)^\ell = (1+x)^{k+\ell} = \sum_{i=0}^{k+\ell} C_{k+\ell}^i x^i.$$

D'autre part,

$$\begin{aligned} (1+x)^k (1+x)^\ell &= \left(\sum_{j=0}^k C_k^j x^j \right) \left(\sum_{j'=0}^{\ell} C_\ell^{j'} x^{j'} \right) \\ &= \sum_{j=0}^k \sum_{j'=0}^{\ell} x^{j+j'} C_k^j C_\ell^{j'} \\ &= \sum_{i=0}^{k+\ell} x^i \sum_{j+j'=i} C_k^j C_\ell^{j'} = \sum_{i=0}^{k+\ell} x^i \sum_{j=0}^i C_k^j C_\ell^{i-j}. \end{aligned}$$

Par identification des coefficients de ce polynôme, on obtient donc

$$\sum_{j=0}^i C_k^j C_\ell^{i-j} = C_{k+\ell}^i.$$

Finalement,

$$\mathbb{P}([S + T = i]) = C_{k+\ell}^i p^i (1-p)^{\ell+k-i},$$

ce qui prouve le résultat attendu. □

Remarque 23 Si les variables binomiales indépendantes S et T de paramètres (k, p) et (ℓ, p) sont obtenues comme somme de variables de Bernoulli indépendantes de paramètre p , alors $S + T$ est une somme de $k + \ell$ variables de Bernoulli indépendantes de paramètre p , et suit donc une loi binomiale de paramètres $(k + \ell, p)$, ce qui corrobore le résultat précédent.

Définition 52 On dit qu'une variable aléatoire $X : \Omega \rightarrow \mathbb{N}$ suit une loi de Poisson de paramètre $\lambda > 0$, notée $\mathcal{P}(\lambda)$, si

$$\forall k \in \mathbb{N}, \quad \mathbb{P}(X = k) = e^{-\lambda} \frac{\lambda^k}{k!}.$$

Exercice 22 Une variable X suivant une loi de Poisson de paramètre λ est intégrable et de second moment fini. De plus,

$$\mathbb{E}(X) = \lambda, \quad \mathbb{V}(X) = \lambda.$$

Dans la proposition suivante, on montre que sous certaines conditions, une suite de variables aléatoires suivant une loi binomiale converge en un certain sens vers une variable aléatoire suivant une loi de Poisson.

Proposition 59 Soit $(p_n)_{n \geq 1}$ une suite de réels strictement positifs telle que $p_n \sim \frac{\lambda}{n}$ lorsque $n \rightarrow +\infty$. Soit $(X_n)_{n \geq 1}$ une suite de lois binomiales de paramètres n et p_n . Alors

$$\forall k \in \mathbb{N}, \quad C_n^k p_n^k (1-p_n)^{n-k} \rightarrow e^{-\lambda} \frac{\lambda^k}{k!}$$

Preuve : On fixe $k \in \mathbb{N}$. On écrit

$$C_n^k p_n^k (1-p_n)^{n-k} = \frac{1}{k!} \frac{n!}{(n-k)! n^k} (np_n)^k \left(1 - \frac{np_n}{n}\right)^{n-k}.$$

Par hypothèse, $\lim_{n \rightarrow +\infty} (np_n)^k = \lambda^k$. De plus,

$$\begin{aligned} \left(1 - \frac{np_n}{n}\right)^{n-k} &= e^{(n-k) \ln\left(1 - \frac{np_n}{n}\right)} \\ &= e^{(n+O(1)) \ln\left(1 - \frac{\lambda}{n} + o\left(\frac{1}{n}\right)\right)} \\ &= e^{(n+O(1))\left(-\frac{\lambda}{n} + o\left(\frac{1}{n}\right)\right)} \\ &= e^{(-\lambda + o(1))}. \end{aligned}$$

Ainsi, $\lim_{n \rightarrow +\infty} \left(1 - \frac{np_n}{n}\right)^{n-k} = e^{-\lambda}$. Enfin,

$$\frac{n!}{(n-k)! n^k} = \frac{n(n-1) \cdots (n-k+1)}{n^k} \sim \frac{n \cdot n \cdots n}{n^k}, \quad n \rightarrow +\infty.$$

On en déduit $\lim_{n \rightarrow +\infty} \frac{n!}{(n-k)! n^k} = 1$. Finalement,

$$\lim_{n \rightarrow +\infty} C_n^k p_n^k (1-p_n)^{n-k} = e^{-\lambda} \frac{\lambda^k}{k!},$$

ce qu'il fallait démontrer. □

La dernière proposition peut recevoir l'interprétation suivante. La somme S_n d'un grand nombre de variables de Bernoulli indépendantes de petit paramètre suit approximativement la loi de Poisson de paramètre $\mathbb{E}(S_n)$.

La proposition précédente fait partie de la grande famille des *théorèmes centraux limites*, qui sont l'un des objets principaux de la section suivante.

6.6 Statistiques

Soit X une variable aléatoire (discrète ou à densité). Soient $X_i, i \in \mathbb{N}^*$ des variables aléatoires indépendantes de même loi que X . On considère les variables aléatoires

$$\bar{X}_n = \frac{1}{n}(X_1 + \dots + X_n) = \frac{S_n}{n}.$$

$$\mathbb{E}(\bar{X}_n) = \mathbb{E}(X), \quad \mathbb{V}(\bar{X}_n) = \frac{1}{n}\mathbb{V}(X).$$

6.6.1 Loi faible des grands nombres

Proposition 60 Soient X_1, \dots, X_n des variables aléatoires de même loi de second moment fini. Alors pour tout $\varepsilon > 0$,

$$\mathbb{P} [|\bar{X}_n - \mathbb{E}(X_1)| > \varepsilon] \leq \frac{\mathbb{V}(X_1)}{\varepsilon^2 n}.$$

Preuve : on calcule $\mathbb{V}(\bar{X}_n - \mathbb{E}(X_1))$ puis on applique l'inégalité de Bienaymé-Tchebychev. □

Remarque 24 (Intervalle de confiance) La probabilité que l'intervalle

$$\left[\bar{X}_n - \sqrt{\frac{\mathbb{V}(X_1)}{na}}, \bar{X}_n + \sqrt{\frac{\mathbb{V}(X_1)}{na}} \right]$$

contienne $\mathbb{E}(X)$ est supérieure ou égale à $1 - a$.

Remarque 25 Application pratique : la variance est souvent inconnue mais on peut la majorer :

1. si $|X| \leq M$, alors $\mathbb{V}(X_1) \leq M^2$,
2. si X est de Bernoulli, $\mathbb{V}(X_1) \leq 1/4$.

Application numérique : pour $n = 1000$, au seuil de confiance 90% (= $1 - a$), l'incertitude est de 5% (= $\sqrt{\frac{\mathbb{V}(X_1)}{na}}$) pour une variable de Bernoulli.

6.6.2 Théorème de Moivre-Laplace

Nous admettrons le dernier résultat de ce chapitre :

Théorème 28 (Théorème de Moivre-Laplace) Soit $(X_n)_{n \geq 1}$ une suite de variables aléatoires de Bernoulli indépendantes¹ de même paramètre $0 < p < 1$. Alors pour tout $a < b$,

$$\lim_{n \rightarrow +\infty} \mathbb{P} \left(a \leq \frac{\sqrt{n}}{\sqrt{p(1-p)}} (\bar{X}_n - p) \leq b \right) = \int_a^b \frac{1}{\sqrt{2\pi}} \exp\left(-\frac{x^2}{2}\right) dx.$$

1. Ce qui signifie que pour tout $n \geq 1$, X_1, \dots, X_n sont indépendantes.

Comme la variable aléatoire

$$S_n : \omega \mapsto \sum_{k=1}^n X_k(\omega)$$

suit une loi binomiale de paramètre (n, p) , on peut reformuler cet énoncé en termes de variables aléatoires suivant une loi binomiale. En effet, soit $(Z_n)_{n \geq 1}$ une suite de variables aléatoires telles que pour tout $n \geq 1$, Z_n suit une loi binomiale de paramètres (n, p) : pour tout $k \in \mathbb{N}$, si $k > n$, alors $\mathbb{P}(Z_n = k) = 0$ tandis que si $0 \leq k \leq n$, $\mathbb{P}(Z_n = k) = C_n^k p^k (1-p)^{n-k}$. L'espérance de Z_n est np et son écart-type est $\sqrt{np(1-p)}$. Alors pour tout $\alpha < \beta$, pour tout $n \geq 1$,

$$\mathbb{P}(\alpha \leq Z_n \leq \beta) = \mathbb{P}(\alpha \leq S_n \leq \beta).$$

En prenant $\alpha = a\sqrt{np(1-p)} + np$ et $\beta = b\sqrt{np(1-p)} + np$, on a donc

$$\mathbb{P}\left(a \leq \frac{Z_n - \mathbb{E}(Z_n)}{\sqrt{\mathbb{V}(Z_n)}} \leq b\right) = \mathbb{P}\left(a \leq \frac{S_n - np}{\sqrt{p(1-p)}\sqrt{n}} \leq b\right).$$

Le théorème de Moivre-Laplace s'écrit donc

$$\lim_{n \rightarrow +\infty} \mathbb{P}\left(a \leq \frac{Z_n - \mathbb{E}(Z_n)}{\sqrt{\mathbb{V}(Z_n)}} \leq b\right) = \int_a^b \frac{1}{\sqrt{2\pi}} \exp\left(-\frac{x^2}{2}\right) dx.$$

Le premier intérêt de ce théorème est de donner une estimation de $\mathbb{P}(\alpha \leq Z_n \leq \beta)$ lorsque Z_n suit une loi binomiale, qui serait impossible à calculer en pratique si on utilisait la formule

$$\mathbb{P}(\alpha \leq \frac{1}{\sigma_n}(Z_n - \mathbb{E}(Z_n)) \leq \beta) = \sum_{k=\alpha}^{\beta} C_n^k p^k (1-p)^{n-k}.$$

Pour mettre en évidence un autre intérêt de ce théorème, on peut imaginer un lancer de pièce tombant sur pile avec une probabilité p et tombant sur face avec une probabilité $(1-p)$. On cherche à donner une bonne estimation de p . Soit $0 < \varepsilon < 1$. Comme la fonction

$$\eta \in [0, +\infty[\mapsto \int_{-\eta}^{\eta} \frac{1}{\sqrt{2\pi}} \exp\left(-\frac{x^2}{2}\right) dx$$

est continue, strictement croissante (car l'intégrande est > 0 en tout point), s'annule en 0 et tend vers 1 en $+\infty$, on en déduit à l'aide du théorème des valeurs intermédiaires qu'il existe un unique $\eta > 0$ tel que

$$\int_{-\eta}^{\eta} \frac{1}{\sqrt{2\pi}} \exp\left(-\frac{x^2}{2}\right) dx = 1 - \varepsilon.$$

Alors avec une probabilité asymptotique de $1 - \varepsilon$, la moyenne empirique \bar{X}_n s'écarte du paramètre p de moins de $\eta \frac{\sqrt{p(1-p)}}{\sqrt{n}} \leq \frac{\eta}{2\sqrt{n}}$, ce qui sera une approximation d'autant meilleure que n est grand. Naturellement, plus on est exigeant sur la probabilité (autrement dit, plus on réduit ε), plus le paramètre η grandit, et donc plus on doit faire de lancers (i.e. plus on doit augmenter n) pour préserver une marge d'erreur identique. On a ainsi un moyen d'obtenir une bonne approximation de p avec une grande probabilité.

Chapitre 7

Variables aléatoires à densité

Définition 53 Soit $\rho : \mathbb{R} \rightarrow \mathbb{R}^+$ une fonction continue à valeurs positives. La fonction $X : \Omega \rightarrow \mathbb{R}$ est une variable aléatoire de densité ρ si, pour tout intervalle $]a, b[$ de \mathbb{R} , l'ensemble

$$[X \in]a, b[= \{\omega \in \Omega : a < X(\omega) < b\}$$

est un événement et

$$\mathbb{P}(X \in]a, b[) = \int_a^b \rho(x) dx.$$

Remarque 26

$$\int_{-\infty}^{\infty} \rho = \mathbb{P}(\Omega) = 1.$$

Remarque 27 Si X est une variable à densité, alors $\mathbb{P}([X = x]) = 0$ pour tout $x \in \mathbb{R}$. Un événement de probabilité nulle n'est pas forcément impossible.

Exemple 51 Soit X une variable aléatoire de densité $f : \mathbb{R} \rightarrow \mathbb{R}$.

1. Lorsque $f = \frac{1}{b-a} 1_{[a,b]}$, où $a < b$ sont deux réels, on dit que X suit une loi uniforme. Elle intervient par exemple lorsqu'on tire au hasard un nombre entre a et b à l'aide d'un ordinateur.
2. Lorsque $f : x \in \mathbb{R} \mapsto \lambda 1_{]0, +\infty[}(x) \exp(-\lambda x)$, où λ est un réel > 0 fixé, on dit que X suit une loi exponentielle de paramètre λ .

Définition 54 (Fonction de répartition) Soit $X : \Omega \rightarrow \mathbb{R}$ une variable aléatoire. On appelle fonction de répartition de X la fonction

$$F_X : t \in \mathbb{R} \mapsto \mathbb{P}(X \leq t).$$

Par exemple, si X suit une loi exponentielle de paramètre λ ,

$$F(t) = 1 - e^{-\lambda t} \quad \text{si } t \geq 0 \quad \text{et} \quad F(t) = 0 \quad \text{sinon.}$$

Proposition 61 La fonction de répartition est croissante, tend vers 0 en $-\infty$, et vers 1 en $+\infty$.

Proposition 62 Si $X : \Omega \rightarrow \mathbb{R}$ est une variable aléatoire à densité f , alors

$$F_X(t) = \int_{-\infty}^t \rho(x) dx.$$

La fonction F_X est dérivable et $F'_X = \rho$.

Inverse de la fonction de répartition et quantiles.

Exemples de calcul de fonctions de répartition et de densités de variables images $f \circ X$ avec f monotone.

Pour une variable aléatoire X continue de densité f , on dit qu'elle est intégrable si

$$\int_{\mathbb{R}} |x|f(x) dx < \infty$$

et dans ce cas, on définit son espérance par

$$\mathbb{E}(X) = \int_{\mathbb{R}} xf(x) dx.$$

Moments :

$$\int_{\mathbb{R}} |x|^k \rho(x) dx$$

$$\mathbb{E}(X) = \int_{\mathbb{R}} x\rho(x) dx.$$

Linéarité et monotonie de l'espérance.

Proposition 63 Soit $X : \Omega \rightarrow \mathbb{R}$ une variable aléatoire et $\varphi : \mathbb{R} \rightarrow \mathbb{R}$ une fonction continue. Si $\varphi \circ X$ est intégrable, alors

$$\mathbb{E}(\varphi(X)) = \int_{\mathbb{R}} \varphi(x)\rho(x) dx.$$

On définit comme avant la variance $\mathbb{V}(X) = \mathbb{E}(X^2) - (\mathbb{E}(X))^2$.

$$\mathbb{V}(aX + b) = a^2\mathbb{V}(X).$$

Ecart-type $\sigma(X) = \sqrt{\mathbb{V}(X)}$. La variable

$$X^* := \frac{X - \mathbb{E}(X)}{\sigma(X)}$$

est centrée réduite. Elle a pour densité

$$\rho^*(x) = \sigma(X)\rho(\mathbb{E}(X) + \sigma(X)x).$$

Lorsque X est de densité f , on a donc :

$$\mathbb{V}(X) = \int_{\mathbb{R}} x^2 f(x) dx - \left(\int_{\mathbb{R}} xf(x) dx \right)^2.$$

Toutes les propriétés énoncées précédemment restent vraies : linéarité, inégalité triangulaire, monotonie, inégalité de Cauchy-Schwarz, lien avec l'indépendance, loi faible des grands nombres.

Exemple 52 Soit X une variable aléatoire de densité

$$f : x \in \mathbb{R} \mapsto \frac{1}{\sqrt{2\pi}\sigma} \exp\left(-\frac{(x-m)^2}{2\sigma^2}\right)$$

où $m \in \mathbb{R}$ et $\sigma > 0$ sont deux paramètres fixés. On dit que X suit une loi gaussienne de moyenne m et de variance σ^2 .

Dans le dernier chapitre, on démontrera que

$$\frac{1}{\sqrt{2\pi}\sigma} \int_{\mathbb{R}} \exp\left(-\frac{(x-m)^2}{2\sigma^2}\right) dx = 1.$$

Il est en revanche facile de vérifier que l'espérance de X vaut m (on reconnaît une primitive à vue) et sa variance vaut σ^2 (on utilise une intégration par parties).

Proposition 64 1. **Inégalité de Markov** Si X est une variable aléatoire à valeurs positives et $a > 0$, alors

$$\mathbb{P}([X \geq a]) \leq \frac{\mathbb{E}(X)}{a}$$

2. **Inégalité de Bienaymé-Tchebychev** Si X est une variable aléatoire de second moment fini et $a > 0$, alors

$$\mathbb{P}(|X - \mathbb{E}(X)| \geq a) \leq \frac{\mathbb{V}(X)}{a^2}.$$

7.1 Loi normale, loi exponentielle

Définition 55 Une variable aléatoire gaussienne (ou normale) centrée réduite est une variable X admettant la densité

$$\rho(x) = \frac{1}{\sqrt{2\pi}} e^{-\frac{x^2}{2}}.$$

$\mathbb{E}(X) = 0, \mathbb{V}(X) = 1$. La variable $Y = \sigma X + m$ est alors une variable gaussienne de moyenne m et de variance σ^2 . Elle admet la densité

$$\rho(x) = \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{(x-m)^2}{2\sigma^2}}.$$

Loi exponentielle $\rho(x) = \lambda e^{-\lambda x}$ sur \mathbb{R}^+ .

$$\mathbb{E}(X) = \frac{1}{\lambda}, \mathbb{V}(X) = \frac{1}{\lambda^2}.$$

Pour tout $x \geq 0$,

$$\mathbb{P}([X > x]) = e^{-\lambda x}.$$

Les lois exponentielles satisfont la propriété

$$\forall s, t \geq 0, \mathbb{P}(X > t + s | X > t) = \mathbb{P}(X > s).$$

Cette propriété caractérise les lois exponentielles. Les lois exponentielles s'interprètent comme les lois de durée de vie sans vieillissement. Ce sont des variantes continues des lois géométriques.

7.2 Indépendance de variables à densité

Définition 56 Les variables aléatoires à densité X_1, \dots, X_n sont dites indépendantes si

$$\mathbb{P}([X_1 \in I_1, \dots, X_n \in I_n]) = \mathbb{P}([X_1 \in I_1]) \dots \mathbb{P}([X_n \in I_n])$$

pour tous intervalles I_k de \mathbb{R} .

Proposition 65 Si X_1, X_2 sont indépendantes et admettent un moment d'ordre 2 fini,

$$\mathbb{E}(X_1 X_2) = \mathbb{E}(X_1) \mathbb{E}(X_2) \quad , \quad \mathbb{V}(X_1 + X_2) = \mathbb{V}(X_1) + \mathbb{V}(X_2).$$

Reste vrai pour n variables aléatoires indépendantes.

On définit comme pour les variables discrètes la covariance et le coefficient de corrélation.

Chapitre 8

Compléments

8.1 L'équation de la chaleur

Historiquement, les séries de Fourier ont été introduites pour résoudre l'équation de la chaleur. On considère le cas simple suivant : on connaît à l'instant initial la température en chaque point d'une barre, et aussi la température aux extrémités à tout instant. Est-on capable de connaître la température de la barre en tout point de celle-ci, et à tout instant ?

On assimile la barre à un segment $[0, L]$ et on note $u(x, t)$ la température à l'instant $t > 0$ et au point d'abscisse x de la barre. L'équation de la chaleur est posée sur l'ouvert $Q =]0, L[\times]0, +\infty[$. On cherche une fonction u sur $[0, L] \times [0, +\infty[$ telle que

$$\frac{\partial u}{\partial t} - \frac{\partial^2 u}{\partial x^2} = 0 \quad \text{dans } Q,$$

avec pour conditions aux limites

$$u(0, t) = u(L, t) = 0 \quad \text{à tout instant } t \geq 0,$$

et avec condition initiale

$$u(x, 0) = h(x) \quad \text{en tout point de la barre } x \in [0, L].$$

On suppose ici que h est une fonction continue sur $[0, L]$ et de plus, $h(0) = 0 = h(L)$ (condition nécessaire de compatibilité entre les conditions au bord et la condition initiale).

Voilà beaucoup de conditions à vérifier pour une seule fonction ! Dans un premier temps, on va simplifier le problème en oubliant la condition initiale, et en cherchant des fonctions qui vérifient seulement l'équation de la chaleur avec les conditions au bord.

On commence par chercher une solution par la méthode de séparation des variables : supposons donc que u soit de la forme :

$$u(x, t) = f(x)g(t)$$

où f et g sont des fonctions C^2 à déterminer. On injecte cette fonction u dans l'équation et on obtient :

$$g'(t)f(x) = f''(x)g(t).$$

Supposons de plus qu'on puisse obtenir une solution en exigeant de plus que ni f ni g ne s'annule (sur $]0, L[$ et $]0, +\infty[$ respectivement). Alors

$$\frac{f''(x)}{f(x)} = \frac{g'(t)}{g(t)} \quad \forall x \in]0, L[, \forall t \in]0, +\infty[.$$

Ceci n'est possible que si les deux membres sont constants, i.e. s'il existe $\lambda \in \mathbb{R}$ tel que

$$\forall x \in]0, L[, f''(x) = \lambda f(x) \quad , \quad \forall t \in]0, +\infty[, g'(t) = \lambda g(t).$$

On résout ces équation en fonction du signe de λ . La première équation est une équation différentielle linéaire homogène du deuxième ordre à coefficients constants. Lorsque $\lambda > 0$, sa solution générale est de la forme

$$f(x) = Ae^{\sqrt{\lambda}x} + Be^{-\sqrt{\lambda}x}$$

pour certaines constantes A et B à déterminer. En exploitant les conditions au bord, on obtient :

$$A + B = 0 \quad , \quad Ae^{\sqrt{\lambda}L} + Be^{-\sqrt{\lambda}L} = 0.$$

On en déduit $A = B = 0$ et donc $u = 0$.

De même, si $\lambda = 0$, la solution générale de l'équation pour f est de la forme

$$f(x) = Ax + B$$

et les conditions au bord conduisent de même à $A = B = 0$. Enfin, si $\lambda < 0$, on obtient

$$f(x) = A \cos(\sqrt{-\lambda}x) + B \sin(\sqrt{-\lambda}x),$$

avec, à cause des conditions au bord,

$$A = 0 \quad , \quad B \sin(\sqrt{-\lambda}L) = 0.$$

Si on exclut la solution nulle (et donc $B = 0$), cela implique que $\sqrt{-\lambda}$ est de la forme $\ell\pi/L$ pour un certain $\ell \in \mathbb{N}$. La fonction g vérifie une équation linéaire homogène du premier ordre à coefficients constants et on obtient

$$g(t) = Ce^{\lambda t} = Ce^{-\frac{\ell^2\pi^2}{L^2}t}.$$

On a donc trouvé pour chaque valeur de $\ell \in \mathbb{N}$ une solution de la forme

$$(x, t) \mapsto BCe^{-\frac{\ell^2\pi^2}{L^2}t} \sin \frac{\ell\pi}{L}x.$$

Conclusion partielle : pour tout $\ell \in \mathbb{N}$, pour tout $a_\ell \in \mathbb{R}$, la fonction

$$u_\ell(x, t) = a_\ell e^{-\frac{\ell^2\pi^2}{L^2}t} \sin \frac{\ell\pi}{L}x$$

est solution de l'équation de la chaleur, qui satisfait les conditions au bord. En revanche, aucune de ces fonctions ne satisfait *a priori* la condition initiale.

Mais comme l'équation est linéaire, une somme de solutions est encore une solution du problème. L'idée est alors de chercher une solution sous la forme

$$u(x, t) = \sum_{\ell=0}^{+\infty} a_\ell e^{-\frac{\ell^2\pi^2}{L^2}t} \sin \frac{\ell\pi}{L}x.$$

Supposons que cette série de fonctions soit convergente (cela dépend *a priori* de la suite a_ℓ qui reste à déterminer). Alors $u(x, t)$ est bien défini pour chaque $x \in [0, L]$ et pour chaque $t \geq 0$. Cette fonction vérifiera les conditions au bord. Il s'agit de comprendre comment choisir les coefficients a_ℓ pour que cette fonction u vérifie la condition initiale $u(x, 0) = h(x)$, autrement dit

$$h(x) = \sum_{\ell=0}^{+\infty} a_\ell \sin \frac{\ell\pi}{L}x.$$

La question devient donc : peut-on trouver une suite $(a_\ell)_{\ell \in \mathbb{N}}$ qui permette d'écrire ainsi h comme la somme d'une série. Cela fait penser à un développement en série de Fourier d'une fonction impaire...

Supposons dans un premier temps que $L = \pi$ (on verra comment traiter le cas général ultérieurement). On prolonge h par imparité en posant $\tilde{h}(x) = -h(-x)$ pour tout $x \in [-\pi, 0]$ (noter que $\tilde{h}(-\pi) = \tilde{h}(0) = \tilde{h}(\pi) = 0$) puis par 2π périodicité : $\tilde{h}(x + 2k\pi) = \tilde{h}(x)$ pour tout $x \in [-\pi, \pi]$ et tout $k \in \mathbb{Z}$. On obtient donc une fonction \tilde{h} qui est continue et 2π périodique sur \mathbb{R} . Une telle fonction n'est PAS en général égale à la limite de série de Fourier. En revanche, **si on suppose de plus h de classe C^1** , alors la fonction \tilde{h} est C^1 (elle est bien sûr C^1 sur $]0, \pi[$, on vérifie aussi en utilisant l'imparité et le fait que $h(0) = 0$ que la dérivée de \tilde{h} est continue en 0 et π puis en tous les points de la forme $k\pi$, $k \in \mathbb{Z}$). Dans ce cas, on sait que la série de Fourier de \tilde{h} converge normalement vers \tilde{h} , autrement dit, la série $\sum_\ell |b_\ell(\tilde{h})|$ converge et pour tout $x \in [0, \pi]$:

$$\tilde{h}(x) = \sum_{\ell=0}^{+\infty} b_\ell(\tilde{h}) \sin \ell x.$$

On pose alors pour tout $\ell \in \mathbb{N}$:

$$a_\ell = b_\ell(\tilde{h}).$$

Soit $x \in [0, \pi]$ et $t \in [0, +\infty[$. On veut montrer que la série

$$\sum_{\ell=0}^{+\infty} a_\ell e^{-\ell^2 t} \sin \ell x$$

converge. Lorsque $t = 0$, on reconnaît la série de Fourier de \tilde{h} dont on sait qu'elle converge (et même absolument). Lorsque $t > 0$, on peut écrire

$$|a_\ell e^{-\ell^2 t} \sin \ell x| \leq |a_\ell| e^{-\ell^2 t} \leq |a_\ell|.$$

La série $\sum_\ell |a_\ell|$ converge. Par comparaison, la série $\sum_\ell a_\ell e^{-\ell^2 t} \sin \ell x$ converge. On peut donc poser pour tout $x \in [0, L]$ et pour tout $t \geq 0$

$$u(x, t) = \sum_{\ell=0}^{+\infty} a_\ell e^{-\ell^2 t} \sin \ell x.$$

Cette fonction vérifie bien la condition initiale et la condition au bord. Il reste à montrer que c'est aussi une solution de l'équation. On fixe $x \in]0, \pi[$ et on montre que la fonction $t \mapsto u(x, t)$ est dérivable sur $]0, +\infty[$. Pour cela, il suffit de montrer qu'elle est dérivable sur chaque intervalle de la forme $]\varepsilon, A[$ où $0 < \varepsilon < A$. On va utiliser le théorème de dérivation des séries de fonctions : lorsqu'une série de fonctions dérivables converge simplement, et que la série des dérivées converge uniformément, alors la somme est dérivable et égale à la somme des dérivées.

La série des $u_\ell(x, t) = a_\ell e^{-\ell^2 t} \sin \ell x$ converge simplement et chaque terme u_ℓ est une fonction dérivable. Donc il suffit de montrer que la série $\sum_\ell \frac{\partial u_\ell}{\partial t}$ converge uniformément sur $]\varepsilon, A[$. On a

$$\frac{\partial u_\ell}{\partial t}(x, t) = a_\ell (-\ell^2) e^{-\ell^2 t} \sin \ell x.$$

Donc

$$\sup_{t \in]\varepsilon, A[} \left| \frac{\partial u_\ell}{\partial t}(x, t) \right| \leq |a_\ell| \ell^2 e^{-\ell^2 \varepsilon}.$$

Pour tout ℓ assez grand, $|\ell^2 e^{-\ell^2 \varepsilon}| \leq 1$. Donc par comparaison, la série $\sum_{\ell} |a_{\ell}| \ell^2 e^{-\ell^2 \varepsilon}$ converge, donc la série $\sum_{\ell} \frac{\partial u_{\ell}}{\partial t}(x, t)$ converge normalement sur $] \varepsilon, A[$ et donc uniformément sur cet intervalle. Comme c'est vrai pour tout $\varepsilon > 0$, on peut conclure que u admet une dérivée partielle par rapport à t et

$$\frac{\partial u}{\partial t}(x, t) = \sum_{\ell=0}^{+\infty} \frac{\partial u_{\ell}}{\partial t}(x, t).$$

On montre de même que u admet une dérivée partielle seconde par rapport à x et de plus

$$\frac{\partial^2 u}{\partial x^2}(x, t) = \sum_{\ell=0}^{+\infty} \frac{\partial^2 u_{\ell}}{\partial x^2}(x, t).$$

On en déduit que

$$\frac{\partial u}{\partial t}(x, t) - \frac{\partial^2 u}{\partial x^2}(x, t) = \sum_{\ell=0}^{+\infty} \frac{\partial u_{\ell}}{\partial t}(x, t) - \frac{\partial^2 u_{\ell}}{\partial x^2}(x, t) = 0,$$

ce qui achève le problème lorsque $L = \pi$.

Dans le cas général, on doit expliquer comment développer en série de Fourier une fonction périodique qui n'est pas 2π périodique. Plus précisément, à partir d'une fonction $h : [0, L] \rightarrow \mathbb{R}$ s'annulant en 0 et π , on peut la prolonger par imparité en posant $\tilde{h}(x) = -h(-x)$ si $x \in [-L, 0]$, puis par $2L$ périodicité en posant $\tilde{h}(x + 2kL) = \tilde{h}(x)$ pour tout $x \in [-L, L]$ et tout $k \in \mathbb{Z}$. On obtient ainsi une fonction \tilde{h} qui est $2L$ périodique.

On pose ensuite $H(x) = \tilde{h}(\frac{L}{\pi}x)$. Alors pour tout $k \in \mathbb{Z}$,

$$H(x + 2k\pi) = \tilde{h}(\frac{L}{\pi}(x + 2k\pi)) = \tilde{h}(\frac{L}{\pi}x + 2kL) = \tilde{h}(\frac{L}{\pi}x) = H(x).$$

Comme \tilde{h} est C^1 , on en déduit que H est également C^1 . On peut donc appliquer la théorie de Fourier à H :

$$H(x) = \sum_{\ell=0}^{+\infty} b_{\ell}(H) \sin \ell x.$$

On en déduit

$$\tilde{h}(x) = H(\frac{\pi}{L}x) = \sum_{\ell=0}^{+\infty} b_{\ell}(H) \sin \frac{\pi \ell}{L} x$$

et il suffit alors de poser $a_{\ell} = b_{\ell}(H)$ dans la définition de u . La suite de la preuve est sans changement.

Remarque 28 On peut aussi montrer que la fonction u ainsi obtenue est continue sur $[0, L] \times [0, +\infty[$ et C^{∞} sur $]0, L[\times]0, +\infty[$.