Chapitre 19

Espaces préhilbertiens réels

19.1 Introduction historique

De tous les principes qu'on peut proposer pour cet objet, je pense qu'il n'en est pas de plus général, de plus exact, ni d'une application plus facile que celui dont nous avons fait usage dans les recherches précédentes, et qui consiste à rendre minimum la somme des quarrés des erreurs. Par ce moyen, il s'établit entre les erreurs une sorte d'équilibre qui empêchant les extrêmes de prévaloir, est très-propre à faire connoître l'état du système le plus proche de la vérité.

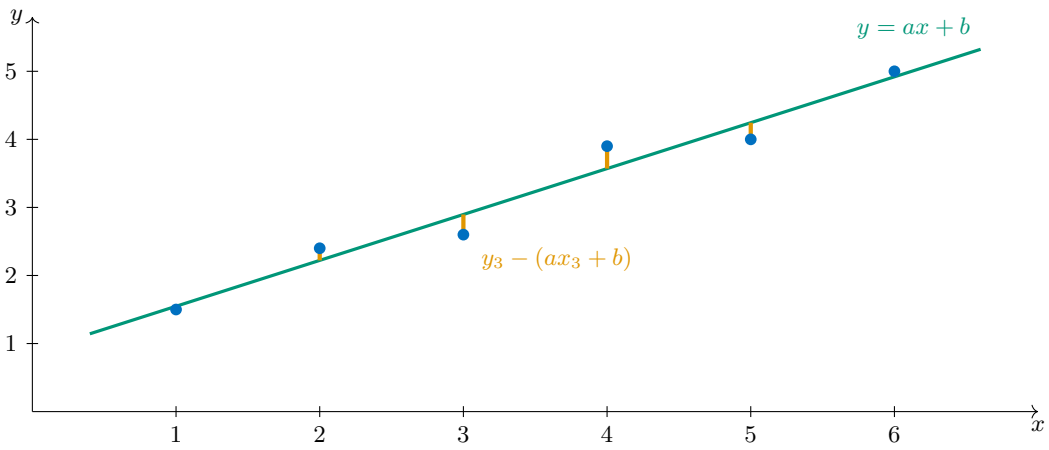
Adrien-Marie Legendre - *Nouvelles méthodes pour la détermination des orbites des comètes*, Paris, 1805

19.1.1 Un principe d'astronome

En 1805, dans l'appendice de son traité sur les orbites des comètes, **Legendre** affronte un problème d'observateur : les mesures fournissent plus d'équations que d'inconnues, et, entachées d'erreurs, elles ne peuvent être satisfaites toutes à la fois. Il propose de choisir les inconnues qui rendent minimale la somme des carrés des erreurs. Le présent chapitre montre que ce principe est géométrique : une somme de carrés est le carré d'une distance, et la meilleure solution est le pied d'une perpendiculaire, c'est-à-dire un projeté orthogonal (section 19.6).

19.1.2 Des sommes aux fonctions

En 1821, dans les notes de son *Cours d'analyse*, **Cauchy** démontre pour des sommes l'inégalité qui porte aujourd'hui son nom ; **Bouniakovski** en 1859, puis **Schwarz** en 1888, en donnent la version pour les intégrales. En 1883, **Gram** orthogonalise une famille de fonctions pour les développer en série par la méthode des moindres carrés ; **Schmidt**, en 1907, reprend le procédé dans ses travaux sur les équations intégrales, ouverts par **Hilbert** à partir de 1904, où une fonction est traitée comme un vecteur d'un espace de dimension infinie. Le chapitre suit ce mouvement : il définit le produit scalaire (section 19.2), en tire la norme et l'inégalité de Cauchy-Schwarz (19.3), l'orthogonalité et l'algorithme de Gram-Schmidt (19.4), les bases orthonormées (19.5), et revient enfin au problème de Legendre. La figure en montre le cas le plus simple : la droite verte rend minimale la somme des carrés des écarts verticaux, en orange, entre les points mesurés et la droite.



19.2 Produit scalaire

Dans tout le chapitre, les espaces vectoriels sont réels : E désigne un $\mathbb{R}$-espace vectoriel. Le produit scalaire du lycée, défini dans le plan et l'espace à partir des longueurs et des angles, est ici caractérisé par ses propriétés algébriques, comme le déterminant l'a été au chapitre 17 ; il devient alors possible de parler de longueur et d'angle droit dans des espaces de matrices ou de fonctions.

19.2.1 Définition

Définition : Un **produit scalaire** sur E est une application $\varphi : E \times E \to \mathbb{R}$ qui est
- **bilinéaire** : pour tout $y \in E$, l'application $x \mapsto \varphi(x,y)$ est linéaire, et pour tout $x \in E$, l'application $y \mapsto \varphi(x,y)$ est linéaire ;
- **symétrique** : $\varphi(y,x) = \varphi(x,y)$ pour tous x, $y \in E$;
- **définie positive** : $\varphi(x,x) \geqslant 0$ pour tout $x \in E$, et $\varphi(x,x) = 0 \implies x = 0$.

On note $\langle x,y\rangle$ le réel $\varphi(x,y)$; on rencontre aussi les notations $(x \mid y)$ et $x \cdot y$. Un **espace préhilbertien réel** est un $\mathbb{R}$-espace vectoriel muni d'un produit scalaire ; un **espace euclidien** est un espace préhilbertien réel de dimension finie.

Pour vérifier la bilinéarité d'une application symétrique, il suffit d'établir la linéarité par rapport à la première variable : la symétrie la transporte sur la seconde. La linéarité donne $\langle 0,y\rangle = 0$ pour tout y, puisqu'une application linéaire envoie 0 sur 0 ; la condition de définition dit que 0 est le seul vecteur orthogonal à lui-même, et on la vérifie toujours en dernier, car c'est elle qui demande un argument.

Proposition 1 : *(Produits scalaires canoniques.)*

(i) Sur $\mathbb{R}^n$, la formule $\langle x,y\rangle = \sum_{k=1}^{n} x_k y_k$ définit un produit scalaire. En identifiant x et y aux colonnes X, $Y \in \mathcal{M}_{n,1}(\mathbb{R})$, on a $\langle x,y\rangle = X^T Y$.

(ii) Sur $\mathcal{M}_{n,p}(\mathbb{R})$, la formule $\langle A,B\rangle = \mathrm{tr}(A^T B) = \sum_{i=1}^{n}\sum_{j=1}^{p} a_{ij} b_{ij}$ définit un produit scalaire.

∘ Démonstration :

▷ *(i)* L'application est symétrique, et linéaire par rapport à x. Pour tout x, $\langle x,x\rangle = \sum x_k^2$ est une somme de carrés de réels, donc positive, et elle n'est nulle que si chaque terme l'est, c'est-à-dire si $x = 0$. Le produit $X^T Y$ est une matrice 1×1, dont l'unique coefficient est $\sum x_k y_k$: on l'identifie à ce réel.

▷ *(ii)* Le coefficient d'indice (j,j) de $A^T B$ est $\sum_i a_{ij} b_{ij}$, puisque la ligne j de A^T est la colonne j de A ; en sommant sur j, on obtient la formule double. C'est le produit scalaire canonique de $\mathbb{R}^{np}$, appliqué aux familles des coefficients, et la démonstration de (i) s'applique. □

Remarques :

- Pour $A = \begin{pmatrix} 1 & 2 & 0 \\ -1 & 3 & 1 \end{pmatrix}$ et $B = \begin{pmatrix} 2 & 0 & 1 \\ 1 & 1 & -1 \end{pmatrix}$, on trouve $\langle A,B\rangle = 2+0+0-1+3-1 = 3$: on multiplie les coefficients de même position et l'on additionne, sans effectuer le produit matriciel $A^T B$ en entier.
- L'exercice 6 de la feuille du chapitre 16 établissait que $A^T A = 0$ entraîne $A = 0$ pour une matrice réelle : c'est le caractère défini de ce produit scalaire, puisque $\mathrm{tr}(A^T A) = \|A\|^2$ avec la notation de la section suivante. Sa question 3 montrait que l'argument tombe en défaut sur $\mathbb{C}$, où $1^2 + \mathrm{i}^2 = 0$. C'est pourquoi le chapitre ne considère que des espaces réels : sur $\mathbb{C}$, on remplace A^T par la transposée conjuguée et $\sum x_k y_k$ par $\sum \overline{x_k}\, y_k$, ce que l'on étudiera en seconde année.

19.2.2 Produits scalaires définis par une intégrale

Proposition 2 : Soit $a < b$ deux réels et $w : [a\,,b] \to \mathbb{R}$ une fonction continue, positive, qui ne s'annule qu'en un nombre fini de points, appelée **poids**.

(i) La formule $\langle f, g\rangle = \displaystyle\int_a^b f(t)g(t)w(t)\,\mathrm{d}t$ définit un produit scalaire sur $\mathcal{C}([a\,,b],\mathbb{R})$. Pour $w = 1$, c'est le produit scalaire $\langle f, g\rangle = \displaystyle\int_a^b fg$.

(ii) La même formule, appliquée aux fonctions polynomiales, définit un produit scalaire sur $\mathbb{R}[X]$.

∘ Démonstration :

▷ *(i)* L'application est symétrique, et linéaire par rapport à f par linéarité de l'intégrale (chapitre 18, théorème 3 (i)). Pour tout f, la fonction f^2w est continue et positive, donc $\langle f, f\rangle \geqslant 0$ par positivité de l'intégrale. Si $\langle f, f\rangle = 0$, le théorème 4 du chapitre 18 montre que f^2w est nulle sur $[a\,,b]$: $f(t) = 0$ en tout point t où $w(t) \neq 0$, c'est-à-dire partout sauf en un nombre fini de points. Chacun de ces points est limite de points où f est nulle, et f y est continue : f y est nulle aussi.

▷ *(ii)* On a les mêmes propriétés de bilinéarité, de symétrie et de positivité. Si $\langle P, P\rangle = 0$, le point (i) montre que la fonction polynomiale $t \mapsto P(t)$ est nulle sur $[a\,,b]$: le polynôme P a une infinité de racines, donc il est nul (chapitre 13, théorème 3). □

C'est la promesse du chapitre 18 : le théorème 4 de ce chapitre, sur les fonctions continues positives d'intégrale nulle, est exactement ce qui rend ce produit scalaire défini.

Remarques :

- L'erreur la plus répandue est d'appliquer la proposition 2 aux fonctions continues par morceaux. La formule $\int_a^b fg$ y est encore bilinéaire, symétrique et positive, mais elle n'est pas définie : l'indicatrice de $\{a\}$ n'est pas nulle, et $\int_a^b \mathbf{1}_{\{a\}}^2 = 0$ (chapitre 18, proposition 6). Sur l'espace $\mathcal{C}_m([a\,,b],\mathbb{R})$ des fonctions continues par morceaux, ce n'est donc pas un produit scalaire.
- Le produit $\langle P, Q\rangle = \sum_{k=0}^n P(x_k)Q(x_k)$, où $x_0, \ldots, x_n$ sont $n+1$ réels distincts, est un produit scalaire sur $\mathbb{R}_n[X]$: si $\langle P, P\rangle = 0$, le polynôme P, de degré au plus n, a $n+1$ racines, donc il est nul. Il n'en est pas un sur $\mathbb{R}[X]$: pour $x_k = k$ et $n = 3$, le polynôme non nul $X(X-1)(X-2)(X-3)$ vérifie $\langle P, P\rangle = 0$. L'espace sur lequel on travaille fait partie de la donnée.
- Le poids $w(t) = 1-t^2$ sur $[-1\,,1]$ s'annule aux deux extrémités, mais il satisfait les hypothèses de la proposition 2. Il donne, par exemple, $\langle 1, X^2\rangle = \int_{-1}^1 t^2(1-t^2)\,\mathrm{d}t = \frac{4}{15}$, au lieu de $\frac{2}{3}$ pour le poids constant : changer de poids, c'est changer la géométrie de l'espace.

19.2.3 Produits scalaires sur $\mathbb{R}^n$ et matrices symétriques

Soit $(e_1, \ldots, e_n)$ la base canonique de $\mathbb{R}^n$. En développant par bilinéarité, on voit que toute application bilinéaire φ sur $\mathbb{R}^n$ s'écrit

$$\varphi(x, y) = \varphi\Big(\sum_{i=1}^n x_i e_i, \sum_{j=1}^n y_j e_j\Big) = \sum_{i=1}^n \sum_{j=1}^n x_i y_j\, \varphi(e_i, e_j) = X^T A Y, \qquad \text{avec} \qquad A = \big(\varphi(e_i, e_j)\big)_{i,j}.$$

L'application φ est symétrique si et seulement si A l'est, puisque $\varphi(e_i, e_j) = a_{ij}$.

Exemples :

- Avec $S = \begin{pmatrix} 2 & 1 \\ 1 & 1 \end{pmatrix}$, l'application $\varphi(x, y) = X^T S Y = 2x_1y_1 + x_1y_2 + x_2y_1 + x_2y_2$ est bilinéaire et symétrique, et
$$\varphi(x, x) = 2x_1^2 + 2x_1x_2 + x_2^2 = (x_1 + x_2)^2 + x_1^2.$$
Cette somme de carrés est positive, et nulle seulement si $x_1 = 0$ et $x_1 + x_2 = 0$, c'est-à-dire si $x = 0$: c'est un produit scalaire sur $\mathbb{R}^2$, différent du produit canonique. Pour lui, $\varphi(e_1, e_2) = 1$: les vecteurs de la base canonique ne sont plus « perpendiculaires ».

- L'application $\varphi(x,y) = x_1y_1 - x_2y_2$ est bilinéaire et symétrique, mais $\varphi(e_2,e_2) = -1$: elle n'est pas positive. L'application $\varphi(x,y) = x_1y_2 + x_2y_1$ ne l'est pas non plus, puisque $\varphi(x,x) = 2x_1x_2$ vaut -2 pour $x = (1,-1)$; et elle n'est pas définie, puisque $\varphi(e_1,e_1) = 0$.

Remarque : L'exercice 7 de la feuille du chapitre 16 montrait que $X \mapsto X^TAX$ ne dépend que de la partie symétrique $\frac{1}{2}(A + A^T)$ de A. C'est la raison pour laquelle on ne cherche les produits scalaires de $\mathbb{R}^n$ que parmi les X^TSY avec S symétrique : la symétrie est imposée par la définition, et la positivité ne verrait de toute façon pas la partie antisymétrique. Reconnaître les matrices S pour lesquelles $X^TSX > 0$ dès que $X \neq 0$ est une question de seconde année.

19.3 Norme associée et inégalité de Cauchy-Schwarz

Dans cette section, E est un espace préhilbertien réel.

19.3.1 Norme et distance

Définition : La **norme** associée au produit scalaire est l'application $x \mapsto \|x\| = \sqrt{\langle x,x\rangle}$, bien définie puisque $\langle x,x\rangle \geqslant 0$. Un vecteur x est **unitaire** si $\|x\| = 1$. La **distance** de x à y est le réel $d(x,y) = \|y - x\|$.

Pour le produit canonique de $\mathbb{R}^2$ ou de $\mathbb{R}^3$, on retrouve la longueur du lycée, $\|x\| = \sqrt{x_1^2 + x_2^2 + x_3^2}$. Pour celui de $\mathcal{M}_{n,p}(\mathbb{R})$, $\|A\|^2 = \sum_{i,j} a_{ij}^2$. Pour le produit intégral de la proposition 2, avec $w = 1$, la norme $\|f\| = \left(\int_a^b f^2\right)^{1/2}$ est la quantité $N_2(f)$ de l'exercice 15 de la feuille du chapitre 18 : c'est une moyenne quadratique de $|f|$, à un facteur $\sqrt{b-a}$ près.

Proposition 3 : Pour tous x, $y \in E$ et tout $\lambda \in \mathbb{R}$:

(i) $\|x\| \geqslant 0$, avec égalité si et seulement si $x = 0$, et $\|\lambda x\| = |\lambda|\,\|x\|$. En particulier, si $x \neq 0$, le vecteur $\frac{x}{\|x\|}$ est unitaire.

(ii) $\|x+y\|^2 = \|x\|^2 + \|y\|^2 + 2\langle x,y\rangle$ et $\|x-y\|^2 = \|x\|^2 + \|y\|^2 - 2\langle x,y\rangle$.

(iii) *Identités de polarisation :* $\langle x,y\rangle = \frac{1}{2}\big(\|x+y\|^2 - \|x\|^2 - \|y\|^2\big) = \frac{1}{4}\big(\|x+y\|^2 - \|x-y\|^2\big)$.

(iv) *Identité du parallélogramme :* $\|x+y\|^2 + \|x-y\|^2 = 2\big(\|x\|^2 + \|y\|^2\big)$.

∘ Démonstration :

▷ *(i)* La racine carrée d'un réel positif est positive, et nulle seulement pour 0 : le cas d'égalité est le caractère défini. Par bilinéarité, $\langle \lambda x, \lambda x\rangle = \lambda^2\langle x,x\rangle$, et $\sqrt{\lambda^2} = |\lambda|$. Pour $x \neq 0$, on applique ce qui précède avec $\lambda = \frac{1}{\|x\|}$.

▷ *(ii)* On développe par bilinéarité, puis on utilise la symétrie :

$$\langle x+y, x+y\rangle = \langle x,x\rangle + \langle x,y\rangle + \langle y,x\rangle + \langle y,y\rangle = \|x\|^2 + 2\langle x,y\rangle + \|y\|^2.$$

La seconde formule s'en déduit en remplaçant y par $-y$.

▷ *(iii) et (iv)* La première identité de polarisation est (ii) résolue en $\langle x,y\rangle$. En retranchant, puis en ajoutant, les deux formules de (ii), on obtient la seconde identité de polarisation et l'identité du parallélogramme. □



À gauche, les diagonales du parallélogramme construit sur x et y sont $x+y$ et $x-y$; ici

$$\|x+y\|^2 + \|x-y\|^2 = 21{,}64 + 6{,}28 = 27{,}92 = 2\,(10{,}4 + 3{,}56) = 2\big(\|x\|^2 + \|y\|^2\big) :$$

la somme des carrés des diagonales est égale à la somme des carrés des quatre côtés. À droite, les mêmes vecteurs mis bout à bout illustrent l'inégalité triangulaire démontrée plus loin : $\|x+y\| \approx 4{,}65$ est plus court que $\|x\| + \|y\| \approx 5{,}11$.

Remarques :

- Les identités de polarisation disent que la norme détermine le produit scalaire : deux produits scalaires qui ont la même norme associée sont égaux. On les utilise pour transformer un énoncé sur les longueurs en un énoncé sur les produits scalaires.
- L'erreur la plus répandue est d'écrire $\|x+y\| = \|x\| + \|y\|$, ou $\|x+y\|^2 = \|x\|^2 + \|y\|^2$ sans condition. La formule (ii) est l'analogue de $(a+b)^2 = a^2 + 2ab + b^2$, et le « double produit » $2\langle x, y\rangle$ ne disparaît que si x et y sont orthogonaux : c'est le théorème de Pythagore de la section 19.4.

19.3.2 L'inégalité de Cauchy-Schwarz

Théorème 1 : Inégalité de Cauchy-Schwarz :
Pour tous x, $y \in E$,

$$|\langle x, y\rangle| \leqslant \|x\| \, \|y\|,$$

avec égalité si et seulement si la famille (x, y) est liée.

∘ Démonstration :

▷ *Le cas $y = 0$.* Les deux membres sont nuls, et la famille $(x, 0)$ est liée : l'énoncé est vrai.

▷ *Un trinôme positif.* Supposons $y \neq 0$. Pour tout réel t, la proposition 3 (ii) donne

$$P(t) = \|x + ty\|^2 = \|y\|^2 t^2 + 2\langle x, y\rangle \, t + \|x\|^2 \geqslant 0.$$

Comme $\|y\|^2 > 0$, la fonction P est un trinôme du second degré qui ne prend que des valeurs positives : il a au plus une racine réelle, donc son discriminant est négatif ou nul,

$$4\langle x, y\rangle^2 - 4\|x\|^2\|y\|^2 \leqslant 0,$$

ce qui est l'inégalité annoncée, après passage à la racine carrée.

▷ *Le cas d'égalité.* Si (x, y) est liée, avec $y \neq 0$, on a $x = \lambda y$, et $|\langle \lambda y, y\rangle| = |\lambda| \, \|y\|^2 = \|x\| \, \|y\|$. Réciproquement, s'il y a égalité, le discriminant est nul : le trinôme a une racine réelle t_0, et $\|x + t_0 y\| = 0$. Par la proposition 3 (i), $x = -t_0 y$: la famille est liée. □

Remarques :

- La démonstration n'utilise le caractère défini que pour le cas d'égalité. L'inégalité reste donc vraie pour une application bilinéaire symétrique positive φ, par exemple $(f, g) \mapsto \int_a^b fg$ sur les fonctions continues par morceaux : si $\varphi(y, y) = 0$ avec $y \neq 0$, la fonction affine $t \mapsto \varphi(x, x) + 2t\varphi(x, y)$ est positive sur $\mathbb{R}$, donc constante, et $\varphi(x, y) = 0$. C'est la difficulté que l'exercice 15 de la feuille du chapitre 18 contournait en introduisant un paramètre λ.
- Pour x et y non nuls, l'inégalité dit que $\frac{\langle x, y\rangle}{\|x\| \, \|y\|}$ appartient à $[-1\,,1]$: il existe un unique $\theta \in [0\,, \pi]$ dont c'est le cosinus. On l'appelle l'**écart angulaire** de x et y ; dans le plan, c'est l'angle géométrique du lycée, et la formule $\langle x, y\rangle = \|x\| \, \|y\| \cos\theta$ devient une définition de l'angle dans tout espace préhilbertien.

Proposition 4 : *(Deux cas particuliers.)*

(i) Pour tous réels $x_1, \ldots, x_n$ et $y_1, \ldots, y_n$,

$$\Big(\sum_{k=1}^{n} x_k y_k\Big)^2 \leqslant \Big(\sum_{k=1}^{n} x_k^2\Big)\Big(\sum_{k=1}^{n} y_k^2\Big),$$

avec égalité si et seulement si les n-uplets (x_k) et (y_k) sont proportionnels.

(ii) Pour toutes fonctions f, g continues sur $[a\,,b]$, avec $a < b$,

$$\Big(\int_a^b fg\Big)^2 \leqslant \Big(\int_a^b f^2\Big)\Big(\int_a^b g^2\Big),$$

avec égalité si et seulement si f et g sont proportionnelles.

∘ Démonstration : On applique le théorème 1, élevé au carré, aux produits scalaires canoniques de $\mathbb{R}^n$ (proposition 1) et de $\mathcal{C}([a\,,b],\mathbb{R})$ (proposition 2, avec $w=1$). □

Exemples :

- Soit $x_1,\dots,x_n$ des réels strictement positifs. Appliquée à $\sqrt{x_k}$ et $\frac{1}{\sqrt{x_k}}$, la proposition 4 (i) donne
$$n^2=\Big(\sum_{k=1}^n \sqrt{x_k}\cdot\frac{1}{\sqrt{x_k}}\Big)^2\leqslant\Big(\sum_{k=1}^n x_k\Big)\Big(\sum_{k=1}^n\frac{1}{x_k}\Big),$$
avec égalité si et seulement si $\sqrt{x_k}=\frac{\lambda}{\sqrt{x_k}}$ pour tout k, c'est-à-dire si les x_k sont tous égaux. C'est l'exercice 21 de la feuille du chapitre 3, démontré alors par une somme double : la moyenne arithmétique des x_k est au moins égale à leur moyenne harmonique. Pour $x_k=k$ et $n=4$, le produit vaut $10\times\frac{25}{12}=\frac{125}{6}\approx 20{,}8$, au-dessus de 16.
- Avec $y_k=1$, on obtient $\sum_{k=1}^n k\leqslant\sqrt{n\sum_{k=1}^n k^2}$. Le rapport des deux membres vaut 0,886 pour $n=10$ et 0,868 pour $n=100$; il tend vers $\frac{\sqrt{3}}{2}$, ce qui signifie que l'écart angulaire entre $(1,2,\dots,n)$ et $(1,1,\dots,1)$ tend vers $\frac{\pi}{6}$: l'inégalité est loin d'être une égalité.
- Sur $[0\,,1]$, avec $f(t)=t$ et $g(t)=\mathrm{e}^t$, une intégration par parties donne $\int_0^1 t\,\mathrm{e}^t\,\mathrm{d}t=1$, et la proposition 4 (ii) donne la majoration $\sqrt{\frac{1}{3}\cdot\frac{\mathrm{e}^2-1}{2}}\approx 1{,}032$. Elle est très fine : sur $[0\,,1]$, les fonctions t et e^t sont « presque proportionnelles », leur écart angulaire vaut environ 14°.
- Pour f continue et strictement positive sur $[a\,,b]$, la proposition 4 (ii) appliquée à $\sqrt{f}$ et $\frac{1}{\sqrt{f}}$ donne $\int_a^b f\int_a^b\frac{1}{f}\geqslant(b-a)^2$: c'est la version intégrale du premier exemple. Pour $f=\exp$ sur $[0\,,1]$, le produit vaut $(\mathrm{e}-1)(1-\mathrm{e}^{-1})=\mathrm{e}+\mathrm{e}^{-1}-2\approx 1{,}086$.

L'exercice 58 de la feuille du chapitre 5, avec l'inégalité de Hölder, et l'exercice 15 de celle du chapitre 18 retrouvaient ces deux inégalités pour l'exposant $p=2$; le théorème 1 en donne la raison commune : ce sont deux instances d'une même inégalité, valable dans tout espace préhilbertien.

19.3.3 Inégalité triangulaire

Théorème 2 : Inégalité triangulaire :
Pour tous x, $y\in E$,
$$\|x+y\|\leqslant\|x\|+\|y\|,$$
avec égalité si et seulement si $x=0$ ou s'il existe un réel $\lambda\geqslant 0$ tel que $y=\lambda x$. En conséquence, pour tous x, y, $z\in E$,
$$\big|\|x\|-\|y\|\big|\leqslant\|x-y\|\qquad\text{et}\qquad d(x,z)\leqslant d(x,y)+d(y,z).$$

∘ Démonstration :

▷ *L'inégalité.* Par la proposition 3 (ii), puis par l'inégalité de Cauchy-Schwarz,
$$\|x+y\|^2=\|x\|^2+\|y\|^2+2\langle x,y\rangle\leqslant\|x\|^2+\|y\|^2+2\|x\|\,\|y\|=\big(\|x\|+\|y\|\big)^2,$$
et l'on conclut par croissance de la racine carrée sur $\mathbb{R}_+$.

▷ *Le cas d'égalité.* Il y a égalité si et seulement si $\langle x,y\rangle=\|x\|\,\|y\|$. Si $x=0$, c'est le cas. Si $y=\lambda x$ avec $\lambda\geqslant 0$, alors $\langle x,y\rangle=\lambda\|x\|^2=\|x\|\,\|\lambda x\|$. Réciproquement, supposons $\langle x,y\rangle=\|x\|\,\|y\|$ et $x\neq 0$: c'est un cas d'égalité de Cauchy-Schwarz, donc (x,y) est liée, et, comme $x\neq 0$, $y=\lambda x$ pour un réel λ. Alors $\lambda\|x\|^2=|\lambda|\,\|x\|^2$, d'où $\lambda=|\lambda|\geqslant 0$.

▷ *Les conséquences.* On applique l'inégalité à $x=(x-y)+y$, ce qui donne $\|x\|-\|y\|\leqslant\|x-y\|$, puis on échange x et y. Enfin, $z-x=(z-y)+(y-x)$. □

Remarques :

- L'erreur la plus répandue est de confondre les deux cas d'égalité. Dans Cauchy-Schwarz, $|\langle x,y\rangle|=\|x\|\,\|y\|$ pour tous vecteurs colinéaires ; dans l'inégalité triangulaire, il faut de plus qu'ils soient de même sens : pour $y=-x\neq 0$, on a $\|x+y\|=0<2\|x\|$.
- Pour la norme intégrale, le théorème 2 est l'inégalité de Minkowski de l'exercice 15 de la feuille du chapitre 18, dans le cas $p=2$; sa démonstration demandait l'inégalité de Hölder, et elle se réduit ici à trois lignes.

19.4 Orthogonalité et algorithme de Gram-Schmidt

Dans cette section, E est un espace préhilbertien réel.

19.4.1 Vecteurs orthogonaux, orthogonal d'une partie

Définition :

— Deux vecteurs x et y sont **orthogonaux**, ce que l'on note $x \perp y$, si $\langle x, y\rangle = 0$.

— L'**orthogonal** d'une partie A de E est l'ensemble des vecteurs orthogonaux à tous les éléments de A :

$$A^\perp = \{y \in E \mid \forall x \in A,\ \langle x, y\rangle = 0\}.$$

— Deux sous-espaces F et G sont **orthogonaux** si tout vecteur de F est orthogonal à tout vecteur de G, c'est-à-dire si $G \subset F^\perp$.

Proposition 5 : Soit A et B deux parties de E et F un sous-espace de E.

(i) $A^\perp$ est un sous-espace vectoriel de E, même si A n'en est pas un.

(ii) Si $A \subset B$, alors $B^\perp \subset A^\perp$; de plus, $A^\perp = (\operatorname{Vect} A)^\perp$ et $A \subset (A^\perp)^\perp$.

(iii) $\{0\}^\perp = E$ et $E^\perp = \{0\}$.

(iv) $F \cap F^\perp = \{0\}$: la somme $F + F^\perp$ est directe.

∘ Démonstration :

▷ *(i)* Pour chaque $x \in A$, l'application $y \mapsto \langle x, y\rangle$ est une forme linéaire ; $A^\perp$ est l'intersection de leurs noyaux, qui sont des sous-espaces (chapitre 14, propositions 17 et 4). Si A est vide, $A^\perp = E$.

▷ *(ii)* Un vecteur orthogonal à tous les éléments de B l'est à ceux de A. Comme $A \subset \operatorname{Vect} A$, on en déduit $(\operatorname{Vect} A)^\perp \subset A^\perp$. Réciproquement, si $y \in A^\perp$, la linéarité de $x \mapsto \langle x, y\rangle$ montre que y est orthogonal à toute combinaison linéaire d'éléments de A. Enfin, tout $x \in A$ est orthogonal à tout $y \in A^\perp$, donc appartient à $(A^\perp)^\perp$.

▷ *(iii) et (iv)* Tout vecteur est orthogonal à 0. Si $y \in E^\perp$, ou si $y \in F \cap F^\perp$, alors y est orthogonal à lui-même : $\langle y, y\rangle = 0$, et $y = 0$ par le caractère défini. La somme est alors directe par la proposition 12 du chapitre 14. □

Exemples :

- Dans $\mathbb{R}^3$ muni du produit canonique, le plan d'équation $x + y + z = 0$ est l'orthogonal du vecteur $(1, 1, 1)$. Plus généralement, dans $\mathbb{R}^n$, l'hyperplan d'équation $a_1x_1 + \cdots + a_nx_n = 0$ est l'orthogonal du vecteur $a = (a_1, \ldots, a_n)$: l'équation d'un hyperplan, au chapitre 15, se lit comme une condition d'orthogonalité.
- Dans $\mathcal{M}_n(\mathbb{R})$ muni de $\langle A, B\rangle = \operatorname{tr}(A^TB)$, les sous-espaces $\mathcal{S}_n(\mathbb{R})$ et $\mathcal{A}_n(\mathbb{R})$ sont orthogonaux. Pour S symétrique et A antisymétrique, la trace d'une matrice est celle de sa transposée, et $\operatorname{tr}(AS) = \operatorname{tr}(SA)$ (chapitre 16, proposition 14), donc
$$\langle S, A\rangle = \operatorname{tr}(SA) = \operatorname{tr}\left((SA)^T\right) = \operatorname{tr}(A^TS^T) = -\operatorname{tr}(AS) = -\operatorname{tr}(SA),$$
et ce réel est nul. On verra (proposition 11) que $\mathcal{A}_n(\mathbb{R})$ est exactement l'orthogonal de $\mathcal{S}_n(\mathbb{R})$.
- Dans $\mathcal{C}([-1\,,1], \mathbb{R})$ muni de $\langle f, g\rangle = \int_{-1}^{1} fg$, toute fonction paire est orthogonale à toute fonction impaire : leur produit est impair, et son intégrale sur $[-1\,,1]$ est nulle par le changement de variable $t \mapsto -t$.

19.4.2 Familles orthogonales et théorème de Pythagore

Définition : Une famille $(x_i)_{i\in I}$ de vecteurs de E est **orthogonale** si ses vecteurs sont deux à deux orthogonaux : $\langle x_i, x_j\rangle = 0$ pour $i \neq j$. Elle est **orthonormée** si de plus ses vecteurs sont unitaires, c'est-à-dire si $\langle x_i, x_j\rangle = \delta_{ij}$ pour tous i, j, où δ_{ij} vaut 1 si $i = j$ et 0 sinon.

Proposition 6 : Une famille orthogonale de vecteurs non nuls est libre. En particulier, une famille orthonormée est libre.

∘ Démonstration : Soit $(x_1,\dots,x_n)$ une sous-famille finie et $\lambda_1,\dots,\lambda_n$ des réels tels que $\sum_k \lambda_k x_k = 0$. Pour chaque j, on prend le produit scalaire avec x_j : tous les termes $\lambda_k\langle x_k, x_j\rangle$ avec $k \neq j$ sont nuls, et il reste $\lambda_j\|x_j\|^2 = 0$. Comme $x_j \neq 0$, on obtient $\lambda_j = 0$. □

Théorème 3 : Théorème de Pythagore :

(i) Deux vecteurs x et y sont orthogonaux si et seulement si $\|x+y\|^2 = \|x\|^2 + \|y\|^2$.

(ii) Si $(x_1,\dots,x_n)$ est une famille orthogonale, alors

$$\Big\|\sum_{k=1}^{n} x_k\Big\|^2 = \sum_{k=1}^{n}\|x_k\|^2.$$

∘ Démonstration : Le point (i) est la proposition 3 (ii). Pour (ii), on développe par bilinéarité :

$$\Big\|\sum_{k=1}^{n} x_k\Big\|^2 = \sum_{j=1}^{n}\sum_{k=1}^{n}\langle x_j, x_k\rangle = \sum_{k=1}^{n}\langle x_k, x_k\rangle,$$

puisque les termes d'indices $j \neq k$ sont nuls. □

Remarque : L'erreur la plus répandue est de croire que la réciproque de (ii) vaut pour plus de deux vecteurs. Dans $\mathbb{R}^2$, les vecteurs $(1,0)$, $(0,1)$ et $(1,-1)$ ont pour somme $(2,0)$, et $\|(2,0)\|^2 = 4 = 1+1+2$; pourtant $\langle(1,0),(1,-1)\rangle = 1 \neq 0$. Les doubles produits non nuls, ici $0+2-2$, se compensent sans être nuls.

Exemple : Dans $\mathcal{C}([0\,,2\pi],\mathbb{R})$ muni de $\langle f,g\rangle = \int_0^{2\pi} fg$, la famille formée de la fonction constante 1 et des fonctions $t\mapsto\cos(kt)$ et $t\mapsto\sin(kt)$, pour $k \geqslant 1$, est orthogonale. Les formules de linéarisation du chapitre 4, par exemple $\cos(kt)\cos(\ell t) = \frac{1}{2}\big(\cos((k-\ell)t)+\cos((k+\ell)t)\big)$, ramènent chaque produit à une somme de fonctions $\cos(mt)$ ou $\sin(mt)$ avec m entier, dont l'intégrale sur $[0\,,2\pi]$ est nulle si $m \neq 0$. On trouve de même $\|1\|^2 = 2\pi$ et $\|\cos(k\,\cdot)\|^2 = \|\sin(k\,\cdot)\|^2 = \pi$ pour $k \geqslant 1$. Le théorème de Pythagore calcule alors sans intégration

$$\int_0^{2\pi}\big(1+2\cos t - 3\sin(2t)\big)^2\,\mathrm{d}t = 2\pi + 4\pi + 9\pi = 15\pi.$$

L'exercice 52 de la feuille du chapitre 6 en donnait la version complexe, avec les fonctions $\theta\mapsto \mathrm{e}^{\mathrm{i}k\theta}$: sa question 1 en est l'orthogonalité, et sa question 3 le théorème de Pythagore, où le conjugué joue le rôle annoncé à la remarque qui suit la proposition 1.

19.4.3 L'algorithme de Gram-Schmidt

Une famille libre engendre un sous-espace ; on peut la remplacer par une famille orthonormée qui engendre, étape par étape, les mêmes sous-espaces. L'idée est de retirer à chaque nouveau vecteur sa composante le long des précédents.

Théorème 4 : Orthonormalisation de Gram-Schmidt :
Soit $(e_1,\dots,e_n)$ une famille libre de E. Il existe une unique famille orthonormée $(u_1,\dots,u_n)$ telle que, pour tout $k \in [\![1\,,n]\!]$,

$$\mathrm{Vect}(u_1,\dots,u_k) = \mathrm{Vect}(e_1,\dots,e_k) \qquad \text{et} \qquad \langle u_k, e_k\rangle > 0.$$

Elle est donnée par $u_1 = \dfrac{e_1}{\|e_1\|}$ et, pour $k \geqslant 2$,

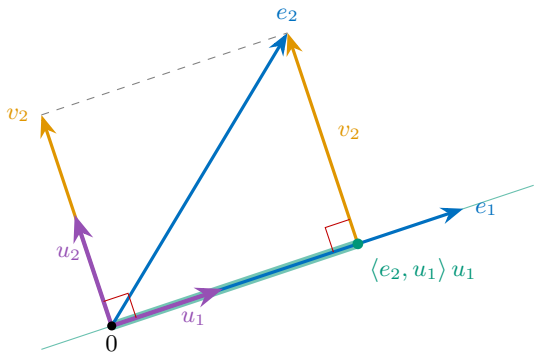
$$v_k = e_k - \sum_{j=1}^{k-1}\langle e_k, u_j\rangle\, u_j \qquad \text{et} \qquad u_k = \frac{v_k}{\|v_k\|}.$$

∘ Démonstration : On note $F_k = \mathrm{Vect}(e_1, \dots, e_k)$, et l'on pose $v_1 = e_1$.

▷ *Existence, par récurrence sur k.* Comme $e_1 \neq 0$, u_1 est unitaire, $\mathrm{Vect}(u_1) = F_1$ et $\langle u_1, e_1\rangle = \|e_1\| > 0$. Supposons construits $u_1, \dots, u_{k-1}$, orthonormés, avec $\mathrm{Vect}(u_1, \dots, u_{k-1}) = F_{k-1}$, et définissons v_k comme dans l'énoncé.

- Le vecteur v_k n'est pas nul : sinon $e_k = \sum_{j<k}\langle e_k, u_j\rangle u_j$ appartiendrait à F_{k-1}, et la famille $(e_1, \dots, e_k)$ serait liée. Le vecteur u_k est donc bien défini, et unitaire.
- Pour $i < k$, l'orthonormalité de $(u_1, \dots, u_{k-1})$ donne $\langle v_k, u_i\rangle = \langle e_k, u_i\rangle - \langle e_k, u_i\rangle = 0$: la famille $(u_1, \dots, u_k)$ est orthonormée.
- Le vecteur v_k appartient à $e_k + F_{k-1} \subset F_k$, donc u_k aussi, et $\mathrm{Vect}(u_1, \dots, u_k) \subset F_k$; réciproquement, $e_k = \|v_k\|\, u_k + \sum_{j<k}\langle e_k, u_j\rangle u_j$ appartient à $\mathrm{Vect}(u_1, \dots, u_k)$, qui contient aussi F_{k-1}.
- Enfin, u_k est orthogonal aux u_j pour $j < k$, donc $\langle u_k, e_k\rangle = \langle u_k, \|v_k\|\, u_k\rangle = \|v_k\| > 0$.

▷ *Unicité.* Soit $(u'_1, \dots, u'_n)$ une famille ayant les mêmes propriétés ; montrons par récurrence que $u'_k = u_k$. Supposons-le pour les indices $j < k$. Le vecteur u'_k appartient à $F_k = \mathrm{Vect}(u_1, \dots, u_k)$: $u'_k = \sum_{j\leqslant k}\lambda_j u_j$. Pour $j < k$, $\lambda_j = \langle u'_k, u_j\rangle = \langle u'_k, u'_j\rangle = 0$, donc $u'_k = \lambda_k u_k$; comme u'_k est unitaire, $\lambda_k = \pm 1$, et la condition $\langle u'_k, e_k\rangle = \lambda_k\langle u_k, e_k\rangle > 0$ impose $\lambda_k = 1$. □



Le vecteur e_2, en bleu, se décompose en sa composante $\langle e_2, u_1\rangle\, u_1$ le long de la droite $\mathrm{Vect}(e_1)$, en vert, et en un vecteur v_2, en orange, orthogonal à cette droite ; le vecteur v_2, ramené à l'origine puis normalisé, donne u_2. La base orthonormée obtenue, en violet, engendre la même droite que e_1 au premier pas et le même plan que (e_1, e_2) au second.

♡ **Méthode : Pour orthonormaliser une famille libre $(e_1, \dots, e_n)$:**

- ▶ On calcule les vecteurs orthogonaux $w_1 = e_1$ et, pour $k \geqslant 2$, $w_k = e_k - \displaystyle\sum_{j=1}^{k-1} \frac{\langle e_k, w_j\rangle}{\|w_j\|^2}\, w_j$, qui sont les v_k du théorème 4, sans racine carrée.
- ▶ On vérifie à chaque étape que w_k est orthogonal aux vecteurs précédents : c'est le contrôle qui détecte presque toutes les erreurs de calcul.
- ▶ On divise enfin chaque w_k par sa norme.

Si seule l'orthogonalité est demandée, on s'arrête avant la dernière étape ; on peut aussi remplacer un w_k par un multiple non nul, pour éviter les fractions.

Exemple : Dans $\mathbb{R}^3$ muni du produit canonique, orthonormalisons $e_1 = (1,1,0)$, $e_2 = (1,0,1)$ et $e_3 = (0,1,1)$. On a $u_1 = \frac{1}{\sqrt{2}}(1,1,0)$. Puis $\langle e_2, u_1\rangle = \frac{1}{\sqrt{2}}$, et

$$v_2 = (1,0,1) - \frac{1}{2}(1,1,0) = \Big(\frac{1}{2}, -\frac{1}{2}, 1\Big), \qquad \|v_2\|^2 = \frac{3}{2}, \qquad u_2 = \frac{1}{\sqrt{6}}(1,-1,2).$$

Enfin $\langle e_3, u_1\rangle = \frac{1}{\sqrt{2}}$ et $\langle e_3, u_2\rangle = \frac{1}{\sqrt{6}}$, d'où

$$v_3 = (0,1,1) - \frac{1}{2}(1,1,0) - \frac{1}{6}(1,-1,2) = \Big(-\frac{2}{3}, \frac{2}{3}, \frac{2}{3}\Big), \qquad u_3 = \frac{1}{\sqrt{3}}(-1,1,1).$$

On contrôle que $\langle u_3, e_1\rangle = \langle u_3, e_2\rangle = 0$: le vecteur u_3 est orthogonal au plan $\mathrm{Vect}(e_1, e_2)$, ce qui servira à la section 19.6.

19.4.4 Polynômes orthogonaux

Munissons $\mathbb{R}_3[X]$ du produit scalaire de la proposition 2 sur $[-1\,,1]$, avec le poids constant,

$$\langle P, Q\rangle = \int_{-1}^{1} P(t)Q(t)\,\mathrm{d}t,$$

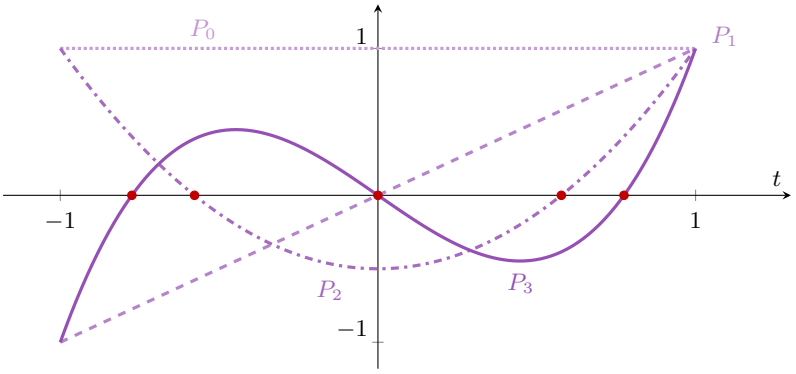
et appliquons la méthode, sans normaliser, à sa base canonique $(1, X, X^2, X^3)$. Le produit d'un polynôme pair et d'un polynôme impair est d'intégrale nulle, et $\langle X^i, X^j\rangle = \int_{-1}^{1} t^{i+j}\,\mathrm{d}t$ vaut $\frac{2}{i+j+1}$ si $i+j$ est pair. On trouve successivement $w_0 = 1$, $w_1 = X$, puis

$$w_2 = X^2 - \frac{\langle X^2, 1\rangle}{\|1\|^2} = X^2 - \frac{2/3}{2} = X^2 - \frac{1}{3}, \qquad w_3 = X^3 - \frac{\langle X^3, X\rangle}{\|X\|^2}\,X = X^3 - \frac{2/5}{2/3}\,X = X^3 - \frac{3}{5}X,$$

les autres produits scalaires étant nuls par parité. Leurs normes au carré valent 2, $\frac{2}{3}$, $\frac{8}{45}$ et $\frac{8}{175}$. Les multiples de ces polynômes qui prennent la valeur 1 en 1 sont les **polynômes de Legendre**

$$P_0 = 1, \qquad P_1 = X, \qquad P_2 = \frac{3X^2-1}{2}, \qquad P_3 = \frac{5X^3-3X}{2},$$

de normes au carré $\|P_n\|^2 = \frac{2}{2n+1}$. Legendre les avait rencontrés, dans un mémoire sur l'attraction des sphéroïdes publié en 1785, bien avant que leur orthogonalité ne soit lue comme une propriété géométrique.



Les quatre polynômes valent 1 en $t = 1$, et P_n a exactement n racines, marquées en rouge, toutes simples et situées dans $]-1\,,1[$: celles de P_2 sont $\pm\frac{1}{\sqrt{3}} \approx \pm 0{,}577$, celles de P_3 sont 0 et $\pm\sqrt{\frac{3}{5}} \approx \pm 0{,}775$. Ce n'est pas une coïncidence, comme le montre la proposition suivante.

Proposition 7 : Soit $a < b$, w un poids sur $[a\,,b]$ et $\langle P, Q\rangle = \int_a^b PQw$ le produit scalaire de la proposition 2 sur $\mathbb{R}[X]$. Soit $(Q_0, \dots, Q_n)$ la famille orthogonale obtenue par la méthode à partir de $(1, X, \dots, X^n)$. Pour tout $k \in [\![0\,, n]\!]$:

(i) $\deg Q_k = k$, et Q_k est orthogonal à $\mathbb{R}_{k-1}[X]$ si $k \geqslant 1$;

(ii) Q_k possède k racines réelles distinctes, toutes simples et situées dans $]a\,,b[$.

∘ Démonstration :

▷ *(i)* Par le théorème 4, $\mathrm{Vect}(Q_0, \dots, Q_k) = \mathrm{Vect}(1, \dots, X^k) = \mathbb{R}_k[X]$: Q_k est de degré au plus k. S'il était de degré au plus $k-1$, il appartiendrait à $\mathrm{Vect}(Q_0, \dots, Q_{k-1})$, et la famille orthogonale $(Q_0, \dots, Q_k)$ de vecteurs non nuls serait liée, contrairement à la proposition 6. Enfin, Q_k est orthogonal à $Q_0, \dots, Q_{k-1}$, donc à $\mathbb{R}_{k-1}[X]$ qu'ils engendrent (proposition 5 (ii)).

▷ *(ii)* Soit $z_1 < \dots < z_s$ les racines de Q_k dans $]a\,,b[$, de multiplicités $m_1, \dots, m_s$. Par la proposition 10 du chapitre 13, $Q_k = (X - z_1)^{m_1} \cdots (X - z_s)^{m_s}\,T$, où T n'a pas de racine dans $]a\,,b[$, car une telle racine serait un z_i de multiplicité supérieure à m_i ; par le théorème des valeurs intermédiaires (chapitre 8, théorème 8), la fonction T garde un signe constant sur $]a\,,b[$. Soit R le produit des $X - z_i$ pour les i tels que m_i est impair, avec $R = 1$ s'il n'y en a pas. Dans Q_kR, chaque facteur $X - z_i$ figure avec un exposant pair, donc Q_kR garde le signe de T sur $]a\,,b[$. La fonction Q_kRw est continue, de signe constant et non identiquement nulle, puisqu'elle ne s'annule qu'en un nombre fini de points : par le théorème 4 du chapitre 18, $\langle Q_k, R\rangle \neq 0$. Par (i), cela impose $\deg R \geqslant k$. Or $\deg R \leqslant s \leqslant k$: donc $s = k$, et chaque m_i est impair ; comme $m_1 + \dots + m_k \leqslant k$, tous les m_i valent 1. □

Remarques :

- Sur $\mathbb{R}[X]$ tout entier, l'algorithme appliqué à $(X^k)_{k\in\mathbb{N}}$ fournit une suite (Q_k), puisque chaque Q_k ne dépend que des précédents ; par la proposition 10 du chapitre 14, c'est une base de $\mathbb{R}[X]$, formée de polynômes deux à deux orthogonaux.
- L'erreur la plus répandue est de croire que les polynômes orthogonaux ne dépendent que de l'espace $\mathbb{R}[X]$: ils dépendent du produit scalaire. Pour le poids $w(t) = 1 - t^2$, on trouve encore $w_0 = 1$ et $w_1 = X$ par parité, mais $w_2 = X^2 - \frac{1}{5}$, puisque $\frac{\langle X^2, 1\rangle}{\|1\|^2} = \frac{4/15}{4/3} = \frac{1}{5}$.

19.5 Bases orthonormées

Dans cette section, E est un espace euclidien de dimension $n \geqslant 1$.

19.5.1 Existence

Définition : Une **base orthonormée** de E est une base de E qui est une famille orthonormée.

Par la proposition 6 et le théorème 4 du chapitre 15, une famille orthonormée de n vecteurs de E est libre, donc c'est une base orthonormée : il n'y a pas à vérifier qu'elle engendre E.

Théorème 5 : Existence de bases orthonormées :

(i) Tout espace euclidien possède une base orthonormée.

(ii) *Théorème de la base orthonormée incomplète :* toute famille orthonormée $(e_1, \dots, e_p)$ de E peut être complétée en une base orthonormée $(e_1, \dots, e_n)$ de E.

∘ Démonstration :

▷ *(ii)* La famille $(e_1, \dots, e_p)$ est libre (proposition 6). Par le théorème de la base incomplète (chapitre 15, théorème 2), on la complète en une base $(e_1, \dots, e_p, f_{p+1}, \dots, f_n)$ de E, à laquelle on applique l'algorithme de Gram-Schmidt. Ses p premiers vecteurs ne sont pas modifiés : si $u_j = e_j$ pour $j < k \leqslant p$, alors $v_k = e_k - \sum_{j<k}\langle e_k, e_j\rangle e_j = e_k$, qui est unitaire, donc $u_k = e_k$. On obtient une famille orthonormée de n vecteurs qui commence par $e_1, \dots, e_p$: c'est une base orthonormée.

▷ *(i)* On applique l'algorithme de Gram-Schmidt à une base de E, qui existe par le théorème 2 du chapitre 15, ou bien (ii) avec $p = 0$. □

Exemples :

- La base canonique de $\mathbb{R}^n$ est orthonormée pour le produit canonique ; celle de $\mathcal{M}_{n,p}(\mathbb{R})$, formée des matrices élémentaires E_{ij}, l'est pour $\langle A, B\rangle = \mathrm{tr}(A^T B)$.
- La base canonique de $\mathbb{R}^2$ n'est pas orthonormée pour le produit $X^T S Y$ de la section 19.2, avec $S = \begin{pmatrix} 2 & 1 \\ 1 & 1 \end{pmatrix}$: être orthonormée dépend du produit scalaire, et non de la base seule.

19.5.2 Calculs dans une base orthonormée

Proposition 8 : Soit $(e_1, \dots, e_n)$ une base orthonormée de E, et $x, y \in E$, de coordonnées $X = (x_k)$ et $Y = (y_k)$ dans cette base. Alors

$$x = \sum_{k=1}^{n} \langle x, e_k\rangle\, e_k, \qquad \langle x, y\rangle = \sum_{k=1}^{n} x_k y_k = X^T Y \qquad \text{et} \qquad \|x\|^2 = \sum_{k=1}^{n} x_k^2.$$

∘ Démonstration : On écrit $x = \sum_k x_k e_k$ et l'on prend le produit scalaire avec e_j : par orthonormalité, $\langle x, e_j\rangle = \sum_k x_k \langle e_k, e_j\rangle = x_j$. Par bilinéarité,

$$\langle x, y\rangle = \sum_{j=1}^{n}\sum_{k=1}^{n} x_j y_k \,\langle e_j, e_k\rangle = \sum_{k=1}^{n} x_k y_k,$$

et la norme s'en déduit avec $y = x$. □

Dans une base orthonormée, tout produit scalaire se calcule donc comme le produit canonique de $\mathbb{R}^n$: un espace euclidien de dimension n « ressemble » à $\mathbb{R}^n$, l'application qui associe à un vecteur ses coordonnées conservant le produit scalaire. Les coordonnées s'obtiennent par des produits scalaires, sans résoudre de système.

Remarque : L'erreur la plus répandue est d'appliquer la formule $\langle x, y\rangle = \sum x_k y_k$ dans une base qui n'est pas orthonormée. Dans $\mathbb{R}^2$ muni du produit canonique, le vecteur $(1,1)$ a pour coordonnées $(0,1)$ dans la base $\big((1,0),(1,1)\big)$, et sa norme au carré vaut 2, et non $0^2+1^2=1$.

♡ **Méthode : Pour décomposer un vecteur x dans une base orthonormée $(e_1,\dots,e_n)$:**

- ▶ On calcule les n produits scalaires $\langle x, e_k\rangle$, qui sont les coordonnées cherchées.
- ▶ On contrôle le résultat par la formule $\|x\|^2 = \sum\langle x, e_k\rangle^2$, qui ne demande pas de reconstituer x.

Exemples :

- Avec la base orthonormée (u_1,u_2,u_3) obtenue à la section 19.4, le vecteur $x=(1,2,3)$ a pour coordonnées

$$\langle x,u_1\rangle = \frac{3}{\sqrt{2}}, \qquad \langle x,u_2\rangle = \frac{1-2+6}{\sqrt{6}} = \frac{5}{\sqrt{6}}, \qquad \langle x,u_3\rangle = \frac{-1+2+3}{\sqrt{3}} = \frac{4}{\sqrt{3}},$$

  et l'on vérifie que $\frac{9}{2}+\frac{25}{6}+\frac{16}{3} = 14 = 1^2+2^2+3^2$. Dans la base non orthonormée (e_1,e_2,e_3), il aurait fallu résoudre un système de trois équations.

- Soit x_0, x_1, x_2 trois réels distincts, et $\mathbb{R}_2[X]$ muni du produit $\langle P,Q\rangle = \sum_{k=0}^{2} P(x_k)Q(x_k)$ de la section 19.2. Les polynômes de Lagrange L_0, L_1, L_2 du chapitre 13 (théorème 13), caractérisés par $L_j(x_k) = \delta_{jk}$, vérifient $\langle L_i, L_j\rangle = \sum_k \delta_{ik}\delta_{jk} = \delta_{ij}$: ils forment une base orthonormée. La proposition 8 donne les coordonnées $\langle P, L_j\rangle = P(x_j)$, c'est-à-dire la formule d'interpolation $P = \sum_j P(x_j)L_j$. Pour $(x_0,x_1,x_2) = (-1,0,2)$, on trouve $L_0 = \frac{X^2-2X}{3}$, $L_1 = \frac{-X^2+X+2}{2}$ et $L_2 = \frac{X^2+X}{6}$.

19.5.3 Matrice de Gram et déterminant

Proposition 9 : Soit e une base orthonormée de E, $(x_1,\dots,x_p)$ une famille de vecteurs de E, et $M \in \mathcal{M}_{n,p}(\mathbb{R})$ la matrice de ses coordonnées dans e. La **matrice de Gram** $G = \big(\langle x_i, x_j\rangle\big)_{1\leqslant i,j\leqslant p}$ vérifie :

(i) $G = M^T M$;

(ii) si $p = n$, $\det G = \det_e(x_1,\dots,x_n)^2$; ainsi $\det G \geqslant 0$, avec égalité si et seulement si la famille est liée ;

(iii) si e' est une autre base orthonormée de E, $\big|\det_{e'}(x_1,\dots,x_n)\big| = \big|\det_e(x_1,\dots,x_n)\big|$, et $\det_e(e') = \pm 1$.

∘ Démonstration :

▷ *(i)* Le coefficient d'indice (i,j) de M^TM est le produit de la ligne i de M^T, c'est-à-dire de la colonne i de M, par la colonne j de M : c'est $\sum_k m_{ki}m_{kj}$, qui vaut $\langle x_i, x_j\rangle$ par la proposition 8.

▷ *(ii)* Par le théorème 6 du chapitre 17, et par définition du déterminant d'une famille dans une base,

$$\det(M^TM) = \det(M^T)\det(M) = (\det M)^2 \qquad \text{et} \qquad \det M = \det_e(x_1,\dots,x_n).$$

Ce déterminant est nul si et seulement si la famille n'est pas une base (chapitre 17, théorème 4), c'est-à-dire, pour n vecteurs, si elle est liée.

▷ *(iii)* La matrice G ne dépend que des produits scalaires, et non de la base orthonormée choisie : par (ii), $\det_{e'}(x_1,\dots,x_n)^2 = \det G = \det_e(x_1,\dots,x_n)^2$. Pour la famille e', la matrice de Gram est I_n, donc $\det_e(e')^2 = 1$. □

Le chapitre 17 interprétait $|\det(u,v,w)|$, calculé dans la base canonique de $\mathbb{R}^3$, comme le volume du parallélépipède construit sur u, v et w. Le point (iii) dit que ce volume ne dépend pas de la base orthonormée dans laquelle on calcule : c'est une notion euclidienne, que l'on peut écrire $\sqrt{\det G}$ sans choisir de base. Pour la famille $e_1 = (1,1,0)$, $e_2 = (1,0,1)$, $e_3 = (0,1,1)$ de la section 19.4,

$$G = \begin{pmatrix} 2 & 1 & 1 \\ 1 & 2 & 1 \\ 1 & 1 & 2 \end{pmatrix}, \qquad \det G = 4 = (-2)^2,$$

où -2 est le déterminant de (e_1,e_2,e_3) dans la base canonique : le parallélépipède a pour volume 2.

Remarques :

- Une matrice $A \in \mathcal{M}_n(\mathbb{R})$ a des colonnes qui forment une base orthonormée de $\mathbb{R}^n$ si et seulement si sa matrice de Gram est l'identité, c'est-à-dire si $A^TA = I_n$. De telles matrices, dites **orthogonales**, sont exactement les matrices de passage entre deux bases orthonormées ; par (iii), leur déterminant vaut ± 1. Celles de déterminant 1 représentent les rotations, annoncées au chapitre 17 et rencontrées dans l'exercice 51 de sa feuille. Leur étude, avec celle des isométries, relève de la seconde année ; la matrice de la base (u_1,u_2,u_3) de la section 19.4, par exemple, est orthogonale, de déterminant -1.
- Pour une base quelconque $(f_1,\dots,f_n)$, de matrice de Gram G, le calcul de la proposition 8 donne $\langle x,y\rangle = X^TGY$: on retrouve la forme X^TSY de la section 19.2, avec $S = G$, et la formule $\langle x,y\rangle = X^TY$ caractérise les bases orthonormées.

19.6 Projection orthogonale sur un sous-espace de dimension finie

Dans cette section, E est un espace préhilbertien réel, de dimension finie ou non, et F un sous-espace de E *de dimension finie*. Le chapitre 14 a défini le projecteur sur F parallèlement à un supplémentaire G, et remarqué qu'il dépend du choix de G ; il annonçait que la projection « à angle droit » du lycée demande un produit scalaire. Le produit scalaire désigne en effet un supplémentaire privilégié.

19.6.1 Supplémentaire orthogonal

Théorème 6 : Supplémentaire orthogonal :
Soit F un sous-espace de dimension finie de E. Alors

$$E = F \oplus F^\perp.$$

Plus précisément, si $(e_1,\dots,e_p)$ est une base orthonormée de F, la décomposition de $x \in E$ est

$$x = \underbrace{\sum_{k=1}^{p} \langle x,e_k\rangle\, e_k}_{\in F} + \underbrace{\left(x - \sum_{k=1}^{p} \langle x,e_k\rangle\, e_k\right)}_{\in F^\perp}.$$

∘ Démonstration : Si $F = \{0\}$, alors $F^\perp = E$ et il n'y a rien à démontrer. Sinon, F, muni de la restriction du produit scalaire, est un espace euclidien : il possède une base orthonormée $(e_1,\dots,e_p)$ (théorème 5). La somme $F + F^\perp$ est directe (proposition 5 (iv)). Soit $x \in E$ et $y = \sum_k \langle x,e_k\rangle e_k$, qui appartient à F. Pour tout j, l'orthonormalité donne

$$\langle x-y, e_j\rangle = \langle x,e_j\rangle - \sum_{k=1}^{p} \langle x,e_k\rangle\langle e_k,e_j\rangle = \langle x,e_j\rangle - \langle x,e_j\rangle = 0.$$

Ainsi $x-y$ est orthogonal à $e_1,\dots,e_p$, donc à $F = \mathrm{Vect}(e_1,\dots,e_p)$ (proposition 5 (ii)), et $x = y + (x-y)$ appartient à $F + F^\perp$. □

Définition : Soit F un sous-espace de dimension finie de E. Le sous-espace $F^\perp$ est le **supplémentaire orthogonal** de F. La **projection orthogonale** sur F est le projecteur p_F sur F parallèlement à $F^\perp$; le vecteur $p_F(x)$ est le **projeté orthogonal** de x sur F.

Proposition 10 : Soit F un sous-espace de dimension finie de E et $x \in E$.

(i) Pour toute base orthonormée $(e_1, \dots, e_p)$ de F, $p_F(x) = \sum_{k=1}^{p} \langle x, e_k \rangle\, e_k$. En particulier, pour $u \neq 0$, le projeté sur la droite $\mathrm{Vect}(u)$ est $\dfrac{\langle x, u\rangle}{\|u\|^2}\, u$.

(ii) Le projeté $p_F(x)$ est l'unique vecteur $y \in F$ tel que $x - y \in F^\perp$; si $(f_1, \dots, f_p)$ engendre F, cette condition s'écrit $\langle x - y, f_i\rangle = 0$ pour tout i.

(iii) $(F^\perp)^\perp = F$, et $\mathrm{id}_E - p_F$ est le projecteur sur $F^\perp$ parallèlement à F.

(iv) *Inégalité de Bessel :* $\|p_F(x)\|^2 = \sum_{k=1}^{p} \langle x, e_k\rangle^2 \leqslant \|x\|^2$, avec égalité si et seulement si $x \in F$.

∘ Démonstration :

▷ *(i) et (ii)* Le point (i) est le théorème 6, et (ii) est l'unicité de la décomposition $x = y + (x - y)$ dans $F \oplus F^\perp$, avec la proposition 5 (ii) pour se ramener à une famille génératrice. Pour une droite, on applique (i) à la base orthonormée $\left(\frac{u}{\|u\|}\right)$.

▷ *(iii)* On sait que $F \subset (F^\perp)^\perp$ (proposition 5 (ii)). Soit $x \in (F^\perp)^\perp$, décomposé en $x = y + z$, avec $y \in F$ et $z \in F^\perp$. Alors $0 = \langle x, z\rangle = \langle y, z\rangle + \|z\|^2 = \|z\|^2$, donc $z = 0$ et $x \in F$. Ainsi $E = F^\perp \oplus (F^\perp)^\perp$, et la proposition 21 du chapitre 14 dit que $\mathrm{id}_E - p_F$ est le projecteur sur $F^\perp$ parallèlement à F.

▷ *(iv)* Par le théorème de Pythagore, $\|x\|^2 = \|p_F(x)\|^2 + \|x - p_F(x)\|^2$, et la proposition 8 calcule $\|p_F(x)\|^2$ dans la base (e_k). L'égalité a lieu si et seulement si $x - p_F(x) = 0$, c'est-à-dire $x \in F$ (chapitre 14, proposition 21 (ii)). □

Remarques :

- L'erreur la plus répandue est d'appliquer la formule (i) avec une base de F qui n'est qu'orthogonale, ou quelconque. Pour une base orthogonale $(f_1, \dots, f_p)$, la formule correcte est $p_F(x) = \sum_k \frac{\langle x, f_k\rangle}{\|f_k\|^2} f_k$, qui se déduit de (i) en normalisant ; pour une base quelconque, on revient à (ii).
- Les vecteurs w_k de la méthode de Gram-Schmidt s'interprètent maintenant : w_k est la composante de e_k orthogonale au sous-espace F_{k-1} engendré par les vecteurs précédents, c'est-à-dire $w_k = e_k - p_{F_{k-1}}(e_k)$, comme le montrait la figure de la section 19.4.

19.6.2 En dimension finie

Proposition 11 : On suppose E euclidien, de dimension n. Pour tout sous-espace F de E,

$$\dim F^\perp = n - \dim F \qquad \text{et} \qquad (F^\perp)^\perp = F.$$

Si $(e_1, \dots, e_p)$ est une base orthonormée de F, complétée en une base orthonormée $(e_1, \dots, e_n)$ de E, alors $F^\perp = \mathrm{Vect}(e_{p+1}, \dots, e_n)$.

∘ Démonstration : Le théorème 6 donne $E = F \oplus F^\perp$, et la proposition 3 du chapitre 15 donne alors $\dim F + \dim F^\perp = n$. L'égalité $(F^\perp)^\perp = F$ est la proposition 10 (iii). Les vecteurs $e_{p+1}, \dots, e_n$ sont orthogonaux à $e_1, \dots, e_p$, donc à F : $\mathrm{Vect}(e_{p+1}, \dots, e_n)$ est un sous-espace de $F^\perp$, de dimension $n - p = \dim F^\perp$, donc égal à $F^\perp$ (chapitre 15, théorème 5). □

Exemple : Dans $\mathcal{M}_n(\mathbb{R})$, on a vu que $\mathcal{A}_n(\mathbb{R}) \subset \mathcal{S}_n(\mathbb{R})^\perp$. Par la proposition 7 du chapitre 16, $\dim \mathcal{A}_n(\mathbb{R}) = \frac{n(n-1)}{2} = n^2 - \frac{n(n+1)}{2} = \dim \mathcal{S}_n(\mathbb{R})^\perp$: ainsi $\mathcal{S}_n(\mathbb{R})^\perp = \mathcal{A}_n(\mathbb{R})$, et la décomposition $\mathcal{M}_n(\mathbb{R}) = \mathcal{S}_n(\mathbb{R}) \oplus \mathcal{A}_n(\mathbb{R})$ du chapitre 16 est une décomposition orthogonale. Le projeté orthogonal de M sur $\mathcal{S}_n(\mathbb{R})$ est sa partie symétrique $\frac{1}{2}(M + M^T)$, puisque $M - \frac{1}{2}(M + M^T) = \frac{1}{2}(M - M^T)$ est antisymétrique.

Remarque : En dimension infinie, l'hypothèse « F de dimension finie » ne peut pas être supprimée. Munissons $\mathbb{R}[X]$ du produit scalaire $\langle P, Q\rangle = \sum_k a_k b_k$, où (a_k) et (b_k) sont les suites des coefficients de P et Q ; la somme est finie, et le produit est défini comme celui de $\mathbb{R}^n$. Soit $H = \{P \in \mathbb{R}[X] \mid P(1) = 0\}$, hyperplan de $\mathbb{R}[X]$ (noyau de la forme linéaire non nulle $P \mapsto P(1)$). Si $Q = \sum b_k X^k$ appartient à $H^\perp$, il est orthogonal à $X^k - X^{k+1} \in H$, d'où $b_k = b_{k+1}$ pour tout k ; la suite (b_k), constante et nulle à partir d'un certain rang, est nulle. Ainsi $H^\perp = \{0\}$: la somme $H \oplus H^\perp = H$ n'est pas $\mathbb{R}[X]$, il n'existe pas de projection orthogonale sur H, et $(H^\perp)^\perp = \mathbb{R}[X] \neq H$.

Proposition 12 : *(Hyperplans et vecteurs normaux.)* On suppose E euclidien, de dimension $n \geqslant 1$.

(i) Pour tout vecteur $u \neq 0$, $\mathrm{Vect}(u)^\perp = \{x \in E \mid \langle u, x\rangle = 0\}$ est un hyperplan de E.

(ii) Réciproquement, pour tout hyperplan H, il existe $u \neq 0$ tel que $H = \mathrm{Vect}(u)^\perp$; un tel vecteur est dit **normal** à H, et il est unique à un facteur non nul près.

(iii) Si, dans une base orthonormée, H a pour équation $a_1 x_1 + \cdots + a_n x_n = 0$, le vecteur de coordonnées $(a_1, \ldots, a_n)$ est normal à H.

∘ Démonstration :

▷ *(i)* L'application $x \mapsto \langle u, x\rangle$ est une forme linéaire, non nulle puisque $\langle u, u\rangle > 0$; son noyau est un hyperplan.

▷ *(ii)* Par la proposition 11, $\dim H^\perp = 1$: $H^\perp = \mathrm{Vect}(u)$ avec $u \neq 0$, et $H = (H^\perp)^\perp = \mathrm{Vect}(u)^\perp$. Si $\mathrm{Vect}(u)^\perp = \mathrm{Vect}(u')^\perp$, on prend l'orthogonal des deux membres : $\mathrm{Vect}(u) = \mathrm{Vect}(u')$, toujours par la proposition 11.

▷ *(iii)* Par la proposition 8, $a_1 x_1 + \cdots + a_n x_n = \langle u, x\rangle$, où u a pour coordonnées $(a_1, \ldots, a_n)$. □

Le chapitre 15 (théorème 12) disait qu'un hyperplan a une équation, unique à un facteur près ; la proposition 12 en donne le sens géométrique : les coefficients de l'équation, lue dans une base orthonormée, sont les coordonnées d'un vecteur normal.

19.6.3 Distance à un sous-espace

Définition : La **distance** d'un vecteur $x \in E$ à une partie non vide A de E est la borne inférieure $d(x, A) = \inf_{a \in A} \|x - a\|$, qui existe puisque l'ensemble de ces réels est non vide et minoré par 0 (chapitre 7).

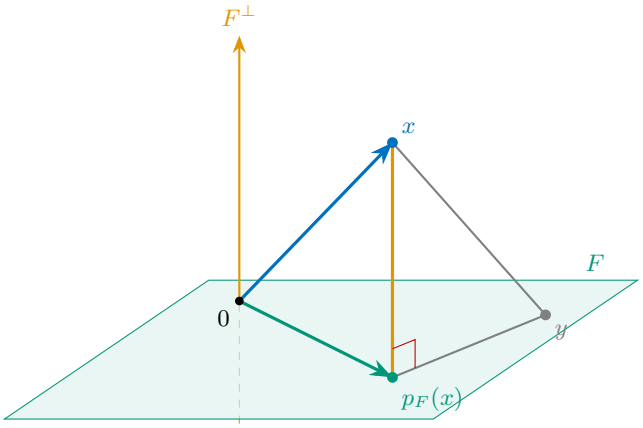
Théorème 7 : Meilleure approximation :
Soit F un sous-espace de dimension finie de E et $x \in E$. Pour tout $y \in F$,

$$\|x - y\|^2 = \|x - p_F(x)\|^2 + \|p_F(x) - y\|^2.$$

Par conséquent, la distance de x à F est atteinte en un unique point de F, le projeté orthogonal $p_F(x)$, et

$$d(x, F) = \|x - p_F(x)\| \qquad \text{et} \qquad d(x, F)^2 = \|x\|^2 - \|p_F(x)\|^2.$$

∘ Démonstration : On écrit $x-y=\big(x-p_F(x)\big)+\big(p_F(x)-y\big)$: le premier vecteur appartient à $F^\perp$, le second à F, car $p_F(x)$ et y sont dans F. Ils sont orthogonaux, et le théorème de Pythagore donne l'égalité. On en déduit $\|x-y\|\geqslant\|x-p_F(x)\|$ pour tout $y\in F$, avec égalité si et seulement si $\|p_F(x)-y\|=0$, c'est-à-dire $y=p_F(x)$. La borne inférieure est donc un minimum, atteint seulement en $p_F(x)$. La dernière formule est la proposition 10 (iv), écrite sous la forme $\|x-p_F(x)\|^2=\|x\|^2-\|p_F(x)\|^2$. □



Le vecteur x, en bleu, se décompose en son projeté $p_F(x)$ sur le plan F, en vert, et en la composante $x-p_F(x)$, en orange, parallèle à la droite $F^\perp$. Pour un autre point y de F, le triangle de sommets x, $p_F(x)$ et y est rectangle en $p_F(x)$: son hypoténuse $\|x-y\|\approx 2{,}98$ est plus longue que le côté $\|x-p_F(x)\|=2{,}3$, ce qu'exprime le théorème 7.

Proposition 13 : *(Distance à un hyperplan.)* On suppose E euclidien. Soit $u\neq 0$ et $H=\mathrm{Vect}(u)^\perp$. Pour tout $x\in E$,

$$p_H(x)=x-\frac{\langle x,u\rangle}{\|u\|^2}\,u \qquad \text{et} \qquad d(x,H)=\frac{|\langle x,u\rangle|}{\|u\|}.$$

∘ Démonstration : Par la proposition 11, $H^\perp=\mathrm{Vect}(u)$. La proposition 10 (iii) donne $p_H=\mathrm{id}_E-p_{\mathrm{Vect}(u)}$, et la proposition 10 (i) donne le projeté sur la droite $\mathrm{Vect}(u)$, d'où la formule de p_H. Par le théorème 7,

$$d(x,H)=\|x-p_H(x)\|=\left\|\frac{\langle x,u\rangle}{\|u\|^2}\,u\right\|=\frac{|\langle x,u\rangle|}{\|u\|^2}\,\|u\|=\frac{|\langle x,u\rangle|}{\|u\|}.$$

□

Exemples :

- Dans $\mathbb{R}^3$ muni du produit canonique, le plan H d'équation $x+y+z=0$ a pour vecteur normal $u=(1,1,1)$. Pour $x=(1,2,3)$, on a $\langle x,u\rangle=6$ et $\|u\|^2=3$, donc

$$p_H(x)=(1,2,3)-2\,(1,1,1)=(-1,0,1) \qquad \text{et} \qquad d(x,H)=\frac{6}{\sqrt{3}}=2\sqrt{3}.$$

  On contrôle que $(-1,0,1)$ vérifie l'équation de H.

- Soit $F=\mathrm{Vect}(e_1,e_2)$, avec $e_1=(1,1,0)$ et $e_2=(1,0,1)$, et $x=(1,2,3)$. La section 19.4 a fourni la base orthonormée (u_1,u_2) de F, et la section 19.5 les produits scalaires $\frac{3}{\sqrt{2}}$ et $\frac{5}{\sqrt{6}}$; donc

$$p_F(x)=\frac{3}{2}(1,1,0)+\frac{5}{6}(1,-1,2)=\left(\frac{7}{3},\frac{2}{3},\frac{5}{3}\right), \qquad x-p_F(x)=\frac{4}{3}(-1,1,1).$$

  Le résidu est colinéaire à u_3, vecteur normal à F, et $d(x,F)=\frac{4}{\sqrt{3}}$, ce que la proposition 13 donne directement : $\frac{|\langle x,u_3\rangle|}{\|u_3\|}=\frac{4}{\sqrt{3}}$, la troisième coordonnée de x dans la base (u_1,u_2,u_3).

♡ **Méthode : Pour calculer le projeté orthogonal de x sur F :**

- Si l'on connaît une base orthonormée, ou orthogonale, de F, on applique la formule de la proposition 10 (i) ; sinon, on peut en construire une par l'algorithme de Gram-Schmidt.
- Si l'on ne connaît qu'une base $(f_1, \dots, f_p)$ de F, on écrit $p_F(x) = \sum_j \lambda_j f_j$, et l'on résout le système des p **équations normales** $\langle x - \sum_j \lambda_j f_j, f_i\rangle = 0$, de matrice la matrice de Gram de $(f_1, \dots, f_p)$, inversible par la proposition 9.
- Si F est un hyperplan d'un espace euclidien, on cherche un vecteur normal u et l'on applique la proposition 13 ; si $\dim F^\perp$ est petite, on peut aussi projeter sur $F^\perp$, puis utiliser $p_F = \mathrm{id}_E - p_{F^\perp}$.

19.6.4 Deux problèmes de moindres carrés

Le théorème 7 transforme un problème de minimisation en un calcul de projeté. On revient d'abord au problème de Legendre, dans l'espace $\mathbb{R}^m$ muni du produit canonique.

Proposition 14 : *(Droite des moindres carrés.)* Soit $m \geqslant 2$ et $(x_1, y_1), \dots, (x_m, y_m)$ des points du plan, dont les abscisses ne sont pas toutes égales. On note $\bar{x} = \frac{1}{m}\sum x_i$ et $\bar{y} = \frac{1}{m}\sum y_i$. La quantité

$$S(a,b) = \sum_{i=1}^{m} (y_i - ax_i - b)^2$$

atteint son minimum sur $\mathbb{R}^2$ en un unique couple (a,b), donné par

$$a = \frac{\displaystyle\sum_{i=1}^{m}(x_i - \bar{x})(y_i - \bar{y})}{\displaystyle\sum_{i=1}^{m}(x_i - \bar{x})^2} \qquad \text{et} \qquad b = \bar{y} - a\bar{x}.$$

○ Démonstration :

▷ *Interprétation.* On considère les vecteurs $X = (x_i)$, $Y = (y_i)$ et $U = (1, \dots, 1)$ de $\mathbb{R}^m$, de sorte que

$$S(a,b) = \|Y - aX - bU\|^2.$$

Les abscisses n'étant pas toutes égales, X n'est pas colinéaire à U, et la famille (X,U) est libre : l'application $(a,b) \mapsto aX + bU$ est une bijection de $\mathbb{R}^2$ sur le plan $F = \mathrm{Vect}(X,U)$. Minimiser S, c'est minimiser $\|Y - Z\|$ pour $Z \in F$: par le théorème 7, le minimum est atteint en un seul point, $Z = p_F(Y)$, donc pour un seul couple (a,b).

▷ *Calcul.* Par la proposition 10 (ii), ce couple est caractérisé par les deux équations normales

$$\langle Y - aX - bU, U\rangle = 0 \qquad \text{et} \qquad \langle Y - aX - bU, X\rangle = 0.$$

Comme $\langle X, U\rangle = m\bar{x}$ et $\langle Y, U\rangle = m\bar{y}$, la première s'écrit $m\bar{y} - am\bar{x} - mb = 0$, ou encore $b = \bar{y} - a\bar{x}$. Le vecteur $R = Y - aX - bU = (Y - \bar{y}U) - a(X - \bar{x}U)$ est alors orthogonal à U et à X, donc à $X - \bar{x}U$:

$$\langle Y - \bar{y}U, X - \bar{x}U\rangle - a\|X - \bar{x}U\|^2 = 0,$$

et $\|X - \bar{x}U\|^2 = \sum(x_i - \bar{x})^2$ n'est pas nul, puisque les x_i ne sont pas tous égaux. On obtient la valeur de a. □

Exemple : Pour les six points de la figure de l'introduction, d'abscisses $1, \dots, 6$ et d'ordonnées 1,5 ; 2,4 ; 2,6 ; 3,9 ; 4 ; 5, on a $\sum x_i = 21$, $\sum x_i^2 = 91$, $\sum y_i = 19{,}4$ et $\sum x_i y_i = 79{,}7$. Les équations normales forment le système

$$\begin{cases} 91a + 21b &= 79{,}7 \\ 21a + 6b &= 19{,}4 \end{cases} \qquad \text{de solution} \qquad a = \frac{118}{175} \approx 0{,}674, \quad b = \frac{131}{150} \approx 0{,}873.$$

Le minimum vaut $S(a,b) = \frac{779}{2\,625} \approx 0{,}297$. La seconde équation dit que la droite passe par le point moyen $(\bar{x}, \bar{y}) = (3{,}5\,;\, 3{,}233)$, et que la somme des écarts $y_i - ax_i - b$ est nulle : les écarts positifs et négatifs se compensent, c'est l'« équilibre » dont parle Legendre. Le coefficient a s'interprète comme l'augmentation moyenne de y quand x augmente d'une unité.

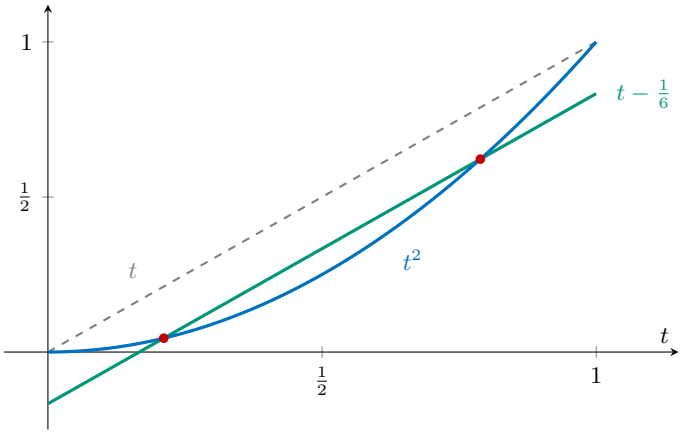
Remarque : L'erreur la plus répandue est de croire que la droite des moindres carrés minimise la somme des carrés des distances des points à la droite. Elle minimise les écarts *verticaux*, mesurés parallèlement à l'axe des ordonnées : les rôles de x et de y ne sont pas symétriques. En ajustant x en fonction de y, on trouverait ici la droite $x \approx 1{,}430\,y - 1{,}123$, soit $y \approx 0{,}699\,x + 0{,}785$, qui n'est pas la même.

Le même principe s'applique aux fonctions, avec le produit scalaire intégral. Cherchons la fonction affine $t \mapsto at + b$ la plus proche de $t \mapsto t^2$ au sens de la norme de $\mathcal{C}([0\,,1], \mathbb{R})$ associée à $\langle f, g\rangle = \int_0^1 fg$, c'est-à-dire le couple (a, b) qui rend minimale l'intégrale

$$\int_0^1 (t^2 - at - b)^2 \,\mathrm{d}t = \|t^2 - (at + b)\|^2.$$

Il s'agit donc de calculer la distance de t^2 au sous-espace F des fonctions affines, qui est de dimension 2, et le point où elle est atteinte. Une base orthogonale de F s'obtient par l'algorithme de Gram-Schmidt : $w_0 = 1$ et $w_1 = t - \frac{\langle t,1\rangle}{\|1\|^2} = t - \frac{1}{2}$, avec $\|1\|^2 = 1$ et $\|w_1\|^2 = \int_0^1 \left(t - \frac{1}{2}\right)^2 \mathrm{d}t = \frac{1}{12}$. Comme $\langle t^2, 1\rangle = \frac{1}{3}$ et $\langle t^2, w_1\rangle = \frac{1}{4} - \frac{1}{6} = \frac{1}{12}$, la proposition 10 (i), sous la forme adaptée aux bases orthogonales, donne

$$p_F(t^2) = \frac{1}{3} + \frac{1/12}{1/12}\left(t - \frac{1}{2}\right) = t - \frac{1}{6}, \qquad d(t^2, F)^2 = \|t^2\|^2 - \|p_F(t^2)\|^2 = \frac{1}{5} - \frac{7}{36} = \frac{1}{180}.$$



La droite verte $t \mapsto t - \frac{1}{6}$ coupe la parabole en deux points, marqués en rouge, et répartit les écarts de part et d'autre de la courbe, alors que la sécante grise $t \mapsto t$, qui interpole t^2 en 0 et en 1, reste toujours au-dessus. Le carré de la distance vaut $\frac{1}{180} \approx 0{,}0056$ pour la droite verte, $\frac{1}{30}$ pour la sécante et $\frac{1}{80}$ pour la tangente en $\frac{1}{2}$, d'équation $t - \frac{1}{4}$: l'erreur quadratique moyenne $\frac{1}{\sqrt{180}} \approx 0{,}075$ est plus de deux fois plus petite que celle de la sécante, $\frac{1}{\sqrt{30}} \approx 0{,}183$.

Remarques :

- La meilleure approximation dépend de la norme choisie. La droite verte s'écarte de t^2 de $\frac{1}{6}$ en 0 et en 1 ; la droite $t \mapsto t - \frac{1}{8}$ ne s'en écarte jamais de plus de $\frac{1}{8}$, mais son intégrale $\int_0^1 \left(t^2 - t + \frac{1}{8}\right)^2 \mathrm{d}t = \frac{7}{960}$ est plus grande que $\frac{1}{180}$. Le produit scalaire mesure un écart moyen, et non un écart maximal.
- En remplaçant F par $\mathbb{R}_n[X]$ et $[0\,,1]$ par $[-1\,,1]$, la base orthogonale à utiliser est celle des polynômes de Legendre de la section 19.4 : le projeté de f est $\sum_{k=0}^{n} \frac{\langle f, P_k\rangle}{\|P_k\|^2} P_k$, et augmenter n ajoute des termes sans modifier les précédents. C'est l'avantage décisif d'une base orthogonale sur la base canonique, pour laquelle il faudrait résoudre un nouveau système à chaque étape.

19.7 Bilan

Le chapitre a transporté dans un espace vectoriel réel quelconque la géométrie des longueurs et des angles droits, à partir de la seule donnée d'un produit scalaire.

Notion	Énoncé	Référence
produits scalaires	X^TY, $\mathrm{tr}(A^TB)$, $\int_a^b fgw$; défini grâce au chapitre 18	propositions 1 et 2
norme	$\|x+y\|^2 = \|x\|^2 + \|y\|^2 + 2\langle x,y\rangle$, polarisation	proposition 3
Cauchy-Schwarz	$\|\langle x,y\rangle\| \leqslant \|x\|\,\|y\|$, égalité si (x,y) liée	théorème 1, proposition 4
inégalité triangulaire	$\|x+y\| \leqslant \|x\| + \|y\|$, égalité si même sens	théorème 2
orthogonalité	$A^\perp$ sous-espace, famille orthogonale libre, Pythagore	propositions 5 et 6, théorème 3
Gram-Schmidt	$v_k = e_k - \sum_{j<k}\langle e_k,u_j\rangle u_j$, polynômes orthogonaux	théorème 4, proposition 7
bases orthonormées	existence, $x = \sum\langle x,e_k\rangle e_k$, $\langle x,y\rangle = X^TY$, $G = M^TM$	théorème 5, propositions 8 et 9
projection orthogonale	$E = F \oplus F^\perp$, $p_F(x) = \sum\langle x,e_k\rangle e_k$, $\dim F^\perp = n - \dim F$	théorème 6, propositions 10 et 11
distance	$d(x,F) = \|x - p_F(x)\|$, $d(x,H) = \frac{\|\langle x,u\rangle\|}{\|u\|}$, moindres carrés	théorème 7, propositions 12 à 14

Trois idées sont à retenir. Le produit scalaire est caractérisé par des propriétés, la bilinéarité, la symétrie et la positivité définie, et non par une formule : c'est ce qui permet de traiter des matrices et des fonctions comme des vecteurs du plan. L'inégalité de Cauchy-Schwarz, conséquence de la seule positivité, contient à la fois les inégalités sur les sommes et celles sur les intégrales. Enfin, une base orthonormée rend tous les calculs immédiats, coordonnées, normes et projections, et l'algorithme de Gram-Schmidt en fournit toujours une en dimension finie ; la projection orthogonale résout alors tout problème de moindres carrés. Le chapitre 20 étudiera les séries numériques : appliquée à une famille orthonormée infinie, l'inégalité de Bessel montrera que la série $\sum\langle x,e_k\rangle^2$ converge, ce qui ouvre la voie aux séries de Fourier, dont les coefficients sont les coordonnées d'une fonction sur la famille trigonométrique de la section 19.4. Le chapitre 21 retrouvera l'inégalité de Cauchy-Schwarz en probabilités, et la seconde année étudiera les transformations qui conservent le produit scalaire.