Espaces préhilbertiens réels

Dans ce chapitre, on travaille seulement avec le corps de base $\mathbb{R}$.

1 Produit scalaire et norme

Définition (Produit scalaire, espace préhilbertien réel, espace euclidien)

- Soit E un $\mathbb{R}$-espace vectoriel. On appelle *produit scalaire sur E* toute forme bilinéaire symétrique définie positive, i.e. toute application $\langle \cdot,\cdot \rangle : E\times E \longrightarrow \mathbb{R}$:
  — *bilinéaire* : $\forall x,y,z\in E,\quad \forall \lambda,\mu\in\mathbb{R},\quad \langle \lambda x+\mu y, z\rangle = \lambda\langle x,z\rangle + \mu\langle y,z\rangle$ et $\langle x, \lambda y+\mu z\rangle = \lambda\langle x,y\rangle + \mu\langle x,z\rangle$,
  — *symétrique* : $\forall x,y\in E,\quad \langle y,x\rangle = \langle x,y\rangle$,
  — *définie* : $\forall x\in E,\quad \langle x,x\rangle = 0 \implies x = 0_E$ (propriété de *séparation*),
  — *positive* : $\forall x\in E,\quad \langle x,x\rangle \geqslant 0$.

  Le produit scalaire $\langle x,y\rangle$ est aussi parfois noté : $(x|y)$ ou $x\cdot y$.
- Un espace vectoriel réel muni d'un produit scalaire est appelé un *espace préhilbertien réel*. Un espace préhilbertien réel DE DIMENSION FINIE est appelé un *espace euclidien*.

Explication Définition déroutante ! Nous n'avons à ce stade encore jamais parlé en algèbre linéaire d'angles et de normes. Mine de rien, nous sommes donc en train de définir le concept de produit scalaire indépendamment de toute relation du type : $\overrightarrow{u}\cdot\overrightarrow{v} = \left\|\overrightarrow{u}\right\|.\left\|\overrightarrow{v}\right\|\cos\left(\overrightarrow{u},\overrightarrow{v}\right)$. En réalité, dans la théorie que nous nous apprêtons à développer, le produit scalaire est premier et ce sont les notions de norme et d'angle qui viennent après.

En pratique Pour montrer la bilinéarité d'un produit scalaire potentiel, la linéarité par rapport à une variable seulement est suffisante si on a pris la peine de démontrer la symétrie avant.

Petite remarque au passage : $\langle x, 0_E\rangle = \langle 0_E, x\rangle = 0$ pour tout $x\in E$, par bilinéarité du produit scalaire.

Définition-théorème (Produit scalaire canonique sur $\mathbb{R}^n$) L'application $(X,Y)\longmapsto {}^{\mathrm{t}}XY = \sum_{k=1}^n x_k y_k$ est un produit scalaire sur $\mathbb{R}^n$ appelé son *produit scalaire canonique*.

Explication Ouf, nous retrouvons bien ici les produits scalaires auxquels nous sommes habitués dans le plan $\mathbb{R}^2$ et l'espace $\mathbb{R}^3$. Par exemple, pour tous vecteurs $\overrightarrow{u} = (x,y)$ et $\overrightarrow{u}' = (x',y')$ de $\mathbb{R}^2$: $\overrightarrow{u}\cdot\overrightarrow{u}' = xx'+yy'$.

Démonstration

- **Symétrie :** Pour tous $X,Y\in\mathbb{R}^n$: ${}^{\mathrm{t}}XY = \sum_{k=1}^n x_k y_k = \sum_{k=1}^n y_k x_k = {}^{\mathrm{t}}YX$.
- **Bilinéarité :** Par symétrie, la linéarité par rapport à la deuxième variable suffit. Pour tous $X,Y,Z\in\mathbb{R}^n$ et $\lambda,\mu\in\mathbb{R}$, par bilinéarité du produit matriciel : ${}^{\mathrm{t}}X(\lambda Y+\mu Z) = \lambda({}^{\mathrm{t}}XY)+\mu({}^{\mathrm{t}}XZ)$.
- **Positivité et séparation :** Pour tout $X\in\mathbb{R}^n$: ${}^{\mathrm{t}}XX = \sum_{k=1}^n x_k^2 \geqslant 0$, et si : ${}^{\mathrm{t}}XX = \sum_{k=1}^n \overbrace{x_k^2}^{\geqslant 0} = 0$, alors : $x_k = 0$ pour tout $k\in[\![1,n]\!]$, i.e. : $X = 0$.

Exemple Il peut exister beaucoup produits scalaires sur un même espace vectoriel. L'application $(X,Y) \longmapsto {}^{t}X \begin{pmatrix} 2 & 1 \\ 1 & 2 \end{pmatrix} Y$ est par exemple un produit scalaire sur $\mathbb{R}^2$ distinct du produit scalaire usuel.

Démonstration Symétrie et bilinéarité évidentes. Ensuite, pour tout $X = (x,y) \in \mathbb{R}^2$:

$$ {}^{t}X \begin{pmatrix} 2 & 1 \\ 1 & 2 \end{pmatrix} X = \begin{pmatrix} x & y \end{pmatrix} \begin{pmatrix} 2x+y \\ x+2y \end{pmatrix} = 2x^2 + 2xy + 2y^2 = x^2 + y^2 + (x+y)^2 \geqslant 0, $$

et si : ${}^{t}X \begin{pmatrix} 2 & 1 \\ 1 & 2 \end{pmatrix} X = 0$, alors : $x = y = x+y = 0$, donc : $X = (0,0)$.

Exemple Soient $a, b \in \mathbb{R}$ avec : $a < b$. L'application $(f,g) \longmapsto \int_a^b f(t)g(t)\,\mathrm{d}t$ est un produit scalaire sur $\mathscr{C}\big([a,b],\mathbb{R}\big)$.

Démonstration

- **Symétrie et bilinéarité :** Évidentes.
- **Positivité et séparation :** Pour tout $f \in \mathscr{C}\big([a,b],\mathbb{R}\big)$: $\int_a^b f(t)^2\,\mathrm{d}t \geqslant 0$ et si : $\int_a^b f(t)^2\,\mathrm{d}t = 0$, alors comme f est CONTINUE et POSITIVE OU NULLE : $f^2 = 0$ sur $[a,b]$ et donc : $f = 0$.

✖ ATTENTION ! ✖ Muni du produit scalaire défini ci-dessus, $\mathscr{C}\big([a,b],\mathbb{R}\big)$ n'est pas un espace euclidien car ce n'est pas un $\mathbb{R}$-espace vectoriel de dimension finie. C'est seulement un espace préhilbertien réel.

Exemple Soient $x_0, \ldots, x_n \in \mathbb{R}$ DISTINCTS. L'application $(P,Q) \longmapsto \sum_{k=0}^{n} P(x_k)Q(x_k)$ est un produit scalaire sur $\mathbb{R}_n[X]$.

Démonstration

- **Symétrie et bilinéarité :** Symétrie évidente, donc la linéarité par rapport à la première variable suffit. Pour tous $P,Q,R \in \mathbb{R}_n[X]$ et $\lambda, \mu \in \mathbb{R}$: $\sum_{k=0}^{n} (\lambda P + \mu Q)(x_k) R(x_k) = \lambda \sum_{k=0}^{n} P(x_k)R(x_k) + \mu \sum_{k=0}^{n} Q(x_k)R(x_k)$.
- **Positivité et séparation :** Pour tout $P \in \mathbb{R}_n[X]$: $\sum_{k=0}^{n} P(x_k)^2 \geqslant 0$, et si : $\langle P,P \rangle = \sum_{k=0}^{n} P(x_k)^2 = 0$, alors : $P(x_k) = 0$ pour tout $k \in [\![0,n]\!]$, autrement dit $x_0, \ldots, x_n$ sont des racines de P. Le polynôme P possède alors au moins $n+1$ racines distinctes, or : $\deg P \leqslant n$, donc : $P = 0$.

Exemple L'application $(A,B) \longmapsto \mathrm{tr}\big({}^{t}AB\big)$ est un produit scalaire sur $\mathscr{M}_n(\mathbb{R})$ — concrètement : $\mathrm{tr}\big({}^{t}AB\big) = \sum_{1 \leqslant i,j \leqslant n} a_{ij} b_{ij}$. Si on n'oublie pas que : $\mathscr{M}_n(\mathbb{R}) = \mathbb{R}^{n^2}$, ce produit scalaire n'est jamais que le produit scalaire canonique de $\mathbb{R}^{n^2}$.

Démonstration

- **Symétrie et bilinéarité :** Symétrie évidente, donc la linéarité par rapport à la première variable suffit. Pour tous $A, B, C \in \mathscr{M}_n(\mathbb{R})$ et $\lambda, \mu \in \mathbb{R}$, par bilinéarité du produit matriciel et linéarité de la trace :
$$ \mathrm{tr}\Big({}^{t}A\big(\lambda B + \mu C\big)\Big) = \mathrm{tr}\Big(\lambda\big({}^{t}AB\big) + \mu\big({}^{t}AC\big)\Big) = \lambda\,\mathrm{tr}\big({}^{t}AB\big) + \mu\,\mathrm{tr}\big({}^{t}AC\big). $$
- **Positivité et séparation :** Pour tout $A \in \mathscr{M}_n(\mathbb{R})$: $\mathrm{tr}\big({}^{t}AA\big) = \sum_{1 \leqslant i,j \leqslant n} a_{ij}^2 \geqslant 0$, et si : $\mathrm{tr}\big({}^{t}AA\big) = 0$, alors : $a_{ij} = 0$ pour tous $i, j \in [\![1,n]\!]$, i.e. : $A = 0$.

Définition (Norme et distance associées à un produit scalaire) Soit E un espace préhilbertien réel.

- On appelle *norme (euclidienne) sur E associée au produit scalaire* $\langle \cdot,\cdot \rangle$ l'application $\|\cdot\| : E \longrightarrow \mathbb{R}_+$ définie pour tout $x \in E$ par : $\|x\| = \sqrt{\langle x,x \rangle}$. On dit qu'un vecteur x de E est *unitaire* si : $\|x\| = 1$.
- On appelle *distance (euclidienne) sur E associée au produit scalaire* $\langle \cdot,\cdot \rangle$ l'application $d : E \times E \longrightarrow \mathbb{R}_+$ définie pour tous $x, y \in E$ par : $d(x,y) = \|x - y\|$.

✖ ATTENTION ! ✖ La notion de distance n'est pas forcément celle qu'on croit ! La distance dépend d'un CHOIX de produit scalaire. Par exemple, pour le produit scalaire $(X,Y) \longmapsto {}^{t}X \begin{pmatrix} 2 & 1 \\ 1 & 2 \end{pmatrix} Y$ sur $\mathbb{R}^2$: $\big\|(1,0)\big\| = \sqrt{2} \neq 1$.

Explication Qui dit « bilinéarité » dit « identités remarquables ». En l'occurrence, pour tous $x, y \in E$:

$$\boxed{\|x+y\|^2 = \|x\|^2 + 2\langle x, y\rangle + \|y\|^2} \qquad \text{et} \qquad \boxed{\langle x+y, x-y\rangle = \|x\|^2 - \|y\|^2.}$$

On peut aussi « inverser » ces relations et récupérer le produit scalaire en fonction de la norme. On obtient alors ce qu'on appelle des *identités de polarisation*. Par exemple : $\boxed{\langle x, y\rangle = \frac{1}{2}\left(\|x+y\|^2 - \|x\|^2 - \|y\|^2\right) = \frac{1}{4}\left(\|x+y\|^2 - \|x-y\|^2\right).}$

Théorème (Inégalité de Cauchy-Schwarz, inégalité triangulaire) Soient E un espace préhilbertien réel et $x, y, z \in E$.

(i) **Inégalité de Cauchy-Schwarz :** $\left|\langle x, y\rangle\right| \leqslant \|x\|.\|y\|$, avec égalité si et seulement si x et y sont colinéaires.

(ii) **Inégalité triangulaire, version norme :** $\Big|\|x\| - \|y\|\Big| \leqslant \|x+y\| \leqslant \|x\| + \|y\|$.

L'inégalité de droite est une égalité si et seulement si x et y sont colinéaires DE MÊME SENS.

Inégalité triangulaire, version distance : $\left|d(x,y) - d(y,z)\right| \leqslant d(x,z) \leqslant d(x,y) + d(y,z)$.

Explication Si nous avions une définition propre du produit scalaire en termes de normes et d'angles, l'inégalité de Cauchy-Schwarz serait une pure trivialité : $|\vec{u}\cdot\vec{v}| = \left|\|\vec{u}\|.\|\vec{v}\|\cos(\vec{u}, \vec{v})\right| \leqslant \|\vec{u}\|.\|\vec{v}\|$. Dans notre contexte, cette inégalité est justement remarquable parce que nous n'avons pas encore de définition propre des angles orientés.

Démonstration

(i) Si $y = 0_E$: $\left|\langle x, y\rangle\right| = 0 \leqslant 0 = \|x\|.\|y\|$ et dans ce cas d'égalité, x et y sont clairement colinéaires.

Supposons à présent $y \neq 0_E$. La fonction $t \longmapsto \left\|x + ty\right\|^2 = \|x\|^2 + 2t\langle x, y\rangle + t^2\|y\|^2$ est alors polynomiale de degré EXACTEMENT 2, et comme elle est positive ou nulle sur tout $\mathbb{R}$, son discriminant est négatif ou nul : $4\langle x, y\rangle^2 - 4\|x\|^2\|y\|^2 \leqslant 0$, d'où le résultat : $\left|\langle x, y\rangle\right| \leqslant \|x\|.\|y\|$.

À quelle condition a-t-on en fait égalité ? L'inégalité est une égalité si et seulement si le discriminant calculé est nul, i.e. si et seulement si la fonction $t \longmapsto \left\|x + ty\right\|^2$ s'annule. Cela revient à dire que : $x + t_0 y = 0_E$ pour un certain $t_0 \in \mathbb{R}$, i.e. que x et y sont colinéaires.

(ii) D'abord : $\|x+y\|^2 = \|x\|^2 + 2\langle x, y\rangle + \|y\|^2 \overset{\text{(i)}}{\leqslant} \|x\|^2 + 2\|x\|.\|y\| + \|y\|^2 = \left(\|x\| + \|y\|\right)^2$, ensuite on passe à la racine carrée.

À quelle condition a-t-on en fait une égalité ? Si et seulement si : $\langle x, y\rangle = \|x\|.\|y\|$. Les vecteurs x et y sont alors colinéaires d'après (i), et quitte à les permuter, on peut supposer que : $y = \lambda x$ pour un certain $\lambda \in \mathbb{R}$. Aussitôt : $\lambda\|x\|^2 = \langle x, \lambda x\rangle = \langle x, y\rangle = \|x\|.\|y\| = \|x\|.\|\lambda x\| = |\lambda|.\|x\|^2$, donc soit x est nul, soit : $\lambda = |\lambda|$, i.e. : $\lambda \geqslant 0$. Dans les deux cas, x et y sont colinéaires DE MÊME SENS. Réciproque immédiate.

Pour l'inégalité généralisée : $\|x\| = \left\|(x+y) + (-y)\right\| \leqslant \|x+y\| + \|-y\| = \|x+y\| + \|y\|$, donc : $\|x\| - \|y\| \leqslant \|x+y\|$, et de même : $\|y\| - \|x\| \leqslant \|x+y\|$. ∎

Exemple Pour tous $x_1, \dots, x_n \in \mathbb{R}$: $\left(\sum_{k=1}^{n} x_k\right)^2 \leqslant n\sum_{k=1}^{n} x_k^2$, avec égalité si et seulement si : $x_1 = \ldots = x_n$.

Démonstration Simple application de l'inégalité de Cauchy-Schwarz aux vecteurs $(x_1, \dots, x_n)$ et $(1, \dots, 1)$ de $\mathbb{R}^n$ pour le produit scalaire canonique.

Exemple Soient $a, b \in \mathbb{R}$ avec : $a < b$. Pour tout $f \in \mathscr{C}^1\big([a,b], \mathbb{R}\big)$: $f(b)^2 - f(a)^2 \leqslant 2\sqrt{\int_a^b f(t)^2\,\mathrm{d}t}\sqrt{\int_a^b f'(t)^2\,\mathrm{d}t}$.

Démonstration Appliquons l'inégalité de Cauchy-Schwarz à f et f' dans l'espace préhilbertien réel $\mathscr{C}\big([a,b], \mathbb{R}\big)$ muni du produit scalaire $(u, v) \longmapsto \int_a^b u(t)v(t)\,\mathrm{d}t$: $\left|\int_a^b f(t)f'(t)\,\mathrm{d}t\right| \leqslant \sqrt{\int_a^b f(t)^2\,\mathrm{d}t}\sqrt{\int_a^b f'(t)^2\,\mathrm{d}t}$. On conclut en calculant simplement l'intégrale de gauche.

Exemple Soit (Ω, P) un espace probabilisé fini. L'application $(X,Y) \longmapsto \mathrm{E}(XY)$ est une forme bilinéaire symétrique positive sur le $\mathbb{R}$-espace vectoriel $\mathbb{R}^\Omega$ des variables aléatoires réelles sur Ω, MAIS CE N'EST PAS FORCÉMENT UN PRODUIT SCALAIRE. C'en est un si et seulement si pour tout $\omega \in \Omega$: $P\big(\{\omega\}\big) > 0$.

Démonstration L'application $(X,Y) \longmapsto \mathrm{E}(XY)$ a-t-elle la propriété de séparation ? C'est toute la question.

— Si : $P\big(\{\omega\}\big) > 0$ pour tout $\omega \in \Omega$, alors pour toute variable aléatoire $X \in \mathbb{R}^\Omega$, si : $\mathrm{E}\big(X^2\big) = 0$, alors : $\displaystyle\sum_{\omega\in\Omega} P\big(\{\omega\}\big)\, X(\omega)^2 = 0$, donc : $\forall \omega \in \Omega,\quad X(\omega) = 0$, donc : $X = 0$.

— Si $(X,Y) \longmapsto \mathrm{E}(XY)$ est un produit scalaire, alors pour tout $\omega \in \Omega$: $P\big(\{\omega\}\big) = \mathrm{E}\big(\mathbb{1}_{\{\omega\}}\big) = \mathrm{E}\big(\mathbb{1}_{\{\omega\}}^2\big) > 0$.

Théorème (Inégalité de Cauchy-Schwarz pour les variables aléatoires) Soient (Ω, P) un espace probabilisé fini et X et Y deux variables aléatoires sur Ω. Alors :

$$\big|\mathrm{E}(XY)\big| \leqslant \sqrt{\mathrm{E}\big(X^2\big)}\,\sqrt{\mathrm{E}\big(Y^2\big)}. \qquad \text{En particulier : } \quad \big|\mathrm{cov}(X,Y)\big| \leqslant \sigma(X)\,\sigma(Y).$$

Démonstration

- **Cas où $\mathrm{E}\big(Y^2\big) \neq 0$:** On peut reprendre ici à l'identique la preuve précédente de l'inégalité de Cauchy-Schwarz — la propriété de séparation du produit scalaire n'y est pas utilisée.
- **Cas où $\mathrm{E}\big(Y^2\big) = 0$:** La fonction $t \longmapsto \mathrm{E}\big((X+tY)^2\big) = \mathrm{E}\big(X^2\big) + 2t\mathrm{E}(XY)$ est à la fois affine et positive ou nulle sur tout $\mathbb{R}$, donc son coefficient directeur est nul : $\mathrm{E}(XY) = 0$, ce qui nous donne comme voulu : $\big|\mathrm{E}(XY)\big| = 0 \leqslant 0 = \sqrt{\mathrm{E}\big(X^2\big)}\,\sqrt{\mathrm{E}\big(Y^2\big)}$.

Enfin : $\big|\mathrm{cov}(X,Y)\big| = \Big|\mathrm{E}\Big(\big(X-\mathrm{E}(X)\big)\big(Y-\mathrm{E}(Y)\big)\Big)\Big| \leqslant \sqrt{\mathrm{E}\Big(\big(X-\mathrm{E}(X)\big)^2\Big)}\,\sqrt{\mathrm{E}\Big(\big(Y-\mathrm{E}(Y)\big)^2\Big)} = \sigma(X)\,\sigma(Y)$. ■

Exemple Pour toute variable aléatoire centrée X : $\mathrm{E}\big(|X|\big) \leqslant \sqrt{\mathrm{V}(X)}$.

Démonstration D'après l'inégalité de Cauchy-Schwarz : $\mathrm{E}\big(|X|\big) = \mathrm{E}\big(|X| \times 1\big) \leqslant \sqrt{\mathrm{E}\big(X^2\big)}\,\sqrt{\mathrm{E}(1)}$, et comme X est centrée : $\mathrm{E}\big(|X|\big) \leqslant \sqrt{\mathrm{E}\big(X^2\big) - \mathrm{E}(X)^2} = \sqrt{\mathrm{V}(X)}$.

2 ORTHOGONALITÉ

2.1 VECTEURS ORTHOGONAUX, FAMILLES ORTHOGONALES/ORTHONORMALES

Définition (Vecteurs orthogonaux, parties orthogonales, familles orthogonales/orthonormales) Soient E un espace préhilbertien réel, $x, y \in E$, X et Y deux parties de E et $(x_i)_{i\in I}$ une famille de vecteurs de E.

- On dit que x et y sont *orthogonaux* si : $\big\langle x, y\big\rangle = 0$, ce qu'on note : $x \perp y$.
- On dit que les parties X et Y sont *orthogonales*, ce qu'on note : $X \perp Y$, si pour tous $x \in X$ et $y \in Y$: $\big\langle x, y\big\rangle = 0$.
- On dit que la famille $(x_i)_{i\in I}$ est *orthogonale* si pour tous $i, j \in I$ DISTINCTS : $\big\langle x_i, x_j\big\rangle = 0$.
- On dit que la famille $(x_i)_{i\in I}$ est *orthonormale* (ou *orthonormée*) si elle est orthogonale et constituée de vecteurs unitaires, i.e. si pour tous $i, j \in I$: $\big\langle x_i, x_j\big\rangle = \delta_{ij}$.

✌ **Explication** ✌ Notre théorie géométrique a définitivement la tête en bas. Jusqu'ici, pour vous, la notion d'orthogonalité était première et le produit scalaire second. C'est le contraire qui est vrai à présent, la notion d'orthogonalité repose sur la définition préalable d'un produit scalaire. En particulier, à chaque produit produit scalaire est associée une notion d'orthogonalité, ce qui fait que les angles droits ne sont pas droits absolument, mais relativement.

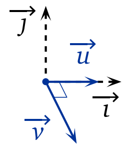
Exemple Pour le produit scalaire canonique de $\mathbb{R}^n$, la base canonique $(E_i)_{1\leqslant i\leqslant n}$ de $\mathbb{R}^n$ est orthonormale car comme on le vérifie aisément, pour tous $i, j \in [\![1, n]\!]$: ${}^tE_iE_j = \delta_{ij}$.

Le résultat suivant est à la fois trivial et essentiel.

Théorème (Vecteurs orthogonaux à tout vecteur) Dans un espace préhilbertien réel, le vecteur nul est le seul vecteur orthogonal à tout vecteur.

Démonstration Soient E un espace préhilbertien réel et $x \in E$. Si x est orthogonal à tout vecteur de E, alors x est en particulier orthogonal à… lui-même : $\langle x, x \rangle = 0$, et ainsi : $x = 0_E$ par séparation. ■

Exemple Pour le produit scalaire $(X, Y) \longmapsto {}^t X \begin{pmatrix} 2 & 1 \\ 1 & 2 \end{pmatrix} Y$ sur $\mathbb{R}^2$, la base canonique n'est pas orthonormale, mais la famille $\left(\dfrac{(1,0)}{\sqrt{2}}, \dfrac{(1,-2)}{\sqrt{6}} \right)$ en est une comme on le vérifie aisément. Cela nous fait apparemment un drôle d'angle droit mais vous devez considérer que c'en est bel et bien un — c'est simplement votre œil qui n'est pas adapté.



Exemple La famille des fonctions $t \longmapsto \sin(nt)$, n décrivant $\mathbb{N}^*$, est orthonormale dans $\mathscr{C}\big([0, 2\pi], \mathbb{R}\big)$ pour le produit scalaire $(f, g) \longmapsto \dfrac{1}{\pi} \displaystyle\int_0^{2\pi} f(t)g(t)\,\mathrm{d}t$.

Démonstration Pour tout $n \in \mathbb{N}^*$: $\dfrac{1}{\pi}\displaystyle\int_0^{2\pi} \sin^2(nt)\,\mathrm{d}t = \dfrac{1}{\pi}\int_0^{2\pi} \dfrac{1-\cos(2nt)}{2}\,\mathrm{d}t = 1 - \dfrac{1}{\pi}\left[\dfrac{\sin(2nt)}{4n}\right]_{t=0}^{t=2\pi} = 1,$

et pour tous $m, n \in \mathbb{N}^*$ **DISTINCTS** :

$$\frac{1}{\pi}\int_0^{2\pi} \sin(mt)\sin(nt)\,\mathrm{d}t = \frac{1}{2\pi}\int_0^{2\pi} \Big(\cos\big((m-n)t\big) - \cos\big((m+n)t\big) \Big)\,\mathrm{d}t = \frac{1}{2\pi}\left[\frac{\sin\big((m-n)t\big)}{m-n} - \frac{\sin\big((m+n)t\big)}{m+n} \right]_{t=0}^{t=2\pi} = 0.$$

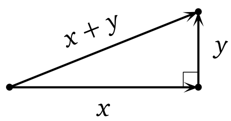
Exemple Dans $\mathscr{C}\big([-1, 1], \mathbb{R}\big)$, l'ensemble des fonctions paires et l'ensemble des fonctions impaires sont deux sous-espaces vectoriels orthogonaux pour le produit scalaire $(f, g) \longmapsto \displaystyle\int_{-1}^{1} f(t)g(t)\,\mathrm{d}t$.

Démonstration Pour toutes $p \in \mathscr{C}\big([-1, 1], \mathbb{R}\big)$ paire et $i \in \mathscr{C}\big([-1, 1], \mathbb{R}\big)$ impaire : $\displaystyle\int_{-1}^{1} \underbrace{p(t)i(t)}_{\text{impaire}}\,\mathrm{d}t = 0.$

Théorème (Propriétés des familles orthogonales) Soit E un espace préhilbertien réel.

(i) **Théorème de Pythagore :** Pour tous $x, y \in E$, x et y sont orthogonaux si et seulement si : $\|x + y\|^2 = \|x\|^2 + \|y\|^2$.

En outre, pour toute famille orthogonale $(x_1, \ldots, x_n)$ de E : $\left\| \displaystyle\sum_{i=1}^n x_i \right\|^2 = \sum_{i=1}^n \|x_i\|^2.$



(ii) Toute famille orthogonale de vecteurs **NON NULS** de E est libre.

En particulier, si E est de dimension finie $n \neq 0$, toute famille famille orthonormale de n vecteurs de E est déjà une **BASE** orthonormale de E.

Démonstration

(i) Tout simplement : $\left\| \displaystyle\sum_{i=1}^n x_i \right\|^2 = \sum_{i=1}^n \|x_i\|^2 + 2 \sum_{1 \leqslant i < j \leqslant n} \underbrace{\langle x_i, x_j \rangle}_{=0} = \sum_{i=1}^n \|x_i\|^2.$

(ii) Soient $(x_1, \ldots, x_n)$ une famille orthogonale de vecteurs non nuls de E et $\lambda_1, \ldots, \lambda_n \in \mathbb{R}$ pour lesquels : $\displaystyle\sum_{k=1}^n \lambda_k x_k = 0_E$. Pour tout $i \in [\![1, n]\!]$: $0 = \langle 0_E, x_i \rangle = \left\langle \displaystyle\sum_{k=1}^n \lambda_k x_k, x_i \right\rangle = \sum_{k=1}^n \lambda_k \langle x_k, x_i \rangle = \lambda_i \|x_i\|^2$, donc comme $x_i \neq 0_E$: $\|x_i\| \neq 0$, et donc : $\lambda_i = 0$. La famille $(x_1, \ldots, x_n)$ est ainsi libre.

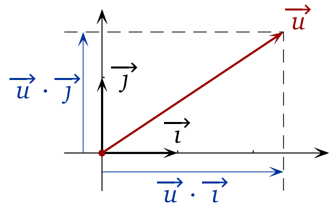
Pour le cas particulier, toute famille orthonormale de E est libre comme on vient de le voir puisque ses vecteurs — unitaires — sont non nuls, et bien sûr une telle famille est une base de E si elle a pour cardinal la dimension de E. ■

2.2 Coordonnées dans une base orthonormale

Théorème (Coordonnées dans une base orthonormale) Soient $E \neq \{0_E\}$ un espace euclidien, $(e_1, \dots, e_n)$ une base **ORTHONORMALE** de E et $x \in E$.

Alors : $x = \displaystyle\sum_{k=1}^{n} \langle x, e_k \rangle e_k$. En d'autres termes, les coordonnées de x dans $(e_1, \dots, e_n)$ sont $\big(\langle x, e_1 \rangle, \dots, \langle x, e_n \rangle\big)$.

Explication Vous connaissez bien ce résultat dans le plan et dans l'espace !



Démonstration Notons $(x_1, \dots, x_n)$ les coordonnées de x dans $(e_1, \dots, e_n)$.

Pour tout $k \in [\![1, n]\!]$: $\langle x, e_k \rangle = \left\langle \displaystyle\sum_{i=1}^{n} x_i e_i, e_k \right\rangle = \displaystyle\sum_{i=1}^{n} x_i \langle e_i, e_k \rangle = \sum_{i=1}^{n} x_i \delta_{ki} = x_k$. ■

Théorème (Expression du produit scalaire et de la norme dans une base orthonormale) Soient $E \neq \{0_E\}$ un espace euclidien et $x, y \in E$ de coordonnées respectives $X = (x_1, \dots, x_n)$ et $Y = (y_1, \dots, y_n)$ dans une certaine base **ORTHONORMALE** de E.

$$\langle x, y \rangle = \sum_{k=1}^{n} x_k y_k = {}^t XY \qquad \text{et} \qquad \|x\| = \sqrt{\sum_{k=1}^{n} x_k^2} = \sqrt{{}^t XX}.$$

Explication Ce résultat montre que finalement, le produit scalaire canonique sur $\mathbb{R}^n$ est un modèle pour tous les produits scalaires des espaces euclidiens. Calculer le produit scalaire $\langle x, y \rangle$ dans un espace euclidien abstrait revient à calculer le produit scalaire canonique des coordonnées des vecteurs x et y dans une base **ORTHONORMALE** quelconque.

✖ **Attention !** ✖ Ces formules sont fausses en général pour des coordonnées dans une base **NON** orthonormale.

Démonstration Notons $(e_1, \dots, e_n)$ la base orthonormale considérée.

Tout simplement : $\langle x, y \rangle = \left\langle \displaystyle\sum_{i=1}^{n} x_i e_i, \sum_{j=1}^{n} y_j e_j \right\rangle = \displaystyle\sum_{1 \leqslant i,j \leqslant n} x_i y_j \langle e_i, e_j \rangle = \sum_{1 \leqslant i,j \leqslant n} x_i y_j \delta_{ij} = \sum_{k=1}^{n} x_k y_k = {}^t XY$. ■

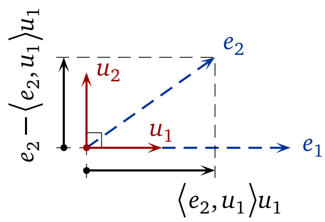
2.3 Algorithme d'orthonormalisation de Gram-Schmidt

Théorème (Algorithme d'orthonormalisation de Gram-Schmidt) Soient E un espace préhilbertien réel et $(e_1, \dots, e_n)$ une famille **LIBRE** de E. On peut transformer $(e_1, \dots, e_n)$ en une famille orthonormale $(u_1, \dots, u_n)$ de E telle que :

$$\forall k \in [\![1, n]\!], \quad \mathrm{Vect}(e_1, \dots, e_k) = \mathrm{Vect}(u_1, \dots, u_k).$$

Les vecteurs $u_1, \dots, u_n$ peuvent être construits de proche en proche depuis u_1 jusqu'à u_n, et pour tout $k \in [\![1, n]\!]$, on n'a que deux choix possibles pour u_k, u_k est soit le vecteur : $\dfrac{e_k - \displaystyle\sum_{i=1}^{k-1} \langle e_k, u_i \rangle u_i}{\left\| e_k - \displaystyle\sum_{i=1}^{k-1} \langle e_k, u_i \rangle u_i \right\|}$, soit son opposé.

✌ **Explication** ✌ On a tout compris quand on a compris le cas $n=2$. On veut transformer une famille libre quelconque (e_1,e_2) en une famille orthonormale (u_1,u_2) avec : $u_1=\dfrac{e_1}{\|e_1\|}$ et $u_2=\dfrac{e_2-\langle e_2,u_1\rangle u_1}{\left\|e_2-\langle e_2,u_1\rangle u_1\right\|}$.



- La première formule normalise e_1, c'est-à-dire le rend unitaire.
- La seconde commence par transformer e_2 en $e_2-\langle e_2,u_1\rangle u_1$, c'est-à-dire à retrancher à e_2 sa composante selon u_1, qui vaut $\langle e_2,u_1\rangle u_1$. Le vecteur ainsi obtenu est orthogonal à u_1, mais il peut ne pas être unitaire, c'est pourquoi on le divise par sa norme.

Plus généralement, l'algorithme de Gram-Schmidt construit u_k en ôtant de e_k ses composantes selon $u_1,\dots,u_{k-1}$, puis en rendant le tout unitaire.

Démonstration

- La construction commence simplement. La famille (e_1) est libre, donc : $e_1\neq 0_E$. On pose : $u_1=\dfrac{e_1}{\|e_1\|}$. Clairement : $\mathrm{Vect}(e_1)=\mathrm{Vect}(u_1)$.
- Soit $k\in[\![1,n]\!]$. Supposons qu'on ait réussi à construire une famille orthonormale $(u_1,\dots,u_{k-1})$ pour laquelle : $\mathrm{Vect}(e_1,\dots,e_p)=\mathrm{Vect}(u_1,\dots,u_p)$ pour tout $p\in[\![0,k-1]\!]$. Nous sommes en quête d'un vecteur u_k pour lequel $(u_1,\dots,u_k)$ est orthonormale et : $\mathrm{Vect}(e_1,\dots,e_k)=\mathrm{Vect}(u_1,\dots,u_k)$.
- **Analyse :** Si un tel u_k existe, il est combinaison linéaire de $e_1,\dots,e_k$ car : $\mathrm{Vect}(e_1,\dots,e_k)=\mathrm{Vect}(u_1,\dots,u_k)$, mais comme : $\mathrm{Vect}(e_1,\dots,e_{k-1})=\mathrm{Vect}(u_1,\dots,u_{k-1})$, u_k est aussi combinaison linéaire de $u_1,\dots,u_{k-1},e_k$, de sorte que pour certains $a_{1k},\dots,a_{kk}\in\mathbb{R}$: $u_k=a_{kk}e_k+\displaystyle\sum_{i=1}^{k-1}a_{ik}u_i$.
Dans ces conditions, pour tout $i\in[\![1,k-1]\!]$, sachant que u_k et u_i sont orthogonaux :
$$0=\langle u_k,u_i\rangle=\left\langle a_{kk}e_k+\sum_{j=1}^{k-1}a_{jk}u_j,u_i\right\rangle=a_{kk}\langle e_k,u_i\rangle+\sum_{j=1}^{k-1}a_{jk}\langle u_j,u_i\rangle=a_{kk}\langle e_k,u_i\rangle+a_{ik},$$
ou encore : $a_{ik}=-a_{kk}\langle e_k,u_i\rangle$. Conclusion : $u_k=a_{kk}\widehat{u}_k$ avec : $\widehat{u}_k=e_k-\displaystyle\sum_{i=1}^{k-1}\langle e_k,u_i\rangle u_i$, mais par ailleurs $\|u_k\|=1$, donc : $|a_{kk}|=\dfrac{1}{\|\widehat{u}_k\|}$, i.e. : $a_{kk}=\dfrac{\pm 1}{\|\widehat{u}_k\|}$. Conclusion : $u_k=\pm\dfrac{\widehat{u}_k}{\|\widehat{u}_k\|}$. Le vecteur u_k, s'il existe, est finalement déterminé AU SIGNE PRÈS par $u_1,\dots,u_{k-1}$ et e_k.
- **Synthèse :** Réciproquement, posons : $\widehat{u}_k=e_k-\displaystyle\sum_{i=1}^{k-1}\langle e_k,u_i\rangle u_i$.
Se peut-il que $\widehat{u}_k$ soit nul ? La liberté de $(e_1,\dots,e_n)$ s'en trouverait contredite car e_k serait combinaison linéaire de $u_1,\dots,u_{k-1}$, et donc de $e_1,\dots,e_{k-1}$ puisque : $\mathrm{Vect}(e_1,\dots,e_{k-1})=\mathrm{Vect}(u_1,\dots,u_{k-1})$. Nous pouvons ainsi poser : $u_k=\dfrac{\widehat{u}_k}{\|\widehat{u}_k\|}$ — on pourrait aussi choisir : $u_k=-\dfrac{\widehat{u}_k}{\|\widehat{u}_k\|}$.
Montrons que la famille $(u_1,\dots,u_k)$ est orthonormale. Comme $(u_1,\dots,u_{k-1})$ l'est par hypothèse de récurrence, et comme u_k est unitaire, il nous suffit de montrer que u_k est orthogonal à $u_1,\dots,u_{k-1}$. Le caractère unitaire de u_k est limpide. Or pour tout $j\in[\![1,k-1]\!]$:
$$\langle u_k,u_j\rangle=\frac{1}{\|\widehat{u}_k\|}\langle\widehat{u}_k,u_j\rangle=\frac{1}{\|\widehat{u}_k\|}\left\langle e_k-\sum_{i=1}^{k-1}\langle e_k,u_i\rangle u_i,u_j\right\rangle=\frac{1}{\|\widehat{u}_k\|}\Big(\langle e_k,u_j\rangle-\langle e_k,u_j\rangle\langle u_j,u_j\rangle\Big)=0.$$
Enfin, dans la mesure où : $\mathrm{Vect}(e_1,\dots,e_{k-1})=\mathrm{Vect}(u_1,\dots,u_{k-1})$, et puisque u_k est combinaison linéaire de $u_1,\dots,u_{k-1}$ et e_k, alors comme voulu : $\mathrm{Vect}(e_1,\dots,e_k)=\mathrm{Vect}(u_1,\dots,u_k)$. ■

Exemple La famille $\left(1\,,\ \sqrt{3}\,(2X-1)\,,\ \sqrt{5}\left(6X^2-6X+1\right)\right)$ est orthonormale pour le produit scalaire $(P,Q)\longmapsto\displaystyle\int_0^1 P(t)Q(t)\,\mathrm{d}t$ sur $\mathbb{R}[X]$.

Démonstration Nous allons, grâce à l'algorithme de Gram-Schmidt, construire une famille orthonormale (P_0,P_1,P_2) à partir de la famille LIBRE $\left(1,X,X^2\right)$.

- **Construction de P_0 :** Posons : $P_0=\dfrac{1}{\|1\|}$. Comme : $\|1\|^2=\displaystyle\int_0^1\mathrm{d}t=1$, en fait : $P_0=1$.

- **Construction de P_1 :** Posons : $P_1 = \dfrac{X - \langle X, P_0\rangle P_0}{\left\| X - \langle X, P_0\rangle P_0 \right\|}$. Comme : $\langle X, P_0\rangle = \displaystyle\int_0^1 t \, \mathrm{d}t = \frac{1}{2}$, le polynôme au numérateur vaut : $X - \dfrac{1}{2}$, mais : $\left\| X - \dfrac{1}{2} \right\|^2 = \displaystyle\int_0^1 \left(t - \frac{1}{2} \right)^2 \mathrm{d}t = \frac{1}{12}$, donc : $P_1 = \sqrt{3}\,(2X-1)$. La famille (P_0, P_1) est orthonormale.
- **Construction de P_2 :** Posons : $P_2 = \dfrac{X^2 - \langle X^2, P_0\rangle P_0 - \langle X^2, P_1\rangle P_1}{\left\| X^2 - \langle X^2, P_0\rangle P_0 - \langle X^2, P_1\rangle P_1 \right\|}$. Comme : $\langle X^2, P_0\rangle = \displaystyle\int_0^1 t^2 \, \mathrm{d}t = \frac{1}{3}$ et $\langle X^2, P_1\rangle = \displaystyle\int_0^1 t^2 \times \sqrt{3}(2t-1) \, \mathrm{d}t = \frac{1}{2\sqrt{3}}$, le polynôme au numérateur vaut : $X^2 - X + \dfrac{1}{6}$, mais : $\left\| X^2 - X + \dfrac{1}{6} \right\|^2 = \displaystyle\int_0^1 \left(t^2 - t + \frac{1}{6} \right)^2 \mathrm{d}t = \frac{1}{180}$, donc : $P_2 = \sqrt{5}\left(6X^2 - 6X + 1\right)$. La famille (P_0, P_1, P_2) est orthonormale.

Essentiel en pratique, l'algorithme d'orthonormalisation de Gram-Schmidt a aussi des conséquences théoriques.

Théorème (Existence de bases orthonormales en dimension finie) Soit $E \neq \{0_E\}$ un espace euclidien. Alors E possède une base orthonormale.

Démonstration De dimension finie non nulle, E possède une base, donc une base orthonormale grâce à l'algorithme de Gram-Schmidt. ■

Théorème (Théorème de la base orthonormale incomplète en dimension finie) Soit $E \neq \{0_E\}$ un espace euclidien. Toute famille orthonormale de E peut être complétée en une base orthonormale de E.

Démonstration Soit $(e_1, \ldots, e_p)$ une famille orthonormale — donc libre — de E. On peut la compléter en une base de E car E est de dimension finie non nulle, puis orthonormaliser cette base grâce à l'algorithme de Gram-Schmidt. L'algorithme n'affecte pas les premiers vecteurs $e_1, \ldots, e_p$ qui forment déjà une famille orthonormale, donc au fond nous avons complété notre famille de départ en une base orthonormale de E. ■

2.4 Supplémentaire orthogonal d'un sous-espace vectoriel

Définition-théorème (Orthogonal d'une partie) Soient E un espace préhilbertien réel et X une partie de E. On appelle *orthogonal de X dans E* l'ensemble : $X^\perp = \left\{ t \in E /\quad \forall x \in X, \quad \langle t, x\rangle = 0 \right\}$.

(i) $X^\perp$ est un sous-espace vectoriel de E orthogonal à X.

(ii) Si X est un sous-espace vectoriel de E, X et $X^\perp$ sont en somme directe.

(iii) $X^\perp = \mathrm{Vect}(X)^\perp$ et $X \subset X^{\perp\perp}$.

En pratique L'égalité : $X^\perp = \mathrm{Vect}(X)^\perp$ signifie que pour déterminer l'orthogonal d'un sous-espace vectoriel engendré par un ensemble X de vecteurs, il est suffisant d'exiger l'orthogonalité à ces vecteurs.

✖ **Attention !** ✖ Si F un sous-espace vectoriel de E, ce théorème montre en particulier que F et $F^\perp$ sont en somme directe. Il n'est pas vrai en général, en revanche, que F et $F^\perp$ sont supplémentaires dans E, ils sont seulement en somme directe. De même, il n'est pas vrai en général que : $F^{\perp\perp} = F$. Nous verrons tout de même que ces résultats sont vrais si F est de dimension finie, donc en particulier si E est euclidien.

Démonstration

(i) D'abord : $0_E \in X^\perp$ car pour tout $x \in X$: $\langle 0_E, x\rangle = 0$. Ensuite, soient $t, t' \in X^\perp$ et $\lambda, \lambda' \in \mathbb{R}$. Pour tout $x \in X$: $\langle \lambda t + \lambda' t', x\rangle = \lambda\langle x, t\rangle + \lambda'\langle t', x\rangle = \lambda.0 + \lambda'.0 = 0$, donc : $\lambda t + \lambda' t' \in X^\perp$.

(i) Si X est un sous-espace vectoriel de E : $0_E \in X \cap X^\perp$. Montrons inversement que : $X \cap X^\perp \subset \{0_E\}$. Or pour tout $x \in X \cap X^\perp$: $x \perp x$ donc : $\langle x, x \rangle = 0$, donc : $x = 0_E$ par séparation.

(iii) Par linéarité à droite du produit scalaire, il est équivalent pour un vecteur d'être orthogonal à X ou à ses combinaisons linéaires, autrement à $\mathrm{Vect}(X)$, donc en effet : $X^\perp = \mathrm{Vect}(X)^\perp$. Enfin, l'inclusion : $X \subset X^{\perp\perp}$ signifie juste que tout vecteur de X est orthogonal à tout vecteur de $X^\perp$ — ce qui est évident par définition de $X^\perp$. ■

Exemple Pour tout espace préhilbertien réel E : $\{0_E\}^\perp = E$ et $E^\perp = \{0_E\}$.
Pour la seconde égalité, rappelons que 0_E EST LE SEUL VECTEUR DE E ORTHOGONAL À TOUT VECTEUR — et pourquoi ?

Exemple Pour le produit scalaire $(P,Q) \longmapsto P(-1)Q(-1) + P(0)Q(0) + P(1)Q(1)$ sur $\mathbb{R}_2[X]$: $\mathbb{R}_1[X]^\perp = \mathrm{Vect}\big(3X^2 - 2\big)$.

Démonstration Pour tout $P = aX^2 + bX + c \in \mathbb{R}_2[X]$: $P \in \mathbb{R}_1[X]^\perp \overset{\mathbb{R}_1[X]=\mathrm{Vect}(1,X)}{\Longleftrightarrow} \langle P, 1 \rangle = 0$ et $\langle P, X \rangle = 0$

$$\Longleftrightarrow \left\{ \begin{array}{rcccccl} (a-b+c)\times 1 & + & c \times 1 & + & (a+b+c)\times 1 & = & 0 \\ (a-b+c)\times(-1) & + & c\times 0 & + & (a+b+c)\times 1 & = & 0 \end{array} \right. \Longleftrightarrow 2a + 3c = 0 \quad \text{et} \quad 2b = 0.$$

Définition-théorème (Supplémentaire orthogonal d'un sous-espace vectoriel de dimension finie) Soient E un espace préhilbertien réel et F un sous-espace vectoriel DE DIMENSION FINIE de E.

(i) $F^\perp$ est un supplémentaire de F dans E orthogonal à F et c'est même le seul. On l'appelle par conséquent LE *supplémentaire orthogonal de F dans E*.

(ii) $F^{\perp\perp} = F$.

✖ ATTENTION ! ✖

- Il est ici essentiel que F soit DE DIMENSION FINIE.
- Il existe UN unique supplémentaire orthogonal mais tout plein de supplémentaires « généraux », ne l'oubliez pas.

Démonstration

(i) Nous savons déjà que : $F \cap F^\perp = \{0_E\}$.

- Montrons que : $E = F + F^\perp$, i.e. que : $E \subset F + F^\perp$. Nous pouvons supposer : $F \neq \{0_E\}$ car : $\{0_E\}^\perp = E$, et comme F est de dimension finie, nous donner une base orthonormale $(f_1, \ldots, f_n)$ de F. Soit $x \in E$. Pour tout $i \in [\![1,n]\!]$:
$$\left\langle x - \sum_{k=1}^n \langle x, f_k \rangle f_k, f_i \right\rangle = \langle x, f_i \rangle - \sum_{k=1}^n \langle x, f_k \rangle \langle f_k, f_i \rangle = \langle x, f_i \rangle - \sum_{k=1}^n \langle x, f_k \rangle \delta_{ki} = \langle x, f_i \rangle - \langle x, f_i \rangle = 0,$$
donc $x - \sum_{k=1}^n \langle x, f_k \rangle f_k$ est orthogonal à $f_1, \ldots, f_n$, donc élément de $F^\perp$. Comme voulu : $x \in F + F^\perp$.
- Montrons que $F^\perp$ est le seul supplémentaire de F dans E orthogonal à F. Soit F' un tel supplémentaire. Aussitôt : $F' \subset F^\perp$ car : $F \perp F'$. Inversement, soit $x \in F^\perp$. Comme : $E = F + F'$, alors : $x = f + f'$ pour certains $f \in F$ et $f' \in F'$, donc : $\langle f, f \rangle = \langle f + f', f \rangle = \langle x, f \rangle = 0$, puis : $f = 0_E$, et enfin : $x = f' \in F'$.

(ii) Pour l'inclusion : $F^{\perp\perp} \subset F$, soit $x \in F^{\perp\perp}$, disons : $x = f + f'$ avec $f \in F$ et $f' \in F^\perp$. Aussitôt : $\langle f', f' \rangle = \langle f + f', f' \rangle = \langle x, f' \rangle = 0$, donc : $f' = 0_E$ par séparation, i.e. : $x = f \in F$. ■

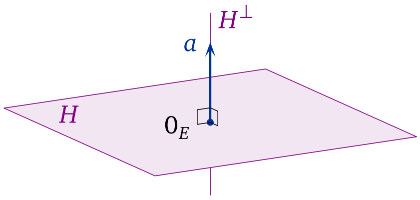
2.5 VECTEURS NORMAUX À UN HYPERPLAN ET ORIENTATION

Définition-théorème (Vecteurs normaux à un hyperplan) Soient $E \neq \{0_E\}$ un espace euclidien de dimension n et H un hyperplan de E. Le sous-espace $H^\perp$ est une droite dont tout vecteur non nul est appelé un *vecteur normal à H*.

Par extension, pour tout hyperplan affine $\mathscr{H}$ de E de direction H, les vecteurs normaux à H sont aussi qualifiés de *vecteurs normaux à $\mathscr{H}$*.

Démonstration Comme E est de dimension finie : $E = H \oplus H^\perp$, et par conséquent, comme H est un hyperplan : $\dim H^\perp = \dim E - \dim H = 1$. ■

Explication Rappelons qu'un hyperplan H est par définition le noyau d'une forme linéaire non nulle, autrement dit un ensemble décrit par une unique équation linéaire scalaire non nulle. Comment les notions de forme linéaire et de vecteur normal sont-elles alors liées ? Le plus simplement du monde à vrai dire, car :



$$H = \mathrm{Vect}(a)^\perp = \left\{x \in E / \quad \langle x, a \rangle = 0\right\},$$

H est le noyau de la forme linéaire non nulle $x \longmapsto \langle x, a \rangle$. En outre, si nous nous donnons une base orthornormale $(e_1, \dots, e_n)$ de E et si a a pour coordonnées $(a_1, \dots, a_n)$ dans cette base, l'équation : $\langle x, a \rangle = 0$ qui peut servir à définir H s'écrit aussi : $a_1 x_1 + \dots + a_n x_n = 0$ sous forme analytique. Nous retrouvons là l'équation typique d'un hyperplan et le fait que les coefficients $a_1, \dots, a_n$ d'une telle équation définissent naturellement un vecteur normal.

Exemple

- Dans l'espace euclidien canonique $\mathbb{R}^2$, l'hyperplan AFFINE $\mathscr{H}$ d'équation : $3x + 2y = 1$ admet pour direction l'hyperplan VECTORIEL H d'équation : $3x + 2y = 0$, équation qu'on peut réécrire comme un produit scalaire : $(x, y) \cdot (3, 2) = 0$. A fortiori, H est l'ensemble des vecteurs de $\mathbb{R}^2$ orthogonaux à $(3, 2)$: $H = (3, 2)^\perp$. On qualifie le vecteur $(3, 2)$ de *vecteur normal à H (ou à $\mathscr{H}$)*.
- De même, dans l'espace euclidien canonique $\mathbb{R}^3$, l'hyperplan AFFINE $\mathscr{H}$ d'équation : $x - y + 2z = 3$ admet pour direction l'hyperplan VECTORIEL H d'équation : $x - y + 2z = 0$, équation qu'on peut réécrire comme un produit scalaire : $(x, y, z) \cdot (1, -1, 2) = 0$. A fortiori, H est l'ensemble des vecteurs de $\mathbb{R}^3$ orthogonaux à $(1, -1, 2)$: $H = (1, -1, 2)^\perp$. On qualifie le vecteur $(1, -1, 2)$ de *vecteur normal à H (ou à $\mathscr{H}$)*.

Définition-théorème (Orientation d'un hyperplan par un vecteur normal) Soient $E \neq \{0_E\}$ un espace euclidien ORIENTÉ de dimension n et H un hyperplan de E.

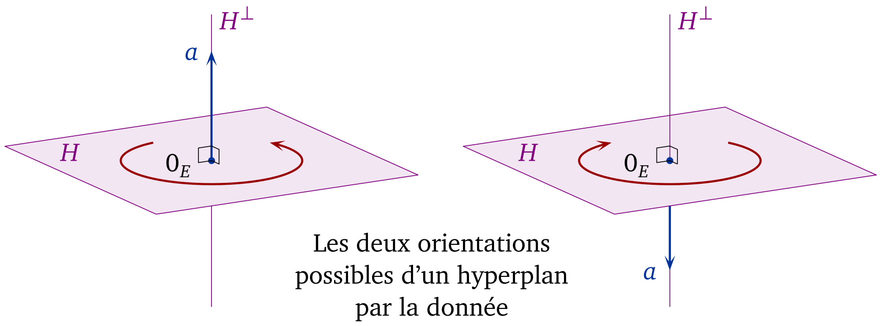
L'espace vectoriel E a beau être orienté, son sous-espace vectoriel H ne l'est pas a priori. Le choix supplémentaire d'un vecteur normal a à H donne en revanche à H une orientation de la manière suivante :

Une base $(h_1, \dots, h_{n-1})$ de H est tenue pour directe si et seulement si la base $(h_1, \dots, h_{n-1}, a)$ de E l'est.

Démonstration On a défini au chapitre « Déterminants » une relation « avoir la même orientation » sur l'ensemble des bases d'un $\mathbb{R}$-espace vectoriel quelconque de dimension finie. L'orientation de cet espace consistait alors en le choix arbitraire de l'une des deux classes d'équivalence de cette relation, considérée comme la classe des bases directes. Notre orientation de H à partir de l'orientation de E et d'un vecteur normal a à H respecte-t-elle cette définition ?

Donnons-nous deux bases $\mathscr{B} = (h_1, \dots, h_{n-1})$ et $\mathscr{B}' = (h'_1, \dots, h'_{n-1})$ de H. Les familles $\mathscr{B}_a = (h_1, \dots, h_{n-1}, a)$ et $\mathscr{B}'_a = (h'_1, \dots, h'_{n-1}, a)$ sont alors deux bases de E et : $\mathrm{Mat}_{\mathscr{B}_a}(\mathscr{B}'_a) = \begin{pmatrix} \mathrm{Mat}_{\mathscr{B}}(\mathscr{B}') & 0 \\ 0 & 1 \end{pmatrix}$. En particulier : $\det_{\mathscr{B}_a}(\mathscr{B}'_a) = \det_{\mathscr{B}}(\mathscr{B}')$, donc ces deux déterminants ont le même signe. En termes d'orientation, cela signifie que $\mathscr{B}'$ a la même orientation que $\mathscr{B}$ si et seulement si $\mathscr{B}'_a$ a la même orientation que $\mathscr{B}_a$. Via le choix d'un vecteur normal a, cette équivalence justifie l'idée que l'orientation de E se transmet à H. ■

Explication Les hyperplans d'un $\mathbb{R}$-espace vectoriel orienté ne sont pas naturellement orientés quant à eux, car par exemple en dimension 3, on ne perçoit pas l'orientation d'un plan de la même manière selon le demi-espace depuis lequel on observe ce plan. Se donner un vecteur normal d'un hyperplan, cela revient justement à choisir un demi-espace parmi les deux que l'hyperplan délimite, donc à choisir un point de vue à partir duquel on peut envisager une orientation de l'hyperplan.



Les deux orientations possibles d'un hyperplan par la donnée d'un vecteur normal

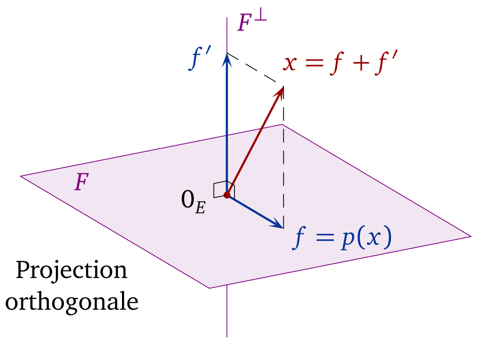
3 Projection orthogonale sur un sous-espace vectoriel de dimension finie

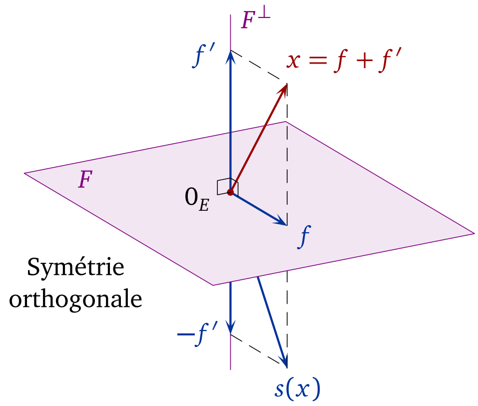
3.1 Projections et symétries orthogonales

Définition (Projection orthogonale, symétrie orthogonale, réflexion) Soient E un espace préhilbertien réel et F un sous-espace vectoriel DE DIMENSION FINIE de E.

- On appelle *projection orthogonale sur F* ou *projecteur orthogonal sur F* la projection sur F de direction $F^\perp$.
- On appelle *symétrie orthogonale par rapport à F* la symétrie par rapport à F parallèlement à $F^\perp$.
  On parle plutôt de *réflexion par rapport à F* lorsque F est un hyperplan de E.

Explication Comme F est de dimension finie : $E = F \oplus F^\perp$.





Théorème (Expression d'une projection orthogonale dans une base orthonormale) Soient E un espace préhilbertien réel, F un sous-espace vectoriel DE DIMENSION FINIE non nulle de E et $(f_1, \ldots, f_n)$ une base orthonormale de F.

Si on note p la projection orthogonale sur F, alors pour tout $x \in E$: $p(x) = \displaystyle\sum_{k=1}^n \langle x, f_k \rangle f_k$.

Explication Dans ce résultat, « $\langle x, f_k \rangle f_k$ » représente la composante de x selon le vecteur f_k.

Démonstration Soit $x \in E$. Le vecteur $x - \displaystyle\sum_{k=1}^n \langle x, f_k \rangle f_k$ est orthogonal à $f_1, \ldots, f_n$, donc est élément de $F^\perp$.

Ainsi : $x = \underbrace{\displaystyle\sum_{k=1}^n \langle x, f_k \rangle f_k}_{\in F} + \underbrace{\left(x - \displaystyle\sum_{k=1}^n \langle x, f_k \rangle f_k \right)}_{\in F^\perp}$, et donc : $p(x) = \displaystyle\sum_{k=1}^n \langle x, f_k \rangle f_k$. ■

En pratique On peut calculer essentiellement de deux manières un projeté orthogonal. Donnons-nous E un espace préhilbertien réel, F un sous-espace vectoriel de dimension finie non nulle de E, $(f_1, \ldots, f_n)$ une base PAS FORCÉMENT ORTHONORMALE de F et $x \in E$. Comment calculer le projeté orthogonal $p(x)$ de x sur F ? Si la base $(f_1, \ldots, f_n)$ est orthonormale, on peut bien sûr utiliser le théorème précédent, mais sinon ?

- **Première stratégie :** On orthonormalise $(f_1, \ldots, f_n)$ grâce à l'algorithme de Gram-Schmidt et du coup on peut utiliser le théorème précédent.
- **Deuxième stratégie :** On introduit les coordonnées $(\lambda_1, \ldots, \lambda_n)$ de $p(x)$ dans $(f_1, \ldots, f_n)$: $p(x) = \displaystyle\sum_{i=1}^n \lambda_i f_i$, puis on les calcule grâce aux relations : $\langle x - p(x), f_j \rangle = 0$ pour tout $j \in [\![1, n]\!]$. Ces relations expriment l'appartenance de $x - p(x)$ à $F^\perp$ et fournissent un système de n équations à n inconnues que l'on n'a plus qu'à résoudre.

Exemple On note F le sous-espace vectoriel $\mathrm{Vect}(\sin,\cos)$ de $\mathscr{C}\big([0,2\pi],\mathbb{R}\big)$. Le projeté orthogonal de l'identité Id sur F pour le produit scalaire $(f,g)\longmapsto \displaystyle\int_0^{2\pi} f(t)g(t)\,\mathrm{d}t$ est la fonction $t\longmapsto -2\sin t$.

Démonstration Nous aurons besoin d'un certain nombre de produits scalaires, calculons-les de prime abord pour que l'essentiel des stratégies adoptées soit bien lisible ensuite.

$$\|\sin\|^2=\int_0^{2\pi}\sin^2 t\,\mathrm{d}t=\int_0^{2\pi}\frac{1-\cos(2t)}{2}\,\mathrm{d}t=\left[\frac{t}{2}-\frac{\sin(2t)}{4}\right]_{t=0}^{t=2\pi}=\pi,\qquad \text{et de même :}\quad \|\cos\|^2=\pi.$$

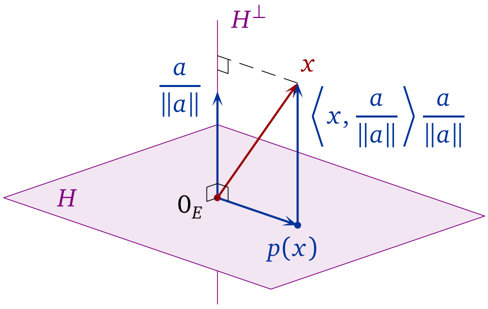
En outre : $\big\langle \sin,\cos\big\rangle=\displaystyle\int_0^{2\pi}\sin t\cos t\,\mathrm{d}t=\left[\frac{\sin^2 t}{2}\right]_{t=0}^{t=2\pi}=0$, donc la famille $(\sin,\cos)$ est orthogonale.

Enfin : $\displaystyle\int_0^{2\pi} t\mathrm{e}^{\mathrm{i}t}\,\mathrm{d}t=\Big[t\times(-\mathrm{i})\mathrm{e}^{\mathrm{i}t}\Big]_{t=0}^{t=2\pi}-\int_0^{2\pi}(-\mathrm{i})\mathrm{e}^{\mathrm{i}t}\,\mathrm{d}t=-2\mathrm{i}\pi+\mathrm{i}\Big[(-\mathrm{i})\mathrm{e}^{\mathrm{i}t}\Big]_{t=0}^{t=2\pi}=-2\mathrm{i}\pi+0=-2\mathrm{i}\pi$, donc :

$$\big\langle \mathrm{Id},\sin\big\rangle=\int_0^{2\pi}t\sin t\,\mathrm{d}t=\mathrm{Im}\left(\int_0^{2\pi}t\mathrm{e}^{\mathrm{i}t}\,\mathrm{d}t\right)=\mathrm{Im}(-2\mathrm{i}\pi)=-2\pi\qquad\text{et de même :}\quad\big\langle \mathrm{Id},\cos\big\rangle=0.$$

- **Première stratégie :** On commence par orthonormaliser la famille **LIBRE** $(\sin,\cos)$ de F grâce à l'algorithme de Gram-Schmidt. Cette famille étant déjà orthogonale, on n'a par chance qu'à la normaliser, et en l'occurrence $\left(\dfrac{\sin}{\|\sin\|},\dfrac{\cos}{\|\cos\|}\right)=\left(\dfrac{\sin}{\sqrt{\pi}},\dfrac{\cos}{\sqrt{\pi}}\right)$ est une base orthonormale de F. Le projeté orthogonal de Id sur F est finalement la fonction :
$$\left\langle \mathrm{Id},\frac{\sin}{\sqrt{\pi}}\right\rangle\frac{\sin}{\sqrt{\pi}}+\left\langle \mathrm{Id},\frac{\cos}{\sqrt{\pi}}\right\rangle\frac{\cos}{\sqrt{\pi}}=\frac{\big\langle \mathrm{Id},\sin\big\rangle}{\pi}\sin+\frac{\big\langle \mathrm{Id},\cos\big\rangle}{\pi}\cos=-2\sin.$$
- **Deuxième stratégie :** Notons $p(\mathrm{Id})$ le projeté orthogonal de Id sur F et (λ,μ) ses coordonnées dans la base $(\sin,\cos)$ de F : $p(\mathrm{Id})=\lambda\sin+\mu\cos$. Appuyons-nous sur le fait que : $\mathrm{Id}-p(\mathrm{Id})\in F^\perp$.
$$0=\big\langle \mathrm{Id}-p(\mathrm{Id}),\sin\big\rangle=\big\langle \mathrm{Id}-\lambda\sin-\mu\cos,\sin\big\rangle=\big\langle \mathrm{Id},\sin\big\rangle-\lambda\|\sin\|^2-\mu\big\langle \cos,\sin\big\rangle=-2\pi-\lambda\pi$$
$$\text{et}\qquad 0=\big\langle \mathrm{Id}-p(\mathrm{Id}),\cos\big\rangle=\big\langle \mathrm{Id}-\lambda\sin-\mu\cos,\cos\big\rangle=\big\langle \mathrm{Id},\cos\big\rangle-\lambda\big\langle \sin,\cos\big\rangle-\mu\|\cos\|^2=-\mu\pi.$$
Conclusion : $\lambda=-2$ et $\mu=0$, donc : $p(\mathrm{Id})=-2\sin$.

Exemple Soient $E\neq\{0_E\}$ un espace euclidien et H un hyperplan de E de vecteur normal a. On note p la projection orthogonale sur H. Alors pour tout $x\in E$: $p(x)=x-\dfrac{\big\langle x,a\big\rangle\,a}{\|a\|^2}$. Si de plus a est unitaire, cette expression devient : $p(x)=x-\big\langle x,a\big\rangle a$. Tout simplement, **PROJETER x SUR H REVIENT À LUI ÔTER SA COMPOSANTE SELON $H^\perp$.**



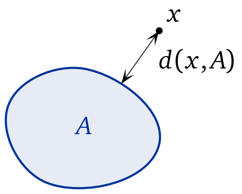
Démonstration Notons p' la projection orthogonale sur $\mathrm{Vect}(a)=H^\perp$.
Comme $\left(\dfrac{a}{\|a\|}\right)$ est une base orthonormale de $\mathrm{Vect}(a)$: $p'(x)=\left\langle x,\dfrac{a}{\|a\|}\right\rangle\dfrac{a}{\|a\|}=\dfrac{\big\langle x,a\big\rangle\,a}{\|a\|^2}$ pour tout $x\in E$, d'où l'expression de p puisque : $p+p'=\mathrm{Id}_E$.

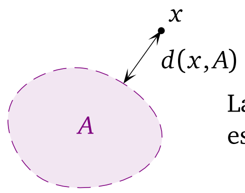
3.2 Distance à un sous-espace vectoriel de dimension finie

Définition-théorème (Distance à une partie) Soient E un espace préhilbertien réel, A une partie non vide de E et $x\in E$. On appelle *distance de x à A*, notée $d(x,A)$, le réel : $d(x,A)=\inf\limits_{a\in A}d(x,a)$.

✌ **Explication** ✌ Intuitivement, la distance d'un vecteur x à une partie A est la plus petite distance séparant x d'un élément de A. Mais comment savoir si une telle « plus petite distance » existe ? En fait, elle n'existe pas nécessairement et c'est pourquoi on n'a surtout pas posé : $d(x,A)=\min\limits_{a\in A}d(x,a)$, on a utilisé une borne inférieure.

La distance de x à A est ici un minimum (i.e. elle est atteinte).





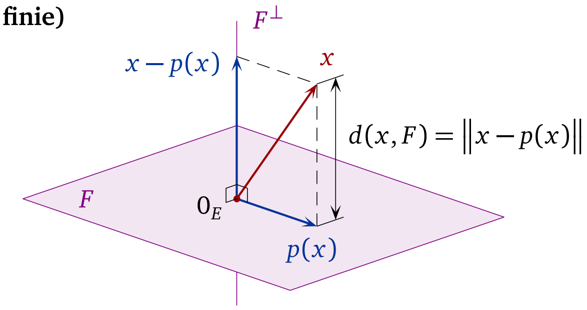
La distance de x à A est ici seulement une borne inférieure.

Démonstration Le réel $d(x,A)$ est bien défini d'après la propriété de la borne inférieure, car l'ensemble $\big\{d(x,a)\big\}_{a\in A}$ est une partie de $\mathbb{R}$ non vide car : $A \neq \varnothing$, et minorée par 0. ■

Théorème (Distance à un sous-espace vectoriel de dimension finie) Soient E un espace préhilbertien réel, F un sous-espace vectoriel DE DIMENSION FINIE de E et $x \in E$. On note p la projection orthogonale sur F.

- Tout d'abord : $d(x,F) = \big\|x-p(x)\big\| = d\big(x,p(x)\big)$.

  La distance de x à F est donc un minimum. Ce minimum n'est atteint qu'en le projeté orthogonal de x sur F.
- Par ailleurs : $d(x,F)^2 = \|x\|^2 - \big\|p(x)\big\|^2$.



Démonstration

- Pour tout $f \in F$: $x - f = \underbrace{\big(x-p(x)\big)}^{\in \operatorname{Ker} p = F^\perp} + \underbrace{\big(p(x)-f\big)}^{\in \operatorname{Im} p = F}$, donc aussitôt d'après le théorème de Pythagore : $\|x-f\|^2 = \big\|x-p(x)\big\|^2 + \big\|p(x)-f\big\|^2$.
- Montrons à présent l'égalité : $d(x,F) = \big\|x-p(x)\big\| = d\big(x,p(x)\big)$. Posons : $\mathscr{D} = \big\{d(x,f)\big\}_{f\in F}$.
  — Pour commencer : $\big\|x-p(x)\big\| = d\big(x,p(x)\big) \in \mathscr{D}$ car $p(x) \in F$.
  — Ensuite, pour tout $f \in F$: $d(x,f) = \|x-f\| = \sqrt{\big\|x-p(x)\big\|^2 + \big\|p(x)-f\big\|^2} \geqslant \big\|x-p(x)\big\|$, donc $\big\|x-p(x)\big\|$ minore $\mathscr{D}$.

  Conclusion : $d\big(x,p(x)\big) = \min \mathscr{D}$, donc a fortiori : $d\big(x,p(x)\big) = \inf \mathscr{D} = d(x,F)$.
- Enfin, pour tout $f \in F \setminus \big\{p(x)\big\}$: $\big\|p(x)-f\big\| > 0$, donc :

$$d(x,f) = \sqrt{\big\|x-p(x)\big\|^2 + \big\|p(x)-f\big\|^2} > \big\|x-p(x)\big\|.$$

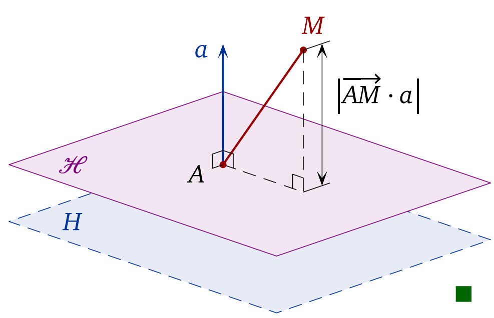
  Comme voulu, la distance $d(x,F)$ n'est atteinte qu'en $p(x)$. ■

Théorème (Distance à un hyperplan affine) Soient E un espace euclidien, $\mathscr{H}$ un hyperplan affine passant par A et de vecteur normal UNITAIRE a. Pour tout $M \in E$: $d(M,\mathscr{H}) = \big|\overrightarrow{AM}\cdot a\big|$.

Démonstration Notons H la direction de $\mathscr{H}$ et p la projection orthogonale sur H. Nous avons vu dans un exemple précédent que pour tout $x \in E$: $p(x) = x - \langle x,a\rangle a$. Du coup :

$$\begin{aligned} d(M,\mathscr{H}) &= d(M, A+H) = \inf_{h\in H} \big\|M-(A+h)\big\| \\ &= d\big(\overrightarrow{AM},H\big) = \Big\|\overrightarrow{AM} - p\big(\overrightarrow{AM}\big)\Big\| = \Big\|\big\langle \overrightarrow{AM},a\big\rangle a\Big\|, \end{aligned}$$

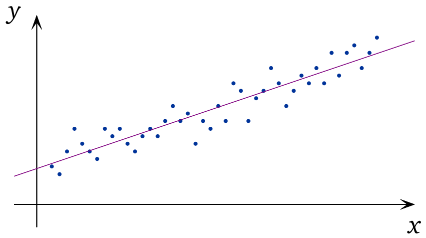
donc comme voulu : $d(M,\mathscr{H}) = \Big|\big\langle\overrightarrow{AM},a\big\rangle\Big|$. ■



Exemple Dans l'espace euclidien canonique $\mathbb{R}^3$, le point $(4,3,2)$ est à distance 4 du plan d'équation : $2x+y+2z=3$.

Démonstration Le plan $\mathscr{P}$ d'équation : $2x+y+2z=3$ est un hyperplan affine de $\mathbb{R}^3$ passant par $A=(1,1,0)$ et de vecteur normal unitaire $a = \dfrac{(2,1,2)}{3}$. Comme voulu, si on pose $M=(4,3,2)$: $\overrightarrow{AM} = (3,2,2)$, donc : $d(M,\mathscr{P}) = \big|\overrightarrow{AM}\cdot a\big| = 4$.

Exemple



- Il est courant qu'on ait à expliquer — en physique, en économie ou dans n'importe quelle discipline expérimentale — une certaine quantité y par une autre quantité x. En vue de cette explication, supposons qu'on ait fait n mesures expérimentales $(x_1, y_1), \ldots, (x_n, y_n)$ du couple (x, y). Le monde étant bien fait, il arrive souvent que la variable y, dite *expliquée*, soit une fonction affine de la variable x, dite *explicative*. Le nuage des couples $(x_1, y_1), \ldots, (x_n, y_n)$ est alors assez proche d'une droite.

  Cela dit, comment déduit-on proprement d'un nuage de points l'équation d'une *droite de meilleure approximation* comme celle que nous avons représentée ci-dessus ? Nous cherchons deux réels m et p pour lesquels la droite d'équation : $y = mx + p$ est la plus proche possible du nuage de points. Ce problème est appelé un problème de *régression linéaire simple* — *linéaire* en raison de la forme : $y = mx+p$ cherchée, *simple* parce que y est supposée ne dépendre que d'une seule variable explicative x.

- Pour tout $i \in [\![1, n]\!]$, notre mesure de y pour $x = x_i$ nous a fourni la valeur y_i, mais la valeur « sans erreur » que nous aurions dû trouver est $mx_i + p$. L'écart entre les deux vaut $\left|y_i - mx_i - p\right|$, mais ce n'est pas cet écart ponctuel qui nous intéresse, nous nous intéressons plutôt à un écart global entre le nuage de points et la droite d'équation : $y = mx+p$. Or comment définir cet écart ? Plusieurs définitions sont possibles, par exemple : $\max\limits_{1\leqslant i\leqslant n} \left|y_i - mx_i - p\right|$, ou : $\displaystyle\sum_{i=1}^n \left|y_i - mx_i - p\right|$, ou : $\displaystyle\sqrt{\sum_{i=1}^n (y_i - mx_i - p)^2}$. Nous travaillerons désormais dans le cadre de cette troisième possibilité. La méthode de régression linéaire correspondante est appelée la *méthode des moindres carrés*.

- Nous cherchons donc des réels m et p — s'il en existe — pour lesquels la quantité $\displaystyle\sqrt{\sum_{i=1}^n (y_i - mx_i - p)^2}$ est minimale, et c'est précisément maintenant que les produits scalaires entrent en scène. Dans l'espace euclidien canonique $\mathbb{R}^n$, posons : $X = (x_1, \ldots, x_n)$, $Y = (y_1, \ldots, y_n)$, $U = \left(\dfrac{1}{n}, \ldots, \dfrac{1}{n}\right)$ et $F = \mathrm{Vect}(X, U)$.

$$\inf_{m,p\in\mathbb{R}} \sqrt{\sum_{i=1}^n (y_i - mx_i - p)^2} = \inf_{m,p\in\mathbb{R}} \left\|Y - mX - npU\right\| = \inf_{m,p\in\mathbb{R}} d(Y, mX + npU) = \inf_{Z\in F} d(Y, Z) = d(Y, F).$$

  Or d'après le théorème précédent, si nous notons p la projection orthogonale de $\mathbb{R}^n$ sur F, la distance $d(Y, F)$ est atteinte en un et un seul point, à savoir : $p(Y) = mX + npU$ où m et p sont les deux réels que nous cherchons.

- Par définition de $p(Y)$: $Y - p(Y) \in F^\perp$, donc : $\big\langle X, Y - p(Y)\big\rangle = 0$ et $\big\langle U, Y - p(Y)\big\rangle = 0$, ce qu'on peut aussi écrire ainsi, sachant que $\|U\|^2 = \dfrac{1}{n}$: $m\|X\|^2 + np\big\langle X, U\big\rangle = \big\langle X, Y\big\rangle$ et $m\big\langle X, U\big\rangle + p = \big\langle Y, U\big\rangle$ — deux équations, deux inconnues. En particulier : $m = \dfrac{\big\langle X, Y\big\rangle - n\big\langle X, U\big\rangle\big\langle Y, U\big\rangle}{\|X\|^2 - n\big\langle X, U\big\rangle^2}$ et $p = \big\langle Y, U\big\rangle - m\big\langle X, U\big\rangle$. Posons alors :

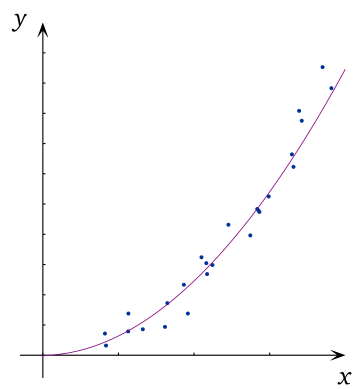
$$\mathrm{E}(x) = \big\langle X, U\big\rangle = \frac{1}{n}\sum_{i=1}^n x_i \quad (\textit{moyenne empirique des } x_i), \qquad \mathrm{E}(y) = \big\langle Y, U\big\rangle = \frac{1}{n}\sum_{i=1}^n y_i \quad (\textit{moyenne empirique des } y_i),$$

$$\mathrm{V}(x) = \frac{1}{n}\left\|X - n\mathrm{E}(x)U\right\|^2 = \frac{1}{n}\sum_{i=1}^n \big(x_i - \mathrm{E}(x)\big)^2 \quad (\textit{variance empirique des } x_i)$$

$$\text{et} \quad \mathrm{cov}(x, y) = \frac{1}{n}\big\langle X - n\mathrm{E}(x)U, Y - n\mathrm{E}(y)U\big\rangle = \frac{1}{n}\sum_{i=1}^n \big(x_i - \mathrm{E}(x)\big)\big(y_i - \mathrm{E}(y)\big) \quad (\textit{covariance empirique des } x_i \textit{ et des } y_i).$$

  Il n'est alors pas dur de vérifier que : $n\mathrm{V}(x) = \|X\|^2 - n\big\langle X, U\big\rangle^2$ et $n\,\mathrm{cov}(x, y) = \big\langle X, Y\big\rangle - n\big\langle X, U\big\rangle\big\langle Y, U\big\rangle$.

  Conclusion : $\boxed{m = \dfrac{\mathrm{cov}(x, y)}{\mathrm{V}(x)} \quad \text{et} \quad p = \mathrm{E}(y) - m\mathrm{E}(x).}$



- Le problème de ce qui précède, c'est que toute variable expliquée y ne dépend pas de manière affine de sa variable explicative x. La méthode des moindres carrés est-elle caduque au-delà ? Heureusement non, et nous allons nous en convaincre sur un exemple.

  Sur la figure ci-contre, on peut émettre l'hypothèse d'une relation de la forme : $y = \lambda x^\alpha$ entre x et y où λ et α sont deux réels à déterminer. On peut se ramener ici au cas affine en réécrivant les choses ainsi : $\ln y = \alpha \ln x + \ln \lambda$. Dans ce cas simple, ce sont les réels α et $\ln\lambda$ qu'on calculera par régression linéaire simple.