# 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)**

- **Produit scalaire :** 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 $\langle x|y\rangle$, $(x|y)$ ou bien sûr $x\cdot y$.
- **Espace préhilbertien réel, espace euclidien :** 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*.

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

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 (Produits scalaires canoniques sur $\mathbb{R}^n$ et $\mathscr{M}_{n,p}(\mathbb{R})$)**

(i) L'application $(X,Y)\longmapsto X^\top Y = \sum_{k=1}^n x_k y_k$ est un produit scalaire sur $\mathbb{R}^n$ appelé son *produit scalaire canonique*.

(ii) L'application $(A,B)\longmapsto \mathrm{tr}\big(A^\top B\big) = \sum_{\substack{1\leqslant i\leqslant n\\ 1\leqslant j\leqslant p}} a_{ij}b_{ij}$ est un produit scalaire sur $\mathscr{M}_{n,p}(\mathbb{R})$ appelé son *produit scalaire canonique*.

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

Pour tous $X,Y\in\mathbb{R}^n$, le produit matriciel $X^\top Y$ est une matrice carrée de taille 1, i.e. un réel. Plus généralement, $A^\top B$ est une matrice carrée de taille $p$ pour tous $A,B\in\mathscr{M}_{n,p}(\mathbb{R})$, que l'on convertit en un réel grâce à la trace.

Retenez bien que le produit scalaire canonique de $\mathbb{R}^n$ peut être calculé matriciellement.

**Démonstration** Nous nous contenterons de démontrer l'assertion (ii), car comme on vient de le remarquer, le produit scalaire canonique de $\mathbb{R}^n$ n'est jamais que le produit scalaire canonique de $\mathscr{M}_{n,1}(\mathbb{R})$. Pour commencer, pour tous $A,B\in\mathscr{M}_{n,p}(\mathbb{R})$ : $A^\top B\in\mathscr{M}_p(\mathbb{R})$, donc $\mathrm{tr}\big(A^\top B\big) = \sum_{j=1}^p \big(A^\top B\big)_{jj} = \sum_{j=1}^p\sum_{i=1}^n \big(A^\top\big)_{ji} b_{ij} = \sum_{i,j} a_{ij}b_{ij}$.

- **Symétrie :** Pour tous $A,B\in\mathscr{M}_{n,p}(\mathbb{R})$ : $\mathrm{tr}\big(A^\top B\big) = \mathrm{tr}\Big(\big(A^\top B\big)^\top\Big) = \mathrm{tr}\big(B^\top A\big)$.
- **Bilinéarité :** Par symétrie, la linéarité par rapport à la première variable suffit. Pour tous $A,B,C\in\mathscr{M}_n(\mathbb{R})$ et $\lambda\in\mathbb{R}$, par bilinéarité du produit matriciel et linéarité de la trace :
$$\mathrm{tr}\Big(A^\top(\lambda B+C)\Big) = \mathrm{tr}\Big(\lambda\big(A^\top B\big)+\big(A^\top C\big)\Big) = \lambda\,\mathrm{tr}\big(A^\top B\big)+\mathrm{tr}\big(A^\top C\big).$$
- **Positivité et séparation :** Pour tout $A\in\mathscr{M}_{n,p}(\mathbb{R})$ : $\mathrm{tr}\big(A^\top A\big) = \sum_{i,j} a_{ij}^2 \geqslant 0$, et si $\mathrm{tr}\big(A^\top A\big)=0$, alors $a_{ij}=0$ pour tous $i,j\in[\![1,n]\!]$ par positivité des réels sommés.

**Exemple** De nombreux produits scalaires peuvent exister sur un même espace vectoriel. L'application $(X,Y) \longmapsto X^\top \begin{pmatrix} 2 & 1 \\ 1 & 2 \end{pmatrix} Y$ est par exemple un produit scalaire sur $\mathbb{R}^2$ distinct du produit scalaire canonique.

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

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

et si $X^\top \begin{pmatrix} 2 & 1 \\ 1 & 2 \end{pmatrix} X = 0$, alors $x = y = x+y = 0$ car $x^2$, $y^2$ et $(x+y)^2$ sont positifs, 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. Ensuite, 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 $\int_a^b f(t)^2\,\mathrm{d}t = 0$, $f$ étant CONTINUE et POSITIVE, alors $f^2$ est nulle sur $[a,b]$, 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 $\sum_{k=0}^n P(x_k)^2 = 0$, alors $P(x_k) = 0$ pour tout $k \in [\![0,n]\!]$ par positivité des réels sommés, i.e. $x_0, \ldots, x_n$ sont des racines de $P$. Le polynôme $P$ de degré inférieur ou égal à $n$ possède ainsi $n+1$ racines distinctes, donc est nul.

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

- **Norme :** On appelle *norme 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 $\|x\| = 1$.
- **Distance :** Pour tous $x, y \in E$, le réel $\|x - y\|$, souvent noté $d(x,y)$, est appelé la *distance entre $x$ et $y$ (associée au produit scalaire $\langle \cdot, \cdot \rangle$)*.

**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 $(X,Y) \longmapsto X^\top \begin{pmatrix} 2 & 1 \\ 1 & 2 \end{pmatrix} Y$ sur $\mathbb{R}^2$ : $\big\|(1,0)\big\| = \sqrt{2} \neq 1$.

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 : $\boxed{\langle x,y \rangle = \frac{1}{2}\Big(\|x+y\|^2 - \|x\|^2 - \|y\|^2\Big) = \frac{1}{4}\Big(\|x+y\|^2 - \|x-y\|^2\Big).}$

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

(i) **Inégalité de Cauchy-Schwarz :** $\big|\langle x,y \rangle\big| \leqslant \|x\|.\|y\|$, 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.

Si nous avions défini le produit scalaire à partir de normes et d'angles, l'inégalité de Cauchy-Schwarz serait une pure trivialité : $\big|\vec{u} \cdot \vec{v}\big| = \big\|\vec{u}\big\| \times \big\|\vec{v}\big\| \times \big|\cos\big(\vec{u}, \vec{v}\big)\big| \leqslant \big\|\vec{u}\big\|.\big\|\vec{v}\big\|$, mais dans le contexte de ce chapitre, cette inégalité est remarquable justement parce que nous n'avons pas encore de définition propre des angles.

Dans certains contextes, il est utile de disposer d'une inégalité de Cauchy-Schwarz un peu plus souple que la précédente.

**Théorème (Inégalité de Cauchy-Schwarz généralisée)** Soient $E$ un $\mathbb{R}$-espace vectoriel, $\varphi$ une forme bilinéaire symétrique positive — MAIS PAS FORCÉMENT DÉFINIE POSITIVE — et $x, y \in E$. Alors : $\big|\varphi(x,y)\big| \leqslant \sqrt{\varphi(x,x)}\sqrt{\varphi(y,y)}$.

**Démonstration**

(i) **Inégalité généralisée :** Pour tout $t \in \mathbb{R}$ : $\overbrace{\varphi(x+ty,x+ty)}^{\geqslant 0} = \overbrace{\varphi(x,x)}^{\geqslant 0} + 2t\,\varphi(x,y) + t^2\overbrace{\varphi(y,y)}^{\geqslant 0}$.

— Si $\varphi(y,y) = 0$, la fonction $t \longmapsto \varphi(x+ty,x+ty) = \varphi(x,x) + 2t\,\varphi(x,y)$ est affine et positive sur $\mathbb{R}$ tout entier, donc son coefficient directeur $\varphi(x,y)$ est nul et c'est fini.

— Si $\varphi(y,y) \neq 0$, la fonction $t \longmapsto \varphi(x+ty,x+ty)$ est polynomiale de degré EXACTEMENT 2 et positive sur $\mathbb{R}$ tout entier, donc son discriminant est négatif : $\varphi(x,y)^2 - \varphi(x,x)\,\varphi(y,y) \leqslant 0$ et c'est fini.

**Cas d'égalité dans le cas préhilbertien :** Le cas d'égalité est trivial si $y = 0_E$. Dans le cas contraire, en reprenant la preuve qui précède, l'inégalité est une égalité si et seulement si le discriminant calculé est nul, i.e. si et seulement si la fonction $t \longmapsto \|x+ty\|^2$ s'annule. Or cette annulation est équivalente à l'existence d'un réel $t_0 \in \mathbb{R}$ pour lequel $x + t_0 y = 0_E$, i.e. à la colinéarité de $x$ et $y$.

(ii) D'abord : $\|x+y\|^2 = \|x\|^2 + 2\langle x,y\rangle + \|y\|^2 \overset{(i)}{\leqslant} \|x\|^2 + 2\|x\|.\|y\| + \|y\|^2 = \big(\|x\| + \|y\|\big)^2$, 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 : $\|x\| = \big\|(x+y)+(-y)\big\| \leqslant \|x+y\| + \|-y\| = \|x+y\| + \|y\|$, donc $\|x\| - \|y\| \leqslant \|x+y\|$, et de même $\|y\| - \|x\| \leqslant \|x+y\|$.

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

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

**Exemple** Pour toute fonction $f \in \mathscr{C}^1\big([0,1],\mathbb{R}\big)$ pour laquelle $f(0) = 0$ : $f(1)^2 \leqslant 2\sqrt{\int_0^1 f(t)^2\,dt}\sqrt{\int_0^1 f'(t)^2\,dt}$.

**Démonstration** Appliquons l'inégalité de Cauchy-Schwarz à $f$ et $f'$ dans l'espace préhilbertien réel $\mathscr{C}\big([0,1],\mathbb{R}\big)$ muni du produit scalaire $(u,v) \longmapsto \int_0^1 u(t)v(t)\,dt$ : $\left|\int_0^1 f(t)f'(t)\,dt\right| \leqslant \sqrt{\int_0^1 f(t)^2\,dt}\sqrt{\int_0^1 f'(t)^2\,dt}$. On conclut en calculant l'intégrale de gauche, qui vaut $\frac{1}{2}\big(f(1)^2 - f(0)^2\big) = \frac{1}{2}f(1)^2$.

**Théorème (Inégalité de Cauchy-Schwarz pour les variables aléatoires)** Soient $(\Omega, P)$ un espace probabilisé fini.

(i) L'application $(X,Y) \longmapsto \mathrm{E}(XY)$ est une forme bilinéaire symétrique positive sur $\mathbb{R}^\Omega$, MAIS N'EST PAS FORCÉMENT UN PRODUIT SCALAIRE. Elle en est un si et seulement si $P\big(\{\omega\}\big) > 0$ pour tout $\omega \in \Omega$.

(ii) **Inégalité de Cauchy-Schwarz :** Soient $X$ et $Y$ deux variables aléatoires sur $\Omega$. Alors :

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

**Démonstration**

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

- Si $(X,Y) \longmapsto \mathrm{E}(XY)$ est un produit scalaire, alors pour tout $\omega \in \Omega$ : $P\big(\{\omega\}\big) = \mathrm{E}\big(\mathbb{1}_{\{\omega\}}\big) = \mathrm{E}\big(\mathbb{1}_{\{\omega\}}^2\big) > 0$.
- Faisons l'hypothèse que $P\big(\{\omega\}\big) > 0$ pour tout $\omega \in \Omega$. Pour toute variable aléatoire $X \in \mathbb{R}^\Omega$ pour laquelle $\mathrm{E}\big(X^2\big) = 0$ : $\sum_{\omega\in\Omega} P\big(\{\omega\}\big)\,X(\omega)^2 = 0$, donc $X(\omega) = 0$ pour tout $\omega \in \Omega$, donc $X = 0$.

(ii) La première inégalité est une simple application de l'inégalité de Cauchy-Schwarz généralisée. Ensuite :
$$\left|\mathrm{cov}(X,Y)\right| = \left|\mathrm{E}\Big(\big(X-\mathrm{E}(X)\big)\big(Y-\mathrm{E}(Y)\big)\Big)\right| \leqslant \sqrt{\mathrm{E}\Big(\big(X-\mathrm{E}(X)\big)^2\Big)}\,\sqrt{\mathrm{E}\Big(\big(Y-\mathrm{E}(Y)\big)^2\Big)} = \sigma(X)\sigma(Y).$$

**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*, ce qu'on note $x\perp y$, si $\langle x,y\rangle = 0$,

— les parties $X$ et $Y$ sont *orthogonales*, ce qu'on note $X\perp Y$, si pour tous $x\in X$ et $y\in Y$ : $\langle x,y\rangle = 0$,

— la famille $(x_i)_{i\in I}$ est *orthogonale* si pour tous $i,j\in I$ distincts : $\langle x_i,x_j\rangle = 0$,

— 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 $i,j\in I$ : $\langle x_i,x_j\rangle = \delta_{ij}$.

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 scalaire est associée une notion d'orthogonalité, ce qui fait que les angles droits ne sont pas droits absolument, mais relativement.

Le petit résultat suivant est à la fois trivial et essentiel.

La propriété de séparation énonce que LE VECTEUR NUL EST LE SEUL VECTEUR ORTHOGONAL À LUI-MÊME.
En particulier, seul le vecteur nul est orthogonal à tout vecteur.

**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 $E_i^\top E_j = \delta_{ij}$ pour tous $i,j\in[\![1,n]\!]$.

**Exemple** La base canonique de $\mathbb{R}^2$ n'est pas orthonormale pour le produit scalaire $(X,Y)\longmapsto X^\top\begin{pmatrix}2&1\\1&2\end{pmatrix}Y$ sur $\mathbb{R}^2$, mais la famille $\left(\dfrac{(1,0)}{\sqrt{2}},\dfrac{(1,-2)}{\sqrt{6}}\right)$ l'est comme on le vérifie aisément.

**Exemple** Pour le produit scalaire canonique de $\mathscr{M}_{n,p}(\mathbb{R})$, la base canonique de $\mathscr{M}_{n,p}(\mathbb{R})$ est orthonormale.

**Démonstration** Pour tous $i\in[\![1,n]\!]$ et $j\in[\![1,p]\!]$, notons $E_{ij}$ la matrice de $\mathscr{M}_{n,p}(\mathbb{R})$ dont tous les coefficients sont nuls à l'exception du coefficient de position $(i,j)$, égal à 1. Fixons $i,k\in[\![1,n]\!]$ et $j,l\in[\![1,p]\!]$. Seules la $l^{\text{ème}}$ colonne et la $j^{\text{ème}}$ ligne de la matrice $E_{ij}^\top E_{kl}$ sont éventuellement non nulles. Cela revient à dire que tous les coefficients de cette matrice sont nuls, sauf peut-être son coefficient de position $(j,l)$. Et que vaut-il ? Réponse : $\delta_{ik}$, donc $E_{ij}^\top E_{kl} = \delta_{ik}E_{jl}$. CE CALCUL EST IMPORTANT ET MÉRITE QUE VOUS LE REFASSIEZ SEULS. Finalement : $\big\langle E_{ij},E_{kl}\big\rangle = \mathrm{tr}\big(E_{ij}^\top E_{kl}\big) = \mathrm{tr}\big(\delta_{ik}E_{jl}\big) = \delta_{ik}\delta_{jl} = \delta_{(i,j),(k,l)}$.

**Exemple** La famille des fonctions $t\xrightarrow{s_n}\sin(nt)$, $n$ 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 tous $m, n \in \mathbb{N}^*$ :

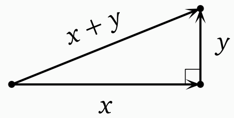
$$\langle s_m, s_n\rangle = \begin{cases} \|s_n\|^2 = \dfrac{1}{\pi}\displaystyle\int_0^{2\pi} \dfrac{1-\cos(2nt)}{2}\,\mathrm{d}t = 1 - \dfrac{1}{\pi}\left[\dfrac{\sin(2nt)}{4n}\right]_{t=0}^{t=2\pi} = 1 & \text{si } m = n \\ \dfrac{1}{2\pi}\displaystyle\int_0^{2\pi}\Big(\cos\big((m-n)t\big) - \cos\big((m+n)t\big)\Big)\,\mathrm{d}t = \dfrac{1}{2\pi}\left[\dfrac{\sin\big((m-n)t\big)}{m-n} - \dfrac{\sin\big((m+n)t\big)}{m+n}\right]_{t=0}^{t=2\pi} = 0 & \text{si } m \neq n. \end{cases}$$

**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 : $\langle p, i\rangle = \displaystyle\int_{-1}^1 \overbrace{p(t)\,i(t)}^{\text{Impaire}}\,\mathrm{d}t = 0$.

**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 $E$ : $\displaystyle\left\|\sum_{i=1}^n x_i\right\|^2 = \sum_{i=1}^n \|x_i\|^2$.

(ii) **Liberté :** Toute famille orthogonale de vecteurs NON NULS de $E$ est libre. En particulier, toute famille orthonormale de $E$ est libre.

Aviez-vous compris avant que le théorème de Pythagore n'est qu'un simple commentaire de l'identité remarquable :

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

**Démonstration**

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

(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 $i \in [\![1,n]\!]$ : $\quad 0 = \langle 0_E, x_i\rangle = \left\langle \displaystyle\sum_{k=1}^n \lambda_k x_k\ ,\ x_i\right\rangle = \displaystyle\sum_{k=1}^n \lambda_k \langle x_k, x_i\rangle = \lambda_i \|x_i\|^2$, or $\|x_i\| \neq 0$ par séparation, donc $\lambda_i = 0$. La famille $(x_1, \ldots, x_n)$ est ainsi libre.

## 2.2 Composante selon un vecteur et coordonnées dans une base orthonormale

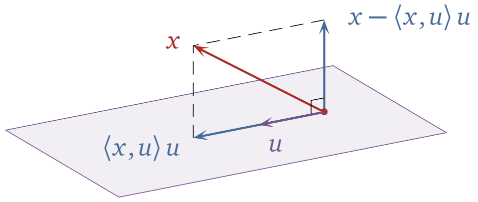
**Définition-théorème (Composante selon un vecteur)** Soient $E$ un espace préhilbertien réel et $x \in E$.

(ii) **Cas d'un vecteur unitaire :** Soit $u \in E$ UNITAIRE. Le vecteur $x - \langle x, u\rangle\, u$ est orthogonal à $u$ et $\langle x, u\rangle\, u$ est appelé la *composante de $x$ selon $u$*.

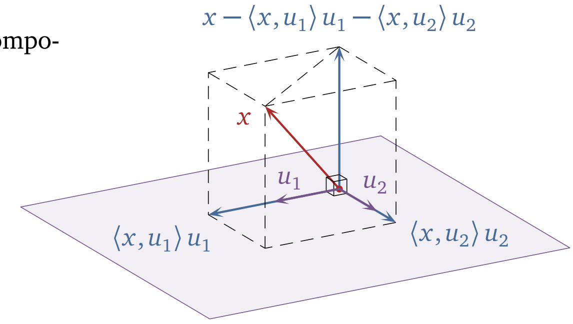
**Cas général :** Soit $a \in E$ NON NUL. La composante de $x$ selon $\dfrac{a}{\|a\|}$ est aussi appelé sa *composante selon $a$*. Elle vaut $\left\langle x, \dfrac{a}{\|a\|}\right\rangle \dfrac{a}{\|a\|} = \dfrac{\langle x, a\rangle\, a}{\|a\|^2}$.

(i) **Et si on retire plein de composantes ?** Soit $(u_1, \ldots, u_n)$ une famille ORTHONORMALE de $E$.

Le vecteur $x - \displaystyle\sum_{k=1}^n \langle x, u_k\rangle\, u_k$ est orthogonal à $u_1, \ldots, u_n$.




Dans l'assertion (i), $x$ n'a aucune raison d'être orthogonal à $u$, mais si on le redresse correctement en lui retirant sa composante selon $u$, il le devient.




Dans l'assertion (ii), on rend $x$ orthogonal à $u_1,\dots,u_n$ en lui ôtant ses composantes selon $u_1,\dots,u_n$.

**Démonstration**

(i) $\Big\langle x-\langle x,u\rangle\,u\ ,\ u\Big\rangle=\langle x,u\rangle-\langle x,u\rangle\,\|u\|^2=0.$

(ii) Pour tout $i\in[\![1,n]\!]$ :

$$\left\langle x-\sum_{k=1}^n\langle x,u_k\rangle\,u_k\ ,\ u_i\right\rangle=\langle x,u_i\rangle-\sum_{k=1}^n\langle x,u_k\rangle\,\langle u_k,u_i\rangle$$
$$=\langle x,u_i\rangle-\sum_{k=1}^n\langle x,u_k\rangle\,\delta_{ki}=\langle x,u_i\rangle-\langle x,u_i\rangle=0.$$

Nous démontrerons bientôt que tout espace euclidien possède une base orthonormale, mais en attendant, quelques remarques simples sur les coordonnées dans une telle base.

**Théorème (Coordonnées dans une base orthonormale et expression du produit scalaire et de la norme)** Soient $E$ un espace euclidien, $(e_1,\dots,e_n)$ une base ORTHONORMALE de $E$ et $x,y\in E$.

(i) **Coordonnées :** $\displaystyle x=\sum_{k=1}^n\langle x,e_k\rangle\,e_k$. Ainsi, les coordonnées de $x$ dans $(e_1,\dots,e_n)$ sont $\big(\langle x,e_1\rangle,\dots,\langle x,e_n\rangle\big)$.

(ii) **Produit scalaire et norme :** En notant $X=(x_1,\dots,x_n)$ et $Y=(y_1,\dots,y_n)$ les coordonnées respectives de $x$ et $y$ dans $(e_1,\dots,e_n)$ :

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

Dans l'assertion (i), attention de ne pas confondre les coordonnées de $x$ dans $(e_1,\dots,e_n)$, qui sont des scalaires, et ses composantes selon $e_1,\dots,e_n$, qui sont des vecteurs.

L'assertion (ii) montre que le produit scalaire canonique sur $\mathbb{R}^n$ est l'archétype de 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 !** Tout ceci est très faux si la base n'est pas ORTHONORMALE !

**Démonstration**

(i) D'après le théorème précédent, $\displaystyle x-\sum_{k=1}^n\langle x,e_k\rangle\,e_k$ est orthogonal à $e_1,\dots,e_n$ donc à tout vecteur de $E$, donc est nul par séparation.

On peut aussi voir les choses autrement. En notant $(x_1,\dots,x_n)$ les coordonnées de $x$ dans $(e_1,\dots,e_n)$, pour tout $k\in[\![1,n]\!]$ : $\displaystyle\langle x,e_k\rangle=\left\langle\sum_{i=1}^n x_ie_i\ ,\ e_k\right\rangle=\sum_{i=1}^n x_i\,\langle e_i,e_k\rangle=\sum_{i=1}^n x_i\delta_{ki}=x_k.$

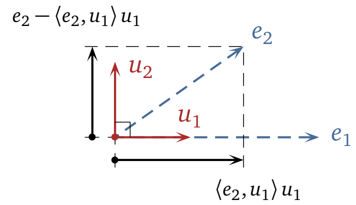
(ii) $$\langle x,y\rangle=\left\langle\sum_{i=1}^n x_ie_i\ ,\ \sum_{j=1}^n y_je_j\right\rangle=\sum_{1\leqslant i,j\leqslant n}x_iy_j\,\langle e_i,e_j\rangle=\sum_{1\leqslant i,j\leqslant n}x_iy_j\delta_{ij}=\sum_{k=1}^n x_ky_k=X^\top Y.$$

## 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 $E$. Il existe alors une famille orthonormale $(u_1,\dots,u_n)$ de $E$ pour laquelle pour tout $k\in[\![1,n]\!]$ : $\mathrm{Vect}(u_1,\dots,u_k)=\mathrm{Vect}(e_1,\dots,e_k)$.

Les vecteurs $u_1,\dots,u_n$ peuvent être construits de proche en proche depuis $u_1$ jusqu'à $u_n$. Pour tout $k\in[\![1,n]\!]$, si on a construit $u_1,\dots,u_{k-1}$ et si on pose $\displaystyle\widehat{u}_k=e_k-\sum_{i=1}^{k-1}\langle e_k,u_i\rangle\,u_i$, $u_k$ est soit le vecteur $\dfrac{\widehat{u}_k}{\|\widehat{u}_k\|}$, soit son opposé.

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 $u_2=\dfrac{e_2-\langle e_2,u_1\rangle\,u_1}{\big\|e_2-\langle e_2,u_1\rangle\,u_1\big\|}$.




— La première formule normalise $e_1$, c'est-à-dire le rend unitaire.

— La seconde redresse $e_2$ en lui retranchant sa composante $\langle e_2, u_1 \rangle u_1$ selon $u_1$ pour le rendre orthogonal à $u_1$. Le vecteur obtenu n'a aucune raison d'être unitaire, c'est pourquoi on le divise par sa norme.

Plus généralement, l'algorithme de Gram-Schmidt construit $u_k$ en redressant $e_k$, i.e. en le rendant orthogonal à $u_1, \ldots, u_{k-1}$, puis en le rendant unitaire.

**Démonstration**

- Montrons par récurrence que pour tout $k \in [\![1, n]\!]$, il existe une famille orthonormale $(u_1, \ldots, u_k)$ de $E$ pour laquelle pour tout $i \in [\![1, k]\!]$ : $\mathrm{Vect}(u_1, \ldots, u_i) = \mathrm{Vect}(e_1, \ldots, e_i)$.

  **Initialisation :** La famille $(e_1)$ est libre, donc $e_1 \neq 0_E$, et si on pose $u_1 = \dfrac{e_1}{\|e_1\|}$, alors $\mathrm{Vect}(e_1) = \mathrm{Vect}(u_1)$.

  **Hérédité :** Soit $k \in [\![2, n]\!]$. Faisons l'hypothèse qu'il existe une famille orthonormale $(u_1, \ldots, u_{k-1})$ de $E$ pour laquelle pour tout $i \in [\![1, k-1]\!]$ : $\mathrm{Vect}(u_1, \ldots, u_i) = \mathrm{Vect}(e_1, \ldots, e_i)$. Nous sommes en quête d'un vecteur $u_k$ pour lequel $(u_1, \ldots, u_k)$ est orthonormale et $\mathrm{Vect}(u_1, \ldots, u_k) = \mathrm{Vect}(e_1, \ldots, e_k)$.

  Le vecteur $\widehat{u}_k = e_k - \displaystyle\sum_{i=1}^{k-1} \langle e_k, u_i \rangle u_i$ est orthogonal à $u_1, \ldots, u_{k-1}$. Se peut-il qu'il soit nul ? La liberté de $(e_1, \ldots, e_n)$ s'en trouverait contredite car $e_k$ appartiendrait à $\mathrm{Vect}(u_1, \ldots, u_{k-1}) = \mathrm{Vect}(e_1, \ldots, e_{k-1})$. Nous pouvons ainsi poser $u_k = \dfrac{\widehat{u}_k}{\|\widehat{u}_k\|}$ — ou choisir son opposé. La famille $(u_1, \ldots, u_k)$ est orthonormale. Enfin : $\mathrm{Vect}(u_1, \ldots, u_{k-1}) = \mathrm{Vect}(e_1, \ldots, e_{k-1})$, et $u_k$ est combinaison linéaire de $u_1, \ldots, u_{k-1}$ et $e_k$ avec un coefficient non nul devant $e_k$, donc $\mathrm{Vect}(u_1, \ldots, u_k) = \mathrm{Vect}(e_1, \ldots, e_k)$.

- Tâchons enfin de comprendre pourquoi, au rang $k$, notre choix de vecteur $u_k$ était le seul possible au signe près. Soit $u_k$ un vecteur pour lequel $(u_1, \ldots, u_k)$ est orthonormale et $\mathrm{Vect}(u_1, \ldots, u_k) = \mathrm{Vect}(e_1, \ldots, e_k)$. Comme $\mathrm{Vect}(u_1, \ldots, u_k) = \mathrm{Vect}(e_1, \ldots, e_k) \overset{\mathrm{HDR}}{=} \mathrm{Vect}(u_1, \ldots, u_{k-1}, e_k)$, $u_k$ est combinaison linéaire de $u_1, \ldots, u_{k-1}, e_k$, autrement dit $u_k = a_1 u_1 + \ldots + a_{k-1} u_{k-1} + a_k e_k$ pour certains $a_1, \ldots, a_k \in \mathbb{R}$. Ainsi, pour tout $i \in [\![1, k-1]\!]$ : $0 = \langle u_k, u_i \rangle = \big\langle a_1 u_1 + \ldots + a_{k-1} u_{k-1} + a_k e_k \,,\, u_i \big\rangle = a_i + a_k \langle e_k, u_i \rangle$, donc $a_i = -a_k \langle e_k, u_i \rangle$, donc $u_k = a_k \left( e_k - \displaystyle\sum_{i=1}^{k-1} \langle e_k, u_i \rangle u_i \right)$, et on conclut en n'oubliant pas que $u_k$ est unitaire. ■

**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 $\mathbb{R}[X]$. Au passage, savez-vous justifier que c'est un produit scalaire ?

**Démonstration** Exploitons l'algorithme de Gram-Schmidt pour construire une famille orthonormale $(U_0, U_1, U_2)$ à partir de la famille LIBRE $\left(1, X, X^2\right)$.

- **Construction de $U_0$ :** $\|1\|^2 = \displaystyle\int_0^1 \mathrm{d}t = 1$, donc on pose $U_0 = \dfrac{1}{\|1\|} = 1$.

- **Construction de $U_1$ :** $\langle X, U_0 \rangle = \displaystyle\int_0^1 t\,\mathrm{d}t = \frac{1}{2}$, donc $X - \langle X, U_0 \rangle U_0 = X - \dfrac{1}{2}$.

  Ensuite : $\left\| X - \dfrac{1}{2} \right\|^2 = \displaystyle\int_0^1 \left( t - \frac{1}{2} \right)^2 \mathrm{d}t = \frac{1}{12}$, donc on pose $U_1 = \dfrac{X - \langle X, U_0 \rangle U_0}{\left\| X - \langle X, U_0 \rangle U_0 \right\|} = \sqrt{3}\,(2X - 1)$.

- **Construction de $U_2$ :** $\left\langle X^2, U_0 \right\rangle = \displaystyle\int_0^1 t^2\,\mathrm{d}t = \frac{1}{3}$ et $\left\langle X^2, U_1 \right\rangle = \displaystyle\int_0^1 t^2 \times \sqrt{3}(2t-1)\,\mathrm{d}t = \frac{1}{2\sqrt{3}}$, donc $X^2 - \left\langle X^2, U_0 \right\rangle U_0 - \left\langle X^2, U_1 \right\rangle U_1 = X^2 - X + \dfrac{1}{6}$. Ensuite : $\left\| X^2 - X + \dfrac{1}{6} \right\|^2 = \displaystyle\int_0^1 \left( t^2 - t + \frac{1}{6} \right)^2 \mathrm{d}t = \frac{1}{180}$, donc on pose $U_2 = \dfrac{X^2 - \left\langle X^2, U_0 \right\rangle U_0 - \left\langle X^2, U_1 \right\rangle U_1}{\left\| X^2 - \langle X^2, U_0 \rangle U_0 - \langle X^2, U_1 \rangle U_1 \right\|} = \sqrt{5}\left(6X^2 - 6X + 1\right)$.

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

**Théorème (Existence de bases orthonormales en dimension finie)**

(i) Tout espace euclidien possède une base orthonormale.

(ii) **Théorème de la base orthonormale incomplète :** Soit $E$ un espace euclidien. Toute famille orthonormale de $E$ peut être complétée en une base orthonormale de $E$.

**Démonstration**

(i) Tout $\mathbb{R}$-espace vectoriel de dimension finie possède une base — éventuellement vide — donc une base orthonormale grâce à l'algorithme de Gram-Schmidt.

(ii) Libre, toute famille orthonormale $\mathscr{F}$ de $E$ peut être complétée en une base de $E$. Or appliqué à cette base, l'algorithme de Gram-Schmidt n'affecte pas les vecteurs de $\mathscr{F}$ et on peut donc considérer que $\mathscr{F}$ a été complétée 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 $X^\perp = \left\{t \in E \mid \quad \forall x \in X, \quad \langle t, x\rangle = 0\right\}$.

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

L'égalité $X^\perp = \mathrm{Vect}(X)^\perp$ signifie que pour déterminer l'orthogonal d'un sous-espace vectoriel $F$, il n'est pas nécessaire d'exiger l'orthogonalité à TOUS les vecteurs de $F$, l'orthogonalité à tout vecteur d'une partie génératrice suffit.

**Attention !** Pour un sous-espace vectoriel $F$ de $E$, $F$ et $F^\perp$ sont TOUJOURS en somme directe. En revanche, il n'est pas vrai en général que $F$ et $F^\perp$ sont supplémentaires dans $E$, ils sont seulement en somme directe. Contrairement aux apparences, il n'est pas non plus vrai en général que $F^{\perp\perp} = F$. Nous y reviendrons.

**Démonstration**

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

(ii) Si $X$ est un sous-espace vectoriel de $E$, $X \cap X^\perp$ contient $0_E$. Inversement, $x \perp x$ pour tout $x \in X \cap X^\perp$, 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 $X^\perp = \mathrm{Vect}(X)^\perp$. 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 $E$ : $\quad \left\{0_E\right\}^\perp = E \quad$ et $\quad E^\perp = \left\{0_E\right\}$. $\quad$ Pour la seconde égalité, rappelons que $0_E$ est le seul vecteur de $E$ orthogonal à tout vecteur.

**Exemple** Dans l'espace euclidien canonique $\mathbb{R}^3$, le plan vectoriel d'équation $3x - y + 2z = 0$ n'est jamais que l'orthogonal $\left\{(3, -1, 2)\right\}^\perp$ car pour tout $(x, y, z) \in \mathbb{R}^3$ : $\quad \left\langle (3, -1, 2), (x, y, z)\right\rangle = 3x - y + 2z$. $\quad$ En termes géométriques, le vecteur $(3, -1, 2)$ est normal au plan d'équation $3x - y + 2z = 0$. Remarque essentielle, nous y reviendrons.

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

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

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

Le point important de cet exemple, c'est que l'orthogonalité à 1 et $X$ a suffi car $\mathbb{R}_1[X] = \mathrm{Vect}(1, X)$.

**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) **Supplémentaire orthogonal :** $F^\perp$ 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 $E$*. Pour résumer, on écrit souvent les choses ainsi : $\quad E = F \overset{\perp}{\oplus} F^\perp$.

En particulier, si $E$ lui-même est de dimension finie : $\quad \dim F^\perp = \dim E - \dim F$.

(ii) **Bi-orthogonal :** $\quad F^{\perp\perp} = F$.

✖ **Attention !** Il est ici essentiel que $F$ soit **DE DIMENSION FINIE**. Par ailleurs, il existe **UN SEUL** supplémentaire **ORTHOGONAL** mais tout plein de supplémentaires en toute généralité.

**Démonstration**

(i) Nous savons déjà que $F \cap F^{\perp} = \{0_E\}$, et bien sûr $F \perp F^{\perp}$. Pour montrer que $E = F \overset{\perp}{\oplus} F^{\perp}$, il nous reste à montrer que $E \subset F + F^{\perp}$. Comme $F$ est de dimension finie, nous pouvons nous en donner une base orthonormale $(f_1, \ldots, f_n)$ — éventuellement vide. Pour tout $x \in E$, le vecteur $x - \langle x, f_1 \rangle f_1 - \ldots - \langle x, f_n \rangle f_n$ est orthogonal à $f_1, \ldots, f_n$, donc appartient à $\mathrm{Vect}(f_1, \ldots, f_n)^{\perp} = F^{\perp}$, donc $x \in F + F^{\perp}$.

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 $E = F + F'$, $x = f + f'$ 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 $f = 0_E$ par séparation, et enfin $x = f' \in F'$.

(ii) Nous savons déjà que $F \subset F^{\perp\perp}$. Inversement, pour tout $x \in F^{\perp\perp}$, disons $x = f + f'$ avec $f \in F$ et $f' \in F^{\perp}$ : $\langle f', f' \rangle = \langle f + f', f' \rangle = \langle x, f' \rangle = 0$, donc $f' = 0_E$ par séparation, i.e. $x = f \in F$.

**Exemple** Dans l'espace euclidien canonique $\mathbb{R}^4$, le plan vectoriel $P$ d'équations $x - y - z - t = 0$ et $2x + y + z - t = 0$ admet $\mathrm{Vect}\big((1,-1,-1,-1),(2,1,1,-1)\big)$ pour orthogonal.

**Démonstration** La première égalité suivante mérite d'être parfaitement comprise.

$$\begin{aligned} P &= \Big\{(x,y,z,t) \in \mathbb{R}^4 \mid \big\langle (1,-1,-1,-1),(x,y,z,t) \big\rangle = 0 \quad \text{et} \quad \big\langle (2,1,1,-1),(x,y,z,t) \big\rangle = 0 \Big\} \\ &= \Big\{(1,-1,-1,-1),(2,1,1,-1)\Big\}^{\perp} = \mathrm{Vect}\big((1,-1,-1,-1),(2,1,1,-1)\big)^{\perp}, \end{aligned}$$

or $\mathbb{R}^4$ est euclidien, donc $P^{\perp} = \mathrm{Vect}\big((1,-1,-1,-1),(2,1,1,-1)\big)^{\perp\perp} = \mathrm{Vect}\big((1,-1,-1,-1),(2,1,1,-1)\big)$.

**Exemple** On munit $\mathscr{M}_n(\mathbb{R})$ de sa structure euclidienne canonique — cela veut dire qu'on travaille avec le produit scalaire canonique de $\mathscr{M}_n(\mathbb{R})$. Dans ce contexte : $\mathscr{S}_n(\mathbb{R})^{\perp} = \mathscr{A}_n(\mathbb{R})$.

**Démonstration** Pour tous $S \in \mathscr{S}_n(\mathbb{R})$ et $A \in \mathscr{A}_n(\mathbb{R})$ :

$$\langle A, S \rangle = \mathrm{tr}\big(A^{\top} S\big) = -\mathrm{tr}(AS) = -\mathrm{tr}(SA) = -\mathrm{tr}\big(S^{\top} A\big) = -\langle S, A \rangle = -\langle A, S \rangle, \qquad \text{donc } \langle A, S \rangle = 0,$$

donc $\mathscr{S}_n(\mathbb{R})$ et $\mathscr{A}_n(\mathbb{R})$ sont orthogonaux — donc en somme directe. Bref : $\mathscr{A}_n(\mathbb{R}) \subset \mathscr{S}_n(\mathbb{R})^{\perp}$, et notez bien que ce n'est qu'une inclusion à ce stade. Cela dit, pour tout $M \in \mathscr{M}_n(\mathbb{R})$ : $M = \dfrac{M + M^{\top}}{2} + \dfrac{M - M^{\top}}{2}$ avec $\dfrac{M + M^{\top}}{2} \in \mathscr{S}_n(\mathbb{R})$ et $\dfrac{M - M^{\top}}{2} \in \mathscr{A}_n(\mathbb{R})$, donc $\mathscr{M}_n(\mathbb{R}) = \mathscr{S}_n(\mathbb{R}) + \mathscr{A}_n(\mathbb{R})$. Finalement $\mathscr{M}_n(\mathbb{R}) = \mathscr{S}_n(\mathbb{R}) \overset{\perp}{\oplus} \mathscr{A}_n(\mathbb{R})$, donc par unicité du supplémentaire orthogonal : $\mathscr{A}_n(\mathbb{R}) = \mathscr{S}_n(\mathbb{R})^{\perp}$.

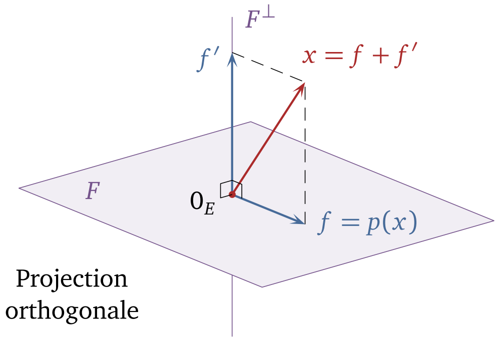
# 3 PROJECTION ORTHOGONALE SUR UN SOUS-ESPACE VECTORIEL DE DIMENSION FINIE

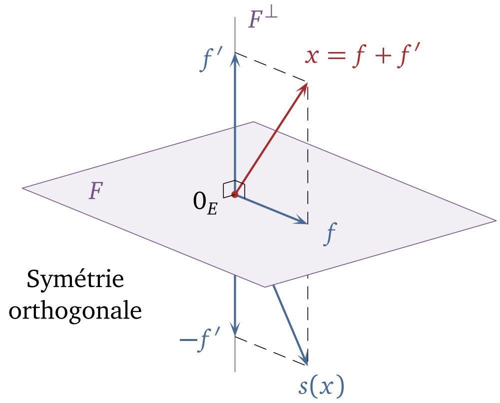
## 3.1 PROJECTIONS ET SYMÉTRIES ORTHOGONALES

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

- **Projection orthogonale :** On appelle *projection orthogonale sur $F$* ou *projecteur orthogonal sur $F$* la projection sur $F$ de direction $F^{\perp}$.
- **Symétrie orthogonale :** On appelle *symétrie orthogonale par rapport à $F$* la symétrie par rapport à $F$ parallèlement à $F^{\perp}$.

Comme $F$ est de dimension finie : $\quad E = F \overset{\perp}{\oplus} F^\perp$.




**Théorème (Expression d'un projeté orthogonal dans une base orthonormale)** Soient $E$ un espace préhilbertien réel, $F$ un sous-espace vectoriel **de dimension finie** de $E$ et $(f_1, \ldots, f_n)$ une base orthonormale de $F$.

Pour tout $x \in E$, le projeté orthogonal de $x$ sur $F$ vaut $\displaystyle\sum_{k=1}^n \langle x, f_k\rangle\, f_k$.

Dans ce résultat, $\langle x, f_k\rangle\, f_k$ est la composante de $x$ selon le vecteur $f_k$ dans la base $(f_1, \ldots, f_n)$.

**Démonstration** D'après nos résultats précédents : $\quad x = \underbrace{\sum_{k=1}^n \langle x, f_k\rangle\, f_k}_{\in F} + \underbrace{\left(x - \sum_{k=1}^n \langle x, f_k\rangle\, f_k\right)}_{\in F^\perp}$.

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 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 s'y ramène de force en orthonormalisant $(f_1, \ldots, f_n)$ grâce à l'algorithme de Gram-Schmidt.
- **Deuxième stratégie :** Le projeté $p(x)$ est caractérisé par deux assertions : $\boxed{p(x) \in F \quad \text{et} \quad x - p(x) \in F^\perp.}$ que je vous conseille de bien visualiser. Concrètement, on introduit les coordonnées $(\lambda_1, \ldots, \lambda_n)$ de $p(x)$ dans $(f_1, \ldots, f_n)$ : $p(x) = \lambda_1 f_1 + \ldots + \lambda_n f_n$, puis on les calcule grâce aux relations $\left\langle x - p(x), f_j\right\rangle = 0$, $j$ décrivant $[\![1, n]\!]$. Ces relations expriment l'appartenance de $x - p(x)$ à $F^\perp$ et fournissent un système de $n$ équations à $n$ inconnues qu'on n'a plus qu'à résoudre.

Finalement, quelle stratégie vaut-il mieux privilégier ? La deuxième me paraît souvent plus agréable à pratiquer.

**Exemple** On note $F$ le sous-espace vectoriel $\mathrm{Vect}(\cos, \sin)$ de $\mathscr{C}\big([0, 2\pi], \mathbb{R}\big)$. Le projeté orthogonal de l'identité $\mathrm{Id}$ 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.

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

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

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

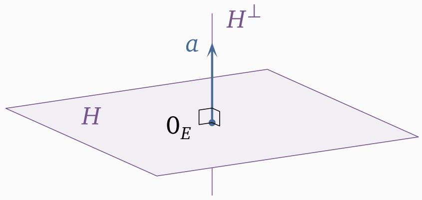
$$\langle \mathrm{Id}, \sin\rangle = \int_0^{2\pi} t \sin t\, \mathrm{d}t = \mathrm{Im}\left(\int_0^{2\pi} t\mathrm{e}^{\mathrm{i}t}\, \mathrm{d}t\right) = \mathrm{Im}(-2\mathrm{i}\pi) = -2\pi \qquad \text{et de même } \langle \mathrm{Id}, \cos\rangle = 0.$$

- **Première stratégie :** On orthonormalise la **base** $(\cos, \sin)$ de $F$ grâce à l'algorithme de Gram-Schmidt. Cette famille étant déjà orthogonale, $\left(\dfrac{\cos}{\|\cos\|}, \dfrac{\sin}{\|\sin\|}\right) = \left(\dfrac{\cos}{\sqrt{\pi}}, \dfrac{\sin}{\sqrt{\pi}}\right)$ est une base orthonormale de $F$. Le projeté orthogonal de $\mathrm{Id}$ sur $F$ est finalement la fonction :
$$\left\langle \mathrm{Id}, \frac{\cos}{\sqrt{\pi}}\right\rangle \frac{\cos}{\sqrt{\pi}} + \left\langle \mathrm{Id}, \frac{\sin}{\sqrt{\pi}}\right\rangle \frac{\sin}{\sqrt{\pi}} = \frac{\langle \mathrm{Id}, \cos\rangle}{\pi} \cos + \frac{\langle \mathrm{Id}, \sin\rangle}{\pi} \sin = -2\sin.$$

- **Deuxième stratégie :** Notons $p(\mathrm{Id})$ le projeté orthogonal de Id sur $F$ et $(\lambda, \mu)$ ses coordonnées dans la base $(\cos, \sin)$ de $F$ : $p(\mathrm{Id}) = \lambda \cos + \mu \sin$. Appuyons-nous sur le fait que $\mathrm{Id} - p(\mathrm{Id}) \in F^\perp$.
$$0 = \big\langle \mathrm{Id} - p(\mathrm{Id})\ ,\ \cos \big\rangle = \big\langle \mathrm{Id} - \lambda \cos - \mu \sin\ ,\ \cos \big\rangle = \langle \mathrm{Id}, \cos \rangle - \lambda \|\cos\|^2 - \mu \langle \sin, \cos \rangle = -\lambda \pi$$
et $$0 = \big\langle \mathrm{Id} - p(\mathrm{Id})\ ,\ \sin \big\rangle = \big\langle \mathrm{Id} - \lambda \cos - \mu \sin\ ,\ \sin \big\rangle = \langle \mathrm{Id}, \sin \rangle - \lambda \langle \cos, \sin \rangle - \mu \|\sin\|^2 = -2\pi - \mu \pi.$$
Conclusion : $\lambda = 0$ et $\mu = -2$, donc $p(\mathrm{Id}) = -2 \sin$.

## 3.2 Cas particulier des hyperplans

**Définition-théorème (Vecteurs normaux à un hyperplan, projection orthogonale et réflexion)** Soient $E$ un espace euclidien de dimension non nulle et $H$ un hyperplan de $E$.



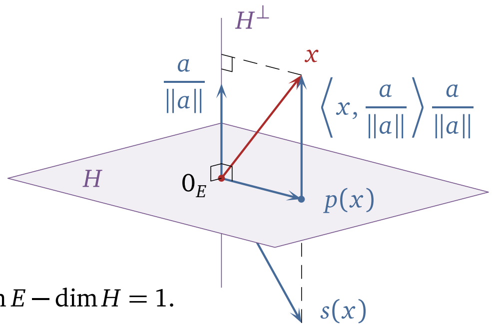

(i) **Vecteurs normaux à un hyperplan :** Le sous-espace $H^\perp$ est une droite dont tout vecteur non nul est appelé un *vecteur normal à $H$*.

Soit $a$ un vecteur normal à $H$ fixé.

(ii) **Projection orthogonale :** Pour tout $x \in E$, le projeté orthogonal de $x$ sur $H$ vaut $x - \dfrac{\langle x, a \rangle\, a}{\|a\|^2}$, ou encore $x - \langle x, a \rangle\, a$ si $a$ est unitaire.

(iii) **Réflexion :** On appelle *réflexion de $E$* toute symétrie orthogonale par rapport à un hyperplan de $E$. La réflexion de $E$ par rapport à $H$ a pour expression $x \longmapsto x - 2\, \dfrac{\langle x, a \rangle\, a}{\|a\|^2}$, ou encore $x \longmapsto x - 2 \langle x, a \rangle\, a$ si $a$ est unitaire.

Pour projeter sur $H$ qui est gros, il vaut mieux projeter sur $H^\perp$ qui est petit. Projeter $x$ sur $H$ revient ainsi à retrancher à $x$ sa composante selon $a$, et calculer son symétrique orthogonal par rapport à $H$ revient à retrancher deux fois cette composante.




**Démonstration**

(i) $H$ est de dimension finie, donc $E = H \oplus H^\perp$, mais par ailleurs $E$ est de dimension et $H$ en est un hyperplan, donc $\dim H^\perp = \dim E - \dim H = 1$.

(ii) La famille $\left( \dfrac{a}{\|a\|} \right)$ est une base orthonormale de $\mathrm{Vect}(a)$, donc le projeté orthogonal de $x$ sur $\mathrm{Vect}(a)$ vaut $\left\langle x, \dfrac{a}{\|a\|} \right\rangle \dfrac{a}{\|a\|} = \dfrac{\langle x, a \rangle\, a}{\|a\|^2}$. A fortiori, son projeté orthogonal sur $H = \mathrm{Vect}(a)^\perp$ vaut $x - \dfrac{\langle x, a \rangle}{\|a\|^2}$.

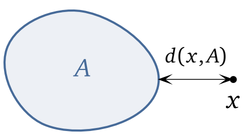
Pour finir, n'oublions pas que par définition des hyperplans, $H$ est le noyau d'une forme linéaire non nulle, i.e. peut être décrit par une unique équation linéaire scalaire non nulle. Quel lien avec la notion de vecteur normal ? Tout simplement, pour tout vecteur normal $a$ à $H$ : $H = \{a\}^\perp = \big\{ x \in E \mid \ \langle x, a \rangle = 0 \big\}$, donc $H$ est le noyau de la forme linéaire non nulle $x \longmapsto \langle x, a \rangle$. Donnons-nous maintenant une base orthornormale $(e_1, \ldots, e_n)$ de $E$ et notons $(a_1, \ldots, a_n)$ les coordonnées de $a$ dans cette base. L'équation $\langle x, a \rangle = 0$ s'écrit aussi $a_1 x_1 + \ldots + a_n x_n = 0$ sous forme analytique. Nous retrouvons là l'équation typique d'un hyperplan et le fait bien connu que les coefficients $a_1, \ldots, a_n$ y décrivent un vecteur normal.

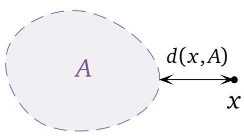
## 3.3 Distance à un sous-espace vectoriel de dimension finie

**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 $d(x, A) = \inf\limits_{a \in A} \|x - a\|$.

Intuitivement, la distance d'un vecteur $x$ à une partie $A$ est la plus petite distance séparant $x$ d'un élément de $A$, mais qui nous dit qu'une telle plus petite distance existe ? Elle n'existe justement pas forcément et c'est pour cela que la définition introduit 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.

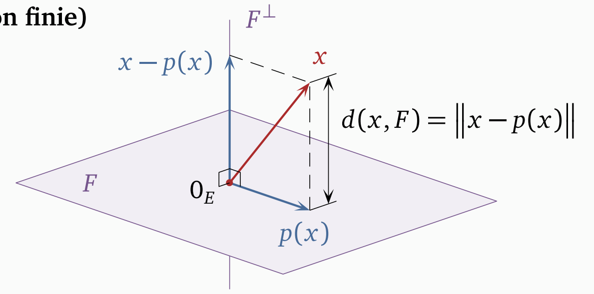
**Démonstration** L'ensemble $\big\{\|x-a\| \ \mid \ a\in A\big\}$ est une partie de $\mathbb{R}$ non vide car $A\neq\varnothing$ et minorée par 0, donc possède une borne inférieure d'après la propriété de la borne inférieure.

**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 $F$.

La distance de $x$ à $F$ est mieux qu'une borne inférieure, c'est un minimum : $d(x,F)=\big\|x-p(x)\big\|$, et ce minimum n'est atteint qu'en $p(x)$.

Par ailleurs : $d(x,F)^2=\|x\|^2-\big\|p(x)\big\|^2$.




**Démonstration**

- Pour tout $f\in F$ : $x-f=\underbrace{\big(x-p(x)\big)}_{\in\,\mathrm{Ker}\,p\,=\,F^\perp}+\underbrace{\big(p(x)-f\big)}_{\in\,\mathrm{Im}\,p\,=\,F}$, donc $\|x-f\|^2=\big\|x-p(x)\big\|^2+\big\|p(x)-f\big\|^2$ d'après le théorème de Pythagore.
- Notons à présent $\mathscr{D}$ l'ensemble $\big\{\|x-f\| \ \mid \ f\in F\big\}$. Montrons que $\big\|x-p(x)\big\|=\min\mathscr{D}$, cela prouvera que $\big\|x-p(x)\big\|=\inf\mathscr{D}=d(x,F)$. Or d'une part $\big\|x-p(x)\big\|\in\mathscr{D}$ car $p(x)\in F$, et d'autre part, pour tout $f\in F$ : $\|x-f\|=\sqrt{\big\|x-p(x)\big\|^2+\big\|p(x)-f\big\|^2}\geqslant\big\|x-p(x)\big\|$, donc $\mathscr{D}$ est minoré par $\big\|x-p(x)\big\|$.
- Pour finir, pour tout $f\in F\setminus\big\{p(x)\big\}$ : $\|x-f\|=\sqrt{\big\|x-p(x)\big\|^2+\big\|p(x)-f\big\|^2}>\big\|x-p(x)\big\|$ car $\big\|p(x)-f\big\|>0$ par séparation. Comme voulu, la distance $d(x,F)$ n'est atteinte qu'en $p(x)$.

**Exemple** On souhaite calculer $I=\inf\limits_{a,b\in\mathbb{R}}\displaystyle\int_0^{2\pi}\big(t-a\cos t-b\sin t\big)^2\,\mathrm{d}t$. Étudiez soigneusement cet exemple !

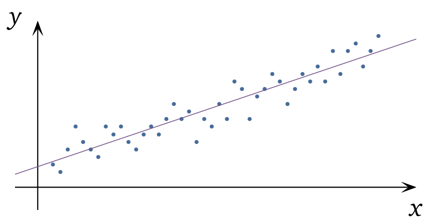
À ce stade du chapitre, une telle borne inférieure évoque une distance au carrée, mais ce point de vue mérite des précisions.

Intéressons-nous au produit scalaire $(f,g)\longmapsto\displaystyle\int_0^{2\pi}f(t)f(t)\,\mathrm{d}t$ sur $\mathscr{C}\big([0,2\pi],\mathbb{R}\big)$.

Première reformulation du problème : $I=\inf\limits_{a,b\in\mathbb{R}}\big\|\mathrm{Id}-a\cos-b\sin\big\|^2$, mais nous pouvons aller plus loin en notant $F$ le sous-espace vectoriel $\mathrm{Vect}(\cos,\sin)$ de $\mathscr{C}\big([0,2\pi],\mathbb{R}\big)$ : $I=\inf\limits_{f\in F}\big\|\mathrm{Id}-f\big\|^2=d(\mathrm{Id},F)^2$.

Sachant que $F$ est de dimension finie, il ne nous reste maintenant plus qu'à calculer le projeté orthogonal de Id sur $F$, puis $I=\big\|\mathrm{Id}-p(\mathrm{Id})\big\|^2=\|\mathrm{Id}\|^2-\big\|p(\mathrm{Id})\big\|^2$. Or nous avons calculé $p(\mathrm{Id})$ dans un exemple précédent : $p(\mathrm{Id})=-2\sin$. Finalement :

$$I=\|\mathrm{Id}\|^2-4\|\sin\|^2=\int_0^{2\pi}t^2\,\mathrm{d}t-4\int_0^{2\pi}\sin^2t\,\mathrm{d}t=\frac{8\pi^3}{3}-4\pi.$$




**Exemple** Il est courant qu'on ait à expliquer une quantité $y$ par une quantité $x$, en physique ou en économie par exemple. Faisons l'hypothèse que $y$ est une fonction affine de $x$ : $y=mx+p$ pour certains $m,p\in\mathbb{R}$. Comment déterminer $m$ et $p$ quand on a effectué $n$ mesures expérimentales $(x_1,y_1),\dots,(x_n,y_n)$ du couple $(x,y)$ ? Le nuage des mesures $(x_1,y_1),\dots,(x_n,y_n)$ est en principe assez proche de la droite idéale visée, mais il n'est pas inclus dedans a priori, ne serait-ce qu'en raison des erreurs de mesure. Nous cherchons donc 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* — *linéaire* en raison de la relation $y=mx+p$ qui lie $x$ et $y$.

Mais la droite « la plus proche » en quel sens ? Pour tout $i\in[\![1,n]\!]$, notre mesure de $y$ pour $x=x_i$ a fourni la valeur $y_i$, mais en l'absence d'erreurs de mesure, nous aurions trouvé $mx_i+p$. L'erreur de la $i^{\text{ème}}$ mesure vaut donc $\big|y_i-mx_i-p\big|$, 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 $\max\limits_{1\leqslant i\leqslant n}\big|y_i-mx_i-p\big|$, ou $\displaystyle\sum_{i=1}^n\big|y_i-mx_i-p\big|$, ou $\sqrt{\displaystyle\sum_{i=1}^n\big(y_i-mx_i-p\big)^2}$. 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 finalement deux réels $m$ et $p$ pour lesquels la quantité $\sqrt{\displaystyle\sum_{i=1}^n\big(y_i-mx_i-p\big)^2}$ est minimale — s'ils existent.

Dans l'espace euclidien canonique $\mathbb{R}^n$, posons : $x = (x_1, \ldots, x_n)$, $y = (y_1, \ldots, y_n)$, $u = \left(\frac{1}{n}, \ldots, \frac{1}{n}\right)$ et $F = \mathrm{Vect}(x, u)$.

Aussitôt : $\displaystyle \inf_{m,p\in\mathbb{R}} \sqrt{\sum_{i=1}^{n} \big(y_i - mx_i - p\big)^2} = \inf_{m,p\in\mathbb{R}} \big\| y - mx - npu \big\| = \inf_{f\in F} \| y - f \| = d(y, F)$. Or d'après le théorème précédent, en notant $p$ la projection orthogonale de $\mathbb{R}^n$ sur $F$, la distance $d(y,F)$ est atteinte en un et un seul point, en l'occurrence en $p(y) = mx + npu$ où $m$ et $p$ sont les deux réels que nous cherchons. Calculons donc $p(y)$ ! Or $y - p(y) \in F^\perp$, donc $\big\langle x, y - p(y)\big\rangle = 0$ et $\big\langle u, y - p(y)\big\rangle = 0$, i.e. : $m\,\|x\|^2 + np\,\langle x, u\rangle = \langle x, y\rangle$ et $m\,\langle x, u\rangle + p = \langle y, u\rangle$ — deux équations, deux inconnues. Après calcul : $m = \dfrac{\langle x, y\rangle - n\,\langle x, u\rangle\,\langle y, u\rangle}{\|x\|^2 - n\,\langle x, u\rangle^2}$ et $p = \langle y, u\rangle - m\,\langle x, u\rangle$. Posons alors :

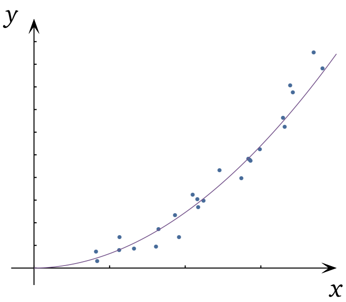
$$\mathrm{E}(x) = \langle x, u\rangle = \frac{1}{n}\sum_{i=1}^{n} x_i \quad \textit{(moyenne empirique des } x_i\textit{)}, \qquad \mathrm{E}(y) = \langle y, u\rangle = \frac{1}{n}\sum_{i=1}^{n} y_i \quad \textit{(moyenne empirique des } y_i\textit{)},$$

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

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

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

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




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 $\lambda$ et $\alpha$ sont deux réels à déterminer. On se ramène au cas affine en réécrivant les choses ainsi : $\ln y = \alpha \ln x + \ln \lambda$. Dans ce cas simple, ce sont les réels $\alpha$ et $\ln \lambda$ qu'on estimera par régression linéaire simple.