Chapitre 34

Espaces préhilbertiens réels

Dans ce chapitre, on ne manipule que des $\mathbb{R}$-espaces vectoriels. On se donne dans la suite un $\mathbb{R}$-espace vectoriel E (pas forcément de dimension finie dans un premier temps) et un entier $n \geqslant 1$.

Les scalaires seront donc toujours des réels.

I Produit scalaire

Un produit scalaire est un outil permettant de parler d'orthogonalité, et donc d'introduire un certain nombre de concepts permettant de généraliser la géométrie euclidienne du plan ou de l'espace à des situations plus « abstraites » (grrrrrr). Pour commencer, rappelons certains résultats concernant la bilinéarité (qui n'est qu'un cas particulier de la n-linéarité, mais un cas particulier très fréquent donc que l'on peut étudier à part entière).

I.1 Rappels sur la bilinéarité

Définition. Soit $f : E^2 \to \mathbb{R}$. On dit que f est une forme bilinéaire si elle est linéaire par rapport à chacune de ses variables, les autres étant fixées quelconques, c'est-à-dire si pour tout $(x_1, x_2) \in E^n$, les applications

$$f_1 : \left\{ \begin{array}{ccc} E & \to & \mathbb{R} \\ x & \mapsto & f(x, x_2) \end{array} \right. \qquad \text{et} \qquad f_2 : \left\{ \begin{array}{ccc} E & \to & \mathbb{R} \\ x & \mapsto & f(x_1, x) \end{array} \right.$$

sont linéaires.

Comme dans le chapitre 30, une forme est une application à valeurs dans $\mathbb{R}$ (rappelons que, dans ce chapitre, on se place toujours sur le corps $\mathbb{R}$). Attention, dans ce chapitre, nous ne parlerons presque jamais de formes linéaires, mais plutôt de formes bilinéaires symétriques définies positives.

Exemple : L'application produit

$$p : \left\{ \begin{array}{ccc} \mathbb{R}^2 & \to & \mathbb{R} \\ (x, y) & \mapsto & xy \end{array} \right.$$

est bilinéaire.

Nous reverrons de nombreux exemples dans la suite du cours.

Remarque : On peut définir des applications bilinéaires (pas forcément des formes, i.e. pas forcément à valeurs dans $\mathbb{R}$) mais cela n'est pas le cadre de ce chapitre. D'ailleurs, on l'a déjà fait ! En effet, la composition d'applications linéaires est bilinéaire, c'est-à-dire que

$$\circ : \left\{ \begin{array}{ccc} \mathscr{L}(E)^2 & \to & \mathscr{L}(E) \\ (u, v) & \mapsto & u \circ v \end{array} \right.$$

est bilinéaire. La linéarité à gauche est valable pour toutes fonctions, c'est-à-dire que l'égalité $(\lambda_1 u_1 + \lambda_2 u_2) \circ v = \lambda_1 u_1 \circ v + \lambda_2 u_2 \circ v$ est automatique pour toutes fonctions u_1, u_2, v (pas forcément linéaires), mais la linéarité à droite découle de la linéarité des applications, c'est-à-dire que l'égalité $v \circ (\lambda_1 u_1 + \lambda_2 u_2) = \lambda_1 v \circ u_1 + \lambda_2 v \circ u_2$ découle de la linéarité de v. Ces propriétés font que l'on note la composition d'applications linéaires comme un produit, cf. chapitre 29. Cependant, dans ce chapitre, on ne s'intéressera qu'à des formes i.e. des applications à valeurs dans $\mathbb{R}$.

Proposition. Soit $f : E^2 \to \mathbb{R}$ bilinéaire. Soient $x_1, \dots, x_n, y_1, \dots, y_n$ des vecteurs et $\alpha_1, \dots, \alpha_n, \beta_1, \dots, \beta_n$ des réels. Alors :

$$\begin{aligned} f\left(\sum_{k=1}^n \alpha_k x_k, \sum_{k=1}^n \beta_k y_k\right) &= \sum_{i=1}^n \alpha_i f\left(x_i, \sum_{k=1}^n \beta_k y_k\right) \\ &= \sum_{i=1}^n \sum_{j=1}^n \alpha_i \beta_j f(x_i, y_j) \end{aligned}$$

cf. chapitre 33.

Linéarité par rapport à la première variable.

Linéarité par rapport à la deuxième variable.

Remarques :

- En d'autres termes, on développe comme un produit « normal ».
- Comme dit dans le chapitre 33, il est impératif de changer d'indice de sommation ! En effet, si on écrit :

$$\sum_{k=1}^n \sum_{k=1}^n \alpha_k \beta_k f(x_k, y_k)$$

alors il y a confusion sur les indices. De plus, cela pourrait laisser croire qu'on ne trouve que les termes $\alpha_1\beta_1, \dots, \alpha_n\beta_n$ alors qu'en fait on trouve tous les $\alpha_i\beta_j$! Cela ne posait pas de problème au début car les deux sommes étaient disjointes, mais si on a une somme double, il faut impérativement deux indices différents.

Proposition. Soit $f : E^2 \to \mathbb{R}$ bilinéaire. Soient $(x_1, x_2) \in E^2$ et $\lambda \in \mathbb{R}$. Alors :

$$f(\lambda x_1, \lambda x_2) = \lambda^2 f(x_1, x_2)$$

Attention, f n'est pas linéaire, donc ce n'est pas λ qui sort mais λ^2, un λ par coordonnée !

I.2 Formes symétriques et définies positives

Définition. Soit $f : E^2 \to \mathbb{R}$. On dit que f est une forme

- symétrique si : $\forall (x,y) \in E^2, f(x,y) = f(y,x)$.
- définie positive si : $\forall x \in E, f(x,x) \geqslant 0$, avec égalité si et seulement si $x = 0$.

Exemple : L'application produit

$$p : \left\{ \begin{array}{ccc} \mathbb{R}^2 & \to & \mathbb{R} \\ (x,y) & \mapsto & xy \end{array} \right.$$

est symétrique : en effet, pour tous x et y appartenant à $\mathbb{R}$, $xy = yx$. Cette forme est de plus définie positive : en effet, pour tout $x \in \mathbb{R}$, $x^2 \geqslant 0$ avec égalité si et seulement si $x = 0$.

Nous reverrons de nombreux exemples dans la suite du cours.

Remarques :

- Certains ouvrages séparent la positivité de la définition, c'est-à-dire définissent une forme positive ($\forall x \in \mathbb{R}, f(x,x) \geqslant 0$) et une forme définie ($\forall x \in E, f(x,x) = 0 \iff x = 0$). Cependant, puisque nous ne manipulerons que des formes définies positives, nous utiliserons toujours ces deux termes ensembles afin de ne pas créer de confusion avec le terme « défini« ou « définition », qui pourrait être confondu avec un éventuel domaine de définition.

- Attention, une forme positive n'est pas à valeurs positives ! En effet, une forme est positive lorsqu'elle ne prend que des valeurs positives **lorsqu'on l'applique deux fois au même élément »** : ce ne sont que les images du type $f(x,x)$ qui sont positives ! Avec l'exemple précédent, pour tout x, on a bien $p(x,x) = x^2 \geqslant 0$ mais on n'a pas $xy \geqslant 0$ pour tous x et y !
- De la même façon, le caractère défini d'une forme ne signifie pas que seul le vecteur nul à une image nulle, ni que $f(x,y) = 0$ si et seulement si x ou y est nul (même si c'est le cas avec l'exemple très particulier ci-dessus). Nous verrons dans la suite des vecteurs x et y tous les deux non nuls tels que $f(x,y) = 0$ (et nous dirons que x et y sont orthogonaux, cf. paragraphe III).

I.3 Produits scalaires

I.3.a Définition, premières propriétés

Définition. Un produit scalaire est une forme bilinéaire, symétrique, définie positive.

Remarques :

- En d'autres termes, un produit scalaire sur un espace vectoriel E est une application $f : E^2 \to \mathbb{R}$ vérifiant les trois propriétés suivantes :
  - ⋆ bilinéarité : $\forall (x_1, x_2) \in E^2, \forall (\lambda_1, \lambda_2) \in \mathbb{R}^2, \forall y \in E$,
  $$f(\lambda_1 x_1 + \lambda_2 x_2, y) = \lambda_1 f(x_1, y) + \lambda_2 f(x_2, y) \qquad \text{et} \qquad f(y, \lambda_1 x_1 + \lambda_2 x_2) = \lambda_1 f(y, x_1) + \lambda_2 f(y, x_2)$$
  - ⋆ symétrie : $\forall (x,y) \in E^2, f(x,y) = f(y,x)$.
  - ⋆ définie positivité : $\forall x \in E, f(x,x) \geqslant 0$ avec égalité si et seulement si $x = 0_E$.
- Par linéarité à gauche, pour tout x_2, $x \mapsto f(x, x_2)$ est linéaire donc $f(0_E, x_2) = 0$ et, de même, $f(x_1, 0_E) = 0$ pour tout x_1. En particulier, $f(0_E, 0_E) = 0$ donc, pour montrer la définie positivité, on peut se contenter de l'implication : $f(x,x) = 0 \Rightarrow x = 0_E$, l'implication réciproque étant toujours vraie.
- Un produit scalaire prenant en arguments des vecteurs et renvoyant des scalaires (réels), il n'y aura aucune ambiguïté quant à la nature des objets que nous manipulerons et donc nous noterons le vecteur nul 0, comme le zéro réel. Par exemple, la définie positivité se réécrit : $\forall x \in E, f(x,x) \geqslant 0$ avec égalité si et seulement si $x = 0$.

On dira dans le paragraphe III.1 que le vecteur nul est orthogonal à tout vecteur.

Notation : Un produit scalaire est en général noté $\langle\, , \rangle$ ou $(\,|\,)$, c'est-à-dire que le produit scalaire de deux vecteurs x et y est noté $\langle x,y\rangle$ ou $(x|y)$. On le note parfois avec un point, c'est-à-dire que le produit scalaire de x et y est noté $x.y$, mais cette notation est plus rare et réservée en général au produit scalaire canonique (voir la suite) sur $\mathbb{R}^2$ ou $\mathbb{R}^3$, pour reprendre les notations que vous aviez au lycée.

On a donc prouvé ci-dessus que, pour tout $x \in E, \langle x,0\rangle = \langle 0,x\rangle = 0$.

Remarque : En d'autres termes, un produit scalaire est un moyen de définir une multiplication sur un espace vectoriel, mais attention : comme son nom l'indique, il renvoie un scalaire c'est-à-dire (dans ce chapitre) un réel, et non pas un vecteur. Sauf cas particuliers (par exemple des matrices), un produit de vecteurs n'a aucun sens. La bilinéarité fait qu'on peut développer un produit scalaire comme un produit normal :

Un produit scalaire est bien plus qu'une simple « multiplication », la remarque ci-contre n'est là que pour vous rappeler qu'on ne peut en général pas définir un produit de vecteurs, et surtout pour vous faire retenir les propositions ci-contre par analogie avec le produit classique.

Proposition. Soit $\langle\, , \rangle$ un produit scalaire sur E. Soient $x_1, \dots, x_n, y_1, \dots, y_n$ des vecteurs et $\alpha_1, \dots, \alpha_n, \beta_1, \dots, \beta_n$ des réels. Alors :
$$\left\langle \sum_{k=1}^n \alpha_k x_k, \sum_{k=1}^n \beta_k y_k \right\rangle = \sum_{i=1}^n \sum_{j=1}^n \alpha_i \beta_j \langle x_i, x_j \rangle$$

Démonstration. Découle du paragraphe I.1 puisqu'un produit scalaire est une forme bilinéaire. Rappelons qu'il faut changer l'indice des sommes quand on développe.

La symétrie du produit scalaire permet d'obtenir des « identités remarquables » :

Proposition (Identités remarquables). Soit $\langle\, ,\rangle$ un produit scalaire sur E.

- Soit $(x, y) \in E^2$. Alors : $\langle x+y, x+y\rangle = \langle x,x\rangle + 2\langle x,y\rangle + \langle y,y\rangle$.
- Plus généralement, si $n \geqslant 1$ et $(x_1, \dots, x_n) \in E^n$:

$$\left\langle \sum_{i=1}^n x_i, \sum_{i=1}^n x_i \right\rangle = \sum_{i=1}^n \langle x_i, x_i\rangle + 2\sum_{i<j} \langle x_i, x_j\rangle$$

Remarque : En d'autres termes, c'est comme pour un produit classique : tous les carrés (i.e. le produit d'un vecteur avec lui-même, mais le terme de « carré » deviendra plus clair dans le paragraphe II.1) et tous les doubles produits.

Démonstration.

- On a successivement :

$$\begin{aligned} \langle x+y, x+y\rangle &= \langle x, x+y\rangle + \langle y, x+y\rangle \\ &= \langle x,x\rangle + \langle x,y\rangle + \langle y,x\rangle + \langle y,y\rangle \\ &= \langle x,x\rangle + \langle x,y\rangle + \langle x,y\rangle + \langle y,y\rangle \end{aligned} \quad \square$$

Linéarité à gauche.

Linéarité à droite.

Symétrie.

ce qui donne le résultat voulu.
- Analogue au cas réel en utilisant la symétrie du produit scalaire : $\rightsquigarrow$ Exercice.

Remarques :

- Attention, comme on l'a vu ci-dessus, quand on veut développer une somme dans un produit scalaire, il faut faire attention : on peut développer comme un produit classique avec la proposition ci-dessus, on on développe d'abord par rapport à une coordonnée avant de développer par rapport à l'autre. Dans tous les cas, il faut faire attention et ne pas sortir d'ânerie du genre : « par linéarité, $\langle x+y, x+y\rangle = \langle x,x\rangle + \langle y,y\rangle$ » !
- On a également l'identité suivante, qu'on prouve de la même façon (exo) :

$$\forall (x,y) \in E^2, \langle x+y, x-y\rangle = \langle x,x\rangle - \langle y,y\rangle$$

mais elle est moins utile en pratique.

Définition.

- Un espace vectoriel E muni d'un produit scalaire est appelé un espace préhilbertien.
- Un espace vectoriel E **de dimension finie** muni d'un produit scalaire est appelé un espace euclidien.

Préhilbertien, sous-entendu : il lui manque un truc pour être un espace hilbertien, ou un espace de Hilbert... C'est effectivement le cas, mais les espaces de Hilbert sont HP... On se contentera donc des (gentils) espaces préhilbertiens.

Remarques :

- De la même façon qu'un groupe est un ensemble muni d'une loi interne, ou qu'un espace vectoriel est un ensemble muni d'une loi interne et d'une loi externe, un espace préhilbertien est un espace vectoriel muni d'un produit scalaire, c'est-à-dire un couple $(E, \langle\, ,\rangle)$ où $\langle\, ,\rangle$ est un produit scalaire sur E.
- De plus, de même qu'on n'explicite pas toujours la loi d'un groupe en parlant d'un groupe G au lieu d'un groupe $(G, *)$, et donc on écrit généralement la loi d'un groupe de façon multiplicative, on n'explicitera pas le produit scalaire d'un espace vectoriel en général : on parlera d'un espace préhilbertien E au lieu d'un espace préhilbertien $(E, \langle\, ,\rangle)$, le produit scalaire sera alors noté de façon habituelle $\langle\, ,\rangle$ ou $(\,|\,)$.

- Un espace euclidien n'est rien d'autre qu'un espace préhilbertien (i.e. muni d'un produit scalaire) de dimension finie. Ainsi, quand on commencera un exercice par : « soit E un espace euclidien », il est sous-entendu que E est muni d'un produit scalaire qu'on notera comme ci-dessus, et également qu'il est de dimension finie.

I.3.b Exemples

Certains exemples sont explicitement au programme et doivent former un « stock » que vous pouvez utiliser n'importe quand (et aussi que l'énoncé peut vous demander d'utiliser sans vous rappeler leur définition) :

Proposition.

- L'application

$$\langle\,,\rangle : \begin{cases} \mathbb{R}^n \times \mathbb{R}^n & \to \quad \mathbb{R} \\ (X,Y) & \mapsto \quad X^\top \times Y \end{cases}$$

est un produit scalaire sur $\mathbb{R}^n$, appelé produit scalaire canonique.

- L'application

$$\langle\,,\rangle : \begin{cases} \mathscr{M}_n(\mathbb{R}) \times \mathscr{M}_n(\mathbb{R}) & \to \quad \mathbb{R} \\ (A,B) & \mapsto \quad \operatorname{tr}\left(A^\top \times B\right) \end{cases}$$

est un produit scalaire sur $\mathscr{M}_n(\mathbb{R})$, appelé produit scalaire canonique.

- Soient $a < b$ deux réels. L'application

$$\langle\,,\rangle : \begin{cases} \mathscr{C}([\,a\,;b\,],\mathbb{R}) \times \mathscr{C}([\,a\,;b\,],\mathbb{R}) & \to \quad \mathbb{R} \\ (f,g) & \mapsto \quad \displaystyle\int_a^b f(t) \times g(t)\,\mathrm{d}t \end{cases}$$

est un produit scalaire sur $\mathscr{C}([\,a\,;b\,],\mathbb{R})$.

$\mathscr{C}([\,a\,;b\,],\mathbb{R})$ est donc un espace préhilbertien (mais pas euclidien !), et $\mathbb{R}^n$ et $\mathscr{M}_n(\mathbb{R})$ sont donc des espaces euclidiens car sont de dimension finie.

Bon... celui-là n'est pas appelé canonique, mais il est tout de même important et nous l'utiliserons sans arrêt : nous l'appellerons donc « produit scalaire habituel » (appellation personnelle...).

Remarques :

- Si

$$X = \begin{pmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{pmatrix} \qquad \text{et} \qquad Y \begin{pmatrix} y_1 \\ y_2 \\ \vdots \\ y_n \end{pmatrix}$$

sont deux éléments de $\mathbb{R}^n$, alors

$$X^\top \times Y = \begin{pmatrix} x_1 & x_2 & \cdots & x_n \end{pmatrix} \times \begin{pmatrix} y_1 \\ y_2 \\ \vdots \\ y_n \end{pmatrix} = x_1y_1 + x_2y_2 + \cdots + x_ny_n$$

En d'autres termes, le produit scalaire canonique sur $\mathbb{R}^n$ est l'application :

$$\left(\begin{pmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{pmatrix}, \begin{pmatrix} y_1 \\ y_2 \\ \vdots \\ y_n \end{pmatrix}\right) \longmapsto x_1y_1 + \cdots + x_ny_n = \sum_{i=1}^n x_iy_i$$

c'est-à-dire qu'on fait le produit coordonnée par coordonnée, et on somme. Pour $n=2$ ou $n=3$, on retrouve le produit scalaire du lycée ! On le note parfois également avec un point, c'est-à-dire qu'on note parfois (rarement, sauf si $n=2$ ou $n=3$ pour ne pas confondre avec la loi externe) $X.Y$ le produit scalaire de X et de Y.

- Comme pour les bases, un produit scalaire est dit canonique lorsque c'est le produit scalaire qu'on utilise le plus souvent, que c'est celui qui est universellement reconnu comme le plus simple, mais il y en a beaucoup d'autres (voir la suite). Pour faire simple : là aussi, c'est moi qui ai le super pouvoir pour décider qu'un produit scalaire est canonique (le dernier ne l'est pas, par exemple), pas vous !

Démonstration.

- Pour le produit scalaire sur $\mathbb{R}^n$:
  - ⋆ Soient $X=(x_1,\dots,x_n)$ et $Y=(y_1,\dots,y_n)\in\mathbb{R}^n$. Alors :

$$\begin{aligned}\langle X,Y\rangle &= \sum_{i=1}^n x_i y_i \\ &= \sum_{i=1}^n y_i x_i \\ &= \langle Y,X\rangle\end{aligned}$$

    En d'autres termes, $\langle\,,\rangle$ est symétrique.
  - ⋆ Soient $Z=(z_1,\dots,z_n)$ et $(\lambda,\mu)\in\mathbb{R}^2$.

$$\begin{aligned}\langle X,\lambda Y+\mu Z\rangle &= \sum_{i=1}^n x_i(\lambda y_i+\mu z_i) \\ &= \lambda\sum_{i=1}^n x_i y_i+\mu\sum_{i=1}^n y_i z_i \\ &= \lambda\langle X,Y\rangle+\mu\langle X,Z\rangle\end{aligned}$$

    c'est-à-dire que $\langle\,,\rangle$ est linéaire à droite et, par symétrie, linéaire à gauche donc bilinéaire.

On pourrait dire directement que $\langle\,,\rangle$ est bilinéaire par linéarité de la somme.

  - ⋆ Tout d'abord,

$$\langle X,X\rangle=\sum_{i=1}^n {x_i}^2\geqslant 0$$

    Supposons que $\langle X,X\rangle=0$. Alors $\displaystyle\sum_{i=1}^n {x_i}^2=0$. On a une somme de termes positifs (rappelons qu'on est sur $\mathbb{R}$) nulle donc tous les termes sont nuls, si bien que $X=0$: $\langle\,,\rangle$ est définie positive, c'est bien un produit scalaire.

- Pour le produit scalaire sur $\mathscr{M}_n(\mathbb{R})$: rappelons que si M et N appartiennent à $\mathscr{M}_n(\mathbb{R})$, alors, pour tous i et j :

$$(MN)_{i,j}=\sum_{k=1}^n M_{i,k}N_{k,j}$$

⋆ Soient A et $B \in \mathscr{M}_n(\mathbb{R})$.

$$\begin{aligned} \langle A,B\rangle &= \operatorname{tr}\left(A^\top B\right) \\ &= \operatorname{tr}\left(\left(A^\top B\right)^\top\right) \\ &= \operatorname{tr}\left(B^\top A\right) \\ &= \langle B,A\rangle \end{aligned}$$

La trace d'une matrice est égale à la trace de sa transposée.

donc $\langle\,,\rangle$ est symétrique.

⋆ Soient $C \in \mathscr{M}_n(\mathbb{R})$ et $(\lambda,\mu) \in \mathbb{R}^2$.

On pourrait dire que $\langle\,,\rangle$ est bilinéaire par linéarité de la trace.

$$\begin{aligned} \langle A,\lambda B+\mu C\rangle &= \operatorname{tr}\left(A^\top(\lambda B+\mu C)\right) \\ &= \lambda \operatorname{tr}\left(A^\top B\right) + \mu \operatorname{tr}\left(A^\top C\right) \\ &= \lambda\langle A,B\rangle + \mu\langle A,C\rangle \end{aligned}$$

Linéarité de la trace.

c'est-à-dire que $\langle\,,\rangle$ est linéaire à droite. Par symétrie, $\langle\,,\rangle$ est bilinéaire.

⋆ On a :

$$\begin{aligned} \langle A,A\rangle &= \operatorname{tr}\left(A^\top A\right) \\ &= \sum_{i=1}^n \left(A^\top A\right)_{i,i} \\ &= \sum_{i=1}^n \sum_{k=1}^n A^\top_{i,k} A_{k,i} \\ &= \sum_{i=1}^n \sum_{k=1}^n A_{k,i} A_{k,i} \\ &= \sum_{i=1}^n \sum_{k=1}^n {A_{k,i}}^2 \end{aligned}$$

On en déduit que $\langle A,A\rangle \geqslant 0$. De plus, on a une somme de termes positifs donc :

$$\langle A,A\rangle = 0 \iff \forall i \in [\![1\,;\,n]\!], \forall k \in [\![1\,;\,n]\!], A_{i,k} = 0$$

En d'autres termes, $\langle A,A\rangle = 0$ si et seulement si tous les coefficients de A sont nuls, si et seulement si $A = 0$: $\langle\,,\rangle$ est définie positive, c'est bien un produit scalaire.

- Pour le produit scalaire sur $\mathscr{C}([\,a\,;b\,],\mathbb{R})$:

⋆ Soient f et $g \in \mathscr{C}([\,a\,;b\,],\mathbb{R})$.

$$\begin{aligned} \langle f,g\rangle &= \int_a^b f(t)g(t)\,\mathrm{d}t \\ &= \int_a^b g(t)f(t)\,\mathrm{d}t \\ &= \langle g,f\rangle \end{aligned}$$

donc $\langle\,,\rangle$ est symétrique.

⋆ Soient $h \in \mathscr{C}([\,a\,;b\,],\mathbb{R})$ et $(\lambda,\mu) \in \mathbb{R}^2$.

$$\begin{aligned}\langle f,\lambda g+\mu h\rangle &= \int_a^b f(t)(\lambda g(t)+\mu h(t))\,\mathrm{d}t\\ &= \lambda\int_a^b f(t)g(t)\,\mathrm{d}t+\mu\int_a^b f(t)h(t)\,\mathrm{d}t\\ &= \lambda\langle f,g\rangle+\mu\langle f,h\rangle\end{aligned}$$

On pourrait dire que $\langle\,,\rangle$ est bilinéaire par linéarité de l'intégrale.

Par linéarité de l'intégrale.

c'est-à-dire que $\langle\,,\rangle$ est linéaire à droite. Par symétrie, $\langle\,,\rangle$ est bilinéaire.

⋆ On a :

$$\langle f,f\rangle = \int_a^b f(t)^2\,\mathrm{d}t$$

Les bornes sont dans l'ordre croissant et f^2 est une fonction positive donc, par positivité de l'intégrale, $\langle f,f\rangle \geqslant 0$. Supposons que $\langle f,f\rangle = 0$. Alors

$$\int_a^b f(t)^2\,\mathrm{d}t = 0$$

□

On a l'intégrale d'une fonction positive, continue, avec $a<b$, qui vaut 0, donc $f=0$: $\langle\,,\rangle$ est définie positive, c'est bien un produit scalaire.

La continuité est indispensable pour affirmer que f est nulle. On rencontre souvent des produits scalaires de cette forme i.e. avec des intégrales (voir ci-dessous), et la définie positivité se prouve toujours de la même façon, à l'aide de la continuité de f.

Remarques :

- Comme on l'a vu, prouver la symétrie avant la bilinéarité permet de ne montrer la linéarité que d'un seul côté. Si on veut prouver la bilinéarité avant la symétrie, il faut la prouver des deux côtés, ce qui est une perte de temps (même si, en général, un simple « de même » suffit).
- Comme on l'a dit, ces trois produits scalaires sont au programme et doivent être connus. Ils peuvent également être utiles dans des exercices n'ayant a priori rien à voir avec des produits scalaires. Par exemple, on trouve souvent, sous une forme ou sous une autre (cf. par exemple l'exercice 37 du chapitre 31) la question suivante : « Soit $A \in \mathscr{M}_n(\mathbb{R})$ non nulle. Donner la dimension de $H = \{M \mid \mathrm{tr}\,(AM) = 0\}$ ».

  On pense évidemment à un hyperplan et on dit que H est le noyau de $\varphi : M \mapsto \mathrm{tr}\,(AM)$ qui est une forme linéaire non nulle donc est un hyperplan de $\mathscr{M}_n(\mathbb{R})$ donc est de dimension n^2-1. Seulement voilà : comment prouver que cette forme linéaire (c'en est bien une par linéarité de la trace) est non nulle ? Il faut exhiber une matrice M pour laquelle $\mathrm{tr}\,(AM) \neq 0$, et ce n'est pas si simple si on ne pense pas au produit scalaire ci-dessus : on sait que c'est un produit scalaire donc une forme définie positive donc $\mathrm{tr}\left(A^\top \times A\right) = \mathrm{tr}\left(A \times A^\top\right) \neq 0$ puisque $A \neq 0$ donc cette forme linéaire est non nulle car sa valeur en $A^\top$ est non nulle.

Rappelons que $\mathrm{tr}\,(AB) = \mathrm{tr}\,(BA)$.

Donnons d'autres exemples de produits scalaires (plus ou moins classiques donc il est bon de les connaître, mais ils ne sont pas explicitement au programme comme les précédents).

- Le dernier exemple peut aussi servir à définir un produit scalaire sur $\mathbb{R}[X]$: prouvons que

$$\langle\,,\rangle : \left\{\begin{array}{ccc}\mathbb{R}[X]^2 & \to & \mathbb{R}\\ (P,Q) & \mapsto & \displaystyle\int_a^b P(t)Q(t)\,\mathrm{d}t\end{array}\right.$$

  est un produit scalaire sur $\mathbb{R}[X]$.

- ⋆ La symétrie est immédiate.
- ⋆ La bilinéarité découle de la linéarité de l'intégrale.
- ⋆ Enfin, si $P \in \mathbb{R}[X]$, alors

$$\langle P,P\rangle = f(0)^2 \int_0^1 P(t)^2 \,\mathrm{d}t \geqslant 0$$

par positivité de l'intégrale. Si $\langle P,P\rangle = 0$ alors

$$\int_0^1 P(t)^2 \,\mathrm{d}t = 0$$

Or, P est continu positif d'intégrale nulle donc est nul sur $[\,a\,;b\,]$: il en découle que P admet une infinité de racines donc est le polynôme nul. On en déduit que $\langle\,,\rangle$ est définie positive, c'est un produit scalaire.

On identifie polynôme et fonction polynomiale associée.

- Montrons que

$$\langle\,,\rangle : \begin{cases} \mathscr{C}([\,-1\,;1\,],\mathbb{R})^2 & \to & \mathbb{R} \\ (f,g) & \mapsto & \displaystyle\int_{-1}^1 f(t)g(t)\sqrt{1-t^2}\,\mathrm{d}t \end{cases}$$

est un produit scalaire sur $\mathscr{C}([\,-1\,;1\,],\mathbb{R})$.

- ⋆ La symétrie est immédiate.
- ⋆ La bilinéarité découle de la linéarité de l'intégrale.
- ⋆ Enfin, si $f \in \mathscr{C}([\,-1\,;1\,],\mathbb{R})$, alors

$$\langle f,f\rangle = \int_{-1}^1 f(t)^2\sqrt{1-t^2}\,\mathrm{d}t \geqslant 0$$

par positivité de l'intégrale. Si $\langle f,f\rangle = 0$ alors $t \mapsto f(t)^2\sqrt{1-t^2}$ est une fonction continue positive d'intégrale nulle donc est continue sur $[\,-1\,;1\,]$. Or, $\sqrt{1-t^2} \neq 0$ sur $]\,-1\,;1\,[$ donc f est nulle sur $]\,-1\,;1\,[$ donc sur $[\,-1\,;1\,]$ par continuité : $\langle\,,\rangle$ est définie positive, c'est un produit scalaire.

Remarque : On peut définir de même un produit scalaire sur $\mathbb{R}[X]$:

$$\langle\,,\rangle : \begin{cases} \mathbb{R}[X]^2 & \to & \mathbb{R} \\ (P,Q) & \mapsto & \displaystyle\int_{-1}^1 P(t)Q(t)\sqrt{1-t^2}\,\mathrm{d}t \end{cases}$$

Plus généralement, si $a < b$ sont deux réels et si w est une fonction continue sur $[\,a\,;b\,]$ strictement positive sur $]\,a\,;b\,[$ (appelée fonction de poids), alors

$$\langle\,,\rangle : \begin{cases} \mathscr{C}([\,a\,;b\,],\mathbb{R})^2 & \to & \mathbb{R} \\ (f,g) & \mapsto & \displaystyle\int_a^b f(t)g(t)w(t)\,\mathrm{d}t \end{cases}$$

est un produit scalaire sur $\mathscr{C}([\,a\,;b\,],\mathbb{R})$, et on peut définir de même un produit scalaire sur $\mathbb{R}[X]$. Ce genre de produit scalaire est très utile pour définir ce qu'on appelle des polynômes orthogonaux, polynômes qui dépendent évidemment de la fonction de poids w : tout l'intérêt consiste à choisir une fonction w pour avoir des polynômes « intéressants », cf. exercices 43, 44, 45.

- On peut définir d'autres produits scalaires sur $\mathbb{R}[X]$: rappelons que tout polynôme $P \in \mathbb{R}[X]$ s'écrit $P = \displaystyle\sum_{k=0}^{+\infty} a_k X^k$, cette somme étant en fait finie (et donc il n'y a aucun problème de convergence, contrairement à l'exemple suivant). . On montre de même que pour le produit scalaire canonique de $\mathbb{R}^n$ que

Rappelons qu'un polynôme est, par définition, une suite presque nulle.

$$\langle\,,\rangle : \left\{ \begin{array}{ccc} \mathbb{R}[X]^2 & \to & \mathbb{R} \\ \left(\displaystyle\sum_{k=0}^{+\infty} a_k X^k, \sum_{k=0}^{+\infty} b_k X^k \right) & \mapsto & \displaystyle\sum_{k=0}^{+\infty} a_k b_k \end{array} \right.$$

est un produit scalaire sur $\mathbb{R}[X]$.

- Rappelons (cf. exercice 32 du chapitre 25 et exercice 24 du chapitre 30) que

$$\ell^2(\mathbb{N}) = \left\{ (u_n)_{n\in\mathbb{N}} \mid \text{ la série } \sum {u_n}^2 \text{ converge} \right\}$$

est un espace vectoriel. Montrons que

$$\langle\,,\rangle : \left\{ \begin{array}{ccc} \ell^2(\mathbb{N}) \times \ell^2(\mathbb{N}) & \to & \mathbb{R} \\ ((u_n)_{n\in\mathbb{N}}, (v_n)_{n\in\mathbb{N}}) & \mapsto & \displaystyle\sum_{n=0}^{+\infty} u_n v_n \end{array} \right.$$

est un produit scalaire.

  ⋆ Prouvons déjà (ce qui n'était pas nécessaire dans les exemples précédents) que ce produit scalaire est bien défini, c'est-à-dire que la somme infinie ci-dessus existe bien. Soient donc (u_n) et (v_n) deux éléments de $\ell^2(\mathbb{N})$. Alors, pour tout n,

Inégalité classique ! cf. chapitre 2 et exercice 32 du chapitre 25.

$$|a_n b_n| \leqslant \frac{1}{2}\left({a_n}^2 + {b_n}^2\right)$$

  Or, par hypothèse, les deux séries $\sum {a_n}^2$ et $\sum {b_n}^2$ convergent donc la série $\sum \frac{1}{2}\left({a_n}^2 + {b_n}^2\right)$ converge : par théorème de comparaison pour les séries à termes positifs, $\sum |a_n b_n|$ converge donc la série $\sum a_n b_n$ converge absolument, et donc converge : la fonction est bien définie.

  ⋆ La symétrie est immédiate.

  ⋆ La bilinéarité découle de la linéarité de la somme (infinie).

  ⋆ Enfin, si $(u_n) \in \ell^2(\mathbb{N})$, alors :

$$\langle (u_n)_{n\in\mathbb{N}}, (u_n)_{n\in\mathbb{N}} \rangle = \sum_{n=0}^{+\infty} {u_n}^2 \geqslant 0$$

  De plus, si cette quantité est nulle, alors ${u_n}^2 = 0$ pour tout n (somme de termes positifs) donc $u_n = 0$ pour tout n : (u_n) est donc la suite nulle, $\langle\,,\rangle$ est définie positive, c'est un produit scalaire.

- Montrons que

$$\langle\,,\rangle : \left\{ \begin{array}{ccc} \mathscr{C}^1([\,0\,;1\,], \mathbb{R})^2 & \to & \mathbb{R} \\ (f, g) & \mapsto & f(0)g(0) + \displaystyle\int_0^1 f'(t)g'(t)\,\mathrm{d}t \end{array} \right.$$

est un produit scalaire sur $\mathscr{C}^1([\,0\,;1\,], \mathbb{R})$.

- ⋆ La symétrie est immédiate.
- ⋆ Soient $(f,g,h) \in \mathscr{C}^1([\,0\,;1\,],\mathbb{R})^3$ et $(\lambda,\mu) \in \mathbb{R}^2$.

$$\begin{aligned}
\langle f,\lambda g+\mu h\rangle &= f(0)\times(\lambda g+\mu h)(0)+\int_0^1 f'(t)(\lambda g+\mu h)'(t)\,\mathrm{d}t\\
&= f(0)\times(\lambda g(0)+\mu h(0))+\int_0^1 f'(t)(\lambda g'(t)+\mu h'(t))\,\mathrm{d}t\\
&= \lambda f(0)g(0)+\mu f(0)h(0)+\lambda\int_0^1 f'(t)g'(t)\,\mathrm{d}t+\mu\int_0^1 f'(t)h'(t)\,\mathrm{d}t\\
&= \lambda\langle f,g\rangle+\mu\langle f,h\rangle
\end{aligned}$$

Linéarité de l'évaluation et de la dérivation.

Linéarité de l'intégrale.

c'est-à-dire que $\langle\,,\rangle$ est linéaire à droite. Par symétrie, elle est aussi linéaire à gauche donc bilinéaire.

- ⋆ Enfin, si $f \in \mathscr{C}^1([\,0\,;1\,],\mathbb{R})$, alors

$$\langle f,f\rangle = f(0)^2+\int_0^1 f'(t)^2\,\mathrm{d}t \geqslant 0$$

par positivité de l'intégrale. Si $\langle f,f\rangle = 0$ alors

$$f(0)^2 = \int_0^1 f'(t)^2\,\mathrm{d}t = 0$$

puisqu'on a une somme de termes positifs. On en déduit que f' est nulle puisque f'^2 est continue (car f est $\mathscr{C}^1$) positive d'intégrale nulle donc est la fonction nulle : il en découle que f est constante, et puisque $f(0)=0$, alors f est constante égale à 0 : $\langle\,,\rangle$ est définie positive, c'est un produit scalaire.

- • On peut évidemment (comme on l'a déjà vu sur $\mathscr{C}([\,-1\,;1\,],\mathbb{R})$ ou sur $\mathbb{R}[X]$) définir plusieurs produits scalaires sur un même espace vectoriel. Montrons par exemple que

$$\langle\,,\rangle : \begin{cases} \mathbb{R}^2\times\mathbb{R}^2 & \to \quad \mathbb{R}\\ (x,y) & \mapsto \quad x^\top\times\begin{pmatrix}2&1\\1&2\end{pmatrix}\times y\end{cases}$$

est un produit scalaire (différent du produit scalaire canonique) sur $\mathbb{R}^2$.

- ⋆ Soient $x=\begin{pmatrix}x_1\\x_2\end{pmatrix}$ et $y=\begin{pmatrix}y_1\\y_2\end{pmatrix} \in \mathbb{R}^2$. D'une part :

$$\begin{aligned}
\langle x,y\rangle &= \begin{pmatrix}x_1 & x_2\end{pmatrix}\begin{pmatrix}2&1\\1&2\end{pmatrix}\begin{pmatrix}y_1\\y_2\end{pmatrix}\\
&= 2x_1x_2+x_1y_2+x_2y_1+2x_2y_2
\end{aligned}$$

On obtient de même que cette quantité est égale à $\langle y,x\rangle$ donc $\langle\,,\rangle$ est symétrique.

- ⋆ La bilinéarité est immédiate.
- ⋆ Enfin, d'après ce qui précède, $\langle x,x\rangle = 2{x_1}^2+2x_1x_2+2{x_2}^2 = {x_1}^2+(x_1+x_2)^2+{x_2}^2 \geqslant 0$ puisqu'on a une somme de termes positifs. Si $\langle x,x\rangle=0$ alors ${x_1}^2={x_2}^2=0$ car ce sont des termes positifs donc $x_1=x_2=0$ si bien que x est le vecteur nul : $\langle\,,\rangle$ est définie positive, c'est un produit scalaire.

- Cependant, si $(\Omega, \mathscr{P}(\Omega), P)$ est un espace probabilité fini, les applications

$$\begin{cases} \mathbb{R}^\Omega \times \mathbb{R}^\Omega & \to & \mathbb{R} \\ (X,Y) & \mapsto & E(XY) \end{cases} \quad \text{et} \quad \mathrm{Cov} : \begin{cases} \mathbb{R}^\Omega \times \mathbb{R}^\Omega & \to & \mathbb{R} \\ (X,Y) & \mapsto & \mathrm{Cov}(X,Y) \end{cases}$$

ne sont pas des produits scalaires car ne sont pas définies positives : elles sont symétriques, bilinéaires, mais pas définies positives. Plus précisément, elles sont positives (i.e. l'image de (X,X) est positive) mais pas définies car on peut avoir l'image de (X,X) qui est nulle sans que X soit nulle. Par exemple, si X est constante égale à 1 presque sûrement, alors $\mathrm{Cov}(X,X) = V(X) = 0$.

Cela nous permettra tout de même d'appliquer l'inégalité de Cauchy-Schwarz et le théorème de Pythagore : cf. paragraphes II.2.a et IV.4.

II Norme et inégalités (très) classiques

On suppose dans la suite du cours que E est un espace préhilbertien muni d'un produit scalaire $\langle\, ,\rangle$.

Préhilbertien, pas euclidien : E n'est pas supposé de dimension finie.

II.1 Norme associée à un produit scalaire

Définition. On appelle norme euclidienne associée à $\langle\, ,\rangle$ l'application :

$$\|.\| : \begin{cases} E \longrightarrow \mathbb{R}_+ \\ x \longmapsto \sqrt{\langle x,x\rangle} \end{cases}$$

Remarques :

- La norme étant une racine carrée, elle est bien à valeurs positives.
- La norme est bien définie puisque, par définition d'un produit scalaire, $\langle x,x\rangle \geqslant 0$ pour tout x.
- Attention : $\langle x,x\rangle = \|x\|^2$! Il y a un carré, attention à ne pas l'oublier. Par exemple (cf. exercice 28), $\langle x,x\rangle^2 = \|x\|^4$.
- Si $x \in E$, $\|x\|$ est appelée norme (euclidienne) de x. Une norme est un moyen de définir la « taille » d'un objet : plus sa norme est grande, plus l'objet est « grand ». En prenant la différence, une norme peut également servir à définir une distance et donc est un moyen de définir la distance entre deux objets, et donc de définir rigoureusement le fait que deux objets sont éloignés l'un de l'autre. Nous en reparlerons dans le paragraphe II.4.
- Pour l'instant, elle permet de donner une forme relativement simple à certaines propriétés du produit scalaire :

Proposition. Soient $(x,y) \in E^2$ et $\lambda \in \mathbb{R}$.

- **(Positivité et séparation)** $\|x\| \geqslant 0$, avec égalité si et seulement si $x = 0$.
- **(Absolue homogénéité)** $\|\lambda x\| = |\lambda| \times \|x\|$.
- **(Identités remarquables)** $\|x+y\|^2 = \|x\|^2 + 2\langle x,y\rangle + \|y\|^2$. Plus généralement, si $n \geqslant 1$ et $(x_1, \dots, x_n) \in E^n$:

$$\left\| \sum_{i=1}^n x_i \right\|^2 = \sum_{i=1}^n \|x_i\|^2 + 2\sum_{i<j} \langle x_i,x_j\rangle$$

Ces identités remarquables nous permettront par exemple de démontrer le théorème de Pythagore dans le paragraphe III.2.

Démonstration. • On a déjà vu que $\|x\| \geqslant 0$. De plus, le produit scalaire étant défini positif : $\|x\| = 0 \iff \langle x,x\rangle = 0 \iff x = 0$.

- $\|\lambda x\| = \sqrt{\langle \lambda x,\lambda x\rangle}$. Or, par bilinéarité du produit scalaire, $\langle \lambda x,\lambda x\rangle = \lambda^2\langle x,x\rangle$ ce qui permet de conclure.

- Découle du paragraphe I.3.a.

Définition. Un vecteur unitaire est un vecteur de norme 1.

Tout vecteur non nul peut « se ramener » à un vecteur unitaire :

cf. paragraphe II.4 pour un dessin.

Proposition. Si $x \neq 0$, alors $\dfrac{x}{\|x\|}$ est unitaire.

Démonstration. Il suffit d'appliquer l'absolue homogénéité avec $\lambda = 1/\|x\| \geqslant 0$:

$$\left\| \frac{x}{\|x\|} \right\| = \frac{1}{\|x\|} \times \|x\| = 1$$

□

On définit la norme en fonction du produit scalaire. On peut se poser la question suivante : peut-on retrouver le produit scalaire en fonction de la norme ? Deux produits scalaires distincts donnent-ils deux normes distinctes ? La réponse est oui :

Corollaire (Formule de polarisation).

$$\forall (x,y) \in E^2, \qquad \langle x,y \rangle = \frac{\|x+y\|^2 - \|x\|^2 - \|y\|^2}{2}$$

Démonstration. Découle de l'identité remarquable ci-dessus.

Remarque : Par conséquent, en partant du produit scalaire, on obtient une norme, et en partant de la norme, on peut retrouver le produit scalaire. On en déduit plusieurs choses :

- Si on a deux produits scalaires distincts, cela donne deux normes distinctes : en d'autres termes, l'application qui à un produit scalaire donne sa norme associée est injective. Encore en d'autres termes : un produit scalaire est entièrement déterminé par sa norme euclidienne associée.
- Un résultat vrai pour l'un aura de grandes chances d'être vrai pour l'autre. Donnons un exemple explicite de cette idée.

Exemple : Soit $f \in \mathscr{L}(E)$. On dit que f préserve la norme si :

$$\forall x \in E, \|f(x)\| = \|x\|$$

et on dit que f préserve le produit scalaire si :

$$\forall x, y, \langle f(x), f(y) \rangle = \langle x,y \rangle$$

Montrons que f préserve la norme si et seulement si elle préserve le produit scalaire.

Un tel endormorphisme est dit orthogonal : cf. deuxième année ! Par exemple : une rotation (de centre O : on cherche des applications linéaires).

Supposons que f préserve la norme, montrons qu'elle préserve le produit scalaire. Soit $(x,y) \in E^2$. D'après la formule de polarisation ci-dessus :

$$\begin{aligned}
\langle f(x), f(y) \rangle &= \frac{\|f(x)+f(y)\|^2 - \|f(x)\|^2 - \|f(y)\|^2}{2} \\
&= \frac{\|f(x+y)\|^2 - \|f(x)\|^2 - \|f(y)\|^2}{2} \\
&= \frac{\|x+y\|^2 - \|x\|^2 - \|y\|^2}{2} \\
&= \langle x,y \rangle
\end{aligned}$$

Linéarité de f.

f préserve la norme.

donc f préserve le produit scalaire. Réciproquement, supposons que f préserve le produit scalaire et prouvons qu'elle préserve la norme. Soit $x \in \mathbb{E}$. Alors $\langle f(x), f(x) \rangle = \langle x,x \rangle$ donc, en prenant la racine carrée : $\|f(x)\| = \|x\|$.

II.2 Inégalité de Cauchy-Schwarz

II.2.a Inégalité proprement dite

Théorème (Inégalité de Cauchy-Schwarz). Soit $(x,y) \in E^2$. Alors :

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

avec égalité si et seulement si x et y sont colinéaires.

Schwarz, sans « t » ! On ne confondra pas Hermann Schwarz avec Laurent Schwartz ! Comme moyen mnémotechnique, vous pouvez introduire un nouveau rituel pour l'apéro (l'abus d'alcool est dangereux pour la santé, à consommer avec modération) : Cauchy-Schwarz, santé !

Remarques :

- Sous une forme équivalente :
$$\forall (x,y) \in E^2, \langle x,y\rangle^2 \leqslant \|x\|^2 \times \|y\|^2$$
ou encore :
$$\forall (x,y) \in E^2, \langle x,y\rangle^2 \leqslant \langle x,x\rangle \times \langle y,y\rangle$$
ou encore :
$$\forall (x,y) \in E^2, |\langle x,y\rangle| \leqslant \sqrt{\langle x,x\rangle} \times \sqrt{\langle y,y\rangle}$$
C'est d'ailleurs sous l'une de ces formes qu'on l'utilise le plus, mais je préfère que vous reteniez l'autre et que vous reveniez à celle-ci car, sinon, il y a un risque d'oubli de la racine carrée, et l'énoncé du théorème est très simple à retenir car il est immédiat dans le cas particulier du produit scalaire canonique sur $\mathbb{R}^2$ ou sur $\mathbb{R}^3$, car on a $\langle u,v\rangle = \|u\| \times \|v\| \times \cos(u,v)$.
- L'inégalité de Cauchy-Schwarz est la plus importante du chapitre et sans doute l'une des plus importantes des mathématiques. Quand on vous demande de prouver une inégalité, demandez-vous s'il n'y aurait pas du Cauchy-Schwarz caché derrière, même si a priori cela n'a rien à voir avec un produit scalaire : il peut y avoir un produit scalaire sous-jacent ! Voir les exemples au paragraphe suivant.

Démonstration. Le résultat est immédiat si $y = 0$ puisque $\langle x,0\rangle = 0$ pour tout $x \in E$, donc on a bien $|\langle x,0\rangle| \leqslant \|x\| \times \|0\|$. Supposons donc que $y \neq 0$.Soit

$$\varphi : \begin{cases} \mathbb{R} \longrightarrow \mathbb{R} \\ t \longmapsto \|x+ty\|^2 \end{cases}$$

Soit $t \in \mathbb{R}$.

$$\begin{aligned} \varphi(t) &= \|x\|^2 + 2t\langle x,y\rangle + \|ty\|^2 \\ &= \|x\|^2 + 2t\langle x,y\rangle + t^2\|y\|^2 \end{aligned}$$

c'est-à-dire que φ est une fonction polynôme de degré 2 (car son coefficient dominant, $\|y\|$, est non nul). Or, φ est à valeurs positives donc $\Delta \leqslant 0$. Par conséquent :

En effet, si $\Delta > 0$, alors φ change de signe entre ses deux racines ce qui est exclu.

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

si bien que $\langle x,y\rangle^2 \leqslant \|x\|^2\|y\|^2$ et on conclut en appliquant la racine carrée (fonction croissante). Cherchons quand il y a égalité (quand $y \neq 0$ puisqu'il y a évidemment égalité quand $y = 0$).

$$\begin{aligned} \text{Il y a égalité} &\iff \Delta = 0 \\ &\iff \varphi \text{ admet une racine} \\ &\iff \exists t_0 \in \mathbb{R}, \varphi(t_0) = 0 \\ &\iff \exists t_0 \in \mathbb{R}, \|x + t_0 y\| = 0 \\ &\iff \exists t_0 \in \mathbb{R}, x + t_0 y = 0 \\ &\iff \exists t_0 \in \mathbb{R}, x = -t_0 y \\ &\iff x \text{ et } y \text{ sont proportionnels.} \end{aligned}$$

$\|u\| = 0 \iff u = 0.$

$\square$

II.2.b Exemples

- Avec le produit scalaire canonique sur $\mathbb{R}^n$, l'inégalité de Cauchy-Schwarz devient :

$$\forall (x_1,\dots,x_n,y_1,\dots,x_n)\in\mathbb{R}^{2n}, \left|\sum_{i=1}^n x_iy_i\right| \leqslant \sqrt{\sum_{i=1}^n {x_i}^2}\times\sqrt{\sum_{i=1}^n {y_i}^2}$$

ou encore :

$$\left(\sum_{i=1}^n x_iy_i\right)^2 \leqslant \left(\sum_{i=1}^n {x_i}^2\right)\times\left(\sum_{i=1}^n {y_i}^2\right)$$

Exemple : Montrer que pour tous $x_1,\dots,x_n\in\mathbb{R}$:

$$\left(\sum_{i=1}^n x_i\right)^2 \leqslant n\left(\sum_{i=1}^n {x_i}^2\right)$$

Il suffit d'appliquer l'inégalité de Cauchy-Schwarz avec tous les y_i égaux à 1.

- Avec le produit scalaire habituel sur $\mathscr{C}([\,a\,;b\,],\mathbb{R})$ (cf. paragraphe I.3.b), l'inégalité de Cauchy-Schwarz devient celle vue au chapitre 22 :

$$\forall (f,g)\in\mathscr{C}([\,a\,;b\,],\mathbb{R})^2, \left|\int_a^b f(t)g(t)\,\mathrm{d}t\right| \leqslant \sqrt{\int_a^b f(t)^2\,\mathrm{d}t}\times\sqrt{\int_a^b g(t)^2\,\mathrm{d}t}$$

ou encore :

$$\forall (f,g)\in\mathscr{C}([\,a\,;b\,],\mathbb{R})^2, \left(\int_a^b f(t)g(t)\,\mathrm{d}t\right)^2 \leqslant \left(\int_a^b f(t)^2\,\mathrm{d}t\right)\times\left(\int_a^b g(t)^2\,\mathrm{d}t\right)$$

Exemple : Soit $f\in\mathscr{C}([\,0\,;1\,],\mathbb{R})$. Montrer que :

$$\left(\int_0^1 f(t)\,\mathrm{d}t\right)^2 \leqslant \int_0^1 f(t)^2\,\mathrm{d}t$$

Il suffit d'appliquer l'inégalité de Cauchy-Schwarz avec g constante égale à 1 sur $[\,0\,;1\,]$ (l'intégrale de g^2 est alors égale à 1).

> Comme on l'a vu deux fois, lorsqu'il n'y a qu'une seule somme ou une seule fonction, on peut tout de même utiliser l'inégalité de Cauchy-Schwarz en prenant l'autre vecteur du produit scalaire constant (pour une fonction, ou avec tous ses termes constants pour un vecteur), mais l'inégalité de Cauchy-Schwarz ne s'applique pas que dans ce cas (simple) mais dans des situations variées : cf. TD.

Remarque : Nous n'avons pas utilisé le caractère « défini » du produit scalaire dans la preuve de l'inégalité de Cauchy-Schwarz (mais nous l'avons utilisé dans le cas d'égalité), nous n'avons utilisé que la positivité (pour la positivité de φ), la symétrie et la bilinéarité (pour l'identité remarquable). Par conséquent, elle est toujours vraie pour une forme bilinéaire, symétrique et positive (pas forcément définie positive). Bien que cela ne soit pas explicitement au programme, donnons une application intéressante en probabilités.

Exemple : Soient X et Y deux variables aléatoires réelles définies sur un même Ω. Alors :

$$|E(XY)| \leqslant \sqrt{E\left(X^2\right)}\times\sqrt{E\left(Y^2\right)}$$

Cette inégalité découle de l'inégalité de Cauchy-Schwarz (pour une forme bilinéaire symétrique positive, voir ci-dessus). En particulier :

- en appliquant cette inégalité à $X-E(X)$ et $Y-E(Y)$ à la place de X et Y (penser à truc et machin), il vient :

$$|E((X-E(X))(Y-E(Y)))| \leqslant \sqrt{E\left((X-E(X))^2\right)}\times\sqrt{E\left((Y-E(Y))^2\right)}$$

c'est-à-dire que $|\mathrm{Cov}(X,Y)|\leqslant\sigma(X)\times\sigma(Y)$: la valeur absolue de la covariance est inférieure au produit des écarts-types. Cela permet par exemple de prouver que le coefficient de corrélation est toujours inférieur à 1.

- En appliquant cette inégalité à $|X|$, où X est une variable aléatoire centrée, et à Y constante égale à 1, on obtient :

$$|E(|X|)| \leqslant \sqrt{E(X^2)}$$

et puisque $E(|X|)\geqslant 0$ par positivité de l'espérance, il vient : $E(|X|)\leqslant\sqrt{V(X)}$.

II.3 Inégalité triangulaire

Revenons dans le monde tranquille des espaces préhilbertiens et des produits scalaires.

Proposition (Inégalité triangulaire). Soit $(x,y) \in E^2$. Alors :

$$\|x+y\| \leqslant \|x\| + \|y\|$$

avec égalité si et seulement si x et y sont colinéaires de même sens, i.e. s'il existe $\lambda \in \mathbb{R}_+$ tel que $x = \lambda y$ ou $y = \lambda x$.

Cette inégalité est parfois aussi appelée inégalité de Minkowski.

Démonstration. On a :

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

Démonstration analogue à celle du chapitre 7.

Or, un réel est inférieur à sa valeur absolue donc $\langle x,y\rangle \leqslant |\langle x,y\rangle| \leqslant \|x\|\|y\|$ d'après l'inégalité de Cauchy-Schwarz. Finalement :

$$\|x+y\|^2 \leqslant \|x\|^2 + 2\|x\|\|y\| + \|y\|^2 = (\|x\| + \|y\|)^2 \qquad \square$$

Par croissance de la racine carrée, et une norme étant positive, on a le résultat voulu. Cherchons à présent quand il y a égalité.

Comme dans le chapitre 7 : comparer deux nombres **positifs** revient à comparer leurs carrés.

L'inégalité triangulaire résulte des deux inégalités $\langle x,y\rangle \leqslant |\langle x,y\rangle| \leqslant \|x\|\|y\|$. Ainsi, on a égalité si et seulement si ces deux inégalités sont des égalités. La première (qu'on notera donc l'inégalité (1)) est une égalité si et seulement si $\langle x,y\rangle \in \mathbb{R}_+$ et la deuxième si et seulement si x et y sont colinéaires d'après le paragraphe précédent.

Le raisonnement par équivalences paraît difficile : on fait un sens et puis l'autre.

Supposons que x et y sont colinéaires de même sens : il existe $\lambda \in \mathbb{R}_+$ tel que $x = \lambda y$ ou $y = \lambda x$. Sans perte de généralité, on suppose qu'il existe $\lambda \in \mathbb{R}_+$ tel que $y = \lambda x$, si bien que $\langle x,y\rangle = \lambda\langle x,x\rangle \geqslant 0$ et x et y sont colinéaires donc il y a égalité dans l'inégalité de Cauchy-Schwarz : les deux conditions sont vérifiées donc il y a égalité dans l'inégalité triangulaire.

Réciproquement, supposons qu'on ait égalité : on a alors égalité dans l'inégalité de Cauchy-Schwarz donc x et y sont colinéaires. Sans perte de généralité, on suppose qu'il existe $\lambda \in \mathbb{R}$ tel que $y = \lambda x$ si bien que $\langle x,y\rangle = \lambda\langle x,x\rangle$. Or, il y a égalité dans l'inégalité (1) donc $\langle x,y\rangle \geqslant 0$: soit $x \neq 0$ alors $\langle x,x\rangle > 0$ donc $\lambda \geqslant 0$, soit $x = 0$ et alors $y = 0 = x$ donc, dans tous les cas, il existe $\lambda \in \mathbb{R}_+^*$ tel que $y = \lambda x$ i.e. x et y sont colinéaires de même sens.

On a les mêmes corollaires que pour les complexes, et les démonstrations sont analogues :

Corollaire. Soit $n \geqslant 2$, soit $(x_1, \dots, x_n) \in E^n$. Alors :

$$\left\|\sum_{k=1}^n x_k\right\| \leqslant \sum_{k=1}^n \|x_k\|$$

Corollaire. Soit $(x_1, x_2) \in E^2$. On a les inégalités suivantes :

- $\|x_1\| - \|x_2\| \leqslant \|x_1 + x_2\|$.
- $\|x_2\| - \|x_1\| \leqslant \|x_1 + x_2\|$.
- $\|x_1\| - \|x_2\| \leqslant \|x_1 - x_2\|$.
- $\|x_2\| - \|x_1\| \leqslant \|x_1 - x_2\|$.

II.4 Norme et distance

Définition (Deuxième année). Une norme sur un espace vectoriel E est une application $N : E \to \mathbb{R}_+$ vérifiant les trois conditions suivantes :

- **(Séparation)** $\forall x \in E, N(x) = 0 \iff x = 0$.
- **(Absolue homogénéïté)** $\forall x \in E, \forall \lambda \in \mathbb{R}, N(\lambda x) = |\lambda| \times N(x)$.
- **(Inégalité triangulaire)** $\forall (x,y) \in E^2, N(x+y) \leqslant N(x) + N(y)$.

On a donc prouvé dans les paragraphes précédents que la norme euclidienne associée à un produit scalaire est bien une norme (d'où son nom).

Par contre, certaines normes ne sont pas des normes issues d'un produit scalaire : on peut prouver qu'une norme est issue d'un produit scalaire si et seulement si elle vérifie l'identité du parallélogramme, cf. exercice 48.

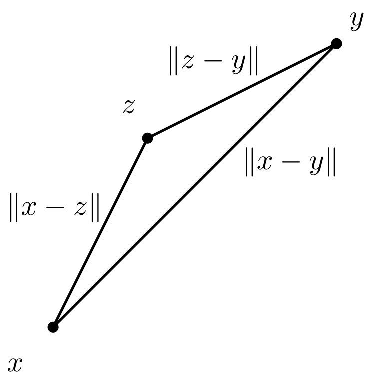
La norme permet de définir rigoureusement la notion de taille, de longueur d'un vecteur, et les trois conditions ci-dessus sont alors totalement naturelles pour que cela soit conforme à l'intuition :

- Le seul vecteur ayant une « taille », une « longueur » nulle est le vecteur nul.
- Quand on multiplie un vecteur par un réel λ, on multiplie la taille par $|\lambda|$.
- Enfin, comme pour une distance dans le plan, on s'attend à ce qu'une distance vérifie l'inégalité triangulaire.

Car, oui, une norme peut servir à définir une distance : plus précisément, si x et y sont deux éléments de E, $\|x-y\|$ représente la distance entre x et y. On peut donc généraliser la notion de distance à un espace vectoriel préhilbertien quelconque et, grâce aux propriétés vérifiées par la norme, elle vérifie les propriétés intuitives (qu'on s'attend à voir vérifiées par une distance, et qui sont même la définition d'une distance générale, mais cela dépasse le cadre du programme) suivante :

- **(Séparation)** $\forall (x,y) \in E^2, \|x-y\| = 0 \iff x = y$.
- **(Inégalité triangulaire)** $\forall (x,y,z) \in E^3, \|x-y\| \leqslant \|x-z\| + \|z-y\|$.

L'inégalité triangulaire ci-contre découle de celle démontrée plus haut en remarquant que $x-y = (x-z)+(z-y)$.



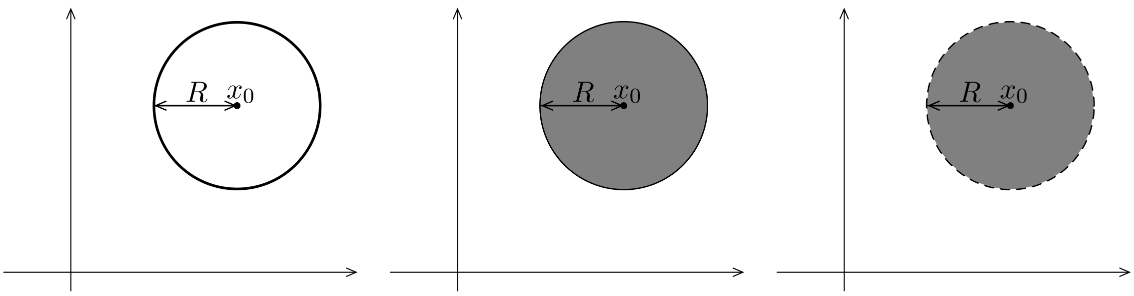
Maintenant qu'on a défini une norme et une distance, on peut définir des objets géométriques de façon analogue à ce qui existe dans le plan ou l'espace :

Définition. Soit $x_0 \in E$, soit $R \in \mathbb{R}_+$. On définit les ensembles suivants :

- $\{x \in E \mid \|x-x_0\| = R\}$ est la sphère de centre x_0 de rayon R. On la note $S(x_0, R)$.
- $\{x \in E \mid \|x-x_0\| \leqslant R\}$ est la boule (fermée) de centre x_0 de rayon R. On la note $B_F(x_0, R)$.
- $\{x \in E \mid \|x-x_0\| < R\}$ est la boule ouverte (c'est-à-dire privée de la sphère) de centre x_0 de rayon R. On la note $B_O(x_0, R)$.

Rappelons qu'un vecteur est dit unitaire lorsqu'il est de norme 1. Par conséquent, $S(0,1)$ est appelée la sphère unité, $B_F(0,1)$ est la boule unité fermée et $B_O(0,1)$ la boule unité ouverte.

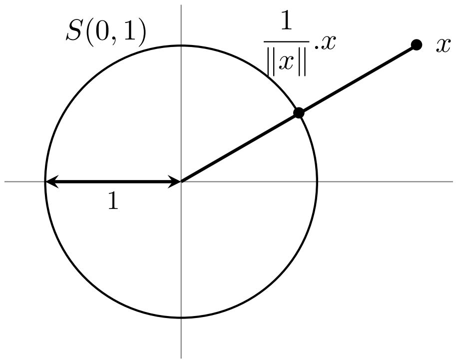
Interprétation géométrique :



Sphère Boule fermée Boule ouverte

Les boules peuvent être moins « rondes et lisses » que ci-contre si on prend des normes différentes : par exemple, la boule unité de $\mathbb{R}^2$ pour la norme infinie (cf. deuxième année) est un carré !

Cette représentation géométrique permet d'illustrer facilement que si x est non nul, alors $x/\|x\|$ est unitaire :



La norme est donc un outil permettant de parler de taille, de longueur, de distance dans des espaces vectoriels. Il faut cependant garder en tête que la norme est liée au produit scalaire : en prenant un produit scalaire différent, la norme change également. Par exemple, pour le produit scalaire canonique sur $\mathbb{R}^2$, pour tout $x = (x_1, x_2) \in \mathbb{R}^2$,

Pour l'instant, on ne sait le faire que dans des espaces préhilbertiens (puisque notre norme est définie grâce à un produit scalaire), vous pourrez faire ça dans d'autres espaces vectoriels l'an prochain !

$$\begin{aligned}\|x\| &= \sqrt{\langle x,x\rangle} \\ &= \sqrt{{x_1}^2 + {x_2}^2}\end{aligned}$$

et on retrouve notre distance bien connue, mais si on prend le produit scalaire

$$\langle\,,\rangle : \left\{\begin{array}{ccc}\mathbb{R}^2 \times \mathbb{R}^2 & \to & \mathbb{R} \\ (x,y) & \mapsto & x^\top \times \begin{pmatrix} 2 & 1 \\ 1 & 2 \end{pmatrix} \times y\end{array}\right.$$

alors on a vu que $\|x\| = \sqrt{2{x_1}^2 + 2x_1x_2 + 2{x_2}^2}$: on a donc une distance différente (ce qui est logique). Par exemple, pour cette nouvelle distance, $\|(1,0)\| = \sqrt{2}$. On voit donc qu'on a parfois des résultats contre-intuitifs, mais il suffit de retenir que tout résultat est relatif au produit scalaire choisi.

Ce sera la même chose avec l'orthogonalité dans le paragraphe suivant : des vecteurs orthogonaux pour un produit scalaire ne le seront peut-être plus pour un autre !

III Vecteurs orthogonaux

Nous arrivons à l'utilisation principale du produit scalaire : pouvoir parler de vecteur orthogonaux, et généraliser cette notion à des espaces vectoriels plus généraux que $\mathbb{R}^2$ ou $\mathbb{R}^3$. On rappelle qu'on se place sur un espace préhilbertien $(E, \langle\,,\rangle)$.

III.1 Définition

Définition. Soit $(x, y) \in E^2$. On dit que x et y sont orthogonaux lorsque $\langle x,y \rangle = 0$. On note alors : $x \perp y$.

Remarque : Le produit scalaire étant symétrique, $\langle x,y \rangle = 0$ si et seulement si $\langle y,x \rangle = 0$. Par conséquent, cela revient au même d'avoir $\langle x,y \rangle = 0$ ou $\langle y,x \rangle = 0$ (ou différent de 0 si on veut prouver que les vecteurs ne sont pas orthogonaux).

Exemple : On a déjà vu que, pour tout x, $\langle x,0_E \rangle = 0$ c'est-à-dire que le vecteur nul est orthogonal à tout vecteur. C'est même le seul :

Proposition. Le vecteur nul est le seul vecteur orthogonal à tout vecteur. En d'autres termes, si $x \in E$ vérifie :

$$\forall y \in E, \langle x,y \rangle = 0$$

alors $x = 0_E$.

Démonstration. Il suffit de voir que x est alors orthogonal à lui-même et alors $\langle x,x \rangle = 0$ et on conclut à l'aide de la définie positivité du produit scalaire.

Il est donc équivalent de dire que seul le vecteur nul est orthogonal à lui-même.

Remarque : Là aussi, il faut bien comprendre que l'orthogonalité dépend du produit scalaire choisi : deux vecteurs orthogonaux pour un produit scalaire ne le seront peut-être plus pour un autre.

Exemple : Pour le produit scalaire canonique sur $\mathbb{R}^2$, les vecteurs $e_1 = (1, 0)$ et $e_2 = (0, 1)$ sont orthogonaux puisque

$$\begin{aligned} \langle e_1,e_2 \rangle &= 1 \times 0 + 0 \times 2 \\ &= 0 \end{aligned}$$

tandis que les vecteurs $e_1 = (1, 0)$ et $e_3 = (1, -2)$ ne le sont pas car $\langle e_1,e_2 \rangle = 1$. Cependant, si on prend encore le produit scalaire

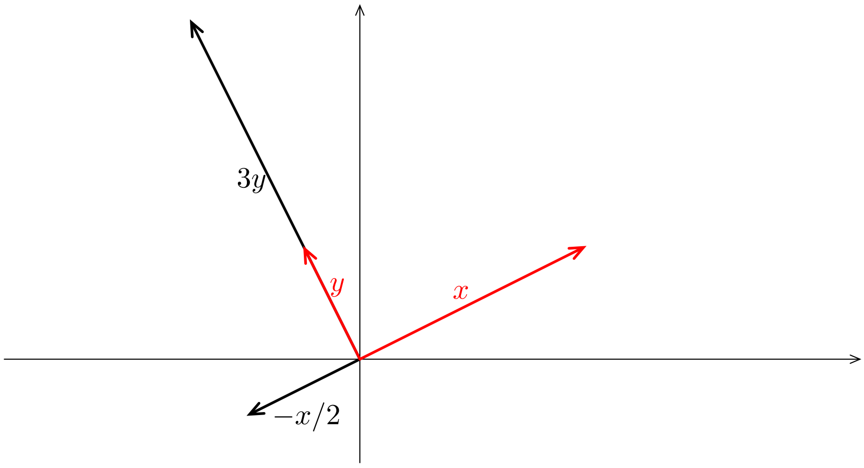
$$\langle \, , \rangle : \begin{cases} \mathbb{R}^2 \times \mathbb{R}^2 & \to & \mathbb{R} \\ (x, y) & \mapsto & x^\top \times \begin{pmatrix} 2 & 1 \\ 1 & 2 \end{pmatrix} \times y \end{cases}$$

alors e_1 et e_2 ne sont plus orthogonaux car leur produit scalaire vaut 1 mais e_1 et e_3 le sont car leur produit scalaire est nul. Cela peut paraître étrange car, si on représente ces vecteurs dans le plan, « cela se voit bien qu'ils ne sont pas orthogonaux ». Et, en effet, ils ne le sont pas pour le produit scalaire canonique, mais si on change de produit scalaire, alors ils deviennent orthogonaux !

Proposition. Soient $(x, y) \in E^2$ et $(\lambda, \mu) \in \mathbb{R}^2$. Si x et y sont orthogonaux, alors λx et μy sont orthogonaux.

Démonstration. Par bilinéarité du produit scalaire, $\langle \lambda x,\mu y \rangle = \lambda\mu \langle x,y \rangle = 0$.

Remarque : Ce résultat se voit très bien : quand x et y sont orthogonaux, ils restent orthogonaux quand on les multiplie par des scalaires.



Ce résultat permettra de prouver que l'orhogonal d'une partie est un espace vectoriel : cf. paragraphe IV.1.

III.2 Familles orthogonales et orthonormées/orthonormales

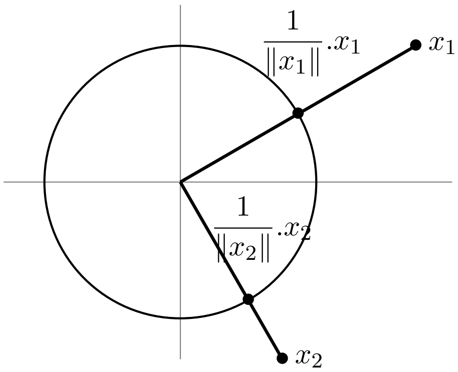
Définition. Une famille de vecteurs est :

- orthogonale si les vecteurs de cette famille sont deux à deux orthogonaux.
- orthonormée ou orthonormale si elle est orthogonale et si tous les vecteurs sont unitaires, c'est-à-dire si les vecteurs de cette famille sont deux à deux orthogonaux et de norme 1.

Remarques :

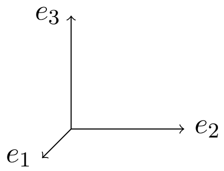
- En d'autres termes :
  - ⋆ $(x_i)_{i\in I}$ est orthogonale si : $\forall i \neq j, x_i \perp x_j$ i.e. $\langle x_i, x_j\rangle = 0$.
  - ⋆ $(x_i)_{i\in I}$ est orthonormale si : $\forall i \neq j, x_i \perp x_j$ et $\forall i \in I, \|x_i\| = \sqrt{\langle x_i, x_i\rangle} = 1$, cette dernière condition étant équivalente à : $\forall i \in I, \langle x_i, x_i\rangle = 1$, une norme étant positive (et donc $\|x_i\| = 1 \iff \|x_i\|^2 = 1$).
- Une famille orthonormale est formée de vecteurs tous non nuls, mais une famille orthogonale peut contenir le vecteur nul.
- Un vecteur non nul divisé par sa norme étant unitaire (cf. paragraphe II.1), on peut facilement obtenir une famille orthonormale à partir d'une famille orthogonale formée de vecteurs **tous non nuls** : en effet, si $(x_i)_{i\in I}$ est une famille orthogonale de vecteurs tous non nuls, alors la famille $(x_i/\|x_i\|)_{i\in I}$ est encore orthogonale (cf. paragraphe III.1), et est formée de vecteurs unitaires (cf. paragraphe II.1) donc est une famille orthonormale.

Ne pas oublier de supposer $i \neq j$: seul le vecteur nul est orthogonal à lui-même (même si une famille orthogonale, contrairement à une famille orthonormale, peut contenir le vecteur nul).



Exemples :

- Pour le produit scalaire canonique sur $\mathbb{R}^n$, la base canonique $(e_1, \dots, e_n)$ est une famille orthonormale (c'est donc une base orthonormale, cf. paragraphe III.3). Ci-dessous un dessin dans $\mathbb{R}^3$:



Attention, encore une fois, ce n'est plus vrai si on prend un produit scalaire différent.

- Montrons que les matrices élémentaires $(E_{i,j})_{1\leqslant i,j\leqslant n}$ forment une famille orthonormale de $\mathscr{M}_n(\mathbb{R})$ muni de son produit scalaire canonique. Rappelons (cf. chapitre 21) que, pour tous i,j,k,l, $E_{i,j} \times E_{k,l} = \delta_{j,k} E_{i,l}$. Soient donc (i,j) et (k,l) appartenant à $[\![1\,;\,n]\!]^2$. On a :

$$\begin{aligned}
\langle E_{i,j}, E_{k,l}\rangle &= \operatorname{tr}\left(E_{i,j}^\top \times E_{k,l}\right)\\
&= \operatorname{tr}\left(E_{j,i} \times E_{k,l}\right)\\
&= \operatorname{tr}\left(\delta_{i,k} E_{j,l}\right)\\
&= \delta_{i,k} \operatorname{tr}\left(E_{j,l}\right)
\end{aligned}$$

par linéarité de la trace. Il y a plusieurs cas :

  - ⋆ Si le couple (i,j) est distinct du couple (k,l) : alors soit $i \neq k$, et alors $\delta_{i,k} = 0$, soit $j \neq l$, et alors $E_{j,l}$ est de trace nulle puisque son seul élément non nul n'est pas sur la diagonale. Dans tous les cas, $\langle E_{i,j}, E_{k,l}\rangle = 0$: la famille est orthogonale.
  - ⋆ Si $(i,j) = (k,l)$, alors $\delta_{i,k} = 1$ et $\operatorname{tr}(E_{j,l}) = 1$ puisque $j = l$ et donc l'unique 1 de la matrice est sur la diagonale. On en déduit que $\langle E_{i,j}, E_{i,j}\rangle = 1$: la famille est orthonormale.

- Pour tout $n \in \mathbb{N}$, notons $a_n : t \mapsto \cos(nt)$ et, pour tout $n \in \mathbb{N}^*$, notons $b_n : t \mapsto \sin(nt)$. Montrons que $(a_n)_{n\in\mathbb{N}} \cup (b_n)_{n\in\mathbb{N}^*}$ est une famille orthogonale pour le produit scalaire habituel sur $\mathscr{C}([\,0\,;2\pi\,], \mathbb{R})$ vu au paragraphe I.3.b.

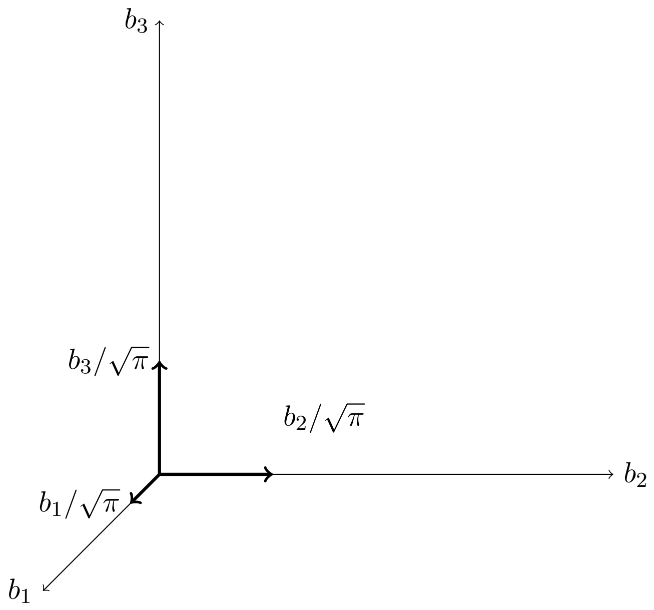
  - ⋆ Soient n et p deux entiers naturels distincts. Alors $n - p \neq 0$ et l'un au moins des entiers n et p est non nul donc $n + p \neq 0$. Dès lors,

$$\begin{aligned}
\langle a_n, a_p\rangle &= \int_0^{2\pi} \cos(nt)\cos(pt)\,\mathrm{d}t\\
&= \frac{1}{2}\int_0^{2\pi} \cos((n-p)t) - \cos((n+p)t)\,\mathrm{d}t\\
&= \frac{1}{2}\left[\frac{\sin((n+p)t)}{n+p} + \frac{\sin((n-p)t)}{n-p}\right]_0^{2\pi}\\
&= 0
\end{aligned}$$

  - ⋆ On prouve de même que si n et p sont deux entiers naturels distincts, alors $\langle b_n, b_p\rangle = 0$.
  - ⋆ Enfin, si $n \in \mathbb{N}$ et $p \in \mathbb{N}^*$ sont deux entiers (pas forcément distincts), on prouve de même que $\langle a_n, b_p\rangle = 0$ (il faut tout de même faire deux cas selon que $n = p$ ou non, exo).

C'est donc bien une famille orthogonale, on peut donc représenter les trois premiers termes de la façon suivante :

Pour visualiser les autres, c'est plus difficile car on est en dimension infinie...



Cependant, ce n'est pas une famille orthonormale :

- ⋆ Soit $n \in \mathbb{N}^*$.

$$\begin{aligned}\langle b_n, b_n\rangle &= \int_0^{2\pi} \sin^2(nt)\,\mathrm{d}t \\ &= \int_0^{2\pi} \frac{1-\cos(2nt)}{2}\,\mathrm{d}t \\ &= \pi - \frac{1}{2}\left[\frac{\sin(2nt)}{2n}\right]_0^{2\pi} \\ &= \pi\end{aligned}$$

si bien que $\|b_n\| = \sqrt{\pi}$. On prouve de même (attention, séparer les cas $n = 0$ et $n \neq 0$) que $\|a_0\| = \sqrt{2\pi}$ et, si $n \neq 0$, $\|a_n\| = \sqrt{\pi}$. Comme on l'a vu, on peut « transformer » cette famille en famille orthonormale : la famille $\left(a_0/\sqrt{2\pi}\right) \cup (a_n/\sqrt{\pi})_{n\in\mathbb{N}^*} \cup (b_n/\sqrt{\pi})_{n\in\mathbb{N}^*}$ est orthonormale.

Théorème (Théorème de Pythagore). Soient $n \geqslant 1$ et $(x_1, \dots, x_n)$ une famille orthogonale de E. Alors :

$$\left\|\sum_{i=1}^n x_i\right\|^2 = \sum_{i=1}^n \|x_i\|^2$$

En particulier, si x et y sont deux vecteurs orthogonaux : $\|x+y\|^2 = \|x\|^2 + \|y\|^2$.

Démonstration. Découle de l'identité remarquable vue au paragraphe II.1 puisque tous les produits scalaires sont nuls.

Remarque : Ce n'est pas équivalent, c'est-à-dire que si on a l'égalité ci-dessus, la famille n'est pas forcément orthogonale (sauf dans le cas $n = 2$ où, là, il y a équivalence).

III.3 Bases orthonormales

III.3.a Résultat préliminaire

On le voit avec l'exemple des fonctions trigonométriques : il est difficile de dessiner plus de trois vecteurs orthogonaux en dimension 3. Cela découle du résultat suivant :

Proposition. Une famille orthogonale de vecteurs **tous non nuls** est libre. En particulier, une famille orthonormale est libre.

On avait prouvé ce résultat dans le cas particulier de $\mathscr{C}([0\,;1],\mathbb{R})$ dans l'exercice 22 du chapitre 28.

Démonstration. Soit F une famille orthogonale composée de vecteurs tous non nuls. Soient $n \geqslant 1$ et $x_1, \dots, x_n$ des vecteurs distincts de F. Soit $(\lambda_1, \dots, \lambda_n) \in \mathbb{R}^n$ tel que $\lambda_1 x_1 + \cdots + \lambda_n x_n = 0$. Soit $i \in [\![1\,;n]\!]$.

$$\langle \lambda_1 x_1 + \cdots + \lambda_n x_n, x_i \rangle = \langle 0, x_i \rangle = 0$$

Par linéarité à gauche du produit scalaire :

$$\sum_{j=1}^{n} \lambda_j \langle x_j, x_i \rangle = 0$$

□

Or, la famille est orthogonale donc, parmi les produits scalaires ci-dessus, tous sont nuls sauf celui d'indice i qui vaut $\|x_i\|^2$ si bien que $\lambda_i \|x_i\|^2 = 0$. Or, $x_i \neq 0$ donc $\|x_i\| \neq 0$ si bien que $\lambda_i = 0$: i étant quelconque, la famille $(x_1, \dots, x_n)$ est libre, donc tout sous-famille finie est libre, F est libre.

Remarque : Il en découle qu'une famille orthogonale (de vecteurs tous non nuls), dans un espace de dimension n, est de cardinal inférieur ou égal à n. Si on a une famille à n éléments, on a donc une base.

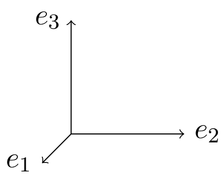
III.3.b Bases orthonormales

Définition. On appelle base orthonormée ou orthonormale... une base qui est orthonormale.

Remarque : En d'autres termes, une base orthonormale (qu'on abrège souvent en BON) est une base de vecteurs deux à deux orthogonaux, et unitaires.

Exemples : On a prouvé précédemment les résultats suivants :

- La base canonique $(e_1, \dots, e_n)$ est une base orthonormale.



- La base canonique $(E_{i,j})_{1 \leqslant i,j \leqslant n}$ de $\mathscr{M}_n(\mathbb{R})$ est aussi une base orthonormale.

Rappelons le résultat suivant, vu au chapitre 28 :

Théorème (Existence et unicité de la décomposition selon une base). Soit $B = (e_i)_{i \in I}$ base de E. Alors, pour tout $x \in E$, il existe une unique famille de scalaires presque nulle $(\alpha_i)_{i \in I}$ telle que $x = \displaystyle\sum_{i \in I} \alpha_i e_i$.

Les α_i sont alors appelées coordonnées de x dans la base B. Dans le cas où la base B est orthonormale, on a une expression explicite de ces coordonnées :

Théorème (Expression des coordonnées dans une base orthonormale). Soit $B = (e_i)_{i\in I}$ une base orthonormale de E. Alors, pour tout $x \in E$, les coordonnées de x dans la base B sont les $(\langle x,e_i\rangle)_{i\in I}$. En particulier :

$$x = \sum_{i\in I} \langle x,e_i\rangle e_i \qquad \text{et} \qquad \|x\| = \sqrt{\sum_{i\in I}\langle x,e_i\rangle^2}$$

Attention à ne pas oublier les vecteurs dans x et à ne pas les mettre dans $\|x\|$: x est un vecteur, $\|x\|$ est un réel.

Démonstration. Soit $x \in E$. Notons $(x_i)_{i\in I}$ les coordonnées de x, si bien que :

$$x = \sum_{i\in I} x_i e_i$$

Soit $j \in I$.

$$\begin{aligned} \langle x,e_j\rangle &= \langle \sum_{i\in I} x_i e_i, e_j\rangle \\ &= \sum_{i\in I} x_i \langle e_i,e_j\rangle \end{aligned}$$

Linéarité à gauche du produit scalaire.

Or, la famille $(e_i)_{i\in I}$ est orthonormale donc $\langle e_i,e_j\rangle = \delta_{i,j}$: il en découle que $\langle x,e_j\rangle = x_j$, c'est-à-dire que les coordonnées sont bien les $(\langle x,e_j\rangle)_{j\in I}$. L'expression de x en découle. Dès lors, par bilinéarité du produit scalaire :

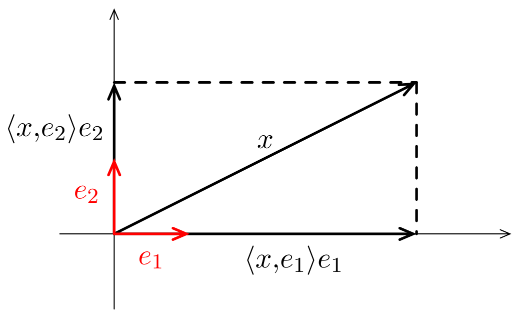
$$\begin{aligned} \|x\|^2 &= \langle x,x\rangle \\ &= \langle \sum_{i\in I} x_i e_i, \sum_{i\in I} x_i e_i\rangle \\ &= \sum_{i\in I}\sum_{j\in I} x_i x_j \langle e_i,e_j\rangle \\ &= \sum_{i\in I}\sum_{j\in I} x_i x_j \delta_{i,j} \\ &= \sum_{i\in I} {x_i}^2 \end{aligned}$$

Changer l'indice de sommation !

□

ce qui permet de conclure.

Remarque : Attention, ce résultat n'est pas valable si la base n'est pas orthonormale, et en particulier si elle est uniquement orthogonale. De plus, ce résultat se voit bien géométriquement :



L'idée sous-jacente est celle de projeté orthogonal : cf. paragraphe IV.3.

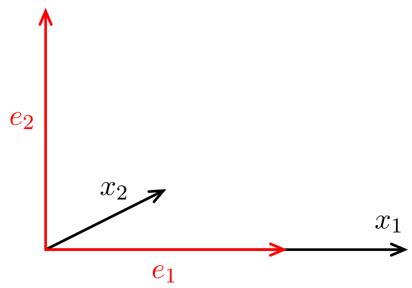
III.3.c Procédé d'orthonormalisation de Gram-Schmidt

Le procédé (ou algorithme car c'est en effet un procédé algorithmique) d'orthonormalisation de Gram-Schmidt est un procédé permettant de « rendre orthonormale » une base ou, plus simplement, une famille libre.

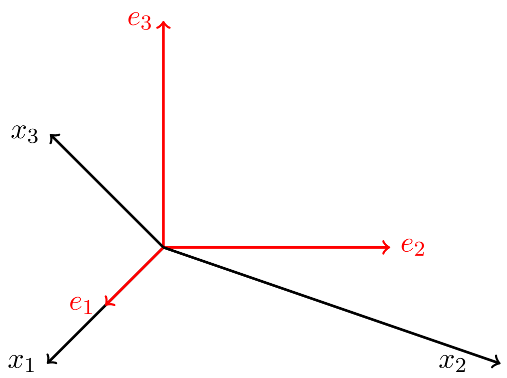
Proposition (Procédé d'orthonormalisation de Gram-Schmidt). Soit $(x_1, \dots, x_n)$ une famille libre de E. Il existe une famille orthonormale $(e_1, \dots, e_n)$ telle que, pour tout $k \in [\![1 \,;\, n]\!]$, $\text{Vect}(e_1, \dots, e_k) = \text{Vect}(x_1, \dots, x_k)$.

Remarques :

- Faisons un dessin dans $\mathbb{R}^2$ (i.e. dans le plan) :



  x_1 et x_2 forment une famille libre (et même une base de $\mathbb{R}^2$ car famille libre à deux éléments en dimension 2) mais ne forment pas une famille orthonormale (ni même orthogonale). On commence par x_1, qu'on divise par sa norme pour le rendre unitaire, puis on choisit un vecteur e_2 orthogonal à e_1 et unitaire : (e_1, e_2) est orthonormale, $\text{Vect}(e_1) = \text{Vect}(x_1)$ et $\text{Vect}(e_1, e_2) = \text{Vect}(x_1, x_2) = \mathbb{R}^2$.
- Faisons un dessin dans le cas $n = 3$ (i.e. dans l'espace) :



  On a bien (e_1, e_2, e_3) orthonormale, $\text{Vect}(x_1) = \text{Vect}(e_1)$, $\text{Vect}(x_1, x_2) = \text{Vect}(e_1, e_2)$ (le plan engendré par x_1 et x_2) et enfin $\text{Vect}(x_1, x_2, x_3) = \text{Vect}(e_1, e_2, e_3)$ ($\mathbb{R}^3$ tout entier). Ce procédé est en fait extrêmement simple à comprendre :

  ⋆ On divise le premier vecteur par sa norme pour le rendre unitaire.
  ⋆ On prend ensuite un vecteur e_2 dans $\text{Vect}(x_1, x_2)$ unitaire et orthogonal à e_1.
  ⋆ On prend ensuite un vecteur e_3 dans $\text{Vect}(x_1, x_2, x_3)$ orthogonal à e_1 et e_2.
  ⋆ etc.

- Rien qu'avec ces dessins, on voit qu'il n'y a pas unicité : dans l'exemple de $\mathbb{R}^2$, on pourrait prendre e_1 et $-e_2$ par exemple, ou $-e_1$ et $-e_2$ etc. Ce sont cependant les seuls cas de non unicité (voir ci-dessous) : remplacer un vecteur par son opposé. On peut imposer une condition d'unicité en demandant que, pour tout k, $\langle x_k, e_k \rangle$ soit positif : on en reparle.

Démonstration. Montrons par récurrence sur $k \in [\![1 \,;\, n]\!]$ qu'une telle famille existe.

- Si $k \in [\![1 \,;\, n]\!]$, notons H_k : « il existe une famille $(e_1, \dots, e_k)$ orthonormale telle que, pour tout $i \in [\![1 \,;\, k]\!]$, $\text{Vect}(e_1, \dots, e_i) = \text{Vect}(x_1, \dots, x_i)$ ».

On généralise aisément (avec une récurrence classique au lieu d'une récurrence finie) à une suite de vecteurs libres (si E est de dimension infinie) : plus précisément, si $(x_n)_{n \in \mathbb{N}}$ est une famille libre, alors il existe une famille orthonormale $(x_n)_{n \in \mathbb{N}}$ telle que, pour tout $n \in \mathbb{N}$, $\text{Vect}(e_1, \dots, e_n) = \text{Vect}(x_1, \dots, x_n)$. Nous appliquerons principalement ce procédé en dimension finie, donc j'ai préféré mettre le cas d'une famille finie dans la proposition.

- x_1 est non nul car $(x_1, \dots, x_n)$ est libre donc ne contient pas le vecteur nul, si bien que $e_1 = x_1/\|x_1\|$ est unitaire. De plus, e_1 est colinéaire à x_1 et est non nul donc $\text{Vect}(e_1) = \text{Vect}(x_1)$: H_1 est vraie.

Il n'y a pas d'orthogonalité à vérifier puisqu'il n'y a qu'un vecteur.

- Soit $k \in [\![1\,;\,n-1]\!]$. Supposons H_k vraies et prouvons que H_{k+1} est vraie. Par hypothèse de récurrence, il existe $(e_1, \dots, e_k)$ orthonormale telle que, pour tout $i \in [\![1\,;\,k]\!]$, $\text{Vect}(e_1, \dots, e_i) = \text{Vect}(x_1, \dots, x_i)$. Il suffit donc de trouver e_{k+1} unitaire orthogonal à $e_1, \dots, e_k$ et vérifiant $\text{Vect}(e_1, \dots, e_{k+1}) = \text{Vect}(x_1, \dots, x_{k+1})$.

  ⋆ **Analyse :** Si e_{k+1} convient, alors $e_{k+1} \in \text{Vect}(x_1, \dots, x_{k+1})$ donc il existe $(\alpha_1, \dots, \alpha_{k+1}) \in \mathbb{R}^{k+1}$ tel que

Même si on l'utilise la plupart du temps pour cela, un raisonnement par analyse synthèse ne sert pas qu'à montrer l'existence et l'unicité d'un objet : on l'utilise lorsqu'un raisonnement par équivalences est difficile ou délicat (cf. chapitre 1). Ici, l'analyse-synthèse permet de montrer qu'il existe exactement deux solutions : voir plus bas.

$$e_{k+1} = \sum_{i=1}^{k} \alpha_i x_i + \alpha_{k+1} x_{k+1}$$

Or, $\displaystyle\sum_{i=1}^{k} \alpha_i x_i \in \text{Vect}(x_1, \dots, x_k) = \text{Vect}(e_1, \dots, e_k)$ donc il existe $(\beta_1, \dots, \beta_k) \in \mathbb{R}^k$ tel que $\displaystyle\sum_{i=1}^{k} \alpha_i x_i = \sum_{i=1}^{k} \beta_i e_i$ si bien que

$$e_{k+1} = \sum_{i=1}^{k} \beta_i e_i + \alpha_{k+1} x_{k+1}$$

Or, e_{k+1} est orthogonal à $(e_1, \dots, e_k)$ donc, pour tout $j \in [\![1\,;\,k]\!]$,

$$\begin{aligned} \langle e_{k+1}, e_j \rangle = 0 &= \sum_{i=1}^{k} \beta_i \underbrace{\langle e_i, e_j \rangle}_{=\delta_{i,j}} + \alpha_{k+1} \langle x_{k+1}, e_j \rangle \\ &= \beta_j + \alpha_{k+1} \langle x_{k+1}, e_j \rangle \end{aligned}$$

On utilise la linéarité à gauche du produit scalaire.

si bien que $\beta_j = -\alpha_{k+1} \langle x_{k+1}, e_j \rangle$.
En d'autres termes, $e_{k+1} = \alpha_{k+1} u_{k+1}$ où

$$u_{k+1} = x_{k+1} - \sum_{i=1}^{k} \langle x_{k+1}, e_i \rangle e_i$$

Or, e_{k+1} est unitaire donc $\|e_{k+1}\| = 1 = |\alpha_{k+1}| \times \|u_{k+1}\|$ si bien que $\alpha_{k+1} = \pm 1/\|u_{k+1}\|$ c'est-à-dire que

$$e_{k+1} = \frac{\pm 1}{\|u_{k+1}\|}.u_{k+1} \qquad \text{où} \qquad u_{k+1} = x_{k+1} - \sum_{i=1}^{k} \langle x_{k+1}, e_i \rangle e_i$$

  ⋆ **Synthèse :** Soit

$$u_{k+1} = x_{k+1} - \sum_{i=1}^{k} \langle x_{k+1}, e_i \rangle e_i$$

et soit $e_{k+1} = \dfrac{1}{\|u_{k+1}\|}.u_{k+1}$. En effet, u_{k+1} est non nul : si $u_{k+1} = 0$, alors x_{k+1} est CL de $e_1, \dots, e_k$ donc appartient à $\text{Vect}(e_1, \dots, e_k) = \text{Vect}(x_1, \dots, x_k)$ ce qui est absurde car $(x_1, \dots, x_n)$ est une famille libre. On en déduit que $u_{k+1} \neq 0$ si bien que $e_{k+1} = u_{k+1}/\|u_{k+1}\|$ est bien défini et unitaire. Il est de plus orthogonal à $e_1, \dots, e_k$: en effet, pour tout $j \in [\![1\,;\,k]\!]$,

$$\begin{aligned}\langle e_{k+1},e_j\rangle &= \langle x_{k+1},e_j\rangle - \sum_{i=1}^{k}\langle x_{k+1},e_i\rangle\underbrace{\langle e_i,e_j\rangle}_{=\delta_{i,j}}\\ &= \langle x_{k+1},e_j\rangle - \langle x_{k+1},e_j\rangle\\ &= 0\end{aligned}$$

c'est-à-dire que e_{k+1} est unitaire, orthogonal à $e_1,\dots,e_k$, et $(e_1,\dots,e_k)$ est une famille orthonormale donc $(e_1,\dots,e_{k+1})$ est orthonormale : e_{k+1} convient.

Montrons enfin que $\text{Vect}(e_1,\dots,e_{k+1}) = \text{Vect}(x_1,\dots,x_{k+1})$. Les familles $(e_1,\dots,e_{k+1})$ et $(x_1,\dots,x_{k+1})$ sont libres (la première car orthonormale et la deuxième par hypothèse) donc ces deux espaces sont de dimension $k+1$: ils ont la même dimension, il suffit donc de prouver que l'un des deux est inclus dans l'autre, ce qui est immédiat puisque, pour tout $i \in [\![1 ; k]\!]$,

$$e_i \in \text{Vect}(e_1,\dots,e_k) = \text{Vect}(x_1,\dots,x_k) \subset \text{Vect}(x_1,\dots,x_{k+1})$$

et

$$e_{k+1} = \frac{1}{\|u_{k+1}\|}\left(x_{k+1} - \sum_{i=1}^{k}\langle x_{k+1},e_i\rangle e_i\right) \in \text{Vect}(e_1,\dots,e_k,x_{k+1}) = \text{Vect}(x_1,\dots,x_{k+1})$$

□

Finalement, $\text{Vect}(e_1,\dots,e_{k+1}) \subset \text{Vect}(x_1,\dots,x_{k+1})$ ce qui permet de conclure : H_{k+1} est vraie.

- D'après le principe de récurrence, H_k est vraie pour tout $k \in [\![1 ; n]\!]$.

Remarques :

- On a même montré le résultat annoncé plus haut : les seules solutions sont les vecteurs $e_k = \pm u_k/\|u_k\|$ où, pour tout k,

$$u_k = x_k - \sum_{i=1}^{k-1}\langle x_k,ei\rangle e_i$$

  On peut même imposer l'unicité si on demande que $\langle e_k,x_k\rangle > 0$ pour tout k : cela revient à demander que, pour tout k, les bases $(x_1,\dots,x_k)$ et $(e_1,\dots,e_k)$ de $\text{Vect}(e_1,\dots,e_k)$ aient la même orientation.
- Car, oui, pour tout k, ce sont bien des bases de $\text{Vect}(e_1,\dots,e_k)$ car libres et génératrices. En particulier, si $(x_1,\dots,x_n)$ est une base de E, alors $(e_1,\dots,e_n)$ est aussi une base de E : on peut donc « transformer » une base en base orthonormale grâce à cet algorithme.
- Dans la synthèse, on a pris $u_k/\|u_k\|$ et non pas son opposé (ou les deux) car le but de la démonstration était de prouver l'existence d'un vecteur qui convenait.
- De même, en pratique, on cherche en général à donner des vecteurs qui conviennent, ou simplement à prouver l'existence d'une famille qui convient : il est inutile de se souvenir de toutes les solutions.
- L'expression de u_k est totalement naturelle une fois qu'on connaît l'expression du projeté orthogonal sur un sous-espace de dimension finie, cf. paragraphe IV.3.
- Si la famille est déjà orthogonale, alors $e_1 = x_1/\|x_1\|$ et il découle de la récurrence que $u_{k+1} = x_{k+1}$ pour tout k donc que $e_{k+1} = x_{k+1}/\|x_{k+1}\|$: l'algorithme ci-dessus ne fait que diviser les vecteurs par leur norme. En particulier, si elle est déjà orthonormale, elle ne fait rien, puisque la famille est déjà unitaire ! Nous utiliserons ce point dans la démonstration du théorème de la base orthonormale incomplète.

- Chaque vecteur e_k est CL de $(x_1, \dots, x_k)$. En d'autres termes, les coordonnées selon $e_{k+1}, \dots, e_n$ sont nulles. Si $B_1 = (x_1, \dots, x_n)$ est une base de E, alors $B_2 = (e_1, \dots, e_n)$ est une base (orthonormale) de E, et ce qui précède implique donc que la matrice de passage P_{B_1,B_2} constituée des coordonnées des vecteurs de B_2 dans la base B_1 est triangulaire supérieure. On utilise ce résultat par exemple dans l'exercice 52.

III.3.d Exemples

Les exercices demandant d'utiliser (ou de redémontrer en adaptant) le procédé d'orthonormalisation de Gram-Schmidt sont la plupart du temps de trois types :

- on demande d'orthonormaliser une famille libre explicite en petite dimension (3 ou 4 grand maximum). Pour ces cas de figures simples (qui sont la majorité des cas), il y a différentes sectes : on peut utiliser la formule explicite des vecteurs u_k directement, mais on peut également chercher de proche en proche u_k sous la forme

$$u_k = x_k - \sum_{i=1}^{k-1} \alpha_i e_i$$

  puis chercher les α_i pour que u_k soit orthogonal à $e_1, \dots, e_{k-1}$. On peut même se contenter de chercher de proche en proche, de la même façon, les vecteurs $u_1, \dots, u_k$, qui vont former une famille orthogonale, puis normaliser à la fin. Quelle que soit la stratégie choisie, on obtient la même chose et on divise ensuite par $\|u_k\|$ pour obtenir e_k.
- on demande juste d'appliquer le résultat sous sa forme théorique, c'est-à-dire qu'il suffit de dire que, d'après le procédé de Gram-Schmidt, on peut se donner une famille orthonormale vérifiant les bonnes conditions.
- on demande d'orthonormaliser une famille infinie libre, ou on impose une condition d'unicité. On a déjà vu qu'à chaque étape, il y a deux solutions possibles, donc si on impose une condition particulière, cela permet souvent de prouver l'unicité de la famille. On trouve en général la condition : $\forall k, \langle x_k, e_k \rangle > 0$. Il y a parfois des variantes pour ce genre d'exos : on demande par exemple de prouver l'existence et l'unicité d'une famille orthogonale de polynômes (P_n) de coefficient dominant égal à 1 telle que, pour tout n, P_n soit de degré n. Dans tous les cas, deux méthodes : refaire (en adaptant) la preuve ci-dessus, ou utiliser l'orthogonal d'une partie, cf. paragraphe IV.2.

Attention, quand on manipule à la fois des polynômes et des produits scalaires (ce qui peut tout de même arriver assez souvent), le mot « unitaire » devient ambigu : désigne-t-il un polynôme de norme 1 (pour le produit scalaire choisi) ou un polynôme de coefficient dominant égal à 1 ? On se mettra d'accord au début de l'exercice (par exemple, on ne parlera que de polynômes de coefficient dominant égal à 1 et on gardera le terme unitaire pour les vecteurs de norme 1) et on fera attention de ne pas changer d'avis en chemin.

Donnons différents exemples du premier et du troisième cas de figure (nous verrons des exemples du deuxième en TD et dans le théorème de la base orthonormale incomplète, voir ci-dessous).

Exemples :

- Donner une base orthonormale (pour le produit scalaire canonique) de H, l'hyperplan de $\mathbb{R}^3$ d'équation $5x - 3y + z = 0$.

  De même que dans le chapitre 28, on trouve que $x_1 = (0, 1, 3)$ et $x_2 = (1, 0, -5)$ forment une base de H. Appliquons l'algorithme d'orthonormalisation de Gram-Schmidt. Posons $e_1 = x_1/\|x_1\|$. Or, $\|x_1\| = \sqrt{10}$ donc

$$e_1 = \frac{1}{\sqrt{10}}(0, 1, 3)$$

  Posons à présent $u_2 = x_2 - \langle x_2, e_1 \rangle e_1$. On a :

$$\langle x_2, e_1 \rangle = \frac{-15}{\sqrt{10}}$$

si bien que

$$\begin{aligned} u_2 &= (1,0,-5) + \frac{15}{\sqrt{10}} \times \frac{1}{\sqrt{10}}(0,1,3) \\ &= \frac{1}{10}(10,15,-5) \\ &= \frac{1}{2}(2,3,-1) \end{aligned}$$

De plus,

$$\begin{aligned} \|u_2\| &= \frac{1}{2}\sqrt{2^2+3^2+(-1)^2} \\ &= \frac{1}{2} \times \sqrt{14} \end{aligned}$$

Posons finalement

$$\begin{aligned} e_2 &= \frac{2}{\sqrt{14}} u_2 \\ &= \frac{1}{\sqrt{14}}(2,3,-1) \end{aligned}$$

(e_1, e_2) est une base orthonormale de H.

- On se place sur $\mathbb{R}_2[X]$ muni du produit scalaire

$$\langle\,,\rangle : \begin{cases} \mathbb{R}_2[X]^2 & \to & \mathbb{R} \\ (P,Q) & \mapsto & \displaystyle\int_{-1}^{1} P(t)Q(t)\,\mathrm{d}t \end{cases}$$

Donnons une base orthonormale de cet espace. Appliquons l'orthonormalisation de Gram-Schmidt à la base canonique $(1, X, X^2)$ (qui n'est pas orthonormale pour ce produit scalaire).

Posons $x_1 = 1, x_2 = X$ et $x_3 = X^2$ les éléments de la base canonique. $\|x_1\|^2 = \langle 1{,}1\rangle = 2$: posons $e_1 = x_1/\|x_1\| = 1/\sqrt{2}$. À présent, soit

$$u_2 = x_2 - \langle x_2{,}e_1\rangle e_1$$

Or :

$$\begin{aligned} \langle x_2{,}e_1\rangle &= \frac{1}{\sqrt{2}}\int_{-1}^{1} t\,\mathrm{d}t \\ &= 0 \end{aligned}$$

Par conséquent, $u_2 = X$ et

$$\begin{aligned} \|u_2\|^2 &= \langle X{,}X\rangle \\ &= \int_{-1}^{1} t^2\,\mathrm{d}t \\ &= \frac{2}{3} \end{aligned}$$

Posons donc $e_2 = X \times \sqrt{\dfrac{3}{2}}$. Enfin, posons

$$u_3 = x_3 - \langle x_3,e_1\rangle e_1 - \langle x_3,e_2\rangle e_2$$

On a tout d'abord :

$$\begin{aligned}\langle x_3,e_1\rangle &= \frac{1}{\sqrt{2}}\int_{-1}^{1} t^2\,\mathrm{d}t \\ &= \frac{1}{\sqrt{2}} \times \frac{2}{3} \\ &= \frac{\sqrt{2}}{3}\end{aligned}$$

De plus :

$$\begin{aligned}\langle x_3,e_2\rangle &= \sqrt{\frac{3}{2}}\int_{-1}^{1} t^3\,\mathrm{d}t \\ &= 0\end{aligned}$$

Dès lors,

$$\begin{aligned}u_3 &= X^2 - \frac{\sqrt{2}}{3} \times \frac{1}{\sqrt{2}} \\ &= X^2 - \frac{1}{3}\end{aligned}$$

Finalement :

$$\begin{aligned}\|u_3\|^2 &= \langle u_3,u_3\rangle \\ &= \int_{-1}^{1} \left(t^2 - \frac{1}{3}\right)^2 \mathrm{d}t \\ &= \int_{-1}^{1} t^4 - \frac{2t^2}{3} + \frac{1}{9}\,\mathrm{d}t \\ &= \frac{2}{5} - \frac{4}{9} + \frac{2}{9} \\ &= \frac{8}{45}\end{aligned}$$

et donc on pose finalement

$$e_3 = \sqrt{\frac{45}{8}}\left(X^2 - \frac{1}{3}\right)$$

Finalement, $(e_1, e_2, e_3) = \left(\dfrac{1}{\sqrt{2}}, X\sqrt{\dfrac{3}{2}}, \sqrt{\dfrac{45}{8}}\left(X^2 - \dfrac{1}{3}\right)\right)$ est une base orthonormale de $\mathbb{R}_2[X]$. On peut itérer le processus : en orthonormalisant les polynômes de la base canonique, on obtient (à une constante près) des polynômes appelés polynômes de Legendre. Il existe de nombreuses familles de polynômes orthogonaux, chacun pour un produit scalaire différent : polynômes de Legendre, de Tchebychev de première espèce, de Tchebychev de deuxième espèce, de Laguerre, de Hermite etc.

Remarque : Comme dit plus haut, on peut aussi (si on a un trou de mémoire le jour J) refaire un peu la démonstration et chercher de proche en proche u_k sous la forme

$$u_k = x_k - \sum_{i=1}^{k-1} \alpha_i u_i$$

puis chercher les α_k pour que u_k soit orthogonal à $u_1, \dots, u_{k-1}$, et diviser par la norme à la fin. Refaisons le deuxième exemple avec cette méthode.

On pose $u_1 = 1$. On cherche ensuite u_2 sous la forme $u_2 = X + \alpha_1 u_1$ et on cherche α_1 pour que u_2 soit orthogonal à u_1. On a :

$$\begin{aligned} u_1 \perp u_2 \quad & \iff \quad \langle u_2, u_1 \rangle = 0 \\ & \iff \quad \langle X, u_1 \rangle + \alpha_1 \langle u_1, u_1 \rangle = 0 \\ & \iff \quad \alpha_1 = -\frac{\langle X, u_1 \rangle}{\|u_1\|^2} \end{aligned}$$

On pose donc $u_2 = X - \dfrac{\langle X, u_1 \rangle}{\|u_1\|^2}$. Enfin, on pose $u_3 = X^2 + \alpha_1 u_1 + \alpha_2 u_2$ avec α_1, α_2 réels et on cherche α_1 et α_2 pour avoir u_3 orthogonal à u_1 et u_2. Commençons par u_1 :

$$\begin{aligned} u_1 \perp u_3 \quad & \iff \quad \langle u_3, u_1 \rangle = 0 \\ & \iff \quad \langle X^2 + \alpha_1 u_1 + \alpha_2 u_2, u_1 \rangle = 0 \\ & \iff \quad \langle X^2, u_1 \rangle + \alpha_1 u_1 u_1 + \alpha_2 u_2 u_1 = 0 \\ & \iff \quad \langle X^2, u_1 \rangle + \alpha_1 \|u_1\|^2 = 0 \\ & \iff \quad \alpha_1 = \frac{-\langle X^2, u_1 \rangle}{\|u_1\|^2} \end{aligned}$$

Car u_1 et u_2 sont orthogonaux.

On trouve de même que $u_2 \perp u_3$ si et seulement si $\alpha_2 = -\langle X^2, u_2 \rangle / \|u_2\|^2$. On pose donc

$$u_3 = X^2 - \frac{\langle X^2, u_1 \rangle}{\|u_1\|^2} - \frac{\langle X^2, u_2 \rangle}{\|u_2\|^2}$$

C'est la même formule que ci-dessus, mais attention, les u_i ne sont pas forcément unitaires !

On déroule ensuite tous les calculs : $u_1 = 1$ donc $\langle X, 1 \rangle = 0$ donc $u_2 = X$, $\langle X^2, u_1 \rangle = 2/3$, $\|u_1\|^2 = 2$ et $\langle X^2, u_2 \rangle = 0$ donc $u_3 = X^2 - 1/3$. Par conséquent, $(1, X, X^2 - 1/3)$ est orthogonale et il suffit de diviser par la norme de chacun des vecteurs pour aboutir au même résultat.

Donnons à présent un exemple du dernier cas de figure.

Exemple : On se place sur $\mathbb{R}[X]$ muni d'un produit scalaire $\langle , \rangle$. Montrer l'existence et l'unicité d'une famille de polynômes $(P_n)_{n \in \mathbb{N}}$ orthogonale telle que, pour tout n, $\deg(P_n) = n$ et P_n soit unitaire (i.e. de coefficient dominant égal à 1 dans cet exemple uniquement). Il suffit alors d'adapter la preuve ci-dessus (mais on peut faire une preuve plus courte, cf. IV.2).

- Si $n \in \mathbb{N}$, notons H_n : « il existe une unique famille $(P_0, \dots, P_n)$ orthogonale telle que, pour tout $k \in [\![0 \,;\, n]\!]$, $\deg(P_k) = k$ et P_k soit unitaire. »
- P_0 convient si et seulement si P_0 est constant et unitaire, donc si et seulement si $P_0 = 1$: il y a bien existence et unicité de P_0 donc H_0 est vraie.
- Soit $n \in \mathbb{N}$. Supposons H_n vraie et prouvons que H_{n+1} est vraie. Par hypothèse de récurrence, il existe $(P_0, \dots, P_n)$ orthogonale telle que, pour tout $k \in [\![0 \,;\, n]\!]$, P_k soit unitaire de degré k. On cherche donc P_{n+1} unitaire orthogonal à $P_0, \dots, P_n$, unitaire de degré $n+1$.

Les variations sur ce thème sont nombreuses. On demande par exemple, dans certains exercices, de prouver l'existence et l'unicité d'une famille (P_n) orthonormale telle que, pour tout n, $\langle P_n, X^n \rangle > 0$ (cf. exercice 43).

* **Analyse :** Si P_{k+1} convient. La famille $(P_0,\dots,P_n,X^{n+1})$ étant échelonnée en degré, c'est une famille libre de $\mathbb{R}_n[X]$ et c'est une famille libre à $n+1$ éléments en dimension $n+1$ donc c'est une base : $(\alpha_1,\dots,\alpha_{n+1})\in\mathbb{R}^{n+1}$ tel que :

$$P_{n+1}=\sum_{i=0}^n \alpha_i P_i+\alpha_{n+1}X^{n+1}$$

Or, P_{n+1} est orthogonal à $(P_0,\dots,P_n)$ donc, pour tout $j\in[\![\,0\,;\,n\,]\!]$,

$$\begin{aligned}\langle P_{n+1},P_j\rangle=0 &= \sum_{i=0}^n \alpha_i\langle P_i,P_j\rangle+\alpha_{n+1}\langle X^{n+1},P_j\rangle\\ &= \alpha_j\|P_j\|^2+\alpha_{n+1}\langle X^{n+1},e_j\rangle\end{aligned}$$

On a en effet utilisé le fait que les P_i sont deux à deux orthogonaux donc $\langle P_i,P_j\rangle=0$ si $i\neq j$ (mais n'est pas forcément égal à 1 si $i=j$: la famille n'est pas orthonormale !). On en déduit que $\alpha_j=-\alpha_{n+1}\langle X^{n+1},P_j\rangle/\|P_j\|^2$.
En d'autres termes, $P_{n+1}=\alpha_{n+1}u_{n+1}$ où

$$u_{n+1}=X^{n+1}-\sum_{i=0}^n\frac{\langle X^{n+1},P_i\rangle}{\|P_i\|^2}P_i$$

Or, P_{n+1} est unitaire donc α_{n+1} est l'inverse du coefficient dominant de u_{n+1} : α_{n+1} est uniquement déterminé, donc P_{n+1} également.

* **Synthèse :** Soit

$$u_{n+1}=X^{n+1}-\sum_{i=0}^n\frac{\langle X^{n+1},P_i\rangle}{\|P_i\|^2}P_i$$

et soit $P_{n+1}=\alpha_{n+1}u_{n+1}$ où α_{n+1} est l'inverse du coefficient dominant de u_{n+1}. En effet, u_{n+1} est non nul : si $u_{n+1}=0$, alors X^{n+1} est CL de $P_0,\dots,P_n$ donc appartient à $\mathbb{R}_n[X]$ ce qui est absurde. On en déduit que $u_{n+1}\neq 0$ si bien que $P_{n+1}=\alpha_{n+1}u_{n+1}$ est bien défini et unitaire. Il est de plus orthogonal à $P_0,\dots,P_n$: en effet, pour tout $j\in[\![\,0\,;\,n\,]\!]$,

$$\begin{aligned}\langle P_{n+1},P_j\rangle &= \alpha_{n+1}\langle X^{n+1},P_j\rangle-\alpha_{n+1}\sum_{i=0}^n\frac{\langle X^{n+1},P_i\rangle}{\|P_i\|^2}\langle P_i,P_j\rangle\\ &= \alpha_{n+1}\langle X^{n+1},P_j\rangle-\alpha_{n+1}\frac{\langle X^{n+1},P_j\rangle}{\|P_j\|^2}\langle P_j,P_j\rangle\\ &= 0\end{aligned}$$

On utilise encore une fois le fait que $\langle P_i,P_j\rangle=0$ si $i\neq j$.

c'est-à-dire que P_{n+1} est unitaire et orthogonal à $P_0,\dots,P_n$. Pour conclure, il suffit de voir que P_{n+1} est de degré $n+1$ car $P_0,\dots,P_n$ sont de degré inférieur ou égal à n : H_{n+1} est vraie.

* D'après le principe de récurrence, H_n est vraie pour tout $n\in\mathbb{N}$.

III.3.e En dimension finie

On suppose dans ce paragraphe que E est un espace euclidien (donc de dimension finie) de dimension $n\geqslant 1$.

Le dernier résultat du paragraphe précédent devient alors :

Théorème (Expression des coordonnées dans une base orthonormale). Soit $(e_1,\dots,e_n)$ une base orthonormale de E. Alors, pour tout $x\in E$, les coordonnées de x dans la base B sont les $(\langle x,e_1\rangle,\dots,\langle x,e_n\rangle)$. En particulier :

$$x=\sum_{i=1}^n \langle x,e_i\rangle e_i \qquad \text{et} \qquad \|x\|=\sqrt{\langle x,e_1\rangle^2+\cdots+\langle x,e_n\rangle^2}$$

Nous avons parlé de bases orthogonales dans le paragraphe précédent, il est donc légitime de se poser la question suivante : les résultats du chapitre 30 sont-ils encore valables avec des bases orthonormales ? Par exemple : le théorème de la base incomplète est-il encore valable avec des bases orthonormales ? Tout espace (euclidien) admet-il une base orthonormale ? Le procédé d'orthonormalisation de Gram-Schmidt permet de répondre par l'affirmative (en adaptant un peu).

Théorème (Théorème de la base orthonormale incomplète). Tout famille orthonormale de E peut être complétée en base orthonormale de E. En particulier, E admet une base orthonormale.

Démonstration. Soit $(e_1,\dots,e_p)$ une famille orthonormale de E. Alors $(e_1,\dots,e_p)$ est une famille libre donc, d'après le théorème de la base incomplète, on peut la compléter en base : il existe donc $(e_{p+1},\dots,e_n)$ tels que $(e_1,\dots,e_n)$ soit une base (pas encore orthonormale). On peut orthonormaliser cette base grâce à l'orthonormalisation de Gram-Schmidt, ce qui donne une base orthonormale mais, puisque les vecteurs $(e_1,\dots,e_p)$ forment déjà une famille orthonormale, ils restent inchangés par cet algorithme (cf. paragraphe III.3.c) : on a donc complété notre famille en base orthonormale de E.

Enfin, E étant de dimension $n\geqslant 1$, E contient un vecteur non nul x. Dès lors, $x/\|x\|$ est une famille orthonormale, qu'on peut donc compléter en base orthonormale (et donc E admet une base orthonormale).

Remarques :

- Attention, le théorème de la base orthonormale incomplète ne dit pas qu'on peut compléter une famille orthonormale en BON avec des vecteurs issus d'une famille génératrice (contrairement au théorème de la base incomplète) puisque celle-ci n'a aucune raison d'être une famille orthogonale ou orthonormale. On peut compléter en base, mais après on « orthonormalise » ces vecteurs, donc ce ne sont plus (après modification) les vecteurs qu'on a piochés pour compléter en base.
- Il n'existe pas de théorème de la base extraite orthonormale. En effet, une famille orthonormale étant déjà libre, il n'est pas possible d'avoir une famille orthonormale avec strictement plus d'éléments qu'une base, donc on ne peut pas extraire une base d'une famille orthonormale (à moins d'avoir déjà une base au départ, donc de prendre tous les vecteurs, et dans ce cas l'extraction a un intérêt limité...).

IV Orthogonal d'une partie

On revient au cas général, c'est-à-dire que E est un espace préhilbertien quelconque (pas forcément de dimension finie). On se donne X une partie non vide de E (pas forcément un espace vectoriel).

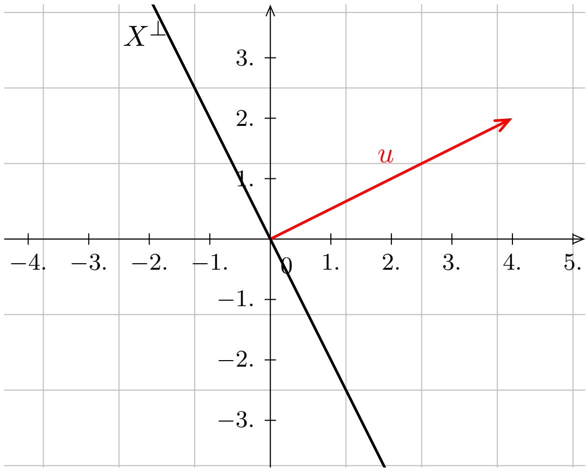
IV.1 Définition

Définition. On appelle orthogonal de X l'ensemble noté $X^\perp$ défini par : $X^\perp=\{y\in E\,|\,\forall x\in X, \langle x,y\rangle=0\}$. En d'autres termes, $X^\perp$ est l'ensemble des vecteurs orthogonaux à tous les éléments de X.

Un produit scalaire étant symétrique, il est équivalent de définir $X^\perp$ par la condition : $\forall x\in E, \langle y,x\rangle=0$.

Exemple : Si X est la partie de $\mathbb{R}^2$ uniquement formée du vecteur $u = (4,2)$, alors $X^\perp$ est la droite d'équation $y = -2x$. En effet, si $v = (x,y) \in \mathbb{R}^2$, alors :

$$\begin{aligned} v \in X^\perp & \iff \langle u,v \rangle = 0 \\ & \iff 4x + 2y = 0 \\ & \iff y = -2x \end{aligned}$$



On prend évidemment le produit scalaire canonique.

Proposition. $X^\perp$ est un sous-espace vectoriel de E.

Même lorsque X n'est pas un espace vectoriel ! Cela se voit d'ailleurs très bien sur l'exemple ci-dessus.

DÉMONSTRATION.
- Le vecteur nul est orthogonal à tout vecteur donc à tout élément de X : on en déduit que $0_E \in X^\perp$, $X^\perp$ est donc non vide.
- Soient $(y_1, y_2) \in \left(X^\perp\right)^2$ et $(\lambda_1, \lambda_2) \in \mathbb{R}^2$. Soit $x \in E$. Par linéarité à droite du produit scalaire, $\langle x,\lambda_1 y_1 + \lambda_2 y_2 \rangle = \lambda_1 \langle x,y_1 \rangle + \lambda_2 \langle x,y_2 \rangle$. Or, y_1 et y_2 appartiennent à $X^\perp$ donc $\langle x,y_1 \rangle = \langle y_2, = \rangle 0$ si bien que $\langle x,\lambda_1 y_1 + \lambda_2 y_2 \rangle = 0$: $\lambda_1 y_1 + \lambda_2 y_2 \in X^\perp$, $X^\perp$ est stable par CL, c'est donc un sev de E.

Lemme. Soit Y une partie non vide de E. Si $X \subset Y$ alors $Y^\perp \subset X^\perp$.

« Quand l'ensemble grandit, l'orthogonal rétrécit ». C'est intuitif : « être orthogonal à tout les éléments de l'ensemble » est une propriété très restrictive : si on ajoute des vecteurs auxquels il faut être orthogonal, on obtiendra moins de vecteurs qui vérifient cette propriété, donc un orthogonal plus petit.

DÉMONSTRATION. Soit $z \in Y^\perp$. Montrons que $z \in X^\perp$. Soit donc $x \in X$. Alors $x \in Y$ et $z \in Y^\perp$ donc $\langle z,x \rangle = 0$ ce qui permet de conclure.

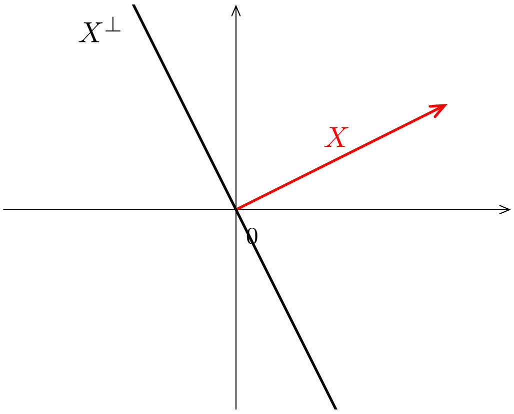
Proposition. $X^\perp = \mathrm{Vect}(X)^\perp$.

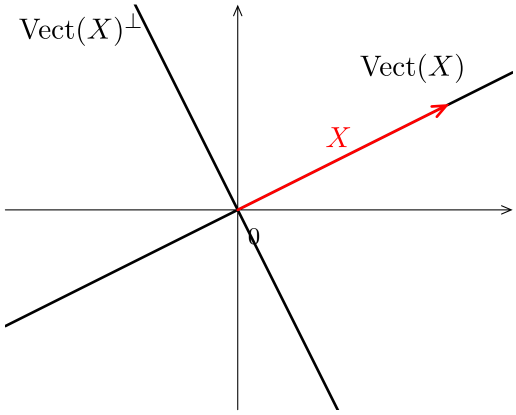
DÉMONSTRATION. L'inclusion $\mathrm{Vect}(X)^\perp \subset X^\perp$ découle du lemme précédent (puisque $X \subset \mathrm{Vect}(X)$). Réciproquement, soit $y \in X^\perp$. Soit $x \in \mathrm{Vect}(X)$: il existe donc $n \geqslant 1$, $(x_1, \dots, x_n) \in X^n$, $(\alpha_1, \dots \alpha_n) \in \mathbb{K}^n$ tels que $x = \alpha_1 x_1 + \dots + \alpha_n x_n$. Par linéarité à droite du produit scalaire,

$$\begin{aligned} \langle y,x \rangle & = \langle y,\alpha_1 x_1 + \dots + \alpha_n x_n \rangle \\ & = \alpha_1 \langle y,x_1 \rangle + \dots + \alpha_n \langle y,x_n \rangle \\ & = 0 \end{aligned}$$ □

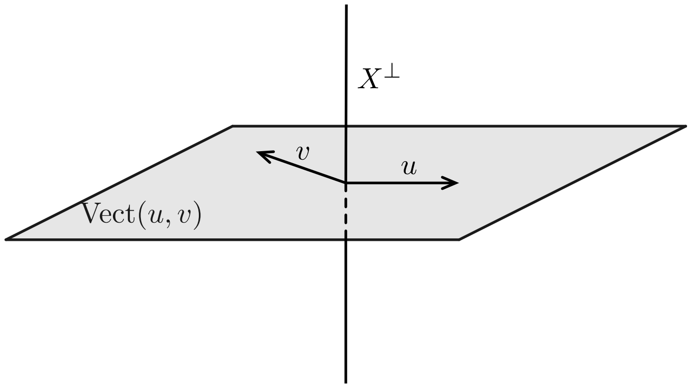
car y est orthogonal à tous les éléments de X donc est orthogonal aux x_i. On en déduit que $y \in \mathrm{Vect}(X)^\perp$. D'où l'inclusion réciproque, d'où l'égalité.

Remarque : Cela se voit très bien sur le dessin ci-dessous : cela revient au même d'être orthogonal à tous les éléments de X et d'être orthogonal à tous les éléments de $\mathrm{Vect}(X)$.





De même, un vecteur est orthogonal à deux vecteurs u et v si et seulement s'il est orthogonal au plan engendré par ces vecteurs : exemple ci-dessous avec X la partie formée par les vecteurs u et v.

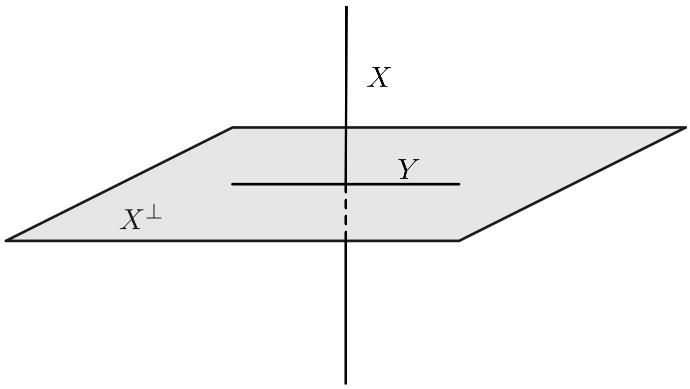


On peut définir une notion un peu plus faible :

Définition. Deux parties non vides de E sont orthogonales si : $\forall (x,y) \in X \times Y, \langle x,y \rangle = 0$, c'est-à-dire si tout élément de l'une est orthogonal à tout élément de l'autre.

On dit parfois également qu'elles sont en position orthogonale.

Remarque : Attention, cela ne signifie pas que $Y = X^\perp$ car il n'y a aucune raison que Y contienne **tous** les vecteurs orthogonaux à X. Cependant, on a alors $Y \subset X^\perp$.



IV.2 Orthogonal d'un sous-espace vectoriel

Rappelons qu'on peut définir l'orthogonal d'une partie quelconque (pas forcément un espace vectoriel). Cependant, puisque l'orthogonal d'une partie est égal à l'orthogonal de son espace engendré, on s'intéresse plus particulièrement dans ce paragraphe à l'orthogonal espace vectoriel. On se donne dans cette partie un sous-espace vectoriel F de E.

Proposition. Deux espaces vectoriels orthogonaux sont en somme directe. En particulier, F et $F^\perp$ sont en somme directe.

Démonstration. Soient E_1 et E_2 deux sous-espaces vectoriels de E, orthogonaux. Il suffit de prouver que $E_1 \cap E_2 = \{0\}$. Soit $x \in E_1 \cap E_2$. Alors $x \in E_1$ et puisque $x \in E_2$, alors x est orthogonal à tous les éléments de E_1 (car E_1 et E_2 sont orthogonaux) donc est orthogonal à lui-même, c'est-à-dire que $\langle x,x \rangle = 0$ donc $x = 0$ par définie positivé du produit scalaire.

Remarques :

- Lorsque deux sev de E sont orthogonaux, ils sont donc en somme directe, ce qu'on note $E_1 \overset{\perp}{\oplus} E_2$.
  Attention, contrairement à l'idée intuitive qu'on se fait en regardant les exemples précédents dans $\mathbb{R}^2$ ou dans $\mathbb{R}^3$, F et $F^\perp$ ne sont pas forcément supplémentaires (voir un contre-exemple ci-dessous), c'est-à-dire qu'on n'a pas forcément $F \overset{\perp}{\oplus} F^\perp = E$: il faut une condition supplémentaire.

Proposition. On suppose que F est de dimension finie. Alors F et $F^\perp$ sont supplémentaires : $F^\perp$ est alors appelé le supplémentaire orthogonal de F, c'est-à-dire que $F \overset{\perp}{\oplus} F^\perp = E$. Enfin, $\left(F^\perp\right)^\perp = F$.

Précisons que E n'a pas besoin d'être de dimension finie. En particulier, F admet un supplémentaire, ce qu'on ne savait pas jusqu'à présent : on ne pouvait affirmer l'existence d'un supplémentaire que quand tout l'espace était de dimension finie.

Démonstration. Soit $x \in E$. Montrons par analyse-synthèse qu'il existe $x_1 \in F$ et $x_2 \in F^\perp$ tel que $x = x_1 + x_2$.

Analyse : si x_1 et x_2 conviennent. F étant de dimension finie, il admet une base orthonormale notée $(f_1, \dots, f_n)$. Ainsi, il existe $(\alpha_1, \dots, \alpha_n) \in \mathbb{R}^n$ tel que $x_1 = \alpha_1 e_1 + \dots + \alpha_n e_n$ et donc

$$x = \alpha_1 e_1 + \dots + \alpha_n e_n + x_2$$

Soit $i \in [\![1 \,;\, n]\!]$. Par linéarité à gauche du produit scalaire :

$$\begin{aligned} \langle x,e_i \rangle &= \langle \alpha_1 e_1 + \dots + \alpha_n e_n + x_2, e_i \rangle \\ &= \alpha_1 \langle e_1,e_i \rangle + \dots + \alpha_n \langle e_n,e_i \rangle + \langle x_2,e_i \rangle \\ &= \alpha_i \end{aligned}$$

En effet, la base $(e_1, \dots, e_n)$ est orthonormale donc, si $j \neq i$, alors $\langle e_j,e_i \rangle = 0$, et $\langle e_i,e_i \rangle = 1$. Enfin, $x_2 \in F^\perp$ donc $\langle x_2,e_i \rangle = 0$. Par conséquent,

$$x_1 = \sum_{i=1}^n \langle x,e_i \rangle e_i \qquad \text{et} \qquad x_2 = x - x_1$$

Synthèse : Soient

$$x_1 = \sum_{i=1}^n \langle x,e_i \rangle e_i \qquad \text{et} \qquad x_2 = x - x_1$$

Alors, par construction, $x_1 \in F$ et $x = x_1 + x_2$. Il suffit donc de prouver que $x_2 \in F^\perp$. Or, $F = \text{Vect}(e_1, \dots, e_n)$ donc $F^\perp = \{e_1; \dots; e_n\}^\perp$: il suffit donc de prouver que x_2 est orthogonal à $e_1, \dots, e_n$. Soit $i \in [\![1 \,;\, n]\!]$. Par linéarité à gauche du produit scalaire :

$$\begin{aligned}\langle x_2, e_i\rangle &= \langle x - \sum_{j=1}^{n} \langle x, e_j\rangle e_j, e_i\rangle \\ &= \langle x, e_i\rangle - \sum_{j=1}^{n} \langle x, e_j\rangle \times \langle e_j, e_i\rangle \\ &= \langle x, e_i\rangle - \langle x, e_i\rangle \\ &= 0\end{aligned}$$

Car $\langle e_j, e_i\rangle = \delta_{i,j} = 0$.

Dès lors, $x_2 \in F^\perp$: on a prouvé l'existence et l'unicité de x_1 et x_2. Tout vecteur s'écrit de façon unique comme somme d'un élément de F et d'un élément de $F^\perp$ donc F et $F^\perp$ sont bien supplémentaires.

Montrons enfin que $F = \left(F^\perp\right)^\perp$ par double inclusion. Si $x \in F$ alors, par définition, x est orthogonal à tous les éléments de $F^\perp$ donc $x \in \left(F^\perp\right)^\perp$: $F \subset \left(F^\perp\right)^\perp$. Réciproquement, supposons que $x \in \left(F^\perp\right)^\perp$. En particulier, $x \in E = F \oplus F^\perp$ donc il existe $x_1 \in F$ et $x_2 \in F^\perp$ (uniques) tels que $x = x_1 + x_2$. Or, $x \in \left(F^\perp\right)^\perp$ donc $x \perp x_2$ donc :

$$\begin{aligned}\langle x, x_2\rangle = 0 &= \langle x_1, x_2\rangle + \langle x_2, x_2\rangle \\ &= 0 + \langle x_2, x_2\rangle\end{aligned}$$

□

On en déduit que $\langle x_2, x_2\rangle = 0$ donc, par définie positivité du produit scalaire, $x_2 = 0$ donc $x = x_1 \in F$: d'où l'inclusion réciproque, d'où l'égalité.

Remarques :

- On a supposé sans le dire que F était de dimension supérieure ou égale à 1 quand on en a donné une base, mais ce résultat est encore vrai si $F = \{0\}$: en effet, le vecteur nul est orthogonal à tout vecteur donc $\{0\}^\perp = E$. On en déduit que $\{0\}^\perp$ est un supplémentaire de $\{0\}$ (c'est même le seul), et puisque le vecteur nul est le seul orthogonal à tout vecteur, $E^\perp = \{0\}$ donc on a encore $\left(\{0\}^\perp\right)^\perp = \{0\}$.
- On dit « le » supplémentaire orthogonal car il est unique (quand il existe, ce qui est donc le cas en particulier quand F est de dimension finie), ce qui se voit très bien sur un dessin. Montrons en effet que si E_1 et E_2 sont supplémentaires et orthogonaux, alors $E_2 = {E_1}^\perp$. Tout d'abord, $E_2 \subset {E_1}^\perp$ car E_1 et E_2 sont orthogonaux. Réciproquement, soit $x \in {E_1}^\perp$. Alors $x \in E = E_1 \oplus E_2$ donc il existe x_1 et x_2 (uniques) tels que $x = x_1 + x_2$. Or, $x \in {E_1}^\perp$ donc $x \perp x_1$. De plus, $x_1 \perp x_2$ car E_1 et E_2 sont orthogonaux. Dès lors, par linéarité à gauche du produit scalaire :

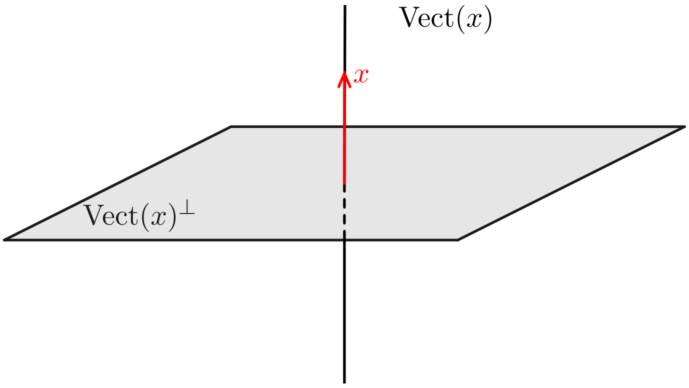
$$\begin{aligned}\langle x, x_1\rangle = 0 &= \langle x_1, x_1\rangle + \langle x_1, x_2\rangle \\ &= \langle x_1, x_1\rangle\end{aligned}$$

  Par définie positivité du produit scalaire, $x_1 = 0$ si bien que $x = x_2 \in E_2$: d'où l'inclusion réciproque, d'où l'égalité.

Corollaire. Si E est de dimension finie et si F est un sous-espace vectoriel de E, alors $F \overset{\perp}{\oplus} F^\perp = E$, $\dim\left(F^\perp\right) = \dim(E) - \dim(F)$, et $\left(F^\perp\right)^\perp = F$.

Exemples :

- Si $x \in E$ est un vecteur non nul, $\text{Vect}(x)$ est de dimension finie donc $H = \text{Vect}(x)^\perp$ est un supplémentaire de $\text{Vect}(x)$. C'est également un hyperplan car H admet un supplémentaire de dimension 1.



- On se place dans $\mathscr{M}_n(\mathbb{R})$ muni de son produit scalaire canonique. Montrons que $S_n(\mathbb{R})^\perp = A_n(\mathbb{R})$.

  Soient $S \in S_n(\mathbb{R})$ et $A \in A_n(\mathbb{R})$. D'une part, S étant symétrique :

$$\begin{aligned} \langle S,A\rangle &= \text{tr}\left(S^\top \times A\right) \\ &= \text{tr}\,(SA) \end{aligned}$$

  et d'autre part, le produit scalaire étant symétrique et A antisymétrique :

$$\begin{aligned} \langle S,A\rangle &= \langle A,S\rangle \\ &= \text{tr}\left(A^\top \times S\right) \\ &= \text{tr}\,(-AS) \\ &= -\text{tr}\,(AS) \\ &= -\text{tr}\,(SA) \\ &= -\langle S,A\rangle \end{aligned}$$

Linéarité de la trace

$\text{tr}\,(AB) = \text{tr}\,(BA)$.

  et donc $\langle S,A\rangle = 0$: A et S sont orthogonales. Par conséquent, $S_n(\mathbb{R})$ et $A_n(\mathbb{R})$ sont orthogonaux donc $A_n(\mathbb{R}) \subset S_n(\mathbb{R})^\perp$. De plus, on est en dimension finie donc $S_n(\mathbb{R})^\perp$ est un supplémentaire de $S_n(\mathbb{R})$. Or, on sait que $A_n(\mathbb{R})$ est un supplémentaire de $S_n(\mathbb{R})$: ces deux espaces ont donc même dimension, et puisque l'un des deux est inclus dans l'autre, alors ils sont égaux.

On peut donc visualiser ces deux espaces de la même façon que le dessin précédent... même si ce n'est pas fidèle à la réalité : $A_n(\mathbb{R})$ n'est pas de dimension 1 !

- On se place dans $E = \mathscr{C}([-1\,;1], \mathbb{R})$ muni du produit scalaire classique vu au paragraphe I.3.b. On note P le sous-espace vectoriel de E formé par les fonctions paires (définies et continues sur $[-1\,;1]$) et I le sev de E formé par les fonctions impaires. Montrons que $P \overset{\perp}{\oplus} I = E$. Tout d'abord, soient $p \in P$ et $i \in I$. Par produit, $p \times i$ est impaire donc :

Attention, les résultats précédents ne s'appliquent pas car P et I ne sont pas de dimension finie !

$$\begin{aligned} \langle p,i\rangle &= \int_{-1}^{1} p(t)i(t)\,\mathrm{d}t \\ &= 0 \end{aligned}$$

  car intégrale d'une fonction impaire sur un intervalle symétrique par rapport à 0 : P et I sont donc orthogonaux. Attention, on ne peut pas refaire le raisonnement précédent car on est en dimension infinie. Cependant, on sait déjà que P et I sont supplémentaires donc on a bien $P \overset{\perp}{\oplus} I = E$, c'est-à-dire que $I = P^\perp$.

Là aussi, on peut donc visualiser P et I comme sur le dessin ci-dessus... en faisant encore une fois attention car I n'est pas de dimension 1 : I est de dimension infinie !

- On se place toujours sur $E = \mathscr{C}([-1\,;1]\,,\mathbb{R})$ muni du produit scalaire intégral habituel. On note

$$H = \{f \in E \mid f(0) = 0\}$$

Alors H est un hyperplan de E car est le noyau d'une forme linéaire non nulle. Donnons $H^\perp$. Soit $f \in H^\perp$. Alors $g : t \mapsto t^2 f(t)$ est dans H donc $\langle f,g\rangle = 0$. Or,

$$\langle f,g\rangle = \int_{-1}^{1} t^2 f(t)^2 \,\mathrm{d}t$$

On a l'intégrale d'une fonction continue, positive et $-1 < 1$ donc la fonction intégrée est nulle : on en déduit que, pour tout $t \neq 0$, $f(t) = 0$ donc f est la fonction nulle sur $[-1\,;1]$ par continuité de f. On en déduit donc que $H^\perp = \{0\}$: $H^\perp$ n'est pas un supplémentaire de H, et on n'a pas $H = \left(H^\perp\right)^\perp$ puisque $\{0\}^\perp = E$. D'où l'importance que F soit de dimension finie !

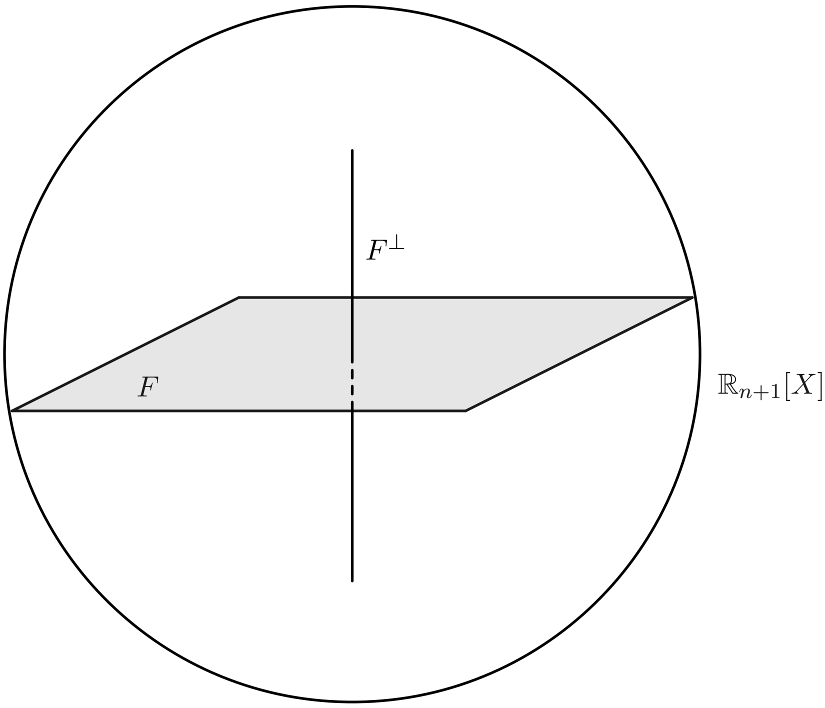
Même si cela marche parfois avec des sev de dimension infinie, voir l'exemple précédent.

Remarque : Grâce à ce résultat, on peut prouver de façon plus courte l'existence d'une famille orthogonale ou orthonormale (ainsi que, le cas échéant, l'unicité si on ajoute une condition supplémentaire) en partant d'une famille libre : l'astuce consiste à chercher x_{k+1} dans l'orthogonal de $\text{Vect}(e_1,\dots,e_k)$ dans $E = \text{Vect}(x_1,\dots,x_{k+1})$. Un exemple vaut mieux qu'un long discours : reprenons l'exemple du paragraphe III.3.d pour montrer l'existence et l'unicité d'une famille orthogonale de polynômes $(P_n)_{n\in\mathbb{N}}$ telle que, pour tout n, P_n soit de degré n et unitaire (c'est-à-dire, dans cet exemple uniquement, de coefficient dominant égal à 1).

Reprenons uniquement l'hérédité : on suppose donc que $n \in \mathbb{N}$ et qu'il existe $(P_0,\dots,P_n)$ qui conviennent, et on cherche à prouver l'existence et l'unicité de P_{n+1}.

- **Analyse :** Si P_{n+1}, alors $P_{n+1} \in \mathbb{R}_{n+1}[X]$ et P_{n+1} est orthogonal à $P_0,\dots,P_n$: si on pose $F = \text{Vect}(P_0,\dots,P_n)$, alors $P_{n+1} \in F^\perp$ dans $E = \mathbb{R}_{n+1}[X]$.

On ne prend pas l'orthogonal dans $\mathbb{R}[X]$ mais uniquement dans $\mathbb{R}_{n+1}[X]$ qui devient l'espace ambiant. L'avantage est que, dans cet espace, $F^\perp$ est de dimension 1.



$(P_0,\dots,P_n)$ étant une famille libre (car échelonnée en degré), F est de dimension $n+1$ et E est de dimension $n+2$ donc $F^\perp$ est de dimension 1 : il existe donc Q tel que $F^\perp = \text{Vect}(Q)$, c'est-à-dire qu'il existe α tel que $P_{n+1} = \alpha Q$, et α est égal à l'inverse du coefficient dominant de Q donc est unique donc P_{n+1} également.

- **Synthèse :** Soit $P_{n+1} = \alpha Q$ avec Q une base de $F^\perp$ et α l'inverse de son coefficient dominant. Tout d'abord, $Q \neq 0$ car $F^\perp$ est de dimension 1 donc α est bien défini et P_{n+1} est unitaire (et donc non nul). De plus, pour tout $j \in [\![0 \, ; n]\!]$, $P_j \in F$ donc P_j est orthogonal à P_{n+1} : la famille $(P_0, \dots, P_{n+1})$ est donc orthogonale, et enfin, $\deg(P_{n+1}) = n+1$ car, si $\deg(P_{n+1}) \leqslant n$, alors $P_n \in F$ donc est orthogonal à lui-même ce qui implique que $\langle P_{n+1}, P_{n+1} \rangle = 0$ ce qui implique que P_{n+1} est nul par définie positivité du produit scalaire, ce qui est absurde, ce qui clôt la récurrence.

Remarque : « Mais... puisque c'est si simple, pourquoi on s'est embêtés avec une récurrence lourde pour démontrer la validité du procédé de Gram-Schmidt ? » Parce que le résultat donnant la dimension de $F^\perp$ en dimension finie repose sur l'existence d'une BON, résultat... obtenu grâce au procédé de Gram-Schmidt ! On ne pouvait donc pas utiliser cette méthode pour démontrer ce résultat. Cependant, rétrospectivement, on comprend bien pourquoi il y a deux possibilités dans le procédé de Gram-Schmidt (et pourquoi il y a unicité avec une condition supplémentaire) : à chaque fois, on veut un vecteur orthogonal aux précédents, l'orthogonal (dans $E = \text{Vect}(x_1, \dots, x_{k+1})$) est de dimension 1, donc est une droite vectorielle, et il n'y a que deux vecteurs unitaires (qui sont opposés) sur une droite vectorielle (et, pour le dernier exemple, il n'y a qu'un polynôme de coefficient dominant égal à 1 sur une droite vectorielle puisque tous les polynômes de cette droite sont proportionnels).

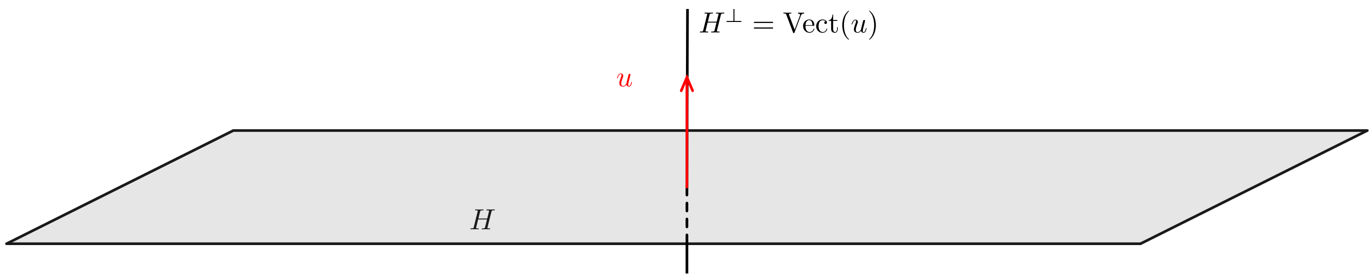
Faites un dessin !

Revenons aux vecteurs normaux : en dimension finie, on a le cas particulier suivant :

Proposition/Définition. On suppose que E est de dimension finie. Soit H un hyperplan de E. Alors $H^\perp$ est une droite vectorielle, et tout élément **non nul** de $H^\perp$ est appelé un vecteur normal à H, et est une base de $H^\perp$. En particulier, tous les vecteurs normaux à H sont proportionnels.

Un vecteur normal à un hyperplan est non nul par définition (même si le vecteur nul est orthogonal à tout le monde, il n'est pas un considéré comme un vecteur normal).

Démonstration. Il n'y a rien à prouver : on est en dimension finie donc $H^\perp$ est un supplémentaire de H et est de dimension 1 car H est un hyperplan, et tout vecteur non nul de $H^\perp$ est une famille libre à un élément donc une base de $H^\perp$.



Remarque : Si H est un hyperplan de E et si $(f_1, \dots, f_{n-1})$ est une base de H, alors $H^\perp = \text{Vect}(f_1, \dots, f_{n-1})^\perp = \{f_1; \dots; f_{n-1}\}^\perp$. Par conséquent, un vecteur est normal à H si et seulement s'il est non nul et orthogonal à tous les f_i. À part dans le cas particulier de $\mathbb{R}^n$ muni du produit scalaire canonique où on peut donner un vecteur normal directement (voir ci-dessous), pour trouver un vecteur normal, on trouve d'abord $H^\perp$ grâce aux relations $\langle x, f_i \rangle = 0$, puis on en prend un élément non nul.

Rappelons que $X^\perp = \text{Vect}(X)^\perp$.

Exemples :

- Dans $\mathbb{R}^2$ muni du produit scalaire canonique, si H est l'hyperplan d'équation $3x - y + 2z = 0$ (hyperplan car noyau de $f : (x, y, z) \mapsto 3x - y + 2z$ forme linéaire non nulle), alors $u = (3, -1, 2)$ est un vecteur normal à H. Plus généralement, dans $\mathbb{R}^n$ (muni du produit scalaire canonique), si H est un hyperplan d'équation $a_1x_1 + \cdots + a_nx_n = 0$

Ce n'est bien sûr pas aussi simple si on prend un autre produit scalaire.

(avec $a_1, \dots, a_n$ non tous nuls), alors $u = (a_1, \dots, a_n)$ est un vecteur normal à H. En effet, soit $x = (x_1, \dots, x_n) \in H$. Alors

$$\langle u,x \rangle = a_1x_1 + \cdots + a_nx_n = 0$$

donc $u \in H^\perp$: u étant non nul, c'est un vecteur normal à H.

- On se place dans $E = \mathbb{R}_2[X]$ muni du produit scalaire

$$\langle \, , \rangle : \begin{cases} \mathbb{R}_2[X]^2 & \to & \mathbb{R} \\ (P,Q) & \mapsto & \displaystyle\int_0^1 P(t)Q(t)\,\mathrm{d}t \end{cases}$$

Donnons un vecteur normal à $H = \mathbb{R}_1[X] = \mathrm{Vect}(1,X)$. Soit $P \in \mathbb{R}_2[X]$. Il existe $(a,b,c) \in \mathbb{R}^3$ tel que $P = aX^2 + bX + c$. Puisque $\mathrm{Vect}(1,X)^\perp = \{1;X\}^\perp$, alors :

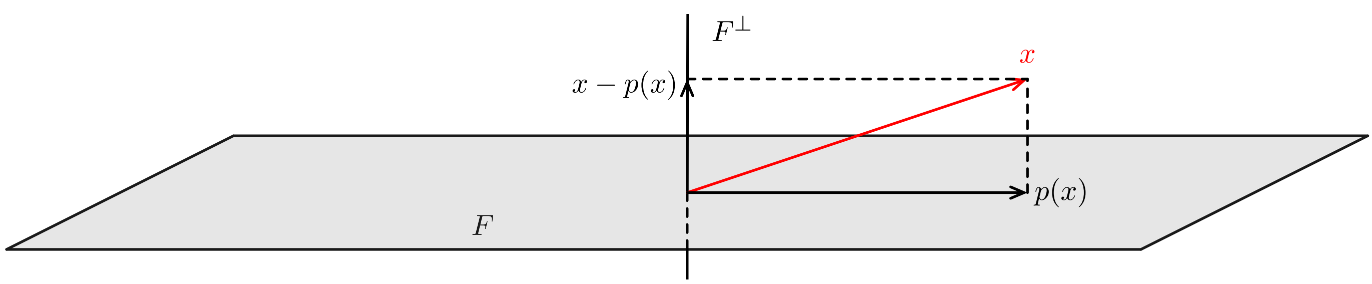
$$\begin{aligned}
P \in \mathbb{R}_1[X]^\perp \quad & \iff \quad \langle P,1 \rangle = \langle P,X \rangle = 0 \\
& \iff \quad \int_0^1 at^2 + bt + c\,\mathrm{d}t \; = \int_0^1 at^3 + bt^2 + ct\,\mathrm{d}t \; = 0 \\
& \iff \quad \begin{cases} a/3 & + & b/2 & + & c & = & 0 \\ a/4 & + & b/3 & + & c/2 & = & 0 \end{cases} \\
& \iff \quad \begin{cases} 2a & + & 3b & + & 6c & = & 0 \\ 3a & + & 4b & + & 6c & = & 0 \end{cases} \\
& \iff \quad \begin{cases} 2a & + & 3b & + & 6c & = & 0 \\ a & + & b & & & = & 0 \end{cases} \\
& \iff \quad \begin{cases} & & b & = & -6c \\ a & & & = & -b \end{cases} \\
& \iff \quad a = 6c \qquad \text{et} \qquad b = -6c
\end{aligned}$$

Par conséquent, $\mathbb{R}_1[X] = \{6cX^2 - 6cX + c \,|\, c \in \mathbb{R}\} = \mathrm{Vect}(6X^2 - 6X + 1)$ donc $6X^2 - 6X + 1$ est un vecteur normal à $\mathbb{R}_1[X]$.

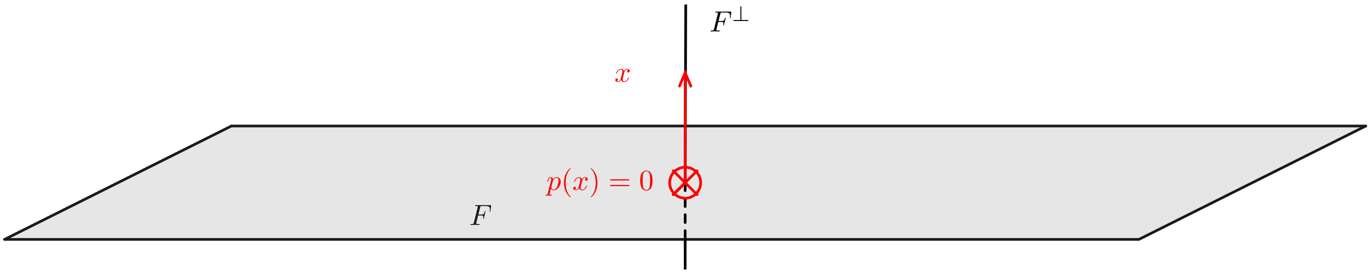
IV.3 Projection orthogonale sur un sous-espace de dimension finie

On se donne dans ce paragraphe un sous-espace vectoriel F de E **de dimension finie**. D'après le paragraphe précédent, les deux espaces F et $F^\perp$ sont supplémentaires : on peut donc définir la projection sur F parallèlement à $F^\perp$.

Définition. La projection sur F parallèlement à $F^\perp$ est appelée projection orthogonale sur F (parallèlement à $F^\perp$).



Remarque : Si p est la projection orthogonale sur F parallèlement à $F^\perp$, alors $\ker(p) = F^\perp$. Par conséquent, si $x \in F^\perp$, alors $p(x) = 0$: quand on projette un vecteur orthogonal à F sur F, on tombe sur 0, ce qui se voit bien sur le dessin ci-dessous.



Proposition. Si p est la projection orthogonale sur F alors, pour tout $x \in E$, $x - p(x) \in F^\perp$.

Ce résultat est utile pour calculer le projeté orthogonal explicitement, voir ci-dessous.

Démonstration. Soit $x \in E = F \oplus F^\perp$ donc il existe $x_1 \in F$ et $x_2 \in F^\perp$ uniques tels que $x = x_1 + x_2$. Or, par définition de p, $p(x) = x_1$ donc $x_2 = x - x_1 = x - p(x) \in F^\perp$.

Théorème (Expression du projeté orthogonal d'un vecteur x dans une BON.). Soit p la projection orthogonale sur F. Soit $B = (e_1, \dots, e_n)$ une base orthonormale de F. Alors :

$$\forall x \in E, p(x) = \sum_{i=1}^{n} \langle x,e_i \rangle e_i$$

Rappelons que F est de dimension finie donc admet une base orthonormale (finie).

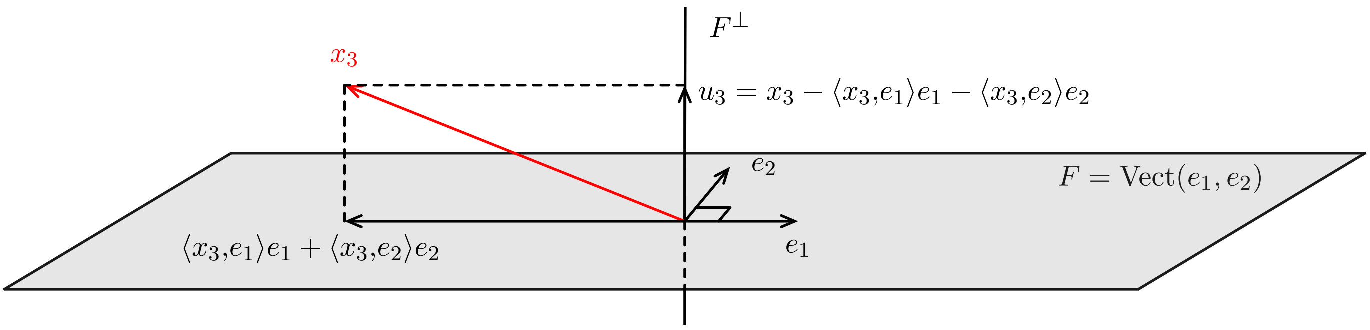
Démonstration. Soit $x \in E$. F et $F^\perp$ étant supplémentaires, il existe $x_1 \in F$ et $x_2 \in F^\perp$ tels que $x = x_1 + x_2$. De plus, par définition de la projection orthogonale sur F, $p(x) = x_1$. $p(x) \in F$ donc il existe $(\alpha_1, \dots, \alpha_n) \in \mathbb{R}^n$ tel que $p(x) = \alpha_1 e_1 + \cdots + \alpha_n e_n$ si bien que :

$$x = \sum_{i=1}^{n} \alpha_i e_i + \underbrace{x_2}_{\in F^\perp}$$

□

On prouve de même que précédemment (en calculant $\langle x,e_i \rangle$) que $\alpha_i = \langle x,e_i \rangle$ pour tout i ce qui est le résultat voulu.

Remarque : L'expression du vecteur u_k, dans l'algorithme d'orthonormalisation de Gram-Schmidt, est alors totalement naturelle : c'est le vecteur x_k auquel on a retiré l'expression du projeté orthogonal sur $\mathrm{Vect}(e_1, \dots, e_{k-1})$. Il est alors orthogonal à $e_1, \dots, e_{k-1}$, et il suffit de le diviser par sa norme pour obtenir e_k.



Remarque : Par conséquent, pour calculer le projeté orthogonal d'un vecteur x sur F, il y a en gros deux stratégies. Tout d'abord, on se donne une base (pas forcément orthonormale) $B = (f_1, \dots, f_n)$ de F, puis :

- on rend cette base orthonormale grâce au procédé d'orthonormalisation de Gram-Schmidt, et on applique le théorème précédent. On utilise en général cette méthode lorsqu'on dispose déjà d'une base orthonormale ou, à la limite, orthogonale, ou en petite dimension (1 ou 2) car, au-delà, le procédé d'orthonormalisation rend les calculs assez lourds.
- ou on exprime x comme CL d'éléments des f_i, on utilise le fait que $x - p(x) \in F^\perp$ donc $\langle x - p(x), f_i \rangle = 0$ et cela donne un système que l'on résout. On utilise cette méthode dans le cas général, quand on a une base qui n'est pas orthogonale car les calculs sont souvent plus simples.

Exemples :

- On se place sur $E = \mathscr{C}([\,0\,;2\pi\,], \mathbb{R})$ muni du produit scalaire habituel. Donnons le projeté orthogonal de $f = \mathrm{Id}_{[\,0\,;2\pi\,]}$ sur $F = \mathrm{Vect}(\sin, \cos)$ (qui est bien de dimension finie même si E ne l'est pas).

  La famille $(\sin, \cos)$ est déjà orthogonale (cf. paragraphe III.2) donc on utilise plutôt la première méthode : l'orthonormalisation de Gram-Schmidt consiste simplement à la normaliser, et ces fonctions sont de norme $\sqrt{\pi}$ (cf. paragraphe III.2) c'est-à-dire que $(\sin/\sqrt{\pi}, \cos/\sqrt{\pi})$ est une base orthonormale de F, si bien que :

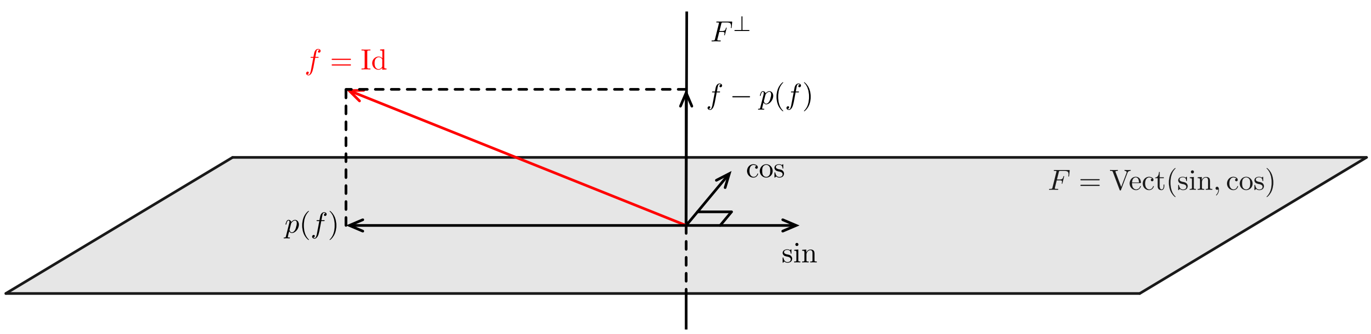
$$\begin{aligned} p(f) &= \langle f, \frac{\cos}{\sqrt{\pi}} \rangle . \frac{\cos}{\sqrt{\pi}} + \langle f, \frac{\sin}{\sqrt{\pi}} \rangle . \frac{\sin}{\sqrt{\pi}} \\ &= \langle f, \cos \rangle . \frac{\cos}{\pi} + \langle f, \sin \rangle . \frac{\sin}{\pi} \end{aligned}$$

Sortir le $\sqrt{\pi}$ du produit scalaire en utilisant la bilinéarité et ne pas le garder permet de simplifier les calculs.

  Or :

$$\begin{aligned} \langle f, \sin \rangle &= \int_0^{2\pi} t \sin(t)\,\mathrm{d}t \\ &= \left[-t\cos(t)\right]_0^{2\pi} + \int_0^{2\pi} \cos(t)\,\mathrm{d}t \\ &= -2\pi + \left[\sin(t)\right]_0^{2\pi} \\ &= -2\pi \end{aligned}$$

  et on trouve de même que l'autre produit scalaire est nul, si bien que $p(f) = -2\sin$.



- On se place sur $\mathscr{C}([\,0\,;1\,], \mathbb{R})$ muni du produit scalaire habituel. Donnons le projeté orthogonal de l'exponentielle sur l'ensemble des fonctions affines.

  Ici, il paraît plus simple d'appliquer la deuxième méthode. Notons $f_1 : x \mapsto 1$ et $f_2 : x \mapsto x$ si bien que $F = \mathrm{Vect}(f_1, f_2)$ est de dimension 2 (donc de dimension finie). Notons p la projection orthogonale sur F. (f_1, f_2) est une base de F donc il existe $(a_1, a_2) \in \mathbb{R}^2$ tel que $p(\exp) = a_1 f_1 + a_2 f_2$. $\exp - p(\exp) \in F^\perp$ donc $\langle \exp - p(\exp), f_1 \rangle = \langle \exp - p(\exp), f_2 \rangle = 0$. D'une part :

On considère dans cet exemple que les fonctions sont définies sur $[\,0\,;1\,]$, c'est-à-dire qu'on s'intéresse aux restrictions de l'exponentielle, etc. à $[\,0\,;1\,]$.

$$\begin{aligned}\langle \exp - p(\exp), f_1\rangle = 0 &= \langle \exp - af_1 - bf_2, f_1\rangle \\ &= \int_0^1 e^t - a - bt\,\mathrm{d}t \\ &= \left[e^t - at - \frac{bt^2}{2}\right]_0^1 \\ &= e - a - \frac{b}{2} - 1\end{aligned}$$

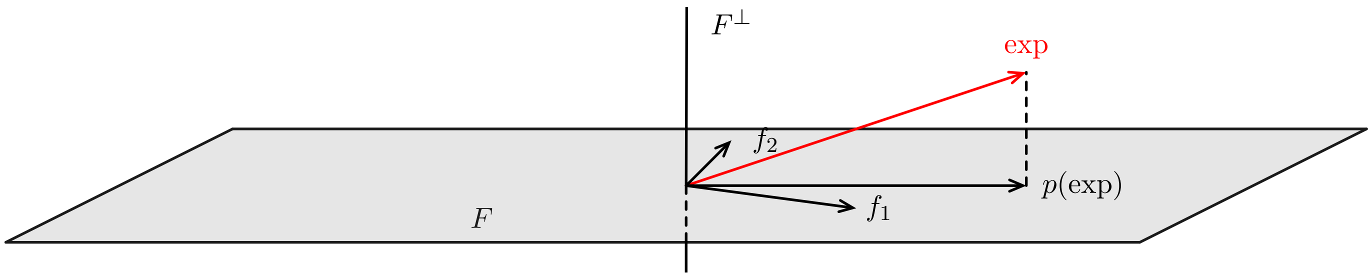
et d'autre part :

$$\begin{aligned}\langle \exp - p(\exp), f_2\rangle = 0 &= \langle \exp - af_1 - bf_2, f_2\rangle \\ &= \int_0^1 te^t\,\mathrm{d}t - \int_0^1 at + bt^2\,\mathrm{d}t \\ &= \left[te^t\right]_0^1 - \int_0^1 e^t\,\mathrm{d}t - \left[\frac{at^2}{2} + \frac{bt^3}{3}\right]_0^1 \\ &= e - (e-1) - \frac{a}{2} - \frac{b}{3} \\ &= 1 - \frac{a}{2} - \frac{b}{3}\end{aligned}$$

La résolution du système

$$\begin{cases} a + b/2 = e - 1 \\ a/2 + b/3 = 1 \end{cases}$$

donne $a = 4e - 10$ et $b = 18 - 6e$ si bien que $p(\exp) : x \mapsto (4e-10)x + 18 - 6e$.



- On se place sur $\mathscr{C}([0;2\pi],\mathbb{R})$ muni du produit scalaire habituel. Donnons le projeté orthogonal du cosinus sur l'ensemble des fonctions affines.

  Notons $f_1 : x \mapsto 1$ et $f_2 : x \mapsto x$ si bien que $F = \mathrm{Vect}(f_1, f_2)$ est de dimension 2 (donc de dimension finie). Notons p la projection orthogonale sur F. (f_1, f_2) est une base de F donc il existe $(a_1, a_2) \in \mathbb{R}^2$ tel que $p(\cos) = a_1 f_1 + a_2 f_2$. $\cos - p(\cos) \in F^\perp$ donc $\langle \cos - p(\cos), f_1\rangle = \langle \cos - p(\cos), f_2\rangle = 0$. D'une part :

On considère dans cet exemple que les fonctions sont définies sur $[0;2\pi]$, c'est-à-dire qu'on s'intéresse aux restrictions du cosinus, etc. à $[0;2\pi]$.

$$\begin{aligned}\langle \exp - p(\cos), f_1\rangle = 0 &= \langle \cos - af_1 - bf_2, f_1\rangle \\ &= \int_0^{2\pi} \cos(t) - a - bt\,\mathrm{d}t \\ &= \left[\sin(t) - at - \frac{bt^2}{2}\right]_0^{2\pi} \\ &= 0 - a \times 2\pi - b \times 2\pi^2\end{aligned}$$

et d'autre part :

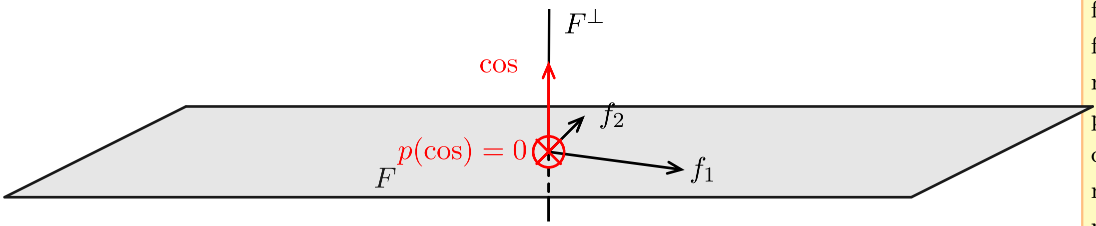
$$\begin{aligned}
\langle \cos - p(\cos), f_2 \rangle = 0 &= \langle \cos - af_1 - bf_2, f_2 \rangle \\
&= \int_0^{2\pi} t\cos(t)\,\mathrm{d}t - \int_0^{2\pi} at + bt^2\,\mathrm{d}t \\
&= [t\sin(t)]_0^{2\pi} - \int_0^{2\pi} \sin(t)\,\mathrm{d}t - \left[\frac{at^2}{2} + \frac{bt^3}{3}\right]_0^{2\pi} \\
&= 0 - a \times 2\pi^2 - b \times \frac{8\pi^3}{3}
\end{aligned}$$

On peut résoudre ce système, mais il suffit de voir que la matrice de ce système est

$$A = \begin{pmatrix} -2\pi & -2\pi^2 \\ -2\pi^2 & -8\pi^3/3 \end{pmatrix}$$

de déterminant $-4\pi^4/3 \neq 0$: on a un système de Cramer, donc qui a une unique solution, la solution nulle est solution évidente donc c'est la seule. En d'autres termes, $a = b = 0$ donc $p(\cos) = 0$: le projeté orthogonal du cos est le vecteur nul, ce qu'on aurait pu trouver directement (si on y avait pensé) en remarquant que $\langle \cos, f_1 \rangle = \langle \cos, f_2 \rangle$: cos est orthogonal à f_1 et f_2 donc à leur espace engendré (rappelons que $X^\perp = \mathrm{Vect}(X^\perp)$) donc appartient à $F^\perp$: son projeté orthogonal est alors le vecteur nul.

Le cosinus est donc orthogonal à l'espace des fonctions affines, mais il faut faire attention : ce n'est valable que pour le produit scalaire qu'on a choisi ! Par exemple, ce n'est plus vrai si on se place sur $\mathscr{C}\left(\left[0\,;\frac{\pi}{2}\right], \mathbb{R}\right)$ muni du produit scalaire habituel (exo).



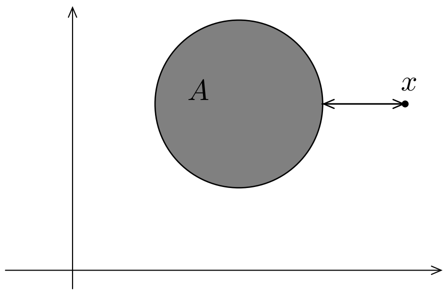
On aurait donc pu conclure plus rapidement dans ce cas : on pourra donc, dans ce genre d'exercice, commencer par calculer $\langle x, f_i \rangle$ pour tout i, puisque de toute façon il faudra les calculer pour trouver le projeté orthogonal.

IV.4 Distance d'un point à un sous-espace de dimension finie

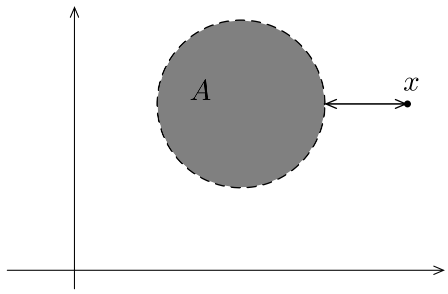
Définition. Soit A une partie non vide de E et soit $x \in E$. On appelle distance de x à A, notée $d(x, A)$, le réel : $d(x, A) = \inf_{a \in A} \|x - a\|$.

Remarques :

- Cette distance est bien définie : en effet, l'ensemble $\{\|x - a\| \mid a \in A\}$ est non vide car A est non vide, et il est minoré par 0 car une norme est positive.
- Cette définition est intuitive : on a envie de définir la distance d'un point à un ensemble comme le plus court chemin pour aller de ce point à cet ensemble. On prend donc toutes les distances entre x et un élément de a (rappelons que la distance est définie par la norme de la différence) et on prend ensuite la plus petite, c'est-à-dire la borne inférieure.
- Pourquoi pas le minimum ? Car, pour une partie quelconque, il n'existe pas forcément, c'est-à-dire que la borne inférieure n'est pas forcément atteinte.



Distance atteinte.



Distance non atteinte.

Activité : Montrons que, pour tous x et y, $|d(x, A) - d(y, A)| \leqslant \|x - y\|$.

Soit $a \in A$. D'après l'inégalité triangulaire, $\|x-a\| \leqslant \|x-y\|+\|y-a\|$. Or, $d(x, A) \leqslant \|x-a\|$ (la borne inférieure est un minorant) donc $d(x, A) \leqslant \|x - y\| + \|y - a\|$ donc

$$d(x, A) - \|x - y\| \leqslant \|y - a\|$$

En d'autres termes, $d(x, A) - \|x - y\|$ est un minorant de $\{\|y - a\| \,|\, a \in A\}$ donc est inférieur à sa borne inférieure (la borne inférieure est le plus grand des minorants), si bien que $d(x, A) - \|x - y\| \leqslant d(y, A)$. En d'autres termes : $d(x, A) - d(y, A) \leqslant \|x - y\|$. Par symétrie des rôles, $d(y, A) - d(x, A) \leqslant \|x - y\|$ ce qui permet de conclure.

Si on étend la notion de fonction lipschitzienne au cas des fonctions définies sur E, on en déduit que $x \mapsto d(x, A)$ est 1-lipschitzienne donc continue : cf. cours de l'an prochain !

Rappelons que $|\alpha| = \max(\alpha, -\alpha)$ donc, pour montrer une inégalité du type $|\alpha| \leqslant B$, il suffit de prouver que $\alpha \leqslant B$ et $-\alpha \leqslant B$.

On a dit que la distance n'était pas forcément atteinte, d'où la nécessité de la borne inférieure, mais dans le cas où A est un sous-espace vectoriel de dimension finie, « tout va bien ».

Proposition. Soit F un sous-espace vectoriel de E de dimension finie. Soit $x \in E$. Alors la distance de x à F est atteinte, et elle est atteinte uniquement en $p(x)$, où p est la projection orthogonale sur F. Plus précisément, $\|x - p(x)\| = \min_{a \in F} \|x - a\|$, et $p(x)$ est l'unique élément de F qui réalise la distance de x à F.

Le fait que F est de dimension est nécessaire car, sinon, F et $F^\perp$ ne sont pas forcément supplémentaires, et alors la projection orthogonale sur F n'est pas forcément définie.

Démonstration. Il suffit donc de prouver que, pour tout $a \in F$, $\|x - p(x)\| \leqslant \|x - a\|$. Soit $a \in F$.

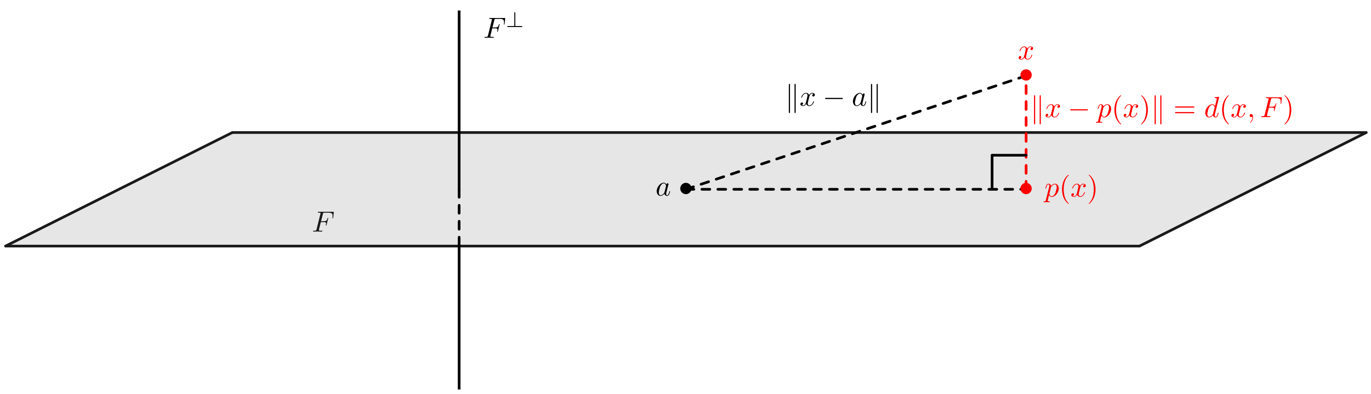
$$x - a = (x - p(x)) + (p(x) - a)$$

Or, $x - p(x) \in F^\perp$ et $p(x) - a \in F$ car F est un sev de E donc est stable par somme. Par conséquent, $x - p(x)$ et $(p(x) - a)$ sont orthogonaux : d'après le théorème de Pythagore,

$$\|x - a\|^2 = \|x - p(x)\|^2 + \|p(x) - a\|^2 \geqslant \|x - p(x)\|^2 \qquad \square$$

avec égalité si et seulement si $\|p(x) - a\|^2 = 0$, si et seulement si $p(x) = a$. En particulier, $\|x - a\| \geqslant \|x - p(x)\|$ avec égalité si et seulement si $a = p(x)$, ce qui permet de conclure.

Remarque : Cela se voit bien sur un dessin :



Remarque : On peut interpréter l'écart-type comme la distance entre une variable aléatoire X et l'espace vectoriel F des fonctions constantes (qui est bien de dimension finie car est de dimension 1). Plus précisément, si on se place sur $E = \mathbb{R}^\Omega$, alors on a vu que l'application $(X, Y) \mapsto E(XY)$ n'est pas un produit scalaire mais est tout de même positive, bilinéaire et symétrique. Par conséquent, l'égalité prouvée dans l'exercice 1 du chapitre 27, à savoir :

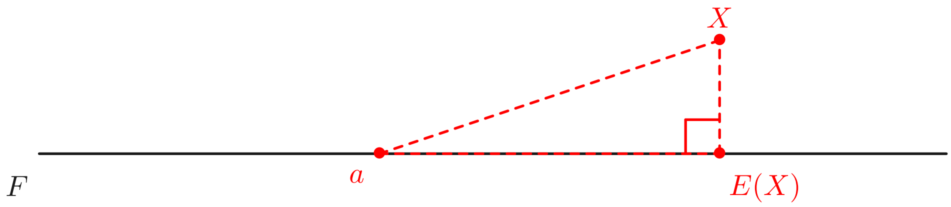
$$\forall a \in \mathbb{R}, E\left[(X-a)^2\right] = E\left[(X-E(X))^2\right] + (E(X)-a)^2$$

peut être interprétée comme :

Attention, ce n'est pas rigoureux : comme dit ci-dessus, la fonction $(X, Y) \mapsto E(XY)$ n'est pas un produit scalaire, donc on n'a pas vraiment une norme, on travaille avec les mains.

$$\|X-a\|^2 = \|X-E(X)\|^2 + \|E(X)-a\|^2$$

On est dans la même configuration que ci-dessus, avec X une variable aléatoire donc un élément de l'espace vectoriel des variables aléatoires $\mathbb{R}^\Omega$, $E(X)$ le projeté orthogonal de X sur l'espace vectoriel des fonctions constantes, et a un réel quelconque.



En se souvenant que $E\left[(X-E(X))^2\right] = V(X) = \sigma(X)^2$, on peut donc interpréter $E(X)$ comme le projeté orthogonal de X sur F et $\sigma(X)$ la distance entre X et F donc la distance entre X et l'ensemble des fonctions constantes, et on trouve alors que $E(X)$ est la constante qui minimise la distance. C'est intuitif ! La meilleure approximation d'une variable aléatoire par une constante est sa moyenne !

Remarque : Il est fréquent de voir des exercices demandant des distances de points à des espaces vectoriels de façon cachée, sans parler de produit scalaire. Il faut alors y penser soi-même. Typiquement, il faut se poser la question lorsqu'on voit une borne inférieure (ou un minimum), une différence, et quelque-chose qui ressemble à une norme (il y a un carré) ou qui ressemble à un produit scalaire connu. Des exemples valent mieux qu'on long discours.

Exemples :

- Calculer

$$m = \min_{(a,b)\in\mathbb{R}^2} \int_0^{2\pi} (\cos(t) - (at+b))^2 \, \mathrm{d}t$$

On se place sur $\mathscr{C}([0\,;2\pi], \mathbb{R})$ muni du produit scalaire habituel et de la norme associée. On cherche donc $\min_{f\in F} \|\cos - f\|^2$ avec F l'ensemble des fonctions affines. On cherche donc $d(\cos, F)^2$. Or, $d(\cos, F) = \|\cos - p(\cos)\|$, où p est la projection orthogonale sur F. Or, on a prouvé au paragraphe IV.3 que $p(\cos) = 0$ donc

$$\begin{aligned} m &= \|\cos\|^2 \\ &= \langle \cos, \cos \rangle \\ &= \int_0^{2\pi} \cos^2(t) \, \mathrm{d}t \\ &= \int_0^{2\pi} \frac{\cos(2t)+1}{2} \, \mathrm{d}t \\ &= \pi \end{aligned}$$

En particulier, le minimum est bien atteint, et il l'est uniquement pour $a = b = 0$.

- Calculer

$$m = \min_{(a,b)\in\mathbb{R}^2} \int_0^{2\pi} (t - (a\cos(t) + b\sin(t)))^2 \,\mathrm{d}t$$

On se place sur $\mathscr{C}([\,0\,;2\pi\,],\mathbb{R})$ muni du produit scalaire habituel et de la norme associée. On cherche donc $\min_{f\in F} \|\operatorname{Id} - f\|^2$ avec $F = \operatorname{Vect}(\cos, \sin)$. On cherche donc $d(\operatorname{Id}, F)^2$. Or, $d(\operatorname{Id}, F) = \|\operatorname{Id} - p(\operatorname{Id})\|$, où p est la projection orthogonale sur F. Or, on a prouvé au paragraphe IV.3 que $p(\operatorname{Id}) = -2\sin$ donc

$$\begin{aligned} m &= \|\operatorname{Id} - 2\sin\|^2 \\ &= \langle \operatorname{Id} - 2\sin, \operatorname{Id} - 2\sin \rangle \\ &= \int_0^{2\pi} (t - 2\sin(t))^2 \,\mathrm{d}t \\ &= \int_0^{2\pi} (t^2 - 4\sin(t) + 4\sin^2(t)) \,\mathrm{d}t \\ &= \frac{8\pi^3}{3} + 4\pi \end{aligned}$$

En particulier, le minimum est bien atteint, et il l'est uniquement pour $a = 0$ et $b = -2$.

- Calculer

$$m = \min_{(a,b)\in\mathbb{R}^2} \int_0^1 (e^t - (at + b))^2 \,\mathrm{d}t$$

On se place sur $\mathscr{C}([\,0\,;1\,],\mathbb{R})$ muni du produit scalaire habituel et de la norme associée. On cherche donc $\min_{f\in F} \|\exp - f\|^2$ avec F l'ensemble des fonctions affines. On cherche donc $d(\exp, F)^2$. Or, $d(\exp, F) = \|\exp - p(\exp)\|$, où p est la projection orthogonale sur F. Or, on a prouvé au paragraphe IV.3 que $p(\exp) : t \mapsto (4e - 10)t + (18 - 6e)$ donc

$$\begin{aligned} m &= \|\exp - p(\exp)\|^2 \\ &= \langle \exp - p(\exp), \exp - p(\exp) \rangle \\ &= \int_0^1 \left(e^t - (4e - 10)t - (18 - 6e)\right)^2 \,\mathrm{d}t \end{aligned}$$

Tous calculs faits (c'est très moche), on trouve :

$$\min_{(a,b)\in\mathbb{R}^2} \int_0^1 (e^t - (at + b))^2 \,\mathrm{d}t = \frac{1397 - 1000e + 179e^2}{6}$$

En particulier, le minimum est bien atteint, et il l'est uniquement pour $a = 4e - 10$ et $b = 18 - 6e$.

Morale de l'histoire : Ce genre de minimum (ou de borne inférieure) est égal à $d(x, F)^2$ pour un certain x et un certain F. On commence donc par calculer le projeté orthogonal de x sur F, et on calcule ensuite $\|x - p(x)\|^2$.

Corollaire. On suppose que E est de dimension finie n. Soit $u \in E$ non nul. Alors $H = \operatorname{Vect}(u)^\perp$ est un hyperplan de E et, pour tout $x \in E$:

- le projeté orthogonal de x sur H (parallèlement à $\operatorname{Vect}(u)$) est $p(x) = x - \dfrac{\langle x, u\rangle}{\|u\|^2}.u$.
- la distance de x à H est $d(x, H) = \dfrac{|\langle x, u\rangle|}{\|u\|}$.

Il y a un carré !

En particulier, si u est unitaire, alors $p(x) = x - \langle x,u\rangle.u$ et $d(x,H) = |\langle x,u\rangle|$.

Démonstration. Notons $\tilde{p}$ la projection orthogonale sur $\text{Vect}(u) = H^\perp$. $u/\|u\|$ étant une base orthonormale de $\text{Vect}(u)$, d'après le théorème du paragraphe IV.3,

$$\begin{aligned}\tilde{p}(x) &= \langle x, \frac{u}{\|u\|}\rangle.\frac{u}{\|u\|} \\ &= \frac{\langle x,u\rangle}{\|u\|^2}.u\end{aligned}$$

Or, H et $\text{Vect}(u)$ étant supplémentaires, x s'écrit d'une façon unique sous la forme $x = x_1 + x_2$ avec $x_1 \in H$ et $x_2 \in \text{Vect}(u)$ et, par définition de ces projections, $x_1 = p(x)$ et $p_2 = \tilde{p}(x)$ si bien que $x = p(x) + \tilde{p}(x)$, et on a finalement $p(x) = x - \tilde{p}(x)$ ce qui est le résultat voulu.

Enfin, d'après ce qui précède :

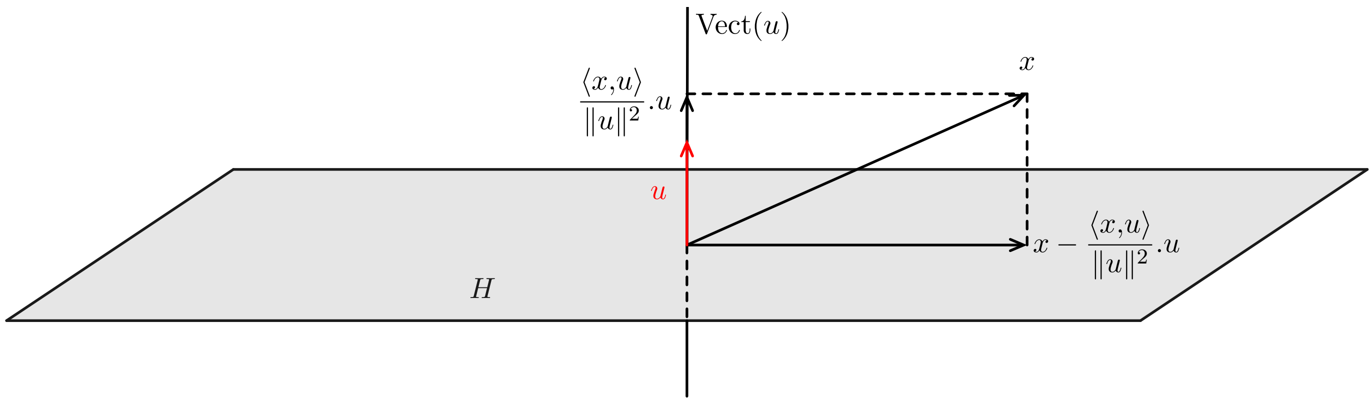
$$\begin{aligned}d(x,H) &= \|x - p(x)\| \\ &= \|\tilde{p}(x)\| \\ &= \left\|\frac{\langle x,u\rangle}{\|u\|^2}.u\right\| \\ &= \frac{|\langle x,u\rangle|}{\|u\|^2} \times \|u\|\end{aligned}$$

□

Absolue homogénéité de la norme.

ce qui permet de conclure.

Remarque : Ce résultat pouvant paraître compliqué est en fait extrêmement simple lorsqu'on fait un dessin :



Alors : **le projeté sur H est juste égal à x auquel on a retiré sa composante en** u, et la distance à H est simplement la longueur de cette composante (cette composante qui sépare x de H) donc est égale à sa norme.