

Chap V - L'ESPACE DE HILBERT $L^2(\mu)$

“Les gens peuvent choisir n'importe quelle valeur pour leur angle θ , du moment qu'ils sont droits” - H. Ford (à peu près)



1 Introduction

Il ne vous aura pas échappé que

- les espaces vectoriels, c'est fantastique.
- les espaces euclidiens (ou préhilbertiens), c'est mieux.

et le secret des espaces vectoriels euclidiens, c'est l'orthogonalité.

Un produit scalaire permet d'enrichir considérablement la “géométrie” qu'on peut faire sur un espace vectoriel E : en plus de calculer des distances, on va maintenant pouvoir parler d'angles (droits, surtout).

Pour cela, on définit l'angle θ entre deux vecteurs non nuls u, v d'un espace préhilbertien $(E, \langle \cdot, \cdot \rangle)$ par

$$\cos(\theta) = \frac{\langle u, v \rangle}{\|u\| \|v\|} \in [-1, 1]^1$$

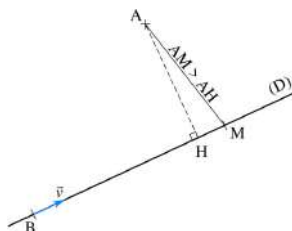
ce qui nous donne une généralisation du *théorème d'Al Kashi*:

$$\|u + v\|^2 = \|u\|^2 - 2\|u\| \|v\| \cos(\theta) + \|v\|^2$$

Exercice: Montrer cette formule. Et, pendant qu'on y est, montrer *l'identité du parallélogramme*:

$$\forall u, v \in E, \|u + v\|^2 + \|u - v\|^2 = 2\|u\|^2 + 2\|v\|^2. \quad (1)$$

En dimension finie, l'orthogonalité nous permet notamment de faire des projections orthogonales: étant donné un sous-espace vectoriel $F \subset E$, on va pouvoir trouver, pour chaque $u \in E$, quel est le point le plus proche dans F .



Le produit scalaire nous permet aussi de représenter toute forme linéaire $\phi \in E^*$ par un vecteur de E . Dans \mathbb{R}^n avec le produit scalaire habituelle, cette opération consiste à prendre une matrice ligne (qui représente ϕ) et à la mettre en colonne (et pouf, un vecteur !). Vous n'êtes donc peut-être pas impressionnés, mais c'est plus impressionnant en dimension infinie, où ce résultat permet de démontrer d'importants théorèmes de probabilités (Radon-Nikodym notamment).

Enfin, dans un espace vectoriel de dimension finie, on a des bases: des familles finies de vecteurs (e_1, \dots, e_n) telles que tout vecteur v de l'espace vectoriel E est combinaison linéaire des vecteurs e_i :

$$v = \sum_{i=1}^n \lambda_i e_i$$

¹D'après Cauchy-Schwartz !

et, si E est un espace préhilbertien et (e_1, \dots, e_n) est une *base orthonormée*, on peut trouver facilement les λ_i : pour chaque i , $\lambda_i = \langle v, e_i \rangle$.

Malheureusement, tous les espaces vectoriels ne sont pas de dimension finie, et donc ne disposent pas de bases orthonormées. Et la résolution de beaucoup de problèmes mathématiques requiert de travailler dans des espaces de fonctions, qui ne sont furieusement pas de dimension finie.

Pour compenser, en dimension infinie, on ajoute une hypothèse plus puissante qu'elle n'en a l'air: la complétude.

Définition 1. *Un **espace de Hilbert** est un espace vectoriel muni d'un produit scalaire $\langle \cdot, \cdot \rangle$, qui est complet pour la norme associée $\|x\| = \sqrt{\langle x, x \rangle}$.*

Un espace de Hilbert de dimension infinie, comme par exemple ℓ^2 ou $L^2(\mathbb{R})$ qu'on a croisés au chapitre IV, ne peut pas être muni d'une base orthonormée finie.

Toutefois, on va voir que les espaces de Hilbert *séparables* ont des presque-bases orthonormées, qui donnent accès à de très beaux résultats.

Définition 2. *Un e.v.n. $(E, \|\cdot\|)$ est **séparable** s'il a un sous-ensemble dénombrable dense: autrement dit s'il existe $(\alpha_n)_n \in E^{\mathbb{N}}$ tel que*

$$\overline{\{\alpha_n, n \in \mathbb{N}\}} = E$$

ou encore $\forall v \in E, \inf_{n \in \mathbb{N}} \|\alpha_n - v\| = 0$ i.e.

$$\forall \varepsilon > 0, \exists n_0 \in \mathbb{N} \text{ t.q. } \|\alpha_{n_0} - v\| < \varepsilon$$

Exercice: Montrer que s'il existe une famille $\{e_n, n \in \mathbb{N}\}$ dans E telle que

$$\overline{\text{Vect}(e_n, n \in \mathbb{N})} = E$$

alors E est séparable.

2 Orthogonalité

On travaille dans l'espace vectoriel $L^2(\mu)$ muni du produit scalaire

$$\langle f, g \rangle = \int Xfg d\mu$$

ce qui, on l'a vu, en fait un espace de Hilbert.

Définition 3. • On dit que $f, g \in L^2(\mu)$ sont *orthogonaux* si $\langle f, g \rangle = 0$.

• Soit $A \subset L^2(\mu)$, on définit *l'orthogonal de A* par

$$A^\perp = \{v \in L^2(\mu), \forall u \in A, \langle u, v \rangle = 0\}.$$

Exemples:

- quel que soit $v \in E$, 0_E est orthogonal à v ; et c'est le seul vecteur qui soit orthogonal à tous les autres.
- $t \mapsto \cos(3t)$ et $t \mapsto \sin(25t)$ sont orthogonaux dans $L^2([-\pi, \pi])$, et ils sont tous deux orthogonaux à 1.
- Dans ℓ^2 , posons

$$F = \{(a_n)_n \in \ell^2, \forall k \in \mathbb{N}, a_{2k} = 0\}.$$

$$\text{Alors } F^\perp = \{(a_n)_n \in \ell^2, \forall k \in \mathbb{N}, a_{2k+1} = 0\}.$$

- Dans $L^2(\mathbb{R})$, posons

$$F = \{f \in L^2(\mathbb{R}), f(x) = 0 \text{ } \lambda_1\text{-p.p sur } \mathbb{R}_+^*\}$$

alors F est un sev fermé, et

$$F^\perp = \{f \in L^2(\mathbb{R}), f(x) = 0 \text{ } \lambda_1\text{-p.p sur } \mathbb{R}_-\}$$

Remarques:

- Deux vecteurs orthogonaux u et v vérifient le *théorème de Pythagore*:

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

- On n'a pas besoin que A soit un s.e.v. de $L^2(\mu)$ pour calculer A^\perp .

En revanche, même si A n'en est pas un, A^\perp est un s.e.v fermé.

Proposition 1. *L'orthogonal d'un sous-ensemble $A \subset L^2(\mu)$ a les propriétés suivantes:*

$$\text{Si } A \subset B, B^\perp \subset A^\perp$$

et

$$A \cap A^\perp = \{0\}; \quad (\overline{A})^\perp = A^\perp; \quad A \subset (A^\perp)^\perp$$

Projection orthogonale

L'orthogonalité a, dans les espaces vectoriels de dimension finie, des applications très agréables, notamment la *projection orthogonale* sur un sous-espace vectoriel fermé (ou plus généralement sur un convexe fermé).

Pour disposer de cette projection en dimension infinie, on va voir que l'hypothèse de *complétude* joue un rôle clé (et peut-être inattendu, car la complétude n'a, a priori, que peu de lien avec la dimension finie).

La question est la suivante: si $F \subset L^2(\mu)$ est un s.e.v. *fermé*, et $f \in L^2(\mu)$ quelconque, quel est le point de F le plus proche de f ?²

Remarque: Contrairement à l'impression qu'on pourrait avoir en se basant sur la dimension finie, tous les s.e.v. d'un e.v.n. ne sont pas nécessairement fermés.

Contre-exemple: Considérons le s.e.v. $\mathcal{C}^0([-1, 1])$ de $L^2([-1, 1])$. Alors la suite de fonctions

$$f_n : x \in [-1, 1] \mapsto \begin{cases} -1 & \text{si } x \leq -1/n \\ nx & \text{si } -1/n < x < 1/n \\ 1 & \text{si } x \geq 1/n \end{cases}$$

est une suite convergente de $\mathcal{C}^0([-1, 1])$, mais sa limite n'est pas continue.

Donc, plus précisément, étant donné un sev fermé F et $f \in L^2(\mu)$, on cherche $g_0 \in F$ tel que, pour tout $g \in F$,

$$\|f - g_0\| \leq \|f - g\|$$

Remarque: Dans un e.v.n., même complet, un tel g_0 n'existe pas forcément: c'est, on va le voir, une spécificité des espaces de Hilbert.

Contre-exemple: Considérons l'espace vectoriel $E = \mathcal{C}^0([0, 1])$ muni de la norme $\|f\|_\infty = \sup_{t \in [0, 1]} |f(t)|$.

$\rightsquigarrow (E, \|\cdot\|_\infty)$ est un espace de Banach, et le sous-ensemble

$$F = \left\{ g \in E, \int_0^1 g(t) dt = 0 \text{ et } g(1) = 0 \right\}$$

²C'est une question particulièrement intéressante dans le cadre des théorèmes d'approximation: c'est le cas où F est un sous-ensemble de fonctions gentilles, par exemple les polynômes trigonométriques dans le cas de la théorie de Fourier. C'est aussi ce qui sous-tend la technique d'erreur quadratique moyenne minimale (MMSE).

est un sous-espace vectoriel fermé.

Prenons la fonction $f : x \in [0, 1] \mapsto 1 - x$. Alors, si on pose, pour tout $n \in \mathbb{N}$,

$$g_n : x \in [0, 1] \mapsto \frac{1}{2} - x + \frac{x^n}{2} + \frac{x-1}{n+1}$$

alors $\|f - g_n\|_\infty \rightarrow \frac{1}{2}$. On en déduit que $d(f, F) \leq \frac{1}{2}$.

D'un autre côté, pour tout $g \in F$, $(f - g)(1) = 0$ et $\int_0^1 f(t) - g(t) dt = \frac{1}{2}$, donc on doit avoir $\|f - g\|_\infty > \frac{1}{2}$.

Donc, aucune fonction $g_0 \in F$ ne réalise la distance minimale, sinon on aurait $\frac{1}{2} < \|f - g_0\| \leq \frac{1}{2}$, ce qui est impossible.

Dans un espace de Hilbert, ce problème ne se pose pas:

Théorème 4 (Théorème de projection orthogonale). *Soit $F \subset L^2(\mu)$ un sous-espace vectoriel fermé³. Pour tout $f \in L^2(\mu)$, on définit la **distance entre f et F** par:*

$$d(f, F) = \inf\{\|f - g\|, g \in F\}.$$

Alors il existe un unique $g_0 \in F$ tel que $\|f - g_0\| = d(f, F)$.

On appelle g_0 le **projeté orthogonal de f sur F** , noté $p_F(f)$.

L'application $p_F : f \in L^2(\mu) \mapsto p_F(f) \in F$ est appelée la **projection orthogonale sur F** .

Preuve:

Existence Par définition de la borne inférieure, il existe, pour chaque $n \in \mathbb{N}$, $g_n \in F$ tel que

$$d(f, F) \leq \|f - g_n\| \leq d(f, F) + \frac{1}{n+1}$$

de sorte que $(g_n)_n \in F^{\mathbb{N}}$ vérifie $\|f - g_n\| \rightarrow d(f, F)$.

▲ Ca n'implique pas que la suite $(g_n)_n$ elle-même converge vers quoi que ce soit: voir Contre-exemple ci-dessus.

Pour montrer que la suite $(g_n)_n$ est bel est bien convergente, on va utiliser l'hypothèse de complétude. On veut donc montrer que $(g_n)_n$ est de Cauchy.

³En fait, ce résultat est vrai pour tout sous-ensemble convexe fermé, pas uniquement pour les s.e.v.

D'après l'identité du parallélogramme (1), on a, pour tous $p, q \in \mathbb{N}$,

$$\begin{aligned}\|g_p - g_q\|^2 &= \|(f - g_p) - (f - g_q)\|^2 \\ &= 2\|f - g_p\|^2 + 2\|f - g_q\|^2 - \|(f - g_p) + (f - g_q)\|^2 \\ &= 2\|f - g_p\|^2 + 2\|f - g_q\|^2 - \|2f - (g_p + g_q)\|^2 \\ &= 2\|f - g_p\|^2 + 2\|f - g_q\|^2 - 4\left\|f - \frac{g_p + g_q}{2}\right\|^2\end{aligned}$$

Or, F est un s.e.v, donc $\frac{g_p + g_q}{2} \in F$; donc $\left\|f - \frac{g_p + g_q}{2}\right\| \leq d(f, F)$. Ce qui donne

$$\|g_p - g_q\|^2 \leq 2\|f - g_p\|^2 + 2\|f - g_q\|^2 - 4d(f, F) \xrightarrow{p, q \rightarrow \infty} 0$$

On en déduit que la suite $(g_n)_n \in F^{\mathbb{N}}$ est de Cauchy. Or F est fermé dans $L^2(\mu)$ complet, donc F est complet: on en déduit qu'il existe $g_\infty \in F$ tel que $g_n \xrightarrow{L^2(\mu)} g_\infty$.

De plus on a, par continuité de la norme, $\|f - g_n\| \xrightarrow{n \rightarrow \infty} \|f - g_\infty\|$; par unicité de la limite, ceci implique $\|f - g_\infty\| = d(f, F)$.

Unicité Supposons qu'il existe deux éléments $g, \tilde{g} \in F$ tels que $\|f - g\| = \|f - \tilde{g}\| = d(f, F)$. Alors, en utilisant l'identité du parallélogramme, on trouve comme précédemment

$$\begin{aligned}\|g - \tilde{g}\|^2 &= 2\|f - g\|^2 + 2\|f - \tilde{g}\|^2 - 4\left\|f - \frac{g + \tilde{g}}{2}\right\|^2 \\ &\leq 2\|f - g\|^2 + 2\|f - \tilde{g}\|^2 - 4d(f, F)^2 = 0\end{aligned}$$

d'où $g = \tilde{g}$.

□

Remarque: On a $p_F \circ p_F = p_F$.

Remarque: Dans \mathbb{R}^n , muni de son produit scalaire habituel, un sous-espace vectoriel F est décrit par un certain nombre d'équations linéaires:

$$x \in F \iff \begin{cases} \alpha_{1,1}x_1 + \dots + \alpha_{1,n}x_n = 0 \\ \vdots \\ \alpha_{p,1}x_1 + \dots + \alpha_{p,n}x_n = 0 \end{cases} \iff \langle x, u_1 \rangle = \dots = \langle x, u_p \rangle = 0,$$

où, pour chaque $j = 1, \dots, p$, $u_j = (\alpha_{j,1}, \dots, \alpha_{j,n}) \in \mathbb{R}^n$.

Dans ce cas, étant donné $x_0 \in \mathbb{R}^n$, trouver le point de F le plus proche de x est un problème d'optimisation sous contraintes

$$(\mathcal{P}_F) \begin{cases} \text{Minimiser } \|x_0 - x\|^2 \text{ t.q.} \\ \langle x, u_1 \rangle = \dots = \langle x, u_p \rangle = 0 \end{cases}$$

avec des contraintes linéaires $g_j(x) = \langle x, u_j \rangle$

On trouve que si $x \in F$ est solution, alors, par le théorème des extrema liés,

$$\nabla f(x) = \sum_{j=1}^p \lambda_j \nabla g_j(x) \iff x_0 - x = \sum_{j=1}^p \lambda_j u_j$$

\rightsquigarrow Pour tout $y \in F$,

$$\langle x - x_0, y \rangle = \sum_{j=1}^p \lambda_j \langle u_j, y \rangle = 0$$

i.e. $x - x_0 \in F^\perp$.

\rightsquigarrow La projection orthogonale dans $L^2(\mu)$ a des propriétés similaires:

Exercice: Démontrer les propriétés suivantes: Soit $F \in \mathcal{L}^2(\mu)$ un s.e.v. fermé.

1. $\forall f \in L^2(\mu)$, $f - p_F(f) \in F^\perp$.
2. Réciproquement, si $h \in F$ vérifie $\langle f - h, g \rangle = 0$ pour tout $g \in F$ alors $h = p_F(f)$.

En déduire:

1. $p_F : L^2(\mu) \rightarrow L^2(\mu)$ est une application linéaire continue de norme 1.
2. $\|p_F(f)\| = \|f\|$ ssi $f \in F$

Proposition 2 (Propriétés de l'orthogonal d'un sev dans un Hilbert). *Soit $F \subset L^2(\mu)$ un s.e.v, pas nécessairement fermé. Alors*

1. $\overline{F} = (F^\perp)^\perp$

2. F est dense dans $L^2(\mu)$ ssi $F^\perp = \{0\}$.

Preuve:

1. On procède par double inclusion:

$\boxed{\subset}$ On sait que $(F^\perp)^\perp$ est un s.e.v. fermé de $L^2(\mu)$. De plus, on montre sans trop de difficulté que $F \subset (F^\perp)^\perp$. En prenant l'adhérence de part et d'autre de l'inclusion, on trouve $\overline{F} \subset \overline{(F^\perp)^\perp} = (F^\perp)^\perp$.

$\boxed{\supset}$ Prenons $f \in (F^\perp)^\perp$. Alors $p_{\overline{F}}(f) \in \overline{F} \subset (F^\perp)^\perp$, donc $f - p_{\overline{F}}(f) \in (F^\perp)^\perp$. Mais on sait aussi que $f - p_{\overline{F}}(f) \in \overline{F}^\perp = F^\perp$.
Donc $f - p_{\overline{F}}(f) \in \overline{F}^\perp \cap (F^\perp)^\perp$, ce qui implique $f - p_{\overline{F}}(f) = 0$.
Donc $f = p_{\overline{F}}(f) \in \overline{F}$.

\rightsquigarrow On obtient l'égalité souhaitée.

2. On procède par double implication:

$\boxed{\Rightarrow}$ Supposons que F est un s.e.v. dense, c'est-à-dire que $\overline{F} = L^2(\mu)$.
Alors

$$F^\perp = \overline{F}^\perp = L^2(\mu)^\perp = \{0\}.$$

$\boxed{\Leftarrow}$ Supposons que $F^\perp = \{0\}$. Alors par 1.,

$$\overline{F} = ((F^\perp)^\perp) = \{0\}^\perp = L^2(\mu)$$

□

Remarque: Attention, dans un espace de Hilbert, on n'a pas forcément $(F^\perp)^\perp = F$! Par exemple considérons le sous-espace vectoriel des fonctions étagées dans $[0, 1]$.

$$F = \text{Vect}(\mathbf{1}_B, B \in \mathcal{B}([0, 1]))$$

Alors

$$F^\perp = \left\{ g \in L^2([0, 1]), \forall B \in \mathcal{B}([0, 1]), \int_B g d\lambda_1 = 0 \right\} = \{0\}$$

donc $(F^\perp)^\perp = \{0\}^\perp = L^2(\mu)$. Mais $F \neq L^2([0, 1])$: il y a des fonctions L^2 qui ne sont pas étagées !

Exemple: Reprenons l'exemple de

$$F = \{f \in L^2(\mathbb{R}), f(x) = 0 \text{ } \lambda_{1\text{-p.p}} \text{ sur } \mathbb{R}_+^*\}, F^\perp = \{f \in L^2(\mathbb{R}), f(x) = 0 \text{ } \lambda_{1\text{-p.p}} \text{ sur } \mathbb{R}_-\}$$

alors pour tout $f \in L^2(\mu)$,

$$p_F(f) = f\mathbf{1}_{\mathbb{R}_+}, p_{F^\perp}(f) = f\mathbf{1}_{\mathbb{R}_-}.$$

Théorème de représentation de Riesz

Pour chaque vecteur $v \in \mathbb{R}^n$, on peut définir une forme linéaire $\phi_v : u \in \mathbb{R}^n \mapsto \langle v, u \rangle$. Et inversement, toute forme linéaire $\phi : \mathbb{R}^n \rightarrow \mathbb{R}$ définit un vecteur $v_\phi = (\phi(e_1), \dots, \phi(e_n)) \in \mathbb{R}^n$ qui vérifie $\phi(u) = \langle v_\phi, u \rangle$ pour tout $u \in \mathbb{R}^n$.

\rightsquigarrow C'est comme ça qu'on définit le *gradient* d'une application $\mathcal{C}^1(U)$, $f : U \rightarrow \mathbb{R}$:

$$\forall x \in U, \forall h \in \mathbb{R}^n, df(x)(h) = \langle \nabla f(x), h \rangle$$

De même, à chaque $f \in L^2(\mu)$, on peut associer une forme linéaire

$$\phi_f : g \in L^2(\mu) \mapsto \int_X fg d\mu$$

\rightsquigarrow D'après l'inégalité de Cauchy-Schwarz, ϕ_f est bien définie et vérifie, pour tout $g \in L^2(\mu)$,

$$|\phi_f(g)| \leq \|f\| \|g\|$$

donc c'est une forme linéaire continue.

Exercice: Montrer que $\|\phi_f\|_{\mathcal{L}(L^2, \mathbb{R})} = \|f\|$.

Le théorème de représentation de Riesz affirme que, comme dans \mathbb{R}^n , toutes les formes linéaires *continues* sur $L^2(\mu)$ sont de ce type.

Contre-exemple: Considérons l'e.v.n. $\mathcal{C}^0([0, 1])$ muni du produit scalaire

$$\langle f, g \rangle = \int_0^1 f(t)g(t) dt$$

(c'est un sous-espace vectoriel pas fermé de $L^2([0, 1])$).

\rightsquigarrow Toutes les formes linéaires ne sont pas continues: ainsi, l'application

$$\phi : f \in \mathcal{C}^0([0, 1]) \mapsto f(0)$$

n'est pas continue pour la norme donnée par le produit scalaire. Il n'y a donc aucune chance de représenter ϕ par un vecteur de $\mathcal{C}^0([0, 1])$

\rightsquigarrow $(\mathcal{C}^0([0, 1]), \langle \cdot, \cdot \rangle)$ n'est pas un Hilbert, et ça se sent: l'application

$$\psi : f \in \mathcal{C}^0([0, 1]) \mapsto \int_0^{\frac{1}{2}} f(t) dt$$

est linéaire, continue, mais il n'existe aucune fonction $g \in \mathcal{C}^0([0, 1])$ telle que $\psi(f) = \langle f, g \rangle$ pour tout f .

(Et si on ne forçait pas g à être continue ?)

Théorème 5 (Théorème de représentation de Riesz). Soit $\phi \in \mathcal{L}(\mathcal{L}^2(\mu), \mathbb{R})$ une forme linéaire continue.

Alors il existe un unique⁴ $f_\phi \in L^2(\mu)$ tel que, pour tout $g \in L^2(\mu)$,

$$\phi(h) = \langle f_\phi, h \rangle = \int_X f_\phi h \, d\mu$$

Preuve: Soit $\phi \in \mathcal{L}(\mathcal{L}^2(\mu), \mathbb{R})$ une forme linéaire continue. Remarquons que la fonction f_ϕ que l'on cherche vérifie

$$h \in \text{Ker } \phi \iff \phi(h) = \int_X f_\phi h \, d\mu = 0 \iff h \in \{f\}^\perp$$

Ce qui nous donne l'idée de creuser du côté de $(\text{Ker } \phi)^\perp$.

Commençons par évacuer le cas simple: si $\phi = 0_{\mathcal{L}(L^2(\mu), \mathbb{R})}$ est la forme linéaire constante égale à 0, alors on peut prendre $f = 0_{L^2(\mu)}$.

Supposons donc que $\phi \neq 0_{\mathcal{L}(L^2(\mu), \mathbb{R})}$. Alors $\text{Ker } \phi \neq L^2(\mu)$. Puisque $\text{Ker } \phi = \phi^{-1}(\{0\})$ est un sous-espace vectoriel fermé⁵ de $L^2(\mu)$, $\text{Ker } \phi$ ne peut pas non plus être dense dans $L^2(\mu)$; on en déduit que

$$(\text{Ker } \phi)^\perp \neq \{0_{L^2}\}$$

On peut donc trouver $g \in (\text{Ker } \phi)^\perp$ tel que $g \neq 0_{L^2}$. On a alors $g \notin \text{Ker } \phi$ (Pourquoi ?) et on peut donc poser $f = \frac{g}{\phi(g)} \in (\text{Ker } \phi)^\perp$. L'intérêt, c'est que f vérifie $\phi(f) = 1$, ce qui est très pratique car alors, pour tout $h \in L^2(\mu)$,

$$\phi(h) = \phi(h)\phi(f) = \phi(\phi(h)f) \text{ i.e. } \phi(h - \phi(h)f) = 0$$

Donc, pour tout $h \in L^2(\mu)$, $h - \phi(h)f \in \text{Ker } \phi$. Puisque $f \in (\text{Ker } \phi)^\perp$, on a donc

$$\langle h - \phi(h)f, f \rangle = 0 \iff \phi(h) = \frac{\langle h, f \rangle}{\langle f, f \rangle} = \left\langle h, \frac{f}{\|f\|^2} \right\rangle$$

La fonction $f_\phi = \frac{f}{\|f\|^2}$ fonctionne.

Montrons que c'est la seule: soit $\tilde{f}_\phi \in L^2(\mu)$ une autre fonction telle que, pour tout h ,

$$\phi(h) = \langle f_\phi, h \rangle = \langle \tilde{f}_\phi, h \rangle$$

⁴Rappelons que l'unicité dans $L^2(\mu)$ signifie que, si deux fonctions f, \tilde{f} vérifient cette propriété, alors $f = \tilde{f}$ μ -p.p.

⁵Pourquoi ? D'ailleurs, à ce sujet, revoir l'exercice 3 du TD8 d'Analyse S5

On a alors

$$\begin{aligned}\|f_\phi - \tilde{f}_\phi\|^2 &= \langle f_\phi - \tilde{f}_\phi, f_\phi - \tilde{f}_\phi \rangle \\ &= \langle f_\phi, f_\phi - \tilde{f}_\phi \rangle - \langle \tilde{f}_\phi, f_\phi - \tilde{f}_\phi \rangle \\ &= \phi(f_\phi - \tilde{f}_\phi) - \phi(f_\phi - \tilde{f}_\phi) = 0\end{aligned}$$

donc $f_\phi = \tilde{f}_\phi$ dans $L^2(\mu)$. \square

Remarque: En utilisant l'inégalité de Hölder, on voit que, si p et q sont exposants conjugués, alors pour tout $g \in L^q(\mu)$, l'application

$$\phi_g : f \in L^p(\mu) \mapsto \int_X fg \, d\mu \in \mathbb{R}$$

est une forme linéaire continue sur $L^p(\mu)$. On en déduit une application (linéaire continue de norme 1)

$$g \in L^q(\mu) \mapsto \phi_g \in \mathcal{L}(L^p(\mu), \mathbb{R})$$

Ce qu'on a montré avec le théorème de représentation de Riesz, c'est que cette application (dite de dualité) est bijective pour $p = q = 2$.

C'est en fait vrai en général pour les espaces mesurés σ -finis, mais la preuve est plus difficile !

Application: le théorème de Radon-Nikodym

Définition 6. Soient μ et ν deux mesures positives sur un espace mesurable (X, \mathcal{T}) . On dit que ν est *absolument continue par rapport à μ* , ce que l'on note $\nu \ll \mu$, si, pour tout ensemble mesurable $A \in \mathcal{T}$,

$$(\mu(A) = 0) \implies (\nu(A) = 0)$$

Un cas (pas si) particulier: les mesures à densité

Soit (X, \mathcal{T}, μ) un espace mesuré et $f \in \mathcal{L}^0((X, \mathcal{T}), \mathbb{R}_+)$ une fonction mesurable positive, alors la mesure ν de densité f par rapport à μ , donnée par

$$\nu : A \in \mathcal{T} \mapsto \int_A^* f \, d\mu$$

est absolument continue par rapport à μ .

Exemples:

1. Soit $U \subset \mathbb{R}$ un borélien tel que $0 < \lambda_1(U) < \infty$.

Alors la mesure $\nu : B \in \mathcal{B}(\mathbb{R}) \mapsto \frac{\lambda_1(U \cap B)}{\lambda_1(U)}$ est absolument continue par rapport à λ_1 .

\rightsquigarrow A-t-elle une densité par rapport à λ_1 ? (Essayer avec $U = [a, b]$).

2. Soit X une variable aléatoire suivant une loi géométrique de paramètre $p \in]0, 1[$:

$$\mathbb{P}(X = k) = (1 - p)^{k-1} p$$

Alors la loi de X est la mesure de probabilités P_X , de densité $f(k) = (1 - p)^{k-1} p$ par rapport à la mesure de comptage sur \mathbb{N} , et elle est absolument continue par rapport à la mesure de comptage.

3. ...Ce qui n'est pas dur: toute mesure sur (X, \mathcal{F}) est absolument continue par rapport à la mesure de comptage.

Contre-exemple: Soit δ_0 la mesure de Dirac en 0 sur \mathbb{R} . Alors δ_0 n'est pas absolument continue par rapport à λ_1 puisque $\delta_0(\{0\}) = 1$ alors que $\lambda_1(\{0\}) = 0$. Et λ_1 n'est pas absolument continue par rapport à δ puisque $\delta([34, 34 + \pi]) = 0$ mais $\lambda_1([34, 34 + \pi]) = \pi > 0$.

Si X suit une loi géométrique, on peut aussi la représenter comme une variable aléatoire sur \mathbb{R} , avec la mesure de probabilités

$$P_X = \sum_{k \geq 1} (1 - p)^{k-1} p \delta_k$$

où δ_k est la mesure de Dirac en k . Mais dans ce cas, X n'a pas de densité par rapport à la mesure de Lebesgue sur \mathbb{R} : en effet, si c'était le cas, alors on aurait $P_X \ll \lambda_1$; or ce n'est pas le cas (même raison que Dirac ci-dessus).

\rightsquigarrow Quand est-ce qu'une v.a. admet une densité par rapport à λ ?

Ce que dit le théorème de Radon-Nikodym, c'est que réciproquement, une mesure absolument continue par rapport à une mesure μ σ -finie a une densité par rapport à μ . On rappelle:

Définition 7. Une mesure $\mu : \mathcal{F} \rightarrow [0, +\infty]$ est σ -finie s'il existe une suite d'ensembles mesurables disjoints $(E_n)_n \in \mathcal{F}^{\mathbb{N}}$ telle que

$$\forall n \in \mathbb{N}, \mu(E_n) < \infty \text{ et } \bigcup_{n \in \mathbb{N}} E_n = X$$

Exemple: La mesure de comptage sur $(\mathbb{N}, \mathcal{P}(\mathbb{N}))$ et la mesure de Lebesgue λ_1 sur $(\mathbb{R}, \mathcal{B}(\mathbb{R}))$ sont σ -finies. En revanche, la mesure de comptage sur $(\mathbb{R}, \mathcal{P}(\mathbb{R}))$ ne l'est pas.

Théorème 8. Soit μ une mesure positive sur un espace mesuré (X, \mathcal{F}) . On suppose que μ est σ -finie.

Soit ν une mesure sur (X, \mathcal{F}) telle que $\nu \ll \mu$. Alors il existe une unique⁶ fonction mesurable positive $h \in \mathcal{L}^0(\mu)$ telle que $d\nu = h d\mu$.

On appelle parfois h la *dérivée de Radon-Nikodym* de ν par rapport à μ , ce que, dans un abus scandaleux de notations, on note $h = \frac{d\nu}{d\mu}$.

Remarque: Si ν est une mesure finie (par exemple une mesure de probabilité), alors $h \in L^1(\mu)$.

Preuve: Comme souvent, l'hypothèse que μ est σ -finie va nous permettre de prouver d'abord le résultat dans le cas plus facile des mesures finies, avant de passer au cas général.

Cas particulier: Mesures finies Dans un premier temps, donc, supposons que μ et ν sont deux mesures finies sur (X, \mathcal{F}) . Remarquons qu'alors, $\mathbb{1}_X$ est intégrable à la fois par rapport à μ et ν , et donc, comme $|\mathbb{1}_X|^2 = \mathbb{1}_X$, $\mathbb{1}_X \in \mathcal{L}^2(\mu + \nu)$.

Par l'inégalité de Hölder, on en déduit que, pour tout $f \in \mathcal{L}^2(\mu + \nu)$, $f\mathbb{1}_X = f \in \mathcal{L}^1(\mu + \nu)$ voir aussi TD1 ex 8 donc

$$\int_X^* |f| d\nu \leq \int_X^* |f| d(\mu + \nu) = \int_X^* \mathbb{1}_X |f| d(\mu + \nu) \leq \sqrt{\mu(X) + \nu(X)} \|f\|_{L^2(\mu + \nu)}$$

Remarquons de plus que, si $f = \tilde{f}$ $(\mu + \nu)$ -p.p., alors

$$\nu(\{x \in X, f(x) \neq \tilde{f}(x)\}) \leq (\mu + \nu)(\{x \in X, f(x) \neq \tilde{f}(x)\}) = 0$$

donc $f = \tilde{f}$ ν -p.p., et donc leurs intégrales sur X par rapport à ν sont égales.

Ainsi, l'application

$$\phi : f \in L^2(\mu + \nu) \mapsto \int_X f d\nu \in \mathbb{R}$$

est bien définie. C'est une forme linéaire, et elle est continue puisque, pour tout $f \in L^2(\mu + \nu)$,

$$|\phi(f)| \leq \sqrt{\mu(X) + \nu(X)} \|f\|_{L^2(\mu + \nu)}.$$

⁶Comme toujours, au sens de l'égalité μ -presque partout.

D'après le théorème de représentation de Riesz, il existe donc $\tilde{g} \in L^2(\mu + \nu)$ telle que , pour tout $f \in L^2(\mu + \nu)$,

$$\phi(f) = \int_X f d\nu = \int_X f \tilde{g} d(\mu + \nu)$$

d'où, en réarrangeant,

$$\int_X f(1 - \tilde{g})d\nu = \int_X f \tilde{g} d\mu$$

Que peut-on apprendre sur \tilde{g} ?

- Soit $A = \{x \in X, \tilde{g}(x) < 0\} \in \mathcal{T}$, et $f = \mathbf{1}_A$. Alors $f \in L^2(\mu + \nu)$ et

$$\int_X f(1 - \tilde{g})d\nu \geq 0, \quad \int_X f \tilde{g} d\mu \leq 0$$

donc, puisque ces deux quantités sont égales, $\int_X f \tilde{g} d\mu = \int_A \tilde{g} d\mu = 0$. Ce qui permet de conclure que $\mu(\{x \in X, \tilde{g}(x) < 0\}) = 0$, donc $\tilde{g}(x) \geq 0$ μ -p.p.

- Prenons maintenant $B = \{x \in X, \tilde{g}(x) \geq 1\} \in \mathcal{T}$, et $f = \mathbf{1}_B$. Alors $f \in L^2(\mu + \nu)$ et

$$\int_X f(1 - \tilde{g})d\nu \leq 0, \quad \int_X f \tilde{g} d\mu \geq 0$$

donc $\int_X f \tilde{g} d\mu = \int_B \tilde{g} d\mu = 0$. Ce qui permet de conclure que $\mu(\{x \in X, \tilde{g}(x) \geq 1\}) = 0$, donc $\tilde{g}(x) < 1$ μ -p.p.

Modifions un peu \tilde{g} : on pose, pour tout $x \in X$,

$$g(x) = \begin{cases} \frac{1}{2} & \text{si } x \in A \cup B \\ \tilde{g}(x) & \text{sinon.} \end{cases}$$

Alors $\tilde{g} = g$ μ -p.p; puisque $\nu \ll \mu$ on en déduit que $\tilde{g} = g$ $(\mu + \nu)$ -p.p. Ces deux fonctions sont donc égales dans $L^2(\mu + \nu)$ et, pour tout $f \in L^2(\mu + \nu)$

$$\int_X f(1 - g)d\nu = \int_X f g d\mu \tag{2}$$

On a de plus, pour tout $x \in X$, $0 \leq g(x) < 1$, ce qui nous permet de définir

$$h : x \in X \mapsto \frac{g(x)}{1-g(x)} \in [0, +\infty[$$

\rightsquigarrow On va voir que h vérifie $d\nu = hd\mu$: il faut donc qu'on montre que, pour tout ensemble mesurable $E \in \mathcal{F}$,

$$\nu(E) = \int_X \mathbf{1}_E d\nu = \int_X \mathbf{1}_E h d\mu$$

Pour obtenir ça, on serait tenté de poser $f = \frac{\mathbf{1}_E}{1-g}$ dans (2), et ça semble plié. Mais rien ne garantit que cette fonction soit dans $\mathcal{L}^2(\mu + \nu)$!

Il va falloir être un peu plus prudent. L'avantage des mesures finies, c'est que toute fonction bornée est intégrable: on va donc approcher $f = \frac{\mathbf{1}_E}{1-g(x)}$ en s'assurant de ne pas dépasser les bornes. Pour ça, on pose, pour tout $n \in \mathbb{N}^*$,

$$f_n : x \in X \mapsto \begin{cases} \frac{\mathbf{1}_E(x)}{1-g(x)} & \text{si } \frac{\mathbf{1}_E}{1-g(x)} \leq n \\ 0 & \text{sinon.} \end{cases}$$

Alors $f_n \in L^2(\mu + \nu)$ et donc, pour tout n ,

$$\int_X f_n(1-g)d\nu = \int_X f_n g d\mu$$

Or, $(f_n)_n$ est une suite croissante qui converge simplement vers $\frac{\mathbf{1}_E}{1-g}$ sur X . Donc on peut appliquer le théorème de Beppo Levi, et en passant à la limite dans l'égalité précédente, on trouve

$$\int_X \frac{\mathbf{1}_E}{1-g}(1-g)d\nu = \int_X \frac{\mathbf{1}_E g}{1-g} d\mu, \text{ i.e. } \nu(E) = \int_X \mathbf{1}_E h d\mu$$

On en déduit que ν est la mesure de densité h par rapport à μ .

Cas général: Mesures σ -finies. On relâche l'hypothèse de finitude de μ et ν et on suppose maintenant seulement que μ et ν sont σ -finies. Soit donc $(E_n)_n \in \mathcal{F}^{\mathbb{N}}$, $(F_n)_n \in \mathcal{F}^{\mathbb{N}}$ deux suites d'ensembles mesurables disjoints telles que

$$\bigcup_{n \in \mathbb{N}} E_n = \bigcup_{n \in \mathbb{N}} F_n = X \text{ et } \forall n \in \mathbb{N}, \mu(E_n) < \infty, \nu(F_n) < \infty$$

En posant $A_{j,k} = E_j \cap F_k$ et en renumérotant⁷, on obtient une suite $(A_n)_n$ d'ensembles mesurables disjoints qui "marche" pour les deux mesures en même temps:

$$\bigcup_{n \in \mathbb{N}} A_n X \text{ et } \forall n \in \mathbb{N}, \mu(A_n) < \infty, \nu(A_n) < \infty$$

Pour chaque $n \in \mathbb{N}$, on peut alors définir deux mesures sur \mathcal{T} par

$$\mu_n(B) = \mu(B \cap A_n), \nu_n(B) = \nu(B \cap A_n).$$

Ce sont deux mesures finies, et, comme $\mu \ll \nu$,

$$\mu(B \cap A_n) = 0 \Rightarrow \nu(B \cap A_n) = 0 \text{ d'où } \nu_n \ll \mu_n$$

D'après le premier cas, il existe donc, pour chaque $n \in \mathbb{N}$, h_n mesurable positive telle que $d\nu_n = h_n d\mu_n$.

On pose $h_n = \sum_{n \in \mathbb{N}} h_n \mathbb{1}_{A_n}$ (pourquoi est-elle à valeurs dans \mathbb{R} ?). Alors, par le corollaire de Beppo-Levi sur les séries, pour tout $B \in \mathcal{T}$

$$\begin{aligned} \int_B^* h d\mu &= \int_X^* \mathbb{1}_B \sum_{n \in \mathbb{N}} h_n \mathbb{1}_{A_n} d\mu = \sum_{n \in \mathbb{N}} \int_X^* \mathbb{1}_B h_n \mathbb{1}_{A_n} d\mu \\ &= \sum_{n \in \mathbb{N}} \int_X^* \mathbb{1}_B h_n d\mu_n \\ &= \sum_{n \in \mathbb{N}} \nu_n(B) = \sum_{n \in \mathbb{N}} \nu(B \cap A_n) \\ &= \nu(B) \end{aligned}$$

comme souhaité.

Unicité: Supposons qu'il existe deux fonctions mesurables positives h, \tilde{h} telles que $d\nu = h d\mu = \tilde{h} d\mu$. Alors $A = \{x \in X, h(x) > \tilde{h}(x)\} \in \mathcal{T}$ et

$$\nu(A) = \int_A h d\mu = \int_A \tilde{h} d\mu \text{ donc } \int_A (h - \tilde{h}) d\mu = 0$$

Or $h - \tilde{h} > 0$ sur A donc ceci implique $\mu(A) = 0$.

On obtient de même que $\mu(\{x \in X, h(x) < \tilde{h}(x)\}) = 0$, et de là, $\mu(\{h \neq \tilde{h}\}) = 0$, autrement dit $h = \tilde{h}$ μ -p.p. □

⁷c'est-à-dire en notant $n = \psi(i, j)$ où $\psi : \mathbb{N}^2 \rightarrow \mathbb{N}$ est une bijection.

Application de l'application: Espérance conditionnelle

Soit $(\Omega, \mathcal{F}, \mathbb{P})$ un espace probabilisé, et X une variable aléatoire intégrable. Soit de plus $\mathcal{G} \subset \mathcal{F}$ une sous-tribu (qui représente l'information acquise).

Définition 9. Il existe une fonction mesurable $\mathbb{E}(X|\mathcal{G}) : (\Omega, \mathcal{G}) \rightarrow (\overline{\mathbb{R}}, \mathcal{B}(\overline{\mathbb{R}}))$, unique aux ensembles de mesure nulle près, et telle que, pour tout $E \in \mathcal{G}$,

$$\int_E \mathbb{E}(X|\mathcal{G}) d\mathbb{P} = \int_E X d\mathbb{P}$$

On l'appelle *espérance conditionnelle de X sachant \mathcal{G}* .

Si $X = \mathbf{1}_A$, on note $\mathbb{E}(X|\mathcal{G}) = \mathbb{P}(A|\mathcal{G})$.

\rightsquigarrow Pour montrer ce résultat, on applique Radon-Nikodym à la mesure de densité X par rapport à \mathbb{P} .

Exemple: Prenons $X = \mathbf{1}_A$. Soit $B \in \mathcal{F}$ est tel que $\mathbb{P}(B) > 0$, alors $\mathcal{G} = \{\emptyset, B, B^c, \Omega\}$ est une sous-tribu de \mathcal{F} . Alors on obtient que Y est \mathcal{G} -mesurable ssi il existe des constantes α, β telles que $Y = \alpha \mathbf{1}_B + \beta \mathbf{1}_{B^c}$. De là, on trouve

$$\mathbb{P}(A|\mathcal{G}) : \omega \in \Omega \mapsto \begin{cases} \frac{\mathbb{P}(A \cap B)}{\mathbb{P}(B)} & \text{si } \omega \in B \\ \frac{\mathbb{P}(A \cap B^c)}{\mathbb{P}(B^c)} & \text{si } \omega \in B^c \end{cases}$$

\rightsquigarrow On retrouve la probabilité conditionnelle "classique".

3 Bases orthonormées

Définition 10. Une famille $\{f_p, p \in \mathbb{N}\} \subset L^2(\mu)$ est

- *orthogonale* si pour tous $i \neq j$, $\langle f_i, f_j \rangle = 0$
 - *orthonormée* si de plus, pour tout i , $\|f_i\|_2 = 1$.
- \rightsquigarrow On a alors, pour tous i, j , $\langle f_i, f_j \rangle = \delta_{ij}$.

Exemple:

- La famille $\{\sin(nx), n \in \mathbb{N}\}$ est orthogonale dans $L^2([-\pi, \pi])$. Du coup, la famille $\{\frac{1}{\sqrt{\pi}} \sin(nx), n \in \mathbb{N}^*\}$ est orthonormée.

- Dans ℓ^2 , on considère, pour chaque $k \in \mathbb{N}$, la suite $(e_{k,n})_n$ définie par

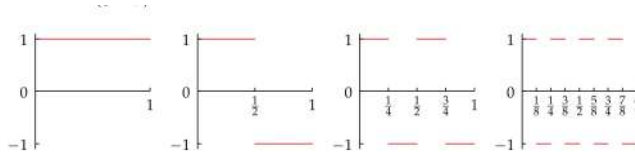
$$e_{k,n} = \begin{cases} 0 & \text{si } n \neq k \\ 1 & \text{sinon.} \end{cases}$$

Alors $\{(e_{k,n})_n, k \in \mathbb{N}\}$ est une famille orthonormée de ℓ^2 .

- (*Fonctions de Haar*) Pour chaque $k \in \mathbb{N}$, on pose

$$e_k : x \in [0, 1[\mapsto \begin{cases} 1 & \text{si } x \in [\frac{n-1}{2^k}, \frac{n}{2^k}[\text{ pour un } n \text{ impair} \\ -1 & \text{si } x \in [\frac{n-1}{2^k}, \frac{n}{2^k}[\text{ pour un } n \text{ pair} \end{cases}$$

Alors $\{e_k, k \in \mathbb{N}\}$ est une famille orthonormée de $L^2([0, 1])$.



Exercice 1:

1. Montrer que si $\{e_p, p \in \mathbb{N}\}$ est orthonormée, alors c'est une famille libre. En déduire que $L^2(\mu)$ n'est pas de dimension finie.
2. Montrer que si $\{e_p, p \in \mathbb{N}\}$ est orthonormée, alors, pour tous $i \neq j$, $\|e_i - e_j\|_2 = \sqrt{2}$.
En déduire que la boule unité fermée $\overline{B}(0, 1)$ de $L^2([-\pi, \pi])$ n'est pas compacte (et plus généralement la boule unité fermée de $L^2(I)$).

Exercice 2:

Soit $\{e_1, \dots, e_n\}$ une famille orthonormée finie de $L^2(\mu)$. Montrer que pour tous réels $\alpha_1, \dots, \alpha_n$

$$\left\| \sum_{j=1}^n \alpha_j e_j \right\|^2 = \sum_{j=1}^n |\alpha_j|^2$$

Soient $\{e_n, n \in \mathbb{N}\}$ une famille orthonormée dénombrable de $L^2(\mu)$ et $(a_n)_n \in \mathbb{R}^{\mathbb{N}}$. Montrer que la suite $(\sum_{k=0}^n a_k e_k)_n$ converge dans $L^2(I)$ vers une fonction f ssi $(a_n)_n \in \ell^2$, et qu'on a

$$\|f\|_2 = \sum_{k=0}^{\infty} |a_k|^2 = \|(a_n)_n\|_{\ell^2}$$

De plus, pour tout k ,

$$a_k = \langle f, e_k \rangle$$

Dans le dernier cas, attention, f n'est pas combinaison linéaire des $(e_n)_n$! Une combinaison linéaire a seulement un nombre fini de termes.

En particulier, on ne peut pas écrire

$$f = \sum_{k=0}^{\infty} \langle f, e_k \rangle e_k$$

Par contre, f est limite dans $L^2(\mu)$ de la suite $(\sum_{k=0}^n \langle f, e_k \rangle e_k)_n$.

Dans $L^2(\mu)$ qui est de dimension infinie, il n'y a aucun espoir de trouver une base finie, et encore moins une base orthonormée finie.

On pourrait alors espérer trouver une base infinie (dénombrable) $\{e_n, n \in \mathbb{N}\}$ de sorte que tout vecteur de $L^2(\mu)$ soit combinaison linéaire (finie) des e_n ; mais cette tentative est également vouée à l'échec.

Exemples

1. La famille $(e_{k,n})_k$ de ℓ^2 définie à l'exemple précédent n'engendrent pas ℓ^2 : sinon, toutes les suites ℓ^2 seraient nulles à partir d'un certain rang (et ce n'est pas le cas: donner un contre-exemple).
2. La famille $\{\frac{1}{\sqrt{\pi}} \sin(nx), n \in \mathbb{N}^*\}$ n'engendre pas $L^2([0, 1])$: sinon, toutes les fonctions de $\mathcal{L}^2([0, 1])$ seraient égales presque partout à une fonction \mathcal{C}^∞ . (Et ce n'est pas le cas: donner un contre-exemple)
3. La famille des fonctions de Haar n'engendre pas $L^2([0, 1])$ non plus: sinon, toutes les fonctions de $\mathcal{L}^2([0, 1])$ seraient en escalier.

On va donc se servir du procédé de passage à la limite de l'Exercice 2 pour sauver ce qu'on peut du merveilleux cadre des bases orthonormées.

On commence par montrer:

Proposition 3 (Inégalité de Bessel). *Si $\{e_p, p \in \mathbb{N}\}$ est une famille orthonormée de $L^2(I)$, alors, pour tout $f \in L^2(I)$*

$$\sum_{n=0}^{\infty} |\langle f, e_n \rangle|^2 \leq \|f\|_2^2$$

Exercice: Montrer cette inégalité, en écrivant

$$f = \left(\sum_{n=0}^{\infty} \langle f, e_n \rangle e_n \right) + \left(f - \sum_{n=0}^{\infty} \langle f, e_n \rangle e_n \right)$$

puis en appliquant le théorème de Pythagore.

Avec l'aide de Bessel, on va obtenir:

Proposition 4. Soient $\{e_n, n \in \mathbb{N}\}$ une famille orthonormée dénombrable de $L^2(\mu)$. Alors

$$\overline{\text{Vect}(e_n, n \in \mathbb{N})} = \left\{ \sum_{k \in \mathbb{N}} \alpha_k e_k, (\alpha_n)_n \in \ell^2 \right\}$$

De plus, pour tout $f \in \overline{\text{Vect}(e_n, n \in \mathbb{N})}$, $f = \lim_{n \rightarrow \infty} \sum_{k=0}^n \langle f, e_k \rangle e_k$.

Exercice: Montrons cette proposition.

1. $\boxed{\supset}$ Montrer que pour toute suite $(a_n)_n \in \ell^2$, $\sum_{k \in \mathbb{N}} \alpha_k e_k \in \overline{\text{Vect}(e_n, n \in \mathbb{N})}$.
2. $\boxed{\subset}$ Soit $f \in \overline{\text{Vect}(e_n, n \in \mathbb{N})}$. On pose $g = \sum_{k \in \mathbb{N}} \langle f, e_k \rangle e_k$. Montrer que $f - g \in \overline{\text{Vect}(e_n, n \in \mathbb{N})}^\perp$, puis que $f - g \in \text{Vect}(e_n, n \in \mathbb{N})^\perp$. En déduire que $f = g$.

Ce qui nous amène à formaliser une généralisation des bases orthonormées pour $L^2(\mu)$:

Définition 11. On dit qu'une famille orthonormée $\{e_n, n \in \mathbb{N}\}$ est une *base orthonormée de $L^2(\mu)$* si

$$\overline{\text{Vect}(e_n, n \in \mathbb{N})} = L^2(\mu)$$

autrement dit, si, pour tout $f \in L^2(\mu)$, $f = \lim_{n \rightarrow \infty} \sum_{k=0}^n \langle f, e_k \rangle e_k$ (convergence dans $L^2(\mu)$)

Remarque: Au risque d'insister lourdement, cette définition constitue un abus du terme "base" tel que vous le connaissez.

En algèbre linéaire, on dit qu'une famille (finie ou non) de vecteurs $\{e_j, j \in J\}$ est une base d'un espace vectoriel E si tout élément de E peut s'écrire, de manière unique, comme combinaison linéaire (finie, donc) des e_j :

$$\forall u \in E, \exists n \in \mathbb{N}, \exists \lambda_1, \dots, \lambda_n \in \mathbb{R}, u = \sum_{k=1}^n \lambda_k e_k$$

Ici, on dira que $\{e_j, j \in J\}$ est une base de $L^2(\mu)$ si, pour tout $f \in L^2(\mu)$, f est limite d'une suite de combinaisons linéaires des e_i :

$$\forall f \in L^2(\mu), \forall \varepsilon > 0 \exists n \in \mathbb{N}, \exists \lambda_1, \dots, \lambda_n \in \mathbb{R}, \left\| f - \sum_{k=1}^n \lambda_k e_k \right\| < \varepsilon$$

Pour les amateurs de noms, les bases "classiques" sont appelées bases de Hamel et les nouvelles, bases de Schauder.

Exemples:

1. Dans ℓ^2 , on considère, pour chaque $k \in \mathbb{N}$, la suite $(e_{k,n})_n$ définie par

$$e_{k,n} = \begin{cases} 0 & \text{si } n \neq k \\ 1 & \text{sinon.} \end{cases}$$

Alors $\{(e_{k,n})_n, k \in \mathbb{N}\}$ est une famille orthonormée de ℓ^2 .

2. La famille des fonctions de Haar, ainsi que la famille

$$\{1, \sin(2\pi nx), \cos(2\pi nx), n \in \mathbb{N}^*\}$$

sont des bases orthonormées de $L^2([0, 1])$.

Remarque:

- Pour montrer qu'une famille est une base au sens de la définition précédente, on peut utiliser la proposition 2: $\text{Vect}\{e_n, n \in \mathbb{N}\}$ est dense dans $L^2(\mu)$ ssi $\text{Vect}\{e_n, n \in \mathbb{N}\}^\perp = \{e_n, n \in \mathbb{N}\}^\perp = \{0\}$.

\rightsquigarrow Il suffit de montrer que, si $f \in \mathcal{L}^2(\mu)$ vérifie $\langle f, e_n \rangle = 0$ pour tout n , alors $f = 0$ μ -p.p.

\rightsquigarrow Dans le cas de $L^2(I)$, on peut de plus supposer f continue, car $\mathcal{C}^0(I)$ est dense dans $L^2(I)$.

- Pour la famille des fonctions trigonométriques, le fait qu'elles forment une base n'est pas évident à montrer, mais donne lieu à une des plus belles et plus utiles théories mathématiques: *l'analyse de Fourier*.

L'intérêt des bases dans un espace vectoriel de dimension finie, c'est qu'une fois qu'on a une base, alors notre espace vectoriel se comporte exactement comme \mathbb{R}^n .

L'intérêt des bases orthonormées de $L^2(\mu)$, c'est qu'une fois qu'on en a une, notre espace se comporte comme ℓ^2 :

Proposition 5 (Identité de Parseval). Soit $\{e_n, n \in \mathbb{N}\}$ une base orthonormée de $L^2(\mu)$, et $f, g \in L^2(\mu)$. Alors

1. $f = \sum_{n \in \mathbb{N}} \langle f, e_n \rangle e_n$
2. $\langle f, g \rangle = \sum_{n \in \mathbb{N}} \langle f, e_n \rangle \langle g, e_n \rangle$
3. $\|f\|^2 = \sum_{n \in \mathbb{N}} |\langle f, e_n \rangle|^2$

\rightsquigarrow Quand un espace $L^2(\mu)$ admet-il des bases orthonormées ?

Dans \mathbb{R}^n , vous connaissez le procédé de Gram-Schmidt, qui permet d'obtenir une b.o.n. à partir de n'importe quelle base.

Ce procédé marche aussi dans $L^2(\mu)$! Mais il faut l'adapter un peu, puisqu'on a changé la définition d'une base.

Notamment, on sait que, la plupart du temps, $L^2(\mu)$ n'admet pas de base au sens classique du terme. On ne peut donc pas partir d'une base, puis la Gram-Schmidter en une base orthonormée.

A la place, on va se concentrer sur ce qu'on a changé: la densité.

Définition 12. Un e.v.n. $(E, \|\cdot\|)$ est *séparable* s'il a un sous-ensemble dénombrable dense: autrement dit s'il existe $(\alpha_n)_n \in E^{\mathbb{N}}$ tel que

$$\overline{\{\alpha_n, n \in \mathbb{N}\}} = E$$

ou encore $\forall v \in E, \inf_{n \in \mathbb{N}} \|\alpha_n - v\| = 0$ i.e.

$$\forall \varepsilon > 0, \exists n_0 \in \mathbb{N} \text{ t.q. } \|\alpha_{n_0} - v\| < \varepsilon$$

Exemple 1. 1. ℓ^2 est séparable: l'ensemble dénombrable

$$\begin{aligned} & \bigcup_{n \in \mathbb{N}} \{(a_k)_{k \in \mathbb{N}} \in \ell^2 \mid \forall k, a_k \in \mathbb{Q} \text{ et } \forall k \geq n, a_k = 0\} \\ &= \bigcup_{n \in \mathbb{N}} \{(a_0, a_1, \dots, a_n, 0, \dots) \in \ell^2, a_j \in \mathbb{Q} \forall j\} \end{aligned}$$

est dense dans ℓ^2 .

2. $L^2([0, 1], L^2(\mathbb{R}))$ sont séparables: l'ensemble des combinaisons linéaires à coefficients rationnels de fonctions $\mathbf{1}_{]a, b[}$ avec $a, b \in \mathbb{Q}$ est un ensemble dénombrable dense.

Théorème 13 (Orthonormalisation de Gram-Schmidt, version L^2). Si $L^2(\mu)$ est séparable, alors il admet une base orthonormée.

Preuve: Soit $(f_n)_n \in L^2(\mu)^{\mathbb{N}}$ un sous-ensemble dénombrable dense. Quitte à supprimer les 0, on suppose que, pour tout n , $f_n \neq 0$. On construit par récurrence une famille orthonormée $(e_n)_n$ tq, pour tout n ,

$$\text{Vect}(f_1, \dots, f_n) \subset \text{Vect}(e_1, \dots, e_n)$$

ce qui nous donnera $\overline{\text{Vect}(e_n, n \in \mathbb{N})} = \overline{\text{Vect}(f_n, n \in \mathbb{N})} = L^2(\mu)$.

Initialisation On pose $e_1 = \frac{f_1}{\|f_1\|}$.

Hérédité Supposons qu'on a construit e_1, \dots, e_n , deux à deux orthogonaux et de norme 1, tq $\text{Vect}(f_1, \dots, f_n) = \text{Vect}(e_1, \dots, e_n)$.

\rightsquigarrow Si, pour tout $k \in \mathbb{N}$, $f_k \in \text{Vect}(e_1, \dots, e_n)$, alors (e_1, \dots, e_n) est une base de $L^2(\mu)$ et on s'arrête là. Sinon, on pose

$$m = \min\{p \in \mathbb{N}, e_p \notin \text{Vect}(e_1, \dots, e_n)\}$$

On pose alors

$$e_{n+1} = \frac{f_{n+1} - \sum_{k=1}^n \langle f_{n+1}, e_k \rangle e_k}{\|f_{n+1} - \sum_{k=1}^n \langle f_{n+1}, e_k \rangle e_k\|}$$

$\rightsquigarrow e_{n+1}$ est de norme 1 et orthogonal à tous les $e_k, k = 1, \dots, n$. De plus, puisqu'on a choisit pour m le plus petit entier tel que $f_m \notin \text{Vect}(e_1, \dots, e_n)$, on a bien

$$\text{Vect}(f_1, \dots, f_{n+1}) \subset \text{Vect}(e_1, \dots, e_{n+1})$$

□

Retour sur la projection orthogonale

Proposition 6. Soit $F \subset L^2(\mu)$ un s.e.v. fermé, et $(e_n, n \in \mathbb{N})$ une base orthonormée de F . Alors, pour tout $f \in L^2(\mu)$,

$$p_F(f) = \sum_{k \in \mathbb{N}} \langle f, e_k \rangle e_k$$

Application: Trouver le polynôme $P \in \mathbb{R}_{10}[X]$ qui minimise

$$\int_{-1}^1 |\sqrt{|x|} - P(x)|^2 dx$$

Recette: On travaille d'abord dans $L^2([-1, 1])$, et on cherche le projeté orthogonal de $f : x \mapsto \sqrt{|x|}$ sur le sous-espace $F = \mathbb{R}_{10}[X]$.

1. On vérifie que F est fermé.
2. On applique Gram-Schmidt à la base $(1, X, \dots, X^{10})$ de F , ce qui nous donne une base orthonormée de polynômes P_0, \dots, P_{10} (on les appelle *polynômes de Legendre*, voir https://fr.wikipedia.org/wiki/Polyn%C3%B4me_de_Legendre pour une liste):

$$P_0 = \frac{1}{\sqrt{2}},$$

$$P_1 = \frac{\sqrt{6}}{2}X, \dots$$

$$P_{10} = \frac{\sqrt{42}}{512}(46189X^{10} - 109395X^8 + 90090X^6 - 30030X^4 + 3465X^2 - 63)$$

(Si, si, vraiment.)

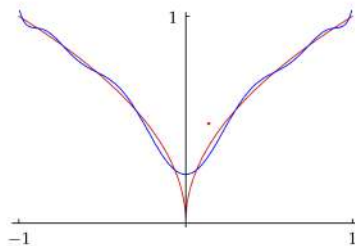
3. On calcule, pour $k = 0, \dots, 10$,

$$\langle f, P_k \rangle = \int_{-1}^1 \sqrt{|x|} P_k(x) dx$$

et on obtient que le polynôme recherché est

$$P(X) = \frac{1}{2944}(51051X^{10} - 138567X^8 + 139230X^6 - 64350X^4 + 15015X^2 + 693)^8$$

ce qui nous donne une plutôt bonne approximation de f :



Retour sur la projection orthogonale

Proposition 7. Soit $\phi \in \mathcal{L}(L^2(\mu), \mathbb{R})$ une forme linéaire continue et $(e_n, n \in \mathbb{N})$ une base orthonormée de $L^2(\mu)$. Alors la somme

$$\sum_{k \in \mathbb{N}} \phi(e_k) e_k$$

⁸(Si, si, vraiment.)

converge dans $L^2(\mu)$. Notons h la limite, alors, pour tout $f \in L^2(\mu)$,

$$\phi(f) = \langle f, h \rangle.$$

De plus, $\|\phi\|_{\mathcal{L}(L^2(\mu), \mathbb{R})} = (\sum_{k \in \mathbb{N}} |\phi(e_k)|^2)^{\frac{1}{2}}$.