

Algèbre bilinéaire, séries de Fourier (Mat404-ex244)

2019

Table des matières

1 Motivations	7
1.1 L'équation de la chaleur.	7
1.2 L'équation des ondes.	9
2 Séries numériques.	11
2.1 Convergence des séries	12
2.2 Les séries et le calcul sur machine.	21
3 Rappels d'algèbre linéaire.	25
3.1 Espaces vectoriels : définitions et exemples.	26
3.2 Bases et coordonnées.	27
3.3 Applications linéaires.	30
3.4 Calcul Matriciel.	31
3.5 Matrices carrées	34
4 Formes bilinéaires.	37
4.1 Le produit scalaire canonique sur \mathbb{R}^3 .	37
4.2 Formes bilinéaires : définitions et exemples.	38
4.3 Formes bilinéaires : représentation matricielle.	39
4.4 Orthogonalité.	41
4.5 Calcul effectif d'une base φ -orthogonale.	45
5 Produits scalaires.	53
5.1 Produit scalaires, définitions et exemples.	53
5.2 Produits scalaires : géométrie.	55
5.3 Procédé d'orthonormalisation de Gram-Schmidt.	60
5.4 Diagonalisation orthogonale des matrices symétriques.	67
5.5 Matrices orthogonales et unitaires.	70
6 Séries de Fourier.	73
6.1 Approximants de Fourier, coefficients de Fourier et séries de Fourier : définitions et exemples.	74
6.2 Séries en sin et cos.	82
6.3 Convergence des séries de Fourier.	85
6.4 Solutions d'équations aux dérivées partielles.	86
6.4.1 L'équation de la chaleur.	86
6.4.2 L'équation des ondes.	89
6.5 La transformée de Fourier.	92
A Appendice : espace-temps, bases et forme de Minkowski.	95
B Appendice : le tenseur d'inertie d'un corps rigide.	99
B.1 Application : rotation libre d'un objet avec symétrie rotationnelle.	101

C Appendice : les coniques et quadriques.	103
D Appendice : Formes hermitiennes.	107

Index

absolument convergente, 15

chaleur, équation de la, 7
convergente, absolument, 15
critère de d'Alembert, 16
critère de Riemann, 18

d'Alembert, critère de, 16

équation de la chaleur, 7
équation des ondes, 9

forme quadratique, 39
Fourier, séries de, 73

général, terme, 12

ondes, équation des, 9

partielle, somme, 12
produit scalaire, 53

quadratique, forme, 39

Riemann, critère de, 18

série, 12
séries de Fourier, 73
scalaire, produit, 53
somme partielle, 12

terme général, 12

Chapitre 1

Motivations

Dans cette section, nous allons considérer deux problèmes de physique dont la solution semble nécessiter des techniques mathématiques plus sophistiquées que celles que vous avez utilisées jusqu'à présent.

1.1 L'équation de la chaleur.

Supposons donnée une barre chauffée de façon inhomogène. Comment se diffuse la chaleur dans cette barre ?

On considère une barre d'un matériau homogène de longueur finie L , la température initiale (au temps $t = 0$) étant donnée par une fonction $\varphi : [0, L] \rightarrow \mathbb{R}$, ou $x \mapsto \varphi(x)$.

On suppose que les échanges de chaleur entre la barre et l'air sont négligeables et que les extrémités de la barre sont au contact d'un parfait isolant, ce qui implique qu'il n'y a pas de flux de chaleur à travers ces extrémités. En particulier le gradient de la chaleur y est nul. On veut comprendre comment la chaleur se diffuse dans la barre avec le temps ; autrement dit, si $T(x, t)$ est la température dans la barre au point x en un temps t , alors on veut comprendre l'évolution de la valeur de $T(x, t)$ avec t .

Des considérations physiques montrent que T doit satisfaire à l'équation, dite équation de la chaleur :

$$\frac{\partial T}{\partial t} = k \frac{\partial^2 T}{\partial x^2}$$

où k est une constante positive (la *conductivité thermique*) qui dépend du matériau. Nous avons en plus les conditions au bord

$$\frac{\partial T}{\partial x}(0, t) = \frac{\partial T}{\partial x}(L, t) = 0 \text{ pour tout } t,$$

qui traduisent l'absence de flux de chaleur à travers les extrémités, et la condition initiale

$$T(x, 0) = \varphi(x).$$

Oublions d'abord la condition $T(x, 0) = \varphi(x)$. Autrement dit, on cherche les solutions vérifiant seulement les conditions au bord

$$\frac{\partial T}{\partial x}(0, t) = \frac{\partial T}{\partial x}(L, t) = 0 \text{ pour tout } t.$$

L'équation étant beaucoup trop compliquée pour être résolue avec les méthodes dont nous disposons actuellement, nous allons commencer par simplement chercher des *exemples* de fonctions qui la satisfont. Les fonctions à *variables séparés* (c'est-à-dire s'écrivant dans la forme $T(x, t) = f(x)g(t)$) sont une source féconde d'exemples satisfaisant à des équations aux dérivées partielles, puisque de telles équations se simplifient souvent dans ce cas. Nous commencerons donc par chercher des solutions de la forme $T(x, t) = f(x)g(t)$. On a alors que

$$f(x)g'(t) = kf''(x)g(t),$$

soit

$$\frac{f''(x)}{f(x)} = \frac{g'(t)}{kg(t)},$$

au moins sur la région où ni f ni g ne s'annule. Notons que le membre de gauche est une fonction qui ne dépend que de x et le membre de droite est une fonction qui ne dépend que de t : comme x et t sont indépendantes, cela implique qu'il existe $\alpha \in \mathbb{R}$ tel que

$$\frac{f''(x)}{f(x)} = \frac{g'(t)}{kg(t)} = \alpha.$$

Ainsi, on a

$$f''(x) - \alpha f(x) = 0$$

et

$$g'(t) - k\alpha g(t) = 0.$$

On a donc $g(t) = \lambda e^{k\alpha t}$ pour $\lambda \in \mathbb{R}$, et donc $g(t) \neq 0$ pour tout $t \geq 0$ (car on cherche T non identiquement nulle). La contrainte

$$\frac{\partial T}{\partial x}(0, t) = \frac{\partial T}{\partial x}(L, t) = 0$$

entraîne alors $f'(0) = f'(L) = 0$. Pour résoudre l'équation en f il nous faut maintenant distinguer 3 cas.

1. Cas 1 : $\alpha = 0$. On a alors $f''(x) = 0$, et donc $f(x) = b_0x + a_0$. Les conditions $f'(0) = f'(L) = 0$ imposent alors facilement $f(x) = a_0$ pour tout x . On a donc une première solution de base

$$T_0(x, t) = 1.$$

2. Cas 2 : $\alpha > 0$. On peut alors poser $\alpha = \omega^2$ et f est de la forme $f(x) = ae^{\omega x} + be^{-\omega x}$. Les conditions que $f'(0) = 0$ et $f'(L) = 0$ impliquent alors $a = b = 0$, et f est identiquement nulle, ce qui est exclu.
3. Cas 3 : $\alpha < 0$. On peut alors poser $\alpha = -\omega^2$ et

$$f(x) = a \cos(\omega x) + b \sin(\omega x), \quad a, b, \in \mathbb{R}.$$

Puisque $f'(0) = 0$ on a $b = 0$, et puisque $f'(L) = 0$ on a $a \sin(\omega L) = 0$. Puisque l'on cherche T non nulle, on a $a \neq 0$ et donc $\sin(\omega L) = 0$.

Ainsi $\omega L = \pi n$ pour $n \geq 0$, et donc pour chaque n , on a une solution de la forme

$$T_n(x, t) = \cos\left(\frac{n\pi x}{L}\right) e^{-\frac{\pi^2 n^2}{L^2} kt}.$$

Pour chaque entier positif $n \geq 0$ nous avons donc une solution de l'équation de la chaleur

$$T_n(x, t) = \cos\left(\frac{n\pi x}{L}\right) e^{-\frac{\pi^2 n^2}{L^2} kt}.$$

(Nous pouvons intégrer la solution $T_0(x, t) = 1$ dans cette famille de solutions en considérant qu'il s'agit de $T_0(x, t) = \cos(0x)e^{-0t}$.) La condition initiale $\varphi_n(x)$ correspondant à la solution $T_n(x, t)$ est donnée par $\varphi_n(x) = T_n(x, 0)$, c'est à dire

$$\varphi_n(x) = \cos\left(\frac{n\pi x}{L}\right).$$

Nous avons donc trouvé une solution à l'équation de la chaleur pour certaines conditions initiales bien particulières, c'est à dire certains cosinus. Est ce qu'on peut en construire d'autres solutions pour d'autres conditions initiales ?

Notons tout d'abord que l'équation de la chaleur a une propriété très utile :

Remarque 1.1.1 (Linéarité de l'équation de la chaleur.) Si $T_1(x, t)$ et $T_2(x, t)$ sont deux solutions à l'équation de la chaleur alors pour tous réels $\lambda, \mu \in \mathbb{R}$

$$T(x, t) = \lambda T_1(x, t) + \mu T_2(x, t)$$

est encore une solution de cette équation. (Une telle fonction est appelée une combinaison linéaire de T_1 et T_2). On dit alors que l'équation de la chaleur est une **équation linéaire**.

Exercice. Démontrer que l'équation de la chaleur est une équation linéaire.

En particulier, toute fonction qui est une combinaison linéaire finie

$$T(x, t) = \lambda_0 T_0(x, t) + \lambda_1 T_1(x, t) + \lambda_2 T_2(x, t) + \dots + \lambda_n T_n(x, t)$$

avec des nombres réels $\lambda_0, \dots, \lambda_n$ est encore une solution de l'équation de la chaleur. Cette solution correspond à la condition initiale

$$\varphi(x) = T(x, 0)$$

c'est à dire

$$\varphi(x) = \lambda_0 + \lambda_1 \cos\left(\frac{\pi x}{L}\right) + \lambda_2 \cos\left(\frac{2\pi x}{L}\right) + \dots + \lambda_n \cos\left(\frac{n\pi x}{L}\right).$$

Nous savons donc trouver une solution pour l'équation de la chaleur pour certaines conditions initiales bien particulières : celles qui s'écrivent comme des sommes finies de cosinus.

Et il vient assez naturellement l'idée : *Peut-on résoudre cette équation de la même façon pour une condition initiale φ quelconque en l'écrivant comme une "somme infinie" de cosinus ?*

1.2 L'équation des ondes.

Nous nous penchons maintenant sur un autre cas, en apparence différent. Un fil horizontal de longueur L , soumis à une tension T et de densité linéaire μ , est tenu aux deux extrémités. Au temps $t = 0$ il est relâché et se met à osciller librement dans un plan vertical.

Soit $D(x, t)$ la fonction égale au déplacement vertical¹ à l'instant t de la partie du fil qui se trouve (à l'équilibre) à une distance x d'une des extrémités.

Nous avons cette fois les conditions aux bords

$$D(0, t) = D(L, t) = 0,$$

qui traduisent le fait que le fil est attaché aux extrémités. Si le déplacement initial du fil est décrit par la fonction $\phi(x)$ alors nous avons aussi les conditions initiales

$$D(x, 0) = \phi(x) \text{ et } \frac{\partial D}{\partial t}(x, 0) = 0,$$

cette dernière condition traduisant le fait que le fil est relâché à l'instant $t = 0$ et se trouve donc à ce moment-là au repos. Des considérations physiques montrent que l'évolution de D est décrite par l'équation des ondes

$$\frac{\partial^2 D}{\partial t^2} = k \frac{\partial^2 D}{\partial x^2}$$

où k est la constante positive $k = \frac{T}{\mu}$. Cherchons comme ci-dessus des solutions de la forme $f(x)g(t)$. On a alors

$$f(x)g''(t) = kf''(x)g(t),$$

1. par rapport à l'équilibre

soit

$$\frac{f''(x)}{f(x)} = \frac{g''(t)}{kg(t)}.$$

Notons que le membre de gauche est une fonction qui ne dépend que de x et le membre de droite est une fonction qui ne dépend que de t : comme x et t sont deux variables indépendantes, cela implique qu'il existe $\alpha \in \mathbb{R}$ tel que

$$\frac{f''(x)}{f(x)} = \frac{g''(t)}{kg(t)} = \alpha.$$

Ainsi, on a

$$f''(x) - \alpha f(x) = 0 \quad \text{et} \quad g''(t) - k\alpha g(t) = 0.$$

Le même raisonnement que ci-dessus nous montre que cette équation a une solution telle que $D(0, t) = D(L, t) = 0$ si et seulement si il existe un entier n tel que $\alpha = \frac{n^2\pi^2}{L^2}$ et dans ce cas on a une solution donnée par

$$D_n(x, t) = \sin\left(\frac{n\pi x}{L}\right) \cos\left(\frac{\sqrt{kn}\pi t}{L}\right).$$

Ceci nous donne une solution au problème pour une condition initiale

$$\varphi_n(x) = \sin\left(\frac{n\pi x}{L}\right).$$

Remarque 1.2.1 *L'équation des ondes est encore une équation linéaire,*

Exercice Démontrer que l'équation des ondes est linéaire.

Puisque la fonction $D_n(x, t)$ est une solution pour chaque n , toute combinaison linéaire finie

$$D(x, t) = \lambda_1 D_1(x, t) + \lambda_2 D_2(x, t) + \dots + \lambda_k D_k(x, t)$$

ou les λ_k sont des nombres réels est encore une solution de l'équation de la chaleur. Cette solution correspond à la condition initiale

$$\varphi(x) = \lambda_1 \sin\left(\frac{\pi x}{L}\right) + \lambda_2 \sin\left(\frac{2\pi x}{L}\right) + \dots + \lambda_n \sin\left(\frac{n\pi x}{L}\right).$$

Nous savons donc trouver une solution à cette équation pour des conditions initiales bien particulières : celles qui s'écrivent comme des sommes finies de sinus.

Il vient la même idée que dans le cas de l'équation de la chaleur : *Peut-on résoudre cette équation pour une condition initiale quelconque φ en écrivant φ comme une "somme infinie" de sinus ?*

Avant de se lancer dans des spéculations sur les sommes infinies de fonctions, il faudrait déjà savoir ce que veut dire une somme infinie de nombres. Dans le prochain chapitre, nous allons étudier les séries² numériques.

2. C'est le nom que les mathématiciens donnent aux sommes infinies.

Chapitre 2

Séries numériques.

Vous avez déjà rencontré au cours de vos études l'équation suivante

$$1 + \frac{1}{2} + \frac{1}{4} + \frac{1}{8} + \dots = 2$$

ou le symbole “...” se comprend comme “et ainsi de suite jusqu’à l’infini”. Quel sens donner à cette équation, et en particulier, quel sens donner à son membre de gauche $1 + \frac{1}{2} + \frac{1}{4} + \frac{1}{8} + \dots$? Ca ne peut pas signifier “le résultat qu’on obtient en effectuant une infinité d’additions” puisqu’il est impossible de faire une infinité d’additions.

La somme infinie à gauche doit être comprise comme une limite. En écrivant cette équation, nous disons la chose suivante :

En prenant n assez grand, nous pouvons rendre la somme finie

$$1 + \frac{1}{2} + \frac{1}{4} + \dots + \frac{1}{2^n}$$

aussi proche qu’on veut de 2.

La somme infinie

$$1 + \frac{1}{2} + \frac{1}{4} + \frac{1}{8} + \dots,$$

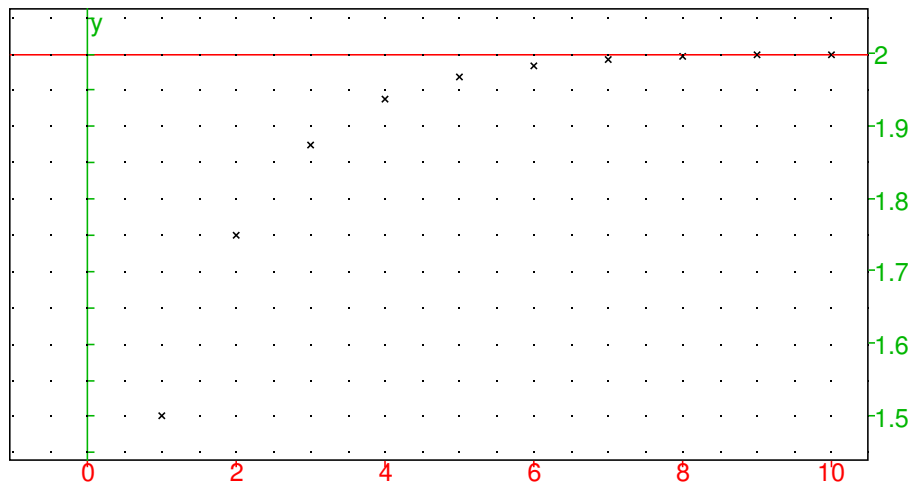
que l’on écrit aussi $\sum_{n=0}^{\infty} \frac{1}{2^n}$, doit être compris comme la **limite de la suite des sommes partielles** $s_k = \sum_{n=0}^k \frac{1}{2^n}$.

```
S(N) := sum(1/2^n, n, 0, N); evalf(S(20)); evalf(S(100)); sum(1/2^n, n, 0, inf)
```

$$N \rightarrow \sum \left(\frac{1}{2^n}, n, 0, N \right), 1.99999904633, 2.0, 2$$

Représentation graphique des premières sommes partielles

```
S(N) := sum(1/2^n, n, 0, N); seq(point(N, S(N)), N, 0, 10); line(y=2, color=red)
```



2.1 Convergence des séries

Définition 2.1.1 Soit (u_n) une suite de nombres réels ou complexes. On définit la suite de sommes partielles $(s_k)_k$ (également notée $(\sum_{n \geq 0} u_n)$)

$$s_k = u_0 + u_1 + u_2 + \dots + u_k = \sum_{n=0}^k u_n$$

Nous appelons cette suite la **série** de terme général u_n .

On peut adapter la définition lorsqu'on commence la somme à un rang $m \neq 0$, par exemple $m = 1$ si u_0 n'est pas défini, $(\sum_{n \geq m} u_n)$ est la suite des sommes partielles $s_k = u_m + u_{m+1} + \dots + u_{k+m}$.

Exemples

1. Si on pose, comme ci-dessus, $u_n = \frac{1}{2^n}$ et on considère la série $(\sum_{n \geq 0} u_n)$ alors la somme partielle $s_k = \sum_{n=0}^k u_n$ est donnée par

$$s_k = 1 + \frac{1}{2} + \dots + \frac{1}{2^k} = 2 - \frac{1}{2^k}.$$

Si on considère la série $(\sum_{n \geq 3} u_n)$ alors la somme partielle

$$s_k = \sum_{n=3}^{k+3} u_n$$

est donnée par

$$s_k = \frac{1}{8} + \frac{1}{16} + \dots + \frac{1}{2^{k+3}} = \frac{1}{4} - \frac{1}{2^{k+3}}.$$

2. Si on pose $u_n = 1$ pour tout n et on considère la série $(\sum_{n \geq 0} u_n)$ alors les sommes partielles $s_k = \sum_{n=0}^k u_n$ sont données pour tout k par

$$s_k = 1 + 1 + \dots + 1 = k + 1.$$

3. Si on pose¹ $u_n = (-1)^n$ et on considère la série $(\sum_{n \geq 0} u_n)$ alors les sommes partielles $s_k = \sum_{n=0}^k u_n$ sont données par

$$s_0 = 1$$

1. c'est à dire $u_0 = 1, u_1 = -1, u_2 = 1, u_3 = -1$ et ainsi de suite.

$$s_1 = 1 - 1 = 0$$

$$s_2 = 1 - 1 + 1 = 1$$

et ainsi de suite, c'est à dire que pour tout k paire nous avons que $s_k = 1$ et pour tout k impaire nous avons que $s_k = 0$.

4. Si on pose $u_n = \frac{1}{n^2}$ et on considère la série $(\sum_{n \geq 1} u_n)$ alors la somme partielle s_k est le nombre réel

$$s_k = 1 + \frac{1}{4} + \frac{1}{9} + \dots + \frac{1}{k^2}.$$

Contrairement aux autres cas, nous ne disposons d'aucune formule générale pour cette somme partielle.

Lorsque cette suite de sommes partielles $(s_k)_{k \geq 0}$ est convergente, on dit que sa limite est la valeur de la "somme infinie"

$$\sum_{n=0}^{\infty} u_n = u_0 + u_1 + u_2 + \dots$$

Définition 2.1.2 Soit $(u_n)_{n \geq m}$ une suite infinie et considérons la série $(\sum_{n \geq m} u_n)$. Nous disons que la série $(\sum_{n \geq m} u_n)$ admet comme limite le nombre fini l si la suite $(s_k)_{k \geq 0}$ de sommes partielles converge vers l ,

$$\lim_{k \rightarrow \infty} s_k = l.$$

Dans ce contexte, nous disons que l est la somme de la série $(\sum_{n \geq m} u_n)$ et nous écrivons

$$\sum_{n \geq m}^{\infty} u_n = l.$$

Attention ! Les deux notations

$$\left(\sum_{n \geq m} u_n \right)$$

et

$$\sum_{n=m}^{\infty} u_n,$$

qui sont très proches, désignent quand même des choses différentes. Lorsque nous écrivons $(\sum_{n \geq m} u_n)$ nous parlons de la **suite** de sommes partielles $(s_k)_{k \geq 0}$ alors que $\sum_{n=m}^{\infty} u_n$ désigne la **limite** de cette suite (en supposant, bien sur, qu'elle existe).

Proposition 2.1.3 Le terme général d'une série convergente tend vers 0.

En effet, soit $(\sum_{n \geq m} u_n)$ une série et soit $(s_k)_{k \geq 0}$ ses sommes partielles. Si $(\sum_{n \geq m} u_n)$ converge vers l alors on a que

$$s_k \rightarrow_{k \rightarrow \infty} l$$

$$s_{k-1} \rightarrow_{k \rightarrow \infty} l$$

donc

$$s_k - s_{k-1} \rightarrow_{k \rightarrow \infty} 0$$

Or $s_k - s_{k-1} = u_{k+m}$ donc $u_k \rightarrow_{k \rightarrow \infty} 0$.

Remarque 2.1.4 – Par contre-apposition, si le terme général d'une suite ne tend pas vers 0 alors la série diverge. Par exemple $\sum_{n \geq 0} (-1)^n$ diverge parce que son terme général $(-1)^n$ ne tend pas vers 0. De même une suite géométrique de raison λ diverge lorsque $|\lambda| \geq 1$.

– Attention ! La réciproque est fautive. Il existe des séries divergentes dont le terme général ne tend pas vers 0, par exemple on montre que $\sum_{n>0} \frac{1}{n}$ diverge alors que son terme général $\frac{1}{n}$ tend vers 0 (cf. proposition 2.1.13 plus bas)

Exemples.

1. Pour la série $(\sum_{n \geq 0} \frac{1}{2^n})$ nous avons que la somme partielle

$$s_k = 2 - \frac{1}{2^k} \xrightarrow{k \rightarrow \infty} 2.$$

On peut donc écrire

$$\sum_{n=0}^{\infty} \frac{1}{2^n} = 2.$$

2. Soit maintenant λ un nombre réel ou complexe tel que $|\lambda| < 1$, et considérons la série $(\sum_{n \geq 0} \lambda^n)$. La somme partielle

$$s_k = 1 + \lambda + \dots + \lambda^k$$

peut être calculée par l'astuce suivante :

$$(1 - \lambda)s_k = s_k - \lambda s_k$$

$$(1 - \lambda)s_k = (1 + \lambda + \dots + \lambda^k) - (\lambda + \lambda^2 + \dots + \lambda^{k+1})$$

$$(1 - \lambda)s_k = 1 - \lambda^{k+1}$$

$$s_k = \frac{1 - \lambda^{k+1}}{1 - \lambda}.$$

Puisque $|\lambda| < 1$ nous avons que $\lambda^k \xrightarrow{k \rightarrow \infty} 0$ donc

$$s_k \xrightarrow{k \rightarrow \infty} \frac{1}{1 - \lambda}.$$

Autrement dit, la série géométrique de raison $|\lambda| < 1$ converge et on a

$$\sum_{n=0}^{\infty} \lambda^n = \frac{1}{1 - \lambda}.$$

`sum(lambda^n, n, 0, k); assume(-1 < lambda < 1)`
`); sum(lambda^n, n, 0, inf)`

$$\frac{\lambda^{k+1}}{(\lambda - 1)} - \frac{1}{\lambda - 1}, \lambda, -\frac{1}{\lambda - 1}$$

3. La série $(\sum_{n \geq 0} 1)$ a pour sommes partielles $s_k = k + 1$. Cette suite n'est pas convergente : sa limite n'est pas finie. On dit alors que la série $(\sum_{n \geq 0} 1)$ est **divergente**².
4. La série $(\sum_{n \geq 0} (-1)^n)$ a pour sommes partielles

$$s_k = 1 \text{ si } k \text{ paire, } s_k = 0 \text{ si } k \text{ impaire.}$$

Cette suite de sommes partielles, bien que bornée (les sommes partielles n'approchent pas ∞) ne converge pas. On dit encore une fois que la série $(\sum_{n \geq 0} (-1)^n)$ est divergente.³

2. On aurait pu aussi remarquer que $u_n \not\xrightarrow{n \rightarrow \infty} 0$ et donc cette série, par Remarque 2.1.3, ne converge pas.

3. De même, cette série ne peut pas converger par Remarque 2.1.3.

5. Même si nous ne disposons pas de formule pour les sommes partielles $s_k = \sum_{n=1}^k \frac{1}{n^2}$ il est possible de montrer que cette suite converge vers une limite finie. Nous verrons à la fin du semestre que

$$\lim_{k \rightarrow \infty} s_k = \frac{\pi^2}{6}$$

que nous pouvons aussi écrire

$$\sum_{n=1}^{\infty} \frac{1}{n^2} = \frac{\pi^2}{6}.$$

`s := sum(1/n^2, n, 1, inf); evalf(s); evalf(sum(1/n^2, n, 1, 100));`

$$\frac{1}{6} \cdot \pi^2, 1.64493406685, 1.63498390018$$

La remarque suivante, qui suit des propriétés de linéarité des suites, est souvent utile dans l'étude des séries.

Proposition 2.1.5 (Linéarité de la convergence des séries) Soient $(\sum_{n \geq m} u_n)$ et $(\sum_{n \geq m} v_n)$ deux séries convergentes réelles ou complexes, de limites u et v respectivement. Alors pour tout $\lambda, \mu \in \mathbb{C}$, la série

$$\left(\sum_{n \geq m} \lambda u_n + \mu v_n \right)$$

est convergente, avec limite $\lambda u + \mu v$.

Le cas des séries réelles à termes positifs est assez simple.

Lemme 2.1.6 Soit $(\sum_{n \geq m} u_n)$ une série réelle dont toutes les termes u_n sont positifs. Pour tout $k \geq m$ soit s_k la somme partielle

$$s_k = \sum_m^{k+m} u_n.$$

Il y a alors deux possibilités

1. la suite $(s_k)_{k \geq 0}$ converge vers une limite finie l . Autrement dit, la série $(\sum_{n \geq m} u_n)$ est convergente
2. la suite de sommes partielles $(s_k)_{k \geq 0}$ tend vers $+\infty$.

En effet la différence entre deux sommes partielles consécutives est un terme de la suite u_n donc est positif ($s_k - s_{k-1} = u_{k+m} \geq 0$) donc la suite s_k est croissante. Si elle est majorée, elle converge vers une limite finie (toute suite croissante majorée est convergente). Sinon, elle n'est pas majorée et tend donc vers $+\infty$.

Pour appliquer ce lemme, il sera utile de se ramener à des séries à termes positifs. On introduit la notion de série absolument convergente.

Définition 2.1.7 Soit $(\sum_{n \geq m} u_n)$ une série. On dit que $(\sum_{n \geq m} u_n)$ est absolument convergente si la série $(\sum_{n \geq m} |u_n|)$ est convergente.

On a le résultat suivant.

Proposition 2.1.8 Toute série absolument convergente est convergente.

Idée de la preuve (hors programme) : cela résulte de l'inégalité triangulaire sur les sommes partielles

$$\left| \sum_{n=m}^M u_n \right| \leq \sum_{n=m}^M |u_n|$$

Comme $\sum_n |u_n|$ est convergente, le terme de droite peut être rendu aussi petit que l'on veut pourvu que l'on choisisse m assez grand. Cela permet d'établir rigoureusement la convergence de la suite des sommes partielles de u_n (c'est ce qu'on appelle une suite de Cauchy).

Attention : la réciproque de cette proposition est fautive : il existe des séries réelles convergentes qui ne sont pas absolument convergentes. Leur comportement est parfois surprenant – par exemple, en permutant les termes d’une telle série on peut la rendre divergente, ou la faire converger vers n’importe quel nombre réel. De plus ces séries convergent lentement, il faut calculer des sommes partielles à des rangs d’indice élevé pour avoir une valeur approchée de la somme. Les séries absolument convergentes sont donc plus intéressantes ! Mais on n’a pas toujours le choix (par exemple certaines séries de Fourier).

Remarque 2.1.9 Le comportement de la série de terme général $u_n = (-1)^n$, qui diverge sans tendre vers $+\infty$, n’est possible que parce que certains termes de cette série sont négatifs.

Le critère de d’Alembert traite le cas des séries qui se comportent comme des séries géométriques.

Proposition 2.1.10 Soit $\sum u_k$ une série telle que $\frac{u_{k+1}}{u_k} \rightarrow_{k \rightarrow \infty} \lambda$. Si $|\lambda| < 1$ alors la série $\sum u_k$ est absolument convergente. Si $|\lambda| > 1$ alors la série $\sum u_k$ diverge.

Preuve :

Si $|\lambda| > 1$, le terme général de la série ne tend pas vers 0, donc elle diverge. Si $|\lambda| < 1$, on choisit $\epsilon > 0$ tel que $|\lambda| + \epsilon < 1$ (par exemple $\epsilon = \frac{1-|\lambda|}{2}$). Puis on applique la définition de la convergence de la suite $|u_{n+1}/u_n|$ vers $|\lambda|$, il existe un rang l tel que

$$\forall n > l, \quad |u_{n+1}|/|u_n| \leq |\lambda| + \epsilon$$

Donc pour tout $m \geq n$, on a :

$$|u_m| \leq (|\lambda| + \epsilon)^{m-n} |u_n|$$

. Comme $|\lambda| + \epsilon < 1$ la série

$$\sum_m |u_n| (\lambda + \epsilon)^{m-n} = |u_n| \sum_m (\lambda + \epsilon)^{m-n}$$

converge. Comme $|u_m|$ est positive, la série $\sum_{m \geq l} |u_m|$ converge.

Les séries à convergence de type géométrique convergent assez rapidement, ces séries sont très utiles pour calculer des valeurs approchées de fonctions trigonométriques, exponentielles, logarithmes etc. (cf. la section 2.2). Mais toutes les séries ne convergent pas aussi rapidement, par exemple les séries de Fourier qui seront abordées en fin de cours. Pour déterminer leur nature, on va utiliser un critère plus fin, le critère des équivalents.

Le lemme 2.1.6 entraîne le corollaire suivant.

Corollaire 2.1.11 Soient $(\sum_{n \geq m} u_n)$, $(\sum_{n \geq m} v_n)$ des séries avec un nombre fini de termes négatifs. Alors :

1. Si $u_n \leq v_n$ pour tout n et $(\sum_{n \geq m} v_n)$ converge alors $(\sum_{n \geq m} u_n)$ converge aussi.
2. Si $u_n \sim_{n \rightarrow \infty} u'_n$ alors la série $(\sum_{n \geq m} u_n)$ converge si et seulement si la série $(\sum_{n \geq m} v_n)$ converge aussi.

Quitte à commencer la somme à un indice $m > 0$, on peut supposer que toutes les termes des deux suites sont positifs.

1. Soit $(s_k)_{k \geq 0}$ la suite de sommes partielles de la série $(\sum_{n \geq m} u_n)$. Soit $(t_k)_{k \geq 0}$ la suite de sommes partielles de la série $(\sum_{n \geq m} v_n)$. Puisque $(\sum_{n \geq m} v_n)$ converge, i.e. la suite $(t_k)_{k \geq 0}$ converge, donc est majorée. Puisque $u_n \leq v_n$ pour tout n , $s_k \leq t_k$ pour tout k , donc la suite s_k est majorée. Par Lemme 2.1.6 il suit que la série $(\sum_{n \geq m} u_n)$ converge.
2. Puisque $u_n \sim_{n \rightarrow \infty} v_n$ et $u_n, v_n > 0$ il existe des nombres réels strictement positifs α, β tels que, pour tout n ,

$$\alpha v_n \geq u_n.$$

$$\beta u_n \geq v_n.$$

Nous avons donc par (1) que

$$\left(\sum_{n \geq m} u_n\right) \text{ converge} \Rightarrow \left(\sum_{n \geq m} \beta u_n\right) \text{ converge} \Rightarrow \left(\sum_{n \geq m} v_n\right) \text{ converge}.$$

De même

$$\left(\sum_{n \geq m} v_n\right) \text{ converge} \Rightarrow \left(\sum_{n \geq m} \alpha v_n\right) \text{ converge} \Rightarrow \left(\sum_{n \geq m} u_n\right) \text{ converge}.$$

Ceci termine la démonstration du Corollaire 2.1.11.

Exercice. Montrer que le Corollaire 2.1.11 est toujours valable lorsque : les deux séries sont à termes négatifs,

Remarque 2.1.12 Si la suite u_n ne comprend qu'un nombre fini de termes négatifs et la suite v_n satisfait $v_n \sim u_n$ alors la suite v_n ne comprend qu'un nombre fini de termes négatifs. Il suffit donc de vérifier cette condition sur une seule des deux suites.

Vous avez étudié en 1ère année les développements limités. Utilisant ces développements, il est très souvent possible de montrer qu'une suite donnée est équivalente à une suite de la forme $(\frac{1}{n^s})_{n \geq 1}$.

Exemples

1. Considérons $u_n = \sin(\frac{1}{n})$. Nous avons que

$$\sin\left(\frac{1}{n}\right) =_{n \rightarrow \infty} \frac{1}{n} + o\left(\frac{1}{n}\right),$$

c'est à dire que $\sin(\frac{1}{n}) \sim_{n \rightarrow \infty} \frac{1}{n}$.

taylor(sin(1/n), n=inf, 2)

$$n^{-1} + (n^{-1})^3 \text{ order_size}(n^{-1})$$

2. Considérons $u_n = e^{1/n} - 1$. Nous avons que

$$e^{1/n} =_{n \rightarrow \infty} 1 + \frac{1}{n} + o\left(\frac{1}{n}\right)$$

et donc $(e^{1/n} - 1) =_{n \rightarrow \infty} \frac{1}{n} + o(\frac{1}{n})$, c'est à dire $e^{1/n} \sim_{n \rightarrow \infty} \frac{1}{n}$.

taylor(exp(1/n)-1, n=inf, 2)

$$n^{-1} + \frac{1}{2} \cdot (n^{-1})^2 + (n^{-1})^3 \text{ order_size}(n^{-1})$$

3. Considérons $u_n = \frac{\cos(\frac{1}{n}) - 1}{n}$. Nous avons que

$$\cos\left(\frac{1}{n}\right) =_{n \rightarrow \infty} 1 - \frac{1}{2n^2} + o\left(\frac{1}{n^2}\right)$$

et donc $\cos(\frac{1}{n}) - 1 \sim_{n \rightarrow \infty} \frac{-1}{2n^2}$, d'où il vient que

$$\frac{\cos(\frac{1}{n}) - 1}{n} \sim_{n \rightarrow \infty} \frac{-1}{2n^3}.$$

taylor((cos(1/n)-1)/n, n=inf, 3)

$$-\frac{1}{2} \cdot (n^{-1})^3 + (n^{-1})^5 \text{ order_size}(n^{-1})$$

4. Considérons $u_n = \frac{\cos(\frac{1}{n})-1}{\sin(\frac{1}{n})} \sim_{n \rightarrow \infty} \frac{-\frac{1}{2n^2}}{\frac{1}{n}} = \frac{-1}{2n}$.

$$\text{taylor}((\cos(1/n)-1)/\sin(1/n), n=\text{inf}, 3)$$

$$-\frac{1}{2 \cdot n} + (n^{-1})^3 \text{order_size}(n^{-1})$$

La corollaire 2.1.11 sera donc un outil très puissant pour établir la convergence des séries, à condition de savoir quand la série

$$\left(\sum_{n>0} \frac{1}{n^s} \right)$$

converge.

Proposition 2.1.13 (Critère de Riemann.) Pour tout nombre réel positif $s > 0$ la suite infinie

$$\left(\sum_{n \geq 1} \frac{1}{n^s} \right)$$

diverge si $s \leq 1$ et converge si $s > 1$.

Preuve

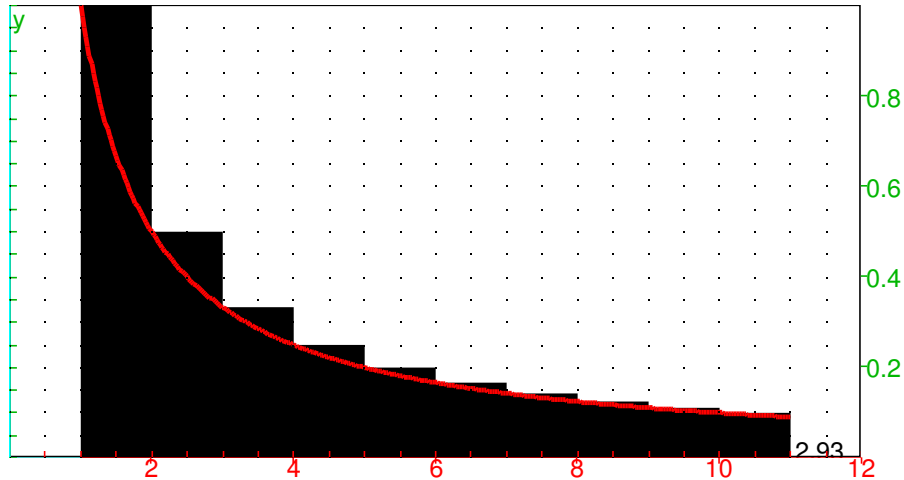
On doit déterminer quand la suite de sommes partielles

$$s_k = \sum_{n=1}^k \frac{1}{n^s}$$

converge. Puisque la suite (u_n) est à termes positifs il suffit par le lemme 2.1.6 de savoir quand la suite s_k est majorée. Nous allons faire cela par une technique très puissante : comparaison d'une somme avec une intégrale. Il y a en effet un lien fort entre l'intégrale $\int_1^k f(x)dx$ et la somme $\sum_{n=1}^k f(n)$.

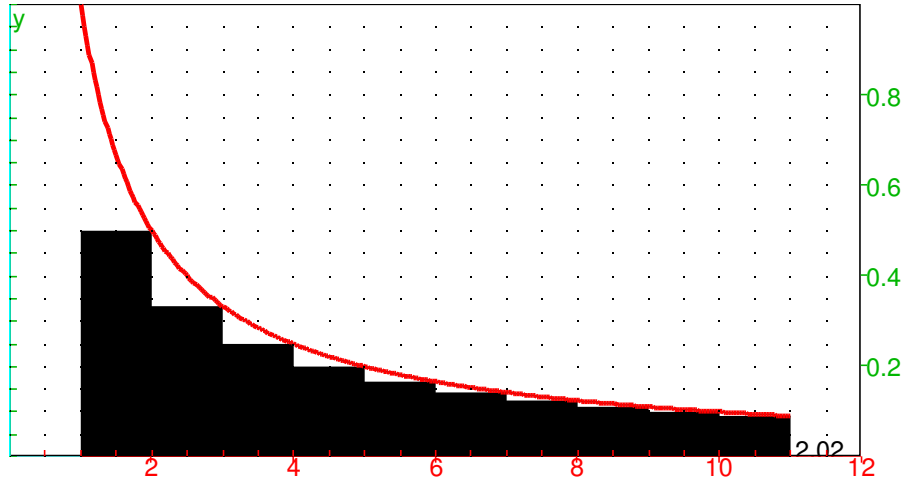
Illustration graphique de la nature identique de la convergence d'une série ou de l'intégrale correspondante pour une fonction décroissante telle que $f(x) = 1/x$. En noir la somme des rectangles dont l'aire est une somme partielle de la série (car chaque rectangle a pour dimension horizontale 1 et verticale un des $f(n)$). On compare avec l'aire sous la courbe représentative de f (en rouge), de deux façons, en prenant les rectangles à gauche ou à droite.

```
f(x) := 1/x; gl_y=0..1; gl_x=0..12;
plotarea(f(x), x=1..11, 10, rectangle_gauche);
```



Si l'aire des rectangles gauches noirs tend vers une limite finie, alors l'aire sous la courbe aussi (puisque'elle est majorée par l'aire des rectangles).

```
f(x):=1/x; gl_y=0..1; gl_x=0..12;
plotarea(f(x),x=1..11,10,rectangle_droit);
```



Réciproquement, si l'aire sous la courbe tend vers une limite, l'aire des rectangles droits aussi. On calcule ensuite

$$\int_1^k \frac{1}{t^s} dt = 1 - \frac{1}{(s-1)k^{s-1}}, \quad s \neq 1$$

et on fait tendre k vers $+\infty$, il y a convergence lorsque $s > 1$. Attention au cas particulier $s = 1$, on a alors un logarithme qui ne tend pas vers une limite finie.

Une autre preuve plus formelle. Puisque $s > 0$ pour tout x tel que $x \in [n, n+1]$ nous avons que

$$\frac{1}{n^s} \geq \frac{1}{x^s} \geq \frac{1}{(n+1)^s}.$$

Il en suit que

$$\int_n^{n+1} \frac{1}{n^s} dx \geq \int_n^{n+1} \frac{1}{x^s} dx \geq \int_n^{n+1} \frac{1}{(n+1)^s} dx.$$

cest à dire que pour tout entier positif n nous avons que

$$\frac{1}{n^s} \geq \int_n^{n+1} \frac{1}{x^s} dx \geq \frac{1}{(n+1)^s}.$$

En sommant ces inégalités, nous obtenons que

$$1 + \frac{1}{2^s} + \dots + \frac{1}{n^s} \geq \int_1^2 \frac{1}{x^s} dx + \int_2^3 \frac{1}{x^s} dx + \dots + \int_n^{n+1} \frac{1}{x^s} dx \geq \frac{1}{2^s} + \frac{1}{3^s} + \dots + \frac{1}{(n+1)^s}.$$

Autrement écrit

$$\sum_{n=1}^k \frac{1}{n^s} \geq \int_1^{k+1} \frac{1}{x^s} dx \geq \sum_{n=2}^{k+1} \frac{1}{n^s},$$

c'est à dire

$$s_k \geq \int_1^{k+1} \frac{1}{x^s} dx \geq s_k - \frac{k}{k+1}.$$

Re-organisant ces équations, nous obtenons que

$$\int_1^k \frac{1}{x^s} dx \leq s_k \leq \int_1^k \frac{1}{x^s} dx + 1.$$

Nous allons maintenant distinguer 3 cas, selon que s est plus grand que, égale à ou plus petit que 1.

1. Cas 1 : $s > 1$. Nous avons alors que

$$\int_1^k x^{-s} dx = \left[\frac{x^{1-s}}{1-s} \right]_1^k = \frac{1 - k^{1-s}}{s-1} \leq \frac{1}{s-1}.$$

On a donc que pour tout k

$$s_k \leq \frac{1}{s-1} + 1.$$

La suite s_k est donc majorée et la série $(\sum_{n \geq 1} \frac{1}{n^s})$ converge.

2. Cas 2 : $s = 1$. Nous avons alors que

$$\int_1^k x^{-1} dx = [\log(x)]_1^k = \log(k) \rightarrow_{k \rightarrow \infty} \infty.$$

Nous avons donc que $s_k \geq \log(k)$ pour tout k : la suite s_k ne peut donc pas être majorée et par le lemma 2.1.6 la suite $(\sum_{n \geq m} \frac{1}{n})$ doit diverger.

3. Cas 3 : $s < 1$. Pour tout entier positif n nous avons alors que $\frac{1}{n^s} \geq \frac{1}{n} > 0$. Comme $(\sum_{n \geq 1} \frac{1}{n})$ ne converge pas, il résulte du Lemma 2.1.6 que $(\sum_{n \geq 1} \frac{1}{n^s})$ ne converge pas non plus.

Ceci termine démonstration de la proposition 2.1.13

Exemples :

Les exemples qui suivent montrent à quel point l'attelage du Corollaire 2.1.11 avec la Proposition 2.1.13 est un outil puissant pour déterminer si des séries positives convergent ou divergent.

1. Soit $u_n = \sin(\frac{1}{n})$ pour tout $n \geq 1$. Nous avons vu que

$$u_n \sim_{n \rightarrow \infty} \frac{1}{n}.$$

Puisque la série $(\sum_{n \geq 1} \frac{1}{n})$ diverge par la proposition 2.1.13, il résulte du corollaire 2.1.11 que $(\sum_{n \geq 1} u_n)$ diverge aussi.

2. Soit $u_n = 1 - \cos(\frac{1}{n})$ pour tout $n \geq 1$. Nous avons vu que $u_n \sim_{n \rightarrow \infty} \frac{1}{2n^2}$. Puisque la série $(\sum_{n \geq 1} \frac{1}{n^2})$ converge par la Proposition 2.1.13, il résulte de la corollaire 2.1.11 que $(\sum_{n \geq 1} u_n)$ converge aussi.

3. Soit $u_n = \frac{1 - \cos(\frac{1}{n})}{\sqrt{n}}$. Nous avons que

$$u_n \sim_{n \rightarrow \infty} \frac{1}{2n^2 \sqrt{n}} = 2n^{-\frac{5}{2}}.$$

`taylor((1-cos(1/n))/sqrt(n), n=inf, 3)`

$$\frac{1}{2} \cdot \sqrt{n^{-1}} (n^{-1})^2 + \sqrt{n^{-1}} (n^{-1})^4 \text{ order_size}(n^{-1})$$

Puisque la série $(\sum_{n \geq 1} \frac{1}{n^{5/2}})$ converge par la Proposition 2.1.13, il résulte de la corollaire 2.1.11 que $(\sum_{n \geq 1} u_n)$ converge aussi.

4. Soit $u_n = \sin(\frac{1}{n}) (e^{\frac{1}{\sqrt{n}}} - 1)$. Par les développements limités, on a que

$$\sin(\frac{1}{n}) \sim_{n \rightarrow \infty} \frac{1}{n}$$

et

$$e^{\frac{1}{\sqrt{n}}} - 1 \sim_{n \rightarrow \infty} n^{-1/2}.$$

Il en suit que

$$u_n \sim_{n \rightarrow \infty} \frac{1}{n * n^{1/2}} = n^{-3/2}.$$

```
taylor(sin(1/n) * (exp(1/sqrt(n)) - 1), n=inf, 3)
```

$$\frac{\sqrt{n^{-1}}}{n} + \frac{1}{2} \cdot (n^{-1})^2 + \frac{1}{6} \cdot \sqrt{n^{-1}} (n^{-1})^2 + (n^{-1})^3 \text{ order_size } (n^{-1})$$

Puisque la série $(\sum_{n \geq 1} \frac{1}{n^{3/2}})$ converge par la Proposition 2.1.13, il résulte du corollaire 2.1.11 que $(\sum_{n \geq 1} u_n)$ converge aussi.

2.2 Les séries et le calcul sur machine.

Les séries peuvent être utilisées dans les logiciels de calcul pour calculer des fonctions transcendentes (sinus, cosinus, exponentielle, logarithme...) ou spéciales que vous verrez peut-être dans vos études (error function, fonctions de Bessel, de Airy, W de Lambert, etc.).

Un processeur est capable de faire des opérations arithmétiques de base⁴ :

1. addition, soustraction de deux nombres
2. multiplication, division de deux nombres.

A l'aide de ces opérations, il s'agit de coder des fonctions plus avancées citées ci-dessus. Mais la plupart des fonctions qu'on souhaite calculer - cos, sin, exp, log et ainsi de suite - ne s'expriment pas exactement à l'aide de opérations de base + - * / d'ailleurs. Les seules fonctions qui peuvent être calculées exactement utilisant seulement ces opérations sont les fractions de polynômes :

$$P(x) = a_0 + a_1x + a_2x^2 + \dots + a_nx^n$$

où x est une variable et a_0, \dots, a_n sont des nombres.

Alors, comment faire pour faire calculer $\sin(x)$ (par exemple) à un ordinateur ? Il y a plusieurs méthodes : l'une d'elles utilise les séries de Taylor⁵. Voilà comment cela se passe dans le cas de la fonction $\sin(x)$.

1. On applique la théorie de la série de Taylor pour écrire $\sin(x)$ comme une somme infinie de puissances de x . En l'occurrence, nous savons que

$$\sin(x) = x - \frac{x^3}{3!} + \frac{x^5}{5!} - \frac{x^7}{7!} + \dots = \sum_{n=0}^{\infty} \frac{(-1)^n x^{2n+1}}{(2n+1)!}.$$

(Le lecteur pourra vérifier que la série converge en appliquant le critère de d'Alembert).

2. Les sommes partielles $s_k(x)$ de cette série de Taylor sont des polynômes et peuvent être calculées en utilisant seulement les opérations d'addition, multiplication, division. Dans le cas de $\sin(x)$ nous avons que

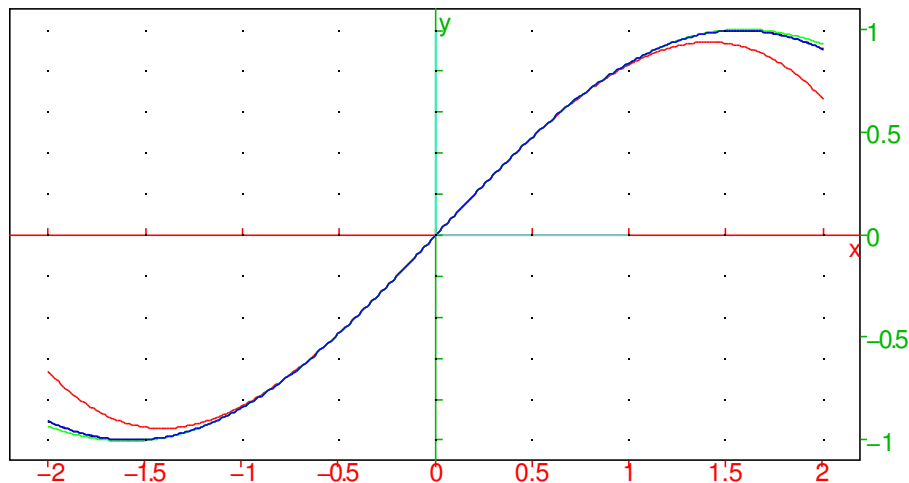
$$s_k(x) = x - \frac{x^3}{3!} + \frac{x^5}{5!} + \dots + \frac{(-1)^k x^{2k+1}}{(2k+1)!}.$$

3. Pour k assez grand le polynôme $s_k(x)$ sera une approximation acceptable pour $\sin(x)$. **Attention** : ici le k "assez grand" dépendra de x : plus x sera petit, plus on pourra utiliser une valeur basse pour k . On le voit sur la représentation graphique ci-dessous de la fonction sinus et des sommes partielles de la série de Taylor en 0.

```
S(n, x) := sum((-1)^k * x^(2*k+1) / (2*k+1)!, k, 0, n);
plot([sin(x), S(1, x), S(2, x), S(3, x)], x, -2, 2, color=[black, red, green, blue])
```

4. Aujourd'hui, les coprocesseurs arithmétiques sont capables de calculer les fonctions sinus, cosinus, exponentielles, etc. en décomposant le calcul en opérations arithmétiques de base

5. Une autre méthode populaire est l'algorithme CORDIC



La question plus difficile à laquelle il faudrait répondre est la suivante : si je cherche (par exemple) à calculer $\sin(2)$ à 10^{-3} près, quelle valeur de k dois-je utiliser ? Pour minimiser le temps de calcul de la machine, il est souhaitable de prendre k le plus petit possible.

L'erreur $E_k(x)$ que je commets lorsque j'approche $\sin(x)$ par

$$s_k(x) = x - \frac{x^3}{3!} + \frac{x^5}{5!} + \dots + \frac{(-1)^k x^{2k+1}}{(2k+1)!}$$

est donnée par le formule⁶

$$E_k(x) = \frac{(-1)^k x^{2k+1}}{(2k+1)!} + \dots$$

On peut démontrer que dès lors que $x^2 < (2k+2)(2k+3)$ nous avons que

$$|E_k(x)| \leq \frac{|x|^{2k+3}}{(2k+3)!}$$

Pour garantir que $E_k(x)$ est petit il suffit donc de s'assurer que

$$|x|^{2k+3}/(2k+3)!$$

est petit. Ici, comme on voudrait calculer $\sin(2)$ à 10^{-3} près, il suffit de trouver k tel que $2^{2k+3}/(2k+3)! < 10^{-3}$. Des calculs nous donnent que

$$2^3/3! = 4/3$$

$$2^5/5! = 4/15$$

$$2^7/7! = 8/315$$

$$2^9/9! = 4/2835$$

$$2^{11}/11! = 8/155925 < 10^{-3}.$$

`seq([k, evalf(2^(2*k+1)/(2*k+3)!), 3], k, 1, 7)`

$$\begin{pmatrix} 1 & 1.33 \\ 2 & 0.267 \\ 3 & 0.0254 \\ 4 & 0.00141 \\ 5 & 5.13e-05 \\ 6 & 1.32e-06 \\ 7 & 2.51e-08 \end{pmatrix}$$

6. La terme générale de la séries $E_k(x)$ a un coefficient $\frac{1}{(2k+1)!}$ qui devient très petit très vite, ce qui fait qu'il est typiquement possible de calculer $\sin(x)$ avec relativement peu de termes.

Nous savons donc que les deux expressions

$$\sin(2), \quad \text{et} \quad 2 - 2^3/3! + 2^5/5! - 2^7/7! + 2^9/9!$$

diffèrent par moins de 10^{-3} .

```
evalf(sin(2))
```

```
0.909297426826
```

```
sin2:=S(4,2); evalf(sin2);
```

```
 $\frac{2578}{2835}, 0.909347442681$ 
```

Dans la pratique, on peut optimiser en utilisant les propriétés de la fonction sinus (périodicité, règles telles que $\sin(x) = \cos(\frac{\pi}{2} - x)$) et se ramener à un argument plus proche de 0.

Chapitre 3

Rappels d'algèbre linéaire.

Nous avons vu dans le chapitre précédent comment on peut donner un sens à une somme infinie de nombres – mais notre but initial ne concernait pas les nombres, mais les fonctions. Nous voudrions prendre une fonction $\phi(x)$, défini sur une intervalle $[0, L]$, et l'écrire comme une somme infini de fonctions trigonométriques, dans l'espoir que cela nous permettra de résoudre l'équation de la chaleur avec condition initiale ϕ .

Notons tout d'abord que la définition que nous avons donnée d'une somme infinie de nombres ne s'applique pas naturellement aux fonctions. En effet, la valeur d'une somme infinie s'exprime comme une limite d'une suite, et pouvoir parler de la limite d'une suite on a besoin d'une notion de distance - il faut pouvoir dire quand deux objets sont "proches". Or, si cette notion est facile pour des nombres réels ou complexes, c'est beaucoup plus délicat de dire quand deux fonctions sont "proches" ou définir une "distance" entre deux fonctions.

Mettons brièvement de côté cette difficulté. Si on nous donne une fonction ϕ sur une intervalle $[0, L]$, comment pourrait-on essayer d'écrire ϕ comme une somme infinie de fonctions trigonométriques ? Une première idée pourrait être de calculer cette somme par approximations successives : pour chaque entier k , on pourrait essayer de calculer $S_k(\phi)$, qui serait le "meilleur approximant" à ϕ de la forme

$$a_0 + a_1 \cos(\pi x/L) + a_2 \cos(2\pi x/L) + \dots + a_k \cos(k\pi x/L).$$

Peut être en prenant des valeurs de k de plus en plus grandes, on trouvera des $S_k(\phi)$, sommes trigonométriques finies, de plus en plus proches à ϕ ? Peut être lorsque k tend vers ∞ , les $S_k(\phi)$ convergeront vers une somme infinie de fonctions trigonométriques dont le résultat est ϕ ?

Cette idée d'écrire ϕ comme une somme infinie de fonctions trigonométriques par approximations successives est séduisante, mais pose plus de questions qu'elle ne résout :

1. Quel sens donner à une somme infinie de fonctions ?
2. Qu'est ce que ça veut dire, quand on dit que deux fonctions sont "proches" ?
3. Plus précisément, comment quantifier la "distance" entre deux fonctions ?
4. Comment calculer effectivement cette "meilleure approximation" $S_k(\phi)$?
5. Qu'est ce que cela signifie quand on dit qu'une suite de fonctions converge vers une autre fonction ?

Nous avons déjà commencer, dans le chapitre précédent, de répondre à la question 1), au moins dans le cas simple quie st celui d'une somme infinie de nombres. Nous chercherons maintenant à comprendre ce que peut vouloir une "bonne approximation" pour des fonctions. En effet, le premier problème que l'on rencontre lorsqu'on essaie de résoudre ces deux équations par une méthode d'approximations successives et celui de définir ce qu'on veut dire par une "bonne approximation", ou une "distance" entre deux fonctions.¹

1. Nous serons particulièrement attentifs à la question de l'approximation d'une fonction quelconque par des sommes de fonctions trigonométriques $\sin(n\pi x)$ et $\cos(n\pi x)$, puisque ces fonctions, qui représentent mathématiquement les phénomènes physiques *ondulatoires*, occupent une place très importante dans les mathématiques au service de la physique.

Nous allons en particulier regarder de près la question suivante :

Supposons donnée sur un intervalle $[0, L]$ une fonction f . Comment faire pour trouver la meilleure approximation pour f de la forme

$$S_n(f) = a_0 + \sum_{j=1}^n a_j \cos(j\pi x/L) + b_j \sin(j\pi x/L) ?$$

Si nous ne disposons pas actuellement d'une bonne notion de distance entre des fonctions² il existe bien des espaces pour lesquels la notion de distance existe. Ce sont les espaces géométriques \mathbb{R}^2 et \mathbb{R}^3 . De plus, dans ces espaces, il existe des algorithmes efficaces qui permettent, étant donné un point x et un plan ou droite S , de calculer le point de S le plus proche de x .

Nous allons nous baser sur ce que nous savons sur \mathbb{R}^2 et \mathbb{R}^3 pour définir des distances entre fonctions (et plein d'autres choses). Dans ce chapitre, nous allons étudier la notion d'espace vectoriel, qui réunit (entre autres) les espaces géométriques et les espaces de fonctions.

3.1 Espaces vectoriels : définitions et exemples.

Définition 3.1.1 Un \mathbb{R} -espace vectoriel est un ensemble V muni d'une loi interne

$$V \times V \rightarrow V, (x, y) \mapsto x + y,$$

et d'une loi externe

$$\mathbb{R} \times V \rightarrow V, (\lambda, x) \mapsto \lambda \cdot x,$$

appelée parfois multiplication par un scalaire, satisfaisant aux propriétés suivantes :

1. Il existe un élément $0_V \in V$ tel que $0_V + x = x + 0_V = x$ pour tout $x \in V$.
2. $x + (y + z) = (x + y) + z$ pour tout $x, y, z \in V$
3. $x + y = y + x$ pour tout $x, y \in V$
4. Pour tout $x \in V$, il existe un élément $x' \in V$ tel que $x + x' = x' + x = 0_V$.
Cet élément x' est alors unique, et est noté $-x$.
5. $1 \cdot x = x$ pour tout $x \in M$
6. $(\lambda\mu) \cdot x = \lambda \cdot (\mu \cdot x)$ pour tout $\lambda, \mu \in \mathbb{R}, x \in V$
7. $\lambda \cdot (x + y) = \lambda \cdot x + \lambda \cdot y$ pour tout $x, y \in V, \lambda \in \mathbb{R}$
8. $(\lambda + \mu) \cdot x = \lambda \cdot x + \mu \cdot x$ pour tout $x \in V, \lambda, \mu \in \mathbb{R}$.

Remarque 3.1.2 Grosso modo, un \mathbb{R} -espace vectoriel est un ensemble dans lequel :

- on peut additionner des éléments,
- on peut multiplier des éléments par des scalaires réels,
- cette addition et cette multiplication satisfont les mêmes règles algébriques que l'addition et multiplication habituelle dans \mathbb{R} .

Dans un \mathbb{C} -espace vectoriel il existe une multiplication par des scalaires complexes.

Remarque 3.1.3 On écrira λx pour $\lambda \cdot x$.

Exemples 3.1.4 1. \mathbb{R}^n , l'espace de vecteurs colonnes $\underline{X} = \begin{pmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{pmatrix}$ avec $x_i \in \mathbb{R}$, est un espace vectoriel réel. L'espace \mathbb{C}^n de vecteurs colonnes complexes est un espace vectoriel complexe.

2. Et encore moins d'un algorithme permettant de calculer ce "meilleur approximant" pour une fonction donnée...

2. $\mathbb{R}[X]$, l'espace de polynômes réels en une variable X , est un espace vectoriel réel. De même, $\mathbb{C}[Y]$, l'espace de polynômes complexes en une variable Y est un espace vectoriel complexe.
3. $\mathbb{R}_n[X]$, l'espace de polynômes réels en une variable X de degré $\leq n$, est un espace vectoriel réel. De même, $\mathbb{C}_n[Y]$, l'espace de polynômes complexes en une variable Y de degré $\leq n$, est un espace vectoriel complexe.
4. $M_n(\mathbb{R})$, l'espace de matrices $n \times n$ à coefficients réels, est un espace vectoriel réel,
5. Pour tout $a < b \in \mathbb{R}$ l'espace $C^0([a, b], \mathbb{R})$ de toutes les fonctions continues réelles sur l'intervalle $[a, b]$, est un espace vectoriel réel.
6. Pour tout $a < b \in \mathbb{R}$ et tout entier $i > 0$ l'espace $C^i([a, b], \mathbb{C})$ de toutes les fonctions continûment i -fois dérivables complexes sur l'intervalle $[a, b]$, est un espace vectoriel complexe.

Dans la pratique, nous travaillerons souvent avec des espaces vectoriels qui sont inclus dans d'autres.

Définition 3.1.5 Soit V un \mathbb{R} -espace vectoriel. Un **sous-espace vectoriel** W de V est un sous-ensemble de $W \subset V$ **non vide**, tel que

1. pour tout $w_1, w_2 \in W$ nous avons que $w_1 + w_2 \in W$
2. pour tout $w_1 \in W$ et $\lambda \in \mathbb{R}$ nous avons que $\lambda w_1 \in W$

L'ensemble W est alors toujours un espace vectoriel avec l'addition et la multiplication héritées de V .

Exercice 3.1.6 Montrer que les sous-ensembles suivants sont tous des sous-espaces vectoriels.

1. L'ensemble de tous les $(x, y) \in \mathbb{C}^2$ tels que $x + y = 0$.
2. Un plan d'équation $ax + by + cz = 0$ ($a, b, c \in \mathbb{R}$) dans \mathbb{R}^3 .
3. L'ensemble $\{P \in \mathbb{R}[X] \mid P(1) = 0\}$ des polynômes qui s'annulent en 1 dans $\mathbb{R}[X]$.
4. L'ensemble $\{M \in M_n(\mathbb{C}) \mid {}^t M = M\}$ des matrices symétriques dans $M_n(\mathbb{C})$.
5. L'ensemble de toutes les fonctions deux fois dérivables $f \in C^2(\mathbb{R}, \mathbb{R})$ telles que $f'' = -2f$ dans $C^2(\mathbb{R}, \mathbb{R})$.

Remarque 3.1.7 (Remarque utile.) Un sous-espace vectoriel W de V contient toujours 0_V , car si $x \in W$, alors $0 \cdot x = 0_V \in W$.

3.2 Bases et coordonnées.

Une notion clé dans ce qui suit sera la notion de base, qui permet de *représenter* un élément d'un espace vectoriel par un vecteur colonne.

Définition 3.2.1 Soit V un espace vectoriel réel. Une famille ordonnée d'éléments de V , $\mathbf{e} = (e_1, \dots, e_n)$ est une base (finie) pour V si pour tout élément $v \in V$ il existe des uniques scalaires $\lambda_1, \lambda_2, \dots, \lambda_n$ tels que

$$v = \lambda_1 e_1 + \lambda_2 e_2 + \dots + \lambda_n e_n.$$

Définition 3.2.2 Avec les notations de la définition 3.2.1, nous dirons que le vecteur colonne

$$\begin{pmatrix} \lambda_1 \\ \lambda_2 \\ \vdots \\ \lambda_n \end{pmatrix}$$

est le vecteur des coordonnées de v dans la base \mathbf{e} .

Remarque 3.2.3 (Attention !) Le vecteur de coordonnées de v dans une base \mathbf{e} dépend autant de la base \mathbf{e} que du vecteur v .

Remarque 3.2.4 (Notation) Dans ce qui suit il sera très important de distinguer l'élément v dans un espace vectoriel V de dimension finie n (qui peut être un vecteur colonne, ou une matrice, ou une fonction, ou un polynôme, ou plein d'autres choses) et le vecteur colonne $\underline{v} \in \mathbb{R}^n$ qui le représente dans une base donnée.

Pour bien distinguer ces deux objets, nous soulignerons systématiquement les noms des variables qui sont des vecteurs colonnes, et ne soulignerons pas ceux qui ne le sont pas.

Exemples 3.2.5

1. Les vecteurs

$$\begin{pmatrix} 1 \\ 0 \\ \vdots \\ 0 \end{pmatrix}, \dots, \begin{pmatrix} 0 \\ \vdots \\ 0 \\ 1 \end{pmatrix}$$

forment une base de \mathbb{R}^n , appelée la **base canonique**.

Si $\begin{pmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{pmatrix}$ est un élément de \mathbb{R}^n alors on peut écrire

$$\begin{pmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{pmatrix} = x_1 \begin{pmatrix} 1 \\ 0 \\ \vdots \\ 0 \end{pmatrix} + \dots + x_n \begin{pmatrix} 0 \\ \vdots \\ 0 \\ 1 \end{pmatrix};$$

autrement dit, le vecteur de coordonnées de $\begin{pmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{pmatrix}$ dans la base canonique est $\begin{pmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{pmatrix}$. Ceci

est une source importante de confusion.

2. Montrons que $B = \left(\begin{pmatrix} 1 \\ 1 \end{pmatrix}, \begin{pmatrix} 1 \\ 2 \end{pmatrix} \right)$ est une base de \mathbb{C}^2 . Nous considérons pour un vecteur arbitraire $\begin{pmatrix} x \\ y \end{pmatrix}$ l'équation

$$\begin{pmatrix} x \\ y \end{pmatrix} = \lambda_1 \begin{pmatrix} 1 \\ 1 \end{pmatrix} + \lambda_2 \begin{pmatrix} 1 \\ 2 \end{pmatrix}$$

c'est-à-dire

$$x = \lambda_1 + \lambda_2$$

$$y = \lambda_1 + 2\lambda_2$$

ce qui (après pivot de Gauss) nous donne l'unique solution

$$\lambda_1 = 2x - y,$$

$$\lambda_2 = y - x.$$

Cette famille est donc une base et le vecteur de coordonnées de $\begin{pmatrix} x \\ y \end{pmatrix}$ dans la base B est

$$\begin{pmatrix} 2x - y \\ y - x \end{pmatrix}.$$

3. La famille $B = (1, X, \dots, X^n)$ forme une base de l'espace vectoriel $\mathbb{R}[X]_n$ des polynômes à coefficients dans \mathbb{R} de degré au plus n . Si $P = a_0 + a_1X + \dots + a_nX^n$ est un élément de $\mathbb{R}_n[X]$ alors son vecteur de coefficients dans la base B est

$$\begin{pmatrix} a_0 \\ a_1 \\ \vdots \\ a_n \end{pmatrix}.$$

4. On considère $M_2(\mathbb{C})$, l'espace de matrices carrées complexes 2×2 . Elle a une base

$$B = \left(\begin{pmatrix} 1 & 0 \\ 0 & 0 \end{pmatrix}, \begin{pmatrix} 0 & 1 \\ 0 & 0 \end{pmatrix}, \begin{pmatrix} 0 & 0 \\ 1 & 0 \end{pmatrix}, \begin{pmatrix} 0 & 0 \\ 0 & 1 \end{pmatrix} \right),$$

et dans cette base la matrice $M = \begin{pmatrix} a & b \\ c & d \end{pmatrix}$ a pour vecteur de coefficients $\begin{pmatrix} a \\ b \\ c \\ d \end{pmatrix}$.

5. On considère l'espace de fonctions réelles deux fois dérivables sur \mathbb{R} qui satisfont l'équation $f'' = -2f$. Vous avez vu dans MAT128 que cette espace est de dimension 2 et la famille

$$(\cos(\sqrt{2}x), \sin(\sqrt{2}x))$$

en est une base. Le vecteur de coordonnées de la fonction $f = a \cos(\sqrt{2}x) + b \sin(\sqrt{2}x)$ dans cette base est $\begin{pmatrix} a \\ b \end{pmatrix}$.

Définition 3.2.6 Lorsqu'un espace vectoriel V possède une base finie on dit que V est de dimension finie. Toutes les bases de V ont alors le même nombre d'éléments (nous admettrons ce théorème) : ce nombre s'appelle la **dimension** de V .

Exemples 3.2.7 1. L'espace \mathbb{R}^n est de dimension n .

2. L'espace $\mathbb{R}_n[X]$ est de dimension $n + 1$.

3. L'espace $M_2(\mathbb{R})$ est de dimension 4.

Remarque 3.2.8 Tout sous-espace d'un espace de dimension finie est de dimension finie.

Le résultat suivant, que nous rappelons sans démonstration, sera souvent utilisé pour vérifier qu'une famille de vecteurs est une base.

Lemme 3.2.9 Soit V un espace vectoriel de dimension n et soit (e_1, \dots, e_n) une famille de n vecteurs dans V . Si la famille (e_1, \dots, e_n) est libre (c'est à dire que $\sum \lambda_i e_i = 0_V \Rightarrow \lambda_i = 0 \forall i$) alors elle est une base.

Les coordonnées d'un élément $v \in V$ dans une base seront essentielles dans la suite, car elles nous permettront de ramener tous nos calculs à de simples multiplications de matrices. Il nous sera, d'ailleurs, souvent utile de simplifier nos calculs au maximum en choisissant une base bien adaptée. Pour faire cela, il nous faut comprendre comment le vecteur \underline{v} des coordonnées d'un élément $v \in V$ dans une base \mathbf{e} se transforme lorsqu'on change de base.

Définition 3.2.10 Soit V un espace vectoriel de dimension n et soient $\mathbf{e} = (e_1, \dots, e_n)$ et $\mathbf{f} = (f_1, \dots, f_n)$ des bases de V . Soit \underline{v}_i le vecteur de coordonnées de f_i dans la base (e_1, \dots, e_n) . Alors, la matrice de passage de \mathbf{e} vers \mathbf{f} est la matrice

$$P = (\underline{v}_1, \dots, \underline{v}_n).$$

Remarque 3.2.11 (Attention) Il y a un cas spécial quand \mathbf{e} est la base canonique de \mathbb{R}^n : nous avons alors que la matrice de passage P est donnée par

$$P = (\underline{f}_1, \dots, \underline{f}_n).$$

Encore une fois, ce cas spécial donne lieu à beaucoup de confusion.

L'importance de la matrice de passage vient du théorème fondamental suivant, que nous rappelons :

Théorème 3.2.12 Soient \mathbf{e}_1 et \mathbf{e}_2 des bases de V et soit v un élément de V . Soient \underline{V}_1 et \underline{V}_2 les vecteurs de coordonnées de v dans les bases \mathbf{e}_1 et \mathbf{e}_2 . Soit P la matrice de passage de B_1 vers B_2 . Alors

$$\underline{V}_1 = P\underline{V}_2$$

ou, de façon équivalente

$$\underline{V}_2 = P^{-1}\underline{V}_1$$

Remarque 3.2.13 Attention le terminologie de “matrice de passage” est assez trompeuse ! C’est P^{-1} , et pas P , qui convertit le vecteur qui représente v dans la base \mathbf{e} dans le vecteur qui représente v dans la base \mathbf{f} .

Il y a une généralisation de la notion de base qui sera utile dans la démonstration d’un théorème ultérieur.

Définition 3.2.14 Soient V_1, \dots, V_m des sous-espaces vectoriels de V . On dit que V est la **somme directe** des sous-espaces V_1, \dots, V_m , et on écrit $V = V_1 \oplus V_2 \oplus \dots \oplus V_m$, si et seulement si pour tout $v \in V$ il existe des uniques éléments $v_1 \in V_1, \dots, v_m \in V_m$ tels que

$$v = v_1 + \dots + v_m.$$

Le résultat suivant, que nous admettrons, sera aussi utile dans un résultat ultérieur :

Proposition 3.2.15 Si $V = V_1 \oplus V_2 \oplus \dots \oplus V_m$ et pour chaque i nous avons que \mathbf{e}_i est une base de V_i alors la concaténation $(\mathbf{e}_1, \mathbf{e}_2, \dots, \mathbf{e}_m)$ est une base de V .

3.3 Applications linéaires.

Considérons maintenant la classe des applications qui préservent la structure d’un espace vectoriel.

Définition 3.3.1 Soient V et V' deux \mathbb{R} -espaces vectoriels.

Une **application linéaire** de V dans V' est une application $f : V \rightarrow V'$ vérifiant

1. $f(v_1 + v_2) = f(v_1) + f(v_2)$ pour tous $v_1, v_2 \in V$
2. $f(\lambda v) = \lambda f(v)$ pour tous $\lambda \in \mathbb{R}, v \in V$

Dans le cas où l’espace d’arrivée est \mathbb{R} on dira que f est une **forme linéaire**.

Remarque 3.3.2 Pour toute application linéaire f on a nécessairement $f(0) = 0$.

Exemples 3.3.3 1. L’application $\mathbb{R}^3 \rightarrow \mathbb{R}^2$ donnée par $\begin{pmatrix} x \\ y \\ z \end{pmatrix} \mapsto \begin{pmatrix} x \\ y \end{pmatrix}$ est linéaire.

2. L’application $\mathbb{C}^3 \rightarrow \mathbb{C}^2$ donnée par $\begin{pmatrix} x \\ y \\ z \end{pmatrix} \mapsto \begin{pmatrix} x \\ y + 1 \end{pmatrix}$ n’est pas linéaire.

3. L’application $C^1(\mathbb{R}, \mathbb{R}) \mapsto C^0(\mathbb{R}, \mathbb{R})$, $f \mapsto f' - 2f$ est linéaire.

4. L’application $M_n(\mathbb{C}) \mapsto M_n(\mathbb{C})$ donnée par $M \mapsto {}^t M$ est linéaire.

5. L’application $\mathbb{R}_3[X] \mapsto \mathbb{R}_1[X]$, $P \mapsto P''$, est une application linéaire.

Exercice 3.3.4 Démontrer que les applications 1, 3, 4, 5 de 3.3.3 sont linéaires et que 2 ne l'est pas.

Définition 3.3.5 Le **noyau** de f , noté $\text{Ker}(f)$, est l'ensemble

$$\text{Ker}(f) = \{v \in V \mid f(v) = 0\} (\subseteq V).$$

C'est un sous-espace vectoriel de V .

Définition 3.3.6 L'**image** de f , notée $\text{Im}(f)$, est l'ensemble

$$\text{Im}(f) = \{f(v), v \in V\} \subseteq V'.$$

C'est un sous-espace vectoriel de V' .

Exemples 3.3.7 1. Montrer que le noyau et l'image d'une application linéaire sont des sous-espaces vectoriels.

2. Pour les applications linéaires données dans 3.3.3, calculer leur image et noyau.

On rappelle le théorème du rang, dont nous aurons besoin dans une démonstration ultérieure.

Théorème 3.3.8 Soit $f : V \rightarrow W$ une application linéaire. On suppose que V est de dimension finie. Alors on a que $\text{Im}(f)$ est de dimension infinie et

$$\dim(V) = \dim(\text{Ker}(f)) + \dim(\text{Im}(f)).$$

3.4 Calcul Matriciel.

Dans cette section nous ferons des rappels sur les matrices et leurs manipulations. Celles-ci seront un élément clé de notre travail ce semestre.

Définition 3.4.1 Etant donnés deux entiers m et n strictement positifs, une matrice à m lignes et n colonnes est un tableau rectangulaire de réels $A = (a_{i,j})$. L'indice de ligne i va de 1 à m , l'indice de colonne j va de 1 à n .

$$A = (a_{i,j}) = \begin{pmatrix} a_{1,1} & \cdots & a_{1,j} & \cdots & a_{1,n} \\ \vdots & & \vdots & & \vdots \\ a_{i,1} & \cdots & a_{i,j} & \cdots & a_{i,n} \\ \vdots & & \vdots & & \vdots \\ a_{m,1} & \cdots & a_{m,j} & \cdots & a_{m,n} \end{pmatrix}.$$

Les entiers m et n sont les dimensions de la matrice, $a_{i,j}$ est son coefficient d'ordre (i, j) .

Notons qu'une matrice A peut être précisée en donnant une expression pour ses coefficients $a_{i,j}$. Par exemple, la matrice A de taille 2×2 donnée par le formule $a_{i,j} = i + j$ est la matrice

$$A = \begin{pmatrix} 1+1 & 1+2 \\ 2+1 & 2+2 \end{pmatrix} = \begin{pmatrix} 2 & 3 \\ 3 & 4 \end{pmatrix}.$$

L'ensemble des matrices à m lignes et n colonnes et à coefficients réels est noté $\mathcal{M}_{m,n}(\mathbb{R})$. Ce qui suit s'applique aussi, si on remplace \mathbb{R} par \mathbb{C} , à l'ensemble des matrices à coefficients complexes.

Notons trois cas spéciaux :

1. Un vecteur colonne à n éléments $\begin{pmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{pmatrix}$ est une matrice $n \times 1$.

2. Un vecteur ligne à n éléments (x_1, x_2, \dots, x_n) est une matrice $1 \times n$.
3. Un nombre réel x est une matrice 1×1 .

Du point de vue du calcul matriciel - en particulier lorsqu'il s'agit de faire des multiplications - un vecteur ligne ne se comporte pas comme un vecteur colonne. Nous ferons cette distinction en considérant, par exemple, que les vecteurs

$$(1 \quad 2 \quad 3) \text{ et } \begin{pmatrix} 1 \\ 2 \\ 3 \end{pmatrix}$$

sont différents, même s'ils contiennent les mêmes nombres dans le même ordre.

Notation. Si \underline{X} est un vecteur colonne à n éléments, on notera le coefficient $\underline{X}_{1,i}$ par \underline{X}_i .

L'ensemble $\mathcal{M}_{m,n}(\mathbb{R})$ est naturellement muni d'une addition (on peut ajouter deux matrices de mêmes dimensions terme à terme) et de multiplication par des scalaires (on peut multiplier une matrice par un réel terme à terme).

- *Addition* : Si $A = (a_{i,j})$ et $B = (b_{i,j})$ sont deux matrices de $\mathcal{M}_{m,n}(\mathbb{R})$, leur somme $A + B$ est la matrice $(a_{i,j} + b_{i,j})$. Par exemple :

$$\begin{pmatrix} 1 & 1 \\ 2 & 3 \\ 1 & -1 \end{pmatrix} + \begin{pmatrix} -3 & 1 \\ 5 & -3 \\ 0 & 2 \end{pmatrix} = \begin{pmatrix} -2 & 2 \\ 7 & 0 \\ 1 & 1 \end{pmatrix}$$

- *Multiplication par un scalaire* : Si $A = (a_{i,j})$ est une matrice de $\mathcal{M}_{m,n}(\mathbb{R})$, et λ est un réel, le produit λA est la matrice $(\lambda a_{i,j})$. Par exemple :

$$-2 \begin{pmatrix} 1 & 1 \\ 2 & 3 \\ 1 & -1 \end{pmatrix} = \begin{pmatrix} -2 & -2 \\ -4 & -6 \\ -2 & 2 \end{pmatrix}$$

Observons que ces opérations auraient le même effet si les matrices étaient disposées comme des mn -uplets de réels (toutes les lignes étant concaténées, par exemple)

Les matrices de taille $m \times n$ peuvent agir sur des vecteurs colonnes de taille n pour produire un vecteur de taille m , par la formule suivante :

$$(M\underline{X})_j = \sum_{i=1}^n M_{j,i} \underline{X}_i.$$

L'opération la plus importante est le *produit matriciel*.

Définition 3.4.2 Soient m, n, p trois entiers strictement positifs. Soit $A = (a_{i,j})$ une matrice de $\mathcal{M}_{m,n}(\mathbb{R})$ et soit $B = (b_{j,k})$ une matrice de $\mathcal{M}_{n,p}(\mathbb{R})$. On appelle produit matriciel de A par B la matrice $C \in \mathcal{M}_{m,p}(\mathbb{R})$ dont le terme général $c_{i,k}$ est défini, pour tout $i = 1, \dots, m$ et pour tout $k \in 1, \dots, p$ par :

$$c_{i,k} = \sum_{j=1}^n a_{i,j} b_{j,k}.$$

Nous insistons sur le fait que le produit AB de deux matrices n'est défini que si le nombre de colonnes de A et le nombre de lignes de B sont les mêmes. Dans le cas particulier où B est un vecteur colonne de taille

$n \times 1$ cette opération nous fournit un vecteur colonne de taille $m \times 1$.

$$\begin{pmatrix} a_{1,1} & \cdots & & \cdots & a_{1,n} \\ \vdots & & \vdots & & \vdots \\ a_{i,1} & \cdots & a_{i,j} & \cdots & a_{i,n} \\ \vdots & & \vdots & & \vdots \\ a_{m,1} & \cdots & & \cdots & a_{m,n} \end{pmatrix} \begin{pmatrix} b_{1,1} & \cdots & b_{1,k} & \cdots & b_{1,n} \\ \vdots & & \vdots & & \vdots \\ & \cdots & b_{j,k} & \cdots & \\ \vdots & & \vdots & & \vdots \\ b_{n,1} & \cdots & b_{n,k} & \cdots & b_{n,p} \\ c_{1,1} & & \vdots & & c_{1,p} \\ & & \vdots & & \\ \cdots & \cdots & c_{i,k} & & \\ c_{m,1} & & & & c_{m,p} \end{pmatrix}$$

Posons par exemple :

$$A = \begin{pmatrix} 1 & 1 \\ 2 & 3 \\ 1 & -1 \end{pmatrix} \quad \text{et} \quad B = \begin{pmatrix} 0 & 1 & -1 & -2 \\ -3 & -2 & 0 & 1 \end{pmatrix}.$$

La matrice A a 3 lignes et 2 colonnes, la matrice B a 2 lignes et 4 colonnes. Le produit AB a donc un sens : c'est une matrice à 3 lignes et 4 colonnes.

$$\begin{pmatrix} 1 & 1 \\ 2 & 3 \\ 1 & -1 \end{pmatrix} \begin{pmatrix} 0 & 1 & -1 & -2 \\ -3 & -2 & 0 & 1 \end{pmatrix} = \begin{pmatrix} -3 & -1 & -1 & -1 \\ -9 & -4 & -2 & -1 \\ 3 & 3 & -1 & -3 \end{pmatrix}$$

Le produit matriciel a toutes les propriétés que l'on attend d'un produit, sauf qu'il n'est pas commutatif.

Proposition 3.4.3 *Le produit matriciel possède les propriétés suivantes.*

1. Associativité : Si les produits AB et BC sont définis, alors les produits $A(BC)$ et $(AB)C$ le sont aussi et ils sont égaux.

$$A(BC) = (AB)C .$$

2. Linéarité à droite : Si B et C sont deux matrices de mêmes dimensions, si λ et μ sont deux réels et si A a autant de colonnes que B et C ont de lignes, alors

$$A(\lambda B + \mu C) = \lambda AB + \mu AC .$$

3. Linéarité à gauche : Si A et B sont deux matrices de mêmes dimensions, si λ et μ sont deux réels et si C a autant de lignes que A et B ont de colonnes, alors

$$(\lambda A + \mu B)C = \lambda AC + \mu BC .$$

Ces propriétés se démontrent à partir de la définition 3.4.2.

La transposition est une notion importante, dont la justification provient de la dualité, qui dépasse le cadre de ce cours.

Définition 3.4.4 *Étant donnée une matrice $A = (a_{i,j})$ de $\mathcal{M}_{m,n}(\mathbb{R})$, sa transposée est la matrice de $\mathcal{M}_{n,m}(\mathbb{R})$ dont le coefficient d'ordre (j, i) est $a_{i,j}$.*

Pour écrire la transposée d'une matrice, il suffit de transformer ses lignes en colonnes. Par exemple :

$$A = \begin{pmatrix} 1 & 1 \\ 2 & 3 \\ 1 & -1 \end{pmatrix}, \quad {}^tA = \begin{pmatrix} 1 & 2 & 1 \\ 1 & 3 & -1 \end{pmatrix}.$$

Observons que la transposée de la transposée est la matrice initiale.

$${}^t({}^tA) = A.$$

La transposée d'un produit est le produit des transposées, mais il faut inverser l'ordre des facteurs.

Proposition 3.4.5 Soient m, n, p trois entiers strictement positifs. Soient $A = (a_{i,j})$ une matrice de $\mathcal{M}_{m,n}(\mathbb{R})$ et $B = (b_{j,k})$ une matrice de $\mathcal{M}_{n,p}(\mathbb{R})$. La transposée du produit de A par B est le produit de la transposée de B par la transposée de A .

$${}^t(AB) = {}^tB {}^tA.$$

Par exemple, en reprenant les matrices A et B définies ci-dessus :

$$\begin{pmatrix} 0 & -3 \\ 1 & -2 \\ -1 & 0 \\ -2 & 1 \end{pmatrix} \begin{pmatrix} 1 & 2 & 1 \\ 1 & 3 & -1 \end{pmatrix} = \begin{pmatrix} -3 & -9 & 3 \\ -1 & -4 & 3 \\ -1 & -2 & -1 \\ -1 & -1 & -3 \end{pmatrix}$$

Définition 3.4.6 Soit n un entier strictement positif et A une matrice carrée à n lignes et n colonnes. On dit que A est symétrique si pour tous $i, j = 1, \dots, n$, ses coefficients d'ordre $a_{i,j}$ et $a_{j,i}$ sont égaux, ce qui est équivalent à dire que A est égale à sa transposée.

Le produit d'une matrice par sa transposée est toujours une matrice symétrique. En effet :

$${}^t(A {}^tA) = {}^t({}^tA) {}^tA = A {}^tA.$$

3.5 Matrices carrées

En général si le produit AB est défini, le produit BA n'a aucune raison de l'être. Le produit d'une matrice par sa transposée est une exception, les matrices carrées en sont une autre : si A et B sont deux matrices à n lignes et n colonnes, les produits AB et BA sont tous deux définis et ils ont les mêmes dimensions que A et B . En général ils ne sont pas égaux. Par exemple,

$$\begin{pmatrix} 0 & -1 \\ 1 & 0 \end{pmatrix} \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix} = \begin{pmatrix} 1 & 0 \\ 0 & -1 \end{pmatrix} \quad \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix} \begin{pmatrix} 0 & -1 \\ 1 & 0 \end{pmatrix} = \begin{pmatrix} -1 & 0 \\ 0 & 1 \end{pmatrix}$$

Nous noterons simplement $\mathcal{M}_n(\mathbb{R})$ l'ensemble $\mathcal{M}_{n,n}(\mathbb{R})$ des matrices carrées à n lignes et n colonnes, à coefficients réels. Parmi elles la *matrice identité*, notée I_n , joue un rôle particulier.

$$I_n = \begin{pmatrix} 1 & 0 & \cdots & \cdots & 0 \\ 0 & 1 & \ddots & & \vdots \\ \vdots & \ddots & \ddots & \ddots & \vdots \\ \vdots & & \ddots & 1 & 0 \\ 0 & \cdots & \cdots & 0 & 1 \end{pmatrix}$$

En effet, elle est l'élément neutre du produit matriciel : pour toute matrice $A \in \mathcal{M}_{n,m}(\mathbb{R})$,

$$A I_n = I_m A = A.$$

On le vérifie facilement à partir de la définition 3.4.2.

Définition 3.5.1 Soit A une matrice de \mathcal{M}_n . On dit que A est inversible s'il existe une matrice de \mathcal{M}_n , notée A^{-1} , telle que

$$A A^{-1} = A^{-1} A = I_n .$$

Par exemple :

$$\begin{pmatrix} 1 & 0 & -1 \\ 1 & -1 & 0 \\ 1 & -1 & 1 \end{pmatrix} \begin{pmatrix} 1 & -1 & 1 \\ 1 & -2 & 1 \\ 0 & -1 & 1 \end{pmatrix} = \begin{pmatrix} 1 & -1 & 1 \\ 1 & -2 & 1 \\ 0 & -1 & 1 \end{pmatrix} \begin{pmatrix} 1 & 0 & -1 \\ 1 & -1 & 0 \\ 1 & -1 & 1 \end{pmatrix} = \begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{pmatrix}$$

Observons que l'inverse, s'il existe, est nécessairement unique. En effet, soient B_1 et B_2 deux matrices telles que $A B_1 = B_1 A = I_n$ et $A B_2 = B_2 A = I_n$. En utilisant l'associativité, le produit $B_1 A B_2$ vaut $B_1 (A B_2) = B_1 I_n = B_1$, mais aussi $(B_1 A) B_2 = I_n B_2 = B_2$. Donc $B_1 = B_2$.

Nous rappelons la proposition suivante, qui nous dit qu'il suffit de trouver une matrice B telle que $A B = I_n$ pour être sûr que A est inversible et que son inverse est B .

Proposition 3.5.2 Soit A une matrice de \mathcal{M}_n . Supposons qu'il existe une matrice B telle que $A B = I_n$ ou bien $B A = I_n$. Alors A est inversible et $B = A^{-1}$.

Si A et B sont deux matrices inversibles de \mathcal{M}_n , leur produit est inversible.

Proposition 3.5.3 Soient A et B deux matrices inversibles de $\mathcal{M}_n(\mathbb{R})$. Le produit AB est inversible et son inverse est $B^{-1}A^{-1}$.

Nous utilisons le théorème 3.5.2, ainsi que l'associativité du produit :

$$(B^{-1}A^{-1})(AB) = B^{-1}(A^{-1}A)B = B^{-1}I_n B = B^{-1}B = I_n .$$

Enfin, nous aurons parfois besoin du lemme suivant :

Lemme 3.5.4 Soit $M \in M_n(\mathbb{R})$ une matrice carrée $n \times n$. Si pour tout $\underline{X}, \underline{Y} \in \mathbb{R}^n$ nous avons que ${}^t \underline{X} M \underline{Y} = 0$ alors $M = 0$.

Soit pour tout i le vecteur colonne $\underline{e}_i \in \mathbb{R}^n$ défini par

$$(\underline{e}_i)_j = 1 \text{ si } i = j, 0 \text{ si } i \neq j.$$

Alors pour tout $1 \leq i, j \leq n$ on a que

$${}^t \underline{e}_i M \underline{e}_j = M_{i,j} = 0$$

et donc $M = 0$. Récrivons maintenant notre problème initial dans le langage des espaces vectoriels. Nous considérons une fonction réelle continue f , définie sur un intervalle $[0, L]$. Autrement dit, f est un élément de l'espace vectoriel réel $C^0([0, L], \mathbb{R}) = V$. Nous voulons chercher une fonction g_n qui est de la forme

$$g_n(x) = a_0 + \sum_{k=1}^n a_k \cos\left(\frac{k\pi x}{L}\right) + b_k \sin\left(\frac{k\pi x}{L}\right)$$

et qui doit être "aussi proche que possible" de f .

Dans le langage des espaces vectoriels on pourrait écrire la chose suivante :

Soit W le sous-espace de tous les éléments $g \in V$ qui peuvent s'écrire sous la forme

$$g_n(x) = a_0 + \sum_{k=1}^n a_k \cos\left(\frac{k\pi x}{L}\right) + b_k \sin\left(\frac{k\pi x}{L}\right).$$

W est alors un sous-espace vectoriel de V (**exercice** : démontrez-le !) : de plus, W est de dimension finie et admet pour base finie la famille

$$\mathbf{e} = \left(1, \cos \pi x/L, \sin \pi x/L, \dots, \cos \frac{n\pi x}{L}, \sin \frac{n\pi x}{L}\right).$$

Nous cherchons à identifier un élément $g \in W$ qui est “le plus proche que possible” de $f \in V$.

Notre problème initial est donc un exemple particulier du problème suivant :

Question. J'ai un espace vectoriel V et un élément $v \in V$. Il y a dans V un sous-espace spécial de dimension finie $W \subset V$. Je veux approcher au mieux v par un élément $w \in W$. Comment faire ? Et tout d'abord, qu'est ce que ça veut dire “approcher au mieux” ?

Dans les deux prochains chapitres, nous aborderons surtout la question : qu'est ce que ça veut dire “approcher au mieux” ?

Chapitre 4

Formes bilinéaires.

4.1 Le produit scalaire canonique sur \mathbb{R}^3 .

Dans le chapitre précédent, nous avons étudié la notion d'espace vectoriel. Cette notion est utile parce qu'elle englobe à la fois des espaces géométriques tels que \mathbb{R}^2 et \mathbb{R}^3 et des espaces de fonctions tels que $\mathbb{R}_n[X]$ et $C^0([0, 1], \mathbb{R})$. Notre but est maintenant d'utiliser cette notion pour étendre des idées géométriques (distance et angle, par exemple) à des espaces de fonctions. Pour faire cela, il nous sera nécessaire d'identifier une formule purement algébrique qui permet de calculer distances et angles dans \mathbb{R}^3 .

Ceci sera possible utilisant le produit scalaire canonique sur \mathbb{R}^3 .

Définition 4.1.1 *Le produit scalaire canonique sur \mathbb{R}^3 est la fonction de deux vecteurs $\underline{X} = \begin{pmatrix} x_1 \\ x_2 \\ x_3 \end{pmatrix}$ et*

$$\underline{Y} = \begin{pmatrix} y_1 \\ y_2 \\ y_3 \end{pmatrix}$$

$$\underline{X}, \underline{Y} \rightarrow \langle \underline{X}, \underline{Y} \rangle$$

donnée par $\langle \underline{X}, \underline{Y} \rangle = x_1y_1 + x_2y_2 + x_3y_3$.

Le produit scalaire canonique est donc une fonction de deux vecteurs donnée par une formule simple sur les coordonnées des vecteurs. Il tire son intérêt du fait qu'il encode la géométrie de l'espace \mathbb{R}^3 .

Théorème 4.1.2 *Soient \underline{X} et \underline{Y} deux vecteurs dans \mathbb{R}^3 , soit θ l'angle entre ces deux vecteurs et soit d la distance entre elles : nous avons alors que*

$$d = (\langle \underline{X} - \underline{Y}, \underline{X} - \underline{Y} \rangle)^{1/2}, \quad \theta = \arccos \left(\frac{\langle \underline{X}, \underline{Y} \rangle}{(\langle \underline{X}, \underline{X} \rangle, \langle \underline{Y}, \underline{Y} \rangle)^{1/2}} \right).$$

Il existe donc une formule qui permet de calculer la distance et l'angle entre deux vecteurs utilisant seulement le produit scalaire. Nous allons donc essayer de définir des classes de fonctions sur des espaces vectoriels qui ressemblent au produit scalaire sur \mathbb{R}^3 dans l'espoir qu'elles nous livrent une bonne notion de "distance".

Une des propriétés clés du produit scalaire est qu'il se comporte effectivement comme un produit sous les opérations algébriques de base sur les vecteurs, c'est-à-dire qu'on a, pour tout $\underline{X}, \underline{Y}, \underline{Z} \in \mathbb{R}^3$ et pour tout $\lambda \in \mathbb{R}$

1. $\langle \underline{X} + \underline{Y}, \underline{Z} \rangle = \langle \underline{X}, \underline{Z} \rangle + \langle \underline{Y}, \underline{Z} \rangle$
2. $\langle \underline{X}, \underline{Y} + \underline{Z} \rangle = \langle \underline{X}, \underline{Y} \rangle + \langle \underline{X}, \underline{Z} \rangle$
3. $\langle \underline{X}, \lambda \underline{Y} \rangle = \langle \lambda \underline{X}, \underline{Y} \rangle = \lambda \langle \underline{X}, \underline{Y} \rangle$

Nous allons donc commencer par étudier les fonctions de deux vecteurs qui respectent ces conditions.

4.2 Formes bilinéaires : définitions et exemples.

Dans cette section, de nouveau, nous présenterons la théorie des formes bilinéaires réelles, mais tous nos résultats seront valables pour des formes complexes.

Définition 4.2.1 Soient V et V' deux \mathbb{R} -espaces vectoriels, et soit $\varphi : V \times V' \rightarrow \mathbb{R}$ une application de $V \times V'$ dans \mathbb{R} .

On dit que $\varphi : V \times V' \rightarrow \mathbb{R}$ est **une forme bilinéaire** si :

1. pour tout $v_1, v_2 \in V$ et $v' \in V'$ nous avons que $\varphi(v_1 + v_2, v') = \varphi(v_1, v') + \varphi(v_2, v')$
2. pour tout $v \in V$ et $v'_1, v'_2 \in V'$ nous avons que $\varphi(v, v'_1 + v'_2) = \varphi(v, v'_1) + \varphi(v, v'_2)$
3. pour tout $v \in V$, $v' \in V'$ et $\lambda \in \mathbb{R}$ nous avons que $\varphi(\lambda v, v') = \varphi(v, \lambda v') = \lambda \varphi(v, v')$.

Dans le cas où $V = V'$, on dit que $\varphi : V \times V \rightarrow \mathbb{R}$ est **symétrique** si $\varphi(y, x) = \varphi(x, y)$ pour tout $x, y \in V$, et on dit que $\varphi : V \times V \rightarrow \mathbb{R}$ est **antisymétrique** si $\varphi(y, x) = -\varphi(x, y)$ pour tout $x, y \in V$.

Exemples 4.2.2

(1) L'application

$$\varphi : \mathbb{R} \times \mathbb{R} \rightarrow \mathbb{R}, (x, y) \mapsto xy$$

est une forme bilinéaire symétrique.

(2) Le produit scalaire

$$\varphi : \mathbb{R}^n \times \mathbb{R}^n \rightarrow \mathbb{R}, \left(\begin{pmatrix} x_1 \\ \vdots \\ x_n \end{pmatrix}, \begin{pmatrix} y_1 \\ \vdots \\ y_n \end{pmatrix} \right) \mapsto x \cdot y = \sum_{i=1}^n x_i y_i$$

est une forme bilinéaire symétrique. Lorsque $n = 2$ ou 3 , on retrouve le produit scalaire étudié ci-dessus. Nous appelons cette forme le produit scalaire canonique sur \mathbb{R}^n .

(3) L'application

$$\varphi : \mathbb{C}[X] \times \mathbb{C}[X] \rightarrow \mathbb{C}, (P, Q) \mapsto P(0)Q(1)$$

est une forme bilinéaire ni symétrique ni antisymétrique.

(4) L'application

$$\varphi : M_n(\mathbb{R}) \times M_n(\mathbb{R}) \rightarrow \mathbb{R}, (M, N) \mapsto \text{tr}(MN)$$

est une forme bilinéaire symétrique.

(5) L'application

$$\varphi : \mathbb{C}^2 \times \mathbb{C}^2 \rightarrow \mathbb{C}, \left(\begin{pmatrix} x_1 \\ x_2 \end{pmatrix}, \begin{pmatrix} y_1 \\ y_2 \end{pmatrix} \right) \mapsto x_1 x_2 + 2x_1 y_2$$

n'est pas bilinéaire.

En effet, posons $\underline{U} = \begin{pmatrix} x_1 \\ x_2 \end{pmatrix}$, $\underline{V} = \begin{pmatrix} y_1 \\ y_2 \end{pmatrix}$. On a

$$\varphi(\lambda \underline{U}, \underline{V}) = (\lambda x_1)(\lambda x_2) + 2(\lambda x_1)y_2 = \lambda^2 x_1 x_2 + 2\lambda x_1 y_2 \neq \lambda \varphi(\underline{U}, \underline{V}).$$

(6) L'application

$$\varphi : C^0([0, 1], \mathbb{R}) \times C^0([0, 1], \mathbb{R}) \rightarrow \mathbb{R}, f, g \mapsto \int_0^1 f(x)g(x)dx$$

est une forme bilinéaire symétrique.

(7) Pour toute fonction continue $p : [0, 1] \rightarrow \mathbb{R}$, l'application

$$\varphi : C^0([0, 1], \mathbb{R}) \times C^0([0, 1], \mathbb{R}) \rightarrow \mathbb{R}, f, g \mapsto \int_0^1 p(x)f(x)g(x)dx$$

est une forme bilinéaire symétrique.

Un cas particulier intéressant est celui où on applique une forme bilinéaire à deux vecteurs identiques.

Définition 4.2.3 Soit V un espace vectoriel sur \mathbb{R} et soit $\varphi : V \times V \rightarrow \mathbb{R}$ une forme bilinéaire symétrique. Alors la forme quadratique associée à φ , notée q_φ , est la fonction

$$q_\varphi : V \rightarrow \mathbb{R}, \quad q_\varphi(v) = \varphi(v, v).$$

La forme quadratique associée à une forme bilinéaire est un analogue de la fonction carrée d'un nombre réel, ou de $|v|^2$ quand v est un vecteur dans \mathbb{R}^2 ou \mathbb{R}^3 . Les formules suivantes (dit "formule de polarisation" et "formule du parallélogramme") permettent entre autres de récupérer une forme bilinéaire symétrique à partir de sa forme quadratique.

Lemme 4.2.4 Soit V un espace vectoriel, φ une forme bilinéaire sur $V \times V$ et q_φ la forme quadratique associée. Alors pour tout $v, w \in V$ on a que

$$\begin{aligned} \varphi(v, w) &= \frac{1}{2}(q_\varphi(v + w) - q_\varphi(v) - q_\varphi(w)) \\ q_\varphi(v + w) + q_\varphi(v - w) &= 2(q_\varphi(v) + q_\varphi(w)). \end{aligned}$$

La démonstration de ce lemme est laissée en exercice.

Remarque 4.2.5 Ces formules sont les généralisations des relations suivantes sur \mathbb{R} :

$$\begin{aligned} xy &= \frac{1}{2}((x + y)^2 - x^2 - y^2). \\ (x + y)^2 + (x - y)^2 &= 2(x^2 + y^2). \end{aligned}$$

4.3 Formes bilinéaires : représentation matricielle.

Nous allons maintenant définir la matrice d'une forme bilinéaire dans une base, qui va nous permettre, modulo le choix d'une base, de réduire le calcul des formes bilinéaires sur des espaces de dimension finie à des multiplications de matrices.

Définition 4.3.1 Soit V un \mathbb{R} -espace vectoriel de dimension finie n , soit $\mathbf{e} = (e_1, \dots, e_n)$ une base de V , et soit $\varphi : V \times V \rightarrow \mathbb{R}$ une forme bilinéaire. La **matrice représentative** de φ dans la base \mathbf{e} est la matrice $n \times n$, M , dont les coefficients sont donnés par

$$M_{i,j} = (\varphi(e_i, e_j))_{1 \leq i, j \leq n}.$$

Lemme 4.3.2 Soit V un espace vectoriel de dimension finie n . Soient $x, y \in V$, et soit $\mathbf{e} = (e_1, \dots, e_n)$

une base de V . Finalement, soient $\underline{X} = \begin{pmatrix} x_1 \\ \vdots \\ x_n \end{pmatrix}$ et $\underline{Y} = \begin{pmatrix} y_1 \\ \vdots \\ y_n \end{pmatrix}$ les vecteurs coordonnées de x et y dans la

base \mathbf{e} (autrement dit $x = \sum_{i=1}^n x_i e_i, y = \sum_{i=1}^n y_i e_i$). Soit $\varphi : V \times V \rightarrow \mathbb{R}$ une forme bilinéaire, et soit M la matrice qui représente φ dans la base \mathbf{e} . Alors on a

$$\varphi(x, y) = {}^t \underline{X} M \underline{Y} = \sum_{i,j} \varphi(e_i, e_j) x_i y_j.$$

On a

$$\varphi(x, y) = \varphi\left(\sum_{i=1}^n x_i e_i, \sum_{j=1}^n y_j e_j\right) = \sum_{j=1}^n \varphi\left(\sum_{i=1}^n x_i e_i, y_j e_j\right) = \sum_{j=1}^n y_j \varphi\left(\sum_{i=1}^n x_i e_i, e_j\right),$$

puisque φ est linéaire en y . Or on a aussi

$$\varphi\left(\sum_{i=1}^n x_i e_i, e_j\right) = \sum_{i=1}^n \varphi(x_i e_i, e_j) = \sum_{i=1}^n x_i \varphi(e_i, e_j).$$

Ainsi, on obtient

$$\varphi(x, y) = \sum_{j=1}^n y_j \left(\sum_{i=1}^n x_i \varphi(e_i, e_j) \right) = \sum_{i,j} \varphi(e_i, e_j) x_i y_j.$$

On a aussi

$$M\underline{Y} = \begin{pmatrix} \vdots \\ \sum_{j=1}^n \varphi(e_i, e_j) y_j \\ \vdots \end{pmatrix},$$

et donc

$${}^t \underline{X} B \underline{Y} = \left(\cdots \quad x_i \quad \cdots \right) \begin{pmatrix} \vdots \\ \sum_{j=1}^n \varphi(e_i, e_j) y_j \\ \vdots \end{pmatrix} = \sum_{i,j} x_i \varphi(e_i, e_j) y_j = \sum_{i,j} \varphi(e_i, e_j) x_i y_j.$$

Corollaire 4.3.3 Soit V un espace vectoriel de dimension finie n . Soit $\varphi : V \times V \rightarrow \mathbb{R}$ une forme bilinéaire. Les propositions suivantes sont équivalentes.

1. φ est symétrique
2. Pour toute base \mathbf{e} de V , la matrice M de φ dans la base \mathbf{e} est symétrique.
3. Il existe une base \mathbf{e} de V telle que la matrice M de φ dans la base \mathbf{e} est symétrique.

Soit $\varphi : V \times V \rightarrow \mathbb{R}$ une forme bilinéaire, et soit \mathbf{e} une base de V .

Si φ est symétrique, alors on a

$$\varphi(e_i, e_j) = \varphi(e_j, e_i) \text{ pour tout } i, j,$$

et ceci s'écrit matriciellement ${}^t M = M$, par définition de la matrice représentative. On a donc (1) \Rightarrow (2). L'implication (2) \Rightarrow (3) étant claire, il reste à montrer (3) \Rightarrow (1).

Supposons qu'il existe une base \mathbf{e} de V telle que M est symétrique. Soient $x, y \in V$, et soient $\underline{X}, \underline{Y}$ leurs vecteurs de coordonnées dans la base \mathbf{e} . On a alors que

$$\varphi(x, y) = {}^t \underline{X} M \underline{Y}$$

Le membre de droit est une matrice 1×1 : elle est donc égale à sa propre transposée et on a

$$\varphi(x, y) = {}^t \underline{X} M \underline{Y} = {}^t ({}^t \underline{X} M \underline{Y}) = {}^t \underline{Y} {}^t M \underline{X} = {}^t \underline{Y} M \underline{X} = \varphi(y, x)$$

CQFD.

Le lemme précédent admet une réciproque, bien utile pour démontrer qu'une application est bilinéaire et donner sa matrice représentative dans une base fixée.

Lemme 4.3.4 Soit V un \mathbb{R} -espace vectoriel de dimension finie, et soit $\mathbf{e} = (e_1, \dots, e_n)$ une base de V . Pour tout $a_{ij} \in \mathbb{R}, 1 \leq i, j \leq n$, l'application

$$\varphi : V \times V \rightarrow \mathbb{R}, \left(\sum_{i=1}^n x_i e_i, \sum_{j=1}^n y_j e_j \right) \mapsto \sum_{1 \leq i, j \leq n} a_{ij} x_i y_j$$

est une forme bilinéaire, dont la matrice A dans la base \mathbf{e} est donnée par $A_{ij} = (a_{ij})$.

Exemples 4.3.5(1) *L'application*

$$\varphi : \mathbb{R}^2 \times \mathbb{R}^2 \rightarrow \mathbb{R}, \left(\begin{pmatrix} x_1 \\ x_2 \end{pmatrix}, \begin{pmatrix} y_1 \\ y_2 \end{pmatrix} \right) \mapsto x_1y_1 + x_2y_2 + 3x_1y_2 - x_2y_1$$

est bilinéaire, et sa matrice représentative dans la base canonique de \mathbb{R}^2 est

$$M = \begin{pmatrix} 1 & 3 \\ -1 & 1 \end{pmatrix}.$$

(2) *Considérons l'application*

$$\varphi : \mathbb{C}_2[X] \times \mathbb{C}_2[X] \rightarrow \mathbb{C}, (P, Q) \mapsto P(1)Q(0).$$

On peut vérifier directement que φ est bilinéaire, mais on peut aussi utiliser la remarque précédente. Pour cela, considérons la base $1, X, X^2$ de $\mathbb{R}_2[X]$. On écrit

$$P = x_1 + x_2X + x_3X^2, Q = y_1 + y_2X + y_3X^2.$$

On vérifie alors que $\varphi(P, Q) = x_1y_1 + x_2y_1 + x_3y_1$. Donc φ est bilinéaire et sa matrice représentative dans la base $1, X, X^2$ est

$$M = \begin{pmatrix} 1 & 0 & 0 \\ 1 & 0 & 0 \\ 1 & 0 & 0 \end{pmatrix}.$$

Regardons maintenant ce qui se passe lorsque l'on effectue un changement de base.

Proposition 4.3.6 Soit V un \mathbb{R} -espace vectoriel de dimension finie n , soient \mathbf{e} et \mathbf{e}' deux bases de V , et soit P la matrice de passage de la base \mathbf{e} à la base \mathbf{e}' , c'est-à-dire la matrice des coordonnées des vecteurs de \mathbf{e}' dans la base \mathbf{e} . Soit $\varphi : V \times V \rightarrow \mathbb{R}$ une forme bilinéaire, soit M sa matrice dans la base \mathbf{e} et soit N sa matrice dans la base \mathbf{e}' . Alors on a $N = {}^tPMP$.

Soient $x, y \in V$, soient $\underline{X}, \underline{Y}$ leur vecteurs de coordonnées dans la base \mathbf{e} et soient $\underline{X}', \underline{Y}'$ leurs coordonnées dans la base \mathbf{e}' . On a alors $\underline{X} = P\underline{X}'$ et $\underline{Y} = P\underline{Y}'$ pour tout x, y et donc

$$\varphi(x, y) = {}^t\underline{X}M\underline{Y} = {}^t(P\underline{X}')MP\underline{Y}' = {}^t\underline{X}'{}^tPMP\underline{Y}' = {}^t\underline{X}'N\underline{Y}'.$$

c'est à dire que $N = {}^tPMP$ par 3.5.4.

Nous sommes prêts à définir la notion de rang.

Définition 4.3.7 Soit $\varphi : V \times V \rightarrow \mathbb{R}$ une forme bilinéaire. Le **rang** de φ est le rang de n'importe quelle matrice représentative de φ dans une base de V .

Le rang est bien défini et ne dépend pas de la base choisie d'après la proposition précédente.

4.4 Orthogonalité.

Les expressions permettant de calculer $\varphi(x, y)$ peuvent se simplifier grandement lorsque la base \mathbf{e} est choisie convenablement. Par exemple, il est souvent utile de se débarrasser des termes croisés lorsque c'est possible. On introduit pour cela la notion d'orthogonalité.

Définition 4.4.1 Soit V un espace vectoriel de dimension n sur \mathbb{R} , et soit $\varphi : V \times V \rightarrow \mathbb{R}$ une forme bilinéaire symétrique.

On dit que deux vecteurs $x, y \in V$ sont φ -orthogonaux si $\varphi(x, y) = 0$.

On le note $x \underset{\varphi}{\perp} y$, ou $x \perp y$ s'il n'y a pas de confusion possible.

On dit que la base $\mathbf{e} = (e_1, \dots, e_n)$ est φ -orthogonale si les vecteurs de la base sont φ -orthogonaux deux à deux, c'est-à-dire si on a

$$\varphi(e_i, e_j) = 0 \text{ pour tout } i \neq j.$$

Lemme 4.4.2 La base \mathbf{e} est φ -orthogonale si et seulement si M , la matrice de φ dans la base \mathbf{e} , est diagonale.

La base \mathbf{e} est φ -orthogonale $\Leftrightarrow \varphi(e_i, e_j) = 0$ si $i \neq j \Leftrightarrow M_{i,j} = 0$ si $i \neq j \Leftrightarrow M$ est diagonale.

On dit que \mathbf{e} est φ -**orthonormée** si on a

$$\varphi(e_i, e_j) = \begin{cases} 0 & \text{si } i \neq j \\ 1 & \text{si } i = j \end{cases}$$

Lemme 4.4.3 La base \mathbf{e} est φ -orthonormée si et seulement si $\text{Mat}(\varphi, \mathbf{e})$ est la matrice identité.

Laissée en exercice.

Définition 4.4.4 On dit que deux sous-espaces W, W' de V sont **orthogonaux** si on a

$$\varphi(w, w') = 0 \text{ pour tout } w \in W, w' \in W'.$$

On dit que V est la somme directe **orthogonale** des sous-espaces V_1, \dots, V_m si $V = V_1 \oplus \dots \oplus V_m$ et les sous-espaces V_1, \dots, V_m sont orthogonaux deux à deux. On note alors

$$V = V_1 \underset{\perp}{\oplus} \dots \underset{\perp}{\oplus} V_m.$$

Nous aurons besoin du résultat suivant, que nous admettrons :

Lemme 4.4.5 Soit V un espace vectoriel et soit φ une forme bilinéaire sur V . Soient V_1, \dots, V_k des sous-espaces de V tels que $V = \underset{\perp}{\oplus} V_i$. Si pour chaque i on a que \mathbf{v}_i est une base orthonormée de V_i alors la concatenation $(\mathbf{v}_1, \mathbf{v}_2, \dots, \mathbf{v}_k)$ est une base orthonormée de V

Exemples 4.4.6

(2) L'application

$$\varphi : \mathbb{R}_2[X] \times \mathbb{R}_2[X] \rightarrow \mathbb{R}, (P, Q) \mapsto \int_{-1}^1 P(t)Q(t)dt$$

est bilinéaire symétrique. De plus, $1 \underset{\varphi}{\perp} X$ et $X \underset{\varphi}{\perp} X^2$. Par contre, 1 et X^2 ne sont pas φ -orthogonaux, puisque l'on a $\varphi(1, X^2) = \frac{2}{3}$; la base $1, X, X^2$ n'est donc pas φ -orthogonale. Par contre, on peut vérifier que la base

$$1, X, X^2 - \frac{1}{3}$$

est φ -orthogonale. Elle n'est pas φ -orthonormée puisque

$$\varphi(1, 1) = 2, \varphi(X, X) = 2/3, \varphi(X^2 - \frac{1}{3}, X^2 - \frac{1}{3}) = 8/45.$$

Par contre, on peut la rendre φ -orthonormée en multipliant chaque élément de la base par une constante bien choisie. Plus précisément, la base

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

est une base φ -orthonormée.

(3) La base canonique de \mathbb{R}^n est φ -orthonormée pour la forme bilinéaire symétrique

$$\varphi : \mathbb{R}^n \times \mathbb{R}^n \rightarrow \mathbb{R}, \left(\begin{pmatrix} x_1 \\ \vdots \\ x_n \end{pmatrix}, \begin{pmatrix} y_1 \\ \vdots \\ y_n \end{pmatrix} \right) \mapsto x \cdot y = \sum_{i=1}^n x_i y_i$$

(4) Soit $V = C^0([-1, 1], \mathbb{R})$, et soient \mathcal{P} et \mathcal{I} le sous-espace des fonctions paires et impaires respectivement. On sait que l'on a

$$V = \mathcal{P} \oplus \mathcal{I}.$$

Considérons l'application

$$\varphi : V \times V \rightarrow \mathbb{R}, (f, g) \mapsto \int_{-1}^1 f(t)g(t)dt$$

Alors, on a

$$\varphi(f, g) = 0 \text{ pour tout } f \in \mathcal{P}, g \in \mathcal{I}.$$

On a donc

$$V = \mathcal{P} \oplus_{\perp} \mathcal{I}.$$

Le lemme 4.3.2 entraîne immédiatement :

Lemme 4.4.7 Soit V un espace vectoriel de dimension finie n , soit $\mathbf{e} = (e_1, \dots, e_n)$ une base de V , et soient

$$x = \sum_{i=1}^n x_i e_i, y = \sum_{i=1}^n y_i e_i$$

deux vecteurs de V . Soit $\varphi : V \times V \rightarrow \mathbb{R}$ une forme bilinéaire symétrique. Si \mathbf{e} est φ -orthogonale, on a

$$\varphi(x, y) = \sum_{i=1}^n \varphi(e_i, e_i) x_i y_i.$$

En particulier, si \mathbf{e} est φ -orthonormée, on a

$$\varphi(x, y) = \sum_{i=1}^n x_i y_i.$$

Une base φ -orthonormée n'existe pas toujours. En effet, si $\varphi : V \times V \rightarrow \mathbb{R}$ est bilinéaire symétrique et il existe une base φ -orthonormée alors le lemme précédent montre que $\varphi(x, x) > 0$ pour tout $x \neq 0$.

Par exemple, la forme bilinéaire symétrique

$$\varphi : \mathbb{R}^2 \times \mathbb{R}^2 \rightarrow \mathbb{R}, ((x_1, x_2), (y_1, y_2)) \mapsto x_1 y_1 - x_2 y_2.$$

n'admet pas de base φ -orthonormée, puisque $\varphi((0, 1), (0, 1)) = -1 < 0$.

En revanche, on a le théorème suivant :

Théorème 4.4.8 Soit V un espace vectoriel de dimension finie sur \mathbb{R} , et soit $\varphi : V \times V \rightarrow \mathbb{R}$ une forme bilinéaire symétrique. Alors il existe une base de V qui est φ -orthogonale.

On démontre l'existence d'une base φ -orthogonale par récurrence sur $n = \dim(V)$.

Soit (P_n) la propriété :

(P_n) Pour tout \mathbb{R} -espace vectoriel de dimension n et tout $\varphi : V \times V \rightarrow \mathbb{R}$, il existe une base φ -orthogonale.

Si $n = 1$, il n'y a rien à démontrer.

Supposons que (P_n) soit vraie, et soit $\varphi : V \times V \rightarrow \mathbb{R}$ une forme bilinéaire symétrique avec $\dim(V) = n + 1$.

Si $\varphi = 0$, toute base est φ -orthogonale, et on a fini. On suppose donc que $\varphi \neq 0$. Soit q la forme quadratique associée. Par le formule de polarisation, si $q = 0$ alors $\varphi = 0$, ce qui n'est pas le cas. Il existe donc un e_0 tel que $q(e_0) \neq 0$, c'est à dire, $\varphi(e_0, e_0) \neq 0$.

L'application

$$f : V \rightarrow \mathbb{R} \quad y \mapsto \varphi(e_0, y)$$

est alors une application linéaire non nulle, puisque

$$f(e_0) = \varphi(e_0, e_0) \neq 0$$

et son image est donc $= \mathbb{R}$. Par le théorème du rang,

$$\dim \text{Ker}(f) = n + 1 - 1 = n.$$

Par hypothèse de récurrence, il existe une base (e_1, \dots, e_n) de $\text{Ker}(f)$ qui est orthogonale pour la forme

$$\varphi' : \text{Ker}(f) \times \text{Ker}(f) \rightarrow \mathbb{R}, (x, y) \mapsto \varphi(x, y)$$

Montrons que $\mathbf{e} = (e_0, e_1, \dots, e_n)$ est une base de V . Puisque $\dim(V) = n + 1$, il suffit de montrer que la famille (e_0, \dots, e_n) est libre. Soient $\lambda_0, \dots, \lambda_n \in \mathbb{R}$ tels que

$$\lambda_0 e_0 + \lambda_1 e_1 + \dots + \lambda_n e_n = 0.$$

En appliquant f à cette égalité et en utilisant la linéarité, on obtient

$$\lambda_0 f(e_0) + \lambda_1 f(e_1) + \dots + \lambda_n f(e_n) = 0.$$

Puisque $e_1, \dots, e_n \in \text{Ker}(f)$, on obtient $\lambda_0 f(e_0) = 0$. Comme $f(e_0) \neq 0$, on obtient $\lambda_0 = 0$. On a donc

$$\lambda_1 e_1 + \dots + \lambda_n e_n = 0.$$

Comme (e_1, \dots, e_n) est une base de $\text{Ker}(f)$, ils sont linéairement indépendants, et on obtient donc

$$\lambda_1 = \dots = \lambda_n = 0.$$

Ceci prouve que \mathbf{e} est une base de V . Il reste à vérifier que cette base est φ -orthogonale.

Par choix des e_i , on a

$$\varphi(e_i, e_j) = \varphi'(e_i, e_j) = 0 \text{ pour tout } i \neq j, 1 \leq i, j \leq n$$

et aussi

$$\varphi(e_0, e_j) = f(e_j) = 0 \text{ pour tout } j > 0$$

parce que $e_j \in \text{Ker}(f)$. On a donc que

$$\varphi(e_i, e_j) = 0 \text{ pour tout } 0 \leq i \neq j \leq n.$$

Ainsi, (e_0, e_1, \dots, e_n) est une base φ -orthogonale. Ceci achève la récurrence.

Remarques 4.4.9

Le résultat précédent peut être faux si φ n'est pas bilinéaire symétrique. Par exemple, si $\varphi : V \times V \rightarrow \mathbb{R}$ est antisymétrique, c'est-à-dire si on a

$$\varphi(y, x) = -\varphi(x, y) \text{ pour tout } x, y \in V,$$

et si φ est **non nulle**, alors il n'existe pas de base de V qui est φ -orthogonale.

En effet, si φ est une telle forme, alors on a

$$\varphi(x, x) = -\varphi(x, x) \text{ pour tout } x \in V.$$

On a donc

$$\varphi(x, x) = 0 \text{ pour tout } x \in V.$$

Supposons maintenant que $\mathbf{e} = (e_1, \dots, e_n)$ est une base φ -orthogonale. On a donc

$$\varphi(e_i, e_i) = 0 \text{ pour tout } i = 1, \dots, n.$$

Comme $\varphi(e_i, e_j) = 0$ pour tout $i \neq j$ puisque \mathbf{e} est φ -orthogonale, on en déduit que si M est la matrice de φ dans \mathbf{e} alors $M = 0$.

Le Lemme 4.3.2 entraîne alors que l'on a

$$\varphi(x, y) = 0 \text{ pour tout } x, y \in V,$$

ce qui contredit le fait que φ est non nulle.

Un exemple d'une telle forme bilinéaire φ est donné par exemple par

$$\varphi : \mathbb{R}^2 \times \mathbb{R}^2 \rightarrow \mathbb{R}, \left(\begin{pmatrix} x_1 \\ x_2 \end{pmatrix}, \begin{pmatrix} y_1 \\ y_2 \end{pmatrix} \right) \mapsto x_1 y_2 - x_2 y_1.$$

4.5 Calcul effectif d'une base φ -orthogonale.

Nous allons calculer une base φ -orthogonale en exploitant la forme quadratique q qui lui est associée. Rappelons que la forme bilinéaire symétrique φ peut être reconstruite de la forme quadratique s via la formule de polarisation

$$\varphi(x, y) = \frac{1}{2}(q(x + y) - q(x) - q(y)).$$

Nous disons alors que φ est la forme polaire de q , que nous noterons parfois φ_q .

Exemples 4.5.1

(1) L'application

$$q : \mathbb{R}^n \rightarrow \mathbb{R}, x = \begin{pmatrix} x_1 \\ \vdots \\ x_n \end{pmatrix} \mapsto x_1^2 + \dots + x_n^2$$

est une forme quadratique, de forme polaire

$$\varphi_q : \mathbb{R}^n \times \mathbb{R}^n \rightarrow \mathbb{R}, \left(\begin{pmatrix} x_1 \\ \vdots \\ x_n \end{pmatrix}, \begin{pmatrix} y_1 \\ \vdots \\ y_n \end{pmatrix} \right) \mapsto x_1 y_1 + \dots + x_n y_n.$$

En effet, l'application

$$\varphi : \mathbb{R}^n \times \mathbb{R}^n \rightarrow \mathbb{R}, \left(\begin{pmatrix} x_1 \\ \vdots \\ x_n \end{pmatrix}, \begin{pmatrix} y_1 \\ \vdots \\ y_n \end{pmatrix} \right) \mapsto x_1 y_1 + \dots + x_n y_n$$

est bilinéaire symétrique et on a clairement $\varphi(x, x) = q(x)$.

Vérifions la formule de polarisation. On a que

$$q(x + y) = \sum_{i=1}^n (x_i + y_i)^2 = \sum_{i=1}^n x_i^2 + 2x_i y_i + y_i^2 = q(x) + q(y) + 2\varphi(x, y).$$

(2) L'application

$$q : C^0([0, 1], \mathbb{R}) \rightarrow \mathbb{R}, f \mapsto \int_0^1 f(t)^2 dt$$

est une forme quadratique, de forme polaire

$$\varphi_q : C^0([0, 1], \mathbb{R}) \times C^0([0, 1], \mathbb{R}) \rightarrow \mathbb{R}, (f, g) \mapsto \int_0^1 f(t)g(t)dt.$$

En effet, l'application

$$\varphi : C^0([0, 1], \mathbb{R}) \times C^0([0, 1], \mathbb{R}) \rightarrow \mathbb{R}, (f, g) \mapsto \int_0^1 f(t)g(t)dt$$

est bilinéaire symétrique et on a clairement $\varphi(f, f) = q(f)$.

Vérifions de nouveau la formule de polarisation.

$$\begin{aligned} q(f + g) &= \int_0^1 (f(t) + g(t))^2 dt = \int_0^1 f(t)^2 + 2f(t)g(t) + g(t)^2 dt \\ &= q(f) + q(g) + 2 \int_0^1 f(t)g(t)dt. \end{aligned}$$

Définition 4.5.2 Soit V un \mathbb{R} -espace vectoriel de dimension finie n , et soit $q : V \rightarrow \mathbb{R}$ une forme quadratique. Soit \mathbf{e} une base de V . La matrice M de q dans la base \mathbf{e} est la matrice de la forme polaire φ_q dans la base \mathbf{e} . C'est une matrice symétrique par le Corollaire 4.3.3.

Le rang de q , noté $\text{rg}(q)$, est le rang de sa forme polaire.

On dit que \mathbf{e} est q -orthogonale (resp. q -orthonormée) si elle est φ_q -orthogonale (resp. φ_q -orthonormée).

L'égalité $q(x) = \varphi_q(x, x)$ et le Lemme 4.3.2 donnent immédiatement :

Lemme 4.5.3 Soit V un espace vectoriel de dimension finie n et \mathbf{e} une base pour V . Soit $x \in V$, et soit \underline{X} le vecteur coordonnées de x dans la base \mathbf{e} .

Soit $q : V \rightarrow \mathbb{R}$ une forme quadratique, et soit M sa matrice dans la base \mathbf{e} . Alors on a

$$q(x) = {}^t \underline{X} M \underline{X}.$$

En particulier, si \mathbf{e} est q -orthogonale, c'est à dire si M est symétrique, alors on a

$$q(x) = \sum_{i=1}^n q(e_i) x_i^2.$$

Le lemme suivant nous permet de passer directement de la forme quadratique q à sa matrice M sans calculer le forme polaire φ .

Lemme 4.5.4 Soit V un espace vectoriel de dimension finie n . Soient $x, y \in V$, et soit $\mathbf{e} = (e_1, \dots, e_n)$ une base de V . Alors pour tout $a_{ij} \in \mathbb{R}$, $1 \leq i \leq j \leq n$, l'application

$$q : V \rightarrow \mathbb{R}, \sum_{i=1}^n x_i e_i \mapsto \sum_{i=1}^n a_{ii} x_i^2 + 2 \sum_{1 \leq i < j \leq n} a_{ij} x_i x_j$$

est une forme quadratique, et sa matrice A dans la base \mathbf{e} est donnée par

$$A = (a_{ij}).$$

La démonstration est laissée en exercice au lecteur. **Attention au facteur 2 !**

Exemple 4.5.5 L'application

$$q : \mathbb{R}^2 \times \mathbb{R}^2 \rightarrow \mathbb{R}, \begin{pmatrix} x_1 \\ x_2 \end{pmatrix} \mapsto 3x_1^2 + 4x_1 x_2 + 5x_2^2$$

est une forme quadratique, et sa matrice représentative dans la base canonique de \mathbb{R}^2 est donnée par

$$\begin{pmatrix} 3 & 2 \\ 2 & 5 \end{pmatrix}.$$

Soient maintenant φ une forme bilinéaire sur un espace V , q sa forme polaire, \mathbf{e} une base pour V . Soit

$x \in V$ un élément arbitraire et $\underline{X} = \begin{pmatrix} x_1 \\ \vdots \\ x_n \end{pmatrix}$ son vecteur de coordonnées dans la base \mathbf{e} . Alors

\mathbf{e} est φ -orthogonale

\Leftrightarrow

la matrice de φ dans la base \mathbf{e} est diagonale

\Leftrightarrow

la matrice de q dans la base \mathbf{e} est diagonale

⇕

$$\exists a_i \in \mathbb{R} \text{ tel que } q(x) = \sum_{i=1}^n a_i x_i^2.$$

Nous allons maintenant décrire un algorithme, dit algorithme de Gauss, qui permet de trouver une base q -orthogonale. Nous ne justifierons pas complètement toutes les étapes de cet algorithme. L'idée de base sera la suivante : on cherche des coordonnées $x_i(v)$ telles que

$$q(v) = \sum_{i=1}^n a_i (x_i(v))^2$$

et une fois trouvée les coordonnées $x_i(v)$ on cherchera la base orthogonale e_i correspondante, c-a-d, telle que

$$v = \sum_i x_i(v) e_i.$$

Algorithme de Gauss

Soit V un \mathbb{R} -espace vectoriel de dimension finie n , et soit \mathbf{e} une base de V . Soit $q : V \rightarrow \mathbb{R}$ une forme quadratique, et soit $M = (a_{ij})_{1 \leq i, j \leq n}$ sa matrice représentative dans la base \mathbf{e} . Si $x = \sum_{i=1}^n x_i e_i$, on a donc

$$q(x) = \sum_{i=1}^n a_{ii} x_i^2 + 2 \sum_{1 \leq i < j \leq n} a_{ij} x_i x_j =: P(x_1, \dots, x_n).$$

On procède par récurrence sur le nombre de variables. A chaque étape, il y a deux cas.

(1) S'il existe un indice k tel que $a_{kk} \neq 0$, on regroupe tous les termes faisant intervenir la variable x_k , et on complète le carré. On écrit

$$P(x_1, \dots, x_n) = a_{kk} x_k^2 + 2f_k x_k + P_0,$$

où f_k est une forme linéaire en les variables $x_i, i \neq k$, et P_0 est une forme quadratique en les variables $x_i, i \neq k$.

On a alors

$$\begin{aligned} P(x_1, \dots, x_n) &= a_{kk} \left(x_k^2 + \frac{2}{a_{kk}} f_k x_k \right) + P_0 \\ &= a_{kk} \left(\left(x_k + \frac{f_k}{a_{kk}} \right)^2 - \frac{f_k^2}{a_{kk}^2} \right) + P_0. \end{aligned}$$

On peut donc écrire

$$P(x_1, \dots, x_n) = a_{kk} \left(x_k + \frac{f_k}{a_{kk}} \right)^2 + P_1,$$

où P_1 est une forme quadratique en les variables $x_i, i \neq k$.

(2) Si $a_{kk} = 0$ pour tout k , mais qu'il existe k et ℓ tels que $k < \ell$ et $a_{k\ell} \neq 0$. C'est le cas pénible.

On écrit

$$P(x_1, \dots, x_n) = 2a_{k\ell} x_k x_\ell + 2f_k x_k + 2f_\ell x_\ell + P_0,$$

où f_k et f_ℓ sont des formes linéaires en les variables $x_i, i \neq k, \ell$, et P_0 est une forme quadratique en les variables $x_i, i \neq k, \ell$.

On a ainsi

$$P(x_1, \dots, x_n) = 2a_{k\ell} \left(x_k + \frac{1}{a_{k\ell}} f_\ell \right) \left(x_\ell + \frac{1}{a_{k\ell}} f_k \right) - \frac{2}{a_{k\ell}} f_k f_\ell + P_0.$$

On a donc

$$P(x_1, \dots, x_n) = 2a_{k\ell} AB + P_1,$$

avec $A = x_k + \frac{1}{a_{k\ell}} f_\ell$, $B = x_\ell + \frac{1}{a_{k\ell}} f_k$, et P_1 est une forme quadratique en les variables x_i , $i \neq k, \ell$. Pour la suite de la récurrence : si $P_1 = 0$ on arrête, sinon on recommence le procédé avec P_1 .

On a alors

$$P(x_1, \dots, x_n) = \frac{a_{k\ell}}{2} ((A+B)^2 - (A-B)^2) + P_1.$$

* Si $P_1 = 0$, on arrête. Sinon, on recommence le procédé avec P_1 .

On peut montrer que l'on obtient alors une écriture de la forme

$$q(x) = \alpha_1(L_1(x))^2 + \dots + \alpha_r(L_r(x))^2,$$

où :

1. chaque $\alpha_i \in \mathbb{R}^*$
2. chaque L_i est une forme linéaire sur V
3. la famille de formes (L_1, \dots, L_r) est indépendante.

* Ensuite, on choisit des formes linéaires $L_{r+1}, L_{r+2}, \dots, L_n$ de telle façon à ce que la famille (L_1, \dots, L_n) soit libre et on écrit

$$q(x) = \alpha_1(L_1(x))^2 + \dots + \alpha_r(L_r(x))^2 + 0(L_{r+1}(x))^2 + \dots + 0(L_n(x))^2.$$

* Enfin, on cherche une base $\mathbf{e}' = (e'_1, e'_2, \dots, e'_n)$ telle que pour tout $x \in V$ le vecteur de coordonnées de x dans la base \mathbf{e}' soit

$$\underline{X} = \begin{pmatrix} L_1(x) \\ L_2(x) \\ \vdots \\ L_n(x) \end{pmatrix}.$$

Autrement dit, on doit avoir que

$$x = L_1(x)e'_1 + L_2(x)e'_2 + \dots + L_n(x)e'_n,$$

ce qui est un système d'équations qu'on peut résoudre par le pivot de Gauss pour trouver e'_1, \dots, e'_n .

Si x'_1, \dots, x'_n sont les coordonnées de $x \in V$ dans cette nouvelle base, on a

$$q(x) = \alpha_1 x_1'^2 + \dots + \alpha_r x_r'^2.$$

par définition.

Remarque 4.5.6 En particulier, $r = \text{rg}(q)$.

il résulte du lemme 4.5.4 que la matrice de q dans la base \mathbf{e}' est la matrice

$$M = \text{diag}(\alpha_1, \alpha_2, \dots, \alpha_r, 0, \dots, 0)$$

Remarque 4.5.7 Si $\phi : V \times V \rightarrow \mathbb{R}$ est bilinéaire symétrique, alors en appliquant l'algorithme de Gauss à la forme quadratique

$$q_b : V \rightarrow \mathbb{R}, x \mapsto \phi(x, x),$$

on trouve une base \mathbf{v} qui est q_ϕ -orthogonale. Mais par définition, \mathbf{v} est donc orthogonale pour la forme polaire de q_ϕ , qui est ϕ .

Cet algorithme permet donc de trouver une base ϕ -orthogonale pour n'importe quelle forme bilinéaire symétrique ϕ , ainsi que son rang.

Exemple 4.5.8 Soit $q : \mathbb{R}^4 \rightarrow \mathbb{R}$ l'application qui à $\mathbf{u} = \begin{pmatrix} x \\ y \\ z \\ t \end{pmatrix}$ associe

$$q(\mathbf{u}) = x^2 + 2xy + 2xz + 2xt + y^2 + 6yz - 2yt + z^2 + 10zt + t^2.$$

L'application q est une forme quadratique car c'est un polynôme de degré 2 homogène.

Appliquons l'algorithme de Gauss à q pour trouver une base q -orthogonale.

On a

$$\begin{aligned} q(\mathbf{u}) &= x^2 + 2(y+z+t)x + y^2 + 6yz - 2yt + z^2 + 10zt + t^2 \\ &= (x+y+z+t)^2 - (y+z+t)^2 + y^2 + 6yz - 2yt + z^2 + 10zt + t^2 \\ &= (x+y+z+t)^2 + 4yz - 4yt + 8zt. \end{aligned}$$

On a maintenant

$$\begin{aligned} 4yz - 4yt + 8zt &= 4(yz + (-t)y + (2t)z) \\ &= 4((y+2t)(z-t) + 2t^2) \\ &= 4(y+2t)(z-t) + 8t^2 \\ &= (y+z+t)^2 - (y-z+3t)^2 + 8t^2 \end{aligned}$$

Finalemnt, on obtient

$$q(\mathbf{u}) = (x+y+z+t)^2 + (y+z+t)^2 - (y-z+3t)^2 + 8t^2.$$

On a donc $\text{rg}(q) = 4$. On a

$$L_1(u) = x + y + z + t; L_2(u) = y + z + t; L_3(u) = y - z + t; L_4(u) = t$$

Nous cherchons e'_1, \dots, e'_4 tels que pour tout $\begin{pmatrix} x \\ y \\ z \\ t \end{pmatrix} \in \mathbb{R}^4$

$$(x+y+z+t)e'_1 + (y+z+t)e'_2 + (y-z+t)e'_3 + te'_4 = \begin{pmatrix} x \\ y \\ z \\ t \end{pmatrix}$$

ce qui donne, en isolant chaque variable x, y, z, t

$$e'_1 = \begin{pmatrix} 1 \\ 0 \\ 0 \\ 0 \end{pmatrix}, e'_1 + e'_2 + e'_3 = \begin{pmatrix} 0 \\ 1 \\ 0 \\ 0 \end{pmatrix}, e'_1 + e'_2 - e'_3 = \begin{pmatrix} 0 \\ 0 \\ 1 \\ 0 \end{pmatrix}, e'_1 + e'_2 + e'_3 + e'_4 = \begin{pmatrix} 0 \\ 0 \\ 0 \\ 1 \end{pmatrix}.$$

et donc après résolution par le pivot de Gauss on trouve que

$$e'_1 = \begin{pmatrix} 1 \\ 0 \\ 0 \\ 0 \end{pmatrix}, e'_2 = \begin{pmatrix} -1 \\ 1/2 \\ 1/2 \\ 0 \end{pmatrix}, e'_3 = \begin{pmatrix} 0 \\ 1/2 \\ -1/2 \\ 0 \end{pmatrix}, e'_4 = \begin{pmatrix} 0 \\ -1 \\ 0 \\ 1 \end{pmatrix}$$

Ces vecteurs (e'_1, e'_2, e'_3, e'_4) forment donc une base q -orthogonale.

Le théorème suivant n'est valable que pour des formes réelles.

Théorème 4.5.9 (Théorème d'inertie de Sylvester) Soit V un \mathbb{R} -espace vectoriel de dimension finie n , et soit $q : V \rightarrow \mathbb{R}$ une forme quadratique. Soit \mathbf{e} une base q -orthogonale. Soit

$$r_+ = \text{card}\{i | q(e_i) > 0\} \text{ et } r_- = \text{card}\{i | q(e_i) < 0\}.$$

Alors le couple (r_+, r_-) ne dépend pas de la base q -orthogonale choisie. De plus, $r_+ + r_- = \text{rg}(q)$.

Soit $\mathbf{e} = (e_1, \dots, e_n)$ une base q -orthogonale. Posons $\alpha_i = q(e_i) = \varphi_q(e_i, e_i)$ et $r = r_+ + r_-$. Changer l'ordre des vecteurs de \mathbf{e} ne change pas r_+ et r_- , ni le fait que la base soit q -orthogonale. On peut donc supposer sans perte de généralité que l'on a

$$q(e_i) > 0, i = 1, \dots, r_+, q(e_i) < 0, i = r_+ + 1, \dots, r, q(e_i) = 0, i = r + 1, \dots, n.$$

Puisque \mathbf{e} est q -orthogonale (c'est-à-dire φ_q -orthogonale), on obtient que M , la matrice de q dans la base \mathbf{e} , s'écrit

$$M = \begin{pmatrix} q(e_1) & & \\ & \ddots & \\ & & q(e_n) \end{pmatrix}.$$

Or, seuls les réels $q(e_1), \dots, q(e_r)$ sont non nuls. Le rang d'une matrice diagonale étant le nombre de termes diagonaux non nuls, on a bien $\text{rg}(q) = r = r_+ + r_-$.

Soit maintenant \mathbf{e}' une autre base q -orthogonale. Soient (r'_+, r'_-) le couple d'entiers correspondant. Remarquons que l'on a $r'_+ + r'_- = \text{rg}(q) = r$ par le point précédent. Comme précédemment, quitte à changer l'ordre des vecteurs, on peut supposer que

$$q(e'_i) > 0, i = 1, \dots, r'_+, q(e'_i) < 0, i = r'_+ + 1, \dots, r, q(e'_i) = 0, i = r + 1, \dots, n.$$

Montrons que $e_1, \dots, e_{r_+}, e'_{r'_++1}, \dots, e'_n$ sont linéairement indépendants. Supposons que l'on ait une relation

$$\lambda_1 e_1 + \dots + \lambda_{r_+} e_{r_+} + \lambda_{r'_++1} e'_{r'_++1} + \dots + \lambda_n e'_n = 0.$$

On a donc

$$\lambda_1 e_1 + \dots + \lambda_{r_+} e_{r_+} = -(\lambda_{r'_++1} e'_{r'_++1} + \dots + \lambda_n e'_n).$$

En appliquant q des deux côtés, et en utilisant le fait que les bases \mathbf{e} et \mathbf{e}' sont q -orthogonales, on obtient

$$\sum_{i=1}^{r_+} q(e_i) \lambda_i^2 = \sum_{i=r'_++1}^n q(e'_i) \lambda_i^2.$$

Par choix de \mathbf{e} et de \mathbf{e}' , le membre de gauche est ≥ 0 et le membre de droite est ≤ 0 .

On en déduit que l'on a

$$\sum_{i=1}^{r_+} q(e_i) \lambda_i^2 = 0,$$

et puisque $q(e_i) > 0$ pour $i = 1, \dots, r_+$, on en déduit

$$\lambda_1 = \dots = \lambda_{r_+} = 0.$$

Mais alors, on a

$$\lambda_{r'_++1} e'_{r'_++1} + \dots + \lambda_n e'_n = 0,$$

et comme \mathbf{e}' est une base, on en déduit

$$\lambda_{r'_++1} = \dots = \lambda_n = 0.$$

Ainsi, $e_1, \dots, e_{r_+}, e'_{r_++1}, \dots, e'_n$ sont $r_+(n-r_+)$ vecteurs linéairement indépendants dans un espace vectoriel de dimension n . On a donc

$$r_+ + (n - r'_+) \leq n,$$

et donc $r_+ \leq r'_+$. En échangeant les rôles de e et e' , on a de même $r'_+ \leq r_+$.

On a donc $r_+ = r'_+$, et comme on a $\text{rg}(q) = r_+ + r_- = r'_+ + r'_-$, on en déduit $r_- = r'_-$. Ceci achève la démonstration.

Cela conduit à la définition suivante.

Définition 4.5.10 Soit V un \mathbb{R} -espace vectoriel de dimension finie n , et soit $q : V \rightarrow \mathbb{R}$ une forme quadratique. Le couple (r_+, r_-) est appelé la **signature** de q .

Remarque 4.5.11 Pour calculer la signature d'une forme quadratique q , il suffit d'utiliser l'algorithme de Gauss pour écrire $q(x)$ sous la forme

$$\alpha_1(u_{11}x_1 + \dots + u_{1n}x_n)^2 + \dots + \alpha_r(u_{r1}x_1 + \dots + u_{rn}x_n)^2,$$

et de compter le nombre de coefficients α_i qui sont strictement plus grand que 0 et strictement plus petit que 0.

En effet, on a vu que si $\mathbf{v} = (v_1, \dots, v_n)$ est la base q -orthogonale obtenue à la fin de l'algorithme de Gauss, et M est la matrice de q dans cette base, alors

$$M = \text{diag}(\alpha_1, \dots, \alpha_r, 0, \dots, 0).$$

Mais les coefficients diagonaux de M sont exactement les réels $q(v_i)$, et on conclut en utilisant la définition de r_+ et r_- .

Exemple 4.5.12 La signature de la forme quadratique q de l'exemple précédent est $(3, 1)$.

Chapitre 5

Produits scalaires.

Contrairement aux premiers chapitres, cette section n'est valable que pour les espaces vectoriels réels. Nous survolerons en fin de chapitre sa généralisation à des espaces complexes.

5.1 Produit scalaires, définitions et exemples.

Nous voulons maintenant généraliser la notion de produit scalaire - et donc de longueur, de distance et d'angle - à un espace vectoriel réel arbitraire. Sur \mathbb{R}^2 ou \mathbb{R}^3 , et plus généralement \mathbb{R}^n , l'application produit scalaire canonique

$$\mathbb{R}^n \times \mathbb{R}^n \rightarrow \mathbb{R}, \left(\left(\begin{array}{c} x_1 \\ \vdots \\ x_n \end{array} \right), \left(\begin{array}{c} y_1 \\ \vdots \\ y_n \end{array} \right) \right) \mapsto x \cdot y = \sum_{i=1}^n x_i y_i$$

est bilinéaire et symétrique. Notons que la longueur d'un vecteur $x \in \mathbb{R}^3$ peut être calculée par $\|x\| = \sqrt{x \cdot x}$: nous aurions donc envie d'associer une notion de longueur à une forme bilinéaire φ en posant $\|x\| = \sqrt{\varphi(x, x)}$. Malheureusement, il n'est pas sûr que cette quantité existe (si $\varphi(x, x) < 0$ elle n'existe pas) ni qu'elle soit strictement positive pour un x non-nul (or nous ne voulons pas une distance 0 entre deux vecteurs distincts).

Ces considérations inspirent les définitions suivantes :

Définition 5.1.1 Soit V un espace vectoriel réel. On dit qu'une forme bilinéaire symétrique $\varphi : V \times V \rightarrow \mathbb{R}$ est **positive** si $\varphi(x, x) \geq 0$ pour tout $x \in V$, et **définie positive** si $\varphi(x, x) > 0$ pour tout $x \in V, x \neq 0$.

Remarquons que φ est définie positive si et seulement si

- (i) φ est positive et
- (ii) $\varphi(x, x) = 0 \Rightarrow x = 0_V$.

C'est en général cette reformulation de la définition que l'on utilise en pratique pour vérifier si oui ou non une forme bilinéaire donnée est définie positive.

Définition 5.1.2 Soit V un \mathbb{R} -espace vectoriel (non nécessairement de dimension finie). Un **produit scalaire** sur V est une forme bilinéaire

$$\langle \cdot, \cdot \rangle : V \times V \rightarrow \mathbb{R}, (x, y) \mapsto \langle x, y \rangle$$

symétrique et définie positive.

On dit alors que le couple $(V, \langle \cdot, \cdot \rangle)$ est un espace **prehilbertien**.

Exemples 5.1.3

(1) L'application

$$\mathbb{R}^n \times \mathbb{R}^n \rightarrow \mathbb{R}, \left(\left(\begin{array}{c} x_1 \\ \vdots \\ x_n \end{array} \right), \left(\begin{array}{c} y_1 \\ \vdots \\ y_n \end{array} \right) \right) \mapsto \sum_{i=1}^n x_i y_i$$

est un produit scalaire.

(2) L'application

$$\langle \cdot, \cdot \rangle : C^0([a, b], \mathbb{R}) \times C^0([a, b], \mathbb{R}) \rightarrow \mathbb{R}, (f, g) \mapsto \langle f, g \rangle = \int_a^b f(t)g(t)dt$$

est un produit scalaire.

Nous allons traiter cet exemple en détail.

(a) Montrons que $\langle \cdot, \cdot \rangle$ est symétrique. En effet, pour tout $f, g \in C^0([a, b], \mathbb{R})$, on a

$$\langle g, f \rangle = \int_a^b g(t)f(t)dt = \int_a^b f(t)g(t)dt = \langle f, g \rangle.$$

(b) Montrons que $\langle \cdot, \cdot \rangle$ est bilinéaire. Pour tout $f_1, f_2, f, g \in C^0([a, b], \mathbb{R})$, $\lambda \in \mathbb{R}$, on a

$$\begin{aligned} \langle f_1 + f_2, g \rangle &= \int_a^b (f_1 + f_2)(t)g(t)dt \\ &= \int_a^b (f_1(t) + f_2(t))g(t)dt \\ &= \int_a^b f_1(t)g(t)dt + \int_a^b f_2(t)g(t)dt \\ &= \langle f_1, g \rangle + \langle f_2, g \rangle \end{aligned}$$

et aussi

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

Par (a) il découle que

$$\langle f, g_1 + g_2 \rangle = \langle f, g_1 \rangle + \langle f, g_2 \rangle \text{ et } \langle f, \lambda g \rangle = \lambda \langle f, g \rangle$$

pour tout $f, g, g_1, g_2 \in \mathbb{R}[X]$, $\lambda \in \mathbb{R}$ (**Vérifiez-le vous-même !!!**).

Ainsi, $\langle \cdot, \cdot \rangle$ est bilinéaire.

(b) Montrons que $\langle \cdot, \cdot \rangle$ est symétrique. En effet, pour tout $f, g \in \mathbb{R}[X]$, on a

$$\langle g, f \rangle = \int_a^b g(t)f(t)dt = \int_a^b f(t)g(t)dt = \langle f, g \rangle.$$

(c) Montrons enfin que $\langle \cdot, \cdot \rangle$ est définie positive. On va utiliser pour cela la reformulation de la définition 5.1.1.

Pour tout $f \in C^0([a, b], \mathbb{R})$, on a

$$\langle f, f \rangle = \int_a^b f(t)^2 dt.$$

Or, l'intégrale d'une fonction positive est positive.

Comme la fonction $f^2(t)$ est positive, on en déduit que

$$\langle f, f \rangle \geq 0 \text{ pour tout } f \in C^0([a, b], \mathbb{R}).$$

Supposons maintenant que l'on a $\langle f, f \rangle = 0$, c'est à dire que

$$\int_a^b f(t)^2 dt = 0.$$

Or l'intégrale d'une fonction **positive et continue** $f : [a, b] \rightarrow \mathbb{R}$ est nulle si et seulement si f est identiquement nulle. **Exercice : démontrez-le vous même.** Comme la fonction

$$[a, b] \rightarrow \mathbb{R}, t \mapsto f(t)^2$$

est positive et continue, on en déduit

$$f(t)^2 = 0 \text{ pour tout } t \in [a, b],$$

c'est-à-dire $f = 0$: CQFD.

(3). Pour tout fonction continue et strictement positive sur $[a, b]$, $p : [a, b] \rightarrow \mathbb{R}$, la fonction

$$\langle \cdot, \cdot \rangle : C^0([a, b], \mathbb{R}) \times C^0([a, b], \mathbb{R}) \rightarrow \mathbb{R}, (f, g) \mapsto \langle f, g \rangle = \int_a^b p(t)f(t)g(t)dt$$

est un produit scalaire.

(4) L'application $M_n(\mathbb{R}) \times M_n(\mathbb{R}) \rightarrow \mathbb{R}, (M, N) \mapsto \text{Tr}({}^tMN)$ est un produit scalaire sur $M_n(\mathbb{R})$.

(5) L'application

$$\varphi : \mathbb{R}^2 \times \mathbb{R}^2 \rightarrow \mathbb{R}, \left(\begin{pmatrix} x_1 \\ x_2 \end{pmatrix}, \begin{pmatrix} y_1 \\ y_2 \end{pmatrix} \right) \mapsto x_1y_1 - x_2y_2$$

n'est pas un produit scalaire. Elle est bien bilinéaire symétrique, mais elle n'est pas positive, comme on l'a vu précédemment.

(6) L'application

$$\varphi : \mathbb{R}[X] \times \mathbb{R}[X] \rightarrow \mathbb{R}, (P, Q) \mapsto P(0)Q(0)$$

n'est pas un produit scalaire. Elle est bien bilinéaire, symétrique, positive, mais pas définie positive. Par exemple, on a $\varphi(X, X) = 0$, mais $X \neq 0$.

5.2 Produits scalaires : géométrie.

Les propriétés du produit scalaire permettent alors, comme dans le cas classique, de définir la "longueur", ou *norme* d'un vecteur de V .

Définition 5.2.1 Soit $(V, \langle \cdot, \cdot \rangle)$ un espace préhilbertien. Pour tout $x \in V$, on définit la **norme** de x , notée $\|x\|$, par

$$\|x\| = \sqrt{\langle x, x \rangle}.$$

Notons que par définition d'un produit scalaire, $\|x\| \geq 0$, et $\|x\| = 0$ si et seulement si $x = 0$.

Définition 5.2.2 Soit $(V, \langle \cdot, \cdot \rangle)$ un espace préhilbertien. Soient $v, w \in V$. On définit la distance entre v et w par

$$d(v, w) = \|v - w\|.$$

Encore une fois, la distance entre v et w est positive et n'est 0 que si $v = w$. Nous aurions envie de poser la définition suivante :

Définition 5.2.3 Soit $(V, \langle \cdot, \cdot \rangle)$ un espace préhilbertien. Soient $v, w \in V$ avec $v, w \neq 0$. On définit l'angle entre v et w par

$$\theta = \arccos \left(\frac{\langle v, w \rangle}{\|v\| \times \|w\|} \right).$$

Remarque 5.2.4 Avec cette définition de θ , l'angle entre v et w , nous avons automatiquement $\theta \in [0, \pi]$. Par ailleurs, il s'agit d'une angle non-orienté : θ ne dépend pas de l'ordre de v et w .

Malheureusement, ce n'est pas évident que cette définition soit bien posée. En effet, la fonction arccos n'est définie que pour des nombres réels x satisfaisant la condition $-1 \leq x \leq 1$ ou autrement dit $|x| \leq 1$. Nous devons donc vérifier la proposition suivante :

Proposition 5.2.5 (Inégalité de Cauchy-Schwarz) *Soit $(V, \langle \cdot, \cdot \rangle)$ un espace préhilbertien. Alors pour tout $x, y \in V$, on a*

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

et on a égalité dans cette expression si et seulement si la famille x, y est liée sur \mathbb{R} , c'est-à-dire s'il existe $\lambda, \mu \in \mathbb{R}$, $(\lambda, \mu) \neq (0, 0)$ tels que $\lambda x + \mu y = 0$.

Le résultat étant immédiat si x ou y est égal à 0, on peut supposer $x, y \neq 0$: si $x, y \neq 0$ nous avons qu'il existe $\lambda, \mu \in \mathbb{R}$, $(\lambda, \mu) \neq (0, 0)$ tels que $\lambda x + \mu y = 0$ si et seulement si il existe $t \in \mathbb{R}$ tel que $x + ty = 0$. Considérons la fonction de t

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

Ceci est une fonction quadratique de t qui ne prend pas de valeurs négatives : elle a donc un discriminant $\Delta \leq 0$, c'est à dire

$$\Delta = 4(\langle x, y \rangle)^2 - 4\|x\|^2\|y\|^2 \leq 0.$$

On a donc que

$$(\langle x, y \rangle)^2 \leq \|x\|^2\|y\|^2$$

et

$$|\langle x, y \rangle| \leq \|x\|\|y\|.$$

De plus, on a égalité dans cette expression si et seulement si $\Delta = 0$, c'est-à-dire si et seulement si il existe t tel que $f(t) = 0$. Par définition de $f(t)$, nous avons égalité dans cette expression si et seulement si il existe t tel que $x + ty = 0$. CQFD. L'inégalité de Cauchy-Schwarz est donc valable et notre définition de θ est bien posée.

Un certain nombre de formules de la géométrie dans l'espace sont toujours valables dans ce contexte :

Lemme 5.2.6 (Théorème de Pythagore) *Soit $(V, \langle \cdot, \cdot \rangle)$ un espace préhilbertien et soient $v, w \in V$ avec $v, w \neq 0_V$. Soit θ l'angle entre v et w . Alors on a*

$$d(v, w)^2 = d(0_V, v)^2 + d(0_V, w)^2 \Leftrightarrow \theta = \pi/2.$$

On note tout d'abord que par définition $\theta = \pi/2$ si et seulement si $\langle v, w \rangle = 0$. Par définition,

$$\begin{aligned} d(v, w)^2 &= \langle v - w, v - w \rangle \\ &= \langle v, v \rangle + \langle w, w \rangle - 2\langle v, w \rangle \\ &= d(v, 0_V)^2 + d(w, 0_V)^2 - 2\langle v, w \rangle \end{aligned}$$

et donc

$$d(v, w)^2 = d(0, v)^2 + d(0, w)^2 \Leftrightarrow \langle v, w \rangle = 0 \Leftrightarrow \theta = \pi/2.$$

Lemme 5.2.7 (Identité du parallélogramme) *Soit $(V, \langle \cdot, \cdot \rangle)$ un espace préhilbertien et soient $v, w \in V$. On a alors*

$$\|v + w\|^2 + \|v - w\|^2 = 2(\|v\|^2 + \|w\|^2).$$

Exercice pour le lecteur.

Lemme 5.2.8 (Inégalité triangulaire) *Soit $(V, \langle \cdot, \cdot \rangle)$ un espace préhilbertien et soient $v, w \in V$. On a alors*

$$\|v + w\| \leq \|v\| + \|w\|.$$

On a que

$$\|v + w\|^2 = \|v\|^2 + \|w\|^2 + 2\langle v, w \rangle.$$

Par l'inégalité de Cauchy-Schwarz on a que

$$\|v + w\|^2 \leq \|v\|^2 + \|w\|^2 + 2\|v\| \times \|w\| = (\|v\| + \|w\|)^2.$$

Puisque $\|v + w\|$ et $\|v\| + \|w\|$ sont positifs, on peut prendre la racine carrée des deux membres pour déduire que

$$\|v + w\| \leq \|v\| + \|w\|.$$

Les deux lemmes suivants sont souvent très utiles.

Lemme 5.2.9 Soit $(V, \langle \cdot, \cdot \rangle)$ un espace préhilbertien, et soient $x_1, \dots, x_k \in V$ une famille de vecteurs deux à deux orthogonaux. Alors on a

$$\|x_1 + \dots + x_k\|^2 = \|x_1\|^2 + \dots + \|x_k\|^2.$$

Supposons $x_1, \dots, x_k \in V$ deux à deux orthogonaux. On a donc

$$\langle x_i, x_j \rangle = 0 \text{ pour tout } i \neq j.$$

Par ailleurs, on a que

$$\|x_1 + \dots + x_k\|^2 = \langle x_1 + \dots + x_k, x_1 + \dots + x_k \rangle = \sum_{i,j=1}^k \langle x_i, x_j \rangle.$$

Mais puisque $\langle x_i, x_j \rangle = 0$ pour tout $i \neq j$, on obtient

$$\|x_1 + \dots + x_k\|^2 = \sum_{i=1}^k \langle x_i, x_i \rangle = \sum_{i=1}^k \|x_i\|^2,$$

ce que l'on voulait démontrer.

Lemme 5.2.10 Soit $(V, \langle \cdot, \cdot \rangle)$ un espace préhilbertien, et soient $x_1, \dots, x_k \in V$ des vecteurs **non nuls** deux à deux orthogonaux. Alors (x_1, \dots, x_k) est une famille libre.

Soient $\lambda_1, \dots, \lambda_k \in \mathbb{R}$ tels que

$$\lambda_1 x_1 + \dots + \lambda_k x_k = 0_V.$$

Soit $j \in \{1, \dots, k\}$. On a

$$\langle x_j, \lambda_1 x_1 + \dots + \lambda_k x_k \rangle = \langle x_j, 0_V \rangle = 0,$$

et donc

$$\sum_{i=1}^k \lambda_i \langle x_j, x_i \rangle = 0.$$

Puisque les x_i sont deux à deux orthogonaux, cela s'écrit

$$\lambda_j \langle x_j, x_j \rangle = 0.$$

Puisque par hypothèse $x_j \neq 0$, on a $\langle x_j, x_j \rangle > 0$, et donc $\lambda_j = 0$. Ceci achève la démonstration.

Revenons maintenant à l'existence de bases orthonormées.

Proposition 5.2.11 Soit $(V, \langle \cdot, \cdot \rangle)$ un espace préhilbertien de dimension finie. Alors V possède une base (v_1, \dots, v_n) orthonormée pour le produit scalaire.

De plus, si (v_1, \dots, v_n) est une base orthonormée, alors pour tout $x \in V$, on a

$$x = \langle v_1, x \rangle v_1 + \dots + \langle v_n, x \rangle v_n.$$

D'après le Théorème 4.4.8, il existe une base $\mathbf{e} = (e_1, \dots, e_n)$ de V qui est orthogonale pour le produit scalaire. Puisque \mathbf{e} est une base, $e_i \neq 0$ pour tout i , et on a donc $\|e_i\| \neq 0$.

Pour tout $i = 1, \dots, n$, on pose

$$v_i = \frac{1}{\|e_i\|} e_i.$$

Il est clair que (v_1, \dots, v_n) est une base de V .

De plus, on a

$$\langle v_i, v_j \rangle = \left\langle \frac{1}{\|e_i\|} e_i, \frac{1}{\|e_j\|} e_j \right\rangle = \frac{1}{\|e_i\| \times \|e_j\|} \langle e_i, e_j \rangle \text{ pour tout } i, j.$$

Puisque \mathbf{e} est une base orthogonale, on obtient

$$\langle v_i, v_j \rangle = 0 \text{ pour tout } i \neq j.$$

De plus, pour tout i , on a

$$\langle v_i, v_i \rangle = \frac{1}{\|e_i\|^2} \langle e_i, e_i \rangle = \frac{\langle e_i, e_i \rangle}{\langle e_i, e_i \rangle} = 1.$$

Ainsi, (v_1, \dots, v_n) est une base orthonormée.

Soit maintenant (v_1, \dots, v_n) une base orthonormée, et soit $x \in V$. Puisque v_1, \dots, v_n est une base, on peut écrire

$$x = \lambda_1 v_1 + \dots + \lambda_n v_n.$$

Pour tout j , on a alors

$$\langle v_j, x \rangle = \sum_{i=1}^n \lambda_i \langle v_j, v_i \rangle = \lambda_j,$$

la dernière égalité provenant du fait que v_1, \dots, v_n est une base orthonormée. On a donc bien l'égalité annoncée.

Nous avons donc maintenant une notion satisfaisante de la distance entre deux éléments d'un espace vectoriel muni d'un produit scalaire. Rappelons que la question qui a motivé ce travail est la suivante : je veux construire dans un espace vectoriel V un "bon approximant" w pour un élément v sous la contrainte que w doit être contenu dans un sous-espace W . Nous savons maintenant ce qu'on veut dire exactement par un "bon approximant" : on veut que la distance $d(v, w)$ entre v et w soit la plus petite possible. Le lemme suivant nous donne un critère numérique pour que $w \in W$ soit le "meilleur approximant" pour v .

Lemme 5.2.12 *Soit V un espace préhilbertien, W un sous-espace de V et v un élément de V . Si $w \in W$ a la propriété que $\langle v - w, w' \rangle = 0$ pour tout $w' \in W$ alors pour tout $w' \in W$ on a que $d(v, w) \leq d(v, w')$, avec égalité si et seulement si $w' = w$.*

Autrement dit, si la droite qui relie v à $w \in W$ est perpendiculaire à W alors w est le point de W le plus proche de v . Ce résultat vous est familier depuis le lycée pour le cas où $v \in \mathbb{R}^3$ et W est un plan dans \mathbb{R}^3 . On a que

$$d(v, w') = \|v - w'\| = \|(v - w) + (w - w')\|.$$

Maintenant, $w - w' \in W$ donc par hypothèse $(v - w) \perp (w - w')$ et par le théorème de Pythagore

$$d(v, w')^2 = \|(v - w)\|^2 + \|(w - w')\|^2 \geq d(v, w)^2$$

avec égalité si et seulement si $\|w - w'\| = 0$, c'est-à-dire $w = w'$.

Notre critère est que $(v - w)$ doit être orthogonal à tous les éléments de W . Etudions donc l'ensemble constitué de tels éléments.

Définition 5.2.13 *Soit $(V, \langle \cdot, \cdot \rangle)$ un espace préhilbertien et soit S un sous-ensemble de V . L'orthogonal de S , noté S^\perp , est le sous-ensemble de V défini par*

$$S^\perp = \{x \in V \mid \langle s, x \rangle = 0 \text{ pour tout } s \in S\}.$$

Exercice. Démontrer que S^\perp est toujours un sous-espace vectoriel de W .

Théorème 5.2.14 Soit $(V, \langle \cdot, \cdot \rangle)$ un espace préhilbertien et soit W un sous-espace vectoriel de V . Alors :

1. Pour tout $w \in W$ et tout $w' \in W^\perp$, on a $w \perp w'$. De plus, $W \cap W^\perp = \{0_V\}$.
2. Si W est de dimension finie, on a $V = W \oplus W^\perp$. Autrement dit, tout $x \in V$ s'écrit de manière unique sous la forme

$$x = w + w', w \in W, w' \in W^\perp.$$

De plus, si (v_1, \dots, v_k) est une base orthonormée pour W alors on a $w = \sum_{i=1}^k \langle x, v_i \rangle v_i$.

(1) Si $w \in W$ et $w' \in W^\perp$, alors on a $\langle w, w' \rangle = 0$ par définition de W^\perp . On a donc $w \perp w'$. Soit maintenant $w \in W \cap W^\perp$. Puisque $w \in W^\perp$ et $w \in W$ on a que $\langle w, w \rangle = 0$ et donc $w = 0$ d'après les propriétés du produit scalaire.

Ainsi, on a $W \cap W^\perp = \{0\}$, ce qu'il fallait vérifier.

(2) D'après (1), il reste à démontrer que $V = W + W^\perp$, c'est-à-dire que tout vecteur $v \in V$ peut s'écrire $v = w + w'$ avec $w \in W$ et $w' \in W^\perp$.

Si $W = \{0\}$, on a $W^\perp = V$, et il n'y a rien à faire. On peut donc supposer que W n'est pas l'espace trivial $\{0_V\}$. La restriction à W du produit scalaire sur V est encore un produit scalaire. Puisque W est de dimension finie, W possède une base orthonormée (v_1, \dots, v_k) d'après la proposition précédente.

Soit $v \in V$. On pose

$$w = \sum_{i=1}^k \langle v_i, v \rangle v_i.$$

Alors $w \in W$. D'autre part, on a

$$\begin{aligned} \langle v_j, v - w \rangle &= \langle v_j, v \rangle - \langle v_j, w \rangle \\ &= \langle v_j, v \rangle - \langle v_j, \sum_{i=1}^k \langle v_i, v \rangle v_i \rangle \\ &= \langle v_j, v \rangle - \sum_{i=1}^k \langle v_i, v \rangle \langle v_j, v_i \rangle. \end{aligned}$$

Puisque v_1, \dots, v_k est orthonormée, on en déduit :

$$\langle v_j, v - w \rangle = \langle v_j, v \rangle - \langle v_j, v \rangle = 0,$$

et ceci pour tout $j = 1, \dots, k$.

Soit $s \in W$. Alors on peut écrire $s = s_1 v_1 + \dots + s_k v_k$, et donc

$$\langle s, v - w \rangle = \sum_{i=j}^k s_j \langle v_j, v - w \rangle = 0.$$

Ainsi, $v - w \in W^\perp$, et donc on a la décomposition voulue en posant $w' = v - w$. Si maintenant on a deux décompositions

$$v = w_1 + w'_1 = w_2 + w'_2, w_i \in W, w'_i \in W^\perp,$$

on a

$$w_1 - w_2 = w'_2 - w'_1 \in W \cap W^\perp,$$

car W et W^\perp sont des sous-espaces vectoriels de V . Par le premier point, on en déduit $w_1 - w_2 = w'_2 - w'_1 = 0_V$, et donc $w_1 = w_2, w'_1 = w'_2$, CQFD.

Remarque 5.2.15 Le point (2) est faux sans hypothèse de finitude de la dimension de W .

D'après le deuxième point du théorème, lorsque W est de dimension finie, tout $x \in V$ se décompose de manière unique sous la forme

$$x = w + w', w \in W, w' \in W^\perp.$$

Cela conduit à la définition suivante :

Définition 5.2.16 Soit $(V, \langle \cdot, \cdot \rangle)$ un espace préhilbertien, et soit W un sous-espace de V de dimension finie. Pour tout $x = w + w' \in V$ avec $w \in W$ et $w' \in W^\perp$ on pose

$$p_W(x) = w.$$

Le vecteur $p_W(x) \in W$ est appelé la **projection orthogonale** de x sur W . Si (v_1, \dots, v_k) est une base orthonormée de W alors on a

$$p_W(x) = \sum_{i=1}^k \langle x, v_i \rangle v_i.$$

Le lecteur vérifiera à titre d'exercice les propriétés suivantes :

1. L'application $p_W : V \rightarrow V$ est linéaire.
2. Pour tout $x \in V$, on a $p_W(x) \in W$, $(x - p_W(x)) \in W^\perp$.

Notons que la projection orthogonale a surtout la propriété intéressante suivante :

$$p_W(x) \text{ est le point de } W \text{ le plus proche de } x.$$

Il nous est donc important d'avoir une formule explicite pour calculer cette projection orthogonale si on veut l'utiliser pour des calculs effectifs. Les résultats ci-dessous nous livrent une telle formule,

$$p_W(x) = \sum_i \langle x, v_i \rangle v_i,$$

à condition de disposer d'une base orthonormée $(v_1 \dots v_n)$ pour W . Il nous est donc important de pouvoir construire de telles bases orthonormées, ce que nous faisons dans le prochain paragraphe.

5.3 Procédé d'orthonormalisation de Gram-Schmidt.

Soit $(V, \langle \cdot, \cdot \rangle)$ un espace préhilbertien de dimension finie. On suppose donnée une base pour V , $\mathbf{e} = (e_1, \dots, e_n)$. Nous allons construire récursivement une nouvelle base (orthonormée) $\mathbf{v} = (v_1, \dots, v_n)$ à partir de \mathbf{e} par la méthode suivante :

1. On pose $v_1 = \frac{e_1}{\|e_1\|}$,
2. On suppose donnés (v_1, \dots, v_k) , et on se propose de construire v_{k+1} .
3. (Orthogonalisation). On introduit un vecteur auxiliaire f_{k+1} défini par

$$f_{k+1} = e_{k+1} - \sum_{i=1}^k \langle e_{k+1}, v_i \rangle v_i.$$

Par construction f_{k+1} est orthogonal aux vecteurs v_1, \dots, v_k . Il n'est pas, par contre, de longueur 1.

4. (Normalisation). On pose

$$v_{k+1} = \frac{f_{k+1}}{\|f_{k+1}\|}.$$

5. Nous avons maintenant l'ensemble (v_1, \dots, v_{k+1}) . On revient à l'étape (2) pour construire v_{k+2} .

Vérifions maintenant que cette construction donne bien une base orthonormée.

Proposition 5.3.1 *Les vecteurs de la famille \mathbf{v} construite ci-dessus forment une base orthonormée pour V .*

Il suffira de montrer que cette famille est orthonormée : elle est alors automatiquement libre et comme elle contient $n = \dim V$ éléments c'est une base.

il résulte directement de $v_k = f_k/\|f_k\|$ que $\|v_k\| = 1$. On va montrer par récurrence sur $k \geq 2$ la proposition

$P(k)$: les éléments (v_1, \dots, v_k) sont orthogonaux deux à deux $\forall k \in \{2, \dots, n\}$.

Démontrons d'abord $P(2)$. On doit montrer que $\langle v_1, v_2 \rangle = 0$. Puisque $\langle v_1, v_2 \rangle = \frac{1}{\|e_1\| \|f_2\| \langle e_1, f_2 \rangle}$ il suffira de montrer que $\langle e_1, f_2 \rangle = 0$.

Mais

$$\begin{aligned} \langle e_1, f_2 \rangle &= \langle e_1, e_2 - \langle e_2, v_1 \rangle v_1 \rangle \\ &= \langle e_1, e_2 \rangle - \langle e_1, \langle e_2, v_1 \rangle v_1 \rangle \\ &= \langle e_1, e_2 \rangle - \langle e_2, v_1 \rangle \langle e_1, v_1 \rangle \\ &= \langle e_1, e_2 \rangle - \frac{\langle e_2, e_1 \rangle}{\|e_1\|} \|e_1\| = 0. \end{aligned}$$

$P(2)$ est donc vrai. Supposons que $P(k-1)$ est vraie et déduisons $P(k)$. Il suffira de démontrer que $\langle v_i, v_k \rangle = 0$ pour tout $i < k$. Mais on a

$$\begin{aligned} \langle v_i, v_k \rangle &= \frac{1}{\|f_k\|} \langle v_i, f_k \rangle \\ &= \frac{1}{\|f_k\|} \langle v_i, e_k - \sum_{j=1}^{k-1} \langle e_k, v_j \rangle v_j \rangle \\ &= \frac{1}{\|f_k\|} \left(\langle v_i, e_k \rangle - \sum_{j=1}^{k-1} \langle e_k, v_j \rangle \langle v_i, v_j \rangle \right) \end{aligned}$$

et maintenant puisque $\langle v_i, v_j \rangle = 0$ si $i \neq j$ et $\langle v_i, v_j \rangle = 1$ si $i = j$ on a

$$\frac{1}{\|f_k\|} \left(\langle v_i, e_k \rangle - \sum_{j=1}^{k-1} \langle e_k, v_j \rangle \langle v_i, v_j \rangle \right) = \frac{1}{\|f_k\|} (\langle v_i, e_k \rangle - \langle e_k, v_i \rangle) = 0.$$

On a donc $P(k)$ pour tout $k \leq n$: en particulier $P(n)$ est vraie et cette famille est une base orthonormée.

Exemple 5.3.2 *On considère la base de \mathbb{R}^3*

$$e_1 = \begin{pmatrix} 1 \\ 1 \\ 0 \end{pmatrix}, e_2 = \begin{pmatrix} 1 \\ 0 \\ 1 \end{pmatrix}, e_3 = \begin{pmatrix} 0 \\ 1 \\ 1 \end{pmatrix}.$$

Appliquons le procédé de Gram-Schmidt à cette base afin d'obtenir une base orthonormée pour le produit scalaire. On pose $v_1 = e_1/\|e_1\| = \begin{pmatrix} 1/\sqrt{2} \\ 1/\sqrt{2} \\ 0 \end{pmatrix}$. On a

$$f_2 = e_2 - \langle v_1, e_2 \rangle v_1 = \begin{pmatrix} \frac{1}{2} \\ -\frac{1}{2} \\ 1 \end{pmatrix}.$$

On pose

$$v_2 = f_2 / \|f_2\| = \begin{pmatrix} \frac{1}{\sqrt{6}} \\ -\frac{1}{\sqrt{6}} \\ \frac{2}{\sqrt{6}} \end{pmatrix}.$$

Enfin,

$$f_3 = e_3 - \langle v_1, e_3 \rangle v_1 - \langle v_2, e_3 \rangle v_2 = \begin{pmatrix} -2/3 \\ 2/3 \\ 2/3 \end{pmatrix},$$

et donc

$$v_3 = \frac{f_3}{\|f_3\|} = \frac{\sqrt{3}}{2} \begin{pmatrix} -2/3 \\ 2/3 \\ 2/3 \end{pmatrix}.$$

On a donc

$$v_1 = \frac{1}{\sqrt{2}} \begin{pmatrix} 1 \\ 0 \\ 0 \end{pmatrix}, v_2 = \sqrt{\frac{2}{3}} \begin{pmatrix} 1/2 \\ 1/2 \\ 1 \end{pmatrix}, v_3 = \frac{\sqrt{3}}{2} \begin{pmatrix} -2/3 \\ 2/3 \\ 2/3 \end{pmatrix}.$$

Exemple 5.3.3 Construisons une base orthonormée pour le plan d'équation $x + y + z = 0$ dans \mathbb{R}^3 . Il a une base non orthonormée (e_1, e_2) donnée par

$$e_1 = \begin{pmatrix} 1 \\ -1 \\ 0 \end{pmatrix}, e_2 = \begin{pmatrix} 1 \\ 0 \\ -1 \end{pmatrix}.$$

On pose $v_1 = \frac{e_1}{\|e_1\|} = \begin{pmatrix} 1/\sqrt{2} \\ -1/\sqrt{2} \\ 0 \end{pmatrix}$. On introduit alors

$$f_2 = e_2 - \langle v_1, e_2 \rangle v_1 = e_2 - \frac{1}{\sqrt{2}} v_1 = \begin{pmatrix} 1/2 \\ 1/2 \\ -1 \end{pmatrix}$$

et on pose

$$v_2 = f_2 / \|f_2\| = \begin{pmatrix} 1/\sqrt{6} \\ 1/\sqrt{6} \\ -2/\sqrt{6} \end{pmatrix}.$$

Ceci nous donne la base (v_1, v_2) avec

$$v_1 = \begin{pmatrix} 1/\sqrt{2} \\ -1/\sqrt{2} \\ 0 \end{pmatrix}, v_2 = \begin{pmatrix} 1/\sqrt{6} \\ 1/\sqrt{6} \\ -2/\sqrt{6} \end{pmatrix}.$$

Remarque 5.3.4 Le procédé de Gram-Schmidt permet de calculer la projection orthogonale de tout vecteur $x \in V$ sur un sous-espace W de dimension finie, en calculant une base orthonormée (v_1, \dots, v_k) de W à partir d'une base quelconque e_1, \dots, e_k de W (pour le produit scalaire sur W obtenu par restriction du produit scalaire sur V). On aura alors

$$p_W(x) = \sum_{j=1}^k \langle v_j, x \rangle v_j.$$

Rappelons que $p_W(x)$ est le meilleur approximant de x dans W .

Exemple 5.3.5 Utilisons cette méthode pour construire pour tout $v \in \mathbb{R}^3$ le point le plus proche de v dans W , le plan d'équation $x + y + z = 0$. Nous avons vu qu'une base orthonormée pour ce plan est donnée

par $v_1 = \begin{pmatrix} 1/\sqrt{2} \\ -1/\sqrt{2} \\ 0 \end{pmatrix}$, $v_2 = \begin{pmatrix} 1/\sqrt{6} \\ 1/\sqrt{6} \\ -2/\sqrt{6} \end{pmatrix}$. Soit $v = \begin{pmatrix} x \\ y \\ z \end{pmatrix}$: on a donc

$$\begin{aligned} p_W(v) &= \langle v, v_1 \rangle v_1 + \langle v, v_2 \rangle v_2 \\ &= \frac{(x-y)}{\sqrt{2}} v_1 + \frac{(x+y-2z)}{\sqrt{6}} v_2 \\ &= \begin{pmatrix} (x-y)/2 \\ (-x+y)/2 \\ 0 \end{pmatrix} + \begin{pmatrix} (x+y-2z)/6 \\ (x+y-2z)/6 \\ -2x-2y+4z/6 \end{pmatrix} \\ &= \begin{pmatrix} (2x-y-z)/3 \\ (-x+2y-z)/3 \\ (-x-y+2z)/3 \end{pmatrix}. \end{aligned}$$

Nous pouvons utiliser ces techniques pour résoudre des problèmes de minimisation.

Exemple 5.3.6 Considérons le problème suivant. On veut mesurer une donnée y (pH d'une solution, température) en fonction d'un paramètre x (concentration d'un ion, temps). Considérons les n points (avec $n \geq 2$) $P_1 := (x_1, y_1), \dots, P_n := (x_n, y_n)$ de \mathbb{R}^2 représentant par exemple le résultat de n expérimentations. On suppose que les x_i s sont deux à deux distincts. Supposons que la théorie nous dise que y varie de façon affine en fonction de x . A cause des erreurs de manipulation, de mesure, les n points P_1, \dots, P_n ne sont pas alignés.

Comment trouver la droite de meilleure approximation, c'est-à-dire la droite d'équation $y = ax + b$ telle que les points théoriques $Q_1 := (x_1, ax_1 + b), \dots, Q_n := (x_n, ax_n + b)$ soient le plus proche possible des points expérimentaux P_1, \dots, P_n ?

Plus précisément, comment choisir la droite $y = ax + b$ telle que l'erreur quadratique

$$e := P_1 Q_1^2 + \dots + P_n Q_n^2$$

soit minimale ?

On veut donc trouver $(a, b) \in \mathbb{R}^2$ tels que

$$e := (y_1 - (ax_1 + b))^2 + \dots + (y_n - (ax_n + b))^2$$

soit minimale. Posons

$$\underline{X} = \begin{pmatrix} x_1 \\ \vdots \\ x_n \end{pmatrix}, \underline{Y} = \begin{pmatrix} y_1 \\ \vdots \\ y_n \end{pmatrix} \text{ et } \underline{U} = \begin{pmatrix} 1 \\ \vdots \\ 1 \end{pmatrix}.$$

On a facilement que

$$\underline{Y} - (a\underline{X} + b\underline{U}) = \begin{pmatrix} y_1 - (ax_1 + b) \\ \vdots \\ y_n - (ax_n + b) \end{pmatrix},$$

et donc

$$d = \|\underline{Y} - (a\underline{X} + b\underline{U})\|^2,$$

où nous utilisons la norme associée au produit scalaire canonique sur \mathbb{R}^n . Soit W le sous-espace vectoriel dans \mathbb{R}^n formé de tous les vecteurs de la forme $a\underline{X} + b\underline{U}$ lorsque (a, b) décrit \mathbb{R}^2 . On veut donc minimiser $\|\underline{Y} - w\|$, lorsque w décrit W . D'après les propriétés de la projection orthogonale, le minimum est obtenu pour $w = p_W(\underline{Y})$.

On doit donc calculer $p_W(\underline{Y})$. Les coefficients a et b seront alors donnés par la relation

$$p_W(\underline{Y}) = a\underline{X} + b\underline{U}$$

car $(\underline{X}, \underline{U})$ est une base de W . Posons

$$\bar{x} = \frac{x_1 + \dots + x_n}{n}, \bar{y} = \frac{y_1 + \dots + y_n}{n}.$$

Appliquons l'algorithme de Gram-Schmidt à la base $e_1 = \underline{U}, e_2 = \underline{X}$ de W . On a $v_1 = \underline{U}/\|\underline{U}\| = \frac{1}{\sqrt{n}}\underline{U}$. On a aussi

$$f_2 = e_2 - \langle v_1, e_2 \rangle v_1 = \underline{X} - \bar{x}\underline{U}$$

et $v_2 = f_2/\|f_2\|$. On a alors

$$\begin{aligned} p_W(\underline{Y}) &= \langle v_1, \underline{Y} \rangle v_1 + \langle v_2, \underline{Y} \rangle v_2 \\ &= \bar{y}\underline{U} + \frac{\sum_{i=1}^n x_i y_i - \bar{x}\bar{y}}{\sum_{i=1}^n (x_i - \bar{x})^2} (\underline{X} - \bar{x}\underline{U}). \end{aligned}$$

Remarquons que l'on a

$$\sum_{i=1}^n (x_i y_i - \bar{x}\bar{y}) = \left(\sum_{i=1}^n x_i y_i \right) - n\bar{x}\bar{y} = \sum_{i=1}^n (x_i y_i - \bar{x}\bar{y}).$$

On a donc

$$p_W(\underline{Y}) = \frac{\sum_{i=1}^n (x_i y_i - \bar{x}\bar{y})}{\sum_{i=1}^n (x_i - \bar{x})^2} \underline{X} + \left(\bar{y} - \bar{x} \frac{\sum_{i=1}^n (x_i y_i - \bar{x}\bar{y})}{\sum_{i=1}^n (x_i - \bar{x})^2} \right) \underline{U}.$$

Ainsi, la droite de meilleure approximation est donnée par

$$y = \frac{\sum_{i=1}^n (x_i y_i - \bar{x}\bar{y})}{\sum_{i=1}^n (x_i - \bar{x})^2} (x - \bar{x}) + \bar{y}.$$

$$c'est-à-dire qu'on a $a = \frac{\sum_{i=1}^n (x_i y_i - \bar{x}\bar{y})}{\sum_{i=1}^n (x_i - \bar{x})^2}$ et $b = -\frac{\sum_{i=1}^n (x_i y_i - \bar{x}\bar{y})}{\sum_{i=1}^n (x_i - \bar{x})^2} \bar{x} + \bar{y}$.$$

Exemple 5.3.7 On peut aussi vouloir approximer une fonction continue $f : [a, b] \rightarrow \mathbb{R}$ par une fonction affine $y = \alpha x + \beta$. Dans ce cas, la méthode précédente ne marche plus, puisque l'on doit considérer une infinité de points.

L'idée est de considérer un grand nombre de points sur le graphe de f , dont les abscisses sont régulièrement espacés, $P_1 = (x_1, f(x_1)), \dots, P_n = (x_n, f(x_n))$, avec $x_i = a + \frac{(b-a)i}{n}$, et de considérer la droite

de meilleure approximation pour ces points. Bien sûr, plus n est grand, meilleure est l'approximation. L'entier n étant fixé, on doit donc minimiser

$$d := (f(x_1) - (\alpha x_1 + \beta))^2 + \dots + (f(x_n) - (\alpha x_n + \beta))^2.$$

Ceci revient aussi à minimiser

$$S_n := \frac{1}{n} \sum_{i=1}^n (f(x_i) - (\alpha x_i + \beta))^2, \text{ avec } x_i = a + \frac{(b-a)i}{n}.$$

On voit graphiquement (et on peut démontrer rigoureusement) que S_n converge vers $\int_a^b (f(t) - (\alpha t + \beta))^2 dt$. En particulier, S_n est très proche de cette intégrale lorsque n est suffisamment grand.

Il est alors naturel de définir la droite de meilleure approximation $y = \alpha x + \beta$ comme celle qui minimise l'intégrale $\int_a^b (f(t) - (\alpha t + \beta))^2 dt$.

Ce genre d'intégrale s'interprète souvent comme l'énergie d'un système. Ainsi, le problème de minimisation précédent revient à demander de minimiser cette énergie.

Exemple 5.3.8 *Considérons le problème de minimisation suivant : trouver $a, b \in \mathbb{R}$ tels que $\int_0^{\frac{\pi}{2}} (\cos(x) - a - bx)^2 dx$ soit minimale.*

Soit V l'espace $C^0([0, \frac{\pi}{2}], \mathbb{R})$ avec son produit scalaire

$$\langle \cdot, \cdot \rangle : V \times V \rightarrow \mathbb{R}, (f, g) \mapsto \int_0^{\frac{\pi}{2}} f(x)g(x)dx.$$

On vérifie que $\langle \cdot, \cdot \rangle$ est un produit scalaire sur V . Considérons maintenant le sous-espace W de V défini par

$$W = \text{Vect}(1, x) = \{f | f : x \mapsto a + bx, a, b \in \mathbb{R}\}.$$

Le problème de minimisation se reformule alors ainsi :

Trouver $g \in W$ tel que $\langle \cos(x) - g(x), \cos(x) - g(x) \rangle$ soit minimal. Autrement dit, on cherche $g \in W$ tel que $\|\cos(x) - g(x)\|$ soit minimal. Cela revient à dire que $g = p_W(\cos(x))$. On cherche donc à calculer la projection orthogonale de $\cos(x)$ sur $W = \text{Vect}(1, x)$.

Appliquons le procédé de Gram-Schmidt à la base $e_1 = 1, e_2 = x$ de W . On a $v_1 = e_1 / \|e_1\| = \sqrt{\frac{2}{\pi}}$. On pose

$$f_2 = e_2 - \langle v_1, e_2 \rangle v_1 = (x - \frac{\pi}{4})$$

et on prend $v_2 = (x - \frac{\pi}{4}) / \|x - \frac{\pi}{4}\|$

On a alors

$$g = p_W(\cos(x)) = \frac{\langle 1, \cos(x) \rangle}{\langle 1, 1 \rangle} 1 + \frac{\langle x - \frac{\pi}{4}, \cos(x) \rangle}{\langle x - \frac{\pi}{4}, x - \frac{\pi}{4} \rangle} (x - \frac{\pi}{4}) = ax + b$$

avec $a = (\frac{24}{\pi^2} - \frac{96}{\pi^3})$ et $b = (\frac{-4}{\pi} + \frac{24}{\pi^2})$.

Exemple 5.3.9 *On peut aussi vouloir approximer une fonction $f : [a, b] \rightarrow \mathbb{R}$ par une fonction autre qu'une droite. Par exemple, on peut vouloir approximer f par une fonction g appartenant à un sous-espace vectoriel W de $C^0([a, b], \mathbb{R})$, de façon à ce que l'intégrale*

$$\int_a^b (f(t) - g(t))^2 dt$$

soit minimale, lorsque g décrit W .

Considérons le problème posé dans l'introduction, celui d'approcher une fonction par des sommes trigonométriques. Soit $f : [-L, L] \rightarrow \mathbb{R}$ une fonction que l'on supposera continue : on veut approximer f par une somme finie de fonctions trigonométriques

$$S_n(f) := a_0 + \sum_{k=1}^n a_k \cos\left(\frac{2k\pi x}{L}\right) + b_k \sin\left(\frac{2k\pi x}{L}\right).$$

On veut trouver les coefficients a_k et b_k tels que l'intégrale

$$\int_0^L (f(t) - S_n(t))^2 dt$$

soit minimale.

Soit V l'espace vectoriel $C^0([-L, L], \mathbb{R})$ et W le sous-espace vectoriel de V engendré par

$$\frac{1}{L} \cos\left(\frac{2k\pi x}{L}\right), \sin\left(\frac{2k\pi x}{L}\right), k = 1, \dots, n.$$

Autrement dit, W est l'ensemble de fonctions de la forme

$$g(x) = a_0 + \sum_{k=1}^n a_k \cos\left(\frac{k\pi x}{L}\right) + b_k \sin\left(\frac{k\pi x}{L}\right).$$

Considérons le produit scalaire

$$\langle \cdot, \cdot \rangle : V \rightarrow \mathbb{R}, (f, g) \mapsto \langle f, g \rangle = \int_{-L}^L f(t)g(t)dt.$$

Le raisonnement précédent montre que la meilleure approximation $S_n(f)$ est donnée par $p_W(f)$. Or, on peut vérifier que

$$\frac{1}{\sqrt{2L}}, \sqrt{\frac{1}{L}} \cos\left(\frac{2k\pi x}{L}\right), \sqrt{\frac{1}{L}} \sin\left(\frac{2k\pi x}{L}\right), k = 1, \dots, n$$

fournit une base orthonormée de W – nous reviendrons en détail sur ce calcul dans le dernier chapitre.

La formule pour la projection orthogonale $p_W(f)$ nous donne alors

$$\begin{aligned} p_W(f) &= \langle 1, f \rangle \frac{1}{2L} + \sum_{k=1}^n \frac{1}{L} \langle \cos\left(\frac{k\pi x}{L}\right), f \rangle \cos\left(\frac{k\pi x}{L}\right) + \frac{1}{L} \langle \sin\left(\frac{k\pi x}{L}\right), f \rangle \sin\left(\frac{k\pi x}{L}\right) \\ &= \frac{1}{2L} \int_{-L}^L f(t)dt + \frac{1}{L} \int_{-L}^L f(t) \cos\left(\frac{k\pi t}{L}\right)dt \cos\left(\frac{k\pi x}{L}\right) + \frac{1}{L} \int_{-L}^L f(t) \sin\left(\frac{k\pi t}{L}\right)dt \sin\left(\frac{k\pi x}{L}\right). \end{aligned}$$

Les choix de coefficients a_0, a_k, b_k qui minimisent cet intégrale sont donc donnés par

$$\begin{aligned} a_0 &= \frac{1}{2L} \int_{-L}^L f(t)dt \\ a_k &= \frac{1}{L} \int_{-L}^L f(t) \cos\left(\frac{k\pi t}{L}\right)dt, \\ b_k &= \frac{1}{L} \int_{-L}^L f(t) \sin\left(\frac{k\pi t}{L}\right)dt. \end{aligned}$$

5.4 Diagonalisation orthogonale des matrices symétriques.

Nous présentons ici un théorème sur la diagonalisation des matrices symétriques, que vous avez étudiée en MAT234.

On commence par un lemme.

Lemme 5.4.1 Soit (V, \langle, \rangle) un espace préhilbertien de dimension n , et soit $\mathbf{e} = (e_1, \dots, e_n)$ une base orthonormée. Soit $\mathbf{v} = (v_1, \dots, v_n)$ une autre base de V , et soit P la matrice de passage correspondante (c'est-à-dire la matrice dont les colonnes sont les vecteurs coordonnées de (v_1, \dots, v_n) dans la base (e_1, \dots, e_n)). La base (v_1, \dots, v_n) est orthonormée si et seulement si ${}^t P P = I_n$, c'est-à-dire si $P^{-1} = {}^t P$.

Soient M et N les matrices de la forme \langle, \rangle dans les bases \mathbf{e} et \mathbf{v} . On sait que $N = {}^t P M P$: puisque \mathbf{e} est supposée orthonormée nous avons $M = I_n$ et $N = {}^t P P$. La base \mathbf{v} est orthonormée si et seulement si $N = I_n$ c'est à dire ssi

$$I_n = {}^t P P.$$

CQFD.

Théorème 5.4.2 Soit $B \in M_n(\mathbb{R})$ une matrice symétrique, c'est-à-dire vérifiant ${}^t B = B$. Alors il existe une base de \mathbb{R}^n formée de vecteurs propres de B qui est orthonormée pour le produit scalaire usuel sur \mathbb{R}^n .

La démonstration repose sur le critère suivant.

Lemme 5.4.3 Soient B une matrice carrée $n \times n$. Alors si B est symétrique on a pour tout $\underline{X}, \underline{Y} \in \mathbb{R}^n$

$$\langle \underline{X}, B \underline{Y} \rangle = \langle B \underline{X}, \underline{Y} \rangle.$$

ou \langle, \rangle est le produit scalaire canonique.

On a que

$$\langle \underline{X}, B \underline{Y} \rangle = {}^t \underline{X} B \underline{Y} = {}^t ({}^t B \underline{X}) \underline{Y} = {}^t (B \underline{X}) \underline{Y} = \langle B \underline{X}, \underline{Y} \rangle$$

CQFD.

Remarque 5.4.4 En particulier, si $B \underline{X} = \lambda \underline{X}$ (c'est-à-dire \underline{X} est un vecteur propre de B) et $B \underline{Y} = \mu \underline{Y}$ alors

$$\lambda \langle \underline{X}, \underline{Y} \rangle = \mu \langle \underline{X}, \underline{Y} \rangle$$

et si $\lambda \neq \mu$ alors \underline{X} et \underline{Y} doivent être orthogonaux

Preuve du Théorème 4.28

Vous avez vu dans MAT234 que toute matrice symétrique réelle est diagonalisable sur \mathbb{R} . Soient $\lambda_1, \dots, \lambda_k$ ses valeurs propres distincts avec E_{λ_i} le sous-espace propre associé à λ_i . Nous avons alors

$$\mathbb{R}^n = E_{\lambda_1} \oplus_{\perp} \dots \oplus_{\perp} E_{\lambda_k},$$

et par la remarque 5.4.4, E_{λ_i} est orthogonal à E_{λ_j} si $i \neq j$. Pour tout i soit \mathbf{e}_i une base orthonormée pour E_{λ_i} et soit \mathbf{e} la concaténation $(\mathbf{e}_1, \dots, \mathbf{e}_k)$. Par le Lemme 4.4.5, \mathbf{e} est une base orthonormée pour \mathbb{R}^n composée de vecteurs propres de B .

Ceci se traduit en termes de formes bilinéaires de la façon suivante :

Théorème 5.4.5

Soit (V, \langle, \rangle) un espace préhilbertien de dimension finie, et soit $\varphi : V \times V \rightarrow \mathbb{R}$ une forme bilinéaire symétrique. Alors il existe une base orthonormée pour \langle, \rangle qui est aussi φ -orthogonale.

Soit $\mathbf{e} = (e_1, \dots, e_n)$ orthonormée pour \langle, \rangle , et soit B sa matrice dans cette base. Alors B est une matrice symétrique d'après le Lemme 4.3.3.

D'après le théorème précédent, il existe une base $(\underline{V}_1, \dots, \underline{V}_n)$ de \mathbb{R}^n formée de vecteurs propres de B qui est orthonormée pour le produit scalaire usuel de \mathbb{R}^n .

Si $\underline{V}_j = \begin{pmatrix} v_{1j} \\ \vdots \\ v_{nj} \end{pmatrix}$, posons $v_j = \sum_{i=1}^n v_{ij}e_i$, de telle façon que \underline{V}_j est le vecteur de coordonnées de v_j

dans la base \mathbf{e} .

Nous allons montrer que $\mathbf{v} = (v_1, \dots, v_n)$ est une base de V qui possède les propriétés voulues.

Puisque \mathbf{e} est orthonormée, on a

$$\langle v_i, v_j \rangle = {}^t \underline{V}_i \underline{V}_j$$

d'après le Lemme 4.4.7. Par choix de $\underline{V}_1, \dots, \underline{V}_n$, on en déduit que

$$\langle v_i, v_j \rangle = 0 \text{ si } i \neq j$$

et

$$\langle v_i, v_i \rangle = 1 \text{ pour tout } i.$$

Il reste à voir que \mathbf{v} est φ -orthogonale. Soit P la matrice de passage de \mathbf{v} à \mathbf{e} . La matrice N qui représente φ dans la base \mathbf{v} est donc ${}^t P B P$. Or \mathbf{v} étant orthonormée, on a ${}^t P P = I_n$. On a ainsi $N = P^{-1} B P$.

Mais \mathbf{v} étant formée de vecteurs propres de B , nous avons que $P^{-1} B P$ est diagonale. N est donc diagonale, ce qui revient à dire que \mathbf{v} est φ -orthogonale. Cette démonstration nous donne en plus que nous pouvons construire une telle base orthonormée et φ -orthogonale en prenant des vecteurs propres de B .

Méthode pratique pour trouver une base de vecteurs orthonormée et φ -orthogonale.

(1) Soit M une matrice symétrique réelle.

- Pour chaque valeur propre $\lambda \in \mathbb{R}$ de M , on calcule une base de E_λ , puis on applique l'algorithme de Gram-Schmidt pour obtenir une base orthonormée de E_λ .

- On recolle les bases orthonormées précédentes pour obtenir une base $(\underline{V}_1, \dots, \underline{V}_n)$ de \mathbb{R}^n formée de vecteurs propres de M , orthonormée **pour le produit scalaire usuel** sur \mathbb{R}^n .

(2) Soit $\varphi : V \times V \rightarrow \mathbb{R}$ une forme bilinéaire symétrique. On se fixe une base \mathbf{e} de V orthonormée pour \langle, \rangle .

Soit M la matrice de φ dans la base \mathbf{e} . M est une matrice symétrique. On applique la méthode précédente pour obtenir une base $(\underline{V}_1, \dots, \underline{V}_n)$ orthonormée de \mathbb{R}^n formée de vecteurs propres de M . On prend alors v_i l'unique vecteur dans V qui admet pour vecteur coordonnées dans la base \mathbf{e} le vecteur \underline{V}_i . La base (v_1, \dots, v_n) est alors la base recherchée.

Exemples 5.4.6

(1) Soit $B = \begin{pmatrix} 3 & 4 \\ 4 & -3 \end{pmatrix}$.

On vérifie que les valeurs propres sont 5 et -5, et que

$$E_5 = \text{Vect}\left\{\begin{pmatrix} 2 \\ 1 \end{pmatrix}\right\}, E_{-5} = \text{Vect}\left\{\begin{pmatrix} 1 \\ -2 \end{pmatrix}\right\}.$$

Une base orthonormée pour E_5 est donc

$$\frac{1}{\sqrt{5}} \begin{pmatrix} 2 \\ 1 \end{pmatrix},$$

et une base orthonormée pour E_{-5} est donc

$$\frac{1}{\sqrt{5}} \begin{pmatrix} 1 \\ -2 \end{pmatrix}.$$

La base recherchée est donc donnée par

$$\left(\frac{1}{\sqrt{5}} \begin{pmatrix} 2 \\ 1 \end{pmatrix}, \frac{1}{\sqrt{5}} \begin{pmatrix} 1 \\ -2 \end{pmatrix} \right).$$

(2) Munissons \mathbb{R}^3 de son produit scalaire usuel, et soit

$$\varphi : \mathbb{R}^3 \times \mathbb{R}^3 \rightarrow \mathbb{R}, \left(\begin{pmatrix} x_1 \\ \vdots \\ x_3 \end{pmatrix}, \begin{pmatrix} y_1 \\ \vdots \\ y_3 \end{pmatrix} \right) \mapsto \sum_{i,j \leq 3} x_i y_j.$$

Soit \mathbf{e} la base canonique de \mathbb{R}^3 . C'est une base orthonormée pour le produit scalaire usuel. La matrice M de φ dans la base canonique est alors

$$M = \begin{pmatrix} 1 & 1 & 1 \\ 1 & 1 & 1 \\ 1 & 1 & 1 \end{pmatrix}.$$

On vérifie que les valeurs propres sont 3 et 0, que E_3 admet comme base la famille $\left(\begin{pmatrix} 1 \\ 1 \\ 1 \end{pmatrix} \right)$ et que E_0 admet comme base la famille

$$\left(\begin{pmatrix} 1 \\ -1 \\ 0 \end{pmatrix}, \begin{pmatrix} 1 \\ 0 \\ -1 \end{pmatrix} \right).$$

Une base orthonormée pour E_1 est donc

$$\frac{1}{\sqrt{3}} \begin{pmatrix} 1 \\ 1 \\ 1 \end{pmatrix}.$$

Pour trouver une base orthonormée de E_0 , on applique Gram-Schmidt. On pose $v_1 = \frac{1}{\sqrt{2}} \begin{pmatrix} 1 \\ -1 \\ 0 \end{pmatrix}$.

Ensuite on pose

$$f_2 = \begin{pmatrix} 1 \\ 0 \\ -1 \end{pmatrix} - \frac{1}{2} \begin{pmatrix} 1 \\ -1 \\ 0 \end{pmatrix} = \begin{pmatrix} 1/2 \\ 1/2 \\ -1 \end{pmatrix}.$$

Enfin on pose $v_2 = f_2 / \|f_2\| = \sqrt{\frac{2}{3}} \begin{pmatrix} 1/2 \\ 1/2 \\ -1 \end{pmatrix}$

Une base orthonormée pour E_0 est donc

$$\left(\frac{1}{\sqrt{2}} \begin{pmatrix} 1 \\ -1 \\ 0 \end{pmatrix}, \sqrt{\frac{2}{3}} \begin{pmatrix} \frac{1}{2} \\ \frac{1}{2} \\ -1 \end{pmatrix} \right).$$

La base recherchée est donc donnée par

$$v_1 = \frac{1}{\sqrt{3}} \begin{pmatrix} 1 \\ 1 \\ 1 \end{pmatrix}, v_2 = \frac{1}{\sqrt{2}} \begin{pmatrix} 1 \\ -1 \\ 0 \end{pmatrix}, v_3 = \sqrt{\frac{2}{3}} \begin{pmatrix} \frac{1}{2} \\ \frac{1}{2} \\ -1 \end{pmatrix}.$$

Si $x = x'_1 v_1 + x'_2 v_2 + x'_3 v_3$ et $y = x y'_1 v_1 + y'_2 v_2 + y'_3 v_3$, on a

$$b(x, y) = 3x'_3 y'_3.$$

5.5 Matrices orthogonales et unitaires.

Nous avons vu ci-dessus que les matrices réelles M telles que ${}^tMM = I_n$, ou les matrices complexes telles que ${}^t\overline{M}M = I_n$, sont très importantes puisqu'elles encodent des changements de bases orthonormées. Nous allons maintenant étudier ces matrices en détail.

Définition 5.5.1 Soit M une matrice réelle de taille $n \times n$. Les conditions suivantes sont équivalentes :

1. ${}^tMM = I_n$;
2. pour tous $v, w \in \mathbb{R}^n$ nous avons $\langle Mv, Mw \rangle = \langle v, w \rangle$, où $\langle \cdot, \cdot \rangle$ est le produit scalaire canonique.
3. pour tout $v \in \mathbb{R}^n$ nous avons $\|Mv\| = \|v\|$, où $\|v\|$ est la norme de v pour le produit scalaire canonique.

On dit qu'une matrice qui satisfait à ces conditions est orthogonale.

Si ${}^tMM = I_n$ alors

$$\langle Mv, Mw \rangle = {}^t(Mv)Mw = {}^t v {}^tMMw = {}^t v I_n w = {}^t v w = \langle v, w \rangle.$$

Donc (1) implique (2). (2) implique (3) en prenant $v = w$ et (3) implique (2) par la formule de polarisation. Reste à montrer que (2) implique (1). Si $\langle Mv, Mw \rangle = \langle v, w \rangle$ alors pour tout $v, w \in \mathbb{R}^n$ nous avons

$${}^t v {}^tMMw = {}^t v I_n w = {}^t v w = \langle v, w \rangle$$

et donc ${}^tMM = I_n$. On peut bien sûr faire la même chose pour des matrices complexes.

Définition 5.5.2 Soit M une matrice complexe de taille $n \times n$. Les conditions suivantes sont équivalentes :

1. ${}^t\overline{M}M = I_n$;
2. pour tous $v, w \in \mathbb{C}^n$ nous avons $h(Mv, Mw) = h(v, w)$ où h est le produit hermitien canonique ;
3. pour tout $v \in \mathbb{C}^n$ nous avons $\|Mv\| = \|v\|$, où $\|v\|$ est la norme associée au produit hermitien canonique.

On dit qu'une matrice qui satisfait à ces conditions est unitaire.

Si ${}^t\overline{M}M = I_n$ alors

$$h(Mv, Mw) = {}^t(\overline{Mv})Mw = {}^t\overline{v} {}^t\overline{M}Mw = {}^t\overline{v} I_n w = {}^t\overline{v} w = h(v, w).$$

Donc (1) implique (2). (2) implique (3) en prenant $v = w$ et (3) implique (2) par la formule de polarisation. Reste à montrer que (2) implique (1). Si $h(Mv, Mw) = h(v, w)$ alors pour tout $v, w \in \mathbb{C}^n$ nous avons que

$${}^t\overline{v} {}^t\overline{M}Mw = {}^t\overline{v} I_n w$$

et donc ${}^t\overline{M}M = I_n$. Nous finissons cette section avec une étude des matrices orthogonales de taille 2×2 . Nous allons démontrer le théorème suivant :

Proposition 5.5.3 Soit M une matrice 2×2 orthogonale. Alors l'application $\mathbb{R}^2 \mapsto \mathbb{R}^2$ donnée par $v \mapsto Mv$ est

1. une rotation autour de l'origine
2. * ou
3. une symétrie par rapport à une droite passant par l'origine.

Soit $M = \begin{pmatrix} a & b \\ c & d \end{pmatrix}$ une matrice orthogonale. On a alors

$$\|M \begin{pmatrix} 1 \\ 0 \end{pmatrix}\| = \left\| \begin{pmatrix} a \\ c \end{pmatrix} \right\| = 1$$

et, en utilisant des coordonnées polaires, il existe un θ tel que $\begin{pmatrix} a \\ c \end{pmatrix} = \begin{pmatrix} \cos(\theta) \\ \sin(\theta) \end{pmatrix}$. De même $\begin{pmatrix} b \\ d \end{pmatrix} = \begin{pmatrix} \cos(\phi) \\ \sin(\phi) \end{pmatrix}$ et on peut écrire

$$M = \begin{pmatrix} \cos \theta & \cos \phi \\ \sin \theta & \sin \phi \end{pmatrix}.$$

On a alors

$$\begin{aligned} {}^tMM &= \begin{pmatrix} \cos^2 \theta + \sin^2 \theta & \cos \theta \cos \phi + \sin \theta \sin \phi \\ \cos \theta \cos \phi + \sin \theta \sin \phi & \cos^2 \phi + \sin^2 \phi \end{pmatrix} \\ &= \begin{pmatrix} 1 & \cos(\theta - \phi) \\ \cos(\theta - \phi) & 1 \end{pmatrix} \end{aligned}$$

et nous avons donc M orthogonale si et seulement si $\cos(\theta - \phi) = 0$, c'est-à-dire si et seulement si

$$\phi = \theta + \pi/2 \text{ ou } \phi = \theta - \pi/2.$$

Dans le premier cas nous avons

$$M = \begin{pmatrix} \cos \theta & -\sin \theta \\ \sin \theta & \cos \theta \end{pmatrix}$$

et on reconnaît la matrice d'une rotation d'angle θ autour de l'origine. Dans le deuxième cas on a

$$M = \begin{pmatrix} \cos \theta & +\sin \theta \\ \sin \theta & -\cos \theta \end{pmatrix}.$$

Après calcul le polynôme caractéristique de M est $\lambda^2 - 1$, qui a pour solutions 1 et -1 . Après calcul des vecteurs propres par pivot de Gauss, on voit que

$$e_1 = \begin{pmatrix} \cos \theta/2 \\ \sin \theta/2 \end{pmatrix}, e_2 = \begin{pmatrix} -\sin \theta/2 \\ \cos \theta/2 \end{pmatrix}$$

sont des vecteurs propres de M de valeur propre 1 et -1 respectivement. Autrement dit, on a $Me_1 = e_1$ et $Me_2 = -e_2$. Puisque $\langle e_1, e_2 \rangle = 0$, ou autrement dit les vecteurs e_1 et e_2 sont perpendiculaires, M représente une symétrie par rapport à la droite engendrée par e_1 .

Chapitre 6

Séries de Fourier.

Nous allons maintenant revenir sur la question posée en début de semestre. Rappelons que nous cherchions à résoudre l'équation de la chaleur :

$$\frac{\partial T}{\partial t} = D \frac{\partial^2 T}{\partial x^2}$$

sur le domaine $\{(x, t) | x \in [0, L], t \geq 0\}$ en respectant les conditions initiales

$$T(x, 0) = \phi(x)$$

(ou ϕ est une fonction donnée) et les conditions aux bords

$$\frac{\partial T}{\partial x}(0, t) = \frac{\partial T}{\partial x}(L, t) = 0$$

pour tout $t > 0$. Nous avons remarqué que lorsque la condition initiale ϕ était une somme finie de cosinus

$$\phi = a_0 + \sum_{k=1}^n a_k \cos\left(\frac{k\pi x}{L}\right)$$

cette équation possède une solution

$$T(x, t) = a_0 + \sum_{k=1}^n a_k \cos\left(\frac{k\pi x}{L}\right) e^{-Dk^2\pi^2 x/L^2}.$$

Nous allons maintenant chercher à résoudre cette équation pour une condition initiale ϕ quelconque en approchant ϕ par des sommes finies de la forme

$$\phi = a_0 + \sum_{k=1}^n a_k \cos\left(\frac{k\pi x}{L}\right).$$

Pour l'équation des ondes, on cherche plutôt à approcher ϕ par une somme de sinus. On va présenter une méthode pour approcher une fonction quelconque ϕ par une somme trigonométrique - c'est à dire, une fonction g de la forme

$$g(x) = a_0 + \sum_{k=1}^n a_k \cos\left(\frac{k\pi x}{L}\right) + b_k \sin\left(\frac{k\pi x}{L}\right)$$

Le cas de l'équation de la chaleur correspond au cas où tous les b_k sont nuls, celui de l'équation des ondes au cas où les a_k sont nuls.

6.1 Approximants de Fourier, coefficients de Fourier et séries de Fourier : définitions et exemples.

Dans ce paragraphe, nous allons appliquer la méthode de la projection orthogonale pour approcher une fonction par une somme trigonométrique.

Supposons donnée une fonction f , continue, réelle et définie sur une intervalle $[-L, L]$. On cherche à approcher f par une somme trigonométrique

$$a_0 + \sum_{k=1}^n a_k \cos\left(\frac{k\pi x}{L}\right) + b_k \sin\left(\frac{k\pi x}{L}\right).$$

Pour chaque n , nous allons considérer la fonction $S_n^{\mathbb{R}}(f)$, qui sera le meilleur approximant de f de la forme

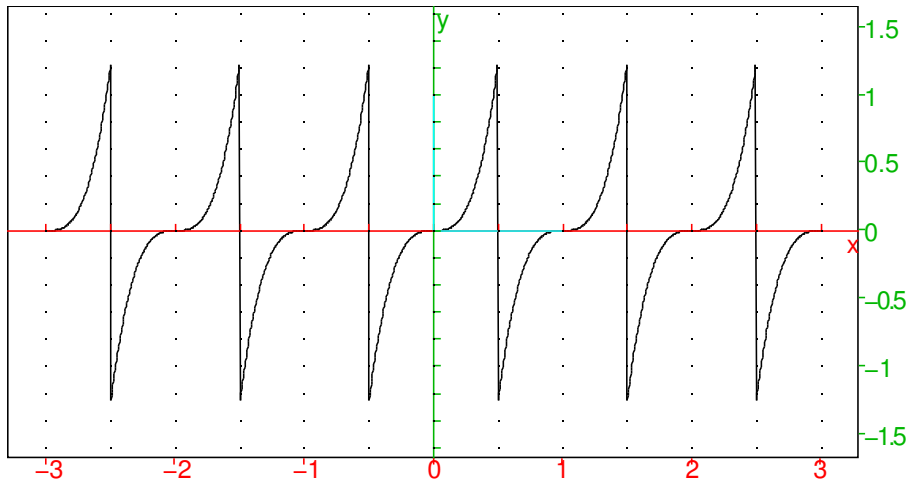
$$a_0 + \sum_{k=1}^n a_k \cos\left(\frac{k\pi x}{L}\right) + b_n \sin\left(\frac{k\pi x}{L}\right)$$

par rapport à la distance définie par le produit scalaire

$$\langle f, g \rangle = \int_{-L}^L f(x)g(x)dx.$$

Dans ce chapitre nous aurons souvent besoin de travailler avec des fonctions qui ne sont pas continues sur $[-L, L]$ mais presque, au sens où elles peuvent avoir un nombre fini de “sauts”, par exemple

```
f(x) := 10 * (x - floor(x + 1/2)) ^ 3;
plot(f(x), x, -3, 3, xstep=1/256);
```



Plus formellement :

Définition 6.1.1 Soit $i \geq 0$ un entier.

Une fonction $f : [a, b] \rightarrow \mathbb{C}$ est dite C^i par morceaux s'il existe

$$a = a_0 < a_1 < \dots < a_p = b$$

tels que

1. la fonction f est de classe C^i sur chaque intervalle $]a_{k-1}, a_k[$.

6.1. APPROXIMANTS DE FOURIER, COEFFICIENTS DE FOURIER ET SÉRIES DE FOURIER : DÉFINITIONS ET EXEMPLES

2. Pour tout $m = 0 \dots i$ et tout $j = 0 \dots p$ les limites à gauche et à droite

$$\lim_{x \rightarrow a_j^-} f^{(m)}(x)$$

et

$$\lim_{x \rightarrow a_j^+} f^{(m)}(x)$$

existent et sont finies.

Une fonction $f : \mathbb{R} \rightarrow \mathbb{C}$ est dite C^i par morceaux si elle est C^i par morceaux sur $[-n, n]$ pour tout n (i.e. les points de discontinuité sont en nombre fini sur un intervalle borné)

Une fonction $f : [a, b] \rightarrow \mathbb{C}$ est C^∞ par morceaux si elle est C^i par morceaux pour tout $i \geq 0$. On note l'espace vectoriel de toute les fonctions réelles (resp. complexes) C^i par morceaux sur une intervalle $[a, b]$ par $C_{\text{mor}}^i([a, b], \mathbb{R})$ (resp. $C_{\text{mor}}^i([a, b], \mathbb{C})$.)

Remarque

Si $f : [a, b] \rightarrow \mathbb{C}$ est continue par morceaux, alors $\int_a^b f(x)dx$ est bien définie. Si $a = a_0, a_1, \dots, a_p = b$ est la subdivision correspondante alors on a

$$\int_a^b f(x)dx = \sum_{j=0}^{p-1} \int_{a_j}^{a_{j+1}} f(x)dx.$$

En particulier, on peut donc définir sur $C_{\text{mor}}^i([a, b], \mathbb{R})$ le produit scalaire

$$\langle f, g \rangle = \int_{-L}^L f(x)g(x)dx$$

Soit maintenant $f : [-L, L] \rightarrow \mathbb{R}$ une fonction continue par morceaux. On considère $C_{\text{mor}}^0([-L, L], \mathbb{R})$ avec son produit scalaire :

$$\langle f, g \rangle = \int_{-L}^L f(x)g(x)dx.$$

Pour tout n , notons W_n l'ensemble des fonctions $g(x)$ de la forme

$$g(x) = a_0 + \sum_{k=1}^n a_k \cos\left(\frac{k\pi x}{L}\right) + b_k \sin\left(\frac{k\pi x}{L}\right).$$

(On vérifie facilement que W_n est un sous-espace vectoriel de $C_{\text{mor}}^0([-L, L], \mathbb{R})$.)

Définition 6.1.2 Le n -ième approximant trigonométrique de Fourier de f , noté $S_n^{\mathbb{R}}(f)$, est alors défini par la projection de f sur W_n

$$S_n^{\mathbb{R}}(f) = p_{W_n}(f).$$

Autrement dit, $S_n^{\mathbb{R}}(f)$ est la fonction dans W_n qui minimise la distance euclidienne

$$d(S_n^{\mathbb{R}}(f), f) = \sqrt{\int_{-L}^L (S_n^{\mathbb{R}}(f)(x) - f(x))^2 dx}$$

Nous allons maintenant procéder au calcul explicite des approximants de Fourier. Il faut pour cela commencer par identifier une base orthonormée de W_n . Par définition de cet espace,

$$B = \left\{ 1, \cos\left(\frac{\pi x}{L}\right), \sin\left(\frac{\pi x}{L}\right), \cos\left(\frac{2\pi x}{L}\right), \dots, \cos\left(\frac{n\pi x}{L}\right), \sin\left(\frac{n\pi x}{L}\right) \right\}$$

est une famille génératrice de W_n .

Lemme 6.1.3 *La famille B est orthogonale pour W_n par rapport au produit scalaire $\langle f, g \rangle = \int_{-L}^L f(x)g(x)dx$. C'est donc une base orthogonale de W_n .*

Il nous faut démontrer que

1. $\int_{-L}^L 1 \sin(\frac{k\pi x}{L})dx = \int_{-L}^L 1 \cos(\frac{k\pi x}{L})dx = 0$ pour tout $k \neq 0$
2. $\int_{-L}^L \cos(\frac{j\pi x}{L}) \sin(\frac{k\pi x}{L})dx = 0$ pour tout $k, j > 0$.
3. $\int_{-L}^L \sin(\frac{j\pi x}{L}) \sin(\frac{k\pi x}{L})dx = \int_{-L}^L \cos(\frac{j\pi x}{L}) \cos(\frac{k\pi x}{L})dx = 0$ pour $k \neq j > 0$.

On a que

$$\int_{-L}^L \sin(\frac{k\pi x}{L})dx = \left[\frac{-L}{k\pi} \cos(\frac{k\pi x}{L}) \right]_{-L}^L = 0$$

$$\int_{-L}^L \cos(\frac{k\pi x}{L})dx = \left[\frac{L}{k\pi} \sin(\frac{k\pi x}{L}) \right]_{-L}^L = 0$$

donc (1) est vrai. Ensuite,

$$\int_{-L}^L \cos(\frac{j\pi x}{L}) \sin(\frac{k\pi x}{L})dx = \int_{-L}^L \frac{1}{2} (\sin(\frac{(k+j)\pi x}{L}) + \sin(\frac{(k-j)\pi x}{L}))dx.$$

Par (1) cet intégrale est nulle si $k \neq j$. Lorsque $k = j$ on a que

$$\int_{-L}^L \cos(\frac{k\pi x}{L}) \sin(\frac{k\pi x}{L})dx = \int_{-L}^L \frac{1}{2} \sin(2\frac{k\pi x}{L})dx = 0.$$

Dans tous les cas, (2) est vérifié. Enfin, pour $k \neq j$ et $k, j > 0$ on a que

$$\int_{-L}^L \sin(\frac{j\pi x}{L}) \sin(\frac{k\pi x}{L})dx = \int_{-L}^L \frac{1}{2} (-\cos(\frac{(k+j)\pi x}{L}) + \cos(\frac{(k-j)\pi x}{L}))dx = 0$$

par (1). De même

$$\int_{-L}^L \cos(\frac{j\pi x}{L}) \cos(\frac{k\pi x}{L})dx = \int_{-L}^L \frac{1}{2} (\cos(\frac{(k+j)\pi x}{L}) + \cos(\frac{(k-j)\pi x}{L}))dx = 0$$

par (1). La condition (3) est donc vérifiée. Ceci termine la démonstration de l'orthogonalité de B .

Pour obtenir une base orthonormée pour $W_n(\mathbb{R})$ il suffira donc de normaliser la base

$$B = (1, \cos(\frac{\pi x}{L}), \sin(\frac{\pi x}{L}), \cos(\frac{2\pi x}{L}), \dots, \cos(\frac{n\pi x}{L}), \sin(\frac{n\pi x}{L}))$$

On a que

$$\|1\| = \sqrt{\int_{-L}^L 1^2 dx} = \sqrt{2L}$$

et

$$\begin{aligned} \|\cos \frac{k\pi x}{L}\| &= \sqrt{\int_{-L}^L \cos^2(\frac{k\pi x}{L})dx} \\ &= \sqrt{\int_{-L}^L \frac{1}{2} (1 + \cos(2\frac{k\pi x}{L}))dx} = \sqrt{L}. \end{aligned}$$

Par un calcul similaire

$$\|\sin(\frac{k\pi x}{L})\| = \sqrt{L}.$$

6.1. APPROXIMANTS DE FOURIER, COEFFICIENTS DE FOURIER ET SÉRIES DE FOURIER : DÉFINITIONS ET EXEMPLES

Nous avons donc une base orthonormée \tilde{B} de W_n donnée par

$$\tilde{B} = \left(\frac{1}{\sqrt{2L}}, \frac{\cos(\frac{\pi x}{L})}{\sqrt{L}}, \frac{\sin(\frac{\pi x}{L})}{\sqrt{L}}, \frac{\cos(\frac{2\pi x}{L})}{\sqrt{L}}, \dots, \frac{\cos(\frac{n\pi x}{L})}{\sqrt{L}}, \frac{\sin(\frac{n\pi x}{L})}{\sqrt{L}} \right)$$

Nous pouvons donc calculer l'approximant de Fourier trigonométriques $S_n^{\mathbb{R}}(f)$ en utilisant le formule de la projection orthogonale. Ce formule nous dit que

$$\begin{aligned} S_n^{\mathbb{R}}(f) &= \left\langle \frac{1}{\sqrt{2L}}, f \right\rangle \frac{1}{\sqrt{2L}} + \sum_{k=1}^n \left\langle \frac{\cos(\frac{k\pi x}{L})}{\sqrt{L}}, f \right\rangle \frac{\cos(\frac{k\pi x}{L})}{\sqrt{L}} + \left\langle \frac{\sin(\frac{k\pi x}{L})}{\sqrt{L}}, f \right\rangle \frac{\sin(\frac{k\pi x}{L})}{\sqrt{L}} \\ &= \frac{1}{2L} \int_{-L}^L f(x) dx + \sum_{k=1}^n a_k(f) \cos\left(\frac{k\pi x}{L}\right) + b_k(f) \sin\left(\frac{k\pi x}{L}\right) \end{aligned}$$

ou les coefficients $a_k(f)$ et $b_k(f)$ sont définis par

$$a_k(f) = \frac{1}{L} \int_{-L}^L \cos\left(\frac{k\pi x}{L}\right) f(x) dx$$

et

$$b_k(f) = \frac{1}{L} \int_{-L}^L \sin\left(\frac{k\pi x}{L}\right) f(x) dx.$$

Définition 6.1.4 Soit f une fonction réelle ou complexe continue par morceaux définie sur une intervalle $[-L, L]$. Les coefficients de Fourier trigonométriques de f sont les nombres $a_0(f)$, $a_k(f)$, $b_k(f)$ ($k > 0$) définis par

$$a_0(f) = \frac{1}{2L} \int_{-L}^L f dx$$

$$a_k(f) = \frac{1}{L} \int_{-L}^L \cos\left(\frac{k\pi x}{L}\right) f(x) dx$$

$$b_k(f) = \frac{1}{L} \int_{-L}^L \sin\left(\frac{k\pi x}{L}\right) f(x) dx.$$

Le n -ième approximation de Fourier trigonométrique est alors donnée par

$$S_n^{\mathbb{R}}(f) = a_0(f) + \sum_k a_k(f) \cos\left(\frac{k\pi x}{L}\right) + b_k \sin\left(\frac{k\pi x}{L}\right).$$

Remarque 6.1.5 il résulte des formules d'Euler :

$$e^{i \frac{k\pi x}{L}} = \cos\left(\frac{k\pi x}{L}\right) + i \sin\left(\frac{k\pi x}{L}\right)$$

$$\cos\left(\frac{k\pi x}{L}\right) = \frac{1}{2} (e^{i \frac{k\pi x}{L}} + e^{-i \frac{k\pi x}{L}})$$

$$\sin\left(\frac{k\pi x}{L}\right) = \frac{1}{2i} (e^{i \frac{k\pi x}{L}} - e^{-i \frac{k\pi x}{L}})$$

qu'approcher f par une somme de fonctions trigonométriques équivaut à l'approcher par une somme d'exponentielles complexes. Plus précisément, considérons la somme trigonométrique

$$a_0 + \sum_{k=1}^n \left(a_k \cos\left(\frac{k\pi x}{L}\right) + b_k \sin\left(\frac{k\pi x}{L}\right) \right).$$

Si on pose $c_0 = a_0$ et pour tout $k > 0$

$$c_k = \frac{1}{2}(a_k - ib_k), \quad c_{-k} = \frac{1}{2}(a_k + ib_k)$$

alors on a

$$a_0 + \sum_{k=1}^n \left(a_k \cos\left(\frac{k\pi x}{L}\right) + b_k \sin\left(\frac{k\pi x}{L}\right) \right) = \sum_{k=-n}^n c_k e^{ik\pi x/L}.$$

Pour des fonctions complexes, cette version des approximants de Fourier est particulièrement utile.

Définition 6.1.6 Soit $f : [-L, L] \rightarrow \mathbb{C}$ une fonction continue par morceaux. On définit le n -ième approxi-
mant exponentiel de Fourier de f , noté $S_n^{\mathbb{C}}(f)$, par

$$S_n^{\mathbb{C}}(f) = \sum_{k=-n}^n c_k e^{i\frac{k\pi x}{L}}$$

ou c_k sont les approximants exponentiels de f définis par

$$c_k = \frac{1}{2}(a_k - ib_k), \quad c_{-k} = \frac{1}{2}(a_k + ib_k)$$

ou a_k et b_k sont les coefficients trigonométriques de Fourier de f .

En utilisant le formule pour les a_k et b_k s ci-dessus, on obtient la formule suivante.

Définition 6.1.7 Soit f une fonction réelle ou complexe continue par morceaux et définie sur $[-L, L]$. Les
coefficients de Fourier exponentiels de f sont les nombres $c_k(f)$ définis par

$$c_k(f) = \frac{1}{2L} \int_{-L}^L e^{-i\frac{k\pi x}{L}} f(x) dx$$

Nous avons donc que l'approximant de Fourier trigonométrique est

$$S_n^{\mathbb{R}}(f) = a_0(f) + \sum_{k=1}^n a_k(f) \cos\left(\frac{k\pi x}{L}\right) + b_k(f) \sin\left(\frac{k\pi x}{L}\right).$$

et que l'approximant exponentiel est

$$S_n^{\mathbb{C}}(f) = \sum_{k=-n}^n c_k(f) e^{i\frac{k\pi x}{L}}.$$

Il est pratique de mettre toutes ces termes ensemble dans une somme infinie, i.e. une série, même s'il n'est pas encore très clair qu'on peut donner un sens à cette somme formelle.

Définition 6.1.8 Soit f une fonction continue par morceaux sur une intervalle $[-L, L]$. La **série de Fourier trigonométrique** de f est alors la somme infinie

$$S^{\mathbb{R}}(f) = \frac{a_0(f)}{2} + \sum_{k=1}^{\infty} a_k(f) \cos\left(\frac{k\pi x}{L}\right) + b_k(f) \sin\left(\frac{k\pi x}{L}\right).$$

La **série de Fourier exponentielle** de f est la somme infinie

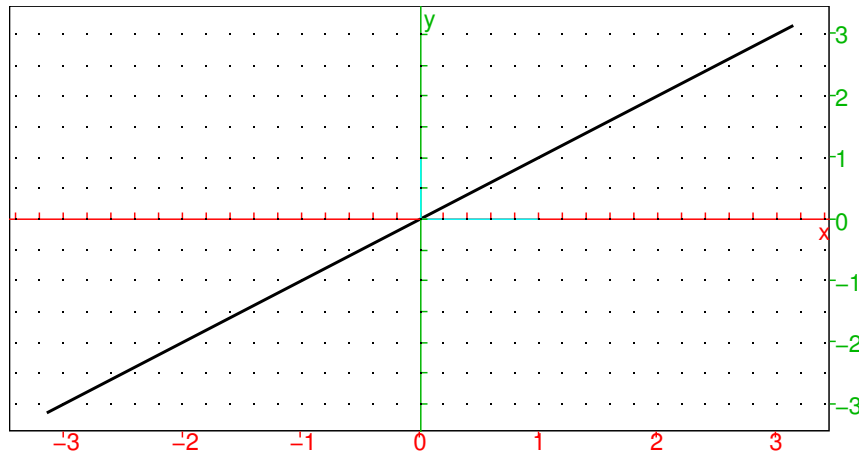
$$S^{\mathbb{C}}(f) = \sum_{k=-\infty}^{\infty} c_k(f) e^{\frac{2ik\pi}{L}x}.$$

Exemples

6.1. APPROXIMANTS DE FOURIER, COEFFICIENTS DE FOURIER ET SÉRIES DE FOURIER : DÉFINITIONS ET EXEMPLES

1. On considère la fonction $f(x) = x$ sur l'intervalle $[-\pi, \pi]$

```
plot(x, x=-pi..pi, display=line_width_2);
```



Nous avons alors

$$a_0(f) = \frac{1}{2\pi} \int_{-\pi}^{\pi} x dx = 0.$$

Ensuite, pour tout $k > 0$, en intégrant par parties, on voit que

$$\begin{aligned} a_k(f) &= \frac{1}{\pi} \int_{-\pi}^{\pi} x \cos(kx) dx \\ &= \frac{1}{\pi} \left[\frac{x}{k} \sin(kx) \right]_{-\pi}^{\pi} - \frac{1}{k\pi} \int_{-\pi}^{\pi} \sin(kx) dx \\ &= -\frac{1}{k\pi} \int_{-\pi}^{\pi} \sin(kx) dx = 0. \end{aligned}$$

On pouvait d'ailleurs éviter le calcul en appliquant un argument de symétrie (voir plus bas).

```
fourier_an(x, x, 2*pi, k, -pi)
```

0

De même, l'intégration par parties montre que

$$\begin{aligned} b_k(f) &= \frac{1}{\pi} \int_{-\pi}^{\pi} x \sin(kx) dx \\ &= \left[-\frac{x}{k\pi} \cos(kx) \right]_{-\pi}^{\pi} + \frac{1}{k\pi} \int_{-\pi}^{\pi} \cos(kx) dx \\ &= -\frac{2(-1)^k}{k} + \frac{1}{k\pi} \int_{-\pi}^{\pi} \cos(kx) dx = \frac{2(-1)^{k+1}}{k}. \end{aligned}$$

```
assume(k, integer) ;; fourier_bn(x, x, 2*pi, k, -pi)
```

$$\text{Done, } -\frac{2 \cdot (-1)^k}{k}$$

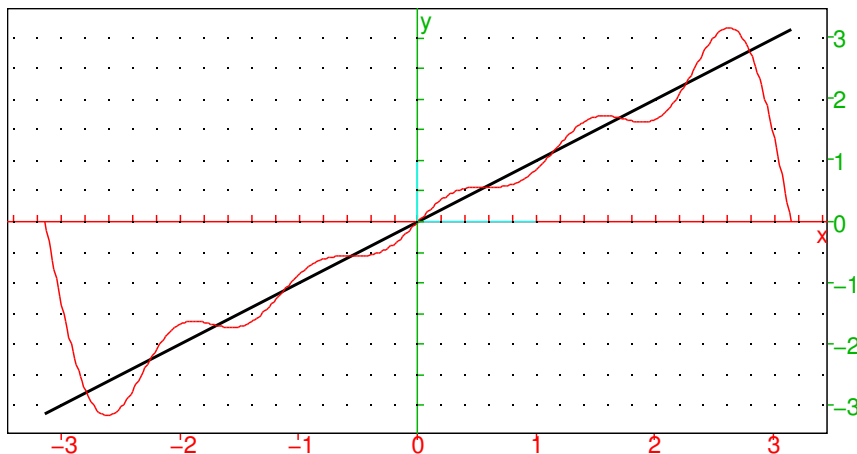
La série de Fourier trigonométrique de f est donc la somme infinie

$$S^{\mathbb{R}}(f) = \sum_{k=1}^{\infty} \frac{2(-1)^{k+1}}{k} \sin(kx).$$

(Attention, ce n'est pas une série absolument convergente, par exemple en $x = \pi/2$).

Représentation graphique de f et de la somme partielle jusqu'à un l'ordre 5 de sa série de Fourier (vous pouvez modifier 5 en un autre ordre et tester dans la version HTML de ce document)

```
S(n) := sum(2 * (-1)^(k+1) / k * sin(k*x), k, 1, n); plot
([x, S(5)], x=-pi..pi, color=[line_width_2, red])
```



Calculons maintenant la série de Fourier exponentielle de f . L'intégration par parties nous donne que

$$\begin{aligned} c_k(f) &= \frac{1}{2\pi} \int_{-\pi}^{\pi} e^{-ikx} x dx = \frac{1}{2\pi} \left[\frac{x}{-ik} e^{-ikx} \right]_{-\pi}^{\pi} + \int_{-\pi}^{\pi} \frac{1}{2ik\pi} e^{-ikx} dx \\ &= \frac{(-1)^k}{-2ik} = \frac{i(-1)^k}{2k}. \end{aligned}$$

Vérification assume(k, integer) : ; fourier_cn(x, x, 2*pi, k, -pi)

$$\text{Done, } \frac{i \cdot (-1)^k}{k}$$

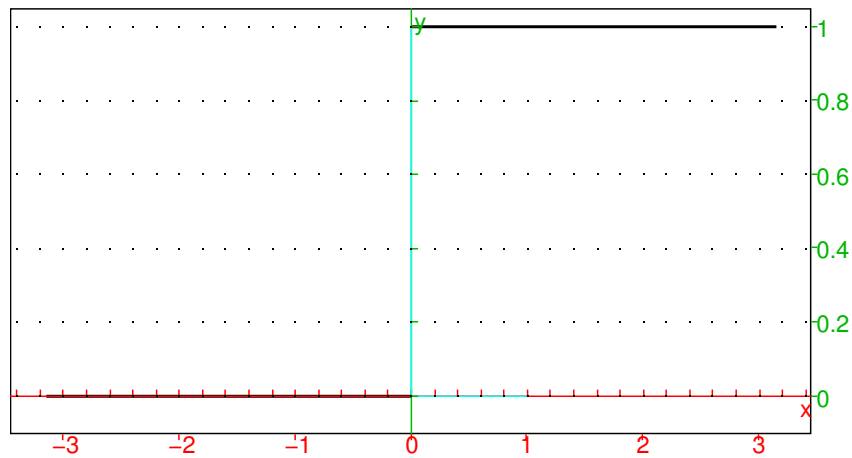
La série des Fourier exponentielle de f est alors la somme infinie

$$S^C(f) = \sum_{k=-\infty}^{\infty} \frac{i(-1)^k}{2k} e^{ikx}.$$

2. On considère la fonction f définie sur $[-\pi, \pi]$ telle que $f(x) = 1$ pour $x \geq 0$ et $f(x) = 0$ pour $x < 0$.

```
f(x) := piecewise(x >= 0, 1, 0); plot(f(x), x=-pi..pi,
display=line_width_2);
```


6.1. APPROXIMANTS DE FOURIER, COEFFICIENTS DE FOURIER ET SÉRIES DE FOURIER : DÉFINITIONS ET EXEMPLES



On a que

$$a_0(f) = \frac{1}{2\pi} \int_{-\pi}^{\pi} f(x) dx = \frac{1}{2}.$$

Par ailleurs, pour tout $k > 0$,

$$\begin{aligned} a_k(f) &= \frac{1}{\pi} \int_{-\pi}^{\pi} \cos(kx) f(x) dx = \frac{1}{\pi} \int_0^{\pi} \cos(kx) dx \\ &= \frac{1}{k\pi} [\sin(kx)]_0^{\pi} = 0 \end{aligned}$$

et

$$\begin{aligned} b_k(f) &= \frac{1}{\pi} \int_{-\pi}^{\pi} \sin(kx) f(x) dx = \frac{1}{\pi} \int_0^{\pi} \sin(kx) dx \\ &= \frac{1}{k\pi} [-\cos(kx)]_0^{\pi} = \frac{1 - (-1)^k}{k\pi}. \end{aligned}$$

Vérification

```
assume(k, integer) ; f(x) := piecewise(x >= 0, 1, 0) ;
fourier_an(f(x), x, 2*pi, k, -pi) ; fourier_bn
(f(x), x, 2*pi, k, -pi) ;
```

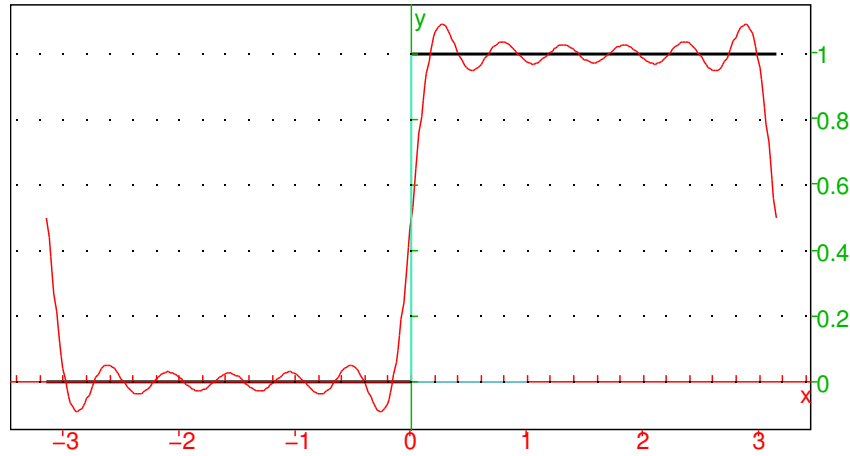
$$\text{Done, Done, } 0, \frac{(-(-1)^k + 1)}{k \cdot \pi}$$

On note que $(1 - (-1)^k) = 2$ si k est impair et 0 si k est pair. En écrivant tout k impair dans la forme $k = 2l + 1$, on obtient que la série de Fourier trigonométrique de f est

$$S^{\mathbb{R}}(f) = \frac{1}{2} + \sum_{l=0}^{+\infty} \frac{2}{(2l+1)\pi} \sin((2l+1)x).$$

Représentation graphique :

```
f(x) := piecewise(x >= 0, 1, 0) ; S(n) := 1/2 + sum(2/
(2*l+1)/pi * sin((2*l+1)*x), l, 0, n) ; plot([f
(x), S(5)], x = -pi..pi, color=[line_width_2, red])
```



Calculons maintenant la série de Fourier exponentielle de f . Pour $k \neq 0$ on a que

$$\begin{aligned} c_k(f) &= \frac{1}{2\pi} \int_{-\pi}^{\pi} e^{-ikx} f(x) dx = \frac{1}{2\pi} \int_0^{\pi} e^{-ikx} dx \\ &= \frac{1}{-2ik\pi} [e^{-ikx}]_0^{\pi} = \frac{-1 + (-1)^k}{-2ik\pi}. \end{aligned}$$

Vérification

```
assume(k, integer) ; f(x) := piecewise(x >= 0, 1, 0) ;
fourier_cn(f(x), x, 2*pi, k, -pi) ;
```

$$\text{Done, Done, } \frac{(i \cdot (-1)^k - i)}{2 \cdot k \cdot \pi}$$

Comme nous avons déjà calculé que $c_0 = a_0 = \frac{1}{2}$, la série de Fourier exponentielle de f est alors

$$S^C(f) = \frac{1}{2} + \sum_{l=-\infty}^{\infty} \frac{-i}{(2l+1)\pi} e^{i(2l+1)x}.$$

6.2 Séries en sin et cos.

La série de Fourier est une série qui mélange des termes de sin et en cos. Nous montrerons dans cette section comment modifier cette construction pour obtenir des séries en sin ou en cos approchant une fonction f donnée. Notre point de départ sera la proposition suivante :

Proposition 6.2.1 Soit $f : [-L, L] \rightarrow \mathbb{R}$ une fonction réelle. Alors

1. si f est paire alors $b_k(f) = 0$ pour tout k .
2. si f est impaire alors $a_k(f) = 0$ pour tout k .

On note que pour toute fonction g , impaire sur $[-L, L]$, nous avons que

$$\int_{-L}^L g(x) dx = 0.$$

Par ailleurs,

1. Le produit d'une fonction impaire et une fonction paire est lui-même une fonction impaire.
2. Pour tout $k > 0$ la fonction $\cos(kx)$ est une fonction paire et la fonction $\sin(kx)$ est une fonction impaire.

Si la fonction $f : [-L, L] \rightarrow \mathbb{R}$ est une fonction paire alors pour tout k la fonction $f(x) \sin(k\pi x/L)$ est une fonction impaire et donc

$$b_k(f) = \frac{1}{L} \int_{-L}^L f(x) \sin\left(\frac{k\pi x}{L}\right) dx = 0$$

Par contre, si la fonction $f : [-L, L] \rightarrow \mathbb{R}$ est une fonction impaire alors pour tout k la fonction $f(x) \cos(\frac{k\pi x}{L})$ est une fonction impaire et donc

$$a_k(f) = \frac{1}{L} \int_{-L}^L f(x) \cos\left(\frac{k\pi x}{L}\right) dx = 0$$

Ceci termine la démonstration de la proposition. Avec cette proposition, nous pourrons construire des sommes de cos (resp. de sin) approchant une fonction donnée. Notre méthode sera la suivante :

1. Etant donnée une fonction $f : [0, L] \rightarrow \mathbb{R}$, on construit une extension g sur $[-L, L]$ qui est paire (si on veut construire une série en cos) ou impaire (si on veut construire une série en sin.)
2. Nous construisons alors la série de Fourier de cette nouvelle fonction g .
3. Puisque g est paire (resp. impaire) sa série de Fourier ne contient que des termes en cos (resp. en sin.) C'est cette série de Fourier qui sera la série en cos (resp. en sin) de f .

Définition 6.2.2 Soit $f : [0, L] \rightarrow \mathbb{R}$ une fonction continue par morceaux. On définit sur $[-L, L]$ les extensions impaire et paire de f par

$$f_{\text{paire}}(x) = \begin{cases} f(x) & \text{si } x \geq 0, \\ f(-x) & \text{si } x < 0. \end{cases}$$

$$f_{\text{impaire}}(x) = \begin{cases} f(x) & \text{si } x > 0, \\ -f(-x) & \text{si } x < 0, \\ 0 & \text{si } x = 0. \end{cases}$$

Notons que f_{paire} est paire et f_{impaire} est impaire par construction. il résulte de la proposition 6.2.1 que

– pour tout k , $a_k(f_{\text{impaire}}) = 0$

– pour tout k , $b_k(f_{\text{paire}}) = 0$.

Nous pouvons maintenant définir les séries en sin et cos de notre fonction f .

Définition 6.2.3 Soit $f : [0, L] \rightarrow \mathbb{R}$ une fonction continue par morceaux. La série en sin de f est la série trigonométrique

$$S^{\text{sin}}(f) = S^{\mathbb{R}}(f_{\text{impaire}}) = \sum_{k=1}^{\infty} b_k(f_{\text{impaire}}) \sin\left(\frac{k\pi x}{L}\right).$$

La série en cos de f est la série trigonométrique

$$S^{\text{cos}}(f) = S^{\mathbb{R}}(f_{\text{paire}}) = a_0(f_{\text{paire}}) + \sum_{k=1}^{\infty} a_k(f_{\text{paire}}) \cos\left(\frac{k\pi x}{L}\right).$$

Remarque

Notons que par parité

$$b_k(f_{\text{impaire}}) = \frac{1}{L} \int_{-L}^L f_{\text{impaire}}(x) \sin\left(\frac{k\pi x}{L}\right) dx = \frac{2}{L} \int_0^L f(x) \sin\left(\frac{k\pi x}{L}\right) dx.$$

De même

$$a_k(f_{\text{paire}}) = \frac{1}{L} \int_{-L}^L f_{\text{paire}}(x) \cos\left(\frac{k\pi x}{L}\right) dx = \frac{2}{L} \int_0^L f(x) \cos\left(\frac{k\pi x}{L}\right) dx.$$

et

$$a_0(f) = \frac{1}{L} \int_0^L f(x) dx.$$

Exemple

On considère la fonction $f(x) = e^x$ sur l'intervalle $[0, \pi]$. Nous cherchons à calculer sa série en sin. Nous avons que

$$S^{\sin}(f) = \sum_{k=1}^{\infty} b_k \sin(kx)$$

avec

$$\begin{aligned} b_k &= \frac{2}{\pi} \int_0^{\pi} e^x \sin(kx) dx = \\ &= \frac{2}{\pi} \int_0^{\pi} e^x \frac{e^{ikx} - e^{-ikx}}{2i} dx = \\ &= \frac{1}{i\pi} \int_0^{\pi} e^{x(1+ik)} - e^{x(1-ik)} dx \\ &= \frac{1}{i\pi} \left[\frac{e^{x(1+ik)}}{1+ik} - \frac{e^{x(1-ik)}}{1-ik} \right]_0^{\pi} \\ &= \frac{1}{i\pi} \left[\frac{(1-ik)e^{x(1+ik)} - (1+ik)e^{x(1-ik)}}{1+k^2} \right]_0^{\pi} \\ &= \frac{1}{i\pi} \left[\frac{e^x(-2ik \cos(kx) + 2i \sin(kx))}{1+k^2} \right]_0^{\pi} \\ &= \frac{2k(1 - (-1)^k e^{\pi})}{(1+k^2)\pi}. \end{aligned}$$

Vérification

```
assume(k, integer) ;; 2/pi*int(exp(x)*sin(k*x)
), x, 0, pi)
```

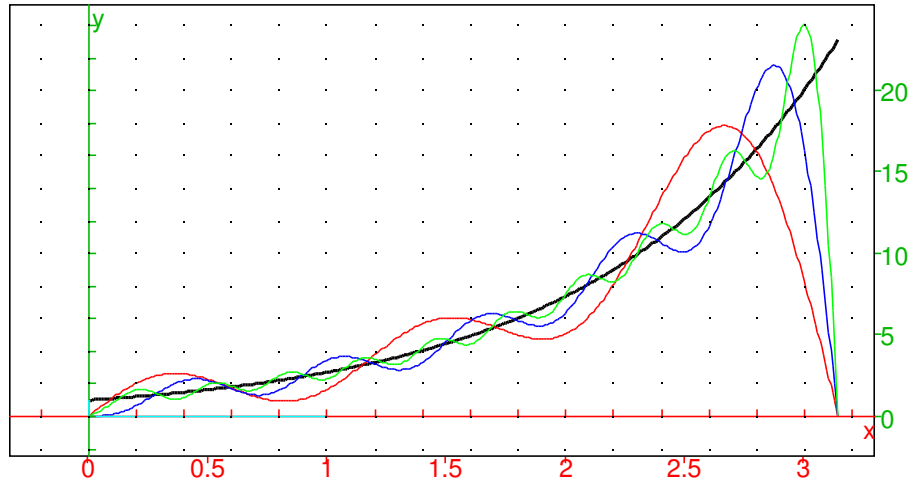
$$\text{Done, } \frac{2}{\pi} \cdot \left(-\frac{k \cdot (-1)^k e^{\pi}}{(k^2 + 1)} + \frac{k}{(k^2 + 1)} \right)$$

La série en sin de f est donc

$$S^{\sin}(f) = \sum_{k>0} \frac{2k(1 - (-1)^k e^{\pi})}{(1+k^2)\pi} \sin(kx).$$

Représentation graphique de f et des sommes partielles d'ordre 5 10 et 20.

```
f(x) := exp(x);
assume(k, integer) ;;
b := unapply(2/pi*int(f(x)*sin(k*x), x, 0, pi), k);
S(n) := sum(b(k)*sin(k*x), k, 1, n);
plot([f(x), S(5), S(10), S(20)], x=0..pi, color=[line_width_2, red, blue, green])
```



6.3 Convergence des séries de Fourier.

Nous avons donc créé, pour chaque fonction f continue par morceaux définie sur un intervalle $[-L, L]$, une suite de fonctions $S_n^{\mathbb{R}}(f)$. Chaque élément dans cette suite de fonctions est “plus proche” de la fonction f que celle qui la précède. Mais pour nous être utile, il faudrait s’assurer que, quitte à prendre n très grand, la fonction $S_n^{\mathbb{R}}(f)$ est aussi proche que l’on veut de la fonction f . Cette question sera le sujet de ce paragraphe.

Le théorème suivant, dont la démonstration dépasse le cadre de ce cours, nous assure que si la fonction f est continument dérivable alors en tout point x la série de Fourier converge vers f , plus précisément :

Théorème 6.3.1 (Dirichlet) Soit $f : [-L, L] \rightarrow \mathbb{C}$ une fonction C^1 par morceaux. Alors pour tout point $x \in]-L, L[$ où f est continue, on a :

$$\lim_{n \rightarrow \infty} S_n^{\mathbb{R}}(f)(x) \rightarrow f(x)$$

Pour les valeurs de $x \in]-L, L[$ où f effectue un saut, on a convergence de la série vers la moyenne des limites à droite et à gauche :

$$\lim_{n \rightarrow \infty} S_n^{\mathbb{R}}(f)(x) = \frac{1}{2} \left(\lim_{y \rightarrow x^+} f(y) + \lim_{y \rightarrow x^-} f(y) \right).$$

Ce résultat s’étend aux extrémités :

$$\lim_{n \rightarrow \infty} S_n^{\mathbb{R}}(f)(-L) = \lim_{n \rightarrow \infty} S_n^{\mathbb{R}}(f)(L) = \frac{\lim_{x \rightarrow L^-} f(x) + \lim_{x \rightarrow -L^+} f(x)}{2}.$$

En particulier, si f est continue et C^1 par morceaux, $S_n^{\mathbb{R}}(f)$ converge vers f sur $] -L, L[$.

Ce théorème est faux si la fonction f n’est pas supposée dérivable.

Il y a une autre forme de convergence qui nous sera utile, c’est la convergence en distance. Encore une fois, la démonstration de ce théorème dépasse le cadre de ce cours.

Théorème 6.3.2 (Théorème de Parseval.) Soit f une fonction réelle continue par morceaux sur un intervalle $[-L, L]$. Soit d_n la distance de f à la somme partielle d’ordre n de sa série de Fourier :

$$d_n = d(f, S_n^{\mathbb{R}}(f)) = \sqrt{\int_{-L}^L (f - S_n^{\mathbb{R}}(f))^2 dx}$$

entre f et $S_n^{\mathbb{R}}(f)$. Alors $d_n \rightarrow 0$ quand $n \rightarrow \infty$.

En appliquant le fait que la somme partielle de la série de Fourier est une projection orthogonale de f pour la norme $\|g\|^2 = \int_{-L}^L g(t)^2 dt$ et en appliquant l'inégalité triangulaire, on a

$$0 \leq \|f\| - \|S_n^{\mathbb{R}}(f)\| \leq d_n$$

D'où :

Corollaire 6.3.3 (Égalité de Parseval) *Soit f une fonction réelle continue par morceaux sur une intervalle $[-L, L]$. Alors*

$$\lim_{n \rightarrow \infty} \|S_n^{\mathbb{R}}(f)\| = \lim_{n \rightarrow \infty} (2La_0^2 + \sum_{k=1}^n L(a_k^2 + b_k^2)) = \int_{-L}^L f(x)^2 dx.$$

Autrement dit,

$$(2La_0^2 + \sum_{k \geq 1} L(a_k^2 + b_k^2)) = \int_{-L}^L f(x)^2 dx.$$

Exemple

Considérons $f(x) = x$ sur $[-\pi, \pi]$. On a vu que pour cette fonction $a_k = 0$ pour tout k et $b_k = \frac{2(-1)^k}{k}$. Par ailleurs

$$\int_{-\pi}^{\pi} x^2 dx = \frac{2\pi^3}{3}.$$

Il en suit que

$$\pi \sum_{k=1}^{\infty} \frac{4}{k^2} = \frac{2\pi^3}{3}$$

et donc

$$\sum_{k=1}^{\infty} \frac{1}{k^2} = \frac{\pi^2}{6}.$$

Corollaire 6.3.4 *Soient $f, g : \mathbb{R} \rightarrow \mathbb{C}$ deux fonctions réelles continues par morceaux et définies sur $[-L, L]$. Si $a_n(f) = a_n(g)$ et $b_n(f) = b_n(g)$ pour tout n alors pour tout point x où f et g sont toutes les deux continues $f(x) = g(x)$.*

L'égalité de Parseval implique que si $a_n(f) = a_n(g)$ et $b_n(f) = b_n(g)$ pour tout n alors

$$\int_{-L}^L |f(x) - g(x)|^2 dx = 0.$$

Ceci n'est possible que si $f(x) = g(x)$ en tout point x où f et g sont continues.

6.4 Solutions d'équations aux dérivées partielles.

Dans cette section, nous serons amenés à faire quelques manipulations dont nous ne pourrons pas donner une justification complète.

6.4.1 L'équation de la chaleur.

On rappelle que l'on cherche une fonction C^∞

$$T : [0, L] \times \mathbb{R}^+ \rightarrow \mathbb{R}$$

telle que

$$\frac{\partial T}{\partial t} = D \frac{\partial^2 T}{\partial x^2}, x \in [0, L], t > 0$$

avec les conditions aux bords

$$\frac{\partial T}{\partial x}(0, t) = \frac{\partial T}{\partial x}(L, t) = 0$$

et la condition initiale,

$$T(x, 0) = \varphi(x) \text{ pour tout } x \in [0, L], t = 0.$$

Remarquons que puisque nous cherchons une solution T qui est C^∞ , la condition initiale φ doit aussi être C^∞ sur $[0, L]$.

Nous allons montrer le théorème suivant :

Théorème 6.4.1 Soit $\varphi : [0, L] \rightarrow \mathbb{R}$ une fonction C^∞ sur $[0, L]$ telle que $\varphi'(0) = \varphi'(L) = 0$. Alors l'équation de la chaleur

$$\frac{\partial T}{\partial t} = D \frac{\partial^2 T}{\partial x^2}, x \in [0, L], t \geq 0$$

admet une unique solution C^∞

$$T : [0, L] \times \mathbb{R}^+ \rightarrow \mathbb{R}$$

vérifiant les conditions aux bords

$$\frac{\partial T}{\partial x}(0, t) = \frac{\partial T}{\partial x}(L, t) = 0$$

et la condition initiale,

$$T(x, 0) = \varphi(x) \text{ pour tout } x \in [0, L], t = 0.$$

Cette solution est donnée par le formule

$$T(x, t) = a_0 + \sum_{k=1}^{\infty} a_k \cos\left(\frac{k\pi}{L}x\right) e^{-\frac{\pi^2 k^2}{L^2}Dt},$$

où $a_0 = \frac{1}{L} \int_0^L \varphi(x) dx$ et pour tout $k > 0$ on a que $a_k = \frac{2}{L} \int_0^L \varphi(x) \cos\left(\frac{k\pi}{L}x\right) dx$.

La démonstration que la fonction T est bien définie et C^∞ dépasse le cadre de ce cours ¹

Montrons que T satisfait la condition initiale. Puisque φ est C^1 , et $\varphi'(0) = \varphi'(L) = 0$, son extension paire, φ_{paire} , est C^1 . Par le théorème de Dirichelet il suit que sur $[-L, L]$ nous avons que

$$\varphi_{\text{paire}}(x) = a_0 + \sum_{k=1}^{\infty} a_k \cos\left(\frac{k\pi x}{L}\right).$$

et en particulier pour tout $x \in [0, L]$

$$\varphi(x) = a_0 + \sum_{k=1}^{\infty} a_k \cos\left(\frac{k\pi x}{L}\right).$$

où $a_0 = \frac{1}{L} \int_0^L f(x) dx$ et pour tout $k > 0$ on a que $a_k = \frac{2}{L} \int_0^L \varphi(x) \cos\left(\frac{k\pi}{L}x\right) dx$.

Autrement dit, le théorème de Dirichelet nous dit dans ce cas que φ est égale à la somme (infinie) de sa série en cos.

Par analogie avec le cas où φ est égale à une somme finie de cosinus, nous allons poser

$$T(x, t) = a_0 + \sum_{k=1}^{\infty} a_k \cos\left(\frac{k\pi}{L}x\right) e^{-\frac{\pi^2 k^2}{L^2}Dt},$$

1. La régularité d'une fonction périodique peut se lire sur la décroissance de ses coefficients de Fourier a_k et b_k lorsque k tend vers l'infini. Si f est régulière, on peut intégrer par parties dans le calcul de a_k et b_k et faire apparaître autant de puissances de k négatives que l'on veut, a_k et b_k tendent vers 0 plus vite que n'importe quelle puissance négative de k . Réciproquement, les puissances de k qui apparaissent quand on dérive sous le signe somme ne génèrent pas la convergence de la série. On peut d'ailleurs observer qu'il suffit de prendre la condition initiale C^2 en x pour obtenir une solution C^∞ en tout instant non nul grâce à l'exponentielle qui décroît plus vite que toute puissance négative de k .

et vérifier que T est solution de l'équation de la chaleur. Pour tout $x \in [0, L]$, on a

$$T(x, 0) = a_0 + \sum_{k=1}^{\infty} a_k \cos\left(\frac{k\pi}{L}x\right) = \varphi(x).$$

et la condition initiale est donc vérifiée. De plus, $\frac{\partial T}{\partial x}(0, t) = \frac{\partial T}{\partial x}(L, t) = 0$ d'après les propriétés du cosinus : les conditions aux bords sont donc vérifiées.

Reste à établir que T satisfait $\frac{\partial T}{\partial t} = -D \frac{\partial^2 T}{\partial x^2}$. Posons

$$u_k(x, t) = a_k \cos\left(\frac{k\pi}{L}x\right) e^{-\frac{\pi^2 k^2 D}{L^2} t},$$

de telle façon que $T(x, t) = a_0 + \sum_k u_k(x, t)$. On a T qui est une somme infinie de fonctions

$$T(x, t) = \sum_{k=1}^{\infty} u_k(x, t)$$

On peut sous certaines conditions intervertir la dérivation et le signe \sum (ceci est mathématiquement assez délicat à justifier), de sorte que :

$$\frac{\partial T}{\partial x} = \sum_{k=1}^{\infty} \frac{\partial u_k}{\partial x}$$

et

$$\frac{\partial T}{\partial t} = \sum_{k=1}^{\infty} \frac{\partial u_k}{\partial t}$$

On admettra qu'on peut ici le faire. Alors

$$\begin{aligned} \frac{\partial T}{\partial t} &= \sum_{k \geq 1} a_k \frac{\partial}{\partial t} \left(\cos\left(\frac{k\pi}{L}x\right) e^{-\frac{\pi^2 k^2 D}{L^2} t} \right) \\ &= \sum_{k \geq 1} -D \frac{\pi^2 k^2}{L^2} a_k \left(\cos\left(\frac{k\pi}{L}x\right) e^{-\frac{\pi^2 k^2 D}{L^2} t} \right). \end{aligned}$$

Mais on a aussi

$$\begin{aligned} \frac{\partial^2 T}{\partial x^2} &= \sum_{k \geq 1} a_k \frac{\partial^2}{\partial x^2} \left(\cos\left(\frac{k\pi}{L}x\right) e^{-\frac{\pi^2 k^2 D}{L^2} t} \right) \\ &= \sum_{k \geq 1} -\frac{\pi^2 k^2}{L^2} a_k \left(\cos\left(\frac{k\pi}{L}x\right) e^{-\frac{\pi^2 k^2 D}{L^2} t} \right). \end{aligned}$$

On a donc bien

$$\frac{\partial T}{\partial t} = D \frac{\partial^2 T}{\partial x^2}$$

et la fonction donnée est donc une solution de notre équation.

Montrons maintenant que cette solution est unique. Pour cela, supposons que l'on ait deux solutions du problème, disons T_1 et T_2 , et posons $u = T_1 - T_2$. Il est facile de vérifier que l'on a (linéarité) :

$$\frac{\partial u}{\partial t} = D \frac{\partial^2 u}{\partial x^2}, x \in [0, L], t \in \mathbb{R}^+,$$

et que

$$\frac{\partial u}{\partial x}(0, t) = \frac{\partial u}{\partial x}(L, t) = u(x, 0) = 0 \text{ pour tout } x \in [0, L], t \in \mathbb{R}^+.$$

Considérons la fonction $J(t) = \frac{1}{2D} \int_0^L u(x, t)^2 dx$. On a (modulo la justification de la dérivation sous l'intégrale)

$$J'(t) = \frac{1}{D} \int_0^L u(x, t) \frac{\partial u(x, t)}{\partial t} dx = \int_0^L u(x, t) \frac{\partial^2 u(x, t)}{\partial x^2} dx.$$

En intégrant par parties, il vient, en tenant compte des conditions aux bords,

$$J'(t) = - \int_0^L \left(\frac{\partial u(x, t)}{\partial x} \right)^2 dx.$$

On a donc $J'(t) \leq 0$ pour tout $t \geq 0$. Ainsi J est décroissante, et on a

$$J(t) \leq J(0) \text{ pour tout } t \geq 0.$$

Or on a

$$J(0) = \frac{1}{2D} \int_0^L u(x, 0)^2 dx = 0,$$

d'après les conditions au bord vérifiées par u . Ainsi, $J(t) \leq 0$ pour tout $t \geq 0$. Mais J est l'intégrale d'une fonction positive ou nulle, et donc on a aussi $J(t) \geq 0$ pour tout $t \geq 0$. On en déduit que $J = 0$, et donc $u(x, t) = 0$ pour tout $t > 0$ et tout $x \in [0, L]$. Comme on a aussi $u(x, 0) = 0$ pour tout $x \in [0, L]$, u est identiquement nulle, c'est-à-dire $T_1 = T_2$.

6.4.2 L'équation des ondes.

On rappelle que l'on cherche une fonction

$$T : [0, L] \times \mathbb{R}^+ \rightarrow \mathbb{R}$$

telle que

$$\frac{\partial^2 T}{\partial t^2} = D \frac{\partial^2 T}{\partial x^2}, x \in [0, L], t > 0$$

avec les conditions aux bords

$$T(0, t) = T(L, t) = 0$$

et les conditions initiales,

$$T(x, 0) = \varphi(x) \text{ pour tout } x \in [0, L], t = 0.$$

$$\frac{\partial T}{\partial t}(x, 0) = 0 \text{ pour tout } x \in [0, L], t = 0.$$

Remarquons que puisque la solution T est C^∞ , la condition initiale φ doit aussi être C^∞ sur $[0, L]$.

Nous allons montrer le théorème suivant :

Théorème 6.4.2 Soit $\varphi : [0, L] \rightarrow \mathbb{R}$ une fonction C^∞ sur $[0, L]$ telle que $\varphi(0) = \varphi(L) = 0$. Alors l'équation des ondes

$$\frac{\partial T^2}{\partial t^2} = D \frac{\partial^2 T}{\partial x^2}, x \in [0, L], t \geq 00$$

admet une unique solution C^∞

$$T : [0, L] \times \mathbb{R}^+ \rightarrow \mathbb{R}$$

vérifiant les conditions aux bords

$$T(0, t) = T(L, t) = 0$$

et les conditions initiales,

$$T(x, 0) = \varphi(x) \text{ pour tout } x \in [0, L], t = 0.$$

$$\frac{\partial T}{\partial t}(x, 0) = 0 \text{ pour tout } x \in [0, L], t = 0.$$

Cette solution est donnée par la formule

$$T(x, t) = \sum_{k=1}^{\infty} b_k \sin\left(\frac{k\pi}{L}x\right) \cos\left(\frac{\pi k\sqrt{D}}{L}t\right),$$

où pour tout $k > 0$ on a que $b_k = \frac{2}{L} \int_0^L \varphi(x) \sin\left(\frac{k\pi}{L}x\right) dx$.

Puisque φ est C^1 , et $\varphi(0) = \varphi(L) = 0$, son extension impaire, φ_{impaire} est C^1 . Par le théorème de Dirichelet il suit que sur $[-L, L]$ nous avons que

$$\varphi_{\text{impaire}}(x) = \sum_{k=1}^{\infty} b_k(\varphi_{\text{impaire}}) \sin\left(\frac{k\pi x}{L}\right).$$

et en particulier pour tout $x \in [0, L]$

$$\varphi(x) = \sum_{k=1}^{\infty} b_k \sin\left(\frac{k\pi x}{L}\right).$$

où $b_k = \frac{2}{L} \int_0^L \varphi(x) \sin\left(\frac{k\pi}{L}x\right) dx$.

Autrement dit, le théorème de Dirichelet nous dit dans ce cas que φ est égale à la somme (infinie) de sa série en sin.

Par analogie avec le cas où φ est donnée par une somme finie de sinus, nous allons poser

$$T(x, t) = \sum_{k \geq 1} b_k \sin\left(\frac{k\pi}{L}x\right) \cos\left(\frac{\pi k\sqrt{D}}{L}t\right),$$

et vérifier que T est solution de l'équation des ondes. Pour tout $x \in [0, L]$, on a

$$T(x, 0) = \sum_{k \geq 1} b_k \sin\left(\frac{k\pi}{L}x\right) = \varphi(x)$$

et la première condition initiale est donc vérifiée. De plus, $T(0, t) = T(L, t) = 0$ d'après les propriétés du sinus : les conditions aux bords sont donc vérifiées.

Posons

$$u_k(x, t) = b_k \sin\left(\frac{k\pi}{L}x\right) \cos\left(\frac{\pi k\sqrt{D}}{L}t\right),$$

de telle façon que $T(x, t) = \sum_k u_k(x, t)$. Comme pour l'équation de la chaleur², on admettra qu'on peut intervertir la dérivation et le signe \sum , i.e. en posant

$$T(x, t) = \sum_k u_k(x, t)$$

on a bien que

$$\frac{\partial T}{\partial x} = \sum_{k=1}^{\infty} \frac{\partial u_k}{\partial x}$$

et

$$\frac{\partial T}{\partial t} = \sum_{k=1}^{\infty} \frac{\partial u_k}{\partial t}$$

2. Mais contrairement à l'équation de la chaleur, on a un facteur en cosinus qui oscille au lieu d'une exponentielle décroissante, il n'y a donc pas de régularisation en temps $t > 0$ pour une condition initiale éventuellement non régulière

Vérifions maintenant la deuxième condition initiale. Puisque pour tout k , $\frac{\partial u_k}{\partial t}(x, 0) = 0$, il suit que $\frac{\partial T}{\partial t}(x, 0) = 0$ pour tout $x \in [0, L]$.

Il reste à vérifier que T satisfait l'équation $\frac{\partial^2 T}{\partial t^2} = D \frac{\partial^2 T}{\partial x^2}$. On a que

$$\begin{aligned} \frac{\partial^2 T}{\partial t^2} &= \sum_{k \geq 1} b_k \frac{\partial^2}{\partial t^2} \left(\sin\left(\frac{k\pi}{L}x\right) \cos\left(\frac{\pi k \sqrt{D}}{L}t\right) \right) \\ &= \sum_{k \geq 1} -D \frac{\pi^2 k^2}{L^2} b_k \left(\sin\left(\frac{k\pi}{L}x\right) \cos\left(\frac{\pi k \sqrt{D}}{L}t\right) \right). \end{aligned}$$

Mais on a aussi

$$\begin{aligned} \frac{\partial^2 T}{\partial x^2} &= \sum_{k \geq 1} b_k \frac{\partial^2}{\partial x^2} \left(\sin\left(\frac{k\pi}{L}x\right) \cos\left(\frac{\pi k \sqrt{D}}{L}t\right) \right) \\ &= \sum_{k \geq 1} -\frac{\pi^2 k^2}{L^2} b_k \left(\sin\left(\frac{k\pi}{L}x\right) \cos\left(\frac{\pi k \sqrt{D}}{L}t\right) \right). \end{aligned}$$

On a donc bien

$$\frac{\partial^2 T}{\partial t^2} = D \frac{\partial^2 T}{\partial x^2}$$

et la fonction donnée est donc une solution de notre équation.

Montrons maintenant que cette solution est unique. Pour cela, supposons que l'on ait deux solutions C^∞ du problème, disons T_1 et T_2 , et posons $u = T_1 - T_2$. Il est facile de vérifier que l'on a

$$\frac{\partial u}{\partial t} = D \frac{\partial^2 u}{\partial x^2}, \quad x \in [0, L], t \in \mathbb{R}^+,$$

et que

$$u(0, t) = u(L, t) = u(x, 0) = 0 \text{ pour tout } x \in [0, L], t \in \mathbb{R}^+.$$

Considérons la fonction

$$b_k(t) = \frac{1}{2D} \int_0^L u(x, t) \sin\left(\frac{k\pi x}{L}\right) dx$$

Autrement dit, $b_k(t)$ est le k -ième coefficient dans la série en sin de la fonction $x \rightarrow u(x, t)$. Par le corollaire 6.3.4 il suffira de montrer que $b_k(t)$ est nul pour tout t . Par les conditions initiales on a que

$$b_k(0) = b_k, \quad \frac{\partial b_k}{\partial t}(0) = \int_0^L \frac{\partial u_k(x, 0)}{\partial t} \sin\left(\frac{k\pi x}{L}\right) dx = 0$$

Calculons

$$\begin{aligned} b_k''(t) &= \int_0^L \frac{\partial^2 u(x, t)}{\partial t^2} \sin\left(\frac{k\pi x}{L}\right) dx \\ &= D \int_0^L \frac{\partial^2 u(x, t)}{\partial x^2} \sin\left(\frac{k\pi x}{L}\right) dx = 0 \end{aligned}$$

ce qui est égal, après une double intégration par parties et en utilisant les conditions aux bords, à

$$-\frac{Dk^2\pi^2}{L^2} \int_0^L \frac{\partial^2 u(x, t)}{\partial x^2} \sin\left(\frac{k\pi x}{L}\right) dx = -\frac{Dk^2\pi^2}{L^2} b_k(t).$$

Autrement dit,

$$b_k''(t) = -\frac{Dk^2\pi^2}{L^2} b_k(t).$$

Mais la seule solution C^∞ de cette fonction telle que $b_k(0) = b_k'(0) = 0$ est la fonction nulle. On a donc $b_k(t) = 0$ pour tout k et tout t , ce qui donne bien que $u(x, t) = 0$ pour tout x, t .

6.5 La transformée de Fourier.

La série de Fourier que nous avons étudié ci-dessus s'adapte bien pour des fonctions définies sur un intervalle, c'est à dire, à des phénomènes de physique bornées. Or, en physique, il nous arrive souvent de vouloir analyser une fonction qui n'est pas a priori bornée dans le temps ou l'espace. Ceci nous inspire la question suivante :

Que pourrait être la série de Fourier d'une fonction sur \mathbb{R} ?

Soit $f : \mathbb{R} \rightarrow \mathbb{R}$ une fonction : pour plus de simplicité, nous ferons l'hypothèse que l'intégrale $\int_{\mathbb{R}} |f(x)| dx$ a une valeur finie. Nous allons essayer de trouver une analogue de la série de Fourier pour f en regardant ce qui se passe quand on prend, pour L de plus en plus grand, la série de Fourier de la fonction restreinte $f_L : [-L, L] \rightarrow \mathbb{R}$ donnée par

$$f_L(x) = f(x) \forall x \in [-L, L].$$

Il convient d'utiliser la série de Fourier exponentielle, pour ne pas avoir à considérer séparément les termes en cosinus et en sinus.

Le coefficient de $e^{i\lambda x}$ dans la série de Fourier exponentielle de f_L ³ est donnée par c_k ou $k = \frac{L\lambda}{\pi}$. Autrement dit, le coefficient de $e^{i\lambda x}$ dans la série de Fourier de f_L est

$$\frac{1}{2L} \int_{-L}^L f(x) e^{-i\lambda x} dx.$$

Appelons cette valeur $c_L(\lambda)$. Que se passe-t-il pour des valeurs de L de plus en plus grande ? Notre premier instinct pourrait être de considérer le comportement de $c_L(\lambda)$ quand $L \rightarrow \infty$ mais ce n'est pas très intéressant : il converge toujours vers 0. Comment contourner cette difficulté pour arriver à une limite intéressante ?

Observation clé. Lorsque nous essayons de construire une "série de Fourier" pour une fonction sur \mathbb{R} , a priori toutes les fonctions de la forme

$$e^{i\lambda x}$$

peuvent y apparaître, puisqu'il n'y a plus de conditions aux bords qui pourraient restreindre les valeurs possibles de λ . Notre "séries de Fourier" devraient donc être une "somme" d'éléments de la forme

$$c(\lambda) e^{i\lambda x}$$

indexé sur **tous** les nombre réels λ : je mets le mot somme en guillemets, car c'est mathématiquement impossible de faire une somme sur tous les nombres réels. En mathématiques, qu'est ce qui joue le rôle de "somme" sur les nombres réels ? C'est l'intégrale. On s'attend donc à ce que la "série de Fourier" d'une fonction f sur \mathbb{R} soit une expression **intégrale** pour f de la forme

$$f(x) = \int_{\mathbb{R}} c(\lambda) e^{i\lambda x}.$$

On veut comprendre ce qui devrait être la fonction $c(\lambda)$ qui va jouer le rôle des coefficients de Fourier $c_k(f)$. Ce raisonnement heuristique nous suggère l'idée suivante :

Si on veut considérer les limites quand $L \rightarrow \infty$ de la série de Fourier de la fonction f_L , il faudrait commencer par écrire cette série comme une intégrale.

Rappelons que

$$S_{\mathbb{C}}(f_L) = \sum_{k=-\infty}^{\infty} c_k e^{ik\pi x/L}$$

3. On fait ici l'hypothèse que le réel $L\lambda/\pi$ est un entier, et donc il existe un terme $e^{i\lambda x}$ dans la série de Fourier de f

que j'écris comme une intégrale de la façon suivante

$$S_{\mathbb{C}} = \sum_{k=-\infty}^{\infty} \frac{L}{\pi} \int_{k\pi/L}^{(k+1)\pi/L} c_k e^{ik\pi x/L} d\lambda.$$

Ici, j'ai écrit le terme $c_k e^{ik\pi x/L}$ comme l'intégrale par rapport à une variable λ d'une constante sur une intervalle de longueur $\frac{\pi}{L}$. Appliquant la relation de Chasles à cette somme d'intégrales, j'obtiens

$$S_{\mathbb{C}}(f)(x) = \int_{\mathbb{R}} \frac{L}{\pi} c_L(\lambda) e_L(\lambda, x) d\lambda$$

ou ici, $c_L(\lambda)$ est la fonction en escalier telle que pour tout $\lambda \in \left[\frac{k\pi}{L}, \frac{(k+1)\pi}{L} \right]$

$$c_L(\lambda) = \int_{-L}^L e^{-ik\pi x/L} f(x) dx$$

et $e_L(\lambda, x)$ est la fonction en escalier telle que pour tout $\lambda \in \left[\frac{k\pi}{L}, \frac{(k+1)\pi}{L} \right]$

$$e_L(\lambda, x) = e^{\frac{k\pi i x}{L}}$$

Que se passe-t-il dans cette expression quand $L \rightarrow \infty$, pour x et λ fixes ? Les intervalles $\left[\frac{k\pi}{L}, \frac{(k+1)\pi}{L} \right]$ sont alors de plus en plus petit, et si $\lambda \in \left[\frac{k\pi}{L}, \frac{(k+1)\pi}{L} \right]$ alors $\lambda \sim \frac{k\pi}{L}$. Quand $L \rightarrow \infty$

1. la fonction $e_L(\lambda, x)$ est approchée par la fonction $e^{i\lambda x}$
2. la fonction $c_L(\lambda)$ est approchée par la fonction $c(\lambda) = \int_{\mathbb{R}} e^{-i\lambda x} f(x) dx$.

Ces considérations inspirent la définition suivante.

Définition 6.5.1 Soit f une fonction continue sur \mathbb{R} telle que $\int_{-\infty}^{\infty} |f(x)| dx$ est finie. Alors on définit la transformée de Fourier de f , $\hat{f}(\lambda)$ par⁴

$$\hat{f} : \lambda \rightarrow \int_{-\infty}^{\infty} e^{-i\lambda x} f(x) dx.$$

Exemples

1. Soit f la fonction définie par $f(x) = 1$ si $x \in [a, b]$; $f = 0$ sinon. Alors

$$\hat{f}(\lambda) = \int_a^b e^{-i\lambda x} dx = \frac{e^{ib\lambda} - e^{ia\lambda}}{\lambda}$$

2. Soit f la fonction définie par $f(x) = e^{-x}$ si $x \geq 0$ et $f(x) = 0$ sinon. Alors

$$\hat{f}(\lambda) = \int_0^{\infty} e^{-i\lambda x} e^{-x} dx = \frac{1}{1 + i\lambda}.$$

Première propriétés des transformées de Fourier.

Lemme 6.5.2 Soient f, g deux fonctions de $\mathcal{C}^0(\mathbb{R}, \mathbb{R})$ telles que $\int_{\mathbb{R}} |f(x)| dx$ et $\int_{\mathbb{R}} |g(x)| dx$ soient finies. Alors pour tout scalaire a nous avons que

$$\widehat{f + ag} = \hat{f} + a\hat{g}.$$

Ce lemme est une conséquence immédiate de la linéarité de l'intégrale.

4. Nous avons choisi ici de supprimer le terme $\frac{1}{2\pi}$ conformément aux conventions habituellement utilisées en physique.

Lemme 6.5.3 Soit f une fonction de $C^0(\mathbb{R}, \mathbb{R})$ telle que $\int_{\mathbb{R}} |f(x)| dx$ et $\int_{\mathbb{R}} \left| \frac{\partial f}{\partial x} \right| dx$ soient finies. On suppose en plus que

$$\lim_{x \rightarrow -\infty} f(x) = \lim_{x \rightarrow \infty} f(x) = 0.$$

Alors

$$\widehat{\frac{\partial f}{\partial x}} = i\lambda \hat{f}(\lambda).$$

Démonstration.

Par définition

$$\widehat{\frac{\partial f}{\partial x}}(\lambda) = \lim_{L \rightarrow \infty} \int_{-L}^L f'(x) e^{-i\lambda x} dx.$$

En faisant une IPP sur cette expression on obtient

$$\frac{1}{2\pi} \left([f(x) e^{-i\lambda x}]_{-L}^L + \int_{-L}^L f(x) (i\lambda) e^{-i\lambda x} dx \right)$$

ce qui compte tenu de l'hypothèse $\lim_{x \rightarrow -\infty} f(x) = \lim_{x \rightarrow \infty} f(x) = 0$ nous donne

$$\begin{aligned} \lim_{L \rightarrow \infty} \int_{-L}^L i\lambda f(x) e^{-i\lambda x} dx \\ = i\lambda \hat{f}(\lambda). \end{aligned}$$

Notre dernière proposition, dont la démonstration dépasse le cadre de ce cours, dit qu'une équivalente du théorème de Dirichelet vaut pour cette transformée de Fourier.

Proposition 6.5.4 (Inversion de Fourier.) Soit f une fonction réelle et continue sur \mathbb{R} . Supposons que $\int_{\mathbb{R}} |f(x)| dx$ et $\int_{\mathbb{R}} |\hat{f}(\lambda)| d\lambda$ sont tous deux finis. Alors, nous pouvons reconstruire f de sa transformée de Fourier par la formule⁵

$$f(x) = \frac{1}{2\pi} \int_{-\infty}^{\infty} e^{i\lambda x} \hat{f}(\lambda) d\lambda.$$

Ces deux propriétés font de la transformée de Fourier un outil pratique pour la solution d'équations différentielles.

Exemple On considère l'équation différentielle

$$\frac{\partial f}{\partial x} - f = e^{-|x|}.$$

Appliquant la transformée de Fourier on obtient que

$$(i\lambda - 1)\hat{f}(\lambda) = \frac{2}{1 + \lambda^2}$$

Après réorganisation on a que

$$\hat{f} = \frac{2}{(1 + \lambda^2)(i\lambda - 1)}$$

et en utilisant le formule d'inversion de Fourier on obtient que

$$f(x) = \int_{-\infty}^{\infty} \frac{2e^{i\lambda x}}{(1 + \lambda^2)(i\lambda - 1)} d\lambda.$$

5. La terme $\frac{1}{2\pi}$ supprimée dans notre définition de la transformée de Fourier réapparaît ici.

Annexe A

Appendice : espace-temps, bases et forme de Minkowski.

La théorie de la relativité pose que la séparation que nous observons entre l'espace et le temps est une illusion et qu'en réalité les événements sont placés dans un continuum de dimension 4 que l'on appelle l'espace-temps. Cet espace-temps dont les points représentent des événements est de dimension 4, puisque pour préciser un événement il faut donner :

1. le lieu où il s'est produit (précisé dans des coordonnées cartésiennes par 3 données numériques) et
2. l'heure à laquelle il s'est produit (précisée par 1 donnée numérique).

Donc, pour préciser un événement il faut 4 coordonnées.

Contrairement à la physique Newtonienne, il n'est plus possible dans la relativité restreinte de donner une décomposition de l'espace temps en une partie spatiale plus une partie temps. Plus précisément, la physique Newtonienne, si elle place bien les événements dans un continuum de dimension 4 (lieu où l'événement a eu lieu, plus l'heure à laquelle il s'est produit) pose aussi l'existence d'une décomposition intrinsèque

$$\text{Espace-temps} = \text{espace} \oplus \text{temps}.$$

Deux observateurs Newtoniens, même en mouvement, seront toujours d'accord pour dire que deux événements ont lieu au même moment (c'est à dire, que le vecteur qui les sépare dans l'espace-temps est contenu dans le sous-espace "espace") ou qu'ils ont lieu au même endroit (c'est à dire, que le vecteur qui les sépare dans l'espace-temps est contenu dans le sous-espace "temps").

Ceci n'est plus vrai dans la relativité restreinte : selon les observateurs la décomposition de l'espace temps en espace et temps va varier. Lorsqu'un observateur O , non-soumis à une accélération, observe un événement, il va le mesurer utilisant son *référentiel*, et le résultat de ce mesure sera un *quadrivecteur* (T_O, X_O, Y_O, Z_O) où

- $\frac{T_O}{c}$ est l'heure à laquelle l'évènement s'est produit, mesuré par l'observateur O ,
- (X_O, Y_O, Z_O) est la position de l'évènement, mesuré par O .

Ces mesures n'ont plus rien d'absolu ; ils varieront selon l'observateur.

Qu'est ce que c'est un référentiel ? **C'est la donnée d'un origine A dans l'espace-temps¹ et d'une base de l'espace temps.** Après avoir fixé une origine A , l'observateur O mesure l'espace temps utilisant une base qui lui est propre (t_O, x_O, y_O, z_O) . Si v est un élément de l'espace-temps (que l'on considère comme un espace vectoriel avec origine A), les coordonnées de v dans la base (t_O, x_O, y_O, z_O) sont précisément les coefficients du quadri-vecteur (T_O, X_O, Y_O, Z_O) .

1. Ce qui permet de considérer l'espace temps comme un espace vectoriel, en identifiant un point P avec le vecteur \vec{AP}

Attention : si tous les référentiels ont des bases, seulement certaines bases spéciales peuvent être utilisées dans des référentiels.

Comment parler des événements ? On peut, bien sur, choisir un observateur et identifier un événement avec son quadrivecteur dans le référentiel de cet observateur. Cette solution est peu satisfaisante, puisqu'elle nous oblige à choisir un référentiel spécial auquel nous donnons une signification particulière, alors que le principe fondamental de la relativité restreinte est que toutes les référentiels se valent. Mais il est aussi possible avec les éléments ci-dessus de donner une description de la relativité restreinte sans référentiel distingué.

Recapitulons :

1. L'espace-temps, est un espace *affine*² de 4 dimensions dont les points représentent des événements. Il existe indépendamment de la choix du référentiel et du système de coordonnées. Nous l'appellerons *ET*.
2. Chaque observateur, O , mesure l'espace temps utilisant son propre référentiel, consistant en une choix d'une origine A et une base de *ET*, (t_O, x_O, y_O, z_O) . (*ET* est un espace vectoriel après choix de A .)
3. Soit $v \in ET$ un événement et soient (T_O, X_O, Y_O, Z_O) les coordonnées de v dans la base (t_O, x_O, y_O, z_O) . L'observateur O verra l'événement v en un temps T_O et une position (X_O, Y_O, Z_O) . *Le temps et la position d'un événement dans le référentiel de O sont simplement ses coordonnées dans cette base particulière.*

Une question naturelle se pose :

Question : y a-t-il des propriétés d'un vecteur $v \in ET$ qui ne dépendent pas du choix d'observateur ?

Si v est un vecteur de *ET* – pensons-le comme la séparation entre deux événements, A et B – alors ni le temps ni la distance représenté par v n'est indépendant de l'observateur. Mais il y a une notion intrinsèque, au moins pour les vecteurs de type temporels, c'est celui du *temps propre*.

Définition A.0.5 Soit $v = \overrightarrow{AB}$ un vecteur de *ET*. Le temps propre de v est le temps vécu par un observateur qui voyage de A et B sans accélération.

De nombreuses expériences ont établie comme donnée expérimentale la relation suivante entre le temps propre $TP(v)$ d'un vecteur temporel et ses coordonnées (T_O, X_O, Y_O, Z_O) mesurés par un observateur O :

$$TP(v)^2 = T_O^2 - X_O^2 - Y_O^2 - Z_O^2.$$

On reconnaît dans le membre de droite la forme quadratique associée à une forme bilinéaire, auxquelles nous donnons le nom de "produit scalaire de Minkowski".

Définition A.0.6 Soit $v, v' \in ET$, soit O un observateur, soit le référentiel de O donné par (t_O, x_O, y_O, z_O) . Soient (T_O, X_O, Y_O, Z_O) et (T'_O, X'_O, Y'_O, Z'_O) les quadrivecteurs de v et w mesurés par O . Le produit scalaire de Minkowski est la forme bilinéaire sur *ET* donnée par

$$M(v, v') = T_O T'_O - X_O X'_O - Y_O Y'_O - Z_O Z'_O.$$

Sa forme quadratique associée $q_M(v) = M(v, v)$ a la propriété que pour tout v temporel

$$q_M(v) = (\text{temps propre de } v)^2.$$

Attention : il y a ici un abus de notation. Puisqu'il existe des vecteurs pour lesquels $q_M(v) < 0$, la forme de Minkowski n'est pas un produit scalaire au sens des mathématiciens.

Plusieurs notions de la relativité restreinte admettent une interprétation en termes de la forme de Minkowski.

2. c'est à dire, un espace qui devient un espace vectoriel après choix d'une origine.

1. Un vecteur $v \in ET$ est temporel si $q_M(v) > 0$, lumineuse si $q_M(v) = 0$ et spatiale si $q_M(v) < 0$.
2. Les transformations de Lorentz sont des matrices P de changement de base (ou de changement de référentiel en gardant la même origine) qui laissent invariante la forme de Minkowski.

Annexe B

Appendice : le tenseur d'inertie d'un corps rigide.

Dans cette annexe, pour rester plus proche des notations utilisées en physique, nous dénoterons les quantités vectorielles dans \mathbb{R}^3 par des lettres en gras.

Soit C un corps rigide massif dans l'espace \mathbb{R}^3 . Nous voudrions comprendre le moment d'inertie de ce corps, c'est à dire, la résistance qu'elle oppose à être mise en rotation.

Pour plus de simplicité nous nous plaçons dans un référentiel inertiel ¹ dont l'origine est le centre de gravité G du corps C , et considérons une rotation du corps C autour de son centre de gravité G qui resterait fixe. Supposons que le corps C possède une vitesse angulaire ω . ² Quel serait le moment angulaire produit par cette rotation autour d'un autre axe ν ?

Dans un premier temps, considérons un réseau rigide de points massifs P_i , où chaque P_i a une masse m_i . En un temps t , soit \mathbf{r}_i le vecteur de position du point massif P_i (c'est à dire qu'on a $\mathbf{r}_i = \overrightarrow{GP_i}$) Soit \mathbf{v}_i la vitesse (normale) du point P_i . Le moment angulaire du point massif P_i autour de l'axe ν est donc

$$m_i \nu \cdot (\mathbf{r}_i \wedge \mathbf{v}_i).$$

Mais on sait par ailleurs par la définition de la vitesse angulaire que

$$\mathbf{v}_i = \omega \wedge \mathbf{r}_i$$

où ω est la vitesse angulaire. Lorsque P_i tourne avec une vitesse angulaire ω autour de son centre de gravité, le moment angulaire autour de l'axe ν du point massif P_i est donné par l'expression

$$m_i \nu \cdot (\mathbf{r}_i \wedge (\omega \wedge \mathbf{r}_i)).$$

Le moment total de ce réseau de points est alors donné par la somme

$$\sum_i m_i \nu \cdot (\mathbf{r}_i \wedge (\omega \wedge \mathbf{r}_i))$$

Approchons maintenant C par un maillage de points massifs P_i , chacun de masse m_i , en une position \mathbf{r}_i . Le moment angulaire de C autour de l'axe ν est alors approché par la quantité

$$\sum_i m_i \nu \cdot (\mathbf{r}_i \wedge (\omega \wedge \mathbf{r}_i))$$

1. c'est à dire la donnée d'une origine dans l'espace et une base de \mathbb{R}^3 ne variant pas avec le temps.

2. La vitesse angulaire d'un corps en rotation dont le centre de gravité est fixe est une quantité vectorielle ω telle que la vitesse d'un point $P \in C$ est donnée par $\omega \wedge \overrightarrow{GP}$. La direction de ω donne l'axe de rotation et sa longueur donne la vitesse.

ou ω est la vitesse angulaire du corps C .

Pour un maillage assez fini de C , la masse m_i du point massif P_i est très proche de ρdV , ou ρ est la densité locale de C au point P_i et dV est l'élément de volume autour de P_i . On obtient, en approchant C par un maillage toujours plus fin de points P_i que le moment angulaire de C autour de ν est donné par

$$\begin{aligned} & \int_C \rho \nu \cdot (\mathbf{r}_i \wedge (\omega \wedge \mathbf{r}_i)) dV \\ &= \nu \cdot \int_C \rho (\mathbf{r} \wedge (\mathbf{r} \wedge \omega)) dV \\ &= \nu \cdot \int_C \rho (\omega \cdot \|\mathbf{r}\|^2 - \mathbf{r}(\omega \cdot \mathbf{r})) dV \\ &= \nu \cdot \omega \left(\int_C \rho \|\mathbf{r}\|^2 dV \right) - \int_C (\nu \cdot \mathbf{r})(\omega \cdot \mathbf{r}). \end{aligned}$$

ou ω est la vitesse angulaire du corps C .

Pour résumer, nous avons la proposition suivante.

Proposition B.0.7 Soit I_C la fonction de deux vecteurs dans \mathbb{R}^3 , ω et ν , telle que $I_C(\omega, \nu)$ le moment d'inertie du corps rigide C autour de l'axe ν lorsque C tourne autour de G avec vitesse angulaire ω .

Alors nous avons que

$$I_C(\omega, \nu) = \nu \cdot \omega \left(\int_C \rho \|\mathbf{r}\|^2 dV \right) - \int_C (\nu \cdot \mathbf{r})(\omega \cdot \mathbf{r}).$$

L'exercice suivant est laissé au lecteur.

Exercice L'application I_C est une forme bilinéaire symétrique.

Définition B.0.8 Soit C un corps rigide massif dans l'espace. La forme bilinéaire symétrique I_C s'appelle le tenseur d'inertie du corps C .

Remarque B.0.9 L'expression du tenseur d'inertie dans une base inertielle dépend de l'orientation de C dans l'espace. En particulier, lorsque l'orientation de C varie dans le temps³ l'expression de I_C dans une base inertielle ne sera pas constante.

Remarque B.0.10 Le théorème sur la diagonalisation orthonormée des formes bilinéaires symétrique nous garantit l'existence d'une base orthonormée pour \mathbb{R}^3 , $\mathbf{i}, \mathbf{j}, \mathbf{k}$, qui est une base orthogonale pour le tenseur de l'inertie.

Définition B.0.11 Les éléments $\mathbf{i}, \mathbf{j}, \mathbf{k}$ de la base qui est orthonormée pour \mathbb{R}^3 et orthogonale pour I_C s'appellent les axes principaux d'inertie du corps C .

Remarque B.0.12 Ces axes principaux dépendent, bien sur, de l'orientation du corps C .

Calculons maintenant le moment total de C , μ_C .⁴ il résulte immédiatement de notre formule pour $I_C(\omega, \nu)$ que μ_C est donné par le formule

$$\mu_C = \omega \left(\int_C \rho \|\mathbf{r}\|^2 dV \right) - \int_C \rho \mathbf{r}(\omega \cdot \mathbf{r}) dV.$$

3. Notamment, lorsque C est en rotation.

4. Ce moment total d'un corps en rotation est une quantité vectorielle μ_C qui a la propriété que pour tout axe ν le moment de C autour de ν est donné par le produit scalaire $\nu \cdot \mu_C$.

B.1 Application : rotation libre d'un objet avec symmétrie rotationnelle.

Nous savons qu'il existe une base $\mathbf{i}, \mathbf{j}, \mathbf{k}$ telle que à tout moment, $\mathbf{i}, \mathbf{j}, \mathbf{k}$ est la base orthonormée de \mathbb{R}^3 qui est aussi orthogonale pour I_C . Cette base est variable dans le temps - elle n'est pas un référentielle inertielle - mais elle est constante du point de vue du corps rigide et en particulier, si C est en rotation autour de son centre de gravité avec vitesse angulaire ω alors on a que

$$\dot{\mathbf{i}} = \mathbf{i} \wedge \omega \mathbf{j}' = \mathbf{j} \wedge \omega \mathbf{k}' = \mathbf{k} \wedge \omega.$$

Ici, pour toute quantité vectorielle \mathbf{a} nous notons \mathbf{a}' sa dérivée dans une base inertielle. Par ailleurs, nous notons $\frac{\partial \mathbf{a}}{\partial t'}$ sa dérivée dans la base $\mathbf{i}, \mathbf{j}, \mathbf{k}$. Nous avons alors que

$$\frac{\partial \mathbf{a}}{\partial t'} + \mathbf{a} \wedge \omega = \mathbf{a}'.$$

Dans le référentiel $(\mathbf{i}, \mathbf{j}, \mathbf{k})$ la matrice de I_C est constante et diagonale. Considérons le cas où C a une symmétrie rotationnelle autour de l'axe \mathbf{i} : la matrice de I_C dans la base $\mathbf{i}, \mathbf{j}, \mathbf{k}$ est alors de la forme

$$\begin{pmatrix} a & 0 & 0 \\ 0 & b & 0 \\ 0 & 0 & b \end{pmatrix}.$$

Supposons maintenant que le corps C tourne librement, sans torque extérieure. On sait alors que son moment d'inertie μ_C satisfait l'équation

$$\mu'_C = 0$$

En passant dans le repère $\mathbf{i}, \mathbf{j}, \mathbf{k}$ on obtient

$$\frac{\partial \mu_C}{\partial t'} = -\mu_C \wedge \omega$$

ce qui, en utilisant le fait que $\mu_C = I_C(\omega)$ nous donne

$$I_C \frac{\partial \omega}{\partial t'} = -I_C \omega \wedge \omega.$$

Ecrivant $\omega = x\mathbf{i} + y\mathbf{j} + z\mathbf{k}$ on voit que

$$ax' = 0$$

$$by' = (a - b)xz$$

$$bz' = (b - a)xy$$

La valeur x est donc une constante. Posons $\frac{(b-a)x}{b} = C$, nous avons alors les équations

$$y' = -Cz$$

$$z' = Cy$$

qui a pour solution générale

$$y = \lambda \cos Ct + \theta$$

$$z = \lambda \sin Ct + \theta.$$

L'axe de rotation ω n'est donc pas constante dans le référentiel corporel. Il tourne autour de l'axe de symmétrie rotationnelle \mathbf{i} avec une vitesse qui dépend de

1. l'angle entre ω et \mathbf{i}
2. le rapport entre a et b .

Annexe C

Appendice : les coniques et quadriques.

On va maintenant appliquer les résultats précédents à l'étude des coniques et des quadriques.

Jusqu'à la fin de ce paragraphe, on se place dans l'espace affine \mathbb{R}^2 ou \mathbb{R}^3 , muni de son repère orthonormé usuel.

Définition C.0.1 Une **conique** est le lieu géométrique de \mathbb{R}^2 défini par une équation de la forme

$$ax^2 + by^2 + 2cxy + dx + ey + f = 0,$$

où au moins un des termes quadratiques est non nul.

Une **quadrique** est le lieu géométrique de \mathbb{R}^3 défini par une équation de la forme

$$ax^2 + by^2 + cz^2 + 2dxy + 2exz + 2fyz + \alpha x + \beta y + \gamma z + \delta = 0,$$

où au moins un des termes quadratiques est non nul.

On veut classer les différents types de coniques et quadriques. Puisque l'on veut conserver le lieu géométrique, on ne s'autorise qu'à faire des changements de variables qui remplacent le repère canonique par un autre repère orthonormé (pour les produit scalaire canonique), donc des translations ou des isométries (i.e. qui conservent les distances et les angles non orientés).

Ceci est nécessaire : en effet, l'équation

$$\frac{x^2}{9} + \frac{y^2}{4} = 1$$

représente une ellipse de centre O , alors que le changement de variables $x' = \frac{x}{3}$ et $y' = \frac{y}{2}$ donne l'équation

$$x'^2 + y'^2 = 1$$

représente le cercle unité, qui n'est pas le même lieu géométrique.

Cas des coniques.

Considérons la conique

$$ax^2 + by^2 + 2cxy + dx + ey + f = 0,$$

et soit

$$q : \mathbb{R}^2 \rightarrow \mathbb{R}, (x, y) \mapsto ax^2 + by^2 + 2cxy.$$

D'après le théorème 4.32, il existe une base orthonormée (v_1, v_2) qui est ϕ_q -orthogonale, donc q -orthogonale.

Dans le nouveau repère (O, v_1, v_2) , l'équation de la conique s'écrit

$$a'x'^2 + c'y'^2 + d'x + e'y + f' = 0.$$

On se débarrasse ensuite d'un ou deux termes linéaires en complétant les carrés et en effectuant une translation d'origine.

Après éventuellement permutation des nouvelles variables et/ou changements du type $X \leftrightarrow -X$ ou $Y \leftrightarrow -Y$ on obtient une équation d'un des types suivants, selon la signature de q , pour des réels $U, V > 0$

(1) Signature $(2, 0)$ ou $(0, 2)$:

1. $\frac{X^2}{U^2} + \frac{Y^2}{V^2} = 1$: c'est une ellipse (ou un cercle si $U = V$).

2. $\frac{X^2}{U^2} + \frac{Y^2}{V^2} = 0$: c'est le point $(X, Y) = (0, 0)$.

3. $\frac{X^2}{U^2} + \frac{Y^2}{V^2} = -1$: c'est l'ensemble vide.

(2) Signature $(1, 1)$:

1. $\frac{X^2}{U^2} - \frac{Y^2}{V^2} = 1$: c'est une hyperbole.

2. $\frac{X^2}{U^2} - \frac{Y^2}{V^2} = 0$: c'est la réunion des deux droites d'équation $X/U = Y/V$ et $X/U = -Y/V$.

(3) Signature $(1, 0)$ ou $(0, 1)$: $Y^2 = 2\lambda X$: c'est une parabole.

Exemple C.0.2 *Considérons la conique d'équation*

$$3x^2 - 3y^2 + 8xy + 6\sqrt{5}x + 2\sqrt{5}y + 5 = 0.$$

Soit $q : \mathbb{R}^2 \rightarrow \mathbb{R}, \begin{pmatrix} x \\ y \end{pmatrix} \mapsto 3x^2 - 3y^2 + 8xy$.

Sa matrice représentative dans la base canonique est

$$\begin{pmatrix} 3 & 4 \\ 4 & -3 \end{pmatrix}.$$

D'après un exemple précédent, une base orthonormée qui est aussi q -orthogonale est donnée par

$$\frac{1}{\sqrt{5}} \begin{pmatrix} 2 \\ 1 \end{pmatrix}, \frac{1}{\sqrt{5}} \begin{pmatrix} 1 \\ -2 \end{pmatrix},$$

les vecteurs étant respectivement des vecteurs propres pour 5 et -5 .

Soient x', y' les coordonnées dans cette nouvelle base. On a donc

$$\begin{pmatrix} x \\ y \end{pmatrix} = \frac{1}{\sqrt{5}} \begin{pmatrix} 2 & 1 \\ 1 & -2 \end{pmatrix} \begin{pmatrix} x' \\ y' \end{pmatrix} = \frac{1}{\sqrt{5}} \begin{pmatrix} 2x' + y' \\ x' - 2y' \end{pmatrix}.$$

Par construction de cette base, la forme q dans cette base s'écrit

$$5x'^2 - 5y'^2.$$

Elle est donc de signature $(1, 1)$, et on a donc une hyperbole (sauf cas dégénéré).

On a alors

$$5x'^2 - 5y'^2 + 10x' + 10y' + 5 = 0,$$

soit

$$x'^2 - y'^2 + 2x' + 2y' + 1 = 0.$$

On a donc

$$(x' + 1)^2 - (y' - 1)^2 + 1 = 0.$$

En posant $X = x' - 1, Y = y' - 1$, on obtient $X^2 - Y^2 = -1$. En posant $X' = Y$ et $Y' = X$, on obtient finalement l'équation réduite de l'hyperbole

$$X'^2 - Y'^2 = 1.$$

Cas des quadriques.

Comme précédemment, on se ramène au cas d'une équation sans termes croisés, et on se débarrasse d'un ou plusieurs termes linéaires.

Après éventuellement permutation des nouvelles variables et/ou changements du type $X \leftrightarrow -X$, $Y \leftrightarrow -Y$, $Z \leftrightarrow -Z$, et éventuellement une nouvelle translation/rotation,

on obtient la classification suivante :

(1) Signature (3, 0) ou (0, 3). On obtient 3 cas :

(a) $\frac{X^2}{U^2} + \frac{Y^2}{V^2} + \frac{Z^2}{W^2} = 1$: c'est un ellipsoïde.

(b) $\frac{X^2}{U^2} + \frac{Y^2}{V^2} + \frac{Z^2}{W^2} = 0$: c'est le point $(X, Y, Z) = (0, 0, 0)$

(c) $\frac{X^2}{U^2} + \frac{Y^2}{V^2} + \frac{Z^2}{W^2} = -1$: c'est l'ensemble vide.

(2) Signature (2, 1) ou (1, 2). On obtient 3 cas :

(a) $\frac{X^2}{U^2} + \frac{Y^2}{V^2} - \frac{Z^2}{W^2} = -1$: c'est un hyperboloïde à deux nappes

(b) $\frac{X^2}{U^2} + \frac{Y^2}{V^2} - \frac{Z^2}{W^2} = 1$: c'est un hyperboloïde à une nappe

(c) $\frac{X^2}{U^2} + \frac{Y^2}{V^2} = \frac{Z^2}{W^2}$: c'est un cône.

(3) Signature (2, 0) ou (0, 2). On obtient quatre cas :

(a) $\frac{X^2}{U^2} + \frac{Y^2}{V^2} = 1$: c'est un cylindre elliptique

(b) $\frac{X^2}{U^2} + \frac{Y^2}{V^2} = 0$: c'est la droite $X = Y = 0$.

(c) $\frac{X^2}{U^2} + \frac{Y^2}{V^2} = -1$: c'est l'ensemble vide.

(d) $\frac{X^2}{U^2} + \frac{Y^2}{V^2} = \frac{Z}{W}$: c'est un paraboloides elliptique.

(4) Signature (1, 1) :

(a) $\frac{X^2}{U^2} - \frac{Y^2}{V^2} = 1$: c'est un cylindre hyperbolique.

(b) $\frac{X^2}{U^2} - \frac{Y^2}{V^2} = 0$: c'est la réunion des deux plans d'équation $X/U - Y/V = 0$ et $X/U + Y/V = 0$

(c) $\frac{X^2}{U^2} - \frac{Y^2}{V^2} = -1$: c'est l'ensemble vide.

(d) $\frac{X^2}{U^2} - \frac{Y^2}{V^2} = \frac{Z}{W}$: c'est un paraboloides hyperbolique.

Signature (1, 0) ou (0, 1) :

(a) $X^2 = 2pY$: cylindre parabolique.

(b) $X^2/U^2 = 1$: réunion de deux plans parallèles d'équation $X = 1$ et $X = -1$.

(c) $X^2/U^2 = 0$: plan $X = 0$

(d) $X^2/U^2 = -1$: ensemble vide.

Exemple C.0.3 Soit la quadrique

$$x^2 + y^2 + z^2 + 2xy + 2xz + 2yz + \sqrt{3}x + \sqrt{3}y + 2 = 0.$$

Soit $q : \mathbb{R}^3 \rightarrow \mathbb{R}$, $\begin{pmatrix} x \\ y \\ z \end{pmatrix} \mapsto x^2 + y^2 + z^2 + 2xy + 2xz + 2yz$.

Sa matrice représentative dans la base canonique est

$$\begin{pmatrix} 1 & 1 & 1 \\ 1 & 1 & 1 \\ 1 & 1 & 1 \end{pmatrix}.$$

D'après un exemple précédent, une base orthonormée qui est aussi q -orthogonale est donnée par

$$\frac{1}{\sqrt{3}} \begin{pmatrix} 1 \\ 1 \\ 1 \end{pmatrix}, \frac{1}{\sqrt{2}} \begin{pmatrix} 1 \\ -1 \\ 0 \end{pmatrix}, \sqrt{\frac{2}{3}} \begin{pmatrix} \frac{1}{2} \\ \frac{1}{2} \\ -1 \end{pmatrix},$$

ces vecteurs étant respectivement des vecteurs propres pour 1, 0 et 0.

Soient x', y', z' les coordonnées dans cette nouvelle base.

Par construction de cette base, la forme q dans cette base s'écrit

$$3x'^2.$$

Elle est donc de signature $(1, 0)$, et on a donc un cylindre parabolique, un ensemble vide, un plan ou une réunion de deux plans.

On vérifie que l'équation de cette quadrique dans cette base est

$$3x'^2 + 2x' + \sqrt{2}z' + 2 = 0,$$

soit

$$3\left(x' + \frac{1}{3}\right)^2 + \sqrt{2}z' + \frac{5}{3} = 0.$$

Si on pose $X = x' + \frac{1}{3}$, $Y = z' + \frac{5}{3\sqrt{2}}$, $Z = y'$, on obtient

$$X^2 = -\sqrt{2}Y.$$

Annexe D

Appendice : Formes hermitiennes.

Dans beaucoup d'applications en physique - notamment en mécanique quantique, mais pas exclusivement - nous avons besoin d'utiliser des espaces complexes. Par exemple, la fonction d'onde qui représente une particule dans la représentation de Schrödinger, est un élément de $C^0(\mathbb{R}^3, \mathbb{C})$, c'est-à-dire, une fonction complexe sur l'espace \mathbb{R}^3 .

Mais si on essaie de définir une notion de longueur sur un espace complexe V utilisant des formes bilinéaires complexes on se rend rapidement compte que c'est impossible. En effet, aucune forme bilinéaire complexe φ ne peut avoir une forme quadratique associée qui est réelle positive partout, puisque si

$$\varphi(v, v) > 0$$

alors on a

$$\varphi(iv, iv) = i^2\varphi(v, v) = -\varphi(v, v) < 0$$

Par contre, on sait que la fonction

$$f(z) = \bar{z}z$$

est partout réelle et positive sur \mathbb{C} : d'ailleurs, la distance euclidienne sur \mathbb{C} , vu comme un \mathbb{R} -espace vectoriel, est donnée par

$$d(z_1, z_2) = \sqrt{(z_1 - z_2)(\bar{z}_1 - \bar{z}_2)}.$$

Nous allons donc essayer de définir des distances sur des espaces complexes en utilisant des formes *hermitiennes*, c'est-à-dire, des fonctions de deux variables se comportant comme la fonction

$$(z_1, z_2) \mapsto \bar{z}_1 z_2.$$

Définition D.0.4 Soit V un espace vectoriel complexe. Une fonction

$$h : V \times V \rightarrow \mathbb{C}$$

est une forme hermitienne si et seulement si

1. $h(x + y, z) = h(x, z) + h(y, z)$
2. $h(x, \lambda y) = \lambda h(x, y)$
3. $h(x, y) = \overline{h(y, x)}$.

Notez qu'il résulte de (3) que $h(x, x)$ est réel pour tout x . Nous avons, par ailleurs, que $h(\lambda x, y) = \bar{\lambda}h(x, y)$.

Exemples D.0.5 1. La forme h définie sur $\mathbb{C}^n \times \mathbb{C}^n$ par

$$h \left(\begin{pmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{pmatrix}, \begin{pmatrix} y_1 \\ y_2 \\ \vdots \\ y_n \end{pmatrix} \right) = \sum \overline{x_i} y_i$$

est une forme hermitienne. Celle-ci s'appelle la forme hermitienne canonique sur \mathbb{C}^n .

2. La forme h définie sur $C^0([a, b], \mathbb{C}) \times C^0([a, b], \mathbb{C})$ par

$$h(f, g) = \int_a^b \overline{f(x)} g(x) dx$$

est une forme hermitienne sur $C^0([-1, 1], \mathbb{C})$.

3. La forme h définie sur $M_n(\mathbb{C}) \times M_n(\mathbb{C})$ par

$$h(M, N) = \text{Tr}({}^t \overline{M} N)$$

est une forme hermitienne.

On dit qu'une forme hermitienne h est définie positive si pour tout $x \in V \setminus 0_V$ nous avons que

$$h(x, x) > 0.$$

Définition D.0.6 Un espace hermitien (V, h) est la donnée d'un espace vectoriel complexe V et d'une forme hermitienne h , définie positive sur E .

Bien sûr, tout ce que nous avons fait pour les formes bilinéaires symétriques peut aussi se faire pour les formes hermitiennes.

Proposition D.0.7 Soit E un espace vectoriel complexe et soit h une forme hermitienne sur E . Soit $\mathbf{e} = (e_1 \dots e_n)$ une base pour E et soit M la matrice définie par $M_{i,j} = h(e_i, e_j)$. La matrice M est appelée la matrice de h dans la base \mathbf{e} . Soient $x, y \in E$ et soient $\underline{X}, \underline{Y}$ leurs vecteurs de coordonnées dans la base \mathbf{e} . Alors nous avons

$$h(x, y) = {}^t \overline{\underline{X}} M \underline{Y}$$

Proposition D.0.8 (Règles de changement de la base) Soit E un espace vectoriel complexe et soit h une forme hermitienne sur E . Soient $\mathbf{e} = (e_1 \dots e_n)$ et $\mathbf{f} = (f_1 \dots f_n)$ deux bases pour E . Soit M la matrice de h dans la base \mathbf{e} et soit N la matrice de h dans la base \mathbf{f} . Soit P la matrice de passage de \mathbf{e} vers \mathbf{f} . Alors

$$N = {}^t \overline{P} M P.$$

Nous notons que condition (3) de la définition des formes hermitiennes implique la proposition suivante.

Proposition D.0.9 Soit h une forme hermitienne sur un espace complexe, V . Soit M la matrice de h dans une base \mathbf{e} . Alors

$${}^t \overline{M} = M$$

Cette proposition inspire la définition suivante.

Définition D.0.10 Soit M une matrice complexe. L'adjointe M^* de M est la matrice définie par

$${}^t \overline{M} = M^*.$$

Définition D.0.11 Une matrice complexe M de taille $n \times n$ est dite hermitienne (ou auto-adjointe) si et seulement si

$$M^* = {}^t \overline{M} = M.$$

Nous pouvons définir le longueur d'un vecteur et la distance entre deux éléments dans un espace hermitien comme pour un espace prehilbertien.

Définition D.0.12 Soit (V, h) un espace hermitien et soient $v, w \in V$. On définit la longueur de v par

$$\|v\| = \sqrt{h(v, v)},$$

et la distance entre v et w par $d(v, w) = \|v - w\|$.

Avec cette notion de distance et de longueur, une version du procédé de Gram-Schmidt et la projection orthogonale sont valables aussi sur des espaces hermitiens.

Proposition D.0.13 (Projection orthogonale dans les espaces hermitiens.) Soit (V, h) un espace hermitien et soit $W \in E$ un sous-espace vectoriel de dimension finie. Soit (v_1, \dots, v_n) une base orthonormée pour W . Alors pour tout $v \in V$ on définit la projection orthogonale de v sur W par

$$p_W(v) = \sum_{i=1}^n h(v_i, v)v_i.$$

La projection $p_W(v)$ est alors l'élément de W qui minimise la distance $d(v, w)$ lorsque w parcourt W .

La démonstration est identique à celle donnée dans le cas des espaces prehilbertiens.

Proposition D.0.14 Soit (V, h) un espace hermitien de dimension finie et soit $\mathbf{e} = (e_1, \dots, e_n)$ une base. On construit une nouvelle base (v_1, \dots, v_n) récursivement par l'algorithme suivant :

1. On pose $v_1 = \frac{e_1}{\|e_1\|}$.
2. La famille (v_1, \dots, v_k) étant construite, nous posons

$$f_{k+1} = e_{k+1} - \sum_{i=1}^k h(v_i, e_{k+1})v_i.$$

3. On pose $v_{k+1} = f_{k+1}/\|f_{k+1}\|$.
4. On a maintenant construit (v_1, \dots, v_{k+1}) et on revient à l'étape (2) pour construire v_{k+1} .

La base de (V, h) ainsi construite est orthonormée.

La démonstration est identique à celle donnée dans le cas des espaces prehilbertiens.