

Chapitre 4

Formes quadratiques

Ces notes ont été rédigées par Mélanie GUENAI (2020-2022), et reprises par Damien THOMINE (2022-2026).

Ce texte contient de nombreux passages en gris. Ceux-ci pourront-être passé en première lecture. Ces passages contiennent cependant de nombreuses démonstrations; s'il n'est pas important (ni souhaitable) de les apprendre, cela peut être un exercice instructif d'essayer de les faire soi-même.

I Introduction

Dans ce chapitre, nous étudierons l'espace des matrices symétriques (ou hermitiennes lorsqu'on se place dans \mathbb{C}) par l'intermédiaire des formes bilinéaires symétriques et des formes quadratiques. Cette étude conduit à revisiter la notion de matrices congruentes, déjà rencontrée lors de l'étude des espaces euclidiens.

Dans tout le chapitre, nous nous placerons dans un \mathbb{R} -espace vectoriel E de dimension n finie. L'hypothèse que E est de dimension finie n'est pas toujours nécessaire aux propriétés énoncées; nous expliciterons parfois les passages où elle ne l'est pas.

Les vecteurs seront repérés par des flèches. L'objectif est d'éviter la confusion entre vecteurs et coordonnées, qui peut arriver avec les formes quadratiques.

I.1 Formes bilinéaires et symétrie

Définition

Définition I.1 (Forme bilinéaire).

Une **forme bilinéaire** sur E est une application $\varphi : E \times E \rightarrow \mathbb{R}$ telle que :

- ▷ Pour tout $\vec{x}_0 \in E$, l'application $\varphi_{1,\vec{x}_0} : \vec{y} \in E \mapsto \varphi(\vec{x}_0, \vec{y}) \in \mathbb{R}$ est une forme linéaire sur E (linéarité à droite);
- ▷ Pour tout $\vec{y}_0 \in E$, l'application $\varphi_{2,\vec{y}_0} : \vec{x} \in E \mapsto \varphi(\vec{x}, \vec{y}_0) \in \mathbb{R}$ est une forme linéaire sur E (linéarité à gauche).

Symétrie et antisymétrie

Définition I.2 (Formes symétriques et antisymétriques).

Soit φ une forme bilinéaire. On dit que

- ▷ φ est **symétrique** si $\varphi(\vec{x}, \vec{y}) = \varphi(\vec{y}, \vec{x})$ pour tous $\vec{x}, \vec{y} \in E$.
- ▷ φ est **antisymétrique** si $\varphi(\vec{x}, \vec{y}) = -\varphi(\vec{y}, \vec{x})$ pour tous $\vec{x}, \vec{y} \in E$.

Exemple I.3.

- ▷ Un produit scalaire sur E est une forme bilinéaire symétrique (qui vérifie des conditions supplémentaires : être définie et positive).
- ▷ La forme bilinéaire nulle est à la fois symétrique et antisymétrique.
- ▷ L'application $\varphi : (\vec{x}, \vec{y}) \mapsto \det(\vec{x}, \vec{y})$, où l'on identifie $\vec{x} = (x_1, x_2)$ avec ses coordonnées dans la base canonique $\begin{bmatrix} x_1 \\ x_2 \end{bmatrix}$, est une forme bilinéaire antisymétrique. On peut l'expliciter :

$$\varphi(\vec{x}, \vec{y}) = x_1 y_2 - x_2 y_1.$$

- ▷ L'exemple précédent se généralise en dimension paire. Posons, pour tous $\vec{x}, \vec{y} \in \mathbb{R}^{2n}$,

$$\varphi(\vec{x}, \vec{y}) := (x_1 y_2 - x_2 y_1) + (x_3 y_4 - x_4 y_3) + \cdots + (x_{2n-1} y_{2n} - x_{2n} y_{2n-1}).$$

Alors φ est une forme bilinéaire antisymétrique sur \mathbb{R}^{2n} .

Propriété I.4.

Toute forme bilinéaire se décompose de manière unique sous la forme d'une somme d'une forme bilinéaire symétrique et d'une forme bilinéaire antisymétrique.

PREUVE : Soit φ une forme bilinéaire. Exhibons tout d'abord une décomposition : pour tous $\vec{x}, \vec{y} \in E$, on pose

$$\varphi_S(\vec{x}, \vec{y}) := \frac{\varphi(\vec{x}, \vec{y}) + \varphi(\vec{y}, \vec{x})}{2}, \quad \varphi_A(\vec{x}, \vec{y}) := \frac{\varphi(\vec{x}, \vec{y}) - \varphi(\vec{y}, \vec{x})}{2}.$$

Alors φ_S est bilinéaire symétrique, φ_A est bilinéaire antisymétrique, et $\varphi = \varphi_S + \varphi_A$.

Il reste à vérifier que cette décomposition est unique. Soit $\varphi = \psi_1 + \psi_2$ une décomposition, où ψ_1 est bilinéaire symétrique et ψ_2 bilinéaire antisymétrique. Alors $(\psi_1)_S = \psi_1$ et $(\psi_2)_S = 0$, donc $\varphi_S = (\psi_1)_S = \psi_1$. De même, on obtient $\varphi_A = \psi_2$. La décomposition est donc bien unique. \square

Remarque I.5.

L'espace des formes bilinéaires sur E est un espace vectoriel, et l'on vérifiera qu'il est de dimension finie. Les espaces des formes bilinéaires symétriques (respectivement, antisymétriques) en sont des sous-espaces vectoriels. La propriété ci-dessus revient à dire que l'espace des formes bilinéaires est **somme directe** des espaces des formes symétriques et des formes antisymétriques.

Exemple I.6.

Soit

$$\varphi(\vec{x}, \vec{y}) := x_1 y_1 - x_1 y_2 + 4x_2 y_2.$$

Alors φ est une forme bilinéaire sur \mathbb{R}^2 . On utilise les formules précédentes pour déterminer ses parties paires et impaires. Pour tous $\vec{x}, \vec{y} \in \mathbb{R}^2$,

$$\begin{aligned}\varphi_S(\vec{x}, \vec{y}) &= \frac{\varphi(\vec{x}, \vec{y}) + \varphi(\vec{y}, \vec{x})}{2} \\ &= \frac{x_1y_1 - x_1y_2 + 4x_2y_2 + y_1x_1 - y_1x_2 + 4y_2x_2}{2} \\ &= x_1y_1 - \frac{1}{2}x_1y_2 - \frac{1}{2}x_2y_1 + 4x_2y_2,\end{aligned}$$

et

$$\begin{aligned}\varphi_A(\vec{x}, \vec{y}) &= \frac{\varphi(\vec{x}, \vec{y}) - \varphi(\vec{y}, \vec{x})}{2} \\ &= \frac{x_1y_1 - x_1y_2 + 4x_2y_2 - y_1x_1 + y_1x_2 - 4y_2x_2}{2} \\ &= -\frac{1}{2}x_1y_2 + \frac{1}{2}x_2y_1.\end{aligned}$$

I.2 Matrices associées à une forme bilinéaire

On munit E d'une base \mathcal{B} .

Définition

Propriété I.7 (Matrice associée à une forme bilinéaire).

Soit φ une forme bilinéaire sur E (de dimension finie) et \mathcal{B} une base de E . Pour $\vec{x} \in E$ (respectivement, $\vec{y} \in E$), on note X (respectivement, Y) le vecteur colonne des coordonnées de \vec{x} (respectivement, \vec{y}) dans la base \mathcal{B} .

Il existe une unique matrice $A \in \mathcal{M}_n(\mathbb{R})$ telle que, pour tous $\vec{x}, \vec{y} \in E$,

$$\varphi(\vec{x}, \vec{y}) = X^T A Y.$$

On appelle A la matrice associée à φ dans la base \mathcal{B} . Lorsqu'on ne précise pas la base, celle-ci est sous-entendue (par exemple, $E = \mathbb{R}^n$ et \mathcal{B} est la base canonique).

PREUVE : On reprend exactement la preuve présentée dans le Chapitre 2 pour les produits scalaires.

□

Conséquence I.8.

L'espace des formes bilinéaires sur un espace vectoriel E de dimension n est un espace vectoriel de dimension n^2 .

Symétrie et antisymétrie

Rappelons qu'une matrice A de $\mathcal{M}_n(\mathbb{R})$ est dite symétrique si $A^T = A$. On note $\mathcal{S}_n(\mathbb{R})$ l'ensemble des matrices symétriques réelles, et $\mathcal{A}_n(\mathbb{R})$ l'ensemble des matrices antisymétriques réelles.

Remarque I.9.

- ▷ De manière équivalente, $A = (a_{i,j})_{1 \leq i,j \leq n}$ est symétrique si et seulement si $a_{i,j} = a_{j,i}$ pour tous $1 \leq i, j \leq n$.
- ▷ Pour tout A de $\mathcal{M}_n(\mathbb{R})$, on a $(A^T)^T = A$ (on dit que l'opération de transposition est involutive).
- ▷ Pour tout A de $\mathcal{M}_{p,n}(\mathbb{R})$, on a $(A^T)A \in \mathcal{S}_n(\mathbb{R})$ (**Attention aux dimensions !**)
- ▷ L'ensemble $\mathcal{S}_n(\mathbb{R})$ est un sous-espace vectoriel de $\mathcal{M}_n(\mathbb{R})$.

Il faut savoir démontrer ces propriétés !

Propriété I.10.

Soit φ une forme bilinéaire de E (de dimension finie). Les propriétés suivantes sont équivalentes :

- ▷ φ est symétrique.
- ▷ Il existe une base dans laquelle la matrice de φ est symétrique.
- ▷ La matrice de φ est symétrique dans toute base.

Autrement dit, une forme bilinéaire φ est symétrique si et seulement si ses matrices le sont.

PREUVE : **I implique III** : Supposons φ symétrique. Soient $\mathcal{B} = (\vec{e}_1, \dots, \vec{e}_n)$ une base de E et A la matrice de φ dans cette base. Alors, pour tous $1 \leq i, j \leq n$,

$$A_{ij} = \varphi(\vec{e}_i, \vec{e}_j) = \varphi(\vec{e}_j, \vec{e}_i) = A_{ji},$$

donc A est symétrique.

III implique II : Tout espace vectoriel de dimension finie a une base.

II implique I : Soit \mathcal{B} une base dans laquelle la matrice de φ est symétrique. Soient $\vec{x}, \vec{y} \in E$, de coordonnées respectives X et Y dans \mathcal{B} . Alors

$$\varphi(\vec{x}, \vec{y}) = X^T A Y = X^T A^T Y = (A X)^T Y = Y^T A X = \varphi(\vec{y}, \vec{x}),$$

donc φ est symétrique. □

Remarque I.11.

Cette propriété a un analogue évident pour les formes bilinéaires et matrices antisymétriques.

Remarque I.12.

On peut de plus vérifier que :

- ▷ La matrice de la partie symétrique d'une forme bilinéaire est la partie symétrique de sa matrice.
- ▷ La matrice de la partie antisymétrique d'une forme bilinéaire est la partie antisymétrique de sa matrice.

Il est beaucoup moins risqué, pour éviter les erreurs de calculs, de passer par une matrice de φ pour la décomposer en parties symétrique et antisymétrique.

Exemple I.13.

Reprenons l'exemple

$$\varphi(\vec{x}, \vec{y}) := x_1 y_1 - x_1 y_2 + 4x_2 y_2.$$

Dans la base canonique de \mathbb{R}^2 , cette forme bilinéaire a pour matrice

$$\begin{pmatrix} 1 & -1 \\ 0 & 4 \end{pmatrix}.$$

Les parties symétrique et antisymétrique de cette matrice sont respectivement

$$\begin{pmatrix} 1 & -\frac{1}{2} \\ -\frac{1}{2} & 4 \end{pmatrix} \quad \text{et} \quad \begin{pmatrix} 0 & -\frac{1}{2} \\ \frac{1}{2} & 0 \end{pmatrix}.$$

On retrouve beaucoup plus rapidement les formes bilinéaires calculées précédemment. On profite ici du fait que la matrice condense l'information essentielle définissant une forme bilinéaire, et est donc plus facile à manipuler.

I.3 Endomorphismes associés à une forme bilinéaire

On munit E d'un produit scalaire $\langle \cdot, \cdot \rangle$.

Définition

Soit $f \in L(E)$. Alors l'application $\varphi : (\vec{x}, \vec{y}) \mapsto \langle \vec{x}, f(\vec{y}) \rangle$ définie sur $E \times E$ est une forme bilinéaire sur E .

Le point clef est que toute forme bilinéaire peut s'écrire ainsi, et que l'endomorphisme f caractérise φ . On peut ainsi associer un endomorphisme à toute forme bilinéaire.

Propriété I.14.

Soient E un espace euclidien. Pour toute forme bilinéaire φ sur E , il existe un unique endomorphisme $f \in L(E)$ tel que, pour tous $\vec{x}, \vec{y} \in E$,

$$\varphi(\vec{x}, \vec{y}) = \langle \vec{x}, f(\vec{y}) \rangle.$$

Nous pouvons en donner plusieurs démonstrations, de niveaux d'abstraction variés.

PREUVE : Munissons E d'une base orthonormée \mathcal{B} . Soit A la matrice de φ dans \mathcal{B} , et f l'endomorphisme de matrice A dans \mathcal{B} . Alors, pour tous $\vec{x}, \vec{y} \in E$,

$$\varphi(\vec{x}, \vec{y}) = X^T AY = \langle \vec{x}, f(\vec{y}) \rangle.$$

Il reste à montrer qu'un tel endomorphisme est unique. Soient f, g deux endomorphismes associés à la même forme bilinéaire. Soient A et B leurs matrices respectives dans \mathcal{B} . Alors, pour tous $\vec{x}, \vec{y} \in E$

$$X^T AY = \varphi(\vec{x}, \vec{y}) = X^T BY,$$

donc $A = B$, donc $f = g$. □

PREUVE : Donnons une autre démonstration. $L(E)$ est un espace vectoriel de dimension n^2 , de même que l'espace vectoriel $\text{Bil}(E)$ des formes bilinéaires sur E .

À tout endomorphisme f , on associe une forme bilinéaire $\varphi_f : (\vec{x}, \vec{y}) \mapsto \langle \vec{x}, f(\vec{y}) \rangle$. L'application $f \mapsto \varphi_f$ est linéaire. On veut montrer que cette application est bijective ; par égalité des dimension, il suffit de montrer son injectivité.

Soit $f \in L(E)$ telle que $\varphi_f = 0$. On veut montrer que $f = 0$. Soit $\vec{y} \in E$. Alors

$$\|f(\vec{y})\|^2 = \langle f(\vec{y}), f(\vec{y}) \rangle = \varphi_f(f(\vec{y}), \vec{y}) = 0,$$

donc $f(\vec{y}) = \vec{0}$, ce qu'il fallait montrer. □

PREUVE : Donnons une troisième démonstration, à un niveau d'abstraction supérieur.

Soit φ une forme bilinéaire. Par linéarité à gauche, pour tout $\vec{y} \in E$, l'application $\ell_{\vec{y}} : \vec{x} \mapsto \varphi(\vec{x}, \vec{y})$ est une forme linéaire. Par linéarité à droite, l'application $\ell : \vec{y} \mapsto \ell_{\vec{y}}$ est linéaire de E dans son dual E^* .

Par le théorème de représentation de Riesz, il existe un isomorphisme $\Theta : E^* \rightarrow E$ tel que toute forme linéaire $\ell \in E^*$ et de la forme $\langle \vec{x}, \cdot \rangle$, où $\vec{x} = \Theta(\ell)$.

Posons $f := \Theta \circ \ell \in L(E)$. Alors, pour tous $\vec{x}, \vec{y} \in E$,

$$\langle \vec{x}, f(\vec{y}) \rangle = \langle \vec{x}, \Theta(\ell(\vec{y})) \rangle = \langle \vec{x}, \Theta(\ell_{\vec{y}}) \rangle = \ell_{\vec{y}}(\vec{x}) = \varphi(\vec{x}, \vec{y}).$$

On remarquera au passage que la seconde démonstration de ce fait ressemble à la démonstration du théorème de représentation de Riesz ; ce n'est pas un hasard ! \square

Remarque I.15.

Cette représentation d'une forme bilinéaire a plusieurs avantages :

- ▷ On a associé à φ et $\langle \cdot, \cdot \rangle$ un unique endomorphisme f . Les constructions liées aux endomorphismes seront donc naturelles : valeurs propres, espaces propres... Elles ne sont pas artificiellement associées à φ via une de ses matrices.
- ▷ Elle ne nécessite pas de passer par une matrice, et donc restera valable sous conditions en dimension infinie, par exemple en analyse hilbertienne.

Pour conclure : On dispose donc de trois points de vue distincts sur les formes bilinéaires :

- ▷ Il s'agit par définition d'une application $\varphi : E \times E \rightarrow \mathbb{R}$ vérifiant certaines propriétés (linéarité à gauche et à droite).
- ▷ Si on dispose d'une base \mathcal{B} de E , on peut lui associer une matrice $A \in \mathcal{M}_n(\mathbb{R})$.
- ▷ Si on dispose d'un produit scalaire $\langle \cdot, \cdot \rangle$ sur E , on peut lui associer un endomorphisme $f \in L(E)$.

La matrice A dépend de la base \mathcal{B} , et l'endomorphisme f dépend du produit scalaire. Ceci dit, si \mathcal{B} est orthonormée¹, ces deux représentations sont cohérentes : la matrice de φ est alors aussi la matrice de f . Dans ce cas, pour tous $\vec{x}, \vec{y} \in E$,

$$\varphi(\vec{x}, \vec{y}) = X^T A Y = \langle \vec{x}, f(\vec{y}) \rangle.$$

Savoir-faire :

- ▷ Identifier les formes bilinéaires ;
- ▷ Passer d'une représentation à une autre².

1. C'est-à-dire que la base est "compatible" avec le produit scalaire.
 2. Par exemple en passant par une matrice.

Symétrie et antisymétrie

Nous venons de voir que, pour tout $f \in L(E)$, l'application $(\vec{x}, \vec{y}) \mapsto \langle \vec{x}, f(\vec{y}) \rangle$ est bilinéaire. Cette application est symétrique si et seulement si, pour tous $\vec{x}, \vec{y} \in E$,

$$\langle \vec{x}, f(\vec{y}) \rangle = \langle f(\vec{x}), \vec{y} \rangle.$$

Ceci motive la définition suivante.

Définition I.16 (Endomorphisme adjoint).

Soient E un espace euclidien et $f \in L(E)$.

▷ On appelle **adjoint** de f (pour le produit scalaire donné) l'application linéaire $f^* \in L(E)$ telle que, pour tous $\vec{x}, \vec{y} \in E$,

$$\langle \vec{x}, f(\vec{y}) \rangle = \langle f^*(\vec{x}), \vec{y} \rangle.$$

▷ On dit que f est **autoadjoint** s'il est égal à son adjoint.

Cette définition cache une propriété à démontrer : celle que l'application linéaire f^* existe et est unique. Nous allons utiliser la propriété suivante :

Propriété I.17.

Soient E un espace euclidien et $f \in L(E)$. Il existe une unique application linéaire $f^* \in L(E)$ telle que $\langle \vec{x}, f(\vec{y}) \rangle = \langle f^*(\vec{x}), \vec{y} \rangle$ pour tous $\vec{x}, \vec{y} \in E$.

Si \mathcal{B} est une base **orthonormée** de E et A la matrice de f dans \mathcal{B} , alors la matrice de f^* dans \mathcal{B} est A^T .

PREUVE : Montrons l'existence et l'unicité de l'adjoint. Soit $f \in L(E)$. Alors l'application $(\vec{x}, \vec{y}) \mapsto \langle f(\vec{x}), \vec{y} \rangle$ est une forme bilinéaire. Il existe donc un unique endomorphisme f^* tel que, pour tous $\vec{x}, \vec{y} \in E$,

$$\langle f(\vec{x}), \vec{y} \rangle = \langle \vec{x}, f^*(\vec{y}) \rangle.$$

La symétrie du produit scalaire permet de conclure.

Soit \mathcal{B} une base orthonormée. On note X les coordonnées du vecteur \vec{x} et Y celles de \vec{y} . Alors

$$\langle x, f(y) \rangle = X^T(AY) = X^T(A^T)^T Y = (A^T X)Y.$$

Il existe un unique endomorphisme f^* tel que $\langle f^*(\vec{x}), \vec{y} \rangle = (A^T X)Y$ pour tous $\vec{x}, \vec{y} \in E$; c'est l'endomorphisme dont la matrice dans la base \mathcal{B} est A^T . \square

Propriété I.18. Soient \mathcal{B} est une base **orthonormée** de E et A la matrice de f dans \mathcal{B} . Il y a équivalence entre :

- ▷ la forme bilinéaire $(\vec{x}, \vec{y}) \mapsto \langle \vec{x}, f(\vec{y}) \rangle$ est symétrique ;
- ▷ f est autoadjoint ;
- ▷ $A^T = A$.

PREUVE : L'équivalence entre les deux premiers points découle de la définition d'endomorphisme autoadjoint ; l'équivalence entre les deux derniers points de ce qui précède ($f = f^*$ si et seulement si $A = A^T$). \square

Exemple I.19.

Soit r_θ la rotation d'angle θ dans \mathbb{R}^2 . Alors $r_{-\theta}$ est une isométrie, donc préserve le produit scalaire. Par conséquent, pour tous $x, y \in \mathbb{R}^2$,

$$\langle x, r_\theta(y) \rangle = \langle r_{-\theta}(x), r_{-\theta} \circ r_\theta(y) \rangle = \langle r_{-\theta}(x), y \rangle.$$

Donc l'adjoint de la rotation d'angle θ est la rotation d'angle $-\theta$. Par conséquent, r_θ est autoadjoint si et seulement si $r_\theta = \text{id}$ ou $r_\theta = -\text{id}$.

On peut prendre le point de vue matriciel : la base canonique de \mathbb{R}^2 est orthonormée. La matrice de l'adjoint de r_θ est la transposée de la matrice de r_θ , et on vérifie qu'il s'agit bien de la matrice de $r_{-\theta}$.

Exercice 1. Montrez matriciellement qu'une symétrie axiale de \mathbb{R}^2 est autoadjointe.

Exercice 2. Montrez plus généralement qu'une symétrie orthogonale est autoadjointe.

Propriété I.20.

Si $f \in L(E)$, alors :

- ▷ $(f^*)^* = f$.
- ▷ f et f^* ont les mêmes valeurs propres.
- ▷ $\text{Ker}(f^*) = (\text{Im}(f))^\perp$.
- ▷ $\text{Im}(f^*) = (\text{Ker}(f))^\perp$.

PREUVE : **Premier point :** Cela est équivalent au fait que $(A^T)^T = A$. On peut démontrer le même résultat sans utiliser de matrices. Pour tous $\vec{x}, \vec{y} \in E$,

$$\langle \vec{x}, (f^*)^*(\vec{y}) \rangle = \langle (f^*)^*(\vec{y}), \vec{x} \rangle = \langle \vec{y}, f^*(\vec{x}) \rangle = \langle f^*(\vec{x}), \vec{y} \rangle = \langle \vec{x}, f(\vec{y}) \rangle.$$

Ceci étant vrai pour tous $\vec{x}, \vec{y} \in E$, on obtient $(f^*)^* = f$.

Deuxième point : Les valeurs propres de f (respectivement f^*) sont par définition celles de A (respectivement A^T). Or une matrice et sa transposée ont les mêmes valeurs propres.

Troisième point : Montrons tout d'abord que $\text{Ker}(f^*) \subset (\text{Im}(f))^\perp$. Soit $\vec{x} \in \text{Ker}(f^*)$ et $\vec{y} \in \text{Im}(f)$; on veut montrer que $\langle \vec{x}, \vec{y} \rangle = 0$. Soit \vec{z} tel que $\vec{y} = f(\vec{z})$. Alors

$$\langle \vec{x}, \vec{y} \rangle = \langle \vec{x}, f(\vec{z}) \rangle = \langle f^*(\vec{x}), \vec{z} \rangle = \langle \vec{0}, \vec{z} \rangle = 0.$$

Ceci étant valable pour tout $\vec{y} \in \text{Im}(f)$, on a bien $\vec{x} \in (\text{Im}(f))^\perp$. Ceci étant valable pour tout $\vec{x} \in \text{Ker}(f^*)$, on a bien l'inclusion attendue.

Réciproquement, soit $\vec{x} \in (\text{Im}(f))^\perp$. Alors

$$\|f^*(\vec{x})\|^2 = \langle f^*(\vec{x}), f^*(\vec{x}) \rangle = \langle \vec{x}, f(f^*(\vec{x})) \rangle = 0,$$

car $f(f^*(\vec{x})) \in \text{Im}(f)$ et $\vec{x} \in (\text{Im}(f))^\perp$. Donc $\|f^*(\vec{x})\| = 0$, donc $f^*(\vec{x}) = \vec{0}$, donc $\vec{x} \in \text{Ker}(f^*)$, ce qu'il fallait montrer.

Quatrième point : On applique le point précédent à f^* , et on utilise le premier point.

$$(\text{Im}(f^*))^\perp = \text{Ker}((f^*)^*) = \text{Ker}(f).$$

Enfin, on prend les orthogonaux :

$$(\text{Ker}(f))^\perp = ((\text{Im}(f^*))^\perp)^\perp = \text{Im}(f^*)$$

□

I.4 Formes quadratiques

La donnée d'un produit scalaire est équivalente à la donnée d'une norme euclidienne. Nous généralisons cela en étudiant le lien entre formes bilinéaires et formes quadratiques.

Définition I.21 (Forme quadratique).

On dit qu'une application $q : E \rightarrow \mathbb{R}$ est **forme quadratique** s'il existe une forme bilinéaire φ sur E telle que, pour tout $\vec{x} \in E$,

$$q(\vec{x}) = \varphi(\vec{x}, \vec{x}).$$

On dit que q est la forme quadratique associée à φ .

Exemple I.22.

Si $\langle \cdot, \cdot \rangle$ est un produit scalaire, alors $q(\vec{x}) = \|\vec{x}\|^2 = \langle \vec{x}, \vec{x} \rangle$ est sa forme quadratique associée. En particulier, $q(\vec{x}) = \sum_{i=1}^n x_i^2$ est une forme quadratique sur \mathbb{R}^n .

La forme quadratique nulle est associée à la forme bilinéaire nulle.

Soit $\varphi(\vec{x}, \vec{y}) = x_1 y_2 - x_2 y_1$ la forme bilinéaire "déterminant" sur \mathbb{R}^2 . La forme quadratique associée est $q(\vec{x}) = x_1 x_2 - x_2 x_1 = 0$.

Soit $\varphi(\vec{x}, \vec{y}) = x_1 y_1 - x_1 y_2 + 4x_2 y_2$. La forme quadratique associée est $q(\vec{x}) = x_1^2 - x_1 x_2 + 4x_2^2$. On voit apparaître des signes négatifs ainsi que des termes croisés (monôme $x_1 x_2$).

On remarque qu'une même forme quadratique peut être associée à plusieurs formes bilinéaires : c'est le cas de la forme quadratique nulle sur \mathbb{R}^2 , associée à la fois à la forme bilinéaire nulle et au déterminant. Plus généralement :

Lemme I.23.

Soient E un espace vectoriel réel, φ une forme bilinéaire sur E , et q la forme quadratique associée à φ . Alors $q = 0$ si et seulement si φ est antisymétrique.

PREUVE : Supposons φ antisymétrique. Soit $\vec{x} \in E$. Alors $q(\vec{x}) = \varphi(\vec{x}, \vec{x}) = -\varphi(\vec{x}, \vec{x}) = -q(\vec{x})$, donc $q(\vec{x}) = 0$.

Réciproquement, supposons que $q = 0$. Alors $q(\vec{x}) = \varphi(\vec{x}, \vec{x}) = 0$ pour tout $x \in E$. Maintenant, pour tous $\vec{x}, \vec{y} \in E$,

$$q(\vec{x} + \vec{y}) = \varphi(\vec{x} + \vec{y}, \vec{x} + \vec{y}) = \varphi(\vec{x}, \vec{x}) + \varphi(\vec{x}, \vec{y}) + \varphi(\vec{y}, \vec{x}) + \varphi(\vec{y}, \vec{y}).$$

Or $q(\vec{x} + \vec{y}) = q(\vec{x}) = q(\vec{y}) = 0$, donc $\varphi(\vec{x}, \vec{y}) + \varphi(\vec{y}, \vec{x}) = 0$ pour tous $\vec{x}, \vec{y} \in E$. Donc φ est antisymétrique. \square

On en déduit que l'on peut représenter les formes quadratiques par des matrices symétriques.

Théorème I.24.

Soit q une forme quadratique de E . Il existe une unique forme bilinéaire symétrique φ qui lui est associée. Cette forme bilinéaire φ satisfait la **formule de polarisation** :

$$\forall \vec{x}, \vec{y} \in E, \quad \varphi(\vec{x}, \vec{y}) = \frac{1}{2} (q(\vec{x} + \vec{y}) - q(\vec{x}) - q(\vec{y})).$$

On appelle cette forme bilinéaire φ la **forme polaire** associée à q .

PREUVE : Soit φ une forme bilinéaire associée à q . On peut décomposer φ en une partie symétrique φ_S et une partie antisymétrique φ_A . Alors, pour tout $\vec{x} \in E$,

$$q(\vec{x}) = \varphi(\vec{x}, \vec{x}) = \varphi_S(\vec{x}, \vec{x}) + \varphi_A(\vec{x}, \vec{x}) = \varphi_S(\vec{x}, \vec{x}),$$

donc q est associée à φ_S . Cela montre l'existence d'une telle forme bilinéaire symétrique.

Si φ est une forme bilinéaire symétrique associée à q , on vérifie par le calcul qu'elle coïncide avec la forme polaire, ce qui montre l'unicité d'une telle forme bilinéaire symétrique. \square

Exemple I.25.

- ▷ Dans le cas d'un produit scalaire, on retrouve la formule de polarisation associée.
- ▷ Si $q(\vec{x}) = x_1^2 - x_1x_2 + 4x_2^2$ sur \mathbb{R}^2 , alors q est associée à la forme bilinéaire de matrice $\begin{pmatrix} 1 & -1 \\ 0 & 4 \end{pmatrix}$.
Cette matrice n'est pas symétrique, mais sa partie symétrique est $\begin{pmatrix} 1 & -1/2 \\ -1/2 & 4 \end{pmatrix}$, qui est la matrice symétrique associée à q .
- ▷ Si $q(\vec{x}) = x_1^2 + x_1x_2 + x_2^2$ sur \mathbb{R}^2 , on trouve $\varphi(\vec{x}, \vec{y}) = x_1y_1 + \frac{1}{2}x_1y_2 + \frac{1}{2}x_2y_1 + x_2y_2$.
- ▷ Si $q(\vec{x}) = x_1^2 - x_2^2$ sur \mathbb{R}^2 , on trouve $\varphi(\vec{x}, \vec{y}) = x_1y_1 - x_2y_2$.

Conséquence I.26. L'ensemble des formes quadratiques sur E forme un espace vectoriel, de même dimension de $\mathcal{S}_n(\mathbb{R})$, soit $\frac{n(n+1)}{2}$.

Pour conclure : On dispose donc de quatre points de vue distincts sur les formes bilinéaires symétriques :

- ▷ Il s'agit par définition d'une application $\varphi : E \times E \rightarrow \mathbb{R}$ vérifiant certaines propriétés (linéarité à gauche et à droite et *symétrie*).
- ▷ Si on dispose d'une base \mathcal{B} de E , on peut lui associer une matrice *symétrique* $A \in \mathcal{S}_n(\mathbb{R})$.
- ▷ Si on dispose d'un produit scalaire $\langle \cdot, \cdot \rangle$ sur E , on peut lui associer un endomorphisme *autoadjoint* $f \in L(E)$.
- ▷ Finalement, on peut lui associer une forme quadratique q .

La matrice A dépend de la base \mathcal{B} , et l'endomorphisme f dépend du produit scalaire. Ceci dit, si \mathcal{B} est orthonormée³, ces deux représentations sont cohérentes : la matrice de φ est alors aussi la matrice de f . En particulier, la matrice de f est alors symétrique.

Savoir-faire : Passer d'une représentation à une autre⁴. Les trois premiers points ont déjà été vus pour des formes bilinéaires générales ; la nouveauté concerne les formes quadratiques.

II Réductions des formes quadratiques

II.1 Matrices d'une forme quadratique

Définition

Définition II.1 (Matrice associée à une forme quadratique dans une base).

3. C'est-à-dire que la base est "compatible" avec le produit scalaire.

4. Par exemple en passant par une matrice.

Soit \mathcal{B} une base de E . Soit q une forme quadratique de E et φ sa forme bilinéaire symétrique associée. La **matrice** de q dans \mathcal{B} est la matrice $A \in \mathcal{S}_n(\mathbb{R})$ de φ dans \mathcal{B} .

Alors, pour tout $\vec{x} \in E$, en notant X ses coordonnées dans \mathcal{B} et $A = (A_{ij})_{1 \leq i, j \leq n}$,

$$q(\vec{x}) = X^T A X = \sum_{i=1}^n A_{ii} X_i^2 + 2 \sum_{1 \leq i < j \leq n} A_{ij} X_i X_j.$$

Si l'on n'impose pas la symétrie de A , alors elle on perd l'unicité de la matrice dans une base donnée.

PREUVE : Il ne reste à démontrer que la seconde égalité. Or, la matrice A étant symétrique,

$$\begin{aligned} \sum_{i,j=1}^n A_{ij} X_i X_j &= \sum_{1 \leq i < j \leq n} A_{ij} X_i X_j + \sum_{1 \leq i=j \leq n} A_{ij} X_i X_j + \sum_{1 \leq j < i \leq n} A_{ij} X_i X_j \\ &= \sum_{1 \leq i < j \leq n} A_{ij} X_i X_j + \sum_{1 \leq i \leq n} A_{ii} X_i^2 + \sum_{1 \leq i < j \leq n} A_{ji} X_i X_j \\ &= 2 \sum_{1 \leq i < j \leq n} A_{ij} X_i X_j + \sum_{1 \leq i \leq n} A_{ii} X_i^2. \end{aligned}$$

□

Exemple II.2.

▷ Si $q(\vec{x}) = 2x_1^2 + 2x_1x_2 + x_2^2$ sur \mathbb{R}^2 , sa matrice dans la base canonique est

$$\begin{pmatrix} 2 & 1 \\ 1 & 1 \end{pmatrix}.$$

Les coefficients devant les termes quadratiques x_i^2 se retrouvent sur la diagonale, tandis que les termes croisés $x_i x_j$ se retrouvent divisés par deux et hors de la diagonale.

▷ Si $q(\vec{x}) = x_1x_2$ sur \mathbb{R}^2 , sa matrice dans la base canonique est

$$\begin{pmatrix} 0 & 1/2 \\ 1/2 & 0 \end{pmatrix}.$$

La représentation matricielle permet d'identifier facilement les formes quadratiques sans avoir à expliciter une forme bilinéaire.

Propriété II.3.

Soit $q : E \rightarrow \mathbb{R}$. Soit \mathcal{B} une base de E . L'application q est une forme quadratique si et seulement si on peut trouver une matrice $A = (A_{ij})_{1 \leq i, j \leq n} \in \mathcal{M}_n(\mathbb{R})$ telle que, pour tout $\vec{x} \in E$,

$$q(\vec{x}) = X^T A X = \sum_{\substack{1 \leq i \leq n \\ 1 \leq j \leq n}} A_{ij} X_i X_j.$$

PREUVE : Si q est une forme quadratique, soit φ une forme bilinéaire telle que $q(\vec{x}) = \varphi(\vec{x}, \vec{x})$ pour tout $\vec{x} \in E$. Soit A la matrice de φ dans la base \mathcal{B} . Alors A convient.

Réciproquement, si A est une telle matrice, alors on pose $\varphi(\vec{x}, \vec{y}) := X^T A Y$. Cette application est une forme bilinéaire, et $q(\vec{x}) = \varphi(\vec{x}, \vec{x})$ pour tout $\vec{x} \in E$. □

Congruence

Comme dans les premiers chapitres, on s'intéresse maintenant à l'ensemble de toutes les matrices associées à une forme quadratique donnée. Cet ensemble forme une famille de matrices de $\mathcal{S}_n(\mathbb{R})$ qui définit une classe d'équivalence dans l'ensemble des matrices symétriques réelles : on parle de matrices congruentes.

Propriété II.4.

Soit q une forme quadratique sur E et $\mathcal{B}, \mathcal{B}'$ deux bases de E . Soient A, B respectivement les matrices de q dans $\mathcal{B}, \mathcal{B}'$. Soit P la matrice de passage de \mathcal{B} à \mathcal{B}' .

Alors $B = P^T A P$.

PREUVE : Soit $\vec{x} \in E$. Notons X (respectivement X') ses coordonnées dans \mathcal{B} (respectivement \mathcal{B}'). Rappelons que $X = P X'$. Par conséquent,

$$q(\vec{x}) = X^T A X = (P X')^T A (P X') = (X')^T (P^T A P) X'.$$

Alors $P^T A P$ est une matrice symétrique réelle, et donc la matrice associée à q dans \mathcal{B}' (qui est unique). On en déduit que $B = P^T A P$. \square

Définition II.5 (Rappel : Congruence).

On dit que deux matrices symétriques réelles A et B de $\mathcal{S}_n(\mathbb{R})$ sont **congruentes** s'il existe $P \in GL_n(\mathbb{R})$ telle que $A = P^T B P$.

Cette relation définit une relation d'équivalence sur l'ensemble $\mathcal{S}_n(\mathbb{R})$

PREUVE : Voir le cours sur le produit scalaire. \square

Deux matrices symétriques sont congruentes si elles représentent la même forme quadratique, éventuellement dans des bases différentes.

Il s'agit maintenant de déterminer les "formes normales" de ces classes d'équivalence. Le premier avantage est de pouvoir bénéficier, si nécessaire, d'une écriture matricielle particulièrement simple d'une forme quadratique donnée, quitte à devoir changer de base.

Le second avantage est de trouver des invariants (si possible complets) de congruence. Si une fonctionnelle $F : \mathcal{S}_n(\mathbb{R}) \rightarrow \mathbb{R}$ (ou, plus généralement, à valeurs dans un ensemble quelconque) est constante sur les classes de congruence, alors pour toute forme quadratique q et matrice associée A , la valeur de $F(A)$ ne dépendra pas de la base choisie. Ainsi, elle ne dépend que de q , et est donc définie pour la forme quadratique.

Remarque II.6.

Vous rappelez-vous les autres relations d'équivalences déjà vues ? Leurs invariants ? Leur système d'invariants total ? Les représentations "normales" ?

Propriété II.7.

Le rang est un invariant de congruence.

PREUVE : Deux matrices congruentes sont équivalentes, et donc de même rang. \square

On peut donc définir le rang associé à une forme quadratique.

Définition II.8 (Rang d'une forme quadratique).

Soit q une forme quadratique de matrice associée $A \in \mathcal{S}_n(\mathbb{R})$ dans une base donnée. Alors on appelle **rang** de q , noté $\text{rg}(q)$ le rang de sa matrice associée A .

De plus si $\text{rg}(A) = \dim(E)$, on dit que q est **non dégénérée**.

Exemple II.9.

Si q est le carré d'une norme euclidienne, alors sa matrice associée est à valeurs propres strictement positives, donc est inversible. Donc q est non dégénérée.

Une forme quadratique non dégénérée peut néanmoins s'annuler ailleurs qu'en 0. Par exemple, choisissons $q(\vec{x}) = x_1^2 - x_2^2$ dans \mathbb{R}^2 . Alors la matrice associée dans la base canonique est $\begin{pmatrix} 1 & 0 \\ 0 & -1 \end{pmatrix}$, qui est de rang 2. La forme quadratique q est donc non dégénérée. Pourtant, $q(\vec{x}) = 0$ pour tout vecteur \vec{x} de la forme (t, t) ou $(t, -t)$, avec $t \in \mathbb{R}$.

Le fait de ne pas être dégénéré peut se définir directement à partir de la forme bilinéaire symétrique associée à une forme quadratique.

Lemme II.10.

Soit q une forme quadratique de forme bilinéaire symétrique associée φ . Alors q est non dégénérée si et seulement si, pour tout $\vec{x} \in E \setminus \{\vec{0}\}$, il existe $\vec{y} \in E$ tel que $\varphi(\vec{x}, \vec{y}) \neq 0$.

PREUVE : Soit q une forme quadratique, φ la forme bilinéaire symétrique associée, \mathcal{B} une base de E et A la matrice associée à q dans la base \mathcal{B} . On note X les coordonnées d'un vecteur x dans la base \mathcal{B} .

Supposons q non dégénérée. Soit $x \in E \setminus \{\vec{0}\}$. La matrice A étant inversible, soit Y tel que $AY = X$, et soit \vec{y} le vecteur associé. Alors $\varphi(\vec{x}, \vec{y}) = X^T AY = X^T X > 0$.

Réciproquement, supposons que φ vérifie la propriété énoncée. Soit $X \in \text{Ker}(A)$. Alors, pour tout $\vec{y} \in E$,

$$\varphi(\vec{x}, \vec{y}) = X^T AY = (AX)^T Y = 0Y = 0.$$

Donc $\varphi(\vec{x}, \vec{y}) = 0$ pour tout $\vec{y} \in E$. Ce n'est possible que si $\vec{x} = \vec{0}$, donc $\text{Ker}(A) = \{0\}$, donc A est inversible. \square

II.2 Réduction des endomorphismes autoadjoints réels

Nous démontrons dans ce paragraphe le “théorème spectral”⁵ qui dit que les matrices symétriques à coefficients réels sont diagonalisables dans une base orthonormée.

Propriété II.11.

Soit $A \in \mathcal{S}_n(\mathbb{R})$. Alors les valeurs propres (sur \mathbb{C}) de A sont réelles.

PREUVE : Dans le chapitre 2, nous avons montré que si A est la matrice d'un produit scalaire, alors ses valeurs propres sont réelles. Cette démonstration n'utilisait que le fait que A était symétrique, et s'applique donc sans modification. \square

5. En réalité, le théorème spectral est un théorème valable aussi en dimension infinie, que vous apprendrez peut-être plus tard.

Conséquence II.12.

Soit $f \in L(E)$. Si f est autoadjoint, alors toutes ses valeurs propres sont réelles.

Pour trouver une base orthonormée de vecteurs propres, il est nécessaire de vérifier que les sous-espaces propres sont orthogonaux. C'est le but de la proposition suivante :

Propriété II.13.

Soit $f \in L(E)$. Si f est autoadjoint, alors ses sous-espaces propres sont orthogonaux.

PREUVE : Soient $\lambda \neq \mu$ deux valeurs propres distinctes de f , et E_λ, E_μ les sous-espaces propres associés. Soient $\vec{v}_\lambda \in E_\lambda$ et $\vec{v}_\mu \in E_\mu$. Alors

$$\begin{aligned}\langle \vec{v}_\lambda, f(\vec{v}_\mu) \rangle &= \langle \vec{v}_\lambda, \mu \vec{v}_\mu \rangle = \mu \langle \vec{v}_\lambda, \vec{v}_\mu \rangle, \\ \langle \vec{v}_\lambda, f(\vec{v}_\mu) \rangle &= \langle f(\vec{v}_\lambda), \vec{v}_\mu \rangle = \lambda \langle \vec{v}_\lambda, \vec{v}_\mu \rangle.\end{aligned}$$

Donc $\mu \langle \vec{v}_\lambda, \vec{v}_\mu \rangle = \lambda \langle \vec{v}_\lambda, \vec{v}_\mu \rangle$. Or $\lambda \neq \mu$, donc $\langle \vec{v}_\lambda, \vec{v}_\mu \rangle = 0$. Ceci étant valable pour tous $\vec{v}_\lambda \in E_\lambda$ et $\vec{v}_\mu \in E_\mu$, les espaces E_λ, E_μ sont orthogonaux. \square

Théorème II.14.

En dimension finie, tout endomorphisme autoadjoint de E est diagonalisable dans une base orthonormée.

PREUVE : Le plus efficace est de procéder par récurrence sur la dimension de E .

Pour $n = 1$, tout endomorphisme est diagonalisable et la propriété est vraie.

Soit $n > 1$. Supposons la propriété vraie pour tout espace euclidien F de dimension $k < n$ et tout endomorphisme autoadjoint $g \in L(F)$. Soit $f \in L(E)$ un endomorphisme autoadjoint. On sait qu'il admet des valeurs propres réelles et que ses sous-espaces propres sont orthogonaux.

Soit λ une valeur propre de f , et $E_\lambda := \text{Ker}(f - \lambda \text{id})$ son sous-espace propre associé. Soit $d := \dim(E_\lambda) \geq 1$. Alors $E = E_\lambda \oplus E_\lambda^\perp$ avec $\dim(E_\lambda^\perp) = n - d < n$. Pour appliquer l'hypothèse de récurrence, il reste à démontrer que E_λ^\perp est stable par f .

Soit $\vec{v} \in E_\lambda^\perp$. On veut montrer que $f(\vec{v}) \in E_\lambda^\perp$, c'est à dire que $\langle f(\vec{v}), \vec{y} \rangle = 0$ pour tout $\vec{y} \in E_\lambda$. Or, comme f est autoadjointe, pour tout $\vec{y} \in E_\lambda$,

$$\langle f(\vec{x}), \vec{y} \rangle = \langle \vec{x}, f(\vec{y}) \rangle = \langle \vec{x}, \lambda \vec{y} \rangle = \lambda \langle \vec{x}, \vec{y} \rangle = 0.$$

Donc E_λ^\perp est donc bien stable par f . De plus, la restriction de f à E_λ^\perp est aussi autoadjointe. Donc, par l'hypothèse de récurrence, on peut trouver une base orthonormée de E_λ^\perp dans laquelle la matrice de la restriction de f est diagonale. On la complète par une base orthonormée de E_λ en une base orthonormée de E , ce qui est possible car E_λ et E_λ^\perp sont orthogonaux ; la matrice de f est alors diagonale dans la base orthonormée ainsi créée. \square

II.3 Réduction d'une forme quadratique dans un base orthonormée

On munit E d'un produit scalaire. On s'intéresse pour l'instant aux matrices représentant q dans une base orthonormée⁶.

6. De façon équivalente, on ajoute la contrainte $P \in O_n(\mathbb{R})$ dans la définition de la notion de congruence.

À q est associé un endomorphisme autoadjoint f tel que, pour tout $\vec{x} \in E$,

$$q(\vec{x}) = \langle \vec{x}, f(\vec{x}) \rangle.$$

Soient $\lambda_1 \leq \dots \leq \lambda_n$ les valeurs propres de f avec multiplicité.

Soit A la matrice de q dans une base orthonormée. Alors A est aussi la matrice de f dans la même base, donc a les mêmes valeurs propres avec la même multiplicité. Autrement dit, tant que l'on travaille avec des bases orthonormées, le spectre de la matrice de A ne dépend pas de la base.

On peut aussi retrouver directement cette propriété sans introduire d'endomorphisme.

Soient q une forme quadratique sur E et $\mathcal{B}, \mathcal{B}'$ deux bases orthonormées de E . Soient A, B respectivement les matrices de q dans $\mathcal{B}, \mathcal{B}'$.

Soit P la matrice de passage de \mathcal{B} à \mathcal{B}' . Alors $B = P^T A P$. Or P est une matrice de passe entre deux bases orthonormées, donc est une matrice orthogonale. Donc $P^T = P^{-1}$, et $B = P^{-1} A P$. Donc A et B sont semblables, et ont les mêmes valeurs propres avec la même multiplicité.

Le théorème spectral implique que l'on peut trouver une base orthonormée dans laquelle la matrice de q est aussi simple que possible, étant donnée cette contrainte.

Théorème 1.

Soit $q : E \rightarrow \mathbb{R}$ une forme quadratique et A la matrice de q dans une base orthonormée \mathcal{B} . Alors il existe une base **orthonormée** \mathcal{B}' et des réels $\lambda_1 \leq \dots \leq \lambda_n$ tels que, pour tout $\vec{x} \in E$ de coordonnées X dans \mathcal{B}' ,

$$q(\vec{x}) = \sum_{i=1}^n \lambda_i X_i^2.$$

De plus, les $(\lambda_i)_{1 \leq i \leq n}$ sont les valeurs propres de A avec multiplicité.

PREUVE : Soit f l'endomorphisme autoadjoint associé à q . Comme \mathcal{B} est orthonormée, les λ_i sont les valeurs propres de f avec multiplicité. Soit \mathcal{B}' une base orthonormée dans laquelle la matrice D de f est diagonale avec coefficients $\lambda_1 \leq \dots \leq \lambda_n$. Alors, pour tout $x \in E$, en notant X ses coordonnées dans la base \mathcal{B}' ,

$$q(\vec{x}) = \langle \vec{x}, f(\vec{x}) \rangle = X^T D X = \sum_{i=1}^d \lambda_i X_i^2$$

□

Conséquence II.15.

Soit A la matrice d'une forme quadratique q dans une base orthonormée. Soient λ_{\max} la plus grande valeur propre de A et λ_{\min} la plus petite valeur propre de A . Alors, pour tout $\vec{x} \in E$,

$$\lambda_{\min} \|\vec{x}\|^2 \leq q(\vec{x}) \leq \lambda_{\max} \|\vec{x}\|^2.$$

Exemple II.16.

La forme quadratique $q(\vec{u}) = xy$ est représentée par la matrice $\begin{pmatrix} 0 & \frac{1}{2} \\ \frac{1}{2} & 0 \end{pmatrix}$ dans la base canonique (qui est orthonormée). Cette matrice a pour valeurs propres $\pm \frac{1}{2}$, donc pour tous $x, y \in \mathbb{R}$,

$$-\frac{x^2 + y^2}{2} \leq xy \leq \frac{x^2 + y^2}{2}.$$

La forme quadratique $q(\vec{u}) = 2x^2 + 2xy + y^2$ est représentée par la matrice $\begin{pmatrix} 2 & 1 \\ 1 & 1 \end{pmatrix}$ dans la base canonique. Cette matrice a pour valeurs propres $\frac{3 \pm \sqrt{5}}{2}$, donc pour tous $x, y \in \mathbb{R}$,

$$\frac{3 - \sqrt{5}}{2}(x^2 + y^2) \leq 2x^2 + 2xy + y^2 \leq \frac{3 + \sqrt{5}}{2}(x^2 + y^2).$$

II.4 Réduction d'une forme quadratique dans un base quelconque

Si l'espace E est muni d'un produit scalaire (et donc de notions de longueurs, d'angles...) privilégié, le Théorème 1 est tout à fait satisfaisant. Si ce n'est pas le cas, il n'y a aucune raison de se restreindre aux bases orthonormées. La liberté que cela nous offre nous laisse espérer pouvoir trouver des formes normales encore plus simples.

Par le Théorème 1, toute classe de congruence contient une matrice diagonale. La question est de savoir maintenant quels types de matrices diagonales on peut trouver dans la même classe de congruence. Autrement dit, peut-on trouver d'autres formes diagonalisées de q en changeant de base ?

Exercice 3. Posons $q(\vec{x}) := x_1^2 + x_2^2$ dans \mathbb{R}^2 . Trouvez d'autres bases dans laquelle q s'écrit sous forme diagonale. Que peut-on dire des matrices diagonales congruentes à id ?

Définition II.17.

Soient n, p, q trois entiers tels que $p + q \leq n$. On pose

$$I_{p,q,n} := \begin{pmatrix} I_p & \cdots & 0 \\ \vdots & -I_q & \vdots \\ 0 & \cdots & 0 \end{pmatrix}.$$

Théorème 2.

Soit A une matrice symétrique réelle. Soient p est le nombre de valeurs propres strictements positives de A et q est le nombre de valeurs propres strictements négatives de A . Alors A est congruente à $I_{p,q,n}$.

De manière équivalente, toute forme quadratique de E admet dans une base \mathcal{B} bien choisie une écriture de la forme

$$q(\vec{x}) = \sum_{i=1}^p X_i^2 - \sum_{i=p+1}^{p+q} X_i^2,$$

pour tout $\vec{x} \in E$, en notant X les coordonnées de \vec{x} dans \mathcal{B} .

Enfin, si E est muni d'un produit scalaire, on peut choisir \mathcal{B} orthogonale.

PREUVE : Soit $\mathcal{B} = (\vec{e}_1, \dots, \vec{e}_n)$ une base orthonormée dans laquelle q s'écrit de la forme $q(\vec{x}) = \sum_{i=1}^n \lambda_i X_i^2$. Soit p le nombre de λ_i strictement positifs, et q le nombre de λ_i strictement négatifs.

Sans perte de généralité, on peut réordonner la base \mathcal{B} de telle sorte que $\lambda_1, \dots, \lambda_p$ soient strictement positifs, $\lambda_{p+1}, \dots, \lambda_{p+q}$ soient strictement négatifs, et $\lambda_{p+q+1}, \dots, \lambda_n$ soient nuls. Alors :

$$q(\vec{x}) = \sum_{i=1}^p \lambda_i X_i^2 - \sum_{i=p+1}^{p+q} (-\lambda_i) X_i^2 = \sum_{i=1}^p (\sqrt{\lambda_i} X_i)^2 - \sum_{i=p+1}^{p+q} (\sqrt{-\lambda_i} X_i)^2.$$

Posons $\vec{e}'_i = \frac{\vec{e}_i}{\sqrt{\lambda_i}}$ pour $1 \leq i \leq p$, puis $\vec{e}'_i = \frac{\vec{e}_i}{\sqrt{-\lambda_i}}$ pour $p+1 \leq i \leq p+q$ et $\vec{e}'_i = \vec{e}_i$ pour $i > p+q$. Alors la matrice de q dans la base $(e'_i)_{1 \leq i \leq n}$ est $I_{p,q,n}$. \square

Remarque II.18.

La donnée du couple (p, q) permet de caractériser la matrice apparaissant dans le Théorème 2. Toute classe de congruence contient au moins une matrice de cette forme.

On peut remarquer que $\text{rg}(A) = p + q$. On sait déjà que le rang est un invariant de congruence, donc si $I_{p,q,n}$ et $I_{p',q',n}$ sont dans la même classe de congruence, alors $p + q = p' + q'$. En fait, on va montrer que toute classe de congruence contient une unique matrice de cette forme, et donc que le couple (p, q) est un invariant complet de congruence.

Propriété II.19.

Soit q une forme quadratique. Alors il existe une unique matrice $I_{p,q,n}$ représentant q dans une base.

PREUVE : Soit q une forme quadratique. Soient $\mathcal{B} = (\vec{e}_1, \dots, \vec{e}_n)$, $\mathcal{B}' = (\vec{f}_1, \dots, \vec{f}_n)$ deux bases de E dans lesquelles les matrices représentant q sont $I_{p,q,n}$ et $I_{p',q',n}$ respectivement.

Soient $E_{\leq 0} := \text{Vect}(\vec{e}_{p+1}, \dots, \vec{e}_n)$ et $E'_{>0} := \text{Vect}(\vec{f}_1, \dots, \vec{f}_{p'})$. Soit $\vec{x} \in E_{\leq 0} \cap E'_{>0}$ non nul. Alors $q(\vec{x}) = -\sum_{i=p+1}^{p+q} X_i^2 \leq 0$ car $x \in E_{\leq 0}$, et $q(\vec{x}) = \sum_{i=1}^{p'} (X'_i)^2 > 0$ car $x \in E'_{>0}$ et est non nul. C'est impossible. Donc $E_{\leq 0} \cap E'_{>0}$ ne contient pas de vecteur non nul, ou encore $E_{\leq 0} \cap E'_{>0} = \{\vec{0}\}$: E_- et E'_+ sont transverses.

Par conséquent, $\dim(E_{\leq 0}) + \dim(E'_{>0}) \leq n$, ou encore $(n-p) + p' \leq n$, soit finalement $p' \leq p$.

En intervertissant le rôle de \mathcal{B} et \mathcal{B}' , on trouve de même $p \leq p'$. Donc $p = p'$. Or $p + q = p' + q' = \text{rg}(q)$, donc $q = q'$. \square

Définition II.20 (Signature).

On appelle **signature** d'une forme quadratique le couple d'entiers (p, q) , où $I_{p,q,n}$ est l'unique matrice de cette forme représentant q .

Si A est une matrice représentant q dans une base quelconque, alors p est le nombre de valeurs propres strictement positives de A et q le nombre de valeurs propres strictement négatives de A .

Attention : il y a 2 objets notés q ici, qui n'ont rien à voir !

Exercice 4. Déterminez les signatures de matrices symétriques de votre choix.

On en déduit la fin de la recherche des invariants pour la classe de congruence d'une matrice symétrique réelle : **la signature est un invariant complet de congruence.**

Définition II.21.

Une forme quadratique q est dite

- ▷ **définie** si pour tout $\vec{x} \in E$, $q(\vec{x}) = 0 \implies x = 0$
- ▷ **positive** si pour tout $\vec{x} \in E$, $q(\vec{x}) \geq 0$
- ▷ **négative** si pour tout $\vec{x} \in E$, $q(\vec{x}) \leq 0$

Exercice 5. Montrer que si une forme quadratique q est dégénérée, alors elle n'est pas définie. Que pensez-vous de la réciproque ?

Exercice 6. Montrer que si q est définie, alors elle est de signe constant

On a plus précisément la propriété suivante :

Propriété II.22.

Soit q une forme quadratique sur un espace vectoriel de dimension n , alors q est définie si et seulement si sa signature est égale à $(n, 0)$ ou $(0, n)$.

PREUVE : Supposons que la signature de q soit $(n, 0)$. Alors $q(x) = \sum_{i=1}^n X_i^2$ dans une base adaptée, donc q est définie positive. De même, si q est de signature $(0, n)$, alors q est définie négative.

Supposons que q est de signature (p, q) différente de $(n, 0)$ ou $(0, n)$. Alors :

- ▷ Si $p + q < n$: soit A la matrice représentant q dans une base \mathcal{B} quelconque. On sait que $\text{rang}(A) = p + q < n$. Soit donc $X \in \text{Ker}(A)$ non nul, et \vec{x} le vecteur de coordonnées X . Alors $q(\vec{x}) = X^T A X = X^T 0 = 0$ avec $\vec{x} \neq \vec{0}$, donc q n'est pas définie.
- ▷ Si $p + q = n$: alors $p \geq 1$ et $q \geq 1$. Soit \mathcal{B} une base telle que la matrice représentant q soit $I_{p,q,n}$. Alors $q(e_1 + e_{p+1}) = 1^2 - 1^2 = 0$ et $e_1 + e_{p+1} \neq 0$, donc q n'est pas définie.

□

Remarque II.23.

Si l'on travaille dans \mathbb{C} et non dans \mathbb{R} , on peut réduire encore la forme quadratique q . En effet, dans une base réduite $\mathcal{B} = (\vec{e}_1, \dots, \vec{e}_n)$,

$$q(\vec{x}) = \sum_{i=1}^p X_i^2 - \sum_{i=p+1}^{p+q} X_i^2 = \sum_{i=1}^p X_i^2 + \sum_{i=p+1}^{p+q} (iX_i)^2.$$

Après multiplication des vecteurs $\vec{e}_{p+1}, \dots, \vec{e}_{p+q}$ par $-i$, dans la nouvelle base, q s'écrit

$$q(\vec{x}) = \sum_{i=1}^p X_i^2 + \sum_{i=p+1}^{p+q} X_i^2 = \sum_{i=1}^{p+q} X_i^2.$$

Autrement dit, le seul invariant de congruence dans \mathbb{C} est $p + q$, c'est-à-dire le rang.

II.5 Algorithme de réduction de Gauss

L'objectif de cette partie est de déterminer rapidement la signature d'une forme quadratique, sans avoir à calculer le spectre d'une matrice. Pour cela, il faut trouver sa forme réduite : c'est ce que fournit l'algorithme de la réduction de Gauss pour les formes quadratiques.

Définition II.24.

*Soit q une forme quadratique de E de signature (p, q) . Une base de E est une **base réduite** pour q si dans cette base la matrice associée à q est la matrice $I_{p,q,n}$. Autrement dit, dans une base réduite \mathcal{B} , pour tout $\vec{x} \in E$ de coordonnées X dans \mathcal{B} ,*

$$q(\vec{x}) = \sum_{i=1}^p X_i^2 - \sum_{i=p+1}^{p+q} X_i^2 = X^T I_{p,q,n} X.$$

Cette dernière formule est l'écriture réduite de q .

Attention : Il y a une unique écriture réduite, mais plusieurs bases réduites.

Pour trouver une base réduite pour q , on peut procéder en trois temps :

1. Écrire la matrice A de q dans une base quelconque.
2. La matrice A est symétrique réelle. La diagonaliser.
3. Multiplier les vecteurs de la base par des constantes pour obtenir une écriture réduite (voir la fin de la démonstration du Théorème 2).

La partie laborieuse est celle consistant à diagonaliser A . L'algorithme de Gauss fournit une méthode plus efficace.

Proposition II.25.

Soit q une forme quadratique sur E et \mathcal{B} une base de E . Soit A la matrice symétrique représentant q dans \mathcal{B} . Alors on obtient une base réduite pour q en procédant récursivement de la manière suivante :

1. Si $A_{11} \neq 0$, on met sous forme canonique l'équation en X_1 :

$$q(\vec{x}) = A_{11} \left(X_1 + \sum_{j=2}^n \frac{A_{1j}X_j}{A_{11}} \right)^2 - \left(\sum_{j=2}^n \frac{A_{1j}X_j}{A_{11}} \right)^2 + \sum_{i=2}^n A_{ii}X_i^2 + 2 \sum_{2 \leq i < j \leq n} A_{ij}X_iX_j.$$

On pose ensuite $X'_1 = \sqrt{|A_{11}|} \left(X_1 + \sum_{j=2}^n \frac{A_{1j}X_j}{A_{11}} \right)$ et $X'_i = X_i$ si $i \geq 2$. Alors seul le premier terme (en bleu) dépend de X'_1 , et on peut poursuivre l'algorithme pour les termes suivants.

2. Si $A_{11} = 0$, mais qu'il existe $i \in \{2, n\}$ tel que $A_{ii} \neq 0$: on échange alors les première et i -ème coordonnées⁷, puis on se ramène au cas précédent.
3. Si $A_{ii} = 0$ pour tout $i \in \{1, n\}$ mais que $A \neq 0$, alors on fait apparaître des carrés en utilisant l'identité $xy = \left(\frac{x+y}{2}\right)^2 - \left(\frac{x-y}{2}\right)^2$. Quitte à permuter les éléments de la base, on peut supposer que $A_{12} \neq 0$. On obtient alors

$$\begin{aligned} q(\vec{x}) = & 2A_{12} \left[\left(\frac{X_1 + X_2}{2} \right)^2 - \left(\frac{X_1 - X_2}{2} \right)^2 \right] \\ & + 2 \sum_{j=3}^n (A_{1j} + A_{2j}) \frac{X_1 + X_2}{2} X_j \\ & + 2 \sum_{j=3}^n (A_{1j} - A_{2j}) \frac{X_1 - X_2}{2} X_j \end{aligned}$$

On pose alors $X'_1 = \frac{X_1 + X_2}{2}$, $X'_2 = \frac{X_1 - X_2}{2}$, $X'_i = X_i$ si $i \geq 3$. On est donc ramené au premier cas.

4. L'algorithme termine quand il ne reste plus aucun terme.

PREUVE : Pour justifier qu'on obtient bien une forme réduite de q , il suffit de démontrer que les changements de coordonnées correspondent bien à des changements de base. Le cas 2 est évident, puisqu'il correspond à la permutation de deux vecteurs de la base.

7. Donc on échange les lignes et les colonnes !

Le cas 1 est analogue à la première étape de l'algorithme du pivot de Gauss. C'est la composée de deux transformations : on ajoute à X_1 une combinaison linéaire des autres coordonnées, puis on factorise par $\sqrt{|A_{11}|}$. Chacune de ces étapes est inversible, donc la première étape est inversible.

Pour le cas 3, on peut remarquer qu'on obtient X_1 et X_2 comme somme et différence de $\frac{X_1+X_2}{2}$ et de $\frac{X_1-X_2}{2}$, ce qui assure l'inversibilité du changement de coordonnées. \square

Remarque II.26.

Dans la proposition précédente, l'algorithme ne fournit pas nécessairement une base réduite dans le bon ordre (d'abord les positifs, puis les négatifs, puis le reste) : on peut être amené à changer l'ordre de la base obtenue pour se ramener à l'écriture réduite, et donc à une forme réduite.

Exemple II.27.

Soit $q(\vec{x}) = x_1^2 + 2x_1x_2 - x_1x_3 - 2x_2^2 + \frac{x_3^2}{6} + 2x_3x_4$ une forme quadratique sur \mathbb{R}^4 . On applique l'algorithme :

$$\begin{aligned} q(\vec{x}) &= \left(x_1 + x_2 - \frac{x_3}{2}\right)^2 - x_2^2 - \frac{x_3^2}{4} + x_2x_3 - 2x_2^2 + \frac{x_3^2}{6} + 2x_3x_4 \\ &= \left(x_1 + x_2 - \frac{x_3}{2}\right)^2 - 3x_2^2 + x_2x_3 - \frac{x_3^2}{12} + 2x_3x_4 \\ &= \left(x_1 + x_2 - \frac{x_3}{2}\right)^2 - 3\left(x_2 - \frac{x_3}{6}\right)^2 + 2x_3x_4 \\ &= \left(x_1 + x_2 - \frac{x_3}{2}\right)^2 - 3\left(x_2 - \frac{x_3}{6}\right)^2 + 2\left(\frac{x_3+x_4}{2}\right)^2 - 2\left(\frac{x_3-x_4}{2}\right)^2. \end{aligned}$$

Il y a deux termes positifs et deux termes négatifs ; la signature de q est donc $(2, 2)$. On peut mettre q sous forme réduite :

$$q(\vec{x}) = \left(x_1 + x_2 - \frac{x_3}{2}\right)^2 + \left(\frac{x_3+x_4}{\sqrt{2}}\right)^2 - \left(\sqrt{3}x_2 - \frac{x_3}{2\sqrt{3}}\right)^2 - \left(\frac{x_3-x_4}{\sqrt{2}}\right)^2.$$

Exercice 7. Trouvez la signature de la forme quadratique $q(\vec{x}) = 2x_1^2 + 3x_1x_2 - 4x_3^2$ de \mathbb{R}^3 , puis donner une base réduite associée. Même question pour $q(\vec{x}) = 4x_1x_3 + 3x_2x_3$. Les bases trouvées sont-elles des bases de vecteurs propres pour la matrice associée à q dans la base canonique ? Expliquez.

III Formes quadratiques et quadriques

On peut aussi s'appuyer sur la géométrie pour rechercher des bases réduites, ou bien utiliser la signature pour analyser des objets géométriques : les quadriques. Si \vec{u} est un vecteur d'une base réduite, alors on a $q(\vec{u}) \in \{1, 0, -1\}$, ce qui conduit à l'étude des ensembles vérifiant $q(\vec{u}) = c$ pour $c \in \mathbb{R}$: un tel ensemble est appelé **ligne de niveau** de q . En général, on doit distinguer deux cas, selon que $c = 0$ ou non.

III.1 Lignes de niveau et cône isotrope

Définition III.1 (Cône isotrope). Le **cône isotrope** d'une forme quadratique q est l'ensemble

$$\mathcal{C}_0(q) := \{\vec{u} \in E, \quad q(\vec{u}) = 0\}.$$

Remarque III.2.

Une forme quadratique q est définie si et seulement si le cône isotrope est réduit à $\{0\}$. Cet ensemble est donc intéressant pour les formes quadratiques non dégénérées.

Exercice 8. Déterminer le cône isotrope de $q(\vec{u}) = x^2 - y^2$ sur \mathbb{R}^2 . Dessiner l'ensemble sur lequel $q > 0$.

Exercice 9. Déterminer le cône isotrope de $q(\vec{u}) = x^2 + y^2 - z^2$ sur \mathbb{R}^3 . Dessiner l'ensemble sur lequel $q > 0$.

Exercice 10. Soit q une forme quadratique non dégénérée et non définie de \mathbb{R}^2 . Montrer que q est de signature $(1, 1)$, puis que son cône isotrope est l'union de deux droites vectorielles.

Dans cette partie, on étudie les lignes de niveau pour les formes quadratiques non dégénérées. Cette étude nous permet de visualiser différentes bases réduites obtenues par réduction de Gauss.

Définition III.3.

Soit \mathcal{C} une partie non vide de E . On dit que \mathcal{C} est une **quadrique** s'il existe une forme quadratique $q \neq 0$ de E et $r \in \mathbb{R}$ tels que

$$\mathcal{C} = \mathcal{C}_r(q) := \{\vec{u} \in E : q(\vec{u}) = r\}.$$

On dit que \mathcal{C} est **non dégénérée** si q est non dégénérée et $r \neq 0$.

Remarque III.4.

Si $\dim(E) = 2$, on parlera non pas de quadrique mais de conique.

Remarque III.5.

Si $\mathcal{C} = \mathcal{C}_r(q)$ est une quadrique, alors \mathcal{C} est symétrique par rapport à l'origine ; en effet, en reprenant la définition,

$$\vec{u} \in \mathcal{C} \iff q(\vec{u}) = r \iff q(-\vec{u}) = r \iff -\vec{u} \in \mathcal{C}.$$

En particulier, cette définition ne couvre pas toutes les sections coniques possibles ; par exemple, avec cette définition, les paraboles ne sont pas des quadriques.

Propriété III.6.

Soit \mathcal{C} une quadrique non dégénérée. Alors il existe une unique forme quadratique q telle que $\mathcal{C} = \{\vec{u} \in E : q(\vec{u}) = 1\}$, et cette forme quadratique est non dégénérée. On dit que cette forme quadratique est **associée** à \mathcal{C} .

PREUVE : Existence : Soient q non dégénérée et $r \in \mathbb{R}^*$ tels que $\mathcal{C} = \{\vec{u} \in E : q(\vec{u}) = r\}$. Alors $q' := q/r$ convient, et est non dégénérée.

Unicité : Soit q'' une forme quadratique associée à \mathcal{C} , et montrons que $q'' = q'$. Remarquons que, pour tout $\vec{u} \in \mathcal{C}$ et $\lambda \in \mathbb{R}$, on a $q'(u) = q''(u) = 1$, donc $q'(\lambda u) = \lambda^2 = q''(\lambda u)$. Par conséquent, q' et q'' coïncident sur $\text{Vect}(\vec{u})$. Ceci étant valable pour tout $\vec{u} \in \mathcal{C}$, les formes quadratiques q' et q'' coïncident sur $\text{Vect}(\mathcal{C})$. Il reste à montrer que $\text{Vect}(\mathcal{C}) = E$.

Soient $\vec{u} \in \mathcal{C}$ et $\mathcal{B} = (\vec{e}_1, \dots, \vec{e}_n)$ une base de E . Soit $1 \leq i \leq n$. La fonction $t \mapsto q'(\vec{u} + t\vec{e}_i)$ est quadratique en t et vaut 1 en 0, donc il existe $\varepsilon_i > 0$ tel que $q'(\vec{u} + \varepsilon_i\vec{e}_i) > 0$. Mais alors $\frac{\vec{u} + \varepsilon_i\vec{e}_i}{\sqrt{q'(\vec{u} + \varepsilon_i\vec{e}_i)}} \in \mathcal{C}$. Par conséquent, \mathcal{C} contient la famille libre de vecteurs $(\vec{u} + \varepsilon_i\vec{e}_i)_{1 \leq i \leq n}$, donc $\text{Vect}(\mathcal{C}) = E$, donc $q' = q''$.

□

III.2 Ellipsoïdes

Définition III.7 (Ellipsoïde).

Soit \mathcal{C} une quadrique non dégénérée et q la forme quadratique associée. Si q est de signature $(n, 0)$, on dit que \mathcal{C} est un **ellipsoïde**.

Proposition III.8.

Soit \mathcal{C} une partie non vide d'un espace euclidien E . Alors \mathcal{C} est un ellipsoïde si et seulement s'il existe $f \in GL(E)$ telle que $\mathcal{C} = f(\mathbb{S}^{n-1})$, où \mathbb{S}^{n-1} est la sphère unité de E .

Autrement dit, les ellipsoïdes sont des sphères déformées par une application linéaire.

PREUVE :

Soit \mathcal{C} un ellipsoïde. Soit q la forme quadratique associée ; alors q est définie positive. Soient \mathcal{B} une base orthonormée de E et A la matrice de q dans \mathcal{B} . Alors A est symétrique définie positive. D'après le théorème de Cholevski, il existe une matrice P triangulaire supérieure inversible telle que $A = P^T P$. Soit f l'endomorphisme de E de matrice P . Alors, pour tout $\vec{u} \in E$,

$$q(\vec{u}) = \langle f(\vec{u}), f(\vec{u}) \rangle = \|f(\vec{u})\|^2.$$

Alors

$$\vec{u} \in \mathcal{C} \iff \|f(\vec{u})\|^2 = 1 \iff f(\vec{u}) \in \mathbb{S}^{n-1}.$$

Autrement dit, $\vec{u} \in \mathbb{S}^{n-1}$ si et seulement si $f^{-1}(\vec{u}) \in \mathcal{C}$: l'ellipse \mathcal{C} est l'image de \mathbb{S}^{n-1} par l'application linéaire f^{-1} .

Réciproquement, soit f une application linéaire inversible. Posons $q(\vec{u}) := \|f^{-1}(\vec{u})\|^2$. Alors :

- ▷ q est une forme quadratique, associée à la forme bilinéaire symétrique $\varphi(\vec{u}, \vec{v}) := \langle f^{-1}(\vec{u}), f^{-1}(\vec{v}) \rangle$.
- ▷ φ est un produit scalaire. Par conséquent, q est définie positive, donc de signature $(n, 0)$.

Par le même calcul que précédemment, $\vec{u} \in f(\mathbb{S}^{n-1}) \iff \|f^{-1}(\vec{u})\|^2 = 1$, donc $f(\mathbb{S}^{n-1})$ est l'ellipsoïde $\mathcal{C} := \{\vec{u} \in E : q(\vec{u}) = 1\}$. \square

Exemple III.9.

Considérons l'ellipsoïde de \mathbb{R}^2 :

$$\mathcal{C} := \{(x, y) \in \mathbb{R}^2 : 2x^2 + 3y^2 = 1\}.$$

L'équation le définissant se réécrit $(\sqrt{2}x)^2 + (\sqrt{3}y)^2 = 1$. Soit $f(xy) := (\frac{x}{\sqrt{2}}, \frac{y}{\sqrt{3}})$. Alors

$$(x, y) \in \mathcal{C} \iff 2x^2 + 3y^2 = 1 \iff f^{-1}(x, y) \in \mathbb{S}_1 \iff (x, y) \in f(\mathbb{S}_1),$$

donc \mathcal{C} est l'image du cercle unité par f .

III.3 Étude des ellipses

Soit q la forme quadratique associée à une conique $\mathcal{C} \subset \mathbb{R}^2$. Comme q est non dégénérée, sa signature est soit $(2, 0)$, soit $(1, 1)$, soit $(0, 2)$. Dans le troisième cas, q est définie négative, donc \mathcal{C} est vide, ce qui est absurde. Il reste donc deux types de coniques :

Définition III.10.

- ▷ Si q est de signature $(2,0)$, on dit que $\mathcal{C} \subset \mathbb{R}^2$ est une **ellipse**. Les ellipses sont des cas particuliers d'ellipsoïdes.
- ▷ Si q est de signature $(1,1)$, on dit que $\mathcal{C} \subset \mathbb{R}^2$ est une **hyperbole**.

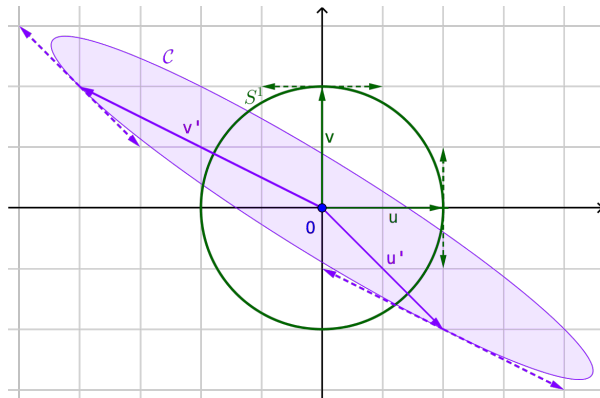
Exemple III.11.

- ▷ Lorsque $q(u) = x^2 + y^2$, la conique associée est le cercle unité $\mathbb{S}^1 = \{(x, y) \in \mathbb{R}^2; x^2 + y^2 = 1\}$.
- ▷ Lorsque $q(u) = xy$, la conique associée est l'hyperbole classique $\mathcal{H} = \{(x, y) \in \mathbb{R}^2; xy = 1\}$. Son cône isotrope est la réunion des deux axes.

Exercice 11. Déterminer la nature de la conique associée à $q(u) = x^2 - y^2$. Quel lien géométrique peut-on faire avec \mathcal{H} ? On précisera son cône isotrope et les directions propres de sa matrice associée.

Rappelons qu'une ellipse est l'image du cercle unité par une application linéaire inversible f . Cette propriété permet d'identifier géométriquement les tangentes à l'ellipse à partir de l'image des vecteurs de la base canonique par f .

Exercice 12. Déterminer sur le schéma suivant la matrice associée à l'application f ci-dessus. En déduire l'expression de la forme quadratique associée à l'ellipse représentée. Déterminer les axes puis l'équation réduite de l'ellipse (voir le théorème suivant).



La propriété suivante explicite dans le cas de \mathbb{R}^2 l'existence de différentes bases réduites pour q et la signification géométrique de la base obtenue par diagonalisation de A .

Théorème III.12.

Soit $\mathcal{C} \subset \mathbb{R}^2$ une ellipse et q la forme quadratique associée. Alors il existe une **base orthonormée** \mathcal{B} et $\lambda_1 \geq \lambda_2 > 0$ tels que

$$\mathcal{C} = \{ \vec{u} : \lambda_1 X^2 + \lambda_2 Y^2 = 1 \}.$$

Si $\lambda_1 > \lambda_2$, alors cette base orthonormée est unique au signe près et définit les **axes** de l'ellipse. Elle correspond à la base des vecteurs propres de la matrice associée à q . Le grand axe est associé à λ_2 et le petit axe à λ_1 .

PREUVE : Il suffit de choisir une base orthonormée dans laquelle $q(\vec{u}) = \lambda_1 X^2 + \lambda_2 Y^2$. Quitte à échanger les deux vecteurs de la base, on peut supposer que $\lambda_1 \geq \lambda_2$. De plus, comme \mathcal{C} est une ellipse, q est définie positive, donc $\lambda_2 > 0$. \square

Attention : Le **grand** axe est associé à la plus **petite** valeur propre, et réciproquement.

Attention : On peut avoir une équation de la forme $aX^2 + bY^2 = 1$ qui n'est pas en base orthogonale : dans ce cas, on ne peut pas en déduire d'information sur les valeurs propres et directions propres ! En revanche, si la base est orthonormée, alors elle est nécessairement une base propre.

Remarque III.13.

Dans l'écriture précédente, λ_1 et λ_2 sont les valeurs propres de la matrice associée à q dans une base orthonormée.

Remarque III.14.

En général, on peut trouver une base orthonormée dans laquelle l'équation d'une conique \mathcal{C} est de la forme

$$\sum_{i=1}^n \lambda_i X_i^2 = 1,$$

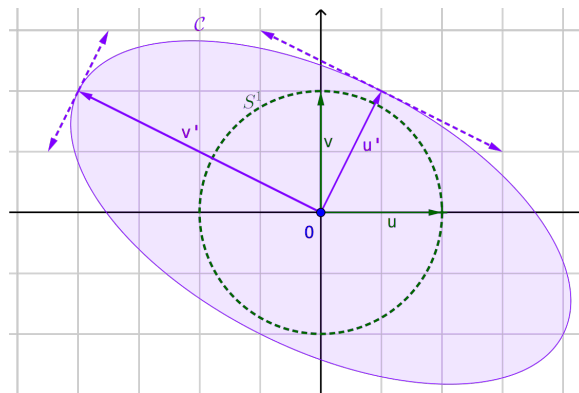
avec $\lambda_1 \leq \dots \leq \lambda_n$. Si de plus les λ_i sont deux à deux distincts, alors la base est unique à changement de signes des vecteurs près, donc les directions propres sont uniques. On dispose donc de n axes "remarquables". Cependant, ces axes sont plus difficiles à visualiser dans un ellipsoïde, et beaucoup plus difficiles à voir si q n'est pas définie positive.

Proposition III.15.

Les axes d'une ellipse sont aussi ses axes de symétrie. Par conséquent, pour trouver les directions propres de la matrice associée, il suffit de déterminer les axes de symétrie de l'ellipse.

PREUVE : On se place dans une base orthonormée dans laquelle l'équation de l'ellipse \mathcal{C} est $\lambda_1 X^2 + \lambda_2 Y^2 = 1$. Alors (X, Y) satisfait cette équation si et seulement si $(X, -Y)$ en est aussi solution, donc la réflexion d'axe $\{Y = 0\}$ est bien une symétrie de l'ellipse. De même, la réflexion d'axe $\{X = 0\}$ est une symétrie de l'ellipse. \square

Exercice 13. Pour l'ellipse donnée sur le graphique ci-dessous, préciser son équation réduite et donner les valeurs et vecteurs propres de la matrice associée. En déduire son équation cartésienne dans la base canonique.



Exemple III.16. Nature et représentation de la courbe $\mathcal{C} = \{(x, y) \in \mathbb{R}^2, 5x^2 - 4xy + 5y^2 = 1\}$: il s'agit de la conique associée à $q(\vec{u}) = 5x^2 - 4xy + 5y^2$.

1. **Nature de la courbe** : on utilise la réduction de Gauss.

$$q(\vec{u}) = 5x^2 - 4xy + 5y^2 = 5 \left(x + \frac{2}{5}y \right)^2 - \frac{4}{5}y^2 + 5y^2 = 5 \left(x + \frac{2}{5}y \right)^2 + \frac{21}{5}y^2.$$

On en déduit que q est de signature $(2, 0)$, donc \mathcal{C} est une ellipse.

2. **Réduction de Gauss et signature** : La réduction précédente fournit une forme réduite de q sous la forme $x'^2 + y'^2$ avec comme changement de variable

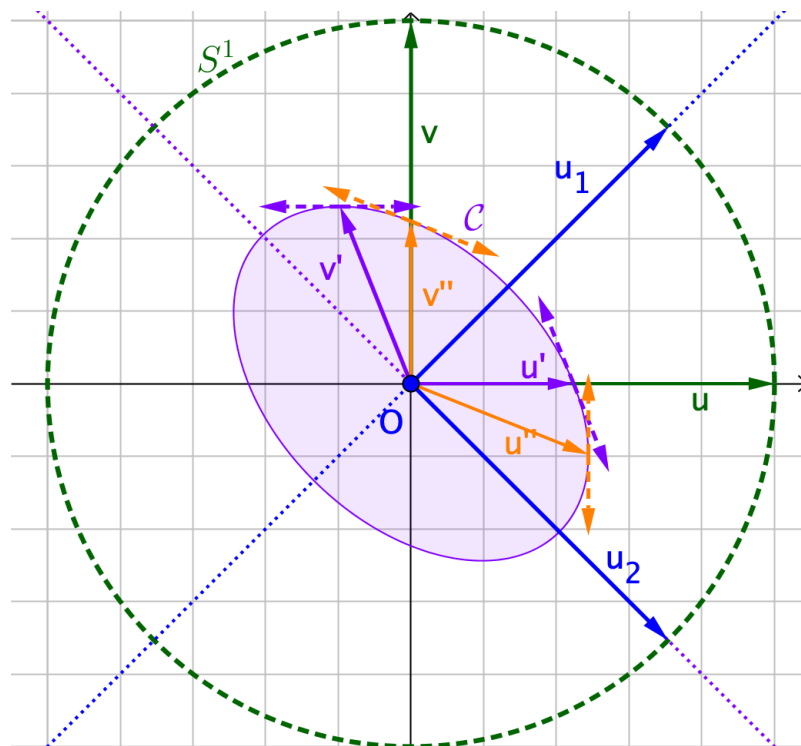
$$x' = \sqrt{5} \left(x + \frac{2}{5}y \right) \quad \text{et} \quad y' = \frac{\sqrt{105}}{5}y, \quad \text{i.e.} \quad \begin{pmatrix} x' \\ y' \end{pmatrix} = P^{-1} \begin{pmatrix} x \\ y \end{pmatrix} \quad \text{avec} \quad P^{-1} = \frac{\sqrt{5}}{5} \begin{pmatrix} 5 & 2 \\ 0 & \sqrt{21} \end{pmatrix}.$$

L'ellipse \mathcal{C} est donc l'image du cercle unité par $P = \frac{1}{\sqrt{105}} \begin{pmatrix} \sqrt{21} & -2 \\ 0 & 5 \end{pmatrix}$. Notons (\vec{u}', \vec{v}') la base dont la matrice de passage est P . De même, en procédant à la réduction en commençant par le terme $5y^2$, on aurait trouvé une base (\vec{u}'', \vec{v}'') .

3. **Diagonalisation et axes de l'ellipse** : La matrice associée à q dans la base canonique est $A = \begin{pmatrix} 5 & 2 \\ 2 & 5 \end{pmatrix}$. Pour déterminer les valeurs propres, on peut chercher les racines du polynôme caractéristique : on trouve $\lambda_1 = 7$ et $\lambda_2 = 3$. Ces valeurs propres étant strictement positives, on retrouve bien que q est de signature $(2, 0)$, et que \mathcal{C} est une ellipse. De plus, comme $\lambda_1 \neq \lambda_2$, cette ellipse n'a que deux axes de symétrie. $(1, 1)$ est un vecteur propre pour la valeur propre 7. Comme les directions propres sont orthogonales, $(1, -1)$ est un vecteur propre pour la valeur propre 3. Les axes de l'ellipse sont donc les droites dirigées par $(1, 1)$ (petit axe) et $(1, -1)$ (grand axe). Soit $\mathcal{B} = (\vec{u}_1, \vec{v}_1)$ une base orthonormée de vecteurs propres, avec $\vec{u}_1 := \frac{\sqrt{2}}{2}(1, 1)$ et $\vec{v}_1 := \frac{\sqrt{2}}{2}(1, -1)$. Alors, en notant (X', Y') les coordonnées dans \mathcal{B} ,

$$\mathcal{C} = \{(\vec{u} : 3X'^2 + 7Y'^2 = 1\}.$$

4. **Représentation graphique** des différentes bases obtenues dans les questions précédentes :



III.4 Cadre des espaces affines

Les quadriques se définissent aussi dans des espaces affines.

Définition III.17.

Soit E un espace affine de dimension n muni d'un repère. Une partie \mathcal{C} de E est une **quadrique** s'il existe une forme quadratique q , une forme linéaire ℓ et une constante r telles que

$$\mathcal{C} = \{a \in E : q(X) + \ell(X) + r = 0\},$$

où X sont les coordonnées de a dans le repère.

Les objets q , ℓ et r dépendent du repère, mais l'ensemble des quadriques ne dépend pas du repère choisi.

Exemple III.18.

Concrètement, les fonctions de la forme $q(X) + \ell(X) + r$ sont les polynômes de degré au plus 2 en les coefficients, comme par exemple

$$(x, y, z) \mapsto x^2 - 2xy + 2y^2 - z^2 - 2z + 5 ;$$

La forme quadratique correspond aux termes de degré exactement 2, donc ici $q(\vec{u}) = x^2 - 2xy + 2y^2 - z^2$. La forme linéaire correspond aux termes de degré exactement 1, donc ici $\ell(\vec{u}) = -2z$. Enfin, le terme constant est le terme de degré exactement 0, donc ici 5.

Soit A la matrice de q dans le repère, et L celle de ℓ . Supposons q non dégénérée. Alors A est inversible ; soit y le vecteur de coordonnées $-\frac{1}{2}A^{-1}L^T$. On trouve :

$$\begin{aligned} X^TAX + LX + r &= X^TAX + \frac{1}{2}LA^{-1}AX + \frac{1}{2}X^TAA^{-1}L^T + r \\ &= X^TAX - Y^TAX - X^TAY + r \\ &= (X - Y)^T A(X - Y) + r - Y^TAY \\ &= (X - Y)^T A(X - Y) + r - \frac{1}{4}LA^{-1}L^T. \end{aligned}$$

Par conséquent, en translatant de Y l'origine du repère, la courbe obtenue est d'équation $q(X) + r' = 0$, où $r' = r - \frac{1}{4}LA^{-1}L^T$. En particulier, la forme quadratique associée est la même. On obtient donc :

Proposition III.19.

Soit \mathcal{C} une quadrique affine. Supposons que sa partie quadratique q est non dégénérée. Alors, à translation près, \mathcal{C} est :

- ▷ ou bien une quadrique non dégénérée, dont la forme quadratique associée est proportionnelle à q ;
- ▷ ou bien le cône isotrope de q .

Exemple III.20. *On reprend l'exemple précédent. On utilise l'algorithme de réduction de Gauss, mais en absorbant les termes de degré 1 dans les termes quadratiques. Dans \mathbb{R}^3 ,*

$$x^2 - 2xy + 2y^2 - z^2 - 2z + 5 = (x - y)^2 + y^2 - (z + 1)^2 + 6,$$

donc en posant $(x', y', z') = (x, y, z + 1)$, c'est-à-dire en translatant le repère de $(0, 0, -1)$, l'équation de la quadrique affine devient $x'^2 - 2x'y' + 2y'^2 - z'^2 + 6 = 0$. L'étude se poursuit alors comme dans les parties précédentes.

Si l'on considère dans \mathbb{R}^2 la quadrique \mathcal{C} d'équation $xy - x - y + 1 = 0$, on obtient après factorisation $(x - 1)(y - 1) = 0$; cette fois-ci, à translation près, \mathcal{C} est le cône isotrope de $q(u) = xy$.

Si la forme quadratique est dégénérée, on peut obtenir des quadriques qui n'ont pas d'équivalents dans le cadre linéaire. C'est par exemple le cas de la conique d'équation $x^2 - y$ dans \mathbb{R}^2 ; comment s'appelle-t-elle ?

IV Introduction aux problèmes d'optimisation

L'objectif de ce paragraphe est d'appliquer les résultats des précédents chapitres aux problèmes de modélisation linéaire.

Souvent, des problèmes concrets se ramènent à des questions d'optimisation (en économie, pour le transport, pour le stockage de données...). Ces problèmes doivent prendre en compte un certain nombre de contraintes et proposer des solutions "les meilleures possibles" compte tenu des choix de données possibles. On parle de problèmes d'optimisation sous contraintes.

Dans les modèles linéaires, les contraintes sont souvent des contraintes de normes.

IV.1 Exemple d'une modélisation de type Leontief

Présentons un modèle de type Leontief. On dispose de p produits. Ces produits sont fabriqués par n industries, qui consomment du capital. Plus précisément, pour chaque unité monétaire qu'elle consomme, l'industrie j produit une quantité m_{ij} de chaque produit i . Pour simplifier, on supposera que $m_{ij} > 0$ pour tous i, j ; les industries ne consomment pas d'un produit au cours de leur production. On pourra penser par exemple à n régions industrielles, chacune étant plus ou moins spécialisée dans différents produits, et donc les produisant en différentes proportions et avec des efficacités différentes.

On dispose d'un capital fixé C , et on veut répartir ce capital entre les différentes industries pour produire "le plus possible". Notons x (X en colonne) le vecteur donnant le capital alloué à chaque industrie. Le capital étant fixé,

- ▷ $x_j \geq 0$ pour tout j ,
- ▷ $\sum_{j=1}^n x_j = C$.

De plus, la quantité de chaque produit fabriquée par une industrie donnée dépend linéairement du capital alloué. Autrement dit, la quantité total de produit i obtenue est

$$\sum_{j=1}^n a_{ij}x_j = (MX)_i.$$

Soit $f \in L(\mathbb{R}^n, \mathbb{R}^p)$ l'application linéaire de matrice $M = (m_{ij})$ dans les bases canoniques. On cherche donc à maximiser $f(x)$.

D'une part, la norme $\|x\|_1 = \sum_{i=1}^n |x_i|$ se prête moins aux calculs algébriques que la norme euclidienne. On va donc faire une entorse à notre modèle, et utilisant à la place la contrainte

$$\sum_{i=1}^n x_i^2 = X^T X = C^2,$$

c'est-à-dire la norme euclidienne. On espère avoir ainsi une solution peut-être moins précise, mais raisonnable et beaucoup plus facile à obtenir.

D'autre part, la notion de "maximiser $f(x)$ " est mal définie. Il n'y a pas de relation d'ordre naturelle sur \mathbb{R}^p ; on sait donc ce que veut dire maximiser une fonction à valeurs réelles, pas une fonction à valeurs dans \mathbb{R}^p . Il y a ici une part d'arbitraire; dans le cadre de cet exemple, on va chercher à obtenir suffisamment de chaque produit (par exemple, si les produits sont de l'eau et de l'électricité, on va vouloir éviter une solution qui consisterait à produire seulement de l'électricité, quitte à mourir de soif). Pour cela, une bonne solution consiste à chercher à maximiser la norme euclidienne $\|f(x)\|$, ce qui est équivalent à maximiser son carré $\|f(x)\|^2 = \sum_{i=1}^p (f(x))_i^2$.

Remarquons que l'on peut vouloir pondérer chaque ressource par une certaine valeur (sinon, par exemple, pourquoi accorder la même importance à 1 m^3 d'eau et à 1 MWh d'énergie électrique?). Cela peut se faire en modifiant la fonction f . Cependant, cela ne change rien au développement mathématique qui va suivre.

Le problème d'optimisation de production revient finalement à trouver x pour que $f(x)$ soit maximal, c'est-à-dire :

$$\text{On cherche } x_0 \text{ tel que } \|x_0\| = 1 \text{ et } \|f(x_0)\| = \max\{\|f(x)\|; \|x\| = 1\}.$$

Enfin, on remarque que

$$\|f(x)\|^2 = \langle f(x), f(x) \rangle = (MX)^T MX = X^T (M^T M) X.$$

Si on note $A = M^T M$, alors $q(x) = \|f(x)\|^2$ est la forme quadratique associée à A dans la base canonique. Il reste donc à étudier cette forme quadratique pour résoudre le problème d'optimisation.

IV.2 Valeurs singulières et théorème de minimax pour les formes quadratiques

Propriété IV.1.

Soit $M \in \mathcal{M}_{p,n}(\mathbb{R})$. Alors $A = M^T M \in \mathcal{S}_n(\mathbb{R})$, et toutes ses valeurs propres sont positives.

PREUVE : On calcule $A^T (M^T M)^T = M^T (M^T)^T = M^T M = A$, donc A est symétrique.

Soit λ une valeur propre de A et X un vecteur propre associé. Alors, d'une part, $X^T A X = X^T (\lambda X) = \lambda X^T X$; et, d'autre part, $X^T A X = X^T M^T M X = (MX)^T (MX) \geq 0$. Donc $\lambda X^T X \geq 0$, donc $\lambda \geq 0$. \square

Définition IV.2 (Valeurs singulières).

Avec les notations précédentes, on note $0 \leq \lambda_1 \leq \lambda_2 \leq \dots \leq \lambda_n$ les valeurs propres de $A = M^T M$. On appelle **valeurs singulières** de la matrice M les racines de valeurs propres de A , c'est-à-dire la famille $(\sqrt{\lambda_i})_{1 \leq i \leq n}$.

Remarque IV.3.

Le rang de A est inférieur au rang de M , lui-même inférieur à $\min\{n, p\}$. Par conséquent, si $p < n$, alors au moins $n - p$ des valeurs singulières sont nulles. Dans le modèle de type Leontief, cela arrive s'il y a moins de produits que d'industries.

La forme quadratique associée à A est donc positive (mais pas nécessairement définie positive!), et la solution au problème est donné par le résultat suivant :

Théorème 3.

Soit q une forme quadratique positive de \mathbb{R}^n . Notons A sa matrice associée dans la base canonique et $0 \leq \lambda_1 \leq \lambda_2 \leq \dots \leq \lambda_n$ ses valeurs propres. Alors :

- ▷ $\min\{q(x); \|x\| = 1\} = \lambda_1$, et ce minimum est atteint exactement par les vecteurs propres associés à $\lambda_1 = \lambda_{\min}$ et de norme 1.
- ▷ $\max\{q(x); \|x\| = 1\} = \lambda_n$, et ce maximum est atteint exactement par les vecteurs propres associés à $\lambda_n = \lambda_{\max}$ et de norme 1.

PREUVE : Le raisonnement est semblable à celui de la Conséquence II.15. On se place dans une base orthonormée pour le produit scalaire et dans laquelle la matrice $P^T A P$ est diagonale de coefficients $\lambda_1, \dots, \lambda_n$. Alors, pour tout x de coordonnées X dans cette base,

$$q(x) = \sum_{j=1}^n \lambda_j X_j^2 \geq \sum_{j=1}^n \lambda_{\min} X_j^2 = \lambda_{\min} \|x\|^2 ;$$

en particulier, si $\|x\| = 1$, alors $q(x) \geq \lambda_{\min}$. De plus, il y a égalité si et seulement s'il y a égalité dans la suite d'inégalités ci-dessus, c'est-à-dire si, pour tout j ,

$$\lambda_j X_j^2 = \lambda_{\min} X_j^2.$$

Mais cela implique que $X_j = 0$ si $\lambda_j > \lambda_{\min}$, donc que x est combinaison linéaire de vecteurs propres pour la valeur propre λ_{\min} , donc que x est vecteur propre pour la valeur propre λ_{\min} .

L'analyse du maximum est similaire. \square

Dans le problème initial, on va donc chercher le vecteur X de norme C et propre pour la plus grande valeur propre de $A = M^T M$.

IV.3 Analyse en Composantes Principales (ACP)

Un autre exemple de problème peut être présenté en lien avec l'analyse d'images. Pour stocker des images, l'objectif est de construire une banque d'images servant de bibliothèque avec laquelle on pourra stocker le mieux possible les images réelles⁸. Autrement dit, chaque image sera remplacée par une combinaison linéaire des images du dictionnaire. La question est de trouver le meilleur dictionnaire possible.

Modélisation

Une image est donnée par un vecteur de \mathbb{R}^p , où p est généralement très grand⁹. Ce vecteur s'appelle un signal.

On dispose d'une banque d'image, c'est-à-dire d'un grand nombre de signaux, s_1, \dots, s_k . On souhaite choisir une famille de signaux qui soit les "meilleurs" représentants de l'ensemble de ces signaux ; meilleur étant compris au sens où cette famille permettra de reconstruire le plus fidèlement possible l'image de départ, à l'aide de combinaison linéaires des images de cette famille. C'est bien un problème linéaire

Décrivons une méthode permettant de trouver le premier vecteur $b \in \mathbb{R}^p$ de "la bibliothèque". Comme il s'agit d'une base, l'optimisation porte sur l'ensemble des vecteurs b de même norme, donc de norme 1. On dit alors que le meilleur b est celui dont la direction est en moyenne "la plus proche" de tous les s_i . On calcule alors la distance de chaque s_i à la droite $\mathbb{R}b$, et on cherche à minimiser la somme des carrés de ces distances¹⁰.

$$\text{On cherche } b \text{ tel que } \|b\| = 1 \text{ et } \sum_{i=1}^k d^2(s_i, \mathbb{R}b) \text{ est minimale.}$$

Résolution

On peut calculer ces distances à l'aide de projections orthogonales π_b sur $\mathbb{R}b$:

$$d^2(s_i, \mathbb{R}b) = \|b - \pi_b(s_i)\|^2 = \|b - \langle b, s_i \rangle b\|^2 = \|b\|^2 - \langle b, s_i \rangle^2.$$

8. Par exemple, les formats JPEG vont utiliser comme images de base des ondes ou des ondelettes.

9. Un appareil photo peut proposer des images de l'ordre de 10 Mégapixels, soit $p = 10^7$ pixels.

10. Là encore, il serait a priori raisonnable de minimiser une autre norme. Cela est plus facile avec la norme euclidienne – et quand on travaille avec des vecteurs de plusieurs millions de coordonnées, cela peut changer beaucoup de chose en termes de temps de calcul !

En notant S_i et B les vecteurs de coordonnées associées, on doit donc maximiser :

$$\sum_{i=1}^k \langle s_i, b \rangle^2 = \sum_{i=1}^k B^T S_i S_i^T B = B^T \left(\sum_{i=1}^k S_i^T S_i \right) B$$

En posant $A = \sum_{i=1}^k S_i^T S_i$, on reconnaît la forme quadratique associée à A . Le théorème précédent permet donc de choisir pour b un vecteur propre associé à la plus grande valeur propre de A .

Pour construire les vecteurs suivants de la bibliothèque, il suffit de choisir des vecteurs propres normés correspondant à la seconde, troisième, etc. plus grande valeur propre.

IV.4 Problème des moindres carrés et équations normales

Problème des moindres carrés

On se donne maintenant une bibliothèque d'images b_1, \dots, b_n de \mathbb{R}^p , et on souhaite encoder un signal $b \in \mathbb{R}^p$ à l'aide de cette bibliothèque. Cela revient à chercher ses coordonnées dans (b_1, \dots, b_n) . Le problème, c'est que l'on a justement choisi la bibliothèque pour qu'elle ne soit pas trop lourde, c'est à dire que n est petit devant p , et donc (b_1, \dots, b_n) est très loin d'être une base de \mathbb{R}^p .

Soit A la matrice dont les colonnes sont données par les coordonnées des b_i , et B le vecteur colonne des coordonnées de b . Alors, trouver les coordonnées de b dans (b_1, \dots, b_n) , c'est trouver un vecteur colonne X de taille n tel que $AX = B$. Comme $n \ll p$, ce système n'a typiquement pas de solution. On va donc chercher une solution approchée, la meilleure possible. On cherche donc à trouver le vecteur X tel que AX soit le plus proche possible de B , donc minimisant $\|AX - B\|^2$. C'est ce qu'on appelle le **problème des moindres carrés**.

Définition IV.4.

Avec les notations ci-dessus, on dit que X_0 est **solution du problème des moindres carrés** si

$$\|AX_0 - B\| = \min\{\|AX - B\| ; X \in \mathbb{R}^n\}.$$

Equations normales

Le problème des moindres carrés admet toujours des solutions, que l'on peut expliciter grâce aux observations suivantes.

- ▷ Comme $AX \in \mathfrak{S}(A)$, minimiser $\|AX - B\|$ revient à chercher X tel que $AX = \pi_{\mathfrak{S}(A)}(B)$: il existe donc un tel X .
- ▷ $AX_0 = \pi_{\mathfrak{S}(A)}(B)$ si et seulement si, pour tout X , on a $(AX_0 - B) \perp AX$, ou encore que $\langle B, AX \rangle = \langle AX_0, AX \rangle$.
- ▷ De la condition précédente on déduit que $B^T AX = X_0^T A^T AX$ pour tout X , donc que $B^T A = X_0^T A^T A$.
- ▷ Si X_0 est une solution au problème des moindres carrés, alors l'ensemble des solution est $X_0 + \text{Ker}(A)$.

Exercice 14. Montrer que $\text{Ker}(A^T A) = \text{Ker}(A)$.

Théorème 4.

Soient $A \in \mathcal{M}_{p,n}(\mathbb{R})$, $B \in \mathcal{M}_{1,p}(\mathbb{R})$ et $X \in \mathcal{M}_{1,n}(\mathbb{R})$. Les assertions suivantes sont équivalentes :

1. Le vecteur X est solution du problème des moindres carrés.
2. Le vecteur X est solution du système $AX = \pi_{\mathfrak{S}(A)}(B)$.
3. Le vecteur X est solution du système des “équations normales” $A^T AX = A^T B$.

Enfin, on peut normaliser les solutions au moindres carrés en minimisant la norme des solutions au problème des moindres carrés :

Proposition IV.5.

Soient $A \in \mathcal{M}_{p,n}(\mathbb{R})$ et $B \in \mathcal{M}_{1,p}(\mathbb{R})$. Il existe un unique $X_* \in \mathcal{M}_{1,n}(\mathbb{R})$ de norme minimale parmi les solutions du problème des moindres carrés, c’est à dire tel que

$$\|X_*\|^2 = \min\{\|X\|_2; X \text{ solution au problème des moindres carrés}\}.$$

PREUVE : Soit X_0 une solution au problème des moindres carrés. Alors l’ensemble des solutions à ce problème est $S = X_0 + \text{Ker}(A)$. On cherche donc un vecteur $Y \in \text{Ker}(A)$ qui minimise $\|Y + X_0\|^2$. Soit C une matrice paramétrisant A , c’est-à-dire la matrice d’une application linéaire bijective de $\mathbb{R}^{\dim(\text{Ker}(A))}$ dans $\text{Ker}(A)$, ou encore une matrice dont les colonnes forment une base de $\text{Ker}(A)$. On cherche donc le vecteur Y' qui minimise $\|CY' + X_0\|^2$. Il s’agit encore d’un problème des moindres carrés, mais avec $A = C$ et $B = -X_0$.

Ce problème admet une solution. De plus, cette fois-ci, la matrice C est bijective de $\mathbb{R}^{\dim(\text{Ker}(A))}$ dans $\text{Ker}(A)$, donc injective de $\mathbb{R}^{\dim(\text{Ker}(A))}$ dans \mathbb{R}^p . La solution à ce nouveau problème est donc unique. \square