$$Q = \begin{bmatrix} \frac{1}{\sqrt{2}} & \frac{1}{\sqrt{10}} & -\frac{1}{\sqrt{115}} \\ \frac{1}{\sqrt{2}} & -\frac{1}{\sqrt{10}} & \frac{1}{\sqrt{115}} \\ 0 & \frac{2}{\sqrt{10}} & -\frac{2}{\sqrt{115}} \\ 0 & 0 & \frac{10}{\sqrt{115}} \\ 0 & \frac{2}{\sqrt{10}} & \frac{3}{\sqrt{115}} \end{bmatrix}.$$

Étape 3 : matrice $R = Q^T A$.
Les coefficients de R sont $r_{ij} = \overrightarrow{q}_i \cdot \overrightarrow{a}_j$. On obtient

$$R = \begin{bmatrix} \sqrt{2} & \frac{1}{\sqrt{2}} & \frac{1}{\sqrt{2}} \\ 0 & \frac{\sqrt{10}}{2} & -\frac{3}{\sqrt{10}} \\ 0 & 0 & \frac{\sqrt{115}}{5} \end{bmatrix}.$$

On vérifie que $A = QR$ et que Q a des colonnes orthonormées et R est triangulaire supérieure : c'est une factorisation QR de A.

6.9 La méthode des moindres carrés

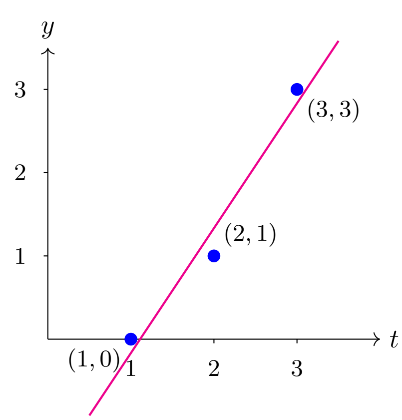
La méthode des moindres carrés permet de trouver la « meilleure » solution approximative d'un système d'équations linéaires qui n'a pas de solution exacte.

Motivation

Considérons trois points du plan :

$$(1,0), (2,1), (3,3)$$

On constate qu'ils ne sont pas alignés !
Question :
Quelle est l'équation de la droite du plan la plus « proche » de ces points ?



Soit $y = x_1 + x_2 t$ l'équation cherchée, où x_1, x_2 sont à déterminer.
On aimerait que les trois points donnés soient sur la droite :

$$\begin{aligned}(t_1, y_1) = (1, 0) : \quad & 0 = x_1 + 1 \cdot x_2 \\ (t_2, y_2) = (2, 1) : \quad & 1 = x_1 + 2 \cdot x_2 \\ (t_3, y_3) = (3, 3) : \quad & 3 = x_1 + 3 \cdot x_2\end{aligned}$$

Équation matricielle associée : $A\overrightarrow{x} = \overrightarrow{b}$ où $A = \begin{bmatrix} 1 & 1 \\ 1 & 2 \\ 1 & 3 \end{bmatrix}$ et $\overrightarrow{b} = \begin{bmatrix} 0 \\ 1 \\ 3 \end{bmatrix}$

Matrice augmentée :

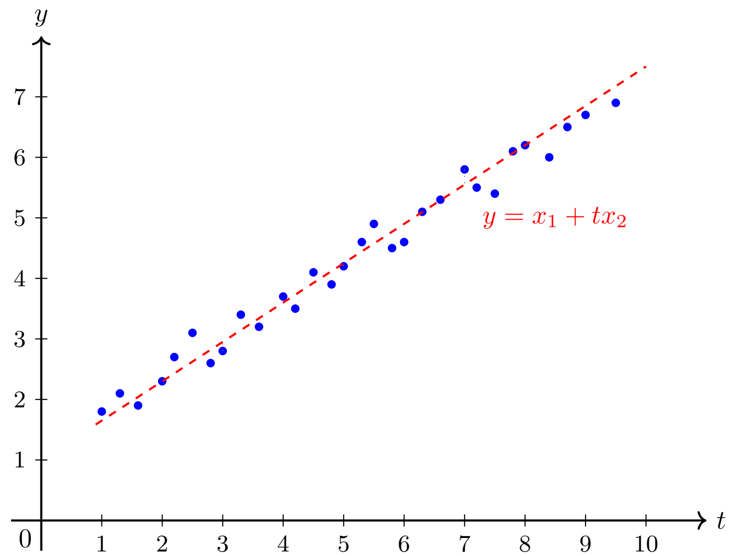
$$\begin{bmatrix} 1 & 1 & 0 \\ 1 & 2 & 1 \\ 1 & 3 & 3 \end{bmatrix} \sim \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix}$$

Cette équation n'a donc pas de solution (la dernière ligne donne $0 = 1$) ce qui confirme que les points ne sont pas alignés. Aucune droite ne passe donc par ces trois points. Nous allons donc non pas « résoudre » mais le « résoudre approximativement ». C'est un problème très classique de régression linéaire.

Cas général

Soit A une matrice de taille $m \times n$ et $\overrightarrow{b} \in \mathbb{R}^m$. Nous avons vu que le système $A\overrightarrow{x} = \overrightarrow{b}$ possède des solutions si et seulement si $\overrightarrow{b} \in \mathrm{Im}(A)$. En particulier, si $\overrightarrow{b} \notin \mathrm{Im}(A)$ alors $A\overrightarrow{x} = \overrightarrow{b}$ ne possède pas de solution !

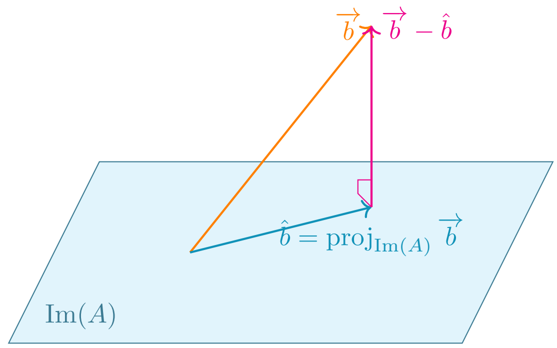
En pratique, on s'intéresse à des situations où $m \gg n$ (m beaucoup plus grand que n). Il est alors, dans les cas pratiques, très probable que $\overrightarrow{b} \notin \mathrm{Im}(A)$. Dans l'exemple précédent, cela revient à avoir un très grand nombre de points (beaucoup plus que 3), et à chercher une droite qui les relie. Si ces points sont issus de données expérimentales, il est peu probable qu'ils soient alignés. On va alors chercher la droite « la plus proche globalement » de tous ces points.



Ici, on cherche à résoudre un système dont la matrice augmentée est $A = \begin{bmatrix} 1 & 1.0 & 1.8 \\ 1 & 1.3 & 2.1 \\ 1 & 1.6 & 1.9 \\ 1 & 2.0 & 2.3 \\ 1 & 2.2 & 2.7 \\ 1 & 2.5 & 3.1 \\ 1 & 2.8 & 2.6 \\ 1 & 3.0 & 2.8 \\ 1 & 3.3 & 3.4 \\ 1 & 3.6 & 3.2 \\ 1 & 4.0 & 3.7 \\ 1 & 4.2 & 3.5 \\ 1 & 4.5 & 4.1 \\ 1 & 4.8 & 3.9 \\ 1 & 5.0 & 4.2 \\ 1 & 5.3 & 4.6 \\ 1 & 5.5 & 4.9 \\ 1 & 5.8 & 4.5 \\ 1 & 6.0 & 4.6 \\ 1 & 6.3 & 5.1 \\ 1 & 6.6 & 5.3 \\ 1 & 7.0 & 5.8 \\ 1 & 7.2 & 5.5 \\ 1 & 7.5 & 5.4 \\ 1 & 7.8 & 6.1 \\ 1 & 8.0 & 6.2 \\ 1 & 8.4 & 6.0 \\ 1 & 8.7 & 6.5 \\ 1 & 9.0 & 6.7 \\ 1 & 9.5 & 6.9 \end{bmatrix}$.

Considérons $\hat{b} = \mathrm{proj}_{\mathrm{Im}(A)}\,\overrightarrow{b}$.

Géométriquement :



Comme par construction $\hat{b} \in \mathrm{Im}(A)$, le système $A\overrightarrow{x} = \hat{b}$ est consistant et il existe au moins un $\hat{x} \in \mathbb{R}^n$ tel que $A\hat{x} = \hat{b}$.

Comme par ailleurs, $\hat{b}$ est d'après le théorème de la meilleure approximation 6.36, le vecteur de $\mathrm{Im}(A)$ le plus proche de $\overrightarrow{b}$, en résolvant $A\hat{x} = \hat{b}$ au lieu de $A\overrightarrow{x} = \overrightarrow{b}$, on a la meilleure approximation des solutions.

Définition 6.52 *Solution des moindres carrés*

Soit A une matrice $m \times n$ et $\overrightarrow{b} \in \mathbb{R}^m$.
Une *solution des moindres carrés* de $A\overrightarrow{x} = \overrightarrow{b}$ est un vecteur $\hat{x} \in \mathbb{R}^n$ qui minimise :

$$\|A\overrightarrow{x} - \overrightarrow{b}\|$$

Autrement dit : $\|A\hat{x} - \overrightarrow{b}\| \leq \|A\overrightarrow{x} - \overrightarrow{b}\|$ pour tout $\overrightarrow{x} \in \mathbb{R}^n$.

La distance $\|A\hat{x} - \overrightarrow{b}\|$ est appelée *erreur de l'approximation.*

Comment calculer $\hat{x}$?

Théorème 6.53 $A\hat{x} = \hat{b} \Leftrightarrow A^T A\hat{x} = A^T\overrightarrow{b}$

Le vecteur $\hat{x}$ est une solution des moindres carrés de $A\overrightarrow{x} = \overrightarrow{b}$ si et seulement si :

$$A^T A\hat{x} = A^T\overrightarrow{b}$$

.

Démonstration. ($\Rightarrow$) Commençons par montrer que si $\hat{x}$ est une solution des moindres carrés de $A\overrightarrow{x} = \overrightarrow{b}$ alors $A^T A\hat{x} = A^T\overrightarrow{b}$.

En notant $\hat{b} = A\hat{x}$ où $\hat{x}$ est la solution des moindres carrés, on a $\overrightarrow{b} - \hat{b} \in (\mathrm{Im}(A))^\perp$.

Donc $\overrightarrow{b} - A\hat{x} \in (\mathrm{Im}(A))^\perp$, ce qui signifie que $\overrightarrow{b} - A\hat{x}$ est orthogonal à chaque colonne de

A, donc que le produit scalaire entre chaque colonne de A et $\overrightarrow{b} - A\hat{x}$ est nul.

Par définition du produit scalaire, ceci équivaut à $A^T(\overrightarrow{b} - A\hat{x}) = \overrightarrow{0}$, soit $A^T A\hat{x} = A^T\overrightarrow{b}$.

($\Leftarrow$) Montrons maintenant qu'une solution de $A^T A\hat{x} = A^T\overrightarrow{b}$ est une solution au sens des moindres carrés de $A\overrightarrow{x} = \overrightarrow{b}$.

Si on a $A^T A\hat{x} = A^T\overrightarrow{b}$, alors $A^T(A\hat{x} - \overrightarrow{b}) = \overrightarrow{0}$.

Par conséquent, $\overrightarrow{b} - A\hat{x} \in \mathrm{Ker}(A^T) = (\mathrm{Im}(A))^\perp$ (propriété 6.16).

Comme $\overrightarrow{b} = \underbrace{(\overrightarrow{b} - A\hat{x})}_{\in(\mathrm{Im}(A))^\perp} + \underbrace{A\hat{x}}_{\in\mathrm{Im}(A)}$, l'unicité de la décomposition orthogonale nous dit que

$$A\hat{x} = \mathrm{proj}_{\mathrm{Im}(A)}\overrightarrow{b} = \hat{b}$$

autrement dit, $\hat{x}$ est une solution au sens des moindres carrés de $A\overrightarrow{x} = \overrightarrow{b}$.

$\square$

Remarques 6.6.0.54. *Par construction de $\hat{b}$ et $\hat{x}$, l'équation $A^T A\hat{x} = A^T\overrightarrow{b}$ admet au moins une solution.*

Définition 6.55 *Système normal*

Soit A une matrice $m \times n$, $\overrightarrow{b} \in \mathbb{R}^m$. Le système $(A^T A)\overrightarrow{x} = A^T\overrightarrow{b}$ est appelé système normal associé au système $A\overrightarrow{x} = \overrightarrow{b}$.

Exemples. Reprenons l'exemple introductif de régression linéaire.

Soit $A = \begin{bmatrix} 1 & 1 \\ 1 & 2 \\ 1 & 3 \end{bmatrix}$ et $\overrightarrow{b} = \begin{bmatrix} 0 \\ 1 \\ 3 \end{bmatrix}$

Trouver une solution au sens des moindres carrés du système $A\overrightarrow{x} = \overrightarrow{b}$.

On a

$$A^T A = \begin{bmatrix} 1 & 1 & 1 \\ 1 & 2 & 3 \end{bmatrix} \begin{bmatrix} 1 & 1 \\ 1 & 2 \\ 1 & 3 \end{bmatrix} = \begin{bmatrix} 3 & 6 \\ 6 & 14 \end{bmatrix}$$

et

$$A^T\overrightarrow{b} = \begin{bmatrix} 1 & 1 & 1 \\ 1 & 2 & 3 \end{bmatrix} \begin{bmatrix} 0 \\ 1 \\ 3 \end{bmatrix} = \begin{bmatrix} 4 \\ 11 \end{bmatrix}$$

Matrice augmentée associée au système normal :

$$\begin{bmatrix} 3 & 6 & 4 \\ 6 & 14 & 11 \end{bmatrix} \sim \begin{bmatrix} 1 & 0 & -5/3 \\ 0 & 1 & 3/2 \end{bmatrix}$$

Donc $\begin{bmatrix}\hat{x}_1\\ \hat{x}_2\end{bmatrix} = \begin{bmatrix}-5/3\\ 3/2\end{bmatrix}$ et la droite qui s'approche le mieux (au sens des moindres carrés) des points $(1,0),(2,1),(3,3)$ est $y = -\dfrac{5}{3} + \dfrac{3}{2}t$.

De manière générale, la droite de régression linéaire pour les points

$$(t_1,y_1),(t_2,y_2),\ldots,(t_N,y_N) \in \mathbb{R}^2$$

peut être trouvée à l'aide de la solution au sens des moindres carrés du système $A\overrightarrow{x} = \overrightarrow{b}$ où

$$A = \begin{bmatrix}1 & t_1\\ 1 & t_2\\ \vdots & \vdots\\ 1 & t_N\end{bmatrix} \in \mathcal{M}_{N,2}(\mathbb{R}) \quad \text{et} \quad \overrightarrow{b} = \begin{bmatrix}y_1\\ y_2\\ \vdots\\ y_N\end{bmatrix} \in \mathbb{R}^N$$

Si $\hat{x} = \begin{bmatrix}\hat{x}_1\\ \hat{x}_2\end{bmatrix}$ est la solution au sens des moindres carrés, alors l'équation de la droite de régression linéaire est :

$$y = \hat{x}_1 + \hat{x}_2 t$$

Exemples. Soit $A = \begin{bmatrix}1 & -2\\ 1 & -1\\ 1 & 0\\ 1 & 1\\ 1 & 2\end{bmatrix}$ et $\overrightarrow{b} = \begin{bmatrix}4\\ 2\\ 1\\ 3\\ 7\end{bmatrix}$

On constate que le système $A\overrightarrow{x} = \overrightarrow{b}$ n'a pas de solution.

Trouver une solution au sens des moindres carrés du système $A\overrightarrow{x} = \overrightarrow{b}$.

On a

$$A^T A = \begin{bmatrix}1 & 1 & 1 & 1 & 1\\ -2 & -1 & 0 & 1 & 2\end{bmatrix}\begin{bmatrix}1 & -2\\ 1 & -1\\ 1 & 0\\ 1 & 1\\ 1 & 2\end{bmatrix} = \begin{bmatrix}5 & 0\\ 0 & 10\end{bmatrix}$$

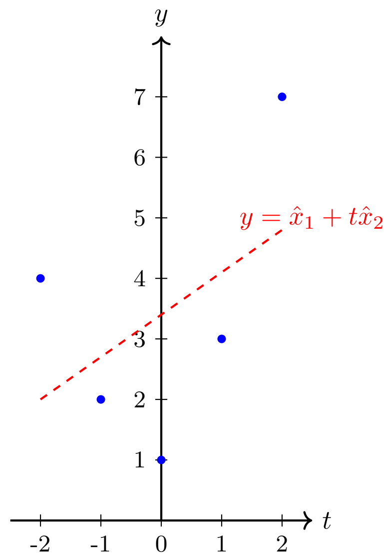
et

$$A^T \overrightarrow{b} = \begin{bmatrix}1 & 1 & 1 & 1 & 1\\ -2 & -1 & 0 & 1 & 2\end{bmatrix}\begin{bmatrix}4\\ 2\\ 1\\ 3\\ 7\end{bmatrix} = \begin{bmatrix}17\\ 7\end{bmatrix}$$

d'où

$$\begin{bmatrix}5 & 0 & 17\\ 0 & 10 & 7\end{bmatrix} \sim \begin{bmatrix}1 & 0 & 17/5\\ 0 & 1 & 7/10\end{bmatrix}$$

Donc la solution est $\begin{bmatrix}\hat{x}_1\\ \hat{x}_2\end{bmatrix} = \begin{bmatrix}17/5\\ 7/10\end{bmatrix}$



Exemples. 1. Soit $A = \begin{bmatrix} 1 & 1 & 0 \\ 1 & 0 & 1 \\ 0 & 1 & -1 \\ 0 & 0 & 2 \\ 0 & 1 & 0 \end{bmatrix}$ et $\overrightarrow{b} = \begin{bmatrix} 1 \\ 0 \\ 2 \\ 1 \\ -1 \end{bmatrix}$.

Trouver une solution au sens des moindres carrés du système $A\overrightarrow{x} = \overrightarrow{b}$.

On a

$$A^T A = \begin{bmatrix} 1 & 1 & 0 & 0 & 0 \\ 1 & 0 & 1 & 0 & 1 \\ 0 & 1 & -1 & 2 & 0 \end{bmatrix} \begin{bmatrix} 1 & 1 & 0 \\ 1 & 0 & 1 \\ 0 & 1 & -1 \\ 0 & 0 & 2 \\ 0 & 1 & 0 \end{bmatrix} = \begin{bmatrix} 2 & 1 & 1 \\ 1 & 3 & -1 \\ 1 & -1 & 6 \end{bmatrix}.$$

De même,

$$A^T \overrightarrow{b} = \begin{bmatrix} 1 & 1 & 0 & 0 & 0 \\ 1 & 0 & 1 & 0 & 1 \\ 0 & 1 & -1 & 2 & 0 \end{bmatrix} \begin{bmatrix} 1 \\ 0 \\ 2 \\ 1 \\ -1 \end{bmatrix} = \begin{bmatrix} 1 \\ 2 \\ 0 \end{bmatrix}.$$

Le système normal est donc

$$(A^T A)\overrightarrow{x} = A^T \overrightarrow{b}, \quad \text{soit} \quad \begin{bmatrix} 2 & 1 & 1 \\ 1 & 3 & -1 \\ 1 & -1 & 6 \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \\ x_3 \end{bmatrix} = \begin{bmatrix} 1 \\ 2 \\ 0 \end{bmatrix}.$$

Matrice augmentée associée :

$$\begin{bmatrix} 2 & 1 & 1 & 1 \\ 1 & 3 & -1 & 2 \\ 1 & -1 & 6 & 0 \end{bmatrix} \sim \begin{bmatrix} 1 & 0 & 0 & 3/23 \\ 0 & 1 & 0 & 15/23 \\ 0 & 0 & 1 & 2/23 \end{bmatrix}.$$

Donc

$$\begin{bmatrix} \hat{x}_1 \\ \hat{x}_2 \\ \hat{x}_3 \end{bmatrix} = \begin{bmatrix} 3/23 \\ 15/23 \\ 2/23 \end{bmatrix}.$$

2. Soit $A = \begin{bmatrix} 1 & 0 & 1 \\ 1 & 1 & 0 \\ 1 & 0 & 2 \\ 1 & 2 & 0 \\ 1 & 1 & 1 \end{bmatrix}$ et $\overrightarrow{b} = \begin{bmatrix} 3 \\ 2 \\ 5 \\ 4 \\ 4 \end{bmatrix}$

On constate que le système $A\overrightarrow{x} = \overrightarrow{b}$ n'a pas de solution.

Trouver une solution au sens des moindres carrés du système $A\overrightarrow{x} = \overrightarrow{b}$.

On a

$$\begin{aligned} A^T A &= \begin{bmatrix} 1 & 1 & 1 & 1 & 1 \\ 0 & 1 & 0 & 2 & 1 \\ 1 & 0 & 2 & 0 & 1 \end{bmatrix} \begin{bmatrix} 1 & 0 & 1 \\ 1 & 1 & 0 \\ 1 & 0 & 2 \\ 1 & 2 & 0 \\ 1 & 1 & 1 \end{bmatrix} \\ &= \begin{bmatrix} 5 & 4 & 4 \\ 4 & 6 & 1 \\ 4 & 1 & 6 \end{bmatrix} \end{aligned}$$

et

$$A^T \overrightarrow{b} = \begin{bmatrix} 1 & 1 & 1 & 1 & 1 \\ 0 & 1 & 0 & 2 & 1 \\ 1 & 0 & 2 & 0 & 1 \end{bmatrix} \begin{bmatrix} 3 \\ 2 \\ 5 \\ 4 \\ 4 \end{bmatrix} = \begin{bmatrix} 18 \\ 14 \\ 17 \end{bmatrix}$$

Matrice augmentée associée au système normal :

$$\begin{bmatrix} 5 & 4 & 4 & 18 \\ 4 & 6 & 1 & 14 \\ 4 & 1 & 6 & 17 \end{bmatrix} \sim \begin{bmatrix} 1 & 0 & 0 & \frac{2}{3} \\ 0 & 1 & 0 & \frac{23}{15} \\ 0 & 0 & 1 & \frac{32}{15} \end{bmatrix}$$

Donc $\begin{bmatrix} \hat{x}_1 \\ \hat{x}_2 \\ \hat{x}_3 \end{bmatrix} = \begin{bmatrix} 2/3 \\ 23/15 \\ 32/15 \end{bmatrix}$

Théorème 6.56

Soit A une matrice de taille $m \times n$.
Les affirmations suivantes sont équivalentes :

1. $A\overrightarrow{x} = \overrightarrow{b}$ admet une unique solution au sens des moindres carrés pour tout $\overrightarrow{b} \in \mathbb{R}^m$.
2. Les colonnes de A sont linéairement indépendantes.
3. $\text{rang}(A) = n$
4. $\text{rang}(A^T A) = n$
5. La matrice $A^T A$ est inversible.

Dans ce cas, la solution au sens des moindres carrés s'écrit

$$\hat{x} = (A^T A)^{-1} A^T \overrightarrow{b}$$

Démonstration. Les équivalences (1) $\Leftrightarrow$ (2) $\Leftrightarrow$ (3) découlent de ce qui précède et du théorème du rang.

L'équivalence (4) $\Leftrightarrow$ (5) découle du théorème du rang appliqué à $A^T A$.

Le seul point à montrer est (3) $\Leftrightarrow$ (4). Nous allons montrer que $\text{Ker}(A^T A) = \text{Ker}(A)$ et conclure grâce au théorème du rang.

— $\text{Ker}(A) \subset \text{Ker}(A^T A)$

Si $\overrightarrow{x} \in \text{Ker}(A)$, alors $A\overrightarrow{x} = \overrightarrow{0}$, donc

$$A^T A\overrightarrow{x} = A^T \overrightarrow{0} = \overrightarrow{0},$$

donc $\overrightarrow{x} \in \text{Ker}(A^T A)$.

Ainsi, $\text{Ker}(A) \subset \text{Ker}(A^T A)$.

— $\text{Ker}(A^T A) \subset \text{Ker}(A)$.

Soit maintenant $\overrightarrow{x} \in \text{Ker}(A^T A)$, c'est-à-dire

$$A^T A\overrightarrow{x} = \overrightarrow{0}.$$

On prend le produit scalaire avec $\overrightarrow{x}$:

$$\overrightarrow{x} \cdot (A^T A\overrightarrow{x}) = \overrightarrow{x} \cdot \overrightarrow{0} = 0.$$

En notation matricielle :

$$\overrightarrow{x} \cdot (A^T A\overrightarrow{x}) = \overrightarrow{x}^T (A^T A\overrightarrow{x}) = (A\overrightarrow{x}) \cdot (A\overrightarrow{x}) = \|A\overrightarrow{x}\|^2.$$

On obtient donc

$$\|A\overrightarrow{x}\|^2 = 0.$$

Par positivité de la norme, cela implique

$$A\overrightarrow{x} = \overrightarrow{0},$$

c'est-à-dire $\overrightarrow{x} \in \mathrm{Ker}(A)$.

On a ainsi montré $\mathrm{Ker}(A^TA) \subset \mathrm{Ker}(A)$.

□

Théorème 6.57

Soit A une matrice de taille $m \times n$ dont les colonnes sont linéairement indépendantes. Soit $A = QR$ une factorisation QR de A. La solution au sens des moindres carrés de $A\overrightarrow{x} = \overrightarrow{b}$ est unique et s'écrit :

$$\hat{x} = R^{-1}Q^T\overrightarrow{b}$$

Démonstration. Comme les colonnes de A sont linéairement indépendantes, la solution au sens des moindres carrés est unique :

$$\hat{x} = (A^TA)^{-1}A^T\overrightarrow{b}$$

Par hypothèse, A peut s'écrire de manière unique sous la forme $A = QR$ où Q est une matrice de taille $m \times n$ telle que $Q^TQ = \mathbb{I}_n$ et R est une matrice de taille $n \times n$ triangulaire supérieure inversible (théorème 6.49).

On a $A^T = R^TQ^T$ et

$$A^TA = (R^TQ^T)(QR) = R^T(Q^TQ)R = R^TR,$$

d'où

$$(A^TA)^{-1}A^T = (R^TR)^{-1}R^TQ^T = R^{-1}(R^T)^{-1}R^TQ^T = R^{-1}Q^T$$

et $\hat{x} = R^{-1}Q^T\overrightarrow{b}$ □

Exemple. Soit $A = \begin{bmatrix} 1 & 0 \\ 1 & 1 \\ 0 & 1 \end{bmatrix}$ et soit

$$\overrightarrow{b} = \begin{bmatrix} 2 \\ 3 \\ 1 \end{bmatrix}.$$

On a obtenu précédemment une factorisation

$$A = QR$$

avec

$$Q = \begin{bmatrix} \frac{1}{\sqrt{2}} & -\frac{1}{\sqrt{6}} \\ \frac{1}{\sqrt{2}} & \frac{1}{\sqrt{6}} \\ 0 & \frac{2}{\sqrt{6}} \end{bmatrix}, \qquad R = \begin{bmatrix} \sqrt{2} & \frac{1}{\sqrt{2}} \\ 0 & \sqrt{\frac{3}{2}} \end{bmatrix}.$$

Le théorème assure que la solution des moindres carrés de $A\overrightarrow{x} = \overrightarrow{b}$ est

$$\hat{x} = R^{-1}Q^T\overrightarrow{b}.$$

Calcul de $Q^T \overrightarrow{b}$.

$$Q^T \overrightarrow{b} = \begin{bmatrix} \frac{1}{\sqrt{2}} & \frac{1}{\sqrt{2}} & 0 \\ -\frac{1}{\sqrt{6}} & \frac{1}{\sqrt{6}} & \frac{2}{\sqrt{6}} \end{bmatrix} \begin{bmatrix} 2 \\ 3 \\ 1 \end{bmatrix} = \begin{bmatrix} \frac{5}{\sqrt{2}} \\ \frac{3}{\sqrt{6}} \end{bmatrix}.$$

Calcul de $R^{-1} Q^T \overrightarrow{b}$.
Comme

$$R^{-1} = \begin{bmatrix} \frac{1}{\sqrt{2}} & -\frac{1}{\sqrt{6}} \\ 0 & \sqrt{\frac{2}{3}} \end{bmatrix},$$

on obtient

$$\hat{x} = R^{-1} \begin{bmatrix} \frac{5}{\sqrt{2}} \\ \frac{3}{\sqrt{6}} \end{bmatrix} = \begin{bmatrix} \frac{5}{2} - \frac{3}{6} \\ \frac{2}{3} \end{bmatrix} = \begin{bmatrix} 2 \\ 1 \end{bmatrix}.$$

C'est la solution unique au sens des moindres carrés du système $A\overrightarrow{x} = \overrightarrow{b}$.

Exemple. Reprenons la matrice

$$A = \begin{bmatrix} 1 & 1 & 0 \\ 1 & 0 & 1 \\ 0 & 1 & -1 \\ 0 & 0 & 2 \\ 0 & 1 & 0 \end{bmatrix}, \qquad \overrightarrow{b} = \begin{bmatrix} 3 \\ 2 \\ -1 \\ 4 \\ 1 \end{bmatrix}.$$

La factorisation QR obtenue précédemment est

$$A = QR$$

avec

$$Q = \begin{bmatrix} \frac{1}{\sqrt{2}} & \frac{1}{\sqrt{10}} & -\frac{1}{\sqrt{115}} \\ \frac{1}{\sqrt{2}} & -\frac{1}{\sqrt{10}} & \frac{1}{\sqrt{115}} \\ 0 & \frac{2}{\sqrt{10}} & -\frac{2}{\sqrt{115}} \\ 0 & 0 & \frac{10}{\sqrt{115}} \\ 0 & \frac{2}{\sqrt{10}} & \frac{3}{\sqrt{115}} \end{bmatrix},$$

et

$$R = \begin{bmatrix} \sqrt{2} & \frac{1}{\sqrt{2}} & \frac{1}{\sqrt{2}} \\ 0 & \frac{\sqrt{10}}{2} & -\frac{3}{\sqrt{10}} \\ 0 & 0 & \frac{\sqrt{115}}{5} \end{bmatrix}.$$

Le théorème dit :
$$\hat{x} = R^{-1}Q^T\overrightarrow{b}.$$

Étape 1 : calcul de $Q^T\overrightarrow{b}$.
On obtient
$$Q^T\overrightarrow{b} = \begin{bmatrix} \frac{5}{\sqrt{2}} \\ \frac{1}{\sqrt{10}} \\ \frac{44}{\sqrt{115}} \end{bmatrix}.$$

Étape 2 : calcul de $R^{-1}Q^T\overrightarrow{b}$.
L'inverse de R est
$$R^{-1} = \begin{bmatrix} \frac{1}{\sqrt{2}} & -\frac{1}{\sqrt{10}} & -\frac{4}{\sqrt{115}} \\ 0 & \frac{2}{\sqrt{10}} & \frac{3}{\sqrt{115}} \\ 0 & 0 & \frac{5}{\sqrt{115}} \end{bmatrix}.$$

On obtient donc
$$\hat{x} = R^{-1}Q^T\overrightarrow{b} = \begin{bmatrix} \frac{1}{\sqrt{2}} & -\frac{1}{\sqrt{10}} & -\frac{4}{\sqrt{115}} \\ 0 & \frac{2}{\sqrt{10}} & \frac{3}{\sqrt{115}} \\ 0 & 0 & \frac{5}{\sqrt{115}} \end{bmatrix} \begin{bmatrix} \frac{5}{\sqrt{2}} \\ \frac{1}{\sqrt{10}} \\ \frac{44}{\sqrt{115}} \end{bmatrix}.$$

Un calcul direct donne
$$\hat{x} = \begin{bmatrix} \frac{20}{23} \\ \frac{31}{23} \\ \frac{44}{23} \end{bmatrix}.$$

C'est la solution unique au sens des moindres carrés du système $A\overrightarrow{x} = \overrightarrow{b}$.

Remarque importante 6.58

Nous avons défini $\hat{b} = \text{proj}_{\text{Im}(A)}\overrightarrow{b}$

Comme $\hat{x} \in \mathbb{R}^n$ est tel que $A\hat{x} = \hat{b}$, pour trouver la projection orthogonale de $\overrightarrow{b}$ sur $\text{Im}(A)$ il suffit de trouver une solution $\hat{x}$ du système normal $(A^TA)\overrightarrow{x} = A^T\overrightarrow{b}$ et la multiplier par A :
$$\text{proj}_{\text{Im}(A)}\overrightarrow{b} = A\hat{x}$$
plutôt que de construire une base orthogonale à l'aide du procédé d'orthogonalisation de Gram-Schmidt.

Théorème 6.59 *Matrice associée à* proj_W

Soit W un sous-espace vectoriel de $\mathbb{R}^n$ de dimension $k > 0$.
Soit $(\vec{w}_1, \dots, \vec{w}_k)$ une base quelconque de W (pas forcément orthogonale). Soit A la matrice de taille $n \times k$ suivante

$$A = \begin{bmatrix} \vec{w}_1 & \vec{w}_2 & \cdots & \vec{w}_k \end{bmatrix}$$

alors la matrice canoniquement associée à l'application linéaire $\mathrm{proj}_W : \mathbb{R}^n \to \mathbb{R}^n$ est

$$A_{\mathrm{proj}_W} = A(A^TA)^{-1}A^T.$$

Démonstration. Comme $(\vec{w}_1, \dots, \vec{w}_k)$ est une base de W et $A = \begin{bmatrix} \vec{w}_1 & \vec{w}_2 & \cdots & \vec{w}_k \end{bmatrix}$, nous avons $W = \mathrm{Im}(A)$.

Par conséquent, $W^\perp = (\mathrm{Im}(A))^\perp = \mathrm{Ker}(A^T)$

Le théorème de la projection orthogonale 6.32 nous dit que tout vecteur $\vec{v} \in \mathbb{R}^n$ s'écrit de manière unique

$$\vec{v} = \vec{w} + \vec{w}', \text{ où } \vec{w} \in W \text{ et } \vec{w}' \in W^\perp.$$

Comme $\vec{w} \in W = \mathrm{Im}(A)$, on a $\vec{w} = A\vec{x}$, avec $\vec{x} \in \mathbb{R}^k$.

Comme $\vec{w}' \in W^\perp = \mathrm{Ker}(A^T)$, on a $A^T\vec{w}' = \vec{0}$.

Ainsi $A^T\vec{v} = A^T(\vec{w} + \vec{w}') = A^T\vec{w} + A^T\vec{w}' = A^TA\vec{x} + \vec{0}$.

d'où $A^T\vec{v} = A^TA\vec{x}$.

Comme A^TA est une matrice carrée de taille $k \times k$ et $\mathrm{rang}(A^TA) = \mathrm{rang}(A)$ (théorème 6.56) la matrice A^TA est inversible car $\mathrm{rang}(A^TA) = k$.

Nous avons donc $\vec{x} = (A^TA)^{-1}A^T\vec{v}$, d'où

$$\mathrm{proj}_W \vec{v} = \vec{w} = A\vec{x} = A(A^TA)^{-1}A^T\vec{v}$$

□

Exemple. Considérons le sous-espace vectoriel

$$W = \mathrm{Vect}\left(\begin{bmatrix}1\\1\\0\end{bmatrix}, \begin{bmatrix}1\\0\\1\end{bmatrix}\right).$$

Une base de W étant donnée par

$$\vec{w}_1 = \begin{bmatrix}1\\1\\0\end{bmatrix}, \qquad \vec{w}_2 = \begin{bmatrix}1\\0\\1\end{bmatrix},$$

la matrice

$$A = \begin{bmatrix} 1 & 1 \\ 1 & 0 \\ 0 & 1 \end{bmatrix}$$

a pour colonnes les vecteurs de cette base.

Le théorème affirme que la matrice associée à la projection orthogonale $\mathrm{proj}_W : \mathbb{R}^3 \to \mathbb{R}^3$ est

$$A_{\mathrm{proj}_W} = A(A^T A)^{-1} A^T.$$

Calculons chaque terme.

$$A^T A = \begin{bmatrix} 1 & 1 & 0 \\ 1 & 0 & 1 \end{bmatrix} \begin{bmatrix} 1 & 1 \\ 1 & 0 \\ 0 & 1 \end{bmatrix} = \begin{bmatrix} 2 & 1 \\ 1 & 2 \end{bmatrix}.$$

$$(A^T A)^{-1} = \frac{1}{3} \begin{bmatrix} 2 & -1 \\ -1 & 2 \end{bmatrix}.$$

Ainsi,

$$A(A^T A)^{-1} = \frac{1}{3} \begin{bmatrix} 1 & 1 \\ 1 & 0 \\ 0 & 1 \end{bmatrix} \begin{bmatrix} 2 & -1 \\ -1 & 2 \end{bmatrix} = \frac{1}{3} \begin{bmatrix} 1 & 1 \\ 2 & -1 \\ -1 & 2 \end{bmatrix}.$$

Enfin,

$$A_{\mathrm{proj}_W} = A(A^T A)^{-1} A^T = \frac{1}{3} \begin{bmatrix} 1 & 1 \\ 2 & -1 \\ -1 & 2 \end{bmatrix} \begin{bmatrix} 1 & 1 & 0 \\ 1 & 0 & 1 \end{bmatrix} = \frac{1}{3} \begin{bmatrix} 2 & 1 & 1 \\ 1 & 2 & -1 \\ 1 & -1 & 2 \end{bmatrix}.$$

Ainsi, pour tout $\overrightarrow{x} \in \mathbb{R}^3$:

$$\mathrm{proj}_W(\overrightarrow{x}) = A_{\mathrm{proj}_W} \overrightarrow{x} = \frac{1}{3} \begin{bmatrix} 2 & 1 & 1 \\ 1 & 2 & -1 \\ 1 & -1 & 2 \end{bmatrix} \overrightarrow{x}.$$

6.10 Angle entre deux vecteurs

Théorème 6.60 *Inégalité de Cauchy-Schwarz*

Pour tous vecteurs $\overrightarrow{u}, \overrightarrow{v} \in \mathbb{R}^n$:

$$|\overrightarrow{u} \cdot \overrightarrow{v}| \leq \|\overrightarrow{u}\| \, \|\overrightarrow{v}\|$$

L'égalité a lieu si et seulement si $\overrightarrow{u}$ et $\overrightarrow{v}$ sont colinéaires.

Démonstration. Si $\overrightarrow{u} = \overrightarrow{0}$, l'inégalité est vérifiée (avec égalité).

Supposons $\overrightarrow{u} \neq \overrightarrow{0}$. La norme de la projection orthogonale de $\overrightarrow{v}$ sur $\overrightarrow{u}$ vaut :

$$\| \operatorname{proj}_{\overrightarrow{u}} \overrightarrow{v} \| = \left\| \frac{\overrightarrow{v} \cdot \overrightarrow{u}}{\overrightarrow{u} \cdot \overrightarrow{u}} \overrightarrow{u} \right\| = \frac{|\overrightarrow{u} \cdot \overrightarrow{v}|}{\|\overrightarrow{u}\|}$$

Le théorème de Pythagore appliqué à $\overrightarrow{v} = \operatorname{proj}_{\overrightarrow{u}} \overrightarrow{v} + (\overrightarrow{v} - \operatorname{proj}_{\overrightarrow{u}} \overrightarrow{v})$ donne :

$$\|\overrightarrow{v}\|^2 = \| \operatorname{proj}_{\overrightarrow{u}} \overrightarrow{v}\|^2 + \|\overrightarrow{v} - \operatorname{proj}_{\overrightarrow{u}} \overrightarrow{v}\|^2 \geq \| \operatorname{proj}_{\overrightarrow{u}} \overrightarrow{v}\|^2$$

Donc $\| \operatorname{proj}_{\overrightarrow{u}} \overrightarrow{v}\| \leq \|\overrightarrow{v}\|$, c'est-à-dire $\dfrac{|\overrightarrow{u} \cdot \overrightarrow{v}|}{\|\overrightarrow{u}\|} \leq \|\overrightarrow{v}\|$, d'où l'inégalité.

L'égalité a lieu si et seulement si $\overrightarrow{v} = \operatorname{proj}_{\overrightarrow{u}} \overrightarrow{v} \in \operatorname{Vect}(\overrightarrow{u})$. □

Définition 6.61 *Angle entre deux vecteurs*

Soit $\overrightarrow{u}, \overrightarrow{v} \in \mathbb{R}^n$ deux vecteurs non nuls. L'inégalité de Cauchy-Schwarz garantit que $\dfrac{\overrightarrow{u} \cdot \overrightarrow{v}}{\|\overrightarrow{u}\| \, \|\overrightarrow{v}\|} \in [-1, 1]$.

L'*angle* entre $\overrightarrow{u}$ et $\overrightarrow{v}$ est l'unique $\theta \in [0, \pi]$ tel que :

$$\cos\theta = \frac{\overrightarrow{u} \cdot \overrightarrow{v}}{\|\overrightarrow{u}\| \, \|\overrightarrow{v}\|}$$

De manière équivalente : $\overrightarrow{u} \cdot \overrightarrow{v} = \|\overrightarrow{u}\| \, \|\overrightarrow{v}\| \cos\theta$.

Exemples.

1. Angle entre $\overrightarrow{u} = \begin{bmatrix} 1 \\ 0 \end{bmatrix}$ et $\overrightarrow{v} = \begin{bmatrix} 1 \\ 1 \end{bmatrix}$ dans $\mathbb{R}^2$:

$$\cos\theta = \frac{1}{1 \cdot \sqrt{2}} = \frac{\sqrt{2}}{2} \quad \Rightarrow \quad \theta = \frac{\pi}{4}$$

2. Angle entre $\overrightarrow{u} = \begin{bmatrix} 1 \\ 2 \\ 2 \end{bmatrix}$ et $\overrightarrow{v} = \begin{bmatrix} 3 \\ -4 \\ 0 \end{bmatrix}$ dans $\mathbb{R}^3$:

$$\cos\theta = \frac{3 - 8 + 0}{3 \cdot 5} = -\frac{1}{3} \quad \Rightarrow \quad \theta = \arccos\left(-\frac{1}{3}\right) \approx 109.47$$

3. Les vecteurs $\overrightarrow{u} = \begin{bmatrix} 1 \\ 1 \\ 1 \end{bmatrix}$ et $\overrightarrow{v} = \begin{bmatrix} 1 \\ -1 \\ 0 \end{bmatrix}$ sont orthogonaux car $\overrightarrow{u} \cdot \overrightarrow{v} = 0$, donc $\theta = \dfrac{\pi}{2}$.

6.11 Espaces préhilbertiens

Définition 6.62 *Produit scalaire sur un espace vectoriel*

Soit V un espace vectoriel réel. Un *produit scalaire* sur V est une application $\langle \cdot, \cdot \rangle : V \times V \to \mathbb{R}$ vérifiant pour tout $u, v, w \in V$ et tout $c \in \mathbb{R}$:

Symétrie $\langle u, v \rangle = \langle v, u \rangle$

Linéarité $\langle u + v, w \rangle = \langle u, w \rangle + \langle v, w \rangle$

$\langle cu, v \rangle = c\langle u, v \rangle$

Positivité $\langle u, u \rangle \geq 0$, avec égalité si et seulement si $u = 0$

Un espace vectoriel muni d'un produit scalaire est appelé *espace préhilbertien.* S'il est de dimension finie, on parle d'*espace euclidien.*

Exemples (Produit scalaire pondéré sur $\mathbb{R}^2$)**.** Soit $a, b > 0$. Pour $\overrightarrow{u} = \begin{bmatrix} u_1 \\ u_2 \end{bmatrix}$ et $\overrightarrow{v} = \begin{bmatrix} v_1 \\ v_2 \end{bmatrix}$ dans $\mathbb{R}^2$, on pose :

$$\langle \overrightarrow{u}, \overrightarrow{v} \rangle = a\, u_1 v_1 + b\, u_2 v_2$$

Vérifions les axiomes :

— **Symétrie :** $\langle \overrightarrow{u}, \overrightarrow{v} \rangle = a\, u_1 v_1 + b\, u_2 v_2 = \langle \overrightarrow{v}, \overrightarrow{u} \rangle$

— **Linéarité :** $\langle \overrightarrow{u} + \overrightarrow{v}, \overrightarrow{w} \rangle = a(u_1 + v_1)w_1 + b(u_2 + v_2)w_2 = \langle \overrightarrow{u}, \overrightarrow{w} \rangle + \langle \overrightarrow{v}, \overrightarrow{w} \rangle$

— $\langle c\overrightarrow{u}, \overrightarrow{v} \rangle = a(cu_1)v_1 + b(cu_2)v_2 = c\langle \overrightarrow{u}, \overrightarrow{v} \rangle$

— **Positivité :** $\langle \overrightarrow{u}, \overrightarrow{u} \rangle = a\, u_1^2 + b\, u_2^2 \geq 0$, et $\langle \overrightarrow{u}, \overrightarrow{u} \rangle = 0$ ssi $\overrightarrow{u} = \overrightarrow{0}$

Par exemple, avec $a = 4$, $b = 5$, $\overrightarrow{u} = \begin{bmatrix} 2 \\ -1 \end{bmatrix}$ et $\overrightarrow{v} = \begin{bmatrix} 3 \\ 2 \end{bmatrix}$:

$$\langle \overrightarrow{u}, \overrightarrow{v} \rangle = 4 \cdot 2 \cdot 3 + 5 \cdot (-1) \cdot 2 = 24 - 10 = 14$$

Exemples. Produit scalaire sur les polynômes

Soit $t_0, t_1, \ldots, t_n$ des réels distincts. Pour $p, q \in \mathbb{P}_n$, on définit :

$$\langle p, q \rangle = \sum_{k=0}^{n} p(t_k)q(t_k)$$

Les axiomes de symétrie, linéarité et homogénéité sont immédiats. Pour la positivité : $\langle p, p \rangle = \sum_{k=0}^{n} [p(t_k)]^2 \geq 0$, et si $\langle p, p \rangle = 0$, alors p s'annule en $n + 1$ points distincts, donc $p = 0$.

Par exemple, dans $\mathbb{P}_2$ avec $t_0 = 0$, $t_1 = \frac{1}{2}$, $t_2 = 1$, pour $p(t) = 12t^2$ et $q(t) = 2t - 1$:

$$\langle p, q \rangle = p(0)q(0) + p\left(\tfrac{1}{2}\right) q\left(\tfrac{1}{2}\right) + p(1)q(1) = (0)(-1) + (3)(0) + (12)(1) = 12$$

Exemples. Produit scalaire intégral sur $C([a, b])$

Pour $f, g \in C([a, b])$, on définit :

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

Les axiomes de symétrie, linéarité et homogénéité découlent des propriétés de l'intégrale. Pour la positivité : $\langle f, f\rangle = \int_a^b [f(t)]^2\, dt \geq 0$, et si cette intégrale est nulle avec f continue, alors $f = 0$.

Par exemple, dans $C([0, 1])$ avec $f(t) = t$ et $g(t) = t^2$:

$$\langle f, g\rangle = \int_0^1 t^3\, dt = \frac{1}{4}, \quad \langle f, f\rangle = \int_0^1 t^2\, dt = \frac{1}{3}$$

Remarque 6.6.0.63. *Dans un espace préhilbertien de dimension finie, on définit la norme par* $\|v\| = \sqrt{\langle v, v\rangle}$*, la distance par* $d(u, v) = \|u - v\|$*, et l'orthogonalité par* $u \perp v \Leftrightarrow \langle u, v\rangle = 0$*. Le procédé de Gram-Schmidt et les théorèmes de projection orthogonale s'étendent à ce cadre.*

Chapitre 7 : Matrices symétriques et décomposition en valeurs singulières

7.1 Diagonalisation des matrices symétriques

Dans les chapitres précédents, nous avons étudié la diagonalisation des matrices carrées et l'orthogonalité dans $\mathbb{R}^n$. Ce chapitre combine ces deux notions pour établir un résultat fondamental : les matrices symétriques sont toujours diagonalisables, et de plus, elles le sont dans une base orthonormée. Ce résultat, appelé théorème spectral, a de nombreuses applications en analyse de données, en physique et en optimisation.

Rappels

Une matrice carrée A de taille $n \times n$ est *symétrique* si $A^T = A$. Autrement dit, les coefficients de A vérifient $a_{ij} = a_{ji}$ pour tout i, j. Une matrice symétrique est donc nécessairement carrée.

Exemples. Les matrices suivantes sont symétriques :

$$\mathbb{I}_2 = \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}, \quad A = \begin{bmatrix} 1 & 2 & 3 \\ 2 & 4 & 5 \\ 3 & 5 & 6 \end{bmatrix}, \quad B = \begin{bmatrix} \pi & e & \sqrt{2} \\ e & \pi & e \\ \sqrt{2} & e & \pi \end{bmatrix}$$

Par contre, les matrices suivantes ne sont pas symétriques :

$$\begin{bmatrix} 1 & -3 \\ 3 & 0 \end{bmatrix}, \quad \begin{bmatrix} 1 & -4 & 0 \\ -6 & 1 & -4 \\ 0 & -6 & 1 \end{bmatrix}, \quad \begin{bmatrix} 5 & 4 & 3 & 2 \\ 4 & 3 & 2 & 1 \\ 3 & 2 & 1 & 0 \end{bmatrix}$$

Une matrice A de taille $n \times n$ est *diagonalisable* s'il existe une matrice inversible P et une matrice diagonale D telles que $A = PDP^{-1}$. Les coefficients diagonaux de D sont les valeurs propres de A, et les colonnes de P sont des vecteurs propres associés, linéairement indépendants.

Nous avons vu que si A possède n valeurs propres distinctes, alors A est forcément diagonalisable. Mais qu'une matrice peut être non diagonalisable si la multiplicité algébrique d'une valeur propre est strictement plus grande que la dimension du sous-espace propre associé (multiplicité géométrique).

7.1.1 Orthogonalité des vecteurs propres

Nous allons démontrer une propriété remarquable des matrices symétriques : les vecteurs propres associés à des valeurs propres distinctes sont automatiquement orthogonaux.

Théorème 7.1 *Orthogonalité des vecteurs propres*

Soit A une matrice symétrique de taille $n \times n$. Si $\overrightarrow{v_1}$ et $\overrightarrow{v_2}$ sont deux vecteurs propres de A associés à des valeurs propres distinctes $\lambda_1 \neq \lambda_2$, alors $\overrightarrow{v_1}$ et $\overrightarrow{v_2}$ sont orthogonaux :

$$\overrightarrow{v_1} \cdot \overrightarrow{v_2} = 0$$

Démonstration. Par hypothèse, on a $A\overrightarrow{v_1} = \lambda_1 \overrightarrow{v_1}$ et $A\overrightarrow{v_2} = \lambda_2 \overrightarrow{v_2}$, avec $\lambda_1 \neq \lambda_2$.

Calculons $\lambda_1(\overrightarrow{v_1} \cdot \overrightarrow{v_2})$ de deux façons différentes.

D'une part :

$$\lambda_1(\overrightarrow{v_1} \cdot \overrightarrow{v_2}) = (\lambda_1 \overrightarrow{v_1}) \cdot \overrightarrow{v_2} = (A\overrightarrow{v_1}) \cdot \overrightarrow{v_2}$$

Or, $(A\overrightarrow{v_1}) \cdot \overrightarrow{v_2} = (A\overrightarrow{v_1})^T \overrightarrow{v_2} = \overrightarrow{v_1}^T A^T \overrightarrow{v_2}$.

Comme A est symétrique, $A^T = A$, donc :

$$\overrightarrow{v_1}^T A^T \overrightarrow{v_2} = \overrightarrow{v_1}^T A \overrightarrow{v_2} = \overrightarrow{v_1}^T (\lambda_2 \overrightarrow{v_2}) = \lambda_2(\overrightarrow{v_1}^T \overrightarrow{v_2}) = \lambda_2(\overrightarrow{v_1} \cdot \overrightarrow{v_2})$$

On obtient donc :

$$\lambda_1(\overrightarrow{v_1} \cdot \overrightarrow{v_2}) = \lambda_2(\overrightarrow{v_1} \cdot \overrightarrow{v_2})$$

Ce qui implique :

$$(\lambda_1 - \lambda_2)(\overrightarrow{v_1} \cdot \overrightarrow{v_2}) = 0$$

Comme $\lambda_1 \neq \lambda_2$, on a nécessairement $\overrightarrow{v_1} \cdot \overrightarrow{v_2} = 0$. □

Exemple. Considérons la matrice symétrique $A = \begin{bmatrix} 5 & 2 \\ 2 & 5 \end{bmatrix}$.

Polynôme caractéristique :

$$p_A(\lambda) = \det(A - \lambda \mathbb{I}_2) = \det \begin{bmatrix} 5-\lambda & 2 \\ 2 & 5-\lambda \end{bmatrix} = (5-\lambda)^2 - 4 = \lambda^2 - 10\lambda + 21$$

$$= (\lambda - 7)(\lambda - 3)$$

Valeurs propres : $\lambda_1 = 7$ et $\lambda_2 = 3$ (toutes deux de multiplicité 1).
Espaces propres :
Pour $\lambda_1 = 7$:

$$A - 7\mathbb{I}_2 = \begin{bmatrix} -2 & 2 \\ 2 & -2 \end{bmatrix} \sim \begin{bmatrix} 1 & -1 \\ 0 & 0 \end{bmatrix}$$

Le système $x - y = 0$ donne $x = y$. Donc $E_7 = \text{Vect}\left(\begin{bmatrix} 1 \\ 1 \end{bmatrix}\right)$.

Pour $\lambda_2 = 3$:

$$A - 3\mathbb{I}_2 = \begin{bmatrix} 2 & 2 \\ 2 & 2 \end{bmatrix} \sim \begin{bmatrix} 1 & 1 \\ 0 & 0 \end{bmatrix}$$

Le système $x+y=0$ donne $x=-y$. Donc $E_3 = \text{Vect}\left(\begin{bmatrix}1\\-1\end{bmatrix}\right)$.

Vérification de l'orthogonalité :

$$\begin{bmatrix}1\\1\end{bmatrix} \cdot \begin{bmatrix}1\\-1\end{bmatrix} = 1\cdot 1 + 1\cdot(-1) = 0$$

Les vecteurs propres sont bien orthogonaux, conformément au théorème 7.1.
Donc par conséquent, $E_7^{\perp} = E_3$ et $E_3^{\perp} = E_7$.

Exemple. Considérons la matrice symétrique $A = \begin{bmatrix}6 & -2 & -1\\-2 & 6 & -1\\-1 & -1 & 5\end{bmatrix}$.

On vérifie que $A^T = A$.
Polynôme caractéristique :

$$p_A(\lambda) = \det(A - \lambda\mathbb{I}_3) = \begin{vmatrix}6-\lambda & -2 & -1\\-2 & 6-\lambda & -1\\-1 & -1 & 5-\lambda\end{vmatrix}$$

Après développement (par exemple selon la troisième colonne), on trouve :

$$p_A(\lambda) = -\lambda^3 + 17\lambda^2 - 90\lambda + 144 = -(\lambda-8)(\lambda-6)(\lambda-3)$$

Valeurs propres : $\lambda_1 = 8$, $\lambda_2 = 6$ et $\lambda_3 = 3$.
Espaces propres :
Pour $\lambda_1 = 8$: on résout $(A - 8\mathbb{I}_3)\overrightarrow{v} = \overrightarrow{0}$.

$$A - 8\mathbb{I}_3 = \begin{bmatrix}-2 & -2 & -1\\-2 & -2 & -1\\-1 & -1 & -3\end{bmatrix} \sim \begin{bmatrix}1 & 1 & 0\\0 & 0 & 1\\0 & 0 & 0\end{bmatrix}$$

Donc $E_8 = \text{Vect}\left(\begin{bmatrix}-1\\1\\0\end{bmatrix}\right)$.

Pour $\lambda_2 = 6$: $E_6 = \text{Vect}\left(\begin{bmatrix}-1\\-1\\2\end{bmatrix}\right)$.

Pour $\lambda_3 = 3$: $E_3 = \text{Vect}\left(\begin{bmatrix}1\\1\\1\end{bmatrix}\right)$.

Vérification de l'orthogonalité :

$$\begin{bmatrix}-1\\1\\0\end{bmatrix} \cdot \begin{bmatrix}-1\\-1\\2\end{bmatrix} = 1 - 1 + 0 = 0$$

$$\begin{bmatrix}-1\\1\\0\end{bmatrix} \cdot \begin{bmatrix}1\\1\\1\end{bmatrix} = -1 + 1 + 0 = 0$$

$$\begin{bmatrix}-1\\-1\\2\end{bmatrix}\cdot\begin{bmatrix}1\\1\\1\end{bmatrix}=-1-1+2=0$$

Les trois vecteurs propres sont deux à deux orthogonaux.

7.1.2 Matrices orthogonales

Rappel. Une matrice carrée U de taille $n\times n$ est dite *orthogonale* si elle est inversible et si $U^{-1}=U^T$.

Remarque 7.7.0.2. *Comme par définition nous avons $U^TU=\mathbb{I}_n$, les colonnes de U sont orthonormées et forment une base orthonormée de $\mathbb{R}^n$. De plus, comme $UU^T=\mathbb{I}_n$, les colonnes de U^T (c'est-à-dire les lignes de U) sont aussi orthonormées et forment une autre base orthonormée de $\mathbb{R}^n$.*

7.1.3 Diagonalisation orthogonale

Définition 7.3 *Diagonalisation orthogonale*

Soit A une matrice carrée de taille $n\times n$. On dit que A est *diagonalisable orthogonalement* (ou *diagonalisable en base orthonormée*) s'il existe une matrice orthogonale P et une matrice diagonale D telles que :

$$A=PDP^T$$

Remarques 7.7.0.4.

1. *Si A est diagonalisable orthogonalement, alors A est diagonalisable (car $P^{-1}=P^T$). La réciproque est fausse en général.*
2. *La condition $A=PDP^T$ signifie que les colonnes de P forment une base orthonormée de $\mathbb{R}^n$ constituée de vecteurs propres de A.*

Propriété 7.5

Si A est diagonalisable orthogonalement, c'est-à-dire $A=PDP^T$ où P est orthogonale et D est diagonale, alors A est symétrique.

Démonstration. On calcule A^T :

$$A^T=(PDP^T)^T=(P^T)^TD^TP^T=PDP^T$$

car $D^T=D$ (une matrice diagonale est symétrique).

Donc $A^T=PDP^T=A$, ce qui montre que A est symétrique. □

Le résultat fondamental de ce chapitre est la réciproque : toute matrice symétrique est diagonalisable orthogonalement. C'est le théorème spectral.

Exemple. Reprenons l'exemple de $A=\begin{bmatrix}5&2\\2&5\end{bmatrix}$.

Nous avons trouvé les vecteurs propres $\overrightarrow{v_1}=\begin{bmatrix}1\\1\end{bmatrix}$ pour $\lambda_1=7$ et $\overrightarrow{v_2}=\begin{bmatrix}1\\-1\end{bmatrix}$ pour $\lambda_2=3$.

Ces vecteurs sont orthogonaux mais pas de norme 1. Normalisons-les :

$$\|\overrightarrow{v_1}\| = \sqrt{1^2+1^2} = \sqrt{2}, \quad \|\overrightarrow{v_2}\| = \sqrt{1^2+(-1)^2} = \sqrt{2}$$

Les vecteurs propres unitaires sont :

$$\overrightarrow{u_1} = \frac{1}{\sqrt{2}}\begin{bmatrix}1\\1\end{bmatrix}, \quad \overrightarrow{u_2} = \frac{1}{\sqrt{2}}\begin{bmatrix}1\\-1\end{bmatrix}$$

En posant $P = \begin{bmatrix}\frac{1}{\sqrt{2}} & \frac{1}{\sqrt{2}}\\ \frac{1}{\sqrt{2}} & -\frac{1}{\sqrt{2}}\end{bmatrix}$ et $D = \begin{bmatrix}7 & 0\\0 & 3\end{bmatrix}$, on a :

$$A = PDP^T$$

Vérifions :

$$PDP^T = \begin{bmatrix}\frac{1}{\sqrt{2}} & \frac{1}{\sqrt{2}}\\ \frac{1}{\sqrt{2}} & -\frac{1}{\sqrt{2}}\end{bmatrix}\begin{bmatrix}7 & 0\\0 & 3\end{bmatrix}\begin{bmatrix}\frac{1}{\sqrt{2}} & \frac{1}{\sqrt{2}}\\ \frac{1}{\sqrt{2}} & -\frac{1}{\sqrt{2}}\end{bmatrix}$$

$$= \begin{bmatrix}\frac{7}{\sqrt{2}} & \frac{3}{\sqrt{2}}\\ \frac{7}{\sqrt{2}} & -\frac{3}{\sqrt{2}}\end{bmatrix}\begin{bmatrix}\frac{1}{\sqrt{2}} & \frac{1}{\sqrt{2}}\\ \frac{1}{\sqrt{2}} & -\frac{1}{\sqrt{2}}\end{bmatrix}$$

$$= \begin{bmatrix}\frac{7}{2}+\frac{3}{2} & \frac{7}{2}-\frac{3}{2}\\ \frac{7}{2}-\frac{3}{2} & \frac{7}{2}+\frac{3}{2}\end{bmatrix} = \begin{bmatrix}5 & 2\\2 & 5\end{bmatrix} = A$$

7.2 Le théorème spectral

7.2.1 Énoncé et conséquences

Théorème 7.6 *Théorème spectral*

Soit A une matrice carrée de taille $n \times n$. Les assertions suivantes sont équivalentes :

1. A est symétrique.
2. A est diagonalisable en base orthonormée.

De plus, si A est symétrique, alors :

(a) A admet n valeurs propres réelles (comptées avec leur multiplicité).

(b) Pour chaque valeur propre λ, la multiplicité algébrique et la multipolicité géométrique de λ sont égales.

(c) Les sous-espaces propres sont deux à deux orthogonaux. Donc notamment, deux vecteurs propres associés à des valeurs propres distinctes sont orthogonaux.

(d) Il existe une base orthonormée de $\mathbb{R}^n$ formée de vecteurs propres de A.

Remarques 7.7.0.7. *1. L'ensemble des valeurs propres de A est appelé le* spectre *de A. C'est pourquoi ce théorème s'appelle le théorème spectral.*

2. *La propriété (a) garantit que les valeurs propres sont réelles (pas complexes), ce qui n'est pas le cas pour les matrices non symétriques en général.*
3. *La propriété (b) dit que la multiplicité géométrique égale la multiplicité algébrique pour toute valeur propre. C'est ce qui garantit la diagonalisabilité.*
4. *Si l'une des propriétés (a), (b), (c) ou (d) n'est pas vérifiée, alors A n'est pas symétrique.*

Démonstration. [Preuve partielle] Nous avons déjà démontré :
— (2) $\Rightarrow$ (1) : Si $A = PDP^T$ avec P orthogonale, alors A est symétrique (propriété démontrée précédemment).
— La propriété (c) découle du théorème 7.1.

La preuve complète de (1) $\Rightarrow$ (2) et des propriétés (a) et (b) nécessite des outils avancés d'algèbre linéaire (factorisation de Schur) et dépasse le cadre de ce cours. Nous l'admettrons. □

7.2.2 Valeurs propres réelles des matrices symétriques 2×2

Pour les matrices symétriques 2×2, nous pouvons démontrer que les valeurs propres sont toujours réelles.

Propriété 7.8

Soit $A = \begin{bmatrix} a & b \\ b & d \end{bmatrix}$ une matrice symétrique 2×2. Alors les valeurs propres de A sont réelles.

Démonstration. Le polynôme caractéristique de A est :

$$p_A(\lambda) = \det(A - \lambda \mathbb{I}_2) = (a - \lambda)(d - \lambda) - b^2 = \lambda^2 - (a + d)\lambda + (ad - b^2)$$

Le discriminant est :

$$\begin{aligned} \Delta = (a + d)^2 - 4(ad - b^2) = a^2 + 2ad + d^2 - 4ad + 4b^2 &= a^2 - 2ad + d^2 + 4b^2 \\ &= (a - d)^2 + 4b^2 \end{aligned}$$

Comme $(a - d)^2 \geq 0$ et $4b^2 \geq 0$, on a $\Delta \geq 0$.
Donc le polynôme caractéristique a toujours deux racines réelles (éventuellement égales). □

7.2.3 Exemples de diagonalisation orthogonale

Méthode 7.9 *Diagonalisation orthogonale d'une matrice symétrique*

Pour diagonaliser orthogonalement une matrice symétrique A :
1. Calculer les valeurs propres de A (racines du polynôme caractéristique).
2. Pour chaque valeur propre λ, calculer une base de l'espace propre E_λ.
3. Si un espace propre est de dimension ≥ 2, appliquer (si nécessaire) le procédé de Gram-Schmidt pour obtenir une base orthogonale de cet espace propre.
4. Normaliser tous les vecteurs propres pour obtenir des vecteurs unitaires.
5. Les colonnes de P sont les vecteurs propres unitaires. La matrice D est diagonale avec les valeurs propres correspondantes sur la diagonale.

Exemple. Diagonaliser orthogonalement $A = \begin{bmatrix} 5 & -4 & -2 \\ -4 & 5 & -2 \\ -2 & -2 & 8 \end{bmatrix}$.

Étape 1 : Valeurs propres.
On calcule $p_A(\lambda) = \det(A - \lambda \mathbb{I}_3)$:

$$A - \lambda \mathbb{I}_3 = \begin{bmatrix} 5-\lambda & -4 & -2 \\ -4 & 5-\lambda & -2 \\ -2 & -2 & 8-\lambda \end{bmatrix}$$

Après calcul, $p_A(\lambda) = -\lambda(\lambda - 9)^2$
Les valeurs propres sont $\lambda_1 = 0$ (multiplicité 1) et $\lambda_2 = 9$ (multiplicité 2).
Étape 2 : Espaces propres.
Pour $\lambda_1 = 0$:

$$A - 0 \cdot \mathbb{I}_3 = A = \begin{bmatrix} 5 & -4 & -2 \\ -4 & 5 & -2 \\ -2 & -2 & 8 \end{bmatrix}$$

Réduction :

$$\sim \begin{bmatrix} 1 & 0 & -2 \\ 0 & 1 & -2 \\ 0 & 0 & 0 \end{bmatrix}$$

Solution : $x_1 = 2x_3$, $x_2 = 2x_3$, x_3 libre.
Donc $E_0 = \text{Vect}\left(\begin{bmatrix} 2 \\ 2 \\ 1 \end{bmatrix}\right)$. Notons $\overrightarrow{v_1} = \begin{bmatrix} 2 \\ 2 \\ 1 \end{bmatrix}$.
Pour $\lambda_2 = 9$:

$$A - 9\mathbb{I}_3 = \begin{bmatrix} -4 & -4 & -2 \\ -4 & -4 & -2 \\ -2 & -2 & -1 \end{bmatrix} \sim \begin{bmatrix} 2 & 2 & 1 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \end{bmatrix}$$

Le système $2x_1 + 2x_2 + x_3 = 0$ a deux variables libres (x_2 et x_3).
Donc $E_9 = \text{Vect}\left(\begin{bmatrix} -1 \\ 1 \\ 0 \end{bmatrix}, \begin{bmatrix} -1 \\ 0 \\ 2 \end{bmatrix}\right)$. Posons $\overrightarrow{v_2} = \begin{bmatrix} -1 \\ 1 \\ 0 \end{bmatrix}$ et $\overrightarrow{v_3} = \begin{bmatrix} -1 \\ 0 \\ 2 \end{bmatrix}$.

Étape 3 : Orthogonalisation dans E_9.
Vérifions d'abord si $\overrightarrow{v_2}$ et $\overrightarrow{v_3}$ sont orthogonaux :

$$\overrightarrow{v_2} \cdot \overrightarrow{v_3} = 1 \neq 0$$

Ils ne sont pas orthogonaux. Appliquons le procédé de Gram-Schmidt.
Calculons la projection :

$$\text{proj}_{\overrightarrow{v_2}}(\overrightarrow{v_3}) = \frac{\overrightarrow{v_3} \cdot \overrightarrow{v_2}}{\overrightarrow{v_2} \cdot \overrightarrow{v_2}} \overrightarrow{v_2} = \frac{1}{2} \begin{bmatrix} -1 \\ 1 \\ 0 \end{bmatrix} = \begin{bmatrix} -1/2 \\ 1/2 \\ 0 \end{bmatrix}$$

Alors :

$$\overrightarrow{w_3} = \overrightarrow{v_3} - \text{proj}_{\overrightarrow{v_2}}(\overrightarrow{v_3}) = \begin{bmatrix} -1 \\ 0 \\ 2 \end{bmatrix} - \begin{bmatrix} -1/2 \\ 1/2 \\ 0 \end{bmatrix} = \begin{bmatrix} -1/2 \\ -1/2 \\ 2 \end{bmatrix}.$$

Étape 4 : Normalisation.

$$\|\overrightarrow{v_1}\| = \sqrt{4+4+1} = 3, \quad \overrightarrow{u_1} = \frac{1}{3}\begin{bmatrix} 2 \\ 2 \\ 1 \end{bmatrix}$$

$$\|\overrightarrow{v_2}\| = \sqrt{1+1+0} = \sqrt{2}, \quad \overrightarrow{u_2} = \frac{1}{\sqrt{2}}\begin{bmatrix} -1 \\ 1 \\ 0 \end{bmatrix}$$

$$\|\overrightarrow{w_3}\| = \sqrt{\frac{1}{4}+\frac{1}{4}+4} = \frac{3}{\sqrt{2}}, \quad \overrightarrow{u_3} = \frac{1}{3\sqrt{2}}\begin{bmatrix} -1 \\ -1 \\ 4 \end{bmatrix}$$

Résultat :

$$P = \begin{bmatrix} \frac{2}{3} & -\frac{1}{\sqrt{2}} & -\frac{1}{3\sqrt{2}} \\ \frac{2}{3} & \frac{1}{\sqrt{2}} & -\frac{1}{3\sqrt{2}} \\ \frac{1}{3} & 0 & \frac{4}{3\sqrt{2}} \end{bmatrix}, \quad D = \begin{bmatrix} 0 & 0 & 0 \\ 0 & 9 & 0 \\ 0 & 0 & 9 \end{bmatrix}$$

Et on a $A = PDP^T$.

7.3 Décomposition spectrale

Le théorème spectral nous permet d'écrire une matrice symétrique sous une forme particulièrement élégante appelée *décomposition spectrale*.

Définition 7.10 *Décomposition spectrale*

Soit A une matrice symétrique de taille $n \times n$, diagonalisable orthogonalement avec $A = PDP^T$, où $P = \begin{bmatrix} \overrightarrow{u_1} & \overrightarrow{u_2} & \cdots & \overrightarrow{u_n} \end{bmatrix}$ est une matrice orthogonale dont les colonnes forment une base orthonormée de vecteurs propres, et $D = \begin{bmatrix} \lambda_1 & 0 & \cdots & 0 \\ 0 & \lambda_2 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & \lambda_n \end{bmatrix}$ est la matrice diagonale des valeurs propres associées.

La *décomposition spectrale* de A est l'écriture :

$$A = \lambda_1 \overrightarrow{u_1}\overrightarrow{u_1}^T + \lambda_2 \overrightarrow{u_2}\overrightarrow{u_2}^T + \cdots + \lambda_n \overrightarrow{u_n}\overrightarrow{u_n}^T = \sum_{i=1}^{n} \lambda_i \overrightarrow{u_i}\overrightarrow{u_i}^T$$