Chapitre VI

Méthode des moindres carrés

1. Introduction
2. Equations normales
3. Exemple d'application

1. Introduction

Exemple introductif

- On s'intéresse à une grandeur $\mathbf{b}$ pouvant prendre les valeurs $b_1, b_2, \dots, b_m$; on notera :

$$\mathbf{b} = (b_1, b_2, \dots, b_m)^T$$

- Les valeurs b_i dépendent a priori d'un paramètre x_i. Cela signifie que chaque valeur b_i de la grandeur $\mathbf{b}$ est reliée à une valeur x_i par une fonction f a priori inconnue* ; on notera :

$$\forall i \in [\![1, m]\!], \qquad b_i = f(x_i)$$

- Supposons que f soit un polynôme de degré $n-1$, alors :

$$\begin{cases} b_1 = a_0 + a_1 x_1 + a_2 x_1^2 + \cdots + a_{n-1} x_1^{n-1} \\ b_2 = a_0 + a_1 x_2 + a_2 x_2^2 + \cdots + a_{n-1} x_2^{n-1} \\ \vdots \\ b_m = a_0 + a_1 x_m + a_2 x_m^2 + \cdots + a_{n-1} x_m^{n-1} \end{cases}$$

- On a alors un système de m équations à n inconnues qui sont les coefficients du polynôme, c'est-à-dire $a_0, a_1, a_2, \dots, a_{n-1}$. Le plus souvent le nombre de mesures (m) est plus grand que le nombre d'inconnues (n).

** on peut imaginer que b_i est une température mesurée pour une pression donnée x_i (ou alors une tension mesurée en fonction de l'intensité du courant, ou encore l'intensité d'une source lumineuse en fonction de la longueur d'onde)*

1. Introduction

- Le système précédent s'écrit sous forme matricielle :

$$\mathbf{Ax} = \mathbf{b} \quad \text{avec} \quad \mathbf{A} = \begin{pmatrix} 1 & x_1 & x_1^2 & \dots & x_1^{n-1} \\ 1 & x_2 & x_2^2 & \dots & x_2^{n-1} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 1 & x_m & x_m^2 & \dots & x_m^{n-1} \end{pmatrix} \in \mathbb{R}^{m \times n}$$

$$\mathbf{x} = \begin{pmatrix} a_0 \\ a_1 \\ a_2 \\ \vdots \\ a_{n-1} \end{pmatrix} \in \mathbb{R}^{\boldsymbol{n}} \quad \text{et} \quad \mathbf{b} = \begin{pmatrix} b_1 \\ b_2 \\ \vdots \\ b_m \end{pmatrix} \in \mathbb{R}^{\boldsymbol{m}}$$

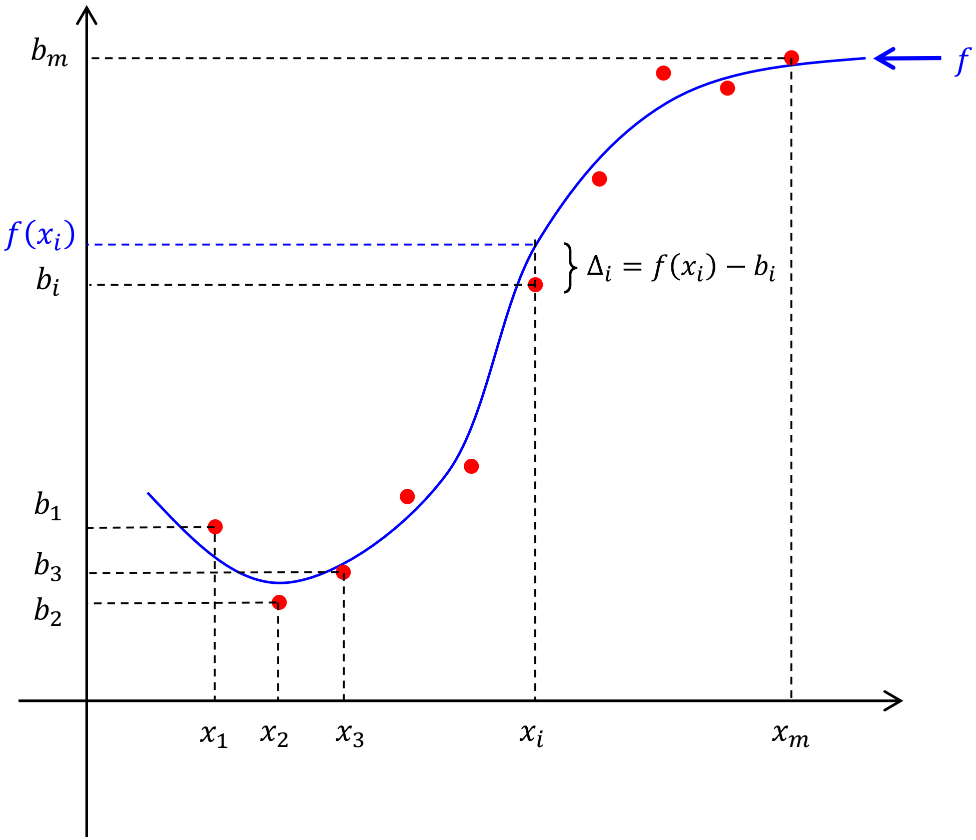
Attention, les notations sont trompeuses : $\mathbf{x} = (a_0, a_1, \dots, a_{n-1})^T$ *et les* x_i *apparaissent dans la matrice* $\mathbf{A}$

- Le système ci-dessus étant surdimensionné ($m \geq n$), il n'admet pas de solution. Cependant il doit bien exister un vecteur $\mathbf{x}^*$ qui minimise la distance entre les vecteurs $\mathbf{Ax}$ et $\mathbf{b}$, c'est-à-dire un vecteur $\mathbf{x}^*$ tel que $\|\mathbf{Ax}^* - \mathbf{b}\|$ soit le plus petit possible. On note :

$$\mathbf{x}^* = \underset{\mathbf{x} \in \mathbb{R}^n}{\operatorname{argmin}} \|\mathbf{Ax} - \mathbf{b}\|$$

1. Introduction

- Schématiquement :



$$\mathbf{Ax} - \mathbf{b} = \begin{pmatrix} a_0 + a_1 x_1 + \cdots a_{n-1} x_1^{n-1} - b_1 \\ a_0 + a_1 x_2 + \cdots a_{n-1} x_2^{n-1} - b_2 \\ \vdots \\ a_0 + a_1 x_m + \cdots a_{n-1} x_m^{n-1} - b_m \end{pmatrix}$$

$$= \begin{pmatrix} f(x_1) - b_1 \\ f(x_2) - b_2 \\ \vdots \\ f(x_m) - b_m \end{pmatrix}$$

$$\Rightarrow \|\mathbf{Ax} - \mathbf{b}\|^2 = \sum_{i=1}^{m} [f(x_i) - b_i]^2$$

Minimiser $\|\mathbf{Ax} - \mathbf{b}\|$ *revient donc à minimiser la somme des carrés des écarts* $\Delta_i = f(x_i) - b_i$

1. Introduction

C'est le **problème des moindres carrés** qui de façon générale s'énonce de la façon suivante :

On considère le système d'équations linéaires $\mathbf{Ax} = \mathbf{b}$ avec $\mathbf{A} \in \mathbb{R}^{m \times n}, \mathbf{x} \in \mathbb{R}^n, \mathbf{b} \in \mathbb{R}^m$ et $m \geq n$ (plus d'équations que d'inconnues). Ce système n'ayant pas de solution on a forcément $\mathbf{Ax} - \mathbf{b} \neq \mathbf{0}$. On cherche alors un vecteur $\mathbf{x}^* \in \mathbb{R}^n$ qui minimise la distance $\|\mathbf{Ax} - \mathbf{b}\|$ entre les vecteur $\mathbf{Ax}$ et $\mathbf{b}$.

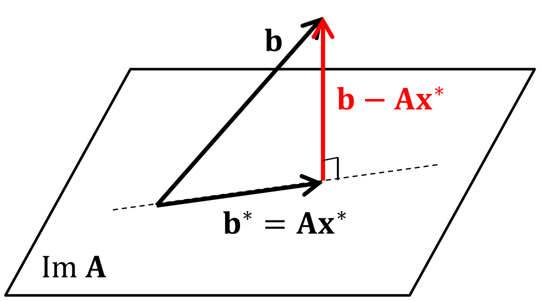
Le vecteur $\mathbf{x}^*$ est appelé **solution du système $\mathbf{Ax} = \mathbf{b}$ au sens des moindres carrés** et il est noté :

$$\mathbf{x}^* = \underset{\mathbf{x} \in \mathbb{R}^n}{\operatorname{argmin}} \|\mathbf{Ax} - \mathbf{b}\|^2$$

Interprétation géométrique

On note $\mathbf{b}^* = \mathbf{Ax}^*$ alors :

- $\mathbf{b}^* \in \operatorname{Im} \mathbf{A}$ (évident)
- $\mathbf{b}^*$ est le vecteur qui minimise la distance entre $\mathbf{b}$ et $\operatorname{Im} \mathbf{A}$ (qui est un sous-espace vectoriel de$\mathbb{R}^n$)



Intuitivement, $\mathbf{b}^*$ est la projection orthogonale de $\mathbf{b}$ sur $\operatorname{Im} \mathbf{A}$.

2. Equations normales

Soit $\mathbf{x} \in \mathbb{R}^n$ alors $\langle \mathbf{Ax}, \mathbf{b} - \mathbf{b}^* \rangle = 0$ car $\mathbf{Ax} \in \mathrm{Im}\,\mathbf{A}$ et $\mathbf{b} - \mathbf{b}^* \perp \mathrm{Im}\,\mathbf{A}$ (voir figure précédente).

La solution au sens des moindres carrés, notée $\mathbf{x}^*$, du problème $\mathbf{Ax} = \mathbf{b}$ est donc la solution du système d'équations $\mathbf{A}^T\mathbf{Ax}^* = \mathbf{A}^T\mathbf{b}$ où $\mathbf{A}^T\mathbf{A}$ est une matrice carrée de dimension $n \times n$ (encore faut-il que cette matrice soit inversible !). On montre (*voir démonstration page suivante*) que si $\mathrm{rg}\,\mathbf{A} = n$ alors la matrice $\mathbf{A}^T\mathbf{A}$ est inversible.

2. Equations normales

Démonstration

Soit $\mathbf{z} \in \mathrm{Ker}\,(\mathbf{A}^T\mathbf{A})$ *alors* $\mathbf{A}^T\mathbf{A}\mathbf{z} = 0$ *donc* $\mathbf{A}\mathbf{z} \in \mathrm{Ker}\,\mathbf{A}^T$.

Or $\mathrm{Ker}\,\mathbf{A}^T = (\mathrm{Im}\,\mathbf{A})^{\perp}$ *(admis) donc* $\mathbf{A}\mathbf{z} \in (\mathrm{Im}\,\mathbf{A})^{\perp}$.

Mais par définition $\mathbf{A}\mathbf{z} \in \mathrm{Im}\,\mathbf{A}$ *donc* $\mathbf{A}\mathbf{z} \in \mathrm{Im}\,\mathbf{A} \cap (\mathrm{Im}\,\mathbf{A})^{\perp} = \{\mathbf{0}\}$ *donc* $\mathbf{A}\mathbf{z} = \mathbf{0}$ *donc* $\mathbf{z} \in \mathrm{Ker}\,\mathbf{A}$.

Comme par hypothèse $\mathrm{rg}\,\mathbf{A} = n$ *on a* $\dim \mathrm{Ker}\,\mathbf{A} = 0$, *c'est-à-dire* $\mathrm{Ker}\,\mathbf{A} = \{\mathbf{0}\}$ *donc* $\mathbf{A}^T\mathbf{A}$ *est inversible.*

2. Equations normales

A retenir

Soient $\mathbf{A} \in \mathbb{R}^{m \times n}$ de rang n avec $m \geq n$ et $\mathbf{b} \in \mathbb{R}^{m}$. La solution au sens des moindres carrés du système $\mathbf{Ax} = \mathbf{b}$, où $\mathbf{x} \in \mathbb{R}^{n}$, est l'<u>unique</u> solution des équations $\mathbf{A}^T\mathbf{Ax} = \mathbf{A}^T\mathbf{b}$, c'est-à-dire:

$$\mathbf{x}^* \stackrel{\text{def}}{=} \underset{\mathbf{x} \in \mathbb{R}^n}{\operatorname{argmin}} \|\mathbf{Ax} - \mathbf{b}\|^2 = (\mathbf{A}^T\mathbf{A})^{-1}\mathbf{A}^T\mathbf{b}$$

Les équations $\mathbf{A}^T\mathbf{Ax}^* = \mathbf{A}^T\mathbf{b}$ sont appelées les **équations normales**.

La matrice $\mathbf{A}^\dagger = (\mathbf{A}^T\mathbf{A})^{-1}\mathbf{A}^T$ est appelée **matrice pseudo-inverse** de $\mathbf{A}$.

On parle de matrice pseudo-inverse car la matrice $\mathbf{A}$ n'étant pas carrée, elle ne peut être inversible.

On remarquera cependant que :

- $\mathbf{A}^\dagger\mathbf{A} = (\mathbf{A}^T\mathbf{A})^{-1}\mathbf{A}^T\mathbf{A} = \mathbf{M}^{-\mathbf{1}}\mathbf{M} = \mathbf{I}$
- $\mathbf{AA}^\dagger = \mathbf{A}(\mathbf{A}^T\mathbf{A})^{-1}\mathbf{A}^T \neq \mathbf{I}$

2. Equations normales

Remarque

Si $\mathbf{A} \in \mathbb{R}^{m\times n}$ *avec* $m \geq n$ *et* $\operatorname{rg} \mathbf{A} = r < n$ *alors la solution n'est pas unique. En effet :*

Soit $\mathbf{x}^*$ *une solution au sens des moindres carrés de* $\mathbf{Ax} = \mathbf{b}$ *et soit* $\mathbf{z} \in \operatorname{Ker} \mathbf{A}$ *:*

$$\mathbf{A}(\mathbf{x}^* + \mathbf{z}) - \mathbf{b} = \mathbf{Ax}^* + \mathbf{Az} - \mathbf{b} = \mathbf{Ax}^* - \mathbf{b}$$

Donc si $\mathbf{x}^*$ *minimise* $\|\mathbf{Ax} - \mathbf{b}\|$ *alors il en est de même pour* $\mathbf{x}^* + \mathbf{z}$*.*

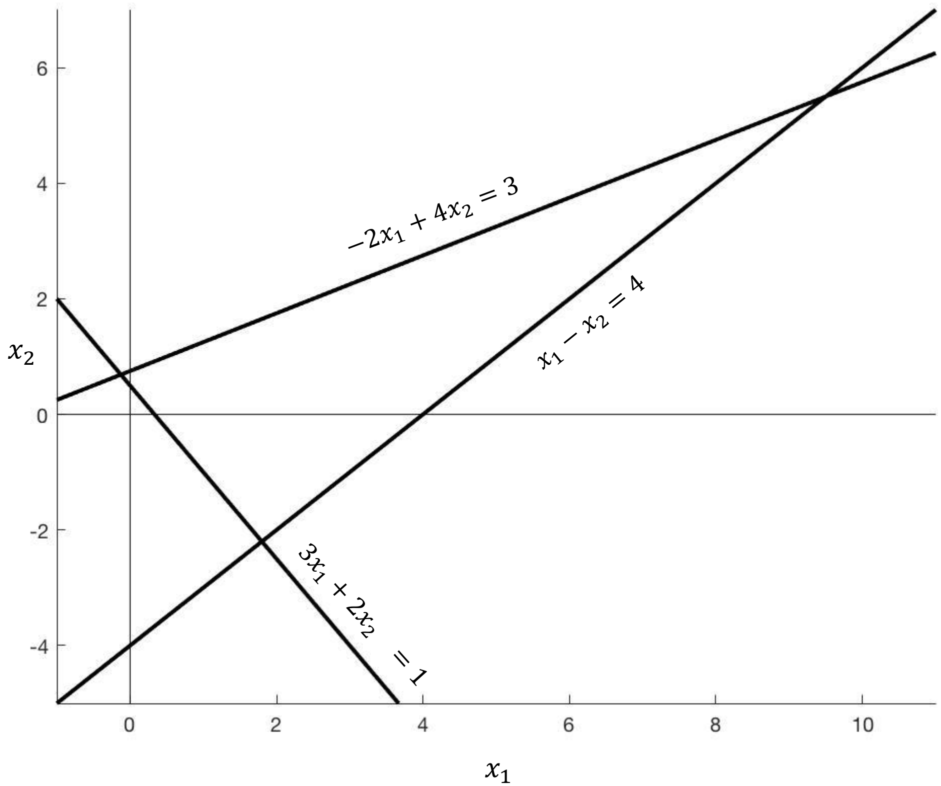
On retiendra que si $\operatorname{rg} \mathbf{A} = r < n$ *et si* $\mathbf{x}^*$ *est une solution au sens des moindres carrés de* $\mathbf{Ax} = \mathbf{b}$ *alors il existe une infinité de solutions ; elles sont de la forme* $\mathbf{x}^* + \mathbf{z}$ *où* $\mathbf{z} \in \operatorname{Ker} \mathbf{A}$*.*

2. Equations normales

Exercice

Calculer la solution au sens des moindres carrés du système défini par :

$$\begin{cases} x_1 - x_2 = 4 \\ 3x_1 + 2x_2 = 1 \\ -2x_1 + 4x_2 = 3 \end{cases}$$



2. Equations normales

Solution

Ici $m = 3$, $n = 2$, $\mathbf{A} = \begin{pmatrix} 1 & -1 \\ 3 & 2 \\ -2 & 4 \end{pmatrix}$, $\mathbf{b} = \begin{pmatrix} 4 \\ 1 \\ 3 \end{pmatrix}$ et $\mathbf{x} = \begin{pmatrix} x_1 \\ x_2 \end{pmatrix}$

La matrice $\mathbf{A}$ est de rang 2 (évident car les deux vecteurs colonnes sont libres), le problème $\mathbf{Ax} = \mathbf{b}$ admet donc une solution unique (au sens des moindres carrés) $\mathbf{x}^*$ qui vérifie les équations normales :

$$\mathbf{A}^T\mathbf{A}\mathbf{x} = \mathbf{A}^T\mathbf{b}$$

avec $\mathbf{A}^T\mathbf{A} = \begin{pmatrix} 14 & -3 \\ -3 & 21 \end{pmatrix}$ et $\mathbf{A}^T\mathbf{b} = \begin{pmatrix} 1 \\ 10 \end{pmatrix}$.

Il suffit donc d'inverser la matrice $\mathbf{A}^T\mathbf{A}$:

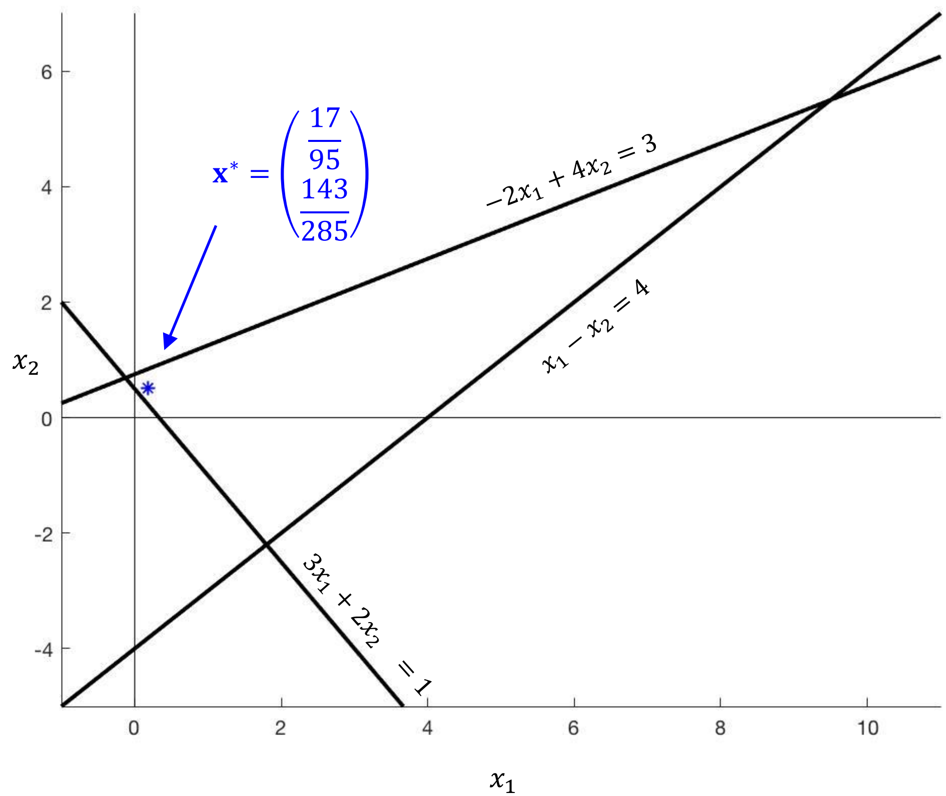
$$(\mathbf{A}^T\mathbf{A})^{-\mathbf{1}} = \frac{1}{14 \times 21 - 9}\begin{pmatrix} 21 & 3 \\ 3 & 14 \end{pmatrix} = \begin{pmatrix} \frac{7}{95} & \frac{1}{95} \\ \frac{1}{95} & \frac{14}{285} \end{pmatrix}$$

La solution est donc :

$$\mathbf{x}^* = (\mathbf{A}^T\mathbf{A})^{-1}\mathbf{A}^T\mathbf{b} = \begin{pmatrix} \frac{7}{95} & \frac{1}{95} \\ \frac{1}{95} & \frac{14}{285} \end{pmatrix}\begin{pmatrix} 1 \\ 10 \end{pmatrix} = \begin{pmatrix} \frac{17}{95} \\ \frac{143}{285} \end{pmatrix}$$

2. Equations normales

Graphiquement :

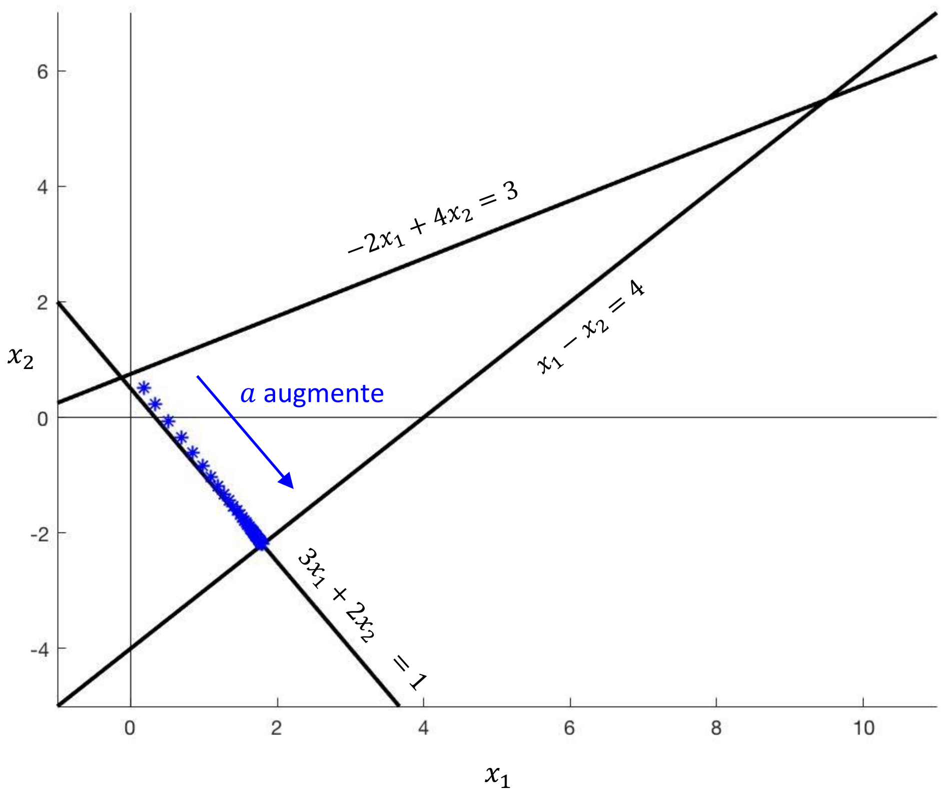


2. Equations normales

Si on divise la dernière ligne par un nombre positif a que l'on fait varier de 1 à 100 :

$$\begin{cases} x_1 - x_2 & = 4 \\ 3x_1 + 2x_2 & = 1 \\ -\dfrac{2}{a}x_1 + \dfrac{4}{a}x_2 & = \dfrac{3}{a} \end{cases}$$

La résolution au sens des moindres carrés des systèmes $\mathbf{A}_a\mathbf{x} = \mathbf{b}$ avec $\mathbf{A}_a = \begin{pmatrix} 1 & -1 \\ 3 & 2 \\ -2/a & 4/a \end{pmatrix}$ et $\mathbf{b} = \begin{pmatrix} 4 \\ 1 \\ 3/a \end{pmatrix}$, donne :



Interprétation ?

2. Equations normales

Autre méthode pour démontrer la formule des équations normales

On définit la fonction suivante :

$$E(\mathbf{x}) = \|\mathbf{Ax} - \mathbf{b}\|^2 \qquad \text{avec } \mathbf{x} = (x_1, x_2, \ldots, x_n)^T$$

$E(\mathbf{x})$ est l'écart (au carré) entre $\mathbf{Ax}$ et $\mathbf{b}$, c'est une fonction réelle des variables réelles $x_1, x_2, \ldots, x_n$.

Résoudre le problème $\mathbf{Ax} = \mathbf{b}$ au sens des moindres carrés revient à déterminer le vecteur $\mathbf{x}^* = (x_1^*, x_2^*, \ldots, x_n^*)^T$ qui minimise $E(\mathbf{x})$, c'est-à-dire tel que :

$$\overrightarrow{\text{grad}}\, E(\mathbf{x}) = \mathbf{0}_{\mathbb{R}^n}$$

2. Equations normales

Or : $E(\mathbf{x}) = \|\mathbf{A}\mathbf{x} - \mathbf{b}\|^2$

Donc (cf. chapitre II) :

Finalement :

2. Equations normales

Factorisation QR appliquée aux problèmes des moindres carrés

Reprenons les équations normales dans lesquelles on remplace la matrice $\mathbf{A}$ par sa décomposition QR :

$$\mathbf{A}^T\mathbf{A}\mathbf{x} = \mathbf{A}^T\mathbf{b}$$

A retenir

Soient $\mathbf{A} \in \mathbb{R}^{m \times n}$ de rang n avec $m \geq n$ et $\mathbf{b} \in \mathbb{R}^m$. Soit $\mathbf{x}^*$ la solution au sens des moindres carrés du système $\mathbf{A}\mathbf{x} = \mathbf{b}$, où $\mathbf{x} \in \mathbb{R}^n$. Alors :

$$\mathbf{x}^* = \mathbf{R}^{-1}\mathbf{Q}^T\mathbf{b}$$

où $\mathbf{Q}$ et $\mathbf{R}$ sont les matrices de la factorisation QR de $\mathbf{A}$.

3. Exemple

Déterminer la solution au sens des moindres carrés du système $\mathbf{Ax} = \mathbf{b}$ où $\mathbf{A} = \begin{pmatrix} 1 & -2 & -1 \\ 2 & 0 & 1 \\ 2 & -4 & 2 \\ 4 & 0 & 0 \end{pmatrix}$ et $\mathbf{b} = \begin{pmatrix} -1 \\ 1 \\ 1 \\ -2 \end{pmatrix}$.

La décomposition QR de $\mathbf{A}$ donne :

$$\mathbf{A} = \mathbf{QR} \quad \text{avec} \quad \mathbf{Q} = \frac{1}{5}\begin{pmatrix} 1 & -2 & -4 \\ 2 & 1 & 2 \\ 2 & -4 & 2 \\ 4 & 2 & -1 \end{pmatrix} \quad \text{et} \quad \mathbf{R} = \begin{pmatrix} 5 & -2 & 1 \\ 0 & 4 & -1 \\ 0 & 0 & 2 \end{pmatrix}$$

On a vu que $\mathbf{x}^* = \mathbf{R^{-1}Q}^T\mathbf{b}$ mais plutôt que d'inverser la matrice $\mathbf{R}$ on résout le système triangulaire $\mathbf{Rx}^* = \mathbf{Q}^T\mathbf{b}$

avec $\mathbf{Q}^T\mathbf{b} = \frac{1}{5}\begin{pmatrix} 1 & -2 & -4 \\ 2 & 1 & 2 \\ 2 & -4 & 2 \\ 4 & 2 & -1 \end{pmatrix}\begin{pmatrix} -1 \\ 1 \\ 1 \\ -2 \end{pmatrix} = \begin{pmatrix} -1 \\ -1 \\ 2 \end{pmatrix}$

On a alors : $\mathbf{Rx}^* = \mathbf{Q}^T\mathbf{b} \quad \Leftrightarrow \quad \begin{pmatrix} 5 & -2 & 1 \\ 0 & 4 & -1 \\ 0 & 0 & 2 \end{pmatrix}\begin{pmatrix} x_1 \\ x_2 \\ x_3 \end{pmatrix} = \begin{pmatrix} -1 \\ -1 \\ 2 \end{pmatrix} \quad \Leftrightarrow \quad \begin{cases} 5x_1 - 2x_2 + x_3 = -1 \\ 4x_2 - x_3 = -1 \\ 2x_3 = 2 \end{cases}$

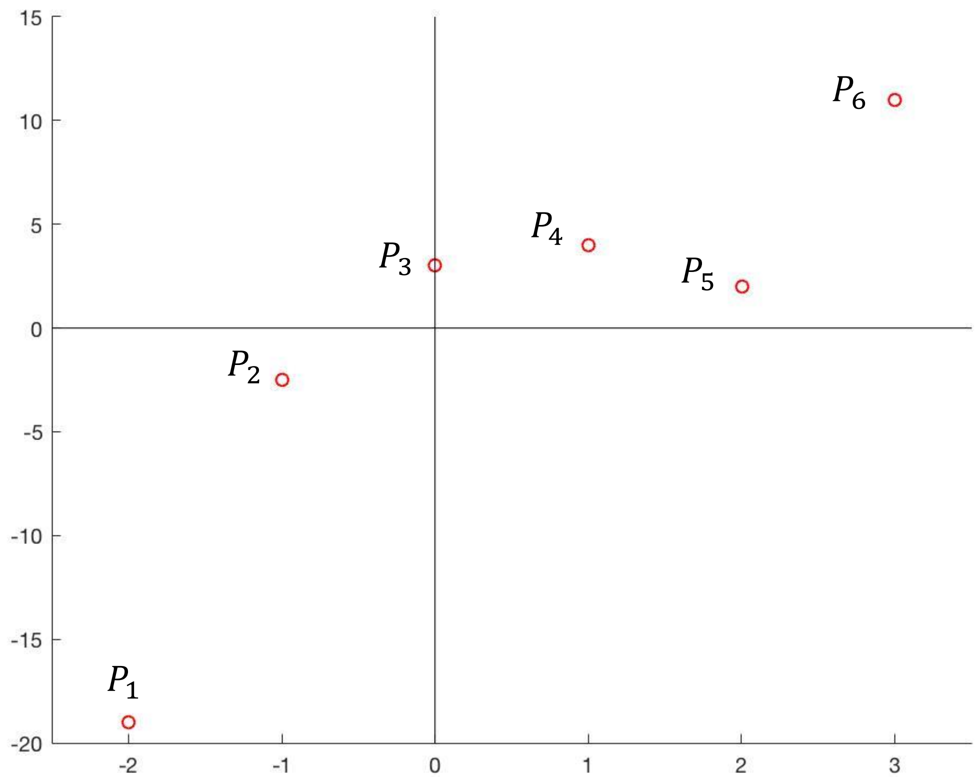
On a donc un système triangulaire qui se résout facilement, on trouve : $\mathbf{x}^* = \begin{pmatrix} -\frac{2}{5} & 0 & 1 \end{pmatrix}^T$

3. Exemple

Exercice

Déterminer l'équation de la droite passant au plus près des points suivants :

$$P_1 = \begin{pmatrix} -2 \\ -19 \end{pmatrix}, \quad P_2 = \begin{pmatrix} -1 \\ -\frac{5}{2} \end{pmatrix}, \quad P_3 = \begin{pmatrix} 0 \\ 3 \end{pmatrix}, \quad P_4 = \begin{pmatrix} 1 \\ 4 \end{pmatrix}, \quad P_5 = \begin{pmatrix} 2 \\ 2 \end{pmatrix}, \quad P_6 = \begin{pmatrix} 3 \\ 11 \end{pmatrix}$$



3. Exemple

Soit $y = a_0 + a_1 x$ l'équation de la droite recherchée alors :

$$\begin{cases} -2a_1 + a_0 = -19 \\ -1a_1 + a_0 = -5/2 \\ 0a_1 + a_0 = 3 \\ 1a_1 + a_0 = 4 \\ 2a_1 + a_0 = 2 \\ 3a_1 + a_0 = 11 \end{cases} \quad \Leftrightarrow \quad \begin{pmatrix} 1 & -2 \\ 1 & -1 \\ 1 & 0 \\ 1 & 1 \\ 1 & 2 \\ 1 & 3 \end{pmatrix} \begin{pmatrix} a_0 \\ a_1 \end{pmatrix} = \begin{pmatrix} -19 \\ -5/2 \\ 3 \\ 4 \\ 2 \\ 11 \end{pmatrix}$$

$$\Leftrightarrow \quad \mathbf{Ax} = \mathbf{b} \quad \text{avec} \quad \mathbf{A} = \begin{pmatrix} 1 & -2 \\ 1 & -1 \\ 1 & 0 \\ 1 & 1 \\ 1 & 2 \\ 1 & 3 \end{pmatrix}, \qquad \mathbf{x} = \begin{pmatrix} a_0 \\ a_1 \end{pmatrix} \quad \text{et} \quad \mathbf{b} = \begin{pmatrix} -19 \\ -5/2 \\ 3 \\ 4 \\ 2 \\ 11 \end{pmatrix}$$

3. Exemple

Les équations normales s'écrivent $\mathbf{A}^T\mathbf{A}\mathbf{x} = \mathbf{A}^T\mathbf{b}$ avec :

$$\mathbf{A}^T\mathbf{A} = \begin{pmatrix} 6 & 3 \\ 3 & 19 \end{pmatrix} \quad \text{et} \quad \mathbf{A}^T\mathbf{b} = \begin{pmatrix} -3/2 \\ 163/2 \end{pmatrix}$$

On inverse $\mathbf{A}^T\mathbf{A}$:

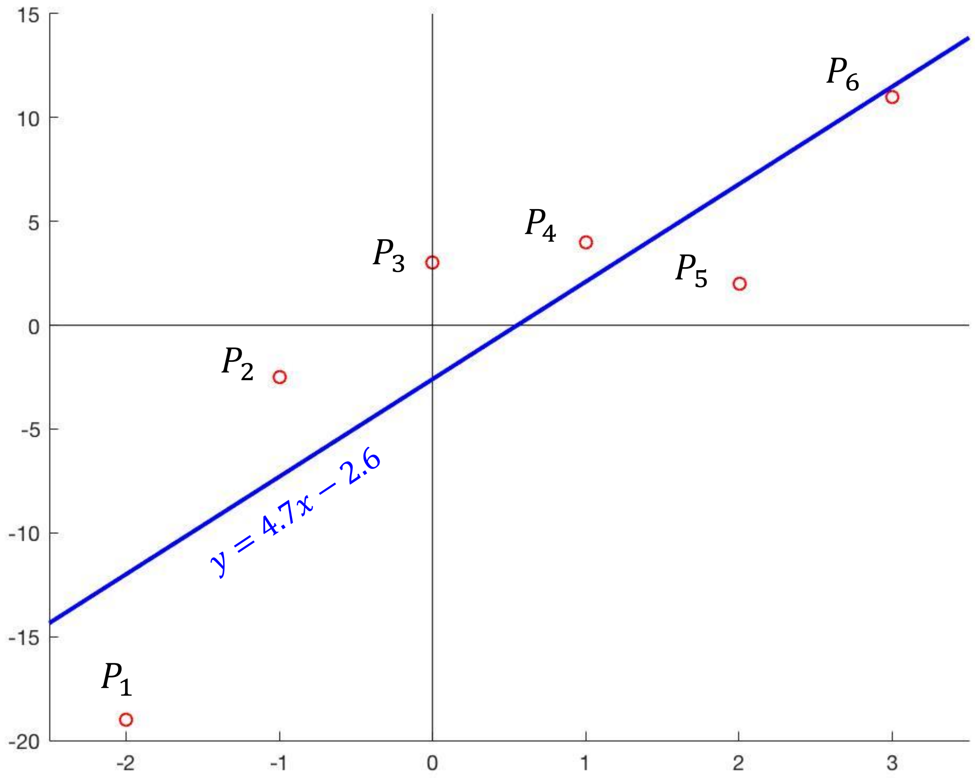
$$(\mathbf{A}^T\mathbf{A})^{-1} = \frac{1}{105}\begin{pmatrix} 19 & -3 \\ -3 & 6 \end{pmatrix}$$

La solution au sens des moindre carrés est alors :

$$\mathbf{x}^* = ({a_0}^* \quad {a_1}^*)^T = (\mathbf{A}^T\mathbf{A})^{-1}\mathbf{A}^T\mathbf{b} = \frac{1}{105}\begin{pmatrix} 19 & -3 \\ -3 & 6 \end{pmatrix}\begin{pmatrix} -\dfrac{3}{2} \\ \dfrac{163}{2} \end{pmatrix} = \begin{pmatrix} -\dfrac{91}{35} = -2{,}6 \\ \dfrac{329}{70} = 4{,}7 \end{pmatrix}$$

3. Exemple

L'équation de la droite passant au plus près des points est donc : $y = 4.7x - 2.6$



Question : quelle est l'équation du polynôme du second degré passant au plus près des 6 points ? (à faire en exercice)

Réponse : $y = 0.614 + 5.905x - 1.205x^2$