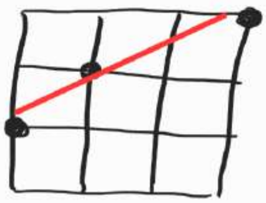
Fiche 5 Moindres carrés, valeurs propres.

Exo 1

Trouver la droite qui passe au plus près (au sens des moindres carrés) des points $\begin{pmatrix}0\\1\end{pmatrix}, \begin{pmatrix}1\\2\end{pmatrix}, \begin{pmatrix}3\\3\end{pmatrix}$

"meilleure droite"



droite $y = \alpha x + \beta$

pas de solution exacte (points non alignés).

$$\begin{cases} 0 \times \alpha + \beta = 1 \\ \alpha + \beta = 2 \\ 3\alpha + \beta = 3 \end{cases} \Leftrightarrow \begin{pmatrix} 0 & 1 \\ 1 & 1 \\ 3 & 1 \end{pmatrix} \begin{pmatrix} \alpha \\ \beta \end{pmatrix} = \begin{pmatrix} 1 \\ 2 \\ 3 \end{pmatrix}$$

Ce système peut se réécrire sous la forme

$$A\vec{x} = \vec{b} \qquad A = \begin{pmatrix} 0 & 1 \\ 1 & 1 \\ 3 & 1 \end{pmatrix}$$

$$\vec{b} = \begin{pmatrix} 1 \\ 2 \\ 3 \end{pmatrix} \quad \vec{x} = \begin{pmatrix} \alpha \\ \beta \end{pmatrix}$$

On résoud ce système au sens des moindres carrés, on cherche le (ou les) vecteur(s) $\vec{x}$ qui minimisent

$$E = \| A\vec{x} - \vec{b} \|_2^2 = (\beta - 1)^2 + (\alpha + \beta - 2)^2 + (3\alpha + \beta - 3)^2$$

E est l'erreur. Si $E = 0$ alors on a une solution exacte (les points sont alignés) mais ce n'est pas le cas ici.

Solution (cf cours)

On résoud l'équation normale

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

- $A^TA = \begin{pmatrix} 0 & 1 & 3 \\ 1 & 1 & 1 \end{pmatrix} \begin{pmatrix} 0 & 1 \\ 1 & 1 \\ 3 & 1 \end{pmatrix} = \begin{pmatrix} 10 & 4 \\ 4 & 3 \end{pmatrix}$

- $A^T\vec{b} = \begin{pmatrix} 0 & 1 & 3 \\ 1 & 1 & 1 \end{pmatrix} \begin{pmatrix} 1 \\ 2 \\ 3 \end{pmatrix} = \begin{pmatrix} 11 \\ 6 \end{pmatrix}$

d'où l'équation $\begin{pmatrix} 10 & 4 \\ 4 & 3 \end{pmatrix} \begin{pmatrix} \alpha \\ \beta \end{pmatrix} = \begin{pmatrix} 11 \\ 6 \end{pmatrix}$

$$\begin{cases} 10\alpha + 4\beta = 11 & (1) \\ 4\alpha + 3\beta = 6 & (2) \end{cases}$$

(ou bien l'inverse des matrices 2x2 $\frac{1}{\det A} {}^t\mathrm{Com}\, A$.)

(1) $3\times(1) - 4\times(2)$: $30\alpha - 16\alpha = 33 - 24$

$14\alpha = 9 \Rightarrow \alpha = \frac{9}{14}$

$\beta = 8/7$

donc $\vec{x} = \begin{pmatrix} 9/14 \\ 8/7 \end{pmatrix}$

donc $y = \frac{9}{14}x + \frac{8}{7}$ est la droite recherchée.

La solution est unique ici.

Vu en cours :

- L'ensemble des solutions est $\{A^+\vec{b} + \vec{x}', \vec{x}' \in \ker A\}$

où A^+ est l'inverse généralisé (ou pseudo-inverse)

Dans notre exemple, on vérifie facilement $\ker A = \{0\}$

compatible avec l'unicité que l'on a trouvé.

- S'il y a plusieurs solutions, celle de norme minimale est $A^+\vec{b}$.

Pour trouver A^+ en pratique, écrire la svd de A^*

$$A = V D U^* \quad , \quad A^* = U D^T V^* ,$$

inverser les éléments non nuls dans D^T pour obtenir D',

et $\quad A^+ = U D' V^*$

par ex si $\quad D = \begin{pmatrix} \mu_1 & 0 & 0 \\ 0 & \mu_2 & 0 \end{pmatrix} \quad , \quad D^T = \begin{pmatrix} \mu_1 & 0 \\ 0 & \mu_2 \\ 0 & 0 \end{pmatrix}$

$$D' = \begin{pmatrix} 1/\mu_1 & 0 \\ 0 & 1/\mu_2 \\ 0 & 0 \end{pmatrix}$$

Rq (Philosophique)

Pourquoi minimise-t-on les moindres carrés ? (basé sur la $\|\cdot\|_2$)

parce que

- solution simple (équation normale, inverse généralisé) qui ne marche pas pour d'autre normes.
- Lien avec l'estimateur du maximum de vraissemblance en stat (dans le cas gaussien).

Exo 2 Approximation polynomiale.

$\mathbb{R}_d[X]$ = ens des polynômes de degré $\leq d$.

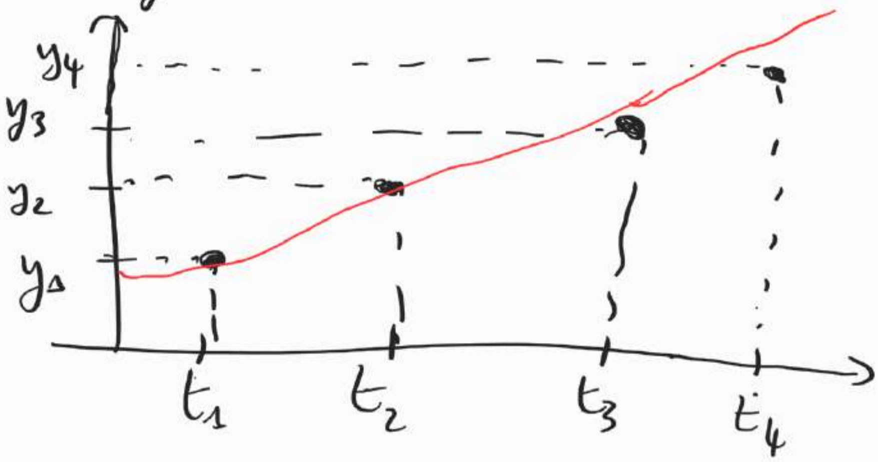
On se donne n mesures $y_1, \dots, y_n$ en n points $t_1, \dots, t_n$

et l'on cherche $P \in \mathbb{R}_d[X]$ qui minimise

$$E = \left\| \begin{pmatrix} y_1 \\ \vdots \\ y_n \end{pmatrix} - \begin{pmatrix} Q(t_1) \\ \vdots \\ Q(t_n) \end{pmatrix} \right\|_2$$

1) Ecrire ce problème comme la résolution au sens moindres carrés (m.c) d'une certaine Matrice $A \in \mathcal{M}_{n,d+1}(\mathbb{R})$ et un vecteur $b \in \mathbb{R}^n$.

1) C'est une généralisation de l'exo précédent $(d=1, n=3)$



• Rappel : interpolation de Lagrange (L2)

soit $y_1, \dots, y_n$ quelconques.

et $t_1, \dots, t_n$ distincts 2 à 2

alors $\exists! P \in \mathbb{R}_{n-1}[X], \ P(t_i) = y_i$

ce polynôme est le polynôme interpolateur de Lagrange.

• On peut aussi trouver des polynômes de degré plus élevé qui passent exactement par ces points

(par exemple en rajoutant des contraintes supplémentaires)

$\longrightarrow$ infinité de solutions dans ce cas.

Donc, si l'on revient à notre problème.

si $\underline{d \geq n-1}$: il y a au moins une solution exacte ($E=0$), qui est unique si $d=n-1$, pas unique si $d>n-1$.

$\underline{\text{si } d<n-1}$: pas nécessairement de solution exacte. On en cherche une au sens m.c.

Maintenant, on répond à la question :

Soit $P \in \mathbb{R}_d[X]$, alors $P(X) = a_0 + a_1 X + \dots + a_d X^d$

$$\begin{pmatrix} P(t_1) \\ \vdots \\ P(t_n) \end{pmatrix} = \begin{pmatrix} a_0 + a_1 t_1 + \dots + a_d t_1^d \\ \vdots \\ a_0 + a_1 t_n + \dots + a_d t_n^d \end{pmatrix}$$

$$= \underbrace{\begin{pmatrix} 1 & t_1 & t_1^2 & \dots & t_1^d \\ \vdots & & & & \\ 1 & t_n & t_n^2 & \dots & t_n^d \end{pmatrix}}_{\text{matrice } A,\ n \text{ lignes},\ d+1 \text{ colonnes}} \underbrace{\begin{pmatrix} a_0 \\ a_1 \\ \vdots \\ a_d \end{pmatrix}}_{\vec{x} \text{ vecteur à } d+1 \text{ composantes}} \quad (*)$$

En posant $\vec{b} = \begin{pmatrix} y_1 \\ \vdots \\ y_n \end{pmatrix} \in \mathbb{R}^n$

On cherche bien le ou les $\vec{x}$ qui minimisent

$$E = \| A\vec{x} - \vec{b} \|_2$$

donc c'est bien un moindre carré.

2) Pour $d = 0$ et $d = 1$, résoudre explicitement lorsqu'il y a une unique solution.

2) • <u>$d = 0$</u>

$$A = \begin{pmatrix} 1 \\ 1 \\ \vdots \\ 1 \end{pmatrix}, \quad \vec{b} = \begin{pmatrix} y_1 \\ \vdots \\ y_n \end{pmatrix}, \quad \vec{x} = (a_0)$$

Solution unique car $\ker A = \{0\}$

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

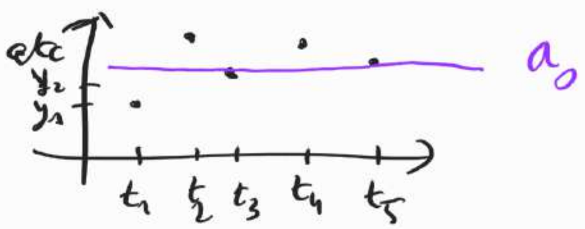
$$A^T A = (1 \cdots 1) \begin{pmatrix} 1 \\ \vdots \\ 1 \end{pmatrix} = \underbrace{1 + \cdots + 1}_{n \text{ fois}} = n$$

$$\Rightarrow n a_0 = (1 \cdots 1) \begin{pmatrix} y_1 \\ \vdots \\ y_n \end{pmatrix}$$

$$n a_0 = \sum_{i=1}^{n} y_i$$

$$a_0 = \frac{1}{n} \sum_{i=1}^{n} y_i$$

<u>Interprétation</u> :



La meilleure approximation par une fonction constante est la moyenne de $y_1, \ldots, y_n$.

$n \geq 1$

• $\underline{d = 1}$: Approximation par des fonctions affines.

$$A = \begin{pmatrix} 1 & t_1 \\ \vdots & \vdots \\ 1 & t_n \end{pmatrix}$$

Solution unique si $\ker A = \{0\}$

$\Leftrightarrow \begin{pmatrix} 1 \\ \vdots \\ 1 \end{pmatrix}$ et $\begin{pmatrix} t_1 \\ \vdots \\ t_n \end{pmatrix}$ non colinéaires

$\Leftrightarrow$ Tous les t_i ne sont pas égaux.

$$A^T A = \begin{pmatrix} 1 & \cdots & 1 \\ t_1 & \cdots & t_n \end{pmatrix} \begin{pmatrix} 1 & t_1 \\ \vdots & \vdots \\ 1 & t_n \end{pmatrix}$$

$$= \begin{pmatrix} n & \sum_{i=1}^n t_i \\ \sum_{i=1}^n t_i & \sum_{i=1}^n t_i^2 \end{pmatrix}$$

$$A^T y = \begin{pmatrix} 1 & \cdots & 1 \\ t_1 & \cdots & t_n \end{pmatrix} \begin{pmatrix} y_1 \\ \vdots \\ y_n \end{pmatrix}$$

$$= \begin{pmatrix} \sum_{i=1}^n y_i \\ \sum_{i=1}^n t_i y_i \end{pmatrix}$$

Au final
$$\begin{pmatrix} n & \sum t_i \\ \sum t_i & \sum t_i^2 \end{pmatrix} \begin{pmatrix} a_0 \\ a_1 \end{pmatrix} = \begin{pmatrix} \sum y_i \\ \sum t_i y_i \end{pmatrix}$$

$$\begin{cases} n a_0 + a_1 \sum t_i = \sum y_i & (1) \\ a_0 \sum t_i + a_1 \sum t_i^2 = \sum t_i y_i & (2) \end{cases}$$

$(\sum t_i)(1) - n(2)$:

$$a_1\left(\sum t_i\right)^2 - n\left(\sum t_i^2\right) a_1 = \left(\sum t_i\right)\left(\sum y_i\right) - n \sum t_i y_i$$

$$a_1 = \frac{\left(\sum t_i\right)\left(\sum y_i\right) - n \sum t_i y_i}{\left(\sum t_i\right)^2 - n \sum t_i^2} \quad \left(= \frac{\mathrm{cov}(x,t)}{\mathrm{var}\, t}\right)$$

cf cours.

$(\sum t_i^2)(1) - (\sum t_i)(2)$:

On trouve après calcul

$$a_0 = \frac{\left(\sum t_i^2\right)\left(\sum y_i\right) - \left(\sum t_i\right)\left(\sum t_i y_i\right)}{n \sum t_i^2 - \sum t_i^2} \quad \left(= \bar{y} - a_1 \bar{x}\right)$$

cf cours.

L'exercice 1 est un cas particulier de ce résultat

3) On suppose que les $t_1, \dots, t_n$ sont distincts 2 à 2.

A quelle condition sur d la matrice $A^T A$ est-elle inversible ?

(exo intéressant, voir à la fin)

3) $A^T A$ inversible $\iff \ker A = \{0\}$

$\iff$ il existe une unique solution au problème $(*)$ au sens des moindres carrés

$\hookrightarrow A\vec{x} = \vec{b}$.

- $\underline{d > n-1}$: On a plusieurs polynômes interpolateurs (exacts) donc plusieurs solutions (exactes) au sens des moindres carrés.

donc $A^T A$ n'est pas inversible.

• <u>$d = n-1$</u> : Il y a un unique polynôme interpolateur (le polynôme de Lagrange)

donc la solution est unique

donc $A^T A$ est inversible.

<u>Autre argument</u> : $A = \begin{pmatrix} 1 & t_1 & \cdots & t_1^d \\ \vdots & \vdots & & \vdots \\ 1 & t_d & \cdots & t_{d+1}^d \end{pmatrix}$ car $d = n-1 \Leftrightarrow d+1 = n$

donc A est une matrice carrée $((d+1)\times(d+1))$

et $$\det A = \prod_{1 \leq i < j \leq d+1} (t_j - t_i)$$

(déterminant de Vandermonde, cf L2)

Par hypothèse les t_i sont distincts 2 à 2, donc $\det A \neq 0$

donc A inversible donc $A^T A$ inversible
carrée.

• <u>$d < n-1$</u> : pas de solution exacte en général, mais on cherche au sens moindres carrés.

$$A = \begin{pmatrix} 1 & t_1 & \cdots & t_1^d \\ \vdots & \vdots & & \vdots \\ \vdots & \vdots & & \vdots \\ 1 & t_n & \cdots & t_n^d \end{pmatrix}$$

"plus de lignes que de colonnes"

on considère le bloc $(d+1)\times(d+1)$

La sous matrice en rouge est inversible (par exemple, car son déterminant calculé plus haut (1) est non nul.)

donc ses vecteurs colonnes sont linéairement indépendants.

donc $\begin{pmatrix} 1 \\ \vdots \\ 1 \end{pmatrix}, \begin{pmatrix} t_1 \\ \vdots \\ t_{d+1} \end{pmatrix}, \cdots \begin{pmatrix} t_1^d \\ \vdots \\ t_{d+1}^d \end{pmatrix}$ sont linéairement indépendants.

donc, comme $n > d+1$, $\begin{pmatrix} 1 \\ \vdots \\ 1 \end{pmatrix}, \begin{pmatrix} t_1 \\ \vdots \\ t_n \end{pmatrix}, \cdots \begin{pmatrix} t_1^d \\ \vdots \\ t_n^d \end{pmatrix}$

sont linéairement indépendants.

$\Rightarrow \ker A = \{0\}$

$\Rightarrow A^T A$ inversible

$\boxed{\text{Exo 4}}$ Théorème de Gerschgorin.

$A \in \mathcal{M}_n(\mathbb{C})$, $n \in \mathbb{N}^*$

Montrer que $\sigma(A) \subset \bigcup_{i=1}^{n} D_i$

$$D_i = \left\{ z \in \mathbb{C}, |z - A_{ii}| \leq \sum_{j \neq i} |A_{ij}| \right\}$$

↳ "disque de Gerschgorin"

(ou bien : toute valeur propre de A est située dans l'un de ces disques)

Soit λ une valeur propre de A, et $x = \begin{pmatrix} x_1 \\ \vdots \\ x_n \end{pmatrix}$ le vecteur propre associé : $Ax = \lambda x$

$$(*) \quad \sum_{j=1}^{n} A_{ij} x_j = \lambda x_i \qquad \text{pour tout } i \in \{1, \dots, n\}$$

ou bien
$$\begin{pmatrix} A_{11}x_1 + A_{12}x_2 + \cdots A_{1n}x_n \\ \\ \\ A_{n1}x_1 + A_{n2}x_2 + \cdots + A_{nn}x_n \end{pmatrix} = \begin{pmatrix} \lambda x_1 \\ \lambda x_2 \\ \vdots \\ \lambda x_n \end{pmatrix}$$

On choisit l'indice i_0 t.q $|x_{i_0}| = \max_{1 \leq i \leq n} |x_i|$
$= \|x\|_\infty$

Pour $i = i_0$, la relation $(*)$ devient

$$\sum_{j=1}^{n} A_{i_0 j} x_j = \lambda x_{i_0}$$

$$A_{i_0 i_0} x_{i_0} + \sum_{\substack{j=1 \\ j \neq i_0}}^{n} A_{i_0 j} x_j = \lambda x_{i_0}$$

$$(\lambda - A_{i_0 i_0}) x_{i_0} = \sum_{\substack{j=1 \\ j \neq i_0}}^{n} A_{i_0 j} x_j$$

$$\Rightarrow |\lambda - A_{i_0 i_0}| |x_{i_0}| = \left| \sum_{j \neq i_0} A_{i_0 j} x_j \right|$$

$$|\lambda - A_{i_0 i_0}|\,|x_{i_0}| \leq \sum_{j \neq i_0} |A_{i_0 j}|\,|x_j|$$

$$\forall j \in \{1, \dots, n\}, \quad |x_j| \leq |x_{i_0}|$$

$$\Rightarrow |\lambda - A_{i_0 i_0}| \leq \sum_{j \neq i_0} |A_{i_0 j}| \frac{|x_j|}{|x_{i_0}|}$$

$$\leq \sum_{j \neq i_0} |A_{i_0 j}|$$

(somme des éléments hors diagonaux sur la même ligne i_0)

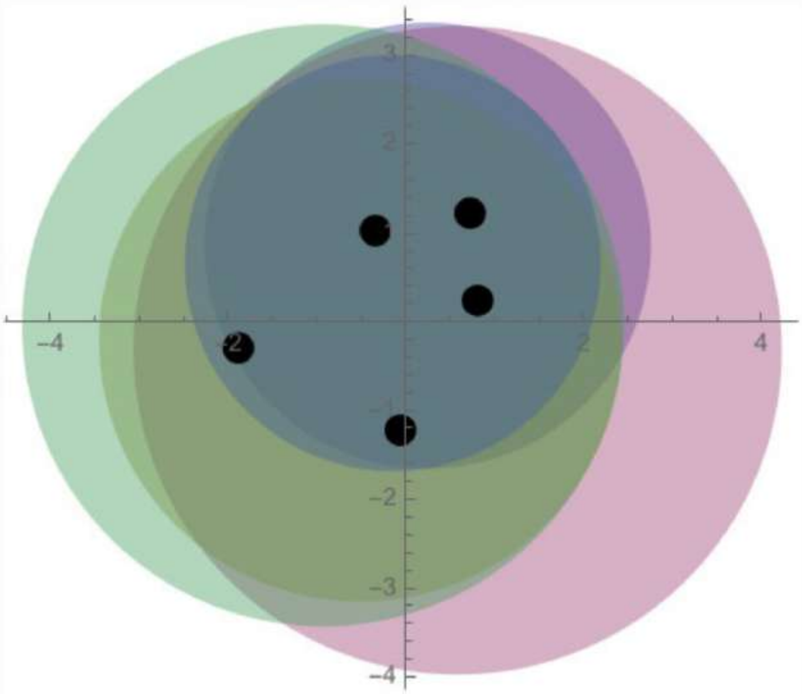
donc $\lambda \in D_{i_0}$

donc cette valeur propre est bien dans l'un des disques.

Exemples pour des matrices de $\mathcal{M}_5(\mathbb{C})$:

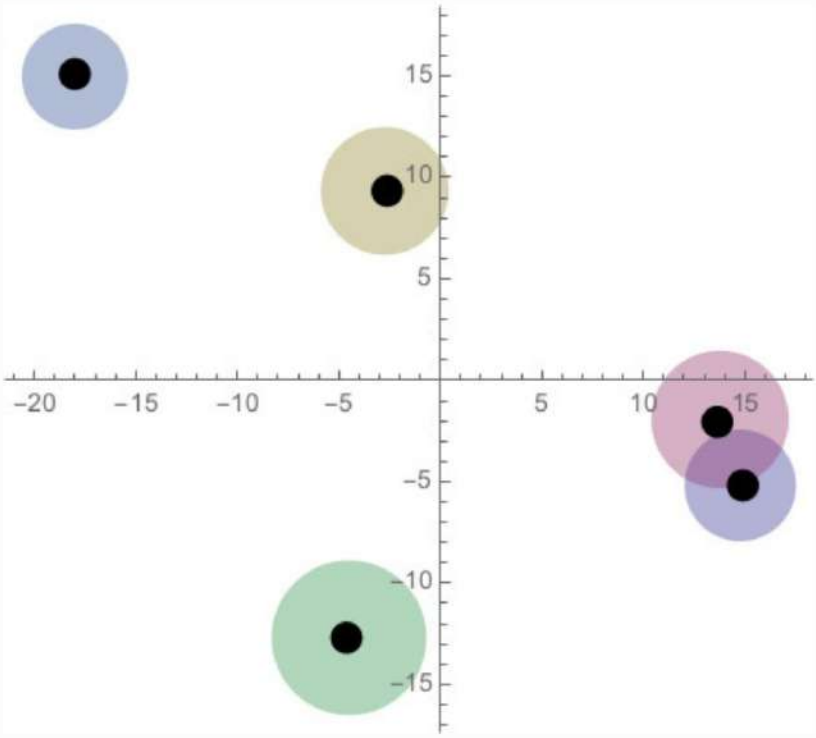
Une matrice avec coefficients tirés au hasard.

Les points noirs sont les valeurs propres
En couleurs les disques de Gerschgorin



Une matrice à diagonale dominante (cf fiche suivante)

Les valeurs propres sont proches des éléments diagonaux (les centres des disques)



Pourquoi $A^T A$ inversible $\Leftrightarrow$ $\ker A = \{0\}$?

On peut montrer que $\ker A^T A = \ker A$

- $\underline{\ker A \subset \ker A^T A}$: Soit $x \in \ker A$

$$\text{Alors } Ax = 0 \Rightarrow A^T A x = 0$$
$$\Rightarrow x \in \ker A^T A$$

- $\underline{\ker A^T A \subset \ker A}$: Soit $x \in \ker A^T A$

$$\text{Alors } A^T A x = 0 \Rightarrow x^T A^T A x = 0$$
$$\Rightarrow \|Ax\|_2^2 = 0$$
$$\Rightarrow Ax = 0$$
$$\Rightarrow x \in \ker A.$$