Cours 8 MT09. Résolution des problèmes de moindres carrés

Antoine Zurek

LMAC - UTC

I. Formulation du problème de moindres carrés
1. Position du problème
2. Meilleure solution au sens des moindres carrés
3. Résumé

II. Résolution théorique du problème de moindres carrés
1. Théorème d'existence et d'unicité
2. Preuve analytique des équations normales
3. Interprétation géométrique
4. Preuve algébrique pour le problème des moindres carrés

III. Résolution pratique du problème de moindres carrés
1. Décomposition QR
2. Application

Table of contents

I. Formulation du problème de moindres carrés
1. Position du problème
2. Meilleure solution au sens des moindres carrés
3. Résumé

II. Résolution théorique du problème de moindres carrés
1. Théorème d'existence et d'unicité
2. Preuve analytique des équations normales
3. Interprétation géométrique
4. Preuve algébrique pour le problème des moindres carrés

III. Résolution pratique du problème de moindres carrés
1. Décomposition QR
2. Application

1. Position du problème

Problème. Soit $A \in \mathcal{M}_{m,n}$ une matrice rectangulaire. On veut résoudre le système linéaire $Ax = b$ avec $x \in M_{n,1}$ et $b \in \mathcal{M}_{m,1}$ dans le cas où

$$m \gg n$$

- Ici on dispose de plus d'équations (m) que d'inconnues (n) !
- Il n'existe pas généralement de vecteur $x \in M_{n,1}$ satisfaisant toutes les équations du système $Ax = b$ (problème mal posé).
- Autrement dit, $b \notin \mathrm{Im}(A)$ en général.

1. Position du problème

Problème. Soit $A \in \mathcal{M}_{m,n}$ une matrice rectangulaire. On veut résoudre le système linéaire $Ax = b$ avec $x \in M_{n,1}$ et $b \in \mathcal{M}_{m,1}$ dans le cas où

$$m \gg n$$

- Ici on dispose de plus d'équations (m) que d'inconnues (n) !
- Il n'existe pas généralement de vecteur $x \in M_{n,1}$ satisfaisant toutes les équations du système $Ax = b$ (problème mal posé).
- Autrement dit, $b \notin \mathrm{Im}(A)$ en général.

Idée. Se contenter de trouver le "meilleur" $\bar{x}$ tq $A\bar{x} - b \approx 0$.

Question. Sens du terme meilleur dans ce contexte ?

2. Meilleure solution au sens des moindres carré

Définition 3.1.1.

Soit $A \in \mathcal{M}_{m,n}$ et $b \in \mathcal{M}_{m,1}$ donnés. On appelle problèmes de moindres carrés le problème

$$\min_{x \in \mathbb{R}^n} \|Ax - b\|_{2,m}^2,$$

où pour rappel, si $y \in \mathbb{R}^p$ $(p \geq 1)$, alors

$$\|y\|_{2,p}^2 = \sum_{i=1}^{p} |y_i|^2 = y^\top y.$$

Autrement dit $\|\cdot\|_{2,p}$ désigne la norme euclidienne dans $\mathbb{R}^p$.

2. Meilleure solution au sens des moindres carré

Définition 3.1.1.

Soit $A \in \mathcal{M}_{m,n}$ et $b \in \mathcal{M}_{m,1}$ donnés. On appelle problèmes de moindres carrés le problème

$$\min_{x \in \mathbb{R}^n} \|Ax - b\|_{2,m}^2,$$

où pour rappel, si $y \in \mathbb{R}^p$ $(p \geq 1)$, alors

$$\|y\|_{2,p}^2 = \sum_{i=1}^{p} |y_i|^2 = y^\top y.$$

Autrement dit $\|\cdot\|_{2,p}$ désigne la norme euclidienne dans $\mathbb{R}^p$.

Si le problème de moindres carrés à une solution, notée $\bar{x}$, alors $\bar{x}$ est la meilleure solution approchée du système linéaire $Ax = b$ au sens des moindres carrés. Autrement dit

$$(0 \leq)\|A\bar{x} - b\|_{2,m}^2 \leq \|Ax - b\|_{2,m}^2 \quad \forall x \in \mathbb{R}^m.$$

3. Résumé

- On remplace un problème (complexe) de la forme $Ax = b$, qui n'admet pas nécessairement de solutions, par un problème plus "simple" d'optimisation.
- On formule ce problème d'optimisation via la norme euclidienne, d'où le nom problème de moindres carrés.
- On cherche donc $\bar{x} \in \mathbb{R}^m$ tq $\|A\bar{x} - b\|_{2,m}^2$ soit la quantité positive ou nulle la plus petite possible.

À retenir. le problème des moindres carrés consiste à remplacer un système d'équations qui n'admet pas nécessairement de solution exacte par un problème d'optimisation qui cherche la solution minimisant l'erreur quadratique.

3. Résumé

- On remplace un problème (complexe) de la forme $Ax = b$, qui n'admet pas nécessairement de solutions, par un problème plus "simple" d'optimisation.
- On formule ce problème d'optimisation via la norme euclidienne, d'où le nom problème de moindres carrés.
- On cherche donc $\bar{x} \in \mathbb{R}^m$ tq $\|A\bar{x} - b\|_{2,m}^2$ soit la quantité positive ou nulle la plus petite possible.

À retenir. le problème des moindres carrés consiste à remplacer un système d'équations qui n'admet pas nécessairement de solution exacte par un problème d'optimisation qui cherche la solution minimisant l'erreur quadratique.

Objectif pour la suite. Assurer l'existence d'une solution au problème de moindres carrés.

Table of contents

I. Formulation du problème de moindres carrés
1. Position du problème
2. Meilleure solution au sens des moindres carrés
3. Résumé

II. Résolution théorique du problème de moindres carrés
1. Théorème d'existence et d'unicité
2. Preuve analytique des équations normales
3. Interprétation géométrique
4. Preuve algébrique pour le problème des moindres carrés

III. Résolution pratique du problème de moindres carrés
1. Décomposition QR
2. Application

1. Théorème d'existence et d'unicité

Théorème 3.1.1.

Soit $A \in \mathcal{M}_{m,n}$ et $b \in \mathbb{R}^m$ donnés. Si $\mathrm{rang}(A) = n$, alors le problème de moindres carrés

$$\min_{x \in \mathbb{R}^n} \|Ax - b\|_{2,m}^2,$$

admet une unique solution notée $\bar{x}$ satisfaisant les équations, dites normales, suivantes :

$$A^\top A \bar{x} = A^\top b.$$

1. Théorème d'existence et d'unicité

Reformulation du Théorème 3.1.1.

Soit $A \in \mathcal{M}_{m,n}$ et $b \in \mathbb{R}^m$ donnés. Si $\mathrm{rang}(A) = n$, alors $\bar{x} \in \mathbb{R}^n$ est l'unique solution du problème de moindres carrés

$$\|A\bar{x} - b\|_{2,m}^2 = \min_{x \in \mathbb{R}^n} \|Ax - b\|_{2,m}^2,$$

ssi $\bar{x}$ est l'unique solution du systèmes linéaire :

$$A^\top A\,\bar{x} = A^\top b.$$

Remarque 1. comme $\mathrm{rang}(A) \leq \min(n, m)$, alors imposer que $\mathrm{rang}(A) = n$ signifie que les colonnes de A forment une base de $\mathrm{Im}(A)$. De plus, le théorème du rang implique que

$$\dim(\mathrm{Ker}(A)) = n - \mathrm{rang}(A) = 0.$$

1. Théorème d'existence et d'unicité

Reformulation du Théorème 3.1.1.

Soit $A \in \mathcal{M}_{m,n}$ et $b \in \mathbb{R}^m$ donnés. Si $\mathrm{rang}(A) = n$, alors $\bar{x} \in \mathbb{R}^n$ est l'unique solution du problème de moindres carrés

$$\|A\bar{x} - b\|_{2,m}^2 = \min_{x \in \mathbb{R}^n} \|Ax - b\|_{2,m}^2,$$

ssi $\bar{x}$ est l'unique solution du systèmes linéaire :

$$A^\top A\,\bar{x} = A^\top b.$$

Remarque 2. $A^\top A \in \mathcal{M}_{n,n}$. Si $\mathrm{rang}(A) = n$ alors $\mathrm{rang}(A^\top A) = n$ et donc le système

$$A^\top A\,\bar{x} = A^\top b,$$

admet une unique solution.

2. Preuve analytique des équations normales

Dans un premier temps, on a que

$$E(x) = \|Ax - b\|_{2,m}^2 = x^\top A^\top Ax - 2\left(A^\top b\right)^\top x + \|b\|_{2,m}^2.$$

Ainsi, trouver un $\bar{x}$ qui minimise $E(x)$ revient à trouver un $\bar{x}$ tq

$$J(\bar{x}) = \min_{x\in\mathbb{R}^n} J(x) = \min_{x\in\mathbb{R}^n} \left(x^\top Gx - 2h^\top x\right),$$

avec $G = A^\top A$ symétrique semi-définie positive et $h = A^\top b$ un vecteur. Comme dans le cas des fonction de $\mathbb{R} \to \mathbb{R}$, chercher un minimum de $J : \mathbb{R}^n \to R$ revient à chercher un $\bar{x} \in \mathbb{R}^n$ vérifiant $\nabla J(\bar{x}) = \vec{0}$. Pour rappel

$$\nabla J(x) = \left(\frac{\partial J}{\partial x_1}, \frac{\partial J}{\partial x_2}, \ldots, \frac{\partial J}{\partial x_n}\right)^\top.$$

Ici on obtient

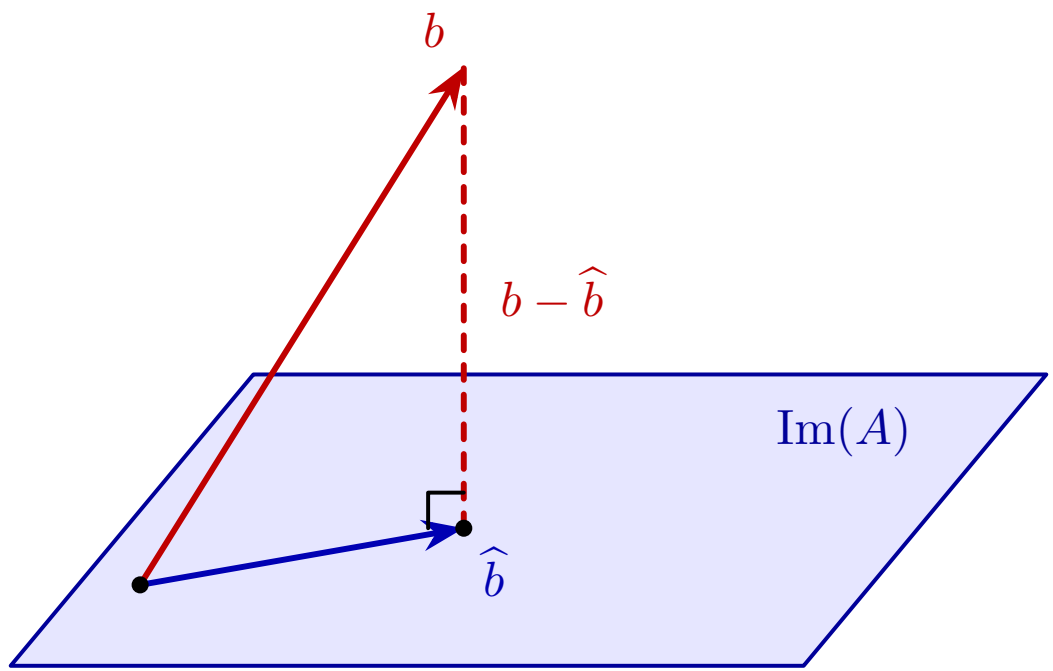
$$\nabla J(\bar{x}) = \vec{0} \Leftrightarrow 2(G\bar{x} - h) = \vec{0} \Leftrightarrow A^\top A\bar{x} = A^\top b.$$

3. Interprétation géométrique

Résoudre le problème

$$\min_{x \in \mathbb{R}^n} \|Ax - b\|_{2,m}^2,$$

revient à chercher $\hat{b} \in \mathrm{Im}(A)$ tq $\|\hat{b} - b\|_{2,m}^2$ soit le plus petit possible.



4. Preuve algébrique pour le problème des moindres carrés

Proposition 3.2.7.

Soit S un s.e.v. de $\mathbb{R}^m$ et $y \in \mathbb{R}^m$ un vecteur donné. Le problème d'optimisation

$$\min_{z \in S} \|z - y\|_{2,m}^2,$$

admet une unique solution $\hat{y} \in S$ projection orthogonale de y sur S.

Rappel/Remarque. Le projeté orthogonale $\hat{y} \in S$ de y est caractérisé par

$$\langle y - \hat{y},\, s \rangle = 0 \quad \forall s \in S.$$

En particulier, $\hat{y}$ vérifie

$$\|y - \hat{y}\|_{2,m}^2 < \|y - z\|_{2,m}^2 \quad \forall z \in S \text{ avec } z \neq \hat{y}.$$

Conséquence. On applique le résultat précédent à $S = \mathrm{Im}(A)$ et $y = b$, ce qui permet de déduire l'existence et l'unicité d'un $\hat{b} \in \mathrm{Im}(A)$ tq $\|\hat{b} - b\|_{2,m}^2$ soit le plus petit possible.

4. Preuve algébrique pour le problème des moindres carrés

Corollaire 3.2.3.

Soit $A \in \mathcal{M}_{m,n}$ et $b \in \mathbb{R}^m$ donnés. Le problème de moindres carrés, trouver $\bar{x} \in \mathbb{R}^n$ tel que

$$\|A\bar{x} - b\|_{2,m}^2 \leq \|Ax - b\|_{2,m}^2 \quad \forall x \in \mathbb{R}^n, \tag{1}$$

est équivalent à trouver $\bar{x} \in \mathbb{R}^n$ tel que

$$A^\top A\bar{x} = A^\top b. \tag{2}$$

Conséquences. L'inégalité (1) est une reformulation du problème de moindres carrés initial. Ainsi, trouver une solution au problème de moindres carrés revient à trouver une solutions aux équations normales (2). Or, on sait que si $\mathrm{rang}(A) = n$ alors (2) admet une unique solution dans $\mathbb{R}^n$!

Table of contents

I. Formulation du problème de moindres carrés
1. Position du problème
2. Meilleure solution au sens des moindres carrés
3. Résumé

II. Résolution théorique du problème de moindres carrés
1. Théorème d'existence et d'unicité
2. Preuve analytique des équations normales
3. Interprétation géométrique
4. Preuve algébrique pour le problème des moindres carrés

III. Résolution pratique du problème de moindres carrés
1. Décomposition QR
2. Application

1. Décomposition QR

D'après ce qui précède, si $\mathrm{rang}(A) = n$, résoudre le problème de moindres carrés revient à résoudre le système linéaire $n \times n$

$$A^\top A\bar{x} = A^\top b.$$

Cependant, on peut montrer que $\chi_2(A^\top A) = \chi_2(A)^2$ et donc le conditionnement de $A^\top A$ peut être très mauvais !
En pratique, on va utiliser le résultat suivant :

Théorème 3.3.4.

Soit $A \in \mathcal{M}_{m,n}$ avec $m \geq n$, alors il existe $Q \in \mathcal{M}_{m,m}$ avec $Q^\top Q = I_m$ et une matrice triangulaire supérieure $\tilde{R} \in \mathcal{M}_{n,n}$ telles que

$$A = Q \begin{pmatrix} \tilde{R} \\ 0_{m-n,n} \end{pmatrix} = QR.$$

Remarque. Si $\mathrm{rang}(A) = n$ alors $\tilde{R}$ est inversible.

2. Application

Si $\mathrm{rang}(A) = n$ alors

$$A^\top A\bar{x} = A^\top b \quad \Leftrightarrow \quad R^\top \underbrace{Q^\top Q}_{=I_m} R\,\bar{x} = R^\top Q^\top b.$$

Soit $c \in \mathbb{R}^n$ et $d \in \mathbb{R}^{m-n}$ alors en notant

$$Q^\top b = \begin{pmatrix} c \\ d \end{pmatrix},$$

on obtient

$$R^\top R\,\bar{x} = R^\top Q^\top b \quad \Leftrightarrow (\tilde{R}^\top \quad 0) \begin{pmatrix} \tilde{R} \\ 0 \end{pmatrix} \bar{x} = (\tilde{R}^\top \quad 0) \begin{pmatrix} c \\ d \end{pmatrix}.$$

2. Application

Ainsi, le problème initial est équivalent à résoudre le système

$$\tilde{R}^\top \tilde{R}\bar{x} = \tilde{R}^\top c \quad \Leftrightarrow \quad \tilde{R}\bar{x} = c,$$

car $\tilde{R}$ est inversible, donc $\tilde{R}^\top$ aussi. Ce dernier système est un système tri. sup. à résoudre !

À retenir (en conservant les notations précédentes)

Soit $A \in \mathcal{M}_{m,n}$ et $b \in \mathbb{R}^m$ donnée. Résoudre le problème de moindres carrés associé à A et b est équivalent à inverser le système triangulaire supérieur :

$$\tilde{R}\bar{x} = c.$$

Remarque. On peut montrer que $\chi_2(\tilde{R}) = \chi_2(A)$.

Retour sur les espaces orthogonaux

On considère $\mathbb{R}^m$ doté du p.s. euclidien/canonique/usuel

$$\langle x, y \rangle_m = x^\top y = y^\top x \quad \forall x,\, y \in \mathbb{R}^m,$$

on retrouve bien $\|x\|_{2,m}^2 = \langle x, x \rangle_m$. Soit $(x_1, \ldots, x_k)$ une famille de vecteurs de $\mathbb{R}^m$. On dit que cette famille est orthonormée pour le p.s. euclidien si

$$\langle x_i, x_j \rangle_m = 0 \quad \forall i \neq j \quad \text{et} \quad \|x_i\|_{2,m} = 1 \quad \forall i.$$

Pour rappel, une famille orthonormée (ou juste orthogonale) est libre (voir Proposition 3.2.3 et Corollaire 3.2.1).

Retour sur les espaces orthogonaux

Soit S un s.e.v. de $\mathbb{R}^m$, alors l'ensemble

$$S^{\perp} = \{x \in \mathbb{R}^m \ : \ \langle x, y \rangle_m = \vec{0} \quad \forall y \in S\},$$

est un s.e.v. de $\mathbb{R}^m$ appelé orthogonal de S. Voici des résultats sur S et $S^{\perp}$ (voir Prop 3.2.5, Coro 3.2.2 et Prop 3.2.6) :

- $S \cap S^{\perp} = \{\vec{0}\}$ et $\mathbb{R}^m = S \oplus S^{\perp}$.
- Soit $y \in \mathbb{R}^m$, il existe un unique $\hat{y} \in S$ tel que $y - \hat{y} \in S^{\perp}$.
- Le vecteur $\hat{y}$ est le projeté orthogonal de y sur S et vérifie

$$\langle y - \hat{y}, s \rangle_m = 0 \quad \forall s \in S.$$

- Pour $A \in \mathcal{M}_{m,n}$ alors $(\mathrm{Im}(A))^{\perp} = \mathrm{Ker}\left(A^{\top}\right)$.