

11. Méthode des moindres carrés et procédé de Gram-Schmidt

MTH1007

Sébastien Le Digabel
Polytechnique Montréal

A26

2026-07-28

v0

Plan

Approximations par moindres carrés

- Moindres carrés : Définition

- Droite d'ajustement (régression linéaire)

- Trois points de vue

- Solution analytique

- Généralisation

- Exercices

Procédé de Gram-Schmidt

Décomposition QR

Liens avec le manuel *alfa*

- ▶ 9.2 : Moindres carrés
- ▶ 9.3 : Régression linéaire
- ▶ 10 : Orthogonalisation

Approximations par moindres carrés

Moindres carrés : Définition

Droite d'ajustement (régression linéaire)

Trois points de vue

Solution analytique

Généralisation

Exercices

Procédé de Gram-Schmidt

Décomposition QR

Introduction

- ▶ $\mathbf{A} \in \mathbb{R}^{m \times n}$ avec $m > n$ et $\text{rg}(\mathbf{A}) = n$
- ▶ Le SÉL $\mathbf{Ax} = \mathbf{b}$ est de plein rang colonne et ne possède pas toujours de solution
- ▶ Au lieu on résout le système $\mathbf{A}^\top \mathbf{Ax} = \mathbf{A}^\top \mathbf{b}$ qui possède toujours une solution (ce sont les *équations normales*)
- ▶ $\mathbf{p} = \mathbf{Ax}$ est la projection de $\mathbf{b}$ dans $\text{Im}(\mathbf{A})$: C'est donc le point de $\text{Im}(\mathbf{A})$ le plus proche de $\mathbf{b}$
- ▶ $\hat{\mathbf{x}}$ minimise l'erreur

$$\|\mathbf{e}\|^2 = \|\mathbf{b} - \mathbf{p}\|^2 = \|\mathbf{b} - \mathbf{Pb}\|^2 = \|\mathbf{b} - \mathbf{Ax}\|^2$$

- ▶ On peut donc voir $\hat{\mathbf{x}}$ comme la *meilleure "solution" possible* à $\mathbf{Ax} = \mathbf{b}$
- ▶ $\hat{\mathbf{x}}$ est appelée la solution de $\mathbf{Ax} = \mathbf{b}$ au sens des *moindres carrés*

Application : Droite d'ajustement (1/3)

- ▶ Étant donnés $m > 2$ *points de données* $(t_1, b_1), (t_2, b_2), \dots, (t_m, b_m)$ de $\mathbb{R}^2$, on essaie de trouver l'équation d'une droite qui passe par les m points
- ▶ Cette équation est $b(t) = c + dt$ avec $c, d \in \mathbb{R}$
- ▶ c et d devraient être les solutions du système $\mathbf{A}\mathbf{x} = \mathbf{b}$ avec

$$\mathbf{A} = \begin{bmatrix} 1 & t_1 \\ 1 & t_2 \\ \vdots & \vdots \\ 1 & t_m \end{bmatrix}, \mathbf{x} = \begin{bmatrix} c \\ d \end{bmatrix}, \mathbf{b} = \begin{bmatrix} b_1 \\ b_2 \\ \vdots \\ b_m \end{bmatrix}$$

Droite d'ajustement (2/3)

- ▶ Or le système $A\mathbf{x} = \mathbf{b}$ n'a a priori pas de solution car il est peu probable que les m points soient alignés
- ▶ Au lieu, on résout

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

- ▶ La solution $\hat{\mathbf{x}} = (\hat{c}, \hat{d})$ donnera la *droite d'ajustement* ou de *régression* qui minimise la somme des erreurs verticales avec les points de données : $\hat{\mathbf{x}}$ est tel que $E(\mathbf{x}) = \|A\mathbf{x} - \mathbf{b}\|^2$ est le plus petit possible :

$$\|\mathbf{e}\|^2 = E(\hat{\mathbf{x}}) = \min_{\mathbf{x} \in \mathbb{R}^2} E(\mathbf{x})$$

Droite d'ajustement (3/3)

Explicitement, le système d'équations 2×2 qui définit la droite d'ajustement est

$$\underbrace{\begin{bmatrix} m & \sum_{i=1}^m t_i \\ \sum_{i=1}^m t_i & \sum_{i=1}^m t_i^2 \end{bmatrix}}_{\mathbf{A}^\top \mathbf{A}} \underbrace{\begin{bmatrix} \hat{c} \\ \hat{d} \end{bmatrix}}_{\hat{\mathbf{x}}} = \underbrace{\begin{bmatrix} \sum_{i=1}^m b_i \\ \sum_{i=1}^m t_i b_i \end{bmatrix}}_{\mathbf{A}^\top \mathbf{b}}$$

Point de vue géométrique

- ▶ On cherche le vecteur de $\text{Im}(\mathbf{A})$ qui est le plus proche de $\mathbf{b}$
- ▶ Il s'agit du vecteur $\mathbf{p}$ qui minimise (le carré de) la distance

$$\|\mathbf{e}\|^2 = \|\mathbf{p} - \mathbf{b}\|^2$$

- ▶ Cette distance représente la somme des carrés des erreurs verticales entre les points et la droite

Point de vue algébrique

- ▶ Chaque vecteur $\mathbf{b} \in \mathbb{R}^m$ se décompose en $\mathbf{b} = \mathbf{p} + \mathbf{e}$ où $\mathbf{p} \in \text{Im}(\mathbf{A})$ et $\mathbf{e} \in \text{Im}(\mathbf{A})^\perp = \text{Ker}(\mathbf{A}^\top)$
- ▶ Le système $\mathbf{Ax} = \mathbf{p} + \mathbf{e}$ n'a pas de solution mais $\mathbf{A}\hat{\mathbf{x}} = \mathbf{p}$ possède une solution
- ▶ $\mathbf{Ax} \in \mathbb{R}^m$ et $\mathbf{p} \in \mathbb{R}^m$ sont dans $\text{Im}(\mathbf{A})$, donc $\mathbf{Ax} - \mathbf{p} \in \text{Im}(\mathbf{A})$. Ainsi $(\mathbf{Ax} - \mathbf{p}) \perp \mathbf{e}$
- ▶ Cette solution minimise l'erreur

$$E(\mathbf{x}) = \|\mathbf{Ax} - \mathbf{b}\|^2 = \|\mathbf{Ax} - \mathbf{p}\|^2 + \|\mathbf{e}\|^2$$

et

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

Point de vue de l'optimisation

- ▶ On minimise la fonction d'erreur

$$E(\mathbf{x}) = \|\mathbf{Ax} - \mathbf{b}\|^2$$

- ▶ Pour cela, on résout $\nabla E(\mathbf{x}) = \mathbf{0}$ (donne $\hat{\mathbf{x}}$) et on montre que $\nabla^2 E(\hat{\mathbf{x}})$ est *définie-positive*

- ▶ Ce qui revient à

$$\mathbf{A}^\top \mathbf{A} \hat{\mathbf{x}} = \mathbf{A}^\top \mathbf{b}$$

Solution analytique

- ▶ Il n'est pas nécessaire de résoudre explicitement $\mathbf{A}^\top \mathbf{A} \hat{\mathbf{x}} = \mathbf{A}^\top \mathbf{b}$, car on peut trouver une solution analytique
- ▶ Par exemple, par l'optimisation, on obtient

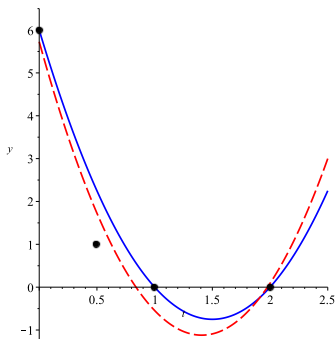
$$\begin{cases} \hat{d} &= \frac{\sum_{i=1}^m t_i b_i - m \hat{t} \hat{b}}{\sum_{i=1}^m t_i^2 - m \hat{t}^2} = \frac{\sum_{i=1}^m (t_i - \hat{t})(b_i - \hat{b})}{\sum_{i=1}^m (t_i - \hat{t})^2} \\ \hat{c} &= \hat{b} - \hat{d} \hat{t} \end{cases}$$

avec $\hat{b} = \frac{1}{m} \sum_{i=1}^m b_i$ et $\hat{t} = \frac{1}{m} \sum_{i=1}^m t_i$ (moyennes des b_i et des t_i)

Généralisation

Le même principe s'applique à l'ajustement d'une courbe de forme connue à un ensemble de points donnés

Par exemple : Ajuster une parabole sur les points $(0, 6)$, $(1, 0)$, $(2, 0)$ et $(1/2, 1)$:



Exercice 1

Ajuster une droite aux points $(t, b) = (-1, 5), (1, 13), (2, 17)$.
Calculer l'erreur de la régression et conclure

Exercice 2

Ajuster une surface aux points

$$(x, y, z) = (1, 0, 0), (0, 1, 1), (-1, 0, 3), (0, -1, 4)$$

Approximations par moindres carrés

Moindres carrés : Définition

Droite d'ajustement (régression linéaire)

Trois points de vue

Solution analytique

Généralisation

Exercices

Procédé de Gram-Schmidt

Décomposition QR

Procédé de Gram-Schmidt (1/2)

Soit $\mathbf{u}_1, \mathbf{u}_2, \dots, \mathbf{u}_n$ des vecteurs linéairement indépendants de $\mathbb{R}^m$ et W le sous-espace engendré par ces vecteurs ($n \leq m$).

La procédure suivante produit une base orthonormale de W :

$$(1) \quad \mathbf{v}_1 = \mathbf{u}_1$$

$$(2) \quad \mathbf{v}_2 = \mathbf{u}_2 - \frac{\mathbf{v}_1^\top \mathbf{u}_2}{\mathbf{v}_1^\top \mathbf{v}_1} \mathbf{v}_1 \quad (\mathbf{v}_2 \perp \mathbf{v}_1 \text{ car obtenu par } \mathbf{u}_2 - \mathbf{p} \text{ avec } \mathbf{p} \text{ la projection de } \mathbf{u}_2 \text{ sur } \mathbf{v}_1)$$

$$(3) \quad \mathbf{v}_3 = \mathbf{u}_3 - \frac{\mathbf{v}_1^\top \mathbf{u}_3}{\mathbf{v}_1^\top \mathbf{v}_1} \mathbf{v}_1 - \frac{\mathbf{v}_2^\top \mathbf{u}_3}{\mathbf{v}_2^\top \mathbf{v}_2} \mathbf{v}_2$$

$$\vdots$$

Procédé de Gram-Schmidt (2/2)

$$(j) \quad \mathbf{v}_j = \mathbf{u}_j - \sum_{k=1}^{j-1} \frac{\mathbf{v}_k^\top \mathbf{u}_j}{\mathbf{v}_k^\top \mathbf{v}_k} \mathbf{v}_k$$

$$\vdots$$

$$(n) \quad \mathbf{v}_n = \mathbf{u}_n - \sum_{k=1}^{n-1} \frac{\mathbf{v}_k^\top \mathbf{u}_n}{\mathbf{v}_k^\top \mathbf{v}_k} \mathbf{v}_k$$

$$(n+1) \quad \mathbf{q}_j = \frac{\mathbf{v}_j}{\|\mathbf{v}_j\|} \text{ pour tout } j \in \llbracket 1; n \rrbracket$$

Les vecteurs $\mathbf{q}_1, \mathbf{q}_2, \dots, \mathbf{q}_n$ forment la matrice $\mathbf{Q}$ et la base orthonormale recherchée. On a donc $W = \text{Im}(\mathbf{Q}) = \text{Im}(\mathbf{A})$ si $\mathbf{A} = [\mathbf{u}_1 \ \mathbf{u}_2 \ \cdots \ \mathbf{u}_n]$

Procédé de Gram-Schmidt : Remarques

- ▶ Si $m = n$, on obtient une base orthonormale de $\mathbb{R}^n$
- ▶ Le coût du procédé est en $\mathcal{O}(mn^2)$
- ▶ La transformation de Householder permet aussi d'obtenir une base orthonormale, mais à moindre coût
- ▶ La forme matricielle du procédé de Gram-Schmidt est la **décomposition QR**

Approximations par moindres carrés

Moindres carrés : Définition

Droite d'ajustement (régression linéaire)

Trois points de vue

Solution analytique

Généralisation

Exercices

Procédé de Gram-Schmidt

Décomposition QR

Factorisation QR (1/2)

- ▶ Si $\mathbf{A} = [\mathbf{u}_1 \ \mathbf{u}_2 \ \cdots \ \mathbf{u}_n] \in \mathbb{R}^{m \times n}$ et $\mathbf{Q} = [\mathbf{q}_1 \ \mathbf{q}_2 \ \cdots \ \mathbf{q}_n] \in \mathbb{R}^{m \times n}$ est obtenue par le procédé de Gram-Schmidt alors les $\mathbf{u}_i$ sont des combinaisons linéaires des $\mathbf{q}_i$ (et vice versa), de sorte que

$$\mathbf{A} = \mathbf{QR} \text{ avec } \mathbf{R} \in \mathbb{R}^{n \times n}$$

- ▶ $\mathbf{R}$ est triangulaire supérieure et peut s'obtenir par $\mathbf{R} = \mathbf{Q}^\top \mathbf{A}$
- ▶ Ses éléments diagonaux correspondent aux normes des $\mathbf{v}_j$ du procédé de Gram-Schmidt

Factorisation QR (2/2)

- ▶ La décomposition fonctionne pour des matrices carrées ou rectangulaires ($n \leq m$), à condition que les colonnes de $\mathbf{A}$ soient indépendantes
- ▶ Illustration avec $m = n = 3$:

$$\begin{bmatrix} \mathbf{u}_1 & \mathbf{u}_2 & \mathbf{u}_3 \end{bmatrix} = \begin{bmatrix} \mathbf{q}_1 & \mathbf{q}_2 & \mathbf{q}_3 \end{bmatrix} \begin{bmatrix} \mathbf{q}_1^\top \mathbf{u}_1 & \mathbf{q}_1^\top \mathbf{u}_2 & \mathbf{q}_1^\top \mathbf{u}_3 \\ & \mathbf{q}_2^\top \mathbf{u}_2 & \mathbf{q}_2^\top \mathbf{u}_3 \\ & & \mathbf{q}_3^\top \mathbf{u}_3 \end{bmatrix}$$

- ▶ Les formules des moindres carrés se simplifient en $\hat{\mathbf{x}} = \mathbf{R}^{-1} \mathbf{Q}^\top \mathbf{b}$ où $\mathbf{R}^{-1}$ se calcule facilement par remontée triangulaire

Exercice 3

Effectuer la décomposition QR de $\mathbf{A} = \begin{bmatrix} 1 & 2 & 3 \\ -1 & 0 & -3 \\ 0 & -2 & 3 \end{bmatrix}$