

Moindres Carrés

Slides 1ère année SN

Jean-Yves Tournet⁽¹⁾

(1) University of Toulouse, ENSEEIHT-IRIT-TéSA
jyt@n7.fr <http://perso.tesa.prd.fr/jyt/>

Année 2022 – 2023

Bibliographie

Quelques références

- ▶ **Joseph Salmon**, Notes de cours “Modèles linéaires”, <http://josephsalmon.eu/MDI720.html>
- ▶ **Stephen Boyd and Lieven Vandenberghe**, Introduction to Applied Linear Algebra. Vectors, Matrices, and Least Squares, Cambridge University Press, Cambridge, UK, 2018.
- ▶ **John Fox**, Robust Regression. Appendix to An R and S-PLUS Companion to Applied Regression, Sage Publications, Thousands Oaks, USA, 2002.

Plan du cours

Résumé

- ▶ **Introduction**
- ▶ Moindres carrés ordinaires
- ▶ Moindres carrés totaux
- ▶ Exemples

Notation “Boîte Noire”

Modèle de régression

$$\begin{aligned} f : \mathbb{R}^m \times \mathbb{R}^p &\longrightarrow \mathbb{R}^q \\ (\mathbf{x}, \boldsymbol{\beta}) &\longmapsto \mathbf{y} = f(\mathbf{x}, \boldsymbol{\beta}) \end{aligned}$$

- ▶ **modèle** défini par la fonction f
- ▶ **variables d'entrée** $\mathbf{x} \in \mathbb{R}^m$ (covariables, régresseurs, variables explicatives)
- ▶ **variables de sortie** $\mathbf{y} \in \mathbb{R}^q$ (réponses, variables expliquées)
- ▶ **paramètres** $\boldsymbol{\beta} \in \mathbb{R}^p$

Mesures

$$\mathbf{y}_i = f(\mathbf{x}_i, \boldsymbol{\beta}), \quad i \in \{1, \dots, n\}$$

Problème : comment estimer $\boldsymbol{\beta}$ à partir de $(\mathbf{x}_1, \dots, \mathbf{x}_n)$ et $(\mathbf{y}_1, \dots, \mathbf{y}_n)$?

Modèles de régression

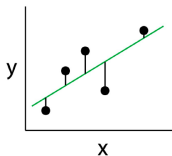
Modèle linéaire

$$y = f(x, \beta) = C(x)\beta + D(x)$$

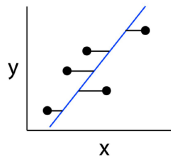
avec $C : \mathbb{R}^m \rightarrow \mathbb{R}^{q \times p}$, $D : \mathbb{R}^m \rightarrow \mathbb{R}^q$ et $\beta \in \mathbb{R}^p$.

Régression linéaire - erreurs de modèle

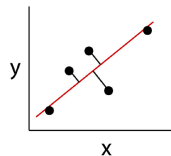
A Vertical residuals:
x independent,
y dependent



B Horizontal residuals:
x dependent,
y independent



C Perpendicular residuals



Erreurs de modèle

Fonctions de coût

- ▶ Erreurs **verticales** : $y_i = \beta_1 x_i + \beta_2 + \epsilon_i$. Critère à optimiser

$$J(\beta) = \sum_{i=1}^n (y_i - \beta_1 x_i - \beta_2)^2$$

Régression linéaire

- ▶ Erreurs **horizontales** : $y_i = \beta_1(x_i + \epsilon_i) + \beta_2$. Critère à optimiser

$$J(\beta) = \sum_{i=1}^n \left(\frac{y_i - \beta_2}{\beta_1} - x_i \right)^2$$

Régression non linéaire

- ▶ Erreurs **orthogonales**. Critère à optimiser

$$J(\beta) = \sum_{i=1}^n d_i^2 = \sum_{i=1}^n \frac{(y_i - \beta_1 x_i - \beta_2)^2}{1 + \beta_1^2}$$

Régression non linéaire

Plan du cours

Résumé

- ▶ Introduction
- ▶ **Moindres carrés ordinaires**
- ▶ Moindres carrés totaux
- ▶ Exemples

Moindre carrés ordinaires (ordinary least squares)

Fonction de coût à optimiser

$$J(\boldsymbol{\beta}) = \frac{1}{2} \sum_{i=1}^n \|\mathbf{y}_i - f(\mathbf{x}_i, \boldsymbol{\beta})\|^2, \boldsymbol{\beta} \in \mathbb{R}^p$$

Dans le cas **linéaire unidimensionnel**, on a

$y_i = f(\mathbf{x}_i, \boldsymbol{\beta}) = \mathbf{C}(\mathbf{x}_i)\boldsymbol{\beta} + \mathbf{D}(\mathbf{x}_i) \in \mathbb{R}$. Donc

$$J(\boldsymbol{\beta}) = \frac{1}{2} \|\mathbf{A}\boldsymbol{\beta} - \mathbf{b}\|^2, \mathbf{A} \in \mathbb{R}^{n \times p}, \mathbf{b} \in \mathbb{R}^n, \boldsymbol{\beta} \in \mathbb{R}^p$$

avec

$$\mathbf{A} = \begin{bmatrix} \mathbf{C}(\mathbf{x}_1) \\ \vdots \\ \mathbf{C}(\mathbf{x}_n) \end{bmatrix} \in \mathbb{R}^{n \times p}, \mathbf{b} = \begin{bmatrix} y_1 - \mathbf{D}(\mathbf{x}_1) \\ \vdots \\ y_n - \mathbf{D}(\mathbf{x}_n) \end{bmatrix} \in \mathbb{R}^n,$$

Cas linéaire

Fonction de coût à optimiser

$$J(\boldsymbol{\beta}) = \frac{1}{2} \|\mathbf{A}\boldsymbol{\beta} - \mathbf{b}\|^2, \quad \mathbf{A} \in \mathbb{R}^{n \times p}, \mathbf{b} \in \mathbb{R}^n, \boldsymbol{\beta} \in \mathbb{R}^p$$

Utilisation de la pseudo-inverse

$$\begin{aligned} J(\boldsymbol{\beta}) &= \frac{1}{2} (\mathbf{A}\boldsymbol{\beta} - \mathbf{b})^T (\mathbf{A}\boldsymbol{\beta} - \mathbf{b}) \\ &= \frac{1}{2} \boldsymbol{\beta}^T (\mathbf{A}^T \mathbf{A}) \boldsymbol{\beta} - \boldsymbol{\beta}^T \mathbf{A}^T \mathbf{b} + \frac{1}{2} \mathbf{b}^T \mathbf{b}. \end{aligned}$$

Donc

$$\frac{\partial J(\boldsymbol{\beta})}{\partial \boldsymbol{\beta}} = (\mathbf{A}^T \mathbf{A}) \boldsymbol{\beta} - \mathbf{A}^T \mathbf{b}$$

Si $n \geq p$ et $\text{rang}(\mathbf{A}) = p$, alors

$$\hat{\boldsymbol{\beta}}_{\text{OLS}} = (\mathbf{A}^T \mathbf{A})^{-1} \mathbf{A}^T \mathbf{b} = \mathbf{A}^+ \mathbf{b}$$

où $\mathbf{A}^+ = (\mathbf{A}^T \mathbf{A})^{-1} \mathbf{A}^T$ est la matrice **pseudo inverse** de \mathbf{A} .

Remarque : $\mathbf{P} = \mathbf{A}(\mathbf{A}^T \mathbf{A})^{-1} \mathbf{A}^T$ est une matrice de projection (elle vérifie $\mathbf{P}^2 = \mathbf{P}$ et $\mathbf{P}^T = \mathbf{P}$). Le terme $\mathbf{A}\hat{\boldsymbol{\beta}}_{\text{OLS}}$ peut s'interpréter comme la projection de \mathbf{b} sur l'espace engendré par \mathbf{A} (espace image de \mathbf{A}) noté souvent $R(\mathbf{A})$.

Interprétation géométrique

Equations normales

Quand on annule le gradient de $J(\beta)$, on obtient les équations normales

$$(\mathbf{A}^T \mathbf{A})\beta = \mathbf{A}^T \mathbf{b}.$$

On peut retrouver ce résultat en remarquant qu'on cherche le point β tel que $\mathbf{A}\beta$ est le plus proche de \mathbf{b} . La solution de ce problème est la projection orthogonale de \mathbf{b} sur l'espace image de \mathbf{A} noté $R(\mathbf{A})$ qui est telle que $\mathbf{A}\beta - \mathbf{b}$ est orthogonal à $R(\mathbf{A})$:

$$\langle \mathbf{A}\mathbf{w}, \mathbf{A}\beta - \mathbf{b} \rangle = 0, \forall \mathbf{w} \Leftrightarrow \langle \mathbf{w}, \mathbf{A}^T (\mathbf{A}\beta - \mathbf{b}) \rangle = 0, \forall \mathbf{w}$$

ce qui conduit à

$$\mathbf{A}^T (\mathbf{A}\beta - \mathbf{b}) = 0 \Leftrightarrow (\mathbf{A}^T \mathbf{A})\beta = \mathbf{A}^T \mathbf{b}.$$

Utilisation de la décomposition en valeurs singulières (SVD)

SVD de la matrice \mathbf{A} de rang p

$$\hat{\beta}_{\text{SVD}} = \left(\sum_{i=1}^p \frac{1}{\sigma_i} \mathbf{u}_i \mathbf{v}_i^T \right) \mathbf{b}$$

avec $\mathbf{A} = \mathbf{U} \mathbf{\Sigma} \mathbf{V}^T = \sum_{i=1}^p \sigma_i \mathbf{u}_i \mathbf{v}_i^T$.

SVD de la matrice $[\mathbf{A}, \mathbf{b}]$ de rang $p + 1$

$$\hat{\beta}_{\text{SVD}} = \frac{-1}{v_{p+1,p+1}} \begin{pmatrix} v_{1,p+1} \\ \vdots \\ v_{p,p+1} \end{pmatrix} \quad (1)$$

avec $[\mathbf{A} \ \mathbf{b}] = \mathbf{U} \mathbf{\Sigma} \mathbf{V}^T = \sum_{k=1}^{p+1} \sigma_k \mathbf{u}_k \mathbf{v}_k^T$.

Remarques

Calcul de la solution des moindres carrés

- ▶ **Détermination de la pseudo-inverse** : pour des problèmes de dimensionnalité réduite. Le conditionnement de la matrice $A^T A$ est proche de $(\text{cond} A)^2$, ce qui peut poser problème pour de grandes valeurs de n et p .
- ▶ **Utilisation de la SVD de A** : solution très précise mais lente en grande dimension.
- ▶ **Factorisation QR de A** : si A est de rang p , on peut décomposer A sous la forme $A = QR$, où Q est une matrice orthogonale et R est une matrice triangulaire supérieure, d'où la solution
 - ▶ Factorisation : $A = QR$
 - ▶ Transformation du problème en : $R\beta = Q^T b$. Résolution simple et rapide car R triangulaire supérieure.

Avantage : plus rapide que la SVD.

Preuve de (1)

Si le rang de la matrice $[\mathbf{A} \ \mathbf{b}]$ est $p + 1$, alors sa SVD est $[\mathbf{A} \ \mathbf{b}] = \sum_{k=1}^{p+1} \sigma_k \mathbf{u}_k \mathbf{v}_k^T$. Donc, en posant $\boldsymbol{\beta}' = \begin{pmatrix} \beta \\ -1 \end{pmatrix}$ et $[\mathbf{A}' \ \mathbf{b}'] = \sum_{k=1}^p \sigma_k \mathbf{u}_k \mathbf{v}_k^T$, on a

$$\begin{aligned} J(\boldsymbol{\beta}) &= \frac{1}{2} \|[\mathbf{A} \ \mathbf{b}]\boldsymbol{\beta}'\|^2, \\ &= \frac{1}{2} \|[\mathbf{A}' \ \mathbf{b}']\boldsymbol{\beta}'\|^2 + \frac{1}{2} \|\sigma_{p+1} \mathbf{u}_{p+1} \mathbf{v}_{p+1}^T \boldsymbol{\beta}'\|^2, \\ &\geq \frac{1}{2} \|\sigma_{p+1} \mathbf{u}_{p+1} \mathbf{v}_{p+1}^T \boldsymbol{\beta}'\|^2, \end{aligned}$$

avec égalité si

$$[\mathbf{A}' \ \mathbf{b}']\boldsymbol{\beta}' = \mathbf{0}.$$

Il suffit donc de chercher un vecteur $\boldsymbol{\beta}'$ de la forme $\begin{pmatrix} \beta \\ -1 \end{pmatrix}$ appartenant au noyau de la matrice $[\mathbf{A}' \ \mathbf{b}']$. Comme par construction les vecteurs singuliers sont orthogonaux, $\mathbf{v}_{p+1} = [v_{1,p+1}, \dots, v_{p+1,p+1}]^T$ appartient à ce noyau et donc une solution est

$$\hat{\boldsymbol{\beta}}_{\text{SVD}} = \frac{-1}{v_{p+1,p+1}} \begin{pmatrix} v_{1,p+1} \\ \vdots \\ v_{p,p+1} \end{pmatrix}$$

Comme la solution du problème des moindres carrés ordinaires est unique pour $n \geq p$ et $\text{rang}(\mathbf{A}) = p$, cette solution est identique à celle obtenue avec la pseudo-inverse de \mathbf{A} .

Rappels sur la décomposition en valeurs singulières (SVD)

Soit $\mathbf{X} \in \mathbb{R}^{n \times p}$ avec $n \geq p$, alors la décomposition en valeurs singulières ou SVD de \mathbf{X} est

$$\mathbf{X} = \mathbf{U}\mathbf{\Sigma}\mathbf{V}^T$$

avec

- ▶ $\mathbf{U} \in \mathbb{R}^{n \times n}$ formée de n vecteurs propres orthonormés de $\mathbf{X}^T \mathbf{X}$
- ▶ $\mathbf{V} \in \mathbb{R}^{p \times p}$ formée de p vecteurs propres orthonormés de $\mathbf{X} \mathbf{X}^T$
- ▶ $\mathbf{\Sigma} \in \mathbb{R}^{n \times p}$ est une matrice rectangulaire dont les éléments non nuls sur la diagonale sont les valeurs singulières s_1, \dots, s_p (les racines carrées des valeurs propres positives de $\mathbf{X}^T \mathbf{X}$ ou de $\mathbf{X} \mathbf{X}^T$).

$$\begin{array}{c} \xrightarrow{p} \\ \begin{array}{|c|} \hline \phantom{\mathbf{X}} \\ \hline \end{array} \\ \xleftarrow{n} \end{array} \quad = \quad \begin{array}{c} \xrightarrow{n} \\ \begin{array}{|c|} \hline \phantom{\mathbf{U}} \\ \hline \end{array} \\ \xleftarrow{n} \end{array} \cdot \begin{array}{c} \xrightarrow{p} \\ \begin{array}{|c|} \hline s_1 \\ \vdots \\ s_p \\ 0 \\ \hline \end{array} \\ \xleftarrow{p} \end{array} \cdot \begin{array}{c} \xrightarrow{p} \\ \begin{array}{|c|} \hline \phantom{\mathbf{V}^T} \\ \hline \end{array} \\ \xleftarrow{p} \end{array}$$

$$\mathbf{X} = \mathbf{U} \cdot \mathbf{\Sigma} \cdot \mathbf{V}^T$$

Propriétés de la SVD

► Valeurs singulières de $\mathbf{X}\mathbf{X}^T$

$\text{rang}(\mathbf{X}) = p$ donc les $q - p$ plus petites valeurs propres de $\mathbf{X}\mathbf{X}^T$ vérifient $s_{p+1} = \dots = s_n = 0$.

► SVD de \mathbf{X}^T

Si la SVD de \mathbf{X} est $\mathbf{X} = \mathbf{U}\mathbf{\Sigma}\mathbf{V}^T$, alors $\mathbf{X}^T = \mathbf{V}\mathbf{\Sigma}^T\mathbf{U}$ est la SVD de \mathbf{X}^T .

► Meilleure approximation de rang k de \mathbf{X}

Soit $\mathbf{X} \in \mathbb{R}^{n \times p}$ de SVD $\mathbf{X} = \sum_{i=1}^p \sigma_i \mathbf{u}_i \mathbf{v}_i^T$ avec $p = \text{rang}(\mathbf{X})$. Si $k < p$, alors $\mathbf{X}_k = \sum_{i=1}^k \sigma_i \mathbf{u}_i \mathbf{v}_i^T$ est la meilleure approximation de rang k de \mathbf{X} , i.e.,

$$\|\mathbf{X} - \mathbf{X}_k\|_F = \arg \min_{\text{rang}(\mathbf{D})=k} \|\mathbf{X} - \mathbf{X}_k\|_F$$

► Expression de la pseudo-inverse de \mathbf{X} à l'aide de la SVD

Soit $\mathbf{X} \in \mathbb{R}^{n \times p}$ de SVD $\mathbf{X} = \sum_{i=1}^p \sigma_i \mathbf{u}_i \mathbf{v}_i^T$ avec $p = \text{rang}(\mathbf{X})$ et $n \geq p$, alors

$$(\mathbf{X}^T \mathbf{X})^{-1} \mathbf{X} = \sum_{i=1}^p \frac{1}{\sigma_i} \mathbf{u}_i \mathbf{v}_i^T$$

Plan du cours

Résumé

- ▶ Introduction
- ▶ Moindres carrés ordinaires
- ▶ **Moindres carrés totaux**
- ▶ Exemples

Moindre carrés totaux ou “Orthogonal Distance Regression (ODR)”

Fonction de coût à optimiser

$$J(\beta) = \sum_{i=1}^n \|y_i - f(x_i - \delta_i, \beta)\|^2, \delta_i \in \mathbb{R}^m, \beta \in \mathbb{R}^p, \lambda > 0$$

Cas particulier d'une fonction linéaire

$$y_i = (x_i - \delta_i)^T \beta + q_0^T \beta + q_1 + \epsilon_i$$

avec

$$y_i \in \mathbb{R} (q = 1), x_i \in \mathbb{R}^m, \beta \in \mathbb{R}^m (p = m), q_0 \in \mathbb{R}^m, q_1 \in \mathbb{R}, \epsilon_i \in \mathbb{R}$$

Moindre carrés totaux

Écriture matricielle

$$(\mathbf{A} + \mathbf{E})\boldsymbol{\beta} = \mathbf{b} + \boldsymbol{\epsilon}$$

avec

$$\mathbf{A} = \begin{bmatrix} \mathbf{x}_1^T + \mathbf{q}_0^T \\ \vdots \\ \mathbf{x}_n^T + \mathbf{q}_0^T \end{bmatrix}, \mathbf{E} = \begin{bmatrix} -\boldsymbol{\delta}_1^T \\ \vdots \\ -\boldsymbol{\delta}_n^T \end{bmatrix}, \mathbf{b} = \begin{bmatrix} y_1 - q_1 \\ \vdots \\ y_n - q_1 \end{bmatrix}, \boldsymbol{\epsilon} = \begin{bmatrix} -\epsilon_1 \\ \vdots \\ -\epsilon_n \end{bmatrix}$$

Problème à résoudre

$$\min \|\mathbf{E} \boldsymbol{\epsilon}\|^2$$

avec

$$(\mathbf{A} + \mathbf{E})\boldsymbol{\beta} = \mathbf{b} + \boldsymbol{\epsilon}, \quad \mathbf{A}, \mathbf{E} \in \mathbb{R}^{n \times p}, \boldsymbol{\epsilon} \in \mathbb{R}^n, \mathbf{b} \in \mathbb{R}^p$$

La résolution de ce problème utilise la SVD de $[\mathbf{A} \ \mathbf{b}]$.

Plan du cours

Résumé

- ▶ Introduction
- ▶ Moindres carrés ordinaires
- ▶ Moindres carrés totaux
- ▶ Exemples

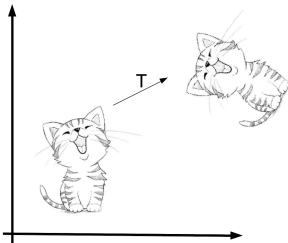
Exemple 1 : transformation affine du plan

► Problème

On cherche une transformation du plan notée T permettant de transformer des points $(u_i, v_i) \in \mathbb{R}^2$ en d'autres points $(u'_i, v'_i) \in \mathbb{R}^2$. Ce problème peut se ramener à déterminer les vecteurs $\mathbf{a} = (a_1, a_2, a_3)^T$ et $\mathbf{b} = (b_1, b_2, b_3)^T$ tels que

$$\begin{pmatrix} u'_i \\ v'_i \end{pmatrix} \approx \begin{pmatrix} a_1 & a_2 \\ b_1 & b_2 \end{pmatrix} \begin{pmatrix} u_i \\ v_i \end{pmatrix} + \begin{pmatrix} a_3 \\ b_3 \end{pmatrix}, \quad i = 1, \dots, n$$

► Illustration



Exemple 1 : transformation affine du plan

► Questions

1. Montrer que ce problème peut se formuler à l'aide d'un problème de régression linéaire $\arg \min_{\beta} \|\mathbf{A}\beta - \mathbf{b}\|^2$ avec $\beta = (a_1, a_2, a_3, b_1, b_2, b_3)^T$ et des matrices \mathbf{A} et \mathbf{b} que l'on précisera. On construira la matrice \mathbf{A} de manière à obtenir une matrice diagonale par blocs.
2. Quelle est la solution de ce problème ?

Exemple 1 : transformation affine du plan

► Réponses

1. On a

$$\mathbf{y}_i = \begin{pmatrix} u'_i \\ v'_i \end{pmatrix} = f(\mathbf{x}_i, \beta) = \mathbf{C}(\mathbf{x}_i)\beta + \mathbf{d}(\mathbf{x}_i)$$

avec

$$\mathbf{x}_i = \begin{pmatrix} u_i \\ v_i \end{pmatrix}, \mathbf{C}(\mathbf{x}_i) = \begin{bmatrix} u_i & v_i & 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & u_i & v_i & 1 \end{bmatrix}, \mathbf{d}(\mathbf{x}_i) = \begin{bmatrix} 0 \\ 0 \end{bmatrix}$$

donc

$$\mathbf{A} = \begin{bmatrix} u_1 & v_1 & 1 & 0 & 0 & 0 \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\ u_n & v_n & 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & u_1 & v_1 & 1 \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\ 0 & 0 & 0 & u_n & v_n & 1 \end{bmatrix} \quad \text{et } \mathbf{b} = \begin{bmatrix} u'_1 \\ \vdots \\ u'_n \\ v'_1 \\ \vdots \\ v'_n \end{bmatrix}$$

Exemple 1 : transformation affine du plan

► Réponses

1. voir slide précédent
2. En posant

$$\mathbf{A} = \begin{bmatrix} \mathbf{A}_1 & \mathbf{0} \\ \mathbf{0} & \mathbf{A}_1 \end{bmatrix}, \quad \mathbf{b} = \begin{bmatrix} \mathbf{b}_1 \\ \mathbf{b}_2 \end{bmatrix}, \quad \mathbf{b}_1 = \begin{bmatrix} u'_1 \\ \vdots \\ u'_n \end{bmatrix}, \quad \mathbf{b}_2 = \begin{bmatrix} v'_1 \\ \vdots \\ v'_n \end{bmatrix}$$

on obtient

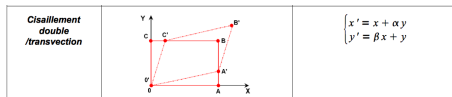
$$\boldsymbol{\beta} = \begin{bmatrix} (\mathbf{A}_1^T \mathbf{A}_1)^{-1} \mathbf{A}_1 \mathbf{b}_1 \\ (\mathbf{A}_1^T \mathbf{A}_1)^{-1} \mathbf{A}_1 \mathbf{b}_2 \end{bmatrix}$$

Exemple 2 : transformation géométrique 2D (recalage d'images)

► Illustration



Fig 1 - Recalage géométrique d'une image



► Modélisation

$$\mathbf{u}' = \begin{pmatrix} x' \\ y' \end{pmatrix} = f(\mathbf{u}, \gamma) \in \mathbb{R}^2 \text{ avec } \mathbf{u} = \begin{pmatrix} x \\ y \end{pmatrix}, \gamma = \begin{pmatrix} \alpha \\ \beta \end{pmatrix}$$

Exemple 2 : transformation géométrique

► Questions

1. Écrire le problème d'optimisation permettant d'estimer le vecteur γ en considérant un ensemble de n points transformés par la même transvection
2. Quelle est la solution de ce problème ?

Exemple 2 : transformation géométrique

► Réponses

1. On a

$$\mathbf{u}' = f(\mathbf{u}, \gamma) = \begin{bmatrix} y & 0 \\ 0 & x \end{bmatrix} \begin{bmatrix} \alpha \\ \beta \end{bmatrix} + \begin{bmatrix} x \\ y \end{bmatrix}$$

donc en considérant n points, on a

$$\begin{bmatrix} x'_1 \\ \vdots \\ x'_n \\ y'_1 \\ \vdots \\ y'_n \end{bmatrix} = \begin{bmatrix} y_1 & 0 \\ \vdots & \vdots \\ y_n & 0 \\ 0 & x_1 \\ \vdots & \vdots \\ 0 & x_n \end{bmatrix} \begin{bmatrix} \alpha \\ \beta \end{bmatrix} + \begin{bmatrix} x_1 \\ \vdots \\ x_n \\ y_1 \\ \vdots \\ y_n \end{bmatrix}$$

soit

$$\mathbf{b} = \begin{bmatrix} x'_1 - x_1 \\ \vdots \\ x'_n - x_n \\ y'_1 - y_1 \\ \vdots \\ y'_n - y_n \end{bmatrix} = \mathbf{A}\gamma$$

Exemple 2 : transformation géométrique

► Réponses

1. Le problème d'optimisation consiste donc à rechercher le vecteur γ qui minimise $\|\mathbf{A}\gamma - \mathbf{b}\|^2$:

$$\min_{\gamma} \frac{1}{2} \|\mathbf{A}\gamma - \mathbf{b}\|^2$$

2. La solution du problème précédent est

$$\hat{\gamma} = (\mathbf{A}^T \mathbf{A})^{-1} \mathbf{A}^T \mathbf{b} = \mathbf{A}^+ \mathbf{b}.$$

En utilisant la structure particulière de \mathbf{A} , on obtient

$$\mathbf{A}^T \mathbf{A} = \begin{bmatrix} \sum_{i=1}^n y_i^2 & 0 \\ 0 & \sum_{i=1}^n x_i^2 \end{bmatrix}$$

En supposant $\sum_{i=1}^n x_i^2 \neq 0$ et $\sum_{i=1}^n y_i^2 \neq 0$, on en déduit

$$(\mathbf{A}^T \mathbf{A})^{-1} = \begin{bmatrix} \frac{1}{\sum_{i=1}^n y_i^2} & 0 \\ 0 & \frac{1}{\sum_{i=1}^n x_i^2} \end{bmatrix}$$

Exemple 2 : transformation géométrique

► Réponses

- 1.
2. De plus

$$\mathbf{A}^T \mathbf{b} = \begin{bmatrix} y_1 & \dots & y_n & 0 & \dots & 0 \\ 0 & \dots & 0 & x_1 & \dots & x_n \end{bmatrix} \begin{bmatrix} x'_1 - x_1 \\ \vdots \\ x'_n - x_n \\ y'_1 - y_1 \\ \vdots \\ y'_n - y_n \end{bmatrix} = \begin{bmatrix} \sum_{i=1}^n (x'_i - x_i) y_i \\ \sum_{i=1}^n (y'_i - y_i) x_i \end{bmatrix}$$

d'où

$$\hat{\gamma} = \begin{bmatrix} \hat{\alpha} \\ \hat{\beta} \end{bmatrix} = \begin{bmatrix} \frac{1}{\sum_{i=1}^n y_i^2} & 0 \\ 0 & \frac{1}{\sum_{i=1}^n x_i^2} \end{bmatrix} \begin{bmatrix} \sum_{i=1}^n (x'_i - x_i) y_i \\ \sum_{i=1}^n (y'_i - y_i) x_i \end{bmatrix} = \begin{bmatrix} \frac{\sum_{i=1}^n (x'_i - x_i) y_i}{\sum_{i=1}^n y_i^2} \\ \frac{\sum_{i=1}^n (y'_i - y_i) x_i}{\sum_{i=1}^n x_i^2} \end{bmatrix}$$

Que faut-il savoir ?

- ▶ Modéliser un problème d'estimation de paramètres à l'aide d'un **problème des moindres carrés ordinaires**
- ▶ Déterminer la solution d'un problème des moindres carrés ordinaires à l'aide du calcul d'une **pseudo inverse** ou d'une **SVD**
- ▶ Connaître l'existence de la **méthode des moindres carrés totaux** appelée **Total Least Squares** ou **Orthogonal Distance Regression (ODR)**.

Pour aller plus loin

Régression robuste

Pour combattre la présence d'éléments aberrants ("outliers"), on peut utiliser des méthodes de **régression robuste**. Par exemple, dans le cas où la variable de sortie $y_i \in \mathbb{R}$ est univariée (ou unidimensionnelle), on peut citer les méthodes suivantes

- ▶ **L_1 -norm regression or Least Absolute Deviation Regression**

$$\arg \min_{\beta} |y_i - f(\mathbf{x}_i, \beta)|$$

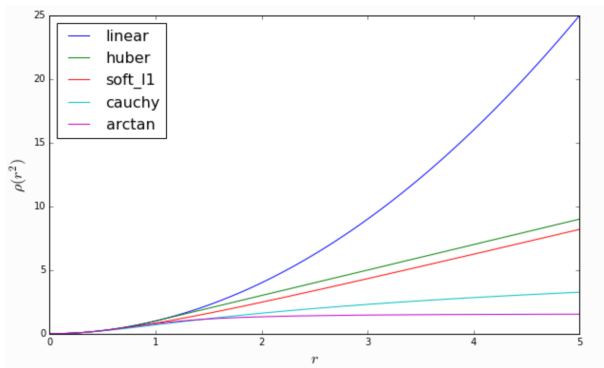
correspond à une loi de Laplace pour le bruit ϵ_i tel que $y_i = f(\mathbf{x}_i, \beta) + \epsilon_i$.

- ▶ **M Estimators**

$$\arg \min_{\beta} \rho(y_i - f(\mathbf{x}_i, \beta))$$

où ρ est une fonction qui pénalise les fortes valeurs.

Exemples de fonctions ρ



Exemple de résultats (M estimateur avec fonction “soft ℓ_1 ”)

