Chapitre 13

La méthode des moindres carrés

13.1 Introduction

La méthode des *moindres carrés* (également appelée *régression linéaire*, **linear regression** (lien web), ou *least squares* en anglais) est une technique qui permet de modéliser des données expérimentales à l'aide d'un modèle linéaire optimal (dans un sens que nous préciserons). Elle est utilisée dans beaucoup de domaines, et constitue en particulier un des piliers des méthodes de base rencontrées en *machine learning*.

De notre point de vue, la méthode des moindres carrés sera une application de l'algèbre linéaire à des problèmes d'optimisation.

Avant de la décrire en toute généralité, nous allons la motiver sur un exemple simple, de petite dimension, qui nous permettra de comprendre l'idée de base, qui sera ensuite généralisée.

Celcius vs Fahrenheit ?

Supposons que l'on souhaite étudier la relation permettant de convertir les unités de mesure d'une température, de **Celcius** (notée T_C) en **Fahrenheit** (notée T_F).

On se souvient que cette relation est du type suivant :

$$(t): \qquad T_F = \alpha T_C + \beta\,,$$

mais on ne se souvient plus des valeurs de α et β.

Si on dispose de deux thermomètre, un qui mesure en Celcius, l'autre en Fahrenheit, on peut prendre des mesures et les utiliser pour essayer de retrouver les valeurs des coefficients α et β. Si ces thermomètres permettaient de faire des mesures "parfaites", il suffirait de faire deux mesures de températures assez différentes, $(T_C^{(1)}, T_F^{(1)})$ (au milieu du laboratoire par exemple) et $(T_C^{(2)}, T_F^{(2)})$ (dans le frigo par exemple), de les injecter dans (t),

$$\begin{aligned} \alpha T_C^{(1)} + \beta &= T_F^{(1)} \\ \alpha T_C^{(2)} + \beta &= T_F^{(2)}\,, \end{aligned}$$

et de résoudre ce système pour trouver α et β.

Mais on sait que des mesures empiriques ne sont par définition pas parfaites : un processus de mesure de ce genre peut contenir de multiples sources d'erreur : mauvaise calibration des appareils, minivariations de températures entre les points où la température est mesurée, imprécisions lors de la lecture de la température sur les thermomètres, etc.

Supposons pour simplifier que l'on fasse trois mesures. On les reporte dans un tableau :

T_C	T_F
2	30
12	52
65	147

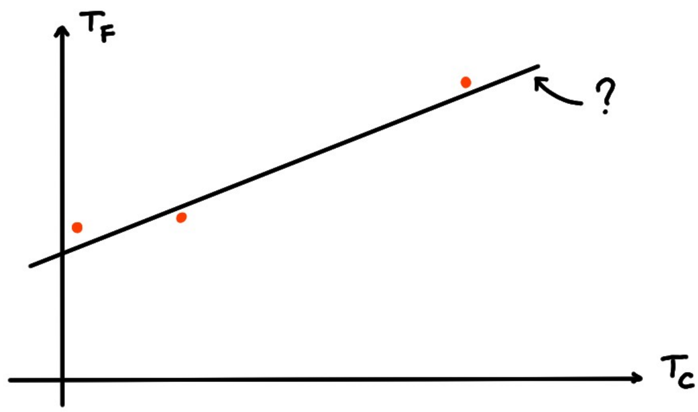
Encore une fois, comme nos mesures ne sont pas exactes, il est très peu probable que les trois points satisfassent simultanément la relation $T_F = \alpha T_C + \beta$, pour des coefficients α, β bien définis. En d'autres termes, le système

$$\begin{cases} 2\alpha & + & \beta & = & 30 \\ 12\alpha & + & \beta & = & 52 \\ 65\alpha & + & \beta & = & 147 \end{cases}$$

est incompatible.

Mais on ne doit pas pour autant abandonner la recherche de la vraie relation qui lie ces températures ! Car si des mesures expérimentales ne permettent pas de retrouver exactement une relation théorique, elles permettent néanmoins de s'en *approcher*.

D'un point de vue graphique, le problème rencontré ci-dessus peut s'exprimer comme suit : les trois paires (T_C, T_F) mesurées en laboratoire peuvent être représentées comme des points dans le plan :



Si ces points ne sont pas sur une même droite, ils doivent quand-même être *proches* de la droite théorique "$T_F = \alpha T_C + \beta$". Et on peut donc se poser la question de savoir si il est possible, à partir de nos trois mesures, de calculer une paire (α_*, β_*) qui donne une droite $T_F = \alpha_* T_C + \beta_*$ qui *approxime au mieux* ce nuage de trois points. Comment définir cette droite ?

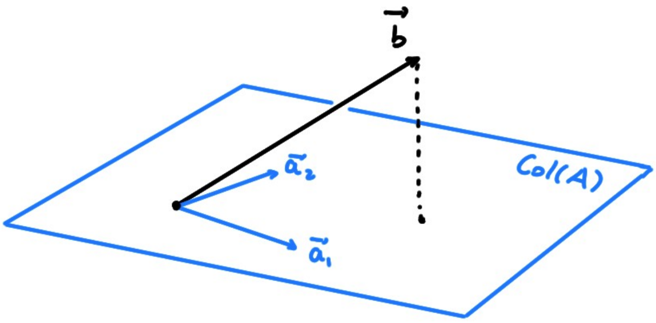
Pour répondre à cette question, utilisons le langage de l'algèbre linéaire pour formuler précisément le problème. On l'a dit, avec nos trois mesures, on est mené au système 3×2

$$\underbrace{\begin{pmatrix} 2 & 1 \\ 12 & 1 \\ 65 & 1 \end{pmatrix}}_{=A} \underbrace{\begin{pmatrix} \alpha \\ \beta \end{pmatrix}}_{=\mathbf{x}} = \underbrace{\begin{pmatrix} 30 \\ 52 \\ 147 \end{pmatrix}}_{=\mathbf{b}},$$

qui est incompatible, et qui le sera en général dès que ces trois mesures sont faites en laboratoire. Il est utile de formuler géométriquement l'absence de solution au problème $A\mathbf{x} = \mathbf{b}$ ci-dessus, en reprenant la définition de base du produit matriciel :

$$\alpha \underbrace{\begin{pmatrix} 2 \\ 12 \\ 65 \end{pmatrix}}_{=:\mathbf{a}_1} + \beta \underbrace{\begin{pmatrix} 1 \\ 1 \\ 1 \end{pmatrix}}_{=:\mathbf{a}_2} = \underbrace{\begin{pmatrix} 30 \\ 52 \\ 147 \end{pmatrix}}_{=:\mathbf{b}}$$

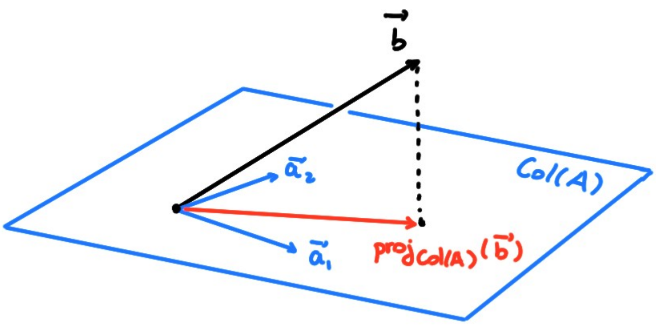
Ce système posséderait une solution (α, β) si $\mathbf{b}$ appartenait à $\mathrm{Col}(A)$, c'est-à-dire au plan engendré par $\mathbf{a}_1$ et $\mathbf{a}_2$. Mais le plus probable est que $\mathbf{b}$ ne soit *pas* dans ce plan :



Cette image suggère que malgré tout, si on ne peut pas trouver de paire telle que la combinaison linéaire $\alpha\mathbf{a}_1 + \beta\mathbf{a}_2$ soit exactement égale à $\mathbf{b}$, on pourrait chercher la paire telle que la combinaison linéaire $\alpha\mathbf{a}_1 + \beta\mathbf{a}_2$ soit *aussi proche que possible de* $\mathbf{b}$, c'est à dire la paire (α, β) qui minimise la distance

$$\|(\alpha\mathbf{a}_1 + \beta\mathbf{a}_2) - \mathbf{b}\| .$$

On sait, par les résultats démontrés dans le chapitre précédent, que la combinaison linéaire qui réalise ce minimum est précisément celle qui est égale à la projection de $\mathbf{b}$ sur l'espace engendré par $\mathbf{a}_1$ et $\mathbf{a}_2$, à savoir $\mathrm{Col}(A)$:



Pour résumer, au lieu de résoudre le système incompatible

$$(*) : \qquad A\mathbf{x} = \mathbf{b}\,,$$

on cherche le $\mathbf{x}$ qui minimise la distance

$$\|A\mathbf{x} - \mathbf{b}\|\,,$$

et on sait que ce $\mathbf{x}$ correspond à la solution de

$$(*)_{MC} : \qquad A\mathbf{x} = \mathrm{proj}_{\mathrm{Col}(A)}(\mathbf{b})\,.$$

Ce deuxième système possède *toujours* une solution $\mathbf{x}$, puisque par définition, la projection $\mathrm{proj}_{\mathrm{Col}(A)}(\mathbf{b}) \in \mathrm{Col}(A)$.

Calculons donc la projection de $\mathbf{b}$ sur $W = \mathrm{Col}(A) = \mathrm{Vect}\{\mathbf{a}_1, \mathbf{a}_2\}$.

Informel 13.1. Attention, les calculs qui suivent sont simples, mais mènent à des fractions que l'on ne peut pas forcément simplifier. Pas grave, c'est la vie ! La plupart du temps, dès qu'on s'attaque à un problème venu d'une situation pratique, il apparaît toujours des nombres moins jolis que ceux qu'on est habitués à trouver dans les séries d'exercices. (Et le plus probable est que l'on implémente l'algorithme sur un ordinateur, donc on ne fera pas à la main ces calculs de fractions.)

Comme les colonnes de A ne sont pas orthogonales, on peut d'abord faire (Gram-Schmidt) :

$$\mathbf{a}_2' := \mathbf{a}_2 - \mathrm{proj}_{\mathbf{a}_1}(\mathbf{a}_2) = \begin{pmatrix}1\\1\\1\end{pmatrix} - \frac{79}{4373}\begin{pmatrix}2\\12\\65\end{pmatrix} = \begin{pmatrix}4215/4373\\3425/4373\\-762/4373\end{pmatrix}.$$

La projection peut maintenant se calculer :

$$\begin{aligned}\mathrm{proj}_{\mathrm{Col}(A)}(\mathbf{b}) &= \frac{\mathbf{b}\cdot\mathbf{a}_1}{\|\mathbf{a}_1\|^2}\mathbf{a}_1 + \frac{\mathbf{b}\cdot\mathbf{a}_2'}{\|\mathbf{a}_2'\|^2}\mathbf{a}_2' \\ &= \frac{10239}{4373}\begin{pmatrix}2\\12\\65\end{pmatrix} + \frac{192536}{30077494}\begin{pmatrix}4215\\3425\\-762\end{pmatrix} \\ &= \begin{pmatrix}31.6644\dots\\50.0215\dots\\147.3140\dots\end{pmatrix}\end{aligned}$$

Maintenant, on peut résoudre le système $A\mathbf{x} = \mathrm{proj}_{\mathrm{Col}(A)}(\mathbf{b})$:

$$\begin{pmatrix}2 & 1\\12 & 1\\65 & 1\end{pmatrix}\begin{pmatrix}\alpha\\\beta\end{pmatrix} = \begin{pmatrix}31.6644\dots\\50.0215\dots\\147.3140\dots\end{pmatrix}$$

On trouve :

$$\begin{aligned}\alpha &= 1.8357\dots\\ \beta &= 27.9930\dots\end{aligned}$$

Donc nos trois mesures, et le raisonnement géométrique menant à projeter sur les colonnes de la matrice A, nous ont mené à la version suivante de la relation entre degrés Celsius et Fahrenheit :

$$T_F = 1.8357\dots T_C + 27.9930\dots\,,$$

Cette droite est celle qui approxime le mieux nos mesures, **au sens des moindre carrés** (voir la section suivante pour l'explication de cette terminologie).

Pour information, la vraie relation, que l'on trouve par exemple **ici** (lien web), est la suivante :

$$T_F = \frac{9}{5}T_C + 32 = 1.8T_C + 32\,.$$

Avec seulement trois points, notre méthode fournit donc des coefficients dont l'erreur avec la relation théorique est d'environ 2% pour α, et 13% pour β.

Informel 13.2. Bien-sûr, on obtiendrait un bien meilleur résultat en faisant beaucoup plus que trois mesures ! Si on faisait 100 mesures par exemple, l'erreur sur α et β serait bien plus petite. Pourtant, on traiterait le problème exactement de la même façon : avec 100 mesures, on devrait considérer un système incompatible

$$\underbrace{A}_{100\times 2}\underbrace{\mathbf{x}}_{\in\mathbb{R}^2} = \underbrace{\mathbf{b}}_{\in\mathbb{R}^{100}}\,.$$

On projetterait alors $\mathbf{b} \in \mathbb{R}^{100}$ sur le plan $\mathrm{Col}(A) \subset \mathbb{R}^{100}$, pour finalement obtenir une droite qui approxime notre nuage, formé par les 100 points des mesures faites en laboratoire.

13.2 Méthode générale

Considérons un système $m \times n$,

$$(*) : \quad A\mathbf{x} = \mathbf{b}\,,$$

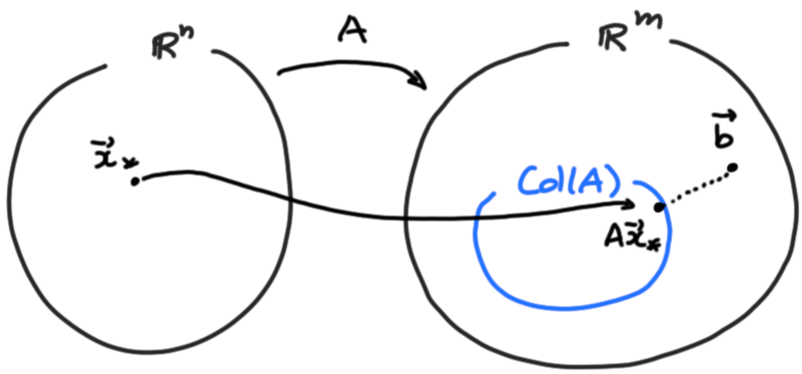
que l'on supposera incompatible, ce qui signifie

$$\min_{\mathbf{x}\in\mathbb{R}^n} \|A\mathbf{x} - \mathbf{b}\| > 0\,.$$

Définition 13.3. On dit que $\mathbf{x}_* \in \mathbb{R}^n$ est une **pseudo-solution de** $(*)$, ou **solution de** $(*)$ **au sens des moindres carrés** si

$$\|A\mathbf{x}_* - \mathbf{b}\| = \min_{\mathbf{x}\in\mathbb{R}^n} \|A\mathbf{x} - \mathbf{b}\|\,.$$

Shématiquement :



Remarque 13.4. À propos de la terminologie “moindres carrés” : Trouver le $\mathbf{x}$ qui minimise une certaine norme revient au même que de trouver le $\mathbf{x}$ qui minimise le *carré* de cette norme , donc la recherche d'une pseudo-solution revient à minimiser la fonction

$$\mathbf{x} \mapsto \|A\mathbf{x} - \mathbf{b}\|^2 = \sum_{k=1}^{m} \big((A\mathbf{x})_k - b_k\big)^2 ,$$

qui est une somme de *carrés*. ◇

Exemple 13.5. (Généralisation du problème de l'exemple de motivation.) Supposons que l'on ait un nuage de points dans le plan, obtenu en prenant des mesures

$$\mathcal{P} = \{(x_1, y_1), \dots, (x_N, y_N)\} ,$$

sensées obéir à une relation affine théorique de la forme

$$y = \alpha x + \beta .$$

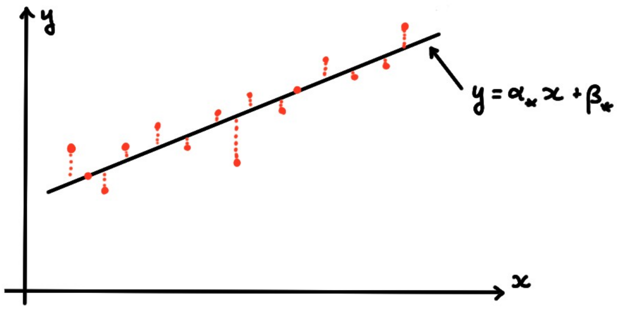
Dans ce cas, le système $N \times 2$

$$\left\{\begin{array}{ccccc} \alpha x_1 & + & \beta & = & y_1 \\ \alpha x_2 & + & \beta & = & y_2 \\ & & & \vdots & \\ \alpha x_N & + & \beta & = & y_N \end{array}\right.$$

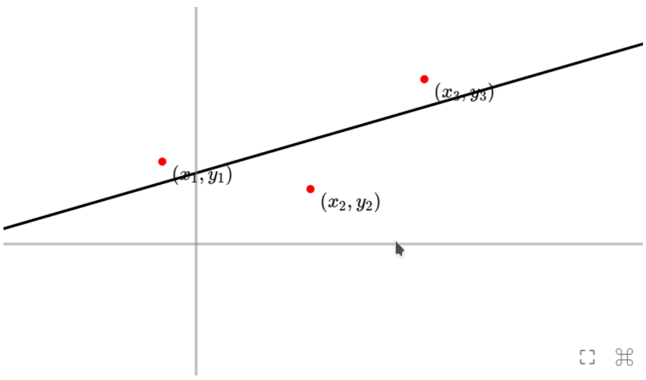
est en général incompatible, et la solution au sens des moindres carrés correspond à minimiser

$$\mathbf{x} = \begin{pmatrix} \alpha \\ \beta \end{pmatrix} \mapsto \sum_{k=1}^{N} \big((\alpha x_k + \beta) - y_k\big)^2 .$$

La paire (α_*, β_*) qui minimise cette fonction fournit donc une droite qui approxime le nuage $\mathcal{P}$, *au sens des moindres carrés* :



Par exemple, avec seulement $N = 3$ (comme dans l'exemple de motivation) :



Pour une animation semblable, mais fonctionnant avec un nombre arbitraire de points, cliquer **ici (Stats applets)** (lien web). ◇

L'équation normale

Considérons une pseudo-solution $\mathbf{x}_*$:

$$\|A\mathbf{x}_* - \mathbf{b}\| = \min_{\mathbf{x}\in\mathbb{R}^n} \|A\mathbf{x} - \mathbf{b}\| .$$

Comme on sait, considérer tous les produits $\mathbf{q} := A\mathbf{x}$ possibles, lorsque $\mathbf{x}$ varie, revient à considérer toutes les combinaisons linéaires possibles des colonnes de A, et donc à parcourir tout le sous-espace $\mathrm{Col}(A)$. Donc on peut tout aussi bien écrire

$$\min_{\mathbf{x}\in\mathbb{R}^n} \|A\mathbf{x} - \mathbf{b}\| = \min_{\mathbf{q}\in\mathrm{Col}(A)} \|\mathbf{q} - \mathbf{b}\| .$$

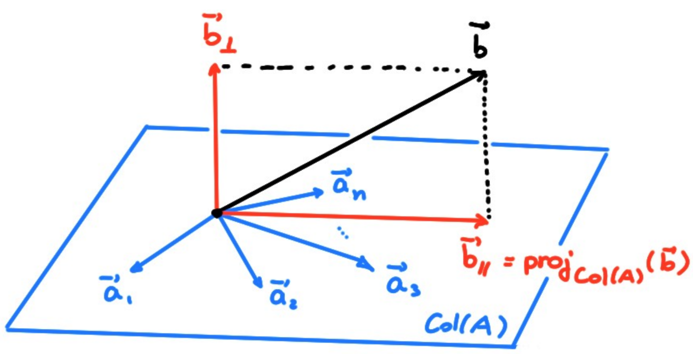
Or on a vu que le minimum de cette distance est réalisé lorsque $\mathbf{q}$ est la projection de $\mathbf{b}$ sur $\mathrm{Col}(A)$:

$$\mathbf{q}_* = \mathrm{proj}_{\mathrm{Col}(A)}(\mathbf{b})$$

On peut toujours calculer cette projection, typiquement en extrayant une base de $\mathrm{Col}(A)$, et en l'orthogonalisant avec le procédé de Gram-Schmidt. (C'est ce que nous avons fait dans l'exemple de l'introduction.)

Mais nous allons voir qu'il est possible de passer outre le calcul explicite de cette projection.

En effet, commençons par rappeler que $\mathbf{b}_\parallel = \mathrm{proj}_{\mathrm{Col}(A)}(\mathbf{b})$ est caractérisé par le fait que $\mathbf{b}_\perp := \mathbf{b} - \mathbf{b}_\parallel$ est orthogonal à $\mathrm{Col}(A)$, i.e. $\mathbf{b}^\perp \in \mathrm{Col}(A)^\perp$:



Or on a **montré** (lien web) précédemment que

$$\mathrm{Col}(A)^\perp = \ker(A^T) .$$

Ainsi, $\mathbf{b}^\perp$ doit satisfaire $A^T\mathbf{b}^\perp = \mathbf{0}$, qui donne

$$A^T(\mathbf{b} - \mathbf{b}_\parallel) = \mathbf{0} ,$$

c'est-à-dire

$$A^T\mathbf{b}_\parallel = A^T\mathbf{b} .$$

Comme $\mathbf{b}_\parallel = A\mathbf{x}_*$, on a démontré :

Théorème 13.6. *Un vecteur $\mathbf{x} \in \mathbb{R}^n$ est solution de $A\mathbf{x} = \mathbf{b}$ au sens des moindres carrés si et seulement si $\mathbf{x}$ est solution de l'**équation normale**, donnée par*

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

Exemple 13.7. Considérons l'exemple de l'introduction, où le système incompatible $A\mathbf{x} = \mathbf{b}$ de départ était

$$\begin{pmatrix} 2 & 1 \\ 12 & 1 \\ 65 & 1 \end{pmatrix} \begin{pmatrix} \alpha \\ \beta \end{pmatrix} = \begin{pmatrix} 30 \\ 52 \\ 147 \end{pmatrix} .$$

Donnons la solution de cette équation sans passer par la projection, en utilisant le théorème ci-dessus. On obtient l'équation normale en multipliant des deux côtés par A^T, qui donne

$$\begin{pmatrix} 2 & 12 & 65 \\ 1 & 1 & 1 \end{pmatrix} \begin{pmatrix} 2 & 1 \\ 12 & 1 \\ 65 & 1 \end{pmatrix} \begin{pmatrix} \alpha \\ \beta \end{pmatrix} = \begin{pmatrix} 2 & 12 & 65 \\ 1 & 1 & 1 \end{pmatrix} \begin{pmatrix} 30 \\ 52 \\ 147 \end{pmatrix} ,$$

c'est-à-dire

$$\begin{pmatrix} 4373 & 79 \\ 79 & 3 \end{pmatrix} \begin{pmatrix} \alpha \\ \beta \end{pmatrix} = \begin{pmatrix} 10239 \\ 229 \end{pmatrix}$$

La solution de ce dernier est donnée par

$$\begin{aligned} \alpha &= \frac{12626}{6878} = 1.8357\dots \\ \beta &= \frac{192536}{6878} = 27.9930\dots , \end{aligned}$$

comme nous avions trouvé en utilisant la projection. ◇

Informel 13.8. Si A est $m \times n$, $A^T A$ est $n \times n$, et donc l'équation normale représente un système carré $n \times n$ qui possède *toujours* une solution.

Dans l'exemple précédent, la solution de l'équation normale était unique. Mais il peut arriver que l'équation normale possède plus d'une solution, ce que l'on aimerait éviter dans les problèmes pratiques. Voyons comment garantir l'unicité de la pseudo-solution :

Théorème 13.9. *Soit A une matrice $m \times n$. Sont équivalents :*

a) Pour tout $\mathbf{b} \in \mathbb{R}^m$, la pseudo-solution de $A\mathbf{x} = \mathbf{b}$ est unique.

b) Pour tout $\mathbf{b} \in \mathbb{R}^m$, la solution de l'équation normale $A^T A\mathbf{x} = A^T\mathbf{b}$ est unique.

c) $A^T A$ (qui est $n \times n$) est inversible.

d) Les colonnes de A sont linéairement indépendantes.

Nous aurons besoin du résultat préliminaire suivant :

Lemme 19. *Pour toute matrice A ($m \times n$), $A\mathbf{x} = \mathbf{0}$ si et seulement si $A^T A\mathbf{x} = \mathbf{0}$.*

Preuve: Il est évident que si $A\mathbf{x} = \mathbf{0}$, alors $A^T A\mathbf{x} = \mathbf{0}$.

Inversément, si $A^T A\mathbf{x} = \mathbf{0}$, alors

$$\|A\mathbf{x}\|^2 = (A\mathbf{x})^T (A\mathbf{x}) = \mathbf{x}^T (A^T A\mathbf{x}) = \mathbf{0} ,$$

ce qui implique $\|A\mathbf{x}\| = 0$, c'est-à-dire $A\mathbf{x} = \mathbf{0}$. □

Passons maintenant à la preuve du théorème :
Preuve: 1. $\Leftrightarrow$ 2. : clair par ce que nous avons montré ci-dessus (un $\mathbf{x}$ est pseudo-solution si et seulement si c'est une solution de l'équation normale).

2. $\Leftrightarrow$ 3. : On sait qu'un système carré $M\mathbf{x} = \mathbf{y}$ possède une unique solution pour tout $\mathbf{y}$ si et seulement si M est inversible.

3. $\Leftrightarrow$ 4. : En effet, les colonnes de A sont indépendantes si et seulement si l'équation $A\mathbf{x} = \mathbf{0}$ ne possède que la solution triviale, et par le lemme ci-dessus, ceci est équivalent à dire que $A^T A\mathbf{x} = \mathbf{0}$ ne possède que la solution triviale, qui encore une fois est équivalent à dire que $A^T A$ est inversible, qui est 2. □

Exemple 13.10. Le système incompatible

$$\begin{pmatrix} 1 & 2 & 3 \\ 1 & -3 & -2 \\ 0 & 5 & 5 \\ -1 & 1 & 0 \end{pmatrix} \begin{pmatrix} x_1 \\ x_2 \\ x_3 \end{pmatrix} = \begin{pmatrix} 1 \\ 2 \\ 3 \\ 4 \end{pmatrix}$$

possède une infinité de pseudo-solutions. En effet, la troisième colonne de A est égale à la somme des deux premières; par le théorème, ceci implique que la solution n'est pas unique. On conclut que le nombre de solutions est infini, par le Théorème "$0, 1, \infty$" appliqué à $A^T A\mathbf{x} = A^T\mathbf{b}$. ◇

Plus tard, nous appliquerons la méthode des moindres carrés pour résoudre d'autres problèmes d'optimisation, inspirés de l'analyse.

13.3 Utilisation de la décomposition QR

La décomposition QR intervient dans la recherche des solutions d'un système au sens des moindres carrés.

En effet, considérons un système incompatible

$$A\mathbf{x} = \mathbf{b}\,.$$

Théorème 13.11. *Soit A une matrice $m \times n$ quelconque, et soit $A = QR$ une décomposition QR de A. Alors un vecteur $\mathbf{x} \in \mathbb{R}$ est la solution de $A\mathbf{x} = \mathbf{b}$ au sens des moindres carrés si et seulement si il est solution du système*

$$R\mathbf{x} = Q^T\mathbf{b}\,.$$

Remarque 13.12. L'avantage du système $R\mathbf{x} = Q^T\mathbf{b}$ est qu'il est *triangulaire*. ◇

Preuve: Supposons d'abord que $\mathbf{x}$ est pseudo-solution de $A\mathbf{x} = \mathbf{b}$. On sait que cela signifie que

$$\mathbf{b} - A\mathbf{x} \in \mathrm{Col}(A)^\perp = \mathrm{Col}(Q)^\perp = \ker(Q^T)\,.$$

Dans la première égalité, on a utilisé le fait que les colonnes de Q, par définition, engendrent le même sous-espace que celles de A.

On a donc

$$Q^T(\mathbf{b} - A\mathbf{x}) = \mathbf{0}\,,$$

et comme $Q^T A = R$, cette dernière implique que $\mathbf{x}$ est solution de

$$R\mathbf{x} = Q^T\mathbf{b}\,.$$

Inversément, supposons que $\mathbf{x}$ est solution de ce dernier système, que l'on écrit plutôt

$$Q^T A\mathbf{x} = Q^T\mathbf{b}\,.$$

En multipliant des deux côtés par R^T et en utilisant $R^T Q^T = (QR)^T = A^T$, on obtient

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

donc $\mathbf{x}$ est solution de l'équation normale. □

Exemple 13.13. Considérons le système $A\mathbf{x} = \mathbf{b}$ incompatible suivant :

$$\begin{pmatrix} 2 & 1 \\ 2 & 0 \\ 1 & 1 \end{pmatrix} \begin{pmatrix} x_1 \\ x_2 \end{pmatrix} = \begin{pmatrix} 1 \\ -5 \\ 2 \end{pmatrix}.$$

Puisque les colonnes de A sont indépendantes, la solution au sens des moindres carrés est unique, et on va la calculer en utilisant le théorème ci-dessus.

Le procédé de Gram-Schmidt appliqué aux colonnes de A, suivi d'une normalisation, donne

$$Q = \begin{pmatrix} 2/3 & 1/3 \\ 2/3 & -2/3 \\ 1/3 & 2/3 \end{pmatrix}.$$

On a donc

$$R = Q^T A = \begin{pmatrix} 3 & 1 \\ 0 & 1 \end{pmatrix}, \qquad \text{et} \qquad Q^T \mathbf{b} = \begin{pmatrix} -2 \\ 5 \end{pmatrix}.$$

Ainsi, $R\mathbf{x} = Q^T \mathbf{b}$ est le système triangulaire donné par

$$\begin{pmatrix} 3 & 1 \\ 0 & 1 \end{pmatrix} \begin{pmatrix} x_1 \\ x_2 \end{pmatrix} = \begin{pmatrix} -2 \\ 5 \end{pmatrix}.$$

La solution de ce dernier est $x_1 = -7/3$, $x_2 = 5$. On peut bien-sûr vérifier que cette solution est la même que celle de l'équation normale associée au système incompatible initial. ◇