La méthode QR pour le problème des moindres carrés

Leçons 160,233

Théorème (Méthode QR et problème des moindres carrés)

1. Soient $m, n \in \mathbb{N}^*$ avec $n < m$. Soit $A \in \mathcal{M}_{m,n}(\mathbb{R})$ avec $rg(A) = n$. Alors ils existent $Q \in \mathcal{O}_m(\mathbb{R})$ et $R \in GL_n(\mathbb{R})$ telles que:
$$A = Q \begin{bmatrix} R \\ 0_{m-n,n} \end{bmatrix}$$

2. Soit $B \in \mathbb{R}^m$. Alors le problème suivant, dit des moindres carrés:
$$\text{Trouver } X \in \mathbb{R}^n \text{ tel que } \quad \|AX - B\| = \underset{Y\in\mathbb{R}^n}{Inf} \|AY - B\|$$
admet une unique solution, donnée par $X = R^{-1}C_1$, où $Q^T B = \begin{bmatrix} C_1 \\ C_2 \end{bmatrix}$ avec $C_1 \in \mathbb{R}^n$ et $C_2 \in \mathbb{R}^{m-n}$ (composantes 1 à n, puis $n+1$ à m)

Remarques. *1. $\|\cdot\|$ et $\langle\cdot|\cdot\rangle$ désignent respetivement la norme eucidienne et le produit scalaire canonique.*

*2. Pour tout vecteur $V \in \mathbb{R}^m$, on note $V^{(j)}$ sa $j-$ième coordonnée**

Voici le plan de la démonstration:

1. Étudier la décomposition QR sur une matrice $\widetilde{A} \in \mathcal{M}_m(\mathbb{R})$ (que l'on précisera) dont les colonnes forment une base de $\mathbb{R}^m$.

2. Utiliser cette décomposition pour en déduire X

Démonstration. *1. Comme $rg(A) = n$, les colonnes de $A = \begin{bmatrix} U_1 & \dots & U_n \end{bmatrix}$ forment une famille libre de $\mathbb{R}^m$. Par le théorème de la base incomplète, on peut compléter la famille $\{U_1, \dots, U_n\}$ en une base $\{U_1, \dots, U_n, U_{n+1}, \dots, U_m\}$ de $\mathbb{R}^m$. Posons alors $\widetilde{A} = \begin{bmatrix} U_1 & \dots & U_n & U_{n+1} & \dots & U_m \end{bmatrix}$.*

On orthonormalise ensuite la base $\{U_1, \dots, U_m\}$ via l'algorithme de Gram-Schmidt. On pose $E_1 = \frac{U_1}{\|U_1\|}$, puis, pour tout $k \in [\![2, n]\!]$,
$$E_k = \frac{U_k - \sum_{j=1}^{k-1} \langle U_j | E_j \rangle E_j}{\left\| U_k - \sum_{j=1}^{k-1} \langle U_j | E_j \rangle E_j \right\|}$$

On obtient ainsi:

$$\left\{\begin{array}{ccccccccc} U_1 & = & r_{1,1}E_1 & & & & & & \\ U_2 & = & r_{2,1}E_1 & + & r_{2,2}E_2 & & & & \\ \vdots & & \vdots & & \vdots & & \ddots & & \\ U_m & = & r_{m,1}E_1 & + & r_{m,2}E_2 & + & \dots & + & r_{m,m}E_m \end{array}\right.$$

avec, pour tout $j \in [\![1,m]\!]$, $r_{j,j} \neq 0$.

Posons alors $Q = \left[\begin{array}{cccc} E_1 & E_2 & \dots & E_m \end{array}\right] \in \mathcal{O}_m(\mathbb{R})$
ainsi que:

$$\widetilde{R} = \left[\begin{array}{cccc} r_{1,1} & r_{2,1} & \dots & r_{m,1} \\ & r_{2,2} & \dots & r_{m,2} \\ & & \ddots & \vdots \\ (0) & & & r_{m,m} \end{array}\right] \in GL_m(\mathbb{R})$$

On a ainsi:

$$\begin{aligned} Q\widetilde{R} &= \left[\begin{array}{cccc} E_1^{(1)} & E_2^{(1)} & \dots & E_m^{(1)} \\ E_1^{(2)} & E_2^{(2)} & \dots & E_m^{(2)} \\ \vdots & \vdots & \ddots & \vdots \\ E_1^{(m)} & E_2^{(m)} & \dots & E_m^{(m)} \end{array}\right] \left[\begin{array}{cccc} r_{1,1} & r_{2,1} & \dots & r_{m,1} \\ & r_{2,2} & \dots & r_{m,2} \\ & & \ddots & \vdots \\ (0) & & & r_{m,m} \end{array}\right] \\ &= \left[\begin{array}{cccc} r_{1,1}E_1^{(1)} & r_{2,1}E_1^{(1)} + r_{2,2}E_2^{(1)} & \dots & r_{m,1}E_1^{(1)} + r_{m,2}E_2^{(1)} + \dots + r_{m,m}E_m^{(1)} \\ r_{1,1}E_1^{(2)} & r_{2,1}E_1^{(2)} + r_{2,2}E_2^{(2)} & \dots & r_{m,1}E_1^{(2)} + r_{m,2}E_2^{(2)} + \dots + r_{m,m}E_m^{(2)} \\ \vdots & \vdots & \ddots & \vdots \\ r_{1,1}E_1^{(m)} & r_{2,1}E_1^{(m)} + r_{2,2}E_2^{(m)} & \dots & r_{m,1}E_1^{(m)} + r_{m,2}E_2^{(m)} + \dots + r_{m,m}E_m^{(m)} \end{array}\right] \\ &= \left[\begin{array}{cccc} r_{1,1}E_1 & r_{2,1}E_1 + r_{2,2}E_2 & \dots & r_{m,1}E_1 + r_{m,2}E_2 + \dots + r_{m,m}E_m \end{array}\right] \\ &= \left[\begin{array}{cccc} U_1 & U_2 & \dots & U_m \end{array}\right] \end{aligned}$$

On a donc construit nos matrices Q *et* $\widetilde{R}$ *telles que:*

$$\widetilde{A} = Q\widetilde{R} \tag{1}$$

Posons la matrice:

$$R = \left[\begin{array}{cccc} r_{1,1} & r_{2,1} & \dots & r_{n,1} \\ & r_{2,2} & \dots & r_{n,2} \\ & & \ddots & \vdots \\ (0) & & & r_{n,n} \end{array}\right] \in GL_n(\mathbb{R})$$

La matrice $\left[\begin{array}{c} R \\ 0_{m-n,n} \end{array}\right] \in \mathcal{M}_{m,n}(\mathbb{R})$ *est ainsi constituée des n premières colonnes de* $\widetilde{R}$. *On obtient ainsi:*

$$
\begin{aligned}
Q \begin{bmatrix} R \\ 0_{m-n,n} \end{bmatrix} &= \begin{bmatrix} U_1 & \dots & U_n \end{bmatrix} \\
&= A
\end{aligned}
$$

ce qui montre la première partie du théorème

2. *Soit $Y \in \mathbb{R}^n$. On a:*

$$
\begin{aligned}
||AY - B||^2 &= \left|\left| Q \begin{bmatrix} R \\ 0_{m-n,n} \end{bmatrix} - B \right|\right|^2 \\
&\overset{Q \in \mathcal{O}_m(\mathbb{R})}{=} \left|\left| \begin{bmatrix} R \\ 0_{m-n,n} \end{bmatrix} - Q^T B \right|\right|^2 \\
&= \left|\left| \begin{bmatrix} RY \\ 0_{\mathbb{R}^{m-n}} \end{bmatrix} - \begin{bmatrix} C_1 \\ C_2 \end{bmatrix} \right|\right|^2 \\
&= \left|\left| \begin{bmatrix} RY - C_1 \\ 0_{\mathbb{R}^{m-n}} \end{bmatrix} - \begin{bmatrix} 0_{\mathbb{R}^n} \\ C_2 \end{bmatrix} \right|\right|^2 \qquad (2) \\
&= ||RY - C_1||^2 + ||C_2||^2 \qquad (3)
\end{aligned}
$$

Ainsi, le vecteur X minimisant la quantité (3) est donné par $X = R^{-1}C_1$, ce qui a bien un sens car R est inversible.

■

Remarques. 1. *Dans la décomposition QR donnée par (1), on a montré l'existence, mais on a également unicité du couple $(Q, \widetilde{R})$. En effet, soient $Q_1, Q_2 \in \mathcal{O}_m(\mathbb{R})$ et $\widetilde{R}_1, \widetilde{R}_2 \in Gl_m(\mathbb{R})$ triangulaires supérieures telles que $A = Q_1\widetilde{R}_1 = Q_2\widetilde{R}_2$. La matrice $Q_1^T Q_2 = \widetilde{R}_2 \widetilde{R}_1^{-1}$ est donc à la fois orthogonale et triangulaire supérieure, donc vaut l'identité, ce qui donne l'unicité.*

2. *Le passage de (2) à (3) utilise la décomposition de $\mathbb{R}^m$ en somme directe de deux sous espaces orthogonaux: $\mathbb{R}^m = \mathbb{R}^n \overset{\perp}{\oplus} \mathbb{R}^{m-n}$, puis le théorème de Pythagore. Ainsi, les vecteurs $RY - C_1$ et C_2 sont respectivement dans $\mathbb{R}^n$ et $\mathbb{R}^{m-n}$, et la norme euclidienne $\|\cdot\|$ de (2) est celle sur $\mathbb{R}^m$ alors que celle dans (3) est celle sur $\mathbb{R}^{m-n}$ ou $\mathbb{R}^n$*

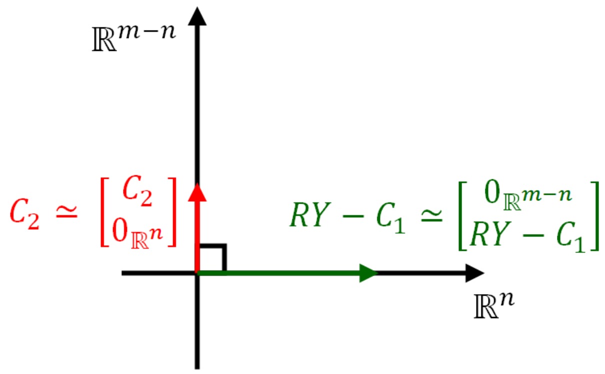

Figure 1: Illustration du passage de (2) à (3). Le symbole $\simeq$ indique un abus de notation mais permet de visualiser les choses