TD 7 – Projection orthogonale et factorisation QR

Exercice 1 – Projection orthogonale sur une droite

Si $\boldsymbol{a} = 0$, la fonction est constante et tout α est minimiseur. On suppose donc $\boldsymbol{a} \neq 0$ pour obtenir l'unicité.

1. Par bilinéarité du produit scalaire,

$$\varphi(\alpha) = \frac{1}{2}\langle \alpha\boldsymbol{a} - \boldsymbol{y}, \alpha\boldsymbol{a} - \boldsymbol{y}\rangle = \frac{1}{2}\big(\alpha^2 \|\boldsymbol{a}\|_2^2 - 2\alpha\langle \boldsymbol{a}, \boldsymbol{y}\rangle + \|\boldsymbol{y}\|_2^2\big).$$

D'où

$$\varphi'(\alpha) = \alpha\|\boldsymbol{a}\|_2^2 - \langle \boldsymbol{a}, \boldsymbol{y}\rangle, \qquad \varphi''(\alpha) = \|\boldsymbol{a}\|_2^2 > 0.$$

Ainsi φ est strictement convexe : son unique point critique est son minimum global.

2. En annulant la dérivée,

$$\varphi'(\alpha^\star) = 0 \iff \alpha^\star\|\boldsymbol{a}\|_2^2 = \langle \boldsymbol{a}, \boldsymbol{y}\rangle \iff \alpha^\star = \frac{\langle \boldsymbol{a}, \boldsymbol{y}\rangle}{\|\boldsymbol{a}\|_2^2}.$$

3. On en déduit

$$\boldsymbol{y}^\star = \alpha^\star\boldsymbol{a} = \left\langle \boldsymbol{y}, \frac{\boldsymbol{a}}{\|\boldsymbol{a}\|_2}\right\rangle \frac{\boldsymbol{a}}{\|\boldsymbol{a}\|_2}.$$

De plus,

$$\langle \boldsymbol{y} - \boldsymbol{y}^\star, \boldsymbol{a}\rangle = \langle \boldsymbol{y}, \boldsymbol{a}\rangle - \alpha^\star\|\boldsymbol{a}\|_2^2 = 0 \quad \Longrightarrow \quad \boldsymbol{y} - \boldsymbol{y}^\star \perp \mathrm{Vect}(\boldsymbol{a}).$$

Comme $\boldsymbol{y}^\star \in \mathrm{Vect}(\boldsymbol{a})$, c'est la projection orthogonale cherchée.

Exercice 2 – Projection sur un sous-espace vectoriel

Minimiser $\|\boldsymbol{z} - \boldsymbol{y}\|_2$ ou son carré revient au même, car la fonction $t \mapsto t^2$ est croissante sur $[0, +\infty[$.

1. Soit $\boldsymbol{z}^\star \in F$ un minimiseur. Pour $\boldsymbol{d} \in F$ et $\varepsilon \in \mathbb{R}$, le vecteur $\boldsymbol{z}^\star + \varepsilon\boldsymbol{d}$ appartient encore à F. Par minimalité,

$$\|\boldsymbol{z}^\star + \varepsilon\boldsymbol{d} - \boldsymbol{y}\|_2^2 - \|\boldsymbol{z}^\star - \boldsymbol{y}\|_2^2 \geq 0.$$

Développons le membre de gauche :

$$\|(\boldsymbol{z}^\star - \boldsymbol{y}) + \varepsilon\boldsymbol{d}\|_2^2 - \|\boldsymbol{z}^\star - \boldsymbol{y}\|_2^2 = 2\varepsilon\langle \boldsymbol{z}^\star - \boldsymbol{y}, \boldsymbol{d}\rangle + \varepsilon^2\|\boldsymbol{d}\|_2^2.$$

Le quotient par ε et le passage à la limite donnent

$$\begin{cases} \varepsilon \to 0^+ : & \langle \boldsymbol{z}^\star - \boldsymbol{y}, \boldsymbol{d}\rangle \geq 0, \\ \varepsilon \to 0^- : & \langle \boldsymbol{z}^\star - \boldsymbol{y}, \boldsymbol{d}\rangle \leq 0, \end{cases} \quad \Longrightarrow \quad \langle \boldsymbol{y} - \boldsymbol{z}^\star, \boldsymbol{d}\rangle = 0 \quad (\forall \boldsymbol{d} \in F).$$

Avec $\boldsymbol{z}^\star \in F$, c'est la caractérisation de la projection orthogonale.

2. Réciproquement, soit $\boldsymbol{z}^\star \in F$ tel que $\boldsymbol{y} - \boldsymbol{z}^\star \perp F$. Pour tout $\boldsymbol{z} \in F$,

$$\boldsymbol{y} - \boldsymbol{z} = (\boldsymbol{y} - \boldsymbol{z}^\star) + (\boldsymbol{z}^\star - \boldsymbol{z}).$$

Comme $\boldsymbol{z}^\star - \boldsymbol{z} \in F$ et $\boldsymbol{y} - \boldsymbol{z}^\star \perp F$, le théorème de Pythagore donne

$$\|\boldsymbol{y} - \boldsymbol{z}\|_2^2 = \|\boldsymbol{y} - \boldsymbol{z}^\star\|_2^2 + \|\boldsymbol{z}^\star - \boldsymbol{z}\|_2^2 \geq \|\boldsymbol{y} - \boldsymbol{z}^\star\|_2^2.$$

D'où $\|\boldsymbol{y} - \boldsymbol{z}\|_2^2 \geq \|\boldsymbol{y} - \boldsymbol{z}^\star\|_2^2$, avec égalité si et seulement si $\boldsymbol{z} = \boldsymbol{z}^\star$: minimum et unicité.

3. Pour $F = \mathrm{Im}(A)$, la projection $\boldsymbol{z}^\star$ existe (F est un sous-espace fermé de $\mathbb{R}^m$). Puis $\boldsymbol{z}^\star \in \mathrm{Im}(A) \Rightarrow \exists \boldsymbol{u} \in \mathbb{R}^n,\ \boldsymbol{z}^\star = A\boldsymbol{u}$. La condition d'orthogonalité devient

$$\langle \boldsymbol{y} - A\boldsymbol{u}, \boldsymbol{z}\rangle = 0 \qquad (\forall \boldsymbol{z} \in \mathrm{Im}(A)),$$

c'est-à-dire

$$\langle A\boldsymbol{u} - \boldsymbol{y}, \boldsymbol{z}\rangle = 0 \quad (\forall \boldsymbol{z} \in \mathrm{Im}(A)).$$

Le vecteur $A\boldsymbol{u}$ est unique, mais pas nécessairement $\boldsymbol{u}$ si $\ker(A) \neq \{0\}$.

4. Tout $\boldsymbol{z} \in \mathrm{Im}(A)$ est de la forme $A\boldsymbol{v}$, $\boldsymbol{v} \in \mathbb{R}^n$. Ainsi

$$\begin{aligned}\langle A\boldsymbol{u} - \boldsymbol{y}, A\boldsymbol{v}\rangle = 0 \ (\forall \boldsymbol{v}) &\iff (A\boldsymbol{u} - \boldsymbol{y})^T A\boldsymbol{v} = 0 \ (\forall \boldsymbol{v}) \iff \boldsymbol{v}^T A^T (A\boldsymbol{u} - \boldsymbol{y}) = 0 \ (\forall \boldsymbol{v}) \\ &\iff A^T(A\boldsymbol{u} - \boldsymbol{y}) = 0 \iff A^T A\boldsymbol{u} = A^T\boldsymbol{y}.\end{aligned}$$

La troisième équivalence utilise $\boldsymbol{w} \perp \mathbb{R}^n \iff \boldsymbol{w} = 0$: ce sont les *équations normales*.

Exercice 3 – Reformulation d'un système linéaire

Puisque A est inversible, A^T l'est également. Multiplier $A\boldsymbol{u} = \boldsymbol{b}$ par A^T ne change donc pas son ensemble de solutions :

$$A\boldsymbol{u} = \boldsymbol{b} \iff A^T A\boldsymbol{u} = A^T\boldsymbol{b}.$$

1. On vérifie d'abord la symétrie :

$$(A^T A)^T = A^T (A^T)^T = A^T A.$$

Puisque A est inversible,

$$\boldsymbol{x} \neq 0 \implies A\boldsymbol{x} \neq 0 \implies \boldsymbol{x}^T A^T A\boldsymbol{x} = \|A\boldsymbol{x}\|_2^2 > 0.$$

Ainsi $A^T A$ est symétrique définie positive : méthode directe de *Cholesky*, ou itérative de *Gauss–Seidel*.

2. Pour une matrice inversible M, on rappelle que $\mathrm{Cond}_2(M) = \|M\|_2 \left\|M^{-1}\right\|_2$. Si $\sigma_1 \geq \cdots \geq \sigma_n > 0$ sont les valeurs singulières de A, alors

$$\|A\|_2 = \sigma_1, \qquad \left\|A^{-1}\right\|_2 = \frac{1}{\sigma_n}, \qquad \mathrm{Cond}_2(A) = \frac{\sigma_1}{\sigma_n}.$$

Les valeurs propres de $A^T A$ sont $\sigma_1^2, \ldots, \sigma_n^2$. Comme cette matrice est symétrique définie positive,

$$\left\|A^T A\right\|_2 = \sigma_1^2, \qquad \left\|(A^T A)^{-1}\right\|_2 = \frac{1}{\sigma_n^2}.$$

On obtient

$$\mathrm{Cond}_2(A^T A) = \frac{\sigma_1^2}{\sigma_n^2} = \left(\frac{\sigma_1}{\sigma_n}\right)^2 = \mathrm{Cond}_2(A)^2.$$

Le conditionnement est donc élevé au carré.

Exercice 4 – Factorisation QR réduite

On suppose $A \in \mathbb{R}^{m\times n}$ de rang n ; pour la question 2, on considère $n \geq 2$. Toutes ses colonnes sont donc linéairement indépendantes.

1. La première colonne ne peut pas être nulle. Posons

$$\alpha_{11} = \|A_1\|_2 > 0, \qquad \boldsymbol{q}_1 = \frac{A_1}{\alpha_{11}}.$$

Ainsi $\|\boldsymbol{q}_1\|_2 = 1$ et $A_1 = \alpha_{11}\boldsymbol{q}_1$.

2. On cherche $A_2 = \alpha_{12}\boldsymbol{q}_1 + \alpha_{22}\boldsymbol{q}_2$ avec $\boldsymbol{q}_2 \perp \boldsymbol{q}_1$ et $\|\boldsymbol{q}_2\|_2 = 1$. En prenant le produit scalaire avec $\boldsymbol{q}_1$,

$$\langle A_2, \boldsymbol{q}_1\rangle = \alpha_{12}\|\boldsymbol{q}_1\|_2^2 + \alpha_{22}\langle \boldsymbol{q}_2, \boldsymbol{q}_1\rangle = \alpha_{12}.$$

On soustrait la projection sur $\boldsymbol{q}_1$:

$$\boldsymbol{r}_2 = A_2 - \langle A_2, \boldsymbol{q}_1\rangle \boldsymbol{q}_1, \qquad \langle \boldsymbol{r}_2, \boldsymbol{q}_1\rangle = 0.$$

Comme A_1 et A_2 sont indépendantes, $\boldsymbol{r}_2 \neq 0$. On pose donc

$$\alpha_{22} = \|\boldsymbol{r}_2\|_2 > 0, \qquad \boldsymbol{q}_2 = \frac{\boldsymbol{r}_2}{\alpha_{22}}.$$

D'où $\|\boldsymbol{q}_2\|_2 = 1$, $\langle \boldsymbol{q}_1, \boldsymbol{q}_2\rangle = 0$ et $A_2 = \alpha_{12}\boldsymbol{q}_1 + \alpha_{22}\boldsymbol{q}_2$.

3. Le procédé se poursuit par *Gram–Schmidt*. Supposons $\boldsymbol{q}_1, \dots, \boldsymbol{q}_{j-1}$ déjà construits et orthonormés. On retire de A_j ses projections sur ces vecteurs :

$$\alpha_{ij} = \langle A_j, \boldsymbol{q}_i \rangle \ \ (1 \le i < j), \quad \boldsymbol{r}_j = A_j - \sum_{i=1}^{j-1} \alpha_{ij} \boldsymbol{q}_i, \quad \alpha_{jj} = \|\boldsymbol{r}_j\|_2 > 0, \quad \boldsymbol{q}_j = \frac{\boldsymbol{r}_j}{\alpha_{jj}}.$$

Pour $i < j$,

$$\langle \boldsymbol{r}_j, \boldsymbol{q}_i \rangle = \alpha_{ij} - \alpha_{ij} = 0, \qquad \operatorname{rang}(A) = n \implies \boldsymbol{r}_j \neq 0.$$

Les $\boldsymbol{q}_j$ sont donc orthonormés et

$$A_j = \sum_{i=1}^{j} \alpha_{ij} \boldsymbol{q}_i.$$

En juxtaposant ces n identités colonne par colonne, on obtient

$$A = QR, \qquad Q = (\boldsymbol{q}_1 \ \cdots \ \boldsymbol{q}_n) \in \mathbb{R}^{m \times n}, \qquad Q^T Q = I_n,$$

où $R = (\alpha_{ij}) \in \mathbb{R}^{n \times n}$ est triangulaire supérieure, avec $\alpha_{ij} = 0$ pour $i > j$ et $\alpha_{jj} > 0$. Elle est donc inversible.

4. En remplaçant A par QR dans les équations normales,

$$A^T A \boldsymbol{u} = A^T \boldsymbol{y} \iff R^T (Q^T Q) R \boldsymbol{u} = R^T Q^T \boldsymbol{y} \iff R^T R \boldsymbol{u} = R^T Q^T \boldsymbol{y} \iff R \boldsymbol{u} = Q^T \boldsymbol{y}.$$

Ici $Q^T Q = I_n$ et R^T est inversible. Le système $R\boldsymbol{u} = Q^T \boldsymbol{y}$ se résout par remontée.

5. L'identité $A = QR$ et l'orthonormalité des colonnes de Q donnent

$$A^T A = R^T Q^T Q R = R^T R.$$

Si $\lambda_{\min}$ et $\lambda_{\max}$ sont les valeurs propres extrêmes de $R^T R = A^T A$, alors

$$\operatorname{Cond}_2(R) = \frac{\sigma_{\max}(R)}{\sigma_{\min}(R)} = \frac{\sqrt{\lambda_{\max}}}{\sqrt{\lambda_{\min}}} = \sqrt{\frac{\lambda_{\max}}{\lambda_{\min}}} = \sqrt{\operatorname{Cond}_2(A^T A)}.$$

C'est la relation souhaitée.