Mathématiques pour l'Intelligence Artificielle

Notes de cours – CM3

Enseignant : Thomas Schatz

Rédigé par :

Sajjad Ghasemian
Raissa Mezine

16 septembre 2025

Contents

1 Lien entre A^TA et AA^T

Soit $A \in \mathbb{R}^{m\times n}$. Alors :

- A^TA est une matrice symétrique de taille $n \times n$.
- AA^T est une matrice symétrique de taille $m \times m$.

Propriétés principales :

1. **Même rang :**
$$\operatorname{rang}(A^TA) = \operatorname{rang}(AA^T) = \operatorname{rang}(A).$$
2. **Valeurs propres :** Les valeurs propres non nulles de A^TA et AA^T sont identiques. Elles correspondent aux carrés des valeurs singulières de A.
3. **Lien avec la SVD :** Si $A = U\Sigma V^T$ est une décomposition en valeurs singulières (SVD), alors
$$A^TA = V\Sigma^T\Sigma V^T = V\,\operatorname{diag}(\sigma_1^2, \dots, \sigma_r^2, 0, \dots, 0)\,V^T,$$
$$AA^T = U\Sigma\Sigma^T U^T = U\,\operatorname{diag}(\sigma_1^2, \dots, \sigma_r^2, 0, \dots, 0)\,U^T,$$
où $\sigma_1, \dots, \sigma_r$ sont les valeurs singulières non nulles de A.

2 Jacobienne et Hessienne

2.1 Gradient et Jacobienne

Gradient . Pour $f : \mathbb{R}^n \to \mathbb{R}$, le **gradient** est le vecteur colonne
$$\nabla f(x) = \begin{bmatrix} \frac{\partial f}{\partial x_1}(x) \\ \vdots \\ \frac{\partial f}{\partial x_n}(x) \end{bmatrix} \in \mathbb{R}^{n\times 1}.$$

Jacobienne . Pour $g : \mathbb{R}^n \to \mathbb{R}^m$ avec $g(x) = \big(g_1(x), \dots, g_m(x)\big)^T$, la **matrice Jacobienne** est
$$J_g(x) = \begin{bmatrix} \frac{\partial g_1}{\partial x_1} & \cdots & \frac{\partial g_1}{\partial x_n} \\ \vdots & \ddots & \vdots \\ \frac{\partial g_m}{\partial x_1} & \cdots & \frac{\partial g_m}{\partial x_n} \end{bmatrix} \in \mathbb{R}^{m\times n}.$$

Chaque *ligne* est le gradient (transposé) d'un des composantes : $J_g(x) = \begin{bmatrix} \nabla g_1(x)^T \\ \cdots \\ \nabla g_m(x)^T \end{bmatrix}$.

2.2 Hessienne

Pour $f : \mathbb{R}^n \to \mathbb{R}$, la **Hessienne** est la matrice carrée des dérivées secondes
$$\nabla^2 f(x) \;=\; \begin{bmatrix} \frac{\partial^2 f}{\partial x_1^2} & \cdots & \frac{\partial^2 f}{\partial x_1 \partial x_n} \\ \vdots & \ddots & \vdots \\ \frac{\partial^2 f}{\partial x_n \partial x_1} & \cdots & \frac{\partial^2 f}{\partial x_n^2} \end{bmatrix} \in \mathbb{R}^{n\times n}.$$

3 Ensembles ouverts et fermés

- Un **ensemble ouvert** $U \subset \mathbb{R}^n$ est un ensemble qui ne contient pas ses bords. Autrement dit : si tu prends un point de l'ensemble, tu peux toujours trouver un petit voisinage (un petit intervalle autour, ou un petit disque autour en dimension 2) qui reste entièrement dans l'ensemble
- Un **ensemble fermé** $F \subset \mathbb{R}^n$ est un ensemble qui contient tous ses points frontières. Autrement dit : si une suite de points de l'ensemble converge vers une limite, alors cette limite appartient aussi à l'ensemble..
- Un ensemble peut être ni ouvert ni fermé, ou bien à la fois. Exemple : $[0, 1)$ n'est ni ouvert ni fermé ; $\mathbb{R}$ est à la fois ouvert et fermé.

4 Optimisation

4.1 Sans contrainte

Problème : minimiser $f(x)$ avec $x \in \mathbb{R}^n$.
Condition nécessaire pour un minimum local :

$$\nabla f(x^*) = 0$$

Cela signifie que le gradient s'annule au point critique.

4.2 Avec contraintes

Quand on ajoute des contraintes, le problème devient plus complexe. On distingue deux types de contraintes :

- **Contrainte d'égalité :** $g(x) = 0$
- **Contrainte d'inégalité :** $h(x) \leq 0$

Multiplicateurs de Lagrange : Pour résoudre un problème avec contraintes, on introduit des multiplicateurs λ (aussi appelés multiplicateurs de Lagrange). L'idée est de former une *fonction de Lagrange* :

$$\mathcal{L}(x, \lambda, \mu) = f(x) + \sum_{i \in \mathcal{E}} \lambda_i g_i(x) + \sum_{j \in \mathcal{I}} \mu_j h_j(x)$$

où :

- $\mathcal{E}$ = ensemble des contraintes d'égalité,
- $\mathcal{I}$ = ensemble des contraintes d'inégalité,
- $\lambda_i \in \mathbb{R}$ (sans signe imposé),
- $\mu_j \geq 0$ (toujours positifs).

Conditions de Karush-Kuhn-Tucker (KKT) : Un point x^* est solution si et seulement si :

1. **Stationnarité :**
$$\nabla_x \mathcal{L}(x'\lambda'\mu^*) = 0$$
2. **Primalité :**
$$g_i(x^*) = 0 \quad \forall i \in \mathcal{E}, \qquad h_j(x^*) \leq 0 \quad \forall j \in \mathcal{I}$$
3. **Dualité :**
$$\mu_j^* \geq 0 \quad \forall j \in \mathcal{I}$$
4. **Complémentarité :**
$$\mu_j^* \, h_j(x^*) = 0 \quad \forall j \in \mathcal{I}$$
(c'est-à-dire : soit la contrainte est active $h_j(x^*) = 0$, soit le multiplicateur associé vaut 0).

Interprétation intuitive :

- Sans contraintes : on cherche où le gradient est nul.
- Avec contraintes : le gradient de f doit être *aligné* avec les gradients des contraintes actives.
- Les multiplicateurs λ, μ mesurent l'« importance » de chaque contrainte dans la solution.

5 Condition nécessaire d'optimalité : cas général

5.1 Qualification de Mangasarian-Fromovitz (MFCQ)

La **qualification de contraintes de Mangasarian-Fromovitz (MFCQ)** s'énonce ainsi :

- Les gradients des contraintes d'égalité sont linéairement indépendants en x^*.
- Il existe un vecteur $d \in \mathbb{R}^n$ tel que :
$$\nabla c_i(x^*)^T d < 0 \quad \text{pour toutes les contraintes inégalités actives,}$$
et
$$\nabla e_j(x^*)^T d = 0 \quad \text{pour toutes les contraintes égalités.}$$

Théorème. Si x^* est une solution locale et que la qualification de contraintes de Mangasarian-Fromovitz est respectée en x^*, alors les conditions KKT sont vérifiées en x^*.

L'idée de MFCQ est la suivante :

- Les contraintes d'égalité doivent être « bien posées » (leurs gradients ne sont pas redondants).
- Pour les contraintes d'inégalité qui sont actives en x^*, il doit exister une direction d qui permet de *s'éloigner* de ces contraintes (en les rendant moins restrictives).