Méthode de Gradient à pas optimal

(*rédigé par Rémi Moreau et Vincent Béhani*)

Leçons : 162 ; 219 ; 226 ; 229 ; 233

Références : RAMIS–WARUSFEL, *Cours de Mathématiques pures et appliquées* (p.412) ou HERON, *Analyse numérique* (p.15 et 263)

Prérequis :

- le théorème spectral
- $\mathrm{Cond}_2(A) = \dfrac{\mu_n}{\mu_1}$ où μ_1 et μ_n sont respectivement les valeurs propres minimales et maximales de A.

Introduction :

Comme on va le voir dans le théorème suivant et l'algorithme du gradient à pas optimal, on cherche à trouver une solution de $AX = B$ en l'approchant par une suite. Ainsi on veut trouver le minimum de la fonctionnelle convexe f présentée ci dessous. Le théorème 1 et l'algorithme sont des éléments à mettre dans le plan de la leçon et le développement consiste en la démonstration du théorème 2 avec l'inégalité de Kantorovich.

Théorème 1. Si $A \in \mathcal{S}_n^{++}(\mathbb{R})$ alors :

$$AX_0 = B \iff X_0 \text{ minimise l'application } f : \begin{cases} \mathbb{R}^n & \to \quad \mathbb{R} \\ X & \mapsto \quad \frac{1}{2}\langle AX, X\rangle - \langle B, X\rangle \end{cases}$$

Démonstration. Soit $X_0 \in \mathbb{R}^n$ solution du système $AX = B$. Remarquons que pour tout $H \in \mathbb{R}^n$, puisque $A \in \mathcal{S}_n^{++}(\mathbb{R})$ on a :

$$\begin{aligned} f(X_0 + H) &= \tfrac{1}{2}\langle AX_0, X_0\rangle + \tfrac{1}{2}\langle AH, X_0\rangle + \tfrac{1}{2}\langle AX_0, H\rangle + \tfrac{1}{2}\langle AH, H\rangle - \langle B, X_0\rangle - \langle B, H\rangle \\ &= f(X_0) + \tfrac{1}{2}\langle AX_0, H\rangle + \tfrac{1}{2}\langle AX_0, H\rangle + \tfrac{1}{2}\langle AH, H\rangle - \langle B, H\rangle \\ &= f(X_0) + \langle B, H\rangle + \tfrac{1}{2}\langle AH, H\rangle - \langle B, H\rangle \\ &= f(X_0) + \tfrac{1}{2}\langle AH, H\rangle \\ &\geq f(X_0) \end{aligned}$$

avec $\frac{1}{2}\langle AH, H\rangle > 0$ si et seulement si $H \neq 0$. Alors X_0 est l'unique minimum global de f.

Réciproquement, supposons que $X_0 \in \mathbb{R}^n$ minimise f. Alors $\nabla f(X_0) = 0$. Or le calcul précédent indique que pour tout $H \in \mathbb{R}^n$ et tout $X \in \mathbb{R}^n$:

$$f(X + H) \underset{\|H\| \to 0}{=} f(X) + \langle AX - B, H\rangle + \mathcal{O}(\|H\|^2)$$

Ainsi $\nabla f(X_0) = AX_0 - B = 0$ implique que X_0 est solution du système $AX = B$. □

Algorithme du gradient à pas optimal :

Soit $X_0 \in \mathbb{R}^n$ et $R_0 = AX_0 - B$. Soit $\varepsilon > 0$:
tant que $\frac{\|R_n\|}{\|R_0\|} \geq \varepsilon$:

- $\alpha_{n+1} = \frac{\|R_n\|^2}{\langle AR_n, R_n\rangle}$;
- $X_{n+1} = X_n - \alpha_{n+1} R_n$;

- $R_{n+1} = AX_{n+1} - B$;

Résultat : $\frac{\|R_n\|}{\|R_0\|} = \frac{\|AX_n - B\|}{\|AX_0 - B\|} < \varepsilon$.

Cherchant la solution du système $AX = B$, on cherche à minimiser la fonction :

$$f : \left\{ \begin{array}{ccc} \mathbb{R}^n & \to & \mathbb{R} \\ X & \mapsto & \frac{1}{2}\langle AX, X\rangle - \langle B, X\rangle \end{array} \right.$$

Partant d'un vecteur $X_0 \in \mathbb{R}^n$, on construit une suite $(X_k)_{k\in\mathbb{N}}$ convergeant vers son unique minimum global. Or puisque $\nabla f(X_k)$ est la direction selon laquelle f croît le plus à partir de X_k, cherchons X_{k+1} de la forme $X_{k+1} = X_k - \alpha_{k+1}\nabla f(X_k)$ tel que $f(X_k - \alpha_{k+1}\nabla f(X_k))$ soit minimal. Considérons l'application :

$$\phi_k : \left\{ \begin{array}{ccc}]0, +\infty[& \to & \mathbb{R} \\ t & \mapsto & f(X_k - t\nabla f(X_k)) \end{array} \right.$$

Pour tout $t > 0$, on a :

$$\begin{array}{rcl} \phi_k(t) & = & f(X_k) + \langle AX_k, -t\nabla f(X_k)\rangle + \frac{1}{2}\langle A(-t\nabla f(X_k)), -t\nabla f(X_k)\rangle - \langle B, -t\nabla f(X_k)\rangle \\ & = & f(X_k) - t\langle AX_k - B, \nabla f(X_k)\rangle + \frac{t^2}{2}\langle A\nabla f(X_k), \nabla f(X_k)\rangle \end{array}$$

Or $R_k = AX_k - B = \nabla f(X_k)$ donc :

$$\phi_k(t) = f(X_k) - t\|R_k\|^2 + \frac{t^2}{2}\langle AR_k, R_k\rangle$$

Ainsi pour tout $t > 0$ on a :

$$\phi_k'(t) = -\|R_k\|^2 + t\langle AR_k, R_k\rangle \geq 0 \quad \Longleftrightarrow \quad t \geq \frac{\|R_k\|^2}{\langle AR_k, R_k\rangle} = \alpha_{k+1}$$

Donc ϕ_k est décroissante sur $]0, \alpha_{k+1}]$ et croissante sur $[\alpha_{k+1}, +\infty[$, elle atteint alors son minimum en $\alpha_{k+1} = \frac{\|R_k\|^2}{\langle AR_k, R_k\rangle}$.

Théorème 2 (Méthode du gradient à pas optimal)**.** L'algorithme du gradient converge vers la solution $\overline{X}$ du système $AX = B$ et pour tout $k \in \mathbb{N}$:

$$\|X_k - \overline{X}\| \leq \sqrt{\mathrm{Cond}(A)} \left(\frac{\mathrm{Cond}(A) - 1}{\mathrm{Cond}(A) + 1}\right)^k \|X_0 - \overline{X}\|$$

Lemme 1 (Inégalité de Kantorovich)**.** Toute matrice $A \in \mathcal{S}_n^{++}(\mathbb{R})$, de valeurs propres minimale et maximale respectivement μ_1 et μ_n, vérifie pour tout $X \in \mathbb{R}^n\backslash\{0\}$:

$$\frac{\|X\|^4}{\|X\|_{A^{-1}}^2\|X\|_A^2} \geq 4\frac{\mu_1\mu_n}{(\mu_1 + \mu_n)^2}$$

où $\|X\|_A^2 = \langle AX, X\rangle$.

Démonstration du lemme. Puisque $A \in \mathcal{S}_n^{++}(\mathbb{R})$, il existe une base $(e_i)_{i\in\{1,\ldots,n\}}$ orthonormale de $\mathbb{R}^n$ constituée de vecteurs propres de A associés aux valeurs propres $(\mu_i)_{i\in\{1,\ldots,n\}}$. Soit un vecteur non nul $X = \sum_{i=1}^n x_i e_i \in \mathbb{R}^n$ exprimé dans la base tel que $AX = A\left(\sum_{i=1}^n x_i e_i\right) = \sum_{i=1}^n \mu_i x_i e_i$. Alors puisque $\sqrt{ab} \leq \frac{1}{2}(a+b)$ pour tout a et b positifs, on a :

$$\begin{aligned} \sqrt{\|X\|_{A^{-1}}^2 \|X\|_A^2} &= \sqrt{\left(\sum_{i=1}^n \frac{1}{\mu_i} x_i^2\right)\left(\sum_{j=1}^n \mu_j x_j^2\right)} = \sqrt{\frac{\mu_1}{\mu_n}\left(\sum_{i=1}^n \frac{\mu_n}{\mu_i} x_i^2\right)\left(\sum_{j=1}^n \frac{\mu_j}{\mu_1} x_j^2\right)} \\ &\leq \frac{1}{2}\sqrt{\frac{\mu_1}{\mu_n}}\left(\sum_{i=1}^n \frac{\mu_n}{\mu_i} x_i^2 + \sum_{j=1}^n \frac{\mu_j}{\mu_1} x_j^2\right) \leq \frac{1}{2}\sqrt{\frac{\mu_1}{\mu_n}} \sum_{i=1}^n \left(\frac{\mu_n}{\mu_i} + \frac{\mu_i}{\mu_1}\right) x_i^2 \end{aligned}$$

Or l'application $\psi : \left\{ \begin{array}{ccc} [\mu_1, \mu_n] & \to & \mathbb{R} \\ x & \mapsto & \frac{\mu_n}{x} + \frac{x}{\mu_1} \end{array} \right.$ est dérivable et de dérivée $\psi'(x) = \frac{1}{\mu_1} - \frac{\mu_n}{x^2}$ pour tout $x \in [\mu_1, \mu_n]$. Ainsi pour tout $x \in [\mu_1, \mu_n]$:

$$\psi'(x) \geq 0 \iff \frac{1}{\mu_1} \geq \frac{\mu_n}{x^2} \iff x \geq \sqrt{\mu_1 \mu_n}$$

Donc ψ est décroissante sur $[\mu_1, \sqrt{\mu_1\mu_n}]$ et croissante sur $[\sqrt{\mu_1\mu_n}, \mu_n]$. L'application ψ atteint dès lors son maximum en μ_1 et μ_n car $\psi(\mu_1) = 1 + \frac{\mu_n}{\mu_1} = \psi(\mu_n)$. D'où :

$$\begin{aligned} \sqrt{\|X\|_{A^{-1}}^2 \|X\|_A^2} &\leq \frac{1}{2}\sqrt{\frac{\mu_1}{\mu_n}} \sum_{i=1}^n \left(1 + \frac{\mu_n}{\mu_1}\right) x_i^2 \leq \frac{1}{2}\left(\sqrt{\frac{\mu_1}{\mu_n}} + \sqrt{\frac{\mu_n}{\mu_1}}\right) \sum_{i=0}^n x_i^2 \\ &\leq \frac{1}{2}\left(\sqrt{\frac{\mu_1}{\mu_n}} + \sqrt{\frac{\mu_n}{\mu_1}}\right) \|X\|^2 \end{aligned}$$

Or remarquons que :

$$\sqrt{\frac{\mu_1}{\mu_n}} + \sqrt{\frac{\mu_n}{\mu_1}} = \frac{\mu_1 + \mu_n}{\sqrt{\mu_1 \mu_n}}$$

ce qui donne :

$$\|X\|_{A^{-1}}^2 \|X\|_A^2 \leq \frac{1}{4} \frac{(\mu_1 + \mu_n)^2}{\mu_1 \mu_n} \|X\|^4 \implies \frac{\|X\|^4}{\|X\|_{A^{-1}}^2 \|X\|_A^2} \geq 4 \frac{\mu_1 \mu_n}{(\mu_1 + \mu_n)^2}$$

□

Démonstration du théorème. Montrons que la suite $(X_k)_{k \in \mathbb{N}}$ converge vers la solution $\overline{X}$ du système $AX = B$. Remarquons tout d'abord que pour tout $k \in \mathbb{N}$:

$$\begin{aligned} \langle R_{k+1}, R_k \rangle &= \langle AX_{k+1} - B, R_k \rangle = \langle AX_k - B - \alpha_{k+1} AR_k, R_k \rangle = \langle R_k, R_k \rangle - \alpha_{k+1} \langle AR_k, R_k \rangle \\ &= \|R_k\|^2 - \frac{\|R_k\|^2}{\|R_k\|_A^2} \|R_k\|_A^2 = 0 \end{aligned}$$

Ainsi on calcule :

$$\begin{aligned} \|X_{k+1} - \overline{X}\|_A^2 &= \langle A(X_{k+1} - \overline{X}), X_{k+1} - \overline{X} \rangle \\ &= \langle AX_{k+1} - B, X_{k+1} - X_k \rangle + \langle A(X_{k+1} - \overline{X}), X_k - \overline{X} \rangle \\ &= \langle R_{k+1}, -\alpha_{k+1} R_k \rangle + \langle A(X_{k+1} - X_k), X_k - \overline{X} \rangle + \langle A(X_k - \overline{X}), X_k - \overline{X} \rangle \\ &= \langle X_{k+1} - X_k, A(X_k - \overline{X}) \rangle + \|X_k - \overline{X}\|_A^2 \\ &= -\alpha_{k+1} \langle R_k, AX_k - B \rangle + \|X_k - \overline{X}\|_A^2 \\ &= -\alpha_{k+1} \langle R_k, R_k \rangle + \|X_k - \overline{X}\|_A^2 \\ &= -\frac{\|R_k\|^4}{\|R_k\|_A^2} + \|X_k - \overline{X}\|_A^2 \end{aligned}$$

Remarquons que :

$$\|X_k - \overline{X}\|_A^2 = \langle A(X_k - \overline{X}), X_k - \overline{X} \rangle = \langle AX_k - B, A^{-1}(AX_k - B) \rangle = \langle R_k, A^{-1} R_k \rangle = \|R_k\|_{A^{-1}}^2$$

Ainsi :

$$\begin{aligned}\|X_{k+1}-\overline{X}\|_A^2 &= \left(1-\frac{\|R_k\|^4}{\|R_k\|_A^2\|R_k\|_{A^{-1}}^2}\right)\|X_k-\overline{X}\|_A^2 \leq \left(1-\frac{4\mu_1\mu_n}{(\mu_1+\mu_n)^2}\right)\|X_k-\overline{X}\|_A^2 \\ &\leq \frac{(\mu_n-\mu_1)^2}{(\mu_n+\mu_1)^2}\|X_k-\overline{X}\|_A^2\end{aligned}$$

En itérant cette inégalité, on a pour tout $k \in \mathbb{N}$:

$$\|X_k-\overline{X}\|_A \leq \left(\frac{\mu_n-\mu_1}{\mu_n+\mu_1}\right)^k \|X_0-\overline{X}\|_A$$

En remarquant que $0 < \frac{\mu_n-\mu_1}{\mu_n+\mu_1} < 1$, on a convergence de la suite $(X_k)_{k\in\mathbb{N}}$ vers $\overline{X}$. De plus pour tout $X = \sum_{i=1}^n x_i e_i \in \mathbb{R}^n$ exprimé dans une base orthonormale de vecteurs propres de A :

$$\mu_1\|X\|^2 = \sum_{i=1}^n \mu_1 x_i^2 \leq \sum_{i=1}^n \mu_i x_i^2 = \|X\|_A^2 \leq \sum_{i=1}^n \mu_n x_i^2 = \mu_n\|X\|^2$$

Alors :

$$\|X_k-\overline{X}\| \leq \frac{1}{\sqrt{\mu_1}}\|X_k-\overline{X}\|_A \leq \frac{1}{\sqrt{\mu_1}}\left(\frac{\mu_n-\mu_1}{\mu_n+\mu_1}\right)^k\|X_0-\overline{X}\|_A \leq \sqrt{\frac{\mu_n}{\mu_1}}\left(\frac{\mu_n-\mu_1}{\mu_n+\mu_1}\right)^k\|X_0-\overline{X}\|$$

Or on sait que $\mathrm{Cond}_2(A) = \frac{\mu_n}{\mu_1}$, d'où :

$$\sqrt{\frac{\mu_n}{\mu_1}}\frac{\mu_n-\mu_1}{\mu_n+\mu_1} = \sqrt{\frac{\mu_n}{\mu_1}}\frac{\frac{\mu_n}{\mu_1}-1}{\frac{\mu_n}{\mu_1}+1} = \sqrt{\mathrm{Cond}_2(A)}\frac{\mathrm{Cond}_2(A)-1}{\mathrm{Cond}_2(A)+1}$$

D'où l'inégalité :

$$\|X_k-\overline{X}\| \leq \sqrt{\mathrm{Cond}_2(A)}\left(\frac{\mathrm{Cond}_2(A)-1}{\mathrm{Cond}_2(A)+1}\right)^k\|X_0-\overline{X}\|$$

$\square$

Astuces de l'agrégatif :
Comme je l'ai dit plus haut, le développement se concentre sur le theorème 2 et le lemme 1. S'il reste du temps, il est bien d'expliquer pourquoi on fait tout cela avec notamment le théorème 1 et la raison pour laquelle α_{k+1} est défini comme ça.

Questions possibles :

- Est-ce qu'on peut changer le rôle de μ_1 et μ_n ?
- Vaut-il mieux que les valeurs propres soient rapprochées ou éloignées pour améliorer la vitesse de convergence ?
- Quel est le lien entre l'algorithme et le théorème qui dit que $\bar{X}$ est solution de $AX = b$ SSI $\bar{X}$ minimise la fonctionnelle convexe

$$f : \left\{\begin{array}{ccc} \mathbb{R}^n & \to & \mathbb{R} \\ X & \mapsto & \frac{1}{2}\langle AX, X\rangle - \langle b, X\rangle \end{array}\right.$$