Méthode du gradient à pas optimal

Ce développement est extrait du *Cours de mathématiques pures et appliquées, volume 1* de Ramis, Warusfel et Moulin. La preuve de l'inégalité de Kantorovitch se trouve dans les oraux X-ENS.
Soit A une matrice symétrique définie positive. On utilise sans démonstration le résultat suivant : $\overline{x}$ est la solution du système $Ax = b$ si et seulement s'il minimise la fonction $f : x \mapsto \frac{1}{2}\langle Ax, x\rangle - \langle b, x\rangle$. Il s'agit d'appliquer l'algorithme de minimisation du gradient à pas optimal à la fonction f : on construit la suite (x_n) en posant

$$x_{n+1} = x_n + \alpha_{n+1} r_n$$

où $r_n = -\nabla f(x_n) = b - Ax_n$ (direction de plus grande pente) et α_{n+1} est le pas optimal, c'est-à-dire que

$$f(x_n + \alpha_{n+1} r_n) = \min_{\alpha>0} f(x_n + \alpha r_n).$$

On vérifie que $\alpha_{n+1} = \frac{\|r_n\|^2}{\langle Ar_n, r_n\rangle}$, défini tant que $x_n \neq \overline{x}$.
On rappelle l'algorithme :

> Soit x_0 donné et $r_0 = b - Ax_0$. On se fixe un seuil de tolérance TOL.
> Tant que $\frac{\|r_n\|}{\|r_0\|} \geqslant TOL$, faire

$$\alpha_{n+1} = \frac{\|r_n\|^2}{\langle Ar_n, r_n\rangle} \tag{1}$$
$$x_{n+1} = x_n + \alpha_{n+1} r_n \tag{2}$$
$$r_{n+1} = b - Ax_{n+1}. \tag{3}$$

Théorème. *L'algorithme du gradient converge vers la solution $\overline{x}$ du système $Ax = b$ et, pour tout $n \in \mathbb{N}$*

$$\|x_n - \overline{x}\| \leqslant \left(\frac{\mathrm{Cond}(A) - 1}{\mathrm{Cond}(A) + 1}\right)^n \sqrt{\mathrm{Cond}(A)}\|x_0 - \overline{x}\|.$$

Démonstration.

Lemme. *Inégalité de Kantorovitch Soit $A \in \mathcal{M}_N(\mathbb{R})$ symétique définie positive, de valeurs propres minimale et maximale λ_m et λ_M. Alors*

$$\forall x \in \mathbb{R}, \quad \frac{\|x\|^4}{\|x\|^2_{A^{-1}}\|x\|^2_A} \geqslant 4\frac{\lambda_m\lambda_M}{(\lambda_M + \lambda_m)^2}$$

où $\|x\|_A = \sqrt{\langle Ax, x\rangle}$.

Démonstration. Il suffit de montrer cette inégalité pour x de norme 1, ce qu'on supposera dans la suite. A est symétrique donc il existe une base $(e_1, \cdots, e_n)$ orthonormale de vecteurs propres de A. On note $x = \sum x_i e_i$ et $0 < \lambda_1 \leqslant \cdots \leqslant \lambda_n$ les valeurs propres de A.

$$\begin{aligned}\langle Ax, x\rangle \langle A^{-1}x, x\rangle &= \sum_{i=1}^n \lambda_i x_i^2 \sum_{i=1}^n \frac{1}{\lambda_i} x_i^2 = \frac{\lambda_1}{\lambda_n} \sum_{i=1}^n \frac{\lambda_i}{\lambda_n} x_i^2 \sum_{i=1}^n \frac{\lambda_1}{\lambda_i} x_i^2 \\ &\leqslant \frac{\lambda_1}{4\lambda_n}\left(\sum_{i=1}^n \left(\frac{\lambda_i}{\lambda_n} + \frac{\lambda_1}{\lambda_i}\right) x_i^2\right)^2.\end{aligned}$$

Après étude de la fonction $x \mapsto x + \frac{1}{x}$ qui atteint son maximum $1 + \frac{\lambda_n}{\lambda_1}$ en λ_1 et en λ_n, on en déduit que

$$\langle Ax, x\rangle \langle A^{-1}x, x\rangle \leqslant \frac{\lambda_1}{4\lambda_n}\left(\left(1+\frac{\lambda_n}{\lambda_1}\right)\sum_{i=1}^{n} x_i^2\right)^2 =\leqslant \frac{\lambda_1}{4\lambda_n}\left(1+\frac{\lambda_n}{\lambda_1}\right)^2 = \frac{(\lambda_1+\lambda_n)^2}{4\lambda_1\lambda_n}.$$

Il suffit d'inverser cette formule pour obtenir l'inégalité escomptée.

□

Pour la preuve générale, remarquons que

$$f(x) = \frac{1}{2}\langle Ax, x\rangle - \langle b, x\rangle = \frac{1}{2}\langle Ax, x\rangle - \langle A\overline{x}, x\rangle = \frac{1}{2}\langle A(x-\overline{x}), x-\overline{x}\rangle - \frac{1}{2}\langle A\overline{x}, \overline{x}\rangle,$$

c'est-à-dire $f(x) = \frac{1}{2}\|x-\overline{x}\|_A - \frac{1}{2}\langle A\overline{x}, \overline{x}\rangle$. L'algorithme garantit la décroissance de f à chaque itération. Reste à voir qu'on a en fait une contractance de f :

$$\frac{\|x_{n+1}-\overline{x}\|_A^2}{\|x_n-\overline{x}\|_A^2} = \frac{\|x_n+\alpha_{n+1}r_n-\overline{x}\|_A^2}{\|x_n-\overline{x}\|_A^2} = \frac{\|x_n-\overline{x}\|_A^2 + \alpha_{n+1}^2\|r_n\|_A^2 + 2\langle A(x_n-\overline{x}), \alpha_{n+1}r_n\rangle}{\|x_n-\overline{x}\|_A^2} \tag{4}$$

$$= 1 + \frac{\|r_n\|^4/\|r_n\|_A^2 + 2\alpha_{n+1}\langle Ax_n-b, b-Ax_n\rangle}{\|x_n-\overline{x}\|_A^2} \tag{5}$$

$$= 1 - \frac{\|r_n\|^4/\|r_n\|_A^2}{\|x_n-\overline{x}\|_A^2} = 1 - \frac{\|r_n\|^4}{\|r_n\|_A^2\|A^{-1}r_n\|_A^2} = 1 - \frac{\|r_n\|^4}{\langle Ar_n, r_n\rangle \langle A^{-1}r_n, r_n\rangle}. \tag{6}$$

On peut alors appliquer l'inégalité de Kantorovitch qui donne

$$\|x_{n+1}-\overline{x}\|_A^2 \leqslant \left(1 - 4\frac{\lambda_1\lambda_n}{(\lambda_1+\lambda_n)^2}\right)\|x_n-\overline{x}\|_A^2 \tag{7}$$

$$\leqslant \left(\frac{\lambda_1-\lambda_n}{\lambda_1+\lambda_n}\right)^2\|x_n-\overline{x}\|_A^2. \tag{8}$$

$$\leqslant \left(\frac{1-\mathrm{Cond}(A)}{1+\mathrm{Cond}(A)}\right)^2\|x_n-\overline{x}\|_A^2. \tag{9}$$

Ceci montre la convergence linéaire de (x_n) vers $\overline{x}$ pour $\|\cdot\|_A$.
Or, pour tout x, $\leqslant \lambda_1\|x\|^2 \langle Ax, x\rangle = \|x\|_A^2 \leqslant \lambda_n\|x\|^2$, c'est-à-dire

$$\|x_n-\overline{x}\| \leqslant \left(\frac{1-\mathrm{Cond}(A)}{1+\mathrm{Cond}(A)}\right)^n\sqrt{\lambda_n/\lambda_1}\|x_0-\overline{x}\| = \left(\frac{1-\mathrm{Cond}(A)}{1+\mathrm{Cond}(A)}\right)^n\sqrt{\mathrm{Cond}(A)}\|x_0-\overline{x}\|,$$

ce qu'il fallait montrer.

□