Descente de gradient optimale d'une fonctionnelle quadratique

Lemme. *(Inégalité de Kantorovitch) Soit $A \in S_n^{++}(\mathbb{R})$, on note $\lambda_1 \leq .. \leq \lambda_n$ ses valeurs propres et $\|x\|_A^2 = \langle Ax, x\rangle$ la norme induite par A. Alors pour tout $x \in \mathbb{R}^n$,*

$$\|x\|_A \|x\|_{A^{-1}} \leq \frac{1}{2}\left(\sqrt{\frac{\lambda_n}{\lambda_1}} + \sqrt{\frac{\lambda_1}{\lambda_n}}\right)\|x\|^2$$

Démonstration. Supposons que $\|x\| = 1$. Alors $\|x\|_A^2$ est un barycentre en les λ_i.
Soit $t > 0$, une inégalité arithmético-géométrique sur le terme de droite donne

$$\|x\|_A \|x\|_{A^{-1}} \leq \frac{1}{2}\|x\|^2_{\frac{A}{t}+\frac{t}{A}}$$

Les valeurs propres de $\frac{A}{t} + \frac{t}{A}$ sont les $\frac{\lambda_i}{t} + \frac{t}{\lambda_i}$. $\lambda \mapsto \frac{\lambda}{t} + \frac{t}{\lambda}$ est une fonction convexe, donc pour tout i,

$$\frac{\lambda_i}{t} + \frac{t}{\lambda_i} \leq \max\left(\frac{\lambda_n}{t} + \frac{t}{\lambda_n}, \frac{\lambda_1}{t} + \frac{t}{\lambda_1}\right)$$

Avec $t = \sqrt{\lambda_1 \lambda_n}$ et la remarque précédente, on obtient l'inégalité recherchée. □

Théorème. *Soit $A \in S_n^{++}(\mathbb{R})$, $b \in \mathbb{R}^n$, on note $x_0 \in \mathbb{R}^n$, $\bar{x} = A^{-1}b$, $c = \lambda_n/\lambda_1$, et on défini la descente de gradient optimale $(x_k)_k$ par*

$$\begin{cases} x_{k+1} = x_k - \frac{\|d_k\|^2}{\|d_k\|_A^2} d_k \\ d_k = Ax_k - b \end{cases}$$

Alors

$$\|x_k - \bar{x}\| \leq c\left(\frac{c-1}{c+1}\right)^k \|x_0 - \bar{x}\|$$

Démonstration. Soit $f(x) := \frac{1}{2}\langle x, Ax\rangle - \langle b, x\rangle$. f est une fonction convexe coercive, telle que son unique minimum soit atteint en $\bar{x}$. On a $\nabla f(x) = Ax - b$ et :

$$f(x) - f(\bar{x}) = \frac{1}{2}\|x\|_A^2$$

Soit $x_0 \in \mathbb{R}^n$, on définit la suite $(x_k)_k$ de descente de gradient à pas optimal par récurrence ; si x_k est définie, on pose $d_k = \nabla f(x_k) = Ax_k - b$, et on cherche à minimiser la fonction $t \mapsto f(x_k - td_k) - f(\bar{x})$. Cette expression se réécrit :

$$\begin{aligned} f(x_k - td_k) - f(\bar{x}) &= \frac{1}{2}\left(\|x_k - \bar{x}\|_A^2 - 2t\langle A(x_k - \bar{x}), d_k\rangle + t^2\|d_k\|_A^2\right) \\ &= \frac{1}{2}\left(\|x_k - \bar{x}\|_A^2 - 2t\|d_k\|^2 + t^2\|d_k\|_A^2\right) \end{aligned}$$

C'est un polynôme du second degré qui admet un unique minimum en $t = \frac{\|d_k\|^2}{\|d_k\|_A^2}$; on en déduit la formule de récurrence.

Estimons maintenant la vitesse de convergence ; pour cela, notons que :

$$\begin{aligned}
\|x_{k+1} - \bar{x}\|_A^2 &= \|x_k - \bar{x} - \frac{\|d_k\|^2}{\|d_k\|_A^2} d_k\|_A^2 \\
&= \|x_k - \bar{x}\|_A^2 - 2\frac{\|d_k\|^2}{\|d_k\|_A^2} \langle A(x_k - \bar{x}), d_k\rangle + \left(\frac{\|d_k\|^2}{\|d_k\|_A^2}\right)^2 \|d_k\|_A^2 \\
&= \|x_k - \bar{x}\|_A^2 - \frac{\|d_k\|^4}{\|d_k\|_A^2} \\
&= \|x_k - \bar{x}\|_A^2 \left(1 - \frac{\|d_k\|^4}{\|d_k\|_A^2 \|x_k - \bar{x}\|_A^2}\right)
\end{aligned}$$

Enfin, $\|x_k - \bar{x}\|_A^2 = \|d_k\|_{A^{-1}}$. On utilise alors l'inégalité de Kantorovich :

$$\begin{aligned}
\|x_{k+1} - \bar{x}\|_A^2 &\leq \|x_k - \bar{x}\|_A^2 \left(1 - \frac{4}{\left(\sqrt{\frac{\lambda_n}{\lambda_1}} + \sqrt{\frac{\lambda_1}{\lambda_n}}\right)^2}\right) \\
&= \left(\frac{c-1}{c+1}\right)^2 \|x_k - \bar{x}\|_A^2
\end{aligned}$$

□