Algorithme du gradient à pas optimal

Référence : Hiriart-Urruty Optimisation et analyse convexe exercice II.8. p.52 et FGNal3
Leçons : 162,232,233

Lemme (Inégalité de Kantorovitch)

Soit $A \in \mathcal{S}_n^{++}(\mathbb{R})$ de valeurs propres minimale et maximale $\lambda_{\min}$ et $\lambda_{\max}$. Alors en posant $c(A) := \dfrac{\lambda_{\min}}{\lambda_{\max}}$ le conditionnement de A,

$$\forall x \in \mathbb{R}, \quad \frac{\|x\|^4}{\langle Ax, x\rangle \langle A^{-1}x, x\rangle} \geqslant 4\left(\sqrt{\frac{\lambda_{\max}}{\lambda_{\min}}} + \sqrt{\frac{\lambda_{\min}}{\lambda_{\max}}}\right)^{-2} = 4\frac{c(A)}{(c(A)+1)^2}$$

Preuve :

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 $x \mapsto \dfrac{x}{\lambda_1} + \dfrac{\lambda_n}{x}$ qui atteint son maximum $1 + \dfrac{\lambda_n}{\lambda_1}$ en λ_1 et en λ_n, on en déduit que

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

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

■

Théorème

Problème : Soit $A \in \mathcal{S}_n^+ + (\mathbb{R})$ et $b \in \mathbb{R}^n$. On cherche à minimiser $f : x \mapsto \frac{1}{2}\langle Ax, x\rangle + \langle b, x\rangle$ quand [a] x parcourt $\mathbb{R}^n$.
Il existe une unique solution à ce problème, caractérisée par $\nabla f(\bar{x}) = 0$.
De plus, l'algorithme défini par : $x_0 \in \mathbb{R}^n$ et pour $k \geqslant 0$:

$$\begin{cases} d_k = -\nabla f(x_k) = -Ax_k - b \\ x_{k+1} = x_k + t_k d_k \end{cases}$$

où t_k est l'unique réel (positif) minimisant $g : t \mapsto f(x_k + td_k)$ sur $\mathbb{R}$
converge vers $\bar{x}$.

a. On enlève le c du livre qui n'apporte rien

Preuve :

Conditions de minimalité (question 1)

Soit $\bar{x}$ un point minimal, alors $\nabla f(\bar{x}) = 0$.
Or

$$\begin{aligned} f(x+h) &= \frac{1}{2}\langle A(x+h), x+h\rangle + \langle b, x+h\rangle \\ &= \frac{1}{2}\langle Ax, x\rangle + \langle b, x\rangle + \frac{1}{2}\left(\langle Ax, h\rangle + \langle Ah, x\rangle\right) + \langle b, h\rangle + \frac{1}{2}\langle Ah, h\rangle \\ &= \frac{1}{2}f(x) + \langle Ax, h\rangle + \langle b, h\rangle + \frac{1}{2}\langle Ah, h\rangle \qquad \text{car } A \text{ symétrique} \\ &= \frac{1}{2}f(x) + \langle Ax + b, h\rangle + o(\|h\|) \end{aligned}$$

Donc $\nabla f(x) = Ax + b$ et donc $\bar{x} = -A^{-1}b$.
Minimiser f revient à inverser une matrice, de manière peut-être plus efficace.
D'autre part,

$$\begin{aligned} \bar{f} := \min_{x \in \mathbb{R}^n} f(x) = f(\bar{x}) &= \frac{1}{2}\langle A\bar{x}, \bar{x}\rangle + \langle b, x\rangle \\ &= \frac{1}{2}\langle A\bar{x}, \bar{x}\rangle + \langle -A\bar{x}, \bar{x}\rangle \\ &= -\frac{1}{2}\langle A\bar{x}, \bar{x}\rangle \\ &= -\frac{1}{2}\langle A(-A^{-1}b), -A^{-1}b\rangle \\ &= -\frac{1}{2}\langle -b, -A^{-1}b\rangle \\ &= -\frac{1}{2}\langle A^{-1}b, b\rangle \end{aligned}$$

De plus, $\bar{x}$ est bien le minimum car f est strictement convexe [1], coercive [2] sur $\mathbb{R}^n$ donc il existe un unique minimum pour f.
Dans toute la suite de la preuve, on suppose $d_k \neq 0$ car sinon $Ax_k = -b$ et l'algorithme converge en temps fini.

1. car la matrice hessienne de f est A qui est définie positive.
2.

$$\begin{aligned} |f(x)| &\geqslant \frac{1}{2}\left|\langle Ax, x\rangle\right| - \left|\langle b, x\rangle\right| \\ &\geqslant \frac{1}{2}\lambda_{\min}\|x\|^2 - \|b\|\,\|x\| \\ &\xrightarrow[\|x\| \to \infty]{} \infty \end{aligned}$$

Calcul de t_k (question 2)

$$\begin{aligned}
g(t) = f(x_k + td_k) &= \frac{1}{2}\langle A(x_k+td_k), x_k+td_k\rangle + \langle b, x_k+td_k\rangle \\
&= \frac{1}{2}\langle Ax_k, x_k\rangle + \frac{1}{2}\langle Atd_k, x_k\rangle + \frac{1}{2}\langle Ax_k, td_k\rangle + \frac{1}{2}\langle Atd_k, td_k\rangle + \langle b, x_k\rangle + \langle b, td_k\rangle \\
&= f(x_k) + \langle \underbrace{Ax_k + b}_{-d_k}, td_k\rangle + \frac{1}{2}\langle Atd_k, td_k\rangle \\
&= f(x_k) - t\left\|d_k\right\|^2 + \frac{1}{2}t^2\langle Ad_k, d_k\rangle
\end{aligned}$$

Donc $g'(t) = -\left\|d_k\right\|^2 + t\langle Ad_k, d_k\rangle$ et $g'(t) = 0 \Leftrightarrow t = t_k = \dfrac{\left\|d_k\right\|^2}{\langle Ad_k, d_k\rangle} > 0$ (on peut diviser par $\langle Ad_k, d_k\rangle$ car $d_k \neq 0$ et A définie positive).

Calcul de l'erreur sur $f(x_k) - \bar{f}$ (questions 2 et 3)

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

Donc

$$f(x_{k+1}) - \bar{f} = f(x_k) - \bar{f} - \frac{1}{2}\frac{\left\|d_k\right\|^4}{\langle Ad_k, d_k\rangle} = \left(f(x_k) - \bar{f}\right)\left(1 - \frac{\left\|d_k\right\|^4}{2\langle Ad_k, d_k\rangle\left(f(x_k) - \bar{f}\right)}\right)$$

Or :

$$\begin{aligned}
\langle A^{-1}d_k, d_k\rangle &= \langle A^{-1}(Ax_k + b), Ax_k + b\rangle \\
&= \langle x_k + A^{-1}b, Ax_k + b\rangle \\
&= \langle x_k, Ax_k\rangle + \langle x_k, b\rangle + \underbrace{\langle A^{-1}b, Ax_k\rangle}_{\langle b, x_k\rangle} + \underbrace{\langle A^{-1}b, b\rangle}_{-2\bar{f}} \\
&= 2\left(\underbrace{\frac{1}{2}\langle x_k, Ax_k\rangle + \langle x_k, b\rangle}_{f(x_k)} - \bar{f}\right) = 2\left(f(x_k) - \bar{f}\right)
\end{aligned}$$

Donc finalement

$$f(x_{k+1}) - \bar{f} = \left(f(x_k) - \bar{f}\right)\left(1 - \frac{\left\|d_k\right\|^4}{\langle Ad_k, d_k\rangle\langle A^{-1}d_k, d_k\rangle}\right)$$

L'inégalité de Kantorovitch donne,

$$\frac{\left\|d_k\right\|^4}{\langle Ad_k, d_k\rangle\langle A^{-1}d_k, d_k\rangle} \geqslant 4\frac{c(A)}{(c(A)+1)^2}$$

Donc

$$f(x_{k+1}) - \bar{f} \leqslant \left(f(x_k) - \bar{f}\right)\left(\frac{c(A)-1}{c(A)+1}\right)^2 \leqslant \cdots \leqslant \left(f(x_0) - \bar{f}\right)\left(\frac{c(A)-1}{c(A)+1}\right)^{2k}$$

Calcul de l'erreur sur $\|x_k - \bar{x}\|$ (question 3)

Par ailleurs, on a

$$\begin{aligned}
\langle A(x_k - \bar{x}), x_k - \bar{x}\rangle &= \langle Ax_k, x_k - \bar{x}\rangle - \langle A\bar{x}, x_k - \bar{x}\rangle \\
&= \langle Ax_k, x_k\rangle - \langle Ax_k, \bar{x}\rangle - \langle A\bar{x}, x_k\rangle + \langle A\bar{x}, \bar{x}\rangle \\
&= \langle Ax_k, x_k\rangle + \langle Ax_k, A^{-1}b\rangle + \langle AA^{-1}b, x_k\rangle + \langle AA^{-1}b, A^{-1}b\rangle \\
&= \underbrace{\langle Ax_k, x_k\rangle + \langle x_k, b\rangle + \langle b, x_k\rangle}_{2f(x_k)} + \underbrace{\langle b, A^{-1}b\rangle}_{-2\bar{f}}
\end{aligned}$$

Donc

$$f(x_k) - \bar{f} = \frac{1}{2}\langle A(x_k - \bar{x}), x_k - \bar{x}\rangle \geqslant \frac{1}{2}\lambda_{\min} \|x_k - \bar{x}\|^2$$

Et finalement

$$\|x_k - \bar{x}\| \leqslant \sqrt{\frac{2\left(f(x_0) - \bar{f}\right)}{\lambda_{\min}}} \left(\frac{c(A) - 1}{c(A) + 1}\right)^k$$

Donc $x_k \underset{n \to +\infty}{\longrightarrow} \bar{x}$. ■

<u>Notes :</u>

✓ **A l'oral**, on ne fait pas Kantorovitch. 14'36 allure normale (énoncer Kanto avec le conditionnement !)
✓ Lorsque $n = 2$ x_k converge en zigzags vers $\bar{x}$.
✓ Plus $c(A)$ est proche de 1, plus la méthode converge rapidement. A l'inverse lorsque $c(A)$ est loin de 1, la méthode est très lente.
Le cas limite serait celui où $c(A) = 1$ (donc A homothétie), ce qui suppose $2(f(x) - \bar{f}) = \lambda \|x - \bar{x}\|^2$ auquel cas on atteint $\bar{x}$ dès la première itération. En effet, $d_0 = -Ax_0 - b = -\lambda x_0 + \lambda \bar{x}$, donc $t_0 = \frac{1}{\lambda}$ et on obtient bien $x_1 = \bar{x}$ (ou sinon on utilise la dernière majoration...).