Algorithme du gradient à pas optimal

Référence(s) :
– J.-B. Hiriart-Urruty - *Optimisation et analyse convexe*

Lemme 1 (*Kantorovitch*)

Soit $A \in \mathcal{S}_n^{++}(\mathbb{R})$; Soient λ_1 et λ_n ses plus petite et plus grande valeurs propres. Alors

$$\forall x \in \mathbb{R}^n, \langle Ax, x\rangle\langle A^{-1}x, x\rangle \leqslant \frac{1}{4}\left(\sqrt{\frac{\lambda_1}{\lambda_n}} + \sqrt{\frac{\lambda_n}{\lambda_1}}\right) \|x\|^4$$

Théorème 1

Soit $A \in \mathcal{S}_n^{++}(\mathbb{R})$; Soit $b \in \mathbb{R}^n$. Soit $f : \left| \begin{array}{ccc} \mathbb{R}^n & \longrightarrow & \mathbb{R} \\ x & \longmapsto & \frac{1}{2}\langle Ax, x\rangle + \langle b, x\rangle \end{array} \right.$.
On veut minimiser $f(x)$, quand x parcourt $\mathbb{R}^n$.
Il existe une unique solution $\bar{x}$ à ce problème, qui est caractérisée par $\nabla f(\bar{x}) = 0$
La suite définie par $\left\{ \begin{array}{l} x_0 \in \mathbb{R}^n \\ \forall k \in \mathbb{N}, x_{k+1} := x_k + t_k d_k \end{array} \right.$ *, où $d_k = -\nabla f(x_k)$ et t_k est l'unique réel minimisant la fonction $t \mapsto f(x_k + t_k d_k)$, converge vers $\bar{x}$.*

Étape 1

La solution $\bar{x}$ existe et est unique. De plus : $\bar{x} = -A^{-1}b$ et $f(\bar{x}) = -\frac{1}{2}\langle A^{-1}b, b\rangle$

Soit $\bar{x}$ un point minimal de f. Alors, nécessairement $\nabla f(\bar{x}) = 0$. Or, pour $x, h \in \mathbb{R}^n$, on a :

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

On en déduit que f est différentiable et $\forall x \in \mathbb{R}^n, \nabla f(x) = Ax + b$. De plus, la hessienne de f est A qui est définie positive, donc f est strictement convexe. On en déduit que $\bar{x}$ existe, est unique et même $\bar{x} = -A^{-1}b$. Du coup, on peut calculer :

$$\bar{f} = f(\bar{x}) = \frac{1}{2}\langle -b, -A^{-1}b\rangle + \langle b, -A^{-1}b\rangle = \frac{1}{2}\langle A^{-1}b, \rangle - \langle A^{-1}b, b\rangle = -\frac{1}{2}\langle A^{-1}b, b\rangle$$

Étape 2

On suppose que pour tout $k \in \mathbb{N}$, $d_k \neq 0$.

En effet, sinon on aurait $f(x_k) = f(\bar{x})$ et l'algorithme converge en temps fini.
On remarque que tant que $d_k \neq 0$, on $f(x_k) - \bar{f} > 0$

Étape 3

Pour tout $k \in \mathbb{N}$, $t_k = \frac{\|d_k\|^2}{\langle Ad_k, d_k\rangle}$

Pour tout $t \in \mathbb{R}$:

$$f(x_k + td_k) = f(x_k) + \frac{1}{2}t^2\langle Ad_k, d_k\rangle + t\langle \underbrace{Ax_k + b}_{=-d_k}, d_k\rangle = f(x_k) - t\|d_k\|^2 + \frac{t^2}{2}\langle Ad_k, d_k\rangle$$

Ainsi, $t \mapsto f(x_k + td_k)$ admet un unique minimum, qui est atteint en $t_k = \frac{\|d_k\|^2}{\langle Ad_k, d_k\rangle}$ (> 0 et bien défini car $A \in \mathcal{S}_n^{++}$ et $d_k \neq 0$)

Étape 4

Pour $k \in \mathbb{N}$, on a : $f(x_{k+1}) - \bar{f} = (f(x_k) - \bar{f})\left(1 - \frac{\|d_k\|^4}{\langle A^{-1}d_k, d_k\rangle\langle Ad_k, d_k\rangle}\right)$

Soit $k \in \mathbb{N}$. Alors

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

On en déduit

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

Mais on remarque que

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

Et donc :

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

Étape 5

On obtient une erreur : $f(x_k) - \bar{f} \leqslant (f(x_0) - \bar{f})\left(\frac{c(A)-1}{c(A)+1}\right)^{2k}$

On pose $c(A) = \frac{\lambda_n}{\lambda_1}$, où λ_1 est la plus petit valeur propre de A et λ_n est la plus grande.
On applique l'inégalité de Kantorovitch au résultat précédent :

$$\begin{aligned} f(x_{k+1} - \bar{f}) &\leqslant (f(x_k) - \bar{f})\left(1 - \frac{4}{\left(\sqrt{c(A)} + \frac{1}{\sqrt{c(A)}}\right)^2}\right) \\ &= (f(x_k) - \bar{f})\left(1 - \frac{4c(A)}{(c(A)+1)^2}\right) \\ &\leqslant (f(x_k) - \bar{f})\left(\frac{c(A)-1}{c(A)+1}\right)^2 \end{aligned}$$

Puis par récurrence on obtient, pour tout $k \in \mathbb{N}$:

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

Étape 6

Finalement $\|x_k - \bar{x}\| \leqslant \sqrt{\frac{2}{\lambda_1}(f(x_0 - \bar{f})}\left(\frac{c(A)-1}{c(A)+1}\right)^k$; et la suite (x_k) converge vers $\bar{x}$

$$
\begin{aligned}
\|x_k - \bar{x}\|^2 &\leqslant \frac{1}{\lambda_1}\langle A(x_k - \bar{x}), x_k - \bar{x}\rangle \\
&= \frac{1}{\lambda_1}\left(\langle Ax_k, x_k\rangle - \langle Ax_k, \bar{x}\rangle - \langle A\bar{x}, x_k\rangle + \langle A\bar{x}, \bar{x}\rangle\right) \\
&= \frac{1}{\lambda_1}\left(\langle Ax_k, x_k\rangle - 2\langle x_k, A\bar{x}\rangle - 2\bar{f}\right) \\
&= \frac{2}{\lambda_1}(f(x_k) - \bar{f})
\end{aligned}
$$

Ainsi, on a :

$$
\|x_k - \bar{x}\| \leqslant \sqrt{\frac{2}{\lambda_1}}\sqrt{f(x_k) - \bar{f}} \leqslant \sqrt{\frac{2}{\lambda_1}(f(x_0) - \bar{f})}\,(fracc(A) - 1c(A) + 1)^k
$$

Comme $|fracc(A) - 1c(A) + 1| < 1$, on en déduit que la suite (x_k) converge vers $\bar{x}$.[1]

Preuve du lemme de Kantorovitch

1. La convergence est rapide lorsque $c(A)$ est proche de 1 et lente dans le cas contraire.