Algorithme de gradient à pas optimal

Rappelons que, pour $J : \mathbb{R} \to \mathbb{R}$ différentiable, l'algorithme du gradient à pas optimal est défini par la suite :

$$u_0 \in \mathbb{R}^n \quad \text{et} \quad \forall\, k \in \mathbb{N},\ u^{k+1} = u^k - \rho^k \nabla J(u^k) \quad \text{où} \quad J\left(u^k - \rho^k \nabla J(u^k)\right) = \inf_{\rho \in \mathbb{R}^n} J\left(u^k - \rho \nabla J(u^k)\right)$$

Théorème 1. *Si J est α-convexe et différentiable, et que ∇J est L-lipschitzienne, alors la méthode de gradient à pas optimal converge vers l'unique minimum de J.*

Démonstration.

Étape 1 : Montrons que le problème de minimisation est bien posé.

Commençons par noter que l'α-convexité de J assure l'existence d'un unique minimum global $u \in \mathbb{R}^n$, caractérisé par l'équation d'Euler $\nabla J(u) = 0$, le problème de minimisation est donc bien posé.

Étape 2 : Montrons que le problème de minimisation intermédiaire est bien posé.

On peut supposer que, pour tout $k \in \mathbb{N}$, $\nabla J(u^k) \neq 0$. En effet, si $\nabla J(u^k) = 0$ alors la suite est stationnaire, et donc convergente en un nombre fini d'itérations. On considère maintenant la fonction suivante, que l'on cherche à minimiser à chaque étape :

$$\varphi_k : \left|\begin{array}{ccc} \mathbb{R} & \longrightarrow & \mathbb{R} \\ \rho & \longmapsto & J(u^k - \rho \nabla J(u^k)) \end{array}\right.$$

Cette fonction est dérivable, et on a :

$$\varphi_k'(\rho) = -\left\langle \nabla J(u^k - \rho \nabla J(u^k)), \nabla J(u^k) \right\rangle$$

De plus, pour tous $\rho, \sigma \in \mathbb{R}$, on a :

$$\begin{aligned}
(\varphi_k'(\rho) - \varphi_k'(\sigma))(\rho - \sigma) &= \left\langle \nabla J(u^k - \sigma \nabla J(u^k)) - \nabla J(u^k - \rho \nabla J(u^k)), \nabla J(u^k) \right\rangle (\rho - \sigma) \\
&= \left\langle \nabla J(u^k - \sigma \nabla J(u^k)) - \nabla J(u^k - \rho \nabla J(u^k)), (u^k - \sigma \nabla J(u^k)) - (u^k - \rho \nabla J(u^k)) \right\rangle \\
&\geqslant \alpha \left\| (u^k - \sigma \nabla J(u^k)) - (u^k - \rho \nabla J(u^k)) \right\|^2 \\
&\geqslant \alpha |\rho - \sigma|^2 \left\| \nabla J(u^k) \right\|^2
\end{aligned}$$

Donc φ_k est $\alpha \left\| \nabla J(u^k) \right\|^2$-convexe, et le problème de minimisation intermédiaire admet une unique solution ρ^k caractérisée par l'équation d'Euler :

$$\varphi_k'(\rho) = -\left\langle \nabla J(u^k - \rho \nabla J(u^k)), \nabla J(u^k) \right\rangle = \left\langle \nabla J(u^{k+1}), \nabla J(u^k) \right\rangle = 0$$

En particulier, deux directions de descente successives $\nabla J(u^k)$ et $\nabla J(u^{k+1})$ son orthogonales.

Étape 3 : Montrons que la suite $(J(u^k))_{k \in \mathbb{N}}$ est convergente.

Par α-convexité de J, on a :

$$\begin{aligned}
J(u^k) &\geqslant J(u^{k+1}) + \left\langle \nabla J(u^{k+1}), u^k - u^{k+1} \right\rangle + \frac{\alpha}{2} \left\| u^k - u^{k+1} \right\|^2 \\
&\geqslant J(u^{k+1}) + \rho^k \left\langle \nabla J(u^{k+1}), \nabla J(u^k) \right\rangle + \frac{\alpha}{2} \left\| u^k - u^{k+1} \right\|^2 \\
&\geqslant J(u^{k+1}) + \frac{\alpha}{2} \left\| u^k - u^{k+1} \right\|^2
\end{aligned}$$

La suite $(J(u^k))_{k \in \mathbb{N}}$ est donc décroissante et minorée par $J(u)$ par définition : elle est convergente.

Étape 4 : Montrons que la suite $(u^k)_{k\in\mathbb{N}}$ converge vers u.

Comme la suite $(J(u^k))_{k\in\mathbb{N}}$ est convergente, $J(u^k) - J(u^{k+1})$ tend vers 0 quand k tends vers $+\infty$, et donc $u^k - u^{k+1}$ également par l'inégalité précédente. Par α-convexité de J, on a :

$$\begin{aligned}
\alpha \left\|u^k - u\right\|^2 &\leqslant \left\langle \nabla J(u^k) - \nabla J(u), u^k - u \right\rangle \\
&= \left\langle \nabla J(u^k), u^k - u \right\rangle \\
&\leqslant \left\|\nabla J(u^k)\right\| \left\|u^k - u\right\|
\end{aligned}$$

Ainsi $\alpha \left\|u^k - u\right\| \leqslant \left\|\nabla J(u^k)\right\|$. Pour montrer la convergence de $(u^k)_{k\in\mathbb{N}}$ vers u, il suffit donc de montrer celle de $(\nabla J(u^k))_{k\in\mathbb{N}}$ vers 0. Or, on a :

$$\begin{aligned}
\left\|\nabla J(u^k)\right\|^2 &= \left\langle \nabla J(u^k), \nabla J(u^k) \right\rangle \\
&= \left\langle \nabla J(u^k) - \nabla J(u^{k+1}), \nabla J(u^k) \right\rangle \\
&\leqslant \left\|\nabla J(u^k) - \nabla J(u^{k+1})\right\| \left\|\nabla J(u^k)\right\| \\
&\leqslant L \left\|u^k - u^{k+1}\right\| \left\|\nabla J(u^k)\right\|
\end{aligned}$$

Donc $\left\|\nabla J(u^k)\right\| \leqslant L \left\|u^k - u^{k+1}\right\|$, et $(\nabla J(u^k))_{k\in\mathbb{N}}$ tend vers 0 lorsque k tend vers $+\infty$. On en déduit alors que la suite $(u^k)_{k\in\mathbb{N}}$ converge vers u.

$\square$

Références

[Cia88] Philippe Ciarlet. *Introduction à l'analyse numérique et à l'optimisation.* Masson, 1988