44. Méthode du gradient à pas optimal [FGN12, §1.21, p39–41]

ÉNONCÉ

THÉORÈME. [MÉTHODE DE GRADIENT À PAS OPTIMAL]
Soit $f : \mathbb{R}^p \longrightarrow \mathbb{R}$ elliptique, c'est-à-dire f est $\mathcal{C}^1$ et telle qu'il existe $\alpha > 0$ satisfaisant

$$\forall x, y \in \mathbb{R}^p, \langle \nabla f(x) - \nabla f(y) \mid x - y\rangle \geq \alpha \|x - y\|^2$$

Alors la suite $(x_n)_{n\in\mathbb{N}}$ définie par $x_0 \in \mathbb{R}^p$ et $x_{n+1} = x_n - \rho_n \nabla f(x_n)$ où $\rho_n = \operatorname{argmin}_{\rho>0} f(x_n - \rho\nabla f(x_n))$ converge vers l'unique minimum global de f.

DÉVELOPPEMENT

Soit f satisfaisant les hypothèses du théorème.

1. Montrons d'abord que

$$\forall x, y \in \mathbb{R}^p, f(y) - f(x) \geq \langle \nabla f(x) \mid y - x\rangle + \frac{\alpha}{2} \|x - y\|^2$$

En effet, soient $x, y \in \mathbb{R}^p$. On a par la formule de TAYLOR avec reste intégral :

$$\begin{aligned}
f(y) - f(x) &= \int_0^1 \langle \nabla f(x + t(y - x)) \mid y - x\rangle dt \\
&= \int_0^1 \langle \nabla f(x) \mid y - x\rangle dt + \int_0^1 \langle \nabla f(x + t(y - x)) - \nabla f(x) \mid y - x\rangle dt \\
&= \langle \nabla f(x) \mid y - x\rangle + \int_0^1 \frac{1}{t}\langle \nabla f(x + t(y - x)) - \nabla f(x) \mid t(y - x)\rangle dt \\
&\geq \langle \nabla f(x) \mid y - x\rangle + \int_0^1 \frac{1}{t}\alpha \|t(y - x)\|^2 dt \\
&\geq \langle \nabla f(x) \mid y - x\rangle + \frac{\alpha}{2} \|y - x\|^2
\end{aligned}$$

2. f est continue et coercive puisque :

$$f(y) \geq f(0) + \langle \nabla f(0) \mid y\rangle + \frac{\alpha}{2} \|y\|^2 = O(\|y\|) + \frac{\alpha}{2} \|y\|^2 \underset{\|y\|\to+\infty}{\longrightarrow} +\infty$$

Donc f atteint son minimum global en un point $x_\star$.
Si x_1, x_2 sont deux minimums locaux, on a $0 \geq \langle 0 \mid x_1 - x_2\rangle + \frac{\alpha}{2} \|x_1 - x_2\|^2$, ce qui implique $x_1 = x_2$. Ainsi $x_\star$ est le seul minimum global et aussi local de f.

Notons par ailleurs que si $\nabla f(x) = 0$, alors $f(x_\star) - f(x) \geq \frac{\alpha}{2} \|x_\star - x\|^2$ et donc $x = x_\star$.

3. Soit $x \in \mathbb{R}^p$ tel que $\nabla f(x) \neq 0$ ($x \neq x_\star$). Posons $\varphi_x : \rho \in \mathbb{R}_+ \longmapsto f(x - \rho\nabla f(x))$. C'est une fonction $\mathcal{C}^1$ qui tend vers $+\infty$ en $+\infty$. Elle atteint donc son minimum en un point ρ_x annulant la dérivée de φ_x. Calculons :

$$0 = \varphi_x'(\rho_x) = -\langle \nabla f(x + \rho_x\nabla f(x)) \mid \nabla f(x)\rangle$$

Donc pour tout $\rho \neq \rho_x$:

$$\varphi_x(\rho) - \varphi_x(\rho_x) \geq \underbrace{\langle \nabla f(x + \rho_x\nabla f(x)) \mid (\rho - \rho_x)\nabla f(x)\rangle}_{=0} + \frac{\alpha}{2} \|(\rho - \rho_x)\nabla f(x)\|^2 > 0$$

Donc ρ_x est l'unique minimum de φ_x.

4. Notre suite $(x_n)_{n\in\mathbb{N}}$ est ainsi bien définie tant que $x_n \neq x_\star$.
Si elle atteint $x_\star$, elle est stationnaire en ce point et le résultat est clair.

Si ce n'est pas le cas, montrons que $(x_n)_{n\in\mathbb{N}}$ converge tout de même vers $x_\star$.

La suite $(f(x_n))_{n\in\mathbb{N}}$ est strictement décroissante et minorée, donc converge.
Remarquons par ce qui précède que $\nabla f(x_n)$ et $\nabla f(x_{n+1})$ sont orthogonaux pour tout $n \in \mathbb{N}$. Comme $\nabla f(x_n)$ et $x_{n+1} - x_n$ sont colinéaires, on a donc :

$$f(x_n) - f(x_{n+1}) \geq \frac{\alpha}{2} \|x_{n+1} - x_n\|^2$$

Donc $\|x_{n+1} - x_n\| \underset{n\to+\infty}{\longrightarrow} 0$.

Or par coercivité, la suite $(x_n)_{n\in\mathbb{N}}$ vit dans un compact de $\mathbb{R}^p$, donc admet une valeur d'adhérence x. Par continuité de ∇f, on a $\nabla f(x_{\psi(n)}) \underset{n\to+\infty}{\longrightarrow} \nabla f(x)$. Et comme $(x_{\psi(n)+1})_{n\in\mathbb{N}}$ converge aussi vers x puisque $\|x_{n+1} - x_n\| \underset{n\to+\infty}{\longrightarrow} 0$, on a finalement :

$$0 = \langle \nabla f(x_{\psi(n)}) \mid \nabla f(x_{\psi(n)+1})\rangle \underset{n\to+\infty}{\longrightarrow} \|\nabla f(x)\|^2$$

de sorte que nécessairement $x = x_\star$.

COMMENTAIRES

On peut prendre n'importe quel produit scalaire sur $\mathbb{R}^p$!

Bien penser à faire un dessin en annexe.