Descente de gradient

Geoffrey Deperle

Leçons associées :
— 215 : Applications différentiables définies sur un ouvert de $\mathbb{R}^n$. Exemples et applications.
— 223 : Extremums : existence, caractérisation, recherche. Exemples et applications.
— 226 : Suites vectorielles et réelles définies par une relation de récurrence $u_{n+1} = f(u_n)$. Exemples. Applications.
— 229 : Fonctions monotones. Fonctions convexes. Exemples et applications.
— 253 : Utilisation de la notion de convexité en analyse.

Le but de ce développement est de montrer le théorème suivant :

Théorème. *Soit $G : \mathbb{R}^n \to \mathbb{R}$ de classe C^2 et $\alpha > 0$ tels que pour tout $x, y \in \mathbb{R}$,*

$$G(y) - G(x) \geq \langle \nabla G(x), y - x \rangle + \frac{\alpha}{2}||y - x||^2$$

G admet une borne inférieure sur $\mathbb{R}^n$ et la suite (x_k) définie par :

$$\begin{cases} x_0 \in \mathbb{R}^n \\ \forall k \in \mathbb{N}, x_{k+1} = x_k + \lambda_k \nabla G(x_k) \end{cases}$$

où λ_k est tel que $x_k + \lambda_k \nabla G(x_k)$ est un minimum de G sur la droite $x_k + \mathbb{R}\nabla G(x_k)$.
La suite (x_k) converge vers cette borne inférieure.

Preuve :

Étape 1 : Montrons l'existence de $\inf G$

Pour $x = 0$, et $y \in \mathbb{R}^n$, l'hypothèse donne

$$G(y) \geq G(0) + \langle \nabla G(0), y \rangle + \frac{\alpha}{2}||y||^2 = \frac{\alpha}{2}||y||^2 + O(||y||)$$

donc $\lim\limits_{||y|| \to +\infty} G(y) = +\infty$ et G est coercive.
Rappelons les arguments justifiant qu'une fonction coercive sur $\mathbb{R}^n$ admet un minimum global :
Il existe $A > 0$ tel que pour $||x|| > A$, $G(x) > G(0)$. Comme $\mathbb{R}^n$ est de dimension finie, la boule de centre 0 et de rayon A est un compact sur lequel G est continue donc borné et atteint son minimum en un point a.
Pour tout $||x|| > A$, on a $G(x) > G(0) \geq G(a)$ donc G admet un minimum global sur $\mathbb{R}^n$ en a.
Ce point est nécessairement un point critique donc $\nabla G(a) = 0$. Ainsi, pour $y \neq a$, on a

$$G(y) \geq G(a) + \frac{\alpha}{2}||y - a||^2 > G(a)$$

donc G admet son minimum uniquement en a.
Ce minimum est caractérisé par la condition $\nabla G(a) = 0$. En effet, soit $x \in \mathbb{R}^n$ tel que $\nabla G(x) = 0$, alors $G(a) - G(x) \geq \frac{\alpha}{2}||a - x||^2$ d'où $G(x) \leq G(a) - \frac{\alpha}{2}||a - x||^2$ donc comme le minimum de G est atteint en un unique point a, on a $||a - x||^2 = 0$ d'où $x = a$.

Étape 2 : Montrons que $\forall x \in \mathbb{R}^n$, $t \mapsto G(x + t\nabla G(x))$ atteint son minimum en un unique point

Soit $\varphi_x : t \mapsto G(x + t\nabla G(x))$, φ_x est C^1 et par composition des limites, $\varphi_x(t) \xrightarrow[t \to +\infty]{} +\infty$ donc φ_x possède un minimum en un point τ vérifiant $\varphi_x'(\tau) = 0$.
Or,

$$\varphi_x'(\tau) = \mathrm{d}G(x + \tau\nabla G(x)).\nabla G(x) = \langle \nabla G(x + \tau\nabla G(x)), \nabla G(x) \rangle$$

donc $\nabla G(x + \tau\nabla G(x)) \perp \nabla G(x)$ car $\varphi_x'(\tau) = 0$.
Montrons l'unicité, soit $t \in \mathbb{R}^n$, si $t \neq \tau$, on a

$$\begin{aligned}\varphi_x(t) = G(x + t\nabla G(x)) &\geq G(x + \tau\nabla G(x)) + \langle \nabla G(x + t\nabla G(x)), (\tau - t)\nabla G(x) \rangle + \frac{\alpha}{2}|\tau - t|^2||\nabla G(x)||^2 \\ &> \varphi_x(\tau)\end{aligned}$$

donc φ_x atteint son minimum en un unique point.

Étape 3 : Montrons la convergence de (x_k)

On a $\forall k \in \mathbb{N}, \nabla G(x_k) \perp \nabla G(x_{k+1})$.
Si l'un des $x_k = a$, alors la suite est stationnaire et est égale à a donc le résultat est acquis.
Sinon, la suite $(G(x_k))$ est décroissante par construction et minorée par $G(a)$ donc elle converge vers un réel l donc en particulier $G(x_k) - G(x_{k+1}) \xrightarrow[k \to +\infty]{} 0$. De plus,

$$\begin{aligned}G(x_k) - G(x_{k+1}) &\geq \langle \nabla G(x_{k+1}), x_k - x_{k+1} \rangle + \frac{\alpha}{2}||x_k - x_{k+1}||^2 \\ &= \lambda_k \langle \nabla G(x_{k+1}), \nabla G(x_k) \rangle + \frac{\alpha}{2}||x_k - x_{k+1}||^2 \\ &= \frac{\alpha}{2}||x_k - x_{k+1}||^2\end{aligned}$$

donc $(x_k - x_{k+1})$ converge également vers 0 par théorème d'encadrement.

Comme il existe $A > 0$ tel que si $||x|| > A, G(x) > G(x_0)$, par décroissance de $(G(x_k))$, on a nécessairement pour tout $k \in \mathbb{N}$, $x_k \in \mathrm{B}(0, A)$.
Ainsi, comme (x_k) est une suite définie sur un espace de dimension finie, d'après le théorème de Bolzano-Weierstrass, (x_k) possède une valeur d'adhérence : il existe $\phi : \mathbb{N} \to \mathbb{N}$ strictement croissante et $x \in \mathbb{R}^n$ tel que $\lim\limits_{k \to +\infty} x_{\phi(k)} = x$ et par continuité de ∇G, on a $\nabla G(x_{\phi(k)}) \xrightarrow[k \to +\infty]{} \nabla G(x)$.
Or, $(x_{\phi(k)+1})$ converge vers x car $||x_{\phi(k)+1} - x_{\phi(k)}|| \leq \frac{2}{\alpha}(G(x_{\phi(k)+1}) - G(x_{\phi(k)}))$.
D'où

$$0 = \langle G(x_{\phi(k)+1}), G(x_{\phi(k)}) \rangle \xrightarrow[k \to +\infty]{} ||\nabla G(x)||^2$$

d'où $\nabla G(x) = 0$ et $x = a$.
Donc la suite (x_k) admet une unique valeur d'adhérence, donc la suite (x_k) converge vers a. □

Références

[1] Serge Francinou, Hervé Gianella et Serge Nicolas. *Oraux X-ENS Analyse 4*. Cassini, 2016.