

Méthode de descente du gradient

Aurélien Guerder

2023-2024

Source : LESESVRE, MONTAGNON, LE BARBENCHON, PIERRON, 131 développements pour l'oral

Attention, la preuve dans la source n'a ni queue ni tête, mais j'ai repris les idées pour la corriger.

Théorème 1. Soit $f : \mathbb{R}^n \rightarrow \mathbb{R}$ de classe C^1 strictement convexe dont le gradient est L -lipschitzien, et admettant un minimum en x^* (unique car f est strictement convexe). pour $t > 0$ et $x_0 \in \mathbb{R}^n$, on définit

$$\forall k \in \mathbb{N}, x_{k+1} = x_k - t \nabla f(x_k)$$

Alors, si $t \leq \frac{1}{L}$, $x_k \xrightarrow{k \rightarrow +\infty} x^*$.

Lemme 1. f est coercive, c'est-à-dire $\forall A \in \mathbb{R}, \exists r > 0, \forall x \in \mathbb{R}^n, \|x\| > r \Rightarrow \|f(x)\| > A$.

Remarque : on peut admettre ce lemme, ou bien ajouter la coercivité de f en hypothèse du théorème.

Démonstration. Posons $K = \{x \in \mathbb{R}^n | f(x) \leq A\} = f^{-1}([-\infty, A])$. On va montrer que K est borné. Si $K = \emptyset$, c'est clair. Supposons K non vide. Alors, $x^* \in K$.

Supposons K non borné : il existe une suite $(y_k)_k$ de K telle que $\|y_k\| \rightarrow +\infty$ et

$\forall k \in \mathbb{N}, \|y_k - x^*\| > 1$.

Posons $\forall k \in \mathbb{N}, z_k = x^* + \frac{y_k - x^*}{\|y_k - x^*\|} \in C(x^*, 1)$. On a

$$\begin{aligned} f(z_k) &= f\left(x^* + \frac{y_k - x^*}{\|y_k - x^*\|}\right) \\ &\leq \frac{1}{\|y_k - x^*\|} f(y_k) + \left(1 - \frac{1}{\|y_k - x^*\|}\right) f(x^*) && \text{par la convexité} \\ &\xrightarrow{k \rightarrow +\infty} f(x^*) && \text{car } f(y_k) \text{ est bornée} \end{aligned}$$

On a donc $\inf_{C(x^*, 1)} f \leq f(x^*)$ et c'est un minimum car f est continue sur le compact $C(x^*, 1)$. D'où $f(x^*) = \min_{C(x^*, 1)} f$ car $f(x^*)$ est le minimum de f . Ceci est absurde car le minimum est atteint en un unique point qui est x^* . □

Lemme 2. Pour $t \leq \frac{1}{L}$, on a $\forall x \in \mathbb{R}^n, f(x - t \nabla f(x)) - f(x) \leq -\frac{t}{2} \|\nabla f(x)\|^2$.

Démonstration. Soit $x, y \in \mathbb{R}^n$.

Si on pose $g(t) = f(x + t(y - x))$, on a $g'(t) = \langle \nabla f(x + t(y - x)), y - x \rangle$.

Le théorème fondamental de l'analyse donne alors :

$$\begin{aligned}
 f(y) - f(x) &= \int_0^1 \langle \nabla f(x + t(y - x)), y - x \rangle dt \\
 &= \int_0^1 \langle \nabla f(x), y - x \rangle dt + \int_0^1 \langle \nabla f(x + t(y - x)) - \nabla f(x), y - x \rangle dt \\
 &= \langle \nabla f(x), y - x \rangle + \int_0^1 \langle \nabla f(x + t(y - x)) - \nabla f(x), y - x \rangle dt \\
 &\leq \langle \nabla f(x), y - x \rangle + \int_0^1 \|\nabla f(x + t(y - x)) - \nabla f(x)\| \cdot \|y - x\| dt \\
 &\leq \langle \nabla f(x), y - x \rangle + L\|y - x\| \int_0^1 \|t(y - x)\| dt \\
 &= \langle \nabla f(x), y - x \rangle + \frac{L}{2} \|y - x\|^2
 \end{aligned}$$

Maintenant, en prenant $y = x - t\nabla f(x)$,

$$\begin{aligned}
 f(x - t\nabla f(x)) - f(x) &\leq -t\|\nabla f(x)\|^2 + \frac{Lt^2}{2} \|\nabla f(x)\|^2 \\
 &= -t \left(1 - \frac{Lt}{2}\right) \|\nabla f(x)\|^2 \\
 &\leq -\frac{t}{2} \|\nabla f(x)\|^2
 \end{aligned}$$

pour $t \leq \frac{1}{L}$. □

Venons-en à la preuve du Théorème 1 :

Démonstration. Par le lemme 2, on a

$$\forall k \in \mathbb{N}, f(x_{k+1}) - f(x_k) \leq -\frac{t}{2} \|\nabla f(x_k)\|^2 \leq 0 \quad (1)$$

$(f(x_k))_k$ est alors décroissante, et minorée par $f(x^*)$ donc elle converge.

Dans l'inégalité ci-dessus, faisons tendre k vers l'infini : on obtient alors que $\nabla f(x_k) \rightarrow 0$.

Posons $K = f^{-1}(]-\infty, f(x_0)])$, qui est compact car fermé et borné (lemme 1). $(x_k)_k$ est alors une suite de K .

Si $l \in K$ en est une valeur d'adhérence, par continuité de ∇f , on obtient $\nabla f(l) = \lim_{k \rightarrow +\infty} \nabla f(x_k) = 0$, donc $l = x^*$ car c'est l'unique point critique de f qui est strictement convexe. D'où (x_k) a une unique valeur d'adhérence et est dans un compact, d'où $(x_k)_k$ converge vers x^* . □

Théorème 2. On a la majoration $\forall k \in \mathbb{N}^*, f(x_k) - f(x^*) \leq \frac{1}{2tk} \|x_0 - x^*\|^2$.

Remarque : Si le développement est déjà trop long, on peut laisser de côté cette preuve. Il est quand même bon de mettre ce résultat dans le plan, car il donne une vitesse de convergence : cela montre qu'on sait que c'est une des problématiques des méthodes d'approximations numériques.

Démonstration. Par la convexité de f , on a $f(x) \leq f(x^*) + \langle \nabla f(x), x - x^* \rangle$. En l'introduisant dans (1) avec $x = x_k$:

$$f(x_{k+1}) \leq f(x^*) + \langle \nabla f(x_k), x_k - x^* \rangle - \frac{t}{2} \|\nabla f(x_k)\|^2$$

Donc :

$$\begin{aligned} f(x_{k+1}) - f(x^*) &\leq \frac{1}{2t} (2t \langle \nabla f(x_k), x_k - x^* \rangle - t^2 \|\nabla f(x_k)\|^2) \\ &\leq \frac{1}{2t} (\|x_k - x^*\|^2 - \|x_k - x^* - t \nabla f(x_k)\|^2) \\ &\leq \frac{1}{2t} (\|x_k - x^*\|^2 - \|x_{k+1} - x^*\|^2) \end{aligned}$$

En sommant de 0 à $k-1$,

$$\sum_{i=0}^{k-1} (f(x_{i+1}) - f(x^*)) \leq \frac{1}{2t} \sum_{i=0}^{k-1} (\|x_i - x^*\|^2 - \|x_{i+1} - x^*\|^2) = \frac{1}{2t} (\|x_0 - x^*\|^2 - \|x_k - x^*\|^2)$$

Et comme $(f(x_k))_k$ est décroissante,

$$f(x_k) - f(x^*) \leq \frac{1}{k} \sum_{i=0}^{k-1} (f(x_{i+1}) - f(x^*)) \leq \frac{1}{2kt} (\|x_0 - x^*\|^2 - \|x_k - x^*\|^2) \leq \frac{1}{2tk} \|x_0 - x^*\|^2$$

□