Gradient à pas optimal

Léo Gayral

2017-2018

ref : FGN – Oraux X-ENS, Analyse 4 – p.39

Définition 1 (Fonction fortement convexe)**.** Soient $\Phi \in \mathcal{C}^0\left(\mathbb{R}^d, \mathbb{R}\right)$ et $\alpha > 0$. On dit que Φ est α-convexe lorsque, pour $x, y \in \mathbb{R}^n$ et $t \in]0,1[$ quelconques :

$$\Phi(tx + (1-t)y) \leq t\Phi(x) + (1-t)\Phi(y) - \alpha \times t(1-t)\left\|x-y\right\|^2 .$$

Une telle fonction est en particulier strictement convexe.

Lemme 1. Soit $\Phi \in \mathcal{C}^1\left(\mathbb{R}^d, \mathbb{R}\right)$. On a équivalence entre :
- Φ est α-convexe,
- $\forall x, y \in \mathbb{R}^d, \Phi(y) \geq \Phi(x) + \langle \nabla\Phi(x), y-x\rangle + \alpha\left\|y-x\right\|^2$.

Démonstration.
Supposons Φ α-convexe. Dans ce cas, on a :

$$\Phi(x + t(y-x)) \leq \Phi(x) + t[\Phi(y) - \Phi(x)] - \alpha \times t(1-t)\left\|y-x\right\|^2$$

$$\frac{\Phi(x+t(y-x)) - \Phi(x)}{t} \leq \Phi(y) - \Phi(x) - \alpha \times (1-t)\left\|y-x\right\|^2$$

donc en passant à la limite $t \to 0^+$ on en déduit :

$$\partial_t[\Phi(x+t(y-x))]_{t=0} = \langle \nabla\Phi(x), y-x\rangle \leq \Phi(y) - \Phi(x) - \alpha\left\|y-x\right\|^2 .$$

Réciproquement, posons $z = tx + (1-t)y$. On a $z - x = (1-t)(y-x)$, d'où la majoration :

$$\begin{array}{rcl}\Phi(z) & \leq & \Phi(x) + \langle \nabla\Phi(z), z-x\rangle - \alpha\left\|z-x\right\|^2 \\ & \leq & \Phi(x) + (1-t)\langle \nabla\Phi(z), y-x\rangle - \alpha \times (1-t)^2\left\|y-x\right\|^2\end{array}$$

et de même $z-y=-t(y-x)$ d'où :

$$\begin{array}{rcl}\Phi(z) & \leq & \Phi(y)+\langle\nabla\Phi(z),z-y\rangle-\alpha\left\|z-y\right\|^2\\ & \leq & \Phi(y)-t\left\langle\nabla\Phi(z),y-x\right\rangle-\alpha\times t^2\left\|y-x\right\|^2\end{array}$$

donc, en considérant la moyenne pondérée des deux majorations :

$$\begin{array}{rcl}\Phi(z) & \leq & t\Phi(x)+(1-t)\Phi(y)+[t(1-t)-t(1-t)]\left\langle\nabla\Phi(z),y-x\right\rangle\\ & & -\alpha\left\|z-y\right\|^2\times[t(1-t)^2+(1-t)t^2]\\ & \leq & t\Phi(x)+(1-t)\Phi(y)-\alpha\left\|y-x\right\|^2\times t(1-t)\end{array}$$

c'est-à-dire le résultat voulu. □

Théorème 1. Soit $\Phi\in\mathcal{C}^1\left(\mathbb{R}^d,\mathbb{R}\right)$ une fonction α-convexe. Φ admet un unique minimum (global *et* local) atteint en x^*.

Soit $x_0\in\mathbb{R}^d$ fixé. On considère la suite $x_{n+1}=x_n+\lambda_n\nabla\Phi\left(x_n\right)$, où :

$$\lambda_n=\left\{\begin{array}{ll}\underset{\mu\in\mathbb{R}}{\arg\min}\,\Phi\left[x_n+\mu\nabla\Phi\left(x_n\right)\right] & \text{si}\quad x_n\neq x^*\\ 0 & \text{si}\quad x_n=x^*\end{array}\right.$$

de sorte que la suite stationne si elle atteint effectivement x^*. Pour tout x_0, cette suite est bien définie et $x_n\to x^*$.

Démonstration.
D'après le lemme, on a :

$$\Phi(x)\geq\Phi(0)+\langle\nabla\Phi(0),x\rangle+\alpha\left\|x\right\|^2$$

donc $\Phi(x)=\Omega\left(\left\|x\right\|^2\right)$ est une fonction coercive. Par continuité sur le compact $\Phi^{-1}\left(]-\infty,\Phi(0)]\right)$, Φ admet un minimum global en $x^*\in\mathbb{R}^d$.

Ce minimum global est *a fortiori* un minimum local, donc $\nabla\Phi\left(x^*\right)=0$. En outre, on a l'inégalité :

$$\Phi\left(x^*\right)\geq\Phi(y)+\left\langle\nabla\Phi(y),x^*-y\right\rangle+\alpha\left\|y-x^*\right\|^2$$

donc $\left\langle\nabla\Phi(y),y-x^*\right\rangle\geq\alpha\left\|y-x^*\right\|^2>0$ dès que $y\neq x^*$. En particulier, $\nabla\Phi(y)\neq0$ donc x^* est le seul point critique, le seul minimum local.

On considère la fonction $\varphi_n(\lambda)=\Phi\left(x_n+\nabla\Phi\left(x_n\right)\times\lambda\right)$. On a :

$$\begin{array}{rcl}\varphi_n(t\lambda+(1-t)\mu) & = & \Phi\left(t\left[x_n+\lambda\nabla\Phi\left(x_n\right)\right]+(1-t)\left[x_n+\mu\nabla\Phi\left(x_n\right)\right]\right)\\ & \leq & t\varphi_n(\lambda)+(1-t)\varphi_n(\mu)-\alpha\left\|\nabla\Phi\left(x_n\right)\right\|\times t(1-t)\left|\lambda-\mu\right|\end{array}$$

donc φ_n est une fonction $\alpha \times \|\nabla\Phi\,(x_n)\|$-convexe. Dès lors que $x_n \neq x^*$, on a bien φ_n fortement convexe. Par ce qui précède, φ_n admet un unique minimum global-local en λ_n. La suite (x_n) est donc bien définie.

L'application $\varphi_n \in \mathcal{C}^1(\mathbb{R}, \mathbb{R})$ se dérive en :

$$\varphi_n'(\lambda) = \langle \nabla\Phi\left[x_n + \lambda\nabla\Phi\,(x_n)\right], \nabla\Phi\,(x_n)\rangle$$

donc $\varphi_n'(0) = \|\nabla\Phi\,(x_n)\| \neq 0$ d'où $\Phi\,(x_{n+1}) < \Phi\,(x_n)$ lorsque $x_n \neq x^*$. En toute généralité, La suite $(\Phi\,[x_n])$ est décroissante, minorée par $\Phi\,(x^*)$, donc convergente. Il en découle $\Phi\,(x_{n+1}) - \Phi\,(x_n) \to 0$.

φ_n admet un minimum en λ_n donc $\langle \nabla\Phi\,(x_{n+1}), \nabla\Phi\,(x_n)\rangle = \varphi_n'(\lambda_n) = 0$. Comme $\nabla\Phi\,(x_n)$ et $x_{n+1} - x_n$ sont colinéaires, on en déduit :

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

donc la suite $(x_{n+1} - x_n)$ converge vers 0. Quitte à se restreindre au compact $\Phi^{-1}\left(]-\infty, \Phi\,(x_0)]\right)$, l'application $\nabla\Phi$ est uniformément continue, donc en particulier $\nabla\Phi\,(x_{n+1}) - \nabla\Phi\,(x_n) \to 0$. Par orthogonalité :

$$\|\nabla\Phi\,(x_n)\|^2 \leq \|\nabla\Phi\,(x_n)\|^2 + \|\nabla\Phi\,(x_{n+1})\|^2 = \|\nabla\Phi\,(x_n) - \nabla\Phi\,(x_{n+1})\|^2 \to 0$$

donc $\nabla\Phi\,(x_n) \to 0$.

Soit $\overline{x}$ une valeur d'adhérence de (x_n). Par continuité de $\nabla\Phi$, on a $\nabla\Phi(\overline{x}) = 0$, d'où nécessairement $\overline{x} = x^*$. La suite (x_n) est bornée et admet une unique valeur d'adhérence, donc $x_n \to x^*$. □

Remarque 1. Soient $A \in S_n^{++}$ et $b \in \mathbb{R}^d$. On veut résoudre le système linéaire $Ax = b$.

On considère alors $\Phi(x) = \frac{1}{2}\langle Ax, x\rangle - \langle b, x\rangle$. On vérifie aisément que $\nabla\Phi(x) = Ax - b$, donc x est solution du système précédent ssi $\nabla\Phi(x) = 0$.

On peut montrer que Φ est fortement convexe. On en déduit que x^*, comme défini précédemment, est l'unique l'unique solution du système $Ax = b$, et qu'on peut l'approcher par la méthode du gradient à pas optimal.