Théorème de convergence pour l'algorithme du gradient projeté

$$\boxed{x^{k+1} = T(x^k) = \mathrm{proj}_X(x^k - \alpha \nabla f(x^k))}$$

Lemme 1 *Soit f une fonction convexe dont le gradient ∇f est lipschitzien de constante L. On a alors :*

$$\langle \nabla f(x) - \nabla f(y), x - y \rangle \geq \frac{1}{L} \|\nabla f(x) - \nabla f(y)\|^2$$

pour tout x, y dans l'ensemble de définition de la fonction f.

Preuve Soit la fonction

$$h(x) = f(x) - f(y) - \langle f(y), x - y \rangle.$$

La fonction h est non négative, convexe et l'on a :

$$\nabla h(x) = \nabla f(x) - \nabla f(y).$$

On a :

$$\begin{aligned} 0 &\leq h(x - \frac{1}{L} \nabla h(x)) \\ &\leq h(x) - \frac{1}{L} \|\nabla h(x)\|^2 + \frac{1}{2L} \|\nabla h(x)\|^2 \\ &= h(x) - \frac{1}{2L} \|\nabla h(x)\|^2 \\ &= f(x) - f(y) - \langle \nabla f(y), x - y \rangle - \frac{1}{2L} \|\nabla f(y) - \nabla f(x)\|^2. \end{aligned}$$

En invertissant les rôles de x et de y on obtient :

$$0 \leq f(y) - f(x) - \langle \nabla f(x), y - x \rangle - \frac{1}{2L} \nabla f(x) - \nabla f(y)\|^2.$$

L'addition de ces deux inégalits mène au résultat.

Théorème 1 *Soit f une fonction convexe de gradient ∇f lipschitzien de constante L. Soit $X^* \neq \emptyset$ l'ensemble des points satisfaisant les conditions nécessaires d'optimalité du premier ordre. Si $0 < \alpha < 2/L$, la suite $\{x^k\}$ converge vers un point x^* de X^*. Si la fonction f est fortement convexe, alors la convergence est géométrique.*

Preuve (i) Soit $x^* \in X^*$. Par construction de x^{k+1} on a :

$$\begin{aligned} & \langle x^{k+1} - (x^k - \alpha \nabla f(x^k)), x^{k+1} - x^* \rangle \leq 0 \\ \Rightarrow \quad & \langle x^{k+1} - x^k, x^{k+1} - x^* \rangle + \alpha \langle \nabla f(x^k), x^{k+1} - x^* \rangle \leq 0 \\ \Rightarrow \quad & \langle x^{k+1} - x^k, x^{k+1} - x^* \rangle + \alpha \langle \nabla f(x^k) - \nabla f(x^*), x^{k+1} - x^* \rangle \leq 0 \end{aligned}$$

puisque $\langle \nabla f(x^*), x^* - x^{k+1}\rangle \leq 0$ (rappelez-vous que x^* satisfait les conditions du premier ordre). Maintenant, analysons les deux termes de gauche de cette inégalité :

- $\langle x^{k+1} - x^k, x^{k+1} - x^*\rangle = \frac{1}{2}[-\|x^k - x^*\|^2 + \|x^{k+1} - x^k\|^2 + \|x^{k+1} - x^*\|^2]$
-

$$\begin{aligned}
\langle \nabla f(x^k) - \nabla f(x^*), x^{k+1} - x^*\rangle &= \langle \nabla f(x^k) - \nabla f(x^*), x^k - x^*\rangle + \langle \nabla f(x^k) - \nabla f(x^*), x^{k+1} - x^k\rangle \\
&\geq \frac{1}{L}\|\nabla f(x^k) - \nabla f(x^*)\|^2 + \langle \nabla f(x^k) - \nabla f(x^*), x^{k+1} - x^k\rangle \\
&\geq -\frac{L}{4}\|x^{k+1} - x^k\|^2 \quad \text{puisque} \quad \|a\|^2 + \langle a, b\rangle \geq -\frac{\|b\|^2}{4}
\end{aligned}$$

On en déduit que :

$$-\|x^k - x^*\|^2 + \|x^{k+1} - x^*\|^2 + \|x^{k+1} - x^k\|^2 \left(1 - \frac{\alpha L}{2}\right) \leq 0$$

et enfin :

$$\|x^{k+1} - x^*\|^2 \leq \|x^k - x^*\|^2 - \left(1 - \frac{\alpha L}{2}\right)\|x^{k+1} - x^k\|^2. \tag{1}$$

Puisque la suite $\{\|x^k - x^*\|^2\}$ est décroissante et bornée, elle doit converger. On a aussi que:

$$\lim_{k\to\infty} \|x^{k+1} - x^k\| = 0.$$

Soit $\tilde{x}$ un point d'accumulation de la suite $\{x^k\}$. Il est facile de voir que $\tilde{x}$ est un point de fixe de l'application T et donc un minimum global de la fonction f. En utilisant (1), il est alors possible (exercice) de démontrer que *toute* la suite $\{x^k\}$ converge vers la solution $\tilde{x}$. Attention ! Il se peut fort bien que $\tilde{x} \neq x^*$.

(ii) Une fonction f est *fortement convexe de module* β sur X si l'on a :

$$f(y) \geq f(x) + \langle \nabla f(x), y - x\rangle + \frac{1}{2}\beta(1-\beta)\|y - x\|^2.$$

Cette condition implique que la solution existe et est unique. Si f est différentiable, cette condition prend la forme plus sympathique :

$$\langle \nabla f(y) - \nabla f(x), y - x\rangle \geq \beta\|y - x\|^2.$$

On peut alors écrire, puisque $x^* = T(x^*)$:

$$\begin{aligned}
\|x^{k+1} - x^*\|^2 &= \|T(x^k) - T(x^*)\|^2 \\
&\leq \|(x^k - \alpha\nabla f(x^k)) - (x^* - \alpha\nabla f(x^*))\|^2 \\
&\leq \|x^k - x^*\|^2 - 2\alpha\langle \nabla f(x^k) - \nabla f(x^*), x^k - x^*\rangle + \alpha^2\|\nabla f(x^k) - \nabla f(x^*)\|^2 \\
&\leq \|x^k - x^*\|^2 - 2\alpha\langle \nabla f(x^k) - \nabla f(x^*), x^k - x^*\rangle + L\langle \nabla f(x^k) - \nabla f(x^*), x^k - x^*\rangle \\
&\qquad \text{par le lemme 1} \\
&= \|x^k - x^*\|^2 - \alpha(2 - \alpha L)\rangle\nabla f(x^k) - \nabla f(x^*), x^k - x^*\rangle \\
&\leq (1 - \alpha\beta(2 - \alpha L))\|x^k - x^*\|^2 \\
&\qquad \text{en utilisant la forte convexité de la fonction } f.
\end{aligned}$$