# Méthode du gradient à pas optimal.

Rémi Lajugie

On note $\|\cdot\|$ la norme euclidienne usuelle de $\mathbb{R}^n$. On considère le problème de la minimisation de la fonction quadratique suivante, $o$ù$b \in \mathbb{R}^n, A \in S_n^{++}$ de valeurs propres $0 < \lambda_1 \leq \lambda_2 \leq \ldots \leq \lambda_n$ :

$$\frac{1}{2}x^\top Ax - b^\top x.$$

On appelle méthode du gradient à pas optimal l'algorithme suivant :

**Result**: approximation de l'argument du minimum de la fonction $x^*$
Initialisation : $x_0, r_0 = Ax_0 - b$ ;
**while** $\|r_k\| > \epsilon$ **do**
  $x_{k+1} = x_k + \alpha_k r_k$ ;
  $r_{k+1} = r_k - \alpha_k A r_k$ ;
  $\alpha_k = \frac{\|r_k\|^2}{r_k^\top A r_k}$
**end**

**Algorithm 1:** Méthode du gradient à pas optimal.

**Proposition 1** *(Inégalité de Kantorovitch) Sous nos hypothèses, on a :*

$$\forall x \in \mathbb{R}^n, \|x\|^4 \leq (x^\top A^{-1}x)(x^\top Ax) \leq \frac{(\lambda_1 + \lambda_n)^2}{4\lambda_1\lambda_n}\|x\|^4.$$

Preuve :

*Etape 1 : inégalité de gauche.*

**Fait 1** $\forall x > 0, x + \frac{1}{x} \geq 2.$

Comme $A$ est supposée symétrique définie positive on écrit alors $x^\top Ax = \sum_{i=1}^n(\lambda_i x_i^2)$, $x^\top A^{-1}x = \sum_{i=1}^n(\frac{1}{\lambda_i}x_i^2)$.

On fait un calcul

$$(x^\top Ax)(x^\top A^{-1}x) = \sum_{i=1}^n x_i^4 + \sum_{i<j}(\frac{\lambda_i}{\lambda_j} + \frac{\lambda_j}{\lambda_i})x_i x_j \geq (\sum_{i=1}^n x_i^4 + \sum_{i<j}(\frac{\lambda_i}{\lambda_j} + \frac{\lambda_j}{\lambda_i}x_i x_j) = \|x\|^4).$$

*Etape 2 : Inégalité de droite.*

**Fait 2** *La matrice $B = (A - \lambda_1 Id)(A - \lambda_n Id)A^-1$ est négative.*

Ainsi $\forall x, x^\top Bx \leq 0$.

Or on remarque que $x^\top Bx = x^\top Ax + \lambda_1\lambda_n x^\top A^{-1}x - (\lambda_1 + \lambda_n)x^\top x \neq 0$.

On sait aussi que $\forall a, b \in \mathbb{R}, \frac{ab}{2} \leq a^2 + b^2$. Donc

$$\forall x, \sqrt{(x^\top Ax)(x^\top A^{-1}x)(\lambda_1\lambda_n)} \leq x^\top Ax + \lambda_1\lambda_n x^\top A^{-1}x \leq (\lambda_1 + \lambda_n)x^\top x.$$

L'inégalité de Kantorovitch en découle en élevant au carré et divisant par la quantité strictement positive $\lambda_1\lambda_n$.

**Théorème 1** *La méthode du gradient à pas optimal converge et a un taux de convergence vérifiant*

$$\|x_k - x^*\| \leq C(\frac{\lambda_1 - \lambda_n}{\lambda_n + \lambda_1})^k,$$

*où $x^*$ est l'argument du minimum.*

Preuve :

$$\begin{aligned} r_{k+1}^\top A^{-1} r_{k+1} &= r_k A^{-1} r_k - 2 r_k^\top r_k \frac{r_k^\top r_k}{r_k^\top A r_k} + \alpha_k^2 r_k^\top A r_k. \qquad (1) \\ \frac{r_{k+1}^\top A^{-1} r_{k+1}}{r_k^\top A^{-1} r_k} &= 1 - \frac{\|r_k\|^4}{(r_k^\top A^{-1} r_k)(r_k^\top A r_k)}.\text{simplifications} \end{aligned}$$

On conclut par inégalité de Kantorovitch et en notant que $\sqrt{x^\top A^{-1} x} \geq \lambda_n \|x\|$.

# Références


— Di Menza (dernier chapitre).