Option Calcul Scientifique: Optimisation Numérique

Préparation Agrégation de Mathématiques
Université de Rennes 1
Isabelle Gruais

5 décembre 2023

1 Méthodes de gradient

Soit $K \subset \mathbb{R}^N$ et soit $J : K \to \mathbb{R}$, α-convexe. Soit à résoudre : trouver $x^* \in K$ solution de :

$$x^* \in K \quad \text{et} \quad J(x^*) = \min_{x \in K} J(x)$$

Hypothèse 1.0.1. *On suppose que J est α-convexe, différentiable et que ∇J est localement lipschitzienne.*

1.1 Gradient avec pas optimal

Définition 1.1.1 (Algorithme du gradient avec pas optimal)**.** On appelle suite engendrée par l'Algorithme du gradient avec pas optimal la suite $(u_n)_{n\geq 0} \in (\mathbb{R}^N)^{\mathbb{N}}$ définie par la relation de récurrence :

$$u_0 \in \mathbb{R}^N \quad \text{arbitraire,}$$

$$u_{n+1} = u_n - \mu_n \nabla J(u_n)$$

avec

$$J(u_n - \mu_n \nabla J(u_n)) = \min_{\mu \in \mathbb{R}} J(u_n - \mu \nabla J(u_n)).$$

Définition 1.1.2. On appelle direction de descente à l'étape n le vecteur $r_n := -\nabla J(u_n)$.

Proposition 1.1.1. *Dans l'algorithme 1.2.1, deux directions de descente consécutives sont orthogonales.*

Démonstration. Soit $n \geq 1$. On pose :

$$f_n(\mu) = J(u_n + \mu r_n) \quad \text{où} \quad r_n := -\nabla J(u_n). \tag{1}$$

Par définition :

$$f_n'(\mu_n) = 0 = \langle \underbrace{\nabla J(u_n + \mu_n r_n)}_{=-r_{n+1}}, r_n \rangle = -\langle r_{n+1}, r_n \rangle.$$

□

Théorème 1.1.2. *Sous l'Hypothèse 1.0.1, la suite $(u_n)_{n\geq 0}$ générée par l l'algorithme du gradient avec pas optimal converge vers la solution $u \in \mathbb{R}^N$ du problème de minimisation :*

$$J(u) = \min_{v \in \mathbb{R}^N} J(v)$$

et on a l'estimation :

$$\|u_n - u\| \leq \frac{1}{\alpha} \|\nabla J(u_n)\|.$$

Démonstration. Par construction, la suite $(J(u_n)_{n\geq 0}$ est décroissante, minorée par coercivité :

$$J(u_n) \geq \beta \|u_n\|^2 + b \geq b > -\infty$$

donc convergente. On en déduit qu'il existe $C > 0$ t.q. : $\forall n \geq 0$,

$$C \geq J(u_n) \geq \beta \|u_n\|^2 + b \Rightarrow \|u_n\|^2 \leq \frac{C}{\beta} =: M.$$

i.e. la suite $(u_n)_{n\geq 0}$ est majorée.

Comme J est α-convexe, on aussi :

$$J(u_n) \geq J(u_{n+1}) + \langle \nabla J(u_{n+1}), u_n - u_{n+1} \rangle + \frac{\alpha}{2} \|u_n - u_{n+1}\|^2 =$$

$$= J(u_{n+1}) + \mu_n \underbrace{\langle r_n, r_{n+1} \rangle}_{=0} + \frac{\alpha}{2} \|u_n - u_{n+1}\|^2 = J(u_{n+1}) + \frac{\alpha}{2} \|u_n - u_{n+1}\|^2$$

$$\Rightarrow \|u_n - u_{n+1}\|^2 \leq \frac{2}{\alpha}(J(u_n) - J(u_{n+1})) \underset{n\to+\infty}{\rightarrow} 0$$

Il en résulte :

$$\begin{aligned}\|\nabla J(u_n)\|^2 = \langle \nabla J(u_n), \nabla J(u_n)\rangle &= \langle \nabla J(u_n), \nabla J(u_n) - \nabla J(u_{n+1})\rangle \\ &\leq C_M \|\nabla J(u_n)\| \|u_n - u_{n+1}\|\end{aligned}$$

d'où :

$$\|\nabla J(u_n)\| \leq C_M \| \|u_n - u_{n+1}\| \underset{n\to+\infty}{\rightarrow} 0.$$

Par α-convexité de J :

$$\alpha\|u_n - u\|^2 \leq \langle \nabla J(u_n) - \nabla J(u), u_n - u\rangle = \langle \nabla J(u_n), u_n - u\rangle \leq \|\nabla J(u_n)\|\|u_n - u\|$$

$$\Rightarrow \|u_n - u\| \leq \frac{1}{\alpha}\|\nabla J(u_n)\| \underset{n\to+\infty}{\rightarrow} 0.$$

$\square$

Proposition 1.1.3 (Cas particulier d'une fonctionnelle quadratique)**.** *On suppose que J est définie par :*

$$J(x) = \frac{1}{2}\langle Ax, x\rangle - \langle b, x\rangle, \quad \forall x \in \mathbb{R}^N$$

où $A \in \mathbb{R}^{N\times N}$ est symétrique, définie positive, et qu'il existe $\alpha > 0$ t.q.

$$\langle Ax, x\rangle \geq \alpha\|x\|^2, \quad \forall x \in \mathbb{R}^N.$$

Alors :

$$\mu_n = \frac{\|r_n\|^2}{\langle Ar_n, r_n\rangle}, \quad \forall n \geq 0. \tag{2}$$

Démonstration. Avec la définition (1) :

$$\begin{aligned}f_n'(\mu_n) = 0 = \langle A(u_n + \mu_n r_n) - b, r_n\rangle = \langle \underbrace{Au_n - b}_{=-r_n}, r_n\rangle + \mu_n\langle Ar_n, r_n\rangle = \\ = -\|r_n\|^2 + \mu_n\langle Ar_n, r_n\rangle\end{aligned}$$

$\square$

Théorème 1.1.4. *Sous les hypothèses de la Proposition 1.1.3, on note $0 < \lambda_1 \leq \cdots \leq \lambda_N$ la suite des valeurs propres de A comptées avec leur ordre de multiplicité. Alors, l'erreur à l'étape n définie par*

$$e_n := u_n - u$$

vérifie :

$$\frac{\langle Ae_n, e_n\rangle}{\langle Ae_0, e_0\rangle} \leq C\left(\frac{\lambda_N - \lambda_1}{\lambda_N + \lambda_1}\right)^{2n}, \quad \forall n \geq 0$$

Démonstration. Le démonstration utilise le résultat suivant qui sera admis :

Lemme 1.1.5 (Estimation de Kantorovitch)**.**

$$\|x\|^4 \leq \langle Ax, x\rangle \langle A^{-1}x, x\rangle \leq \frac{1}{4}\left(\sqrt{c} + \frac{1}{\sqrt{c}}\right)^2 \|x\|^4 \tag{3}$$

où

$$c := \frac{\lambda_N}{\lambda_1}$$

Soit $n \geq 0$. On a :

$$e_{n+1} = u_{n+1} - u = \mu_n r_n + e_n$$

donc

$$\langle Ae_{n+1}, e_{n+1}\rangle = \mu_n \langle \underbrace{Ae_{n+1}}_{=-r_{n+1}}, r_n\rangle + \langle Ae_{n+1}, e_n\rangle = \langle Ae_{n+1}, e_n\rangle$$

$$= \langle Ae_n, e_n\rangle + \mu_n \langle Ar_n, e_n\rangle = \langle Ae_n, e_n\rangle - \mu_n \|r_n\|^2$$

$$\underset{(2)}{=} \langle Ae_n, e_n\rangle - \frac{\|r_n\|^4}{\langle Ar_n, r_n\rangle}$$

avec :

$$\langle Ae_n, e_n\rangle = \langle r_n, A^{-1}r_n\rangle.$$

On en déduit :

$$\frac{\langle Ae_{n+1}, e_{n+1}\rangle}{\langle Ae_n, e_n\rangle} = 1 - \frac{\|r_n\|^4}{\langle r_n, A^{-1}r_n\rangle \langle A\langle Ar_n, r_n\rangle}$$

$$\underset{(3)}{\leq} 1 - 4\left(\sqrt{c} + \frac{1}{\sqrt{c}}\right)^{-2} = \left(\frac{c-1}{c+1}\right)^2 = \left(\frac{\lambda_N - \lambda_1}{\lambda_N + \lambda_1}\right)^2.$$

Il en résulte :

$$\frac{laAe_{n+1}, e_{n+1}\rangle}{\langle Ae_0, e_0\rangle} = \Pi_{k=0}^n \frac{laAe_{k+1}, e_{k+1}\rangle}{\langle Ae_k, e_k\rangle} \leq \left(\frac{\lambda_N - \lambda_1}{\lambda_N + \lambda_1}\right)^{2(n+1)}.$$

□

1.2 Gradient avec pas fixe

Définition 1.2.1 (Algorithme du gradient avec pas fixe)**.** On appelle suite engendrée par l'Algorithme du gradient avec pas fixe $\mu \in \mathbb{R}$ la suite $(u_n)_{n\geq 0} \in (\mathbb{R}^N)^{\mathbb{N}}$ définie par la relation de récurrence :

$$u_0 \in \mathbb{R}^N \quad \text{arbitraire,}$$

$$u_{n+1} = u_n - \mu \nabla J(u_n)$$

Théorème 1.2.1. *Sous l'hypothèse 1.0.1, la suite $(u_n)_{n\geq 0}$ définie par l'Algorithme 1.2.1 converge vers la solution du problème de minimisation :*

$$J(u) = \min_{v \in \mathbb{R}^N} J(v)$$

pour tout $\mu > 0$ suffisamment petit. En particulier, elle converge pour tout $\mu \in]0, \frac{2\alpha}{C^2}[$ où $C > 0$ est la constante de Lipschitz de ∇J sur la boule fermée $B := \bar{B}(u, \|u - u_0\|)$.

Démonstration. Soit $\mathcal{P}(n)$ la propriété : $u_n \in B$. Alors $\mathcal{P}(0)$ est vraie par définition de B. On suppose $\mathcal{P}(n)$ vraie.

De la relation :

$$e_{n+1} = u_{n+1} - u_n + e_n = -\mu \nabla J(u_n) + e_n = -\mu(\nabla J(u_n) - \nabla J(u)) + e_n$$

on déduit que :

$$\begin{aligned} \|e_{n+1}\|^2 &= \|e_n\|^2 - 2\mu \langle \nabla J(u_n) - \nabla J(u), e_n \rangle + \mu^2 \|\nabla J(u_n) - \nabla J(u)\|^2 \\ &\leq \|e_n\|^2 \underbrace{(1 - 2\mu\alpha + C^2\mu^2)}_{=:\theta(\mu)} \end{aligned}$$

avec : $\forall u, v \in \mathbb{R}^N$,

$$\alpha \|u - v\|^2 \leq \langle \nabla J(u) - \nabla J(v), u - v \rangle \leq C\alpha \|u - v\|^2$$

donc $0 < \alpha < C$. L'étude des variations de $\theta : \mu \mapsto 1 - 2\mu\alpha + C^2\mu^2$ montre que

$$0 < \theta(\mu) < 1 \iff \mu \in]0, \frac{2\alpha}{C^2}[.$$

On en déduit que si $\mu \in]0, \frac{2\alpha}{C^2}[$ alors

$$\|e_{n+1}\| < \|e_n\| \underset{\mathcal{P}(n)}{<} \|u - u_0\|$$

i.e. $u_{n+1} \in B$. Par récurrence sur $n \geq 0$:

$$\|e_n\| \leq \theta(\mu)^n \|e_0\| \underset{n \to +\infty}{\to} 0.$$

$\square$

Cas des fonctionnelles quadratiques

Soit $J(x) = \frac{1}{2}\langle Ax, x\rangle$, $\forall x \in \mathbb{R}^N$. avec A symétrqie définie positive de valeurs propres :

$$0 < \lambda_1 \leq \cdots \leq \lambda_N.$$

Après diagonalisation dans une bon de vecteurs propres, J se réécrit sous la forme :

$$J(x) = \frac{1}{2}\sum_{i=1}^{N} \lambda_i x_i^2.$$

Soit $\mu > 0$. L'algorithme de gradient avec pas fixe μ se réécrit :

$$u^0 \in \mathbb{R}^N, \quad u_i^{n+1} = (1 - \mu\lambda_i)u_i^n, \quad i = 1, \cdots, N.$$

La suite $(u^n)_{n\geq 0}$ converge ssi : $\forall i \in [[1, N]]$,

$$|1 - \mu\lambda_i| < 1 \iff 0 < \mu < \frac{2}{\lambda_i}$$

i.e. ssi $0 < \mu < \frac{2}{\lambda_N}$. Le taux de convergence est optimal si $\mu = \mu_{\text{opt}}$ est solution de

$$\max_{1\leq i\leq N} |1 - \mu\lambda_i|$$

ce qui est réalisé pour

$$1 - \mu\lambda_1 = -1 + \mu\lambda_N \iff \mu = \frac{2}{\lambda_1 + \lambda_N} =: \mu_{\text{opt}}$$