Gradient à pas optimale

réf: Ciarlet X-ENS, Analyse 4. p.39 + Site de Fabien Argelier.
Ciarlet, Analyse numérique p.183.

Définition: Soit $C \longrightarrow \mathbb{R}$ où C est une partie convexe et elle dite α-convexe pour $\alpha > 0$ si elle vérifie pour tout $u, v \in C$, $t \in [0,1]$:

$$J((1-t)u + tv) \leq (1-t)J(u) + tJ(v) - \frac{\alpha}{2} t(1-t)\|u-v\|^2.$$

On a la caractérisation suivante.

Lemme: Soit $J: C \longrightarrow \mathbb{R}$ une fonction différentiable et $\alpha > 0$. Alors on a équivalences entre:

1) J est α-convexe
2) Pour tous $u, v \in C$, $J(v) \geq J(u) + \langle \nabla J(u), v-u \rangle + \frac{\alpha}{2}\|v-u\|^2$
3) Pour tous $u, v \in C$, $\langle \nabla J(v) - \nabla J(u), v-u \rangle \leq \alpha \|v-u\|^2$.

Preuve:

1) $\Rightarrow$ 2): Soient $u, v \in C$ et $t \in [0,1]$ on a:

$$\frac{1}{t}\left(J(u + t(v-u)) - J(u)\right) \leq J(v) - J(u) - \alpha \frac{(1-t)}{2}\|u-v\|^2$$

$$\xrightarrow[t \to 0]{} \quad \langle \nabla J(u), v-u \rangle \leq J(v) - J(u) - \frac{\alpha}{2}\|u-v\|^2$$

2) $\Rightarrow$ 3) Soient $u, v \in C$, on a:

$$J(v) \geq J(u) + \langle \nabla J(u), v-u \rangle + \frac{\alpha}{2}\|u-v\|^2$$

$$J(u) \geq J(v) + \langle \nabla J(v), u-v \rangle + \frac{\alpha}{2}\|u-v\|^2$$

Il vient donc

$$J(v) - \left[J(v) + \langle \nabla J(v), u-v \rangle + \frac{\alpha}{2}\|u-v\|^2\right] \geq \langle \nabla J(u), v-u \rangle + \frac{\alpha}{2}\|u-v\|^2$$

Ce qui donne:

$$-\frac{\alpha}{2}\|u-v\|^2 + \langle \nabla J(v), v-u \rangle \geq \langle \nabla J(u), v-u \rangle + \frac{\alpha}{2}\|u-v\|^2$$

On a bien:

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

3)⇒1) : Soient $u, v \in \mathbb{R}^n$ et $t \in [0,1]$. On pose $\varphi : t \mapsto J(u+t(v-u))$. Alors φ est dérivable et

$$\forall t \in \mathbb{R}, \quad \varphi'(t) = \langle \nabla J(u+t(v-u)), v-u \rangle .$$

Ainsi par hypothèses, on a pour $t, s \in \mathbb{R}$ tels que $s < t$:

$$\varphi'(t) - \varphi'(s) = \langle \nabla J(u+t(v-u)) - \nabla J(u+s(v-u)), (t-s)(v-u) \rangle \frac{1}{t-s}$$

$$\geq \alpha (t-s) \|v-u\|^2 > 0.$$

On obtient donc pour $\theta \in [1-t]$ en intégrant de $t=0$ à $t=1$ et $s=0$ à $s=\theta$:

$$\theta\left(\varphi(1) - \varphi(\theta)\right) - (1-\theta)\left(\varphi(\theta) - \varphi(0)\right) \geq \alpha \frac{\theta(1-\theta)}{2} \|u-v\|^2$$

donc $$(1-\theta) J(u) + \theta J(v) - \frac{\alpha}{2} \theta(1-\theta) \|u-v\|^2 \geq J((1-\theta)u + \theta v).$$

Théorème : Soit $G : \mathbb{R}^n \to \mathbb{R}$ une fonction de classe $\mathcal{C}^1$. On suppose que G est α-convexe, alors :

1) G admet un unique minimum global sur $\mathbb{R}^n$ en un point $a \in \mathbb{R}^n$. De plus a est l'unique point tel que $\nabla g(a) = 0$

2) Pour tout point initiale $x^0 \in \mathbb{R}^n$, la suite $(x_n)_{n \in \mathbb{N}}$ définie par :

- $x_0 = x^0$
- $\forall n \in \mathbb{N}, \quad x_{n+1} = x_n + \rho_n \nabla G(x_n)$ où $\rho_n = \underset{\rho \in \mathbb{R}}{\operatorname{argmin}} \left(G(x_n + \rho \nabla G(x_n))\right)$

converge vers le minimiseur a de G.

Preuve :

1) On utilise le fait classique que toutes fonctions continue coercive admettant un minimum globale.

En effet comme $\lim_{\|x\| \to +\infty} f(x) = +\infty$, il existe $A > 0$ tel que

$$\forall x \in \mathbb{R}^n \setminus \overline{B(0,A)}, \quad f(x) > f(0).$$

De plus $\overline{B(0,A)}$ est compact et f continue donc il existe $x_0 \in \overline{B(0,A)}$ telle que $f(x_0) = \min_{x \in \overline{B(0,A)}} f(x)$.

Ainsi $f(x_0) \leq f(0)$.
On a donc $\forall x \in \mathbb{R}^n$, $f(x_0) \leq f(x)$ d'où le résultat.

Il suffit de montrer que G est coercive or pour $x \in \mathbb{R}^n$ on a :

$$G(y) \geqslant G(0) + \langle \nabla G(0), y \rangle + \frac{\alpha}{2}\|y\|^2 \xrightarrow[\|y\| \to +\infty]{} +\infty.$$

car $|\langle \nabla G(0), y \rangle| \leq \|\nabla G(0)\| \|y\|$ donc $\langle \nabla G(0), y \rangle = O(\|y\|)$

On note a ce minimum global, comme G est de classe $\mathcal{C}^1$ on a : $\nabla G(a) = 0$

Pour l'unicité on utilise la stricte convexité, soit $b \in \mathbb{R}^n$ alors

$$G(b) - G(a) \geqslant \langle \nabla G(a), b-a \rangle + \frac{\alpha}{2}\|b-a\|^2$$

donc $G(b) \geqslant G(a) + \underbrace{\frac{\alpha}{2}\|b-a\|^2}_{>0 \text{ si } b \neq 0} > G(b)$.

On a donc l'unicité.

De même soit $c \in \mathbb{R}^n$, si $\nabla G(c) = 0$ alors

$$G(c) \geqslant G(a) \geqslant G(c) + \frac{\alpha}{2}\|c-a\|^2$$

Ainsi $c = a$.

c) Montrons par récurrence que la suite $(x_n)_{n \in \mathbb{N}}$ est bien définie.
Soit $n \in \mathbb{N}$, on suppose x_n définit.
Si $\nabla G(x_n) = 0$ alors pour le premier point $x_n = a$, la suite converge en un nombre fini d'opération.
Si $\nabla G(x_n) \neq 0$ il suffit de montrer que la fonction $\varphi_n : \rho \longmapsto G(x_n + \rho \nabla G(x_n))$ admette un unique minimum global, pour cela il suffit de montrer que φ_n est β-coercive avec $\beta > 0$ et d'appliquer le point ①.

Or on a pour tout $t, s \in \mathbb{R}$,

$$\varphi_n(t) - \varphi_n(s) = G(x_n + t\nabla G(x_n)) - G(x_n + s\nabla G(x_n))$$
$$\geqslant \langle \nabla G(x_n + s\nabla G(x_n)), (t-s)\nabla G(x_n) \rangle + \underbrace{\frac{\alpha\|\nabla G(x_n)\|^2}{2}}_{>0} |t-s|^2$$
$$= \varphi_n'(s)(t-s) + \frac{\alpha\|\nabla G(x_n)\|^2}{2}|t-s|^2.$$

Car $\varphi_n'(s) = \langle \nabla G(x_n + s\nabla G(x_n)), \nabla G(x_n) \rangle$.

Donc φ_n est $\alpha\|\nabla G(x_n)\|^2$ - coercive, de classe $\mathcal{C}^1$. Donc elle admet un unique minimum global.

Montrons que les gradients sont orthogonaux

On sait que la suite $(x_n)_{n\in\mathbb{N}}$ est bien définie et de plus on a la caractérisation, pour $n\in\mathbb{N}$, ρ_n est l'unique réelle tel que

$$0 = \varphi_n'(\rho_n) = \langle \nabla G(x_n + \rho_n \nabla G(x_n)), \nabla G(x_n)\rangle.$$

Ainsi on a :

$$\langle \nabla G(x_{n+1}), \nabla G(x_n)\rangle = 0.$$

Montrons que la suite converge :

Soit a l'unique minimum de G.

Si $\exists n\in\mathbb{N}$ tel que $x_n = a \Leftrightarrow \nabla G(x_n) = 0$ la suite est stationnaire à a.

Supposons que ça ne soit pas le cas.

Par définition la suite $(G(x_n))_{n\in\mathbb{N}}$ est minorée décroissante, en effet :

$$\forall n\in\mathbb{N}, \qquad G(a) \le G(x_{n+1}) = \min_{\rho\in\mathbb{R}} G(x_n + \rho \nabla G(x_n)) \le G(x_n).$$

Donc elle converge, soit ℓ sa limite.

De plus pour $n\in\mathbb{N}$ on a, $\nabla G(x_n) = \dfrac{x_{n+1} - x_n}{\rho_n}$ différent de 0.

et $\langle \nabla G(x_{n+1}), \nabla G(x_n)\rangle = 0$

On a donc $G(x_n) - G(x_{n+1}) \ge \langle \nabla G(x_{n+1}), x_n - x_{n+1}\rangle + \dfrac{\alpha}{2}\|x_{n+1} - x_n\|^2$

$$= \frac{\alpha}{2}\|x_{n+1} - x_n\|^2.$$

Or $G(x_n) \longrightarrow \ell$, $G(x_{n+1}) \longrightarrow \ell$.

On a donc que $\lim\limits_{n\to+\infty} \|x_{n+1} - x_n\|^2 = 0$.

De plus G est coercive donc il existe $A>0$ tel que $\forall \|x\| > A$, on a $G(x) > G(x_0)$. Ainsi la suite $(x_n)_{n\in\mathbb{N}}$ est bornée, soit ψ une extractrice et x^* une valeur d'adhérence de la suite.

Par le point précédent, $x_{\psi(n)} \longrightarrow x^*$

$$x_{\psi(n)+1} \longrightarrow x^*$$

Or G est de classe $\mathcal{C}^1$ donc ∇G est continue ainsi

$$0 = \langle \nabla G(x_{\psi(n)+1}), \nabla G(x_{\psi(n)})\rangle \longrightarrow \|\nabla G(x^*)\|^2$$

On a donc $\nabla G(x^*) = 0$ donc $x^* = a$.

Ainsi $(x_n)_{n\in\mathbb{N}}$ admet une unique valeur d'adhérence donc converge vers a.

On a même une estimation de l'erreur en effet

$$\langle \nabla G(x_n), x_n - a \rangle \geq \alpha \| x_n - a \|^2$$

Ce qui donne :

$$\| x_n - a \| \leq \frac{1}{\alpha} \| \nabla G(x_n) \| .$$

□

Prop : Dans le cas d'une fonctionnelle quadratique (avec $A \in S_n^{++}(\mathbb{R})$)

$$G(x) = \frac{1}{2} \langle Ax, x \rangle - \langle b, x \rangle .$$

On peut calculer le pas optimal,

$$\rho_n = - \frac{\| x_n \|^2}{\langle Ax_n, x_n \rangle} .$$

Et par l'inégalité de Kantorovitch, on a l'estimation de vitesse :

$$\| x_n - a \| \leq \sqrt{\frac{\lambda_{max}}{\lambda_{min}}} \left(\frac{\lambda_{max} - \lambda_{min}}{\lambda_{max} + \lambda_{min}} \right)^n \| x_0 - a \|$$

Ce qui se reformule, $\| x_n - a \| \leq \sqrt{cond(A)} \left(\frac{cond(A) - 1}{cond(A) + 1} \right) \| x_0 - a \|$.

Plan preuve : ① G admet une unique minimum

② x_{n+1} est bien définie en montrant que $\varphi_n : \rho \longmapsto G(x_n + \rho \nabla G(x_n))$ admet un unique minimum par le point ①

③ $\langle \nabla G(x_{n+1}), \nabla G(x_n) \rangle = 0$

④ $G(x_n) \longrightarrow \ell$ et $\nabla G(x_{n+1}) = \frac{x_{n+1} - x_n}{\rho_n}$ donc $\| x_{n+1} - x_n \| \to 0$

⑤ x_n admet une unique valeur d'adhérence et est bornée.