Référence :
FGN4 p. 39

Developpement :

Gradient à pas optimal :

leçons 219
223
226
229
253

Introduction

Les méthodes du gradient sont des méthodes itératives où on remplace un système linéaire par un problème de minimalité. L'algo du gradient à pas optimal consiste à déterminer un pas optimal pour minimiser une fonction.

But : . Minimiser une distance (optimisation)
. trouver état d'équilibre d'un système

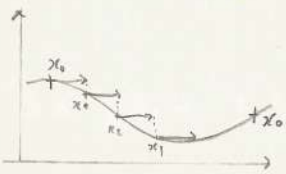
Ex : (Méthodes sans gradient)

→ Dichotomie
→ méthode de la règle d'or.

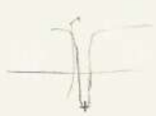
Ex (Méthodes du gradient)

On connait x_0 et on a $x_{k+1} = x_k + \underbrace{d_k}_{\text{pas de descente}} \cdot \underbrace{g_k}_{\text{direction de descente}}$ $\quad g_k = -\frac{df}{dx}\Big|_{x_k}$

① Gradient à pas fixe :



Tant que $\|\nabla f(x)\| > \varepsilon$
$\quad$ L $x \mapsto x - \text{pas} \times \nabla f(x)$
retourner x



la suite des itérés pourrait ne pas "voir" le trou ici

problème : marche seulement pour fonctions douces et pour x_0 proche de la solution.

② Gradient à pas optimal :

on choisit la longueur du pas à chaque étape !

la contrainte : il faut connaître la dérivée seconde.

astuce : DL de $f(x)$ au point x_k :

$$f(x_k + dx_k) = f(x_k) + dx_k \cdot \frac{\delta f}{\delta x}(x_k) + \frac{dx_k^2}{2} \cdot \frac{\delta^2 f}{\delta x^2}(x_k) + o(dx_k^2)$$

on dérive

$$f'(x_k + dx_k) = f'(x_k) + dx_k \frac{\delta^2 f}{\delta x^2}(x_k) + \cdots$$

et on pose $f'(x_k + dx_k) \simeq f'(x_k) + dx_k \frac{\delta^2 f}{\delta x^2}(x_k) = 0$ pour avoir un point critique.

Ainsi $dx_k = \dfrac{-f'(x_k)}{\left(\frac{\delta^2 f}{\delta^2 x}(x_k)\right)}$

étape 1 : Place x_0, on trace la droite

étape 2 : on regarde le min de δ. Et on réitère !

On a convergence linéaire.

Définition : (fonction α-convexe)

Soit $\varphi \in \mathcal{C}^0(\mathbb{R}^d, \mathbb{R})$. Soit $\alpha > 0$. On dit que φ est α-convexe lorsque $\forall x, y \in \mathbb{R}^n$, $\forall t \in]0,1[$,

$$\varphi(tx + (1-t)y) \leq t\,\varphi(x) + (1-t)\,\varphi(y) - \alpha . t . (1-t) \|x-y\|^2.$$

Une telle fonction est en particulier strictement convexe.

Développement

Lemme : Soit $\varphi \in \mathcal{C}^1(\mathbb{R}^d, \mathbb{R})$. Alors (i) et (ii) sont équivalents :

(i) φ est α-convexe

(ii) $\forall x, y \in \mathbb{R}^d$, $\varphi(y) \geqslant \varphi(x) + \langle \nabla \varphi(x), y - x \rangle + \alpha . \|y - x\|^2$.

$\boxed{i \Rightarrow ii}$ Soit φ α-convexe.

Alors $\varphi(x + t(y-x)) \leq \varphi(x) + t(\varphi(y) - \varphi(x)) - \alpha . t . (1-t) \|y-x\|^2$

donc $\dfrac{\varphi(x + t(y-x)) - \varphi(x)}{t} \leq \varphi(y) - \varphi(x) - \alpha(1-t) . \|y-x\|^2$

calculer ce terme pour s'en convaincre

donc en passant à la limite, pour $t \to 0^+$:

$$\partial_t [\varphi(x + t(y-x))]_{t=0} \leq \varphi(y) - \varphi(x) - \alpha \|y-x\|^2$$

i.e. $\langle \nabla \varphi(x), y - x \rangle \leq \varphi(y) - \varphi(x) - \alpha \|y-x\|^2$

$\boxed{ii \Rightarrow i}$ Posons $z = tx + (1-t)y$. On a $z - x = (1-t)(y-x)$, d'où la majoration :

$\varphi(x) \geqslant \varphi(z) + \langle \nabla \varphi(z), x - z \rangle + \alpha \|x - z\|^2$ (ii) pour x et z

$$\Rightarrow \varphi(z) \leq \varphi(x) + \langle \nabla\varphi(z), z-x\rangle - \alpha \|z-x\|^2$$
$$\leq \varphi(x) + (1-t)\langle \nabla\varphi(z), y-x\rangle - \alpha\cdot(1-t)^2 \cdot \|y-x\|^2$$

et de même pour z et y, $z-y = -t(y-x)$ d'où

$$\varphi(z) \leq \varphi(y) + \langle \nabla\varphi(z), z-y\rangle - \alpha\|z-y\|^2$$
$$\leq \varphi(y) - t\cdot\langle \nabla\varphi(z), y-x\rangle - \alpha\cdot t^2\cdot\|y-x\|^2.$$

donc en considérant la moyenne pondérée de deux applications,

$$\varphi(z) \leq t\,\varphi(x) + (1-t)\,\varphi(y) + [t(1-t) - t(1-t)]\cdot\langle\nabla\varphi(z), y-x\rangle$$
$$- [t(1-t)^2 + (1-t)t^2]\cdot\alpha\cdot\|y-x\|^2$$
$$\leq t\,\varphi(x) + (1-t)\,\varphi(y) - \alpha\cdot\|y-x\|^2\cdot t(1-t)$$ ■

Théorème : Soit $\varphi \in \mathcal{C}^1(\mathbb{R}^d, \mathbb{R})$ une fonction α-convexe. φ admet un unique minimum (local et global) atteint en x^*.
Soit $x_0 \in \mathbb{R}^d$ fixé. On considère la suite $x_{n+1} = x_n + \lambda_n \nabla\varphi(x_n)$
où $\lambda_n = 0$ si $x_n = x^*$ ou λ_n tq $\varphi(x_n + \lambda_n\nabla\varphi(x_n)) = \min_{\mu\in\mathbb{R}}\{\varphi(x_n + \mu\nabla\varphi(x_n))\}$.
Pour tout $x_0 \in \mathbb{R}^d$, cette suite est bien définie et $(x_n) \xrightarrow[n\to+\infty]{} x^*$.

Preuve :

▶ Étape 1 : φ admet un unique minimum atteint en x^*, qui est le seul point critique de φ :

D'après le lemme, on a $\varphi(x) \geq \varphi(0) + \langle\nabla\varphi(0), x\rangle + \alpha\cdot\|x\|^2$
donc φ est une fonction coercive. Ainsi, φ admet un minimum en x^* (par coercivité). Ce minimum global est a fortiori un minimum local, donc $\nabla\varphi(x^*) = 0$. En outre on a l'inégalité

$$\varphi(x^*) \geq \varphi(y) + \langle\nabla\varphi(y), x^*-y\rangle + \alpha\cdot\|y-x^*\|^2$$

donc
$$\langle\nabla\varphi(y), y-x^*\rangle \geq \underbrace{\varphi(y) - \varphi(x^*)}_{\geq 0} + \alpha\cdot\|y-x^*\|^2$$
$$\geq \alpha\cdot\|y-x^*\|^2$$
$$> 0 \text{ dès que } y \neq x^*.$$

En particulier $\nabla\varphi(y) \neq 0$ donc x^* est le seul point critique, donc le seul minimum local.

▶ Étape 2 : La suite est bien définie et orthogonalité de $\nabla\varphi(x_n)$, $\nabla\varphi(x_{n+1})$

* Considérons la fonction $\boxed{\varphi_n(\lambda) = \varphi(x_n + \nabla\varphi(x_n).\lambda)}$

On a $\varphi_n(t\lambda + (1-t)\mu) = \varphi\left(x_n + \nabla\varphi(x_n)(t\lambda + (1-t)\mu)\right)$

$$= \varphi(t x_n + (1-t)x_n + \nabla\varphi(x_n)\cdot(t\lambda + (1-t)\mu))$$
$$= \varphi(t[x_n + \lambda\nabla\varphi(x_n)] + (1-t)[x_n + \mu\nabla\varphi(x_n)])$$
$$\varphi\ \alpha\text{-convexe} \to \leq t.\varphi_n(\lambda) + (1-t).\varphi_n(\mu) - \alpha.\|\nabla\varphi(x_n)\|^2.t.(1-t).\|\lambda-\mu\|^2$$

donc φ_n est $\alpha.\|\nabla\varphi(x_n)\|^2$-convexe. Dès lors que $x_n \neq x^*$, φ_n admet un unique minimum (global et local) en λ_n. La suite (x_n) est donc bien définie.

* φ_n étant C^1, elle se dérive en $\varphi_n'(\lambda) = \langle\nabla\varphi(x_n + \lambda\nabla\varphi(x_n)), \nabla\varphi(x_n)\rangle$

(en effet $\varphi_n'(\lambda) = \nabla\varphi(x_n).\varphi'(x_n + \nabla\varphi(x_n).\lambda) = \langle\nabla\varphi(x_n + \nabla\varphi(x_n).\lambda), \nabla\varphi(x_n)\rangle$

donc $\varphi_n'(0) = \|\nabla\varphi(x_n)\|^2 \neq 0$ d'où $\varphi(x_{n+1}) < \varphi(x_n)$ pour $x \neq x^*$.

(car la suite $x_{n+1} = x_n + \lambda_n\cdot\nabla\varphi(x_n) \Rightarrow \varphi(x_{n+1}) = \varphi(x_n + \lambda_n\nabla\varphi(x_n)) \Rightarrow \varphi(x_{n+1}) = \varphi_n(\lambda_n)$ or λ_n le min global de φ_n. Donc $\underset{\varphi(x_n)}{\underset{\parallel}{\varphi_n(0)}} > \varphi_n(\lambda_n)$.)

De manière générale, $(\varphi(x_n))_{n\geq1}$ est décroissante, minorée par $\varphi(x^*)$ (car x^* minimum de φ) et donc $(\varphi(x_n))_{n\geq1}$ converge, donc $\varphi(x_{n+1}) - \varphi(x_n) \underset{n\to+\infty}{\longrightarrow} 0$

φ_n admet un minimum en λ_n donc $\langle\nabla\varphi(x_{n+1}), \nabla\varphi(x_n)\rangle \underset{\text{déf } \varphi_n'}{=} \varphi_n'(\lambda_n) \underset{\varphi_n \text{ admet min en } \lambda_n}{=} 0$

Ainsi, pour $x \neq x^*$, $\langle\nabla\varphi(x_{n+1}), \nabla\varphi(x_n)\rangle = 0$.

▶ Étape 3

$x_{n+1} - x_n$ et $\nabla\varphi(x_n)$ étant colinéaires (car $x_{n+1} = x_n + \lambda_n\nabla\varphi(x_n)$)

on en déduit que $\langle x_{n+1} - x_n, \nabla\varphi(x_{n+1})\rangle = 0$.

Ainsi, φ α-convexe et donc (par le lemme)

$$\varphi(x_n) \geq \varphi(x_{n+1}) + \langle\nabla\varphi(x_{n+1}), x_n - x_{n+1}\rangle + \alpha.\|x_{n+1} - x_n\|^2$$
$$\Leftrightarrow \varphi(x_n) - \varphi(x_{n+1}) \geq \alpha.\|x_{n+1} - x_n\|^2$$

donc la suite $(x_{n+1} - x_n)_{n \geq 0}$ converge vers 0. Quitte à se restreindre au compact $\varphi^{-1}(]-\infty, \varphi(x_0)])$ (car ), $\nabla\varphi$ est uniformément continue, donc en particulier $\nabla\varphi(x_{n+1}) - \nabla\varphi(x_n) \underset{n\to+\infty}{\longrightarrow} 0$.
(car $x_{n+1} - x_n \underset{n\to+\infty}{\longrightarrow} 0$)

$$\|\nabla\varphi(x_n)\|^2 \leq \|\nabla\varphi(x_n)\|^2 + \|\nabla\varphi(x_{n+1})\|^2 \overset{\text{orthogonalité}}{=} \|\nabla\varphi(x_n) - \nabla\varphi(x_{n+1})\|^2 \underset{n\to+\infty}{\longrightarrow} 0$$

donc $\nabla\varphi(x_n) \underset{n\to+\infty}{\longrightarrow} 0$.

Soit $\bar{x}$ une valeur d'adhérence de (x_n). Par continuité de $\nabla\varphi$, on a $\nabla\varphi(\bar{x}) = 0$, d'où nécessairement $\bar{x} = x^*$. La suite (x_n) est bornée et admet une unique valeur d'adhérence, donc $x_n \underset{n\to+\infty}{\longrightarrow} x^*$. ■

Application:

Soit $A \in S_n^{++}(\mathbb{R})$. $b \in \mathbb{R}^d$. On veut résoudre le système linéaire $Ax = b$.

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

φ α-convexe donc $\exists!$ solution de $Ax = b$ que l'on peut approcher avec la méthode du gradient à pas optimal.

Question du jury:

Thm: Une suite bornée avec une unique valeur d'adhérence converge.

Preuve: Par l'absurde

Soit (U_n) suite bornée admettant une unique valeur d'adhérence ℓ et ne convergeant pas vers ℓ:

$\exists \varepsilon > 0$ tq $\forall N \in \mathbb{N}$, $\exists n \in \mathbb{N}$ tq $n \geq N$ tq $|U_n - \ell| \geq \varepsilon$.

Alors l'ensemble $A := \{n \in \mathbb{N} ; |U_n - \ell| \geq \varepsilon\}$ est infini.

Soit $\sigma : \mathbb{N} \to A$ strictement croissante.

Ainsi $\forall n \in \mathbb{N}$ $|U_{\sigma(n)} - \ell| > \varepsilon$.

Or $(U_{\sigma(n)})$ bornée car (U_n) bornée donc on peut extraire une sous suite convergente:

$\exists \sigma' : \mathbb{N} \to \mathbb{N}$ ↗ tq $U_{\sigma \circ \sigma'(n)} \longrightarrow \ell'$.

On a alors $\forall n \in \mathbb{N}$, $|U_{\sigma \circ \sigma'(n)} - \ell| \geq \varepsilon$ et par passage à la limite $|\ell' - \ell| \geq \varepsilon$. Donc $\ell' \neq \ell$. (U_n) aurait alors deux valeurs d'adhérence distinctes... $\blacksquare$

Exercice 1

$f : \mathbb{R}^2 \longrightarrow \mathbb{R}$

$\binom{x}{y} \longmapsto 2x_1^2 + x_2^2 - x_1 x_2 - 3x_1 - x_2 + 4$

1) Mq $\exists !\ \bar{x} \in \mathbb{R}^2$; $\bar{x} = \min\limits_{x \in \mathbb{R}^2} f(x)$ et le calculer.

existence

$Df(x,y) = [4x_1 - x_2 - 3 \ ; 2x_2 - x_1 - 1]$

$Hess_{(x,y)} f = \begin{bmatrix} 4 & -1 \\ -1 & 2 \end{bmatrix} = \begin{bmatrix} r & s \\ s & t \end{bmatrix}$

rappel : $|Hess| > 0 \Rightarrow f$ convexe ou concave ↗ convexe si $r > 0$ (min global) ↘ concave si $r < 0$ (max global)

$|Hess| < 0$. on ne peut rien dire

donc f est convexe.

Si Hess est déf pos ($\forall p > 0$) alors strictement convexe.

Donc f strictement convexe.

De plus, f est clairement coercive. Donc $\exists !$ sol à $\bar{x} = \min\limits_{x \in \mathbb{R}^2} f(x)$.

Trouvons-la ! Trouvons les pts critiques de f :

soit (x,y) tq $[4x_1 - x_2 - 3 \ ; 2x_2 - x_1 - 1] = (0,0)$

$\begin{cases} 4x_1 - x_2 = 3 \\ 2x_2 - x_1 = 1 \end{cases} \Longrightarrow \begin{cases} 4x_1 - x_2 = 3 \\ -4x_1 + 8x_2 = 4 \end{cases} \Longrightarrow \begin{cases} x_2 = 1 \\ x_1 = 1 \end{cases}$

$\hookrightarrow (x_1, x_2) = (1 ; 1)$ unique pt fixe

c'est le minimum.

2) Calculer le premier itéré donné par l'algo du gradient à pas optimal (en partant de $(x_1^{(0)}, x_2^{(0)}) = (0,0)$)

Algo : • $x^{(k)}$ connu
• on calcule $w^{(k)} = -\nabla f(x^{(k)})$
• trouver $\rho_k \geq 0$ tq $f(x^{(k)} + \rho_k \cdot w^{(k)}) \leq f(x^{(k)} + t \cdot w^{(k)})$ $\forall t \geq 0$
• $x^{(k+1)} = x^{(k)} + \rho_k w^{(k)}$

Ici $w^{(0)} = (3, 1)$. ρ_0 minimise $t \mapsto \varphi(t) = f(x^{(0)} + t . w^{(0)})$
$= f((0,0) + t(3,1))$
$= f(3t, t)$

De plus, il faut $\varphi'(\rho_0) = 0$

$\hookrightarrow \varphi'(t) = f'(x^{(0)} + t\, w^{(0)}) \cdot w^{(0)}$

$\varphi'(\rho_0) = f'(3\rho_0, \rho_0) \cdot \begin{bmatrix} 3 \\ 1 \end{bmatrix}$

$= \begin{bmatrix} 11 & \rho_0 - 3 \\ -\rho_0 & -1 \end{bmatrix} \cdot \begin{bmatrix} 3 \\ 1 \end{bmatrix}$

$= 32\rho_0 - 10$

$\Rightarrow \rho_0 = \frac{5}{16}$.

Ainsi $x^{(1)} = (0,0) + \frac{5}{16}(3,1) = \left(\frac{15}{16}, \frac{5}{16}\right)$

Quest: Pourriez vous l'appliquer avec l'algo du gradient à pt fixe ? (ds par $\alpha = 0{,}5$)

algo: • $x^{(k)}$ connu
• $w^{(k)} = -\nabla f(x^{(k)})$
• $x^{(k+1)} = x^{(k)} + \underset{\text{fixe} = 5}{\rho}\, w^{(k)}$

Ici $w^{(0)} = (+3, +1)$
$x^{(1)} = (3, 1) \cdot \frac{1}{2} = \left(\frac{3}{2}, \frac{1}{2}\right)$ MOINS PRÉCIS