

Gradient à pas optimal

Leçons: 162, 219, 223, 226, 229, 233, 253

Ref.: Bernis, Analyse pour l'optimisation : 40 développements p 159

Cadre:

Soit $A \in \mathcal{S}_n^{++}(\mathbb{R})$, $b \in \mathbb{R}^n \setminus \{0\}$ et $\bar{x} \in \mathbb{R}^n$ l'unique solution de $Ax = b$.

On pose $\Phi : \mathbb{R}^n \rightarrow \mathbb{R}$

$$x \mapsto \frac{1}{2} \langle Ax, x \rangle - \langle b, x \rangle$$

et on notera : $\langle x, y \rangle_A = {}^t x A y$; $\|x\|_A = \sqrt{\langle Ax, x \rangle}$; $\|x\| = \|x\|_2$

Prop.:

Soit $a \in \mathbb{R}^n$. On définit la suite $(x_k)_{k \in \mathbb{N}}$ par :

$$\begin{cases} x_0 = a \\ x_k = \frac{\|\nabla \Phi(x_k)\|^2}{\|\nabla \Phi(x_k)\|_A^2} \text{ si } x_k \neq \bar{x}, 0 \text{ sinon} \\ x_{k+1} = x_k - x_k \nabla \Phi(x_k) \end{cases}$$

Alors, $x_k \xrightarrow[k \rightarrow +\infty]{} \bar{x}$

1) Rappels

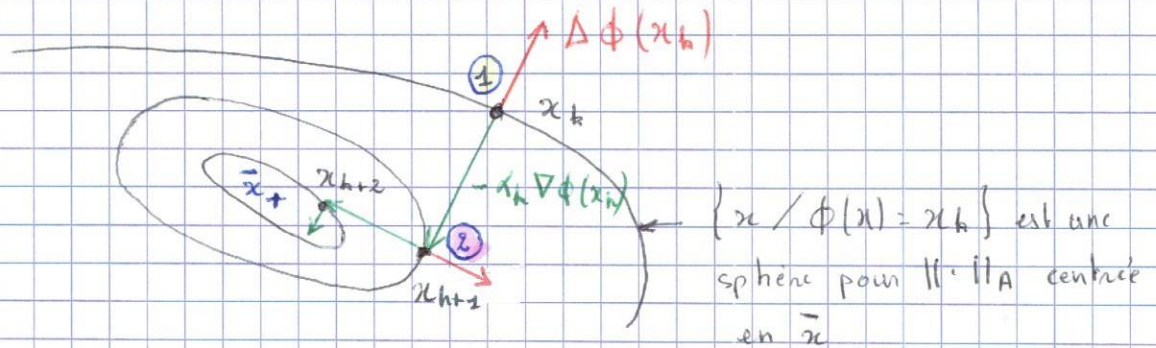
Φ est concave (grâce au $\|x\|_A^2$) et strictement convexe ($d^2 \Phi(x) = A$) sur $\mathbb{R}^n$ qui est convexe et fermé, donc Φ admet un unique point de minimum.

$$d\Phi(x)(h) = \frac{1}{2} \times 2 \langle Ax, h \rangle - \langle b, h \rangle = \langle Ax - b, h \rangle \quad \forall x, h \in \mathbb{R}^n$$

donc $\nabla \Phi(x) = Ax - b = A(x - \bar{x})$

$\nabla \phi(x) = 0 \Leftrightarrow x = \bar{x}$ donc l'unique point de minimum de ϕ est $\bar{x}$

2) Explication de l'algorithme



① En x_k , $\nabla \phi(x_k)$ donne la direction (et le sens...) dans laquelle ϕ croît le plus vite, il est donc naturel de se déplacer dans la direction de $-\nabla \phi(x_k)$ pour approcher le point de minimum de ϕ .

Plus précisément, on cherche $x_k \in \mathbb{R}$ réalisant le minimum de

$$f: \mathbb{R} \rightarrow \mathbb{R}$$

$$t \mapsto \phi(x_k - t \nabla \phi(x_k))$$

f est un polynôme de degré 2 en t de coefficient dominant > 0

donc f admet un unique point de minimum.

$$\begin{aligned} \text{En effet, } f(t) &= \phi(x_k) - t \langle Ax_k - b, \nabla \phi(x_k) \rangle + \frac{t^2}{2} \|\nabla \phi(x_k)\|_A^2 \\ &= \phi(x_k) - \|\nabla \phi(x_k)\|^2 t + \frac{\|\nabla \phi(x_k)\|_A^2}{2} t^2 \end{aligned}$$

$$f \text{ atteint son minimum en } x_k = \frac{\|\nabla \phi(x_k)\|^2}{\|\nabla \phi(x_k)\|_A^2}$$

② On a alors $f'(x_k) = 0 = \langle \nabla \phi(x_k - x_k \nabla \phi(x_k)), -\nabla \phi(x_k) \rangle$

$$\text{soit } \langle \nabla \phi(x_{k+1}), \nabla \phi(x_k) \rangle = 0$$

3) Lemme: (Kantorovich)

Soient $\lambda_1 \dots \lambda_n > 0$ les vp. de A , $\lambda_{\min}$ et $\lambda_{\max}$ sont... et $x \in \mathbb{R}^n \setminus \{0\}$

Alors:

$$\frac{\|x\|^4}{\|x\|_A^2 \|x\|_{A^{-1}}^2} \geq 4 \times \frac{\lambda_{\min} \times \lambda_{\max}}{(\lambda_{\min} + \lambda_{\max})^2}$$

$A \in \mathcal{S}_n^{++}(\mathbb{R})$, donc par le théorème spectral il existe $e = (e_1 \dots e_n)$ une base orthonormée de $\mathbb{R}^n$ constituée de $\vec{v}_p$ de A associés à $(\lambda_1 \dots \lambda_n)$. C'est également une base de $\vec{v}_p$ de A^{-1} associés à $(\frac{1}{\lambda_1} \dots \frac{1}{\lambda_n})$

Soit $x = (x_1 \dots x_n)$ dans e . Alors

$$\begin{aligned} \|x\|_A \|x\|_{A^{-1}} &= \left(\sum_{i=1}^n \lambda_i x_i^2 \right)^{1/2} \left(\sum_{i=1}^n \frac{1}{\lambda_i} x_i^2 \right)^{1/2} \\ &= \sqrt{\frac{\lambda_{\max}}{\lambda_{\min}}} \left(\sum_{i=1}^n \frac{\lambda_i}{\lambda_{\max}} x_i^2 \right)^{1/2} \left(\sum_{i=1}^n \frac{\lambda_{\min}}{\lambda_i} x_i^2 \right)^{1/2} \end{aligned}$$

$$\|x\|_A \|x\|_{A^{-1}} \leq \sqrt{\frac{\lambda_{\max}}{\lambda_{\min}}} \times \frac{1}{2} \sum_{i=1}^n \left(\frac{\lambda_i}{\lambda_{\max}} + \frac{\lambda_{\min}}{\lambda_i} \right) x_i^2 \quad \leftarrow ab \leq \frac{1}{2}(a^2 + b^2)$$

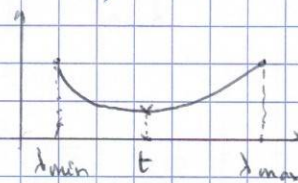
Soit $g : [\lambda_{\min}, \lambda_{\max}] \rightarrow \mathbb{R}$

$$t \mapsto \frac{t}{\lambda_{\max}} + \frac{\lambda_{\min}}{t}$$

g est convexe comme somme de deux fonctions convexes, et

$$g(\lambda_{\min}) = 1 + \frac{\lambda_{\min}}{\lambda_{\max}} = g(\lambda_{\max})$$

donc : $\forall t \in [\lambda_{\min}, \lambda_{\max}], g(t) \leq 1 + \frac{\lambda_{\min}}{\lambda_{\max}}$



$$\text{donc } \|x\|_A \|x\|_{A^{-1}} \leq \frac{1}{2} \sqrt{\frac{\lambda_{\max}}{\lambda_{\min}}} \times \frac{\lambda_{\max} + \lambda_{\min}}{\lambda_{\max}} \times \|x\|^2$$

et en élevant au carré,

$$\frac{\|x\|^4}{\|x\|_A^2 \|x\|_{A^{-1}}^2} \geq 4 \times \frac{\lambda_{\min} \lambda_{\max}}{(\lambda_{\min} + \lambda_{\max})^2}$$

4) Pour montrer que $x_k \rightarrow \bar{x}$, m.g. $\|x_{k+1} - \bar{x}\| \leq \frac{\lambda_{\max} - \lambda_{\min}}{\lambda_{\max} + \lambda_{\min}} \|x_k - \bar{x}\|$

On pose $g_k = \nabla \Phi(x_k) = A(x_k - \bar{x})$. On a alors:

(2) $x_{k+1} = x_k - \alpha_k g_k$

(3) $\alpha_k = \frac{\|g_k\|^2}{\|g_k\|_A^2}$

(4) $\langle g_{k+1}, g_k \rangle = 0$

$$\|x_{k+1} - \bar{x}\|_A^2 = \langle A(x_{k+1} - \bar{x}), x_{k+1} - \bar{x} \rangle$$

$-x_k + x_k$

on veut se débarrasser des x_{k+1}
en faisant apparaître des x_k

$$= \langle \overset{(1)}{g_{k+1}}, -\overset{(2)}{x_k g_k} \rangle + \langle A(x_{k+1} - \bar{x}), x_k - \bar{x} \rangle$$

$-x_k + x_k$

$= 0$ (4)

A symétrique

$$= \langle A(x_{k+1} - x_k), A(x_k - \bar{x}) \rangle + \langle A(x_k - \bar{x}), x_k - \bar{x} \rangle$$

$$= -\alpha_k \overset{(2)}{\langle g_k, g_k \rangle} + \overset{(1)}{\|x_k - \bar{x}\|_A^2}$$

$$\|x_{k+1} - \bar{x}\|_A^2 = \overset{(3)}{-\frac{\|g_k\|^4}{\|g_k\|_A^2}} + \|x_k - \bar{x}\|_A^2$$

et $\|x_k - \bar{x}\|_A^2 = \langle A(x_k - \bar{x}), x_k - \bar{x} \rangle = \langle \overset{(1)}{A(x_k - \bar{x})}, \overset{(1)}{A^{-1} A(x_k - \bar{x})} \rangle$

donc $\|x_k - \bar{x}\|_A^2 = \|g_k\|_{A^{-1}}^2$

$$\text{donc } \|x_{k+1} - \bar{x}\|_A^2 \leq \left(1 - \frac{\|g_k\|^4}{\|g_k\|_A^2 \|g_k\|_{A^{-1}}^2} \right) \|x_k - \bar{x}\|_A^2$$

$$\leq 1 \times \frac{\lambda_{\max} \lambda_{\min}}{(\lambda_{\min} + \lambda_{\max})^2}$$

donc $\|x_{k+1} - \bar{x}\|_A^2 \leq \left(\frac{\lambda_{\max} - \lambda_{\min}}{\lambda_{\max} + \lambda_{\min}} \right)^2 \|x_k - \bar{x}\|_A^2$

6) Conclusion et estimation en norme euclidienne

En itérant l'inégalité précédente, on a :

$$\sqrt{\lambda_{\min}} \|\cdot\| \leq \|x_k - \bar{x}\|_A \leq \underbrace{\left(\frac{\lambda_{\max} - \lambda_{\min}}{\lambda_{\max} + \lambda_{\min}} \right)^k}_{< 1} \|x_0 - \bar{x}\|_A \leq \sqrt{\lambda_{\max}} \|\cdot\|$$

donc $x_k \xrightarrow[k \rightarrow \infty]{} \bar{x}$ et la convergence est géométrique.

Pour avoir une estimation en $\|\cdot\|$ plutôt qu'en $\|\cdot\|_A$, on note que

$$\sqrt{\lambda_{\min}} \|\cdot\| \leq \|\cdot\|_A \leq \sqrt{\lambda_{\max}} \|\cdot\|$$

d'où

$$\|x_k - \bar{x}\| \leq \sqrt{\frac{\lambda_{\max}}{\lambda_{\min}}} \times \left(\frac{\lambda_{\max} - \lambda_{\min}}{\lambda_{\max} + \lambda_{\min}} \right)^k \|x_0 - \bar{x}\|$$

Compléments

1) Lignes de niveaux de Φ

On veut expliciter $\{x \in \mathbb{R}^n / \Phi(x) = \Phi(x_k) = c\}$

$$\Phi(x) = c \Leftrightarrow \frac{1}{2} \|x\|_A^2 - \langle b, x \rangle = c$$

$$\Leftrightarrow \|x\|_A^2 - 2 \langle A\bar{x}, x \rangle = c$$

$$\Leftrightarrow \|x - \bar{x}\|_A^2 = c + \| \bar{x} \|_A^2$$

Les lignes de niveau de Φ sont bien des sphères pour $\|\cdot\|_A$ centrées en $\bar{x}$