Gradient à pas optimal pour la résolution de systèmes linéaires:

Référence : Bernis

Énoncé : Soit $A \in S_n^{++}(\mathbb{R})$, $B \in \mathbb{R}^n$, on note $\langle \cdot ; \cdot \rangle_A$ le p.s.

$\langle x ; y \rangle_A = {}^t x A y = \langle x ; Ay \rangle_{\mathbb{R}^n}$

Alors $\varphi : \mathbb{R}^n \to \mathbb{R}$ atteint son minimum

$$x \mapsto \frac{1}{2} \|x\|_A^2 + {}^t x b$$

en x^* l'unique solution de $Ax = B$, et uniquement en x^*.

Si $a \in \mathbb{R}^n$ $a \neq x^*$, la suite définie par

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

Alors $(x_k)_{k \geq 0}$ converge vers x^*

Étapes : (E1) Reformuler en recherche de minimum

(E2) Construction de la suite minimisante

(E3) Convergence de la suite.

E1 Reformuler en pb de recherche de minimum

Mq Φ admet un unique point de minimum : x^*

$\forall x \in \mathbb{R}^n, \forall h \in \mathbb{R}^n,$

$$\Phi(x+h) = \frac{1}{2} \langle x+h ; x+h \rangle_A - {}^t x B - {}^t h B$$

Donc

$$\Phi(x+h) = \frac{1}{2} \|x\|_A^2 + \frac{1}{2} \|h\|_A^2 + \langle h ; x \rangle_A - {}^t x B - {}^t h B$$

$$= \Phi(x) + {}^t h A x - {}^t h B + \frac{1}{2} \|h\|_A^2$$

$$= \Phi(x) + \langle h ; Ax - B \rangle + \frac{1}{2} \|h\|_A^2$$

Ainsi, Φ est différentiable sur $\mathbb{R}$ et $\forall x \in \mathbb{R}^n$, $\nabla \Phi(x) = Ax - B$

<u>Lemme</u> : si $f : \mathbb{R}^n \to \mathbb{R}$ est différentiable et f atteint un extremum en x^* alors $\nabla f(x^*) = 0$

<u>Preuve</u> : $f(x^* + ty) = f(x^*) + t \langle \nabla f(x^*) ; y \rangle + t \|y\| \varepsilon(t)$

Donc si c'est un min, $f(x^* + ty) - f(x^*) \geq 0 \quad \forall y, \forall t$

Donc $\langle \nabla f(x^*) ; y \rangle + \|y\| \varepsilon(t) \geq 0 \quad \forall y, \forall t$

Donc $\langle \nabla f(x^*) ; y \rangle \geq 0 \quad \forall y$

Donc $\langle \nabla f(x^*) ; -y \rangle \geq 0 \quad \forall y$

Donc $\langle \nabla f(x^*) ; y \rangle = 0 \quad \forall y$,

Donc $\nabla f(x^*) = 0$.

$\nabla \Phi(x) = 0$ ssi $Ax = B$ ssi $x = x^*$.

Donc x^* est le seul extremum possible pour Φ.

et $\forall y \in \mathbb{R}^n$

$$\phi(x^* + y) = \phi(x^*) + \frac{1}{2}\|y\|^2$$

Donc $\phi(x^* + y) \geq \phi(x^*)$ $\forall y \in \mathbb{R}^n$ et x^* est bien un point de minimum.

Ainsi, approcher x^*, solution de $Ax = B$ peut se faire en minimisant ϕ.

E2 Construction de la suite minimisante

Soit $x \in \mathbb{R}^n$, $f : \mathbb{R} \to \mathbb{R}$, $t \mapsto \phi(x - t\nabla\phi(x))$ est dérivable comme composée de fonctions diff et pour $t \in \mathbb{R}$

$$f'(t) = d\phi_{x - t\nabla\phi(x)}(-\nabla\phi(x)) = -\langle \nabla\phi(x) ; \nabla\phi(x - t\nabla\phi(x))\rangle$$

$$= \langle -\nabla\phi(x) ; A(x - t\nabla\phi(x)) - B\rangle$$

$$= -\langle \underbrace{\nabla\phi(x)}_{= Ax - B} ; Ax - B\rangle + t\langle \nabla\phi(x) ; A\nabla\phi(x)\rangle$$

$$= -\|\nabla\phi(x)\|^2 + t\|\nabla\phi(x)\|_A^2$$

Donc $f'(\alpha) = 0$ ssi $\alpha = \dfrac{\|\nabla\phi(x)\|^2}{\|\nabla\phi(x)\|_A^2}$ si $\nabla\phi(x) \neq 0$

si $\nabla\phi(x) = 0$ alors $x = x^*$ et on a trouvé le min.

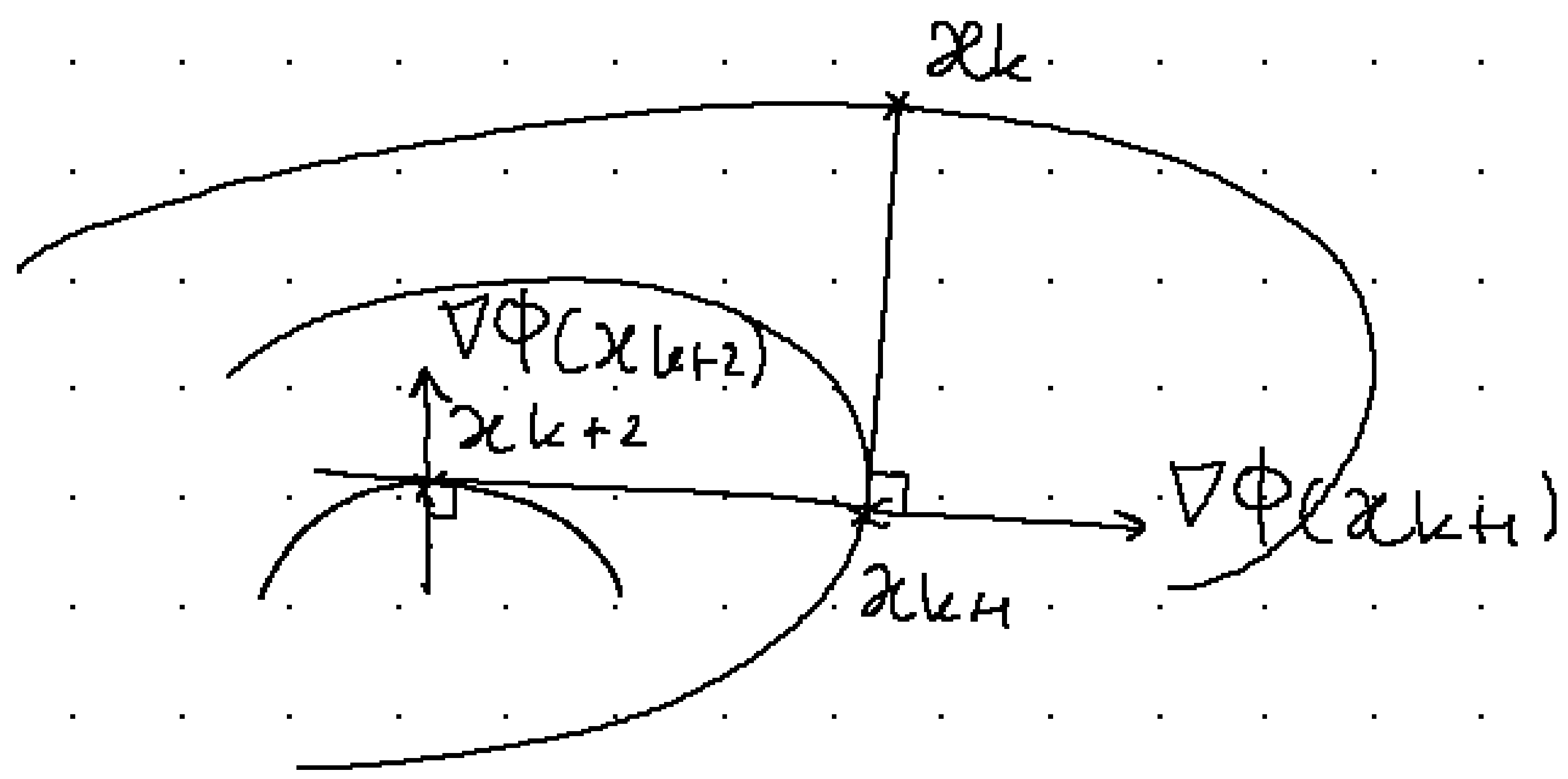
On remarque au passage que au point α minimisant f, $\langle \nabla\phi(x) ; \nabla\phi(x - \alpha\nabla\phi(x))\rangle = 0$. $(*)$

Ainsi, pour tout $x \in \mathbb{R}^n$ $x \neq x^*$, f a un unique point de minimum α.

On prend alors $a \in \mathbb{R}^n$ $a \neq x^*$ et on construit

$$\begin{cases} x_0 = a \\ \alpha_k = \begin{cases} \dfrac{\|\nabla\phi(x_k)\|^2}{\|\nabla\phi(x_k)\|_A^2} & \text{si } \nabla\phi(x_k) \neq 0, \\ 0 & \text{sinon} \end{cases} \\ x_{k+1} = \begin{cases} x_k - \alpha_k \nabla\phi(x_k) & \text{si } x_k \neq x^* \\ x_k & \text{sinon} \end{cases} \end{cases}$$

et la rq $(*)$ nous indique que deux directions de descente consécutives sont $\perp$:



(E3) Étude de la suite:

Si $\exists p \in \mathbb{N}$ tq $x_p = x^*$ alors la suite est stationnaire à partir de p et on a trouvé le minimum.

Sinon, supposons que $\forall k \in \mathbb{N}$ $x_k \neq x^*$.

(E3.1) Lemme : Inégalité de Kantorovitch

Soit $x \in \mathbb{R}^n$, $x \neq 0$,

$A \in S_n^{++}(\mathbb{R})$ donc par le thm spectral, il existe une b.o.n. $\mathcal{B} = (e_1, \dots, e_n)$ de vecteurs propres.

On note $\forall i \in [\![1; n]\!]$ λ_i la valeur propre associée à λ_i.

Ainsi, $\forall i \in [\![1; n]\!]$ $Ae_i = \lambda_i e_i$

et donc $e_i = \lambda_i A^{-1} e_i$

comme A est inversible, $\forall i, \lambda_i \neq 0$ donc $A^{-1} e_i = \frac{1}{\lambda_i} e_i$

[Et $A^{-1} A = I_n$ donc $\underbrace{{}^tA}_{=A} {}^t(A^{-1}) = I_n$ donc ${}^tA^{-1} = A^{-1}$ et $A^{-1} \in S_n^{++}(\mathbb{R})$, on peut alors définir $\|\cdot\|_{A^{-1}} = \langle \cdot ; A^{-1} \cdot \rangle$]

Donc si $x = x_1 e_1 + \dots + x_n e_n$, alors

$$\begin{aligned}
\|x\|_A^2 \|x\|_{A^{-1}}^2 &= \langle x ; Ax \rangle \langle x ; A^{-1} x \rangle \\
&= \langle x ; \lambda_1 x_1 e_1 + \dots + \lambda_n x_n e_n \rangle \\
&\quad \times \langle x ; \frac{1}{\lambda_1} x_1 e_1 + \dots + \frac{1}{\lambda_n} x_n e_n \rangle \\
&= \left(\sum_{i=1}^{n} \lambda_i x_i^2 \right) \times \left(\sum_{j=1}^{n} \frac{x_j^2}{\lambda_j} \right) \\
&= \frac{\lambda_{max}}{\lambda_{min}} \left(\sum_{i=1}^{n} \frac{\lambda_i}{\lambda_{max}} x_i^2 \right) \left(\sum_{i=1}^{n} \frac{\lambda_{min}}{\lambda_i^2} x_i^2 \right)
\end{aligned}$$

or $ab \leq \frac{1}{2}(a^2 + b^2)$ $\forall (a;b) \in \mathbb{R}^2$, donc on a :

$$\sqrt{\sum_{i=1}^{n} \frac{\lambda_i}{\lambda_{max}} x_i^2} \times \sqrt{\sum_{i=1}^{n} \frac{\lambda_{min}}{\lambda_i^2} x_i^2} \leq \frac{1}{2} \left(\sum_{i=1}^{n} \left(\frac{\lambda_i}{\lambda_{max}} + \frac{\lambda_{min}}{\lambda_i} \right) x_i^2 \right)$$

or $g : [\lambda_{min} ; \lambda_{max}] \longrightarrow \mathbb{R}$ est convexe

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

(somme de deux f^c convexes) alors

$$g(\varepsilon\lambda_{max} + (1-\varepsilon)\lambda_{min}) \leq \varepsilon g(\lambda_{max}) + (1-\varepsilon) g(\lambda_{min})$$

$$\leq \varepsilon\left(1+\frac{\lambda_{min}}{\lambda_{max}}\right) + (1-\varepsilon)\left(\frac{\lambda_{min}}{\lambda_{max}}+1\right)$$

$$\leq 1 + \frac{\lambda_{min}}{\lambda_{max}}$$

D'où $\forall t \in [\lambda_{min} ; \lambda_{max}]$ $g(t) \leq 1 + \frac{\lambda_{min}}{\lambda_{max}}$

Donc

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

d'où $$\frac{\|x\|^2}{\|x\|_A \|x\|_{A^{-1}}} \geq 2\frac{\sqrt{\lambda_{min}\lambda_{max}}}{\lambda_{min}+\lambda_{max}}$$ □

(E3.2) Convergence de la suite soit $k \in \mathbb{N}$,

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

$$= \langle x_{k+1} - x_k ; Ax_{k+1} - B\rangle$$

$$+ \langle x_k - x^* ; Ax_{k+1} - Ax^*\rangle$$

$$= \underbrace{\langle -\alpha_k \nabla\Phi(x_k) ; \nabla\Phi(x_{k+1})\rangle}_{=0}$$

$$+ \langle x_k - x^* ; Ax_{k+1} - Ax^*\rangle$$

$$= \langle x_k - x^* ; A(x_{k+1} - x_k)\rangle +$$

$$+ \langle x_k - x^* ; A(x_k - x^*)\rangle$$

Donc

$$\|x_{k+1}-x^*\|_A^2 = \|x_k - x^*\|_A^2 + \langle x_k - x^* ; -\alpha_k A \nabla\varphi(x_k)\rangle$$

$$= \|x_k - x^*\|_A^2 - \langle \underbrace{A(x_k - x^*)}_{=\nabla\varphi(x_k)} ; \alpha_k \nabla\varphi(x_k)\rangle$$

$$= \|x_k - x^*\|_A^2 - \frac{\|\nabla\varphi(x_k)\|^2}{\|\nabla\varphi(x_k)\|_A^2} \times \|\nabla\varphi(x_k)\|^2$$

$$= \|x_k - x^*\|_A^2 - \frac{\|\nabla\varphi(x_k)\|^4}{\|\nabla\varphi(x_k)\|_A^2}$$

et $\|x_k - x^*\|_A^2 = \langle x_k - x^* ; A(x_k - x^*)\rangle$

$$= \langle A(x_k - x^*) ; A^{-1}A(x_k - x^*)\rangle$$

$$= \|\nabla\varphi(x_k)\|_{A^{-1}}^2$$

donc $\|x_{k+1}-x^*\|_A^2 = \|x_k - x^*\|_A^2 \underbrace{\left(1 - \frac{\|\nabla\varphi(x_k)\|^4}{\|\nabla\varphi(x_k)\|_A^2 \|\nabla\varphi(x_k)\|_{A^{-1}}^2}\right)}$

Kantorovitch $\rightarrow$ $\leq \underbrace{1 - 4\frac{\lambda_{max}\lambda_{min}}{(\lambda_{max}+\lambda_{min})^2}}$

$$= \frac{(\lambda_{max}-\lambda_{min})^2}{(\lambda_{max}+\lambda_{min})^2}$$

Donc $\|x_{k+1}-x^*\|_A \leq \|x_k - x^*\|_A \dfrac{\lambda_{max}-\lambda_{min}}{\lambda_{max}+\lambda_{min}}$ □