Méthode du gradient à pas optimal.

Bernis

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

On pose $\phi : \mathbb{R}^n \longrightarrow \mathbb{R}$
$$x \longmapsto \frac{1}{2}\langle Ax, x\rangle - \langle b, x\rangle$$

1) ϕ atteint son minimum en $\bar{x}$ et uniquement en ce point.

2) Soit $a \in \mathbb{R}^n \setminus \{\bar{x}\}$ et $(x_k)_{k\geq 0}$ définie par

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

La suite (x_k) converge vers $\bar{x}$ et on a $\forall k \in \mathbb{N}$, $\|x_{k+1} - \bar{x}\| \leq \sqrt{\dfrac{\lambda_{max}}{\lambda_{min}}}\left(\dfrac{\lambda_{max} - \lambda_{min}}{\lambda_{max} + \lambda_{min}}\right)^{k+1} \|x_0 - \bar{x}\|$

Lemme : $\forall x \in \mathbb{R}^n \setminus \{0\}$, $\dfrac{\|x\|^4}{\|x\|_{A^{-1}}^2 \|x\|_A^2} \geq 4 \dfrac{\lambda_{min}\lambda_{max}}{(\lambda_{max} + \lambda_{min})^2}$

Preuve Lemme : Par le Théorème Spectral, A se diagonalise dans une base orthonormée de vecteurs propres $(e_1, \dots, e_n)$ et on note λ_i la valeur propre associée à e_i. Comme $A \in S_n^{++}(\mathbb{R})$, e_i est aussi vecteur propre de A^{-1} de valeur propre $\frac{1}{\lambda_i}$.

On a alors $\|x\|_A \|x\|_{A^{-1}} = \sqrt{\sum_{i=1}^n \lambda_i x_i^2}\sqrt{\sum_{i=1}^n \frac{1}{\lambda_i} x_i^2} = \sqrt{\frac{\lambda_{max}}{\lambda_{min}}}\sqrt{\sum_{i=1}^n \frac{\lambda_i}{\lambda_{max}} x_i^2}\sqrt{\sum_{i=1}^n \frac{\lambda_{min}}{\lambda_i} x_i^2}$

$$\leq \frac{1}{2}\sqrt{\frac{\lambda_{max}}{\lambda_{min}}} \sum_{i=1}^n \left(\frac{\lambda_i}{\lambda_{max}} + \frac{\lambda_{min}}{\lambda_i}\right) x_i^2 \quad \text{car } ab \leq \frac{1}{2}(a^2+b^2)$$

Or $g : t \in [\lambda_{min}, \lambda_{max}] \longmapsto \dfrac{t}{\lambda_{max}} + \dfrac{\lambda_{min}}{t}$ est convexe, comme somme de 2 fonctions convexes

Donc comme $g(\lambda_{min}) = g(\lambda_{max}) = 1 + \dfrac{\lambda_{min}}{\lambda_{max}}$, on a $g(t) \leq 1 + \dfrac{\lambda_{min}}{\lambda_{max}}$

D'où $\|x\|_A \|x\|_{A^{-1}} \leq \dfrac{1}{2}\sqrt{\dfrac{\lambda_{max}}{\lambda_{min}}}\left(1 + \dfrac{\lambda_{min}}{\lambda_{max}}\right)\|x\|^2$

D'où, en, élevant au carré, on obtient le résultat.

Preuve Théorème :

1) $\forall h \in \mathbb{R}^n$, $\phi(x+h) = \frac{1}{2} \langle A(x+h), x+h \rangle - \langle b, x+h \rangle$

$$= \phi(x) + \langle Ax - b, h \rangle + \frac{1}{2} \|h\|_A^2$$

$$= \phi(x) + \langle Ax - b, h \rangle + \underset{h\to 0}{o}(\|h\|)$$

Ainsi, $\phi(\bar{x}+h) = \phi(\bar{x}) + \frac{1}{2}\|h\|_A^2$ et comme ϕ est différentiable (car polynomiale), on a

$$D_x\phi(h) = {}^t(Ax-b)h \quad \text{et} \quad D_{\bar{x}}\phi(h) = 0.$$

Comme $\nabla\phi = Ax - b$ et $\nabla^2\phi = A \in \mathcal{S}_n^{++}(\mathbb{R})$, alors ϕ est strictement convexe.

Donc $\bar{x}$ est l'unique minimum de ϕ.

2) On pose $f : \mathbb{R} \to \mathbb{R}$, $t \mapsto \phi(x_k - t\nabla\phi(x_k))$ qui atteint son minimum en α_k.

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

$= \phi(x_k) - t\|\nabla\phi(x_k)\|^2 + \frac{t^2}{2}\|\nabla\phi(x_k)\|_A^2$ et l'étude du polynome donne un minimum en α_k.

On a alors $f'(\alpha_k) = 0$ d'où $0 = -\|\nabla\phi(x_k)\|^2 + \alpha_k\|\nabla\phi(x_k)\|_A^2$

$$= -\langle \nabla\phi(x_k), \nabla\phi(x_k)\rangle + \alpha_k\langle A\nabla\phi(x_k), \nabla\phi(x_k)\rangle$$

$$= -\langle \nabla\phi(x_k) - \alpha_k A\nabla\phi(x_k), \nabla\phi(x_k)\rangle$$

$$= -\langle A(x_k - \alpha_k\nabla\phi(x_k)) - b, \nabla\phi(x_k)\rangle$$

$$= -\langle \nabla\phi(x_{k+1}), \nabla\phi(x_k)\rangle \quad (*)$$

Ainsi, les gradients en x_{k+1} et en x_k sont orthogonaux.

On note $g_k = \nabla\phi(x_k)$

Soit $p \in \mathbb{N}$, $\|x_{p+1} - \bar{x}\|_A^2 = \langle A(x_{p+1} - \bar{x}), x_{p+1} - x_p\rangle + \langle A(x_{p+1} - \bar{x}), x_p - \bar{x}\rangle$

Or $\langle A(x_{p+1} - \bar{x}), x_{p+1} - x_p\rangle = \langle Ax_{p+1} - b, -\alpha_p g_p\rangle = -\alpha_p\langle g_{p+1}, g_p\rangle = 0$ d'après $(*)$

Donc par symétrie de A,

$$\|x_{p+1}-\bar{x}\|_A^2 = \langle A(x_{p+1}-\bar{x}), x_p-\bar{x}\rangle = \langle A(x_{p+1}-x_p), x_p-\bar{x}\rangle + \langle A(x_p-\bar{x}), x_p-\bar{x}\rangle$$
$$= \langle x_{p+1}-x_p, A(x_p-\bar{x})\rangle + \langle A(x_p-\bar{x}), A^{-1}A(x_p-\bar{x})\rangle$$
$$= -\alpha_p\|g_p\|^2 + \|g_p\|_{A^{-1}}^2$$
$$= -\frac{\|g_p\|^4}{\|g_p\|_A^2} + \|g_p\|_{A^{-1}}^2$$

D'où
$$\|x_{p+1}-\bar{x}\|_A^2 = \left(1-\frac{\|g_p\|^4}{\|g_p\|_A^2\|g_p\|_{A^{-1}}^2}\right)\|x_p-\bar{x}\|_A^2$$
$$\leq \left(1-\frac{4\lambda_{min}\lambda_{max}}{(\lambda_{max}+\lambda_{min})^2}\right)\|x_p-\bar{x}\|_A^2 \quad \text{par le lemme.}$$
$$\leq \left(\frac{\lambda_{max}-\lambda_{min}}{\lambda_{max}+\lambda_{min}}\right)^2\|x_p-\bar{x}\|_A^2.$$

D'où $\|x_{p+1}-\bar{x}\|_A \leq \left(\frac{\lambda_{max}-\lambda_{min}}{\lambda_{max}+\lambda_{min}}\right)\|x_p-\bar{x}\|_A^2 \rightarrow$ permet d'affirmer la convergence de (x_k)

De plus, on a $\sqrt{\lambda_{min}}\,\|.\| \leq \|.\|_A \leq \sqrt{\lambda_{max}}\,\|.\|$

D'où $\|x_{p+1}-\bar{x}\|_A \leq \left(\frac{\lambda_{max}-\lambda_{min}}{\lambda_{max}+\lambda_{min}}\right)^{p+1}\|x_0-\bar{x}\|_A$

D'où $\|x_{p+1}-\bar{x}\| \leq \sqrt{\frac{\lambda_{max}}{\lambda_{min}}}\left(\frac{\lambda_{max}-\lambda_{min}}{\lambda_{max}+\lambda_{min}}\right)^{p+1}\|x_0-\bar{x}\|$