18min
donc faire vite voir sauter des calculs

ALGORITHME DU GRADIENT À PAS OPTIMAL

Hiriart Urruty
Optimisation et analyse convexe
II.2 p52 + FGN AL3

13'20.

Prérequis :
- calcul de différentielle, extremums $\Rightarrow$ pt critique.
- Si f est C^2, a est un point critique et $d^2f(a) > 0$ alors a est un min local de f
- diagonalisation des matrices symétriques réelles en b.o.n.
- f convexe dérivable, $f'(c) = 0 \Rightarrow f(c) = \min f$.

Théorème : Soit $A \in S_n^{++}(\mathbb{R})$, soit $b \in \mathbb{R}^n$. Soit $f : \mathbb{R}^n \longrightarrow \mathbb{R}$, $x \longmapsto \frac{1}{2}\langle Ax, x\rangle + \langle b, x\rangle$.

On veut minimiser $f(x)$ quand x parcourt $\mathbb{R}^n$.
Il existe une unique solution $\bar{x}$ à ce problème, caractérisée par $\nabla f(\bar{x}) = 0$.
De plus, l'algorithme défini par : $x_0 \in \mathbb{R}^n$ et pour $k \geq 0$ $\begin{cases} d_k = -\nabla f(x_k) \\ x_{k+1} = x_k + t_k d_k \end{cases}$,
où t_k est l'unique réel (positif) minimisant $t \mapsto f(x_k + t d_k)$ sur $\mathbb{R}$, converge vers $\bar{x}$.

Dém :

Etape 1 : Existence et unicité de $\bar{x}$ + formules.

Soit x un point minimal de f, alors nécessairement $\nabla f(x) = 0$.
Soient $x, h \in \mathbb{R}^n$, on a

$$f(x+h) - f(x) = \frac{1}{2}\langle Ax+h, x+h\rangle + \langle b, x+h\rangle - \frac{1}{2}\langle Ax, x\rangle - \langle b, x\rangle.$$
$$= \frac{1}{2}\langle Ax, h\rangle + \frac{1}{2}\langle Ah, x\rangle + \frac{1}{2}\langle Ah, h\rangle + \langle b, h\rangle.$$
$$= \langle Ax, h\rangle + \langle b, h\rangle + \frac{1}{2}\langle Ah, h\rangle \quad \text{car } A^* = A. \qquad (*)$$
$$= \underbrace{\langle Ax+b, h\rangle}_{\text{linéaire continue}} + o(\|h\|) \quad \text{car } |\langle Ah, h\rangle| \overset{CYS}{\leq} \|A\|\|h\|\,\|h\|$$

donc f est différentiable et pour tout $x \in \mathbb{R}^n$ $\nabla f(x) = Ax + b$. et $d^2_x f(h_1, h_2) = {}^t h_2 A h_1$ terme linéaire + terme cst
donc la hessienne de f est $A \in S_n^{++}(\mathbb{R})$ donc si $\nabla f(\bar{x}) = 0$ alors $\bar{x}$ est minimum ~~local~~. Or il y a un seul point critique donc unicité :
$\nabla f(\bar{x}) = 0 \iff \bar{x} = -A^{-1}b$

On peut calculer $f(\bar{x}) = \frac{1}{2}\langle -b, -A^{-1}b\rangle + \langle b, -A^{-1}b\rangle = \frac{1}{2}\langle A^{-1}b, b\rangle - \langle A^{-1}b, b\rangle =$
$$= -\frac{1}{2}\langle A^{-1}b, b\rangle.$$

Etape 2 : On suppose pour tout $k \in \mathbb{N}$ $d_k \neq 0$.

Sinon, on aurait $\nabla f(x_k) = 0$ et donc $x_k = \bar{x}$ et l'algorithme converge en temps fini.
On remarque que tant que $d_k \neq 0$, $f(x_k) - f(\bar{x}) > 0$. ← minimum ! → plus besoin de valeur absolue par la suite

Etape 3 : Calcul de t_k :

Soit $t \in \mathbb{R}$, par $(*)$, on a $h = td_k$

$$g(t) = f(x_k + td_k) = f(x_k) + \langle Ax_k, td_k\rangle + \langle b, td_k\rangle + \frac{1}{2}\langle A td_k, td_k\rangle.$$
$$= f(x_k) + \langle \underbrace{Ax_k + b}_{-d_k}, d_k\rangle t + \frac{1}{2}\langle Ad_k, d_k\rangle t^2.$$
$$= f(x_k) - \|d_k\|^2 t + \frac{1}{2}\langle Ad_k, d_k\rangle t^2. \qquad (**)$$

$g'(t)=0 \iff -\|d_k\|^2 + t\langle Ad_k, d_k\rangle = 0 \iff t_k = \dfrac{\|d_k\|^2}{\langle Ad_k, d_k\rangle}$ (>0) ← on peut diviser car $d_k \neq 0$ et $A \in S_n^{++}$

g est convexe donc g admet un (unique) minimum en t_k.
← uniq pt critique

Etape 4: Pour $k\in\mathbb{N}$ on a $f(x_{k+1}) - f(\bar{x}) = (f(x_k) - f(\bar{x})) \times \dots$

soit $k \in \mathbb{N}$

On a $f(x_{k+1}) - f(\bar{x}) = f(x_k + t d_k) - f(\bar{x}) = f(x_k) - t_k\|d_k\|^2 + \dfrac{t_k^2}{2}\langle Ad_k, d_k\rangle - f(\bar{x})$ par $(**)$

$$= f(x_k) - \frac{\|d_k\|^4}{\langle Ad_k,d_k\rangle} + \frac{\|d_k\|^4}{2\langle Ad_k,d_k\rangle} - f(\bar{x}) = f(x_k) - \frac{1}{2}\frac{\|d_k\|^4}{\langle Ad_k,d_k\rangle} - f(\bar{x})$$

$$= \left(f(x_k) - f(\bar{x})\right)\left[1 - \frac{\|d_k\|^4}{2\left(f(x_k)-f(\bar{x})\right)\langle Ad_k,d_k\rangle}\right].$$

Or $\langle A^{-1}d_k, d_k\rangle = \langle A^{-1}(Ax_k+b), Ax_k+b\rangle = \langle x_k, Ax_k\rangle + \langle x_k, b\rangle + \underbrace{\langle A^{-1}b, Ax_k\rangle}_{=\langle b, x_k\rangle} + \underbrace{\langle A^{-1}b, b\rangle}_{=-2f(\bar{x})}$ $A = A^*$

$$= 2\left(\frac{1}{2}\langle Ax_k, x_k\rangle + \langle b, x_k\rangle - f(\bar{x})\right)$$

$$= 2\left(f(x_k) - f(\bar{x})\right)$$

donc $f(x_{k+1}) - f(\bar{x}) = \left(f(x_k) - f(\bar{x})\right)\left(1 - \dfrac{\|d_k\|^4}{\langle A^{-1}d_k, d_k\rangle\langle Ad_k, d_k\rangle}\right)$

Etape 5: Erreur géométrique par application du lemme.

D'après l'inégalité de Kantorovitch,

$$\frac{\|d_k\|^4}{\langle A^{-1}d_k,d_k\rangle\langle Ad_k,d_k\rangle} \geq 4\frac{c(A)}{(c(A)+1)^2} \quad \text{donc}$$

$$f(x_{k+1}) - f(\bar{x}) \leq \left(f(x_k) - f(\bar{x})\right)\left[1 - \frac{4c(A)}{(c(A)+1)^2}\right] = \left(f(x_k) - f(\bar{x})\right)\left[\frac{c(A)-1}{c(A)+1}\right]^2.$$

$$\underset{\text{récurrence}}{\leq \dots} \leq \left(f(x_0) - f(\bar{x})\right)\left[\frac{c(A)-1}{c(A)+1}\right]^{2(k+1)}$$

Etape 6: Calcul de l'erreur.

On a $\langle A(x_k - \bar{x}), x_k - \bar{x}\rangle = \langle Ax_k, x_k\rangle - \langle Ax_k, \bar{x}\rangle - \langle A\bar{x}, x_k\rangle + \langle A\bar{x}, \bar{x}\rangle$.

ou $= \langle Ax_k + b - \underbrace{A\bar{x} - b}_{=0}, A^{-1}(Ax_k+b) - A^{-1}(A\bar{x}+b)\rangle = \langle d_k, A^{-1}d_k\rangle$

$$= \langle Ax_k, x_k\rangle + \langle Ax_k, A^{-1}b\rangle + \langle b, x_k\rangle + \langle b, A^{-1}b\rangle$$

$$= \underbrace{\langle Ax_k, x_k\rangle + \langle x_k, b\rangle + \langle b, x_k\rangle}_{2f(x_k)} + \underbrace{\langle b, A^{-1}b\rangle}_{=-2f(\bar{x})}$$

$$= 2\left(f(x_k) - f(\bar{x})\right)$$

b.o.n pour diagonalisat° de A sym réelle

donc $f(x_k) - f(\bar{x}) = \dfrac{1}{2}\langle A(x_k - \bar{x}), x_k - \bar{x}\rangle \geq \dfrac{1}{2}\lambda_{min}\|x_k - \bar{x}\|^2$.

Donc

$$\|x_k - \bar{x}\|^2 \leq \frac{2\cdot\left(f(x_0) - f(\bar{x})\right)}{\lambda_{min}}\underbrace{\left(\frac{c(A)-1}{c(A)+1}\right)}_{<1}^{2k} \xrightarrow[k\to+\infty]{} 0.$$

Donc $x_k \xrightarrow[k\to+\infty]{} \bar{x}$.

ne pas le dém à l'oral.

Inégalité de Kantorovitch : Soit $A \in S_n^{++}(\mathbb{R})$ avec $0 < \lambda_1 \leq \dots \leq \lambda_n$ les vp. On pose $c(A) = \frac{\lambda_1}{\lambda_n}$ le conditionnement de A, alors

$$\forall x \in \mathbb{R} \quad \frac{\|x\|^4}{\langle Ax, x\rangle \langle A^{-1}x, x\rangle} \geq 4\left(\sqrt{\frac{\lambda_n}{\lambda_1}} + \sqrt{\frac{\lambda_1}{\lambda_n}}\right)^{-2} = 4\,\frac{c(A)}{(c(A)+1)^2}$$

__Dém :__ On peut se ramener au cas où x est de norme 1.

A est symétrique réelle donc il existe une b.o.n. $(e_1, .., e_n)$ de vecteurs propres de A. On note $x = \sum x_i e_i$. On a

$$\langle Ax, x\rangle \langle A^{-1}x, x\rangle = \sum_{i=1}^n \lambda_i x_i^2 \sum_{j=1}^n \frac{1}{\lambda_j} x_j^2 = \frac{\lambda_1}{\lambda_n} \sum_{i=1}^n \sum_{j=1}^n \frac{\lambda_i}{\lambda_n} \frac{\lambda_1}{\lambda_j} x_i^2 x_j^2.$$

$$\leq \frac{1}{4}\frac{\lambda_1}{\lambda_n} \sum_{i=1}^n \sum_{j=1}^n \left(\frac{\lambda_i}{\lambda_n} + \frac{\lambda_1}{\lambda_i}\right)\left(\frac{\lambda_j}{\lambda_n} + \frac{\lambda_1}{\lambda_j}\right) x_i^2 x_j^2$$

car $\left(\frac{\lambda_i}{\lambda_n} + \frac{\lambda_1}{\lambda_i}\right)\left(\frac{\lambda_j}{\lambda_n} + \frac{\lambda_1}{\lambda_j}\right) = \frac{\lambda_i \lambda_j}{\lambda_n \lambda_n} + \frac{\lambda_i \lambda_1}{\lambda_n \lambda_j} + \frac{\lambda_1 \lambda_j}{\lambda_i \lambda_n} + \frac{\lambda_1 \lambda_1}{\lambda_i \lambda_j}$

$$\geq \frac{\lambda_i \lambda_1}{\lambda_n \lambda_j} \times 4.$$

$$\leq \frac{1}{4}\frac{\lambda_1}{\lambda_n}\left(\sum_{i=1}^n \left(\frac{\lambda_i}{\lambda_n} + \frac{\lambda_1}{\lambda_i}\right) x_i^2\right)^2$$

Or $\frac{\lambda_i}{\lambda_n} + \frac{\lambda_1}{\lambda_i} = \frac{\lambda_i^2 + \lambda_1 \lambda_n}{\lambda_n \lambda_i} \leq \frac{\lambda_n^2 + \lambda_1 \lambda_n}{\lambda_n \lambda_i} = \frac{\lambda_n + \lambda_1}{\lambda_i} \leq \frac{\lambda_n + \lambda_1}{\lambda_1} = 1 + \frac{\lambda_n}{\lambda_1}$

d'où $\langle Ax, x\rangle \langle A^{-1}x, x\rangle \leq \frac{\lambda_1}{4\lambda_n}\left(\sum_{i=1}^n \left(1 + \frac{\lambda_n}{\lambda_1}\right) x_i^2\right)^2 = \frac{\lambda_1}{4\lambda_n}\left(1 + \frac{\lambda_n}{\lambda_1}\right)^2 \underbrace{\left(\sum x_i^2\right)^2}_{=1 \text{ car } \|x\| = 1}$

$$\leq \frac{\lambda_1}{4\lambda_n}\left(1 + \frac{\lambda_n^2}{\lambda_1^2} + 2\frac{\lambda_n}{\lambda_1}\right) = \frac{1}{4}\left(\frac{\lambda_1}{\lambda_n} + \frac{\lambda_n}{\lambda_1} + 2\right) = \frac{1}{4}\left(\sqrt{\frac{\lambda_1}{\lambda_n}} + \sqrt{\frac{\lambda_n}{\lambda_1}}\right)^2$$

Il suffit maintenant d'inverser l'inégalité.

$$\frac{c(A)}{(c(A)+1)^2} = \frac{\lambda_1}{\lambda_n\left(\frac{\lambda_1}{\lambda_n} + 1\right)^2} = \frac{1}{\frac{\lambda_n}{\lambda_1}\left(\frac{\lambda_1^2}{\lambda_n^2} + 2\frac{\lambda_1}{\lambda_n} + 1\right)} = \frac{4}{\left(\frac{\lambda_1}{\lambda_n} + 2 + \frac{\lambda_n}{\lambda_1}\right)} = \frac{1}{\left(\sqrt{\frac{\lambda_1}{\lambda_n}} + \sqrt{\frac{\lambda_n}{\lambda_1}}\right)^2}$$

✓ **A l'oral**, on ne fait pas Kantorovitch. 14'36 allure normale (énoncer Kanto avec le conditionnement !)
✓ Lorsque $n = 2$ x_k converge en zigzags vers $\bar{x}$.
✓ Plus $c(A)$ est proche de 1, plus la méthode converge rapidement. A l'inverse lorsque $c(A)$ est loin de 1, la méthode est très lente.
Le cas limite serait celui où $c(A) = 1$ (donc A homothétie), ce qui suppose $2(f(x) - \bar{f}) = \lambda \|x - \bar{x}\|^2$ auquel cas on atteint $\bar{x}$ dès la première itération. En effet, $d_0 = -Ax_0 - b = -\lambda x_0 + \lambda \bar{x}$, donc $t_0 = \frac{1}{\lambda}$ et on obtient bien $x_1 = \bar{x}$ (ou sinon on utilise la dernière majoration...).

chercher l'algorithme avec dessin + explicat° sur Adrien Laurent