Algorithme du gradient à pas optimal

Mohamed NASSIRI

Référence :

Analyse numérique des équations aux dérivées partielles, Laurent Di Menza, p.208 → 212
Exercices de mathématiques : Oraux X-ENS, XXX, Serge Francinou, Hervé Gianella et Serge Nicolas, p.208 → 212

Recasage :

●●●●● 219 : Extremums : existence, caractérisation, recherche. Exemples et applications.
●●●●○ 223 : Suites numériques. Convergence, valeurs d'adhérence. Exemples et applications.
●●●●○ 226 : Suites vectorielles et réelles définies par une relation de récurrence $u_{n+1} = f(u_n)$. Exemples. Applications à la réesolution approchée d'équations.
●●●●○ 229 : Fonctions monotones. Fonctions convexes. Exemples et applications.
●●●●○ 253 : Utilisation de la notion de convexité en analyse.
●●●○○ 203 : Utilisation de la notion de compacité.

Résumé :

Les méthodes de gradient sont des méthodes itératives où l'on a remplacé un système linéaire par un problème de minimisation. L'algorithme du gradient à pas optimal consiste à déterminer un pas optimal minimisant une certaine fonction.

Prérequis :

Méthodes itératives - Convexité - Matrices symétriques définies positives -

---

Théorème : (i) Soient A une matrice symétrique définie positive, $b \in \mathbb{R}^n$ et $c \in \mathbb{R}$. Alors les problèmes

$$\text{Trouver } x \in \mathbb{R}^n \text{ tel que } Ax = b \qquad (\dagger)$$

et

$$\text{Trouver } x \in \mathbb{R}^n \text{ tel que } f(x) = \min_{y \in \mathbb{R}^n} f(y) \text{ avec } f(y) = \frac{1}{2}\langle Ay, y\rangle - \langle b, y\rangle + c, \quad y \in \mathbb{R}^n \qquad (\dagger\dagger)$$

sont équivalents.
(ii) Pour une matrice A symétrique définie positive, l'algorithme itératif

$$\begin{cases} x^0 \in \mathbb{R}^n, \ r^0 = b - Ax^0 \\ x^{k+1} = x^k + \alpha_k r^k \\ r^{k+1} = r^k - \alpha_k A r^k \\ \alpha_k = \frac{||r_k||^2}{\langle Ar^k, r^k\rangle} \end{cases}$$

appelé *méthode du gradient à pas optimal*, converge pour tout matrice symétrique définie positive.

Démonstration.
Etape 1 - Equivalence des deux problèmes :

$\Rightarrow$: Soit $x \in \mathbb{R}^n$ solution de (†). Pour tout $h \in \mathbb{R}^n$, on a

$$\begin{aligned} f(x+h) &= \frac{1}{2}\langle A(x+h), x+h\rangle - \langle b, x+h\rangle + c \\ &= \frac{1}{2}\langle Ax, x\rangle + \frac{1}{2}\langle Ax, h\rangle + \frac{1}{2}\langle Ah, x\rangle + \frac{1}{2}\langle Ah, h\rangle - \langle b, x\rangle - \langle b, h\rangle + c \qquad \text{or } A = {}^tA \\ &\overset{\blacklozenge}{=} f(x) + \frac{1}{2}\langle Ax, h\rangle + \frac{1}{2}\langle {}^tAh, x\rangle - \langle b, h\rangle + \frac{1}{2}\langle Ah, h\rangle \qquad \text{or } \langle {}^tAh, x\rangle = \langle Ax, h\rangle \\ &= f(x) + \langle \underbrace{Ax-b}_{=0\ (\dagger)}, h\rangle + \frac{1}{2}\underbrace{\langle Ah, h\rangle}_{>0 \text{ si } h\neq 0} > f(x) \end{aligned}$$

Donc x est bien un minimum pour la fonction f.

$\Leftarrow$: Soit $x \in \mathbb{R}^n$ solution de (††). Alors la différentielle de la fonction f est nulle en x. Donc pour tout $h \in \mathbb{R}^n$, on a

$$0 = Df(x)(h) \overset{\blacklozenge}{=} \frac{1}{2}\langle Ax, h\rangle + \frac{1}{2}\langle {}^tAh, x\rangle - \langle b, h\rangle = \langle Ax - b, h\rangle$$

Comme cette égalité est valable pour tout $h \in \mathbb{R}^n$, on en déduit que $Ax - b = 0$. Donc x est bien solution de (†).

Etape 2 -Une relation de récurrence pour la suite $(\langle A^{-1}r^k, r^k\rangle)_{k\in\mathbb{N}}$:

On rappelle que la valeur de α_k est telle que $\langle r^k, r^{k+1}\rangle = 0$. En effet, x^{k+1} est calculé à partir de x^k de telle façon que $f(x^{k+1})$ soit minimal (*i.e.*) on souhaite donc minimiser la fonction, de la variable α,

$$g(\alpha) := f(x^{k+1}) = f(x^k + \alpha r^k)$$

On obtient donc

$$\alpha_k = \frac{||r^k||^2}{\langle Ar^k, r^k\rangle}$$

Par suite, on a bien

$$\langle r^k, r^{k+1}\rangle = \langle r^k, r^k - \alpha_k A r^k\rangle = \underbrace{\langle r^k, r^k\rangle}_{=||r^k||^2} - \alpha_k\langle r^k, Ar^k\rangle = 0$$

Par conséquent, on a

$$\begin{aligned} \langle A^{-1}r^{k+1}, r^{k+1}\rangle &= \langle A^{-1}(r^k - \alpha_k A r^k), r^{k+1}\rangle \\ &= \langle A^{-1}r^k, r^{k+1}\rangle - \alpha_k \underbrace{\langle r^k, r^{k+1}\rangle}_{=0} \\ &= \langle A^{-1}r^k, r^k - \alpha_k A r^k\rangle \\ &= \langle A^{-1}r^k, r^k\rangle - \alpha_k\langle A^{-1}r^k, Ar^k\rangle \\ &= \langle A^{-1}r^k, r^k\rangle - \alpha_k||r^k||^2 \end{aligned}$$

Ainsi, on en déduit l'appétissante relation de récurrence :

$$\frac{\langle A^{-1}r^{k+1}, r^{k+1}\rangle}{\langle A^{-1}r^k, r^k\rangle} = 1 - \frac{||r^k||^4}{\langle Ar^k, r^k\rangle\langle A^{-1}r^k, r^k\rangle} \qquad (\sharp)$$

Lemme : Inégalité de Kantorovitch
Soient A une matrice symétrique définie positive, de valeurs propres $0 < \lambda_1 \leq \lambda_2 \leq ... \leq \lambda_n$. Alors, pour $x \in \mathbb{R}^n$, on a l'inégalité

$$\langle Ax, x\rangle \langle A^{-1}x, x\rangle \leq \frac{(\lambda_1 + \lambda_n)^2}{4\lambda_1\lambda_n} ||x||^4$$

Démonstration :
En notant $(x_1, ..., x_n)$ les coordonées de x dans une base orthonormale de vecteurs propres de A, et en se rappelant que $\sqrt{ab} \leq \frac{1}{2}(a+b)$ pour $a, b \geq 0$, on a

$$\begin{aligned}
\sqrt{\langle Ax, x\rangle \langle A^{-1}x, x\rangle} &= \sqrt{\left(\sum_{i=1}^n \lambda_i x_i^2\right)\left(\sum_{i=1}^n \frac{1}{\lambda_i} x_i^2\right)} \\
&= \sqrt{\frac{\lambda_1}{\lambda_n}\left(\sum_{i=1}^n \frac{\lambda_i}{\lambda_1} x_i^2\right)\left(\sum_{i=1}^n \frac{\lambda_n}{\lambda_i} x_i^2\right)} \\
&= \frac{1}{2}\sqrt{\frac{\lambda_1}{\lambda_n}}\left(\left(\sum_{i=1}^n \frac{\lambda_i}{\lambda_1} x_i^2\right) + \left(\sum_{i=1}^n \frac{\lambda_n}{\lambda_i} x_i^2\right)\right) \\
&= \frac{1}{2}\sqrt{\frac{\lambda_1}{\lambda_n}}\left(\sum_{i=1}^n \left(\frac{\lambda_i}{\lambda_1} + \frac{\lambda_n}{\lambda_i}\right) x_i^2\right)
\end{aligned}$$

Pour aller plus loin, on va tenter de maximiser la fonction $\varphi : x \to \frac{x}{\lambda_1} + \frac{\lambda_n}{x}$ sur l'intervalle $[\lambda_1, \lambda_n]$. Cette fonction est décroissante sur $[\lambda_1, \sqrt{\lambda_1\lambda_n}]$ et croissante sur $[\sqrt{\lambda_1\lambda_n}, \lambda_n]$. Elle admet donc un maximum en λ_1 ou en λ_n, et on a $\varphi(\lambda_1) = \varphi(\lambda_n) = 1 + \frac{\lambda_n}{\lambda_1}$. Ainsi, on peut majorer

$$\begin{aligned}
\sqrt{\langle Ax, x\rangle \langle A^{-1}x, x\rangle} &= \frac{1}{2}\sqrt{\frac{\lambda_1}{\lambda_n}}\left(\sum_{i=1}^n \left(1 + \frac{\lambda_n}{\lambda_1}\right) x_i^2\right) \\
&= \frac{1}{2}\sqrt{\frac{\lambda_1}{\lambda_n}}\left(1 + \frac{\lambda_n}{\lambda_1}\right)\sum_{i=1}^n x_i^2 \\
&= \frac{1}{2}\left(\sqrt{\frac{\lambda_1}{\lambda_n}} + \sqrt{\frac{\lambda_n}{\lambda_1}}\right) ||x||^2
\end{aligned}$$

En élevant au carré, on obtient l'inégalité voulue.

$\square$

Etape 3 - On termine ... :

Grâce à l'inégalité de Kantorovitch, on a donc, à partir de $(\sharp)$:

$$\begin{aligned}
\frac{\langle A^{-1}r^{k+1}, r^{k+1}\rangle}{\langle A^{-1}r^k, r^k\rangle} = 1 - \frac{||r^k||^4}{\langle Ar^k, r^k\rangle \langle A^{-1}r^k, r^k\rangle} &\leq 1 - ||r^k||^4 \frac{4\lambda_1\lambda_n}{(\lambda_1+\lambda_n)^2}\frac{1}{||r^k||^4} \\
&= \left(\frac{\lambda_n - \lambda_1}{\lambda_n + \lambda_1}\right)^2
\end{aligned}$$

Par récurrence, on obtient donc

$$\langle A^{-1}r^k, r^k\rangle \leq \left(\frac{\lambda_n - \lambda_1}{\lambda_n + \lambda_1}\right)^{2k} \langle A^{-1}r^0, r^0\rangle := \xi^{2k}\langle A^{-1}r^0, r^0\rangle$$

avec $\xi = \frac{\lambda_n - \lambda_1}{\lambda_n + \lambda_1} < 1$. On en déduit donc que

$$\langle A^{-1}r^k, r^k\rangle \underset{k\to+\infty}{\longrightarrow} 0$$

Or, puisque $\langle A^{-1}r^k, r^k\rangle \geq \frac{1}{\lambda_n}||r^k||^2$, on a donc

$$||r^k|| \underset{k\to+\infty}{\longrightarrow} 0$$

□

Remarques :

- **Explications de la méthode :**
- **Illustration :**
- **Importance de la symétrie :**
- **Propriétés de f :** strictement cvx, coercive avec les 1/2 mais aussi les A-1
- **Différentielle de f :** $f(y) = \frac{1}{2}\langle Ay, y\rangle - \langle b, y\rangle + c, \quad y \in \mathbb{R}^n$

$$f(x+h) = f(x) + \frac{1}{2}\langle Ax, h\rangle + \frac{1}{2}\langle {}^tAh, x\rangle - \langle b, h\rangle + \frac{1}{2}\langle Ah, h\rangle$$

- **Minimum de g :** Dans la démonstration, on souhaitait minimiser la fonction, de la variable α,

$$g(\alpha) := f(x^{k+1}) = f(x^k + \alpha r^k)$$

et on a affirmé que $g'(\alpha) = \langle A(x^k + \alpha r^k) - b, r^k\rangle$. En effet,

$$\begin{aligned} f(x^k + \alpha r^k) &= \frac{1}{2}\langle A(x^k + \alpha r^k), x^k + \alpha r^k\rangle - \langle b, x^k + \alpha r^k\rangle + c \\ &= \frac{1}{2}\langle Ax^k, x^k\rangle + \frac{\alpha}{2}\langle Ax^k, r^k\rangle + \frac{\alpha}{2}\langle Ar^k, x^k\rangle + \frac{\alpha^2}{2}\langle Ar^k, r^k\rangle - \langle b, x^k\rangle - \alpha\langle b, r^k\rangle + c \\ &= f(x^k) + \alpha\langle \underbrace{Ax^k - b}_{=-r^k}, r^k\rangle + \frac{\alpha^2}{2}\langle Ar^k, r^k\rangle \\ &= f(x^k) - \alpha||r^k||^2 + \frac{\alpha^2}{2}\langle Ar^k, r^k\rangle \end{aligned}$$

Il s'agit d'un bon vieux polynôme de degré 2 en α donc le coefficient dominant est positif puisque A est symétrique définie positive. Le minimum est donc atteint en

$$\alpha = \alpha_k = \frac{||r^k||^2}{\langle Ar^k, r^k\rangle}$$

- **Minoration pour Kantorovitch :** En notant $(x_1, ..., x_n)$ les coordonées de x dans une base orthonormale de vecteurs propres de A et en appliquant l'inégalité de Cauchy-Schwarz, on a

$$\langle Ax, x\rangle\langle A^{-1}x, x\rangle = \left(\sum_{i=1}^{n} \lambda_i x_i^2\right)\left(\sum_{i=1}^{n} \frac{1}{\lambda_i} x_i^2\right) \geq \left(\sum_{i=1}^{n} \frac{\sqrt{\lambda_i}}{\sqrt{\lambda_i}} x_i^2\right)^2 = ||x||^4$$

Théorème :

Démonstration :

$\square$