Deux méthodes de gradient

Leçons : 158, *162*, 219, *226*, 233 (gradient conjugué)

On considère $A \in \mathcal{S}_n^{++}(\mathbb{R})$.

Proposition 1

La résolution de $Ax = b$ équivaut à trouver le point qui minimise la fonctionnelle :

$$\Phi(y) = \frac{1}{2} y^T A y - y^T b.$$

Démonstration. Il est facile de voir que

$$\nabla\Phi(y) = \frac{1}{2}(A^T + A)y - b = Ay - b. \tag{1}$$

Et si x est solution du système linéaire, alors $\Phi(y) = \Phi(x + (y-x)) = \Phi(x) + \frac{1}{2}(y-x)^T A(y-x)$ *i.e* $\frac{1}{2}\|y-x\|_A^2 = \Phi(y) - \Phi(x)$, où $\|z\|_A^2 = z^T A z$ est la norme associée à A que l'on utilisera toujours par la suite.

□

Définition 2

Une méthode de gradient consiste à partir d'un point $x_0 \in \mathbb{R}^n$ et à construire la suite

$$x_{k+1} = x_k + \alpha_k d_k \tag{2}$$

où $d_k \in \mathbb{R}^n$ est une direction à choisir et $\alpha_k \in \mathbb{R}$.

Une idée naturelle est de choisir α_k de sorte à optimiser $\Phi(x_{k+1})$ dans la direction d_k, c'est à dire tel que $\frac{d}{d\alpha_k}\Phi(x_k + \alpha_k d_k) = -d_k^T r_k + \alpha_k d_k^T A d_k = 0$, où $-r_k := \nabla\Phi(x_k) = Ax_k - b$. On trouve :

$$\alpha_k = \frac{\langle d_k, r_k\rangle}{\|d_k\|_A^2} \tag{3}$$

(c'est bien défini lorsque $d_k \neq 0$ car $A \in \mathcal{S}_n^{++}(\mathbb{R})$).

Méthode de gradient conjugué

Remarquons que pour tout $k \in \mathbb{N}$:

$$r_{k+1} = r_k - \alpha_k A d_k \tag{4}$$

et α_k est choisi de sorte à ce que

$$\langle r_{k+1}, d_k\rangle = 0. \tag{5}$$

Idée. Construire des directions (d_k) deux à deux A-orthogonales ; ainsi, r_{k+1} sera orthogonal à $\mathrm{Vect}(d_0, \ldots, d_k)$.

Notations. Pour $x, y \in \mathbb{R}^n$, on note $x \perp y$ lorsque x et y sont orthogonaux pour le produit scalaire euclidien et $x \perp_A y$ lorsque x et y sont orthogonaux pour le produit scalaire donné par A. On étend naturellement cette notation à des sous-espaces de $\mathbb{R}^n$.

On pose $d_0 = r_0$ et pour $k \in \mathbb{N}$, on construit d_{k+1} comme l'orthogonalisé de Gram-Schmidt pour le produit scalaire donné par A de r_{k+1} relativement à $\text{Vect}(d_k)$:

$$d_{k+1} = r_{k+1} - \beta_k d_k \tag{6}$$

où

$$\beta_k = \frac{\langle r_{k+1}, Ad_k \rangle}{\|d_k\|_A^2} \quad \text{si } d_k \neq 0, \quad \beta_k = 0 \text{ sinon.} \tag{7}$$

Remarquons que si $d_k = 0$ alors r_k et d_{k-1} sont colinéaires et comme ils sont aussi orthogonaux par (5), $r_k = 0$.

Lemme 3

Avec le choix (7)*, les directions* (6) *vérifient pour tout $k \in \mathbb{N}$ la propriété suivante : si $r_0, \dots, r_k$ ne sont pas nuls alors,*

1 $\text{Vect}(r_0, \dots, r_k) = \text{Vect}(d_0, \dots, d_k)$

2 $r_{k+1} \perp \text{Vect}(d_0, \dots, d_k)$

3 $d_{k+1} \perp_A \text{Vect}(d_0, \dots, d_k)$

Démonstration. On procède par récurrence sur $k \in \mathbb{N}$. Lorsque $k = 0$, 1, 2 et 3 sont vrais grâce aux relations $r_0 = d_0$, (5) et (6) et bien sûr $r_0 \neq 0$ sinon il n'y a rien à faire. Supposons donc le résultat vrai au rang $k-1$, $k \in \mathbb{N}^*$.

1 Par (6), on a : $d_k = r_k - \beta_{k-1} d_{k-1}$.

2 Par (5), on a déjà $r_{k+1} \perp d_k$ et si $j \in \{0, \dots, k-1\}$, la relation (4) couplée à l'hypothèse de récurrence 2 et 3 donne $r_{k+1} \perp d_j$.

3 Par (6), on a déjà $d_{k+1} \perp_A d_k$ (c'est la définition) et si $j \in \{0, \dots, k-1\}$, la relation (6) couplée à l'hypothèse de récurrence 3 donne $\langle d_{k+1}, Ad_j \rangle = \langle r_{k+1}, Ad_j \rangle$.

Montrons que $Ad_j \in \text{Vect}(r_0, \dots, r_k)$, ce qui conclura grâce aux relations 1 et 2 que l'on vient de prouver. Grâce à la relation (4) avec $k = j$, il suffit de montrer que $\alpha_j \neq 0$. Or, $\alpha_j = 0 \overset{(3)}{\iff} \langle r_j, d_j \rangle = 0 \overset{(6)}{\iff} r_j = 0$ puisque $\langle r_j, r_j \rangle = \langle d_j, r_j \rangle + \beta_{j-1} \langle d_{j-1}, r_j \rangle = \langle d_j, r_j \rangle$ selon 2. Donc comme on a supposé $r_j \neq 0$, on a $\alpha_j \neq 0$.

□

Théorème 4

La méthode de gradient associée aux directions (6) *avec le choix* (7) *converge vers la solution x du problème $Ax = b$ en au plus n itérations.*

Démonstration. Les conditions 1 et 2 du lemme précédent assurent que tant que $r_l \neq 0$, la famille $(r_0, \dots, r_l)$ est une famille orthogonale donc libre. On est en dimension n donc nécessairement $l + 1 \leqslant n$ et si $r_l = 0$, x_l est solution du système. □

Méthode de gradient à pas optimal

On choisit pour direction la « plus grande pente » , c'est à dire $d_k = -\nabla\Phi(x_k) = -Ax_k + b = r_k$.

Dans ce cas, $d_k \neq 0$ tant que la solution n'est pas atteinte. La convergence découle essentiellement de l'inégalité de Kantorovich :

Lemme 5 (*Inégalité de Kantorovich*)

En notant $0 < \lambda_1 \leq \ldots \leq \lambda_n$ les valeurs propres de A, on a pour tout $y \in \mathbb{R}^n$,

$$\frac{\|y\|^4}{\|y\|_A^2 \|y\|_{A^{-1}}^2} \geq \frac{4\lambda_n \lambda_1}{(\lambda_n + \lambda_1)^2}.$$

Démonstration. On va montrer l'inégalité équivalente :

$$\forall y \in \mathbb{R}^n, \ \|y\|^4 \leq \frac{1}{4}\left(\sqrt{\frac{\lambda_n}{\lambda_1}} + \sqrt{\frac{\lambda_1}{\lambda_n}}\right)^2.$$

On peut même supposer que $\|y\| = 1$ et commencer par remarquer :

$$1 = \|y\|^2 = \langle y, AA^{-1}y\rangle \leq \|y\|_A \|A^{-1}y\|_A = \|y\|_A \|y\|_{A^{-1}}$$

Et dans une base orthonormale de vecteurs propres :

$$\begin{aligned}
\|y\|_A \|y\|_{A^{-1}} &= \sqrt{\left(\sum_{i=1}^n \lambda_i y_i^2\right)\left(\sum_{i=1}^n \frac{1}{\lambda_i} y_i^2\right)} = \sqrt{\frac{\lambda_1}{\lambda_n}\left(\sum_{i=1}^n \frac{\lambda_i}{\lambda_1} y_i^2\right)\left(\sum_{i=1}^n \frac{\lambda_n}{\lambda_i} y_i^2\right)} \\
&\leq \frac{1}{2}\sqrt{\frac{\lambda_1}{\lambda_n}}\left(\left(\sum_{i=1}^n \frac{\lambda_i}{\lambda_1} y_i^2\right) + \left(\sum_{i=1}^n \frac{\lambda_n}{\lambda_i} y_i^2\right)\right) \\
&\leq \frac{1}{2}\sqrt{\frac{\lambda_1}{\lambda_n}}\left(\sum_{i=1}^n \left(\frac{\lambda_i}{\lambda_1} + \frac{\lambda_n}{\lambda_i}\right) y_i^2\right)
\end{aligned}$$

La fonction $x \mapsto \dfrac{x}{\lambda_1} + \dfrac{\lambda_n}{x}$ admet un maximum en λ_1 ou en λ_n et il vaut dans les deux cas : $1 + \dfrac{\lambda_n}{\lambda_1}$. Ainsi,

$$\|y\|_A \|y\|_{A^{-1}} \leq \frac{1}{2}\sqrt{\frac{\lambda_1}{\lambda_n}}\left(\sum_{i=1}^n \left(1 + \frac{\lambda_n}{\lambda_1}\right) y_i^2\right) \leq \frac{1}{2}\left(\sqrt{\frac{\lambda_n}{\lambda_1}} + \sqrt{\frac{\lambda_1}{\lambda_n}}\right),$$

et le résultat suit en élevant au carré. □

Et sachant que $\mathrm{cond}(A) = \lambda_n/\lambda_1$, on obtient le résultat suivant :

Théorème 6

Avec les choix précédents et $d_k = r_k$, la suite (2) converge vers x avec :

$$\|x_{k+1} - x\|_A \leq \frac{\lambda_n - \lambda_1}{\lambda_n + \lambda_1} \|x_k - x\|_A.$$

Plus précisément,

$$\|x_k - x\| \leq \sqrt{\mathrm{cond}(A)}\left(\frac{\mathrm{cond}(A) - 1}{\mathrm{cond}(A) + 1}\right)^k \|x_0 - x\|.$$

Démonstration. La première inégalité découle directement de l'inégalité de Kantorovich. Pour la seconde, on remarque que pour tout $y \in \mathbb{R}^n$, $\lambda_1 \|y\|^2 \leq \|y\|_A^2 \leq \lambda_n \|y\|^2$. □

Avec la dernière inégalité, on voit que la convergence peut être lente lorsque la matrice est mal conditionnée.

Référence : Alfio QUARTERONI, Ricardo SACCO et Fausto SALERI (2007). *Méthodes numériques : Algorithmes, analyse et applications*. Springer, pp. 138-145.

Merci à Antoine Diez pour ce développement.