Méthodes de descente pour la résolution de systèmes linéaires

5 novembre 2000

1 Principe des méthodes de descente

1.1 Minimisation de fonctions quadratiques

Le but est de construire des méthodes itératives pour résoudre un système linéaire

$$Ax = b, \tag{1}$$

en lui associant un problème de minimisation équivalent et en construisant une suite minimisante : on va travailler avec des matrices réelles symétriques et définies positives, mais on pourrait tout généraliser au cas de matrices hermitiennes définies positives. Considérons la forme quadratique F sur $\mathbb{R}^d$ définie par

$$F(x) = \frac{1}{2}x^T Ax - x^T b. \tag{2}$$

La fonction F est continue et $\lim_{||x||\to\infty} F(x) = +\infty$. La fonction admet donc un minimum dans $\mathbb{R}^d$. La fonction F est de plus différentiable, et son gradient vaut

$$\forall y \in R^n, \quad DF(y) = Ay - b. \tag{3}$$

Exercice Démontrer l'assertion précédente.

Comme la solution de (1) est unique car A est inversible, le gradient de F ne s'annule qu'en un seul point qui réalise le minimum de F.

Exercice Montrer que F est strictement convexe, c'est à dire que

$$F(\alpha x + (1-\alpha)y) - \alpha F(x) - (1-\alpha)F(y) \geq \frac{1}{2}\alpha(1-\alpha)\lambda_{min}(A)$$

où $\lambda_{min}(A)$ est la plus petite valeur propre de A.

On voit donc que le problème (1) est équivalent à la minimisation de F. L'idée va donc de construire des suites minimisantes de F pour approcher la solution de (1).

1.2 Méthodes de descente

Une méthode de descente consiste à construire une suite minimisante sous la forme

$$x_{n+1} = x_n + \alpha_n p_n, \tag{4}$$

où $p_n \in \mathbb{R}^d$, $p_n \neq 0$ et où le scalaire α_n est choisi pour que $F(x_{n+1}) < F(x_n)$. La convergence de la méthode dépend bien sûr des choix des p_n et α_n.

Définition 1 *L'erreur à l'étape n est le vecteur*

$$e_n = x - x_n. \tag{5}$$

On appelle résidu à l'étape n le vecteur

$$r_n = b - Ax_n = Ae_n. \tag{6}$$

Remarque 1 *Le résidu à la nième étape vérifie $r_n = -DF(x_n)$.*

1.2.1 Choix optimal de α_n pour p_n fixé

Supposons choisie la direction p_n : on peut choisir α_n de manière à minimiser la fonction ϕ de $\mathbb{R}^+$ dans $\mathbb{R}$ donnée par

$$\phi(t) = F(x_n + tp_n).$$

Cette fonction à un minimum unique pour

$$\alpha_{opt} = \frac{r_n^T p_n}{r_n^T A p_n}, \tag{7}$$

et on a

$$x_{n+1} = x_n + \frac{r_n^T p_n}{r_n^T A p_n} p_n$$

Proposition 1 *Pour tout p_n, et si on choisit $\alpha_n = \alpha_{opt}$, on a*

$$r_{n+1}^T p_n = 0. \tag{8}$$

2 Méthodes de gradient

2.1 Principe des méthodes de gradient

L'idée des méthodes de gradient va être de choisir comme direction de descente le gradient de F en x_n, ou (voir Remarque 1), de manière équivalente, le résidu r_n :

$$\begin{aligned} x_{n+1} &= x_n - \rho_n DF(x_n) \\ &= x_n + \rho_n r_n. \end{aligned} \tag{9}$$

Pour tout $x \in \mathbb{R}^n$, il existe un nombre réel $\rho_{\max}(x) > 0$, tel que

$$\rho \in]0, \quad \rho_{\max}(x)[\Leftrightarrow F(x - \rho DF(x)) < F(x).$$

Exercice Démontrer cette assertion.

Le pas ρ_n doit donc être un réel positif choisi tel que $F(x_{n+1}) < F(x_n)$.

Remarque 2 *On peut généraliser ces méthodes à des fonctions strictement convexes et différentiables.*

2.2 Interprétation géométrique en dimension deux

Soit A une matrice d'ordre deux symétrique et définie positive. On sait que A est diagonalisable dans une base orthonormale. Quitte à changer les coordonnées on peut supposer que A est diagonale et que $b = 0$:

$$A = \begin{pmatrix} \lambda_1 & 0 \\ 0 & \lambda_2 \end{pmatrix}. \tag{10}$$

La fonction F s'écrit alors $F(x) = \lambda_1 x_1^2 + \lambda_2 x_2^2$, et les lignes de niveaux de F : $F(x) = r^2$ sont les ellipses concentriques

$$\lambda_1 x_1^2 + \lambda_2 x_2^2 = r^2. \tag{11}$$

Le gradient de F au point x est orthogonal à l'ellipse d'équation (11) passant par x. La figure suivante donne un exemple des premiers itérés d'une méthode de gradient :

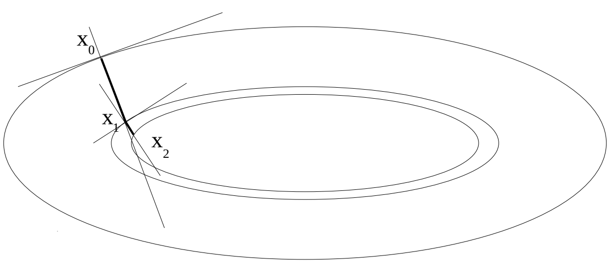

FIG. 1 – *deux itérés d'une méthode de gradient*

2.3 Méthodes du gradient à pas fixe

Si A est défini positif, on peut obtenir une méthode convergente en fixant ρ_n à une valeur bien choisie: la méthode du gradient à pas fixe consiste à constuire la suite récurrente

$$x_{n+1} = x_n - \alpha(Ax_n - b).$$

L'erreur vérifie

$$e_{n+1} = (I - \alpha A)e_n = (I - \alpha A)^{n+1} e_0,$$

ce qui montre que la méthode du gradient à pas fixe converge si et seulement si

$$\rho(I - \alpha A) < 1$$

où $\tau(\alpha) = \rho(I - \alpha A)$ est le rayon spectral de $I - \alpha A$. On appelle $\tau(\alpha)$ le taux de convergence de la méthode. On a donc

Théorème 1 *Si A est symétrique définie positive, la méthode du gradient à pas fixe converge vers la solution x de (1) si et seulement si*

$$\alpha < \frac{2}{\lambda_{\max}(A)}. \tag{12}$$

De plus la valeur de α minimisant le taux de convergence $\tau(\alpha)$ est

$$\alpha_{opt} = \frac{2}{\lambda_{\min}(A) + \lambda_{\max}(A)}.$$

et le taux de convergence vaut alors

$$\tau_{opt} = \frac{cond_2(A) - 1}{cond_2(A) + 1}$$

Remarque 3 *On voit donc que la méthode du gradient à pas fixe converge d'autant plus lentement que le conditionnement de A est grand. En effet, avec le choix optimal pour ρ, il faut de l'ordre de*

$$\frac{|\log(\epsilon)|}{|\log(\frac{cond_2(A)-1}{cond_2(A)+1})|}$$

itérations pour réduire l'erreur d'un facteur ϵ; quand $cond_2(A) \gg 1$, le nombre d' itérations est donc de l'ordre de

$$|\log(\epsilon)|\frac{cond_2(A)}{2}.$$

2.4 Méthode du gradient à pas optimal

Comme dans § 1.2.1, on peut construire

$$x_{n+1} = x_n - \rho_n(Ax_n - b),$$

en choisissant ρ_n de manière à minimiser la fonction ϕ de $\mathbb{R}^+$ dans $\mathbb{R}$ donnée par

$$\phi(t) = F(x_n - t(Ax_n - b)).$$

Cette fonction à un minimum unique pour

$$\rho_{opt} = \frac{\|Ax_n - b\|_2^2}{\|Ax_n - b\|_A^2} = \frac{\|r_n\|_2^2}{\|r_n\|_A^2}, \tag{13}$$

où

$$\|y\|_A^2 = y^T Ay.$$

Il est important de noter que dans la méthode du gradient à pas optimal, d'après (8), les résidus successifs sont orthogonaux :

$$r_{n+1}^T r_n = 0. \tag{14}$$

Théorème 2 *Si A est symétrique et définie positive, la méthode du gradient à pas optimal converge vers la solution x de (1).*

Démonstration La suite des $J(x_n)$ est décroissante par construction, et bornée inférieurement par $J(x)$, donc converge. On en déduit que

$$J(x_{n+1}) - J(x_n) \to 0,$$

ce qui implique que

$$-\rho_n \|r_n\|_2^2 + \frac{1}{2}\rho_n^2 r_n^T A r_n \to 0,$$

et comme $\rho_n = \frac{\|r_n\|_2^2}{r_n^T A r_n}$, on trouve finalement que

$$\frac{\|r_n\|_2^4}{r_n^T A r_n} \to 0.$$

Comme A est définie positive, ceci implique que $r_n \to 0$. Toujours grâce au caractère défini positif de A, on en déduit que $e_n \to 0$.

Remarque 4 *Cette démonstration se généralise à la méthode du gradient à pas optimal appliqué à la minimisation d'une fonction F fortement convexe et de gradient Lipchitzien.*

Pour déterminer la vitesse de convergence de la méthode de gradient à pas optimal, on utilise l'inégalité de Kantorovitch

Lemme 1 *Soient d réels strictement positifs,*

$$0 < \ell_1 < \ldots < \ell_i < \ldots < \ell_d$$

et d réels positifs β_i tels que $\sum_1^d \beta_i = 1$. On note $\ell = \sum_1^d \beta_i \ell_i$. On a

$$\sum_1^d \frac{\beta_i}{\ell_i} \le \frac{\ell_1 + \ell_d - \ell}{\ell_1 \ell_d}, \tag{15}$$

et

$$\frac{1}{\ell \sum_1^d \frac{\beta_i}{\ell_i}} \ge \frac{4\ell_1 \ell_d}{(\ell_1 + \ell_d)^2}. \tag{16}$$

Démonstration Pour prouver (15), on doit montrer que

$$\sum_1^d \beta_i \left(\frac{1}{\ell_i} + \frac{\ell_i}{\ell_1 \ell_d}\right) \le \frac{\ell_1 + \ell_d}{\ell_1 \ell_d}$$

Pour cela, on voit que la fonction qui à $x \in [\ell_1, \ell_d]$ associe $\frac{1}{x} + \frac{x}{\ell_1 \ell_d}$ atteint son maximum en $x = \ell_1$ et en $x = \ell_d$ et le maximum vaut $\frac{1}{\ell_1} + \frac{1}{\ell_d}$. On conclut en utilisant le fait que $\sum_{i=1}^d \beta_i = 1$. Après, (16) s'obtient facilement en cherchant le minimum de $\ell \frac{\ell_1 + \ell_d - \ell}{\ell_1 \ell_d}$ sur l'intervalle $[\ell_1, \ell_d]$.

Proposition 2 *(Kantorovitch) Soit $A \in \mathbb{R}^{d\times d}$ symétrique définie positive dont les valeurs propres vérifient $0 < \lambda_{\min} = \lambda_1 < \ldots < \lambda_i < \ldots < \lambda_d = \lambda_{\max}$. On a*

$$\inf_{y \neq 0} \frac{\|y\|_2^4}{(y^T A y)\ (y^T A^{-1} y)} = \frac{4\lambda_{\min}\lambda_{\max}}{(\lambda_{\min} + \lambda_{\max})^2}.$$

Démonstration On note $\lambda_1 \le \lambda_2 \le \cdots \le \lambda_d$ les valeurs propres de A et $(v_i)_{1\le i\le d}$ une base orthonormale de vecteurs propres: $Av_i = \lambda_i v_i$. Pour $y \in \mathbb{R}^d$, on note $\beta_i = \frac{(y^T v_i)^2}{\|y\|_2^2}$:

$$\sum_{i=1}^d \beta_i = 1.$$

On a aussi

$$\frac{(Ay,y)}{\|y\|_2^2} = \sum_{i=1}^{d} \beta_i \lambda_i \quad \text{et} \quad \frac{(A^{-1}y,y)}{\|y\|_2^2} = \sum_{i=1}^{d} \frac{1}{\lambda_i} \beta_i.$$

et on applique le lemme précédent, et on obtient

$$\inf_{y \neq 0} \frac{\|y\|_2^4}{(y^T A y)\,(y^T A^{-1} y)} \geq \frac{4\lambda_{\min}\lambda_{\max}}{(\lambda_{\min} + \lambda_{\max})^2}.$$

Enfin, cet infimum est atteint par $y = v_1 + v_d$.

Théorème 3 *Pour la méthode du gradient à pas optimal, on a l'estimation*

$$\|e_n\|_A \leq \left(\frac{cond_2(A) - 1}{cond_2(A) + 1}\right)^n \|e_0\|_A$$

Démonstration On a $\rho_n = \frac{\|r_n\|_2^2}{r_n^T A r_n}$, $e_{n+1} = e_n - \frac{\|r_n\|_2^2}{r_n^T A r_n} r_n$, et $r_{n+1} = r_n - \frac{\|r_n\|_2^2}{r_n^T A r_n} A r_n$. Donc

$$\begin{aligned}
\|e_{n+1}\|_A^2 = e_{n+1}^T A e_{n+1} \quad &= e_{n+1}^T r_{n+1} \\
&= e_n^T r_n - \tfrac{\|r_n\|_2^4}{r_n^T A r_n} \\
&= (1 - \tfrac{\|r_n\|_2^4}{(r_n^T A r_n)(r_n^T A^{-1} r_n)}) e_n^T A e_n \\
&\leq \tfrac{(\lambda_{\max} - \lambda_{\min})^2}{(\lambda_{\max} + \lambda_{\min})^2} e_n^T A e_n \\
&= \left(\tfrac{cond_2(A) - 1}{cond_2(A) + 1}\right)^2 \|e_n\|_A^2
\end{aligned}$$

Interprétation géométrique en dimension deux On reprend la matrice A donnée par (10). D'après (14), on peut construire graphiquement la suite des itérés, car le point x_{n+1} est à la fois sur la droite de direction r_n passant par x_n, et sur l'ellipse d'équation (11) tangentée par cette droite. La figure suivante donne un exemple des premiers itérés d'une méthode de gradient à pas optimal : Dire que la matrice A est mal conditionnée, c'est dire que les lignes de niveaux

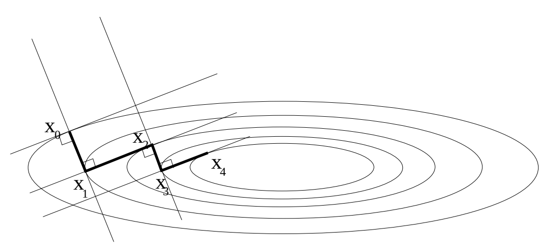

FIG. 2 – *méthode de gradient à pas optimal*

de F sont des ellipses très allongées, ou encore à fort rapport d'aspect. Dans ce cas, on voit que la suite des x_n se rapproche de sa limite x en zigzaguant beaucoup, et on comprend que la convergence est lente.