Université Claude Bernard - Lyon 1 Semestre de printemps 2019-2020

Analyse matricielle-L3

Feuille d'exercices n° 5

MÉTHODES DE DESCENTE.

Dans cette feuille, on notera $\| \cdot \|$ la norme euclidienne sur $\mathbf{R}^n$. On rappelle que pour tous vecteurs x et y, $y^T x$ est le produit scalaire entre x et y, que l'on peut aussi noter $\langle x, y \rangle$.

Exercice 1. *Méthode de Jacobi*

Soit $n \in \mathbf{N}^*$. Pour minimiser une fonction $f : \mathbf{R}^n \longrightarrow \mathbf{R}$, on appelle méthode de descente à pas fixe $\alpha \in \mathbf{R}_+^*$ la méthode itérative définie par la récurrence

$$\begin{cases} x^{(0)} \in \mathbf{R}^n, \\ x^{(k+1)} = x^{(k)} + \alpha w^{(k)} \text{ pour tout } k \in \mathbf{N}, \end{cases}$$

où $w^{(k)} \in \mathbf{R}^n$, appelé alors *direction de descente stricte*, satisfait pour tout k à ${w^{(k)}}^T \nabla f(x^{(k)}) < 0$ tant que $x^{(k)}$ ne réalise pas un minimum local de f.

Dans toute la suite, on considère le cas quadratique où pour tout $x \in \mathbf{R}^n$, $f(x) = x^T Ax/2 - x^T b$ pour une matrice symétrique définie positive $A \in \mathcal{M}_n(\mathbf{R})$ et un vecteur $b \in \mathbf{R}^n$ donnés.

1. Montrer que pour tout $x \in \mathbf{R}^n$, $\nabla f(x) = Ax - b$.

   Pour cela, développer pour tout x et tout h, $f(x+h)$. Quelle valeur de $\nabla f(x)$ donne le calcul si A n'est pas symétrique ?

2. On veut montrer que la méthode de Jacobi pour la résolution itérative du système $Ax = b$ est une telle méthode de descente. On utilise les notations de la feuille de TP 1 pour décomposer $A = D - (E + F)$ où D est diagonale, E triangulaire inférieure stricte et F triangulaire supérieure stricte.

   (a) Justifier que D est inversible.

   (b) Montrer que la méthode de Jacobi se réécrit sous la forme : pour tout $k \in \mathbf{N}$,

   $$x^{(k+1)} = x^{(k)} + w^{(k)}, \quad \text{où } w^{(k)} = D^{-1} r^{(k)} \text{ et } r^{(k)} = b - Ax^{(k)}.$$

   (c) En déduire que la méthode de Jacobi est une méthode de descente (avec $\alpha = 1$).

3. Pour améliorer la méthode de Jacobi, on va remplacer le pas fixe $\alpha = 1$ par un pas *optimal*, $\alpha^{(k)}$, dépendant de l'itération, défini comme réalisant le minimum sur $\mathbf{R}_+$ de $\alpha \longmapsto f(x^{(k)} + \alpha w^{(k)})$, pour le $w^{(k)}$ trouvé à la question précédente.

   (a) Soit $k \in \mathbf{N}$ tel que $w^{(k)} \neq 0$. On note $g_k : \mathbf{R}^+ \to \mathbf{R}$, $\alpha \longmapsto f(x^{(k)} + \alpha w^{(k)})$. Montrer que g_k admet un unique minique global, au point $\alpha^{(k)} = \dfrac{|{w^{(k)}}^T r^{(k)}|^2}{{w^{(k)}}^T A w^{(k)}}$.

   Dans la suite de l'exercice, on étudie la suite définie par

   $$\begin{cases} x^{(0)} \in \mathbf{R}^n, \\ x^{(k+1)} = x^{(k)} + \alpha^{(k)} w^{(k)} \text{ pour tout } k \in \mathbf{N}. \end{cases}$$

(b) Soit $k \in \mathbf{N}$ tel que $w^{(k)} \neq 0$.

i. Montrer que $|f(x^{(k+1)}) - f(x^{(k)})| = \dfrac{|w^{(k)^T} r^{(k)}|^2}{2 w^{(k)^T} A w^{(k)}}$.

ii. En déduire qu'il existe $C > 0$ tel que $|f(x^{(k+1)}) - f(x^{(k)})| \geq C \|r^{(k)}\|^2$.

(c) Montrer que la suite réelle $\big(f(x^{(k)})\big)_{k\in\mathbf{N}}$ converge. En déduire que $r^{(k)}$ tend vers $0 \in \mathbf{R}^n$ lorsque k tend vers l'infini.

(d) Montrer que la suite $(x^{(k)})_{k\in\mathbf{N}}$ converge vers la solution x du système $Ax = b$ quand k tend vers l'infini.

(e) Quel résultat de convergence vient-on de démontrer sur la méthode de Jacobi à pas optimal ? Comparer les conditions (conditions sur A) de convergence de la méthode de Jacobi et de la méthode de Jacobi à pas optimal. On pourra considérer le cas de la matrice

$$A = \begin{pmatrix} 1 & 3/4 & 3/4 \\ 3/4 & 1 & 3/4 \\ 3/4 & 3/4 & 1 \end{pmatrix}$$

et se souvenir de l'exercice 9 de la feuille 2.

Exercice 2. ($\star$) *Méthode du gradient conjugué*

On considère $n \in \mathbf{N}^*$, $A \in \mathcal{M}_n(\mathbf{R})$ symétrique définie positive, et $b \in \mathbf{R}^n$. La méthode du gradient conjugué pour approcher la solution du système $Ax = b$ est une méthode de descente (appliquée à la minimisation de la fonctionnelle quadratique $f : \mathbf{R}^n \to \mathbf{R}$, $x \longmapsto f(x) = x^T A x/2 - x^T b$) qui peut être vue comme une généralisation de la méthode du gradient à pas optimal, mais où la *direction de descente* est elle-même choisie pour minimiser le résidu (dans un espace de dimension 1 pour que cette minimisation soit facile), à chaque itération.

On se donne un vecteur initial $x^{(0)} \in \mathbf{R}^n$. La suite minimisante est définie par

$$\forall k \in \mathbf{N}, \quad \left\{ \begin{array}{l} p^{(k)} = r^{(k)} + \beta^{(k)} p^{(k-1)}, \text{ où } r^{(k)} = -\nabla f(x^{(k)}) = b - Ax^{(k)}, \\ x^{(k+1)} = x^{(k)} + \alpha^{(k)} p^{(k)}, \end{array} \right.$$

où les suites de réels $(\alpha^{(k)})_{k\in\mathbf{N}}$ et $(\beta^{(k)})_{k\in\mathbf{N}}$ sont calculées pour minimiser à chaque étape la quantité $f(x^{(k)} + \alpha p^{(k)}) = f(x^{(k)} + \alpha(r^{(k)} + \beta p^{(k-1)}))$, vue ici comme une fonction de α et β (à k fixé).

Par convention, on choisit $p^{(-1)} = 0 \in \mathbf{R}^n$.

Si $r^{(k)} = 0$, l'algorithme s'arrête (en effet dans ce cas, $x^{(k)}$ est solution).

Dans la suite, on note pour tout $k \in \mathbf{N}$, $j_k : \mathbf{R}^2 \to \mathbf{R}$, $(\alpha, \beta) \mapsto j_k(\alpha, \beta) = f\big(x^{(k)} + \alpha(r^{(k)} + \beta p^{(k-1)})\big)$.

A) Définition de l'algorithme.

1. Soit $k \in \mathbf{N}$. Calculer pour tout $(\alpha, \beta) \in \mathbf{R}^2$, $\dfrac{\partial j_k}{\partial \alpha}(\alpha, \beta)$ et $\dfrac{\partial j_k}{\partial \beta}(\alpha, \beta)$.

   Quelle condition nécessaire d'optimalité a-t-on sur $\dfrac{\partial j_k}{\partial \alpha}(\alpha^{(k)}, \beta^{(k)})$ et $\dfrac{\partial j_k}{\partial \beta}(\alpha^{(k)}, \beta^{(k)})$?

2. Soit $k \in \mathbf{N}$. Montrer que $p^{(k-1)^T} r^{(k)} = 0$. *On remarquera que* $\frac{\partial j_k}{\partial \alpha}(\alpha^{(k)}, \beta^{(k)}) = -p^{(k)^T} r^{(k+1)}$.

   En déduire que $p^{(k)}$ n'est nul que si $r^{(k)}$ l'est.

   Dans la suite, il sera entendu que $r^{(k)} \neq 0$ pour les calculs.

3. Soit $k \in \mathbf{N}$. Justifier que ${p^{(k)}}^T A p^{(k)} \neq 0$, puis montrer que

$$\alpha^{(k)} = \frac{{p^{(k)}}^T r^{(k)}}{{p^{(k)}}^T A p^{(k)}} = \frac{{r^{(k)}}^T r^{(k)}}{{p^{(k)}}^T A p^{(k)}}.$$

4. Soit $k \geq 1$. À l'aide de la deuxième condition d'optimalité sur j_k, montrer que

$$\beta^{(k)} = -\frac{{p^{(k-1)}}^T A r^{(k)}}{{p^{(k-1)}}^T A p^{(k-1)}}.$$

B) Convergence de l'algorithme.

1. Pour tout $k \in \mathbf{N}$, on note $K_k = \text{Vect}\,\{r^{(0)}, \ldots, r^{(k)}\}$. Ce sous-espace vectoriel est appelé k^e espace de Krylov associé à A et $x^{(0)}$. Montrer que, pour tout $k \in \mathbf{N}$, $p^{(k)} \in K_k$.
2. Montrer que pour tout $k \in \mathbf{N}$, ${p^{(k-1)}}^T A p^{(k)} = 0$.
3. En déduire que pour tout $k \in \mathbf{N}$, ${p^{(k)}}^T A p^{(k)} = {r^{(k)}}^T A p^{(k)}$ et $\alpha^{(k)} = \dfrac{{r^{(k)}}^T r^{(k)}}{{r^{(k)}}^T A p^{(k)}}$,

   puis que pour tout $k \in \mathbf{N}$, ${r^{(k)}}^T r^{(k+1)} = 0$.
4. Soit $m \in \mathbf{N}^*$, on suppose que pour tout $0 \leq k \leq m$, $r^{(k)} \neq 0$.

   Soit $k \leq m$. Montrer, par récurrence sur l, que pour tout $0 \leq l \leq m$,

$$\begin{aligned} {p^{(l)}}^T A p^{(k)} = 0 &\quad \text{si } l \neq k, \\ {r^{(l)}}^T r^{(k)} = 0 &\quad \text{si } l \neq k, \\ {r^{(l)}}^T p^{(k)} = 0 &\quad \text{si } l > k. \end{aligned}$$

5. En déduire que l'algorithme converge en au plus n itérations.

Remarque Dans la pratique cependant, cette méthode étant assez sensible aux erreurs numériques, il est préférable de ne pas considérer le gradient conjugué comme une méthode exacte, et d'arrêter l'algorithme lorsqu'un critère classique, utilisant la norme du résidu par exemple, est satisfait (et ceci conduit souvent à effectuer un nombre d'itérations supérieur à n).