Méthodes numériques : optimisation.
Examen du 13 mai 2015

Recommandations

Le sujet est long, mais le barème en tiendra compte. Il n'est pas nécessaire de tout avoir résolu pour avoir la note maximale. Les trois exercices sont indépendants. Les questions de chaque exercice ne sont pas bloquantes pour les suivantes. N'hésitez donc pas à admettre des résultats et passer à la suite.

Pour chaque exercice, deux questions ou sous-questions commencent par l'intitulé « *Cours :* ». On ne demande pas de démonstration pour ces questions-là, mais de prendre soin de donner le résultat du cours en l'adaptant au contexte. Un nombre non-négligeable de points sera affecté à ces questions.

1 Règle de Wolfe, méthode BFGS et positivité.

On veut minimiser une fonction $f \in C^1(\mathbb{R}^n, \mathbb{R})$, dont on sait calculer le gradient, par la méthode BFGS : on se donne un point initial x_0, une matrice initiale H_0 symétrique définie positive (par exemple $H_0 = I_n$), et pour $k \geqslant 0$, on effectue les étapes suivantes.

- On pose $d_k = -H_k \nabla f(x_k)$ (et si $\nabla f(x_k) = 0$, on s'arrête).
- On choisit un pas α_k satisfaisant la règle de Wolfe au point x_k dans la direction d_k, en essayant d'abord le pas $\alpha_k = 1$.
- On pose $x_{k+1} = x_k + \alpha_k d_k$, puis $s_k = x_{k+1} - x_k$ et $y_k = \nabla f(x_{k+1}) - \nabla f(x_k)$.
- On calcule enfin H_{k+1} par la formule BFGS :

$$H_{k+1} = \left(I_n - \frac{s_k y_k^T}{\langle s_k, y_k \rangle} \right) H_k \left(I_n - \frac{y_k s_k^T}{\langle s_k, y_k \rangle} \right) + \frac{s_k s_k^T}{\langle s_k, y_k \rangle}. \tag{1}$$

1. Montrer qu'en dimension 1, si on suppose que le pas $\alpha_k = 1$ satisfait toujours la règle de Wolfe et que les réels y_k et s_k sont toujours non nuls, alors les itérées pour $k \geqslant 1$ correspondent exactement à la méthode de la sécante.
2. On rappelle que le pas α_k satisfait la deuxième condition de Wolfe avec la constante $0 < c_2 < 1$ si on a
$$\langle \nabla f(x_{k+1}), d_k \rangle \geqslant c_2 \langle \nabla f(x_k), d_k \rangle.$$
   (a) *Cours :* Quelle est l'autre condition à satisfaire pour la règle de Wolfe (pour la constante c_1) et que doivent satisfaire les constantes c_1 et c_2 ? Donner brièvement une interprétation de ces deux critères. À quelle condition est-on certain de pouvoir satisfaire la règle de Wolfe ?
   (b) Pour $k \geqslant 0$, si on suppose H_k définie positive et $\nabla f(x_k) \neq 0$, montrer, en utilisant la deuxième condition de Wolfe, qu'on a $\langle s_k, y_k \rangle > 0$.
3. Si on suppose que H_k est définie positive et que $\nabla f(x_k) \neq 0$, montrer que pour $v \in \mathbb{R}^n$, on a $\langle v, H_{k+1} v \rangle = \langle w, H_k w \rangle + \frac{\langle s_k, v \rangle^2}{\langle s_k, y_k \rangle}$ avec $w = v - \frac{\langle s_k, v \rangle}{\langle s_k, y_k \rangle} y_k$, et que la matrice H_{k+1} est symétrique définie positive.
4. En déduire que pour tout $k \geqslant 0$ tel que $\nabla f(x_k) \neq 0$, la matrice H_{k+1} est bien définie par la formule (1), symétrique définie positive, et vérifie la condition « de la sécante » :
$$\nabla f(x_{k+1}) - \nabla f(x_k) = H_{k+1}^{-1}(x_{k+1} - x_k).$$
   Cours : En observant que cette condition correspond à dire que H_{k+1}^{-1} est une approximation de la Hessienne de f en x_{k+1}, justifier l'appellation « quasi-Newton » de la méthode BFGS.

2 Cas test de la méthode de descente de gradient à pas optimal en dimension deux

On considère la fonction $f : \mathbb{R}^2 \to \mathbb{R}$ donnée par $f(x) = \frac{1}{2}\langle x, Ax\rangle + \langle B, x\rangle$, où $A \in M_2(\mathbb{R})$ est une matrice symétrique définie positive, et $B \in \mathbb{R}^2$.

On se donne $x_0 \in \mathbb{R}^2$ et on définit les suites $(x_k)_{k\in\mathbb{N}}$ et $(r_k)_{k\in\mathbb{N}}$ par

$$\begin{cases} r_k = \nabla f(x_k), \\ x_{k+1} = x_k - \alpha_k r_k. \end{cases}$$

où α_k minimise la fonction $t \mapsto f(x_k - t\, r_k)$ (c'est la méthode de descente du gradient à pas optimal).

1. Soit $x_* \in \mathbb{R}^2$ l'unique point de minimum de f. Montrer que pour tout $k \geqslant 0$,
$$f(x_k) - f(x_*) = \frac{1}{2}\langle x_k - x_*, A(x_k - x_*)\rangle = \frac{1}{2}\langle r_k, A^{-1} r_k\rangle.$$
2. Soit $x \in \mathbb{R}^2$, $d \in \mathbb{R}^2 \setminus \{0\}$ et $h : \mathbb{R} \to \mathbb{R}$ la fonction auxiliaire définie par $h(t) = f(x + td)$. Montrer que h admet un unique minimum en $t = \frac{-\langle Ax+B, d\rangle}{\langle d, Ad\rangle}$.
3. Montrer que si $r_k \neq 0$, on a $\alpha_k = \frac{\langle r_k, r_k\rangle}{\langle r_k, Ar_k\rangle}$ et qu'alors $r_{k+1} = r_k - \frac{\langle r_k, r_k\rangle}{\langle r_k, Ar_k\rangle} Ar_k$.
Vérifier à l'aide de cette formule que pour tout $k \geqslant 0$, les vecteurs r_{k+1} et r_k sont orthogonaux.
Cours : dans quel cadre plus général a-t-on ce résultat ?
4. On note L et ℓ les deux valeurs propres de A, avec $\ell \leqslant L$ et on note (e_L, e_ℓ) une base orthonormale de vecteurs propres pour A. On décompose r_0 dans cette base : $r_0 = a\, e_L + b\, e_\ell$.
   (a) Que se passe-t-il pour r_1 si $\ell = L$ ou si $a = 0$ ou $b = 0$?
   (b) On suppose que $\ell < L$. Montrer que pour tout $k \geqslant 0$, il existe $c_k \in \mathbb{R}$ tel que
$$r_k = \begin{cases} c_k(a\, e_L + b\, e_\ell) & \text{si } k \text{ est pair} \\ c_k(-b\, e_L + a\, e_\ell) & \text{si } k \text{ est impair} \end{cases} \tag{2}$$
   (c) Montrer que si $a \neq 0$ et $b \neq 0$, alors
$$\alpha_k = \begin{cases} \frac{a^2+b^2}{La^2+\ell b^2} & \text{si } k \text{ est pair} \\ \frac{a^2+b^2}{Lb^2+\ell a^2} & \text{si } k \text{ est impair,} \end{cases} \tag{3}$$
   (d) (*calculatoire*) Montrer que si $a \neq 0$ et $b \neq 0$, alors
$$c_{k+1} = \begin{cases} \frac{(L-\ell)ab}{La^2+\ell b^2} c_k & \text{si } k \text{ est pair} \\ \frac{(L-\ell)ab}{Lb^2+\ell a^2} c_k & \text{si } k \text{ est impair} \end{cases} \tag{4}$$
5. (*calculatoire*) À l'aide des formules (2) et (4), montrer que si k est pair ou impair, on a
$$\langle r_{k+1}, A^{-1} r_{k+1}\rangle = \frac{(L-\ell)^2 a^2 b^2}{(La^2+\ell b^2)(\ell a^2 + Lb^2)} \langle r_k, A^{-1} r_k\rangle.$$
6. On veut comparer cette méthode avec celle de descente de gradient à pas fixe.
   (a) *Cours :* quel est le taux de convergence ρ_{fixe} de la méthode de descente de gradient à pas fixe, lorsqu'on a choisi le meilleur pas fixe α possible (donner aussi la valeur de ce pas fixe) ?
   (b) En notant $\|x\|_A^2 = \langle x, Ax\rangle$, et $\rho_{\text{opt}} = \dfrac{L-\ell}{\sqrt{L^2+\ell^2+(\frac{a^2}{b^2}+\frac{b^2}{a^2})L\ell}}$, déduire des questions 1. et 5. que $\|x_k - x_*\|_A = \rho_{\text{opt}}^k \|x_0 - x_*\|_A$.
   (c) Montrer que $\rho_{\text{opt}} \leqslant \rho_{\text{fixe}}$. Dans quel cas particulier ces deux taux sont-ils égaux ? Dans ce cas, que peut-on remarquer à l'aide de la formule (3) ?
   Indication : on pourra utiliser l'inégalité $t + \frac{1}{t} \geqslant 2$, valable pour tout $t > 0$, et qui est stricte dès que $t \neq 1$.

3 Gradient conjugué pour les moindres carrés linéaires

Soient $M \in M_{m,n}(\mathbb{R})$ et $y \in \mathbb{R}^m$. Dans la suite $\|\cdot\|$ correspondra toujours à la norme euclidienne sur $\mathbb{R}^n$ ou $\mathbb{R}^m$, suivant le contexte. On s'intéresse au problème suivant :

$$\inf_{x\in\mathbb{R}^n} \|Mx - y\|. \tag{5}$$

1. Montrer que le problème (5) admet toujours au moins une solution. Montrer que les solutions sont caractérisées par l'équation suivante (appelée « équation normale ») :

$$M^T Mx = M^T y. \tag{6}$$

Indication : on pourra penser à une projection orthogonale sur un sous-espace vectoriel bien choisi.

On suppose pour la suite que la matrice M, vue comme une application de $\mathbb{R}^n$ dans $\mathbb{R}^m$, est injective.

2. Pourquoi a-t-on nécessairement $m \geqslant n$? Montrer que la solution du problème (5) est unique.
3. En écrivant $A = M^T M \in M_n(\mathbb{R})$ et $B = -M^T y \in \mathbb{R}^n$, montrer que le problème (5) revient à minimiser la fonction $f : x \mapsto \frac{1}{2}\langle x, Ax\rangle + \langle B, x\rangle$.
   Cours : pourquoi peut-on appliquer la méthode du gradient conjugué, et dans quels cas a-t-on intérêt à le faire, plutôt que de résoudre directement le système linéaire (6) ?
4. *Cours :* écrire l'initialisation et les équations standard donnant les itérées de la méthode du gradient conjugué pour cette fonction f, en fonction uniquement de M, y et d'un vecteur initial x_0. On notera comme dans le cours x_k pour le vecteur que l'on considère, r_k pour le gradient, p_k pour la direction de descente, et α_k le pas.

Les deux questions suivantes ne sont pas plus difficiles, mais un peu moins proches du cours.

On s'intéresse maintenant à savoir quel est le bon ordre des opérations à faire pour calculer le vecteur $M^T M\, p_k$ dont on a besoin dans l'algorithme, pour un cas particulier de matrice creuse ne contenant que de l'ordre de n éléments non-nuls.

5. * On considère pour $d_1,\ldots,d_n$ et $c_1,\ldots,c_n$ des réels non-nuls la matrice de $M_{n+1,n}(\mathbb{R})$ suivante :

$$M = \begin{pmatrix} d_1 & 0 & \cdots & 0 \\ 0 & d_2 & \ddots & \vdots \\ \vdots & \ddots & \ddots & 0 \\ 0 & \cdots & 0 & d_n \\ c_1 & c_2 & \cdots & c_n \end{pmatrix}.$$

   Montrer que M est injective et que $M^T M$ n'a que des éléments non-nuls. En déduire que pour calculer $(M^T M)\, p_k$, même en ayant déjà calculé $M^T M$, cela nécessite n^2 multiplications de réels. Montrer que pour calculer $M^T(M\, p_k)$ dans cet ordre, cela ne nécessite que $4n$ multiplications. Quelle est l'autre raison qui nous pousse à ne pas vouloir calculer initialement $M^T M$ et la stocker pour les calculs suivants, si n est grand ?

Si on a déjà calculé les vecteurs $M\, p_k$ et $M^T M\, p_k$, le calcul $\langle p_k, M^T M\, p_k\rangle = \|M\, p_k\|^2$ dont on a besoin pour le calcul de α_k nécessite n ou m opérations suivant la formule utilisée. On veut montrer qu'il y a un intérêt à utiliser la deuxième formule en terme de précision numérique, même si elle est plus coûteuse.

On rappelle que si η est la précision machine, alors l'erreur faite par la machine en calculant $a + b$ est d'ordre $\max(|a|, |b|)\eta$, et l'erreur faite en calculant ab est d'ordre $|ab|\eta$.

6. * On prend $M = \begin{pmatrix} \varepsilon & 0 \\ 1 & 1 \end{pmatrix}$ et $p = \begin{pmatrix} 1 \\ -1+\delta \end{pmatrix}$ avec $0 < \varepsilon \ll 1$ et $0 < \delta \ll 1$. Si η est la précision machine et que $\eta \ll \varepsilon$ et $\eta \ll \delta$, alors montrer que l'erreur faite par la machine en calculant Mp, puis $\|Mp\|^2$ est d'ordre $O((\varepsilon^2 + \delta)\eta)$. Montrer que l'erreur est bien plus grande, d'ordre $O(\eta)$, lorsque la machine calcule Mp, puis $M^T(Mp)$, puis $\langle p, M^T(Mp)\rangle$.