UFR
DE MATHÉMATIQUES
ET INFORMATIQUE

UNIVERSITÉ
PARIS
DESCARTES

Optimisation (MML1E31) (M1 MM, 2017-2018)

Examen du vendredi 12 janvier 2018

Nombre de pages du sujet : 4. Durée 2h.

Les documents suivants sont autorisés :
— Polycopiés distribués en cours et notes de cours manuscrites correspondantes,
— Sujets de TP imprimés et notes manuscrites correspondantes.
La consultation de tout autre document (livres, etc.) et l'utilisation de tout appareil électronique sont interdites.

Exercice 1. *(sur environ 14 points)*

On rappelle que l'algorithme de Newton correspond au pseudo-code suivant :

Algorithme 1 : Algorithme de descente de Newton

Données : Un point initial $x^{(0)} \in \mathbb{R}^n$, un seuil de tolérance $\varepsilon > 0$, des paramètres $\alpha \in]0, \frac{1}{2}[$ et $\beta \in]0, 1[$ pour la méthode de rebroussement
Résultat : Un point $x \in \mathbb{R}^n$ proche de $x^\star$
Initialiser x :
$x \leftarrow x^{(0)}$;
$k \leftarrow 0$;
Calculer la première direction de descente :
$d^{(0)} = -\nabla^2 f(x^{(0)})^{-1} \nabla f(x^{(0)})$;
$\Lambda^{(0)} = -\langle d^{(0)}, \nabla f(x^{(0)}) \rangle$;
tant que $\Lambda^{(k)} > \varepsilon^2$ **faire**

1. Déterminer un pas de descente $t^{(k)} > 0$ au point $x^{(k)}$ selon la direction $d^{(k)}$ par la méthode de rebroussement avec les paramètres α et β.
2. Mettre à jour x :
   $x \leftarrow x^{(k+1)} = x^{(k)} + t^{(k)} d^{(k)}$;
   $k \leftarrow k + 1$;
3. Calculer la nouvelle direction de descente :
   $d^{(k)} = -\nabla^2 f(x^{(k)})^{-1} \nabla f(x^{(k)})$;
   $\Lambda^{(k)} = -\langle d^{(k)}, \nabla f(x^{(k)}) \rangle$;

fin

On rappelle également le théorème de convergence pour la méthode de Newton.

Théorème 1 (Convergence de la méthode de Newton)**.** Soit $f : \mathbb{R}^n \to \mathbb{R}$ une fonction fortement convexe telle que

$$\forall x, h \in \mathbb{R}^n, \quad m\|h\|^2 \leqslant \langle \nabla^2 f(x)h, h \rangle \leqslant M\|h\|^2$$

avec $0 < m \leqslant M$ et dont la matrice hessienne est lipschitzienne pour la constante $L > 0$

$$\forall x, y \in \mathbb{R}^n, \quad \|\nabla^2 f(x) - \nabla^2 f(y)\|_{\mathcal{M}_n(\mathbb{R})} \leqslant L\|x - y\|.$$

Soit $x^{(0)}$ un point quelconque de $\mathbb{R}^n$. On pose

$$\eta = \min(1, 3(1-2\alpha))\frac{m^2}{L} \quad \text{et} \quad \gamma = \alpha\beta\eta^2\frac{m}{M^2}.$$

Alors on a :

— Si $\|\nabla f(x^{(k)})\| \geqslant \eta$, alors

$$f(x^{(k+1)}) - f(x^{(k)}) \leqslant -\gamma.$$

— Si $\|\nabla f(x^{(k)})\| < \eta$, alors la méthode de rebroussement retourne le pas $t^{(k)} = 1$ et

$$\frac{L}{2m^2}\|\nabla f(x^{(k+1)})\| \leqslant \left(\frac{L}{2m^2}\|\nabla f(x^{(k)})\|\right)^2.$$

En particulier, l'algorithme de Newton converge et atteint un régime de convergence quadratique au bout d'un nombre fini d'itérations.

Le but de cet exercice est d'implémenter la méthode de Newton et de prouver la première partie du théorème de convergence. **Dans toute la suite de l'exercice on considère une fonction f vérifiant les hypothèses du Théorème 1.**

Partie I : Implémentation de l'Algorithme 1 *(sur environ 5 points)*

1. Rappeler le pseudo-code de la méthode de rebroussement appliquée à une fonction f en un point x selon une direction de descente d et les paramètres $\alpha \in]0, \frac{1}{2}[$ et $\beta \in]0, 1[$.
2. Exprimer le critère de sortie de la méthode de rebroussement pour la direction de descente de Newton en fonction de $x^{(k)}$, $d^{(k)}$, α et $\Lambda^{(k)}$.
3. On suppose que les fonctions `f = @(x)`, `gradf = @(x)` et `hessf = @(x)` qui évaluent respectivement $f(x)$, $\nabla f(x)$ et $\nabla^2 f(x)$ sont définies dans Octave. Implémenter une fonction
   `function [Xk, FXk] = desc_newton(x0,eps,alph,bet,f,gradf,hessf)`
   qui applique la méthode de Newton en appelant les trois fonctions `f`, `gradf`, et `hessf`. Cette fonction renvoie la matrice `Xk` (de taille n par nombre d'itérations) contenant la liste des vecteurs $x^{(k)}$ ainsi que la liste des valeurs `Fxk` de la fonctionnelle f aux points $x^{(k)}$.

Partie II : Quelques inégalités *(sur environ 3 points)*

4. Justifier qu'en tout point x la matrice hessienne $\nabla^2 f(x)$ est inversible et vérifie

$$\forall h \in \mathbb{R}^n, \quad \frac{1}{M}\|h\|^2 \leqslant \langle \nabla^2 f(x)^{-1}h, h\rangle \leqslant \frac{1}{m}\|h\|^2.$$

5. Soit $x^{(k)}$ un point de $\mathbb{R}^n$, et $d^{(k)} = -\nabla^2 f(x^{(k)})^{-1}\nabla f(x^{(k)})$ et $\Lambda^{(k)} = -\langle d^{(k)}, \nabla f(x^{(k)})\rangle$ les quantités associées. Montrer que

$$\|d^{(k)}\|^2 \leqslant \frac{1}{m}\Lambda^{(k)}$$

et que

$$\frac{1}{M}\|\nabla f(x^{(k)})\|^2 \leqslant \Lambda^{(k)}.$$

Partie III : Phase compensée *(sur environ 6 points)*

On suppose dans cette partie que l'algorithme a atteint l'itération k et on pose

$$d^{(k)} = -\nabla^2 f(x^{(k)})^{-1} \nabla f(x^{(k)}) \quad \text{et} \quad \Lambda^{(k)} = -\langle d^{(k)}, \nabla f(x^{(k)}) \rangle.$$

6. Montrer que pour tout $t \in \mathbb{R}$,

$$f(x^{(k)} + td^{(k)}) \leqslant f(x^{(k)}) - t\Lambda^{(k)} + \frac{M}{2m} t^2 \Lambda^{(k)}.$$

7. En déduire que pour tout t tel que $0 < t \leqslant \dfrac{m}{M}$, on a

$$f(x^{(k)} + td^{(k)}) \leqslant f(x^{(k)}) - \alpha \Lambda^{(k)}.$$

8. Soit $t^{(k)}$ le pas renvoyé par la méthode de rebroussement permettant de définir $x^{(k+1)} = x^{(k)} + t^{(k)} d^{(k)}$. Montrer que $t^{(k)} \geqslant \beta \frac{m}{M}$ et que

$$f(x^{(k+1)}) \leqslant f(x^{(k)}) - \alpha t^{(k)} \Lambda^{(k)}.$$

9. En déduire que

$$f(x^{(k+1)}) - f(x^{(k)}) \leqslant -\alpha\beta \frac{m}{M^2} \|\nabla f(x^{(k)})\|^2.$$

10. Conclure.

Exercice 2. *(sur environ 8 points)*

Dans cet exercice on considère $f : \mathbb{R}^n \to \mathbb{R}$ une fonctionnelle quadratique

$$f(x) = \frac{1}{2} \langle Ax, x \rangle - \langle b, x \rangle + c,$$

où $A \in \mathcal{M}_n(\mathbb{R})$ est une matrice carrée symétrique définie positive, $b \in \mathbb{R}^n$ et $c \in \mathbb{R}$. On rappelle que f est de classe $\mathcal{C}^\infty$ sur $\mathbb{R}^n$ et que pour tout $x \in \mathbb{R}^n$,

$$\nabla f(x) = Ax - b \quad \text{et} \quad \nabla^2 f(x) = A.$$

1. Démontrer que f admet un unique minimum global $x^\star$ donné par $x^\star = A^{-1} b$.

Dans la suite de l'exercice on étudie la convergence de l'algorithme de descente de gradient à pas fixe appliqué à cette fonctionnelle quadratique. Pour une fonction quelconque, cet algorithme est décrit par le pseudo-code suivant :

Algorithme 2 : Algorithme de descente de gradient à pas fixe

Données : Un point initial $x^{(0)} \in \mathbb{R}^n$, un seuil de tolérance $\varepsilon > 0$, un pas fixe $t > 0$
Résultat : Un point $x \in \mathbb{R}^n$ proche de $x^\star$
Initialiser x :
  $x \leftarrow x^{(0)}$;
  $k \leftarrow 0$;
tant que $\|\nabla f(x)\| > \varepsilon$ **faire**
  1. Mettre à jour x avec le pas fixe t dans la direction de descente $-\nabla f(x^{(k)})$:
    $x \leftarrow x^{(k+1)} = x^{(k)} - t\nabla f(x^{(k)})$;
    $k \leftarrow k + 1$;
fin

2. Soit $k \in \mathbb{N}$. Montrer que

$$x^{(k+1)} - x^\star = (I_n - tA)(x^{(k)} - x^\star).$$

   En déduire une expression de $x^{(k)} - x^\star$ en fonction A, t, et $x^{(0)} - x^\star$.

3. Soit $0 < \lambda_1 \leqslant \lambda_2 \leqslant \cdots \leqslant \lambda_n$ les valeurs propres de A et $u_1, u_2, \ldots, u_n$ des vecteurs propres associés formant une base orthonormée de $\mathbb{R}^n$. En introduisant la décomposition du vecteur $x^{(0)} - x^\star$ dans la base $(u_1, u_2, \ldots, u_n)$, montrer que la suite $(x^{(k)})_{k\in\mathbb{N}}$ converge vers $x^\star$ quelque soit la valeur de $x^{(0)}$ si et seulement si

$$\forall i \in \{1, \ldots, n\}, \quad |1 - t\lambda_i| < 1.$$

4. Montrer que

$$\forall i \in \{1, \ldots, n\}, \quad |1 - t\lambda_i| \leqslant \max(|1 - t\lambda_1|, |1 - t\lambda_n|).$$

5. En déduire l'intervalle des valeurs de t pour lesquelles l'Algorithme 2 converge et montrer que

$$\|x^{(k)} - x^\star\| \leqslant \max(|1 - t\lambda_1|, |1 - t\lambda_n|)^k \|x^{(0)} - x^\star\|.$$

6. Tracer l'allure des graphiques des fonctions $t \mapsto |1 - t\lambda_1|$ et $t \mapsto |1 - t\lambda_n|$. Pour quelle valeur de t le membre de droite de l'inégalité précédente tend-il le plus vite vers 0 ?