UFR
DE MATHÉMATIQUES
ET INFORMATIQUE

UNIVERSITÉ
PARIS
DESCARTES

Optimisation, algorithmique (MML1E31) (M1 Maths, 2016-2017)

Examen partiel du mercredi 16 novembre 2016
Corrigé

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 4 points)*
Soient $n, p \in \mathbb{N}^*$ deux entiers non nuls. Soient $A \in \mathcal{M}_{p,n}(\mathbb{R})$ une matrice rectangulaire, $b \in \mathbb{R}^p$ un vecteur et $\lambda > 0$. On définit la fonction $f : \mathbb{R}^n \to \mathbb{R}$ par

$$f(x) = \|Ax - b\|^2 + \lambda\|x\|^2$$

(attention on utilise la même notation pour la norme dans $\mathbb{R}^n$ et dans $\mathbb{R}^p$, mais les vecteurs ne sont a priori pas de la même taille). On admet que f est deux fois différentiable sur $\mathbb{R}^n$.

1. Donner l'expression du gradient $\nabla f(x)$ et de la matrice hessienne $\nabla^2 f(x)$ de f en tout point $x \in \mathbb{R}^n$.
2. Montrer que la fonction f est fortement convexe.
3. Est-ce que l'algorithme de descente de gradient converge pour cette fonction f ?

Solution de l'exercice 1.

1. On peut soit dériver chacun des deux termes et pour le premier utiliser les formules de dérivation des fonctions composées, soit mettre la fonctionnelle sous la forme d'une fonctionnelle quadratique. On détaille cette deuxième solution.
   En développant et regroupant les termes selon leur degré en x, on a
   $$f(x) = \frac{1}{2}\langle(2\lambda I_n + 2A^T A)x, x\rangle - \langle 2A^T b, x\rangle + \|b\|^2.$$
   Donc f est une fonctionnelle quadratique pour la matrice $A' = 2\lambda I_n + 2A^T A$, le vecteur $b' = 2A^T b$, et la constante $c' = \|b\|^2$. On en déduit que le gradient de f est
   $$\nabla f(x) = A'x - b' = (2\lambda I_n + 2A^T A)x - 2A^T b = 2\lambda x + 2A^T(Ax - b)$$
   et la matrice hessienne de f est
   $$\nabla^2 f(x) = 2\lambda I_n + 2A^T A.$$

2. Soient $x, h \in \mathbb{R}^n$. Alors

$$\langle \nabla^2 f(x)h, h\rangle = \langle (2\lambda I_n + 2A^T A)h, h\rangle = 2\lambda \|h\|^2 + 2\|Ah\|^2 \geqslant 2\lambda \|h\|^2.$$

   Donc f est fortement convexe pour la constante $m = 2\lambda$.

3. Pour que l'algorithme de gradient converge il faut que f soit fortement convexe et également qu'il existe une constante $M > 0$ telle que

$$\langle \nabla^2 f(x)h, h\rangle \leqslant M.$$

   Comme ici la matrice hessienne est constante, on peut prendre M comme étant la norme subordonnée de la matrice hessienne $A' = 2\lambda I_n + 2A^T A$. Donc les hypothèses du théorème de converge de l'algorithme de descente de gradient sont vérifiés. Cet algorithme converge linéairement.

Exercice 2. *(sur environ 6 points)*

1. Rappeler l'algorithme de calcul de pas de descente par la méthode de rebroussement en spécifiant le domaine de validité des paramètres α et β.
2. On suppose que les fonctions

   `function fx = f(x)` et `function gfx = gradf(x)`

   sont définies dans scilab. Donner le code d'une fonction scilab

   `function t = rebroussement(x, d, alph, bet)`

   qui prend en entrées les variables représentant le point $x \in \mathbb{R}^n$, la direction de descente $d \in \mathbb{R}^n$ et les paramètres α et β et renvoie le pas de descente t calculé par la méthode de rebroussement pour la fonction f.

Dans la suite de l'exercice on suppose que f est une fonction deux fois différentiable sur $\mathbb{R}$ et 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 $0 < m \leqslant M$. On note $x^\star$ le point où f atteint son minimum global.

Soient $x \in \mathbb{R}^n$, $x \neq x^\star$, et $d \in \mathbb{R}^n$ une direction de descente au point x pour la fonction f.

3. Montrer que

$$f(x + td) \leqslant f(x) + t\langle \nabla f(x), d\rangle + \frac{t^2}{2} M\|d\|^2.$$

4. Montrer que le critère d'arrêt de la méthode de rebroussement est vérifié pour tout t tel que

$$0 < t \leqslant -2(1-\alpha)\frac{\langle \nabla f(x), d\rangle}{M\|d\|^2}.$$

5. En déduire une majoration du nombre d'itérations de la méthode de rebroussement au point x pour la direction de descente d.

Solution de l'exercice 2.

1. D'après le cours :

Algorithme 1 : Algorithme de calcul du pas de descente par méthode de rebroussement
Données : Un point $x \in \mathbb{R}^n$, une direction de descente associée $d \in \mathbb{R}^n$, deux réels $\alpha \in]0, \frac{1}{2}[$ et $\beta \in]0, 1[$ **Résultat :** Un pas de descente $t > 0$ Initialiser t : $t \leftarrow 1$; **tant que** $f(x + td) > f(x) + \alpha t\langle \nabla f(x), d\rangle$ **faire** Réduire t d'un facteur β : $t \leftarrow \beta t$; **fin**

2. La fonction est la suivante :

```
function t = rebroussement(x, d, alph, bet)
fx = f(x)
gfx = gradf(x)
t = 1
while( f(x+t*d) > fx + alph*t*gfx'*d )
   t = bet*t
end
endfunction
```

3. Comme f est deux fois différentiable sur $[x, x + td]$, d'après la formule de Taylor-Maclaurin, il existe $u \in]0, t[$ tel que

$$f(x + td) = f(x) + t\langle \nabla f(x), d\rangle + \frac{t^2}{2}\langle \nabla^2 f(x + ud)d, d\rangle.$$

Or par hypothèse, $\langle \nabla^2 f(x + ud)d, d\rangle \leqslant M\|d\|^2$. Ainsi,

$$f(x + td) \leqslant f(x) + t\langle \nabla f(x), d\rangle + \frac{t^2}{2}M\|d\|^2.$$

4. t vérifie le critère d'arrêt si

$$f(x + td) \leqslant f(x) + \alpha t\langle \nabla f(x), d\rangle.$$

D'après la majoration de la question précédente, le critère d'arrêt est vérifié pour tout t tel que

$$f(x) + t\langle \nabla f(x), d\rangle + \frac{t^2}{2}M\|d\|^2 \leqslant f(x) + \alpha t\langle \nabla f(x), d\rangle,$$

soit par équivalence,

$$t\langle \nabla f(x), d\rangle + \frac{t^2}{2}M\|d\|^2 \leqslant \alpha t\langle \nabla f(x), d\rangle$$

$$\frac{t}{2}M\|d\|^2 \leqslant -(1 - \alpha)\langle \nabla f(x), d\rangle$$

$$t \leqslant -2(1 - \alpha)\frac{\langle \nabla f(x), d\rangle}{M\|d\|^2}.$$

5. Dans la méthode de rebroussement $t = \beta^n$ où $n \in \mathbb{N}$ est le nombre d'itérations de la boucle "tant que/while". On vient de montrer que si
$$t \leqslant -2(1-\alpha)\frac{\langle \nabla f(x), d\rangle}{M\|d\|^2},$$
alors le critère d'arrêt de la boucle est vérifié. Ainsi dès lors que
$$\beta^n \leqslant -2(1-\alpha)\frac{\langle \nabla f(x), d\rangle}{M\|d\|^2}$$
soit
$$n \geqslant \frac{\log\left(-2(1-\alpha)\frac{\langle \nabla f(x), d\rangle}{M\|d\|^2}\right)}{\log(\beta)}$$
le critère d'arrêt est vérifié. On en déduit que n vaut au plus la partie entière supérieure de la fraction
$$\frac{\log\left(-2(1-\alpha)\frac{\langle \nabla f(x), d\rangle}{M\|d\|^2}\right)}{\log(\beta)}.$$
En particulier l'algorithme converge en un nombre fini d'itérations.

Exercice 3. *(sur environ 10 points)*

1. Soit $f : \mathbb{R}^n \to \mathbb{R}$ une fonction convexe différentiable. Montrer que pour tous $x, y \in \mathbb{R}^n$,
$$\langle \nabla f(y) - \nabla f(x), y - x\rangle \geqslant 0.$$

Dans la suite de l'exercice on suppose que f est une fonction deux fois différentiable sur $\mathbb{R}$ et 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 $0 < m \leqslant M$. On note $x^\star$ le point où f atteint son minimum global.

2. Montrer que la fonction $g(x) = f(x) - \frac{m}{2}\|x\|^2$ est convexe.
3. En déduire que pour tous $x, y \in \mathbb{R}^n$,
$$\langle \nabla f(y) - \nabla f(x), y - x\rangle \geqslant m\|y - x\|^2.$$

Dans la suite de l'exercice on cherche à étudier la convergence de l'algorithme de descente de gradient à pas fixe, à savoir

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

On suppose tout d'abord que l'algorithme converge.

4. Justifier en quelques mots l'intérêt de cet algorithme par rapport à l'algorithme de descente de gradient utilisant la méthode de rebroussement pour le calcul du pas de descente ?
5. On suppose que les fonctions

   `function fx = f(x)` et `function gfx = gradf(x)`

   sont définies dans scilab. Donner le code d'une fonction scilab

   `function x = desc_gradient_pas_fixe(x0, eps, t)`

   qui prend en entrées un point initial $x^{(0)} \in \mathbb{R}^n$, un seuil de tolérance $\varepsilon > 0$, et un pas fixe $t > 0$ et renvoie le point x calculé par l'algorithme de descente de gradient à pas fixe.

On passe maintenant à l'étude de la convergence. Soient $t > 0$ et $x^{(0)} \in \mathbb{R}^n$. On considère la suite de points définie par $x^{(k+1)} = x^{(k)} - t\nabla f(x^{(k)})$. On rappelle que comme pour tous $x, h \in \mathbb{R}^n$, $\langle \nabla^2 f(x)h, h\rangle \leqslant M\|h\|^2$, la fonction $x \mapsto \nabla f(x)$ est M-lipschitzienne, c'est-à-dire,

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

6. Justifier que l'on a

$$x^{(k+1)} - x^\star = x^{(k)} - x^\star - t(\nabla f(x^{(k)}) - \nabla f(x^\star)).$$

7. En déduire que

$$\|x^{(k+1)} - x^\star\|^2 \leqslant (1 - 2mt + M^2t^2)\|x^{(k)} - x^\star\|^2.$$

8. En déduire que pour tout $t \in \left]0, \frac{2m}{M^2}\right[$, l'algorithme de descente de gradient à pas fixe t converge linéairement.

Solution de l'exercice 3.

1. Comme f est convexe et différentiable, pour tous $x, y \in \mathbb{R}^n$,

$$f(y) \geqslant f(x) + \langle \nabla f(x), y - x\rangle.$$

   En échangeant le rôle de x et y, on a

$$f(x) \geqslant f(y) + \langle \nabla f(y), x - y\rangle.$$

   Ainsi en combinant ces deux inégalités on a

$$f(y) \geqslant f(y) + \langle \nabla f(y), x - y\rangle + \langle \nabla f(x), y - x\rangle,$$

   soit

$$0 \geqslant \langle \nabla f(y) - \nabla f(x), x - y\rangle \Leftrightarrow \langle \nabla f(y) - \nabla f(x), y - x\rangle \geqslant 0.$$

2. La fonction $g(x) = f(x) - \frac{m}{2}\|x\|^2$ est deux fois différentiable en tant que somme de fonctions deux fois différentiables. Par linéarité, pour tout $x \in \mathbb{R}^n$,

$$\nabla^2 g(x) = \nabla^2 f(x) - mI_n.$$

   Ainsi, pour tous $x, h \in \mathbb{R}^n$,

$$\langle \nabla^2 g(x)h, h\rangle = \langle \nabla^2 f(x)h, h\rangle - m\langle h, h\rangle \geqslant 0.$$

   Ainsi la matrice hessienne de g est positive et donc g est convexe d'après le théorème de caractérisation des fonctions convexe deux fois différentiable.

3. Soient $x, y \in \mathbb{R}^n$. Comme g est convexe, d'après la question 1 on a
$$\langle \nabla g(y) - \nabla g(x), y - x \rangle \geqslant 0$$
Or par linéarité, $\nabla g(x) = \nabla f(x) - mx$. D'où
$$\begin{aligned}\langle \nabla g(y) - \nabla g(x), y - x \rangle &= \langle \nabla f(y) - my - (\nabla f(x) - mx), y - x \rangle \\ &= \langle \nabla f(y) - \nabla f(x), y - x \rangle - m\|x - y\|^2\end{aligned}$$
D'où
$$\langle \nabla f(y) - \nabla f(x), y - x \rangle \geqslant m\|x - y\|^2.$$
4. A chaque itération cet algorithme fait zéro appel à la fonction $x \mapsto f(x)$ (et un seul appel à la fonction $x \mapsto \nabla f(x)$). En comparaison, avec la méthode de rebroussement, à chaque itération on doit calculer $f(x + td)$ pour tous les pas de descente candidats de la forme β^p (et également un seul appel à la fonction $x \mapsto \nabla f(x)$)
5. Le code est le suivant :
```
function x = desc_gradient_pas_fixe(x0, eps, t)
x = x0
gfx = gradf(x)
while(gfx'*gfx>eps^2)
   x = x - t*gfx
   gfx = gradf(x)
end
endfunction
```
6. Comme $x^\star$ est un point de minimum global on a $\nabla f(x^\star) = 0$. Donc
$$x^{(k+1)} - x^\star = x^{(k)} - x^\star - t\nabla f(x^{(k)}) = x^{(k)} - x^\star - t(\nabla f(x^{(k)}) - \nabla f(x^\star)).$$
7. On développe la norme carrée en utilisant $\|u - v\|^2 = \|u\|^2 - 2\langle u, v\rangle + \|v\|^2$. D'où
$$\begin{aligned}&\|x^{(k+1)} - x^\star\|^2 \\ &= \|x^{(k)} - x^\star\|^2 - 2t\langle x^{(k)} - x^\star, \nabla f(x^{(k)}) - \nabla f(x^\star)\rangle + t^2\|\nabla f(x^{(k)}) - \nabla f(x^\star)\|^2.\end{aligned}$$
Or d'après l'inégalité démontrée à la question 3
$$\langle x^{(k)} - x^\star, \nabla f(x^{(k)}) - \nabla f(x^\star)\rangle \geqslant m\|x^{(k)} - x^\star\|^2$$
et comme ∇f est M-lipschitzienne,
$$\|\nabla f(x^{(k)}) - \nabla f(x^\star)\|^2 \leqslant M^2\|x^{(k)} - x^\star\|^2.$$
Ainsi,
$$\|x^{(k+1)} - x^\star\|^2 \leqslant (1 - 2mt + M^2t^2)\|x^{(k)} - x^\star\|^2.$$
8. D'après l'inégalité précédente, la suite $(\|x^{(k)} - x^\star\|^2)_{k\in\mathbb{N}}$ tend vers 0 dès lors que $1 - 2mt + M^2t^2 < 1$. Or on a les équivalences (on rappelle que $t > 0$),
$$1 - 2mt + M^2t^2 < 1 \Leftrightarrow M^2t^2 < 2mt \Leftrightarrow t < \frac{2m}{M^2}.$$
Ainsi pour tout $t < \frac{2m}{M^2}$, on a
$$\|x^{(k)} - x^\star\| \leqslant \gamma^n\|x^{(0)} - x^\star\|$$
avec $\gamma = \sqrt{1 - 2mt + M^2t^2} \in [0, 1[$ et l'algorithme de gradient à pas constant converge donc linéairement.