UFR
DE MATHÉMATIQUES
ET INFORMATIQUE

UNIVERSITÉ
PARIS
DESCARTES

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

## Examen du mercredi 4 janvier 2017

*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 12 points)*

**Rappel sur les fonctionnelles quadratiques :** On rappelle qu'une fonction $g : \mathbb{R}^n \to \mathbb{R}$ est une fonctionnelle quadratique s'il existe une matrice carrée symétrique $A \in \mathcal{M}_n(\mathbb{R})$, un vecteur $b \in \mathbb{R}^n$ et une constante $c \in \mathbb{R}$ tels que

$$\forall x \in \mathbb{R}^n, \quad g(x) = \frac{1}{2}\langle Ax, x\rangle - \langle b, x\rangle + c$$

(et alors la matrice $A$, le vecteur $b$ et la constante $c$ sont uniques).

**Fonctionnelle des moindres carrés :** 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. On définit la fonction $f : \mathbb{R}^n \to \mathbb{R}$ par

$$f(x) = \frac{1}{2}\|Ax - b\|^2.$$

Attention on utilisera la même notation pour la norme et le produit scalaire dans $\mathbb{R}^n$ et dans $\mathbb{R}^p$, mais les vecteurs ne sont a priori pas de la même taille !

1. Montrer que $f$ est une fonctionnelle quadratique et préciser la matrice carrée $A' \in \mathcal{M}_n(\mathbb{R})$, le vecteur $b' \in \mathbb{R}^n$ et la constante $c' \in \mathbb{R}$ associés à $f$.
2. Donner l'expression du gradient $\nabla f(x)$ et de la matrice hessienne $\nabla^2 f(x)$ de $f$ en tout point $x \in \mathbb{R}^n$ *(on pourra se référer à un résultat du cours plutôt que de faire les calculs)*.
3. Montrer que $f$ est convexe.
4. Montrer que $f$ est fortement convexe si et seulement si $\ker(A) = \{0\}$.
5. Soit $x \in \mathbb{R}^n$ tel que $\nabla f(x) \neq 0$. Rappeler la définition d'une direction de descente $d \in \mathbb{R}^n$ au point $x$ et montrer que $d \in \mathbb{R}^n$ est une direction de descente si et seulement si

$$\langle Ax - b, Ad\rangle < 0$$

(en particulier $Ad \neq 0$).

6. Soit $x \in \mathbb{R}^n$ tel que $\nabla f(x) \neq 0$ et $d$ une direction de descente au point $x$. Montrer que le problème d'optimisation
$$\inf_{t \in \mathbb{R}} f(x + td)$$
admet une unique solution donnée par
$$t^\star = -\frac{\langle \nabla f(x), d \rangle}{\|Ad\|^2}. \tag{1}$$

**Algorithme de gradient conjugué pour les moindres carrés :** On suppose désormais que $\ker(A) = \{0\}$. On rappelle ci-dessous le pseudo-code de l'algorithme du gradient conjugué afin de minimiser une fonctionnelle quadratique

$$g(x) = \frac{1}{2}\langle A'x, x\rangle - \langle b', x\rangle + c'$$

lorsque $A'$ est symétrique définie positive (ce qui revient à résoudre le système linéaire $A'x = b'$). On rappelle que, par convention, pour cet algorithme ce sont les vecteurs opposés $-d^{(k)}$ qui sont des directions de descente.

**Algorithme 1 :** Algorithme du gradient conjugué

**Données :** Un point initial $x^{(0)} \in \mathbb{R}^n$, un seuil de tolérance $\varepsilon > 0$
**Résultat :** Un point $x \in \mathbb{R}^n$ proche de $x^\star$
Initialiser $x$ :
  $x \leftarrow x^{(0)}$ ;
  $k \leftarrow 0$ ;
Première itération :
1. Calculer $d^{(0)} = \nabla g(x) \quad (d^{(0)} = \nabla g(x^{(k)}))$
2. Calculer le pas de descente optimal $t^{(0)}$ dans la direction de descente $-d^{(0)}$
3. Mettre à jour $x$ :
   $x \leftarrow x^{(1)} = x^{(0)} - t^{(0)}d^{(0)}$ ; (attention c'est bien un signe $-$)
   $k \leftarrow k + 1$ ;

**tant que** $\|\nabla f(x)\|^2 > \varepsilon^2$ **faire**
1. Calculer $d^{(k)} = \nabla g(x^{(k)}) + \frac{\|\nabla g(x^{(k)})\|^2}{\|\nabla g(x^{(k-1)})\|^2} d^{(k-1)}$.
2. Calculer le pas de descente optimal $t^{(k)}$ dans la direction de descente $-d^{(k)}$
3. Mettre à jour $x$ :
   $x \leftarrow x^{(k+1)} = x^{(k)} - t^{(k)}d^{(k)}$ ; (attention c'est bien un signe $-$)
   $k \leftarrow k + 1$ ;

**fin**

7. En utilisant l'équation (1), donner l'expression des pas de descente $t^{(0)}$ et $t^{(k)}$, $k \geqslant 1$, pour la fonction $f(x) = \frac{1}{2}\|Ax - b\|^2$ en fonction de $\nabla f(x^{(0)})$, $A$ et $d^{(0)}$ pour $t^{(0)}$ et $\nabla f(x^{(k)})$, $A$ et $d^{(k)}$ pour $t^{(k)}$ (en faisant attention à la convention sur les directions de descentes).

8. Ecrire une fonction scilab

```
function x = grad_conj_moindres_carres(A, b, x0, eps)
```

qui applique l'algorithme du gradient conjugué à la fonctionnelle $f(x) = \frac{1}{2}\|Ax - b\|^2$ en partant du point $x_0$ et avec une tolérance $\varepsilon$ pour le critère d'arrêt. On pourra introduire une fonction intermédiaire qui calcule le gradient de $f$ (non obligatoire).

**Exercice 2.** *(sur environ 9 points)*

Soit $n \geqslant 2$ un entier. On note $e_1, e_2, \dots, e_n$ les vecteurs de la base canonique de $\mathbb{R}^n$. On note $\|x\| = \sqrt{\langle x, x\rangle}$ la norme euclidienne usuelle de $\mathbb{R}^n$ et

$$\|x\|_\infty = \max_{i \in \{1,\dots,n\}} |x_i|.$$

Dans la suite de l'exercice, $f : \mathbb{R}^n \to \mathbb{R}$ désigne une fonction deux fois différentiable sur $\mathbb{R}^n$ 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 et $p^\star = f(x^\star)$ la valeur minimale de $f$.

Le but de l'exercice est d'étudier la convergence d'un algorithme de descente qui prend pour direction de descente à l'étape $k$ le vecteur

$$d^{(k)} = -\frac{\partial f}{\partial x_i}(x^{(k)})e_i$$

où $i$ est le plus petit indice pour lequel

$$\left|\frac{\partial f}{\partial x_i}(x^{(k)})\right| = \|\nabla f(x^{(k)})\|_\infty.$$

Plus précisément, on considère l'Algorithme 2 ci-dessous.

**Algorithme 2 :** Algorithme de descente selon la dérivée partielle maximale

**Données :** Un point initial $x^{(0)} \in \mathbb{R}^n$, un seuil de tolérance $\varepsilon > 0$
**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. Déterminer le plus petit indice $i$ tel que $\left|\frac{\partial f}{\partial x_i}(x^{(k)})\right| = \|\nabla f(x^{(k)})\|_\infty$
2. Déterminer un pas de descente $t^{(k)} > 0$ par la méthode exacte pour la direction de descente $d^{(k)} = -\frac{\partial f}{\partial x_i}(x^{(k)})e_i$
(ou par la méthode de rebroussement de paramètres $\alpha$ et $\beta$).
3. Mettre à jour $x$ :
$x \leftarrow x^{(k+1)} = x^{(k)} + t^{(k)}d^{(k)}$ ;
$k \leftarrow k + 1$ ;

**fin**

**Implémentation de l'Algorithme 2 avec la méthode de rebroussement** On suppose que les fonctions

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

sont définies dans scilab. On rappelle qu'en Scilab si `u` est un vecteur $u = (u_1, \dots, u_n)$, alors

— `abs(u)` renvoie le vecteur $(|u_1|, \dots, |u_n|)$.

— `[v,i] = max(u)` renvoie la valeur maximale $v = \max_i u_i$ des coordonnées de $u$ ainsi que le plus petit indice $i$ tel que $u_i = v = \max_i u_i$.

1. Ecrire une fonction scilab

```
function x = desc_der_partielle_max(x0, eps, alph, bet)
```

qui prend en entrées un point initial $x^{(0)} \in \mathbb{R}^n$, un seuil de tolérance $\varepsilon > 0$, et deux paramètres $\alpha$ et $\beta$, et renvoie le point $x$ calculé par l'algorithme de descente selon la dérivée partielle maximale utilisant la méthode de rebroussement pour le calcul du pas de descente (avec les paramètres $\alpha \in [0, \frac{1}{2}[$ et $\beta \in ]0, 1[$).

**Etude de la convergence de l'Algorithme 2 avec la méthode exacte** Le but de la fin de l'exercice est de démontrer la convergence de l'Algorithme 2 avec la méthode exacte pour le calcul du pas de descente. On rappelle que la forte convexité de $f$ implique que pour tout $x \in \mathbb{R}^n$,

$$\|\nabla f(x)\|^2 \geqslant 2m(f(x) - p^\star).$$

On suppose qu'à l'étape $k$ de l'Algorithme 2 le point $x^{(k)}$ est différent de $x^\star$. On note $i$ le plus petit indice tel que $\left|\frac{\partial f}{\partial x_i}(x^{(k)})\right| = \|\nabla f(x^{(k)})\|_\infty$ et on pose $d^{(k)} = -\frac{\partial f}{\partial x_i}(x^{(k)}) e_i$.

2. Montrer que $\langle \nabla f(x^{(k)}), d^{(k)} \rangle = -\|\nabla f(x^{(k)})\|_\infty^2$ (ce qui montre que $d^{(k)}$ est bien une direction de descente pour le point $x$).
3. Montrer que pour tout $t > 0$,

$$f(x^{(k+1)}) \leqslant f(x^{(k)} + td^{(k)}) \leqslant f(x^{(k)}) - t\|\nabla f(x^{(k)})\|_\infty^2 + t^2\frac{M}{2}\|\nabla f(x^{(k)})\|_\infty^2.$$

4. En déduire que

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

5. Démontrer que pour tout $x \in \mathbb{R}^n$, $\|x\|_\infty^2 \geqslant \frac{1}{n}\|x\|^2$.
6. Montrer, à l'aide des deux questions précédentes, que

$$f(x^{(k+1)}) - p^\star \leqslant \left(1 - \frac{m}{nM}\right)\left(f(x^{(k)}) - p^\star\right).$$

7. Conclure sur la convergence de l'Algorithme 2.