UFR
DE MATHÉMATIQUES
ET INFORMATIQUE

UNIVERSITÉ
PARIS
DESCARTES

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

Examen partiel du lundi 20 novembre 2017
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 6 points)*

Soient m et n deux entiers non nuls. Soient $f : \mathbb{R}^m \to \mathbb{R}$ une fonction, $A \in M_{m,n}(\mathbb{R})$ une matrice rectangulaire et $b \in \mathbb{R}^m$ un vecteur colonne. On pose $g : \mathbb{R}^n \to \mathbb{R}$ la fonction définie par

$$\forall x \in \mathbb{R}^n, \quad g(x) = f(Ax + b).$$

1. Rappeler la définition d'une fonction strictement convexe sur $\mathbb{R}^n$.
2. On suppose que $\mathrm{Ker}(A) \neq \{0\}$. Montrer que g n'est pas strictement convexe sur $\mathbb{R}^n$.
3. Montrer que si f est strictement convexe sur $\mathbb{R}^m$ et $\mathrm{Ker}(A) = \{0\}$ alors la fonction g est également strictement convexe sur $\mathbb{R}^n$.
4. On considère la fonction $h : \mathbb{R}^2 \to \mathbb{R}$ définie par

$$h(x) = e^{2x_1 - x_2 - 1} + e^{-x_1 + 2} + e^{x_2}.$$

   Démontrer que h est strictement convexe sur $\mathbb{R}^2$ à l'aide de la question précédente.
5. Donner un script Octave qui :
   — Définit une fonction `h = @(x)` qui évalue $h(x)$ en s'assurant que cette fonction puisse être évaluée sur des listes de p points $(x^{(1)}, x^{(2)}, \dots, x^{(p)})$ données sous la forme d'une matrice de taille $n \times p$.
   — Affiche le graphe de la fonction h sur le domaine $[0, 2] \times [0, 2]$.

Solution de l'exercice 1.

1. f est *strictement convexe* sur $\mathbb{R}^n$ si

$$\forall x, y \in \mathbb{R}^n,\ x \neq y,\ \forall \theta \in]0, 1[, \quad f(\theta x + (1 - \theta)y) < \theta f(x) + (1 - \theta)f(y).$$

2. Pour tout $x \in \mathrm{Ker}(A)$, $g(x) = f(Ax + b) = g(b)$. En particulier, comme $\mathrm{Ker}(A)$ n'est pas réduit à $\{0\}$ il existe $x, y \in \mathrm{Ker}(A)$ tel que $x \neq y$, et alors pour tout $\theta \in]0, 1[$,

$$g(\theta x + (1 - \theta)y) = f(b) = \theta f(b) + (1 - \theta)f(b) = \theta g(x) + (1 - \theta)g(y).$$

   Donc g n'est pas strictement convexe.

3. Soient $x, y \in \mathbb{R}^n$, $x \neq y$ et $\theta \in]0,1[$. On a,

$$g(\theta x + (1-\theta)y) = f(A(\theta x + (1-\theta)y) + b) = f(\theta(Ax+b) + (1-\theta)(Ay+b))$$

Comme $\mathrm{Ker}(A) \neq 0$, $x \mapsto Ax$ est injective, et donc $x \neq y$ implique $Ax \neq Ay$. On peut donc appliquer l'inégalité de strict convexité de f aux points $Ax + b \neq Ay + b \in \mathbb{R}^m$, ce qui donne

$$g(\theta x + (1-\theta)y) < \theta f(Ax+b) + (1-\theta) f(Ay+b) = \theta g(x) + (1-\theta) g(y).$$

Donc g est bien strictement convexe.

4. h est la composée de la fonction

$$\begin{array}{ccc} \mathbb{R}^2 & \to & \mathbb{R}^3 \\ x & \mapsto & Ax + b \end{array}$$

avec

$$A = \begin{pmatrix} 2 & -1 \\ -1 & 0 \\ 0 & 1 \end{pmatrix} \quad \text{et} \quad b = \begin{pmatrix} -1 \\ 2 \\ 0 \end{pmatrix}$$

et la fonction

$$\begin{array}{ccc} \mathbb{R}^3 & \to & \mathbb{R} \\ \begin{pmatrix} y_1 \\ y_2 \\ y_3 \end{pmatrix} & \mapsto & e^{y_1} + e^{y_2} + e^{y_3}. \end{array}.$$

On a clairement $\mathrm{Ker}(A) = \{0\}$ car les deux vecteurs colonnes ne sont pas colinéaires. Par ailleurs, on a

$$\nabla^2 f(y) = \begin{pmatrix} e^{y_1} & 0 & 0 \\ 0 & e^{y_2} & 0 \\ 0 & 0 & e^{y_3} \end{pmatrix}$$

qui est définie positive pour tout $y \in \mathbb{R}^3$, donc f est bien strictement convexe sur $\mathbb{R}^3$. D'après la question 3), $h = f(Ax+b)$ est bien strictement convexe sur $\mathbb{R}^2$.

5.
```
x1 = linspace(0,2,200); x2 = x1; [X1,X2] = meshgrid(x1,x2); Z
= f([X1(:),X2(:)]'); Z=reshape(Z,size(X1));
mesh(X1,X2,Z)
```

Exercice 2. *(sur environ 17 points)*

Soit $n \geqslant 2$ un entier. On note $\|x\| = \sqrt{\langle x, x\rangle}$ la norme euclidienne usuelle de $\mathbb{R}^n$,

$$\|x\|_\infty = \max_{i \in \{1,\dots,n\}} |x_i| \quad \text{et} \quad \|x\|_1 = \sum_{i=1}^n |x_i|.$$

Pour $x \in \mathbb{R}^n$ on note $\mathrm{sign}(x)$ le vecteur de $\{-1,1\}^n$ tel que pour tout $i \in \{1,\dots,n\}$

$$(\mathrm{sign}(x))_i = \begin{cases} 1 & \text{si } x_i \geqslant 0, \\ -1 & \text{si } x_i < 0. \end{cases}$$

Dans la suite de l'exercice, $f : \mathbb{R}^n \to \mathbb{R}$ désigne une fonction deux fois différentiable sur $\mathbb{R}^n$, fortement convexe et 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

$$d^{(k)} = -\|\nabla f(x^{(k)})\|_1 \operatorname{sign}(\nabla f(x^{(k)})).$$

Plus précisément, on considère l'Algorithme 1 ci-dessous.

Algorithme 1 : Algorithme de descente selon la direction de plus grande pente pour $\|\cdot\|_1$

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. Calculer la direction de descente $d^{(k)} = -\|\nabla f(x^{(k)})\|_1 \operatorname{sign}(\nabla f(x^{(k)}))$.
2. Déterminer un pas de descente $t^{(k)} > 0$ par la méthode exacte pour la direction de descente $d^{(k)} = -\|\nabla f(x^{(k)})\|_1 \operatorname{sign}(\nabla f(x^{(k)}))$
(ou par la méthode de rebroussement de paramètres α et β).
3. Mettre à jour x :
$x \leftarrow x^{(k+1)} = x^{(k)} + t^{(k)} d^{(k)}$;
$k \leftarrow k + 1$;

fin

Partie I : Implémentation de l'Algorithme 1 avec la méthode de rebroussement *(sur environ 4 points)*

On suppose que les fonctions `f = @(x)` et `gradf = @(x)` qui évaluent respectivement $f(x)$ et $\nabla f(x)$ sont définies dans Octave. On rappelle qu'en Octave si `u` est un vecteur $u = (u_1, \dots, u_n)$, alors `abs(u)` renvoie le vecteur $(|u_1|, \dots, |u_n|)$.

1. Ecrire une fonction Octave `sign` qui prend en entrée un vecteur colonne u et renvoie le vecteur $\operatorname{sign}(u) \in \{-1, 1\}^n$.
2. Ecrire une fonction Octave

```
function x = descente_norme_un(x0, eps, alph, bet, f, gradf)
```

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

Partie II : Quelques propriétés générales *(sur environ 6 points)*

3. Montrer que pour tout $x \in \mathbb{R}^n$, $\|x\|_1 \geqslant \|x\|$.
4. Soit $x \in \mathbb{R}^n$. Montrer que

$$\langle x, -\operatorname{sign}(x)\rangle = \min_{u\in\mathbb{R}^n,\ \|u\|_\infty \leqslant 1} \langle x, u\rangle.$$

5. Montrer que pour tout $y \in \mathbb{R}^n$,

$$f(y) \geqslant f(x) + \langle \nabla f(x), y-x\rangle + \frac{m}{2}\|y-x\|^2.$$

6. Montrer que $g : y \mapsto f(x)+\langle \nabla f(x), y-x\rangle + \frac{m}{2}\|y-x\|^2$ est une fonctionnelle quadratique pour laquelle on précisera la matrice A, le vecteur b et la constante c.
7. Déterminer le minimum de g et en déduire que

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

Partie III : Etude de la convergence de l'Algorithme 1 avec la méthode exacte *(sur environ 7 points)*

Le but de la fin de l'exercice est de démontrer la convergence de l'Algorithme 1 **avec la méthode exacte** pour le calcul du pas de descente. On suppose qu'à l'étape k de l'Algorithme 1 le point $x^{(k)}$ est différent de $x^\star$ et on pose $d^{(k)} = -\|\nabla f(x^{(k)})\|_1 \operatorname{sign}(\nabla f(x^{(k)}))$.

8. Montrer que $\langle \nabla f(x^{(k)}), d^{(k)}\rangle = -\|\nabla f(x^{(k)})\|_1^2$ et justifier que $d^{(k)}$ est bien une direction de descente pour le point x.
9. Montrer que pour tout $t > 0$,

$$f(x^{(k+1)}) \leqslant f(x^{(k)} + td^{(k)}) \leqslant f(x^{(k)}) - t\|\nabla f(x^{(k)})\|_1^2 + t^2\frac{M}{2}n\|\nabla f(x^{(k)})\|_1^2.$$

10. En déduire que

$$f(x^{(k+1)}) \leqslant f(x^{(k)}) - \frac{1}{2nM}\|\nabla f(x^{(k)})\|_1^2.$$

11. En déduire à l'aide des inégalités de la Partie II que

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

12. Conclure sur la convergence de l'Algorithme 1.

Solution de l'exercice 2.

1.
```
function v = sign(u)
v = ones(size(u));
v(find(u<0)) = -1;
end
```
2.

```
function v = sign(u)
v = ones(size(u));
v(find(u<0)) = -1;
end


function x = descente_norme_un(x0, eps, alph, bet, f, gradf)
x=x0;
gfx=gradf(x);
sqngfx = gfx'*gfx;
while(sqngfx > eps^2)
    normungfx = sum(abs(gfx));
    d = -normungfx*(sign(gfx));
    t=1;
    fx = f(x);
    while(f(x+t*d) > fx - alph*t*normungfx^2)
        % ou fx + alph*t*gfx'*d sans utiliser la question 6
        t = bet*t;
    end
    x = x+t*d;
    gfx = gradf(x);
    sqngfx = gfx'*gfx;
end
endfunction

% Parametres
A = [ 1 3 ; 1 -3 ; -1 0 ]
b = [ -0.1 ; -0.1 ; -0.1 ]

% Fonction somme d'exponentielle
f = @(x) sum(exp(A*x + b));

% Fonction gradient
gradf = @(x) A'*exp(A*x + b);


x0 = [0;0];
alph = 1/3;
bet = 0.8;
eps = 10^(-8);

% Solution
[xstar, vg_min] = fsolve(gradf,x0)
pstar = f(xstar)

x = descente_norme_un(x0, eps, alph, bet, f, gradf)
f(x)
```

3. $$\left(\sum_{i=1}^{n}|x_i|\right)^2 = \sum_{i=1}^{n} x_i^2 + 2\sum_{i<j}|x_i||x_j| \geqslant \sum_{i=1}^{n} x_i^2 = \|x\|^2.$$

4. On a bien $\|-\operatorname{sign}(x)\|_\infty = 1$ et pour tout $u = (u_1, u_2, \ldots, u_n)$ tel que $\|u\|_\infty \leqslant 1$ on a $1 \leqslant u_i \leqslant -1$, d'où
$$\langle x, u\rangle = \sum_{i=1}^{n} x_i u_i \geqslant \sum_{i=1}^{n} -|x_i||u_i| \geqslant \sum_{i=1}^{n} -|x_i| = \langle x, -\operatorname{sign}(x)\rangle.$$

5. D'après la formule de Taylor-Maclaurin, il existe $z \in [x, y]$ tel que
$$f(y) = f(x) + \langle \nabla f(x), y-x\rangle + \frac{1}{2}\langle \nabla^2 f(z)(y-x), y-x\rangle.$$
Or par hypothèse,
$$\langle \nabla^2 f(z)(y-x), y-x\rangle \geqslant m\|y-x\|^2$$
d'où l'inégalité.

6. Un simple calcul montre que $y \mapsto g(y)$ est une fonctionnelle quadratique avec $A = mI_n$ et $b = mx - \nabla f(x)$ et $c = g(0) = f(x) + \langle \nabla f(x), x\rangle + \frac{m}{2}\|x\|^2$.

7. On a
$$p^\star = \min_{y\in\mathbb{R}^n} f(y) \geqslant \min_{y\in\mathbb{R}^n} f(x) + \langle \nabla f(x), y-x\rangle + \frac{m}{2}\|y-x\|^2.$$
Or $y \mapsto g(y)$ est une fonctionnelle quadratique avec $A = mI_n$ et $b = mx - \nabla f(x)$. $A = mI_n$ est définie positive donc g admet un unique minimum global atteint en $y^\star = A^{-1}b = \frac{1}{m}(mx - \nabla f(x)) = x - \frac{1}{m}\nabla f(x)$. On a alors la valeur
$$g(y^\star) = f(x) - \frac{1}{m}\|\nabla f(x)\|^2 + \frac{1}{2m}\|\nabla f(x)\|^2 = f(x) - \frac{1}{2m}\|\nabla f(x)\|^2.$$
Ainsi
$$p^\star \geqslant f(x) - \frac{1}{2m}\|\nabla f(x)\|^2$$
et donc
$$\|\nabla f(x)\|^2 \geqslant 2m(f(x) - p^\star).$$

8. On a
$$\begin{aligned}\langle \nabla f(x^{(k)}), d^{(k)}\rangle &= -\|\nabla f(x^{(k)})\|_1 \langle \nabla f(x^{(k)}), \operatorname{sign}(\nabla f(x^{(k)}))\rangle \\ &= -\|\nabla f(x^{(k)})\|_1 \sum_{i=1}^{n}\left|\frac{\partial f}{\partial x_i}(x^{(k)})\right| = -\|\nabla f(x^{(k)})\|_1^2\end{aligned}$$
On a donc $\langle \nabla f(x^{(k)}), d^{(k)}\rangle < 0$, $d^{(k)}$ est bien une direction de descente au point $x^{(k)}$.

9. Par définition de la méthode exacte pour le calcul du pas de descente,
$$f(x^{(k+1)}) = \min_{t>0} f(x^{(k)} + td^{(k)}),$$
donc pour tout $t > 0$, on a bien
$$f(x^{(k+1)}) \leqslant f(x^{(k)} + td^{(k)}).$$

D'après la formule de Taylor-Maclaurin, il existe $z \in [x^{(k)}, x^{(k)} + td^{(k)}]$ tel que

$$f(x^{(k)} + td^{(k)}) = f(x^{(k)}) + t\langle \nabla f(x^{(k)}), d^{(k)}\rangle + t^2\frac{1}{2}\langle \nabla^2 f(z)d^{(k)}, d^{(k)}\rangle.$$

Or par hypothèse,

$$\langle \nabla^2 f(z)d^{(k)}, d^{(k)}\rangle \leqslant M\|d^{(k)}\|^2.$$

Enfin, on remarque que

$$\|d^{(k)}\|^2 = \|\nabla f(x^{(k)})\|_1^2 \| \operatorname{sign}(\nabla f(x^{(k)}))\|^2 = \|\nabla f(x^{(k)})\|_1^2 n$$

et on vient de monter que

$$\langle \nabla f(x^{(k)}), d^{(k)}\rangle = \|\nabla f(x^{(k)})\|_1^2.$$

10. On choisit le $t > 0$ qui minimise l'expression $f(x^{(k)}) - t\|\nabla f(x^{(k)})\|_1^2 + t^2\frac{M}{2}n\|\nabla f(x^{(k)})\|_1^2$. Il est donné pour t tel que

$$-\|\nabla f(x^{(k)})\|_1^2 + tMn\|\nabla f(x^{(k)})\|_1^2 = 0$$

soit

$$t = \frac{1}{Mn}.$$

Pour cette valeur de t on a

$$f(x^{(k)}) - t\|\nabla f(x^{(k)})\|_1^2 + t^2\frac{M}{2}n\|\nabla f(x^{(k)})\|_1^2 = f(x^{(k)}) - \frac{1}{2nM}\|\nabla f(x^{(k)})\|_1^2.$$

Ainsi, on a bien

$$f(x^{(k+1)}) \leqslant f(x^{(k)}) - \frac{1}{2Mn}\|\nabla f(x^{(k)})\|_1^2.$$

11. On a

$$\|\nabla f(x^{(k)})\|_1^2 \geqslant \|\nabla f(x^{(k)})\|^2 \geqslant 2m(f(x) - p^\star),$$

d'où

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

12. Par récurrence immédiate on a

$$f(x^{(k)}) - p^\star \leqslant \left(1 - \frac{m}{nM}\right)^k \left(f(x^{(0)}) - p^\star\right)$$

et $1 - \frac{m}{nM} \in [0, 1[$ donc l'Algorithme 1 converge linéairement.