UFR
DE MATHÉMATIQUES
ET INFORMATIQUE

UNIVERSITÉ
PARIS
DESCARTES

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

Examen du mercredi 4 janvier 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 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 ε pour le critère d'arrêt. On pourra introduire une fonction intermédiaire qui calcule le gradient de f (non obligatoire).

Solution de l'exercice 1.

1. $$\begin{aligned} f(x) &= \frac{1}{2}\|Ax - b\|^2 \\ &= \frac{1}{2}\langle Ax - b, Ax - b\rangle \\ &= \frac{1}{2}\langle Ax, Ax\rangle - \langle b, Ax\rangle + \frac{1}{2}\langle b, b\rangle \\ &= \frac{1}{2}\langle A^T Ax, x\rangle - \langle A^T b, x\rangle + \frac{1}{2}\|b\|^2. \end{aligned}$$

   Donc f est quadratique avec

   $$A' = A^T A, \quad b' = A^T b, \quad c = \frac{1}{2}\|b\|^2.$$

2. D'après les expression du gradient et de la matrice hessienne des fonctionnelles quadratiques,

   $$\nabla f(x) = A'x - b' = A^T(Ax - B)$$

   et

   $$\nabla^2 f(x) = A' = A^T A.$$

3. Pour tout $h \in \mathbb{R}^n$,

   $$\langle \nabla^2 f(x)h, h\rangle = \langle A^T Ah, h\rangle = \|Ah\|^2 \geqslant 0.$$

   Donc la matrice hessienne de f est positive en tout point $x \in \mathbb{R}^n$, f est donc convexe. On peut également justifier que f est convexe car c'est la composée de l'application affine $x \mapsto Ax - b$ et de la fonction convexe $y \mapsto \|y\|^2$.

4. Supposons que f soit fortement convexe. Alors il existe $m > 0$ tel que

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

   Or $\langle \nabla^2 f(x)h, h\rangle = \|Ah\|^2$. D'où, $\|Ah\|^2 \geqslant m\|h\|^2$ pour tout $h \in \mathbb{R}^n$, en particulier, $h \neq 0$ implique $Ah \neq 0$, donc $\ker(A) = \{0\}$.
   Supposons que $\ker(A) = \{0\}$. Alors $A' = A^T A$ est une matrice réelle symétrique définie positive. Elle admet une valeur propre minimale $\lambda > 0$ et on a

   $$\langle A^T Ah, h\rangle \geqslant \lambda\|h\|^2,$$

   donc f est fortement convexe pour $m = \lambda$.

5. $d \in \mathbb{R}^n$ est une *direction de descente* au point x si $\langle \nabla f(x), d\rangle < 0$. Ainsi d est une direction de descente ssi
$$\langle A^T(Ax-b), d\rangle < 0$$
soit ssi
$$\langle Ax-b, Ad\rangle < 0.$$
6. La fonction $\varphi : t \mapsto f(x+td)$ est un trinôme du second degré pour la variable t. En effet, on a
$$\varphi(t) = f(x+td) = f(x) + t\langle \underbrace{A^T(Ax-b)}_{=\nabla f(x)}, d\rangle + \frac{t^2}{2}\|Ad\|^2.$$
Cette fonction est minimale au point t tel que $\varphi'(t) = 0$, soit
$$\langle \nabla f(x), d\rangle + t\|Ad\|^2 = 0$$
d'où l'expression de $t^\star$.
7. On en appliquant l'équation (1) avec $d = -d^{(0)} = -\nabla f(x^{(0)})$,
$$t^{(0)} = \frac{\|d^{(0)}\|^2}{\|Ad^{(0)}\|^2}.$$
$$t^{(k)} = \frac{\langle \nabla f(x^{(x)}), d^{(k)}\rangle}{\|Ad^{(k)}\|^2}.$$
8.

```
function x = grad_conj_moindres_carres(A, b, x0, eps)
x = x0
k = 0
gfx = A'*(A*x-b)
sqngfx = gfx'*gfx
d = gfx;
Ad = A*d;
t = sqngfx/(Ad'*Ad)
x = x - t*d
sqngfxold = sqngfx
gfx = A'*(A*x - b)
sqngfx = gfx'*gfx
k = k+1

while(sqngfx>eps^2 & k < 1000)
   d = gfx + sqngfx/sqngfxold*d
   Ad = A*d;
   t = gfx'*d/(Ad'*Ad)
   x = x - t*d
   sqngfxold = sqngfx
   gfx = A*x - b
   sqngfx = gfx'*gfx
   k = k+1
```

```
  disp([k,sqngfx]);
end

endfunction
```

Exercice 2. *(sur environ 9 points)*

Soit $n \geqslant 2$ un entier. On note $e_1, e_2, \ldots, e_n$ 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

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

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 $\left| \dfrac{\partial f}{\partial x_i}(x^{(k)}) \right| = \|\nabla f(x^{(k)})\|_\infty$
2. Déterminer un pas de descente $t^{(k)} > 0$ par la méthode exacte pour la direction de descente $d^{(k)} = -\dfrac{\partial f}{\partial x_i}(x^{(k)}) e_i$ (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

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 α et β, 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.

Solution de l'exercice 2.

1.

```
function x = desc_der_partielle_max(x0, eps, alph, bet)
x=x0;
gfx=gradf(x);
sqngfx = gfx'*gfx;
while(sqngfx > eps^2)
    [ninfgfx,i] = max(abs(gfx));
    d = zeros(gfx);
    d(i) = -gfx(i);

    t=1;
    xk = x;
    fxk = f(xk);
    x = xk + t*d;
    fx = f(x);
    while(fx > fxk - alph*t*ninfgfx^2)
        // ou fxk + alph*t*gfx'*d sans utiliser la question 2)
        t = bet*t;
        x = xk + t*d;
        fx = f(x);
    end
    gfx=gradf(x);
    sqngfx = gfx'*gfx;
    disp(fx)
end
endfunction


// Test:
function fx = objectif(x,A,b)
    fx = sum(exp(A*x + b));
endfunction

function gfx = grad_obj(x,A,b)
    gfx = A'*exp(A*x + b);
endfunction

// Attention : Besoin de definir les variables A et b avant d'appeler ces
deff('fx=f(x)','fx = objectif(x,A,b)');
deff('gfx=gradf(x)','gfx = grad_obj(x,A,b)');

// Definition de la fonctionnelle
A = [ 1 3 ; 1 -3 ; -1 0 ]
b = [ -0.1 ; -0.1 ; -0.1 ]
x0 = [-2 ; 3]
eps = 10^(-8)
alph = 0.4;
bet = 0.8;
```

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

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

2. On a
$$\langle \nabla f(x^{(k)}), d^{(k)} \rangle = -\frac{\partial f}{\partial x_i}(x^{(k)})^2 = -\|\nabla f(x^{(k)})\|_\infty^2.$$

3. 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,
$$\langle \nabla^2 f(z) d^{(k)}, d^{(k)} \rangle \leqslant M\|d^{(k)}\|^2 = M\frac{\partial f}{\partial x_i}(x^{(k)})^2 = M\|\nabla f(x^{(k)})\|_\infty^2,$$
d'où la deuxième inégalité.

4. On choisit le $t > 0$ qui minimise l'expression $f(x^{(k)}) - t\|\nabla f(x^{(k)})\|_\infty^2 + t^2 \frac{M}{2}\|\nabla f(x^{(k)})\|_\infty^2$. Il est donné pour t tel que
$$-\|\nabla f(x^{(k)})\|_\infty^2 + tM\|\nabla f(x^{(k)})\|_\infty^2 = 0$$
soit
$$t = \frac{1}{M}.$$
Pour cette valeur de t on a
$$f(x^{(k)}) - t\|\nabla f(x^{(k)})\|_\infty^2 + t^2 \frac{M}{2}\|\nabla f(x^{(k)})\|_\infty^2 = f(x^{(k)}) - \frac{1}{2M}\|\nabla f(x^{(k)})\|_\infty^2.$$
Ainsi, on a bien
$$f(x^{(k+1)}) \leqslant f(x^{(k)}) - \frac{1}{2M}\|\nabla f(x^{(k)})\|_\infty^2.$$

5. Soit $x \in \mathbb{R}^n$. Alors pour tout $i \in \{1, \ldots, n\}$, $x_i^2 \leqslant \|x\|_\infty^2$, d'où
$$\|x\|^2 = \sum_{i=1}^{n} x_i^2 \leqslant n\|x\|_\infty^2.$$
Donc on a bien $\|x\|_\infty^2 \geqslant \frac{1}{n}\|x\|^2$.

6. On a
$$\|\nabla f(x^{(k)})\|_\infty^2 \geqslant \frac{1}{n}\|\nabla f(x^{(k)})\|^2 \geqslant \frac{2m}{n}(f(x) - p^\star),$$
d'où
$$f(x^{(k+1)}) - p^\star \leqslant f(x^{(k)}) - p^\star - \frac{1}{2M}\|\nabla f(x^{(k)})\|_\infty^2 \leqslant \left(1 - \frac{m}{nM}\right)\left(f(x^{(k)}) - p^\star\right).$$

7. 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 2 converge linéairement.