Optimisation algorithmique - Examen du 10 janvier 2019 - Correction

Exercice 1. On considère la fonction suivante de $\mathbb{R}^2$ dans $\mathbb{R}$:

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

1. *question préliminaire.* Montrer que l'équation $t+4+2e^t = 0$ admet une unique solution dans $\mathbb{R}$, qu'on notera γ (on ne cherchera pas à calculer sa valeur).

2. Calculer le gradient de f, trouver ses points critiques et montrer que f admet un unique minimiseur global.

Correction

1. La fonction $g(t) = t + 4 + 2e^t$ est dérivable sur $\mathbb{R}$, de dérivée $g'(t) = 1 + 2e^t > 0$; donc g est strictement croissante et continue sur $\mathbb{R}$. De plus,

   - lorsque $t \to -\infty$, $e^t \to 0$ donc $g(t) = t + 4 + 2e^t \to -\infty$;
   - lorsque $t \to +\infty$, $e^t \to +\infty$ donc $g(t) = t + 4 + 2e^t \to +\infty$.

   Par conséquent l'équation $g(t) = 0$ admet une solution unique sur $\mathbb{R}$, que l'on note γ.

2. $$\nabla f(x) = \begin{pmatrix} 2(x_1 - x_2 + 2) + x_2^2 e^{x_1} \\ -2(x_1 - x_2 + 2) + 2x_2 e^{x_1} \end{pmatrix}.$$

   Les points critiques de f vérifient

$$\begin{aligned}
\nabla f(x) = 0 \quad &\Leftrightarrow \quad \begin{cases} 2(x_1 - x_2 + 2) + x_2^2 e^{x_1} = 0 \\ -2(x_1 - x_2 + 2) + 2x_2 e^{x_1} = 0 \end{cases} \\
&\Leftrightarrow \quad \begin{cases} 2(x_1 - x_2 + 2) + x_2^2 e^{x_1} = 0 \\ x_2^2 e^{x_1} + 2x_2 e^{x_1} = 0 \end{cases} \\
&\Leftrightarrow \quad \begin{cases} 2(x_1 - x_2 + 2) + x_2^2 e^{x_1} = 0 \\ x_2^2 + 2x_2 = 0 \end{cases} \\
&\Leftrightarrow \quad \begin{cases} 2(x_1 - x_2 + 2) + x_2^2 e^{x_1} = 0 \\ x_2 = 0 \text{ ou } x_2 = -2 \end{cases} \\
&\Leftrightarrow \quad \begin{cases} x_1 = -2 \\ x_2 = 0 \end{cases} \text{ ou } \begin{cases} 2(x_1 + 2 + 2) + (-2)^2 e^{x_1} = 0 \\ x_2 = -2 \end{cases} \\
&\Leftrightarrow \quad \begin{cases} x_1 = -2 \\ x_2 = 0 \end{cases} \text{ ou } \begin{cases} x_1 + 4 + 2e^{x_1} = 0 \\ x_2 = -2 \end{cases} \\
&\Leftrightarrow \quad \begin{cases} x_1 = -2 \\ x_2 = 0 \end{cases} \text{ ou } \begin{cases} x_1 = \gamma \\ x_2 = -2 \end{cases}
\end{aligned}$$

   Ainsi f admet deux points critiques : $(-2, 0)$ et $(\gamma, -2)$.

Pour montrer que f admet un unique minimiseur global, le plus simple est de voir directement que $f(x)$ est positif pour tout $x \in \mathbb{R}^2$ (car les carrés et l'exponentielle sont positifs), et qu'elle est nulle si et seulement si les deux carrés sont nuls, i.e. si on a à la fois $x_1 - x_2 + 2 = 0$ et $x_2 = 0$, c'est-à-dire si $x = (-2, 0)$. Par conséquent $(-2, 0)$ est l'unique minimiseur global de f.

Exercice 2. On considère la fonction suivante de $\mathbb{R}^n$ dans $\mathbb{R}$:

$$f(x) = \sum_{i=1}^{n-2} (x_{i+2} - 2x_{i+1} + x_i)^2 + \sum_{i=1}^{n} (x_i - y_i)^2,$$

où $y \in \mathbb{R}^n$ est fixé.

1. Identifier la matrice A, le vecteur b et le réel c tels que f s'écrive sous la forme standard d'une fonctionnelle quadratique :

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

   où $\langle\, \cdot \,,\, \cdot \,\rangle$ désigne le produit scalaire canonique de $\mathbb{R}^n$.

2. Montrer que A est une matrice symétrique définie positive.

3. En déduire que f est une fonction fortement convexe et qu'elle admet un minimiseur unique x^*, et écrire le système d'équations linéaires vérifié par x^*.

4. Écrire une fonction Octave ou Matlab `function xstar = minim_f(y)` renvoyant la solution x^* en fonction du vecteur y donné en entrée.

Correction

1. On peut tout d'abord écrire

   $$f(x) = \|Mx\|^2 + \|x - y\|^2,$$

   où $\|\cdot\|$ est la norme euclidienne canonique sur $\mathbb{R}^n$, et M est la matrice de taille $(n-2) \times n$:

   $$M = \begin{pmatrix} 1 & -2 & 1 & 0 & \cdots & 0 \\ 0 & \ddots & \ddots & \ddots & \ddots & \vdots \\ \vdots & \ddots & \ddots & \ddots & \ddots & 0 \\ 0 & \dots & 0 & 1 & -2 & 1 \end{pmatrix}.$$

   Ensuite on a

   $$\begin{aligned} f(x) &= \|Mx\|^2 + \|x\|^2 = \langle\, Mx \,,\, Mx \,\rangle + \langle\, x \,,\, x \,\rangle + \langle\, y \,,\, y \,\rangle - 2 + \langle\, x \,,\, y \,\rangle \\ &= \left\langle\, M^T Mx \,,\, x \,\right\rangle + \langle\, x \,,\, x \,\rangle - 2 + \langle\, x \,,\, y \,\rangle + \langle\, y \,,\, y \,\rangle \\ &= \left\langle\, (M^T M + I_n)x \,,\, x \,\right\rangle - \langle\, x \,,\, 2y \,\rangle + \langle\, y \,,\, y \,\rangle . \end{aligned}$$

   Donc $A = 2(M^T M + I_n)$, $b = 2y$ et $c = \langle\, y \,,\, y \,\rangle$.

2. A est symétrique car $A^T = 2((M^TM)^T + I_n^T) = 2(M^TM + I_n) = A$. De plus pour tout x on a

$$\begin{aligned}\langle\ Ax\ ,\ x\ \rangle &= \left\langle\ 2(M^TM + I_n)x\ ,\ x\ \right\rangle = 2\left\langle\ M^TMx\ ,\ x\ \right\rangle + 2\langle\ x\ ,\ x\ \rangle \\ &= 2\langle\ Mx\ ,\ Mx\ \rangle + 2\langle\ x\ ,\ x\ \rangle = 2\|Mx\|^2 + 2\|x\|^2 .\end{aligned}$$

   Donc $\langle\ Ax\ ,\ x\ \rangle \geq 0$ et donc A est positive ; et de plus $\langle\ Ax\ ,\ x\ \rangle = 0$ si et seulement si $Mx = 0$ et $x = 0$, autrement dit si et seulement si $x = 0$. Ainsi A est bien définie positive.

3. La matrice hessienne de f est $Hf(x) = A$ pour tout $x \in \mathbb{R}^n$, et comme cette matrice est définie positive, elle a toutes ses valeurs propres strictement positives, donc il existe $\alpha > 0$ tel que toutes les valeurs propres de $Hf(x)$, quel que soit x, soient supérieures ou égales à α : il suffit de prendre pour α la plus petites des valeurs propres de A. Ainsi par définition f est fortement convexe. Par conséquent f admet un minimiseur unique x^*, qui est aussi l'unique point critique. Le système linéaire vérifié par x^* correspond à l'équation
$$\nabla f(x) = 0 \Leftrightarrow Ax - b = 0 \Leftrightarrow Ax = b.$$

4.
```
function xstar = minim_f(y)
    n = length(y);
    M = zeros(n-2,n);
    for i=1:n-2
        M(i:i+2) = [1,-2,1];
    end
    A = 2*(M'*M+eye(n));
    b = 2*y;
    xstar = A\b;
end
```

Exercice 3. Soit $f : \mathbb{R}^2 \to \mathbb{R}$ définie par
$$f(x) = x_1^2 + x_2^2 + e^{x_1}.$$

1. Montrer que f est fortement convexe, de constante m à déterminer.

2. Décrire de manière générale l'algorithme de descente de gradient avec méthode de rebroussement pour la minimisation d'une fonction.

3. Écrire une fonction Matlab ou Octave `function xstar = grad_desc(xinit,alpha,beta,eps)` réalisant la descente de gradient avec méthode de rebroussement pour la minimisation de la fonction f, avec initialisation x_{init}, constantes α et β pour la méthode de rebroussement, et condition d'arrêt ϵ sur la norme du gradient.

4. Rappeler les conditions de convergence de l'algorithme précédent. Sont-elles satisfaites ici ?

5. Montrer que l'on peut trouver une constante $R > 0$ en fonction de x_{init} telle que la suite des itérés x^k de l'algorithme de descente soit telle que $\|x^k\| < R$ pour tout $k \geq 0$. En déduire que les conditions de convergence sont satisfaites en considérant la fonction f restreinte à la boule ouverte $B(0, R)$.

6. Décrire de manière générale l'algorithme de Newton pour la minimisation d'une fonction. Écrire explicitement l'étape de mise à jour $x^k \mapsto x^{k+1}$ dans le cas de la fonction f.

Correction

1. Le gradient et la hessienne de f s'écrivent

$$\nabla f(x) = (2x_1 + e^{x_1}, 2x_2),$$

et

$$Hf(x) = \begin{pmatrix} 2 + e^{x_1} & 0 \\ 0 & 2 \end{pmatrix}.$$

Les valeurs propres de $Hf(x)$ sont toujours strictement positives, donc f est strictement convexe. De plus ces valeurs propres sont toujours supérieures ou égales à 2 puisque l'exponentielle est positive. donc f est fortement convexe, de constante $m = 2$.

2. L'algorithme de descente de gradient avec méthode de rebroussement consiste, à partir d'une fonction différentiable $f : \mathbb{R}^n \to \mathbb{R}$, d'un point initial $x_{init} \in \mathbb{R}^n$, et de deux constantes $\alpha \in]0, \frac{1}{2}[$ et $\beta \in]0, 1[$, à calculer la suite de points $x^k \in \mathbb{R}^n$ définie itérativement ainsi :

   - $x^0 = x_{init}$,
   - pour tout $k \geq 0$ (ou jusqu'à une condition d'arrêt spécifique), on pose :

$$x^{k+1} = x^k - \eta^k \nabla f(x^k),$$

   où η^k est défini par l'algorithme suivant :
     - On pose $\eta = 1$.
     - Tant que $f(x^k - \eta\nabla f(x^k)) > f(x^k) - \alpha\eta \left\|\nabla f(x^k)\right\|^2$,
       on met à jour η : $\eta \leftarrow \beta\eta$.
     - On pose $\eta^k = \eta$.

3.
```
func = @(x)x(1)^2+x(2)^2+exp(x(1));
gradfunc = @(x)[2*x(1)+exp(x(1)),2*x(2)];

function xstar = grad_desc(xinit,alpha,beta,eps)
    x = xinit;
    g = gradf(x);
    sqnorm_g = dot(g,g);
    while sqnorm_g > eps
        eta = 1;
        fx = f(x);
```

```
        while f(x-eta*g) > fx-alpha*eta*sqnorm_g
            eta = eta*beta;
        end
    end
    x = x-eta*g;
end
```

4. L'algorithme de descente de gradient avec méthode de rebroussement converge lorsque f est une fonction deux fois différentiable telle qu'il existe deux constantes m, M, $0 < m \leq M$, telles que

$$\forall x, h \in \mathbb{R}^n, \quad m \left\| h \right\|^2 \leq \langle\ Hf(x)h\ ,\ h\ \rangle \leq M \left\| h \right\|^2 .$$

Ici on a vu déjà que f est fortement convexe donc la première inégalité est vérifiée avec $m = 2$. Par contre il ne peut pas exister de constante M telle que la deuxième inégalité soit vérifiée pour tout $x \in \mathbb{R}^n$. En effet on a vu que

$$Hf(x) = \begin{pmatrix} 2 + e^{x_1} & 0 \\ 0 & 2 \end{pmatrix}$$

donc en posant $h = (1, 0)$ on a $\left\| h \right\|^2 = 1$, donc $M \left\| h \right\|^2 = M$, et $\langle\ Hf(x)h\ ,\ h\ \rangle = 2 + e^{x_1} \to +\infty$ lorsque $x_1 \to +\infty$, ce qui contredit le fait que $\langle\ Hf(x)h\ ,\ h\ \rangle$ soit borné par M. Ainsi les conditions de convergence ne sont pas satisfaites ici.

5. Par définition de la méthode de rebroussement, à chaque itération on a nécessairement (sauf si $x^k = x^*$) :

$$f(x^{k+1}) = f(x^k - \eta \nabla f(x^k)) \leq f(x^k) - \alpha \eta \| \nabla f(x^k) \|^2,$$

puisque c'est à cette condition que la boucle définissant η s'arrête (*remarque : en toute rigueur il faudrait prouver à nouveau ici que cette boucle se termine bien comme ça avait été fait dans le cours*). Par conséquent, $f(x^{k+1}) < f(x^k)$ et donc la suite $f(x^k)$ est décroissante. Ainsi,

$$\left\| x^k \right\|^2 < \left\| x^k \right\|^2 + e^{x_1^k} = f(x^k) \leq f(x_{init}),$$

et donc en posant $R = \sqrt{f(x_{init})} + 1$ on a bien que $\left\| x^k \right\| < R$ pour tout $k \geq 0$. A présent en considérant la fonction f restreinte à $B(0, R)$, d'après ce qui précède on est assuré que le suite des itérés x^k reste dans $B(0, R)$, et sur ce domaine on a

$$\langle\ Hf(x)h\ ,\ h\ \rangle \leq (2 + e^{x_1}) \left\| h \right\|^2 \leq (2 + e^R) \left\| h \right\|^2 ,$$

donc l'inégalité est vérifiée pour $M = 2 + e^R$, et les conditions de convergence sont donc satisfaites.

6. La méthode de Newton pour la minimisation d'une fonction f est une méthode de descente dans laquelle la direction de descente calculée à chaque étape k est égale à

$$d^k = -Hf(x^k)^{-1} \nabla f(x^k).$$

L'étape de mise à jour est donc

$$x^{k+1} = x^k - \eta^k Hf(x^k)^{-1} \nabla f(x^k),$$

où η^k est le pas de descente qui peut être calculé par n'importe quelle méthode de recherche de pas. Dans le cas de la fonction f de cet exercice, la matrice hessienne est diagonale et donc s'inverse facilement :

$$Hf(x)^{-1} = \begin{pmatrix} \frac{1}{2+e^{x_1}} & 0 \\ 0 & \frac{1}{2} \end{pmatrix}$$

On a donc

$$x^{k+1} = x^k - \eta^k \begin{pmatrix} (2x_1^k + e^{x_1^k})/(2 + e^{x_1^k}) \\ x_2^k \end{pmatrix},$$