Professeur : Michel Bierlaire, Assistant responsable : Virginie Lurkin, Nikola Obrenovic

Optimisation non linéaire – corrigé (8 décembre 2017)

Solution de la question 1:

Nous commençons par calculer le gradient $\nabla f(x)$ et la matrice hessienne $\nabla^2 f(x)$ de notre fonction $f(x)$. En utilisant les définitions du cours, on obtient :

$$\nabla f(x) = \begin{pmatrix} 6x_1^2 - 24 \\ 12x_2^3 - 12 \end{pmatrix} \tag{1}$$

$$\nabla^2 f(x) = \begin{pmatrix} 12x_1 & 0 \\ 0 & 36x_2^2 \end{pmatrix} \tag{2}$$

1. Afin de déterminer la nature de chacun des points donnés, il faut tester les conditions d'optimalité nécessaires et suffisantes en chacun de ces points. Les sections 5.1 et 5.2 du livre donnent un complément d'information, si nécessaire.

   Point $x_a = (1, 2)^T$

   Le gradient, donné dans l'équation (1), au point x_a vaut :

   $$\nabla f(x_a) = \begin{pmatrix} -18 \\ 84 \end{pmatrix} \neq 0.$$

   Puisque le gradient n'est pas nul, cela signifie que le point $x_a = (1, 2)^T$ n'est pas un point stationnaire.

   Point $x_b = (2, 1)^T$

   Le gradient, donné dans l'équation (1), au point x_b vaut :

   $$\nabla f(x) = \begin{pmatrix} 0 \\ 0 \end{pmatrix}$$

   Le gradient étant nul, on sait que le point x_b est un point stationnaire. Afin de déterminer sa nature, il faut utiliser la matrice hessienne (2) :

   $$\nabla^2 f(x) = \begin{pmatrix} 24 & 0 \\ 0 & 36 \end{pmatrix}$$

   On voit que la matrice hessienne est une matrice définie positive. Cela signifie donc que le point $x_b = (2, 1)^T$ est un minimum local.

Point $x_c = (-2, 1)^T$

Le gradient, donné dans l'équation (1), au point x_x vaut :

$$\nabla f(x) = \begin{pmatrix} 0 \\ 0 \end{pmatrix}$$

Le gradient étant nul, on sait que le point x_c est un point stationnaire. Afin de déterminer sa nature, il faut utiliser la matrice hessienne (2) :

$$\nabla f^2(x) = \begin{pmatrix} -24 & 0 \\ 0 & 36 \end{pmatrix}$$

La matrice hessienne n'étant ni définie positive, ni définie négative, cela signifie que le point $x_c = (-2, 1)^T$ est un point de selle.

2. Le point de départ est $x^0 = (2, 2)^T$. On sait, par définition, que :

$$x^{k+1} = x^k + d^k$$

Où d^k correspond à la direction que doit prendre l'algorithme. On peut trouver cette direction en résolvant le système suivant :

$$\nabla^2 f(x^k) d^k = -\nabla f(x^k) \tag{3}$$

Il nous suffit alors d'utiliser les équations (1) et (2) appliquées au point x^0 pour obtenir le système que nous devons résoudre :

$$\begin{pmatrix} 12 \cdot 2 & 0 \\ 0 & 36 \cdot 2^2 \end{pmatrix} \begin{pmatrix} d_1^0 \\ d_2^0 \end{pmatrix} = - \begin{pmatrix} 6 \cdot 2^2 - 24 \\ 12 \cdot 2^3 - 12 \end{pmatrix}$$

$$\begin{pmatrix} 24 & 0 \\ 0 & 144 \end{pmatrix} \begin{pmatrix} d_1^0 \\ d_2^0 \end{pmatrix} = - \begin{pmatrix} 0 \\ 12 \cdot 7 \end{pmatrix}$$

En résolvant ce système, on obtient :

$$d^0 = \begin{pmatrix} 0 \\ -{}^7\!/_{12} \end{pmatrix}$$

On peut donc calculer la valeur de la première itération :

$$\begin{aligned} x^1 &= x^0 + d^0 \\ &= \begin{pmatrix} 2 \\ 2 \end{pmatrix} + \begin{pmatrix} 0 \\ -7/12 \end{pmatrix} \\ &= \begin{pmatrix} 2 \\ 2 - 7/12 \end{pmatrix} \\ &= \begin{pmatrix} 2 \\ 17/12 \end{pmatrix} \simeq \begin{pmatrix} 2 \\ 1.4167 \end{pmatrix} \end{aligned}$$

Pour la deuxième itération, il nous suffit de reprendre la même procédure. On part de l'équation (3) et on utilise les valeurs de x^1 cette fois-ci :

$$\begin{pmatrix} 12 \cdot 2 & 0 \\ 0 & 36 \cdot (17/12)^2 \end{pmatrix} \begin{pmatrix} d_1^1 \\ d_2^1 \end{pmatrix} = - \begin{pmatrix} 6 \cdot 2^2 - 24 \\ 12 \cdot (17/12)^3 - 12 \end{pmatrix}$$

On voit tout de suite que $d_1^1 = 0$. Il nous faut, par contre, résoudre le système pour d_2^1 :

$$36 \cdot (17/12)^2 \cdot d_2^1 = - \left(12 \cdot (17/12)^3 - 12\right)$$

On le résout en isolant d_2^1 d'un côté de l'équation :

$$\begin{aligned} d_2^1 &= - \left(\frac{17^3}{12^2} \cdot \frac{12^2}{17^2} \cdot \frac{1}{36} - 12 \cdot \frac{12^2}{17^2} \cdot \frac{1}{36} \right) \\ &= - \left(\frac{17}{36} - \frac{12^3}{17^2} \cdot \frac{1}{36} \right) \simeq -0.3061 \end{aligned}$$

La direction vaut donc :

$$d^1 = \begin{pmatrix} 0 \\ -0.3061 \end{pmatrix}$$

On peut donc calculer la valeur de la deuxième itération :

$$\begin{aligned} x^2 &= x^1 + d^1 \\ &= \begin{pmatrix} 2 \\ 1.4167 \end{pmatrix} + \begin{pmatrix} 0 \\ -0.3061 \end{pmatrix} \\ &\simeq \begin{pmatrix} 2 \\ 1.1106 \end{pmatrix} \end{aligned}$$

Note : On voit qu'après deux itérations, l'algorithme tend à nous rapprocher du point $(2,1)^T$ qui est un minimum local. Si on continue plusieurs itérations de l'algorithme, on verra que $\lim_{k\to\infty} x^k = (2,1)^T$

Solution de la question 2:

1. Afin de calculer le gradient de la fonction f aux points a et b, il nous faut d'abord calculer les dérivées partielles par rapport à x_1 et x_2 :

$$\begin{aligned}
\frac{\partial f}{\partial x_1}(x) &= 7 \cdot 2\left(x_1^2 - x_2\right) \cdot 2x_1 + 3\left(x_1 - x_2\right)^2 \\
&= 28x_1\left(x_1^2 - x_2\right) + 3\left(x_1 - x_2\right)^2 \\
\frac{\partial f}{\partial x_2}(x) &= 7 \cdot 2\left(x_1^2 - x_2\right) \cdot (-1) + 3\left(x_1 - x_2\right)^2 \cdot (-1) \\
&= -14\left(x_1^2 - x_2\right) - 3\left(x_1 - x_2\right)^2
\end{aligned}$$

Le gradient vaut donc :

$$\nabla f(x) = \begin{pmatrix} \dfrac{\partial f}{\partial x_1} \\ \dfrac{\partial f}{\partial x_2} \end{pmatrix}(x) = \begin{pmatrix} 28x_1\left(x_1^2 - x_2\right) + 3\left(x_1 - x_2\right)^2 \\ \\ -14\left(x_1^2 - x_2\right) - 3\left(x_1 - x_2\right)^2 \end{pmatrix}$$

On peut donc calculer la valeur de la fonction et le gradient aux points a et b.

Point $a = (0,0)^T$

$$\begin{aligned}
f(a) &= f\left(\begin{pmatrix} 0 \\ 0 \end{pmatrix}\right) = 0 \\
\nabla f(a) &= \nabla f\left(\begin{pmatrix} 0 \\ 0 \end{pmatrix}\right) = \begin{pmatrix} 0 \\ 0 \end{pmatrix}
\end{aligned}$$

Point $b = (-1,1)^T$

$$f(b) = f\left(\left(\begin{array}{c} -1 \\ 1 \end{array}\right)\right) = 7\,(1-1)^2 + (-1-1)^3 = -8$$

$$\nabla f(b) = \nabla f\left(\left(\begin{array}{c} -1 \\ 1 \end{array}\right)\right) = \begin{pmatrix} 28 \cdot (-1) \cdot (1-1) + 3\,(-1-1))^2 \\ -14\,(-1-1) - 3\,(-1-1)^2 \end{pmatrix}$$

$$= \begin{pmatrix} 3 \cdot (-2)^2 \\ -3 \cdot (-2)^2 \end{pmatrix} = \left(\begin{array}{c} 12 \\ -12 \end{array}\right)$$

2. La direction d est donnée par

$$d = a - b = \left(\begin{array}{c} 1 \\ -1 \end{array}\right)$$

Afin de savoir s'il s'agit d'une direction de descente en un point donné, il faut regarder si la direction en ce point vérifie la condition suivante :

$$\nabla f(x)^T \cdot d < 0$$

Si $\nabla f(x)^T \cdot d > 0$, cela signifie que d est une direction de montée.

Dans notre cas, nous avons :

$$\nabla f(b)^T \cdot d = \begin{pmatrix} 12 & -12 \end{pmatrix} \cdot \begin{pmatrix} 1 \\ -1 \end{pmatrix} = 24 > 0$$

On voit donc que la direction d au point b est une direction de montée, et non pas une direction de descente.

Figure 1 montrent justement les courbes de niveaux de la fonction $f(x_1, x_2)$. Les points a et b y sont dessinés ainsi que la direction de montée d.

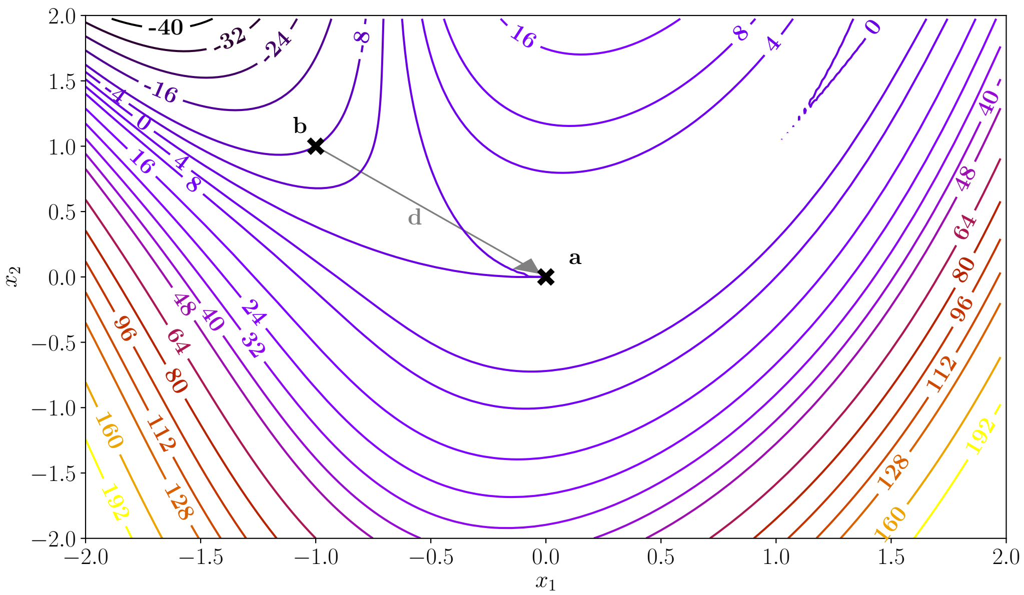

FIGURE 1 – Courbes de niveaux de la fonction $f(x_1, x_2)$. Les points a et b sont montrés ainsi que la direction d. On voit bien que la direction est une direction de montée.

Solution de la question 3:

Avant d'appliquer les algorithmes de la plus forte pente ou de Newton, commençons par calculer le gradient et la matrice hessienne[1] de la fonction f.

On a

$$\nabla f(x) = \begin{pmatrix} 6x_1 \\ 4x_2 \end{pmatrix} \tag{4}$$

$$\nabla^2 f(x) = \begin{pmatrix} 6 & 0 \\ 0 & 4 \end{pmatrix} \tag{5}$$

De plus, il peut être intéressant de voir que la valeur de la fonction va effectivement diminuer lors des itérations de ces deux algorithmes. Nous calculons

1. L'algorithme de la plus forte pente est un algorithme de premier ordre. Il n'utilise donc que le gradient. Par contre, la méthode de Newton est une méthode de deuxième ordre. Nous avons donc besoin de la matrice hessienne pour cet algorithme.

donc la valeur de la fonction au point x_0 :

$$f(x_0) = 147 + 8 = 155$$

On rappelle aussi que la première condition de Wolfe s'écrit :

$$f(x^k + \alpha_k d_k) \leq f(x^k) + \alpha_k \beta \nabla f(x_k)^T d_k \tag{6}$$

1. On commence par trouver la direction de la plus forte pente. Celle-ci est donnée par :

$$d_k = -\nabla f(x_k)$$

Au point x_0, on obtient :

$$d_0 = -\begin{pmatrix} 6x_1 \\ 4x_2 \end{pmatrix}$$

On a donc que :

$$x_{k+1} = x_k + \alpha_k d_k = \begin{pmatrix} x_1 \\ x_2 \end{pmatrix} - \alpha_k \begin{pmatrix} 6x_1 \\ 4x_2 \end{pmatrix} = \begin{pmatrix} (1-6\alpha_k)x_1 \\ (1-4\alpha_k)x_2 \end{pmatrix}.$$

Appliqué à x_0, on obtient :

$$x_1 = x_0 + \alpha_0 d_0 = \begin{pmatrix} 7 - 42\alpha_0 \\ 2 - 8\alpha_0 \end{pmatrix}.$$

On peut maintenant calculer la partie de gauche de la condition de Wolfe (6) :

$$\begin{aligned} f(x_1) &= f(x^0 + \alpha_0 d_0) \\ &= 3\,(7 - 42\alpha_0)^2 + 2\,(2 - 8\alpha_0)^2 \\ &= 5420\alpha_0^2 - 1828\alpha_0 + 155. \end{aligned}$$

Afin de calculer la partie droite de la condition de Wolfe (6), il nous faut d'abord identifier $\nabla f(x_k)^T d_k$:

$$\nabla f(x_k)^T d_k = -\nabla f(x_k)^T \nabla f(x_k) = -||\nabla f(x_k)||^2 = -\left(36x_1^2 + 16x_2^2\right).$$

Appliqué à x_0, on obtient :

$$\nabla f(x_0)^T d_0 = -\left(36 \cdot 7^2 + 16 \cdot 2^2\right) = -1828.$$

La condition de Wolfe, avec $\beta = 1/100$, s'écrit donc

$$5420\alpha_0^2 - 1828\alpha_0 + 155 \leq 155 - \frac{1828}{100}\alpha_0. \tag{7}$$

Il nous faut maintenant trouver la valeur de α_0 de sorte que cette inégalité soit vérifiée. En pratique, on commence avec $\alpha_0 = 1$ et si l'inégalité n'est pas vérifiée, on divise α_0 par 2. Et on continue à le diviser par 2 jusqu'à ce que l'inégalité soit vérifiée. On commence donc avec $\alpha_0 = 1$. On obtient :

$$3747 \overset{?}{\leq} 138.72.$$

On voit que la condition n'est pas vérifiée. En effet, la valeur de la fonction objectif a augmenté, passant de 155 à 3747. On divise donc le pas par 2. Pour $\alpha_0 = 1/2$, on a :

$$596 \overset{?}{\leq} 147.86.$$

On voit que la condition n'est toujours pas vérifée. Par contre l'augmentation de la valeur de la fonction objectif a diminué puisqu'elle passe de 155 à 596. On divise donc le pas une nouvelle fois par 2. Pour $\alpha_0 = 1/4$, on a :

$$36.75 \overset{?}{\leq} 152.43.$$

Cette fois, la condition de Wolfe est vérifiée. La valeur de la fonction objectif diminue, passant de 155 à 36.75. Le nouvel itéré vaut donc :

$$x_1 = \begin{pmatrix} (1 - 6/4)7 \\ (1 - 4/4)2 \end{pmatrix} = \begin{pmatrix} -3.5 \\ 0 \end{pmatrix}.$$

Note : La condition de Wolfe que l'on obtient pour ce problème (7) peut être résolue d'une autre manière. Celle que nous montrons est la méthode standard lorsque l'on implémente un algorithme avec la condition de Wolfe. Mais on peut aussi résoudre le problème à la main. En effet, on a que :

$$
\begin{aligned}
5420\alpha_0^2 - 1828\alpha_0 + \cancel{155} &\leq \cancel{155} - \frac{1828}{100}\alpha_0 \\
\Rightarrow \qquad 5420\alpha_0^{\cancel{2}} - 1828\cancel{\alpha_0} &\leq -\frac{1828}{100}\cancel{\alpha_0} \\
\Rightarrow \qquad 5420\alpha_0 &\leq 1828 + \frac{1828}{100} \\
\Rightarrow \qquad \alpha_0 &\leq \frac{99}{100}\frac{1828}{5420} \simeq 0.33.
\end{aligned}
$$

On voit bien que la valeur de α_0 trouvé précédemment fonctionne puisque $1/4 \leq 0.33$. Cependant, ce n'est pas le plus grand pas possible qui vérife la condition de Wolfe. On choisit donc $\alpha_0 = \alpha_0^* = 99/100 \, 1828/5420$ On pourrait donc avoir :

$$
\begin{aligned}
x_1 &= \begin{pmatrix} (1-\alpha_0^* \cdot 6)7 \\ (1-\alpha_0^* \cdot 4)2) \end{pmatrix} \simeq \begin{pmatrix} -7.023 \\ -0.671 \end{pmatrix} \\
f(x_1) &= 148.896.
\end{aligned}
$$

On voit ici qu'en utilisant un plus grand pas, la fonction objectif diminue très peu. En effet, la fonction étant convexe, on est en train d'osciller autour du minimum. C'est un problème connu de l'algorithme de la plus forte pente. On peut donc utiliser des algorithmes plus puissants, comme la méthode de Newton.

2. De la même manière que pour l'algorithme de la plus forte pente, il nous faut d'abord calculer la direction de descente. Elle est définie de la manière suivante :

$$
\begin{aligned}
d_k &= -\nabla^2 f(x_k)^{-1} \nabla f(x_k) \\
&= -\begin{pmatrix} 6 & 0 \\ 0 & 4 \end{pmatrix}^{-1} \begin{pmatrix} 6x_1 \\ 4x_2 \end{pmatrix} \\
&= -\begin{pmatrix} 1/6 & 0 \\ 0 & 1/4 \end{pmatrix} \begin{pmatrix} 6x_1 \\ 4x_2 \end{pmatrix} \\
&= -\begin{pmatrix} x_1 \\ x_2 \end{pmatrix}
\end{aligned}
$$

On peut directement calculer le pas suivant :

$$
x_{k+1} = x_k + \alpha_k d_k = x_k - \alpha_k x_k = (1-\alpha_k)\, x_k
$$

On calcule alors la partie gauche de la condition de Wolfe (6) :

$$
\begin{aligned}
f(x_{k+1}) &= f\left((1-\alpha_k)\, x_k\right) \\
&= 3\left((1-\alpha_k)\, x_1\right)^2 + 2\left((1-\alpha_k)\, x_2\right)^2 \\
&= (1-\alpha_k)^2 \cdot \left(3x_1^2 + 2x_2^2\right) \\
&= (1-\alpha_k)^2\, f(x_k)
\end{aligned}
$$

Pour la partie de droite, il nous faut encore calculer $\nabla f(x_k)^T d_k$:

$$
\nabla f(x_k)^T d_k = -\begin{pmatrix} 6x_1 & 4x_2 \end{pmatrix} \begin{pmatrix} x_1 \\ x_2 \end{pmatrix} = -\left(6x_1^2 + 4x_2^2\right)
$$

On peut maintenant appliquer le tout à x_0 et donner la condition de Wolfe :

$$
155\left(1-\alpha_0^2\right) \leq 155 - \alpha \frac{310}{100}
$$

Comme précédemment, on commence avec $\alpha_0 = 1$. On a :

$$
0 \leq 155.
$$

La valeur de la fonction objectif passe de 155 à 0. Le nouvel itéré x_1 vaut :

$$
x_1 = \begin{pmatrix} 0 \\ 0 \end{pmatrix}.
$$

On peut maintenant passer à la deuxième itération. On voit que la nouvelle direction ainsi que le gradient sont nuls. Cela signifie donc que nous avons atteint un extremum local qui s'avère être un minimum local. Il n'est donc pas nécessaire de continuer et la solution optimale est :

$$
x^* = \begin{pmatrix} 0 \\ 0 \end{pmatrix}
$$