Professeur : Michel Bierlaire, Assistant responsable : Nikola Obrenovic, Nourelhouda Dougui

Optimisation non linéaire – corrigé (7 décembre 2018)

Solution de la question 1:

a) Nous commençons par calculer le gradient $\nabla f(x_1, x_2)$ et la matrice hessienne $\nabla f^2(x_1, x_2)$ de la fonction $f(x_1, x_2)$. En utilisant les définitions du cours, on obtient :

$$\nabla f(x_1, x_2) = \begin{pmatrix} 4x_1^3 - 4x_1 \\ 3x_2^2 - 3 \end{pmatrix}$$

$$\nabla f^2(x_1, x_2) = \begin{pmatrix} 12x_1^2 - 4 & 0 \\ 0 & 6x_2 \end{pmatrix}$$

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** $(2, 2)^T$: Le gradient au point $(2, 2)^T$ vaut :

$$\nabla f(2, 2) = \begin{pmatrix} 24 \\ 9 \end{pmatrix}$$

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

— **Point** $(-1, 1)^T$ **:** Le gradient au point $(-1, 1)^T$ vaut :

$$\nabla f(-1, 1) = \begin{pmatrix} 0 \\ 0 \end{pmatrix}$$

Le gradient étant nul, on sait que le point $(-1, 1)^T$ est un point stationnaire. Afin de déterminer sa nature, il faut calculer la matrice hessienne :

$$\nabla f^2(-1, 1) = \begin{pmatrix} 8 & 0 \\ 0 & 6 \end{pmatrix}$$

La matrice hessienne est une matrice définie positive car ses valeurs propres sont strictement positives. Cela signifie donc que le point $(-1, 1)^T$ est un minimum local.

— **Point** $(0,-1)^T$ Le gradient au point $(0,-1)^T$ vaut :

$$\nabla f(0,-1) = \begin{pmatrix} 0 \\ 0 \end{pmatrix}$$

Le gradient étant nul, on sait que le point $(0,-1)^T$ est un point stationnaire. Afin de déterminer sa nature, on calcule la matrice hessienne

$$\nabla f^2(0,-1) = \begin{pmatrix} -4 & 0 \\ 0 & -6 \end{pmatrix}$$

La matrice hessienne est définie négative car ses valeurs propres sont strictement négatifs. Cela signifie donc que le point $(0,-1)^T$ est un maximum local.

b) 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 :

$$\nabla f^2(x^k)d^k = -\nabla f(x^k)$$

Il nous suffit alors de résoudre le système au point x^0 :

$$\begin{pmatrix} 44 & 0 \\ 0 & 12 \end{pmatrix} \begin{pmatrix} d_1^0 \\ d_2^0 \end{pmatrix} = - \begin{pmatrix} 24 \\ 9 \end{pmatrix}$$

En résolvant ce système, on obtient :

$$d^0 = \begin{pmatrix} -6/11 \\ -3/4 \end{pmatrix}$$

On peut donc calculer la valeur de la première itération :

$$x^1 = x^0 + d^0 = \begin{pmatrix} 2 \\ 2 \end{pmatrix} + \begin{pmatrix} -6/11 \\ -3/4 \end{pmatrix} = \begin{pmatrix} 16/11 \\ 5/4 \end{pmatrix}$$

Pour la deuxième itération, il nous suffit de reprendre la même procédure. On résout à présent le système pour le point x_1 :

$$\begin{pmatrix} 21.388 & 0 \\ 0 & 15/2 \end{pmatrix} \begin{pmatrix} d_1^1 \\ d_2^1 \end{pmatrix} = - \begin{pmatrix} 6.491 \\ 1.687 \end{pmatrix}$$

En résolvant ce système, on obtient :

$$d^1 = \begin{pmatrix} -0.3035 \\ -0.225 \end{pmatrix}$$

On peut donc calculer la valeur de la deuxième itération :

$$x^2 = x^1 + d^1 = \begin{pmatrix} 16/11 \\ 5/4 \end{pmatrix} + \begin{pmatrix} -0.3035 \\ -0.225 \end{pmatrix} = \begin{pmatrix} 1.151 \\ 1.025 \end{pmatrix}$$

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

Solution de la question 2:

1. The starting point is $x^0 = (-1, 0)^T$. We know that :

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

where d^k is the direction that the algorithm must take. We can find this direction by resolving the following system :

$$\nabla^2 f(x^k) d^k = -\nabla f(x^k).$$

Therefore, we now need to calculate the gradient and the Hessian.

$$\nabla f(x_1, x_2) = \begin{pmatrix} -2x_1 + 4x_2^2 \\ 8x_1x_2 - 2x_2 \end{pmatrix}$$

$$\nabla^2 f(x_1, x_2) = \begin{pmatrix} -2 & 8x_2 \\ 8x_2 & 8x_1 - 2 \end{pmatrix}.$$

Now we resolve the system $\nabla^2 f(x^k) d^k = -\nabla f(x^k)$ for point x^0. We obtain :

$$\begin{pmatrix} -2 & 0 \\ 0 & -10 \end{pmatrix} \begin{pmatrix} d_1^0 \\ d_2^0 \end{pmatrix} = - \begin{pmatrix} 2 \\ 0 \end{pmatrix}$$

Solving this system, we obtain :

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

We can now calculate the value of the first iteration :

$$\begin{aligned} x^1 &= x^0 + d^0 \\ &= \begin{pmatrix} -1 \\ 0 \end{pmatrix} + \begin{pmatrix} 1 \\ 0 \end{pmatrix} \\ &= \begin{pmatrix} 0 \\ 0 \end{pmatrix} \end{aligned}$$

Note : Here the algorithm converges in one iteration to the point $(0,0)^T$ which is a stationary point. However, this point doesn't satisfy the second-order necessary optimality conditions and isn't consequently a local minimum. Actually, the hessian matrix is definite negative which means that $(0,0)^T$ is a local maximum.

2. Given a function which is twice differentiable and given point x_k, the Cauchy point of f in x_k is the point x_C that minimizes the quadratic model of f in the direction with the steepest decent,

$$x_C = x_k - \alpha_C \nabla f(x_k),$$

where

$$\alpha_C \in \text{argmin}_{\alpha \in \mathbb{R}_0^+} m_{x_k}(x_k - \alpha \nabla f(x_k)).$$

We are going to compute α_C :

$$m_{x^0}(x^0 - \alpha \nabla f(x^0)) = f(x^0) - \alpha \nabla^T f(x^0) \nabla f(x^0) + \frac{1}{2}\alpha^2 \nabla^T f(x^0) \nabla^2 f(x^0) \nabla f(x^0)$$

$$m_{x^0}(x^0 - \alpha \nabla f(x^0)) = -1 - \alpha \begin{pmatrix} 2 & 0 \end{pmatrix} \begin{pmatrix} 2 \\ 0 \end{pmatrix} + \frac{1}{2}\alpha^2 \begin{pmatrix} 2 & 0 \end{pmatrix} \begin{pmatrix} -2 & 0 \\ 0 & -10 \end{pmatrix} \begin{pmatrix} 2 \\ 0 \end{pmatrix}$$

$$m_{x^0}(x^0 - \alpha \nabla f(x^0)) = -1 - 4\alpha - 4\alpha^2$$

Since this function is concave, there is no $\alpha \in \mathbb{R}_0^+$ that minimizes the function. Consequently, there is no Cauchy point of f in x^0.

Solution de la question 3:

1. We are looking for the change of variables

$$x' = L_k^T x,$$

where L_k^T is such that the Hessian matrix at the step k is $H_k^T = L_k L_k^T$. It corresponds to its cholesky decomposition.
Here the Hessian matrix at step k is

$$H_k^T = \begin{pmatrix} 2 & 1 \\ 1 & 22 \end{pmatrix}.$$

We want to find a, b, c such that

$$\begin{pmatrix} 2 & 1 \\ 1 & 22 \end{pmatrix} = \begin{pmatrix} a & 0 \\ b & c \end{pmatrix} \begin{pmatrix} a & b \\ 0 & c \end{pmatrix} = \begin{pmatrix} a^2 & ab \\ ab & b^2 + c^2 \end{pmatrix}.$$

Thus we get

$$a = \sqrt{2}, \quad b = \frac{1}{\sqrt{2}}, \quad c = \sqrt{\frac{43}{2}}.$$

The change of variables is thus

$$x' = L_k^T x = \begin{pmatrix} \sqrt{2}x_1 + \frac{1}{\sqrt{2}}x_2 \\ \sqrt{\frac{43}{2}}x_2 \end{pmatrix}.$$

The function becomes

$$\begin{aligned} \bar{f}(x') &= (\sqrt{2}x_1 + \frac{1}{\sqrt{2}}x_2)^2 + 11(\sqrt{\frac{43}{2}}x_2)^2 + (\sqrt{2}x_1 + \frac{1}{\sqrt{2}}x_2)\sqrt{\frac{43}{2}}x_2 \\ &= 2x_1^2 + 2x_1x_2 + \frac{x_2^2}{2} + 236.5x_2^2 + \sqrt{43}x_1x_2 + \frac{\sqrt{43}}{2}x_2^2 \\ &= 2x_1^2 + (2 + \sqrt{43})x_1x_2 + (237 + \frac{\sqrt{43}}{2})x_2^2 \end{aligned}$$

2. The direction for the first step is defined as

$$d_0 = -H_0^{-1}\nabla f(x_0),$$

with

$$H_0^{-1} = \frac{1}{43}\begin{pmatrix} 22 & -1 \\ -1 & 2 \end{pmatrix}.$$

Thus we get

$$\begin{aligned} d_0 &= -H_0^{-1}\nabla f(x_0) \\ &= -\frac{1}{43}\begin{pmatrix} 22 & -1 \\ -1 & 2 \end{pmatrix}\begin{pmatrix} 2*4+1 \\ 22*1+4 \end{pmatrix} \\ &= -\frac{1}{43}\begin{pmatrix} 22 & -1 \\ -1 & 2 \end{pmatrix}\begin{pmatrix} 9 \\ 26 \end{pmatrix} \\ &= \begin{pmatrix} -4 \\ -1 \end{pmatrix} \end{aligned}$$

Solution de la question 4:

We have

$$\nabla f(x) = \begin{pmatrix} 2x_1 \\ 4x_2 \end{pmatrix}$$

and

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

and

$$f(x^0) = 81 + 2 = 83.$$

1. When applying the steepest descent algorithm starting from x^0, the next iterate is :

$$x^1 = \begin{pmatrix} x_1^0 \\ x_2^0 \end{pmatrix} - \alpha_0 \begin{pmatrix} 2x_1^0 \\ 4x_2^0 \end{pmatrix} = \begin{pmatrix} x_1^0 - 2\alpha_0 x_1^0 \\ x_2^0 - 4\alpha_0 x_2^0 \end{pmatrix} = \begin{pmatrix} 9 - 18\alpha_0 \\ 1 - 4\alpha_0 \end{pmatrix}.$$

The value of the function in that iterate x^0 is :

$$f(x^1) = (9 - 18\alpha_0)^2 + 2(1 - 4\alpha_0)^2 = 356\alpha_0^2 - 340\alpha_0 + 83.$$

As $d_k = -\nabla f(x^k)$, we have

$$\nabla f(x^0)^T d_0 = -(18^2 + 4^2) = -340.$$

Recall that the first Wolfe condition is defined as

$$f(x^k + \alpha_k d_k) \leq f(x^k) + \alpha_k \beta_1 \nabla f(x^k)^T d_k.$$

Thus, here it is written as :

$$356\alpha_k^2 - 340\alpha_k + 83 \leq 83 - \alpha_k \frac{1}{100} 340.$$

If we start the line search algorithm with $\alpha_0 = 1$, we get

$$99 \leq 79.6$$

and the first Wolfe condition is not satisfied. In fact, the value of the function increases from 83 to 99.

The step α_0 being too long, we define the step $\alpha_1 = \frac{\alpha_l + \alpha_0}{2}$ where $\alpha_l = 0$. Thus, $\alpha_1 = \frac{1}{2}$, and the first wolf condition becomes

$$2 \leq 81.3.$$

This time, the condition is satisfied.
Lets check whether this step $\alpha_1 = \frac{1}{2}$ satisfy the second Wolfe condition.

The second Wolfe condition is :

$$\frac{\nabla f(x^k + \alpha_k d_k)^T d_k}{\nabla f(x^k)^T d_k} \leq \beta_2.$$

Here we have

$$\begin{aligned}
\frac{\nabla f(x^0 + \alpha_1 d_0)^T d_0}{\nabla f(x^0)^T d_0} &= \frac{\nabla f((9 - 0.5 * 18, 1 - 0.5 * 4))^T \begin{pmatrix} -18 \\ -4 \end{pmatrix}}{\nabla f((9,1))^T \begin{pmatrix} -18 \\ -4 \end{pmatrix}} \\
&= \frac{\begin{pmatrix} 0 & -4 \end{pmatrix} \begin{pmatrix} -18 \\ -4 \end{pmatrix}}{\begin{pmatrix} 18 & 4 \end{pmatrix} \begin{pmatrix} -18 \\ -4 \end{pmatrix}} \\
&= \frac{16}{340} \\
&= 0.047 \\
&\leq 0.99 = \beta_2.
\end{aligned}$$

Thus, the second Wolfe condition is satisfied and the step α_1 is valid. We get a sufficient decrease from 83 to 2. The new iterate is then :

$$x^1 = \begin{pmatrix} 0 \\ -1 \end{pmatrix}.$$

2. Recall that Newton's direction is defined as

$$d_k = -\nabla^2 f(x^k)^{-1} \nabla f(x^k).$$

In that case, since the Hessian matrix is positive definite, we don't need to apply the modified Cholesky factorization algorithm. Newton's direction corresponds to :

$$d_k = -\begin{pmatrix} 1/2 & 0 \\ 0 & 1/4 \end{pmatrix} \begin{pmatrix} 2x_1^k \\ 4x_2^k \end{pmatrix} = -\begin{pmatrix} x_1^k \\ x_2^k \end{pmatrix}.$$

One iteration of Newton's method is :

$$\begin{aligned} x^{k+1} &= x^k + \alpha_k \cdot d^k \\ &= \begin{pmatrix} x_1^k \\ x_2^k \end{pmatrix} + \alpha_k d^k \\ &= (1-\alpha_k) \begin{pmatrix} x_1^k \\ x_2^k \end{pmatrix}. \end{aligned}$$

Starting in $x^0 = (9,1)$, we have :

$$x^1 = (1-\alpha_0) \begin{pmatrix} x_1^0 \\ x_2^0 \end{pmatrix} = \begin{pmatrix} 9(1-\alpha_0) \\ (1-\alpha_0) \end{pmatrix}$$

and the value of the function in that iterate is :

$$f(x^1) = 81(1-\alpha_0)^2 + 2(1-\alpha_0)^2 = 83(1-\alpha_0)^2.$$

Furthermore,

$$\nabla f(x^0)^T d_0 = \begin{pmatrix} 18 & 4 \end{pmatrix} \begin{pmatrix} -9 \\ -1 \end{pmatrix} = -166.$$

The first Wolfe condition is thus written as

$$83(1-\alpha_0)^2 \leq 83 - \alpha_0 \frac{1}{100} 166.$$

It is satisfied for $\alpha_0 = 1$.

Futhermore, we have

$$\nabla f(x^0 + \alpha_0 d_0)^T d_0 = -9(18 - 18\alpha_0) - (4 - 4\alpha_0) = 166\alpha_0 - 166.$$

The second Wolfe condition is written as :

$$\frac{166\alpha_0 - 166}{-166} = -\alpha_0 + 1 \leq 0.99,$$

which is satisfied for $\alpha_0 = 1$.

The value of the function decreases from 83 to 0. The new iterate is :

$$x^1 = \begin{pmatrix} 0 \\ 0 \end{pmatrix}.$$

However, this is the optimal solution of the problem : we do not need any additional iteration.