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

Optimisation non linéaire – corrigé (15 décembre 2017)

Solution de la question 1:

Afin de calculer le pas de la méthode de Newton, il nous faut calculer le gradient et la matrice hessienne de la fonction $f(x)$. En effet, on rappelle que la direction de Newton est donnée par :

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

Le gradient est donc donné par :

$$\nabla f(x^k) = \begin{pmatrix} 2x_1^k - 4x_2^k \\ -4x_1^k + 10x_2^k \end{pmatrix}$$

La matrice hessienne est, quant à elle, donnée par :

$$\nabla^2 f(x^k) = \begin{pmatrix} 2 & -4 \\ -4 & 10 \end{pmatrix}$$

On peut alors calculer l'inverse de la matrice hessienne :

$$\left(\nabla^2 f(x^k)\right)^{-1} = \begin{pmatrix} 5/2 & 1 \\ 1 & 1/2 \end{pmatrix}$$

Finalement, il est possible de calculer la direction de Newton :

$$\begin{aligned} d_k &= -\left(\nabla^2 f(x^k)\right)^{-1} \nabla f(x^k) \\ &= -\begin{pmatrix} 5/2 & 1 \\ 1 & 1/2 \end{pmatrix} \begin{pmatrix} 2x_1^k - 4x_2^k \\ -4x_1^k + 10x_2^k \end{pmatrix} \\ &= -\begin{pmatrix} x_1 \\ x_2 \end{pmatrix} \\ &= -x^k \end{aligned}$$

Ainsi, lorsque nous calculons $x^{k+1} \in \mathbb{R}^2$ à partir de n'importe quel point $x^k \in \mathbb{R}^2$, on obtient :

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

On voit donc qu'on atteint une certaine solution. Afin de vérifier que cette solution est bien la solution optimale du problème, il faut vérifier que le gradient est nul. On a bien que :

$$\nabla f\left(x^{k+1}\right) = \nabla f(0) = 0$$

De plus, la matrice hessienne est définie positive. Pour voir cela, il nous suffit d'appliquer la définition `B.9` du livre, page 690 :

$$\begin{aligned}
x^T \nabla^2 f(x) x &= (x_1, x_2) \begin{pmatrix} 5/2 & 1 \\ 1 & 1/2 \end{pmatrix} \begin{pmatrix} x_1 \\ x_2 \end{pmatrix} \\
&= (2x_1 - 4x_2, -4x_1 + 10x_2) \begin{pmatrix} x_1 \\ x_2 \end{pmatrix} \\
&= 2x_1^2 - 4x_1x_2 - 4x_1x_2 + 10x_2^2 \\
&= 2\left(x_1^2 - 4x_1x_2 + 4x_2^2\right) + 2x_2^2 \\
&= 2\left(x_1 - 2x_2\right)^2 + 2x_2^2
\end{aligned}$$

On voit que $x^T \nabla^2 f(x) x$ est équivalent à une somme de carré. On a donc :

$$x^T \nabla^2 f(x) x > 0, \quad \forall x \in \mathbb{R}^2, x \neq 0$$

Puisque la matrice hessienne est définie positive, cela nous permet donc de dire que nous avons atteint une solution optimale et que cette solution optimale est minimale.

Solution de la question 2:

Cet exercice demande qu'on utilise la direction de Newton définie par :

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

Il nous faut donc calculer le gradient et la matrice hessienne de la fonction $f(x)$. Le gradient est donné par :

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

La matrice hessienne est donc donnée par :

$$\nabla^2 f(x^k) = \begin{pmatrix} 8 & 0 \\ 0 & 2 \end{pmatrix}$$

Son inverse peut être facilement calculer :

$$\left(\nabla^2 f(x^k)\right)^{-1} = \begin{pmatrix} 1/8 & 0 \\ 0 & 1/2 \end{pmatrix}$$

1. On peut maintenant calculer la direction de Newton en utilisant le gradient et la matrice hessienne calculés auparavant :

$$\begin{aligned} d_k &= -\left(\nabla^2 f(x^k)\right)^{-1} \nabla f(x^k) \\ &= -\begin{pmatrix} 1/8 & 0 \\ 0 & 1/2 \end{pmatrix} \begin{pmatrix} 8x_1^k - 4 \\ 2x_2^k + 2 \end{pmatrix} \\ &= -\begin{pmatrix} x_1^k - 1/2 \\ x_2^k + 1 \end{pmatrix} \end{aligned}$$

   On rappelle maintenant les deux conditions de Wolfe, dans l'ordre :

$$f(x^k + \alpha_k d_k) \leq f(x^k) + \alpha_k \beta_1 \nabla f(x^k)^T d_k \tag{1}$$

$$\frac{\nabla f(x^k + \alpha_k d_k)^T d_k}{\nabla f(x^k)^T d_k} \leq \beta_2 \tag{2}$$

   Le point de départ $x^0 = (0,0)^T$ nous est donné. On peut alors calculer la prochaine position, donnée par :

$$x^1 = x^0 + \alpha_0 d_0$$

   Notre but est donc de trouver les bornes du pas α_0 en utilisant ces deux conditions, sachant que la première condition donne une borne supérieure et la deuxième une borne inférieure, afin que la position x^1 soit admissible. Il faut donc étudier ces bornes séparement.

   (a) **Première condition de Wolfe**
   On peut commencer par calculer la valeur de f et de son gradient au point x^0 ainsi que la direction d_0 :

$$\begin{aligned} f(x^0) &= 0 \\ \nabla f(x^0) &= \begin{pmatrix} -4 \\ 2 \end{pmatrix} \\ d_0 &= \begin{pmatrix} 1/2 \\ -1 \end{pmatrix} \end{aligned}$$

On peut donc calculer la valeur de la position après le premier pas :

$$\begin{aligned} x^1 &= x^0 + \alpha_0 d_0 \\ &= \begin{pmatrix} 0 \\ 0 \end{pmatrix} + \alpha_0 \begin{pmatrix} 1/2 \\ -1 \end{pmatrix} \\ &= \begin{pmatrix} \alpha_0/2 \\ -\alpha_0 \end{pmatrix} \end{aligned}$$

Il ne nous reste plus qu'à calculer la valeur de f au point x_1 :

$$f(x^1) = 2\alpha_0^2 - 4\alpha_0$$

Si on étudie la valeur de la fonction à la prochaine itération, on voit que le pas α_0 doit prendre une valeur plus petite que 2 afin d'obtenir une diminution de la fonction objectif, *i.e.* $f(x^1) < f(x^0)$. Figure 1 montre la fonction $f(x^1)$ en fonction du paramètre α_0. On voit bien que pour avoir une diminution de la fonction, *i.e.* que la fonction soit négative, α_0 doit être plus grand que 0 et plus petit que 2.

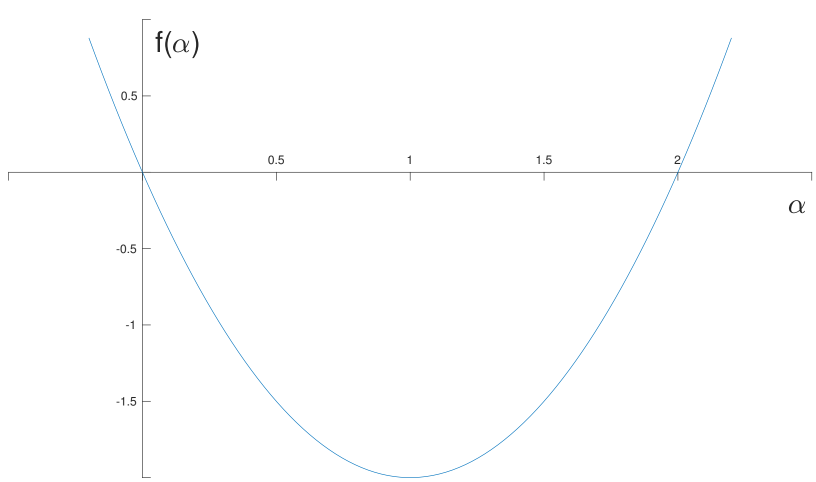

FIGURE 1 – Fonction $f(x^1)$ en fonction du pas α_0.

La première condition de Wolfe nous donne donc :

$$
\begin{array}{rrcl}
 & f(x^0+\alpha_0 d_0) & \leq & f(x^0)+\alpha_0\beta_1\nabla f(x^0)^T d_0 \\
\Rightarrow & 2\alpha_0^2-4\alpha_0 & \leq & 0+\alpha_0\beta_1(-4,2)\begin{pmatrix} 1/2 \\ -1 \end{pmatrix} \\
\Rightarrow & 2\alpha_0^{\cancel{2}}-4\cancel{\alpha_0} & \leq & -4\cancel{\alpha_0}\beta_1 \\
\Rightarrow & 2\alpha_0-4 & \leq & -4\beta_1 \\
\Rightarrow & \alpha_0 & \leq & 2(1-\beta_1)
\end{array}
$$

Si on utilise $\beta_1 = 10^{-4}$, on obtient alors que :

$$\alpha_0 \leq 1.9998$$

Comme attendu, α_0 doit être, au maximum, légèrement plus petit que 2.

(b) **Deuxième condition de Wolfe**
Afin de calculer la deuxième condition de Wolfe, il nous manque la valeur du gradient au point x^1 :

$$
\begin{aligned}
\nabla f(x^0+\alpha_0 d_0) &= \nabla f(x^1) \\
&= \begin{pmatrix} 8x_1^1-4 \\ 2x_2^1+2 \end{pmatrix} \\
&= \begin{pmatrix} 4(\alpha_0-1) \\ 2(1-\alpha_0) \end{pmatrix}
\end{aligned}
$$

On peut maintenant calculer le numérateur et le dénominateur de la deuxième condition de Wolfe :

$$
\begin{aligned}
\nabla f(x^0+\alpha_0 d_0)^T d_0 &= (4(\alpha_0-1), 2(1-\alpha_0))\begin{pmatrix} 1/2 \\ -1 \end{pmatrix} \\
&= 2(\alpha_0-1)-2(1-\alpha_0) \\
&= 4(\alpha_0-1)
\end{aligned}
$$

$$
\begin{aligned}
\nabla f(x^0)^T d_0 &= (-4,2)\begin{pmatrix} 1/2 \\ -1 \end{pmatrix} \\
&= -2-2=-4
\end{aligned}
$$

La deuxième condition de Wolfe vaut donc :

$$\begin{aligned} & \frac{\nabla f(x^k + \alpha_k d_k)^T d_k}{\nabla f(x^k)^T d_k} \leq \beta_2 \\ \Rightarrow \quad & \frac{4(\alpha_0 - 1)}{-4} \leq \beta_2 \\ \Rightarrow \quad & -\alpha_0 + 1 \leq \beta_2 \\ \Rightarrow \quad & \alpha_0 \geq 1 - \beta_2 \end{aligned}$$

Si on utilise $\beta_2 = 0.99$, on obtient alors que :

$$\alpha_0 \geq 0.01$$

Grâce aux deux conditions de Wolfe, on peut maintenant trouver une valeur pour α_0, les bornes étant :

$$0.01 \leq \alpha_0 \leq 1.9998$$

Figure 2 montre à nouveau la fonction $f(x^1)$. Mais cette fois-ci nous avons marqué les deux bornes pour α_0. Si la borne supérieure est choisie comme valeur pour α_0, la fonction va diminuer que très légèrement. En effet, on voit que $f(x^1) = -0.00079992$ pour $\alpha = 1.9998$.

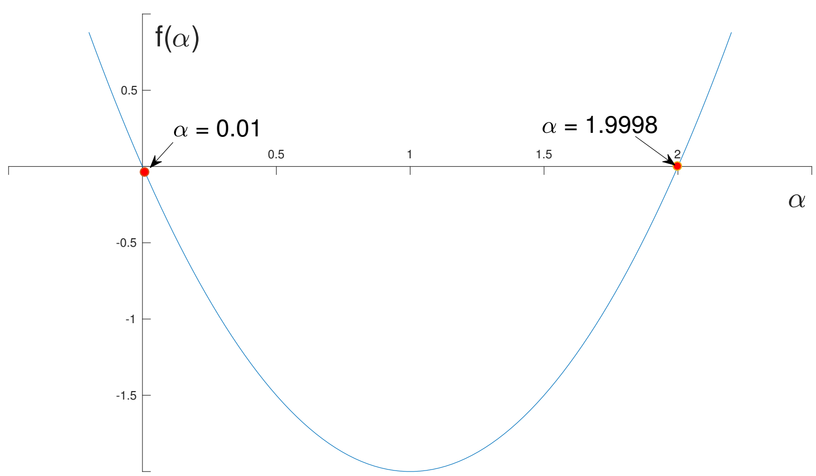

Figure 2 – Fonction $f(x^1)$ en fonction du pas α_0. Les bornes du paramètre α_0 sont marquées sur le graphe.

2. Comme la direction d nous est déjà donnée, on peut déjà calculer le prochain pas. On note juste que $x^k = (0,0)^T$:

$$\begin{aligned} x^{k+1} &= x^k + \alpha_k d \\ &= \begin{pmatrix} 0 \\ 0 \end{pmatrix} + \alpha_k \begin{pmatrix} -1 \\ 2 \end{pmatrix} \\ &= \begin{pmatrix} -\alpha_k \\ 2\alpha_k \end{pmatrix} \end{aligned}$$

   Notre but est de trouver α_k afin que la fonction f diminue. Au point x^k, la fonction objectif est nulle. Il nous faut donc trouver α_k afin que $f(x^{k+1})$ soit négative. Si on calcule $f(x^{k+1})$, on obtient :

$$\begin{aligned} f(x^{k+1}) &= 4(-\alpha_k)^2 - 4(-\alpha_k) + (2\alpha_k)^2 + 2(2\alpha_k) \\ &= 4\alpha_k^2 + 4\alpha_k + 4\alpha_k^2 + 4\alpha_k \\ &= 8\alpha_k^2 + 8\alpha_k \end{aligned}$$

   Etant donné que le pas α_k ne peut pas être négatif, la fonction $f(x^{k+1})$ ne peut qu'augmenter. Il n'est donc pas possible de calculer une valeur du pas α_k minimisant la fonction f dans la direction d.

Solution de la question 3:

1. La matrice hessienne de f est donnée par Q pour tout $x \in \mathbb{R}^3$. Q étant une matrice diagonale avec des éléments diagonaux strictement positifs, on en conclut qu'elle est définie positive et que, par conséquent, f est strictement convexe sur $\mathbb{R}^3$.

   L'unique minimum de f sur $\mathbb{R}^3$ est donc donné par la solution unique du système d'équations $\nabla f(x) = 0$, i.e. $Qx = -b$, qui peut se réécrire comme suit :

$$\begin{cases} 2x_1 = -17 \\ 4x_2 = -13 \\ 8x_3 = -23 \end{cases}$$

   On trouve alors :

$$\begin{cases} x_1^* = -17/2 \\ x_2^* = -13/4 \\ x_3^* = -23/8 \end{cases}$$

L'unique minimum de f sur $\mathbb{R}^3$ est donc :

$$x^* = (x_1^* \; x_2^* \; x_3^*)^T = \begin{pmatrix} -17/2 \\ -13/4 \\ -23/8 \end{pmatrix}$$

2. Le gradient de f est donné par :

$$\nabla f(x) = \begin{pmatrix} 2x_1 + 17 \\ 4x_2 + 13 \\ 8x_3 + 23 \end{pmatrix} \qquad \forall x \in \mathbb{R}^3$$

La direction de la plus forte pente pour f en x^0 est par définition :

$$d_0 = -\nabla f(x^0) = \begin{pmatrix} -17 \\ -13 \\ -23 \end{pmatrix}$$