Optimisation algorithmique - feuille de TD 1 - Correction

Exercice 1.

Exercice 2. Un réel $\alpha > 0$ étant fixé, on définit

$$\forall x \in \mathbb{R}^n, \quad J_\alpha(x) = \sum_{i=2}^{n-1} N_\alpha(x_{i+1} + x_{i-1} - 2x_i), \quad \text{où} \quad N_\alpha(t) = \sqrt{\alpha + t^2}.$$

a) Prouver que J_α est différentiable et calculer, pour tout $x, h \in \mathbb{R}^n$ l'expression $DJ_\alpha(x).h$ de trois façons différentes:

- en calculant $\frac{\partial J_\alpha}{\partial x_j}$ pour tout j;
- en calculant les dérivées directionnelles de J_α
- en calculant directement la forme différentielle DJ_α.

b) Écrire une fonction `gradientJ(x,alpha)` qui prend en argument un vecteur colonne `x`, un réel strictement positif `alpha`, et renvoie le gradient de J_α en `x` sous forme de vecteur colonne.

Correction

a) - Il peut être plus simple d'écrire J_α sous forme développée :

$$\begin{aligned} J_\alpha(x) &= N_\alpha(x_3 + x_1 - 2x_2) + N_\alpha(x_4 + x_2 - 2x_3) + N_\alpha(x_5 + x_3 - 2x_4) + N_\alpha(x_6 + x_4 - 2x_3) \\ &\quad + N_\alpha(x_7 + x_5 - 2x_6) + \cdots + N_\alpha(x_{n-3} + x_{n-5} - 2x_{n-4}) \\ &\quad + N_\alpha(x_{n-2} + x_{n-4} - 2x_{n-3}) + N_\alpha(x_{n-1} + x_{n-3} - 2x_{n-2}) + N_\alpha(x_n + x_{n-2} - 2x_{n-1}). \end{aligned}$$

Les premières dérivées partielles s'écrivent donc :

$$\begin{aligned} \frac{\partial J_\alpha(x)}{\partial x_1} &= N'_\alpha(x_3 + x_1 - 2x_2), \\ \frac{\partial J_\alpha(x)}{\partial x_2} &= -2N'_\alpha(x_3 + x_1 - 2x_2) + N'_\alpha(x_4 + x_2 - 2x_3) \\ \frac{\partial J_\alpha(x)}{\partial x_3} &= N'_\alpha(x_3 + x_1 - 2x_2) - 2N'_\alpha(x_4 + x_2 - 2x_3) + N'_\alpha(x_5 + x_3 - 2x_4) \\ \frac{\partial J_\alpha(x)}{\partial x_4} &= N'_\alpha(x_4 + x_2 - 2x_3) - 2N'_\alpha(x_5 + x_3 - 2x_4) + N'_\alpha(x_6 + x_4 - 2x_3) \\ &\cdots \end{aligned}$$

avec $N'_\alpha(t) = t/\sqrt{\alpha + t^2}$. Pour les dérivées partielles suivantes, jusqu'à $\frac{\partial J_\alpha(x)}{\partial x_{n-2}}$, il y a également trois termes. La formule générale est

$$\frac{\partial J_\alpha(x)}{\partial x_i} = N'_\alpha(x_i + x_{i-2} - 2x_{i-1}) - 2N'_\alpha(x_{i+1} + x_{i-1} - 2x_i) + N'_\alpha(x_{i+2} + x_i - 2x_{i+1}),$$

ceci étant donc valide pour $3 \leq i \leq n-2$. Pour les derniers nous avons

$$\begin{aligned}
\frac{\partial J_\alpha(x)}{\partial x_{n-1}} &= N'_\alpha(x_{n-1} + x_{n-3} - 2x_{n-2}) - 2N'_\alpha(x_n + x_{n-2} - 2x_{n-1}) \\
\frac{\partial J_\alpha(x)}{\partial x_n} &= N'_\alpha(x_n + x_{n-2} - 2x_{n-1})
\end{aligned}$$

Pour retrouver l'expression $DJ_\alpha(x).h$ il suffit ensuite d'écrire

$$DJ_\alpha(x).h = \sum_{i=1}^{n} \frac{\partial J_\alpha(x)}{\partial x_i} h_i.$$

- Pour calculer la dérivée directionnelle on pose

$$f(t) = J_\alpha(x + th),$$

où $x, h \in \mathbb{R}^n$ et $t \in \mathbb{R}$. Nous devons calculer $f'(0)$. Nous avons

$$\begin{aligned}
f(t) &= J_\alpha(x + th) = \sum_{i=2}^{n-1} N_\alpha((x_{i+1} + th_{i+1}) + (x_{i-1} + th_{i-1}) - 2(x_i + th_i)) \\
&= \sum_{i=2}^{n-1} N_\alpha((x_{i+1} + x_{i-1} - 2x_i) + t(h_{i+1} + h_{i-1} - 2h_i)),
\end{aligned}$$

dont la dérivée est

$$f'(t) = \sum_{i=2}^{n-1} N'_\alpha((x_{i+1} + x_{i-1} - 2x_i) + t(h_{i+1} + h_{i-1} - 2h_i))(h_{i+1} + h_{i-1} - 2h_i).$$

Donc,

$$f'(0) = DJ_\alpha(x).h = \sum_{i=2}^{n-1} N'_\alpha(x_{i+1} + x_{i-1} - 2x_i)(h_{i+1} + h_{i-1} - 2h_i).$$

remarque : il n'estpas évident de voir que cette formule égale bien celle obtenue avec les dérivées partielles. Pour le faire il faut séparer les somme précédente en trois, puis réindexer les sommes ($i+1 \to i$, $i-1 \to i$) avant de rassembler à nouveau les sommes, ce qui est fastidieux.

- On revient à la formule définissant la fonction :

$$J_\alpha(x) = \sum_{i=2}^{n-1} N_\alpha(x_{i+1} + x_{i-1} - 2x_i).$$

On obtient directement la forme différentielle en utilisant la règle de dérivation composée :

$$DJ_\alpha(x).h = \sum_{i=2}^{n-1} N'_\alpha(x_{i+1} + x_{i-1} - 2x_i)(h_{i+1} + h_{i-1} - 2h_i).$$

b) On suit simplement la dernière formule pour écrire le code de calcul du gradient :

```
def gradientJ(x,alpha):
    n = x.size
    G = np.zeros((n,1))
    for i in range(1,n-1):
        t = x[i+1] + x[i-1] - 2*x[i]
        c = t / np.sqrt(alpha+t**2)
        G[i+1] +=  c
        G[i-1] += c
        G[i] -= 2*c
    return G
```

Ce code peut être testé avec les commandes suivantes :

```
import numpy as np
x = np.random.randn(10,1)
alpha = 0.01
G = gradientJ(x,alpha)
print(G)
```

Exercice 3.

Exercice 4. Soit $f : \mathbb{R}^2 \to \mathbb{R}$ définie par :

$$f(x, y) = x^2 + y^3$$

Cette fonction admet-elle un extremum global sur $\mathbb{R}^2$? Admet-elle un extremum local ?

Correction

On peut ici répondre directement, sans avoir à étudier les points critiques de f : puisque la fonction $y \mapsto y^3$ est strictement croissante, on peut dire directement qu'il n'y a aucun extremum local (et donc a fortiori au extremum global) de f. En effet si (x, y) était un tel point, en considérant $(x, y + t)$ pour $t > 0$ aussi petit que l'on veut, la valeur $f(x, y + t) = x^2 + (y + t)^3 > x^2 + y^3 = f(x, y)$, donc (x, y) ne peut pas être un maximiseur local de f, et à l'inverse en considérant $(x, y-t)$ on a $f(x, y-t) < f(x, y)$ donc (x, y) ne peut pas être un minimiseur local.

Ceci conclut l'exercice, mais no peut aussi essayer d'y répondre via l'étude classique des points critiques et des valeurs propres de la hessienne. On a $\nabla f(x, y) = (2x, 3y^2)$, donc $\nabla f(x, y) = 0$ si et seulement si $x = 0$ et $y = 0$. Donc le seul point critique est $(0, 0)$. La hessienne s'écrit

$$Hf(x, y) = \begin{pmatrix} 2 & 0 \\ 0 & 6y \end{pmatrix}$$

donc

$$Hf(0,0) = \begin{pmatrix} 2 & 0 \\ 0 & 0 \end{pmatrix}$$

dont les valeurs propres sont 2 et 0. Comme une valeur propre est nulle et l'autre positive, on ne peut pas conclure directement : $(0,0)$ pourrait être un minimiseur local ou un point selle. Donc il faut utiliser d'autres méthodes pour conclure. En fait ici $f(0,t) = t^3$ est négatif pour $t < 0$ et positif pour $t > 0$, ce qui implique que $(0,0)$ est un point selle. Le seul point critique de f étant un point selle, f n'admet donc pas d'extremum local et donc encore moins d'extremum global.

Exercice 5. Etudier les extrema locaux et globaux des fonctions suivantes de $\mathbb{R}^2$ dans $\mathbb{R}$:

1. $f_1(x,y) = xy(x+y-1)$
2. $f_2(x,y) = 4xy + \frac{1}{x} + \frac{1}{y}$
3. $f_3(x,y) = x^3 + y^3 - 9xy + 27$
4. $f_4(x,y) = x^2 - 3x^2y + 2x^4$
5. $f_5(x,y) = (y^2 - x^2)(y^2 - 2x^2)$
6. $f_6(x,y) = (x+y)^2 - (x^4 + y^4)$
7. $f_7(x,y) = x \ln y + y \ln x$

Correction

1. $f_1(x,y) = xy(x+y-1) = x^2y + xy^2 - xy$

   On a $\nabla f_1(x,y) = (2xy + y^2 - y, x^2 + 2xy - x)$, donc

$$\begin{aligned}
\nabla f_1(x,y) &= 0 \Leftrightarrow \left\{ \begin{array}{l} 2xy + y^2 - y = 0 \\ x^2 + 2xy - x = 0 \end{array} \right. \\
&\Leftrightarrow \left\{ \begin{array}{l} (2x+y-1)y = 0 \\ (x+2y-1)x = 0 \end{array} \right. \\
&\Leftrightarrow \left\{ \begin{array}{l} 2x+y-1 = 0 \\ x+2y-1 = 0 \end{array} \right. \text{ou} \left\{ \begin{array}{l} y = 0 \\ x-2y-1 = 0 \end{array} \right. \text{ou} \left\{ \begin{array}{l} 2x+y-1 = 0 \\ x = 0 \end{array} \right. \text{ou} \left\{ \begin{array}{l} y = 0 \\ x = 0 \end{array} \right. \\
&\Leftrightarrow \left\{ \begin{array}{l} x = 1/3 \\ y = 1/3 \end{array} \right. \text{ou} \left\{ \begin{array}{l} y = 0 \\ x = 1 \end{array} \right. \text{ou} \left\{ \begin{array}{l} y = 1 \\ x = 0 \end{array} \right. \text{ou} \left\{ \begin{array}{l} y = 0 \\ x = 0 \end{array} \right.
\end{aligned}$$

   Il y a 4 points critiques : $(1/3, 1/3)$, $(0,1)$, $(1,0)$ et $(0,0)$. La hessienne de f_1 est

$$Hf_1(x,y) = \begin{pmatrix} 2y & 2x+2y-1 \\ 2x+2y-1 & 2x \end{pmatrix}.$$

En $(1/3, 1/3)$ on a :

$$Hf_1(1/3, 1/3) = \begin{pmatrix} 2/3 & 1/3 \\ 1/3 & 2/3 \end{pmatrix}.$$

Nous savons que les valeurs propres λ_1, λ_2 sont des réels et doivent satisfaire

$$\begin{cases} \lambda_1 + \lambda_2 = {}^t(Hf_1(1/3, 1/3)) = 4/3 > 0, \\ \lambda_1\lambda_2 = \det(Hf_1(1/3, 1/3)) = (2/3) \times (2/3) - (1/3) \times (1/3) = 1/3 > 0 \end{cases}$$

La somme des deux nombres est positive, ainsi que leur produit. Ceci est possible si et seulement si les deux nombres sont positifs. Donc les deux valeurs propres sont positives, ce qui implique que $(1/3, 1/3)$ est un minimiseur local.

En $(0, 1)$ on a :

$$Hf_1(0, 1) = \begin{pmatrix} 2 & 1 \\ 1 & 0 \end{pmatrix}.$$

Les deux valeurs propres doivent satisfaire $\lambda_1\lambda_2 = -1 < 0$ ce qui est possible si et seulement si l'une est positive et l'autre négative. Donc $(0, 1)$ est un point selle.

En $(1, 0)$ on a :

$$Hf_1(1, 0) = \begin{pmatrix} 0 & 1 \\ 1 & -2 \end{pmatrix}.$$

Ici également, les deux valeurs propres satisfont $\lambda_1\lambda_2 = -1 < 0$, donc $(1, 0)$ est aussi un point selle.

En $(0, 0)$ on a :

$$Hf_1(1, 0) = \begin{pmatrix} 0 & -1 \\ -1 & 0 \end{pmatrix}.$$

Iici aussi, les deux valeurs propres satisfont $\lambda_1\lambda_2 = -1 < 0$, donc $(0, 0)$ est aussi un point selle.

Finalement, à propos des extrema globaux : d'abord, puisque f_1 n'a pas de maximum local, elle ne peut pas avoir le maximum global. Ensuite f_1 a exactement un minimisuer local, en $(1/3, 1/3)$, donc la question est de savoir si ce point est aussi un minimiseur global. En d'autres termes, la question est:

(Q): Est-il vrai que $f_1(x, y) \geq f_1(1/3, 1/3) = -1/27$ pour tout $(x, y) \in \mathbb{R}^2$?

En fait nous pouvons remarquer qu'en rpenant $x = y$ on a $f_1(x, x) = x^2(2x - 1)$, ce qui tend vers $-\infty$ lorsque x tend vers $-\infty$. Donc cette fonction n'a pas de minimum global. Pour trouver un contre-exemple explicite à (Q) on peut choisir $x = y = -2$:

$$f_1(-2, -2) = (-2)(-2)(-2 - 2 - 1) = -20 < -1/27.$$

2. $f_2(x, y) = 4xy + \frac{1}{x} + \frac{1}{y}$

Premièrement nous remarquons que $f_2(x, y)$ n'est pas défini lorsque $x = 0$ ou $y = 0$. Aussi lorsque x ou y s'approche de 0, $f_2(x, y)$ tend vers $+\infty$ ou $-\infty$ suivant le signe de x ou y. Donc ici non plus, f_2 n'a pas d'extremum global.

Le gradient s'écrit

$$\nabla f_2(x,y) = \begin{pmatrix} 4y - \frac{1}{x^2} \\ 4x - \frac{1}{y^2} \end{pmatrix}$$

Les points critiques satisfont

$$\begin{cases} 4y - \frac{1}{x^2} = 0 \\ 4x - \frac{1}{y^2} = 0 \end{cases} \Leftrightarrow \begin{cases} y = \frac{1}{4x^2} \\ 4x - 16x^4 = 0 \end{cases} \Leftrightarrow \begin{cases} y = \frac{1}{4x^2} \\ 1 - 4x^3 = 0 \end{cases} \Leftrightarrow \begin{cases} y = \frac{1}{4x^2} = 4^{-1/3} \\ x = 4^{-1/3} \end{cases}.$$

Donc $(4^{-1/3}, 4^{-1/3})$ est le seul point critique. La hessienne s'écrit

$$Hf_2(x,y) = \begin{pmatrix} 2/x^3 & 4 \\ 4 & 2/y^3 \end{pmatrix}.$$

Donc,

$$Hf_2(4^{-1/3}, 4^{-1/3}) = \begin{pmatrix} 8 & 4 \\ 4 & 8 \end{pmatrix}.$$

Les valeurs propres λ_1, λ_2 de cette matrice satisfont $\lambda_1 + \lambda_2 = 8 + 8 = 16 > 0$ and $\lambda_1\lambda_2 = 8 * 8 - 4 * 4 = 48 > 0$, donc les deux valeurs propres sont positives.. Par conséquent $(4^{-1/3}, 4^{-1/3})$ est un minimiseur local.

3. $f_3(x,y) = x^3 + y^3 - 9xy + 27$

$f_3(x,y)$ tend vers $+\infty$ ou $-\infty$ lorsque x ou y tend vers $+\infty$ ou $-\infty$, à cause des termes x^3 et y^3. Donc $f_3(x,y)$ n'ap as d'extremum global. Son gradient s'écrit

$$\nabla f_3(x,y) = \begin{pmatrix} 3x^2 - 9y \\ 3y^2 - 9x \end{pmatrix}$$

Les points critiques satisfont

$$\begin{aligned} \begin{cases} 3x^2 - 9y = 0 \\ 3y^2 - 9x = 0 \end{cases} \quad & \Leftrightarrow \quad \begin{cases} x^2 = 3y \\ y^2 = 3x \end{cases} \Leftrightarrow \begin{cases} y = x^2/3 \\ x^4/9 = 3x \end{cases} \\ & \Leftrightarrow \quad x = y = 0 \quad \text{ou} \quad \begin{cases} y = x^2/3 \\ x^3 = 27 \end{cases} \\ & \Leftrightarrow \quad x = y = 0 \quad \text{ou} \quad x = y = 3. \end{aligned}$$

Il y a donc 2 points critiques : $(0,0)$ et $(3,3)$. La matrice hessienne vaut

$$Hf_3(x,y) = \begin{pmatrix} 6x & -9 \\ -9 & 6y \end{pmatrix},$$

donc pour $(0,0)$ on a

$$Hf_3(0,0) = \begin{pmatrix} 0 & -9 \\ -9 & 0 \end{pmatrix}.$$

Les valeurs propres vérifient $\lambda_1\lambda_2 = -81 < 0$, donc une est positive et l'autre négative. Donc $(0,0)$ est un point selle. Pour $(3,3)$ on a

$$Hf_3(3,3) = \begin{pmatrix} 18 & -9 \\ -9 & 18. \end{pmatrix}$$

Les valeurs propres satisfont $\lambda_1\lambda_2 = 18^2 - 81 > 0$ et $\lambda_1 + \lambda_2 = 36 > 0$, donc elles sont positives. Donc $(3,3)$ est un minimiseur local.

4. $f_4(x,y) = x^2 - 3x^2y + 2x^4$.

   On peut d'abord remarquer que $f_4(x,y)$ tend vers $+\infty$ lorsque y tend vers $-\infty$ et vers $-\infty$ lorsque y tend vers $+\infty$, ceci tant que x est fixé et non nul. Ceci montre que f_4 n'a pas d'extremum global.

   Pour étudier les extrema locaux, calculons le gradient de f_4 :

$$\nabla f_4(x,y) = \begin{pmatrix} 2x - 6yx + 8x^3 \\ -3x^2 \end{pmatrix}$$

   Les points critiques dont les solutions de $\nabla f_4(x,y) = 0$:

$$\nabla f_4(x,y) = 0 \quad \Leftrightarrow \quad \left\{ \begin{array}{l} 2x - 6yx + 8x^3 = 0 \\ -3x^2 = 0 \end{array} \right. \quad \Leftrightarrow \quad x = 0,$$

   ce qui signifie que tous les points $(0,y)$ avec $y \in \mathbb{R}$ sont points critiques ; et donc tous ces points sont des potentiels extrema locaux. La hessienne s'écrit

$$Hf_4(x,y) = \begin{pmatrix} 2 - 6y + 24x^2 & -6x \\ -6x & 0 \end{pmatrix},$$

   et donc pour les points critiques :

$$Hf_4(0,y) = \begin{pmatrix} 2 - 6y & 0 \\ 0 & 0 \end{pmatrix}.$$

   Cette matrice a une valeur propre nulle, donc nous ne pouvons pas conclure directement sur le type de point critique. Pour répondre ici, nous remarquons que $f_4(x,y) = x^2(1 - 3y + 2x^2)$: lorsque x est proche de 0, x^2 est positif et $1 - 3y + 2x^2$ a le même signe que $1 - 3y$, excepté lorsque $1 - 3y = 0$ auquel cas il est positif. Donc pour $y < \frac{1}{3}$, $1 - 3y > 0$ donc $f_4(x,y) \geq 0 = f_4(0,y)$ donc $(0,y)$ est un minimiseur local dans ce cas. Pour $y > \frac{1}{3}$, $1 - 3y < 0$ donc $f_4(x,y) \leq 0 = f_4(0,y)$ donc $(0,y)$ est un maximiseur local. Finalement lorsque $y = \frac{1}{3}$, on a un point selle parce que $f_4(x,y)$ prend des valeurs positives et négatives dans le voisinage.

5. $f_5(x,y) = (y^2 - x^2)(y^2 - 2x^2)$

   On a $f_5(x,y) = y^4 - 3x^2y^2 + 2x^4$. Le gradient s'écrit

$$\nabla f_5(x,y) = \begin{pmatrix} -6y^2x + 8x^3 \\ 4y^3 - 6x^2y \end{pmatrix}$$

Les points critiques satisfont

$$\left\{\begin{array}{l} -6y^2x+8x^3=0 \\ 4y^3-6x^2y=0 \end{array}\right. \quad \Leftrightarrow \quad \left\{\begin{array}{l} x=0 \\ y=0 \end{array}\right. \quad \text{ou} \quad \left\{\begin{array}{l} -6y^2+8x^2=0 \\ 4y^2-6x^2=0 \end{array}\right.$$
$$\Leftrightarrow \quad \left\{\begin{array}{l} x=0 \\ y=0 \end{array}\right. \quad \text{ou} \quad \left\{\ 4x^2=3y^2+2y^2=3x^2\right. \quad .$$

Le dernier sutème implique $x^2/y^2 = 3/4 = 2/3$ ce qui est faux, donc il n'a pas de solution. Ainsi $(0,0)$ est le seul point critique. La matrice hessienne est

$$Hf_5(x,y) = \begin{pmatrix} -6y^2+24x^2 & -12yx \\ -12yx & 12y^2-6x^2 \end{pmatrix},$$

donc $Hf_5(0,0)$ est la matrice nulle. Donc nous ne pouvons rien conclure de ceci. Pour conclure nous revenons à la première écriture de f_5 : $f_5(x,y) = (y^2-x^2)(y^2-2x^2)$. Lorsque le point (x,y) est situé sur la ligne $y=\alpha x$ on a $f_5(x,\alpha x) = (\alpha^2x^2-x^2)(\alpha^2x^2-2x^2) = (\alpha^2-1)(\alpha^2-2)x^4$, ce qui est négatif lorsque α^2 est entre 1 et 2 et positif sinon. Donc nous voyons que $f_5(x,y)$ prend des valeurs positives et négatives dans le voisinage de $(0,0)$, tandis que $f_5(0,0)=0$. Ceci montre que $(0,0)$ est un point selle. Par conséquent f_5 n'a pas d'extremum local, et donc pas d'extremum global non plus.

6. $f_6(x,y) = (x+y)^2 - (x^4+y^4)$

Ici le terme x^4+y^4 domine sur le premier lorsque x ou y est grand, ce qui montre que $f_6(x,y)$ tend vers $-\infty$ quand x ou y tend vers $\pm\infty$. Donc $-f_6$ étant continue et coercive, elle admet un minimum global et pas de maximum global ; et donc à l'inverse f_6 admet un maximum global mais pas de minimum local.

Le gradient d'écrit

$$\nabla f_6(x,y) = \begin{pmatrix} 2x+2y-4x^3 \\ 2y+2x-4y^3 \end{pmatrix}$$

On a

$$\nabla f_6(x,y)=0 \quad \Leftrightarrow \quad \left\{\begin{array}{l} 2x+2y-4x^3=0 \\ 2y+2x-4y^3=0 \end{array}\right. \quad \Leftrightarrow \quad \left\{\begin{array}{l} y=2x^3-x \\ 2x^3=2y^3 \end{array}\right. \quad \Leftrightarrow \quad \left\{\begin{array}{l} x=2x^3-x \\ y=x \end{array}\right.$$
$$\Leftrightarrow \quad x=y=0 \quad \text{ou} \quad \left\{\begin{array}{l} 2=2x^2 \\ y=x \end{array}\right.$$
$$\Leftrightarrow \quad x=y=0 \quad \text{ou} \quad \left\{\begin{array}{l} x=\pm 1 \\ y=x \end{array}\right.$$

Donc il y a 3 points critiques : $(0,0)$, $(1,1)$ et $(-1,-1)$. La hessienne s'écrit

$$Hf_6(x,y) = \begin{pmatrix} 2-12x^2 & 2 \\ 2 & 2-12y^2 \end{pmatrix}.$$

Donc pour $(0,0)$ on a

$$Hf_6(0,0) = \begin{pmatrix} 2 & 2 \\ 2 & 2 \end{pmatrix}.$$

Cette matrice est dégénérée, ce qui signifie qu'elle a une valeur propre nulle. Donc on ne peut pas conclure sur le type de ce point critique à partir de l'analyse de la hessienne. Cependant on peut remarquer que lorsque (x, y) approche $(0, 0)$le long de la ligne $y = x$, $f_6(x, y) = f_6(x, x) = 4x^2 - 2x^4 = 2x^2(2 - x^2) > 0$; et quand (x, y) approche $(0, 0)$ le long de $y = -x$, $f_6(x, y) = f_6(x, -x) = -2x^4 < 0$. Donc f_6 prend à la fois des valeurs positives et négatives au voisinage de $(0, 0)$, tandis que $f_6(0, 0) = 0$. Donc $(0, 0)$ est un point selle.

Pour $(1, 1)$ et $(-1, -1)$ on a

$$Hf_6(1, 1) = Hf_6(-1, -1) = \begin{pmatrix} -10 & 2 \\ 2 & -10 \end{pmatrix}.$$

LEs deux valeurs propres satisfont $\lambda_1\lambda_2 = (-10)^2 - 2^2 = 96 > 0$ et $\lambda_1 + \lambda_2 = -10 - 10 = -20 < 0$, donc les deux valeurs propres sont négatives. Par conséquent $(1, 1)$ et $(-1, -1)$ sont tous les deux des maxima locaux.

7. $f_7(x, y) = x \ln y + y \ln x$

$f_7(x, y)$ est défini seulement pour $x > 0$ et $y > 0$, à cause des logarithmes. La fonction tend vers $-\infty$ lorsque x ou y tend vers 0 et vers $+\infty$ quand x ou y tend vers $+\infty$. Donc f_7 n'a pas d'extremum global. Le gradient de f_7 est

$$\nabla f_7(x, y) = \begin{pmatrix} \ln y + \frac{y}{x} \\ \frac{x}{y} + \ln x \end{pmatrix}$$

Maintenant,

$$\nabla f_7(x, y) = 0 \quad \Leftrightarrow \quad \begin{cases} \ln y + \frac{y}{x} = 0 \\ \frac{x}{y} + \ln x = 0 \end{cases} \quad \Leftrightarrow \quad \begin{cases} x = -\frac{y}{\ln y} \\ y = -\frac{x}{\ln x} \end{cases}$$

On peut montrer que ce système admet comme solution unique $(1/e, 1/e)$: en effet, tout d'abord si (x, y) vérifie le système on a forcément $0 < x < 1$ et $0 < y < 1$, car si $x > 1$, $\ln x > 0$ et donc $y = -\frac{x}{\ln x} < 0$, ce qui est exclu ; et de même $y > 1$ est impossible. A présent posons $v(t) = -\frac{t}{\ln t}$. Si (x, y) est solution du système on doit avoir $y = v(x)$ et $x = v(y) = v(v(x))$. On va montrer que seul $x = 1/e$ vérifie $x \in]0, 1[$, $v(x) \in]0, 1[$ et $x = v(v(x))$. Pour cela posons $u(t) := -\frac{t}{\ln t} - t$ et étudions ses variations sur $]0, 1[$. On a $u'(t) = \frac{1-\ln t}{(\ln t)^2} - 1 = \frac{1-\ln t-(\ln t)^2}{(\ln t)^2}$, qui est du signe de $1 - \ln t - (\ln t)^2$. Or le polynôme $-X^2 - X + 1$ a pour racines $\frac{-1-\sqrt{5}}{2} < 0$ et $\frac{-1+\sqrt{5}}{2} > 0$, et est positif entre ces deux valeurs, négatif en dehors, et donc par croissance de l'exponentielle, $1 - \ln t - (\ln t)^2$ est positif entre $t = e^{\frac{-1-\sqrt{5}}{2}} (< 1)$ et $t = e^{\frac{-1+\sqrt{5}}{2}} (> 1)$. Puisqu'on s'intéresse aux valeurs de $t \in]0, 1[$, ça implique que $u'(t) < 0$ sur $]0, e^{\frac{-1-\sqrt{5}}{2}}[$ et positif sur $]e^{\frac{-1-\sqrt{5}}{2}}, 1[$, et donc u est décroissante sur $]0, e^{\frac{-1-\sqrt{5}}{2}}[$ et croissante sur $]e^{\frac{-1-\sqrt{5}}{2}}, 1[$. De plus $u(t)$ tend vers 0 lorsque t tend vers 0 ; et on a $u(1/e) = 0$, avec $1/e > e^{\frac{-1-\sqrt{5}}{2}}$. Par conséquent, l'étude des variations de $u(t)$ montre que $u(t)$ est négatif sur $]0, 1/e[$ et positif sur $]1/e, 1[$. Supposons alors que $x \in]0, 1[$, avec $v(x) \in]0, 1[$. On a trois cas :

- si $0 < x < 1/e$, alors $u(x) < 0$ donc $v(x) < x$, et donc $0 < v(x) < 1/e$, et donc $v(v(x)) < v(x) < x$. On ne peut donc pas avoir $x = v(v(x))$ dans ce cas.
- si $1/e < x < 1$, alors $u(x) > 0$ donc $x < v(x)$, et donc $1/e < v(x) < 1$ et donc $x < v(x) < v(v(x))$. On ne peut donc toujours pas avoir $x = v(v(x))$.
- si $x = 1/e$ alors on a bien $x = v(x)$ et donc $x = v(v(x))$.

Au final, $(1/e, 1/e)$ est le seul point critique de f_7.

A présent la matrice hessienne s'écrit

$$Hf_7(x,y) = \begin{pmatrix} -\frac{y}{x^2} & \frac{1}{y} + \frac{1}{x} \\ \frac{1}{y} + \frac{1}{x} & -\frac{x}{y^2} \end{pmatrix}.$$

Donc

$$Hf_7(1/e, 1/e) = \begin{pmatrix} -e & 2e \\ 2e & -e \end{pmatrix}.$$

Son déterminant est $e^2 - 4e^2 = -3e^2 < 0$, donc les deux valeurs propres ont des signes opposés. Ainsi $(1/e, 1/e)$ est un point selle.

Exercice 6. Soit $p^1 = (p_1^1, p_2^1)$, $p^2 = (p_1^2, p_2^2)$, ..., $p^n = (p_1^n, p_2^n)$ n points dans $\mathbb{R}^2$. Le but de cet exercice est de trouver le point qui minimise la somme des distances au carré à tous les points p^i : ainsi, pour tout $x = (x_1, x_2)$ nous définissons la fonctionnelle suivante :

$$J(x) = \sum_{i=1}^{n} \|x - p^i\|^2.$$

1. Calculer $\frac{\partial J}{\partial x_1}(x)$ et $\frac{\partial J}{\partial x_2}(x)$, et en déduire que J admet un unique point critique x^* sur $\mathbb{R}^2$ à déterminer. Comment s'appelle ce point en termes géométriques ?
2. Montrer que x^* est un minimiseur local de J.
3. Expliquer pourquoi x^* est aussi l'unique minimiseur global de J.

Correction

1. On a

$$J(x) = \sum_{i=1}^{n} \|x - p^i\|^2 = \sum_{i=1}^{n} (x_1 - p_1^i)^2 + (x_2 - p_2^i)^2,$$

donc

$$\begin{aligned} \frac{\partial J}{\partial x_1}(x) &= \sum_{i=1}^{n} 2(x_1 - p_1^i) \\ &= 2\sum_{i=1}^{n} x_1 - 2\sum_{i=1}^{n} p_1^i \\ &= 2nx_1 - 2\sum_{i=1}^{n} p_1^i, \end{aligned}$$

et de même

$$\frac{\partial J}{\partial x_2}(x) \quad = \quad 2nx_2 - 2\sum_{i=1}^{n} p_2^i.$$

Les points critiques de J sont les points pour lesquels le gradient s'annule, i.e. les points pour lesquels toutes les dérivées partielles s'annulent ensemble.On doit donc avoir

$$\left\{ \begin{array}{l} \frac{\partial J}{\partial x_1}(x) = 0 \\ \frac{\partial J}{\partial x_2}(x) = 0 \end{array} \right. \quad \Leftrightarrow \quad \left\{ \begin{array}{l} 2nx_1 - 2\sum_{i=1}^{n} p_1^i = 0 \\ 2nx_2 - 2\sum_{i=1}^{n} p_2^i = 0 \end{array} \right.$$

$$\Leftrightarrow \quad \left\{ \begin{array}{l} x_1 = \frac{1}{n}\sum_{i=1}^{n} p_1^i \\ x_2 = \frac{1}{n}\sum_{i=1}^{n} p_2^i. \end{array} \right.$$

Par conséquent le seul point critique est $x^* = (\frac{1}{n}\sum_{i=1}^{n} p_1^i, \frac{1}{n}\sum_{i=1}^{n} p_2^i) = \frac{1}{n}\sum_{i=1}^{n}(p_1^i, p_2^i)$. Donc

$$x^* = \frac{1}{n}\sum_{i=1}^{n} p^i.$$

Ce point est appelé barycentre des points p^i dans $\mathbb{R}^2$.

2. On calcule la matrice hessienne de la fonctionnelle J:

$$\begin{aligned} H_J(x) &= \begin{pmatrix} \frac{\partial^2 J}{\partial x_1^2}(x) & \frac{\partial^2 J}{\partial x_1 \partial x_2}(x) \\ \frac{\partial^2 J}{\partial x_1 \partial x_2}(x) & \frac{\partial^2 J}{\partial x_1^2}(x) \end{pmatrix} \\ &= \begin{pmatrix} 2n & 0 \\ 0 & 2n \end{pmatrix}. \end{aligned}$$

   $H_J(x^*)$ admet $2n > 0$ comme valeur propre doubles, donc x^* est un minimiseur local de J.

3. $H_J(x)$ a ses valeurs propres positives quel que soit x, donc J est une fonction convexe. Donc tout minimiseur local est en fait un minimiseur global. Ainsi x^* est minimiseur global de J. Remarquons que même sans avoir étudié les points critiques de J, on aurait pu dire directement que J admet un minimiseur global unique car J est une fonction elliptique : il existe $\alpha > 0$ tel que pour tout $x \in \mathbb{R}^2$, les valeurs propres de $H_J(x)$ sont toujours $\geq \alpha$ (il suffit de prendre $\alpha = 2n$).

Exercice 7.

Exercice 8. Modèle linéaire Soient $(x_i, y_i)_{i=1}^n$ un n-uplet de points de $\mathbb{R}^2$. Nous cherchons à trouver une relation entre les variables x_i et y_i. On considère le modèle

$$f(x) = \sum_{j=0}^{k} \beta_j w_j(x),$$

où les w_j sont des fonctions de $\mathbb{R}$ dans $\mathbb{R}$ et les β_j sont des coefficients. On cherche les coefficients β_j qui s'adaptent le mieux au modèle $y_i = f(x_i)$ pour $i = 1, \cdots, n$ en minimisant

$$J(\beta_0, \cdots, \beta_k) = \sum_{i=1}^{n} \left(\sum_{j=0}^{k} \beta_j w_j(x_i) - y_i \right)^2 .$$

On supposera que $n \geq k + 1$.

1. Montrer que l'on peut écrire
$$J(\beta) = \|M\beta - y\|^2,$$
où $\beta = (\beta_0, \cdots, \beta_k)^T$, $y = (y_1, \cdots, y_n)^T$, et M est une matrice à définir.
2. Montrer que le vecteur β des coefficients optimaux satisfait $M^T M\beta = M^T y$.
3. Montrer que si M est de rang maximal alors la solution est unique.
4. On considère le cas de la régression polynomiale : $w_j(x) = x^j$. Écrire une fonction RegressionPolynomiale(x,y,k) qui calcule cette solution à partir de vecteurs x, y donnant les coordonnées des points, et de l'ordre k. La fonction doit renvoyer un vecteur beta dans $\mathbb{R}^{k+1}$ correspondant aux coefficients calculés. Tester cette fonction pour

```
x = np.random.rand(10,1)
y = -5 + 12*x + np.random.randn(10,1)
k = 3
```

et afficher sur le même graphique les points et la courbe de régression.

Correction

1. Si on pose M la matrice de taille $(n, k+1)$, de coefficients $M_{ij} = w_j(x_i)$, alors on a
$$J(\beta) = \sum_{i=1}^{n} \left(\sum_{j=0}^{k} M_{ij}\beta_j - y_i \right)^2 = \sum_{i=1}^{n} ((M\beta)_i - y_i)^2 = \|M\beta - y\|^2.$$
2. Calculons le gradient de J : on a
$$DJ(\beta).h = 2 \langle Mh , M\beta - y \rangle = \left\langle h , 2M^T(M\beta - y) \right\rangle ,$$
et donc $\nabla J(\beta) = 2M^T(M\beta - y)$. Par conséquent les points critiques de J sont les β vérifiant
$$\nabla J(\beta) = 0 \quad \Leftrightarrow \quad M^T M\beta = M^T y$$

Montrons à présent que J est une fonction convexe. Soient $\beta, \beta' \in \mathbb{R}^{k+1}$; on va montrer que $J(\beta') - J(\beta) - \langle \nabla J(\beta) , \beta' - \beta \rangle \geq 0$. Notons Δ cette quantité. On a

$$\begin{aligned}
\Delta &= \|M\beta' - y\|^2 - \|M\beta - y\|^2 - 2 \left\langle M^T(M\beta - y) , \beta' - \beta \right\rangle \\
&= \left\langle (M\beta' - y) - (M\beta - y) , (M\beta' - y) + (M\beta - y) \right\rangle - 2 \left\langle M\beta - y , M(\beta' - \beta) \right\rangle \\
&= \left\langle M(\beta' - \beta) , (M(\beta' + \beta) - 2y \right\rangle - 2 \left\langle M\beta - y , M(\beta' - \beta) \right\rangle \\
&= \left\langle M(\beta' - \beta) , (M(\beta' + \beta) - 2y - 2M\beta + 2y \right\rangle \\
&= \left\langle M(\beta' - \beta) , (M(\beta' - \beta) \right\rangle \\
&= \|M(\beta' - \beta)\|^2.
\end{aligned}$$

Par conséquent $\Delta \geq 0$ et donc J est une fonction convexe. Ceci implique que les minimiseurs de J sont exactement ses points critiques, et donc les solutions de $M^T M\beta = M^T y$.

3. Puisque $n \geq k + 1$, si M est de rang maximal c'est que son rang vaut $k + 1$/ Montrons qu'alors $M^T M$ est inversible. En effet,

$$M^T Mh = 0 \Rightarrow \left\langle M^T Mh, h\right\rangle = 0 \Leftrightarrow \langle Mh, Mh\rangle = 0 \Leftrightarrow \|Mh\|^2 = 0 \Leftrightarrow Mh = 0.$$

   Or M étant de rang $k + 1$, l'application $h \mapsto Mh$ est injective d'après le théorème du rang, et donc $Mh = 0$ implique $h = 0$. Donc $M^T M$ est alors inversible, et donc le système $M^T M\beta = M^T y$ admet une solution unique.

4. cf fichier exo8.py