Optimisation algorithmique - Examen partiel du 4 novembre 2019 - Correction

Exercice 1. On considère la fonction $f : \mathbb{R}^2 \to \mathbb{R}$ définie par $f(x,y) = x^3 + 6x^2 + 3y^2 - 12xy + 9x$.

1. Calculer le gradient de f et déterminer les points critiques de la fonction.

Correction $\nabla f(x,y) = (3x^2 + 12x - 12y + 9\ ,\ 6y - 12x)$. Les points critiques de f vérifient donc

$$\left\{\begin{array}{l} 3x^2 + 12x - 12y + 9 = 0 \\ 6y - 12x = 0 \end{array}\right. \Leftrightarrow \left\{\begin{array}{l} x^2 + 4x - 4y + 3 = 0 \\ y = 2x \end{array}\right. \Leftrightarrow \left\{\begin{array}{l} x^2 - 4x + 3 = 0 \\ y = 2x \end{array}\right.$$

L'équation $x^2 - 4x + 3 = 0$ a pour solutions $x = 1$ et $x = 3$. Les points critiques de f sont donc $(1,2)$ et $(3,6)$.

2. Pour chacun des points critiques, dire s'il s'agit d'un minimiseur local, d'un maximiseur local, ou d'un point selle.

Correction La matrice hessienne de f vaut $H_f(x,y) = \begin{pmatrix} 6x+12 & -12 \\ -12 & 6 \end{pmatrix}$. On a donc :

- en $(1,2)$: $H_f(1,2) = \begin{pmatrix} 18 & -12 \\ -12 & 6 \end{pmatrix} = 6\begin{pmatrix} 3 & -2 \\ -2 & 1 \end{pmatrix}$. Les deux valeurs propres λ_1, λ_2 de cette dernière matrice vérifient $\lambda_1\lambda_2 = 3 \times 1 - (-2)^2 = -1 < 0$, donc elles ont des signes opposés. Par conséquent le point critique $(1,2)$ est un point selle.
- en $(3,6)$: $H_f(3,6) = \begin{pmatrix} 30 & -12 \\ -12 & 6 \end{pmatrix} = 6\begin{pmatrix} 5 & -2 \\ -2 & 1 \end{pmatrix}$. Les deux valeurs propres λ_1, λ_2 de cette dernière matrice vérifient $\lambda_1\lambda_2 = 5 \times 1 - (-2)^2 = 1 > 0$, donc elles sont de même signe. De plus leur somme vaut $\lambda_1 + \lambda_2 = 5 + 1 = 6 > 0$, donc il s'agit de deux valeurs propres strictement positives. Par conséquent le point critique $(3,6)$ est un minimiseur local.

3. La fonction f admet-elle un minimum global ?

Correction On peut facilement voir que f n'admet pas de minimum global. En effet le long de la droite $y = 0$ on a : $f(x,0) = x^3 + 6x^2 + 9x$, ce qui est un polynôme de degré 3 en x, et ne peut donc pas admettre d'extremum. Plus précisément, $f(x,0)$ tend vers $-\infty$ lorsque x tend vers $-\infty$, le terme x^3 étant dominant. Ainsi f n'admet pas de minimum global.

Exercice 2. *Partie A (questions préliminaires).*

1. Montrer qu'une somme de fonctions convexes est convexe.

Correction En toute généralité, une "somme" ne contient pas que deux termes, ni même un nombre fini, et donc pour répondre le plus rigoureusement possible il faut considérer une fonction de la forme

$$f(x) = \sum_{n\in\mathbb{N}} f_n(x),$$

où chaque f_n est une fonction convexe, définie sur un même sous-ensemble convexe C d'un espace vectoriel E, et à valeurs dans $\mathbb{R}$. La fonction f est supposée bien définie, au sens où chaque série $\sum_{n\in\mathbb{N}} f_n(x)$ converge (la convergence simple suffit). On considère alors $x, y \in C$ quelconques, et $\lambda \in [0,1]$. Par convexité de C, le point $\lambda x+(1-\lambda)y$ est bien aussi un élément de C, et comme chaque f_n est convexe, on a

$$\forall n \in \mathbb{N}, \quad f_n(\lambda x + (1-\lambda)y) \leq \lambda f_n(x) + (1-\lambda) f_n(y),$$

et donc en sommant, d'abord sur une nombre fini de termes, on a (puisque λ et $1-\lambda$ sont positifs) :

$$\forall N \in \mathbb{N}, \quad \sum_{n=0}^{N} f_n(\lambda x + (1-\lambda)y) \leq \lambda \sum_{n=0}^{N} f_n(x) + (1-\lambda) \sum_{n=0}^{N} f_n(y),$$

puis en passant à la limite en N,

$$f(\lambda x + (1-\lambda)y) \leq \lambda f(x) + (1-\lambda) f(y).$$

Ainsi f est bien aussi une fonction convexe.

2. Montrer que si une fonction $f : \mathbb{R} \to \mathbb{R}$ est convexe et constante sur un intervalle $[a, b]$, alors tout $x \in [a, b]$ minimise f.

Correction *remarque* : L'énoncé aurait du préciser qu'on suppose $a \neq b$, sinon c'est évidemment faux. Soit $y \in \mathbb{R}$ quelconque. On distingue trois cas :

- si $y < a$: alors $y < a < b$ et donc il existe $\lambda \in]0, 1[$ tel que $a = \lambda y + (1-\lambda)b$ (plus précisément, $\lambda = (b-a)/(b-y)$). On a donc par convexité de f,

  $$m = f(a) = f(\lambda y + (1-\lambda)b) \leq \lambda f(y) + (1-\lambda) f(b) = \lambda f(y) + (1-\lambda)m,$$

  et donc $\lambda m \leq \lambda f(y)$, et donc puisque $\lambda > 0$, $m \leq f(y)$.

- si $a \leq y \leq b$: alors $f(y) = m$,

- si $y > b$: il existe alors $\lambda \in]0, 1[$ tel que $b = \lambda a+(1-\lambda)y$, et donc comme précédemment, on peut écrire

  $$m = f(b) = f(\lambda a + (1-\lambda)y) \leq \lambda f(a) + (1-\lambda) f(y) = \lambda m + (1-\lambda) f(y),$$

  ce qui implique $(1-\lambda)m \leq (1-\lambda) f(y)$, et comme $1-\lambda > 0$, on obtient $m \leq f(y)$.

Ainsi dans tous les cas, $f(y) \geq m$, ce qui prouve que m est la valeur minimale de f sur $\mathbb{R}$. Par conséquent, tout $x \in [a,b]$ est un minimiseur de f.

Soit $d \geq 1$ un entier et $y \in \mathbb{R}^d$ fixé. On considère la fonction $f : \mathbb{R}^d \to \mathbb{R}$ définie par $f(x) = \|x - y\|$, où $\|\cdot\|$ désigne la norme euclidienne.

3. Sur quel domaine f est-elle différentiable ?

Correction On a $f(x) = \sqrt{\sum_{i=1}^d (x_i - y_i)^2} = h \circ g(x)$ où $g(x) = \sum_{i=1}^d (x_i - y_i)^2$ et $h(t) = \sqrt(t)$ pour $t \in \mathbb{R}_+$. La fonction g est polynomiale donc différentiable partout, et h est dérivable sur $\mathbb{R}_+^*$ mais pas en $t = 0$. Or $g(x) = 0$ si et seulement si $x = y$, donc par composition de fonctions, on voit que f est différentiable pour tout $x \in \mathbb{R}^d \setminus \{y\}$. En $x = y$ la fonction f n'est pas différentiable, mais ceci doit être prouvé. Pour cela il suffit par exemple de montrer que les dérivées directionnelles de f en $y = x$ n'existent pas. On pose donc pour $h \in \mathbb{R}^d$ quelconque (non nul), $\psi(t) = f(y + th)$ et on veut montrer que $\psi'(0)$ n'existe pas. Or

$$\psi(t) = f(y + th) = \|y + th - y\| = |t|\|h\|,$$

et on sait que la fonction valeur absolue n'est pas dérivable en 0. Ainsi ψ n'est pas dérivable en $t = 0$. En conclusion f est différentiable partout sauf en $x = y$.

4. Montrer que f est une fonction coercive.

Correction On a $f(x) = \|x - y\| \geq \|x\| - \|y\|$, ce qui tend vers $+\infty$ lorsque $\|x\|$ tend vers $+\infty$. Donc f est coercive.

5. Montrer que f est convexe, mais pas strictement convexe.

Correction Soient $x, z \in \mathbb{R}^d$ et $\lambda \in [0,1]$. On a

$$f(\lambda x + (1-\lambda)z) = \|\lambda x + (1-\lambda)z - y\| = \|\lambda(x-y) + (1-\lambda)(z-y)\| \leq \lambda\|x-y\| + (1-\lambda)\|z-y)\| = \lambda f(x) + (1-\lambda) f(z)$$

donc f est convexe.

Pour montrer que f n'est pas strictement convexe, on choisit $z = x$ et $\lambda \in]0,1[$. On a alors

$$f(\lambda x + (1-\lambda)y) = \|\lambda x + (1-\lambda)y - y\| = \|\lambda(x-y) \ = \lambda\|x-y\| = \lambda f(x) = \lambda f(x) + (1-\lambda) f(y)$$

(puisque $f(y) = 0$). Donc l'inégalité de stricte convexité n'est pas vérifiée. Ainsi f n'est pas strictement convexe.

Partie B. On considère à présent n points fixés $x^{(i)} \in \mathbb{R}^d$, et $J : \mathbb{R}^d \to \mathbb{R}$ la fonction

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

6. Sur quel domaine la fonction J est-elle différentiable ?

Correction D'après la question 3, chaque fonction $x \mapsto \|x - x^{(i)}\|$ est différentiable partout sauf en $x^{(i)}$, donc on en déduit que J est différentiable sur $\mathbb{R}^d \setminus \{x^{(1)}, \dots, x^{(n)}\}$. Montrons maintenant que f n'est pas différentiable en l'un des points $x^{(i_0)}$. Si les points $x^{(i)}$ sont tous distincts, tous les $x \mapsto \|x - x^{(i)}\|$ sont différentiables en $x^{(i_0)}$, sauf pour $i = i_0$, et donc J est la somme de fonctions différentiables en $x^{(i_0)}$ et d'une fonction non différentiable en $x^{(i_0)}$, et donc elle n'est pas différentiable en $x^{(i_0)}$. Dans le cas où les points en sont pas nécessairement distincts, et qu'un ou plusieurs $x^{(i)}$ sont confondus avec $x^{(i_0)}$, il suffit de voir qu'alors J s'écrit $J(x) = A\|x - x^{(i_0)}\| + g(x)$, où $A \geq 2$ entier et g différentiable en $x^{(i_0)}$, et donc de la même manière J ne peut pas être différentiable en $x^{(i_0)}$. Ainsi au final J est différentiable partout sauf aux points $x^{(i)}$, $1 \leq i \leq n$.

7. Montrer que J est coercive et convexe.

Correction D'après la question 5, chaque $x \mapsto \|x - x^{(i)}\|$ est convexe, et comme une somme de fonctions convexes est convexe (question 1), J est bien une fonction convexe. De même d'après la question 4, chaque $x \mapsto \|x - x^{(i)}\|$ est coercive, est il est clair qu'une somme de fonctions coercives est coercive (chaque fonction dans la somme tend vers $+\infty$, donc la somme également).

8. Montrer que si les points $x^{(i)}$ ne sont pas tous alignés, alors la fonction J est strictement convexe. En déduire que dans ce cas la fonction J admet un minimiseur unique.

Correction Supposons que les points $x^{(i)}$ ne sont pas tous alignés. Soient $x, y \in \mathbb{R}^d$ distincts et $\lambda \in]0,1[$. On a

$$\begin{aligned} J(\lambda x + (1-\lambda)y) &= \sum_{i=1}^{n} \|\lambda x + (1-\lambda)y - x^{(i)}\| \\ &= \sum_{i=1}^{n} \|\lambda(x - x^{(i)}) + (1-\lambda)(y - x^{(i)})\| \end{aligned}$$

Or $\|\lambda(x - x^{(i)}) + (1-\lambda)(y - x^{(i)})\| \leq \|\lambda(x - x^{(i)})\| + \|(1-\lambda)(y - x^{(i)})\|$ avec égalité possible seulement si $\lambda(x - x^{(i)})$ et $(1-\lambda)(y - x^{(i)})$ sont colinéaires, donc seulement si $x - x^{(i)}$ et $y - x^{(i)}$ sont colinéaires, autrement dit si $x^{(i)}$ est situé sur la droite passant par x et y. D'après l'hypothèse, les $x^{(i)}$ n'étant pas tous alignés, il existe forcément un indice i pour lequel $x^{(i)}$ n'est pas situé sur cette droite, et donc pour lequel $\|\lambda(x - x^{(i)}) + (1-\lambda)(y - x^{(i)})\| < \|\lambda(x - x^{(i)})\| + \|(1-\lambda)(y - x^{(i)})\|$. Alors nécessairement $J(\lambda x + (1-\lambda)y) < \lambda J(x) + (1-\lambda)J(y)$, et donc J est strictement convexe.

La fonction J est donc d'une part continue et coercive, ce qui implique qu'elle admet un minimiseur unique, et d'autre part strictement convexe, ce qui implique que ce minimiseur est unique.

9. Dans cette question on suppose $d = 1$, que les $x^{(i)}$ sont tous distincts, et que n est pair. Montrer qu'alors la fonction J est minimisée sur un certain intervalle à déterminer.

Correction On peut remarquer d'abord que comme $d = 1$ les points $x^{(i)}$ sont nécessairement alignés, et donc l'hypothèse précédente ne tient pas. Il s'agit donc ici de regarder un contre-exemple au cas d'unicité du minimiseur.

Comme $d = 1$, la fonction J est une fonction d'une variable réelle qui s'écrit

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

On peut trouver une permutation σ des indices de telle manière que $x^{(\sigma(1))} < x^{(\sigma(2))} < \cdots < x^{(\sigma(n))}$. Si $x^{(\sigma(n/2))} < x < x^{(\sigma(n/2+1))}$ on a alors

$$\begin{aligned}
J(x) &= \sum_{i=1}^{n/2}(x - x^{(i)}) + \sum_{i=n/2+1}^{n}(x^{(i)} - x) \\
&= (n/2)x - \sum_{i=1}^{n/2} x^{(i)} + \sum_{i=n/2+1}^{n} x^{(i)} - (n/2)x \\
&= -\sum_{i=1}^{n/2} x^{(i)} + \sum_{i=n/2+1}^{n} x^{(i)}.
\end{aligned}$$

Cette expression est indépendante de x, et donc on voit que J est constante sur l'intervalle $[x^{(\sigma(n/2))}, x^{(\sigma(n/2+1))}]$. J étant de plus convexe, d'après la question 2 on en conclut que J est minimisée sur cet intervalle.

10. On revient au cas d quelconque. Montrer que si le point x minimise J et qu'il est distinct de tous les $x^{(i)}$, alors on a

$$\sum_{i=1}^{n} \frac{x - x^{(i)}}{\|x - x^{(i)}\|} = 0. \qquad (*)$$

Correction Si le point x est distinct de tous les $x^{(i)}$, alors J est différentiable en x d'après la question 6. Donc si de plus J est minimisée en x, alors x est un point critique de J. On doit donc calculer le gradient de J. Pour cela on écrit que $J = \sum_{i=1}^{n} h \circ g_i$ où $h(t) = \sqrt{t}$ et $g_i(x) = \|x - x^{(i)}\|^2$. On a alors pour tout $h \in \mathbb{R}^d$,

$$DJ(x).h = \sum_{i=1}^{n} h'(g_i(x))Dg_i(x).h.$$

Or $h'(t) = \frac{1}{2\sqrt{t}}$ et $Dg_i(x).h = 2\left\langle\, x - x^{(i)}\ ,\ h \,\right\rangle$. D'où

$$DJ(x).h = \sum_{i=1}^{n} \frac{2\left\langle\, x - x^{(i)}\ ,\ h \,\right\rangle}{2\sqrt{g_i(x)}} = \sum_{i=1}^{n} \frac{\left\langle\, x - x^{(i)}\ ,\ h \,\right\rangle}{\|x - x^{(i)}\|} = \left\langle\, \sum_{i=1}^{n} \frac{x - x^{(i)}}{\|x - x^{(i)}\|}\ ,\ h \,\right\rangle.$$

Ainsi par identification, $\nabla J(x) = \sum_{i=1}^{n} \frac{x-x^{(i)}}{\|x-x^{(i)}\|}$, et donc comme x est un point critique,

$$\sum_{i=1}^{n} \frac{x-x^{(i)}}{\|x-x^{(i)}\|} = 0.$$

11. Montrer que réciproquement, si un point $x \in \mathbb{R}^d$ est distinct de tous les $x^{(i)}$ et vérifie $(*)$, alors J est minimisée en x.

Correction J est différentiable en x, avec $\nabla J(x) = 0$, et par ailleurs on sait que J est convexe. Donc tout point critique est un minimiseur, et ainsi x minimise J.

12. Un algorithme classique de calcul d'une approximation du minimiseur de J consiste à effectuer les itérations suivantes :

$$y^{(k+1)} \quad = \quad \sum_{i=1}^{n} \frac{x^{(i)}}{\|x^{(i)} - y^{(k)}\|} \quad \Big/ \quad \sum_{i=1}^{n} \frac{1}{\|x^{(i)} - y^{(k)}\|}$$

en partant d'un point $y^{(0)}$ quelconque. Écrire une fonction Python MinimiseJ(x,N) qui effectue N itérations de cet algorithme et renvoie l'approximation du minimiseur obtenue. x est supposé être un tableau de taille $n \times d$ contenant les coordonnées des points $x^{(i)}$.

Correction

```
def MinimiseJ(x,N):
    d = x.shape[1]
    y = np.zeros((n,1))
    for k in range(N):
        # calcul des ||x^(i)-y|| :
        norms = np.linalg.norm(x-y,axis=1).reshape((n,1))
        # mise à jour de y :
        y = np.sum(x/norms,axis=0) / np.sum(1/norms)
    return y
```

Voici un code pour tester : on choisit $n = 10$ points aléatoires dans $\mathbb{R}^2$; on calcule le minimiseur de J grâce à l'algorithme (ce point s'appelle **médiane géométrique** des points $x^{(i)}$) et on affiche les points, la médiane ainsi que le barycentre pour comparer.

```
import numpy as np
import matplotlib.pyplot as plt
n = 10
x = np.random.rand(n,2)
med_x = MinimiseJ(x,N=100)
```

```
bar_x = np.mean(x,axis=0)
plt.plot(x[:,0],x[:,1],'o', label='points x^(i)')
plt.plot(med_x[0],med_x[1],'xr', label='médiane')
plt.plot(bar_x[0],bar_x[1],'+g', label='barycentre')
plt.legend()
plt.show()
```