Université Paris Cité
UFR de Mathématiques et Informatique
45, rue des Saints-Pères, 75006, Paris.

Master $1^{\text{ère}}$ année, MMA, 2022-2023
Optimisation

Examen du 11/01/2024

Durée 2h. Aucun document n'est autorisé. Les exercices sont indépendants.

Total : 43 dont 4.5pt hors barème.

Exercice 1 (19.5pt)

Soit $n \in \mathbb{N}^*$ et $A \in \mathcal{M}_n(\mathbb{R})$ une matrice symétrique, $b \in \mathbb{R}^n$ et $c \in \mathbb{R}$. On pose $g : x \in \mathbb{R}^n \mapsto \frac{1}{2}\langle Ax,\, x\rangle - \langle b,\, x\rangle + c$ et $f = e^g$.

1. On suppose dans cette question que A est définie positive.
   a) Déterminer pour tout $x \in \mathbb{R}^n$, $\nabla f(x)$.
   b) Déterminer pour tout $x \in \mathbb{R}^n$, $\nabla^2 f(x)$.
   c) Montrer que f admet un minimum global sur $\mathbb{R}^n$.
   d) Déduire des précédentes questions qu'il existe $m > 0$ tel que f est m-fortement convexe sur $\mathbb{R}^n$.

2. On considère dans cette question que $n = 3$, $A = \begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & -1 \end{pmatrix}$, $b = (1, -2, 1)$, $c = \frac{5}{2}$. Soit $C = \{x = (x_1, x_2, x_3) \in \mathbb{R}^3 \,/\, x_3 = 0\}$ et $U = \mathrm{Vect}((1, 1, 0))$.
   a) Montrer que C et U sont des convexes fermés.
   b) Montrer que f est strictement convexe sur C. En déduire que f est strictement convexe sur U.
   c) Est-ce que f est convexe sur $\mathbb{R}^3$?
   d) En vous aidant des ensembles de niveau de f dans C, proposez un candidat pour le minimum global de f sur U.
   e) Montrer que c'est bien l'unique minimum global de f sur U.

Correction.
19.5 = (1.5 + 2.5 + 2.5 + 2.5) + (0.5 + 2.5 + 2 + 3.5 + 2)

1. a) g est polynomiale en ses coefficients donc g est $\mathcal{C}^\infty$ sur $\mathbb{R}^n$. Par composée de fonctions $\mathcal{C}^\infty$, on a donc f $\mathcal{C}^\infty$ sur $\mathbb{R}^n$.
   Soit $x, h \in \mathbb{R}^n$, par formule de la différentielle d'une composée, on a $df(x)(h) = e^{g(x)} dg(x)(h) = f(x)\langle Ax - b,\, h\rangle$.
   Par conséquent $\nabla f(x) = f(x)(Ax - b)$ (1.5pt).
   b) Soit $x, h \in \mathbb{R}^n$, on a

$$\begin{aligned}
\nabla f(x+h) = f(x+h)(A(x+h) - b) &= \nabla f(x) + df(x)(h)(Ax - b) + f(x)Ah + o_{h\to 0}(h),\\
&= \nabla f(x) + f(x)\langle Ax - b,\, h\rangle (Ax - b) + f(x)Ah + o_{h\to 0}(h),\\
&= \nabla f(x) + f(x)\left(\underbrace{(Ax-b)^T h}_{\in\mathbb{R}}(Ax - b) + Ah\right) + o_{h\to 0}(h),\\
&= \nabla f(x) + f(x)\left(\underbrace{(Ax-b)(Ax-b)^T}_{\in\mathcal{M}_n(\mathbb{R})} h + Ah\right) + o_{h\to 0}(h),\\
&= \nabla f(x) + f(x)\left((Ax-b)(Ax-b)^T + A\right)h + o_{h\to 0}(h).
\end{aligned}$$

   Donc par identification, on a $\nabla^2 f(x) = f(x)\left((Ax-b)(Ax-b)^T + A\right)$ (2.5pt).

c) <u>Méthode 1.</u> Comme A est symétrique définie positive, A est diagonalisable dans une base orthonormée et en notant λ la plus petite valeur propre de A, on a $\lambda > 0$ et pour tout $x \in \mathbb{R}^n$, $\langle Ax,\, x\rangle \geqslant \lambda \|x\|_2^2$.
Pour tout $x \in \mathbb{R}^n$, par l'inégalité de Cauchy-Schwarz, on a $-\langle b,\, x\rangle \geqslant -\|b\|_2 \|x\|_2$.
Soit $x \in \mathbb{R}^n$, alors comme la fonction exponentielle est croissante, on a $f(x) \geqslant e^{\lambda\|x\|_2^2 - \|b\|_2\|x\|_2 + c} \underset{\|x\|_2 \to +\infty}{\longrightarrow} +\infty$. D'où f est coercive. De plus f est continue sur $\mathbb{R}^n$, donc f admet un minimum global (2.5pt).
<u>Méthode 2.</u> Comme A est symétrique définie positive, A est inversible et donc $A^{-1}b$ est un point critique de f (0.5pt).
Soit $x \in \mathbb{R}^n$. La matrice $(Ax-b)(Ax-b)^T$ est symétrique positive car pour tout $h \in \mathbb{R}^n$, $h^T(Ax-b)(Ax-b)^T h = \langle Ax-b,\, h\rangle^2 \geqslant 0$. Donc comme $f(x) > 0$, on a $\nabla^2 f(x) \succeq f(x)A$. Et comme A est symétrique définie positive, on a $\nabla^2 f(x) \succ 0$ i.e. $\nabla^2 f(x)$ est symétrique définie positive. Ainsi f est strictement convexe sur $\mathbb{R}^n$ (1.5pt).
Comme f admet un point critique, ce point critique est donc nécessairement un minimum global de f sur $\mathbb{R}^n$ (0.5pt).

> Grâce à cette deuxième méthode, nous savons même que ce minimum global est unique.

Suite à la correction de vos copies, voici une troisième méthode que vous avez pu proposer. La fonction g est quadratique et A est définie positive, donc g admet un unique minimum global sur $\mathbb{R}^n$. Notons le $x^\star$. Alors pour tout $x \in \mathbb{R}^n$, $g(x) \geqslant g(x^\star)$. Et comme la fonction exponentielle est croissante, on a donc pour tout $x \in \mathbb{R}^n$, $f(x) \geqslant f(x^\star)$ i.e. $x^\star$ est aussi un minimum global de f sur R^n.

d) Notons alors I la valeur minimale de f sur $\mathbb{R}^n$ (on sait qu'elle existe et est atteinte par f puisque f admet un minimum global sur $\mathbb{R}^n$). Comme f est à valeurs strictement positives, on a $I > 0$. De plus comme A est symétrique définie positive, A est diagonalisable dans une base orthonormée et en notant λ la plus petite valeur propre de A, on a $\lambda > 0$ et pour tout $h \in \mathbb{R}^n$, $\langle Ah,\, h\rangle \geqslant \lambda\|h\|_2^2$.
Soit $x \in \mathbb{R}^n$ et $h \in \mathbb{R}^n$, on a $\left\langle \nabla^2 f(x)h,\, h\right\rangle = f(x)\left\langle (Ax-b)(Ax-b)^T h + Ah,\, h\right\rangle \geqslant I(\langle Ax-b,\, h\rangle^2 + \langle Ah,\, h\rangle) \geqslant I\lambda\|h\|_2^2$. En notant $m = I\lambda > 0$, on a montré que $\nabla^2 f(x) \succeq mI_n$, i.e. f est m-fortement convexe sur $\mathbb{R}^n$ (2.5pt).

2. a) C et U sont des sous espaces vectoriels de dimension finie donc sont des ensembles convexes fermés (de $\mathbb{R}^3$) (0.5pt).
Remarque post-correction de copies : vous devez savoir qu'un $\mathbb{R}$-sev de dimension finie est fermé. Et comme un $\mathbb{R}$-sev est stable par combinaison linéaire, c'est nécessairement un ensemble convexe.
b) Soit $x, y \in C$, $x \neq y$, alors $x_3 = y_3 = 0$ et donc $A(y-x) = (y_1 - x_1, y_2 - x_2, 0) = y - x$. D'où
$$\left\langle \nabla^2 f(x)(y-x),\, y-x\right\rangle = f(x)(\langle Ax-b,\, y-x\rangle^2 + \langle A(y-x),\, y-x\rangle) \geqslant f(x)\|y-x\|_2^2 > 0.$$
D'où f est strictement convexe sur C (1.75pt).
Comme $U \subset C$, U convexe et f strictement convexe sur C, alors f est strictement convexe sur U (0.75pt).
c) Supposons par l'absurde que f est convexe sur $\mathbb{R}^3$, alors f est convexe sur le sous ensemble convexe $\mathrm{Vect}((0,0,1))$. Pour tout $t \in \mathbb{R}$, $f(0,0,t) = e^{-\frac{t^2}{2} - t + \frac{5}{2}} \underset{|t| \to +\infty}{\longrightarrow} 0$. Ainsi il existe $t_1 < 0$ et $t_2 > 0$ tels que $f(0,0,t_1) < e^{\frac{5}{2}}$ et $f(0,0,t_2) < e^{\frac{5}{2}}$. Pour $\mu = \frac{-t_1}{t_2 - t_1} \in]0,1[$, on a $(1-\mu)t_1 + \mu t_2 = 0$ et donc $f((1-\mu)(0,0,t_1) + \mu(0,0,t_2)) = f(0,0,0) = e^{\frac{5}{2}} > (1-\mu)f(0,0,t_1) + \mu f(0,0,t_2)$. Contradiction avec la convexité de f sur $\mathrm{Vect}((0,0,1))$, d'où f n'est pas convexe sur $\mathbb{R}^3$ (2pt).
d) Soit $x = (x_1, x_2, x_3) \in C$, alors $x_3 = 0$ et $g(x) = \frac{1}{2}\left(x_1^2 + x_2^2\right) - x_1 + 2x_2 + \frac{5}{2} = \frac{1}{2}(x_1 - 1)^2 + \frac{1}{2}(x_2 + 2)^2$.
Soit $\lambda \in \mathbb{R}$, alors l'ensemble de niveau λ de f dans C est
$$L_\lambda \cap C = \{(x_1, x_2, 0) \,/\, x_1, x_2 \in \mathbb{R},\, e^{\frac{1}{2}(x_1-1)^2 + \frac{1}{2}(x_2+2)^2} = \lambda\}.$$
Cet ensemble est non vide si et seulement si $\lambda \geqslant 1$ et alors $L_\lambda \cap C$ représente un cercle dans C, de rayon $2\log(\lambda)$ et de centre $(1, -2, 0)$ (1.5pt).
2pt pour le dessin et trouver $x^\star$.
<u>Analyse :</u> si $x^\star \in U \subset C$ est un minimum global de f sur U, alors comme U est un sous espace vectoriel, on a nécessairement $\nabla f(x^\star) \perp U$. Comme les lignes de niveau de f dans C sont des cercles et que $\nabla f(x^\star)$ est orthogonal en $x^\star$ à la ligne de niveau de f qui passe par $x^\star$, alors graphiquement $x^\star$ est le point de U qui se situe sur le cercle de centre $(1, -2, 0)$ dont la droite U est une tangente. Voir la figure 1. Analytiquement, cela mène donc à chercher un point $x^\star = (x_1^\star, x_2^\star, 0) \in U$, donc satisfaisant $x_1^\star = x_2^\star$, tel que $x^\star - (1, -2, 0) \perp U$. Or un vecteur normal à U dans C est par exemple $(-1, 1, 0)$, il faut donc trouver

$x^\star = (x_1^\star, x_2^\star, 0)$, avec $x_1^\star = x_2^\star$ tel que $x^\star - (1,-2,0) = (x_1^\star - 1, x_2^\star + 2, 0)$ soit colinéaire à $(-1,1,0)$ i.e. il existe $t \in \mathbb{R}$ satisfaisant

$$\begin{aligned} x_1^\star &= x_2^\star, \\ x_1^\star - 1 &= -t, \\ x_2^\star + 2 &= t. \end{aligned}$$

En résolvant, on trouve nécessairement $x_1^\star = x_2^\star = -0.5$. Donc le candidat pour un minimum global de f sur U est le point $(-0.5,-0.5,0)$.

> Il ne manquerait pas grand chose à ajouter pour que le raisonnement par analyse fait au-dessus devienne une preuve que $(-0.5,-0.5,0)$ est l'unique minimum global de f sur U. Cette analyse est en fait très détaillée par soucis pédagogique. En pratique on peut aller très vite lors de cette phase et comme on va le voir dans la question suivante, la démonstration effective une fois la conjecture faite est simple et rapide.

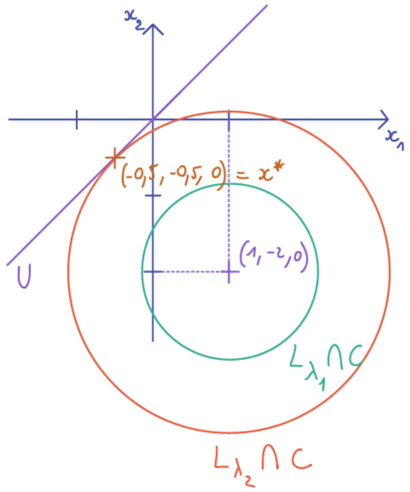

Figure 1. Représentation des lignes de niveau de f dans C (plan d'équation cartésienne $x_3 = 0$). Par construction on a $\lambda_2 > \lambda_1 > 1$ et λ_2 est choisi de sorte que le cercle associé $L_{\lambda_2} \cap C$ ait pour tangente la droite U. Le point d'intersection entre $L_{\lambda_2} \cap C$ est alors nécessairement l'unique minimum global de f sur U comme cela est démontré rigoureusement dans la question 2.e).

e) Notons donc $x^\star = (-0.5,-0.5,0)$. Comme f est strictement convexe sur U, $x^\star$ est l'unique minimum global de f sur U si et seulement si il satisfait la condition d'optimalité

$$\forall y \in U, \quad \langle \nabla f(x^\star),\, y - x^\star \rangle = 0. \quad (1pt) \tag{1}$$

Soit $y \in U$, donc $y = (u,u,0)$ avec $u \in \mathbb{R}$. On a

$$\begin{aligned} \langle \nabla f(x^\star),\, y - x^\star \rangle &= f(x^\star)\,\langle Ax^\star - b,\, (u+0.5, u+0.5, 0)\rangle\,, \\ &= f(x^\star)\,\langle (-0.5,-0.5,0) - (1,-2,1),\, (u+0.5,u+0.5,0)\rangle\,, \\ &= f(x^\star)\,(-1.5(u+0.5) + 1.5(u+0.5)) = 0. \ (1pt) \end{aligned}$$

> En toute généralité la condition d'optimalité sur un convexe U s'écrit $\forall y \in U$, $\langle \nabla f(x^\star),\, y - x^\star\rangle \geqslant 0$. Mais comme ici U est un sous espace vectoriel, c'est équivalent à remplacer le "$\geqslant 0$" par "$= 0$" dans (1).

Exercice 2 (23.5pt)

Soit $M, n \in \mathbb{N}^*$. On considère une collection de points $(x_i)_{1 \leqslant i \leqslant M}$ de $[0,1]$ deux à deux distincts et des valeurs réelles associées $(y_i)_{1 \leqslant i \leqslant M}$. Une telle collection de couples (x_i, y_i) est représentée en exemple sur la figure 2 suivante.

On souhaite trouver une fonction $f : [0,1] \to \mathbb{R}$ de la forme suivante

$$\forall x \in [0,1], \quad f(x) = \sum_{k=1}^{n} a_k e^{-\frac{(x-p_k)^2}{2\sigma^2}}, \tag{2}$$

où

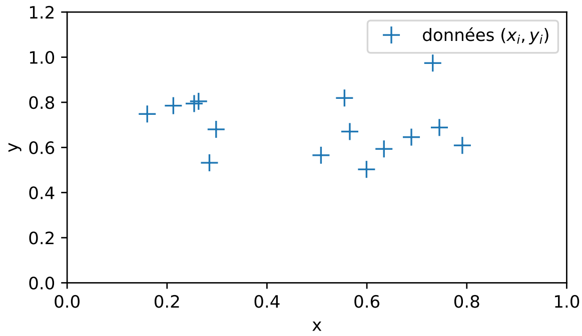

FIGURE 2. Représentation des points x_i de $[0, 1]$ et des valeurs y_i associées, pour $1 \leqslant i \leqslant M$.

— $\sigma > 0$, $0 \leqslant p_1 < p_2 < \ldots < p_n \leqslant 1$ sont des données fixées du problème,

— les poids $a_1, \ldots, a_n \in \mathbb{R}$ sont les inconnus du problème paramétrant la fonction f,

qui interpole le mieux possible les données précédentes et de manière simple. Plus précisément, on souhaite que f satisfasse les contraintes $f(x_i) \simeq y_i$ pour tout $i \in \{1, \ldots, M\}$ avec le maximum de coefficients a_k nuls. Voir la figure 3 pour des représentations de fonctions du type donné dans (2).

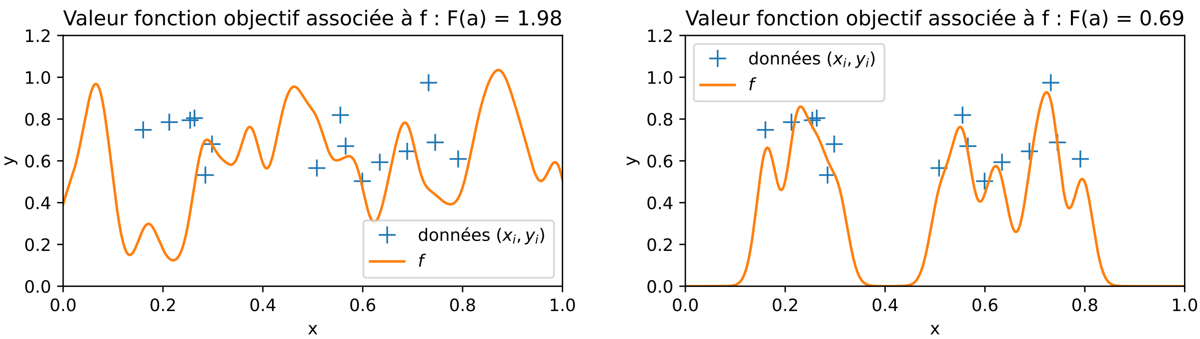

FIGURE 3. Représentation de deux fonctions de la forme donnée dans (2). A gauche les poids $a_1, a_2, \ldots, a_n \in \mathbb{R}$ sont aléatoires, à droite les poids $a = (a_k)_{1 \leqslant k \leqslant n}$ sont bien choisis pour assurer une interpolation satisfaisant nos critères.

La fonction f de (2), qui est donc une somme pondérée par les poids $a_k \in \mathbb{R}$ de gaussiennes centrées en les points $p_k \in [0, 1]$, peut se réécrire de la manière suivante

$$\forall x \in [0,1], \quad f(x) = \sum_{k=1}^{n} a_k e^{-\frac{(x-p_k)^2}{2\sigma^2}} = \sum_{k=1}^{n} a_k u_k(x) = \langle u(x),\, a\rangle\,,$$

où pour tout $k \in \{1, \ldots, n\}$, $u_k : x \in [0, 1] \mapsto e^{-\frac{(x-p_k)^2}{2\sigma^2}}$ est la gaussienne centrée en p_k et pour tout $x \in [0, 1]$, $u(x) = (u_k(x))_{1 \leqslant k \leqslant n} \in \mathbb{R}^n$.

Pour trouver la meilleure fonction interpolatrice comme décrite précédemment, on va s'intéresser au problème d'optimisation suivant

$$\inf_{a \in \mathbb{R}^n} F(a), \tag{3}$$

où

$$\forall a \in \mathbb{R}^n, \quad F(a) = \frac{1}{2} \sum_{i=1}^{M} (\langle u(x_i),\, a\rangle - y_i)^2 + \lambda \sum_{k=1}^{n} N_\varepsilon(a_k). \tag{4}$$

avec $\lambda > 0$ et $N_\varepsilon : t \in \mathbb{R} \mapsto \sqrt{t^2 + \varepsilon}$, $\varepsilon > 0$.

Résumé.

— Données du problème :

— les points à interpoler $(x_i, y_i) \in [0, 1] \times \mathbb{R}$, $i \in \{1, \ldots, M\}$,

— $\sigma > 0$, $0 \leqslant p_1 < p_2 < \ldots < p_n \leqslant 1$ les paramètres des gaussiennes.

— $\lambda > 0$, $\varepsilon > 0$.

— Objectif : trouver le meilleur vecteur $a \in \mathbb{R}^n$ définissant une fonction interpolante $f : x \in [0,1] \mapsto \langle u(x),\, a\rangle$ en résolvant le problème (3).

Indication : on pourra utiliser les notations N_ε, N'_ε, N''_ε quand cela est pertinent afin d'alléger les calculs.

1. On suppose, dans cette question uniquement, que $n = 5$, $p_1 = 0$, $p_2 = 1/4$, $p_3 = 1/2$, $p_4 = 3/4$ et $p_5 = 1$. Représenter sur un même graphe l'allure des fonctions u_2 et $2u_4$ lorsque $\sigma = 0.1$.
2. On pose $G : a \in \mathbb{R}^n \mapsto \frac{1}{2}\sum_{i=1}^M (\langle u(x_i),\, a\rangle - y_i)^2$. Quel est l'intérêt, vis à vis d'une éventuelle solution de (3), d'avoir le terme $G(a)$ dans l'expression de F ?
3. On pose $H : a \in \mathbb{R}^n \mapsto \sum_{k=1}^n N_\varepsilon(a_k)$. Montrer que H est une fonction coercive.
4. Déterminer pour tout $a \in \mathbb{R}^n$, $\nabla F(a)$. *Indication : on pourra utiliser le fait que pour tout $k \in \{1, \dots, n\}$, $a_k = \langle e_k,\, a\rangle$, où e_k est le k-ème vecteur de la base canonique de $\mathbb{R}^n$.*
5. Déterminer pour tout $a \in \mathbb{R}^n$, $\nabla^2 F(a)$.
6. Montrer que F est strictement convexe sur $\mathbb{R}^n$.
7. Montrer que F admet un unique minimum global sur $\mathbb{R}^n$.
8. Montrer que les ensembles de sous-niveau de F, lorsqu'ils sont non vides, sont des ensembles convexes et compacts de $\mathbb{R}^n$.
9. Soit $a \in \mathbb{R}^n$, $d \in \mathbb{R}^n$ telle que $\langle \nabla F(a),\, d\rangle < 0$.
   a) Montrer que d est une direction de descente pour F en a.
   b) Rappeler la définition de la condition d'Armijo en a dans la direction d pour F.
   c) Illustrer cette condition par un dessin.
10. Écrire en Python une fonction renvoyant la valeur du pas d'Armijo en $a \in \mathbb{R}^n$ dans la direction de descente $-\nabla F(a)$ pour F. Cette fonction prendra en argument les variables que vous jugerez nécessaires.
11. Montrer que la convergence de l'algorithme de descente de gradient à pas d'Armijo vers l'unique minimum global de F est assurée.
12. Bonus. Écrire en Python deux fonctions renvoyant respectivement $F(a)$ et $\nabla F(a)$ pour $a \in \mathbb{R}^n$ de sorte que leurs implémentations sont vectorielles, i.e. sans aucunes boucles for. Bonus**2. On suppose que $M > n$. Trouvez des implémentations qui minimisent le nombre d'opérations. Vous pourrez précalculer en amont de vos fonctions les quantités que vous jugerez nécessaires.

    Indication : vous pourrez écrire sur votre feuille les calculs que vous jugerez pertinents.

Correction.

$21.5 = 0.5 + 1 + 1 + 2.5 + 2 + 1.75 + 1.25 + 2.5 + (1 + 0.5 + 1) + 2.5 + 2 + 4$

1. Voir la figure 4 (0.5pt).

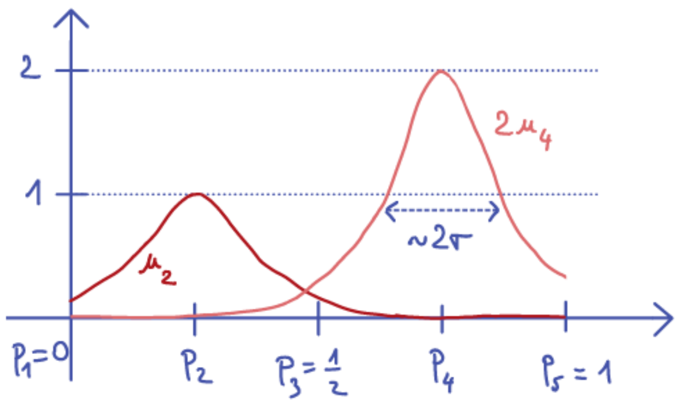

FIGURE 4. Représentation des fonctions u_2 et $2u_4$.

2. Comme F est somme de termes positifs, si F admet un minimum global, noté $a^\star$ alors $G(a^\star)$ sera "petit" et "proche" de 0, i.e. pour tout $i \in \{1, \dots, M\}$, on aura $f(x_i) = \langle u(x_i),\, a^\star\rangle \simeq y_i$. C'est ce que l'on souhaite : interpoler approximativement les valeurs y_i en les x_i par une fonction du type de f (1pt).

3. On a pour tout $a \in \mathbb{R}^n$, $H(a) = \sum_{i=1}^n \sqrt{a_i^2 + \varepsilon} \geqslant \sum_{i=1}^n |a_i| = \|a\|_1 \underset{\|a\|_1 \to +\infty}{\longrightarrow} +\infty$. Or comme toutes les normes sont équivalentes en dimension finie, on a bien $H(a) \underset{\|a\|_2 \to +\infty}{\longrightarrow} +\infty$, i.e. H est coercive (1pt).

4. La fonction N_ε est $\mathcal{C}^\infty$ sur $\mathbb{R}$ donc H est $\mathcal{C}^\infty$ sur $\mathbb{R}^n$ par composée. Pour $a \in \mathbb{R}^n$, $G(a)$ est polynomiale en les coefficients de a donc G est $\mathcal{C}^\infty$ sur $\mathbb{R}^n$. Par somme F est $\mathcal{C}^\infty$ sur $\mathbb{R}^n$ (0.5pt).

   Soit $a, h \in \mathbb{R}^n$, en différenciant $G(a+h)$ et $H(a+h)$ on trouve que

$$\nabla G(a) = \sum_{i=1}^{M} (\langle u(x_i),\, a\rangle - y_i) u(x_i), \ (1pt)$$

et en utilisant le fait que pour tout k, $a_k = \langle e_k,\, a\rangle$, on obtient

$$\nabla H(a) = \sum_{k=1}^{n} N_\varepsilon'(a_k) e_k, \ (1pt)$$

d'où

$$\nabla F(a) = \nabla G(a) + \lambda \nabla H(a) = \sum_{i=1}^{M} (\langle u(x_i),\, a\rangle - y_i) u(x_i) + \lambda \sum_{k=1}^{n} N_\varepsilon'(a_k) e_k.$$

5. Soit $a, h \in \mathbb{R}^n$, en différenciant $\nabla G(a+h)$ et $\nabla H(a+h)$ on trouve que

$$\nabla^2 G(a) = \sum_{i=1}^{M} u(x_i) u(x_i)^T, \ (1pt)$$

et

$$\nabla^2 H(a) = \sum_{k=1}^{n} N_\varepsilon''(a_k) e_k e_k^T, \ (1pt)$$

d'où

$$\nabla^2 F(a) = \nabla^2 G(a) + \lambda \nabla^2 H(a) = \sum_{i=1}^{M} u(x_i) u(x_i)^T + \lambda \sum_{k=1}^{n} N_\varepsilon''(a_k) e_k e_k^T.$$

6. Soit $a, h \in \mathbb{R}^n$, $h \neq 0$, on a

$$\left\langle \nabla^2 F(a) h,\, h \right\rangle = h^T \nabla^2 F(a) h = \sum_{i=1}^{M} \langle u(x_i),\, h\rangle^2 + \lambda \sum_{k=1}^{n} N_\varepsilon''(a_k) h_k^2 \geqslant \lambda \sum_{k=1}^{n} N_\varepsilon''(a_k) h_k^2.$$

Or pour tout $t \in \mathbb{R}$, $N_\varepsilon'(t) = \frac{t}{N_\varepsilon(t)}$ et $N_\varepsilon''(t) = \frac{\varepsilon}{N_\varepsilon(t)^3} > 0$ (en particulier la fonction N_ε est strictement convexe sur $\mathbb{R}$). Ainsi

$$\left\langle \nabla^2 F(a) h,\, h \right\rangle > 0,$$

et donc F est strictement convexe sur $\mathbb{R}^n$ (1.75pt).

7. On a pour tout $a \in \mathbb{R}^n$, $F(a) \geqslant \lambda H(a)$ car $G(a) \geqslant 0$. Donc comme H est coercive, on déduit que F est coercive également. Puisque F est continue sur $\mathbb{R}^n$, F admet donc un minimum global sur $\mathbb{R}^n$. Il est unique par stricte convexité de F sur $\mathbb{R}^n$ (1.25pt).

8. Soit $\gamma \in \mathbb{R}$, tel que l'ensemble de sous niveau γ : $C_\gamma = \{a \in \mathbb{R}^n \,/\, F(a) \leqslant \gamma\} = F^{-1}(]-\infty, \gamma])$ soit non vide.

   Voir le cours pour la démonstration que C_γ est convexe (il suffit de montrer que la définition d'un ensemble convexe est satisfaite pour C_γ) (1pt).

   Comme F est continue, C_γ est fermé dans $\mathbb{R}^n$ comme l'image réciproque d'un fermé (de $\mathbb{R}$) (0.5pt).

   Comme F est coercive, il existe $r > 0$ tel que pour tout $a \in \mathbb{R}^n$ satisfaisant $\|a\|_2 > r$, $F(a) > \gamma$. D'où $C_\gamma \subset \bar{B}(0, r)$, i.e. C_γ est borné. Ainsi C_γ est un fermé borné de $\mathbb{R}^n$, donc est compact dans $\mathbb{R}^n$ (1pt).

9. a) Grâce à la condition de l'énoncé on montre que $g : t \in \mathbb{R}_+ \mapsto F(a + td)$ (dérivable) est strictement décroissante au voisinage de 0 car $g'(0) = \langle \nabla F(a),\, d\rangle < 0$ (voir le cours pour le calcul complet), donc d est bien une direction de descente pour F en a (1pt).

   b) Soit $\alpha \in]0, \frac{1}{2}[$. On dit que $t \geqslant 0$ satisfait la condition d'Armijo si l'inégalité

$$F(a + td) \leqslant F(a) + \alpha t \langle \nabla F(a),\, d\rangle, \ (0.5pt)$$

   est satisfaite.

c) La pente de la droite $\kappa : t \in \mathbb{R} \mapsto F(a) + \alpha t \langle \nabla F(a),\, d\rangle$ vaut $\alpha \langle \nabla F(a),\, d\rangle$, donc est strictement négative mais supérieure à la pente de la tangente en 0 de la fonction g définie dans la question précédente (puisque $0 < \alpha < \frac{1}{2}$). Un pas $t \geqslant 0$ satisfait la condition d'Armijo si la graphe de g en t est en dessous du graphe de κ en t. Voir la figure 5 (1pt).

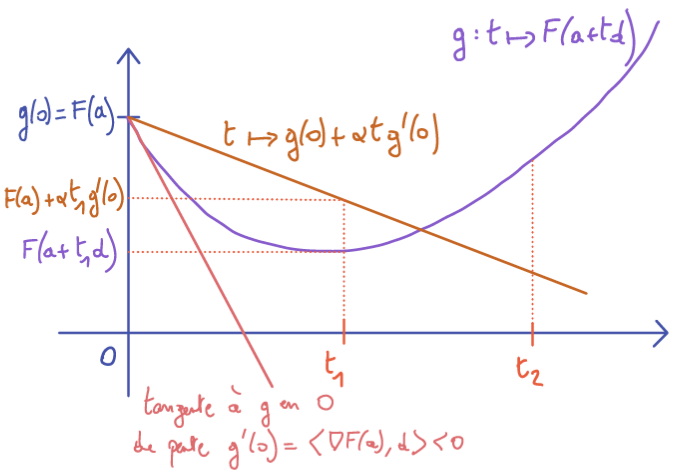

FIGURE 5. Illustration de la condition d'Armijo. Le réel t_1 satisfait la condition d'Armijo, mais pas t_2.

10. Voir le cours et TP.
0.5pt pour les paramètres $\alpha \in]0, \frac{1}{2}[$ et $\beta \in]0, 1[$,
0.5pt pour les autres paramètres : F, a, ∇F,
1pt pour la boucle,
Bonus +0.5pt si implémentation "efficace" où notamment les paramètres F, ∇F sont remplacés par F, a, d et θ, avec d et θ définies au préalable comme $d = -\nabla F(a)$ et $\theta = \langle \nabla F(a),\, \nabla F(a)\rangle$. Donner ces paramètres à la fonction évitent de calculer, à chaque passage dans la boucle, l'évaluation du gradient de F en a et le produit scalaire $\langle \nabla F(a),\, d\rangle$. De plus dans le cas où cet algorithme est intégré dans une descente de gradient, les quantités $d = -\nabla F(a)$ et θ sont de toute façon à calculer (θ pour la condition d'arrêt globale de l'algorithme).

11. Soit $a^{(0)} \in \mathbb{R}^n$ qui représente la localisation d'initialisation de la descente de gradient. On a vu que F est $\mathcal{C}^2$ et (strictement) convexe sur $\mathbb{R}^n$. On a vu également que l'ensemble de sous niveau $C_{F(a^{(0)})}$ est un compact de $\mathbb{R}^n$. On a démontré que pour tout $a \in \mathbb{R}^n$, $\nabla^2 F(a) \succ 0$, donc c'est aussi le cas sur $C_{F(a^{(0)})}$. Par conséquent, on sait d'après le cours que les hypothèses du théorème de convergence de la méthode de descente de gradient à pas d'Armijo sont satisfaites : il existe $0 < m \leqslant M$ tels que pour tout $a \in C_{F(a^{(0)})}$ (qui est compact)

$$mI_n \preceq \nabla^2 F(a) \preceq MI_n.$$

Ainsi la suite $(a^{(k)})_{k\in\mathbb{N}}$ construite par récurrence grâce à la méthode converge vers l'unique minimum global de F (2pt).

> On sait alors que la suite $(F(a^{(k)}))_{k\in\mathbb{N}}$ converge linéairement vers la valeur minimale de F sur $\mathbb{R}^n$.

12. Voir le code fourni dans le notebook (4pt).

A noter que l'implémentation fourni de $F(a)$ et $\nabla F(a)$ n'est finalement qu'une implémentation vectorielle qui pourrait encore être amélioré.