Partiel du 03/03/2020 : éléments de correction

Les documents sont interdits. La qualité de la rédaction sera prise en compte dans la correction. Vous pouvez bien sûr admettre les résultats d'une question pour traiter les questions suivantes.

La partie numérique est disponible dans le répertoire `/commun/doc/merigot/M315.ipynb`. Elle est constituée de 3 parties indépendantes, qui peuvent être traitées après les Questions 3, 5 et 6.

- **Le fichier `M315.ipynb` doit être copié dans le répertoire `reMise_copie` sur le bureau et renommé `M315-NOMPRENOM.ipynb`.**
- **Vous devez également indiquer votre n° de poste (`pXXeXX`) sur votre copie.**

I. Méthode de Newton pure en dimension 1

Soit $g \in \mathcal{C}^3(\mathbb{R})$ une fonction vérifiant

(i) $\exists \eta > 0, \quad g'' \geq \eta$ sur $\mathbb{R}$,

(ii) $g''' \geq 0$ sur $]-\infty, p]$ et $g''' \leq 0$ sur $[p, +\infty[$.

On s'intéresse au problème de minimisation suivant

$$\min_{t \in \mathbb{R}} g(t). \tag{1}$$

que l'on cherche à résoudre par une méthode de Newton pure, soit

$$t^{(0)} = p, \qquad t^{(k+1)} = N(t^{(k)}) \text{ où } N(t) = t - g'(t)/g''(t). \tag{2}$$

Question 1. En utilisant l'hypothèse (i), montrer que $\lim_{t \to \pm\infty} g(t) = +\infty$. Justifier l'existence d'un unique minimiseur t^* de (1), puis que $N(t^*) = t^*$.

Réponse: *Par formule de Taylor-Lagrange, pour tout $t \geq 0$, il existe $s \in [0, t]$ tel que*

$$g(t) = g(0) + tg'(0) + \frac{t^2}{2} g''(s) \quad \geq g(0) + tg'(0) + \frac{t^2}{2} \eta$$

On en déduit que $\lim_{t \to +\infty} g(t) = +\infty$. La limite en $-\infty$ se montre de la même anière. Comme la fonction g est continue et vérifie $\lim_{|x| \to +\infty} g = +\infty$, par proposition du cours, elle admet un minimiseur sur $\mathbb{R}$. Comme $g'' > 0$, la fonction g est strictement convexe et le minimiseur est donc unique, on le note t^*. Enfin, comme g est convexe t^* est minimiseur si et seulement si $g'(t^*) = 0$ ssi $N(t^*) = t^*$ par définition de N.

Question 2. *[Convergence de l'algorithme de Newton]*
Dans cette question, on suppose $t^* > p$.

1. Montrer que $g' \leq 0$ sur $I = [p, t^*]$.
   Réponse: *Par croissance des pentes d'une fonction convexe, pour $t \leq t^*$ on a $g'(t) \leq g'(t^*) = 0$.*

2. En déduire que $t^{(0)} \leq t^{(1)} = N(t^{(0)})$.
   Réponse: *On a*
   $$t^{(1)} = t^{(0)} - \frac{g'(t^{(0)})}{g''(t^{(0)})} \leq t^{(0)}$$
   car $g'(t^{(0)}) = g'(p) \leq 0$ (question précédente et hypothèse $p < t^*$) et $g'' > 0$.
3. Calculer N', et montrer que N est croissant sur I.
   Réponse: *Un calcul simple montre que*
   $$N'(t) = 1 - \frac{g''(t)^2 - g'(t)g'''(t)}{g''(t)^2} = \frac{g'(t)g'''(t)}{g''(t)^2}.$$
   Or, par l'hypothèse (ii), $g'' \leq 0$ sur $[p, +\infty[$, et par la question Q2.1, $g'(t) \leq 0$ sur $[p, t^*]$. On en déduit que $N' \geq 0$ sur $[p, t^*]$, prouvant ainsi que N est croissante sur cet ensemble.
4. Montrer par récurrence que $\forall k \geq 1, t^{(k)} \leq t^{(k+1)} \leq t^*$.
   Réponse: *Par Q2.2., on a $t^{(0)} \leq t^{(1)}$. D'autre part, comme $p = t^{(0)} \leq t^*$ (hypothèse), la croissance de N sur $[p, t^*]$ nous donne $t^{(1)} = N(p) \leq N(t^*)$, et par Q1, $N(t^*) = t^*$, donc $t^{(1)} \leq t^*$. On a montré l'initialisation de la récurrence.*

   Supposant donc la propriété vraie jusqu'au rang $k > 0$: ceci implique
   $$p = t^{(0)} \leq t^{(1)} \leq \ldots \leq t^{(k-1)} \leq t^{(k)} \leq t^*.$$
   Ainsi, $t^{(k)} \in [p, t^*]$, et N est croissante sur cet intervalle, donc
   $$N(t^{(k-1)}) \leq N(t^{(k)}) \leq N(t^*),$$
   soit $t^{(k)} \leq t^{(k+1)} \leq t^*$.
5. Déduire de ce qui précède que la suite $(t^{(k)})$ admet une limite $\bar{t}$ lorsque $k \to +\infty$.
   Réponse: *La suite est croissante et majorée et admet donc une limite, notée $\bar{t}$*
6. Montrer que $N(\bar{t}) = \bar{t}$, et en déduire que $\bar{t} = t^*$. Conclure.
   Réponse: *La suite $(t^{(k)})$ converge vers $\bar{t}$, ainsi,*
   $$\bar{t} = \lim_{k \to +\infty} t^{(k)} = \lim_{k \to +\infty} t^{(k+1)} = \lim_{k \to +\infty} N(t^{(k)}) = N(\bar{t}),$$
   où l'on a utilisé la continuité de N. Ainsi,
   $$\bar{t} = N(\bar{t}) = \bar{t} - \frac{g'(\bar{t})}{g''(\bar{t})},$$
   ce qui implique $g'(\bar{t}) = 0$. Comme g est convexe, $\bar{t}$ doit être égal à l'unique minimiseur de g sur $\mathbb{R}$ (cf Q1), donc $\bar{t} = t^*$.

Question 3. *[Cas particulier]* On considère la fonction g définie par
$$g(t) = at^2 + bt + \sqrt{\varepsilon + (c+t)^2} \text{ où } \varepsilon > 0, a > 0 \tag{3}$$

1. Calculer g', g'', g'''.
   Réponse:
   $$g'(t) = 2at + b + \frac{c+t}{\sqrt{\varepsilon + (c+t)^2}}$$
   $$g''(t) = 2a + \frac{\varepsilon}{(\varepsilon + (c+t)^2)^{3/2}}$$
   $$g'''(t) = -\frac{3}{2} \frac{2\varepsilon(c+t)}{(\varepsilon + (c+t)^2)^{5/2}}$$

2. Montrer que les hypothèses (i) et (ii) sont vérifiées, avec $p = -c$.
   Réponse: *(i) est vérifiée car $g''(t) \geq 2a$ et (ii) est vérifiée car si $t \geq p := -c$, alors $c+t \geq 0$ et $g'''(t) \leq 0$; de même, si $t \leq p, g'''(t) \geq 0$.*

   La question précédente montre donc que, dans ce cas particulier, la méthode de Newton pure converge vers la solution de (1).

Mettre en œuvre l'algorithme de Newton pur 1D sur ordinateur.

II. Application en dimension $d \geq 1$: problème LASSO

On considère un problème de minimisation sans contrainte sur $\mathbb{R}^d$, issu des statistiques. La matrice A possède n lignes et d colonnes et **vérifie** $Ae_i \neq 0$ **pour tout** $1 \leq i \leq d$, où e_i est le $i$ème vecteur de la base canonique ; y est un vecteur fixé de $\mathbb{R}^n$ et $\mu \in \mathbb{R}_+^*$:

$$\min_{x\in\mathbb{R}^d} f(x) \text{ où } f(x) = \frac{\mu}{2}\|Ax-y\|^2 + \sum_{1\leq i\leq d} s(x_i), s(t) = \sqrt{\varepsilon + t^2} \text{ et } \varepsilon > 0. \tag{4}$$

Question 4. *[Existence et unicité]*

1. Montrer que le problème (4) admet au moins un minimiseur.
   Réponse: *La fonction f est continue (car quadratique), et*

$$\begin{aligned} f(x) &= \frac{\mu}{2}\|Ax-y\|^2 + \sum_{1\leq i\leq d}\sqrt{\varepsilon + x_i^2} \\ &\geq \sum_{1\leq i\leq d}\sqrt{x_i^2} = \|x\|_1 \end{aligned}$$

   Comme $\|\cdot\|_1$ est équivalente à la norme euclidienne, on en déduit que

$$\lim_{\|x\|\to+\infty} f(x) = +\infty.$$

   Ainsi, par proposition du cours, le problème de minimisation (4) admet un minimiseur.
2. Montrer que $\nabla f(x) = \mu(A^TAx - A^Ty) + G(x)$ et $\mathrm{D}^2 f(x) = \mu A^TA + H(x)$ où

$$G(x) = (s'(x_1), \ldots, s'(x_d)) \in \mathbb{R}^d, \qquad H(x) = \begin{pmatrix} s''(x_1) & & \\ & \ddots & \\ & & s''(x_d) \end{pmatrix}.$$

   Donner également l'expression de $s'(t)$ et $s''(t)$.
   Réponse: *Posons $f = g + h$ où*

$$g(x) = \frac{\mu}{2}\|Ax-y\|^2 \text{ et } h(x) = \sum_{1\leq i\leq d} s(x_i).$$

   Nous calculons les dérivées de g et h séparément :

$$\begin{aligned} g(x) &= \frac{\mu}{2}(\|Ax\|^2 + \|y\|^2 - 2\langle Ax|y\rangle) \\ &= \frac{\mu}{2}(\langle Ax|Ax\rangle + \|y\|^2 - 2\langle x|A^Ty\rangle) \\ &= \frac{\mu}{2}(\langle A^TAx|x\rangle + \|y\|^2 - 2\langle x|A^Ty\rangle). \end{aligned}$$

La fonction g est donc quadratique, et ses dérivées sont donc (TD 1) :

$$\nabla g(x) = \mu(A^T A - A^T y) \qquad \mathrm{D}^2 g(x) = \mu A^T A.$$

Pour h, on calcule les dérivées partielles :

$$\frac{\partial h}{\partial x_i} = s'(x_i) \qquad \frac{\partial^2 h}{\partial x_j \partial x_i} = \delta_{ij} s''(x_i),$$

où

$$\begin{aligned} s'(t) &= \frac{t}{\sqrt{\varepsilon + t^2}}, \\ s''(t) &= \frac{\sqrt{\varepsilon + t^2} - t\frac{t}{\sqrt{\varepsilon+t^2}}}{\varepsilon + t^2} = \frac{\varepsilon + t^2 - t^2}{(\varepsilon + t^2)^{\frac{3}{2}}} \\ &= \frac{\varepsilon}{(\varepsilon + t^2)^{\frac{3}{2}}} \end{aligned}$$

3. Démontrer que la matrice $H(x)$ est définie positive, puis que $D^2 f(x)$ l'est aussi.
   Réponse: *La matrice diagonale H est définie positive : en effet, si $v \in \mathbb{R}^d$ et $v \neq 0$, alors il existe i_0 tel que $v_{i_0} \neq 0$, soit*

$$\langle v | H(x) v\rangle = \sum_{1 \leq i \leq d} v_i^2 H(x)_{ii}^2 \geq s''(x_{i_0}) v_{i_0}^2 > 0.$$

   On en déduit que $H(x)$ est définie positive. De plus, $\mu A^T A$ est aussi positif car

$$\forall v \in \mathbb{R}^d, \langle v | \mu A^T A v\rangle = \mu \langle Av | Av\rangle = \mu \left\|Av\right\|^2 \geq 0.$$

   On en déduit que $D^2 f(x)$ est définie positive, comme somme d'une matrice semi-définie positive et d'une matrice définie positive.
4. En déduire que le problème (4) admet une unique solution qu'on notera désormais x^*.
   Réponse: *Par la question précédente, la fonction f est strictement convexe, et elle admet au plus un minimiseur. Comme on sait déjà qu'elle en possède au moins un un (Q4.1), on a bien existence et unicité.*

Question 5. *[Descente de gradient avec pas d'Armijo]*

1. Démontrer que si $f(x) \leq R := f(x^{(0)})$, alors $\sum_{1 \leq i \leq d} |x_i| \leq f(x^{(0)})$. En déduire que l'ensemble $S = \{x \in \mathbb{R}^d \mid f(x) \leq R\}$ est inclus dans $[-R, R]^d$ et qu'il est compact.
   Réponse: *Supposons que $f(x) \leq R$. Alors,*

$$R \geq f(x) = \frac{\mu}{2} \left\|Ax - y\right\|^2 + \sum_{1 \leq i \leq d} s(x_i) \quad \geq \sum_{1 \leq i \leq d} \sqrt{\varepsilon + x_i^2} \geq \sum_{1 \leq i \leq d} |x_i| \geq \max_{1 \leq i \leq d} |x_i|$$

   Ainsi, $x \in [-R, R]^d$. Ceci montre que l'ensemble

$$S = \{x \in \mathbb{R}^d \mid f(x) \leq f(x^{(0)})\}$$

   est inclus dans $[-R, R]^d$ et est donc borné. De plus, il est fermé car f est continue (cela se démontre facilement en utilisant la caractérisation séquentielle de la continuité, ou en utilisant qu'une fonction est continue si et seulement si l'image inverse d'un fermé est fermée).

2. Démontrer que $\forall x \in S, \forall v \in \mathbb{R}^d$, $\frac{\varepsilon}{(\varepsilon+R^2)^{\frac{3}{2}}} \|v\|^2 \leq \langle \mathrm{D}^2 f(x)v|v\rangle \leq \left(\mu \|A\|^2 + \frac{1}{\sqrt{\varepsilon}}\right) \|v\|^2$.
Réponse: *On commence par la majoration*

$$\begin{aligned}
\langle D^2 f(x)v|v\rangle &= \langle \mu A^T Av|v\rangle + \sum_{1\leq i\leq d} s''(x_i)v_i^2 \\
&= \mu \|Av\|^2 + \sum_{1\leq i\leq d} \frac{\varepsilon}{(\varepsilon + x_i^2)^{\frac{3}{2}}} v_i^2 \\
&\leq \left\|A^2\right\| \|v\|^2 + \frac{1}{\varepsilon^{1/2}} \|v\|^2,
\end{aligned}$$

où l'on a utilisé $\|Ax\| \leq \|A\| \|x\|$ par définition de la norme matricielle. Pour la minoration, le calcul est similaire :

$$\begin{aligned}
\langle D^2 f(x)v|v\rangle &= \mu \|Av\|^2 + \sum_{1\leq i\leq d} \frac{\varepsilon}{(\varepsilon + x_i^2)^{\frac{3}{2}}} v_i^2 \\
&\geq \frac{\varepsilon}{(\varepsilon + R^2)^{\frac{3}{2}}} \sum_{1\leq i\leq d} v_i^2,
\end{aligned}$$

où l'on a utilisé $|x_i| \leq R$.

3. En invoquant et citant un théorème du cours, montrer que les itérées $x^{(k)}$ de l'algorithme de descente de gradient avec pas d'Armijo convergent vers x^*. À quelle vitesse ?

Mettre en œuvre l'algorithme de descente de gradient sur ordinateur.

Question 6. *[Minimisation sur une coordonnée]* Soit $g(t) = f(x + te_i)$.

1. Montrer que $g(t) = at^2 + bt + \sqrt{\varepsilon + (c+t)^2} + D$, avec $a = \frac{\mu}{2} \|Ae_i\|^2$, $b = \mu\langle Ax - y|Ae_i\rangle$, $c = x_i$ et D une constante à déterminer.
2. Déduire de la partie I que le problème de minimisation

$$\min_{t\in\mathbb{R}} f(x + te_i). \tag{5}$$

admet un unique minimiseur, que l'on notera $T(x, i) \in \mathbb{R}$.

Algorithme de descente par coordonnée : Pour $k \in \mathbb{N}$, on notera $i_k = 1 + \mathrm{mod}(k, d) \in \{1, \ldots, d\}$, où $\mathrm{mod}(k, d)$ est le reste de la division euclidienne de k par d. On considère les itérées

$$x^{(0)} \in \mathbb{R}^d, \quad x^{(k+1)} = x^{(k)} + T(x^{(k)}, i_k)e_{i_k}.$$

Mettre en œuvre l'algorithme de descente par coordonnée sur ordinateur.

Question 7. *[Propriété des itérées et convergence]* Démontrer les propriétés (i) et (ii) suivantes :

$$\text{(i) } f(x^{(k+1)}) \leq f(x^{(k)}), \qquad \text{(ii) } \frac{\partial f}{\partial e_{i_k}}(x^{(k+1)}) = 0, \qquad \text{(iii) } f(x^{(k+1)}) \leq f(x^{(k)}) - \frac{1}{2M} \left|\frac{\partial f}{\partial e_{i_k}}(x^{(k)})\right|^2.$$

En admettant (iii), démontrer que $\lim_{k\to+\infty} \left\|\nabla f(x^{(k)})\right\|^2 = 0$.