Algorithmes d'optimisation – L3 MINT/DL 2019/2020 – Université Paris-Sud

Partiel du 03/03/2020 (3h)

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$^{\text{o}}$ 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 $N(t^*) = t^*$.

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

1. Montrer que $g' \leq 0$ sur $I = [p, t^*]$.
2. En déduire que $t^{(0)} \leq t^{(1)} = N(t^{(0)})$.
3. Calculer N', et montrer que N est croissant sur I.
4. Montrer par récurrence que $\forall k \geq 1, t^{(k)} \leq t^{(k+1)} \leq t^*$.
5. Déduire de ce qui précède que la suite $(t^{(k)})$ admet une limite $\bar{t}$ lorsque $k \to +\infty$.
6. Montrer que $N(\bar{t}) = \bar{t}$, et en déduire que $\bar{t} = t^*$. Conclure.

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'''.
2. Montrer que les hypothèses (i) et (ii) sont vérifiées, avec $p = -c$.
   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.
2. Montrer que $\nabla f(x) = \mu(A^T Ax - A^T y) + G(x)$ et $\mathrm{D}^2 f(x) = \mu A^T A + H(x)$ 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 $s''(t)$.
3. Démontrer que la matrice $H(x)$ est définie positive, puis que $D^2 f(x)$ l'est aussi.
4. En déduire que le problème (4) admet une unique solution qu'on notera désormais x^*.

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.
2. Démontrer que $\forall x \in S, \forall v \in \mathbb{R}^d$, $\frac{\varepsilon}{(\varepsilon+R^2)^{\frac{3}{2}}} \|v\|^2 \leq \langle \mathrm{D}^2 f(x) v | v \rangle \leq \left(\mu \|A\|^2 + \frac{1}{\sqrt{\varepsilon}} \right) \|v\|^2$.
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 $\lim_{k \to +\infty} \left\| \nabla f(x^{(k)}) \right\|^2 = 0$.