5MAS02 : Méthodes du premier ordre pour l'optimisation non convexe et non lisse

Corrigé de l'examen final du 14 janvier 2021 – Durée : 3 heures

Exercice 1

Soit $(n, m) \in (\mathbb{N}^*)^2$, $A \in \mathcal{M}_{n,m}(\mathbb{R})$ et $b \in \mathbb{R}^m$. Considérons le problème d'optimisation suivant :

$$\min_{x \in \mathbb{R}^n} \frac{1}{2} \|A\,x - b\|_2^2 + \|x\|_1$$

1. Appliquer l'algorithme *Forward-Backward Splitting* pour résoudre ce problème. Justifier **soigneusement** de la convergence de cette méthode dans le cas considéré.
2. Appliquer l'algorithme de Chambolle–Pock pour résoudre ce problème. Quelle est la formulation primale-duale du problème associée ? Justifier **soigneusement** de la convergence de cette méthode dans le cas considéré.
3. Discuter des avantages et inconvénients éventuels des deux méthodes appliquées.

Exercice 2

Soit $n \in \mathbb{N}^*$. On note $\langle \cdot, \cdot \rangle$ le produit scalaire associé à la base canonique de $\mathbb{R}^n$ et $\| \cdot \|$ la norme euclidienne associée.

Première partie – Soit $n \in \mathbb{N}^*$ et $\Sigma \in \mathcal{M}_{n,n}(\mathbb{R})$ une matrice **symétrique définie positive**, *i.e.* $\Sigma^\top = \Sigma$ et $\langle \Sigma\,x, x \rangle > 0$ pour tout $x \in \mathbb{R}^n \setminus \{0\}$. On notera que Σ est inversible et que Σ^{-1} est symétrique définie positive, et qu'elle admet des valeurs propres **strictement positives**. On considère la fonction suivante :

$$\| \cdot \|_\Sigma : \left\{ \begin{array}{ccc} \mathbb{R}^n & \to & \mathbb{R} \\ x & \mapsto & \sqrt{\langle x, \Sigma^{-1} x \rangle} \end{array} \right.$$

1. Montrer que $(x, y) \mapsto \|x - y\|_\Sigma^2$ définit une norme.
2. En déduire que $\| \cdot \|_\Sigma^2$ est convexe.
3. Montrer que $\| \cdot \|_\Sigma^2$ est différentiable et déterminer son gradient.

   Corrigé – Soit $x \in \mathbb{R}^n$ et $h \in \mathbb{R}^n$.

   $$\begin{aligned} \|x + h\|_\Sigma^2 &= \langle x + h, \Sigma^{-1}(x + h) \rangle \\ &= \langle x + h, \Sigma^{-1} x + \Sigma^{-1} h \rangle \\ \|x + h\|_\Sigma^2 &= \langle x, \Sigma^{-1} x \rangle + \langle x, \Sigma^{-1} h \rangle + \langle h, \Sigma^{-1} x \rangle + \langle h, \Sigma^{-1} h \rangle \end{aligned}$$

   Puisque Σ^{-1} est symétrique et que les normes sont équivalentes en dimension finie, on a

   $$\begin{aligned} \|x + h\|_\Sigma^2 &= \|x\|_\Sigma^2 + 2\,\langle h, \Sigma^{-1} x \rangle + \|h\|_\Sigma^2 \\ &= \|x\|_\Sigma^2 + 2\,\langle h, \Sigma^{-1} x \rangle + o(\|h\|) \end{aligned}$$

   $$\boxed{\forall\, x \in \mathbb{R}^n, \qquad \nabla(\| \cdot \|_\Sigma^2)(x) = 2\,\Sigma^{-1} x}$$

4. Montrer que $\| \cdot \|_\Sigma^2$ est fortement convexe.

Pour tout $f : \mathbb{R}^n \to \mathbb{R} \cup \{+\infty\}$ une fonction convexe, propre et semi-continue inférieurement et $x^0 \in \mathbb{R}^n$, on considère le problème d'optimisation suivant :

$$\min_{x \in \mathbb{R}^n} \left\{ f(x) + \frac{1}{2} \|x - x^0\|_\Sigma^2 \right\} \tag{$\mathcal{P}$}$$

5. Combien de solutions le problème $(\mathcal{P})$ admet-il ?

   Corrigé – La fonction f est convexe et $\|\cdot\|_\Sigma^2$ est fortement convexe ; comme $1/2$ est strictement positif, on en déduit que le problème $(\mathcal{P})$ est fortement convexe. Par conséquent,

$$\boxed{(\mathcal{P}) \text{ admet exactement une solution.}}$$

6. Justifier la notation

$$\forall\, x^0 \in \mathbb{R}^n, \qquad (\mathrm{Id} + \Sigma\, \partial f)^{-1}(x^0) = \arg\min_{x \in \mathbb{R}^n} \left\{ f(x) + \frac{1}{2} \|x - x^0\|_\Sigma^2 \right\}$$

   Préciser l'ensemble de définition et l'image de l'application $(\mathrm{Id} + \Sigma\, \partial f)^{-1}$.

   Corrigé – Puisque le problème $(\mathcal{P})$ est fortement convexe, la règle de Fermat assure que

$$x^+ \in \arg\min_{x \in \mathbb{R}^n} \left\{ f(x) + \frac{1}{2} \|x - x^0\|_\Sigma^2 \right\} \quad \iff \quad 0 \in \partial \left(f + \frac{1}{2} \|\cdot - x^0\|_\Sigma^2 \right)(x^+)$$

   Puisque $\|\cdot - x^0\|_\Sigma^2/2$ est différentiable et que $f + \|\cdot - x^0\|_\Sigma^2/2$ est convexe, on a la décomposition suivante :

$$\partial \left(f + \frac{1}{2} \|\cdot - x^0\|_\Sigma^2 \right)(x^+) = \partial f(x^+) + \nabla \left(\frac{1}{2} \|\cdot - x^0\|_\Sigma^2 \right)(x^+) = \partial f(x^+) + \Sigma^{-1}(x^+ - x^0)$$

   Il s'ensuit que, puisque Σ est inversible,

$$\begin{aligned}
x^+ \in \arg\min_{x \in \mathbb{R}^n} \left\{ f(x) + \frac{1}{2} \|x - x^0\|_\Sigma^2 \right\} \quad &\iff \quad 0 \in \partial f(x^+) + \Sigma^{-1}(x^+ - x^0) \\
&\iff \quad 0 \in \Sigma \partial f(x^+) + x^+ - x^0 \\
&\iff \quad x^0 \in x^+ + \Sigma \partial f(x^+) \\
x^+ \in \arg\min_{x \in \mathbb{R}^n} \left\{ f(x) + \frac{1}{2} \|x - x^0\|_\Sigma^2 \right\} \quad &\iff \quad x^0 \in (\mathrm{Id} + \Sigma \partial f)(x^+)
\end{aligned}$$

   Autrement dit, x^0 est dans l'image directe de $\{x^+\}$ par $\mathrm{Id} + \Sigma \partial f$; ainsi, x^+ appartient à l'antécédent de x^0 par $\mathrm{Id} + \Sigma \partial f$:

$$x^+ \in (\mathrm{Id} + \Sigma \partial f)^{-1}(x^0)$$

   On vient donc de démontrer que

$$\boxed{\arg\min_{x \in \mathbb{R}^n} \left\{ f(x) + \frac{1}{2} \|x - x^0\|_\Sigma^2 \right\} = (\mathrm{Id} + \Sigma \partial f)^{-1}(x^0)}$$

   Par existence et unicité de la solution du problème $(\mathcal{P})$,

$$\boxed{(\mathrm{Id} + \Sigma \partial f)^{-1} \text{ est défini sur } \mathbb{R}^n \text{ et est à valeurs dans } \mathbb{R}^n \text{ (opérateur mono-valué).}}$$

7. Dans quels cas particuliers reconnaît-on un opérateur connu ?

   Corrigé –

$$\boxed{\text{Lorsque } \Sigma = \sigma\, \mathrm{Id} \text{ avec } \sigma > 0, \text{ on reconnaît l'opérateur proximal de } f.}$$

8. Déterminer les points fixes de $(\mathrm{Id} + \Sigma\,\partial f)^{-1}$.

   Corrigé – D'après les calculs précédents,

$$\begin{aligned} x^+ = (\mathrm{Id} + \Sigma\partial f)^{-1}(x^+) &\iff x^+ \in (\mathrm{Id} + \Sigma\partial f)(x^+) \\ &\iff x^+ \in x^+ + \Sigma\partial f(x^+) \\ &\iff 0 \in \Sigma\partial f(x^+) \\ x^+ = (\mathrm{Id} + \Sigma\partial f)^{-1}(x^+) &\iff 0 \in \partial f(x^+) \end{aligned}$$

   Les points fixes de $(\mathrm{Id}+\Sigma\,\partial f)^{-1}$ sont donc les points critiques de f ; puisque f est convexe,

$$\boxed{\text{Les points fixes de } (\mathrm{Id} + \Sigma\,\partial f)^{-1} \text{ sont les minimiseurs de } f.}$$

Deuxième partie – *On conserve les notations de la première partie.* On s'intéresse à la résolution du problème d'optimisation suivant :

$$\min_{x\in\mathbb{R}^n} f(x) \qquad (\mathcal{P}')$$

où $f : \mathbb{R}^n \to \mathbb{R} \cup \{+\infty\}$ est un fonction convexe, propre et continue sur son domaine, que l'on suppose fermé. On suppose que f est coercive. On s'intéresse à l'algorithme suivant : pour $x_0 \in \mathbb{R}^n$, on définit pour tout $k \in \mathbb{N}$

$$x_{k+1} = (\mathrm{Id} + \Sigma\,\partial f)^{-1}(x_k)$$

1. À quoi correspondent les points fixes de cet algorithme ?

   Corrigé – Les points fixes de l'algorithme sont les points $x^* \in \mathbb{R}^n$ tels que

$$x^* = (\mathrm{Id} + \Sigma\,\partial f)^{-1}(x^*)$$

   Ainsi, d'après la question 8 de la première partie,

$$\boxed{\text{Les points fixes de l'algorithme sont les minimiseurs de } f.}$$

2. Montrer qu'il existe $C > 0$ tel que, pour tout $k \in \mathbb{N}$,

$$f(x_{k+1}) \le f(x_{k+1}) + \frac{C}{2}\,\|x_{k+1} - x_k\|^2 \le f(x_k)$$

   En déduire que la suite $(f(x_k))_{k\in\mathbb{N}}$ est une suite convergente.

   Corrigé – Par optimalité,

$$f(x_{k+1}) + \frac{1}{2}\,\|x_{k+1} - x_k\|_\Sigma^2 \le f(x_k) + \frac{1}{2}\,\|x_k - x_k\|_\Sigma^2 = f(x_k)$$

   Or, sur $\mathbb{R}^n$, toutes les normes sont équivalentes ; donc il existe $C > 0$ tel que

$$\forall\,(x,y) \in \mathbb{R}^n, \qquad C\,\|x-y\|^2 \le \|x-y\|_\Sigma^2$$

   Par ailleurs, puisque $\qquad 0 \le \dfrac{C}{2}\,\|x_{k+1} - x_k\|^2$

$$\boxed{\text{Pour tout } k \in \mathbb{N},\ f(x_{k+1}) \le f(x_{k+1}) + \frac{C}{2}\,\|x_{k+1} - x_k\|^2 \le f(x_k)}$$

   Ainsi, on a montré que la suite $(f(x_k))_{k\in\mathbb{N}}$ est décroissante ; puisque f est s.c.i. et coercive, elle admet un minimiseur, donc elle est *minorée*, ce qui implique donc que la suite $(f(x_k))_{k\in\mathbb{N}}$ est également minorée.

$$\boxed{\text{La suite } (f(x_k))_{k\in\mathbb{N}} \text{ converge.}}$$

3. Montrer que la suite des $x_{k+1} - x_k$ converge vers 0.

   Corrigé – Soit $K \in \mathbb{N}^*$. On somme les inégalités établies dans la question précédentes pour k entre 0 et K :

$$\sum_{k=0}^{K} f(x_{k+1}) + \frac{C}{2} \|x_{k+1} - x_k\|^2 \leq \sum_{k=0}^{K} f(x_k)$$

   Après télescopage des $f(x_k)$, on obtient

$$f(x_{K+1}) + \frac{C}{2} \sum_{k=0}^{K} \|x_{k+1} - x_k\|^2 \leq f(x_0)$$

   La fonction f est s.c.i. et coercive, donc elle admet un minimiseur ; par conséquent, on a la minoration suivante

$$\min_{x\in\mathbb{R}^n} f(x) + \frac{C}{2} \sum_{k=0}^{K} \|x_{k+1} - x_k\|^2 \leq f(x_0)$$

   Par conséquent,

$$\sum_{k=0}^{K} \|x_{k+1} - x_k\|^2 \leq \frac{2}{C} \left(f(x_0) - \min_{x\in\mathbb{R}^n} f(x) \right)$$

   Il s'ensuit que la suite $\left(\sum_{k=0}^{K} \|x_{k+1} - x_k\|^2 \right)_{K\in\mathbb{N}^*}$ est croissante et majorée, donc convergente. Autrement dit, la série de terme général $\|x_{k+1} - x_k\|^2$ est absolument convergente, donc

$$\boxed{\text{La suite } (\|x_{k+1} - x_k\|)_{k\in\mathbb{N}} \text{ converge vers } 0.}$$

4. Pour tout $k \in \mathbb{N}^*$, déterminer un sous-gradient p_k de f en x_k. Montrer que la suite $(p_k)_{k\in\mathbb{N}}$ converge vers 0.

   Corrigé – On écrit la règle de Fermat pour la définition de x_{k+1} :

$$0 \in \partial f(x_{k+1}) + \Sigma^{-1}(x_{k+1} - x_k)$$

   Il s'ensuit que

$$\Sigma^{-1}(x_k - x_{k+1}) \in \partial f(x_{k+1})$$

$$\boxed{\forall\, k \in \mathbb{N}, \qquad p_k = \Sigma^{-1}(x_k - x_{k+1}) \in \partial f(x_{k+1})}$$

   D'après la question précédente, et par continuité de l'application linéaire Σ^{-1},

$$\boxed{\lim_{k\to+\infty} p_k = 0}$$

5. Justifier que la suite des x_k admet une valeur d'adhérence.

   Corrigé – D'après la question 2, on a

$$\forall\, k \in \mathbb{N}, \qquad f(x_k) \leq f(x_0)$$

   Autrement dit, les x_k appartiennent à l'ensemble de sous-niveau $f(x_0)$. Puisque f est coercive et continue sur son domaine fermé, cet ensemble est borné.

$$\boxed{\text{La suite } (x_k)_{k\in\mathbb{N}} \text{ admet une valeur d'adhérence.}}$$

6. Montrer que les valeurs d'adhérence de $(x_k)_{k\in\mathbb{N}}$ sont des minimiseurs de f.

Corrigé – Commençons par vérifier que les x_k appartiennent au domaine de f ; en effet, x_k est minimiseur de $x \mapsto f(x) + \|x - x_{k-1}\|_\Sigma^2/2$, donc

$$f(x_k) + \frac{1}{2}\,\|x_k - x_{k-1}\|_\Sigma^2$$

est fini ; il s'ensuit que $f(x_k)$ est fini. Soit x^* une valeur d'adhérence de la suite $(x_k)_{k\in\mathbb{N}}$. Ainsi, il existe une sous-suite $(x_{k_j})_{j\in\mathbb{N}}$ de limite x^*. Par hypothèse, f est continue sur son domaine, donc

$$\lim_{j\to+\infty} f(x_{k_j}) = f(x^*)$$

Par ailleurs, d'après ce qui précède,

$$\lim_{j\to+\infty} p_{k_j} = 0 \qquad \text{avec} \qquad p_{k_j} \in \partial f(x_{k_j})$$

Par fermeture du sous-différentiel,

$$\boxed{0 \in \partial f(x^*)}$$

Autrement dit, x^* est un point critique de f. Puisque f est convexe, il s'ensuit que

$$\boxed{x^* \text{ est un minimiseur de } f.}$$

7. En déduire que, si f est strictement convexe, alors l'algorithme considéré converge vers une solution du problème $(\mathcal{P}')$.

   Corrigé – Si f est strictement convexe, f admet au plus un minimiseur ; puisque f est coercive, elle admet exactement un minimiseur. Or, la suite $(x_k)_{k\in\mathbb{N}}$ possède au moins une valeur d'adhérence, lesquelles sont des minimiseurs de f. Donc elle admet exactement une valeur d'adhérence et est donc convergente, de limite x^*. D'après la question précédente, x^* est minimiseur de f.