5Mas02 : Méthodes du premier ordre pour l'optimisation non convexe et non lisse

Examen final du 8 janvier 2020 – Durée : 3 heures

Consignes – *Aucun document,* ***à l'exception de deux pages format A4 de notes manuscrites****, n'est autorisé. L'utilisation de calculatrices, téléphones portables (ainsi que tout autre appareil électronique) est interdite. Cet énoncé comporte* ***trois*** *exercices largement indépendants qui peuvent être traités dans l'ordre de votre choix. Les réponses doivent être soigneusement justifiées, sauf mention contraire. La qualité de la rédaction et la rigueur des raisonnements seront prises en compte dans la notation.* ***Le candidat est autorisé à rédiger en anglais.***

Exercice 1

Première partie – Soit $n \in \mathbb{N}^*$. Considérons la fonction suivante :

$$\forall\, x = \big(x_{(i)}\big)_{1 \leq i \leq n} \in \mathbb{R}^n, \qquad f(x) = \|x\| = \sqrt{\sum_{i=1}^{n} x_{(i)}^2}$$

1. Montrer que f est convexe.
2. Calculer en tout point $x \in \mathbb{R}^n$ le sous-différentiel $\partial f(x)$ de f.
3. Déterminer sa conjuguée convexe f^*.
4. Justifier que l'opérateur proximal prox_f est monovalué (c'est-à-dire à valeurs dans $\mathbb{R}^n$). Déterminer pour tout point $x^0 \in \mathbb{R}^n$ le point proximal de f en x^0.
5. Calculer $\mathrm{prox}_{\tau f}(x^0)$ pour tout point $x^0 \in \mathbb{R}^n$ et tout réel $\tau > 0$.

Deuxième partie – Soit $\mathcal{X}$ un espace vectoriel de dimension finie muni du produit scalaire $\langle \cdot, \cdot \rangle$. Démontrer les deux résultats suivants :

1. *Lemme de descente.* Soit $L > 0$ et $f : \mathcal{X} \to \mathbb{R}$ une fonction L-régulière. Alors on a

$$\forall\, (x, z) \in \mathcal{X}^2, \qquad f(z) \leq f(x) + \langle \nabla f(x), z - x \rangle + \frac{L}{2} \|x - z\|^2$$

2. *Inégalité proximale.* Soit $g : \mathcal{X} \to \mathbb{R} \cup \{+\infty\}$ une fonction convexe, s.c.i. et propre. Soit $x^0 \in \mathcal{X}$ et $\tau > 0$. On pose

$$x^+ \in \mathrm{prox}_{\tau g}(x^0)$$

Alors on a

$$\forall\, x \in \mathcal{X}, \qquad g(x) \geq g(x^+) - \frac{1}{\tau} \langle x - x^+, x^+ - x^0 \rangle$$

Exercice 2

Soit $\mathcal{X}$ et $\mathcal{Y}$ deux espaces vectoriels de dimension finie munis de produits scalaires notés $\langle \cdot, \cdot \rangle$. Considérons le problème d'optimisation :

$$\min_{(x,y)\in\mathcal{X}\times\mathcal{Y}} J(x,y) = f(x,y) + g(x) + h(y) \tag{$\mathcal{P}$}$$

avec $f : \mathcal{X}\times\mathcal{Y} \to \mathbb{R}$ une fonction L-régulière ($L > 0$) et $g : \mathcal{X} \to \mathbb{R}\cup\{+\infty\}$ et $h : \mathcal{Y} \to \mathbb{R}\cup\{+\infty\}$ deux fonctions continues sur leur domaine non vide et fermé. On suppose que J est une fonction minorée et coercive, de domaine non vide. On se propose d'étudier le comportement de l'algorithme $\mathcal{A}$ suivant :

$$(x_0, y_0) \in \mathcal{X}\times\mathcal{Y} \qquad \text{et} \qquad \forall\, k \in \mathbb{N}, \qquad \begin{cases} x_{k+1} \in \operatorname{prox}_{\tau g}\left(x_k - \tau\,\dfrac{\partial f}{\partial x}(x_k, y_k)\right) \\ y_{k+1} \in \operatorname{prox}_{\tau(f(x_{k+1},\cdot)+h)}(y_k) \end{cases}$$

pour $\tau > 0$, où l'on rappelle que

$$f(x,\cdot) : y \mapsto f(x,y)$$

1. Expliquer, sans calcul, que si $\tau \in \left]\,0\,;L^{-1}\right[$, alors $\mathcal{A}$ est un algorithme de descente, c'est-à-dire que, pour tout $k \in \mathbb{N}$, on a

   $$J(x_{k+1}, y_{k+1}) \leq J(x_{k+1}, y_k) \leq J(x_k, y_k)$$

   En déduire la convergence en valeur de l'algorithme $\mathcal{A}$.
2. Sans justification, citer deux avantages qu'apporterait la convexité des fonctions f, g et h à l'algorithme $\mathcal{A}$.
3. **On suppose désormais que f, g et h sont convexes.** Quels sont les points fixes de l'algorithme $\mathcal{A}$?
4. On suppose que $\tau < 2\,L^{-1}$. Démontrer que les suites réelles $(\|x_k - x_{k+1}\|)_{k\in\mathbb{N}}$ et $(\|y_k - y_{k+1}\|)_{k\in\mathbb{N}}$ convergent vers zéro.
5. Déterminer, en fonction de $(x_k, x_{k+1}, y_k, y_{k+1})$, un sous-gradient p_{k+1} de J en (x_{k+1}, y_{k+1}). En déduire que la suite des p_k converge vers zéro.
6. Démontrer que $J(x_1, y_1)$ est fini.
7. Justifier que la suite $((x_k, y_k))_{k\in\mathbb{N}}$ admet une valeur d'adhérence dans le domaine de J et démontrer que toute valeur d'adhérence de cette suite est un minimiseur de J.
8. Quelle condition ajouter pour assurer la convergence forte des itérées (x_k, y_k) vers un minimiseur de J ?

Exercice 3

Les trois parties sont très largement indépendantes.

Première partie – Soit $n \in \mathbb{N}^*$ et $(M,m) \in \mathbb{R}^2$ tels que $M \geq m$. Considérons le problème d'optimisation sous contraintes suivant :

$$\min_{\substack{x=(x_{(i)})_{0\leq i\leq n-1}\\ x_{(0)}=1}} \frac{M-m}{2}\,\|K\,x\|^2 + \frac{m}{2}\,\|x\|^2 \qquad (\mathcal{P})$$

avec $K : \mathbb{R}^n \to \mathbb{R}^{n-1}$ une application linéaire définie par

$$\forall\, j \in [\![\, 1\,;\, n-1\,]\!]\,, \qquad (K\,x)_{(j)} = \frac{x_{(j)} - x_{(j-1)}}{2}$$

1. Justifier que ce problème est fortement convexe, et déterminer une valeur pour le module de forte convexité. En déduire le nombre de solutions du problème $(\mathcal{P})$.

2. Posons $$\forall\, x \in \mathbb{R}^n, \qquad J(x) = \frac{M-m}{2}\,\|K\,x\|^2 + \frac{m}{2}\,\|x\|^2$$

   Justifier que J est différentiable et calculer son gradient en tout point de $\mathbb{R}^n$. En déduire que J est régulière.

2* *Question subsidiaire :* En utilisant la majoration $(a-b)^2 \leq 2\,a^2 + 2\,b^2$ valable pour tout $(a,b) \in \mathbb{R}^2$, donner une estimation numérique du paramètre de régularité L lorsque $M = 10$, $m = 1$ et $n = 4$. *On admet que*

$$\forall\, i \in [\![\, 0\,;\, n-1\,]\!]\,, \quad (K^*K\,x)_{(i)} = \begin{cases} \dfrac{x_{(0)} - x_{(1)}}{4} & \text{si } i = 0 \\ \dfrac{2\,x_{(j)} - x_{(j-1)} - x_{(j+1)}}{4} & \text{si } i \in [\![\, 1\,;\, n-2\,]\!] \\ \dfrac{x_{(n-1)} - x_{(n-2)}}{4} & \text{si } i = n-1 \end{cases}$$

3. Posons $$\mathcal{C} = \left\{ x = (x_{(i)})_{0\leq i\leq n-1} \in \mathbb{R}^n \mid x_{(0)} = 1 \right\}$$

   Définir et justifier que la projection orthogonale sur cet ensemble existe et est définie de manière unique.

4. Calculer la projection orthogonale $\mathrm{proj}_{\mathcal{C}}$ sur $\mathcal{C}$.

5. Écrire l'algorithme du gradient projeté pour ce problème. Comment le pas de temps doit-il être choisi pour garantir la convergence de cet algorithme ? *Ce choix ne devra pas être justifié.*

Deuxième partie – *On conserve les notations de la première partie.* Considérons le problème d'optimisation sous contraintes suivant :

$$\min_{\substack{(x,z)\in\mathbb{R}^n\times\mathbb{R}^{n-1}\\ Kx=z}} \frac{M-m}{2}\,\|z\|^2 + \frac{m}{2}\,\|x\|^2 + \chi_{\mathcal{C}}(x) \qquad (\mathcal{P}')$$

1. Combien de solutions admet le problème $(\mathcal{P}')$? Expliciter le lien entre les solutions du problème $(\mathcal{P})$ et celles du problème $(\mathcal{P}')$.
2. Écrire le lagrangien augmenté $\mathcal{L}_\tau$ associé au problème $(\mathcal{P}')$, en introduisant un multiplicateur de LAGRANGE pour la contrainte d'égalité $K\,x = z$.
3. Écrire les conditions nécessaires et suffisantes pour qu'un point $(\bar{x}, \bar{z}, \bar{\lambda})$ soit un point-selle du lagrangien augmenté $\mathcal{L}_\tau$.
4. Écrire les itérations de l'ADMM (*Alternating Direction Method of Multipliers*) pour résoudre le problème $(\mathcal{P})$, en donnant une formule explicite pour chaque itérée. *Indication : pour la mise-à-jour de la variable x, on pourra introduire le sous-vecteur $\tilde{x} = (x_{(i)})_{1\leq i\leq n-1}$.*

Posons
$$\forall\, x \in \mathbb{R}^n, \qquad f(x) = \frac{M-m}{2}\,\|K\,x\|^2$$

5. À l'aide de la conjuguée convexe f^* de f, proposer une autre formulation primale-duale pour le problème $(\mathcal{P})$.
6. Écrire les itérations de l'algorithme de CHAMBOLLE–POCK pour la recherche d'un point-selle de la fonction de couplage $\mathcal{L}$.

Troisième partie – *On conserve les notations de la première et de la deuxième partie.* Considérons le problème d'optimisation sous contraintes suivant :

$$\min_{(x,y,z)\in\mathbb{R}^3} \frac{9}{8}\left((x-1)^2 + (x-y)^2 + (y-z)^2\right) + \frac{1}{2}\left(x^2+y^2+z^2\right) \qquad (\mathcal{P}'')$$

1. Expliciter le lien entre les solutions des problèmes $(\mathcal{P}'')$ et $(\mathcal{P})$.
2. Résoudre le problème $(\mathcal{P}'')$.
3. Écrire les itérations de l'algorithme BCD (*Block Coordinate Descent*) pour résoudre le problème $(\mathcal{P}'')$. On donnera une formulation explicite des itérées.

FIN DE L'ÉNONCÉ