UniversitéParis-Saclay
M2 MathSV et Math de l'Aléatoire
Octobre 2015

Exercices d'optimisation convexe

Exercice 1. Considérer les ensembles

$$A = \{(x, y) \in \mathbb{R}^2 \,:\, |x + y| \leq 1\} \quad \text{et} \quad B = \overline{B(0, 1)} = \{(x, y) \in \mathbb{R}^2 \,:\, x^2 + y^2 \leq 1\}.$$

1. Dessiner A, B et $A \cap B$. A et B sont-ils convexes ?
2. Écrire une formule pour la projection sur $A \cap B$, du type $P_{A\cap B}(x, y) = \dots$, en distinguant éventuellement des cas (il suffit de la justifier avec un dessin).
3. Lesquelles des relations suivantes sont-elles vraies ?

$$P_{A\cap B} = P_A \circ P_B, \quad P_A \circ P_B = P_B \circ P_A, \quad P_{A\cap B} = P_{A\cap B} \circ P_B.$$

Exercice 2. Considérer le problème

$$\min\{J(f) := \int_0^1 \left[\frac{1}{2}f'(t)^2 + tf(t) + \frac{1}{2}tf(t)^2\right] dt\,;\ f \in \mathcal{A}\}, \quad \text{où} \quad \mathcal{A} := \{f \in C^1([0, 1]) \,:\, f(0) = 0\},$$

qui n'est pas évident à résoudre explicitement. Suggérer une discrétisation et une méthode numérique pour le résoudre de manière approchée. Écrire précisément la fonction qu'on veut minimiser dans la discrétisation (avec les matrices et/ou les vecteurs qui apparaissent dans cette fonction) et expliquer la méthode que vous choisissez d'utiliser.

Solution On peut discrétiser ce problème en fixant une valeur n et en cherchant, au lieu de fonctions C^1, des fonctions affines par morceaux, chaque morceaux étant un intervalle de type $[\frac{i}{n}, \frac{i+1}{n}]$. On cherche donc des valeurs f_i, $i = 1, \dots, n$ (on prend $f_0 = 0$). Dans ce cas, le terme $\frac{1}{2}\int f'(t)^2dt$ s'écrit $\sum_{i=1}^n \tau(\frac{f_i - f_{i-1}}{\tau})^2$, où $\tau = 1/n$. Pour les autres termes, il serait possible de calculer exactment l'intégrale, mais par simplicité on considère une approximation, où on calcule l'intégrale comme si la fonction était constante et égale à f_i sur chaque intervalle $[\frac{i-1}{n}, \frac{i}{n}]$ (et la valeur de t sera aussi prise égale à i/n). Dans ce cas là les deux derniers terms s'écrivent $\sum_{i=1}^n \tau \cdot i\tau \cdot (f_i + f_i^2)$.

Il faut donc minimiser la fonction quadratique $\frac{1}{2}Af \cdot f + b \cdot f$, où $f \in \mathbb{R}^n$, $b = (\tau^2, 2\tau^2, \dots, n\tau^2)$ et la matrice A est obtenue comme $A = B + C$, où B est une matrice diagonale avec $B^{i,i} = i\tau^2$ et C est une matrice avec seulement des coefficients non nuls sur la diagonale principale et les deu diagonales à côté, et plus précisément $C^{i,i} = 1/\tau$ pour $i < n$, $C^{n,n} = 1/(2\tau)$, $C^{i,i+1} = -1/(2\tau)$ pour $i < n$, $C^{i,i-1} = -1/(2\tau)$ pour $i > 1$.

L'optimisation peut se faire par les formules qu'on a vues en cours, avec l'algorithme du gradient à pas fixe, ou du gradient à pas optimal.

Exercice 3. Considérons un échiquier 8×8 standard. Dans cet ćhiquier il ya 64 cases carrées et 144 segments qui constituent soit le bord entre deux cases adjacentes, soit des portions de bord de l'échiquier. On indiquera avec l'indice k les cases, et avec l'indice i les segments. Pour une case k donnée, $i = N(k), E(k), S(k), O(k)$ représentent les quatre segments (au nord, est, sud et ouest, respectivement) de bord de cette case. Dans chaque case on va écrire un nombre f_k (qu'on considère donné) et sur chaque bord un nombre w_i (qui est notre inconnue). Sur les segments qui sont sur le bord de l'équiquier (c'est-à-dire les i tels qu'il existe k avec $i = N(k)$ mais pas k' avec $i = S(k')$, et de même pour les autres points cardinaux), on écrit forcement $w_i = 0$ (c'est une contrainte). On met également la contrainte que pour chaque k on ait $w_{N(k)} - w_{S(k)} + w_{E(k)} - w_{O(k)} = f_k$.

Considérer le problème de minimisation

$$\min\left\{\sum_i \frac{1}{p}|w_i|^p \ : \ w \text{ satisfait les contraintes}\right\}.$$

Prouver que le problème est convexe, qu'il admet une solution, et trouver son dual.

Quel serait le correspondant "continue" de ce problème ?

Exercice 4. Soit $L : \mathbb{R}^n \times \mathbb{R}^m \to \mathbb{R}$ une fonction C^1, avec $L(x,y) \geq f(x) - C(y)$, où f est coercive ($\lim_{|x|\to\infty} f(x) = +\infty$) et C est localement bornée. Soit $G(y) := \min_x L(x,y)$. Supposons qu'un certain y_0 est tel que le minimum dans la définition de $G(y_0)$ est atteint en un point unique x_0. Prouver que $\nabla G(y_0) = \nabla_y L(x_0,y_0)$. Appliquer ce résultat pour prouver que h^* est C^1, et calculer son gradient, dès que h est strictement convexe et telle que $\lim_{|x|\to\infty} h(x)/|x| = +\infty$. Expliquer où toutes les hypothèses sont utilisées.

Solution On veut prouver $\lim_{y\to y_0} \frac{G(y)-G(y_0)-\nabla_y L(x_0,y_0)\cdot(y-y_0)}{|y-y_0|} = 0$. Une inégalité est simple : en effet $G(y) \leq L(x_0,y)$ par définition de G, et $G(y_0) = L(x_0,y_0)$, donc

$$\frac{G(y)-G(y_0)-\nabla_y L(x_0,y_0)\cdot(y-y_0)}{|y-y_0|} \leq \frac{L(x_0,y)-L(x_0,y_0)-\nabla_y L(x_0,y_0)\cdot(y-y_0)}{|y-y_0|} = 0,$$

où la limite vaut 0 par définition de différentiel de $y \mapsto L(x_0,y)$.

Pour l'inégalité opposée, prenons une suite $y_n \to y_0$ et appelons x_n les points (des points, s'ils sont pas uniques) qui réalisent le minimum dans la définition de $G(y_n)$: on a donc $G(y_n) = L(x_n,y_n)$ mais aussi $G(y_0) \leq L(x_n,y_0)$. On a donc

$$\frac{G(y_n)-G(y_0)-\nabla_y L(x_0,y_0)\cdot(y_n-y_0)}{|y_n-y_0|} \geq \frac{L(x_n,y_n)-L(x_n,y_0)-\nabla_y L(x_0,y_0)\cdot(y_n-y_0)}{|y_n-y_0|}.$$

Or, $L(x_n,y_n) - L(x_n,y_0) = \nabla_y L(x_n,z_n)\cdot(y_n-y_0)$, où $z_n \in [y_n,y_0$ est un point intermédiaire donné par le théorème des accroissements finis. On a donc

$$\frac{G(y_n)-G(y_0)-\nabla_y L(x_0,y_0)\cdot(y_n-y_0)}{|y_n-y_0|} \geq (\nabla_y L(x_n,z_n) - \nabla_y L(x_0,y_0))\cdot\frac{y_n-y_0}{|y_n-y_0|} \geq -|\nabla_y L(x_n,z_n) - \nabla_y L(x_0,y_0)|.$$

Comme L est C^1, il suffit de montrer $(x_n,z_n) \to (x_0,y_0)$ pour obtenir 0 dans la limite. Or, $z_n \to y_0$ parce que $y_n \to y_0$. Quant à x_n, on utilise $L(x_n,y_y) \geq f(x_n) - C(y_n)$, avec $L(x_n,y_n) \leq L(\bar{x},y_n)$ (pour un $\bar{x}$ quelconque, par optimalité de x_n), et le fait que y_n est une suite bornée. On en trouve que $f(x_n)$ est bornée, et comme f est coercive, x_n aussi. A une sous-suite près, on peut dire $x_n \to x$. Mais $y \mapsto G(y) := \min_x L(x,y)$ est semi-continue supérieurement (inf de fonctions continues), donc $G(y_0) \geq \limsup_n G(y_n) = \limsup_n L(x_n,y_n) = L(x,y_0)$. Alos x est optimal pour y_0, donc $x = x_0$. Comme cela est vrai pour toute sous-suite, on trouve $x_n \to x_0$, ce qui permet de conclure.

Pour répondre à la dernière question, on utilise $L(x,y) = h(x) - x\cdot y$ et $-h^*(y) = \min_x L(x,y)$. Si h est strictement convexe ce minimum est réalisé en un seul point $x(y)$, quel que soit y. Si on peut appliquer le raisonnement ci-dessus, on obtiendrait $-\nabla h^*(y) = -x(y)$. Ceci montrerait que h est différentiable, mais également C^1, parce que $x(y)$ est une fonction continue de y (pour le prouver, on fait comme tout à l'heure : on démontre que, si $x_n = x(y_n)$ et y_n est bornée, alors x_n est bornée aussi, on prend une sous-suite, et on prouve que la limite est $x(y)$. Dans la preuve ci-dessus, il faut juste trouver une minoration de L. On a $x\cdot y = \frac{1}{2}x\cdot 2y \leq \frac{1}{2}(h(x) + h^2(2y))$. On peut donc dire $L(x,y) \geq \frac{1}{2}h(x) - h^*(2y)$, donc $f(x) = h(x)/2$ et $g(y) = h^*(2y)$. Où utilise-t-on que h est superlinéaire ? pour dire que f est coercive, mais aussi que g est localement bornée. En effet, si h est supérlinéaire,h^* est une fonction convexe finie partout, et donc elle est continue, et donc localement bornée.