Introduction à l'optimisation continue

Contrôle

(9 janvier 2018)

Exercice I

On rappelle que pour une fonction convexe $f : X \to \mathbb{R} \cup \{+\infty\}$,

$$\text{prox}_{\tau f}(x) = \arg\min_y f(y) + \frac{1}{2\tau}\|y - x\|^2.$$

Calculer $\text{prox}_{\tau f}(x)$ pour $\tau > 0$, et

1. $X = \mathbb{R}^n$, $f(x) = \|x\|^2/2$;
2. $X = \mathbb{R}^n$, $f(x) = \|x\|_\infty = \max_{i=1,\dots,n} |x_i|$:
   (a) Montrer d'abord que
   $$\min_y \|y\|_\infty + \frac{1}{2\tau}\|y - x\|^2 = \min_{t\geq 0} t + \min_{y:|y_i|\leq t\,\forall i} \frac{1}{2\tau}\|y - x\|^2.$$
   (b) Déduire que
   $$\min_y \|y\|_\infty + \frac{1}{2\tau}\|y - x\|^2 = \min_{t\geq 0} t + \frac{1}{2\tau}\sum_{i=1}^n [(|x_i| - t)^+]^2$$
   où pour $s \in \mathbb{R}$, $s^+ := \max\{s, 0\}$. En déduire un algorithme pour trouver y.

Exercice II

On considère dans $E = \mathbb{R}^{n\times n}$ ensemble des matrices $n \times n$ la fonction

$$f(X) = \begin{cases} -\log\det X & \text{si } X \text{ est symétrique définie positive,} \\ +\infty & \text{sinon.} \end{cases}$$

1. Montrer que le domaine de f est un ensemble convexe.

2. Soit X une matrice symétrique définie positive ($X \in \text{dom}\, f$). Montrer que si H est une matrice symétrique et t est assez petit,

$$f(X + tH) = f(X) - t\left\langle X^{-1}, H\right\rangle + O(t)$$

où $\langle A, B\rangle = \text{Tr}\,(A^T B)$ ($= \text{Tr}\,(AB)$ bien entendu lorsque A, B sont symétriques. On rappelle que c'est un produit scalaire Euclidien, vérifiant notamment l'inégalité de Cauchy-Schwartz).

On a

$$-\log\det(X + tH) = -\log[\det(X)\det(I + tX^{-1}H)]$$
$$= -\log\det X - \log[1 + t\text{Tr}\,(X^{-1}H) + O(t^2)] = -\log\det X - t\text{Tr}\,(X^{-1}H) + o(t).$$

Donc $\nabla f(X) = -X^{-1}$.

3. Montrer que si X, Y sont des matrices symétriques définies positives,

$$n \leq \frac{1}{2}\left(\mathrm{Tr}\,(XY^{-1}) + \mathrm{Tr}\,(X^{-1}Y)\right).$$

(Indication : on partira de l'identité $I = X^{1/2}Y^{-1/2}Y^{1/2}X^{-1/2}$.)

On écrit

$$\begin{aligned} n = \mathrm{Tr}\,(I) = \mathrm{Tr}\,X^{1/2}Y^{-1/2}Y^{1/2}X^{-1/2} &= \left\langle X^{1/2}Y^{-1/2}, X^{-1/2}Y^{1/2}\right\rangle \\ &\leq \|X^{1/2}Y^{-1/2}\|\|X^{-1/2}Y^{1/2}\| \leq \frac{1}{2}\|X^{1/2}Y^{-1/2}\|^2 + \frac{1}{2}\|X^{-1/2}Y^{1/2}\|^2. \end{aligned}$$

On remarque ensuite que

$$\|X^{1/2}Y^{-1/2}\|^2 = \mathrm{Tr}\,(X^{1/2}Y^{-1/2}Y^{-1/2}X^{1/2}) = \mathrm{Tr}\,(XY^{-1})$$

et on conclut.

4. En déduire que f est une fonction convexe. On montrera (en se restreignant aux matrice symétriques définies positives) que $\nabla f(X)$ est un opérateur monotone.

On voit que

$$\begin{aligned} \langle \nabla f(X) - \nabla f(Y), X - Y\rangle &= -\mathrm{Tr}\,((X^{-1} - Y^{-1})(X - Y)) \\ &= -\mathrm{Tr}\,(2I - X^{-1}Y - Y^{-1}X) \geq 0 \end{aligned}$$

d'après la question précédente.

5. Calculer $Y = \mathrm{prox}_{\tau f}(X)$.

On écrit pour X une matrice et $\bar{X} = (X + X^T)/2$ sa projection sur les matrice symétriques que Y doit minimiser

$$\min_{Y \mathrm{s.d.p.}} f(Y) + \frac{1}{2\tau}(\|Y - \bar{X}\|^2 + \|\bar{X} - X\|^2).$$

Donc Y est solution de

$$-\tau Y^{-1} + Y - \bar{X} = 0 \quad \Leftrightarrow \quad -\tau + Y^2 - \bar{X}Y = 0.$$

On résout cette équation facilement en diagonalisant $\bar{X}$ (de valeurs propres $(\bar{\mu}_i)_{i=1}^n \in \mathbb{R}^n$) et en cherchant Y comme matrice diagonale dans la même base : les valeurs propres (μ_i) de Y doivent vérifier

$$-\tau + \mu_i^2 - \bar{\mu}_i\mu_i = 0$$

et être positives, elles sont donc données par

$$\mu_i = \frac{\bar{\mu}_i + \sqrt{\bar{\mu}_i^2 + 4\tau}}{2}.$$

Exercice III (Descente de gradient stochastique).

On note $x = (x_i)_{i=1}^n$ un point générique de $\mathbb{R}^n$, $n \geq 2$. On considère $f : \mathbb{R}^n \to \mathbb{R}$ une fonction convexe de classe C^2 avec $|\partial^2_{i,i} f| \leq L_i$ pour $i = 1, \dots, n$. (En d'autres termes, $\partial_i f$ est L_i-Lipschitz par rapport à x_i.)

On introduit les opérateurs de descente suivants, pour tout i :

$$T^i_\tau x = \begin{pmatrix} x_1 \\ \vdots \\ x_{i-1} \\ x_i - \tau \partial_i f(x) \\ x_{i+1} \\ \vdots \\ x_n \end{pmatrix}.$$

1. Montrer que si $y = T^i_\tau x$

$$f(y) \leq f(x) - \tfrac{\tau}{2}(2 - \tau L_i)|\partial_i f(x)|^2.$$

On écrit un développement de Taylor :

$$\begin{aligned} f(y) &= f(x) + \int_0^1 \nabla f(x + s(y-x)) \cdot (y-x) ds \\ &= f(x) + \nabla f(x) \cdot (y-x) + \int_0^1 [\nabla f(x + s(y-x)) - \nabla f(x)] \cdot (y-x) ds \end{aligned}$$

Si $y - x = -\tau \partial_i f(x) e_i$, on déduit

$$f(y) \leq f(x) - \tau |\partial_i f(x)|^2 + L_i \int_0^1 s\|y-x\|^2 ds \leq f(x) - \tfrac{\tau}{2}(1 - \tau L_i)|\partial_i f(x)|^2.$$

On considère maintenant l'algorithme "stochastique" suivant : lorsque x^k est calculé, on choisit l'indice i avec probabilité p_i ($p_i > 0, \sum_i p_i = 1$) et on pose $x^{k+1} = T^i_{\tau_i} x^k$ où $\tau_i = \theta / L_i$, $\theta \in]0, 2[$ fixé.

On rappelle que l'espérance de $g(x^{k+1})$ sachant x^k est donnée par

$$\mathbb{E}_k(g(x^{k+1})) = \sum_{i=1}^n p_i [\text{valeur de } g(x^{k+1}) \text{ si } i \text{ est choisi}],$$

pour n'importe quelle fonction g.

2. Montrer que

$$\mathbb{E}_k(f(x^{k+1})) \leq f(x^k) - \frac{\theta(2-\theta)}{2} \|\nabla f(x^k)\|^2_M$$

où M est la matrice diagonale telle que $m_{ii} = p_i / L_i$ et

$$\|\xi\|_M := \langle M\xi, \xi \rangle = \sum_i m_{ii} |\xi_i|^2.$$

3. Montrer que si x^* est un minimiseur de f, pour tout $x \in \mathbb{R}^n$,

$$f(x) - f(x^*) \leq - \langle \nabla f(x), x^* - x \rangle \leq \|\nabla f(x)\|_M \|x^* - x\|_{M^{-1}}.$$

En déduire que

$$\mathbb{E}_k(f(x^{k+1}) - f(x^*)) \leq f(x^k) - f(x^*) - \frac{\theta(2-\theta)}{2\|x^* - x^k\|^2_{M^{-1}}}(f(x^k) - f(x^*))^2.$$

4. On suppose qu'on sache trouver une constante $C > 0$ telle que $\|x^k - x^*\|_{M^{-1}} \leq C$ pour tout k (et tous les tirages aléatoires possibles). En déduire, en prenant l'espérance sur tous les x^k possibles, que

$$\Delta_{k+1} \leq \Delta_k - \frac{\theta(2-\theta)}{2C^2}\Delta_k^2$$

où l'on a noté $\Delta^k = \mathbb{E}(f(x^k) - f(x^*))$.

5. En déduire que

$$\Delta_k \leq \frac{2C^2}{\theta(2-\theta)} \frac{1}{k+1}.$$

Conclure.

6. Quel avantage peut-on trouver par rapport à un algorithme de descente de gradient déterministe (notamment dans le choix du pas de descente) ? Y a-t-il a priori un choix des $(p_i)_i$ meilleur qu'un autre ?

Exercice IV

On considère, pour $x \in \mathbb{R}^n$, $K : \mathbb{R}^n \to \mathbb{R}^m$ $(n, m \geq 1)$ opérateur linéaire continu et $f : \mathbb{R}^m \to [0, +\infty]$, $g : \mathbb{R}^n \to [0, +\infty]$ fonctions convexes, semi-continues inférieurement, le problème :

$$\min_x f(Kx) + g(x) = \min_x \sup_y \langle Kx, y \rangle - f^*(y) + g(x).$$

On choisit $M \in \mathbb{R}^{n \times n}$, $N \in \mathbb{R}^{m \times m}$ des “métriques”, c'est à dire des matrices symétriques définies positives définissant des normes Euclidiennes

$$\|x\|_M^2 := \langle Mx, x \rangle, \quad \|y\|_N^2 := \langle My, y \rangle.$$

0. Montrer que si A est un opérateur (maximal) monotone dans $\mathbb{R}^n$, alors $M^{-1}A$ est un opérateur (maximal) monotone dans $\mathbb{R}^n$ muni de la norme Euclidienne $\|\cdot\|_M$. Que déduit-on alors du théorème de Minty ?

On suppose maintenant qu'on sait résoudre les problèmes

$$\min_x \frac{1}{2}\|x - \bar{x}\|_M^2 + g(x),$$

$$\min_y \frac{1}{2}\|y - \bar{y}\|_N^2 + f^*(y).$$

On définit alors l'algorithme $(x^k, y^k) \mapsto (x^{k+1}, y^{k+1})$ par le calcul

$$\begin{pmatrix} M & -K^* \\ -K & N \end{pmatrix} \begin{pmatrix} x^{k+1} - x^k \\ y^{k+1} - y^k \end{pmatrix} + \begin{pmatrix} \partial g(x^{k+1}) \\ \partial f^*(y^{k+1}) \end{pmatrix} + \begin{pmatrix} 0 & K^* \\ -K & 0 \end{pmatrix} \begin{pmatrix} x^{k+1} \\ y^{k+1} \end{pmatrix} \ni 0_{\mathbb{R}^{n+m}}. \tag{1}$$

1. Montrer que la matrice

$$\begin{pmatrix} M & -K^* \\ -K & N \end{pmatrix}$$

est définie positive si et seulement si $\|N^{-1/2}KM^{-1/2}\| < 1$. Que se passe-t-il si $\|N^{-1/2}KM^{-1/2}\| = 1$?

2. Montrer que si $\|N^{-1/2}KM^{-1/2}\| < 1$, (1) définit un algorithme de point proximal associé à un opérateur maximal monotone, et montrer qu'il est effectivement implémentable (c'est-à-dire qu'on peut calculer, connaissant (x^k, y^k), le point (x^{k+1}, y^{k+1})).

Pour simplifier les notations, on écrit l'algorithme

$$P(\xi^{k+1} - \xi^k) + C\xi^{k+1} \ni 0$$

où P est symétrique, définie positive et C un opérateur maximal monotone (dans $\mathbb{R}^{n+m}$, donc). On suppose qu'il existe ξ^* avec $0 \in C\xi^*$. On rappelle que

$$2\langle P\xi, \eta\rangle = \|\xi\|_P^2 + \|\eta\|_P^2 - \|\xi - \eta\|_P^2.$$

3.a. Montrer que pour tout $k \geq 0$,

$$\begin{cases} \|\xi^{k+1} - \xi^*\|_P^2 + \|\xi^{k+1} - \xi^k\|_P^2 \leq \|\xi^k - \xi^*\|_P^2, \\ \|\xi^{k+2} - \xi^{k+1}\|_P^2 + \|\xi^{k+2} - 2\xi^{k+1} + \xi^k\|_P^2 \leq \|\xi^{k+1} - \xi^k\|_P^2. \end{cases} \tag{2}$$

En déduire que $\eta^k := P(\xi^{k-1} - \xi^k) \in C\xi^k$ vérifie

$$\|\eta^k\|_{P^{-1}}^2 \leq \frac{\|\xi^0 - \xi^*\|_P^2}{k},$$

puis que (ξ^k) a des sous-suite qui convergent vers des solutions de $C\xi \ni 0$.

3.b. Déduire ensuite de (2) que toute la suite (ξ^k) converge (vers une solution).

4. Que peut-on encore dire si $P \geq 0$? (P positive non définie).

On a encore $C\xi^k \to 0$. Si (ξ^k) a une sous-suite convergente on peut encore déduire que la limite est une solution.

On s'intéresse maintenant au cas où $K = I$, $n = m$, $M = N^{-1}$ (et donc $\|N^{-1/2}KM^{-1/2}\| = 1$). Pour simplifier on note $A = \partial g$, $B = \partial f$ et l'algorithme s'écrit donc

$$\begin{pmatrix} N^{-1} & -I \\ -I & N \end{pmatrix} \begin{pmatrix} x^{k+1} - x^k \\ y^{k+1} - y^k \end{pmatrix} + \begin{pmatrix} Ax^{k+1} \\ B^{-1}y^{k+1} \end{pmatrix} + \begin{pmatrix} 0 & I \\ -I & 0 \end{pmatrix} \begin{pmatrix} x^{k+1} \\ y^{k+1} \end{pmatrix} \ni 0_{\mathbb{R}^{2n}}. \tag{3}$$

5. Montrer que si

$$N(\hat{y} - \bar{y}) + B^{-1}\bar{y} \ni 0$$

alors

$$N\hat{y} = N\bar{y} - (I + NB)^{-1}(N\bar{y}).$$

6. En déduire que le point $v^k = x^k - Ny^k$ suit une itération de type “Douglas-Rachford”, que l’on précisera. Que peut-on conclure ?