Introduction à l'optimisation continue
Contrôle
(3 janvier 2017)

Exercice I

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

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

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

1. $X = \mathbb{R}$, $f(x) = -\ln x$ si $x > 0$, $+\infty$ si $x < 0$.
2. $f(x) = \psi(\|x\|)$ où $\psi : \mathbb{R} \to \mathbb{R} \cup \{+\infty\}$ est une fonction convexe, paire avec $\psi(0) = 0$. On montrera d'abord que f est une fonction convexe, on calculera ensuite $\mathrm{prox}_{\tau f}$ en fonction de $\mathrm{prox}_{\tau\psi}$.
3. $f(x) = \|x\|^3/3$.

Exercice II

On considère un espace de Hilbert X et une fonction strictement convexe, semi-continue inférieurement (sci) $\psi : X \to \mathbb{R} \cup \{+\infty\}$ telle que l'intérieur de $\mathrm{dom}\,\psi$, noté D, est non vide, $\overline{D} = \mathrm{dom}\,\psi$, $\psi \in C^1(D) \cap C^0(\overline{D})$, et $\partial\psi(x) = \emptyset$ pour tout $x \notin D$. En d'autres termes, $\partial\psi(x)$ est soit $\emptyset$ (si $x \notin D$), ou un singleton $\{\nabla\psi(x)\}$ (si $x \in D$). On suppose aussi que

$$\lim_{\|x\|\to\infty} \psi(x) = +\infty.$$

On définit la "distance de Bregman associée à ψ", notée $D_\psi(x,y)$, pour $y \in D$ et $x \in X$, de la manière suivante:

$$D_\psi(x,y) := \psi(x) - \psi(y) - \langle \nabla\psi(y), x-y \rangle .$$

1. Montrer que $D_\psi(x,y) \geq 0$, et que $D_\psi(x,y) = 0 \Rightarrow y = x$. Quel autre inégalité peut-on écrire si, en outre, ψ est fortement convexe ? Pourquoi est-ce que D_ψ *n'est pas* une distance au sens classique ?

2. Exprimer D_ψ lorsque $D = X$, $\psi(x) = \|x\|^2/2$. Lorsque $X = \mathbb{R}^n$, $D =]0,+\infty[^n$, $\psi(x) = \sum_{i=1}^n x_i \ln x_i$.

3. Soit $f : X \to \mathbb{R} \cup \{+\infty\}$ une fonction convexe, propre, sci. Soit $\bar{x} \in D$. On suppose qu'il existe $x \in D$ tel que $f(x) < +\infty$. Montrer qu'il existe un unique point $\hat{x} \in X$ tel que

$$f(\hat{x}) + D_\psi(\hat{x},\bar{x}) \leq f(x) + D_\psi(x,\bar{x}) \quad \forall x \in X. \tag{1}$$

4. Expliquer pourquoi $\partial(f+\psi) = \partial f + \partial\psi$. Écrire la condition d'optimalité du premier ordre pour $\hat{x}$. En déduire que $\hat{x} \in D$.

5. Montrer (à partir des conditions d'optimalité) que pour tout $x \in X$,

$$f(x) + D_\psi(x, \bar{x}) \geq f(\hat{x}) + D_\psi(\hat{x}, \bar{x}) + D_\psi(x, \hat{x}). \tag{2}$$

Un algorithme de descente de gradient "nonlinéaire". On considère le problème de minimisation:

$$\min_{x \in \overline{D}} f(x) + g(x), \tag{P}$$

pour f, g convexes, sci, propres, et où f est C^1 dans D et g est "simple" au sens suivant: on suppose qu'on sait résoudre

$$\min_x g(x) + \langle p, x \rangle + \frac{1}{\tau} D_\psi(x, y)$$

pour tout $\tau > 0$, $p \in X$ et $y \in D$. On suppose en outre qu'il existe $L > 0$ tel que pour tout $y \in D$, $x \in X$

$$D_f(x, y) \leq L D_\psi(x, y). \tag{3}$$

(Ici $D_f(x, y) = f(x) - f(y) - \langle \nabla f(y), x - y \rangle$.) On suppose que le problème de minimisation a une solution. On note $F(x) = f(x) + g(x)$.

6. Montrer que si ψ est 1-convexe (fortement convexe avec paramètre 1) et f a un gradient L-Lipschitzien, alors (3) est vraie.

Étant donné $\bar{x} \in D$, $\tau > 0$, on définit maintenant l'opérateur T_τ suivant : $\hat{x} = T_\tau(\bar{x})$ est la solution du problème de minimisation

$$\min_{x \in D} f(\bar{x}) + \langle \nabla f(\bar{x}), x - \bar{x} \rangle + g(x) + \frac{1}{\tau} D(x, \bar{x}). \tag{4}$$

7. Expliquer pourquoi ce problème est simple à résoudre. Montrer que si τ est assez petit, on a la règle de descente suivante: pour tout $x \in X$,

$$F(x) + \frac{1}{\tau} D_\psi(x, \bar{x}) \geq F(\hat{x}) + \frac{1}{\tau} D_\psi(x, \hat{x}).$$

8. On définit l'algorithme suivant : on choisit $x^0 \in D$, et pour tout $k \geq 0$, on pose $x^{k+1} = T_\tau x^k$, où $\tau \leq L$ est fixé. Montrer que pour tout $k \geq 0$, $F(x^{k+1}) \leq F(x^k)$. Si x^* est un minimiseur de F dans $\overline{D}$, montrer que

$$F(x^k) - F(x^*) \leq \frac{1}{k\tau} D_\psi(x^*, x^0).$$

9. On suppose que $F(x) \to +\infty$ quand $\|x\| \to +\infty$. Pourquoi peut-on trouver $\tilde{x} \in \overline{D}$ et une sous-suite x^{k_l} tels que $x^{k_l} \to \tilde{x}$ quand $l \to \infty$? Pourquoi est-ce que $\tilde{x}$ est une solution de (P)?

Application: minimisation dans le simplexe unité. On considère le cas où $X = \mathbb{R}^d$,

$$\Sigma = \left\{ x \in X : x_i \geq 0 \, \forall i = 1, \ldots, d; \sum_{i=1}^{d} x_i = 1 \right\}$$

est le simplexe unité et

$$g(x) = \begin{cases} 0 & \text{si } x \in \Sigma \\ +\infty & \text{sinon.} \end{cases}$$

On choisit $\psi(x) = \sum_{i=1}^{d} x_i \ln x_i$ et $D =]0, +\infty[^d$.

10. Exprimer $D_\psi(x, y)$ pour $x \in \Sigma, y \in \Sigma \cap D$.

11. Montrer que l'algorithme décrit précédemment est implémentable : donner les détails du calcul de chaque itération. Indication : introduire le multiplicateur de Lagrange pour la contrainte $\sum_i x_i = 1$.

Exercice III

On considère un opérateur maximal monotone A dans une espace de Hilbert (réel) X. On considère aussi une "métrique" M, c'est-à-dire, un opérateur continu, *coercif*, et symétrique :

$$\|Mx\| \leq \|M\| \|x\| \; \forall x \in X, \quad \langle Mx, x \rangle \geq \delta \|x\|^2, \quad \langle Mx, y \rangle = \langle x, My \rangle$$

pour tout $x, y \in X$, où $\delta > 0$.

1. Montrer que $(x, y) \mapsto \langle Mx, y \rangle =: \langle x, y \rangle_M$ définit un produit scalaire qui est équivalent au produit scalaire $\langle \cdot, \cdot \rangle$. Montrer que pour $y \in X$, le problème

$$\min_x \frac{1}{2} \|x\|_M^2 - \langle y, x \rangle$$

a une solution unique. En déduire que M est inversible. On a noté $\|.\|_M$ la norme Hilbertienne induite par le M-produit scalaire.

2. Montrer que $(M^{-1}A)$ est un opérateur maximal monotone dans le M-produit scalaire. Déduire du théorème de Minty que pour tout $y \in X$, il existe un unique x tel que

$$M(x - y) + Ax \ni 0.$$

3. On considère A, B deux opérateurs maximaux monotones et $K \in \mathcal{L}(X, X)$ un opérateur continu, linéaire dans X. On définit dans $X \times X$ la métrique, pour $\tau, \sigma > 0$,

$$M := \begin{pmatrix} \frac{I}{\tau} & -K^* \\ -K & \frac{I}{\sigma} \end{pmatrix}.$$

Ici $I \in \mathcal{L}(X, X)$ est l'opérateur identité. Montrer que si $\tau\sigma < 1/\|K\|^2$, M est bien continue et coercive dans $X \times X$.

4. En déduire que (pour de tels τ, σ) on peut définir l'algorithme suivant: on choisit $(x^0, y^0) \in X \times X$ et on définit pour tout $k \geq 0$ le nouveau point (x^{k+1}, y^{k+1}) de la manière suivante :

$$M \begin{pmatrix} x^{k+1} - x^k \\ y^{k+1} - y^k \end{pmatrix} + \begin{pmatrix} 0 & K^* \\ -K & 0 \end{pmatrix} \begin{pmatrix} x^{k+1} \\ y^{k+1} \end{pmatrix} + \begin{pmatrix} Ax^{k+1} \\ B^{-1}y^{k+1} \end{pmatrix} \ni 0.$$

Exprimer cette itération sous la forme d'une première formule donnant x^{k+1} à partir de x^k, y^k, puis une seconde donnant y^{k+1} à partir de x^k, x^{k+1}, y^k.

5. Dans quels cas sait-on que (x^k, y^k) converge ? (et en quel sens ?) Dans ce cas, que satisfait la limite $(\bar{x}, \bar{y})$? Écrire en particulier une équation pour $\bar{x}$.

6. On considère maintenant un opérateur maximal monotone C et le nouveau schéma itératif :

$$M \begin{pmatrix} x^{k+1} - x^k \\ y^{k+1} - y^k \end{pmatrix} + \begin{pmatrix} 0 & K^* \\ -K & 0 \end{pmatrix} \begin{pmatrix} x^{k+1} \\ y^{k+1} \end{pmatrix} + \begin{pmatrix} Ax^{k+1} \\ B^{-1}y^{k+1} \end{pmatrix} \ni \begin{pmatrix} Cx^k \\ 0 \end{pmatrix}.$$

Sous quelles conditions sur τ, σ, C est-ce que cet algorithme converge, et vers quelle limite ?