Continuous optimization, an introduction

Assessment

(3rd January 2017)

Exercise I

We recall that for a convex function $f : X \to \mathbb{R} \cup \{+\infty\}$,

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

Evaluate $\mathrm{prox}_{\tau f}(x)$ for $\tau > 0$, and

1. $X = \mathbb{R}$, $f(x) = -\ln x$ for $x > 0$, $+\infty$ for $x < 0$.
2. $f(x) = \psi(\|x\|)$ where $\psi : \mathbb{R} \to \mathbb{R} \cup \{+\infty\}$ is a convex, even (paire) function with $\psi(0) = 0$. Show first that f is a convex function, then evaluate $\mathrm{prox}_{\tau f}$ in terms of $\mathrm{prox}_{\tau\psi}$.
3. $f(x) = \|x\|^3/3$.

Exercise II

We consider X a Hilbert space and a strictly convex lower-semicontinuous (lsc) function $\psi : X \to \mathbb{R} \cup \{+\infty\}$ such that the interior of $\mathrm{dom}\,\psi$, denoted D, is not empty, $\overline{D} = \mathrm{dom}\,\psi$, $\psi \in C^1(D) \cap C^0(\overline{D})$, and $\partial\psi(x) = \emptyset$ for all $x \notin D$. In other words, $\partial\psi(x)$ is either $\emptyset$ (if $x \notin D$), or a singleton $\{\nabla\psi(x)\}$ (if $x \in D$). We also assume that

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

We define the "Bregman distance associated to ψ", denoted $D_\psi(x, y)$, as, for $y \in D$ and $x \in X$,

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

1. Show that $D_\psi(x, y) \geq 0$, and that $D_\psi(x, y) = 0 \Rightarrow y = x$. What other estimate can we write if in addition ψ is strongly convex? Why is D_ψ *not* a distance in the classical sense?

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

3. Let $f : X \to \mathbb{R} \cup \{+\infty\}$ a proper, convex, lsc function. Let $\bar{x} \in D$. We assume that there exists $x \in D$ with $f(x) < +\infty$. Show that there exists a unique point $\hat{x} \in X$ such that

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

4. Explain why $\partial(f + \psi) = \partial f + \partial\psi$. Write the first order optimality condition for $\hat{x}$. Deduce that $\hat{x} \in D$.

5. Show (from the first order optimality condition) that for all $x \in X$,

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

A "nonlinear" descent algorithm. We consider a minimisation problem

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

for f, g convex, lsc, proper functions, where f is C^1 in D and g is "simple" in the following sense: one assume that one knows how to solve

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

for any $\tau > 0$, $p \in X$ and $y \in D$. We suppose in addition that there exists $L > 0$ such that for any $y \in D$, $x \in X$

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

(Here $D_f(x, y) = f(x) - f(y) - \langle \nabla f(y), x - y \rangle$.) We assume that the minimisation problem has a solution. We denote $F(x) = f(x) + g(x)$.

6. Show that if ψ is 1-convex (strongly convex with parameter 1) and f has L-Lipschitz gradient, then (3) is true.

Given $\bar{x} \in D$, $\tau > 0$, we now define the following operator: we let $\hat{x} = T_\tau(\bar{x})$ be the solution of the minimisation problem

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

7. Explain why this problem is easy to solve. Show that if τ is small enough, one has the following descent rule: for all $x \in X$,

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

8. We define the following algorithm: we choose $x^0 \in D$, and for all $k \geq 0$, let $x^{k+1} = T_\tau x^k$, where $\tau \leq L$ is fixed. Show that for all $k \geq 0$, $F(x^{k+1}) \leq F(x^k)$. If x^* is a minimiser of F in $\overline{D}$, show that

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

9. We assume that $F(x) \to +\infty$ when $\|x\| \to +\infty$. Why can we find $\tilde{x} \in \overline{D}$ and extract a subsequence x^{k_l} such that $x^{k_l} \to \tilde{x}$ as $l \to \infty$? Why is $\tilde{x}$ a solution of (P)?

Application: minimisation in the unit simplex. One considers the case where $X = \mathbb{R}^d$,

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

is the unit simplex and

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

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

10. Give the expression of $D_\psi(x, y)$ for $x \in \Sigma, y \in \Sigma \cap D$.

11. Show that the algorithm described in the previous part is implementable: express in detail the computation of the iterations. Hint: introduce the Lagrange multiplier for the constraint $\sum_i x_i = 1$.

Exercise III

We consider a maximal monotone operator A in a (real) Hilbert space X. We consider also a "metric" M, that is, a continuous, *coercive*, and symmetric operator:

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

for all $x, y \in X$, where $\delta > 0$.

1. Show that $(x, y) \mapsto \langle Mx, y \rangle =: \langle x, y \rangle_M$ defines a scalar product which is equivalent to the scalar product $\langle \cdot, \cdot \rangle$. Show that for all $y \in X$, the problem

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

has a unique solution. Deduce that M is invertible. We have denoted $\|.\|_M$ the Hilbertian norm induced by the M-scalar product.

2. Show that $(M^{-1}A)$ is a maximal monotone operator in the M-scalar product. Deduce from Minty's theorem that for any $y \in X$, there exists a unique x such that

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

3. We consider A, B two maximal monotone operators and $K \in \mathcal{L}(X, X)$ a continuous, linear operator in X. We define in $X \times X$ the metric, for $\tau, \sigma > 0$,

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

Here $I \in \mathcal{L}(X, X)$ is the identity operator. Show that if $\tau\sigma < 1/\|K\|^2$, M is continuous and coercive in $X \times X$.

4. Deduce that (for such τ, σ) one can define the following algorithm: we let $(x^0, y^0) \in X \times X$ and define for each $k \geq 0$ the new point (x^{k+1}, y^{k+1}) as follows:

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

Express this as a first iteration defining x^{k+1} from x^k, y^k and then an iteration defining y^{k+1} from x^k, x^{k+1}, y^k.

5. In what case does (x^k, y^k) converge? (and in what sense?) In this case, what does the limit $(\bar{x}, \bar{y})$ satisfy? Write, in particular, an equation for $\bar{x}$.

6. We now consider a maximal monotone operator Cx and the new iterative scheme:

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

Under which condition on τ, σ, C will this iterative scheme be converging? To which limit?