Introduction to Continuous optimization

Assessment

(6th January 2022)

Durée: 3h (It is not necessary to do all *4 exercises!)*

Exercise I: non-linear forward-backward descent

We consider a space X (to simplify, finite-dimensional, yet everything below is dimension independent), with a norm $\|\cdot\|$, and dual X^* with dual norm, for all $u \in X^*$,

$$\|u\|_* = \sup\left\{\langle u, x\rangle_{X^*,X} : \|x\| \leq 1\right\}$$

and we recall (admit) that

$$\|x\| = \sup\left\{\langle u, x\rangle_{X^*,X} : \|u\|_* \leq 1\right\}.$$

In particular, $\langle u, x\rangle_{X^*,X} \leq \|u\|_*\|x\|$ for any $x \in X$, $u \in X^*$. Here, $\langle u, x\rangle_{X^*,X}$ denotes the linear form $u \in X^*$ evaluated at the vector $x \in X$. (In practice, one identifies $X \sim \mathbb{R}^d$, $X^* \sim \mathbb{R}^d$, and $\langle u, x\rangle_{X^*,X} = \sum_{i=1}^d u_i x_i$, where $d \geq 1$ is the dimension. In this case, one can use the standard Euclidean stucture of $\mathbb{R}^d$ to define the convex conjugate, etc.)

1. Let $\mathcal{N}(x) := \|x\|^2/2$. Show that the conjugate

$$\mathcal{N}^*(u) = \sup_x \langle u, x\rangle_{X^*,X} - \mathcal{N}(x)$$

is given by $\|u\|_*^2/2$.

Important remark: We recall that $u \in \partial\mathcal{N}(x) \Leftrightarrow x \in \partial\mathcal{N}^*(u) \Leftrightarrow \langle u, x\rangle_{X^*,X} = \mathcal{N}(x) + \mathcal{N}^*(u)$, with moreover, in that case, using that $\mathcal{N}$ and $\mathcal{N}^*$ are 2-homogeneous, $\langle u, x\rangle_{X^*,X} = 2\mathcal{N}(x) = 2\mathcal{N}^*(u)$ (Legendre-Fenchel's identity plus Euler's identity for homogeneous functions), therefore $\|x\| = \|u\|_*$.

2. We consider a convex, lower-semicontinuous function $F(x) = f(x) + g(x)$, where f, g are convex and where f has L-Lipschitz differential $df : X \to X^*$:

$$\|df(x) - df(y)\|_* \leq L\|x - y\|.$$

We introduce the "Bregman divergence" of f, defined by:

$$D_f(y, x) := f(y) - f(x) - \langle df(x), y - x\rangle_{X^*,X}.$$

Show that $D_f(y, x) \leq L\|y - x\|^2/2 = L\mathcal{N}(y - x)$.

3. (Implicit-explicit algorithm.) We define an iterative algorithm by choosing $x^0 \in X$, $\tau > 0$, and letting, for $k \geq 0$, x^{k+1} be a minimizer of:

$$\min_x g(x) + \left\langle df(x^k), x\right\rangle_{X^*,X} + \frac{1}{\tau}\mathcal{N}(x - x^k).$$

We admit that it exists (it is not difficult), and assume that it can be computed (this is an assumption on g). Write the equation satisfied by x^{k+1}, and show that there is $q^{k+1} \in \partial g(x^{k+1})$ such that:

$$\|x^{k+1} - x^k\| = \tau\|q^{k+1} + df(x^k)\|_*$$

4. Show that for all $x \in X$,

$$F(x) + \frac{1}{\tau}\mathcal{N}(x - x^k) \geq F(x^{k+1}) + \frac{1}{\tau}\mathcal{N}(x^{k+1} - x^k) - D_f(x^{k+1}, x^k).$$

Deduce that if $\tau = \theta/L$ for some $\theta \in]0,1[$, one has:

$$F(x^k) \geq F(x^{k+1}) + \frac{1-\theta}{2\tau}\|x^{k+1} - x^k\|^2.$$

5. Using the convexity of g, f, show that, considering $q^{k+1} \in \partial g(x^{k+1})$ with $\tau\|q^{k+1} + df(x^k)\|_* = \|x^{k+1} - x^k\|$, one has, for any $x^* \in X$:

$$F(x^{k+1}) - F(x^*) \leq \left(\frac{1}{\tau} + L\right)\|x^{k+1} - x^k\|\|x^* - x^{k+1}\|.$$

6. We denote for $k \geq 0$, $\Delta_k := F(x^k) - F(x^*)$, where x^* is a minimizer of F. We now assume that there exists $D > 0$ such that $\|x^k - x^*\| \leq D$ for all $k \geq 0$ (this is clear for instance if the domain of g is bounded). Deduce from the Questions **5.** and **4.** (still using $\tau = \theta/L$) that for all $k \geq 0$:

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

7. Letting $a_k := \frac{1-\theta}{2(1+\theta)^2}\frac{\tau}{D^2}\Delta_k$, one has therefore $a_{k+1} + a_{k+1}^2 \leq a_k$, and $a_k \geq 0$ for all k (assuming x^* is a minimizer of F).

i. show that if $a_0 \geq 2$ and $k \geq \log_2 \log_2 a_0$, then $a_k \leq 2$ (We recall $\log_2 x = \ln x / \ln 2$, so that $2^{\log_2 x} = x$). [Remark: it means for instance that $a_{10} \leq 2$ if $a_0 \approx 10^{300}$.]

ii. show that if $a_{k_0} \leq 2$, for some $k_0 \geq 1$, then:

$$a_k \leq \frac{2}{k - k_0 + 1}.$$

[*Hint:* introduce $b_k := 1/a_k \geq 1/2$ and show that $b_{k+1} \geq b_k + \lambda$, considering the alternatives $b_{k+1}/b_k \geq \lambda$ and $b_{k+1}/b_k \leq \lambda$, for some $\lambda \in (0,1)$ to be determined.]

8. Conclude by giving a convergence rate for the algorithm. Show that (with this analysis) the best choice for θ is $\theta = 1/3$ which gives the rate:

$$F(x^k) - F(x^*) \leq \frac{32D^2L}{1 + k - k_0}.$$

Exercice II - conjugates

1. Let $A \in \mathbb{R}^{n\times n}$ be invertible, and consider

$$f(x) = \frac{1}{2}\|Ax\|^2, \quad (x \in \mathbb{R}^n)$$

Evaluate $\nabla f(x)$. Deduce that $f^*(y) = \left\langle (A^*A)^{-1}y, y\right\rangle/2 = \|(A^*)^{-1}y\|^2/2$.

2. For $x \in \mathbb{R}$, let $f(x) = -\ln(1 - |x|)$ if $|x| < 1$, $+\infty$ if $|x| \geq 1$. Show that $f(x) \geq |x|$. Deduce that $f^*(y) = 0$ if $|y| \leq 1$. Show then that $f^*(y) = (|y| - 1)^+ - \ln(1 + (|y| - 1)^+)$, where $t^+ = \max\{t, 0\}$.

Exercice III - prox and conjugate of entropy and max functions

Let $\Sigma = \{x \in \mathbb{R}^n : \sum_{i=1}^n x_i = 1, x_i \geq 0 \,\forall\, i = 1, \dots, n\}$ be the unit simplex in $\mathbb{R}^n$.

1. Compute the convex conjugate of $g : x \mapsto \sum_{i=1}^n x_i \ln x_i$ if $x \in \Sigma$, and $+\infty$ else, where $0 \ln 0 = 0$.

2. For $\varepsilon > 0$ one considers the "soft-max" function $\varepsilon - \max(y)$, $y \in \mathbb{R}^n$, given by

$$\varepsilon - \max(y) = \varepsilon \ln \sum_{i=1}^n e^{y_i/\varepsilon}.$$

Show that $\max\{y_1, \dots, y_n\} \leq \varepsilon - \max(y) \leq \max\{y_1, \dots, y_n\} + \varepsilon \ln n$.

3. Show that $(\varepsilon - \max)^*(x) = \varepsilon g(x)$ (with g defined in Question **1.**).

4. If $\max(y)$ denotes the function $\max\{y_1, \dots, y_n\}$, deduce that

$$\max{}^*(x) = \begin{cases} 0 & \text{if } \sum_{i=1}^n x_i = 1, x_i \geq 0 \;\forall\; i = 1, \dots, n\,, \\ +\infty & \text{else} \end{cases} \qquad = \delta_\Sigma(x)$$

the characteristic function of the set Σ.

5. One wishes to compute $\operatorname{prox}_{\tau \max}(\bar{x})$ for $\tau > 0$, $\bar{x} \in \mathbb{R}^n$, that is:

$$\arg\min_x \frac{1}{2\tau} \sum_{i=1}^n (x_i - \bar{x}_i)^2 + \max_{i=1}^n x_i$$

Show first that it is equivalent to solve:

$$\min_{t \in \mathbb{R}} \min_{x_i \leq t \,\forall i} t + \frac{1}{2\tau} \sum_{i=1}^n (x_i - \bar{x}_i)^2$$

and then to solve:

$$\min_{t \in \mathbb{R}} t + \frac{1}{2\tau} \sum_{i=1}^n [(\bar{x}_i - t)^+]^2$$

where $z^+ := \max\{z, 0\}$ denotes the "positive part" of $z \in \mathbb{R}$.

6. Show that the optimal t exists and satisfies:

$$\sum_{i=1}^n (\bar{x}_i - t)^+ = \tau.$$

Deduce that $t < \max_{i=1}^n \bar{x}_i$.

7. Can you imagine an algorithm to compute t?

8. Assuming the previous question is solved, deduce an algorithm for projecting onto the unit simplex Σ.

Exercice IV: epi-convergence

Let $(C_n)_n$ be a sequence of closed, convex subsets of $\mathbb{R}^d$, $C \subset \mathbb{R}^d$. $\mathbb{R}^d$ is equipped with the Euclidean norm.
We say that $C_n \overset{K}{\to} C$ (convergence in the sense of Kuratowski) if and only if:

i. for all $x \in C$, there exists a sequence $(x_n)_n$ with $x_n \in C_n$ for all n and such that $x_n \overset{n\to\infty}{\longrightarrow} x$;
ii. if $x_{n_k} \in C_{n_k}$ (for a subsequence) and if $x_{n_k} \overset{k\to\infty}{\longrightarrow} x$, then $x \in C$.

1. Distance function We introduce $d_n(x) = \text{dist}\,(x, C_n) = \min_{y \in C_n} \|x - y\| \geq 0$. Why is there a unique $y \in C$ with $d_n(x) = \|x - y\|$? Show that for each n, d_n is 1-Lipschitz, and convex.

2. We recall Ascoli-Arzelà's theorem:

Theorem 1 (Ascoli-Arzelà)**.** *If $f_n : \mathbb{R}^d \to \mathbb{R}$ are functions which are uniformly equi-continuous, and uniformly bounded in some point, then there is a subsequence f_{n_k} which converges locally uniformly.*

Uniformly equi-continuous means that $\forall \varepsilon > 0$, $\exists \eta > 0$, $\forall x, x' \in \mathbb{R}^d$, $\|x - x'\| \leq \eta \Rightarrow (\forall n,\ |f_n(x) - f_n(x')| \leq \varepsilon)$. Show that either $d_n(x) \to \infty$ for all $x \in \mathbb{R}^d$, or there exists a function d and a subsequence d_{n_k} such that $d_{n_k} \to d$ locally uniformly.

3. We assume $d_{n_k} \to d$ locally uniformly. Let $C := \{x \in \mathbb{R}^d : d(x) = 0\}$. Show that $C_{n_k} \xrightarrow{K} C$, and that C is closed and convex.

4. Show that in this case, $d(x) = \text{dist}\,(x, C)$ for all x.

5. Show that if $d_n \to +\infty$, then $C_n \xrightarrow{K} \emptyset$.
We have shown that the K-convergence is compact on the set of closed, convex[1] sets: given any sequence (C_n) of closed (convex) sets, there is a subsequence which converges to a closed, convex set (but possibly empty).

6. Let $f_n : \mathbb{R}^d \to \mathbb{R} \cup \{+\infty\}$ be convex, proper lower semi-continuous functions. Let $C_n = \text{epi}\, f_n = \{(x,t) \in \mathbb{R}^{d+1}, t \geq f_n(x)\}$. By the previous result, there exists C, closed and convex and a subsequence with $C_{n_k} \xrightarrow{K} C$ (in $\mathbb{R}^{d+1}$). Show that $C = \text{epi}\, f$ for some convex, lower-semicontinuous function f. When is f not proper? We say that f_{n_k} "epi-converges" to f.

7. We assume now that $f_n \geq 0$ for all n, and that $\sup_n \min_{\overline{B}(0,1)} f_n < +\infty$. Show that f is proper. ($\overline{B}(0,1) = \{x \in \mathbb{R}^d : \|x\| \leq 1\}$.)

8. We assume f_n epi-converges to f which is proper. Show that f_n "Γ-converges" to f, that is:

(Γ^-) for all x and $x_n \to x$, $f(x) \leq \liminf_n f_n(x_n)$;

(Γ^+) for all x, there exists $x_n \to x$ such that $\limsup_n f_n(x_n) \leq f(x)$ (so that, by (Γ^-), $\lim_n f_n(x_n) = f(x)$).

[Hint: use (ii) for (Γ^-) and (i) for (Γ^+).]

9. In the case of the previous question, assuming in addition (to simplify) $f_n \geq 0$, let $x \in \mathbb{R}^d$ and $x_n \to x$: show, using properties (Γ^+) and (Γ^-), that $\lim_{n\to\infty} \text{prox}_{f_n}(x_n) = \text{prox}_f(x)$.

(One has to show (1) that $\text{prox}_{f_n}(x_n)$ is bounded, (2) that any limit point has to be $\text{prox}_f(x)$.)

10. We consider f_n convex, proper, lsc, which Γ-converges to f, convex, proper, lsc. Show that f_n^* (the convex conjugate) Γ-converges to f^*.

(a.) show first, using (Γ^+) for f_n, that (Γ^-) holds for f_n^*;

(b.) to show (Γ^+), we first admit that it is enough to show the property for $y \in \mathbb{R}^d$ such that $\partial f^*(y) \neq \emptyset$, so that there is $x \in \partial f^*(y) \Leftrightarrow y \in \partial f(x)$.

What is the minimizer of $z \mapsto f(z) - \langle y, z \rangle + \|z - x\|^2/2$? Introduce z_n as the minimizer of $f_n(z) - \langle y, z \rangle + \|z - x\|^2/2$ and show (using Question **9.**) that $z_n \to x$. Then, let $y_n = y - z_n + x \to y$: use Legendre-Fenchel's inequality to show that $f_n^*(y_n) \to f^*(y)$.

[1] In fact it is compact on the set of closed sets.