**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 $\|u\|_*^2/2$.

One has, letting $x = ty$ for $t = \|x\|$:

$$\mathcal{N}^*(u) = \sup_{t\geq 0, \|y\|\leq 1} \langle u, ty\rangle_{X^*,X} - \frac{t}{2}^2 = \sup_{t\geq 0} t\|u\|_* - \frac{t}{2}^2 = \frac{1}{2}\|u\|_*^2.$$

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

This is the classical Taylor expansion:

$$\begin{aligned} f(y) = f(x) + \int_0^1 \langle df(x + s(y - x)), y - x\rangle_{X^*,X}\, ds \\ = f(x) + \langle df(x), y - x\rangle_{X^*,X} + \int_0^1 \langle df(x + s(y - x)) - df(x), y - x\rangle_{X^*,X}\, ds, \end{aligned}$$

so that:

$$D_f(y,x) = \int_0^1 \langle df(x+s(y-x)) - df(x), y-x\rangle_{X^*,X}\, ds.$$

Then, we use

$$\langle df(x+s(y-x)) - df(x), y-x\rangle_{X^*,X} \le \|df(x+s(y-x)) - df(x)\|_* \|y-x\| \le Ls\|y-x\|^2,$$

and $\int_0^1 s ds = 1/2$ to conclude.

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

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

Since $\left\langle df(x^k), x\right\rangle + \frac{1}{\tau}\mathcal{N}(x - x^k)$ is continuous with full domain in $E \sim \mathbb{R}^d$, one can apply the results from the lecture notes and one has at $x^{k+1}$:

$$0 \in \partial g(x^{k+1}) + df(x^k) + \frac{1}{\tau}\partial\mathcal{N}(x^{k+1} - x^k)$$

which means that there exists $q^{k+1} \in \partial g(x^{k+1})$ such that

$$-\tau(q^{k+1} + df(x^k)) \in \partial\mathcal{N}(x^{k+1} - x^k).$$

By the results in Question **1.**, it implies in particular that:

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

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

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

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

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

By definition, $x^{k+1}$ satisfies that for all $x \in X$,

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

Now, by convexity,

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

and we deduce:

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

We conclude using that $g(x^{k+1}) + f(x^k) + \left\langle df(x^k), x^{k+1} - x^k\right\rangle_{X^*,X} = F(x^{k+1}) - D_f(x^{k+1}, x^k)$. Then, using Question **2.**, it follows, if we take $x = x^k$:

$$F(x^k) \ge F(x^{k+1}) + \frac{1}{\tau}\mathcal{N}(x^{k+1} - x^k) - \frac{L}{2}\|x^{k+1} - x^k\|^2 = F(x^{k+1}) + \left(\frac{1}{\tau} - L\right)\mathcal{N}(x^{k+1} - x^k)$$

so that if $\tau \le \theta/L \Leftrightarrow L \le \theta/\tau$, we can show the required inequality.

**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 $x^* \in X$:

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

One has

$$F(x^*) = f(x^*) + g(x^*) \geq f(x^{k+1}) + g(x^{k+1}) + \langle df(x^{k+1}) + q^{k+1}, x^* - x^{k+1} \rangle_{X^*,X}$$

Hence

$$\begin{aligned} F(x^{k+1}) - F(x^*) &\leq -\langle df(x^k) + q^{k+1}, x^* - x^{k+1} \rangle_{X^*,X} - \langle df(x^{k+1}) - df(x^k), x^* - x^{k+1} \rangle_{X^*,X} \\ &\leq \|df(x^k) + q^{k+1}\|_* \|x^* - x^{k+1}\| + \|df(x^{k+1}) - df(x^k)\|_* \|x^* - x^{k+1}\| \\ &\leq \frac{1}{\tau}\|x^{k+1} - x^k\| \|x^* - x^{k+1}\| + L\|x^{k+1} - x^k\| \|x^* - x^{k+1}\| \end{aligned}$$

**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 $k \geq 0$:

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

We have from **5.**:

$$\Delta_{k+1} \leq \frac{1+\theta}{\tau} D\|x^{k+1} - x^k\| \;\Rightarrow\; \mathcal{N}(x^{k+1} - x^k) \geq \frac{1}{2} \frac{\tau^2}{(1+\theta)^2 D^2} \Delta_{k+1}^2.$$

From **4.**,

$$\Delta_{k+1} + \frac{1-\theta}{\tau} \mathcal{N}(x^{k+1} - x^k) \leq \Delta_k.$$

Hence,

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

**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.]

For the first point (i.), we use that $a_{k+1} \leq \sqrt{a_k}$, that is, $\log_2 a_{k+1} \leq \frac{1}{2}\log_2 a_k$. By induction it follows $\log_2 a_k \leq 2^{-k}\log_2 a_0$. In particular, $\log_2 a_k \leq 1$ as soon as $2^{-k}\log_2 a_0 \leq 1$, that is $2^k \geq \log_2 a_0$, or $k \geq \log_2 \log_2 a_0$.

For the second point, following the hint, we introduce $b_k = 1/a_k$ and write that

$$\frac{1}{b_{k+1}}\left(1 + \frac{1}{b_{k+1}}\right) \leq \frac{1}{b_k} \;\Leftrightarrow\; b_k\left(1 + \frac{1}{b_{k+1}}\right) \leq b_{k+1} \;\Leftrightarrow\; b_{k+1} \geq b_k + \frac{b_k}{b_{k+1}}.$$

We remark that $a_k \leq 2$ for any $k \geq k_0$ so that $b_k \geq 1/2$.

Let $\lambda \in (0,1)$: either $b_k \geq \lambda b_{k+1}$ and one has $b_{k+1} \geq b_k + \lambda$, or $b_k \leq \lambda b_{k+1}$ and one has $b_{k+1} \geq b_k + (\frac{1}{\lambda} - 1)b_k \geq b_k + \frac{1-\lambda}{2\lambda}$. The best choice for $\lambda$ is to choose $\lambda = (1-\lambda)/(2\lambda)$, that is

$$2\lambda^2 + \lambda - 1 = 0, \lambda \in (0,1) \Leftrightarrow \lambda = \frac{1}{2}.$$

We deduce by induction that $b_k \geq b_{k_0} + (k - k_0)/2 \geq (1 + k - k_0)/2$, hence the result.

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

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

It follows that

$$F(x^k) - F(x^*) \leq \frac{4D^2}{\tau} \frac{(1+\theta)^2}{1-\theta} \frac{1}{1+k-k_0} \leq \frac{(1+\theta)^2}{\theta(1-\theta)} \frac{4D^2L}{1+k-k_0}$$

for $k_0 \geq \log_2 \log_2[(1-\theta)\tau/((1+\theta)^2D^2)(F(x_0) - F(x^*))]$. Minimizing this rate with respect to $\theta$ gives $\theta = 1/3$ and $(1+\theta)^2/(\theta(1-\theta)) = 8$.

## 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 $f^*(y) = \left\langle (A^*A)^{-1}y, y\right\rangle/2 = \|(A^*)^{-1}y\|^2/2$.

Observe that

$$f(x+ty) = \frac{1}{2}\|A(x+ty)\|^2 = f(x) + t\langle Ax, Ay\rangle + t^2 f(y) = f(x) + t\langle A^*Ax, y\rangle + o(t)$$

which shows that $\nabla f(x) = A^*Ax$.

Then, to compute $\sup_x \langle x, y\rangle - \|Ax\|^2/2$, one sees that at the maximum,

$$y - A^*Ax = 0$$

Since $A$ is invertible, one has $x = (A^*A)^{-1}y$. One deduces that

$$\langle x, y\rangle - \frac{1}{2}\|Ax\|^2 = \left\langle (A^*A)^{-1}y, y\right\rangle - \frac{1}{2}\|A(A^{-1}(A^*)^{-1}y\|^2 = \frac{1}{2}\|(A^*)^{-1}y\|^2$$

**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 $t^+ = \max\{t, 0\}$.

One has $-\ln(1+t) \geq -\ln(1) - t = -t$ by convexity hence $f(x) \geq |x|$. In particular $f^*(y) \leq \delta_{[-1,1]}(y)$ and $f^* \leq 0$ on $[-1,1]$. Since $f^*(y) \geq 0 \cdot y - f(0) = 0$ we find that $f^*(y) = 0$ if $|y| \leq 1$.

Now for $|y| > 1$, we compute $\sup_x xy + \ln(1-|x|)$: one has $y - \frac{\partial|\cdot|(x)}{1-|x|} = 0$ (or one can easily check that the sup is not at $x = 0$, so that $|x|' = \pm 1$), i.e., for $|y| > 1$, $y = \operatorname{sign}(x)/(1-|x|)$ for some $|x| < 1$. In particular $|y| = 1/(1-|x|)$ and $|x| = (|y|-1)/|y|$ so that $x = (|y|-1)/y$. Then, $xy + \ln(1-|x|) = |y| - 1 - \ln|y| = f^*(y)$.

All-in-all, $f^*(y) = (|y|-1)^+ - \ln(1 + (|y|-1)^+)$.

## 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 $0 \ln 0 = 0$.

One has to compute

$$\sup_{\sum_i x_i = 1} \sum_{i=1}^n x_i y_i - x_i \ln x_i .$$

At the maximum, one has $y_i - \ln x_i - 1 = \lambda$ where $\lambda$ is the multiplier for $\sum_i x_i = 1$. In addition, one has $\sum_i x_i y_i - x_i \ln x_i = (\lambda + 1) \sum_i x_i = \lambda + 1$. One has

$$x_i = e^{-1-\lambda} e^{y_i}, \quad e^{-1-\lambda} \sum_{i=1}^n e^{y_i} = 1$$

so that

$$g^*(y) = \lambda + 1 = \ln \sum_{i=1}^n e^{y_i} .$$

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

If $\bar{y} = \max_i y_i$, one has $\sum_{i=1} e^{y_i/\varepsilon} \geq e^{\bar{y}/\varepsilon}$, and $\sum_{i=1} e^{y_i/\varepsilon} \leq n e^{\bar{y}/\varepsilon}$. The result follows.

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

One sees that

$$(\varepsilon g)^*(y) = \sup_x x \cdot y - \varepsilon g(x) = \varepsilon \sum_x x \cdot (y/\varepsilon) - g(x) = \varepsilon g^*(y/\varepsilon)$$

and the results follows from the first question, and the fact $(\varepsilon g)^{**} = \varepsilon g$.

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

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

the characteristic function of the set $\Sigma$.

In fact it is easy to see that the conjugate of the right-hand side is the max function. But one also has, from question **2.**,

$$\max{}^*(x) \geq (\varepsilon - \max)^*(x) = \varepsilon g(x) \geq \max{}^*(x) - \varepsilon \ln n,$$

so that $\max^*(x) = \lim_{\varepsilon \to 0} \varepsilon g(x) = 0$ for $x \in \operatorname{dom} g = \Sigma$, $+\infty$ else.

**5.** One wishes to compute $\mathrm{prox}_{\tau \max}(\bar{x})$ 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:

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

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 $z \in \mathbb{R}$.

The first statement is obvious, since $\max_{i=1}^n x_i = \min_{t: x_i \le t\, \forall i} t$ by definition. Then, solving $\min_{x_i \le t}(x_i - \bar{x}_i)^2$ yields $x_i = \bar{x}_i$ if $\bar{x}_i \le t$, $x_i = t$ else. In particular, $\bar{x}_i - x_i = 0$ if $\bar{x}_i \le t$, and $\bar{x}_i - x_i = \bar{x}_i - t$ if $\bar{x}_i \ge t$, that is: $\bar{x}_i - x_i = (\bar{x}_i - t)^+$.

**6.** Show that the optimal $t$ exists and satisfies:

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

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

In fact the function in the min in the previous question is $C^1$, goes to $\infty$ as $t \to +\infty$, and if $t \le \bar{x}_i$ for all $i$, it is $t + \sum_i (x_i - t)^2/(2\tau)$ which also goes to $+\infty$ as $t \to -\infty$. Hence it reaches a minimum at some $t \in \mathbb{R}$ where the derivative vanishes.

The derivative is $1 - \sum_{i=1}^n (\bar{x}_i - t)^+/\tau$, hence the equation. In addition, the left-hand side of the equation can be positive only if $t < \max_i \bar{x}_i$.

**7.** Can you imagine an algorithm to compute $t$?

To compute $t$, the best method is to first sort the values $(\bar{x}_i)$ by a sorting algorithm. Then, we assume $\bar{x}_1 \ge \bar{x}_2 \ge \cdot \ge \bar{x}_n$, and we have to guess $k \ge 1$ such that $t \ge \bar{x}_i$ for $i \ge k+1$, $t < \bar{x}_k$. If $k = 1$, then one should have

$$(\bar{x}_1 - t) = \tau \quad \Leftrightarrow \quad t = \bar{x}_1 - \tau.$$

If $\bar{x}_1 - \tau \ge \bar{x}_2$ then the value is admissible and the problem is solved, otherwise we try $k = 2$. In general, for a given $k$, one should have

$$t = \frac{1}{k}\sum_{i=1}^k \bar{x}_i - \frac{\tau}{k}$$

which is admissible if $\bar{x}_i \ge t$ for $i = 1, \dots, k$ and $\bar{x}_k \le t$. Note that (denoting $t_k$ the value of $t$ for a guess $k$)

$$t_k = \frac{1}{k}\bar{x}_k + \frac{k-1}{k} t_{k-1}$$

so computing the successive value of $t_k$ does not need more than one additional operation, hence the overall complexity is the time for sorting, plus $O(n)$.

**8.** Assuming the previous question is solved, deduce an algorithm for projecting onto the unit simplex $\Sigma$.

We have $\max^* = \delta_\Sigma$ the characteristic of $\Sigma$, and $\Pi_\Sigma(x) = \mathrm{prox}_{\delta_\Sigma}(x)$. Hence, one has by Moreau's identity:

$$\Pi_\Sigma(x) = x - \mathrm{prox}_{\max}(x)$$

which is computed in the previous question.

## 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 $C_n \xrightarrow{K} C$ (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 $x_n \xrightarrow{n\to\infty} x$;

ii. if $x_{n_k} \in C_{n_k}$ (for a subsequence) and if $x_{n_k} \xrightarrow{k\to\infty} x$, then $x \in C$.

**1. Distance function** We introduce $d_n(x) = \operatorname{dist}(x, C_n) = \min_{y\in C_n} \|x - y\| \geq 0$. 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.

The projection theorem says there is a unique $y = \Pi_{C_n}(x)$, projection of $x$ onto the closed convex set $C_n$. Now, let $x, x' \in \mathbb{R}^d$ and $y \in C_n$ such that $\|x - y\| = d_n(x)$. One has $d_n(x') - d_n(x) \leq \|x' - y\| - \|x - y\| \leq \|x - x'\|$.

Eventually, $d_n$ is convex as the minimizer with respect to $y$ of the jointly convex function $(x, y) \mapsto \|x - y\| + \delta_{C_n}(y)$ ($\delta_{C_n}$ the characteristic function of $C_n$). This is easily proved by considering $x, x', y, y' \in C_n$ with $d_n(x) = \|x-y\|$, $d_n(x') = \|x'-y'\|$, $t \in [0, 1]$, and using that $d_n(tx+(1-t)x') \leq \|tx+(1-t)x'-(ty+(1-t)y')\|$ since $ty + (1 - t)y' \in C_n$.

**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.

If for some $x$, $d_n(x) \not\to +\infty$, then there exists a subsequence such that $d_{n_k}(x)$ is bounded (recall $d_n \geq 0$). Then, $d_{n_k}$ is uniformly bounded at the point $x$. In addition, it is uniformly equi-continuous because all the functions are 1-Lipschitz (the property holds with $\eta = \varepsilon$). Hence by Ascoli-Arzelà's theorem, there is a further subsequence (still denoted $d_{n_k}$) which converges locally uniformly to some limit function $d$.

**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.

Let $x \in C$: then $d_{n_k}(x) \to 0$. It means there exists $x_k \in C_{n_k}$ with $d_{n_k}(x) = \|x - x_k\| \to 0$. This shows property (i) of the K-limit. For (ii): let now $x_k \in C_{n_k}$ for all $k$ and and assume a subsequence $x_{k_l}$ converges to a point $x \in \mathbb{R}^d$. Then $0 = d_{n_{k_l}}(x_{k_l}) \to d(x)$ by uniform convergence, and $d(x) = 0$, hence $x \in C$.

$C$ is closed and convex as the minimal level set of $d$ which is continuous and convex.

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

Let $x \in \mathbb{R}^d$ First, let $y \in C$ (which is closed and convex) such that $\|x - y\| = \operatorname{dist}(x, C)$. By (i.) there exists $y_k \in C_{n_k}$ with $y_k \to y$, so that $d_{n_k}(x) \leq \|x - y_k\|$, and in the limit we find $d(x) \leq \operatorname{dist}(x, C)$.

On the other hand, let now $y_k$ such that $d_{n_k}(x) = \|x - y_k\| \to d(x)$. We have that $\|y_k\| \leq \|x\| + d_{n_k}(x)$ is bounded hence up to a subsequence, it has a limit $y$. By (ii), $y \in C$ and one has $d_{n_k}(x) \to d(x) = \|x - y\| \geq \operatorname{dist}(x, C)$.

Hence $d(x) = \operatorname{dist}(x, C)$.

**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).

Point (i): is true since there is no $x \in \emptyset$. Point (ii): is true since given $x_n \in C_n$, $d_n(0) \leq \|x_n\|$ and since $d_n(0) \to \infty$, $x_n$ has no converging subsequence.

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

**6.** Let $f_n : \mathbb{R}^d \to \mathbb{R} \cup \{+\infty\}$ be convex, proper lower semi-continuous functions. Let $C_n = \operatorname{epi} f_n = \{(x,t) \in \mathbb{R}^{d+1}, t \ge f_n(x)\}$. 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 $C = \operatorname{epi} f$ for some convex, lower-semicontinuous function $f$. When is $f$ not proper? We say that $f_{n_k}$ "epi-converges" to $f$.

First, we know that $C$ is closed and convex. Then, it is the epigraph of some $f$ if and only if $(x,t) \in C \Rightarrow (x,t') \in C$ for all $t' \ge t$. In that case, $f(x)$ is given by $\inf\{t \in \mathbb{R} : (x,t) \in C\}$ which is either $+\infty$ (if the set is empty), $-\infty$ (if the set is $\mathbb{R}$), or a min.

Let $(x,t) \in C$. Then, by (i) there is $(x_n, t_n) \in C_n$ with $(x_n, t_n) \to (x,t)$. If $t' \ge t$, one has $t_n + t' - t \ge t_n \ge f_n(x_n)$ hence $(x_n, t_n + t' - t) \in C_n$. By (ii), since $(x_n, t_n + t' - t) \to (x, t')$, one deduces $(x,t') \in C$. Hence if $f(x) := \inf\{t \in \mathbb{R} : (x,t) \in C\}$ one has $C = \{(x,t) : t \ge f(x)\}$.

If $C = \emptyset$, $f \equiv +\infty$ is not proper. If $C$ contains $\{x\} \times \mathbb{R}$ for some $x \in \mathbb{R}^d$, $f(x) = -\infty$ and $f$ and is not proper either. Otherwise, $f$ must be proper.

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

$C_n \subset \mathbb{R}^d \times \mathbb{R}_+$ since $f_n \ge 0$, so that $C \subset \mathbb{R}^d \times \mathbb{R}_+$, that is $f \ge 0$, and one cannot have $f = -\infty$. On the other hand, letting $t = \sup_n \min_{\overline{B}(0,1)} f_n$, for each $n$ there is $x_n \in B(0,1)$ such that $(x_n, t) \in \operatorname{epi} f_n = C_n$, and one has $d_n((0,0)) \le \|x_n\| + t \le 1 + t$ which is bounded: hence we are in the situation where some subsequence $d_{n_k} \to d = \operatorname{dist}(\cdot, C)$ locally uniformly and $C$ is not empty.

**8.** We assume $f_n$ epi-converges to $f$ which is proper. Show that $f_n$ "$\Gamma$-converges" to $f$, that is:

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

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

[Hint: use (ii) for ($\Gamma^-$) and (i) for ($\Gamma^+$).]

($\Gamma^-$): Let $x \in \mathbb{R}^d$ and $x_n \to x$. Assume $\liminf_n f_n(x_n) < +\infty$. Consider a subsequence $f_{n_k}(x_{n_k}) \to \liminf_n f_n(x_n)$, we also may assume that $f_{n_k}(x_{n_k}) < +\infty$ for all $k$. Let $t > \liminf_n f_n(x_n)$, then for $k$ large enough, $f_{n_k}(x_{n_k}) \le t$ and $(x_{n_k}, t) \in \operatorname{epi} f_{n_k}$. By (ii), we deduce that $(x,t) \in \operatorname{epi} f$, hence $f(x) \le t$. Since this is true for all $t > \liminf_n f_n(x_n)$ it follows that $f(x) \le \liminf_n f_n(x_n)$.

($\Gamma^+$): Let $x \in \mathbb{R}^d$, assume $f(x) < +\infty$, and let $t \ge f(x)$ so that $(x,t) \in \operatorname{epi} f$. Then by (i), there exists $(x_n, t_n) \in \operatorname{epi} f_n$ with $(x_n, t_n) \to (x,t)$. In particular, $\limsup_n f_n(x_n) \le \limsup_n t_n = t$. Since $f$ is proper, lsc, one can take $t = f(x)$ and the property is proved.

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

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

Let $z_n = \operatorname{prox}_{f_n}(x_n)$, which minimizes $f_n(z) + \|z - x_n\|^2/2$. First, $f$ is proper, hence there is $\bar{x}$ with $f(\bar{x}) \in \mathbb{R}$, and there is $\bar{x}_n$ with $\lim_n f_n(\bar{x}_n) = f(\bar{x})$.

In particular, $f_n(z_n) + \|z_n - x_n\|^2 \le f_n(\bar{x}_n) + \|\bar{x}_n - x_n\|^2$ is bounded, and since $f_n(z_n) \ge 0$ it shows that $z_n$ is bounded. Up to a subsequence we may assume that $z_{n_k} \to z$.

Now, we have by ($\Gamma^-$):

$$f(z) + \frac{1}{2}\|z - x\|^2 \le \liminf_k f_{n_k}(z_{n_k}) + \frac{1}{2}\|z_{n_k} - x_{n_k}\|^2.$$

On the other hand, if $z' \in \mathbb{R}^d$, by ($\Gamma^+$) there exists $z'_n$ with $\limsup_n f_n(z'_n) \le f(z)$. By minimality, one has

$$f_{n_k}(z_{n_k}) + \frac{1}{2}\|z_{n_k} - x_{n_k}\|^2 \le f_{n_k}(z'_{n_k}) + \frac{1}{2}\|z'_{n_k} - x_{n_k}\|^2$$

hence $\liminf_k f_{n_k}(z_{n_k}) + \|z_{n_k} - x_{n_k}\|^2/2 \le \limsup_k f_{n_k}(z'_{n_k}) + \|z'_{n_k} - x_{n_k}\|^2/2 \le f(z') + \|z' - x\|^2/2.$

We deduce that

$$f(z) + \frac{1}{2}\|z - x\|^2 \le f(z') + \frac{1}{2}\|z' - x\|^2,$$

since $z'$ is arbitrary it shows that $z = \mathrm{prox}_f(x)$. Since the limit of any converging subsequence is the same, we deduce that $z_n \to z$.

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

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

(b.) to show $(\Gamma^+)$, 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 $f_n^*(y_n) \to f^*(y)$.

Let $(y_n)$ with $\liminf_n f^*(y_n) < +\infty$, and assume $y_n \to y$. Let $x \in \mathbb{R}^d$, let $x_n \to x$ with $\limsup_n f_n(x_n) \le f(x)$. Then,

$$f_n^*(y_n) \ge \langle x_n, y_n\rangle - f_n(x_n) \quad \Rightarrow \quad \liminf_n f^*(y_n) \ge \liminf_n \langle x_n, y_n\rangle - f_n(x_n)$$
$$= \lim_n \langle x_n, y_n\rangle - \limsup_n f_n(x_n) \ge \langle x, y\rangle - f(x).$$

We deduce, taking the sup over $x$, that

$$\liminf_n f_n^*(y_n) \ge f^*(y).$$

For the limsup: as suggested, let $z_n$ minimize $f_n(z) - \langle y, z\rangle + \|z - x\|^2/2$, that is $z_n = \mathrm{prox}_{f_n - \langle y, \cdot\rangle}(x)$. Then by Question **9.**, $z_n \to \mathrm{prox}_{f - \langle y, \cdot\rangle}(x) = x$ (since $y \in \partial f(x)$).

We have

$$\partial f_n(z_n) - y + z_n - x \ni 0 \Leftrightarrow y_n := y - z_n + x \in \partial f_n(z_n).$$

In particular $y_n \to y$. We write:

$$f_n(z_n) + f_n^*(y_n) = \langle z_n, y_n\rangle.$$

In the limit using (a.), we find $f(x) + f^*(y) \le \langle x, y\rangle$ but then, one has:

$$\langle x, y\rangle \le f(x) + f^*(y) \le \liminf_n f_n(z_n) + \liminf_n f_n^*(y_n) \le \liminf_n (f_n(z_n) + f_n^*(y_n))$$
$$\le \limsup_n (f_n(z_n) + f_n^*(y_n)) \le \lim_n \langle z_n, y_n\rangle = \langle x, y\rangle$$

showing that $f_n(z_n) + f_n^*(y_n) \to f(x) + f^*(y)$. But then,

$$f(x) + \limsup_n f_n^*(y_n) \le \liminf_n f_n(z_n) + \limsup_n f_n^*(y_n) \le \limsup_n (f_n(z_n) + f_n^*(y_n)) = f(x) + f^*(y),$$

and we deduce that $f_n^*(y_n) \to f^*(y)$, as we needed to show (and also that $f_n(z_n) \to f(x)$).