Introduction to Continuous optimization
Assessment
(19th January 2024)
Durée: 3h

Exercise I: Frank-Wolfe algorithm

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 $x \in X$.

Let C be a bounded, closed, convex set and $f : C \to [0, +\infty)$ a convex function, C^1, such that df is L-Lipschitz. This means that for any $x, y \in C$:

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

We want to solve

$$\min_{x \in C} f(x), \tag{P}$$

and we assume there exists a solution $x^* \in C$.

We make the assumption that for any $p \in X^*$, it is easy to find $x \in \arg\min_{x \in C} \langle p, x \rangle_{X^*,X}$.

1. Show that for any $x, y \in C$,

$$f(y) \leq f(x) + \langle df(x), y - x \rangle_{X^*,X} + \frac{L}{2}\|y - x\|^2.$$

This is the classical Taylor expansion:

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

so that:

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

Then, we use

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

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

2. We introduce the Frank-Wolfe algorithm: given $x^0 \in C$, we solve for each $k \geq 1$:

$$\begin{cases} s^k \in \arg\min_{x \in C} f(x^{k-1}) + \langle df(x^{k-1}), x - x^{k-1} \rangle_{X^*,X} , \\ x^k = x^{k-1} + \omega_k (s^k - x^{k-1}) \end{cases}$$

where ω_k is a given step (in practice we will use $\omega_k = 2/(k+1)$, $k \geq 1$ and in particular, $\omega_1 = 1$).

We denote $\Delta_k := f(x^k) - f(x^*)$. Show that for $k \geq 1$,

$$\Delta_k \leq \varepsilon_k := -\langle df(x^k), s^{k+1} - x^k \rangle_{X^*,X}$$

This is just convexity, indeed for all $x \in C$ we have if $k \geq 1$

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

by definition of s^{k+1}. Choose then $x = x^*$.

3. Using **1.** for $y = x^{k+1}$, $x = x^k$, deduce from **2.** that for all $k \geq 0$,

$$\Delta_{k+1} \leq (1 - \omega_{k+1})\Delta_k + \frac{L}{2}\omega_{k+1}^2 \|s^{k+1} - x^k\|^2$$

We have for $k \geq 0$,

$$f(x^{k+1}) \leq f(x^k) + \langle df(x^k), x^{k+1} - x^k \rangle + \frac{L}{2}\|x^{k+1} - x^k\|^2.$$

Using $x^{k+1} - x^k = \omega_{k+1}(s^{k+1} - s^k)$ we deduce:

$$f(x^{k+1}) \leq f(x^k) + \omega_{k+1} \langle df(x^k), s^{k+1} - x^k \rangle + \frac{L\omega_{k+1}^2}{2}\|s^{k+1} - x^k\|^2.$$

From **2.** we have $\langle df(x^k), s^{k+1} - x^k \rangle_{X^*,X} = -\varepsilon_k \leq -\Delta_k$. Hence, removing $f(x^*)$ on both sides of the previous inequality, we find:

$$\Delta_{k+1} \leq \Delta_k - \omega_{k+1}\Delta_k + \frac{L\omega_{k+1}^2}{2}\|s^{k+1} - x^k\|^2.$$

4. We call $D = \max_{x,y \in C} \|x - y\|$ the diameter of C. We assume that $\omega_k = 2/(k+1)$ for $k \geq 1$. Deduce that $\Delta_1 \leq \frac{LD^2}{2}$. We let $M := LD^2/2$. Show then by induction that

$$\Delta_k \leq \frac{4M}{k+3}$$

We have $\omega_1 = 1$ so that from **3.**, with $k = 0$, we find:

$$\Delta_1 \leq (1 - \omega_1)\Delta_0 + \frac{LD^2}{2}\omega_1^2 = \frac{LD^2}{2} =: M.$$

Then, in particular, $\Delta_1 \leq \frac{4M}{1+3}$.

Assume $\Delta_k \leq \frac{4M}{k+3}$. We find:

$$\begin{aligned} \Delta_{k+1} \leq & (1 - \omega_{k+1})\frac{4M}{k+3} + \omega_{k+1}^2 M \\ &= M\left(\frac{4k}{(k+2)(k+3)} + \frac{4}{(k+2)^2}\right) \\ &= \frac{4M}{k+2}\left(\frac{k}{k+3} + \frac{1}{k+2}\right) = \frac{4M}{k+4}\frac{k+4}{k+2}\left(\frac{k}{k+3} + \frac{1}{k+2}\right). \end{aligned}$$

Now, we observe that

$$(k+4)\left(\frac{k}{k+3}+\frac{1}{k+2}\right)=k\left(1+\frac{1}{k+3}\right)+1+\frac{2}{k+2}=k+1+\frac{k}{k+3}+\frac{2+k-k}{k+2}$$
$$=k+2+\frac{k}{k+3}-\frac{k}{k+2}\le k+2$$

hence

$$\frac{4M}{k+4}\frac{k+4}{k+2}\left(\frac{k}{k+3}+\frac{1}{k+2}\right)\le\frac{4M}{k+4}.$$

5. Application: we consider $X=\mathbb{R}^d$ with norms either:

$$\|x\|=|x|_\infty=\max_{i=1,\dots,d}|x_i|\,,x\in X=\mathbb{R}^d \text{ and } \|p\|_*=|p|_1=\sum_{i=1}^d|p_i|\,,p\in X^*=\mathbb{R}^d,$$

((ℓ_1,ℓ_∞) case) or

$$\|x\|=|x|_2=\sqrt{\sum_{i=1}^d x_i^2}\,,x\in X=\mathbb{R}^d \text{ and } \|p\|_*=|p|_2\,,p\in X^*=\mathbb{R}^d$$

((ℓ_2,ℓ_2) or Euclidean case).

Let $f:\mathbb{R}^d\to[0,+\infty)$ be C^1, convex. We wish to solve:

$$\min_{|x|_\infty\le 1} f(x)$$

that is, minimize f on $C:=\{x\in\mathbb{R}^d:|x_i|\le 1\ \forall i=1,\dots,d\}=[-1,1]^d$.

We denote L_∞ the (best) Lipschitz constant of df on $[-1,1]^d$ in the (ℓ_1,ℓ_∞) duality, and L_2 the (best) Lipschitz constant in the Euclidean norms. Show first that

$$L_2\le L_\infty\le dL_2.$$

(In a first step, show the inequalities comparing $|\cdot|_1$ and $|\cdot|_2$, and comparing $|\cdot|_\infty$ and $|\cdot|_2$.)

First, we show that for all $x\in\mathbb{R}^d$, $|x|_\infty\le|x|_2\le\sqrt{d}|x|_\infty$, indeed:

$$|x|_\infty=\max_i|x_i|\le\sqrt{\sum_{i=1}^d x_i^2},\quad \sqrt{\sum_{i=1}^d x_i^2}\le\sqrt{\sum_{i=1}^d(\max_{i'}x_{i'}^2)}\le\sqrt{d}\max_{i'}|x_{i'}|.$$

By duality, one finds $|x|_2\le|x|_1\le\sqrt{d}|x|_2$. This also follows easily from the fact that $|x|_2=|\sum_i x_ie_i|_2\le\sum_i|x_ie_i|_2=\sum_i|x_i|$ where $(e_i)_{i=1}^d$ is the canonical basis of $\mathbb{R}^d$: $(e_i)_j=0$ if $j\ne i$, $(e_i)_i=1$. On the other hand, $|x|_1=\sum_{i=1}^d 1\times|x_i|\le\sqrt{\sum_{i=1}^d 1^2}\sqrt{\sum_{i=1}^d|x_i|^2}=\sqrt{d}|x|_2$ by Cauchy-Schwartz.

Hence if df is L_∞-Lipschitz in (ℓ_1,ℓ_∞), one has for any x,x':

$$|df(x)-df(x')|_2\le|df(x)-df(x')|_1\le L_\infty|x-x'|_\infty\le L_\infty|x-x'|_2$$

so that $L_2\le L_\infty$ while if df is L_2 Lipschitz for the Euclidean norm, one has for any x,x':

$$|df(x)-df(x')|_1\le\sqrt{d}|df(x)-df(x')|_2\le\sqrt{d}L_2|x-x'|_2\le dL_2|x-x'|_\infty,$$

hence $L_\infty\le dL_2$.

6. We want to compare the Frank-Wolfe algorithm and the Forward-Backward splitting algorithm (or projected gradient) for solving f. We recall that the FB consists here in performing one step of gradient descent $\tilde{x}_i^{k+1} = x_i^k - \tau\partial_i f(x^k)$, and then projecting $\tilde{x} = (\tilde{x}_i^{k+1})_{i=1}^d$ on the box $C = [-1,1]^d$ to obtain the new point $x^{k+1} = (x_i^{k+1})_{i=1}^d$. First, describe a step of each algorithm.

For Frank-Wolfe, one step consists in minimizing $\sum_{i=1}^d \partial_i f(x^k)x_i$ for $x \in C$: a solution is given by $s_i^{k+1} = 1$ if $\partial_i f(x^k) < 0$, $s_i^{k+1} = -1$ if > 0, any value $s_i^{k+1} \in [-1,1]$ if $= 0$. In the latter case a possible choice is $s_i^{k+1} = x_i^k$. Then, one has $x_i^{k+1} = x_i^k$ if $\partial_i f(x_i^k) = 0$, $x_i^{k+1} = (1-\omega_k)x_i^k \pm \omega_k$ if $\mp\partial_i f(x^k) > 0$.

For FB, as said one lets:

$$x_i^{k+1} = \min\{1, \max\{x_i^k - \tau\partial_i f(x^k), -1\}\}$$

for $\tau = 1/L_2$.

7. We recall that for the FB splitting with step $\tau = 1/L_2$, the rate is

$$f(x^k) - f(x^*) \le \frac{L_2|x^* - x^0|_2^2}{2k}.$$

What is the diameter D_2 of $C = [-1,1]^d$ in the $|\cdot|_2$ norm? What is the diameter D_∞ in the $|\cdot|_\infty$ norm? Write the rate of the Frank-Wolfe method for the norms (ℓ_1, ℓ_∞) and for the norms (ℓ_2, ℓ_2) (Euclidean).

For D_2, the diameter is $|(1,\dots,1)^T - (-1,\dots,-1)^T|_2 = 2\sqrt{d}$. For D_∞, the diameter is

$$\max_{|x_i|\le 1, |y_i|\le 1} \max_i |x_i - y_i| = 2.$$

Hence in the Euclidean norms, the rate of FW is:

$$f(x^k) - f(x^*) \le \frac{8dL_2}{k+3}$$

and in the $(\infty, 1)$ norms, it is:

$$f(x^k) - f(x^*) \le \frac{8L_\infty}{k+3}.$$

8. What algorithm seems better for this problem? (In general). How could we improve?

The rate for the FB is

$$f(x^k) - f(x^*) \le \frac{L_2|x^* - x^0|_2^2}{2k} \le \frac{dL_2}{2k}$$

if $x^0 = 0$. Since $L_2 \le L_\infty$, we have also $f(x^k) - f(x^*) \le \frac{dL_\infty}{2k}$. On the other hand, the rate for FW in the $(\infty, 1)$ norms is

$$f(x^k) - f(x^*) \le \frac{8L_\infty}{k+3}.$$

which is not (much) worse than the rate for the FB, since using $L_\infty \le dL_2$ it yields $f(x^k) - f(x^*) \le \frac{8dL_2}{k+3}$ which is 16 times worse. In case $L_\infty \approx dL_2$ the FB method is preferable, but in case L_∞ is much smaller, the FW is better.

However, using "FISTA" we could improve the convergence of the projected gradient to $f(x^k) - f(x^*) \le 2L_2|x^0 - x^*|^2/k^2 \le 2dL_2/k^2$, which is better (at least if d is not too large and one is ok with $k \gtrsim d$ iterations).

Exercise II - limits of convex functions

Let $C \subset \mathbb{R}^d$, $d \geq 1$, a convex set with nonempty interior and let $f_n(x) : C \to \mathbb{R}$ be convex functions such that for any $x \in C$, $\sup_{n\geq 1} |f_n(x)| < +\infty$.

1. Show that for any $x \in \mathring{C}$ (the interior of C), there exists $\delta > 0$ such that $\sup_{n\geq 1} \sup_{y\in B(x,\delta)} |f_n(y)| < +\infty$. [Hint: show that for $\delta > 0$, $B(x, \delta)$ is in the convex envelope of a finite set of points, and use the boundedness of f_n at each of these points, and at x.]

Let $x \in \mathring{C}$ and let $z_i^{\pm} = x \pm te_i$, $i \in 1, \dots, d$ where $(e_i)_i$ is a basis and $t > 0$. First, since x is in the interior of C, then for t small enough, $x \pm te_i \in C$ for all $i = 1, \dots, d$. We choose such a $t > 0$ and let $Z = \{z_1^-, \dots, z_d^-, z_1^+, \dots, z_d^+\}$).

Then, let $M \geq \max_{z\in Z} \sup_{n\geq 1} |f(z)|$ and $M \geq \sup_{n\geq 1} |f(x)|$. Then for any $y \in \mathrm{conv}(Z)$, that is with $y = \sum_{z\in Z} t_z z$ with $t_z \geq 0$, $\sum_{z\in Z} t_z = 1$, one has $f_n(y) \leq \sum_{z\in Z} t_z f_n(z) \leq M$.

On the other hand, letting $y' = 2x - y$ so that $x = (y + y')/2$, $f_n(x) \leq (f_n(y) + f_n(y'))/2$ so that $f_n(y) \geq 2f_n(x) - f_n(y') \geq -2M - M = -3M$. Hence, $\sup_n |f_n(y)| \leq 3M$. We choose δ so that $B(x, \delta) \subset \mathrm{conv}(Z)$ to conclude.

2. Deduce that for any compact set K with $K \subset \mathring{C}$ (the interior of C), then

$$\sup_{n\geq 1} \sup_{x\in K} |f_n(x)| < +\infty.$$

For each $z \in K$, there is a ball $B(z, \delta_z) \subset C$ where f_n is uniformly bounded. By compactness, there is a finite sequence $(z_i)_{i=1}^N$ such that

$$K \subset \bigcup_{i=1}^N B(z_i, \delta_{z_i}).$$

Then

$$\sup_{x\in K} \sup_{n\geq 1} |f_n(x)| \leq \max_{i=1}^N \sup_{x\in B(z_i,\delta_{z_i})} |f_n(x)| < +\infty.$$

3. Let $R' > R$ with $B(x, R') \subset C$. Show that there exists $L > 0$ such that for any $n \geq 1$, f_n is L-Lipschitz in $B(x, R)$.

Let R' as in the question and $R'' = (R + R')/2$. Then, by the previous question **2.**, there exists M such that $|f_n(y)| \leq M$ for all $y \in \overline{B(x, R'')}$ and all $n \geq 1$. Then, it is shown in the lecture notes that f_n is $L := 2M/(R'' - R)$ Lipschitz in $B(x, R)$.

4. Deduce that there exists a subsequence f_{n_k} with $f_{n_k} \to f$ uniformly in $\overline{B(x, R)}$. Show that f is convex. We recall the:

> **Theorem** (Ascoli-Arzelà) *Let K be compact in $\mathbb{R}^d$. If $f_n : K \to \mathbb{R}$ 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 $\forall \varepsilon > 0, \exists \eta > 0, \forall x, x' \in K, \|x-x'\| \leq \eta \Rightarrow (\forall n,\ |f_n(x)-f_n(x')| \leq \varepsilon)$.

This is an obvious application of the theorem. In addition, if $x, y \in C$ and $t \in [0, 1]$, $f_n(tx + (1 - t)y) \leq tf_n(x) + (1 - t)f_n(y)$ for all n and passing to the limit we find that f is also convex.

5. (Difficult.) Deduce that there exists f convex and a subsequence f_{n_k} such that $f_{n_k} \to f$ uniformly on any compact set $K \subset \mathring{C}$. [Hint: consider a countable family $B(x_i, \delta_i) \subset C$, $i \geq 1$, such that f_n are uniformly bounded and Lipschitz in all balls $B(x_i, \delta_i)$ and $(x_i)_{i\geq 1}$ is dense in C. We choose $\delta_i = \mathrm{dist}(x_i, \mathbb{R}^d \setminus C)/2$ so that in particular, $\bigcup_{i\geq 1} B(x_i, \delta_i) = \mathring{C}$.]

We extract a first subsequence and find f^1 such that $f_{n_k^1} \to f^1$ uniformly in $\overline{B(x_1, \delta_1)}$. Given the subsequence n_k^i, we find f^{i+1} and a subsequence n_k^{i+1} extracted from n_k^i such that $f_{n_k^{i+1}} \to f^{i+1}$ uniformly in $B(x_{i+1}, \delta_{i+1})$. We first observe that if $B(x_i, \delta_i) \cap B(x_j, \delta_j) \neq \emptyset$ and $j > i$, then since $f_{n_k^j}$ is a subsequence of $f_{n_k^i}$, $f^j = f^i$ in the intersection. Therefore we can drop the superscript and call f the limit function which is defined in all the balls $B(x_i, \delta_i)$, and therefore in $\mathring{C}$.

Then, we let $f_{n_k} = f_{n_k^k}$ which is, for $k \geq i$, a subsequence of $f_{n_k^i}$, so that it converges to f uniformly in all $\overline{B(x_i, \delta_i)}$. (This is called a "diagonal" subsequence.)

If $K \subset \mathring{C}$ is a compact set, then by construction it is covered by finitely many balls $B(x_i, \delta_i)$, so that $f_{n_k} \to f$ uniformly on K.

We assume, from now on, that there exists f convex such that $f_n \to f$ uniformly on any compact $K \subset \mathring{C}$.

6. Show that if $x \in \mathring{C}$, for any $x_n \to x$, $f_n(x_n) \to f(x)$.

First, $f_n \to f$ uniformly on a ball $\overline{B(x, \delta)}$ where all f_n, and in particular also f, are L-Lipschitz with some constant f. It follows

$$|f_n(x_n) - f(x)| \leq |f_n(x_n) - f(x_n)| + L|x_n - x| \to 0$$

as $n \to \infty$. (We use that f is Lipschitz, but continuity would be sufficient.)

7. Let $x_n \to x$, $x_n \in C$, $x \in \mathring{C}$, and let $p_n \in \partial f_n(x_n)$ for all n. Show that p_n is bounded, and that any limit p of a subsequence is such that $p \in \partial f(x)$. What happens if f is of class C^1?

Let $B(x, \delta)$ a ball where all f_n are bounded and uniformly L-Lipschitz. Then $x_n \in B(x, \delta)$ for n large enough. One has $L|z - x_n| \geq f_n(z) - f_n(x_n) \geq \langle p_n, z - x_n \rangle$ for all $z \in C$, which shows that $|p_n| \leq L$. Hence p_n is bounded and has subsequences which converge. Assume that $p_{n_k} \to p$. Passing in the limit in the subgradient inequality and using **6.** we get $f(z) - f(x) \geq \langle p, z - x \rangle$ which shows that $p \in \partial f(x)$.

If f is C^1 then $\partial f(x) = \{\nabla f(x)\}$ and we find that $p_n \to \nabla f(x)$, since all converging subsequences have the same limit.

8. Assume that f_n, f are all C^1. Deduce that $\nabla f_n \to \nabla f$ uniformly on compact sets $K \subset \mathring{C}$. [Hint: assume this is false.]

If not there is K compact such that there exist $\varepsilon > 0$, for any N, there is $n \geq N$ and $x_n \in K$ with $|\nabla f_n(x_n) - \nabla f(x)| \geq \varepsilon$.

Since K is compact, without loss of generality we may assume that x_n converges to a point $x \in K$. Then, by the previous question, we have $\nabla f_n(x_n) \to \nabla f(x)$. On the other hand, $|\nabla f_n(x_n) - \nabla f(x)| \geq \varepsilon$ for all $n \geq 1$. This is a contradiction.

Exercice III - conjugates

Find the convex (Legendre-Fenchel) conjugate $f^*(y) = \sup_x \langle x, y \rangle - f(x)$ in the following cases:

1. $f(x) = \frac{1}{2}(|x|_\infty)^2$, where $x \in \mathbb{R}^d$ and $|x|_\infty = \max_{i=1}^d |x_i|$.

$$\sup_x \langle x, y \rangle - \frac{1}{2}|x|_\infty^2 = \sup_{t>0} \max_{|x_i|\leq t} \sum_{i=1}^d x_i y_i - \frac{t^2}{2} = \sup_{t>0} \sum_{i=1}^d t|y_i| - \frac{t^2}{2} = \frac{1}{2}\left(\sum_{i=1}^d |y_i|\right)^2 = \frac{|y|_1^2}{2}.$$

2.a. $f(x) = -\ln(x)$ for $x > 0$ and $+\infty$ if $x \le 0$.

We have $f^*(t) = \sup_{x>0} tx + \ln(x)$: if there exists a maximum then $t + 1/x = 0$, hence $x = -1/t$. If $t \ge 0$, obviously $tx + \ln(x) \to +\infty$ as $x \to +\infty$ and $f^*(t) = +\infty$. If $t < 0$ then the maximum is reached at $x = -1/t$ and one has $f^*(t) = -1 + \ln(-1/t) = -1 - \ln(-t)$.

2.b. Find a function f such that $f^*(y) = f(-y)$ for all $y \in \mathbb{R}$.

Let $f_a(x) = -\ln(x) - a$ for $a \in \mathbb{R}$ ($+\infty$ if $x \le 0$). Then by **2.a.**, $f_a^*(y) = a - 1 - \ln(-y)$ for $y < 0$. To have $f_a^*(y) = f_a(-y)$ we just need to have $-a = a - 1$, that is, $a = 1/2$.

3 $f(x) = x \ln x - x$ for $x \ge 0$ (and $0 \ln 0 = 0$), and $+\infty$ for $x < 0$.

$f'(x) = \ln(x)$, hence $(f')^{-1}(y) = \exp(y)$, and this should be $(f^*)'$, which shows that $f^*(y) = \exp(y)$ (plus a constant, but $f^*(0) = -\min_x f$ is reached for $\ln x = 0$, that is $x = 1$, and the value is 1).

Exercice IV - perspective function

Let f be convex, proper, lower semi-continuous on a Hilbert space X, with conjugate f^*. We assume, for simplicity, that

$$\boxed{f(0) = 0.}$$

We define, for $(x, t) \in X \times \mathbb{R}$,

$$h(x, t) = \sup_{s,y:s+f^*(y)\le 0} \langle x, y\rangle + st.$$

("The support function of the symmetric of the epigraph of f with respect to the horizontal axis.")

1. Show that if $t < 0$, $h(x, t) = +\infty$.

Indeed, s is unbounded from below and sending $s \to -\infty$ one has $h = +\infty$.

2. Show that if $t > 0$, $h(x, t) = tf(x/t)$.

If $t > 0$ then

$$\sup_{s,y:s+f^*(y)\le 0} \langle x, y\rangle + st = \sup_{s,y:f^*(y)\le s} \langle x, y\rangle - st$$
$$= \sup_{y:f^*(y)<+\infty} \langle x, y\rangle - tf^*(y) = t \sup_y \langle x/t, y\rangle - f^*(y) = tf(x/t).$$

3. If $t = 0$ check that $h(x, 0) = \sup_{f^*(y)<+\infty} \langle x, y\rangle$. We want to show that this is $\lim_{t\downarrow 0} tf(x/t)$.
First, show that if $0 < t < s$, $sf(x/s) \le tf(x/t)$, so that $\lim_{t\downarrow 0} tf(x/t) = \sup_{t>0} tf(x/t)$ exists in $\mathbb{R} \cup \{+\infty\}$.

First, $h(x, 0) = \sup_{(y,s):s+f^*(y)\le 0} \langle x, y\rangle$. If $f^*(y) = +\infty$ there is no s such that $s + f^*(y) \le 0$, while if $f^*(y) < +\infty$, it is enough to take $s = -f^*(y)$.

Then one has, using convexity, $sf(x/s) = sf((t/s)(x/t) + (1 - t/s)0) \le s(t/s)f(x/t) + (1 - t/s)f(0) = tf(x/t)$.

4. Why do we have $f^* \ge 0$? Deduce that $tf(x/t) \le h(x, 0)$ for any $t > 0$.

One has $f^*(y) = \sup_x \langle x, y\rangle - f(x) \ge \langle 0, y\rangle - f(0) = 0$. Then, for $t > 0$,

$$tf(x/t) = t \sup_y \langle x/t, y\rangle - f^*(y) \le \sup_{y:f^*(y)<+\infty} \langle x, y\rangle - tf^*(y) \le \sup_{y:f^*(y)<+\infty} \langle x, y\rangle = h(x, 0).$$

5. Show eventually that $h(x,0) = \sup_{f^*(y)<+\infty} \langle x, y\rangle \leq \lim_{t\downarrow 0} tf(x/t)$ and conclude. The function $f^\infty(x) = \lim_{t\downarrow 0} tf(x/t)$ is called the "recession function of f".

If $f^*(y) < +\infty$, then

$$\langle x, y\rangle = \lim_{t\to 0} \langle x, y\rangle - tf^*(y) = \lim_{t\to 0} t\left(\langle x/t, y\rangle - f^*(y)\right) \leq \lim_{t\to 0} tf(x/t).$$

Hence, $h(x,0) \leq \lim_{t\to 0} tf(x/t)$. By **3.** we deduce it is equal.

6. What can we say of the function (called the "perspective function of f"), defined on $X \times \mathbb{R}$:

$$(x,t) \mapsto \begin{cases} tf\left(\frac{x}{t}\right) & \text{if } t > 0\,, \\ +\infty & \text{if } t < 0\,, \\ 0 & \text{if } (x,t) = (0,0)\,, \\ f^\infty(x) & \text{if } t = 0 \end{cases} \qquad ?$$

It is a convex, lower-semicontinuous and positively one-homogeneous function, as the support function of a set (or conjugate of a "characteristic function").