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 $x^* \in C$.

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

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

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

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

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

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 $k \geq 1$,

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

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

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

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

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

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

$$L_2 \le L_\infty \le d L_2.$$

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

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.

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).

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

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.]

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

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

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 $B(x, R)$.

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 $\forall \varepsilon > 0, \exists \eta > 0, \forall x, x' \in K, \|x-x'\| \leq \eta \Rightarrow (\forall n,\ |f_n(x)-f_n(x')| \leq \varepsilon)$.

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 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 $x_n \to x$, $f_n(x_n) \to f(x)$.

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?

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.]

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 $|x|_\infty = \max_{i=1}^d |x_i|$.

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

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

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

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

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

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

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

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

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 $0 < t < s$, $sf(x/s) \leq tf(x/t)$, so that $\lim_{t\downarrow 0} tf(x/t) = \sup_{t>0} tf(x/t)$ exists in $\mathbb{R}\cup\{+\infty\}$.

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

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".

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

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