Continuous Optimization
Introduction à l'optimisation continue
Assessment
(4th January 2021)

1. Convex analysis: Exercise

1. Evaluate the convex conjugate (Legendre-Fenchel conjugate) of the functions:

1. $x \mapsto |x|^3/3$;
2. $x \mapsto 3x$;
3. $x \mapsto \langle Ax, x \rangle /2$ where $x \in \mathbb{R}^n$ and A is a symmetric, positive definite operator;
4. $x \mapsto -\sqrt{x}$ if $x \geq 0$, $+\infty$ if $x < 0$.

2. Evaluate $y = \mathrm{prox}_{\tau f}(x)$ for $f(x) = |x|^3/3$, $\tau > 0$.

3. (More difficult) Evaluate the convex conjugate of the "Entropy" function:

$$S \ : \ \mathbb{R}^n \to [0, +\infty] \ ; \qquad x \mapsto \begin{cases} \sum_{i=1}^n x_i \ln x_i & \text{if } x_i \geq 0 \ \forall i, \sum_i x_i = 1 \,, \\ +\infty & \text{else,} \end{cases}$$

where here by convention we let $t \ln t = 0$ when $t = 0$. (Hint: introduce a Lagrange multiplier for the constraint $\sum_i x_i = 1$.)

2. Convex analysis: Moreau-Yosida regularization

Given $f : \mathbb{R}^n \to]-\infty, +\infty]$ a convex, lower-semicontinuous function, which is proper (that is, $f > -\infty$ and $\mathrm{dom}\, f \neq \emptyset$), we recall that the Moreau-Yosida regularization of f with parameter $\tau > 0$ is given by:

$$f_\tau(x) = \min_y f(y) + \frac{1}{2\tau} \|y - x\|^2$$

We recall that for any x, this problem has a unique minimizer y (because the function to minimize is strongly convex, lower-semicontinuous) and that the minimizer is also known as $y = \mathrm{prox}_{\tau f}(x)$, the "proximity operator" of τf evaluated at x. In particular, $f_\tau(x) \in \mathbb{R}$ and $\mathrm{dom}\, f_\tau = \mathbb{R}^n$. Further properties of the $\mathrm{prox}_{\tau f}$ operator are described in the lecture notes.

1. Show (by giving a proof or invoking the appropriate result in the lecture notes) that f_τ is convex, lower-semicontinuous.

2. Let $x \in \mathbb{R}^n$ and $p \in \partial f_\tau(x)$. Show that for any $h \in \mathbb{R}^n$,

$$p \cdot h \leq \left(\frac{x - \mathrm{prox}_{\tau f}(x)}{\tau} \right) \cdot h.$$

(Hint: bound from below and above $f_\tau(x + th)$, for $t > 0$ small, then send t to zero.) Deduce that f_τ is differentiable at x, with $\nabla_\tau f(x) = (x - \mathrm{prox}_{\tau f}(x))/\tau$.

3. Recall why $\mathrm{prox}_{\tau f}$ is "firmly non-expansive". Deduce that ∇f_τ is $(1/\tau)$-Lipschitz.

4. Recall "Moreau's" identity. Deduce that $\nabla f_\tau(x) = \mathrm{prox}_{\frac{1}{\tau} f^*}(x/\tau)$ where f^* is the convex conjugate (Legendre-Fenchel transform) of f.

In what follows, ****to simplify**** we let $\tau = 1$.

5. Deduce from the previous results that for any x,

$$\nabla f_1(x) + \nabla (f^*)_1(x) = x$$

(here $(f^*)_1$ is the Moreau-Yosida regularization with parameter $\tau = 1$ of the conjugate f^* of f, and *not!!* the conjugate of f_1).

6. Therefore by integration one finds: $f_1(x) + (f^*)_1(x) = \|x\|^2/2 + C$ for some constant C, with $C = f_1(0) + (f^*)_1(0)$. Let $y = \mathrm{prox}_f(0)$, $z = \mathrm{prox}_{f^*}(0)$. Show that $y = -z$. Deduce that $C = 0$.

3. Optimization: Nonlinear gradient descent

Let $\|\cdot\|$ be a norm on $\mathbb{R}^n$, possibly different from the standard Euclidean 2-norm: for instance, $\|x\| = \sum_{i=1}^n |x_i|$ (the 1-norm), or $\|x\| = \max\{|x_1|, \dots, |x_n|\}$ (the ∞-norm). (A norm is any convex, 1-homogeneous, even, function with values in $[0, +\infty[$ and which is strictly positive except in 0.) We define the dual (or polar) norm $\|y\|_*$ by the formula:

$$\|y\|_* = \sup_{x : \|x\| \leq 1} y \cdot x$$

where $y \cdot x$ is the standard dot product $y \cdot x = \sum_{i=1}^n y_i x_i$. In particular, one has $y \cdot x \leq \|y\|_* \|x\|$ for all y, x. (The "right" point of view should be that y is in the dual E^* of $E = \mathbb{R}^n$ (which is also $E^* = \mathbb{R}^n$) and that $y \cdot x$ is the evaluation of the linear form y at x. Then, $\|\cdot\|$ is the norm on E while $\|\cdot\|_*$ is the norm on E^*.)

1. Show that if $\mathcal{F}(x) := \|x\|^2/2$, then its convex conjugate is $\mathcal{F}^*(y) = \|y\|_*^2/2$. Deduce that the dual norm of $\|\cdot\|_*$ is $\|\cdot\|$, that is, for all x,

$$\|x\| = \sup_{y : \|y\|_* \leq 1} y \cdot x.$$

2. Compute $\|\cdot\|_*$ in the following cases:

i. 1-norm: $\|x\| = \sum_{i=1}^n |x_i|$;

ii. 2-norm: $\|x\| = \left(\sum_{i=1}^n |x_i|^2\right)^{\frac{1}{2}} = \sqrt{x \cdot x}$.

Now, we consider a function f whose differential is L-Lipschitz in the normed space $E = (\mathbb{R}^n, \|\cdot\|)$, which means precisely that for any $x, x' \in \mathbb{R}^n$,

$$\|\nabla f(x) - \nabla f(x')\|_* \leq L\|x - x'\|$$

where $\nabla f(x) \in E^*$ is the vector of partial derivatives $(\partial f/\partial x_i)_{i=1}^n$.

3. Show that, as in the Euclidean case, one has for $x, x' \in E$,

$$f(x') \leq f(x) + \nabla f(x) \cdot (x' - x) + \frac{L}{2}\|x - x'\|^2.$$

We want to define a "gradient descent" method in the norms $\|\cdot\|, \|\cdot\|_*$. We choose $x^0 \in E$. Given x^k, $k \geq 0$, we define $x^{k+1} = x^k - p^k$ and we find the descent direction p^k as follows: we observe that

$$f(x^{k+1}) \leq f(x^k) - \nabla f(x^k) \cdot p^k + \frac{L}{2}\|p^k\|^2.$$

Then, we choose a p^k which minimizes the expression in the right-hand side of this equation.

4. Show that one has to choose $p^k \in \partial\mathcal{F}^*(\frac{1}{L}\nabla f^*)$, and that one obtains, for such a choice:

$$f(x^{k+1}) \leq f(x^k) - \frac{1}{2L}\|\nabla f(x^k)\|_*^2.$$

5. We assume the set $X = \{f \leq f(x^0)\}$ is bounded, and observe that $x^k \in X$ for any $k \geq 1$. We also assume that f has a minimizer x^* (obviously, $x^* \in X$). As in the Lecture notes, show that for all $k \geq 0$:

$$f(x^{k+1}) - f(x^*) \leq f(x^k) - f(x^*) - \frac{(f(x^k) - f(x^*))^2}{2L\|x^k - x^*\|^2},$$

and:

$$f(x^k) - f(x^*) \leq \frac{2LC}{k+1}$$

where $C = \max_{x \in X} \|x - x^*\|^2$.

4. Optimization: Polyak's subgradient descent method

In his book from 1987, Boris T. Polyak suggests the following variant of the subgradient descent method, which can be used whenever the optimal value of a problem is known. One consider a convex function $f : \mathbb{R}^n \to \mathbb{R}$ ($\mathrm{dom}\, f = \mathbb{R}^n$), which has a non empty set of minimizer(s) X^*, and we assume that the minimal value f^* is known. For instance:

$$f(x) = \max_{1 \le i \le p} |a^i \cdot x - b_i|$$

where $a^i \in \mathbb{R}^n, b \in \mathbb{R}^p$ are such that $a^i \cdot x = b_i$, $i = 1, \dots, p$ has a solution: in that case $f^* = 0$.

Then, one chooses $x^0 \in \mathbb{R}^n$ and computes a subgradient descent method by picking for all $k \ge 0$, $p^k \in \partial f(x^k)$ and

$$x^{k+1} = x^k - \frac{f(x^k) - f^*}{\|p^k\|^2} p^k.$$

1. Show that, if $x^* \in X^*$ is any minimizer,

$$\|x^{k+1} - x^*\|^2 \le \|x^k - x^*\|^2 - \frac{(f(x^k) - f^*)^2}{\|p^k\|^2}.$$

What do we deduce for the sequence $(x^k)_{k \ge 0}$?

2. Why is it true that $C := \sup_k \|p^k\| < +\infty$? Deduce that $\displaystyle\sum_{k=0}^{\infty} (f(x^k) - f^*)^2 < +\infty$.

3. We deduce that $f(x^k) \to f^*$. Show that there is one minimizer $x^* \in X^*$, such that $x^k \to x^*$.

4. We now assume that the function is "α-sharp", $\alpha \ge 1$, meaning that for some $\gamma > 0$,

$$f(x) - f^* \ge \gamma \mathrm{dist}\,(x, X^*)^\alpha.$$

Show that

$$\mathrm{dist}\,(x^{k+1}, X^*)^2 \le \mathrm{dist}\,(x^k, X^*)^2 - \frac{\gamma^2 \mathrm{dist}\,(x^k, X^*)^{2\alpha}}{C^2}.$$

In case $\alpha = 1$ (which is the situation in the example mentioned in the introduction of this exercise), what do we deduce?

5. We consider a sequence a_k, $k \ge 0$, with for all $k \ge 0$, $a_k \ge 0$ and $a_{k+1} \le a_k - c^{-1} a_k^{1+\beta}$, $c > 0$, $\beta > 0$. Show that:

$$a_k \le \left(\frac{c}{\max\{\beta, 1\}(k+1)} \right)^{1/\beta}.$$

Hint: introduce $b_k := a_k^\beta$, and depending on whether $\beta \ge 1$ or $\beta \le 1$, try to show that $b_k \le b_k - c'^{-1} b_k^2$ for some c' (depending on c, β). Use then the Lecture notes.

6. Deduce the rate of convergence for the distance from x^k to the set X^* in case $\alpha > 1$.