Exam M2 Datascience 2019 : Convex analysis and advanced optimization

Course book and written notes are authorized. All electronic devices are prohibited.
Questions can be answered either in French or in English. Duration : 3h00

The sets $\mathcal{X}$ and $\mathcal{Y}$ respectively represent the Euclidean spaces $\mathbb{R}^d$ and $\mathbb{R}^m$ respectively, where d, m are positive integers. The set $\Gamma_0(\mathcal{X})$ represents the set of proper, lower semi-continuous and convex functions on $\mathcal{X} \to (-\infty, +\infty]$. If $t \in \mathbb{R}$, then $\lfloor t \rfloor$ is the largest integer less or equal to t and $\lceil t \rceil$ is the smallest integer greater or equal to t.

If $\|\,.\,\|$ is an arbitrary norm (not necessarily the Euclidean norm), the corresponding *dual norm* is the norm

$$\|y\|_* = \sup_{x \in \mathcal{X} : \|x\| \le 1} \langle x, y \rangle \,.$$

In particular, for every x, y, it holds that $\langle x, y \rangle \le \|x\| \, \|y\|_*$.

Exercise 1. Let C be a non-empty compact convex subset of $\mathcal{X}$. A *regularizer* on C is a mapping $h : \mathcal{X} \to (-\infty, +\infty]$ such that $\mathrm{dom}(h) = C$ and the restriction of h to C is strictly convex and continuous. The aim of the exercice is to show the following theorem :

> **Theorem 1.** If h is a regularizer on C, then h^* is differentiable and for every $y \in \mathcal{X}$, $\nabla h^*(y) = \arg\max_{x \in C}(\langle x, y \rangle - h(x))$.

1. Prove that for every $y \in \mathcal{X}$, there exists a unique point $x_y \in C$ such that

$$h^*(y) = \langle x_y, y \rangle - h(x_y) \,.$$

2. Show that $\partial h^*(y) = \{x_y\}$.

Exercise 2. Let h be a regularizer on a nonempty compact convex set $C \subset \mathcal{X}$ (see previous exercise). We set $h_{\min} = \inf\{h(x) : x \in C\}$ and $h_{\max} = \sup\{h(x) : x \in C\}$. Consider an arbitrary locally integrable function $t \mapsto u(t)$ defined on $\mathbb{R}_+ \to \mathcal{X}$, and a positive, nonincreasing, piecewise continuous function $t \mapsto \eta(t)$. Consider the function defined for every $t \ge 0$ by :

$$x(t) = \nabla h^* \left(\eta(t) \int_0^t u(s) ds \right) \,.$$

The aim of this exercise is to prove the following Lemma.

> **Lemma 1.** For every $x \in C$,
>
> $$\int_0^t \langle u(s), x \rangle ds - \int_0^t \langle u(s), x(s) \rangle ds \le \frac{h_{\max} - h_{\min}}{\eta(t)} \,.$$

In the following, we simplify the problem by assuming that $t \mapsto \eta(t)$ is **continuously differentiable**, although this assumption is in fact not needed.

1. We set $y(t) = \eta(t) \int_0^t u(s)ds$. Show that for every $x \in C$, one has :
$$h^*(y(t)) + h(x) \geq \eta_t \int_0^t \langle u(s), x \rangle ds\,.$$
2. Show that $h^*(y(t)) + h(x(t)) = \eta_t \int_0^t \langle u(s), x(t) \rangle ds$.
3. Using the chain rule of derivation, prove that :
$$\frac{d}{dt}\frac{h^*(y(t))}{\eta(t)} = \langle u(t), x(t) \rangle + \frac{\eta'(t)}{\eta(t)^2} h(x(t))\,,$$
where $\eta'(t)$ is the derivative of $\eta(t)$.
4. Prove that :
$$\frac{h^*(y(t))}{\eta(t)} - \frac{h^*(0)}{\eta_0} \leq \int_0^t \langle u(s), x(s) \rangle ds - \frac{h_{\min}}{\eta(t)} + \frac{h_{\min}}{\eta_0}\,.$$
5. Deduce that :
$$\frac{h^*(y(t))}{\eta(t)} \leq \int_0^t \langle u(s), x(s) \rangle ds - \frac{h_{\min}}{\eta(t)}\,.$$
6. Prove Lemma 1.

Exercise 3. Let h be a regularizer on the non-empty compact convex set C. We assume that ∇h^* is $\frac{1}{\mu}$-Lipschitz continuous w.r.t. some arbitrary norm $\|\,.\,\|$, which reads :

$$\forall x, y \in \mathcal{X}, \quad \|\nabla h^*(x) - \nabla h^*(y)\| \leq \frac{1}{\mu}\|x - y\|_*$$

where $\|\,.\,\|_*$ is the dual norm. An algorithm called *Follow the Regularized Leader* (FtRL) works as follows. At every iteration $n \in \mathbb{N}^*$, the *environment* generates a vector $u_n \in \mathcal{X}$. Observing the past sequence of observations $u_1, \ldots, u_n$, the *agent* generates an action $x_{n+1} \in C$ given the following rule :

$$\begin{aligned} U_n &= U_{n-1} + u_n \\ x_{n+1} &= \nabla h^*(\eta_n U_n)\,, \end{aligned}$$

where $(\eta_n)_{n \in \mathbb{N}^*}$ is a sequence of positive and non-increasing step sizes. By definition, the *reward* obtained by the agent at time n is $\langle u_n, x_n \rangle$. Hence the cumulated reward up to time n is $\sum_{k=1}^n \langle u_k, x_k \rangle$. We compare this to a fixed strategy : if an agent had played the same fixed action $x \in C$ at each iteration n, the cumulated reward would have been $\sum_{k=1}^n \langle u_k, x \rangle$. The *regret* of the algorithm with respect to $x \in C$ is defined by :

$$\mathrm{Reg}_n(x) = \sum_{k=1}^n \langle u_k, x \rangle - \sum_{k=1}^n \langle u_k, x_k \rangle\,.$$

The aim is to bound the regret uniformly in x. To that end, we define the following piecewise constant mappings on $\mathbb{R}_+$ by :

$$\bar{u}(t) = u_{\lceil t \rceil}$$
$$\bar{\eta}(t) = \eta_{\lfloor t \rfloor}$$

where we set $\eta_0 := \eta_1$. We define $y(t) = \bar{\eta}(t) \int_0^t \bar{u}(s)ds$ and finally, $\bar{x}(t) = \nabla h^*(y(t))$.

1. What does $\sup_{x \in C} \operatorname{Reg}_n(x)$ represent, and why are we trying to upper-bound this quantity ? (Short answer in 1 or 2 lines)
2. Prove that for every $k \in \mathbb{N}^*$ and every $t \in (k-1, k)$,

$$|\langle \bar{u}(t), \bar{x}(t)\rangle - \langle u_k, x_k\rangle| \leq \frac{1}{\mu}\|u_k\|_* \|y(t) - y_{k-1}\|_* .$$

   Hint : Note that for every $k \geq 1$, $x_k = \bar{x}(k-1)$.
3. Prove that for every $k \in \mathbb{N}^*$ and every $t \in (k-1, k)$,

$$\|y(t) - y_{k-1}\|_* \leq \eta_{k-1}\|u_k\|_*(t - k + 1) .$$

4. Prove that for every $n \in \mathbb{N}^*$,

$$\left| \int_0^n \langle \bar{u}(t), \bar{x}(t)\rangle dt - \sum_{k=1}^n \langle u_k, x_k\rangle \right| \leq \sum_{k=1}^n \beta_k \|u_k\|_*^2 ,$$

   where β_k is a constant to be expressed as a function of μ and η_{k-1}.
5. Using Lemma 1, find a relevant bound $\mathcal{B}_n$ depending on μ and the sequences $(\eta_1, \ldots, \eta_n)$, $(u_1, \ldots, u_n)$, $h_{\min}$, $h_{\max}$, such that $\operatorname{Reg}_n(x) \leq \mathcal{B}_n$ for every $x \in C$ and every $n \in \mathbb{N}^*$.
6. Assume that $\|u_k\|_* \leq 1$ for every k. Assume that the agent uses a constant step $\eta_k = \eta$ for all k. If the number n of iteration is known in advance, what value of η would you suggest to choose, in order to ensure a low regret after n iterations ? Express η as a function of n, $h_{\min}$, $h_{\max}$ and μ.

Exercise 4. Let $f : \mathbb{R} \to (-\infty, +\infty]$ be the function defined by

$$f(x) = \begin{cases} x \ln x & \text{if } x > 0 \\ 0 & \text{if } x = 0 \\ +\infty & \text{otherwise.} \end{cases}$$

1. Prove that $f \in \Gamma_0(\mathbb{R})$.
2. Compute $\partial f(x)$ for every $x \in \mathbb{R}$.

Exercise 5. Define $\Delta = \left\{x \in \mathcal{X} : \sum_{i=1}^d x_i = 1 \text{ and } \forall i, x_i \geq 0\right\}$ and set

$$h(x) = \begin{cases} \sum_{i=1}^d x_i \ln(x_i) & \text{if } x \in \Delta \\ +\infty & \text{otherwise}\,, \end{cases}$$

with the convention that $(0 \times \ln 0) = 0$. For a fixed $y \in \mathcal{X}$, we consider the constrained optimization problem

$$\min_{x \in \Delta} h(x) - \langle x, y \rangle$$

1. How equality constraints has this problem ? How many inequality constraints ?
2. Write the Lagrangian function.
3. By solving the KKT conditions, provide the primal solution.
4. Using Theorem 1, compute ∇h^* and provide the explicit expression of the FtRL algorithm.

Exercise 6. Let C be a non-empty compact convex set of $\mathcal{X}$. We define

$$h(x) = \begin{cases} \|x\|^2/2 & \text{if } x \in C \\ +\infty & \text{otherwise,} \end{cases}$$

where $\|\,.\,\|$ is the Euclidean norm. Let $(f_n)_{n \in \mathbb{N}^*}$ be a sequence of convex functions on $\mathcal{X} \to \mathbb{R}$. We consider the sequence (x_n) generated by the FtRL algorithm. We assume that for each $n \in \mathbb{N}^*$,

$$u_n \in \partial f_n(x_n)\,.$$

1. Using Theorem 1, compute ∇h^* and provide the explicit expression of the sequence (x_n) generated by the FtRL algorithm.
2. Find which of the following inequalities holds true for every $x \in C$:

$$\mathrm{Reg}_n(x) \leq \sum_{k=1}^n f_k(x_k) - \sum_{k=1}^n f_k(x)$$
$$\text{or } \mathrm{Reg}_n(x) \geq \sum_{k=1}^n f_k(x_k) - \sum_{k=1}^n f_k(x)\,.$$

Exercise 7. We consider the elastic net regression problem defined as

$$\min_{x \in \mathcal{X}} \frac{1}{2} \|Ax - b\|_2^2 + \lambda_1 \|x\|_1 + \frac{\lambda_2}{2} \|x\|_2^2 \tag{7.1}$$

where $A : \mathcal{X} \to \mathcal{Y}$ is a linear operator, $b \in \mathcal{Y}$, $\lambda_1 > 0$ and $\lambda_2 > 0$.

1. Define $g : x \mapsto \lambda_1 \|x\|_1 + \frac{\lambda_2}{2} \|x\|_2^2$. Show that g is convex and compute its Fenchel conjugate.
2. Give a dual problem to the elastic net regression problem (7.1).
3. Propose an algorithm for the resolution of (7.1). What are the advantages and drawback of your solution ?