Labs
Optimization for Machine Learning
Spring 2025

EPFL
School of Computer and Communication Sciences
Nicolas Flammarion
github.com/epfml/OptML_course

Problem Set 10, May 16, 2025 (Lower Bounds & Convex conjugate)

1 Lower Bounds for a Non-smooth Function

In the lecture, we have seen the lower bounds for a smooth and covex function. In this exercise, we will show a similar lowerbound for a non-smooth function.

1. Consider the function $f : \mathbb{R}^d \to \mathbb{R}$ given by
$$f(\mathbf{x}) = \gamma \max_{1\le i\le t} \{\mathbf{x}_i\} + \frac{\alpha}{2}\|\mathbf{x}\|^2.$$
Is the function f (strongly-)convex? Is f smooth? Show that the sub-differentiable of f at $\mathbf{x}$ is given by
$$\partial f(\mathbf{x}) = \alpha\mathbf{x} + \gamma \operatorname{conv}\left\{e_i \,:\, i \in \operatorname*{argmax}_{1\le j\le t} \mathbf{x}_j\right\}.$$

2. For the function f defined in the previous question, for any sub-gradient-based algorithm, i.e., $\mathbf{x}_s \in \mathbf{x}_0 + \mathrm{Span}\{\partial f(\mathbf{x}_0), \partial f(\mathbf{x}_1), \dots, \partial f(\mathbf{x}_{s-1})\}$ initialized at $\mathbf{x}_0 = 0$. Show that
$$\mathrm{Span}\{\partial f(\mathbf{x}_0), \partial f(\mathbf{x}_1), \dots, \partial f(\mathbf{x}_{s-1})\} = \mathrm{Span}\{e_1, e_2, \dots, e_s\}$$
where $\mathbf{x}_s$ is the point after s iterations of the algorithm. For the sub-gradient, assume that the gradient oracle returns $\partial f(\mathbf{x}) = \alpha\mathbf{x} + \gamma e_i$ where i is the smallest index such that $\mathbf{x}_i = \max_{1\le j\le t} \mathbf{x}_j$.

3. For any $R \ge 0$, consider the set $B(R) = \{\mathbf{x} \in \mathbb{R}^d : \|\mathbf{x}\|_2 \le R\}$. Show that for any $\mathbf{x} \in B(R)$, f is Lipschitz.

4. For the point y defined by
$$y[i] = \begin{cases} -\frac{\gamma}{\alpha t} & \text{if } 0 \le i \le t \\ 0 & \text{otherwise.} \end{cases},$$
show that $0 \in \partial f(y)$.

5. Using the point y and the previous properties, show that for $R > 0$, there exists a $L-$Lipschitz function f, such that for iterates $\{x_i\}$'s given by a gradient-based algorithm initialized at $\mathbf{x}_0 = 0$, the following holds:
$$\min_{1\le x\le s} f(x_s) - \min_{\|x\|\le R} f(x) \ge \frac{RL}{2(\sqrt{t}+1)},$$
for $s < t$.
Hint: Choose α and γ appropriately for R and the Lipschitz constant.

2 Convex Conjugate

For a function $f : \mathbb{R}^d \to \mathbb{R} \cup \{+\infty\}$ (which is not necessarily convex !), we consider its **convex conjugate** which for $y \in \mathbb{R}^d$ is defined as
$$f^*(y) = \sup_{x\in\mathbb{R}^d} (\langle x, y\rangle - f(x)) \in \mathbb{R} \cup \{+\infty\}$$
Prove the following properties.

1. Show that f^* is convex.
2. Show that for $x, y \in \mathbb{R}^d$, $f(x) + f^*(y) \geq \langle x, y \rangle$. This is known as Fenchel's inequality.
3. Show that the biconjugate f^{**} (the conjugate of the conjugate) is such that $f^{**} \leq f$.

The Fenchel-Moreau theorem (which we will not prove here) states that $f = f^{**}$ if and only if f is convex and closed. It will turn out to be useful to show the following property.

4. Assume that f is closed and convex. Then show that for any $x, y \in \mathbb{R}^d$,

$$\begin{aligned} y \in \partial f(x) &\Leftrightarrow x \in \partial f^*(y) \\ &\Leftrightarrow f(x) + f^*(y) = \langle x, y \rangle \end{aligned}$$