Tutorial 8 : Nonsmooth optimization - Subgradient Descent

Exercise 1 (Subgradient of functions)**.** Compute the subgradients of the following functions:

1. $\mathrm{ReLU}(x) = \max(x, 0)$.
2. ℓ_1 norm, i.e. $\|x\|_1 = \sum_{i=1}^d |x_i|$.
3. Generalize the previous question for decomposable functions, i.e., compute the subgradient $f(x) = \sum_{i=1}^d f_i(x_i)$ with those of $f_i, i = 1, \dots, d$.
4. ℓ_2 norm, i.e., $\|x\|_2 = \sqrt{\sum_{i=1}^d x_i^2}$.
5. Total variation function, i.e., $\mathrm{TV}(x) = \sum_{i=1}^{d-1} |x_{i+1} - x_i|$.
6. $f(x) = \|\mathbf{A}_1 x + b\|_1 + \|x\|_2, \mathbf{A} \in \mathbb{R}^{m \times d}, b \in \mathbb{R}^m$.

Exercise 2 (Normal cone introduction)**.** A very important concept in optimization is the normal cone. More formally, a normal cone to a convex set A at a point $x \in A$ is defined as:

$$\mathcal{N}(x, A) := \{u \in \mathbb{R}^n \mid \langle v - x, u\rangle \le 0, \forall v \in A\}.$$

Answer the following questions:

1. $\mathcal{N}(x, A)$ is a non-empty, closed and convex cone (a set A is a cone if for any $\alpha \ge 0$ and any $x \in A$, we have: $\alpha x \in A$).
2. If the set A is closed, prove that: $z \in \mathcal{N}(x, A)$ if and only if $\|z\|_2 = \mathrm{dist}(x + z, A)$.
3. If the set A is an epigraph of a convex function $f : \mathbb{R}^d \to \mathbb{R}$, i.e.,

$$\mathtt{epi}\, f = \{(x, \gamma) \mid \gamma \ge f(x)\} \subseteq \mathbb{R}^{d+1},$$

   and the point $y = (x, f(x))$, then $z \in \mathcal{N}(y, A)$ if and only if $z = (\alpha g, -\alpha)$ where $g \in \partial f(x)$ and $\alpha \ge 0$.

Exercise 3 (Normal cones computing)**.** Compute the normal cones of the following sets:

1. $A = \mathbb{R}_+^d$ is the positive orthant, at the origin $x = 0$.
2. $A = \bar{\mathcal{B}}(0, 1)$ is the closed unit ball and $x \in \mathcal{B}(0, 1)$ in the interior of A and $x \in \bar{\mathcal{B}}(0, 1) \setminus \mathcal{B}(0, 1)$ at the boundary of A.

3. A set $A \in \mathbb{R}^{d+1}$ and $x = 0 \in \mathbb{R}^d$ where $A = \{(x, y) \in \mathbb{R}^{d+1} \mid y \geq \|x\|_1\}$.

Exercise 4 (Domain assumption necessity)**.** We already see in the lecture that for two convex functions $f_1, f_2 : \mathbb{R}^d \to \mathbb{R}$, $\partial(f_1 + f_2)(x) = \partial f_1(x) + \partial f_2(x)$. However, this is not necessarily true for *extended-valued* convex functions, i.e., $f : \mathbb{R}^d \to \mathbb{R} \cup \{+\infty\}$. Consider the indicator function of a convex set A, defined as:

$$\mathbf{1}_A(x) = \begin{cases} 0, & \text{if } x \in A \\ +\infty, & \text{otherwise} \end{cases}.$$

Consider A_1 and A_2 be two closed balls and $A_1 \cap A_2 = \{x\}$ is a singleton. Answer the following questions:

1. General question: Prove that for general convex functions f_1, f_2, we always have: $\partial f_1 + \partial f_2 \subseteq \partial(f_1 + f_2)$.
2. What is the subgradient of the functions $\mathbf{1}_{A_1}$ and $\mathbf{1}_{A_2}$ at x?
3. What is the function $\mathbf{1}_{A_1} + \mathbf{1}_{A_2}$? What is its set of subgradients at the point x? Conclude that $\partial \mathbf{1}_{A_1} + \partial \mathbf{1}_{A_2} \subsetneq \partial(\mathbf{1}_{A_1} + \mathbf{1}_{A_2})$.

Exercise 5 (Prove subgradient calculus under assumptions)**.** In this section, we will prove that for two convex functions $f_1, f_2 : \mathbb{R}^d \to \mathbb{R}$, we have:

$$\partial(f_1 + f_2)(x) = \partial f_1(x) + \partial f_2(x).$$

Answer the following questions:

1. Consider the sets:
$$\Omega_1 := \{(x, \lambda_1, \lambda_2) \mid \lambda_1 \geq f_1(x)\} \subseteq \mathbb{R}^{d+2},$$
$$\Omega_2 := \{(x, \lambda_1, \lambda_2) \mid \lambda_2 \geq f_2(x)\} \subseteq \mathbb{R}^{d+2}.$$
Prove that $\mathcal{N}((x, \lambda_1, \lambda_2), \Omega_k) = \mathcal{N}((x, \lambda_k), \Omega_k) \times \{0\}, k = 1, 2$.
2. Prove that $v \in \partial(f_1 + f_2)(x)$ if and only if $(v, -1, -1) \in \mathcal{N}((x, f_1(x), f_2(x)), \Omega_1 \cap \Omega_2)$.
3. Admit that $\mathcal{N}(z, \Omega_1 \cap \Omega_2) = \mathcal{N}(z, \Omega_1) + \mathcal{N}(z, \Omega_2)$ (warning: this is not always true, as we see in the previous exercise), conclude the proof.

Exercise 6 (Alternative projection as Polyak step-sizes)**.** Finding the interesetion of $\Omega_1 \cap \ldots \cap \Omega_n$ where $\Omega_k, k = 1, \ldots, n$ are non-empty, closed, convex sets, is an important computer science question. Assume that we can project onto each Ω_k efficiently, consider the following algorithm: at the ℓth iteration, we compute $\mathrm{dist}(x_\ell, \Omega_k), k = 1, \ldots, n$ and choose $p := \arg\max_k \mathrm{dist}(x_\ell, \Omega_k)$, then:

$$x_{\ell+1} = \mathrm{argmin}_{x \in \Omega_p} \mathrm{dist}(x, \Omega_p).$$

Answer the following questions:

1. Let $f(x) = \max(\mathrm{dist}(x, \Omega_1), \ldots, \mathrm{dist}(x, \Omega_n))$. Prove that f is convex, and 1-Lipschitz.
2. Prove that $\frac{x_{\ell+1} - x_\ell}{\|x_{\ell+1} - x_\ell\|}$ is a subgradient of f at x_ℓ. Hint: What is the subgradient of the distance function to a non-empty, closed and convex set?
3. Prove that $x_{\ell+1}$ is the update of the subgradient method for the function f with Polyak step-size $\Omega_1 \cap \ldots \cap \Omega_n \neq \emptyset$.
4. If $\Omega_1 \cap \ldots \cap \Omega_n \neq \emptyset$, what can we conclude? Can we find a point $x \in \Omega_1 \cap \ldots \cap \Omega_n$ at the end of the algorithm?