ELE 539 / COS 512: Homework 1

due Mar. 3, 2021 (11:59 PM Blackboard)

1 Operations that Perserve Convexity

Prove that following operations perserve the convexity of the functions. For simplicity, you can always assume the domain of the function is $\mathbb{R}^d$.

(a) [1 point] If $f_1, \ldots, f_n$ are convex functions, $\alpha_1, \ldots, \alpha_n$ are nonnegative scalars, show that $f := \sum_{i=1}^n \alpha_i f_i$ is also a convex function.

(b) [1 point] If f_θ is a convex function for all $\theta \in \Theta$, show that $f := \sup_{\theta \in \Theta} f_\theta$ is also a convex function.

(c) [1 point] If g is a convex function, show that for arbitrary matrix $A \in \mathbb{R}^{d \times d}$, and vector $b \in \mathbb{R}^d$, function $f(x) := g(Ax + b)$ is also convex.

2 Equivalent Characterizations of Smooth Functions

There are two characterizations of ℓ-smooth functions.

Condition 1: For any $x, y \in \mathbb{R}^d$, $\|\nabla f(x) - \nabla f(y)\| \leq \ell \|x - y\|$.

Condition 2: For any $x, y \in \mathbb{R}^d$, $|f(y) - f(x) - \nabla f(x)^\top (y - x)| \leq (\ell/2)\|y - x\|^2$.

[3 points] In the lecture, we have shown that condition 1 implies condition 2. To show their equivalence, prove that condition 2 also implies condition 1. Note in this question, f is not necessarily convex or twice-differentiable. [Hint: Prove that for any $x, y, z, z' \in \mathbb{R}^d$, we have $[\nabla f(x) - \nabla f(y)]^\top (z - z') \leq (\ell/2) \cdot [\|x - z\|^2 + \|x - z'\|^2 + \|y - z\|^2 + \|y - z'\|^2]$.]

3 Recover Strongly Convex Rate by Convex Results

Suppose we have an algorithm $\mathcal{A}$ (which is not necessarily gradient descent). The algorithm takes an initial point x_1, and a integer $t \in \mathbb{N}$ as input, and has the following guarantee: for any ℓ-smooth, convex function f, after quering the gradient oracle t times, the output x_t satisfies:

$$f(x_t) - f(x^\star) \leq \frac{\ell \|x_1 - x^\star\|^2}{t}$$

[3 points] Prove that, for any ℓ-smooth, α-strongly convex function f, to find a point $\hat{x}$ such that $f(\hat{x}) - f(x^\star) \le \epsilon$, it suffices to query the gradient oracle $\tilde{\mathcal{O}}(\ell/\alpha)$ times, by smart uses of the algorithm $\mathcal{A}$. [hint: compute how many gradients that algorithm $\mathcal{A}$ requires to guarantee $\|x_t - x^\star\|^2 \le \|x_1 - x^\star\|^2/2$.]

4 Last Iterate Guarantee of GD for Convex Lipschitz Functions

In this question, we prove the following Theorem.

Theorem 1. *Suppose that convex set $\mathcal{X}$ satisfies $\sup_{x,x' \in \mathcal{X}} \|x - x'\| \le R$, and function f is convex and L-Lipschitz. Then, by running projected subgradient descent with learning rate $\eta_s = R/(L\sqrt{s})$ for all $s \in \mathbb{N}$, we have that for any $t \ge 1$:*

$$f(x_t) - f(x^\star) \le \mathcal{O}\left(\frac{RL \log t}{\sqrt{t}}\right)$$

(a) [2 point] Prove using the similar techniques taught in the lecture that for any $t, k \in \mathbb{N}$, $k \le t$:

$$\sum_{s=t-k}^{t} [f(x_s) - f(x_{t-k})] \le \mathcal{O}(RL(\sqrt{t} - \sqrt{t-k-1})). \tag{1}$$

(b) [1 point] For a fixed $t \in \mathbb{N}$, let $S_k = \frac{1}{k+1}\sum_{s=t-k}^{t} f(x_s)$, prove the following using (1).

$$S_{k-1} \le S_k + \mathcal{O}\left(\frac{RL}{k\sqrt{t}}\right)$$

(c) [1 point] Use above results to prove Theorem 1.