Optimization for Machine Learning 50.579

Instructor: Ioannis Panageas

Homework

Question 1. Show that a twice *differentiable* function $f(x)$ is convex if and only if the domain $\mathrm{dom}(f)$ is a convex set and $\forall x \in \mathrm{dom}(f)$

$$\nabla^2 f(x) \succeq 0 \text{ i.e., the Hessian is positive semi-definite.}$$

Question 2. Suppose $f(x)$ is differentiable and α-strongly convex. $\forall x, y \in \mathrm{dom}(f)$ show that

$$f(y) - f(x) \geq \nabla f(x)^\top (y-x) + \frac{\alpha}{2} \|y-x\|_2^2 .$$

Question 3. Suppose $f(x)$ is L-Lipschitz continous and $\partial f(x) \neq \emptyset$. Show that $\forall x \in \mathrm{dom}(f)$ it holds

$$\|g_x\|_2 \leq L \text{ where } g_x \in \partial f(x).$$

Question 4. Let $f : \mathbb{R}^n \to \mathbb{R}$ be a *differentiable*, convex and L-smooth. Let x^* be a minimizer of f and set $x_{t+1} = x_t - \frac{1}{L}\nabla f(x_t)$ (aka GD) with x_0 some initial condition. Show that $\|x_t - x^*\|$ is decreasing in t that is

$$\|x_{t+1} - x^*\| \leq \|x_t - x^*\| \text{ for all } t \geq 0.$$

Question 5. Let $f(x) := \mathbb{E}_{z \sim \mathcal{N}(\mu,1)}\left[-\log\left(\frac{1}{2\sqrt{2\pi}}e^{(z-x)^2/2} + \frac{1}{2\sqrt{2\pi}}e^{(z+x)^2/2}\right)\right]$ where $\mu \neq 0$ and $\mathcal{N}(\mu, 1)$ denotes the Gaussian with mean μ and variance 1. Show that $f(x)$ is not convex.

Question 6. Assume A is a $n \times n$ matrix with entries in [-1,1]. Moreover, assume that we run MWUA for the zero sum game with payoff matrix A for T iterations, starting from uniform distribution for both players. Let $\tilde{x} = \frac{1}{T}\sum_t p_x^t$ and $\tilde{y} = \frac{1}{T}\sum_t p_y^t$. For $T = \Theta\left(\frac{\log n}{\epsilon^2}\right)$ show that $(\tilde{x}, \tilde{y})$ is an ϵ-approximate NE that is

$$\tilde{x}^\top A\tilde{y} \leq x'^\top A\tilde{y} + \epsilon \text{ for all } x' \in \Delta_n \text{ and } \tilde{x}^\top A\tilde{y} \geq \tilde{x}^\top Ay' - \epsilon \text{ for all } y' \in \Delta_n{}^1.$$

Question 7. Let $f_k : \Delta_n \to \mathbb{R}$ be such that $f_k(x) = x^\top c_k$ (linear function). Moreover, let $R := \|x\|_2^2$ (Euclidean norm squared which is strongly convex). Show that

$$x_t = \mathrm{argmin}_{x \in \Delta_n}\{\epsilon \cdot \sum_{k=0}^{t-1} f_k(x) + R(x)\}$$

is the same dynamics (algorithm) as MWUA.

[1] Δ_n denotes the simplex of size n.

Question 8. Suppose that $f : \mathbb{R}^n \to \mathbb{R}$ is a twice differentiable, convex function that is locally α-strongly convex around x, in the sense that $f(y) \geq f(x) + (y-x)^\top \nabla f(x) + \frac{\alpha}{2} \|y - x\|_2^2$ holds for all vectors y in the ball $\mathbb{B}_2 = \{y : \|y - x\|_2 \leq \rho\}$. Show that

$$(y-x)^\top (\nabla f(y) - \nabla f(x)) \geq \rho\alpha \|y - x\|_2 \text{ for all } y \in \mathbb{R}^n \backslash \mathbb{B}_2.$$

Due 30th April (23:59pm) on edimension.