SI251 Convex Optimization, Spring 2019
Final Exam

Note: We are interested in the reasoning underlying the solution, as opposed to simply the answer. Thus, solutions with the correct answer but without adequate explanation will not receive full credit; on the other hand, partial solutions with explanation will receive partial credit. Within a given problem, you can assume the results of previous parts in proving later parts (e.g., it is fine to solve part 3) first, assuming the results of parts 1) and 2)). Your use of resources should be limited to printed lecture slides, lecture notes, homework, homework solutions, general resources, class reading and textbooks, and other related textbooks on optimization. You should not discuss the final exam problems with anyone or use any electronic devices. Detected violations of this policy will be processed according to ShanghaiTech's code of academic integrity. Please hand in the exam papers and answer sheets at the end of exam.

I. Basic Knowledge

1. The polar of $\mathcal{C} \subseteq \mathbb{R}^n$ is defined as the set

$$\mathcal{C}^\circ = \{\boldsymbol{y} \in \mathbb{R}^n | \boldsymbol{y}^T \boldsymbol{x} \leq 1 \text{ for all } \boldsymbol{x} \in \mathcal{C}\}.$$

   Determine whether $\mathcal{C}^\circ$ is convex or not? Give necessary explanations.(5 points)

2. How to determine the strict saddle points for a twice continuously differentiable function $f : \mathbb{R}^n \to \mathbb{R}$ (i.e., $f \in \mathcal{C}^2$)? Provide a method to escape the saddle points. (5 points)

3. Is the following statement true or false? Give necessary explanations.
   Consider an arbitrary set $\mathcal{S}$. For any $\boldsymbol{x}$ and $\boldsymbol{z}$ we always have $\|\mathcal{P}_\mathcal{S}(\boldsymbol{x}) - \mathcal{P}_\mathcal{S}(\boldsymbol{z})\|_2 \leq \|\boldsymbol{x} - \boldsymbol{z}\|_2$, where $\mathcal{P}_\mathcal{S}(\cdot)$ represents Euclidean projection onto set S. (5 points)

4. Compute the proximal operator prox_f, where $f = -\sum_{j=1}^d \log(x_j)$ with $x_j \in \mathbb{R}_+$. (5 points)

II. Water Filling

We consider the following optimization problem

$$\begin{aligned} \min_{\boldsymbol{x} \in \mathbb{R}^n} \quad & -\sum_{i=1}^n \log(\alpha_i + x_i) \\ \text{subject to} \quad & \mathbf{1}^T \boldsymbol{x} = 1 \\ & x_i \geq 0, \; i = 1, ..., n \end{aligned} \tag{1}$$

where $\alpha_i > 0$. This problem arises in information theory, in allocating power to a set of n communication channels. The variable x_i represents the transmitter power allocated to the ith channel, and $\log(\alpha_i + x_i)$ gives the capacity or communication rate of the channel, so the problem is to allocate a total power of one to the channels, in order to maximize the total communication rate.

1. Determin that this problem is convex or not and provide your argument. (5 points)
2. Write down the dual problem of (1). (5 points)
3. Derive the KKT conditions of (1). (5 points)
4. Derive the expression of the optimal solution of (1). (5 points)

III. Non-smooth Optimization

Define the function $f : \mathbb{R} \to \mathbb{R}$ via

$$f(x) = \max_{|y| \leq 1} \; yx. \tag{2}$$

1. What is the subdifferential of f at x? Is f differentiable? (5 points) (Hint: Consider Danskin's Theorem.)

2. For some $\mu > 0$, define the function

$$f_\mu(x) := \max_{|y|\le 1}\{yx - \frac{\mu}{2}y^2\}. \tag{3}$$

   (a) Show that $f_\mu(x)$ is convex and differentiable. (2 points)
   (b) Prove that $f_\mu(x)$ is μ-smooth. (3 points)
3. Show that $f_\mu(x) \le f(x) \le f_\mu(x) + \frac{\mu}{2}$. (5 points)
4. Use the previous parts to suggest a gradient-based algorithm for minimizing f up to accuracy ϵ in function value. Specify the number of iterations $T(\epsilon)$ that are needed. (5 points)

IV. Gradient Descent and Convergence Analysis

Let $\boldsymbol{z} = (x, y)$ and

$$f(\boldsymbol{z}) = \frac{1}{2}e^{2x-1} + x\sin y.$$

1. Compute the gradient and the Hessian for function f.(5 points)
2. Find all minima of function f.(5 points)

For the remaining part of this problem we discuss one step of a line search method, starting from $\boldsymbol{z}_0 = (0.5, -\pi/3)$, with search direction $\boldsymbol{p}_0 = (0, -1)$.

3. Confirm that $\boldsymbol{p}_0$ is a descent direction from $\boldsymbol{z}_0$. Give the condition that a direction $\boldsymbol{p}$ is descent starting from $\boldsymbol{z}_0$. (5 points)
4. Do one step of the line search method (i.e., $\boldsymbol{z}_1 = \boldsymbol{z}_0 + \mu_0\boldsymbol{p}_0$), using the exact value for the step size μ_0.(5 points)

V. Proximal Gradient Algorithm and Conditional Gradient Method

1. Consider an unconstrained optimization problem $\min_{\boldsymbol{x}}\{f(\boldsymbol{x}) + h(\boldsymbol{x})\}$, where f is convex and has a continuous derivative. Compute the form of the updates of proximal gradient algorithm for the following cases:
   (a) $h(\boldsymbol{x}) = \|\boldsymbol{x}\|_1$ with $\boldsymbol{x} \in \mathbb{R}^d$, and $f : \mathbb{R}^d \to \mathbb{R}$. (4 points)
   (b) $h(\boldsymbol{X}) = \sum_{j=1}^{d} \sigma_j(\boldsymbol{X})$ with $\boldsymbol{X} \in \mathbb{R}^{d\times d}$, where $\sigma_j(\boldsymbol{X})$ is the j^{th} singular value, and $f : \mathbb{R}^{d\times d} \to \mathbb{R}$. (4 points)
2. Consider a constrained optimization problem $\min_{\boldsymbol{x}\in\mathcal{C}} f(\boldsymbol{x})$ where $\mathcal{C}$ is a compact convex set, and f is convex and has a continuous derivative. The conditional gradient method with stepsizes $\{\alpha^\ell\}_{\ell=1}^{\infty}$ generates a sequence of the form

$$\boldsymbol{x}^{\ell+1} = \left(1 - \alpha^\ell\right)\boldsymbol{x}^\ell + \alpha^\ell \boldsymbol{z}^\ell,$$

   where $\boldsymbol{z}^\ell \in \arg\min_{\boldsymbol{z}\in\mathcal{C}} \left\langle \nabla f\left(\boldsymbol{x}^\ell\right), \boldsymbol{z}\right\rangle$. Compute the form of these updates for the following cases:
   (c) $\mathcal{C} = \left\{\boldsymbol{x} \in \mathbb{R}^d \middle| \|\boldsymbol{x}\|_1 \le 1\right\}$, and $f : \mathbb{R}^d \to \mathbb{R}$. (4 points)
   (d) $\mathcal{C} = \{\boldsymbol{X} \in \mathbb{R}^{d\times d} | \sum_{j=1}^{d} \sigma_j(\boldsymbol{X}) \le 1\}$ where $\sigma_j(\boldsymbol{X})$ is the j^{th} singular value, and $f : \mathbb{R}^{d\times d} \to \mathbb{R}$. (4 points)
3. Comment on the complexity between the updates in part (a) and in part (c), and the complexity between the updates in part (b) and in part (d). (4 points)