SI251 - Convex Optimization, Fall 2021
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. Is the set $\{\boldsymbol{a} \in \mathbb{R}^k \mid p(0) = 1,\ |p(t)| \le 1 \text{ for } \alpha \le t \le \beta\}$, where $p(t) = a_1 + a_2 t + \cdots + a_k t^{k-1}$ convex? (10 points)
2. Prove that $\boldsymbol{x}^* = (1, \frac{1}{2}, -1)$ is optimal for the optimization problem

$$\begin{aligned} \text{minimize} \quad & \frac{1}{2}\boldsymbol{x}^T \boldsymbol{P}\boldsymbol{x} + \boldsymbol{q}^T\boldsymbol{x} + r \\ \text{subject to} \quad & -1 \le x_i \le 1, \quad i = 1, 2, 3, \end{aligned}$$

   where

$$\boldsymbol{P} = \begin{bmatrix} 13 & 12 & -2 \\ 12 & 17 & 6 \\ -2 & 6 & 12 \end{bmatrix}, \ \boldsymbol{q} = \begin{bmatrix} -22 \\ -14.5 \\ 13 \end{bmatrix}, \ r = 1.$$

   (10 points)
3. Equivalently reformulate the following problem into a standard convex optimization form, i.e., Linear Programming (LP), Second-Order Cone Programming (SOCP) and Semidefinite Programming (SDP).

$$\text{Minimize} \quad ||\boldsymbol{A}\boldsymbol{x} - \boldsymbol{b}||_\infty$$

   (10 points)

II. KKT Conditions

Let $n \ge 2$. Consider the following minimization problem over $\boldsymbol{x} = [x_1, \ldots, x_n]^T$,

$$\begin{aligned} \min_{\boldsymbol{x}} \quad & \sum_{j=1}^{n} \frac{c_j}{x_j} \\ \text{s.t.} \quad & \sum_{j=1}^{n} x_j = 1, \\ & x_j \ge \epsilon, j = 1, \ldots, n, \end{aligned} \tag{1}$$

where $c_j > 0,\ \forall j, \epsilon > 0$ are parameters.

(1) Determine whether 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). (10 points)

III. **Gradient Methods**

1. Let $f : \mathbb{R}^d \to \mathbb{R}$ be a convex and differentiable function, it is said to be M-smooth if

$$f(\boldsymbol{y}) \leq f(\boldsymbol{x}) + \langle \nabla f(\boldsymbol{x}),\ \boldsymbol{y} - \boldsymbol{x} \rangle + \frac{M}{2} \|\boldsymbol{y} - \boldsymbol{x}\|_2^2, \quad \forall \boldsymbol{x}, \boldsymbol{y} \in \mathbb{R}^d. \tag{2}$$

Prove that the above M-smooth condition holds *if and only if*

$$\|\nabla f(\boldsymbol{x}) - \nabla f(\boldsymbol{y})\|_2 \leq M \|\boldsymbol{x} - \boldsymbol{y}\|_2, \quad \forall \boldsymbol{x}, \boldsymbol{y} \in \mathbb{R}^d. \tag{3}$$

(20 points)

2. The subgradient method updates are given by $\boldsymbol{x}^{\ell+1} = \boldsymbol{x}^{\ell} - \alpha^{\ell} \boldsymbol{g}^{\ell}$, where $\boldsymbol{g}^{\ell} \in \partial f(\boldsymbol{x}^{\ell})$. Suppose f is convex with finite unconstrained minimum $f^\star$, and is L-Lipschitz. Letting $\mathcal{X}^\star$ denote the space of minimizers (taken to be non-empty and closed), define the distance function $\mathrm{dist}_{\mathcal{X}^\star}(\boldsymbol{u}) = \inf_{\boldsymbol{z} \in \mathcal{X}^\star} \|\boldsymbol{z} - \boldsymbol{u}\|_2$, and suppose there exists some $\mu > 0$ such that $\mathrm{dist}_{\mathcal{X}^\star}(\boldsymbol{y}) \leq \frac{1}{\mu}(f(\boldsymbol{y}) - f^\star)$ for all $\boldsymbol{y} \in \mathrm{dom}(f)$. Prove that

$$\mathrm{dist}^2_{\mathcal{X}^\star}(\boldsymbol{x}^{\ell+1}) \leq \left(1 - \frac{\mu^2}{L^2}\right)^{\ell+1} \mathrm{dist}^2_{\mathcal{X}^\star}(\boldsymbol{x}^0).$$

(15 points)

IV. **Proximal Algorithm**

For $\boldsymbol{0} \prec \boldsymbol{\alpha}, \boldsymbol{\beta} \prec \boldsymbol{1}$, define $h_{\boldsymbol{\alpha},\boldsymbol{\beta}} : \mathbb{R}^n \to \mathbb{R}$ as

$$h_{\boldsymbol{\alpha},\boldsymbol{\beta}}(\boldsymbol{x}) = \boldsymbol{\alpha}^T \boldsymbol{x}_+ + \boldsymbol{\beta}^T \boldsymbol{x}_-,$$

where $\boldsymbol{x}_+ = \max\{\boldsymbol{x}, \boldsymbol{0}\}$ and $\boldsymbol{x}_- = \max\{-\boldsymbol{x}, \boldsymbol{0}\}$, the maximum is taken elementwise.

Give a simple expression for the proximal operator of $h_{\boldsymbol{\alpha},\boldsymbol{\beta}}$. (10 points)