SI251 - Convex Optimization, Fall 2022
Final Exam

Note: We are interested in the reasoning underlying the solution, as opposed to simply the answer. Thus, solutions with the correct answer without adequate explanation will not receive full credit; on the other hand, partial solutions with an 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)). The resources you use 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 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 the exam.

I. **Basic Knowledge**

1. (20 points) Determine the following statements are true or false and explain your reasons.

(a) Suppose $f : \mathbb{R}^n \to \mathbb{R}$ is convex and $g : \mathbb{R} \to \mathbb{R}$ is concave and non-decreasing, then the function $g \circ f$ is convex.

(b) $f(x) = e^{-\|x\|^2}$ is convex, $x \in \mathbb{R}^n$.

(c) The set $\{x \in \mathbb{R}^n : x_n^2 \geq \sum_{i=1}^{n-1} x_i^2, x_n \geq 0\}$ is a self-dual convex cone.

(d) $f(x) = x_1^2 - 2x_1x_2 + \frac{1}{3}x_2^3 - 8x_2$ has no global maximizers or minimizers.

2. (20 points) Derive the conjugate function $f^*(y)$ for each of the following functions:

(a) $f(x) = -\ln x + 2$; (b) $f(x) = x^p$ on $[0, +\infty)$, where $p > 1$.

3. (10 points) Consider the function $f : \mathbb{R} \to (-\infty, \infty]$ given for any $x \in \mathbb{R}$ by

$$f(x) = \begin{cases} \mu x, & 0 \leq x \leq \alpha \\ \infty, & \text{else} \end{cases}$$

where $\mu \in \mathbb{R}$ and $\alpha \in [0, \infty]$. Please represent and compute the proximal operator of f.

II. **Advanced Knowledge**

4. (15 points) Derive the KKT conditions for the problem

$$\begin{array}{ll} \text{minimize} & \operatorname{tr} X - \log\det X \\ \text{subject to} & Xs = y, \end{array}$$

with variable $X \in \mathbf{S}^n$ and domain $\mathbf{S}^n_{++} \cdot y \in \mathbf{R}^n$ and $s \in \mathbf{R}^n$ are given, with $s^T y = 1$. Verify that the optimal solution is given by

$$X^\star = I + yy^T - \frac{1}{s^T s} ss^T.$$

[hint: $\frac{\partial det(X)}{\partial X} = det(X)(X^{-1})^T$]

5. (10 points) Prove: suppose that f is a convex function, $f(x)$ is L-Lipschitz if and only if the subgradient of f is bounded, i.e.,

$$\|g\| \leq L, \quad \forall g \in \partial f(x), \quad x \in R^n. \tag{1}$$

6. (10 points) Convex-Concave Programming (CCP) is a well-known algorithm to handle the difference between two convex functions. Suppose f_i and g_i are convex functions with $i = 0, 1, \cdots, m$. Please show the monotone property of the CCP iteration, i.e. $f_0(x_{k+1}) - g_0(x_{k+1}) \le f_0(x_k) - g_0(x_k)$.

DCP problems:

$$\begin{array}{ll} \text{minimize} & f_0(x) - g_0(x) \\ \text{subject to} & f_i(x) - g_i(x) \le 0, \quad i = 1, \ldots, m. \end{array}$$

Algorithm 1: CCP algorithm

1: given an initial feasible point x_0.
2: $k := 0$.
3: **repeat**
4: **Convexify**. Form $\hat{g}_i(x; x_k) \triangleq g_i(x_k) + \nabla g_i(x_k)^T (x - x_k)$ for $i = 0, \ldots, m$.
5: **Solve**. Set the value of x_{k+1} to a solution of the convex problem

$$\begin{array}{l} \text{minimize } f_0(x) - \hat{g}_0(x; x_k) \\ \text{subject to } f_i(x) - \hat{g}_i(x; x_k) \le 0, \quad i = 1, \ldots, m. \end{array}$$

6: **Update iteration**. $k := k + 1$.
7: **until** stopping criterion is satisfied.

7. (15 points) Consider the least squares regression problem with ℓ_2-norm regularization,

$$\underset{\boldsymbol{x} \in \mathbb{R}^d}{\text{minimize}} \quad \frac{1}{2} \sum_{i=1}^{N} \|\boldsymbol{A}_i \boldsymbol{x} - \boldsymbol{b}_i\|_2^2 + (\lambda/2)\|\boldsymbol{x}\|_2^2$$

where $\boldsymbol{A}_i \in \mathbb{R}^{n_i \times d}$, $\boldsymbol{b}_i \in \mathbb{R}^{n_i}$, λ is the regularization parameter. Please write the **exact** ADMM update steps for this problem. (Note: the *argmin* operator should not be kept.)