# 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 whether the following statements are true or false and explain your reasons.

   (a) Convex optimization problems are tractable because they always have a unique global minimum.

   (b) Gradient descent method with constant or diminishing step size can only ensure convergence to a stationary point, but with backtracking line search, a globally optimal solution is guaranteed.

   (c) $f(\boldsymbol{x}) = \sum_{i=1}^{m} e^{-1/f_i(\boldsymbol{x})}$ is convex if the functions $f_i$ are convex, $\mathbf{dom} f = \{\boldsymbol{x} \mid f_i(\boldsymbol{x}) < 0, i = 1, \cdots, m\}$.

   (d) We cannot maximize a convex function with respect to a convex set because the objective function will be unbounded above and no valid solution exists.

2. (20 points) Consider the function $f(\boldsymbol{x}) = x_1 \ln x_1 + x_2 \ln x_2$, defined for $\boldsymbol{x} = (x_1, x_2)^T \in (0, \infty)^2$ (i.e. $x_1, x_2 > 0$).

   (a) Compute $\boldsymbol{p}_0 = -\nabla f(\boldsymbol{x}_0)$ when $\boldsymbol{x}_0 = (1, 1)^T$, and find the value $\alpha_0^*$ which minimizes the function $\varphi(\alpha) = f(\boldsymbol{x}_0 + \alpha \boldsymbol{p}_0)$.

   (b) Write down the gradient descent algorithm. How many steps does it take to converge for optimal $\alpha^*$ and $\boldsymbol{x}_0 = (1, 1)^T$ ?

3. (10 points) Consider the LP

$$\begin{array}{ll} \text{minimize} & c^T x \\ \text{subject to} & Ax \preceq b \end{array}$$

   with $A \in \mathbb{R}^{M \times N}$ square and nonsingular, $x, c \in \mathbb{R}^N$, $b \in \mathbb{R}^M$. Compute the optimal value $p^*$.

II. **Advanced Knowledge**

4. (10 points) Let $f : \mathbb{R}^2 \to \mathbb{R}$ be defined by $f(x_1, x_2) = |x_1| + 2|x_2|$. Compute $\partial f(x_1, x_2)$ for each $x_1, x_2 \in \mathbb{R}^2$.

5. (10 points) Let $f : \mathbb{R}^n \to \mathbb{R}$ be a convex function of $x \in \mathbb{R}^n$, and let its associated proximal operator be $\operatorname{prox}_f(x)$ (which is a mapping from $\mathbb{R}^n$ to $\mathbb{R}^n$ ). Assuming that $f(x)$ is separable,

$$f(x) = \sum_{i=1}^{n} f_i(x_i),$$

   derive an expression for $\operatorname{prox}_f(x)$ in terms of $\operatorname{prox}_{f_i}(x_i)$.

6. (15 points) Consider the following regularized least square problem with one scalar inequality constraint:

$$\begin{array}{ll} \underset{\boldsymbol{\beta}}{\text{minimize}} & \|A\boldsymbol{\beta} - b\|_2^2 + \tau \|\boldsymbol{\beta} - \hat{\boldsymbol{\beta}}\|_2^2 \\ \text{subject to} & \beta_1 \geq \epsilon, \end{array}$$

where $A \in \mathbb{R}^{n\times n}, \boldsymbol{\beta}, b, \hat{\boldsymbol{\beta}} \in \mathbb{R}^n$, and $\beta_1$ is the first component of vector $\boldsymbol{\beta}$. Please find the optimal $\boldsymbol{\beta}^*$ by solving the KKT system.

7. (15 points) **Quantile Regression.** For $\alpha \in (0,1)$, define the quantile loss function, $h_\alpha : \mathbb{R}^N \to \mathbb{R}$ , as

$$h_\alpha(\mathbf{x}) = \alpha \mathbf{1}^T \mathbf{x}_+ + (1-\alpha)\mathbf{1}^T \mathbf{x}_-,$$

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

(a) Find the subdifferential of $h_\alpha$ at $\mathbf{x}$. For $\alpha = 0.2$, plot the quantile function and its subdifferential $\partial h_\alpha(\mathbf{x})$ as a function of $\mathbf{x}$ assuming that $N = 1$. Explain what happens when $\alpha = 0.5$.

(b) The quantile regression problem is

$$\text{minimize}\, h_\alpha(A\mathbf{x} - b)$$

with variable $\mathbf{x} \in \mathbb{R}^N$, and parameters $A \in \mathbb{R}^{M\times N}$ and $b \in \mathbb{R}^M$. Explain how to write the quantile regression problem as a Linear Program (LP).

(c) Explain how to use the Alternating Direction Method of Multipliers (ADMM) algorithm to solve the quantile regression problem. Give the detailed update equations (not just the update optimization problems).