Management Sciences Topics: Convex Optimization
Midterm Exam: Due March 9 (at noon)

(You can directly use any properties, theorems, examples or facts from the lectures.)

Problem 1 (15 points): Are the following functions convex? You need to provide the reason only if your answer is Yes.

a. $f(x, y) = \log(x^2 + xy + y^2 + 1)$, $x, y \in \mathbb{R}$

b. $f(x) = \exp\left(-\|Ax + b\|_2^2\right)$

c. $f(x) = \exp\left(\|Ax + b\|_2^2\right)$

where $A \in \mathbb{R}^{n\times d}$ and $b \in \mathbb{R}^n$.

Problem 2 (20 points): Write down the KKT condition and the dual problem of the following minimization problem? Your formulation of the dual problem should not contain the primal decision variable x any more.

$$\begin{aligned} \min_{\mathbf{x}\in\mathbb{R}^n} \quad & \sum_{i=1}^n x_i \ln(x_i/y_i) \\ \text{s.t.} \quad & \sum_{i=1}^n x_i = 1 \\ & a_i \le x_i \le b_i, i = 1, 2, \ldots, n, \end{aligned}$$

where $y_i > 0$, $0 < a_i < b_i$ for $i = 1, 2, \ldots, n$.

Problem 3 (15 points): Suppose $h = \lambda\|\mathbf{x}\|_2$ with $\lambda > 0$ and $\mathbf{z} \in \mathbb{R}^d$. Prove that

$$\mathrm{Prox}_h(\mathbf{z}) = \begin{cases} (\|\mathbf{z}\|_2 - \lambda) \cdot \frac{\mathbf{z}}{\|\mathbf{z}\|_2} & \text{if } \|\mathbf{z}\|_2 \ge \lambda \\ \mathbf{0} & \text{if } \|\mathbf{z}\|_2 < \lambda \end{cases}$$

where $\mathrm{Prox}_h(\mathbf{z})$ is the proximal mapping with respect to h, i.e., $\mathrm{Prox}_h(\mathbf{z}) = \arg\min_{\mathbf{x}\in\mathbb{R}^d} \frac{1}{2}\|\mathbf{x} - \mathbf{z}\| + h(\mathbf{x})$. You can use the fact that the $\mathbf{x}$ is optimal if and only if the subdifferential of the objective function at $\mathbf{x}$ contains zero.

Problem 4 (25 points): Implement the Accelerated Proximal Gradient (APG) method **without** line search to solve the following problem (elastic net regularized logistic regression).

$$\min_{\mathbf{x}\in\mathbb{R}^d} \qquad \frac{1}{n}\sum_{i=1}^{n}\log(1+\exp(-b_i\mathbf{a}_i^\top\mathbf{x}))+\lambda_1\|x\|_1+\frac{\lambda_2}{2}\|x\|_2^2$$

where $\lambda_1 = 0.1$, $\lambda_2 = 0.001$, $\mathbf{a}_i \in \mathbb{R}^d$ is the feature vector of data point i and $b_i \in \{1,-1\}$ is its class label. The dataset $\{\mathbf{a}_i, b_i\}_{i=1}^n$ is the *news20.binary* dataset from the LIBSVM library for (`https://www.csie.ntu.edu.tw/~cjlin/libsvmtools/datasets/binary.html#news20.binary`). You can manually tune η and γ_0 to achieve a good convergence performance. However, you must choose a reasonable value for $\mu > 0$ such that it is the strong convexity parameter of the problem. It does not have to be the largest strong convexity parameter. You must run the algorithm for at least 200 iteration.

Problem 5 (25 points): Implement the Stochastic Subgradient (SSG) method to solve the following problem (sparse hinge loss minimization).

$$\min_{\mathbf{x}\in\mathbb{R}^d} \qquad \frac{1}{n}\sum_{i=1}^{n}\max\{1-b_i\mathbf{a}_i^\top\mathbf{x},0\}+\lambda\|x\|_1$$

where $\lambda = 0.01$, $\mathbf{a}_i \in \mathbb{R}^d$ is the feature vector of data point i and $b_i \in \{1,-1\}$ is its class label. The dataset $\{\mathbf{a}_i, b_i\}_{i=1}^n$ is the same as in Problem 4. Sample i from $\{1,2,\ldots,n\}$ uniformly randomly to construct the stochastic subgradient in each iteration. You can manually tune C in the step size $\eta_k = \frac{C}{\sqrt{K}}$ to achieve a good convergence performance. You must run the algorithm for at least $K \geq 200$ iteration. For the purpose of this exam, you are allowed to use the entire dataset to compute the objective value in every iteration.