Management Sciences Topics: Convex Optimization
Homework 3: Due April 4rd (11:59 pm)

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

Problem 1: Consider the unconstrained min-max problem

$$\min_{\mathbf{x}} \max_{\mathbf{y}} f(\mathbf{x}, \mathbf{y})$$

where f real-valued, μ-strongly convex in $\mathbf{x}$ for any fixed $\mathbf{y}$, and μ-strongly concave in $\mathbf{y}$ for any fixed $\mathbf{x}$. Modify the primal-dual subgradient method for finding a saddle-point point of f such that the algorithm has a convergence rate of $O(\frac{1}{k})$ and does not require a bounded domain. Prove the convergence rate of your algorithm by following the analysis of the stochastic subgradient method with $\mu > 0$.

Problem 2: Apply Nesterov's smoothing method to the following problem (overlapping group regularized logistic regression).

$$\min_{\mathbf{x} \in \mathcal{X}} \qquad \frac{1}{n} \sum_{i=1}^{n} \log(1 + \exp(-b_i \mathbf{a}_i^\top \mathbf{x})) + \lambda \sum_{g \in \mathcal{G}} \|\mathbf{x}_g\|_2$$

where $\mathcal{X} = \{\mathbf{x} \in \mathbb{R}^d | \|\mathbf{x}\|_2 \leq 10\}$, g is a subset of $\{1, 2, \ldots, d\}$, $\mathbf{x}_g$ is the sub-vector of $\mathbf{x}$ that consists of the coordinates indexed by g, $\lambda = 0.0001$, $\mathbf{a}_i \in \mathbb{R}^d$ 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 *covtype.libsvm.binary.scale.bz2* from the LIBSVM library for (`https://www.csie.ntu.edu.tw/~cjlin/libsvmtools/datasets/binary.html#covtype.binary`). Note that b_i in the original dataset is in $\{1, 2\}$ instead of $\{1, -1\}$. You will need to first convert 1 to -1 and 2 to 1. This dataset has a dimension $d = 54$. The set $\mathcal{G}$ contains 16 g's and is defined as follows

$$\left\{ \begin{array}{c} g_1 = \{1, 2, 3, 4, 5, 6, 7, 8, 9\} \\ g_2 = \{4, 5, 6, 7, 8, 9, 10, 11, 12\} \\ g_3 = \{7, 8, 9, 10, 11, 12, 13, 14, 15\} \\ \cdots\cdots\cdots \\ g_{16} = \{46, 47, 48, 49, 50, 51, 52, 53, 54\} \end{array} \right\},$$

namely, the indexes in g_i increase by three when i increases by one. Represent $h(\mathbf{x}) = \lambda \sum_{g \in \mathcal{G}} \|\mathbf{x}_g\|_2$ as $h(\mathbf{x}) = \max_{\mathbf{y} \in \mathcal{Y}} \mathbf{y}^\top A\mathbf{x}$ for some A where $\mathcal{Y}$ is a Cartesian product of Euclidean balls. Use APG without line search to solve the smooth approximation problem with a smoothing parameter μ. You can manually tune η and μ to achieve a good convergence performance. Plot the objective value in each iteration.

Problem 3: Apply the primal-dual subgradient method to the same problem in Problem 2 with the same dataset and parameters. You can manually tune the constant C in η_k to achieve a good convergence performance. Plot the objective value of the average iterate in each iteration.