Homework 4 – Due by 11:59 PM on Sunday April 6

Instruction: For questions 1-3, you need to use the definition of subgradient. For question 4, use the first order optimality condition of prox. For question 5 (b,c,d), you need to submit your MATLAB (.m) files or Python (.py or .txt) files. When we run your code, the desired output should be displayed; otherwise, you lose points.

(1) Show that $\partial(\lambda f(x)) = \lambda \partial f(x)$ for any $\lambda > 0$.

(2) For a convex function f and subgradients $g_x \in \partial f(x)$, $g_y \in \partial f(y)$, prove that

$$(g_x - g_y)^T(x - y) \geq 0.$$

(3) x^* is the global minimum of f, iff $0 \in \partial f(x^*)$.

(4) Show that x^* minimizes a non-differentiable function f if and only if $x^* = prox_f(x^*)$.

(5) Let $g(x) = \sum_{i=1}^n \log(1 + \exp(-a_i^T x))$ and consider the following logistic regression problem:

$$\min_{x \in \mathbb{R}^m} \ f(x) \triangleq g(x) + \lambda \|x\|_1,$$

Note that $\nabla g(x) = \sum_{i=1}^n \dfrac{-a_i(\exp(-a_i^T x))}{1 + \exp(-a_i^T x)}$.

(a) Show that Lipschitz constant of $\nabla g(x)$ is $L = \dfrac{\| \sum_{i=1}^n a_i a_i^T \|}{4} = \dfrac{\|A\|^2}{4}$, where $A = [a_i^T]_{i=1}^n \in \mathbb{R}^{n \times m}$.

Problem Setup for parts (b)-(d). Fix the seed to 123. Generate $A = [a_i^T]_{i=1}^n \in \mathbb{R}^{n \times m}$ randomly with normal distribution, i.e., `A = randn(n,m);`. Let $n = 500$, $m = 100$, $\lambda = 10^{-3}$, $x_0 = \mathbf{0} \in \mathbb{R}^m$ and set the total number of iterations as $maxiter = 200$.

(b) Solve the problem by subgradient method with diminising stepsize $\alpha_k = 1/\sqrt{k}$. The ouptut of the algorithm should be the best objective value.

(c) Solve the problem by proximal gradient method with stepsize $\alpha = 1/L$. The output of the algorithm should be the objective value of the last iterate.

(d) Solve the problem by FISTA with stepsize $\alpha = 1/L$. The output of the algorithm should be the objective value of the last iterate.

(e) Compare the output of the three algorithms. What can you conclude? i.e., Which one has the fastest and which one has the slowest convergence?

Students enrolled in SIE 549 must solve the following problem.
Students in SIE 449 will get extra credit by solving it.

(6) Let f be a continuously differentiable convex function over a closed and convex set $C \subseteq \mathbb{R}^n$. Show that if $x^* \in C$ is an optimal solution of

$$\min \{f(x) \mid x \in C\}$$

then $\nabla_x f(x)^T(x^* - x) \leq 0$ for all $x \in C$.

Hint: Use that fact that if f is convex then $(\nabla f(y) - \nabla f(x))^T(x - y) \leq 0$ for any $x, y \in C$.

Extra Credit Question. Consider the problem

$$\begin{aligned} \min \quad & f(x) \\ \text{s.t.} \quad & g(x) \leq 0 \\ & x \in X, \end{aligned} \tag{P}$$

where f and g are convex functions over $\mathbb{R}^n$ and $X \subseteq \mathbb{R}^n$ is a convex set. Suppose x^* is an optimal solution of (P) that satisfies $g(x^*) < 0$. Show that x^* is an optimal solution of

$$\begin{aligned} \min \quad & f(x) \\ \text{s.t.} \quad & x \in X. \end{aligned} \tag{Q}$$

Hint: Prove by contradiction.