The University of Texas at Austin
Department of Electrical and Computer Engineering

EE381V-11: Large Scale Optimization — Fall 2012

PROBLEM SET EIGHT

Caramanis/Sanghavi — Due: Thursday, November 16, 2012.

Written Problems

1. Consider the ℓ_1-regularized regression problem

$$\min_{x} \quad \frac{1}{2}\|y - Ax\|_2^2 \, + \, \lambda\|x\|_1 \tag{1}$$

Where $x \in \mathbb{R}^n$, $y \in \mathbb{R}^m$. Show that a point $\bar{x}$ is an optimum of this problem if and only if there exists a $z \in \mathbb{R}^n$ such that both the following hold:
(a) $-A'(y - A\bar{x}) + \lambda z = 0$
(b) For every $i \in [n]$, $z_i = sign(\bar{x}_i)$ if $\bar{x}_i \neq 0$, and $|z_i| \leq 1$ if $\bar{x}_i = 0$.

2. Make a stochastic sub-gradient algorithm for solving (1), such that each step represents an unbiased sub gradient, and each step only uses only k randomly selected rows of A (out of the total of m rows). That is, clearly specify the update rule, and show it is unbiased.

3. Make a stochastic sub-gradient algorithm for solving (1), such that each step represents an unbiased sub gradient, and each step only uses only a size $k \times l$ random sub-matrix of A. That is, clearly specify the update rule, and show it is unbiased.

4. In class we learnt that the convergence of sub gradient descent is given by

$$f_{k,best} - f^* \;\leq\; \frac{R^2 + G^2 \sum_{i \leq k} h_i^2}{\sum_{i \leq k} h_i}$$

where h_i were the step sizes. We also learnt that, for a *fixed* k, the lowest this bound could be is $\frac{RG}{\sqrt{k+1}}$; this is achieved by choosing every $h_i = \frac{R/G}{\sqrt{k+1}}$. Thus, each k needs a different sequence to achieve the best lower bound.

Suggest a *single* sequence of h_i's that makes the above bound decay as $O(\frac{\log k}{\sqrt{k}})$, for *every* k. Prove your result.

5. In class we saw how to generate an unbiased stochastic sub gradient by doing coordinate descent on a coordinate picked uniformly at random. Suppose now we want to pick coordinates non-uniformly; in particular, suppose that at each point x we have a probability $p_i(x) > 0$ for each coordinate i, such that $\sum_i p_i(x) = 1$. Devise an update rule which generates an unbiased sub gradient at x, by sampling a single coordinate according to the $p(x)$ distribution.