Solutions to hw3 homework on Convex Optimization https://web.stanford.edu/class/ee364b/homework.html

Andrei Keino

March 2, 2021

3.1 (4 points)

Consider the optimization problem

$$minimize_{\{x_j\}_{j=1}^{J}}\ f(x_1, \dots, x_j) := \frac{1}{2}\|b - \sum_{j=1}^{J} A_j x_j\|_2^2 + \lambda \cdot \sum_{j=1}^{J} \|x_j\|_2,$$
$$s.t.\ A_j x_j \geq 0,\ \forall j \in \{1, 2, \dots, J\}$$

with variable $x_1, \dots, x_J \in R^n$, and problem data $A_1, \dots, A_J \in R^{m \times n}$, $b \in R^m$, and $\lambda > 0$. for constrained optimization given on page 11 (really p. 12) of the lecture slides for subgradient methods for constrained problems.

Let $J = 3$, $n = 100$, $m = 10$ and $\lambda = 0.5$ Generate random matrices $A_1, \dots, A_J \in R^{m \times n}$ with independent uniformly distributed entries in the interval $[0, \frac{1}{\sqrt{m}})$, and, random vectors $x_1, \dots, x_J \in 2R^n$uniformly distributed entries in the interval $[0, \frac{1}{\sqrt{n}})$, then set $b = \sum_{j=1}^{J} A_j x_j$. Plot convergence in terms of the objective $f(x_1^{(k)}, \dots x_1^{(J)})$. Try different step length schedules. Also, plot the maximal violation for the linear constraints at each step.

Solution:
The code is in the file `solution_3_1_b.m`. The code is nearly the same as one for the task 2.4, the only difference that on every step of the gradient descent we are calculating the constraint violation vector and if there are at least one violation, we replace the gradient of the minimized function with the gradient of any constraint violation found.

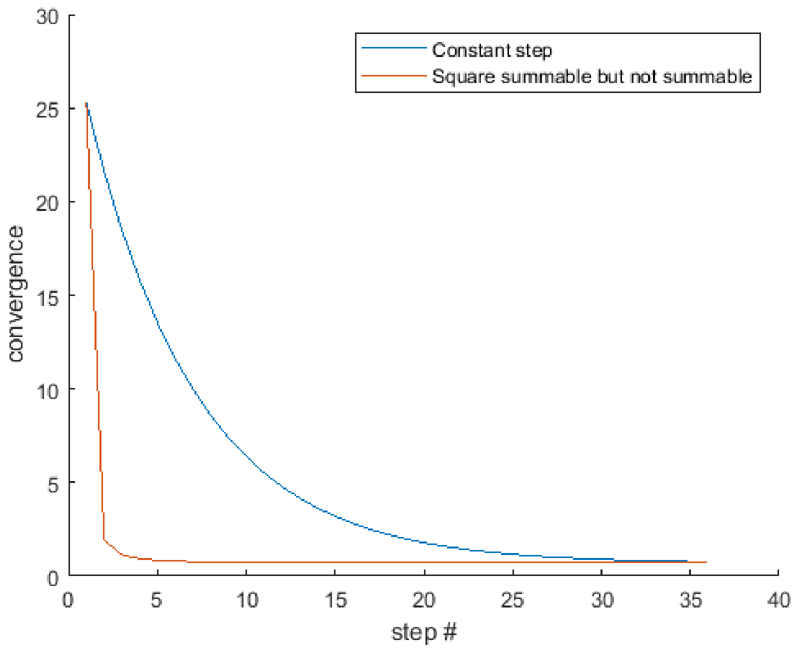

Figure 1: Convergence with different step length.

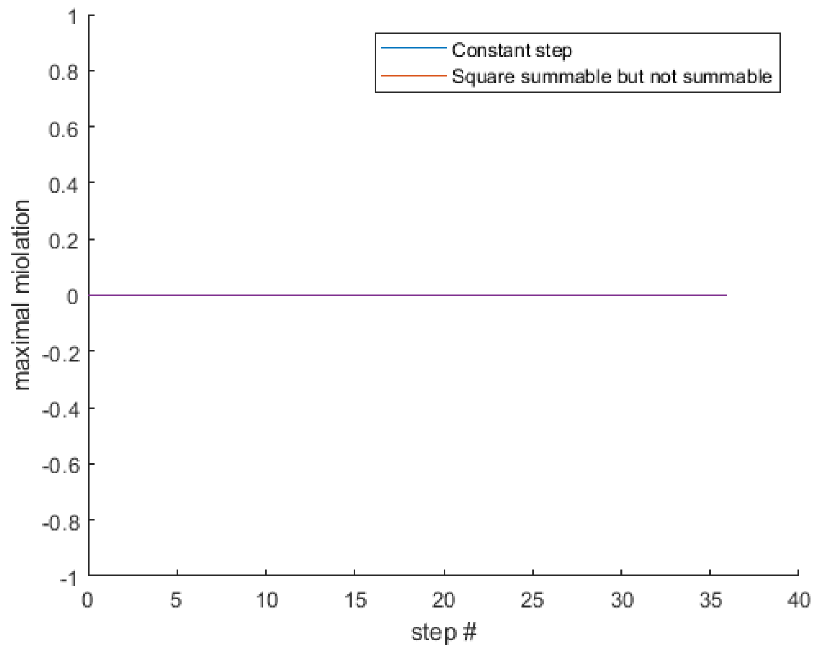

Figure 2: Maximal violation for the linear constraints.

3.2 (4 points)

A randomized least squares solver. Consider the Least Squares minimization problem

$$\begin{aligned} &\text{minimize } \frac{1}{2m}\underbrace{\sum_{i=1}^{m}(b_i - a_i^T x)^2}_{f(x)} \\ &\text{subject to } x \in R^n \end{aligned}$$

where $a_1, \ldots, a_m$ are rows of matrix A. We will consider the stochastic subgradient descent iterates

$$x^{t+1} = x^t - \alpha_t g_t$$

where g_t is a noisy unbiased gradient of the objective function, i.e., $E[g^T | x^T] \in \partial f(x^t)$.

a) (1 point) Let j will be a random index chosen from $\{1, \ldots, m\}$ such that for every index $i \in \{1, \ldots, m\}$ the probability that $j = i$ is p_i, i.e.,

$$P(i = j) = p_i,$$

for a given discrete probability distribution $p_1, \ldots, p_m \geq 0$, $\sum_{i=1}^{m} = 1$. Show that

$$E(\frac{(a_j^T x - b_j)}{mp_j} a_j) \in \partial f(x)$$

where the expectation is taken over the random variable j.

Solution:

$$f(x) = \frac{1}{2m}(b - Ax)^T(b - Ax)$$

The gradient of $f(x)$ is

$$\frac{(Ax - b)^T}{m} A$$

or, component-wize

$$e_i \frac{(a_i^T x - b_i)}{m} a_i$$

there e_i is the i-th component of the unit vector.
The expectation is

$$\begin{aligned}
&E\frac{(a_j^T x - b_j)}{mp_j} a_j = \\
&\frac{1}{m}(E(a_j^T x - b_j)a_j/p_j) = \\
&\frac{1}{m}(E(a_j^T x a_j/p_j) - E(b_j a_j/p_j))
\end{aligned}$$

For the second member of equation:

$$\begin{aligned}
E(b_j a_j/p_j) = \sum_{j=1}^{m} e_j p_j b_j a_j/p_j = \sum_{j=1}^{m} e_j b_j a_j \\
= e_j b_j a_j.
\end{aligned}$$

there e_j is the j-th component of the unit vector.
So, $E(b_j a_j/p_j) = e_j b_j a_j$.
The equality for the first member of the equation:

$$E(a_j^T x a_j/p_j) = e_j a_j^T x a_j$$

can be proved in the same way. So, we're done.