Homework 4

Total 20 points

Due November 12th at 11:59pm

Problem 1. (5 points) If the convex function f satisfies

$$\langle \nabla f(\mathbf{x}), \mathbf{x} - \mathbf{x}^* \rangle \geq \frac{\mu}{2} \|\mathbf{x} - \mathbf{x}^*\|_2^2 + \frac{1}{2L} \|\nabla f(\mathbf{x})\|_2^2, \quad \forall \mathbf{x}$$

where $\mathbf{x}^*$ is the minimizer, show that gradient descent with $\eta_t = \eta = \frac{1}{L}$ outputs

$$\|\mathbf{x}_t - \mathbf{x}^*\|_2^2 \leq \left(1 - \frac{\mu}{L}\right)^t \|\mathbf{x}_0 - \mathbf{x}^*\|_2^2 .$$

Solution.

$$\begin{aligned}
\|\mathbf{x}_{t+1} - \mathbf{x}^*\|_2^2 &= \left\| \mathbf{x}_t - \mathbf{x}^* - \frac{1}{L} \nabla f(\mathbf{x}_t) \right\|_2^2 \\
&= \|\mathbf{x}_t - \mathbf{x}^*\|_2^2 + \frac{1}{L^2} \|\nabla f(\mathbf{x}_t)\|_2^2 - \frac{2}{L} \langle \mathbf{x}_t - \mathbf{x}^*, \nabla f(\mathbf{x}_t) \rangle \\
&\leq \|\mathbf{x}_t - \mathbf{x}^*\|_2^2 - \frac{\mu}{L} \|\mathbf{x}_t - \mathbf{x}^*\|_2^2 \\
&= \left(1 - \frac{\mu}{L}\right) \|\mathbf{x}_t - \mathbf{x}^*\|_2^2
\end{aligned}$$

where the last inequality comes from the condition mentioned in the problem. Thus we have

$$\|\mathbf{x}_t - \mathbf{x}^*\|_2^2 \leq \left(1 - \frac{\mu}{L}\right)^t \|\mathbf{x}_0 - \mathbf{x}^*\|_2^2 .$$

Problem 2. (15 points) Consider the objection function of regularized logistic regression:

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

The homework ZIP file contains two text files, labeled `A.txt` and `b.txt`, that contains an $n \times d$ matrix $\mathbf{A} = [\mathbf{a}_1, \ldots, \mathbf{a}_n]^\top$ and an n-dimensional vector $\mathbf{b} = [b_1; \ldots; b_n]$, with $n = 2000$, $d = 112$, $b_i \in \{-1, +1\}$. For initial point $\mathbf{x}_0 = \mathbf{0}$, solve this problem with $\lambda = 0, 10^{-6}, 10^{-3}, 10^{-1}$ using the following algorithms:

(a) Gradient descent with backtracking line search;

(b) Gradient descent with constant stepsize.

For each setting, plot two figures: the function value $f(\mathbf{x})$ versus the iteration number, and the logarithm of gradient norm $\log \|\nabla f(\mathbf{x})\|_2$ versus the iteration number. Hand in your code and a report showing the figures and how the stepsizes are chosen. The code can be implemented in Matlab or Python. You should try your best to use matrix operations rather than vector or scalar operations.
Hint: the gradient of f is $\nabla f(\mathbf{x}) = \frac{1}{n} \sum_{i=1}^{n} -\frac{1}{1+\exp(b_i \mathbf{a}_i^\top \mathbf{x})} b_i \mathbf{a}_i + \lambda \mathbf{x}$.