Homework 8

Total 30 points

Due December 31st at 11:59pm

Problem 1. (5 points) Suppose $F(\mathbf{x}) \triangleq \mathbb{E}_\xi[f(\mathbf{x};\xi)]$ is L-smooth and μ-strongly convex, $g(\mathbf{x}_t, \xi_t)$ is an unbiased estimator of $\nabla F(\mathbf{x}_t)$, with bounded variance σ^2. Show that the stochastic gradient method with fixed step size $\eta \le 1/(2L)$ achieves

$$\mathbb{E}[F(\mathbf{x}_t) - F(\mathbf{x}^*)] \le (1-2\eta\mu)^t (F(\mathbf{x}_0) - F(\mathbf{x}^*)) + \frac{\eta\sigma^2 L}{4\mu}.$$

Problem 2. (5 points) In this problem, we study a stochastic gradient method with a projection step. Let $F : \mathbb{R}^d \to \mathbb{R}$ be differentiable and μ-strongly convex, and let $\mathcal{C}$ be a closed, convex set. Consider the projected stochastic gradient method

$$\mathbf{x}_{t+1} = \mathcal{P}_{\mathcal{C}}(\mathbf{x}_t - \eta_t G(\mathbf{x}_t)),$$

where $G(\mathbf{x}_t)$ is an unbiased estimate of $\nabla F(\mathbf{x}_t)$. Assume that the randomness in $G(\mathbf{x}_t)$ is independent of all past randomness in the algorithm. Letting $\mathbf{x}^* = \arg\min_{\mathbf{x}\in\mathcal{C}} F(\mathbf{x})$, prove that the iterates satisfy the bound

$$\mathbb{E}[\|\mathbf{x}_{t+1} - \mathbf{x}^*\|_2^2] \le (1 - 2\eta_t\mu)\mathbb{E}[\|\mathbf{x}_t - \mathbf{x}^*\|_2^2] + \eta_t^2 B^2$$

where $B^2 = \sup_{\mathbf{x}\in\mathcal{C}} \mathbb{E}\,\|G(\mathbf{x})\|_2^2$.

Problem 3. (5 points) Let $F(\mathbf{x}) = \frac{1}{n}\sum_{i=1}^n f_i(\mathbf{x})$, where $f_i(\mathbf{x})$ is differentiable and L-smooth. Suppose j is uniformly sampled from $\{1, 2, \ldots, n\}$. Show that

$$\mathbb{E}[\|\nabla f_j(\mathbf{x})\|_2^2] \le L^2 \mathbb{E}[\|\mathbf{x} - \mathbf{x}^*\|_2^2] + \mathbb{E}[\|\nabla f_j(\mathbf{x}) - \nabla F(\mathbf{x}))\|_2^2]$$

where $\mathbf{x}^*$ is a minimizer of $F(\mathbf{x})$.

Problem 4. (15 points) In this problem, you are required to use stochastic gradient method to solve the following quadratic problem:

$$f(\mathbf{x}) = \frac{1}{2n}\sum_{i=1}^n (\mathbf{a}_i^\top \mathbf{x} - b_i)^2 = \frac{1}{2n}\|\mathbf{A}\mathbf{x} - \mathbf{b}\|_2^2,$$

where $\mathbf{A} \in \mathbb{R}^{n\times d}$ and $\mathbf{b} \in \mathbb{R}^n$. The homework ZIP file contains two text files, labeled `A.txt` and `b.txt`, that contains an $n \times d$ matrix $\mathbf{A}$ and an n-dimensional vector $\mathbf{b}$, with $n = 500$, $d = 50$.

(a) (2 points) Compute the closed form of $\mathbf{x}^* = \arg\min_{\mathbf{x}} f(\mathbf{x})$.

(b) (4 points) Implement the stochastic gradient method for minimizing f with constant step size and diminishing step size.

(c) (3 points) Plot the error $\|\mathbf{x}_t - \mathbf{x}^*\|_2$ versus the iteration number, where $\mathbf{x}^*$ is computed by (a).

(d) (3 points) Now suppose that after every $T = 10$ iterations, you are allowed to evaluate the exact gradient $f(\mathbf{z})$, where $\mathbf{z}$ is the current iterate. Construct a better stochastic gradient estimate and implement it.

(e) (3 points) Plot the error $\|\mathbf{x}_t - \mathbf{x}^*\|_2$ and compare it to the naive scheme from (b).