Solution to Homework 8

Problem 1. 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

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

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

$$\begin{aligned}
\mathbb{E}[\|\nabla f_j(\mathbf{x})\|_2^2] &= \mathbb{E}[\|\nabla f_j(\mathbf{x}) - \nabla F(\mathbf{x}) + \nabla F(\mathbf{x})\|_2^2] \\
&= \mathbb{E}[\|\nabla f_j(\mathbf{x}) - \nabla F(\mathbf{x})\|_2^2] + \mathbb{E}[(\nabla f_j(\mathbf{x}) - \nabla F(\mathbf{x}))^\top \nabla F(\mathbf{x})] + \mathbb{E}[\|\nabla F(\mathbf{x})\|_2^2] \\
&= \mathbb{E}[\|\nabla f_j(\mathbf{x}) - \nabla F(\mathbf{x})\|_2^2] + \mathbb{E}[\|\nabla F(\mathbf{x}) - \nabla F(\mathbf{x}^*)\|_2^2] \\
&\leq L^2\mathbb{E}[\|\mathbf{x} - \mathbf{x}^*\|_2^2] + \mathbb{E}[\|\nabla f_j(\mathbf{x}) - \nabla F(\mathbf{x}))\|_2^2]
\end{aligned}$$

The last equation is due to $\nabla F(\mathbf{x}^*) = 0$ and $\mathbb{E}[\nabla f_j(\mathbf{x})] = \nabla F(\mathbf{x})$.

Problem 2. 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

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

where $B^2 = \sup_{\mathbf{x}\in\mathcal{C}} \mathbb{E}\,\|G(\mathbf{x})\|_2^2$.
Solution. We use non-expansiveness of the projection operator and the fact that $\mathbf{x}_t \in \mathcal{C}$ to obtain

$$\begin{aligned}
\|\mathbf{x}_{t+1} - \mathbf{x}^*\|_2^2 &= \|\mathcal{P}_{\mathcal{C}}(\mathbf{x}_t - \eta_t G(\mathbf{x}_t)) - \mathcal{P}_{\mathcal{C}}(\mathbf{x}^*)\|_2^2 \\
&\leq \|\mathbf{x}_t - \mathbf{x}^* - \eta_t G(\mathbf{x}_t)\|_2^2 \\
&= \|\mathbf{x}_t - \mathbf{x}^*\|_2^2 + \eta_t^2\,\|G(\mathbf{x}_t)\|_2^2 - 2\eta_t\langle G(\mathbf{x}_t), \mathbf{x}_t - \mathbf{x}^*\rangle \\
&\leq \|\mathbf{x}_t - \mathbf{x}^*\|_2^2 + \eta_t^2\,\|G(\mathbf{x}_t)\|_2^2 - 2\eta_t\langle G(\mathbf{x}_t) - G(\mathbf{x}^*), \mathbf{x}_t - \mathbf{x}^*\rangle
\end{aligned}$$

where the last inequality follows from optimality of $\mathbf{x}^*$. Now taking the expectations on both sides conditioned on $\mathbf{x}_t$, we have

$$\begin{aligned}\mathbb{E}_t[\|\mathbf{x}_{t+1}-\mathbf{x}^*\|_2^2] &\le \|\mathbf{x}_t-\mathbf{x}^*\|_2^2+\eta_t^2B^2-2\eta_t\langle\nabla f(\mathbf{x}_t)-\nabla f(\mathbf{x}^*),\mathbf{x}_t-\mathbf{x}^*\rangle \\ &\le (1-2\eta_t\mu)\,\|\mathbf{x}_t-\mathbf{x}^*\|_2^2+\eta_t^2B^2\end{aligned}$$

where the second line follows by μ-strong convexity of f. By taking expectation on both side, we can get the conclusion.

The original slides has a typo: "$\eta \le \frac{\theta}{t+1}$" should be "$\eta = \frac{\theta}{t+1}$". So Problem 3 does not count towards the score.

Problem 3. Prove the conclusion on page 10 of the slides.

Solution. Similar to the solution of Problem 2, we have

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

Then we prove this conclusion by induction. When $t=0$, we can get $\mathbb{E}[\|\mathbf{x}_0-\mathbf{x}^*\|_2^2] \le \frac{\alpha_\theta}{t+1}$ by the definition of α_θ. Now we suppose $\mathbb{E}[\|\mathbf{x}_t-\mathbf{x}^*\|_2^2] \le \frac{\alpha_\theta}{t+1}$. For the case of "$t+1$", we know

$$\begin{aligned}\mathbb{E}[\|\mathbf{x}_{t+1}-\mathbf{x}^*\|_2^2] &\le (1-2\eta\mu)\mathbb{E}[\|\mathbf{x}_t-\mathbf{x}^*\|_2^2]+\eta^2\sigma^2 \\ &\le \left(1-\frac{2\mu\theta}{t+1}\right)\frac{\alpha_\theta}{t+1}+\frac{\theta^2\sigma^2}{t+1} \\ &\le \left(1-\frac{2\mu\theta}{t+1}\right)\frac{\alpha_\theta}{t+1}+\frac{2\mu\theta-1}{2(t+1)^2}\alpha_\theta \\ &= \left(\frac{1}{t+1}-\frac{2\mu\theta+1}{2(t+1)^2}\right)\alpha_\theta \\ &= \left(\frac{t+2}{t+1}-\frac{(2\mu\theta+1)(t+2)}{2(t+1)^2}\right)\frac{\alpha_\theta}{t+2} \\ &\le \left(\frac{t+2}{t+1}-\frac{t+2}{(t+1)^2}\right)\frac{\alpha_\theta}{t+2} \\ &= \left(1-\frac{1}{2(t+1)^2}\right)\frac{\alpha_\theta}{t+2} \\ &\le \frac{\alpha_\theta}{t+2}\end{aligned}$$

where we use the fact that $2\mu\theta > 1$. By induction, it completes the proof.