HOMEWORK 6. IMPORTANCE SAMPLING

Kun Yuan

November 9, 2023

Attention: Turn in your homework at the beginning of our lecture on Nov. 14, 2023

1 Gradient variance in importance sampling

Given a dataset with finite samples $\{\xi_1, \cdots, \xi_N\}$, we consider solving the following finite-sum minimization problem:

$$\min_{x \in \mathbb{R}^d} \quad f(x) = \frac{1}{N} \sum_{i=1}^{N} F(x; \xi_i). \tag{1}$$

Suppose each data is sampled from distribution $\mathcal{D}_p$ in which

$$\mathbb{P}(\boldsymbol{\xi} = \xi_i) = p_i \in [0, 1], \quad \forall i \in \{1, \cdots, N\} \tag{2}$$

and $\sum_{i=1}^{N} p_i = 1$. The bold symbol $\boldsymbol{\xi}$ indicates a random variable. Prove the following results.

- Problem (1) is equivalent to

$$\min_{x \in \mathbb{R}^d} \quad \mathbb{E}_{\boldsymbol{\xi} \sim \mathcal{D}_p}[F_p(x; \boldsymbol{\xi})] \tag{3}$$

  where $F_p(x; \boldsymbol{\xi}) = \frac{1}{N p_i} F(x; \xi_i)$ if $\boldsymbol{\xi} = \xi_i$.

- Suppose $F(x, \xi_i)$ is L-smooth in terms of x for any $\xi_i \in \{\xi_1, \cdots, \xi_N\}$, it holds that

$$\mathbb{E}_{\boldsymbol{\xi} \sim \mathcal{D}_p}[\nabla_x F_p(x; \boldsymbol{\xi})] = \nabla f(x) \tag{4}$$

$$\mathbb{E}_{\boldsymbol{\xi} \sim \mathcal{D}_p}[\|\nabla_x F_p(x; \boldsymbol{\xi}) - \nabla f(x)\|^2] \le L_p^2 \|x - x^\star\|^2 + \sigma_p^2 \tag{5}$$

where $x^\star$ is a stationary point of $f(x)$ and

$$L_p^2 = \sum_{i=1}^{N} \frac{2L^2}{p_i N}, \quad \sigma_p^2 = \sum_{i=1}^{N} \frac{2}{p_i N^2} \|\nabla F(x^\star; \xi_i)\|^2 \tag{6}$$