Optimization Theory and Algorithm **October 28, 2021**

Homework 3

Lecturer:Xiangyu Chang *Scribe: Xiangyu Chang*

Edited by: Xiangyu Chang

HW 1 *We consider the following optimization problem:*

$$\min_{\mathbf{x}} \ f(\mathbf{x}) = \frac{1}{m}\sum_{i=1}^{m} f_i(\mathbf{x}).$$

And assume that f is β-smooth and α-strong convex. Using mini-bath SGD with fixed learning rate to solve it as

$$\mathbf{x}^{t+1} = \mathbf{x}^t - \frac{s}{n_b}\sum_{i_t \in D_t} \nabla f_{i_t}(\mathbf{x}^t),$$

where $D_t \subset \{1, 2, \ldots, m\}$ are drawn randomly and $|D_t| = n_b$ is the size of D_t. We further suppose that

(1) The index D_t does not depended from the previous $D_0, D_1, \ldots, D_{t-1}$.

(2) $\mathbb{E}_{i_t \in D_t}[\nabla f_{i_t}(\mathbf{x}^t)] = \nabla f(\mathbf{x}^t)$ (Unbiased Estimation).

(3) $\mathbb{E}_{i_t \in D_t}[\|\nabla f_{i_t}(\mathbf{x}^t)\|^2] = \sigma^2 + \|\nabla f(\mathbf{x}^t)\|^2$ (control the variance).

Prove

(i) $\mathbb{E}_{D_t}\|g^t\|^2 = \frac{\sigma^2}{n_b} + \|\nabla f(\mathbf{x}^t)\|^2$, *where* $\mathbf{g}^t = \frac{1}{n_b}\sum_{i \in D_t} \nabla f_i(\mathbf{x}^t)$.

(ii)

$$\mathbb{E}_{D_t}[f(\mathbf{x}^{t+1})] \le f(\mathbf{x}^t) - s\nabla f(\mathbf{x}^t)^\top \mathbb{E}_{D_t}[\mathbf{g}^t] + \frac{\beta s^2}{2}\mathbb{E}_{D_t}[\|\mathbf{g}^t\|^2].$$

(iii)

$$\mathbb{E}_{D_t}[f(\mathbf{x}^{t+1}) - f(\mathbf{x}^t)] \le -(s - \frac{\beta s^2}{2})\|\nabla f(\mathbf{x}^t)\|^2 + \frac{\beta s^2}{2n_b}\sigma^2.$$

(iv) Then

$$\mathbb{E}[f(\mathbf{x}^{t+1}) - f^*] - \frac{\beta s}{2n_b\alpha(2-\beta s)}\sigma^2 \le (1 - \alpha s(2-\beta s))\left[\mathbb{E}[f(\mathbf{x}^t) - f^*] - \frac{\beta s}{2n_b\alpha(2-\beta s)}\sigma^2\right].$$

HW 2 *Derive BCD algorithm for LASSO problem.*

HW 3 *Derive ADMM algorithm for Fused LASSO problem.*

HW 4 *Consider the following Basis Pursuit problem as:*

$$\min_{\mathbf{x}} \ \|\mathbf{x}\|_1$$
$$s.t. \ A\mathbf{x} = \mathbf{b}.$$

Derive ADMM algorithm for it.

Hit: using the indicator function of $\Omega = \{\mathbf{x} | A\mathbf{x} = \mathbf{b}\}$.

References