Note of Lemma 4 in "Stochastic Gradient Descent with Only One Projection"

Authors *

The Lemma 4 in the paper should be read as follows, which is obvious from the proof in the supplement.

Lemma 1. *For any fixed* $\mathbf{x} \in \mathcal{B}$*, define* $D_t = \sum_{t=1}^T \|\mathbf{x}_t - \mathbf{x}\|_2^2$*,* $\Lambda_T = \sum_{t=1}^T \zeta_t(\mathbf{x})$*, and* $m = 2\lceil \log_2 T \rceil$*. We have*

$$\Pr\Bigg(\underbrace{D_t \leq \frac{4}{T}, \Lambda_T \geq 4G_1\sqrt{D_T \ln \frac{m}{\delta}} + 4G_1 \ln \frac{m}{\delta}}_{\mathcal{A}_1}\Bigg) + \Pr\Bigg(\underbrace{\Lambda_T < 4G_1\sqrt{D_T \ln \frac{m}{\delta}} + 4G_1 \ln \frac{m}{\delta}}_{\mathcal{A}_2}\Bigg) \geq 1 - \delta$$

Based on the above lemma, then the proof of Theorem 2 in the papers goes as following. Conditioned on the event of $\mathcal{A}_1$, we derive an upper bound of $\sum_{t=1}^T \zeta_t(\mathbf{x}_*) \leq 4G_1$, and conditioned on the event of $\mathcal{A}_2$, we derive another upper bound of $\sum_{t=1}^T \zeta_t(\mathbf{x}_*) \leq \frac{\beta}{4} D_t + \left(\frac{16G_1^2}{\beta} + 4G_1\right) \ln \frac{m}{\delta}$. Since $\mathcal{A}_1 \cap \mathcal{A}_2 = \emptyset$, then we have

$$\begin{aligned}
&\Pr\Bigg(\underbrace{\sum_{t=1}^T \zeta_t(\mathbf{x}_*) \leq \frac{\beta}{4} D_t + \left(\frac{16G_1^2}{\beta} + 4G_1\right) \ln \frac{m}{\delta} + 4G_1}_{\mathcal{C}}\Bigg) \geq \Pr(\mathcal{C}, \mathcal{A}_1) + \Pr(\mathcal{C}, \mathcal{A}_2) \\
&= \Pr(\mathcal{C}|\mathcal{A}_1)\Pr(\mathcal{A}_1) + \Pr(\mathcal{C}|\mathcal{A}_2)\Pr(\mathcal{A}_2) \\
&= \Pr(\mathcal{A}_1) + \Pr(\mathcal{A}_2) \geq 1 - \delta
\end{aligned}$$

Then the proof follows similarly as in the paper.

*Please send email to tianbao-yang@uiowa.edu if you have questions.