Appendix A
Proof of the Theorem 1

Proof. Under the Assumption of L-smooth of loss function F, we have,

$$\begin{aligned} & F(w^{t+1}) - F(w^t) \\ & \leq \left\langle \triangledown F(w^t), w^{t+1} - w^t \right\rangle + \frac{L}{2} \left\| w^{t+1} - w^t \right\|_2^2 \\ & = \left\langle \triangledown F(w^t), -\frac{\eta}{N} \sum_{i=1}^{N} g_i^t \right\rangle + \frac{L}{2} \left\| \frac{\eta}{N} \sum_{i=1}^{N} g_i^t \right\|_2^2 . \end{aligned} \tag{54}$$

Taking the expectation at t-th iteration, we have

$$\begin{aligned} & \mathbb{E}\left[F(\boldsymbol{w}^{t+1})\right] - \mathbb{E}\left[F(\boldsymbol{w}^t)\right] \\ & \leq \underbrace{-\eta \left\langle \triangledown F(w^t), \mathbb{E}\left[\frac{1}{N} \sum_{i=1}^{N} g_i^t\right] \right\rangle}_{A1} + \underbrace{\frac{L\eta^2}{2} \mathbb{E} \left\| \frac{1}{N} \sum_{i=1}^{N} g_i^t \right\|_2^2}_{A2} . \end{aligned} \tag{55}$$

For $A1$, we have,

$$\begin{aligned} A1 := & \left\langle \triangledown F(w^t), \mathbb{E}\left[\frac{\eta}{N} \sum_{i=1}^{N} g_i^t\right] - \mathbb{E}\left[\frac{\eta}{N} \sum_{i=1}^{N} \triangledown F_i(\widetilde{\boldsymbol{w}}_i^t)\right] \right\rangle \\ & + \eta \left\langle \triangledown F(w^t), \mathbb{E}\left[\frac{1}{N} \sum_{i=1}^{N} \triangledown F_i(\widetilde{\boldsymbol{w}}_i^t)\right] \right\rangle \\ \overset{(a)}{=} & \underbrace{-\frac{\eta}{2} \mathbb{E} \left\| \triangledown F(w^t) - \mathbb{E}\left[\frac{1}{N} \sum_{i=1}^{N} \triangledown F_i(\widetilde{\boldsymbol{w}}_i^t)\right] \right\|_2^2}_{A3} \\ & + \frac{\eta}{2} \mathbb{E} \left\| \triangledown F(w^t) \right\|_2^2 + \frac{\eta}{2} \mathbb{E} \left\| \frac{1}{N} \sum_{i=1}^{N} \triangledown F_i(\widetilde{\boldsymbol{w}}_i^t) \right\|_2^2 \end{aligned} \tag{56}$$

where (a) is otain by $\mathbb{E}[g_i^t] = \triangledown F(\widetilde{\boldsymbol{w}}^t)$.

For $A3$, we have,

$$\begin{aligned} A3 & \leq \frac{\eta}{2N^2} \mathbb{E} \left\| \sum_{i=1}^{N} \triangledown F_i(\widetilde{\boldsymbol{w}}_i^t) - \triangledown F(\boldsymbol{w}^t) \right\|_2^2 \\ & \leq \frac{\eta}{2N} \sum_{i=1}^{N} \mathbb{E} \left\| \triangledown F_i(\widetilde{\boldsymbol{w}}_i^t) - \triangledown F_i(\boldsymbol{w}^t) \right\|_2^2 \\ & \leq \frac{dL^2\eta}{8N} \sum_{i=1}^{N} \delta_i^2 . \end{aligned} \tag{57}$$

For $A2$, we derive,

$$\begin{aligned} A2 = & \frac{L\eta^2}{2} \mathbb{E} \left\| \frac{1}{N} \sum_{i=1}^{N} g_i^t \right\|_2^2 \leq \frac{L\eta^2}{2N} \sum_{i=1}^{N} \mathbb{E} \left\| g_i^t \right\|_2^2 \\ = & \frac{L\eta^2}{2N} \sum_{i=1}^{N} \mathbb{E} \left\| g_i^t - \triangledown F(\boldsymbol{w}^t) + \triangledown F(\boldsymbol{w}^t) \right\|_2^2 \\ \leq & \underbrace{\frac{L\eta^2}{N} \sum_{i=1}^{N} \mathbb{E} \left\| g_i^t - \triangledown F(\boldsymbol{w}^t) \right\|_2^2}_{A4} + L\eta^2 \mathbb{E} \left\| \triangledown F(\boldsymbol{w}^t) \right\|_2^2 . \end{aligned} \tag{58}$$

For $A4$, we obtain,

$$\begin{aligned} A4 = & \frac{L\eta^2}{N} \sum_{i=1}^{N} \mathbb{E} \left\| g_i^t - \triangledown F_i(\widetilde{\boldsymbol{w}}_i^t) + \triangledown F_i(\widetilde{\boldsymbol{w}}_i^t) - \triangledown F(\boldsymbol{w}^t) \right\|_2^2 \\ \leq & \frac{2L\eta^2}{N} \sum_{i=1}^{N} \mathbb{E} \left\| g_i^t - \triangledown F_i(\widetilde{\boldsymbol{w}}_i^t) \right\|_2^2 \\ & + \frac{2L\eta^2}{N} \sum_{i=1}^{N} \mathbb{E} \left\| \triangledown F_i(\widetilde{\boldsymbol{w}}_i^t) - \triangledown F(\boldsymbol{w}^t) \right\|_2^2 \\ \leq & \frac{2L\eta^2 \sum_{i=1}^{N} \tau_i^2}{MN} + \frac{4L\eta^2}{N} \sum_{i=1}^{N} \mathbb{E} \left\| \triangledown F_i(\widetilde{\boldsymbol{w}}_i^t) - \triangledown F_i(\boldsymbol{w}^t) \right\|_2^2 \\ & + \frac{4L\eta^2}{N} \sum_{i=1}^{N} \mathbb{E} \left\| \triangledown F_i(\boldsymbol{w}^t) - \triangledown F(\boldsymbol{w}^t) \right\|_2^2 \\ \leq & \left(\frac{2L\eta^2 \sum_{i=1}^{N} \tau_i^2}{MN} + 4L\eta^2 \varphi^2 \right) + \frac{dL^3\eta^2}{N} \sum_{i=1}^{N} \delta_i^2 . \end{aligned} \tag{59}$$

If we plug (59) back into (58), we can get:

$$\begin{aligned} A2 \leq & L\eta^2 \mathbb{E} \left\| \triangledown F(\boldsymbol{w}^t) \right\|_2^2 + \frac{dL^3\eta^2}{N} \sum_{i=1}^{N} \delta_i^2 \\ & + \left(\frac{2L\eta^2 \sum_{i=1}^{N} \tau_i^2}{MN} + 4L\eta^2 \varphi^2 \right) . \end{aligned} \tag{60}$$

From (57) and (60), we have the bound for (55):

$$\begin{aligned} & \mathbb{E}\left[F(\boldsymbol{w}^{t+1})\right] - \mathbb{E}\left[F(\boldsymbol{w}^t)\right] \\ & \leq \frac{2\eta^2 L - \eta}{2} \mathbb{E} \left\| \triangledown F(\boldsymbol{w}^t) \right\|_2^2 + \frac{dL^2\eta + 8dL^3\eta^2}{8N} \sum_{i=1}^{I} \delta_i^2 \\ & \quad + \left(\frac{2L\eta^2 \sum_{i=1}^{N} \tau_i^2}{MN} + 4L\eta^2 \varphi^2 \right) \end{aligned} \tag{61}$$

If we sum up the above inequality for $t = 1, 2, \ldots, T$, we get,

$$\begin{aligned} \frac{(\eta - 2\eta^2 L)}{2} \sum_{t=0}^{T-1} \mathbb{E} \left\| \triangledown F(\boldsymbol{w}^t) \right\|_2^2 & \leq F(\boldsymbol{w}^0) - F(\boldsymbol{w}^T) + TH \\ & \leq F(\boldsymbol{w}^0) - F^\star + TH \end{aligned} \tag{62}$$

where

$$H = \frac{\eta L^2 d + 8\eta^2 L^3 d}{8N} \sum_{i=1}^{N} \delta_i^2 + \frac{2L\eta^2 \sum_{i=1}^{N} \tau_i^2}{MN} + 4L\eta^2\varphi^2. \tag{63}$$

$\square$