Appendix A PROOF OF THEOREM 1

For each party k, the update rule is

$$\theta_{t+1}^{k} = \theta_{t}^{k} - \eta_t \sum_{\mu=0}^{E_k-1} \nabla_k F(\tilde{\theta}_{t,\mu}^{k}) \tag{1}$$

where $\tilde{\theta}_{t,\mu}^{k}$ is the concatenated model parameters in which $\tilde{\theta}_{t,\mu}^{k} = (\theta_t^1, \theta_t^2, \cdots, \theta_{t,\mu}^{k}, \cdots, \theta_t^K)$, E_k is the number of local epochs in party k.

$$R_e = \frac{1}{T} \sum_{t=1}^{T} F_t\left(\theta_t\right) - F\left(\theta_*\right). \tag{2}$$

For the purpose of the analysis, we make the following assumptions.

Assumption 1. We assume the following for for any θ_1, θ_2, $f(\theta)$ is L-smooth, i.e.,

$$\left\|\nabla_j F\left(\theta_1\right) - \nabla_j F\left(\theta_2\right)\right\| \leq L_j \left\|\theta_1 - \theta_2\right\|. \tag{3}$$

Assumption 2. We assume the $F(\theta)$ is convex.

Assumption 3. (Bounded Solution Space) We assume there exists a $D > 0$, for any $\forall t$.

$$\mathrm{D}_t := \frac{1}{2}\left\|\theta_t - \theta_*\right\|_2^2 \leq \frac{1}{2} D^2. \tag{4}$$

Assumption 4. (Bounded Gradient) We assume the gradient in k-th party is bounded, i.e.,

$$\left\|\nabla_k F\left(x^k\right)\right\|_2^2 \leq R_k \cdot G. \tag{5}$$

where R_k denotes the number of features in k-th party.

Theorem 1. Under the assumption 1, 2, 3, 4, we can set the step size $\eta_t = \frac{\eta}{\sqrt{t}}$. The upper bound of the regret error R_e satisfies

$$R_e \leq \frac{\alpha \sum_{k=1}^{K} E_k^2 R_k + D^2 + \beta \sqrt{\sum_{k=1}^{K} [\sqrt{R_k} \cdot E_k \cdot (E_k - 1)]^2}}{\eta \sqrt{T} E_{\min}} \tag{6}$$

where $\alpha = \eta^2 G, \beta = \eta D L_{\max} \sqrt{G}$, $L_{\max} = \max\{L_k | k \in \mathcal{M}\}$.

Proof: Proof of proposition goes here. ∎

Before analyze the convergence of the proposed Algorithm, we first propose a Lemma in order to help proving the further results.

Lemma 1.

$$\begin{aligned}
&< \theta_t - \theta_*, \nabla F(\theta_t) > \\
&\leq \frac{\eta_t}{2E_{\min}} \sum_{k=1}^{K} \| \sum_{\mu=0}^{E_k-1} \nabla_k F(\theta_{t,\mu}^{k}) \|^2 - \frac{D_{t+1} - D_t}{\eta_t E_{min}} \\
&+ \frac{1}{E_{\min}} \sum_{k=1}^{K} < \theta_t^k - \theta_*^k, E_k \nabla_k F(\theta_t) - \sum_{\mu=0}^{E_k-1} \nabla_k F(\theta_{t,\mu}^{k}) >
\end{aligned} \tag{7}$$

$$
\begin{aligned}
&D_{t+1} - D_t \\
&= \frac{1}{2}\left(\|\theta_{t+1} - \theta_*\|^2 - \|\theta_t - \theta_*\|^2\right) && (8)\\
&= \frac{1}{2}\sum_{k=1}^{K}\left(\|\theta_{t+1}^k - \theta_*^k\|^2 - \|\theta_t^k - \theta_*^k\|^2\right) && (9)\\
&= \frac{1}{2}\sum_{k=1}^{K}\left(\|\theta_t^k - \eta_t \sum_{\mu=0}^{E_k-1}\nabla_k F(\theta_{t,\mu}^k) - \theta_*^k\|^2 - \|\theta_t^k - \theta_*^k\|^2\right) && (10)\\
&= \frac{1}{2}\sum_{k=1}^{K}\left(\|\eta_t \sum_{\mu=0}^{E_k-1}\nabla_k F(\theta_{t,\mu}^k)\|^2 - 2\eta_t\|\theta_t^k - \theta_*^k\| \cdot \|\sum_{\mu=0}^{E_k-1}\nabla_k F(\theta_{t,\mu}^k)\|\right) && (11)\\
&= \frac{1}{2}\eta_t^2\sum_{k=1}^{K}\|\sum_{\mu=0}^{E_k-1}\nabla_k F(\theta_{t,\mu}^k)\|^2 - \eta_t\sum_{k=1}^{K} < \theta_t^k - \theta_*^k, \sum_{\mu=0}^{E_k-1}\nabla_k F(\theta_{t,\mu}^k) > && (12)\\
&= \frac{1}{2}\eta_t^2\sum_{k=1}^{K}\|\sum_{\mu=0}^{E_k-1}\nabla_k F(\theta_{t,\mu}^k)\|^2 \\
&- \eta_t\sum_{k=1}^{K} < \theta_t^k - \theta_*^k, \sum_{\mu=0}^{E_k-1}\nabla_k F(\theta_{t,\mu}^k) - \sum_{\mu=0}^{E_k-1}\nabla_k F(\theta_t) + \sum_{\mu=0}^{E_k-1}\nabla_k F(\theta_t) > && (13)\\
&= \frac{1}{2}\eta_t^2\sum_{k=1}^{K}\|\sum_{\mu=0}^{E_k-1}\nabla_k F(\theta_{t,\mu}^k)\|^2 - \eta_t\sum_{k=1}^{K} < \theta_t^k - \theta_*^k, \sum_{\mu=0}^{E_k-1}\nabla_k F(\theta_t) > \\
&+ \eta_t\sum_{k=1}^{K} < \theta_t^k - \theta_*^k, \sum_{\mu=0}^{E_k-1}\nabla_k F(\theta_t) - \sum_{\mu=0}^{E_k-1}\nabla_k F(\theta_{t,\mu}^k) > && (14)\\
&= \frac{1}{2}\eta_t^2\sum_{k=1}^{K}\|\sum_{\mu=0}^{E_k-1}\nabla_k F(\theta_{t,\mu}^k)\|^2 - \eta_t\sum_{k=1}^{K} E_k < \theta_t^k - \theta_*^k, \nabla_k F(\theta_t) > \\
&+ \eta_t\sum_{k=1}^{K} < \theta_t^k - \theta_*^k, \sum_{\mu=0}^{E_k-1}\nabla_k F(\theta_t) - \sum_{\mu=0}^{E_k-1}\nabla_k F(\theta_{t,\mu}^k) > && (15)\\
&\leq \frac{1}{2}\eta_t^2\sum_{k=1}^{K}\|\sum_{\mu=0}^{E_k-1}\nabla_k F(\theta_{t,\mu}^k)\|^2 - \eta_t E_{min} < \theta_t - \theta_*, \nabla F(\theta_t) > \\
&+ \eta_t\sum_{k=1}^{K} < \theta_t^k - \theta_*^k, \sum_{\mu=0}^{E_k-1}\nabla_k F(\theta_t) - \sum_{\mu=0}^{E_k-1}\nabla_k F(\theta_{t,\mu}^k) > && (16)
\end{aligned}
$$

Dividing the above equation by $\eta_t E_{min}$ and rearranging it, we can obtain:

$$
\begin{aligned}
< \theta_t - \theta_*, \nabla F(\theta_t) > &\leq \frac{\eta_t}{2E_{min}}\sum_{k=1}^{K}\|\sum_{\mu=0}^{E_k-1}\nabla_k F(\theta_{t,\mu}^k)\|^2 - \frac{D_{t+1} - D_t}{\eta_t E_{min}} \\
&+ \frac{1}{E_{min}}\sum_{k=1}^{K} < \theta_t^k - \theta_*^k, \sum_{\mu=0}^{E_k-1}\nabla_k F(\theta_t) - \sum_{\mu=0}^{E_k-1}\nabla_k F(\theta_{t,\mu}^k) > \qquad (17)
\end{aligned}
$$

We now come to evaluate the regret R_e up to iteration T . By the definition in (2) and, we have

$$
\begin{aligned}
R_e &= \frac{1}{T}\sum_t F(\theta_t) - F(\theta_*) = \frac{1}{T}\sum_{t=1}^{T} F(\theta_t) - \frac{1}{T}\sum_{t=1}^{T} F(\theta_*) && (18)\\
&= \frac{1}{T}\sum_{t=1}^{T}\left(F(\theta_t) - F(\theta_*)\right) \leq \frac{1}{T}\sum_{t=1}^{T} < \theta_t - \theta_*, \nabla F(\theta_t) > && (19)
\end{aligned}
$$

where (19) follows from the convexity of the loss functions.

$$
\begin{aligned}
T \cdot R_e &\leq \sum_{t=1}^{T} < \theta_t - \theta_*, \nabla F(\theta_t) > \\
&= \sum_{t=1}^{T} \Big(\frac{\eta_t}{2E_{min}} \sum_{k=1}^{K} \| \sum_{\mu=0}^{E_k-1} \nabla_k F(\theta_{t,\mu}^k) \|^2 - \frac{D_{t+1} - D_t}{\eta_t E_{min}} && (20) \\
&+ \frac{1}{E_{min}} \sum_{k=1}^{K} < \theta_t^k - \theta_*^k, \sum_{\mu=0}^{E_k-1} \nabla_k F(\theta_t) - \sum_{\mu=0}^{E_k-1} \nabla_k F(\theta_{t,\mu}^k) > \Big) && (21) \\
&= \underbrace{\sum_{t=1}^{T} \frac{\eta_t}{2E_{min}} \sum_{k=1}^{K} \| \sum_{\mu=0}^{E_k-1} \nabla_k F(\theta_{t,\mu}^k) \|^2}_{T_1} && (22) \\
&\underbrace{- \sum_{t=1}^{T} \frac{D_{t+1} - D_t}{\eta_t E_{min}}}_{T_2} && (23) \\
&+ \underbrace{\sum_{t=1}^{T} \frac{1}{E_{min}} \sum_{k=1}^{K} < \theta_t^k - \theta_*^k, \sum_{\mu=0}^{E_k-1} \nabla_k F(\theta_t) - \sum_{\mu=0}^{E_k-1} \nabla_k F(\theta_{t,\mu}^k) >}_{T_3} && (24)
\end{aligned}
$$

Bound T_1:

$$
\begin{aligned}
T_1 &= \sum_{t=1}^{T} \frac{\eta_t}{2E_{min}} \sum_{k=1}^{K} \| \sum_{\mu=0}^{E_k-1} \nabla_k F(\theta_{t,\mu}^k) \|^2 && (25) \\
&\leq \sum_{t=1}^{T} \frac{\eta_t}{2E_{min}} \sum_{k=1}^{K} E_k \sum_{\mu=0}^{E_k-1} \| \nabla_k F(\theta_{t,\mu}^k) \|^2 && (26) \\
&\leq \sum_{t=1}^{T} \frac{\eta_t}{2E_{min}} \sum_{k=1}^{K} E_k^2 R_k G && (27) \\
&= \frac{G}{E_{min}} \sum_{k=1}^{K} E_k^2 R_k \sum_{t=1}^{T} \frac{1}{2} \frac{\eta}{\sqrt{t}} && (28) \\
&\leq \frac{\eta G \sqrt{T}}{E_{min}} \sum_{k=1}^{K} E_k^2 R_k && (29)
\end{aligned}
$$

where the second inequality is due to the Cauchy-Schwarz inequality, the last equality is because eq. (30).

$$
\sum_{t=a}^{b} \frac{1}{\sqrt{t}} \leq \int_{a-1}^{b} \frac{1}{\sqrt{t}} dt = 2(\sqrt{b} - \sqrt{a-1}) \tag{30}
$$

Bound T_2:

$$T_2 = -\sum_{t=1}^{T} \frac{D_{t+1} - D_t}{\eta_t E_{min}} \tag{31}$$

$$= \frac{1}{E_{min}}[\frac{D_1}{\eta_1} - \frac{D_{T+1}}{\eta_T} + \sum_{t=2}^{T} D_t(\frac{1}{\eta_t} - \frac{1}{\eta_{t-1}})] \tag{32}$$

$$\leq \frac{1}{E_{min}}[\frac{D^2}{\eta} - 0 + \sum_{t=2}^{T} \frac{D^2}{\eta}(\sqrt{t} - \sqrt{t-1})] \tag{33}$$

$$= \frac{D^2\sqrt{T}}{\eta E_{min}} \tag{34}$$

Bound T_3:

$$T_3 = \sum_{t=1}^{T} \frac{1}{E_{min}} \sum_{k=1}^{K} < \theta_t^k - \theta_*^k, \sum_{\mu=0}^{E_k-1} \nabla_k F(\theta_t) - \sum_{\mu=0}^{E_k-1} \nabla_k F(\theta_{t,\mu}^k) > \tag{35}$$

$$= \sum_{t=1}^{T} \frac{1}{E_{min}} \sum_{k=1}^{K} \sum_{\mu=0}^{E_k-1} < \theta_t^k - \theta_*^k, \nabla_k F(\theta_t) - \nabla_k F(\theta_{t,\mu}^k) > \tag{36}$$

$$\leq \sum_{t=1}^{T} \frac{1}{E_{min}} \sum_{k=1}^{K} \sum_{\mu=0}^{E_k-1} \|\theta_t^k - \theta_*^k\| \cdot L_k \|\theta_t - \theta_{t,\mu}^k\| \tag{37}$$

$$\leq \sum_{t=1}^{T} \frac{1}{E_{min}} \sum_{k=1}^{K} L_k \|\theta_t^k - \theta_*^k\| \cdot \sum_{\mu=0}^{E_k-1} \underbrace{\|\theta_t - \theta_{t,\mu}^k\|}_{T_4} \tag{38}$$

For T_4:

$$T_4 = \|\theta_t - \theta_{t,\mu}^k\| \tag{39}$$

$$= \|\eta_t \sum_{q=0}^{\mu-1} \nabla_k F(\theta_{t,q}^k)\| = \eta_t \| \sum_{q=0}^{\mu-1} \nabla_k F(\theta_{t,q}^k)\| \tag{40}$$

$$\leq \eta_t \sum_{q=0}^{\mu-1} \|\nabla_k F(\theta_{t,q}^k)\| \tag{41}$$

$$\leq \eta_t \mu \sqrt{R_k G} \tag{42}$$

Putting T_4 into T_3, we have

$$
\begin{aligned}
T_3 &\leq \sum_{t=1}^{T} \frac{1}{E_{min}} \sum_{k=1}^{K} L_k \|\theta_t^k - \theta_*^k\| \cdot \sum_{\mu=0}^{E_k-1} \|\theta_t - \theta_{t,\mu}^k\| && (43)\\
&\leq \sum_{t=1}^{T} \frac{1}{E_{min}} \sum_{k=1}^{K} L_k \|\theta_t^k - \theta_*^k\| \cdot \eta_t \sqrt{R_k G} \sum_{\mu=0}^{E_k-1} \mu && (44)\\
&= \sum_{t=1}^{T} \frac{\eta_t}{2E_{min}} \sum_{k=1}^{K} L_k \|\theta_t^k - \theta_*^k\| \cdot \sqrt{R_k G} \cdot E_k \cdot (E_k - 1) && (45)\\
&\leq \sum_{t=1}^{T} \frac{\eta_t L_{max} \sqrt{G}}{2E_{min}} \sum_{k=1}^{K} \|\theta_t^k - \theta_*^k\| \cdot \sqrt{R_k} \cdot E_k \cdot (E_k - 1) && (46)\\
&\leq \sum_{t=1}^{T} \frac{\eta_t L_{max} \sqrt{G}}{2E_{min}} \sqrt{\sum_{k=1}^{K} \|\theta_t^k - \theta_*^k\|^2 \cdot \sum_{k=1}^{K} [\sqrt{R_k} \cdot E_k \cdot (E_k - 1)]^2} && (47)\\
&= \sum_{t=1}^{T} \frac{\eta_t D L_{max} \sqrt{G}}{2E_{min}} \sqrt{\sum_{k=1}^{K} [\sqrt{R_k} \cdot E_k \cdot (E_k - 1)]^2} && (48)\\
&= \frac{D L_{max} \sqrt{G}}{2E_{min}} \sqrt{\sum_{k=1}^{K} [\sqrt{R_k} \cdot E_k \cdot (E_k - 1)]^2} \sum_{t=1}^{T} \eta_t && (49)\\
&\leq \frac{D L_{max} \sqrt{GT}}{E_{min}} \sqrt{\sum_{k=1}^{K} [\sqrt{R_k} \cdot E_k \cdot (E_k - 1)]^2} && (50)
\end{aligned}
$$

where the eq. (48) is due to

$$
\sum_{k=1}^{K} \|\theta_t^k - \theta_*^k\|^2 = \|\theta_t - \theta_*\|^2 \tag{51}
$$

Combining T_1, T_2, T_3,

$$
T \cdot R_e \leq \frac{\eta G \sqrt{T}}{E_{min}} \sum_{k=1}^{K} E_k^2 R_k + \frac{D^2 \sqrt{T}}{\eta E_{min}} + \frac{D L_{max} \sqrt{GT}}{E_{min}} \sqrt{\sum_{k=1}^{K} [\sqrt{R_k} \cdot E_k \cdot (E_k - 1)]^2} \tag{52}
$$

So,

$$
\begin{aligned}
R_e &\leq \frac{\eta G}{\sqrt{T} E_{min}} \sum_{k=1}^{K} E_k^2 R_k + \frac{D^2}{\eta \sqrt{T} E_{min}} + \frac{D L_{max} \sqrt{G}}{\sqrt{T} E_{min}} \sqrt{\sum_{k=1}^{K} [\sqrt{R_k} \cdot E_k \cdot (E_k - 1)]^2} && (53)\\
&= \frac{\eta^2 G \sum_{k=1}^{K} E_k^2 R_k + D^2 + \eta D L_{max} \sqrt{G} \sqrt{\sum_{k=1}^{K} [\sqrt{R_k} \cdot E_k \cdot (E_k - 1)]^2}}{\eta \sqrt{T} E_{\min}} && (54)
\end{aligned}
$$

Define $\alpha = \eta^2 G, \beta = \eta D L_{\max} \sqrt{G}$, then,

$$
R_e \leq \frac{\alpha \sum_{k=1}^{K} E_k^2 R_k + D^2 + \beta \sqrt{\sum_{k=1}^{K} [\sqrt{R_k} \cdot E_k \cdot (E_k - 1)]^2}}{\eta \sqrt{T} E_{\min}} \tag{55}
$$