Communication-Efficient Federated Learning with Adaptive Aggregation for Heterogeneous Client-Edge-Cloud Network - Proof of Theorem 1

Long Luo, Chi Zhang, Hongfang Yu, Gang Sun and Shouxi Luo
University of Electronic Science and Technology of China, Southwest Jiaotong University.

Assumption 1 (Lipschitz Continuous Gradient). *There exists a constant $L > 0$, such that:*

$$\|\nabla F(\boldsymbol{\omega}_1) - \nabla F(\boldsymbol{\omega}_2)\| \leq L\|\boldsymbol{\omega}_1 - \boldsymbol{\omega}_2\|,\ \forall \boldsymbol{\omega}_1, \boldsymbol{\omega}_2 \tag{1}$$

Assumption 2 (Unbiased Estimated Gradient). *Let ξ be a random sample from local dataset $\mathcal{D}$, then the estimated local gradient is unbiased:*

$$\mathbb{E}\left[\nabla F(\boldsymbol{\omega}; \xi)\right] = \nabla F(\boldsymbol{\omega}) \tag{2}$$

Assumption 3 (Bounded Variance). *There exists a constant σ, such that the variance of the estimated local gradient can be bounded by:*

$$\mathbb{E}\left[\|\nabla F(\boldsymbol{\omega}; \xi) - \nabla F(\boldsymbol{\omega})\|^2\right] \leq \sigma^2 \tag{3}$$

Before stating the proof of Theorem 1 of FedAda, we first mention the following three lemmas.

Lemma 1. *According to Assumption 1, the expected inner product between full batch gradient and stochastic gradient can be bounded with:*

$$\begin{aligned}
&\mathbb{E}\left\langle \nabla f(\boldsymbol{\omega}^r), g^r \right\rangle \\
&\leq \frac{L^2}{2N} \sum_{i=1}^{M} \alpha_i^r \sum_{s=1}^{q} \sum_{j \in \mathcal{A}_i} \alpha_j^r \sum_{t=1}^{p} \|\boldsymbol{\omega}^r - \boldsymbol{\omega}_j^{t,s,r}\|^2 \\
&- \frac{1}{2N} \sum_{i=1}^{M} \alpha_i^r \sum_{s=1}^{q} \sum_{j \in \mathcal{A}_i} \alpha_j^r \sum_{t=1}^{p} \|\nabla f(\boldsymbol{\omega}^r)^2\|^2 \\
&- \frac{1}{2N} \sum_{i=1}^{M} \alpha_i^r \sum_{s=1}^{q} \sum_{j \in \mathcal{A}_i} \alpha_j^r \sum_{t=1}^{p} \|\nabla F_j\left(\boldsymbol{\omega}_j^{t,s,r}\right)\|^2
\end{aligned} \tag{4}$$

Proof.

$$\begin{aligned}
&\mathbb{E}\left\langle \nabla f(\boldsymbol{\omega}^r), g^r \right\rangle \\
&= \frac{1}{N} \sum_{i=1}^{M} \alpha_i^r \sum_{s=1}^{q} \sum_{j \in \mathcal{A}_i} \alpha_j^r \sum_{t=1}^{p} \left\langle \nabla f(\boldsymbol{\omega}^r), \nabla F_j\left(\boldsymbol{\omega}_j^{t,s,r}\right) \right\rangle
\end{aligned} \tag{5}$$

$$\overset{①}{=} \frac{1}{2N} \sum_{i=1}^{M} \alpha_i^r \sum_{s=1}^{q} \sum_{j \in \mathcal{A}_i} \alpha_j^r \sum_{t=1}^{p} \left[-\|\nabla f(\boldsymbol{\omega}^r)\|^2 - \|\nabla F_j\left(\boldsymbol{\omega}_j^{t,s,r}\right)\|^2 + \|\nabla f(\boldsymbol{\omega}^r) - \nabla F_j\left(\boldsymbol{\omega}_j^{t,s,r}\right)\|^2 \right]$$

$$\overset{②}{\leq} \frac{1}{2N} \sum_{i=1}^{M} \alpha_i^r \sum_{s=1}^{q} \sum_{j \in \mathcal{A}_i} \alpha_j^r \sum_{t=1}^{p} \left[-\|\nabla f(\boldsymbol{\omega}^r)\|^2 - \|\nabla F_j\left(\boldsymbol{\omega}_j^{t,s,r}\right)\|^2 + L^2 \|\boldsymbol{\omega}^r - \boldsymbol{\omega}_j^{t,s,r}\|^2 \right]$$

where $g^r = \frac{1}{N} \sum_{i=1}^{M} \alpha_i^r \sum_{s=1}^{q} \sum_{j \in \mathcal{A}_i} \alpha_j^r \sum_{t=1}^{p} \nabla F_j\left(\boldsymbol{\omega}_j^{t,s,r}\right)$, ① is due to $2\langle \boldsymbol{x}, \boldsymbol{y} \rangle = \|\boldsymbol{x}\|^2 + \|\boldsymbol{y}\|^2 - \|\boldsymbol{x} - \boldsymbol{y}\|^2$, and ② based on Assumption 1. □

Lemma 2. *Under Assumption 2 and Assumption 3, we have the following bound:*

$$\mathbb{E}\|g^r\|^2 \leq \frac{1}{N^2} \sum_{i=1}^{M} (\alpha_i^r)^2 \sum_{s=1}^{q} \sum_{j \in \mathcal{A}_i} \left(\alpha_j^r\right)^2 \sum_{t=1}^{p} \sigma^2 + \frac{p}{N^2} \sum_{i=1}^{M} (\alpha_i^r)^2 \sum_{s=1}^{q} \sum_{j \in \mathcal{A}_i} \left(\alpha_j^r\right)^2 \sum_{t=1}^{p} \|\nabla F_j\left(\boldsymbol{\omega}_j^{t,s,r}\right)\|^2 \quad (6)$$

Proof. Let $\sum_{t=1}^{p} \nabla F_j\left(\boldsymbol{\omega}_j^{t,s,r}\right)$ be $\tilde{F}_j$. Then we have:

$$\mathbb{E}\|\frac{1}{N} \sum_{i=1}^{M} \alpha_i^r \sum_{s=1}^{q} \sum_{j \in \mathcal{A}_i} \alpha_j^r \sum_{t=1}^{p} \nabla F_j\left(\boldsymbol{\omega}_j^{t,s,r}\right)\|^2 = \mathbb{E}\|\frac{1}{N} \sum_{i=1}^{M} \alpha_i^r \sum_{s=1}^{q} \sum_{j \in \mathcal{A}_i} \alpha_j^r \nabla \tilde{F}_j\|^2$$

$$\overset{①}{=} Var\left(\frac{1}{N} \sum_{i=1}^{M} \alpha_i^r \sum_{s=1}^{q} \sum_{j \in \mathcal{A}_i} \alpha_j^r \nabla \tilde{F}_j\right) + \|\frac{1}{N} \sum_{i=1}^{M} \alpha_i^r \sum_{s=1}^{q} \sum_{j \in \mathcal{A}_i} \alpha_j^r \nabla \tilde{F}_j\|^2$$

$$\leq \frac{1}{N^2} \left(\alpha_i^r \alpha_j^r\right)^2 \sum_{i=1}^{M} \sum_{s=1}^{q} \sum_{j \in \mathcal{A}_i} Var\left(\nabla \tilde{F}_j\right) + \frac{1}{N^2} \left(\alpha_i^r \alpha_j^r\right)^2 \sum_{i=1}^{M} \sum_{s=1}^{q} \sum_{j \in \mathcal{A}_i} \nabla \tilde{F}_j \quad (7)$$

$$\overset{②}{\leq} \frac{1}{N^2} \left(\alpha_i^r \alpha_j^r\right)^2 \sum_{i=1}^{M} \sum_{s=1}^{q} \sum_{j \in \mathcal{A}_i} p\sigma^2 + \frac{1}{N^2} \left(\alpha_i^r \alpha_j^r\right)^2 \sum_{i=1}^{M} \sum_{s=1}^{q} \sum_{j \in \mathcal{A}_i} \nabla \tilde{F}_j$$

$$\overset{③}{\leq} \frac{1}{N^2} \sum_{i=1}^{M} (\alpha_i^r)^2 \sum_{s=1}^{q} \sum_{j \in \mathcal{A}_i} \left(\alpha_j^r\right)^2 \sum_{t=1}^{p} \left[p\|\nabla F_j\left(\boldsymbol{\omega}_j^{t,s,r}\right)\|^2 + \sigma^2\right]$$

where ① is due to $\mathbb{E}\left[\|x\|^2\right] = Var(x) + \|\mathbb{E}[x]\|^2$ and follows Assumption 2, ② follows Assumption 3 and has $Var\left(\nabla \tilde{F}_j\right) = Var\left(\sum_{t=1}^{p} \nabla F_j\left(\boldsymbol{\omega}_j^{t,s,r}\right)\right) = \sum_{t=1}^{p} Var\left(\nabla F_j\left(\boldsymbol{\omega}_j^{t,s,r}\right)\right) \leq p\sigma^2$, and ③ is because $\|\nabla \tilde{F}_j\|^2 = \|\sum_{t=1}^{p} \nabla F_j\left(\boldsymbol{\omega}_j^{t,s,r}\right)\|^2 \leq p \sum_{t=1}^{p} \|\nabla F_j\left(\boldsymbol{\omega}_j^{t,s,r}\right)\|^2$. □

Lemma 3. *Under Assumptions 3 and according to [1], we have the following bound:*

$$\mathbb{E}\|\boldsymbol{\omega}^r - \boldsymbol{\omega}_j^{t,s,r}\|^2 \leq \eta^2 \left[\left(\frac{q-1}{Nq}\right)^2 + \frac{(p-1)^2}{Np^2q^2}\right] \sum_{i=1}^{M} \sum_{s=1}^{q} \sum_{j \in \mathcal{A}_i} \sum_{t=1}^{p} \left[p\|\nabla F_j\left(\boldsymbol{\omega}_j^{t,s,r}\right)\|^2 + \sigma^2\right] \quad (8)$$

Proof. According to [1], we have the following relationship:

$$\frac{1}{N} \sum_{j=1}^{N} \|\nabla F_j(\boldsymbol{\omega}) - \nabla f(\boldsymbol{\omega})\|^2 = \sum_{i=1}^{M} \frac{N_i}{N} \|\nabla f_i(\boldsymbol{\omega}) - \nabla f(\boldsymbol{\omega})\|^2 + \sum_{i=1}^{M} \frac{1}{N} \sum_{j \in \mathcal{A}_i} \|\nabla F_j(\boldsymbol{\omega}) - \nabla f_i(\boldsymbol{\omega})\|^2 \quad (9)$$

Under the Law of Large Numbers, we have $\mathbb{E}\left[\boldsymbol{\omega}^r - \eta \nabla f(\boldsymbol{\omega}^r)\right] = \mathbb{E}\left[\boldsymbol{\omega}_j^{t,s,r} - \eta \nabla f\left(\boldsymbol{\omega}_j^{t,s,r}\right)\right]$, so that we have $\|\boldsymbol{\omega}^r - \boldsymbol{\omega}_j^{t,s,r}\|^2 = \eta^2 \|\nabla F_j(\boldsymbol{\omega}) - \nabla f(\boldsymbol{\omega})\|^2$, and then we can get:

$$\frac{1}{N}\sum_{j=1}^{N}\|\boldsymbol{\omega}^r - \boldsymbol{\omega}_j^{t,s,r}\|^2 = \eta^2 \sum_{i=1}^{M}\frac{N_i}{N}\|\nabla f_i(\boldsymbol{\omega}) - \nabla f(\boldsymbol{\omega})\|^2 + \eta^2 \sum_{i=1}^{M}\frac{1}{N}\sum_{j\in\mathcal{A}_i}\|\nabla F_j(\boldsymbol{\omega}) - \nabla f_i(\boldsymbol{\omega})\|^2 \tag{10}$$

Thus, we can prove the Lemma 3 by

$$\begin{aligned}
(1)\quad & \sum_{i=1}^{M}\frac{N_i}{N}\|\nabla f_i(\boldsymbol{\omega}) - \nabla f(\boldsymbol{\omega})\|^2 \\
&= \sum_{i=1}^{M}\frac{N_i}{N}\|\frac{1}{N_i}\sum_{j\in\mathcal{A}_i}\sum_{t=1}^{p}\nabla F_j\left(\boldsymbol{\omega}_j^{t,s,r}\right) - \frac{1}{N}\sum_{i=1}^{M}\sum_{s=1}^{q}\sum_{j\in\mathcal{A}_i}\sum_{t=1}^{p}\nabla F_j\left(\boldsymbol{\omega}_j^{t,s,r}\right)\|^2 \\
&= \sum_{i=1}^{M}\frac{N_i}{N}\|\frac{1}{Nq}\sum_{i=1}^{M}\sum_{s=1}^{q}\sum_{j\in\mathcal{A}_i}\sum_{t=1}^{p}\nabla F_j\left(\boldsymbol{\omega}_j^{t,s,r}\right) - \frac{1}{N}\sum_{i=1}^{M}\sum_{s=1}^{q}\sum_{j\in\mathcal{A}_i}\sum_{t=1}^{p}\nabla F_j\left(\boldsymbol{\omega}_j^{t,s,r}\right)\|^2 \\
&= \sum_{i=1}^{M}\frac{N_i}{N}\|\frac{q-1}{Nq}\sum_{i=1}^{M}\sum_{s=1}^{q}\sum_{j\in\mathcal{A}_i}\sum_{t=1}^{p}\nabla F_j\left(\boldsymbol{\omega}_j^{t,s,r}\right)\|^2 \\
&\overset{①}{\leq} \sum_{i=1}^{M}\frac{N_i}{N}\left\{\left(\frac{q-1}{Nq}\right)^2\sum_{i=1}^{M}\sum_{s=1}^{q}\sum_{j\in\mathcal{A}_i}\left[Var\left(\nabla\tilde{F}_j\right) + \|\nabla\tilde{F}_j\|^2\right]\right\} \\
&\overset{②}{\leq} \left(\frac{q-1}{Nq}\right)^2\sum_{i=1}^{M}\sum_{s=1}^{q}\sum_{j\in\mathcal{A}_i}\sum_{t=1}^{p}\left[p\|\nabla F_j\left(\boldsymbol{\omega}_j^{t,s,r}\right)\|^2 + \sigma^2\right]
\end{aligned} \tag{11}$$

$$\begin{aligned}
(2)\quad & \sum_{i=1}^{M}\frac{1}{N}\sum_{j\in\mathcal{A}_i}\|\nabla F_j(\boldsymbol{\omega}) - \nabla f_i(\boldsymbol{\omega})\|^2 \\
&= \sum_{i=1}^{M}\frac{1}{N}\sum_{j\in\mathcal{A}_i}\|\nabla F_j\left(\boldsymbol{\omega}_j^{t,s,r}\right) - \frac{1}{N_i}\sum_{j\in\mathcal{A}_i}\sum_{t=1}^{p}\nabla F_j\left(\boldsymbol{\omega}_j^{t,s,r}\right)\|^2 \\
&= \sum_{i=1}^{M}\frac{1}{N}\sum_{j\in\mathcal{A}_i}\|\frac{1}{Npq}\sum_{i=1}^{M}\sum_{s=1}^{q}\sum_{j\in\mathcal{A}_i}\sum_{t=1}^{p}\nabla F_j\left(\boldsymbol{\omega}_j^{t,s,r}\right) - \frac{1}{Nq}\sum_{i=1}^{M}\sum_{s=1}^{q}\sum_{j\in\mathcal{A}_i}\sum_{t=1}^{p}\nabla F_j\left(\boldsymbol{\omega}_j^{t,s,r}\right)\|^2 \\
&= \sum_{i=1}^{M}\frac{1}{N}\sum_{j\in\mathcal{A}_i}\|\frac{p-1}{Npq}\sum_{i=1}^{M}\sum_{s=1}^{q}\sum_{j\in\mathcal{A}_i}\sum_{t=1}^{p}\nabla F_j\left(\boldsymbol{\omega}_j^{t,s,r}\right)\|^2 \\
&\overset{①}{\leq} \sum_{i=1}^{M}\frac{1}{N}\sum_{j\in\mathcal{A}_i}\left(\frac{p-1}{Npq}\right)^2\sum_{i=1}^{M}\sum_{s=1}^{q}\sum_{j\in\mathcal{A}_i}\left[Var\left(\nabla\tilde{F}_j\right) + \|\nabla\tilde{F}_j\|^2\right] \\
&\overset{②}{\leq} N\left(\frac{p-1}{Npq}\right)^2\sum_{i=1}^{M}\sum_{s=1}^{q}\sum_{j\in\mathcal{A}_i}\sum_{t=1}^{p}\left[p\|\nabla F_j\left(\boldsymbol{\omega}_j^{t,s,r}\right)\|^2 + \sigma^2\right]
\end{aligned}$$

Combining (1) and (2), we can obtain:

$$\begin{aligned}
&\frac{1}{N}\sum_{j=1}^{N}\|\boldsymbol{\omega}^{r}-\boldsymbol{\omega}_{j}^{t,s,r}\|^{2}=\eta^{2}\sum_{i=1}^{M}\frac{N_i}{N}\|\nabla f_i\left(\boldsymbol{\omega}\right)-\nabla f\left(\boldsymbol{\omega}\right)\|^{2}+\eta^{2}\sum_{i=1}^{M}\frac{1}{N}\sum_{j\in\mathcal{A}_i}\|\nabla F_j\left(\boldsymbol{\omega}\right)-\nabla f_i\left(\boldsymbol{\omega}\right)\|^{2}\\
&\leq\eta^{2}\left[\left(\frac{q-1}{Nq}\right)^{2}+\frac{(p-1)^{2}}{Np^{2}q^{2}}\right]\sum_{i=1}^{M}\sum_{s=1}^{q}\sum_{j\in\mathcal{A}_i}\sum_{t=1}^{p}\left[p\|\nabla F_j\left(\boldsymbol{\omega}_j^{t,s,r}\right)\|^{2}+\sigma^{2}\right]
\end{aligned}\tag{12}$$

where ① and ② refer to the Lemma 2. □

Now, based on the above assumptions and lemmas, let's prove Theorem 1. First, according to Assumption 1 and the SGD model update formula we can obtain:

$$\begin{aligned}
\mathbb{E}\left\{f\left(\boldsymbol{\omega}^{r+1}\right)\right\}&=\mathbb{E}\left\{f\left(\boldsymbol{\omega}^{r}-\eta\frac{1}{N}\sum_{i=1}^{M}\alpha_i^{r}\sum_{s=1}^{q_i^r}\sum_{j\in\mathcal{A}_i}\alpha_j^{r}\sum_{t=1}^{p_j^r}\nabla F_j\left(\boldsymbol{\omega}_j^{t,s,r}\right)\right)\right\}\\
&\leq\mathbb{E}\left\{f\left(\boldsymbol{\omega}^{r}-\eta\frac{1}{N}\sum_{i=1}^{M}\alpha_1\sum_{s=1}^{q}\sum_{j\in\mathcal{A}_i}\alpha_2\sum_{t=1}^{p}\nabla F_j\left(\boldsymbol{\omega}_j^{t,s,r}\right)\right)\right\}\\
&\leq\mathbb{E}\left\{f\left(\boldsymbol{\omega}^{r}\right)\right\}-\eta\mathbb{E}\left\{\left\langle\nabla f\left(\boldsymbol{\omega}^{r}\right),\frac{1}{N}\sum_{i=1}^{M}\alpha_1\sum_{s=1}^{q}\sum_{j\in\mathcal{A}_i}\alpha_2\sum_{t=1}^{p}\nabla F_j\left(\boldsymbol{\omega}_j^{t,s,r}\right)\right\rangle\right\}\\
&+\frac{\eta^{2}L}{2}\mathbb{E}\|\frac{1}{N}\sum_{i=1}^{M}\alpha_1\sum_{s=1}^{q}\sum_{j\in\mathcal{A}_i}\alpha_2\sum_{t=1}^{p}\nabla F_j\left(\boldsymbol{\omega}_j^{t,s,r}\right)\|^{2}
\end{aligned}\tag{13}$$

where $p=\max\left\{p_j^r\right\}$, $q=\max\left\{q_i^r\right\}$, $\alpha_1=\max\left\{\alpha_j^r\right\}$, $\alpha_2=\max\left\{\alpha_i^r\right\}$.

Then, based on the Lemma 1, Lemma 2, Lemma 3 and (13), we can obtain:

$$\begin{aligned}
&\mathbb{E}\left\{f\left(\boldsymbol{\omega}^{r+1}\right)-f\left(\boldsymbol{\omega}^{r}\right)\right\}\\
&\leq\frac{\eta}{2N}\sum_{i=1}^{M}\alpha_1\sum_{s=1}^{q}\sum_{j\in\mathcal{A}_i}\alpha_2\sum_{t=1}^{p}\left[-\|\nabla f\left(\boldsymbol{\omega}^{r}\right)\|^{2}-\|\nabla F_j\left(\boldsymbol{\omega}_j^{t,s,r}\right)\|^{2}+L^{2}\|\boldsymbol{\omega}^{r}-\boldsymbol{\omega}_j^{t,s,r}\|^{2}\right]\\
&+\frac{\eta^{2}L}{2}\frac{1}{N^{2}}\sum_{i=1}^{M}\alpha_1^{2}\sum_{s=1}^{q}\sum_{j\in\mathcal{A}_i}\alpha_2^{2}\sum_{t=1}^{p}\left[p\|\nabla F_j\left(\boldsymbol{\omega}_j^{t,s,r}\right)\|^{2}+\sigma^{2}\right]\\
&\leq\frac{\eta}{2N}\sum_{i=1}^{M}\alpha_1\sum_{s=1}^{q}\sum_{j\in\mathcal{A}_i}\alpha_2\sum_{t=1}^{p}\left[-\|\nabla f\left(\boldsymbol{\omega}^{r}\right)\|^{2}-\|\nabla F_j\left(\boldsymbol{\omega}_j^{t,s,r}\right)\|^{2}\right]\\
&+\frac{L^{2}\eta^{3}}{2N}\frac{\left[p^{2}\left(q-1\right)^{2}+N\left(p-1\right)^{2}\right]}{Npq}\alpha_1\alpha_2\sum_{i=1}^{M}\sum_{s=1}^{q}\sum_{j\in\mathcal{A}_i}\sum_{t=1}^{p}\left[p\|\nabla F_j\left(\boldsymbol{\omega}_j^{t,s,r}\right)\|^{2}+\sigma^{2}\right]\\
&+\frac{L\eta^{2}}{2N^{2}}\alpha_1^{2}\alpha_2^{2}\sum_{i=1}^{M}\sum_{s=1}^{q}\sum_{j\in\mathcal{A}_i}\sum_{t=1}^{p}\left[p\|\nabla F_j\left(\boldsymbol{\omega}_j^{t,s,r}\right)\|^{2}+\sigma^{2}\right]\\
&=-\frac{\eta\alpha_1\alpha_2pq}{2}\|\nabla f\left(\boldsymbol{\omega}^{r}\right)\|^{2}+\frac{L^{2}\eta^{3}\alpha_1\alpha_2\left[p^{2}\left(q-1\right)^{2}+N\left(p-1\right)^{2}\right]}{2N}\sigma^{2}+\frac{L\eta^{2}\alpha_1^{2}\alpha_2^{2}pq}{2N}\sigma^{2}
\end{aligned}\tag{14}$$

$$
-\left\{\frac{\eta\alpha_1\alpha_2}{2N}-\frac{L^2\eta^3\alpha_1\alpha_2\left[p^2\left(q-1\right)^2+N\left(p-1\right)^2\right]}{2N^2q}-\frac{L\eta^2\alpha_1^2\alpha_2^2p}{2N^2}\right\}\sum_{i=1}^{M}\sum_{s=1}^{q}\sum_{j\in\mathcal{A}_i}\sum_{t=1}^{p}\|\nabla F_j\left(\boldsymbol{\omega}_j^{t,s,r}\right)\|^2
$$

$$
=-\frac{\eta\alpha_1\alpha_2pq}{2}\|\nabla f\left(\boldsymbol{\omega}^r\right)\|^2+\frac{L\eta^2\alpha_1\alpha_2}{2N}\left\{L\eta\left[p^2\left(q-1\right)^2+N\left(p-1\right)^2\right]+\alpha_1\alpha_2pq\right\}\sigma^2
$$

$$
-\frac{\eta\alpha_1\alpha_2}{2N}\left\{1-\frac{L^2\eta^2\left[p^2\left(q-1\right)^2+N\left(p-1\right)^2\right]}{Nq}-\frac{L\eta\alpha_1\alpha_2p}{N}\right\}\sum_{i=1}^{M}\sum_{s=1}^{q}\sum_{j\in\mathcal{A}_i}\sum_{t=1}^{p}\|\nabla F_j\left(\boldsymbol{\omega}_j^{t,s,r}\right)\|^2
$$

$$
\leq-\frac{\eta\alpha_1\alpha_2pq}{2}\|\nabla f\left(\boldsymbol{\omega}^r\right)\|^2+\frac{L\eta^2\alpha_1\alpha_2}{2N}\left\{L\eta\left[p^2\left(q-1\right)^2+N\left(p-1\right)^2\right]+\alpha_1\alpha_2pq\right\}\sigma^2
$$

where p, q, α_1, α_2 and η satisfy:

$$
\frac{L^2\eta^2\left[p^2\left(q-1\right)^2+N\left(p-1\right)^2\right]}{Nq}-\frac{L\eta\alpha_1\alpha_2p}{N}\leq 1 \tag{15}
$$

Next, based on (14), we can obtain:

$$
\begin{aligned}
&\frac{1}{R}\sum_{r=1}^{R}\|\nabla f\left(\boldsymbol{\omega}^r\right)\|^2\\
&\leq\frac{2\left[f\left(\boldsymbol{\omega}^0\right)-f\left(\boldsymbol{\omega}^*\right)\right]}{\eta\alpha_1\alpha_2pqR}+\frac{L\eta}{Npq}\left\{L\eta\left[p^2\left(q-1\right)^2+N\left(p-1\right)^2\right]+\alpha_1\alpha_2pq\right\}\sigma^2\\
&\leq\frac{2\left[f\left(\boldsymbol{\omega}^0\right)-f\left(\boldsymbol{\omega}^*\right)\right]}{\eta\alpha_1\alpha_2pqR}+\frac{L\eta}{Npq}\left\{L\eta\left[p^2q^2+Np^2\right]+\alpha_1\alpha_2pq\right\}\sigma^2\\
&\leq\frac{2\left[f\left(\boldsymbol{\omega}^0\right)-f\left(\boldsymbol{\omega}^*\right)\right]}{\eta\alpha_1\alpha_2pqR}+\frac{L\eta\alpha_1\alpha_2}{N}\sigma^2+\frac{L^2\eta^2pq}{N}\sigma^2+\frac{L^2\eta^2p}{q}\sigma^2
\end{aligned}
\tag{16}
$$

Thus, the proof is completed.

References

[1] Jiayi Wang, Shiqiang Wang, Rong-Rong Chen, et al. "Demystifying why local aggregation helps: Convergence analysis of hierarchical SGD". In: *AAAI*. Vol. 36. 8. 2022, pp. 8548–8556.