APPENDIX

A. *Proof of Theorem 1*

We first define three virtual sequences of global models $\{\overline{\boldsymbol{w}}_{t+1},\overline{\boldsymbol{x}}_{t+1},\overline{\boldsymbol{p}}_{t+1}\}$, $\overline{\boldsymbol{w}}_{t+1}=\overline{\boldsymbol{x}}_{t+1}+\overline{\boldsymbol{p}}_{t+1}$, and $\overline{\boldsymbol{p}}_{t+1}=\gamma(\overline{\boldsymbol{x}}_{t+1}-\overline{\boldsymbol{x}}_{t})$ for all $t=1,2,...,T$ computed by

$$\begin{aligned}&\overline{\boldsymbol{w}}_0=\boldsymbol{w}_{t_0}=\overline{\boldsymbol{x}}_0=\boldsymbol{x}_{t_0},\overline{\boldsymbol{x}}_1=\boldsymbol{x}_{t_1},\overline{\boldsymbol{p}}_1=\boldsymbol{p}_{t_1}\\&\overline{\boldsymbol{w}}_t=\boldsymbol{w}_{t_0}-\eta_t\sum_{m=1}^{M}d_m\sum_{j=0}^{t-1}\nabla F_m(\boldsymbol{x}_j^m,\boldsymbol{\xi}_j^m)\end{aligned}\tag{1}$$

Then we define two vectors $\overline{\boldsymbol{g}}_t=\sum_{m=1}^{M}d_m\nabla F_m(\boldsymbol{x}_t^m)$ and $\boldsymbol{g}_t=\sum_{m=1}^{M}d_m\nabla F_m(\boldsymbol{x}_t^m,\boldsymbol{\xi}_t^m)$. Given the above definitions, we have $\mathbb{E}[\boldsymbol{g}_t]=\overline{\boldsymbol{g}}_t$. Consequently, we have $\overline{\boldsymbol{w}}_{t+1}=\overline{\boldsymbol{w}}_t-\eta\boldsymbol{g}_t$. Next, we give the bound of $\|\overline{\boldsymbol{w}}_{t+1}-\boldsymbol{x}_*\|^2$:

$$\begin{aligned}&\|\overline{\boldsymbol{w}}_{t+1}-\boldsymbol{x}_*\|^2=\|\overline{\boldsymbol{w}}_t-\eta_t\boldsymbol{g}_t-\boldsymbol{x}_*\|^2=\|\overline{\boldsymbol{w}}_t-\boldsymbol{x}_*-\eta_t\overline{\boldsymbol{g}}_t+\eta_t\overline{\boldsymbol{g}}_t-\eta\boldsymbol{g}_t\|^2=\underbrace{\|\overline{\boldsymbol{w}}_t-\boldsymbol{x}_*-\eta_t\overline{\boldsymbol{g}}_t\|^2}_{\text{A1}}+\underbrace{\eta_t^2\|\overline{\boldsymbol{g}}_t-\boldsymbol{g}_t\|^2}_{\text{A2}}\\&+\underbrace{2\eta_t\left\langle\overline{\boldsymbol{w}}_t-\boldsymbol{x}_*-\eta_t\overline{\boldsymbol{g}}_t,\overline{\boldsymbol{g}}_t-\boldsymbol{g}_t\right\rangle}_{\text{A3}}\end{aligned}\tag{2}$$

In order to bound Eqn. (2), we bound A1, A2, A3 separately. First, A3 is bounded as $\mathbb{E}[\text{A3}]=0$ since $\mathbb{E}[\boldsymbol{g}_t]=\overline{\boldsymbol{g}}_t$. Then to bound A1 and A2, we first give the following lemmas.

Lemma 1. *Assume Assumption 1 $\sim$ 2, $\beta\in(0,1)$, and $\alpha_t=\overline{\alpha}\le\min\{\frac{1-\beta}{\mu},\frac{1-\beta}{4L}\}$ hold, we have*

$$\mathbb{E}[\text{A1}]\le(1-\mu\overline{\eta})\|\overline{\boldsymbol{w}}_t-\boldsymbol{x}_*\|^2+\underbrace{\frac{1}{2}\sum_{m=1}^{M}d_m\|\boldsymbol{x}_t^m-\boldsymbol{w}_t^m\|^2}_{\text{H1}}+\underbrace{2\sum_{m=1}^{M}d_m\|\overline{\boldsymbol{w}}_t-\boldsymbol{w}_t^m\|^2}_{\text{H2}}+10L\overline{\eta}^2\Gamma_1\tag{3}$$

where $\overline{\eta}=\frac{\overline{\alpha}}{1-\beta}$, and $\Gamma_1=F_*-\sum_{m=1}^{M}d_mF_m^*\ge0$

Lemma 2. *Assume Assumption 4 holds, we have*

$$\text{H1}=\tfrac{1}{2}\sum_{m=1}^{M}d_m\|\boldsymbol{x}_t^m-\boldsymbol{w}_t^m\|^2\le\tfrac{\beta\overline{\eta}^2}{2\overline{\alpha}}\sum_{m=1}^{M}d_m\left[\|\boldsymbol{x}_1^m-\boldsymbol{x}_0^m\|+\tfrac{\overline{\alpha}G}{1-\beta}\right]^2=\tfrac{\beta\overline{\eta}^2}{2\overline{\alpha}}\left[\|\overline{\boldsymbol{x}}_1-\overline{\boldsymbol{x}}_0\|+\tfrac{\overline{\alpha}G}{1-\beta}\right]^2\tag{4}$$

Lemma 3. *Assume Assumption 1 and 3, $\beta\in(0,1)$, and $\alpha_t=\overline{\alpha}\le\min\{\frac{1-\beta}{\mu},\frac{1-\beta}{4L}\}$ hold, we have*

$$\left\|\boldsymbol{g}_{t+1}^m\right\|\le(1+(1-\beta)L\overline{\eta})^t\left\|\boldsymbol{g}_0^m\right\|+\underbrace{(1+\overline{\alpha}L)^T\frac{\Xi_1}{\overline{\alpha}L}}_{\Xi_2}\tag{5}$$

where Ξ_1 and Ξ_2 are the constant.

Lemma 4. *Assume Assumption 3 holds, we have*

$$\text{H2}=2\sum_{m=1}^{M}d_m\|\overline{\boldsymbol{w}}_t-\boldsymbol{w}_t^m\|^2\le2\overline{\eta}^2\chi^2\ell\sum_{m=1}^{M}d_m(\lambda_m^2\ell-\lambda_m)\tag{6}$$

Lemma 5. *Assume Assumption 3, $\beta\in(0,1)$, and $\alpha_t=\overline{\alpha}\le\min\{\frac{1-\beta}{\mu},\frac{1-\beta}{4L}\}$ hold, we have*

$$\mathbb{E}\left[\text{A2}\right]=\mathbb{E}\left[\overline{\eta}^2\|\overline{\boldsymbol{g}}_t-\boldsymbol{g}_t\|^2\right]=\mathbb{E}\left[\overline{\eta}^2\left\|\sum_{m=1}^{M}d_m(\nabla F_m(\boldsymbol{x}_t^m,\boldsymbol{\xi}_t^m)-\boldsymbol{g}_t^m)\right\|^2\right]=\overline{\eta}^2\sum_{i=1}^{M}d_m^2\delta_m^2\tag{7}$$

We firstly give the proof process of **Lemma 1**.

Proof. We can split A1 into three parts, which are shown as follows:

$$\mathrm{A1} = \|\overline{\boldsymbol{w}}_t - \boldsymbol{x}_*\|^2 + \underbrace{\eta_t^2 \|\overline{\boldsymbol{g}}_t\|^2}_{\mathrm{B1}} \underbrace{-2\eta_t \left\langle \overline{\boldsymbol{w}}_t - \boldsymbol{x}_*, \overline{\boldsymbol{g}}_t \right\rangle}_{\mathrm{B2}} \tag{8}$$

According to the **Assumption 1**, we can obtain the bound of B1 as follows:

$$\begin{aligned}
&\mathrm{B1} = \eta_t^2 \textstyle\sum_{m=1}^M d_m \|\nabla F_m(\boldsymbol{x}_t^m)\|^2 = \eta_t^2 \sum_{m=1}^M d_m \|\nabla F_m(\boldsymbol{x}_t^m) - \nabla F_m(\boldsymbol{w}_t^m) + \nabla F_m(\boldsymbol{w}_t^m)\|^2 \\
&\leq 2\eta_t^2 \textstyle\sum_{m=1}^M d_m \|\nabla F_m(\boldsymbol{x}_t^m) - \nabla F_m(\boldsymbol{w}_t^m)\|^2 + 2\eta_t^2 \sum_{m=1}^M d_m \|\nabla F_m(\boldsymbol{w}_t^m)\|^2 \\
&(\text{because of} \|a+b\|^2 \leq 2a^2 + 2b^2) \\
&\leq 2L^2\eta_t^2 \textstyle\sum_{m=1}^M d_m \|\boldsymbol{x}_t^m - \boldsymbol{w}_t^m\|^2 + 4L\eta_t^2 \sum_{m=1}^M d_m (F_m(\boldsymbol{w}_t^m) - F_*^m)
\end{aligned} \tag{9}$$

where $F_*^m = F_m(\boldsymbol{x}_*^m)$, $\boldsymbol{x}_*^m$ is the local optimal value for local function F_m. Next, we give the bound of B2, which is shown as follows:

$$\begin{aligned}
&\mathrm{B2} = -2\eta_t \textstyle\sum_{m=1}^M d_m \left\langle \overline{\boldsymbol{w}}_t - \boldsymbol{x}_*, \nabla F_m(\boldsymbol{x}_t^m) \right\rangle = -2\eta_t \sum_{m=1}^M d_m \left\langle \overline{\boldsymbol{w}}_t - \boldsymbol{w}_t^m, \nabla F_m(\boldsymbol{x}_t^m) \right\rangle - 2\eta_t \sum_{m=1}^M d_m \left\langle \boldsymbol{w}_t^m - \boldsymbol{x}_*, \nabla F_m(\boldsymbol{x}_t^m) \right\rangle \\
&= \textstyle\sum_{m=1}^M d_m \underbrace{-2\eta_t \left\langle \overline{\boldsymbol{w}}_t - \boldsymbol{w}_t^m, \nabla F_m(\boldsymbol{x}_t^m) \right\rangle}_{\mathrm{C1}} + \sum_{m=1}^M d_m \underbrace{-2\eta_t \left\langle \boldsymbol{x}_t^m - \boldsymbol{x}_*, \nabla F_m(\boldsymbol{x}_t^m) \right\rangle}_{\mathrm{C2}} + \sum_{m=1}^M d_m \underbrace{-2\eta_t \left\langle \boldsymbol{w}_t^m - \boldsymbol{x}_t^m, \nabla F_m(\boldsymbol{x}_t^m) \right\rangle}_{\mathrm{C3}}
\end{aligned} \tag{10}$$

Now, according to the Cauchy-Schwarz inequality and **Assumption 2**, we give the bound of C1, C2, and C3, which are shown as follows:

$$\begin{aligned}
&\mathrm{C1} \leq 2\eta_t \left|\overline{\boldsymbol{w}}_t - \boldsymbol{w}_t^m\right| \left|\nabla F_m(\boldsymbol{x}_t^m)\right| \leq \|\overline{\boldsymbol{w}}_t - \boldsymbol{w}_t^m\|^2 + \eta_t^2 \|\nabla F_m(\boldsymbol{x}_t^m)\|^2 \\
&\leq \|\overline{\boldsymbol{w}}_t - \boldsymbol{w}_t^m\|^2 + \eta_t^2 \|\nabla F_m(\boldsymbol{x}_t^m) - \nabla F_m(\boldsymbol{w}_t^m) + \nabla F_m(\boldsymbol{w}_t^m)\|^2 \\
&(\text{because of} \|a+b\|^2 \leq 2a^2 + 2b^2) \\
&\leq \|\overline{\boldsymbol{w}}_t - \boldsymbol{w}_t^m\|^2 + 2\eta_t^2 \|\nabla F_m(\boldsymbol{x}_t^m) - \nabla F_m(\boldsymbol{w}_t^m)\|^2 + 2\eta_t^2 \|\nabla F_m(\boldsymbol{w}_t^m)\|^2 \\
&\leq \|\overline{\boldsymbol{w}}_t - \boldsymbol{w}_t^m\|^2 + 2L^2\eta_t^2 \|\boldsymbol{x}_t^m - \boldsymbol{w}_t^m\|^2 + 4L\eta_t^2 (F_m(\boldsymbol{w}_t^m) - F_*^m)
\end{aligned} \tag{11}$$

$$\mathrm{C2} \leq -2\eta_t (F_m(\boldsymbol{x}_t^m) - F_m(\boldsymbol{x}_*)) - \mu\eta_t \|\boldsymbol{x}_t^m - \boldsymbol{x}_*\|^2 \tag{12}$$

$$\mathrm{C3} \leq 2\eta_t (F_m(\boldsymbol{x}_t^m) - F_m(\boldsymbol{w}_t^m)) + L\eta_t \|\boldsymbol{x}_t^m - \boldsymbol{w}_t^m\|^2 \tag{13}$$

Thus, we have

$$\textstyle\sum_{m=1}^M d_m \mathrm{C1} \leq \sum_{m=1}^M d_m \|\overline{\boldsymbol{w}}_t - \boldsymbol{w}_t^m\|^2 + 2L^2\eta_t^2 \sum_{m=1}^M d_m \|\boldsymbol{x}_t^m - \boldsymbol{w}_t^m\|^2 + 4L\eta_t^2 \sum_{m=1}^M d_m (F_m(\boldsymbol{w}_t^m) - F_*^m) \tag{14}$$

$$\textstyle\sum_{m=1}^M d_m \mathrm{C2} \leq -2\eta_t \sum_{m=1}^M d_m (F_m(\boldsymbol{x}_t^m) - F_m(\boldsymbol{x}_*)) - \mu\eta_t \sum_{m=1}^M d_m \|\boldsymbol{x}_t^m - \boldsymbol{x}_*\|^2 \tag{15}$$

$$\textstyle\sum_{m=1}^M d_m \mathrm{C3} \leq 2\eta_t \sum_{m=1}^M d_m (F_m(\boldsymbol{x}_t^m) - F_m(\boldsymbol{w}_t^m)) + L\eta_t \sum_{m=1}^M d_m \|\boldsymbol{x}_t^m - \boldsymbol{w}_t^m\|^2 \tag{16}$$

Therefore, we can obtain the bound of B2 as shown in follows:

$$\begin{aligned}
&\mathrm{B2} \leq \textstyle\sum_{m=1}^M d_m \|\overline{\boldsymbol{w}}_t - \boldsymbol{w}_t^m\|^2 + 2L^2\eta_t^2 \sum_{m=1}^M d_m \|\boldsymbol{x}_t^m - \boldsymbol{w}_t^m\|^2 + 4L\eta_t^2 \sum_{m=1}^M d_m (F_m(\boldsymbol{w}_t^m) - F_*^m) \\
&-2\eta_t \textstyle\sum_{m=1}^M d_m (F_m(\boldsymbol{w}_t^m) - F_m(\boldsymbol{x}_*)) - \mu\eta_t \sum_{m=1}^M d_m \|\boldsymbol{x}_t^m - \boldsymbol{x}_*\|^2 + L\eta_t \sum_{m=1}^M d_m \|\boldsymbol{x}_t^m - \boldsymbol{w}_t^m\|^2
\end{aligned} \tag{17}$$

Let $\theta_t = 2\eta_t(1-4L\eta_t)$, where $\eta_t \le \frac{1}{4L}$ (i.e., $\alpha_t \le \frac{1-\beta}{4L}$), we can bound A1 as follows:

$$\begin{aligned}
&\mathrm{A1} \le \|\overline{\boldsymbol{w}}_t - \boldsymbol{x}_*\|^2 + \mathrm{B1} + \mathrm{B2} = \|\overline{\boldsymbol{w}}_t - \boldsymbol{x}_*\|^2 + 2L^2\eta_t^2 \textstyle\sum_{m=1}^M d_m \|\boldsymbol{x}_t^m - \boldsymbol{w}_t^m\|^2 + 4L\eta_t^2 \sum_{m=1}^M d_m (F_m(\boldsymbol{w}_t^m) - F_*^m) \\
&+ \textstyle\sum_{m=1}^M d_m \|\overline{\boldsymbol{w}}_t - \boldsymbol{w}_t^m\|^2 + 2L^2\eta_t^2 \sum_{m=1}^M d_m \|\boldsymbol{x}_t^m - \boldsymbol{w}_t^m\|^2 + 4L\eta_t^2 \sum_{m=1}^M d_m (F_m(\boldsymbol{w}_t^m) - F_*^m) \\
&-2\eta_t \textstyle\sum_{m=1}^M d_m (F_m(\boldsymbol{w}_t^m) - F_m(\boldsymbol{x}_*)) - \mu\eta_t \sum_{m=1}^M d_m \|\boldsymbol{x}_t^m - \boldsymbol{x}_*\|^2 + L\eta_t \sum_{m=1}^M d_m \|\boldsymbol{x}_t^m - \boldsymbol{w}_t^m\|^2 \\
&= \|\overline{\boldsymbol{w}}_t - \boldsymbol{x}_*\|^2 - \mu\eta_t \|\overline{\boldsymbol{x}}_t - \boldsymbol{x}_*\|^2 + \underbrace{8L\eta_t^2 \sum_{m=1}^M d_m (F_m(\boldsymbol{w}_t^m) - F_*^m) - 2\eta_t \sum_{m=1}^M d_m (F_m(\boldsymbol{w}_t^m) - F_m(\boldsymbol{x}_*))}_{\mathrm{D1}} \\
&+ \underbrace{(4L\eta_t + 1)L\eta_t}_{\le\frac{1}{2}} \textstyle\sum_{m=1}^M d_m \|\boldsymbol{x}_t^m - \boldsymbol{w}_t^m\|^2 + \sum_{m=1}^M d_m \|\overline{\boldsymbol{w}}_t - \boldsymbol{w}_t^m\|^2 \\
&\le \|\overline{\boldsymbol{w}}_t - \boldsymbol{x}_*\|^2 - \mu\eta_t \|\overline{\boldsymbol{x}}_t - \boldsymbol{x}_*\|^2 + \underbrace{8L\eta_t^2 \sum_{m=1}^M d_m (F_m(\boldsymbol{w}_t^m) - F_*^m) - 2\eta_t \sum_{m=1}^M d_m (F_m(\boldsymbol{w}_t^m) - F_m(\boldsymbol{x}_*))}_{\mathrm{D1}} \\
&+ \tfrac{1}{2} \textstyle\sum_{m=1}^M d_m \|\boldsymbol{x}_t^m - \boldsymbol{w}_t^m\|^2 + \sum_{m=1}^M d_m \|\overline{\boldsymbol{w}}_t - \boldsymbol{w}_t^m\|^2
\end{aligned} \tag{18}$$

Next, we can bound D1 in the following manner:

$$\begin{aligned}
\mathrm{D1} &= -2\eta_t(1-4L\eta_t) \textstyle\sum_{m=1}^M d_m (F_m(\boldsymbol{w}_t^m) - F_*^m) + 2\eta_t \sum_{m=1}^M d_m (F_m(\boldsymbol{x}_*) - F_*^m) \\
&= -\theta_t \textstyle\sum_{m=1}^M d_m (F_m(\boldsymbol{w}_t^m) - F_*) + (2\eta_t - \theta_t) \sum_{m=1}^M d_m (F_* - F_*^m) \\
&= \underbrace{-\theta_t \sum_{m=1}^M d_m (F_m(\boldsymbol{w}_t^m) - F_*)}_{\mathrm{E1}} + 8L\eta_t^2 \Gamma_1
\end{aligned} \tag{19}$$

where $\Gamma_1 = \sum_{m=1}^M d_m (F_* - F_*^m) = F_* - \sum_{m=1}^M d_m F_*^m$. Before bounding E1, we firstly give the following bound:

$$\begin{aligned}
&\textstyle\sum_{m=1}^M d_m (F_m(\boldsymbol{w}_t^m) - F_*) = \sum_{m=1}^M d_m (F_m(\boldsymbol{w}_t^m) - F_m(\overline{\boldsymbol{w}}_t)) + \sum_{m=1}^M d_m (F_m(\overline{\boldsymbol{w}}_t) - F_*) \\
&\ge \textstyle\sum_{m=1}^M d_m \left\langle \nabla F_m(\overline{\boldsymbol{w}}_t), \boldsymbol{w}_t^m - \overline{\boldsymbol{w}}_t \right\rangle + (F(\overline{\boldsymbol{w}}_t) - F_*) \\
&\ge -\tfrac{1}{2} \textstyle\sum_{m=1}^M d_m \left[\eta_t \|\nabla F_m(\overline{\boldsymbol{w}}_t)\|^2 + \frac{1}{\eta_t} \|\boldsymbol{w}_t^m - \overline{\boldsymbol{w}}_t\|^2\right] + (F(\overline{\boldsymbol{w}}_t) - F_*) \\
&\ge -\textstyle\sum_{m=1}^M d_m \left[\eta_t L (F_m(\overline{\boldsymbol{w}}_t) - F_*^m) + \frac{1}{2\eta_t} \|\boldsymbol{w}_t^m - \overline{\boldsymbol{w}}_t\|^2\right] + (F(\overline{\boldsymbol{w}}_t) - F_*)
\end{aligned} \tag{20}$$

Thus, we can obtain the bound of E1 as follows:

$$\begin{aligned}
\mathrm{E1} &\le \theta_t \textstyle\sum_{m=1}^M d_m \left[\eta_t L (F_m(\overline{\boldsymbol{w}}_t) - F_*^m) + \frac{1}{2\eta_t} \|\boldsymbol{w}_t^m - \overline{\boldsymbol{w}}_t\|^2\right] - \theta_t (F(\overline{\boldsymbol{w}}_t) - F_*) \\
&= \theta_t \eta_t L \textstyle\sum_{m=1}^M d_m (F_m(\overline{\boldsymbol{w}}_t) - F_*) + \theta_t \eta_t L \sum_{m=1}^M d_m (F_* - F_*^m) \\
&\quad -\theta_t \textstyle\sum_{m=1}^M d_m (F_m(\overline{\boldsymbol{w}}_t) - F_*) + \frac{\theta_t}{2\eta_t} \sum_{m=1}^M d_m \|\boldsymbol{w}_t^m - \overline{\boldsymbol{w}}_t\|^2 \\
&= \underbrace{\theta_t (\underbrace{\eta_t L - 1}_{\le 0}) \sum_{m=1}^M d_m (\underbrace{F_m(\overline{\boldsymbol{w}}_t) - F_*}_{\ge 0})}_{\le 0} + \underbrace{\theta_t \eta_t L}_{\le 2L\eta_t^2} \Gamma_1 + \underbrace{\frac{\theta_t}{2\eta_t}}_{\le 1} \textstyle\sum_{m=1}^M d_m \|\boldsymbol{w}_t^m - \overline{\boldsymbol{w}}_t\|^2 \\
&\le 2L\eta_t^2 \Gamma_1 + \textstyle\sum_{m=1}^M d_m \|\boldsymbol{w}_t^m - \overline{\boldsymbol{w}}_t\|^2
\end{aligned} \tag{21}$$

Then, we bound D1 in the following manner:

$$\mathrm{D1} \le 10L\eta_t^2 \Gamma_1 + \sum_{m=1}^M d_m \|\overline{\boldsymbol{w}}_t - \boldsymbol{w}_t^m\|^2 \tag{22}$$

Therefore, we can rewrite the bound of A1 as follows:

$$\mathrm{A1} \le \|\overline{\boldsymbol{w}}_t - \boldsymbol{x}_*\|^2 - \mu\eta_t \|\overline{\boldsymbol{x}}_t - \boldsymbol{x}_*\|^2 + \underbrace{\frac{1}{2} \sum_{m=1}^M d_m \|\boldsymbol{x}_t^m - \boldsymbol{w}_t^m\|^2}_{\mathrm{H1}} + \underbrace{2 \sum_{m=1}^M d_m \|\overline{\boldsymbol{w}}_t - \boldsymbol{w}_t^m\|^2}_{\mathrm{H2}} + 10L\eta_t^2 \Gamma_1 \tag{23}$$

Because of $\mathbb{E}\|\overline{\boldsymbol{x}}_t - \boldsymbol{x}_*\|^2 = \|\overline{\boldsymbol{w}}_t - \boldsymbol{x}_*\|^2$, taking the expectation on A1 and holding the value of $\alpha_t = \overline{\alpha}$, we can obtain the Eqn. (3). Thus, we finish the proof of **Lemma 1**. $\square$

Secondly, we give the proof process of **Lemma 2**.

Proof. To bound H1, we first give the following inequality.

$$\|\boldsymbol{x}_t^m - \boldsymbol{w}_t^m\| = \gamma \left\|\boldsymbol{x}_t^m - \boldsymbol{x}_{t-1}^m\right\| = \frac{\beta}{\overline{\alpha}}\overline{\eta}\left\|\boldsymbol{x}_t^m - \boldsymbol{x}_{t-1}^m\right\| \tag{24}$$

where, we can bound the $\left\|\boldsymbol{x}_t^m - \boldsymbol{x}_{t-1}^m\right\|$ as follows:

$$\begin{aligned}&\left\|\boldsymbol{x}_t^m - \boldsymbol{x}_{t-1}^m\right\| = \left\|-\overline{\alpha}\boldsymbol{g}_{t-1}^m + \beta(\boldsymbol{x}_{t-1}^m - \boldsymbol{x}_{t-2}^m)\right\| \leq \overline{\alpha}\left\|\boldsymbol{g}_{t-1}^m\right\| + \beta\left\|\boldsymbol{x}_{t-1}^m - \boldsymbol{x}_{t-2}^m\right\| \leq \overline{\alpha}G + \beta\left\|\boldsymbol{x}_{t-1}^m - \boldsymbol{x}_{t-2}^m\right\| \\ &(\text{According to the difference equation}) \\ &\leq \left(\|\boldsymbol{x}_1^m - \boldsymbol{x}_0^m\| - \tfrac{\overline{\alpha}G}{1-\beta}\right)\underbrace{\beta^{t-1}}_{\leq 1} + \tfrac{\overline{\alpha}G}{1-\beta} \leq \|\boldsymbol{x}_1^m - \boldsymbol{x}_0^m\| + \tfrac{\overline{\alpha}G}{1-\beta}\end{aligned} \tag{25}$$

Now, we can obtain the bound of H1 as shown in follows:

$$\mathrm{H1} = \frac{1}{2}\sum\nolimits_{m=1}^{M} d_m\|\boldsymbol{x}_t^m - \boldsymbol{w}_t^m\|^2 \leq \frac{\beta\overline{\eta}^2}{2\overline{\alpha}}\sum\nolimits_{m=1}^{M} d_m\left[\|\boldsymbol{x}_1^m - \boldsymbol{x}_0^m\| + \frac{\overline{\alpha}G}{1-\beta}\right]^2 = \frac{\beta\overline{\eta}^2}{2\overline{\alpha}}\left[\|\overline{\boldsymbol{x}}_1 - \overline{\boldsymbol{x}}_0\| + \frac{\overline{\alpha}G}{1-\beta}\right]^2 \tag{26}$$

□

Then, we give the proof process of **Lemma 3**.

Proof. The detailed proof process is shown as follows:

$$\begin{aligned}&\left\|\boldsymbol{g}_{t+1}^m\right\| = \left\|\boldsymbol{g}_{t+1}^m - \boldsymbol{g}_t^m + \boldsymbol{g}_t^m\right\| \leq \left\|\boldsymbol{g}_{t+1}^m - \boldsymbol{g}_t^m\right\| + \|\boldsymbol{g}_t^m\| \\ &= \left\|\nabla F_m(\boldsymbol{x}_t^m - \overline{\alpha}\boldsymbol{g}_t^m + \beta(\boldsymbol{x}_t^m - \boldsymbol{x}_{t-1}^m)) - \nabla F_m(\boldsymbol{x}_t^m)\right\| + \|\boldsymbol{g}_t^m\| \\ &(\text{According to the } \textbf{Assumption 2}) \\ &\leq L\left\|-\overline{\alpha}\boldsymbol{g}_t^m + \beta(\boldsymbol{x}_t^m - \boldsymbol{x}_{t-1}^m)\right\| + \|\boldsymbol{g}_t^m\| \leq L\left\|-\overline{\alpha}\boldsymbol{g}_t^m\right\| + L\beta\left\|\boldsymbol{x}_t^m - \boldsymbol{x}_{t-1}^m\right\| + \|\boldsymbol{g}_t^m\| \\ &\leq (1+\overline{\alpha}L)\left\|\boldsymbol{g}_t^m\right\| + L\beta\left\|\boldsymbol{x}_t^m - \boldsymbol{x}_{t-1}^m\right\| \leq (1+\overline{\alpha}L)\left\|\boldsymbol{g}_t^m\right\| + \underbrace{L\beta\left[\|\boldsymbol{x}_1^m - \boldsymbol{x}_0^m\| + \frac{\overline{\alpha}G}{1-\beta}\right]}_{\Xi_1} \\ &(\text{According to the difference equation}) \\ &\leq \underbrace{-\frac{\Xi_1}{\overline{\alpha}L}}_{\leq 0} + \left[\|\boldsymbol{g}_0^m\| + \tfrac{\Xi_1}{\overline{\alpha}L}\right](1+\overline{\alpha}L)^t \leq (1+(1-\beta)L\overline{\eta})^t\left\|\boldsymbol{g}_0^m\right\| + \underbrace{(1+\overline{\alpha}L)^T\frac{\Xi_1}{\overline{\alpha}L}}_{\Xi_2}\end{aligned} \tag{27}$$

□

Next, we give the proof process of **Lemma 4**.

Proof. To bound H2, we firstly bound $\|\overline{\boldsymbol{w}}_t - \boldsymbol{w}_t^m\|$ in the following manner:

$$\begin{aligned}&\|\overline{\boldsymbol{w}}_t - \boldsymbol{w}_t^m\| = \left\|(\boldsymbol{w}_t^m - \boldsymbol{w}_{t_{m_0}}) - (\overline{\boldsymbol{w}}_t - \boldsymbol{w}_{t_{m_0}})\right\| = \sum\nolimits_{l=t_{m_0}}^{t}\left\|\overline{\eta}(\boldsymbol{g}_l^m - \sum\nolimits_{m=1}^{M} d_m\boldsymbol{g}_l^m)\right\| \\ &\leq \sum\nolimits_{l=t_{m_0}}^{t}(1+(1-\beta)L\overline{\eta})^{l-t_{m_0}}\left\|\overline{\eta}(\boldsymbol{g}_{t_{m_0}}^m - \sum\nolimits_{m=1}^{M} d_m\boldsymbol{g}_{t_{m_0}}^m)\right\| \leq \left\|\tfrac{(1+(1-\beta)L\overline{\eta})^{t-t_{m_0}}-1}{(1-\beta)L}(\boldsymbol{g}_{t_{m_0}}^m - \sum\nolimits_{m=1}^{M} d_m\boldsymbol{g}_{t_{m_0}}^m)\right\|\end{aligned} \tag{28}$$

The t_{m_0} is the last time step when device m uploads its model to the server. Note that the Ξ_2 can be eliminated as a constant term. According to the **Assumption 3** and $t - t_{m_0} \leq \lambda_m\ell - 1$, we can obtain the following inequality by using Taylor expansion.

$$\|\overline{\boldsymbol{w}}_t - \boldsymbol{w}_t^m\|^2 \leq \left\|\frac{(1+(1-\beta)L\overline{\eta})^{\lambda_m\ell-1}-1}{(1-\beta)L}\right\|^2\chi^2 \leq \frac{2}{2}\overline{\eta}^2(\lambda_m\ell-1)^2\chi^2 \leq \overline{\eta}^2\ell(\lambda_m^2\ell - \lambda_m)\chi^2 \tag{29}$$

Thus, we can bound the H2 as follows:

$$\mathrm{H2} = 2\sum_{m=1}^{M} d_m \|\overline{\boldsymbol{w}}_t - \boldsymbol{w}_t^m\|^2 \leq 2\overline{\eta}^2\chi^2\ell\sum_{m=1}^{M} d_m(\lambda_m^2\ell - \lambda_m) \tag{30}$$

□

Therefore, we can obtain the expectation of A1 in the following manner:

$$\mathbb{E}\left[\mathrm{A1}\right] \leq (1-\mu\overline{\eta})\|\overline{\boldsymbol{w}}_t - \boldsymbol{x}_*\|^2 + \frac{\beta\overline{\eta}^2}{2\overline{\alpha}}\left[\|\overline{\boldsymbol{x}}_1 - \overline{\boldsymbol{x}}_0\| + \frac{\overline{\alpha}G}{1-\beta}\right]^2 + 2\overline{\eta}^2\chi^2\ell\sum_{m=1}^{M} d_m(\lambda_m^2\ell - \lambda_m) + 10L\overline{\eta}^2\Gamma_1 \tag{31}$$

Finally, we give the proof process of **Lemma 5**.

Proof. We can take the expectation of A2 in the following manner:

$$\mathbb{E}\left[\mathrm{A2}\right] = \mathbb{E}\left[\overline{\eta}^2\|\overline{\boldsymbol{g}}_t - \boldsymbol{g}_t\|^2\right] = \mathbb{E}\left[\overline{\eta}^2\left\|\sum_{m=1}^{M} d_m(\nabla F_m(\boldsymbol{x}_t^m, \boldsymbol{\xi}_t^m) - \boldsymbol{g}_t^m)\right\|^2\right] = \overline{\eta}^2\sum_{i=1}^{M} d_m^2\delta_m^2 \tag{32}$$

□

Now, we have obtained the bound of A1, A2, and A3, we define $\Delta_{t+1} = \mathbb{E}(\|\overline{\boldsymbol{w}}_{t+1} - \boldsymbol{x}_*\|^2)$, taking the expectation of both side Eqn. (2) we have

$$\begin{aligned}\Delta_{t+1} &\leq (1-\mu\overline{\eta})\Delta_t + \tfrac{\beta\overline{\eta}^2}{2\overline{\alpha}}\left[\|\overline{\boldsymbol{x}}_1 - \overline{\boldsymbol{x}}_0\| + \tfrac{\overline{\alpha}G}{1-\beta}\right]^2 + 2\overline{\eta}^2\chi^2\ell\textstyle\sum_{m=1}^{M} d_m(\lambda_m^2\ell - \lambda_m) + 10L\overline{\eta}^2\Gamma_1 + \overline{\eta}^2\sum_{i=1}^{M} d_m^2\delta_m^2 \\ &= (1-\mu\overline{\eta})\Delta_t + \overline{\eta}^2\underbrace{\left[\underbrace{\frac{\beta}{2\overline{\alpha}}\left[\|\overline{\boldsymbol{x}}_1 - \overline{\boldsymbol{x}}_0\| + \frac{\overline{\alpha}G}{1-\beta}\right]^2}_{\Gamma_2} + 2\chi^2\ell\sum_{m=1}^{M} d_m(\lambda_m^2\ell - \lambda_m) + 10L\Gamma_1 + \sum_{i=1}^{M} d_m^2\delta_m^2\right]}_{\Lambda} \\ &= (1-\mu\overline{\eta})\Delta_t + \overline{\eta}^2\Lambda\end{aligned} \tag{33}$$

Taking the recursion, we have

$$\Delta_{t+1} - \frac{\overline{\eta}\Lambda}{\mu} \leq (1-\mu\overline{\eta})^{t+1}(\Delta_0 - \frac{\overline{\eta}\Lambda}{\mu}) \tag{34}$$

According to the **Assumption 1**, we have

$$\mathbb{E}(F(\overline{\boldsymbol{w}}_t)) - F_* \leq \frac{L}{2}\Delta_t \tag{35}$$

Finally, we can obtain the following conclusion

$$\mathbb{E}\left[F(\boldsymbol{x}_T)\right] - F_* \leq \frac{L(1-\mu\overline{\eta})^T}{2}(\|\boldsymbol{x}_{t_0} - \boldsymbol{x}_*\|^2 - 2\Phi) + \Phi \tag{36}$$

where $\Phi = \frac{\overline{\eta}L\Lambda}{2\mu} = \frac{\overline{\eta}L}{2\mu}[\Gamma_2 + 2\chi^2\ell\sum_{m=1}^{M} d_m(\lambda_m^2\ell - \lambda_m) + 10L\Gamma_1 + \sum_{m=1}^{M} d_m^2\delta_m^2]$, $\Gamma_1 = F_* - \sum_{m=1}^{M} d_m F_m^*$, and $\Gamma_2 = \frac{\beta}{2\overline{\alpha}}\left[\|\overline{\boldsymbol{x}}_1 - \overline{\boldsymbol{x}}_0\| + \frac{\overline{\alpha}G}{1-\beta}\right]^2$.

*B. **Proof of Theorem 2***

In order to bound Eqn. (2), we bound A1, A2, A3 separately. First, A3 is bounded as $\mathbb{E}[\mathrm{A3}] = 0$ since $\mathbb{E}[\boldsymbol{g}_t] = \overline{\boldsymbol{g}}_t$. Then to bound A1 and A2, we first give the following lemmas.

Lemma 6. *Assume Assumption 1 $\sim$ Assumption 2, $\beta \in (0,1)$, and $\alpha_t = \frac{2(1-\beta)}{\mu(\upsilon+t)}$ hold, where $\upsilon = \max\{8\rho, \max_i(\lambda_i)\ell\}$ and $\rho = \frac{L}{\mu}$, we have*

$$\mathbb{E}[\mathrm{A1}] \leq (1-\mu\eta_t)\|\overline{\boldsymbol{w}}_t - \boldsymbol{x}_*\|^2 + \underbrace{\frac{1}{2}\sum_{m=1}^{M} d_m\|\boldsymbol{x}_t^m - \boldsymbol{w}_t^m\|^2}_{\mathrm{H1}'} + \underbrace{2\sum_{m=1}^{M} d_m\|\overline{\boldsymbol{w}}_t - \boldsymbol{w}_t^m\|^2}_{\mathrm{H2}'} + 10L\eta_t^2\Gamma_1 \tag{37}$$

where $\eta_t = \frac{\alpha_t}{1-\beta}$, and $\Gamma_1 = F_* - \sum_{m=1}^{M} d_m F_m^* \geq 0$

Proof. The process of **Lemma 6** is same as the **Lemma 1**, only replacing $\overline{\eta}$ with η_t. ☐

Lemma 7. *Assume Assumption 4 and $\alpha_{\min} = \min(\alpha_t)$, and $\alpha_{\max} = \max(\alpha_t)$ hold, we have*

$$\mathrm{H1}' = \tfrac{1}{2}\sum_{m=1}^{M} d_m\|\boldsymbol{x}_t^m - \boldsymbol{w}_t^m\|^2 \leq \tfrac{\beta\eta_t^2}{2\alpha_{min}}\sum_{m=1}^{M} d_m\left[\|\boldsymbol{x}_1^m - \boldsymbol{x}_0^m\| + \tfrac{\alpha_{max}G}{1-\beta}\right]^2 = \tfrac{\beta\eta_t^2}{2\alpha_{min}}\left[\|\overline{\boldsymbol{x}}_1 - \overline{\boldsymbol{x}}_0\| + \tfrac{\alpha_{max}G}{1-\beta}\right]^2 \tag{38}$$

Proof. The process of **Lemma 7** is same as the **Lemma 2**, only replacing $\overline{\alpha}$ with α_{min} and α_{max}. ☐

Lemma 8. *Assume Assumption 1 and 3, $\beta \in (0,1)$, $\alpha_t = \frac{2(1-\beta)}{\mu(\upsilon+t)}$, $\alpha_{\min} = \min(\alpha_t)$, and $\alpha_{\max} = \max(\alpha_t)$ hold, we have*

$$\left\|\boldsymbol{g}_{t+1}^m\right\| \leq (1+(1-\beta)L\eta_t)^t\left\|\boldsymbol{g}_0^m\right\| + \underbrace{(1+\alpha_{max}L)^T\frac{\Xi 1'}{\alpha_{min}L}}_{\Xi 2'} \tag{39}$$

where $\Xi 1'$ and $\Xi 2'$ are the constant.

Proof. The process of **Lemma 8** is same as the **Lemma 3**, only replacing $\overline{\eta}$ with η_t, and $\overline{\alpha}$ with α_{min} and α_{max}. ☐

Lemma 9. *Assume Assumption 3, $\beta \in (0,1)$, $\alpha_t = \frac{2(1-\beta)}{\mu(\upsilon+t)}$, $\alpha_{\min} = \min(\alpha_t)$, and $\alpha_{\max} = \max(\alpha_t)$ hold. Moreover, we have $\eta_t \leq 2\eta_{t+\lambda_m\ell}$ (i.e., $\eta_{t_{m_0}} \leq 2\eta_t$), thus we have*

$$\mathrm{H2}' = 2\sum_{m=1}^{M} d_m\|\overline{\boldsymbol{w}}_t - \boldsymbol{w}_t^m\|^2 \leq 8\eta_t^2\chi^2\ell\sum_{m=1}^{M} d_m(\lambda_m^2\ell - \lambda_m) \tag{40}$$

Proof. Inspired by the process of **Lemma 4**, using the inequality $\eta_t \leq 2\eta_{t+\lambda_m\ell}$ (i.e., $\eta_{t_{m_0}} \leq 2\eta_t$), and replacing $\overline{\eta}$ with η_t, the process of **Lemma 9** is finished. ☐

Lemma 10. *Assume Assumption 3, $\beta \in (0,1)$, $\alpha_t = \frac{2(1-\beta)}{\mu(\upsilon+t)}$, $\alpha_{\min} = \min(\alpha_t)$, and $\alpha_{\max} = \max(\alpha_t)$ hold. We have*

$$\mathbb{E}\left[\mathrm{A2}\right] = \mathbb{E}\left[\eta_t^2\|\overline{\boldsymbol{g}}_t - \boldsymbol{g}_t\|^2\right] = \mathbb{E}\left[\eta_t^2\left\|\sum_{m=1}^{M} d_m(\nabla F_m(\boldsymbol{x}_t^m, \boldsymbol{\xi}_t^m) - \boldsymbol{g}_t^m)\right\|^2\right] = \eta_t^2\sum_{i=1}^{M} d_m^2\delta_m^2 \tag{41}$$

Proof. The process of **Lemma 10** is same as the **Lemma 5**, only replacing $\overline{\eta}$ with η_t. ☐

Using the new bound and replacing Λ with Λ' in inequality (33), we have

$$\Delta_{t+1} \leq (1-\mu\eta_t)\Delta_t + \eta_t^2\Lambda' \tag{42}$$

The rest of the proof can refer to the literature [1]. Then we finish the proof of **Theorem 2**.

C. *Proof of Theorem 3*

The expression of NAG in an asynchronous federated learning system is shown as follows:

$$\begin{aligned} &\boldsymbol{w}_{t+1}^m = \boldsymbol{x}_t^m - \alpha_t \nabla F_m(\boldsymbol{x}_t^m, \xi_t^m) \\ &\boldsymbol{x}_{t+1}^m = \begin{cases} \boldsymbol{w}_{t+1}^m + \beta_t(\boldsymbol{w}_{t+1}^m - \boldsymbol{w}_t^m) & \text{if } \mathrm{t}+1 \notin \boldsymbol{\Theta}_m \\ \boldsymbol{x}_{t+1} & \text{if } \mathrm{t}+1 \in \boldsymbol{\Theta}_m \end{cases} \end{aligned} \tag{43}$$

According to the literature [2], we first give the following lemmas.

Lemma 11. *Assume Assumption 1 ∼ 4, let* $\overline{\boldsymbol{w}}_t = \sum_{m=1}^M d_m \boldsymbol{w}_t^m$, $\alpha_t = \frac{6}{\mu(t+\upsilon)}$, $\beta_{t-1} = \frac{3}{14(t+\upsilon)(1-\frac{6}{t+\upsilon})\max\{\mu,1\}}$, $\upsilon = \max\{32\rho, \max_m(\lambda_m)\ell\}$, *and* $\rho{=}\frac{L}{\mu}$ *hold, we have*

$$\begin{aligned} &\mathbb{E}\|\overline{\boldsymbol{w}}_{t+1} - x_*\|^2 \le \mathbb{E}(1-\mu\alpha_t)(1+\beta_{t-1})^2\|\overline{\boldsymbol{w}}_t - x_*\|^2 + 20L\alpha_t^3 G^2 \textstyle\sum_{m=1}^M d_m \lambda_m^2 \ell^2 \\ &+(1-\mu\alpha_t)\beta_{t-1}^2\|\overline{\boldsymbol{w}}_{t-1} - \boldsymbol{x}_*\|^2 + \frac{\alpha_t^2 \nu_{\max}}{M}\textstyle\sum_{m=1}^M d_m^2 \delta_m^2 + 2\beta_{t-1}(1+\beta_{t-1})(1-\mu\alpha_t)\,\|\overline{\boldsymbol{w}}_t - \boldsymbol{x}_*\|\,\|\overline{\boldsymbol{w}}_{t-1} - \boldsymbol{x}_*\| \end{aligned} \tag{44}$$

where $\nu_{\max} = \max_m M d_m$

Proof. The process of **Lemma 11** can be deduced from combining the proof process of **Theorem 1** and the Lemma 7 of [2]. □

Lemma 12. *Assume Assumption 1 ∼ 4, let* $\overline{\boldsymbol{w}}_t = \sum_{m=1}^M d_m \boldsymbol{w}_t^m$ *and assume that* α_t *is non-increasing,* $\alpha_t \le 2\alpha_{t+\lambda_m \ell}$ *(i.e.,* $\alpha_{t_{m_0}} \le 2\alpha_t$*),* $t - t_{m_0} \le \lambda_m \ell - 1$, *and* $2\beta_{t-1}^2 + 2\alpha_t^2 \le \frac{1}{2}$ *for all* $t \ge 0$, *where the definition of* t_{m_0} *is given in the Eqn. (28), we have*

$$\mathbb{E}\sum_{m=1}^M d_m \|\overline{\boldsymbol{x}}_{t+1} - \boldsymbol{x}_t^m\|^2 \le 16\alpha_{t_0}^2 G^2 \sum_{m=1}^M d_m (\lambda_m \ell - 1)^2 \tag{45}$$

Proof. We can take the expectation of $\sum_{m=1}^M d_m \|\overline{\boldsymbol{x}}_{t+1} - \boldsymbol{x}_t^m\|^2$ as follows:

$$\begin{aligned} &\mathbb{E}\textstyle\sum_{m=1}^M d_m \|\overline{\boldsymbol{x}}_{t+1} - \boldsymbol{x}_t^m\|^2 \le \mathbb{E}\sum_{m=1}^M d_m \|\overline{\boldsymbol{x}}_{t+1} - \boldsymbol{x}_{t_{m0}}\|^2 \\ &= \mathbb{E}\textstyle\sum_{m=1}^M d_m \left\|\sum_{i=t_{m0}}^{t-1} \beta_i(\boldsymbol{w}_{i+1}^m - \boldsymbol{w}_i^m) - \sum_{i=t_{m0}}^{t-1} \alpha_i \boldsymbol{g}_i^m\right\|^2 \\ &\le 2\textstyle\sum_{m=1}^M d_m \mathbb{E}\sum_{i=t_{m0}}^{t-1} (\lambda_m \ell - 1)^2 \beta_i^2 \left\|(\boldsymbol{w}_{i+1}^m - \boldsymbol{w}_i^m)\right\|^2 + 2\sum_{i=t_{m0}}^{t-1} d_m \mathbb{E}\sum_{i=t_{m0}}^{t-1} (\lambda_m \ell - 1)\alpha_i^2 \|\boldsymbol{g}_i^m\|^2 \\ &\le 2\textstyle\sum_{m=1}^M d_m \mathbb{E}\sum_{i=t_{m0}}^{t-1} (\lambda_m \ell - 1)^2 \alpha_i^2 \left[\|\boldsymbol{g}_i^m\|^2 + \left\|(\boldsymbol{w}_{i+1}^m - \boldsymbol{w}_i^m)\right\|^2\right] \\ &\le 4\textstyle\sum_{m=1}^M d_m \mathbb{E}\sum_{i=t_{m0}}^{t-1} (\lambda_m \ell - 1)^2 \alpha_i^2 G^2 \le 4\sum_{m=1}^M d_m (\lambda_m \ell - 1)^2 \alpha_{t_{m0}}^2 G^2 \le 16\alpha_{t_0}^2 G^2 \sum_{m=1}^M d_m (\lambda_m \ell - 1)^2 \end{aligned} \tag{46}$$

□

Then, we can refer to the literature [2] and combine the **Lemma 11** and **Lemma 12** to finish the rest proof of **Theorem 3**.

References

[1] X. Li, K. Huang, W. Yang, S. Wang, and Z. Zhang, "On the convergence of fedavg on non-iid data," in *International Conference on Learning Representations (ICLR)*, 2019.

[2] Z. Qu, K. Lin, Z. Li, J. Zhou, and Z. Zhou, "A unified linear speedup analysis of federated averaging and nesterov fedavg," *Journal of Artificial Intelligence Research*, vol. 78, pp. 1143–1200, 2023.