

**Lemma D.8.** Fix the number of communication rounds  $T$ , local update rounds  $\tau$ , and the number of participating clients  $P$  per communication round. Under Assumptions 3.1 and 3.2, the client drifts  $\Delta_x^{(t)}$ ,  $\Delta_y^{(t)}$  and  $\Delta_\theta^{(t)}$  defined in (97) can be bounded as follows:

$$\Delta_x^{(t)} \leq 3\eta_x^2 \tau \left( \frac{\delta_f^2}{c_t^2} + 2\delta_g^2 \right) + 3\eta_x^2 \tau \left( \frac{L_f^2}{c_t^2} + 2L_g^2 \right), \quad (74)$$

$$\begin{aligned} \Delta_\theta^{(t)} \leq & \frac{1}{\frac{(1-12\tau\eta_y^2(\frac{1}{\gamma^2}+L_2^2))}{6\tau\eta_\theta^2\frac{1}{\gamma^2}}(1-6\tau\eta_\theta^2(L_2^2+\frac{1}{\gamma^2}))-12\tau\eta_y^2(\frac{1}{\gamma^2}+L_2^2)} \left\{ \frac{(1-12\tau\eta_y^2(\frac{1}{\gamma^2}+L_2^2))}{6\tau\eta_\theta^2\frac{1}{\gamma^2}} \left( \eta_\theta^2 \tau \delta_g^2 \right. \right. \\ & \left. \left. + 6\tau\eta_\theta^2 L_2^2 \Delta_x^{(t)} + 4\eta_\theta^2 \tau \Delta^2 + 4\tau\eta_\theta^2 \left( \frac{1}{\gamma^2} + L_2^2 \right) \mathbb{E} \left\| \theta^{(t)} - \theta_\gamma^*(x^{(t)}, y^{(t)}) \right\|^2 \right) \right. \\ & \left. + 2\eta_y^2 \tau \left( \frac{\delta_f^2}{c_t^2} + \delta_g^2 \right) + 12L_2^2 \tau \eta_y^2 \Delta_x^{(t)} + 18\tau\eta_y^2 \frac{L_f^2}{c_t^2} + 8\eta_y^2 \tau \Delta^2 + 16\frac{\eta_y^2}{\lambda_y^2} \tau \mathbb{E} \left\| y^{(t+1)} - y^{(t)} \right\|^2 \right\}, \end{aligned} \quad (75)$$

$$\begin{aligned} \Delta_y^{(t)} \leq & \frac{1}{(1-12\tau\eta_y^2(\frac{1}{\gamma^2}+L_2^2)) - \frac{12\tau\eta_y^2(\frac{1}{\gamma^2}+L_2^2)}{(1-6\tau\eta_\theta^2(L_2^2+\frac{1}{\gamma^2}))} 6\tau\eta_\theta^2 \frac{1}{\gamma^2}} \left\{ \frac{12\tau\eta_y^2(\frac{1}{\gamma^2}+L_2^2)}{(1-6\tau\eta_\theta^2(L_2^2+\frac{1}{\gamma^2}))} \left( \eta_\theta^2 \tau \delta_g^2 \right. \right. \\ & \left. \left. + 6\tau\eta_\theta^2 L_2^2 \Delta_x^{(t)} + 4\eta_\theta^2 \tau \Delta^2 + 4\tau\eta_\theta^2 \left( \frac{1}{\gamma^2} + L_2^2 \right) \mathbb{E} \left\| \theta^{(t)} - \theta_\gamma^*(x^{(t)}, y^{(t)}) \right\|^2 \right) \right. \\ & \left. + 2\eta_y^2 \tau \left( \frac{\delta_f^2}{c_t^2} + \delta_g^2 \right) + 12L_2^2 \tau \eta_y^2 \Delta_x^{(t)} + 18\tau\eta_y^2 \frac{L_f^2}{c_t^2} + 8\eta_y^2 \tau \Delta^2 + 16\frac{\eta_y^2}{\lambda_y^2} \tau \mathbb{E} \left\| y^{(t+1)} - y^{(t)} \right\|^2 \right\}. \end{aligned} \quad (76)$$

*Proof.* For  $\Delta_x^{(t)}$ ,

$$\begin{aligned} \Delta_x^{(t)} &= \sum_{i=1}^n \frac{1}{\tau} \sum_{k=0}^{\tau-1} \mathbb{E} \|x_i^{t,k} - x^{(t)}\|^2 \\ &= \sum_{i=1}^n \frac{1}{n} \eta_x^2 \sum_{k=0}^{\tau-1} \mathbb{E} \left\| \sum_{j=0}^{k-1} (\tilde{h}_{i,x}^{(t,j)} - h_{i,x}^{(t,j)}) \right\|^2 \\ &\quad + \sum_{i=1}^n \frac{1}{n} \eta_x^2 \sum_{k=0}^{\tau-1} \mathbb{E} \left\| \sum_{j=0}^{k-1} \tilde{h}_{i,x}^{(t,j)} \right\|^2 \\ &\stackrel{(a)}{\leq} \eta_x^2 \frac{1}{\tau} \sum_{k=0}^{\tau-1} 3k \left( \frac{\delta_f^2}{c_t^2} + 2\delta_g^2 \right) + \eta_x^2 \frac{1}{\tau} \sum_{k=0}^{\tau-1} 3k \left( \frac{L_f^2}{c_t^2} + 2L_g^2 \right) \\ &\leq 3\eta_x^2 \tau \left( \frac{\delta_f^2}{c_t^2} + 2\delta_g^2 \right) + 3\eta_x^2 \tau \left( \frac{L_f^2}{c_t^2} + 2L_g^2 \right), \end{aligned}$$

where the first term in (a) comes from L-smoothness of  $f_i$  and  $g_i$ , the second term in (a) comes from the Lipschitz continuity of  $f_i$  and  $g_i$ .

Next, we will analyze  $\Delta_y^{(t)}$  and  $\Delta_\theta^{(t)}$ . For  $\Delta_\theta^{(t)}$ ,

$$\begin{aligned} \Delta_\theta^{(t)} &:= \sum_{i=1}^n \frac{1}{n} \frac{1}{\tau} \sum_{k=0}^{\tau-1} \mathbb{E} \|\theta_i^{t,k} - \theta^{(t)}\|^2 \\ &= \sum_{i=1}^n \frac{1}{n} \frac{\eta_\theta^2}{\tau} \sum_{k=0}^{\tau-1} \mathbb{E} \left\| \sum_{j=0}^{k-1} h_{i,\theta}^{(t,j)} \right\|^2 \end{aligned}$$

$$\begin{aligned}
&= \sum_{i=1}^n \frac{1}{n} \eta_\theta^2 \frac{1}{\tau} \sum_{k=0}^{\tau-1} \mathbb{E} \left\| \sum_{j=0}^{k-1} \tilde{h}_{i,\theta}^{(t,j)} - h_{i,\theta}^{(t,j)} \right\|^2 + \sum_{i=1}^n \frac{1}{n} \eta_\theta^2 \frac{1}{\tau} \sum_{k=0}^{\tau-1} \mathbb{E} \left\| \sum_{j=0}^{k-1} \tilde{h}_{i,\theta}^{(t,j)} \right\|^2 \\
&\stackrel{(a)}{\leq} \eta_\theta^2 \tau \delta_g^2 + \sum_{i=1}^n \frac{1}{n} \eta_\theta^2 \frac{1}{\tau} \sum_{k=0}^{\tau-1} \mathbb{E} \left\| \sum_{j=0}^{k-1} \tilde{h}_{i,\theta}^{(t,j)} \right\|^2,
\end{aligned} \tag{77}$$

where the first term in (a) comes from L-smoothness of  $f_i$  and  $g_i$ . For the second term in (77),

$$\begin{aligned}
&\sum_{i=1}^n \frac{1}{n} \eta_\theta^2 \frac{1}{\tau} \sum_{k=0}^{\tau-1} \mathbb{E} \left\| \sum_{j=0}^{k-1} \tilde{h}_{i,\theta}^{(t,j)} \right\|^2 \\
&\leq \sum_{i=1}^n \frac{1}{n} \eta_\theta^2 \frac{1}{\tau} \sum_{k=0}^{\tau-1} \mathbb{E} \left\| \sum_{j=0}^{k-1} \nabla_y g_i(x_i^{(t,j)}, \theta_i^{(t,j)}) + \frac{1}{\gamma} (\theta_i^{(t,j)} - y_i^{(t,j)}) \right\|^2 \\
&\stackrel{(a)}{\leq} 2 \sum_{i=1}^n \frac{1}{n} \eta_\theta^2 \frac{1}{\tau} \sum_{k=0}^{\tau-1} \sum_{j=0}^{k-1} \left( \mathbb{E} \left\| \nabla_y g_i(x_i^{(t,j)}, \theta_i^{(t,j)}) + \frac{1}{\gamma} (\theta_i^{(t,j)} - y_i^{(t,j)}) \right. \right. \\
&\quad \left. \left. - \nabla_y g_i(x^{(t)}, \theta^{(t)}) - \frac{1}{\gamma} (\theta^{(t)} - y^{(t)}) \right\|^2 \right) \\
&\quad + 2 \sum_{i=1}^n \frac{1}{n} \eta_\theta^2 \frac{1}{\tau} \sum_{k=0}^{\tau-1} \sum_{j=0}^{k-1} \left( \mathbb{E} \left\| \nabla_y g_i(x^{(t)}, \theta^{(t)}) + \frac{1}{\gamma} (\theta^{(t)} - y^{(t)}) - \sum_{i=1}^n \nabla_y g_i(x^{(t)}, \theta^{(t)}) \right. \right. \\
&\quad \left. \left. - \frac{1}{\gamma} (\theta^{(t)} - y^{(t)}) + \sum_{i=1}^n \nabla_y g_i(x^{(t)}, \theta^{(t)}) + \frac{1}{\gamma} (\theta^{(t)} - y^{(t)}) - \sum_{i=1}^n \nabla_y g_i(x^{(t)}, \theta_\gamma^*(x^{(t)}, y^{(t)})) \right. \right. \\
&\quad \left. \left. - \frac{1}{\gamma} (\theta_\gamma^*(x^{(t)}, y^{(t)}) - y^{(t)}) \right\|^2 \right) \\
&\stackrel{(b)}{\leq} 6L_2^2 \eta_\theta^2 \sum_{i=1}^n \frac{1}{n} \frac{1}{\tau} \sum_{k=0}^{\tau-1} \sum_{j=0}^{k-1} \mathbb{E} \left\| x_i^{(t,j)} - x^{(t)} \right\|^2 + 6\eta_\theta^2 \frac{1}{\gamma^2} \sum_{i=1}^n \frac{1}{n} \frac{1}{\tau} \sum_{k=0}^{\tau-1} \sum_{j=0}^{k-1} \mathbb{E} \left\| y_i^{(t,j)} - y^{(t)} \right\|^2 \\
&\quad + 6\eta_\theta^2 (L_2^2 + \frac{1}{\gamma^2}) \sum_{i=1}^n \frac{1}{n} \frac{1}{\tau} \sum_{k=0}^{\tau-1} \sum_{j=0}^{k-1} \mathbb{E} \left\| \theta_i^{(t,j)} - \theta^{(t)} \right\|^2 + 4\eta_\theta^2 \tau \Delta^2 \\
&\quad + 4\tau \eta_\theta^2 (\frac{1}{\gamma^2} + L_2^2) \mathbb{E} \left\| \theta^{(t)} - \theta_\gamma^*(x^{(t)}, y^{(t)}) \right\|^2 \\
&\leq 6\tau \eta_\theta^2 L_2^2 \Delta_x^{(t)} + 6\tau \eta_\theta^2 (L_2^2 + \frac{1}{\gamma^2}) \Delta_\theta^{(t)} + 6\tau \eta_\theta^2 \frac{1}{\gamma^2} \Delta_y^{(t)} + 4\eta_\theta^2 \tau \Delta^2 \\
&\quad + 4\tau \eta_\theta^2 (\frac{1}{\gamma^2} + L_2^2) \mathbb{E} \left\| \theta^{(t)} - \theta_\gamma^*(x^{(t)}, y^{(t)}) \right\|^2,
\end{aligned}$$

where (a) comes from the definition of  $\theta_\gamma^*(x^{(t)}, y^{(t)})$  and (b) comes from Assumption 3.2 (iii) and L-smoothness of  $f_i(x, y)$  and  $g_i(x, y)$  in Assumption 3.1 (i) and Assumption 3.2 (i). Then we have

$$\begin{aligned}
\Delta_\theta^{(t)} &\leq \eta_\theta^2 \tau \delta_g^2 + 6\tau \eta_\theta^2 L_2^2 \Delta_x^{(t)} + 6\tau \eta_\theta^2 (L_2^2 + \frac{1}{\gamma^2}) \Delta_\theta^{(t)} + 6\tau \eta_\theta^2 \frac{1}{\gamma^2} \Delta_y^{(t)} + 4\eta_\theta^2 \tau \Delta^2 \\
&\quad + 4\tau \eta_\theta^2 (\frac{1}{\gamma^2} + L_2^2) \mathbb{E} \left\| \theta^{(t)} - \theta_\gamma^*(x^{(t)}, y^{(t)}) \right\|^2.
\end{aligned} \tag{78}$$

For  $\Delta_y^{(t)}$ ,

$$\Delta_y^{(t)} = \sum_{i=1}^n \frac{1}{n} \frac{1}{\tau} \sum_{k=0}^{\tau-1} \mathbb{E} \|y_i^{t,k} - y^{(t)}\|^2$$

$$\begin{aligned}
&= \sum_{i=1}^n \frac{1}{n} \frac{\eta_y^2}{\tau} \sum_{k=0}^{\tau-1} \mathbb{E} \left\| \sum_{j=0}^{k-1} h_{i,y}^{(t,j)} \right\|^2 \\
&= \sum_{i=1}^n \frac{1}{n} \frac{\eta_y^2}{\tau} \sum_{k=0}^{\tau-1} \mathbb{E} \left\| \sum_{j=0}^{k-1} \tilde{h}_{i,y}^{(t,j)} - h_{i,y}^{(t,j)} \right\|^2 + \sum_{i=1}^n \frac{1}{n} \frac{\eta_y^2}{\tau} \sum_{k=0}^{\tau-1} \mathbb{E} \left\| \sum_{j=0}^{k-1} \tilde{h}_{i,y}^{(t,j)} \right\|^2 \\
&\stackrel{(a)}{\leq} 2\eta_y^2 \tau \left( \frac{\delta_f^2}{c_t^2} + \delta_g^2 \right) + \sum_{i=1}^n \frac{1}{n} \frac{\eta_y^2}{\tau} \sum_{k=0}^{\tau-1} \mathbb{E} \left\| \sum_{j=0}^{k-1} \tilde{h}_{i,y}^{(t,j)} \right\|^2, \tag{79}
\end{aligned}$$

where the first term in (a) comes from L-smoothness of  $f_i$  and  $g_i$ . For the second term in (79),

$$\begin{aligned}
&\sum_{i=1}^n \frac{1}{n} \frac{\eta_y^2}{\tau} \sum_{k=0}^{\tau-1} \mathbb{E} \left\| \sum_{j=0}^{k-1} \tilde{h}_{i,y}^{(t,j)} \right\|^2 \\
&\leq \sum_{i=1}^n \frac{1}{n} \frac{\eta_y^2}{\tau} \sum_{k=0}^{\tau-1} \mathbb{E} \left\| \sum_{j=0}^{k-1} \frac{1}{c_t} \nabla_y f_i(x_i^{(t,j)}, y_i^{(t,j)}) + \nabla_y g_i(x_i^{(t,j)}, y_i^{(t,j)}) - \frac{1}{\gamma} (y_i^{(t,j)} - \theta_i^{(t,j)}) \right\|^2 \\
&\leq 4 \sum_{i=1}^n \frac{1}{n} \frac{\eta_y^2}{\tau} \sum_{k=0}^{\tau-1} \mathbb{E} \left\| \sum_{j=0}^{k-1} \nabla_y g_i(x_i^{(t,j)}, y_i^{(t,j)}) - \frac{1}{\gamma} (y_i^{(t,j)} - \theta_i^{(t,j)}) - (\nabla_y g_i(x^{(t)}, y^{(t)}) - \frac{1}{\gamma} (y^{(t)} - \theta^{(t)})) \right\|^2 \\
&\quad + 2 \sum_{i=1}^n \frac{1}{n} \frac{\eta_y^2}{\tau} \sum_{k=0}^{\tau-1} \mathbb{E} \left\| \sum_{j=0}^{k-1} \frac{1}{c_t} \nabla_y f_i(x_i^{(t,j)}, y_i^{(t,j)}) \right\|^2 + 4 \sum_{i=1}^n \frac{1}{n} \frac{\eta_y^2}{\tau} \sum_{k=0}^{\tau-1} \mathbb{E} \left\| \nabla_y g_i(x^{(t)}, y^{(t)}) - \frac{1}{\gamma} (y^{(t)} - \theta^{(t)}) \right\|^2 \\
&\stackrel{(a)}{\leq} 2\tau\eta_y^2 \frac{L_f^2}{c_t^2} + 12L_2^2\tau\eta_y^2\Delta_x^{(t)} + 12\left(\frac{1}{\gamma^2} + L_2^2\right)\tau\eta_y^2\Delta_y^{(t)} + \frac{12}{\gamma^2}\tau\eta_y^2\Delta_\theta^{(t)} \\
&\quad + 8 \sum_{i=1}^n \frac{\eta_y^2}{\tau} \sum_{k=0}^{\tau-1} \mathbb{E} \left\| \frac{1}{n} \sum_{i=1}^n \nabla_y g_i(x^{(t)}, y^{(t)}) + \frac{1}{\gamma} (y^{(t)} - \theta^{(t)}) \right\|^2 \\
&\quad + 8 \sum_{i=1}^n \frac{1}{n} \frac{\eta_y^2}{\tau} \sum_{k=0}^{\tau-1} \mathbb{E} \left\| \nabla_y g_i(x^{(t)}, y^{(t)}) - \frac{1}{\gamma} (y^{(t)} - \theta^{(t)}) - \frac{1}{n} \sum_{i=1}^n \nabla_y g_i(x^{(t)}, y^{(t)}) + \frac{1}{\gamma} (y^{(t)} - \theta^{(t)}) \right\|^2 \\
&\stackrel{(b)}{\leq} 2\tau\eta_y^2 \frac{L_f^2}{c_t^2} + 12L_2^2\tau\eta_y^2\Delta_x^{(t)} + 12\left(\frac{1}{\gamma^2} + L_2^2\right)\tau\eta_y^2\Delta_y^{(t)} + \frac{12}{\gamma^2}\tau\eta_y^2\Delta_\theta^{(t)} + 16 \sum_{i=1}^n \frac{\eta_y^2}{\tau} \sum_{k=0}^{\tau-1} \mathbb{E} \left\| \frac{1}{n} \sum_{i=1}^n \nabla_y \frac{1}{c_t} f_i(x^{(t)}, y^{(t)}) \right\|^2 \\
&\quad + 8\eta_y^2\tau\Delta^2 + 16 \sum_{i=1}^n \frac{\eta_y^2}{\tau} \sum_{k=0}^{\tau-1} \mathbb{E} \left\| \frac{1}{n} \sum_{i=1}^n \nabla_y \frac{1}{c_t} f_i(x^{(t)}, y^{(t)}) + \nabla_y g_i(x^{(t)}, y^{(t)}) + \frac{1}{\gamma} (y^{(t)} - \theta^{(t)}) \right\|^2, \\
&\leq 18\tau\eta_y^2 \frac{L_f^2}{c_t^2} + 12L_2^2\tau\eta_y^2\Delta_x^{(t)} + 12\left(\frac{1}{\gamma^2} + L_2^2\right)\tau\eta_y^2\Delta_y^{(t)} + \frac{12}{\gamma^2}\tau\eta_y^2\Delta_\theta^{(t)} + 8\eta_y^2\tau\Delta^2 + 16 \frac{\eta_y^2}{\lambda_y^2} \tau \mathbb{E} \left\| y^{(t+1)} - y^{(t)} \right\|^2
\end{aligned}$$

where (a) comes from Lipschitz continuity of  $f$  and the L-smoothness of  $f_i(x, y)$  and  $g_i(x, y)$  in Assumption 3.1 (ii) and Assumption 3.2 (i) and (b) comes from Assumption 3.2 (iii). Then we have

$$\begin{aligned}
\Delta_y^{(t)} &\leq 2\eta_y^2 \tau \left( \frac{\delta_f^2}{c_t^2} + \delta_g^2 \right) + 12L_2^2\tau\eta_y^2\Delta_x^{(t)} + 12\left(\frac{1}{\gamma^2} + L_2^2\right)\tau\eta_y^2\Delta_y^{(t)} + \frac{12}{\gamma^2}\tau\eta_y^2\Delta_\theta^{(t)} \\
&\quad + 18\tau\eta_y^2 \frac{L_f^2}{c_t^2} + 8\eta_y^2\tau\Delta^2 + 16 \frac{\eta_y^2}{\lambda_y^2} \tau \mathbb{E} \left\| y^{(t+1)} - y^{(t)} \right\|^2. \tag{80}
\end{aligned}$$

Recall the inequality (78),

$$\begin{aligned}
\Delta_\theta^{(t)} &\leq \eta_\theta^2 \tau \delta_g^2 + 6\tau\eta_\theta^2 L_2^2 \Delta_x^{(t)} + 6\tau\eta_\theta^2 (L_2^2 + \frac{1}{\gamma^2}) \Delta_\theta^{(t)} + 6\tau\eta_\theta^2 \frac{1}{\gamma^2} \Delta_y^{(t)} + 4\eta_\theta^2 \tau \Delta^2 \\
&\quad + 4\tau\eta_\theta^2 \left( \frac{1}{\gamma^2} + L_2^2 \right) \mathbb{E} \left\| \theta^{(t)} - \theta_\gamma^*(x^{(t)}, y^{(t)}) \right\|^2.
\end{aligned}$$

Next, we will prove that  $\Delta_y^{(t)}$  and  $\Delta_\theta^{(t)}$  can be bounded by  $\mathbb{E} \left\| \theta^{(t)} - \theta_\gamma^*(x^{(t)}, y^{(t)}) \right\|^2$  and some constants. Let's reformulate the inequalities (78) and (80),

$$\begin{aligned} (1 - 6\tau\eta_\theta^2(L_2^2 + \frac{1}{\gamma^2}))\Delta_\theta^{(t)} &\leq \eta_\theta^2\tau\delta_g^2 + 6\tau\eta_\theta^2L_2^2\Delta_x^{(t)} + 6\tau\eta_\theta^2\frac{1}{\gamma^2}\Delta_y^{(t)} + 4\eta_\theta^2\tau\Delta^2 \\ &\quad + 4\tau\eta_\theta^2(\frac{1}{\gamma^2} + L_2^2)\mathbb{E} \left\| \theta^{(t)} - \theta_\gamma^*(x^{(t)}, y^{(t)}) \right\|^2, \end{aligned} \quad (81)$$

and

$$\begin{aligned} (1 - 12\tau\eta_y^2(\frac{1}{\gamma^2} + L_2^2))\Delta_y^{(t)} &\leq 2\eta_y^2\tau(\frac{\delta_f^2}{c_t^2} + \delta_g^2) + 12L_2^2\tau\eta_y^2\Delta_x^{(t)} + \frac{12}{\gamma^2}\tau\eta_y^2\Delta_\theta^{(t)} \\ &\quad + 18\tau\eta_y^2\frac{L_f^2}{c_t^2} + 8\eta_y^2\tau\Delta^2 + 16\frac{\eta_y^2}{\lambda_y^2}\tau\mathbb{E} \left\| y^{(t+1)} - y^{(t)} \right\|^2 \end{aligned} \quad (82)$$

Multiply both sides of the inequality (81) by  $\frac{(1 - 12\tau\eta_y^2(\frac{1}{\gamma^2} + L_2^2))}{6\tau\eta_\theta^2\frac{1}{\gamma^2}}$ ,

$$\begin{aligned} &\frac{(1 - 12\tau\eta_y^2(\frac{1}{\gamma^2} + L_2^2))}{6\tau\eta_\theta^2\frac{1}{\gamma^2}}(1 - 6\tau\eta_\theta^2(L_2^2 + \frac{1}{\gamma^2}))\Delta_\theta^{(t)} \\ &\leq \frac{(1 - 12\tau\eta_y^2(\frac{1}{\gamma^2} + L_2^2))}{6\tau\eta_\theta^2\frac{1}{\gamma^2}} \left( \eta_\theta^2\tau\delta_g^2 + 6\tau\eta_\theta^2L_2^2\Delta_x^{(t)} + 6\tau\eta_\theta^2\frac{1}{\gamma^2}\Delta_y^{(t)} + 4\eta_\theta^2\tau\Delta^2 \right. \\ &\quad \left. + 4\tau\eta_\theta^2(\frac{1}{\gamma^2} + L_2^2)\mathbb{E} \left\| \theta^{(t)} - \theta_\gamma^*(x^{(t)}, y^{(t)}) \right\|^2 \right). \end{aligned} \quad (83)$$

Adding the inequality (83) to the inequality (82), we get:

$$\begin{aligned} &\frac{(1 - 12\tau\eta_y^2(\frac{1}{\gamma^2} + L_2^2))}{6\tau\eta_\theta^2\frac{1}{\gamma^2}}(1 - 6\tau\eta_\theta^2(L_2^2 + \frac{1}{\gamma^2}))\Delta_\theta^{(t)} \\ &\leq \frac{(1 - 12\tau\eta_y^2(\frac{1}{\gamma^2} + L_2^2))}{6\tau\eta_\theta^2\frac{1}{\gamma^2}} \left( \eta_\theta^2\tau\delta_g^2 + 6\tau\eta_\theta^2L_2^2\Delta_x^{(t)} + 4\eta_\theta^2\tau\Delta^2 \right. \\ &\quad \left. + 4\tau\eta_\theta^2(\frac{1}{\gamma^2} + L_2^2)\mathbb{E} \left\| \theta^{(t)} - \theta_\gamma^*(x^{(t)}, y^{(t)}) \right\|^2 \right) \\ &\quad + 2\eta_y^2\tau(\frac{\delta_f^2}{c_t^2} + \delta_g^2) + 12L_2^2\tau\eta_y^2\Delta_x^{(t)} \\ &\quad + 18\tau\eta_y^2\frac{L_f^2}{c_t^2} + 8\eta_y^2\tau\Delta^2 + 16\frac{\eta_y^2}{\lambda_y^2}\tau\mathbb{E} \left\| y^{(t+1)} - y^{(t)} \right\|^2. \end{aligned}$$

Then we can obtain

$$\begin{aligned} \Delta_\theta^{(t)} &\leq \frac{1}{\frac{(1 - 12\tau\eta_y^2(\frac{1}{\gamma^2} + L_2^2))}{6\tau\eta_\theta^2\frac{1}{\gamma^2}}(1 - 6\tau\eta_\theta^2(L_2^2 + \frac{1}{\gamma^2})) - 12\tau\eta_y^2(\frac{1}{\gamma^2} + L_2^2)} \left\{ \frac{(1 - 12\tau\eta_y^2(\frac{1}{\gamma^2} + L_2^2))}{6\tau\eta_\theta^2\frac{1}{\gamma^2}} \left( \eta_\theta^2\tau\delta_g^2 \right. \right. \\ &\quad \left. \left. + 6\tau\eta_\theta^2L_2^2\Delta_x^{(t)} + 4\eta_\theta^2\tau\Delta^2 + 4\tau\eta_\theta^2(\frac{1}{\gamma^2} + L_2^2)\mathbb{E} \left\| \theta^{(t)} - \theta_\gamma^*(x^{(t)}, y^{(t)}) \right\|^2 \right) \right. \\ &\quad \left. + 2\eta_y^2\tau(\frac{\delta_f^2}{c_t^2} + \delta_g^2) + 12L_2^2\tau\eta_y^2\Delta_x^{(t)} + 18\tau\eta_y^2\frac{L_f^2}{c_t^2} + 8\eta_y^2\tau\Delta^2 + 16\frac{\eta_y^2}{\lambda_y^2}\tau\mathbb{E} \left\| y^{(t+1)} - y^{(t)} \right\|^2 \right\}. \end{aligned} \quad (84)$$

Then the  $\Delta_\theta^{(t)}$  is bounded by  $\mathbb{E} \left\| \theta^{(t)} - \theta_\gamma^*(x^{(t)}, y^{(t)}) \right\|^2$  and some constants. Similarly, we can bound the  $\Delta_y^{(t)}$ ,

$$\begin{aligned} \Delta_y^{(t)} \leq & \frac{1}{(1 - 12\tau\eta_y^2(\frac{1}{\gamma^2} + L_2^2)) - \frac{12\tau\eta_y^2(\frac{1}{\gamma^2} + L_2^2)}{(1 - 6\tau\eta_\theta^2(L_2^2 + \frac{1}{\gamma^2}))} 6\tau\eta_\theta^2 \frac{1}{\gamma^2}} \left\{ \frac{12\tau\eta_y^2(\frac{1}{\gamma^2} + L_2^2)}{(1 - 6\tau\eta_\theta^2(L_2^2 + \frac{1}{\gamma^2}))} \left( \eta_\theta^2 \tau \delta_g^2 \right. \right. \\ & + 6\tau\eta_\theta^2 L_2^2 \Delta_x^{(t)} + 4\eta_\theta^2 \tau \Delta^2 + 4\tau\eta_\theta^2 \left( \frac{1}{\gamma^2} + L_2^2 \right) \mathbb{E} \left\| \theta^{(t)} - \theta_\gamma^*(x^{(t)}, y^{(t)}) \right\|^2 \Big) \\ & \left. + 2\eta_y^2 \tau \left( \frac{\delta_f^2}{c_t^2} + \delta_g^2 \right) + 12L_2^2 \tau \eta_y^2 \Delta_x^{(t)} + 18\tau\eta_y^2 \frac{L_f^2}{c_t^2} + 8\eta_y^2 \tau \Delta^2 + 16 \frac{\eta_y^2}{\lambda_y^2} \tau \mathbb{E} \left\| y^{(t+1)} - y^{(t)} \right\|^2 \right\}. \end{aligned} \quad (85)$$

□

Note that we can choose proper  $\eta_y, \eta_\theta$  to simplify inequalities (84) and (85). Now we choose  $\eta_y, \eta_\theta$  such that  $1 - 12\tau\eta_y^2(\frac{1}{\gamma^2} + L_2^2) \geq \frac{2}{3}, 1 - 6\tau\eta_\theta^2(L_2^2 + \frac{1}{\gamma^2}) \geq \frac{2}{3}$ , namely,  $12\tau\eta_y^2(\frac{1}{\gamma^2} + L_2^2) \leq \frac{1}{3}, 6\tau\eta_\theta^2(L_2^2 + \frac{1}{\gamma^2}) \leq \frac{1}{3}$ .

For the term  $\frac{1}{\frac{(1 - 12\tau\eta_y^2(\frac{1}{\gamma^2} + L_2^2))}{6\tau\eta_\theta^2 \frac{1}{\gamma^2}} (1 - 6\tau\eta_\theta^2(L_2^2 + \frac{1}{\gamma^2})) - 12\tau\eta_y^2(\frac{1}{\gamma^2} + L_2^2)}$  in inequality (84),

$$\begin{aligned} & \frac{1}{\frac{(1 - 12\tau\eta_y^2(\frac{1}{\gamma^2} + L_2^2))}{6\tau\eta_\theta^2 \frac{1}{\gamma^2}} (1 - 6\tau\eta_\theta^2(L_2^2 + \frac{1}{\gamma^2})) - 12\tau\eta_y^2(\frac{1}{\gamma^2} + L_2^2)} \\ & \leq \frac{6\tau\eta_\theta^2 \frac{1}{\gamma^2} + 6\tau\eta_\theta^2 \frac{1}{\gamma^2}}{(1 - 12\tau\eta_y^2(\frac{1}{\gamma^2} + L_2^2))(1 - 6\tau\eta_\theta^2(L_2^2 + \frac{1}{\gamma^2})) - 12\tau\eta_y^2(\frac{1}{\gamma^2} + L_2^2) 6\tau\eta_\theta^2 \frac{1}{\gamma^2} + 6\tau\eta_\theta^2 \frac{1}{\gamma^2}} \\ & \leq \frac{12\tau\eta_\theta^2 \frac{1}{\gamma^2}}{(1 - 12\tau\eta_y^2(\frac{1}{\gamma^2} + L_2^2))(1 - 6\tau\eta_\theta^2(L_2^2))} \\ & \stackrel{(a)}{\leq} \frac{12\tau\eta_\theta^2(L_2^2 + \frac{1}{\gamma^2})}{(1 - 12\tau\eta_y^2(\frac{1}{\gamma^2} + L_2^2))(1 - 6\tau\eta_\theta^2(L_2^2 + \frac{1}{\gamma^2}))} \leq 2, \end{aligned}$$

where (a) comes from the inequalities  $1 - 12\tau\eta_y^2(\frac{1}{\gamma^2} + L_2^2) \geq \frac{2}{3}, 1 - 6\tau\eta_\theta^2(L_2^2 + \frac{1}{\gamma^2}) \geq \frac{2}{3}$ .

For the term  $\frac{1}{\frac{(1 - 12\tau\eta_y^2(\frac{1}{\gamma^2} + L_2^2))}{6\tau\eta_\theta^2 \frac{1}{\gamma^2}} (1 - 6\tau\eta_\theta^2(L_2^2 + \frac{1}{\gamma^2})) - 12\tau\eta_y^2(\frac{1}{\gamma^2} + L_2^2)}$  in inequality (84), we have

$$\begin{aligned} & \frac{1}{\frac{(1 - 12\tau\eta_y^2(\frac{1}{\gamma^2} + L_2^2))}{6\tau\eta_\theta^2 \frac{1}{\gamma^2}} (1 - 6\tau\eta_\theta^2(L_2^2 + \frac{1}{\gamma^2})) - 12\tau\eta_y^2(\frac{1}{\gamma^2} + L_2^2)} \frac{(1 - 12\tau\eta_y^2(\frac{1}{\gamma^2} + L_2^2))}{6\tau\eta_\theta^2 \frac{1}{\gamma^2}} \\ & \leq \frac{12\tau\eta_\theta^2 \frac{1}{\gamma^2}}{(1 - 12\tau\eta_y^2(\frac{1}{\gamma^2} + L_2^2))(1 - 6\tau\eta_\theta^2(L_2^2))} \frac{(1 - 12\tau\eta_y^2(\frac{1}{\gamma^2} + L_2^2))}{6\tau\eta_\theta^2 \frac{1}{\gamma^2}} \\ & \leq \frac{2}{1 - 6\tau\eta_\theta^2(L_2^2)} \leq \frac{2}{1 - 6\tau\eta_\theta^2(L_2^2 + \frac{1}{\gamma^2})} \leq 3. \end{aligned}$$

Then the  $\Delta_\theta^{(t)}$  can be simplified,

$$\begin{aligned} & \Delta_\theta^{(t)} \\ & \leq 2 \left\{ 3 \left( \eta_\theta^2 \tau \delta_g^2 + 6\tau\eta_\theta^2 L_2^2 \Delta_x^{(t)} + 4\eta_\theta^2 \tau \Delta^2 + 4\tau\eta_\theta^2 \left( \frac{1}{\gamma^2} + L_2^2 \right) \mathbb{E} \left\| \theta^{(t)} - \theta_\gamma^*(x^{(t)}, y^{(t)}) \right\|^2 \right) \right. \end{aligned}$$

$$\begin{aligned}
& + 2\eta_y^2 \tau \left( \frac{\delta_f^2}{c_t^2} + \delta_g^2 \right) + 12L_2^2 \tau \eta_y^2 \Delta_x^{(t)} + 18\tau \eta_y^2 \frac{L_f^2}{c_t^2} + 8\eta_y^2 \tau \Delta^2 + 16 \frac{\eta_y^2}{\lambda_y^2} \tau \mathbb{E} \left\| y^{(t+1)} - y^{(t)} \right\|^2 \Big\} \\
& \leq 6\eta_\theta^2 \tau \delta_g^2 + 36\tau \eta_\theta^2 L_2^2 \Delta_x^{(t)} + 24\eta_\theta^2 \tau \Delta^2 + 24\tau \eta_\theta^2 \left( \frac{1}{\gamma^2} + L_2^2 \right) \mathbb{E} \left\| \theta^{(t)} - \theta_\gamma^*(x^{(t)}, y^{(t)}) \right\|^2 \\
& \quad + 4\eta_y^2 \tau \left( \frac{\delta_f^2}{c_t^2} + \delta_g^2 \right) + 24L_2^2 \tau \eta_y^2 \Delta_x^{(t)} + 18\tau \eta_y^2 \frac{L_f^2}{c_t^2} + 16\eta_y^2 \tau \Delta^2 + 32 \frac{\eta_y^2}{\lambda_y^2} \tau \mathbb{E} \left\| y^{(t+1)} - y^{(t)} \right\|^2 \\
& \leq 32 \frac{\eta_y^2}{\lambda_y^2} \tau \left\| \mathbb{E}[y^{(t+1)} - y^{(t)}] \right\|^2 + 24 \left( \frac{1}{\gamma^2} + L_2^2 \right) \eta_\theta^2 \tau \mathbb{E} \left\| \theta^{(t)} - \theta_\gamma^*(x^{(t)}, y^{(t)}) \right\|^2 \\
& \quad + \tau \eta_\theta^2 \left( 6\delta_g^2 + 24\Delta^2 \right) + \tau \eta_y^2 \left( 8 \left( \frac{\delta_f^2}{c_t^2} + \delta_g^2 \right) + 18 \frac{L_f^2}{c_t^2} + 16\eta_y^2 \tau \Delta^2 \right) + \tau \eta_\theta^2 \tau \eta_x^2 \mathcal{O}(1). \tag{86}
\end{aligned}$$

For the term  $\frac{1}{(1-12\tau\eta_y^2(\frac{1}{\gamma^2}+L_2^2)) - \frac{12\tau\eta_y^2(\frac{1}{\gamma^2}+L_2^2)}{(1-6\tau\eta_\theta^2(L_2^2+\frac{1}{\gamma^2}))} 6\tau\eta_\theta^2 \frac{1}{\gamma^2}}$  in inequality (85),

$$\begin{aligned}
& \frac{1}{(1-12\tau\eta_y^2(\frac{1}{\gamma^2}+L_2^2)) - \frac{12\tau\eta_y^2(\frac{1}{\gamma^2}+L_2^2)}{(1-6\tau\eta_\theta^2(L_2^2+\frac{1}{\gamma^2}))} 6\tau\eta_\theta^2 \frac{1}{\gamma^2}} \\
& \leq \frac{1}{(1-12\tau\eta_y^2(\frac{1}{\gamma^2}+L_2^2)) - \frac{12\tau\eta_y^2(\frac{1}{\gamma^2}+L_2^2)}{(1-6\tau\eta_\theta^2(L_2^2+\frac{1}{\gamma^2}))} 6\tau\eta_\theta^2 \frac{1}{\gamma^2}} \\
& \leq \frac{1-6\tau\eta_\theta^2(L_2^2+\frac{1}{\gamma^2})}{(1-12\tau\eta_y^2(\frac{1}{\gamma^2}+L_2^2))(1-6\tau\eta_\theta^2(L_2^2+\frac{1}{\gamma^2})) - 12\tau\eta_y^2(\frac{1}{\gamma^2}+L_2^2)6\tau\eta_\theta^2 \frac{1}{\gamma^2}} \\
& \leq \frac{(1-12\tau\eta_y^2(\frac{1}{\gamma^2}+L_2^2))(1-6\tau\eta_\theta^2(L_2^2+\frac{1}{\gamma^2})) + 12\tau\eta_y^2(\frac{1}{\gamma^2}+L_2^2)(1-6\tau\eta_\theta^2(L_2^2+\frac{1}{\gamma^2}))}{(1-12\tau\eta_y^2(\frac{1}{\gamma^2}+L_2^2))(1-6\tau\eta_\theta^2(L_2^2+\frac{1}{\gamma^2})) - 12\tau\eta_y^2(\frac{1}{\gamma^2}+L_2^2)6\tau\eta_\theta^2 \frac{1}{\gamma^2}} \\
& \leq 1 + \frac{12\tau\eta_y^2(\frac{1}{\gamma^2}+L_2^2) + 12\tau\eta_y^2(\frac{1}{\gamma^2}+L_2^2)}{(1-12\tau\eta_y^2(\frac{1}{\gamma^2}+L_2^2))(1-6\tau\eta_\theta^2(L_2^2+\frac{1}{\gamma^2})) + 12\tau\eta_y^2(\frac{1}{\gamma^2}+L_2^2)(1-6\tau\eta_\theta^2(L_2^2+\frac{1}{\gamma^2}))} \\
& = 1 + \frac{24\tau\eta_y^2 \frac{1}{\gamma^2}}{1-6\tau\eta_\theta^2(L_2^2+\frac{1}{\gamma^2})} \leq 2.
\end{aligned}$$

Then we have

$$\begin{aligned}
& \Delta_y^{(t)} \\
& \leq 2 \left\{ \left( \eta_\theta^2 \tau \delta_g^2 + 6\tau \eta_\theta^2 L_2^2 \Delta_x^{(t)} + 4\eta_\theta^2 \tau \Delta^2 + 4\tau \eta_\theta^2 \left( \frac{1}{\gamma^2} + L_2^2 \right) \mathbb{E} \left\| \theta^{(t)} - \theta_\gamma^*(x^{(t)}, y^{(t)}) \right\|^2 \right) \right. \\
& \quad \left. + 2\eta_y^2 \tau \left( \frac{\delta_f^2}{c_t^2} + \delta_g^2 \right) + 12L_2^2 \tau \eta_y^2 \Delta_x^{(t)} + 18\tau \eta_y^2 \frac{L_f^2}{c_t^2} + 8\eta_y^2 \tau \Delta^2 + 16 \frac{\eta_y^2}{\lambda_y^2} \tau \mathbb{E} \left\| y^{(t+1)} - y^{(t)} \right\|^2 \right\} \\
& \leq 2\eta_\theta^2 \tau \delta_g^2 + 12\tau \eta_\theta^2 L_2^2 \Delta_x^{(t)} + 8\eta_\theta^2 \tau \Delta^2 + 8\tau \eta_\theta^2 \left( \frac{1}{\gamma^2} + L_2^2 \right) \mathbb{E} \left\| \theta^{(t)} - \theta_\gamma^*(x^{(t)}, y^{(t)}) \right\|^2 \\
& \quad + 4\eta_y^2 \tau \left( \frac{\delta_f^2}{c_t^2} + \delta_g^2 \right) + 24L_2^2 \tau \eta_y^2 \Delta_x^{(t)} + 36\tau \eta_y^2 \frac{L_f^2}{c_t^2} + 16\eta_y^2 \tau \Delta^2 + 16 \frac{\eta_y^2}{\lambda_y^2} \tau \mathbb{E} \left\| y^{(t+1)} - y^{(t)} \right\|^2 \\
& \leq 8\tau \eta_\theta^2 \left( \frac{1}{\gamma^2} + L_2^2 \right) \mathbb{E} \left\| \theta^{(t)} - \theta_\gamma^*(x^{(t)}, y^{(t)}) \right\|^2 + 16 \frac{\eta_y^2}{\lambda_y^2} \tau \mathbb{E} \left\| y^{(t+1)} - y^{(t)} \right\|^2 \\
& \quad + \tau \eta_\theta^2 \left( 2\delta_g^2 + 8\Delta^2 \right) + \tau \eta_y^2 \left( 4 \left( \frac{\delta_f^2}{c_t^2} + \delta_g^2 \right) + 18 \frac{L_f^2}{c_t^2} + 16\eta_y^2 \tau \Delta^2 \right) + \tau \eta_\theta^2 \tau \eta_x^2 \mathcal{O}(1). \tag{87}
\end{aligned}$$