Lemma D.8. *Fix the number of communication rounds T, local update rounds τ, 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:*

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

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

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

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

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

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 $\Delta_\theta^{(t)}$,

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

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

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} \Big\| \sum_{j=0}^{k-1} \widetilde{h}_{i,\theta}^{(t,j)} \Big\|^2 \\
&\leq \sum_{i=1}^{n} \frac{1}{n} \frac{{\eta_\theta}^2}{\tau} \sum_{k=0}^{\tau-1} \mathbb{E} \Big\| \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)}) \Big\|^2 \\
&\overset{(a)}{\leq} 2 \sum_{i=1}^{n} \frac{1}{n} \frac{{\eta_\theta}^2}{\tau} \sum_{k=0}^{\tau-1} \sum_{j=0}^{k-1} \Big(\mathbb{E} \Big\| \nabla_y g_i(x_i^{(t,j)}, \theta_i^{(t,j)}) + \frac{1}{\gamma}(\theta_i^{(t,j)} - y_i^{(t,j)}) \\
&\quad - \nabla_y g_i(x^{(t)}, \theta^{(t)}) - \frac{1}{\gamma}(\theta^{(t)} - y^{(t)}) \Big\|^2 \Big) \\
&\quad + 2 \sum_{i=1}^{n} \frac{1}{n} \frac{{\eta_\theta}^2}{\tau} \sum_{k=0}^{\tau-1} \sum_{j=0}^{k-1} \Big(\mathbb{E} \Big\| \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)}) \\
&\quad - \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)})) \\
&\quad - \frac{1}{\gamma}(\theta_\gamma^*(x^{(t)}, y^{(t)}) - y^{(t)}) \Big\|^2 \Big) \\
&\overset{(b)}{\leq} 6 L_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} \Big\| x_i^{(t,j)} - x^{(t)} \Big\|^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} \Big\| y_i^{(t,j)} - y^{(t)} \Big\|^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} \Big\| \theta_i^{(t,j)} - \theta^{(t)} \Big\|^2 + 4 {\eta_\theta}^2 \tau \Delta^2 \\
&\quad + 4 \tau {\eta_\theta}^2 (\frac{1}{\gamma^2} + L_2^2) \mathbb{E} \Big\| \theta^{(t)} - \theta_\gamma^*(x^{(t)}, y^{(t)}) \Big\|^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} \Big\| \theta^{(t)} - \theta_\gamma^*(x^{(t)}, y^{(t)}) \Big\|^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

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

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

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

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

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} {\eta_y}^2 \frac{1}{\tau} \sum_{k=0}^{\tau-1} \mathbb{E} \Big\| \sum_{j=0}^{k-1} \widetilde{h}_{i,y}^{(t,j)} \Big\|^2 \\
\leq & \sum_{i=1}^{n} \frac{1}{n} \frac{{\eta_y}^2}{\tau} \sum_{k=0}^{\tau-1} \mathbb{E} \Big\| \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)}) \Big\|^2 \\
\leq & 4 \sum_{i=1}^{n} \frac{1}{n} \frac{{\eta_y}^2}{\tau} \sum_{k=0}^{\tau-1} \mathbb{E} \Big\| \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)}) - \big(\nabla_y g_i(x^{(t)}, y^{(t)}) - \frac{1}{\gamma}(y^{(t)} - \theta^{(t)}) \big) \Big\|^2 \\
& + 2 \sum_{i=1}^{n} \frac{1}{n} \frac{{\eta_y}^2}{\tau} \sum_{k=0}^{\tau-1} \mathbb{E} \Big\| \sum_{j=0}^{k-1} \frac{1}{c_t} \nabla_y f_i(x_i^{(t,j)}, y_i^{(t,j)}) \Big\|^2 + 4 \sum_{i=1}^{n} \frac{1}{n} \frac{{\eta_y}^2}{\tau} \sum_{k=0}^{\tau-1} \mathbb{E} \Big\| \nabla_y g_i(x^{(t)}, y^{(t)}) - \frac{1}{\gamma}(y^{(t)} - \theta^{(t)}) \Big\|^2 \\
\overset{(a)}{\leq} & 2\tau {\eta_y}^2 \frac{L_f^2}{c_t^2} + 12 L_2^2 \tau {\eta_y}^2 \Delta_x^{(t)} + 12 \big(\frac{1}{\gamma^2} + L_2^2 \big) \tau {\eta_y}^2 \Delta_y^{(t)} + \frac{12}{\gamma^2} \tau {\eta_y}^2 \Delta_\theta^{(t)} \\
& + 8 \sum_{i=1}^{n} \frac{{\eta_y}^2}{\tau} \sum_{k=0}^{\tau-1} \mathbb{E} \Big\| \frac{1}{n} \sum_{i=1}^{n} \nabla_y g_i(x^{(t)}, y^{(t)}) + \frac{1}{\gamma}(y^{(t)} - \theta^{(t)}) \Big\|^2 \\
& + 8 \sum_{i=1}^{n} \frac{1}{n} \frac{{\eta_y}^2}{\tau} \sum_{k=0}^{\tau-1} \mathbb{E} \Big\| \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)}) \Big\|^2 \\
\overset{(b)}{\leq} & 2\tau {\eta_y}^2 \frac{L_f^2}{c_t^2} + 12 L_2^2 \tau {\eta_y}^2 \Delta_x^{(t)} + 12 \big(\frac{1}{\gamma^2} + L_2^2 \big) \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} \Big\| \frac{1}{n} \sum_{i=1}^{n} \nabla_y \frac{1}{c_t} f_i(x^{(t)}, y^{(t)}) \Big\|^2 \\
& + 8 \eta_y^2 \tau \Delta^2 + 16 \sum_{i=1}^{n} \frac{{\eta_y}^2}{\tau} \sum_{k=0}^{\tau-1} \mathbb{E} \Big\| \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)}) \Big\|^2, \\
\leq & 18 \tau {\eta_y}^2 \frac{L_f^2}{c_t^2} + 12 L_2^2 \tau {\eta_y}^2 \Delta_x^{(t)} + 12 \big(\frac{1}{\gamma^2} + L_2^2 \big) \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 \Big\| \mathbb{E}[y^{(t+1)} - y^{(t)}] \Big\|^2
\end{aligned}
$$

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

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

Recall the inequality (78),

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

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),

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

and

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

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

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

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

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

Then we can obtain

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

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 $\Delta_y^{(t)}$,

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

□

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

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

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

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

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

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

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

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

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

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

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

Then we have

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