Supplementary Materials

Appendix G
Proofs for Results in Section III

A. Proof of Theorem 2

Following the similar proof of Theorem 1, we first analyze the descent between each outer iteration. Notice throughout the proof, we assume that $p=0$, that is, there is no delayed communication. It follows that the following holds:

$$\mathbf{x}_{0,i}^{r+1} = \frac{1}{N}\sum_{j=1}^{N}\mathbf{x}_{0,j}^{r+}, \quad \forall i=1,\ldots,N.$$

We also recall that r is the (outer) stage index, and q is the local update index. First we provide a series of lemmas.

Lemma 9. *Under Assumption 1, consider FedPD with Algorithm 4 (Oracle II) as the update rule. The difference of the local AL is bounded by* (92).

Then we deal with the variance of the stochastic gradients.

Lemma 10. *Suppose A1 holds and the samples are randomly sampled according to* (7)*, consider FedPD with Algorithm 4 (Oracle II) as the update rule. The expected norm square of the difference between $g_i^{r,q+1}$ and $\nabla f_i(\mathbf{x}_i^{r,q+1})$ is bounded by*

$$\mathbb{E}\left\|g_i^{r,q+1}-\nabla f_i(\mathbf{x}_i^{r,q+1})\right\|^2 \le \frac{L^2}{B}\sum_{\tau=\{r_0,1\}}^{\{r,q+1\}}\mathbb{E}\left\|\mathbf{x}_i^{\tau}-\mathbf{x}_i^{\tau-1}\right\|^2. \tag{93}$$

Lastly we upper bound the original loss function.

Lemma 11. *Under A1 and A2, the difference between the original loss and the AL is bounded as below:*

$$\begin{aligned}
&\mathbb{E}\, f(\mathbf{x}_0^r)\\
&\le\ \mathbb{E}\,\mathcal{L}(\mathbf{x}_0^r,\mathbf{x}_1^r,\ldots,\mathbf{x}_N^r,\lambda_1^r,\ldots,\lambda_N^r) - \frac{1-3L\eta}{2N\eta}\sum_{i=1}^{N}\mathbb{E}\left\|\mathbf{x}_i^r-\mathbf{x}_0^r\right\|^2\\
&+\frac{(1+L\gamma)^2+L^2\gamma^2}{4L\gamma^2}\Big[\frac{1}{B}\sum_{\tau=\{r_0,1\}}^{\{r-1,Q-1\}}\mathbb{E}\left\|\mathbf{x}_i^{\tau}-\mathbf{x}_i^{\tau-1}\right\|^2\\
&+\mathbb{E}\left\|\mathbf{x}_i^{r-1,Q}-\mathbf{x}_i^{r-1,Q-1}\right\|^2\Big].
\end{aligned}\tag{94}$$

B. Proof of Lemma 9

Let us first express the difference of the local AL as:

$$\begin{aligned}
\mathcal{L}_i^{r+1}-\mathcal{L}_i^{r} &= \mathcal{L}_i(\mathbf{x}_i^{r+1},\mathbf{x}_{0,i}^r,\lambda_i^r)-\mathcal{L}_i^r+\mathcal{L}_i(\mathbf{x}_i^{r+1},\mathbf{x}_{0,i}^r,\lambda_i^{r+1})\\
&\quad -\mathcal{L}_i(\mathbf{x}_i^{r+1},\mathbf{x}_{0,i}^r,\lambda_i^r)\ +\mathcal{L}_i^{r+1}-\mathcal{L}_i(\mathbf{x}_i^{r+1},\mathbf{x}_{0,i}^r,\lambda_i^{r+1}),
\end{aligned}\tag{95}$$

where the above three differences respectively correspond to the three steps in the algorithm's update steps.

Let us bound the above three differences one by one. First, note that we have the following decomposition (by using the fact that $\mathbf{x}_i^{r,Q+1}=\mathbf{x}_i^{r+1}$ and $\mathbf{x}_i^{r,1}=\mathbf{x}_i^r$):

$$\begin{aligned}
&\mathcal{L}_i(\mathbf{x}_i^{r+1},\mathbf{x}_{0,i}^r,\lambda_i^r)-\mathcal{L}_i^r\\
&=\sum_{q=1}^{Q}\left(\mathcal{L}_i(\mathbf{x}_i^{r,q+1},\mathbf{x}_{0,i}^r,\lambda_i^r)-\mathcal{L}_i(\mathbf{x}_i^{r,q},\mathbf{x}_{0,i}^r,\lambda_i^r)\right).
\end{aligned}\tag{96}$$

Each term on the right hand side (RHS) of the above equality can be bounded by (see a similar arguments in (46)):

$$\begin{aligned}
&\mathcal{L}_i(\mathbf{x}_i^{r,q+1},\mathbf{x}_{0,i}^r,\lambda_i^r)-\mathcal{L}_i(\mathbf{x}_i^{r,q},\mathbf{x}_{0,i}^r,\lambda_i^r)\\
&\le\left\langle\nabla f_i(\mathbf{x}_i^{r,q})+\lambda_i^r+\frac{1}{\eta}(\mathbf{x}^{r,q+1}-\mathbf{x}_{0,i}^r),\mathbf{x}_i^{r,q+1}-\mathbf{x}_i^{r,q}\right\rangle\\
&-\frac{1-L\eta}{2\eta}\left\|\mathbf{x}_i^{r,q+1}-\mathbf{x}_i^{r,q}\right\|^2\\
&\overset{(a)}{=}\left\langle\nabla f_i(\mathbf{x}_i^{r,q})-g_i^{r,q}-\frac{1}{\gamma}(\mathbf{x}^{r,q+1}-\mathbf{x}_i^{r,q}),\mathbf{x}_i^{r,q+1}-\mathbf{x}_i^{r,q}\right\rangle\\
&-(\frac{1}{2\eta}-\frac{L}{2})\left\|\mathbf{x}_i^{r,q+1}-\mathbf{x}_i^{r,q}\right\|^2\\
&=\ \left\langle\nabla f_i(\mathbf{x}_i^{r,q})-g_i^{r,q},\mathbf{x}_i^{r,q+1}-\mathbf{x}_i^{r,q}\right\rangle-\\
&(\frac{1}{2\eta}+\frac{1}{\gamma}-\frac{L}{2})\left\|\mathbf{x}_i^{r,q+1}-\mathbf{x}_i^{r,q}\right\|^2\\
&\overset{(b)}{\le}\ \frac{1}{2L}\left\|\nabla f_i(\mathbf{x}_i^{r,q})-g_i^{r,q}\right\|^2-(\frac{1}{2\eta}+\frac{1}{\gamma}-L)\left\|\mathbf{x}_i^{r,q+1}-\mathbf{x}_i^{r,q}\right\|^2,
\end{aligned}$$

where in (a) we use the optimal condition that $\nabla_{\mathbf{x}_i}\tilde{\mathcal{L}}_i(\mathbf{x}_i^{r,q+1},\mathbf{x}_{0,i}^r,\lambda_i^r;\mathbf{x}_i^{r,q},g_i^{r,q})=0$ which gives us the following relation

$$\lambda_i^r+\frac{1}{\eta}(\mathbf{x}_i^{r,q+1}-\mathbf{x}_{0,i}^r)+g_i^{r,q}+\frac{1}{\gamma}(\mathbf{x}_i^{r,q+1}-\mathbf{x}_i^{r,q})=0; \tag{97}$$

in (b) we use the fact that $2\langle a,b\rangle\le L\|a\|^2+\frac{1}{L}\|b\|^2$. Therefore, the first difference in the RHS of (95) is given by

$$\begin{aligned}
\mathcal{L}_i(\mathbf{x}_i^{r+1},\mathbf{x}_{0,i}^r,\lambda_i^r)-\mathcal{L}_i^r &\le \frac{1}{2L}\sum_{q=1}^{Q}\left\|\nabla f_i(\mathbf{x}_i^{r,q})-g_i^{r,q}\right\|^2\\
&\quad -(\frac{1}{2\eta}+\frac{1}{\gamma}-L)\sum_{q=1}^{Q}\left\|\mathbf{x}_i^{r,q+1}-\mathbf{x}_i^{r,q}\right\|^2.
\end{aligned}\tag{98}$$

The other two differences in (95) can be expressed as:

$$\mathcal{L}_i(\mathbf{x}_i^{r+1},\mathbf{x}_{0,i}^r,\lambda_i^{r+1})-\mathcal{L}_i(\mathbf{x}_i^{r+1},\mathbf{x}_{0,i}^r,\lambda_i^r)=\eta\left\|\lambda_i^{r+1}-\lambda_i^r\right\|^2, \tag{99}$$

$$\begin{aligned}
\mathcal{L}_i^{r+1}-\mathcal{L}_i(\mathbf{x}_i^{r+1},\mathbf{x}_{0,i}^r,\lambda_i^{r+1}) &= -\frac{1}{2\eta}\left\|\mathbf{x}_{0,i}^{r+1}-\mathbf{x}_{0,i}^r\right\|^2\\
&+\left\langle\lambda_i^{r+1}+\frac{1}{\eta}(\mathbf{x}_i^{r+1}-\mathbf{x}_{0,i}^{r+1}),\mathbf{x}_{0,i}^{r+1}-\mathbf{x}_{0,i}^r\right\rangle.
\end{aligned}\tag{100}$$

Next we bound $\left\|\lambda_i^{r+1}-\lambda_i^r\right\|^2$. Notice that the from the update rule the following holds:

$$\lambda_i^{r+1}=\lambda_i^r+\frac{1}{\eta}(\mathbf{x}_i^{r,Q}-\mathbf{x}_{0,i}^r)\overset{(97)}{=}-\frac{1}{\gamma}(\mathbf{x}_i^{r,Q}-\mathbf{x}_i^{r,Q-1})-g^{r,Q-1}. \tag{101}$$

Using the above property, we have

$$\begin{aligned}
\left\|\lambda_i^{r+1}-\lambda_i^r\right\|^2 &= \Big\|\frac{1}{\gamma}(\mathbf{x}_i^{r,Q}-\mathbf{x}_i^{r,Q-1})+g_i^{r,Q-1}\\
&\quad -\frac{1}{\gamma}(\mathbf{x}_i^{r-1,Q}-\mathbf{x}_i^{r-1,Q-1})-g_i^{r-1,Q-1}\Big\|^2\\
&\overset{(a)}{\le} 3\left\|g_i^{r,Q-1}-g_i^{r-1,Q-1}\right\|^2+\frac{3}{\gamma^2}\left\|\mathbf{x}_i^{r,Q}-\mathbf{x}_i^{r,Q-1}\right\|^2\\
&\quad +\frac{3}{\gamma^2}\left\|\mathbf{x}_i^{r-1,Q}-\mathbf{x}_i^{r-1,Q-1}\right\|^2.
\end{aligned}\tag{102}$$

where in (a) we apply Cauchy-Schwarz inequality. Next we bound $\left\|g_i^{r,Q-1}-g_i^{r-1,Q-1}\right\|^2$ by (103), where in (a) and (b)

$$
\begin{aligned}
\mathcal{L}_i^{r+1}-\mathcal{L}_i^{r} \leq & -\frac{1}{2\eta}\left\|\mathbf{x}_{0,i}^{r+1}-\mathbf{x}_{0,i}^{r}\right\|^2-\left(\frac{1}{2\eta}+\frac{1}{\gamma}-L-\frac{3\eta}{\gamma^2}\right)\left\|\mathbf{x}_i^{r,Q}-\mathbf{x}_i^{r,Q-1}\right\|^2-(\frac{1}{2\eta}+\frac{1}{\gamma}-L-9Q^2L^2\eta)\sum_{q=1}^{Q-1}\left\|\mathbf{x}_i^{r,q}-\mathbf{x}_i^{r,q-1}\right\|^2 \\
& +\left(9Q^2L^2\eta+\frac{3\eta}{\gamma^2}\right)\left\|\mathbf{x}_i^{r-1,Q}-\mathbf{x}_i^{r-1,Q-1}\right\|^2+\frac{1}{2L}\sum_{q=0}^{Q-2}\|\nabla f_i(\mathbf{x}_i^{r,q})-g_i^{r,q}\|^2+\left\langle\lambda_i^{r+1}+\frac{1}{\eta}(\mathbf{x}_i^{r+1}-\mathbf{x}_{0,i}^{r+1}),\mathbf{x}_{0,i}^{r+1}-\mathbf{x}_{0,i}^{r}\right\rangle \\
& +(\frac{1}{2L}+9\eta)\left\|g_i^{r,Q-1}-\nabla f_i(\mathbf{x}_i^{r,Q-1})\right\|^2+9\eta\left\|g_i^{r-1,Q-1}-\nabla f_i(\mathbf{x}_i^{r-1,Q-1})\right\|^2
\end{aligned}
\tag{92}
$$

$$
\begin{aligned}
\left\|g_i^{r,Q-1}-g_i^{r-1,Q-1}\right\|^2 = & \left\|g_i^{r,Q-1}-\nabla f_i(\mathbf{x}_i^{r,Q-1})+\nabla f_i(\mathbf{x}_i^{r,Q-1})-\nabla f_i(\mathbf{x}_i^{r-1,Q-1})+\nabla f_i(\mathbf{x}_i^{r-1,Q-1})-g_i^{r-1,Q-1}\right\|^2 \qquad (103)\\
\overset{(a)}{\leq} & \; 3\left\|g_i^{r,Q-1}-\nabla f_i(\mathbf{x}_i^{r,Q-1})\right\|^2+3\left\|g_i^{r-1,Q-1}-\nabla f_i(\mathbf{x}_i^{r-1,Q-1})\right\|^2+3L^2\left\|\mathbf{x}_i^{r,Q-1}-\mathbf{x}_i^{r-1,Q-1}\right\|^2 \\
\overset{(b)}{\leq} & \; 3\left\|g_i^{r,Q-1}-\nabla f_i(\mathbf{x}_i^{r,Q-1})\right\|^2+3\left\|g_i^{r-1,Q-1}-\nabla f_i(\mathbf{x}_i^{r-1,Q-1})\right\|^2+3Q^2L^2\sum_{q=1}^{Q-1}\left\|\mathbf{x}_i^{r,q}-\mathbf{x}_i^{r,q-1}\right\|^2+3Q^2L^2\left\|\mathbf{x}_i^{r-1,Q}-\mathbf{x}_i^{r-1,Q-1}\right\|^2,
\end{aligned}
$$

we both apply Cauchy-Schwarz inequality, in (a) we use A1 to the last term and in (b) we notice $\mathbf{x}_i^{r-1,Q}=\mathbf{x}_i^{r,0}$.

Substitute (103) to (102) and sum the three parts, we have (104), which complete the proof of Lemma 9.

C. Proof of Lemma 10

To study $\mathbb{E}\|g_i^{r,q}-\nabla f_i(\mathbf{x}_i^{r,q})\|^2$, we denote the latest iteration before r that computes full gradients as r_0. That is, in r_0 we have $g_i^{r_0,0}=\nabla f_i(\mathbf{x}_i^{r_0,0})$. By the description of the algorithm we know

$$r_0=kI,\;\; k\in\mathbb{N},\;\; rQ+q-r_0Q\leq IQ.$$

That is, r_0 is a multiple of I and there is no more than IQ local update steps between step $\{r_0,0\}$ and step $\{r,q\}$. By the update rule of $g_i^{r,q}$, we have

$$
\begin{aligned}
& g_i^{r,q+1}-\nabla f_i(\mathbf{x}_i^{r,q+1}) \\
& =g_i^{r,q}-\nabla f_i(\mathbf{x}_i^{r,q+1})+\frac{1}{B}\sum_{b=1}^{B}(h_i(\mathbf{x}_i^{r,q+1};\xi_{i,b}^{r,q})-h_i(\mathbf{x}_i^{r,q};\xi_{i,b}^{r,q})).
\end{aligned}
\tag{105}
$$

Take expectation on both sides, we have

$$
\begin{aligned}
& \mathbb{E}_{\{\xi_{i,b}^{r,q}\}_{b=1}^B}[g_i^{r,q+1}-\nabla f_i(\mathbf{x}_i^{r,q+1})] \\
& =g_i^{r,q}-\nabla f_i(\mathbf{x}_i^{r,q+1}) \\
& \quad+\mathbb{E}_{\{\xi_{i,b}^{r,q}\}_{b=1}^B}[\frac{1}{B}\sum_{b=1}^{B}(h_i(\mathbf{x}_i^{r,q+1};\xi_{i,b}^{r,q})-h_i(\mathbf{x}_i^{r,q};\xi_{i,b}^{r,q}))] \\
& = g_i^{r,q}-\nabla f_i(\mathbf{x}_i^{r,q+1})+\nabla f_i(\mathbf{x}_i^{r,q+1})-\nabla f_i(\mathbf{x}_i^{r,q}) \\
& = g_i^{r,q}-\nabla f_i(\mathbf{x}_i^{r,q}).
\end{aligned}
\tag{106}
$$

By using the fact that $\mathbb{E}[X^2]=[\mathbb{E}X]^2+\mathbb{E}[[X-\mathbb{E}X]^2]$ and substitute (106) we obtain (107), where (a) comes form the fact that we view $h_i(\mathbf{x}_i^{r,q+1};\xi_{i,b}^{r,q})-h_i(\mathbf{x}_i^{r,q};\xi_{i,b}^{r,q})$ as X and by identically random sampling strategy we have $\mathbb{E}X=\nabla f_i(\mathbf{x}_i^{r,q+1})-\nabla f_i(\mathbf{x}_i^{r,q})$ and $\mathbb{E}[[X-\mathbb{E}X]^2\leq\mathbb{E}[X]^2$, in (b) we use A1.

Iteratively taking expectation until $\{r,q\}=\{r_0,0\}$, we have

$$\mathbb{E}\left\|g_i^{r,q+1}-\nabla f_i(\mathbf{x}_i^{r,q+1})\right\|^2\leq\frac{L^2}{B}\sum_{\tau=\{r_0,1\}}^{\{r,q+1\}}\mathbb{E}\left\|\mathbf{x}_i^{\tau}-\mathbf{x}_i^{\tau-1}\right\|^2, \tag{108}$$

which completes the proof.

D. Proof of Lemma 11

Applying A1, we have

$$
\begin{aligned}
f_i(\mathbf{x}_0^r) \leq & f_i(\mathbf{x}_i^r)+\langle\nabla f_i(\mathbf{x}_i^r),\mathbf{x}_0^r-\mathbf{x}_i^r\rangle+\frac{L}{2}\left\|\mathbf{x}_0^r-\mathbf{x}_i^r\right\|^2 \\
= & \mathcal{L}_i(\mathbf{x}_i^r,\mathbf{x}_0^r,\lambda_i^r)-\langle\nabla f_i(\mathbf{x}_i^r)+\lambda_i^r,\mathbf{x}_0^r-\mathbf{x}_i^r\rangle \\
& -\frac{1-L\eta}{2\eta}\left\|\mathbf{x}_0^r-\mathbf{x}_i^r\right\|^2 \\
\leq & \mathcal{L}_i(\mathbf{x}_i^r,\mathbf{x}_0^r,\lambda_i^r)+\frac{1}{4L}\left\|\nabla f_i(\mathbf{x}_i^r)+\lambda_i^r\right\|^2 \\
& -\frac{1-3L\eta}{2\eta}\left\|\mathbf{x}_0^r-\mathbf{x}_i^r\right\|^2.
\end{aligned}
\tag{109}
$$

Then notice $\mathbf{x}_i^r=\mathbf{x}_i^{r-1,Q}$ and apply (101), we can bound $\mathbb{E}\left\|\nabla f_i(\mathbf{x}_i^r)+\lambda_i^r\right\|^2$ by the following:

$$
\begin{aligned}
& \mathbb{E}\left\|\nabla f_i(\mathbf{x}_i^r)+\lambda_i^r\right\|^2 \\
& \overset{(101)}{=} \mathbb{E}\left\|\nabla f_i(\mathbf{x}_i^{r-1,Q})-g_i^{r-1,Q-1}-\frac{1}{\gamma}(\mathbf{x}_i^{r-1,Q}-\mathbf{x}_i^{r-1,Q-1})\right\|^2 \\
& \overset{(a)}{\leq} (1+\frac{(1+L\gamma)^2}{L^2\gamma^2})\,\mathbb{E}\left\|\nabla f_i(\mathbf{x}_i^{r-1,Q-1})-g_i^{r-1,Q-1}\right\|^2 \\
& +(1+\frac{L^2\gamma^2}{(1+L\gamma)^2})(1+\frac{1}{L\gamma})\,\mathbb{E}\left\|\nabla f_i(\mathbf{x}_i^{r-1,Q})-\nabla f_i(\mathbf{x}_i^{r-1,Q-1})\right\|^2 \\
& +\frac{(1+\frac{L^2\gamma^2}{(1+L\gamma)^2})(1+L\gamma)}{\gamma^2}\,\mathbb{E}\left\|\mathbf{x}_i^{r-1,Q}-\mathbf{x}_i^{r-1,Q-1}\right\|^2 \\
& \overset{(b)}{\leq} \frac{(1+L\gamma)^2+L^2\gamma^2}{B\gamma^2}\sum_{\tau=\{r_0,1\}}^{\{r-1,Q-1\}}\mathbb{E}\left\|\mathbf{x}_i^{\tau}-\mathbf{x}_i^{\tau-1}\right\|^2 \\
& +(1+\frac{L^2\gamma^2}{(1+L\gamma)^2})\left((1+\frac{1}{L\gamma})L^2+\frac{1+L\gamma}{\gamma^2}\right) \\
& \times\mathbb{E}\left\|\mathbf{x}_i^{r-1,Q}-\mathbf{x}_i^{r-1,Q-1}\right\|^2 \\
& = \frac{(1+L\gamma)^2+L^2\gamma^2}{B\gamma^2}\sum_{\tau=\{r_0,1\}}^{\{r-1,Q-1\}}\mathbb{E}\left\|\mathbf{x}_i^{\tau}-\mathbf{x}_i^{\tau-1}\right\|^2 \\
& +\frac{(1+L\gamma)^2+L^2\gamma^2}{\gamma^2}\,\mathbb{E}\left\|\mathbf{x}_i^{r-1,Q}-\mathbf{x}_i^{r-1,Q-1}\right\|^2,
\end{aligned}
\tag{110}
$$

$$\mathcal{L}_i^{r+1}-\mathcal{L}_i^r \leq -\frac{1}{2\eta}\left\|\mathbf{x}_{0,i}^{r+1}-\mathbf{x}_{0,i}^r\right\|^2-(\frac{1}{2\eta}+\frac{1}{\gamma}-L-\frac{3\eta}{\gamma^2})\left\|\mathbf{x}_i^{r,Q}-\mathbf{x}_i^{r,Q-1}\right\|^2-(\frac{1}{2\eta}+\frac{1}{\gamma}-L-9Q^2L^2\eta)\sum_{q=1}^{Q-1}\left\|\mathbf{x}_i^{r,q}-\mathbf{x}_i^{r,q-1}\right\|^2$$
$$+(9Q^2L^2\eta+\frac{3\eta}{\gamma^2})\left\|\mathbf{x}_i^{r-1,Q}-\mathbf{x}_i^{r-1,Q-1}\right\|^2+\frac{1}{2L}\sum_{q=0}^{Q-2}\|\nabla f_i(\mathbf{x}_i^{r,q})-g_i^{r,q}\|^2+\left\langle\lambda_i^{r+1}+\frac{1}{\eta}(\mathbf{x}_i^{r+1}-\mathbf{x}_{0,i}^{r+1}),\mathbf{x}_{0,i}^{r+1}-\mathbf{x}_{0,i}^r\right\rangle$$
$$+(\frac{1}{2L}+9\eta)\left\|g_i^{r,Q-1}-\nabla f_i(\mathbf{x}_i^{r,Q-1})\right\|^2+9\eta\left\|g_i^{r-1,Q-1}-\nabla f_i(\mathbf{x}_i^{r-1,Q-1})\right\|^2 \tag{104}$$

$$\mathbb{E}_{\{\xi_{i,b}^{r,q}\}_{b=1}^B}\left\|g_i^{r,q+1}-\nabla f_i(\mathbf{x}_i^{r,q+1})\right\|^2 \tag{107}$$
$$=\left\|\mathbb{E}_{\{\xi_{i,b}^{r,q}\}_{b=1}^B}[g_i^{r,q+1}-\nabla f_i(\mathbf{x}_i^{r,q+1})]\right\|^2+\mathbb{E}_{\{\xi_{i,b}^{r,q}\}_{b=1}^B}\left\|g_i^{r,q+1}-\nabla f_i(\mathbf{x}_i^{r,q+1})-\mathbb{E}_{\{\xi_{i,b}^{r,q}\}_{b=1}^B}[g_i^{r,q+1}-\nabla f_i(\mathbf{x}_i^{r,q+1})]\right\|^2$$
$$\overset{(106)}{=}\|g_i^{r,q}-\nabla f_i(\mathbf{x}_i^{r,q})\|^2+\mathbb{E}_{\{\xi_{i,b}^{r,q}\}_{b=1}^B}\left\|\frac{1}{B}\sum_{b=1}^B(h_i(\mathbf{x}_i^{r,q+1};\xi_{i,b}^{r,q}-h_i(\mathbf{x}_i^{r,q};\xi_{i,b}^{r,q}))-\nabla f_i(\mathbf{x}_i^{r,q+1})+\nabla f_i(\mathbf{x}_i^{r,q})\right\|^2$$
$$\overset{(a)}{\leq}\|g_i^{r,q}-\nabla f_i(\mathbf{x}_i^{r,q})\|^2+\frac{1}{B^2}\sum_{b=1}^B\mathbb{E}_{\{\xi_{i,b}^{r,q}\}_{b=1}^B}\left\|h_i(\mathbf{x}_i^{r,q+1};\xi_{i,b}^{r,q})-h_i(\mathbf{x}_i^{r,q};\xi_{i,b}^{r,q}))\right\|^2\overset{(b)}{\leq}\|g_i^{r,q}-\nabla f_i(\mathbf{x}_i^{r,q})\|^2+\frac{L^2}{B}\left\|\mathbf{x}_i^{r,q+1}-\mathbf{x}_i^{r,q}\right\|^2,.$$

where in (a) we apply Cauchy-Schwarz inequality twice:

$$\|x+y+z\|^2\leq(1+\frac{1}{a})\|x\|^2+(1+a)\|y+z\|^2$$
$$\leq(1+\frac{1}{a})\|x\|^2+(1+a)(1+b)\|y\|^2+(1+a)(1+\frac{1}{b})\|z\|^2;$$

in (b) we apply Lemma 10 to the first term and apply A1 to the second term.

Substitute (110) to (109) and average over the agents, Lemma 11 is proved.

E. Proof of Theorem 2

By the update step of $\mathbf{x}_0^r$, following (28) we have

$$\left\|\frac{1}{N}\sum_{i=1}^N\nabla_{\mathbf{x}_{0,i}}\mathcal{L}_i(\mathbf{x}_i^r,\mathbf{x}_{0,i}^r,\lambda_i^r)\right\|$$
$$=\left\|\frac{1}{N}\sum_{i=1}^N(\frac{1}{\eta}(\mathbf{x}_i^r-\mathbf{x}_{0,i}^r)+\lambda_i^r)\right\|=0.$$

We also have (111) where in (a), the first term is obtained by plugging in (101) given below

$$\lambda_i^r=-g_i^{r,0}-\frac{1}{\gamma}(\mathbf{x}_i^{r,1}-\mathbf{x}_i^r)-\frac{1}{\eta}(\mathbf{x}_i^{r,1}-\mathbf{x}_{0,i}^r).$$

Next we take expectation and substitute (102), (103) to obtain (112), where we substitute Lemma 10 and (103) in (a).

Taking expectation of (92), summing over $r=0$ to $r=T-1$ and average over the agents, we obtain (113) where in (a) we apply Lemma 10 and (28).

Finally, in the last equation of (113), we have defined the constant C_{10} as

$$C_{10}:=\frac{1}{2\eta}+\frac{1}{\gamma}-L-\frac{6\eta}{\gamma^2}-9Q^2L^2\eta-\frac{(1+18L\eta)LIQ}{2B}.$$

Then by taking expectation and applying Lemma 11, we obtain

$$\mathbb{E}[f(\mathbf{x}_0^T)-f(\mathbf{x}_0^0)]$$
$$\leq\ -\frac{C_{10}-\frac{(1+L\gamma)^2+L^2\gamma^2}{4BL\gamma^2}}{N}\sum_{i=1}^N\sum_{q=0}^{Q-1}\sum_{r=0}^{T-1}\mathbb{E}\left\|\mathbf{x}_i^{r,q+1}-\mathbf{x}_i^{r,q-1}\right\|^2$$
$$-\frac{1}{2\eta}\sum_{r=0}^{T-1}\mathbb{E}\left\|\mathbf{x}_0^{r+1}-\mathbf{x}_0^r\right\|^2,$$

where by the initialization that $\mathbf{x}_i^0=\mathbf{x}_0^0$ we have $f_(\mathbf{x}_0^0)=\frac{1}{N}\sum_{i=1}^N\mathcal{L}_i(\mathbf{x}_i^0,\mathbf{x}_{0,i}^0,\lambda_i^0)$.

Combine (112) and (114), we can find a positive constant C_{11} satisfying

$$C_{11}\leq\min\left\{C_{12}/C_{13},1/(4\eta)\right\}, \tag{114}$$

where we have defined

$$C_{12}\triangleq C_{10}-\frac{(1+L\gamma)^2+L^2\gamma^2}{4BL\gamma^2},$$
$$C_{13}\triangleq Q\left(2(\frac{\eta+\gamma}{\eta\gamma})^2+\frac{2I(1+18\eta^2)L^2}{B}\right)$$
$$+Q\left(\frac{3L(1+9L\eta)\eta^2}{2B\gamma^2}+18Q^2L^2\eta^2\right) \tag{115}$$

so that the following holds

$$\frac{C_{11}}{NT}\sum_{r=0}^T\sum_{i=1}^N\mathbb{E}\left\|\nabla\mathcal{L}_i(\mathbf{x}_i^r,\mathbf{x}_{0,i}^r,\lambda_i^r)\right\|^2$$
$$\leq\frac{C_{10}-\frac{(1+L\gamma)^2+L^2\gamma^2}{4BL\gamma^2}}{NT}\sum_{i=1}^N\sum_{q=0}^{Q-1}\sum_{r=0}^{T-1}\mathbb{E}\left\|\mathbf{x}_i^{r,q+1}-\mathbf{x}_i^{r,q-1}\right\|^2$$
$$+\frac{1}{2\eta T}\sum_{r=0}^{T-1}\mathbb{E}\left\|\mathbf{x}_0^{r+1}-\mathbf{x}_0^r\right\|^2$$
$$\leq\frac{1}{T}(f(\mathbf{x}_0^0)-\mathbb{E}f(\mathbf{x}_0^T))\leq\frac{1}{T}(f(\mathbf{x}_0^0)-f(\mathbf{x}^\star)). \tag{116}$$

Similar to the proof of Theorem 1, we can bound $\|\nabla f(\mathbf{x}_0^r)\|^2$ by $\frac{1}{N}\sum_{i=1}^N\|\nabla\mathcal{L}_i(\mathbf{x}_i^r,\mathbf{x}_0^r,\lambda_i^r)\|^2$, therefore Theorem 2 is proved.

Note that during the proof we need the following constants C_9, C_{10}, C_{11} in (117) to be positive By selecting $\gamma>\frac{5}{B\sqrt{L}}\eta$, and $0<\eta<\frac{1}{3(Q+\sqrt{QI/B})L}$, this is guaranteed.

$$\left\|\nabla\mathcal{L}_i(\mathbf{x}_i^r,\mathbf{x}_{0,i}^r,\lambda_i^r)\right\|^2 = \left\|\nabla_{\mathbf{x}_i}\mathcal{L}_i(\mathbf{x}_i^r,\mathbf{x}_{0,i}^r,\lambda_i^r)\right\|^2 + \left\|\nabla_{\lambda_i}\mathcal{L}_i(\mathbf{x}_i^r,\mathbf{x}_{0,i}^r,\lambda_i^r)\right\|^2 = \left\|\nabla f_i(\mathbf{x}_i^r)+\lambda_i^r+\frac{1}{\eta}(\mathbf{x}_i^r-\mathbf{x}_{0,i}^r)\right\|^2 + \left\|\mathbf{x}_i^r-\mathbf{x}_{0,i}^r\right\|^2$$
$$\overset{(a)}{=} \left\|\nabla f_i(\mathbf{x}_i^r) - g_i^{r,0} - \frac{\eta+\gamma}{\eta\gamma}(\mathbf{x}_i^{r,1}-\mathbf{x}_i^r)\right\|^2 + \left\|\mathbf{x}_i^r-\mathbf{x}_{0,i}^r+\mathbf{x}_{0,i}^{r-1}-\mathbf{x}_{0,i}^{r-1}\right\|^2$$
$$\leq \left\|\nabla f_i(\mathbf{x}_i^r) - g_i^{r,0} - \frac{\eta+\gamma}{\eta\gamma}(\mathbf{x}_i^{r,1}-\mathbf{x}_i^r)\right\|^2 + 2\left\|\mathbf{x}_i^r-\mathbf{x}_{0,i}^{r-1}\right\|^2 + 2\left\|\mathbf{x}_{0,i}^r-\mathbf{x}_{0,i}^{r-1}\right\|^2$$
$$\leq 2\left\|\nabla f_i(\mathbf{x}_i^r)-g_i^{r,0}\right\|^2 + 2(\frac{\eta+\gamma}{\eta\gamma})^2\left\|\mathbf{x}_i^{r,1}-\mathbf{x}_i^r\right\|^2 + 2\eta^2\left\|\lambda_i^r-\lambda_i^{r-1}\right\|^2 + 2\left\|\mathbf{x}_{0,i}^r-\mathbf{x}_{0,i}^{r-1}\right\|^2. \tag{111}$$

$$\mathbb{E}\left\|\nabla\mathcal{L}_i(\mathbf{x}_i^r,\mathbf{x}_{0,i}^r,\lambda_i^r)\right\|^2 \leq 2\,\mathbb{E}\left\|\nabla f_i(\mathbf{x}_i^r)-g_i^{r,0}\right\|^2 + 2(\frac{\eta+\gamma}{\eta\gamma})^2\,\mathbb{E}\left\|\mathbf{x}_i^{r,1}-\mathbf{x}_i^r\right\|^2 + 2\,\mathbb{E}\left\|\mathbf{x}_{0,i}^r-\mathbf{x}_{0,i}^{r-1}\right\|^2$$
$$+\frac{6\eta^2}{\gamma^2}(\gamma^2\,\mathbb{E}\left\|g_i^{r-1,Q-1}-g_i^{r-2,Q-1}\right\|^2 + \mathbb{E}\left\|\mathbf{x}_i^{r-1,Q}-\mathbf{x}_i^{r-1,Q-1}\right\|^2 + E\left\|\mathbf{x}_i^{r-2,Q}-\mathbf{x}_i^{r-2,Q-1}\right\|^2)$$
$$\overset{(a)}{\leq}\frac{2L^2}{B}\sum_{\tau=\{r_0,1\}}^{\{r,0\}}\mathbb{E}\left\|\mathbf{x}_i^\tau-\mathbf{x}_i^{\tau-1}\right\|^2 + 2(\frac{\eta+\gamma}{\eta\gamma})^2\,\mathbb{E}\left\|\mathbf{x}_i^{r,1}-\mathbf{x}_i^r\right\|^2 + 2\,\mathbb{E}\left\|\mathbf{x}_{0,i}^r-\mathbf{x}_{0,i}^{r-1}\right\|^2$$
$$+\frac{6\eta^2}{\gamma^2}(\mathbb{E}\left\|\mathbf{x}_i^{r-1,Q}-\mathbf{x}_i^{r-1,Q-1}\right\|^2 + \mathbb{E}\left\|\mathbf{x}_i^{r-2,Q}-\mathbf{x}_i^{r-2,Q-1}\right\|^2)$$
$$+18\eta^2\left(\mathbb{E}\left\|g_i^{r-1,Q-1}-\nabla f_i(\mathbf{x}_i^{r-1,Q-1})\right\|^2 + \mathbb{E}\left\|g_i^{r-2,Q-1}-\nabla f_i(\mathbf{x}_i^{r-2,Q-1})\right\|^2\right)$$
$$+18\eta^2Q^2L^2\left(\sum_{q=1}^{Q-1}\mathbb{E}\left\|\mathbf{x}_i^{r-1,q}-\mathbf{x}_i^{r-1,q-1}\right\|^2 + \mathbb{E}\left\|\mathbf{x}_i^{r-2,Q}-\mathbf{x}_i^{r-2,Q-1}\right\|^2\right), \tag{112}$$

$$\frac{1}{N}\sum_{i=1}^N\mathbb{E}[\mathcal{L}_i(\mathbf{x}_i^T,\mathbf{x}_{0,i}^T,\lambda_i^T)-\mathcal{L}_i(\mathbf{x}_i^0,\mathbf{x}_{0,i}^0,\lambda_i^0)]$$
$$\leq -\frac{1}{2\eta}\sum_{r=0}^{T-1}\mathbb{E}\left\|\mathbf{x}_0^{r+1}-\mathbf{x}_0^r\right\|^2 - (\frac{1}{2\eta}+\frac{1}{\gamma}-L-\frac{6\eta}{\gamma^2}-9Q^2L^2\eta)\frac{1}{N}\sum_{i=1}^N\sum_{q=0}^{Q-1}\sum_{r=0}^{T-1}\mathbb{E}\left\|\mathbf{x}_i^{r,q+1}-\mathbf{x}_i^{r,q-1}\right\|^2$$
$$+(\frac{1}{2L}+18\eta)\frac{1}{N}\sum_{i=1}^N\sum_{r=0}^{T-1}\sum_{q=0}^{Q-1}\mathbb{E}\left\|\nabla f_i(\mathbf{x}_i^{r,q})-g_i^{r,q}\right\|^2 + \sum_{r=0}^{T-1}\frac{1}{N}\,\mathbb{E}\left\langle\sum_{i=1}^N(\lambda_i^{r+1}+\frac{1}{\eta}(\mathbf{x}_i^{r+1}-\mathbf{x}_{0,i}^{r+1})),\mathbf{x}_{0,i}^{r+1}-\mathbf{x}_{0,i}^r\right\rangle$$
$$\overset{(a)}{\leq} -(\frac{1}{2\eta}+\frac{1}{\gamma}-L-\frac{6\eta}{\gamma^2}-9Q^2L^2\eta)\frac{1}{N}\sum_{i=1}^N\sum_{q=0}^{Q-1}\sum_{r=0}^{T-1}\mathbb{E}\left\|\mathbf{x}_i^{r,q+1}-\mathbf{x}_i^{r,q-1}\right\|^2 - \frac{1}{2\eta}\sum_{r=0}^{T-1}\mathbb{E}\left\|\mathbf{x}_0^{r+1}-\mathbf{x}_0^r\right\|^2$$
$$+\frac{(1+18L\eta)LIQ}{2B}\frac{1}{N}\sum_{i=1}^N\sum_{r=0}^{T-1}\sum_{q=0}^{Q-1}\mathbb{E}\left\|\mathbf{x}_i^{r,q+1}-\mathbf{x}_i^{r,q-1}\right\|^2$$
$$=-\frac{C_{10}}{N}\sum_{i=1}^N\sum_{q=0}^{Q-1}\sum_{r=0}^{T-1}\mathbb{E}\left\|\mathbf{x}_i^{r,q+1}-\mathbf{x}_i^{r,q-1}\right\|^2 - \frac{1}{2\eta}\sum_{r=0}^{T-1}\mathbb{E}\left\|\mathbf{x}_0^{r+1}-\mathbf{x}_0^r\right\|^2, \tag{113}$$

$$C_9 = 4L^2/C_{11}, \quad C_{10} = \frac{1}{2\eta}+\frac{1}{\gamma}-L-\frac{6\eta}{\gamma^2}-9Q^2L^2\eta-\frac{(1+18L\eta)LIQ}{2B}, \tag{117}$$
$$C_{11} \leq \min\left\{\frac{\left(C_{10}-\frac{(1+L\gamma)^2+L^2\gamma^2}{4BL\gamma^2}\right)}{Q\left(2(\frac{\eta+\gamma}{\eta\gamma})^2+\frac{2I(1+18\eta^2)L^2}{B}+\frac{3L(1+9L\eta)\eta^2}{2B\gamma^2}+18Q^2L^2\eta^2\right)},\frac{1}{4\eta}\right\}$$