I. Appendix

A. Notations

As defined in Eq.(28), the stationary gap $\mathcal{G}^t$ is denoted by

$$\mathcal{G}^t = \begin{bmatrix} \{P(\mathbf{x}_i, \bar{\nabla}_{\mathbf{x}} L'_p(\{\mathbf{x}_i^t\}, \{\mathbf{y}_i^t\}, \{\lambda_{i,l}^t\}, \{\boldsymbol{\theta}_{i,j}^t\}), \eta_{i,x}^t)\} \\ \{\bar{\nabla}_{\mathbf{y}} L_p(\{\mathbf{x}_i^t\}, \{\mathbf{y}_i^t\}, \{\lambda_{i,l}^t\}, \{\boldsymbol{\theta}_{i,j}^t\})\} \\ \{\nabla_{\lambda_{i,l}} L_p(\{\mathbf{x}_i^t\}, \{\mathbf{y}_i^t\}, \{\lambda_{i,l}^t\}, \{\boldsymbol{\theta}_{i,j}^t\})\} \\ \{\nabla_{\boldsymbol{\theta}_{i,j}} L_p(\{\mathbf{x}_i^t\}, \{\mathbf{y}_i^t\}, \{\lambda_{i,l}^t\}, \{\boldsymbol{\theta}_{i,j}^t\})\} \end{bmatrix}. \tag{S1}$$

We further define that

$$\begin{aligned} &(\mathcal{G}^t)_{\mathbf{x}_i} = P(\mathbf{x}_i, \bar{\nabla}_{\mathbf{x}} L'_p(\{\mathbf{x}_i^t\}, \{\mathbf{y}_i^t\}, \{\lambda_{i,l}^t\}, \{\boldsymbol{\theta}_{i,j}^t\}), \eta_{i,x}^t), \\ &(\mathcal{G}^t)_{\mathbf{y}} = \bar{\nabla}_{\mathbf{y}_i} L_p(\{\mathbf{x}_i^t\}, \{\mathbf{y}_i^t\}, \{\lambda_{i,l}^t\}, \{\boldsymbol{\theta}_{i,j}^t\}), \\ &(\mathcal{G}^t)_{\lambda_{i,l}} = \nabla_{\lambda_{i,l}} L_p(\{\mathbf{x}_i^t\}, \{\mathbf{y}_i^t\}, \{\lambda_{i,l}^t\}, \{\boldsymbol{\theta}_{i,j}^t\}), \\ &(\mathcal{G}^t)_{\boldsymbol{\theta}_{i,j}} = \nabla_{\boldsymbol{\theta}_{i,j}} L_p(\{\mathbf{x}_i^t\}, \{\mathbf{y}_i^t\}, \{\lambda_{i,l}^t\}, \{\boldsymbol{\theta}_{i,j}^t\}). \end{aligned} \tag{S2}$$

Similarly, we define $\tilde{\mathcal{G}}^t$ as

$$\tilde{\mathcal{G}}^t = \begin{bmatrix} \{P(\mathbf{x}_i, \bar{\nabla}_{\mathbf{x}} \tilde{L}_p(\{\mathbf{x}_i^t\}, \{\mathbf{y}_i^t\}, \{\lambda_{i,l}^t\}, \{\boldsymbol{\theta}_{i,j}^t\}), \eta_{i,x}^t)\} \\ \{\bar{\nabla}_{\mathbf{y}} \tilde{L}_p(\{\mathbf{x}_i^t\}, \{\mathbf{y}_i^t\}, \{\lambda_{i,l}^t\}, \{\boldsymbol{\theta}_{i,j}^t\})\} \\ \{\nabla_{\lambda_{i,l}} \tilde{L}_p(\{\mathbf{x}_i^t\}, \{\mathbf{y}_i^t\}, \{\lambda_{i,l}^t\}, \{\boldsymbol{\theta}_{i,j}^t\})\} \\ \{\nabla_{\boldsymbol{\theta}_{i,j}} \tilde{L}_p(\{\mathbf{x}_i^t\}, \{\mathbf{y}_i^t\}, \{\lambda_{i,l}^t\}, \{\boldsymbol{\theta}_{i,j}^t\})\} \end{bmatrix}, \tag{S3}$$

and define

$$\begin{aligned} &(\tilde{\mathcal{G}}^t)_{\mathbf{x}_i} = P(\mathbf{x}_i, \bar{\nabla}_{\mathbf{x}} \tilde{L}_p(\{\mathbf{x}_i^t\}, \{\mathbf{y}_i^t\}, \{\lambda_{i,l}^t\}, \{\boldsymbol{\theta}_{i,j}^t\}), \eta_{i,x}^t), \\ &(\tilde{\mathcal{G}}^t)_{\mathbf{y}_i} = \bar{\nabla}_{\mathbf{y}} \tilde{L}_p(\{\mathbf{x}_i^t\}, \{\mathbf{y}_i^t\}, \{\lambda_{i,l}^t\}, \{\boldsymbol{\theta}_{i,j}^t\}), \\ &(\tilde{\mathcal{G}}^t)_{\lambda_{i,l}} = \nabla_{\lambda_{i,l}} \tilde{L}_p(\{\mathbf{x}_i^t\}, \{\mathbf{y}_i^t\}, \{\lambda_{i,l}^t\}, \{\boldsymbol{\theta}_{i,j}^t\}), \\ &(\tilde{\mathcal{G}}^t)_{\boldsymbol{\theta}_{i,j}} = \nabla_{\boldsymbol{\theta}_{i,j}} \tilde{L}_p(\{\mathbf{x}_i^t\}, \{\mathbf{y}_i^t\}, \{\lambda_{i,l}^t\}, \{\boldsymbol{\theta}_{i,j}^t\}). \end{aligned} \tag{S4}$$

As introduced in Definition 3, the convergence metric Ψ^t can be denoted by

$$\begin{aligned} \Psi^t =& \sum_{i=1}^{N} ||(\mathcal{G}^t)_{\mathbf{x}_i}||^2 + L^2 \sum_{i=1}^{N} ||\mathbf{x}_i - \bar{\mathbf{x}}_i||^2 + \sum_{i=1}^{N} ||(\mathcal{G}^t)_{\mathbf{y}_i}||^2 + L^2 ||\mathbf{y}_i - \bar{\mathbf{y}}_i||^2 \\ &+ \sum_{i=1}^{N} \sum_{l=1}^{|\mathcal{P}_i|} ||(\mathcal{G}^t)_{\lambda_{i,j}}||^2 + \sum_{i=1}^{N} \sum_{j=1}^{N} ||(\mathcal{G}^t)_{\boldsymbol{\theta}_{i,j}}||^2, \end{aligned} \tag{S5}$$

where N is the number of agents, M is the maximum number of cutting planes.

Similarly, we define that

$$\begin{aligned} \tilde{\Psi}^t =& \sum_{i=1}^{N} ||(\tilde{\mathcal{G}}^t)_{\mathbf{x}_i}||^2 + L^2 \sum_{i=1}^{N} ||\mathbf{x}_i - \bar{\mathbf{x}}_i||^2 + \sum_{i=1}^{N} ||(\tilde{\mathcal{G}}^t)_{\mathbf{y}_i}||^2 + L^2 ||\mathbf{y}_i - \bar{\mathbf{y}}_i||^2 \\ &+ \sum_{i=1}^{N} \sum_{l=1}^{|\mathcal{P}_i|} ||(\tilde{\mathcal{G}}^t)_{\lambda_{i,j}}||^2 + \sum_{i=1}^{N} \sum_{j=1}^{N} ||(\tilde{\mathcal{G}}^t)_{\boldsymbol{\theta}_{i,j}}||^2. \end{aligned} \tag{S6}$$

B. Convergence Analysis

1) Lemma 1 (Descending Inequality of x Variables): Under Assumptions 1 and 2, the following inequality holds,

$$\begin{aligned} &\mathbb{E}[L_p(\{\bar{\mathbf{x}}^{t+1}\}, \{\mathbf{y}_i^t\}, \{\lambda_{i,l}^t\}, \{\boldsymbol{\theta}_{i,j}^t\}) - L_p(\{\bar{\mathbf{x}}^t\}, \{\mathbf{y}_i^t\}, \{\lambda_{i,l}^t\}, \{\boldsymbol{\theta}_{i,j}^t\})] \\ \le& (NL - \frac{N}{2\eta_x}) \mathbb{E}[\sum_{i=1}^{N} ||\mathbf{x}_i^{t+1} - \bar{\mathbf{x}}^t||^2] - \frac{N}{2\eta_x} \mathbb{E}[\sum_{i=1}^{N} ||\mathbf{x}_i^{t+1} - \mathbf{d}_i^t||^2] + (\frac{N}{2\eta_x} + N^2 L) \mathbb{E}[\sum_{i=1}^{N} ||\bar{\mathbf{x}}^t - \mathbf{x}_i^t||^2] \\ &+ 2\tau k_1 N^2 L (\mathbb{E}[\sum_{i=1}^{N} (||\mathbf{x}_i^{t+1} - \mathbf{x}_i^t||^2] + \mathbb{E}[\sum_{i=1}^{N} ||\mathbf{y}_i^{t+1} - \mathbf{y}_i^t||^2])), \end{aligned} \tag{S7}$$

where $\bar{\mathbf{x}}^{t+1} = \frac{1}{N} \sum_{i=1}^{N} \mathbf{x}_i^{t+1}$, $\bar{\mathbf{x}}^t = \frac{1}{N} \sum_{i=1}^{N} \mathbf{x}_i^t$.

Proof. Firstly, based on the definitions of L_p and L_{pi} functions, we have

$$\begin{aligned}
&L_p(\{\bar{\mathbf{x}}^{t+1}\},\{\mathbf{y}_i^t\},\{\lambda_{i,l}^t\},\{\boldsymbol{\theta}_{i,j}^t\})-L_p(\{\bar{\mathbf{x}}^{t}\},\{\mathbf{y}_i^t\},\{\lambda_{i,l}^t\},\{\boldsymbol{\theta}_{i,j}^t\})\\
=&\sum_{i=1}^{N}(L_{pi}(\{\bar{\mathbf{x}}^{t+1}\},\{\mathbf{y}_i^t\},\{\lambda_{i,l}^t\},\{\boldsymbol{\theta}_{i,j}^t\})-(L_{pi}(\{\bar{\mathbf{x}}^{t}\},\{\mathbf{y}_i^t\},\{\lambda_{i,l}^t\},\{\boldsymbol{\theta}_{i,j}^t\}))\\
\leq&\sum_{i=1}^{N}(L'_{pi}(\{\bar{\mathbf{x}}^{t+1}\},\{\mathbf{y}_i^t\},\{\lambda_{i,l}^t\},\{\boldsymbol{\theta}_{i,j}^t\})-L'_{pi}(\{\bar{\mathbf{x}}^{t}\},\{\mathbf{y}_i^t\},\{\lambda_{i,l}^t\},\{\boldsymbol{\theta}_{i,j}^t\})+(R(\bar{\mathbf{x}}^{t+1})-R(\bar{\mathbf{x}}^{t})))\\
\overset{(a)}{\leq}&\sum_{i=1}^{N}(L'_{pi}(\{\bar{\mathbf{x}}^{t+1}\},\{\mathbf{y}_i^t\},\{\lambda_{i,l}^t\},\{\boldsymbol{\theta}_{i,j}^t\})-L'_{pi}(\{\bar{\mathbf{x}}^{t}\},\{\mathbf{y}_i^t\},\{\lambda_{i,l}^t\},\{\boldsymbol{\theta}_{i,j}^t\}))+N(\frac{1}{N}\sum_{i=1}^{N}R(\mathbf{x}_i^{t+1})-R(\bar{\mathbf{x}}^{t}))\\
\overset{(b)}{\leq}&\frac{NL}{2}\sum_{i=1}^{N}||\bar{\mathbf{x}}^{t+1}-\bar{\mathbf{x}}^{t}||^2+N\sum_{i=1}^{N}\langle\nabla_{\mathbf{x}_i}L'_{pi}(\{\bar{\mathbf{x}}^{t}\},\{\mathbf{y}_i^t\},\{\lambda_{i,l}^t\},\{\boldsymbol{\theta}_{i,j}^t\}),\bar{\mathbf{x}}^{t+1}-\bar{\mathbf{x}}^{t}\rangle+\sum_{i=1}^{N}(R(\mathbf{x}_i^{t+1})-R(\bar{\mathbf{x}}^{t})),
\end{aligned} \tag{S8}$$

where (a) is because of the convexity of $R(\cdot)$, (b) utilizs the Lipschitz properties in Assumption 1.

According to the updating rules of $\mathbf{x}$ variables in Eq.(17), we have

$$\mathbf{0}\in\mathbb{E}[\frac{\tilde{\eta}_{i,x}^t}{N}\partial R(\mathbf{x}_i^{t+1})+\mathbf{x}_i^{t+1}-(\mathbf{d}_i^t-\tilde{\eta}_{i,x}^t\nabla_{\mathbf{x}_i}\tilde{L}_{pi}(\{\mathbf{x}_i^{\hat{t}_i}\},\{\mathbf{y}_i^{\hat{t}_i}\},\{\lambda_{i,l}^{\hat{t}_i}\},\{\boldsymbol{\theta}_{i,j}^{\hat{t}_i}\}))], \tag{S9}$$

where ∂ here denotes the subgradient. $\tilde{\eta}_{i,x}^t$ is a virtual learning rate defined as following:

$$\tilde{\eta}_{i,x}^t=\left\{\begin{array}{l}\eta_{i,x}^t, i\in Q^{t+1}\\ 0, i\notin Q^{t+1}\end{array}\right. . \tag{S10}$$

Similar to [1], let $\mathbb{E}[\tilde{\eta}_{i,x}^t]=\eta_x$, $\mathbb{E}[\tilde{\eta}_{i,y}^t]=\eta_y$, $\mathbb{E}[\tilde{\eta}_{i,\lambda}^t]=\eta_\lambda$, and $\mathbb{E}[\tilde{\eta}_{i,\theta}^t]=\eta_\theta$.

Combining Eq.(S10) and Eq.(S9), we have

$$\mathbf{0}\in\eta_x\partial R(\mathbf{x}_i^{t+1})+\mathbf{x}_i^{t+1}-(\mathbf{d}_i^t-\eta_x\nabla_{\mathbf{x}_i}\tilde{L}_{pi}(\{\mathbf{x}_i^{\hat{t}_i}\},\{\mathbf{y}_i^{\hat{t}_i}\},\{\lambda_{i,l}^{\hat{t}_i}\},\{\boldsymbol{\theta}_{i,j}^{\hat{t}_i}\})), \tag{S11}$$

According to the convexity of R, for some $\tilde{\nabla}R(\mathbf{x}_i^{t+1})\in\{\partial R(\mathbf{x}_i^{t+1})\}$ and any $\mathbf{x}_i\in\mathbb{R}^n$, we have

$$\begin{aligned}
&R(\mathbf{x}_i^{t+1})-R(\mathbf{x}_i)+\langle\mathbf{x}_i^{t+1}-\bar{\mathbf{x}}^t,N\nabla_{\mathbf{x}_i}\tilde{L}_{pi}(\{\mathbf{x}_i^{\hat{t}_i}\},\{\mathbf{y}_i^{\hat{t}_i}\},\{\lambda_{i,l}^{\hat{t}_i}\},\{\boldsymbol{\theta}_{i,j}^{\hat{t}_i}\})\rangle\\
\leq&\langle\mathbf{x}_i^{t+1}-\mathbf{x}_i,\tilde{\nabla}R(\mathbf{x}_i^{t+1})+N\nabla_{\mathbf{x}_i}\tilde{L}_{pi}(\{\mathbf{x}_i^{\hat{t}_i}\},\{\mathbf{y}_i^{\hat{t}_i}\},\{\lambda_{i,l}^{\hat{t}_i}\},\{\boldsymbol{\theta}_{i,j}^{\hat{t}_i}\})\rangle\\
=&-\frac{N}{\eta_x}\langle\mathbf{x}_i^{t+1}-\mathbf{x}_i,\mathbf{x}_i^{t+1}-\mathbf{d}_i^t\rangle\\
\overset{(a)}{=}&-\frac{N}{2\eta_x}(||\mathbf{x}_i^{t+1}-\mathbf{x}_i||^2+||\mathbf{x}_i^{t+1}-\mathbf{d}_i^t||^2-||\mathbf{x}_i-\mathbf{d}_i^t||^2),
\end{aligned} \tag{S12}$$

where (a) is from $\langle\mathbf{a},\mathbf{b}\rangle=\frac{1}{2}(\|\mathbf{a}\|_2^2+\|\mathbf{b}\|_2^2-\|\mathbf{a}-\mathbf{b}\|_2^2)$.

Setting $\mathbf{x}_i=\bar{\mathbf{x}}^t$, we have that for all $i=1,\cdots,N$,

$$\begin{aligned}
&R(\mathbf{x}_i^{t+1})-R(\bar{\mathbf{x}}^t)+\langle\mathbf{x}_i^{t+1}-\bar{\mathbf{x}}^t,N\nabla_{\mathbf{x}_i}\tilde{L}_{pi}(\{\mathbf{x}_i^{\hat{t}_i}\},\{\mathbf{y}_i^{\hat{t}_i}\},\{\lambda_{i,l}^{\hat{t}_i}\},\{\boldsymbol{\theta}_{i,j}^{\hat{t}_i}\})\rangle\\
\leq&-\frac{N}{2\eta_x}(||\mathbf{x}_i^{t+1}-\bar{\mathbf{x}}^t||^2+||\mathbf{x}_i^{t+1}-\mathbf{d}_i^t||^2-||\bar{\mathbf{x}}^t-\mathbf{d}_i^t||^2).
\end{aligned} \tag{S13}$$

Plugging Eq.(S13) into Eq.(S8) yields

$$\begin{aligned}
&L_p(\{\bar{\mathbf{x}}^{t+1}\},\{\mathbf{y}_i^t\},\{\lambda_{i,l}^t\},\{\boldsymbol{\theta}_{i,j}^t\})-L_p(\{\bar{\mathbf{x}}^{t}\},\{\mathbf{y}_i^t\},\{\lambda_{i,l}^t\},\{\boldsymbol{\theta}_{i,j}^t\})\\
\leq&\frac{N^2L}{2}||\bar{\mathbf{x}}^{t+1}-\bar{\mathbf{x}}^{t}||^2+N\sum_{i=1}^{N}\langle\nabla_{\mathbf{x}_i}L'_{pi}(\{\bar{\mathbf{x}}^{t}\},\{\mathbf{y}_i^t\},\{\lambda_{i,l}^t\},\{\boldsymbol{\theta}_{i,j}^t\}),\bar{\mathbf{x}}^{t+1}-\bar{\mathbf{x}}^{t}\rangle\\
&-N\sum_{i=1}^{N}\langle\mathbf{x}_i^{t+1}-\bar{\mathbf{x}}^t,\nabla_{\mathbf{x}_i}\tilde{L}_{pi}(\{\mathbf{x}_i^{\hat{t}_i}\},\{\mathbf{y}_i^{\hat{t}_i}\},\{\lambda_{i,l}^{\hat{t}_i}\},\{\boldsymbol{\theta}_{i,j}^{\hat{t}_i}\})\rangle\\
&-\frac{N}{2\eta_x}\sum_{i=1}^{N}(||\mathbf{x}_i^{t+1}-\bar{\mathbf{x}}^t||^2+||\mathbf{x}_i^{t+1}-\mathbf{d}_i^t||^2-||\bar{\mathbf{x}}^t-\mathbf{d}_i^t||^2).
\end{aligned} \tag{S14}$$

In Eq.(S14), according to the linearity of the inner product, we have

$$\begin{aligned}
&N\sum_{i=1}^{N}\langle\nabla_{\mathbf{x}_i}L'_{pi}(\{\bar{\mathbf{x}}^t\},\{\mathbf{y}_i^t\},\{\lambda_{i,l}^t\},\{\boldsymbol{\theta}_{i,j}^t\}),\bar{\mathbf{x}}^{t+1}-\bar{\mathbf{x}}^t\rangle-N\sum_{i=1}^{N}\langle\mathbf{x}_i^{t+1}-\bar{\mathbf{x}}^t,\nabla_{\mathbf{x}_i}\tilde{L}_{pi}(\{\mathbf{x}_i^{\hat{t}_i}\},\{\mathbf{y}_i^{\hat{t}_i}\},\{\lambda_{i,l}^{\hat{t}_i}\},\{\boldsymbol{\theta}_{i,j}^{\hat{t}_i}\})\rangle\\
=&N\sum_{i=1}^{N}\langle\nabla_{\mathbf{x}_i}L'_{pi}(\{\bar{\mathbf{x}}^t\},\{\mathbf{y}_i^t\},\{\lambda_{i,l}^t\},\{\boldsymbol{\theta}_{i,j}^t\}),\mathbf{x}_i^{t+1}-\bar{\mathbf{x}}^t\rangle-N\sum_{i=1}^{N}\langle\mathbf{x}_i^{t+1}-\bar{\mathbf{x}}^t,\nabla_{\mathbf{x}_i}\tilde{L}_{pi}(\{\mathbf{x}_i^{\hat{t}_i}\},\{\mathbf{y}_i^{\hat{t}_i}\},\{\lambda_{i,l}^{\hat{t}_i}\},\{\boldsymbol{\theta}_{i,j}^{\hat{t}_i}\})\rangle\\
=&N\sum_{i=1}^{N}\langle\nabla_{\mathbf{x}_i}L'_{pi}(\{\bar{\mathbf{x}}^t\},\{\mathbf{y}_i^t\},\{\lambda_{i,l}^t\},\{\boldsymbol{\theta}_{i,j}^t\})-\nabla_{\mathbf{x}_i}\tilde{L}_{pi}(\{\mathbf{x}_i^{\hat{t}_i}\},\{\mathbf{y}_i^{\hat{t}_i}\},\{\lambda_{i,l}^{\hat{t}_i}\},\{\boldsymbol{\theta}_{i,j}^{\hat{t}_i}\}),\mathbf{x}_i^{t+1}-\bar{\mathbf{x}}^t\rangle\\
\le&\frac{NL}{2}\sum_{i=1}^{N}||\mathbf{x}_i^{t+1}-\bar{\mathbf{x}}^t||^2+\frac{N}{2L}\sum_{i=1}^{N}||\nabla_{\mathbf{x}_i}L'_{pi}(\{\bar{\mathbf{x}}^t\},\{\mathbf{y}_i^t\},\{\lambda_{i,l}^t\},\{\boldsymbol{\theta}_{i,j}^t\})-\nabla_{\mathbf{x}_i}L'_{pi}(\{\mathbf{x}_i^{\hat{t}_i}\},\{\mathbf{y}_i^{\hat{t}_i}\},\{\lambda_{i,l}^{\hat{t}_i}\},\{\boldsymbol{\theta}_{i,j}^{\hat{t}_i}\})||^2.
\end{aligned}\tag{S15}$$

Looking at the last term in Eq.(S15), combining the Cauchy-Schwarz inequality with Assumption 1, we have

$$\begin{aligned}
&||\nabla_{\mathbf{x}_i}L'_{pi}(\{\bar{\mathbf{x}}^t\},\{\mathbf{y}_i^t\},\{\lambda_{i,l}^t\},\{\boldsymbol{\theta}_{i,j}^t\})-\nabla_{\mathbf{x}_i}L'_{pi}(\{\mathbf{x}_i^{\hat{t}_i}\},\{\mathbf{y}_i^{\hat{t}_i}\},\{\lambda_{i,l}^{\hat{t}_i}\},\{\boldsymbol{\theta}_{i,j}^{\hat{t}_i}\})||^2\\
=&||\nabla_{\mathbf{x}_i}L'_{pi}(\{\bar{\mathbf{x}}^t\},\{\mathbf{y}_i^t\},\{\lambda_{i,l}^t\},\{\boldsymbol{\theta}_{i,j}^t\})-\nabla_{\mathbf{x}_i}L'_{pi}(\{\mathbf{x}_i^t\},\{\mathbf{y}_i^t\},\{\lambda_{i,l}^t\},\{\boldsymbol{\theta}_{i,j}^t\})\\
+&\nabla_{\mathbf{x}_i}L'_{pi}(\{\mathbf{x}_i^t\},\{\mathbf{y}_i^t\},\{\lambda_{i,l}^t\},\{\boldsymbol{\theta}_{i,j}^t\})-\nabla_{\mathbf{x}_i}L'_{pi}(\{\mathbf{x}_i^{\hat{t}_i}\},\{\mathbf{y}_i^{\hat{t}_i}\},\{\lambda_{i,l}^{\hat{t}_i}\},\{\boldsymbol{\theta}_{i,j}^{\hat{t}_i}\})||^2\\
\le&2L^2\sum_{i=1}^{N}||\bar{\mathbf{x}}^t-\mathbf{x}_i^t||^2+2L^2(\sum_{i=1}^{N}||\mathbf{x}_i^t-\mathbf{x}_i^{\hat{t}_i}||^2+\sum_{i=1}^{N}||\mathbf{y}_i^t-\mathbf{y}_i^{\hat{t}_i}||^2)\\
\le&2L^2\sum_{i=1}^{N}||\bar{\mathbf{x}}^t-\mathbf{x}_i^t||^2+4L^2\tau k_1(\sum_{i=1}^{N}||\mathbf{x}_i^{t+1}-\mathbf{x}_i^t||^2+\sum_{i=1}^{N}||\mathbf{y}_i^{t+1}-\mathbf{y}_i^t||^2).
\end{aligned}\tag{S16}$$

It follows from Jensens inequality that

$$||\bar{\mathbf{x}}^{t+1}-\bar{\mathbf{x}}^t||^2\le\frac{1}{N}\sum_{i=1}^{N}||\mathbf{x}_i^{t+1}-\bar{\mathbf{x}}^t||^2.\tag{S17}$$

According to Assumption 2 and the updating rules of $\mathbf{x}$ variables, we have:

$$\begin{aligned}
||\frac{1}{N}\mathbf{e}\mathbf{e}^\top\mathbf{x}^t-\mathbf{d}^t||_F^2&=||(\mathbf{W}^t-\frac{1}{N}\mathbf{e}\mathbf{e}^\top)(\mathrm{I}-\frac{1}{N}\mathbf{e}\mathbf{e}^\top)\mathbf{x}^t||_F^2\\
&\le\rho^2||\frac{1}{N}\mathbf{e}\mathbf{e}^\top\mathbf{x}^t-\mathbf{x}^t||_F^2<||\frac{1}{N}\mathbf{e}\mathbf{e}^\top\mathbf{x}^t-\mathbf{x}^t||_F^2,
\end{aligned}\tag{S18}$$

where $||\cdot||_F$ means the Frobenius norm.

Similarly, it follows that

$$\begin{aligned}
||\frac{1}{N}\mathbf{e}\mathbf{e}^\top\mathbf{y}^t-\mathbf{u}^t||_F^2&=||(\mathbf{W}^t-\frac{1}{N}\mathbf{e}\mathbf{e}^\top)(\mathrm{I}-\frac{1}{N}\mathbf{e}\mathbf{e}^\top)\mathbf{y}^t||_F^2\\
&\le\rho^2||\frac{1}{N}\mathbf{e}\mathbf{e}^\top\mathbf{y}^t-\mathbf{y}^t||_F^2<||\frac{1}{N}\mathbf{e}\mathbf{e}^\top\mathbf{y}^t-\mathbf{y}^t||_F^2.
\end{aligned}\tag{S19}$$

Lemma 1 follows via plugging Eq.(S15), Eq.(S16), Eq.(S17), and Eq.(S18) into Eq.(S14).

□

2) Lemma 2 (Descending Inequality of y Variables): Under Assumption 1 and 2, the following inequality holds,

$$\begin{aligned}
&\mathbb{E}[L_p(\{\bar{\mathbf{x}}^{t+1}\},\{\bar{\mathbf{y}}^{t+1}\},\{\lambda_{i,l}^t\},\{\boldsymbol{\theta}_{i,j}^t\})-L_p(\{\bar{\mathbf{x}}^{t+1}\},\{\bar{\mathbf{y}}^t\},\{\lambda_{i,l}^t\},\{\boldsymbol{\theta}_{i,j}^t\})]\\
\le&\frac{3\eta_yN^2L^2}{2\beta}(\mathbb{E}[\sum_{i=1}^{N}||\mathbf{x}^{t+1}-\bar{\mathbf{x}}^t||^2]+\mathbb{E}[\sum_{i=1}^{N}||\bar{\mathbf{x}}^t-\mathbf{x}_i^t||^2]+\mathbb{E}[\sum_{i=1}^{N}||\bar{\mathbf{y}}^t-\mathbf{y}_i^t||^2])\\
+&\frac{3\eta_yN^2L^2\tau k_1}{\beta}(\mathbb{E}[\sum_{i=1}^{N}||\mathbf{x}_i^{t+1}-\mathbf{x}_i^t||^2]+\mathbb{E}[\sum_{i=1}^{N}||\mathbf{y}_i^{t+1}-\mathbf{y}_i^t||^2)\\
+&(\frac{\eta_yN\beta}{2}+\frac{\eta_y^2NL}{2}-\eta_yN)\mathbb{E}[\sum_{i=1}^{N}||\nabla_{\mathbf{y}_i}\tilde{L}_{pi}(\{\mathbf{x}_i^{\hat{t}_i}\},\{\mathbf{y}_i^{\hat{t}_i}\},\{\lambda_{i,l}^{\hat{t}_i}\},\{\boldsymbol{\theta}_{i,j}^{\hat{t}_i}\})||^2],
\end{aligned}\tag{S20}$$

where $\bar{\mathbf{y}}^{t+1}=\frac{1}{N}\sum_{i=1}^{N}\mathbf{y}_i^{t+1}$, $\bar{\mathbf{y}}^t=\frac{1}{N}\sum_{i=1}^{N}\mathbf{y}_i^t$, β is a parameter that can be tuned later.

Proof. From the definitions, we have

$$
\begin{aligned}
&\mathbb{E}[L_p(\{\bar{\mathbf{x}}^{t+1}\},\{\bar{\mathbf{y}}^{t+1}\},\{\lambda_{i,l}^t\},\{\boldsymbol{\theta}_{i,j}^t\})-L_p(\{\bar{\mathbf{x}}^{t+1}\},\{\bar{\mathbf{y}}^{t}\},\{\lambda_{i,l}^t\},\{\boldsymbol{\theta}_{i,j}^t\})]\\
=&\mathbb{E}[\sum_{i=1}^{N}(L_{pi}(\{\bar{\mathbf{x}}^{t+1}\},\{\bar{\mathbf{y}}^{t+1}\},\{\lambda_{i,l}^t\},\{\boldsymbol{\theta}_{i,j}^t\})-L_{pi}(\{\bar{\mathbf{x}}^{t+1}\},\{\bar{\mathbf{y}}^{t}\},\{\lambda_{i,l}^t\},\{\boldsymbol{\theta}_{i,j}^t\}))]\\
\overset{(a)}{\le}&\mathbb{E}[\sum_{i=1}^{N}(\sum_{i=1}^{N}\langle\nabla_{\mathbf{y}_i}L_{pi}(\{\bar{\mathbf{x}}^{t+1}\},\{\bar{\mathbf{y}}^{t}\},\{\lambda_{i,l}^t\},\{\boldsymbol{\theta}_{i,j}^t\}),\bar{\mathbf{y}}^{t+1}-\bar{\mathbf{y}}^{t}\rangle+\frac{NL}{2}||\bar{\mathbf{y}}^{t+1}-\bar{\mathbf{y}}^{t}||^2)]\\
\overset{(b)}{=}&-\eta_yN\mathbb{E}[\sum_{i=1}^{N}\langle\nabla_{\mathbf{y}_i}L_{pi}(\{\bar{\mathbf{x}}^{t+1}\},\{\bar{\mathbf{y}}^{t}\},\{\lambda_{i,l}^t\},\{\boldsymbol{\theta}_{i,j}^t\}),\bar{\nabla}_{\mathbf{y}_i}\tilde{L}_{pi}(\{\mathbf{x}_i^{\hat{t}_i}\},\{\mathbf{y}_i^{\hat{t}_i}\},\{\lambda_{i,l}^{\hat{t}_i}\},\{\boldsymbol{\theta}_{i,j}^{\hat{t}_i}\})\rangle]\\
&+\frac{\eta_y^2NL}{2}\mathbb{E}[\sum_{i=1}^{N}||\bar{\nabla}_{\mathbf{y}_i}\tilde{L}_{pi}(\{\mathbf{x}_i^{\hat{t}_i}\},\{\mathbf{y}_i^{\hat{t}_i}\},\{\lambda_{i,l}^{\hat{t}_i}\},\{\boldsymbol{\theta}_{i,j}^{\hat{t}_i}\})||^2]\\
\overset{(c)}{\le}&\frac{\eta_yN}{2\beta}\sum_{i=1}^{N}\mathbb{E}[||\nabla_{\mathbf{y}_i}L_{pi}(\{\bar{\mathbf{x}}^{t+1}\},\{\bar{\mathbf{y}}^{t}\},\{\lambda_{i,l}^t\},\{\boldsymbol{\theta}_{i,j}^t\})-\bar{\nabla}_{\mathbf{y}_i}L_{pi}(\{\mathbf{x}_i^{\hat{t}_i}\},\{\mathbf{y}_i^{\hat{t}_i}\},\{\lambda_{i,l}^{\hat{t}_i}\},\{\boldsymbol{\theta}_{i,j}^{\hat{t}_i}\})||^2]\\
&+(\frac{\eta_yN\beta}{2}+\frac{\eta_y^2NL}{2}-\eta_yN)\mathbb{E}[\sum_{i=1}^{N}||\nabla_{\mathbf{y}_i}\tilde{L}_{pi}(\{\mathbf{x}_i^{\hat{t}_i}\},\{\mathbf{y}_i^{\hat{t}_i}\},\{\lambda_{i,l}^{\hat{t}_i}\},\{\boldsymbol{\theta}_{i,j}^{\hat{t}_i}\})||^2],
\end{aligned}
\tag{S21}
$$

where (a) utilizs the Lipschitz properties in Assumption 1. (b) is because of the updating rules of $\mathbf{y}$ variables. (c) uses the variants of the Cauchy-Schwarz inequality $\langle a,b\rangle\le\frac{1}{2\beta}||a||^2+\frac{\beta}{2}||b||^2$.

According to Assumption 1 and the triangle inequality, we have

$$
\begin{aligned}
&\mathbb{E}[||\nabla_{\mathbf{y}_i}L_{pi}(\{\bar{\mathbf{x}}^{t+1}\},\{\bar{\mathbf{y}}^{t}\},\{\lambda_{i,l}^t\},\{\boldsymbol{\theta}_{i,j}^t\})-\bar{\nabla}_{\mathbf{y}_i}L_{pi}(\{\mathbf{x}_i^{\hat{t}_i}\},\{\mathbf{y}_i^{\hat{t}_i}\},\{\lambda_{i,l}^{\hat{t}_i}\},\{\boldsymbol{\theta}_{i,j}^{\hat{t}_i}\})||^2]\\
=&\mathbb{E}[||\bar{\nabla}_{\mathbf{y}_i}L_{pi}(\{\bar{\mathbf{x}}^{t+1}\},\{\bar{\mathbf{y}}^{t}\},\{\lambda_{i,l}^t\},\{\boldsymbol{\theta}_{i,j}^t\})-\bar{\nabla}_{\mathbf{y}_i}L_{pi}(\{\mathbf{x}_i^{\hat{t}_i}\},\{\mathbf{y}_i^{\hat{t}_i}\},\{\lambda_{i,l}^{\hat{t}_i}\},\{\boldsymbol{\theta}_{i,j}^{\hat{t}_i}\})||^2]\\
\le&\mathbb{E}[\frac{1}{N}\sum_{i=1}^{N}||\nabla_{\mathbf{y}_i}L_{pi}(\{\bar{\mathbf{x}}^{t+1}\},\{\bar{\mathbf{y}}^{t}\},\{\lambda_{i,l}^t\},\{\boldsymbol{\theta}_{i,j}^t\})-\nabla_{\mathbf{y}_i}L_{pi}(\{\mathbf{x}_i^{\hat{t}_i}\},\{\mathbf{y}_i^{\hat{t}_i}\},\{\lambda_{i,l}^{\hat{t}_i}\},\{\boldsymbol{\theta}_{i,j}^{\hat{t}_i}\})||^2]\\
=&\frac{1}{N}\mathbb{E}[\sum_{i=1}^{N}||\nabla_{\mathbf{y}_i}L_{pi}(\{\bar{\mathbf{x}}^{t+1}\},\{\bar{\mathbf{y}}^{t}\},\{\lambda_{i,l}^t\},\{\boldsymbol{\theta}_{i,j}^t\})-\nabla_{\mathbf{y}_i}L_{pi}(\{\bar{\mathbf{x}}^{t}\},\{\bar{\mathbf{y}}^{t}\},\{\lambda_{i,l}^t\},\{\boldsymbol{\theta}_{i,j}^t\})\\
&+\nabla_{\mathbf{y}_i}L_{pi}(\{\bar{\mathbf{x}}^{t}\},\{\bar{\mathbf{y}}^{t}\},\{\lambda_{i,l}^t\},\{\boldsymbol{\theta}_{i,j}^t\})-\nabla_{\mathbf{y}_i}L_{pi}(\{\mathbf{x}_i^{t}\},\{\mathbf{y}_i^{t}\},\{\lambda_{i,l}^t\},\{\boldsymbol{\theta}_{i,j}^t\})\\
&+\nabla_{\mathbf{y}_i}L_{pi}(\{\mathbf{x}_i^{t}\},\{\mathbf{y}_i^{t}\},\{\lambda_{i,l}^t\},\{\boldsymbol{\theta}_{i,j}^t\})-\nabla_{\mathbf{y}_i}L_{pi}(\{\mathbf{x}_i^{\hat{t}_i}\},\{\mathbf{y}_i^{\hat{t}_i}\},\{\lambda_{i,l}^{\hat{t}_i}\},\{\boldsymbol{\theta}_{i,j}^{\hat{t}_i}\})||^2]\\
\le&3L^2(\mathbb{E}[\sum_{i=1}^{N}||\bar{\mathbf{x}}^{t+1}-\bar{\mathbf{x}}^{t}||^2]+\mathbb{E}[\sum_{i=1}^{N}||\bar{\mathbf{x}}^{t}-\mathbf{x}_i^{t}||^2]+\mathbb{E}[\sum_{i=1}^{N}||\bar{\mathbf{y}}^{t}-\mathbf{y}_i^{t}||^2]+\mathbb{E}[\sum_{i=1}^{N}||\mathbf{x}_i^{t}-\mathbf{x}_i^{\hat{t}_i}||^2]+\mathbb{E}[\sum_{i=1}^{N}||\mathbf{y}_i^{t}-\mathbf{y}_i^{\hat{t}_i}||^2])\\
\le&3L^2(\mathbb{E}[\sum_{i=1}^{N}||\bar{\mathbf{x}}^{t}-\mathbf{x}_i^{t}||^2]+\mathbb{E}[\sum_{i=1}^{N}||\bar{\mathbf{y}}^{t}-\mathbf{y}_i^{t}||^2]+\mathbb{E}[\sum_{i=1}^{N}||\mathbf{x}^{t+1}-\bar{\mathbf{x}}^{t}||^2])+6L^2\tau k_1(\mathbb{E}[\sum_{i=1}^{N}||\mathbf{x}_i^{t+1}-\mathbf{x}_i^{t}||^2]+\mathbb{E}[\sum_{i=1}^{N}||\mathbf{y}_i^{t+1}-\mathbf{y}_i^{t}||^2).
\end{aligned}
\tag{S22}
$$

Lemma 2 naturally follows via plugging Eq.(S22) into Eq.(S21).

□

3) Lemma 3 (Descending Inequality of the L_p Function): Under Assumption 1 and 2, the following inequality holds,

$$
\begin{aligned}
&\mathbb{E}[L_p(\{\bar{\mathbf{x}}^{t+1}\},\{\bar{\mathbf{y}}^{t+1}\},\{\lambda_{i,l}^{t+1}\},\{\boldsymbol{\theta}_{i,j}^{t+1}\})-L_p(\{\bar{\mathbf{x}}^{t}\},\{\bar{\mathbf{y}}^{t}\},\{\lambda_{i,l}^{t}\},\{\boldsymbol{\theta}_{i,j}^{t}\})]\\
\leq&(NL-\frac{N}{2\eta_x}+\frac{3\eta_y N^2L^2}{2\beta}+\frac{MNL^2\eta_\lambda}{2}+\frac{N^2L^2\eta_{\boldsymbol{\theta}}}{2})\mathbb{E}[\sum_{i=1}^{N}||\mathbf{x}_i^{t+1}-\bar{\mathbf{x}}^t||^2]-\frac{N}{2\eta_x}\mathbb{E}[\sum_{i=1}^{N}||\mathbf{x}_i^{t+1}-\mathbf{d}_i^t||^2]\\
&+(\frac{N}{2\eta_x}+N^2L+\frac{3\eta_y N^2L^2}{2\beta}+\frac{MNL^2\eta_\lambda}{2}+\frac{N^2L^2\eta_{\boldsymbol{\theta}}}{2})\mathbb{E}[\sum_{i=1}^{N}||\bar{\mathbf{x}}^t-\mathbf{x}_i^t||^2]\\
&+(\frac{3\eta_y N^2L^2}{2\beta}+\frac{MNL^2\eta_\lambda}{2}+\frac{N^2L^2\eta_{\boldsymbol{\theta}}}{2})(\mathbb{E}[\sum_{i=1}^{N}||\bar{\mathbf{y}}^t-\mathbf{y}_i^t||^2])\\
&+(2\tau k_1N^2L+\frac{3\eta_y N^2L^2\tau k_1}{\beta})(\mathbb{E}[\sum_{i=1}^{N}(||\mathbf{x}_i^{t+1}-\mathbf{x}_i^t||^2]+\mathbb{E}[\sum_{i=1}^{N}||\mathbf{y}_i^{t+1}-\mathbf{y}_i^t||^2]))\\
&+(\frac{\eta_y N\beta}{2}+\frac{\eta_y^2NL}{2}-\eta_yN+\frac{MNL^2\eta_y^2\eta_\lambda}{2}+\frac{N^2L^2\eta_y^2\eta_{\boldsymbol{\theta}}}{2})\mathbb{E}[\sum_{i=1}^{N}||\nabla_{\mathbf{y}_i}\tilde{L}_{pi}(\{\mathbf{x}_i^{\hat{t}_i}\},\{\mathbf{y}_i^{\hat{t}_i}\},\{\lambda_{i,l}^{\hat{t}_i}\},\{\boldsymbol{\theta}_{i,j}^{\hat{t}_i}\})||^2]\\
&+(\frac{1}{\eta_\lambda}-\frac{c_1^{t-1}-c_1^t}{2})\sum_{i=1}^{N}\sum_{l=1}^{|\mathcal{P}_i^t|}\mathbb{E}[||\lambda_{i,l}^{t+1}-\lambda_{i,l}^t||^2]+(\frac{1}{\eta_\theta}-\frac{c_2^{t-1}-c_2^t}{2})\mathbb{E}[\sum_{i=1}^{N}\sum_{j=1}^{N}||\boldsymbol{\theta}_{i,j}^{t+1}-\boldsymbol{\theta}_{i,j}^t||^2]\\
&+\frac{c_1^{t-1}}{2}\mathbb{E}[\sum_{i=1}^{N}\sum_{l=1}^{|\mathcal{P}_i^t|}(||\lambda_{i,l}^{t+1}||^2-||\lambda_{i,l}^t||^2)]+\frac{1}{2\eta_\lambda}\mathbb{E}[\sum_{i=1}^{N}\sum_{l=1}^{|\mathcal{P}_i^t|}||\lambda_{i,l}^t-\lambda_{i,l}^{t-1}||^2]\\
&+\frac{c_2^{t-1}}{2}\mathbb{E}[\sum_{i=1}^{N}\sum_{j=1}^{N}(||\boldsymbol{\theta}_{i,j}^{t+1}||^2-||\boldsymbol{\theta}_{i,j}^t||^2])+\frac{1}{2\eta_\theta}\sum_{i=1}^{N}\mathbb{E}[\sum_{j=1}^{N}||\boldsymbol{\theta}_{i,j}^t-\boldsymbol{\theta}_{i,j}^{t-1}||^2].
\end{aligned}
\tag{S23}
$$

Proof. We first construct the descending inequalities of λ and $\boldsymbol{\theta}$ variables.
The updating rules of λ variables can be represented as

$$
\lambda_{i,l}^{t+1}=\lambda_{i,l}^t+\tilde{\eta}_{i,\lambda}^t\nabla_{\lambda_{i,l}}\tilde{L}_{pi}(\{\mathbf{x}_i^{t+1}\},\{\mathbf{y}_i^{t+1}\},\{\lambda_{i,l}^{\hat{t}_i}\},\{\boldsymbol{\theta}_{i,j}^{\hat{t}_i}\}).
\tag{S24}
$$

$\forall\lambda$, it follows in the $(t+1)^{th}$ iteration that:

$$
\langle\lambda_{i,l}^{t+1}-\lambda_{i,l}^t-\tilde{\eta}_{i,\lambda}^t\nabla_{\lambda_{i,l}}\tilde{L}_{pi}(\{\mathbf{x}_i^{t+1}\},\{\mathbf{y}_i^{t+1}\},\{\lambda_{i,l}^{\hat{t}_i}\},\{\boldsymbol{\theta}_{i,j}^{\hat{t}_i}\}),\lambda-\lambda_{i,l}^{t+1}\rangle=0.
\tag{S25}
$$

Let $\lambda=\lambda_{i,l}^t$ to have

$$
\langle\nabla_{\lambda_{i,l}}\tilde{L}_{pi}(\{\mathbf{x}_i^{t+1}\},\{\mathbf{y}_i^{t+1}\},\{\lambda_{i,l}^{\hat{t}_i}\},\{\boldsymbol{\theta}_{i,j}^{\hat{t}_i}\})-\frac{1}{\tilde{\eta}_{i,\lambda}^t}(\lambda_{i,l}^{t+1}-\lambda_{i,l}^t),\lambda_{i,l}^t-\lambda_{i,l}^{t+1}\rangle=0.
\tag{S26}
$$

Likewise, in the t^{th} iteration we have:

$$
\langle\nabla_{\lambda_{i,l}}\tilde{L}_{pi}(\{\mathbf{x}_i^t\},\{\mathbf{y}_i^t\},\{\lambda_{i,l}^{\hat{t-1}_i}\},\{\boldsymbol{\theta}_{i,j}^{\hat{t-1}_i}\})-\frac{1}{\tilde{\eta}_{i,\lambda}^{t-1}}(\lambda_{i,l}^t-\lambda_{i,l}^{t-1}),\lambda_{i,l}^{t+1}-\lambda_{i,l}^t\rangle=0.
\tag{S27}
$$

Since $\tilde{L}_{pi}(\{\mathbf{x}_i\},\{\mathbf{y}_i\},\{\lambda_{i,l}\},\{\boldsymbol{\theta}_{i,j}\})$ is concave with respect to $\lambda_{i,l}$, we have

$$
\begin{aligned}
&\widetilde{L}_{pi}(\{\bar{\mathbf{x}}^{t+1}\},\{\bar{\mathbf{y}}^{t+1}\},\{\lambda_{i,l}^{t+1}\},\{\boldsymbol{\theta}_{i,j}^t\})-\widetilde{L}_{pi}(\{\bar{\mathbf{x}}^{t+1}\},\{\bar{\mathbf{y}}^{t+1}\},\{\lambda_{i,l}^t\},\{\boldsymbol{\theta}_{i,j}^t\})\\
\leq&\sum_{l=1}^{|\mathcal{P}_i^t|}\langle\nabla_{\lambda_{i,l}}\widetilde{L}_{pi}(\{\bar{\mathbf{x}}^{t+1}\},\{\bar{\mathbf{y}}^{t+1}\},\{\lambda_{i,l}^t\},\{\boldsymbol{\theta}_{i,j}^t\}),\lambda_{i,l}^{t+1}-\lambda_{i,l}^t\rangle\\
=&\sum_{l=1}^{|\mathcal{P}_i^t|}\langle\nabla_{\lambda_{i,l}}\widetilde{L}_{pi}(\{\bar{\mathbf{x}}^{t+1}\},\{\bar{\mathbf{y}}^{t+1}\},\{\lambda_{i,l}^t\},\{\boldsymbol{\theta}_{i,j}^t\})-\nabla_{\lambda_{i,l}}\widetilde{L}_{pi}(\{\bar{\mathbf{x}}^t\},\{\bar{\mathbf{y}}^t\},\{\lambda_{i,l}^t\},\{\boldsymbol{\theta}_{i,j}^t\}),\lambda_{i,l}^{t+1}-\lambda_{i,l}^t\rangle\\
&+\sum_{l=1}^{|\mathcal{P}_i^t|}\langle\nabla_{\lambda_{i,l}}\widetilde{L}_{pi}(\{\bar{\mathbf{x}}^t\},\{\bar{\mathbf{y}}^t\},\{\lambda_{i,l}^t\},\{\boldsymbol{\theta}_{i,j}^t\}),\lambda_{i,l}^{t+1}-\lambda_{i,l}^t\rangle.
\end{aligned}
\tag{S28}
$$

For the first term in Eq.(S28), we have

$$
\begin{aligned}
&\sum_{l=1}^{|\mathcal{P}_i^t|}\langle\nabla_{\lambda_{i,l}}\widetilde{L}_{pi}(\{\bar{\mathbf{x}}^{t+1}\},\{\bar{\mathbf{y}}^{t+1}\},\{\lambda_{i,l}^t\},\{\boldsymbol{\theta}_{i,j}^t\})-\nabla_{\lambda_{i,l}}\widetilde{L}_{pi}(\{\bar{\mathbf{x}}^t\},\{\bar{\mathbf{y}}^t\},\{\lambda_{i,l}^t\},\{\boldsymbol{\theta}_{i,j}^t\}),\lambda_{i,l}^{t+1}-\lambda_{i,l}^t\rangle\\
=&\sum_{l=1}^{|\mathcal{P}_i^t|}\langle\nabla_{\lambda_{i,l}}L_{pi}(\{\bar{\mathbf{x}}^{t+1}\},\{\bar{\mathbf{y}}^{t+1}\},\{\lambda_{i,l}^t\},\{\boldsymbol{\theta}_{i,j}^t\})-\nabla_{\lambda_{i,l}}L_{pi}(\{\bar{\mathbf{x}}^t\},\{\bar{\mathbf{y}}^t\},\{\lambda_{i,l}^t\},\{\boldsymbol{\theta}_{i,j}^t\}),\lambda_{i,l}^{t+1}-\lambda_{i,l}^t\rangle\\
+&\frac{c_1^{t-1}-c_1^t}{2}\sum_{l=1}^{|\mathcal{P}_i^t|}(||\lambda_{i,l}^{t+1}||^2-||\lambda_{i,l}^t||^2)-\frac{c_1^{t-1}-c_1^t}{2}\sum_{l=1}^{|\mathcal{P}_i^t|}||\lambda_{i,l}^{t+1}-\lambda_{i,l}^t||^2\\
\le&\frac{NL^2}{2a_1}\sum_{l=1}^{|\mathcal{P}_i^t|}(||\bar{\mathbf{x}}^{t+1}-\bar{\mathbf{x}}^t||^2+||\bar{\mathbf{y}}^{t+1}-\bar{\mathbf{y}}^t||^2)+\frac{a_1}{2}\sum_{l=1}^{|\mathcal{P}_i^t|}||\lambda_{i,l}^{t+1}-\lambda_{i,l}^t||^2\\
+&\frac{c_1^{t-1}-c_1^t}{2}\sum_{l=1}^{|\mathcal{P}_i^t|}(||\lambda_{i,l}^{t+1}||^2-||\lambda_{i,l}^t||^2)-\frac{c_1^{t-1}-c_1^t}{2}\sum_{l=1}^{|\mathcal{P}_i^t|}||\lambda_{i,l}^{t+1}-\lambda_{i,l}^t||^2,
\end{aligned}
\tag{S29}
$$

where $a_1>0$ is a constant.

For the second term in Eq.(S28), according to Eq.(S27), we can obtain that

$$
\begin{aligned}
&\sum_{l=1}^{|\mathcal{P}_i^t|}\langle\nabla_{\lambda_{i,l}}\widetilde{L}_{pi}(\{\bar{\mathbf{x}}^t\},\{\bar{\mathbf{y}}^t\},\{\lambda_{i,l}^t\},\{\boldsymbol{\theta}_{i,j}^t\}),\lambda_{i,l}^{t+1}-\lambda_{i,l}^t\rangle\\
\le&\sum_{l=1}^{|\mathcal{P}_i^t|}(\langle\nabla_{\lambda_{i,l}}\widetilde{L}_{pi}(\{\bar{\mathbf{x}}^t\},\{\bar{\mathbf{y}}^t\},\{\lambda_{i,l}^t\},\{\boldsymbol{\theta}_{i,j}^t\})-\nabla_{\lambda_{i,l}}\widetilde{L}_{pi}(\{\mathbf{x}_i^t\},\{\mathbf{y}_i^t\},\{\lambda_{i,l}^{t\hat{-}1_i}\},\{\boldsymbol{\theta}_{i,j}^{t\hat{-}1_i}\}),\lambda_{i,l}^{t+1}-\lambda_{i,l}^t\rangle\\
+&\frac{1}{\tilde{\eta}_{i,\lambda}^{t-1}}\langle\lambda_{i,l}^t-\lambda_{i,l}^{t-1},\lambda_{i,l}^{t+1}-\lambda_{i,l}^t\rangle).
\end{aligned}
\tag{S30}
$$

Denoting $v_{1,i,l}^{t+1}=\lambda_{i,l}^{t+1}-\lambda_{i,l}^t-(\lambda_{i,l}^t-\lambda_{i,l}^{t-1})$ and taking the expectation, we have

$$
\frac{1}{\eta_\lambda}\mathbb{E}[\langle\lambda_{i,l}^t-\lambda_{i,l}^{t-1},\lambda_{i,l}^{t+1}-\lambda_{i,l}^t\rangle]\le\frac{1}{2\eta_\lambda}\mathbb{E}[||\lambda_{i,l}^{t+1}-\lambda_{i,l}^t||^2]-\frac{1}{2\eta_\lambda}\mathbb{E}[||v_{1,i,l}^{t+1}||^2]+\frac{1}{2\eta_\lambda}\mathbb{E}[||\lambda_{i,l}^t-\lambda_{i,l}^{t-1}||^2],
\tag{S31}
$$

and

$$
\begin{aligned}
&\sum_{l=1}^{|\mathcal{P}_i^t|}\langle\nabla_{\lambda_{i,l}}\widetilde{L}_{pi}(\{\bar{\mathbf{x}}^t\},\{\bar{\mathbf{y}}^t\},\{\lambda_{i,l}^t\},\{\boldsymbol{\theta}_{i,j}^t\})-\nabla_{\lambda_{i,l}}\widetilde{L}_{pi}(\{\mathbf{x}_i^t\},\{\mathbf{y}_i^t\},\{\lambda_{i,l}^{t-1}\},\{\boldsymbol{\theta}_{i,j}^{t-1}\}),\lambda_{i,l}^{t+1}-\lambda_{i,l}^t\rangle\\
=&\sum_{l=1}^{|\mathcal{P}_i^t|}\langle\nabla_{\lambda_{i,l}}\widetilde{L}_{pi}(\{\bar{\mathbf{x}}^t\},\{\bar{\mathbf{y}}^t\},\{\lambda_{i,l}^t\},\{\boldsymbol{\theta}_{i,j}^t\})-\nabla_{\lambda_{i,l}}\widetilde{L}_{pi}(\{\mathbf{x}_i^t\},\{\mathbf{y}_i^t\},\{\lambda_{i,l}^t\},\{\boldsymbol{\theta}_{i,j}^t\}),\lambda_{i,l}^{t+1}-\lambda_{i,l}^t\rangle(1a)\\
+&\sum_{l=1}^{|\mathcal{P}_i^t|}\langle\nabla_{\lambda_{i,l}}\widetilde{L}_{pi}(\{\mathbf{x}_i^t\},\{\mathbf{y}_i^t\},\{\lambda_{i,l}^t\},\{\boldsymbol{\theta}_{i,j}^t\})-\nabla_{\lambda_{i,l}}\widetilde{L}_{pi}(\{\mathbf{x}_i^t\},\{\mathbf{y}_i^t\},\{\lambda_{i,l}^{t-1}\},\{\boldsymbol{\theta}_{i,j}^{t-1}\}),v_{1,i,l}^{t+1}\rangle(1b)\\
+&\sum_{l=1}^{|\mathcal{P}_i^t|}\langle\nabla_{\lambda_{i,l}}\widetilde{L}_{pi}(\{\mathbf{x}_i^t\},\{\mathbf{y}_i^t\},\{\lambda_{i,l}^t\},\{\boldsymbol{\theta}_{i,j}^t\})-\nabla_{\lambda_{i,l}}\widetilde{L}_{pi}(\{\mathbf{x}_i^t\},\{\mathbf{y}_i^t\},\{\lambda_{i,l}^{t-1}\},\{\boldsymbol{\theta}_{i,j}^{t-1}\}),\lambda_{i,l}^t-\lambda_{i,l}^{t-1}\rangle(1c).
\end{aligned}
\tag{S32}
$$

It follows from the Cauchy-Schwarz inequality and Assumption 1 that:

$$
\begin{aligned}
&\sum_{l=1}^{|\mathcal{P}_i^t|}\langle\nabla_{\lambda_{i,l}}\widetilde{L}_{pi}(\{\bar{\mathbf{x}}^t\},\{\bar{\mathbf{y}}^t\},\{\lambda_{i,l}^t\},\{\boldsymbol{\theta}_{i,j}^t\})-\nabla_{\lambda_{i,l}}\widetilde{L}_{pi}(\{\mathbf{x}_i^t\},\{\mathbf{y}_i^t\},\{\lambda_{i,l}^t\},\{\boldsymbol{\theta}_{i,j}^t\}),\lambda_{i,l}^{t+1}-\lambda_{i,l}^t\rangle\\
=&\sum_{l=1}^{|\mathcal{P}_i^t|}\langle\nabla_{\lambda_{i,l}}L_{pi}(\{\bar{\mathbf{x}}^t\},\{\bar{\mathbf{y}}^t\},\{\lambda_{i,l}^t\},\{\boldsymbol{\theta}_{i,j}^t\})-\nabla_{\lambda_{i,l}}L_{pi}(\{\mathbf{x}_i^t\},\{\mathbf{y}_i^t\},\{\lambda_{i,l}^t\},\{\boldsymbol{\theta}_{i,j}^t\}),\lambda_{i,l}^{t+1}-\lambda_{i,l}^t\rangle\\
\le&\frac{|\mathcal{P}_i^t|L^2}{2a_2}\sum_{i=1}^{N}(||\bar{\mathbf{x}}^t-\mathbf{x}_i^t||^2+||\bar{\mathbf{y}}^t-\mathbf{y}_i^t||^2)+\frac{a_2}{2}\sum_{l=1}^{|\mathcal{P}_i^t|}||\lambda_{i,l}^{t+1}-\lambda_{i,l}^t||^2,
\end{aligned}
\tag{S33}
$$

where $a_2>0$ is a constant.

According to the Cauchy-Schwarz inequality, (1b) in Eq.(S33) can be expressed as follows:

$$\begin{aligned}&\sum_{l=1}^{|\mathcal{P}_i^t|}\langle\nabla_{\lambda_{i,l}}\widetilde{L}_{pi}(\{\mathbf{x}_i^t\},\{\mathbf{y}_i^t\},\{\lambda_{i,l}^t\},\{\boldsymbol{\theta}_{i,j}^t\})-\nabla_{\lambda_{i,l}}\widetilde{L}_{pi}(\{\mathbf{x}_i^t\},\{\mathbf{y}_i^t\},\{\lambda_{i,l}^{t-1}\},\{\boldsymbol{\theta}_{i,j}^{t-1}\}),v_{1,i,l}^{t+1}\rangle\\ \le&\sum_{l=1}^{|\mathcal{P}_i^t|}(\frac{a_3}{2}||\nabla_{\lambda_{i,l}}\widetilde{L}_{pi}(\{\mathbf{x}_i^t\},\{\mathbf{y}_i^t\},\{\lambda_{i,l}^t\},\{\boldsymbol{\theta}_{i,j}^t\})-\nabla_{\lambda_{i,l}}\widetilde{L}_{pi}(\{\mathbf{x}_i^t\},\{\mathbf{y}_i^t\},\{\lambda_{i,l}^{t-1}\},\{\boldsymbol{\theta}_{i,j}^{t-1}\})||^2+\frac{1}{2a_3}||v_{1,i,l}^{t+1}||^2),\end{aligned}\tag{S34}$$

where $a_3>0$ is a constant.

Defining $L_1'=L+c_1^0$, combining Assumption 1 and the triangle inequality, $\forall\lambda_{i,l}$, we have

$$\begin{aligned}&||\nabla_{\lambda_{i,l}}\widetilde{L}_{pi}(\{\mathbf{x}_i^t\},\{\mathbf{y}_i^t\},\{\lambda_{i,l}^t\},\{\boldsymbol{\theta}_{i,j}^t\})-\nabla_{\lambda_{i,l}}\widetilde{L}_{pi}(\{\mathbf{x}_i^t\},\{\mathbf{y}_i^t\},\{\lambda_{i,l}^{t-1}\},\{\boldsymbol{\theta}_{i,j}^{t-1}\})||\\ =&||\nabla_{\lambda_{i,l}}L_{pi}(\{\mathbf{x}_i^t\},\{\mathbf{y}_i^t\},\{\lambda_{i,l}^t\},\{\boldsymbol{\theta}_{i,j}^t\})-\nabla_{\lambda_{i,l}}L_{pi}(\{\mathbf{x}_i^t\},\{\mathbf{y}_i^t\},\{\lambda_{i,l}^{t-1}\},\{\boldsymbol{\theta}_{i,j}^t\})-c_1^{t-1}(\lambda_{i,l}^t-\lambda_{i,l}^{t-1})||\\ \le&(L+c_1^{t-1})||\lambda_{i,l}^t-\lambda_{i,l}^{t-1}||\\ \le&L_1'||\lambda_{i,l}^t-\lambda_{i,l}^{t-1}||.\end{aligned}\tag{S35}$$

Since $\widetilde{L}_{pi}(\{\mathbf{x}_i\},\{\mathbf{y}_i\},\{\lambda_{i,l}\},\{\boldsymbol{\theta}_{i,j}\})$ is strongly concave with respect to $\lambda_{i,l}$, we have

$$\begin{aligned}&\sum_{l=1}^{|\mathcal{P}_i^t|}\langle\nabla_{\lambda_{i,l}}\widetilde{L}_{pi}(\{\mathbf{x}_i^t\},\{\mathbf{y}_i^t\},\{\lambda_{i,l}^t\},\{\boldsymbol{\theta}_{i,j}^t\})-\nabla_{\lambda_{i,l}}\widetilde{L}_{pi}(\{\mathbf{x}_i^t\},\{\mathbf{y}_i^t\},\{\lambda_{i,l}^{t-1}\},\{\boldsymbol{\theta}_{i,j}^{t-1}\}),\lambda_{i,l}^t-\lambda_{i,l}^{t-1}\rangle\\ \le&\sum_{l=1}^{|\mathcal{P}_i^t|}(-\frac{1}{L_1'+c_1^{t-1}}||\nabla_{\lambda_{i,l}}\widetilde{L}_{pi}(\{\mathbf{x}_i^t\},\{\mathbf{y}_i^t\},\{\lambda_{i,l}^t\},\{\boldsymbol{\theta}_{i,j}^t\})-\nabla_{\lambda_{i,l}}\widetilde{L}_{pi}(\{\mathbf{x}_i^t\},\{\mathbf{y}_i^t\},\{\lambda_{i,l}^{t-1}\},\{\boldsymbol{\theta}_{i,j}^{t-1}\})||^2\\ &-\frac{c_1^{t-1}L_1'}{L_1'+c_1^{t-1}}||\lambda_{i,l}^t-\lambda_{i,l}^{t-1}||^2).\end{aligned}\tag{S36}$$

Combining Eq.(S28), Eq.(S29), Eq.(S30), Eq.(S31), Eq.(S32), Eq.(S33), Eq.(S34), Eq.(S35) and Eq.(S36), setting $a_3=\eta_\lambda$, $a_2=a_1$, $\frac{\eta_\lambda}{2}\le\frac{1}{L_1'+c_1^0}$, and using $|\mathcal{P}_i^t|<M$, we have

$$\begin{aligned}&\mathbb{E}[L_p(\{\bar{\mathbf{x}}^{t+1}\},\{\bar{\mathbf{y}}^{t+1}\},\{\lambda_{i,l}^{t+1}\},\{\boldsymbol{\theta}_{i,j}^t\})-L_p(\{\bar{\mathbf{x}}^{t+1}\},\{\bar{\mathbf{y}}^{t+1}\},\{\lambda_{i,l}^t\},\{\boldsymbol{\theta}_{i,j}^t\})]\\ \le&\frac{MN^2L^2}{2a_1}(\mathbb{E}[||\bar{\mathbf{x}}^{t+1}-\bar{\mathbf{x}}^t||^2]+\mathbb{E}[||\bar{\mathbf{y}}^{t+1}-\bar{\mathbf{y}}^t||^2])+\frac{MNL^2}{2a_1}(\mathbb{E}[\sum_{i=1}^N||\bar{\mathbf{x}}^t-\mathbf{x}_i^t||^2]+\mathbb{E}[\sum_{i=1}^N||\bar{\mathbf{y}}^t-\mathbf{y}_i^t||^2])\\ &+(a_1-\frac{c_1^{t-1}-c_1^t}{2}+\frac{1}{2\eta_\lambda})\mathbb{E}[\sum_{i=1}^N\sum_{l=1}^{|\mathcal{P}_i^t|}||\lambda_{i,l}^{t+1}-\lambda_{i,l}^t||^2]+\frac{c_1^{t-1}}{2}\mathbb{E}[\sum_{i=1}^N\sum_{l=1}^{\mathcal{P}_i^t}(||\lambda_{i,l}^{t+1}||^2-||\lambda_{i,l}^t||^2)]\\ &+\frac{1}{2\eta_\lambda}\mathbb{E}[\sum_{i=1}^N\sum_{l=1}^{|\mathcal{P}_i^t|}||\lambda_{i,l}^t-\lambda_{i,l}^{t-1}||^2].\end{aligned}\tag{S37}$$

Using $||\bar{\mathbf{x}}^{t+1}-\bar{\mathbf{x}}^t||^2\le\frac{1}{N}\sum_{i=1}^N||\mathbf{x}_i^{t+1}-\bar{\mathbf{x}}^t||^2$ and

$$\mathbb{E}[\sum_{i=1}^N||\bar{\mathbf{y}}^{t+1}-\bar{\mathbf{y}}^t||^2]={\eta_y}^2\mathbb{E}[\sum_{i=1}^N||\bar{\nabla}_{\mathbf{y}_i}\tilde{L}_{p\_i}(\{\mathbf{x}_i^{\hat{t}_i}\},\{\mathbf{y}_i^{\hat{t}_i}\},\{\lambda_{i,l}^{\hat{t}_i}\},\{\boldsymbol{\theta}_{i,j}^{\hat{t}_i}\})||^2],\tag{S38}$$

we have

$$\begin{aligned}&\mathbb{E}[L_p(\{\bar{\mathbf{x}}^{t+1}\},\{\bar{\mathbf{y}}^{t+1}\},\{\lambda_{i,l}^{t+1}\},\{\boldsymbol{\theta}_{i,j}^t\})-L_p(\{\bar{\mathbf{x}}^{t+1}\},\{\bar{\mathbf{y}}^{t+1}\},\{\lambda_{i,l}^t\},\{\boldsymbol{\theta}_{i,j}^t\})]\\ \le&\frac{MNL^2}{2a_1}\mathbb{E}[\sum_{i=1}^N||\mathbf{x}_i^{t+1}-\bar{\mathbf{x}}^t||^2]+\frac{MNL^2\eta_y^2}{2a_1}\mathbb{E}[\sum_{i=1}^N||\bar{\nabla}_{\mathbf{y}_i}\tilde{L}_{p\_i}(\{\mathbf{x}_i^{\hat{t}_i}\},\{\mathbf{y}_i^{\hat{t}_i}\},\{\lambda_{i,l}^{\hat{t}_i}\},\{\boldsymbol{\theta}_{i,j}^{\hat{t}_i}\})||^2]\\ &+\frac{MNL^2}{2a_1}(\mathbb{E}[\sum_{i=1}^N||\bar{\mathbf{x}}^t-\mathbf{x}_i^t||^2]+\mathbb{E}[\sum_{i=1}^N||\bar{\mathbf{y}}^t-\mathbf{y}_i^t||^2])+(a_1-\frac{c_1^{t-1}-c_1^t}{2}+\frac{1}{2\eta_\lambda})\sum_{i=1}^N\sum_{l=1}^{|\mathcal{P}_i^t|}\mathbb{E}[||\lambda_{i,l}^{t+1}-\lambda_{i,l}^t||^2]\\ &+\frac{c_1^{t-1}}{2}\mathbb{E}[\sum_{i=1}^N\sum_{l=1}^{|\mathcal{P}_i^t|}(||\lambda_{i,l}^{t+1}||^2-||\lambda_{i,l}^t||^2)]+\frac{1}{2\eta_\lambda}\mathbb{E}[\sum_{i=1}^N\sum_{l=1}^{|\mathcal{P}_i^t|}||\lambda_{i,l}^t-\lambda_{i,l}^{t-1}||^2].\end{aligned}\tag{S39}$$

Then for $\boldsymbol{\theta}$ variables, $\forall\boldsymbol{\theta}$, it follows that in the $(t+1)^{th}$ iteration, we have:

$$\langle \boldsymbol{\theta}_{i,j}^{t+1} - \boldsymbol{\theta}_{i,j}^{t} - \tilde{\eta}_{i,\theta}^{t} \nabla_{\boldsymbol{\theta}_{i,j}} \tilde{L}_{pi}(\{\mathbf{x}_i^{t+1}\}, \{\mathbf{y}_i^{t+1}\}, \{\lambda_{i,l}^{t+1}\}, \{\boldsymbol{\theta}_{i,j}^{t}\}), \boldsymbol{\theta} - \boldsymbol{\theta}_{i,j}^{t+1} \rangle = 0. \tag{S40}$$

Let $\boldsymbol{\theta} = \boldsymbol{\theta}_{i,j}^{t}$ to have

$$\langle \nabla_{\boldsymbol{\theta}_{i,j}} \tilde{L}_{pi}(\{\mathbf{x}_i^{t+1}\}, \{\mathbf{y}_i^{t+1}\}, \{\lambda_{i,l}^{t+1}\}, \{\boldsymbol{\theta}_{i,j}^{t}\}) - \frac{1}{\tilde{\eta}_{i,\theta}^{t}}(\boldsymbol{\theta}_{i,j}^{t+1} - \boldsymbol{\theta}_{i,j}^{t}), \boldsymbol{\theta}_{i,j}^{t} - \boldsymbol{\theta}_{i,j}^{t+1} \rangle = 0. \tag{S41}$$

Likewise, in the t^{th} iteration we have:

$$\langle \nabla_{\boldsymbol{\theta}_{i,j}} \tilde{L}_{pi}(\{\mathbf{x}_i^{t}\}, \{\mathbf{y}_i^{t}\}, \{\lambda_{i,l}^{t}\}, \{\boldsymbol{\theta}_{i,j}^{t-1}\}) - \frac{1}{\tilde{\eta}_{i,\theta}^{t-1}}(\boldsymbol{\theta}_{i,j}^{t} - \boldsymbol{\theta}_{i,j}^{t-1}), \boldsymbol{\theta}_{i,j}^{t+1} - \boldsymbol{\theta}_{i,j}^{t} \rangle = 0. \tag{S42}$$

Since $\tilde{L}_{pi}(\{\mathbf{x}_i\}, \{\mathbf{y}_i\}, \{\lambda_{i,l}\}, \{\boldsymbol{\theta}_{i,j}\})$ is concave with respect to $\boldsymbol{\theta}_{i,j}$, we have

$$\begin{aligned}
&\widetilde{L}_{pi}(\{\bar{\mathbf{x}}^{t+1}\}, \{\bar{\mathbf{y}}^{t+1}\}, \{\lambda_{i,l}^{t+1}\}, \{\boldsymbol{\theta}_{i,j}^{t+1}\}) - \widetilde{L}_{pi}(\{\bar{\mathbf{x}}^{t+1}\}, \{\bar{\mathbf{y}}^{t+1}\}, \{\lambda_{i,l}^{t+1}\}, \{\boldsymbol{\theta}_{i,j}^{t}\}) \\
\le& \sum_{j=1}^{N} \langle \nabla_{\boldsymbol{\theta}_{i,j}} \widetilde{L}_{pi}(\{\bar{\mathbf{x}}^{t+1}\}, \{\bar{\mathbf{y}}^{t+1}\}, \{\lambda_{i,l}^{t+1}\}, \{\boldsymbol{\theta}_{i,j}^{t}\}), \boldsymbol{\theta}_{i,j}^{t+1} - \boldsymbol{\theta}_{i,j}^{t} \rangle \\
=& \sum_{j=1}^{N} \langle \nabla_{\boldsymbol{\theta}_{i,j}} \widetilde{L}_{pi}(\{\bar{\mathbf{x}}^{t+1}\}, \{\bar{\mathbf{y}}^{t+1}\}, \{\lambda_{i,l}^{t+1}\}, \{\boldsymbol{\theta}_{i,j}^{t}\}) - \nabla_{\boldsymbol{\theta}_{i,j}} \widetilde{L}_{pi}(\{\bar{\mathbf{x}}^{t}\}, \{\bar{\mathbf{y}}^{t}\}, \{\lambda_{i,l}^{t}\}, \{\boldsymbol{\theta}_{i,j}^{t}\}), \boldsymbol{\theta}_{i,j}^{t+1} - \boldsymbol{\theta}_{i,j}^{t} \rangle \\
+& \sum_{j=1}^{N} \langle \nabla_{\boldsymbol{\theta}_{i,j}} \widetilde{L}_{pi}(\{\bar{\mathbf{x}}^{t}\}, \{\bar{\mathbf{y}}^{t}\}, \{\lambda_{i,l}^{t}\}, \{\boldsymbol{\theta}_{i,j}^{t}\}), \boldsymbol{\theta}_{i,j}^{t+1} - \boldsymbol{\theta}_{i,j}^{t} \rangle.
\end{aligned} \tag{S43}$$

For the first term in Eq.(S43), we have

$$\begin{aligned}
&\sum_{j=1}^{N} \langle \nabla_{\boldsymbol{\theta}_{i,j}} \widetilde{L}_{pi}(\{\bar{\mathbf{x}}^{t+1}\}, \{\bar{\mathbf{y}}^{t+1}\}, \{\lambda_{i,l}^{t+1}\}, \{\boldsymbol{\theta}_{i,j}^{t}\}) - \nabla_{\boldsymbol{\theta}_{i,j}} \widetilde{L}_{pi}(\{\bar{\mathbf{x}}^{t}\}, \{\bar{\mathbf{y}}^{t}\}, \{\lambda_{i,l}^{t}\}, \{\boldsymbol{\theta}_{i,j}^{t}\}), \boldsymbol{\theta}_{i,j}^{t+1} - \boldsymbol{\theta}_{i,j}^{t} \rangle \\
=& \sum_{j=1}^{N} \langle \nabla_{\boldsymbol{\theta}_{i,j}} L_{pi}(\{\bar{\mathbf{x}}^{t+1}\}, \{\bar{\mathbf{y}}^{t+1}\}, \{\lambda_{i,l}^{t}\}, \{\boldsymbol{\theta}_{i,j}^{t}\}) - \nabla_{\boldsymbol{\theta}_{i,j}} L_{pi}(\{\bar{\mathbf{x}}^{t}\}, \{\bar{\mathbf{y}}^{t}\}, \{\lambda_{i,l}^{t}\}, \{\boldsymbol{\theta}_{i,j}^{t}\}), \boldsymbol{\theta}_{i,j}^{t+1} - \boldsymbol{\theta}_{i,j}^{t} \rangle \\
+& \frac{c_2^{t-1} - c_2^{t}}{2} \sum_{j=1}^{N} (||\boldsymbol{\theta}_{i,j}^{t+1}||^2 - ||\boldsymbol{\theta}_{i,j}^{t}||^2) - \frac{c_2^{t-1} - c_2^{t}}{2} \sum_{j=1}^{N} ||\boldsymbol{\theta}_{i,j}^{t+1} - \boldsymbol{\theta}_{i,j}^{t}||^2 \\
\le& \frac{NL^2}{2a_4} \sum_{j=1}^{N} (||\bar{\mathbf{x}}^{t+1} - \bar{\mathbf{x}}^{t}||^2 + ||\bar{\mathbf{y}}^{t+1} - \bar{\mathbf{y}}^{t}||^2) + \frac{a_4}{2} \sum_{j=1}^{N} ||\boldsymbol{\theta}_{i,j}^{t+1} - \boldsymbol{\theta}_{i,j}^{t}||^2 \\
+& \frac{c_2^{t-1} - c_2^{t}}{2} \sum_{j=1}^{N} (||\boldsymbol{\theta}_{i,j}^{t+1}||^2 - ||\boldsymbol{\theta}_{i,j}^{t}||^2) - \frac{c_2^{t-1} - c_2^{t}}{2} \sum_{j=1}^{N} ||\boldsymbol{\theta}_{i,j}^{t+1} - \boldsymbol{\theta}_{i,j}^{t}||^2,
\end{aligned} \tag{S44}$$

where $a_4 > 0$ is a constant.

For the second term in Eq.(S43), according to Eq.(S42) we have

$$\begin{aligned}
&\sum_{j=1}^{N} \langle \nabla_{\boldsymbol{\theta}_{i,j}} \widetilde{L}_{pi}(\{\bar{\mathbf{x}}^{t}\}, \{\bar{\mathbf{y}}^{t}\}, \{\lambda_{i,l}^{t}\}, \{\boldsymbol{\theta}_{i,j}^{t}\}), \boldsymbol{\theta}_{i,j}^{t+1} - \boldsymbol{\theta}_{i,j}^{t} \rangle \\
\le& \sum_{j=1}^{N} (\langle \nabla_{\boldsymbol{\theta}_{i,j}} \widetilde{L}_{pi}(\{\bar{\mathbf{x}}^{t}\}, \{\bar{\mathbf{y}}^{t}\}, \{\lambda_{i,l}^{t}\}, \{\boldsymbol{\theta}_{i,j}^{t}\}) - \nabla_{\boldsymbol{\theta}_{i,j}} \widetilde{L}_{pi}(\{\mathbf{x}_i^{t}\}, \{\mathbf{y}_i^{t}\}, \{\lambda_{i,l}^{t}\}, \{\boldsymbol{\theta}_{i,j}^{t-1}\}), \boldsymbol{\theta}_{i,j}^{t+1} - \boldsymbol{\theta}_{i,j}^{t} \rangle \\
+& \frac{1}{\tilde{\eta}_{i,\theta}^{t-1}} \langle \boldsymbol{\theta}_{i,j}^{t} - \boldsymbol{\theta}_{i,j}^{t-1}, \boldsymbol{\theta}_{i,j}^{t+1} - \boldsymbol{\theta}_{i,j}^{t} \rangle).
\end{aligned} \tag{S45}$$

Denoting $v_{2,i,j}^{t+1} = \boldsymbol{\theta}_{i,j}^{t+1} - \boldsymbol{\theta}_{i,j}^{t} - (\boldsymbol{\theta}_{i,j}^{t} - \boldsymbol{\theta}_{i,j}^{t-1})$ and taking the expectation, we can obtain

$$\frac{1}{\eta_\theta}\mathbb{E}[\langle \boldsymbol{\theta}_{i,j}^{t} - \boldsymbol{\theta}_{i,j}^{t-1}, \boldsymbol{\theta}_{i,j}^{t+1} - \boldsymbol{\theta}_{i,j}^{t} \rangle] \le \frac{1}{2\eta_\theta}\mathbb{E}[||\boldsymbol{\theta}_{i,j}^{t+1} - \boldsymbol{\theta}_{i,j}^{t}||^2] - \frac{1}{2\eta_\theta}\mathbb{E}[||v_{2,i,j}^{t+1}||^2] + \frac{1}{2\eta_\theta}\mathbb{E}[||\boldsymbol{\theta}_{i,j}^{t} - \boldsymbol{\theta}_{i,j}^{t-1}||^2 \tag{S46}$$

and

$$
\begin{aligned}
&\sum_{j=1}^{N}\langle \nabla_{\boldsymbol{\theta}_{i,j}}\widetilde{L}_{pi}(\{\bar{\mathbf{x}}^t\},\{\bar{\mathbf{y}}^t\},\{\lambda_{i,l}^t\},\{\boldsymbol{\theta}_{i,j}^t\})-\nabla_{\boldsymbol{\theta}_{i,j}}\widetilde{L}_{pi}(\{\mathbf{x}_i^t\},\{\mathbf{y}_i^t\},\{\lambda_{i,l}^t\},\{\boldsymbol{\theta}_{i,j}^{t-1}\}),\boldsymbol{\theta}_{i,j}^{t+1}-\boldsymbol{\theta}_{i,j}^t\rangle\\
=&\sum_{j=1}^{N}\langle \nabla_{\boldsymbol{\theta}_{i,j}}\widetilde{L}_{pi}(\{\bar{\mathbf{x}}^t\},\{\bar{\mathbf{y}}^t\},\{\lambda_{i,l}^t\},\{\boldsymbol{\theta}_{i,j}^t\})-\nabla_{\boldsymbol{\theta}_{i,j}}\widetilde{L}_{pi}(\{\mathbf{x}_i^t\},\{\mathbf{y}_i^t\},\{\lambda_{i,l}^t\},\{\boldsymbol{\theta}_{i,j}^t\}),\boldsymbol{\theta}_{i,j}^{t+1}-\boldsymbol{\theta}_{i,j}^t\rangle(2a)\\
+&\sum_{j=1}^{N}\langle \nabla_{\boldsymbol{\theta}_{i,j}}\widetilde{L}_{pi}(\{\mathbf{x}_i^t\},\{\mathbf{y}_i^t\},\{\lambda_{i,l}^t\},\{\boldsymbol{\theta}_{i,j}^t\})-\nabla_{\boldsymbol{\theta}_{i,j}}\widetilde{L}_{pi}(\{\mathbf{x}_i^t\},\{\mathbf{y}_i^t\},\{\lambda_{i,l}^t\},\{\boldsymbol{\theta}_{i,j}^{t-1}\}),v_{2,i,j}^{t+1}\rangle(2b)\\
+&\sum_{j=1}^{N}\langle \nabla_{\boldsymbol{\theta}_{i,j}}\widetilde{L}_{pi}(\{\mathbf{x}_i^t\},\{\mathbf{y}_i^t\},\{\lambda_{i,l}^t\},\{\boldsymbol{\theta}_{i,j}^t\})-\nabla_{\boldsymbol{\theta}_{i,j}}\widetilde{L}_{pi}(\{\mathbf{x}_i^t\},\{\mathbf{y}_i^t\},\{\lambda_{i,l}^t\},\{\boldsymbol{\theta}_{i,j}^{t-1}\}),\boldsymbol{\theta}_{i,j}^{t}-\boldsymbol{\theta}_{i,j}^{t-1}\rangle(2c).
\end{aligned}
\tag{S47}
$$

According to the Cauchy-Schwarz inequality with Assumption 1, we have the following inequality from $(2a)$ in Eq.(S47):

$$
\begin{aligned}
&\sum_{j=1}^{N}\langle \nabla_{\boldsymbol{\theta}_{i,j}}\widetilde{L}_{pi}(\{\bar{\mathbf{x}}^t\},\{\bar{\mathbf{y}}^t\},\{\lambda_{i,l}^t\},\{\boldsymbol{\theta}_{i,j}^t\})-\nabla_{\boldsymbol{\theta}_{i,j}}\widetilde{L}_{pi}(\{\mathbf{x}_i^t\},\{\mathbf{y}_i^t\},\{\lambda_{i,l}^t\},\{\boldsymbol{\theta}_{i,j}^t\}),\boldsymbol{\theta}_{i,j}^{t+1}-\boldsymbol{\theta}_{i,j}^t\rangle\\
=&\sum_{j=1}^{N}\langle \nabla_{\boldsymbol{\theta}_{i,j}}L_{pi}(\{\bar{\mathbf{x}}^t\},\{\bar{\mathbf{y}}^t\},\{\lambda_{i,l}^t\},\{\boldsymbol{\theta}_{i,j}^t\})-\nabla_{\boldsymbol{\theta}_{i,j}}L_{pi}(\{\mathbf{x}_i^t\},\{\mathbf{y}_i^t\},\{\lambda_{i,l}^t\},\{\boldsymbol{\theta}_{i,j}^t\}),\boldsymbol{\theta}_{i,j}^{t+1}-\boldsymbol{\theta}_{i,j}^t\rangle\\
\le&\frac{NL^2}{2a_5}\sum_{i=1}^{N}(||\bar{\mathbf{x}}^t-\mathbf{x}_i^t||^2+||\bar{\mathbf{y}}^t-\mathbf{y}_i^t||^2)+\frac{a_5}{2}\sum_{j=1}^{N}||\boldsymbol{\theta}_{i,j}^{t+1}-\boldsymbol{\theta}_{i,j}^t||^2,
\end{aligned}
\tag{S48}
$$

where $a_5 > 0$ is a constant.

According to the Cauchy-Schwarz inequality, (2b) in Eq.(S47) is:

$$
\begin{aligned}
&\sum_{j=1}^{N}\langle \nabla_{\boldsymbol{\theta}_{i,j}}\widetilde{L}_{pi}(\{\mathbf{x}_i^t\},\{\mathbf{y}_i^t\},\{\lambda_{i,l}^t\},\{\boldsymbol{\theta}_{i,j}^t\})-\nabla_{\boldsymbol{\theta}_{i,j}}\widetilde{L}_{pi}(\{\mathbf{x}_i^t\},\{\mathbf{y}_i^t\},\{\lambda_{i,l}^t\},\{\boldsymbol{\theta}_{i,j}^{t-1}\}),v_{2,i,j}^{t+1}\rangle\\
\le&\sum_{j=1}^{N}(\frac{a_6}{2}||\nabla_{\boldsymbol{\theta}_{i,j}}\widetilde{L}_{pi}(\{\mathbf{x}_i^t\},\{\mathbf{y}_i^t\},\{\lambda_{i,l}^t\},\{\boldsymbol{\theta}_{i,j}^t\})-\nabla_{\boldsymbol{\theta}_{i,j}}\widetilde{L}_{pi}(\{\mathbf{x}_i^t\},\{\mathbf{y}_i^t\},\{\lambda_{i,l}^t\},\{\boldsymbol{\theta}_{i,j}^{t-1}\})||^2+\frac{1}{2a_6}||v_{2,i,j}^{t+1}||^2),
\end{aligned}
\tag{S49}
$$

where $a_6 > 0$ is a constant.

Defining $L_2' = L + c_2^0$, combining Assumption 1 and the triangle inequality, $\forall \boldsymbol{\theta}_{i,j}$, we have,

$$
\begin{aligned}
&||\nabla_{\boldsymbol{\theta}_{i,j}}\widetilde{L}_{pi}(\{\mathbf{x}_i^t\},\{\mathbf{y}_i^t\},\{\lambda_{i,l}^t\},\{\boldsymbol{\theta}_{i,j}^t\})-\nabla_{\boldsymbol{\theta}_{i,j}}\widetilde{L}_{pi}(\{\mathbf{x}_i^t\},\{\mathbf{y}_i^t\},\{\lambda_{i,l}^t\},\{\boldsymbol{\theta}_{i,j}^{t-1}\})||\\
=&||\nabla_{\boldsymbol{\theta}_{i,j}}L_{pi}(\{\mathbf{x}_i^t\},\{\mathbf{y}_i^t\},\{\lambda_{i,l}^t\},\{\boldsymbol{\theta}_{i,j}^t\})-\nabla_{\boldsymbol{\theta}_{i,j}}L_{pi}(\{\mathbf{x}_i^t\},\{\mathbf{y}_i^t\},\{\lambda_{i,l}^t\},\{\boldsymbol{\theta}_{i,j}^{t-1}\})-c_2^{t-1}(\boldsymbol{\theta}_{i,j}^t-\boldsymbol{\theta}_{i,j}^{t-1})||\\
\le&(L+c_2^{t-1})||\boldsymbol{\theta}_{i,j}^t-\boldsymbol{\theta}_{i,j}^{t-1}||\\
\le&L_2'||\boldsymbol{\theta}_{i,j}^t-\boldsymbol{\theta}_{i,j}^{t-1}||.
\end{aligned}
\tag{S50}
$$

Since $\widetilde{L}_{pi}(\{\mathbf{x}_i\},\{\mathbf{y}_i^t\},\{\lambda_{i,l}\},\{\boldsymbol{\theta}_{i,j}\})$ is strongly concave with respect to $\boldsymbol{\theta}_{i,j}$, we have

$$
\begin{aligned}
&\sum_{j=1}^{N}\langle \nabla_{\boldsymbol{\theta}_{i,j}}\widetilde{L}_{pi}(\{\mathbf{x}_i^t\},\{\mathbf{y}_i^t\},\{\lambda_{i,l}^t\},\{\boldsymbol{\theta}_{i,j}^t\})-\nabla_{\boldsymbol{\theta}_{i,j}}\widetilde{L}_{pi}(\{\mathbf{x}_i^t\},\{\mathbf{y}_i^t\},\{\lambda_{i,l}^t\},\{\boldsymbol{\theta}_{i,j}^{t-1}\}),\boldsymbol{\theta}_{i,j}^{t}-\boldsymbol{\theta}_{i,j}^{t-1}\rangle\\
\le&\sum_{j=1}^{N}(-\frac{1}{L_2'+c_2^{t-1}}||\nabla_{\boldsymbol{\theta}_{i,j}}\widetilde{L}_{pi}(\{\mathbf{x}_i^t\},\{\mathbf{y}_i^t\},\{\lambda_{i,l}^t\},\{\boldsymbol{\theta}_{i,j}^t\})-\nabla_{\boldsymbol{\theta}_{i,j}}\widetilde{L}_{pi}(\{\mathbf{x}_i^t\},\{\mathbf{y}_i^t\},\{\lambda_{i,l}^t\},\{\boldsymbol{\theta}_{i,j}^{t-1}\})||^2.
\end{aligned}
\tag{S51}
$$

Combining Eq.(S43), Eq.(S44), Eq.(S45), Eq.(S46), Eq.(S47), Eq.(S48), Eq.(S49), Eq.(S50) and Eq.(S51), let $a_6 = \eta_\theta$,

$a_5 = a_4$, $\frac{\eta_\theta}{2} \le \frac{1}{L_2'+c_2^0}$, we have

$$\begin{aligned}
&\mathbb{E}[L_{pi}(\{\bar{\mathbf{x}}^{t+1}\},\{\bar{\mathbf{y}}^{t+1}\},\{\lambda_{i,l}^{t+1}\},\{\boldsymbol{\theta}_{i,j}^{t+1}\}) - L_{pi}(\{\bar{\mathbf{x}}^{t+1}\},\{\bar{\mathbf{y}}^{t+1}\},\{\lambda_{i,l}^{t+1}\},\{\boldsymbol{\theta}_{i,j}^{t}\})] \\
\le&\frac{N^3L^2}{2a_4}(\mathbb{E}[||\bar{\mathbf{x}}^{t+1}-\bar{\mathbf{x}}^t||^2] + \mathbb{E}[||\bar{\mathbf{y}}^{t+1}-\bar{\mathbf{y}}^t||^2]) + \frac{N^2L^2}{2a_4}(\mathbb{E}[\sum_{i=1}^N||\bar{\mathbf{x}}^t-\mathbf{x}_i^t||^2] + \mathbb{E}[\sum_{i=1}^N||\bar{\mathbf{y}}^t-\mathbf{y}_i^t||^2]) \\
&+(a_4 - \frac{c_2^{t-1}-c_2^t}{2} + \frac{1}{2\eta_\theta})\mathbb{E}[\sum_{i=1}^N\sum_{j=1}^N||\boldsymbol{\theta}_{i,j}^{t+1}-\boldsymbol{\theta}_{i,j}^t||^2] + \frac{c_2^{t-1}}{2}\mathbb{E}[\sum_{i=1}^N\sum_{j=1}^N(||\boldsymbol{\theta}_{i,j}^{t+1}||^2 - ||\boldsymbol{\theta}_{i,j}^t||^2]) \\
&+\frac{1}{2\eta_\theta}\mathbb{E}[\sum_{i=1}^N\sum_{j=1}^N||\boldsymbol{\theta}_{i,j}^t-\boldsymbol{\theta}_{i,j}^{t-1}||^2].
\end{aligned} \tag{S52}$$

Using $||\bar{\mathbf{x}}^{t+1}-\bar{\mathbf{x}}^t||^2 \le \frac{1}{N}\sum_{i=1}^N||\mathbf{x}_i^{t+1}-\bar{\mathbf{x}}^t||^2$ and Eq.(S38), we have

$$\begin{aligned}
&\mathbb{E}[L_p(\{\bar{\mathbf{x}}^{t+1}\},\{\bar{\mathbf{y}}^{t+1}\},\{\lambda_{i,l}^{t+1}\},\{\boldsymbol{\theta}_{i,j}^{t+1}\}) - L_p(\{\bar{\mathbf{x}}^{t+1}\},\{\bar{\mathbf{y}}^{t+1}\},\{\lambda_{i,l}^{t+1}\},\{\boldsymbol{\theta}_{i,j}^{t}\})] \\
\le&\frac{N^2L^2}{2a_4}\sum_{i=1}^N\mathbb{E}[||\mathbf{x}_i^{t+1}-\bar{\mathbf{x}}^t||^2] + \frac{N^2L^2\eta_y^2}{2a_4}\mathbb{E}[\sum_{i=1}^N||\nabla_{\mathbf{y}_i}\tilde{L}_{pi}(\{\mathbf{x}_i^{\hat{t}_i}\},\{\mathbf{y}_i^{\hat{t}_i}\},\{\lambda_{i,l}^{\hat{t}_i}\},\{\boldsymbol{\theta}_{i,j}^{\hat{t}_i}\})||^2] \\
&+\frac{N^2L^2}{2a_4}(\mathbb{E}[\sum_{i=1}^N||\bar{\mathbf{x}}^t-\mathbf{x}_i^t||^2] + \mathbb{E}[\sum_{i=1}^N||\bar{\mathbf{y}}^t-\mathbf{y}_i^t||^2]) + (a_4 - \frac{c_2^{t-1}-c_2^t}{2} + \frac{1}{2\eta_\theta})\mathbb{E}[\sum_{i=1}^N\sum_{j=1}^N||\boldsymbol{\theta}_{i,j}^{t+1}-\boldsymbol{\theta}_{i,j}^t||^2] \\
&+\frac{c_2^{t-1}}{2}\mathbb{E}[\sum_{i=1}^N\sum_{j=1}^N(||\boldsymbol{\theta}_{i,j}^{t+1}||^2 - ||\boldsymbol{\theta}_{i,j}^t||^2)] + \frac{1}{2\eta_\theta}\sum_{i=1}^N\mathbb{E}[\sum_{j=1}^N||\boldsymbol{\theta}_{i,j}^t-\boldsymbol{\theta}_{i,j}^{t-1}||^2].
\end{aligned} \tag{S53}$$

Combining Eq.(S7), Eq.(S20), Eq.(S39), and Eq.(S53), setting $a_1 = \frac{1}{2\eta_\lambda}$, $a_4 = \frac{1}{2\eta_\theta}$, then lemma 3 can be proved. □

4) Lemma 4 (Iterates Contraction): The following contraction properties of iterates hold:

$$\begin{aligned}
\mathbb{E}[\sum_{i=1}^N||\mathbf{x}_i^{t+1}-\bar{\mathbf{x}}^{t+1}||^2]\le&(\rho+\frac{170\eta_x^2L^2}{1-\rho})\mathbb{E}[\sum_{i=1}^N||\mathbf{x}_i^t-\bar{\mathbf{x}}^t||^2] + \frac{40\eta_x^2L^2}{1-\rho}\mathbb{E}[\sum_{i=1}^N||\mathbf{x}_i^{t+1}-\mathbf{d}_i^t||^2] \\
&+\frac{60\eta_x^2L^2}{(1-\rho)^2}\mathbb{E}[\sum_{i=1}^N||\bar{\mathbf{y}}^t-\mathbf{y}_i^t||^2] + \frac{40\eta_x^2\eta_y^2L^2}{(1-\rho)^2}\mathbb{E}[\sum_{i=1}^N||\nabla_{\mathbf{y}_i}\tilde{L}_{pi}(\{\mathbf{x}_i^{\hat{t}_i}\},\{\mathbf{y}_i^{\hat{t}_i}\},\{\lambda_{i,l}^{\hat{t}_i}\},\{\boldsymbol{\theta}_{i,j}^{\hat{t}_i}\})||^2],
\end{aligned} \tag{S54}$$

$$\begin{aligned}
\mathbb{E}[\sum_{i=1}^N||\mathbf{y}^{t+1}-\bar{\mathbf{y}}^{t+1}||^2]\le&(\rho+\frac{230\eta_y^2L^2}{(1-\rho)^2})\mathbb{E}[\sum_{i=1}^N||\mathbf{y}_i^t-\bar{\mathbf{y}}^t||^2] + \frac{40\eta_y^2L^2}{1-\rho}\mathbb{E}[\sum_{i=1}^N||\mathbf{x}_i^{t+1}-\mathbf{d}_i^t||^2] \\
&+\frac{40\eta_y^4L^2}{(1-\rho)^2}\mathbb{E}[\sum_{i=1}^N||\nabla_{\mathbf{y}_i}\tilde{L}_{pi}(\{\mathbf{x}_i^{\hat{t}_i}\},\{\mathbf{y}_i^{\hat{t}_i}\},\{\lambda_{i,l}^{\hat{t}_i}\},\{\boldsymbol{\theta}_{i,j}^{\hat{t}_i}\})||^2].
\end{aligned} \tag{S55}$$

Proof. First, since $\mathbf{e}\mathbf{e}^\top$ is a projection operator, for any matrix $\mathbf{A} \in \mathbb{R}^{N\times n}$ or $\mathbb{R}^{N\times m}$,

$$\begin{aligned}
||\mathbf{A} - \frac{1}{N}\mathbf{e}\mathbf{e}^\top\mathbf{A}||_F^2 &= \|\mathbf{A}\|_F^2 - 2\left\langle \mathbf{A}, \frac{1}{N}\mathbf{e}\mathbf{e}^\top\mathbf{A}\right\rangle + ||\frac{1}{N}\mathbf{e}\mathbf{e}^\top\mathbf{A}||_F^2 \\
&= \|\mathbf{A}\|_F^2 - 2||\frac{1}{N}\mathbf{e}\mathbf{e}^\top\mathbf{A}||_F^2 + ||\frac{1}{N}\mathbf{e}\mathbf{e}^\top\mathbf{A}||_F^2 = \|\mathbf{A}\|_F^2 - ||\frac{1}{N}\mathbf{e}\mathbf{e}^\top\mathbf{A}||_F^2.
\end{aligned} \tag{S56}$$

Using the compatibility of the Frobenius norm and the 2-norm, and considering Assumption 2, we have

$$||(\mathbf{W}^t - \frac{1}{N}\mathbf{e}\mathbf{e}^\top)(\mathbf{x}^t - \frac{1}{N}\mathbf{e}\mathbf{e}^\top\mathbf{x}^t)||_F^2 \le ||(\mathbf{W}^t - \frac{1}{N}\mathbf{e}\mathbf{e}^\top)||_2^2||\mathbf{x}^t - \frac{1}{N}\mathbf{e}\mathbf{e}^\top\mathbf{x}^t||_F^2 = \rho^2||\mathbf{x}^t - \frac{1}{N}\mathbf{e}\mathbf{e}^\top\mathbf{x}^t||_F^2, \tag{S57}$$

$$||(\mathbf{W}^t - \frac{1}{N}\mathbf{e}\mathbf{e}^\top)(\mathbf{y}^t - \frac{1}{N}\mathbf{e}\mathbf{e}^\top\mathbf{y}^t)||_F^2 \le ||(\mathbf{W}^t - \frac{1}{N}\mathbf{e}\mathbf{e}^\top)||_2^2||\mathbf{y}^t - \frac{1}{N}\mathbf{e}\mathbf{e}^\top\mathbf{y}^t||_F^2 = \rho^2||\mathbf{y}^t - \frac{1}{N}\mathbf{e}\mathbf{e}^\top\mathbf{y}^t||_F^2. \tag{S58}$$

Further, we have

$$\begin{aligned}
&||\mathbf{x}^{t+1} - \mathbf{x}^t||_F^2 \\
=&||\mathbf{x}^{t+1} - \mathbf{d}^t + \mathbf{d}^t - \mathbf{x}^t||_F^2 \\
\le&2||\mathbf{x}^{t+1} - \mathbf{d}^t||_F^2 + 2||\mathbf{W}^t\mathbf{x}^t - \mathbf{W}^t\frac{1}{N}\mathbf{e}\mathbf{e}^\top\mathbf{x}^t + \mathbf{W}^t\frac{1}{N}\mathbf{e}\mathbf{e}^\top\mathbf{x}^t - \mathbf{x}^t||_F^2 \\
=&2||\mathbf{x}^{t+1} - \mathbf{d}^t||_F^2 + 2||(\mathbf{I} - \mathbf{W}^t)(\mathbf{x}^t - \frac{1}{N}\mathbf{e}\mathbf{e}^\top\mathbf{x}^t)||_F^2 \\
\le&2||\mathbf{x}^{t+1} - \mathbf{d}^t||_F^2 + 8||\mathbf{x}^t - \frac{1}{N}\mathbf{e}\mathbf{e}^\top\mathbf{x}^t||_F^2,
\end{aligned} \tag{S59}$$

and

$$
\begin{aligned}
&\mathbb{E}[||\mathbf{y}^{t+1}-\mathbf{y}^{t}||_F^2]\\
=&\mathbb{E}[||\mathbf{W}^t\mathbf{y}^t-\mathbf{y}^t-\tilde{\eta}_y^t\nabla_{\mathbf{y}}\tilde{L}_p(\{\mathbf{x}_i^{\hat{t}_i}\},\{\mathbf{y}_i^{\hat{t}_i}\},\{\lambda_{i,l}^{\hat{t}_i}\},\{\boldsymbol{\theta}_{i,j}^{\hat{t}_i}\})||_F^2]\\
=&\mathbb{E}[||\mathbf{W}^t(\mathbf{y}^t-\frac{1}{N}\mathbf{e}\mathbf{e}^\top\mathbf{y}^t)+\frac{1}{N}\mathbf{e}\mathbf{e}^\top\mathbf{y}^t-\mathbf{y}^t-\tilde{\eta}_y^t\nabla_{\mathbf{y}}\tilde{L}_p(\{\mathbf{x}_i^{\hat{t}_i}\},\{\mathbf{y}_i^{\hat{t}_i}\},\{\lambda_{i,l}^{\hat{t}_i}\},\{\boldsymbol{\theta}_{i,j}^{\hat{t}_i}\})||_F^2]\\
\leq&\rho^2(1+\delta)||\mathbf{y}^t-\frac{1}{N}\mathbf{e}\mathbf{e}^\top\mathbf{y}^t||_F^2+(1+\frac{1}{\delta})\mathbb{E}[||\frac{1}{N}\mathbf{e}\mathbf{e}^\top\mathbf{y}^t-\mathbf{y}^t-\tilde{\eta}_y^t\nabla_{\mathbf{y}}\tilde{L}_p(\{\mathbf{x}_i^{\hat{t}_i}\},\{\mathbf{y}_i^{\hat{t}_i}\},\{\lambda_{i,l}^{\hat{t}_i}\},\{\boldsymbol{\theta}_{i,j}^{\hat{t}_i}\})||_F^2]\\
\leq&\rho^2(1+\delta)||\mathbf{y}^t-\frac{1}{N}\mathbf{e}\mathbf{e}^\top\mathbf{y}^t||_F^2+2(1+\frac{1}{\delta})||\frac{1}{N}\mathbf{e}\mathbf{e}^\top\mathbf{y}^t-\mathbf{y}^t||_F^2+2(\eta_y)^2(1+\frac{1}{\delta})||\nabla_{\mathbf{y}}\tilde{L}_p(\{\mathbf{x}_i^{\hat{t}_i}\},\{\mathbf{y}_i^{\hat{t}_i}\},\{\lambda_{i,l}^{\hat{t}_i}\},\{\boldsymbol{\theta}_{i,j}^{\hat{t}_i}\})||_F^2\\
=&(\rho+\frac{2}{1-\rho})||\mathbf{y}^t-\frac{1}{N}\mathbf{e}\mathbf{e}^\top\mathbf{y}^t||_F^2+\frac{2\eta_y^2}{1-\rho}||\nabla_{\mathbf{y}}\tilde{L}_p(\{\mathbf{x}_i^{\hat{t}_i}\},\{\mathbf{y}_i^{\hat{t}_i}\},\{\lambda_{i,l}^{\hat{t}_i}\},\{\boldsymbol{\theta}_{i,j}^{\hat{t}_i}\})||_F^2,
\end{aligned}
\tag{S60}
$$

where (a) using $\delta=\frac{1-\rho}{\rho}$.

According to the updating rules in Eq.(17) and the Peter-Paul inequality, we have

$$
\begin{aligned}
&||\mathbf{x}^{t+1}-\bar{\mathbf{x}}^{t+1}||_F^2\\
&=||\operatorname{prox}^R(\mathbf{W}^t\mathbf{x}^t-\tilde{\eta}_x^t\nabla_{\mathbf{x}}\tilde{L}_p(\{\mathbf{x}_i^{\hat{t}_i}\},\{\mathbf{y}_i^{\hat{t}_i}\},\{\lambda_{i,l}^{\hat{t}_i}\},\{\boldsymbol{\theta}_{i,j}^{\hat{t}_i}\}))-\frac{1}{N}\mathbf{e}\mathbf{e}^\top\operatorname{prox}_R(\mathbf{W}^t\mathbf{x}^t-\tilde{\eta}_x^t\nabla_{\mathbf{x}}\tilde{L}_p(\{\mathbf{x}_i^{\hat{t}_i}\},\{\mathbf{y}_i^{\hat{t}_i}\},\{\lambda_{i,l}^{\hat{t}_i}\},\{\boldsymbol{\theta}_{i,j}^{\hat{t}_i}\}))||_F^2\\
&\overset{(a)}{=}||\operatorname{prox}^R(\mathbf{W}^t\mathbf{x}^t-\tilde{\eta}_x^t\nabla_{\mathbf{x}}\tilde{L}_p(\{\mathbf{x}_i^{\hat{t}_i}\},\{\mathbf{y}_i^{\hat{t}_i}\},\{\lambda_{i,l}^{\hat{t}_i}\},\{\boldsymbol{\theta}_{i,j}^{\hat{t}_i}\}))-\operatorname{prox}^R(\frac{1}{N}\mathbf{e}\mathbf{e}^\top\mathbf{x}^t-\tilde{\eta}_x^t\bar{\nabla}_{\mathbf{x}}\tilde{L}_p(\{\mathbf{x}_i^{\hat{t}_i}\},\{\mathbf{y}_i^{\hat{t}_i}\},\{\lambda_{i,l}^{\hat{t}_i}\},\{\boldsymbol{\theta}_{i,j}^{\hat{t}_i}\}))||_F^2\\
&-||\frac{1}{N}\mathbf{e}\mathbf{e}^\top\operatorname{prox}^R(\frac{1}{N}\mathbf{e}\mathbf{e}^\top\mathbf{x}^t-\tilde{\eta}_x^t\bar{\nabla}_{\mathbf{x}}\tilde{L}_p(\{\mathbf{x}_i^{\hat{t}_i}\},\{\mathbf{y}_i^{\hat{t}_i}\},\{\lambda_{i,l}^{\hat{t}_i}\},\{\boldsymbol{\theta}_{i,j}^{\hat{t}_i}\}))\\
&-\operatorname{prox}^R(\mathbf{W}^t\mathbf{x}^t-\tilde{\eta}_x^t\nabla_{\mathbf{x}}\tilde{L}_p(\{\mathbf{x}_i^{\hat{t}_i}\},\{\mathbf{y}_i^{\hat{t}_i}\},\{\lambda_{i,l}^{\hat{t}_i}\},\{\boldsymbol{\theta}_{i,j}^{\hat{t}_i}\}))||_F^2\\
&\overset{(b)}{\leq}||\mathbf{W}^t\mathbf{x}^t-\frac{1}{N}\mathbf{e}\mathbf{e}^\top\mathbf{x}^t-\tilde{\eta}_x^t(\nabla_{\mathbf{x}}\tilde{L}_p(\{\mathbf{x}_i^{\hat{t}_i}\},\{\mathbf{y}_i^{\hat{t}_i}\},\{\lambda_{i,l}^{\hat{t}_i}\},\{\boldsymbol{\theta}_{i,j}^{\hat{t}_i}\})-\bar{\nabla}_{\mathbf{x}}\tilde{L}_p(\{\mathbf{x}_i^{\hat{t}_i}\},\{\mathbf{y}_i^{\hat{t}_i}\},\{\lambda_{i,l}^{\hat{t}_i}\},\{\boldsymbol{\theta}_{i,j}^{\hat{t}_i}\}))||_F^2\\
&=||\mathbf{W}^t\mathbf{x}^t-\frac{1}{N}\mathbf{e}\mathbf{e}^\top\mathbf{x}^t||_F^2+||\tilde{\eta}_x^t(\nabla_{\mathbf{x}}\tilde{L}_p(\{\mathbf{x}_i^{\hat{t}_i}\},\{\mathbf{y}_i^{\hat{t}_i}\},\{\lambda_{i,l}^{\hat{t}_i}\},\{\boldsymbol{\theta}_{i,j}^{\hat{t}_i}\})-\bar{\nabla}_{\mathbf{x}}\tilde{L}_p(\{\mathbf{x}_i^{\hat{t}_i}\},\{\mathbf{y}_i^{\hat{t}_i}\},\{\lambda_{i,l}^{\hat{t}_i}\},\{\boldsymbol{\theta}_{i,j}^{\hat{t}_i}\}))||_F^2\\
&-2\langle\mathbf{W}^t\mathbf{x}^t-\frac{1}{N}\mathbf{e}\mathbf{e}^\top\mathbf{x}^t,\tilde{\eta}_x^t(\nabla_{\mathbf{x}}\tilde{L}_p(\{\mathbf{x}_i^{\hat{t}_i}\},\{\mathbf{y}_i^{\hat{t}_i}\},\{\lambda_{i,l}^{\hat{t}_i}\},\{\boldsymbol{\theta}_{i,j}^{\hat{t}_i}\})-\bar{\nabla}_{\mathbf{x}}\tilde{L}_p(\{\mathbf{x}_i^{\hat{t}_i}\},\{\mathbf{y}_i^{\hat{t}_i}\},\{\lambda_{i,l}^{\hat{t}_i}\},\{\boldsymbol{\theta}_{i,j}^{\hat{t}_i}\}))\rangle\\
&\leq||\mathbf{W}^t\mathbf{x}^t-\frac{1}{N}\mathbf{e}\mathbf{e}^\top\mathbf{x}^t||_F^2+||\tilde{\eta}_x^t(\nabla_{\mathbf{x}}\tilde{L}_p(\{\mathbf{x}_i^{\hat{t}_i}\},\{\mathbf{y}_i^{\hat{t}_i}\},\{\lambda_{i,l}^{\hat{t}_i}\},\{\boldsymbol{\theta}_{i,j}^{\hat{t}_i}\})-\bar{\nabla}_{\mathbf{x}}\tilde{L}_p(\{\mathbf{x}_i^{\hat{t}_i}\},\{\mathbf{y}_i^{\hat{t}_i}\},\{\lambda_{i,l}^{\hat{t}_i}\},\{\boldsymbol{\theta}_{i,j}^{\hat{t}_i}\}))||_F^2\\
&+\delta||\mathbf{W}^t\mathbf{x}^t-\frac{1}{N}\mathbf{e}\mathbf{e}^\top\mathbf{x}^t||_F^2+\frac{1}{\delta}||\tilde{\eta}_x^t(\nabla_{\mathbf{x}}\tilde{L}_p(\{\mathbf{x}_i^{\hat{t}_i}\},\{\mathbf{y}_i^{\hat{t}_i}\},\{\lambda_{i,l}^{\hat{t}_i}\},\{\boldsymbol{\theta}_{i,j}^{\hat{t}_i}\})-\bar{\nabla}_{\mathbf{x}}\tilde{L}_p(\{\mathbf{x}_i^{\hat{t}_i}\},\{\mathbf{y}_i^{\hat{t}_i}\},\{\lambda_{i,l}^{\hat{t}_i}\},\{\boldsymbol{\theta}_{i,j}^{\hat{t}_i}\}))||_F^2\\
&\overset{(c)}{=}(1+\delta)||(\mathbf{W}^t-\frac{1}{N}\mathbf{e}\mathbf{e}^\top)(\mathbf{x}^t-\frac{1}{N}\mathbf{e}\mathbf{e}^\top\mathbf{x}^t)||_F^2\\
&+(1+\frac{1}{\delta})(\tilde{\eta}_x^t)^2||\nabla_{\mathbf{x}}\tilde{L}_p(\{\mathbf{x}_i^{\hat{t}_i}\},\{\mathbf{y}_i^{\hat{t}_i}\},\{\lambda_{i,l}^{\hat{t}_i}\},\{\boldsymbol{\theta}_{i,j}^{\hat{t}_i}\})-\bar{\nabla}_{\mathbf{x}}\tilde{L}_p(\{\mathbf{x}_i^{\hat{t}_i}\},\{\mathbf{y}_i^{\hat{t}_i}\},\{\lambda_{i,l}^{\hat{t}_i}\},\{\boldsymbol{\theta}_{i,j}^{\hat{t}_i}\})||_F^2\\
&\overset{(d)}{\leq}\rho||(\mathbf{W}^t-\frac{1}{N}\mathbf{e}\mathbf{e}^\top)(\mathbf{x}^t-\frac{1}{N}\mathbf{e}\mathbf{e}^\top\mathbf{x}^t)||_F^2\\
&+\frac{(\tilde{\eta}_x)^2}{1-\rho}||\nabla_{\mathbf{x}}\tilde{L}_p(\{\mathbf{x}_i^{\hat{t}_i}\},\{\mathbf{y}_i^{\hat{t}_i}\},\{\lambda_{i,l}^{\hat{t}_i}\},\{\boldsymbol{\theta}_{i,j}^{\hat{t}_i}\})-\bar{\nabla}_{\mathbf{x}}\tilde{L}_p(\{\mathbf{x}_i^{\hat{t}_i}\},\{\mathbf{y}_i^{\hat{t}_i}\},\{\lambda_{i,l}^{\hat{t}_i}\},\{\boldsymbol{\theta}_{i,j}^{\hat{t}_i}\})||_F^2,
\end{aligned}
\tag{S61}
$$

where $\bar{\nabla}_{\mathbf{x}}\tilde{L}_p(\{\mathbf{x}_i^{\hat{t}_i}\},\{\mathbf{y}_i^{\hat{t}_i}\},\{\lambda_{i,l}^{\hat{t}_i}\},\{\boldsymbol{\theta}_{i,j}^{\hat{t}_i}\})=\frac{1}{N}\mathbf{e}\mathbf{e}^\top\nabla_{\mathbf{x}}\tilde{L}_p(\{\mathbf{x}_i^{\hat{t}_i}\},\{\mathbf{y}_i^{\hat{t}_i}\},\{\lambda_{i,l}^{\hat{t}_i}\},\{\boldsymbol{\theta}_{i,j}^{\hat{t}_i}\})$. (a) utilizes Eq.(S56). (b) is from $||\operatorname{prox}_r(\mathbf{a})-\operatorname{prox}_r(\mathbf{b})||_2\leq\|\mathbf{a}-\mathbf{b}\|_2$, for any $\mathbf{a},\mathbf{b}\in\mathbb{R}^n$ when r is a closed, convex function. (c) uses $(\mathbf{W}^t-\frac{1}{N}\mathbf{e}\mathbf{e}^\top)=(\mathbf{W}^t-\frac{1}{N}\mathbf{e}\mathbf{e}^\top)(\mathbf{I}-\frac{1}{N}\mathbf{e}\mathbf{e}^\top)$. (d) uses $\delta=\frac{1-\rho}{\rho}$ and Eq.(S57).

According to Assumption 1 and Assumption 2, we have

$$
\begin{aligned}
&\mathbb{E}[||\nabla_{\mathbf{x}}\tilde{L}_p(\{\mathbf{x}_i^{\hat{t}_i}\},\{\mathbf{y}_i^{\hat{t}_i}\},\{\lambda_{i,l}^{\hat{t}_i}\},\{\boldsymbol{\theta}_{i,j}^{\hat{t}_i}\})-\bar{\nabla}_{\mathbf{x}}\tilde{L}_p(\{\mathbf{x}_i^{\hat{t}_i}\},\{\mathbf{y}_i^{\hat{t}_i}\},\{\lambda_{i,l}^{\hat{t}_i}\},\{\boldsymbol{\theta}_{i,j}^{\hat{t}_i}\})||_F^2]\\
\leq&10L^2(\mathbb{E}[\sum_{i=1}^N||\mathbf{x}_i^t-\mathbf{x}_i^{\hat{t}_i}||^2]+\mathbb{E}[\sum_{i=1}^N||\mathbf{y}_i^t-\mathbf{y}_i^{\hat{t}_i}||^2]+\mathbb{E}[\sum_{i=1}^N||\bar{\mathbf{x}}^t-\mathbf{x}_i^t||^2])\\
\leq&20L^2(\mathbb{E}[\sum_{i=1}^N||\mathbf{x}_i^{t+1}-\mathbf{x}_i^t||^2]+\mathbb{E}[\sum_{i=1}^N||\mathbf{y}_i^{t+1}-\mathbf{y}_i^t||^2])+10L^2\mathbb{E}[\sum_{i=1}^N||\bar{\mathbf{x}}^t-\mathbf{x}_i^t||^2].
\end{aligned}
\tag{S62}
$$

Plugging Eq.(S59), Eq.(S60), and Eq.(S62) into Eq.(S61), then Eq.(S54) can be proved.

According to the updating rules in Eq.(18) and the Peter-Paul inequality, we have

$$
\begin{aligned}
&||\mathbf{y}^{t+1}-\bar{\mathbf{y}}^{t+1}||_F^2\\
&=||(\mathbf{W}^t\mathbf{y}^t-\tilde{\eta}_y^t\nabla_{\mathbf{y}}\tilde{L}_p(\{\mathbf{x}_i^{\hat{t}_i}\},\{\mathbf{y}_i^{\hat{t}_i}\},\{\lambda_{i,l}^{\hat{t}_i}\},\{\boldsymbol{\theta}_{i,j}^{\hat{t}_i}\}))-\frac{1}{N}\mathbf{e}\mathbf{e}^\top(\mathbf{W}^t\mathbf{y}^t-\tilde{\eta}_y^t\nabla_{\mathbf{y}}\tilde{L}_p(\{\mathbf{x}_i^{\hat{t}_i}\},\{\mathbf{y}_i^{\hat{t}_i}\},\{\lambda_{i,l}^{\hat{t}_i}\},\{\boldsymbol{\theta}_{i,j}^{\hat{t}_i}\}))||_F^2\\
&\overset{(a)}{=}||(\mathbf{W}^t\mathbf{y}^t-\tilde{\eta}_y^t\nabla_{\mathbf{y}}\tilde{L}_p(\{\mathbf{x}_i^{\hat{t}_i}\},\{\mathbf{y}_i^{\hat{t}_i}\},\{\lambda_{i,l}^{\hat{t}_i}\},\{\boldsymbol{\theta}_{i,j}^{\hat{t}_i}\}))-(\frac{1}{N}\mathbf{e}\mathbf{e}^\top\mathbf{y}^t-\tilde{\eta}_y^t\bar{\nabla}_{\mathbf{y}}\tilde{L}_p(\{\mathbf{x}_i^{\hat{t}_i}\},\{\mathbf{y}_i^{\hat{t}_i}\},\{\lambda_{i,l}^{\hat{t}_i}\},\{\boldsymbol{\theta}_{i,j}^{\hat{t}_i}\}))||_F^2\\
&-||\frac{1}{N}\mathbf{e}\mathbf{e}^\top((\frac{1}{N}\mathbf{e}\mathbf{e}^\top\mathbf{y}^t-\tilde{\eta}_y^t\bar{\nabla}_{\mathbf{y}}\tilde{L}_p(\{\mathbf{x}_i^{\hat{t}_i}\},\{\mathbf{y}_i^{\hat{t}_i}\},\{\lambda_{i,l}^{\hat{t}_i}\},\{\boldsymbol{\theta}_{i,j}^{\hat{t}_i}\}))-(\mathbf{W}^t\mathbf{y}^t-\tilde{\eta}_y^t\nabla_{\mathbf{y}}\tilde{L}_p(\{\mathbf{x}_i^{\hat{t}_i}\},\{\mathbf{y}_i^{\hat{t}_i}\},\{\lambda_{i,l}^{\hat{t}_i}\},\{\boldsymbol{\theta}_{i,j}^{\hat{t}_i}\})))||_F^2\\
&\le||\mathbf{W}^t\mathbf{y}^t-\frac{1}{N}\mathbf{e}\mathbf{e}^\top\mathbf{y}^t-\tilde{\eta}_y^t(\nabla_{\mathbf{y}}\tilde{L}_p(\{\mathbf{x}_i^{\hat{t}_i}\},\{\mathbf{y}_i^{\hat{t}_i}\},\{\lambda_{i,l}^{\hat{t}_i}\},\{\boldsymbol{\theta}_{i,j}^{\hat{t}_i}\})-\bar{\nabla}_{\mathbf{y}}\tilde{L}_p(\{\mathbf{x}_i^{\hat{t}_i}\},\{\mathbf{y}_i^{\hat{t}_i}\},\{\lambda_{i,l}^{\hat{t}_i}\},\{\boldsymbol{\theta}_{i,j}^{\hat{t}_i}\}))||_F^2\\
&=||\mathbf{W}^t\mathbf{y}^t-\frac{1}{N}\mathbf{e}\mathbf{e}^\top\mathbf{y}^t||_F^2+||\tilde{\eta}_y^t(\nabla_{\mathbf{y}}\tilde{L}_p(\{\mathbf{x}_i^{\hat{t}_i}\},\{\mathbf{y}_i^{\hat{t}_i}\},\{\lambda_{i,l}^{\hat{t}_i}\},\{\boldsymbol{\theta}_{i,j}^{\hat{t}_i}\})-\bar{\nabla}_{\mathbf{y}}\tilde{L}_p(\{\mathbf{x}_i^{\hat{t}_i}\},\{\mathbf{y}_i^{\hat{t}_i}\},\{\lambda_{i,l}^{\hat{t}_i}\},\{\boldsymbol{\theta}_{i,j}^{\hat{t}_i}\}))||_F^2\\
&-2\left\langle\mathbf{W}^t\mathbf{y}^t-\frac{1}{N}\mathbf{e}\mathbf{e}^\top\mathbf{y}^t,\tilde{\eta}_y^t(\nabla_{\mathbf{y}}\tilde{L}_p(\{\mathbf{x}_i^{\hat{t}_i}\},\{\mathbf{y}_i^{\hat{t}_i}\},\{\lambda_{i,l}^{\hat{t}_i}\},\{\boldsymbol{\theta}_{i,j}^{\hat{t}_i}\})-\bar{\nabla}_{\mathbf{y}}\tilde{L}_p(\{\mathbf{x}_i^{\hat{t}_i}\},\{\mathbf{y}_i^{\hat{t}_i}\},\{\lambda_{i,l}^{\hat{t}_i}\},\{\boldsymbol{\theta}_{i,j}^{\hat{t}_i}\}))\right\rangle\\
&\le||\mathbf{W}^t\mathbf{y}^t-\frac{1}{N}\mathbf{e}\mathbf{e}^\top\mathbf{y}^t||_F^2+||\tilde{\eta}_y^t(\nabla_{\mathbf{y}}\tilde{L}_p(\{\mathbf{x}_i^{\hat{t}_i}\},\{\mathbf{y}_i^{\hat{t}_i}\},\{\lambda_{i,l}^{\hat{t}_i}\},\{\boldsymbol{\theta}_{i,j}^{\hat{t}_i}\})-\bar{\nabla}_{\mathbf{y}}\tilde{L}_p(\{\mathbf{x}_i^{\hat{t}_i}\},\{\mathbf{y}_i^{\hat{t}_i}\},\{\lambda_{i,l}^{\hat{t}_i}\},\{\boldsymbol{\theta}_{i,j}^{\hat{t}_i}\}))||_F^2\\
&+\delta||\mathbf{W}^t\mathbf{y}^t-\frac{1}{N}\mathbf{e}\mathbf{e}^\top\mathbf{y}^t||_F^2+\frac{1}{\delta}||\tilde{\eta}_y^t(\nabla_{\mathbf{y}}\tilde{L}_p(\{\mathbf{x}_i^{\hat{t}_i}\},\{\mathbf{y}_i^{\hat{t}_i}\},\{\lambda_{i,l}^{\hat{t}_i}\},\{\boldsymbol{\theta}_{i,j}^{\hat{t}_i}\})-\bar{\nabla}_{\mathbf{y}}\tilde{L}_p(\{\mathbf{x}_i^{\hat{t}_i}\},\{\mathbf{y}_i^{\hat{t}_i}\},\{\lambda_{i,l}^{\hat{t}_i}\},\{\boldsymbol{\theta}_{i,j}^{\hat{t}_i}\}))||_F^2\\
&\overset{(b)}{=}(1+\delta)||(\mathbf{W}^t-\frac{1}{N}\mathbf{e}\mathbf{e}^\top)(\mathbf{y}^t-\frac{1}{N}\mathbf{e}\mathbf{e}^\top\mathbf{y}^t)||_F^2\\
&+(1+\frac{1}{\delta})(\tilde{\eta}_y^t)^2||\nabla_{\mathbf{y}}\tilde{L}_p(\{\mathbf{x}_i^{\hat{t}_i}\},\{\mathbf{y}_i^{\hat{t}_i}\},\{\lambda_{i,l}^{\hat{t}_i}\},\{\boldsymbol{\theta}_{i,j}^{\hat{t}_i}\})-\bar{\nabla}_{\mathbf{y}}\tilde{L}_p(\{\mathbf{x}_i^{\hat{t}_i}\},\{\mathbf{y}_i^{\hat{t}_i}\},\{\lambda_{i,l}^{\hat{t}_i}\},\{\boldsymbol{\theta}_{i,j}^{\hat{t}_i}\})||_F^2\\
&\overset{(c)}{\le}\rho||(\mathbf{W}^t-\frac{1}{N}\mathbf{e}\mathbf{e}^\top)(\mathbf{y}^t-\frac{1}{N}\mathbf{e}\mathbf{e}^\top\mathbf{y}^t)||_F^2\\
&+\frac{(\tilde{\eta}_y^t)^2}{1-\rho}||\nabla_{\mathbf{y}}\tilde{L}_p(\{\mathbf{x}_i^{\hat{t}_i}\},\{\mathbf{y}_i^{\hat{t}_i}\},\{\lambda_{i,l}^{\hat{t}_i}\},\{\boldsymbol{\theta}_{i,j}^{\hat{t}_i}\})-\bar{\nabla}_{\mathbf{y}}\tilde{L}_p(\{\mathbf{x}_i^{\hat{t}_i}\},\{\mathbf{y}_i^{\hat{t}_i}\},\{\lambda_{i,l}^{\hat{t}_i}\},\{\boldsymbol{\theta}_{i,j}^{\hat{t}_i}\})||_F^2,
\end{aligned}
\tag{S63}
$$

where (a) utilizes Eq.(S56). (b) uses $(\mathbf{W}^t-\frac{1}{N}\mathbf{e}\mathbf{e}^\top)=(\mathbf{W}^t-\frac{1}{N}\mathbf{e}\mathbf{e}^\top)(\mathbf{I}-\frac{1}{N}\mathbf{e}\mathbf{e}^\top)$. (c) uses $\delta=\frac{1-\rho}{\rho}$ and Eq.(S58).

According to Assumption 1 and Assumption 2, we have

$$
\begin{aligned}
&\mathbb{E}[||\nabla_{\mathbf{y}}L_p(\{\mathbf{x}_i^{\hat{t}_i}\},\{\mathbf{y}_i^{\hat{t}_i}\},\{\lambda_{i,l}^{\hat{t}_i}\},\{\boldsymbol{\theta}_{i,j}^{\hat{t}_i}\})-\bar{\nabla}_{\mathbf{y}}L_p(\{\mathbf{x}_i^{\hat{t}_i}\},\{\mathbf{y}_i^{\hat{t}_i}\},\{\lambda_{i,l}^{\hat{t}_i}\},\{\boldsymbol{\theta}_{i,j}^{\hat{t}_i}\})||_F^2]\\
&\le10L^2(\mathbb{E}[\sum_{i=1}^N||\mathbf{x}_i^t-\mathbf{x}_i^{\hat{t}_i}||^2]+\mathbb{E}[\sum_{i=1}^N||\mathbf{y}_i^t-\mathbf{y}_i^{\hat{t}_i}||^2]+\mathbb{E}[\sum_{i=1}^N||\bar{\mathbf{y}}^t-\mathbf{y}_i^t||^2])\\
&\le20L^2(\mathbb{E}[\sum_{i=1}^N||\mathbf{x}_i^{t+1}-\mathbf{x}_i^t||^2]+\mathbb{E}[\sum_{i=1}^N||\mathbf{y}_i^{t+1}-\mathbf{y}_i^t||^2])+10L^2\mathbb{E}[\sum_{i=1}^N||\bar{\mathbf{y}}^t-\mathbf{y}_i^t||^2].
\end{aligned}
\tag{S64}
$$

Plugging Eq.(S64) into Eq.(S63), then Eq.(S55) can be proved. □

5) Lemma 5:

Proof. Denoting S_1^{t+1}, S_2^{t+1}, and F^{t+1} as,

$$
S_1^{t+1}=\frac{4}{\eta_\lambda^2c_1^{t+1}}\sum_{i=1}^N\sum_{l=1}^{|\mathcal{P}_i^t|}||\lambda_{i,l}^{t+1}-\lambda_{i,l}^t||^2-\frac{4}{\eta_\lambda}(\frac{c_1^{t-1}}{c_1^t}-1)\sum_{i=1}^N\sum_{l=1}^{|\mathcal{P}_i^t|}||\lambda_{i,l}^{t+1}||^2,
\tag{S65}
$$

$$
S_2^{t+1}=\frac{4}{\eta_\theta^2c_2^{t+1}}\sum_{i=1}^N\sum_{j=1}^N||\boldsymbol{\theta}_{i,j}^{t+1}-\boldsymbol{\theta}_{i,j}^t||^2-\frac{4}{\eta_\theta}(\frac{c_2^{t-1}}{c_2^t}-1)\sum_{i=1}^N\sum_{j=1}^N||\boldsymbol{\theta}_{i,j}^{t+1}||^2,
\tag{S66}
$$

$$
\begin{aligned}
F^{t+1}=&L_p(\{\bar{\mathbf{x}}^{t+1}\},\{\bar{\mathbf{y}}^{t+1}\},\{\lambda_{i,l}^{t+1}\},\{\boldsymbol{\theta}_{i,j}^{t+1}\})+S_1^{t+1}+S_2^{t+1}+\gamma_1^t\sum_{i=1}^N||\mathbf{x}_i^{t+1}-\bar{\mathbf{x}}^{t+1}||^2+\gamma_2^t\sum_{i=1}^N||\mathbf{y}_i^{t+1}-\bar{\mathbf{y}}^{t+1}||^2\\
&-\frac{6}{\eta_\lambda}\sum_{i=1}^N\sum_{l=1}^{|\mathcal{P}_i^t|}||\lambda_{i,l}^{t+1}-\lambda_{i,l}^t||^2-\frac{c_1^t}{2}\sum_{i=1}^N\sum_{l=1}^{|\mathcal{P}_i^t|}||\lambda_{i,l}^{t+1}||^2-\frac{6}{\eta_\theta}\sum_{i=1}^N\sum_{j=1}^N||\boldsymbol{\theta}_{i,j}^{t+1}-\boldsymbol{\theta}_{i,j}^t||^2-\frac{c_2^t}{2}\sum_{i=1}^N\sum_{j=1}^N||\boldsymbol{\theta}_{i,j}^{t+1}||^2,
\end{aligned}
\tag{S67}
$$

$\forall t \geq T_1$, we have

$$
\begin{aligned}
&(\frac{N}{2\eta_x} - NL - \frac{3\eta_y N^2 L^2}{2\beta} - \frac{MNL^2\eta_\lambda}{2} - \frac{N^2L^2\eta_{\boldsymbol{\theta}}}{2})\mathbb{E}[\sum_{i=1}^{N}||\mathbf{x}_i^{t+1} - \bar{\mathbf{x}}^t||^2] \\
&+(\frac{N}{2\eta_x} - (4\tau k_1 N^2 L + \frac{8\eta_y N^2 L^2 \tau k_1}{\beta} + \frac{16NML^2}{\eta_\lambda (c_1^t)^2} + \frac{16N^2L^2}{\eta_\theta (c_2^t)^2}) - \frac{40L^2(\eta_x^2\gamma_1^t + \eta_y^2\gamma_2^t)}{1-\rho})\mathbb{E}[\sum_{i=1}^{N}||\mathbf{x}_i^{t+1} - \mathbf{d}_i^t||^2] \\
&+(\gamma_1^{t-1} - \gamma_1^t(\rho + \frac{170\eta_x^2L^2}{1-\rho}) - \frac{N}{2\eta_x} - N^2L - \frac{3\eta_y N^2 L^2}{2\beta} - \frac{MNL^2\eta_\lambda}{2} - \frac{N^2L^2\eta_{\boldsymbol{\theta}}}{2} \\
&\quad - (16\tau k_1 N^2 L + \frac{32\eta_y N^2 L^2 \tau k_1}{\beta} + \frac{64NML^2}{\eta_\lambda (c_1^t)^2} + \frac{64N^2L^2}{\eta_\theta (c_2^t)^2}))\mathbb{E}[\sum_{i=1}^{N}||\bar{\mathbf{x}}^t - \mathbf{x}_i^t||^2] \\
&+(\gamma_2^{t-1} - \gamma_2^t(\rho + \frac{230\eta_y^2L^2}{1-\rho}) - \frac{3\eta_y N^2 L^2}{2\beta} - \frac{MNL^2\eta_\lambda}{2} - \frac{N^2L^2\eta_{\boldsymbol{\theta}}}{2} - \frac{60\eta_x^2L^2\gamma_1^t}{1-\rho} \\
&\quad - (\rho + \frac{2}{1-\rho})(2\tau k_1 N^2 L + \frac{4\eta_y N^2 L^2 \tau k_1}{\beta} + \frac{8NML^2}{\eta_\lambda (c_1^t)^2} + \frac{8N^2L^2}{\eta_\theta (c_2^t)^2}))\mathbb{E}[\sum_{i=1}^{N}||\bar{\mathbf{y}}^t - \mathbf{y}_i^t||^2] \\
&+(\eta_y N - \frac{\eta_y N\beta}{2} - \frac{\eta_y^2 NL}{2} - \frac{MNL^2\eta_y^2\eta_\lambda}{2} - \frac{N^2L^2\eta_y^2\eta_{\boldsymbol{\theta}}}{2} - \frac{40\eta_y^2L^2(\eta_x^2\gamma_1^t + \eta_y^2\gamma_2^t)}{(1-\rho)^2} \\
&\quad - \frac{2\eta_y^2}{1-\rho}(2\tau k_1 N^2 L + \frac{4\eta_y N^2 L^2 \tau k_1}{\beta} + \frac{8NML^2}{\eta_\lambda (c_1^t)^2} + \frac{8N^2L^2}{\eta_\theta (c_2^t)^2}))\mathbb{E}[\sum_{i=1}^{N}||\nabla_{\mathbf{y}_i}\tilde{L}_{pi}(\{\mathbf{x}_i^{\hat{t}_i}\}, \{\mathbf{y}_i^{\hat{t}_i}\}, \{\lambda_{i,l}^{\hat{t}_i}\}, \{\boldsymbol{\theta}_{i,j}^{\hat{t}_i}\})||^2] \\
&+\frac{1}{\eta_\lambda}\mathbb{E}[\sum_{i=1}^{N}\sum_{l=1}^{|\mathcal{P}_i^t|}||\lambda_{i,l}^{t+1} - \lambda_{i,l}^t||^2] + \frac{1}{\eta_\theta}\mathbb{E}[\sum_{i=1}^{N}\sum_{j=1}^{N}||\boldsymbol{\theta}_{i,j}^{t+1} - \boldsymbol{\theta}_{i,j}^t||^2] \\
&\leq \mathbb{E}[F^t - F^{t+1}] + \frac{4}{\eta_\lambda}(\frac{c_1^{t-2}}{c_1^{t-1}} - \frac{c_1^{t-1}}{c_1^t})\mathbb{E}[\sum_{i=1}^{N}\sum_{l=1}^{|\mathcal{P}_i^t|}||\lambda_{i,l}^t||^2] + \frac{4}{\eta_\theta}(\frac{c_2^{t-2}}{c_2^{t-1}} - \frac{c_2^{t-1}}{c_2^t})\mathbb{E}[\sum_{i=1}^{N}\sum_{j=1}^{N}||\boldsymbol{\theta}_{i,j}^t||^2] \\
&+(\frac{c_1^{t-1}}{2} - \frac{c_1^t}{2})\mathbb{E}[\sum_{i=1}^{N}\sum_{l=1}^{|\mathcal{P}_i^t|}||\lambda_{i,l}^{t+1}||^2] + (\frac{c_2^{t-1}}{2} - \frac{c_2^t}{2})\mathbb{E}[\sum_{i=1}^{N}\sum_{j=1}^{N}||\boldsymbol{\theta}_{i,j}^{t+1}||^2].
\end{aligned}
\tag{S68}
$$

Proof: According to the updating rules and take the expectation, $\forall\lambda$, in the $(t+1)^{th}$ iteration, we have

$$\langle \lambda_{i,l}^{t+1} - \lambda_{i,l}^t - \tilde{\eta}_{i,\lambda}^t \nabla_{\lambda_{i,l}}\tilde{L}_{pi}(\{\mathbf{x}_i^{t+1}\}, \{\mathbf{y}_i^{t+1}\}, \{\lambda_{i,l}^t\}, \{\boldsymbol{\theta}_{i,j}^t\}), \lambda_{i,l}^{t+1} - \lambda_{i,l}^t\rangle = 0. \tag{S69}$$

Similar to Eq.(S69), in the t^{th} iteration, we have

$$\langle \lambda_{i,l}^t - \lambda_{i,l}^{t-1} - \tilde{\eta}_{i,\lambda}^{t-1} \nabla_{\lambda_{i,l}}\tilde{L}_{pi}(\{\mathbf{x}_i^t\}, \{\mathbf{y}_i^t\}, \{\lambda_{i,l}^{t-1}\}, \{\boldsymbol{\theta}_{i,j}^{t-1}\}), \lambda_{i,l}^{t+1} - \lambda_{i,l}^t\rangle = 0. \tag{S70}$$

Combining Eq.(S69) and Eq.(S70), it follows that

$$
\begin{aligned}
&\frac{1}{\eta_\lambda}\mathbb{E}[\langle v_{1,i,l}^{t+1}, \lambda_{i,l}^{t+1} - \lambda_{i,l}^t\rangle] \\
=&\frac{1}{\eta_\lambda}\mathbb{E}[\langle \lambda_{i,l}^{t+1} - \lambda_{i,l}^t, \lambda_{i,l}^{t+1} - \lambda_{i,l}^t\rangle] - \frac{1}{\eta_\lambda}\mathbb{E}[\langle \lambda_{i,l}^t - \lambda_{i,l}^{t-1}, \lambda_{i,l}^{t+1} - \lambda_{i,l}^t\rangle] \\
=&\mathbb{E}[\langle \nabla_{\lambda_{i,l}}\tilde{L}_{pi}(\{\mathbf{x}_i^{t+1}\}, \{\mathbf{y}_i^{t+1}\}, \{\lambda_{i,l}^t\}, \{\boldsymbol{\theta}_{i,j}^t\}), \lambda_{i,l}^{t+1} - \lambda_{i,l}^t\rangle] - \mathbb{E}[\langle \nabla_{\lambda_{i,l}}\tilde{L}_{pi}(\{\mathbf{x}_i^t\}, \{\mathbf{y}_i^t\}, \{\lambda_{i,l}^{t-1}\}, \{\boldsymbol{\theta}_{i,j}^{t-1}\}), \lambda_{i,l}^{t+1} - \lambda_{i,l}^t\rangle] \\
=&\mathbb{E}[\langle \nabla_{\lambda_{i,l}}\tilde{L}_{pi}(\{\mathbf{x}_i^{t+1}\}, \{\mathbf{y}_i^{t+1}\}, \{\lambda_{i,l}^t\}, \{\boldsymbol{\theta}_{i,j}^t\}) - \nabla_{\lambda_{i,l}}\tilde{L}_{pi}(\{\mathbf{x}_i^t\}, \{\mathbf{y}_i^t\}, \{\lambda_{i,l}^t\}, \{\boldsymbol{\theta}_{i,j}^t\}), \lambda_{i,l}^{t+1} - \lambda_{i,l}^t\rangle \\
&+\langle \nabla_{\lambda_{i,l}}\tilde{L}_{pi}(\{\mathbf{x}_i^t\}, \{\mathbf{y}_i^t\}, \{\lambda_{i,l}^t\}, \{\boldsymbol{\theta}_{i,j}^t\}) - \nabla_{\lambda_{i,l}}\tilde{L}_{pi}(\{\{\mathbf{x}_i^t\}, \{\mathbf{y}_i^t\}, \{\lambda_{i,l}^{t-1}\}, \{\boldsymbol{\theta}_{i,j}^{t-1}\}), v_{1,i,l}^{t+1}\rangle \\
&+\langle \nabla_{\lambda_{i,l}}\tilde{L}_{pi}(\{\mathbf{x}_i^t\}, \{\mathbf{y}_i^t\}, \{\lambda_{i,l}^t\}, \{\boldsymbol{\theta}_{i,j}^t\}) - \nabla_{\lambda_{i,l}}\tilde{L}_{pi}(\{\{\mathbf{x}_i^t\}, \{\mathbf{y}_i^t\}, \{\lambda_{i,l}^{t-1}\}, \{\boldsymbol{\theta}_{i,j}^{t-1}\}), \lambda_{i,l}^t - \lambda_{i,l}^{t-1}\rangle].
\end{aligned}
\tag{S71}
$$

Since we have

$$\frac{1}{\eta_\lambda}\mathbb{E}[\langle v_{1,i,l}^{t+1}, \lambda_{i,l}^{t+1} - \lambda_{i,l}^t\rangle] = \frac{1}{2\eta_\lambda}\mathbb{E}[||\lambda_{i,l}^{t+1} - \lambda_{i,l}^t||^2] + \frac{1}{2\eta_\lambda}\mathbb{E}[||v_{1,i,l}^{t+1}||^2] - \frac{1}{2\eta_\lambda}\mathbb{E}[||\lambda_{i,l}^t - \lambda_{i,l}^{t-1}||^2], \tag{S72}$$

it follows from Eq. (S71) and Eq. (S72) that,

$$
\begin{aligned}
&\frac{1}{2\eta_\lambda}\mathbb{E}[||\lambda_{i,l}^{t+1}-\lambda_{i,l}^{t}||^2]+\frac{1}{2\eta_\lambda}\mathbb{E}[||v_{1,i,l}^{t+1}||^2]-\frac{1}{2\eta_\lambda}\mathbb{E}[||\lambda_{i,l}^{t}-\lambda_{i,l}^{t-1}||^2]\\
\leq&\frac{L^2}{2b_1^t}(\mathbb{E}[\sum_{i=1}^{N}||\mathbf{x}_i^{t+1}-\mathbf{x}_i^{t}||^2]+\mathbb{E}[\sum_{i=1}^{N}||\mathbf{y}_i^{t+1}-\mathbf{y}_i^{t}||^2])+\frac{b_1^t}{2}\mathbb{E}[||\lambda_{i,l}^{t+1}-\lambda_{i,l}^{t}||^2]\\
&+\frac{c_1^{t-1}-c_1^t}{2}(\mathbb{E}[||\lambda_{i,l}^{t+1}||^2]-\mathbb{E}[||\lambda_{i,l}^{t}||^2])-\frac{c_1^{t-1}-c_1^t}{2}\mathbb{E}[||\lambda_{i,l}^{t+1}-\lambda_{i,l}^{t}||^2]\\
&+\frac{\eta_\lambda}{2}\mathbb{E}[||\nabla_{\lambda_{i,l}}\tilde{L}_{pi}(\{\mathbf{x}_i^t\},\{\mathbf{y}_i^t\},\{\lambda_{i,l}^t\},\{\boldsymbol{\theta}_{i,j}^t\})-\nabla_{\lambda_{i,l}}\tilde{L}_{pi}(\{\mathbf{x}_i^t\},\{\mathbf{y}_i^t\},\{\lambda_{i,l}^{t-1}\},\{\boldsymbol{\theta}_{i,j}^{t-1}\})||^2]+\frac{1}{2\eta_\lambda}\mathbb{E}[||v_{1,i,l}^{t+1}||^2]\\
&-\frac{1}{L_1'+c_1^{t-1}}\mathbb{E}[||\nabla_{\lambda_{i,l}}\tilde{L}_{pi}(\{\mathbf{x}_i^t\},\{\mathbf{y}_i^t\},\{\lambda_{i,l}^t\},\{\boldsymbol{\theta}_{i,j}^t\})-\nabla_{\lambda_{i,l}}\tilde{L}_{pi}(\{\mathbf{x}_i^t\},\{\mathbf{y}_i^t\},\{\lambda_{i,l}^{t-1}\},\{\boldsymbol{\theta}_{i,j}^{t-1}\})||^2]\\
&-\frac{c_1^{t-1}L_1'}{L_1'+c_1^{t-1}}\mathbb{E}[||\lambda_{i,l}^{t}-\lambda_{i,l}^{t-1}||^2],
\end{aligned}
\tag{S73}
$$

where $b_1^t>0$. According to the setting that $c_1^0\leq L_1'$, we have $-\frac{c_1^{t-1}L_1{}'}{L_1{}'+c_1^{t-1}}\leq-\frac{c_1^{t-1}L_1{}'}{2L_1{}'}=-\frac{c_1^{t-1}}{2}\leq-\frac{c_1^t}{2}$. Multiplying both sides of Eq.(S73) by $\frac{8}{\eta_\lambda c_1^t}$, we have

$$
\begin{aligned}
&\frac{4}{\eta_\lambda^2c_1^t}\mathbb{E}[||\lambda_{i,l}^{t+1}-\lambda_{i,l}^{t}||^2]-\frac{4}{\eta_\lambda}(\frac{c_1^{t-1}-c_1^t}{c_1^t})\mathbb{E}[||\lambda_{i,l}^{t+1}||^2]\\
\leq&\frac{4}{\eta_\lambda^2c_1^t}\mathbb{E}[||\lambda_{i,l}^{t}-\lambda_{i,l}^{t-1}||^2]-\frac{4}{\eta_\lambda}(\frac{c_1^{t-1}-c_1^t}{c_1^t})\mathbb{E}[||\lambda_{i,l}^{t}||^2]+\frac{4b_1^t}{\eta_\lambda c_1^t}\mathbb{E}[||\lambda_{i,l}^{t+1}-\lambda_{i,l}^{t}||^2]-\frac{4}{\eta_\lambda}\mathbb{E}[||\lambda_{i,l}^{t}-\lambda_{i,l}^{t-1}||^2]\\
&+\frac{4L^2}{\eta_\lambda c_1^tb_1^t}(\mathbb{E}[\sum_{i=1}^{N}||\mathbf{x}_i^{t+1}-\mathbf{x}_i^{t}||^2]+\mathbb{E}[\sum_{i=1}^{N}||\mathbf{y}_i^{t+1}-\mathbf{y}_i^{t}||^2]).
\end{aligned}
\tag{S74}
$$

Setting $b_1^t=\frac{c_1^t}{2}$, using the definition of S_1^t, we have,

$$
\begin{aligned}
&\mathbb{E}[S_1^{t+1}-S_1^t]\\
\leq&\frac{4}{\eta_\lambda}(\frac{c_1^{t-2}}{c_1^{t-1}}-\frac{c_1^{t-1}}{c_1^t})\mathbb{E}[\sum_{i=1}^{N}\sum_{l=1}^{|\mathcal{P}_i^t|}||\lambda_{i,l}^{t}||^2]+(\frac{2}{\eta_\lambda}+\frac{4}{\eta_\lambda^2}(\frac{1}{c_1^{t+1}}-\frac{1}{c_1^t}))\mathbb{E}[\sum_{i=1}^{N}\sum_{l=1}^{|\mathcal{P}_i^t|}||\lambda_{i,l}^{t+1}-\lambda_{i,l}^{t}||^2]\\
&-\frac{4}{\eta_\lambda}\mathbb{E}[\sum_{i=1}^{N}\sum_{l=1}^{|\mathcal{P}_i^t|}||\lambda_{i,l}^{t}-\lambda_{i,l}^{t-1}||^2]+\frac{8NL^2}{\eta_\lambda(c_1^t)^2}(\mathbb{E}[\sum_{i=1}^{N}\sum_{l=1}^{|\mathcal{P}_i^t|}||\mathbf{x}_i^{t+1}-\mathbf{x}_i^{t}||^2]+\mathbb{E}[\sum_{i=1}^{N}\sum_{l=1}^{|\mathcal{P}_i^t|}||\mathbf{y}_i^{t+1}-\mathbf{y}_i^{t}||^2]).
\end{aligned}
\tag{S75}
$$

Similarly to Eq.(S71), it follows that

$$
\begin{aligned}
&\frac{1}{\eta_\theta}\mathbb{E}[\langle v_{2,i,j}^{t+1},\boldsymbol{\theta}_{i,j}^{t+1}-\boldsymbol{\theta}_{i,j}^{t}\rangle]\\
=&\mathbb{E}[\langle\nabla_{\boldsymbol{\theta}_{i,j}}\tilde{L}_{pi}(\{\mathbf{x}_i^{t+1}\},\{\mathbf{y}_i^{t+1}\},\{\lambda_{i,l}^t\},\{\boldsymbol{\theta}_{i,j}^t\}),\boldsymbol{\theta}_{i,j}^{t+1}-\boldsymbol{\theta}_{i,j}^{t}\rangle]-\mathbb{E}[\langle\nabla_{\boldsymbol{\theta}_{i,j}}\tilde{L}_{pi}(\{\mathbf{x}_i^{t}\},\{\mathbf{y}_i^{t}\},\{\lambda_{i,l}^t\},\{\boldsymbol{\theta}_{i,j}^{t-1}\}),\boldsymbol{\theta}_{i,j}^{t+1}-\boldsymbol{\theta}_{i,j}^{t}\rangle]\\
=&\mathbb{E}[\langle\nabla_{\boldsymbol{\theta}_{i,j}}\tilde{L}_{pi}(\{\mathbf{x}_i^{t+1}\},\{\mathbf{y}_i^{t+1}\},\{\lambda_{i,l}^{t+1}\},\{\boldsymbol{\theta}_{i,j}^t\})-\nabla_{\boldsymbol{\theta}_{i,j}}\tilde{L}_{pi}(\{\mathbf{x}_i^{t}\},\{\mathbf{y}_i^{t}\},\{\lambda_{i,l}^t\},\{\boldsymbol{\theta}_{i,j}^{t}\}),\boldsymbol{\theta}_{i,j}^{t+1}-\boldsymbol{\theta}_{i,j}^{t}\rangle]\\
+&\mathbb{E}[\langle\nabla_{\boldsymbol{\theta}_{i,j}}\tilde{L}_{pi}(\{\mathbf{x}_i^{t}\},\{\mathbf{y}_i^{t}\},\{\lambda_{i,l}^t\},\{\boldsymbol{\theta}_{i,j}^t\})-\nabla_{\boldsymbol{\theta}_{i,j}}\tilde{L}_{pi}(\{\mathbf{x}_i^{t}\},\{\mathbf{y}_i^{t}\},\{\lambda_{i,l}^t\},\{\boldsymbol{\theta}_{i,j}^{t-1}\}),v_{2,i,j}^{t+1}\rangle]\\
+&\mathbb{E}[\langle\nabla_{\boldsymbol{\theta}_{i,j}}\tilde{L}_{pi}(\{\mathbf{x}_i^{t}\},\{\mathbf{y}_i^{t}\},\{\lambda_{i,l}^t\},\{\boldsymbol{\theta}_{i,j}^t\})-\nabla_{\boldsymbol{\theta}_{i,j}}\tilde{L}_{pi}(\{\mathbf{x}_i^{t}\},\{\mathbf{y}_i^{t}\},\{\lambda_{i,l}^t\},\{\boldsymbol{\theta}_{i,j}^{t-1}\}),\boldsymbol{\theta}_{i,j}^{t}-\boldsymbol{\theta}_{i,j}^{t-1}\rangle].
\end{aligned}
\tag{S76}
$$

Since we have

$$
\frac{1}{\eta_\theta}\mathbb{E}[\langle v_{2,i,j}^{t+1},\boldsymbol{\theta}_{i,j}^{t+1}-\boldsymbol{\theta}_{i,j}^{t}\rangle]=\frac{1}{2\eta_\theta}\mathbb{E}[||\boldsymbol{\theta}_{i,j}^{t+1}-\boldsymbol{\theta}_{i,j}^{t}||^2]+\frac{1}{2\eta_\theta}\mathbb{E}[||v_{2,i,j}^{t+1}||^2]-\frac{1}{2\eta_\theta}\mathbb{E}[||\boldsymbol{\theta}_{i,j}^{t}-\boldsymbol{\theta}_{i,j}^{t-1}||^2],
\tag{S77}
$$

it follows from Eq. (S76) and Eq. (S77) that,

$$\begin{aligned}
&\frac{1}{2\eta_\theta}\mathbb{E}[||\boldsymbol{\theta}_{i,j}^{t+1}-\boldsymbol{\theta}_{i,j}^{t}||^2]+\frac{1}{2\eta_\theta}\mathbb{E}[||v_{2,i,j}^{t+1}||^2]-\frac{1}{2\eta_\theta}\mathbb{E}[||\boldsymbol{\theta}_{i,j}^{t}-\boldsymbol{\theta}_{i,j}^{t-1}||^2]\\
\leq&\frac{L^2}{2b_2^t}(\mathbb{E}[\sum_{i=1}^{N}||\mathbf{x}_i^{t+1}-\mathbf{x}_i^t||^2]+\mathbb{E}[\sum_{i=1}^{N}||\mathbf{y}_i^{t+1}-\mathbf{y}_i^t||^2])+\frac{b_2^t}{2}\mathbb{E}[||\boldsymbol{\theta}_{i,j}^{t+1}-\boldsymbol{\theta}_{i,j}^{t}||^2]\\
+&\frac{c_2^{t-1}-c_2^t}{2}(\mathbb{E}[||\boldsymbol{\theta}_{i,j}^{t+1}||^2]-\mathbb{E}[||\boldsymbol{\theta}_{i,j}^{t}||^2])-\frac{c_2^{t-1}-c_2^t}{2}\mathbb{E}[||\boldsymbol{\theta}_{i,j}^{t+1}-\boldsymbol{\theta}_{i,j}^{t}||^2]\\
+&\frac{\eta_\theta}{2}\mathbb{E}[||\nabla_{\boldsymbol{\theta}_{i,j}}\tilde{L}_{pi}(\{\mathbf{x}_i^t\},\{\mathbf{y}_i^t\},\{\lambda_{i,l}^t\},\{\boldsymbol{\theta}_{i,j}^t\})-\nabla_{\boldsymbol{\theta}_{i,j}}\tilde{L}_{pi}(\{\mathbf{x}_i^t\},\{\mathbf{y}_i^t\},\{\lambda_{i,l}^t\},\{\boldsymbol{\theta}_{i,j}^{t-1}\})||^2]+\frac{1}{2\eta_\theta}\mathbb{E}[||v_{2,i,j}^{t+1}||^2]\\
-&\frac{1}{L_2'+c_2^{t-1}}\mathbb{E}[||\nabla_{\boldsymbol{\theta}_{i,j}}\tilde{L}_{pi}(\{\mathbf{x}_i^t\},\{\mathbf{y}_i^t\},\{\lambda_{i,l}^t\},\{\boldsymbol{\theta}_{i,j}^t\})-\nabla_{\boldsymbol{\theta}_{i,j}}\tilde{L}_{pi}(\{\mathbf{x}_i^t\},\{\mathbf{y}_i^t\},\{\lambda_{i,l}^t\},\{\boldsymbol{\theta}_{i,j}^{t-1}\})||^2]\\
-&\frac{c_2^{t-1}L_2'}{L_2'+c_2^{t-1}}\mathbb{E}[||\boldsymbol{\theta}_{i,j}^{t}-\boldsymbol{\theta}_{i,j}^{t-1}||^2],
\end{aligned}\tag{S78}$$

where $b_2^t>0$. According to the setting that $c_2^0\leq L_2'$, we have $-\frac{c_2^{t-1}L_2{}'}{L_2{}'+c_2^{t-1}}\leq-\frac{c_2^{t-1}L_2{}'}{2L_2{}'}=-\frac{c_2^{t-1}}{2}\leq-\frac{c_2^t}{2}$. Multiplying both sides of Eq.(S78) by $\frac{8}{\eta_\theta c_2^t}$, we have

$$\begin{aligned}
&\frac{4}{\eta_\theta^2 c_2^t}\mathbb{E}[||\boldsymbol{\theta}_{i,j}^{t+1}-\boldsymbol{\theta}_{i,j}^{t}||^2]-\frac{4}{\eta_\theta}(\frac{c_2^{t-1}-c_2^t}{c_2^t})\mathbb{E}[||\boldsymbol{\theta}_{i,j}^{t+1}||^2]\\
\leq&\frac{4}{\eta_\theta^2 c_2^t}\mathbb{E}[||\boldsymbol{\theta}_{i,j}^{t}-\boldsymbol{\theta}_{i,j}^{t-1}||^2]-\frac{4}{\eta_\theta}(\frac{c_2^{t-1}-c_2^t}{c_2^t})\mathbb{E}[||\boldsymbol{\theta}_{i,j}^{t}||^2]+\frac{4b_2^t}{\eta_\theta c_2^t}\mathbb{E}[||\boldsymbol{\theta}_{i,j}^{t+1}-\boldsymbol{\theta}_{i,j}^{t}||^2]-\frac{4}{\eta_\theta}\mathbb{E}[||\boldsymbol{\theta}_{i,j}^{t}-\boldsymbol{\theta}_{i,j}^{t-1}||^2]\\
+&\frac{4L^2}{\eta_\theta c_2^t b_2^t}(\mathbb{E}[\sum_{i=1}^{N}||\mathbf{x}_i^{t+1}-\mathbf{x}_i^t||^2]+\mathbb{E}[\sum_{i=1}^{N}||\mathbf{y}_i^{t+1}-\mathbf{y}_i^t||^2]).
\end{aligned}\tag{S79}$$

Setting $b_2^t=\frac{c_2^t}{2}$, using the definition of S_2^t, we have,

$$\begin{aligned}
&\mathbb{E}[S_2^{t+1}-S_2^t]\\
\leq&\frac{4}{\eta_\theta}(\frac{c_2^{t-2}}{c_2^{t-1}}-\frac{c_2^{t-1}}{c_2^t})\mathbb{E}[\sum_{i=1}^{N}\sum_{j=1}^{N}||\boldsymbol{\theta}_{i,j}^{t}||^2]+(\frac{2}{\eta_\theta}+\frac{4}{\eta_\theta^2}(\frac{1}{c_2^{t+1}}-\frac{1}{c_2^t}))\mathbb{E}[\sum_{i=1}^{N}\sum_{j=1}^{N}||\boldsymbol{\theta}_{i,j}^{t+1}-\boldsymbol{\theta}_{i,j}^{t}||^2]\\
-&\frac{4}{\eta_\theta}\mathbb{E}[\sum_{i=1}^{N}\sum_{j=1}^{N}||\boldsymbol{\theta}_{i,j}^{t}-\boldsymbol{\theta}_{i,j}^{t-1}||^2]+\frac{8NL^2}{\eta_\theta(c_2^t)^2}(\mathbb{E}[\sum_{i=1}^{N}\sum_{j=1}^{N}||\mathbf{x}_i^{t+1}-\mathbf{x}_i^t||^2]+\mathbb{E}[\sum_{i=1}^{N}\sum_{j=1}^{N}||\mathbf{y}_i^{t+1}-\mathbf{y}_i^t||^2]).
\end{aligned}\tag{S80}$$

Based on the setting of c_1^t and c_2^t, we can obtain that $\frac{\eta_\lambda}{2}\geq\frac{1}{c_1^{t+1}}-\frac{1}{c_1^t}$, $\frac{\eta_\theta}{2}\geq\frac{1}{c_2^{t+1}}-\frac{1}{c_2^t}$. Add $\gamma_1^t\sum_{i=1}^{N}||\mathbf{x}_i^{t+1}-\bar{\mathbf{x}}^{t+1}||^2$ and $\gamma_2^t\sum_{i=1}^{N}||\mathbf{y}_i^{t+1}-\bar{\mathbf{y}}^{t+1}||^2$ to both sides of Eq.(S23), subtract $\gamma_1^{t-1}\sum_{i=1}^{N}||\mathbf{x}_i^{t}-\bar{\mathbf{x}}^{t}||^2$ and $\gamma_2^{t-1}\sum_{i=1}^{N}||\mathbf{y}_i^{t}-\bar{\mathbf{y}}^{t}||^2$ to both sides of Eq.(S23). Then using the results from Eq.(S59), Eq.(S60), Eq.(S75) and Eq.(S80), Lemma 5 (Eq.(S68)) can be proved. □

6) Proof of Theorem 1:

Proof. According to Definition 1, for $i=1,\cdots,N$, we have

$$\begin{aligned}
&\mathbb{E}[\sum_{i=1}^{N}||P(\mathbf{d}_i^t,\nabla_{\mathbf{x}_i}\tilde{L}_{pi}(\{\mathbf{x}_i^{\hat{t}_i}\},\{\mathbf{y}_i^{\hat{t}_i}\},\{\lambda_{i,l}^{\hat{t}_i}\},\{\boldsymbol{\theta}_{i,j}^{\hat{t}_i}\}),\tilde{\eta}_{i,x}^t)||^2]\\
=&\mathbb{E}[\sum_{i=1}^{N}||\frac{1}{\tilde{\eta}_{i,x}^t}(\mathbf{d}_i^t-\mathrm{prox}_\eta^R(\mathbf{d}_i^t-\tilde{\eta}_{i,x}^t\nabla_{\mathbf{x}_i}\tilde{L}_{pi}(\{\mathbf{x}_i^{\hat{t}_i}\},\{\mathbf{y}_i^{\hat{t}_i}\},\{\lambda_{i,l}^{\hat{t}_i}\},\{\boldsymbol{\theta}_{i,j}^{\hat{t}_i}\})))||^2]\\
=&\frac{2}{\eta_x^2}\mathbb{E}[\sum_{i=1}^{N}||\mathbf{d}_i^t-\mathbf{x}_i^{t+1}||^2].
\end{aligned}\tag{S81}$$

According to the property of the proximal operator, we further have

$$
\begin{aligned}
&\mathbb{E}[\sum_{i=1}^{N}||P(\mathbf{d}_i^t, \nabla_{\mathbf{x}_i}\tilde{L}_{pi}(\{\mathbf{x}_i^{\hat{t}_i}\},\{\mathbf{y}_i^{\hat{t}_i}\},\{\lambda_{i,l}^{\hat{t}_i}\},\{\boldsymbol{\theta}_{i,j}^{\hat{t}_i}\}),\tilde{\eta}_{i,x}^t) - P(\mathbf{d}_i^t, \bar{\nabla}_{\mathbf{d}}L'_{pi}(\{\mathbf{d}_i^t\},\{\mathbf{y}_i^t\},\{\lambda_{i,l}^t\},\{\boldsymbol{\theta}_{i,j}^t\}),\tilde{\eta}_{i,x}^t)||^2]\\
=&\frac{1}{\eta_x^2}\mathbb{E}[\sum_{i=1}^{N}||\mathrm{prox}_{\eta}^{R}(\mathbf{d}_i^t-\eta_x\nabla_{\mathbf{x}_i}\tilde{L}_{pi}(\{\mathbf{x}_i^{\hat{t}_i}\},\{\mathbf{y}_i^{\hat{t}_i}\},\{\lambda_{i,l}^{\hat{t}_i}\},\{\boldsymbol{\theta}_{i,j}^{\hat{t}_i}\}) - \mathrm{prox}_{\eta}^{R}(\mathbf{d}_i^t-\eta_x\bar{\nabla}_{\mathbf{d}}L'_{pi}(\{\mathbf{d}_i^t\},\{\mathbf{y}_i^t\},\{\lambda_{i,l}^t\},\{\boldsymbol{\theta}_{i,j}^t\}))||^2]\\
\leq&\mathbb{E}[\sum_{i=1}^{N}||\nabla_{\mathbf{x}_i}L'_{pi}(\{\mathbf{x}_i^{\hat{t}_i}\},\{\mathbf{y}_i^{\hat{t}_i}\},\{\lambda_{i,l}^{\hat{t}_i}\},\{\boldsymbol{\theta}_{i,j}^{\hat{t}_i}\}) - \bar{\nabla}_{\mathbf{d}}L'_{pi}(\{\mathbf{d}_i^t\},\{\mathbf{y}_i^t\},\{\lambda_{i,l}^t\},\{\boldsymbol{\theta}_{i,j}^t\})||^2].
\end{aligned}
\tag{S82}
$$

According to Young's inequality and Eq.(S81), we have

$$
\begin{aligned}
&\frac{1}{2}\mathbb{E}[\sum_{i=1}^{N}||P(\mathbf{d}_i^t, \bar{\nabla}_{\mathbf{d}}L'_{pi}(\{\mathbf{d}_i^t\},\{\mathbf{y}_i^t\},\{\lambda_{i,l}^t\},\{\boldsymbol{\theta}_{i,j}^t\}),\tilde{\eta}_{i,x}^t)||^2]\\
\leq&\frac{1}{\eta_x^2}\mathbb{E}[\sum_{i=1}^{N}||\mathbf{x}_i^{t+1}-\mathbf{d}_i^t||^2] + \mathbb{E}[\sum_{i=1}^{N}||P(\mathbf{d}_i^t, \nabla_{\mathbf{x}_i}\tilde{L}_{pi}(\{\mathbf{x}_i^{\hat{t}_i}\},\{\mathbf{y}_i^{\hat{t}_i}\},\{\lambda_{i,l}^{\hat{t}_i}\},\{\boldsymbol{\theta}_{i,j}^{\hat{t}_i}\}),\tilde{\eta}_{i,x}^t)\\
&-P(\mathbf{d}_i^t, \bar{\nabla}_{\mathbf{d}}L'_{pi}(\{\mathbf{d}_i^t\},\{\mathbf{y}_i^t\},\{\lambda_{i,l}^t\},\{\boldsymbol{\theta}_{i,j}^t\}),\tilde{\eta}_{i,x}^t)||^2].
\end{aligned}
\tag{S83}
$$

Plugging Eq.(S82) into Eq.(S83) yields

$$
\begin{aligned}
&\frac{1}{2}\mathbb{E}[\sum_{i=1}^{N}||P(\mathbf{d}_i^t, \bar{\nabla}_{\mathbf{d}}L'_{pi}(\{\mathbf{d}_i^t\},\{\mathbf{y}_i^t\},\{\lambda_{i,l}^t\},\{\boldsymbol{\theta}_{i,j}^t\}),\tilde{\eta}_{i,x}^t)||^2]\\
\leq&\frac{1}{\eta_x^2}\mathbb{E}[\sum_{i=1}^{N}||\mathbf{x}_i^{t+1}-\mathbf{d}_i^t||^2]\\
&+\mathbb{E}[\sum_{i=1}^{N}||\nabla_{\mathbf{x}_i}\tilde{L}_{pi}(\{\mathbf{x}_i^{\hat{t}_i}\},\{\mathbf{y}_i^{\hat{t}_i}\},\{\lambda_{i,l}^{\hat{t}_i}\},\{\boldsymbol{\theta}_{i,j}^{\hat{t}_i}\}) - \bar{\nabla}_{\mathbf{d}}L'_{pi}(\{\mathbf{d}_i^t\},\{\mathbf{y}_i^t\},\{\lambda_{i,l}^t\},\{\boldsymbol{\theta}_{i,j}^t\})||^2].
\end{aligned}
\tag{S84}
$$

Next, combining Assumption 1, Eq.(S59), Eq.(S59), and Eq.(S18), we bound

$$
\begin{aligned}
&\mathbb{E}[\sum_{i=1}^{N}||\nabla_{\mathbf{x}_i}\tilde{L}_{pi}(\{\mathbf{x}_i^{\hat{t}_i}\},\{\mathbf{y}_i^{\hat{t}_i}\},\{\lambda_{i,l}^{\hat{t}_i}\},\{\boldsymbol{\theta}_{i,j}^{\hat{t}_i}\}) - \bar{\nabla}_{\mathbf{d}}L'_{pi}(\{\mathbf{d}_i^t\},\{\mathbf{y}_i^t\},\{\lambda_{i,l}^t\},\{\boldsymbol{\theta}_{i,j}^t\})||^2]\\
=&\mathbb{E}[\sum_{i=1}^{N}||\nabla_{\mathbf{x}_i}\tilde{L}_{pi}(\{\mathbf{x}_i^{\hat{t}_i}\},\{\mathbf{y}_i^{\hat{t}_i}\},\{\lambda_{i,l}^{\hat{t}_i}\},\{\boldsymbol{\theta}_{i,j}^{\hat{t}_i}\}) - \nabla_{\mathbf{x}_i}\tilde{L}_{pi}(\{\mathbf{x}_i^t\},\{\mathbf{y}_i^t\},\{\lambda_{i,l}^t\},\{\boldsymbol{\theta}_{i,j}^t\})\\
&+\nabla_{\mathbf{x}_i}\tilde{L}_{pi}(\{\mathbf{x}_i^t\},\{\mathbf{y}_i^t\},\{\lambda_{i,l}^t\},\{\boldsymbol{\theta}_{i,j}^t\}) - \nabla_{\mathbf{x}_i}\tilde{L}_{pi}(\{\bar{\mathbf{x}}^t\},\{\mathbf{y}_i^t\},\{\lambda_{i,l}^t\},\{\boldsymbol{\theta}_{i,j}^t\})\\
&+\nabla_{\mathbf{x}_i}\tilde{L}_{pi}(\{\bar{\mathbf{x}}^t\},\{\mathbf{y}_i^t\},\{\lambda_{i,l}^t\},\{\boldsymbol{\theta}_{i,j}^t\}) - \bar{\nabla}_{\mathbf{x}_i}\tilde{L}_{pi}(\{\bar{\mathbf{x}}^t\},\{\mathbf{y}_i^t\},\{\lambda_{i,l}^t\},\{\boldsymbol{\theta}_{i,j}^t\})\\
&+\bar{\nabla}_{\mathbf{x}_i}\tilde{L}_{pi}(\{\bar{\mathbf{x}}^t\},\{\mathbf{y}_i^t\},\{\lambda_{i,l}^t\},\{\boldsymbol{\theta}_{i,j}^t\}) - \bar{\nabla}_{\mathbf{d}}L'_{pi}(\{\mathbf{d}_i^t\},\{\mathbf{y}_i^t\},\{\lambda_{i,l}^t\},\{\boldsymbol{\theta}_{i,j}^t\})||^2]\\
\leq&4NL^2(\mathbb{E}[\sum_{i=1}^{N}||\mathbf{x}_i^t-\mathbf{x}_i^{\hat{t}_i}||^2] + \mathbb{E}[\sum_{i=1}^{N}||\mathbf{y}_i^t-\mathbf{y}_i^{\hat{t}_i}||^2] + \mathbb{E}[\sum_{i=1}^{N}||\mathbf{x}_i^t-\bar{\mathbf{x}}^t||^2] + \mathbb{E}[\sum_{i=1}^{N}||\mathbf{d}_i^t-\bar{\mathbf{x}}^t||^2])\\
\leq&8NL^2\tau k_1(\mathbb{E}[\sum_{i=1}^{N}||\mathbf{x}_i^{t+1}-\mathbf{x}_i^t||^2] + \mathbb{E}[\sum_{i=1}^{N}||\mathbf{y}_i^{t+1}-\mathbf{y}_i^t||^2]) + 8NL^2\mathbb{E}[\sum_{i=1}^{N}||\mathbf{x}_i^t-\bar{\mathbf{x}}^t||^2]\\
\leq&16NL^2\tau k_1\mathbb{E}[\sum_{i=1}^{N}||\mathbf{x}_i^{t+1}-\mathbf{d}_i^t||^2] + (64NL^2\tau k_1+8NL^2)\mathbb{E}[\sum_{i=1}^{N}||\mathbf{x}_i^t-\bar{\mathbf{x}}^t||^2] + \frac{24NL^2\tau k_1}{1-\rho}\mathbb{E}[\sum_{i=1}^{N}||\mathbf{y}_i^t-\bar{\mathbf{y}}^t||^2]\\
&+\frac{16\eta_y^2NL^2\tau k_1}{1-\rho}\mathbb{E}[\sum_{i=1}^{N}||\nabla_{\mathbf{y}_i}\tilde{L}_{pi}(\{\mathbf{x}_i^{\hat{t}_i}\},\{\mathbf{y}_i^{\hat{t}_i}\},\{\lambda_{i,l}^{\hat{t}_i}\},\{\boldsymbol{\theta}_{i,j}^{\hat{t}_i}\})||^2].
\end{aligned}
\tag{S85}
$$

Adding $\frac{L^2}{2}\mathbb{E}[\sum_{i=1}^{N}||\mathbf{x}_i^t-\bar{\mathbf{x}}^t||^2]$ to both sides of Eq.(S84), according to Eq.(S85) and Assumption 1, results in

$$\begin{aligned}
&\frac{1}{2}\mathbb{E}[\sum_{i=1}^{N}||P(\mathbf{d}_i^t,\bar{\nabla}_{\mathbf{d}}L'_{pi}(\{\mathbf{d}_i^t\},\{\mathbf{y}_i^t\},\{\lambda_{i,l}^t\},\{\boldsymbol{\theta}_{i,j}^t\}),\eta_{i,x})||^2]+\frac{L^2}{2}\mathbb{E}[\sum_{i=1}^{N}||\mathbf{x}_i^t-\bar{\mathbf{x}}^t||^2]\\
\le&\frac{1}{\eta_x^2}\mathbb{E}[\sum_{i=1}^{N}||\mathbf{x}_i^{t+1}-\mathbf{d}_i^t||^2]+16NL^2\tau k_1\mathbb{E}[\sum_{i=1}^{N}||\mathbf{x}_i^{t+1}-\mathbf{d}_i^t||^2]+(64NL^2\tau k_1+16NL^2)\mathbb{E}[\sum_{i=1}^{N}||\mathbf{x}_i^t-\bar{\mathbf{x}}^t||^2]\\
+&\frac{24NL^2\tau k_1}{1-\rho}\mathbb{E}[\sum_{i=1}^{N}||\mathbf{y}_i^t-\bar{\mathbf{y}}^t||^2]+\frac{16\eta_y^2NL^2\tau k_1}{1-\rho}\mathbb{E}[\sum_{i=1}^{N}||\nabla_{\mathbf{y}_i}\tilde{L}_{pi}(\{\mathbf{x}_i^{\hat{t}_i}\},\{\mathbf{y}_i^{\hat{t}_i}\},\{\lambda_{i,l}^{\hat{t}_i}\},\{\boldsymbol{\theta}_{i,j}^{\hat{t}_i}\})||^2].
\end{aligned}\tag{S86}$$

According to Eq.(S86), we have

$$\begin{aligned}
&\mathbb{E}[\sum_{i=1}^{N}||P(\mathbf{d}_i^t,\bar{\nabla}_{\mathbf{d}}L'_{pi}(\{\mathbf{d}_i^t\},\{\mathbf{y}_i^t\},\{\lambda_{i,l}^t\},\{\boldsymbol{\theta}_{i,j}^t\}),\eta_{i,x})||^2]+L^2\mathbb{E}[\sum_{i=1}^{N}||\mathbf{x}_i^t-\bar{\mathbf{x}}^t||^2]\\
\le&(\frac{2}{\eta_x^2}+32NL^2\tau k_1)\mathbb{E}[\sum_{i=1}^{N}||\mathbf{x}_i^{t+1}-\mathbf{d}_i^t||^2]+(128NL^2\tau k_1+32NL^2)\mathbb{E}[\sum_{i=1}^{N}||\mathbf{x}_i^t-\bar{\mathbf{x}}^t||^2]\\
+&\frac{48NL^2\tau k_1}{1-\rho}\mathbb{E}[\sum_{i=1}^{N}||\mathbf{y}_i^t-\bar{\mathbf{y}}^t||^2]+\frac{32\eta_y^2NL^2\tau k_1}{1-\rho}\mathbb{E}[\sum_{i=1}^{N}||\nabla_{\mathbf{y}_i}\tilde{L}_{pi}(\{\mathbf{x}_i^{\hat{t}_i}\},\{\mathbf{y}_i^{\hat{t}_i}\},\{\lambda_{i,l}^{\hat{t}_i}\},\{\boldsymbol{\theta}_{i,j}^{\hat{t}_i}\})||^2].
\end{aligned}\tag{S87}$$

For $\mathbf{y}$ variables, according to Young's inequality, we have

$$\begin{aligned}
&\frac{1}{2}\mathbb{E}[\sum_{i=1}^{N}||\bar{\nabla}_{\mathbf{u}}L'_{pi}(\{\mathbf{x}_i^t\},\{\mathbf{u}_i^t\},\{\lambda_{i,l}^t\},\{\boldsymbol{\theta}_{i,j}^t\})||^2]\\
\le&\mathbb{E}[\sum_{i=1}^{N}||\nabla_{\mathbf{y}_i}\tilde{L}_{pi}(\{\mathbf{x}_i^{\hat{t}_i}\},\{\mathbf{y}_i^{\hat{t}_i}\},\{\lambda_{i,l}^{\hat{t}_i}\},\{\boldsymbol{\theta}_{i,j}^{\hat{t}_i}\})||^2]\\
+&\mathbb{E}[\sum_{i=1}^{N}||\nabla_{\mathbf{y}_i}\tilde{L}_{pi}(\{\mathbf{x}_i^{\hat{t}_i}\},\{\mathbf{y}_i^{\hat{t}_i}\},\{\lambda_{i,l}^{\hat{t}_i}\},\{\boldsymbol{\theta}_{i,j}^{\hat{t}_i}\})-\bar{\nabla}_{\mathbf{u}}L'_{pi}(\{\mathbf{x}_i^t\},\{\mathbf{u}_i^t\},\{\lambda_{i,l}^t\},\{\boldsymbol{\theta}_{i,j}^t\})||^2].
\end{aligned}\tag{S88}$$

Next, we bound

$$\begin{aligned}
&\mathbb{E}[\sum_{i=1}^{N}||\nabla_{\mathbf{y}_i}\tilde{L}_{pi}(\{\mathbf{x}_i^{\hat{t}_i}\},\{\mathbf{y}_i^{\hat{t}_i}\},\{\lambda_{i,l}^{\hat{t}_i}\},\{\boldsymbol{\theta}_{i,j}^{\hat{t}_i}\})-\bar{\nabla}_{\mathbf{u}}L'_{pi}(\{\mathbf{x}_i^t\},\{\mathbf{u}_i^t\},\{\lambda_{i,l}^t\},\{\boldsymbol{\theta}_{i,j}^t\})||^2]\\
=&\mathbb{E}[\sum_{i=1}^{N}||\nabla_{\mathbf{y}_i}\tilde{L}_{pi}(\{\mathbf{x}_i^{\hat{t}_i}\},\{\mathbf{y}_i^{\hat{t}_i}\},\{\lambda_{i,l}^{\hat{t}_i}\},\{\boldsymbol{\theta}_{i,j}^{\hat{t}_i}\})-\nabla_{\mathbf{y}_i}\tilde{L}_{pi}(\{\mathbf{x}_i^t\},\{\mathbf{y}_i^t\},\{\lambda_{i,l}^t\},\{\boldsymbol{\theta}_{i,j}^t\})\\
+&\nabla_{\mathbf{y}_i}\tilde{L}_{pi}(\{\mathbf{x}_i^t\},\{\mathbf{y}_i^t\},\{\lambda_{i,l}^t\},\{\boldsymbol{\theta}_{i,j}^t\})-\nabla_{\mathbf{y}_i}\tilde{L}_{pi}(\{\mathbf{x}_i^t\},\{\bar{\mathbf{y}}^t\},\{\lambda_{i,l}^t\},\{\boldsymbol{\theta}_{i,j}^t\})\\
+&\nabla_{\mathbf{y}_i}\tilde{L}_{pi}(\{\mathbf{x}_i^t\},\{\bar{\mathbf{y}}^t\},\{\lambda_{i,l}^t\},\{\boldsymbol{\theta}_{i,j}^t\})-\bar{\nabla}_{\mathbf{y}_i}\tilde{L}_{pi}(\{\mathbf{x}_i^t\},\{\bar{\mathbf{y}}^t\},\{\lambda_{i,l}^t\},\{\boldsymbol{\theta}_{i,j}^t\})\\
+&\bar{\nabla}_{\mathbf{y}_i}\tilde{L}_{pi}(\{\mathbf{x}_i^t\},\{\bar{\mathbf{y}}^t\},\{\lambda_{i,l}^t\},\{\boldsymbol{\theta}_{i,j}^t\})-\bar{\nabla}_{\mathbf{u}}L'_{pi}(\{\mathbf{x}_i^t\},\{\mathbf{u}_i^t\},\{\lambda_{i,l}^t\},\{\boldsymbol{\theta}_{i,j}^t\})||^2]\\
\le&4NL^2(\mathbb{E}[\sum_{i=1}^{N}||\mathbf{x}_i^t-\mathbf{x}_i^{\hat{t}_i}||^2]+\mathbb{E}[\sum_{i=1}^{N}||\mathbf{y}_i^t-\mathbf{y}_i^{\hat{t}_i}||^2]+\mathbb{E}[\sum_{i=1}^{N}||\mathbf{y}_i^t-\bar{\mathbf{y}}^t||^2]+\mathbb{E}[\sum_{i=1}^{N}||\mathbf{u}_i^t-\bar{\mathbf{y}}^t||^2])\\
\le&8NL^2\tau k_1(\mathbb{E}[\sum_{i=1}^{N}||\mathbf{x}_i^{t+1}-\mathbf{x}_i^t||^2]+\mathbb{E}[\sum_{i=1}^{N}||\mathbf{y}_i^{t+1}-\mathbf{y}_i^t||^2])+8NL^2\mathbb{E}[\sum_{i=1}^{N}||\mathbf{y}_i^t-\bar{\mathbf{y}}^t||^2]\\
\le&16NL^2\tau k_1\mathbb{E}[\sum_{i=1}^{N}||\mathbf{x}_i^{t+1}-\mathbf{d}_i^t||^2]+64NL^2\tau k_1\mathbb{E}[\sum_{i=1}^{N}||\mathbf{x}_i^t-\bar{\mathbf{x}}^t||^2]+(\frac{24NL^2\tau k_1}{1-\rho}+8NL^2)\mathbb{E}[\sum_{i=1}^{N}||\mathbf{y}_i^t-\bar{\mathbf{y}}^t||^2]\\
+&\frac{16\eta_y^2NL^2\tau k_1}{1-\rho}\mathbb{E}[\sum_{i=1}^{N}||\nabla_{\mathbf{y}_i}\tilde{L}_{pi}(\{\mathbf{x}_i^{\hat{t}_i}\},\{\mathbf{y}_i^{\hat{t}_i}\},\{\lambda_{i,l}^{\hat{t}_i}\},\{\boldsymbol{\theta}_{i,j}^{\hat{t}_i}\})||^2].
\end{aligned}\tag{S89}$$

Adding $\frac{L^2}{2}\sum_{i=1}^N \mathbb{E}[||\mathbf{y}_i^t-\bar{\mathbf{y}}^t||^2]$ to both sides of Eq.(S88), according to Eq.(S89) and Eq.(S19), it follows that

$$\begin{aligned}
&\mathbb{E}[\sum_{i=1}^N ||\bar{\nabla}_{\mathbf{u}} L'_{pi}(\{\mathbf{x}_i^t\},\{\mathbf{u}_i^t\},\{\lambda_{i,l}^t\},\{\boldsymbol{\theta}_{i,j}^t\})||^2]+L^2\mathbb{E}[\sum_{i=1}^N ||\mathbf{y}_i^t-\bar{\mathbf{y}}^t||^2]\\
\le & 32NL^2\tau k_1\mathbb{E}[\sum_{i=1}^N ||\mathbf{x}_i^{t+1}-\mathbf{d}_i^t||^2]+128NL^2\tau k_1\mathbb{E}[\sum_{i=1}^N ||\mathbf{x}_i^t-\bar{\mathbf{x}}^t||^2]+(\frac{48NL^2\tau k_1}{1-\rho}+32NL^2)\mathbb{E}[\sum_{i=1}^N ||\mathbf{y}_i^t-\bar{\mathbf{y}}^t||^2]\\
&+(\frac{32\eta_y^2NL^2\tau k_1}{1-\rho}+2)\mathbb{E}[\sum_{i=1}^N ||\nabla_{\mathbf{y}_i}\tilde{L}_{pi}(\{\mathbf{x}_i^{\hat{t}_i}\},\{\mathbf{y}_i^{\hat{t}_i}\},\{\lambda_{i,l}^{\hat{t}_i}\},\{\boldsymbol{\theta}_{i,j}^{\hat{t}_i}\})||^2].
\end{aligned} \tag{S90}$$

Using the definition $(\tilde{\mathcal{G}}^t)_{\lambda_{i,l}}$, the update rules of $\lambda_{i,l}$, trigonometric inequality, and Cauchy-Schwarz inequality, we have

$$\begin{aligned}
\mathbb{E}[||(\tilde{\mathcal{G}}^t)_{\lambda_{i,l}}||^2]\le & 3L^2(\mathbb{E}[\sum_{i=1}^N ||\mathbf{x}_i^{t+1}-\mathbf{x}_i^t||^2]+\mathbb{E}[\sum_{i=1}^N ||\mathbf{y}_i^{t+1}-\mathbf{y}_i^t||^2])+3((c_1^{t-1})^2-(c_1^t)^2)\mathbb{E}[||\lambda_{i,l}^t||^2]+\frac{3}{\eta_\lambda^2}\mathbb{E}[||\lambda_{i,l}^{t+1}-\lambda_{i,l}^t||^2]\\
\le & 6L^2\mathbb{E}[\sum_{i=1}^N ||\mathbf{x}_i^{t+1}-\mathbf{d}_i^t||^2]+24L^2\mathbb{E}[\sum_{i=1}^N ||\mathbf{x}_i^t-\bar{\mathbf{x}}^t||^2]+\frac{9L^2}{1-\rho}\mathbb{E}[\sum_{i=1}^N ||\mathbf{y}_i^t-\bar{\mathbf{y}}^t||^2]\\
&+\frac{6L^2\eta_y^2}{1-\rho}\mathbb{E}[\sum_{i=1}^N ||\nabla_{\mathbf{y}_i}\tilde{L}_{pi}(\{\mathbf{x}_i^{\hat{t}_i}\},\{\mathbf{y}_i^{\hat{t}_i}\},\{\lambda_{i,l}^{\hat{t}_i}\},\{\boldsymbol{\theta}_{i,j}^{\hat{t}_i}\})||^2]+3((c_1^{t-1})^2-(c_1^t)^2)\mathbb{E}[||\lambda_{i,l}^t||^2]+\frac{3}{\eta_\lambda^2}\mathbb{E}[||\lambda_{i,l}^{t+1}-\lambda_{i,l}^t||^2].
\end{aligned} \tag{S91}$$

Using the definition $(\tilde{\mathcal{G}}^t)_{\boldsymbol{\theta}_{i,j}}$, the update rules of $\boldsymbol{\theta}_{i,j}$, trigonometric inequality, and Cauchy-Schwarz inequality, we have

$$\begin{aligned}
\mathbb{E}[||(\tilde{\mathcal{G}}^t)_{\boldsymbol{\theta}_{i,j}}||^2]\le & 3L^2(\mathbb{E}[\sum_{i=1}^N ||\mathbf{x}_i^{t+1}-\mathbf{x}_i^t||^2]+\mathbb{E}[\sum_{i=1}^N ||\mathbf{y}_i^{t+1}-\mathbf{y}_i^t||^2])+3((c_2^{t-1})^2-(c_2^t)^2)\mathbb{E}[||\boldsymbol{\theta}_{i,j}^t||^2]+\frac{3}{\eta_\theta^2}\mathbb{E}[||\boldsymbol{\theta}_{i,j}^{t+1}-\boldsymbol{\theta}_{i,j}^t||^2]\\
\le & 6L^2\mathbb{E}[\sum_{i=1}^N ||\mathbf{x}_i^{t+1}-\mathbf{d}_i^t||^2]+24L^2\mathbb{E}[\sum_{i=1}^N ||\mathbf{x}_i^t-\bar{\mathbf{x}}^t||^2]+\frac{9L^2}{1-\rho}\mathbb{E}[\sum_{i=1}^N ||\mathbf{y}_i^t-\bar{\mathbf{y}}^t||^2]\\
&+\frac{6L^2\eta_y^2}{1-\rho}\mathbb{E}[\sum_{i=1}^N ||\nabla_{\mathbf{y}_i}\tilde{L}_{pi}(\{\mathbf{x}_i^{\hat{t}_i}\},\{\mathbf{y}_i^{\hat{t}_i}\},\{\lambda_{i,l}^{\hat{t}_i}\},\{\boldsymbol{\theta}_{i,j}^{\hat{t}_i}\})||^2]+3((c_2^{t-1})^2-(c_2^t)^2)\mathbb{E}[||\boldsymbol{\theta}_{i,j}^t||^2]+\frac{3}{\eta_\theta^2}\mathbb{E}[||\boldsymbol{\theta}_{i,j}^{t+1}-\boldsymbol{\theta}_{i,j}^t||^2].
\end{aligned} \tag{S92}$$

According to Eq.(S59), Eq.(S60), Eq.(S87), Eq.(S90), Eq.(S91) and Eq.(S92), we can obtain

$$\begin{aligned}
&\mathbb{E}[\tilde{\Psi}^t]\\
\le & (\frac{2}{\eta_x^2}+64NL^2\tau k_1+6NL^2(N+M))\mathbb{E}[\sum_{i=1}^N ||\mathbf{x}_i^{t+1}-\mathbf{d}_i^t||^2]\\
&+(256NL^2\tau k_1+32NL^2+24NL^2(N+M))\mathbb{E}[\sum_{i=1}^N ||\mathbf{x}_i^t-\bar{\mathbf{x}}^t||^2]\\
&+(\frac{96NL^2\tau k_1+9NL^2(N+M)}{1-\rho}+32NL^2)\mathbb{E}[\sum_{i=1}^N ||\mathbf{y}_i^t-\bar{\mathbf{y}}^t||^2]\\
&+(\frac{64\tau k_1\eta_y^2NL^2+6NL^2(N+M)\eta_y^2}{1-\rho}+2)\mathbb{E}[\sum_{i=1}^N ||\nabla_{\mathbf{y}_i}\tilde{L}_{pi}(\{\mathbf{x}_i^{\hat{t}_i}\},\{\mathbf{y}_i^{\hat{t}_i}\},\{\lambda_{i,l}^{\hat{t}_i}\},\{\boldsymbol{\theta}_{i,j}^{\hat{t}_i}\})||^2]\\
&+3((c_1^{t-1})^2-(c_1^t)^2)\mathbb{E}[\sum_{i=1}^N\sum_{l=1}^{|\mathcal{P}_i^t|} ||\lambda_{i,l}^t||^2]+\frac{3}{\eta_\lambda^2}\mathbb{E}[\sum_{i=1}^N\sum_{l=1}^{|\mathcal{P}_i^t|} ||\lambda_{i,l}^{t+1}-\lambda_{i,l}^t||^2]\\
&+3((c_2^{t-1})^2-(c_2^t)^2)\mathbb{E}[\sum_{i=1}^N\sum_{j=1}^N ||\boldsymbol{\theta}_{i,j}^t||^2]+\frac{3}{\eta_\theta^2}\mathbb{E}[\sum_{i=1}^N\sum_{j=1}^N ||\boldsymbol{\theta}_{i,j}^{t+1}-\boldsymbol{\theta}_{i,j}^t||^2].
\end{aligned} \tag{S93}$$

According to the inequality of norms squared differences, we have

$$\begin{aligned}
&\mathbb{E}[\Psi^t]-\mathbb{E}[\tilde{\Psi}^t]\\
=&\sum_{i=1}^{N}\sum_{l=1}^{|\mathcal{P}_i|}(\mathbb{E}[||(\mathcal{G}^t)_{\lambda_{i,j}}||^2]-\mathbb{E}[||(\tilde{\mathcal{G}}^t)_{\lambda_{i,j}}||^2])+\sum_{i=1}^{N}\sum_{j=1}^{N}(\mathbb{E}[||(\mathcal{G}^t)_{\boldsymbol{\theta}_{i,j}}||^2]-\mathbb{E}[||(\tilde{\mathcal{G}}^t)_{\boldsymbol{\theta}_{i,j}}||^2])\\
\leq&\sum_{i=1}^{N}\sum_{l=1}^{|\mathcal{P}_i^t|}\mathbb{E}[||c_1^{t-1}\lambda_{i,l}^t||^2]+\sum_{i=1}^{N}\sum_{j=1}^{N}\mathbb{E}[||c_2^{t-1}\boldsymbol{\theta}_{i,j}^t||^2].
\end{aligned}\tag{S94}$$

Plugging Eq.(S94) into Eq.(S93) to have

$$\begin{aligned}
&\mathbb{E}[\Psi^t]\\
\leq&(\frac{2}{\eta_x^2}+64NL^2\tau k_1+6NL^2(N+M))\mathbb{E}[\sum_{i=1}^{N}||\mathbf{x}_i^{t+1}-\mathbf{d}_i^t||^2]\\
+&(256NL^2\tau k_1+32NL^2+24NL^2(N+M))\mathbb{E}[\sum_{i=1}^{N}||\mathbf{x}_i^t-\bar{\mathbf{x}}^t||^2]\\
+&(\frac{96NL^2\tau k_1+9NL^2(N+M)}{1-\rho}+32NL^2)\mathbb{E}[\sum_{i=1}^{N}||\mathbf{y}_i^t-\bar{\mathbf{y}}^t||^2]\\
+&(\frac{64\tau k_1\eta_y^2NL^2+6NL^2(N+M)\eta_y^2}{1-\rho}+2)\mathbb{E}[\sum_{i=1}^{N}||\nabla_{\mathbf{y}_i}\tilde{L}_{pi}(\{\mathbf{x}_i^{\hat{t}_i}\},\{\mathbf{y}_i^{\hat{t}_i}\},\{\lambda_{i,l}^{\hat{t}_i}\},\{\boldsymbol{\theta}_{i,j}^{\hat{t}_i}\})||^2]\\
+&(4(c_1^{t-1})^2-3(c_1^t)^2)\mathbb{E}[\sum_{i=1}^{N}\sum_{l=1}^{|\mathcal{P}_i^t|}||\lambda_{i,l}^t||^2]+\frac{3}{\eta_\lambda^2}\mathbb{E}[\sum_{i=1}^{N}\sum_{l=1}^{|\mathcal{P}_i^t|}||\lambda_{i,l}^{t+1}-\lambda_{i,l}^t||^2]\\
+&(4(c_2^{t-1})^2-3(c_2^t)^2)\mathbb{E}[\sum_{i=1}^{N}\sum_{j=1}^{N}||\boldsymbol{\theta}_{i,j}^t||^2]+\frac{3}{\eta_\theta^2}\mathbb{E}[\sum_{i=1}^{N}\sum_{j=1}^{N}||\boldsymbol{\theta}_{i,j}^{t+1}-\boldsymbol{\theta}_{i,j}^t||^2]
\end{aligned}\tag{S95}$$

According to Eq.(S68), let $\eta_x=\eta_y=\eta_\lambda=\eta_\theta=\eta$, $\beta=1$ to have:

$$\begin{aligned}
&C_1\mathbb{E}[\sum_{i=1}^{N}||\mathbf{x}_i^{t+1}-\bar{\mathbf{x}}^t||^2]+C_2\mathbb{E}[\sum_{i=1}^{N}||\mathbf{x}_i^{t+1}-\mathbf{d}_i^t||^2]+C_3\mathbb{E}[\sum_{i=1}^{N}||\bar{\mathbf{x}}^t-\mathbf{x}_i^t||^2]+C_4\mathbb{E}[\sum_{i=1}^{N}||\bar{\mathbf{y}}^t-\mathbf{y}_i^t||^2]\\
+&C_5\mathbb{E}[\sum_{i=1}^{N}||\nabla_{\mathbf{y}_i}\tilde{L}_{pi}(\{\mathbf{x}_i^{\hat{t}_i}\},\{\mathbf{y}_i^{\hat{t}_i}\},\{\lambda_{i,l}^{\hat{t}_i}\},\{\boldsymbol{\theta}_{i,j}^{\hat{t}_i}\})||^2]+C_6\mathbb{E}[\sum_{i=1}^{N}\sum_{l=1}^{|\mathcal{P}_i^t|}||\lambda_{i,l}^{t+1}-\lambda_{i,l}^t||^2]+C_7\mathbb{E}[\sum_{i=1}^{N}\sum_{j=1}^{N}||\boldsymbol{\theta}_{i,j}^{t+1}-\boldsymbol{\theta}_{i,j}^t||^2]\\
\leq&\mathbb{E}[F^t-F^{t+1}]+\frac{4}{\eta}(\frac{c_1^{t-2}}{c_1^{t-1}}-\frac{c_1^{t-1}}{c_1^t})\mathbb{E}[\sum_{i=1}^{N}\sum_{l=1}^{|\mathcal{P}_i^t|}||\lambda_{i,l}^t||^2]+\frac{4}{\eta}(\frac{c_2^{t-2}}{c_2^{t-1}}-\frac{c_2^{t-1}}{c_2^t})\mathbb{E}[\sum_{i=1}^{N}\sum_{j=1}^{N}||\boldsymbol{\theta}_{i,j}^t||^2]\\
+&(\frac{c_1^{t-1}}{2}-\frac{c_1^t}{2})\mathbb{E}[\sum_{i=1}^{N}\sum_{l=1}^{|\mathcal{P}_i^t|}||\lambda_{i,l}^{t+1}||^2]+(\frac{c_2^{t-1}}{2}-\frac{c_2^t}{2})\mathbb{E}[\sum_{i=1}^{N}\sum_{j=1}^{N}||\boldsymbol{\theta}_{i,j}^{t+1}||^2],
\end{aligned}\tag{S96}$$

where

$$
\begin{aligned}
&C_1=\frac{N}{2\eta}-NL-\frac{3\eta N^2L^2}{2}-\frac{MNL^2\eta}{2}-\frac{N^2L^2\eta}{2},\\
&C_2=\frac{N}{2\eta}-2d_1^t-\frac{40L^2(\eta^2\gamma_1^t+\eta^2\gamma_2^t)}{1-\rho},\\
&C_3=\gamma_1^{t-1}-\gamma_1^t(\rho+\frac{170\eta^2L^2}{1-\rho})-\frac{N}{2\eta}-N^2L-\frac{3\eta N^2L^2}{2}-\frac{MNL^2\eta}{2}-\frac{N^2L^2\eta}{2}-8d_1^t,\\
&C_4=\gamma_2^{t-1}-\gamma_2^t(\rho+\frac{230\eta^2L^2}{1-\rho})-\frac{3\eta N^2L^2}{2}-\frac{MNL^2\eta}{2}-\frac{N^2L^2\eta}{2}-\frac{60\eta^2L^2\gamma_1^t}{1-\rho}-(\rho+\frac{2}{1-\rho})d_1^t,\\
&C_5=\eta N-\frac{\eta N}{2}-\frac{\eta^2NL}{2}-\frac{MNL^2\eta^3}{2}-\frac{N^2L^2\eta^3}{2}-\frac{40\eta^2L^2(\eta^2\gamma_1^t+\eta^2\gamma_2^t)}{(1-\rho)^2}-\frac{2\eta^2}{1-\rho}d_1^t,\\
&C_6=\frac{1}{\eta},\\
&C_7=\frac{1}{\eta},\\
&d_1^t=2\tau k_1N^2L+4\eta N^2L^2\tau k_1+\frac{8NML^2}{\eta(c_1^t)^2}+\frac{8N^2L^2}{\eta(c_2^t)^2}.
\end{aligned}
\tag{S97}
$$

Similarly, from Eq.(S95), we have

$$
\begin{aligned}
&\mathbb{E}[\Psi^t]\\
\le& C_1'\mathbb{E}[\sum_{i=1}^N||\mathbf{x}_i^{t+1}-\bar{\mathbf{x}}^t||^2]+C_2'\mathbb{E}[\sum_{i=1}^N||\mathbf{x}_i^{t+1}-\mathbf{d}_i^t||^2]+C_3'\mathbb{E}[\sum_{i=1}^N||\mathbf{x}_i^t-\bar{\mathbf{x}}^t||^2]+C_4'(\mathbb{E}[\sum_{i=1}^N||\bar{\mathbf{y}}^t-\mathbf{y}_i^t||^2])\\
&+C_5'\mathbb{E}[\sum_{i=1}^N||\nabla_{\mathbf{y}_i}\tilde{L}_{pi}(\{\mathbf{x}_i^{\hat{t}_i}\},\{\mathbf{y}_i^{\hat{t}_i}\},\{\lambda_{i,l}^{\hat{t}_i}\},\{\boldsymbol{\theta}_{i,j}^{\hat{t}_i}\})||^2]+C_6'\sum_{i=1}^N\sum_{l=1}^{|\mathcal{P}_i^t|}\mathbb{E}[||\lambda_{i,l}^{t+1}-\lambda_{i,l}^t||^2]+C_7'\sum_{i=1}^N\sum_{j=1}^N\mathbb{E}[||\boldsymbol{\theta}_{i,j}^{t+1}-\boldsymbol{\theta}_{i,j}^t||^2]\\
&+(4(c_1^{t-1})^2-3(c_1^t)^2)\sum_{i=1}^N\sum_{l=1}^{|\mathcal{P}_i^t|}\mathbb{E}[||\lambda_{i,l}^t||^2]+4((c_2^{t-1})^2-3(c_2^t)^2)\sum_{i=1}^N\sum_{j=1}^N\mathbb{E}[||\boldsymbol{\theta}_{i,j}^t||^2],
\end{aligned}
\tag{S98}
$$

where

$$
\begin{aligned}
&C_1'=0,\\
&C_2'=\frac{2}{\eta^2}+64NL^2\tau k_1+6NL^2(N+M),\\
&C_3'=256NL^2\tau k_1+32NL^2+24NL^2(N+M),\\
&C_4'=\frac{96NL^2\tau k_1+9NL^2(N+M)}{1-\rho}+32NL^2,\\
&C_5'=\frac{64\tau k_1\eta^2NL^2+6NL^2(N+M)\eta^2}{1-\rho}+2,\\
&C_6'=\frac{3}{\eta^2},\\
&C_7'=\frac{3}{\eta^2}.
\end{aligned}
\tag{S99}
$$

Let $\gamma_1=\gamma_2=\frac{2N}{\eta(1-\rho)}$, and $\eta\le\min\{\frac{1-\rho}{288LNk_1\tau},\frac{\sqrt{1-\rho}}{24L\sqrt{Nk_1\tau}},\frac{\sqrt{1-\rho}}{24L\sqrt{2(N+M)(T_1+T)}}\}$ to have

$$
\frac{3}{1-\rho}d_1^t=\frac{N}{\eta}(\frac{3}{1-\rho})(2\tau k_1NL\eta+4\eta^2NL^2\tau k_1+8(N+M)L^2\eta^2(T_1+T))\le\frac{N}{16\eta}.
\tag{S100}
$$

Let $\eta\le\min\{\frac{1}{8L},\frac{\sqrt{N}}{2L\sqrt{M}}\}$ to have

$$
\begin{aligned}
C_1&=\frac{N}{2\eta}-NL-2\eta N^2L^2-\frac{MNL^2\eta}{2},\\
&=\frac{N}{\eta}(\frac{1}{2}-L\eta-2\eta^2L^2-\frac{ML^2\eta^2}{2N})\\
&\ge\frac{N}{\eta}(\frac{1}{2}-\frac{1}{8}-\frac{1}{32}-\frac{1}{8})>0=C_1'
\end{aligned}
\tag{S101}
$$

Let $\eta < \frac{(1-\rho)^2}{640L^2N}$ to have

$$\begin{aligned} C_2 &= \frac{N}{2\eta} - 2d_1^t - \frac{40L^2(\eta^2\gamma_1^t + \eta^2\gamma_2^t)}{1-\rho}, \\ &= \frac{N}{2\eta} - 2d_1^t - \frac{160NL^2\eta}{(1-\rho)^2}, \\ &\geq \frac{N}{\eta}(\frac{1}{2} - \frac{1}{8} - \frac{1}{4}) > \frac{N}{8\eta}. \end{aligned} \tag{S102}$$

Let $\eta \leq \frac{1}{8L}$ to have

$$\begin{aligned} C_2' &= \frac{2}{\eta^2} + 64NL^2\tau k_1 + 6NL^2(N+M) \\ &\leq \frac{1}{\eta^2}(2 + N\tau k_1 + \frac{3N(N+M)}{32}). \end{aligned} \tag{S103}$$

Let $\eta \leq \{\frac{1-\rho}{2L\sqrt{85}}, \frac{1}{10NL}, \frac{1}{L\sqrt{2M}}\}$ to have

$$\begin{aligned} C_3 &= \gamma_1^{t-1} - \gamma_1^t(\rho + \frac{170\eta^2L^2}{1-\rho}) - \frac{N}{2\eta} - N^2L - \frac{3\eta N^2L^2}{2} - \frac{MNL^2\eta}{2} - \frac{N^2L^2\eta}{2} - 8d_1^t, \\ &\geq \frac{N}{\eta} - \frac{170N\eta L^2}{(1-\rho)^2} - \frac{5N^2L}{4} - \frac{MNL^2\eta}{2} \\ &= \frac{N}{\eta}(1 - \frac{170\eta^2L^2}{(1-\rho)^2} - \frac{5NL\eta}{4} - \frac{ML^2\eta^2}{2}) \\ &\geq \frac{N}{\eta}(1 - \frac{1}{2} - \frac{1}{8} - \frac{1}{4}) = \frac{N}{8\eta}. \end{aligned} \tag{S104}$$

Let $\eta \leq \{\frac{1-\rho}{20L}, \frac{1}{16N^2L^2}, \frac{1}{2L\sqrt{M}}\}$ to have

$$\begin{aligned} C_4 &= \gamma_2^{t-1} - \gamma_2^t(\rho + \frac{230\eta^2L^2}{1-\rho}) - \frac{3\eta N^2L^2}{2} - \frac{MNL^2\eta}{2} - \frac{N^2L^2\eta}{2} - \frac{60\eta^2L^2\gamma_1^t}{1-\rho} - (\rho + \frac{2}{1-\rho})d_1^t, \\ &\geq \frac{2N}{\eta} - \frac{520N\eta L^2}{(1-\rho)^2} - 2\eta N^2L^2 - \frac{MNL^2\eta}{2} - \frac{N}{16\eta} \\ &= \frac{N}{\eta}(2 - \frac{13}{10} - \frac{1}{8} - \frac{1}{8} - \frac{1}{16}) > \frac{N}{8\eta}. \end{aligned} \tag{S105}$$

Let $\eta \leq \{\frac{(1-\rho)^{\frac{3}{2}}}{40L}, \frac{1}{4L\sqrt{M}}, \frac{1}{4L\sqrt{N}}\}$ to have

$$\begin{aligned} \frac{1}{\eta}C_5 &= N - \frac{N}{2} - \frac{\eta NL}{2} - \frac{MNL^2\eta^2}{2} - \frac{N^2L^2\eta^2}{2} - \frac{40\eta L^2(\eta^2\gamma_1^t + \eta^2\gamma_2^t)}{(1-\rho)^2} - \frac{2\eta}{1-\rho}d_1^t, \\ &= \frac{N}{2} - \frac{\eta NL}{2} - \frac{MNL^2\eta^2}{2} - \frac{N^2L^2\eta^2}{2} - \frac{160NL^2\eta^2}{(1-\rho)^3} - \frac{N}{24} \\ &= N(\frac{1}{2} - \frac{\eta L}{2} - \frac{ML^2\eta^2}{2} - \frac{NL^2\eta^2}{2} - \frac{160L^2\eta^2}{(1-\rho)^3} - \frac{1}{24}) \\ &\geq N(\frac{1}{2} - \frac{1}{16} - \frac{1}{32} - \frac{1}{32} - \frac{1}{10} - \frac{1}{24}) > \frac{N}{8}, \end{aligned} \tag{S106}$$

and

$$\begin{aligned} \frac{1}{\eta}C_5' &= \frac{64\tau k_1\eta NL^2 + 6NL^2(N+M)\eta}{1-\rho} + 2 \\ &\leq \frac{8\tau k_1NL + \frac{3}{4}NL(N+M)}{1-\rho} + 2. \end{aligned} \tag{S107}$$

According to Eq.(S101), we have known that $C_1' < C_1$. Then, let $p_2 = \frac{8}{\eta}(\frac{2}{N} + \tau k_1 + \frac{3(N+M)}{32})$, $p_3 = (256L^2\tau k_1 + 32L^2 + 24L^2(N+M))8\eta$, $p_4 = (\frac{96L^2\tau k_1 + 9L^2(N+M)}{1-\rho} + 32L^2)8\eta$, $p_5 = \frac{64\tau k_1 L + 6L(N+M)}{1-\rho} + \frac{16}{N}$, $p_6 = p_7 = \frac{3}{\eta}$, and set $p = \max\{p_2, p_3, p_4, p_5, p_6, p_7\}$, we have

$$C_i' \leq pC_i, i = 2, 3, 4, 5, 6, 7. \tag{S108}$$

Multiply both sides of Eq.(S96) by p, sum Eq.(S96) and Eq.(S98) from $t = T_1 + 2 \cdots T_1 + T$ and divide by $T-1$. According to Eq.(S108) we have

$$
\begin{aligned}
&\frac{1}{T-1}\sum_{t=T_1+2}^{T_1+T}\mathbb{E}[\Psi^t] \\
\leq&\frac{p}{T-1}(\frac{F^{T_1+2}-\underline{L}}{N^2M}+\frac{c_1^1\alpha_1}{2N}+\frac{c_2^1\alpha_2}{2M}+\frac{4}{N\eta}(\frac{c_1^0}{c_1^1}+\frac{c_1^1}{c_1^2})\alpha_1+\frac{4}{M\eta}(\frac{c_2^0}{c_2^1}+\frac{c_2^1}{c_2^2})\alpha_2 \\
&+5(c_1^1)^2\alpha_1+5(c_2^1)^2\alpha_2+\frac{6}{N\eta}\sigma_1^2+\frac{6}{M\eta}{\sigma_2}^2+\frac{c_1^2}{2N}\alpha_1+\frac{c_2^2}{2M}\alpha_2) \\
=&\mathcal{O}(\frac{1}{T})
\end{aligned}
\tag{S109}
$$

where $\sigma_1 = \max\{||\lambda_1 - \lambda_2||\}$, $\sigma_2 = \max\{||\boldsymbol{\mu}_1 - \boldsymbol{\mu}_2||\}$, $\underline{L} = \min L_p(\{\bar{\boldsymbol{\theta}}^t\}, \{\bar{\boldsymbol{\phi}}^t\}, \{\lambda_{i,l}^t\}, \{\boldsymbol{\mu}_{i,j}^t\})$ satisfies $\forall t \geq T_1 + 2$ $F^t \geq \underline{L} - \frac{4}{\eta}\frac{c_1^1}{c_1^2}NM\alpha_1 - \frac{4}{\eta}\frac{c_2^1}{c_2^2}N^2\alpha_2 - \frac{6}{\eta}NM\sigma_1^2 - \frac{6}{\eta}N^2{\sigma_2}^2 - \frac{c_1^2}{2}NM\alpha_1 - \frac{c_2^2}{2}N^2\alpha_2$.

□

C. Communication Complexity Analysis

The communication complexity of Argus consists of two components: the complexity per iteration and the complexity of updating cutting planes.

1) Proof of Theorem 2:

Proof. In each iteration, after updating local parameters according to Eq.(17)-Eq.(18), each agent broadcasts the local variables $\mathbf{x}_i \in \mathbb{R}^n$ and $\mathbf{y}_i \in \mathbb{R}^m$ to neighbors. Then after updating dual variables according to Eq.(19)-Eq.(20), each active agent i broadcasts $\lambda_{i,l} \in \mathbb{R}^1, l \in |\mathcal{P}|$ and $\boldsymbol{\theta}_{i,j} \in \mathbb{R}^n, j \in \mathcal{N}_i$ to neighbors. In summary, the communication complexity of the t^{th} iteration can be calculated as follows:

$$
C_1^t = \underbrace{32d^t(N(m+n)}_{\text{Eq.(17), Eq.(18)}} + \underbrace{\sum_{i=1}^{N} p_i(|\mathcal{P}_i^t| + nd^t)}_{\text{Eq.(19), Eq.(20)}}),
\tag{S110}
$$

where $d^t = \frac{1}{N}\sum_{i=1}^{N}\sum_{j=1}^{N}\mathbf{W}^t$ denotes the average degree across all nodes.

Before the T_1^{th} iteration, Argus updates cutting planes every ι iterations. In each time to update cutting planes, Argus first estimates the lower-level problem as Eq.(5)-Eq.(6) and each agent broadcasts $\mathbf{y'}_i^{(k+1)} \in \mathbb{R}^m$ and $\boldsymbol{\varphi}_{ij}^{(k+1)} \in \mathbb{R}^m$ to neighbors. Then Argus updates cutting planes as Eq.(23) and agents broadcast $\boldsymbol{\mathcal{P}}_i^{t+1}$ and $\{\lambda_i^{t+1}\}$ to neighbors. The communication complexity of updating cutting planes is:

$$
C_2 = 32\sum_{t\in\mathcal{T}}(Nd^t(\underbrace{K(m+md^t)}_{\text{Eq.(5), Eq.(6)}} + \underbrace{(d^t(n+m)+1)}_{\boldsymbol{\mathcal{P}}_i^{t+1},\ \{\lambda_i^{t+1}\}})),
\tag{S111}
$$

where $\mathcal{T} = \{\iota, \cdots, \lfloor\frac{T_1}{\iota}\rfloor\}\cdot\iota$.

According to Eq.(S110) and Eq.(S111), we can obtain that the overall communication complexity of Argus is:

$$
\mathcal{O}(\sum_{t=1}^{T} C_1^t + C_2).
\tag{S112}
$$

□

D. Computational Complexity Analysis

The computational complexity of Argus consists of the computational per iteration and the computational of updating cutting planes.

1) Proof of Theorem 3:

Proof. We first calculate the FLOPs of updating local variables. In Eq.(15) and Eq.(16), each agent aggregates the variables of neighbors. In Eq.(17), active agents perform the proximal operator after a step of gradient descent. According to Assumption 1(b), we assume the proximal operator can be accomplished in one step. In Eq.(18), active agents perform a step of gradient descent. Besides, in Eq.(19) and Eq.(20), active agents perform gradient ascent. In summary, the computational complexity of the t^{th} iteration is:

$$\begin{aligned} C_{P_1}^t &= \underbrace{O\left(Nd^t(n+m)\right)}_{\text{Eq.(15), Eq.(16)}} + \underbrace{O\left(N\left|\mathcal{P}_i^t\right| d^t(n+m)\right)}_{\text{Eq.(17)}} + \underbrace{O\left(N\left|\mathcal{P}_i^t\right|^2 d^t(n+m)\right)}_{\text{Eq.(18)}} + \underbrace{O\left(N{d^t}^2 n\right)}_{\text{Eq.(19), Eq.(20)}} \\ &= O\left(N\left|\mathcal{P}_i^t\right|^2 d^t(n+m)\right) + O\left(N{d^t}^2 n\right) \end{aligned} \tag{S113}$$

where $d^t = \frac{1}{N}\sum_{i=1}^{N}\sum_{j=1}^{N}\mathbf{W}^t$ denotes the average degree across all nodes.

Then we calculate the FLOPs of updating local variables. Before the T_1^{th} iteration, Argus updates cutting planes every ι iterations. When estimating the lower-level problem using K rounds according to Eq.(5)-Eq.(6), Argus utilizes the proximal gradient descent to update $\mathbf{y}_i$ and utilizes the gradient ascent to update the dual variable ϕ_{ij}. After that, Argus calculate the parameters of the new cutting plane according to Eq.(24)-Eq.(26). In summary, the computational complexity of updating cutting planes is:

$$C_{P_2} = \sum_{t\in\mathcal{T}}(\mathcal{O}(Nd^t(n+m) + NmK)), \tag{S114}$$

where $\mathcal{T} = \{\iota, \cdots, \lfloor\frac{T_1}{\iota}\rfloor\}\cdot\iota$.

According to Eq.(S113) and Eq.(S114), the overall computational complexity of Argus is:

$$\sum_{t=1}^{T} C_{P_1}^t + C_{P_2}. \tag{S115}$$

□

References

[1] E. Jeong, M. Zecchin, and M. Kountouris, “Asynchronous decentralized learning over unreliable wireless networks,” in *ICC 2022-IEEE International Conference on Communications*. IEEE, 2022, pp. 607–612.