Appendix Material for "Argus: Federated Non-convex Bilevel Learning over 6G Space-Air-Ground Integrated Network"

Contents

1. Appendix

1.1. Theorem 5.5

(*Convergence of Argus*) Under Assumption 5.1 and Assumption 5.2, when step-sizes satisfy $\mathbb{E}[\eta^t_{i,x}] = \mathbb{E}[\eta^t_{i,y}] = \mathbb{E}[\eta^t_{i,\lambda}] = \mathbb{E}[\eta^t_{i,\theta}] = \eta$, and $\eta \leq \min\{\frac{1}{8L\sqrt{M}}, \frac{M\sqrt{N}}{L\sqrt{\tau k_1}}, \frac{\sqrt{M}}{L}, \frac{1}{64NL}, \frac{1}{12L\sqrt{N}}, \frac{1}{256\tau k_1 NL}, \frac{1}{32L\sqrt{\tau k_1 N}}, \frac{1}{L\sqrt{NM}}, \frac{1}{L\sqrt{NM\tau k_1}}, \frac{1}{64ML}, \frac{\sqrt{(\rho+\frac{2}{1-\rho})}}{8L\sqrt{NM}}, \frac{1-\rho}{64\tau k_1 NL}, \frac{\sqrt{1-\rho}}{8L\sqrt{\tau k_1 N}}, \frac{\sqrt{1-\rho}}{16L\sqrt{2(M+N)(T_1+T+1)}}, \frac{\sqrt{1-\rho}}{L\sqrt{M}}, \frac{\sqrt{1-\rho}}{L\sqrt{N}}\}$, then $(\{\mathbf{x}^t_i\}, \{\mathbf{y}^t_i\}, \{\lambda^t_{i,l}\}, \{\boldsymbol{\theta}^t_{i,j}\})$ generated by Argus satisfies

$$\begin{aligned} \frac{1}{T-1}\sum_{t=T_1+2}^{T_1+T}\Psi^t &\leq \frac{d}{(T-1)} + C_{bias} \\ &= \mathcal{O}(\frac{1}{T-1}) + C_{bias}, \end{aligned} \tag{S1}$$

where

$$\begin{aligned} d =& (\frac{F^{T_1+2}-\underline{L}}{N^2M} + \frac{c_1^1\alpha_1}{2N} + \frac{c_2^1\alpha_2}{2M} + \frac{4}{N\eta}(\frac{c_1^0}{c_1^1} + \frac{c_1^1}{c_1^2})\alpha_1 + \frac{4}{M\eta}(\frac{c_2^0}{c_2^1} + \frac{c_2^1}{c_2^2})\alpha_2 + 4(c_1^1)^2\alpha_1 + 4(c_2^1)^2\alpha_2 \\ &+ \frac{6}{N\eta}\beta_1^2 + \frac{6}{M\eta}{\beta_2}^2)(\max\{\frac{1024M}{\eta}, \frac{2500}{\eta}\}), \end{aligned} \tag{S2}$$

and

$$\begin{aligned} C_{bias} =& (42\sigma_1^2 + 18\varsigma_1^2 + 42\sigma_2^2 + 18\varsigma_2^2 + 4\sigma_3^2 + 4\sigma_4^2 + \frac{2(6\sigma_1^2+3\varsigma_1^2)}{M(1-\rho)^2} + \frac{(82+(\rho+\frac{2}{1-\rho})\frac{13}{512})(6\sigma_2^2+3\varsigma_2^2)}{M(1-\rho)^2} + \frac{\sigma_2^2}{M})(\max\{1024M, 2500\}) \\ &+ (\frac{\alpha_1+\alpha_2}{\eta^2(T_1+2)}). \end{aligned} \tag{S3}$$

Specifically, M is the maximum number of cutting planes, N is the number of agents, τ is the staleness bound, k_1 is a positive constant, $0 < \rho < 1$ represents the connectedness of the network, L is the Lipschitz constant, T_1 is the iteration before which the cutting planes will be updated.

From Theorem 5.5, it can be seen that the convergence performance of Argus is affected by the constant C_{bias}, which is mainly caused by the stochastic gradient estimator. In Argus, the bias term does not increase with the number of network agents N.

1.2. Notations

Definition 1. (Convergence metric) Inspired by [4; 7], we define a stationary gap $\mathcal{G}^t$ as follows:

$$\mathcal{G}^t = \begin{bmatrix} \{P(\mathbf{x}_i, \bar{\nabla}_{\mathbf{x}} L'_p(\{\mathbf{x}_i^t\}, \{\mathbf{y}_i^t\}, \{\lambda_{i,l}^t\}, \{\boldsymbol{\theta}_{i,j}^t\}), \eta_{i,x}^t)\} \\ \{\bar{\nabla}_{\mathbf{y}} L_p(\{\mathbf{x}_i^t\}, \{\mathbf{y}_i^t\}, \{\lambda_{i,l}^t\}, \{\boldsymbol{\theta}_{i,j}^t\})\} \\ \{\nabla_{\lambda_{i,l}} L_p(\{\mathbf{x}_i^t\}, \{\mathbf{y}_i^t\}, \{\lambda_{i,l}^t\}, \{\boldsymbol{\theta}_{i,j}^t\})\} \\ \{\nabla_{\boldsymbol{\theta}_{i,j}} L_p(\{\mathbf{x}_i^t\}, \{\mathbf{y}_i^t\}, \{\lambda_{i,l}^t\}, \{\boldsymbol{\theta}_{i,j}^t\})\} \end{bmatrix}, \tag{S4}$$

we further define that

$$\begin{aligned} (\mathcal{G}^t)_{\mathbf{x}_i} &= P(\mathbf{x}_i, \bar{\nabla}_{\mathbf{x}} L'_p(\{\mathbf{x}_i^t\}, \{\mathbf{y}_i^t\}, \{\lambda_{i,l}^t\}, \{\boldsymbol{\theta}_{i,j}^t\}), \eta_{i,x}^t), \\ (\mathcal{G}^t)_{\mathbf{y}} &= \bar{\nabla}_{\mathbf{y}_i} L_p(\{\mathbf{x}_i^t\}, \{\mathbf{y}_i^t\}, \{\lambda_{i,l}^t\}, \{\boldsymbol{\theta}_{i,j}^t\}), \\ (\mathcal{G}^t)_{\lambda_{i,l}} &= \nabla_{\lambda_{i,l}} L_p(\{\mathbf{x}_i^t\}, \{\mathbf{y}_i^t\}, \{\lambda_{i,l}^t\}, \{\boldsymbol{\theta}_{i,j}^t\}), \\ (\mathcal{G}^t)_{\boldsymbol{\theta}_{i,j}} &= \nabla_{\boldsymbol{\theta}_{i,j}} L_p(\{\mathbf{x}_i^t\}, \{\mathbf{y}_i^t\}, \{\lambda_{i,l}^t\}, \{\boldsymbol{\theta}_{i,j}^t\}), \end{aligned} \tag{S5}$$

where $P(\mathbf{a}, \mathbf{b}, \eta) \triangleq \frac{1}{\eta}(\mathbf{a} - \mathrm{prox}_{\eta}^{r}(\mathbf{a} - \eta \mathbf{b}))$, $\bar{\nabla}_{\mathbf{x}} L'_p(\{\mathbf{x}_i^t\}, \{\mathbf{y}_i^t\}, \{\lambda_{i,l}^t\}, \{\boldsymbol{\theta}_{i,j}^t\}) \triangleq \frac{1}{N}\sum_{i=1}^{N} \nabla_{\mathbf{x}_i} L'_p(\{\mathbf{x}_i^t\}, \{\mathbf{y}_i^t\}, \{\lambda_{i,l}^t\}, \{\boldsymbol{\theta}_{i,j}^t\})$, $\bar{\nabla}_{\mathbf{y}} L_p(\{\mathbf{x}_i^t\}, \{\mathbf{y}_i^t\}, \{\lambda_{i,l}^t\}, \{\boldsymbol{\theta}_{i,j}^t\}) \triangleq \frac{1}{N}\sum_{i=1}^{N} \nabla_{\mathbf{y}_i} L_p(\{\mathbf{x}_i^t\}, \{\mathbf{y}_i^t\}, \{\lambda_{i,l}^t\}, \{\boldsymbol{\theta}_{i,j}^t\})$.

Then we define $\tilde{\mathcal{G}}^t$ as follows:

$$\tilde{\mathcal{G}}^t = \begin{bmatrix} \{P(\mathbf{x}_i, \bar{\nabla}_{\mathbf{x}} \tilde{L}_p(\{\mathbf{x}_i^t\}, \{\mathbf{y}_i^t\}, \{\lambda_{i,l}^t\}, \{\boldsymbol{\theta}_{i,j}^t\}), \eta_{i,x}^t)\} \\ \{\bar{\nabla}_{\mathbf{y}} \tilde{L}_p(\{\mathbf{x}_i^t\}, \{\mathbf{y}_i^t\}, \{\lambda_{i,l}^t\}, \{\boldsymbol{\theta}_{i,j}^t\})\} \\ \{\nabla_{\lambda_{i,l}} \tilde{L}_p(\{\mathbf{x}_i^t\}, \{\mathbf{y}_i^t\}, \{\lambda_{i,l}^t\}, \{\boldsymbol{\theta}_{i,j}^t\})\} \\ \{\nabla_{\boldsymbol{\theta}_{i,j}} \tilde{L}_p(\{\mathbf{x}_i^t\}, \{\mathbf{y}_i^t\}, \{\lambda_{i,l}^t\}, \{\boldsymbol{\theta}_{i,j}^t\})\} \end{bmatrix}, \tag{S6}$$

Similarly, we have

$$\begin{aligned} (\tilde{\mathcal{G}}^t)_{\mathbf{x}_i} &= P(\mathbf{x}_i, \bar{\nabla}_{\mathbf{x}} \tilde{L}_p(\{\mathbf{x}_i^t\}, \{\mathbf{y}_i^t\}, \{\lambda_{i,l}^t\}, \{\boldsymbol{\theta}_{i,j}^t\}), \eta_{i,x}^t), \\ (\tilde{\mathcal{G}}^t)_{\mathbf{y}_i} &= \bar{\nabla}_{\mathbf{y}} \tilde{L}_p(\{\mathbf{x}_i^t\}, \{\mathbf{y}_i^t\}, \{\lambda_{i,l}^t\}, \{\boldsymbol{\theta}_{i,j}^t\}), \\ (\tilde{\mathcal{G}}^t)_{\lambda_{i,l}} &= \nabla_{\lambda_{i,l}} \tilde{L}_p(\{\mathbf{x}_i^t\}, \{\mathbf{y}_i^t\}, \{\lambda_{i,l}^t\}, \{\boldsymbol{\theta}_{i,j}^t\}), \\ (\tilde{\mathcal{G}}^t)_{\boldsymbol{\theta}_{i,j}} &= \nabla_{\boldsymbol{\theta}_{i,j}} \tilde{L}_p(\{\mathbf{x}_i^t\}, \{\mathbf{y}_i^t\}, \{\lambda_{i,l}^t\}, \{\boldsymbol{\theta}_{i,j}^t\}). \end{aligned} \tag{S7}$$

Let Ψ^t represents the convergence metric of our method, we define:

$$\begin{aligned} \Psi^t = &\frac{1}{N}\sum_{i=1}^{N} ||(\mathcal{G}^t)_{\mathbf{x}_i}||^2 + \frac{L^2}{N}\sum_{i=1}^{N} ||\mathbf{x}_i - \bar{\mathbf{x}}_i||^2 + \frac{1}{N}\sum_{i=1}^{N} ||(\mathcal{G}^t)_{\mathbf{y}_i}||^2 + \frac{L^2}{N} ||\mathbf{y}_i - \bar{\mathbf{y}}_i||^2 \\ &+ \frac{1}{NM}\sum_{i=1}^{N}\sum_{l=1}^{|\mathcal{P}_i|} ||(\mathcal{G}^t)_{\lambda_{i,j}}||^2 + \frac{1}{N^2}\sum_{i=1}^{N}\sum_{j=1}^{N} ||(\mathcal{G}^t)_{\boldsymbol{\theta}_{i,j}}||^2, \end{aligned} \tag{S8}$$

and

$$\begin{aligned} \tilde{\Psi}^t = &\frac{1}{N}\sum_{i=1}^{N} ||(\tilde{\mathcal{G}}^t)_{\mathbf{x}_i}||^2 + \frac{L^2}{N}\sum_{i=1}^{N} ||\mathbf{x}_i - \bar{\mathbf{x}}_i||^2 + \frac{1}{N}\sum_{i=1}^{N} ||(\tilde{\mathcal{G}}^t)_{\mathbf{y}_i}||^2 + \frac{L^2}{N} ||\mathbf{y}_i - \bar{\mathbf{y}}_i||^2 \\ &+ \frac{1}{NM}\sum_{i=1}^{N}\sum_{l=1}^{|\mathcal{P}_i|} ||(\tilde{\mathcal{G}}^t)_{\lambda_{i,j}}||^2 + \frac{1}{N^2}\sum_{i=1}^{N}\sum_{j=1}^{N} ||(\tilde{\mathcal{G}}^t)_{\boldsymbol{\theta}_{i,j}}||^2, \end{aligned} \tag{S9}$$

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

1.3. Assumptions

Assumption 5.1. Following [2; 3; 4; 6; 7; 8], we assume that functions and variables satisfy:

a) Lipschitzian gradient: Given that $L'_{p\_i}(\{\mathbf{x}_i\},\{\mathbf{y}_i\},\{\lambda_{i,l}\},\{\boldsymbol{\theta}_{i,j}\}) = G_i(\mathbf{x}_i,\mathbf{y}_i) + \sum_{l=1}^{|\boldsymbol{\mathcal{P}}_i^t|}\lambda_{i,l}(\sum_{j\in\mathcal{N}_i}\boldsymbol{a}_{j,l}^\top\mathbf{x}_j + \sum_{j\in\mathcal{N}_i}\boldsymbol{b}_{j,l}^\top\mathbf{y}_j + c_{i,l}) + \sum_{j\in\mathcal{N}_i}\boldsymbol{\theta}_{i,j}^\top(\mathbf{x}_i-\mathbf{x}_j)$, $i\in[N]$, L'_{p_i} has Lipschitz continuous gradients, i.e., for any $\mathbf{a},\mathbf{b}$, there exists $L>0$ satisfying that $||\nabla L'_{p\_i}(\mathbf{a})-\nabla L'_{p\_i}(\mathbf{b})||\le L||\mathbf{a}-\mathbf{b}||$.

b) Convex proximal operator: R and r are convex, possibly non-smooth functions, such as l_1 norm. They admit proximal mappings that are easily computable.

c) Bounded magnitude: Dual variables are bounded, i.e., $\mathbb{E}||\lambda_{i,l}||^2\le\alpha_1$, $\mathbb{E}||\boldsymbol{\theta}_{i,j}||^2\le\alpha_2$. And we assume that before obtaining the ϵ-stationary point, local variables satisfy that $\mathbb{E}[\sum_{i=1}^N||\mathbf{x}_i^{t+1}-\mathbf{x}_i^t||^2]+\mathbb{E}[\sum_{i=1}^N||\mathbf{y}_i^{t+1}-\mathbf{y}_i^t||^2]\ge\vartheta$, where $\vartheta>0$ is a relative small constant. The change of the local variables is upper bounded within τ iterations: $\mathbb{E}[\sum_{i=1}^N||\mathbf{x}_i^t-\mathbf{x}_i^{t-k}||^2]\le\tau k_1\vartheta$, $\mathbb{E}[\sum_{i=1}^N||\mathbf{y}_i^t-\mathbf{y}_i^{t-k}||^2]\le\tau k_1\vartheta, \forall 1\le k\le\tau$, where $k_1>0$ is a constants.

d) Unbiased estimation:

$$\begin{aligned}
&\mathbb{E}[\nabla_{\mathbf{x}_i}L'_{p_i}(\{\mathbf{x}_i\},\{\mathbf{y}_i\},\{\lambda_{i,l}\},\{\boldsymbol{\theta}_{i,j}\};\xi)] = \nabla_{\mathbf{x}_i}L'_{p_i}(\{\mathbf{x}_i\},\{\mathbf{y}_i\},\{\lambda_{i,l}\},\{\boldsymbol{\theta}_{i,j}\}),\\
&\mathbb{E}[\nabla_{\mathbf{y}_i}L'_{p_i}(\{\mathbf{x}_i\},\{\mathbf{y}_i\},\{\lambda_{i,l}\},\{\boldsymbol{\theta}_{i,j}\};\xi)] = \nabla_{\mathbf{x}_i}L'_{p_i}(\{\mathbf{x}_i\},\{\mathbf{y}_i\},\{\lambda_{i,l}\},\{\boldsymbol{\theta}_{i,j}\}),\\
&\mathbb{E}[\nabla_{\lambda_{i,l}}L'_{p_i}(\{\mathbf{x}_i\},\{\mathbf{y}_i\},\{\lambda_{i,l}\},\{\boldsymbol{\theta}_{i,j}\};\xi)] = \nabla_{\mathbf{x}_i}L'_{p_i}(\{\mathbf{x}_i\},\{\mathbf{y}_i\},\{\lambda_{i,l}\},\{\boldsymbol{\theta}_{i,j}\}),\\
&\mathbb{E}[\nabla_{\mathbf{x}_i}L'_{p_i}(\{\mathbf{x}_i\},\{\mathbf{y}_i\},\{\lambda_{i,l}\},\{\boldsymbol{\theta}_{i,j}\};\xi)] = \nabla_{\boldsymbol{\theta}_{i,j}}L'_{p_i}(\{\mathbf{x}_i\},\{\mathbf{y}_i\},\{\lambda_{i,l}\},\{\boldsymbol{\theta}_{i,j}\}).
\end{aligned}\tag{S10}$$

e) Bounded variance: There exist finite positive constants σ_1^2, σ_2^2, ς_1^2 and ς_2^2 such that

$$\begin{aligned}
&\mathbb{E}[||\nabla_{\mathbf{x}_i}L'_{p_i}(\{\mathbf{x}_i\},\{\mathbf{y}_i\},\{\lambda_{i,l}\},\{\boldsymbol{\theta}_{i,j}\})-\nabla_{\mathbf{x}_i}L'_{p_i}(\{\mathbf{x}_i\},\{\mathbf{y}_i\},\{\lambda_{i,l}\},\{\boldsymbol{\theta}_{i,j}\};\xi)||^2]\le\sigma_1^2,\\
&\mathbb{E}[||\nabla_{\mathbf{y}_i}L'_{p_i}(\{\mathbf{x}_i\},\{\mathbf{y}_i\},\{\lambda_{i,l}\},\{\boldsymbol{\theta}_{i,j}\})-\nabla_{\mathbf{y}_i}L'_{p_i}(\{\mathbf{x}_i\},\{\mathbf{y}_i\},\{\lambda_{i,l}\},\{\boldsymbol{\theta}_{i,j}\};\xi)||^2]\le\sigma_2^2,\\
&\frac{1}{N}\sum_{i=1}^N||\nabla_{\mathbf{x}_i}L'_{p_i}(\{\mathbf{x}_i\},\{\mathbf{y}_i\},\{\lambda_{i,l}\},\{\boldsymbol{\theta}_{i,j}\})-\bar{\nabla}_{\mathbf{x}}L'_{p_i}(\{\mathbf{x}_i\},\{\mathbf{y}_i\},\{\lambda_{i,l}\},\{\boldsymbol{\theta}_{i,j}\})||^2\le\varsigma_1^2,\\
&\frac{1}{N}\sum_{i=1}^N||\nabla_{\mathbf{y}_i}L'_{p_i}(\{\mathbf{x}_i\},\{\mathbf{y}_i\},\{\lambda_{i,l}\},\{\boldsymbol{\theta}_{i,j}\})-\bar{\nabla}_{\mathbf{y}}L'_{p_i}(\{\mathbf{x}\},\{\mathbf{y}_i\},\{\lambda_{i,l}\},\{\boldsymbol{\theta}_{i,j}\})||^2\le\varsigma_2^2,
\end{aligned}\tag{S11}$$

where σ_1^2 and σ_2^2 bound the variance of stochastic gradients in UL and LL problems at each agent. ς_1^2 and ς_2^2 quantify the similarity of data distributions at different agents [1; 6]. If all agents can access all data, then $\varsigma_1=0$ and $\varsigma_2=0$.

Assumption 5.2. Following [5; 6], we assume the mixing matrix $\mathbf{W}^t$ satisfies the following properties:

a) Network-defined sparsity: $\mathbf{W}_{ij}^t>0$ if $(i,j)\in\mathcal{E}^t$; otherwise $\mathbf{W}_{ij}^t=0$.

b) Symmetric: $\mathbf{W}^t={\mathbf{W}^t}^\top$.

c) Null-space property: null $(\mathbf{I}-\mathbf{W}^t)$ = span$\{\mathbf{e}\}$, where $\mathbf{e}\in\mathbb{R}^N$ is the vector of all ones.

d) Spectral property: The eigenvalues of $\mathbf{W}^t$ lie in the range $(-1,1]$ with $\rho\triangleq\left\|\mathbf{W}^t-\frac{1}{N}\mathbf{e}\mathbf{e}^\top\right\|_2<1$, where the value ρ indicates the connectedness of the graph [7].

1.4. Convergence Analysis

1.4.1 Lemma 1 (Descending Inequality of x Variables)

Under Assumptions 5.1 and 5.2, the following inequality holds,

$$\begin{aligned}
&\mathbb{E}[L_p(\{\bar{\mathbf{x}}^{t+1}\},\{\mathbf{y}_i^t\},\{\lambda_{i,l}^t\},\{\boldsymbol{\theta}_{i,j}^t\})-L_p(\{\bar{\mathbf{x}}^t\},\{\mathbf{y}_i^t\},\{\lambda_{i,l}^t\},\{\boldsymbol{\theta}_{i,j}^t\})]\\
\le&(NL-\frac{N}{2\eta_x})\mathbb{E}[\sum_{i=1}^N||\mathbf{x}_i^{t+1}-\bar{\mathbf{x}}^t||^2]-\frac{N}{2\eta_x}\mathbb{E}[\sum_{i=1}^N||\mathbf{x}_i^{t+1}-\mathbf{d}_i^t||^2]+(\frac{N}{2\eta_x}+N^2L)\mathbb{E}[\sum_{i=1}^N||\bar{\mathbf{x}}^t-\mathbf{x}_i^t||^2]\\
&+2\tau k_1N^2L(\mathbb{E}[\sum_{i=1}^N(||\mathbf{x}_i^{t+1}-\mathbf{x}_i^t||^2]+\mathbb{E}[\sum_{i=1}^N||\mathbf{y}_i^{t+1}-\mathbf{y}_i^t||^2])),
\end{aligned}\tag{S12}$$

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_{p_i} functions, we have

$$\begin{aligned}
&L_p(\{\bar{\mathbf{x}}^{t+1}\},\{\mathbf{y}_i^t\},\{\lambda_{i,l}^t\},\{\boldsymbol{\theta}_{i,j}^t\})-L_p(\{\bar{\mathbf{x}}^{t}\},\{\mathbf{y}_i^t\},\{\lambda_{i,l}^t\},\{\boldsymbol{\theta}_{i,j}^t\})\\
=&\sum_{i=1}^{N}(L_{p_i}(\{\bar{\mathbf{x}}^{t+1}\},\{\mathbf{y}_i^t\},\{\lambda_{i,l}^t\},\{\boldsymbol{\theta}_{i,j}^t\})-(L_{p_i}(\{\bar{\mathbf{x}}^{t}\},\{\mathbf{y}_i^t\},\{\lambda_{i,l}^t\},\{\boldsymbol{\theta}_{i,j}^t\}))\\
\le&\sum_{i=1}^{N}(L'_{p_i}(\{\bar{\mathbf{x}}^{t+1}\},\{\mathbf{y}_i^t\},\{\lambda_{i,l}^t\},\{\boldsymbol{\theta}_{i,j}^t\})-L'_{p_i}(\{\bar{\mathbf{x}}^{t}\},\{\mathbf{y}_i^t\},\{\lambda_{i,l}^t\},\{\boldsymbol{\theta}_{i,j}^t\})+(R(\bar{\mathbf{x}}^{t+1})-R(\bar{\mathbf{x}}^{t})))\\
\overset{(a)}{\le}&\sum_{i=1}^{N}(L'_{p_i}(\{\bar{\mathbf{x}}^{t+1}\},\{\mathbf{y}_i^t\},\{\lambda_{i,l}^t\},\{\boldsymbol{\theta}_{i,j}^t\})-L'_{p_i}(\{\bar{\mathbf{x}}^{t}\},\{\mathbf{y}_i^t\},\{\lambda_{i,l}^t\},\{\boldsymbol{\theta}_{i,j}^t\}))+N(\frac{1}{N}\sum_{i=1}^{N}R(\mathbf{x}_i^{t+1})-R(\bar{\mathbf{x}}^{t}))\\
\overset{(b)}{\le}&\frac{NL}{2}\sum_{i=1}^{N}||\bar{\mathbf{x}}^{t+1}-\bar{\mathbf{x}}^{t}||^2+N\sum_{i=1}^{N}\langle\nabla_{\mathbf{x}_i}L'_{p_i}(\{\bar{\mathbf{x}}^{t}\},\{\mathbf{y}_i^t\},\{\lambda_{i,l}^t\},\{\boldsymbol{\theta}_{i,j}^t\}),\bar{\mathbf{x}}^{t+1}-\bar{\mathbf{x}}^{t}\rangle+\sum_{i=1}^{N}(R(\mathbf{x}_i^{t+1})-R(\bar{\mathbf{x}}^{t})),
\end{aligned}\tag{S13}$$

where (a) is because of the convexity of $R(\cdot)$, (b) utilizs the Lipschitz properties in Assumption 5.1.

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

$$\mathbf{0}\in\mathbb{E}[\frac{\tilde{\eta}_{i,x}^t}{N}\partial R(\mathbf{x}_i^{t+1})+\mathbf{x}_i^{t+1}-(\mathbf{d}_i^t-\tilde{\eta}_{i,x}^t\nabla_{\mathbf{x}_i}\tilde{L}_{p\_i}(\{\mathbf{x}_i^{\hat{t}_i}\},\{\mathbf{y}_i^{\hat{t}_i}\},\{\lambda_{i,l}^{\hat{t}_i}\},\{\boldsymbol{\theta}_{i,j}^{\hat{t}_i}\};\xi))],\tag{S14}$$

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

$$\tilde{\eta}_{i,x}^t=\left\{\begin{array}{l}\eta_{i,x}^t, i\in Q^{t+1}\\ 0, i\notin Q^{t+1}\end{array}\right..\tag{S15}$$

Similar to [2], 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.(S15) and Eq.(S14), we have

$$\mathbf{0}\in\frac{\eta_x}{N}\partial R(\mathbf{x}_i^{t+1})+\mathbf{x}_i^{t+1}-(\mathbf{d}_i^t-\eta_x\nabla_{\mathbf{x}_i}\tilde{L}_{p\_i}(\{\mathbf{x}_i^{\hat{t}_i}\},\{\mathbf{y}_i^{\hat{t}_i}\},\{\lambda_{i,l}^{\hat{t}_i}\},\{\boldsymbol{\theta}_{i,j}^{\hat{t}_i}\})),\tag{S16}$$

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

$$\langle\mathbf{x}_i^{t+1}-\mathbf{x}_i,\frac{1}{N}\tilde{\nabla}R(\mathbf{x}_i^{t+1})+\frac{1}{\eta_x}(\mathbf{x}_i^{t+1}-\mathbf{d}_i^t)+\nabla_{\mathbf{x}_i}\tilde{L}_{p\_i}(\{\mathbf{x}_i^{\hat{t}_i}\},\{\mathbf{y}_i^{\hat{t}_i}\},\{\lambda_{i,l}^{\hat{t}_i}\},\{\boldsymbol{\theta}_{i,j}^{\hat{t}_i}\})\rangle=0,\tag{S17}$$

and

$$\langle\mathbf{x}_i^{t+1}-\mathbf{x}_i,\tilde{\nabla}R(\mathbf{x}_i^{t+1})+\frac{N}{\eta_x}(\mathbf{x}_i^{t+1}-\mathbf{d}_i^t)+N\nabla_{\mathbf{x}_i}\tilde{L}_{p\_i}(\{\mathbf{x}_i^{\hat{t}_i}\},\{\mathbf{y}_i^{\hat{t}_i}\},\{\lambda_{i,l}^{\hat{t}_i}\},\{\boldsymbol{\theta}_{i,j}^{\hat{t}_i}\})\rangle=0.\tag{S18}$$

According to the convexity of R, it holds for any $\mathbf{x}_i\in\mathbb{R}^n$ that

$$\begin{aligned}
&R(\mathbf{x}_i^{t+1})-R(\mathbf{x}_i)+\langle\mathbf{x}_i^{t+1}-\bar{\mathbf{x}}^t,N\nabla_{\mathbf{x}_i}\tilde{L}_{p_i}(\{\mathbf{x}_i^{\hat{t}_i}\},\{\mathbf{y}_i^{\hat{t}_i}\},\{\lambda_{i,l}^{\hat{t}_i}\},\{\boldsymbol{\theta}_{i,j}^{\hat{t}_i}\})\rangle\\
\le&\langle\mathbf{x}_i^{t+1}-\mathbf{x}_i,\tilde{\nabla}R(\mathbf{x}_i^{t+1})+N\nabla_{\mathbf{x}_i}\tilde{L}_{p_i}(\{\mathbf{x}_i^{\hat{t}_i}\},\{\mathbf{y}_i^{\hat{t}_i}\},\{\lambda_{i,l}^{\hat{t}_i}\},\{\boldsymbol{\theta}_{i,j}^{\hat{t}_i}\})\rangle\\
=&-\frac{N}{\eta_x}\langle\mathbf{x}_i^{t+1}-\mathbf{x}_i,\mathbf{x}_i^{t+1}-\mathbf{d}_i^t\rangle\\
\overset{(a)}{=}&-\frac{N}{2\eta_x}(||\mathbf{x}_i^{t+1}-\mathbf{x}_i||^2+||\mathbf{x}_i^{t+1}-\mathbf{d}_i^t||^2-||\mathbf{x}_i-\mathbf{d}_i^t||^2),
\end{aligned}\tag{S19}$$

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

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

$$\begin{aligned}
&R(\mathbf{x}_i^{t+1})-R(\bar{\mathbf{x}}^t)+\langle\mathbf{x}_i^{t+1}-\bar{\mathbf{x}}^t,N\nabla_{\mathbf{x}_i}\tilde{L}_{p_i}(\{\mathbf{x}_i^{\hat{t}_i}\},\{\mathbf{y}_i^{\hat{t}_i}\},\{\lambda_{i,l}^{\hat{t}_i}\},\{\boldsymbol{\theta}_{i,j}^{\hat{t}_i}\})\rangle\\
\le&-\frac{N}{2\eta_x}(||\mathbf{x}_i^{t+1}-\bar{\mathbf{x}}^t||^2+||\mathbf{x}_i^{t+1}-\mathbf{d}_i^t||^2-||\bar{\mathbf{x}}^t-\mathbf{d}_i^t||^2).
\end{aligned}\tag{S20}$$

Plugging Eq.(S20) into Eq.(S13) yields

$$\begin{aligned}
&L_p(\{\bar{\mathbf{x}}^{t+1}\},\{\mathbf{y}_i^t\},\{\lambda_{i,l}^t\},\{\boldsymbol{\theta}_{i,j}^t\})-L_p(\{\bar{\mathbf{x}}^t\},\{\mathbf{y}_i^t\},\{\lambda_{i,l}^t\},\{\boldsymbol{\theta}_{i,j}^t\})\\
\le&\frac{N^2L}{2}||\bar{\mathbf{x}}^{t+1}-\bar{\mathbf{x}}^t||^2+N\sum_{i=1}^N\langle\nabla_{\mathbf{x}_i}L'_{p_i}(\{\bar{\mathbf{x}}^t\},\{\mathbf{y}_i^t\},\{\lambda_{i,l}^t\},\{\boldsymbol{\theta}_{i,j}^t\}),\bar{\mathbf{x}}^{t+1}-\bar{\mathbf{x}}^t\rangle\\
&-N\sum_{i=1}^N\langle\mathbf{x}_i^{t+1}-\bar{\mathbf{x}}^t,\nabla_{\mathbf{x}_i}\tilde{L}_{p_i}(\{\mathbf{x}_i^{\hat{t}_i}\},\{\mathbf{y}_i^{\hat{t}_i}\},\{\lambda_{i,l}^{\hat{t}_i}\},\{\boldsymbol{\theta}_{i,j}^{\hat{t}_i}\})\rangle\\
&-\frac{N}{2\eta_x}\sum_{i=1}^N(||\mathbf{x}_i^{t+1}-\bar{\mathbf{x}}^t||^2+||\mathbf{x}_i^{t+1}-\mathbf{d}_i^t||^2-||\bar{\mathbf{x}}^t-\mathbf{d}_i^t||^2).
\end{aligned}\tag{S21}$$

In Eq.(S21), according to the linearity of the inner product, we have

$$\begin{aligned}
&N\sum_{i=1}^N\langle\nabla_{\mathbf{x}_i}L'_{p_i}(\{\bar{\mathbf{x}}^t\},\{\mathbf{y}_i^t\},\{\lambda_{i,l}^t\},\{\boldsymbol{\theta}_{i,j}^t\}),\bar{\mathbf{x}}^{t+1}-\bar{\mathbf{x}}^t\rangle-N\sum_{i=1}^N\langle\mathbf{x}_i^{t+1}-\bar{\mathbf{x}}^t,\nabla_{\mathbf{x}_i}\tilde{L}_{p_i}(\{\mathbf{x}_i^{\hat{t}_i}\},\{\mathbf{y}_i^{\hat{t}_i}\},\{\lambda_{i,l}^{\hat{t}_i}\},\{\boldsymbol{\theta}_{i,j}^{\hat{t}_i}\})\rangle\\
=&N\sum_{i=1}^N\langle\nabla_{\mathbf{x}_i}L'_{p_i}(\{\bar{\mathbf{x}}^t\},\{\mathbf{y}_i^t\},\{\lambda_{i,l}^t\},\{\boldsymbol{\theta}_{i,j}^t\}),\mathbf{x}_i^{t+1}-\bar{\mathbf{x}}^t\rangle-N\sum_{i=1}^N\langle\mathbf{x}_i^{t+1}-\bar{\mathbf{x}}^t,\nabla_{\mathbf{x}_i}\tilde{L}_{p_i}(\{\mathbf{x}_i^{\hat{t}_i}\},\{\mathbf{y}_i^{\hat{t}_i}\},\{\lambda_{i,l}^{\hat{t}_i}\},\{\boldsymbol{\theta}_{i,j}^{\hat{t}_i}\})\rangle\\
=&N\sum_{i=1}^N\langle\nabla_{\mathbf{x}_i}L'_{p_i}(\{\bar{\mathbf{x}}^t\},\{\mathbf{y}_i^t\},\{\lambda_{i,l}^t\},\{\boldsymbol{\theta}_{i,j}^t\})-\nabla_{\mathbf{x}_i}\tilde{L}_{p_i}(\{\mathbf{x}_i^{\hat{t}_i}\},\{\mathbf{y}_i^{\hat{t}_i}\},\{\lambda_{i,l}^{\hat{t}_i}\},\{\boldsymbol{\theta}_{i,j}^{\hat{t}_i}\}),\mathbf{x}_i^{t+1}-\bar{\mathbf{x}}^t\rangle\\
\le&\frac{NL}{2}\sum_{i=1}^N||\mathbf{x}_i^{t+1}-\bar{\mathbf{x}}^t||^2+\frac{N}{2L}\sum_{i=1}^N||\nabla_{\mathbf{x}_i}L'_{p_i}(\{\bar{\mathbf{x}}^t\},\{\mathbf{y}_i^t\},\{\lambda_{i,l}^t\},\{\boldsymbol{\theta}_{i,j}^t\})-\nabla_{\mathbf{x}_i}L'_{p_i}(\{\mathbf{x}_i^{\hat{t}_i}\},\{\mathbf{y}_i^{\hat{t}_i}\},\{\lambda_{i,l}^{\hat{t}_i}\},\{\boldsymbol{\theta}_{i,j}^{\hat{t}_i}\})||^2.
\end{aligned}\tag{S22}$$

Looking at the last term in Eq.(S22), combining the Cauchy-Schwarz inequality with Assumption 5.1, we have

$$\begin{aligned}
&||\nabla_{\mathbf{x}_i}L'_{p_i}(\{\bar{\mathbf{x}}^t\},\{\mathbf{y}_i^t\},\{\lambda_{i,l}^t\},\{\boldsymbol{\theta}_{i,j}^t\})-\nabla_{\mathbf{x}_i}L'_{p_i}(\{\mathbf{x}_i^{\hat{t}_i}\},\{\mathbf{y}_i^{\hat{t}_i}\},\{\lambda_{i,l}^{\hat{t}_i}\},\{\boldsymbol{\theta}_{i,j}^{\hat{t}_i}\})||^2\\
=&||\nabla_{\mathbf{x}_i}L'_{p_i}(\{\bar{\mathbf{x}}^t\},\{\mathbf{y}_i^t\},\{\lambda_{i,l}^t\},\{\boldsymbol{\theta}_{i,j}^t\})-\nabla_{\mathbf{x}_i}L'_{p_i}(\{\mathbf{x}_i^t\},\{\mathbf{y}_i^t\},\{\lambda_{i,l}^t\},\{\boldsymbol{\theta}_{i,j}^t\})\\
&+\nabla_{\mathbf{x}_i}L'_{p_i}(\{\mathbf{x}_i^t\},\{\mathbf{y}_i^t\},\{\lambda_{i,l}^t\},\{\boldsymbol{\theta}_{i,j}^t\})-\nabla_{\mathbf{x}_i}L'_{p_i}(\{\mathbf{x}_i^{\hat{t}_i}\},\{\mathbf{y}_i^{\hat{t}_i}\},\{\lambda_{i,l}^{\hat{t}_i}\},\{\boldsymbol{\theta}_{i,j}^{\hat{t}_i}\})||^2\\
\le&2L^2\sum_{i=1}^N||\bar{\mathbf{x}}^t-\mathbf{x}_i^t||^2+2L^2(\sum_{i=1}^N||\mathbf{x}_i^t-\mathbf{x}_i^{\hat{t}_i}||^2+\sum_{i=1}^N||\mathbf{y}_i^t-\mathbf{y}_i^{\hat{t}_i}||^2)\\
\le&2L^2\sum_{i=1}^N||\bar{\mathbf{x}}^t-\mathbf{x}_i^t||^2+4L^2\tau k_1(\sum_{i=1}^N||\mathbf{x}_i^{t+1}-\mathbf{x}_i^t||^2+\sum_{i=1}^N||\mathbf{y}_i^{t+1}-\mathbf{y}_i^t||^2).
\end{aligned}\tag{S23}$$

It follows from Jensens inequality that

$$||\bar{\mathbf{x}}^{t+1}-\bar{\mathbf{x}}^t||^2\le\frac{1}{N}\sum_{i=1}^N||\mathbf{x}_i^{t+1}-\bar{\mathbf{x}}^t||^2.\tag{S24}$$

According to Assumption 5.2 and the updating rules of $\mathbf{x}$ variables, we have:

$$\begin{aligned}
\left\|\frac{1}{N}\mathbf{e}\mathbf{e}^\top\mathbf{x}^t-\mathbf{d}^t\right\|_F^2&=\left\|\left(\mathbf{W}^t-\frac{1}{N}\mathbf{e}\mathbf{e}^\top\right)\left(\mathrm{I}-\frac{1}{N}\mathbf{e}\mathbf{e}^\top\right)\mathbf{x}^t\right\|_F^2\\
&\le\rho^2\left\|\frac{1}{N}\mathbf{e}\mathbf{e}^\top\mathbf{x}^t-\mathbf{x}^t\right\|_F^2<\left\|\frac{1}{N}\mathbf{e}\mathbf{e}^\top\mathbf{x}^t-\mathbf{x}^t\right\|_F^2,
\end{aligned}\tag{S25}$$

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

Similarly, it follows that

$$\begin{aligned}\left\|\frac{1}{N}\mathbf{e}\mathbf{e}^{\top}\mathbf{y}^{t}-\mathbf{u}^{t}\right\|_F^2&=\left\|\left(\mathbf{W}^{t}-\frac{1}{N}\mathbf{e}\mathbf{e}^{\top}\right)\left(\mathrm{I}-\frac{1}{N}\mathbf{e}\mathbf{e}^{\top}\right)\mathbf{y}^{t}\right\|_F^2\\&\le\rho^2\left\|\frac{1}{N}\mathbf{e}\mathbf{e}^{\top}\mathbf{y}^{t}-\mathbf{y}^{t}\right\|_F^2<\left\|\frac{1}{N}\mathbf{e}\mathbf{e}^{\top}\mathbf{y}^{t}-\mathbf{y}^{t}\right\|_F^2.\end{aligned}\tag{S26}$$

Lemma 1 follows via plugging Eq.(S22), Eq.(S23), Eq.(S24), and Eq.(S25) into Eq.(S21).

1.4.2 Lemma 2 (Descending Inequality of y Variables)

Under Assumptions 5.1 and 5.2, the following inequality holds,

$$\begin{aligned}&\mathbb{E}[L_p(\{\bar{\mathbf{x}}^{t+1}\},\{\bar{\mathbf{y}}^{t+1}\},\{\lambda_{i,l}^t\},\{\boldsymbol{\theta}_{i,j}^t\})-L_p(\{\bar{\mathbf{x}}^{t+1}\},\{\bar{\mathbf{y}}^{t}\},\{\lambda_{i,l}^t\},\{\boldsymbol{\theta}_{i,j}^t\})]\\
\le&\frac{2\eta_y N^2L^2}{\beta}(\mathbb{E}[\sum_{i=1}^N||\mathbf{x}^{t+1}-\bar{\mathbf{x}}^t||^2]+\mathbb{E}[\sum_{i=1}^N||\bar{\mathbf{x}}^{t}-\mathbf{x}_i^t||^2]+\mathbb{E}[\sum_{i=1}^N||\bar{\mathbf{y}}^{t}-\mathbf{y}_i^t||^2])\\
&+\frac{4\eta_y N^2L^2\tau k_1}{\beta}(\mathbb{E}[\sum_{i=1}^N||\mathbf{x}_i^{t+1}-\mathbf{x}_i^t||^2]+\mathbb{E}[\sum_{i=1}^N||\mathbf{y}_i^{t+1}-\mathbf{y}_i^t||^2)\\
&+(\frac{\eta_y N\beta}{2}+\frac{\eta_y^2 2NL}{2}-\eta_y N)\mathbb{E}[\sum_{i=1}^N||\bar{\nabla}_{\mathbf{y}_i}\tilde{L}_{p_i}(\{\mathbf{x}_i^{\hat{t}_i}\},\{\mathbf{y}_i^{\hat{t}_i}\},\{\lambda_{i,l}^{\hat{t}_i}\},\{\boldsymbol{\theta}_{i,j}^{\hat{t}_i}\};\xi)||^2]+\frac{2\eta_y N^2\sigma_2^2}{\beta},\end{aligned}\tag{S27}$$

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

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_{p_i}(\{\bar{\mathbf{x}}^{t+1}\},\{\bar{\mathbf{y}}^{t+1}\},\{\lambda_{i,l}^t\},\{\boldsymbol{\theta}_{i,j}^t\})-L_{p_i}(\{\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_{p_i}(\{\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_y N\mathbb{E}[\sum_{i=1}^N\langle\nabla_{\mathbf{y}_i}L_{p_i}(\{\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}_{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}\};\xi)\rangle]\\
&+\frac{\eta_y^2 2NL}{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}\};\xi)||^2]\\
\overset{(c)}{\le}&\frac{\eta_y N}{2\beta}\sum_{i=1}^N\mathbb{E}[||\nabla_{\mathbf{y}_i}L_{p_i}(\{\bar{\mathbf{x}}^{t+1}\},\{\bar{\mathbf{y}}^{t}\},\{\lambda_{i,l}^t\},\{\boldsymbol{\theta}_{i,j}^t\})-\bar{\nabla}_{\mathbf{y}_i}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}\};\xi)||^2]\\
&+(\frac{\eta_y N\beta}{2}+\frac{\eta_y^2 2NL}{2}-\eta_y N)\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}\};\xi)||^2],\end{aligned}\tag{S28}$$

where (a) utilizs the Lipschitz properties in Assumption 5.1. (b) is because of the updating rules of $\mathbf{y}$ variables. (c) uses the variants of the Cauchy-Schwarz inequality $\langle a,b\rangle\le\frac{1}{2\beta}||a||^2+\frac{\beta}{2}||b||^2$ in which β is a parameter that can be tuned later.

According to Assumption 5.1 and the triangle inequality, we have

$$
\begin{aligned}
&\mathbb{E}[||\nabla_{\mathbf{y}_i} L_{p\_i}(\{\bar{\mathbf{x}}^{t+1}\},\{\bar{\mathbf{y}}^{t}\},\{\lambda_{i,l}^{t}\},\{\boldsymbol{\theta}_{i,j}^{t}\})-\bar{\nabla}_{\mathbf{y}_i} 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}\};\xi)||^2]\\
=&\mathbb{E}[||\bar{\nabla}_{\mathbf{y}_i} L_{p\_i}(\{\bar{\mathbf{x}}^{t+1}\},\{\bar{\mathbf{y}}^{t}\},\{\lambda_{i,l}^{t}\},\{\boldsymbol{\theta}_{i,j}^{t}\})-\bar{\nabla}_{\mathbf{y}_i} 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}\};\xi)||^2]\\
\leq&\mathbb{E}[\frac{1}{N}\sum_{i=1}^{N}||\nabla_{\mathbf{y}_i} L_{p\_i}(\{\bar{\mathbf{x}}^{t+1}\},\{\bar{\mathbf{y}}^{t}\},\{\lambda_{i,l}^{t}\},\{\boldsymbol{\theta}_{i,j}^{t}\})-\nabla_{\mathbf{y}_i} 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}\};\xi)||^2]\\
=&\frac{1}{N}\mathbb{E}[\sum_{i=1}^{N}||\nabla_{\mathbf{y}_i} L_{p\_i}(\{\bar{\mathbf{x}}^{t+1}\},\{\bar{\mathbf{y}}^{t}\},\{\lambda_{i,l}^{t}\},\{\boldsymbol{\theta}_{i,j}^{t}\})-\nabla_{\mathbf{y}_i} L_{p\_i}(\{\bar{\mathbf{x}}^{t}\},\{\bar{\mathbf{y}}^{t}\},\{\lambda_{i,l}^{t}\},\{\boldsymbol{\theta}_{i,j}^{t}\})\\
&+\nabla_{\mathbf{y}_i} L_{p\_i}(\{\bar{\mathbf{x}}^{t}\},\{\bar{\mathbf{y}}^{t}\},\{\lambda_{i,l}^{t}\},\{\boldsymbol{\theta}_{i,j}^{t}\})-\nabla_{\mathbf{y}_i} L_{p\_i}(\{\mathbf{x}_i^{t}\},\{\mathbf{y}_i^{t}\},\{\lambda_{i,l}^{t}\},\{\boldsymbol{\theta}_{i,j}^{t}\})\\
&+\nabla_{\mathbf{y}_i} L_{p\_i}(\{\mathbf{x}_i^{t}\},\{\mathbf{y}_i^{t}\},\{\lambda_{i,l}^{t}\},\{\boldsymbol{\theta}_{i,j}^{t}\})-\nabla_{\mathbf{y}_i} 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}\})\\
&+\nabla_{\mathbf{y}_i} 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}\})-\nabla_{\mathbf{y}_i} 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}\};\xi)]\\
\leq&4L^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])+4\sigma_2^2\\
\leq&4L^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])\\
&+8L^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)+4\sigma_2^2
\end{aligned}
\tag{S29}
$$

Lemma 2 naturally follows via plugging Eq.(S29) into Eq.(S28).

1.4.3 Lemma 3 (Descending Inequality of the L_p Function)

Under Assumptions 5.1 and 5.2, 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{2\eta_y N^2L^2}{\beta}+MNL^2\eta_\lambda+N^2L^2\eta_{\boldsymbol{\theta}})\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{2\eta_y N^2L^2}{\beta}+MNL^2\eta_\lambda+N^2L^2\eta_{\boldsymbol{\theta}})\mathbb{E}[\sum_{i=1}^N||\bar{\mathbf{x}}^t-\mathbf{x}_i^t||^2]\\
&+(\frac{2\eta_y N^2L^2}{\beta}+MNL^2\eta_\lambda+N^2L^2\eta_{\boldsymbol{\theta}})(\mathbb{E}[\sum_{i=1}^N||\bar{\mathbf{y}}^t-\mathbf{y}_i^t||^2])\\
&+(2\tau k_1N^2L+\frac{4\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_yN\beta}{2}+\frac{\eta_y^2 2NL}{2}-\eta_yN+MNL^2\eta_y^2 2\eta_\lambda+N^2L^2\eta_y^2 2\eta_{\boldsymbol{\theta}})\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}\};\xi)||^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]+\frac{2\eta_yN^2\sigma_2^2}{\beta}.
\end{aligned}
\tag{S30}
$$

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

$$
\lambda_{i,l}^{t+1}=\lambda_{i,l}^t+\tilde{\eta}_{i,\lambda}^t\nabla_{\lambda_{i,l}}\tilde{L}_{p\_i}(\{\mathbf{x}_i^{t+1}\},\{\mathbf{y}_i^{t+1}\},\{\lambda_{i,l}^{\hat{t}_i}\},\{\boldsymbol{\theta}_{i,j}^{\hat{t}_i}\};\xi).
\tag{S31}
$$

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

$$
\langle\lambda_{i,l}^{t+1}-\lambda_{i,l}^t-\eta_\lambda^t\nabla_{\lambda_{i,l}}\tilde{L}_{p\_i}(\{\mathbf{x}_i^{t+1}\},\{\mathbf{y}_i^{t+1}\},\{\lambda_{i,l}^{\hat{t}_i}\},\{\boldsymbol{\theta}_{i,j}^{\hat{t}_i}\}),\lambda-\lambda_{i,l}^{t+1}\rangle=0.
\tag{S32}
$$

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

$$
\langle\nabla_{\lambda_{i,l}}\tilde{L}_{p\_i}(\{\mathbf{x}_i^{t+1}\},\{\mathbf{y}_i^{t+1}\},\{\lambda_{i,l}^{\hat{t}_i}\},\{\boldsymbol{\theta}_{i,j}^{\hat{t}_i}\})-\frac{1}{\eta_\lambda^t}(\lambda_{i,l}^{t+1}-\lambda_{i,l}^t),\lambda_{i,l}^t-\lambda_{i,l}^{t+1}\rangle=0.
\tag{S33}
$$

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

$$
\langle\nabla_{\lambda_{i,l}}\tilde{L}_{p\_i}(\{\mathbf{x}_i^{t}\},\{\mathbf{y}_i^{t}\},\{\lambda_{i,l}^{\widehat{t-1}_i}\},\{\boldsymbol{\theta}_{i,j}^{\widehat{t-1}_i}\})-\frac{1}{\eta_\lambda^t}(\lambda_{i,l}^{t}-\lambda_{i,l}^{t-1}),\lambda_{i,l}^{t+1}-\lambda_{i,l}^{t}\rangle=0.
\tag{S34}
$$

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

$$
\begin{aligned}
&\widetilde{L}_{p\_i}(\{\bar{\mathbf{x}}^{t+1}\},\{\bar{\mathbf{y}}^{t+1}\},\{\lambda_{i,l}^{t+1}\},\{\boldsymbol{\theta}_{i,j}^{t}\})-\widetilde{L}_{p\_i}(\{\bar{\mathbf{x}}^{t+1}\},\{\bar{\mathbf{y}}^{t+1}\},\{\lambda_{i,l}^{t}\},\{\boldsymbol{\theta}_{i,j}^{t}\})\\
\leq&\sum_{l=1}^{|\mathcal{P}_i^t|}\langle\nabla_{\lambda_{i,l}}\widetilde{L}_{p\_i}(\{\bar{\mathbf{x}}^{t+1}\},\{\bar{\mathbf{y}}^{t+1}\},\{\lambda_{i,l}^{t}\},\{\boldsymbol{\theta}_{i,j}^{t}\}),\lambda_{i,l}^{t+1}-\lambda_{i,l}^{t}\rangle\\
=&\sum_{l=1}^{|\mathcal{P}_i^t|}\langle\nabla_{\lambda_{i,l}}\widetilde{L}_{p\_i}(\{\bar{\mathbf{x}}^{t+1}\},\{\bar{\mathbf{y}}^{t+1}\},\{\lambda_{i,l}^{t}\},\{\boldsymbol{\theta}_{i,j}^{t}\})-\nabla_{\lambda_{i,l}}\widetilde{L}_{p\_i}(\{\bar{\mathbf{x}}^{t}\},\{\bar{\mathbf{y}}^{t}\},\{\lambda_{i,l}^{t}\},\{\boldsymbol{\theta}_{i,j}^{t}\}),\lambda_{i,l}^{t+1}-\lambda_{i,l}^{t}\rangle\\
+&\sum_{l=1}^{|\mathcal{P}_i^t|}\langle\nabla_{\lambda_{i,l}}\widetilde{L}_{p\_i}(\{\bar{\mathbf{x}}^{t}\},\{\bar{\mathbf{y}}^{t}\},\{\lambda_{i,l}^{t}\},\{\boldsymbol{\theta}_{i,j}^{t}\}),\lambda_{i,l}^{t+1}-\lambda_{i,l}^{t}\rangle.
\end{aligned}
\tag{S35}
$$

For the first term in Eq.(S35), we have

$$
\begin{aligned}
&\sum_{l=1}^{|\mathcal{P}_i^t|}\langle\nabla_{\lambda_{i,l}}\widetilde{L}_{p\_i}(\{\bar{\mathbf{x}}^{t+1}\},\{\bar{\mathbf{y}}^{t+1}\},\{\lambda_{i,l}^{t}\},\{\boldsymbol{\theta}_{i,j}^{t}\})-\nabla_{\lambda_{i,l}}\widetilde{L}_{p\_i}(\{\bar{\mathbf{x}}^{t}\},\{\bar{\mathbf{y}}^{t}\},\{\lambda_{i,l}^{t}\},\{\boldsymbol{\theta}_{i,j}^{t}\}),\lambda_{i,l}^{t+1}-\lambda_{i,l}^{t}\rangle\\
=&\sum_{l=1}^{|\mathcal{P}_i^t|}\langle\nabla_{\lambda_{i,l}}L_{p\_i}(\{\bar{\mathbf{x}}^{t+1}\},\{\bar{\mathbf{y}}^{t+1}\},\{\lambda_{i,l}^{t}\},\{\boldsymbol{\theta}_{i,j}^{t}\})-\nabla_{\lambda_{i,l}}L_{p\_i}(\{\bar{\mathbf{x}}^{t}\},\{\bar{\mathbf{y}}^{t}\},\{\lambda_{i,l}^{t}\},\{\boldsymbol{\theta}_{i,j}^{t}\}),\lambda_{i,l}^{t+1}-\lambda_{i,l}^{t}\rangle\\
+&\frac{c_1^{t-1}-c_1^t}{2}\sum_{l=1}^{|\mathcal{P}_i^t|}\left(\left\|\lambda_{i,l}^{t+1}\right\|^2-\|\lambda_{i,l}^{t}\|^2\right)-\frac{c_1^{t-1}-c_1^t}{2}\sum_{l=1}^{|\mathcal{P}_i^t|}\left\|\lambda_{i,l}^{t+1}-\lambda_{i,l}^{t}\right\|^2\\
\leq&\frac{NL^2}{2a_1}\sum_{l=1}^{|\mathcal{P}_i^t|}(||\bar{\mathbf{x}}^{t+1}-\bar{\mathbf{x}}^{t}||^2+||\bar{\mathbf{y}}^{t+1}-\bar{\mathbf{y}}^{t}||^2)+\frac{a_1}{2}\sum_{l=1}^{|\mathcal{P}_i^t|}||\lambda_{i,l}^{t+1}-\lambda_{i,l}^{t}||^2\\
+&\frac{c_1^{t-1}-c_1^t}{2}\sum_{l=1}^{|\mathcal{P}_i^t|}\left(\left\|\lambda_{i,l}^{t+1}\right\|^2-\|\lambda_{i,l}^{t}\|^2\right)-\frac{c_1^{t-1}-c_1^t}{2}\sum_{l=1}^{|\mathcal{P}_i^t|}\left\|\lambda_{i,l}^{t+1}-\lambda_{i,l}^{t}\right\|^2,
\end{aligned}
\tag{S36}
$$

where $a_1>0$ is a constant.

For the second term in Eq.(S35), according to Eq.(S34) we have

$$
\begin{aligned}
&\sum_{l=1}^{|\mathcal{P}_i^t|}\langle\nabla_{\lambda_{i,l}}\widetilde{L}_{p\_i}(\{\bar{\mathbf{x}}^{t}\},\{\bar{\mathbf{y}}^{t}\},\{\lambda_{i,l}^{t}\},\{\boldsymbol{\theta}_{i,j}^{t}\}),\lambda_{i,l}^{t+1}-\lambda_{i,l}^{t}\rangle\\
\leq&\sum_{l=1}^{|\mathcal{P}_i^t|}(\langle\nabla_{\lambda_{i,l}}\widetilde{L}_{p\_i}(\{\bar{\mathbf{x}}^{t}\},\{\bar{\mathbf{y}}^{t}\},\{\lambda_{i,l}^{t}\},\{\boldsymbol{\theta}_{i,j}^{t}\})-\nabla_{\lambda_{i,l}}\widetilde{L}_{p\_i}(\{\mathbf{x}_i^{t}\},\{\mathbf{y}_i^{t}\},\{\lambda_{i,l}^{t-\hat{1}_i}\},\{\boldsymbol{\theta}_{i,j}^{t-\hat{1}_i}\}),\lambda_{i,l}^{t+1}-\lambda_{i,l}^{t}\rangle\\
+&\frac{1}{\eta_\lambda^t}\langle\lambda_{i,l}^{t}-\lambda_{i,l}^{t-1},\lambda_{i,l}^{t+1}-\lambda_{i,l}^{t}\rangle)\\
\leq&\sum_{l=1}^{|\mathcal{P}_i^t|}\langle\nabla_{\lambda_{i,l}}\widetilde{L}_{p\_i}(\{\bar{\mathbf{x}}^{t}\},\{\bar{\mathbf{y}}^{t}\},\{\lambda_{i,l}^{t}\},\{\boldsymbol{\theta}_{i,j}^{t}\})-\nabla_{\lambda_{i,l}}\widetilde{L}_{p\_i}(\{\boldsymbol{\theta}_i^{t}\},\{\mathbf{y}_i^{t}\},\{\lambda_{i,l}^{t-1}\},\{\boldsymbol{\theta}_i^{t-1}\})\\
+&\nabla_{\lambda_{i,l}}\widetilde{L}_{p\_i}(\{\mathbf{x}_i^{t}\},\{\mathbf{y}_i^{t}\},\{\lambda_{i,l}^{t-1}\},\{\boldsymbol{\theta}_i^{t-1}\})-\nabla_{\lambda_{i,l}}\widetilde{L}_{p\_i}(\{\mathbf{x}_i^{t}\},\{\mathbf{y}_i^{t}\},\{\lambda_{i,l}^{t-\hat{1}_i}\},\{\boldsymbol{\theta}_{i,j}^{t-\hat{1}_i}\}),\lambda_{i,l}^{t+1}-\lambda_{i,l}^{t}\rangle\\
+&\frac{1}{\eta_\lambda^t}\langle\lambda_{i,l}^{t}-\lambda_{i,l}^{t-1},\lambda_{i,l}^{t+1}-\lambda_{i,l}^{t}\rangle).
\end{aligned}
\tag{S37}
$$

According to updating rules, we have

$$
\sum_{l=1}^{|\mathcal{P}_i^t|}\langle\nabla_{\lambda_{i,l}}\widetilde{L}_{p\_i}(\{\mathbf{x}_i^{t}\},\{\mathbf{y}_i^{t}\},\{\lambda_{i,l}^{t-1}\},\{\boldsymbol{\theta}_i^{t-1}\})-\nabla_{\lambda_{i,l}}\widetilde{L}_{p\_i}(\{\mathbf{x}_i^{t}\},\{\mathbf{y}_i^{t}\},\{\lambda_{i,l}^{t-\hat{1}_i}\},\{\boldsymbol{\theta}_{i,j}^{t-\hat{1}_i}\}),\lambda_{i,l}^{t+1}-\lambda_{i,l}^{t}\rangle=0.
\tag{S38}
$$

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

$$\frac{1}{\eta_\lambda}\left\langle \lambda_{i,l}^{t} - \lambda_{i,l}^{t-1}, \lambda_{i,l}^{t+1} - \lambda_{i,l}^{t} \right\rangle \leq \frac{1}{2\eta_\lambda}\left\| \lambda_{i,l}^{t+1} - \lambda_{i,l}^{t} \right\|^2 - \frac{1}{2\eta_\lambda}\left\| v_{1,i,l}^{t+1} \right\|^2 + \frac{1}{2\eta_\lambda}\left\| \lambda_{i,l}^{t} - \lambda_{i,l}^{t-1} \right\|^2, \tag{S39}$$

and

$$\begin{aligned}
&\sum_{l=1}^{|\mathcal{P}_i^t|}\langle \nabla_{\lambda_{i,l}} \widetilde{L}_{p_i}(\{\bar{\mathbf{x}}^t\}, \{\bar{\mathbf{y}}^t\}, \{\lambda_{i,l}^t\}, \{\boldsymbol{\theta}_{i,j}^t\}) - \nabla_{\lambda_{i,l}} \widetilde{L}_{p_i}(\{\mathbf{x}_i^t\}, \{\mathbf{y}_i^t\}, \{\lambda_{i,l}^{t-1}\}, \{\boldsymbol{\theta}_{i,j}^{t-1}\}), \lambda_{i,l}^{t+1} - \lambda_{i,l}^{t}\rangle \\
=&\sum_{l=1}^{|\mathcal{P}_i^t|}\langle \nabla_{\lambda_{i,l}} \widetilde{L}_{p_i}(\{\bar{\mathbf{x}}^t\}, \{\bar{\mathbf{y}}^t\}, \{\lambda_{i,l}^t\}, \{\boldsymbol{\theta}_{i,j}^t\}) - \nabla_{\lambda_{i,l}} \widetilde{L}_{p_i}(\{\mathbf{x}_i^t\}, \{\mathbf{y}_i^t\}, \{\lambda_{i,l}^{t}\}, \{\boldsymbol{\theta}_{i,j}^{t}\}), \lambda_{i,l}^{t+1} - \lambda_{i,l}^{t}\rangle (1a) \\
+&\sum_{l=1}^{|\mathcal{P}_i^t|}\langle \nabla_{\lambda_{i,l}} \widetilde{L}_{p_i}(\{\mathbf{x}_i^t\}, \{\mathbf{y}_i^t\}, \{\lambda_{i,l}^t\}, \{\boldsymbol{\theta}_{i,j}^t\}) - \nabla_{\lambda_{i,l}} \widetilde{L}_{p_i}(\{\mathbf{x}_i^t\}, \{\mathbf{y}_i^t\}, \{\lambda_{i,l}^{t-1}\}, \{\boldsymbol{\theta}_{i,j}^{t-1}\}), v_{1,i,l}^{t+1}\rangle (1b) \\
+&\sum_{l=1}^{|\mathcal{P}_i^t|}\langle \nabla_{\lambda_{i,l}} \widetilde{L}_{p_i}(\{\mathbf{x}_i^t\}, \{\mathbf{y}_i^t\}, \{\lambda_{i,l}^t\}, \{\boldsymbol{\theta}_{i,j}^t\}) - \nabla_{\lambda_{i,l}} \widetilde{L}_{p_i}(\{\mathbf{x}_i^t\}, \{\mathbf{y}_i^t\}, \{\lambda_{i,l}^{t-1}\}, \{\boldsymbol{\theta}_{i,j}^{t-1}\}), \lambda_{i,l}^{t} - \lambda_{i,l}^{t-1}\rangle (1c).
\end{aligned} \tag{S40}$$

It follows from the Cauchy-Schwarz inequality and Assumption 5.1:

$$\begin{aligned}
&\sum_{l=1}^{|\mathcal{P}_i^t|}\langle \nabla_{\lambda_{i,l}} \widetilde{L}_{p_i}(\{\bar{\mathbf{x}}^t\}, \{\bar{\mathbf{y}}^t\}, \{\lambda_{i,l}^t\}, \{\boldsymbol{\theta}_{i,j}^t\}) - \nabla_{\lambda_{i,l}} \widetilde{L}_{p_i}(\{\mathbf{x}_i^t\}, \{\mathbf{y}_i^t\}, \{\lambda_{i,l}^{t}\}, \{\boldsymbol{\theta}_{i,j}^{t}\}), \lambda_{i,l}^{t+1} - \lambda_{i,l}^{t}\rangle \\
=&\sum_{l=1}^{|\mathcal{P}_i^t|}\langle \nabla_{\lambda_{i,l}} L_{p_i}(\{\bar{\mathbf{x}}^t\}, \{\bar{\mathbf{y}}^t\}, \{\lambda_{i,l}^t\}, \{\boldsymbol{\theta}_{i,j}^t\}) - \nabla_{\lambda_{i,l}} L_{p_i}(\{\mathbf{x}_i^t\}, \{\mathbf{y}_i^t\}, \{\lambda_{i,l}^{t}\}, \{\boldsymbol{\theta}_{i,j}^{t}\}), \lambda_{i,l}^{t+1} - \lambda_{i,l}^{t}\rangle \\
\leq&\frac{|\mathcal{P}_i^t|L^2}{2a_2}\sum_{i=1}^{N}(||\bar{\mathbf{x}}^t - \mathbf{x}_i^t||^2 + ||\bar{\mathbf{y}}^t - \mathbf{y}_i^t||^2) + \frac{a_2}{2}\sum_{l=1}^{|\mathcal{P}_i^t|}||\lambda_{i,l}^{t+1} - \lambda_{i,l}^{t}||^2,
\end{aligned} \tag{S41}$$

where $a_2 > 0$ is a constant.

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

$$\begin{aligned}
&\sum_{l=1}^{|\mathcal{P}_i^t|}\langle \nabla_{\lambda_{i,l}} \widetilde{L}_{p_i}(\{\mathbf{x}_i^t\}, \{\mathbf{y}_i^t\}, \{\lambda_{i,l}^t\}, \{\boldsymbol{\theta}_{i,j}^t\}) - \nabla_{\lambda_{i,l}} \widetilde{L}_{p_i}(\{\mathbf{x}_i^t\}, \{\mathbf{y}_i^t\}, \{\lambda_{i,l}^{t-1}\}, \{\boldsymbol{\theta}_{i,j}^{t-1}\}), v_{1,i,l}^{t+1}\rangle \\
\leq&\sum_{l=1}^{|\mathcal{P}_i^t|}(\frac{a_3}{2}||\nabla_{\lambda_{i,l}} \widetilde{L}_{p_i}(\{\mathbf{x}_i^t\}, \{\mathbf{y}_i^t\}, \{\lambda_{i,l}^t\}, \{\boldsymbol{\theta}_{i,j}^t\}) - \nabla_{\lambda_{i,l}} \widetilde{L}_{p_i}(\{\mathbf{x}_i^t\}, \{\mathbf{y}_i^t\}, \{\lambda_{i,l}^{t-1}\}, \{\boldsymbol{\theta}_{i,j}^{t-1}\})||^2 + \frac{1}{2a_3}||v_{1,i,l}^{t+1}||^2),
\end{aligned} \tag{S42}$$

where $a_3 > 0$ is a constant.

Defining $L_1' = L + c_1^0$, combining Assumption 5.1 and the triangle inequality, $\forall \lambda_{i,l}$ we have

$$\begin{aligned}
&||\nabla_{\lambda_{i,l}} \widetilde{L}_{p_i}(\{\mathbf{x}_i^t\}, \{\mathbf{y}_i^t\}, \{\lambda_{i,l}^t\}, \{\boldsymbol{\theta}_{i,j}^t\}) - \nabla_{\lambda_{i,l}} \widetilde{L}_{p_i}(\{\mathbf{x}_i^t\}, \{\mathbf{y}_i^t\}, \{\lambda_{i,l}^{t-1}\}, \{\boldsymbol{\theta}_{i,j}^{t-1}\})|| \\
=&||\nabla_{\lambda_{i,l}} L_{p_i}(\{\mathbf{x}_i^t\}, \{\mathbf{y}_i^t\}, \{\lambda_{i,l}^t\}, \{\boldsymbol{\theta}_{i,j}^t\}) - \nabla_{\lambda_{i,l}} L_{p_i}(\{\mathbf{x}_i^t\}, \{\mathbf{y}_i^t\}, \{\lambda_{i,l}^{t-1}\}, \{\boldsymbol{\theta}_{i,j}^{t}\}) - c_1^{t-1}(\lambda_{i,l}^t - \lambda_{i,l}^{t-1})|| \\
\leq&(L + c_1^{t-1})||\lambda_{i,l}^t - \lambda_{i,l}^{t-1}|| \\
\leq&L_1'||\lambda_{i,l}^t - \lambda_{i,l}^{t-1}||.
\end{aligned} \tag{S43}$$

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

$$\begin{aligned}
&\sum_{l=1}^{|\mathcal{P}_i^t|}\langle\nabla_{\lambda_{i,l}}\widetilde{L}_{p_i}(\{\mathbf{x}_i^t\},\{\mathbf{y}_i^t\},\{\lambda_{i,l}^t\},\{\boldsymbol{\theta}_{i,j}^t\})-\nabla_{\lambda_{i,l}}\widetilde{L}_{p_i}(\{\mathbf{x}_i^t\},\{\mathbf{y}_i^t\},\{\lambda_{i,l}^{t-1}\},\{\boldsymbol{\theta}_{i,j}^{t-1}\}),\lambda_{i,l}^t-\lambda_{i,l}^{t-1}\rangle\\
\le&\sum_{l=1}^{|\mathcal{P}_i^t|}(-\frac{1}{L_1'+c_1^{t-1}}||\nabla_{\lambda_{i,l}}\widetilde{L}_{p_i}(\{\mathbf{x}_i^t\},\{\mathbf{y}_i^t\},\{\lambda_{i,l}^t\},\{\boldsymbol{\theta}_{i,j}^t\})-\nabla_{\lambda_{i,l}}\widetilde{L}_{p_i}(\{\mathbf{x}_i^t\},\{\mathbf{y}_i^t\},\{\lambda_{i,l}^{t-1}\},\{\boldsymbol{\theta}_{i,j}^{t-1}\})||^2\\
&-\frac{c_1^{t-1}L_1'}{L_1'+c_1^{t-1}}||\lambda_{i,l}^t-\lambda_{i,l}^{t-1}||^2).
\end{aligned}\tag{S44}$$

Combining Eq.(S35), Eq.(S36), Eq.(S37), Eq.(S38), Eq.(S39), Eq.(S40), Eq.(S41), Eq.(S42), Eq.(S43) and Eq.(S44), 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

$$\begin{aligned}
&\mathbb{E}[L_{p_i}(\{\bar{\mathbf{x}}^{t+1}\},\{\bar{\mathbf{y}}^{t+1}\},\{\lambda_{i,l}^{t+1}\},\{\boldsymbol{\theta}_{i,j}^t\})-L_{p_i}(\{\bar{\mathbf{x}}^{t+1}\},\{\bar{\mathbf{y}}^{t+1}\},\{\lambda_{i,l}^t\},\{\boldsymbol{\theta}_{i,j}^t\})]\\
\le&\frac{MNL^2}{2a_1}(\mathbb{E}[||\bar{\mathbf{x}}^{t+1}-\bar{\mathbf{x}}^t||^2]+\mathbb{E}[||\bar{\mathbf{y}}^{t+1}-\bar{\mathbf{y}}^t||^2])+\frac{ML^2}{2a_1}(\mathbb{E}[\sum_{i=1}^N||\bar{\mathbf{x}}^t-\mathbf{x}_i^t||^2]+\mathbb{E}[\sum_{i=1}^N||\bar{\mathbf{y}}^t-\mathbf{y}_i^t||^2])\\
&+(a_1-\frac{c_1^{t-1}-c_1^t}{2}+\frac{1}{2\eta_\lambda})\mathbb{E}[\sum_{l=1}^{|\mathcal{P}_i^t|}||\lambda_{i,l}^{t+1}-\lambda_{i,l}^t||^2]+\frac{c_1^{t-1}}{2}\mathbb{E}[\sum_{l=1}^{|\mathcal{P}_i^t|}(||\lambda_{i,l}^{t+1}||^2]-||\lambda_{i,l}^t||^2)\\
&+\frac{1}{2\eta_\lambda}\mathbb{E}[\sum_{l=1}^{|\mathcal{P}_i^t|}||\lambda_{i,l}^t-\lambda_{i,l}^{t-1}||^2],
\end{aligned}\tag{S45}$$

and

$$\begin{aligned}
&\mathbb{E}[L_p(\{\bar{\mathbf{x}}^{t+1}\},\{\bar{\mathbf{y}}^{t+1}\},\{\lambda_{i,l}^{t+1}\},\{\boldsymbol{\theta}_{i,j}^t\})-L_p(\{\bar{\mathbf{x}}^{t+1}\},\{\bar{\mathbf{y}}^{t+1}\},\{\lambda_{i,l}^t\},\{\boldsymbol{\theta}_{i,j}^t\})]\\
\le&\frac{MN^2L^2}{2a_1}(\mathbb{E}[||\bar{\mathbf{x}}^{t+1}-\bar{\mathbf{x}}^t||^2]+\mathbb{E}[||\bar{\mathbf{y}}^{t+1}-\bar{\mathbf{y}}^t||^2])+\frac{MNL^2}{2a_1}(\mathbb{E}[\sum_{i=1}^N||\bar{\mathbf{x}}^t-\mathbf{x}_i^t||^2]+\mathbb{E}[\sum_{i=1}^N||\bar{\mathbf{y}}^t-\mathbf{y}_i^t||^2])\\
&+(a_1-\frac{c_1^{t-1}-c_1^t}{2}+\frac{1}{2\eta_\lambda})\mathbb{E}[\sum_{i=1}^N\sum_{l=1}^{|\mathcal{P}_i^t|}||\lambda_{i,l}^{t+1}-\lambda_{i,l}^t||^2]+\frac{c_1^{t-1}}{2}\mathbb{E}[\sum_{i=1}^N\sum_{l=1}^{\mathcal{P}_i^t}(||\lambda_{i,l}^{t+1}||^2]-\mathbb{E}[||\lambda_{i,l}^t||^2])\\
&+\frac{1}{2\eta_\lambda}\mathbb{E}[\sum_{i=1}^N\sum_{l=1}^{|\mathcal{P}_i^t|}||\lambda_{i,l}^t-\lambda_{i,l}^{t-1}||^2].
\end{aligned}\tag{S46}$$

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

$$\mathbb{E}[\sum_{i=1}^N||\bar{\mathbf{y}}^{t+1}-\bar{\mathbf{y}}^t||^2]=(\eta_y)^2\mathbb{E}[\sum_{i=1}^N||\bar{\nabla}_{\mathbf{y}_i}\tilde{L}_{p\_i}(\{\mathbf{x}_i^{\hat{t}_i}\},\{\mathbf{y}_i^{\hat{t}_i}\},\{\lambda_{i,l}^{\hat{t}_i}\},\{\boldsymbol{\theta}_{i,j}^{\hat{t}_i}\};\xi)||^2],\tag{S47}$$

we have

$$\begin{aligned}
&\mathbb{E}[L_p(\{\bar{\mathbf{x}}^{t+1}\},\{\bar{\mathbf{y}}^{t+1}\},\{\lambda_{i,l}^{t+1}\},\{\boldsymbol{\theta}_{i,j}^t\})-L_p(\{\bar{\mathbf{x}}^{t+1}\},\{\bar{\mathbf{y}}^{t+1}\},\{\lambda_{i,l}^t\},\{\boldsymbol{\theta}_{i,j}^t\})]\\
\le&\frac{MNL^2}{2a_1}\mathbb{E}[\sum_{i=1}^N||\mathbf{x}_i^{t+1}-\bar{\mathbf{x}}^t||^2]+\frac{MNL^2\eta_y^2}{2a_1}\mathbb{E}[\sum_{i=1}^N||\bar{\nabla}_{\mathbf{y}_i}\tilde{L}_{p_i}(\{\mathbf{x}_i^{\hat{t}_i}\},\{\mathbf{y}_i^{\hat{t}_i}\},\{\lambda_{i,l}^{\hat{t}_i}\},\{\boldsymbol{\theta}_{i,j}^{\hat{t}_i}\};\xi)||^2]\\
&+\frac{MNL^2}{2a_1}(\mathbb{E}[\sum_{i=1}^N||\bar{\mathbf{x}}^t-\mathbf{x}_i^t||^2]+\mathbb{E}[\sum_{i=1}^N||\bar{\mathbf{y}}^t-\mathbf{y}_i^t||^2])+(a_1-\frac{c_1^{t-1}-c_1^t}{2}+\frac{1}{2\eta_\lambda})\sum_{i=1}^N\sum_{l=1}^{|\mathcal{P}_i^t|}\mathbb{E}[||\lambda_{i,l}^{t+1}-\lambda_{i,l}^t||^2]\\
&+\frac{c_1^{t-1}}{2}\mathbb{E}[\sum_{i=1}^N\sum_{l=1}^{|\mathcal{P}_i^t|}(||\lambda_{i,l}^{t+1}||^2-||\lambda_{i,l}^t||^2)]+\frac{1}{2\eta_\lambda}\mathbb{E}[\sum_{i=1}^N\sum_{l=1}^{|\mathcal{P}_i^t|}||\lambda_{i,l}^t-\lambda_{i,l}^{t-1}||^2].
\end{aligned}\tag{S48}$$

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

$$\langle \boldsymbol{\theta}_{i,j}^{t+1} - \boldsymbol{\theta}_{i,j}^{t} - \eta_\theta \nabla_{\boldsymbol{\theta}_{i,j}} \tilde{L}_{p\_i}(\{\mathbf{x}_i^{t+1}\},\{\mathbf{y}_i^{t+1}\},\{\lambda_{i,l}^{t+1}\},\{\boldsymbol{\theta}_{i,j}^{t}\}), \boldsymbol{\theta} - \boldsymbol{\theta}_{i,j}^{t+1}\rangle = 0. \tag{S49}$$

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

$$\langle \nabla_{\boldsymbol{\theta}_{i,j}} \tilde{L}_{p\_i}(\{\mathbf{x}_i^{t+1}\},\{\mathbf{y}_i^{t+1}\},\{\lambda_{i,l}^{t+1}\},\{\boldsymbol{\theta}_{i,j}^{t}\}) - \frac{1}{\eta_\theta}(\boldsymbol{\theta}_{i,j}^{t+1} - \boldsymbol{\theta}_{i,j}^{t}), \boldsymbol{\theta}_{i,j}^{t} - \boldsymbol{\theta}_{i,j}^{t+1}\rangle = 0. \tag{S50}$$

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

$$\langle \nabla_{\boldsymbol{\theta}_{i,j}} \tilde{L}_{p\_i}(\{\mathbf{x}_i^{t}\},\{\mathbf{y}_i^{t}\},\{\lambda_{i,l}^{t}\},\{\boldsymbol{\theta}_{i,j}^{t-1}\}) - \frac{1}{\eta_\theta}(\boldsymbol{\theta}_{i,j}^{t} - \boldsymbol{\theta}_{i,j}^{t-1}), \boldsymbol{\theta}_{i,j}^{t+1} - \boldsymbol{\theta}_{i,j}^{t}\rangle = 0. \tag{S51}$$

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

$$\begin{aligned}
&\widetilde{L}_{p_i}(\{\bar{\mathbf{x}}^{t+1}\},\{\bar{\mathbf{y}}^{t+1}\},\{\lambda_{i,l}^{t+1}\},\{\boldsymbol{\theta}_{i,j}^{t+1}\}) - \widetilde{L}_{p_i}(\{\bar{\mathbf{x}}^{t+1}\},\{\bar{\mathbf{y}}^{t+1}\},\{\lambda_{i,l}^{t+1}\},\{\boldsymbol{\theta}_{i,j}^{t}\}) \\
\le& \sum_{j=1}^{N} \langle \nabla_{\boldsymbol{\theta}_{i,j}} \widetilde{L}_{p_i}(\{\bar{\mathbf{x}}^{t+1}\},\{\bar{\mathbf{y}}^{t+1}\},\{\lambda_{i,l}^{t+1}\},\{\boldsymbol{\theta}_{i,j}^{t}\}), \boldsymbol{\theta}_{i,j}^{t+1} - \boldsymbol{\theta}_{i,j}^{t}\rangle \\
=& \sum_{j=1}^{N} \langle \nabla_{\boldsymbol{\theta}_{i,j}} \widetilde{L}_{p_i}(\{\bar{\mathbf{x}}^{t+1}\},\{\bar{\mathbf{y}}^{t+1}\},\{\lambda_{i,l}^{t+1}\},\{\boldsymbol{\theta}_{i,j}^{t}\}) - \nabla_{\boldsymbol{\theta}_{i,j}} \widetilde{L}_{p_i}(\{\bar{\mathbf{x}}^{t}\},\{\bar{\mathbf{y}}^{t}\},\{\lambda_{i,l}^{t}\},\{\boldsymbol{\theta}_{i,j}^{t}\}), \boldsymbol{\theta}_{i,j}^{t+1} - \boldsymbol{\theta}_{i,j}^{t}\rangle \\
&+ \sum_{j=1}^{N} \langle \nabla_{\boldsymbol{\theta}_{i,j}} \widetilde{L}_{p_i}(\{\bar{\mathbf{x}}^{t}\},\{\bar{\mathbf{y}}^{t}\},\{\lambda_{i,l}^{t}\},\{\boldsymbol{\theta}_{i,j}^{t}\}), \boldsymbol{\theta}_{i,j}^{t+1} - \boldsymbol{\theta}_{i,j}^{t}\rangle.
\end{aligned} \tag{S52}$$

For the first term in Eq.(S52), we have

$$\begin{aligned}
&\sum_{j=1}^{N} \langle \nabla_{\boldsymbol{\theta}_{i,j}} \widetilde{L}_{p_i}(\{\bar{\mathbf{x}}^{t+1}\},\{\bar{\mathbf{y}}^{t+1}\},\{\lambda_{i,l}^{t+1}\},\{\boldsymbol{\theta}_{i,j}^{t}\}) - \nabla_{\boldsymbol{\theta}_{i,j}} \widetilde{L}_{p_i}(\{\bar{\mathbf{x}}^{t}\},\{\bar{\mathbf{y}}^{t}\},\{\lambda_{i,l}^{t}\},\{\boldsymbol{\theta}_{i,j}^{t}\}), \boldsymbol{\theta}_{i,j}^{t+1} - \boldsymbol{\theta}_{i,j}^{t}\rangle \\
=& \sum_{j=1}^{N} \langle \nabla_{\boldsymbol{\theta}_{i,j}} L_{p_i}(\{\bar{\mathbf{x}}^{t+1}\},\{\bar{\mathbf{y}}^{t+1}\},\{\lambda_{i,l}^{t}\},\{\boldsymbol{\theta}_{i,j}^{t}\}) - \nabla_{\boldsymbol{\theta}_{i,j}} L_{p_i}(\{\bar{\mathbf{x}}^{t}\},\{\bar{\mathbf{y}}^{t}\},\{\lambda_{i,l}^{t}\},\{\boldsymbol{\theta}_{i,j}^{t}\}), \boldsymbol{\theta}_{i,j}^{t+1} - \boldsymbol{\theta}_{i,j}^{t}\rangle \\
&+ \frac{c_2^{t-1} - c_2^{t}}{2} \sum_{j=1}^{N} \left(\|\boldsymbol{\theta}_{i,j}^{t+1}\|^2 - \|\boldsymbol{\theta}_{i,j}^{t}\|^2 \right) - \frac{c_2^{t-1} - c_2^{t}}{2} \sum_{j=1}^{N} \|\boldsymbol{\theta}_{i,j}^{t+1} - \boldsymbol{\theta}_{i,j}^{t}\|^2 \\
\le& \frac{NL^2}{2a_4} \sum_{j=1}^{N} (||\bar{\mathbf{x}}^{t+1} - \bar{\mathbf{x}}^{t}||^2 + ||\bar{\mathbf{y}}^{t+1} - \bar{\mathbf{y}}^{t}||^2) + \frac{a_4}{2} \sum_{j=1}^{N} ||\boldsymbol{\theta}_{i,j}^{t+1} - \boldsymbol{\theta}_{i,j}^{t}||^2 \\
&+ \frac{c_2^{t-1} - c_2^{t}}{2} \sum_{j=1}^{N} \left(\|\boldsymbol{\theta}_{i,j}^{t+1}\|^2 - \|\boldsymbol{\theta}_{i,j}^{t}\|^2 \right) - \frac{c_2^{t-1} - c_2^{t}}{2} \sum_{j=1}^{N} \|\boldsymbol{\theta}_{i,j}^{t+1} - \boldsymbol{\theta}_{i,j}^{t}\|^2,
\end{aligned} \tag{S53}$$

where $a_4 > 0$ is a constant.

For the second term in Eq.(S52), according to Eq.(S51) we have

$$\begin{aligned}
&\sum_{j=1}^{N} \langle \nabla_{\boldsymbol{\theta}_{i,j}} \widetilde{L}_{p_i}(\{\bar{\mathbf{x}}^{t}\},\{\bar{\mathbf{y}}^{t}\},\{\lambda_{i,l}^{t}\},\{\boldsymbol{\theta}_{i,j}^{t}\}), \boldsymbol{\theta}_{i,j}^{t+1} - \boldsymbol{\theta}_{i,j}^{t}\rangle \\
\le& \sum_{j=1}^{N} (\langle \nabla_{\boldsymbol{\theta}_{i,j}} \widetilde{L}_{p_i}(\{\bar{\mathbf{x}}^{t}\},\{\bar{\mathbf{y}}^{t}\},\{\lambda_{i,l}^{t}\},\{\boldsymbol{\theta}_{i,j}^{t}\}) \\
&- \nabla_{\boldsymbol{\theta}_{i,j}} \widetilde{L}_{p_i}(\{\mathbf{x}_i^{t}\},\{\mathbf{y}_i^{t}\},\{\lambda_{i,l}^{t}\},\{\boldsymbol{\theta}_{i,j}^{t-1}\}), \boldsymbol{\theta}_{i,j}^{t+1} - \boldsymbol{\theta}_{i,j}^{t}\rangle + \frac{1}{\eta_\theta}\langle \boldsymbol{\theta}_{i,j}^{t} - \boldsymbol{\theta}_{i,j}^{t-1}, \boldsymbol{\theta}_{i,j}^{t+1} - \boldsymbol{\theta}_{i,j}^{t}\rangle).
\end{aligned} \tag{S54}$$

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

$$\frac{1}{\eta_\theta} \left\langle \boldsymbol{\theta}_{i,j}^{t} - \boldsymbol{\theta}_{i,j}^{t-1}, \boldsymbol{\theta}_{i,j}^{t+1} - \boldsymbol{\theta}_{i,j}^{t} \right\rangle \leq \frac{1}{2\eta_\theta} \left\| \boldsymbol{\theta}_{i,j}^{t+1} - \boldsymbol{\theta}_{i,j}^{t} \right\|^2 - \frac{1}{2\eta_\theta} \left\| v_{2,i,j}^{t+1} \right\|^2 + \frac{1}{2\eta_\theta} \left\| \boldsymbol{\theta}_{i,j}^{t} - \boldsymbol{\theta}_{i,j}^{t-1} \right\|^2, \tag{S55}$$

and

$$\begin{aligned}
&\sum_{j=1}^{N} \langle \nabla_{\boldsymbol{\theta}_{i,j}} \widetilde{L}_{p_i}(\{\bar{\mathbf{x}}^t\}, \{\bar{\mathbf{y}}^t\}, \{\lambda_{i,l}^t\}, \{\boldsymbol{\theta}_{i,j}^t\}) - \nabla_{\boldsymbol{\theta}_{i,j}} \widetilde{L}_{p_i}(\{\mathbf{x}_i^t\}, \{\mathbf{y}_i^t\}, \{\lambda_{i,l}^t\}, \{\boldsymbol{\theta}_{i,j}^{t-1}\}), \boldsymbol{\theta}_{i,j}^{t+1} - \boldsymbol{\theta}_{i,j}^t \rangle \\
=&\sum_{j=1}^{N} \langle \nabla_{\boldsymbol{\theta}_{i,j}} \widetilde{L}_{p_i}(\{\bar{\mathbf{x}}^t\}, \{\bar{\mathbf{y}}^t\}, \{\lambda_{i,l}^t\}, \{\boldsymbol{\theta}_{i,j}^t\}) - \nabla_{\boldsymbol{\theta}_{i,j}} \widetilde{L}_{p_i}(\{\mathbf{x}_i^t\}, \{\mathbf{y}_i^t\}, \{\lambda_{i,l}^t\}, \{\boldsymbol{\theta}_{i,j}^{t}\}), \boldsymbol{\theta}_{i,j}^{t+1} - \boldsymbol{\theta}_{i,j}^t \rangle (2a) \\
&+\sum_{j=1}^{N} \langle \nabla_{\boldsymbol{\theta}_{i,j}} \widetilde{L}_{p_i}(\{\mathbf{x}_i^t\}, \{\mathbf{y}_i^t\}, \{\lambda_{i,l}^t\}, \{\boldsymbol{\theta}_{i,j}^t\}) - \nabla_{\boldsymbol{\theta}_{i,j}} \widetilde{L}_{p_i}(\{\mathbf{x}_i^t\}, \{\mathbf{y}_i^t\}, \{\lambda_{i,l}^t\}, \{\boldsymbol{\theta}_{i,j}^{t-1}\}), v_{2,i,j}^{t+1} \rangle (2b) \\
&+\sum_{j=1}^{N} \langle \nabla_{\boldsymbol{\theta}_{i,j}} \widetilde{L}_{p_i}(\{\mathbf{x}_i^t\}, \{\mathbf{y}_i^t\}, \{\lambda_{i,l}^t\}, \{\boldsymbol{\theta}_{i,j}^t\}) - \nabla_{\boldsymbol{\theta}_{i,j}} \widetilde{L}_{p_i}(\{\mathbf{x}_i^t\}, \{\mathbf{y}_i^t\}, \{\lambda_{i,l}^t\}, \{\boldsymbol{\theta}_{i,j}^{t-1}\}), \boldsymbol{\theta}_{i,j}^{t} - \boldsymbol{\theta}_{i,j}^{t-1} \rangle (2c).
\end{aligned} \tag{S56}$$

According to the Cauchy-Schwarz inequality with Assumption 5.1, we have the following inequality from $(2a)$:

$$\begin{aligned}
&\sum_{j=1}^{N} \langle \nabla_{\boldsymbol{\theta}_{i,j}} \widetilde{L}_{p_i}(\{\bar{\mathbf{x}}^t\}, \{\bar{\mathbf{y}}^t\}, \{\lambda_{i,l}^t\}, \{\boldsymbol{\theta}_{i,j}^t\}) - \nabla_{\boldsymbol{\theta}_{i,j}} \widetilde{L}_{p_i}(\{\mathbf{x}_i^t\}, \{\mathbf{y}_i^t\}, \{\lambda_{i,l}^t\}, \{\boldsymbol{\theta}_{i,j}^{t}\}), \boldsymbol{\theta}_{i,j}^{t+1} - \boldsymbol{\theta}_{i,j}^t \rangle \\
=&\sum_{j=1}^{N} \langle \nabla_{\boldsymbol{\theta}_{i,j}} L_{p_i}(\{\bar{\mathbf{x}}^t\}, \{\bar{\mathbf{y}}^t\}, \{\lambda_{i,l}^t\}, \{\boldsymbol{\theta}_{i,j}^t\}) - \nabla_{\boldsymbol{\theta}_{i,j}} L_{p_i}(\{\mathbf{x}_i^t\}, \{\mathbf{y}_i^t\}, \{\lambda_{i,l}^t\}, \{\boldsymbol{\theta}_{i,j}^{t}\}), \boldsymbol{\theta}_{i,j}^{t+1} - \boldsymbol{\theta}_{i,j}^t \rangle \\
\leq&\frac{NL^2}{2a_5} \sum_{i=1}^{N} (||\bar{\mathbf{x}}^t - \mathbf{x}_i^t||^2 + ||\bar{\mathbf{y}}^t - \mathbf{y}_i^t||^2) + \frac{a_5}{2} \sum_{j=1}^{N} ||\boldsymbol{\theta}_{i,j}^{t+1} - \boldsymbol{\theta}_{i,j}^t||^2,
\end{aligned} \tag{S57}$$

where $a_5 > 0$ is a constant.

According to the Cauchy-Schwarz inequality, (2b) can be expressed as follows:

$$\begin{aligned}
&\sum_{j=1}^{N} \langle \nabla_{\boldsymbol{\theta}_{i,j}} \widetilde{L}_{p_i}(\{\mathbf{x}_i^t\}, \{\mathbf{y}_i^t\}, \{\lambda_{i,l}^t\}, \{\boldsymbol{\theta}_{i,j}^t\}) - \nabla_{\boldsymbol{\theta}_{i,j}} \widetilde{L}_{p_i}(\{\mathbf{x}_i^t\}, \{\mathbf{y}_i^t\}, \{\lambda_{i,l}^t\}, \{\boldsymbol{\theta}_{i,j}^{t-1}\}), v_{2,i,j}^{t+1} \rangle \\
\leq&\sum_{j=1}^{N} (\frac{a_6}{2} ||\nabla_{\boldsymbol{\theta}_{i,j}} \widetilde{L}_{p_i}(\{\mathbf{x}_i^t\}, \{\mathbf{y}_i^t\}, \{\lambda_{i,l}^t\}, \{\boldsymbol{\theta}_{i,j}^t\}) - \nabla_{\boldsymbol{\theta}_{i,j}} \widetilde{L}_{p_i}(\{\mathbf{x}_i^t\}, \{\mathbf{y}_i^t\}, \{\lambda_{i,l}^t\}, \{\boldsymbol{\theta}_{i,j}^{t-1}\})||^2 + \frac{1}{2a_6} ||v_{2,i,j}^{t+1}||^2),
\end{aligned} \tag{S58}$$

where $a_6 > 0$ is a constant.

Defining $L_2' = L + c_2^0$, combining Assumption 5.1 and the triangle inequality, $\forall \boldsymbol{\theta}_{i,j}$, we have,

$$\begin{aligned}
&||\nabla_{\boldsymbol{\theta}_{i,j}} \widetilde{L}_{p_i}(\{\mathbf{x}_i^t\}, \{\mathbf{y}_i^t\}, \{\lambda_{i,l}^t\}, \{\boldsymbol{\theta}_{i,j}^t\}) - \nabla_{\boldsymbol{\theta}_{i,j}} \widetilde{L}_{p_i}(\{\mathbf{x}_i^t\}, \{\mathbf{y}_i^t\}, \{\lambda_{i,l}^t\}, \{\boldsymbol{\theta}_{i,j}^{t-1}\})|| \\
=&||\nabla_{\boldsymbol{\theta}_{i,j}} L_{p_i}(\{\mathbf{x}_i^t\}, \{\mathbf{y}_i^t\}, \{\lambda_{i,l}^t\}, \{\boldsymbol{\theta}_{i,j}^t\}) - \nabla_{\boldsymbol{\theta}_{i,j}} L_{p_i}(\{\mathbf{x}_i^t\}, \{\mathbf{y}_i^t\}, \{\lambda_{i,l}^t\}, \{\boldsymbol{\theta}_{i,j}^{t-1}\}) - c_2^{t-1}(\boldsymbol{\theta}_{i,j}^t - \boldsymbol{\theta}_{i,j}^{t-1})|| \\
\leq&(L + c_2^{t-1})||\boldsymbol{\theta}_{i,j}^t - \boldsymbol{\theta}_{i,j}^{t-1}|| \\
\leq&L_2'||\boldsymbol{\theta}_{i,j}^t - \boldsymbol{\theta}_{i,j}^{t-1}||.
\end{aligned} \tag{S59}$$

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

$$\begin{aligned}
&\sum_{j=1}^{N} \langle \nabla_{\boldsymbol{\theta}_{i,j}} \widetilde{L}_{p_i}(\{\mathbf{x}_i^t\}, \{\mathbf{y}_i^t\}, \{\lambda_{i,l}^t\}, \{\boldsymbol{\theta}_{i,j}^t\}) - \nabla_{\boldsymbol{\theta}_{i,j}} \widetilde{L}_{p_i}(\{\mathbf{x}_i^t\}, \{\mathbf{y}_i^t\}, \{\lambda_{i,l}^t\}, \{\boldsymbol{\theta}_{i,j}^{t-1}\}), \boldsymbol{\theta}_{i,j}^{t} - \boldsymbol{\theta}_{i,j}^{t-1} \rangle \\
\leq&\sum_{j=1}^{N} (-\frac{1}{L_2' + c_2^{t-1}} ||\nabla_{\boldsymbol{\theta}_{i,j}} \widetilde{L}_{p_i}(\{\mathbf{x}_i^t\}, \{\mathbf{y}_i^t\}, \{\lambda_{i,l}^t\}, \{\boldsymbol{\theta}_{i,j}^t\}) - \nabla_{\boldsymbol{\theta}_{i,j}} \widetilde{L}_{p_i}(\{\mathbf{x}_i^t\}, \{\mathbf{y}_i^t\}, \{\lambda_{i,l}^t\}, \{\boldsymbol{\theta}_{i,j}^{t-1}\})||^2.
\end{aligned} \tag{S60}$$

Combining Eq.(S52), Eq.(S53), Eq.(S54),, Eq.(S55), Eq.(S56), Eq.(S57), Eq.(S58), Eq.(S59) and Eq.(S60), let $a_6 = \eta_\theta$, $a_5 = a_4$, $\frac{\eta_\theta}{2} \le \frac{1}{L_2' + c_2^0}$, we have

$$\begin{aligned}
&\mathbb{E}[L_{p_i}(\{\bar{\mathbf{x}}^{t+1}\}, \{\bar{\mathbf{y}}^{t+1}\}, \{\lambda_{i,l}^{t+1}\}, \{\boldsymbol{\theta}_{i,j}^{t+1}\}) - L_{p_i}(\{\bar{\mathbf{x}}^{t+1}\}, \{\bar{\mathbf{y}}^{t+1}\}, \{\lambda_{i,l}^{t+1}\}, \{\boldsymbol{\theta}_{i,j}^{t}\})] \\
\le& \frac{N^2L^2}{2a_4}(\mathbb{E}[||\bar{\mathbf{x}}^{t+1} - \bar{\mathbf{x}}^t||^2] + \mathbb{E}[||\bar{\mathbf{y}}^{t+1} - \bar{\mathbf{y}}^t||^2]) + \frac{NL^2}{2a_4}(\mathbb{E}[\sum_{i=1}^N ||\bar{\mathbf{x}}^t - \mathbf{x}_i^t||^2] + \mathbb{E}[\sum_{i=1}^N ||\bar{\mathbf{y}}^t - \mathbf{y}_i^t||^2]) \\
&+ (a_4 - \frac{c_2^{t-1} - c_2^t}{2} + \frac{1}{2\eta_\theta})\mathbb{E}[\sum_{j=1}^N ||\boldsymbol{\theta}_{i,j}^{t+1} - \boldsymbol{\theta}_{i,j}^t||^2] + \frac{c_2^{t-1}}{2}\mathbb{E}[\sum_{j=1}^N (||\boldsymbol{\theta}_{i,j}^{t+1}||^2 - ||\boldsymbol{\theta}_{i,j}^t||^2]) \\
&+ \frac{1}{2\eta_\theta}\mathbb{E}[\sum_{j=1}^N ||\boldsymbol{\theta}_{i,j}^t - \boldsymbol{\theta}_{i,j}^{t-1}||^2],
\end{aligned} \tag{S61}$$

and

$$\begin{aligned}
&\mathbb{E}[L_{p_i}(\{\bar{\mathbf{x}}^{t+1}\}, \{\bar{\mathbf{y}}^{t+1}\}, \{\lambda_{i,l}^{t+1}\}, \{\boldsymbol{\theta}_{i,j}^{t+1}\}) - L_{p_i}(\{\bar{\mathbf{x}}^{t+1}\}, \{\bar{\mathbf{y}}^{t+1}\}, \{\lambda_{i,l}^{t+1}\}, \{\boldsymbol{\theta}_{i,j}^{t}\})] \\
\le& \frac{N^3L^2}{2a_4}(\mathbb{E}[||\bar{\mathbf{x}}^{t+1} - \bar{\mathbf{x}}^t||^2] + \mathbb{E}[||\bar{\mathbf{y}}^{t+1} - \bar{\mathbf{y}}^t||^2]) + \frac{N^2L^2}{2a_4}(\mathbb{E}[\sum_{i=1}^N ||\bar{\mathbf{x}}^t - \mathbf{x}_i^t||^2] + \mathbb{E}[\sum_{i=1}^N ||\bar{\mathbf{y}}^t - \mathbf{y}_i^t||^2]) \\
&+ (a_4 - \frac{c_2^{t-1} - c_2^t}{2} + \frac{1}{2\eta_\theta})\mathbb{E}[\sum_{i=1}^N \sum_{j=1}^N ||\boldsymbol{\theta}_{i,j}^{t+1} - \boldsymbol{\theta}_{i,j}^t||^2] + \frac{c_2^{t-1}}{2}\mathbb{E}[\sum_{i=1}^N \sum_{j=1}^N (||\boldsymbol{\theta}_{i,j}^{t+1}||^2 - ||\boldsymbol{\theta}_{i,j}^t||^2]) \\
&+ \frac{1}{2\eta_\theta}\mathbb{E}[\sum_{i=1}^N \sum_{j=1}^N ||\boldsymbol{\theta}_{i,j}^t - \boldsymbol{\theta}_{i,j}^{t-1}||^2].
\end{aligned} \tag{S62}$$

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.(S47), we have

$$\begin{aligned}
&\mathbb{E}[L_p(\{\bar{\mathbf{x}}^{t+1}\}, \{\bar{\mathbf{y}}^{t+1}\}, \{\lambda_{i,l}^{t+1}\}, \{\boldsymbol{\theta}_{i,j}^{t+1}\}) - L_p(\{\bar{\mathbf{x}}^{t+1}\}, \{\bar{\mathbf{y}}^{t+1}\}, \{\lambda_{i,l}^{t+1}\}, \{\boldsymbol{\theta}_{i,j}^{t}\})] \\
\le& \frac{N^2L^2}{2a_4}\sum_{i=1}^N \mathbb{E}[||\mathbf{x}_i^{t+1} - \bar{\mathbf{x}}^t||^2] + \frac{N^2L^2\eta_y^2 2}{2a_4}\mathbb{E}[\sum_{i=1}^N ||\nabla_{\mathbf{y}_i}\tilde{L}_{p_i}(\{\mathbf{x}_i^{\hat{t}_i}\}, \{\mathbf{y}_i^{\hat{t}_i}\}, \{\lambda_{i,l}^{\hat{t}_i}\}, \{\boldsymbol{\theta}_{i,j}^{\hat{t}_i}\}; \xi)||^2] \\
&+ \frac{N^2L^2}{2a_4}(\mathbb{E}[\sum_{i=1}^N ||\bar{\mathbf{x}}^t - \mathbf{x}_i^t||^2] + \mathbb{E}[\sum_{i=1}^N ||\bar{\mathbf{y}}^t - \mathbf{y}_i^t||^2]) + (a_4 - \frac{c_2^{t-1} - c_2^t}{2} + \frac{1}{2\eta_\theta})\mathbb{E}[\sum_{i=1}^N \sum_{j=1}^N ||\boldsymbol{\theta}_{i,j}^{t+1} - \boldsymbol{\theta}_{i,j}^t||^2] \\
&+ \frac{c_2^{t-1}}{2}\mathbb{E}[\sum_{i=1}^N \sum_{j=1}^N (||\boldsymbol{\theta}_{i,j}^{t+1}||^2 - ||\boldsymbol{\theta}_{i,j}^t||^2]) + \frac{1}{2\eta_\theta}\sum_{i=1}^N \mathbb{E}[\sum_{j=1}^N ||\boldsymbol{\theta}_{i,j}^t - \boldsymbol{\theta}_{i,j}^{t-1}||^2].
\end{aligned} \tag{S63}$$

Combining Eq.(S12), Eq.(S27), Eq.(S48), and Eq.(S63), setting $a_1 = \frac{1}{2\eta_\lambda}$, $a_4 = \frac{1}{2\eta_\theta}$, then lemma 3 can be proved.

1.4.4 Lemma 4 (Iterates Contraction)

The following contraction properties of iterates hold:

$$\mathbb{E}[\sum_{i=1}^N ||\mathbf{x}_i^{t+1} - \bar{\mathbf{x}}^{t+1}||^2] \le \rho\mathbb{E}[\sum_{i=1}^N ||\mathbf{x}_i^t - \bar{\mathbf{x}}^t||^2] + \frac{(\eta_x)^2(6\sigma_1^2 + 3\varsigma_1^2)}{1-\rho}, \tag{S64}$$

$$\mathbb{E}[\sum_{i=1}^N ||\mathbf{y}^{t+1} - \bar{\mathbf{y}}^{t+1}||^2] \le \rho\mathbb{E}[\sum_{i=1}^N ||\mathbf{y}_i^t - \bar{\mathbf{y}}^t||^2 + \frac{\eta_y^2 2(6\sigma_2^2 + 3\varsigma_2^2)}{1-\rho}. \tag{S65}$$

Proof:

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

$$\begin{aligned}\left\|\mathbf{A}-\frac{1}{N}\mathbf{e}\mathbf{e}^\top\mathbf{A}\right\|_F^2 &= \|\mathbf{A}\|_F^2 - 2\left\langle \mathbf{A}, \frac{1}{N}\mathbf{e}\mathbf{e}^\top\mathbf{A}\right\rangle + \left\|\frac{1}{N}\mathbf{e}\mathbf{e}^\top\mathbf{A}\right\|_F^2 \\ &= \|\mathbf{A}\|_F^2 - 2\left\|\frac{1}{N}\mathbf{e}\mathbf{e}^\top\mathbf{A}\right\|_F^2 + \left\|\frac{1}{N}\mathbf{e}\mathbf{e}^\top\mathbf{A}\right\|_F^2 = \|\mathbf{A}\|_F^2 - \left\|\frac{1}{N}\mathbf{e}\mathbf{e}^\top\mathbf{A}\right\|_F^2.\end{aligned} \tag{S66}$$

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

$$\left\|\left(\mathbf{W}^t - \frac{1}{N}\mathbf{e}\mathbf{e}^\top\right)(\mathbf{x}^t - \frac{1}{N}\mathbf{e}\mathbf{e}^\top\mathbf{x}^t)\right\|_F^2 \le \left\|\left(\mathbf{W}^t - \frac{1}{N}\mathbf{e}\mathbf{e}^\top\right)\right\|_2^2\left\|\mathbf{x}^t - \frac{1}{N}\mathbf{e}\mathbf{e}^\top\mathbf{x}^t\right\|_F^2 = \rho^2\left\|\mathbf{x}^t - \frac{1}{N}\mathbf{e}\mathbf{e}^\top\mathbf{x}^t\right\|_F^2, \tag{S67}$$

$$\left\|\left(\mathbf{W}^t - \frac{1}{N}\mathbf{e}\mathbf{e}^\top\right)(\mathbf{y}^t - \frac{1}{N}\mathbf{e}\mathbf{e}^\top\mathbf{y}^t)\right\|_F^2 \le \left\|\left(\mathbf{W}^t - \frac{1}{N}\mathbf{e}\mathbf{e}^\top\right)\right\|_2^2\left\|\mathbf{y}^t - \frac{1}{N}\mathbf{e}\mathbf{e}^\top\mathbf{y}^t\right\|_F^2 = \rho^2\left\|\mathbf{y}^t - \frac{1}{N}\mathbf{e}\mathbf{e}^\top\mathbf{y}^t\right\|_F^2. \tag{S68}$$

Further, we have

$$\begin{aligned}&\left\|\mathbf{x}^{t+1}-\mathbf{x}^t\right\|_F^2 \\ =&\left\|\mathbf{x}^{t+1}-\mathbf{d}^t+\mathbf{d}^t-\mathbf{x}^t\right\|_F^2 \\ \le& 2\left\|\mathbf{x}^{t+1}-\mathbf{d}^t\right\|_F^2 + 2\left\|\mathbf{W}^t\mathbf{x}^t - \mathbf{W}^t\frac{1}{N}\mathbf{e}\mathbf{e}^\top\mathbf{x}^t + \mathbf{W}^t\frac{1}{N}\mathbf{e}\mathbf{e}^\top\mathbf{x}^t - \mathbf{x}^t\right\|_F^2 \\ =& 2\left\|\mathbf{x}^{t+1}-\mathbf{d}^t\right\|_F^2 + 2\left\|(\mathbf{I}-\mathbf{W}^t)(\mathbf{x}^t - \frac{1}{N}\mathbf{e}\mathbf{e}^\top\mathbf{x}^t)\right\|_F^2 \\ \le& 2\left\|\mathbf{x}^{t+1}-\mathbf{d}^t\right\|_F^2 + 8\left\|\mathbf{x}^t - \frac{1}{N}\mathbf{e}\mathbf{e}^\top\mathbf{x}^t\right\|_F^2,\end{aligned} \tag{S69}$$

and

$$\begin{aligned}&\mathbb{E}[||\mathbf{y}^{t+1}-\mathbf{y}^t||_F^2] \\ =&\mathbb{E}[||\mathbf{W}^t\mathbf{y}^t - \mathbf{y}^t - \tilde{\eta}_y^t\nabla_{\mathbf{y}}\tilde{L}_p(\{\mathbf{x}_i^{\hat{t}_i}\},\{\mathbf{y}_i^{\hat{t}_i}\},\{\lambda_{i,l}^{\hat{t}_i}\},\{\boldsymbol{\theta}_{i,j}^{\hat{t}_i}\};\xi)||_F^2] \\ =&\mathbb{E}[||\mathbf{W}^t(\mathbf{y}^t - \frac{1}{N}\mathbf{e}\mathbf{e}^\top\mathbf{y}^t) + \frac{1}{N}\mathbf{e}\mathbf{e}^\top\mathbf{y}^t - \mathbf{y}^t - \tilde{\eta}_y^t\nabla_{\mathbf{y}}\tilde{L}_p(\{\mathbf{x}_i^{\hat{t}_i}\},\{\mathbf{y}_i^{\hat{t}_i}\},\{\lambda_{i,l}^{\hat{t}_i}\},\{\boldsymbol{\theta}_{i,j}^{\hat{t}_i}\};\xi)||_F^2] \\ \le&\rho^2(1+\delta)||\mathbf{y}^t - \frac{1}{N}\mathbf{e}\mathbf{e}^\top\mathbf{y}^t||_F^2 + (1+\frac{1}{\delta})\mathbb{E}[||\frac{1}{N}\mathbf{e}\mathbf{e}^\top\mathbf{y}^t - \mathbf{y}^t - \tilde{\eta}_y^t\nabla_{\mathbf{y}}\tilde{L}_p(\{\mathbf{x}_i^{\hat{t}_i}\},\{\mathbf{y}_i^{\hat{t}_i}\},\{\lambda_{i,l}^{\hat{t}_i}\},\{\boldsymbol{\theta}_{i,j}^{\hat{t}_i}\};\xi)||_F^2] \\ \le&\rho^2(1+\delta)||\mathbf{y}^t - \frac{1}{N}\mathbf{e}\mathbf{e}^\top\mathbf{y}^t||_F^2 + 2(1+\frac{1}{\delta})||\frac{1}{N}\mathbf{e}\mathbf{e}^\top\mathbf{y}^t - \mathbf{y}^t||_F^2 + 2\eta_y^2 2(1+\frac{1}{\delta})||\nabla_{\mathbf{y}}\tilde{L}_p(\{\mathbf{x}_i^{\hat{t}_i}\},\{\mathbf{y}_i^{\hat{t}_i}\},\{\lambda_{i,l}^{\hat{t}_i}\},\{\boldsymbol{\theta}_{i,j}^{\hat{t}_i}\};\xi)||_F^2 \\ =&(\rho+\frac{2}{1-\rho})||\mathbf{y}^t - \frac{1}{N}\mathbf{e}\mathbf{e}^\top\mathbf{y}^t||_F^2 + \frac{2\eta_y^2 2}{1-\rho}||\nabla_{\mathbf{y}}\tilde{L}_p(\{\mathbf{x}_i^{\hat{t}_i}\},\{\mathbf{y}_i^{\hat{t}_i}\},\{\lambda_{i,l}^{\hat{t}_i}\},\{\boldsymbol{\theta}_{i,j}^{\hat{t}_i}\};\xi)||_F^2,\end{aligned} \tag{S70}$$

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}_{i,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}\};\xi))-\frac{1}{N}\mathbf{e}\mathbf{e}^\top\operatorname{prox}_R(\mathbf{W}^t\mathbf{x}^t-\tilde{\eta}_{i,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}\};\xi))||_F^2\\
&\overset{(a)}{=}||\operatorname{prox}^R(\mathbf{W}^t\mathbf{x}^t-\tilde{\eta}_{i,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}\};\xi))-\operatorname{prox}^R(\frac{1}{N}\mathbf{e}\mathbf{e}^\top\mathbf{x}^t-\tilde{\eta}_{i,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}\};\xi))||_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}_{i,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}\};\xi))\\
&-\operatorname{prox}^R(\mathbf{W}^t\mathbf{x}^t-\tilde{\eta}_{i,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}\};\xi))||_F^2\\
&\overset{(b)}{\leq}||\mathbf{W}^t\mathbf{x}^t-\frac{1}{N}\mathbf{e}\mathbf{e}^\top\mathbf{x}^t-\tilde{\eta}_{i,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}\};\xi)-\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}\};\xi))||_F^2\\
&=||\mathbf{W}^t\mathbf{x}^t-\frac{1}{N}\mathbf{e}\mathbf{e}^\top\mathbf{x}^t||_F^2+||\tilde{\eta}_{i,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}\};\xi)-\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}\};\xi))||_F^2\\
&-2\langle\mathbf{W}^t\mathbf{x}^t-\frac{1}{N}\mathbf{e}\mathbf{e}^\top\mathbf{x}^t,\tilde{\eta}_{i,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}\};\xi)-\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}\};\xi))\rangle\\
&\leq||\mathbf{W}^t\mathbf{x}^t-\frac{1}{N}\mathbf{e}\mathbf{e}^\top\mathbf{x}^t||_F^2+||\tilde{\eta}_{i,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}\};\xi)-\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}\};\xi))||_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}_{i,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}\};\xi)-\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}\};\xi))||_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}_{i,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}\};\xi)-\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}\};\xi)||_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}_{i,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}\};\xi)-\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}\};\xi)||_F^2,
\end{aligned}
\tag{S71}
$$

where $\bar{\nabla}_{\mathbf{x}}\tilde{L}_p(\{\mathbf{x}_i^{\hat{t}_i}\},\{\mathbf{y}_i^{\hat{t}_i}\},\{\lambda_{i,l}^{\hat{t}_i}\},\{\boldsymbol{\theta}_{i,j}^{\hat{t}_i}\};\xi)=\frac{1}{N}\mathbf{e}\mathbf{e}^\top\nabla_{\mathbf{x}}\tilde{L}_p(\{\mathbf{x}_i^{\hat{t}_i}\},\{\mathbf{y}_i^{\hat{t}_i}\},\{\lambda_{i,l}^{\hat{t}_i}\},\{\boldsymbol{\theta}_{i,j}^{\hat{t}_i}\};\xi)$. (a) utilizes Eq.(S66). (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.(S67).

According to Assumption 5.1, we have

$$
\begin{aligned}
&\mathbb{E}[||\nabla_{\mathbf{x}}\tilde{L}_p(\{\mathbf{x}_i^{\hat{t}_i}\},\{\mathbf{y}_i^{\hat{t}_i}\},\{\lambda_{i,l}^{\hat{t}_i}\},\{\boldsymbol{\theta}_{i,j}^{\hat{t}_i}\};\xi)-\bar{\nabla}_{\mathbf{x}}\tilde{L}_p(\{\mathbf{x}_i^{\hat{t}_i}\},\{\mathbf{y}_i^{\hat{t}_i}\},\{\lambda_{i,l}^{\hat{t}_i}\},\{\boldsymbol{\theta}_{i,j}^{\hat{t}_i}\};\xi)||_F^2]\\
=&\mathbb{E}[||\nabla_{\mathbf{x}}\tilde{L}_p(\{\mathbf{x}_i^{\hat{t}_i}\},\{\mathbf{y}_i^{\hat{t}_i}\},\{\lambda_{i,l}^{\hat{t}_i}\},\{\boldsymbol{\theta}_{i,j}^{\hat{t}_i}\};\xi)-\nabla_{\mathbf{x}}\tilde{L}_p(\{\mathbf{x}_i^{\hat{t}_i}\},\{\mathbf{y}_i^{\hat{t}_i}\},\{\lambda_{i,l}^{\hat{t}_i}\},\{\boldsymbol{\theta}_{i,j}^{\hat{t}_i}\})\\
&+\nabla_{\mathbf{x}}\tilde{L}_p(\{\mathbf{x}_i^{\hat{t}_i}\},\{\mathbf{y}_i^{\hat{t}_i}\},\{\lambda_{i,l}^{\hat{t}_i}\},\{\boldsymbol{\theta}_{i,j}^{\hat{t}_i}\})-\bar{\nabla}_{\mathbf{x}}\tilde{L}_p(\{\mathbf{x}_i^{\hat{t}_i}\},\{\mathbf{y}_i^{\hat{t}_i}\},\{\lambda_{i,l}^{\hat{t}_i}\},\{\boldsymbol{\theta}_{i,j}^{\hat{t}_i}\})\\
&+\bar{\nabla}_{\mathbf{x}}\tilde{L}_p(\{\mathbf{x}_i^{\hat{t}_i}\},\{\mathbf{y}_i^{\hat{t}_i}\},\{\lambda_{i,l}^{\hat{t}_i}\},\{\boldsymbol{\theta}_{i,j}^{\hat{t}_i}\})-\bar{\nabla}_{\mathbf{x}}\tilde{L}_p(\{\mathbf{x}_i^{\hat{t}_i}\},\{\mathbf{y}_i^{\hat{t}_i}\},\{\lambda_{i,l}^{\hat{t}_i}\},\{\boldsymbol{\theta}_{i,j}^{\hat{t}_i}\};\xi)||_F^2]\\
\leq&6N\sigma_1^2+3N\varsigma_1^2.
\end{aligned}
\tag{S72}
$$

Plugging Eq.(S72) into Eq.(S71), then Eq.(S64) 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}_{i,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}\};\xi))-\frac{1}{N}\mathbf{e}\mathbf{e}^\top(\mathbf{W}^t\mathbf{y}^t-\tilde{\eta}_{i,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}\};\xi))||_F^2\\
&\overset{(a)}{=}||(\mathbf{W}^t\mathbf{y}^t-\tilde{\eta}_{i,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}\};\xi))-(\frac{1}{N}\mathbf{e}\mathbf{e}^\top\mathbf{y}^t-\tilde{\eta}_{i,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}\};\xi))||_F^2\\
&-||\frac{1}{N}\mathbf{e}\mathbf{e}^\top((\frac{1}{N}\mathbf{e}\mathbf{e}^\top\mathbf{y}^t-\tilde{\eta}_{i,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}\};\xi))-(\mathbf{W}^t\mathbf{y}^t-\tilde{\eta}_{i,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}\};\xi)))||_F^2\\
&\le||\mathbf{W}^t\mathbf{y}^t-\frac{1}{N}\mathbf{e}\mathbf{e}^\top\mathbf{y}^t-\tilde{\eta}_{i,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}\};\xi)-\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}\};\xi))||_F^2\\
&=||\mathbf{W}^t\mathbf{y}^t-\frac{1}{N}\mathbf{e}\mathbf{e}^\top\mathbf{y}^t||_F^2+||\tilde{\eta}_{i,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}\};\xi)-\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}\};\xi))||_F^2\\
&-2\left\langle\mathbf{W}^t\mathbf{y}^t-\frac{1}{N}\mathbf{e}\mathbf{e}^\top\mathbf{y}^t,\tilde{\eta}_{i,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}\};\xi)-\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}\};\xi))\right\rangle\\
&\le||\mathbf{W}^t\mathbf{y}^t-\frac{1}{N}\mathbf{e}\mathbf{e}^\top\mathbf{y}^t||_F^2+||\tilde{\eta}_{i,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}\};\xi)-\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}\};\xi))||_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}_{i,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}\};\xi)-\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}\};\xi))||_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}_{i,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}\};\xi)-\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}\};\xi)||_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}_{i,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}\};\xi)-\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}\};\xi)||_F^2,
\end{aligned}
\tag{S73}
$$

where (a) utilizes Eq.(S66). (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.(S68).

According to Assumption 5.1, we have

$$
\begin{aligned}
&\mathbb{E}[||\nabla_{\mathbf{y}}\tilde{L}_p(\{\mathbf{x}_i^{\hat{t}_i}\},\{\mathbf{y}_i^{\hat{t}_i}\},\{\lambda_{i,l}^{\hat{t}_i}\},\{\boldsymbol{\theta}_{i,j}^{\hat{t}_i}\};\xi)-\bar{\nabla}_{\mathbf{y}}\tilde{L}_p(\{\mathbf{x}_i^{\hat{t}_i}\},\{\mathbf{y}_i^{\hat{t}_i}\},\{\lambda_{i,l}^{\hat{t}_i}\},\{\boldsymbol{\theta}_{i,j}^{\hat{t}_i}\};\xi)||_F^2]\\
=&\mathbb{E}[||\nabla_{\mathbf{y}}\tilde{L}_p(\{\mathbf{x}_i^{\hat{t}_i}\},\{\mathbf{y}_i^{\hat{t}_i}\},\{\lambda_{i,l}^{\hat{t}_i}\},\{\boldsymbol{\theta}_{i,j}^{\hat{t}_i}\};\xi)-\nabla_{\mathbf{y}}\tilde{L}_p(\{\mathbf{x}_i^{\hat{t}_i}\},\{\mathbf{y}_i^{\hat{t}_i}\},\{\lambda_{i,l}^{\hat{t}_i}\},\{\boldsymbol{\theta}_{i,j}^{\hat{t}_i}\})\\
&+\nabla_{\mathbf{y}}\tilde{L}_p(\{\mathbf{x}_i^{\hat{t}_i}\},\{\mathbf{y}_i^{\hat{t}_i}\},\{\lambda_{i,l}^{\hat{t}_i}\},\{\boldsymbol{\theta}_{i,j}^{\hat{t}_i}\})-\bar{\nabla}_{\mathbf{y}}\tilde{L}_p(\{\mathbf{x}_i^{\hat{t}_i}\},\{\mathbf{y}_i^{\hat{t}_i}\},\{\lambda_{i,l}^{\hat{t}_i}\},\{\boldsymbol{\theta}_{i,j}^{\hat{t}_i}\})\\
&+\bar{\nabla}_{\mathbf{y}}\tilde{L}_p(\{\mathbf{x}_i^{\hat{t}_i}\},\{\mathbf{y}_i^{\hat{t}_i}\},\{\lambda_{i,l}^{\hat{t}_i}\},\{\boldsymbol{\theta}_{i,j}^{\hat{t}_i}\})-\bar{\nabla}_{\mathbf{y}}\tilde{L}_p(\{\mathbf{x}_i^{\hat{t}_i}\},\{\mathbf{y}_i^{\hat{t}_i}\},\{\lambda_{i,l}^{\hat{t}_i}\},\{\boldsymbol{\theta}_{i,j}^{\hat{t}_i}\};\xi)||_F^2]\\
\le&6N\sigma_2^2+3N\varsigma_2^2.
\end{aligned}
\tag{S74}
$$

Plugging Eq.(S74) into Eq.(S73), then Eq.(S65) can be proved.

1.4.5 Lemma 5

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

$$
S_1^{t+1}=\frac{4}{(\eta_\lambda)^2c_1^{t+1}}\sum_{i=1}^{N}\sum_{l=1}^{|\mathcal{P}_i^t|}||\lambda_{i,l}^{t+1}-\lambda_{i,l}^t||^2-\frac{4}{\eta_\lambda}\left(\frac{c_1^{t-1}}{c_1^t}-1\right)\sum_{i=1}^{N}\sum_{l=1}^{|\mathcal{P}_i^t|}||\lambda_{i,l}^{t+1}||^2,
\tag{S75}
$$

$$
S_2^{t+1}=\frac{4}{(\eta_\theta)^2c_2^{t+1}}\sum_{i=1}^{N}\sum_{j=1}^{N}||\boldsymbol{\theta}_{i,j}^{t+1}-\boldsymbol{\theta}_{i,j}^t||^2-\frac{4}{\eta_\theta}\left(\frac{c_2^{t-1}}{c_2^t}-1\right)\sum_{i=1}^{N}\sum_{j=1}^{N}||\boldsymbol{\theta}_{i,j}^{t+1}||^2,
\tag{S76}
$$

$$
\begin{aligned}
F^{t+1} = & L_p(\{\bar{\mathbf{x}}^{t+1}\},\{\bar{\mathbf{y}}^{t+1}\},\{\lambda_{i,l}^{t+1}\},\{\boldsymbol{\theta}_{i,j}^{t+1}\}) + S_1^{t+1} + S_2^{t+1} + \gamma_1^t \sum_{i=1}^{N} ||\mathbf{x}_i^{t+1} - \bar{\mathbf{x}}^{t+1}||^2 + \gamma_2^t \sum_{i=1}^{N} ||\mathbf{y}_i^{t+1} - \bar{\mathbf{y}}^{t+1}||^2 \\
& - \frac{6}{\eta_\lambda} \sum_{i=1}^{N} \sum_{l=1}^{|\mathcal{P}_i^t|} ||\lambda_{i,l}^{t+1} - \lambda_{i,l}^t||^2 - \frac{c_1^t}{2} \sum_{i=1}^{N} \sum_{l=1}^{|\mathcal{P}_i^t|} ||\lambda_{i,l}^{t+1}||^2 - \frac{6}{\eta_\theta} \sum_{i=1}^{N} \sum_{j=1}^{N} ||\boldsymbol{\theta}_{i,j}^{t+1} - \boldsymbol{\theta}_{i,j}^t||^2 - \frac{c_2^t}{2} \sum_{i=1}^{N} \sum_{j=1}^{N} ||\boldsymbol{\theta}_{i,j}^{t+1}||^2,
\end{aligned}
\tag{S77}
$$

$\forall t \geq T_1$, we have

$$
\begin{aligned}
& (\frac{N}{2\eta_x} - NL - \frac{2\eta_y N^2 L^2}{\beta} - MNL^2\eta_\lambda - N^2L^2\eta_{\boldsymbol{\theta}})\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{64NML^2}{\eta_\lambda (c_1^t)^2} + \frac{64N^2L^2}{\eta_\theta (c_2^t)^2}))\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{N}{2\eta_x} - N^2 L - \frac{2\eta_y N^2 L^2}{\beta} - MNL^2\eta_\lambda - N^2L^2\eta_{\boldsymbol{\theta}} \\
& \quad - (16\tau k_1 N^2 L + \frac{32\eta_y N^2 L^2 \tau k_1}{\beta} + \frac{256NML^2}{\eta_\lambda (c_1^t)^2} + \frac{256N^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{2\eta_y N^2 L^2}{\beta} - MNL^2\eta_\lambda - N^2L^2\eta_{\boldsymbol{\theta}} \\
& \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{32NML^2}{\eta_\lambda (c_1^t)^2} + \frac{32N^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} - MNL^2\eta_y^2\eta_\lambda - N^2L^2\eta_y^2\eta_{\boldsymbol{\theta}} \\
& \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{32NML^2}{\eta_\lambda (c_1^t)^2} + \frac{32N^2L^2}{\eta_\theta (c_2^t)^2}))\mathbb{E}[\sum_{i=1}^{N} ||\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}\};\xi)||^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}\left(\frac{c_1^{t-2}}{c_1^{t-1}} - \frac{c_1^{t-1}}{c_1^t}\right)\mathbb{E}[\sum_{i=1}^{N} \sum_{l=1}^{|\mathcal{P}_i^t|} ||\lambda_{i,l}^t||^2] + \frac{4}{\eta_\theta}\left(\frac{c_2^{t-2}}{c_2^{t-1}} - \frac{c_2^{t-1}}{c_2^t}\right)\mathbb{E}[\sum_{i=1}^{N} \sum_{j=1}^{N} \left\|\boldsymbol{\theta}_{i,j}^t\right\|^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] \\
& + \frac{\gamma_1^t \eta_x^2 (6N\sigma_1^2 + 3N\varsigma_1^2)}{1-\rho} + \frac{\gamma_2^t \eta_y^2 (6N\sigma_2^2 + 3N\varsigma_2^2)}{1-\rho} + \frac{2\eta_y N^2 \sigma_2^2}{\beta}.
\end{aligned}
\tag{S78}
$$

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

$$
\langle \lambda_{i,l}^{t+1} - \lambda_{i,l}^t - \eta_\lambda \nabla_{\lambda_{i,l}} \tilde{L}_{p\_i}(\{\mathbf{x}_i^{t+1}\},\{\mathbf{y}_i^{t+1}\},\{\lambda_{i,l}^t\},\{\boldsymbol{\theta}_{i,j}^t\}), \lambda_{i,l}^{t+1} - \lambda_{i,l}^t \rangle = 0. \tag{S79}
$$

Similar to Eq.(S79), in the t^{th} iteration, we have

$$
\langle \lambda_{i,l}^t - \lambda_{i,l}^{t-1} - \eta_\lambda \nabla_{\lambda_{i,l}} \tilde{L}_{p\_i}(\{\mathbf{x}_i^t\},\{\mathbf{y}_i^t\},\{\lambda_{i,l}^{t-1}\},\{\boldsymbol{\theta}_{i,j}^{t-1}\}), \lambda_{i,l}^{t+1} - \lambda_{i,l}^t \rangle = 0. \tag{S80}
$$

Combining Eq.(S79) and Eq.(S80), it follows that

$$
\begin{aligned}
&\frac{1}{\eta_\lambda}\langle v_{1,i,l}^{t+1}, \lambda_{i,l}^{t+1}-\lambda_{i,l}^{t}\rangle \\
=&\frac{1}{\eta_\lambda}\langle \lambda_{i,l}^{t+1}-\lambda_{i,l}^{t}, \lambda_{i,l}^{t+1}-\lambda_{i,l}^{t}\rangle - \frac{1}{\eta_\lambda}\langle \lambda_{i,l}^{t}-\lambda_{i,l}^{t-1}, \lambda_{i,l}^{t+1}-\lambda_{i,l}^{t}\rangle \\
=&\langle \nabla_{\lambda_{i,l}}\tilde{L}_{p\_i}(\{\mathbf{x}_i^{t+1}\},\{\mathbf{y}_i^{t+1}\},\{\lambda_{i,l}^{t}\},\{\boldsymbol{\theta}_{i,j}^{t}\}), \lambda_{i,l}^{t+1}-\lambda_{i,l}^{t}\rangle - \langle \nabla_{\lambda_{i,l}}\tilde{L}_{p\_i}(\{\mathbf{x}_i^{t}\},\{\mathbf{y}_i^{t}\},\{\lambda_{i,l}^{t-1}\},\{\boldsymbol{\theta}_{i,j}^{t-1}\}), \lambda_{i,l}^{t+1}-\lambda_{i,l}^{t}\rangle \\
=&\langle \nabla_{\lambda_{i,l}}\tilde{L}_{p\_i}(\{\mathbf{x}_i^{t+1}\},\{\mathbf{y}_i^{t+1}\},\{\lambda_{i,l}^{t}\},\{\boldsymbol{\theta}_{i,j}^{t}\}) - \nabla_{\lambda_{i,l}}\tilde{L}_{p\_i}(\{\mathbf{x}_i^{t}\},\{\mathbf{y}_i^{t}\},\{\lambda_{i,l}^{t}\},\{\boldsymbol{\theta}_{i,j}^{t}\}), \lambda_{i,l}^{t+1}-\lambda_{i,l}^{t}\rangle \\
+&\langle \nabla_{\lambda_{i,l}}\tilde{L}_{p\_i}(\{\mathbf{x}_i^{t}\},\{\mathbf{y}_i^{t}\},\{\lambda_{i,l}^{t}\},\{\boldsymbol{\theta}_{i,j}^{t}\}) - \nabla_{\lambda_{i,l}}\tilde{L}_{p\_i}(\{\{\mathbf{x}_i^{t}\},\{\mathbf{y}_i^{t}\},\{\lambda_{i,l}^{t-1}\},\{\boldsymbol{\theta}_{i,j}^{t-1}\}), v_{1,i,l}^{t+1}\rangle \\
+&\langle \nabla_{\lambda_{i,l}}\tilde{L}_{p\_i}(\{\mathbf{x}_i^{t}\},\{\mathbf{y}_i^{t}\},\{\lambda_{i,l}^{t}\},\{\boldsymbol{\theta}_{i,j}^{t}\}) - \nabla_{\lambda_{i,l}}\tilde{L}_{p\_i}(\{\{\mathbf{x}_i^{t}\},\{\mathbf{y}_i^{t}\},\{\lambda_{i,l}^{t-1}\},\{\boldsymbol{\theta}_{i,j}^{t-1}\}), \lambda_{i,l}^{t}-\lambda_{i,l}^{t-1}\rangle.
\end{aligned}
\tag{S81}
$$

Since we have

$$
\frac{1}{\eta_\lambda}\left\langle v_{1,i,l}^{t+1}, \lambda_{i,l}^{t+1}-\lambda_{i,l}^{t}\right\rangle = \frac{1}{2\eta_\lambda}||\lambda_{i,l}^{t+1}-\lambda_{i,l}^{t}||^2 + \frac{1}{2\eta_\lambda}||v_{1,i,l}^{t+1}||^2 - \frac{1}{2\eta_\lambda}||\lambda_{i,l}^{t}-\lambda_{i,l}^{t-1}||^2,
\tag{S82}
$$

it follows from Eq. (S81) and Eq. (S82) that,

$$
\begin{aligned}
&\frac{1}{2\eta_\lambda}||\lambda_{i,l}^{t+1}-\lambda_{i,l}^{t}||^2 + \frac{1}{2\eta_\lambda}||v_{1,i,l}^{t+1}||^2 - \frac{1}{2\eta_\lambda}||\lambda_{i,l}^{t}-\lambda_{i,l}^{t-1}||^2 \\
\leq&\frac{L^2}{b_1^t}\sum_{i=1}^{N}(||\mathbf{x}_i^{t+1}-\mathbf{x}_i^{t}||^2 + ||\mathbf{y}_i^{t+1}-\mathbf{y}_i^{t}||^2) + b_1^t||\lambda_{i,l}^{t+1}-\lambda_{i,l}^{t}||^2 \\
+&\frac{c_1^{t-1}-c_1^t}{2}\left(||\lambda_{i,l}^{t+1}||^2 - ||\lambda_{i,l}^{t}||^2\right) - \frac{c_1^{t-1}-c_1^t}{2}||\lambda_{i,l}^{t+1}-\lambda_{i,l}^{t}||^2 \\
+&\frac{\eta_\lambda}{2}||\nabla_{\lambda_{i,l}}\tilde{L}_{p\_i}(\{\mathbf{x}_i^{t}\},\{\mathbf{y}_i^{t}\},\{\lambda_{i,l}^{t}\},\{\boldsymbol{\theta}_{i,j}^{t}\}) - \nabla_{\lambda_{i,l}}\tilde{L}_{p\_i}(\{\mathbf{x}_i^{t}\},\{\mathbf{y}_i^{t}\},\{\lambda_{i,l}^{t-1}\},\{\boldsymbol{\theta}_{i,j}^{t-1}\})||^2 + \frac{1}{2\eta_\lambda}||v_{1,i,l}^{t+1}||^2 \\
-&\frac{1}{L_1'+c_1^{t-1}}||\nabla_{\lambda_{i,l}}\tilde{L}_{p\_i}(\{\mathbf{x}_i^{t}\},\{\mathbf{y}_i^{t}\},\{\lambda_{i,l}^{t}\},\{\boldsymbol{\theta}_{i,j}^{t}\}) - \nabla_{\lambda_{i,l}}\tilde{L}_{p\_i}(\{\mathbf{x}_i^{t}\},\{\mathbf{y}_i^{t}\},\{\lambda_{i,l}^{t-1}\},\{\boldsymbol{\theta}_{i,j}^{t-1}\})||^2 \\
-&\frac{c_1^{t-1}L_1'}{L_1'+c_1^{t-1}}||\lambda_{i,l}^{t}-\lambda_{i,l}^{t-1}||^2,
\end{aligned}
\tag{S83}
$$

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.(S83) by $\frac{8}{\eta_\lambda c_1^t}$, we have

$$
\begin{aligned}
&\frac{4}{\eta_\lambda^2 c_1^t}||\lambda_{i,l}^{t+1}-\lambda_{i,l}^{t}||^2 - \frac{4}{\eta_\lambda}(\frac{c_1^{t-1}-c_1^t}{c_1^t})||\lambda_{i,l}^{t+1}||^2 \\
\leq&\frac{4}{\eta_\lambda^2 c_1^t}||\lambda_{i,l}^{t}-\lambda_{i,l}^{t-1}||^2 - \frac{4}{\eta_\lambda}(\frac{c_1^{t-1}-c_1^t}{c_1^t})||\lambda_{i,l}^{t}||^2 + \frac{8b_1^t}{\eta_\lambda c_1^t}||\lambda_{i,l}^{t+1}-\lambda_{i,l}^{t}||^2 - \frac{4}{\eta_\lambda}||\lambda_{i,l}^{t}-\lambda_{i,l}^{t-1}||^2 \\
+&\frac{8L^2}{\eta_\lambda c_1^t b_1^t}\sum_{i=1}^{N}(||\mathbf{x}_i^{t+1}-\mathbf{x}_i^{t}||^2 + ||\mathbf{y}_i^{t+1}-\mathbf{y}_i^{t}||^2).
\end{aligned}
\tag{S84}
$$

Setting $b_1^t = \frac{c_1^t}{4}$, using the definition of S_1^t, we have,

$$
\begin{aligned}
&S_1^{t+1} - S_1^t \\
\leq&\frac{4}{\eta_\lambda}\left(\frac{c_1^{t-2}}{c_1^{t-1}} - \frac{c_1^{t-1}}{c_1^t}\right)\sum_{i=1}^{N}\sum_{l=1}^{|\mathcal{P}_i^t|}||\lambda_{i,l}^{t}||^2 + (\frac{2}{\eta_\lambda} + \frac{4}{\eta_\lambda^2}(\frac{1}{c_1^{t+1}} - \frac{1}{c_1^t}))\sum_{i=1}^{N}\sum_{l=1}^{|\mathcal{P}_i^t|}||\lambda_{i,l}^{t+1}-\lambda_{i,l}^{t}||^2 \\
-&\frac{4}{\eta_\lambda}\sum_{i=1}^{N}\sum_{l=1}^{|\mathcal{P}_i^t|}||\lambda_{i,l}^{t}-\lambda_{i,l}^{t-1}||^2 + \frac{32NL^2}{\eta_\lambda(c_1^t)^2}\sum_{i=1}^{N}\sum_{l=1}^{|\mathcal{P}_i^t|}(||\mathbf{x}_i^{t+1}-\mathbf{x}_i^{t}||^2 + ||\mathbf{y}_i^{t+1}-\mathbf{y}_i^{t}||^2).
\end{aligned}
\tag{S85}
$$

Similarly to Eq.(S81), it follows that

$$\begin{aligned}
&\frac{1}{\eta_\theta}\langle v_{2,i,j}^{t+1}, \boldsymbol{\theta}_{i,j}^{t+1} - \boldsymbol{\theta}_{i,j}^{t}\rangle \\
=&\langle \nabla_{\boldsymbol{\theta}_{i,j}} \tilde{L}_{p_i}(\{\mathbf{x}_i^{t+1}\}, \{\mathbf{y}_i^{t+1}\}, \{\lambda_{i,l}^{t}\}, \{\boldsymbol{\theta}_{i,j}^{t}\}), \boldsymbol{\theta}_{i,j}^{t+1} - \boldsymbol{\theta}_{i,j}^{t}\rangle - \langle \nabla_{\boldsymbol{\theta}_{i,j}} \tilde{L}_{p_i}(\{\mathbf{x}_i^{t}\}, \{\mathbf{y}_i^{t}\}, \{\lambda_{i,l}^{t}\}, \{\boldsymbol{\theta}_{i,j}^{t-1}\}), \boldsymbol{\theta}_{i,j}^{t+1} - \boldsymbol{\theta}_{i,j}^{t}\rangle \\
=&\langle \nabla_{\boldsymbol{\theta}_{i,j}} \tilde{L}_{p_i}(\{\mathbf{x}_i^{t+1}\}, \{\mathbf{y}_i^{t+1}\}, \{\lambda_{i,l}^{t+1}\}, \{\boldsymbol{\theta}_{i,j}^{t}\}) - \nabla_{\boldsymbol{\theta}_{i,j}} \tilde{L}_{p_i}(\{\mathbf{x}_i^{t}\}, \{\mathbf{y}_i^{t}\}, \{\lambda_{i,l}^{t}\}, \{\boldsymbol{\theta}_{i,j}^{t}\}), \boldsymbol{\theta}_{i,j}^{t+1} - \boldsymbol{\theta}_{i,j}^{t}\rangle \\
+&\langle \nabla_{\boldsymbol{\theta}_{i,j}} \tilde{L}_{p_i}(\{\mathbf{x}_i^{t}\}, \{\mathbf{y}_i^{t}\}, \{\lambda_{i,l}^{t}\}, \{\boldsymbol{\theta}_{i,j}^{t}\}) - \nabla_{\boldsymbol{\theta}_{i,j}} \tilde{L}_{p_i}(\{\mathbf{x}_i^{t}\}, \{\mathbf{y}_i^{t}\}, \{\lambda_{i,l}^{t}\}, \{\boldsymbol{\theta}_{i,j}^{t-1}\}), v_{2,i,j}^{t+1}\rangle \\
+&\langle \nabla_{\boldsymbol{\theta}_{i,j}} \tilde{L}_{p_i}(\{\mathbf{x}_i^{t}\}, \{\mathbf{y}_i^{t}\}, \{\lambda_{i,l}^{t}\}, \{\boldsymbol{\theta}_{i,j}^{t}\}) - \nabla_{\boldsymbol{\theta}_{i,j}} \tilde{L}_{p_i}(\{\mathbf{x}_i^{t}\}, \{\mathbf{y}_i^{t}\}, \{\lambda_{i,l}^{t}\}, \{\boldsymbol{\theta}_{i,j}^{t-1}\}), \boldsymbol{\theta}_{i,j}^{t} - \boldsymbol{\theta}_{i,j}^{t-1}\rangle.
\end{aligned} \tag{S86}$$

Since we have

$$\frac{1}{\eta_\theta}\left\langle v_{2,i,j}^{t+1}, \boldsymbol{\theta}_{i,j}^{t+1} - \boldsymbol{\theta}_{i,j}^{t}\right\rangle = \frac{1}{2\eta_\theta}\left\|\boldsymbol{\theta}_{i,j}^{t+1} - \boldsymbol{\theta}_{i,j}^{t}\right\|^2 + \frac{1}{2\eta_\theta}\left\|v_{2,i,j}^{t+1}\right\|^2 - \frac{1}{2\eta_\theta}\left\|\boldsymbol{\theta}_{i,j}^{t} - \boldsymbol{\theta}_{i,j}^{t-1}\right\|^2, \tag{S87}$$

it follows from Eq. (S86) and Eq. (S87) that,

$$\begin{aligned}
&\frac{1}{2\eta_\theta}\left\|\boldsymbol{\theta}_{i,j}^{t+1} - \boldsymbol{\theta}_{i,j}^{t}\right\|^2 + \frac{1}{2\eta_\theta}\left\|v_{2,i,j}^{t+1}\right\|^2 - \frac{1}{2\eta_\theta}\left\|\boldsymbol{\theta}_{i,j}^{t} - \boldsymbol{\theta}_{i,j}^{t-1}\right\|^2 \\
\leq&\frac{L^2}{b_2^t}\sum_{i=1}^{N}(||\mathbf{x}_i^{t+1} - \mathbf{x}_i^{t}||^2 + ||\mathbf{y}_i^{t+1} - \mathbf{y}_i^{t}||^2) + b_2^t||\boldsymbol{\theta}_{i,j}^{t+1} - \boldsymbol{\theta}_{i,j}^{t}||^2 \\
+&\frac{c_2^{t-1} - c_2^t}{2}\left(\left\|\boldsymbol{\theta}_{i,j}^{t+1}\right\|^2 - \left\|\boldsymbol{\theta}_{i,j}^{t}\right\|^2\right) - \frac{c_2^{t-1} - c_2^t}{2}\left\|\boldsymbol{\theta}_{i,j}^{t+1} - \boldsymbol{\theta}_{i,j}^{t}\right\|^2 \\
+&\frac{\eta_\theta}{2}||\nabla_{\boldsymbol{\theta}_{i,j}} \tilde{L}_{p_i}(\{\mathbf{x}_i^{t}\}, \{\mathbf{y}_i^{t}\}, \{\lambda_{i,l}^{t}\}, \{\boldsymbol{\theta}_{i,j}^{t}\}) - \nabla_{\boldsymbol{\theta}_{i,j}} \tilde{L}_{p_i}(\{\mathbf{x}_i^{t}\}, \{\mathbf{y}_i^{t}\}, \{\lambda_{i,l}^{t}\}, \{\boldsymbol{\theta}_{i,j}^{t-1}\})||^2 + \frac{1}{2\eta_\theta}||v_{2,i,j}^{t+1}||^2 \\
-&\frac{1}{L_2' + c_2^{t-1}}||\nabla_{\boldsymbol{\theta}_{i,j}} \tilde{L}_{p_i}(\{\mathbf{x}_i^{t}\}, \{\mathbf{y}_i^{t}\}, \{\lambda_{i,l}^{t}\}, \{\boldsymbol{\theta}_{i,j}^{t}\}) - \nabla_{\boldsymbol{\theta}_{i,j}} \tilde{L}_{p_i}(\{\mathbf{x}_i^{t}\}, \{\mathbf{y}_i^{t}\}, \{\lambda_{i,l}^{t}\}, \{\boldsymbol{\theta}_{i,j}^{t-1}\})||^2 + \frac{1}{2\eta_\theta}||^2 \\
-&\frac{c_2^{t-1}L_2'}{L_2' + c_2^{t-1}}||\boldsymbol{\theta}_{i,j}^{t} - \boldsymbol{\theta}_{i,j}^{t-1}||^2,
\end{aligned} \tag{S88}$$

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 S88 by $\frac{8}{\eta_\theta c_2^t}$, we have

$$\begin{aligned}
&\frac{4}{\eta_\theta^2 c_2^t}\left\|\boldsymbol{\theta}_{i,j}^{t+1} - \boldsymbol{\theta}_{i,j}^{t}\right\|^2 - \frac{4}{\eta_\theta}(\frac{c_2^{t-1} - c_2^t}{c_2^t})\left\|\boldsymbol{\theta}_{i,j}^{t+1}\right\|^2 \\
\leq&\frac{4}{\eta_\theta^2 c_2^t}\left\|\boldsymbol{\theta}_{i,j}^{t} - \boldsymbol{\theta}_{i,j}^{t-1}\right\|^2 - \frac{4}{\eta_\theta}(\frac{c_2^{t-1} - c_2^t}{c_2^t})\left\|\boldsymbol{\theta}_{i,j}^{t}\right\|^2 + \frac{8b_2^t}{\eta_\theta c_2^t}||\boldsymbol{\theta}_{i,j}^{t+1} - \boldsymbol{\theta}_{i,j}^{t}||^2 - \frac{4}{\eta_\theta}||\boldsymbol{\theta}_{i,j}^{t} - \boldsymbol{\theta}_{i,j}^{t-1}||^2 \\
+&\frac{8L^2}{\eta_\theta c_2^t b_2^t}\sum_{i=1}^{N}(||\mathbf{x}_i^{t+1} - \mathbf{x}_i^{t}||^2 + ||\mathbf{y}_i^{t+1} - \mathbf{y}_i^{t}||^2).
\end{aligned} \tag{S89}$$

Setting $b_2^t = \frac{c_2^t}{4}$, using the definition of S_2^t, we have,

$$\begin{aligned}
&S_2^{t+1} - S_2^t \\
\leq&\frac{4}{\eta_\theta}\left(\frac{c_2^{t-2}}{c_2^{t-1}} - \frac{c_2^{t-1}}{c_2^t}\right)\sum_{i=1}^{N}\sum_{j=1}^{N}\left\|\boldsymbol{\theta}_{i,j}^{t}\right\|^2 + (\frac{2}{\eta_\theta} + \frac{4}{\eta_\theta^2}(\frac{1}{c_2^{t+1}} - \frac{1}{c_2^t}))\sum_{i=1}^{N}\sum_{j=1}^{N}||\boldsymbol{\theta}_{i,j}^{t+1} - \boldsymbol{\theta}_{i,j}^{t}||^2 \\
-&\frac{4}{\eta_\theta}\sum_{i=1}^{N}\sum_{j=1}^{N}||\boldsymbol{\theta}_{i,j}^{t} - \boldsymbol{\theta}_{i,j}^{t-1}||^2 + \frac{32NL^2}{\eta_\theta (c_2^t)^2}\sum_{i=1}^{N}\sum_{j=1}^{N}(||\mathbf{x}_i^{t+1} - \mathbf{x}_i^{t}||^2 + ||\mathbf{y}_i^{t+1} - \mathbf{y}_i^{t}||^2).
\end{aligned} \tag{S90}$$

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.(S30), 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.(S30). Then using the results from Eq.(S69), Eq.(S70), Eq.(S85) and Eq.(S90), Lemma 5 (Eq.(S78)) can be proved.

1.4.6 Proof of Theorem 5.1

According to Definition 5.1, for $i = 1, \cdots, N$, we have

$$
\begin{aligned}
&\mathbb{E}[\sum_{i=1}^{N} ||P(\mathbf{d}_i^t, \nabla_{\mathbf{x}_i} \tilde{L}_{p\_i}(\{\mathbf{x}_i^{\hat{t}_i}\}, \{\mathbf{y}_i^{\hat{t}_i}\}, \{\lambda_{i,l}^{\hat{t}_i}\}, \{\boldsymbol{\theta}_{i,j}^{\hat{t}_i}\}; \xi), \tilde{\eta}_{i,x}^t)||^2] \\
=&\mathbb{E}[\sum_{i=1}^{N} ||\frac{1}{\eta_{i,x}^t}(\mathbf{d}_i^t - \text{prox}_\eta^R(\mathbf{d}_i^t - \eta_{i,x} \nabla_{\mathbf{x}_i} \tilde{L}_{p\_i}(\{\mathbf{x}_i^{\hat{t}_i}\}, \{\mathbf{y}_i^{\hat{t}_i}\}, \{\lambda_{i,l}^{\hat{t}_i}\}, \{\boldsymbol{\theta}_{i,j}^{\hat{t}_i}\}; \xi)))||^2] \\
=&\frac{2}{\eta_x^2}\mathbb{E}[\sum_{i=1}^{N} ||\mathbf{d}_i^t - \mathbf{x}_i^{t+1}||^2].
\end{aligned}
\tag{S91}
$$

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

$$
\begin{aligned}
&\mathbb{E}[\sum_{i=1}^{N} ||P(\mathbf{d}_i^t, \nabla_{\mathbf{x}_i} \tilde{L}_{p\_i}(\{\mathbf{x}_i^{\hat{t}_i}\}, \{\mathbf{y}_i^{\hat{t}_i}\}, \{\lambda_{i,l}^{\hat{t}_i}\}, \{\boldsymbol{\theta}_{i,j}^{\hat{t}_i}\}; \xi), \eta_{i,x}) - P(\mathbf{d}_i^t, \bar{\nabla}_{\mathbf{d}} L'_{p\_i}(\{\mathbf{d}_i^t\}, \{\mathbf{y}_i^t\}, \{\lambda_{i,l}^t\}, \{\boldsymbol{\theta}_{i,j}^t\}), \eta_{i,x})||^2] \\
=&\frac{1}{\eta_x^2}\mathbb{E}[\sum_{i=1}^{N} ||\text{prox}_\eta^R(\mathbf{d}_i^t - \eta_x \nabla_{\mathbf{x}_i} \tilde{L}_{p\_i}(\{\mathbf{x}_i^{\hat{t}_i}\}, \{\mathbf{y}_i^{\hat{t}_i}\}, \{\lambda_{i,l}^{\hat{t}_i}\}, \{\boldsymbol{\theta}_{i,j}^{\hat{t}_i}\}; \xi) - \text{prox}_\eta^R(\mathbf{d}_i^t - \eta_x \bar{\nabla}_{\mathbf{d}} L'_{p\_i}(\{\mathbf{d}_i^t\}, \{\mathbf{y}_i^t\}, \{\lambda_{i,l}^t\}, \{\boldsymbol{\theta}_{i,j}^t\}))||^2] \\
\le&\mathbb{E}[\sum_{i=1}^{N} ||\nabla_{\mathbf{x}_i} L'_{p\_i}(\{\mathbf{x}_i^{\hat{t}_i}\}, \{\mathbf{y}_i^{\hat{t}_i}\}, \{\lambda_{i,l}^{\hat{t}_i}\}, \{\boldsymbol{\theta}_{i,j}^{\hat{t}_i}\}; \xi) - \bar{\nabla}_{\mathbf{d}} L'_{p\_i}(\{\mathbf{d}_i^t\}, \{\mathbf{y}_i^t\}, \{\lambda_{i,l}^t\}, \{\boldsymbol{\theta}_{i,j}^t\})||^2].
\end{aligned}
\tag{S92}
$$

According to Young's inequality and Eq.(S91), we have

$$
\begin{aligned}
&\frac{1}{2}\mathbb{E}[\sum_{i=1}^{N} ||P(\mathbf{d}_i^t, \bar{\nabla}_{\mathbf{d}} L'_{p\_i}(\{\mathbf{d}_i^t\}, \{\mathbf{y}_i^t\}, \{\lambda_{i,l}^t\}, \{\boldsymbol{\theta}_{i,j}^t\}), \eta_{i,x})||^2] \\
\le&\frac{1}{\eta_x^2}\mathbb{E}[\sum_{i=1}^{N} ||\mathbf{x}_i^{t+1} - \mathbf{d}_i^t||^2] + \mathbb{E}[\sum_{i=1}^{N} ||P(\mathbf{d}_i^t, \nabla_{\mathbf{x}_i} \tilde{L}_{p\_i}(\{\mathbf{x}_i^{\hat{t}_i}\}, \{\mathbf{y}_i^{\hat{t}_i}\}, \{\lambda_{i,l}^{\hat{t}_i}\}, \{\boldsymbol{\theta}_{i,j}^{\hat{t}_i}\}; \xi), \eta_{i,x}) \\
&-P(\mathbf{d}_i^t, \bar{\nabla}_{\mathbf{d}} L'_{p\_i}(\{\mathbf{d}_i^t\}, \{\mathbf{y}_i^t\}, \{\lambda_{i,l}^t\}, \{\boldsymbol{\theta}_{i,j}^t\}), \eta_{i,x})||^2].
\end{aligned}
\tag{S93}
$$

Plugging Eq.(S92) into Eq.(S93) yields

$$
\begin{aligned}
&\frac{1}{2}\mathbb{E}[\sum_{i=1}^{N} ||P(\mathbf{d}_i^t, \bar{\nabla}_{\mathbf{d}} L'_{p\_i}(\{\mathbf{d}_i^t\}, \{\mathbf{y}_i^t\}, \{\lambda_{i,l}^t\}, \{\boldsymbol{\theta}_{i,j}^t\}), \eta_{i,x})||^2] \\
\le&\frac{1}{\eta_x^2}\mathbb{E}[\sum_{i=1}^{N} ||\mathbf{x}_i^{t+1} - \mathbf{d}_i^t||^2] \\
&+\mathbb{E}[\sum_{i=1}^{N} ||\nabla_{\mathbf{x}_i} \tilde{L}_{p\_i}(\{\mathbf{x}_i^{\hat{t}_i}\}, \{\mathbf{y}_i^{\hat{t}_i}\}, \{\lambda_{i,l}^{\hat{t}_i}\}, \{\boldsymbol{\theta}_{i,j}^{\hat{t}_i}\}; \xi) - \bar{\nabla}_{\mathbf{d}} L'_{p\_i}(\{\mathbf{d}_i^t\}, \{\mathbf{y}_i^t\}, \{\lambda_{i,l}^t\}, \{\boldsymbol{\theta}_{i,j}^t\})||^2].
\end{aligned}
\tag{S94}
$$

Next, we bound

$$
\begin{aligned}
&\mathbb{E}[\sum_{i=1}^{N}||\nabla_{\mathbf{x}_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}\};\xi)-\bar{\nabla}_{\mathbf{d}}L'_{p\_i}(\{\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}_{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}\};\xi)-\nabla_{\mathbf{x}_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}\})\\
+&\nabla_{\mathbf{x}_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}\})-\bar{\nabla}_{\mathbf{x}_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}\})\\
+&\bar{\nabla}_{\mathbf{x}_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}\})-\bar{\nabla}_{\mathbf{d}}L'_{p}(\{\mathbf{d}_i^t\},\{\mathbf{y}_i^t\},\{\lambda_{i,l}^t\},\{\boldsymbol{\theta}_{i,j}^t\})||^2]\\
\le&3N\sigma_1^2+3\mathbb{E}[\sum_{i=1}^{N}||\nabla_{\mathbf{x}_i}\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}\};\xi)-\bar{\nabla}_{\mathbf{x}_i}\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}\};\xi)||^2]\\
+&3\mathbb{E}[\sum_{i=1}^{N}||\nabla_{\mathbf{x}_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}\})-\nabla_{\mathbf{d}_i}L'_{p\_i}(\{\mathbf{d}_i^t\},\{\mathbf{y}_i^t\},\{\lambda_{i,l}^t\},\{\boldsymbol{\theta}_{i,j}^t\})||^2]
\end{aligned}
\tag{S95}
$$

According to Jensen's inequality, we have

$$
\begin{aligned}
&\mathbb{E}[\sum_{i=1}^{N}||\nabla_{\mathbf{x}_i}\tilde{L}_{p\_i}(\{\mathbf{x}_i^{\hat{t}_i}\},\{\mathbf{y}_i^{\hat{t}_i}\},\{\lambda_{i,l}^{\hat{t}_i}\},\{\boldsymbol{\theta}_{i,j}^{\hat{t}_i}\})-\nabla_{\mathbf{d}_i}L'_{p\_i}(\{\mathbf{d}_i^t\},\{\mathbf{y}_i^t\},\{\lambda_{i,l}^t\},\{\boldsymbol{\theta}_{i,j}^t\})||^2]\\
=&\mathbb{E}[\sum_{i=1}^{N}||\nabla_{\mathbf{x}_i}\tilde{L}_{p\_i}(\{\mathbf{x}_i^{\hat{t}_i}\},\{\mathbf{y}_i^{\hat{t}_i}\},\{\lambda_{i,l}^{\hat{t}_i}\},\{\boldsymbol{\theta}_{i,j}^{\hat{t}_i}\})-\nabla_{\mathbf{x}_i}\tilde{L}_{p\_i}(\{\mathbf{x}_i^{t}\},\{\mathbf{y}_i^{t}\},\{\lambda_{i,l}^{t}\},\{\boldsymbol{\theta}_{i,j}^{t}\})\\
+&\nabla_{\mathbf{x}_i}\tilde{L}_{p\_i}(\{\mathbf{x}_i^{t}\},\{\mathbf{y}_i^{t}\},\{\lambda_{i,l}^{t}\},\{\boldsymbol{\theta}_{i,j}^{t}\})-\nabla_{\mathbf{x}_i}\tilde{L}_{p\_i}(\{\bar{\mathbf{x}}^{t}\},\{\mathbf{y}_i^{t}\},\{\lambda_{i,l}^{t}\},\{\boldsymbol{\theta}_{i,j}^{t}\})\\
+&\nabla_{\mathbf{x}_i}\tilde{L}_{p\_i}(\{\bar{\mathbf{x}}^{t}\},\{\mathbf{y}_i^{t}\},\{\lambda_{i,l}^{t}\},\{\boldsymbol{\theta}_{i,j}^{t}\})-\nabla_{\mathbf{d}_i}L'_{p\_i}(\{\mathbf{d}_i^t\},\{\mathbf{y}_i^t\},\{\lambda_{i,l}^t\},\{\boldsymbol{\theta}_{i,j}^t\})||^2]\\
\le&3NL^2(\mathbb{E}[\sum_{i=1}^{N}||\mathbf{x}_i^t-\mathbf{x}_i^{\hat{t}_i}||^2]+\mathbb{E}[\sum_{i=1}^{N}||\mathbf{y}_i^t-\mathbf{y}_i^{\hat{t}_i}||^2]+\mathbb{E}[\sum_{i=1}^{N}||\bar{\mathbf{x}}^t-\mathbf{x}_i^t||^2]+\mathbb{E}[||\bar{\mathbf{x}}^t-\mathbf{d}_i^t||^2])\\
\le&6NL^2\tau k_1(\mathbb{E}[\sum_{i=1}^{N}||\mathbf{x}_i^{t+1}-\mathbf{x}_i^t||^2]+\mathbb{E}[\sum_{i=1}^{N}||\mathbf{y}_i^{t+1}-\mathbf{y}_i^t||^2])+3NL^2(\mathbb{E}[\sum_{i=1}^{N}||\bar{\mathbf{x}}^t-\mathbf{x}_i^t||^2]+\mathbb{E}[||\bar{\mathbf{x}}^t-\mathbf{d}_i^t||^2]).
\end{aligned}
\tag{S96}
$$

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.(S94), according to Eq.(S95), Eq.(S96) and Assumption 5.1, results in

$$
\begin{aligned}
&\frac{1}{2}\mathbb{E}[\sum_{i=1}^{N}||P(\mathbf{d}_i^t,\bar{\nabla}_{\mathbf{d}}L'_{p\_i}(\{\mathbf{d}_i^t\},\{\mathbf{y}_i^t\},\{\lambda_{i,l}^t\},\{\boldsymbol{\theta}_{i,j}^t\}),\eta_{i,x})||^2]+\frac{L^2}{2}\mathbb{E}[\sum_{i=1}^{N}||\mathbf{x}_i^t-\bar{\mathbf{x}}^t||^2]\\
\le&\frac{1}{\eta_x^2}\mathbb{E}[\sum_{i=1}^{N}||\mathbf{x}_i^{t+1}-\mathbf{d}_i^t||^2]+\frac{19L^2}{2}\sum_{i=1}^{N}\mathbb{E}[||\mathbf{x}_i^t-\bar{\mathbf{x}}^t||^2]+18NL^2\tau k_1(\mathbb{E}[\sum_{i=1}^{N}||\mathbf{x}_i^{t+1}-\mathbf{x}_i^t||^2]+\mathbb{E}[\sum_{i=1}^{N}||\mathbf{y}_i^{t+1}-\mathbf{y}_i^t||^2])\\
&+9NL^2\mathbb{E}[||\bar{\mathbf{x}}^t-\mathbf{d}_i^t||^2]+21N\sigma_1^2+9N\varsigma_1^2.
\end{aligned}
\tag{S97}
$$

According to Eq.(S97) and Eq.(S25), we have

$$
\begin{aligned}
&\mathbb{E}[\sum_{i=1}^{N}||P(\mathbf{d}_i^t,\bar{\nabla}_{\mathbf{d}}L'_{p\_i}(\{\mathbf{d}_i^t\},\{\mathbf{y}_i^t\},\{\lambda_{i,l}^t\},\{\boldsymbol{\theta}_{i,j}^t\}),\eta_{i,x})||^2+L^2\mathbb{E}[\sum_{i=1}^{N}||\mathbf{x}_i^t-\bar{\mathbf{x}}^t||^2]\\
\le&\frac{2}{\eta_x^2}\mathbb{E}[\sum_{i=1}^{N}||\mathbf{x}_i^{t+1}-\mathbf{d}_i^t||^2]+48NL^2\mathbb{E}[\sum_{i=1}^{N}||\mathbf{x}_i^t-\bar{\mathbf{x}}^t||^2]+36NL^2\tau k_1(\mathbb{E}[\sum_{i=1}^{N}||\mathbf{x}_i^{t+1}-\mathbf{x}_i^t||^2]+\mathbb{E}[\sum_{i=1}^{N}||\mathbf{y}_i^{t+1}-\mathbf{y}_i^t||^2])\\
+&42N\sigma_1^2+18N\varsigma_1^2.
\end{aligned}
\tag{S98}
$$

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

$$
\begin{aligned}
&\frac{1}{2}\mathbb{E}[\sum_{i=1}^{N}||\bar{\nabla}_{\mathbf{u}}L'_{p\_i}(\{\mathbf{x}_i^t\},\{\mathbf{u}_i^t\},\{\lambda_{i,l}^t\},\{\boldsymbol{\theta}_{i,j}^t\})||^2]\\
\leq&\mathbb{E}[\sum_{i=1}^{N}||\nabla_{\mathbf{y}_i}\tilde{L}_{p\_i}(\{\mathbf{x}_i^{\hat{t}_i}\},\{\mathbf{y}_i^{\hat{t}_i}\},\{\lambda_{i,l}^{\hat{t}_i}\},\{\boldsymbol{\theta}_{i,j}^{\hat{t}_i}\};\xi)||^2]\\
+&\mathbb{E}[\sum_{i=1}^{N}||\nabla_{\mathbf{y}_i}\tilde{L}_{p\_i}(\{\mathbf{x}_i^{\hat{t}_i}\},\{\mathbf{y}_i^{\hat{t}_i}\},\{\lambda_{i,l}^{\hat{t}_i}\},\{\boldsymbol{\theta}_{i,j}^{\hat{t}_i}\};\xi)-\bar{\nabla}_{\mathbf{u}}L'_{p\_i}(\{\mathbf{x}_i^t\},\{\mathbf{u}_i^t\},\{\lambda_{i,l}^t\},\{\boldsymbol{\theta}_{i,j}^t\})||^2].
\end{aligned}
\tag{S99}
$$

Next, we bound

$$
\begin{aligned}
&\mathbb{E}[\sum_{i=1}^{N}||\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}\};\xi)-\bar{\nabla}_{\mathbf{u}}L'_{p\_i}(\{\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}_{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}\};\xi)-\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}\})\\
+&\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}\})-\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}\})\\
+&\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}\})-\bar{\nabla}_{\mathbf{u}}L'_{p\_i}(\{\mathbf{x}_i^t\},\{\mathbf{u}_i^t\},\{\lambda_{i,l}^t\},\{\boldsymbol{\theta}_{i,j}^t\})||^2]\\
\leq&3N\sigma_2^2+3\mathbb{E}[\sum_{i=1}^{N}||\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}\})-\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]\\
+&3\mathbb{E}[\sum_{i=1}^{N}||\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}\})-\nabla_{\mathbf{u}_i}L'_{p\_i}(\{\mathbf{x}_i^t\},\{\mathbf{u}_i^t\},\{\lambda_{i,l}^t\},\{\boldsymbol{\theta}_{i,j}^t\})||^2].
\end{aligned}
\tag{S100}
$$

According to Jensens inequality and Assumption 5.1, we have

$$
\begin{aligned}
&\mathbb{E}[\sum_{i=1}^{N}||\nabla_{\mathbf{y}_i}\tilde{L}_{p\_i}(\{\mathbf{x}_i^{\hat{t}_i}\},\{\mathbf{y}_i^{\hat{t}_i}\},\{\lambda_{i,l}^{\hat{t}_i}\},\{\boldsymbol{\theta}_{i,j}^{\hat{t}_i}\})-\nabla_{\mathbf{u}_i}L'_{p\_i}(\{\mathbf{x}_i^t\},\{\mathbf{u}_i^t\},\{\lambda_{i,l}^t\},\{\boldsymbol{\theta}_{i,j}^t\})||^2]\\
=&\mathbb{E}[\sum_{i=1}^{N}||\nabla_{\mathbf{y}_i}\tilde{L}_{p\_i}(\{\mathbf{x}_i^{\hat{t}_i}\},\{\mathbf{y}_i^{\hat{t}_i}\},\{\lambda_{i,l}^{\hat{t}_i}\},\{\boldsymbol{\theta}_{i,j}^{\hat{t}_i}\})-\nabla_{\mathbf{y}_i}\tilde{L}_{p\_i}(\{\mathbf{x}^t\},\{\mathbf{y}^t\},\{\lambda_{i,l}^t\},\{\boldsymbol{\theta}_{i,j}^t\})\\
+&\nabla_{\mathbf{y}_i}\tilde{L}_{p\_i}(\{\mathbf{x}^t\},\{\mathbf{y}_i^t\},\{\lambda_{i,l}^t\},\{\boldsymbol{\theta}_{i,j}^t\})-\nabla_{\mathbf{y}_i}L'_{p\_i}(\{\mathbf{x}_i^t\},\{\bar{\mathbf{y}}^t\},\{\lambda_{i,l}^t\},\{\boldsymbol{\theta}_{i,j}^t\})||^2]\\
+&\nabla_{\mathbf{y}_i}\tilde{L}_{p\_i}(\{\mathbf{x}^t\},\{\bar{\mathbf{y}}^t\},\{\lambda_{i,l}^t\},\{\boldsymbol{\theta}_{i,j}^t\})-\nabla_{\mathbf{u}_i}L'_{p\_i}(\{\mathbf{x}_i^t\},\{\mathbf{u}_i^t\},\{\lambda_{i,l}^t\},\{\boldsymbol{\theta}_{i,j}^t\})||^2]\\
\leq&3NL^2(\mathbb{E}[\sum_{i=1}^{N}||\mathbf{x}^t-\mathbf{x}_i^{\hat{t}_i}||^2]+\mathbb{E}[\sum_{i=1}^{N}||\mathbf{y}_i^t-\mathbf{y}_i^{\hat{t}_i}||^2]+\mathbb{E}[\sum_{i=1}^{N}||\bar{\mathbf{y}}^t-\mathbf{y}_i^t||^2]+\mathbb{E}[\sum_{i=1}^{N}||\bar{\mathbf{y}}^t-\mathbf{u}_i^t||^2])\\
\leq&6NL^2\tau k_1(\mathbb{E}[\sum_{i=1}^{N}||\mathbf{x}_i^{t+1}-\mathbf{x}_i^t||^2]+\mathbb{E}[\sum_{i=1}^{N}||\mathbf{y}_i^{t+1}-\mathbf{y}_i^t||^2])+3NL^2(\mathbb{E}[\sum_{i=1}^{N}||\bar{\mathbf{y}}^t-\mathbf{y}_i^t||^2]+\mathbb{E}[\sum_{i=1}^{N}||\bar{\mathbf{y}}^t-\mathbf{u}_i^t||^2]).
\end{aligned}
\tag{S101}
$$

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.(S99), according to Eq.(S100), Eq.(S101) and Eq.(S25), it follows that

$$
\begin{aligned}
&\mathbb{E}[\sum_{i=1}^{N}||\bar{\nabla}_{\mathbf{u}}L'_{p\_i}(\{\mathbf{x}_i^t\},\{\mathbf{u}_i^t\},\{\lambda_{i,l}^t\},\{\boldsymbol{\theta}_{i,j}^t\})||^2]+L^2\mathbb{E}[\sum_{i=1}^{N}||\mathbf{y}_i^t-\bar{\mathbf{y}}^t||^2]\\
\leq&36NL^2\tau k_1(\mathbb{E}[\sum_{i=1}^{N}||\mathbf{x}_i^{t+1}-\mathbf{x}_i^t||^2]+\mathbb{E}[\sum_{i=1}^{N}||\mathbf{y}_i^{t+1}-\mathbf{y}_i^t||^2])+48NL^2(\mathbb{E}[\sum_{i=1}^{N}||\bar{\mathbf{y}}^t-\mathbf{y}_i^t||^2])\\
+&2\mathbb{E}[\sum_{i=1}^{N}||\nabla_{\mathbf{y}_i}\tilde{L}_{p\_i}(\{\mathbf{x}_i^{\hat{t}_i}\},\{\mathbf{y}_i^{\hat{t}_i}\},\{\lambda_{i,l}^{\hat{t}_i}\},\{\boldsymbol{\theta}_{i,j}^{\hat{t}_i}\};\xi)||^2]+42N\sigma_2^2+18N\varsigma_2^2.
\end{aligned}
\tag{S102}
$$

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

$$\begin{aligned}\mathbb{E}[||(\tilde{\mathcal{G}}^t)_{\lambda_{i,l}}||^2] \leq & 4L^2(\mathbb{E}[\sum_{i=1}^{N}||\mathbf{x}_i^{t+1}-\mathbf{x}_i^t||^2]+\mathbb{E}[\sum_{i=1}^{N}||\mathbf{y}_i^{t+1}-\mathbf{y}_i^t||^2])+4((c_1^{t-1})^2-(c_1^t)^2)\mathbb{E}[||\lambda_{i,l}^t||^2] \\ & +\frac{4}{\eta_\lambda^2}\mathbb{E}[||\lambda_{i,l}^{t+1}-\lambda_{i,l}^t||^2]+4\sigma_3^2.\end{aligned} \tag{S103}$$

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

$$\begin{aligned}\mathbb{E}[||(\tilde{\mathcal{G}}^t)_{\boldsymbol{\theta}_{i,j}}||^2] \leq & 4L^2(\mathbb{E}[\sum_{i=1}^{N}||\mathbf{x}_i^{t+1}-\mathbf{x}_i^t||^2]+\mathbb{E}[\sum_{i=1}^{N}||\mathbf{y}_i^{t+1}-\mathbf{y}_i^t||^2])+4((c_2^{t-1})^2-(c_2^t)^2)\mathbb{E}[||\boldsymbol{\theta}_{i,j}^t||^2] \\ & +\frac{4}{\eta_\theta^2}\mathbb{E}[||\boldsymbol{\theta}_{i,j}^{t+1}-\boldsymbol{\theta}_{i,j}^t||^2]+4\sigma_4^2.\end{aligned} \tag{S104}$$

According to Eq.(S69), Eq.(S70), Eq.(S98), Eq.(S102), Eq.(S103) and Eq.(S104), we can obtain

$$\begin{aligned}&\mathbb{E}[\tilde{\Psi}^t] \\ \leq&(\frac{2}{N\eta_x^2}+144L^2\tau k_1+\frac{8L^2}{NM}+\frac{8L^2}{N^2})\mathbb{E}[\sum_{i=1}^{N}||\mathbf{x}_i^{t+1}-\mathbf{d}_i^t||^2] \\ &+(48L^2+576L^2\tau k_1+\frac{32L^2}{NM}+\frac{32L^2}{N^2})\mathbb{E}[\sum_{i=1}^{N}||\mathbf{x}_i^t-\bar{\mathbf{x}}^t||^2] \\ &+(48L^2+(\rho+\frac{2}{1-\rho})(72L^2\tau k_1+\frac{4L^2}{NM}+\frac{4L^2}{N^2}))(\mathbb{E}[\sum_{i=1}^{N}||\bar{\mathbf{y}}^t-\mathbf{y}_i^t||^2]) \\ &+(\frac{2}{N}+\frac{2\eta_y^2}{1-\rho}(72L^2\tau k_1+\frac{4L^2}{NM}+\frac{4L^2}{N^2}))\mathbb{E}[\sum_{i=1}^{N}||\nabla_{\mathbf{y}_i}\tilde{L}_{p\_i}(\{\mathbf{x}_i^{\hat{t}_i}\},\{\mathbf{y}_i^{\hat{t}_i}\},\{\lambda_{i,l}^{\hat{t}_i}\},\{\boldsymbol{\theta}_{i,j}^{\hat{t}_i}\};\xi)||^2] \\ &+\frac{4}{NM}((c_1^{t-1})^2-(c_1^t)^2)\sum_{i=1}^{N}\sum_{l=1}^{|\mathcal{P}_i^t|}\mathbb{E}[||\lambda_{i,l}^t||^2]+\frac{4}{N^2}((c_2^{t-1})^2-(c_2^t)^2)\sum_{i=1}^{N}\sum_{j=1}^{N}\mathbb{E}[||\boldsymbol{\theta}_{i,j}^t||^2] \\ &+\frac{4}{NM\eta_\lambda^2}\sum_{i=1}^{N}\sum_{l=1}^{|\mathcal{P}_i^t|}\mathbb{E}[||\lambda_{i,l}^{t+1}-\lambda_{i,l}^t||^2]+\frac{4}{N^2\eta_\theta^2}\sum_{i=1}^{N}\sum_{j=1}^{N}\mathbb{E}[||\boldsymbol{\theta}_{i,j}^{t+1}-\boldsymbol{\theta}_{i,j}^t||^2] \\ &+42\sigma_1^2+18\varsigma_1^2+42\sigma_2^2+18\varsigma_2^2+4\sigma_3^2+4\sigma_4^2.\end{aligned} \tag{S105}$$

Multiply both sides of Eq.(S78) by $\frac{1}{N^2M}$, let $\eta_x = \eta_y = \eta_\lambda = \eta_\theta = \eta$, $\beta = 1$ to have:

$$
\begin{aligned}
&C_1\mathbb{E}[\sum_{i=1}^{N}||\mathbf{x}_i^{t+1}-\bar{\mathbf{x}}^t||^2] + C_2\mathbb{E}[\sum_{i=1}^{N}||\mathbf{x}_i^{t+1}-\mathbf{d}_i^t||^2] + C_3\mathbb{E}[\sum_{i=1}^{N}||\bar{\mathbf{x}}^t-\mathbf{x}_i^t||^2] + C_4\mathbb{E}[\sum_{i=1}^{N}||\bar{\mathbf{y}}^t-\mathbf{y}_i^t||^2] \\
&+C_5\mathbb{E}[\sum_{i=1}^{N}||\nabla_{\mathbf{y}_i}\tilde{L}_{p\_i}(\{\mathbf{x}_i^{\hat{t}_i}\},\{\mathbf{y}_i^{\hat{t}_i}\},\{\lambda_{i,l}^{\hat{t}_i}\},\{\boldsymbol{\theta}_{i,j}^{\hat{t}_i}\};\xi)||^2] + C_6\mathbb{E}[\sum_{i=1}^{N}\sum_{l=1}^{|\mathcal{P}_i^t|}||\lambda_{i,l}^{t+1}-\lambda_{i,l}^t||^2] + C_7\mathbb{E}[\sum_{i=1}^{N}\sum_{j=1}^{N}||\boldsymbol{\theta}_{i,j}^{t+1}-\boldsymbol{\theta}_{i,j}^t||^2] \\
\le&\frac{\mathbb{E}[F^t-F^{t+1}]}{N^2M} + \frac{4}{\eta N^2M}\left(\frac{c_1^{t-2}}{c_1^{t-1}}-\frac{c_1^{t-1}}{c_1^t}\right)\mathbb{E}[\sum_{i=1}^{N}\sum_{l=1}^{|\mathcal{P}_i^t|}||\lambda_{i,l}^t||^2] + \frac{4}{\eta N^2M}\left(\frac{c_2^{t-2}}{c_2^{t-1}}-\frac{c_2^{t-1}}{c_2^t}\right)\mathbb{E}[\sum_{i=1}^{N}\sum_{j=1}^{N}\left\|\boldsymbol{\theta}_{i,j}^t\right\|^2] \\
&+\frac{1}{N^2M}(\frac{c_1^{t-1}}{2}-\frac{c_1^t}{2})\mathbb{E}[\sum_{i=1}^{N}\sum_{l=1}^{|\mathcal{P}_i^t|}||\lambda_{i,l}^{t+1}||^2] + \frac{1}{N^2M}(\frac{c_2^{t-1}}{2}-\frac{c_2^t}{2})\mathbb{E}[\sum_{i=1}^{N}\sum_{j=1}^{N}||\boldsymbol{\theta}_{i,j}^{t+1}||^2] \\
&+\frac{\gamma_1^t\eta^2(6\sigma_1^2+3\varsigma_1^2)}{MN(1-\rho)} + \frac{\gamma_2^t\eta^2(6\sigma_2^2+3\varsigma_2^2)}{MN(1-\rho)} + \frac{\eta\sigma_2^2}{M},
\end{aligned}
\tag{S106}
$$

where

$$
\begin{aligned}
C_1 &= \frac{1}{N^2M}(\frac{N}{2\eta}-NL-2\eta N^2L^2-MNL^2\eta-N^2L^2\eta), \\
C_2 &= \frac{1}{N^2M}(\frac{N}{2\eta}-(4\tau k_1N^2L+8\eta N^2L^2\tau k_1+\frac{64NML^2}{\eta(c_1^t)^2}+\frac{64N^2L^2}{\eta(c_2^t)^2})), \\
C_3 &= \frac{1}{N^2M}(\gamma_1^{t-1}-\gamma_1^t\rho-\frac{N}{2\eta}-3N^2L-MNL^2\eta-N^2L^2\eta-(16\tau k_1N^2L+32\eta N^2L^2\tau k_1+\frac{256NML^2}{\eta(c_1^t)^2}+\frac{256N^2L^2}{\eta(c_2^t)^2}), \\
C_4 &= \frac{1}{N^2M}(\gamma_2^{t-1}-\gamma_2^t\rho-2\eta N^2L^2-MNL^2\eta-N^2L^2\eta-(\rho+\frac{2}{1-\rho})(2\tau k_1N^2L+4\eta N^2L^2\tau k_1+\frac{32NML^2}{\eta_\lambda^t(c_1^t)^2}+\frac{32N^2L^2}{\eta_\theta(c_2^t)^2})), \\
C_5 &= \frac{\eta}{N^2M}(N-\frac{N}{2}-\frac{\eta NL}{2}-MNL^2\eta^2-N^2L^2\eta^2-\frac{4\eta}{1-\rho}(\tau k_1N^2L+2\eta N^2L^2\tau k_1+\frac{16NML^2}{\eta(c_1^t)^2}+\frac{16N^2L^2}{\eta(c_2^t)^2})), \\
C_6 &= \frac{1}{\eta}, \\
C_7 &= \frac{1}{\eta}.
\end{aligned}
\tag{S107}
$$

From Eq.(S105), we have

$$
\begin{aligned}
&\mathbb{E}[\tilde{\Psi}^t] \\
\le&C_2'\mathbb{E}[\sum_{i=1}^{N}||\mathbf{x}_i^{t+1}-\mathbf{d}_i^t||^2] + C_3'\mathbb{E}[\sum_{i=1}^{N}||\mathbf{x}_i^t-\bar{\mathbf{x}}^t||^2] + C_4'(\mathbb{E}[\sum_{i=1}^{N}||\bar{\mathbf{y}}^t-\mathbf{y}_i^t||^2]) \\
&+C_5'\mathbb{E}[\sum_{i=1}^{N}||\nabla_{\mathbf{y}_i}\tilde{L}_{p\_i}(\{\mathbf{x}_i^{\hat{t}_i}\},\{\mathbf{y}_i^{\hat{t}_i}\},\{\lambda_{i,l}^{\hat{t}_i}\},\{\boldsymbol{\theta}_{i,j}^{\hat{t}_i}\};\xi)||^2] + C_6'\sum_{i=1}^{N}\sum_{l=1}^{|\mathcal{P}_i^t|}\mathbb{E}[||\lambda_{i,l}^{t+1}-\lambda_{i,l}^t||^2] + C_7'\sum_{i=1}^{N}\sum_{j=1}^{N}\mathbb{E}[||\boldsymbol{\theta}_{i,j}^{t+1}-\boldsymbol{\theta}_{i,j}^t||^2] \\
&+\frac{4}{NM}((c_1^{t-1})^2-(c_1^t)^2)\sum_{i=1}^{N}\sum_{l=1}^{|\mathcal{P}_i^t|}\mathbb{E}[||\lambda_{i,l}^t||^2] + \frac{4}{N^2}((c_2^{t-1})^2-(c_2^t)^2)\sum_{i=1}^{N}\sum_{j=1}^{N}\mathbb{E}[||\boldsymbol{\theta}_{i,j}^t||^2] + 42\sigma_1^2+18\varsigma_1^2+42\sigma_2^2+18\varsigma_2^2+4\sigma_3^2+4\sigma_4^2,
\end{aligned}
\tag{S108}
$$

where

$$
\begin{aligned}
C_1' &= 0, \\
C_2' &= \frac{2}{N\eta^2} + 144L^2\tau k_1 + \frac{8L^2}{NM} + \frac{8L^2}{N^2}, \\
C_3' &= 48L^2 + 576L^2\tau k_1 + \frac{32L^2}{NM} + \frac{32L^2}{N^2}, \\
C_4' &= 48L^2 + (\rho + \frac{2}{1-\rho})(72L^2\tau k_1 + \frac{4L^2}{NM} + \frac{4L^2}{N^2}), \\
C_5' &= \frac{2}{N} + \frac{2\eta^2}{1-\rho}(72L^2\tau k_1 + \frac{4L^2}{NM} + \frac{4L^2}{N^2}), \\
C_6 &= \frac{4}{NM\eta^2}, \\
C_7 &= \frac{4}{N^2\eta}{}^2.
\end{aligned}
\tag{S109}
$$

Using $\eta \leq \min\{\frac{1}{8L}, \frac{1}{8L\sqrt{M}}, \frac{1}{12L\sqrt{N}}\}$ to have

$$
\begin{aligned}
C_1 &= \frac{1}{N^2M}(\frac{N}{2\eta_x} - NL - 2\eta N^2L^2 - MNL^2\eta - N^2L^2\eta) \\
&= \frac{1}{\eta NM}(\frac{1}{2} - L\eta - 3\eta^2L^2N - L^2\eta^2M) \\
&\geq \frac{1}{\eta NM}(\frac{1}{2} - \frac{1}{8} - \frac{1}{48} - \frac{1}{64}) = \frac{131}{192}\frac{1}{\eta NM} > 0.
\end{aligned}
\tag{S110}
$$

Using $\eta \leq \min\{\frac{1}{256\tau k_1 NL}, \frac{1}{32L\sqrt{\tau k_1 N}}, \frac{1}{16L\sqrt{M(T_1+T+1)}}, \frac{1}{16L\sqrt{N(T_1+T+1)}}\}$ to have

$$
\begin{aligned}
C_2 &= \frac{1}{N^2M}(\frac{N}{2\eta} - (4\tau k_1N^2L + 8\eta N^2L^2\tau k_1 + \frac{64NML^2}{\eta(c_1^t)^2} + \frac{64N^2L^2}{\eta(c_2^t)^2})) \\
&= \frac{1}{2NM\eta} - \frac{4}{NM\eta}(\tau k_1LN\eta + 2\eta L^2N\tau k_1\eta + 16L^2M\eta(T_1+T+1)\eta + 16L^2N\eta(T_1+T+1)\eta) \\
&\geq \frac{1}{NM\eta}(\frac{1}{2} - \frac{13}{256}) \geq \frac{1}{4NM\eta}.
\end{aligned}
\tag{S111}
$$

Using $\eta \leq \min\{\frac{M\sqrt{N}}{L\sqrt{\tau k_1}}, \frac{\sqrt{M}}{L}, \frac{\sqrt{N}}{L}\}$ to have

$$
\begin{aligned}
C_2' &= \frac{2}{N\eta^2} + 144L^2\tau k_1 + \frac{8L^2}{NM} + \frac{8L^2}{N^2} \\
&= \frac{1}{NM\eta^2}(2M + \frac{144L^2\tau k_1\eta^2}{NM} + 8L^2\eta^2 + \frac{8L^2M\eta^2}{N}) \\
&\leq \frac{1}{NM\eta^2}(2M + 144M + 8M + 8M) = \frac{1}{NM\eta^2}162M.
\end{aligned}
\tag{S112}
$$

Let $\gamma_1 = \frac{2N}{(1-\rho)\eta}$, further using $\eta \leq \min\{\frac{1}{64NL}, \frac{1}{12L\sqrt{N}}\}$ to have

$$
\begin{aligned}
C_3 &= \frac{\gamma_1(1-\rho)}{N^2M} - \frac{1}{NM2\eta} - \frac{L}{M} - \frac{3\eta L^2}{M} - \frac{L^2\eta}{N} \\
&\quad - 16(\frac{\tau k_1L}{M} + \frac{2\eta L^2\tau k_1}{M} + \frac{16L^2\eta(T_1+T+1)}{N} + \frac{16L^2\eta(T_1+T+1)}{M}) \\
&= \frac{\gamma_1(1-\rho)}{N^2M} - \frac{1}{NM\eta}(\frac{1}{2} + NL\eta + 3NL^2\eta^2 + ML^2\eta^2 \\
&\quad + 16(\tau k_1LN\eta + 2\eta L^2N\tau k_1\eta + 16L^2M\eta(T_1+T+1)\eta + 16L^2N\eta(T_1+T+1)\eta)) \\
&\geq \frac{\gamma_1(1-\rho)}{N^2M} - \frac{1}{NM\eta}\frac{37}{48} \geq \frac{1}{NM\eta}.
\end{aligned}
\tag{S113}
$$

Using $\eta = \min\{\frac{1}{L\sqrt{NM}}, \frac{1}{L\sqrt{NM\tau k_1}}, \frac{\sqrt{N}}{8L\sqrt{M}}\}$ to have

$$\begin{aligned} C'_3 &= 48L^2 + 576L^2\tau k_1 + \frac{32L^2}{NM} + \frac{32L^2}{N^2} \\ &= \frac{1}{NM\eta^2}(48L^2NM\eta^2 + 576L^2NM\eta^2\tau k_1 + 32L^2\eta^2 + \frac{32L^2M\eta^2}{N}) \\ &\le \frac{623}{NM\eta^2}. \end{aligned} \tag{S114}$$

Let $\gamma_2 = \frac{N}{\eta(1-\rho)}(82 + (\rho + \frac{2}{1-\rho})\frac{13}{512})$, further using $\eta \le \min\{\frac{1}{64ML}, \frac{\sqrt{(\rho+\frac{2}{1-\rho})}}{8L\sqrt{NM}}\}$ to have

$$\begin{aligned} \frac{1}{(\rho + \frac{2}{1-\rho})}C'_4 &= \frac{48L^2}{(\rho + \frac{2}{1-\rho})} + (72L^2\tau k_1 + \frac{4L^2}{NM} + \frac{4L^2}{N^2}) \\ &= \frac{1}{NM\eta^2}(\frac{48NML^2\eta^2}{(\rho + \frac{2}{1-\rho})} + (72L^2NM\eta^2\tau k_1 + 4L^2\eta^2 + \frac{4L^2M\eta^2}{N})) \\ &\le \frac{81}{NM\eta^2}, \end{aligned} \tag{S115}$$

and

$$C'_4 \le \frac{81(\rho + \frac{2}{1-\rho})}{NM\eta^2}. \tag{S116}$$

Using $\eta \le \min\{\frac{1-\rho}{64\tau k_1 NL}, \frac{\sqrt{1-\rho}}{8L\sqrt{\tau k_1 N}}, \frac{\sqrt{1-\rho}}{16L\sqrt{2(M+N)(T_1+T+1)}}\}$ to have

$$\begin{aligned} \eta C_5 &= \frac{1}{N^2M}(N - \frac{N}{2} - \frac{\eta NL}{2} - MNL^2\eta^2 - N^2L^2\eta^2 \\ &\quad - \frac{4\eta}{1-\rho}(\tau k_1 N^2L + 2\eta N^2L^2\tau k_1 + \frac{16NML^2}{\eta(c_1^t)^2} + \frac{16N^2L^2}{\eta(c_2^t)^2})) \\ &\ge \frac{1}{N^2M}(\frac{N}{2} - \frac{\eta NL}{2} - MNL^2\eta^2 - N^2L^2\eta^2 \\ &\quad - \frac{4\eta}{1-\rho}(\tau k_1 N^2L + 2\eta N^2L^2\tau k_1 + 16L^2NM(T_1+T+1)\eta + 16L^2N^2(T_1+T+1)\eta) \\ &= \frac{1}{NM}(\frac{1}{2} - \frac{\eta L}{2} - ML^2\eta^2 - NL^2\eta^2 - \frac{4\eta}{1-\rho}(\tau k_1 NL + 2\eta NL^2\tau k_1 + 16L^2M(T_1+T+1)\eta + 16L^2N(T_1+T+1)\eta) \\ &\ge \frac{1}{NM}\frac{15}{64}. \end{aligned} \tag{S117}$$

Using $\eta \le \min\{\frac{\sqrt{1-\rho}}{L\sqrt{N\tau k_1}}, \frac{\sqrt{1-\rho}}{L\sqrt{M}}, \frac{\sqrt{1-\rho}}{L\sqrt{N}}\}$ to have

$$\begin{aligned} C'_5 &= \frac{2}{N} + \frac{2\eta_y^2}{1-\rho}(72L^2\tau k_1 + \frac{4L^2}{NM} + \frac{4L^2}{N^2}) \\ &= \frac{1}{NM}(2M + \frac{2\eta^2}{1-\rho}(72NML^2\tau k_1 + 4L^2 + \frac{4ML^2}{N})) \\ &\le \frac{1}{NM}162M. \end{aligned} \tag{S118}$$

Let $p_2 = \frac{1024M}{\eta}, p_3 = \frac{2500}{\eta}, p_4 = \frac{(\rho+\frac{2}{1-\rho})}{\eta}, p_5 = \frac{700}{M\eta}, p_6 = \frac{4}{NM\eta}, p_7 = \frac{4}{N^2\eta}$, and set $p = \max\{p_2, p_3, p_4, p_5, p_6, p_7\} = \max\{\frac{1024M}{\eta}, \frac{2500}{\eta}\}$, then we have

$$C'_i \le pC_i, i = 2,3,4,5,6,7. \tag{S119}$$

Multiply both sides of Eq.(S106) by p, sum Eq.(S106) and Eq.(S108) from $t = T_1 + 2 \cdots T_1 + T$ and divide by $T - 1$. According to Eq.(S119) we have

$$\begin{aligned}
&\frac{1}{(T-1)} \sum_{t=T_1+2}^{T_1+T} \mathbb{E}[\tilde{\Psi}^t] \\
\leq& \frac{1}{(T-1)} \Big(\frac{F^{T_1+2} - \underline{L}}{N^2 M} + \frac{c_1^1 \alpha_1}{2N} + \frac{c_2^1 \alpha_2}{2M} + \frac{4}{N\eta} \big(\frac{c_1^0}{c_1^1} + \frac{c_1^1}{c_1^2} \big) \alpha_1 \Big) \big(\max\{ \frac{1024M}{\eta}, \frac{2500}{\eta} \} \big) \\
&+ \frac{1}{(T-1)} \Big(\frac{4}{M\eta} \big(\frac{c_2^0}{c_2^1} + \frac{c_2^1}{c_2^2} \big) \alpha_2 + 4(c_1^1)^2 \alpha_1 + 4(c_2^1)^2 \alpha_2 + \frac{6}{N\eta} \beta_1^2 + \frac{6}{M\eta} {\beta_2}^2 \Big) \big(\max\{ \frac{1024M}{\eta}, \frac{2500}{\eta} \} \big) \\
&+ \Big(42\sigma_1^2 + 18\varsigma_1^2 + 42\sigma_2^2 + 18\varsigma_2^2 + 4\sigma_3^2 + 4\sigma_4^2 + \frac{2(6\sigma_1^2 + 3\varsigma_1^2)}{M(1-\rho)^2} + \frac{(82 + (\rho + \frac{2}{1-\rho})\frac{13}{512})(6\sigma_2^2 + 3\varsigma_2^2)}{M(1-\rho)^2} + \frac{\sigma_2^2}{M} \Big) (\max\{1024M, 2500\}),
\end{aligned} \tag{S120}$$

where $\beta_1 = \max\{||\lambda_1 - \lambda_2||\}$, $\beta_2 = \max\{||\boldsymbol{\theta}_1 - \boldsymbol{\theta}_2||\}$, $\underline{L} = \min L_p(\{\bar{\mathbf{x}}^t\}, \{\bar{\mathbf{y}}^t\}, \{\lambda_{i,l}^t\}, \{\boldsymbol{\theta}_{i,j}^t\})$ satisfies $\forall t \geq T_1 + 2$ $F^t \geq \underline{L} - \frac{4}{\eta_\lambda} \frac{c_1^1}{c_1^2} NM\alpha_1 - \frac{4}{\eta_\theta} \frac{c_2^1}{c_2^2} N^2 \alpha_2 - \frac{6}{\eta_\lambda} NM\beta_1^2 - \frac{6}{\eta_\theta} N^2 {\beta_2}^2 - \frac{c_1^{T_1+2}}{2} NM\alpha_1 - \frac{c_2^{T_1+2}}{2} N^2 \alpha_2$.

According to the inequality of norms squared differences, we have

$$\begin{aligned}
&\mathbb{E}[\Psi^t] - \mathbb{E}[\tilde{\Psi}^t] \\
=& \frac{1}{NM} \sum_{i=1}^{N} \sum_{l=1}^{|\mathcal{P}_i|} (||(\mathcal{G}^t)_{\lambda_{i,j}}||^2 - ||(\tilde{\mathcal{G}}^t)_{\lambda_{i,j}}||^2) + \frac{1}{N^2} \sum_{i=1}^{N} \sum_{j=1}^{N} (||(\mathcal{G}^t)_{\boldsymbol{\theta}_{i,j}}||^2 - ||(\tilde{\mathcal{G}}^t)_{\boldsymbol{\theta}_{i,j}}||^2) \\
\leq& \frac{1}{NM} \sum_{i=1}^{N} \sum_{l=1}^{|\mathcal{P}_i|} ||c_1^{t-1} \lambda_{i,l}^t||^2 + \frac{1}{N^2} \sum_{i=1}^{N} \sum_{j=1}^{N} ||c_2^{t-1} \boldsymbol{\theta}_{i,j}^t||^2,
\end{aligned} \tag{S121}$$

and

$$\begin{aligned}
&\frac{1}{T-1} \sum_{t=T_1+2}^{T_1+T} \mathbb{E}[\Psi^t] \\
\leq& \frac{1}{T-1} \sum_{t=T_1+2}^{T_1+T} \mathbb{E}[\tilde{\Psi}^t] + \frac{1}{T-1} \sum_{t=T_1+2}^{T_1+T} \mathbb{E}\Big[\frac{1}{MN} \sum_{i=1}^{N} \sum_{l=1}^{|\mathcal{P}_i^t|} ||c_1^{t-1} \lambda_{i,l}^t||^2 + \frac{1}{N^2} \sum_{i=1}^{N} \sum_{j=1}^{N} ||c_2^{t-1} \boldsymbol{\theta}_{i,j}^t||^2) \Big) \Big] \\
\leq& \frac{1}{(T-1)} \Big(\frac{F^{T_1+2} - \underline{L}}{N^2 M} + \frac{c_1^1 \alpha_1}{2N} + \frac{c_2^1 \alpha_2}{2M} + \frac{4}{N\eta} \big(\frac{c_1^0}{c_1^1} + \frac{c_1^1}{c_1^2} \big) \alpha_1 \Big) \big(\max\{ \frac{1024M}{\eta}, \frac{2500}{\eta} \} \big) \\
&+ \frac{1}{(T-1)} \Big(\frac{4}{M\eta} \big(\frac{c_2^0}{c_2^1} + \frac{c_2^1}{c_2^2} \big) \alpha_2 + 4(c_1^1)^2 \alpha_1 + 4(c_2^1)^2 \alpha_2 + \frac{6}{N\eta} \beta_1^2 + \frac{6}{M\eta} {\beta_2}^2 \Big) \big(\max\{ \frac{1024M}{\eta}, \frac{2500}{\eta} \} \big) \\
&+ \Big(42\sigma_1^2 + 18\varsigma_1^2 + 42\sigma_2^2 + 18\varsigma_2^2 + 4\sigma_3^2 + 4\sigma_4^2 + \frac{2(6\sigma_1^2 + 3\varsigma_1^2)}{M(1-\rho)^2} + \frac{(82 + (\rho + \frac{2}{1-\rho})\frac{13}{512})(6\sigma_2^2 + 3\varsigma_2^2)}{M(1-\rho)^2} + \frac{\sigma_2^2}{M} \Big) (\max\{1024M, 2500\}) \\
&+ \Big(\frac{\alpha_1 + \alpha_2}{\eta^2 (T_1 + 2)} \Big).
\end{aligned} \tag{S122}$$

For brevity, we denote

$$\begin{aligned}
d =& \Big(\frac{F^{T_1+2} - \underline{L}}{N^2 M} + \frac{c_1^1 \alpha_1}{2N} + \frac{c_2^1 \alpha_2}{2M} + \frac{4}{N\eta} \big(\frac{c_1^0}{c_1^1} + \frac{c_1^1}{c_1^2} \big) \alpha_1 + \frac{4}{M\eta} \big(\frac{c_2^0}{c_2^1} + \frac{c_2^1}{c_2^2} \big) \alpha_2 + 4(c_1^1)^2 \alpha_1 + 4(c_2^1)^2 \alpha_2 \\
&+ \frac{6}{N\eta} \beta_1^2 + \frac{6}{M\eta} {\beta_2}^2 \Big) \big(\max\{ \frac{1024M}{\eta}, \frac{2500}{\eta} \} \big),
\end{aligned} \tag{S123}$$

and

$$C_{bias} = (42\sigma_1^2 + 18\varsigma_1^2 + 42\sigma_2^2 + 18\varsigma_2^2 + 4\sigma_3^2 + 4\sigma_4^2 + \frac{2(6\sigma_1^2 + 3\varsigma_1^2)}{M(1-\rho)^2} + \frac{(82 + (\rho + \frac{2}{1-\rho})\frac{13}{512})(6\sigma_2^2 + 3\varsigma_2^2)}{M(1-\rho)^2} + \frac{\sigma_2^2}{M})(\max\{1024M, 2500\}) + (\frac{\alpha_1 + \alpha_2}{\eta^2(T_1 + 2)}). \tag{S124}$$

References

[1] Mahmoud Assran, Nicolas Loizou, Nicolas Ballas, and Mike Rabbat. Stochastic gradient push for distributed deep learning. In *International Conference on Machine Learning*, pages 344–353. PMLR, 2019. 3

[2] Eunjeong Jeong, Matteo Zecchin, and Marios Kountouris. Asynchronous decentralized learning over unreliable wireless networks. In *ICC 2022-IEEE International Conference on Communications*, pages 607–612. IEEE, 2022. 3, 4

[3] Kaiyi Ji, Junjie Yang, and Yingbin Liang. Bilevel optimization: Convergence analysis and enhanced design. In *International conference on machine learning*, pages 4882–4892. PMLR, 2021. 3

[4] Yang Jiao, Kai Yang, Tiancheng Wu, Dongjin Song, and Chengtao Jian. Asynchronous distributed bilevel optimization. In *The Eleventh International Conference on Learning Representations*, 2023. 2, 3

[5] Dmitry Kovalev, Egor Shulgin, Peter Richtárik, Alexander V Rogozin, and Alexander Gasnikov. Adom: Accelerated decentralized optimization method for time-varying networks. In *International Conference on Machine Learning*, pages 5784–5793. PMLR, 2021. 3

[6] Xiangru Lian, Wei Zhang, Ce Zhang, and Ji Liu. Asynchronous decentralized parallel stochastic gradient descent. In *International Conference on Machine Learning*, pages 3043–3052. PMLR, 2018. 3

[7] Gabriel Mancino-ball, Shengnan Miao, Yangyang Xu, and Jie Chen. Proximal stochastic recursive momentum methods for nonconvex composite decentralized optimization. In *AAAI Conference on Artificial Intelligence*, 2023. 2, 3

[8] Qi Qian, Shenghuo Zhu, Jiasheng Tang, Rong Jin, Baigui Sun, and Hao Li. Robust optimization over multiple domains. In *Proceedings of the AAAI Conference on Artificial Intelligence*, volume 33, pages 4739–4746, 2019. 3