Suplementary Content

Appendix A
Theorem 1

Theorem 1: (Iteration Complexity) Let step-sizes $\alpha_{\boldsymbol{\omega}} = \alpha_{\epsilon} = \alpha_{\boldsymbol{z}} = \frac{2}{L+\alpha_\lambda ML^2+\alpha_\phi ML^2+8M\beta L^2\left(\frac{1}{\alpha_\lambda \underline{c}_1{}^2}+\frac{1}{\alpha_\phi \underline{c}_2{}^2}\right)}$, $\alpha_\lambda < \min\left\{\frac{2}{L+2c_1^0}, \frac{1}{30\varkappa k_0 NL^2}\right\}$ and $\alpha_\phi \le \frac{2}{L+2c_2^0}$. Let Assumptions 1 and 2 hold. Then, the iteration complexity of our proposed algorithm to obtain Υ-stationary point is bounded by,

$$T(\Upsilon) \sim \mathcal{O}\Big(\max\Big\{ \frac{16M^2}{\Upsilon^2}\big(\frac{\mu_3}{{\alpha_\lambda}^2}+\frac{\mu_4}{{\alpha_\phi}^2}\big)^2, \Big(\frac{\big(4c_6+2\alpha_\phi L^2(M-S)\big)\left(\bar{c}+k_c\varkappa(\varkappa-1)\right)c_5}{\Upsilon}+\sqrt{2}\Big)^2\Big\}\Big), \tag{1}$$

where $\mu_3, \mu_4, \beta, \bar{c}, c_6$ and c_7 are constants.

Appendix B
Definitions and Assumptions

We present the detailed theoretical analysis on the convergence of BAFDP algorithm. Before proceeding with detailed proof, we first give the following definitions and assumptions.

Definition 1: In the asynchronous distributed algorithm, let $\hat{t}$ denote the last iteration when the client was activated and $\tilde{t}$ represents the next iteration when the client will be activated.

Definition 2: Following [1], the stationarity gap *w.r.t* $\mathcal{L}_\lambda(\{\boldsymbol{\omega}_i\},\{\epsilon_i^t\},\boldsymbol{z},\{\lambda_i\},\{\phi_i\})$ at t-th iteration is defined as:

$$\nabla F^t = \begin{bmatrix} \{\nabla_{\boldsymbol{\omega}_i}\mathcal{L}_\lambda\left(\{\boldsymbol{\omega}_i^t\},\{\epsilon_i^t\},\boldsymbol{z}^t,\{\lambda_i^t\},\{\phi_i^t\}\right)\} \\ \{\nabla_{\epsilon_i^t}\mathcal{L}_\lambda\left(\{\boldsymbol{\omega}_i^t\},\{\epsilon_i^t\},\boldsymbol{z}^t,\{\lambda_i^t\},\{\phi_i^t\}\right)\} \\ \nabla_{\boldsymbol{z}}\mathcal{L}_\lambda\left(\{\boldsymbol{\omega}_i^t\},\{\epsilon_i^t\},\boldsymbol{z}^t,\{\lambda_i^t\},\{\phi_i^t\}\right) \\ \{\nabla_{\lambda_i}\mathcal{L}_\lambda\left(\{\boldsymbol{\omega}_i^t\},\{\epsilon_i^t\},\boldsymbol{z}^t,\{\lambda_i^t\},\{\phi_i^t\}\right)\} \\ \{\nabla_{\phi_i}\mathcal{L}_\lambda\left(\{\boldsymbol{\omega}_i^t\},\{\epsilon_i^t\},\boldsymbol{z}^t,\{\lambda_i^t\},\{\phi_i^t\}\right)\} \end{bmatrix}, \tag{2}$$

where ∇F^t is the simplified form of $\nabla F^t\left(\{\boldsymbol{\omega}_i^t\},\{\epsilon_i^t\},\boldsymbol{z}^t,\{\lambda_i^t\},\{\phi_i^t\}\right)$. And for notational simplicity, we define that,

$$\begin{aligned} (\nabla F^t)_{\boldsymbol{\omega}_i} &= \nabla_{\boldsymbol{\omega}_i}\mathcal{L}_\lambda\left(\{\boldsymbol{\omega}_i^t\},\{\epsilon_i^t\},\boldsymbol{z}^t,\{\lambda_i^t\},\{\phi_i^t\}\right), \\ (\nabla F^t)_{\epsilon_i} &= \nabla_{\epsilon_i}\mathcal{L}_\lambda\left(\{\boldsymbol{\omega}_i^t\},\{\epsilon_i^t\},\boldsymbol{z}^t,\{\lambda_i^t\},\{\phi_i^t\}\right), \\ (\nabla F^t)_{\boldsymbol{z}} &= \nabla_{\boldsymbol{z}}\mathcal{L}_\lambda\left(\{\boldsymbol{\omega}_i^t\},\{\epsilon_i^t\},\boldsymbol{z}^t,\{\lambda_i^t\},\{\phi_i^t\}\right), \\ (\nabla F^t)_{\lambda_i} &= \nabla_{\lambda_i}\mathcal{L}_\lambda\left(\{\boldsymbol{\omega}_i^t\},\{\epsilon_i^t\},\boldsymbol{z}^t,\{\lambda_i^t\},\{\phi_i^t\}\right), \\ (\nabla F^t)_{\phi_i} &= \nabla_{\phi_i}\mathcal{L}_\lambda\left(\{\boldsymbol{\omega}_i^t\},\{\epsilon_i^t\},\boldsymbol{z}^t,\{\lambda_i^t\},\{\phi_i^t\}\right), \end{aligned} \tag{3}$$

which follows that,

$$\left\|\nabla F^t\right\|^2 = \sum_{i=1}^{M}\left(\left\|(\nabla F^t)_{\boldsymbol{\omega}_i}\right\|^2+\left\|(\nabla F^t)_{\epsilon_i}\right\|^2\right)+\left\|(\nabla F^t)_{\boldsymbol{z}}\right\|^2+\sum_{i=1}^{M}\left(\left\|(\nabla F^t)_{\lambda_i}\right\|^2+\left\|(\nabla F^t)_{\phi_i}\right\|^2\right). \tag{4}$$

Definition 3: The stationarity gap *w.r.t* $\bar{\mathcal{L}}_\lambda(\{\boldsymbol{\omega}_i\},\{\epsilon_i^t\},\boldsymbol{z},\{\lambda_i\},\{\phi_i\})$ at t-th iteration is defined as,

$$\nabla \bar{F}^t = \begin{bmatrix} \{\nabla_{\boldsymbol{\omega}_i}\bar{\mathcal{L}}_\lambda\left(\{\boldsymbol{\omega}_i^t\},\{\epsilon_i^t\},\boldsymbol{z}^t,\{\lambda_i^t\},\{\phi_i^t\}\right)\} \\ \{\nabla_{\epsilon_i^t}\bar{\mathcal{L}}_\lambda\left(\{\boldsymbol{\omega}_i^t\},\{\epsilon_i^t\},\boldsymbol{z}^t,\{\lambda_i^t\},\{\phi_i^t\}\right)\} \\ \nabla_{\boldsymbol{z}}\bar{\mathcal{L}}_\lambda\left(\{\boldsymbol{\omega}_i^t\},\{\epsilon_i^t\},\boldsymbol{z}^t,\{\lambda_i^t\},\{\phi_i^t\}\right) \\ \{\nabla_{\lambda_i}\bar{\mathcal{L}}_\lambda\left(\{\boldsymbol{\omega}_i^t\},\{\epsilon_i^t\},\boldsymbol{z}^t,\{\lambda_i^t\},\{\phi_i^t\}\right)\} \\ \{\nabla_{\phi_i}\bar{\mathcal{L}}_\lambda\left(\{\boldsymbol{\omega}_i^t\},\{\epsilon_i^t\},\boldsymbol{z}^t,\{\lambda_i^t\},\{\phi_i^t\}\right)\} \end{bmatrix}, \tag{5}$$

where $\nabla \bar{F}^t$ is the simplified form of $\nabla \bar{F}^t\left(\{\boldsymbol{\omega}_i^t\},\{\epsilon_i^t\},\boldsymbol{z}^t,\{\lambda_i^t\},\{\phi_i^t\}\right)$. Also, we define that,

$$\begin{aligned} (\nabla \bar{F}^t)_{\boldsymbol{\omega}_i} &= \nabla_{\boldsymbol{\omega}_i}\bar{\mathcal{L}}_\lambda\left(\{\boldsymbol{\omega}_i^t\},\{\epsilon_i^t\},\boldsymbol{z}^t,\{\lambda_i^t\},\{\phi_i^t\}\right), \\ (\nabla \bar{F}^t)_{\epsilon_i} &= \nabla_{\epsilon_i}\bar{\mathcal{L}}_\lambda\left(\{\boldsymbol{\omega}_i^t\},\{\epsilon_i^t\},\boldsymbol{z}^t,\{\lambda_i^t\},\{\phi_i^t\}\right), \\ (\nabla \bar{F}^t)_{\boldsymbol{z}} &= \nabla_{\boldsymbol{z}}\bar{\mathcal{L}}_\lambda\left(\{\boldsymbol{\omega}_i^t\},\{\epsilon_i^t\},\boldsymbol{z}^t,\{\lambda_i^t\},\{\phi_i^t\}\right), \\ (\nabla \bar{F}^t)_{\lambda_i} &= \nabla_{\lambda_i}\bar{\mathcal{L}}_\lambda\left(\{\boldsymbol{\omega}_i^t\},\{\epsilon_i^t\},\boldsymbol{z}^t,\{\lambda_i^t\},\{\phi_i^t\}\right), \\ (\nabla \bar{F}^t)_{\phi_i} &= \nabla_{\phi_i}\bar{\mathcal{L}}_\lambda\left(\{\boldsymbol{\omega}_i^t\},\{\epsilon_i^t\},\boldsymbol{z}^t,\{\lambda_i^t\},\{\phi_i^t\}\right), \end{aligned} \tag{6}$$

which follows that,

$$\left\|\nabla \bar{F}^t\right\|^2=\sum_{i=1}^{M}\left(\left\|\left(\nabla \bar{F}^t\right)_{\boldsymbol{\omega}_i}\right\|^2+\left\|\left(\nabla \bar{F}^t\right)_{\epsilon_i}\right\|^2\right)+\left\|\left(\nabla \bar{F}^t\right)_{\boldsymbol{z}}\right\|^2+\sum_{i=1}^{M}\left(\left\|\left(\nabla \bar{F}^t\right)_{\lambda_i}\right\|^2+\left\|\left(\nabla \bar{F}^t\right)_{\phi_i}\right\|^2\right). \tag{7}$$

Assumption 1: (Gradient Lipschitz [2]) We assume that loss function $\mathcal{L}(\cdot)$ has Lipschitz continuous gradients, i.e., for any two samples ξ_1,ξ_2, we assume that there exists $G(\boldsymbol{\omega})>0$ satisfying that,

$$|\mathcal{L}(\xi_1;\boldsymbol{\omega})-\mathcal{L}(\xi_2;\boldsymbol{\omega})|\le G(\boldsymbol{\omega})\cdot\|\xi_1-\xi_2\|. \tag{8}$$

Assumption 2: (Boundedness) Building on our previous studies [3], [4], we assume that variables have boundedness, i.e., $\|\boldsymbol{\omega}_i\|^2\le\mu_1$, $\|\epsilon_i\|^2\le\mu_2$, $\|\boldsymbol{z}\|^2\le\mu_1$, $\|\lambda_i\|^2\le\mu_3$, $\|\phi_i^t\|^2\le\mu_4$. Additionally, we assume that before obtaining the Υ-stationary point, the variables in the server satisfy that $\|\boldsymbol{z}^{t+1}-\boldsymbol{z}^t\|^2+\|\lambda^{t+1}-\lambda^t\|^2\ge\nu$, where $\nu>0$ is a small constant. Moreover, for any $k\in\{1,...,\varkappa\}$, the change of the variables in the server is upper bounded within $\varkappa$ iterations, i.e., $\|\boldsymbol{z}^t-\boldsymbol{z}^{t-k}\|^2\le\varkappa k_0\nu$, $\sum_{i=1}^{M}\|\lambda_i^t-\lambda_i^{t-k}\|^2\le\varkappa k_0\nu$, where k_0 is a constant.

Setting 1: Let c_1^t,c_2^t be two nonnegative non-increasing sequences, i.e., $c_1^t=\frac{1}{\alpha_\lambda(t+1)^{\frac{1}{4}}}\ge\underline{c}_1, c_2^t=\frac{1}{\alpha_\phi(t+1)^{\frac{1}{4}}}\ge\underline{c}_2$, where $\alpha_\lambda,\alpha_\phi$ are stepsize. Also, $\underline{c}_1,\underline{c}_2$ satisfy that $0\le\underline{c}_1\le\frac{1}{\alpha_\lambda c_0}, 0\le\underline{c}_2\le\frac{1}{\alpha_\lambda c_0}$,$c=(\frac{16M^2}{\Upsilon^2}(\frac{\mu_3}{\alpha_\lambda{}^2}+\frac{\mu_4}{\alpha_\phi{}^2})^2+1)^{\frac{1}{4}}$.

Appendix C
Lemma 1

Lemma 1: We let Assumption 1 and 2 hold, and set $\alpha_{\boldsymbol{\omega}}=\alpha_\epsilon=\alpha_{\boldsymbol{z}}$=$\frac{2}{L+\alpha_\lambda ML^2+\alpha_\phi ML^2+8M\beta L^2\left(\frac{1}{\alpha_\lambda(a_1^t)^2}+\frac{1}{\alpha_\phi(c_2^t)^2}\right)}$. Then, we have,

$$\begin{aligned}&\mathcal{L}_\lambda\left(\{\boldsymbol{\omega}_i^{t+1}\},\{\epsilon_i^{t+1}\},\boldsymbol{z}^{t+1},\{\lambda_i^t\},\{\phi_i^t\}\right)-\mathcal{L}_\lambda\left(\{\boldsymbol{\omega}_i^t\},\{\epsilon_i^t\},\boldsymbol{z}^t,\{\lambda_i^t\},\{\phi_i^t\}\right)\\&\le\sum_{i=1}^{M}\left(\left(\frac{L+L^2+1}{2}-\frac{1}{\alpha_{\boldsymbol{\omega}}}\right)\left\|\boldsymbol{\omega}_i^{t+1}-\boldsymbol{\omega}_i^t\right\|^2+\left(\frac{L+1}{2}-\frac{1}{\alpha_\epsilon}\right)\left\|\epsilon_i^{t+1}-\epsilon_i^t\right\|^2+3ML^2\varkappa k_0\left\|\lambda_i^{t+1}-\lambda_i^t\right\|^2\right)\\&+\left(\frac{L+6ML^2\varkappa k_0+\psi}{2}-\frac{1}{\alpha_{\boldsymbol{z}}}\right)\left\|\boldsymbol{z}^{t+1}-\boldsymbol{z}^t\right\|^2+4\psi^2+2\psi.\end{aligned} \tag{9}$$

Proof of Lemma 1:

It follows from Assumption 1 that,

$$\begin{aligned}&\mathcal{L}_\lambda\left(\left\{\boldsymbol{\omega}_1^{t+1},\boldsymbol{\omega}_2^t,\cdots,\boldsymbol{\omega}_M^t\right\},\left\{\epsilon_i^t\right\},\boldsymbol{z}^t,\left\{\lambda_i^t\right\},\left\{\phi_i^t\right\}\right)-\mathcal{L}_\lambda\left(\left\{\boldsymbol{\omega}_i^t\right\},\left\{\epsilon_i^t\right\},\boldsymbol{z}^t,\left\{\lambda_i^t\right\},\left\{\phi_i^t\right\}\right)\\&\le\left\langle\nabla_{\boldsymbol{\omega}_1}\mathcal{L}_\lambda\left(\left\{\boldsymbol{\omega}_i^t\right\},\left\{\epsilon_i^t\right\},\boldsymbol{z}^t,\left\{\lambda_i^t\right\},\left\{\phi_i^t\right\}\right),\boldsymbol{\omega}_1^{t+1}-\boldsymbol{\omega}_1^t\right\rangle+\frac{L}{2}\left\|\boldsymbol{\omega}_1^{t+1}-\boldsymbol{\omega}_1^t\right\|^2,\\&\mathcal{L}_\lambda\left(\left\{\boldsymbol{\omega}_1^{t+1},\boldsymbol{\omega}_2^{t+1},\cdots,\boldsymbol{\omega}_M^t\right\},\left\{\epsilon_i^t\right\},\boldsymbol{z}^t,\left\{\lambda_i^t\right\},\left\{\phi_i^t\right\}\right)-\mathcal{L}_\lambda\left(\left\{\boldsymbol{\omega}_1^{t+1},\boldsymbol{\omega}_2^t,\cdots,\boldsymbol{\omega}_M^t\right\},\left\{\epsilon_i^t\right\},\boldsymbol{z}^t,\left\{\lambda_i^t\right\},\left\{\phi_i^t\right\}\right)\\&\le\left\langle\nabla_{\boldsymbol{\omega}_2}\mathcal{L}_\lambda\left(\left\{\boldsymbol{\omega}_i^t\right\},\left\{\epsilon_i^t\right\},\boldsymbol{z}^t,\left\{\lambda_i^t\right\},\left\{\phi_i^t\right\}\right),\boldsymbol{\omega}_2^{t+1}-\boldsymbol{\omega}_2^t\right\rangle+\frac{L}{2}\left\|\boldsymbol{\omega}_2^{t+1}-\boldsymbol{\omega}_2^t\right\|^2,\\&\qquad\vdots\\&\mathcal{L}_\lambda\left(\left\{\boldsymbol{\omega}_i^{t+1}\right\},\left\{\epsilon_i^t\right\},\boldsymbol{z}^t,\left\{\lambda_i^t\right\},\left\{\phi_i^t\right\}\right)-\mathcal{L}_\lambda\left(\left\{\boldsymbol{\omega}_1^{t+1},\cdots,\boldsymbol{\omega}_{M-1}^{t+1},\boldsymbol{\omega}_M^t\right\},\left\{\epsilon_i^t\right\},\boldsymbol{z}^t,\left\{\lambda_i^t\right\},\left\{\phi_i^t\right\}\right)\\&\le\left\langle\nabla_{\boldsymbol{\omega}_M}\mathcal{L}_\lambda\left(\left\{\boldsymbol{\omega}_i^t\right\},\left\{\epsilon_i^t\right\},\boldsymbol{z}^t,\left\{\lambda_i^t\right\},\left\{\phi_i^t\right\}\right),\boldsymbol{\omega}_M^{t+1}-\boldsymbol{\omega}_M^t\right\rangle+\frac{L}{2}\left\|\boldsymbol{\omega}_M^{t+1}-\boldsymbol{\omega}_M^t\right\|^2.\end{aligned} \tag{10}$$

We sum up the above inequalities in Eq. (10),

$$\begin{aligned}&\mathcal{L}_\lambda\left(\left\{\boldsymbol{\omega}_i^{t+1}\right\},\left\{\epsilon_i^t\right\},\boldsymbol{z}^t,\left\{\lambda_i^t\right\},\left\{\phi_i^t\right\}\right)-\mathcal{L}_\lambda\left(\left\{\boldsymbol{\omega}_i^t\right\},\left\{\epsilon_i^t\right\},\boldsymbol{z}^t,\left\{\lambda_i^t\right\},\left\{\phi_i^t\right\}\right)\\&\le\sum_{i=1}^{M}\left(\left\langle\nabla_{\boldsymbol{\omega}_i}\mathcal{L}_\lambda\left(\left\{\boldsymbol{\omega}_i^t\right\},\left\{\epsilon_i^t\right\},\boldsymbol{z}^t,\left\{\lambda_i^t\right\},\left\{\phi_i^t\right\}\right),\boldsymbol{\omega}_i^{t+1}-\boldsymbol{\omega}_i^t\right\rangle+\frac{L}{2}\left\|\boldsymbol{\omega}_i^{t+1}-\boldsymbol{\omega}_i^t\right\|^2\right).\end{aligned} \tag{11}$$

$$\boldsymbol{\omega}_i^{t+1}=\boldsymbol{\omega}_i^t-\alpha_{\boldsymbol{\omega}}\left(\nabla_{\boldsymbol{\omega}_i}\bar{\mathcal{L}}_\lambda(\{\boldsymbol{\omega}_i^t\},\{\epsilon_i^t\},\boldsymbol{z}^t,\{\lambda_i^t\},\{\phi_i^t\}\right)+\psi\text{sign}(\boldsymbol{z}_i^t-\boldsymbol{\omega}_i^t)). \tag{12}$$

Assuming $\nabla_{\boldsymbol{\omega}_i}\mathcal{L}_\lambda\left(\left\{\boldsymbol{\omega}_i^{\hat{t}}\right\},\left\{\epsilon_i^{\hat{t}}\right\},\boldsymbol{z}^{\hat{t}},\left\{\lambda_i^{\hat{t}}\right\},\left\{\phi_i^{\hat{t}}\right\}\right)=\nabla_{\boldsymbol{\omega}_i}\bar{\mathcal{L}}_\lambda\left(\left\{\boldsymbol{\omega}_i^{\hat{t}}\right\},\left\{\epsilon_i^{\hat{t}}\right\},\boldsymbol{z}^{\hat{t}},\left\{\lambda_i^{\hat{t}}\right\},\left\{\phi_i^{\hat{t}}\right\}\right)$ with Eq. (12), we have,

$$\left\langle\boldsymbol{\omega}_i^t-\boldsymbol{\omega}_i^{t+1},\boldsymbol{\omega}_i^{t+1}-\boldsymbol{\omega}_i^t+\alpha_{\boldsymbol{\omega}}\left(\nabla_{\boldsymbol{\omega}_i}\mathcal{L}_\lambda\left(\left\{\boldsymbol{\omega}_i^{\hat{t}}\right\},\left\{\epsilon_i^{\hat{t}}\right\},\boldsymbol{z}^{\hat{t}},\left\{\lambda_i^{\hat{t}}\right\},\left\{\phi_i^{\hat{t}}\right\}\right)+\psi\text{sign}(\boldsymbol{z}_i^{\hat{t}}-\boldsymbol{\omega}_i^{\hat{t}})\right)\right\rangle\ge0. \tag{13}$$

Based on Eq. (13), we have,

$$\left\langle \boldsymbol{\omega}_i^{t+1}-\boldsymbol{\omega}_i^t, \nabla_{\boldsymbol{\omega}_i}\mathcal{L}_\lambda\left(\left\{\boldsymbol{\omega}_i^{\hat{t}}\right\},\left\{\epsilon_i^{\hat{t}}\right\},\boldsymbol{z}^{\hat{t}},\left\{\lambda_i^{\hat{t}}\right\},\left\{\phi_i^{\hat{t}}\right\}\right)+\psi\mathrm{sign}(\boldsymbol{z}_i^{\hat{t}}-\boldsymbol{\omega}_i^{\hat{t}})\right\rangle \le -\frac{1}{\alpha_{\boldsymbol{\omega}}}\left\|\boldsymbol{\omega}_i^{t+1}-\boldsymbol{\omega}_i^t\right\|^2. \tag{14}$$

Next, according to the Cauchy-Schwarz inequality, Assumption 1 and 2, we have,

$$\begin{aligned}
&\left\langle \boldsymbol{\omega}_i^{t+1}-\boldsymbol{\omega}_i^t, \nabla_{\boldsymbol{\omega}_i}\mathcal{L}_\lambda\left(\left\{\boldsymbol{\omega}_i^t\right\},\left\{\epsilon_i^t\right\},\boldsymbol{z}^t,\left\{\lambda_i^t\right\},\left\{\phi_i^t\right\}\right)+\psi\mathrm{sign}(\boldsymbol{z}_i^t-\boldsymbol{\omega}_i^t)-\nabla_{\boldsymbol{\omega}_i}\mathcal{L}_\lambda\left(\left\{\boldsymbol{\omega}_i^{\hat{t}}\right\},\left\{\epsilon_i^{\hat{t}}\right\},\boldsymbol{z}^{\hat{t}},\left\{\lambda_i^{\hat{t}}\right\},\left\{\phi_i^{\hat{t}}\right\}\right)-\psi\mathrm{sign}(\boldsymbol{z}_i^{\hat{t}}-\boldsymbol{\omega}_i^{\hat{t}})\right\rangle\\
&\le \frac{1}{2}\left\|\boldsymbol{\omega}_i^{t+1}-\boldsymbol{\omega}_i^t\right\|^2+\frac{1}{2}\Big\|\nabla_{\boldsymbol{\omega}_i}\mathcal{L}_\lambda\left(\left\{\boldsymbol{\omega}_i^t\right\},\left\{\epsilon_i^t\right\},\boldsymbol{z}^t,\left\{\lambda_i^t\right\},\left\{\phi_i^t\right\}\right)-\nabla_{\boldsymbol{\omega}_i}\mathcal{L}_\lambda\left(\left\{\boldsymbol{\omega}_i^{\hat{t}}\right\},\left\{\epsilon_i^{\hat{t}}\right\},\boldsymbol{z}^{\hat{t}},\left\{\lambda_i^{\hat{t}}\right\},\left\{\phi_i^{\hat{t}}\right\}\right)\\
&\quad+\psi\left(\mathrm{sign}(\boldsymbol{z}_i^t-\boldsymbol{\omega}_i^t)-\mathrm{sign}(\boldsymbol{z}_i^{\hat{t}}-\boldsymbol{\omega}_i^{\hat{t}})\right)\Big\|^2\\
&\le \frac{1}{2}\left\|\boldsymbol{\omega}_i^{t+1}-\boldsymbol{\omega}_i^t\right\|^2+\Big\|\nabla_{\boldsymbol{\omega}_i}\mathcal{L}_\lambda\left(\left\{\boldsymbol{\omega}_i^t\right\},\left\{\epsilon_i^t\right\},\boldsymbol{z}^t,\left\{\lambda_i^t\right\},\left\{\phi_i^t\right\}\right)-\nabla_{\boldsymbol{\omega}_i}\mathcal{L}_\lambda\left(\left\{\boldsymbol{\omega}_i^{\hat{t}}\right\},\left\{\epsilon_i^{\hat{t}}\right\},\boldsymbol{z}^{\hat{t}},\left\{\lambda_i^{\hat{t}}\right\},\left\{\phi_i^{\hat{t}}\right\}\right)\Big\|^2\\
&\quad+\psi^2\left(\mathrm{sign}(\boldsymbol{z}_i^t-\boldsymbol{\omega}_i^t)-\mathrm{sign}(\boldsymbol{z}_i^{\hat{t}}-\boldsymbol{\omega}_i^{\hat{t}})\right)^2\\
&\le \frac{1}{2}\left\|\boldsymbol{\omega}_i^{t+1}-\boldsymbol{\omega}_i^t\right\|^2+\frac{L^2}{2}\Big(\left\|\boldsymbol{z}^t-\boldsymbol{z}^{\hat{t}}\right\|^2+\sum_{i=1}^{M}\left\|\lambda_i^t-\lambda_i^{\hat{t}}\right\|^2\Big)+4\psi^2\\
&\le \frac{1}{2}\left\|\boldsymbol{\omega}_i^{t+1}-\boldsymbol{\omega}_i^t\right\|^2+\frac{3L^2\varkappa k_0}{2}\Big(\left\|\boldsymbol{z}^{t+1}-\boldsymbol{z}^t\right\|^2+\sum_{i=1}^{M}\left\|\lambda_i^{t+1}-\lambda_i^t\right\|^2\Big)+4\psi^2.
\end{aligned} \tag{15}$$

With the combination of Eq. (11), (14), and (15), we have,

$$\begin{aligned}
&\mathcal{L}_\lambda\big(\left\{\boldsymbol{\omega}_i^{t+1}\right\},\left\{\epsilon_i^t\right\},\boldsymbol{z}^t,\left\{\lambda_i^t\right\},\left\{\phi_i^t\right\}\big)-\mathcal{L}_\lambda\left(\left\{\boldsymbol{\omega}_i^t\right\},\left\{\epsilon_i^t\right\},\boldsymbol{z}^t,\left\{\lambda_i^t\right\},\left\{\phi_i^t\right\}\right)\\
&\quad\le\sum_{i=1}^{M}\big(\frac{L+1}{2}-\frac{1}{\alpha_{\boldsymbol{\omega}}}\big)\left\|\boldsymbol{\omega}_i^{t+1}-\boldsymbol{\omega}_i^t\right\|^2+\frac{3ML^2\varkappa k_0}{2}\big(\left\|\boldsymbol{z}^{t+1}-\boldsymbol{z}^t\right\|^2+\sum_{i=1}^{M}\left\|\lambda_i^{t+1}-\lambda_i^t\right\|^2\big)+4\psi^2.
\end{aligned} \tag{16}$$

According to the Lipschitz properties in Assumption 1, we have,

$$\begin{aligned}
&\mathcal{L}_\lambda\left(\left\{\boldsymbol{\omega}_i^{t+1}\right\},\left\{\epsilon_i^{t+1}\right\},\boldsymbol{z}^t,\left\{\lambda_i^t\right\},\left\{\phi_i^t\right\}\right)-\mathcal{L}_\lambda\left(\left\{\boldsymbol{\omega}_i^{t+1}\right\},\left\{\epsilon_i^t\right\},\boldsymbol{z}^t,\left\{\lambda_i^t\right\},\left\{\phi_i^t\right\}\right)\\
&\le\sum_{i=1}^{M}\Big(\left\langle\nabla_{\epsilon_i}\mathcal{L}_\lambda\left(\left\{\boldsymbol{\omega}_i^{t+1}\right\},\left\{\epsilon_i^t\right\},\boldsymbol{z}^t,\left\{\lambda_i^t\right\},\left\{\phi_i^t\right\}\right),\epsilon_i^{t+1}-\epsilon_i^t\right\rangle+\frac{L}{2}\left\|\epsilon_i^{t+1}-\epsilon_i^t\right\|^2\Big),
\end{aligned} \tag{17}$$

$$\epsilon_i^{t+1}=\epsilon_i^t-\alpha_\epsilon\nabla_{\epsilon_i}\bar{\mathcal{L}}_\lambda(\{\boldsymbol{\omega}_i^{t+1}\},\{\epsilon_i^t\},\boldsymbol{z}^t,\{\lambda_i^t\},\{\phi_i^t\}). \tag{18}$$

According to $\nabla_{\epsilon_i}\mathcal{L}_\lambda\left(\left\{\boldsymbol{\omega}_i^{\hat{t}}\right\},\left\{\epsilon_i^{\hat{t}}\right\},\boldsymbol{z}^{\hat{t}},\left\{\lambda_i^{\hat{t}}\right\},\left\{\phi_i^{\hat{t}}\right\}\right)=\nabla_{\epsilon_i}\bar{\mathcal{L}}_\lambda\left(\left\{\boldsymbol{\omega}_i^{\hat{t}}\right\},\left\{\epsilon_i^{\hat{t}}\right\},\boldsymbol{z}^{\hat{t}},\left\{\lambda_i^{\hat{t}}\right\},\left\{\phi_i^{\hat{t}}\right\}\right)$ and the optimal condition for Eq. (18) in our study, we have,

$$\left\langle\epsilon_i^t-\epsilon_i^{t+1},\epsilon_i^{t+1}-\epsilon_i^t+\alpha_\epsilon\nabla_{\epsilon_i}\mathcal{L}_\lambda\left(\left\{\boldsymbol{\omega}_i^{\hat{t}}\right\},\left\{\epsilon_i^{\hat{t}}\right\},\boldsymbol{z}^{\hat{t}},\left\{\lambda_i^{\hat{t}}\right\},\left\{\phi_i^{\hat{t}}\right\}\right)\right\rangle\ge 0. \tag{19}$$

Combining Eq. (17) with Eq. (19), we have,

$$\left\langle\epsilon_i^{t+1}-\epsilon_i^t,\nabla_{\epsilon_i}\mathcal{L}_\lambda\left(\left\{\boldsymbol{\omega}_i^{\hat{t}}\right\},\left\{\epsilon_i^{\hat{t}}\right\},\boldsymbol{z}^{\hat{t}},\left\{\lambda_i^{\hat{t}}\right\},\left\{\phi_i^{\hat{t}}\right\}\right)\right\rangle\le-\frac{1}{\alpha_\epsilon}\left\|\epsilon_i^{t+1}-\epsilon_i^t\right\|^2. \tag{20}$$

According to the Cauchy-Schwarz inequality, Assumption 1 and 2, we have,

$$\begin{aligned}
&\left\langle\epsilon_i^{t+1}-\epsilon_i^t,\nabla_{\epsilon_i}\mathcal{L}_\lambda\left(\left\{\boldsymbol{\omega}_i^{t+1}\right\},\left\{\epsilon_i^t\right\},\boldsymbol{z}^t,\left\{\lambda_i^t\right\},\left\{\phi_i^t\right\}\right)-\nabla_{\epsilon_i}\mathcal{L}_\lambda\left(\left\{\boldsymbol{\omega}_i^{\hat{t}}\right\},\left\{\epsilon_i^{\hat{t}}\right\},\boldsymbol{z}^{\hat{t}},\left\{\lambda_i^{\hat{t}}\right\},\left\{\phi_i^{\hat{t}}\right\}\right)\right\rangle\\
&\le\frac{1}{2}\left\|\epsilon_i^{t+1}-\epsilon_i^t\right\|^2+\frac{L^2}{2}\Big(\left\|\boldsymbol{\omega}_i^{t+1}-\boldsymbol{\omega}_i^t\right\|^2+\left\|\boldsymbol{z}^t-\boldsymbol{z}^{\hat{t}}\right\|^2+\sum_{i=1}^{M}\left\|\lambda_i^t-\lambda_i^{\hat{t}}\right\|^2\Big)\\
&\le\frac{1}{2}\left\|\epsilon_i^{t+1}-\epsilon_i^t\right\|^2+\frac{L^2}{2}\left\|\boldsymbol{\omega}_i^{t+1}-\boldsymbol{\omega}_i^t\right\|^2+\frac{3L^2\varkappa k_0}{2}\Big(\left\|\boldsymbol{z}^{t+1}-\boldsymbol{z}^t\right\|^2+\sum_{i=1}^{M}\left\|\lambda_i^{t+1}-\lambda_i^t\right\|^2\Big).
\end{aligned} \tag{21}$$

Thus, combining Eq. (17), (20) with (21), we have,

$$\begin{aligned}
&\mathcal{L}_\lambda\left(\left\{\boldsymbol{\omega}_i^{t+1}\right\},\left\{\epsilon_i^{t+1}\right\},\boldsymbol{z}^t,\left\{\lambda_i^t\right\},\left\{\phi_i^t\right\}\right)-\mathcal{L}_\lambda\left(\left\{\boldsymbol{\omega}_i^{t+1}\right\},\left\{\epsilon_i^t\right\},\boldsymbol{z}^t,\left\{\lambda_i^t\right\},\left\{\phi_i^t\right\}\right)\\
&\le\sum_{i=1}^{M}\big(\frac{L+1}{2}-\frac{1}{\alpha_\epsilon}\big)\left\|\epsilon_i^{t+1}-\epsilon_i^t\right\|^2+\sum_{i=1}^{M}\frac{L^2}{2}\left\|\boldsymbol{\omega}_i^{t+1}-\boldsymbol{\omega}_i^t\right\|^2+\frac{3ML^2\varkappa k_0}{2}\big(\left\|\boldsymbol{z}^{t+1}-\boldsymbol{z}^t\right\|^2+\sum_{i=1}^{M}\left\|\lambda_i^{t+1}-\lambda_i^t\right\|^2\big).
\end{aligned} \tag{22}$$

Likewise, according to Assumption 1 and Eq. (18), we have,

$$\begin{aligned}&\mathcal{L}_\lambda\left(\left\{\boldsymbol{\omega}_i^{t+1}\right\},\left\{\epsilon_i^{t+1}\right\},\boldsymbol{z}^{t+1},\left\{\lambda_i^t\right\},\left\{\phi_i^t\right\}\right)-\mathcal{L}_\lambda\left(\left\{\boldsymbol{\omega}_i^{t+1}\right\},\left\{\epsilon_i^{t+1}\right\},\boldsymbol{z}^{t},\left\{\lambda_i^t\right\},\left\{\phi_i^t\right\}\right)\\&\leq\left\langle\nabla_{\boldsymbol{z}}\mathcal{L}_\lambda\left(\left\{\boldsymbol{\omega}_i^{t+1}\right\},\left\{\epsilon_i^{t+1}\right\},\boldsymbol{z}^{t},\left\{\lambda_i^t\right\},\left\{\phi_i^t\right\}\right),\boldsymbol{z}^{t+1}-\boldsymbol{z}^t\right\rangle+\frac{L}{2}\left\|\boldsymbol{z}^{t+1}-\boldsymbol{z}^t\right\|^2,\end{aligned}\tag{23}$$

$$\boldsymbol{z}^{t+1}=\boldsymbol{z}^t-\alpha_{\boldsymbol{z}}\Big(\nabla_{\boldsymbol{z}}\bar{\mathcal{L}}_\lambda(\{\boldsymbol{\omega}_i^{t+1}\},\{\epsilon_i^{t+1}\},\boldsymbol{z}^t,\{\lambda_i^t\},\{\phi_i^t\})+\psi\big(\sum_{i\in\mathcal{R}}\text{sign}(\boldsymbol{z}_i^t-\boldsymbol{\omega}_i^{t+1})+\sum_{i\in\mathcal{B}}\text{sign}(\boldsymbol{z}_i^t-\boldsymbol{\omega}_i^{t+1})\big)\Big).\tag{24}$$

Combining $\nabla_{\boldsymbol{z}}\mathcal{L}_\lambda\left(\left\{\boldsymbol{\omega}_i^{\hat{t}}\right\},\left\{\epsilon_i^{\hat{t}}\right\},\boldsymbol{z}^{\hat{t}},\left\{\lambda_i^{\hat{t}}\right\},\left\{\phi_i^{\hat{t}}\right\}\right)=\nabla_{\boldsymbol{z}}\bar{\mathcal{L}}_\lambda\left(\left\{\boldsymbol{\omega}_i^{\hat{t}}\right\},\left\{\epsilon_i^{\hat{t}}\right\},\boldsymbol{z}^{\hat{t}},\left\{\lambda_i^{\hat{t}}\right\},\left\{\phi_i^{\hat{t}}\right\}\right)$ with the optimal condition for Eq. (24), we have,

$$\Big\langle\boldsymbol{z}^t-\boldsymbol{z}^{t+1},\boldsymbol{z}^{t+1}-\boldsymbol{z}^t+\alpha_{\boldsymbol{z}}\Big(\nabla_{\boldsymbol{z}}\mathcal{L}_\lambda\left(\left\{\boldsymbol{\omega}_i^{\hat{t}}\right\},\left\{\epsilon_i^{\hat{t}}\right\},\boldsymbol{z}^{\hat{t}},\left\{\lambda_i^{\hat{t}}\right\},\left\{\phi_i^{\hat{t}}\right\}\right)+\psi\big(\sum_{i\in\mathcal{R}}\text{sign}(\boldsymbol{z}_i^t-\boldsymbol{\omega}_i^{t+1})+\sum_{i\in\mathcal{B}}\text{sign}(\boldsymbol{z}_i^t-\boldsymbol{\omega}_i^{t+1})\big)\Big)\Big\rangle\geq 0.\tag{25}$$

According to the Cauchy-Schwarz inequality, Assumption 1 and 2, we have,

$$\begin{aligned}&\Big\langle\boldsymbol{z}^{t+1}-\boldsymbol{z}^t,\nabla_{\boldsymbol{z}}\mathcal{L}_\lambda\left(\left\{\boldsymbol{\omega}_i^{\hat{t}}\right\},\left\{\epsilon_i^{\hat{t}}\right\},\boldsymbol{z}^{\hat{t}},\left\{\lambda_i^{\hat{t}}\right\},\left\{\phi_i^{\hat{t}}\right\}\right)+\psi\big(\sum_{i\in\mathcal{R}}\text{sign}(\boldsymbol{z}_i^t-\boldsymbol{\omega}_i^{t+1})+\sum_{i\in\mathcal{B}}\text{sign}(\boldsymbol{z}_i^t-\boldsymbol{\omega}_i^{t+1})\big)\Big\rangle\\&=\Big\langle\boldsymbol{z}^{t+1}-\boldsymbol{z}^t,\nabla_{\boldsymbol{z}}\mathcal{L}_\lambda\left(\left\{\boldsymbol{\omega}_i^{\hat{t}}\right\},\left\{\epsilon_i^{\hat{t}}\right\},\boldsymbol{z}^{\hat{t}},\left\{\lambda_i^{\hat{t}}\right\},\left\{\phi_i^{\hat{t}}\right\}\right)\Big\rangle+\psi\Big\langle\boldsymbol{z}^{t+1}-\boldsymbol{z}^t,\sum_{i\in\mathcal{R}}\text{sign}(\boldsymbol{z}_i^t-\boldsymbol{\omega}_i^{t+1})+\sum_{i\in\mathcal{B}}\text{sign}(\boldsymbol{z}_i^t-\boldsymbol{\omega}_i^{t+1})\Big\rangle\\&\leq-\frac{1}{\alpha_{\boldsymbol{z}}}\left\|\boldsymbol{z}^{t+1}-\boldsymbol{z}^t\right\|^2.\end{aligned}\tag{26}$$

Following Eq. (26), we have,

$$\begin{aligned}&\left\langle\boldsymbol{z}^{t+1}-\boldsymbol{z}^t,\nabla_{\boldsymbol{z}}\mathcal{L}_\lambda\left(\left\{\boldsymbol{\omega}_i^{t+1}\right\},\left\{\epsilon_i^{t+1}\right\},\boldsymbol{z}^{t},\left\{\lambda_i^t\right\},\left\{\phi_i^t\right\}\right)\right\rangle\\&\leq-\frac{1}{\alpha_{\boldsymbol{z}}}\left\|\boldsymbol{z}^{t+1}-\boldsymbol{z}^t\right\|^2+\psi\Big\langle\boldsymbol{z}^t-\boldsymbol{z}^{t+1},\sum_{i\in\mathcal{R}}\text{sign}(\boldsymbol{z}_i^t-\boldsymbol{\omega}_i^{t+1})+\sum_{i\in\mathcal{B}}\text{sign}(\boldsymbol{z}_i^t-\boldsymbol{\omega}_i^{t+1})\Big\rangle\\&\leq(\frac{\psi}{2}-\frac{1}{\alpha_{\boldsymbol{z}}})\left\|\boldsymbol{z}^{t+1}-\boldsymbol{z}^t\right\|^2+2\psi.\end{aligned}\tag{27}$$

According to the Cauchy-Schwarz inequality, Assumption 1 and 2, we have,

$$\begin{aligned}&\mathcal{L}_\lambda\left(\left\{\boldsymbol{\omega}_i^{t+1}\right\},\left\{\epsilon_i^{t+1}\right\},\boldsymbol{z}^{t+1},\left\{\lambda_i^t\right\},\left\{\phi_i^t\right\}\right)-\mathcal{L}_\lambda\left(\left\{\boldsymbol{\omega}_i^{t+1}\right\},\left\{\epsilon_i^{t+1}\right\},\boldsymbol{z}^{t},\left\{\lambda_i^t\right\},\left\{\phi_i^t\right\}\right)\\&\leq\left\langle\nabla_{\boldsymbol{z}}\mathcal{L}_\lambda\left(\left\{\boldsymbol{\omega}_i^{t+1}\right\},\left\{\epsilon_i^{t+1}\right\},\boldsymbol{z}^{t},\left\{\lambda_i^t\right\},\left\{\phi_i^t\right\}\right),\boldsymbol{z}^{t+1}-\boldsymbol{z}^t\right\rangle+\frac{L}{2}\left\|\boldsymbol{z}^{t+1}-\boldsymbol{z}^t\right\|^2\\&\leq(\frac{L+\psi}{2}-\frac{1}{\alpha_{\boldsymbol{z}}})\left\|\boldsymbol{z}^{t+1}-\boldsymbol{z}^t\right\|^2+2\psi.\end{aligned}\tag{28}$$

Finally, combining Eq. (16), Eq. (22), and Eq. (28), we can conclude the proof of Lemma 1.

Appendix D
Lemma 2

Lemma 2: Let Assumption 1 and 2 hold, $\forall t\geq 0$, we have:

$$\begin{aligned}&\mathcal{L}_\lambda\big(\{\boldsymbol{\omega}_i^{t+1}\},\{\epsilon_i^{t+1}\},\boldsymbol{z}^{t+1},\{\lambda_i^{t+1}\},\{\phi_i^{t+1}\}\big)-\mathcal{L}_\lambda\left(\{\boldsymbol{\omega}_i^{t}\},\{\epsilon_i^{t}\},\boldsymbol{z}^{t},\{\lambda_i^{t}\},\{\phi_i^{t}\}\right)\\&\leq(\frac{L^2+L+1}{2}-\frac{1}{\alpha_{\boldsymbol{\omega}}}+\frac{L^2}{2}(\frac{M}{b_1}+\frac{S}{b_3}))\sum_{i=1}^{M}\left\|\boldsymbol{\omega}_i^{t+1}-\boldsymbol{\omega}_i^t\right\|^2\\&+(\frac{L+1}{2}-\frac{1}{\alpha_\epsilon}+\frac{L^2}{2}(\frac{M}{b_1}+\frac{S}{b_3}))\sum_{i=1}^{M}\left\|\epsilon_i^{t+1}-\epsilon_i^t\right\|^2\\&+(\frac{L+6\varkappa k_0ML^2+\psi}{2}-\frac{1}{\alpha_{\boldsymbol{z}}}+\frac{L^2}{2}(\frac{M}{b_1}+\frac{S}{b_3}))\left\|\boldsymbol{z}^{t+1}-\boldsymbol{z}^t\right\|^2+4\psi^2+2\psi+\frac{1}{2\alpha_\phi}\sum_{i=1}^{M}\left\|\phi_i^t-\phi_i^{t-1}\right\|^2\\&+\frac{1}{2}\big(b_1+6\varkappa k_0ML^2-a_1^{t-1}+a_1^t+\frac{1}{\alpha_\lambda}\big)\sum_{i=1}^{M}\left\|\lambda_i^{t+1}-\lambda_i^t\right\|^2+\frac{1}{2}\big(b_3-a_2^{t-1}+a_2^t+\frac{1}{\alpha_\phi}\big)\sum_{i=1}^{M}\left\|\phi_i^{t+1}-\phi_i^t\right\|^2\\&+\frac{a_1^{t-1}}{2}\sum_{i=1}^{M}\big(\left\|\lambda_i^{t+1}\right\|^2-\left\|\lambda_i^t\right\|^2\big)+\frac{1}{2\alpha_\lambda}\sum_{i=1}^{M}\left\|\lambda_i^t-\lambda_i^{t-1}\right\|^2+\frac{a_2^{t-1}}{2}\sum_{i=1}^{M}\big(\left\|\phi_i^{t+1}\right\|^2-\left\|\phi_i^t\right\|^2\big),\end{aligned}\tag{29}$$

where $b_1>0$ and $b_3>0$ are constants.

Proof of Lemma 2:

$$\phi_i^{t+1}=\phi_i^t+\alpha_\phi\nabla_{\phi_i}\bar{\mathcal{L}}_\lambda(\boldsymbol{\omega}_i^{t+1}\},\{\epsilon_i^{t+1}\},\boldsymbol{z}^{t+1},\{\lambda_i^{t+1}\},\{\phi_i^t\}). \tag{30}$$

Following Eq. (30), $\forall\lambda\in\Lambda$, we can get the following equation in $(t+1)$-th iteration,

$$\left\langle\lambda_i^{t+1}-\lambda_i^t-\alpha_\lambda\nabla_{\lambda_i}\bar{\mathcal{L}}_\lambda\left(\left\{\boldsymbol{\omega}_i^{t+1}\right\},\left\{\epsilon_i^{t+1}\right\},\boldsymbol{z}^{t+1},\left\{\lambda_i^t\right\},\left\{\phi_i^t\right\}\right),\lambda-\lambda_i^{t+1}\right\rangle=0. \tag{31}$$

We assume $\lambda=\lambda_i^t$ and can achieve,

$$\left\langle\nabla_{\lambda_i}\bar{\mathcal{L}}_\lambda\left(\left\{\boldsymbol{\omega}_i^{t+1}\right\},\left\{\epsilon_i^{t+1}\right\},\boldsymbol{z}^{t+1},\left\{\lambda_i^t\right\},\left\{\phi_i^t\right\}\right)-\frac{1}{\alpha_\lambda}\left(\lambda_i^{t+1}-\lambda_i^t\right),\lambda_i^t-\lambda_i^{t+1}\right\rangle=0. \tag{32}$$

Likewise, in t-th iteration, we can get,

$$\left\langle\nabla_{\lambda_i}\bar{\mathcal{L}}_\lambda\left(\left\{\boldsymbol{\omega}_i^t\right\},\left\{\epsilon_i^t\right\},\boldsymbol{z}^t,\left\{\lambda_i^{t-1}\right\},\left\{\phi_i^{t-1}\right\}\right)-\frac{1}{\alpha_\lambda}\left(\lambda_i^t-\lambda_i^{t-1}\right),\lambda_i^{t+1}-\lambda_i^t\right\rangle=0. \tag{33}$$

As $\bar{\mathcal{L}}_\lambda\left(\left\{\boldsymbol{\omega}_i\right\},\left\{\epsilon_i\right\},\boldsymbol{z},\left\{\lambda_i\right\},\left\{\phi_i\right\}\right)$ is concave in terms of λ_i and combining Eq. (32) with (33), we have,

$$\begin{aligned}
&\bar{\mathcal{L}}_\lambda\left(\left\{\boldsymbol{\omega}_i^{t+1}\right\},\left\{\epsilon_i^{t+1}\right\},\boldsymbol{z}^{t+1},\left\{\lambda_i^{t+1}\right\},\left\{\phi_i^t\right\}\right)-\bar{\mathcal{L}}_\lambda\left(\left\{\boldsymbol{\omega}_i^{t+1}\right\},\left\{\epsilon_i^{t+1}\right\},\boldsymbol{z}^{t+1},\left\{\lambda_i^t\right\},\left\{\phi_i^t\right\}\right)\\
&\le\sum_{i=1}^M\left\langle\nabla_{\lambda_i}\bar{\mathcal{L}}_\lambda\left(\left\{\boldsymbol{\omega}_i^{t+1}\right\},\left\{\epsilon_i^{t+1}\right\},\boldsymbol{z}^{t+1},\left\{\lambda_i^t\right\},\left\{\phi_i^t\right\}\right),\lambda_i^{t+1}-\lambda_i^t\right\rangle\\
&\le\sum_{i=1}^M\Big(\left\langle\nabla_{\lambda_i}\bar{\mathcal{L}}_\lambda\left(\left\{\boldsymbol{\omega}_i^{t+1}\right\},\left\{\epsilon_i^{t+1}\right\},\boldsymbol{z}^{t+1},\left\{\lambda_i^t\right\},\left\{\phi_i^t\right\}\right)-\nabla_{\lambda_i}\bar{\mathcal{L}}_\lambda\left(\left\{\boldsymbol{\omega}_i^t\right\},\left\{\epsilon_i^t\right\},\boldsymbol{z}^t,\left\{\lambda_i^{t-1}\right\},\left\{\phi_i^{t-1}\right\}\right),\lambda_i^{t+1}-\lambda_i^t\right\rangle\\
&\quad+\frac{1}{\alpha_\lambda}\left\langle\lambda_i^t-\lambda_i^{t-1},\lambda_i^{t+1}-\lambda_i^t\right\rangle\Big).
\end{aligned} \tag{34}$$

Assume $\boldsymbol{d}_{1,i}^{t+1}=\lambda_i^{t+1}-\lambda_i^t-\left(\lambda_i^t-\lambda_i^{t-1}\right)$, then, we have,

$$\begin{aligned}
&\sum_{i=1}^M\left\langle\nabla_{\lambda_i}\bar{\mathcal{L}}_\lambda\left(\left\{\boldsymbol{\omega}_i^{t+1}\right\},\left\{\epsilon_i^{t+1}\right\},\boldsymbol{z}^{t+1},\left\{\lambda_i^t\right\},\left\{\phi_i^t\right\}\right)-\nabla_{\lambda_i}\bar{\mathcal{L}}_\lambda\left(\left\{\boldsymbol{\omega}_i^t\right\},\left\{\epsilon_i^t\right\},\boldsymbol{z}^t,\left\{\lambda_i^{t-1}\right\},\left\{\phi_i^{t-1}\right\}\right),\lambda_i^{t+1}-\lambda_i^t\right\rangle\\
&=\sum_{i=1}^M\left\langle\nabla_{\lambda_i}\bar{\mathcal{L}}_\lambda\left(\left\{\boldsymbol{\omega}_i^{t+1}\right\},\left\{\epsilon_i^{t+1}\right\},\boldsymbol{z}^{t+1},\left\{\lambda_i^t\right\},\left\{\phi_i^t\right\}\right)-\nabla_{\lambda_i}\bar{\mathcal{L}}_\lambda\left(\left\{\boldsymbol{\omega}_i^t\right\},\left\{\epsilon_i^t\right\},\boldsymbol{z}^t,\left\{\lambda_i^t\right\},\left\{\phi_i^t\right\}\right),\lambda_i^{t+1}-\lambda_i^t\right\rangle\text{(I)}\\
&+\sum_{i=1}^M\left\langle\nabla_{\lambda_i}\bar{\mathcal{L}}_\lambda\left(\left\{\boldsymbol{\omega}_i^t\right\},\left\{\epsilon_i^t\right\},\boldsymbol{z}^t,\left\{\lambda_i^t\right\},\left\{\phi_i^t\right\}\right)-\nabla_{\lambda_i}\bar{\mathcal{L}}_\lambda\left(\left\{\boldsymbol{\omega}_i^t\right\},\left\{\epsilon_i^t\right\},\boldsymbol{z}^t,\left\{\lambda_i^{t-1}\right\},\left\{\phi_i^{t-1}\right\}\right),\boldsymbol{d}_{1,i}^{t+1}\right\rangle\text{(II)}\\
&+\sum_{i=1}^M\left\langle\nabla_{\lambda_i}\bar{\mathcal{L}}_\lambda\left(\left\{\boldsymbol{\omega}_i^t\right\},\left\{\epsilon_i^t\right\},\boldsymbol{z}^t,\left\{\lambda_i^t\right\},\left\{\phi_i^t\right\}\right)-\nabla_{\lambda_i}\bar{\mathcal{L}}_\lambda\left(\left\{\boldsymbol{\omega}_i^t\right\},\left\{\epsilon_i^t\right\},\boldsymbol{z}^t,\left\{\lambda_i^{t-1}\right\},\left\{\phi_i^{t-1}\right\}\right),\lambda_i^t-\lambda_i^{t-1}\right\rangle\text{(III)}.
\end{aligned} \tag{35}$$

To begin with, we pay attention to the function (I) in Eq. (35), which can be written as,

$$\begin{aligned}
&\left\langle\nabla_{\lambda_i}\bar{\mathcal{L}}_\lambda\left(\left\{\boldsymbol{\omega}_i^{t+1}\right\},\left\{\epsilon_i^{t+1}\right\},\boldsymbol{z}^{t+1},\left\{\lambda_i^t\right\},\left\{\phi_i^t\right\}\right)-\nabla_{\lambda_i}\bar{\mathcal{L}}_\lambda\left(\left\{\boldsymbol{\omega}_i^t\right\},\left\{\epsilon_i^t\right\},\boldsymbol{z}^t,\left\{\lambda_i^t\right\},\left\{\phi_i^t\right\}\right),\lambda_i^{t+1}-\lambda_i^t\right\rangle\\
&=\left\langle\nabla_{\lambda_i}\mathcal{L}_\lambda\left(\left\{\boldsymbol{\omega}_i^{t+1}\right\},\left\{\epsilon_i^{t+1}\right\},\boldsymbol{z}^{t+1},\left\{\lambda_i^t\right\},\left\{\phi_i^t\right\}\right)-\nabla_{\lambda_i}\mathcal{L}_\lambda\left(\left\{\boldsymbol{\omega}_i^t\right\},\left\{\epsilon_i^t\right\},\boldsymbol{z}^t,\left\{\lambda_i^t\right\},\left\{\phi_i^t\right\}\right),\lambda_i^{t+1}-\lambda_i^t\right\rangle\\
&\quad+\left(a_1^{t-1}-a_1^t\right)\left\langle\lambda_i^t,\lambda_i^{t+1}-\lambda_i^t\right\rangle\\
&=\left\langle\nabla_{\lambda_i}\mathcal{L}_\lambda\left(\left\{\boldsymbol{\omega}_i^{t+1}\right\},\left\{\epsilon_i^{t+1}\right\},\boldsymbol{z}^{t+1},\left\{\lambda_i^t\right\},\left\{\phi_i^t\right\}\right)-\nabla_{\lambda_i}\mathcal{L}_\lambda\left(\left\{\boldsymbol{\omega}_i^t\right\},\left\{\epsilon_i^t\right\},\boldsymbol{z}^t,\left\{\lambda_i^t\right\},\left\{\phi_i^t\right\}\right),\lambda_i^{t+1}-\lambda_i^t\right\rangle\\
&\quad+\frac{a_1^{t-1}-a_1^t}{2}\left(\left\|\lambda_i^{t+1}\right\|^2-\left\|\lambda_i^t\right\|^2-\left\|\lambda_i^{t+1}-\lambda_i^t\right\|^2\right).
\end{aligned} \tag{36}$$

According to Cauchy-Schwarz inequality and Assumption 1, we have,

$$\begin{aligned}
&\left\langle\nabla_{\lambda_i}\mathcal{L}_\lambda\left(\left\{\boldsymbol{\omega}_i^{t+1}\right\},\left\{\epsilon_i^{t+1}\right\},\boldsymbol{z}^{t+1},\left\{\lambda_i^t\right\},\left\{\phi_i^t\right\}\right)-\nabla_{\lambda_i}\mathcal{L}_\lambda\left(\left\{\boldsymbol{\omega}_i^t\right\},\left\{\epsilon_i^t\right\},\boldsymbol{z}^t,\left\{\lambda_i^t\right\},\left\{\phi_i^t\right\}\right),\lambda_i^{t+1}-\lambda_i^t\right\rangle\\
&\le\frac{L^2}{2b_1}\sum_{i=1}^M\left(\left\|\boldsymbol{\omega}_i^{t+1}-\boldsymbol{\omega}_i^t\right\|^2+\left\|\epsilon_i^{t+1}-\epsilon_i^t\right\|^2\right)+\frac{L^2}{2b_1}\left\|\boldsymbol{z}^{t+1}-\boldsymbol{z}^t\right\|^2+\frac{b_1}{2}\left\|\lambda_i^{t+1}-\lambda_i^t\right\|^2.
\end{aligned} \tag{37}$$

Combining Eq. (36) with (37), we have,

$$\begin{aligned}
&\sum_{i=1}^{M} \left\langle \nabla_{\lambda_i} \bar{\mathcal{L}}_\lambda \left(\left\{ \boldsymbol{\omega}_i^{t+1} \right\}, \left\{ \epsilon_i^{t+1} \right\}, \boldsymbol{z}^{t+1}, \left\{ \lambda_i^t \right\}, \left\{ \phi_i^t \right\} \right) - \nabla_{\lambda_i} \bar{\mathcal{L}}_\lambda \left(\left\{ \boldsymbol{\omega}_i^t \right\}, \left\{ \epsilon_i^t \right\}, \boldsymbol{z}^t, \left\{ \lambda_i^t \right\}, \left\{ \phi_i^t \right\} \right), \lambda_i^{t+1} - \lambda_i^t \right\rangle \\
&\leq \sum_{i=1}^{M} \Big(\frac{L^2}{2b_1} \sum_{i=1}^{M} \left(\left\| \boldsymbol{\omega}_i^{t+1} - \boldsymbol{\omega}_i^t \right\|^2 + \left\| \epsilon_i^{t+1} - \epsilon_i^t \right\|^2 \right) + \frac{L^2}{2b_1} \left\| \boldsymbol{z}^{t+1} - \boldsymbol{z}^t \right\|^2 \\
&+ \frac{a_1^{t-1} - a_1^t}{2} \left(\| \lambda_i^{t+1} \|^2 - \| \lambda_i^t \|^2 \right) + \frac{b_1 - a_1^{t-1} + a_1^t}{2} \left\| \lambda_i^{t+1} - \lambda_i^t \right\|^2 \Big).
\end{aligned} \tag{38}$$

Then, we pay attention to the function (II) in Eq. (35), which can be expressed as,

$$\begin{aligned}
&\sum_{i=1}^{M} \left\langle \nabla_{\lambda_i} \bar{\mathcal{L}}_\lambda \left(\left\{ \boldsymbol{\omega}_i^t \right\}, \left\{ \epsilon_i^t \right\}, \boldsymbol{z}^t, \left\{ \lambda_i^t \right\}, \left\{ \phi_i^t \right\} \right) - \nabla_{\lambda_i} \bar{\mathcal{L}}_\lambda \left(\left\{ \boldsymbol{\omega}_i^t \right\}, \left\{ \epsilon_i^t \right\}, \boldsymbol{z}^t, \left\{ \lambda_i^{t-1} \right\}, \left\{ \phi_i^{t-1} \right\} \right), \boldsymbol{d}_{1,i}^{t+1} \right\rangle \\
&\leq \sum_{i=1}^{M} \Big(\frac{b_2}{2} \left\| \nabla_{\lambda_i} \bar{\mathcal{L}}_\lambda \left(\left\{ \boldsymbol{\omega}_i^t \right\}, \left\{ \epsilon_i^t \right\}, \boldsymbol{z}^t, \left\{ \lambda_i^t \right\}, \left\{ \phi_i^t \right\} \right) - \nabla_{\lambda_i} \bar{\mathcal{L}}_\lambda \left(\left\{ \boldsymbol{\omega}_i^t \right\}, \left\{ \epsilon_i^t \right\}, \boldsymbol{z}^t, \left\{ \lambda_i^{t-1} \right\}, \left\{ \phi_i^{t-1} \right\} \right) \right\|^2 + \frac{1}{2b_2} \left\| \boldsymbol{d}_{1,i}^{t+1} \right\|^2 \Big),
\end{aligned} \tag{39}$$

where $b_2 > 0$ is a constant.

Additionally, we focus on the function (III) in Eq. (35). we assume $L_1' = L + a_1^0$. Combining Assumption 1 and the trigonometric inequality, we have,

$$\begin{aligned}
&\left\| \nabla_{\lambda_i} \bar{\mathcal{L}}_\lambda \left(\left\{ \boldsymbol{\omega}_i^t \right\}, \left\{ \epsilon_i^t \right\}, \boldsymbol{z}^t, \left\{ \lambda_i^t \right\}, \left\{ \phi_i^t \right\} \right) - \nabla_{\lambda_i} \bar{\mathcal{L}}_\lambda \left(\left\{ \boldsymbol{\omega}_i^t \right\}, \left\{ \epsilon_i^t \right\}, \boldsymbol{z}^t, \left\{ \lambda_i^{t-1} \right\}, \left\{ \phi_i^{t-1} \right\} \right) \right\| \\
&= \left\| \nabla_{\lambda_i} \mathcal{L}_\lambda \left(\left\{ \boldsymbol{\omega}_i^t \right\}, \left\{ \epsilon_i^t \right\}, \boldsymbol{z}^t, \left\{ \lambda_i^t \right\}, \left\{ \phi_i^t \right\} \right) - \nabla_{\lambda_i} \mathcal{L}_\lambda \left(\left\{ \boldsymbol{\omega}_i^t \right\}, \left\{ \epsilon_i^t \right\}, \boldsymbol{z}^t, \left\{ \lambda_i^{t-1} \right\}, \left\{ \phi_i^{t-1} \right\} \right) - a_1^{t-1} \left(\lambda_i^t - \lambda_i^{t-1} \right) \right\| \\
&\leq (L + a_1^{t-1}) \left\| \lambda_i^t - \lambda_i^{t-1} \right\| = L_1' \left\| \lambda_i^t - \lambda_i^{t-1} \right\|.
\end{aligned} \tag{40}$$

Following from Eq. (40) and the strong concavity of $\bar{\mathcal{L}}_\lambda \left(\left\{ \boldsymbol{\omega}_i \right\}, \left\{ \epsilon_i \right\}, \boldsymbol{z}, \left\{ \lambda_i \right\}, \left\{ \phi_i \right\} \right)$ w.r.t λ_i, we have,

$$\begin{aligned}
&\sum_{i=1}^{M} \left\langle \nabla_{\lambda_i} \bar{\mathcal{L}}_\lambda \left(\left\{ \boldsymbol{\omega}_i^t \right\}, \left\{ \epsilon_i^t \right\}, \boldsymbol{z}^t, \left\{ \lambda_i^t \right\}, \left\{ \phi_i^t \right\} \right) - \nabla_{\lambda_i} \bar{\mathcal{L}}_\lambda \left(\left\{ \boldsymbol{\omega}_i^t \right\}, \left\{ \epsilon_i^t \right\}, \boldsymbol{z}^t, \left\{ \lambda_i^{t-1} \right\}, \left\{ \phi_i^{t-1} \right\} \right), \lambda_i^t - \lambda_i^{t-1} \right\rangle \\
&\leq \sum_{i=1}^{M} \Big(- \frac{1}{L_1' + a_1^{t-1}} \left\| \nabla_{\lambda_i} \bar{\mathcal{L}}_\lambda \left(\left\{ \boldsymbol{\omega}_i^t \right\}, \left\{ \epsilon_i^t \right\}, \boldsymbol{z}^t, \left\{ \lambda_i^t \right\}, \left\{ \phi_i^t \right\} \right) - \nabla_{\lambda_i} \bar{\mathcal{L}}_\lambda \left(\left\{ \boldsymbol{\omega}_i^t \right\}, \left\{ \epsilon_i^t \right\}, \boldsymbol{z}^t, \left\{ \lambda_i^{t-1} \right\}, \left\{ \phi_i^{t-1} \right\} \right) \right\|^2 \\
&\quad - \frac{a_1^{t-1} L_1'}{L_1' + a_1^{t-1}} \left\| \lambda_i^t - \lambda_i^{t-1} \right\|^2 \Big).
\end{aligned} \tag{41}$$

Furthermore, we can achieve the following inequality,

$$\frac{1}{\alpha_\lambda} \left\langle \lambda_i^t - \lambda_i^{t-1}, \lambda_i^{t+1} - \lambda_i^t \right\rangle \leq \frac{1}{2\alpha_\lambda} \left\| \lambda_i^{t+1} - \lambda_i^t \right\|^2 - \frac{1}{2\alpha_\lambda} \left\| \boldsymbol{d}_{1,i}^{t+1} \right\|^2 + \frac{1}{2\alpha_\lambda} \left\| \lambda_i^t - \lambda_i^{t-1} \right\|^2. \tag{42}$$

With the combination of Eq. (35), (36), (38), (39), (41), (42), $\frac{\alpha_\lambda}{2} \leq \frac{1}{L_1' + a_1^0}$, and setting $b_2 = \alpha_\lambda$, we have,

$$\begin{aligned}
&\mathcal{L}_\lambda \left(\left\{ \boldsymbol{\omega}_i^{t+1} \right\}, \left\{ \epsilon_i^{t+1} \right\}, \boldsymbol{z}^{t+1}, \left\{ \lambda_i^{t+1} \right\}, \left\{ \phi_i^t \right\} \right) - \mathcal{L}_\lambda \left(\left\{ \boldsymbol{\omega}_i^{t+1} \right\}, \left\{ \epsilon_i^{t+1} \right\}, \boldsymbol{z}^{t+1}, \left\{ \lambda_i^t \right\}, \left\{ \phi_i^t \right\} \right) \\
&\leq \sum_{i=1}^{M} \Big(\left\langle \nabla_{\lambda_i} \bar{\mathcal{L}}_\lambda \left(\left\{ \boldsymbol{\omega}_i^{t+1} \right\}, \left\{ \epsilon_i^{t+1} \right\}, \boldsymbol{z}^{t+1}, \left\{ \lambda_i^t \right\}, \left\{ \phi_i^t \right\} \right) - \nabla_{\lambda_i} \bar{\mathcal{L}}_\lambda \left(\left\{ \boldsymbol{\omega}_i^t \right\}, \left\{ \epsilon_i^t \right\}, \boldsymbol{z}^t, \left\{ \lambda_i^{t-1} \right\}, \left\{ \phi_i^{t-1} \right\} \right), \lambda_i^{t+1} - \lambda_i^t \right\rangle \\
&\quad + \frac{1}{\alpha_\lambda} \left\langle \lambda_i^t - \lambda_i^{t-1}, \lambda_i^{t+1} - \lambda_i^t \right\rangle + \frac{a_1^t}{2} \left(\left\| \lambda_i^{t+1} \right\|^2 - \left\| \lambda_i^t \right\|^2 \right) \Big) \\
&\leq \frac{ML^2}{2b_1} \Big(\sum_{i=1}^{M} \left(\left\| \boldsymbol{\omega}_i^{t+1} - \boldsymbol{\omega}_i^t \right\|^2 + \left\| \epsilon_i^{t+1} - \epsilon_i^t \right\|^2 \right) + \| \boldsymbol{z}^{t+1} - \boldsymbol{z}^t \|^2 \Big) \\
&\quad + \left(\frac{b_1}{2} - \frac{a_1^{t-1} - a_1^t}{2} + \frac{1}{2\alpha_\lambda} \right) \sum_{i=1}^{M} \left\| \lambda_i^{t+1} - \lambda_i^t \right\|^2 + \frac{a_1^{t-1}}{2} \sum_{i=1}^{M} \left(\left\| \lambda_i^{t+1} \right\|^2 - \left\| \lambda_i^t \right\|^2 \right) + \frac{1}{2\alpha_\lambda} \sum_{i=1}^{M} \left\| \lambda_i^t - \lambda_i^{t-1} \right\|^2.
\end{aligned} \tag{43}$$

According to Eq. (30), in $(t+1)$-th iteration, we have,

$$\left\langle \phi_i^{t+1} - \phi_i^t - \alpha_\phi \nabla_{\phi_i} \bar{\mathcal{L}}_\lambda \left(\left\{ \boldsymbol{\omega}_i^{t+1} \right\}, \left\{ \epsilon_i^{t+1} \right\}, \boldsymbol{z}^{t+1}, \left\{ \lambda_i^{t+1} \right\}, \left\{ \phi_i^t \right\} \right), \phi - \phi_i^{t+1} \right\rangle = 0. \tag{44}$$

Let $\phi_i = \phi_i^t$, and we have,

$$\left\langle \nabla_{\phi_i} \bar{\mathcal{L}}_\lambda \left(\left\{ \boldsymbol{\omega}_i^{t+1} \right\}, \left\{ \epsilon_i^{t+1} \right\}, \boldsymbol{z}^{t+1}, \left\{ \lambda_i^{t+1} \right\}, \left\{ \phi_i^t \right\} \right) - \frac{1}{\alpha_\phi} \left(\phi_i^{t+1} - \phi_i^t \right), \phi_i^t - \phi_i^{t+1} \right\rangle = 0. \tag{45}$$

Likewise, in t-th iteration, we can get,

$$\left\langle \nabla_{\phi_i} \bar{\mathcal{L}}_\lambda \left(\left\{ \boldsymbol{\omega}_i^t \right\}, \left\{ \epsilon_i^t \right\}, \boldsymbol{z}^t, \left\{ \lambda_i^t \right\}, \left\{ \phi_i^{t-1} \right\} \right) - \frac{1}{\alpha_\phi} \left(\phi_i^t - \phi_i^{t-1} \right), \phi_i^{t+1} - \phi_i^t \right\rangle = 0. \tag{46}$$

As $\bar{\mathcal{L}}_\lambda \left(\left\{ \boldsymbol{\omega}_i \right\}, \left\{ \epsilon_i \right\}, \boldsymbol{z}, \left\{ \lambda_i \right\}, \left\{ \phi_i \right\} \right)$ is concave in terms of ϕ_i and according to Eq. (46), we have,

$$\begin{aligned}
&\bar{\mathcal{L}}_\lambda \left(\left\{ \boldsymbol{\omega}_i^{t+1} \right\}, \left\{ \epsilon_i^{t+1} \right\}, \boldsymbol{z}^{t+1}, \left\{ \lambda_i^{t+1} \right\}, \left\{ \phi_i^{t+1} \right\} \right) - \bar{\mathcal{L}}_\lambda \left(\left\{ \boldsymbol{\omega}_i^{t+1} \right\}, \left\{ \epsilon_i^{t+1} \right\}, \boldsymbol{z}^{t+1}, \left\{ \lambda_i^{t+1} \right\}, \left\{ \phi_i^{t} \right\} \right) \\
&\leq \sum_{i=1}^{M} \left\langle \nabla_{\phi_i} \bar{\mathcal{L}}_\lambda \left(\left\{ \boldsymbol{\omega}_i^{t+1} \right\}, \left\{ \epsilon_i^{t+1} \right\}, \boldsymbol{z}^{t+1}, \left\{ \lambda_i^{t+1} \right\}, \left\{ \phi_i^{t} \right\} \right), \phi_i^{t+1} - \phi_i^t \right\rangle \\
&\leq \sum_{i=1}^{M} \Big(\left\langle \nabla_{\phi_i} \bar{\mathcal{L}}_\lambda \left(\left\{ \boldsymbol{\omega}_i^{t+1} \right\}, \left\{ \epsilon_i^{t+1} \right\}, \boldsymbol{z}^{t+1}, \left\{ \lambda_i^{t+1} \right\}, \left\{ \phi_i^{t} \right\} \right) - \nabla_{\phi_i} \bar{\mathcal{L}}_\lambda \left(\left\{ \boldsymbol{\omega}_i^t \right\}, \left\{ \epsilon_i^t \right\}, \boldsymbol{z}^t, \left\{ \lambda_i^t \right\}, \left\{ \phi_i^{t-1} \right\} \right), \phi_i^{t+1} - \phi_i^t \right\rangle \\
&\quad + \frac{1}{\alpha_\phi} \left\langle \phi_i^t - \phi_i^{t-1}, \phi_i^{t+1} - \phi_i^t \right\rangle \Big).
\end{aligned} \tag{47}$$

Assume $\boldsymbol{d}_{2,i}^{t+1} = \phi_i^{t+1} - \phi_i^t - \left(\phi_i^t - \phi_i^{t-1} \right)$, then we have,

$$\begin{aligned}
&\sum_{i=1}^{M} \left\langle \nabla_{\phi_i} \bar{\mathcal{L}}_\lambda \left(\left\{ \boldsymbol{\omega}_i^{t+1} \right\}, \left\{ \epsilon_i^{t+1} \right\}, \boldsymbol{z}^{t+1}, \left\{ \lambda_i^{t+1} \right\}, \left\{ \phi_i^{t} \right\} \right) - \nabla_{\phi_i} \bar{\mathcal{L}}_\lambda \left(\left\{ \boldsymbol{\omega}_i^t \right\}, \left\{ \epsilon_i^t \right\}, \boldsymbol{z}^t, \left\{ \lambda_i^t \right\}, \left\{ \phi_i^{t-1} \right\} \right), \phi_i^{t+1} - \phi_i^t \right\rangle \\
&= \sum_{i=1}^{M} \left\langle \nabla_{\phi_i} \bar{\mathcal{L}}_\lambda \left(\left\{ \boldsymbol{\omega}_i^{t+1} \right\}, \left\{ \epsilon_i^{t+1} \right\}, \boldsymbol{z}^{t+1}, \left\{ \lambda_i^{t+1} \right\}, \left\{ \phi_i^{t} \right\} \right) - \nabla_{\phi_i} \bar{\mathcal{L}}_\lambda \left(\left\{ \boldsymbol{\omega}_i^t \right\}, \left\{ \epsilon_i^t \right\}, \boldsymbol{z}^t, \left\{ \lambda_i^t \right\}, \left\{ \phi_i^{t} \right\} \right), \phi_i^{t+1} - \phi_i^t \right\rangle \text{(I)} \\
&\quad + \sum_{i=1}^{M} \left\langle \nabla_{\phi_i} \bar{\mathcal{L}}_\lambda \left(\left\{ \boldsymbol{\omega}_i^t \right\}, \left\{ \epsilon_i^t \right\}, \boldsymbol{z}^t, \left\{ \lambda_i^t \right\}, \left\{ \phi_i^{t} \right\} \right) - \nabla_{\phi_i} \bar{\mathcal{L}}_\lambda \left(\left\{ \boldsymbol{\omega}_i^t \right\}, \left\{ \epsilon_i^t \right\}, \boldsymbol{z}^t, \left\{ \lambda_i^t \right\}, \left\{ \phi_i^{t-1} \right\} \right), \boldsymbol{d}_{2,i}^{t+1} \right\rangle \text{(II)} \\
&\quad + \sum_{i=1}^{M} \left\langle \nabla_{\phi_i} \bar{\mathcal{L}}_\lambda \left(\left\{ \boldsymbol{\omega}_i^t \right\}, \left\{ \epsilon_i^t \right\}, \boldsymbol{z}^t, \left\{ \lambda_i^t \right\}, \left\{ \phi_i^{t} \right\} \right) - \nabla_{\phi_i} \bar{\mathcal{L}}_\lambda \left(\left\{ \boldsymbol{\omega}_i^t \right\}, \left\{ \epsilon_i^t \right\}, \boldsymbol{z}^t, \left\{ \lambda_i^t \right\}, \left\{ \phi_i^{t-1} \right\} \right), \phi_i^t - \phi_i^{t-1} \right\rangle \text{(III)}.
\end{aligned} \tag{48}$$

Firstly, we pay attention to the function (I) in Eq. (48), which can be written as,

$$\begin{aligned}
&\left\langle \nabla_{\phi_i} \bar{\mathcal{L}}_\lambda \left(\left\{ \boldsymbol{\omega}_i^{t+1} \right\}, \left\{ \epsilon_i^{t+1} \right\}, \boldsymbol{z}^{t+1}, \left\{ \lambda_i^{t+1} \right\}, \left\{ \phi_i^{t} \right\} \right) - \nabla_{\phi_i} \bar{\mathcal{L}}_\lambda \left(\left\{ \boldsymbol{\omega}_i^t \right\}, \left\{ \epsilon_i^t \right\}, \boldsymbol{z}^t, \left\{ \lambda_i^t \right\}, \left\{ \phi_i^{t} \right\} \right), \phi_i^{t+1} - \phi_i^t \right\rangle \\
&= \left\langle \nabla_{\phi_i} \mathcal{L}_\lambda \left(\left\{ \boldsymbol{\omega}_i^{t+1} \right\}, \left\{ \epsilon_i^{t+1} \right\}, \boldsymbol{z}^{t+1}, \left\{ \lambda_i^{t+1} \right\}, \left\{ \phi_i^{t} \right\} \right) - \nabla_{\phi_i} \mathcal{L}_\lambda \left(\left\{ \boldsymbol{\omega}_i^t \right\}, \left\{ \epsilon_i^t \right\}, \boldsymbol{z}^t, \left\{ \lambda_i^t \right\}, \left\{ \phi_i^{t} \right\} \right), \phi_i^{t+1} - \phi_i^t \right\rangle \\
&\quad + \frac{a_2^{t-1} - a_2^t}{2} \left(\left\| \phi_i^{t+1} \right\|^2 - \left\| \phi_i^t \right\|^2 \right) - \frac{a_2^{t-1} - a_2^t}{2} \left\| \phi_i^{t+1} - \phi_i^t \right\|^2.
\end{aligned} \tag{49}$$

According to the Cauchy-Schwarz inequality and Assumption 1, we have,

$$\begin{aligned}
&\left\langle \nabla_{\phi_i} \mathcal{L}_\lambda \left(\left\{ \boldsymbol{\omega}_i^{t+1} \right\}, \left\{ \epsilon_i^{t+1} \right\}, \boldsymbol{z}^{t+1}, \left\{ \lambda_i^{t+1} \right\}, \left\{ \phi_i^{t} \right\} \right) - \nabla_{\phi_i} \mathcal{L}_\lambda \left(\left\{ \boldsymbol{\omega}_i^t \right\}, \left\{ \epsilon_i^t \right\}, \boldsymbol{z}^t, \left\{ \lambda_i^t \right\}, \left\{ \phi_i^{t} \right\} \right), \phi_i^{t+1} - \phi_i^t \right\rangle \\
&= \left\langle \nabla_{\phi_i} \mathcal{L}_\lambda \left(\left\{ \boldsymbol{\omega}_i^{t+1} \right\}, \left\{ \epsilon_i^{t+1} \right\}, \boldsymbol{z}^{t+1}, \left\{ \lambda_i^{t} \right\}, \left\{ \phi_i^{t} \right\} \right) - \nabla_{\phi_i} \mathcal{L}_\lambda \left(\left\{ \boldsymbol{\omega}_i^t \right\}, \left\{ \epsilon_i^t \right\}, \boldsymbol{z}^t, \left\{ \lambda_i^t \right\}, \left\{ \phi_i^{t} \right\} \right), \phi_i^{t+1} - \phi_i^t \right\rangle \\
&\leq \frac{L^2}{2b_3} \Big(\sum_{i=1}^{M} \left(\left\| \boldsymbol{\omega}_i^{t+1} - \boldsymbol{\omega}_i^t \right\|^2 + \left\| \epsilon_i^{t+1} - \epsilon_i^t \right\|^2 \right) + \left\| \boldsymbol{z}^{t+1} - \boldsymbol{z}^t \right\|^2 \Big) + \frac{b_3}{2} \left\| \phi_i^{t+1} - \phi_i^t \right\|^2.
\end{aligned} \tag{50}$$

Combining Eq. (49) with (50), we can obtain the upper bound of the function (I) in Eq. (48),

$$\begin{aligned}
&\sum_{i=1}^{M} \left\langle \nabla_{\phi_i} \bar{\mathcal{L}}_\lambda \left(\left\{ \boldsymbol{\omega}_i^{t+1} \right\}, \left\{ \epsilon_i^{t+1} \right\}, \boldsymbol{z}^{t+1}, \left\{ \lambda_i^{t+1} \right\}, \left\{ \phi_i^{t} \right\} \right) - \nabla_{\phi_i} \bar{\mathcal{L}}_\lambda \left(\left\{ \boldsymbol{\omega}_i^t \right\}, \left\{ \epsilon_i^t \right\}, \boldsymbol{z}^t, \left\{ \lambda_i^t \right\}, \left\{ \phi_i^{t} \right\} \right), \phi_i^{t+1} - \phi_i^t \right\rangle \\
&\leq \sum_{i=1}^{S} \Big(\frac{L^2}{2b_3} \big(\sum_{i=1}^{M} (\left\| \boldsymbol{\omega}_i^{t+1} - \boldsymbol{\omega}_i^t \right\|^2 + \left\| \epsilon_i^{t+1} - \epsilon_i^t \right\|^2) + \left\| \boldsymbol{z}^{t+1} - \boldsymbol{z}^t \right\|^2 \big) \\
&\quad + \frac{b_3}{2} \left\| \phi_i^{t+1} - \phi_i^t \right\|^2 + \frac{a_2^{t-1} - a_2^t}{2} \big(\left\| \phi_i^{t+1} \right\|^2 - \left\| \phi_i^t \right\|^2 \big) - \frac{a_2^{t-1} - a_2^t}{2} \left\| \phi_i^{t+1} - \phi_i^t \right\|^2 \Big).
\end{aligned} \tag{51}$$

Then, we focus on the function (II) in Eq. (48). According to the Cauchy-Schwarz inequality, it can be written as,

$$\begin{aligned}
&\sum_{i=1}^{M} \left\langle \nabla_{\phi_i} \bar{\mathcal{L}}_\lambda \left(\left\{ \boldsymbol{\omega}_i^t \right\}, \left\{ \epsilon_i^t \right\}, \boldsymbol{z}^t, \left\{ \lambda_i^t \right\}, \left\{ \phi_i^{t} \right\} \right) - \nabla_{\phi_i} \bar{\mathcal{L}}_\lambda \left(\left\{ \boldsymbol{\omega}_i^t \right\}, \left\{ \epsilon_i^t \right\}, \boldsymbol{z}^t, \left\{ \lambda_i^t \right\}, \left\{ \phi_i^{t-1} \right\} \right), \boldsymbol{d}_{2,i}^{t+1} \right\rangle \\
&\leq \sum_{i=1}^{M} \Big(\frac{b_4}{2} \left\| \nabla_{\phi_i} \bar{\mathcal{L}}_\lambda \left(\left\{ \boldsymbol{\omega}_i^t \right\}, \left\{ \epsilon_i^t \right\}, \boldsymbol{z}^t, \left\{ \lambda_i^t \right\}, \left\{ \phi_i^{t} \right\} \right) - \nabla_{\phi_i} \bar{\mathcal{L}}_\lambda \left(\left\{ \boldsymbol{\omega}_i^t \right\}, \left\{ \epsilon_i^t \right\}, \boldsymbol{z}^t, \left\{ \lambda_i^t \right\}, \left\{ \phi_i^{t-1} \right\} \right) \right\|^2 + \frac{1}{2b_4} \left\| \boldsymbol{d}_{2,i}^{t+1} \right\|^2 \Big),
\end{aligned} \tag{52}$$

where $b_4 > 0$ is a constant.

Additionally, we pay attention to the function (III) in Eq. (48). We assume $L_2' = L + a_2^0$. According to the trigonometric inequality and Assumption 1, we have,

$$\left\|\nabla_{\phi_i}\bar{\mathcal{L}}_\lambda\left(\left\{\boldsymbol{\omega}_i^t\right\},\left\{\epsilon_i^t\right\},\boldsymbol{z}^t,\left\{\lambda_i^t\right\},\left\{\phi_i^t\right\}\right)-\nabla_{\phi_i}\bar{\mathcal{L}}_\lambda\left(\left\{\boldsymbol{\omega}_i^t\right\},\left\{\epsilon_i^t\right\},\boldsymbol{z}^t,\left\{\lambda_i^t\right\},\left\{\phi_i^{t-1}\right\}\right)\right\| \le L_2'\left\|\phi_i^t-\phi_i^{t-1}\right\|. \tag{53}$$

According to Eq. (53) and the strong concavity of $\bar{\mathcal{L}}_\lambda\left(\{\boldsymbol{\omega}_i\},\{\epsilon_i\},\boldsymbol{z},\{\lambda_i\},\{\phi_i\}\right)$ w.r.t ϕ_i, we can obtain the upper bound of the function (III) in Eq. (48),

$$\begin{aligned}
&\sum_{i=1}^M\left\langle\nabla_{\phi_i}\bar{\mathcal{L}}_\lambda\left(\left\{\boldsymbol{\omega}_i^t\right\},\left\{\epsilon_i^t\right\},\boldsymbol{z}^t,\left\{\lambda_i^t\right\},\left\{\phi_i^t\right\}\right)-\nabla_{\phi_i}\bar{\mathcal{L}}_\lambda\left(\left\{\boldsymbol{\omega}_i^t\right\},\left\{\epsilon_i^t\right\},\boldsymbol{z}^t,\left\{\lambda_i^t\right\},\left\{\phi_i^{t-1}\right\}\right),\phi_i^t-\phi_i^{t-1}\right\rangle\\
&\le\sum_{i=1}^M\Big(-\frac{1}{L_2'+a_2^{t-1}}\left\|\nabla_{\phi_i}\bar{\mathcal{L}}_\lambda\left(\left\{\boldsymbol{\omega}_i^t\right\},\left\{\epsilon_i^t\right\},\boldsymbol{z}^t,\left\{\lambda_i^t\right\},\left\{\phi_i^t\right\}\right)-\nabla_{\phi_i}\bar{\mathcal{L}}_\lambda\left(\left\{\boldsymbol{\omega}_i^t\right\},\left\{\epsilon_i^t\right\},\boldsymbol{z}^t,\left\{\lambda_i^t\right\},\left\{\phi_i^{t-1}\right\}\right)\right\|^2-\frac{a_2^{t-1}L_2'}{L_2'+a_2^{t-1}}\left\|\phi_i^t-\phi_i^{t-1}\right\|^2\Big).
\end{aligned} \tag{54}$$

Furthermore, we can obtain the following inequality,

$$\sum_{i=1}^M\frac{1}{\alpha_\phi}\left\langle\phi_i^t-\phi_i^{t-1},\phi_i^{t+1}-\phi_i^t\right\rangle\le\frac{1}{2\alpha_\phi}\sum_{i=1}^M\Big(\left\|\phi_i^{t+1}-\phi_i^t\right\|^2-\left\|\boldsymbol{d}_{2,i}^{t+1}\right\|^2+\left\|\phi_i^t-\phi_i^{t-1}\right\|^2\Big). \tag{55}$$

With the combination of Eq. (47), (48), (51), (52), (54), (55), $\frac{\alpha_\phi}{2}\le\frac{1}{L_2'+a_2^0}$, and setting $b_4=\alpha_\phi$, we have,

$$\begin{aligned}
&\mathcal{L}_\lambda\left(\left\{\boldsymbol{\omega}_i^{t+1}\right\},\left\{\epsilon_i^{t+1}\right\},\boldsymbol{z}^{t+1},\left\{\lambda_i^{t+1}\right\},\left\{\phi_i^{t+1}\right\}\right)-\mathcal{L}_\lambda\left(\left\{\boldsymbol{\omega}_i^{t+1}\right\},\left\{\epsilon_i^{t+1}\right\},\boldsymbol{z}^{t+1},\left\{\lambda_i^{t+1}\right\},\left\{\phi_i^t\right\}\right)\\
&\le\frac{SL^2}{2b_3}\Big(\sum_{i=1}^M\big(\left\|\boldsymbol{\omega}_i^{t+1}-\boldsymbol{\omega}_i^t\right\|^2+\left\|\epsilon_i^{t+1}-\epsilon_i^t\,|\right\|^2\big)+\left\|\boldsymbol{z}^{t+1}-\boldsymbol{z}^t\right\|^2\Big)\\
&+\big(\frac{b_3}{2}-\frac{a_2^{t-1}-a_2^t}{2}+\frac{1}{2\alpha_\phi}\big)\sum_{i=1}^M\left\|\phi_i^{t+1}-\phi_i^t\right\|^2+\frac{a_2^{t-1}}{2}\sum_{i=1}^M\big(\left\|\phi_i^{t+1}\right\|^2-\left\|\phi_i^t\right\|^2\big)+\frac{1}{2\alpha_\phi}\sum_{i=1}^M\left\|\phi_i^t-\phi_i^{t-1}\right\|^2.
\end{aligned} \tag{56}$$

According to Lemma 1, Eq. (43), and (56), we can conclude the proof of Lemma 2.

Appendix E
Lemma 3

Lemma 3: Firstly, we denote K_1^{t+1}, K_2^{t+1} and G^{t+1} as,

$$\begin{aligned}
K_1^{t+1}&=\frac{4}{{\alpha_\lambda}^2a_1^{t+1}}\sum_{i=1}^M\left\|\lambda_i^{t+1}-\lambda_i^t\right\|^2-\frac{4}{\alpha_\lambda}\left(\frac{a_1^{t-1}}{a_1^t}-1\right)\sum_{i=1}^M\left\|\lambda_i^{t+1}\right\|^2,\\
K_2^{t+1}&=\frac{4}{{\alpha_\phi}^2a_2^{t+1}}\sum_{i=1}^M\left\|\phi_i^{t+1}-\phi_i^t\right\|^2-\frac{4}{\alpha_\phi}\left(\frac{a_2^{t-1}}{a_2^t}-1\right)\sum_{i=1}^M\left\|\phi_i^{t+1}\right\|^2,\\
G^{t+1}&=\mathcal{L}_\lambda\left(\left\{\boldsymbol{\omega}_i^{t+1}\right\},\left\{\epsilon_i^{t+1}\right\},\boldsymbol{z}^{t+1},\left\{\lambda_i^{t+1}\right\},\left\{\phi_i^{t+1}\right\}\right)+K_1^{t+1}+K_2^{t+1}\\
&-\frac{7}{2\alpha_\lambda}\sum_{i=1}^M\left\|\lambda_i^{t+1}-\lambda_i^t\right\|^2-\frac{a_1^t}{2}\sum_{i=1}^M\left\|\lambda_i^{t+1}\right\|^2-\frac{7}{2\alpha_\phi}\sum_{i=1}^M\left\|\phi_i^{t+1}-\phi_i^t\right\|^2-\frac{a_2^t}{2}\sum_{i=1}^M\left\|\phi_i^{t+1}\right\|^2.
\end{aligned} \tag{57}$$

Let $b_5=\max\left\{1,1+L^2,6\varkappa k_0ML^2+\psi\right\}$, then we have,

$$\begin{aligned}
G^{t+1}-G^t\le&\Big(\frac{L+b_5}{2}-\frac{1}{\alpha_{\boldsymbol{\omega}}}+\frac{(M\alpha_\lambda+S\alpha_\phi)L^2}{2}+\frac{8ML^2}{\alpha_\lambda\left(a_1^t\right)^2}+\frac{8ML^2}{\alpha_\phi\left(a_2^t\right)^2}\Big)\sum_{i=1}^M\left\|\boldsymbol{\omega}_i^{t+1}-\boldsymbol{\omega}_i^t\,|\right\|^2\\
&+\Big(\frac{L+b_5}{2}-\frac{1}{\alpha_\epsilon}+\frac{(M\alpha_\lambda+S\alpha_\phi)L^2}{2}+\frac{8ML^2}{\alpha_\lambda\left(a_1^t\right)^2}+\frac{8ML^2}{\alpha_\phi\left(a_2^t\right)^2}\Big)\sum_{i=1}^M\left\|\epsilon_i^{t+1}-\epsilon_i^t\right\|^2\\
&+\Big(\frac{L+b_5}{2}-\frac{1}{\alpha_{\boldsymbol{z}}}+\frac{(M\alpha_\lambda+S\alpha_\phi)L^2}{2}+\frac{8ML^2}{\alpha_\lambda\left(a_1^t\right)^2}+\frac{8ML^2}{\alpha_\phi\left(a_2^t\right)^2}\Big)\left\|\boldsymbol{z}^{t+1}-\boldsymbol{z}^t\,|\right\|^2\\
&-\Big(\frac{1}{10\alpha_\lambda}-\frac{6\varkappa k_0ML^2}{2}\Big)\sum_{i=1}^M\left\|\lambda_i^{t+1}-\lambda_i^t\right\|^2-\frac{1}{10\alpha_\phi}\sum_{i=1}^M\|\phi_i^{t+1}-\phi_i^t\|^2+\frac{a_1^{t-1}-a_1^t}{2}\sum_{i=1}^M\|\lambda_i^{t+1}\|^2\\
&+\frac{a_2^{t-1}-a_2^t}{2}\sum_{i=1}^M\left\|\phi_i^{t+1}\right\|^2+\frac{4}{\alpha_\lambda}\Big(\frac{a_1^{t-2}}{a_1^{t-1}}-\frac{a_1^{t-1}}{a_1^t}\Big)\sum_{i=1}^M\left\|\lambda_i^t\,|\right\|^2+\frac{4}{\alpha_\phi}\Big(\frac{a_2^{t-2}}{a_2^{t-1}}-\frac{a_2^{t-1}}{a_2^t}\Big)\sum_{i=1}^M\left\|\phi_i^t\right\|^2.
\end{aligned} \tag{58}$$

Proof of Lemma 3:

We assume $b_1 = \frac{1}{\alpha_\lambda}, b_3 = \frac{1}{\alpha_\phi}$, and substitute them into Lemma 2, we have,

$$
\begin{aligned}
&\mathcal{L}_\lambda\left(\left\{\boldsymbol{\omega}_i^{t+1}\right\},\left\{\epsilon_i^{t+1}\right\},\boldsymbol{z}^{t+1},\left\{\lambda_i^{t+1}\right\},\left\{\phi_i^{t+1}\right\}\right)-\mathcal{L}_\lambda\left(\left\{\boldsymbol{\omega}_i^{t}\right\},\left\{\epsilon_i^{t}\right\},\boldsymbol{z}^{t},\left\{\lambda_i^{t}\right\},\left\{\phi_i^{t}\right\}\right)\\
&\leq\Big(\frac{L+L^2+1}{2}-\frac{1}{\alpha_{\boldsymbol{\omega}}}+\frac{(M\alpha_\lambda+S\alpha_\phi)L^2}{2}\Big)\sum_{i=1}^{M}\left\|\boldsymbol{\omega}_i^{t+1}-\boldsymbol{\omega}_i^{t}\right\|^2\\
&+\Big(\frac{L+1}{2}-\frac{1}{\alpha_\epsilon}+\frac{(M\alpha_\lambda+S\alpha_\phi)L^2}{2}\Big)\sum_{i=1}^{M}\left\|\epsilon_i^{t+1}-\epsilon_i^{t}\right\|^2+\frac{1}{2\alpha_\lambda}\sum_{i=1}^{M}\left\|\lambda_i^{t}-\lambda_i^{t-1}\right\|^2\\
&+\Big(\frac{L+6\varkappa k_0ML^2+\psi}{2}-\frac{1}{\alpha_{\boldsymbol{z}}}+\frac{(M\alpha_\lambda+S\alpha_\phi)L^2}{2}\Big)\left\|\boldsymbol{z}^{t+1}-\boldsymbol{z}^{t}\right\|^2+4\psi^2+2\psi+\frac{1}{2\alpha_\phi}\sum_{i=1}^{M}\left\|\phi_i^{t}-\phi_i^{t-1}\right\|^2\\
&+\Big(\frac{6\varkappa k_0ML^2}{2}-\frac{a_1^{t-1}-a_1^t}{2}+\frac{1}{\alpha_\lambda}\Big)\sum_{i=1}^{M}\left\|\lambda_i^{t+1}-\lambda_i^{t}\right\|^2+\left(\frac{1}{\alpha_\phi}-\frac{a_2^{t-1}-a_2^t}{2}\right)\sum_{i=1}^{M}\left\|\phi_i^{t+1}-\phi_i^{t}\right\|^2\\
&+\frac{a_1^{t-1}}{2}\sum_{i=1}^{M}\left(\left\|\lambda_i^{t+1}\right\|^2-\left\|\lambda_i^{t}\right\|^2\right)+\frac{a_2^{t-1}}{2}\sum_{i=1}^{M}\left(\left\|\phi_i^{t+1}\right\|^2-\left\|\phi_i^{t}\right\|^2\right),
\end{aligned}
\tag{59}
$$

$$
\lambda_i^{t+1}=\lambda_i^t+\alpha_\lambda\nabla_{\lambda_i}\bar{\mathcal{L}}_\lambda(\{\boldsymbol{\omega}_i^{t+1}\},\{\epsilon_i^{t+1}\},\boldsymbol{z}^{t+1},\{\lambda_i^t\},\{\phi_i^t\}).
\tag{60}
$$

According to Eq. (60), we can get the following equation in $(t+1)$-th iteration,

$$
\left\langle\lambda_i^{t+1}-\lambda_i^t-\alpha_\lambda\nabla_{\lambda_i}\bar{\mathcal{L}}_\lambda\left(\left\{\boldsymbol{\omega}_i^{t+1}\right\},\left\{\epsilon_i^{t+1}\right\},\boldsymbol{z}^{t+1},\left\{\lambda_i^t\right\},\left\{\phi_i^t\right\}\right),\lambda_i^t-\lambda_i^{t+1}\right\rangle=0.
\tag{61}
$$

Similar to Eq. (61), we can get the following equation in t-th iteration,

$$
\left\langle\lambda_i^{t}-\lambda_i^{t-1}-\alpha_\lambda\nabla_{\lambda_i}\bar{\mathcal{L}}_\lambda\left(\left\{\boldsymbol{\omega}_i^{t}\right\},\left\{\epsilon_i^{t}\right\},\boldsymbol{z}^{t},\left\{\lambda_i^{t-1}\right\},\left\{\phi_i^{t-1}\right\}\right),\lambda_i^{t+1}-\lambda_i^{t}\right\rangle=0.
\tag{62}
$$

With the combination of Eq. (61) and Eq. (62), we have,

$$
\begin{aligned}
&\frac{1}{\alpha_\lambda}\left\langle\boldsymbol{d}_{1,i}^{t+1},\lambda_i^{t+1}-\lambda_i^t\right\rangle\\
&=\left\langle\nabla_{\lambda_i}\bar{\mathcal{L}}_\lambda\left(\left\{\boldsymbol{\omega}_i^{t+1}\right\},\left\{\epsilon_i^{t+1}\right\},\boldsymbol{z}^{t+1},\left\{\lambda_i^t\right\},\left\{\phi_i^t\right\}\right)-\nabla_{\lambda_i}\bar{\mathcal{L}}_\lambda\left(\left\{\boldsymbol{\omega}_i^{t}\right\},\left\{\epsilon_i^{t}\right\},\boldsymbol{z}^{t},\left\{\lambda_i^{t-1}\right\},\left\{\phi_i^{t-1}\right\}\right),\lambda_i^{t+1}-\lambda_i^t\right\rangle\\
&=\left\langle\nabla_{\lambda_i}\bar{\mathcal{L}}_\lambda\left(\left\{\boldsymbol{\omega}_i^{t+1}\right\},\left\{\epsilon_i^{t+1}\right\},\boldsymbol{z}^{t+1},\left\{\lambda_i^t\right\},\left\{\phi_i^t\right\}\right)-\nabla_{\lambda_i}\bar{\mathcal{L}}_\lambda\left(\left\{\boldsymbol{\omega}_i^{t}\right\},\left\{\epsilon_i^{t}\right\},\boldsymbol{z}^{t},\left\{\lambda_i^{t}\right\},\left\{\phi_i^{t}\right\}\right),\lambda_i^{t+1}-\lambda_i^t\right\rangle\\
&\quad+\left\langle\nabla_{\lambda_i}\bar{\mathcal{L}}_\lambda\left(\left\{\boldsymbol{\omega}_i^{t}\right\},\left\{\epsilon_i^{t}\right\},\boldsymbol{z}^{t},\left\{\lambda_i^t\right\},\left\{\phi_i^t\right\}\right)-\nabla_{\lambda_i}\bar{\mathcal{L}}_\lambda\left(\left\{\boldsymbol{\omega}_i^{t}\right\},\left\{\epsilon_i^{t}\right\},\boldsymbol{z}^{t},\left\{\lambda_i^{t-1}\right\},\left\{\phi_i^{t-1}\right\}\right),\boldsymbol{d}_{1,i}^{t+1}\right\rangle\\
&\quad+\left\langle\nabla_{\lambda_i}\bar{\mathcal{L}}_\lambda\left(\left\{\boldsymbol{\omega}_i^{t}\right\},\left\{\epsilon_i^{t}\right\},\boldsymbol{z}^{t},\left\{\lambda_i^t\right\},\left\{\phi_i^t\right\}\right)-\nabla_{\lambda_i}\bar{\mathcal{L}}_\lambda\left(\left\{\boldsymbol{\omega}_i^{t}\right\},\left\{\epsilon_i^{t}\right\},\boldsymbol{z}^{t},\left\{\lambda_i^{t-1}\right\},\left\{\phi_i^{t-1}\right\}\right),\lambda_i^t-\lambda_i^{t-1}\right\rangle.
\end{aligned}
\tag{63}
$$

Since we have that,

$$
\frac{1}{\alpha_\lambda}\left\langle\boldsymbol{d}_{1,i}^{t+1},\lambda_i^{t+1}-\lambda_i^t\right\rangle=\frac{1}{2\alpha_\lambda}\left\|\lambda_i^{t+1}-\lambda_i^t\right\|^2+\frac{1}{2\alpha_\lambda}\left\|\boldsymbol{d}_{1,i}^{t+1}\right\|^2-\frac{1}{2\alpha_\lambda}\left\|\lambda_i^t-\lambda_i^{t-1}\right\|^2.
\tag{64}
$$

Based on Eq. (63) and Eq. (64), we have,

$$
\begin{aligned}
&\frac{1}{2\alpha_\lambda}\left\|\lambda_i^{t+1}-\lambda_i^t\right\|^2+\frac{1}{2\alpha_\lambda}\left\|\boldsymbol{d}_{1,i}^{t+1}\right\|^2-\frac{1}{2\alpha_\lambda}\left\|\lambda_i^t-\lambda_i^{t-1}\right\|^2\\
&=\frac{L^2}{2h_1^t}\Big(\sum_{i=1}^{M}\left(\left\|\boldsymbol{\omega}_i^{t+1}-\boldsymbol{\omega}_i^t\right\|^2+\left\|\epsilon_i^{t+1}-\epsilon_i^t\right\|^2\right)+\boldsymbol{z}^{t+1}-\boldsymbol{z}^t\|^2\Big)+\frac{h_1^t}{2}\left\|\lambda_i^{t+1}-\lambda_i^t\right\|^2\\
&+\frac{a_1^{t-1}-a_1^t}{2}\left(\left\|\lambda_i^{t+1}\right\|^2-\left\|\lambda_i^t\right\|^2\right)-\frac{a_1^{t-1}-a_1^t}{2}\left\|\lambda_i^{t+1}-\lambda_i^t\right\|^2\\
&+\frac{\alpha_\lambda}{2}\left\|\nabla_{\lambda_i}\bar{\mathcal{L}}_\lambda\left(\left\{\boldsymbol{\omega}_i^{t}\right\},\left\{\epsilon_i^{t}\right\},\boldsymbol{z}^{t},\left\{\lambda_i^t\right\},\left\{\phi_i^t\right\}\right)-\nabla_{\lambda_i}\bar{\mathcal{L}}_\lambda\left(\left\{\boldsymbol{\omega}_i^{t}\right\},\left\{\epsilon_i^{t}\right\},\boldsymbol{z}^{t},\left\{\lambda_i^{t-1}\right\},\left\{\phi_i^{t-1}\right\}\right)\right\|^2+\frac{1}{2\alpha_\lambda}\left\|\boldsymbol{d}_{1,i}^{t+1}\right\|^2\\
&-\frac{1}{L_1'+a_1^{t-1}}\left\|\nabla_{\lambda_i}\bar{\mathcal{L}}_\lambda\left(\left\{\boldsymbol{\omega}_i^{t}\right\},\left\{\epsilon_i^{t}\right\},\boldsymbol{z}^{t},\left\{\lambda_i^t\right\},\left\{\phi_i^t\right\}\right)-\nabla_{\lambda_i}\bar{\mathcal{L}}_\lambda\left(\left\{\boldsymbol{\omega}_i^{t}\right\},\left\{\epsilon_i^{t}\right\},\boldsymbol{z}^{t},\left\{\lambda_i^{t-1}\right\},\left\{\phi_i^{t-1}\right\}\right)\right\|^2-\frac{a_1^{t-1}L_1'}{L_1'+a_1^{t-1}}\left\|\lambda_i^t-\lambda_i^{t-1}\right\|^2,
\end{aligned}
\tag{65}
$$

where $h_1^t > 0$. Let $a_1^0 \leq L_1'$ and then $-\frac{a_1^{t-1}L_1'}{L_1'+a_1^{t-1}} \leq -\frac{a_1^{t-1}L_1'}{2L_1'} = -\frac{a_1^{t-1}}{2} \leq -\frac{a_1^t}{2}$. Additionally, multiplying both sides of Eq. (65) by $\frac{8}{\alpha_\lambda a_1^t}$, we have,

$$\begin{aligned}
&\frac{4}{{\alpha_\lambda}^2 a_1^t}\left\|\lambda_i^{t+1}-\lambda_i^t\right\|^2 - \frac{4}{\alpha_\lambda}\Big(\frac{a_1^{t-1}}{a_1^t}-1\Big)\left\|\lambda_i^{t+1}\right\|^2 \\
&\leq \frac{4}{{\alpha_\lambda}^2 a_1^t}\left\|\lambda_i^t-\lambda_i^{t-1}\right\|^2 - \frac{4}{\alpha_\lambda}\Big(\frac{a_1^{t-1}}{a_1^t}-1\Big)\left\|\lambda_i^t\right\|^2 + \frac{4h_1^t}{\alpha_\lambda a_1^t}\left\|\lambda_i^{t+1}-\lambda_i^t\right\|^2 - \frac{4}{\alpha_\lambda}\left\|\lambda_i^t-\lambda_i^{t-1}\right\|^2 \\
&\quad + \frac{4L^2}{\alpha_\lambda a_1^t h_1^t}\Big(\sum_{i=1}^M \big(\left\|\boldsymbol{\omega}_i^{t+1}-\boldsymbol{\omega}_i^t\right\|^2 + \left\|\epsilon_i^{t+1}-\epsilon_i^t\right\|^2\big) + \left\|\boldsymbol{z}^{t+1}-\boldsymbol{z}^t\right\|^2\Big).
\end{aligned} \tag{66}$$

According to the definition of K_1^t and setting $h_1^t = \frac{a_1^t}{2}$ in Eq. (66), we have,

$$\begin{aligned}
K_1^{t+1}-K_1^t \leq &\sum_{i=1}^M \frac{4}{\alpha_\lambda}\Big(\frac{a_1^{t-2}}{a_1^{t-1}}-\frac{a_1^{t-1}}{a_1^t}\Big)\left\|\lambda_i^t\right\|^2 + \sum_{i=1}^M\Big(\frac{2}{\alpha_\lambda}+\frac{4}{{\alpha_\lambda}^2}\big(\frac{1}{a_1^{t+1}}-\frac{1}{a_1^t}\big)\Big)\left\|\lambda_i^{t+1}-\lambda_i^t\right\|^2 \\
&-\sum_{i=1}^M \frac{4}{\alpha_\lambda}\|\lambda_i^t-\lambda_i^{t-1}\|^2 + \frac{8ML^2}{\alpha_\lambda\,(a_1^t)^2}\Big(\sum_{i=1}^M \big(\left\|\boldsymbol{\omega}_i^{t+1}-\boldsymbol{\omega}_i^t\right\|^2+\left\|\epsilon_i^{t+1}-\epsilon_i^t\right\|^2\big) + \boldsymbol{z}^{t+1}-\boldsymbol{z}^t\|^2\Big).
\end{aligned} \tag{67}$$

Likewise, according to Eq. (30), we have,

$$\begin{aligned}
&\frac{1}{\alpha_\phi}\left\langle \boldsymbol{d}_{2,i}^{t+1}, \phi_i^{t+1}-\phi_i^t\right\rangle \\
&= \left\langle \nabla_{\phi_i}\bar{\mathcal{L}}_\lambda\left(\left\{\boldsymbol{\omega}_i^{t+1}\right\},\left\{\epsilon_i^{t+1}\right\},\boldsymbol{z}^{t+1},\left\{\lambda_i^{t+1}\right\},\left\{\phi_i^t\right\}\right) - \nabla_{\phi_i}\bar{\mathcal{L}}_\lambda\left(\left\{\boldsymbol{\omega}_i^t\right\},\left\{\epsilon_i^t\right\},\boldsymbol{z}^t,\left\{\lambda_i^t\right\},\left\{\phi_i^t\right\}\right), \phi_i^{t+1}-\phi_i^t\right\rangle \\
&+ \left\langle \nabla_{\phi_i}\bar{\mathcal{L}}_\lambda\left(\left\{\boldsymbol{\omega}_i^t\right\},\left\{\epsilon_i^t\right\},\boldsymbol{z}^t,\left\{\lambda_i^t\right\},\left\{\phi_i^t\right\}\right) - \nabla_{\phi_i}\bar{\mathcal{L}}_\lambda\left(\left\{\boldsymbol{\omega}_i^t\right\},\left\{\epsilon_i^t\right\},\boldsymbol{z}^t,\left\{\lambda_i^t\right\},\left\{\phi_i^{t-1}\right\}\right), \boldsymbol{d}_{2,i}^{t+1}\right\rangle \\
&+ \left\langle \nabla_{\phi_i}\bar{\mathcal{L}}_\lambda\left(\left\{\boldsymbol{\omega}_i^t\right\},\left\{\epsilon_i^t\right\},\boldsymbol{z}^t,\left\{\lambda_i^t\right\},\left\{\phi_i^t\right\}\right) - \nabla_{\phi_i}\bar{\mathcal{L}}_\lambda\left(\left\{\boldsymbol{\omega}_i^t\right\},\left\{\epsilon_i^t\right\},\boldsymbol{z}^t,\left\{\lambda_i^t\right\},\left\{\phi_i^{t-1}\right\}\right), \phi_i^t-\phi_i^{t-1}\right\rangle.
\end{aligned} \tag{68}$$

Furthermore, since

$$\frac{1}{\alpha_\phi}\left\langle \boldsymbol{d}_{2,i}^{t+1}, \phi_i^{t+1}-\phi_i^t\right\rangle = \frac{1}{2\alpha_\phi}\left\|\phi_i^{t+1}-\phi_i^t\right\|^2 + \frac{1}{2\alpha_\phi}\left\|\boldsymbol{d}_{2,i}^{t+1}\right\|^2 - \frac{1}{2\alpha_\phi}\left\|\phi_i^t-\phi_i^{t-1}\right\|^2, \tag{69}$$

we can get that,

$$\begin{aligned}
&\frac{1}{2\alpha_\phi}\left\|\phi_i^{t+1}-\phi_i^t\right\|^2 + \frac{1}{2\alpha_\phi}\left\|\boldsymbol{d}_{2,i}^{t+1}\right\|^2 - \frac{1}{2\alpha_\phi}\left\|\phi_i^t-\phi_i^{t-1}\right\|^2 \\
&= \frac{L^2}{2h_2^t}\Big(\sum_{i=1}^M\big(\left\|\boldsymbol{\omega}_i^{t+1}-\boldsymbol{\omega}_i^t\right\|^2+\left\|\epsilon_i^{t+1}-\epsilon_i^t\right\|^2\big)+\left\|\boldsymbol{z}^{t+1}-\boldsymbol{z}^t\right\|^2\Big) + \frac{h_2^t}{2}\left\|\phi_i^{t+1}-\phi_i^t\right\|^2 \\
&+ \frac{a_2^{t-1}-a_2^t}{2}\left(\left\|\phi_i^{t+1}\right\|^2-\left\|\phi_i^t\right\|^2\right) - \frac{a_2^{t-1}-a_2^t}{2}\left\|\phi_i^{t+1}-\phi_i^t\right\|^2 - \frac{a_2^{t-1}L_2'}{L_2'+a_2^{t-1}}\left\|\phi_i^t-\phi_i^{t-1}\right\|^2 + \frac{1}{2\alpha_\phi}\left\|\boldsymbol{d}_{2,i}^{t+1}\right\|^2 \\
&+ \frac{\alpha_\phi}{2}\left\|\nabla_{\phi_i}\bar{\mathcal{L}}_\lambda\left(\left\{\boldsymbol{\omega}_i^t\right\},\left\{\epsilon_i^t\right\},\boldsymbol{z}^t,\left\{\lambda_i^t\right\},\left\{\phi_i^t\right\}\right) - \nabla_{\phi_i}\bar{\mathcal{L}}_\lambda\left(\left\{\boldsymbol{\omega}_i^t\right\},\left\{\epsilon_i^t\right\},\boldsymbol{z}^t,\left\{\lambda_i^t\right\},\left\{\phi_i^{t-1}\right\}\right)\right\|^2 \\
&- \frac{1}{L_2'+a_2^{t-1}}\left\|\nabla_{\phi_i}\bar{\mathcal{L}}_\lambda\left(\left\{\boldsymbol{\omega}_i^t\right\},\left\{\epsilon_i^t\right\},\boldsymbol{z}^t,\left\{\lambda_i^t\right\},\left\{\phi_i^t\right\}\right) - \nabla_{\phi_i}\bar{\mathcal{L}}_\lambda\left(\left\{\boldsymbol{\omega}_i^t\right\},\left\{\epsilon_i^t\right\},\boldsymbol{z}^t,\left\{\lambda_i^t\right\},\left\{\phi_i^{t-1}\right\}\right)\right\|^2.
\end{aligned} \tag{70}$$

Let $a_2^0 \leq L_2'$ and then $-\frac{a_2^{t-1}L_2'}{L_2'+a_2^{t-1}} \leq -\frac{a_2^{t-1}L_2'}{2L_2'} = -\frac{a_2^{t-1}}{2} \leq -\frac{a_2^t}{2}$. Multiplying both sides of Eq. (70) by $\frac{8}{\alpha_\phi a_2^t}$, we have,

$$\begin{aligned}
&\frac{4}{{\alpha_\phi}^2 a_2^t}\left\|\phi_i^{t+1}-\phi_i^t\right\|^2 - \frac{4}{\alpha_\phi}\left(\frac{a_2^{t-1}}{a_2^t}-1\right)\left\|\phi_i^{t+1}\right\|^2 \\
&\leq \frac{4}{{\alpha_\phi}^2 a_2^t}\left\|\phi_i^t-\phi_i^{t-1}\right\|^2 - \frac{4}{\alpha_\phi}\left(\frac{a_2^{t-1}}{a_2^t}-1\right)\left\|\phi_i^t\right\|^2 + \frac{4h_2^t}{\alpha_\phi a_2^t}\left\|\phi_i^{t+1}-\phi_i^t\right\|^2 - \frac{4}{\alpha_\phi}\left\|\phi_i^t-\phi_i^{t-1}\right\|^2 \\
&+ \frac{4L^2}{\alpha_\phi a_2^t h_2^t}\Big(\sum_{i=1}^M\big(\left\|\boldsymbol{\omega}_i^{t+1}-\boldsymbol{\omega}_i^t\right\|^2+\left\|\epsilon_i^{t+1}-\epsilon_i^t\right\|^2\big)+\left\|\boldsymbol{z}^{t+1}-\boldsymbol{z}^t\right\|^2\Big).
\end{aligned} \tag{71}$$

According to the definition of K_2^t and setting $h_2^t = \frac{a_2^t}{2}$ in Eq. (71), we have,

$$\begin{aligned}
K_2^{t+1}-K_2^t \leq &\sum_{i=1}^M \frac{4}{\alpha_\phi}\Big(\frac{a_2^{t-2}}{a_2^{t-1}}-\frac{a_2^{t-1}}{a_2^t}\Big)\left\|\phi_i^t\right\|^2 + \sum_{i=1}^M\Big(\frac{2}{\alpha_\phi}+\frac{4}{{\alpha_\phi}^2}\big(\frac{1}{a_2^{t+1}}-\frac{1}{a_2^t}\big)\Big)\left\|\phi_i^{t+1}-\phi_i^t\right\|^2 \\
&-\sum_{i=1}^M \frac{4}{\alpha_\phi}\left\|\phi_i^t-\phi_i^{t-1}\right\|^2 + \frac{8ML^2}{\alpha_\phi\,(a_2^t)^2}\Big(\sum_{i=1}^M\big(\left\|\boldsymbol{\omega}_i^{t+1}-\boldsymbol{\omega}_i^t\right\|^2+\left\|\epsilon_i^{t+1}-\epsilon_i^t\right\|^2\big)+\left\|\boldsymbol{z}^{t+1}-\boldsymbol{z}^t\right\|^2\Big).
\end{aligned} \tag{72}$$

Following the setting of a_1^t and a_2^t, we can achieve that $\frac{\alpha_\lambda}{10} \geq \frac{1}{a_1^{t+1}} - \frac{1}{a_1^t}, \frac{\alpha_\phi}{10} \geq \frac{1}{a_2^{t+1}} - \frac{1}{a_2^t}$. Additionally, we set $b_5 = \max\left\{1, 1+L^2, 6\varkappa k_0 ML^2 + \psi\right\}$. Based on the definition of G^{t+1}, Eq. (67), and (72), we have,

$$
\begin{aligned}
&G^{t+1} - G^t \\
&\leq \Big(\frac{L+b_5}{2} - \frac{1}{\alpha_{\boldsymbol{\omega}}} + \frac{L^2(M\alpha_\lambda + S\alpha_\phi)}{2} + 8ML^2\big(\frac{1}{\alpha_\lambda\left(a_1^t\right)^2} + \frac{1}{\alpha_\phi\left(a_2^t\right)^2}\big)\Big)\sum_{i=1}^{M}\left\|\boldsymbol{\omega}_i^{t+1} - \boldsymbol{\omega}_i^t\right\|^2 \\
&+ \Big(\frac{L+b_5}{2} - \frac{1}{\alpha_\epsilon} + \frac{L^2(M\alpha_\lambda + S\alpha_\phi)}{2} + 8ML^2\big(\frac{1}{\alpha_\lambda\left(a_1^t\right)^2} + \frac{1}{\alpha_\phi\left(a_2^t\right)^2}\big)\Big)\sum_{i=1}^{M}\left\|\epsilon_i^{t+1} - \epsilon_i^t\right\|^2 \\
&+ \Big(\frac{L+b_5}{2} - \frac{1}{\alpha_{\boldsymbol{z}}} + \frac{L^2(M\alpha_\lambda + S\alpha_\phi)}{2} + 8ML^2\big(\frac{1}{\alpha_\lambda\left(a_1^t\right)^2} + \frac{1}{\alpha_\phi\left(a_2^t\right)^2}\big)\Big)\|\boldsymbol{z}^{t+1} - \boldsymbol{z}^t\|^2 \\
&- \Big(\frac{1}{10\alpha_\lambda} - \frac{6\varkappa k_0 ML^2}{2}\Big)\sum_{i=1}^{M}\|\lambda_i^{t+1} - \lambda_i^t\|^2 - \frac{1}{10\alpha_\phi}\sum_{i=1}^{M}\|\phi_i^{t+1} - \phi_i^t\|^2 + \frac{a_1^{t-1} - a_1^t}{2}\sum_{i=1}^{M}\left\|\lambda_i^{t+1}\right\|^2 \\
&+ \frac{a_2^{t-1} - a_2^t}{2}\sum_{i=1}^{M}\left\|\phi_i^{t+1}\right\|^2 + \frac{4}{\alpha_\lambda}\Big(\frac{a_1^{t-2}}{a_1^{t-1}} - \frac{a_1^{t-1}}{a_1^t}\Big)\sum_{i=1}^{M}\left\|\lambda_i^t\right\|^2 + \frac{4}{\alpha_\phi}\Big(\frac{a_2^{t-2}}{a_2^{t-1}} - \frac{a_2^{t-1}}{a_2^t}\Big)\sum_{i=1}^{M}\left\|\phi_i^t\right\|^2,
\end{aligned} \tag{73}
$$

which concludes the proof of Lemma 3.

Appendix F
Proof of Theorem 1

Combining Lemma 1, Lemma 2, and Lemma 3, we are going to derive Theorem 1 for our proposed algorithm. To begin with, we assume that,

$$
b_6^t = 4ML^2(\beta - 2)\Big(\frac{1}{\alpha_\lambda\left(a_1^t\right)^2} + \frac{1}{\alpha_\phi\left(a_2^t\right)^2}\Big) + \frac{\alpha_\phi(M-S)L^2}{2} - \frac{b_5}{2}, \tag{74}
$$

where β is constant that satisfies $\beta > 2$ and $4ML^2(\beta-2)\big(\frac{1}{\alpha_\lambda\left(a_1^t\right)^2} + \frac{1}{\alpha_\phi\left(a_2^t\right)^2}\big) > \frac{b_5}{2}$. Therefore, $b_6^t > 0$. Based on the setting of $\alpha_{\boldsymbol{\omega}}, \alpha_\epsilon, \alpha_{\boldsymbol{z}}$ and a_1^t, a_2^t, we have,

$$
\frac{L+b_5}{2} - \frac{1}{\alpha_{\boldsymbol{\omega}}} + \frac{L^2(M\alpha_\lambda + S\alpha_\phi)}{2} + 8ML^2\Big(\frac{1}{\alpha_\lambda\left(a_1^t\right)^2} + \frac{1}{\alpha_\phi\left(a_2^t\right)^2}\Big) = -b_6^t, \tag{75}
$$

$$
\frac{L+b_5}{2} - \frac{1}{\alpha_\epsilon} + \frac{L^2(M\alpha_\lambda + S\alpha_\phi)}{2} + 8ML^2\Big(\frac{1}{\alpha_\lambda\left(a_1^t\right)^2} + \frac{1}{\alpha_\phi\left(a_2^t\right)^2}\Big) = -b_6^t, \tag{76}
$$

$$
\frac{L+b_5}{2} - \frac{1}{\alpha_{\boldsymbol{z}}} + \frac{L^2(M\alpha_\lambda + S\alpha_\phi)}{2} + 8ML^2\Big(\frac{1}{\alpha_\lambda\left(a_1^t\right)^2} + \frac{1}{\alpha_\phi\left(a_2^t\right)^2}\Big) = -b_6^t. \tag{77}
$$

With the combination of Eq. (75), (76), (77) with Lemma 3, we have,

$$
\begin{aligned}
&b_6^t\sum_{i=1}^{M}\Big(\left\|\boldsymbol{\omega}_i^{t+1} - \boldsymbol{\omega}_i^t\right\|^2 + \left\|\epsilon_i^{t+1} - \epsilon_i^t\right\|^2\Big) + b_6^t\left\|\boldsymbol{z}^{t+1} - \boldsymbol{z}^t\right\|^2 \\
&\quad + \Big(\frac{1}{10\alpha_\lambda} - 3\varkappa k_0 ML^2\Big)\sum_{i=1}^{M}\left\|\lambda_i^{t+1} - \lambda_i^t\right\|^2 + \frac{1}{10\alpha_\phi}\sum_{i=1}^{M}\left\|\phi_i^{t+1} - \phi_i^t\right\|^2 \\
&\leq G^t - G^{t+1} + \frac{a_1^{t-1} - a_1^t}{2}\sum_{i=1}^{M}\left\|\lambda_i^{t+1}\right\|^2 + \frac{a_2^{t-1} - a_2^t}{2}\sum_{i=1}^{M}\left\|\phi_i^{t+1}\right\|^2 \\
&\quad + \frac{4}{\alpha_\lambda}\Big(\frac{a_1^{t-2}}{a_1^{t-1}} - \frac{a_1^{t-1}}{a_1^t}\Big)\sum_{i=1}^{M}\left\|\lambda_i^t\right\|^2 + \frac{4}{\alpha_\phi}\Big(\frac{a_2^{t-2}}{a_2^{t-1}} - \frac{a_2^{t-1}}{a_2^t}\Big)\sum_{i=1}^{M}\left\|\phi_i^t\right\|^2.
\end{aligned} \tag{78}
$$

Combining the definition of $\left(\nabla\bar{F}^t\right)_{\boldsymbol{\omega}_i}, \left(\nabla\bar{F}^t\right)_{\epsilon_i}$ in Eq. (6), the trigonometric inequality, the Cauchy-Schwarz inequality with Assumption 1 and 2, we have,

$$
\left\|\left(\nabla\bar{F}^t\right)_{\boldsymbol{\omega}_i}\right\|^2 \leq \frac{2}{{\alpha_{\boldsymbol{\omega}}}^2}\left\|\boldsymbol{\omega}_i^{\tilde{t}} - \boldsymbol{\omega}_i^t\right\|^2 + 6L^2\varkappa k_0\left\|\boldsymbol{z}^{t+1} - \boldsymbol{z}^t\right\|^2 + 6L^2\varkappa k_0\sum_{i=1}^{M}\left\|\lambda_i^{t+1} - \lambda_i^t\right\|^2, \tag{79}
$$

$$
\left\|\left(\nabla\bar{F}^t\right)_{\epsilon_i}\right\|^2 \leq \frac{2}{{\alpha_\epsilon}^2}\left\|\epsilon_i^{\tilde{t}} - \epsilon_i^t\right\|^2 + 6L^2\varkappa k_0\left\|\boldsymbol{z}^{t+1} - \boldsymbol{z}^t\right\|^2 + 6L^2\varkappa k_0\sum_{i=1}^{M}\left\|\lambda_i^{t+1} - \lambda_i^t\right\|^2. \tag{80}
$$

Combining the definition of $\left(\nabla\bar{F}^t\right)_{\boldsymbol{z}}$ in Eq. (6), the trigonometric inequality, the Cauchy-Schwarz inequality with Assumption 1 and 2, we have,

$$\left\|\left(\nabla\bar{F}^t\right)_{\boldsymbol{z}}\right\|^2 \le 2L^2\sum_{i=1}^{M}\left(\left\|\boldsymbol{\omega}_i^{t+1}-\boldsymbol{\omega}_i^t\right\|^2+\left\|\epsilon_i^{t+1}-\epsilon_i^t\right\|^2\right)+\frac{2}{\alpha_{\boldsymbol{z}}^2}\left\|\boldsymbol{z}^{t+1}-\boldsymbol{z}^t\right\|^2. \tag{81}$$

Combining the definition of $\left(\nabla\bar{F}^t\right)_{\lambda_i}$ in Eq. (6), the trigonometric inequality, the Cauchy-Schwarz inequality with Assumption 1 and 2, we have,

$$\begin{aligned}\left\|\left(\nabla\bar{F}^t\right)_{\lambda_i}\right\|^2 \le{}& \frac{3}{{\alpha_\lambda}^2}\left\|\lambda_i^{t+1}-\lambda_i^t\right\|^2+3\left(\left(a_1^{t-1}\right)^2-\left(a_1^t\right)^2\right)\left\|\lambda_i^t\right\|^2\\&+3L^2\sum_{i=1}^{M}\left(\left\|\boldsymbol{\omega}_i^{t+1}-\boldsymbol{\omega}_i^t\right\|^2+\left\|\epsilon_i^{t+1}-\epsilon_i^t\right\|^2\right)+3L^2\left\|\boldsymbol{z}^{t+1}-\boldsymbol{z}^t\right\|^2.\end{aligned} \tag{82}$$

Combining the definition of $\left(\nabla\bar{F}^t\right)_{\phi_i}$ in Eq. (6), the Cauchy-Schwarz inequality with Assumption 1 and 2, we have,

$$\begin{aligned}\left\|\left(\nabla\bar{F}^t\right)_{\phi_i}\right\|^2 \le{}& \frac{3}{{\alpha_\phi}^2}\left\|\phi_i^{\tilde{t}}-\phi_i^t\right\|^2+3L^2\sum_{i=1}^{M}\left(\left\|\boldsymbol{\omega}_i^{\tilde{t}}-\boldsymbol{\omega}_i^t\right\|^2+\left\|\epsilon_i^{\tilde{t}}-\epsilon_i^t\right\|^2\right)+3L^2\varkappa k_0\left\|\boldsymbol{z}^{t+1}-\boldsymbol{z}^t\right\|^2\\&+3L^2\varkappa k_0\sum_{i=1}^{M}\left\|\lambda_i^{t+1}-\lambda_i^t\right\|^2+3\Big(\left(a_2^{\hat{t}-1}\right)^2-\left(a_2^{\tilde{t}-1}\right)^2\Big)\left\|\phi_i^t\right\|^2.\end{aligned} \tag{83}$$

With the combination of Definition 3, Eq. (79), (80), (81), (82), and (83), we can obtain the following inequality,

$$\begin{aligned}\left\|\nabla\bar{F}^t\right\|^2 ={}& \sum_{i=1}^{M}\left(\left\|\left(\nabla\bar{F}^t\right)_{\boldsymbol{\omega}_i}\right\|^2+\left\|\left(\nabla\bar{F}^t\right)_{\epsilon_i}\right\|^2+\left\|\left(\nabla\bar{F}^t\right)_{\lambda_i}\right\|^2+\left\|\left(\nabla\bar{F}^t\right)_{\phi_i}\right\|^2\right)+\left\|\left(\nabla\bar{F}^t\right)_{\boldsymbol{z}}\right\|^2\\
\le{}& \Big(\frac{2}{{\alpha_{\boldsymbol{\omega}}}^2}+3ML^2\Big)\sum_{i=1}^{M}\left\|\boldsymbol{\omega}_i^{\tilde{t}}-\boldsymbol{\omega}_i^t\right\|^2+\left(4+3ML^2\right)\sum_{i=1}^{M}\left\|\boldsymbol{\omega}_i^{t+1}-\boldsymbol{\omega}_i^t\right\|^2\\
&+\Big(\frac{2}{{\alpha_\epsilon}^2}+3ML^2\Big)\sum_{i=1}^{M}\left\|\epsilon_i^{\tilde{t}}-\epsilon_i^t\right\|^2+\left(4+3ML^2\right)\sum_{i=1}^{M}\left\|\epsilon_i^{t+1}-\epsilon_i^t\right\|^2\\
&+\Big(\frac{2}{\alpha_{\boldsymbol{z}}^2}+15\varkappa k_0ML^2+3ML^2\Big)\left\|\boldsymbol{z}^{t+1}-\boldsymbol{z}^t\right\|^2\\
&+\sum_{i=1}^{M}\Big(\frac{3}{{\alpha_\lambda}^2}+15\varkappa k_0ML^2\Big)\left\|\lambda_i^{t+1}-\lambda_i^t\right\|^2+\sum_{i=1}^{M}3\left(\left(a_1^{t-1}\right)^2-\left(a_1^t\right)^2\right)\left\|\lambda_i^t\right\|^2\\
&+\sum_{i=1}^{M}\frac{3}{{\alpha_\phi}^2}\left\|\phi_i^{\tilde{t}}-\phi_i^t\right\|^2+\sum_{i=1}^{M}3\Big(\left(a_2^{\hat{t}-1}\right)^2-\left(a_2^{\tilde{t}-1}\right)^2\Big)\left\|\phi_i^t\right\|^2.\end{aligned} \tag{84}$$

We denote that $\underline{b_6}>0$ is a constant that represents the lower bound of b_6^t. Additionally, we assume that c_1,c_2,c_3 are constants,

$$c_1=\frac{2k_\varkappa\varkappa+\left(4+3M+3k_\varkappa\varkappa M\right)L^2{\alpha_{\boldsymbol{\omega}}}^2}{{\alpha_{\boldsymbol{\omega}}}^2\left(b_6^t\right)^2}, \tag{85}$$

$$c_2=\frac{2k_\varkappa\varkappa+\left(4+3M+3k_\varkappa\varkappa M\right)L^2{\alpha_\epsilon}^2}{{\alpha_\epsilon}^2\left(b_6^t\right)^2}, \tag{86}$$

$$c_3=\frac{2+3\left(5\varkappa k_0+1\right)ML^2{\alpha_{\boldsymbol{z}}}^2}{{\alpha_{\boldsymbol{z}}}^2\left(b_6^t\right)^2}, \tag{87}$$

where $k_{\varkappa} > 0$ is a constant. Then, combining Eq. (84), (85), (86) with (87), we have,

$$
\begin{aligned}
\left\|\nabla \bar{F}^t\right\|^2 \leq & \sum_{i=1}^{M} \left(b_6^t\right)^2 \left(c_1 \left\|\boldsymbol{\omega}_i^{t+1} - \boldsymbol{\omega}_i^t\right\|^2 + c_2 \left\|\epsilon_i^{t+1} - \epsilon_i^t\right\|^2\right) + c_3 \left(b_6^t\right)^2 \left\|\boldsymbol{z}^{t+1} - \boldsymbol{z}^t\right\|^2 \\
& + \left(\frac{2}{{\alpha_{\boldsymbol{\omega}}}^2} + 3ML^2\right) \sum_{i=1}^{M} \left\|\boldsymbol{\omega}_i^{\tilde{t}} - \boldsymbol{\omega}_i^t\right\|^2 - \left(\frac{2k_{\varkappa}\varkappa}{{\alpha_{\boldsymbol{\omega}}}^2} + 3k_{\varkappa}\varkappa ML^2\right) \sum_{i=1}^{M} \left\|\boldsymbol{\omega}_i^{t+1} - \boldsymbol{\omega}_i^t\right\|^2 \\
& + \left(\frac{2}{{\alpha_{\epsilon}}^2} + 3ML^2\right) \sum_{i=1}^{M} \left\|\epsilon_i^{\tilde{t}} - \epsilon_i^t\right\|^2 - \left(\frac{2k_{\varkappa}\varkappa}{{\alpha_{\epsilon}}^2} + 3k_{\varkappa}\varkappa ML^2\right) \sum_{i=1}^{M} \left\|\epsilon_i^{t+1} - \epsilon_i^t\right\|^2 \\
& + \sum_{i=1}^{M} 3\left(\frac{1}{{\alpha_\lambda}^2} + 5\varkappa k_0 ML^2\right) \left\|\lambda_i^{t+1} - \lambda_i^t\right\|^2 + \sum_{i=1}^{M} 3 \left(\left(a_1^{t-1}\right)^2 - \left(a_1^t\right)^2\right) \left\|\lambda_i^t\right\|^2 \\
& + \sum_{i=1}^{M} \frac{3}{{\alpha_\phi}^2} \left\|\phi_i^{\tilde{t}} - \phi_i^t\right\|^2 + \sum_{i=1}^{M} 3\left(\left(a_2^{\hat{t}-1}\right)^2 - \left(a_2^{\tilde{t}-1}\right)^2\right) \left\|\phi_i^t\right\|^2 .
\end{aligned} \tag{88}
$$

We denote c_4^t as a nonnegative sequence,

$$
c_4^t = \frac{1}{\max\left\{c_1 b_6^t, c_2 b_6^t, c_3 b_6^t, \frac{\frac{30}{\alpha_\lambda} + 150\alpha_\lambda \varkappa k_0 ML^2}{1 - 30\alpha_\lambda \varkappa k_0 ML^2}, \frac{30\varkappa}{\alpha_\phi}\right\}}. \tag{89}
$$

Additionally, we denote the upper bound of c_4^t as $\overline{c_4}$ and the lower bound of c_4^t as $\underline{c_4}$, i.e., $\overline{c_4} \geq c_4^t \geq \underline{c_4} \geq 0$. $k_{\varkappa}$ is a constant that satisfies $k_{\varkappa} \geq \max\left\{\frac{\overline{c_4}\left(\frac{2}{\overline{\alpha_\epsilon}^2} + 3ML^2\right)}{\underline{c_4}\left(\frac{2}{\overline{\alpha_\epsilon}^2} + 3ML^2\right)}, \frac{\overline{c_4}\left(\frac{2}{\overline{\alpha_{\boldsymbol{\omega}}}^2} + 3ML^2\right)}{\underline{c_4}\left(\frac{2}{\overline{\alpha_{\boldsymbol{\omega}}}^2} + 3ML^2\right)}\right\}$, where $\overline{\alpha_{\boldsymbol{\omega}}}, \overline{\alpha_\epsilon}$ are the step-size of $\boldsymbol{\omega}_i$ and ϵ_i in the first iteration, respectively. According to Eq. (88) and the definition of c_4^t, we have,

$$
\begin{aligned}
c_4^t \|\nabla \bar{F}^t\|^2 \leq & \, b_6^t \sum_{i=1}^{M} \left(\left\|\boldsymbol{\omega}_i^{t+1} - \boldsymbol{\omega}_i^t\right\|^2 + \left\|\epsilon_i^{t+1} - \epsilon_i^t\right\|^2\right) + b_6^t \left\|\boldsymbol{z}^{t+1} - \boldsymbol{z}^t\right\|^2 \\
& + c_4^t \left(\frac{2}{{\alpha_{\boldsymbol{\omega}}}^2} + 3ML^2\right) \sum_{i=1}^{M} \left\|\boldsymbol{\omega}_i^{\tilde{t}} - \boldsymbol{\omega}_i^t\right\|^2 - c_4^t k_{\varkappa}\varkappa \left(\frac{2}{{\alpha_{\boldsymbol{\omega}}}^2} + 3ML^2\right) \sum_{i=1}^{M} \left\|\boldsymbol{\omega}_i^{t+1} - \boldsymbol{\omega}_i^t\right\|^2 \\
& + c_4^t \left(\frac{2}{{\alpha_{\epsilon}}^2} + 3ML^2\right) \sum_{i=1}^{M} \left\|\epsilon_i^{\tilde{t}} - \epsilon_i^t\right\|^2 - c_4^t k_{\varkappa}\varkappa \left(\frac{2}{{\alpha_{\epsilon}}^2} + 3ML^2\right) \sum_{i=1}^{M} \left\|\epsilon_i^{t+1} - \epsilon_i^t\right\|^2 \\
& + \left(\frac{1}{10\alpha_\lambda} - \frac{6\varkappa k_0 ML^2}{2}\right) \sum_{i=1}^{M} \left\|\lambda_i^{t+1} - \lambda_i^t\right\|^2 + \sum_{i=1}^{M} 3c_4^t \left(\left(a_1^{t-1}\right)^2 - \left(a_1^t\right)^2\right) \left\|\lambda_i^t\right\|^2 \\
& + \frac{1}{10\varkappa\alpha_\phi} \sum_{i=1}^{M} \left\|\phi_i^{\tilde{t}} - \phi_i^t\right\|^2 + \sum_{i=1}^{M} 3c_4^t \left(\left(a_2^{\hat{t}-1}\right)^2 - \left(a_2^{\tilde{t}-1}\right)^2\right) \left\|\phi_i^t\right\|^2 .
\end{aligned} \tag{90}
$$

Combining Assumption 2, the definition of c_4^t with Eq. (78), we have,

$$
\begin{aligned}
c_4^t \left\|\nabla \bar{F}^t\right\|^2 \leq & \, G^t - G^{t+1} + \frac{a_1^{t-1} - a_1^t}{2} M\mu_3 + \frac{a_2^{t-1} - a_2^t}{2} M\mu_4 + \frac{4}{\alpha_\lambda}\left(\frac{a_1^{t-2}}{a_1^{t-1}} - \frac{a_1^{t-1}}{a_1^t}\right) M\mu_3 \\
& + \frac{4M\mu_4}{\alpha_\phi}\left(\frac{a_2^{t-2}}{a_2^{t-1}} - \frac{a_2^{t-1}}{a_2^t}\right) + 3\overline{c_4} M\mu_3 \left(\left(a_1^{t-1}\right)^2 - \left(a_1^t\right)^2\right) + 3\overline{c_4}\mu_4 \sum_{i=1}^{M} \left(\left(a_2^{\hat{t}-1}\right)^2 - \left(a_2^{\tilde{t}-1}\right)^2\right) \\
& + \overline{c_4}\left(\frac{2}{\alpha_{\boldsymbol{\omega}^2}} + 3ML^2\right) \sum_{i=1}^{M} \left\|\boldsymbol{\omega}_i^{\tilde{t}_i} - \boldsymbol{\omega}_i^t\right\|^2 - \underline{c_4} k_{\varkappa}\varkappa \left(\frac{2}{\alpha_{\boldsymbol{\omega}^2}} + 3ML^2\right) \sum_{i=1}^{M} \left\|\boldsymbol{\omega}_i^{t+1} - \boldsymbol{\omega}_i^t\right\|^2 \\
& + \overline{c_4}\left(\frac{2}{{\alpha_\epsilon}^2} + 3ML^2\right) \sum_{i=1}^{M} \left\|\epsilon_i^{\tilde{t}} - \epsilon_i^t\right\|^2 - \underline{c_4} k_{\varkappa}\varkappa \left(\frac{2}{{\alpha_\epsilon}^2} + 3ML^2\right) \sum_{i=1}^{M} \left\|\epsilon_i^{t+1} - \epsilon_i^t\right\|^2 \\
& + \frac{1}{10\varkappa\alpha_\phi} \sum_{i=1}^{M} \left\|\phi_i^{\tilde{t}} - \phi_i^t\right\|^2 - \frac{1}{10\alpha_\phi} \sum_{i=1}^{M} \left\|\phi_i^{t+1} - \phi_i^t\right\|^2 .
\end{aligned} \tag{91}
$$

We assume $\widehat{T}(\Upsilon)=\min\left\{t \mid \left\|\nabla\bar{F}^t\right\|^2 \le \frac{\Upsilon}{4}, t\ge 2\right\}$. Summing up Eq. (91) from $t=2$ to $t=\widehat{T}(\Upsilon)$, we have,

$$
\begin{aligned}
\sum_{t=2}^{\widehat{T}(\Upsilon)} c_4^t\|\nabla\bar{F}^t\|^2 \le\, & G^2-\underline{\mathcal{L}}+\frac{4M\mu_3}{\alpha_\lambda}\Big(\frac{a_1^0}{a_1^1}+\frac{a_1^1}{a_1^2}\Big)+\frac{a_1^1M\mu_3}{2}+\frac{7M\sigma_3^2}{2\alpha_\lambda}+3\overline{c_4}M\mu_3\left(a_1^1\right)^2 \\
& +\frac{4M\mu_4}{\alpha_\phi}\Big(\frac{a_1^0}{a_1^1}+\frac{a_1^1}{a_1^2}\Big)+\frac{a_2^1M\mu_4}{2}+\frac{7M\sigma_4^2}{2\alpha_\phi}+\sum_{i=1}^{M}\sum_{t=2}^{\widehat{T}(\Upsilon)}3\overline{c_4}p_4\Big(\big(c_2^{\hat{t}-1}\big)^2-\big(c_2^{\tilde{t}-1}\big)^2\Big) \\
& +\frac{a_1^2M{\sigma_3}^2}{2}+\frac{a_2^2M{\sigma_4}^2}{2}+\frac{1}{10\varkappa\alpha_\phi}\sum_{t=2}^{\widehat{T}(\Upsilon)}\sum_{i=1}^{M}\left\|\phi_i^{\tilde{t}}-\phi_i^t\right\|^2-\frac{1}{10\alpha_\phi}\sum_{t=2}^{\widehat{T}(\Upsilon)}\sum_{i=1}^{M}\left\|\phi_i^{t+1}-\phi_i^t\right\|^2 \\
& +\overline{c_4}\Big(\frac{2}{{\alpha_{\boldsymbol{\omega}}}^2}+3ML^2\Big)\sum_{t=2}^{\widehat{T}(\Upsilon)}\sum_{i=1}^{M}\left\|\boldsymbol{\omega}_i^{\tilde{t}}-\boldsymbol{\omega}_i^t\right\|^2-\underline{c_4}k_\varkappa\varkappa\Big(\frac{2}{\overline{\alpha_{\boldsymbol{\omega}}}^2}+3ML^2\Big)\sum_{t=2}^{\widehat{T}(\Upsilon)}\sum_{i=1}^{M}\left\|\boldsymbol{\omega}_i^{t+1}-\boldsymbol{\omega}_i^t\right\|^2 \\
& +\overline{c_4}\Big(\frac{2}{{\alpha_\epsilon}^2}+3ML^2\Big)\sum_{t=2}^{\widehat{T}(\Upsilon)}\sum_{i=1}^{M}\left\|\epsilon_i^{\tilde{t}}-\epsilon_i^t\right\|^2-\underline{c_4}k_\varkappa\varkappa\Big(\frac{2}{\overline{\alpha_\epsilon}^2}+3ML^2\Big)\sum_{t=2}^{\widehat{T}(\Upsilon)}\sum_{i=1}^{M}\left\|\epsilon_i^{t+1}-\epsilon_i^t\right\|^2,
\end{aligned}
\tag{92}
$$

where $\sigma_3=\max\{||\lambda_1-\lambda_2\|\}, \sigma_4=\max\{\|\phi_1-\phi_2\|\}$, and $\underline{\mathcal{L}}=\min\mathcal{L}_\lambda\left(\{\boldsymbol{\omega}_i^t\},\{\epsilon_i^t\},\boldsymbol{z}^t,\{\lambda_i^t\},\{\phi_i^t\}\right)$. $\forall t\ge 2$, we have,

$$
G^t \ge \underline{\mathcal{L}}-\frac{4a_1^1M\mu_3}{\alpha_\lambda a_1^2}-\frac{4a_2^1M\mu_4}{\alpha_\phi a_2^2}-\frac{M{\sigma_3}^2}{2}\big(\frac{7}{\alpha_\lambda}+a_1^2\big)-\frac{M{\sigma_4}^2}{2}\big(\frac{7}{\alpha_\phi}+a_2^2\big). \tag{93}
$$

According to Definition 1, we assume that the number of iterations between the last active iteration and the next iteration does not exceed $\varkappa$, i.e., $\tilde{t}-\hat{t}\le\varkappa$. Also, we denote $\mathcal{V}_i(T)$ as the iteration index set when the i-th client is active during the $T+\varkappa$ iterations, where the l-th element of $\mathcal{V}_i(T)$ can be denoted as $\hat{v}_i(l)$. Then,

$$
\begin{aligned}
\sum_{t=2}^{\widehat{T}(\Upsilon)}3\overline{c_4}\mu_4\Big(\big(a_2^{\hat{t}-1}\big)^2-\big(a_2^{\tilde{t}-1}\big)^2\Big) &\le \varkappa\sum_{\hat{v}_i(l)\in\mathcal{V}_i(\widehat{T}(\Upsilon)),2\le\hat{v}_i(l)\le\widehat{T}(\Upsilon)}3\overline{c_4}\mu_4\Big(\big(a_2^{\hat{v}_i(l)-1}\big)^2-\big(a_2^{\hat{v}_i(l+1)-1}\big)^2\Big) \\
&\le 3\varkappa\overline{c_4}\mu_4\left(a_2^1\right)^2.
\end{aligned}
\tag{94}
$$

When $\hat{v}_i(l-1)\le t<\hat{v}_i(l)$, we have $\phi_i^t=\phi_i^{\hat{v}_i(l)-1}$ as inactivate clients do not update their variables in each master iteration. Besides, for any t that satisfies $t\notin\mathcal{V}_i(T)$, we have $\left\|\phi_i^t-\phi_i^{t-1}\right\|^2=0$. Based on the setting of $\hat{v}_i(l)-\hat{v}_i(l-1)\le\varkappa$, we have,

$$
\begin{aligned}
\sum_{t=2}^{\widehat{T}(\Upsilon)}\sum_{i=1}^{M}\left\|\phi_i^{\tilde{t}}-\phi_i^t\right\|^2 &\le \varkappa\sum_{\substack{\hat{v}_i(l)\in\mathcal{V}_i(\widehat{T}(\Upsilon))\\ 3\le\hat{v}_i(l)}}\sum_{i=1}^{M}\left\|\phi_i^{\hat{v}_i(l)}-\phi_i^{\hat{v}_i(l)-1}\right\|^2 \\
&= \varkappa\sum_{t=2}^{\widehat{T}(\Upsilon)}\sum_{i=1}^{M}\left\|\phi_i^{t+1}-\phi_i^t\right\|^2+\varkappa\sum_{t=\widehat{T}(\Upsilon)+1}^{\widehat{T}(\Upsilon)+\varkappa-1}\sum_{i=1}^{M}\left\|\phi_i^{t+1}-\phi_i^t\right\|^2 \\
&\le \varkappa\sum_{t=2}^{\widehat{T}(\Upsilon)}\sum_{i=1}^{M}\left\|\phi_i^{t+1}-\phi_i^t\right\|^2+4\varkappa(\varkappa-1)M\mu_4.
\end{aligned}
\tag{95}
$$

Likewise, we have $\boldsymbol{\omega}_i^t=\boldsymbol{\omega}_i^{\hat{v}_i(l)-1}, \epsilon_i^t=\epsilon_i^{\hat{v}_i(l)-1}$ when $\hat{v}_i(l-1)\le t<\hat{v}_i(l)$. And for $t\notin\mathcal{V}_i(T)$, we have $\left\|\boldsymbol{\omega}_i^t-\boldsymbol{\omega}_i^{t-1}\right\|^2=0, \left\|\epsilon_i^t-\epsilon_i^{t-1}\right\|^2=0$. Based on the setting of $\hat{v}_i(l)-\hat{v}_i(l-1)\le\varkappa$, we have,

$$
\begin{aligned}
\sum_{t=2}^{\widehat{T}}\sum_{i=1}^{M}\left\|\boldsymbol{\omega}_i^{\tilde{t}}-\boldsymbol{\omega}_i^t\right\|^2 &\le \varkappa\sum_{\hat{v}_i(l)\in\mathcal{V}_i(\widehat{T}(\Upsilon)),3\le\hat{v}_i(l)}\sum_{i=1}^{M}\left\|\boldsymbol{\omega}_i^{\hat{v}_i(l)}-\boldsymbol{\omega}_i^{\hat{v}_i(l)-1}\right\|^2 \\
&= \varkappa\sum_{t=2}^{\widehat{T}(\Upsilon)}\sum_{i=1}^{M}\left\|\boldsymbol{\omega}_i^{t+1}-\boldsymbol{\omega}_i^t\right\|^2+\varkappa\sum_{t=\widehat{T}(\Upsilon)+1}^{\widehat{T}(\Upsilon)+\varkappa-1}\sum_{i=1}^{M}\left\|\boldsymbol{\omega}_i^{t+1}-\boldsymbol{\omega}_i^t\right\|^2 \\
&\le \varkappa\sum_{t=2}^{\widehat{T}(\Upsilon)}\sum_{i=1}^{M}\left\|\boldsymbol{\omega}_i^{t+1}-\boldsymbol{\omega}_i^t\right\|^2+4\varkappa(\varkappa-1)M\mu_1,
\end{aligned}
\tag{96}
$$

$$
\begin{aligned}
\sum_{t=2}^{\widehat{T}(\epsilon)}\sum_{i=1}^{M}\left\|\epsilon_i^{\tilde{t}}-\epsilon_i^t\right\|^2 &\le \varkappa \sum_{\hat{v}_i(l)\in\mathcal{V}_i(\widehat{T}(\Upsilon)),3\le\hat{v}_i(l)}\sum_{i=1}^{M}\left\|\epsilon_i^{\hat{v}_i(l)}-\epsilon_i^{\hat{v}_i(l)-1}\right\|^2 \\
&= \varkappa\sum_{t=2}^{\widehat{T}(\Upsilon)}\sum_{i=1}^{M}\left\|\epsilon_i^{t+1}-\epsilon_i^t\right\|^2+\varkappa\sum_{t=\widehat{T}(\Upsilon)+1}^{\widehat{T}(\Upsilon)+\varkappa-1}\sum_{i=1}^{M}\left\|\epsilon_i^{t+1}-\epsilon_i^t\right\|^2 \\
&\le \varkappa\sum_{t=2}^{\widehat{T}(\Upsilon)}\sum_{i=1}^{M}\left\|\epsilon_i^{t+1}-\epsilon_i^t\right\|^2+4\varkappa(\varkappa-1)M\mu_2.
\end{aligned}
\tag{97}
$$

It follows from Eq. (92), (94), (95), (96) that,

$$
\begin{aligned}
\sum_{t=2}^{\widehat{T}(\Upsilon)} c_4^t\|\nabla\bar{F}^t\|^2 \le\; & G^2-\underline{\mathcal{L}}+\frac{4}{\alpha_\lambda}\Big(\frac{a_1^0}{a_1^1}+\frac{a_1^1}{a_1^2}\Big)M\mu_3+\frac{a_1^1}{2}M\mu_3+\frac{7}{2\alpha_\lambda}M{\sigma_3}^2+3\overline{c_4}\left(a_1^1\right)^2M\mu_3 \\
&+\frac{4}{\alpha_\phi}\Big(\frac{a_1^0}{a_1^1}+\frac{a_1^1}{a_1^2}\Big)M\mu_4+\frac{a_2^1}{2}M\mu_4+\frac{7}{2\alpha_\phi}M{\sigma_4}^2+3\varkappa\overline{c_4}\left(a_2^1\right)^2M\mu_4+\frac{a_1^2}{2}M\sigma_3^2+\frac{a_2^2}{2}M{\sigma_4}^2 \\
&+\Big(\frac{2M\mu_4}{5\alpha_\phi}+4\overline{c_4}\big(\frac{2}{{\alpha_{\boldsymbol{\omega}}}^2}+3ML^2\big)M\mu_1\varkappa+4\overline{c_4}\big(\frac{2}{{\alpha_\epsilon}^2}+3ML^2\big)M\mu_2\varkappa\Big)(\varkappa-1) \\
=\;& \bar{c}+k_c\varkappa(\varkappa-1),
\end{aligned}
\tag{98}
$$

where $\bar{c}$ and k_c are constants. Besides, we assume c_5 is a constant,

$$
\begin{aligned}
c_5 &= \max\left\{d_1,d_2,d_3,\frac{\frac{30}{\alpha_\lambda}+150\alpha_\lambda\varkappa k_0ML^2}{(1-30\alpha_\lambda\varkappa k_0ML^2)\,\underline{b_6}},\frac{30\varkappa}{\alpha_\phi\underline{b_6}}\right\} \\
&\ge \max\left\{d_1,d_2,d_3,\frac{\frac{30}{\alpha_\lambda}+150\alpha_\lambda\varkappa k_0ML^2}{(1-30\alpha_\lambda\varkappa k_0ML^2)\,b_6^t},\frac{30\varkappa}{\alpha_\phi b_6^t}\right\}=\frac{1}{c_4^tb_6^t}.
\end{aligned}
\tag{99}
$$

Then, we can get that,

$$
\sum_{t=2}^{\widehat{T}(\Upsilon)}\frac{1}{c_5b_5^t}\left\|\nabla F^{\widehat{T}(\Upsilon)}\right\|^2\le\sum_{t=2}^{\widehat{T}(\Upsilon)}\frac{1}{c_5b_6^t}\left\|\nabla F^t\right\|^2\le\sum_{t=2}^{\widehat{T}(\Upsilon)}c_4^t\left\|\nabla F^t\right\|^2\le\bar{c}+k_c\varkappa(\varkappa-1).
\tag{100}
$$

And it follows from Eq. (100) that,

$$
\left\|\nabla\bar{F}^{\widehat{T}(\Upsilon)}\right\|^2\le\frac{(\bar{c}+k_c\varkappa(\varkappa-1))\,c_5}{\sum_{t=2}^{\widehat{T}(\Upsilon)}\frac{1}{b_5^t}}.
\tag{101}
$$

Based on the setting of a_1^t,a_2^t, we can obtain,

$$
\frac{1}{b_6^t}\ge\frac{1}{4L^2M(\beta-2)\left(\alpha_\lambda+\alpha_\phi\right)(t+1)^{\frac{1}{2}}+\frac{\alpha_\phi L^2(M-S)}{2}}.
\tag{102}
$$

Summing up $\frac{1}{b_6^t}$ from $t=2$ to $t=\widehat{T}(\Upsilon)$, we have,

$$
\begin{aligned}
\sum_{t=2}^{\widehat{T}(\Upsilon)}\frac{1}{b_6^t}&\ge\sum_{t=2}^{\widehat{T}(\Upsilon)}\frac{1}{4L^2M(\beta-2)\left(\alpha_\lambda+\alpha_\phi\right)(t+1)^{\frac{1}{2}}+\frac{\alpha_\phi L^2(M-S)}{2}} \\
&\ge\sum_{t=2}^{\widehat{T}(\Upsilon)}\frac{1}{4L^2M(\beta-2)\left(\alpha_\lambda+\alpha_\phi\right)(t+1)^{\frac{1}{2}}+\frac{\alpha_\phi L^2(M-S)}{2}(t+1)^{\frac{1}{2}}} \\
&=\frac{\sum_{t=2}^{\widehat{T}(\Upsilon)}\frac{1}{(t+1)^{\frac{1}{2}}}}{4L^2M(\beta-2)\left(\alpha_\lambda+\alpha_\phi\right)+\frac{\alpha_\phi L^2(M-S)}{2}}\ge\frac{\left(\widehat{T}(\Upsilon)\right)^{\frac{1}{2}}-\sqrt{2}}{4L^2M(\beta-2)\left(\alpha_\lambda+\alpha_\phi\right)+\frac{\alpha_\phi L^2(M-S)}{2}}.
\end{aligned}
\tag{103}
$$

Combining Eq. (103) with Eq. (101), we have,

$$
\left\|\nabla\bar{F}^{\widehat{T}(\Upsilon)}\right\|^2\le\frac{(\bar{c}+k_c\varkappa(\varkappa-1))\,c_5}{\sum_{t=2}^{\widehat{T}(\Upsilon)}\frac{1}{b_6^t}}\le\frac{\left(4L^2M(\beta-2)\left(\alpha_\lambda+\alpha_\phi\right)+\frac{\alpha_\phi L^2(M-S)}{2}\right)(\bar{c}+k_c\varkappa(\varkappa-1))\,c_5}{\left(\widehat{T}(\Upsilon)\right)^{\frac{1}{2}}-\sqrt{2}}.
\tag{104}
$$

We assume c_6 is a constant that satisfies $c_6 = 4L^2M(\beta-2)\left(\alpha_\lambda+\alpha_\phi\right)$. According to the definition of $\widehat{T}(\Upsilon)$, we have,

$$\widehat{T}(\Upsilon) \geq \Big(\frac{\left(4c_6+2\alpha_\phi L^2(M-S)\right)\left(\bar{c}+k_c\varkappa(\varkappa-1)\right)c_5}{\Upsilon}+\sqrt{2}\Big)^2. \tag{105}$$

Combining the definition of ∇F^t and $\nabla \bar{F}^t$ with trigonometric inequality, we have,

$$\left\|\nabla F^t\right\| - \left\|\nabla \bar{F}^t\right\| \leq \left\|\nabla F^t - \nabla \bar{F}^t\right\| \leq \sqrt{\sum_{i=1}^{M}(\left\|a_1^{t-1}\lambda_i^t\right\|^2+\left\|a_2^{t-1}\phi_i^t\right\|^2)}. \tag{106}$$

When $t \geq 16M^2\big(\frac{\mu_3}{{\alpha_\lambda}^2}+\frac{\mu_4}{{\eta_\phi}^2}\big)^2\frac{1}{\epsilon^2}$, we can obtain $\sqrt{\sum_{i=1}^{M}(\left\|a_1^{t-1}\lambda_i^t\right\|^2+\left\|a_2^{t-1}\phi_i^t\right\|^2)} \leq \frac{\sqrt{\epsilon}}{2}$. Combining it with Eq. (105), we can conclude that there exists a

$$T(\Upsilon) \sim \mathcal{O}\Big(\max\Big\{\frac{16M^2}{\Upsilon^2}\big(\frac{\mu_3}{{\alpha_\lambda}^2}+\frac{\mu_4}{{\alpha_\phi}^2}\big)^2, \big(\frac{\left(4c_6+2\alpha_\phi L^2(M-S)\right)\left(\bar{c}+k_c\varkappa(\varkappa-1)\right)c_5}{\Upsilon}+\sqrt{2}\big)^2\Big\}\Big), \tag{107}$$

such that $\left\|\nabla F^t\right\|^2 \leq \Upsilon$, which concludes the proof of Theorem 1.

REFERENCES

[1] Z. Xu, H. Zhang, Y. Xu, and G. Lan, "A unified single-loop alternating gradient projection algorithm for nonconvex–concave and convex–nonconcave minimax problems," *Math. Program.*, pp. 1–72, 2023.
[2] K. Ji, J. Yang, and Y. Liang, "Bilevel optimization: Convergence analysis and enhanced design," in *Proc. Int. Conf. Mach. Learn.*, 2021, pp. 4882–4892.
[3] Y. Jiao, K. Yang, and D. Song, "Distributed distributionally robust optimization with non-convex objectives," in *Proc. Adv. Neural Inf. Proces. Syst.*, vol. 35, 2022, pp. 7987–7999.
[4] Y. Jiao, K. Yang, T. Wu, D. Song, and C. Jian, "Asynchronous distributed bilevel optimization," in *Proc. Int. Conf. Learn. Represent.*, 2023.