Triadic-OCD: Asynchronous Online Change Detection with Provable Robustness, Optimality, and Convergence

Yancheng Huang [1] **Kai Yang** [1] **Zelin Zhu** [1] **Leian Chen** [2 3]

Abstract

The primary goal of online change detection (OCD) is to promptly identify changes in the data stream. OCD problem find a wide variety of applications in diverse areas, e.g., security detection in smart grids and intrusion detection in communication networks. Prior research usually assumes precise knowledge of the system parameters. Nevertheless, this presumption often proves unattainable in practical scenarios due to factors such as estimation errors, system updates, etc. This paper aims to take the first attempt to develop a triadic-OCD framework with certifiable robustness, provable optimality, and guaranteed convergence. In addition, the proposed triadic-OCD algorithm can be realized in a fully asynchronous distributed manner, easing the necessity of transmitting the data to a single server. This asynchronous mechanism could also mitigate the straggler issue that faced by traditional synchronous algorithm. Moreover, the non-asymptotic convergence property of Triadic-OCD is theoretically analyzed, and its iteration complexity to achieve an ϵ-optimal point is derived. Extensive experiments have been conducted to elucidate the effectiveness of the proposed method.

1. Introduction

Detecting distribution changes as quickly as possible while controlling the false alarm rate is the fundamental objective of OCD. And this problem is frequently encountered in a diverse range of fields such as econometrics, climate modeling, and system security (Andersson et al., 2006; Barnett et al., 2001; Tartakovsky et al., 2006; Yang et al., 2016; Dou et al., 2019; Huang et al., 2011; Li et al., 2014; Huang et al., 2014; Kurt et al., 2018b). In contrast to general detection problems, OCD problem in dynamic systems are particularly challenging because changes in the distribution of observed data can be caused by the dynamics of the system itself, even without external changes. Although extensive research has been conducted on this problem, existing studies still face critical issues. The primary challenge is that the parameters associated with systems are assumed to be perfectly estimated. In practice, however, this is often not the case due to factors such as estimation errors, system updates, etc. For instance, the investigated OCD problem within dynamic systems subsumes the false data injection attacks (FDIA) detection problem in smart grids and the blockage detection in MIMO systems —both critical issues that have received extensive attention (Huang et al., 2011; Nishimori et al., 2011). In smart grids, fluctuations in the environment can lead to changes in line admittances, causing inaccuracies in system matrix (Li et al., 2014). Similarly, in MIMO systems, factors such as estimation error, aging, and quantization often prevent the perfect estimation of the channel matrix (Weber et al., 2006). Furthermore, current research focuses on centralized setting, which may incur privacy breaches (Subramanya & Riggio, 2021) and high communication costs. Additionally, the synchronous distributed approach often encounters straggler issues (Jiao et al., 2022) which can cause significant delays in the detection process. Therefore, in this paper, we aim to address these issues, and our contributions are summarized as follows:

- **Certified Robustness:** As opposed to existing works in the literature, Triadic-OCD provides a high degree of confidence across a broad spectrum of system parameter uncertainties. This certifiable robustness assures the reliability of change point detection in real-world applications.
- **Asynchronous Updating:** In order to address concerns related to potential privacy breaches, elevated communication costs, and straggler issues, we propose a novel *asynchronous* distributed algorithm to effectively address the detection problem in the presence of parameter uncertainties. In addition, the proposed approach goes beyond empirical performance by offer-

[1]Tongji University [2]Columbia University [3]This work was done prior to Leian Chen joining Amazon. Correspondence to: Kai Yang <kaiyang@tongji.edu.cn>.

Proceedings of the 41st International Conference on Machine Learning, Vienna, Austria. PMLR 235, 2024. Copyright 2024 by the author(s).

ing theoretical proofs that establish the optimality of Triadic-OCD under certain conditions. This represents a theoretic advancement in ensuring the algorithm's efficiency and effectiveness.

- **Non-asympotic Convergence Analysis:** We not only proves that Triadic-OCD is guaranteed to converge, but also undertake non-asymptotic convergence analysis to establish an upper bound to the iteration complexity of the proposed asynchronous algorithm to attain an ϵ-optimal solution.

2. Related Work

There have been numerous studies that address the OCD problem with uncertain pre- and post-change distributions. (Hare et al., 2021) proposes the Uncertain Likelihood Ratio (ULR) test statistic to tackle the parameters within pre- and post-change distributions, which are completely unknown, or known with limited prior knowledge. On the basis of that, (Hare & Kaplan, 2022) develops a more efficient method called the Windowed Uncertain Likelihood Ratio (W-ULR) test. (Unnikrishnan et al., 2011) consider the OCD problem when the pre-change and post-change distributions belong to known uncertainty sets. It provides a condition under which the detection rule based on the least favorable distributions (LFDs) are minimax robust. (Molloy & Ford, 2017) relaxes this condition and provides the new performance guarantee of misspecified CUSUM rules. (Xie, 2022) proposes a non-parametric method based on Wasserstein uncertainty sets. However, there's a significant distinction between these works and our paper. The general OCD methods considering uncertain distributions assume that the distribution before and after the change remains constant over time, which is completely inconsistent with our problem due to the system's inherent dynamics. As a result, general OCD methods are not applicable to our problem.

The OCD problem in dynamic system has attracted increasing attention recently. Both (Huang et al., 2011) and (Li et al., 2014) proposed CUSUM-type algorithms with prior assumptions on the state variables of the systems. And the method proposed in (Huang et al., 2011) is inefficient for large or negative attacking vectors injected into the system. In contrast, our approach makes no assumptions for the state variables of the system and imposes no restrictions on the sign of the attacking vectors. (Huang et al., 2014) proposes a real-time detection method based on residuals and constructs the decision statistic with the Rao test statistic. However, in some cases, the decision statistic cannot be evaluated due to the covariance singularity of the residuals. All the methods mentioned above, as well as (Zhang, 2019; Zhang & Wang, 2021), assume that the system parameters can be perfectly determined, which is often impractical in reality. In addition, some recent work has studied the OCD in certain dynamic systems in the distributed setting. A distributed algorithm based on the Kalman filter is proposed in (Kurt et al., 2018a). And (Li et al., 2014) propose distributed sequential detectors based on the generalized likelihood ratio. However, these methods are synchronously distributed and may suffer from straggler problems, which could incur significant delays during the detection. As opposed to all previous methods, this paper proposes an asynchronous distributed algorithm to solve the OCD problem in the presence of parameter uncertainties.

3. Problem Statement

In this section, we provide a detailed explanation of the OCD problem with parameter uncertainties. The system of interest can be expressed as follow,

$$\mathbf{y}^{(t)} = \bar{f}(\boldsymbol{\theta}^{(t)} \mid \mathbf{H}) + \mathbf{n}^{(t)}, \mathbf{H} \in \mathcal{U} \tag{1}$$

where $\mathbf{y}^{(t)} \in \mathbb{R}^M$ represents the known observation vector and $\boldsymbol{\theta}^{(t)} \in \mathbb{R}^N$ represents the unknown time-varying system states. $\bar{f}$ models the relationship within the system and the matrix $\mathbf{H} \in \mathbb{R}^{M\times N}$ incorporates the system parameters. In contrast to previous work, the system matrix is assumed to belong to an uncertainty set $\mathcal{U}$ instead of being perfectly known, i.e. $\mathbf{H} \in \mathcal{U}$. When $t \geq t_a$, the time-varying attack vector $\mathbf{a}^{(t)}$ is injected into the system which alters the distribution of $\mathbf{y}^{(t)}$. Our goal is to detect the injected vector as soon as possible. Note that the complexity of this problem is formidable, and its practical applications remain somewhat under-explored. However, in numerous real-world scenarios, including the FDIA detection in smart grids and the blockage detection in MIMO systems, the function $\bar{f}$ degenerates into a linear form. In such case, prevailing methodologies typically rely on precise knowledge of the system parameters, which often proves unattainable in practical settings. To address this limitation, we propose an asynchronous OCD method capable of robust and high-performance detection.

Given that the system states $\boldsymbol{\theta}^{(t)}$ and the injected anomaly vector $\mathbf{a}^{(t)}$ are unknown, we estimate them using their maximum likelihood estimates (MLE) (Tartakovsky et al., 2014). This leads to the generalized CUSUM detector for our problem, which can be written as follows.

$$T_G = \min\{J : \max_{1\leq j\leq J} \Lambda_j^{(J)} \geq \zeta\}. \tag{2}$$

where ζ is the predefined threshold. $\Lambda_j^{(J)}$ is given in (3), where $\mathbf{x}^{(t)}$ represent the component of $\mathbf{a}^{(t)}$ orthogonal to the column space of $\mathbf{H}$. And ρ_H is the upper bound for the absolute value of each component of $\mathbf{x}^{(t)}$. Based on (2), let V_J represents $\max_{1\leq j\leq J} \Lambda_j^{(J)}$, we can obtain that,

$$\Lambda_j^{(J)} \triangleq \sup_{\mathbf{H}\in\mathcal{U}} \ln \frac{\sup\limits_{\boldsymbol{\theta}^{(t)},\mathbf{a}^{(t)}:-\rho_U\mathbf{1}\le\mathbf{x}^{(t)}\le\rho_U\mathbf{1},\mathbf{H}\mathbf{x}^{(t)}=\mathbf{0}} \prod_{t=1}^{j-1} f_p\left(\mathbf{y}^{(t)} \mid \boldsymbol{\theta}^{(t)},\mathbf{H}\right)\prod_{t=j}^{J} f_q\left(\mathbf{y}^{(t)} \mid \boldsymbol{\theta}^{(t)},\mathbf{a}^{(t)},\mathbf{H}\right)}{\sup\limits_{\boldsymbol{\theta}^{(t)}} \prod_{t=1}^{J} f_p\left(\mathbf{y}^{(t)} \mid \boldsymbol{\theta}^{(t)},\mathbf{H}\right)}. \tag{3}$$

$$\begin{aligned} V_J &\triangleq \max_{1\le j\le J} \Lambda_j^{(J)} = \max_{1\le j\le J} \sum\nolimits_{t=j}^{J} \frac{v_t}{2\sigma_n^2} \\ &= \max\{V_{J-1},0\} + \frac{v_t}{2\sigma_n^2}, \text{ with } V_0 = 0, \end{aligned} \tag{4}$$

where the value of v_t can be obtained by solving problem (5). The detailed derivation can be found in Appendix B. It can be seen from (2) and (4) that the change is declared when V_J surpasses ζ. And V_J can be calculated in a recursive way, with the primary challenge being to obtain the value of v_t.

$$\begin{aligned} v_t \triangleq \sup_{\mathbf{H}\in\mathcal{U}} \sup_{-\rho_U\mathbf{1}\le\mathbf{x}^{(t)}\le\rho_U\mathbf{1},\mathbf{H}^T\mathbf{x}^{(t)}=\mathbf{0}} \Bigg\{ \\ \left[2(\mathbf{x}^{(t)})^T\mathbf{y}^{(t)} - \|\mathbf{x}^{(t)}\|_2^2\right]\Bigg\}. \end{aligned} \tag{5}$$

Note that the inherent uncertainties associated with $\mathbf{H}$ makes the constraint $\mathbf{H}^T\mathbf{x}^{(t)} = \mathbf{0}$ in (5) unattainable, which greatly exacerbates the complexity of the problem.

4. Nested Optimization

In this section, we elaborate on the uncertainty of the system matrix and reformulate the problem (5) under the distributed setup as a nested optimization problem.

To address the challenges posed by the inherent uncertainty of $\mathbf{H}$, we adopt the constraint-wise uncertainty model, which decouples the uncertainties between different rows in the matrix. This versatile approach is applicable to a wide range of practical problems (Yang et al., 2014; Liu et al., 2021b). We first represent $\mathbf{h}_i$ as the i-th column of $\mathbf{H}$, each $\mathbf{h}_i$ lies in the uncertainty set $\mathcal{U}_i$. Denote the i-th nominal column as $\bar{\mathbf{h}}_i$. Subsequently, we relax the constraint $\mathbf{H}^T\mathbf{x}^{(t)} = \mathbf{0}$ with $\mathbf{H}^T \in \mathcal{U}$ as follows,

$$\bar{\mathbf{h}}_i^T\mathbf{x}^{(t)} + p_i(\mathbf{x}^{(t)}) \le \delta_i,\ 1 \le i \le 2N, \tag{6}$$

where

$$\begin{aligned} p_i(\mathbf{x}^{(t)}) &= \max_{\mathbf{h}_i\in\mathcal{U}_i} (\mathbf{h}_i - \bar{\mathbf{h}}_i)^T\mathbf{x}^{(t)}, 1 \le i \le N \\ p_i(\mathbf{x}^{(t)}) &= \max_{-\mathbf{h}_i\in\mathcal{U}_i} (\mathbf{h}_i - \bar{\mathbf{h}}_i)^T\mathbf{x}^{(t)}, N+1 \le i \le 2N \end{aligned} \tag{7}$$

are the protection functions. For $i = N+1, \cdots, 2N$, $\bar{\mathbf{h}}_i = \bar{\mathbf{h}}_{i-N}$, $\mathcal{U}_i = \mathcal{U}_{i-N}$, and $\delta_i = \delta_{i-N}$.

The choice of different forms for $\mathcal{U}_i$ will yield distinct protection functions $p_i(\mathbf{x}^{(t)})$, consequently affecting the trade-off between robustness and detection performance. In this paper, we examine a highly versatile uncertain set. Specifically, the uncertainty set $\mathcal{U}_i$ corresponding to each $\mathbf{h}_i$ is assumed to be characterized by U_i differentiable functions, that is,

$$c_{iu}(\mathbf{h}_i) \le 0, 1 \le u \le U_i. \tag{8}$$

Our framework offers the flexibility to customize the uncertainty set chosen in practical applications based on specific requirements, addressing diverse needs related to complexity and performance.

Now we consider the model (1) in a distributed setting, where the system comprises numerous sub-regions (workers) geographically dispersed across a wide area. Each sub-region collects and manages local observation data before communicating with the master node to facilitate collaborative detection. Suppose there are L sub-regions in the system, we rewrite (1) for each sub-region as follows,

$$\mathbf{y}_l^{(t)} = \mathbf{H}_l\boldsymbol{\theta}_l^{(t)} + \mathbf{n}_l^{(t)}, \tag{9}$$

where we utilize the subscript l to denote the local components within the l-th sub-region. Given that the state vectors of neighboring sub-regions may share certain parameters, the variable $\boldsymbol{\theta}_l$ for different l may therefore partially overlap. For clarity, we denote the $\mathbf{x}^{(t)}$ within the l-th sub-region as $\boldsymbol{\mu}_l^{(t)}$. Therefore, we have $\boldsymbol{\mu}_l^{(t)} = \mathbf{B}_l\mathbf{x}^{(t)}, 1 \le l \le L$, where $\mathbf{B}_l$ is the matrix projecting global attacking vector to the local attacking vector on worker l. Below we omit the time superscript (t) for notational simplicity. On the basis of (6)-(9), the problem (5) can be formulated as

$$\begin{aligned} \min \quad & \sum\nolimits_{l=1}^{L} \{\ \|\boldsymbol{\mu}_l\|_2^2 - 2\mathbf{y}_l{}^T\boldsymbol{\mu}_l\ \} \\ \text{s.t.} \quad & -\rho_U\mathbf{1} \le \boldsymbol{\mu}_l \le \rho_U\mathbf{1}, \quad 1 \le l \le L \\ & \boldsymbol{\mu}_l = \mathbf{B}_i\mathbf{x}, \quad 1 \le l \le L \\ & \bar{\mathbf{h}}_i^T\mathbf{x} + p_i(\mathbf{x}) \le \delta_i,\ 1 \le i \le 2N \\ \text{var.} \quad & \{\boldsymbol{\mu}_l\}, \{\mathbf{h}_i\}, \mathbf{x}, \end{aligned} \tag{10}$$

5. Asynchronous Distributed Method

In this section, we provide a detailed explanation of the proposed algorithm named Triadic-OCD, capable of solving our problem in an asynchronous distributed manner while ensuring robustness, and optimality. Triadic-OCD employs a set of cutting planes to approximate the feasible region constrained by (6), leveraging their flexibility in adaptation to problem complexity and efficiency in exploration (Boyd & Vandenberghe, 2007; Franc et al., 2011; Yang et al., 2014).

Subsequently, variables undergo asynchronous distributed updates. These two steps iteratively alternate in the proposed algorithm. The cutting planes continuously refine the approximation, while variable updates on the basis of the newly updated cutting plane sets.

Given that computing the exact value of $\{\mathbf{h}_i\}$ for updating outer-level variables at each iteration is unnecessary and inefficient in terms of time and memory usage (Gould et al., 2016; Yang et al., 2021), we opt to employ the estimates of $\{\mathbf{h}_i\}$ instead of their precise values during algorithm execution. Based on existing methods (Ji et al., 2021; Liu et al., 2021a; Jiao et al., 2022), Triadic-OCD uses D iterations of gradient descent to approximate the optimal value of $\{\mathbf{h}_i\}$ for the inner-level optimization problem (7). We first present the augmented Lagrangian function for each protection function $p_i, i = 1, \cdots, N$ as follows. For $p_i, i = N+1, \cdots, 2N$, the results are similar.

$$L_\sigma(\mathbf{x}, \mathbf{h}_i, \{\phi_{iu}\}) = (\mathbf{h}_i - \bar{\mathbf{h}}_i)^T \mathbf{x} + \frac{\sigma}{2} \sum_{u=1}^{U_i} \left(\max\left\{ \frac{\phi_{iu}}{\sigma} + c_{iu}(\mathbf{h}_i), 0 \right\}^2 - \frac{\phi_{iu}^2}{\sigma^2} \right) \tag{11}$$

where $\phi_{iu} \in \mathbb{R}$ is the dual variable, and $\sigma > 0$ is the penalty parameter. In the $(d+1)^{th}$ iteration, the variables are updated as follows,

$$\begin{aligned} \mathbf{h}_{i,d+1} &= \mathbf{h}_{i,d} - \eta_{\mathbf{h}} \nabla_{\mathbf{h}_i} L_\sigma(\mathbf{x}, \mathbf{h}_{i,d}, \{\phi_{iu,d}\}), \\ \phi_{iu,d+1} &= (\phi_{iu,d} + \eta_\phi c_{iu}(\mathbf{h}_{i,d+1}))^+, \end{aligned} \tag{12}$$

where $(\cdot)^+ = \max\{0, \cdot\}$, and $\eta_{\mathbf{h}}$, η_ϕ are step-sizes. We use the results after D iterations of gradient descent to obtain the estimate of $\mathbf{h}_i$, i.e.,

$$\mathbf{h}_{i,D} = \mathbf{h}_{i,0} - \sum_{d=0}^{D-1} \eta_{\mathbf{h}} \nabla_{\mathbf{h}_i} L_\sigma(\mathbf{x}, \mathbf{h}_{i,d}, \{\phi_{iu,d}\}). \tag{13}$$

Based on the estimated solution, we define,

$$g_i(\mathbf{x}) = (\mathbf{h}_{i,0} - \sum_{d=0}^{D-1} \eta_{\mathbf{h}} \nabla_{\mathbf{h}_i} L_\sigma(\mathbf{x}, \mathbf{h}_{i,d}, \{\phi_{iu,d}\}))^T \mathbf{x}. \tag{14}$$

As such, the problem (10) can be written as

$$\begin{aligned} \min \quad & \sum_{l=1}^{L} \{ \|\boldsymbol{\mu}_l\|_2^2 - 2\mathbf{y}_l{}^T \boldsymbol{\mu}_l \} \\ \text{s.t.} \quad & -\rho_U \mathbf{1} \le \boldsymbol{\mu}_l \le \rho_U \mathbf{1}, \quad 1 \le l \le L \\ & \boldsymbol{\mu}_l = \mathbf{B}_i \mathbf{x}, \quad 1 \le l \le L \\ & g_i(\mathbf{x}) \le \delta_i, \quad 1 \le i \le 2N \\ \text{var.} \quad & \{\boldsymbol{\mu}_l\}, \mathbf{x}. \end{aligned} \tag{15}$$

Since the protection functions are convex (Yang et al., 2014), we employ a set of cutting planes to approximate the feasible region defined by the constraint $g_i(\mathbf{x}) \le \delta_i, 1 \le i \le 2N$. In the $(k+1)^{th}$ iteration, let $\mathcal{P}^k$ denote the feasible region respect to the set of cutting planes, i.e,

$$\mathcal{P}^k = \{\mathbf{b}_s^T \mathbf{x} + \kappa_s \le 0, s = 1, \cdots, |\mathcal{P}^k|\}, \tag{16}$$

where $|\mathcal{P}^k|$ denotes the number of cutting planes in the $(k+1)^{th}$ iteration. And $\mathbf{b}_s \in \mathbb{R}^M$ and $\kappa_s \in \mathbb{R}$ represents the parameters in the sth cutting plane. Therefore, in the $(k+1)^{th}$ iteration, the approximation problem can be formulated as follows :

$$\begin{aligned} \min \quad & \sum_{l=1}^{L} \{ \|\boldsymbol{\mu}_l\|_2^2 - 2\mathbf{y}_l{}^T \boldsymbol{\mu}_l \} \\ \text{s.t.} \quad & -\boldsymbol{\mu}_l - \rho_U \mathbf{1} + \mathbf{r}_l \circ \mathbf{r}_l = \mathbf{0}, \quad 1 \le l \le L \\ & \boldsymbol{\mu}_l - \rho_U \mathbf{1} + \mathbf{p}_l \circ \mathbf{p}_l = \mathbf{0}, \quad 1 \le l \le L \\ & \boldsymbol{\mu}_l = \mathbf{B}_i \mathbf{x}, \quad 1 \le l \le L \\ & \mathbf{b}_s^T \mathbf{x} + \kappa_s + q_s^2 = 0, \ s = 1, \cdots, |\mathcal{P}^k| \\ \text{var.} \quad & \{\boldsymbol{\mu}_l\}, \mathbf{x}, \{\mathbf{r}_l\}, \{\mathbf{p}_l\}, \{q_s\}, \end{aligned} \tag{17}$$

where $\mathbf{r}_l$,$\mathbf{p}_l$,q_s are introduced slack variables. $q_s \in \mathbb{R}$ and $\mathbf{r}_l$, $\mathbf{p}_l$ have the same dimensions as $\boldsymbol{\mu}_l$. Notation $\circ$ represents the Hadamard product. As the algorithm iterates, the set of cutting planes will be continuously updated to better approximate the original feasible region. In Triadic-OCD, we update the set of cutting planes every w iterations to improve the approximation of the original feasible region when $k < K_1$. K_1 and w are the pre-defined positive integers that can be adjusted. If $(k+1) \bmod w = 0$, we compute the value of $g_i(\mathbf{x}^{k+1})$ for each i according to (14). And subsequently, we check if $g_i(\mathbf{x}^{k+1}) \le \delta_i$. If $g_i(\mathbf{x}^{k+1}) > \delta_i$, a new cutting plane will be generated to separate the point from the feasible region defined by $g_i(\mathbf{x}) \le \delta_i$. Following (Boyd & Vandenberghe, 2007), the generated cutting plane cp^{k+1} can be written as,

$$g_i(\mathbf{x}^{k+1}) + \left(\frac{\partial g_i(\mathbf{x}^{k+1})}{\partial \mathbf{x}} \right)^T (\mathbf{x} - \mathbf{x}^{k+1}) \le \delta_i. \tag{18}$$

And the set of cutting planes will be updated as follows,

$$\mathcal{P}^{k+1} = \begin{cases} \mathcal{P}^k \cup \{cp^{k+1}\}, \text{ if } g_i(\mathbf{x}^{k+1}) > \delta_i \\ \mathcal{P}^k, \ \ otherwise \end{cases}. \tag{19}$$

After the new cutting plane is added, its corresponding dual variable set and slack variable set will be updated accordingly,

$$\{q^{k+1}\} = \begin{cases} \{q^k\} \cup q_{|\mathcal{P}^k|+1}^{k+1}, \text{ if } g_i(\mathbf{x}^{k+1}) > \delta_i \\ \{q^k\}, \ \ otherwise \end{cases}, \tag{20}$$

$$\{\gamma^{k+1}\} = \begin{cases} \{\gamma^k\} \cup \gamma_{|\mathcal{P}^k|+1}^{k+1}, \text{ if } g_i(\mathbf{x}^{k+1}) > \delta_i \\ \{\gamma^k\}, \ \ otherwise \end{cases}. \tag{21}$$

$$\boldsymbol{\mu}_l^{k+1} = \begin{cases} \boldsymbol{\mu}_l^k - \eta_{\boldsymbol{\mu}} \nabla_{\boldsymbol{\mu}_l} \widetilde{L}_p(\{\boldsymbol{\mu}_l^{\hat{k}_l}\}, \{\mathbf{r}_l^{\hat{k}_l}\}, \{\mathbf{p}_l^{\hat{k}_l}\}, \{q_s^{\hat{k}_l}\}, \mathbf{x}^{\hat{k}_l}, \{\boldsymbol{\lambda}_l^{\hat{k}_l}\}, \{\boldsymbol{\alpha}_l^{\hat{k}_l}\}, \{\boldsymbol{\beta}_l^{\hat{k}_l}\}, \{\gamma_s^{\hat{k}_l}\}), l \in \mathcal{Q}^{k+1} \\ \boldsymbol{\mu}_l^k, l \notin \mathcal{Q}^{k+1} \end{cases}, \tag{24}$$

$$\mathbf{r}_l^{k+1} = \begin{cases} \mathbf{r}_l^k - \eta_{\mathbf{r}} \nabla_{\mathbf{r}_l} \widetilde{L}_p(\{\boldsymbol{\mu}_l^{\hat{k}_l}\}, \{\mathbf{r}_l^{\hat{k}_l}\}, \{\mathbf{p}_l^{\hat{k}_l}\}, \{q_s^{\hat{k}_l}\}, \mathbf{x}^{\hat{k}_l}, \{\boldsymbol{\lambda}_l^{\hat{k}_l}\}, \{\boldsymbol{\alpha}_l^{\hat{k}_l}\}, \{\boldsymbol{\beta}_l^{\hat{k}_l}\}, \{\gamma_s^{\hat{k}_l}\}), l \in \mathcal{Q}^{k+1} \\ \mathbf{r}_l^k, l \notin \mathcal{Q}^{k+1} \end{cases}, \tag{25}$$

$$\mathbf{p}_l^{k+1} = \begin{cases} \mathbf{p}_l^k - \eta_{\mathbf{p}} \nabla_{\mathbf{p}_l} \widetilde{L}_p(\{\boldsymbol{\mu}_l^{\hat{k}_l}\}, \{\mathbf{r}_l^{\hat{k}_l}\}, \{\mathbf{p}_l^{\hat{k}_l}\}, \{q_s^{\hat{k}_l}\}, \mathbf{x}^{\hat{k}_l}, \{\boldsymbol{\lambda}_l^{\hat{k}_l}\}, \{\boldsymbol{\alpha}_l^{\hat{k}_l}\}, \{\boldsymbol{\beta}_l^{\hat{k}_l}\}, \{\gamma_s^{\hat{k}_l}\}), l \in \mathcal{Q}^{k+1} \\ \mathbf{p}_l^k, l \notin \mathcal{Q}^{k+1} \end{cases}, \tag{26}$$

$$q_s^{k+1} = q_s^k - \eta_q \nabla_{q_s} \widetilde{L}_p(\{\boldsymbol{\mu}_l^{k+1}\}, \{\mathbf{r}_l^{k+1}\}, \{\mathbf{p}_l^{k+1}\}, \{q_s^k\}, \mathbf{x}^k, \{\boldsymbol{\lambda}_l^k\}, \{\boldsymbol{\alpha}_l^k\}, \{\boldsymbol{\beta}_l^k\}, \{\gamma_s^k\}), \tag{27}$$

$$\mathbf{x}^{k+1} = \mathbf{x}^k - \eta_{\mathbf{x}} \nabla_{\mathbf{x}} \widetilde{L}_p(\{\boldsymbol{\mu}_l^{k+1}\}, \{\mathbf{r}_l^{k+1}\}, \{\mathbf{p}_l^{k+1}\}, \{q_s^{k+1}\}, \mathbf{x}^k, \{\boldsymbol{\lambda}_l^k\}, \{\boldsymbol{\alpha}_l^k\}, \{\boldsymbol{\beta}_l^k\}, \{\gamma_s^k\}), \tag{28}$$

$$\boldsymbol{\lambda}_l^{k+1} = \begin{cases} (\boldsymbol{\lambda}_l^k + \eta_{\boldsymbol{\lambda}} \nabla_{\boldsymbol{\lambda}_l} \widetilde{L}_p(\{\boldsymbol{\mu}_l^{k+1}\}, \{\mathbf{r}_l^{k+1}\}, \{\mathbf{p}_l^{k+1}\}, \{q_s^{k+1}\}, \mathbf{x}^{k+1}, \{\boldsymbol{\lambda}_l^k\}, \{\boldsymbol{\alpha}_l^k\}, \{\boldsymbol{\beta}_l^k\}, \{\gamma_s^k\}))^+, l \in \mathcal{Q}^{k+1} \\ \boldsymbol{\lambda}_l^k, l \notin \mathcal{Q}^{k+1} \end{cases}, \tag{29}$$

$$\boldsymbol{\alpha}_l^{k+1} = \begin{cases} (\boldsymbol{\alpha}_l^k + \eta_{\boldsymbol{\alpha}} \nabla_{\boldsymbol{\alpha}_l} \widetilde{L}_p(\{\boldsymbol{\mu}_l^{k+1}\}, \{\mathbf{r}_l^{k+1}\}, \{\mathbf{p}_l^{k+1}\}, \{q_s^{k+1}\}, \mathbf{x}^{k+1}, \{\boldsymbol{\lambda}_l^{k+1}\}, \{\boldsymbol{\alpha}_l^k\}, \{\boldsymbol{\beta}_l^k\}, \{\gamma_s^k\}))^+, l \in \mathcal{Q}^{k+1} \\ \boldsymbol{\alpha}_l^k, l \notin \mathcal{Q}^{k+1} \end{cases}, \tag{30}$$

$$\boldsymbol{\beta}_l^{k+1} = \begin{cases} (\boldsymbol{\beta}_l^k + \eta_{\boldsymbol{\beta}} \nabla_{\boldsymbol{\beta}_l} \widetilde{L}_p(\{\boldsymbol{\mu}_l^{k+1}\}, \{\mathbf{r}_l^{k+1}\}, \{\mathbf{p}_l^{k+1}\}, \{q_s^{k+1}\}, \mathbf{x}^{k+1}, \{\boldsymbol{\lambda}_l^{k+1}\}, \{\boldsymbol{\alpha}_l^{k+1}\}, \{\boldsymbol{\beta}_l^k\}, \{\gamma_s^k\}))^+, l \in \mathcal{Q}^{k+1} \\ \boldsymbol{\beta}_l^k, l \notin \mathcal{Q}^{k+1} \end{cases}, \tag{31}$$

$$\gamma_s^{k+1} = (\gamma_s^k + \eta_\gamma \nabla_{\gamma_s} \widetilde{L}_p(\{\boldsymbol{\mu}_l^{k+1}\}, \{\mathbf{r}_l^{k+1}\}, \{\mathbf{p}_l^{k+1}\}, \{q_s^{k+1}\}, \mathbf{x}^{k+1}, \{\boldsymbol{\lambda}_l^{k+1}\}, \{\boldsymbol{\alpha}_l^{k+1}\}, \{\boldsymbol{\beta}_l^{k+1}\}, \{\gamma_s^k\}))^+. \tag{32}$$

Based on the newly refined cutting plane sets, Triadic-OCD solves the nested optimization problem in an asynchronous distributed way. We first provide the Lagrangian function of (17) as follows,

$$\begin{aligned} &L_p(\{\boldsymbol{\mu}_l\}, \{\mathbf{r}_l\}, \{\mathbf{p}_l\}, \{q_s\}, \mathbf{x}, \{\boldsymbol{\lambda}_l\}, \{\boldsymbol{\alpha}_l\}, \{\boldsymbol{\beta}_l\}, \{\gamma_s\}) \\ &= \sum_{l=1}^{L} \{\|\boldsymbol{\mu}_l\|_2^2 - 2\mathbf{y}_l{}^T \boldsymbol{\mu}_l\} + \sum_{l=1}^{L} \boldsymbol{\lambda}_l^T(-\boldsymbol{\mu}_l - \rho_U \mathbf{1} + \mathbf{r}_l \circ \mathbf{r}_l) \\ &+ \sum_{l=1}^{L} \boldsymbol{\alpha}_l^T(\boldsymbol{\mu}_l - \rho_U \mathbf{1} + \mathbf{p}_l \circ \mathbf{p}_l) + \sum_{l=1}^{L} \boldsymbol{\beta}_l^T(\boldsymbol{\mu}_l - \mathbf{B}_i \mathbf{x}) \\ &+ \sum_{s=1}^{|\mathcal{P}^k|} \gamma_s(\mathbf{b}_s^\top \mathbf{x} + \kappa_s + q_s^2), \end{aligned} \tag{22}$$

where $\boldsymbol{\lambda}_l$, $\boldsymbol{\alpha}_l$, $\boldsymbol{\beta}_l$, γ_s are dual variables. $\gamma_s \in \mathbb{R}$ and $\boldsymbol{\lambda}_l$, $\boldsymbol{\alpha}_l$, $\boldsymbol{\beta}_l$ have the same dimensions as $\boldsymbol{\mu}_l$. Simplify the Lagrangian function of (17) as L_p, we next give the regularized version (Xu et al., 2023a) of L_p as follows,

$$\begin{aligned} &\widetilde{L}_p(\{\boldsymbol{\mu}_l\}, \{\mathbf{r}_l\}, \{\mathbf{p}_l\}, \{q_s\}, \mathbf{x}, \{\boldsymbol{\lambda}_l\}, \{\boldsymbol{\alpha}_l\}, \{\boldsymbol{\beta}_l\}, \{\gamma_s\}) = L_p - \\ &\sum_{l=1}^{L} \frac{c_{\boldsymbol{\lambda}}^k}{2} \|\boldsymbol{\lambda}_l\|^2 - \sum_{l=1}^{L} \frac{c_{\boldsymbol{\alpha}}^k}{2} \|\boldsymbol{\alpha}_l\|^2 - \sum_{l=1}^{L} \frac{c_{\boldsymbol{\beta}}^k}{2} \|\boldsymbol{\beta}_l\|^2 - \sum_{s=1}^{|\mathcal{P}^k|} \frac{c_\gamma^k}{2} \|\gamma_s\|^2, \end{aligned} \tag{23}$$

where $c_{\boldsymbol{\lambda}}^k$, $c_{\boldsymbol{\alpha}}^k$, $c_{\boldsymbol{\beta}}^k$, c_γ^k denote the regularization terms in the $(k+1)^{th}$ iteration. We represent the upper bound of $|\mathcal{P}^k|$ as P. Following the settings in (Zhang & Kwok, 2014), in each iteration, we update the variables in the master once it receives the local variables from S active workers. And we require every worker to communicate with the master at least once every τ iterations. Let $\mathcal{Q}^{k+1}$ denote the set of indexes of active workers in the $(k+1)^{th}$ iteration, the variables in Triadic-OCD are updated as follows,

(1) Local variables in workers are updated according to (24), (25), (26). Notations $\eta_{\boldsymbol{\mu}}$, $\eta_{\mathbf{r}}$ and $\eta_{\mathbf{p}}$ are step-sizes. And $\hat{k}_l$ denotes the last iteration in which worker l was active.

(2) After receiving the updates from active workers, the master updates the variables according to (27), (28), (29), (30), (31), (32). Notations η_q, $\eta_{\mathbf{x}}$, $\eta_{\boldsymbol{\lambda}}$, $\eta_{\boldsymbol{\alpha}}$, $\eta_{\boldsymbol{\beta}}$, η_γ are step-sizes.

The whole process of Triadic-OCD is summarized in Algorithm 1.

6. Theoretical Analysis

Theorem 6.1. *As the set of cutting planes is continuously updated, the optimal objective value of the approximation problem (17) converges monotonically.*

The complete proof is provided in Appendix D.

Let ∇G^k denote the gradient of L_p in the k-th iteration. According to (Boyd & Vandenberghe, 2004), we know that $(\{\boldsymbol{\mu}_l^k\}, \{\mathbf{r}_l^k\}, \{\mathbf{p}_l^k\}, \{q_s^k\}, \mathbf{x}^k, \{\boldsymbol{\lambda}_l^k\}, \{\boldsymbol{\alpha}_l^k\}, \{\boldsymbol{\beta}_l^k\}, \{\gamma_s^k\})$ is the optimal solution of problem (17) if and only if

$$K(\epsilon) \sim \mathcal{O}\left(\max\left\{(16(\frac{Mw_{\boldsymbol{\lambda}}^2}{\eta_\lambda^2}+\frac{Mw_{\boldsymbol{\alpha}}^2}{\eta_{\boldsymbol{\alpha}}^2}+\frac{Mw_{\boldsymbol{\beta}}^2}{\eta_{\boldsymbol{\beta}}^2}+\frac{Pw_\gamma^2}{\eta_\gamma^2})^2\frac{1}{\epsilon^2},(\frac{d_9(d_7+k_d\tau)(\tau-1)d_8}{\epsilon}+(K_1+2)^{\frac{1}{2}})^2\right\}\right). \quad (34)$$

$\|\nabla G^k\|^2 = 0$.

Definition 6.2. $(\{\boldsymbol{\mu}_l^k\}, \{\mathbf{r}_l^k\}, \{\mathbf{p}_l^k\}, \{q_s^k\}, \mathbf{x}^k, \{\boldsymbol{\lambda}_l^k\}, \{\boldsymbol{\alpha}_l^k\}, \{\boldsymbol{\beta}_l^k\}, \{\gamma_s^k\})$ is an ϵ-optimal point of (17) if $\|\nabla G^k\|^2 \le \epsilon$. Define $K(\epsilon)$ as the first iteration index such that $\|\nabla G^k\|^2 \le \epsilon$, i.e., $K(\epsilon) = \min\{k \mid \|\nabla G^k\|^2 \le \epsilon\}$.

Assumption 6.3. Following (Jiao et al., 2022; Qian et al., 2019), we assume that variables are bounded, i.e., $\|\boldsymbol{\mu}_l\|_\infty \le w_{\boldsymbol{\mu}}$, $\|\mathbf{r}_l\|_\infty \le w_{\mathbf{r}}$, $\|\mathbf{p}_l\|_\infty \le w_{\mathbf{p}}$, $\|q_s\|_\infty \le w_q$, $\|\boldsymbol{\lambda}_l\|_\infty \le w_{\boldsymbol{\lambda}}$, $\|\boldsymbol{\alpha}_l\|_\infty \le w_{\boldsymbol{\alpha}}$, $\|\boldsymbol{\beta}_l\|_\infty \le w_{\boldsymbol{\beta}}$, $\|\gamma_s\|_\infty \le w_\gamma$. Before obtaining the ϵ-optimal point, we assume variable $\mathbf{x}$ satisfies that $\|\mathbf{x}^{k+1} - \mathbf{x}^k\|^2 \ge \vartheta$, where $\vartheta > 0$ is a small constant. And the change of the $\mathbf{x}$ is upper bounded within τ iterations, i.e., $\|\mathbf{x}^k - \mathbf{x}^{k-\tau}\|^2 \le \tau k_1 \vartheta$,where $k_1 > 0$ is a constant.

Theorem 6.4. *(Iteration complexity) Suppose Assumption 6.3 holds, set*

$$\begin{aligned}
c_{\boldsymbol{\lambda}}^k &= \frac{1}{\eta_{\boldsymbol{\lambda}}(k+1)^{\frac{1}{4}}} \ge \underline{c_{\boldsymbol{\lambda}}}, \quad c_{\boldsymbol{\alpha}}^k = \frac{1}{\eta_{\boldsymbol{\alpha}}(k+1)^{\frac{1}{4}}} \ge \underline{c_{\boldsymbol{\alpha}}},\\
c_{\boldsymbol{\beta}}^k &= \frac{1}{\eta_{\boldsymbol{\beta}}(k+1)^{\frac{1}{4}}} \ge \underline{c_{\boldsymbol{\beta}}}, \quad c_\gamma^k = \frac{1}{\eta_\gamma(k+1)^{\frac{1}{4}}} \ge \underline{c_\gamma},\\
\eta_{\boldsymbol{\mu}} &= \frac{1}{\frac{8\xi}{\eta_{\boldsymbol{\lambda}}(\underline{c_{\boldsymbol{\lambda}}})^2}+\frac{8\xi}{\eta_{\boldsymbol{\alpha}}(\underline{c_{\boldsymbol{\alpha}}})^2}+\frac{8\xi}{\eta_{\boldsymbol{\beta}}(\underline{c_{\boldsymbol{\beta}}})^2}},\\
\eta_{\mathbf{x}} &= \frac{1}{\frac{8\xi}{\eta_{\boldsymbol{\beta}}(\underline{c_{\boldsymbol{\beta}}})^2}+\frac{8\xi\nu}{\eta_\gamma(\underline{c_\gamma})^2}},\\
\eta_{\mathbf{r}} &= \frac{\eta_{\boldsymbol{\lambda}}(\underline{c_{\boldsymbol{\lambda}}})^2}{32w_{\mathbf{r}}^2\xi}, \quad \eta_{\mathbf{p}} = \frac{\eta_{\boldsymbol{\alpha}}(\underline{c_{\boldsymbol{\alpha}}})^2}{32w_{\mathbf{p}}^2\xi}, \quad \eta_q = \frac{\eta_\gamma(\underline{c_\gamma})^2}{32w_q^2\xi}.
\end{aligned} \quad (33)$$

$\underline{c_{\boldsymbol{\lambda}}} \ge 0$, $\underline{c_\alpha} \ge 0$, $\underline{c_{\boldsymbol{\beta}}} \ge 0$, $\underline{c_\gamma} \ge 0$ represent the lower bound of $c_{\boldsymbol{\lambda}}^k$, $c_{\boldsymbol{\alpha}}^k$, $c_{\boldsymbol{\beta}}^k$, c_γ^k, respectively. For any prescribed ϵ, the iteration complexity of Triadic-OCD to converge to the ϵ-optimal point is shown in (34). Notation ξ, ν, d_7, d_8, d_9 and k_d are all constants.

Due to lack of space, we relegate the complete proof to the Appendix C.

7. Experiment

As discussed before, the problem considered in this paper subsumes the detection of FDIA in smart grids. Therefore, we conduct extensive experiments to show the effectiveness of Triadic-OCD for the FDIA detection task. In this section, we first briefly introduce the background of the smart grid, followed by a presentation of the comprehensive experimental results.

Algorithm 1 Triadic-OCD

Input: Iteration variable $k = 0$. Initialize $\{\boldsymbol{\mu}_l^0\}$, $\{\mathbf{r}_l^0\}$, $\{\mathbf{p}_l^0\}$, $\{q_s^0\}$, $\mathbf{x}^0$, $\{\boldsymbol{\lambda}_l^0\}$, $\{\boldsymbol{\alpha}_l^0\}$, $\{\boldsymbol{\beta}_l^0\}$, $\{\gamma_s^0\}$.

repeat
- **for** *workers* **do**
  - updates local variables $\boldsymbol{\mu}_l^{k+1}$,$\mathbf{r}_l^{k+1}$,$\mathbf{p}_l^{k+1}$ according to (24), (25) and (26);
- **end**
- $\boldsymbol{\mu}_l^{k+1}$, $\mathbf{r}_l^{k+1}$, $\mathbf{p}_l^{k+1}$ in active workers are transmitted to master;
- **for** *master* **do**
  - updates variables $\{q_s^{k+1}\}$, $\mathbf{x}^{k+1}$, $\{\boldsymbol{\lambda}_l^{k+1}\}$, $\{\boldsymbol{\alpha}_l^{k+1}\}$, $\{\boldsymbol{\beta}_l^{k+1}\}$, $\{\gamma_s^{k+1}\}$ according to (27), (28), (29), (30), (31) and (32);
- **end**
- master broadcasts $\{q_s^{k+1}\}$, $\mathbf{x}^{k+1}$, $\{\boldsymbol{\lambda}_l^{k+1}\}$, $\{\boldsymbol{\alpha}_l^{k+1}\}$, $\{\boldsymbol{\beta}_l^{k+1}\}$, $\{\gamma_s^{k+1}\}$ to active workers;
- **if** $(k+1) \bmod w = 0$ *and* $k \le K_1$ **then**
  - master computes $g_i(\mathbf{x}^{k+1})$ according to (14);
  - master updates $\mathcal{P}^{k+1}$, $\{q^{k+1}\}$ and $\{\gamma^{k+1}\}$ according to (19), (20), (21) and broadcasts them to all workers;
- **end**
- $k = k + 1$;

until *termination*;

Output: $\{\boldsymbol{\mu}_l^k\}$, $\{\mathbf{r}_l^k\}$, $\{\mathbf{p}_l^k\}$, $\{q_s^k\}$, $\mathbf{x}^k$, $\{\boldsymbol{\lambda}_l^k\}$, $\{\boldsymbol{\alpha}_l^k\}$, $\{\boldsymbol{\beta}_l^k\}$, $\{\gamma_s^k\}$

7.1. Dateset

A smart grid is an advanced electrical system that integrates information and communication technology to optimize energy supply, demand, and distribution. Given the potential for cyber-attacks to cause catastrophic consequences, the detection of false data injection attacks (FDIA) in smart grids is an important and extensively studied problem (Kosut et al., 2011; Cui et al., 2012). In the context of a smart grid system comprising $(N+1)$ buses and M meters, the dynamic DC power flow model of the system can be exactly expressed as equation (1), where $\boldsymbol{\theta}^{(t)} \in \mathbb{R}^N$ represents the phase angles (one reference angle) and $\mathbf{y}^{(t)} \in \mathbb{R}^M$ represents the meter readings. $\mathbf{H} \in \mathbb{R}^{M\times N}$ is the measurement matrix that is determined by the topology of the system and the admittance of each transmission line. In wide-area monitoring, the power grid is divided into multiple sub-regions, with each having limited access to its own meter measurements and communicating with other sub-regions and the

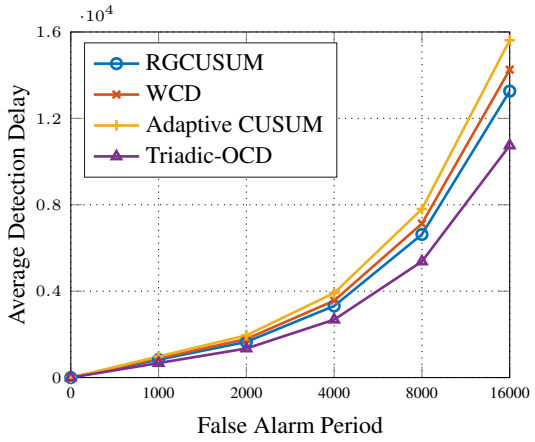

Figure 1. Performance comparison of different detection algorithms.

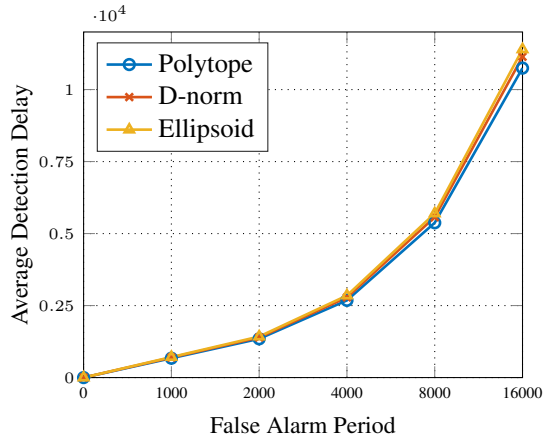

Figure 2. Performance of Triadic-OCD under various uncertainty Sets.

control center through the wireless medium. The centralized setup may not be feasible due to power and bandwidth constraints, leading to a growing interest in distributed detection to minimize communication overhead.

In our experiments, we utilized the measurement data generated from the IEEE-14 bus power system, which can be regarded as a benchmark dataset in current research studies. The IEEE 14-bus system as well as the system matrix can be divided into four sub-regions (Li et al., 2014). In our experiment, we assume that the system matrix can be accurately determined in all areas except for the 4th area. Suppose attackers utilize a sequence of randomly generated fabricated vectors to compromise the meter reading $\mathbf{y}^{(t)}$ from the time instant t_a. The generation process of the injected attacks is as follows,

$$\mathbf{a}^{(t)} = \mathbf{P}_{\mathbf{H}}^{\perp}\mathbf{u}, u_i \sim \mathbf{U}(0.1, 1), \tag{35}$$

where $\mathbf{P}_{\mathbf{H}}^{\perp} \triangleq \mathbf{I} - \mathbf{H}\left(\mathbf{H}^T\mathbf{H}\right)^{-1}\mathbf{H}^T$. Given the fluctuations in admittance caused by factors like environmental disturbances, the detection algorithms employed in the experiment utilize imprecise estimates of $\mathbf{H}$ to identify anomaly vectors injected into smart grid systems. We provide the inaccurate estimates of $\mathbf{H}$ in Appendix A, as well as the corresponding detailed settings of various uncertainty sets.

7.2. Numerical Results

As discussed in Section 2, we emphasized that the conventional OCD methods are built upon various assumptions, such as the assumption of constant distributions before and after the change, as well as the requirement for the system state to adhere to specific conditions. As a result, these methods are unsuitable for the problem considered in this paper and we implemented the Adaptive CUSUM algorithm (Huang et al., 2011) to exemplify this fact. Additionally, we conducted a performance comparison between our algorithm and two others: the method proposed in (Li et al., 2014) and the RGCUSUM algorithm (Zhang & Wang, 2021), which is claimed to be the state-of-the-art detector for the FDIA detection in smart grids. For simplicity, we name the wide-area cyber-attack detection method proposed in (Li et al., 2014) as WCD.

The experimental results after 500 runs are shown in Figure 1. For a fair comparison, we adhere to the conventional metric for OCD task, that is, we compare the average detection delay of different algorithms under the same False Alarm Periods (FAP). FAP refers to the detector's stopping time when no change occurs, serving as a measure to assess the risk of false alarms. From Figure 1, it is evident that our method consistently exhibit a smaller average detection delay for any given FAP. This observation underscores the superior performance achieved by Triadic-OCD. The adaptive CUCUM exhibits the worst performance, primarily due to the inconsistency of its assumptions with the problem of interest. Adaptive CUCUM presupposes the Gaussian distribution of $\boldsymbol{\theta}^{(t)}$, which does not hold in our problem. In our experiment, both RGUCUSM and WCD demonstrate inferior performance compared to Triadic-OCD. This is in line with our expectations since they can achieve commendable performance only when the system matrix $\mathbf{H}$ can be precisely determined. The imperfect knowledge about $\mathbf{H}$ leads to a rapid degradation of their detection performance. In contrast, Triadic-OCD effectively handles uncertainties within the system matrix. We also conduct experiments to show the performance of Triadic-OCD when the system matrix $\mathbf{H}$ is assumed to belong to different uncertainty sets. The uncertainty sets we consider are commonly used in practical applications, including ellipsoid uncertainty set, D-norm uncertainty set, and polyhedron uncertainty set (Yang et al., 2014). The number of Monte Carlo runs is 500. As depicted in Figure 2, Triadic-OCD consistently demonstrates

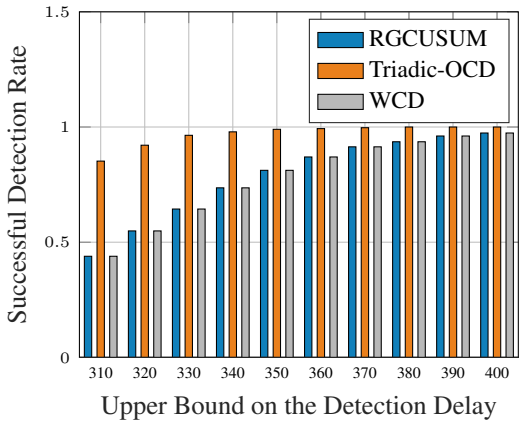

Figure 3. Successful detection rate versus the corresponding upper bound on detection delay.

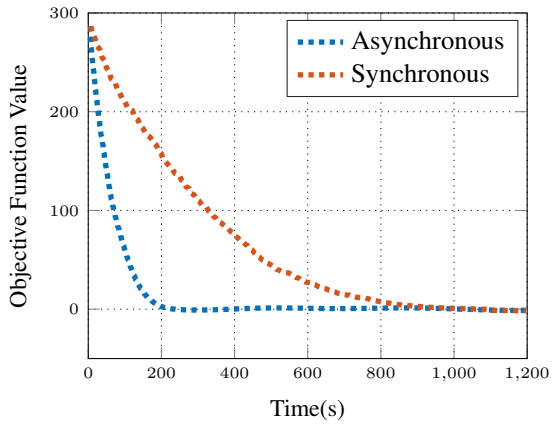

Figure 4. Convergence speed comparison between Triadic-OCD and its synchronous counterpart.

superior performance across diverse uncertainty sets, highlighting its generalization ability.

Furthermore, to demonstrate the robustness of Triadic-OCD against various attacking vectors, we generated 10^3 instances of attacking vectors randomly to encompass a wide array of attack behaviors according to (35). We calculate the success rate of different detectors under varying upper bounds on the detection delay. Specifically, the success rate refers to the proportion of attack vectors that can be successfully detected within the specified upper bound on the detection delay. We then plot the success rate corresponding to different upper bounds on detection delay in Figure 3. It can be seen from Figure 3 that as the upper bound increases, the successful detection rates of all detectors increase as expected. The proposed detector consistently outperforms RGCUSUM and WCD in successful detection rate, thus underscoring the robustness of Triadic-OCD against diverse attacking vectors.

Additionally, we also compare the convergence rate of Triadic-OCD under asynchronous and synchronous conditions to highlight the benefits of asynchronous variable updates in problem-solving. Following (Cohen et al., 2021), the delay of each worker is assumed to obey log-normal distribution LN(1, 0.5). Recall that S denotes the number of active workers required for updating the master node, and τ represents the maximum iteration interval for the communication between each worker and the master. For the asynchronous setting, we set $S = 10$ and $\tau = 10$. In the synchronous setup, the master can update its variables only after receiving updates from all workers. As shown in Figure 4, Triadic-OCD exhibits significantly faster convergence in the asynchronous mode compared to its synchronous counterpart. This is attributed to asynchronous variable updates, which prevent the algorithm from being hindered by the workers with high delays during each iteration, thus demonstrating the efficiency of our approach.

8. Conclusion

Existing studies often conduct online change detection with perfect knowledge of system parameters—a presumption that proves unfeasible in practical scenarios. Moreover, these studies typically focus on either centralized or synchronously distributed settings, which can lead to privacy breaches, straggler issues, and high communication costs. As a remedy, we develop an asynchronous framework for OCD with provable robustness, optimality, and convergence. To our best knowledge, this work represents the first step that tackles OCD with parameter uncertainties in an asynchronous setting. We also provide non-asymptotic theoretical analysis for the convergence property of triadic-OCD. Extensive experiments have been further conducted to elucidate the efficiency and effectiveness of the proposed algorithm.

Acknowledgements

This work was supported in part by the National Natural Science Foundation of China under Grant 12371519 and 61771013.

Impact Statement

This paper presents work whose goal is to advance the field of Machine Learning. There are many potential societal consequences of our work, none which we feel must be specifically highlighted here.

References

Andersson, E., Bock, D., and Frisén, M. Some statistical aspects of methods for detection of turning points in business cycles. *Journal of Applied Statistics*, 33(3):257–278, 2006.

Barnett, T. P., Pierce, D. W., and Schnur, R. Detection of anthropogenic climate change in the world's oceans. *Science*, 292(5515):270–274, 2001.

Boyd, S. and Vandenberghe, L. Localization and cutting-plane methods. *From Stanford EE 364b lecture notes*, 2007.

Boyd, S. P. and Vandenberghe, L. *Convex optimization*. Cambridge university press, 2004.

Cohen, A., Daniely, A., Drori, Y., Koren, T., and Schain, M. Asynchronous stochastic optimization robust to arbitrary delays. *Advances in Neural Information Processing Systems*, 34:9024–9035, 2021.

Cui, S., Han, Z., Kar, S., Kim, T. T., Poor, H. V., and Tajer, A. Coordinated data-injection attack and detection in the smart grid: A detailed look at enriching detection solutions. *IEEE Signal Processing Magazine*, 29(5):106–115, 2012.

Dou, S., Yang, K., and Poor, H. V. Pc2a: Predicting collective contextual anomalies via lstm with deep generative model. *IEEE Internet of Things Journal*, 6(6):9645–9655, 2019.

Franc, V., Sonnenburg, S., and Werner, T. Cutting plane methods in machine learning. *Optimization for Machine Learning*, pp. 185–218, 2011.

Gould, S., Fernando, B., Cherian, A., Anderson, P., Cruz, R. S., and Guo, E. On differentiating parameterized argmin and argmax problems with application to bi-level optimization. *arXiv preprint arXiv:1607.05447*, 2016.

Hare, J. Z. and Kaplan, L. Uncertainty-aware quickest change detection: An experimental study. In *2022 25th International Conference on Information Fusion (FUSION)*, pp. 1–8. IEEE, 2022.

Hare, J. Z., Kaplan, L., and Veeravalli, V. V. Toward uncertainty aware quickest change detection. In *2021 IEEE 24th International Conference on Information Fusion (FUSION)*, pp. 1–8. IEEE, 2021.

Huang, Y., Li, H., Campbell, K. A., and Han, Z. Defending false data injection attack on smart grid network using adaptive cusum test. In *2011 45th Annual Conference on Information Sciences and Systems*, pp. 1–6. IEEE, 2011.

Huang, Y., Tang, J., Cheng, Y., Li, H., Campbell, K. A., and Han, Z. Real-time detection of false data injection in smart grid networks: An adaptive cusum method and analysis. *IEEE Systems Journal*, 10(2):532–543, 2014.

Ji, K., Yang, J., and Liang, Y. Bilevel optimization: Convergence analysis and enhanced design. In *International conference on machine learning*, pp. 4882–4892. PMLR, 2021.

Jiao, Y., Yang, K., Wu, T., Song, D., and Jian, C. Asynchronous distributed bilevel optimization. In *The Eleventh International Conference on Learning Representations*, 2022.

Kosut, O., Jia, L., Thomas, R. J., and Tong, L. Malicious data attacks on the smart grid. *IEEE Transactions on Smart Grid*, 2(4):645–658, 2011.

Kurt, M. N., Yılmaz, Y., and Wang, X. Distributed quickest detection of cyber-attacks in smart grid. *IEEE Transactions on Information Forensics and Security*, 13(8): 2015–2030, 2018a.

Kurt, M. N., Yılmaz, Y., and Wang, X. Real-time detection of hybrid and stealthy cyber-attacks in smart grid. *IEEE Transactions on Information Forensics and Security*, 14 (2):498–513, 2018b.

Li, S., Yılmaz, Y., and Wang, X. Quickest detection of false data injection attack in wide-area smart grids. *IEEE Transactions on Smart Grid*, 6(6):2725–2735, 2014.

Liu, R., Liu, Y., Zeng, S., and Zhang, J. Towards gradient-based bilevel optimization with non-convex followers and beyond. *Advances in Neural Information Processing Systems*, 34:8662–8675, 2021a.

Liu, R., Liu, Y., Zeng, S., and Zhang, J. Towards gradient-based bilevel optimization with non-convex followers and beyond. *Advances in Neural Information Processing Systems*, 34:8662–8675, 2021b.

Molloy, T. L. and Ford, J. J. Misspecified and asymptotically minimax robust quickest change detection. *IEEE Transactions on Signal Processing*, 65(21):5730–5742, 2017.

Nishimori, K., Koide, Y., Kuwahara, D., Honmay, N., Yamada, H., and Hideo, M. Mimo sensor-evaluation on antenna arrangement. In *Proceedings of the 5th European Conference on Antennas and Propagation (EUCAP)*, pp. 2771–2775. IEEE, 2011.

Qian, Q., Zhu, S., Tang, J., Jin, R., Sun, B., and Li, H. Robust optimization over multiple domains. In *Proceedings of the AAAI Conference on Artificial Intelligence*, volume 33, pp. 4739–4746, 2019.

Subramanya, T. and Riggio, R. Centralized and federated learning for predictive vnf autoscaling in multi-domain 5g networks and beyond. *IEEE Transactions on Network and Service Management*, 18(1):63–78, 2021.

Tartakovsky, A., Nikiforov, I., and Basseville, M. *Sequential analysis: Hypothesis testing and changepoint detection*. CRC Press, 2014.

Tartakovsky, A. G., Rozovskii, B. L., Blazek, R. B., and Kim, H. A novel approach to detection of intrusions in computer networks via adaptive sequential and batch-sequential change-point detection methods. *IEEE transactions on signal processing*, 54(9):3372–3382, 2006.

Unnikrishnan, J., Veeravalli, V. V., and Meyn, S. P. Minimax robust quickest change detection. *IEEE Transactions on Information Theory*, 57(3):1604–1614, 2011.

Weber, T., Sklavos, A., and Meurer, M. Imperfect channel-state information in mimo transmission. *IEEE Transactions on Communications*, 54(3):543–552, 2006.

Xie, L. Minimax robust quickest change detection using wasserstein ambiguity sets. In *2022 IEEE International Symposium on Information Theory (ISIT)*, pp. 1909–1914. IEEE, 2022.

Xu, Z., Zhang, H., Xu, Y., and Lan, G. A unified single-loop alternating gradient projection algorithm for nonconvex–concave and convex–nonconcave minimax problems. *Mathematical Programming*, pp. 1–72, 2023a.

Xu, Z., Zhang, H., Xu, Y., and Lan, G. A unified single-loop alternating gradient projection algorithm for nonconvex–concave and convex–nonconcave minimax problems. *Mathematical Programming*, pp. 1–72, 2023b.

Yang, J., Ji, K., and Liang, Y. Provably faster algorithms for bilevel optimization. *Advances in Neural Information Processing Systems*, 34:13670–13682, 2021.

Yang, K., Huang, J., Wu, Y., Wang, X., and Chiang, M. Distributed robust optimization (DRO), part I: Framework and example. *Optimization and Engineering*, 15(1):35–67, 2014.

Yang, K., Liu, R., Sun, Y., Yang, J., and Chen, X. Deep network analyzer (dna): A big data analytics platform for cellular networks. *IEEE Internet of Things Journal*, 4(6): 2019–2027, 2016.

Zhang, J. Quickest detection of time-varying false data injection attacks in dynamic smart grids. In *ICASSP 2019-2019 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, pp. 2432–2436. IEEE, 2019.

Zhang, J. and Wang, X. Low-complexity quickest change detection in linear systems with unknown time-varying pre-and post-change distributions. *IEEE Transactions on Information Theory*, 67(3):1804–1824, 2021.

Zhang, R. and Kwok, J. Asynchronous distributed admm for consensus optimization. In *International conference on machine learning*, pp. 1701–1709. PMLR, 2014.

A. Experimental Settings

The IEEE 14-bus system as well as the system matrix can be divided into four sub-regions . The true system matrix for the 4th area is as follows (Li et al., 2014),

$$\mathbf{H}_4 = \begin{bmatrix} -1 & 3 & -1 \\ 0 & -1 & 0 \\ 0 & 0 & -1 \\ 0 & -1 & 1 \\ 0 & -1 & 2 \end{bmatrix}. \tag{36}$$

In our experiment, we assume that the system matrix can be accurately determined in all areas except for the 4th area.

The polyhedron uncertainty of $\mathbf{H}_4$ is set to be $\mathbf{D}_i\mathbf{h}_i \leq \mathbf{c}_i,\ \forall i$, where $\mathbf{h}_i$ represents the i-th column of $\mathbf{H}_4$. $\mathbf{D}_1$, $\mathbf{D}_2$ and $\mathbf{D}_3$ are set to be

$$\mathbf{D}_1 = \mathbf{D}_2 = \mathbf{D}_3 = \begin{bmatrix} \mathbf{I} \\ -\mathbf{I} \end{bmatrix},$$

where $\mathbf{I}$ represents the 5×5 identity matrix. $\mathbf{c}_1$, $\mathbf{c}_2$ and $\mathbf{c}_3$ are set to be

$$\begin{aligned} \mathbf{c}_1 &= [-0.5, 0.5, 0.5, 0.5, 0.5, 1.5, 0.5, 0.5, 0.5, 0.5]^T, \\ \mathbf{c}_2 &= [3.5, -0.5, 0.5, -0.5, -0.5, -2.5, 1.5, 0.5, 1.5, 1.5]^T, \\ \mathbf{c}_3 &= [-0.5, 0.5, -0.5, 1.5, 2.5, 1.5, 0.5, 1.5, -0.5, -1.5]^T. \end{aligned}$$

As for the ellipsoid uncertainty set, the uncertainty of each $\mathbf{h}_i$ is described as follows,

$$\mathbf{h}_i \in \{\overline{\mathbf{h}}_i + \mathbf{u} \mid \|\mathbf{u}\|_2 \leq 0.36\}, \quad \forall i,$$

where $\overline{\mathbf{h}}_i$ is estimate of the i-th column of $\mathbf{H}_t$. In our experiment, $\overline{\mathbf{h}}_i$ are set to be

$$\begin{aligned} \overline{\mathbf{h}}_1 &= [-1.0, 0.1, 0.3, -0.2, 0.0]^T, \\ \overline{\mathbf{h}}_2 &= [3.0, -0.7, 0.2, -1.3, -0.9]^T, \\ \overline{\mathbf{h}}_3 &= [-1.1, 0.2, -0.6, 0.7, 2.0]^T. \end{aligned}$$

For any vector belonging to the D-norm uncertain set, it is assumed that the vector has at most Γ uncertain components, and each component falls within the error interval determined by $\hat{u}$ (Yang et al., 2014). For all the $\mathbf{h}_i$ in our experiment, the parameter Γ and $\hat{u}$ are set to be 4 and 0.5, respectively.

B. Derivation of (5)

Since $\mathbf{n}^{(t)} \overset{i.i.d.}{\sim} \mathcal{N}\left(\mathbf{0}, \sigma_n^2 \mathbf{I}_M\right)$, according to (1) we have that,

$$f_p\left(\mathbf{y}^{(t)} \middle| \boldsymbol{\theta}^{(t)}, \mathbf{H}\right) = \frac{1}{\left(2\pi\sigma_n^2\right)^{\frac{M}{2}}} \exp\left[-\frac{1}{2\sigma_n^2}\left(\mathbf{y}^{(t)} - \mathbf{H}\boldsymbol{\theta}^{(t)}\right)^T \left(\mathbf{y}^{(t)} - \mathbf{H}\boldsymbol{\theta}^{(t)}\right)\right], \tag{37}$$

$$f_q\left(\mathbf{y}^{(t)} \middle| \boldsymbol{\theta}^{(t)}, \mathbf{a}^{(t)}, \mathbf{H}\right) = \frac{1}{\left(2\pi\sigma_n^2\right)^{\frac{M}{2}}} \exp\left[-\frac{1}{2\sigma_n^2}\left(\mathbf{y}^{(t)} - \mathbf{H}\boldsymbol{\theta}^{(t)} - \mathbf{a}^{(t)}\right)^T \left(\mathbf{y}^{(t)} - \mathbf{H}\boldsymbol{\theta}^{(t)} - \mathbf{a}^{(t)}\right)\right]. \tag{38}$$

According to the definition of $\Lambda_j^{(J)}$ in (3), we can obtain that,

$$\Lambda_j^{(J)} = \sup_{\mathbf{H}\in\mathcal{U}} \sum_{t=j}^{J} \left\{ \sup_{\boldsymbol{\theta}^{(t)}, \mathbf{a}^{(t)}: -\rho_U \mathbf{1} \leq \mathbf{x}^{(t)} \leq \rho_U \mathbf{1}, \mathbf{H}\mathbf{x}^{(t)} = \mathbf{0}} \ln f_q\left(\mathbf{y}^{(t)} \mid \boldsymbol{\theta}^{(t)}, \mathbf{a}^{(t)}, \mathbf{H}\right) - \sup_{\boldsymbol{\theta}^{(t)}} \ln f_p\left(\mathbf{y}^{(t)} \mid \boldsymbol{\theta}^{(t)}, \mathbf{H}\right) \right\} \tag{39}$$

Let $\tilde{\mathbf{y}}^{(t)}$ represents the component of $\mathbf{y}^{(t)}$ orthogonal to the column space of $\mathbf{H}$, i.e.,

$$\tilde{\mathbf{y}}^{(t)} = \mathbf{P}_{\mathbf{H}}^{\perp} \mathbf{y}^{(t)}. \tag{40}$$

Recall that $\mathbf{P}_{\mathbf{H}}^{\perp} \triangleq \mathbf{I} - \mathbf{H}\left(\mathbf{H}^T\mathbf{H}\right)^{-1}\mathbf{H}^T$. Plugging (37) and (38) into (39), we have that,

$$\Lambda_j^{(J)} = \sup_{\mathbf{H}\in\mathcal{U}} \sum_{t=j}^{J} \left\{ \sup_{\mathbf{a}^{(t)}: -\rho_U \mathbf{1} \leq \mathbf{x}^{(t)} \leq \rho_U \mathbf{1}, \mathbf{H}\mathbf{x}^{(t)} = \mathbf{0}} \left[-\frac{1}{2\sigma_n^2} \times \left(\tilde{\mathbf{y}}^{(t)} - \mathbf{x}^{(t)}\right)^T \left(\tilde{\mathbf{y}}^{(t)} - \mathbf{x}^{(t)}\right)\right] + \frac{1}{2\sigma_n^2}\|\tilde{\mathbf{y}}^{(t)}\|^2 \right\} \tag{41}$$

$$= \sup_{\mathbf{H}\in\mathcal{U}} \sum_{t=j}^{J} \sup_{-\rho_U \mathbf{1} \leq \mathbf{x}^{(t)} \leq \rho_U \mathbf{1}, \mathbf{H}\mathbf{x}^{(t)} = \mathbf{0}} \left[\frac{1}{2\sigma_n^2}\left\{2(\mathbf{x}^{(t)})^T \tilde{\mathbf{y}}^{(t)} - \|\mathbf{x}^{(t)}\|_2^2\right\}\right] \tag{42}$$

$$= \sum_{t=j}^{J} \frac{v_t}{2\sigma^2}, \tag{43}$$

where

$$v_t \triangleq \sup_{\mathbf{H}\in\mathcal{U}} \sup_{-\rho_U \mathbf{1} \leq \mathbf{x}^{(t)} \leq \rho_U \mathbf{1}, \mathbf{H}\mathbf{x}^{(t)} = \mathbf{0}} \left[\frac{1}{2\sigma_n^2}\left\{2(\mathbf{x}^{(t)})^T \tilde{\mathbf{y}}^{(t)} - \|\mathbf{x}^{(t)}\|_2^2\right\}\right]. \tag{44}$$

Since $\mathbf{H}\mathbf{x}^{(t)} = \mathbf{0}$ and $\tilde{\mathbf{y}}^{(t)}$ represents the component of $\mathbf{y}^{(t)}$ orthogonal to the column space of $\mathbf{H}$, (44) is equivalent to (5).

C. Complete Proof of Theorem 6.4

In this section, we provide the comprehensive proof of Theorem 6.4. We will start by introducing several definitions. Following that, three lemmas are presented, which are crucial components in the proof of Theorem 6.4. Finally, we provide the complete proof of Theorem 6.4.

Definition C.1. Based on the definition of ∇G^k, we further define :

$$\begin{aligned}
(\nabla G^k)_{\boldsymbol{\mu}_l} &= \nabla_{\boldsymbol{\mu}_l} L_p(\{\boldsymbol{\mu}_l^k\}, \{\mathbf{r}_l^k\}, \{\mathbf{p}_l^k\}, \{q_s^k\}, \mathbf{x}^k, \{\boldsymbol{\lambda}_l^k\}, \{\boldsymbol{\alpha}_l^k\}, \{\boldsymbol{\beta}_l^k\}, \{\gamma_s^k\}),\\
(\nabla G^k)_{\mathbf{r}_l} &= \nabla_{\mathbf{r}_l} L_p(\{\boldsymbol{\mu}_l^k\}, \{\mathbf{r}_l^k\}, \{\mathbf{p}_l^k\}, \{q_s^k\}, \mathbf{x}^k, \{\boldsymbol{\lambda}_l^k\}, \{\boldsymbol{\alpha}_l^k\}, \{\boldsymbol{\beta}_l^k\}, \{\gamma_s^k\}),\\
(\nabla G^k)_{\mathbf{p}_l} &= \nabla_{\mathbf{p}_l} L_p(\{\boldsymbol{\mu}_l^k\}, \{\mathbf{r}_l^k\}, \{\mathbf{p}_l^k\}, \{q_s^k\}, \mathbf{x}^k, \{\boldsymbol{\lambda}_l^k\}, \{\boldsymbol{\alpha}_l^k\}, \{\boldsymbol{\beta}_l^k\}, \{\gamma_s^k\}),\\
(\nabla G^k)_{q_s} &= \nabla_{q_s} L_p(\{\boldsymbol{\mu}_l^k\}, \{\mathbf{r}_l^k\}, \{\mathbf{p}_l^k\}, \{q_s^k\}, \mathbf{x}^k, \{\boldsymbol{\lambda}_l^k\}, \{\boldsymbol{\alpha}_l^k\}, \{\boldsymbol{\beta}_l^k\}, \{\gamma_s^k\}),\\
(\nabla G^k)_{\mathbf{x}} &= \nabla_{\mathbf{x}} L_p(\{\boldsymbol{\mu}_l^k\}, \{\mathbf{r}_l^k\}, \{\mathbf{p}_l^k\}, \{q_s^k\}, \mathbf{x}^k, \{\boldsymbol{\lambda}_l^k\}, \{\boldsymbol{\alpha}_l^k\}, \{\boldsymbol{\beta}_l^k\}, \{\gamma_s^k\}),\\
(\nabla G^k)_{\boldsymbol{\lambda}_l} &= \nabla_{\lambda_l} L_p(\{\boldsymbol{\mu}_l^k\}, \{\mathbf{r}_l^k\}, \{\mathbf{p}_l^k\}, \{q_s^k\}, \mathbf{x}^k, \{\boldsymbol{\lambda}_l^k\}, \{\boldsymbol{\alpha}_l^k\}, \{\boldsymbol{\beta}_l^k\}, \{\gamma_s^k\}),\\
(\nabla G^k)_{\boldsymbol{\alpha}_l} &= \nabla_{\alpha_l} L_p(\{\boldsymbol{\mu}_l^k\}, \{\mathbf{r}_l^k\}, \{\mathbf{p}_l^k\}, \{q_s^k\}, \mathbf{x}^k, \{\boldsymbol{\lambda}_l^k\}, \{\boldsymbol{\alpha}_l^k\}, \{\boldsymbol{\beta}_l^k\}, \{\gamma_s^k\}),\\
(\nabla G^k)_{\boldsymbol{\beta}_l} &= \nabla_{\beta_l} L_p(\{\boldsymbol{\mu}_l^k\}, \{\mathbf{r}_l^k\}, \{\mathbf{p}_l^k\}, \{q_s^k\}, \mathbf{x}^k, \{\boldsymbol{\lambda}_l^k\}, \{\boldsymbol{\alpha}_l^k\}, \{\boldsymbol{\beta}_l^k\}, \{\gamma_s^k\}),\\
(\nabla G^k)_{\gamma_s} &= \nabla_{\gamma_s} L_p(\{\boldsymbol{\mu}_l^k\}, \{\mathbf{r}_l^k\}, \{\mathbf{p}_l^k\}, \{q_s^k\}, \mathbf{x}^k, \{\boldsymbol{\lambda}_l^k\}, \{\boldsymbol{\alpha}_l^k\}, \{\boldsymbol{\beta}_l^k\}, \{\gamma_s^k\}).
\end{aligned} \tag{45}$$

Definition C.2. Similar to the definition of ∇G^k, Denote $\nabla \widetilde{G}^k$ as the gradient of $\widetilde{L}_p$ in the k-th iteration, i.e.,

$$\nabla \widetilde{G}^k = \begin{bmatrix}
\{\nabla_{\boldsymbol{\mu}_l} \widetilde{L}_p(\{\boldsymbol{\mu}_l^k\}, \{\mathbf{r}_l^k\}, \{\mathbf{p}_l^k\}, \mathbf{x}^k, \{q_s^k\}, \{\boldsymbol{\lambda}_l^k\}, \{\boldsymbol{\alpha}_l^k\}, \{\boldsymbol{\beta}_l^k\}, \{\gamma_s^k\})\}\\
\{\nabla_{\mathbf{r}_l} \widetilde{L}_p(\{\boldsymbol{\mu}_l^k\}, \{\mathbf{r}_l^k\}, \{\mathbf{p}_l^k\}, \mathbf{x}^k, \{q_s^k\}, \{\boldsymbol{\lambda}_l^k\}, \{\boldsymbol{\alpha}_l^k\}, \{\boldsymbol{\beta}_l^k\}, \{\gamma_s^k\})\}\\
\{\nabla_{\mathbf{p}_l} \widetilde{L}_p(\{\boldsymbol{\mu}_l^k\}, \{\mathbf{r}_l^k\}, \{\mathbf{p}_l^k\}, \mathbf{x}^k, \{q_s^k\}, \{\boldsymbol{\lambda}_l^k\}, \{\boldsymbol{\alpha}_l^k\}, \{\boldsymbol{\beta}_l^k\}, \{\gamma_s^k\})\}\\
\{\nabla_{q_s} \widetilde{L}_p(\{\boldsymbol{\mu}_l^k\}, \{\mathbf{r}_l^k\}, \{\mathbf{p}_l^k\}, \mathbf{x}^k, \{q_s^k\}, \{\boldsymbol{\lambda}_l^k\}, \{\boldsymbol{\alpha}_l^k\}, \{\boldsymbol{\beta}_l^k\}, \{\gamma_s^k\})\}\\
\nabla_{\mathbf{x}} \widetilde{L}_p(\{\boldsymbol{\mu}_l^k\}, \{\mathbf{r}_l^k\}, \{\mathbf{p}_l^k\}, \mathbf{x}^k, \{q_s^k\}, \{\boldsymbol{\lambda}_l^k\}, \{\boldsymbol{\alpha}_l^k\}, \{\boldsymbol{\beta}_l^k\}, \{\gamma_s^k\})\\
\{\nabla_{\boldsymbol{\lambda}_l} \widetilde{L}_p(\{\boldsymbol{\mu}_l^k\}, \{\mathbf{r}_l^k\}, \{\mathbf{p}_l^k\}, \mathbf{x}^k, \{q_s^k\}, \{\boldsymbol{\lambda}_l^k\}, \{\boldsymbol{\alpha}_l^k\}, \{\boldsymbol{\beta}_l^k\}, \{\gamma_s^k\})\}\\
\{\nabla_{\boldsymbol{\alpha}_l} \widetilde{L}_p(\{\boldsymbol{\mu}_l^k\}, \{\mathbf{r}_l^k\}, \{\mathbf{p}_l^k\}, \mathbf{x}^k, \{q_s^k\}, \{\boldsymbol{\lambda}_l^k\}, \{\boldsymbol{\alpha}_l^k\}, \{\boldsymbol{\beta}_l^k\}, \{\gamma_s^k\})\}\\
\{\nabla_{\boldsymbol{\beta}_l} \widetilde{L}_p(\{\boldsymbol{\mu}_l^k\}, \{\mathbf{r}_l^k\}, \{\mathbf{p}_l^k\}, \mathbf{x}^k, \{q_s^k\}, \{\boldsymbol{\lambda}_l^k\}, \{\boldsymbol{\alpha}_l^k\}, \{\boldsymbol{\beta}_l^k\}, \{\gamma_s^k\})\}\\
\{\nabla_{\gamma_s} \widetilde{L}_p(\{\boldsymbol{\mu}_l^k\}, \{\mathbf{r}_l^k\}, \{\mathbf{p}_l^k\}, \mathbf{x}^k, \{q_s^k\}, \{\boldsymbol{\lambda}_l^k\}, \{\boldsymbol{\alpha}_l^k\}, \{\boldsymbol{\beta}_l^k\}, \{\gamma_s^k\})\}
\end{bmatrix}, \tag{46}$$

with

$$\begin{aligned}
(\nabla \widetilde{G}^k)_{\boldsymbol{\mu}_l} &= \nabla_{\boldsymbol{\mu}_l} \widetilde{L}_p(\{\boldsymbol{\mu}_l^k\}, \{\mathbf{r}_l^k\}, \{\mathbf{p}_l^k\}, \{q_s^k\}, \mathbf{x}^k, \{\boldsymbol{\lambda}_l^k\}, \{\boldsymbol{\alpha}_l^k\}, \{\boldsymbol{\beta}_l^k\}, \{\gamma_s^k\}),\\
(\nabla \widetilde{G}^k)_{\mathbf{r}_l} &= \nabla_{\mathbf{r}_l} \widetilde{L}_p(\{\boldsymbol{\mu}_l^k\}, \{\mathbf{r}_l^k\}, \{\mathbf{p}_l^k\}, \{q_s^k\}, \mathbf{x}^k, \{\boldsymbol{\lambda}_l^k\}, \{\boldsymbol{\alpha}_l^k\}, \{\boldsymbol{\beta}_l^k\}, \{\gamma_s^k\}),\\
(\nabla \widetilde{G}^k)_{\mathbf{p}_l} &= \nabla_{\mathbf{p}_l} \widetilde{L}_p(\{\boldsymbol{\mu}_l^k\}, \{\mathbf{r}_l^k\}, \{\mathbf{p}_l^k\}, \{q_s^k\}, \mathbf{x}^k, \{\boldsymbol{\lambda}_l^k\}, \{\boldsymbol{\alpha}_l^k\}, \{\boldsymbol{\beta}_l^k\}, \{\gamma_s^k\}),\\
(\nabla \widetilde{G}^k)_{q_s} &= \nabla_{q_s} \widetilde{L}_p(\{\boldsymbol{\mu}_l^k\}, \{\mathbf{r}_l^k\}, \{\mathbf{p}_l^k\}, \{q_s^k\}, \mathbf{x}^k, \{\boldsymbol{\lambda}_l^k\}, \{\boldsymbol{\alpha}_l^k\}, \{\boldsymbol{\beta}_l^k\}, \{\gamma_s^k\}),\\
(\nabla \widetilde{G}^k)_{\mathbf{x}} &= \nabla_{\mathbf{x}} \widetilde{L}_p(\{\boldsymbol{\mu}_l^k\}, \{\mathbf{r}_l^k\}, \{\mathbf{p}_l^k\}, \{q_s^k\}, \mathbf{x}^k, \{\boldsymbol{\lambda}_l^k\}, \{\boldsymbol{\alpha}_l^k\}, \{\boldsymbol{\beta}_l^k\}, \{\gamma_s^k\}),\\
(\nabla \widetilde{G}^k)_{\boldsymbol{\lambda}_l} &= \nabla_{\lambda_l} \widetilde{L}_p(\{\boldsymbol{\mu}_l^k\}, \{\mathbf{r}_l^k\}, \{\mathbf{p}_l^k\}, \{q_s^k\}, \mathbf{x}^k, \{\boldsymbol{\lambda}_l^k\}, \{\boldsymbol{\alpha}_l^k\}, \{\boldsymbol{\beta}_l^k\}, \{\gamma_s^k\}),\\
(\nabla \widetilde{G}^k)_{\boldsymbol{\alpha}_l} &= \nabla_{\alpha_l} \widetilde{L}_p(\{\boldsymbol{\mu}_l^k\}, \{\mathbf{r}_l^k\}, \{\mathbf{p}_l^k\}, \{q_s^k\}, \mathbf{x}^k, \{\boldsymbol{\lambda}_l^k\}, \{\boldsymbol{\alpha}_l^k\}, \{\boldsymbol{\beta}_l^k\}, \{\gamma_s^k\}),\\
(\nabla \widetilde{G}^k)_{\boldsymbol{\beta}_l} &= \nabla_{\beta_l} \widetilde{L}_p(\{\boldsymbol{\mu}_l^k\}, \{\mathbf{r}_l^k\}, \{\mathbf{p}_l^k\}, \{q_s^k\}, \mathbf{x}^k, \{\boldsymbol{\lambda}_l^k\}, \{\boldsymbol{\alpha}_l^k\}, \{\boldsymbol{\beta}_l^k\}, \{\gamma_s^k\}),\\
(\nabla \widetilde{G}^k)_{\gamma_s} &= \nabla_{\gamma_s} \widetilde{L}_p(\{\boldsymbol{\mu}_l^k\}, \{\mathbf{r}_l^k\}, \{\mathbf{p}_l^k\}, \{q_s^k\}, \mathbf{x}^k, \{\boldsymbol{\lambda}_l^k\}, \{\boldsymbol{\alpha}_l^k\}, \{\boldsymbol{\beta}_l^k\}, \{\gamma_s^k\}).
\end{aligned} \tag{47}$$

Definition C.3. In the kth iteration of our algorithm, we define the last iteration in which the lth worker was active as $\hat{k}_l$, and the next iteration in which the lth worker will be active as $\overline{k}_l$. Furthermore, we represent the set of iteration indices in which the lth worker is active during the $K_1 + K + \tau$ iteration as $\mathcal{V}_l(K)$. And the jth element in $\mathcal{V}_l(K)$ is represented as $\hat{v}_l(j)$.

Based on the above definitions, we next provide the proof of Lemma C.4, Lemma C.5 and Lemma C.6.

Lemma C.4. *According to Eq. (22), function L_p has Lipschitz continuous Hessian and let L_w denote the Lipschitz constant. Based on definition of $\eta_{\boldsymbol{\mu}}$, $\eta_{\mathbf{x}}$, $\eta_{\mathbf{r}}$, $\eta_{\mathbf{p}}$ and η_q, we set $\eta_{\boldsymbol{\mu}}^k$, $\eta_{\mathbf{x}}^k$, $\eta_{\mathbf{r}}^k$, $\eta_{\mathbf{p}}^k$ and η_q^k to be:*

$$\begin{aligned}
&\eta_{\boldsymbol{\mu}}^k = \frac{1}{\frac{8\xi}{\eta_{\boldsymbol{\lambda}}(c_{\boldsymbol{\lambda}}^k)^2} + \frac{8\xi}{\eta_{\boldsymbol{\alpha}}(c_{\boldsymbol{\alpha}}^k)^2} + \frac{8\xi}{\eta_{\boldsymbol{\beta}}(c_{\boldsymbol{\beta}}^k)^2}}, \quad \eta_{\mathbf{x}}^k = \frac{1}{\frac{8\xi}{\eta_{\boldsymbol{\beta}}(c_{\boldsymbol{\beta}}^k)^2} + \frac{8\xi}{\eta_{\gamma}(c_{\gamma}^k)^2}\sum_{s=1}^{|\mathcal{P}^k|}\|b_s\|^2}, \\
&\eta_{\mathbf{r}}^k = \frac{\eta_{\boldsymbol{\lambda}}(c_{\boldsymbol{\lambda}}^k)^2}{32w_{\mathbf{r}}^2\xi}, \quad \eta_{\mathbf{p}}^k = \frac{\eta_{\boldsymbol{\alpha}}(c_{\boldsymbol{\alpha}}^k)^2}{32w_{\mathbf{p}}^2\xi}, \quad \eta_q^k = \frac{\eta_{\gamma}(c_{\gamma}^k)^2}{32w_q^2\xi}.
\end{aligned} \tag{48}$$

We can obtain that,

$$\begin{aligned}
&L_p(\{\boldsymbol{\mu}_l^{k+1}\}, \{\mathbf{r}_l^{k+1}\}, \{\mathbf{p}_l^{k+1}\}, \{q_s^{k+1}\}, \mathbf{x}^{k+1}, \{\boldsymbol{\lambda}_l^k\}, \{\boldsymbol{\alpha}_l^k\}, \{\boldsymbol{\beta}_l^k\}, \{\gamma_s^k\}) \\
&- L_p(\{\boldsymbol{\mu}_l^k\}, \{\mathbf{r}_l^k\}, \{\mathbf{p}_l^k\}, \{q_s^k\}, \mathbf{x}^k, \{\boldsymbol{\lambda}_l^k\}, \{\boldsymbol{\alpha}_l^k\}, \{\boldsymbol{\beta}_l^k\}, \{\gamma_s^k\}) \\
&\leq (\frac{w_{\boldsymbol{\mu}}L_w M^{\frac{1}{2}}}{3} + 1 - \frac{1}{\eta_{\boldsymbol{\mu}}^k})\sum_{l=1}^{L}\|\boldsymbol{\mu}_l^{k+1} - \boldsymbol{\mu}_l^k\|^2 + (-\frac{1}{\eta_{\mathbf{r}}^k} + \frac{w_{\mathbf{r}}L_w M^{\frac{1}{2}}}{3} + w_{\boldsymbol{\lambda}})\sum_{l=1}^{L}\|\mathbf{r}_l^{k+1} - \mathbf{r}_l^k\|^2 \\
&+ (-\frac{1}{\eta_{\mathbf{p}}^k} + \frac{w_{\mathbf{p}}L_w M^{\frac{1}{2}}}{3} + w_{\boldsymbol{\alpha}})\sum_{l=1}^{L}||\mathbf{p}_l^{k+1} - \mathbf{p}_l^k||^2 + (-\frac{1}{\eta_{\mathbf{q}}^k} + \frac{w_{\mathbf{q}}L_w}{3} + w_{\gamma})\sum_{s=1}^{|\mathcal{P}^k|}||q_s^{k+1} - q_s^k||^2 \\
&+ (\frac{L_w}{6} - \frac{1}{\eta_{\mathbf{x}}^k})\|\mathbf{x}^{k+1} - \mathbf{x}^k\|^2.
\end{aligned} \tag{49}$$

Proof. According to the Lipschitz property of L_p, we can obtain that,

$$\begin{aligned}
&L_p(\{\boldsymbol{\mu}_l^{k+1}\}, \{\mathbf{r}_l^k\}, \{\mathbf{p}_l^k\}, \{q_s^k\}, \mathbf{x}^k, \{\boldsymbol{\lambda}_l^k\}, \{\boldsymbol{\alpha}_l^k\}, \{\boldsymbol{\beta}_l^k\}, \{\gamma_s^k\}) \\
&- L_p(\{\boldsymbol{\mu}_l^k\}, \{\mathbf{r}_l^k\}, \{\mathbf{p}_l^k\}, \{q_s^k\}, \mathbf{x}^k, \{\boldsymbol{\lambda}_l^k\}, \{\boldsymbol{\alpha}_l^k\}, \{\boldsymbol{\beta}_l^k\}, \{\gamma_s^k\}) \\
&\leq \sum_{l=1}^{L}\left(\left\langle \nabla_{\boldsymbol{\mu}_l} L_p(\{\boldsymbol{\mu}_l^k\}, \{\mathbf{r}_l^k\}, \{\mathbf{p}_l^k\}, \{q_s^k\}, \mathbf{x}^k, \{\boldsymbol{\lambda}_l^k\}, \{\boldsymbol{\alpha}_l^k\}, \{\boldsymbol{\beta}_l^k\}, \{\gamma_s^k\}), \boldsymbol{\mu}_l^{k+1} - \boldsymbol{\mu}_l^k \right\rangle + \frac{L_w}{6}||\boldsymbol{\mu}_l^{k+1} - \boldsymbol{\mu}_l^k||^3\right) \\
&+ \frac{1}{2}\sum_{l=1}^{L}\left\langle \nabla^2_{\boldsymbol{\mu}_l} L_p(\{\boldsymbol{\mu}_l^k\}, \{\mathbf{r}_l^k\}, \{\mathbf{p}_l^k\}, \{q_s^k\}, \mathbf{x}^k, \{\boldsymbol{\lambda}_l^k\}, \{\boldsymbol{\alpha}_l^k\}, \{\boldsymbol{\beta}_l^k\}, \{\gamma_s^k\})(\boldsymbol{\mu}_l^{k+1} - \boldsymbol{\mu}_l^k), \boldsymbol{\mu}_l^{k+1} - \boldsymbol{\mu}_l^k \right\rangle.
\end{aligned} \tag{50}$$

By employing (24), since $\eta_{\boldsymbol{\mu}} \leq \eta_{\boldsymbol{\mu}}^k$ for $\forall k$, we have that,

$$\left\langle \nabla_{\boldsymbol{\mu}_l} L_p(\{\boldsymbol{\mu}_l^{\hat{k}_l}\}, \{\mathbf{r}_l^{\hat{k}_l}\}, \{\mathbf{p}_l^{\hat{k}_l}\}, \{q_s^{\hat{k}_l}\}, \mathbf{x}^{\hat{k}_l}, \{\boldsymbol{\lambda}_l^{\hat{k}_l}\}, \{\boldsymbol{\alpha}_l^{\hat{k}_l}\}, \{\boldsymbol{\beta}_l^{\hat{k}_l}\}, \{\gamma_s^{\hat{k}_l}\}), \boldsymbol{\mu}_l^{k+1} - \boldsymbol{\mu}_l^k \right\rangle \leq -\frac{1}{\eta_{\boldsymbol{\mu}}^k}||\boldsymbol{\mu}_l^{k+1} - \boldsymbol{\mu}_l^k||^2. \tag{51}$$

Given that $\boldsymbol{\mu}_l^k = \boldsymbol{\mu}_l^{\hat{k}_l}$, $\boldsymbol{\lambda}_l^k = \boldsymbol{\lambda}_l^{\hat{k}_l}$, $\boldsymbol{\alpha}_l^k = \boldsymbol{\alpha}_l^{\hat{k}_l}$, and $\boldsymbol{\beta}_l^k = \boldsymbol{\beta}_l^{\hat{k}_l}$, according to the function L_p provided in (22), we have

$$\left\langle \nabla_{\boldsymbol{\mu}_l} L_p(\{\boldsymbol{\mu}_l^k\}, \{\mathbf{r}_l^k\}, \{\mathbf{p}_l^k\}, \{q_s^k\}, \mathbf{x}^k, \{\boldsymbol{\lambda}_l^k\}, \{\boldsymbol{\alpha}_l^k\}, \{\boldsymbol{\beta}_l^k\}, \{\gamma_s^k\}), \boldsymbol{\mu}_l^{k+1} - \boldsymbol{\mu}_l^k \right\rangle \leq -\frac{1}{\eta_{\boldsymbol{\mu}}^k}||\boldsymbol{\mu}_l^{k+1} - \boldsymbol{\mu}_l^k||^2. \tag{52}$$

It can also be seen from (22) that

$$\tfrac{1}{2}\left\langle \nabla^2_{\boldsymbol{\mu}_l} L_p(\{\boldsymbol{\mu}_l^k\}, \{\mathbf{r}_l^k\}, \{\mathbf{p}_l^k\}, \{q_s^k\}, \mathbf{x}^k, \{\boldsymbol{\lambda}_l^k\}, \{\boldsymbol{\alpha}_l^k\}, \{\boldsymbol{\beta}_l^k\}, \{\gamma_s^k\})(\boldsymbol{\mu}_l^{k+1} - \boldsymbol{\mu}_l^k), \boldsymbol{\mu}_l^{k+1} - \boldsymbol{\mu}_l^k \right\rangle = \|\boldsymbol{\mu}_l^{k+1} - \boldsymbol{\mu}_l^k\|^2. \tag{53}$$

Based on the upper bound of $\|\boldsymbol{\mu}_l\|_\infty$ provided in Assumption 6.3, by employing trigonometric inequality, we have

$$\frac{L_w}{6}\sum_{l=1}^{L}||\boldsymbol{\mu}_l^{k+1} - \boldsymbol{\mu}_l^k||^3 \leq \frac{w_{\boldsymbol{\mu}}L_w M^{\frac{1}{2}}}{3}\sum_{l=1}^{L}||\boldsymbol{\mu}_l^{k+1} - \boldsymbol{\mu}_l^k||^2. \tag{54}$$

Combining (50), (52), (53) with (54), we can obtain that,

$$\begin{aligned}
&L_p(\{\boldsymbol{\mu}_l^{k+1}\}, \{\mathbf{r}_l^k\}, \{\mathbf{p}_l^k\}, \{q_s^k\}, \mathbf{x}^k, \{\boldsymbol{\lambda}_l^k\}, \{\boldsymbol{\alpha}_l^k\}, \{\boldsymbol{\beta}_l^k\}, \{\gamma_s^k\}) \\
&- L_p(\{\boldsymbol{\mu}_l^k\}, \{\mathbf{r}_l^k\}, \{\mathbf{p}_l^k\}, \{q_s^k\}, \mathbf{x}^k, \{\boldsymbol{\lambda}_l^k\}, \{\boldsymbol{\alpha}_l^k\}, \{\boldsymbol{\beta}_l^k\}, \{\gamma_s^k\}) \\
&\leq (-\frac{1}{\eta_{\boldsymbol{\mu}}^k} + \frac{w_{\boldsymbol{\mu}}L_w M^{\frac{1}{2}}}{3} + 1)\sum_{l=1}^{L}\left(||\boldsymbol{\mu}_l^{k+1} - \boldsymbol{\mu}_l^k||^2\right).
\end{aligned} \tag{55}$$

According to the Lipschitz property of L_p, we can obtain that,

$$\begin{aligned}
&L_p(\{\boldsymbol{\mu}_l^{k+1}\},\{\mathbf{r}_l^{k+1}\},\{\mathbf{p}_l^k\},\{q_s^k\},\mathbf{x}^k,\{\boldsymbol{\lambda}_l^k\},\{\boldsymbol{\alpha}_l^k\},\{\boldsymbol{\beta}_l^k\},\{\gamma_s^k\})\} \\
&-L_p(\{\boldsymbol{\mu}_l^{k+1}\},\{\mathbf{r}_l^k\},\{\mathbf{p}_l^k\},\{q_s^k\},\mathbf{x}^k,\{\boldsymbol{\lambda}_l^k\},\{\boldsymbol{\alpha}_l^k\},\{\boldsymbol{\beta}_l^k\},\{\gamma_s^k\})\} \\
&\leq \sum_{l=1}^{L}\left\langle \nabla_{\mathbf{r}_l} L_p(\{\boldsymbol{\mu}_l^{k+1}\},\{\mathbf{r}_l^k\},\{\mathbf{p}_l^k\},\{q_s^k\},\mathbf{x}^k,\{\boldsymbol{\lambda}_l^k\},\{\boldsymbol{\alpha}_l^k\},\{\boldsymbol{\beta}_l^k\},\{\gamma_s^k\}),\mathbf{r}_l^{k+1}-\mathbf{r}_l^k\right\rangle + \frac{L_w}{6}\sum_{l=1}^{L}||\mathbf{r}_l^{k+1}-\mathbf{r}_l^k||^3 \\
&+\frac{1}{2}\sum_{l=1}^{L}\left\langle \nabla^2_{\mathbf{r}_l} L_p(\{\boldsymbol{\mu}_l^{k+1}\},\{\mathbf{r}_l^k\},\{\mathbf{p}_l^k\},\{q_s^k\},\mathbf{x}^k,\{\boldsymbol{\lambda}_l^k\},\{\boldsymbol{\alpha}_l^k\},\{\boldsymbol{\beta}_l^k\},\{\gamma_s^k\})(\mathbf{r}_l^{k+1}-\mathbf{r}_l^k),\mathbf{r}_l^{k+1}-\mathbf{r}_l^k\right\rangle,
\end{aligned} \tag{56}$$

Similar to (51)-(52), we have

$$\left\langle \nabla_{\mathbf{r}_l} L_p(\{\boldsymbol{\mu}_l^{k+1}\},\{\mathbf{r}_l^k\},\{\mathbf{p}_l^k\},\{q_s^k\},\mathbf{x}^k,\{\boldsymbol{\lambda}_l^k\},\{\boldsymbol{\alpha}_l^k\},\{\boldsymbol{\beta}_l^k\},\{\gamma_s^k\}),\mathbf{r}_l^{k+1}-\mathbf{r}_l^k\right\rangle \leq -\frac{1}{\eta_{\mathbf{r}}^k}||\mathbf{r}_l^{k+1}-\mathbf{r}_l^k||^2. \tag{57}$$

Based on Assumption 6.3, we can obtain that,

$$\tfrac{1}{2}\left(\left\langle \nabla^2_{\mathbf{r}_l} L_p(\{\boldsymbol{\mu}_l^{k+1}\},\{\mathbf{r}_l^k\},\{\mathbf{p}_l^k\},\{q_s^k\},\mathbf{x}^k,\{\boldsymbol{\lambda}_l^k\},\{\boldsymbol{\alpha}_l^k\},\{\boldsymbol{\beta}_l^k\},\{\gamma_s^k\})(\mathbf{r}_l^{k+1}-\mathbf{r}_l^k),\mathbf{r}_l^{k+1}-\mathbf{r}_l^k\right\rangle\right) \leq w_{\boldsymbol{\lambda}}||\mathbf{r}_l^{k+1}-\mathbf{r}_l^k||^2. \tag{58}$$

By employing trigonometric inequality, we have that,

$$\frac{L_w}{6}\sum_{l=1}^{L}||\mathbf{r}_l^{k+1}-\mathbf{r}_l^k||^3 \leq \frac{w_{\mathbf{r}}L_w M^{\frac{1}{2}}}{3}\sum_{l=1}^{L}||\mathbf{r}_l^{k+1}-\mathbf{r}_l^k||^2. \tag{59}$$

Following (56)-(59), we can obtain that,

$$\begin{aligned}
&L_p(\{\boldsymbol{\mu}_l^{k+1}\},\{\mathbf{r}_l^{k+1}\},\{\mathbf{p}_l^k\},\{q_s^k\},\mathbf{x}^k,\{\boldsymbol{\lambda}_l^k\},\{\boldsymbol{\alpha}_l^k\},\{\boldsymbol{\beta}_l^k\},\{\gamma_s^k\}) \\
&-L_p(\{\boldsymbol{\mu}_l^{k+1}\},\{\mathbf{r}_l^k\},\{\mathbf{p}_l^k\},\{q_s^k\},\mathbf{x}^k,\{\boldsymbol{\lambda}_l^k\},\{\boldsymbol{\alpha}_l^k\},\{\boldsymbol{\beta}_l^k\},\{\gamma_s^k\}) \\
&\leq (-\frac{1}{\eta_{\mathbf{r}}^k}+\frac{w_{\mathbf{r}}L_w M^{\frac{1}{2}}}{3}+w_{\boldsymbol{\lambda}})\sum_{l=1}^{L}||\mathbf{r}_l^{k+1}-\mathbf{r}_l^k||^2.
\end{aligned} \tag{60}$$

Likewise, similar results can be obtained for the variable $\mathbf{p}_l$ and q_s.

$$\begin{aligned}
&L_p(\{\boldsymbol{\mu}_l^{k+1}\},\{\mathbf{r}_l^{k+1}\},\{\mathbf{p}_l^{k+1}\},\{q_s^k\},\mathbf{x}^k,\{\boldsymbol{\lambda}_l^k\},\{\boldsymbol{\alpha}_l^k\},\{\boldsymbol{\beta}_l^k\},\{\gamma_s^k\}) \\
&-L_p(\{\boldsymbol{\mu}_l^{k+1}\},\{\mathbf{r}_l^{k+1}\},\{\mathbf{p}_l^k\},\{q_s^k\},\mathbf{x}^k,\{\boldsymbol{\lambda}_l^k\},\{\boldsymbol{\alpha}_l^k\},\{\boldsymbol{\beta}_l^k\},\{\gamma_s^k\}) \\
&\leq (-\frac{1}{\eta_{\mathbf{p}}^k}+\frac{w_{\mathbf{p}}L_w M^{\frac{1}{2}}}{3}+w_{\boldsymbol{\alpha}})\sum_{l=1}^{L}||\mathbf{p}_l^{k+1}-\mathbf{p}_l^k||^2.
\end{aligned} \tag{61}$$

$$\begin{aligned}
&L_p(\{\boldsymbol{\mu}_l^{k+1}\},\{\mathbf{r}_l^{k+1}\},\{\mathbf{p}_l^{k+1}\},\{q_s^{k+1}\},\mathbf{x}^k,\{\boldsymbol{\lambda}_l^k\},\{\boldsymbol{\alpha}_l^k\},\{\boldsymbol{\beta}_l^k\},\{\gamma_s^k\}) \\
&-L_p(\{\boldsymbol{\mu}_l^{k+1}\},\{\mathbf{r}_l^{k+1}\},\{\mathbf{p}_l^{k+1}\},\{q_s^k\},\mathbf{x}^k,\{\boldsymbol{\lambda}_l^k\},\{\boldsymbol{\alpha}_l^k\},\{\boldsymbol{\beta}_l^k\},\{\gamma_s^k\}) \\
&\leq (-\frac{1}{\eta_{\mathbf{q}}^k}+\frac{w_{\mathbf{q}}L_w}{3}+w_{\gamma})\sum_{s=1}^{|\mathcal{P}^k|}||q_s^{k+1}-q_s^k||^2.
\end{aligned} \tag{62}$$

For the variable $\mathbf{x}$, define a constant ν such that $\nu \geq \sum_{s=1}^{|\mathcal{P}^k|}\|\mathbf{b}_s\|^2$. Consequently, we have $\eta_{\mathbf{x}} \leq \eta_{\mathbf{x}}^k$ for all k. Similarly to (56)-(59), we can obtain that,

$$\begin{aligned}
&L_p(\{\boldsymbol{\mu}_l^{k+1}\},\{\mathbf{r}_l^{k+1}\},\{\mathbf{p}_l^{k+1}\},\{q_s^{k+1}\},\mathbf{x}^{k+1},\{\boldsymbol{\lambda}_l^k\},\{\boldsymbol{\alpha}_l^k\},\{\boldsymbol{\beta}_l^k\},\{\gamma_s^k\}) \\
&-L_p(\{\boldsymbol{\mu}_l^{k+1}\},\{\mathbf{r}_l^{k+1}\},\{\mathbf{p}_l^{k+1}\},\{q_s^{k+1}\},\mathbf{x}^k,\{\boldsymbol{\lambda}_l^k\},\{\boldsymbol{\alpha}_l^k\},\{\boldsymbol{\beta}_l^k\},\{\gamma_s^k\}) \\
&\leq (-\frac{1}{\eta_{\mathbf{x}}^k}+\frac{L_w}{6})||\mathbf{x}^{k+1}-\mathbf{x}^k||^2.
\end{aligned} \tag{63}$$

Combining (55), (60), (61), (62) and (63), we conclude the proof of Lemma C.4. □

Lemma C.5. *$\forall k \geq K_1$, we have*

$$
\begin{aligned}
&L_p(\{\boldsymbol{\mu}_l^{k+1}\}, \{\mathbf{r}_l^{k+1}\}, \{\mathbf{p}_l^{k+1}\}, \{q_s^{k+1}\}, \mathbf{x}^{k+1}, \{\boldsymbol{\lambda}_l^{k+1}\}, \{\boldsymbol{\alpha}_l^{k+1}\}, \{\boldsymbol{\beta}_l^{k+1}\}, \{\gamma_s^{k+1}\}) \\
&- L_p(\{\boldsymbol{\mu}_l^{k}\}, \{\mathbf{r}_l^{k}\}, \{\mathbf{p}_l^{k}\}, \{q_s^{k}\}, \mathbf{x}^{k}, \{\boldsymbol{\lambda}_l^{k}\}, \{\boldsymbol{\alpha}_l^{k}\}, \{\boldsymbol{\beta}_l^{k}\}, \{\gamma_s^{k}\}) \\
&\leq (\frac{1}{a_1} + \frac{1}{a_2} + \frac{1}{a_3} + \frac{w_{\boldsymbol{\mu}} L_w M^{\frac{1}{2}}}{3} + 1 - \frac{1}{\eta_{\boldsymbol{\mu}}^k}) \sum_{l=1}^{L} \|\boldsymbol{\mu}_l^{k+1} - \boldsymbol{\mu}_l^k\|^2 \\
&+ (\frac{4w_{\mathbf{r}}^2}{a_1} + \frac{w_{\mathbf{r}} L_w M^{\frac{1}{2}}}{3} + w_{\boldsymbol{\lambda}} - \frac{1}{\eta_{\mathbf{r}}^k}) \sum_{l=1}^{L} \|\mathbf{r}_l^{k+1} - \mathbf{r}_l^k\|^2 \\
&+ (\frac{4w_{\mathbf{p}}^2}{a_2} + \frac{w_{\mathbf{p}} L_w M^{\frac{1}{2}}}{3} + w_{\boldsymbol{\alpha}} - \frac{1}{\eta_{\mathbf{p}}^k}) \sum_{l=1}^{L} \|\mathbf{p}_l^{k+1} - \mathbf{p}_l^k\|^2 \\
&+ (\frac{4w_q^2}{a_4} + \frac{w_q L_w}{3} + w_\gamma - \frac{1}{\eta_q^k}) \sum_{s=1}^{|\mathcal{P}^k|} \|q_s^{k+1} - q_s^k\|^2 \\
&+ (-\frac{1}{\eta_{\mathbf{x}}^k} + \frac{L_w}{6})||\mathbf{x}^{k+1} - \mathbf{x}^k||^2 + \frac{1}{a_3} \sum_{l=1}^{L} \|\mathbf{B}_l \mathbf{x}^{k+1} - \mathbf{B}_l \mathbf{x}^k\|^2 + \frac{1}{a_4} \sum_{s=1}^{|\mathcal{P}^k|} \|\mathbf{b}_s\|^2 \|\mathbf{x}^{k+1} - \mathbf{x}^k\|^2 \\
&+ (\frac{a_1}{2} - \frac{c_{\boldsymbol{\lambda}}^{k-1} - c_{\boldsymbol{\lambda}}^k}{2} + \frac{1}{2\eta_{\boldsymbol{\lambda}}}) \sum_{l=1}^{L} \|\boldsymbol{\lambda}_l^{k+1} - \boldsymbol{\lambda}_l^k\|^2 + \frac{c_{\boldsymbol{\lambda}}^{k-1}}{2} \sum_{l=1}^{L} (\|\boldsymbol{\lambda}_l^{k+1}\|^2 - \|\boldsymbol{\lambda}_l^k\|^2) \\
&+ (\frac{a_2}{2} - \frac{c_{\boldsymbol{\alpha}}^{k-1} - c_{\boldsymbol{\alpha}}^k}{2} + \frac{1}{2\eta_{\boldsymbol{\alpha}}}) \sum_{l=1}^{L} \|\boldsymbol{\alpha}_l^{k+1} - \boldsymbol{\alpha}_l^k\|^2 + \frac{c_{\boldsymbol{\alpha}}^{k-1}}{2} \sum_{l=1}^{L} (\|\boldsymbol{\alpha}_l^{k+1}\|^2 - \|\boldsymbol{\alpha}_l^k\|^2) \\
&+ (\frac{a_3}{2} - \frac{c_{\boldsymbol{\beta}}^{k-1} - c_{\boldsymbol{\beta}}^k}{2} + \frac{1}{2\eta_{\boldsymbol{\beta}}}) \sum_{l=1}^{L} \|\boldsymbol{\beta}_l^{k+1} - \boldsymbol{\beta}_l^k\|^2 + \frac{c_{\boldsymbol{\beta}}^{k-1}}{2} \sum_{l=1}^{L} (\|\boldsymbol{\beta}_l^{k+1}\|^2 - \|\boldsymbol{\beta}_l^k\|^2) \\
&+ (\frac{a_4}{2} - \frac{c_\gamma^{k-1} - c_\gamma^k}{2} + \frac{1}{2\eta_\gamma}) \sum_{s=1}^{|\mathcal{P}^k|} \|\gamma_s^{k+1} - \gamma_s^k\|^2 + \frac{c_\gamma^{k-1}}{2} \sum_{s=1}^{|\mathcal{P}^k|} (\|\gamma_s^{k+1}\|^2 - \|\gamma_s^k\|^2) \\
&+ \frac{1}{2\eta_{\boldsymbol{\lambda}}} \sum_{l=1}^{L} \|\boldsymbol{\lambda}_l^k - \boldsymbol{\lambda}_l^{k-1}\|^2 + \frac{1}{2\eta_{\boldsymbol{\alpha}}} \sum_{l=1}^{L} \|\boldsymbol{\alpha}_l^k - \boldsymbol{\alpha}_l^{k-1}\|^2 \\
&+ \frac{1}{2\eta_{\boldsymbol{\beta}}} \sum_{l=1}^{L} \|\boldsymbol{\beta}_l^k - \boldsymbol{\beta}_l^{k-1}\|^2 + \frac{1}{2\eta_\gamma} \sum_{s=1}^{|\mathcal{P}^k|} \|\gamma_s^k - \gamma_s^{k-1}\|^2,
\end{aligned} \tag{64}
$$

where a_1, a_2, a_3 and a_4 are positive constants.

Proof. According to (29), in the $(k+1)^{th}$ iteration, for $\forall \boldsymbol{\lambda}$ and $\forall l \in \mathcal{Q}^{k+1}$ it follows that,

$$
\left\langle \boldsymbol{\lambda}_l^{k+1} - \boldsymbol{\lambda}_l^k - \eta_{\boldsymbol{\lambda}} \nabla_{\boldsymbol{\lambda}_l} \widetilde{L}_p(\{\boldsymbol{\mu}_l^{k+1}\}, \{\mathbf{r}_l^{k+1}\}, \{\mathbf{p}_l^{k+1}\}, \{q_s^{k+1}\}, \mathbf{x}^{k+1}, \{\boldsymbol{\lambda}_l^k\}, \{\boldsymbol{\alpha}_l^k\}, \{\boldsymbol{\beta}_l^k\}, \{\gamma_s^k\}), \boldsymbol{\lambda} - \boldsymbol{\lambda}_l^{k+1} \right\rangle \geq 0. \tag{65}
$$

Let $\boldsymbol{\lambda} = \boldsymbol{\lambda}_l^k$, we can obtain,

$$
\left\langle \nabla_{\boldsymbol{\lambda}_l} \widetilde{L}_p(\{\boldsymbol{\mu}_l^{k+1}\}, \{\mathbf{r}_l^{k+1}\}, \{\mathbf{p}_l^{k+1}\}, \{q_s^{k+1}\}, \mathbf{x}^{k+1}, \{\boldsymbol{\lambda}_l^k\}, \{\boldsymbol{\alpha}_l^k\}, \{\boldsymbol{\beta}_l^k\}, \{\gamma_s^k\}) - \frac{1}{\eta_{\boldsymbol{\lambda}}} (\boldsymbol{\lambda}_l^{k+1} - \boldsymbol{\lambda}_l^k), \boldsymbol{\lambda}_l^k - \boldsymbol{\lambda}_l^{k+1} \right\rangle \leq 0. \tag{66}
$$

Likewise, in the kth iteration, we have that,

$$
\left\langle \nabla_{\boldsymbol{\lambda}_l} \widetilde{L}_p(\{\boldsymbol{\mu}_l^{k}\}, \{\mathbf{r}_l^{k}\}, \{\mathbf{p}_l^{k}\}, \{q_s^{k}\}, \mathbf{x}^{k}, \{\boldsymbol{\lambda}_l^{k-1}\}, \{\boldsymbol{\alpha}_l^{k-1}\}, \{\boldsymbol{\beta}_l^{k-1}\}, \{\gamma_s^{k-1}\}) - \frac{1}{\eta_{\boldsymbol{\lambda}}} (\boldsymbol{\lambda}_l^{k} - \boldsymbol{\lambda}_l^{k-1}), \boldsymbol{\lambda}_l^{k+1} - \boldsymbol{\lambda}_l^{k} \right\rangle \leq 0. \tag{67}
$$

Since $\boldsymbol{\lambda}_l^{k+1} - \boldsymbol{\lambda}_l^k = \mathbf{0}, l \notin \mathcal{Q}^{k+1}$, inequality (67) holds for l. It can be seen from (23) that $\widetilde{L}_p$ is concave with respect to $\boldsymbol{\lambda}_l$.

Therefore, we can obtain that,

$$\begin{aligned}
&\widetilde{L}_p(\{\boldsymbol{\mu}_l^{k+1}\},\{\mathbf{r}_l^{k+1}\},\{\mathbf{p}_l^{k+1}\},\{q_s^{k+1}\},\mathbf{x}^{k+1},\{\boldsymbol{\lambda}_l^{k+1}\},\{\boldsymbol{\alpha}_l^{k}\},\{\boldsymbol{\beta}_l^{k}\},\{\gamma_s^{k}\})\\
&-\widetilde{L}_p(\{\boldsymbol{\mu}_l^{k+1}\},\{\mathbf{r}_l^{k+1}\},\{\mathbf{p}_l^{k+1}\},\{q_s^{k+1}\},\mathbf{x}^{k+1},\{\boldsymbol{\lambda}_l^{k}\},\{\boldsymbol{\alpha}_l^{k}\},\{\boldsymbol{\beta}_l^{k}\},\{\gamma_s^{k}\})\\
&\leq \sum_{l=1}^{L}\Big\langle \nabla_{\boldsymbol{\lambda}_l}\widetilde{L}_p(\{\boldsymbol{\mu}_l^{k+1}\},\{\mathbf{r}_l^{k+1}\},\{\mathbf{p}_l^{k+1}\},\{q_s^{k+1}\},\mathbf{x}^{k+1},\{\boldsymbol{\lambda}_l^{k}\},\{\boldsymbol{\alpha}_l^{k}\},\{\boldsymbol{\beta}_l^{k}\},\{\gamma_s^{k}\}),\boldsymbol{\lambda}_l^{k+1}-\boldsymbol{\lambda}_l^{k}\Big\rangle\\
&\leq \sum_{l=1}^{L}\Big(\Big\langle \nabla_{\boldsymbol{\lambda}_l}\widetilde{L}_p(\{\boldsymbol{\mu}_l^{k+1}\},\{\mathbf{r}_l^{k+1}\},\{\mathbf{p}_l^{k+1}\},\{q_s^{k+1}\},\mathbf{x}^{k+1},\{\boldsymbol{\lambda}_l^{k}\},\{\boldsymbol{\alpha}_l^{k}\},\{\boldsymbol{\beta}_l^{k}\},\{\gamma_s^{k}\},\boldsymbol{\lambda}_l^{k+1}-\boldsymbol{\lambda}_l^{k}\Big\rangle\\
&+\frac{1}{\eta_{\boldsymbol{\lambda}}}\langle \boldsymbol{\lambda}_l^{k}-\boldsymbol{\lambda}_l^{k-1},\boldsymbol{\lambda}_l^{k+1}-\boldsymbol{\lambda}_l^{k}\rangle\Big)\\
&-\sum_{l=1}^{L}\Big(\Big\langle \nabla_{\boldsymbol{\lambda}_l}\widetilde{L}_p(\{\boldsymbol{\mu}_l^{k}\},\{\mathbf{r}_l^{k}\},\{\mathbf{p}_l^{k}\},\{q_s^{k}\},\mathbf{x}^{k},\{\boldsymbol{\lambda}_l^{k-1}\},\{\boldsymbol{\alpha}_l^{k-1}\},\{\boldsymbol{\beta}_l^{k-1}\},\{\gamma_s^{k-1}\},\boldsymbol{\lambda}_l^{k+1}-\boldsymbol{\lambda}_l^{k}\Big\rangle.
\end{aligned}\tag{68}$$

We have that,

$$\begin{aligned}
&\Big\langle \nabla_{\boldsymbol{\lambda}_l}\widetilde{L}_p(\{\boldsymbol{\mu}_l^{k+1}\},\{\mathbf{r}_l^{k+1}\},\{\mathbf{p}_l^{k+1}\},\{q_s^{k+1}\},\mathbf{x}^{k+1},\{\boldsymbol{\lambda}_l^{k}\},\{\boldsymbol{\alpha}_l^{k}\},\{\boldsymbol{\beta}_l^{k}\},\{\gamma_s^{k}\},\boldsymbol{\lambda}_l^{k+1}-\boldsymbol{\lambda}_l^{k}\Big\rangle\\
&-\Big\langle \nabla_{\boldsymbol{\lambda}_l}\widetilde{L}_p(\{\boldsymbol{\mu}_l^{k}\},\{\mathbf{r}_l^{k}\},\{\mathbf{p}_l^{k}\},\{q_s^{k}\},\mathbf{x}^{k},\{\boldsymbol{\lambda}_l^{k-1}\},\{\boldsymbol{\alpha}_l^{k-1}\},\{\boldsymbol{\beta}_l^{k-1}\},\{\gamma_s^{k-1}\},\boldsymbol{\lambda}_l^{k+1}-\boldsymbol{\lambda}_l^{k}\Big\rangle\\
&=\Big\langle \nabla_{\boldsymbol{\lambda}_l}\widetilde{L}_p(\{\boldsymbol{\mu}_l^{k+1}\},\{\mathbf{r}_l^{k+1}\},\{\mathbf{p}_l^{k+1}\},\{q_s^{k+1}\},\mathbf{x}^{k+1},\{\boldsymbol{\lambda}_l^{k}\},\{\boldsymbol{\alpha}_l^{k}\},\{\boldsymbol{\beta}_l^{k}\},\{\gamma_s^{k}\},\boldsymbol{\lambda}_l^{k+1}-\boldsymbol{\lambda}_l^{k}\Big\rangle\\
&-\Big\langle \nabla_{\boldsymbol{\lambda}_l}\widetilde{L}_p(\{\boldsymbol{\mu}_l^{k}\},\{\mathbf{r}_l^{k}\},\{\mathbf{p}_l^{k}\},\{q_s^{k}\},\mathbf{x}^{k},\{\boldsymbol{\lambda}_l^{k}\},\{\boldsymbol{\alpha}_l^{k}\},\{\boldsymbol{\beta}_l^{k}\},\{\gamma_s^{k}\},\boldsymbol{\lambda}_l^{k+1}-\boldsymbol{\lambda}_l^{k}\Big\rangle\\
&+\Big\langle \nabla_{\boldsymbol{\lambda}_l}\widetilde{L}_p(\{\boldsymbol{\mu}_l^{k}\},\{\mathbf{r}_l^{k}\},\{\mathbf{p}_l^{k}\},\{q_s^{k}\},\mathbf{x}^{k},\{\boldsymbol{\lambda}_l^{k}\},\{\boldsymbol{\alpha}_l^{k}\},\{\boldsymbol{\beta}_l^{k}\},\{\gamma_s^{k}\},\boldsymbol{\lambda}_l^{k+1}-\boldsymbol{\lambda}_l^{k}-(\boldsymbol{\lambda}_l^{k}-\boldsymbol{\lambda}_l^{k-1})\Big\rangle\\
&-\Big\langle \nabla_{\boldsymbol{\lambda}_l}\widetilde{L}_p(\{\boldsymbol{\mu}_l^{k}\},\{\mathbf{r}_l^{k}\},\{\mathbf{p}_l^{k}\},\{q_s^{k}\},\mathbf{x}^{k},\{\boldsymbol{\lambda}_l^{k-1}\},\{\boldsymbol{\alpha}_l^{k-1}\},\{\boldsymbol{\beta}_l^{k-1}\},\{\gamma_s^{k-1}\},\boldsymbol{\lambda}_l^{k+1}-\boldsymbol{\lambda}_l^{k}-(\boldsymbol{\lambda}_l^{k}-\boldsymbol{\lambda}_l^{k-1})\Big\rangle\\
&+\Big\langle \nabla_{\boldsymbol{\lambda}_l}\widetilde{L}_p(\{\boldsymbol{\mu}_l^{k}\},\{\mathbf{r}_l^{k}\},\{\mathbf{p}_l^{k}\},\{q_s^{k}\},\mathbf{x}^{k},\{\boldsymbol{\lambda}_l^{k}\},\{\boldsymbol{\alpha}_l^{k}\},\{\boldsymbol{\beta}_l^{k}\},\{\gamma_s^{k}\},\boldsymbol{\lambda}_l^{k}-\boldsymbol{\lambda}_l^{k-1}\Big\rangle\\
&-\Big\langle \nabla_{\boldsymbol{\lambda}_l}\widetilde{L}_p(\{\boldsymbol{\mu}_l^{k}\},\{\mathbf{r}_l^{k}\},\{\mathbf{p}_l^{k}\},\{q_s^{k}\},\mathbf{x}^{k},\{\boldsymbol{\lambda}_l^{k-1}\},\{\boldsymbol{\alpha}_l^{k-1}\},\{\boldsymbol{\beta}_l^{k-1}\},\{\gamma_s^{k-1}\},\boldsymbol{\lambda}_l^{k}-\boldsymbol{\lambda}_l^{k-1}\Big\rangle.
\end{aligned}\tag{69}$$

According to the definition of $\widetilde{L}_p$ provided in (23) and Cauchy-Schwarz inequality, we can obtain that,

$$\begin{aligned}
&\Big\langle \nabla_{\boldsymbol{\lambda}_l}\widetilde{L}_p(\{\boldsymbol{\mu}_l^{k+1}\},\{\mathbf{r}_l^{k+1}\},\{\mathbf{p}_l^{k+1}\},\{q_s^{k+1}\},\mathbf{x}^{k+1},\{\boldsymbol{\lambda}_l^{k}\},\{\boldsymbol{\alpha}_l^{k}\},\{\boldsymbol{\beta}_l^{k}\},\{\gamma_s^{k}\},\boldsymbol{\lambda}_l^{k+1}-\boldsymbol{\lambda}_l^{k}\Big\rangle\\
&-\Big\langle \nabla_{\boldsymbol{\lambda}_l}\widetilde{L}_p(\{\boldsymbol{\mu}_l^{k}\},\{\mathbf{r}_l^{k}\},\{\mathbf{p}_l^{k}\},\{q_s^{k}\},\mathbf{x}^{k},\{\boldsymbol{\lambda}_l^{k}\},\{\boldsymbol{\alpha}_l^{k}\},\{\boldsymbol{\beta}_l^{k}\},\{\gamma_s^{k}\},\boldsymbol{\lambda}_l^{k+1}-\boldsymbol{\lambda}_l^{k}\Big\rangle\\
&=\Big\langle \nabla_{\boldsymbol{\lambda}_l}L_p(\{\boldsymbol{\mu}_l^{k+1}\},\{\mathbf{r}_l^{k+1}\},\{\mathbf{p}_l^{k+1}\},\{q_s^{k+1}\},\mathbf{x}^{k+1},\{\boldsymbol{\lambda}_l^{k}\},\{\boldsymbol{\alpha}_l^{k}\},\{\boldsymbol{\beta}_l^{k}\},\{\gamma_s^{k}\},\boldsymbol{\lambda}_l^{k+1}-\boldsymbol{\lambda}_l^{k}\Big\rangle\\
&-\Big\langle \nabla_{\boldsymbol{\lambda}_l}L_p(\{\boldsymbol{\mu}_l^{k}\},\{\mathbf{r}_l^{k}\},\{\mathbf{p}_l^{k}\},\{q_s^{k}\},\mathbf{x}^{k},\{\boldsymbol{\lambda}_l^{k}\},\{\boldsymbol{\alpha}_l^{k}\},\{\boldsymbol{\beta}_l^{k}\},\{\gamma_s^{k}\},\boldsymbol{\lambda}_l^{k+1}-\boldsymbol{\lambda}_l^{k}\Big\rangle\\
&+\frac{c_{\boldsymbol{\lambda}}^{k-1}-c_{\boldsymbol{\lambda}}^{k}}{2}(||\lambda_l^{k+1}||^2-||\boldsymbol{\lambda}_l^{k}||^2)-\frac{c_{\boldsymbol{\lambda}}^{k-1}-c_{\boldsymbol{\lambda}}^{k}}{2}||\boldsymbol{\lambda}_l^{k+1}-\boldsymbol{\lambda}_l^{k}||^2\\
&\leq \frac{a_1}{2}\|\boldsymbol{\lambda}_l^{k+1}-\boldsymbol{\lambda}_l^{k}\|+\frac{1}{a_1}\|\boldsymbol{\mu}_l^{k+1}-\boldsymbol{\mu}_l^{k}\|^2+\frac{1}{a_1}\|\mathbf{r}_l^{k+1}\circ\mathbf{r}_l^{k+1}-\mathbf{r}_l^{k}\circ\mathbf{r}_l^{k}\|^2+\frac{c_{\boldsymbol{\lambda}}^{k-1}-c_{\boldsymbol{\lambda}}^{k}}{2}(||\lambda_l^{k+1}||^2-||\boldsymbol{\lambda}_l^{k}||^2)\\
&-\frac{c_{\boldsymbol{\lambda}}^{k-1}-c_{\boldsymbol{\lambda}}^{k}}{2}||\boldsymbol{\lambda}_l^{k+1}-\boldsymbol{\lambda}_l^{k}||^2,
\end{aligned}\tag{70}$$

where $a_1>0$ is a constant. From the definition of $\widetilde{L}_p$, with Cauchy-Schwarz inequality, we also have that,

$$\begin{aligned}
&\Big\langle \nabla_{\boldsymbol{\lambda}_l}\widetilde{L}_p(\{\boldsymbol{\mu}_l^{k}\},\{\mathbf{r}_l^{k}\},\{\mathbf{p}_l^{k}\},\{q_s^{k}\},\mathbf{x}^{k},\{\boldsymbol{\lambda}_l^{k}\},\{\boldsymbol{\alpha}_l^{k}\},\{\boldsymbol{\beta}_l^{k}\},\{\gamma_s^{k}\},\boldsymbol{\lambda}_l^{k+1}-\boldsymbol{\lambda}_l^{k}-(\boldsymbol{\lambda}_l^{k}-\boldsymbol{\lambda}_l^{k-1})\Big\rangle\\
&-\Big\langle \nabla_{\boldsymbol{\lambda}_l}\widetilde{L}_p(\{\boldsymbol{\mu}_l^{k}\},\{\mathbf{r}_l^{k}\},\{\mathbf{p}_l^{k}\},\{q_s^{k}\},\mathbf{x}^{k},\{\boldsymbol{\lambda}_l^{k-1}\},\{\boldsymbol{\alpha}_l^{k-1}\},\{\boldsymbol{\beta}_l^{k-1}\},\{\gamma_s^{k-1}\},\boldsymbol{\lambda}_l^{k+1}-\boldsymbol{\lambda}_l^{k}-(\boldsymbol{\lambda}_l^{k}-\boldsymbol{\lambda}_l^{k-1})\Big\rangle\\
&\leq \frac{1}{2\eta_{\boldsymbol{\lambda}}}\|\boldsymbol{\lambda}_l^{k+1}-\boldsymbol{\lambda}_l^{k}-(\boldsymbol{\lambda}_l^{k}-\boldsymbol{\lambda}_l^{k-1})\|^2+\frac{\eta_{\boldsymbol{\lambda}}}{2}\|c_{\boldsymbol{\lambda}}^{k-1}(\boldsymbol{\lambda}_l^{k}-\boldsymbol{\lambda}_l^{k-1})\|^2.
\end{aligned}\tag{71}$$

Following (Xu et al., 2023b), since $\widetilde{L}_p$ is strong concave with respect to $\boldsymbol{\lambda}$, we have

$$\begin{aligned}&\left\langle \nabla_{\boldsymbol{\lambda}_l}\widetilde{L}_p(\{\boldsymbol{\mu}_l^k\},\{\mathbf{r}_l^k\},\{\mathbf{p}_l^k\},\{q_s^k\},\mathbf{x}^k,\{\boldsymbol{\lambda}_l^k\},\{\boldsymbol{\alpha}_l^k\},\{\boldsymbol{\beta}_l^k\},\{\gamma_s^k\},\boldsymbol{\lambda}_l^k-\boldsymbol{\lambda}_l^{k-1}\right\rangle\\&-\left\langle \nabla_{\boldsymbol{\lambda}_l}\widetilde{L}_p(\{\boldsymbol{\mu}_l^k\},\{\mathbf{r}_l^k\},\{\mathbf{p}_l^k\},\{q_s^k\},\mathbf{x}^k,\{\boldsymbol{\lambda}_l^{k-1}\},\{\boldsymbol{\alpha}_l^{k-1}\},\{\boldsymbol{\beta}_l^{k-1}\},\{\gamma_s^{k-1}\},\boldsymbol{\lambda}_l^k-\boldsymbol{\lambda}_l^{k-1}\right\rangle\\&\leq-\frac{1}{2c_{\boldsymbol{\lambda}}^{k-1}}\|c_{\boldsymbol{\lambda}}^{k-1}(\boldsymbol{\lambda}_l^k-\boldsymbol{\lambda}_l^{k-1})\|^2-\frac{c_{\boldsymbol{\lambda}}^{k-1}}{2}\|\boldsymbol{\lambda}_l^k-\boldsymbol{\lambda}_l^{k-1}\|^2.\end{aligned}\tag{72}$$

Since $\frac{\eta_{\boldsymbol{\lambda}}}{2}\leq\frac{1}{2c_{\boldsymbol{\lambda}}^{k-1}}$, by employing (69)-(72), we have that,

$$\begin{aligned}&\left\langle \nabla_{\boldsymbol{\lambda}_l}\widetilde{L}_p(\{\boldsymbol{\mu}_l^{k+1}\},\{\mathbf{r}_l^{k+1}\},\{\mathbf{p}_l^{k+1}\},\{q_s^{k+1}\},\mathbf{x}^{k+1},\{\boldsymbol{\lambda}_l^k\},\{\boldsymbol{\alpha}_l^k\},\{\boldsymbol{\beta}_l^k\},\{\gamma_s^k\},\boldsymbol{\lambda}_l^{k+1}-\boldsymbol{\lambda}_l^k\right\rangle\\&-\left\langle \nabla_{\boldsymbol{\lambda}_l}\widetilde{L}_p(\{\boldsymbol{\mu}_l^k\},\{\mathbf{r}_l^k\},\{\mathbf{p}_l^k\},\{q_s^k\},\mathbf{x}^k,\{\boldsymbol{\lambda}_l^{k-1}\},\{\boldsymbol{\alpha}_l^{k-1}\},\{\boldsymbol{\beta}_l^{k-1}\},\{\gamma_s^{k-1}\},\boldsymbol{\lambda}_l^{k+1}-\boldsymbol{\lambda}_l^k\right\rangle\\&\leq\frac{a_1}{2}\|\boldsymbol{\lambda}_l^{k+1}-\boldsymbol{\lambda}_l^k\|^2+\frac{1}{a_1}\|\boldsymbol{\mu}_l^{k+1}-\boldsymbol{\mu}_l^k\|^2+\frac{1}{a_1}\|\mathbf{r}_l^{k+1}\circ\mathbf{r}_l^{k+1}-\mathbf{r}_l^k\circ\mathbf{r}_l^k\|^2+\frac{c_{\boldsymbol{\lambda}}^{k-1}-c_{\boldsymbol{\lambda}}^k}{2}(\|\boldsymbol{\lambda}_l^{k+1}\|^2-\|\boldsymbol{\lambda}_l^k\|^2)\\&-\frac{c_{\boldsymbol{\lambda}}^{k-1}-c_{\boldsymbol{\lambda}}^k}{2}\|\boldsymbol{\lambda}_l^{k+1}-\boldsymbol{\lambda}_l^k\|^2+\frac{1}{2\eta_{\boldsymbol{\lambda}}}\|\boldsymbol{\lambda}_l^{k+1}-\boldsymbol{\lambda}_l^k-(\boldsymbol{\lambda}_l^k-\boldsymbol{\lambda}_l^{k-1})\|^2-\frac{c_{\boldsymbol{\lambda}}^{k-1}}{2}\|\boldsymbol{\lambda}_l^k-\boldsymbol{\lambda}_l^{k-1}\|^2.\end{aligned}\tag{73}$$

It can be seen from Assumption 6.3 that,

$$\frac{1}{a_1}\|\mathbf{r}_l^{k+1}\circ\mathbf{r}_l^{k+1}-\mathbf{r}_l^k\circ\mathbf{r}_l^k\|^2\leq\frac{4w_r^2}{a_1}\|\mathbf{r}_l^{k+1}-\mathbf{r}_l^k\|^2.\tag{74}$$

In addition, the following equality can be obtained,

$$\frac{1}{\eta_{\boldsymbol{\lambda}}}\left\langle\boldsymbol{\lambda}_l^k-\boldsymbol{\lambda}_l^{k-1},\boldsymbol{\lambda}_l^{k+1}-\boldsymbol{\lambda}_l^k\right\rangle=\frac{1}{2\eta_{\boldsymbol{\lambda}}}\|\boldsymbol{\lambda}_l^{k+1}-\boldsymbol{\lambda}_l^k\|^2+\frac{1}{2\eta_{\boldsymbol{\lambda}}}\|\boldsymbol{\lambda}_l^k-\boldsymbol{\lambda}_l^{k-1}\|^2-\frac{1}{2\eta_{\boldsymbol{\lambda}}}\|\boldsymbol{\lambda}_l^{k+1}-\boldsymbol{\lambda}_l^k-(\boldsymbol{\lambda}_l^k-\boldsymbol{\lambda}_l^{k-1})\|^2.\tag{75}$$

Combining (68), (73), (74) with (75), we can obtain,

$$\begin{aligned}&L_p(\{\boldsymbol{\mu}_l^{k+1}\},\{\mathbf{r}_l^{k+1}\},\{\mathbf{p}_l^{k+1}\},\{q_s^{k+1}\},\mathbf{x}^{k+1},\{\boldsymbol{\lambda}_l^{k+1}\},\{\boldsymbol{\alpha}_l^k\},\{\boldsymbol{\beta}_l^k\},\{\gamma_s^k\})\\&-L_p(\{\boldsymbol{\mu}_l^{k+1}\},\{\mathbf{r}_l^{k+1}\},\{\mathbf{p}_l^{k+1}\},\{q_s^{k+1}\},\mathbf{x}^{k+1},\{\boldsymbol{\lambda}_l^k\},\{\boldsymbol{\alpha}_l^k\},\{\boldsymbol{\beta}_l^k\},\{\gamma_s^k\})\\&\leq\frac{1}{a_1}\sum_{l=1}^L\|\boldsymbol{\mu}_l^{k+1}-\boldsymbol{\mu}_l^k\|^2+\frac{4w_r^2}{a_1}\sum_{l=1}^L\|\mathbf{r}_l^{k+1}-\mathbf{r}_l^k\|^2+(\frac{a_1}{2}-\frac{c_{\boldsymbol{\lambda}}^{k-1}-c_{\boldsymbol{\lambda}}^k}{2}+\frac{1}{2\eta_{\boldsymbol{\lambda}}})\sum_{l=1}^L\|\boldsymbol{\lambda}_l^{k+1}-\boldsymbol{\lambda}_l^k\|^2\\&+\frac{c_{\boldsymbol{\lambda}}^{k-1}}{2}\sum_{l=1}^L(\|\boldsymbol{\lambda}_l^{k+1}\|^2-\|\boldsymbol{\lambda}_l^k\|^2)+\frac{1}{2\eta_{\boldsymbol{\lambda}}}\sum_{l=1}^L\|\boldsymbol{\lambda}_l^k-\boldsymbol{\lambda}_l^{k-1}\|^2.\end{aligned}\tag{76}$$

Likewise, similar results can be obtained for other variables :

$$\begin{aligned}&L_p(\{\boldsymbol{\mu}_l^{k+1}\},\{\mathbf{r}_l^{k+1}\},\{\mathbf{p}_l^{k+1}\},\{q_s^{k+1}\},\mathbf{x}^{k+1},\{\boldsymbol{\lambda}_l^{k+1}\},\{\boldsymbol{\alpha}_l^{k+1}\},\{\boldsymbol{\beta}_l^k\},\{\gamma_s^k\})\\&-L_p(\{\boldsymbol{\mu}_l^{k+1}\},\{\mathbf{r}_l^{k+1}\},\{\mathbf{p}_l^{k+1}\},\{q_s^{k+1}\},\mathbf{x}^{k+1},\{\boldsymbol{\lambda}_l^{k+1}\},\{\boldsymbol{\alpha}_l^k\},\{\boldsymbol{\beta}_l^k\},\{\gamma_s^k\})\\&\leq\frac{1}{a_2}\sum_{l=1}^L\|\boldsymbol{\mu}_l^{k+1}-\boldsymbol{\mu}_l^k\|^2+\frac{4w_{\mathbf{p}}^2}{a_2}\sum_{l=1}^L\|\mathbf{p}_l^{k+1}-\mathbf{p}_l^k\|^2+(\frac{a_2}{2}-\frac{c_{\boldsymbol{\alpha}}^{k-1}-c_{\boldsymbol{\alpha}}^k}{2}+\frac{1}{2\eta_{\boldsymbol{\alpha}}})\sum_{l=1}^L\|\boldsymbol{\alpha}_l^{k+1}-\boldsymbol{\alpha}_l^k\|^2\\&+\frac{c_{\boldsymbol{\alpha}}^{k-1}}{2}\sum_{l=1}^L(\|\boldsymbol{\alpha}_l^{k+1}\|^2-\|\boldsymbol{\alpha}_l^k\|^2)+\frac{1}{2\eta_{\boldsymbol{\alpha}}}\sum_{l=1}^L\|\boldsymbol{\alpha}_l^k-\boldsymbol{\alpha}_l^{k-1}\|^2,\end{aligned}\tag{77}$$

$$\begin{aligned}&L_p(\{\boldsymbol{\mu}_l^{k+1}\},\{\mathbf{r}_l^{k+1}\},\{\mathbf{p}_l^{k+1}\},\{q_s^{k+1}\},\mathbf{x}^{k+1},\{\boldsymbol{\lambda}_l^{k+1}\},\{\boldsymbol{\alpha}_l^{k+1}\},\{\boldsymbol{\beta}_l^{k+1}\},\{\gamma_s^k\})\\&-L_p(\{\boldsymbol{\mu}_l^{k+1}\},\{\mathbf{r}_l^{k+1}\},\{\mathbf{p}_l^{k+1}\},\{q_s^{k+1}\},\mathbf{x}^{k+1},\{\boldsymbol{\lambda}_l^{k+1}\},\{\boldsymbol{\alpha}_l^{k+1}\},\{\boldsymbol{\beta}_l^k\},\{\gamma_s^k\})\\&\leq\frac{1}{a_3}\sum_{l=1}^L\|\boldsymbol{\mu}_l^{k+1}-\boldsymbol{\mu}_l^k\|^2+\frac{1}{a_3}\sum_{l=1}^L\|\mathbf{B}_l\mathbf{x}^{k+1}-\mathbf{B}_l\mathbf{x}^k\|^2+(\frac{a_3}{2}-\frac{c_{\boldsymbol{\beta}}^{k-1}-c_{\boldsymbol{\beta}}^k}{2}+\frac{1}{2\eta_{\boldsymbol{\beta}}})\sum_{l=1}^L\|\boldsymbol{\beta}_l^{k+1}-\boldsymbol{\beta}_l^k\|^2\\&+\frac{c_{\boldsymbol{\beta}}^{k-1}}{2}\sum_{l=1}^L(\|\boldsymbol{\beta}_l^{k+1}\|^2-\|\boldsymbol{\beta}_l^k\|^2)+\frac{1}{2\eta_{\boldsymbol{\beta}}}\sum_{l=1}^L\|\boldsymbol{\beta}_l^k-\boldsymbol{\beta}_l^{k-1}\|^2,\end{aligned}\tag{78}$$

$$
\begin{aligned}
&L_p(\{\boldsymbol{\mu}_l^{k+1}\},\{\mathbf{r}_l^{k+1}\},\{\mathbf{p}_l^{k+1}\},\{q_s^{k+1}\},\mathbf{x}^{k+1},\{\boldsymbol{\lambda}_l^{k+1}\},\{\boldsymbol{\alpha}_l^{k+1}\},\{\boldsymbol{\beta}_l^{k+1}\},\{\gamma_s^{k+1}\})\\
&-L_p(\{\boldsymbol{\mu}_l^{k+1}\},\{\mathbf{r}_l^{k+1}\},\{\mathbf{p}_l^{k+1}\},\{q_s^{k+1}\},\mathbf{x}^{k+1},\{\boldsymbol{\lambda}_l^{k+1}\},\{\boldsymbol{\alpha}_l^{k+1}\},\{\boldsymbol{\beta}_l^{k+1}\},\{\gamma_s^{k}\})\\
&\leq \frac{1}{a_4}\sum_{s=1}^{|\mathcal{P}^k|}\|\mathbf{b}_s\|^2\|\mathbf{x}^{k+1}-\mathbf{x}^k\|^2+\frac{4w_q^2}{a_4}\sum_{s=1}^{|\mathcal{P}^k|}\|q_s^{k+1}-q_s^k\|^2+(\frac{a_4}{2}-\frac{c_\gamma^{k-1}-c_\gamma^k}{2}+\frac{1}{2\eta_\gamma})\sum_{s=1}^{|\mathcal{P}^k|}\|\gamma_s^{k+1}-\gamma_s^k\|^2\\
&+\frac{c_\gamma^{k-1}}{2}\sum_{s=1}^{|\mathcal{P}^k|}(\|\gamma_s^{k+1}\|^2-\|\gamma_s^k\|^2)+\frac{1}{2\eta_\gamma}\sum_{s=1}^{|\mathcal{P}^k|}\|\gamma_s^k-\gamma_s^{k-1}\|^2,
\end{aligned}
\tag{79}
$$

where $a_2>0$, $a_3>0$ and $a_4>0$ are constant. Combining Lemma C.4 with (76), (77), (78) and (79), we can obtain that,

$$
\begin{aligned}
&L_p(\{\boldsymbol{\mu}_l^{k+1}\},\{\mathbf{r}_l^{k+1}\},\{\mathbf{p}_l^{k+1}\},\{q_s^{k+1}\},\mathbf{x}^{k+1},\{\boldsymbol{\lambda}_l^{k+1}\},\{\boldsymbol{\alpha}_l^{k+1}\},\{\boldsymbol{\beta}_l^{k+1}\},\{\gamma_s^{k+1}\})\\
&-L_p(\{\boldsymbol{\mu}_l^{k}\},\{\mathbf{r}_l^{k}\},\{\mathbf{p}_l^{k}\},\{q_s^{k}\},\mathbf{x}^{k},\{\boldsymbol{\lambda}_l^{k}\},\{\boldsymbol{\alpha}_l^{k}\},\{\boldsymbol{\beta}_l^{k}\},\{\gamma_s^{k}\})\\
&\leq(\frac{1}{a_1}+\frac{1}{a_2}+\frac{1}{a_3}+\frac{w_{\boldsymbol{\mu}}L_wM^{\frac{1}{2}}}{3}+1-\frac{1}{\eta_{\boldsymbol{\mu}}^k})\sum_{l=1}^{L}\|\boldsymbol{\mu}_l^{k+1}-\boldsymbol{\mu}_l^k\|^2+(\frac{4w_{\mathbf{r}}^2}{a_1}+\frac{w_{\mathbf{r}}L_wM^{\frac{1}{2}}}{3}+w_{\boldsymbol{\lambda}}-\frac{1}{\eta_{\mathbf{r}}^k})\sum_{l=1}^{L}\|\mathbf{r}_l^{k+1}-\mathbf{r}_l^k\|^2\\
&+(\frac{4w_{\mathbf{p}}^2}{a_2}+\frac{w_{\mathbf{p}}L_wM^{\frac{1}{2}}}{3}+w_{\boldsymbol{\alpha}}-\frac{1}{\eta_{\mathbf{p}}^k})\sum_{l=1}^{L}\|\mathbf{p}_l^{k+1}-\mathbf{p}_l^k\|^2+(\frac{4w_q^2}{a_4}+\frac{w_qL_w}{3}+w_\gamma-\frac{1}{\eta_q^k})\sum_{s=1}^{|\mathcal{P}^k|}\|q_s^{k+1}-q_s^k\|^2\\
&+(-\frac{1}{\eta_{\mathbf{x}}^k}+\frac{L_w}{6})||\mathbf{x}^{k+1}-\mathbf{x}^k||^2+\frac{1}{a_3}\sum_{l=1}^{L}\|\mathbf{B}_l\mathbf{x}^{k+1}-\mathbf{B}_l\mathbf{x}^k\|^2+\frac{1}{a_4}\sum_{s=1}^{|\mathcal{P}^k|}\|\mathbf{b}_s\|^2\|\mathbf{x}^{k+1}-\mathbf{x}^k\|^2\\
&+(\frac{a_1}{2}-\frac{c_{\boldsymbol{\lambda}}^{k-1}-c_{\boldsymbol{\lambda}}^k}{2}+\frac{1}{2\eta_{\boldsymbol{\lambda}}})\sum_{l=1}^{L}\|\boldsymbol{\lambda}_l^{k+1}-\boldsymbol{\lambda}_l^k\|^2+\frac{c_{\boldsymbol{\lambda}}^{k-1}}{2}\sum_{l=1}^{L}(\|\boldsymbol{\lambda}_l^{k+1}\|^2-\|\boldsymbol{\lambda}_l^k\|^2)+\frac{1}{2\eta_{\boldsymbol{\lambda}}}\sum_{l=1}^{L}\|\boldsymbol{\lambda}_l^k-\boldsymbol{\lambda}_l^{k-1}\|^2\\
&+(\frac{a_2}{2}-\frac{c_{\boldsymbol{\alpha}}^{k-1}-c_{\boldsymbol{\alpha}}^k}{2}+\frac{1}{2\eta_{\boldsymbol{\alpha}}})\sum_{l=1}^{L}\|\boldsymbol{\alpha}_l^{k+1}-\boldsymbol{\alpha}_l^k\|^2+\frac{c_{\boldsymbol{\alpha}}^{k-1}}{2}\sum_{l=1}^{L}(\|\boldsymbol{\alpha}_l^{k+1}\|^2-\|\boldsymbol{\alpha}_l^k\|^2)+\frac{1}{2\eta_{\boldsymbol{\alpha}}}\sum_{l=1}^{L}\|\boldsymbol{\alpha}_l^k-\boldsymbol{\alpha}_l^{k-1}\|^2\\
&+(\frac{a_3}{2}-\frac{c_{\boldsymbol{\beta}}^{k-1}-c_{\boldsymbol{\beta}}^k}{2}+\frac{1}{2\eta_{\boldsymbol{\beta}}})\sum_{l=1}^{L}\|\boldsymbol{\beta}_l^{k+1}-\boldsymbol{\beta}_l^k\|^2+\frac{c_{\boldsymbol{\beta}}^{k-1}}{2}\sum_{l=1}^{L}(\|\boldsymbol{\beta}_l^{k+1}\|^2-\|\boldsymbol{\beta}_l^k\|^2)+\frac{1}{2\eta_{\boldsymbol{\beta}}}\sum_{l=1}^{L}\|\boldsymbol{\beta}_l^k-\boldsymbol{\beta}_l^{k-1}\|^2\\
&+(\frac{a_4}{2}-\frac{c_\gamma^{k-1}-c_\gamma^k}{2}+\frac{1}{2\eta_\gamma})\sum_{s=1}^{|\mathcal{P}^k|}\|\gamma_s^{k+1}-\gamma_s^k\|^2+\frac{c_\gamma^{k-1}}{2}\sum_{s=1}^{|\mathcal{P}^k|}(\|\gamma_s^{k+1}\|^2-\|\gamma_s^k\|^2)+\frac{1}{2\eta_\gamma}\sum_{s=1}^{|\mathcal{P}^k|}\|\gamma_s^k-\gamma_s^{k-1}\|^2,
\end{aligned}
\tag{80}
$$

which concludes the proof of Lemma C.5. □

Lemma C.6. *Define S_1^{k+1}, S_2^{k+1}, S_3^{k+1}, S_4^{k+1}, F^{k+1} and a_5 as :*

$$
\begin{aligned}
S_1^{k+1}&=\frac{4}{\eta_{\boldsymbol{\lambda}}^2c_{\boldsymbol{\lambda}}^{k+1}}\sum_{l=1}^{L}||\lambda_l^{k+1}-\lambda_l^k||^2-\frac{4}{\eta_{\boldsymbol{\lambda}}}(\frac{c_{\boldsymbol{\lambda}}^{k-1}}{c_{\boldsymbol{\lambda}}^k}-1)\sum_{l=1}^{L}||\boldsymbol{\lambda}_l^{k+1}||^2,\\
S_2^{k+1}&=\frac{4}{\eta_{\boldsymbol{\alpha}^2}c_{\boldsymbol{\alpha}}^{k+1}}\sum_{l=1}^{L}||\alpha^{k+1}-\alpha^k||^2-\frac{4}{\eta_{\boldsymbol{\alpha}}}(\frac{c_{\boldsymbol{\alpha}}^{k-1}}{c_{\boldsymbol{\alpha}}^k}-1)\sum_{l=1}^{L}||\boldsymbol{\alpha}_l^{k+1}||^2,\\
S_3^{k+1}&=\frac{4}{\eta_{\boldsymbol{\beta}}^2c_{\boldsymbol{\beta}}^{k+1}}\sum_{l=1}^{L}||\beta^{k+1}-\beta^k||^2-\frac{4}{\eta_{\boldsymbol{\beta}}}(\frac{c_{\boldsymbol{\beta}}^{k-1}}{c_{\boldsymbol{\beta}}^k}-1)\sum_{l=1}^{L}||\boldsymbol{\beta}_l^{k+1}||^2,\\
S_4^{k+1}&=\frac{4}{\eta_\gamma^2c_\gamma^{k+1}}\sum_{s=1}^{|\mathcal{P}^k|}||\gamma_s^{k+1}-\gamma_s^k||^2-\frac{4}{\eta_\gamma}(\frac{c_\gamma^{k-1}}{c_\gamma^k}-1)\sum_{l=1}^{L}||\gamma_s^{k+1}||^2.
\end{aligned}
\tag{81}
$$

$$
\begin{aligned}
F^{k+1} = \quad & L_p(\{\boldsymbol{\mu}_l^{k+1}\}, \{\mathbf{r}_l^{k+1}\}, \{\mathbf{p}_l^{k+1}\}, \{q_s^{k+1}\}, \mathbf{x}^{k+1}, \{\boldsymbol{\lambda}_l^{k+1}\}, \{\boldsymbol{\alpha}_l^{k+1}\}, \{\boldsymbol{\beta}_l^{k+1}\}, \{\gamma_s^{k+1}\}) \\
& + S_1^{k+1} + S_2^{k+1} + S_3^{k+1} + S_4^{k+1} \\
& - \frac{7}{2\eta_{\boldsymbol{\lambda}}} \sum_{l=1}^{L} ||\boldsymbol{\lambda}_l^{k+1} - \boldsymbol{\lambda}_l^k||^2 - \frac{c_{\boldsymbol{\lambda}}^k}{2} \sum_{l=1}^{L} ||\boldsymbol{\lambda}_l^{k+1}||^2 \\
& - \frac{7}{2\eta_{\boldsymbol{\alpha}}} \sum_{l=1}^{L} ||\boldsymbol{\alpha}_l^{k+1} - \boldsymbol{\alpha}_l^k||^2 - \frac{c_{\boldsymbol{\alpha}}^k}{2} \sum_{l=1}^{L} ||\boldsymbol{\alpha}_l^{k+1}||^2 \\
& - \frac{7}{2\eta_{\boldsymbol{\beta}}} \sum_{l=1}^{L} ||\boldsymbol{\beta}_l^{k+1} - \boldsymbol{\beta}_l^k||^2 - \frac{c_{\boldsymbol{\beta}}^k}{2} \sum_{l=1}^{L} ||\boldsymbol{\beta}_l^{k+1}||^2 \\
& - \frac{7}{2\eta_{\gamma}} \sum_{s=1}^{|\mathcal{P}^k|} ||\gamma_s^{k+1} - \gamma_s^k||^2 - \frac{c_{\gamma}^k}{2} \sum_{s=1}^{|\mathcal{P}^k|} ||\gamma_s^{k+1}||^2.
\end{aligned} \tag{82}
$$

$$
\begin{aligned}
a_5 = \max \Bigg\{ & \eta_{\boldsymbol{\lambda}} + \eta_{\boldsymbol{\alpha}} + \eta_{\boldsymbol{\beta}} + \frac{w_{\boldsymbol{\mu}} L_w M^{\frac{1}{2}}}{3} + 1, 4w_{\mathbf{r}}^2 \eta_{\boldsymbol{\lambda}} + \frac{w_{\mathbf{r}} L_w M^{\frac{1}{2}}}{3} + w_{\boldsymbol{\lambda}}, \\
& 4w_{\mathbf{p}}^2 \eta_{\boldsymbol{\alpha}} + \frac{w_{\mathbf{p}} L_w M^{\frac{1}{2}}}{3} + w_{\boldsymbol{\alpha}}, 4w_q^2 \eta_{\gamma} + \frac{w_q L_w}{3} + w_{\gamma}, \eta_{\boldsymbol{\beta}} + \frac{L_w}{6} + \eta_{\gamma} \sum_{s=1}^{|\mathcal{P}^k|} \|\mathbf{b}_s\|^2 \Bigg\}.
\end{aligned} \tag{83}
$$

And then $\forall k \geq K_1$, we have that,

$$
\begin{aligned}
F^{k+1} - F^k \leq & (a_5 - \frac{1}{\eta_{\boldsymbol{\mu}}^k} + \frac{16}{\eta_{\boldsymbol{\lambda}}(c_{\boldsymbol{\lambda}}^k)^2} + \frac{16}{\eta_{\boldsymbol{\alpha}}(c_{\boldsymbol{\alpha}}^k)^2} + \frac{16}{\eta_{\boldsymbol{\beta}}(c_{\boldsymbol{\beta}}^k)^2}) \sum_{l=1}^{L} \|\boldsymbol{\mu}_l^{k+1} - \boldsymbol{\mu}_l^k\|^2 \\
& + (a_5 - \frac{1}{\eta_{\mathbf{r}}^k} + \frac{64w_{\mathbf{r}}^2}{\eta_{\boldsymbol{\lambda}}(c_{\boldsymbol{\lambda}}^k)^2}) \sum_{l=1}^{L} \|\mathbf{r}_l^{k+1} - \mathbf{r}_l^k\|^2 \\
& + (a_5 - \frac{1}{\eta_{\mathbf{p}}^k} + \frac{64w_{\mathbf{p}}^2}{\eta_{\boldsymbol{\alpha}}(c_{\boldsymbol{\alpha}}^k)^2}) \sum_{l=1}^{L} \|\mathbf{p}_l^{k+1} - \mathbf{p}_l^k\|^2 \\
& + (a_5 - \frac{1}{\eta_q^k} + \frac{64w_q^2}{\eta_{\gamma}(c_{\gamma}^k)^2}) \sum_{s=1}^{|\mathcal{P}^k|} \|q_s^{k+1} - q_s^k\|^2 \\
& + (a_5 - \frac{1}{\eta_{\mathbf{x}}^k} + \frac{16}{\eta_{\boldsymbol{\beta}}(c_{\boldsymbol{\beta}}^k)^2} + \frac{16}{\eta_{\gamma}(c_{\gamma}^k)^2} \sum_{s=1}^{|\mathcal{P}^k|} \|b_s\|^2) \|\mathbf{x}^{k+1} - \mathbf{x}^k\|^2 \\
& - \frac{1}{10\eta_{\boldsymbol{\lambda}}} \sum_{l=1}^{L} ||\boldsymbol{\lambda}_l^{k+1} - \boldsymbol{\lambda}_l^k||^2 + \frac{c_{\boldsymbol{\lambda}}^{k-1} - c_{\boldsymbol{\lambda}}^k}{2} \sum_{l=1}^{L} ||\boldsymbol{\lambda}_l^{k+1}||^2 \\
& - \frac{1}{10\eta_{\boldsymbol{\alpha}}} \sum_{l=1}^{L} ||\boldsymbol{\alpha}_l^{k+1} - \boldsymbol{\alpha}_l^k||^2 + \frac{c_{\boldsymbol{\alpha}}^{k-1} - c_{\boldsymbol{\alpha}}^k}{2} \sum_{l=1}^{L} ||\boldsymbol{\alpha}_l^{k+1}||^2 \\
& - \frac{1}{10\eta_{\boldsymbol{\beta}}} \sum_{l=1}^{L} ||\boldsymbol{\beta}_l^{k+1} - \boldsymbol{\beta}_l^k||^2 + \frac{c_{\boldsymbol{\beta}}^{k-1} - c_{\boldsymbol{\beta}}^k}{2} \sum_{l=1}^{L} ||\boldsymbol{\beta}_l^{k+1}||^2 \\
& - \frac{1}{10\eta_{\gamma}} \sum\nolimits_{s=1}^{|\mathcal{P}^k|} ||\gamma_s^{k+1} - \gamma_s^k||^2 + \frac{c_{\gamma}^{k-1} - c_{\gamma}^k}{2} \sum\nolimits_{s=1}^{|\mathcal{P}^k|} ||\gamma_s^{k+1}||^2 \\
& + \frac{4}{\eta_{\boldsymbol{\lambda}}} (\frac{c_{\boldsymbol{\lambda}}^{k-2}}{c_{\boldsymbol{\lambda}}^{k-1}} - \frac{c_{\boldsymbol{\lambda}}^{k-1}}{c_{\boldsymbol{\lambda}}^k}) \sum_{l=1}^{L} ||\boldsymbol{\lambda}_l^k||^2 + \frac{4}{\eta_{\boldsymbol{\alpha}}} (\frac{c_{\boldsymbol{\alpha}}^{k-2}}{c_{\boldsymbol{\alpha}}^{k-1}} - \frac{c_{\boldsymbol{\alpha}}^{k-1}}{c_{\boldsymbol{\alpha}}^k}) \sum_{l=1}^{L} ||\boldsymbol{\alpha}_l^k||^2 \\
& + \frac{4}{\eta_{\boldsymbol{\beta}}} (\frac{c_{\boldsymbol{\beta}}^{k-2}}{c_{\boldsymbol{\beta}}^{k-1}} - \frac{c_{\boldsymbol{\beta}}^{k-1}}{c_{\boldsymbol{\beta}}^k}) \sum_{l=1}^{L} ||\boldsymbol{\beta}_l^k||^2 + \frac{4}{\eta_{\gamma}} (\frac{c_{\gamma}^{k-2}}{c_{\gamma}^{k-1}} - \frac{c_{\gamma}^{k-1}}{c_{\gamma}^k}) \sum_{s=1}^{|\mathcal{P}^k|} ||\gamma_s^k||^2.
\end{aligned} \tag{84}
$$

Proof. Let $a_1 = \frac{1}{\eta_{\boldsymbol{\lambda}}}, a_2 = \frac{1}{\eta_{\boldsymbol{\alpha}}}, a_3 = \frac{1}{\eta_{\boldsymbol{\beta}}}, a_4 = \frac{1}{\eta_\gamma}$, and substitute them into the Lemma C.5, $\forall k \geq K_1$, we have,

$$
\begin{aligned}
&L_p(\{\boldsymbol{\mu}_l^{k+1}\}, \{\mathbf{r}_l^{k+1}\}, \{\mathbf{p}_l^{k+1}\}, \{q_s^{k+1}\}, \mathbf{x}^{k+1}, \{\boldsymbol{\lambda}_l^{k+1}\}, \{\boldsymbol{\alpha}_l^{k+1}\}, \{\boldsymbol{\beta}_l^{k+1}\}, \{\gamma_s^{k+1}\}) \\
&- L_p(\{\boldsymbol{\mu}_l^{k}\}, \{\mathbf{r}_l^{k}\}, \{\mathbf{p}_l^{k}\}, \{q_s^{k}\}, \mathbf{x}^{k}, \{\boldsymbol{\lambda}_l^{k}\}, \{\boldsymbol{\alpha}_l^{k}\}, \{\boldsymbol{\beta}_l^{k}\}, \{\gamma_s^{k}\}) \\
&\leq (\eta_{\boldsymbol{\lambda}} + \eta_{\boldsymbol{\alpha}} + \eta_{\boldsymbol{\beta}} + \frac{w_{\boldsymbol{\mu}} L_w M^{\frac{1}{2}}}{3} + 1 - \frac{1}{\eta_{\boldsymbol{\mu}}^k}) \sum_{l=1}^{L} \|\boldsymbol{\mu}_l^{k+1} - \boldsymbol{\mu}_l^k\|^2 + (4w_{\mathbf{r}}^2 \eta_{\boldsymbol{\lambda}} + \frac{w_{\mathbf{r}} L_w M^{\frac{1}{2}}}{3} + w_{\boldsymbol{\lambda}} - \frac{1}{\eta_{\mathbf{r}}^k}) \sum_{l=1}^{L} \|\mathbf{r}_l^{k+1} - \mathbf{r}_l^k\|^2 \\
&+ (4w_{\mathbf{p}}^2 \eta_{\boldsymbol{\alpha}} + \frac{w_{\mathbf{p}} L_w M^{\frac{1}{2}}}{3} + w_{\boldsymbol{\alpha}} - \frac{1}{\eta_{\mathbf{p}}^k}) \sum_{l=1}^{L} \|\mathbf{p}_l^{k+1} - \mathbf{p}_l^k\|^2 + (4w_q^2 \eta_\gamma + \frac{w_q L_w}{3} + w_\gamma - \frac{1}{\eta_q^k}) \sum_{s=1}^{|\mathcal{P}^k|} \|q_s^{k+1} - q_s^k\|^2 \\
&+ (-\frac{1}{\eta_{\mathbf{x}}^k} + \frac{L_w}{6}) ||\mathbf{x}^{k+1} - \mathbf{x}^k||^2 + \eta_{\boldsymbol{\beta}} \sum_{l=1}^{L} \|\mathbf{B}_l \mathbf{x}^{k+1} - \mathbf{B}_l \mathbf{x}^k\|^2 + \eta_\gamma \sum_{s=1}^{|\mathcal{P}^k|} \|\mathbf{b}_s\|^2 \|\mathbf{x}^{k+1} - \mathbf{x}^k\|^2 \\
&+ (-\frac{c_{\boldsymbol{\lambda}}^{k-1} - c_{\boldsymbol{\lambda}}^k}{2} + \frac{1}{\eta_{\boldsymbol{\lambda}}}) \sum_{l=1}^{L} \|\boldsymbol{\lambda}_l^{k+1} - \boldsymbol{\lambda}_l^k\|^2 + \frac{c_{\boldsymbol{\lambda}}^{k-1}}{2} \sum_{l=1}^{L} (\|\boldsymbol{\lambda}_l^{k+1}\|^2 - \|\boldsymbol{\lambda}_l^k\|^2) + \frac{1}{2\eta_{\boldsymbol{\lambda}}} \sum_{l=1}^{L} \|\boldsymbol{\lambda}_l^k - \boldsymbol{\lambda}_l^{k-1}\|^2 \\
&+ (-\frac{c_{\boldsymbol{\alpha}}^{k-1} - c_{\boldsymbol{\alpha}}^k}{2} + \frac{1}{\eta_{\boldsymbol{\alpha}}}) \sum_{l=1}^{L} \|\boldsymbol{\alpha}_l^{k+1} - \boldsymbol{\alpha}_l^k\|^2 + \frac{c_{\boldsymbol{\alpha}}^{k-1}}{2} \sum_{l=1}^{L} (\|\boldsymbol{\alpha}_l^{k+1}\|^2 - \|\boldsymbol{\alpha}_l^k\|^2) + \frac{1}{2\eta_{\boldsymbol{\alpha}}} \sum_{l=1}^{L} \|\boldsymbol{\alpha}_l^k - \boldsymbol{\alpha}_l^{k-1}\|^2 \\
&+ (-\frac{c_{\boldsymbol{\beta}}^{k-1} - c_{\boldsymbol{\beta}}^k}{2} + \frac{1}{\eta_{\boldsymbol{\beta}}}) \sum_{l=1}^{L} \|\boldsymbol{\beta}_l^{k+1} - \boldsymbol{\beta}_l^k\|^2 + \frac{c_{\boldsymbol{\beta}}^{k-1}}{2} \sum_{l=1}^{L} (\|\boldsymbol{\beta}_l^{k+1}\|^2 - \|\boldsymbol{\beta}_l^k\|^2) + \frac{1}{2\eta_{\boldsymbol{\beta}}} \sum_{l=1}^{L} \|\boldsymbol{\beta}_l^k - \boldsymbol{\beta}_l^{k-1}\|^2 \\
&+ (-\frac{c_\gamma^{k-1} - c_\gamma^k}{2} + \frac{1}{\eta_\gamma}) \sum_{s=1}^{|\mathcal{P}^k|} \|\gamma_s^{k+1} - \gamma_s^k\|^2 + \frac{c_\gamma^{k-1}}{2} \sum_{s=1}^{|\mathcal{P}^k|} (\|\gamma_s^{k+1}\|^2 - \|\gamma_s^k\|^2) + \frac{1}{2\eta_\gamma} \sum_{s=1}^{|\mathcal{P}^k|} \|\gamma_s^k - \gamma_s^{k-1}\|^2.
\end{aligned} \tag{85}
$$

According to (66) ,(67) and (73), in the $(k+1)^{th}$ iteration, we can obtain,

$$
\begin{aligned}
&\frac{1}{\eta_{\boldsymbol{\lambda}}} \langle \boldsymbol{\lambda}_l^{k+1} - \boldsymbol{\lambda}_l^k - (\boldsymbol{\lambda}_l^k - \boldsymbol{\lambda}_l^{k-1}), \boldsymbol{\lambda}_l^{k+1} - \boldsymbol{\lambda}_l^k \rangle \\
&\leq \left\langle \nabla_{\boldsymbol{\lambda}_l} \widetilde{L}_p(\{\boldsymbol{\mu}_l^{k+1}\}, \{\mathbf{r}_l^{k+1}\}, \{\mathbf{p}_l^{k+1}\}, \{q_s^{k+1}\}, \mathbf{x}^{k+1}, \{\boldsymbol{\lambda}_l^k\}, \{\boldsymbol{\alpha}_l^k\}, \{\boldsymbol{\beta}_l^k\}, \{\gamma_s^k\}, \boldsymbol{\lambda}_l^{k+1} - \boldsymbol{\lambda}_l^k \right\rangle \\
&- \left\langle \nabla_{\boldsymbol{\lambda}_l} \widetilde{L}_p(\{\boldsymbol{\mu}_l^{k}\}, \{\mathbf{r}_l^{k}\}, \{\mathbf{p}_l^{k}\}, \{q_s^{k}\}, \mathbf{x}^{k}, \{\boldsymbol{\lambda}_l^{k-1}\}, \{\boldsymbol{\alpha}_l^{k-1}\}, \{\boldsymbol{\beta}_l^{k-1}\}, \{\gamma_s^{k-1}\}, \boldsymbol{\lambda}_l^{k+1} - \boldsymbol{\lambda}_l^k \right\rangle \\
&\leq \frac{1}{b_1^k} \|\boldsymbol{\mu}_l^{k+1} - \boldsymbol{\mu}_l^k\|^2 + \frac{4w_{\mathbf{r}}^2}{b_1^k} \|\mathbf{r}_l^{k+1} - \mathbf{r}_l^k\|^2 + \frac{b_1^k}{2} \|\boldsymbol{\lambda}_l^{k+1} - \boldsymbol{\lambda}_l^k\|^2 + \frac{c_{\boldsymbol{\lambda}}^{k-1} - c_{\boldsymbol{\lambda}}^k}{2} (||\boldsymbol{\lambda}_l^{k+1}||^2 - ||\boldsymbol{\lambda}_l^k||^2) \\
&- \frac{c_{\boldsymbol{\lambda}}^{k-1} - c_{\boldsymbol{\lambda}}^k}{2} ||\boldsymbol{\lambda}_l^{k+1} - \boldsymbol{\lambda}_l^k||^2 + \frac{1}{2\eta_{\boldsymbol{\lambda}}} \|\boldsymbol{\lambda}_l^{k+1} - \boldsymbol{\lambda}_l^k - (\boldsymbol{\lambda}_l^k - \boldsymbol{\lambda}_l^{k-1})\|^2 - \frac{c_{\boldsymbol{\lambda}}^{k-1}}{2} \|\boldsymbol{\lambda}_l^k - \boldsymbol{\lambda}_l^{k-1}\|^2.
\end{aligned} \tag{86}
$$

where $b_1^k > 0$. And we have,

$$
\frac{1}{\eta_{\boldsymbol{\lambda}}} \langle \boldsymbol{\lambda}_l^{k+1} - \boldsymbol{\lambda}_l^k - (\boldsymbol{\lambda}_l^k - \boldsymbol{\lambda}_l^{k-1}), \boldsymbol{\lambda}_l^{k+1} - \boldsymbol{\lambda}_l^k \rangle = \frac{1}{2\eta_{\boldsymbol{\lambda}}} ||\boldsymbol{\lambda}_l^{k+1} - \lambda_l^k||^2 - \frac{1}{2\eta_{\boldsymbol{\lambda}}} ||\boldsymbol{\lambda}_l^k - \boldsymbol{\lambda}_l^{k-1}||^2 + \frac{1}{2\eta_{\boldsymbol{\lambda}}} \|\boldsymbol{\lambda}_l^{k+1} - \boldsymbol{\lambda}_l^k - (\boldsymbol{\lambda}_l^k - \boldsymbol{\lambda}_l^{k-1})\|^2. \tag{87}
$$

Combining (86) with (87), we have

$$
\begin{aligned}
&\frac{1}{2\eta_{\boldsymbol{\lambda}}} ||\boldsymbol{\lambda}_l^{k+1} - \boldsymbol{\lambda}_l^k||^2 - \frac{1}{2\eta_{\boldsymbol{\lambda}}} ||\boldsymbol{\lambda}_l^k - \boldsymbol{\lambda}_l^{k-1}||^2 \leq \frac{1}{b_1^k} \|\boldsymbol{\mu}_l^{k+1} - \boldsymbol{\mu}_l^k\|^2 + \frac{4w_{\mathbf{r}}^2}{b_1^k} \|\mathbf{r}_l^{k+1} - \mathbf{r}_l^k\|^2 + \frac{b_1^k}{2} \|\boldsymbol{\lambda}_l^{k+1} - \boldsymbol{\lambda}_l^k\|^2 \\
&+ \frac{c_{\boldsymbol{\lambda}}^{k-1} - c_{\boldsymbol{\lambda}}^k}{2} (||\boldsymbol{\lambda}_l^{k+1}||^2 - ||\boldsymbol{\lambda}_l^k||^2) - \frac{c_{\boldsymbol{\lambda}}^{k-1} - c_{\boldsymbol{\lambda}}^k}{2} ||\boldsymbol{\lambda}_l^{k+1} - \boldsymbol{\lambda}_l^k||^2 - \frac{c_{\boldsymbol{\lambda}}^{k-1}}{2} \|\boldsymbol{\lambda}_l^k - \boldsymbol{\lambda}_l^{k-1}\|^2.
\end{aligned} \tag{88}
$$

Multiplying both sides of by $\frac{8}{\eta_{\boldsymbol{\lambda}} c_{\boldsymbol{\lambda}}^k}$, we have,

$$\begin{aligned}
&\frac{4}{\eta_{\boldsymbol{\lambda}}^2 c_{\boldsymbol{\lambda}}^k}||\boldsymbol{\lambda}_l^{k+1}-\boldsymbol{\lambda}_l^k||^2-\frac{4}{\eta_{\boldsymbol{\lambda}}}(\frac{c_{\boldsymbol{\lambda}}^{k-1}-c_{\boldsymbol{\lambda}}^k}{c_{\boldsymbol{\lambda}}^k})||\boldsymbol{\lambda}_l^{k+1}||^2\\
&\leq \frac{4}{\eta_{\boldsymbol{\lambda}}^2 c_{\boldsymbol{\lambda}}^k}||\boldsymbol{\lambda}_l^{k}-\boldsymbol{\lambda}_l^{k-1}||^2-\frac{4}{\eta_{\boldsymbol{\lambda}}}(\frac{c_{\boldsymbol{\lambda}}^{k-1}-c_{\boldsymbol{\lambda}}^k}{c_{\boldsymbol{\lambda}}^k})||\boldsymbol{\lambda}_l^{k}||^2\\
&+\frac{4b_1^k}{\eta_{\boldsymbol{\lambda}} c_{\boldsymbol{\lambda}}^k}||\boldsymbol{\lambda}_l^{k+1}-\boldsymbol{\lambda}_l^k||^2-\frac{4}{\eta_{\boldsymbol{\lambda}}}||\boldsymbol{\lambda}_l^{k}-\boldsymbol{\lambda}_l^{k-1}||^2\\
&+\frac{8}{b_1^k\eta_{\boldsymbol{\lambda}} c_{\boldsymbol{\lambda}}^k}\|\boldsymbol{\mu}_l^{k+1}-\boldsymbol{\mu}_l^k\|^2+\frac{32w_{\mathbf{r}}^2}{b_1^k\eta_{\boldsymbol{\lambda}} c_{\boldsymbol{\lambda}}^k}\|\mathbf{r}_l^{k+1}-\mathbf{r}_l^k\|^2.
\end{aligned} \tag{89}$$

Setting $b_1^k=\frac{c_{\boldsymbol{\lambda}}^k}{2}$ in (89) and combine it with the definition of S_1^k, we have,

$$\begin{aligned}
S_1^{k+1}-S_1^k\leq&\sum_{l=1}^L\frac{4}{\eta_{\boldsymbol{\lambda}}}(\frac{c_{\boldsymbol{\lambda}}^{k-2}}{c_{\boldsymbol{\lambda}}^{k-1}}-\frac{c_{\boldsymbol{\lambda}}^{k-1}}{c_{\boldsymbol{\lambda}}^k})||\boldsymbol{\lambda}_l^k||^2+\sum_{l=1}^L(\frac{2}{\eta_{\boldsymbol{\lambda}}}+\frac{4}{\eta_{\boldsymbol{\lambda}}^2}(\frac{1}{c_{\boldsymbol{\lambda}}^{k+1}}-\frac{1}{c_{\boldsymbol{\lambda}}^k}))||\boldsymbol{\lambda}_l^{k+1}-\boldsymbol{\lambda}_l^k||^2\\
&-\sum_{l=1}^L\frac{4}{\eta_{\boldsymbol{\lambda}}}||\boldsymbol{\lambda}_l^k-\boldsymbol{\lambda}_l^{k-1}||^2+\frac{16}{\eta_{\boldsymbol{\lambda}}(c_{\boldsymbol{\lambda}}^k)^2}\sum_{l=1}^L\|\boldsymbol{\mu}_l^{k+1}-\boldsymbol{\mu}_l^k\|^2+\frac{64w_{\mathbf{r}}^2}{\eta_{\boldsymbol{\lambda}}(c_{\boldsymbol{\lambda}}^k)^2}\sum_{l=1}^L\|\mathbf{r}_l^{k+1}-\mathbf{r}_l^k\|^2.
\end{aligned} \tag{90}$$

Similar results can be obtained for S_2^k, S_3^k, S_4^k,

$$\begin{aligned}
S_2^{k+1}-S_2^k\leq&\sum_{l=1}^L\frac{4}{\eta_{\boldsymbol{\alpha}}}(\frac{c_{\boldsymbol{\alpha}}^{k-2}}{c_{\boldsymbol{\alpha}}^{k-1}}-\frac{c_{\boldsymbol{\alpha}}^{k-1}}{c_{\boldsymbol{\alpha}}^k})||\boldsymbol{\alpha}_l^k||^2+\sum_{l=1}^L(\frac{2}{\eta_{\boldsymbol{\alpha}}}+\frac{4}{\eta_{\boldsymbol{\alpha}}^2}(\frac{1}{c_{\boldsymbol{\alpha}}^{k+1}}-\frac{1}{c_{\boldsymbol{\alpha}}^k}))||\boldsymbol{\alpha}_l^{k+1}-\boldsymbol{\alpha}_l^k||^2\\
&-\sum_{l=1}^L\frac{4}{\eta_{\boldsymbol{\alpha}}}||\boldsymbol{\alpha}_l^k-\boldsymbol{\alpha}_l^{k-1}||^2+\frac{16}{\eta_{\boldsymbol{\alpha}}(c_{\boldsymbol{\alpha}}^k)^2}\sum_{l=1}^L\|\boldsymbol{\mu}_l^{k+1}-\boldsymbol{\mu}_l^k\|^2+\frac{64w_{\mathbf{p}}^2}{\eta_{\boldsymbol{\alpha}}(c_{\boldsymbol{\alpha}}^k)^2}\sum_{l=1}^L\|\mathbf{p}_l^{k+1}-\mathbf{p}_l^k\|^2.
\end{aligned} \tag{91}$$

$$\begin{aligned}
S_3^{k+1}-S_3^k\leq&\sum_{l=1}^L\frac{4}{\eta_{\boldsymbol{\beta}}}(\frac{c_{\boldsymbol{\beta}}^{k-2}}{c_{\boldsymbol{\beta}}^{k-1}}-\frac{c_{\boldsymbol{\beta}}^{k-1}}{c_{\boldsymbol{\beta}}^k})||\boldsymbol{\beta}_l^k||^2+\sum_{l=1}^L(\frac{2}{\eta_{\boldsymbol{\beta}}}+\frac{4}{\eta_{\boldsymbol{\beta}}^2}(\frac{1}{c_{\boldsymbol{\beta}}^{k+1}}-\frac{1}{c_{\boldsymbol{\beta}}^k}))||\boldsymbol{\beta}_l^{k+1}-\boldsymbol{\beta}_l^k||^2\\
&-\sum_{l=1}^L\frac{4}{\eta_{\boldsymbol{\beta}}}||\boldsymbol{\beta}_l^k-\boldsymbol{\beta}_l^{k-1}||^2+\frac{16}{\eta_{\boldsymbol{\beta}}(c_{\boldsymbol{\beta}}^k)^2}\sum_{l=1}^L\|\boldsymbol{\mu}_l^{k+1}-\boldsymbol{\mu}_l^k\|^2+\frac{16}{\eta_{\boldsymbol{\beta}}(c_{\boldsymbol{\beta}}^k)^2}\sum_{l=1}^L\|\mathbf{B}_l\mathbf{x}^{k+1}-\mathbf{B}_l\mathbf{x}^k\|^2.
\end{aligned} \tag{92}$$

$$\begin{aligned}
S_4^{k+1}-S_4^k\leq&\sum_{s=1}^{|\mathcal{P}^k|}\frac{4}{\eta_\gamma}(\frac{c_\gamma^{k-2}}{c_\gamma^{k-1}}-\frac{c_\gamma^{k-1}}{c_\gamma^k})||\gamma_s^k||^2+\sum_{s=1}^{|\mathcal{P}^k|}(\frac{2}{\eta_\gamma}+\frac{4}{\eta_\gamma^2}(\frac{1}{c_\gamma^{k+1}}-\frac{1}{c_\gamma^k}))||\gamma_s^{k+1}-\gamma_s^k||^2\\
&-\sum_{s=1}^{|\mathcal{P}^k|}\frac{4}{\eta_\gamma}||\gamma_s^k-\gamma_s^{k-1}||^2+\frac{16}{\eta_\gamma(c_\gamma^k)^2}\sum_{s=1}^{|\mathcal{P}^k|}\|\mathbf{b}_s\|^2\|\mathbf{x}_l^{k+1}-\mathbf{x}_l^k\|^2+\frac{64w_{\mathbf{q}}^2}{\eta_\gamma(c_\gamma^k)^2}\sum_{s=1}^{|\mathcal{P}^k|}\|q_s^{k+1}-q_s^k\|^2.
\end{aligned} \tag{93}$$

Based on the setting of $c_{\boldsymbol{\lambda}}^k$, $c_{\boldsymbol{\alpha}}^k$, $c_{\boldsymbol{\beta}}^k$ and c_γ^k, we can obtain that, $\frac{\eta_{\boldsymbol{\lambda}}}{10}\geq\frac{1}{c_{\boldsymbol{\lambda}}^{k+1}}-\frac{1}{c_{\boldsymbol{\lambda}}^k}$, $\frac{\eta_{\boldsymbol{\alpha}}}{10}\geq\frac{1}{c_{\boldsymbol{\alpha}}^{k+1}}-\frac{1}{c_{\boldsymbol{\alpha}}^k}$, $\frac{\eta_{\boldsymbol{\beta}}}{10}\geq\frac{1}{c_{\boldsymbol{\beta}}^{k+1}}-\frac{1}{c_{\boldsymbol{\beta}}^k}$, $\frac{\eta_\gamma}{10}\geq\frac{1}{c_\gamma^{k+1}}-\frac{1}{c_\gamma^k}$, $\forall k\geq K_1$. In addition, according to the definition of $\mathbf{B}_l$, the following inequality can be obtained,

$$\sum_{l=1}^L\|\mathbf{B}_l\mathbf{x}^{k+1}-\mathbf{B}_l\mathbf{x}^k\|^2=\|\mathbf{x}^{k+1}-\mathbf{x}^k\|^2. \tag{94}$$

According to the definition of a_5, combining (90)-(94) with (85), we can obtain that,

$$
\begin{aligned}
F^{k+1} - F^k \leq & (a_5 - \frac{1}{\eta_{\boldsymbol{\mu}}^k} + \frac{16}{\eta_{\boldsymbol{\lambda}}(c_{\boldsymbol{\lambda}}^k)^2} + \frac{16}{\eta_{\boldsymbol{\alpha}}(c_{\boldsymbol{\alpha}}^k)^2} + \frac{16}{\eta_{\boldsymbol{\beta}}(c_{\boldsymbol{\beta}}^k)^2}) \sum_{l=1}^{L} \|\boldsymbol{\mu}_l^{k+1} - \boldsymbol{\mu}_l^k\|^2 \\
& + (a_5 - \frac{1}{\eta_{\mathbf{r}}^k} + \frac{64 w_{\mathbf{r}}^2}{\eta_{\boldsymbol{\lambda}}(c_{\boldsymbol{\lambda}}^k)^2}) \sum_{l=1}^{L} \|\mathbf{r}_l^{k+1} - \mathbf{r}_l^k\|^2 \\
& + (a_5 - \frac{1}{\eta_{\mathbf{p}}^k} + \frac{64 w_{\mathbf{p}}^2}{\eta_{\boldsymbol{\alpha}}(c_{\boldsymbol{\alpha}}^k)^2}) \sum_{l=1}^{L} \|\mathbf{p}_l^{k+1} - \mathbf{p}_l^k\|^2 \\
& + (a_5 - \frac{1}{\eta_q^k} + \frac{64 w_q^2}{\eta_\gamma(c_\gamma^k)^2}) \sum_{s=1}^{|\mathcal{P}^k|} \|q_s^{k+1} - q_s^k\|^2 \\
& + (a_5 - \frac{1}{\eta_{\mathbf{x}}^k} + \frac{16}{\eta_{\boldsymbol{\beta}}(c_{\boldsymbol{\beta}}^k)^2} + \frac{16}{\eta_\gamma(c_\gamma^k)^2} \sum_{s=1}^{|\mathcal{P}^k|} \|b_s\|^2) \|\mathbf{x}^{k+1} - \mathbf{x}^k\|^2 \\
& - \frac{1}{10\eta_{\boldsymbol{\lambda}}} \sum_{l=1}^{L} ||\boldsymbol{\lambda}_l^{k+1} - \boldsymbol{\lambda}_l^k||^2 + \frac{c_{\boldsymbol{\lambda}}^{k-1} - c_{\boldsymbol{\lambda}}^k}{2} \sum_{l=1}^{L} ||\boldsymbol{\lambda}_l^{k+1}||^2 \\
& - \frac{1}{10\eta_{\boldsymbol{\alpha}}} \sum_{l=1}^{L} ||\boldsymbol{\alpha}_l^{k+1} - \boldsymbol{\alpha}_l^k||^2 + \frac{c_{\boldsymbol{\alpha}}^{k-1} - c_{\boldsymbol{\alpha}}^k}{2} \sum_{l=1}^{L} ||\boldsymbol{\alpha}_l^{k+1}||^2 \\
& - \frac{1}{10\eta_{\boldsymbol{\beta}}} \sum_{l=1}^{L} ||\boldsymbol{\beta}_l^{k+1} - \boldsymbol{\beta}_l^k||^2 + \frac{c_{\boldsymbol{\beta}}^{k-1} - c_{\boldsymbol{\beta}}^k}{2} \sum_{l=1}^{L} ||\boldsymbol{\beta}_l^{k+1}||^2 \\
& - \frac{1}{10\eta_\gamma} \sum_{s=1}^{|\mathcal{P}^k|} ||\gamma_s^{k+1} - \gamma_s^k||^2 + \frac{c_\gamma^{k-1} - c_\gamma^k}{2} \sum_{s=1}^{|\mathcal{P}^k|} ||\gamma_s^{k+1}||^2 \\
& + \frac{4}{\eta_{\boldsymbol{\lambda}}} (\frac{c_{\boldsymbol{\lambda}}^{k-2}}{c_{\boldsymbol{\lambda}}^{k-1}} - \frac{c_{\boldsymbol{\lambda}}^{k-1}}{c_{\boldsymbol{\lambda}}^k}) \sum_{l=1}^{L} ||\boldsymbol{\lambda}_l^k||^2 + \frac{4}{\eta_{\boldsymbol{\alpha}}} (\frac{c_{\boldsymbol{\alpha}}^{k-2}}{c_{\boldsymbol{\alpha}}^{k-1}} - \frac{c_{\boldsymbol{\alpha}}^{k-1}}{c_{\boldsymbol{\alpha}}^k}) \sum_{l=1}^{L} ||\boldsymbol{\alpha}_l^k||^2 \\
& + \frac{4}{\eta_{\boldsymbol{\beta}}} (\frac{c_{\boldsymbol{\beta}}^{k-2}}{c_{\boldsymbol{\beta}}^{k-1}} - \frac{c_{\boldsymbol{\beta}}^{k-1}}{c_{\boldsymbol{\beta}}^k}) \sum_{l=1}^{L} ||\boldsymbol{\beta}_l^k||^2 + \frac{4}{\eta_\gamma} (\frac{c_\gamma^{k-2}}{c_\gamma^{k-1}} - \frac{c_\gamma^{k-1}}{c_\gamma^k}) \sum_{s=1}^{|\mathcal{P}^k|} ||\gamma_s^k||^2.
\end{aligned} \tag{95}
$$

which concludes the proof of Lemma C.6. □

Finally, based on Lemma C.4, Lemma C.5 and Lemma C.6, we provide the proof of Theorem 6.4.

Proof. First, we set

$$a_6^k = \min\left\{\frac{16}{\eta_{\boldsymbol{\lambda}}(c_{\boldsymbol{\lambda}}^k)^2} + \frac{16}{\eta_{\boldsymbol{\alpha}}(c_{\boldsymbol{\alpha}}^k)^2} + \frac{16}{\eta_{\boldsymbol{\beta}}(c_{\boldsymbol{\beta}}^k)^2}, \frac{64w_{\mathbf{r}}^2}{\eta_{\boldsymbol{\lambda}}(c_{\boldsymbol{\lambda}}^k)^2}, \frac{64w_{\mathbf{p}}^2}{\eta_{\boldsymbol{\alpha}}(c_{\boldsymbol{\alpha}}^k)^2}, \frac{64w_q^2}{\eta_\gamma(c_\gamma^k)^2}, \frac{16}{\eta_{\boldsymbol{\beta}}(c_{\boldsymbol{\beta}}^k)^2} + \frac{16}{\eta_\gamma(c_\gamma^k)^2}\sum_{s=1}^{|\mathcal{P}^k|}\|b_s\|^2\right\}\frac{\xi-2}{2} - a_5. \tag{96}$$

where constant $\xi > 2$ and satisfies

$$\min\left\{\frac{16}{\eta_{\boldsymbol{\lambda}}(c_{\boldsymbol{\lambda}}^0)^2} + \frac{16}{\eta_{\boldsymbol{\alpha}}(c_{\boldsymbol{\alpha}}^0)^2} + \frac{16}{\eta_{\boldsymbol{\beta}}(c_{\boldsymbol{\beta}}^0)^2}, \frac{64w_{\mathbf{r}}^2}{\eta_{\boldsymbol{\lambda}}(c_{\boldsymbol{\lambda}}^0)^2}, \frac{64w_{\mathbf{p}}^2}{\eta_{\boldsymbol{\alpha}}(c_{\boldsymbol{\alpha}}^0)^2}, \frac{64w_q^2}{\eta_\gamma(c_\gamma^0)^2}, \frac{16}{\eta_{\boldsymbol{\beta}}(c_{\boldsymbol{\beta}}^0)^2} + \frac{16}{\eta_\gamma(c_\gamma^0)^2}\sum_{s=1}^{|\mathcal{P}^k|}\|b_s\|^2\right\}\frac{\xi-2}{2} > a_5. \tag{97}$$

Thus, we have $a_6^k > 0, \forall k$. According to the setting of $\eta_{\boldsymbol{\lambda}}^k$, $\eta_{\boldsymbol{\alpha}}^k$, $\eta_{\boldsymbol{\beta}}^k$, η_γ^k and $c_{\boldsymbol{\lambda}}^k$, $c_{\boldsymbol{\alpha}}^k$, $c_{\boldsymbol{\beta}}^k$, c_γ^k, we have,

$$\begin{aligned}
a_5 - \frac{1}{\eta_{\boldsymbol{\mu}}^k} + \frac{16}{\eta_{\boldsymbol{\lambda}}(c_{\boldsymbol{\lambda}}^k)^2} + \frac{16}{\eta_{\boldsymbol{\alpha}}(c_{\boldsymbol{\alpha}}^k)^2} + \frac{16}{\eta_{\boldsymbol{\beta}}(c_{\boldsymbol{\beta}}^k)^2} &\le -a_6^k,\\
a_5 - \frac{1}{\eta_{\mathbf{r}}^k} + \frac{64w_{\mathbf{r}}^2}{\eta_{\boldsymbol{\lambda}}(c_{\boldsymbol{\lambda}}^k)^2} &\le -a_6^k,\\
a_5 - \frac{1}{\eta_{\mathbf{p}}^k} + \frac{64w_{\mathbf{p}}^2}{\eta_{\boldsymbol{\alpha}}(c_{\boldsymbol{\alpha}}^k)^2} &\le -a_6^k,\\
(a_5 - \frac{1}{\eta_q^k} + \frac{64w_q^2}{\eta_\gamma(c_\gamma^k)^2}) &\le -a_6^k,\\
a_5 - \frac{1}{\eta_{\mathbf{x}}^k} + \frac{16}{\eta_{\boldsymbol{\beta}}(c_{\boldsymbol{\beta}}^k)^2} + \frac{16}{\eta_\gamma(c_\gamma^k)^2}\sum_{s=1}^{|\mathcal{P}^k|}\|b_s\|^2 &\le -a_6^k.
\end{aligned} \tag{98}$$

Combining it with Lemma C.6, $\forall k \ge K_1$, we can obtain that,

$$\begin{aligned}
&a_6^k\left(\sum_{l=1}^L\|\boldsymbol{\mu}_l^{k+1}-\boldsymbol{\mu}_l^k\|^2 + \sum_{l=1}^L\|\mathbf{r}_l^{k+1}-\mathbf{r}_l^k\|^2 + \sum_{l=1}^L\|\mathbf{p}_l^{k+1}-\mathbf{p}_l^k\|^2 + \sum_{s=1}^{|\mathcal{P}^k|}\|q_s^{k+1}-q_s^k\|^2 + \|\mathbf{x}^{k+1}-\mathbf{x}^k\|^2\right)\\
&\le F^k - F^{k+1} - \frac{1}{10\eta_{\boldsymbol{\lambda}}}\sum_{l=1}^L||\boldsymbol{\lambda}_l^{k+1}-\boldsymbol{\lambda}_l^k||^2 + \frac{c_{\boldsymbol{\lambda}}^{k-1}-c_{\boldsymbol{\lambda}}^k}{2}\sum_{l=1}^L||\boldsymbol{\lambda}_l^{k+1}||^2 + \frac{4}{\eta_{\boldsymbol{\lambda}}}(\frac{c_{\boldsymbol{\lambda}}^{k-2}}{c_{\boldsymbol{\lambda}}^{k-1}} - \frac{c_{\boldsymbol{\lambda}}^{k-1}}{c_{\boldsymbol{\lambda}}^k})\sum_{l=1}^L||\boldsymbol{\lambda}_l^k||^2\\
&- \frac{1}{10\eta_{\boldsymbol{\alpha}}}\sum_{l=1}^L||\boldsymbol{\alpha}_l^{k+1}-\boldsymbol{\alpha}_l^k||^2 + \frac{c_{\boldsymbol{\alpha}}^{k-1}-c_{\boldsymbol{\alpha}}^k}{2}\sum_{l=1}^L||\boldsymbol{\alpha}_l^{k+1}||^2 + \frac{4}{\eta_{\boldsymbol{\alpha}}}(\frac{c_{\boldsymbol{\alpha}}^{k-2}}{c_{\boldsymbol{\alpha}}^{k-1}} - \frac{c_{\boldsymbol{\alpha}}^{k-1}}{c_{\boldsymbol{\alpha}}^k})\sum_{l=1}^L||\boldsymbol{\alpha}_l^k||^2\\
&- \frac{1}{10\eta_{\boldsymbol{\beta}}}\sum_{l=1}^L||\boldsymbol{\beta}_l^{k+1}-\boldsymbol{\beta}_l^k||^2 + \frac{c_{\boldsymbol{\beta}}^{k-1}-c_{\boldsymbol{\beta}}^k}{2}\sum_{l=1}^L||\boldsymbol{\beta}_l^{k+1}||^2 + \frac{4}{\eta_{\boldsymbol{\beta}}}(\frac{c_{\boldsymbol{\beta}}^{k-2}}{c_{\boldsymbol{\beta}}^{k-1}} - \frac{c_{\boldsymbol{\beta}}^{k-1}}{c_{\boldsymbol{\beta}}^k})\sum_{l=1}^L||\boldsymbol{\beta}_l^k||^2\\
&- \frac{1}{10\eta_\gamma}\sum_{s=1}^{|\mathcal{P}^k|}||\gamma_s^{k+1}-\gamma_s^k||^2 + \frac{c_\gamma^{k-1}-c_\gamma^k}{2}\sum_{s=1}^{|\mathcal{P}^k|}||\gamma_s^{k+1}||^2 + \frac{4}{\eta_\gamma}(\frac{c_\gamma^{k-2}}{c_\gamma^{k-1}} - \frac{c_\gamma^{k-1}}{c_\gamma^k})\sum_{s=1}^{|\mathcal{P}^k|}||\gamma_s^k||^2.
\end{aligned} \tag{99}$$

Given the definition of $\nabla\widetilde{G}^k$, we have that,

$$\begin{aligned}
(\nabla\widetilde{G}^k)_{\boldsymbol{\mu}_l} =& \nabla_{\boldsymbol{\mu}_l}\widetilde{L}_p(\{\boldsymbol{\mu}_l^{\hat{k}_l}\},\{\mathbf{r}_l^{\hat{k}_l}\},\{\mathbf{p}_l^{\hat{k}_l}\},\{q_s^{\hat{k}_l}\},\mathbf{x}^{\hat{k}_l},\{\boldsymbol{\lambda}_l^{\hat{k}}\},\{\boldsymbol{\alpha}_l^{\hat{k}}\},\{\boldsymbol{\beta}_l^{\hat{k}}\},\{\gamma_s^{\hat{k}}\})\\
&+\nabla_{\boldsymbol{\mu}_l}\widetilde{L}_p(\{\boldsymbol{\mu}_l^k\},\{\mathbf{r}_l^k\},\{\mathbf{p}_l^k\},\{q_s^k\},\mathbf{x}^k,\{\boldsymbol{\lambda}_l^k\},\{\boldsymbol{\alpha}_l^k\},\{\boldsymbol{\beta}_l^k\},\{\gamma_s^k\})\\
&-\nabla_{\boldsymbol{\mu}_l}\widetilde{L}_p(\{\boldsymbol{\mu}_l^{\hat{k}_l}\},\{\mathbf{r}_l^{\hat{k}_l}\},\{\mathbf{p}_l^{\hat{k}_l}\},\{q_s^{\hat{k}_l}\},\mathbf{x}^{\hat{k}_l},\{\boldsymbol{\lambda}_l^{\hat{k}}\},\{\boldsymbol{\alpha}_l^{\hat{k}}\},\{\boldsymbol{\beta}_l^{\hat{k}}\},\{\gamma_s^{\hat{k}}\}).
\end{aligned} \tag{100}$$

Combining it with (51), we have that,

$$\|(\nabla \widetilde{G}^k)_{\boldsymbol{\mu}_l}\|^2 \leq \frac{1}{\eta_{\boldsymbol{\mu}}^2}\|\boldsymbol{\mu}_l^{\bar{k}_l} - \boldsymbol{\mu}_l^k\|^2. \tag{101}$$

Similar results can be derived for other variables,

$$\begin{aligned}
&\|(\nabla \widetilde{G}^k)_{\mathbf{r}_l}\|^2 \leq \frac{1}{\eta_{\mathbf{r}}^2}\|\mathbf{r}_l^{\bar{k}_l} - \mathbf{r}_l^k\|^2.\\
&\|(\nabla \widetilde{G}^k)_{\mathbf{p}_l}\|^2 \leq \frac{1}{\eta_{\mathbf{p}}^2}\|\mathbf{p}_l^{\bar{k}_l} - \mathbf{p}_l^k\|^2.\\
&\|(\nabla \widetilde{G}^k)_{q_s}\|^2 \leq \frac{1}{\eta_q^2}\|q_s^{k+1} - q_s^k\|^2.\\
&\|(\nabla \widetilde{G}^k)_{\mathbf{x}}\|^2 \leq \frac{1}{\eta_{\mathbf{x}}^2}\|\mathbf{x}^{k+1} - \mathbf{x}^k\|^2.
\end{aligned} \tag{102}$$

According to the definition C.2, we have,

$$\begin{aligned}
(\nabla \widetilde{G}^k)_{\boldsymbol{\lambda}_l} =& \nabla_{\boldsymbol{\lambda}_l}\widetilde{L}_p(\{\boldsymbol{\mu}_l^{\bar{k}_l}\}, \{\mathbf{r}_l^{\bar{k}_l}\}, \{\mathbf{p}_l^{\bar{k}_l}\}, \{q_s^{\bar{k}_l}\}, \mathbf{x}^{\bar{k}_l}, \{\boldsymbol{\lambda}_l^k\}, \{\boldsymbol{\alpha}_l^k\}, \{\boldsymbol{\beta}_l^k\}, \{\gamma_s^k\})\\
&+\nabla_{\boldsymbol{\lambda}_l}\widetilde{L}_p(\{\boldsymbol{\mu}_l^k\}, \{\mathbf{r}_l^k\}, \{\mathbf{p}_l^k\}, \{q_s^k\}, \mathbf{x}^k, \{\boldsymbol{\lambda}_l^k\}, \{\boldsymbol{\alpha}_l^k\}, \{\boldsymbol{\beta}_l^k\}, \{\gamma_s^k\})\\
&-\nabla_{\boldsymbol{\lambda}_l}\widetilde{L}_p(\{\boldsymbol{\mu}_l^{\bar{k}_l}\}, \{\mathbf{r}_l^{\bar{k}_l}\}, \{\mathbf{p}_l^{\bar{k}_l}\}, \{q_s^{\bar{k}_l}\}, \mathbf{x}^{\bar{k}_l}, \{\boldsymbol{\lambda}_l^k\}, \{\boldsymbol{\alpha}_l^k\}, \{\boldsymbol{\beta}_l^k\}, \{\gamma_s^k\}).
\end{aligned} \tag{103}$$

Combining trigonometric inequality, (29) with Assumption 6.3, we can obtain

$$\begin{aligned}
\|(\nabla \widetilde{G}^k)_{\boldsymbol{\lambda}_l}\|^2 \leq& 3\|\nabla_{\boldsymbol{\lambda}_l}\widetilde{L}_p(\{\boldsymbol{\mu}_l^{\bar{k}_l}\}, \{\mathbf{r}_l^{\bar{k}_l}\}, \{\mathbf{p}_l^{\bar{k}_l}\}, \{q_s^{\bar{k}_l}\}, \mathbf{x}^{\bar{k}_l}, \{\boldsymbol{\lambda}_l^k\}, \{\boldsymbol{\alpha}_l^k\}, \{\boldsymbol{\beta}_l^k\}, \{\gamma_s^k\})\|^2\\
&+3((c_{\boldsymbol{\lambda}}^{\hat{k}_l-1})^2 - (c_{\boldsymbol{\lambda}}^{\bar{k}_l-1})^2)||\boldsymbol{\lambda}_l^k||^2 + 3\|\boldsymbol{\mu}_l^{\bar{k}} - \boldsymbol{\mu}_l^k + \mathbf{r}_l^k \circ \mathbf{r}_l^k - \mathbf{r}_l^{\bar{k}} \circ \mathbf{r}_l^{\bar{k}}\|^2\\
\leq& \frac{3}{\eta_{\boldsymbol{\lambda}}^2}||\boldsymbol{\lambda}_l^{\bar{k}_l} - \boldsymbol{\lambda}_l^k||^2 + 3((c_{\boldsymbol{\lambda}}^{\hat{k}_l-1})^2 - (c_{\boldsymbol{\lambda}}^{\bar{k}_l-1})^2)||\boldsymbol{\lambda}_l^k||^2\\
&+6\|\boldsymbol{\mu}_l^{\bar{k}_l} - \boldsymbol{\mu}_l^k\|^2 + 24w_{\mathbf{r}}^2\|\mathbf{r}_l^{\bar{k}_l} - \mathbf{r}_l^k\|^2.
\end{aligned} \tag{104}$$

Similar results can be derived for other variables as well,

$$\begin{aligned}
\|(\nabla \widetilde{G}^k)_{\boldsymbol{\alpha}_l}\|^2 \leq& \frac{3}{\eta_{\boldsymbol{\alpha}}^2}||\boldsymbol{\alpha}_l^{\bar{k}_l} - \boldsymbol{\alpha}_l^k||^2 + 3((c_{\boldsymbol{\alpha}}^{\hat{k}_l-1})^2 - (c_{\boldsymbol{\alpha}}^{\bar{k}_l-1})^2)||\boldsymbol{\alpha}_l^k||^2\\
&+ 6\|\boldsymbol{\mu}_l^{\bar{k}_l} - \boldsymbol{\mu}_l^k\|^2 + 24w_{\mathbf{p}}^2\|\mathbf{p}_l^{\bar{k}_l} - \mathbf{p}_l^k\|^2.\\
\|(\nabla \widetilde{G}^k)_{\boldsymbol{\beta}_l}\|^2 \leq& \frac{3}{\eta_{\boldsymbol{\beta}}^2}||\boldsymbol{\beta}_l^{\bar{k}_l} - \boldsymbol{\beta}_l^k||^2 + 3((c_{\boldsymbol{\beta}}^{\hat{k}_l-1})^2 - (c_{\boldsymbol{\beta}}^{\bar{k}_l-1})^2)||\boldsymbol{\beta}_l^k||^2\\
&+ 6\|\boldsymbol{\mu}_l^{\bar{k}_l} - \boldsymbol{\mu}_l^k\|^2 + 6\|\mathbf{x}^{\bar{k}_l} - \mathbf{x}^k\|^2.\\
\|(\nabla \widetilde{G}^k)_{\gamma_s}\|^2 \leq& \frac{3}{\eta_\gamma^2}||\gamma_s^{k+1} - \gamma_s^k||^2 + 3((c_\gamma^{k-1})^2 - (c_\gamma^k)^2)||\gamma_s^k||^2\\
&+ 6\|\mathbf{b}_s\|^2\|\mathbf{x}^{k+1} - \mathbf{x}^k\|^2 + 24w_q^2\|q_s^{k+1} - q_s^k\|^2.
\end{aligned} \tag{105}$$

According to Assumption 6.3, we have,

$$\|\mathbf{x}^{\bar{k}_l} - \mathbf{x}^k\|^2 \leq \tau k_1 \vartheta \leq \tau k_1\|\mathbf{x}^{k+1} - \mathbf{x}^k\|^2. \tag{106}$$

Combining it with (101)-(105), we can obtain that,

$$
\begin{aligned}
||\nabla \widetilde{G}^k||^2 = & \sum_{l=1}^{L}(||(\nabla \widetilde{G}^k)_{\boldsymbol{\mu}_l}||^2 + ||(\nabla \widetilde{G}^k)_{\mathbf{r}_l}||^2 + ||\nabla G^k)_{\mathbf{p}_l}||^2 + ||\nabla \widetilde{G}^k)_{\boldsymbol{\lambda}_l}||^2 + ||(\nabla G^k)_{\boldsymbol{\alpha}_l}||^2 + ||(\nabla \widetilde{G}^k)_{\boldsymbol{\beta}_l}||^2) \\
& + \sum_{s=1}^{|\mathcal{P}^k|} ||(\nabla \widetilde{G}^k)_{q_s}||^2 + ||(\nabla \widetilde{G}^k)_{\mathbf{x}}||^2 + \sum_{s=1}^{|\mathcal{P}^k|} ||(\nabla \widetilde{G}^k)_{\gamma_s}||^2 \\
& \leq (\frac{1}{\eta_{\boldsymbol{\mu}}^2} + 18) \sum_{l=1}^{L} \|\boldsymbol{\mu}_l^{\bar{k}_l} - \boldsymbol{\mu}_l^k\|^2 + (\frac{1}{\eta_{\mathbf{r}}^2} + 24 w_{\mathbf{r}}^2) \sum_{l=1}^{L} \|\mathbf{r}_l^{\bar{k}_l} - \mathbf{r}_l^k\|^2 \\
& + (\frac{1}{\eta_{\mathbf{p}}^2} + 24 w_{\mathbf{p}}^2) \sum_{l=1}^{L} \|\mathbf{p}_l^{\bar{k}_l} - \mathbf{p}_l^k\|^2 \\
& + (\frac{1}{\eta_q^2} + 24 w_{\mathbf{q}}^2) \sum_{s=1}^{|\mathcal{P}^k|} \|q_s^{k+1} - q_s^k\|^2 \\
& + (\frac{1}{\eta_{\mathbf{x}}^2} + \sum_{s=1}^{|\mathcal{P}^k|} 6\|\mathbf{b}_s\|^2 + 6L\tau k_1)\|\mathbf{x}^{k+1} - \mathbf{x}^k\|^2 \\
& + \frac{3}{\eta_{\boldsymbol{\lambda}}^2} \sum_{l=1}^{L} ||\boldsymbol{\lambda}_l^{\bar{k}_l} - \boldsymbol{\lambda}_l^k||^2 + 3 \sum_{l=1}^{L} ((c_{\boldsymbol{\lambda}}^{\hat{k}_l - 1})^2 - (c_{\boldsymbol{\lambda}}^{\bar{k}_l - 1})^2) ||\boldsymbol{\lambda}_l^k||^2 \\
& + \frac{3}{\eta_{\boldsymbol{\alpha}}^2} \sum_{l=1}^{L} ||\boldsymbol{\alpha}_l^{\bar{k}_l} - \boldsymbol{\alpha}_l^k||^2 + 3 \sum_{l=1}^{L} ((c_{\boldsymbol{\alpha}}^{\hat{k}_l - 1})^2 - (c_{\boldsymbol{\alpha}}^{\bar{k}_l - 1})^2) ||\boldsymbol{\alpha}_l^k||^2 \\
& + \frac{3}{\eta_{\boldsymbol{\beta}}^2} \sum_{l=1}^{L} ||\boldsymbol{\beta}_l^{\bar{k}_l} - \boldsymbol{\beta}_l^k||^2 + 3 \sum_{l=1}^{L} ((c_{\boldsymbol{\beta}}^{\hat{k}_l - 1})^2 - (c_{\boldsymbol{\beta}}^{\bar{k}_l - 1})^2) ||\boldsymbol{\beta}_l^k||^2 \\
& + \frac{3}{\eta_{\gamma}^2} \sum_{s=1}^{|\mathcal{P}^k|} ||\gamma_s^{k+1} - \gamma_s^k||^2 + 3 \sum_{s=1}^{|\mathcal{P}^k|} ((c_{\gamma}^{k-1})^2 - (c_{\gamma}^k)^2) ||\gamma_s^k||^2.
\end{aligned}
\tag{107}
$$

Let constant $\underline{a_6}$ denote the lower bound of a_6^k ($\underline{a_6} > 0$), and we set constants d_1, d_2, d_3, d_4, d_5 that,

$$
\begin{aligned}
d_1 &= \frac{k_\tau \tau + 18 k_\tau \tau \eta_{\boldsymbol{\mu}}^2}{{\eta_{\boldsymbol{\mu}}}^2 (\underline{a_6})^2} \geq \frac{k_\tau \tau + 18 k_\tau \tau \eta_{\boldsymbol{\mu}}^2}{{\eta_{\boldsymbol{\mu}}}^2 (a_6^k)^2}, \\
d_2 &= \frac{k_\tau \tau + 24 w_{\mathbf{r}}^2 k_\tau \tau {\eta_{\mathbf{r}}}^2}{{\eta_{\mathbf{r}}}^2 (\underline{a_6})^2} \geq \frac{k_\tau \tau + 24 w_{\mathbf{r}}^2 k_\tau \tau {\eta_{\mathbf{r}}}^2}{{\eta_{\mathbf{r}}}^2 (a_6^k)^2}, \\
d_3 &= \frac{k_\tau \tau + 24 w_{\mathbf{p}}^2 k_\tau \tau {\eta_{\mathbf{p}}}^2}{{\eta_{\mathbf{p}}}^2 (\underline{a_6})^2} \geq \frac{k_\tau \tau + 24 w_{\mathbf{p}}^2 k_\tau \tau {\eta_{\mathbf{p}}}^2}{{\eta_{\mathbf{p}}}^2 (a_6^k)^2}, \\
d_4 &= \frac{1 + 24 w_q^2 {\eta_q}^2}{{\eta_q}^2 (\underline{a_6})^2} \geq \frac{1 + 24 w_q^2 {\eta_q}^2}{{\eta_q}^2 (a_6^k)^2}, \\
d_5 &= \frac{1 + (\sum_{s=1}^{|\mathcal{P}^k|} 6\|\mathbf{b}_s\|^2 + 6L\tau k_1){\eta_{\mathbf{x}}}^2}{{\eta_{\mathbf{x}}}^2 (\underline{a_6})^2} \geq \frac{1 + (\sum_{s=1}^{|\mathcal{P}^k|} 6\|\mathbf{b}_s\|^2 + 6L\tau k_1){\eta_{\mathbf{x}}}^2}{{\eta_{\mathbf{x}}}^2 (a_6^k)^2}.
\end{aligned}
\tag{108}
$$

where k_τ is a positive constant. By employing (107) and (108), we can obtain,

$$
\begin{aligned}
||\nabla \widetilde{G}^k||^2 \leq & (a_6^k)^2 \left(d_1 \sum_{l=1}^{L} \|\boldsymbol{\mu}_l^{k+1} - \boldsymbol{\mu}_l^k\|^2 + d_2 \sum_{l=1}^{L} \|\mathbf{r}_l^{k+1} - \mathbf{r}_l^k\|^2 + d_3 \sum_{l=1}^{L} \|\mathbf{p}_l^{k+1} - \mathbf{p}_l^k\|^2 \right) \\
& + (a_6^k)^2 \left(d_4 \sum_{s=1}^{|\mathcal{P}^k|} \|q_s^{k+1} - q_s^k\|^2 + d_5 \|\mathbf{x}^{k+1} - \mathbf{x}^k\|^2 \right) \\
& + (\frac{1}{\eta_{\boldsymbol{\mu}}^2} + 18) \sum_{l=1}^{L} \|\boldsymbol{\mu}_l^{\bar{k}_l} - \boldsymbol{\mu}_l^k\|^2 - (\frac{1}{\eta_{\boldsymbol{\mu}}^2} + 18) k_\tau \tau \sum_{l=1}^{L} \|\boldsymbol{\mu}_l^{k+1} - \boldsymbol{\mu}_l^k\|^2 \\
& + (\frac{1}{\eta_{\mathbf{r}}^2} + 24 w_{\mathbf{r}}^2) \sum_{l=1}^{L} \|\mathbf{r}_l^{\bar{k}_l} - \mathbf{r}_l^k\|^2 - (\frac{1}{\eta_{\mathbf{r}}^2} + 24 w_{\mathbf{r}}^2) k_\tau \tau \sum_{l=1}^{L} \|\mathbf{r}_l^{k+1} - \mathbf{r}_l^k\|^2 \\
& + (\frac{1}{\eta_{\mathbf{p}}^2} + 24 w_{\mathbf{p}}^2) \sum_{l=1}^{L} \|\mathbf{p}_l^{\bar{k}_l} - \mathbf{p}_l^k\|^2 - (\frac{1}{\eta_{\mathbf{p}}^2} + 24 w_{\mathbf{p}}^2) k_\tau \tau \sum_{l=1}^{L} \|\mathbf{p}_l^{k+1} - \mathbf{p}_l^k\|^2 \\
& + \frac{3}{\eta_{\boldsymbol{\lambda}}^2} \sum_{l=1}^{L} ||\boldsymbol{\lambda}_l^{\bar{k}_l} - \boldsymbol{\lambda}_l^k||^2 + 3 \sum_{l=1}^{L} ((c_{\boldsymbol{\lambda}}^{\hat{k}_l - 1})^2 - (c_{\boldsymbol{\lambda}}^{\bar{k}_l - 1})^2) ||\boldsymbol{\lambda}_l^k||^2 \\
& + \frac{3}{\eta_{\boldsymbol{\alpha}}^2} \sum_{l=1}^{L} ||\boldsymbol{\alpha}_l^{\bar{k}_l} - \boldsymbol{\alpha}_l^k||^2 + 3 \sum_{l=1}^{L} ((c_{\boldsymbol{\alpha}}^{\hat{k}_l - 1})^2 - (c_{\boldsymbol{\alpha}}^{\bar{k}_l - 1})^2) ||\boldsymbol{\alpha}_l^k||^2 \\
& + \frac{3}{\eta_{\boldsymbol{\beta}}^2} \sum_{l=1}^{L} ||\boldsymbol{\beta}_l^{\bar{k}_l} - \boldsymbol{\beta}_l^k||^2 + 3 \sum_{l=1}^{L} ((c_{\boldsymbol{\beta}}^{\hat{k}_l - 1})^2 - (c_{\boldsymbol{\beta}}^{\bar{k}_l - 1})^2) ||\boldsymbol{\beta}_l^k||^2 \\
& + \frac{3}{\eta_{\gamma}^2} \sum\nolimits_{s=1}^{|\mathcal{P}^k|} ||\gamma_s^{k+1} - \gamma_s^k||^2 + 3 \sum\nolimits_{s=1}^{|\mathcal{P}^k|} ((c_{\gamma}^{k-1})^2 - (c_{\gamma}^k)^2) ||\gamma_s^k||^2.
\end{aligned}
\tag{109}
$$

On the basis of (109), we can further obtain that,

$$
\begin{aligned}
d_6^k ||\nabla \widetilde{G}^k||^2 \leq & a_6^k \left(\sum_{l=1}^{L} \|\boldsymbol{\mu}_l^{k+1} - \boldsymbol{\mu}_l^k\|^2 + \sum_{l=1}^{L} \|\mathbf{r}_l^{k+1} - \mathbf{r}_l^k\|^2 + \sum_{l=1}^{L} \|\mathbf{p}_l^{k+1} - \mathbf{p}_l^k\|^2 + \sum_{s=1}^{|\mathcal{P}^k|} \|q_s^{k+1} - q_s^k\|^2 + \|\mathbf{x}^{k+1} - \mathbf{x}^k\|^2 \right) \\
& + d_6^k (\frac{1}{\eta_{\boldsymbol{\mu}}^2} + 18) \sum_{l=1}^{L} \|\boldsymbol{\mu}_l^{\bar{k}_l} - \boldsymbol{\mu}_l^k\|^2 - d_6^k (\frac{1}{\eta_{\boldsymbol{\mu}}^2} + 18) k_\tau \tau \sum_{l=1}^{L} \|\boldsymbol{\mu}_l^{k+1} - \boldsymbol{\mu}_l^k\|^2 \\
& + d_6^k (\frac{1}{\eta_{\mathbf{r}}^2} + 24 w_{\mathbf{r}}^2) \sum_{l=1}^{L} \|\mathbf{r}_l^{\bar{k}_l} - \mathbf{r}_l^k\|^2 - d_6^k (\frac{1}{\eta_{\mathbf{r}}^2} + 24 w_{\mathbf{r}}^2) k_\tau \tau \sum_{l=1}^{L} \|\mathbf{r}_l^{k+1} - \mathbf{r}_l^k\|^2 \\
& + d_6^k (\frac{1}{\eta_{\mathbf{p}}^2} + 24 w_{\mathbf{p}}^2) \sum_{l=1}^{L} \|\mathbf{p}_l^{\bar{k}_l} - \mathbf{p}_l^k\|^2 - d_6^k (\frac{1}{\eta_{\mathbf{p}}^2} + 24 w_{\mathbf{p}}^2) k_\tau \tau \sum_{l=1}^{L} \|\mathbf{p}_l^{k+1} - \mathbf{p}_l^k\|^2 \\
& + \frac{1}{10 \tau \eta_{\boldsymbol{\lambda}}} \sum_{l=1}^{L} ||\boldsymbol{\lambda}_l^{\bar{k}_l} - \boldsymbol{\lambda}_l^k||^2 + 3 d_6^k \sum_{l=1}^{L} ((c_{\boldsymbol{\lambda}}^{\hat{k}_l - 1})^2 - (c_{\boldsymbol{\lambda}}^{\bar{k}_l - 1})^2) ||\boldsymbol{\lambda}_l^k||^2 \\
& + \frac{1}{10 \tau \eta_{\boldsymbol{\alpha}}} \sum_{l=1}^{L} ||\boldsymbol{\alpha}_l^{\bar{k}_l} - \boldsymbol{\alpha}_l^k||^2 + 3 d_6^k \sum_{l=1}^{L} ((c_{\boldsymbol{\alpha}}^{\hat{k}_l - 1})^2 - (c_{\boldsymbol{\alpha}}^{\bar{k}_l - 1})^2) ||\boldsymbol{\alpha}_l^k||^2 \\
& + \frac{1}{10 \tau \eta_{\boldsymbol{\beta}}} \sum_{l=1}^{L} ||\boldsymbol{\beta}_l^{\bar{k}_l} - \boldsymbol{\beta}_l^k||^2 + 3 d_6^k \sum_{l=1}^{L} ((c_{\boldsymbol{\beta}}^{\hat{k}_l - 1})^2 - (c_{\boldsymbol{\beta}}^{\bar{k}_l - 1})^2) ||\boldsymbol{\beta}_l^k||^2 \\
& + \frac{1}{10 \eta_{\gamma}} \sum_{s=1}^{|\mathcal{P}^k|} ||\gamma_s^{k+1} - \gamma_s^k||^2 + 3 d_6^k \sum_{s=1}^{|\mathcal{P}^k|} ((c_{\gamma}^{k-1})^2 - (c_{\gamma}^k)^2) ||\gamma_s^k||^2,
\end{aligned}
\tag{110}
$$

where d_6^k represents a nonnegative sequence, i.e,

$$d_6^k = \left(\max\left\{d_1 a_6^k, d_2 a_6^k, d_3 a_6^k, d_4 a_6^k, d_5 a_6^k, \frac{30\tau}{\eta_{\boldsymbol{\lambda}}}, \frac{30\tau}{\eta_{\boldsymbol{\alpha}}}, \frac{30\tau}{\eta_{\boldsymbol{\beta}}}, \frac{30}{\eta_\gamma}\right\}\right)^{-1}. \tag{111}$$

And the upper and lower bound of d_6 is denoted as $\overline{d_6}$ and $\underline{d_6}$, respectively. According to Assumption 6.3 and combining (99) with (110), we have,

$$\begin{aligned}
d_6^k||\nabla\widetilde{G}^k||^2 \leq\ & F^k - F^{k+1} \\
& + \overline{d_6}(\frac{1}{\eta_{\boldsymbol{\mu}}^2} + 18)\sum_{l=1}^{L}\|\boldsymbol{\mu}_l^{\bar{k}_l} - \boldsymbol{\mu}_l^k\|^2 - \underline{d_6}(\frac{1}{\eta_{\boldsymbol{\mu}}^2} + 18)k_\tau\tau\sum_{l=1}^{L}\|\boldsymbol{\mu}_l^{k+1} - \boldsymbol{\mu}_l^k\|^2 \\
& + \overline{d_6}(\frac{1}{\eta_{\mathbf{r}}^2} + 24w_{\mathbf{r}}^2)\sum_{l=1}^{L}\|\mathbf{r}_l^{\bar{k}_l} - \mathbf{r}_l^k\|^2 - \underline{d_6}(\frac{1}{\eta_{\mathbf{r}}^2} + 24w_{\mathbf{r}}^2)k_\tau\tau\sum_{l=1}^{L}\|\mathbf{r}_l^{k+1} - \mathbf{r}_l^k\|^2 \\
& + \overline{d_6}(\frac{1}{\eta_{\mathbf{p}}^2} + 24w_{\mathbf{p}}^2)\sum_{l=1}^{L}\|\mathbf{p}_l^{\bar{k}_l} - \mathbf{p}_l^k\|^2 - \underline{d_6}(\frac{1}{\eta_{\mathbf{p}}^2} + 24w_{\mathbf{p}}^2)k_\tau\tau\sum_{l=1}^{L}\|\mathbf{p}_l^{k+1} - \mathbf{p}_l^k\|^2 \\
& + \frac{1}{10\tau\eta_{\boldsymbol{\lambda}}}\sum_{l=1}^{L}||\boldsymbol{\lambda}_l^{\bar{k}_l} - \boldsymbol{\lambda}_l^k||^2 - \frac{1}{10\eta_{\boldsymbol{\lambda}}}\sum_{l=1}^{L}||\boldsymbol{\lambda}_l^{k+1} - \boldsymbol{\lambda}_l^k||^2 \\
& + \frac{1}{10\tau\eta_{\boldsymbol{\alpha}}}\sum_{l=1}^{L}||\boldsymbol{\alpha}_l^{\bar{k}_l} - \boldsymbol{\alpha}_l^k||^2 - \frac{1}{10\eta_{\boldsymbol{\alpha}}}\sum_{l=1}^{L}||\boldsymbol{\alpha}_l^{\bar{k}_l} - \boldsymbol{\alpha}_l^k||^2 \\
& + \frac{1}{10\tau\eta_{\boldsymbol{\beta}}}\sum_{l=1}^{L}||\boldsymbol{\beta}_l^{\bar{k}_l} - \boldsymbol{\beta}_l^k||^2 - \frac{1}{10\eta_{\boldsymbol{\beta}}}\sum_{l=1}^{L}||\boldsymbol{\beta}_l^{\bar{k}_l} - \boldsymbol{\beta}_l^k||^2 \\
& + 3\overline{d_6}\sum_{l=1}^{L}((c_{\boldsymbol{\lambda}}^{\hat{k}_l-1})^2 - (c_{\boldsymbol{\lambda}}^{\bar{k}_l-1})^2)||\boldsymbol{\lambda}_l^k||^2 \\
& + 3\overline{d_6}\sum_{l=1}^{L}((c_{\boldsymbol{\alpha}}^{\hat{k}_l-1})^2 - (c_{\boldsymbol{\alpha}}^{\bar{k}_l-1})^2)||\boldsymbol{\alpha}_l^k||^2 \\
& + 3\overline{d_6}\sum_{l=1}^{L}((c_{\boldsymbol{\beta}}^{\hat{k}_l-1})^2 - (c_{\boldsymbol{\beta}}^{\bar{k}_l-1})^2)||\boldsymbol{\beta}_l^k||^2 \\
& + 3\overline{d_6}\sum_{s=1}^{|\mathcal{P}^k|}((c_\gamma^{k-1})^2 - (c_\gamma^k)^2)||\gamma_s^k||^2 \\
& + \frac{c_{\boldsymbol{\lambda}}^{k-1} - c_{\boldsymbol{\lambda}}^k}{2}Mw_{\boldsymbol{\lambda}}^2 + \frac{c_{\boldsymbol{\alpha}}^{k-1} - c_{\boldsymbol{\alpha}}^k}{2}Mw_{\boldsymbol{\alpha}}^2 \\
& + \frac{c_{\boldsymbol{\beta}}^{k-1} - c_{\boldsymbol{\beta}}^k}{2}Mw_{\boldsymbol{\beta}}^2 + \frac{c_\gamma^{k-1} - c_\gamma^k}{2}Pw_\gamma^2 \\
& + \frac{4}{\eta_{\boldsymbol{\lambda}}}(\frac{c_{\boldsymbol{\lambda}}^{k-2}}{c_{\boldsymbol{\lambda}}^{k-1}} - \frac{c_{\boldsymbol{\lambda}}^{k-1}}{c_{\boldsymbol{\lambda}}^k})\sum_{l=1}^{L}||\boldsymbol{\lambda}_l^k||^2 + \frac{4}{\eta_{\boldsymbol{\alpha}}}(\frac{c_{\boldsymbol{\alpha}}^{k-2}}{c_{\boldsymbol{\alpha}}^{k-1}} - \frac{c_{\boldsymbol{\alpha}}^{k-1}}{c_{\boldsymbol{\alpha}}^k})\sum_{l=1}^{L}||\boldsymbol{\alpha}_l^k||^2 \\
& + \frac{4}{\eta_{\boldsymbol{\beta}}}(\frac{c_{\boldsymbol{\beta}}^{k-2}}{c_{\boldsymbol{\beta}}^{k-1}} - \frac{c_{\boldsymbol{\beta}}^{k-1}}{c_{\boldsymbol{\beta}}^k})\sum_{l=1}^{L}||\boldsymbol{\beta}_l^k||^2 + \frac{4}{\eta_\gamma}(\frac{c_\gamma^{k-2}}{c_\gamma^{k-1}} - \frac{c_\gamma^{k-1}}{c_\gamma^k})\sum_{s=1}^{|\mathcal{P}^k|}||\gamma_s^k||^2.
\end{aligned} \tag{112}$$

Denoting $\widetilde{K}(\epsilon)$ as $\widetilde{K}(\epsilon) = \min\{k \mid ||\nabla\widetilde{G}^{K_1+k}||^2 \leq \frac{\epsilon}{4}, k \geq 2\}$. Summing up (112) from $K_1 + 2$ to $K_1 + \widetilde{K}(\epsilon)$, we can

obtain that,

$$
\begin{aligned}
\sum_{k=K_1+2}^{K_1+\widetilde{K}(\epsilon)} d_6^k ||\nabla \widetilde{G}^k||^2 \leq & F^{K_1+2} - F^{K_1+\widetilde{K}(\epsilon)+1} + \frac{c_{\boldsymbol{\lambda}}^1}{2} M w_{\boldsymbol{\lambda}}^2 + \frac{c_{\boldsymbol{\alpha}}^1}{2} M w_{\boldsymbol{\alpha}}^2 + \frac{c_{\boldsymbol{\beta}}^1}{2} M w_{\boldsymbol{\beta}}^2 + \frac{c_{\gamma}^1}{2} P w_{\gamma}^2 \\
& + \frac{4}{\eta_{\boldsymbol{\lambda}}} \frac{c_{\boldsymbol{\lambda}}^0}{c_{\boldsymbol{\lambda}}^1} M w_{\boldsymbol{\lambda}}^2 + \frac{4}{\eta_{\boldsymbol{\alpha}}} \frac{c_{\boldsymbol{\alpha}}^0}{c_{\boldsymbol{\alpha}}^1} M w_{\boldsymbol{\alpha}}^2 + \frac{4}{\eta_{\boldsymbol{\beta}}} \frac{c_{\boldsymbol{\beta}}^0}{c_{\boldsymbol{\beta}}^1} M w_{\boldsymbol{\beta}}^2 + \frac{4}{\eta_{\gamma}} \frac{c_{\gamma}^0}{c_{\gamma}^1} P w_{\gamma}^2 \\
& + \overline{d_6}(\frac{1}{\eta_{\boldsymbol{\mu}}^2} + 18) \sum_{k=K_1+2}^{K_1+\widetilde{K}(\epsilon)} \sum_{l=1}^{L} \|\boldsymbol{\mu}_l^{\bar{k}_l} - \boldsymbol{\mu}_l^k\|^2 - \underline{d_6}(\frac{1}{\eta_{\boldsymbol{\mu}}^2} + 18) k_\tau \tau \sum_{k=K_1+2}^{K_1+\widetilde{K}(\epsilon)} \sum_{l=1}^{L} \|\boldsymbol{\mu}_l^{k+1} - \boldsymbol{\mu}_l^k\|^2 \\
& + \overline{d_6}(\frac{1}{\eta_{\mathbf{r}}^2} + 24 w_{\mathbf{r}}^2) \sum_{k=K_1+2}^{K_1+\widetilde{K}(\epsilon)} \sum_{l=1}^{L} \|\mathbf{r}_l^{\bar{k}_l} - \mathbf{r}_l^k\|^2 - \underline{d_6}(\frac{1}{\eta_{\mathbf{r}}^2} + 24 w_{\mathbf{r}}^2) k_\tau \tau \sum_{k=K_1+2}^{K_1+\widetilde{K}(\epsilon)} \sum_{l=1}^{L} \|\mathbf{r}_l^{k+1} - \mathbf{r}_l^k\|^2 \\
& + \overline{d_6}(\frac{1}{\eta_{\mathbf{p}}^2} + 24 w_{\mathbf{p}}^2) \sum_{k=K_1+2}^{K_1+\widetilde{K}(\epsilon)} \sum_{l=1}^{L} \|\mathbf{p}_l^{\bar{k}_l} - \mathbf{p}_l^k\|^2 - \underline{d_6}(\frac{1}{\eta_{\mathbf{p}}^2} + 24 w_{\mathbf{p}}^2) k_\tau \tau \sum_{k=K_1+2}^{K_1+\widetilde{K}(\epsilon)} \sum_{l=1}^{L} \|\mathbf{p}_l^{k+1} - \mathbf{p}_l^k\|^2 \\
& + \frac{1}{10\tau\eta_{\boldsymbol{\lambda}}} \sum_{k=K_1+2}^{K_1+\widetilde{K}(\epsilon)} \sum_{l=1}^{L} ||\boldsymbol{\lambda}_l^{\bar{k}_l} - \boldsymbol{\lambda}_l^k||^2 - \frac{1}{10\eta_{\boldsymbol{\lambda}}} \sum_{k=K_1+2}^{K_1+\widetilde{K}(\epsilon)} \sum_{l=1}^{L} ||\boldsymbol{\lambda}_l^{k+1} - \boldsymbol{\lambda}_l^k||^2 \\
& + \frac{1}{10\tau\eta_{\boldsymbol{\alpha}}} \sum_{k=K_1+2}^{K_1+\widetilde{K}(\epsilon)} \sum_{l=1}^{L} ||\boldsymbol{\alpha}_l^{\bar{k}_l} - \boldsymbol{\alpha}_l^k||^2 - \frac{1}{10\eta_{\boldsymbol{\alpha}}} \sum_{k=K_1+2}^{K_1+\widetilde{K}(\epsilon)} \sum_{l=1}^{L} ||\boldsymbol{\alpha}_l^{k+1} - \boldsymbol{\alpha}_l^k||^2 \\
& + \frac{1}{10\tau\eta_{\boldsymbol{\beta}}} \sum_{k=K_1+2}^{K_1+\widetilde{K}(\epsilon)} \sum_{l=1}^{L} ||\boldsymbol{\beta}_l^{\bar{k}_l} - \boldsymbol{\beta}_l^k||^2 - \frac{1}{10\eta_{\boldsymbol{\beta}}} \sum_{k=K_1+2}^{K_1+\widetilde{K}(\epsilon)} \sum_{l=1}^{L} ||\boldsymbol{\beta}_l^{k+1} - \boldsymbol{\beta}_l^k||^2 \\
& + 3\overline{d_6} \sum_{k=K_1+2}^{K_1+\widetilde{K}(\epsilon)} \sum_{l=1}^{L} ((c_{\boldsymbol{\lambda}}^{\hat{k}_l-1})^2 - (c_{\boldsymbol{\lambda}}^{\bar{k}_l-1})^2) ||\boldsymbol{\lambda}_l^k||^2 + 3\overline{d_6} \sum_{k=K_1+2}^{K_1+\widetilde{K}(\epsilon)} \sum_{l=1}^{L} ((c_{\boldsymbol{\alpha}}^{\hat{k}_l-1})^2 - (c_{\boldsymbol{\alpha}}^{\bar{k}_l-1})^2) ||\boldsymbol{\alpha}_l^k||^2 \\
& + 3\overline{d_6} \sum_{k=K_1+2}^{K_1+\widetilde{K}(\epsilon)} \sum_{l=1}^{L} ((c_{\boldsymbol{\beta}}^{\hat{k}_l-1})^2 - (c_{\boldsymbol{\beta}}^{\bar{k}_l-1})^2) ||\boldsymbol{\beta}_l^k||^2 + 3\overline{d_6} (c_{\gamma}^1)^2 P w_{\gamma}^2.
\end{aligned}
\tag{113}
$$

For each worker l, we have that $\overline{k_l} - \hat{k}_l \leq \tau$, thus,

$$
\begin{aligned}
& 3\overline{d_6} \sum_{k=K_1+2}^{K_1+\widetilde{K}(\epsilon)} \sum_{l=1}^{L} ((c_{\boldsymbol{\lambda}}^{\hat{k}_l-1})^2 - (c_{\boldsymbol{\lambda}}^{\bar{k}_l-1})^2) ||\boldsymbol{\lambda}_l^k||^2 \leq 3\tau\overline{d_6} (c_{\boldsymbol{\lambda}}^1)^2 M w_{\boldsymbol{\lambda}}^2, \\
& 3\overline{d_6} \sum_{k=K_1+2}^{K_1+\widetilde{K}(\epsilon)} \sum_{l=1}^{L} ((c_{\boldsymbol{\alpha}}^{\hat{k}_l-1})^2 - (c_{\boldsymbol{\alpha}}^{\bar{k}_l-1})^2) ||\boldsymbol{\alpha}_l^k||^2 \leq 3\tau\overline{d_6} (c_{\boldsymbol{\alpha}}^1)^2 M w_{\boldsymbol{\alpha}}^2, \\
& 3\overline{d_6} \sum_{k=K_1+2}^{K_1+\widetilde{K}(\epsilon)} \sum_{l=1}^{L} ((c_{\boldsymbol{\beta}}^{\hat{k}_l-1})^2 - (c_{\boldsymbol{\beta}}^{\bar{k}_l-1})^2) ||\boldsymbol{\beta}_l^k||^2 \leq 3\tau\overline{d_6} (c_{\boldsymbol{\beta}}^1)^2 M w_{\boldsymbol{\beta}}^2.
\end{aligned}
\tag{114}
$$

In our asynchronous algorithm, inactive workers do not update their variables in each master iteration, Thus, for any k which satisfies $\hat{v}_l(j-1) \leq k < \hat{v}_l(j)$, we have $\boldsymbol{\mu}_l^k = \boldsymbol{\mu}_l^{\hat{v}_l(j)-1}$. And for $k \notin \mathcal{V}_l(K)$, we have $\|\boldsymbol{\mu}_l^k - \boldsymbol{\mu}_l^{k-1}\|^2 = 0$. Since $\hat{v}_l(j) - \hat{v}_l(j-1) \leq \tau$, we can obtain

$$
\sum_{k=K_1+2}^{K_1+\widetilde{K}(\epsilon)} \sum_{l=1}^{L} \|\boldsymbol{\mu}_l^{\bar{k}_l} - \boldsymbol{\mu}_l^k\|^2 \leq \tau \sum_{k=K_1+2}^{K_1+\widetilde{K}(\epsilon)} \sum_{l=1}^{L} ||\boldsymbol{\mu}_l^{k+1} - \boldsymbol{\mu}_l^k||^2 + 4\tau(\tau-1) M w_{\boldsymbol{\mu}}^2. \tag{115}
$$

Similarly, we can obtain

$$\begin{aligned}
&\sum_{k=K_1+2}^{K_1+\widetilde{K}(\epsilon)}\sum_{l=1}^{L}\|\mathbf{r}_l^{\bar{k}_l}-\mathbf{r}_l^k\|^2 \le \tau\sum_{k=K_1+2}^{K_1+\widetilde{K}(\epsilon)}\sum_{l=1}^{L}||\mathbf{r}_l^{k+1}-\mathbf{r}_l^k||^2+4\tau(\tau-1)Mw_{\mathbf{r}}^2,\\
&\sum_{k=K_1+2}^{K_1+\widetilde{K}(\epsilon)}\sum_{l=1}^{L}\|\mathbf{p}_l^{\bar{k}_l}-\mathbf{p}_l^k\|^2 \le \tau\sum_{k=K_1+2}^{K_1+\widetilde{K}(\epsilon)}\sum_{l=1}^{L}||\mathbf{p}_l^{k+1}-\mathbf{p}_l^k||^2+4\tau(\tau-1)Mw_{\mathbf{p}}^2,\\
&\sum_{k=K_1+2}^{K_1+\widetilde{K}(\epsilon)}\sum_{l=1}^{L}\|\boldsymbol{\lambda}_l^{\bar{k}_l}-\boldsymbol{\lambda}_l^k\|^2 \le \tau\sum_{k=K_1+2}^{K_1+\widetilde{K}(\epsilon)}\sum_{l=1}^{L}||\boldsymbol{\lambda}_l^{k+1}-\boldsymbol{\lambda}_l^k||^2+4\tau(\tau-1)Mw_{\boldsymbol{\lambda}}^2,\\
&\sum_{k=K_1+2}^{K_1+\widetilde{K}(\epsilon)}\sum_{l=1}^{L}\|\boldsymbol{\alpha}_l^{\bar{k}_l}-\boldsymbol{\alpha}_l^k\|^2 \le \tau\sum_{k=K_1+2}^{K_1+\widetilde{K}(\epsilon)}\sum_{l=1}^{L}||\boldsymbol{\alpha}_l^{k+1}-\boldsymbol{\alpha}_l^k||^2+4\tau(\tau-1)Mw_{\boldsymbol{\alpha}}^2,\\
&\sum_{k=K_1+2}^{K_1+\widetilde{K}(\epsilon)}\sum_{l=1}^{L}\|\boldsymbol{\beta}_l^{\bar{k}_l}-\boldsymbol{\beta}_l^k\|^2 \le \tau\sum_{k=K_1+2}^{K_1+\widetilde{K}(\epsilon)}\sum_{l=1}^{L}||\boldsymbol{\beta}_l^{k+1}-\boldsymbol{\beta}_l^k||^2+4\tau(\tau-1)Mw_{\boldsymbol{\beta}}^2.
\end{aligned} \tag{116}$$

We set the value of k_τ to satisfy that,

$$k_\tau \ge \max\left\{\frac{\overline{d_6}(\frac{1}{\eta_{\boldsymbol{\mu}}^2}+18)}{\underline{d_6}(\frac{1}{\overline{\eta_{\boldsymbol{\mu}}}^2}+18)},\frac{\overline{d_6}(\frac{1}{\eta_{\mathbf{r}}^2}+24w_{\mathbf{r}}^2)}{\underline{d_6}(\frac{1}{\overline{\eta_{\mathbf{r}}}^2}+24w_{\mathbf{r}}^2)},\frac{\overline{d_6}(\frac{1}{\eta_{\mathbf{p}}^2}+24w_{\mathbf{p}}^2)}{\underline{d_6}(\frac{1}{\overline{\eta_{\mathbf{p}}}^2}+24w_{\mathbf{p}}^2)}\right\}, \tag{117}$$

where $\overline{\eta_{\boldsymbol{\mu}}}$, $\overline{\eta_{\mathbf{r}}}$, and $\overline{\eta_{\mathbf{p}}}$ are the upper bounds of $\eta_{\boldsymbol{\mu}}^k$, $\eta_{\mathbf{r}}^k$, and $\eta_{\mathbf{p}}^k$, respectively. By employing (113), (114), (115) (116) and (117), we can obtain

$$\begin{aligned}
\sum_{k=K_1+2}^{K_1+\widetilde{K}(\epsilon)} d_6^k||\nabla\widetilde{G}^k||^2 \le{}& F^{K_1+2}-F^{K_1+\widetilde{K}(\epsilon)+1}\\
&+\frac{c_{\boldsymbol{\lambda}}^1}{2}Mw_{\boldsymbol{\lambda}}^2+\frac{c_{\boldsymbol{\alpha}}^1}{2}Mw_{\boldsymbol{\alpha}}^2+\frac{c_{\boldsymbol{\beta}}^1}{2}Mw_{\boldsymbol{\beta}}^2+\frac{c_{\gamma}^1}{2}Pw_{\gamma}^2\\
&+\frac{4}{\eta_{\boldsymbol{\lambda}}}\frac{c_{\boldsymbol{\lambda}}^0}{c_{\boldsymbol{\lambda}}^1}Mw_{\boldsymbol{\lambda}}^2+\frac{4}{\eta_{\boldsymbol{\alpha}}}\frac{c_{\boldsymbol{\alpha}}^0}{c_{\boldsymbol{\alpha}}^1}Mw_{\boldsymbol{\alpha}}^2+\frac{4}{\eta_{\boldsymbol{\beta}}}\frac{c_{\boldsymbol{\beta}}^0}{c_{\boldsymbol{\beta}}^1}Mw_{\boldsymbol{\beta}}^2+\frac{4}{\eta_{\gamma}}\frac{c_{\gamma}^0}{c_{\gamma}^1}Pw_{\gamma}^2\\
&+3\tau\overline{d_6}(c_{\boldsymbol{\lambda}}^1)^2Mw_{\boldsymbol{\lambda}}^2+3\tau\overline{d_6}(c_{\boldsymbol{\alpha}}^1)^2Mw_{\boldsymbol{\alpha}}^2+3\tau\overline{d_6}(c_{\boldsymbol{\beta}}^1)^2Mw_{\boldsymbol{\beta}}^2+3\overline{d_6}(c_{\gamma}^1)^2Pw_{\gamma}^2\\
&+\Bigg(\frac{2Mw_{\boldsymbol{\lambda}}^2}{5\eta_{\boldsymbol{\lambda}}}+\frac{2Mw_{\boldsymbol{\alpha}}^2}{5\eta_{\boldsymbol{\alpha}}}+\frac{2Mw_{\boldsymbol{\beta}}^2}{5\eta_{\boldsymbol{\beta}}}+4\overline{d_6}(\frac{1}{\eta_{\boldsymbol{\mu}}^2}+18)Mw_{\boldsymbol{\mu}}^2\tau\\
&+4\overline{d_6}(\frac{1}{\eta_{\mathbf{r}}^2}+24w_{\mathbf{r}}^2)Mw_{\mathbf{r}}^2\tau+4\overline{d_6}(\frac{1}{\eta_{\mathbf{p}}^2}+24w_{\mathbf{p}}^2)Mw_{\mathbf{p}}^2\tau\Bigg)(\tau-1)\\
&=(d_7+k_d\tau)(\tau-1).
\end{aligned} \tag{118}$$

where d_7 and k_d are constants. Set constant d_8 as

$$d_8=\left(\max\left\{d_1,d_2,d_3,d_4,d_5,\frac{30\tau}{\eta_{\boldsymbol{\lambda}}\underline{a_6}},\frac{30\tau}{\eta_{\boldsymbol{\alpha}}\underline{a_6}},\frac{30\tau}{\eta_{\boldsymbol{\beta}}\underline{a_6}},\frac{30}{\eta_{\gamma}\underline{a_6}}\right\}\right)\ge\frac{1}{d_6^k a_6^k}. \tag{119}$$

Thus, we can obtain that

$$\sum_{k=K_1+2}^{K_1+\tilde{K}(\epsilon)}\frac{1}{d_8a_6^k}\|\nabla\widetilde{G}^{K_1+\tilde{K}(\epsilon)}\|^2\le\sum_{k=K_1+2}^{K_1+\tilde{K}(\epsilon)}\frac{1}{d_8a_6^k}\|\nabla\widetilde{G}^k\|^2\le\sum_{k=K_1+2}^{K_1+\tilde{K}(\epsilon)}d_6^k\|\nabla\widetilde{G}^k\|^2\le(d_7+k_d\tau)(\tau-1). \tag{120}$$

We can further obtain

$$||\nabla\widetilde{G}^{K_1+\widetilde{K}(\epsilon)}||^2\le\frac{(d_7+k_d\tau)(\tau-1)d_8}{\sum_{k=K_1+2}^{K_1+\tilde{K}}\frac{1}{a_6^k}}. \tag{121}$$

According to the setting of $c_{\boldsymbol{\lambda}}^k$, $c_{\boldsymbol{\alpha}}^k$, $c_{\boldsymbol{\beta}}^k$ and c_{γ}^k, we have,

$$\frac{1}{a_6^k} \geq \frac{1}{8(\xi-2)(k+1)^{\frac{1}{2}} \min\{\eta_{\boldsymbol{\lambda}}+\eta_{\boldsymbol{\alpha}}+\eta_{\boldsymbol{\beta}}, 4w_{\mathbf{r}}^2\eta_{\boldsymbol{\lambda}}, 4w_{\mathbf{p}}^2\eta_{\boldsymbol{\alpha}}, 4w_{\mathbf{q}}^2\eta_{\gamma}, \eta_{\boldsymbol{\beta}}+\eta_{\gamma}\sum_{s=1}^{|\mathcal{P}^k|}\|b_s\|^2\}}. \tag{122}$$

Summing up a_6^k from $k = K_1+2$ to $k = K_1+\widetilde{K}$, we have

$$\sum_{k=K_1+2}^{K_1+\widetilde{K}(\epsilon)} \frac{1}{a_6^k} \geq \frac{(K_1+\widetilde{K}(\epsilon))^{\frac{1}{2}} - (K_1+2)^{\frac{1}{2}}}{8(\xi-2)\min\{\eta_{\boldsymbol{\lambda}}+\eta_{\boldsymbol{\alpha}}+\eta_{\boldsymbol{\beta}}, 4w_{\mathbf{r}}^2\eta_{\boldsymbol{\lambda}}, 4w_{\mathbf{p}}^2\eta_{\boldsymbol{\alpha}}, 4w_{\mathbf{q}}^2\eta_{\gamma}, \eta_{\boldsymbol{\beta}}+\eta_{\gamma}\sum_{s=1}^{|\mathcal{P}^k|}\|b_s\|^2\}}. \tag{123}$$

Recall that $\widetilde{K}(\epsilon) = \min\{k \mid ||\nabla\widetilde{G}^{K_1+k}||^2 \leq \frac{\epsilon}{4}, k \geq 2\}$. Therefore, by employing (123) and (121), when

$$K_1 + \widetilde{K}(\epsilon) \geq (\frac{d_9(d_7+k_d\tau)(\tau-1)d_8}{\epsilon} + (K_1+2)^{\frac{1}{2}})^2, \tag{124}$$

the value of $||\nabla\widetilde{G}^{K_1+\widetilde{K}(\epsilon)}||^2$ can be guaranteed to be smaller than $\frac{\epsilon}{4}$, where

$$d_9 = 32(\xi-2)\min\{\eta_{\boldsymbol{\lambda}}+\eta_{\boldsymbol{\alpha}}+\eta_{\boldsymbol{\beta}}, 4w_{\mathbf{r}}^2\eta_{\boldsymbol{\lambda}}, 4w_{\mathbf{p}}^2\eta_{\boldsymbol{\alpha}}, 4w_{\mathbf{q}}^2\eta_{\gamma}, \eta_{\gamma}+\eta_{\gamma}\sum_{s=1}^{L}\|b_s\|^2\}. \tag{125}$$

Combining the definition of ∇G^k and $\nabla\widetilde{G}^k$ with trigonometric inequality, we then get:

$$||\nabla G^k|| - ||\nabla\widetilde{G}^k|| \leq ||\nabla G^k - \nabla\widetilde{G}^k|| \leq \sqrt{\sum_{l=1}^{L}||c_{\boldsymbol{\lambda}}^{k-1}\boldsymbol{\lambda}_l^k||^2 + \sum_{l=1}^{L}||c_{\boldsymbol{\alpha}}^{k-1}\boldsymbol{\alpha}_l^k||^2 + + \sum_{l=1}^{L}||c_{\boldsymbol{\beta}}^{k-1}\boldsymbol{\beta}_l^k||^2 + + \sum_{s=1}^{|\mathcal{P}^k|}||c_{\gamma}^{k-1}\gamma_s^k||^2}. \tag{126}$$

If $k \geq 16(\frac{Mw_{\boldsymbol{\lambda}}^2}{\eta_{\lambda}^2} + \frac{Mw_{\boldsymbol{\alpha}}^2}{\eta_{\boldsymbol{\alpha}}^2} + \frac{Mw_{\boldsymbol{\beta}}^2}{\eta_{\boldsymbol{\beta}}^2} + \frac{Pw_{\gamma}^2}{\eta_{\gamma}^2})^2\frac{1}{\epsilon^2}$, we have

$$\sqrt{\sum_{l=1}^{L}||c_{\boldsymbol{\lambda}}^{k-1}\boldsymbol{\lambda}_l^k||^2 + \sum_{l=1}^{L}||c_{\boldsymbol{\alpha}}^{k-1}\boldsymbol{\alpha}_l^k||^2 + + \sum_{l=1}^{L}||c_{\boldsymbol{\beta}}^{k-1}\boldsymbol{\beta}_l^k||^2 + + \sum_{s=1}^{|\mathcal{P}^k|}|||c_{\gamma}^{k-1}\gamma_s^k||^2} \leq \frac{\sqrt{\epsilon}}{2}. \tag{127}$$

Combining it with (124), we can conclude that

$$K(\epsilon) \sim \mathcal{O}\left(\max\left\{(16(\frac{Mw_{\boldsymbol{\lambda}}^2}{\eta_{\lambda}^2} + \frac{Mw_{\boldsymbol{\alpha}}^2}{\eta_{\boldsymbol{\alpha}}^2} + \frac{Mw_{\boldsymbol{\beta}}^2}{\eta_{\boldsymbol{\beta}}^2} + \frac{Pw_{\gamma}^2}{\eta_{\gamma}^2})^2\frac{1}{\epsilon^2}, (\frac{d_9(d_7+k_d\tau)(\tau-1)d_8}{\epsilon} + (K_1+2)^{\frac{1}{2}})^2\right\}\right). \tag{128}$$

which concludes our proof. □

D. Proof of Theorem 6.1

In Algorithm 1, cutting plane set are updated every w iteration, i.e.,

$$\mathcal{P}^0 \supseteq \mathcal{P}^w \supseteq \cdots \supseteq \mathcal{P}^{nw}. \tag{129}$$

The feasible region of problem (17) in the wth iteration is represented as $\mathcal{R}^w$, while the feasible region of problem (15) is represented as $\mathcal{R}^*$. As such, we have

$$\mathcal{R}^0 \supseteq \mathcal{R}^w \supseteq \cdots \supseteq \mathcal{R}^{nw} \supseteq \mathcal{R}^*. \tag{130}$$

Let F^{w*} represent the optimal objective value of the problem (17) in the wth iteration and let $F^* \leq 0$ represent the optimal objective value of the problem in (15). Based on (130), we can obtain

$$F^{0*} \leq F^{w*} \leq \cdots \leq F^{nw*} \leq F^*. \tag{131}$$

Thus, we have

$$\frac{F^*}{F^{0*}} \leq \frac{F^*}{F^{w*}} \leq \cdots \leq \frac{F^*}{F^{nw*}} \leq \Omega, \tag{132}$$

where $\Omega \leq 1$. It can be seen from (132) that as the number of cutting planes increases, the sequence is monotonically non-decreasing. When $nw \to \infty$, $\frac{F^*}{F^{nw*}}$ monotonically converges to Ω.