Decentralized Nonlinear Policy Evaluation Methods for Multi-Agent Reinforcement Learning

Xin Zhang[†] Zhuqing Liu[‡] Jia Liu[‡] Zhengyuan Zhu[*]
[*]Department of Statistics, Iowa State University
[‡]Department of Electrical and Computer Engineering, The Ohio State University

Abstract

Multi-agent reinforcement learning (MARL) has received increasing attention in recent years and has found applications in many scientific and engineering fields. One of the key components in reinforcement learning algorithms (e.g., the actor-critic method) is the policy evaluation problem, whose goal is to evaluate the long-term accumulative reward for a given policy. In this paper, we focus on MARL policy evaluation with smooth nonlinear function approximation over a given transition dataset. We first show that the empirical MARL policy evaluation problem can be reformulated as a decentralized nonconvex-strongly-concave minimax saddle point problem. We then develop a decentralized gradient-based descent ascent algorithm, as well as two decentralized stochastic optimization algorithms enhanced by variance reduction techniques. We show that these algorithms all enjoy an $\mathcal{O}(1/T)$ convergence rate to a stationary point of the minimax problem. We further provide a detailed analysis of sample and communication complexities for these algorithms. In the end, we perform extensive experiments to show the performance of proposed algorithms and verify our theoretical findings.

1 Introduction

Reinforcement learning (RL) is a foundational learning paradigm that has attracted a significant amount of attention in recent years [1]. In the RL framework, an agent interacts with a shared environment and repeats the tasks of observing the current state, performing a policy-based action, receiving rewards, and transition into the next state. The goal is to find an optimal policy to maximize the long-term accumulative reward received through interacting with the environment. A key component in most RL algorithms is the policy evaluation problem, whose goal is to evaluate the expected long-term accumulative reward for a given policy. This problem emerges as an essential step for the agent to find an optimal policy in RL tasks [2, 3]. For example, in the actor-critic learning algorithmic framework, the actor conducts the policy improvement step, while the critic performs the policy evaluation step and estimates the value function. The whole algorithm tries to find an optimal policy by iterating between the policy evaluation and improvement steps. Thus, developing efficient policy evaluation algorithms is critical to the success of RL algorithms based on the actor-critic framework.

In this work, we focus on the policy evaluation problem in a multi-agent reinforcement learning (MARL) framework – an important and challenging RL subfield. MARL has found important applications in many scientific and engineering fields, such as sensor network [4, 5, 6], robotic network [7, 8, 9] and power network [10, 11, 12, 13], etc. In MARL, *multiple* agents observe the current joint state over a network, perform their own actions based on the current state, and transition to the next joint state. Each agent can only observe its local reward, which is a function of the joint states and actions. In this paper, we focus on *cooperative* MARL, where the agents share a common goal to find an optimal global policy to achieve the maximum global accumulative reward [14].

However, developing efficient algorithms for MARL policy evaluation is highly non-trivial. On one hand, the global accumulative reward is not directly observable in an MARL system. A natural idea is to set aside a centralized server node, which collects the local rewards from all agents. However, this approach is not scalable or even feasible in applications with large MARL systems and vulnerable to adversarial attacks. Moreover, due to privacy concerns, the agents are unable to share the local rewards information, which renders the centralized-node approach infeasible. On the other hand, modern MARL tasks have been increasingly complex and often not directly computable. As a result, MARL tasks often use highly nonlinear parametric models (e.g., deep neural network (DNN)) for policy approximation. However, with such nonlinear function approximations, it is often hard to guarantee the convergence performance of RL algorithms [15]. In light of the growing importance of MARL, in this paper, we focus on addressing the above challenges. The main results and contributions of this paper are summarized as follows:

- Via the Fenchel's duality and local reward decomposition, we reformulate the problem as a decentralized non-convex-strongly-concave minimax saddle point problem. To solve the minimax problem in a decentralized fashion, we first propose a gradient-tracking based gradient descent-ascent (GT-GDA) algorithm. We show that GT-GDA enjoys a convergence rate of $\mathcal{O}(1/T)$, which leads to $\mathcal{O}(\epsilon^{-2})$ communication complexity and $\mathcal{O}(mn\epsilon^{-2})$ sample complexity.
- To further reduce the sample complexity, we develop two stochastic variance reduced algorithms, namely gradient-tracking stochastic recursive variance reduction (GT-SRVR) algorithm and its variant with incremental batch size (GT-SRVRI). We show that GT-SRVR and GT-SRVRI achieve the same communication complexity $\mathcal{O}(\epsilon^{-2})$ as GT-GDA, but requiring a lower sample complexity $\mathcal{O}(m\sqrt{n}\epsilon^{-2})$.
- It is worth noting that, in our theoretical analysis, we relax the commonly-used compactness conditions of the feasible set with some mild assumptions on objectives. Thus, the solutions found by our algorithms are exactly the stationary points for the original policy evaluation problem. This result may be of independent interest for general RL problems.

The rest of the paper is organized as follows. In Section 2, we first provide the preliminaries of the multi-agent policy evaluation problem and discuss related works. In Section 3, we first present the GT-GDA algorithm and its convergence guarantees. Then, we propose two stochastic variance reduced algorithms, namely GT-SRVR and GT-SRVRI, and present their theoretical properties in Section 4. Section 5 provides numerical results to verify our theoretical findings, and Section 6 concludes this paper.

2 Preliminaries

To facilitate our technical discussions, in Section 2.1, we first provide an overview on multi-agent Markov decision process (MDP) and formally define the policy evaluation problem in a decentralized multi-agent system. Then, in Section 2.2, we review the recent developments of policy evaluation algorithms and compare them with our work.

2.1 Multi-Agent MDP and Decentralized Policy Evaluation: A Primer

Consider a multi-agent network system $\mathcal{G} = (\mathcal{N}, \mathcal{L})$, where $\mathcal{N}$ and $\mathcal{L}$ denote the sets of agents and edges, respectively, with $|\mathcal{N}| = m$. In the system, the agents cooperatively to perform a learning task. The agents can communicate with each other through edges in $\mathcal{L}$. An MARL problem is formulated based on the multi-agent Markov decision process (MDP), which is characterized by a quintuple $(\mathcal{S}, \mathcal{A}, \mathcal{P}^{\mathbf{a}}_{\mathbf{ss}'}, \{\mathcal{R}_i(\mathbf{s}, \mathbf{a})\}_{i=1}^m, \gamma)$, where $\mathcal{S}$ and $\mathcal{A}$ are the state and action spaces, respectively; $\mathbf{s} \in \mathcal{S}$ and $\mathbf{a} \in \mathcal{A}$ are joint state and action; $\mathcal{P}^{\mathbf{a}}_{\mathbf{ss}'}$ is the transition probability from state $\mathbf{s}$ to state $\mathbf{s}'$ after taking action $\mathbf{a}$; $\mathcal{R}_i(\mathbf{s}, \mathbf{a})$ is the local reward received by agent i after taking action $\mathbf{a}$ in state $\mathbf{s}$; and $\gamma \in (0, 1)$ is a discount factor. Both the joint state $\mathbf{s}$ and action $\mathbf{a}$ are available to all agents, while the local reward $\mathcal{R}_i$ is private to agent i. In the multi-agent system, the global reward function is defined as the average of the local rewards $\frac{1}{m}\sum_{i=1}^m \mathcal{R}_i(\mathbf{s}, \mathbf{a})$. Moreover, a joint policy π specifies sequential decision rules for all agents. Policy $\pi(\mathbf{a}|\mathbf{s})$ is the conditional probability of taking joint action $\mathbf{a}$ given state $\mathbf{s}$. The goal of policy evaluation is to estimate the value function of a given policy π, which is defined as the long-term discounted accumulative reward as follows:

$$\mathcal{V}^\pi(\mathbf{s}_0) = \mathbb{E}\left[\frac{1}{m}\sum_{t=0}^{\infty}\gamma^t \sum_{i=1}^{m}\mathcal{R}_i(\mathbf{s}_t, \mathbf{s}_{t+1})|\mathbf{s}_0, \pi\right], \tag{1}$$

where the expectation is taken over all possible state-action trajectories given policy π and initial state $\mathbf{s}_0$.

To determine $\mathcal{V}^\pi(\cdot)$, one of the most effective methods is the temporal-difference (TD) learning algorithm, which focuses on solving the Bellman equation for $\mathcal{V}^\pi(\cdot)$:

$$\mathcal{V}(\mathbf{s}) = \mathcal{T}^\pi \mathcal{V}(\mathbf{s}) \triangleq \frac{1}{m}\sum_{i=1}^{m}\mathcal{R}_i^\pi(\mathbf{s}) + \gamma \sum_{s'\in\mathcal{S}} \mathcal{P}^\pi_{\mathbf{ss}'}\mathcal{V}(\mathbf{s}'), \tag{2}$$

where $\mathcal{T}^\pi$ denotes the Bellman operator, $\mathcal{R}_i^\pi(\mathbf{s}) = \mathbb{E}_{\mathbf{a}\sim\pi(\cdot|\mathbf{s})}\mathcal{R}(\mathbf{s}, \mathbf{a})$ and $\mathcal{P}^\pi_{\mathbf{ss}'} = \mathbb{E}_{\mathbf{a}\sim\pi(\cdot|\mathbf{s})}\mathcal{P}^a_{\mathbf{ss}'}$. However, $\mathcal{P}^\pi_{\mathbf{ss}'}$ is in general unknown in RL and the size of the state space $\mathcal{S}$ could be infinite. Thus, it is difficult to find the value function that satisfies Eq. (2). To address this challenge, a widely adopted approach is to approximate $\mathcal{V}^\pi(\cdot)$ by a function $\mathcal{V}_{\boldsymbol{\theta}}(\cdot)$ parameterized by $\boldsymbol{\theta} \in \mathbb{R}^p$. According to the formulation in [16, 17], the Bellman equation can be solved by minimizing the following mean-squared projected bellman error (MSPBE):

$$\text{MSPBE}(\boldsymbol{\theta}) \triangleq \frac{1}{2}\left\|\mathbb{E}_{\mathbf{s}\sim d^\pi}\left[\left(\mathcal{T}^\pi \mathcal{V}_{\boldsymbol{\theta}}(\mathbf{s}) - \mathcal{V}_{\boldsymbol{\theta}}\right)\nabla_{\boldsymbol{\theta}}\mathcal{V}_{\boldsymbol{\theta}}(\mathbf{s})^\top\right]\right\|^2_{\mathbf{K}_{\boldsymbol{\theta}}^{-1}},$$

where $\mathbf{K}_{\boldsymbol{\theta}} = \mathbb{E}_{\mathbf{s}\sim d^{\pi}}[\nabla_{\boldsymbol{\theta}}\mathcal{V}_{\boldsymbol{\theta}}(\mathbf{s})\nabla_{\boldsymbol{\theta}}\mathcal{V}_{\boldsymbol{\theta}}(\mathbf{s})^{\top}] \in \mathbb{R}^{p\times p}$ and d^{π} is the stationary distribution of the MDP under policy π. From the Fenchel's duality $\|\mathbf{x}\|^2_{\mathbf{A}^{-1}} = \max_{\mathbf{y}\in\mathbb{R}^p} 2\langle \mathbf{x}, \mathbf{y}\rangle - \mathbf{y}^{\top}\mathbf{A}\mathbf{y}$, we can reformulate the MSPBE minimization problem as the following primal-dual minimax problem:

$$\min_{\boldsymbol{\theta}\in\mathbb{R}^p}\max_{\boldsymbol{\omega}\in\mathbb{R}^p} \mathcal{L}(\boldsymbol{\theta},\boldsymbol{\omega}) \triangleq \mathbb{E}\big[\langle \boldsymbol{\delta}\cdot\nabla_{\boldsymbol{\theta}}\mathcal{V}_{\boldsymbol{\theta}}(\mathbf{s}), \boldsymbol{\omega}\rangle - \frac{1}{2}\boldsymbol{\omega}^{\top}[\nabla_{\boldsymbol{\theta}}\mathcal{V}_{\boldsymbol{\theta}}(\mathbf{s})\nabla_{\boldsymbol{\theta}}\mathcal{V}_{\boldsymbol{\theta}}(\mathbf{s})^{\top}]\boldsymbol{\omega}\big], \quad (3)$$

where the expectation is taken over $\mathbf{s} \sim d^{\pi}(\cdot)$, $\mathbf{a} \sim \pi(\cdot|\mathbf{s})$, $\mathbf{s}' \sim \mathcal{P}^{\mathbf{a}}_{\mathbf{s}\cdot}$, and $\boldsymbol{\delta} = \frac{1}{m}\sum_{i=1}^m \mathcal{R}_i(\mathbf{s},\mathbf{a}) + \gamma\mathcal{V}_{\boldsymbol{\theta}}(\mathbf{s}') - \mathcal{V}_{\boldsymbol{\theta}}(\mathbf{s})$. In practice, we only have access to a finite dataset with n-step trajectories $\mathcal{D} = \big\{(\mathbf{s}_t, \mathbf{a}_t, \{\mathcal{R}_i(\mathbf{s}_t,\mathbf{a}_t)\}_{i=1}^n, \mathbf{s}_{t+1})\big\}_{t=0}^n$. By replacing the unknown expectation with the finite sample average, we have the following empirical minimax problem:

$$\min_{\boldsymbol{\theta}\in\mathbb{R}^p}\max_{\boldsymbol{\omega}\in\mathbb{R}^p} F(\boldsymbol{\theta},\boldsymbol{\omega}) = \frac{1}{m}\sum_{t=1}^m \langle \boldsymbol{\delta}_t\cdot\nabla_{\boldsymbol{\theta}}\mathcal{V}_{\boldsymbol{\theta}}(\mathbf{s}_t), \boldsymbol{\omega}\rangle - \frac{1}{2}\boldsymbol{\omega}^{\top}\widehat{\mathbf{K}}_{\boldsymbol{\theta}}\boldsymbol{\omega}, \quad (4)$$

where $\boldsymbol{\delta}_t \triangleq \frac{1}{m}\sum_{i=1}^m \mathcal{R}_i(\mathbf{s}_t,\mathbf{a}_t) + \gamma\mathcal{V}_{\boldsymbol{\theta}}(\mathbf{s}_{t+1}) - \mathcal{V}_{\boldsymbol{\theta}}(\mathbf{s}_t)$ and $\widehat{\mathbf{K}}_{\boldsymbol{\theta}} \triangleq \frac{1}{n}\sum_{t=1}^n \nabla_{\boldsymbol{\theta}}\mathcal{V}_{\boldsymbol{\theta}}(\mathbf{s}_t)\nabla_{\boldsymbol{\theta}}\mathcal{V}_{\boldsymbol{\theta}}(\mathbf{s}_t)^{\top}$. In this paper, we assume that both $\mathbf{K}_{\boldsymbol{\theta}}$ and its empirical estimate $\hat{\mathbf{K}}_{\boldsymbol{\theta}}$ are positive definite matrices for all $\boldsymbol{\theta}$. Define function $J(\boldsymbol{\theta})$ as:

$$J(\boldsymbol{\theta}) \triangleq F(\boldsymbol{\theta},\boldsymbol{\omega}^*) = \max_{\boldsymbol{\omega}\in\mathbb{R}^p} F(\boldsymbol{\theta},\boldsymbol{\omega}), \quad (5)$$

where $\boldsymbol{\omega}^* = \arg\max_{\boldsymbol{\omega}\in\mathbb{R}^p} F(\boldsymbol{\theta},\boldsymbol{\omega})$. $J(\boldsymbol{\theta})$ can be viewed as the finite empirical version of MSPBE. Here, we aim to minimize $J(\boldsymbol{\theta})$ by finding a stationary point of $F(\boldsymbol{\theta},\boldsymbol{\omega})$.

Recall that under our multi-agent MDP framework, the local reward is only observable for each individual agent and cannot be aggregated to a "centralized" agent. Thus, it is hard to obtain the global reward $\frac{1}{m}\sum_{i=1}^m \mathcal{R}_i(\mathbf{s}_t,\mathbf{a}_t)$ and $\boldsymbol{\delta}_t$ in a multi-agent network. To address this challenge, we define $\boldsymbol{\delta}_{i,t} = \mathcal{R}_i(\mathbf{s}_t,\mathbf{a}_t) + \gamma\mathcal{V}_{\boldsymbol{\theta}}(\mathbf{s}_{t+1}) - \mathcal{V}_{\boldsymbol{\theta}}(\mathbf{s}_t)$ and decompose the minimax problem in (4) as follows:

$$\min_{\boldsymbol{\theta}\in\mathbb{R}^p}\max_{\boldsymbol{\omega}\in\mathbb{R}^p} F(\boldsymbol{\theta},\boldsymbol{\omega}) = \frac{1}{m}\sum_{i=1}^m F_i(\boldsymbol{\theta},\boldsymbol{\omega}) = \frac{1}{mn}\sum_{i=1}^m\sum_{t=1}^n f_{ij}(\boldsymbol{\theta},\boldsymbol{\omega}), \quad (6)$$

where $f_{ij}(\boldsymbol{\theta},\boldsymbol{\omega}) \triangleq \langle \boldsymbol{\delta}_{i,t}\cdot\nabla_{\boldsymbol{\theta}}\mathcal{V}_{\boldsymbol{\theta}}(\mathbf{s}_t), \boldsymbol{\omega}\rangle - \frac{1}{2}\boldsymbol{\omega}^{\top}[\nabla_{\boldsymbol{\theta}}\mathcal{V}_{\boldsymbol{\theta}}(\mathbf{s}_t)\,\nabla_{\boldsymbol{\theta}}\mathcal{V}_{\boldsymbol{\theta}}(\mathbf{s}_t)^{\top}]\boldsymbol{\omega}$. We call this step as local reward decomposition. Further, to solve Problem (6) in a decentralized fashion, we can rewrite it in the following equivalent form:

$$\begin{aligned}&\min_{\{\boldsymbol{\theta}_i\}_{i=1}^m}\max_{\{\boldsymbol{\omega}_i\}_{i=1}^m} \frac{1}{m}\sum_{i=1}^m F_i(\boldsymbol{\theta}_i,\boldsymbol{\omega}_i) = \frac{1}{mn}\sum_{i=1}^m\sum_{t=1}^n f_{ij}(\boldsymbol{\theta}_i,\boldsymbol{\omega}_i),\\ &\text{subject to}\quad \boldsymbol{\theta}_i = \boldsymbol{\theta}_j, \boldsymbol{\omega}_i = \boldsymbol{\omega}_j,\ \ \forall (i,j)\in\mathcal{L},\end{aligned} \quad (7)$$

where $\boldsymbol{\theta}_i$ and $\boldsymbol{\omega}_i$ are the local copies of the original primal-dual parameters at agent i. In (7), the equality constraint ensures that the local copies at all nodes are equal to each other, so the formulation is termed as a "consensus" problem. Clearly, Problems (6) and (7) are equivalent. For a fixed $\boldsymbol{\theta}$, each local function $F_i(\boldsymbol{\theta}_i, \cdot)$ is a strongly concave function of $\boldsymbol{\omega}$. For a fixed $\boldsymbol{\omega}$, $F_i(\cdot, \boldsymbol{\omega})$ is a non-convex function of $\boldsymbol{\theta}$. Therefore, Problem (7) is a decentralized non-convex-strongly-concave minimax consensus optimization problem.

In this paper, we adopt two complexity metrics that are widely used in the decentralized optimization literature (e.g., [18]) to measure the efficiency of an algorithm:

Definition 1 (Sample Complexity)**.** *The sample complexity is defined as the total number of the incremental first-order oracle (IFO) calls required across all the nodes until algorithm converges, where one IFO call evaluates a pair of* $(f_{ij}(\boldsymbol{\theta}, \boldsymbol{\omega}), \nabla f_{ij}(\boldsymbol{\theta}, \boldsymbol{\omega}))$ *at node* i.

Definition 2 (Communication Complexity)**.** *The communication complexity is defined as the total rounds of communications required until algorithm converges, where each node can send and receive a* p*-dimensional vector with its neighboring nodes in one communication round.*

The goal of MARL policy evaluation is to develop efficient decentralized algorithms for solving Problem (7) with both low sample and communication complexities.

2.2 Related Work

1) The Tabular Approach: The study of MARL under the MDP formalization traces its roots to the seminal work by [19]. Motivated by this formalization, several methods have been developed to solve and analyze MARL problems, including [20, 21, 22, 23], etc. However, most of these works approximate the value function in a tabular form. This approach only works for cases where the state and action spaces are relatively small so that the value function can be stored in tables. For complex MARL tasks where the state space is large or even infinite, the tabular approach becomes intractable.

2) Policy Evaluation with Linear Function Approximation: To address such challenge, [28] proposed to estimate the value function with a linear approximation under the multi-agent MDP framework. They developed a distributed gradient temporal-difference (DGTD) learning algorithm, in which the agents can exchange the local information through a (sparse) communication network. However, they only provided asymptotic convergence analysis of DGTD requires diminishing step sizes. In [26], the authors proposed a distributed homotopy primal-dual algorithm (DHPD) for the MARL policy evaluation problem. Also, they cast the MSPBE minimization as a stochastic primal-dual optimization problem, where the objective is convex in primal variables and strongly-concave in dual variables. By using an adaptive restarting scheme (homotopy method), DHPD achieved an $\mathcal{O}(1/T)$ convergence rate in finding stationary points. [29] developed a distributed consensus-based TD(0) algorithm, which integrates the network consensus step and local TD(0) updates. They provided a finite-time analysis and showed that the convergence rate of their algorithm was $\mathcal{O}(1/T)$. To further improve the convergence performance, [27] proposed a primal-dual distributed

Table 1: Comparisons among existing policy evaluation algorithms, where m is the number of agents; n is the size of dataset; ϵ^2 is the convergence error. Our proposed algorithms are marked in bold.

ALGORITHM	REFERENCE	DECENTRALIZED MULTI-AGENT	NONLINEAR APPROXIMATION	CONVEX SETS[1]	SAMPLE COMPLEXITY	COMMUN. COMPLEXITY
PDBG	[24]	✗	✗	✗	$\mathcal{O}(n\log\epsilon^{-2})$	-
SVRG		✗	✗	✗	$\mathcal{O}(n\log\epsilon^{-2})$	-
SAGA		✗	✗	✗	$\mathcal{O}(n\log\epsilon^{-2})$	-
STSG	[25]	✗	✓	✓	$\mathcal{O}(\epsilon^{-4})$	-
ASTSG		✗	✓	✓	$\mathcal{O}(\epsilon^{-3})$	-
DHPD	[26]	✓	✗	✓	$\mathcal{O}(mn\epsilon^{-2})$	$\mathcal{O}(\epsilon^{-2})$
PD-DISTIAG	[27]	✓	✗	✗	$\mathcal{O}(m\log\epsilon^{-2})$	$\mathcal{O}(\log\epsilon^{-2})$
GT-GDA	THEOREM 1	✓	✓	✗	$\mathcal{O}(mn\epsilon^{-2})$	$\mathcal{O}(\epsilon^{-2})$
GT-SRVR	THEOREM 3	✓	✓	✗	$\mathcal{O}(m\sqrt{n}\epsilon^{-2})$	$\mathcal{O}(\epsilon^{-2})$
GT-SRVRI	THEOREM 4					

[1] The feasible parameter spaces are required to be closed convex sets.

incremental aggregated gradient (PD-DistIAG) method, where they adopted the gradient-tracking technique and an incremental aggregated gradient method to reduce the errors of network consensus and stochastic gradient approximation, respectively. They showed that PD-DistIAG converges linearly. However, all these works approximate the value function by linear functions, i.e., $\{\mathcal{V}(\mathbf{s}) \approx \boldsymbol{\phi}(\mathbf{s})^\top\boldsymbol{\theta}, \boldsymbol{\theta} \in \mathbb{R}^p\}$, where $\boldsymbol{\phi}$ denotes a feature mapping $\boldsymbol{\phi} : \mathcal{S} \to \mathbb{R}^p$.

3) Policy Evaluation with Nonlinear Function Approximation: Motivated by the success of deep neural network models, a few recent works consider nonlinear function approximation for policy evaluation. To our knowledge, [30] were the first to study policy evaluation problem with nonlinear approximation and developed a nonlinear TD algorithm. However, the proposed algorithms adopted two-timescale step-sizes, resulting in a slow convergence performance. Recently, [17] showed that the nonlinear approximation policy evaluation problem is equivalent to a non-convex-strongly-concave minimax optimization problem. To find a stationary point for such minimax problem, they proposed a non-convex primal-dual gradient with variance reduced (nPD-VR) algorithm. However, their algorithm require an $\mathcal{O}(1/m)$-step-size, where m is the size of the dataset. This is problematic in cases with a massive amount of transition data. More recently, the authors of [25] proposed two single-timescale first order stochastic algorithms for the nonconvex-strongly-concave minimax optimization. Their two proposed algorithms using the techniques of stochastic gradient with momentum and variance-reduced momentum [31], and achieved $\mathcal{O}(1/\sqrt{T})$ and $\mathcal{O}(1/T^{2/3})$ convergence rates, respectively. However, diminishing step-sizes are required. It is worth noting that all these works focus on the single-agent policy evaluation problem. In this work, we will develop efficient algorithms for the

multi-agent policy evaluation with nonlinear approximation.

3 Gradient-Tracking Gradient Descent Ascent Algorithm.

In this section, we first present a gradient-tracking gradient descent ascent (GT-GDA) method for solving the multi-agent policy evaluation problem in (7). Then, we will provide the main theoretical results for the GT-GDA algorithm.

1) The Algorithm: To solve Problem (7) and reach a consensus, a common approach in the literature is to let agents aggregate neighboring information through a consensus weight matrix $\mathbf{M} \in \mathbb{R}^{m\times m}$. We let $[\mathbf{M}]_{ij}$ represent the element in the i-th row and the j-th column in $\mathbf{M}$. The consensus matrix $\mathbf{M}$ should satisfy the following properties:

(a) *Doubly stochastic:* $\sum_{i=1}^{m}[\mathbf{M}]_{ij} = \sum_{j=1}^{m}[\mathbf{M}]_{ij} = 1$.

(b) *Symmetric:* $[\mathbf{M}]_{ij} = [\mathbf{M}]_{ji}, \forall i, j \in \mathcal{N}$.

(c) *Network-Defined Sparsity:* $[\mathbf{M}]_{ij} > 0$ if $(i, j) \in \mathcal{L}$; otherwise $[\mathbf{M}]_{ij} = 0, \forall i, j \in \mathcal{N}$.

The above properties imply that the eigenvalues of $\mathbf{M}$ are real and can be sorted as $-1 < \lambda_m(\mathbf{M}) \leq \cdots \leq \lambda_2(\mathbf{M}) < \lambda_1(\mathbf{M}) = 1$. We define the second-largest eigenvalue in magnitude of $\mathbf{M}$ as $\lambda \triangleq \max\{|\lambda_2(\mathbf{M})|, |\lambda_m(\mathbf{M})|\}$ for the notation convenience. Later, λ plays an important role in the step-size selection and the algorithm's convergence rate.

Our GT-GDA algorithm for each agent i is illustrated in Algorithm 1. Specifically, in the tth iteration, agent i first calculates the local full gradients as follows:

$$\mathbf{v}_{i,t} = \nabla_{\boldsymbol{\theta}} F_i(\boldsymbol{\theta}_{i,t}, \boldsymbol{\omega}_{i,t}), \quad \mathbf{u}_{i,t} = \nabla_{\boldsymbol{\omega}} F_i(\boldsymbol{\theta}_{i,t}, \boldsymbol{\omega}_{i,t}). \tag{8}$$

Note that $\mathbf{v}_{i,t}$ and $\mathbf{u}_{i,t}$ only contain the gradient information of the local objective function $F_i(\boldsymbol{\theta}, \boldsymbol{\omega})$. Thus, merely updating with $\mathbf{v}_{i,t}$ and $\mathbf{u}_{i,t}$ cannot guarantee the convergence of the global objective function $F(\boldsymbol{\theta}, \boldsymbol{\omega})$. To address this challenge, we introduce two auxiliary variables, $\mathbf{p}_{i,t}$ and $\mathbf{d}_{i,t}$. The agent updates the two variables by performing the following local weighted aggregation:

$$\mathbf{p}_{i,t} = \textstyle\sum_{j\in\mathcal{N}_i}[\mathbf{M}]_{ij}\mathbf{p}_{j,t-1} + \mathbf{v}_{i,t} - \mathbf{v}_{i,t-1}, \tag{9a}$$

$$\mathbf{d}_{i,t} = \textstyle\sum_{j\in\mathcal{N}_i}[\mathbf{M}]_{ij}\mathbf{d}_{j,t-1} + \mathbf{u}_{i,t} - \mathbf{u}_{i,t-1} \tag{9b}$$

where $\mathcal{N}_i \triangleq \{j \in \mathcal{N}, : (i, j) \in \mathcal{L}\}$ denotes agent i's set of neighbors. Technically, $\mathbf{p}_{i,t}$ and $\mathbf{d}_{i,t}$ track the directions of global gradients. With some derivations, it can be shown that $\sum_{i\in\mathcal{N}} \mathbf{p}_{i,t} = \sum_{i\in\mathcal{N}} \nabla_{\boldsymbol{\theta}} F_i(\boldsymbol{\theta}_{i,t}, \boldsymbol{\omega}_{i,t})$ and $\sum_{i\in\mathcal{N}} \mathbf{d}_{i,t} = \sum_{i\in\mathcal{N}} \nabla_{\boldsymbol{\omega}} F_i(\boldsymbol{\theta}_{i,t}, \boldsymbol{\omega}_{i,t})$. Lastly, each agent updates local parameters following the conventional decentralized gradient descent and ascent steps as follows:

$$\boldsymbol{\theta}_{i,t+1} = \textstyle\sum_{j\in\mathcal{N}_i}[\mathbf{M}]_{ij}\boldsymbol{\theta}_{j,t} - \gamma\mathbf{p}_{i,t}, \tag{10a}$$

$$\boldsymbol{\omega}_{i,t+1} = \textstyle\sum_{j\in\mathcal{N}_i}[\mathbf{M}]_{ij}\boldsymbol{\omega}_{j,t} + \eta\mathbf{d}_{i,t}, \tag{10b}$$

Algorithm 1 GT-GDA Algorithm at Agent i.

1: Set prime-dual parameter pair $(\boldsymbol{\theta}_{i,0}, \boldsymbol{\omega}_{i,0}) = (\boldsymbol{\theta}^0, \boldsymbol{\omega}^0)$.
2: Calculate local gradients as $\mathbf{p}_{i,0} = \nabla_{\boldsymbol{\theta}} F_i(\boldsymbol{\theta}_{i,0}, \boldsymbol{\omega}_{i,0})$, and $\mathbf{d}_{i,0} = \nabla_{\boldsymbol{\omega}} F_i(\boldsymbol{\theta}_{i,0}, \boldsymbol{\omega}_{i,0})$;
3: **for** $t = 1, \cdots, T$ **do**
4: Update local parameters $(\boldsymbol{\theta}_{i,t+1}, \boldsymbol{\omega}_{i,t+1})$ as in (10);
5: Calculate local gradients $(\mathbf{v}_{i,t+1}, \mathbf{u}_{i,t+1})$ as in (8);
6: Track global gradients $(\mathbf{p}_{i,t+1}, \mathbf{d}_{i,t+1})$ as in (9);
7: **end for**

where the constants γ and η are the step-sizes for updating the individual primal and dual variables, respectively.

2) Theoretical Results of GT-GDA: In this section, we will establish the convergence behaviours of the proposed GT-GDA algorithm. Toward this end, we first state several assumptions as follows:

Assumption 1. *The function $F(\boldsymbol{\theta}, \boldsymbol{\omega}) = \frac{1}{m}\sum_{i=1}^m F_i(\boldsymbol{\theta}, \boldsymbol{\omega})$ and $J(\boldsymbol{\theta}) = \max_{\boldsymbol{\omega}\in\mathbb{R}^p} F(\boldsymbol{\theta}, \boldsymbol{\omega})$ satisfy:*

(a) (Boundness from Below): There exists a finite lower bound $J^* = \inf_{\boldsymbol{\theta}} J(\boldsymbol{\theta}) > -\infty$;

(b) (Lipschitz Smoothness): Local objective function $F_i(\boldsymbol{\theta}, \cdot)$ is L_F-Lipschitz smooth, i.e., there exists a positive constant L_F such that the gradient $\nabla F_i(\boldsymbol{\theta}, \boldsymbol{\omega}) = [\nabla_{\boldsymbol{\theta}} F_i(\boldsymbol{\theta}, \boldsymbol{\omega})^\top, \nabla_{\boldsymbol{\omega}} F_i(\boldsymbol{\theta}, \boldsymbol{\omega})^\top]^\top$ satisfies $\|\nabla F_i(\boldsymbol{\theta}, \boldsymbol{\omega}) - \nabla F_i(\boldsymbol{\theta}', \boldsymbol{\omega}')\|^2 \le L_F^2\|\boldsymbol{\theta} - \boldsymbol{\theta}'\|^2 + L_F^2\|\boldsymbol{\omega} - \boldsymbol{\omega}'\|^2, \forall\, \boldsymbol{\theta}, \boldsymbol{\theta}', \boldsymbol{\omega}, \boldsymbol{\omega}' \in \mathbb{R}^p, i \in [m]$;

(c) (Strong Concavity): Local objective function $F_i(\boldsymbol{\theta}, \cdot)$ is μ-strongly concave for fixed $\boldsymbol{\theta} \in \mathbb{R}^p$, i.e., there exists a positive constant μ such that $\|\nabla_{\boldsymbol{\omega}} F_i(\boldsymbol{\theta}, \boldsymbol{\omega}) - \nabla_{\boldsymbol{\omega}} F_i(\boldsymbol{\theta}, \boldsymbol{\omega}')\| \ge \mu\|\boldsymbol{\omega} - \boldsymbol{\omega}'\|, \forall\, \boldsymbol{\theta}, \boldsymbol{\omega}, \boldsymbol{\omega}' \in \mathbb{R}^p, i \in [m]$

(d) (Bounded Dual Maximizer): For any primal variable $\boldsymbol{\theta} \in \mathbb{R}^p$, its associated dual maximizer $\boldsymbol{\omega}^*(\boldsymbol{\theta}) \triangleq \arg\max_{\boldsymbol{\omega}\in\mathbb{R}^p} F(\boldsymbol{\theta}, \boldsymbol{\omega})$ is bounded, i.e., $\|\boldsymbol{\omega}^*(\boldsymbol{\theta})\| < \infty$;

(e) (Bounded Gradient at Maximum): The partial derivative at maximum point $\nabla_{\boldsymbol{\theta}} F(\boldsymbol{\theta}, \boldsymbol{\omega}^*(\boldsymbol{\theta}))$ is bounded, i.e., $\|\nabla_{\boldsymbol{\theta}} F(\boldsymbol{\theta}, \boldsymbol{\omega}^*(\boldsymbol{\theta}))\| < \infty$, $\forall\, \boldsymbol{\theta} \in \mathbb{R}^p$.

In these assumptions, (a) and (b) are standard in the literature; it can be verified that (c) holds when the matrix $\widehat{\mathbf{K}}_{\boldsymbol{\theta}}$ is positive definite; (d)-(e) are two assumptions to guarantee $\nabla J(\boldsymbol{\theta}) = \nabla_{\boldsymbol{\theta}} F(\boldsymbol{\theta}, \boldsymbol{\omega}^*(\boldsymbol{\theta}))$ (see Lemma 10 the supplementary material). Most of the existing works [32, 25] adopt the compactness assumption to ensure such gradient equivalence. However, the compactness assumption restricts the feasible parameter space as a closed convex set, which might not include the optimal solution of (3). Although we make boundedness assumptions, the convergence performance of our algorithm is independent of the upper bound of $\|\boldsymbol{\omega}^*(\boldsymbol{\theta})\|$ and $\|\nabla_{\boldsymbol{\theta}} F(\boldsymbol{\theta}, \boldsymbol{\omega}^*(\boldsymbol{\theta}))\|$. To quantify the convergence rate, we use the following metric:

$$\begin{aligned}\mathfrak{M}_t \triangleq & \|\nabla J(\bar{\boldsymbol{\theta}}_t)\|^2 + 2\mathbb{E}\|\boldsymbol{\omega}_t^* - \bar{\boldsymbol{\omega}}_t\|^2 \\ & + \frac{1}{m}\sum_{i=1}^m \left(\|\boldsymbol{\theta}_{i,t} - \bar{\boldsymbol{\theta}}_t\|^2 + \|\boldsymbol{\omega}_{i,t} - \bar{\boldsymbol{\omega}}_t\|^2\right),\end{aligned} \tag{11}$$

where $\boldsymbol{\omega}_t^*$ denotes $\boldsymbol{\omega}^*(\bar{\boldsymbol{\theta}}_t) = \arg\max_{\boldsymbol{\omega}\in\mathbb{R}^p} F(\bar{\boldsymbol{\theta}}_t, \boldsymbol{\omega})$. The first term in (11) measures the convergence of primal variable $\boldsymbol{\theta}$: $\|\nabla J(\bar{\boldsymbol{\theta}}_t)\|^2 = 0$ indicates that $\bar{\boldsymbol{\theta}}_t$ is a first-order stationary point for $J(\cdot)$. The second term in (11) measures $\bar{\boldsymbol{\omega}}_t$'s convergence to the unique maximizer $\boldsymbol{\omega}_t^*$ for $F(\bar{\boldsymbol{\theta}}_t, \cdot)$. The last term in (11) is the average consensus error of local copies. Based on the metric in (11), we have the following:

Theorem 1 (Convergence of GT-GDA)**.** *Under Assumption 1, if the step-sizes satisfy that $\kappa \triangleq \gamma/\eta \leq \mu^2/13L_F^2$ and $\eta \leq \min\{k_1, k_2, k_3, k_4\}$, then GT-GDA has the following convergence result:*

$$\frac{1}{T+1}\sum_{t=0}^{T}\mathbb{E}[\mathfrak{M}_t] \leq \frac{2\mathbb{E}[\mathfrak{P}_0 - \mathfrak{P}_{T+1}]}{\min\{1, L_F^2\}(T+1)\gamma},$$

where $\mathfrak{P}_t$ is the potential function defined as:

$$\begin{aligned}\mathfrak{P}_t \triangleq J(\bar{\boldsymbol{\theta}}_t) + \frac{8\gamma L_F^2}{\mu\eta}\|\bar{\boldsymbol{\omega}}_t - \boldsymbol{\omega}_t^*\|^2 + \frac{1}{m}\sum_{i=1}^{m}\|\boldsymbol{\theta}_{i,t} - \bar{\boldsymbol{\theta}}_t\|^2 \\ + \|\boldsymbol{\omega}_{i,t} - \bar{\boldsymbol{\omega}}_t\|^2 + \gamma\|\mathbf{p}_{i,t} - \bar{\mathbf{p}}_t\|^2 + \eta\|\mathbf{d}_{i,t} - \bar{\mathbf{d}}_t\|^2,\end{aligned}$$

and the constants in the step-size η are as follows:

$$\begin{aligned}k_1 &= \frac{13L_F^2}{2\mu^2}\Big(L_F + \frac{L_F^2}{\mu} + (1-\lambda)\Big),\\ k_2 &= \frac{13L_F^2}{\mu^2\big(1/2 + 1/(1-\lambda)^2\big)}, k_3 = \frac{(1-\lambda)}{6\mu(1+1/\kappa)},\\ k_4 &= \frac{26(1-\lambda)L_F^2}{(\mu^2 + 144L_F^4 + 4L_F^2\mu^2 + \frac{48\mu^2L_F^2(1+1/\kappa)}{1-\lambda})}.\end{aligned}$$

Remark 1. In Theorem 1, the step-sizes and convergence rate are affected by the network topology: for a sparse network, λ is close to (but not exactly) one (recall that $\lambda = \max\{|\lambda_2|, |\lambda_m|\} < 1$). Therefore, k_2 and k_4 are close to zero in this case. Also, the ratio of the step-sizes $\kappa \triangleq \gamma/\eta$ is required to be a non-zero constant. Either too small or too large value of κ might affect the primal or dual convergence of the algorithm. This limitation is due to the consensus error in the decentralized training. In practice, one can first determine κ and then select η and γ.

Based on Theorem 1, we have the following main sample and communication complexity results for GT-GDA:

Corollary 2. *Under the same conditions in Theorem 1, to achieve an ϵ^2-stationary solution, i.e., $\frac{1}{T+1}\sum_{t=0}^{T}\mathbb{E}[\mathfrak{M}_t] \leq \epsilon^2$, the total communication rounds are in the order of $\mathcal{O}(\epsilon^{-2})$ and the total samples evaluated across the network system is in the order of $\mathcal{O}(mn\epsilon^{-2})$.*

4 Gradient-Tracking Stochastic Variance Reduction Algorithms

In the GT-GDA algorithm, agents need to evaluate local full gradients in each iteration, which results in a high sample complexity when the trajectory length n is large. This limitation motivates us to leverage

the stochastic recursive variance-reduced approach (e.g., [33]) to achieve low sample complexity in MARL policy evaluation.

1) The Algorithms: We first propose an algorithm called gradient-tracking stochastic recursive variance reduction (GT-SRVR) algorithm. In iteration t, at agent i, GT-SRVR estimates the local gradient with the following estimators:

$$\mathbf{v}_{i,t} = \begin{cases} \nabla_{\boldsymbol{\theta}} F_i(\boldsymbol{\theta}_{i,t}, \boldsymbol{\omega}_{i,t}), & \text{if } \mod(t,q)=0, \\ \mathbf{v}_{i,t-1} + \frac{1}{|\mathcal{S}_{i,t}|}\sum_{j\in\mathcal{S}_{i,t}}\big(\nabla_{\boldsymbol{\theta}} f_{ij}(\boldsymbol{\theta}_{i,t}, \boldsymbol{\omega}_{i,t}) & \\ \quad -\nabla_{\boldsymbol{\theta}} f_{ij}(\boldsymbol{\theta}_{i,t-1}, \boldsymbol{\omega}_{i,t-1})\big), & \text{o.w.}, \end{cases} \tag{12a}$$

$$\mathbf{u}_{i,t} = \begin{cases} \nabla_{\boldsymbol{\omega}} F_i(\boldsymbol{\theta}_{i,t}, \boldsymbol{\omega}_{i,t}), & \text{if } \mod(t,q)=0, \\ \mathbf{u}_{i,t-1} + \frac{1}{|\mathcal{S}_{i,t}|}\sum_{j\in\mathcal{S}_{i,t}}\big(\nabla_{\boldsymbol{\omega}} f_{ij}(\boldsymbol{\theta}_{i,t}, \boldsymbol{\omega}_{i,t}) & \\ \quad -\nabla_{\boldsymbol{\omega}} f_{ij}(\boldsymbol{\theta}_{i,t-1}, \boldsymbol{\omega}_{i,t-1})\big), & \text{o.w.}, \end{cases} \tag{12b}$$

where $\mathcal{S}_{i,t}$ is a local subsample in the tth iteration for agent i. In (12), the algorithm evaluates full gradients $\nabla F_i(\boldsymbol{\theta}_{i,t}, \boldsymbol{\omega}_{i,t})$ only every q steps. For other iterations with $\mod(t,q) \neq 0$, the algorithm estimates the local gradients with a mini-batch of gradients $\frac{1}{|\mathcal{S}_{i,t}|}\sum_{j\in\mathcal{S}_{i,t}}\nabla_{\boldsymbol{\omega}} f_{ij}(\boldsymbol{\theta}_{i,t}, \boldsymbol{\omega}_{i,t})$ and a recursive correction term $\mathbf{u}_{i,t-1} - \frac{1}{|\mathcal{S}_{i,t}|}\sum_{j\in\mathcal{S}_{i,t}}\nabla_{\boldsymbol{\omega}} f_{ij}(\boldsymbol{\theta}_{i,t-1}, \boldsymbol{\omega}_{i,t-1})$. It will be shown later that thanks to the periodic full gradient and recursive correction term, GT-SRVR is able to achieve the same convergence rate and communication complexity as GT-GDA. Moreover, because of the $\mathcal{S}_{i,t}$ subsampling, GT-SRVR has a lower sample complexity than GT-GDA. The full description of GT-SRVR is shown in Algorithm 2.

Note that in GT-SRVR, full gradients are still required for every q steps, which may incur high computational cost. Also, in the initialization phase (before the main loop), agents need to evaluate full gradients, which could be time-consuming. To address these limitations, we propose an enhanced version of GT-SRVR called GT-SRVR with `Incremental` batch size (GT-SRVR`I`). Specifically, we modify the gradient estimators in (12a) and (12b) for the tth iteration with $\mod(t,q)=0$ as follows :

$$\mathbf{v}_{i,t} = \frac{1}{|\mathcal{R}_{i,t}|}\sum_{j\in\mathcal{R}_{i,t}}\nabla_{\boldsymbol{\theta}} f_{ij}(\boldsymbol{\theta}_{i,t}, \boldsymbol{\omega}_{i,t}), \tag{13a}$$

$$\mathbf{u}_{i,t} = \frac{1}{|\mathcal{R}_{i,t}|}\sum_{j\in\mathcal{R}_{i,t}}\nabla_{\boldsymbol{\omega}} f_{ij}(\boldsymbol{\theta}_{i,t}, \boldsymbol{\omega}_{i,t}), \tag{13b}$$

where $\mathcal{R}_{i,t}$ is a subsample set (sampling without replacement), whose size is chosen as $|\mathcal{R}_{i,t}| = \min\{(t/q+1)^{\alpha}q, c_\epsilon\epsilon^{-2}, n\}$. Here, $\alpha > 0$ is a constant, ϵ is a desired convergence error, and $c_\epsilon > 0$ is a constant that depends on ϵ. Our design of $|\mathcal{R}_{i,t}|$ is motivated by the fact that the periodic full gradient evaluation only plays an important role in the later stage of the convergence process for achieving high accuracy. Later, we will see that under some mild assumptions and parameter settings, GT-SRVR`I` has similar convergence performance as GT-SRVR. The full description of GT-SRVR/GT-SRVR`I` is shown in Algorithm 2.

Remark 2. In GT-SRVR`I`, we increase the batch-size as the iteration index increases. We note that [34] also proposed a batch-size adaptation scheme based on the historical gradient information. Similar idea can

be also adopted in our algorithms, but it requires the exact value of the stochastic gradient variance σ^2 for batch-size selection, as well as extra memory cost to store the history-gradient information. We leave this extension in our future studies.

Algorithm 2 GT-SRVR/GT-SRVRI Algorithm at Agent i.

If GT-SRVR: $|\mathcal{R}_{i,t}| = n$, $|\mathcal{S}_{i,t}| = q$; If GT-SRVRI: $|\mathcal{R}_{i,t}| = \min\{(t/q+1)^{\alpha}q, c_{\epsilon}\epsilon^{-2}, n\}$, $|\mathcal{S}_{i,t}| = q$.

1: Set prime-dual parameter pair $(\boldsymbol{\theta}_{i,0}, \boldsymbol{\omega}_{i,0}) = (\boldsymbol{\theta}^0, \boldsymbol{\omega}^0)$.

2: Draw $\mathcal{R}_{i,0}$ samples without replacement and calculate local stochastic gradient estimators as $\mathbf{p}_{i,0} = \mathbf{v}_{i,0} = \frac{1}{|\mathcal{R}_{i,0}|}\sum_{j\in\mathcal{R}_{i,0}}\nabla_{\boldsymbol{\theta}} f_{ij}(\boldsymbol{\theta}_{i,0}, \boldsymbol{\omega}_{i,0})$, and $\mathbf{d}_{i,0} = \mathbf{u}_{i,0} = \frac{1}{|\mathcal{R}_{i,0}|}\sum_{j\in\mathcal{R}_{i,0}}\nabla_{\boldsymbol{\omega}} f_{ij}(\boldsymbol{\theta}_{i,0}, \boldsymbol{\omega}_{i,0})$;

3: **for** $t = 1, \cdots, T$ **do**

4: Update local parameters $(\boldsymbol{\theta}_{i,t+1}, \boldsymbol{\omega}_{i,t+1})$ as in (10);

5: Calculate local gradient estimators $(\mathbf{v}_{i,t+1}, \mathbf{u}_{i,t+1})$ as in (12);

6: Track global gradients $(\mathbf{p}_{i,t+1}, \mathbf{d}_{i,t+1})$ as in (9);

7: **end for**

2) Theoretical Results of GT-SRVR/GT-SRVRI: Now, we establish the convergence of GT-SRVR/GT-SRVRI. First, we replace Assumption 1(b) with the following individual Lipschitz smoothness assumption:

Assumption 2 (Individual Lipschitz smoothness)**.** *The function $f_{ij}(\boldsymbol{\theta}, \cdot)$ is L_f-Lipschitz smooth, i.e., there exists a constant $L_f > 0$, such that the gradient $\nabla f_{ij}(\boldsymbol{\theta}, \boldsymbol{\omega}) = [\nabla_{\boldsymbol{\theta}} f_{ij}(\boldsymbol{\theta}, \boldsymbol{\omega}]^{\top}, \nabla_{\boldsymbol{\omega}} f_{ij}(\boldsymbol{\theta}, \boldsymbol{\omega})^{\top})^{\top}$ satisfies $\|\nabla f_{ij}(\boldsymbol{\theta}, \boldsymbol{\omega}) - \nabla f_{ij}(\boldsymbol{\theta}', \boldsymbol{\omega}')\|^2 \leq L_f^2\|\boldsymbol{\theta} - \boldsymbol{\theta}'\|^2 + L_F^2\|\boldsymbol{\omega} - \boldsymbol{\omega}'\|^2$, $\forall\, \boldsymbol{\theta}, \boldsymbol{\theta}', \boldsymbol{\omega}, \boldsymbol{\omega}' \in \mathbb{R}^p, i \in [m], j \in [n]$.*

We note that Assumption 2 is a common assumption analyzing stochastic variance reduced methods [33, 18, 25]. Further, we have the following assumption only for GT-SRVRI algorithm:

Assumption 3 (Bounded Variance)**.** *There exists a constant $\sigma^2 > 0$, such that $\mathbb{E}\|\nabla f_{ij}(\boldsymbol{\theta}, \boldsymbol{\omega}) - \nabla F_i(\boldsymbol{\theta}, \boldsymbol{\omega})\|^2 \leq \sigma^2$, $\forall\, \boldsymbol{\theta}, \boldsymbol{\omega}, \in \mathbb{R}^p, i \in [m], j \in [n]$.*

With the metric in (11), the convergence properties of GT-SRVR/GT-SRVRI are characterized as follows:

Theorem 3 (Convergence of GT-SRVR)**.** *Under Assumption 1 (a)&(c)-(e) and Assumption 2, if the step-sizes satisfy $\kappa \triangleq \gamma/\eta \leq \mu^2/13L_f^2$ and $\eta \leq \min\{k_5, k_6, k_7, k_8\}$, then we have the following convergence result for GT-SRVR:*

$$\frac{1}{T+1}\sum_{t=0}^{T}\mathbb{E}[\mathfrak{M}_t] \leq \frac{2\mathbb{E}[\mathfrak{p}_0 - \mathfrak{p}_{T+1}]}{\min\{1, L_f^2\}(T+1)\gamma},$$

where $\mathfrak{p}_t$ is the potential function defined as:

$$\mathfrak{p}_t \triangleq J(\bar{\boldsymbol{\theta}}_t) + \frac{8\gamma L_f^2}{\mu\eta}\|\bar{\boldsymbol{\omega}}_t - \boldsymbol{\omega}_t^*\|^2 + \frac{1}{m}\sum_{i=1}^{m}\big(\|\boldsymbol{\theta}_{i,t} - \bar{\boldsymbol{\theta}}_t\|^2 + \|\boldsymbol{\omega}_{i,t} - \bar{\boldsymbol{\omega}}_t\|^2 + \gamma\|\mathbf{p}_{i,t} - \bar{\mathbf{p}}_t\|^2 + \eta\|\mathbf{d}_{i,t} - \bar{\mathbf{d}}_t\|^2\big),$$

and the constants in the step-size η are:

$$C_0 = \frac{1}{1-\lambda}(1+\frac{1}{\kappa})+\frac{1}{2}+\frac{18L_f^2}{\mu^2},$$
$$k_5 = \frac{13L_f^2}{\mu^2\big(1/2+1/(1-\lambda)^2\big)},\ \ k_6 = \frac{1}{8\mu C_0},$$
$$k_7 = \frac{26(1-\lambda)L_f^2}{(\mu^2+144L_f^4+4L_f^2\mu^2+64C_0L_f^2\mu^2)},$$
$$k_8 = \frac{13L_f^2}{2\mu^2}\Big(L_f+\frac{L_f^2}{\mu}+(1-\lambda)\Big).$$

Theorem 4 (Convergence of GT-SRVRI)**.** *Under Assumption 1 (a)&(c)-(e), Assumption 2, and the same parameter settings and potential function as stated in Theorem 3, we have the following convergence result for GT-SRVRI:*

$$\frac{1}{T+1}\sum_{t=0}^{T}\mathbb{E}[\mathfrak{M}_t] \le \frac{2\mathbb{E}[\mathfrak{p}_0-\mathfrak{p}_{T+1}]}{\min\{1,L_f^2\}(T+1)\gamma}+\frac{1}{\min\{1,L_f^2\}}\times$$
$$\big(\frac{12}{\lambda}(1+\frac{1}{\kappa})+4+\frac{144L_f^2}{\mu^2}\big)\big(\frac{\sigma^2\epsilon^2}{c_\epsilon}+\frac{\sigma^2C(n,q,\alpha)}{T+1}\big), \tag{14}$$

where the constant $C(n,q,\alpha)$ is defined as

$$C(n,q,\alpha) \triangleq \begin{cases} \frac{1}{1-\alpha}(\frac{n}{q})^{(\frac{1}{\alpha}-1)}-\frac{\alpha}{1-\alpha}, & \text{if } \alpha>0 \text{ and } \alpha\neq 1 \\ \log(\frac{n}{q})+1, & \text{if } \alpha=1. \end{cases}$$

Remark 3. In Theorems 3 and 4, it can be seen that the step-sizes and convergence rate depend on the network topology and the ratio of the step-sizes κ. Additionally, the convergence performance of GT-SRVRI is affected by the constant $C(n,q,\alpha)$, which depends on the inexact gradient estimation in tth iteration with $\text{mod}(t,q)=0$.

Following from Theorems 3 and 4, we immediately have the complexity results of GT-SRVR/GT-SRVRI:

Corollary 5. *Under the conditions in Theorems 3 and 4, and with $q=\sqrt{n}$, to achieve an ϵ^2-stationary solution (i.e., $\frac{1}{T+1}\sum_{t=0}^{T}\mathbb{E}[\mathfrak{M}_t]\le\epsilon^2$) with $\sqrt{n}\epsilon^2\le 1$, we have:*

- *for GT-SRVR, the total communication rounds are $\mathcal{O}(\epsilon^{-2})$ and the total samples evaluated across the network are $\mathcal{O}(m\sqrt{n}\epsilon^{-2})$);*
- *GT-SRVRI with $\alpha\ge 1$, the total communication rounds are bounded by $\mathcal{O}(\log(\sqrt{n})\epsilon^{-2})$ and the total samples evaluated across the network are bounded by $\mathcal{O}(m\log(\sqrt{n})\sqrt{n}\epsilon^{-2})$).*

Remark 4. From Corollary 5, we can see that GT-SRVR has the same communication complexity as GT-GDA, but the sample complexity is lower than GT-GDA. For GT-SRVRI with $\alpha\ge 1$, the upper bounds of both the complexities have an additional factor $\log(\sqrt{n})$ factor compared with GT-SRVR. Although the theoretical complexity bounds for GT-SRVRI is weaker than GT-SRVR (due to abandoning full gradients completely), we show in Section 5 that GT-SRVRI outperforms than GT-SRVR in practice.

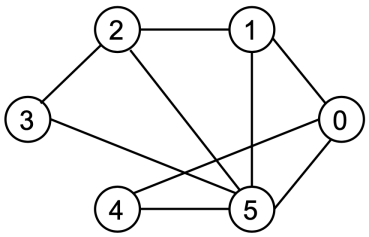


Figure 1: Network topology

5 Experimental Results

In this section, we demonstrate the performance of our proposed GT-GDA and GT-SRVR/GT-SRVRI algorithms for MARL policy evaluation. Due to space limitation, the results of detailed hyper-parameter settings and on different network topologies are relegated to the supplementary material. Since existing works on solving MARL policy evaluation with nonlinear function approximation are quite limited, in our experiments, we compare our algorithms with two stochastic algorithms:

1) *Decentralized Stochastic Gradient Descent Ascent (DSGDA):* This algorithm is motivated by DSGD [35, 36]. Each agent updates its local parameters as $\boldsymbol{\theta}_{i,t+1}=\sum_{j\in\mathcal{N}_i}[\mathbf{M}]_{ij}\boldsymbol{\theta}_{j,t}-\gamma\frac{1}{|\mathcal{S}_{i,t}|}\sum_{j\in\mathcal{S}_{i,t}}\nabla_{\boldsymbol{\theta}}f_{ij}(\boldsymbol{\theta}_{i,t},\boldsymbol{\omega}_{i,t})$ and $\boldsymbol{\omega}_{i,t+1}=\sum_{j\in\mathcal{N}_i}[\mathbf{M}]_{ij}\boldsymbol{\omega}_{j,t}-\eta\frac{1}{|\mathcal{S}_{i,t}|}\sum_{j\in\mathcal{S}_{i,t}}\nabla_{\boldsymbol{\omega}}f_{ij}(\boldsymbol{\theta}_{i,t},\boldsymbol{\omega}_{i,t})$.

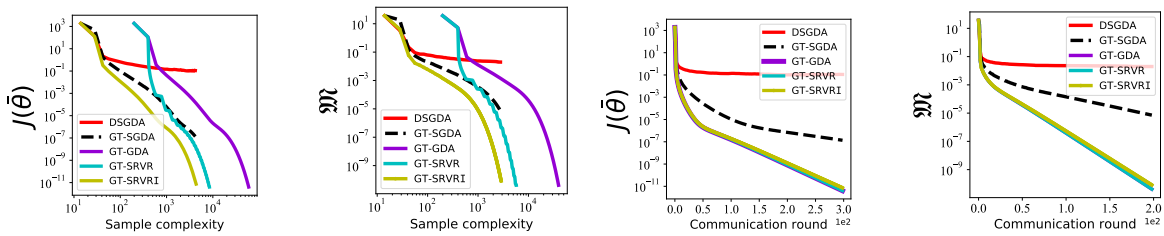
2) *Gradient-Tracking-Based Stochastic Gradient Descent Ascent (GT-SGDA)*: This algorithm is motivated by the GT-SGD algorithm [37, 38]. GT-SGDA has the same structure as that of GT-GDA, but it updates $\mathbf{v}_{i,t}$ and $\mathbf{u}_{i,t}$ using stochastic gradients as follows: $\mathbf{v}_{i,t}=\frac{1}{|\mathcal{S}_{i,t}|}\sum_{j\in\mathcal{S}_{i,t}}\nabla_{\boldsymbol{\theta}}f_{ij}(\boldsymbol{\theta}_{i,t},\boldsymbol{\omega}_{i,t})$ and $\mathbf{u}_{i,t}=\frac{1}{|\mathcal{S}_{i,t}|}\sum_{j\in\mathcal{S}_{i,t}}\nabla_{\boldsymbol{\omega}}f_{ij}(\boldsymbol{\theta}_{i,t},\boldsymbol{\omega}_{i,t})$.

1) RL Environment: We adopt the environment of Cooperative Navigation task in [39], which consists of m agents inhabiting a two-dimensional world with continuous space and discrete time. In the system, agents cooperate with each other to reach their own landmarks.

2) Multi-Agent Networks: We use a six-node multi-agent system, with the communication graph $\mathcal{G}$ being generated by the Erdös-Rènyi graph, where edge connectivity probability is $p_c = 0.35$. The network consensus matrix is chosen as $\mathbf{W} = \mathbf{I} - \frac{2}{3\lambda_{\max}(\mathbf{L})}\mathbf{L}$, where $\mathbf{L}$ is the Laplacian matrix of $\mathcal{G}$, and $\lambda_{\max}(\mathbf{L})$ denotes the largest eigenvalue of $\mathbf{L}$. The generated topology is shown in Figure 1.

3) Data Generation and Model: We first obtain a good policy and then generate a trajectory of the state-action pairs. In all experiments, the trajectory lengh is $n = 200$. We set the discount factor $\gamma = 0.95$. Then, $\mathcal{V}_{\boldsymbol{\theta}}(\cdot)$ is parametrized by a 2-hidden-layer neural network with 20 hidden units, where the Sigmoid activation is used at each unit.

4) Numerical Results: In Figure 2, we compare our GT-GDA and GT-SRVR/GT-SRVRI with two baseline algorithms GT-SGDA and DSGDA in terms of MSPBE $J(\boldsymbol{\theta}) = \max_{\boldsymbol{\omega}\in\mathbb{R}^p} F(\boldsymbol{\theta},\boldsymbol{\omega})$ and the convergence metric in (11). We initialize all algorithms at the same point generated randomly from the normal distribution. We set the constant learning rates to $\gamma = 10^{-1}, \eta = 10^{-1}$ and mini-batch size $q = \lceil\sqrt{n}\rceil$.



(a) Sample complexity vs MSPBE.

(b) Sample complexity vs convergence metric.

(c) Communication complexity vs MSPBE.

(d) Communication complexity vs convergence metric.

Figure 2: Algorithms comparison

From Figure 2(a) and 2(b), it can be seen that GT-SRVRI converges much faster than other algorithms (GT-GDA, GT-SRVR, GT-SGD and DSGD) in terms of the total number of gradient evaluations. We can also observe that both GT-SRVR and GT-SRVRI has better sample efficiency in attaining high accuracy (smaller than 10^{-9}) than the other three algorithms thanks to the variance-reduced techniques. As is shown in Figure 2(c) and 2(d), GT-SRVR and GT-SRVRI have the same communication cost as GT-GDA, which is much lower than those of DSGDA and GT-SGDA. Our experimental results are consistent with our theoretical analysis that GT-SRVR/GT-SRVRI enjoy low sample and communication complexities.

6 Conclusion

In this paper, we studied the multi-agent reinforcement learning (MARL) policy evaluation problem with smooth nonlinear function approximation. By using the Fenchel's duality, we reformulated the MARL MSPBE minimization problem as a decentralized non-convex-strongly-concave minimax problem. To address the minimax problem in the decentralized setting, we first developed and analyzed a gradient tracking based algorithm named GT-GDA. We showed that GT-GDA algorithm has the communication complexity of $\mathcal{O}(\epsilon^{-2})$ and sample complexity of $\mathcal{O}(mn\epsilon^{-2})$. To further reduce the sample complexity while maintaining the communication complexity, we proposed two stochastic variance-reduced methods called GT-SRVR and GT-SRVRI. We showed these the two algorithms can achieve the same communication complexity as GT-GDA, but improve the sample complexity to $\mathcal{O}(m\sqrt{n}\epsilon^{-2})$. We have also conducted extensive numerical studies to verify the performance of our proposed algorithms. Our work opens up several interesting direction for future research. First, in our algorithms, the agents need to send primal-dual parameter pairs to their neighbors. It is interesting to adopt communication-efficient mechanisms to further reduce the communication cost, especially when the parameters are high-dimensional. Second, our model assume that the joint states and actions are available to all agents. In our future work, we will further study MARL problems with partially observable information.

References

[1] R. S. Sutton and A. G. Barto, *Reinforcement learning: An introduction*. MIT press, 2018.

[2] D. P. Bertsekas and J. N. Tsitsiklis, "Neuro-dynamic programming: an overview," in *Proceedings of 1995 34th IEEE Conference on Decision and Control*, vol. 1. IEEE, 1995, pp. 560–564.

[3] M. G. Lagoudakis and R. Parr, "Least-squares policy iteration," *Journal of Machine Learning Research*, vol. 4, no. Dec, pp. 1107–1149, 2003.

[4] J. Cortes, S. Martinez, T. Karatas, and F. Bullo, "Coverage control for mobile sensing networks," *IEEE Transactions on Robotics and Automation*, vol. 20, no. 2, pp. 243–255, 2004.

[5] P. Ogren, E. Fiorelli, and N. E. Leonard, "Cooperative control of mobile sensor networks: Adaptive gradient climbing in a distributed environment," *IEEE Transactions on Automatic Control*, vol. 49, no. 8, pp. 1292–1302, 2004.

[6] M. Rabbat and R. Nowak, "Distributed optimization in sensor networks," in *Proceedings of the 3rd Iternational Symposium on Information Processing in Sensor Networks*, 2004, pp. 20–27.

[7] W. D. Smart and L. P. Kaelbling, "Effective reinforcement learning for mobile robots," in *Proceedings 2002 IEEE International Conference on Robotics and Automation (Cat. No. 02CH37292)*, vol. 4. IEEE, 2002, pp. 3404–3410.

[8] J. Kober, J. A. Bagnell, and J. Peters, "Reinforcement learning in robotics: A survey," *The International Journal of Robotics Research*, vol. 32, no. 11, pp. 1238–1274, 2013.

[9] A. S. Polydoros and L. Nalpantidis, "Survey of model-based reinforcement learning: Applications on robotics," *Journal of Intelligent & Robotic Systems*, vol. 86, no. 2, pp. 153–173, 2017.

[10] D. S. Callaway and I. A. Hiskens, "Achieving controllability of electric loads," *Proceedings of the IEEE*, vol. 99, no. 1, pp. 184–199, 2010.

[11] E. Dall'Anese, H. Zhu, and G. B. Giannakis, "Distributed optimal power flow for smart microgrids," *IEEE Transactions on Smart Grid*, vol. 4, no. 3, pp. 1464–1475, 2013.

[12] D. Ernst, M. Glavic, and L. Wehenkel, "Power systems stability control: reinforcement learning framework," *IEEE Transactions on Power Systems*, vol. 19, no. 1, pp. 427–435, 2004.

[13] M. Glavic, R. Fonteneau, and D. Ernst, "Reinforcement learning for electric power system decision and control: Past considerations and perspectives," *IFAC-PapersOnLine*, vol. 50, no. 1, pp. 6918–6927, 2017.

[14] K. Zhang, Z. Yang, and T. Basar, "Multi-agent reinforcement learning: A selective overview of theories and algorithms," *arXiv preprint arXiv:1911.10635*, 2019.

[15] J. a. o. t. N. Tsitsiklis and B. Van Roy, "An analysis of temporal-difference learning with function approximation," *IEEE Transactions on Automatic Control*, vol. 42, no. 5, pp. 674–690, 1997.

[16] H. R. Maei, C. Szepesvari, S. Bhatnagar, D. Precup, D. Silver, and R. S. Sutton, "Convergent temporal-difference learning with arbitrary smooth function approximation." in *Advances in Neural Information Processing Systems*, 2009, pp. 1204–1212.

[17] H.-T. Wai, M. Hong, Z. Yang, Z. Wang, and K. Tang, "Variance reduced policy evaluation with smooth function approximation," *Advances in Neural Information Processing Systems*, vol. 32, pp. 5784–5795, 2019.

[18] H. Sun, S. Lu, and M. Hong, "Improving the sample and communication complexity for decentralized non-convex optimization: Joint gradient estimation and tracking," in *International Conference on Machine Learning*. PMLR, 2020, pp. 9217–9228.

[19] M. L. Littman, "Markov games as a framework for multi-agent reinforcement learning," in *Machine Learning Proceedings 1994*. Elsevier, 1994, pp. 157–163.

[20] M. Lauer and M. Riedmiller, "An algorithm for distributed reinforcement learning in cooperative multi-agent systems," in *In Proceedings of the Seventeenth International Conference on Machine Learning*. Citeseer, 2000.

[21] X. Wang and T. Sandholm, "Reinforcement learning to play an optimal nash equilibrium in team markov games," *Advances in Neural Information Processing Systems*, vol. 15, pp. 1603–1610, 2002.

[22] J. Hu and M. P. Wellman, "Nash q-learning for general-sum stochastic games," *Journal of Machine Learning Research*, vol. 4, no. Nov, pp. 1039–1069, 2003.

[23] G. Arslan and S. Yüksel, "Decentralized q-learning for stochastic teams and games," *IEEE Transactions on Automatic Control*, vol. 62, no. 4, pp. 1545–1558, 2016.

[24] S. S. Du, J. Chen, L. Li, L. Xiao, and D. Zhou, "Stochastic variance reduction methods for policy evaluation," *arXiv preprint arXiv:1702.07944*, 2017.

[25] S. Qiu, Z. Yang, X. Wei, J. Ye, and Z. Wang, "Single-timescale stochastic nonconvex-concave optimization for smooth nonlinear TD learning," *arXiv preprint arXiv:2008.10103*, 2020.

[26] D. Ding, X. Wei, Z. Yang, Z. Wang, and M. R. Jovanovic, "Fast multi-agent temporal-difference learning via homotopy stochastic primal-dual method," in *Optimization Foundations for Reinforcement Learning Workshop, 33rd Conference on Neural Information Processing Systems*, 2019.

[27] H.-T. Wai, Z. Yang, Z. Wang, and M. Hong, "Multi-agent reinforcement learning via double averaging primal-dual optimization," in *Advances in Neural Information Processing Systems*, 2018, pp. 9649–9660.

[28] D. Lee, H. Yoon, and N. Hovakimyan, "Primal-dual algorithm for distributed reinforcement learning: distributed GTD," in *2018 IEEE Conference on Decision and Control (CDC)*. IEEE, 2018, pp. 1967–1972.

[29] T. Doan, S. Maguluri, and J. Romberg, "Finite-time analysis of distributed TD(0) with linear function approximation on multi-agent reinforcement learning," in *International Conference on Machine Learning*, 2019, pp. 1626–1635.

[30] S. Bhatnagar, D. Precup, D. Silver, R. S. Sutton, H. Maei, and C. Szepesvári, "Convergent temporal-difference learning with arbitrary smooth function approximation," *Advances in Neural Information Processing Systems*, vol. 22, pp. 1204–1212, 2009.

[31] A. Cutkosky and F. Orabona, "Momentum-based variance reduction in non-convex sgd," in *Advances in Neural Information Processing Systems*, 2019, pp. 15 236–15 245.

[32] T. Lin, C. Jin, and M. Jordan, "On gradient descent ascent for nonconvex-concave minimax problems," in *International Conference on Machine Learning*. PMLR, 2020, pp. 6083–6093.

[33] Z. Wang, K. Ji, Y. Zhou, Y. Liang, and V. Tarokh, "Spiderboost and momentum: Faster variance reduction algorithms," in *Advances in Neural Information Processing Systems*, 2019, pp. 2406–2416.

[34] K. Ji, Z. Wang, B. Weng, Y. Zhou, W. Zhang, and Y. Liang, "History-gradient aided batch size adaptation for variance reduced algorithms," in *International Conference on Machine Learning*. PMLR, 2020, pp. 4762–4772.

[35] A. Nedic and A. Ozdaglar, "Distributed subgradient methods for multi-agent optimization," *IEEE Transactions on Automatic Control*, vol. 54, no. 1, p. 48, 2009.

[36] Z. Jiang, A. Balu, C. Hegde, and S. Sarkar, "Collaborative deep learning in fixed topology networks," in *Advances in Neural Information Processing Systems*, 2017, pp. 5904–5914.

[37] R. Xin, U. A. Khan, and S. Kar, "An improved convergence analysis for decentralized online stochastic non-convex optimization," *arXiv preprint arXiv:2008.04195*, 2020.

[38] S. Lu, X. Zhang, H. Sun, and M. Hong, "GNSD: A gradient-tracking based nonconvex stochastic algorithm for decentralized optimization," in *2019 IEEE Data Science Workshop (DSW)*. IEEE, 2019, pp. 315–321.

[39] R. Lowe, Y. Wu, A. Tamar, J. Harb, P. Abbeel, and I. Mordatch, "Multi-agent actor-critic for mixed cooperative-competitive environments," *arXiv preprint arXiv:1706.02275*, 2017.

[40] P. Bernhard and A. Rapaport, "On a theorem of danskin with an application to a theorem of von neumann-sion," *Nonlinear Analysis: Theory, Methods & Applications*, vol. 24, no. 8, pp. 1163–1181, 1995.

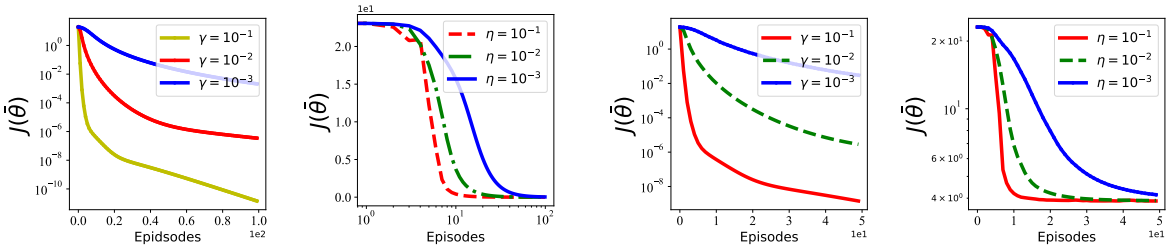
Supplementary Material

A Further Experiments and Additional Results

In the following experiements, we adopt the environment of Cooperative Navigation task in [39], which consists of m nodes inhabiting a two-dimensional world with continuous space and discrete time.

A.1 Learning rate setting

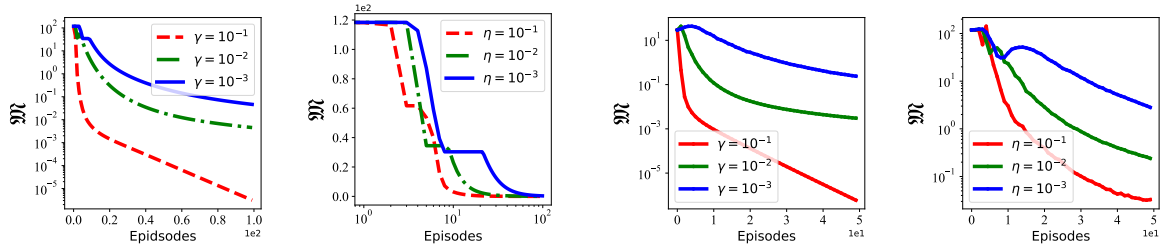
We use a 6-nodes multi-agent system with a generated topology as shown in Figure 1. In this experiment, we choose the trajectory length $n = 200$, discount factor $\gamma = 0.95$ and mini-batch size $q = \lceil \sqrt{n} \rceil$. $\mathcal{V}_{\boldsymbol{\theta}}(\cdot)$ is parametrized by a 2-hidden-layer neural network with 20 hidden units, where the Sigmoid activation is used at each unit. Fig.3-4 illustrate the objective function $J(\boldsymbol{\theta})$ and convergence metric $\mathfrak{M}$ performance of GT-GDA and GT-SRVR with different learning rates γ and η. Since excessive learning rates will result in larger fluctuation in loss value function. Thus, we fixed a relatively small learning rate $\gamma = 10^{-3}$ while comparing η, and set $\eta = 10^{-3}$ while comparing γ. In this experiment, we observe that methods with a smaller learning rate have a smaller slope in the figure, which leads to a slower convergence.



(a) Step-size comparison on GT-GDA.

(b) Step-size comparison on GT-SRVR.

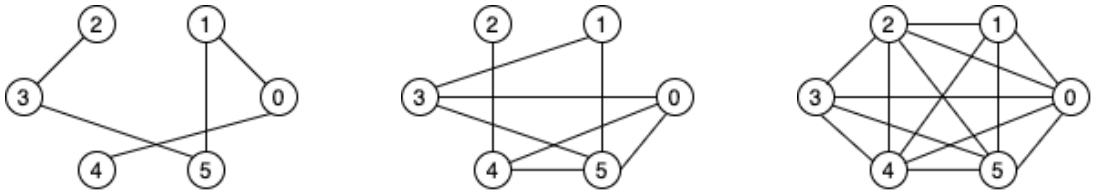
Figure 3: Performance of objective function with different step-size.



(a) Step-size comparison on GT-GDA.

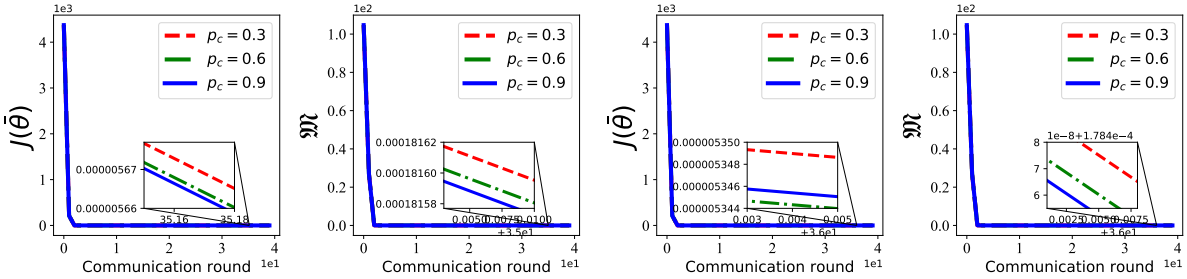
(b) Step-size comparison on GT-SRVR.

Figure 4: Performance of convergence metric $\mathfrak{M}$ with different step-size.



(a) Topology sparsity $p_c = 0.3$. (b) Topology sparsity $p_c = 0.6$. (c) Topology sparsity $p_c = 0.9$.

Figure 5: Topology.



(a) Topology sparsity comparison on GT-GDA. (b) Topology sparsity comparison on GT-SRVR.

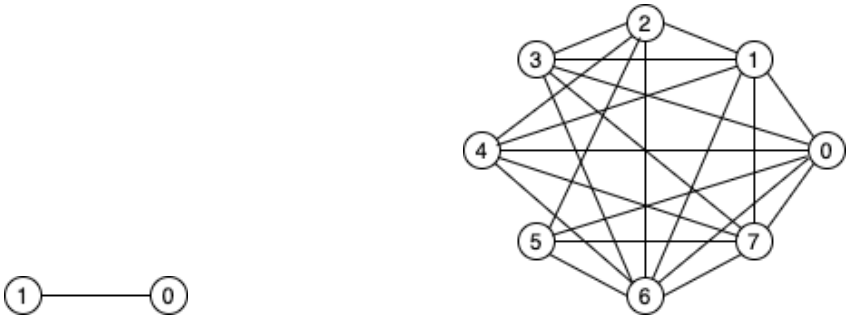
Figure 6: Performance with different topology.

A.2 Topology setting

We use a 6-nodes multi-agent system and experiments on three different topologies. The generated topology with different sparsity are shown in Fig. 5. The trajectory length is $n = 200$ and we set the discount factor $\gamma = 0.95$, constant learning rate $\gamma = 0.1$, $\eta = 0.1$ and mini-batch size $q = \lceil \sqrt{n} \rceil$. We observe that the objective function $J(\theta)$ and convergence metric $\mathfrak{M}$ is insensitive to network topology. The subplot in Fig.6 show that the $J(\theta)$ and $\mathfrak{M}$ slightly increases as p_c decreases.

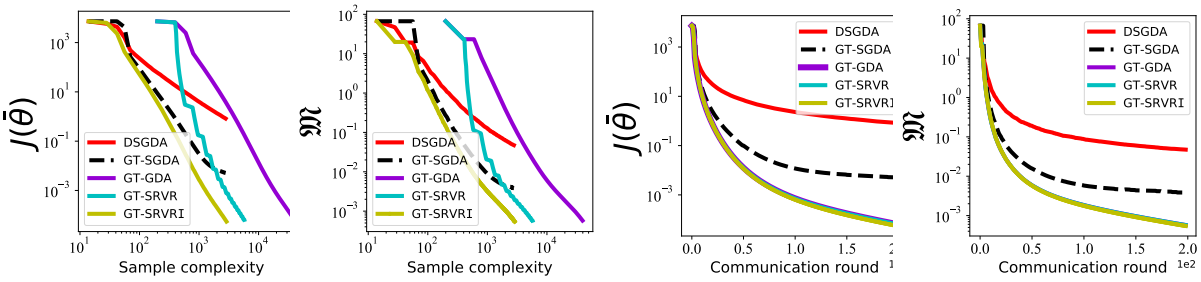
A.3 Node setting

We test the following experiments on different multi-agent systems. The generated topology with a 2-nodes system and a 8-nodes system are shown in Fig. 7. The trajectory length is $n = 200$ and we set the discount factor $\gamma = 0.95$, constant learning rate $\gamma = 0.1$, $\eta = 0.1$ and mini-batch size $q = \lceil \sqrt{n} \rceil$. We compare our proposed algorithm GT-GDA and GT-SRVR/GT-SRVRI with two baseline algorithms GT-SGDA and DSGDA mentioned in Section 5 in terms of MSPBE $J(\boldsymbol{\theta}) = \max_{\boldsymbol{\omega} \in \mathbb{R}^p} F(\boldsymbol{\theta}, \boldsymbol{\omega})$ and the convergence metric in (11). We observe similar results as shown in Section 5. Thus, we can conclude that our proposed algorithms GT-SRVR/GT-SRVRI enjoy low sample and communication complexities in general.



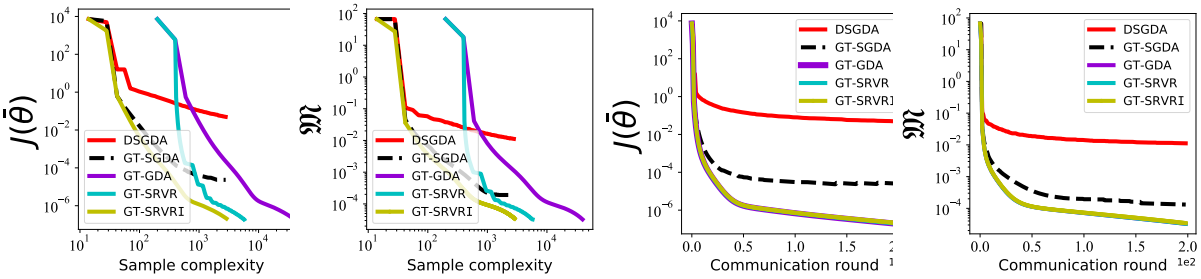
(a) Topology sparsity $p_c = 0.8$ with 2 nodes. (b) Topology sparsity $p_c = 0.8$ with 8 nodes.

Figure 7: Topology.



(a) Sample complexity v.s. MSPBE & convergence metric. (b) Communication complexity v.s. MSPBE & convergence metric.

Figure 8: Performance with 2 Nodes.



(a) Sample complexity v.s. MSPBE & convergence metric. (b) Communication complexity v.s. MSPBE & convergence metric.

Figure 9: Performance with 8 Nodes.

B Proof Outline for Theorem 1

In this section, we organize the proof of Theorem 1 into several key lemmas. Before diving in our theoretical analysis, we first define the following notations:

- $\bar{\mathbf{x}}_t = \frac{1}{m}\sum_{i=1}^m \mathbf{x}_{i,t}$ and $\mathbf{x}_t = [\mathbf{x}_{1,t}^\top, \cdots, \mathbf{x}_{m,t}^\top]^\top$ for any vector $\mathbf{x}$;
- $\nabla_{\boldsymbol{\theta}} F_t = [\nabla_{\boldsymbol{\theta}} F(\boldsymbol{\theta}_{1,t}, \boldsymbol{\omega}_{1,t})^\top, \cdots, \nabla_{\boldsymbol{\theta}} F(\boldsymbol{\theta}_{m,t}, \boldsymbol{\omega}_{m,t})^\top]^\top$;
- $\nabla_{\boldsymbol{\omega}} F_t = [\nabla_{\boldsymbol{\omega}} F(\boldsymbol{\theta}_{1,t}, \boldsymbol{\omega}_{1,t})^\top, \cdots, \nabla_{\boldsymbol{\omega}} F(\boldsymbol{\theta}_{m,t}, \boldsymbol{\omega}_{m,t})^\top]^\top$;
- $\mathcal{E}(\mathbf{x}_t) = \frac{1}{m}\sum_{i=1}^m \|\mathbf{x}_{i,t} - \bar{\mathbf{x}}_t\|^2$ for any vector $\mathbf{x}$.

Our first step is to show the following descent property of GT-GDA algorithm on the function $J(\cdot)$:

Lemma 1 (Descent Inequality on $J(\boldsymbol{\theta})$)**.** *Under Assumption 1, the following descent inequality holds under GT-GDA:*

$$J(\bar{\boldsymbol{\theta}}_{t+1}) - J(\bar{\boldsymbol{\theta}}_t) \le -\frac{\gamma}{2}\|\nabla J(\bar{\boldsymbol{\theta}}_t)\|^2 - (\frac{\gamma}{2} - \frac{L_J\gamma^2}{2})\|\bar{\mathbf{p}}_t\|^2 + \gamma L_F^2\|\boldsymbol{\omega}_t^* - \bar{\boldsymbol{\omega}}_t\|^2 + \gamma\|\nabla_{\boldsymbol{\theta}} F(\bar{\boldsymbol{\theta}}_t, \bar{\boldsymbol{\omega}}_t) - \bar{\mathbf{p}}_t\|^2, \tag{15}$$

where $J(\boldsymbol{\theta}_t) = \max_{\boldsymbol{\omega}} F(\boldsymbol{\theta}_t, \boldsymbol{\omega})$ *and* $\boldsymbol{\omega}_t^* = \arg\max_{\boldsymbol{\omega}} F(\boldsymbol{\theta}_t, \boldsymbol{\omega})$.

Proof. According to the algorithm update, we have:

$$\begin{aligned}
&J(\bar{\boldsymbol{\theta}}_{t+1}) - J(\bar{\boldsymbol{\theta}}_t) \overset{(a)}{\le} \langle \nabla J(\bar{\boldsymbol{\theta}}_t), \bar{\boldsymbol{\theta}}_{t+1} - \bar{\boldsymbol{\theta}}_t\rangle + \frac{L_J}{2}\|\bar{\boldsymbol{\theta}}_{t+1} - \bar{\boldsymbol{\theta}}_t\|^2 \\
&\overset{(b)}{=} -\gamma\langle \nabla J(\bar{\boldsymbol{\theta}}_t), \bar{\mathbf{p}}_t\rangle + \frac{L_J\gamma^2}{2}\|\bar{\mathbf{p}}_t\|^2 \\
&= -\frac{\gamma}{2}\|\nabla J(\bar{\boldsymbol{\theta}}_t)\|^2 - (\frac{\gamma}{2} - \frac{L_J\gamma^2}{2})\|\bar{\mathbf{p}}_t\|^2 + \frac{\gamma}{2}\|\nabla J(\bar{\boldsymbol{\theta}}_t) - \bar{\mathbf{p}}_t\|^2 \\
&= -\frac{\gamma}{2}\|\nabla J(\bar{\boldsymbol{\theta}}_t)\|^2 - (\frac{\gamma}{2} - \frac{L_J\gamma^2}{2})\|\bar{\mathbf{p}}_t\|^2 + \frac{\gamma}{2}\|\nabla J(\bar{\boldsymbol{\theta}}_t) - \nabla_{\boldsymbol{\theta}} F(\bar{\boldsymbol{\theta}}_t, \bar{\boldsymbol{\omega}}_t) + \nabla_{\boldsymbol{\theta}} F(\bar{\boldsymbol{\theta}}_t, \bar{\boldsymbol{\omega}}_t) - \bar{\mathbf{p}}_t\|^2 \\
&\overset{(c)}{\le} -\frac{\gamma}{2}\|\nabla J(\bar{\boldsymbol{\theta}}_t)\|^2 - (\frac{\gamma}{2} - \frac{L_J\gamma^2}{2})\|\bar{\mathbf{p}}_t\|^2 + \gamma\|\nabla J(\bar{\boldsymbol{\theta}}_t) - \nabla_{\boldsymbol{\theta}} F(\bar{\boldsymbol{\theta}}_t, \bar{\boldsymbol{\omega}}_t)\|^2 + \gamma\|\nabla_{\boldsymbol{\theta}} F(\bar{\boldsymbol{\theta}}_t, \bar{\boldsymbol{\omega}}_t) - \bar{\mathbf{p}}_t\|^2 \\
&\overset{(d)}{\le} -\frac{\gamma}{2}\|\nabla J(\bar{\boldsymbol{\theta}}_t)\|^2 - (\frac{\gamma}{2} - \frac{L_J\gamma^2}{2})\|\bar{\mathbf{p}}_t\|^2 + \gamma L_F^2\|\boldsymbol{\omega}_t^* - \bar{\boldsymbol{\omega}}_t\|^2 + \gamma\|\nabla_{\boldsymbol{\theta}} F(\bar{\boldsymbol{\theta}}_t, \bar{\boldsymbol{\omega}}_t) - \bar{\mathbf{p}}_t\|^2,
\end{aligned} \tag{16}$$

where (a) is because of Lipschitz continuous gradients of J, (b) follows from the update rule (9), (c) follows from $\|\mathbf{x}+\mathbf{y}\|^2 \le 2\|\mathbf{x}\|^2 + 2\|\mathbf{y}\|^2\ \forall \mathbf{x}, \mathbf{y}$, and (d) follows from Lemma 10 and Assumption 1 (b). □

Note that in the RHS of (15), there is an error term $\|\boldsymbol{\omega}_t^* - \bar{\boldsymbol{\omega}}_t\|^2$. The following lemma states the contraction property of this term.

Lemma 2 (Error Bound on $\boldsymbol{\omega}^*(\boldsymbol{\theta})$)**.** *Under Assumption 1 and with $\eta \leq 1/2L_F$, the following inequality holds for GT-GDA:*

$$\|\bar{\boldsymbol{\omega}}_{t+1} - \boldsymbol{\omega}^*_{t+1}\|^2 \leq (1 - \frac{\mu\eta}{4})\|\boldsymbol{\omega}^*_t - \bar{\boldsymbol{\omega}}_t\|^2 + \frac{9\eta}{2\mu}\|\nabla_{\boldsymbol{\omega}} F(\bar{\boldsymbol{\theta}}_t, \bar{\boldsymbol{\omega}}_t) - \bar{\mathbf{d}}_t\| - (1 + \frac{\mu\eta}{4})\frac{\eta^2}{4}\|\bar{\mathbf{d}}_t\|^2 + \frac{5L_{\boldsymbol{\omega}}^2\gamma^2}{\mu\eta}\|\bar{\mathbf{p}}_t\|^2, \tag{17}$$

where $\boldsymbol{\omega}^*_t = \boldsymbol{\omega}^*(\bar{\boldsymbol{\theta}}_t) = \arg\max_{\boldsymbol{\omega}} F(\bar{\boldsymbol{\theta}}_t, \boldsymbol{\omega})$.

Proof. Define $\boldsymbol{\omega}^*_t = \boldsymbol{\omega}^*(\bar{\boldsymbol{\theta}}_t) = \arg\max_{\boldsymbol{\omega}} F(\bar{\boldsymbol{\theta}}_t, \boldsymbol{\omega})$. We have:

$$\|\bar{\boldsymbol{\omega}}_{t+1} - \boldsymbol{\omega}^*_t\|^2 = \|\bar{\boldsymbol{\omega}}_t + \eta\bar{\mathbf{d}}_t - \boldsymbol{\omega}^*_t\|^2 = \|\bar{\boldsymbol{\omega}}_t - \boldsymbol{\omega}^*_t\|^2 + \eta^2\|\bar{\mathbf{d}}_t\|^2 + 2\eta\langle\bar{\boldsymbol{\omega}}_t - \boldsymbol{\omega}^*_t, \bar{\mathbf{d}}_t\rangle. \tag{18}$$

To remove the third term in (18), we have:

$$\begin{aligned} &F(\bar{\boldsymbol{\theta}}_t, \boldsymbol{\omega}) - F(\bar{\boldsymbol{\theta}}_t, \bar{\boldsymbol{\omega}}_t) + \frac{\mu}{2}\|\boldsymbol{\omega} - \bar{\boldsymbol{\omega}}_t\|^2 \leq \langle\nabla_{\boldsymbol{\omega}} F(\bar{\boldsymbol{\theta}}_t, \bar{\boldsymbol{\omega}}_t), \boldsymbol{\omega} - \bar{\boldsymbol{\omega}}_t\rangle \\ =&\langle\bar{\mathbf{d}}_t, \boldsymbol{\omega} - \bar{\boldsymbol{\omega}}_{t+1}\rangle + \langle\nabla_{\boldsymbol{\omega}} F(\bar{\boldsymbol{\theta}}_t, \bar{\boldsymbol{\omega}}_t) - \bar{\mathbf{d}}_t, \boldsymbol{\omega} - \bar{\boldsymbol{\omega}}_{t+1}\rangle + \langle\nabla_{\boldsymbol{\omega}} F(\bar{\boldsymbol{\theta}}_t, \bar{\boldsymbol{\omega}}_t), \bar{\boldsymbol{\omega}}_{t+1} - \bar{\boldsymbol{\omega}}_t\rangle. \end{aligned} \tag{19}$$

With $\eta \leq 1/2L_F$ and by Assumption 1 (b), it follows that

$$\begin{aligned} &-\frac{1}{4\eta}\|\bar{\boldsymbol{\omega}}_{t+1} - \bar{\boldsymbol{\omega}}_t\|^2 \leq -\frac{L_F}{2}\|\bar{\boldsymbol{\omega}}_{t+1} - \bar{\boldsymbol{\omega}}_t\|^2 \\ &\leq F(\bar{\boldsymbol{\theta}}_t, \boldsymbol{\omega}_{t+1}) - F(\bar{\boldsymbol{\theta}}_t, \bar{\boldsymbol{\omega}}_t) - \langle\nabla_{\boldsymbol{\omega}} F(\bar{\boldsymbol{\theta}}_t, \bar{\boldsymbol{\omega}}_t), \bar{\boldsymbol{\omega}}_{t+1} - \bar{\boldsymbol{\omega}}_t\rangle. \end{aligned} \tag{20}$$

Combining (19) and (20), with the update $\bar{\boldsymbol{\omega}}_{t+1} - \bar{\boldsymbol{\omega}}_t = \eta\bar{\mathbf{d}}_t$, we have:

$$\begin{aligned} &F(\bar{\boldsymbol{\theta}}_t, \boldsymbol{\omega}) - F(\bar{\boldsymbol{\theta}}_t, \bar{\boldsymbol{\omega}}_{t+1}) + \frac{\mu}{2}\|\boldsymbol{\omega} - \bar{\boldsymbol{\omega}}_t\|^2 \\ \leq&\langle\bar{\mathbf{d}}_t, \boldsymbol{\omega} - \bar{\boldsymbol{\omega}}_{t+1}\rangle + \langle\nabla_{\boldsymbol{\omega}} F(\bar{\boldsymbol{\theta}}_t, \bar{\boldsymbol{\omega}}_t) - \bar{\mathbf{d}}_t, \boldsymbol{\omega} - \bar{\boldsymbol{\omega}}_{t+1}\rangle + \frac{1}{4\eta}\|\bar{\boldsymbol{\omega}}_{t+1} - \bar{\boldsymbol{\omega}}_t\|^2 \\ =&\langle\bar{\mathbf{d}}_t, \boldsymbol{\omega} - \bar{\boldsymbol{\omega}}_t\rangle + \langle\bar{\mathbf{d}}_t, \bar{\boldsymbol{\omega}}_t - \bar{\boldsymbol{\omega}}_{t+1}\rangle + \langle\nabla_{\boldsymbol{\omega}} F(\bar{\boldsymbol{\theta}}_t, \bar{\boldsymbol{\omega}}_t) - \bar{\mathbf{d}}_t, \boldsymbol{\omega} - \bar{\boldsymbol{\omega}}_{t+1}\rangle + \frac{\eta}{4}\|\bar{\mathbf{d}}_t\|^2 \\ =&\langle\bar{\mathbf{d}}_t, \boldsymbol{\omega} - \bar{\boldsymbol{\omega}}_t\rangle - \eta\|\bar{\mathbf{d}}_t\|^2 + \langle\nabla_{\boldsymbol{\omega}} F(\bar{\boldsymbol{\theta}}_t, \bar{\boldsymbol{\omega}}_t) - \bar{\mathbf{d}}_t, \boldsymbol{\omega} - \bar{\boldsymbol{\omega}}_{t+1}\rangle + \frac{\eta}{4}\|\bar{\mathbf{d}}_t\|^2 \\ =&\langle\bar{\mathbf{d}}_t, \boldsymbol{\omega} - \bar{\boldsymbol{\omega}}_t\rangle + \langle\nabla_{\boldsymbol{\omega}} F(\bar{\boldsymbol{\theta}}_t, \bar{\boldsymbol{\omega}}_t) - \bar{\mathbf{d}}_t, \boldsymbol{\omega} - \bar{\boldsymbol{\omega}}_{t+1}\rangle - \frac{3\eta}{4}\|\bar{\mathbf{d}}_t\|^2. \end{aligned} \tag{21}$$

Let $\boldsymbol{\omega} = \boldsymbol{\omega}^*_t$, we have

$$\begin{aligned} &F(\bar{\boldsymbol{\theta}}_t, \boldsymbol{\omega}^*_t) - F(\bar{\boldsymbol{\theta}}_t, \bar{\boldsymbol{\omega}}_{t+1}) + \frac{\mu}{2}\|\boldsymbol{\omega}^*_t - \bar{\boldsymbol{\omega}}_t\|^2 \\ \leq&\langle\bar{\mathbf{d}}_t, \boldsymbol{\omega}^*_t - \bar{\boldsymbol{\omega}}_t\rangle + \langle\nabla_{\boldsymbol{\omega}} F(\bar{\boldsymbol{\theta}}_t, \bar{\boldsymbol{\omega}}_t) - \bar{\mathbf{d}}_t, \boldsymbol{\omega}^*_t - \bar{\boldsymbol{\omega}}_{t+1}\rangle - \frac{3\eta}{4}\|\bar{\mathbf{d}}_t\|^2 \\ \overset{(a)}{\leq}&\langle\bar{\mathbf{d}}_t, \boldsymbol{\omega}^*_t - \bar{\boldsymbol{\omega}}_t\rangle + \frac{2}{\mu}\|\nabla_{\boldsymbol{\omega}} F(\bar{\boldsymbol{\theta}}_t, \bar{\boldsymbol{\omega}}_t) - \bar{\mathbf{d}}_t\| + \frac{\mu}{8}\|\boldsymbol{\omega}^*_t - \bar{\boldsymbol{\omega}}_{t+1}\| - \frac{3\eta}{4}\|\bar{\mathbf{d}}_t\|^2 \\ \overset{(b)}{\leq}&\langle\bar{\mathbf{d}}_t, \boldsymbol{\omega}^*_t - \bar{\boldsymbol{\omega}}_t\rangle + \frac{2}{\mu}\|\nabla_{\boldsymbol{\omega}} F(\bar{\boldsymbol{\theta}}_t, \bar{\boldsymbol{\omega}}_t) - \bar{\mathbf{d}}_t\| + \frac{\mu}{4}\|\boldsymbol{\omega}^*_t - \bar{\boldsymbol{\omega}}_t\|^2 + \frac{\mu}{4}\|\bar{\boldsymbol{\omega}}_t - \bar{\boldsymbol{\omega}}_{t+1}\|^2 - \frac{3\eta}{4}\|\bar{\mathbf{d}}_t\|^2 \\ \overset{(c)}{=}&\langle\bar{\mathbf{d}}_t, \boldsymbol{\omega}^*_t - \bar{\boldsymbol{\omega}}_t\rangle + \frac{2}{\mu}\|\nabla_{\boldsymbol{\omega}} F(\bar{\boldsymbol{\theta}}_t, \bar{\boldsymbol{\omega}}_t) - \bar{\mathbf{d}}_t\| + \frac{\mu}{4}\|\boldsymbol{\omega}^*_t - \bar{\boldsymbol{\omega}}_t\|^2 - (\frac{3\eta}{4} - \frac{\mu\eta^2}{4})\|\bar{\mathbf{d}}_t\|^2, \end{aligned} \tag{22}$$

where (a) follows from $-\langle \mathbf{x}, \mathbf{y}\rangle \leq \frac{1}{2c}\|\mathbf{x}\|^2 + \frac{c}{2}\|\mathbf{y}\|^2$ and $c = \frac{\mu}{4}$, (b) is due to $\|\mathbf{x}+\mathbf{y}\|^2 \leq 2\|\mathbf{x}\|^2 + 2\|\mathbf{y}\|^2$, and (c) is from $\bar{\boldsymbol{\omega}}_{t+1} - \bar{\boldsymbol{\omega}}_t = \eta\bar{\mathbf{d}}_t$.

Since $F(\bar{\boldsymbol{\theta}}_t, \boldsymbol{\omega}_t^*) \geq F(\bar{\boldsymbol{\theta}}_t, \bar{\boldsymbol{\omega}}_{t+1})$, we have

$$\frac{\mu\eta}{2}\|\boldsymbol{\omega}_t^* - \bar{\boldsymbol{\omega}}_t\|^2 \leq 2\eta\langle \bar{\mathbf{d}}_t, \boldsymbol{\omega}_t^* - \bar{\boldsymbol{\omega}}_t\rangle + \frac{4\eta}{\mu}\|\nabla_{\boldsymbol{\omega}} F(\bar{\boldsymbol{\theta}}_t, \bar{\boldsymbol{\omega}}_t) - \bar{\mathbf{d}}_t\| - (\frac{3\eta^2}{2} - \frac{\mu\eta^3}{2})\|\bar{\mathbf{d}}_t\|^2. \tag{23}$$

Combining (18) and (23) and setting $\eta \leq 1/2\mu \leq 1/2L_F$, we have

$$\|\bar{\boldsymbol{\omega}}_{t+1} - \boldsymbol{\omega}_t^*\|^2 \leq (1 - \frac{\mu\eta}{2})\|\boldsymbol{\omega}_t^* - \bar{\boldsymbol{\omega}}_t\|^2 + \frac{4\eta}{\mu}\|\nabla_{\boldsymbol{\omega}} F(\bar{\boldsymbol{\theta}}_t, \bar{\boldsymbol{\omega}}_t) - \bar{\mathbf{d}}_t\| - \frac{\eta^2}{4}\|\bar{\mathbf{d}}_t\|^2. \tag{24}$$

Then, it holds that

$$\begin{aligned}
&\|\bar{\boldsymbol{\omega}}_{t+1} - \boldsymbol{\omega}_{t+1}^*\|^2 = \|\bar{\boldsymbol{\omega}}_{t+1} - \boldsymbol{\omega}_t^* + \boldsymbol{\omega}_t^* - \boldsymbol{\omega}_{t+1}^*\|^2 \\
\overset{(a)}{\leq}&(1 + \frac{\mu\eta}{4})\|\bar{\boldsymbol{\omega}}_{t+1} - \boldsymbol{\omega}_t^*\|^2 + (1 + \frac{4}{\mu\eta})\|\boldsymbol{\omega}_t^* - \boldsymbol{\omega}_{t+1}^*\|^2 \\
\overset{(b)}{\leq}&(1 + \frac{\mu\eta}{4})\|\bar{\boldsymbol{\omega}}_{t+1} - \boldsymbol{\omega}_t^*\|^2 + (1 + \frac{4}{\mu\eta})L_{\boldsymbol{\omega}}^2\|\bar{\boldsymbol{\theta}}_t - \bar{\boldsymbol{\theta}}_{t+1}\|^2 \\
\overset{(c)}{\leq}&(1 + \frac{\mu\eta}{4})(1 - \frac{\mu\eta}{2})\|\boldsymbol{\omega}_t^* - \bar{\boldsymbol{\omega}}_t\|^2 + (1 + \frac{\mu\eta}{4})\frac{4\eta}{\mu}\|\nabla_{\boldsymbol{\omega}} F(\bar{\boldsymbol{\theta}}_t, \bar{\boldsymbol{\omega}}_t) - \bar{\mathbf{d}}_t\| \\
&- (1 + \frac{\mu\eta}{4})\frac{\eta^2}{4}\|\bar{\mathbf{d}}_t\|^2 + (1 + \frac{4}{\mu\eta})L_{\boldsymbol{\omega}}^2\|\bar{\boldsymbol{\theta}}_t - \bar{\boldsymbol{\theta}}_{t+1}\|^2 \\
\overset{(d)}{\leq}&(1 - \frac{\mu\eta}{4})\|\boldsymbol{\omega}_t^* - \bar{\boldsymbol{\omega}}_t\|^2 + \frac{9\eta}{2\mu}\|\nabla_{\boldsymbol{\omega}} F(\bar{\boldsymbol{\theta}}_t, \bar{\boldsymbol{\omega}}_t) - \bar{\mathbf{d}}_t\| - (1 + \frac{\mu\eta}{4})\frac{\eta^2}{4}\|\bar{\mathbf{d}}_t\|^2 + \frac{5L_{\boldsymbol{\omega}}^2\gamma^2}{\mu\eta}\|\bar{\mathbf{p}}_t\|^2,
\end{aligned} \tag{25}$$

where (a) follows from $\|\mathbf{x}+\mathbf{y}\|^2 \leq (1 + 1/c)\|\mathbf{x}\|^2 + (1+c)\|\mathbf{y}\|^2$ and $c = \mu\eta/4$, (b) follows from Lemma 9, (c) follows fromy plugging (24), and (d) due to the facts that:

$$\begin{aligned}
&(1 + \frac{\mu\eta}{4})(1 - \frac{\mu\eta}{2}) = 1 + \frac{\mu\eta}{4} - \frac{\mu\eta}{2} - \frac{\mu^2\eta^2}{8} \leq 1 - \frac{\mu\eta}{4}, \\
&(1 + \frac{\mu\eta}{4})\frac{4\eta}{\mu} \leq (1 + \frac{\mu}{4}\cdot\frac{1}{2\mu})\frac{4\eta}{\mu} = \frac{9\eta}{2\mu}, \\
&1 + \frac{4}{\mu\eta} \leq \frac{1}{\mu\eta} + \frac{4}{\mu\eta} = \frac{5}{\mu\eta} \ \text{, and } \ \bar{\boldsymbol{\theta}}_t - \bar{\boldsymbol{\theta}}_{t+1} = \gamma\bar{\mathbf{p}}_t.
\end{aligned} \tag{26}$$

Plugging (26) into (25) yields:

$$\begin{aligned}
&\|\bar{\boldsymbol{\omega}}_{t+1} - \boldsymbol{\omega}_{t+1}^*\|^2 - \|\boldsymbol{\omega}_t^* - \bar{\boldsymbol{\omega}}_t\|^2 \\
\leq& -\frac{\mu\eta}{4}\|\boldsymbol{\omega}_t^* - \bar{\boldsymbol{\omega}}_t\|^2 + \frac{9\eta}{2\mu}\|\nabla_{\boldsymbol{\omega}} F(\bar{\boldsymbol{\theta}}_t, \bar{\boldsymbol{\omega}}_t) - \bar{\mathbf{d}}_t\| - (1 + \frac{\mu\eta}{4})\frac{\eta^2}{4}\|\bar{\mathbf{d}}_t\|^2 + \frac{5L_{\boldsymbol{\omega}}^2\gamma^2}{\mu\eta}\|\bar{\mathbf{p}}_t\|^2.
\end{aligned} \tag{27}$$

□

Next, by combining the results from Lemma 1-2, we have the following descent result:

Lemma 3. *Under Assumption 1 and with $\eta \leq 1/2L_F$, the following inequality holds for GT-GDA:*

$$
\begin{aligned}
&J(\bar{\boldsymbol{\theta}}_{T+1}) - J(\bar{\boldsymbol{\theta}}_0) + \frac{8\gamma L_F^2}{\mu\eta}\left[\|\bar{\boldsymbol{\omega}}_{T+1} - \boldsymbol{\omega}^*_{T+1}\|^2 - \|\boldsymbol{\omega}^*_0 - \bar{\boldsymbol{\omega}}_0\|^2\right] \\
\leq& -\frac{\gamma}{2}\sum_{t=0}^{T}\|\nabla J(\bar{\boldsymbol{\theta}}_t)\|^2 - \gamma L_F^2\sum_{t=0}^{T}\|\boldsymbol{\omega}^*_t - \bar{\boldsymbol{\omega}}_t\|^2 - (\frac{\gamma}{2} - \frac{L_J\gamma^2}{2} - \frac{40\gamma^3 L_{\boldsymbol{\omega}}^2 L_F^2}{\mu^2\eta^2})\sum_{t=0}^{T}\|\bar{\mathbf{p}}_t\|^2 - \frac{2\gamma\eta L_F^2}{\mu}\sum_{t=0}^{T}\|\bar{\mathbf{d}}_t\|^2 \\
&+ (\frac{72\gamma L_F^4}{\mu^2 m} + \frac{2\gamma L_F^2}{m})\sum_{t=0}^{T}[\mathcal{E}(\boldsymbol{\theta}_t) + \mathcal{E}(\boldsymbol{\omega}_t)] + \frac{2\gamma}{m}\sum_{t=0}^{T}\|\nabla_{\boldsymbol{\omega}} F_t - \mathbf{v}_t\|^2 + \frac{72L_F^2\gamma}{m\mu^2}\sum_{t=0}^{T}\|\nabla_{\boldsymbol{\theta}} F_t - \mathbf{u}_t\|^2.
\end{aligned}
$$

Proof. From Lemma 1-2, we have

$$
\begin{aligned}
&J(\bar{\boldsymbol{\theta}}_{t+1}) - J(\bar{\boldsymbol{\theta}}_t) + \frac{8\gamma L_F^2}{\mu\eta}\left[\|\bar{\boldsymbol{\omega}}_{t+1} - \boldsymbol{\omega}^*_{t+1}\|^2 - \|\boldsymbol{\omega}^*_t - \bar{\boldsymbol{\omega}}_t\|^2\right] \\
\leq& -\frac{\gamma}{2}\|\nabla J(\bar{\boldsymbol{\theta}}_t)\|^2 - \gamma L_F^2\|\boldsymbol{\omega}^*_t - \bar{\boldsymbol{\omega}}_t\|^2 - (\frac{\gamma}{2} - \frac{L_J\gamma^2}{2} - \frac{40\gamma^3 L_{\boldsymbol{\omega}}^2 L_F^2}{\mu^2\eta^2})\|\bar{\mathbf{p}}_t\|^2 - \frac{2\gamma\eta L_F^2}{\mu}\|\bar{\mathbf{d}}_t\|^2 \\
&+ \gamma\|\nabla_{\boldsymbol{\theta}} F(\bar{\boldsymbol{\theta}}_t, \bar{\boldsymbol{\omega}}_t) - \bar{\mathbf{p}}_t\|^2 + \frac{36L_F^2\gamma}{\mu^2}\|\nabla_{\boldsymbol{\omega}} F(\bar{\boldsymbol{\theta}}_t, \bar{\boldsymbol{\omega}}_t) - \bar{\mathbf{d}}_t\|^2.
\end{aligned} \tag{28}
$$

Note that

$$
\begin{aligned}
&\|\nabla_{\boldsymbol{\theta}} F(\bar{\boldsymbol{\theta}}_t, \bar{\boldsymbol{\omega}}_t) - \bar{\mathbf{p}}_t\|^2 \\
=&\|\nabla_{\boldsymbol{\theta}} F(\bar{\boldsymbol{\theta}}_t, \bar{\boldsymbol{\omega}}_t) - \frac{1}{m}\sum_{i=1}^{m}\nabla_{\boldsymbol{\theta}} F_i(\boldsymbol{\theta}_{i,t}, \boldsymbol{\omega}_{i,t}) + \frac{1}{m}\sum_{i=1}^{m}\nabla_{\boldsymbol{\theta}} F_i(\boldsymbol{\theta}_{i,t}, \boldsymbol{\omega}_{i,t}) - \bar{\mathbf{p}}_t\|^2 \\
\leq&2\|\nabla_{\boldsymbol{\theta}} F(\bar{\boldsymbol{\theta}}_t, \bar{\boldsymbol{\omega}}_t) - \frac{1}{m}\sum_{i=1}^{m}\nabla_{\boldsymbol{\theta}} F_i(\boldsymbol{\theta}_{i,t}, \boldsymbol{\omega}_{i,t})\|^2 + 2\|\frac{1}{m}\sum_{i=1}^{m}\nabla_{\boldsymbol{\theta}} F_i(\boldsymbol{\theta}_{i,t}, \boldsymbol{\omega}_{i,t}) - \bar{\mathbf{p}}_t\|^2 \\
\leq&\frac{2}{m}\sum_{i=1}^{m}\|\nabla_{\boldsymbol{\theta}} F(\bar{\boldsymbol{\theta}}_t, \bar{\boldsymbol{\omega}}_t) - \nabla_{\boldsymbol{\theta}} F_i(\boldsymbol{\theta}_{i,t}, \boldsymbol{\omega}_{i,t})\|^2 + 2\|\frac{1}{m}\sum_{i=1}^{m}\nabla_{\boldsymbol{\theta}} F_i(\boldsymbol{\theta}_{i,t}, \boldsymbol{\omega}_{i,t}) - \bar{\mathbf{p}}_t\|^2 \\
\leq&\frac{2L_f^2}{m}\sum_{i=1}^{m}[\|\bar{\boldsymbol{\theta}}_t - \boldsymbol{\theta}_{i,t}\|^2 + \|\bar{\boldsymbol{\omega}}_t - \boldsymbol{\omega}_{i,t}\|^2] + 2\|\frac{1}{m}\sum_{i=1}^{m}\nabla_{\boldsymbol{\theta}} F_i(\boldsymbol{\theta}_{i,t}, \boldsymbol{\omega}_{i,t}) - \bar{\mathbf{p}}_t\|^2.
\end{aligned} \tag{29}
$$

Similarly, we have:

$$
\begin{aligned}
\|\nabla_{\boldsymbol{\omega}} F(\bar{\boldsymbol{\theta}}_t, \bar{\boldsymbol{\omega}}_t) - \bar{\mathbf{d}}_t\|^2 \leq \frac{2L_F^2}{m}\sum_{i=1}^{m}[\|\bar{\boldsymbol{\theta}}_t - \boldsymbol{\theta}_{i,t}\|^2 + \|\bar{\boldsymbol{\omega}}_t - \boldsymbol{\omega}_{i,t}\|^2] \\
+ 2\|\frac{1}{m}\sum_{i=1}^{m}\nabla_{\boldsymbol{\omega}} F_i(\boldsymbol{\theta}_{i,t}, \boldsymbol{\omega}_{i,t}) - \bar{\mathbf{d}}_t\|^2.
\end{aligned} \tag{30}
$$

Thus, we have

$$
\begin{aligned}
&J(\bar{\boldsymbol{\theta}}_{t+1}) - J(\bar{\boldsymbol{\theta}}_t) + \frac{8\gamma L_F^2}{\mu\eta}\big[\|\bar{\boldsymbol{\omega}}_{t+1} - \boldsymbol{\omega}^*_{t+1}\|^2 - \|\boldsymbol{\omega}^*_t - \bar{\boldsymbol{\omega}}_t\|^2\big] \\
\leq& -\frac{\gamma}{2}\|\nabla J(\bar{\boldsymbol{\theta}}_t)\|^2 - \gamma L_F^2\|\boldsymbol{\omega}^*_t - \bar{\boldsymbol{\omega}}_t\|^2 - (\frac{\gamma}{2} - \frac{L_J\gamma^2}{2} - \frac{40\gamma^3 L_{\boldsymbol{\omega}}^2 L_F^2}{\mu^2\eta^2})\|\bar{\mathbf{p}}_t\|^2 - \frac{2\gamma\eta L_F^2}{\mu}\|\bar{\mathbf{d}}_t\|^2 \\
&+ \gamma\|\nabla_{\boldsymbol{\theta}} F(\bar{\boldsymbol{\theta}}_t, \bar{\boldsymbol{\omega}}_t) - \bar{\mathbf{p}}_t\|^2 + \frac{36L_F^2\gamma}{\mu^2}\|\nabla_{\boldsymbol{\omega}} F(\bar{\boldsymbol{\theta}}_t, \bar{\boldsymbol{\omega}}_t) - \bar{\mathbf{d}}_t\|^2 \\
\leq& -\frac{\gamma}{2}\|\nabla J(\bar{\boldsymbol{\theta}}_t)\|^2 - \gamma L_F^2\|\boldsymbol{\omega}^*_t - \bar{\boldsymbol{\omega}}_t\|^2 - (\frac{\gamma}{2} - \frac{L_J\gamma^2}{2} - \frac{40\gamma^3 L_{\boldsymbol{\omega}}^2 L_F^2}{\mu^2\eta^2})\|\bar{\mathbf{p}}_t\|^2 - \frac{2\gamma\eta L_F^2}{\mu}\|\bar{\mathbf{d}}_t\|^2 \\
&+ (\frac{72\gamma L_F^4}{\mu^2 m} + \frac{2\gamma L_F^2}{m})\sum_{i=1}^m[\|\bar{\boldsymbol{\theta}}_t - \boldsymbol{\theta}_{i,t}\|^2 + \|\bar{\boldsymbol{\omega}}_t - \boldsymbol{\omega}_{i,t}\|^2] \\
&+ 2\gamma\|\frac{1}{m}\sum_{i=1}^m \nabla_{\boldsymbol{\omega}} F_i(\boldsymbol{\theta}_{i,t}, \boldsymbol{\omega}_{i,t}) - \bar{\mathbf{v}}_t\|^2 + \frac{72L_F^2\gamma}{\mu^2}\|\frac{1}{m}\sum_{i=1}^m \nabla_{\boldsymbol{\theta}} F_i(\boldsymbol{\theta}_{i,t}, \boldsymbol{\omega}_{i,t}) - \bar{\mathbf{u}}_t\|^2 \\
\overset{(a)}{\leq}& -\frac{\gamma}{2}\|\nabla J(\bar{\boldsymbol{\theta}}_t)\|^2 - \gamma L_F^2\|\boldsymbol{\omega}^*_t - \bar{\boldsymbol{\omega}}_t\|^2 - (\frac{\gamma}{2} - \frac{L_J\gamma^2}{2} - \frac{40\gamma^3 L_{\boldsymbol{\omega}}^2 L_F^2}{\mu^2\eta^2})\|\bar{\mathbf{p}}_t\|^2 - \frac{2\gamma\eta L_F^2}{\mu}\|\bar{\mathbf{d}}_t\|^2 \\
&+ (\frac{72\gamma L_F^4}{\mu^2 m} + \frac{2\gamma L_F^2}{m})[\mathcal{E}(\boldsymbol{\theta}_t) + \mathcal{E}(\boldsymbol{\omega}_t)] + \frac{2\gamma}{m}\|\nabla_{\boldsymbol{\omega}} F_t - \mathbf{v}_t\|^2 + \frac{72L_F^2\gamma}{m\mu^2}\|\nabla_{\boldsymbol{\theta}} F_t - \mathbf{u}_t\|^2, \qquad (31)
\end{aligned}
$$

where (a) due to $\|\frac{1}{m}\sum_{i=1}^m \mathbf{x}_{i,t} - \bar{x}_t\|^2 \leq \frac{1}{m}\sum_{i=1}^m \|\mathbf{x}_{i,t} - \bar{x}_t\|^2$.

Telescoping the above inequality, we have the stated result. □

Next, we prove the contraction of iterations in the following lemma, which is useful in analyzing the decentralized gradient tracking algorithms.

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

$$\|\boldsymbol{\theta}_t - \mathbf{1}\otimes\bar{\boldsymbol{\theta}}_t\|^2 \leq (1+c_1)\lambda^2\|\boldsymbol{\theta}_{t-1} - \mathbf{1}\otimes\bar{\boldsymbol{\theta}}_{t-1}\|^2 + (1+\frac{1}{c_1})\gamma^2\|\mathbf{p}_{t-1} - \mathbf{1}\otimes\bar{\mathbf{p}}_{t-1}\|^2, \qquad (32)$$

$$\|\boldsymbol{\omega}_t - \mathbf{1}\otimes\bar{\boldsymbol{\omega}}_t\|^2 \leq (1+c_2)\lambda^2\|\boldsymbol{\omega}_{t-1} - \mathbf{1}\otimes\bar{\boldsymbol{\omega}}_{t-1}\|^2 + (1+\frac{1}{c_2})\eta^2\|\mathbf{d}_{t-1} - \mathbf{1}\otimes\bar{\mathbf{d}}_{t-1}\|^2, \qquad (33)$$

$$\|\mathbf{p}_t - \mathbf{1}\otimes\bar{\mathbf{p}}_t\|^2 \leq (1+c_1)\lambda^2\|\mathbf{p}_{t-1} - \mathbf{1}\otimes\bar{\mathbf{p}}_{t-1}\|^2 + (1+\frac{1}{c_1})\|\mathbf{v}_t - \mathbf{v}_{t-1}\|^2, \qquad (34)$$

$$\|\mathbf{d}_t - \mathbf{1}\otimes\bar{\mathbf{d}}_t\|^2 \leq (1+c_2)\lambda^2\|\mathbf{d}_{t-1} - \mathbf{1}\otimes\bar{\mathbf{d}}_{t-1}\|^2 + (1+\frac{1}{c_2})\|\mathbf{u}_t - \mathbf{u}_{t-1}\|^2, \qquad (35)$$

where c_1 and c_2 are arbitrary positive constants. Additionally, we have

$$\|\boldsymbol{\theta}_t - \boldsymbol{\theta}_{t-1}\|^2 \leq 8\|(\boldsymbol{\theta}_{t-1} - \mathbf{1}\otimes\bar{\boldsymbol{\theta}}_{t-1})\|^2 + 4\gamma^2\|\mathbf{p}_{t-1} - \mathbf{1}\otimes\bar{\mathbf{p}}_{t-1}\|^2 + 4\gamma^2 m\|\bar{\mathbf{p}}_{t-1}\|^2, \qquad (36)$$

$$\|\boldsymbol{\omega}_t - \boldsymbol{\omega}_{t-1}\|^2 \leq 8\|(\boldsymbol{\omega}_{t-1} - \mathbf{1}\otimes\bar{\boldsymbol{\omega}}_{t-1})\|^2 + 4\gamma^2\|\mathbf{d}_{t-1} - \mathbf{1}\otimes\bar{\mathbf{d}}_{t-1}\|^2 + 4\gamma^2 m\|\bar{\mathbf{d}}_{t-1}\|^2. \qquad (37)$$

Proof. First for the iterates $\boldsymbol{\theta}_t$, we have the following contraction:

$$\|\widetilde{\mathbf{M}}\boldsymbol{\theta}_t - \mathbf{1}\otimes\bar{\boldsymbol{\theta}}_t\|^2 = \|\widetilde{\mathbf{M}}(\boldsymbol{\theta}_t - \mathbf{1}\otimes\bar{\boldsymbol{\theta}}_t)\|^2 \leq \lambda^2\|\boldsymbol{\theta}_t - \mathbf{1}\otimes\bar{\boldsymbol{\theta}}_t\|^2. \qquad (38)$$

This is because $\boldsymbol{\theta}_t - \mathbf{1}\otimes\boldsymbol{\theta}_t$ is orthogonal to $\mathbf{1}$, which is the eigenvector corresponding to the largest eigenvalue of $\widetilde{\mathbf{M}}$, and $\lambda = \max\{|\lambda_2|, |\lambda_m|\}$. Recall that $\bar{\boldsymbol{\theta}}_t = \bar{\boldsymbol{\theta}}_{t-1} - \gamma\bar{\mathbf{p}}_{t-1}$, hence,

$$\begin{aligned}
&\|\boldsymbol{\theta}_t - \mathbf{1}\otimes\bar{\boldsymbol{\theta}}_t\|^2 = \|\widetilde{\mathbf{M}}\boldsymbol{\theta}_{t-1} - \gamma\mathbf{p}_{t-1} - \mathbf{1}\otimes(\bar{\boldsymbol{\theta}}_{t-1} - \gamma\bar{\mathbf{p}}_{t-1})\|^2 \\
&\leq (1+c_1)\|\widetilde{\mathbf{M}}\boldsymbol{\theta}_{t-1} - \mathbf{1}\otimes\bar{\boldsymbol{\theta}}_{t-1}\|^2 + (1+\frac{1}{c_1})\gamma^2\|\mathbf{p}_{t-1} - \mathbf{1}\otimes\bar{\mathbf{p}}_{t-1}\|^2 \\
&\leq (1+c_1)\lambda^2\|\boldsymbol{\theta}_{t-1} - \mathbf{1}\otimes\bar{\boldsymbol{\theta}}_{t-1}\|^2 + (1+\frac{1}{c_1})\gamma^2\|\mathbf{p}_{t-1} - \mathbf{1}\otimes\bar{\mathbf{p}}_{t-1}\|^2. \qquad (39)
\end{aligned}$$

Similarly, we have the result for $\boldsymbol{\omega}_t$,

$$\|\boldsymbol{\omega}_t - \mathbf{1}\otimes\bar{\boldsymbol{\omega}}_t\|^2 \leq (1+c_2)\lambda^2\|\boldsymbol{\omega}_{t-1} - \mathbf{1}\otimes\bar{\boldsymbol{\omega}}_{t-1}\|^2 + (1+\frac{1}{c_2})\eta^2\|\mathbf{d}_{t-1} - \mathbf{1}\otimes\bar{\mathbf{d}}_{t-1}\|^2. \qquad (40)$$

For $\mathbf{p}_t$, we have

$$\begin{aligned}
&\|\mathbf{p}_t - \mathbf{1}\otimes\bar{\mathbf{p}}_t\|^2 = \|\widetilde{\mathbf{M}}\mathbf{p}_{t-1} + \mathbf{v}_t - \mathbf{v}_{t-1} - \mathbf{1}\otimes\big(\bar{\mathbf{p}}_{t-1} + \bar{\mathbf{v}}_t - \bar{\mathbf{v}}_{t-1}\big)\| \\
&\leq (1+c_1)\lambda^2\|\mathbf{p}_{t-1} - \mathbf{1}\otimes\bar{\mathbf{p}}_{t-1}\|^2 + (1+\frac{1}{c_1})\|\mathbf{v}_t - \mathbf{v}_{t-1} - \mathbf{1}\otimes\big(\bar{\mathbf{v}}_t - \bar{\mathbf{v}}_{t-1}\big)\|^2 \\
&\leq (1+c_1)\lambda^2\|\mathbf{p}_{t-1} - \mathbf{1}\otimes\bar{\mathbf{p}}_{t-1}\|^2 + (1+\frac{1}{c_1})\|\big(\mathbf{I} - \frac{1}{n}(\mathbf{1}\mathbf{1}^\top)\otimes\mathbf{I}\big)\big(\mathbf{v}_t - \mathbf{v}_{t-1}\big)\|^2 \\
&\overset{(a)}{\leq} (1+c_1)\lambda^2\|\mathbf{p}_{t-1} - \mathbf{1}\otimes\bar{\mathbf{p}}_{t-1}\|^2 + (1+\frac{1}{c_1})\|\mathbf{v}_t - \mathbf{v}_{t-1}\|^2, \qquad (41)
\end{aligned}$$

where (a) is due to $\|\mathbf{I} - \frac{1}{m}(\mathbf{1}\mathbf{1}^\top)\otimes\mathbf{I}\| \leq 1$. Similarly, we have

$$\|\mathbf{d}_t - \mathbf{1}\otimes\bar{\mathbf{d}}_t\|^2 \leq (1+c_2)\lambda^2\|\mathbf{d}_{t-1} - \mathbf{1}\otimes\bar{\mathbf{d}}_{t-1}\|^2 + (1+\frac{1}{c_2})\|\mathbf{u}_t - \mathbf{u}_{t-1}\|^2. \qquad (42)$$

According to the update, we have

$$\begin{aligned}
&\|\boldsymbol{\theta}_t - \boldsymbol{\theta}_{t-1}\|^2 = \|\widetilde{\mathbf{M}}\boldsymbol{\theta}_{t-1} - \gamma\mathbf{p}_{t-1} - \boldsymbol{\theta}_{t-1}\|^2 \\
=&\|(\widetilde{\mathbf{M}} - \mathbf{I})\boldsymbol{\theta}_{t-1} - \gamma\mathbf{p}_{t-1}\|^2 \leq 2\|(\widetilde{\mathbf{M}} - \mathbf{I})\boldsymbol{\theta}_{t-1}\|^2 + 2\gamma^2\|\mathbf{p}_{t-1}\|^2 \\
=&2\|(\widetilde{\mathbf{M}} - \mathbf{I})(\boldsymbol{\theta}_{t-1} - \mathbf{1}\otimes\bar{\boldsymbol{\theta}}_{t-1})\|^2 + 2\gamma^2\|\mathbf{p}_{t-1}\|^2 \\
\leq&8\|(\boldsymbol{\theta}_{t-1} - \mathbf{1}\otimes\bar{\boldsymbol{\theta}}_{t-1})\|^2 + 4\gamma^2\|\mathbf{p}_{t-1} - \mathbf{1}\otimes\bar{\mathbf{p}}_{t-1}\|^2 + 4\gamma^2 m\|\bar{\mathbf{p}}_{t-1}\|^2 \\
=&8\mathcal{E}(\boldsymbol{\theta}_{t-1}) + 4\gamma^2\mathcal{E}(\mathbf{p}_{t-1}) + 4\gamma^2 m\|\bar{\mathbf{p}}_{t-1}\|^2, \qquad (43)
\end{aligned}$$

and also

$$\|\boldsymbol{\omega}_t - \boldsymbol{\omega}_{t-1}\|^2 \leq 8\mathcal{E}(\boldsymbol{\omega}_{t-1}) + 4\eta^2\mathcal{E}(\mathbf{d}_{t-1}) + 4\eta^2 m\|\bar{\mathbf{d}}_{t-1}\|^2. \qquad (44)$$

□

Lemma 5 (Differential Bound on Estimator for GT-GDA)**.** *Under Assumption 1, the following inequalities holds*

$$\sum_{t=1}^{T}\mathbb{E}\|\mathbf{v}_t - \mathbf{v}_{t-1}\|^2 \leq \sum_{t=1}^{T} 3L_F^2\mathbb{E}\|\boldsymbol{\theta}_{t-1} - \boldsymbol{\theta}_t\|^2 + 3L_F^2\mathbb{E}\|\boldsymbol{\omega}_{t-1} - \boldsymbol{\omega}_t\|^2, \qquad (45)$$

$$\sum_{t=1}^{T}\mathbb{E}\|\mathbf{u}_t - \mathbf{u}_{t-1}\|^2 \leq \sum_{t=1}^{T} 3L_F^2\mathbb{E}\|\boldsymbol{\theta}_{t-1} - \boldsymbol{\theta}_t\|^2 + 3L_F^2\mathbb{E}\|\boldsymbol{\omega}_{t-1} - \boldsymbol{\omega}_t\|^2. \qquad (46)$$

Proof. For $\|\mathbf{v}_t - \mathbf{v}_{t-1}\|^2$, we have

$$\begin{aligned}
&\mathbb{E}\|\mathbf{v}_t - \mathbf{v}_{t-1}\|^2 = \mathbb{E}\|\mathbf{v}_t - \nabla_{\boldsymbol{\theta}}\mathbf{F}_t + \nabla_{\boldsymbol{\theta}}\mathbf{F}_t - \nabla_{\boldsymbol{\theta}}\mathbf{F}_{t-1} + \nabla_{\boldsymbol{\theta}}\mathbf{F}_{t-1} - \mathbf{v}_{t-1}\|^2 \\
\leq& 3\mathbb{E}\|\mathbf{v}_t - \nabla_{\boldsymbol{\theta}}F_t\|^2 + 3\mathbb{E}\|\nabla_{\boldsymbol{\theta}}F_t - \nabla_{\boldsymbol{\theta}}F_{t-1}\|^2 + 3\mathbb{E}\|\nabla_{\boldsymbol{\theta}}F_{t-1} - \mathbf{v}_{t-1}\|^2 \\
\leq& 3L_F\mathbb{E}\|\boldsymbol{\theta}_{t-1} - \boldsymbol{\theta}_t\|^2 + 3L_F^2\mathbb{E}\|\boldsymbol{\omega}_{t-1} - \boldsymbol{\omega}_t\|^2. \qquad (47)
\end{aligned}$$

Thus, we have: $\sum_{t=1}^{T}\mathbb{E}\|\mathbf{v}_t - \mathbf{v}_{t-1}\|^2 \leq \sum_{t=1}^{T} 3L_F^2\mathbb{E}\|\boldsymbol{\theta}_{t-1} - \boldsymbol{\theta}_t\|^2 + 3L_F^2\mathbb{E}\|\boldsymbol{\omega}_{t-1} - \boldsymbol{\omega}_t\|^2$, and similarly, $\sum_{t=1}^{T}\mathbb{E}\|\mathbf{u}_t - \mathbf{u}_{t-1}\|^2 \leq \sum_{t=1}^{T} 3L_F^2\mathbb{E}\|\boldsymbol{\theta}_{t-1} - \boldsymbol{\theta}_t\|^2 + 3L_F^2\mathbb{E}\|\boldsymbol{\omega}_{t-1} - \boldsymbol{\omega}_t\|^2$. □

Now, we show the final step for proving Theorem 1. With Lemmas 3-5 and the defined potential function, we have:

$$\begin{aligned}
\mathbb{E}\mathfrak{P}_{T+1} - \mathfrak{P}_0 \leq& -\frac{\gamma}{2}\sum_{t=0}^{T}\mathbb{E}\|\nabla J(\bar{\boldsymbol{\theta}}_t)\|^2 - \gamma L_f^2\sum_{t=0}^{T}\mathbb{E}\|\boldsymbol{\omega}_t^* - \bar{\boldsymbol{\omega}}_t\|^2 \\
&-(\frac{\gamma}{2} - \frac{L_J\gamma^2}{2} - \frac{40\gamma^3L_{\boldsymbol{\omega}}^2L_F^2}{\mu^2\eta^2})\sum_{t=0}^{T}\mathbb{E}\|\bar{\mathbf{p}}_t\|^2 - \frac{2\gamma\eta L_F^2}{\mu}\sum_{t=0}^{T}\mathbb{E}\|\bar{\mathbf{d}}_t\|^2 \\
&-(1-(1+c_1)\lambda^2 - \frac{72\gamma L_F^4}{\mu^2} - 2\gamma L_F^2)\sum_{t=0}^{T}\frac{\mathcal{E}(\boldsymbol{\theta}_t)}{m} \\
&-(1-(1+c_2)\lambda^2 - \frac{72\gamma L_F^4}{\mu^2} - 2\gamma L_F^2)\sum_{t=0}^{T}\frac{\mathcal{E}(\boldsymbol{\omega}_t)}{m} \\
&-\big(1-(1+c_1)\lambda^2 - (1+\frac{1}{c_1})\gamma\big)\gamma\sum_{t=0}^{T}\frac{\mathcal{E}(\mathbf{p}_t)}{m} \\
&-\big(1-(1+c_2)\lambda^2 - (1+\frac{1}{c_2})\eta\big)\eta\sum_{t=0}^{T}\frac{\mathcal{E}(\mathbf{d}_t)}{m} \\
&\underbrace{+\frac{2\gamma}{m}\sum_{t=0}^{T}\mathbb{E}\|\nabla_{\boldsymbol{\omega}}F_t - \mathbf{v}_t\|^2 + \frac{72L_F^2\gamma}{m\mu^2}\sum_{t=0}^{T}\mathbb{E}\|\nabla_{\boldsymbol{\theta}}F_t - \mathbf{u}_t\|^2}_{R_1} \\
&\underbrace{+(1+\frac{1}{c_1})\frac{\gamma}{m}\sum_{t=1}^{T}\mathbb{E}\|\mathbf{v}_t - \mathbf{v}_{t-1}\|^2 + (1+\frac{1}{c_2})\frac{\eta}{m}\sum_{t=1}^{T}\mathbb{E}\|\mathbf{u}_t - \mathbf{u}_{t-1}\|^2}_{R_2}. \qquad (48)
\end{aligned}$$

First, we have $R_1 = 0$ because of the full gradient evaluation. R_2 can be bounded by

$$
\begin{aligned}
R_2 &= (1+\frac{1}{c_1})\frac{\gamma}{m}\sum_{t=1}^{T}\mathbb{E}\|\mathbf{v}_t-\mathbf{v}_{t-1}\|^2+(1+\frac{1}{c_2})\frac{\eta}{m}\sum_{t=1}^{T}\mathbb{E}\|\mathbf{u}_t-\mathbf{u}_{t-1}\|^2\\
&\overset{(a)}{\leq}\Big((1+\frac{1}{c_1})\frac{\gamma}{m}+(1+\frac{1}{c_2})\frac{\eta}{m}\Big)\sum_{t=1}^{T}3L_F^2\Big(\mathbb{E}\|\boldsymbol{\theta}_{t-1}-\boldsymbol{\theta}_t\|^2+\mathbb{E}\|\boldsymbol{\omega}_{t-1}-\boldsymbol{\omega}_t\|^2\Big)\\
&\overset{(b)}{\leq}3L_F^2\Big((1+\frac{1}{c_1})\frac{\gamma}{m}+(1+\frac{1}{c_2})\frac{\eta}{m}\Big)\sum_{t=1}^{T}\Big(8\mathcal{E}(\boldsymbol{\theta}_{t-1})+4\gamma^2\mathcal{E}(\mathbf{p}_{t-1})+4\gamma^2 m\|\bar{\mathbf{p}}_{t-1}\|^2\Big)\\
&\quad+3L_F^2\Big((1+\frac{1}{c_1})\frac{\gamma}{m}+(1+\frac{1}{c_2})\frac{\eta}{m}\Big)\sum_{t=1}^{T}\Big(8\mathcal{E}(\boldsymbol{\omega}_{t-1})+4\eta^2\mathcal{E}(\mathbf{d}_{t-1})+4\eta^2 m\|\bar{\mathbf{d}}_{t-1}\|^2\Big),
\end{aligned}
\tag{49}
$$

where (a) is from Lemma 5 and (b) is from Lemma 4.

Thus, we have

$$
\begin{aligned}
\mathbb{E}\mathfrak{P}_{T+1}-\mathfrak{P}_0 \leq& -\frac{\gamma}{2}\sum_{t=0}^{T}\mathbb{E}\|\nabla J(\bar{\boldsymbol{\theta}}_t)\|^2-\gamma L_f^2\sum_{t=0}^{T}\mathbb{E}\|\boldsymbol{\omega}_t^*-\bar{\boldsymbol{\omega}}_t\|^2\\
&-\Big(\frac{\gamma}{2}-\frac{L_J\gamma^2}{2}-\frac{40\gamma^3L_{\boldsymbol{\omega}}^2L_F^2}{\mu^2\eta^2}-12\gamma^2L_F^2\big((1+\frac{1}{c_1})\gamma+(1+\frac{1}{c_2})\eta\big)\Big)\sum_{t=0}^{T}\mathbb{E}\|\bar{\mathbf{p}}_t\|^2\\
&-\Big(\frac{2\gamma\eta L_F^2}{\mu}-12\eta^2L_F^2\big((1+\frac{1}{c_1})\gamma+(1+\frac{1}{c_2})\eta\big)\Big)\sum_{t=0}^{T}\mathbb{E}\|\bar{\mathbf{d}}_t\|^2\\
&-\Big(1-(1+c_1)\lambda^2-\frac{72\gamma L_F^4}{\mu^2}-2\gamma L_F^2-24L_F^2\big((1+\frac{1}{c_1})\gamma+(1+\frac{1}{c_2})\eta\big)\Big)\sum_{t=0}^{T}\frac{\mathcal{E}(\boldsymbol{\theta}_t)}{m}\\
&-\Big(1-(1+c_2)\lambda^2-\frac{72\gamma L_F^4}{\mu^2}-2\gamma L_F^2-24L_F^2\big((1+\frac{1}{c_1})\gamma+(1+\frac{1}{c_2})\eta\big)\Big)\sum_{t=0}^{T}\frac{\mathcal{E}(\boldsymbol{\omega}_t)}{m}\\
&-\Big(1-(1+c_1)\lambda^2-(1+\frac{1}{c_1})\gamma-12\gamma L_F^2\big((1+\frac{1}{c_1})\gamma+(1+\frac{1}{c_2})\eta\big)\Big)\gamma\sum_{t=0}^{T}\frac{\mathcal{E}(\mathbf{p}_t)}{m}\\
&-\Big(1-(1+c_2)\lambda^2-(1+\frac{1}{c_2})\eta-12\eta L_F^2\big((1+\frac{1}{c_1})\gamma+(1+\frac{1}{c_2})\eta\big)\Big)\eta\sum_{t=0}^{T}\frac{\mathcal{E}(\mathbf{d}_t)}{m}.
\end{aligned}
\tag{50}
$$

Choosing $c_1 = c_2 = 1/\lambda - 1$, we have

$$
\begin{aligned}
\mathbb{E}\mathfrak{P}_{T+1}-\mathfrak{P}_0 \leq& -\frac{\gamma}{2}\sum_{t=0}^{T}\mathbb{E}\|\nabla J(\bar{\boldsymbol{\theta}}_t)\|^2-\gamma L_f^2\sum_{t=0}^{T}\mathbb{E}\|\boldsymbol{\omega}_t^*-\bar{\boldsymbol{\omega}}_t\|^2-C_{\bar{\mathbf{p}}}\gamma\sum_{t=0}^{T}\mathbb{E}\|\bar{\mathbf{p}}_t\|^2\\
&-C_{\bar{\mathbf{d}}}\gamma\eta\sum_{t=0}^{T}\mathbb{E}\|\bar{\mathbf{d}}_t\|^2-C_{\boldsymbol{\theta}}\sum_{t=0}^{T}\frac{\mathcal{E}(\boldsymbol{\theta}_t)}{m}-C_{\boldsymbol{\omega}}\sum_{t=0}^{T}\frac{\mathcal{E}(\boldsymbol{\omega}_t)}{m}-C_{\mathbf{p}}\gamma\sum_{t=0}^{T}\frac{\mathcal{E}(\mathbf{p}_t)}{m}-C_{\mathbf{d}}\eta\sum_{t=0}^{T}\frac{\mathcal{E}(\mathbf{d}_t)}{m},
\end{aligned}
\tag{51}
$$

where the constants are

$$C_{\bar{\mathbf{p}}} = \frac{1}{2} - \frac{L_J\gamma}{2} - \frac{40\gamma^2 L_{\boldsymbol{\omega}}^2 L_F^2}{\mu^2\eta^2} - \frac{12\gamma L_F^2(\gamma+\eta)}{1-\lambda}, \tag{52}$$

$$C_{\bar{\mathbf{d}}} = \frac{2L_F^2}{\mu} - \frac{12\eta L_F^2(1+\eta/\gamma)}{1-\lambda}, \tag{53}$$

$$C_{\boldsymbol{\theta}} = C_{\boldsymbol{\omega}} = 1-\lambda - \frac{72\gamma L_F^4}{\mu^2} - 2\gamma L_F^2 - \frac{24L_F^2(\gamma+\eta)}{1-\lambda}, \tag{54}$$

$$C_{\mathbf{p}} = 1-\lambda - \frac{\gamma}{1-\lambda} - \frac{12\gamma L_F^2(\gamma+\eta)}{1-\lambda}, \tag{55}$$

$$C_{\mathbf{d}} = 1-\lambda - \frac{\eta}{1-\lambda} - \frac{12\eta L_F^2(\gamma+\eta)}{1-\lambda}. \tag{56}$$

To ensure $C_{\bar{\mathbf{p}}} \geq 0$, we have

$$\begin{aligned} C_{\bar{\mathbf{p}}} &\overset{(a)}{\geq} \frac{1}{4} - \frac{L_J\gamma}{2} - \frac{12\gamma^2 L_F^2(1+\eta/\gamma)}{1-\lambda} \\ &\overset{(b)}{\geq} \frac{1}{4} - \frac{(L_f + L_f^2/\mu)\gamma}{2} - \frac{(1-\lambda)\gamma}{2} \overset{(c)}{\geq} 0, \end{aligned} \tag{57}$$

where (a) follows from $\kappa := \gamma/\eta \leq \mu^2/13L_F^2$ and Lemma 9, (b) is due to $\gamma \leq (1-\lambda)^2/24L_F^2(1+1/\kappa)$ and Lemma 11, and (c) follows from $\gamma \leq 1/2\big((L_F + L_F^2/\mu) + (1-\lambda)\big)$. By setting $\eta \leq (1-\lambda)/6\mu(1+1/\kappa)$, we have $C_{\bar{\mathbf{d}}} \geq 0$. By setting $\gamma \leq (1-\lambda)/(\frac{1}{2} + \frac{72L_F^4}{\mu^2} + 2L_F^2 + \frac{24L_F^2(1+1/\kappa)}{1-\lambda})$, we have $C_{\boldsymbol{\theta}} = C_{\boldsymbol{\omega}} \geq \gamma/2$. To ensure $C_{\mathbf{p}} \geq 0$,

$$\begin{aligned} C_{\mathbf{p}} &= 1-\lambda - \frac{\gamma}{1-\lambda} - \frac{12\gamma^2 L_F^2(1+\eta/\gamma)}{1-\lambda} \\ &\overset{(a)}{\geq} 1-\lambda - \frac{\gamma}{1-\lambda} - \frac{(1-\lambda)\gamma}{2} \overset{(b)}{\geq} 0, \end{aligned} \tag{58}$$

where (a) is by $\gamma \leq (1-\lambda)^2/24L_F^2(1+1/\kappa)$ and (b) is by $\gamma \leq 1/\big(1/2 + 1/(1-\lambda)^2\big)$. Similarly, with $\eta \leq 1/\big(1/2 + 1/(1-\lambda)^2\big)$, we have $C_{\mathbf{d}} \geq 0$.

To summarize, we need the following conditions to ensure $C_{\bar{\mathbf{p}}} \geq 0$, $C_{\bar{\mathbf{d}}} \geq 0$, $C_{\mathbf{p}} \geq 0$, $C_{\mathbf{d}} \geq 0$, $C_{\boldsymbol{\theta}} \geq \gamma/2$, $C_{\boldsymbol{\omega}} \geq \gamma/2$,

$$\kappa = \gamma/\eta \leq \mu^2/13L_F^2 \tag{59}$$

$$\gamma \leq \min\Big\{\frac{1}{2}\Big(L_F + \frac{L_F^2}{\mu} + (1-\lambda)\Big), \frac{1-\lambda}{(\frac{1}{2} + \frac{72L_F^4}{\mu^2} + 2L_F^2 + \frac{24L_F^2(1+1/\kappa)}{1-\lambda})}, \big(1/2 + 1/(1-\lambda)^2\big)^{-1}\Big\}, \tag{60}$$

$$\eta \leq \min\Big\{\frac{(1-\lambda)}{6\mu(1+1/\kappa)}, \big(1/2 + 1/(1-\lambda)^2\big)^{-1}\Big\}, \tag{61}$$

which is satisfied with

$$\kappa = \gamma/\eta \leq \mu^2/13L_F^2 \tag{62}$$

$$\begin{aligned} \eta \leq \min\Big\{&\frac{13L_F^2}{2\mu^2}\Big(L_F + \frac{L_F^2}{\mu} + (1-\lambda)\Big), \frac{26(1-\lambda)L_F^2}{(\mu^2 + 144L_F^4 + 4L_F^2\mu^2 + \frac{48\mu^2 L_F^2(1+1/\kappa)}{1-\lambda})}, \\ &\frac{(1-\lambda)}{6\mu(1+1/\kappa)}, \frac{13L_F^2}{\mu^2\big(1/2 + 1/(1-\lambda)^2\big)}\Big\}. \end{aligned} \tag{63}$$

Note that $\frac{(1-\lambda)}{6\mu(1+1/\kappa)} \leq \frac{1}{2L_F}$, which satisfies the step-size condition in Lemma 2 and Lemma 3.

Thus, with such parameter settings, we have

$$\frac{\gamma}{2}\sum_{t=0}^{T}\Big(\mathbb{E}\|\nabla J(\bar{\boldsymbol{\theta}}_t)\|^2 + 2L_f^2\mathbb{E}\|\boldsymbol{\omega}_t^* - \bar{\boldsymbol{\omega}}_t\|^2 + \frac{\mathcal{E}(\boldsymbol{\theta}_t)}{m} + \frac{\mathcal{E}(\boldsymbol{\omega}_t)}{m}\Big) \leq \mathbb{E}\mathfrak{P}_0 - \mathfrak{P}_{T+1}, \tag{64}$$

which yields the final result by multiplying $2/(T+1)\gamma$ on both sides.

C Proof for Theorem 3 and Theorem 4

In this section, we provide the proofs of Theorem 3-4. Under Assumption 2, we can easily obtain that funciton F_i satisifies L_f-Lipschitz smoothness. Thus, we have the following modified descending result:

Lemma 6. *Under Assumption 1 and set $\eta \leq 1/2L_f$, the following inequality holds for GT-SRVR and GT-SRVR*I*:*

$$\begin{aligned}
&J(\bar{\boldsymbol{\theta}}_{T+1}) - J(\bar{\boldsymbol{\theta}}_0) + \frac{8\gamma L_f^2}{\mu\eta}\big[\|\bar{\boldsymbol{\omega}}_{T+1} - \boldsymbol{\omega}_{T+1}^*\|^2 - \|\boldsymbol{\omega}_0^* - \bar{\boldsymbol{\omega}}_0\|^2\big] \\
\leq& -\frac{\gamma}{2}\sum_{t=0}^{T}\|\nabla J(\bar{\boldsymbol{\theta}}_t)\|^2 - \gamma L_f^2\sum_{t=0}^{T}\|\boldsymbol{\omega}_t^* - \bar{\boldsymbol{\omega}}_t\|^2 - (\frac{\gamma}{2} - \frac{L_J\gamma^2}{2} - \frac{40\gamma^3 L_{\boldsymbol{\omega}}^2 L_f^2}{\mu^2\eta^2})\sum_{t=0}^{T}\|\bar{\mathbf{p}}_t\|^2 - \frac{2\gamma\eta L_f^2}{\mu}\sum_{t=0}^{T}\|\bar{\mathbf{d}}_t\|^2 \\
&+ (\frac{72\gamma L_f^4}{\mu^2 m} + \frac{2\gamma L_f^2}{m})\sum_{t=0}^{T}[\mathcal{E}(\boldsymbol{\theta}_t) + \mathcal{E}(\boldsymbol{\omega}_t)] + \frac{2\gamma}{m}\sum_{t=0}^{T}\|\nabla_{\boldsymbol{\omega}} F_t - \mathbf{v}_t\|^2 + \frac{72L_f^2\gamma}{m\mu^2}\sum_{t=0}^{T}\|\nabla_{\boldsymbol{\theta}} F_t - \mathbf{u}_t\|^2.
\end{aligned}$$

Next, we bound the error of the gradient estimators as the follows:

Lemma 7 (Error of Gradient Estimator)**.** *Under Assumption 2, we have the following error bounds for the estimators:*

$$\begin{aligned}
\sum_{t=0}^{T}\|\mathbf{v}_t - \nabla_{\boldsymbol{\theta}} F_t\|^2 \leq& \sum_{t=1}^{T}\mathbb{E}\|\mathbf{v}_{(n_t-1)q} - \nabla_{\boldsymbol{\theta}} F(\boldsymbol{\theta}_{(n_t-1)q}, \boldsymbol{\omega}_{(n_t-1)q})\|^2 \\
&+ L_f^2\big(\|\boldsymbol{\theta}_t - \boldsymbol{\theta}_{t-1}\|^2 + \|\boldsymbol{\omega}_t - \boldsymbol{\omega}_{t-1}\|^2\big), \qquad (65) \\
\sum_{t=0}^{T}\|\mathbf{u}_t - \nabla_{\boldsymbol{\omega}} F_t\|^2 \leq& \sum_{t=1}^{T}\mathbb{E}\|\mathbf{u}_{(n_t-1)q} - \nabla_{\boldsymbol{\omega}} F(\boldsymbol{\theta}_{(n_t-1)q}, \boldsymbol{\omega}_{(n_t-1)q})\|^2 \\
&+ L_f^2\big(\|\boldsymbol{\theta}_t - \boldsymbol{\theta}_{t-1}\|^2 + \|\boldsymbol{\omega}_t - \boldsymbol{\omega}_{t-1}\|^2\big), \qquad (66)
\end{aligned}$$

where n_t is the largest positive integer that satisfies $(n_t - 1)q \leq t$.

Proof. From the algorithm update, we have:

$$\begin{aligned}
\|\underbrace{\mathbf{v}_{i,t}-\nabla_{\boldsymbol{\theta}}F_{i,t}}_{A_{i,t}}\|^2 &= \|\mathbf{v}_{i,t-1}+\frac{1}{|\mathcal{S}_{i,t}|}\sum_{j\in\mathcal{S}_{i,t}}\nabla_{\boldsymbol{\theta}}f_{i,j}(\boldsymbol{\theta}_{i,t},\boldsymbol{\omega}_{i,t})-\nabla_{\boldsymbol{\theta}}f_{i,j}(\boldsymbol{\theta}_{i,t-1},\boldsymbol{\omega}_{i,t-1})-\nabla_{\boldsymbol{\theta}}F_{i,t}\|^2 \\
&= \|\underbrace{\mathbf{v}_{i,t-1}-\nabla_{\boldsymbol{\theta}}F_{i,t-1}}_{A_{i,t-1}}+\underbrace{\frac{1}{|\mathcal{S}_{i,t}|}\sum_{j\in\mathcal{S}_{i,t}}\nabla_{\boldsymbol{\theta}}f_{i,t}(\boldsymbol{\theta}_{i,t},\boldsymbol{\omega}_{i,t})-\nabla_{\boldsymbol{\theta}}f_{i,t}(\boldsymbol{\theta}_{i,t-1},\boldsymbol{\omega}_{i,t-1})+\nabla_{\boldsymbol{\theta}}F_{i,t-1}-\nabla_{\boldsymbol{\theta}}F_{i,t}}_{B_{i,t}}\|^2 \\
&= \|A_{i,t-1}\|^2+\|B_{i,t}\|^2+2\langle A_{i,t-1},B_{i,t}\rangle. \qquad (67)
\end{aligned}$$

Note that $\mathbb{E}_t[B_{i,t}]=0$, where the expectation is taken over the randomness in tth iteration. Thus,

$$\mathbb{E}_t\|A_{i,t}\|^2 = \|A_{i,t-1}\|^2+\mathbb{E}_t\|B_{i,t}\|^2. \qquad (68)$$

Also, with $|\mathcal{S}_{i,t}|=q$, we have

$$\begin{aligned}
\mathbb{E}_t\|B_{i,t}\|^2 &= \mathbb{E}_t\|\frac{1}{|\mathcal{S}_{i,t}|}\sum_{j\in\mathcal{S}_{i,t}}\nabla_{\boldsymbol{\theta}}f_{i,j}(\boldsymbol{\theta}_{i,t},\boldsymbol{\omega}_{i,t})-\nabla_{\boldsymbol{\theta}}f_{i,j}(\boldsymbol{\theta}_{i,t-1},\boldsymbol{\omega}_{i,t-1})-\nabla_{\boldsymbol{\theta}}F_{i,t}+\nabla_{\boldsymbol{\theta}}F_{i,t-1}\|^2 \\
&\le \frac{1}{|\mathcal{S}_{i,t}|^2}\sum_{j\in\mathcal{S}_{i,t}}\mathbb{E}_t\|\nabla_{\boldsymbol{\theta}}f_{i,j}(\boldsymbol{\theta}_{i,t},\boldsymbol{\omega}_{i,t})-\nabla_{\boldsymbol{\theta}}f_{i,j}(\boldsymbol{\theta}_{i,t-1},\boldsymbol{\omega}_{i,t-1})-\nabla_{\boldsymbol{\theta}}F_{i,t}+\nabla_{\boldsymbol{\theta}}F_{i,t-1}\|^2 \\
&\le \frac{L_f^2}{q}\big(\|\boldsymbol{\theta}_{i,t}-\boldsymbol{\theta}_{i,t-1}\|^2+\|\boldsymbol{\omega}_{i,t}-\boldsymbol{\omega}_{i,t-1}\|^2\big). \qquad (69)
\end{aligned}$$

Taking full expectation and telescoping (69) over t from $(n_t-1)q+1$ to t, where $t\le n_tq-1$, we have

$$\begin{aligned}
\mathbb{E}\|A_t\|^2 &\le \mathbb{E}\|A_{t-1}\|^2+\frac{L_f^2}{q}\mathbb{E}\big(\|\boldsymbol{\theta}_t-\boldsymbol{\theta}_{t-1}\|^2+\|\boldsymbol{\omega}_t-\boldsymbol{\omega}_{t-1}\|^2\big) \\
&\le \mathbb{E}\|A_{(n_t-1)q}\|^2+\sum_{r=(n_t-1)q+1}^{t}\frac{L_f^2}{q}\mathbb{E}\big(\|\boldsymbol{\theta}_r-\boldsymbol{\theta}_{r-1}\|^2+\|\boldsymbol{\omega}_r-\boldsymbol{\omega}_{r-1}\|^2\big). \qquad (70)
\end{aligned}$$

Thus, we have:

$$\begin{aligned}
&\sum_{k=0}^{t}\mathbb{E}\|A_k\|^2=\sum_{k=0}^{q-1}\mathbb{E}\|A_k\|^2+\cdots+\sum_{k=(n_t-1)q}^{t}\mathbb{E}\|A_k\|^2\\
\leq& q\|A_0\|^2+\sum_{k=1}^{q-1}\sum_{r=1}^{k}\frac{L_f^2}{q}\big(\|\boldsymbol{\theta}_r-\boldsymbol{\theta}_{r-1}\|^2+\|\boldsymbol{\omega}_r-\boldsymbol{\omega}_{r-1}\|^2\big)\\
&+\cdots\\
&+\big(t-(n_t-1)q\big)\|A_{(n_t-1)q}\|^2+\sum_{k=(n_t-1)q+1}^{t}\sum_{r=(n_t-1)q+1}^{k}\frac{L_f^2}{q}\big(\|\boldsymbol{\theta}_r-\boldsymbol{\theta}_{r-1}\|^2+\|\boldsymbol{\omega}_r-\boldsymbol{\omega}_{r-1}\|^2\big)\\
\leq& q\|A_0\|^2+\sum_{r=1}^{q-1}\sum_{k=r}^{q-1}\frac{L_f^2}{q}\big(\|\boldsymbol{\theta}_r-\boldsymbol{\theta}_{r-1}\|^2+\|\boldsymbol{\omega}_r-\boldsymbol{\omega}_{r-1}\|^2\big)\\
&+\cdots\\
&+\big(t-(n_t-1)q\big)\|A_{(n_t-1)q}\|^2+\sum_{r=(n_t-1)q+1}^{t}\sum_{k=r}^{t}\frac{L_f^2}{q}\big(\|\boldsymbol{\theta}_r-\boldsymbol{\theta}_{r-1}\|^2+\|\boldsymbol{\omega}_r-\boldsymbol{\omega}_{r-1}\|^2\big)\\
\leq& q\|A_0\|^2+\sum_{r=1}^{q-1}L_f^2\big(\|\boldsymbol{\theta}_r-\boldsymbol{\theta}_{r-1}\|^2+\|\boldsymbol{\omega}_r-\boldsymbol{\omega}_{r-1}\|^2\big)\\
&+\cdots\\
&+\big(t-(n_t-1)q\big)\|A_{(n_t-1)q}\|^2+\sum_{r=(n_t-1)q+1}^{t}L_f^2\big(\|\boldsymbol{\theta}_r-\boldsymbol{\theta}_{r-1}\|^2+\|\boldsymbol{\omega}_r-\boldsymbol{\omega}_{r-1}\|^2\big)\\
=&\sum_{r=0}^{t}\|A_{(n_r-1)q}\|^2+\sum_{r=1}^{t}L_f^2\big(\|\boldsymbol{\theta}_r-\boldsymbol{\theta}_{r-1}\|^2+\|\boldsymbol{\omega}_r-\boldsymbol{\omega}_{r-1}\|^2\big).
\end{aligned}\tag{71}$$

Thus, we have:

$$\sum_{t=0}^{T}\|\mathbf{v}_t-\nabla_{\boldsymbol{\theta}}F_t\|^2\leq\sum_{t=0}^{T}\mathbb{E}\|\mathbf{v}_{(n_t-1)q}-\nabla_{\boldsymbol{\theta}}F_{(n_t-1)q})\|^2+\sum_{t=1}^{T}L_f^2\big(\|\boldsymbol{\theta}_t-\boldsymbol{\theta}_{t-1}\|^2+\|\boldsymbol{\omega}_t-\boldsymbol{\omega}_{t-1}\|^2\big)\tag{72}$$

Similarly, we have:

$$\sum_{t=0}^{T}\|\mathbf{u}_t-\nabla_{\boldsymbol{\omega}}F_t\|^2\leq\sum_{t=0}^{T}\mathbb{E}\|\mathbf{u}_{(n_t-1)q}-\nabla_{\boldsymbol{\omega}}F_{(n_t-1)q})\|^2+\sum_{t=1}^{T}L_f^2\big(\|\boldsymbol{\theta}_t-\boldsymbol{\theta}_{t-1}\|^2+\|\boldsymbol{\omega}_t-\boldsymbol{\omega}_{t-1}\|^2\big).\tag{73}$$

□

Lemma 8 (Differential Bound on Estimator for GT-SRVR and GT-SRVRI)**.** *Under Assumption 2, the following inequalities holds*

$$\begin{aligned}
\sum_{t=1}^{T}\mathbb{E}\|\mathbf{v}_t-\mathbf{v}_{t-1}\|^2\leq&\sum_{t=1}^{T}4L_f^2\mathbb{E}\|\boldsymbol{\theta}_{t-1}-\boldsymbol{\theta}_t\|^2+4L_f^2\mathbb{E}\|\boldsymbol{\omega}_{t-1}-\boldsymbol{\omega}_t\|^2+\sum_{t=0}^{T}6\mathbb{E}\|\mathbf{v}_{(n_t-1)q}-\nabla_{\boldsymbol{\theta}}F_{(n_t-1)q}\|^2,\\
\sum_{t=1}^{T}\mathbb{E}\|\mathbf{u}_t-\mathbf{u}_{t-1}\|^2\leq&\sum_{t=1}^{T}4L_f^2\mathbb{E}\|\boldsymbol{\theta}_{t-1}-\boldsymbol{\theta}_t\|^2+4L_f^2\mathbb{E}\|\boldsymbol{\omega}_{t-1}-\boldsymbol{\omega}_t\|^2+\sum_{t=0}^{T}6\mathbb{E}\|\mathbf{u}_{(n_t-1)q}-\nabla_{\boldsymbol{\omega}}F_{(n_t-1)q}\|^2.
\end{aligned}$$

Proof. For $\|\mathbf{v}_t - \mathbf{v}_{t-1}\|^2$, we have i) when $t \in \big((n_t-1)q, n_tq-1\big] \cap \mathbb{Z}$,

$$\begin{aligned}
&\mathbb{E}\|\mathbf{v}_t - \mathbf{v}_{t-1}\|^2 = \sum_{i=1}^{m} \mathbb{E}\|\frac{1}{|\mathcal{S}_{i,t}|}\sum_{j\in\mathcal{S}_{i,t}} \nabla_{\boldsymbol{\theta}} f_{i,j}(\boldsymbol{\theta}_{i,t}, \boldsymbol{\omega}_{i,t}) - \nabla_{\boldsymbol{\theta}} f_{i,k}(\boldsymbol{\theta}_{i,t-1}, \boldsymbol{\omega}_{i,t-1})\|^2 \\
\leq& \frac{1}{|\mathcal{S}_{i,t}|^2}\sum_{i=1}^{m}\sum_{j\in\mathcal{S}_{i,t}} \mathbb{E}\|\nabla_{\boldsymbol{\theta}} f_{i,j}(\boldsymbol{\theta}_{i,t}, \boldsymbol{\omega}_{i,t}) - \nabla_{\boldsymbol{\theta}} f_{i,j}(\boldsymbol{\theta}_{i,t-1}, \boldsymbol{\omega}_{i,t-1})\|^2 \\
\overset{(a)}{\leq}& L_f^2 \sum_{i=1}^{m} \mathbb{E}\|\boldsymbol{\theta}_{i,t-1} - \boldsymbol{\theta}_{i,t}\|^2 + L_f^2 \sum_{i=1}^{m} \mathbb{E}\|\boldsymbol{\omega}_{i,t-1} - \boldsymbol{\omega}_{i,t}\|^2 \\
=& L_f^2 \mathbb{E}\|\boldsymbol{\theta}_{t-1} - \boldsymbol{\theta}_t\|^2 + L_f^2 \mathbb{E}\|\boldsymbol{\omega}_{t-1} - \boldsymbol{\omega}_t\|^2,
\end{aligned} \tag{74}$$

where (a) is by $q \geq 1$ and Assumption 2.

ii) when $t = n_tq$ and $t > 0$,

$$\begin{aligned}
&\mathbb{E}\|\mathbf{v}_t - \mathbf{v}_{t-1}\|^2 = \mathbb{E}\|\mathbf{v}_t - \nabla_{\boldsymbol{\theta}} F_t + \nabla_{\boldsymbol{\theta}} F_t - \nabla_{\boldsymbol{\theta}} F_{t-1} + \nabla_{\boldsymbol{\theta}} F_{t-1} - \mathbf{v}_{t-1}\|^2 \\
\leq& 3\mathbb{E}\|\mathbf{v}_t - \nabla_{\boldsymbol{\theta}} F_t\|^2 + 3\mathbb{E}\|\nabla_{\boldsymbol{\theta}} F_t - \nabla_{\boldsymbol{\theta}} F_{t-1}\|^2 + 3\mathbb{E}\|\nabla_{\boldsymbol{\theta}} F_{t-1} - \mathbf{v}_{t-1}\|^2 \\
\overset{(a)}{\leq}& 3\mathbb{E}\|\mathbf{v}_t - \nabla_{\boldsymbol{\theta}} F_t\|^2 + 3\mathbb{E}\|\nabla_{\boldsymbol{\theta}} F_{t-1} - \mathbf{v}_{t-1}\|^2 + 3L_f^2\mathbb{E}\|\boldsymbol{\theta}_{t-1} - \boldsymbol{\theta}_t\|^2 + 3L_f^2\mathbb{E}\|\boldsymbol{\omega}_{t-1} - \boldsymbol{\omega}_t\|^2 \\
\overset{(b)}{\leq}& 3\mathbb{E}\|\mathbf{v}_{n_tq} - \nabla_{\boldsymbol{\theta}} F_{n_tq}\|^2 + 3\mathbb{E}\|\mathbf{v}_{(n_t-1)q} - \nabla_{\boldsymbol{\theta}} F_{(n_t-1)q}\|^2 \\
&+ \sum_{r=(n_t-1)q+1}^{n_tq-1} \frac{3L_f^2}{q}\mathbb{E}\big(\|\boldsymbol{\theta}_r - \boldsymbol{\theta}_{r-1}\|^2 + \|\boldsymbol{\omega}_r - \boldsymbol{\omega}_{r-1}\|^2\big) \\
&+ 3L_f^2\mathbb{E}\|\boldsymbol{\theta}_{n_tq-1} - \boldsymbol{\theta}_{n_tq}\|^2 + 3L_f^2\mathbb{E}\|\boldsymbol{\omega}_{n_tq-1} - \boldsymbol{\omega}_{n_tq}\|^2,
\end{aligned} \tag{75}$$

where (a) is by Assumption 2, and (b) is by setting $t = n_tq$ and Lemma 7.

Telescoping (75) from $r = (n_t-1)q+1$ to n_tq,

$$\begin{aligned}
\sum_{r=(n_t-1)q+1}^{n_tq} \mathbb{E}\|\mathbf{v}_r - \mathbf{v}_{r-1}\|^2 \leq& \sum_{r=(n_t-1)q+1}^{n_tq-1} L_f^2\mathbb{E}\|\boldsymbol{\theta}_{r-1} - \boldsymbol{\theta}_r\|^2 + L_f^2\mathbb{E}\|\boldsymbol{\omega}_{r-1} - \boldsymbol{\omega}_r\|^2 \\
&+ 3\mathbb{E}\|\mathbf{v}_{n_tq} - \nabla_{\boldsymbol{\theta}} F_{n_tq}\|^2 + 3\mathbb{E}\|\mathbf{v}_{(n_t-1)q} - \nabla_{\boldsymbol{\theta}} F_{(n_t-1)q}\|^2 \\
&+ \sum_{r=(n_t-1)q+1}^{n_tq-1} \frac{3L_f^2}{q}\mathbb{E}\big(\|\boldsymbol{\theta}_r - \boldsymbol{\theta}_{r-1}\|^2 + \|\boldsymbol{\omega}_r - \boldsymbol{\omega}_{r-1}\|^2\big) \\
&+ 3L_f^2\mathbb{E}\|\boldsymbol{\theta}_{n_tq-1} - \boldsymbol{\theta}_{n_tq}\|^2 + 3L_f^2\mathbb{E}\|\boldsymbol{\omega}_{n_tq-1} - \boldsymbol{\omega}_{n_tq}\|^2 \\
\leq& \sum_{r=(n_t-1)q+1}^{n_tq} 4L_f^2\mathbb{E}\|\boldsymbol{\theta}_{r-1} - \boldsymbol{\theta}_r\|^2 + 4L_f^2\mathbb{E}\|\boldsymbol{\omega}_{r-1} - \boldsymbol{\omega}_r\|^2 \\
&+ 3\mathbb{E}\|\mathbf{v}_{n_tq} - \nabla_{\boldsymbol{\theta}} F_{n_tq}\|^2 + 3\mathbb{E}\|\mathbf{v}_{(n_t-1)q} - \nabla_{\boldsymbol{\theta}} F_{(n_t-1)q}\|^2,
\end{aligned} \tag{76}$$

which leads to that

$$\sum_{t=1}^{T} \mathbb{E}\|\mathbf{v}_t - \mathbf{v}_{t-1}\|^2 \leq \sum_{t=1}^{T} 4L_f^2\mathbb{E}\|\boldsymbol{\theta}_{t-1} - \boldsymbol{\theta}_t\|^2 + 4L_f^2\mathbb{E}\|\boldsymbol{\omega}_{t-1} - \boldsymbol{\omega}_t\|^2 + \sum_{t=0}^{T} 6\mathbb{E}\|\mathbf{v}_{(n_t-1)q} - \nabla_{\boldsymbol{\theta}} F_{(n_t-1)q}\|^2.$$

Similarly, we have

$$\sum_{t=1}^{T}\mathbb{E}\|\mathbf{u}_t-\mathbf{u}_{t-1}\|^2\leq\sum_{t=1}^{T}4L_f^2\mathbb{E}\|\boldsymbol{\theta}_{t-1}-\boldsymbol{\theta}_t\|^2+4L_f^2\mathbb{E}\|\boldsymbol{\omega}_{t-1}-\boldsymbol{\omega}_t\|^2+\sum_{t=0}^{T}6\mathbb{E}\|\mathbf{u}_{(n_t-1)q}-\nabla_{\boldsymbol{\omega}}F_{(n_t-1)q}\|^2.$$

□

With the defined potential function $\mathfrak{p}$, we have

$$\begin{aligned}
\mathbb{E}\mathfrak{p}_{T+1}-\mathfrak{p}_0\leq&-\frac{\gamma}{2}\sum_{t=0}^{T}\mathbb{E}\|\nabla J(\bar{\boldsymbol{\theta}}_t)\|^2-\gamma L_f^2\sum_{t=0}^{T}\mathbb{E}\|\boldsymbol{\omega}_t^*-\bar{\boldsymbol{\omega}}_t\|^2\\
&-(\frac{\gamma}{2}-\frac{L_J\gamma^2}{2}-\frac{40\gamma^3L_{\boldsymbol{\omega}}^2L_f^2}{\mu^2\eta^2})\sum_{t=0}^{T}\mathbb{E}\|\bar{\mathbf{p}}_t\|^2-\frac{2\gamma\eta L_f^2}{\mu}\sum_{t=0}^{T}\mathbb{E}\|\bar{\mathbf{d}}_t\|^2\\
&-(1-(1+c_1)\lambda^2-\frac{72\gamma L_f^4}{\mu^2}-2\gamma L_f^2)\sum_{t=0}^{T}\frac{\mathcal{E}(\boldsymbol{\theta}_t)}{m}\\
&-(1-(1+c_2)\lambda^2-\frac{72\gamma L_f^4}{\mu^2}-2\gamma L_f^2)\sum_{t=0}^{T}\frac{\mathcal{E}(\boldsymbol{\omega}_t)}{m}\\
&-\big(1-(1+c_1)\lambda^2-(1+\frac{1}{c_1})\gamma\big)\gamma\sum_{t=0}^{T}\frac{\mathcal{E}(\mathbf{p}_t)}{m}\\
&-\big(1-(1+c_2)\lambda^2-(1+\frac{1}{c_2})\eta\big)\eta\sum_{t=0}^{T}\frac{\mathcal{E}(\mathbf{d}_t)}{m}\\
&+\underbrace{\frac{2\gamma}{m}\sum_{t=0}^{T}\mathbb{E}\|\nabla_{\boldsymbol{\omega}}F_t-\mathbf{v}_t\|^2+\frac{72L_f^2\gamma}{m\mu^2}\sum_{t=0}^{T}\mathbb{E}\|\nabla_{\boldsymbol{\theta}}F_t-\mathbf{u}_t\|^2}_{R_1}\\
&+\underbrace{(1+\frac{1}{c_1})\frac{\gamma}{m}\sum_{t=1}^{T}\mathbb{E}\|\mathbf{v}_t-\mathbf{v}_{t-1}\|^2+(1+\frac{1}{c_2})\frac{\eta}{m}\sum_{t=1}^{T}\mathbb{E}\|\mathbf{u}_t-\mathbf{u}_{t-1}\|^2}_{R_2}.
\end{aligned}\tag{77}$$

First, for the term R_1, we have

$$\begin{aligned}
&\frac{2\gamma}{m}\sum_{t=0}^{T}\mathbb{E}\|\nabla_{\boldsymbol{\omega}}F_t-\mathbf{v}_t\|^2+\frac{72L_f^2\gamma}{m\mu^2}\sum_{t=0}^{T}\mathbb{E}\|\nabla_{\boldsymbol{\theta}}F_t-\mathbf{u}_t\|^2\\
\leq&\frac{2\gamma}{m}\mathbb{E}\Big(\sum_{t=0}^{T}\|\mathbf{v}_{(n_t-1)q}-\nabla_{\boldsymbol{\theta}}F_{(n_t-1)q}\|^2+\sum_{t=1}^{T}L_f^2\big(\|\boldsymbol{\theta}_t-\boldsymbol{\theta}_{t-1}\|^2+\|\boldsymbol{\omega}_t-\boldsymbol{\omega}_{t-1}\|^2\big)\Big)\\
&+\frac{72L_f^2\gamma}{m\mu^2}\mathbb{E}\Big(\sum_{t=0}^{T}\|\mathbf{u}_{(n_t-1)q}-\nabla_{\boldsymbol{\omega}}F_{(n_t-1)q}\|^2+\sum_{t=1}^{T}L_f^2\big(\|\boldsymbol{\theta}_t-\boldsymbol{\theta}_{t-1}\|^2+\|\boldsymbol{\omega}_t-\boldsymbol{\omega}_{t-1}\|^2\big)\Big)\\
=&L_f^2\big(\frac{2\gamma}{m}+\frac{72L_f^2\gamma}{m\mu^2}\big)\sum_{t=1}^{T}\mathbb{E}\big(\|\boldsymbol{\theta}_t-\boldsymbol{\theta}_{t-1}\|^2+\|\boldsymbol{\omega}_t-\boldsymbol{\omega}_{t-1}\|^2\big)\\
&+\frac{2\gamma}{m}\sum_{t=0}^{T}\mathbb{E}\|\mathbf{v}_{(n_t-1)q}-\nabla_{\boldsymbol{\theta}}F_{(n_t-1)q}\|^2+\frac{72L_f^2\gamma}{m\mu^2}\sum_{t=0}^{T}\mathbb{E}\|\mathbf{u}_{(n_t-1)q}-\nabla_{\boldsymbol{\omega}}F_{(n_t-1)q}\|^2.
\end{aligned}\tag{78}$$

Then, for term R_2, we can bound it as follows:

$$
\begin{aligned}
&(1+\frac{1}{c_1})\frac{\gamma}{m}\sum_{t=1}^{T}\mathbb{E}\|\mathbf{v}_t-\mathbf{v}_{t-1}\|^2+(1+\frac{1}{c_2})\frac{\eta}{m}\sum_{t=1}^{T}\mathbb{E}\|\mathbf{u}_t-\mathbf{u}_{t-1}\|^2\\
\leq&(1+\frac{1}{c_1})\frac{\gamma}{m}\Big(\sum_{t=1}^{T}4L_f^2\mathbb{E}\|\boldsymbol{\theta}_{t-1}-\boldsymbol{\theta}_t\|^2+4L_f^2\mathbb{E}\|\boldsymbol{\omega}_{t-1}-\boldsymbol{\omega}_t\|^2+\sum_{t=0}^{T}6\mathbb{E}\|\mathbf{v}_{(n_t-1)q}-\nabla_{\boldsymbol{\theta}}F_{(n_t-1)q}\|^2\Big)\\
&+(1+\frac{1}{c_2})\frac{\eta}{m}\Big(\sum_{t=1}^{T}4L_f^2\mathbb{E}\|\boldsymbol{\theta}_{t-1}-\boldsymbol{\theta}_t\|^2+4L_f^2\mathbb{E}\|\boldsymbol{\omega}_{t-1}-\boldsymbol{\omega}_t\|^2+\sum_{t=0}^{T}6\mathbb{E}\|\mathbf{u}_{(n_t-1)q}-\nabla_{\boldsymbol{\omega}}F_{(n_t-1)q}\|^2\Big)\\
\leq&(1+\frac{1}{c_1})\frac{6\gamma}{m}\sum_{t=0}^{T}6\mathbb{E}\|\mathbf{v}_{(n_t-1)q}-\nabla_{\boldsymbol{\theta}}F_{(n_t-1)q}\|^2+(1+\frac{1}{c_2})\frac{6\eta}{m}\sum_{t=0}^{T}\mathbb{E}\|\mathbf{u}_{(n_t-1)q}-\nabla_{\boldsymbol{\omega}}F_{(n_t-1)q}\|^2\\
&+\frac{4L_f^2}{m}\Big((1+\frac{1}{c_1})\gamma+(1+\frac{1}{c_2})\eta\Big)\sum_{t=1}^{T}\mathbb{E}\big(\|\boldsymbol{\theta}_{t-1}-\boldsymbol{\theta}_t\|^2+\|\boldsymbol{\omega}_{t-1}-\boldsymbol{\omega}_t\|^2\big).
\end{aligned}
\tag{79}
$$

Thus, we have:

$$
\begin{aligned}
R_1+R_2\leq&\frac{4L_f^2}{m}\Big((1+\frac{1}{c_1})\gamma+(1+\frac{1}{c_2})\eta+\frac{\gamma}{2}+\frac{18L_f^2\gamma}{\mu^2}\Big)\sum_{t=1}^{T}\mathbb{E}\|\boldsymbol{\theta}_{t-1}-\boldsymbol{\theta}_t\|^2\\
&+\frac{4L_f^2}{m}\Big((1+\frac{1}{c_1})\gamma+(1+\frac{1}{c_2})\eta+\frac{\gamma}{2}+\frac{18L_f^2\gamma}{\mu^2}\Big)\sum_{t=1}^{T}\mathbb{E}\|\boldsymbol{\omega}_{t-1}-\boldsymbol{\omega}_t\|^2\\
&+\Big((1+\frac{1}{c_1})\frac{6\gamma}{m}+\frac{72\gamma}{m}\Big)\sum_{t=0}^{T}\mathbb{E}\|\mathbf{v}_{(n_t-1)q}-\nabla_{\boldsymbol{\theta}}F_{(n_t-1)q}\|^2\\
&+\Big((1+\frac{1}{c_2})\frac{6\eta}{m}+\frac{72L_f^2\gamma}{m\mu^2}\Big)\sum_{t=0}^{T}\mathbb{E}\|\mathbf{u}_{(n_t-1)q}-\nabla_{\boldsymbol{\omega}}F_{(n_t-1)q}\|^2\\
\leq&\frac{4L_f^2}{m}\Big((1+\frac{1}{c_1})\gamma+(1+\frac{1}{c_2})\eta+\frac{\gamma}{2}+\frac{18L_f^2\gamma}{\mu^2}\Big)\sum_{t=1}^{T}\mathbb{E}\Big(8\mathcal{E}(\boldsymbol{\theta}_{t-1})+4\gamma^2\mathcal{E}(\mathbf{p}_{t-1})+4\gamma^2m\|\bar{\mathbf{p}}_{t-1}\|^2\Big)\\
&+\frac{4L_f^2}{m}\Big((1+\frac{1}{c_1})\gamma+(1+\frac{1}{c_2})\eta+\frac{\gamma}{2}+\frac{18L_f^2\gamma}{\mu^2}\Big)\sum_{t=1}^{T}\mathbb{E}\Big(8\mathcal{E}(\boldsymbol{\omega}_{t-1})+4\eta^2\mathcal{E}(\mathbf{d}_{t-1})+4\eta^2m\|\bar{\mathbf{d}}_{t-1}\|^2\Big)\\
&+\Big((1+\frac{1}{c_1})\frac{6\gamma}{m}+\frac{2\gamma}{m}\Big)\sum_{t=0}^{T}\mathbb{E}\|\mathbf{v}_{(n_t-1)q}-\nabla_{\boldsymbol{\theta}}F_{(n_t-1)q}\|^2\\
&+\Big((1+\frac{1}{c_2})\frac{6\eta}{m}+\frac{72L_f^2\gamma}{m\mu^2}\Big)\sum_{t=0}^{T}\mathbb{E}\|\mathbf{u}_{(n_t-1)q}-\nabla_{\boldsymbol{\omega}}F_{(n_t-1)q}\|^2.
\end{aligned}
\tag{80}
$$

Plugging the above results, we have

$$\begin{aligned}
\mathbb{E}\mathfrak{p}_{T+1} - \mathfrak{p}_0 \leq & -\frac{\gamma}{2}\sum_{t=0}^{T}\mathbb{E}\|\nabla J(\bar{\boldsymbol{\theta}}_t)\|^2 - \gamma L_f^2\sum_{t=0}^{T}\mathbb{E}\|\boldsymbol{\omega}_t^* - \bar{\boldsymbol{\omega}}_t\|^2 \\
& - c_{\bar{\mathbf{p}}}\sum_{t=0}^{T}\gamma\mathbb{E}\|\bar{\mathbf{p}}_t\|^2 - c_{\bar{\mathbf{d}}}\sum_{t=0}^{T}\gamma\eta\mathbb{E}\|\bar{\mathbf{d}}_t\|^2 - c_{\boldsymbol{\theta}}\sum_{t=0}^{T}\frac{\mathcal{E}(\boldsymbol{\theta}_t)}{m} \\
& - c_{\boldsymbol{\omega}}\sum_{t=0}^{T}\frac{\mathcal{E}(\boldsymbol{\omega}_t)}{m} - c_{\mathbf{p}}\sum_{t=0}^{T}\frac{\gamma\mathcal{E}(\mathbf{p}_t)}{m} - c_{\mathbf{d}}\sum_{t=0}^{T}\frac{\eta\mathcal{E}(\mathbf{d}_t)}{m} \\
& + \Big((1+\frac{1}{c_1})\frac{6\gamma}{m} + \frac{2\gamma}{m}\Big)\sum_{t=0}^{T}\mathbb{E}\|\mathbf{v}_{(n_t-1)q} - \nabla_{\boldsymbol{\theta}}F_{(n_t-1)q}\|^2 \\
& + \Big((1+\frac{1}{c_2})\frac{6\eta}{m} + \frac{72L_f^2\gamma}{m\mu^2}\Big)\sum_{t=0}^{T}\mathbb{E}\|\mathbf{u}_{(n_t-1)q} - \nabla_{\boldsymbol{\omega}}F_{(n_t-1)q}\|^2, \qquad (81)
\end{aligned}$$

where

$$c_{\bar{\mathbf{p}}} = \frac{1}{2} - \frac{L_J\gamma}{2} - \frac{40\gamma^2L_{\boldsymbol{\omega}}^2L_f^2}{\mu^2\eta^2} - 16L_f^2\gamma^2\Big((1+\frac{1}{c_1}) + (1+\frac{1}{c_2})\frac{\eta}{\gamma} + \frac{1}{2} + \frac{18L_f^2}{\mu^2}\Big), \qquad (82)$$

$$c_{\bar{\mathbf{d}}} = \frac{2L_f^2}{\mu} - 16L_f^2\eta\Big((1+\frac{1}{c_1}) + (1+\frac{1}{c_2})\frac{\eta}{\gamma} + \frac{1}{2} + \frac{18L_f^2}{\mu^2}\Big), \qquad (83)$$

$$c_{\boldsymbol{\theta}} = 1 - (1+c_1)\lambda^2 - \frac{72\gamma L_f^4}{\mu^2} - 2\gamma L_f^2 - 32L_f^2\gamma\Big((1+\frac{1}{c_1}) + (1+\frac{1}{c_2})\frac{\eta}{\gamma} + \frac{1}{2} + \frac{18L_f^2}{\mu^2}\Big), \qquad (84)$$

$$c_{\boldsymbol{\omega}} = 1 - (1+c_2)\lambda^2 - \frac{72\gamma L_f^4}{\mu^2} - 2\gamma L_f^2 - 32L_f^2\gamma\Big((1+\frac{1}{c_1}) + (1+\frac{1}{c_2})\frac{\eta}{\gamma} + \frac{1}{2} + \frac{18L_f^2}{\mu^2}\Big), \qquad (85)$$

$$c_{\mathbf{p}} = 1 - (1+c_1)\lambda^2 - (1+\frac{1}{c_1})\gamma - 16L_f^2\gamma^2\big((1+\frac{1}{c_1}) + (1+\frac{1}{c_2})\frac{\eta}{\gamma} + \frac{1}{2} + \frac{18L_f^2}{\mu^2}\big), \qquad (86)$$

$$c_{\mathbf{d}} = 1 - (1+c_2)\lambda^2 - (1+\frac{1}{c_2})\eta - 16L_f^2\eta\gamma\Big((1+\frac{1}{c_1}) + (1+\frac{1}{c_2})\frac{\eta}{\gamma} + \frac{1}{2} + \frac{18L_f^2}{\mu^2}\Big). \qquad (87)$$

Choose $c_1 = c_2 = 1/\lambda - 1$, and define $C_0 = \frac{1}{1-\lambda}(1+\frac{\eta}{\gamma}) + \frac{1}{2} + \frac{18L_f^2}{\mu^2}$. It follows that

$$c_{\bar{\mathbf{p}}} = \frac{1}{2} - \frac{L_J\gamma}{2} - \frac{40\gamma^2L_{\boldsymbol{\omega}}^2L_f^2}{\mu^2\eta^2} - 16C_0L_f^2\gamma^2, \qquad (88)$$

$$c_{\bar{\mathbf{d}}} = \frac{2L_f^2}{\mu} - 16C_0L_f^2\eta, \qquad (89)$$

$$c_{\boldsymbol{\theta}} = c_{\boldsymbol{\omega}} = 1 - \lambda - \frac{72\gamma L_f^4}{\mu^2} - 2\gamma L_f^2 - 32C_0L_f^2\gamma, \qquad (90)$$

$$c_{\mathbf{p}} = 1 - \lambda - \frac{\gamma}{1-\lambda} - 16C_0L_f^2\gamma^2, \qquad (91)$$

$$c_{\mathbf{d}} = 1 - \lambda - \frac{\eta}{1-\lambda} - 16C_0L_f^2\eta\gamma. \qquad (92)$$

To ensure $c_{\bar{\mathbf{p}}} \geq 0$, we have

$$\begin{aligned} c_{\bar{\mathbf{p}}} &\overset{(a)}{\geq} \frac{1}{4} - \frac{L_J\gamma}{2} - 16C_0L_f^2\gamma^2 \\ &\overset{(b)}{\geq} \frac{1}{4} - \frac{(L_f + L_f^2/\mu)\gamma}{2} - \frac{(1-\lambda)\gamma}{2} \overset{(c)}{\geq} 0, \end{aligned} \tag{93}$$

where (a) follows from $\kappa := \gamma/\eta \leq \mu^2/13L_f^2$ and Lemma 9, (b) is due to $\gamma \leq (1-\lambda)/32C_0L_f^2$ and Lemma 11, and (c) is from $\gamma \leq 1/2\big((L_f + L_f^2/\mu) + (1-\lambda)\big)$. By setting $\eta \leq 1/8\mu C_0$, we have $c_{\bar{\mathbf{d}}} \geq 0$. By setting $\gamma \leq (1-\lambda)/(\frac{1}{2} + \frac{72L_f^4}{\mu^2} + 2L_f^2 + 32C_0L_f^2)$, we have $c_{\boldsymbol{\theta}} = c_{\boldsymbol{\omega}} \geq \gamma/2$. To ensure $c_{\mathbf{p}} \geq 0$,

$$\begin{aligned} c_{\mathbf{p}} &= 1 - \lambda - \frac{\gamma}{1-\lambda} - 16C_0L_f^2\gamma^2 \\ &\overset{(a)}{\geq} 1 - \lambda - \frac{\gamma}{1-\lambda} - \frac{(1-\lambda)\gamma}{2} \overset{(b)}{\geq} 0, \end{aligned} \tag{94}$$

where (a) follows from $\gamma \leq (1-\lambda)/32C_0L_f^2$ and (b) is due to $\gamma \leq 1/\big(1/2 + 1/(1-\lambda)^2\big)$. Similarly, with $\eta \leq 1/\big(1/2 + 1/(1-\lambda)^2\big)$, we have $c_{\mathbf{d}} \geq 0$.

To summarize, we need the following conditions to ensure $c_{\bar{\mathbf{p}}} \geq 0$, $C_{\bar{\mathbf{d}}} \geq 0$, $C_{\mathbf{p}} \geq 0$, $C_{\mathbf{d}} \geq 0$, $C_{\boldsymbol{\theta}} \geq \gamma/2$, $c_{\boldsymbol{\omega}} \geq \gamma/2$,

$$\kappa = \gamma/\eta \leq \mu^2/13L_f^2, \tag{95}$$

$$\gamma \leq \min\Big\{\frac{1}{2}\Big(L_f + \frac{L_f^2}{\mu} + (1-\lambda)\Big), \frac{1-\lambda}{(\frac{1}{2} + \frac{72L_f^4}{\mu^2} + 2L_f^2 + 32C_0L_f^2)}, \big(1/2 + 1/(1-\lambda)^2\big)^{-1}\Big\}, \tag{96}$$

$$\eta \leq \min\Big\{\frac{1}{8\mu C_0}, \big(1/2 + 1/(1-\lambda)^2\big)^{-1}\Big\}, \tag{97}$$

which can be satisfied by

$$\kappa = \gamma/\eta \leq \mu^2/13L_f^2, \tag{98}$$

$$\eta \leq \min\Big\{\frac{13L_f^2}{2\mu^2}\Big(L_f + \frac{L_f^2}{\mu} + (1-\lambda)\Big), \frac{26(1-\lambda)L_f^2}{(\mu^2 + 144L_f^4 + 4L_f^2\mu^2 + 64C_0L_f^2\mu^2)}, \frac{1}{8\mu C_0}, \frac{13L_f^2}{\mu^2\big(1/2 + 1/(1-\lambda)^2\big)}\Big\}, \tag{99}$$

where the constant C_0 is defined as $C_0 = \frac{1}{1-\lambda}(1 + \frac{1}{\kappa}) + \frac{1}{2} + \frac{18L_f^2}{\mu^2}$. Also, it can be easily verified that $\frac{1}{8\mu C_0} \leq \frac{1}{2L_f}$.

With the above conditions, we have:

$$\begin{aligned} &\frac{\gamma}{2}\sum_{t=0}^{T}\mathbb{E}\|\nabla J(\bar{\boldsymbol{\theta}}_t)\|^2 + 2L_f^2\mathbb{E}\|\boldsymbol{\omega}_t^* - \bar{\boldsymbol{\omega}}_t\|^2 + \frac{\mathcal{E}(\boldsymbol{\theta}_t)}{m} + \frac{\mathcal{E}(\boldsymbol{\omega}_t)}{m} \\ \leq &\mathbb{E}[\mathfrak{p}_0 - \mathfrak{p}_{T+1}] + \Big(\frac{6\gamma}{m\lambda} + \frac{2\gamma}{m}\Big)\sum_{t=0}^{T}\mathbb{E}\|\mathbf{v}_{(n_t-1)q} - \nabla_{\boldsymbol{\theta}}F_{(n_t-1)q}\|^2 \\ &+ \Big(\frac{6\eta}{m\lambda} + \frac{72L_f^2\gamma}{m\mu^2}\Big)\sum_{t=0}^{T}\mathbb{E}\|\mathbf{u}_{(n_t-1)q} - \nabla_{\boldsymbol{\omega}}F_{(n_t-1)q}\|^2. \end{aligned} \tag{100}$$

For GT-SRVR, the outer loop calculates the full gradients. Thus, we have $\mathbb{E}\|\mathbf{v}_{(n_t-1)q} - \nabla_{\boldsymbol{\theta}} F_{(n_t-1)q}\|^2 = \mathbb{E}\|\mathbf{u}_{(n_t-1)q} - \nabla_{\boldsymbol{\omega}} F_{(n_t-1)q}\|^2 = 0$. Then, we have the stated result in Theorem 3:

$$\frac{1}{T+1}\sum_{t=0}^{T}\mathbb{E}\|\nabla J(\bar{\boldsymbol{\theta}}_t)\|^2 + 2L_f^2\mathbb{E}\|\boldsymbol{\omega}_t^* - \bar{\boldsymbol{\omega}}_t\|^2 + \frac{\mathcal{E}(\boldsymbol{\theta}_t)}{m} + \frac{\mathcal{E}(\boldsymbol{\omega}_t)}{m} \leq \frac{2\mathbb{E}[\mathfrak{p}_0 - \mathfrak{p}_{T+1}]}{(T+1)\gamma}. \tag{101}$$

For GT-SRVRI, we have that

$$\begin{aligned}
&\sum_{t=0}^{T}\mathbb{E}\|\mathbf{v}_{(n_t-1)q} - \nabla_{\boldsymbol{\theta}} F_{(n_t-1)q}\|^2 = \sum_{t=0}^{T}\mathbb{E}\|\mathbf{u}_{(n_t-1)q} - \nabla_{\boldsymbol{\omega}} F_{(n_t-1)q}\|^2 \\
&= \sum_{t=0}^{T}\frac{m\sigma^2\mathbb{1}_{(|\mathcal{R}_{i,(n_t-1)q}|)}}{|\mathcal{R}_{i,(n_t-1)q}|} \leq \sum_{t=0}^{T}\frac{m\sigma^2\mathbb{1}_{(|\mathcal{R}_{i,\lfloor t/q\rfloor q}|<n)}}{\min\{(\lfloor t/q\rfloor+1)^\alpha q, \lceil c_\epsilon\epsilon^{-2}\rceil\}} \\
&\leq \sum_{t=0}^{T} m\sigma^2\max\{\frac{\mathbb{1}_{((\lfloor t/q\rfloor+1)^\alpha q<n)}}{(\lfloor t/q\rfloor+1)^\alpha q}, \frac{\epsilon^2}{c_\epsilon}\} \leq m\sigma^2(\sum_{r=1}^{\infty}\frac{\mathbb{1}_{(\tau^\alpha q<n)}}{\tau^\alpha} + \frac{\epsilon^2(T+1)}{c_\epsilon}) \\
&\leq m\sigma^2(1 + \int_1^{\frac{n}{q}^{1/\alpha}}\frac{1}{\tau^\alpha}d\tau + \frac{\epsilon^2(T+1)}{c_\epsilon}) \\
&\leq \begin{cases} m\sigma^2(\frac{1}{1-\alpha}(\frac{n}{q})^{(\frac{1}{\alpha}-1)} - \frac{\alpha}{1-\alpha} + \frac{\epsilon^2(T+1)}{c_\epsilon}), & \text{if } \alpha>0 \text{ and } \alpha\neq 0, \\ m\sigma^2(1+\log(\frac{n}{q}) + \frac{\epsilon^2(T+1)}{c_\epsilon}), & \text{if } \alpha = 1. \end{cases}
\end{aligned} \tag{102}$$

Thus, for GT-SRVRI, we have the following convergence results:

$$\begin{aligned}
&\frac{1}{T+1}\sum_{t=0}^{T}\mathbb{E}\|\nabla J(\bar{\boldsymbol{\theta}}_t)\|^2 + 2L_f^2\mathbb{E}\|\boldsymbol{\omega}_t^* - \bar{\boldsymbol{\omega}}_t\|^2 + \frac{\mathcal{E}(\boldsymbol{\theta}_t)}{m} + \frac{\mathcal{E}(\boldsymbol{\omega}_t)}{m} \\
\leq& \frac{2\mathbb{E}[\mathfrak{p}_0 - \mathfrak{p}_{T+1}]}{(T+1)\gamma} + (\frac{12}{\lambda}(1+\frac{1}{r}) + 4 + \frac{144L_f^2}{\mu^2})(\frac{\epsilon^2}{c_\epsilon} + \frac{C(n,q,\alpha)}{T+1})\sigma^2,
\end{aligned} \tag{103}$$

where the constant $C(n,q,\alpha)$ is defined as

$$C(n,q,\alpha) \triangleq \begin{cases} \frac{1}{1-\alpha}(\frac{n}{q})^{(\frac{1}{\alpha}-1)} - \frac{\alpha}{1-\alpha}, & \text{if } \alpha>0 \text{ and } \alpha\neq 1 \\ \log(\frac{n}{q}) + 1, & \text{if } \alpha = 1. \end{cases} \tag{104}$$

D Supporting Lemmas

Lemma 9. *Under Assumption 1, $\boldsymbol{\omega}^*(\boldsymbol{\theta}) = \arg\max_{\boldsymbol{\omega}} F(\boldsymbol{\theta},\boldsymbol{\omega})$ is Lipschitz continuous, i.e., there exists a positive constant $L_{\boldsymbol{\omega}}$, such that*

$$\|\boldsymbol{\omega}^*(\boldsymbol{\theta}) - \boldsymbol{\omega}^*(\boldsymbol{\theta}')\| \leq L_{\boldsymbol{\omega}}\|\boldsymbol{\theta} - \boldsymbol{\theta}'\|, \ \ \forall \boldsymbol{\theta},\boldsymbol{\theta}' \in \mathbb{R}^d, \tag{105}$$

where the Lipschitz constant is $L_{\boldsymbol{\omega}} = L_F/\mu$ for Algorithm 1 and $L_{\boldsymbol{\omega}} = L_f/\mu$ for Algorithm 2.

Proof. See Lemma 4.3 in [32]. □

Lemma 10. *Under Assumption 1, the function $J(\boldsymbol{\theta}) = F(\boldsymbol{\theta}, \boldsymbol{\omega}^*(\boldsymbol{\theta}))$ satisfies that $\nabla J(\boldsymbol{\theta}) = \nabla_{\boldsymbol{\theta}} F(\boldsymbol{\theta}, \boldsymbol{\omega}^*(\boldsymbol{\theta}))$.*

Proof. Since $J(\boldsymbol{\theta}) = F(\boldsymbol{\theta}, \boldsymbol{\omega}^*(\boldsymbol{\theta}))$,by chain rule, we have

$$dJ(\boldsymbol{\theta}) = \frac{\partial F(\boldsymbol{\theta}, \boldsymbol{\omega})}{\partial \boldsymbol{\theta}}\Big|_{\boldsymbol{\omega}=\boldsymbol{\omega}^*(\boldsymbol{\theta})} \cdot d\boldsymbol{\theta} + \frac{\partial F(\boldsymbol{\theta}, \boldsymbol{\omega})}{\partial \boldsymbol{\omega}}\Big|_{\boldsymbol{\omega}=\boldsymbol{\omega}^*(\boldsymbol{\theta})} \cdot \frac{\partial \omega^*(\boldsymbol{\theta})}{\partial \boldsymbol{\theta}} \cdot d\boldsymbol{\theta}, \tag{106}$$

where $\partial F(\boldsymbol{\theta}, \boldsymbol{\omega})/\partial \boldsymbol{\theta}$ and $\partial F(\boldsymbol{\theta}, \boldsymbol{\omega})/\partial \boldsymbol{\omega}$ are respectively the partial differential of F w.r.t the first variate $\boldsymbol{\theta}$ and the second variate $\boldsymbol{\omega}$. Note that $\boldsymbol{\omega}^*(\boldsymbol{\theta})$ is the unique optimal point such that $F(\boldsymbol{\theta}, \boldsymbol{\omega})$ reaches the maximums. So, it follows that $\frac{\partial F(\boldsymbol{\theta}, \boldsymbol{\omega})}{\partial \boldsymbol{\omega}}|_{\boldsymbol{\omega}=\boldsymbol{\omega}^*(\boldsymbol{\theta})} = 0$ for all $\boldsymbol{\theta}$. Also, from Lemma 9, we have $\partial \omega^*(\boldsymbol{\theta})/\partial \boldsymbol{\theta}$ is bounded. Thus, it follows that

$$dJ(\boldsymbol{\theta}) = \frac{\partial F(\boldsymbol{\theta}, \boldsymbol{\omega})}{\partial \boldsymbol{\theta}}\Big|_{\boldsymbol{\omega}=\boldsymbol{\omega}^*(\boldsymbol{\theta})} \cdot d\boldsymbol{\theta}, \tag{107}$$

which is $\nabla J(\boldsymbol{\theta}) = \nabla_{\boldsymbol{\theta}} F(\boldsymbol{\theta}, \boldsymbol{\omega}^*(\boldsymbol{\theta}))$.

Additionally, we can follow the detailed derivation of Theorem D2 in [40] to give a more rigorous proof for Eq (107). The difference between our Lemma 10 and Theorem D2 in [40] is that [40] adopted the uniformly bounded gradient assumption (Hypotheses D2.2 in [40]). In the followings, we show that under our assumptions, Eq (107) can be still obtained with similar derivation in [40]. Here we will adopt the notations in [40]. Because of the strong concavity, there is only one set $\{\boldsymbol{\nu}_n\}$ in $\mathcal{W}(\mathbf{u})$ for all $\mathbf{u}$ and $\boldsymbol{\nu}^* = \lim_{n\to} \boldsymbol{\nu}_n$ exists. From Assumption 1 (e), $D_1 J(\mathbf{u}, \boldsymbol{\nu}; h)$ is bounded. Thus, due to Assumption 1 (b), there exists a N_1, with $n \geq N_1$, $D_1 J(\mathbf{u}, \boldsymbol{\nu}_n; h)$ is also bounded. So, Proposition 2 in [40] holds. Meanwhile, because of the continuity of $D_1 J(\mathbf{u}, \boldsymbol{\nu})$ and $D_1 J(\mathbf{u}, \boldsymbol{\nu}^*) < \infty$, there exists N_2 with $n \geq \max\{N_1, N_2\}$, the function $t \mapsto J(\mathbf{u} + th, \boldsymbol{\nu}_n)$ has a bounded directional derivative for all $t \in [0, t_n]$. Thus, Proposition 3 in [40] holds. Because Proposition 2 and Proposition 3 in [40] still holds under our assumptions, we can then reach Eq (107) in Lemma 10. □

Lemma 11. *Under Assumption 1, the funciton $J(\boldsymbol{\theta}) = F(\boldsymbol{\theta}, \omega^*(\boldsymbol{\theta}))$ w.r.t $\boldsymbol{\theta}$ is Lipschitz smooth, i.e., there exists a positive constant L_J, such that*

$$\|\nabla J(\boldsymbol{\theta}) - \nabla J(\boldsymbol{\theta}')\| \leq L_J \|\boldsymbol{\theta} - \boldsymbol{\theta}'\|, \ \ \forall \boldsymbol{\theta}, \boldsymbol{\theta}' \in \mathbb{R}^d, \tag{108}$$

where the Lipschitz constant is $L_J = L_F + L_F^2/\mu$ for Algorithm 1 and $L_J = L_f + L_f^2/\mu$ for Algorithm 2.

Proof. The lemma follows immediately from Lemma 4.3 in [32] and Lemma 10. □