Personalized Federated Reinforcement Learning with Shared Representations

Guojun Xiong
guojun.xiong@stonybrook.edu
Department of Applied Mathematical & Statistics
Department of Computer Science
Stony Brook University

Shufan Wang
shufan.wang@stonybrook.edu
Department of Applied Mathematical & Statistics
Department of Computer Science
Stony Brook Univerisity

Daniel Jiang
danielrjiang@gmail.com
Applied Reinforcement Learning Team, AI at Meta
University of Pittsburgh

Jian Li
jian.li.3@stonybrook.edu
Department of Applied Mathematical & Statistics
Stony Brook University

Abstract

Federated reinforcement learning (FedRL) enables multiple agents to collaboratively learn a policy without sharing their own local trajectories collected during agent-environment interactions. However, in practice, the environments faced by different agents are often heterogeneous, leading to poor performance by the single policy learned by existing FedRL algorithms on individual agents. In this paper, we take a further step and introduce a *personalized* FedRL framework (PFedRL) by taking advantage of possibly shared common structure among agents in heterogeneous environments. Specifically, we develop a class of PFedRL algorithms named PFedRL-Rep that learns (1) a shared feature representation collaboratively among all agents and (2) an agent-specific weight vector personalized to its local environment. We analyze the convergence of PFedTD-Rep, a particular instance of the framework with temporal difference (TD) learning and linear representations. To the best of our knowledge, we are the first to prove a linear convergence speedup with respect to the number of agents in the PFedRL setting. To achieve this, we show that PFedTD-Rep is an example of the federated two-timescale stochastic approximation with Markovian noise. Experimental results demonstrate that PFedTD-Rep, along with an extension to the control setting based on deep Q-networks (DQN), not only improve learning in heterogeneous settings, but also provide better generalization to new environments.

1 Introduction

Federated reinforcement learning (FedRL) Nadiger et al. (2019); Liu et al. (2019); Xu et al. (2021); Zhang et al. (2022a); Jin et al. (2022); Khodadadian et al. (2022); Yuan et al. (2023) has recently emerged as a promising method via blending the distributed nature of federated learning (FL) McMahan et al. (2017) with the sequential decision-making nature of reinforcement learning (RL) Sutton & Barto (2018). In FedRL, multiple agents collaboratively learn *a single policy* without sharing their individual trajectories that are collected during the agent-environment interactions, which protect privacy embedded in their local experiences.

One key challenge of FedRL is the issue of environment heterogeneity among agents, where the collected trajectories among agents may vary to a large extent. To illustrate, consider a few existing applications of FL. These include on-device NLP applications (e.g., next word prediction, web query suggestions, and speech recognition) from internet companies (Hard et al., 2018; Yang et al., 2018;

Table 1: Comparison of settings with existing FedRL frameworks.

Algorithm	Noise	Environment	Representation	Timescale	Local update	Personalization	Linear speedup
FedTD & FedQ Khodadadian et al. (2022)	*Markovian*	*Homogeneous*	✗	*Single*	✓	✗	✓
FedTD Dal Fabbro et al. (2023)	*Markovian*	*Homogeneous*	✗	*Single*	✗	✗	✓
FedTD Wang et al. (2023a)	*Markovian*	*Heterogeneous*	✗	*Single*	✓	✗	✓
QAvg & PAvg Jin et al. (2022)	*i.i.d.*	*Heterogeneous*	✗	*Single*	✗	✗	✗
FedQ Woo et al. (2023)	*Markovian*	*Heterogeneous*	✗	*Single*	✗	✗	✓
A3C Shen et al. (2023)	*Markovian*	*Homogeneous*	✗	*Two*	✗	✗	✓
FedSARSA Zhang et al. (2024)	*Markovian*	*Heterogeneous*	✗	*Single*	✓	✗	✓
PFedRL-Rep (this work)	***Markovian***	***Heterogeneous***	✓	***Two***	✓	✓	✓

Wang et al., 2023b), on-device recommender or ad prediction systems (Maeng et al., 2022; Krichene et al., 2023), and Internet of Things applications like smart healthcare or smart thermostats (Nguyen et al., 2021; Imteaj et al., 2022; Zhang et al., 2022b; Boubouh et al., 2023). Note that *all of them* exist in settings with environment heterogeneity (heterogeneous users, devices, patients, or homes).

As a result, if all agents collaboratively learn a single policy, which most existing FedRL frameworks do, the learned policy might perform poorly on individual agents. This calls for the design of a *personalized* FedRL (PFedRL) framework that can provide personalized policies for agents in different environments. Nevertheless, despite the recent advances in FedRL, the design of PFedRL and its performance analysis remains, to a large extent, an open question. Motivated by this, the first inquiry we aim to answer in this paper is:

> *Can we design a PFedRL framework for agents in heterogenous environments to not only collaboratively learn a useful global model without sharing local trajectories, but also provide a personalized policy in each environment?*

We address this question by viewing the PFedRL problem in heterogeneous environments as N parallel RL tasks with possibly *shared common structure*. This is inspired by observations in centralized learning Bengio et al. (2013); LeCun et al. (2015) and federated/decentralized learning Collins et al. (2021); Xiong et al. (2023); Tziotis et al. (2023), whose success in training multiple tasks simultaneously can be enlarged by leveraging a common (low-dimensional) representation in various machine learning tasks (e.g., image classification).

From a theoretical point of view, questions of the benefits of leveraging such shared representations among heterogeneous agents have received an increased recent emphasis owing to their practical significance, especially in federated/decentralized supervised learning framework Collins et al. (2021); Xiong et al. (2023); Tziotis et al. (2023). However, a theoretical analysis of PFedRL with shared representations is more subtle due to the fact that each agent in PFedRL collects data by following its own Markovian trajectory and simultaneously updates its model parameters, while data is collected before training begins in standard FL paradigm. These considerations motivate the second question we aim to address:

> *How do the shared representations affect the convergence performance of PFedRL under Markovian noise, and is it possible to achieve an N-fold linear convergence speedup?*

Despite some recent progress in federated/decentralized supervised learning framework Collins et al. (2021); Xiong et al. (2023); Tziotis et al. (2023), to the best of our knowledge, this question is still open in the context of learning personalized policies in FedRL under Markovian noise (see Table 1). Motivated by these open questions, we introduce a new PFedRL framework with shared representations, and analyze a class of associated algorithms. Our main contributions are summarized in the following.

• **PFedRL-Rep Algorithm.** We propose PFedRL-Rep, a new PFedRL framework with shared representations by leveraging representation learning theory. PFedRL-Rep learns a global shared feature representation collaboratively among all agents through the aid of a central server, and an agent-specific weight vector that is personalized to its local environment. We note that our PFedRL-

Rep framework can be paired with a wide range of RL algorithms, including both value-function based and policy-gradient based methods with arbitrary feature representation.

• **Linear Speedup for TD Learning.** Within the PFedRL-Rep framework, we further introduce PFedTD-Rep, i.e., the PFedRL-Rep version of TD learning, and analyze its convergence performance in a linear representation setting. We prove that the convergence rate of PFedTD-Rep is $\tilde{\mathcal{O}}\left(\frac{1}{N^{2/3}(T+2)^{2/3}}\right)$, where N is the number of agents and T is the number of communication rounds. This implies a linear convergence speedup for PFedTD-Rep with respect to the number of agents. To our best knowledge, this is the first linear speedup result for PFedRL with shared representations under Markovian noise, and provides a theoretical answer to empirical observations in Mnih et al. (2016) that federated versions of RL algorithms yield faster convergence. This property is highly desirable since it implies that one can efficiently leverage massive parallelism in large-scale systems. To achieve this, we show PFedTD-Rep is an example of the federated two-timescale stochastic approximation, and its convergence analysis is intricate under Markovian noise. We address this challenge by leveraging a Lyapunov drift approach to capture the evolution of two coupled parameters, fundamentally improving upon prior work.

2 Problem Formulation

Notation. Let N and T be the number of agents and communication rounds. Denote $[N]$ as the set of integers $\{1,\ldots,N\}$ and $\|\cdot\|$ as the l_2-norm. We use boldface to denote matrices and vectors.

2.1 Preliminaries: Federated Reinforcement Learning

We consider a FedRL system with N agents interacting with N independent heterogeneous environments. We model the environment of agent i as a MDP, $\mathcal{M}^i = \langle \mathcal{S}, \mathcal{A}, R^i, P^i, \gamma \rangle, \forall i \in [N]$, where $\mathcal{S}$ and $\mathcal{A}$ are finite state and action sets, R^i is the reward function, P^i is the transition kernel, and $\gamma \in (0,1)$ is the discount factor. At each time step k, agent i is in state s_k^i and takes action a_k^i according to a policy $\pi^i(\cdot|s_k^i)$ in hand, which results in reward $R^i(s_k^i, a_k^i)$. In the next time step, the environment transitions to a new state s_{k+1}^i according to the state transition probability $P^i(\cdot|s_k^i, a_k^i)$. The sequence of states and actions constructs a Markov chain, which is the source of Markovian noise. In this paper, this Markov chain is assumed to be unichain, which is known to asymptotically converge to a steady state. We denote the stationary distribution as μ^{i,π^i}.

The state-value function agent i in environment $\mathcal{M}^i$ under policy π^i are defined as $V^{i,\pi^i}(s) = \mathbb{E}_{\pi^i}\left[\sum_{k=0}^{\infty} \gamma^k R^i(s_k^i, a_k^i)|s_0^i = s\right]$. When the state and action spaces are large, it is computationally infeasible to store $V^{i,\pi^i}(s)$ for all states or state-action pairs. One way to deal with is to approximate the value function as $V^{i,\pi^i}(s) \approx \boldsymbol{\Phi}(s)\boldsymbol{\theta}$, where $\boldsymbol{\Phi} \in \mathbb{R}^{|\mathcal{S}|\times d}$ is a feature representation corresponding to states, and $\boldsymbol{\theta} \in \mathbb{R}^d$ is a low-dimensional unknown weight vector. When $\boldsymbol{\Phi}$ is given and known, this falls under the paradigm of RL or FedRL with function approximation.

One intermediate goal in RL is to estimate the value function corresponding to a particular policy π using the trajectory collected from the environment. This task is called *policy evaluation*, and one widely used approach to accomplish this is the Temporal Difference (TD) learning Sutton (1988). Under the FedRL framework, the goal of policy evaluation, or specifically, FedTD Khodadadian et al. (2022); Dal Fabbro et al. (2023); Wang et al. (2023a) is to let N agents collaboratively evaluate a single policy $\pi \equiv \pi^i, \forall i \in [N]$, or precisely, *collaboratively learn a common (non-personalized) weight vector* $\boldsymbol{\theta} \equiv \boldsymbol{\theta}^i, \forall i \in [N]$ *using trajectories collected from* N *different environments when the feature representation* $\boldsymbol{\Phi}(s), \forall s$ *are given.* This can be formulated as the following optimization problem:

$$\mathcal{L}(\boldsymbol{\theta}) := \min_{\boldsymbol{\theta}} \frac{1}{N} \sum_{i=1}^{N} \mathbb{E}_{s\sim\mu^{i,\pi}} \left\|\boldsymbol{\Phi}(s)\boldsymbol{\theta} - V^{i,\pi}(s)\right\|^2. \tag{1}$$

Due to space constraint, we are going to first focus on the policy evaluation problem in RL. Note that policy evaluation is an important part of RL and control, since it is a critical step of policy

Algorithm 1 PFedRL-Rep: A General Description

```
Input: Sampling policy $\pi^i, \forall i$;
Initialize the global feature representation $\mathbf{\Phi}_0$ and local weight vector $\boldsymbol{\theta}_0^i$, $\forall i \in [N]$ randomly;
for round $t = 0, 1, \ldots, T-1$ do
    for agent $1, \ldots, N$ do
        $\boldsymbol{\theta}_{t+1}^i = \text{RL\_ UPDATE}(\mathbf{\Phi}_t, \boldsymbol{\theta}_t^i, \alpha_t, K)$;
        $\mathbf{\Phi}_{t+1/2}^i = \text{RL\_UPDATE}(\mathbf{\Phi}_t, \boldsymbol{\theta}_{t+1}^i, \beta_t)$;
    end for
    Server computes the new global feature representation $\mathbf{\Phi}_{t+1} = \frac{1}{N}\sum_{i=1}^N \mathbf{\Phi}_{t+1/2}^i$.
end for
```

improvement algorithms. However, our proposed framework (e.g., Algorithm 1) can be directly applied to control problem as well, and we relegate the discussions to Section 5 and Appendix C.

2.2 Personalized FedRL with Shared Representations

Since the local environments are heterogeneous across N agents, the aforementioned FedRL methods (in Section 2.1) that aim to learn a single policy or a common weight vector $\boldsymbol{\theta}$ may perform poorly on many individual agents. This necessitates the search for more personalized policies $\{\pi^i, \forall i \in [N]\}$ or personalized local weight vectors $\boldsymbol{\theta}^i$ that can be learned collaboratively among N agents in N heterogeneous environments without sharing their locally collected trajectories. To achieve this, we propose to learn a common representation among agents in heterogeneous environments by viewing the personalized FedRL (PFedRL) problem as N parallel RL tasks with possibly shared common structure. Specifically, the value function of agent i can be approximated as $V^{i,\pi^i} \approx f^i(\boldsymbol{\theta}^i, \mathbf{\Phi})$, where $\mathbf{\Phi}$ is the shared feature representation among all agents, $\boldsymbol{\theta}^i$ is the local unique weight vector, and $f^i(\cdot,\cdot)$ is a general function parameterized by these two *unknown* parameters.

Using these notions, the policy evaluation problem in (1) can be reformulated as:

$$\mathcal{L}(\mathbf{\Phi}, \{\boldsymbol{\theta}^i, \forall i\}) := \min_{\mathbf{\Phi}} \frac{1}{N}\sum_{i=1}^{N} \min_{\{\boldsymbol{\theta}^i, \forall i\}} \mathbb{E}_{s\sim\mu^{i,\pi^i}} \left\| f^i(\boldsymbol{\theta}^i, \mathbf{\Phi}(s)) - V^{i,\pi^i}(s) \right\|^2, \tag{2}$$

where N agents collaboratively learn a shared feature representation $\mathbf{\Phi}$ via a server, and a personalized local weight vector $\{\boldsymbol{\theta}^i, \forall i\}$ using local trajectories at each agent.

Remark 2.1. Our approximation function $f^i(\boldsymbol{\theta}^i, \mathbf{\Phi})$ is general and can take on various forms, such as linear or neural networks. For instance, it can be represented as a linear combination of $\mathbf{\Phi}$ and $\boldsymbol{\theta}^i$, i.e., $f^i(\boldsymbol{\theta}^i, \mathbf{\Phi}) := \mathbf{\Phi}\boldsymbol{\theta}^i$ in TD Bhandari et al. (2018) or Q-learning Chen et al. (2019) with linear function approximation. To further increase the representation capability, $f^i(\boldsymbol{\theta}^i, \mathbf{\Phi})$ can represent a deep neural network, e.g., as for DQN (Q-learning with deep neural networks) Mnih et al. (2015) (see more discussions in Section 5 and Appendix C).

Comparison with Standard FedRL. In FedRL, N agents simultaneously evaluate one policy π over N heterogeneous environments, and the objective in (1) is to collaboratively learn a common weight vector $\boldsymbol{\theta}$ for all agents i. In contrast, consider TD learning with a linear representation under our new PFedRL framework with shared representations. Here, the goal is to collaboratively learn a personalized weight vector $\boldsymbol{\theta}^i$ for each agent i via the optimization problem (2) with $f^i(\boldsymbol{\theta}^i, \mathbf{\Phi}) := \mathbf{\Phi}\boldsymbol{\theta}^i$, leading to a personalized solution for each agent. Since the environments $P^i, \forall i$ are heterogeneous, the learned common weight vector $\boldsymbol{\theta}$ in conventional FedRL is inevitably suboptimal compared with the personalized weight vector $\boldsymbol{\theta}^i$ in environment P^i for agent i.

3 PFedRL-Rep Algorithms

We now propose a class of algorithms called PFedRL-Rep that realize PFedRL with shared representations. Specifically, PFedRL-Rep alternates between three steps among all agents at each

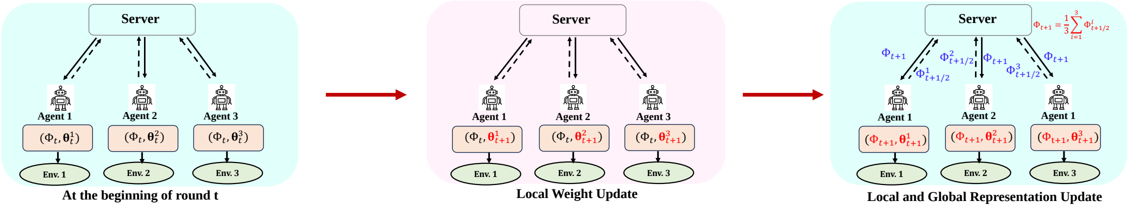


Figure 1: An illustrative example of PFedRL-Rep for 3 agents. (a) At the beginning of round t, each agent $i = 1, 2, 3$ has a local weight vector $\boldsymbol{\theta}_t^i$ and a global feature representation $\boldsymbol{\Phi}_t$. (b) *Local Weight Vector Update:* With $(\boldsymbol{\Phi}_t, \boldsymbol{\theta}_t^i)$, agent i performs a K-step update to obtain $\boldsymbol{\theta}_{t+1}^i$ as in (3). Note that $\boldsymbol{\Phi}_t$ remains unchanged at this step. (c) *Local and Global Feature Representation Update:* Agent i updates the feature representation by executing a one-step update to obtain $\boldsymbol{\Phi}_{t+1/2}^i$ as in (4), which depends on both $\boldsymbol{\theta}_{t+1}^i$ and $\boldsymbol{\Phi}_t$. Then, agent i shares $\boldsymbol{\Phi}_{t+1/2}^i$ with the server, which then executes an averaging step as in (5) to produce the next global feature representation $\boldsymbol{\Phi}_{t+1}$. We highlight the updated parameters in each step in red, and the shared parameters (only the global feature representation) between agents in blue.

communication round: (a) a local weight vector update; (b) a local feature representation update; and (c) a global feature representation update via the server.

Local Weight Vector Update. At round t, agent i performs a RL_UPDATE on its local weight vector given the current global feature representation $\boldsymbol{\Phi}_t$ and local weight vector $\boldsymbol{\theta}_t^i$. In other words, agent i updates its local weight vector as in (3), where RL_UPDATE is a generic notation for an update of the local weight vector $\boldsymbol{\theta}$ using any specific RL algorithm (e.g., TD, Q-learning, DQN, policy gradients, etc, see Appendix D for illustrative examples), and α_t is the learning rate for the local weight vector. To speed up the learning process, we allow each agent to perform K steps local weight vector update based on its local collected trajectory.

$$\boldsymbol{\theta}_{t+1}^i = \text{RL\_UPDATE}(\boldsymbol{\Phi}_t, \boldsymbol{\theta}_t^i, \alpha_t, K), \quad (3)$$

Local Feature Representation Update. Once the updated local weight vector $\boldsymbol{\theta}_{t+1}^i$ is obtained, each agent i executes a one-step local update on their feature representations as in (4), where β_t is the learning rate for the feature representation.

$$\boldsymbol{\Phi}_{t+1/2}^i = \text{RL\_UPDATE}(\boldsymbol{\Phi}_t, \boldsymbol{\theta}_{t+1}^i, \beta_t), \quad (4)$$

Server-based Global Feature Representation Update. The server computes an average of the received local feature representation update $\boldsymbol{\Phi}_{t+1/2}^i$ from all agents $\forall i$ to obtain the next global feature representation $\boldsymbol{\Phi}_{t+1}$ as in (5).

$$\boldsymbol{\Phi}_{t+1} = \frac{1}{N}\sum_{i=1}^{N} \boldsymbol{\Phi}_{t+1/2}^i. \quad (5)$$

PFedRL-Rep alternates between (3), (4) and (5) at each round, and the entire procedure is summarized in Algorithm 1. An example of PFedRL-Rep is illustrated in Figure 1.

Remark 3.1. Similar frameworks by leveraging shared representations have been investigated in federated/decentralized supervised learning (Collins et al., 2021; Xiong et al., 2023; Tziotis et al., 2023). However, in these standard supervised learning frameworks, data is collected before training begins and is often assumed to be i.i.d. In contrast, in our PFedRL framework, each agent collects data by following its own Markovian trajectory, while simultaneously updates its model parameters.

4 PFedTD-Rep with Linear Representation

We present PFedTD-Rep, an instance of PFedRL-Rep paired with TD learning (see its pseudocode in Appendix B), and analyze its convergence performance in a linear representation setting.

4.1 PFedTD-Rep: Algorithm Description

Here, the goal of N agents is to collaboratively solve problem (2) when the underlying RL algorithm is TD learning. To this end, we need to specify the notation of RL_UPDATE in PFedRL-Rep (Algorithm 1) on how to update the local weight vectors $\boldsymbol{\theta}^i$ and global feature representation $\boldsymbol{\Phi}$

using TD. We first consider the notation of RL_UPDATE from the perspective of agent i. At time step k, the state of agent i is s_k^i, and its value function can be denoted as $V(s_k^i) = \boldsymbol{\Phi}(s_k^i)\boldsymbol{\theta}^i$ in a linear representation setting. By one-step Monte Carlo approximation, we approximate $V(s_k^i)$ as $\hat{V}(s_k^i) \approx r_k^i + \gamma\boldsymbol{\Phi}(s_{k+1}^i)\boldsymbol{\theta}^i$. The TD error is defined as

$$\delta_k^i := \hat{V}(s_k^i) - V(s_k^i) = r_k^i + \gamma\,\boldsymbol{\Phi}(s_{k+1}^i)\,\boldsymbol{\theta}^i - \boldsymbol{\Phi}(s_k^i)\,\boldsymbol{\theta}^i. \tag{6}$$

The goal of agent i is to minimize the following loss function for every $s_k^i \in \mathcal{S}$, $\mathcal{L}^i(\boldsymbol{\Phi}(s_k^i), \boldsymbol{\theta}^i) = \frac{1}{2}\left\|V(s_k^i) - \hat{V}(s_k^i)\right\|^2$, with $\hat{V}(s_k^i)$ treated as a constant. We now denote the Markovian observations of agent i at the k-th time step of communication round t as $X_{t,k}^i := (s_{t,k}^i, r_{t,k}^i, s_{t,k+1}^i)$. Note that the observation sequences $\{X_{t,k}^i, \forall t, k\}$ can differ across agents in heterogeneous environments. We assume that $\{X_{t,k}^i, \forall t, k\}$ are statistically independent across all agents.

Local Weight Vector Update: As in Algorithm 1 (i.e., (3)), given the current global feature representation $\boldsymbol{\Phi}_t$, agent i makes K-step local update on its local weight vector $\boldsymbol{\theta}_t^i$ as

$$\boldsymbol{\theta}_{t,k}^i = \boldsymbol{\theta}_{t,k-1}^i + \alpha_t\,\mathbf{g}(\boldsymbol{\theta}_{t,k-1}^i, \boldsymbol{\Phi}_t, X_{t,k-1}^i), \tag{7}$$

for $k \in [K]$, where α_t is the learning rate for the local weight vectors, satisfying $\sum_{t=0}^{\infty}\alpha_t = \infty$ and $\sum_{t=0}^{\infty}\alpha_t^2 < \infty$, and $\mathbf{g}(\boldsymbol{\theta}_{t,k-1}^i, \boldsymbol{\Phi}_t, X_{t,k-1}^i)$ is the negative stochastic gradient of the loss function $\mathcal{L}^i(\boldsymbol{\Phi}_t(s_{t,k-1}^i), \boldsymbol{\theta}_{t,k-1}^i)$ with respect to $\boldsymbol{\theta}$, given the current feature representation $\boldsymbol{\Phi}_t$:

$$\mathbf{g}(\boldsymbol{\theta}_{t,k-1}^i, \boldsymbol{\Phi}_t, X_{t,k-1}^i) := -\nabla_{\boldsymbol{\theta}}\mathcal{L}^i(\boldsymbol{\Phi}_t(s_{t,k-1}^i), \boldsymbol{\theta}_{t,k-1}^i) = \delta_{t,k-1}^i\boldsymbol{\Phi}_t(s_{t,k-1}^i)^\intercal. \tag{8}$$

We allow each agent to perform K-step local updates, and for ease of presentation, we denote $\boldsymbol{\theta}_{t+1}^i := \boldsymbol{\theta}_{t,K}^i$. Since the observations are Markovian, we further add a norm-scaling step for the updated weight vectors $\boldsymbol{\theta}_{t+1}^i$, i.e., enforcing $\|\boldsymbol{\theta}_{t+1}^i\| \leq B$, to stabilize the update. This is essential for finite-time convergence analysis (see Section 4.2), and this technique is widely used in conventional TD learning with linear function approximation Bhandari et al. (2018).

Local Feature Representation Update. As in Algorithm 1 (i.e., (4)), given the updated local weight vector $\boldsymbol{\theta}_{t+1}^i$, agent i executes one-step local update on the global feature representation on its end as

$$\boldsymbol{\Phi}_{t+1/2}^i = \boldsymbol{\Phi}_t + \beta_t\mathbf{h}(\boldsymbol{\theta}_{t+1}^i, \boldsymbol{\Phi}_t, \{X_{t,k-1}^i\}_{k=1}^K), \tag{9}$$

where β_t is the learning rate for global feature representation, satisfying $\sum_{t=0}^{\infty}\beta_t = \infty$, $\sum_{t=0}^{\infty}\beta_t^2 < \infty$, β_t/α_t is non-increasing in t and $\lim_{t\to\infty}\beta_t/\alpha_t = 0$, and $\mathbf{h}(\boldsymbol{\theta}_{t+1}^i, \boldsymbol{\Phi}_t, \{X_{t,k-1}^i\}_{k=1}^K)$ is the set of negative stochastic gradient of the loss function $\mathcal{L}^i(\boldsymbol{\Phi}_t(s_{t,k-1}^i), \boldsymbol{\theta}_{t+1}^i)$ w.r.t. the current global feature representation $\boldsymbol{\Phi}_t$, satisfying

$$\mathbf{h}(\boldsymbol{\theta}_{t+1}^i, \boldsymbol{\Phi}_t, X_{t,k-1}^i) := -\nabla_{\boldsymbol{\Phi}}\mathcal{L}^i(\boldsymbol{\Phi}_t(s_{t,k-1}^i), \boldsymbol{\theta}_{t+1}^i) = \delta_{t,k-1}^i{\boldsymbol{\theta}_{t+1}^i}^\intercal. \tag{10}$$

Server-based Global Feature Representation Update. As in Algorithm 1 (i.e., (5)), the server computes an average of the received local feature representation updates in (9) to obtain the next global feature representation as

$$\boldsymbol{\Phi}_{t+1} = \boldsymbol{\Phi}_t + \beta_t \cdot \frac{1}{N}\sum_{j=1}^{N}\mathbf{h}(\boldsymbol{\theta}_{t+1}^j, \boldsymbol{\Phi}_t, \{X_{t,k-1}^i\}_{k=1}^K). \tag{11}$$

4.2 Convergence Analysis

The coupled updates in (7) and (11) pose a general form as a federated nonlinear two-timescale stochastic approximation (2TSA) Doan (2021) with Markovian noise, with $\boldsymbol{\theta}_t^i$ updating on a faster timescale and $\boldsymbol{\Phi}_t$ on a slower timescale. We aim to establish the finite-time convergence rate of the 2TSA in (7) and (11). This is equivalent to finding a solution pair $(\boldsymbol{\Phi}^*, \{\boldsymbol{\theta}^{i,*}, \forall i\})$ such that[1]

$$\mathbb{E}_{s_t^i\sim\mu^i, s_{t+1}^i\sim P_{\pi^i}^i(\cdot|s_t^i)}[\mathbf{g}(\boldsymbol{\theta}^{i,*}, \boldsymbol{\Phi}^*, X_t^i)] = 0,\ \mathbb{E}_{s_t^i\sim\mu^i, s_{t+1}^i\sim P_{\pi^i}^i(\cdot|s_t^i)}[\mathbf{h}(\boldsymbol{\theta}^{i,*}, \boldsymbol{\Phi}^*, X_t^i)] = 0, \tag{12}$$

[1] The root $(\boldsymbol{\Phi}^*, \{\boldsymbol{\theta}^{i,*}, \forall i\})$ of the nonlinear 2TSA in (7) and (11) can be established by using ODE method following the solution of suitably defined differential equations Doan (2021; 2020); Chen et al. (2019) as in (12).

hold for all Markovian observations X_t^i. In particular, μ^i is a unknown stationary distribution over state s_t^i of agent i at t, and $P_{\pi^i}^i$ is the transition kernel of agent i under policy π^i.

Tsitsiklis & Van Roy (1996) proved that the standard TD iterates converge asymptotically to a vector $\boldsymbol{\theta}^*$ given a fixed feature representation $\mathbf{\Phi}$ almost surely, which is the unique solution of the projected Bellman equation[2] $\Pi_D \mathcal{T}_\mu(\mathbf{\Phi}\boldsymbol{\theta}^*)) = \mathbf{\Phi}\boldsymbol{\theta}^*$. Hence, for agent i, to study the stability of $\boldsymbol{\theta}^i$ when the feature representation $\mathbf{\Phi}$ is fixed, there is a mapping $\boldsymbol{\theta}^i = y^i(\mathbf{\Phi})$ to be the the unique solution to $\mathbb{E}_{s_t^i \sim \mu^i, s_{t+1}^i \sim P_{\pi^i}^i(\cdot|s_t^i)}[\mathbf{g}(\boldsymbol{\theta}^i, \mathbf{\Phi}, X_t^i)] = 0$.

Inspired by Doan (2020), the finite-time analysis of a 2TSA boils down to the choice of two step sizes $\{\alpha_t, \beta_t, \forall t\}$ and a Lyapunov function that couples the two iterates in (7) and (11). Thus, we first define the following two error terms:

$$\tilde{\mathbf{\Phi}}_t = \mathbf{\Phi}_t - \mathbf{\Phi}^*, \qquad \tilde{\boldsymbol{\theta}}_t^i = \boldsymbol{\theta}_t^i - y^i(\mathbf{\Phi}_t),\ \forall i \in [N], \tag{13}$$

which characterize the coupling between $\{\boldsymbol{\theta}_{t+1}^i, \forall i\}$ and $\mathbf{\Phi}_t$. If $\{\tilde{\boldsymbol{\theta}}_{t+1}^i, \forall i\}$ and $\tilde{\mathbf{\Phi}}_t$ go to zero simultaneously, the convergence of $(\{\boldsymbol{\theta}_{t+1}^i, \forall i\}, \mathbf{\Phi}_t)$ to $(\{\boldsymbol{\theta}^{i,*}, \forall i\}, \mathbf{\Phi}^*)$ can be established. Thus, to prove the convergence of $(\{\boldsymbol{\theta}_{t+1}^i, \forall i\}, \mathbf{\Phi}_t)$ of the 2TSA in (7) and (11) to its true value $(\mathbf{\Phi}^*, \{\boldsymbol{\theta}^{i,*}, \forall i\})$, we define the following weighted Lyapunov function to couple the fast and slow iterates

$$M(\{\boldsymbol{\theta}_{t+1}^i, \forall i\}, \mathbf{\Phi}_t) := \|\mathbf{\Phi}_t - \mathbf{\Phi}^*\|^2 + \frac{\beta_{t-1}}{\alpha_t} \cdot \frac{1}{N} \sum_{i=1}^N \|\boldsymbol{\theta}_{t+1}^i - y^i(\mathbf{\Phi}_t)\|^2. \tag{14}$$

To this end, our goal is to characterize the finite-time convergence of $\mathbb{E}[M(\{\boldsymbol{\theta}_{t+1}^i, \forall i\}, \mathbf{\Phi}_t)]$, the Lyapunov function in (14), which is shown in the following theorem.

Theorem 4.1. *For any $T \geq 2\tau_\delta$, and set $\alpha_t = \alpha_0/(t+2)^{5/6}$ and $\beta_t = \beta_0/(t+2)$ with $\beta_0 \leq \omega/2$ and $\alpha_0 \leq \frac{1}{2L\sqrt{2(1+L^2)}}$, where L is a constant, we have*

$$\begin{aligned} M(\{\boldsymbol{\theta}_{T+2}^i\}, \mathbf{\Phi}_{T+1}) \leq \frac{M(\{\boldsymbol{\theta}_1^i\}, \mathbf{\Phi}_0)}{(T+2)^2} &+ C_1(T+2)^{-2/3}\Big(\mathbb{E}[\|\mathbf{\Phi}_0 - \mathbf{\Phi}^*\|^2] + \frac{1}{N}\mathbb{E}\sum_{i=1}^N \|\boldsymbol{\theta}_1^i - y^i(\mathbf{\Phi}_0)\|^2\Big) \\ &+ C_2(T+2)^{-2/3}, \end{aligned} \tag{15}$$

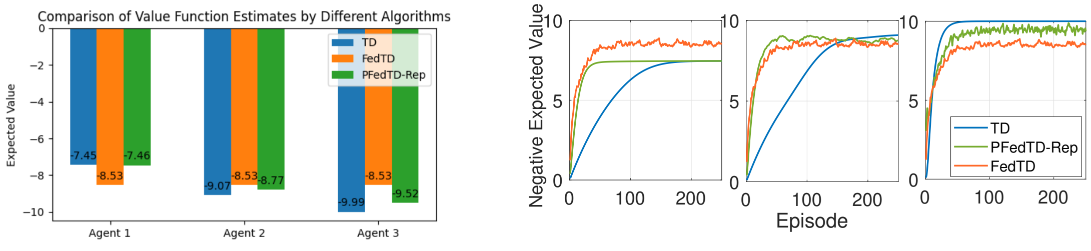
where $C_1 = (144\tau_\delta^2 K^2 L^2 \delta^2 + 4L^2/N)\alpha_0\beta_0$ and $C_2 = (4\alpha_0\beta_0 K^2(3\delta^2(1+B^2) + L^2B^2) + 2\alpha_0^2(3K^2B^2 + 3K^2\delta^2 + 2L^2K^2B^2) + 8\alpha_0\beta_0\delta^2)$.

Remark 4.2. The first term of the right-hand side of (15) corresponds to the bias due to initialization, which goes to zero at a rate $\mathcal{O}(1/T^2)$. The second term corresponds to the accumulated estimation error of the two-timescale update. The third term stands for the variance of Markovian noise. The second and third terms decay at a rate $\mathcal{O}(1/T^{2/3})$, and dominate the overall convergence rate in (15). We leverage a Lyapunov drift approach to capture the evolution of two coupled parameters under Markovian noise, and the characterization of impacts of a norm-scaling step distinguishes our work.

Corollary 4.3. *If $\beta_0 = o(N^{-2/3})$ and $T^2 > N$, we have $M(\{\boldsymbol{\theta}_{t+2}^i\}, \mathbf{\Phi}_{t+1}) \leq \mathcal{O}\Big(\frac{1}{(T+2)^2} + \frac{1}{N^{2/3}(T+2)^{2/3}} + \frac{1}{K^2N^{5/3}(T+2)^{2/3}} + \frac{1}{K^2N^{2/3}(T+2)^{2/3}}\Big)$, which is dominated by $\mathcal{O}\left(\frac{1}{N^{2/3}(T+2)^{2/3}}\right)$.*

Remark 4.4. Corollary 4.3 indicates that to attain an ϵ accuracy, it takes $\mathcal{O}\left(\frac{1}{\epsilon^{3/2}}\right)$ steps with a convergence rate $\mathcal{O}\left(\frac{1}{T^{2/3}}\right)$, while $\mathcal{O}\left(\frac{1}{N\epsilon^{3/2}}\right)$ steps with a convergence rate $\mathcal{O}\left(\frac{1}{N^{2/3}T^{2/3}}\right)$ (the hidden constants in $\mathcal{O}(\cdot)$ are the same). In this sense, we prove that PFedTD-Rep achieves a *linear convergence speedup* w.r.t. the number of agents, i.e., we can proportionally decrease T as N increases while keeping the same convergence rate. To our best knowledge, this is the first linear speedup result for personalized FedRL with shared representations under Markovian noise, and is highly desirable since it implies that one can efficiently leverage the massive parallelism in large-scale systems.

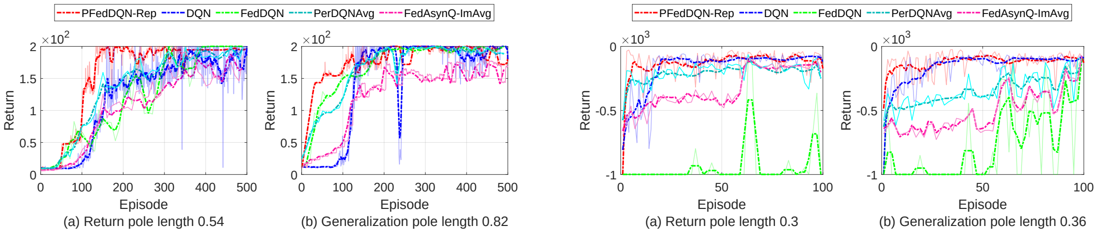
[2] D is a diagonal matrix with entries given by elements of the stationary distribution μ^π of the Markov matrix.



(a) Value function estimates.

(b) Convergence speed of value function learning.

Figure 2: Comparisons in a CliffWalking Environment with 3 agents.



(a) Cartpole environment.

(b) Acrobot environment.

Figure 3: Comparisons in control problems.

4.3 Numerical Evaluation

We empirically evaluate the performance of PFedTD-Rep. In particular, we consider a tabular CliffWalking environment Brockman et al. (2016) with a 4×12 grid world, where 3 agents evaluate 3 different policies. The dimension for the feature representation and weight vector is set to be 6. We compare PFedTD-Rep with (i) "TD": each agent independently leverages the conventional TD without communication; and (ii) "FedTD" without personalization Khodadadian et al. (2022); Dal Fabbro et al. (2023) as listed in Table 1. As shown in Figure 2a, PFedTD-Rep ensures personalization among all agents while FedTD tends to converge uniformly among all agents. Furthermore, PFedTD-Rep attains values much closer to the ground-truth achieved by TD for each agent compared to FedTD; and PFedTD-Rep converges much faster than TD. For instance, agent 1 only needs 50 episodes to converge under PFedTD-Rep, while it takes more than 150 episodes to converge under TD, as illustrated in Figure 2b. The improved convergence performance of PFedTD-Rep further supports our theoretical findings that leveraging shared representations not only provides personalization among agents in heterogeneous environments but yield faster convergence.

5 Application to Control Problems

We evaluate the performance PFedDQN-Rep (see Appendix B) in a modified CartPole environment Brockman et al. (2016). Similar to Jin et al. (2022), we change the length of pole to create different environments. Specifically, we consider 10 agents with varying pole length from 0.38 to 0.74 with a step size of 0.04. We compare PFedDQN-Rep with (i) a conventional DQN that each agent learns its own environment independently; (ii) a federated version DQN (FedDQN) that allows all agents to collaboratively learn a single policy (without personalization), and (iii) two personalized algorithms in state of the arts, i.e., PerDQNAvg Jin et al. (2022) and FedAsynQ-ImAvg Woo et al. (2023). We randomly choose one agent and present its performance in Figure 3a(a). Again, we observe that our PFedDQN-Rep achieves the maximized return much faster than the conventional DQN due to leveraging shared representations among agents; and obtains larger reward than FedDQN, thanks to our personalized policy. We further evaluate the effectiveness of shared representation learned by PFedDQN-Rep when generalizes it to a new agent. As shown in Figure 3a(b), our PFedDQN-Rep generalizes quickly to the new environment. Finally, similar observations can be made from Figure 3b using Acrobot environments (see details in Appendix G).

6 Conclusion and Future Work

In this paper we proposed a novel personalized federated reinforcement learning framework with shared representations. We proved the first linear convergence speedup for PFedTD-Rep, an instance of this framework with TD learning and linear representations. Experimental results demonstrate the superior performance of our proposed framework over existing ones. Several future directions are worth pursuing. First, whether we can provide a finite-time convergence analysis of PFedTD-Rep with neural network feature representation remains an important research direction. Second, giving the promising experimental results on control, whether we can provide a bound for PFedQ-Rep, an instance of our framework with Q-learning, either with linear or neural feature representations remains an open problem.

References

Alekh Agarwal, Sham Kakade, Akshay Krishnamurthy, and Wen Sun. Flambe: Structural complexity and representation learning of low rank mdps. *Advances in neural information processing systems*, 33:20095–20107, 2020.

Manoj Ghuhan Arivazhagan, Vinay Aggarwal, Aaditya Kumar Singh, and Sunav Choudhary. Federated learning with personalization layers. *arXiv preprint arXiv:1912.00818*, 2019.

Yoshua Bengio, Aaron Courville, and Pascal Vincent. Representation learning: A review and new perspectives. *IEEE transactions on pattern analysis and machine intelligence*, 35(8):1798–1828, 2013.

El Houcine Bergou, Konstantin Burlachenko, Aritra Dutta, and Peter Richtárik. Personalized federated learning with communication compression. *arXiv preprint arXiv:2209.05148*, 2022.

Jalaj Bhandari, Daniel Russo, and Raghav Singal. A finite time analysis of temporal difference learning with linear function approximation. In *Conference on learning theory*, pp. 1691–1692. PMLR, 2018.

Vivek S Borkar. *Stochastic approximation: a dynamical systems viewpoint*, volume 48. Springer, 2009.

Karim Boubouh, Robert Basmadjian, Omid Ardakanian, Alexandre Maurer, and Rachid Guerraoui. Efficacy of temporal and spatial abstraction for training accurate machine learning models: A case study in smart thermostats. *Energy and Buildings*, 296:113377, 2023.

Greg Brockman, Vicki Cheung, Ludwig Pettersson, Jonas Schneider, John Schulman, Jie Tang, and Wojciech Zaremba. Openai gym. *arXiv preprint arXiv:1606.01540*, 2016.

Fei Chen, Mi Luo, Zhenhua Dong, Zhenguo Li, and Xiuqiang He. Federated meta-learning with fast convergence and efficient communication. *arXiv preprint arXiv:1802.07876*, 2018.

Zaiwei Chen, Sheng Zhang, Thinh T Doan, Siva Theja Maguluri, and John-Paul Clarke. Performance of q-learning with linear function approximation: Stability and finite-time analysis. *arXiv preprint arXiv:1905.11425*, pp. 4, 2019.

Liam Collins, Hamed Hassani, Aryan Mokhtari, and Sanjay Shakkottai. Exploiting shared representations for personalized federated learning. In *International conference on machine learning*, pp. 2089–2099. PMLR, 2021.

Nicolò Dal Fabbro, Aritra Mitra, and George J Pappas. Federated td learning over finite-rate erasure channels: Linear speedup under markovian sampling. *IEEE Control Systems Letters*, 2023.

Thinh T Doan. Nonlinear two-time-scale stochastic approximation: Convergence and finite-time performance. *arXiv preprint arXiv:2011.01868*, 2020.

Thinh T Doan. Finite-time convergence rates of nonlinear two-time-scale stochastic approximation under markovian noise. *arXiv preprint arXiv:2104.01627*, 2021.

Alireza Fallah, Aryan Mokhtari, and Asuman Ozdaglar. Personalized federated learning: A meta-learning approach. *arXiv preprint arXiv:2002.07948*, 2020.

Andrew Hard, Kanishka Rao, Rajiv Mathews, Swaroop Ramaswamy, Françoise Beaufays, Sean Augenstein, Hubert Eichner, Chloé Kiddon, and Daniel Ramage. Federated learning for mobile keyboard prediction. *arXiv preprint arXiv:1811.03604*, 2018.

Ahmed Imteaj, Khandaker Mamun Ahmed, Urmish Thakker, Shiqiang Wang, Jian Li, and M Hadi Amini. Federated learning for resource-constrained IoT devices: Panoramas and state of the art. *Federated and Transfer Learning*, pp. 7–27, 2022.

Chi Jin, Zhuoran Yang, Zhaoran Wang, and Michael I Jordan. Provably efficient reinforcement learning with linear function approximation. In *Conference on Learning Theory*, pp. 2137–2143. PMLR, 2020.

Hao Jin, Yang Peng, Wenhao Yang, Shusen Wang, and Zhihua Zhang. Federated reinforcement learning with environment heterogeneity. In *International Conference on Artificial Intelligence and Statistics*, pp. 18–37. PMLR, 2022.

Sajad Khodadadian, Pranay Sharma, Gauri Joshi, and Siva Theja Maguluri. Federated reinforcement learning: Linear speedup under markovian sampling. In *International Conference on Machine Learning*, pp. 10997–11057. PMLR, 2022.

Walid Krichene, Nicolas Mayoraz, Steffen Rendle, Shuang Song, Abhradeep Thakurta, and Li Zhang. Private learning with public features. *arXiv preprint arXiv:2310.15454*, 2023.

Yann LeCun, Yoshua Bengio, and Geoffrey Hinton. Deep learning. *nature*, 521(7553):436–444, 2015.

David A Levin and Yuval Peres. *Markov chains and mixing times*, volume 107. American Mathematical Soc., 2017.

Boyi Liu, Lujia Wang, and Ming Liu. Lifelong federated reinforcement learning: a learning architecture for navigation in cloud robotic systems. *IEEE Robotics and Automation Letters*, 4(4):4555–4562, 2019.

Kiwan Maeng, Haiyu Lu, Luca Melis, John Nguyen, Mike Rabbat, and Carole-Jean Wu. Towards fair federated recommendation learning: Characterizing the inter-dependence of system and data heterogeneity. In *Proceedings of the 16th ACM Conference on Recommender Systems*, pp. 156–167, 2022.

Brendan McMahan, Eider Moore, Daniel Ramage, Seth Hampson, and Blaise Aguera y Arcas. Communication-efficient learning of deep networks from decentralized data. In *Artificial intelligence and statistics*, pp. 1273–1282. PMLR, 2017.

Volodymyr Mnih, Koray Kavukcuoglu, David Silver, Andrei A Rusu, Joel Veness, Marc G Bellemare, Alex Graves, Martin Riedmiller, Andreas K Fidjeland, Georg Ostrovski, et al. Human-level control through deep reinforcement learning. *nature*, 518(7540):529–533, 2015.

Volodymyr Mnih, Adria Puigdomenech Badia, Mehdi Mirza, Alex Graves, Timothy Lillicrap, Tim Harley, David Silver, and Koray Kavukcuoglu. Asynchronous methods for deep reinforcement learning. In *International conference on machine learning*, pp. 1928–1937. PMLR, 2016.

Aditya Modi, Jinglin Chen, Akshay Krishnamurthy, Nan Jiang, and Alekh Agarwal. Model-free representation learning and exploration in low-rank mdps. *arXiv preprint arXiv:2102.07035*, 2021.

Chetan Nadiger, Anil Kumar, and Sherine Abdelhak. Federated reinforcement learning for fast personalization. In *2019 IEEE Second International Conference on Artificial Intelligence and Knowledge Engineering (AIKE)*, pp. 123–127. IEEE, 2019.

Dinh C Nguyen, Ming Ding, Pubudu N Pathirana, Aruna Seneviratne, Jun Li, and H Vincent Poor. Federated learning for Internet of Things: A comprehensive survey. *IEEE Communications Surveys & Tutorials*, 23(3):1622–1658, 2021.

Jiaju Qi, Qihao Zhou, Lei Lei, and Kan Zheng. Federated reinforcement learning: Techniques, applications, and open challenges. *arXiv preprint arXiv:2108.11887*, 2021.

John Schulman, Filip Wolski, Prafulla Dhariwal, Alec Radford, and Oleg Klimov. Proximal policy optimization algorithms. *arXiv preprint arXiv:1707.06347*, 2017.

Han Shen, Kaiqing Zhang, Mingyi Hong, and Tianyi Chen. Towards understanding asynchronous advantage actor-critic: Convergence and linear speedup. *IEEE Transactions on Signal Processing*, 2023.

David Silver, Guy Lever, Nicolas Heess, Thomas Degris, Daan Wierstra, and Martin Riedmiller. Deterministic policy gradient algorithms. In *International conference on machine learning*, pp. 387–395. Pmlr, 2014.

Virginia Smith, Chao-Kai Chiang, Maziar Sanjabi, and Ameet S Talwalkar. Federated multi-task learning. *Advances in neural information processing systems*, 30, 2017.

Richard S Sutton. Learning to predict by the methods of temporal differences. *Machine learning*, 3: 9–44, 1988.

Richard S Sutton and Andrew G Barto. *Reinforcement learning: An introduction*. MIT press, 2018.

John Tsitsiklis and Benjamin Van Roy. Analysis of temporal-diffference learning with function approximation. *Advances in neural information processing systems*, 9, 1996.

Isidoros Tziotis, Zebang Shen, Ramtin Pedarsani, Hamed Hassani, and Aryan Mokhtari. Straggler-resilient personalized federated learning. *Transactions on Machine Learning Research*, 2023.

Han Wang, Aritra Mitra, Hamed Hassani, George J Pappas, and James Anderson. Federated temporal difference learning with linear function approximation under environmental heterogeneity. *arXiv preprint arXiv:2302.02212*, 2023a.

Sid Wang, Ashish Shenoy, Pierce Chuang, and John Nguyen. Now it sounds like you: Learning personalized vocabulary on device. *arXiv preprint arXiv:2305.03584*, 2023b.

Christopher JCH Watkins and Peter Dayan. Q-learning. *Machine learning*, 8:279–292, 1992.

Jiin Woo, Gauri Joshi, and Yuejie Chi. The blessing of heterogeneity in federated q-learning: Linear speedup and beyond. In *International Conference on Machine Learning*. PMLR, 2023.

Guojun Xiong, Gang Yan, Shiqiang Wang, and Jian Li. Deprl: Achieving linear convergence speedup in personalized decentralized learning with shared representations. *arXiv preprint arXiv:2312.10815*, 2023.

Minrui Xu, Jialiang Peng, BB Gupta, Jiawen Kang, Zehui Xiong, Zhenni Li, and Ahmed A Abd El-Latif. Multiagent federated reinforcement learning for secure incentive mechanism in intelligent cyber–physical systems. *IEEE Internet of Things Journal*, 9(22):22095–22108, 2021.

Timothy Yang, Galen Andrew, Hubert Eichner, Haicheng Sun, Wei Li, Nicholas Kong, Daniel Ramage, and Françoise Beaufays. Applied federated learning: Improving google keyboard query suggestions. *arXiv preprint arXiv:1812.02903*, 2018.

Zhenyuan Yuan, Siyuan Xu, and Minghui Zhu. Federated reinforcement learning for generalizable motion planning. In *2023 American Control Conference (ACC)*, pp. 78–83. IEEE, 2023.

Chenyu Zhang, Han Wang, Aritra Mitra, and James Anderson. Finite-time analysis of on-policy heterogeneous federated reinforcement learning. *arXiv preprint arXiv:2401.15273*, 2024.

Kaiqing Zhang, Zhuoran Yang, and Tamer Başar. Multi-agent reinforcement learning: A selective overview of theories and algorithms. *Handbook of reinforcement learning and control*, pp. 321–384, 2021.

Sai Qian Zhang, Jieyu Lin, and Qi Zhang. A multi-agent reinforcement learning approach for efficient client selection in federated learning. In *Proceedings of the AAAI Conference on Artificial Intelligence*, volume 36, pp. 9091–9099, 2022a.

Tuo Zhang, Lei Gao, Chaoyang He, Mi Zhang, Bhaskar Krishnamachari, and A Salman Avestimehr. Federated learning for the Internet of Things: Applications, challenges, and opportunities. *IEEE Internet of Things Magazine*, 5(1):24–29, 2022b.

Xuezhou Zhang, Yuda Song, Masatoshi Uehara, Mengdi Wang, Alekh Agarwal, and Wen Sun. Efficient reinforcement learning in block mdps: A model-free representation learning approach. In *International Conference on Machine Learning*, pp. 26517–26547. PMLR, 2022c.

A Related Work

Single-Agent Reinforcement Learning. RL is a machine learning paradigm that trains agents to make sequences of decisions by rewarding desired behaviors and/or penalizing undesired ones in a given environment Sutton & Barto (2018). Starting from Temporal Difference (TD) Learning Sutton (1988), which introduced the concept of learning from the discrepancy between predicted and actual rewards through episodes, the widely used Q-Learning Watkins & Dayan (1992) emerged, advancing the field with an off-policy algorithm that learns action-value functions and enables policy improvement without needing a model of the environment. Later on, the introduction of Deep Q-Networks (DQN) Mnih et al. (2015) marked a significant leap, integrating deep neural networks with Q-Learning to handle high-dimensional state spaces, thus enabling RL to tackle complex problems. Subsequently, policy-based algorithms such as Proximal Policy Optimization (PPO) Schulman et al. (2017) and deep Deterministic Policy Gradients (DDPG) Silver et al. (2014), leverage the Actor-Critic framework to provide more stable and robust ways to directly optimize the policy, overcoming challenges related to action space and variance.

Federated Reinforcement Learning. Jin et al. (2022) introduced a FedRL framework with N agents collaboratively learning a policy by averaging their Q-values or policy gradients. Khodadadian et al. (2022) provided a convergence analysis of federated TD (FedTD) and Q-learning (FedQ) when N agents interact with homogeneous environments. A similar FedTD was considered in Dal Fabbro et al. (2023), and expanded to heterogeneous environments in Wang et al. (2023a). Woo et al. (2023) analyzed (a)synchronous variants of FedQ in heterogeneous settings, and an asynchronous actor-critic method was considered in Shen et al. (2023) with linear speedup guarantee only under i.i.d. samples. Zhang et al. (2024) provided a finite-time analysis of FedSARSA with linear function approximation (i.e., fixed feature representation). To facilitate personalization in heterogeneous settings, Jin et al. (2022) proposed a heuristic personalized FedRL method where agents share a common model, but make use of individual environment embeddings.

Personalized Federated Learning (PFL). In contrast to standard FL where a single model is learned, PFL aims to learn N models specialized for N local datasets. Many PFL methods have been developed, including but not limited to multi-task learning Smith et al. (2017), meta-learning Chen et al. (2018), and various personalization techniques such as local fine-tuning Fallah et al. (2020), layer personalization Arivazhagan et al. (2019), and model compression Bergou et al. (2022). Another line of work Collins et al. (2021); Xiong et al. (2023) leveraged the common representation among agents in heterogeneous environments to guarantee personalized models for federated supervised learning.

Representation Learning in MDP. Representation learning aims to transform high-dimensional observation to low-dimensional embedding to enable efficient learning, and has received increasing attention in Markov decision processs (MDP) settings, such as linear MDPs Jin et al. (2020), low-rank MDPs Modi et al. (2021); Agarwal et al. (2020) and block MDPs Zhang et al. (2022c). However, it is open in the context of leveraging representation learning in PFedFL. In this work, we prove that representation augmented PFedFL forms a general framework as a federated two-timescale stochastic approximation with Markovian noise, which differs significantly from existing works, and hence necessitates different proof techniques.

Multi-Agent Reinforcement Learning vs. Federated Reinforcement Learning. The advent of Multi-Agent Reinforcement Learning (MARL) expanded RL's applications, allowing multiple agents to learn from interactions in cooperative, competitive, or mixed settings, opening new avenues for complex applications and research Zhang et al. (2021). Multi-agent Reinforcement Learning (MARL) addresses scenarios where multiple agents operate within a shared or interrelated environment, potentially engaging in both cooperative and competitive behaviors. The complexity arises from each agent needing to consider the strategies and actions of others, making the learning process highly dynamic. Federated Reinforcement Learning (FedRL)Qi et al. (2021), contrasts with MARL by focusing on privacy-preserving, distributed learning across agents that do not share their raw data.

Algorithm 2 PFedTD-Rep

```
Input: Sampling policy $\pi^i, \forall i \in [N]$;
Initialize $\boldsymbol{\theta}_0^i = \mathbf{0}$, $S_0^i, \forall i \in [N]$, and randomly generate $\mathbf{\Phi} \in \mathbb{R}^{|\mathcal{S}| \times d}$ with each row being unit-norm vector;
for $t = 0, 1, ..., T-1$ do
    for $i = 1, \ldots, N$ do
        for $k = 1, \ldots, K$ do
            Sample observations $X_{t,k-1}^i$;
            Set $\boldsymbol{\theta}_{t,k}^i = \boldsymbol{\theta}_{t,k-1}^i + \alpha_t \mathbf{g}(\boldsymbol{\theta}_{t,k-1}^i, \mathbf{\Phi}_t, X_{t,k-1}^i)$;
        end for
        Scale $\|\boldsymbol{\theta}_{t+1}^i\|$ to $B$ if $\|\boldsymbol{\theta}_{t+1}^i\| > B$, otherwise keep it unchanged;
        Set $\mathbf{\Phi}_{t+1/2}^i = \mathbf{\Phi}_t + \beta_t \mathbf{h}(\boldsymbol{\theta}_{t+1}^i, \mathbf{\Phi}_t, \{X_{t,k-1}^i\}_{k=1}^K)$;
        Normalize $\mathbf{\Phi}_{t+1/2}^i$ as $\mathbf{\Phi}_{t+1/2}^i \leftarrow \frac{\mathbf{\Phi}_{t+1/2}^i}{\|\mathbf{\Phi}_{t+1/2}^i\|}$;
    end for
    $\mathbf{\Phi}_{t+1} = \frac{1}{N} \sum_{i=1}^N \mathbf{\Phi}_{t+1/2}^i$.
end for
```

Instead, these agents might contribute towards a centralized learning model without compromising individual data privacy, addressing the unique challenges of learning from decentralized data sources.

B Pseudocode of PFedTD-Rep

In this section, we present the pseudocode of PFedQ-Rep as summarized in Algorithm 2.

C Application to Control Tasks in RL

The Q-function of agent i in environment $\mathcal{M}^i$ under policy π^i are defined as $Q^{i,\pi^i}(s,a) = \mathbb{E}_{\pi^i}\left[\sum_{k=0}^{\infty} \gamma^k R^i(s_k^i, a_k^i) | s_0^i = s, a_0^i = a\right]$. When the state and action spaces are large, it is computationally infeasible to store $Q^{i,\pi^i}(s,a)$ for all states or state-action pairs. One way to deal with is to approximate the Q-function as $Q^{i,\pi^i}(s,a) \approx \mathbf{\Phi}(s,a)\boldsymbol{\theta}$, where $\mathbf{\Phi} \in \mathbb{R}^{|\mathcal{S}| \times |\mathcal{A}| \times d}$ is a feature representation corresponding to states or state-actions, and $\boldsymbol{\theta} \in \mathbb{R}^d$ is a low-dimensional unknown weight vector. When $\mathbf{\Phi}$ is given and known, this falls under the paradigm of RL or FedRL with function approximation.

C.1 Preliminaries: Control in Federated Reinforcement Learning

Another task in RL is to search for an optimal policy, which is called *a control problem*, and one commonly used approach is Q-learning Watkins & Dayan (1992). Under the FedRL framework, the goal of a control problem is to let N agents collaboratively learn a policy π^* that performs uniformly well across N different environments, i.e., $\pi^* = \arg\max_\pi \frac{1}{N} \sum_{i=1}^N \mathbb{E}_{\pi^i}\left[V^{i,\pi^i}(s_0^i) | s_0^i \sim d_0\right]$, where d_0 is the common initial state distribution in these N environments. Similar to (1), this can be formulated as the optimization problem in (16) to collaboratively learn a common (non-personalized) weight vector $\boldsymbol{\theta} \equiv \boldsymbol{\theta}^i, \forall i \in [N]$ when the feature representation $\mathbf{\Phi}(s,a), \forall s, a$ are given.

$$\mathcal{L}(\boldsymbol{\theta}) := \min_{\boldsymbol{\theta}} \frac{1}{N} \sum_{i=1}^N \mathbb{E}_{\substack{s \sim \mu^{i,\pi^*} \\ a \sim \pi^*(\cdot|s)}} \left\| \mathbf{\Phi}(s,a)\boldsymbol{\theta} - Q^{i,\pi^*}(s,a) \right\|^2. \tag{16}$$

Again, we use the superscript i to highlight heterogeneous environments P^i among agents.

C.2 Control in Personalized FedRL with Shared Representations

The control problem in (16) aims to learn $\boldsymbol{\Phi}$ and $\{\boldsymbol{\theta}^i, \forall i\}$ simultaneously among all N agents via solving the following optimization problem:

$$\mathcal{L}(\boldsymbol{\Phi}, \{\boldsymbol{\theta}^i, \forall i\}) := \min_{\boldsymbol{\Phi}} \frac{1}{N} \sum_{i=1}^{N} \min_{\{\boldsymbol{\theta}^i, \forall i\}} \mathbb{E}_{\substack{s\sim\mu^{i,\pi^{i,*}} \\ a\sim\pi^{i,*}(\cdot|s)}} \left\| f^i(\boldsymbol{\theta}^i, \boldsymbol{\Phi}(s,a)) - Q^{i,\pi^{i,*}}(s,a) \right\|^2. \tag{17}$$

C.3 Algorithms

In this section, we present two realizations of our proposed PFedRL-Rep in Algorithm 1, one is PFedQ-Rep as summarized in Algorithm 3, federated Q-learning with shared representations, and the other is PFedDQN-Rep as outlined in Algorithm 4, federated DQN with shared representations.

Algorithm 3 PFedQ-Rep

```
Input: Sampling policy $\pi^i, \forall i \in [N]$.
Initialize $\boldsymbol{\theta}_0^i = \mathbf{0}$, and $s_0^i, \forall i \in [N]$, and randomly generate $\boldsymbol{\Phi} \in \mathbb{R}^{|\mathcal{S}||\mathcal{A}|\times d}$ with each row being unit-norm vector.
for $t = 0, 1, ..., T-1$ do
  for $i = 1, \dots, N$ do
    for $k = 1, \dots, K$ do
      Sample observations $X_{t,k-1}^i = (s_{t,k}^i, s_{t,k-1}^i, a_{t,k-1}^i)$;
      With fixed $\boldsymbol{\Phi}_t$, update $\boldsymbol{\theta}_{t,k}^i \leftarrow \boldsymbol{\theta}_{t,k-1}^i + \alpha_t \cdot (r_{t,k-1}^i + \gamma \max_a \boldsymbol{\Phi}_t(s_{t,k+1}^i, a)\boldsymbol{\theta}_{t,k-1}^i - \boldsymbol{\Phi}_t(s_{t,k-1}^i)\boldsymbol{\theta}_{t,k-1}^i) \cdot \boldsymbol{\Phi}_t(s_{t,k-1}^i, a_{t,k-1}^i)$;
    end for
    Scale $\|\boldsymbol{\theta}_{t+1}^i\|$ to $B$ if $\|\boldsymbol{\theta}_{t+1}^i\| > B$, otherwise keep it unchanged.
    if $(s,a) \in X_{t,k}^i, \exists k \in \{0, \dots, K-1\}$ then
      Update $\boldsymbol{\Phi}_{t+1/2}^i(s,a) = \boldsymbol{\Phi}_t^i(s,a) + \beta_t(r(s,a) + \gamma \max_a \boldsymbol{\Phi}_t(s',a)\boldsymbol{\theta}_{t+1}^i - \boldsymbol{\Phi}_t(s,a)^\intercal \boldsymbol{\theta}_{t+1}^i) \cdot \boldsymbol{\theta}_{t+1}^i$;
    else
      $\boldsymbol{\Phi}_{t+1/2}^i(s,a) = \boldsymbol{\Phi}_t^i(s,a)$;
    end if
    Normalize $\boldsymbol{\Phi}_{t+1/2}^i$ as $\boldsymbol{\Phi}_{t+1/2}^i \leftarrow \frac{\boldsymbol{\Phi}_{t+1/2}^i}{\|\boldsymbol{\Phi}_{t+1/2}^i\|}$;
  end for
  $\boldsymbol{\Phi}_{t+1} \leftarrow \frac{1}{N}\sum_{i=1}^N \boldsymbol{\Phi}_{t+1/2}^i, \forall i \in [N]$.
end for
```

D Figure Illustrations

We present some figures to further highlight the proposed personalized FedRL (PFedRL) framework with shared representations.

Schematic framework of conventional FedRL. We begin by introducing the conventional FedRL framework Khodadadian et al. (2022), where N agents collaboratively learn a common policy (or optimal value functions) via a server while engaging with homogeneous environments. Each agent generates independent Markovian trajectories, as depicted in Figure 4.

Schematic framework for our proposed PFedRL with shared representations. We introduce our proposed personalized FedRL (PFedRL) framework with shared representations in Figure 5. In PFedRL, N agents independently interact with their own environments and execute actions according to their individual RL component parameterized by $\boldsymbol{\Phi}$ and $\boldsymbol{\theta}^i$. Each agent i performs local update on its local weight vector $\boldsymbol{\theta}_i$, while jointly updating the global shared feature representation $\boldsymbol{\Phi}$ through the server. Similarly, the update follows the Markovian trajectories.

Algorithm 4 PFedDQN-Rep

```
Initialize: The parameters ($\mathbf{\Phi}$,$\boldsymbol{\theta}^i$) for each Q network $Q^i(s,a)$, the replay buffer $\mathcal{R}^i$, and copy the same parameter from Q network to initialize the target Q network $Q^{i,\prime}(s,a)$ for agent $i, \forall i \in [N]$;
for episode $e = 1, \ldots, E$ do
    Get the initial state of the environment;
    for $t = 0, 1, ..., T-1$ do
        for $i = 1, \ldots, N$ do
            for $k = 1, \ldots, K$ do
                Select action $a_{t,k-1}$ according to $\epsilon$-greedy policy with the current network $Q^i(s_{t,k-1}, a)$;
                Execute action $a_{t,k-1}$, receive the reward $r(s_{t,k-1}, a_{t,k-1})$, and the environment state transits to $s_{t,k}$;
                Store the tuple $(s_{t,k-1}, a_{t,k-1}, r(s_{t,k-1}, a_{t,k-1}, s_{t,k})$ into the replay buffer $\mathcal{R}^i$;
                Sample $N$ data tuples from the replay buffer $\mathcal{R}^i$;
                Update the local weight $\boldsymbol{\theta}^i(t,k)$ by minimizing the loss compared with the target network $Q^{i,\prime}$ with fixed representation $\mathbf{\Phi}_t$;
            end for
            Sample $N$ data tuples from replay buffer $\mathcal{R}^i$;
            Update representation model locally by minimizing the loss compared with the target network $Q^{i,\prime}$ with fixed weights $\boldsymbol{\theta}_{t+1}$, and yield $\mathbf{\Phi}^i_{t+1/2}$;
        end for
        Average the representation model from all agents, i.e., $\mathbf{\Phi}_{t+1} := \frac{1}{N}\sum_{i=1}^N \mathbf{\Phi}^i_{t+1/2}$;
    end for
    if $mod(t, T_{target}) = 0$ then
        update the target network $Q^{i,*}$ be copy the up-to-date parameters of Q network $Q^i$, $\forall i \in [N]$;
    end if
end for
```

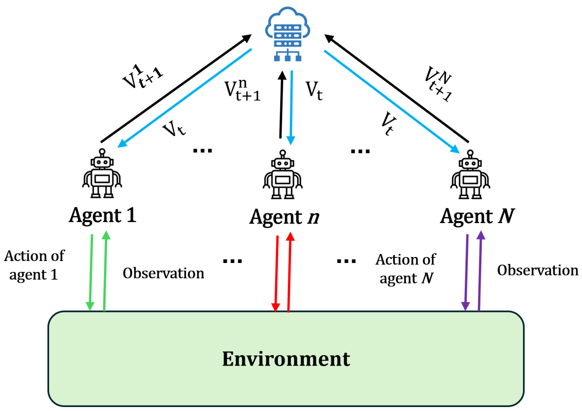


Figure 4: Schematic representation of FedRL, where N agents interact with homogeneous environments.

Motivation of Personalized FedRL. In the following, we also want to provide some examples showing that the conventional FedRL framework may fail, as depicted in Figure 6. In Figure 6a, we provide an example where three agents assess distinct policies within the same environment. In the traditional FedRL framework, agents exchange the evaluated value functions via a central server, leading to a unified consensus on value functions for three different policies. This enforced consensus on value functions, despite the diversity in policies, is not optimal. In another scenario depicted in Figure 6b, three agents each interact with their unique environments. The objective for each agent is to learn an optimal policy tailored to its specific environment. However, within the traditional FedRL framework, the central server mandates a uniform policy across all three agents, which clearly

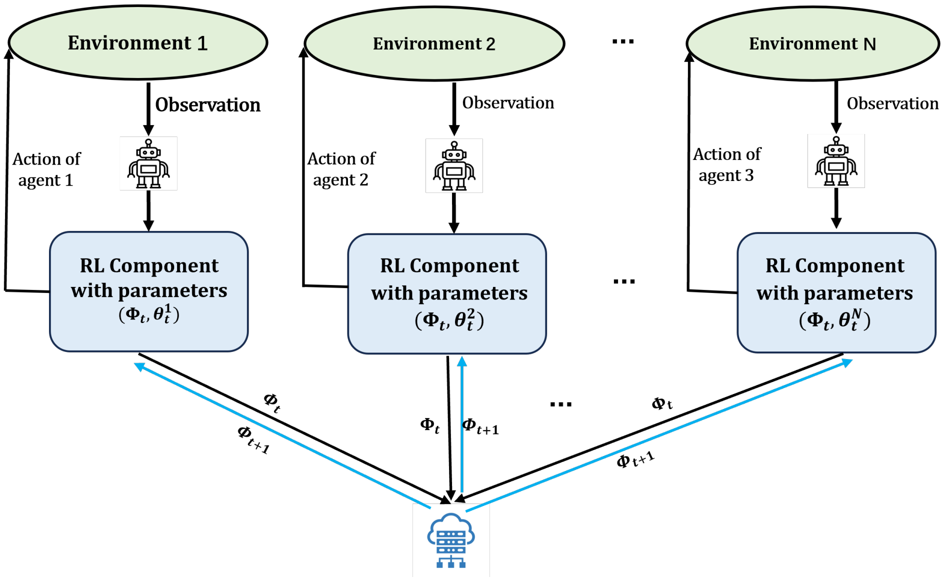
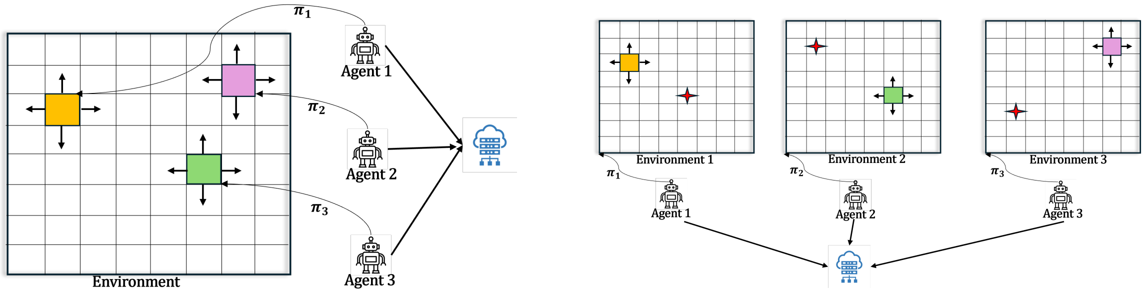


Figure 5: Our proposed PFedRL-Rep framework where N agents independently interact with their own environments and take actions according to their individual RL component parameterized by $\mathbf{\Phi}$ and $\boldsymbol{\theta}^i$. Agent i locally update weight vector $\boldsymbol{\theta}_i$ while jointly updating the shared feature representation $\mathbf{\Phi}$ through the central server. The update follows the Markovian trajectories.

contradicts the intended goal of achieving environment-specific optimization. This highlights the necessity for personalized decision-making, a feature that conventional FedRL frameworks do not accommodate.

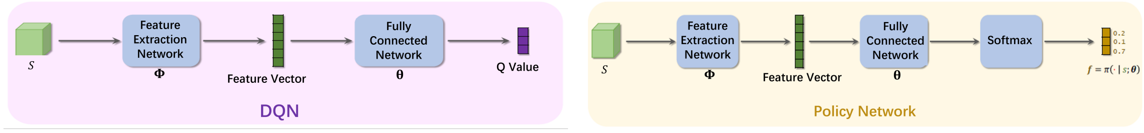


(a) Agents evaluate difference policies in the same environment.

(b) Agents learn optimal policies for heterogeneous environments.

Figure 6: *An illustrative example with three agents that demonstrates the conventional FedRL framework fails to work.*

Example of RL components that fit the proposed PFedRL with shared representations. In the following, we aim to showcase examples of RL components that are compatible with our proposed PFedRL framework featuring shared representations. An illustrative example of this framework is presented in Figure 7. It is important to note that both the DQN architecture in Figure 7a and the policy gradient (PG) approach in Figure 7b seamlessly integrate into our proposed framework. This integration is achieved by designating the parameters of the feature extraction network as the shared feature representation $\mathbf{\Phi}$, and the parameters of the fully connected network, which either predict the Q-values or determine the policy, as the local weight vector $\boldsymbol{\theta}$. This arrangement underscores the adaptability of our framework to various RL methodologies, facilitating personalized learning while maintaining a common foundation of shared representations.



(a) When DQN meets the proposed framework.

(b) When PG meets the proposed framework.

Figure 7: *An illustrative example for the proposed framework. Notice that both the DQN in (a) and policy gradient (PG) in (b) can be fitted into the proposed framework by treating the parameters of the feature extraction network as the shared feature representation* $\mathbf{\Phi}$ *and the parameter of the fully connected network which maps to the Q value of policy as the local weight vector* $\boldsymbol{\theta}$.

E Assumptions, Definitions, and Lemmas

Our goal is to characterize the finite-time convergence of $\mathbb{E}[M(\{\boldsymbol{\theta}_{t+1}^i, \forall i\}, \mathbf{\Phi}_t)]$, the Lyapunov function in (14). We start with some standard assumptions first.

Assumption E.1. Agent i's Markov chain $\{X_t^i\}$ is irreducible and aperiodic. Hence, there exists a unique stationary distribution μ^i Levin & Peres (2017) and constants $C > 0$ and $\rho \in (0,1)$ such that $d_{TV}(P(X_k^i|X_0^i = x), \mu^i) \leq C\rho^k, \forall k \geq 0, x \in \mathcal{X}$, where $d_{TV}(\cdot,\cdot)$ is the total-variation (TV) distance Levin & Peres (2017).

Remark E.2. Assumption E.1 implies that the Markov chain induced by π^i admits a unique stationary distribution μ^i. This assumption is commonly used in the asymptotic convergence analysis of stochastic approximation under Markovian noise Borkar (2009); Chen et al. (2019).

We can define the steady-state local TD update direction as

$$\bar{\mathbf{g}}(\boldsymbol{\theta}^i, \mathbf{\Phi}) := \mathbb{E}_{s_t^i \sim \mu^i, s_{t+1}^i \sim P_{\pi^i}^i(\cdot|s_t^i)}[\mathbf{g}(\boldsymbol{\theta}^i, \mathbf{\Phi}, X_t^i)],$$
$$\bar{\mathbf{h}}(\boldsymbol{\theta}^i, \mathbf{\Phi}) := \mathbb{E}_{s_t^i \sim \mu^i, s_{t+1}^i \sim P_{\pi^i}^i(\cdot|s_t^i)}[\mathbf{h}(\boldsymbol{\theta}^i, \mathbf{\Phi}, X_t^i)]. \tag{18}$$

Definition E.3 (Mixing time Chen et al. (2019))**.** Define $\tau_\delta = \max_{i\in[N]} \min\{t \geq 1 : \{\|\mathbb{E}[\mathbf{g}(\boldsymbol{\theta}^i, \mathbf{\Phi}, X_t^i)|X_0 = x] - \bar{\mathbf{g}}(\boldsymbol{\theta}^i, \mathbf{\Phi})\|,\ \|\mathbb{E}[\mathbf{h}(\boldsymbol{\theta}^i, \mathbf{\Phi}, X_t^i)|X_0 = x] - \bar{\mathbf{h}}(\boldsymbol{\theta}^i, \mathbf{\Phi})\|\} \leq \delta(\|\mathbf{\Phi} - \mathbf{\Phi}^*\| + \|\boldsymbol{\theta}^i - y^i(\mathbf{\Phi}^*)\| + 1)\}, \forall \delta > 0$.

Lemma E.4. $\mathbf{g}(\boldsymbol{\theta}, \mathbf{\Phi}, X)$ *in (8) is globally Lipschitz continuous w.r.t* $\boldsymbol{\theta}$ *and* $\mathbf{\Phi}$ *uniformly in* X*, i.e.,* $\|\mathbf{g}(\boldsymbol{\theta}_1, \mathbf{\Phi}_1, X) - \mathbf{g}(\boldsymbol{\theta}_2, \mathbf{\Phi}_2, X)\| \leq L_g(\|\boldsymbol{\theta}_1 - \boldsymbol{\theta}_2\| + \|\mathbf{\Phi}_1 - \mathbf{\Phi}_2\|), \forall X \in \mathcal{X}$.

Lemma E.5. $\mathbf{h}(\boldsymbol{\theta}, \mathbf{\Phi}, X)$ *in (10) is globally Lipschitz continuous w.r.t* $\boldsymbol{\theta}$ *and* $\mathbf{\Phi}$ *uniformly in* X*, i.e.,* $\|\mathbf{h}(\boldsymbol{\theta}_1, \mathbf{\Phi}_1, X) - \mathbf{h}(\boldsymbol{\theta}_2, \mathbf{\Phi}_2, X)\| \leq L_h(\|\boldsymbol{\theta}_1 - \boldsymbol{\theta}_2\| + \|\mathbf{\Phi}_1 - \mathbf{\Phi}_2\|), \forall X \in \mathcal{X}$.

Lemma E.6. $y^i(\mathbf{\Phi}), \forall i$ *is Lipschitz continuous in* $\mathbf{\Phi}$*, i.e.,* $\|y^i(\mathbf{\Phi}_1) - y^i(\mathbf{\Phi}_2)\| \leq L_y\|\mathbf{\Phi}_1 - \mathbf{\Phi}_2\|$.

Remark E.7. The Lipschitz continuity of $\mathbf{h}$ guarantees the existence of a solution $\mathbf{\Phi}$ to the equilibrium (12) for a fixed $\boldsymbol{\theta}$, while the Lipschitz continuity of $\mathbf{g}$ and y^i ensures the existence of a solution $\boldsymbol{\theta}^i$ of (12) when $\mathbf{\Phi}$ is fixed.

Assumption E.8. There exists a $\omega > 0$ such that $\forall \mathbf{\Phi}$, $\boldsymbol{\theta}$ and $\forall i$:

$$\langle \mathbf{\Phi} - \mathbf{\Phi}^*, \bar{\mathbf{h}}(y^i(\mathbf{\Phi}), \mathbf{\Phi})\rangle \leq -\omega\|\mathbf{\Phi}^* - \mathbf{\Phi}\|^2, \qquad \langle \boldsymbol{\theta}_t^i - y^i(\mathbf{\Phi}_{t-1}), \bar{\mathbf{g}}(\boldsymbol{\theta}_t^i, \mathbf{\Phi}_{t-1})\rangle \leq -\omega\|\boldsymbol{\theta} - y^i(\mathbf{\Phi})\|^2.$$

Remark E.9. Assumption E.8 guarantees the stability of the two-timescale update in (7) and (11), and can be viewed as the monotone property of nonlinear mappings leveraged in Doan (2020); Chen et al. (2019).

Lemma E.10. *Under Assumption E.1, and Lemma E.4 and E.5, there exist constants* $C > 0$*,* $\rho \in (0,1)$ *and* $L_1 = \max(L_g, L_h, \max_X \mathbf{g}(\boldsymbol{\theta}^*, \mathbf{\Phi}^*, X), \max_X \mathbf{h}(\boldsymbol{\theta}^*, \mathbf{\Phi}^*, X))$ *such that* $\tau_\delta \leq \frac{\log(1/\delta) + \log(2L_1Cd)}{\log(1/\rho)}$*, and* $\lim_{\delta\to 0}\delta\tau_\delta = 0$.

E.1 Proof of Lemma E.4

Recall that for any observation $X = (s, a, s')$, the function $\mathbf{g}(\boldsymbol{\theta}, \mathbf{\Phi}, X)$ defined in (8) is expressed as

$$\mathbf{g}(\boldsymbol{\theta}, \mathbf{\Phi}, X) := (r(s,a) + \gamma\mathbf{\Phi}(s')\boldsymbol{\theta} - \mathbf{\Phi}(s)\boldsymbol{\theta}) \cdot \mathbf{\Phi}(s)^\intercal,$$

and hence we have the following inequality for any parameter pairs $(\boldsymbol{\theta}_1, \boldsymbol{\Phi}_1)$ and $(\boldsymbol{\theta}_2, \boldsymbol{\lambda}_2)$ with $X = (s, a, s')$,

$$
\begin{aligned}
&\|\mathbf{g}(\boldsymbol{\theta}_1, \boldsymbol{\Phi}_1, X) - \mathbf{g}(\boldsymbol{\theta}_2, \boldsymbol{\Phi}_2, X)\| \\
&= \|(r(s,a) + \gamma\boldsymbol{\Phi}_1(s')\boldsymbol{\theta}_1 - \boldsymbol{\Phi}_1(s)\boldsymbol{\theta}_1) \cdot \boldsymbol{\Phi}_1(s)^\intercal - (r(s,a) + \gamma\boldsymbol{\Phi}_2(s')\boldsymbol{\theta}_2 - \boldsymbol{\Phi}_2(s)\boldsymbol{\theta}_2) \cdot \boldsymbol{\Phi}_2(s)^\intercal\| \\
&\overset{(a_1)}{\leq} \|(\gamma\boldsymbol{\Phi}_1(s')\boldsymbol{\theta}_1 - \boldsymbol{\Phi}_1(s)\boldsymbol{\theta}_1) \cdot \boldsymbol{\Phi}_1(s)^\intercal - (\gamma\boldsymbol{\Phi}_1(s')\boldsymbol{\theta}_2 - \boldsymbol{\Phi}_1(s)\boldsymbol{\theta}_2) \cdot \boldsymbol{\Phi}_1(s)^\intercal\| \\
&\qquad\qquad\qquad + \|(\gamma\boldsymbol{\Phi}_1(s')\boldsymbol{\theta}_2 - \boldsymbol{\Phi}_1(s)\boldsymbol{\theta}_2) \cdot \boldsymbol{\Phi}_1(s)^\intercal - (\gamma\boldsymbol{\Phi}_2(s')\boldsymbol{\theta}_2 - \boldsymbol{\Phi}_2(s)\boldsymbol{\theta}_2) \cdot \boldsymbol{\Phi}_2(s)^\intercal\| \\
&\overset{(a_2)}{\leq} \|(\gamma\boldsymbol{\Phi}_1(s')\boldsymbol{\theta}_1 - \boldsymbol{\Phi}_1(s)\boldsymbol{\theta}_1) - (\gamma\boldsymbol{\Phi}_1(s')\boldsymbol{\theta}_2 - \boldsymbol{\Phi}_1(s)\boldsymbol{\theta}_2)\| \cdot \|\boldsymbol{\Phi}_1(s)\| \\
&\qquad\qquad\qquad + \|(\gamma\boldsymbol{\Phi}_1(s')\boldsymbol{\theta}_2 - \boldsymbol{\Phi}_1(s)\boldsymbol{\theta}_2) \cdot \boldsymbol{\Phi}_1(s)^\intercal - (\gamma\boldsymbol{\Phi}_2(s')\boldsymbol{\theta}_2 - \boldsymbol{\Phi}_2(s)\boldsymbol{\theta}_2) \cdot \boldsymbol{\Phi}_2(s)^\intercal\| \\
&\overset{(a_3)}{\leq} (1+\gamma)\,\|\boldsymbol{\theta}_1 - \boldsymbol{\theta}_2\| + \|(\gamma\boldsymbol{\Phi}_1(s')\boldsymbol{\theta}_2 - \boldsymbol{\Phi}_1(s)\boldsymbol{\theta}_2) \cdot \boldsymbol{\Phi}_1(s)^\intercal - (\gamma\boldsymbol{\Phi}_2(s')\boldsymbol{\theta}_2 - \boldsymbol{\Phi}_2(s)\boldsymbol{\theta}_2) \cdot \boldsymbol{\Phi}_2(s)^\intercal\| \\
&\overset{(a_4)}{\leq} (1+\gamma)\,\|\boldsymbol{\theta}_1 - \boldsymbol{\theta}_2\| + \|(\gamma\boldsymbol{\Phi}_1(s')\boldsymbol{\theta}_2 - \boldsymbol{\Phi}_1(s)\boldsymbol{\theta}_2) \cdot \boldsymbol{\Phi}_1(s)^\intercal - (\gamma\boldsymbol{\Phi}_1(s')\boldsymbol{\theta}_2 - \boldsymbol{\Phi}_1(s)\boldsymbol{\theta}_2) \cdot \boldsymbol{\Phi}_2(s)^\intercal\| \\
&\qquad\qquad\qquad + \|(\gamma\boldsymbol{\Phi}_1(s')\boldsymbol{\theta}_2 - \boldsymbol{\Phi}_1(s)\boldsymbol{\theta}_2) \cdot \boldsymbol{\Phi}_2(s)^\intercal - (\gamma\boldsymbol{\Phi}_2(s')\boldsymbol{\theta}_2 - \boldsymbol{\Phi}_2(s)\boldsymbol{\theta}_2) \cdot \boldsymbol{\Phi}_2(s)^\intercal\| \\
&\overset{(a_5)}{\leq} (1+\gamma)\,\|\boldsymbol{\theta}_1 - \boldsymbol{\theta}_2\| + \Big\|(\gamma\boldsymbol{\Phi}_1(s')\boldsymbol{\theta}_2 - \boldsymbol{\Phi}_1(s)\boldsymbol{\theta}_2)\Big\| \cdot \Big\|\boldsymbol{\Phi}_1(s) - \boldsymbol{\Phi}_2(s)\Big\| \\
&\qquad\qquad\qquad + \|(\gamma\boldsymbol{\Phi}_1(s')\boldsymbol{\theta}_2 - \boldsymbol{\Phi}_1(s)\boldsymbol{\theta}_2) - (\gamma\boldsymbol{\Phi}_2(s')\boldsymbol{\theta}_2 - \boldsymbol{\Phi}_2(s)\boldsymbol{\theta}_2)\| \cdot \|\boldsymbol{\Phi}_2(s)\| \\
&\overset{(a_6)}{\leq} (1+\gamma)\,\Big\|\boldsymbol{\theta}_1 - \boldsymbol{\theta}_2\Big\| + \Big\|(\gamma\boldsymbol{\Phi}_1(s')\boldsymbol{\theta}_2 - \boldsymbol{\Phi}_1(s)\boldsymbol{\theta}_2)\Big\| \cdot \|\boldsymbol{\Phi}_1(s) - \boldsymbol{\Phi}_2(s)\| \\
&\qquad\qquad\qquad + \|\boldsymbol{\Phi}_1(s') - \boldsymbol{\Phi}_2(s')\| \cdot \|\gamma\boldsymbol{\theta}_2\| + \|\boldsymbol{\Phi}_1(s) - \boldsymbol{\Phi}_2(s)\| \cdot \|\boldsymbol{\theta}_2\| \\
&\leq (1+\gamma)\,\|\boldsymbol{\theta}_1 - \boldsymbol{\theta}_2\| + (2+2\gamma)\,\|\boldsymbol{\theta}_2\| \cdot \|\boldsymbol{\Phi}_1 - \boldsymbol{\Phi}_2\| \\
&\overset{(a_7)}{\leq} L_g\left(\|\boldsymbol{\theta}_1 - \boldsymbol{\theta}_2\| + \|\boldsymbol{\Phi}_1 - \boldsymbol{\Phi}_2\|\right),
\end{aligned}
$$

(a_1) is due to the fact that $\|\mathbf{x}+\mathbf{y}\| \leq \|\mathbf{x}\|+\|\mathbf{y}\|, \forall \mathbf{x}, \mathbf{y} \in \mathbb{R}^d$, (a_2) holds due to $\|\mathbf{x}\cdot\mathbf{y}\| \leq \|\mathbf{x}\|\cdot\|\mathbf{y}\|, \forall \mathbf{x}, \mathbf{y} \in \mathbb{R}^d$, (a_3) comes from the fact and $\|\boldsymbol{\Phi}_1(s)\| \leq 1, \|\boldsymbol{\Phi}_2(s)\| \leq 1 \forall s$. $(a_4) - (a_6)$ holds for the same reason as $(a_1) - (a_3)$. The last inequalty (a_7) comes from the fact that $\boldsymbol{\theta}$ is bounded by norm B and by setting $L_g := \max(1+\gamma, (2+2\gamma)B)$.

E.2 Proof of Lemma E.5

Recall that for any observation $X = (s, a, s')$, the function $\mathbf{h}(\boldsymbol{\theta}, \boldsymbol{\Phi}, X)$ defined in (10) is expressed as

$$
\mathbf{h}(\boldsymbol{\theta}, \boldsymbol{\Phi}, X) := (r(s,a) + \gamma\boldsymbol{\Phi}(s')\boldsymbol{\theta} - \boldsymbol{\Phi}(s)\boldsymbol{\theta}) \cdot \boldsymbol{\theta}^\intercal,
$$

and hence we have the following inequality for any parameter pairs $(\boldsymbol{\theta}_1, \boldsymbol{\Phi}_1)$ and $(\boldsymbol{\theta}_2, \boldsymbol{\lambda}_2)$ with $X = (s, a, s')$,

$$
\begin{aligned}
&\|\mathbf{h}(\boldsymbol{\theta}_1, \boldsymbol{\Phi}_1, X) - \mathbf{h}(\boldsymbol{\theta}_2, \boldsymbol{\Phi}_2, X)\| \\
&= \|(r(s,a) + \gamma\boldsymbol{\Phi}_1(s')\boldsymbol{\theta}_1 - \boldsymbol{\Phi}_1(s)\boldsymbol{\theta}_1) \cdot \boldsymbol{\theta}_1^\intercal - (r(s,a) + \gamma\boldsymbol{\Phi}_2(s')\boldsymbol{\theta}_2 - \boldsymbol{\Phi}_2(s)\boldsymbol{\theta}_2) \cdot \boldsymbol{\theta}_2^\intercal\| \\
&\overset{(b_1)}{\leq} \|(\gamma\boldsymbol{\Phi}_1(s')\boldsymbol{\theta}_1 - \boldsymbol{\Phi}_1(s)\boldsymbol{\theta}_1) \cdot \boldsymbol{\theta}_1^\intercal - (\gamma\boldsymbol{\Phi}_2(s')\boldsymbol{\theta}_1 - \boldsymbol{\Phi}_2(s)\boldsymbol{\theta}_1) \cdot \boldsymbol{\theta}_1^\intercal\| \\
&\qquad\qquad\qquad + \|(\gamma\boldsymbol{\Phi}_2(s')\boldsymbol{\theta}_1 - \boldsymbol{\Phi}_2(s)\boldsymbol{\theta}_1) \cdot \boldsymbol{\theta}_1^\intercal - (\gamma\boldsymbol{\Phi}_2(s')\boldsymbol{\theta}_2 - \boldsymbol{\Phi}_2(s)\boldsymbol{\theta}_2) \cdot \boldsymbol{\theta}_2^\intercal\| \\
&\overset{(b_2)}{\leq} \|(\gamma\boldsymbol{\Phi}_1(s')\boldsymbol{\theta}_1 - \boldsymbol{\Phi}_1(s)\boldsymbol{\theta}_1) - (\gamma\boldsymbol{\Phi}_2(s')\boldsymbol{\theta}_1 - \boldsymbol{\Phi}_2(s)\boldsymbol{\theta}_1)\| \cdot \|\boldsymbol{\theta}_1\| \\
&\qquad\qquad\qquad + \|(\gamma\boldsymbol{\Phi}_2(s')\boldsymbol{\theta}_1 - \boldsymbol{\Phi}_2(s)\boldsymbol{\theta}_1) \cdot \boldsymbol{\theta}_1^\intercal - (\gamma\boldsymbol{\Phi}_2(s')\boldsymbol{\theta}_2 - \boldsymbol{\Phi}_2(s)\boldsymbol{\theta}_2) \cdot \boldsymbol{\theta}_2^\intercal\| \\
&\overset{(b_3)}{\leq} (1+\gamma)\|\boldsymbol{\theta}_1\|^2 \cdot \|\boldsymbol{\Phi}_1 - \boldsymbol{\Phi}_2\| + \|(\gamma\boldsymbol{\Phi}_2(s')\boldsymbol{\theta}_1 - \boldsymbol{\Phi}_2(s)\boldsymbol{\theta}_1) \cdot \boldsymbol{\theta}_1^\intercal - (\gamma\boldsymbol{\Phi}_2(s')\boldsymbol{\theta}_2 - \boldsymbol{\Phi}_2(s)\boldsymbol{\theta}_2) \cdot \boldsymbol{\theta}_2^\intercal\| \\
&\overset{(b_4)}{\leq} (1+\gamma)\|\boldsymbol{\theta}_1\|^2 \cdot \|\boldsymbol{\Phi}_1 - \boldsymbol{\Phi}_2\| + \|(\gamma\boldsymbol{\Phi}_2(s')\boldsymbol{\theta}_1 - \boldsymbol{\Phi}_2(s)\boldsymbol{\theta}_1) \cdot \boldsymbol{\theta}_1^\intercal - (\gamma\boldsymbol{\Phi}_2(s')\boldsymbol{\theta}_1 - \boldsymbol{\Phi}_2(s)\boldsymbol{\theta}_1) \cdot \boldsymbol{\theta}_2^\intercal\|
\end{aligned}
$$

$$
\begin{aligned}
&\quad + \|(\gamma\boldsymbol{\Phi}_2(s')\boldsymbol{\theta}_1 - \boldsymbol{\Phi}_2(s)\boldsymbol{\theta}_1)\cdot\boldsymbol{\theta}_2^\intercal - (\gamma\boldsymbol{\Phi}_2(s')\boldsymbol{\theta}_2 - \boldsymbol{\Phi}_2(s)\boldsymbol{\theta}_2)\cdot\boldsymbol{\theta}_2^\intercal\| \\
&\overset{(b_5)}{\le} (1+\gamma)\|\boldsymbol{\theta}_1\|^2\cdot\|\boldsymbol{\Phi}_1 - \boldsymbol{\Phi}_2\| + \|(\gamma\boldsymbol{\Phi}_2(s')\boldsymbol{\theta}_1 - \boldsymbol{\Phi}_2(s)\boldsymbol{\theta}_1)\|\cdot\|\boldsymbol{\theta}_1 - \boldsymbol{\theta}_2\| \\
&\quad + \|(\gamma\boldsymbol{\phi}_2(s')\boldsymbol{\theta}_1 - \boldsymbol{\Phi}_2(s)\boldsymbol{\theta}_1) - (\gamma\boldsymbol{\Phi}_2(s')\boldsymbol{\theta}_2 - \boldsymbol{\Phi}_2(s)\boldsymbol{\theta}_2)\|\cdot\|\boldsymbol{\theta}_2\| \\
&\overset{(b_6)}{\le} (1+\gamma)\|\boldsymbol{\theta}_1\|^2\cdot\|\boldsymbol{\Phi}_1 - \boldsymbol{\Phi}_2\| + (1+\gamma)\|\boldsymbol{\theta}_1\|\cdot\|\boldsymbol{\theta}_1 - \boldsymbol{\theta}_2\| + (1+\gamma)\|\boldsymbol{\theta}_2\|\cdot\|\boldsymbol{\theta}_1 - \boldsymbol{\theta}_2\| \\
&\le (1+\gamma)\|\boldsymbol{\theta}_1\|^2\cdot\|\boldsymbol{\Phi}_1 - \boldsymbol{\Phi}_2\| + (1+\gamma)(\|\boldsymbol{\theta}_1\| + \|\boldsymbol{\theta}_2\|)\cdot\|\boldsymbol{\theta}_1 - \boldsymbol{\theta}_2\| \\
&\overset{(b_7)}{\le} L_h(\|\boldsymbol{\theta}_1 - \boldsymbol{\theta}_2\| + \|\boldsymbol{\Phi}_1 - \boldsymbol{\Phi}_2\|),
\end{aligned}
$$

(b_1) is due to the fact that $\|\mathbf{x}+\mathbf{y}\| \le \|\mathbf{x}\|+\|\mathbf{y}\|, \forall \mathbf{x}, \mathbf{y} \in \mathbb{R}^d$, (b_2) holds due to $\|\mathbf{x}\cdot\mathbf{y}\| \le \|\mathbf{x}\|\cdot\|\mathbf{y}\|, \forall \mathbf{x}, \mathbf{y} \in \mathbb{R}^d$, (b_3) comes from the fact and $\|\boldsymbol{\Phi}_1(s)\| \le 1, \|\boldsymbol{\Phi}_2(s)\| \le 1 \forall s$. $(b_4) - (b_6)$ holds for the same reason as $(b_1) - (b_3)$. The last inequalty (b_7) comes from by setting $L_h := \max((1+\gamma)B^2, (2+2\gamma)B)$.

E.3 Proof of Lemma E.6

Due to the norm-scale step (step 9) in Algorithm 2, we have

$$\|y^i(\boldsymbol{\Phi}_1) - y^i(\boldsymbol{\Phi}_2)\| \le \max_{(\|\boldsymbol{\theta}\|\le B, \|\boldsymbol{\theta}'\|\le B)} \|\boldsymbol{\theta} - \boldsymbol{\theta}'\| \le 2B. \tag{19}$$

Since the representation matrices $\boldsymbol{\Phi}_1$ and $\boldsymbol{\Phi}_2$ are of unit-norm in each row, there exists a positive constant L_y such that

$$\|y^i(\boldsymbol{\Phi}_1) - y^i(\boldsymbol{\Phi}_2)\| \le L_y\|\boldsymbol{\Phi}_1 - \boldsymbol{\Phi}_2\|. \tag{20}$$

E.4 Proof of Lemma E.10

Proof. Under Lemma E.4, we have

$$\|\mathbf{g}(\boldsymbol{\theta}, \boldsymbol{\Phi}, X) - \mathbf{g}(y^i(\boldsymbol{\Phi}^*), \boldsymbol{\Phi}^*, X)\| \le L(\|\boldsymbol{\theta} - y^i(\boldsymbol{\Phi}^*)\| + \|\boldsymbol{\Phi} - \boldsymbol{\Phi}^*\|), \forall i \in [N]. \tag{21}$$

Similarly, under Lemma E.5, we have

$$\|\mathbf{h}(\boldsymbol{\theta}, \boldsymbol{\Phi}, X) - \mathbf{h}(y^i(\boldsymbol{\Phi}^*), \boldsymbol{\Phi}^*, X)\| \le L(\|\boldsymbol{\theta} - y^i(\boldsymbol{\Phi}^*)\| + \|\boldsymbol{\Phi} - \boldsymbol{\Phi}^*\|), \forall i \in [N]. \tag{22}$$

Let $L_1 = \max(L, \max_X \mathbf{g}(y^i(\boldsymbol{\Phi}^*), \boldsymbol{\Phi}^*, X), \max_X \mathbf{h}(y^i(\boldsymbol{\Phi}^*), \boldsymbol{\Phi}^*, X))$, then according to (21)-(22), we have

$$\|\mathbf{g}(\boldsymbol{\theta}, \boldsymbol{\Phi})\| \le L_1(\|\boldsymbol{\theta} - y^i(\boldsymbol{\Phi}^*)\| + \|\boldsymbol{\Phi} - \boldsymbol{\Phi}^*\| + 1),$$

and

$$\|\mathbf{h}(\boldsymbol{\theta}, \boldsymbol{\Phi})\| \le L_1(\|\boldsymbol{\theta} - y^i(\boldsymbol{\Phi}^*)\| + \|\boldsymbol{\Phi} - \boldsymbol{\Phi}^*\| + 1).$$

Denote $h^j(\boldsymbol{\theta}, \boldsymbol{\phi}, X)$ as the j-th element of $\mathbf{h}(\boldsymbol{\theta}, \boldsymbol{\Phi}, X)$. Following Chen et al. (2019), we can show that $\boldsymbol{\theta} \in \mathbb{R}^d$, $\boldsymbol{\Phi} \in \mathbb{R}^{|\mathcal{S}|\times d}$, and $x \in \mathcal{X}$,

$$
\begin{aligned}
&\|\mathbb{E}[\mathbf{h}(\boldsymbol{\theta}, \boldsymbol{\Phi}, X)|X_0 = x] - \mathbb{E}_\mu[\mathbf{h}(\boldsymbol{\theta}, \boldsymbol{\Phi}, X)]\| \\
&\le \sum_{j=1}^d |\mathbb{E}[h^j(\boldsymbol{\theta}, \boldsymbol{\lambda}, X)|X_0 = x] - \mathbb{E}_\mu[h^j(\boldsymbol{\theta}, \boldsymbol{\Phi}, X)]| \\
&\le 2L_1(\|\boldsymbol{\theta} - y^i(\boldsymbol{\Phi}^*)\| + \|\boldsymbol{\Phi} - \boldsymbol{\Phi}^*\| + 1)\sum_{j=1}^d \Bigg|\mathbb{E}\left[\frac{h^j(\boldsymbol{\theta}, \boldsymbol{\Phi}, X)}{2L_1(\|\boldsymbol{\theta} - y^i(\boldsymbol{\Phi}^*)\| + \|\lambda - \lambda^*\| + 1)}\Big|X_0 = x\right] \\
&\qquad - \mathbb{E}_\mu\left[\frac{h^j(\boldsymbol{\theta}, \boldsymbol{\Phi}, X)}{2L_1(\|\boldsymbol{\theta} - y^i(\boldsymbol{\Phi}^*)\| + \|\lambda - \lambda^*\| + 1)}\right]\Bigg| \\
&\le 2L_1(\|\boldsymbol{\theta} - y^i(\boldsymbol{\Phi}^*)\| + \|\boldsymbol{\Phi} - \boldsymbol{\Phi}^*\| + 1)dC_1\rho_1^k,
\end{aligned}
$$

where the last inequality holds due to Assumption E.1 with constants $C_1 > 0$ and $\rho_1 \in (0, 1)$. To guarantee $2L_1(\|\boldsymbol{\theta} - y^i(\boldsymbol{\Phi}^*)\| + \|\boldsymbol{\Phi} - \boldsymbol{\Phi}^*\| + 1)dC_1\rho_1^k \leq \delta(\|\boldsymbol{\theta} - y^i(\boldsymbol{\Phi}^*)\| + \|\boldsymbol{\Phi} - \boldsymbol{\Phi}^*\| + 1)$, we have

$$\tau_\delta \leq \frac{\log(1/\delta) + \log(2L_1C_1d)}{\log(1/\rho_1)}. \tag{23}$$

Using the same procedures we can show that

$$\|\mathbb{E}[\mathbf{g}(\boldsymbol{\theta}, \boldsymbol{\Phi}, X)|X_0 = x] - \mathbb{E}_\mu[\mathbf{g}(\boldsymbol{\theta}, \boldsymbol{\Phi}, X)]\| \leq 2L_1(\|\boldsymbol{\theta} - y^i(\boldsymbol{\Phi}^*)\| + \|\boldsymbol{\Phi} - \boldsymbol{\Phi}^*\| + 1)dC_2\rho_2^k,$$

hence we have

$$\tau_\delta \leq \frac{\log(1/\delta) + \log(2L_1C_2d)}{\log(1/\rho_2)}. \tag{24}$$

By setting τ_δ as the largest value in (23) and (24), we arrive at the final result in Lemma E.10. □

F Proofs of Main Results

F.1 Proof of Theorem 4.1

For notational simplicity, in the proofs, we use $\mathbf{h}(\boldsymbol{\theta}_{t+1}^i, \boldsymbol{\Phi}_t)$ to denote $\mathbf{h}(\boldsymbol{\theta}_{t+1}^i, \boldsymbol{\Phi}_t, \{X_{t,k-1}^i\}_{k=1}^K)$, and $\mathbf{g}(\boldsymbol{\theta}_{t,k-1}^i, \boldsymbol{\Phi}_t)$ to denote $\mathbf{g}(\boldsymbol{\theta}_{t,k-1}^i, \boldsymbol{\Phi}_t, X_{t,k-1}^i)$. In the following, we first focus on the update of the global representation $\boldsymbol{\Phi}_t$ and characterize the drift of it.

F.1.1 Drift of $\boldsymbol{\Phi}_t$

The drift of $\boldsymbol{\Phi}_t$ is given in the following lemma.

Lemma F.1. *The drift between $\boldsymbol{\Phi}_{t+1}$ and $\boldsymbol{\Phi}_t$ is given by*

$$\begin{aligned}
&\mathbb{E}[\|\boldsymbol{\Phi}_{t+1} - \boldsymbol{\Phi}^*\|^2] \\
&= \mathbb{E}[\|\boldsymbol{\Phi}_t - \boldsymbol{\Phi}^*\|^2] + \underbrace{\frac{\beta_t^2}{N^2}\mathbb{E}\left[\left\|\sum_{i=1}^N \mathbf{h}(\boldsymbol{\theta}_{t+1}^i, \boldsymbol{\Phi}_t)\right\|^2\right]}_{Term_1} + \underbrace{2\beta_t\mathbb{E}\left[\langle \boldsymbol{\Phi}^* - \boldsymbol{\Phi}_t, \frac{-1}{N}\sum_{i=1}^N \bar{\mathbf{h}}(\boldsymbol{\theta}_{t+1}^i, \boldsymbol{\Phi}_t)\rangle\right]}_{Term_2} \\
&\quad + \underbrace{2\beta_t\mathbb{E}\left[\langle \boldsymbol{\Phi}_t - \boldsymbol{\Phi}^*, \frac{1}{N}\sum_{i=1}^N \mathbf{h}(\boldsymbol{\theta}_{t+1}^i, \boldsymbol{\Phi}_t) - \bar{\mathbf{h}}(\boldsymbol{\theta}_{t+1}^i, \boldsymbol{\Phi}_t)\rangle\right]}_{Term_3}.
\end{aligned} \tag{25}$$

Proof. Based on the update of $\boldsymbol{\Phi}_t$ in (11), We have the following equation

$$\begin{aligned}
&\mathbb{E}[\|\boldsymbol{\Phi}_{t+1} - \boldsymbol{\Phi}^*\|^2] - \mathbb{E}[\|\boldsymbol{\Phi}_t - \boldsymbol{\Phi}^*\|^2] \\
&= \mathbb{E}[\|\boldsymbol{\Phi}^*\|^2 + \|\boldsymbol{\Phi}_{t+1}\|^2 - 2\langle \boldsymbol{\Phi}^*, \boldsymbol{\Phi}_{t+1}\rangle] - \mathbb{E}[\|\boldsymbol{\Phi}^*\|^2 + \|\boldsymbol{\Phi}_t\|^2 - 2\langle \boldsymbol{\Phi}^*, \boldsymbol{\Phi}_t\rangle] \\
&= \mathbb{E}[\|\boldsymbol{\Phi}_{t+1}\|^2] - \mathbb{E}[\|\boldsymbol{\Phi}_t\|^2] - 2\langle \boldsymbol{\Phi}^*, \boldsymbol{\Phi}_{t+1} - \boldsymbol{\Phi}_t\rangle] \\
&= \mathbb{E}[\langle \boldsymbol{\Phi}_{t+1} - \boldsymbol{\Phi}_t, \boldsymbol{\Phi}_{t+1} + \boldsymbol{\Phi}_t\rangle] - 2\langle \boldsymbol{\Phi}^*, \boldsymbol{\Phi}_{t+1} - \boldsymbol{\Phi}_t\rangle] \\
&= \mathbb{E}[\langle \boldsymbol{\Phi}_{t+1} - \boldsymbol{\Phi}_t, \boldsymbol{\Phi}_{t+1} - \boldsymbol{\Phi}_t\rangle] + 2\mathbb{E}[\langle \boldsymbol{\Phi}_{t+1} - \boldsymbol{\Phi}_t, \boldsymbol{\Phi}_t\rangle] - 2\langle \boldsymbol{\Phi}^*, \boldsymbol{\Phi}_{t+1} - \boldsymbol{\Phi}_t\rangle] \\
&= \frac{\beta_t^2}{N^2}\mathbb{E}\left[\left\|\sum_{i=1}^N \mathbf{h}(\boldsymbol{\theta}_{t+1}^i, \boldsymbol{\Phi}_t)\right\|^2\right] - 2\beta_t\mathbb{E}\left[\langle \boldsymbol{\Phi}^* - \boldsymbol{\Phi}_t, \frac{1}{N}\sum_{i=1}^N \mathbf{h}(\boldsymbol{\theta}_{t+1}^i, \boldsymbol{\Phi}_t)\rangle\right],
\end{aligned} \tag{26}$$

which directly leads to

$$\begin{aligned}
&\mathbb{E}[\|\boldsymbol{\Phi}_{t+1}-\boldsymbol{\Phi}^*\|^2] \\
&= \mathbb{E}[\|\boldsymbol{\Phi}_t-\boldsymbol{\Phi}^*\|^2] + \frac{\beta_t^2}{N^2}\mathbb{E}\left[\left\|\sum_{i=1}^N \mathbf{h}(\boldsymbol{\theta}_{t+1}^i,\boldsymbol{\Phi}_t)\right\|^2\right] - 2\beta_t\mathbb{E}\left[\langle \boldsymbol{\Phi}^*-\boldsymbol{\Phi}_t, \frac{1}{N}\sum_{i=1}^N \mathbf{h}(\boldsymbol{\theta}_{t+1}^i,\boldsymbol{\Phi}_t)\rangle\right].
\end{aligned} \tag{27}$$

Rearranging the last term yields the desired result. □

In the following, we separately bound $Term_1$ to $Term_3$.

We first bound $Term_1$ as follows.

Lemma F.2. *For any $t \geq \tau$, we have*

$$Term_1 \leq 4\beta_t^2(L^2+L^4)\mathbb{E}[\|\boldsymbol{\Phi}^*-\boldsymbol{\Phi}_t\|^2] + \frac{4\beta_t^2L^2}{N}\mathbb{E}\left[\sum_{i=1}^N \|\boldsymbol{\theta}_{t+1}-y^i(\boldsymbol{\Phi}_t)\|^2\right] + 4\beta_t^2\delta^2 \tag{28}$$

Proof. Note that

$$\begin{aligned}
Term_1 &= \frac{\beta_t^2}{N^2}\mathbb{E}\left[\left\|\sum_{i=1}^N \mathbf{h}(\boldsymbol{\theta}_{t+1}^i,\boldsymbol{\Phi}_t) - \sum_{i=1}^N \mathbf{h}(y^i(\boldsymbol{\Phi}_t),\boldsymbol{\Phi}^*) + \sum_{i=1}^N \mathbf{h}(y^i(\boldsymbol{\Phi}_t),\boldsymbol{\Phi}^*)\right\|^2\right] \\
&\overset{triangular\ inequality}{\leq} \underbrace{\frac{2\beta_t^2}{N^2}\mathbb{E}\left[\left\|\sum_{i=1}^N \mathbf{h}(\boldsymbol{\theta}_{t+1}^i,\boldsymbol{\Phi}_t) - \sum_{i=1}^N \mathbf{h}(y^i(\boldsymbol{\Phi}_t),\boldsymbol{\Phi}^*)\right\|^2\right]}_{\text{Lipschitz of } \mathbf{h}} \\
&\qquad + \frac{2\beta_t^2}{N^2}\mathbb{E}\left[\left\|\sum_{i=1}^N \mathbf{h}(y^i(\boldsymbol{\Phi}_t),\boldsymbol{\Phi}^*)\right\|^2\right] \\
&\overset{(a_1)}{\leq} \frac{2\beta_t^2L^2}{N^2}\mathbb{E}\left[2N\sum_{i=1}^N \left\|(\boldsymbol{\theta}_{t+1}^i-y^i(\boldsymbol{\Phi}_t))\right\|^2 + 2N^2\left\|(\boldsymbol{\Phi}_t-\boldsymbol{\Phi}^*)\right\|^2\right] \\
&\qquad + \frac{2\beta_t^2}{N^2}\mathbb{E}\left[\left\|\sum_{i=1}^N \mathbf{h}(y^i(\boldsymbol{\Phi}_t),\boldsymbol{\Phi}^*) - \sum_{i=1}^N \mathbf{h}(y^i(\boldsymbol{\Phi}^*),\boldsymbol{\Phi}^*) + \sum_{i=1}^N \mathbf{h}(y^i(\boldsymbol{\Phi}^*),\boldsymbol{\Phi}^*)\right\|^2\right] \\
&\leq 4\beta_t^2L^2\mathbb{E}[\|\boldsymbol{\Phi}^*-\boldsymbol{\Phi}_t\|^2] + \frac{4\beta_t^2L^2}{N}\mathbb{E}\left[\sum_{i=1}^N \|\boldsymbol{\theta}_{t+1}^i-y^i(\boldsymbol{\Phi}_t)\|^2\right] \\
&\qquad + \underbrace{\frac{4\beta_t^2}{N^2}\mathbb{E}\left[\left\|\sum_{i=1}^N \mathbf{h}(y^i(\boldsymbol{\Phi}_t),\boldsymbol{\Phi}^*) - \sum_{i=1}^N \mathbf{h}(y^i(\boldsymbol{\Phi}^*),\boldsymbol{\Phi}^*)\right\|^2\right]}_{\text{Lipschitz of } \mathbf{h},\ y^i} \\
&\qquad + \frac{4\beta_t^2}{N^2}\mathbb{E}\left[\left\|\sum_{i=1}^N \mathbf{h}(y^i(\boldsymbol{\Phi}^*),\boldsymbol{\Phi}^*)\right\|^2\right] \\
&\overset{(a_2)}{\leq} 4\beta_t^2L^2\mathbb{E}[\|\boldsymbol{\Phi}^*-\boldsymbol{\Phi}_t\|^2] + \frac{4\beta_t^2L^2}{N}\mathbb{E}\left[\sum_{i=1}^N \|\boldsymbol{\theta}_{t+1}^i-y^i(\boldsymbol{\Phi}_t)\|^2\right]
\end{aligned}$$

$$
\begin{aligned}
&+ 4\beta_t^2 L^4 \mathbb{E}\left[\|\mathbf{\Phi}_t - \mathbf{\Phi}^*\|^2\right] + \frac{4\beta_t^2}{N^2}\mathbb{E}\left[\left\|\sum_{i=1}^N \mathbf{h}(y^i(\mathbf{\Phi}^*), \mathbf{\Phi}^*) - \sum_{i=1}^N \bar{\mathbf{h}}(y^i(\mathbf{\Phi}^*), \mathbf{\Phi}^*)\right\|^2\right] \\
\overset{(a_3)}{\leq}\; & 4\beta_t^2(L^2 + L^4)\mathbb{E}[\|\mathbf{\Phi}^* - \mathbf{\Phi}_t\|^2] + \frac{4\beta_t^2 L^2}{N}\mathbb{E}\left[\sum_{i=1}^N \|\boldsymbol{\theta}_{t+1} - y^i(\mathbf{\Phi}_t)\|^2\right] + 4\beta_t^2\delta^2,
\end{aligned}
$$

where the (a_1) is due to $\|\sum_{i=1}^N \mathbf{x}_i\|^2 \leq N\sum_{i=1}^N \|\mathbf{x}_i\|^2$, (a_2) is due to the Lipschitz of functions $\mathbf{h}$ and y^i, and (a_3) holds based on the mixing time property in Definition 4.3.

□

Next, we bound $Term_2$ in the following lemma.

Lemma F.3. *We have*

$$
Term_2 \leq \beta_t(L/\alpha_t - 2\omega)\mathbb{E}[\|\mathbf{\Phi}^* - \mathbf{\Phi}_t\|^2] + \frac{\beta_t\alpha_t L}{N}\mathbb{E}\left[\sum_{i=1}^N \|\boldsymbol{\theta}_{t+1}^i - y^i(\mathbf{\Phi}_t)\|^2\right]. \tag{29}
$$

Proof. We have

$$
\begin{aligned}
Term_2 &= 2\beta_t\mathbb{E}\left[\langle \mathbf{\Phi}^* - \mathbf{\Phi}_t, \frac{-1}{N}\sum_{i=1}^N \bar{\mathbf{h}}(\boldsymbol{\theta}_{t+1}^i, \mathbf{\Phi}_t)\rangle\right] \\
&= 2\beta_t\mathbb{E}\left[\langle \mathbf{\Phi}^* - \mathbf{\Phi}_t, \frac{-1}{N}\sum_{i=1}^N \bar{\mathbf{h}}(y^i(\mathbf{\Phi}_t), \mathbf{\Phi}_t)\rangle\right] \\
&\quad + 2\beta_t\mathbb{E}\left[\langle \mathbf{\Phi}^* - \mathbf{\Phi}_t, \underbrace{\frac{1}{N}\sum_{i=1}^N \bar{\mathbf{h}}(y^i(\mathbf{\Phi}_t), \mathbf{\Phi}_t) - \bar{\mathbf{h}}(\boldsymbol{\theta}_{t+1}^i, \mathbf{\Phi}_t)}_{\text{Lipschitz of } \mathbf{h}}\rangle\right] \\
&\leq 2\beta_t\mathbb{E}\left[\langle \mathbf{\Phi}^* - \mathbf{\Phi}_t, \frac{-1}{N}\sum_{i=1}^N \bar{\mathbf{h}}(y^i(\mathbf{\Phi}_t), \mathbf{\Phi}_t)\rangle\right] + 2\beta_t L\mathbb{E}\left[\langle \mathbf{\Phi}^* - \mathbf{\Phi}_t, \frac{1}{N}\sum_{i=1}^N (y^i(\mathbf{\Phi}_t) - \boldsymbol{\theta}_{t+1}^i)\rangle\right] \\
&\overset{(b_1)}{\leq} 2\beta_t\mathbb{E}\left[\langle \mathbf{\Phi}^* - \mathbf{\Phi}_t, \frac{-1}{N}\sum_{i=1}^N \bar{\mathbf{h}}(y^i(\mathbf{\Phi}_t), \mathbf{\Phi}_t)\rangle\right] + \beta_t L/\alpha_t\mathbb{E}[\|\mathbf{\Phi}^* - \mathbf{\Phi}_t\|^2] \\
&\quad + \frac{\beta_t\alpha_t L}{N^2}\mathbb{E}\left[\|\sum_{i=1}^N (\boldsymbol{\theta}_{t+1}^i - y^i(\mathbf{\Phi}_t))\|^2\right] \\
&\overset{(b_2)}{\leq} 2\beta_t\mathbb{E}\left[\langle \mathbf{\Phi}_t - \mathbf{\Phi}^*, \frac{1}{N}\sum_{i=1}^N \bar{\mathbf{h}}(y^i(\mathbf{\Phi}_t), \mathbf{\Phi}_t)\rangle\right] + \beta_t L/\alpha_t\mathbb{E}[\|\mathbf{\Phi}^* - \mathbf{\Phi}_t\|^2] \\
&\quad + \frac{\beta_t\alpha_t L}{N}\mathbb{E}\left[\sum_{i=1}^N \|\boldsymbol{\theta}_{t+1}^i - y^i(\mathbf{\Phi}_t)\|^2\right] \\
&\leq \beta_t(L/\alpha_t - 2\omega)\mathbb{E}[\|\mathbf{\Phi}^* - \mathbf{\Phi}_t\|^2] + \frac{\beta_t\alpha_t L}{N}\mathbb{E}\left[\sum_{i=1}^N \|\boldsymbol{\theta}_{t+1}^i - y^i(\mathbf{\Phi}_t)\|^2\right],
\end{aligned}
$$

where (b_1) holds because $2\mathbf{x}^T\mathbf{y} \leq \beta\|\mathbf{x}\|^2 + 1/\beta\|\mathbf{y}\|^2, \forall\beta > 0$, (b_2) is due to $\|\sum_{i=1}^N \mathbf{x}_i\|^2 \leq N\sum_{i=1}^N \|\mathbf{x}_i\|^2$, and the last inequality is due to Assumption E.8.

□

Next, we bound $Term_3$ in the following lemmas.

Lemma F.4. *For all $t \geq \tau$ we have*

$$\begin{aligned} Term_3 \leq\ & (7\beta_t/\alpha_t + 2\beta_t\alpha_t L^2 + 6\beta_t\alpha_t\delta^2)\mathbb{E}[\|\mathbf{\Phi}_{t-\tau} - \mathbf{\Phi}_t\|^2] \\ & + (6\beta_t/\alpha_t + 6\beta_t\alpha_t\delta^2(1+L^2) + 4\beta_t\alpha_t L^2(3+4L^2))\mathbb{E}[\|\mathbf{\Phi}_t - \mathbf{\Phi}^*\|^2] \\ & + \frac{16\beta_t\alpha_t L^2 + 6\beta_t\alpha_t\delta^2}{N}\mathbb{E}\left[\sum_{i=1}^N \|\boldsymbol{\theta}^{i,*} - \boldsymbol{\theta}_{t+1}\|^2\right] + 11\beta_t\alpha_t\delta^2. \end{aligned} \tag{30}$$

Proof. We first decompose $Term_3$ as follows

$$\begin{aligned} Term_3 &= 2\beta_t\mathbb{E}\left[\langle \mathbf{\Phi}_t - \mathbf{\Phi}^*, \frac{1}{N}\sum_{i=1}^N \mathbf{h}(\boldsymbol{\theta}_{t+1}^i, \mathbf{\Phi}_t) - \bar{\mathbf{h}}(\boldsymbol{\theta}_{t+1}^i, \mathbf{\Phi}_t)\rangle\right] \\ &= 2\beta_t\mathbb{E}\left[\langle \mathbf{\Phi}_t - \mathbf{\Phi}_{t-\tau}, \frac{1}{N}\sum_{i=1}^N \mathbf{h}(\boldsymbol{\theta}_{t+1}^i, \mathbf{\Phi}_t) - \bar{\mathbf{h}}(\boldsymbol{\theta}_{t+1}^i, \mathbf{\Phi}_t)\rangle\right] \\ &\quad + 2\beta_t\mathbb{E}\left[\langle \mathbf{\Phi}_{t-\tau} - \mathbf{\Phi}^*, \frac{1}{N}\sum_{i=1}^N \mathbf{h}(\boldsymbol{\theta}_{t+1}^i, \mathbf{\Phi}_t) - \bar{\mathbf{h}}(\boldsymbol{\theta}_{t+1}^i, \mathbf{\Phi}_t)\rangle\right] \\ &= \underbrace{2\beta_t\mathbb{E}\left[\langle \mathbf{\Phi}_t - \mathbf{\Phi}_{t-\tau}, \frac{1}{N}\sum_{i=1}^N \mathbf{h}(\boldsymbol{\theta}_{t+1}^i, \mathbf{\Phi}_t) - \bar{\mathbf{h}}(\boldsymbol{\theta}_{t+1}^i, \mathbf{\Phi}_t)\rangle\right]}_{C_1} \\ &\quad + \underbrace{2\beta_t\mathbb{E}\left[\langle \mathbf{\Phi}_{t-\tau} - \mathbf{\Phi}^*, \frac{1}{N}\sum_{i=1}^N \mathbf{h}(\boldsymbol{\theta}_{t+1}^i, \mathbf{\Phi}_t) - \frac{1}{N}\sum_{i=1}^N \mathbf{h}(\boldsymbol{\theta}_{t+1}^i, \mathbf{\Phi}_{t-\tau})\rangle\right]}_{C_2} \\ &\quad + \underbrace{2\beta_t\mathbb{E}\left[\langle \mathbf{\Phi}_{t-\tau} - \mathbf{\Phi}^*, \frac{1}{N}\sum_{i=1}^N \mathbf{h}(\boldsymbol{\theta}_{t+1}^i, \mathbf{\Phi}_{t-\tau}) - \frac{1}{N}\sum_{i=1}^N \bar{\mathbf{h}}(\boldsymbol{\theta}_{t+1}^i, \mathbf{\Phi}_{t-\tau})\rangle\right]}_{C_3} \\ &\quad + \underbrace{2\beta_t\mathbb{E}\left[\langle \mathbf{\Phi}_{t-\tau} - \mathbf{\Phi}^*, \frac{1}{N}\sum_{i=1}^N \bar{\mathbf{h}}(\boldsymbol{\theta}_{t+1}^i, \mathbf{\Phi}_{t-\tau}) - \frac{1}{N}\sum_{i=1}^N \bar{\mathbf{h}}(\boldsymbol{\theta}_{t+1}^i, \mathbf{\Phi}_t)\rangle\right]}_{C_4}. \end{aligned}$$

Next, we bound C_1 as

$$\begin{aligned} C_1 &= 2\beta_t\mathbb{E}\left[\langle \mathbf{\Phi}_t - \mathbf{\Phi}_{t-\tau}, \frac{1}{N}\sum_{i=1}^N \mathbf{h}(\boldsymbol{\theta}_{t+1}^i, \mathbf{\Phi}_t) - \bar{\mathbf{h}}(\boldsymbol{\theta}_{t+1}^i, \mathbf{\Phi}_t)\rangle\right] \\ &\leq \beta_t/\alpha_t\mathbb{E}[\|\mathbf{\Phi}_t - \mathbf{\Phi}_{t-\tau}\|^2] + \beta_t\alpha_t\mathbb{E}\left[\left\|\frac{1}{N}\sum_{i=1}^N \mathbf{h}(\boldsymbol{\theta}_{t+1}^i, \mathbf{\Phi}_t) - \bar{\mathbf{h}}(\boldsymbol{\theta}_{t+1}^i, \mathbf{\Phi}_t) + \bar{h}(y^i(\mathbf{\Phi}^*), \mathbf{\Phi}^*)\right\|^2\right] \\ &\leq \beta_t/\alpha_t\mathbb{E}[\|\mathbf{\Phi}_t - \mathbf{\Phi}_{t-\tau}\|^2] + 2\beta_t\alpha_t\mathbb{E}\left[\left\|\frac{1}{N}\sum_{i=1}^N \mathbf{h}(\boldsymbol{\theta}_{t+1}^i, \mathbf{\Phi}_t)\right\|^2\right] \\ &\quad + 2\beta_t\alpha_t\mathbb{E}\left[\left\|\frac{1}{N}\sum_{i=1}^N \bar{\mathbf{h}}(y^i(\mathbf{\Phi}^*), \mathbf{\Phi}^*) - \bar{\mathbf{h}}(\boldsymbol{\theta}_{t+1}^i, \mathbf{\Phi}_t)\right\|^2\right] \\ &= \beta_t/\alpha_t\mathbb{E}[\|\mathbf{\Phi}_t - \mathbf{\Phi}_{t-\tau}\|^2] + \frac{2\beta_t\alpha_t}{N^2}\mathbb{E}\left[\left\|\sum_{i=1}^N \mathbf{h}(\boldsymbol{\theta}_{t+1}^i, \mathbf{\Phi}_t)\right\|^2\right] \end{aligned}$$

$$
\begin{aligned}
&\qquad + 2\beta_t\alpha_t\mathbb{E}\left[\left\|\frac{1}{N}\sum_{i=1}^{N}\bar{\mathbf{h}}(y^i(\mathbf{\Phi}^*),\mathbf{\Phi}^*) - \bar{\mathbf{h}}(\boldsymbol{\theta}_{t+1}^i,\mathbf{\Phi}_t)\right\|^2\right] \\
&\overset{\text{Lemma F.2}}{\leq} \beta_t/\alpha_t\mathbb{E}[\|\mathbf{\Phi}_t - \mathbf{\Phi}_{t-\tau}\|^2] + 8\beta_t\alpha_t(L^2+L^4)\mathbb{E}[\|\mathbf{\Phi}^* - \mathbf{\Phi}_t\|^2] \\
&\qquad + \frac{8\beta_t\alpha_t L^2}{N}\mathbb{E}\left[\sum_{i=1}^{N}\|\boldsymbol{\theta}_{t+1}^i - y^i(\mathbf{\Phi}_t)\|^2\right] + 8\beta_t\alpha_t\delta^2 \\
&\qquad + \underbrace{2\beta_t\alpha_t\mathbb{E}\left[\left\|\frac{1}{N}\sum_{i=1}^{N}\bar{\mathbf{h}}(y^i(\mathbf{\Phi}^*),\mathbf{\Phi}^*) - \bar{\mathbf{h}}(\boldsymbol{\theta}_{t+1}^i,\mathbf{\Phi}_t)\right\|^2\right]}_{\text{Lipschitz of }\mathbf{h}} \\
&\leq \beta_t/\alpha_t\mathbb{E}[\|\mathbf{\Phi}_t - \mathbf{\Phi}_{t-\tau}\|^2] + 8\beta_t\alpha_t(L^2+L^4)\mathbb{E}[\|\mathbf{\Phi}^* - \mathbf{\Phi}_t\|^2] \\
&\qquad + \frac{8\beta_t\alpha_t L^2}{N}\mathbb{E}\left[\sum_{i=1}^{N}\|\boldsymbol{\theta}_{t+1}^i - y^i(\mathbf{\Phi}_t)\|^2\right] + 8\beta_t\alpha_t\delta^2 \\
&\qquad + 2\beta_t\alpha_t L^2\mathbb{E}\left[\left\|\frac{1}{N}\sum_{i=1}^{N}2(\mathbf{\Phi}^* - \mathbf{\Phi}_t) + 2(\boldsymbol{\theta}_{t+1}^i - y^i(\mathbf{\Phi}^*))\right\|\right] \\
&\leq \beta_t/\alpha_t\mathbb{E}[\|\mathbf{\Phi}_t - \mathbf{\Phi}_{t-\tau}\|^2] + 8\beta_t\alpha_t(L^2+L^4)\mathbb{E}[\|\mathbf{\Phi}^* - \mathbf{\Phi}_t\|^2] \\
&\qquad + \frac{8\beta_t\alpha_t L^2}{N}\mathbb{E}\left[\sum_{i=1}^{N}\|\boldsymbol{\theta}_{t+1}^i - y^i(\mathbf{\Phi}_t)\|^2\right] + 8\beta_t\alpha_t\delta^2 \\
&\qquad + 4\beta_t\alpha_t L^2\mathbb{E}[\|\mathbf{\Phi}^* - \mathbf{\Phi}_t\|^2] + \frac{4\beta_t\alpha_t L^2}{N}\mathbb{E}\left[\sum_{i=1}^{N}\|\boldsymbol{\theta}_{t+1}^i - y^i(\mathbf{\Phi}^*)\|^2\right] \\
&= \beta_t/\alpha_t\mathbb{E}[\|\mathbf{\Phi}_t - \mathbf{\Phi}_{t-\tau}\|^2] + 8\beta_t\alpha_t(L^2+L^4)\mathbb{E}[\|\mathbf{\Phi}^* - \mathbf{\Phi}_t\|^2] \\
&\qquad + \frac{8\beta_t\alpha_t L^2}{N}\mathbb{E}\left[\sum_{i=1}^{N}\|\boldsymbol{\theta}_{t+1}^i - y^i(\mathbf{\Phi}_t)\|^2\right] + 8\beta_t\alpha_t\delta^2 + 4\beta_t\alpha_t L^2\mathbb{E}[\|\mathbf{\Phi}^* - \mathbf{\Phi}_t\|^2] \\
&\qquad + \frac{4\beta_t\alpha_t L^2}{N}\mathbb{E}\left[\sum_{i=1}^{N}\|\boldsymbol{\theta}_{t+1}^i - y^i(\mathbf{\Phi}_t) + y^i(\mathbf{\Phi}_t) - y^i(\mathbf{\Phi}^*)\|^2\right] \\
&\leq \beta_t/\alpha_t\mathbb{E}[\|\mathbf{\Phi}_t - \mathbf{\Phi}_{t-\tau}\|^2] + 8\beta_t\alpha_t(L^2+L^4)\mathbb{E}[\|\mathbf{\Phi}^* - \mathbf{\Phi}_t\|^2] \\
&\qquad + \frac{8\beta_t\alpha_t L^2}{N}\mathbb{E}\left[\sum_{i=1}^{N}\|\boldsymbol{\theta}_{t+1}^i - y^i(\mathbf{\Phi}_t)\|^2\right] + 8\beta_t\alpha_t\delta^2 \\
&\qquad + 4\beta_t\alpha_t L^2\mathbb{E}[\|\mathbf{\Phi}^* - \mathbf{\Phi}_t\|^2] + \frac{8\beta_t\alpha_t L^2}{N}\mathbb{E}\left[\sum_{i=1}^{N}\|\boldsymbol{\theta}_{t+1}^i - y^i(\mathbf{\Phi}_t)\|^2\right] \\
&\qquad + 8\beta_t\alpha_t L^4\mathbb{E}[\|\mathbf{\Phi}^* - \mathbf{\Phi}_t\|^2] \\
&= \beta_t/\alpha_t\mathbb{E}[\|\mathbf{\Phi}_t - \mathbf{\Phi}_{t-\tau}\|^2] + 4\beta_t\alpha_t L^2(3+4L^2)\mathbb{E}[\|\mathbf{\Phi}^* - \mathbf{\Phi}_t\|^2] \\
&\qquad + \frac{16\beta_t\alpha_t L^2}{N}\mathbb{E}\left[\sum_{i=1}^{N}\|\boldsymbol{\theta}_{t+1}^i - y^i(\mathbf{\Phi}_t)\|^2\right] + 8\beta_t\alpha_t\delta^2,
\end{aligned}
$$

where the last inequality is due to the Lipschitz of the function y^i.

Next, we bound C_2 as follows.

$$
C_2 = 2\beta_t\mathbb{E}\left[\langle\mathbf{\Phi}_{t-\tau} - \mathbf{\Phi}^*, \frac{1}{N}\sum_{i=1}^{N}\mathbf{h}(\boldsymbol{\theta}_{t+1}^i,\mathbf{\Phi}_t) - \frac{1}{N}\sum_{i=1}^{N}\mathbf{h}(\boldsymbol{\theta}_{t+1}^i,\mathbf{\Phi}_{t-\tau})\rangle\right]
$$

$$
\begin{aligned}
&\le \beta_t/\alpha_t\mathbb{E}[\|\boldsymbol{\Phi}_{t-\tau}-\boldsymbol{\Phi}^*\|^2]+\beta_t\alpha_t\mathbb{E}\underbrace{\left[\left\|\frac{1}{N}\sum_{i=1}^{N}\mathbf{h}(\boldsymbol{\theta}^i_{t+1},\boldsymbol{\Phi}_t)-\frac{1}{N}\sum_{i=1}^{N}\mathbf{h}(\boldsymbol{\theta}^i_{t+1},\boldsymbol{\Phi}_{t-\tau})\right\|^2\right]}_{\text{Lipschitz of }\ \mathbf{h}}\\
&\le \beta_t/\alpha_t\mathbb{E}[\|\boldsymbol{\Phi}_{t-\tau}-\boldsymbol{\Phi}^*\|^2]+\beta_t\alpha_t L^2\mathbb{E}[\|\boldsymbol{\Phi}_t-\boldsymbol{\Phi}_{t-\tau}\|^2]\\
&= \beta_t/\alpha_t\mathbb{E}[\|\boldsymbol{\Phi}_{t-\tau}-\boldsymbol{\Phi}_t+\boldsymbol{\Phi}_t-\boldsymbol{\Phi}^*\|^2]+\beta_t\alpha_t L^2\mathbb{E}[\|\boldsymbol{\Phi}_t-\boldsymbol{\Phi}_{t-\tau}\|^2]\\
&\le 2\beta_t/\alpha_t\mathbb{E}[\|\boldsymbol{\Phi}_{t-\tau}-\boldsymbol{\Phi}_t\|^2]+2\beta_t/\alpha_t\mathbb{E}[\|\boldsymbol{\Phi}_t-\boldsymbol{\Phi}^*\|^2]+\beta_t\alpha_t L^2\mathbb{E}[\|\boldsymbol{\Phi}_t-\boldsymbol{\Phi}_{t-\tau}\|^2]\\
&= (2\beta_t/\alpha_t+\beta_t\alpha_t L^2)\mathbb{E}[\|\boldsymbol{\Phi}_{t-\tau}-\boldsymbol{\Phi}_t\|^2]+2\beta_t/\alpha_t\mathbb{E}[\|\boldsymbol{\Phi}_t-\boldsymbol{\Phi}^*\|^2].
\end{aligned}
$$

Similarly, C_4 is bounded exactly same as C_2, i.e.,

$$C_4 \le (2\beta_t/\alpha_t+\beta_t\alpha_t L^2)\mathbb{E}[\|\boldsymbol{\Phi}_{t-\tau}-\boldsymbol{\Phi}_t\|^2]+2\beta_t/\alpha_t\mathbb{E}[\|\boldsymbol{\Phi}_t-\boldsymbol{\Phi}^*\|^2].$$

Next, we bound C_3 as follows.

$$
\begin{aligned}
C_3 &= 2\beta_t\mathbb{E}\left[\langle\boldsymbol{\Phi}_{t-\tau}-\boldsymbol{\Phi}^*,\frac{1}{N}\sum_{i=1}^{N}\mathbf{h}(\boldsymbol{\theta}^i_{t+1},\boldsymbol{\Phi}_{t-\tau})-\frac{1}{N}\sum_{i=1}^{N}\bar{\mathbf{h}}(\boldsymbol{\theta}^i_{t+1},\boldsymbol{\Phi}_{t-\tau})\rangle\right]\\
&\le \beta_t/\alpha_t\mathbb{E}[\|\boldsymbol{\Phi}_{t-\tau}-\boldsymbol{\Phi}^*\|^2]+\beta_t\alpha_t\frac{1}{N^2}\mathbb{E}\left[\left\|\sum_{i=1}^{N}\mathbf{h}(\boldsymbol{\theta}^i_{t+1},\boldsymbol{\Phi}_{t-\tau})-\sum_{i=1}^{N}\bar{\mathbf{h}}(\boldsymbol{\theta}^i_{t+1},\boldsymbol{\Phi}_{t-\tau})\right\|^2\right]\\
&\overset{\text{Definition }4.3}{\le} \beta_t/\alpha_t\mathbb{E}[\|\boldsymbol{\Phi}_{t-\tau}-\boldsymbol{\Phi}^*\|^2]\\
&\qquad+\beta_t\alpha_t\frac{1}{N^2}\mathbb{E}\left[\left(N\delta\left\|\boldsymbol{\Phi}_{t-\tau}-\boldsymbol{\Phi}^*\right\|+N\delta+\delta\sum_{i=1}^{N}\left\|\boldsymbol{\theta}^i_{t+1}-y^i(\boldsymbol{\Phi}^*)\right\|\right)^2\right]\\
&\le \beta_t/\alpha_t\mathbb{E}[\|\boldsymbol{\Phi}_{t-\tau}-\boldsymbol{\Phi}^*\|^2]+3\beta_t\alpha_t\delta^2\mathbb{E}\left[\|\boldsymbol{\Phi}_{t-\tau}-\boldsymbol{\Phi}^*\|^2\right]+3\beta_t\alpha_t\delta^2\\
&\qquad+\frac{3\beta_t\alpha_t\delta^2}{N}\mathbb{E}\left[\sum_{i=1}^{N}\left\|\boldsymbol{\theta}^i_{t+1}-y^i(\boldsymbol{\Phi}^*)\right\|^2\right]\\
&= \beta_t/\alpha_t\mathbb{E}[\|\boldsymbol{\Phi}_{t-\tau}-\boldsymbol{\Phi}^*\|^2]+3\beta_t\alpha_t\delta^2\mathbb{E}\left[\|\boldsymbol{\Phi}_{t-\tau}-\boldsymbol{\Phi}^*\|^2\right]+3\beta_t\alpha_t\delta^2\\
&\qquad+\frac{3\beta_t\alpha_t\delta^2}{N}\mathbb{E}\left[\sum_{i=1}^{N}\left\|\boldsymbol{\theta}^i_{t+1}-y^i(\boldsymbol{\Phi}_t)+y^i(\boldsymbol{\Phi}_t)-y^i(\boldsymbol{\Phi}^*)\right\|^2\right]\\
&\le \beta_t/\alpha_t\mathbb{E}[\|\boldsymbol{\Phi}_{t-\tau}-\boldsymbol{\Phi}^*\|^2]+3\beta_t\alpha_t\delta^2\mathbb{E}\left[\|\boldsymbol{\Phi}_{t-\tau}-\boldsymbol{\Phi}^*\|^2\right]+3\beta_t\alpha_t\delta^2\\
&\qquad+\frac{6\beta_t\alpha_t\delta^2}{N}\mathbb{E}\left[\sum_{i=1}^{N}\left\|\boldsymbol{\theta}^i_{t+1}-y^i(\boldsymbol{\Phi}_t)\right\|^2\right]+6\beta_t\alpha_t L^2\delta^2\mathbb{E}\left[\|\boldsymbol{\Phi}_t-\boldsymbol{\Phi}^*\|^2\right]\\
&\le (2\beta_t/\alpha_t+6\beta_t\alpha_t\delta^2)\mathbb{E}[\|\boldsymbol{\Phi}_{t-\tau}-\boldsymbol{\Phi}_t\|^2]+(2\beta_t/\alpha_t+6\beta_t\alpha_t\delta^2+6\beta_t\alpha_t L^2\delta^2)\mathbb{E}[\|\boldsymbol{\Phi}_t-\boldsymbol{\Phi}^*\|^2]\\
&\qquad+3\beta_t\alpha_t\delta^2+\frac{6\beta_t\alpha_t\delta^2}{N}\mathbb{E}\left[\sum_{i=1}^{N}\left\|\boldsymbol{\theta}^i_{t+1}-y^i(\boldsymbol{\Phi}_t)\right\|^2\right],
\end{aligned}
$$

where the last inequality comes from $\mathbb{E}[\|\boldsymbol{\Phi}_{t-\tau}-\boldsymbol{\Phi}^*\|^2]\le 2\mathbb{E}[\|\boldsymbol{\Phi}_{t-\tau}-\boldsymbol{\Phi}_t\|^2]+2\mathbb{E}[\|\boldsymbol{\Phi}_t-\boldsymbol{\Phi}^*\|^2]$.

Hence, we have $Term_3$ as follows

$$
\begin{aligned}
Term_3 &= C_1+C_2+C_3+C_4\\
&\le \beta_t/\alpha_t\mathbb{E}[\|\boldsymbol{\Phi}_t-\boldsymbol{\Phi}_{t-\tau}\|^2]+4\beta_t\alpha_t L^2(3+4L^2)\mathbb{E}[\|\boldsymbol{\Phi}^*-\boldsymbol{\Phi}_t\|^2]\\
&\qquad+\frac{16\beta_t\alpha_t L^2}{N}\mathbb{E}\left[\sum_{i=1}^{N}\|\boldsymbol{\theta}^i_{t+1}-y^i(\boldsymbol{\Phi}_t)\|^2\right]+8\beta_t\alpha_t\delta^2
\end{aligned}
$$

$$
\begin{aligned}
&+ (2\beta_t/\alpha_t + \beta_t\alpha_t L^2)\mathbb{E}[\|\mathbf{\Phi}_{t-\tau} - \mathbf{\Phi}_t\|^2] + 2\beta_t/\alpha_t\mathbb{E}[\|\mathbf{\Phi}_t - \mathbf{\Phi}^*\|^2] \\
&+ (2\beta_t/\alpha_t + 6\beta_t\alpha_t\delta^2)\mathbb{E}[\|\mathbf{\Phi}_{t-\tau} - \mathbf{\Phi}_t\|^2] \\
&+ (2\beta_t/\alpha_t + 6\beta_t\alpha_t\delta^2 + 6\beta_t\alpha_t L^2\delta^2)\mathbb{E}[\|\mathbf{\Phi}_t - \mathbf{\Phi}^*\|^2] \\
&+ 3\beta_t\alpha_t\delta^2 + \frac{6\beta_t\alpha_t\delta^2}{N}\mathbb{E}\left[\sum_{i=1}^N \left\|\boldsymbol{\theta}_{t+1}^i - y^i(\mathbf{\Phi}_t)\right\|^2\right] \\
&+ (2\beta_t/\alpha_t + \beta_t\alpha_t L^2)\mathbb{E}[\|\mathbf{\Phi}_{t-\tau} - \mathbf{\Phi}_t\|^2] + 2\beta_t/\alpha_t\mathbb{E}[\|\mathbf{\Phi}_t - \mathbf{\Phi}^*\|^2] \\
\leq\ & (7\beta_t/\alpha_t + 2\beta_t\alpha_t L^2 + 6\beta_t\alpha_t\delta^2)\mathbb{E}[\|\mathbf{\Phi}_{t-\tau} - \mathbf{\Phi}_t\|^2] \\
&+ (6\beta_t/\alpha_t + 6\beta_t\alpha_t\delta^2(1+L^2) + 4\beta_t\alpha_t L^2(3+4L^2))\mathbb{E}[\|\mathbf{\Phi}_t - \mathbf{\Phi}^*\|^2] \\
&+ \frac{16\beta_t\alpha_t L^2 + 6\beta_t\alpha_t\delta^2}{N}\mathbb{E}\left[\sum_{i=1}^N \|y^i(\mathbf{\Phi}_t) - \boldsymbol{\theta}_{t+1}^i\|^2\right] + 11\beta_t\alpha_t\delta^2,
\end{aligned}
$$

which completes the proof. □

To bound $Term_3$, we need to bound $\mathbb{E}[\|\mathbf{\Phi}_t - \mathbf{\Phi}_{t-\tau}\|^2]$, which is shown in the following lemma.

Lemma F.5. *we have* $\forall t \geq 2\tau$

$$\mathbb{E}[\|\mathbf{\Phi}_t - \mathbf{\Phi}_{t-\tau}\|^2] \leq 4\tau^2\beta_0^2/\alpha_0^2\mathbb{E}[\|\mathbf{\Phi}^* - \mathbf{\Phi}_t\|^2] + 8\beta_0^2L^2B^2\tau^2 + 8\beta_0^2\delta^2\tau^2. \tag{31}$$

Proof. The proof follows similar procedures of proof for Lemma 3 in Dal Fabbro et al. (2023). Starting with

$$
\begin{aligned}
\|\mathbf{\Phi}^* - \mathbf{\Phi}_{t+1}\|^2 &= \|\mathbf{\Phi}^* - \mathbf{\Phi}_t\|^2 + \frac{\beta_t^2}{N^2}\left\|\sum_{i=1}^N \mathbf{h}(\boldsymbol{\theta}_{t+1}^i, \mathbf{\Phi}_t)\right\|^2 - 2\beta_t\langle\mathbf{\Phi}^* - \mathbf{\Phi}_t, \frac{1}{N}\sum_{i=1}^N \mathbf{h}(\boldsymbol{\theta}_{t+1}^i, \mathbf{\Phi}_t)\rangle \\
&\leq (1 + \beta_t/\alpha_0)\|\mathbf{\Phi}^* - \mathbf{\Phi}_t\|^2 + \frac{(\beta_t\alpha_0 + \beta_t^2)}{N^2}\left\|\sum_{i=1}^N \mathbf{h}_t^i(\boldsymbol{\theta}_{t+1}^i, \mathbf{\Phi}_t)\right\|^2 \\
&\leq (1 + \beta_t/\alpha_0)\|\mathbf{\Phi}^* - \mathbf{\Phi}_t\|^2 + \frac{2\beta_t\alpha_0}{N^2}\left\|\sum_{i=1}^N \mathbf{h}_t^i(\boldsymbol{\theta}_{t+1}^i, \mathbf{\Phi}_t)\right\|^2,
\end{aligned} \tag{32}
$$

where the first inequality holds due to $2\mathbf{x}^T\mathbf{y} \leq \gamma\|\mathbf{x}\|^2 + 1/\gamma\|\mathbf{y}\|^2, \forall\gamma > 0$, and the second inequality holds since $\beta_t\alpha_0 \geq \beta_t^2$. We then have the following inequality according to Lemma F.2,

$$
\begin{aligned}
\mathbb{E}\left[\|\mathbf{\Phi}^* - \mathbf{\Phi}_{t+1}\|^2\right] &\leq (1 + \beta_t/\alpha_0 + 8\beta_t\alpha_0L^2(1+L^2))\mathbb{E}\left[\|\mathbf{\Phi}^* - \mathbf{\Phi}_t\|^2\right] \\
&\quad + \frac{8\beta_t\alpha_0L^2}{N}\mathbb{E}\left[\sum_{i=1}^N \|\boldsymbol{\theta}_{t+1} - y^i(\mathbf{\Phi}_t)\|^2\right] + 8\beta_t\alpha_0\delta^2 \\
&\leq (1 + \beta_t/\alpha_0 + 8\beta_t\alpha_0L^2(1+L^2))\mathbb{E}\left[\|\mathbf{\Phi}^* - \mathbf{\Phi}_t\|^2\right] + 8\beta_t\alpha_0(L^2B^2 + \delta^2).
\end{aligned} \tag{33}
$$

By letting $\alpha_0 \leq \frac{1}{2L\sqrt{2(1+L^2)}}$, we have $\beta_t/\alpha_0 \geq 8\beta_t\alpha_0L^2(1+L^2)$, and hence

$$\mathbb{E}\left[\|\mathbf{\Phi}^* - \mathbf{\Phi}_{t+1}\|^2\right] \leq (1 + 2\beta_0/\alpha_0)\mathbb{E}\left[\|\mathbf{\Phi}^* - \mathbf{\Phi}_t\|^2\right] + 8\beta_0\alpha_0(L^2B^2 + \delta^2). \tag{34}$$

Therefore, for all t' such that $t - \tau \leq t' \leq t$,

$$\mathbb{E}[\|\mathbf{\Phi}^* - \mathbf{\Phi}_{t'}\|^2] \leq (1 + 2\beta_0/\alpha_0)^\tau\mathbb{E}[\|\mathbf{\Phi}^* - \mathbf{\Phi}_{t-\tau}\|^2] + 8\beta_0\alpha_0(L^2B^2 + \delta^2)\sum_{\ell=0}^{\tau-1}(1 + 2\beta_0/\alpha_0)^\ell. \tag{35}$$

Using the fact that $(1+x) \leq e^x$ Dal Fabbro et al. (2023), if we let $\beta_0/\alpha_0 \leq \frac{1}{8\tau}$, we have

$$(1 + 2\beta_0/\alpha_0)^\ell \leq (1 + 2\beta_0/\alpha_0)^\tau \leq e^{0.25} \leq 2,$$

and

$$\sum_{\ell=0}^{\tau-1}(1+32\beta^2)^\ell \leq 2\tau.$$

Hence, we have

$$\mathbb{E}[\|\mathbf{\Phi}^* - \mathbf{\Phi}_{t'}\|^2] \leq 2\mathbb{E}[\|\mathbf{\Phi}^* - \mathbf{\Phi}_{t-\tau}\|^2] + 16\beta_0\alpha_0\tau(L^2B^2+\delta^2).$$

Since $\|\mathbf{\Phi}_t - \mathbf{\Phi}_{t-\tau}\|^2 \leq \tau\sum_{\ell=t-\tau}^{t-1}\|\mathbf{\Phi}_{\ell+1} - \mathbf{\Phi}_\ell\|^2 = \tau\frac{\beta^2}{N^2}\sum_{\ell=t-\tau}^{t-1}\|\sum_{i=1}^N \mathbf{h}_\ell^i(\boldsymbol{\theta}_{\ell+1}^i, \mathbf{\Phi}_\ell)\|^2$, when $t \geq 2\tau$, we have $\ell \geq \tau$ and thus

$$\begin{aligned}
&\mathbb{E}[\|\mathbf{\Phi}_t - \mathbf{\Phi}_{t-\tau}\|^2]\\
&\leq \tau\frac{\beta^2}{N^2}\sum_{\ell=t-\tau}^{t-1}\|\sum_{i=1}^N \mathbf{h}_\ell^i(\boldsymbol{\theta}_{\ell+1}^i, \mathbf{\Phi}_\ell)\|^2\\
&\leq \tau\sum_{\ell=t-\tau}^{t-1}((4\beta_0^2(L^2+L^4)\mathbb{E}[\|\mathbf{\Phi}^* - \mathbf{\Phi}_\ell\|^2] + 4\beta_0^2L^2B^2\tau^2 + 4\beta_0^2\delta^2\tau^2\\
&\leq 4\beta_0^2(L^2+L^4)\tau^2(2\mathbb{E}[\|\mathbf{\Phi}^* - \mathbf{\Phi}_{t-\tau}\|^2] + 16\beta_0\alpha_0\tau(L^2B^2+\delta^2)) + 4\beta_0^2L^2B^2\tau^2 + 4\beta_0^2\delta^2\tau^2\\
&= 8\beta_0^2(L^2+L^4)\tau^2\mathbb{E}[\|\mathbf{\Phi}^* - \mathbf{\Phi}_{t-\tau}\|^2] + 4\beta_0^2L^2B^2\tau^2 + 4\beta_0^2\delta^2\tau^2\\
&\leq \tau^2\beta_0^2/\alpha_0^2\mathbb{E}[\|\mathbf{\Phi}^* - \mathbf{\Phi}_{t-\tau}\|^2] + 4\beta_0^2L^2B^2\tau^2 + 4\beta_0^2\delta^2\tau^2\\
&\leq 2\tau^2\beta_0^2/\alpha_0^2\mathbb{E}[\|\mathbf{\Phi}^* - \mathbf{\Phi}_t\|^2] + 2\tau^2\beta_0^2/\alpha_0^2\mathbb{E}[\|\mathbf{\Phi}_t - \mathbf{\Phi}_{t-\tau}\|^2] + 4\beta_0^2L^2B^2\tau^2 + 4\beta_0^2\delta^2\tau^2.
\end{aligned}$$

Since $2\tau^2\beta_0^2/\alpha_0^2 \leq 1/2$ when $\beta_0/\alpha_0 \leq \frac{1}{8\tau}$, we have

$$\mathbb{E}[\|\mathbf{\Phi}_t - \mathbf{\Phi}_{t-\tau}\|^2] \leq 4\tau^2\beta_0^2/\alpha_0^2\mathbb{E}[\|\mathbf{\Phi}^* - \mathbf{\Phi}_t\|^2] + 8\beta_0^2L^2B^2\tau^2 + 8\beta_0^2\delta^2\tau^2.$$

This completes the proof. □

Lemma F.6. *$Term_3$ is bounded as follows*

$$\begin{aligned}
Term_3 \leq\ &(7\beta_t/\alpha_t + 2\beta_t\alpha_tL^2 + 6\beta_t\alpha_t\delta^2)(4\tau^2\beta_0^2/\alpha_0^2\mathbb{E}[\|\mathbf{\Phi}^* - \mathbf{\Phi}_t\|^2] + 8\beta_0^2L^2B^2\tau^2 + 8\beta_0^2\delta^2\tau^2)\\
&+ (6\beta_t/\alpha_t + 6\beta_t\alpha_t\delta^2(1+L^2) + 4\beta_t\alpha_tL^2(3+4L^2))\mathbb{E}[\|\mathbf{\Phi}_t - \mathbf{\Phi}^*\|^2]\\
&+ \frac{16\beta_t\alpha_tL^2 + 6\beta_t\alpha_t\delta^2}{N}\mathbb{E}\left[\sum_{i=1}^N\|\boldsymbol{\theta}^{i,*} - \boldsymbol{\theta}_{t+1}\|^2\right] + 11\beta_t\alpha_t\delta^2.
\end{aligned}$$

Proof. Substituting the bound of $\mathbb{E}[\|\mathbf{\Phi}_t - \mathbf{\Phi}_{t-\tau}\|^2]$ in (31) into $Term_3$ in Lemma F.4 yield the final results. □

Provided $Term_1$ in Lemma F.2, $Term_2$ in Lemma F.3, and $Term_3$ in Lemma F.6, we have the following lemma to characterize the drift between $\mathbf{\Phi}_{t+1}$ and $\mathbf{\Phi}_t$.

Lemma F.7. *For $t \geq 2\tau$, the following holds*

$$\begin{aligned}
&\mathbb{E}[\|\mathbf{\Phi}^* - \mathbf{\Phi}_{t+1}\|^2]\\
&\leq (1 + 4\beta_t^2(L^2+L^4) + (7\beta_t/\alpha_t + 2\beta_t\alpha_tL^2 + 6\beta_t\alpha_t\delta^2)4\tau^2\beta_0^2/\alpha_0^2\\
&\qquad + (6\beta_t/\alpha_t + 6\beta_t\alpha_t\delta^2(1+L^2) + 4\beta_t\alpha_tL^2(3+4L^2)) + \beta_t(L/\alpha_t - 2\omega))\mathbb{E}[\|\mathbf{\Phi}_t - \mathbf{\Phi}^*\|^2]\\
&\qquad + \frac{4\beta_t^2L^2 + \beta_t\alpha_tL + 16\beta_t\alpha_tL^2 + 6\beta_t\alpha_t\delta^2}{N}\mathbb{E}\left[\sum_{i=1}^N\|\boldsymbol{\theta}^{i,*} - \boldsymbol{\theta}_{t+1}\|^2\right]\\
&\qquad + (7\beta_t/\alpha_t + 2\beta_t\alpha_tL^2 + 6\beta_t\alpha_t\delta^2)(8\beta_0^2L^2B^2\tau^2 + 8\beta_0^2\delta^2\tau^2) + 4\beta_t^2\delta^2 + 11\beta_t\alpha_t\delta^2.
\end{aligned}$$

Proof. Substituting $Term_1, Term_2$ and $Term_3$ back into Lemma F.1, we have

$$
\begin{aligned}
&\mathbb{E}[\|\boldsymbol{\Phi}^* - \boldsymbol{\Phi}_{t+1}\|^2] \\
&\leq \mathbb{E}[\|\boldsymbol{\Phi}^* - \boldsymbol{\Phi}_t\|^2] + 4\beta_t^2(L^2 + L^4)\mathbb{E}[\|\boldsymbol{\Phi}^* - \boldsymbol{\Phi}_t\|^2] + \frac{4\beta_t^2 L^2}{N}\mathbb{E}\left[\sum_{i=1}^N \|\boldsymbol{\theta}_{t+1} - y^i(\boldsymbol{\Phi}_t)\|^2\right] + 4\beta_t^2\delta^2 \\
&\qquad + \beta_t(L/\alpha_t - 2\omega)\mathbb{E}[\|\boldsymbol{\Phi}^* - \boldsymbol{\Phi}_t\|^2] + \frac{\beta_t\alpha_t L}{N}\mathbb{E}\left[\sum_{i=1}^N \|\boldsymbol{\theta}_{t+1}^i - y^i(\boldsymbol{\Phi}_t)\|^2\right] \\
&\qquad + (7\beta_t/\alpha_t + 2\beta_t\alpha_t L^2 + 6\beta_t\alpha_t\delta^2)(4\tau^2\beta_0^2/\alpha_0^2\mathbb{E}[\|\boldsymbol{\Phi}^* - \boldsymbol{\Phi}_t\|^2] + 8\beta_0^2 L^2 B^2\tau^2 + 8\beta_0^2\delta^2\tau^2) \\
&\qquad + (6\beta_t/\alpha_t + 6\beta_t\alpha_t\delta^2(1 + L^2) + 4\beta_t\alpha_t L^2(3 + 4L^2))\mathbb{E}[\|\boldsymbol{\Phi}_t - \boldsymbol{\Phi}^*\|^2] \\
&\qquad + \frac{16\beta_t\alpha_t L^2 + 6\beta_t\alpha_t\delta^2}{N}\mathbb{E}\left[\sum_{i=1}^N \|\boldsymbol{\theta}^{i,*} - \boldsymbol{\theta}_{t+1}\|^2\right] + 11\beta_t\alpha_t\delta^2 \\
&= (1 + 4\beta_t^2(L^2 + L^4) + (7\beta_t/\alpha_t + 2\beta_t\alpha_t L^2 + 6\beta_t\alpha_t\delta^2)4\tau^2\beta_0^2/\alpha_0^2 \\
&\qquad + (6\beta_t/\alpha_t + 6\beta_t\alpha_t\delta^2(1 + L^2) + 4\beta_t\alpha_t L^2(3 + 4L^2)) + \beta_t(L/\alpha_t - 2\omega))\mathbb{E}[\|\boldsymbol{\Phi}_t - \boldsymbol{\Phi}^*\|^2] \\
&\qquad + \frac{4\beta_t^2 L^2 + \beta_t\alpha_t L + 16\beta_t\alpha_t L^2 + 6\beta_t\alpha_t\delta^2}{N}\mathbb{E}\left[\sum_{i=1}^N \|\boldsymbol{\theta}^{i,*} - \boldsymbol{\theta}_{t+1}\|^2\right] \\
&\qquad + (7\beta_t/\alpha_t + 2\beta_t\alpha_t L^2 + 6\beta_t\alpha_t\delta^2)(8\beta_0^2 L^2 B^2\tau^2 + 8\beta_0^2\delta^2\tau^2) + 4\beta_t^2\delta^2 + 11\beta_t\alpha_t\delta^2.
\end{aligned}
$$

This completes the proof.

□

F.1.2 Drift of $\boldsymbol{\theta}_t^i, \forall i$.

Next, we characterize the drift between $\boldsymbol{\theta}_{t+1}^i$ and $\boldsymbol{\theta}_t^i$.

Lemma F.8. *The drift between $\boldsymbol{\theta}_{t+1}^i$ and $\boldsymbol{\theta}_t^i, \forall i$ is given by*

$$
\begin{aligned}
\mathbb{E}[\|\boldsymbol{\theta}_{t+1}^i - y^i(\boldsymbol{\Phi}_t)\|^2] &= \underbrace{\mathbb{E}\left[\left\|\boldsymbol{\theta}_t^i - y^i(\boldsymbol{\Phi}_{t-1}) + \alpha_t\sum_{k=1}^K \mathbf{g}(\boldsymbol{\theta}_{t,k-1}^i, \boldsymbol{\Phi}_t)\right\|^2\right]}_{Term_4} + \underbrace{\mathbb{E}\left[\left\|y^i(\boldsymbol{\Phi}_{t-1}) - y^i(\boldsymbol{\Phi}_t)\right\|^2\right]}_{Term_5} \\
&\quad + \underbrace{2\mathbb{E}\left[\left\langle\boldsymbol{\theta}_t^i - y^i(\boldsymbol{\Phi}_{t-1}) + \alpha_t\sum_{k=1}^K \mathbf{g}(\boldsymbol{\theta}_{t,k-1}^i, \boldsymbol{\Phi}_t), y^i(\boldsymbol{\Phi}_{t-1}) - y^i(\boldsymbol{\Phi}_t)\right\rangle\right]}_{Term_6}. \qquad (36)
\end{aligned}
$$

Proof. According to the update of $\boldsymbol{\theta}_t^i$ in (7), we have

$$
\begin{aligned}
\mathbb{E}[\|\boldsymbol{\theta}_{t+1}^i - y^i(\boldsymbol{\Phi}_t)\|^2] &= \mathbb{E}\left[\left\|\boldsymbol{\theta}_t^i - y^i(\boldsymbol{\Phi}_{t-1}) + \alpha_t\sum_{k=1}^K \mathbf{g}(\boldsymbol{\theta}_{t,k-1}^i, \boldsymbol{\Phi}_t) + y^i(\boldsymbol{\Phi}_{t-1}) - y^i(\boldsymbol{\Phi}_t)\right\|^2\right] \\
&= \underbrace{\mathbb{E}\left[\left\|\boldsymbol{\theta}_t^i - y^i(\boldsymbol{\Phi}_{t-1}) + \alpha_t\sum_{k=1}^K \mathbf{g}(\boldsymbol{\theta}_{t,k-1}^i, \boldsymbol{\Phi}_t)\right\|^2\right]}_{Term_4} + \underbrace{\mathbb{E}\left[\left\|y^i(\boldsymbol{\Phi}_{t-1}) - y^i(\boldsymbol{\Phi}_t)\right\|^2\right]}_{Term_5} \\
&\quad + \underbrace{2\mathbb{E}\left[\left\langle\boldsymbol{\theta}_t^i - y^i(\boldsymbol{\Phi}_{t-1}) + \alpha_t\sum_{k=1}^K \mathbf{g}(\boldsymbol{\theta}_{t,k-1}^i, \boldsymbol{\Phi}_t), y^i(\boldsymbol{\Phi}_{t-1}) - y^i(\boldsymbol{\Phi}_t)\right\rangle\right]}_{Term_6}, \qquad (37)
\end{aligned}
$$

where the second inequality holds due to $\|\mathbf{x}+\mathbf{y}\|^2 = \|\mathbf{x}\|^2 + \|\mathbf{y}\|^2 + 2\langle \mathbf{x}, \mathbf{y}\rangle$.

□

We next analyze each term in (37). **First, we bound $Term_4$ in the following lemma.**

Lemma F.9. *With $t \geq \tau$, we have $Term_4$ bounded as*

$$\begin{aligned} Term_4 \leq\ & (1 + 2\beta_{t-1}/\alpha_t - 2\alpha_t K\omega)\mathbb{E}\left[\left\|\boldsymbol{\theta}_t^i - y^i(\boldsymbol{\Phi}_{t-1})\right\|^2\right] \\ & + (12\alpha_t^2\delta^2K^2 + 6K^2\delta^2\alpha_t^3/\beta_{t-1})\mathbb{E}[\|\boldsymbol{\Phi}_{t-1} - \boldsymbol{\Phi}^*\|^2] \\ & + (12\alpha_t^2\delta^2K^2 + 2L^2\alpha_t^3/\beta_{t-1} + 6K^2\delta^2\alpha_t^3/\beta_{t-1})\mathbb{E}[\|\boldsymbol{\Phi}_t - \boldsymbol{\Phi}_{t-1}\|^2] \\ & + 6\alpha_t^2\delta^2K^2(1+B^2) + 2\alpha_t^2K^2L^2B^2 + 2L^2K^2B^2\alpha_t^3/\beta_{t-1} + \alpha_t^3/\beta_{t-1}(3K^2B^2 + 3K^2\delta^2). \quad (38) \end{aligned}$$

Proof. According to the definition of $Term_4$, we have

$$\begin{aligned} Term_4 &= \mathbb{E}\left[\left\|\boldsymbol{\theta}_t^i - y^i(\boldsymbol{\Phi}_{t-1}) + \alpha_t\sum_{k=1}^K \mathbf{g}(\boldsymbol{\theta}_{t,k-1}^i, \boldsymbol{\Phi}_t)\right\|^2\right] \\ &= \mathbb{E}\left[\left\|\boldsymbol{\theta}_t^i - y^i(\boldsymbol{\Phi}_{t-1})\right\|^2\right] + \alpha_t^2\mathbb{E}\left[\left\|\sum_{k=1}^K \mathbf{g}(\boldsymbol{\theta}_{t,k-1}^i, \boldsymbol{\Phi}_t)\right\|^2\right] \\ &\qquad + 2\alpha_t\left\langle \boldsymbol{\theta}_t^i - y^i(\boldsymbol{\Phi}_{t-1}), \sum_{k=1}^K \mathbf{g}(\boldsymbol{\theta}_{t,k-1}^i, \boldsymbol{\Phi}_t)\right\rangle \\ &= \mathbb{E}\left[\left\|\boldsymbol{\theta}_t^i - y^i(\boldsymbol{\Phi}_{t-1})\right\|^2\right] + 2\alpha_t\mathbb{E}\left[\left\langle \boldsymbol{\theta}_t^i - y^i(\boldsymbol{\Phi}_{t-1}), \sum_{k=1}^K \mathbf{g}(\boldsymbol{\theta}_{t,k-1}^i, \boldsymbol{\Phi}_t)\right\rangle\right] \\ &+ \alpha_t^2\mathbb{E}\left[\left\|\sum_{k=1}^K \mathbf{g}(\boldsymbol{\theta}_{t,k-1}^i, \boldsymbol{\Phi}_t) - \sum_{k=1}^K \bar{\mathbf{g}}(\boldsymbol{\theta}_{t,k-1}^i, \boldsymbol{\Phi}_t) + \sum_{k=1}^K \bar{\mathbf{g}}(\boldsymbol{\theta}_{t,k-1}^i, \boldsymbol{\Phi}_t) - \sum_{k=1}^K \bar{\mathbf{g}}(y^i(\boldsymbol{\Phi}_t), \boldsymbol{\Phi}_t)\right\|^2\right] \\ &\leq \mathbb{E}\left[\left\|\boldsymbol{\theta}_t^i - y^i(\boldsymbol{\Phi}_{t-1})\right\|^2\right] + 2\alpha_t^2\underbrace{\mathbb{E}\left[\left\|\sum_{k=1}^K \mathbf{g}(\boldsymbol{\theta}_{t,k-1}^i, \boldsymbol{\Phi}_t) - \sum_{k=1}^K \bar{\mathbf{g}}(\boldsymbol{\theta}_{t,k-1}^i, \boldsymbol{\Phi}_t)\right\|^2\right]}_{\text{Mixing time property in Definition 4.3}} \\ &\qquad + 2\alpha_t^2\mathbb{E}\left[\left\|\sum_{k=1}^K \bar{\mathbf{g}}(\boldsymbol{\theta}_{t,k-1}^i, \boldsymbol{\Phi}_t) - \sum_{k=1}^K \bar{\mathbf{g}}(y^i(\boldsymbol{\Phi}_t), \boldsymbol{\Phi}_t)\right\|^2\right] \\ &\qquad + 2\alpha_t\mathbb{E}\left[\left\langle \boldsymbol{\theta}_t^i - y^i(\boldsymbol{\Phi}_{t-1}), \sum_{k=1}^K \mathbf{g}(\boldsymbol{\theta}_{t,k-1}^i, \boldsymbol{\Phi}_t)\right\rangle\right] \\ &\leq \mathbb{E}\left[\left\|\boldsymbol{\theta}_t^i - y^i(\boldsymbol{\Phi}_{t-1})\right\|^2\right] + 6\alpha_t^2\delta^2K^2\mathbb{E}\left[\|\boldsymbol{\Phi}_t - \boldsymbol{\Phi}^*\|^2\right] + 6\alpha_t^2\delta^2K^2(1+B^2) + 2\alpha_t^2K^2L^2B^2 \\ &\qquad + \underbrace{2\alpha_t\mathbb{E}\left[\left\langle \boldsymbol{\theta}_t^i - y^i(\boldsymbol{\Phi}_{t-1}), \sum_{k=1}^K \bar{\mathbf{g}}(\boldsymbol{\theta}_{t,k-1}^i, \boldsymbol{\Phi}_t)\right\rangle\right]}_{Term_{41}} \\ &\qquad + \underbrace{2\alpha_t\mathbb{E}\left[\left\langle \boldsymbol{\theta}_t^i - y^i(\boldsymbol{\Phi}_{t-1}), \sum_{k=1}^K \mathbf{g}(\boldsymbol{\theta}_{t,k-1}^i, \boldsymbol{\Phi}_t) - \sum_{k=1}^K \bar{\mathbf{g}}(\boldsymbol{\theta}_{t,k-1}^i, \boldsymbol{\Phi}_t)\right\rangle\right]}_{Term_{42}}, \end{aligned} \quad (39)$$

where the first inequality holds due to the fact that $\|\mathbf{x}+\mathbf{y}\|^2 \leq 2\|\mathbf{x}\|^2 + 2\|\mathbf{y}\|^2$, and the second inequality is due to the mixing time property of function $\mathbf{g}$ as in Definition 4.3.

Next, we bound $Term_{41}$ as

$$\begin{aligned}
Term_{41} &= 2\alpha_t \mathbb{E}\left[\left\langle \boldsymbol{\theta}_t^i - y^i(\boldsymbol{\Phi}_{t-1}), \sum_{k=1}^{K} \bar{\mathbf{g}}(\boldsymbol{\theta}_{t,k-1}^i, \boldsymbol{\Phi}_t) \right\rangle\right] \\
&= 2\alpha_t \mathbb{E}\left[\left\langle \boldsymbol{\theta}_t^i - y^i(\boldsymbol{\Phi}_{t-1}), \sum_{k=1}^{K} \bar{\mathbf{g}}(\boldsymbol{\theta}_t^i, \boldsymbol{\Phi}_{t-1}) \right\rangle\right] \\
&\quad + 2\alpha_t \mathbb{E}\left[\left\langle \boldsymbol{\theta}_t^i - y^i(\boldsymbol{\Phi}_{t-1}), \sum_{k=1}^{K} \bar{\mathbf{g}}(\boldsymbol{\theta}_{t,k-1}^i, \boldsymbol{\Phi}_t) - \sum_{k=1}^{K} \bar{\mathbf{g}}(\boldsymbol{\theta}_t^i, \boldsymbol{\Phi}_{t-1}) \right\rangle\right] \\
&\leq -2\alpha_t K \omega \mathbb{E}\left[\|\boldsymbol{\theta}_t^i - y^i(\boldsymbol{\Phi}_{t-1})\|^2\right] \\
&\quad + 2\alpha_t \mathbb{E}\left[\left\langle \boldsymbol{\theta}_t^i - y^i(\boldsymbol{\Phi}_{t-1}), \sum_{k=1}^{K} \bar{\mathbf{g}}(\boldsymbol{\theta}_{t,k-1}^i, \boldsymbol{\Phi}_t) - \sum_{k=1}^{K} \bar{\mathbf{g}}(\boldsymbol{\theta}_t^i, \boldsymbol{\Phi}_{t-1}) \right\rangle\right] \\
&\leq -2\alpha_t K \omega \mathbb{E}\left[\|\boldsymbol{\theta}_t^i - y^i(\boldsymbol{\Phi}_{t-1})\|^2\right] + \beta_{t-1}/\alpha_t \mathbb{E}\left[\left\|\boldsymbol{\theta}_t^i - y^i(\boldsymbol{\Phi}_{t-1})\right\|^2\right] \\
&\quad + \alpha_t^3/\beta_{t-1} \mathbb{E}\left[\left\| \sum_{k=1}^{K} \bar{\mathbf{g}}(\boldsymbol{\theta}_{t,k-1}^i, \boldsymbol{\Phi}_t) - \sum_{k=1}^{K} \bar{\mathbf{g}}(\boldsymbol{\theta}_t^i, \boldsymbol{\Phi}_{t-1}) \right\|^2\right]. \qquad (40)
\end{aligned}$$

In particular, we can bound $\mathbb{E}\left[\left\|\sum_{k=1}^{K} \bar{\mathbf{g}}(\boldsymbol{\theta}_{t,k-1}^i, \boldsymbol{\Phi}_t) - \sum_{k=1}^{K} \bar{\mathbf{g}}(\boldsymbol{\theta}_t^i, \boldsymbol{\Phi}_{t-1})\right\|^2\right]$ as

$$\begin{aligned}
&\mathbb{E}\left[\left\| \sum_{k=1}^{K} \bar{\mathbf{g}}(\boldsymbol{\theta}_{t,k-1}^i, \boldsymbol{\Phi}_t) - \sum_{k=1}^{K} \bar{\mathbf{g}}(\boldsymbol{\theta}_t^i, \boldsymbol{\Phi}_{t-1}) \right\|^2\right] \\
&\quad \leq 2L^2 \mathbb{E}\left[\left\|\boldsymbol{\Phi}_t - \boldsymbol{\Phi}_{t-1}\right\|^2\right] + 2L^2 \mathbb{E}\left[\left\| \sum_{k=1}^{K} \boldsymbol{\theta}_{t,k-1} - \boldsymbol{\theta}_t \right\|^2\right] \\
&\quad \leq 2L^2 \mathbb{E}\left[\left\|\boldsymbol{\Phi}_t - \boldsymbol{\Phi}_{t-1}\right\|^2\right] + 2L^2 K \mathbb{E}\left[\sum_{k=1}^{K} \|\boldsymbol{\theta}_{t,k-1} - \boldsymbol{\theta}_t\|^2\right] \\
&\quad \leq 2L^2 \mathbb{E}\left[\left\|\boldsymbol{\Phi}_t - \boldsymbol{\Phi}_{t-1}\right\|^2\right] + 2L^2 K^2 B^2. \qquad (41)
\end{aligned}$$

Substituting (41) back into (40), we have $Term_{41}$ bounded as

$$\begin{aligned}
Term_{41} &\leq -2\alpha_t K \omega \mathbb{E}\left[\|\boldsymbol{\theta}_t^i - y^i(\boldsymbol{\Phi}_{t-1})\|^2\right] + \beta_{t-1}/\alpha_t \mathbb{E}\left[\left\|\boldsymbol{\theta}_t^i - y^i(\boldsymbol{\Phi}_{t-1})\right\|^2\right] \\
&\quad + \alpha_t^3/\beta_{t-1}(2L^2 \mathbb{E}\left[\|\boldsymbol{\Phi}_t - \boldsymbol{\Phi}_{t-1}\|^2\right] + 2L^2K^2B^2). \qquad (42)
\end{aligned}$$

We next bound $Term_{42}$ as

$$\begin{aligned}
Term_{42} &= 2\alpha_t \mathbb{E}\left[\left\langle \boldsymbol{\theta}_t^i - y^i(\boldsymbol{\Phi}_{t-1}), \sum_{k=1}^{K} \mathbf{g}(\boldsymbol{\theta}_{t,k-1}^i, \boldsymbol{\Phi}_t) - \sum_{k=1}^{K} \bar{\mathbf{g}}(\boldsymbol{\theta}_{t,k-1}^i, \boldsymbol{\Phi}_t) \right\rangle\right] \\
&\leq \beta_{t-1}/\alpha_t \mathbb{E}\left[\left\|\boldsymbol{\theta}_t^i - y^i(\boldsymbol{\Phi}_{t-1})\right\|^2\right] + \alpha_t^3/\beta_{t-1}(3K^2B^2 + 3K^2\delta^2 + 3K^2\delta^2\mathbb{E}[\|\boldsymbol{\Phi}_t - \boldsymbol{\Phi}^*\|^2]) \\
&\leq \beta_{t-1}/\alpha_t \mathbb{E}\left[\left\|\boldsymbol{\theta}_t^i - y^i(\boldsymbol{\Phi}_{t-1})\right\|^2\right] + \alpha_t^3/\beta_{t-1}(3K^2B^2 + 3K^2\delta^2) \\
&\quad + 6K^2\delta^2\alpha_t^3/\beta_{t-1}\mathbb{E}[\|\boldsymbol{\Phi}_{t-1} - \boldsymbol{\Phi}^*\|^2] + 6K^2\delta^2\alpha_t^3/\beta_{t-1}\mathbb{E}[\|\boldsymbol{\Phi}_t - \boldsymbol{\Phi}_{t-1}\|^2] \qquad (43)
\end{aligned}$$

Substituting $Term_{41}$ and $Term_{42}$ back into (39), we get the final result

$$Term_4 \leq \mathbb{E}\left[\left\|\boldsymbol{\theta}_t^i - y^i(\boldsymbol{\Phi}_{t-1})\right\|^2\right] + 6\alpha_t^2\delta^2K^2\mathbb{E}\left[\|\boldsymbol{\Phi}_t - \boldsymbol{\Phi}^*\|^2\right] + 6\alpha_t^2\delta^2K^2(1+B^2) + 2\alpha_t^2K^2L^2B^2$$

$$\begin{aligned}
&\quad + Term_{41} + Term_{42} \\
&= \mathbb{E}\left[\left\|\boldsymbol{\theta}_t^i - y^i(\boldsymbol{\Phi}_{t-1})\right\|^2\right] + 6\alpha_t^2\delta^2K^2\mathbb{E}\left[\|\boldsymbol{\Phi}_t - \boldsymbol{\Phi}^*\|^2\right] + 6\alpha_t^2\delta^2K^2(1+B^2) + 2\alpha_t^2K^2L^2B^2 \\
&\quad - 2\alpha_t K\omega\mathbb{E}\left[\|\boldsymbol{\theta}_t^i - y^i(\boldsymbol{\Phi}_{t-1})\|^2\right] + \beta_{t-1}/\alpha_t\mathbb{E}\left[\left\|\boldsymbol{\theta}_t^i - y^i(\boldsymbol{\Phi}_{t-1})\right\|^2\right] \\
&\quad + \alpha_t^3/\beta_{t-1}(2L^2\mathbb{E}\left[\|\boldsymbol{\Phi}_t - \boldsymbol{\Phi}_{t-1}\|^2\right] + 2L^2K^2B^2) \\
&\quad + \beta_{t-1}/\alpha_t\mathbb{E}\left[\left\|\boldsymbol{\theta}_t^i - y^i(\boldsymbol{\Phi}_{t-1})\right\|^2\right] + \alpha_t^3/\beta_{t-1}(3K^2B^2 + 3K^2\delta^2) \\
&\quad + 6K^2\delta^2\alpha_t^3/\beta_{t-1}\mathbb{E}[\|\boldsymbol{\Phi}_{t-1} - \boldsymbol{\Phi}^*\|^2] + 6K^2\delta^2\alpha_t^3/\beta_{t-1}\mathbb{E}[\|\boldsymbol{\Phi}_t - \boldsymbol{\Phi}_{t-1}\|^2] \\
&\leq (1 + 2\beta_{t-1}/\alpha_t - 2\alpha_t K\omega)\mathbb{E}\left[\left\|\boldsymbol{\theta}_t^i - y^i(\boldsymbol{\Phi}_{t-1})\right\|^2\right] \\
&\quad + (12\alpha_t^2\delta^2K^2 + 6K^2\delta^2\alpha_t^3/\beta_{t-1})\mathbb{E}[\|\boldsymbol{\Phi}_{t-1} - \boldsymbol{\Phi}^*\|^2] \\
&\quad + (12\alpha_t^2\delta^2K^2 + 2L^2\alpha_t^3/\beta_{t-1} + 6K^2\delta^2\alpha_t^3/\beta_{t-1})\mathbb{E}[\|\boldsymbol{\Phi}_t - \boldsymbol{\Phi}_{t-1}\|^2] \\
&\quad + 6\alpha_t^2\delta^2K^2(1+B^2) + 2\alpha_t^2K^2L^2B^2 + 2L^2K^2B^2\alpha_t^3/\beta_{t-1} \\
&\quad + \alpha_t^3/\beta_{t-1}(3K^2B^2 + 3K^2\delta^2)
\end{aligned} \tag{44}$$

This completes the proof.

□

Next, we bound $Term_5$ in the following lemma.

Lemma F.10. *With $t \geq \tau$, we have $Term_5$ bounded as*

$$Term_5 \leq 4\beta_{t-1}^2(L^4 + L^6)\mathbb{E}[\|\boldsymbol{\Phi}^* - \boldsymbol{\Phi}_{t-1}\|^2] + \frac{4\beta_{t-1}^2L^4}{N}\mathbb{E}\left[\sum_{i=1}^N \|\boldsymbol{\theta}_t - y^i(\boldsymbol{\Phi}_{t-1})\|^2\right] + 4L^2\beta_{t-1}^2\delta^2. \tag{45}$$

Proof. We have

$$\begin{aligned}
Term_5 &= \mathbb{E}\left[\left\|y^i(\boldsymbol{\Phi}_{t-1}) - y^i(\boldsymbol{\Phi}_t)\right\|^2\right] = L^2\mathbb{E}\left[\|\boldsymbol{\Phi}_t - \boldsymbol{\Phi}_{t-1}\|^2\right] \\
&= \frac{L^2\beta_{t-1}^2}{N^2}\mathbb{E}\left[\left\|\sum_{i=1}^N \mathbf{h}(\boldsymbol{\theta}_t^i, \boldsymbol{\Phi}_{t-1})\right\|^2\right] \\
&\leq 4\beta_{t-1}^2(L^4 + L^6)\mathbb{E}[\|\boldsymbol{\Phi}^* - \boldsymbol{\Phi}_{t-1}\|^2] + \frac{4\beta_{t-1}^2L^4}{N}\mathbb{E}\left[\sum_{i=1}^N \|\boldsymbol{\theta}_t - y^i(\boldsymbol{\Phi}_{t-1})\|^2\right] + 4L^2\beta_{t-1}^2\delta^2,
\end{aligned} \tag{46}$$

where the last inequality holds due to Lemma F.2. □

Next, we bound $Term_6$ in the following lemma.

Lemma F.11. *We have $Term_6$ bounded as*

$$Term_6 \leq \beta_{t-1}/\alpha_t Term_4 + \alpha_t/\beta_{t-1}Term_5. \tag{47}$$

Proof.

$$\begin{aligned}
Term_6 &= 2\mathbb{E}\left[\left\langle \boldsymbol{\theta}_t^i - y^i(\boldsymbol{\Phi}_{t-1}) + \alpha_t\sum_{k=1}^K \mathbf{g}(\boldsymbol{\theta}_{t,k-1}^i, \boldsymbol{\Phi}_t), y^i(\boldsymbol{\Phi}_{t-1}) - y^i(\boldsymbol{\Phi}_t)\right\rangle\right] \\
&\leq \beta_{t-1}/\alpha_t\underbrace{\mathbb{E}\left[\left\|\boldsymbol{\theta}_t^i - y^i(\boldsymbol{\Phi}_{t-1}) + \alpha_t\sum_{k=1}^K \mathbf{g}(\boldsymbol{\theta}_{t,k-1}^i\right\|^2\right]}_{Term_4} + \alpha_t/\beta_{t-1}\underbrace{\mathbb{E}\left[\left\|y^i(\boldsymbol{\Phi}_{t-1}) - y^i(\boldsymbol{\Phi}_t)\right\|^2\right]}_{Term_5}
\end{aligned} \tag{48}$$

□

Providing $Term_4$ in Lemma F.9, $Term_5$ in Lemma F.10, and $Term_6$ in Lemma F.11, we have the following result.

Lemma F.12. *For $t \geq \tau$, the following holds*

$$
\begin{aligned}
&\mathbb{E}[\|\boldsymbol{\theta}_{t+1}^i - y^i(\boldsymbol{\Phi}_t)\|^2] \\
&\leq \Bigg[(1+\beta_{t-1}/\alpha_t)\Bigg((1+2\beta_{t-1}/\alpha_t - 2\alpha_t K\omega) \\
&\quad + (12\alpha_t^2\delta^2K^2 + 2L^2\alpha_t^3/\beta_{t-1} + 6K^2\delta^2\alpha_t^3/\beta_{t-1})\frac{4\beta_{t-1}^2L^2}{N}\Bigg) + (1+\alpha_t/\beta_{t-1})\frac{4\beta_{t-1}^2L^4}{N}\Bigg] \\
&\qquad \cdot \mathbb{E}\left[\left\|\boldsymbol{\theta}_t^i - y^i(\boldsymbol{\Phi}_{t-1})\right\|^2\right] \\
&\quad + \Bigg[(1+\beta_{t-1}/\alpha_t)\Bigg((12\alpha_t^2\delta^2K^2 + 6K^2\delta^2\alpha_t^3/\beta_{t-1}) \\
&\quad + (12\alpha_t^2\delta^2K^2 + 2L^2\alpha_t^3/\beta_{t-1} + 6K^2\delta^2\alpha_t^3/\beta_{t-1})(4\beta_{t-1}(L^2+L^4))\Bigg) \\
&\quad + (1+\alpha_t/\beta_{t-1})4\beta_{t-1}^2(L^4+L^6)\Bigg] \cdot \mathbb{E}[\|\boldsymbol{\Phi}^* - \boldsymbol{\Phi}_{t-1}\|^2] \\
&\quad + (1+\beta_{t-1}/\alpha_t)\Big((12\alpha_t^2\delta^2K^2 + 2L^2\alpha_t^3/\beta_{t-1} + 6K^2\delta^2\alpha_t^3/\beta_{t-1})4\beta_{t-1}^2\delta^2 \\
&\quad + 6\alpha_t^2\delta^2K^2(1+B^2) + 2\alpha_t^2K^2L^2B^2 + 2L^2K^2B^2\alpha_t^3/\beta_{t-1} + \alpha_t^3/\beta_{t-1}(3K^2B^2 + 3K^2\delta^2)\Big) \\
&\quad + (1-\alpha_t/\beta_{t+1}) \cdot 4L^2\beta_{t-1}^2\delta^2.
\end{aligned} \tag{49}
$$

Proof. According to (36), we have

$$
\begin{aligned}
&\mathbb{E}[\|\boldsymbol{\theta}_{t+1}^i - y^i(\boldsymbol{\Phi}_t)\|^2] = Term_4 + Term_5 + Term_6 \\
&\overset{\text{Lemma F.11}}{\leq} (1+\beta_{t-1}/\alpha_t)Term_4 + (1+\alpha_t/\beta_{t-1})Term_5 \\
&\leq (1+\beta_{t-1}/\alpha_t)\Bigg((1+2\beta_{t-1}/\alpha_t - 2\alpha_t K\omega)\mathbb{E}\left[\left\|\boldsymbol{\theta}_t^i - y^i(\boldsymbol{\Phi}_{t-1})\right\|^2\right] \\
&\quad + (12\alpha_t^2\delta^2K^2 + 6K^2\delta^2\alpha_t^3/\beta_{t-1})\mathbb{E}[\|\boldsymbol{\Phi}_{t-1} - \boldsymbol{\Phi}^*\|^2] \\
&\quad + (12\alpha_t^2\delta^2K^2 + 2L^2\alpha_t^3/\beta_{t-1} + 6K^2\delta^2\alpha_t^3/\beta_{t-1})\mathbb{E}[\|\boldsymbol{\Phi}_t - \boldsymbol{\Phi}_{t-1}\|^2] \\
&\quad + 6\alpha_t^2\delta^2K^2(1+B^2) + 2\alpha_t^2K^2L^2B^2 + 2L^2K^2B^2\alpha_t^3/\beta_{t-1} + \alpha_t^3/\beta_{t-1}(3K^2B^2 + 3K^2\delta^2)\Bigg) \\
&\quad + (1+\alpha_t/\beta_{t-1})\Bigg(4\beta_{t-1}^2(L^4+L^6)\mathbb{E}[\|\boldsymbol{\Phi}^* - \boldsymbol{\Phi}_{t-1}\|^2] \\
&\quad + \frac{4\beta_{t-1}^2L^4}{N}\mathbb{E}\left[\sum_{i=1}^N \|\boldsymbol{\theta}_t - y^i(\boldsymbol{\Phi}_{t-1})\|^2\right] + 4L^2\beta_{t-1}^2\delta^2\Bigg) \\
&\overset{\text{Lemma F.10}}{\leq} (1+\beta_{t-1}/\alpha_t)\Bigg((1+2\beta_{t-1}/\alpha_t - 2\alpha_t K\omega)\mathbb{E}\left[\left\|\boldsymbol{\theta}_t^i - y^i(\boldsymbol{\Phi}_{t-1})\right\|^2\right] \\
&\quad + (12\alpha_t^2\delta^2K^2 + 6K^2\delta^2\alpha_t^3/\beta_{t-1})\mathbb{E}[\|\boldsymbol{\Phi}_{t-1} - \boldsymbol{\Phi}^*\|^2] \\
&\quad + (12\alpha_t^2\delta^2K^2 + 2L^2\alpha_t^3/\beta_{t-1} + 6K^2\delta^2\alpha_t^3/\beta_{t-1}) \\
&\qquad \cdot \left(4\beta_{t-1}^2(L^2+L^4)\mathbb{E}[\|\boldsymbol{\Phi}^* - \boldsymbol{\Phi}_{t-1}\|^2] + \frac{4\beta_{t-1}^2L^2}{N}\mathbb{E}\left[\sum_{i=1}^N \|\boldsymbol{\theta}_t - y^i(\boldsymbol{\Phi}_{t-1})\|^2\right] + 4\beta_{t-1}^2\delta^2\right)
\end{aligned}
$$

$$
\begin{aligned}
&\quad + 6\alpha_t^2\delta^2K^2(1+B^2) + 2\alpha_t^2K^2L^2B^2 + 2L^2K^2B^2\alpha_t^3/\beta_{t-1} + \alpha_t^3/\beta_{t-1}(3K^2B^2+3K^2\delta^2)\Bigg) \\
&\quad + (1+\alpha_t/\beta_{t-1})\Bigg(4\beta_{t-1}^2(L^4+L^6)\mathbb{E}[\|\mathbf{\Phi}^* - \mathbf{\Phi}_{t-1}\|^2] \\
&\quad + \frac{4\beta_{t-1}^2L^4}{N}\mathbb{E}\left[\sum_{i=1}^{N}\|\boldsymbol{\theta}_t - y^i(\mathbf{\Phi}_{t-1})\|^2\right] + 4L^2\beta_{t-1}^2\delta^2\Bigg) \\
= &\Bigg[(1+\beta_{t-1}/\alpha_t)\Bigg((1+2\beta_{t-1}/\alpha_t - 2\alpha_tK\omega) \\
&\quad + (12\alpha_t^2\delta^2K^2 + 2L^2\alpha_t^3/\beta_{t-1} + 6K^2\delta^2\alpha_t^3/\beta_{t-1})\frac{4\beta_{t-1}^2L^2}{N}\Bigg) + (1+\alpha_t/\beta_{t-1})\frac{4\beta_{t-1}^2L^4}{N}\Bigg] \\
&\qquad \cdot \mathbb{E}\left[\left\|\boldsymbol{\theta}_t^i - y^i(\mathbf{\Phi}_{t-1})\right\|^2\right] \\
&\quad + \Bigg[(1+\beta_{t-1}/\alpha_t)\Bigg((12\alpha_t^2\delta^2K^2 + 6K^2\delta^2\alpha_t^3/\beta_{t-1}) \\
&\quad + (12\alpha_t^2\delta^2K^2 + 2L^2\alpha_t^3/\beta_{t-1} + 6K^2\delta^2\alpha_t^3/\beta_{t-1})(4\beta_{t-1}(L^2+L^4))\Bigg) \\
&\quad + (1+\alpha_t/\beta_{t-1})4\beta_{t-1}^2(L^4+L^6)\Bigg]\cdot\mathbb{E}[\|\mathbf{\Phi}^* - \mathbf{\Phi}_{t-1}\|^2] \\
&\quad + (1+\beta_{t-1}/\alpha_t)\Big((12\alpha_t^2\delta^2K^2 + 2L^2\alpha_t^3/\beta_{t-1} + 6K^2\delta^2\alpha_t^3/\beta_{t-1})4\beta_{t-1}^2\delta^2 \\
&\quad + 6\alpha_t^2\delta^2K^2(1+B^2) + 2\alpha_t^2K^2L^2B^2 + 2L^2K^2B^2\alpha_t^3/\beta_{t-1} + \alpha_t^3/\beta_{t-1}(3K^2B^2+3K^2\delta^2)\Big) \\
&\quad + (1+\alpha_t/\beta_{t-1})\cdot 4L^2\beta_{t-1}^2\delta^2.
\end{aligned}
\tag{50}
$$

This completes the proof. □

F.1.3 Final Step of Proof for Theorem 4.1

Now, we are ready to proof the desired result in Theorem 4.1.

According to the definition of Lyapunov function in (14), We have

$$
\begin{aligned}
&M(\{\boldsymbol{\theta}_{t+2}^i\},\mathbf{\Phi}_{t+1}) = \|\mathbf{\Phi}_{t+1} - \mathbf{\Phi}^*\|^2 + \frac{\beta_t}{\alpha_{t+1}}\cdot\frac{1}{N}\sum_{i=1}^{N}\|\boldsymbol{\theta}_{t+2}^i - y^i(\mathbf{\Phi}_{t+1})\|^2 \\
&\le (1 + 4\beta_t^2(L^2+L^4) + (7\beta_t/\alpha_t + 2\beta_t\alpha_tL^2 + 6\beta_t\alpha_t\delta^2)4\tau^2\beta_0^2/\alpha_0^2 \\
&\qquad + (6\beta_t/\alpha_t + 6\beta_t\alpha_t\delta^2(1+L^2) + 4\beta_t\alpha_tL^2(3+4L^2)) + \beta_t(L/\alpha_t - 2\omega))\mathbb{E}[\|\mathbf{\Phi}_t - \mathbf{\Phi}^*\|^2] \\
&\qquad + \frac{4\beta_t^2L^2 + \beta_t\alpha_tL + 16\beta_t\alpha_tL^2 + 6\beta_t\alpha_t\delta^2}{N}\mathbb{E}\left[\sum_{i=1}^{N}\|\boldsymbol{\theta}_{t+1}^i - y^i(\mathbf{\Phi}_t)\|^2\right] \\
&\qquad + (7\beta_t/\alpha_t + 2\beta_t\alpha_tL^2 + 6\beta_t\alpha_t\delta^2)(8\beta_0^2L^2B^2\tau^2 + 8\beta_0^2\delta^2\tau^2) + 4\beta_t^2\delta^2 + 11\beta_t\alpha_t\delta^2 \\
&+ \frac{\beta_t}{\alpha_{t+1}}\cdot\Bigg[(1+\beta_t/\alpha_{t+1})\Bigg((1+2\beta_t/\alpha_{t+1} - 2\alpha_{t+1}K\omega) \\
&\qquad + (12\alpha_{t+1}^2\delta^2K^2 + 2L^2\alpha_{t+1}^3/\beta_t + 6K^2\delta^2\alpha_{t+1}^3/\beta_t)\frac{4\beta_t^2L^2}{N}\Bigg) + (1+\alpha_{t+1}/\beta_t)\frac{4\beta_t^2L^4}{N}\Bigg] \\
&\qquad\quad \cdot\frac{1}{N}\mathbb{E}\left[\sum_{i=1}^{N}\left\|\boldsymbol{\theta}_{t+1}^i - y^i(\mathbf{\Phi}_t)\right\|^2\right]
\end{aligned}
$$

$$
\begin{aligned}
&+ \Bigg[(1+\beta_t/\alpha_{t+1})\Bigg((12\alpha_{t+1}^2\delta^2K^2 + 6K^2\delta^2\alpha_{t+1}^3/\beta_t) \\
&+ (12\alpha_{t+1}^2\delta^2K^2 + 2L^2\alpha_{t+1}^3/\beta_t + 6K^2\delta^2\alpha_{t+1}^3/\beta_t)(4\beta_t(L^2+L^4))\Bigg) \\
&+ (1+\alpha_{t+1}/\beta_t)4\beta_t^2(L^4+L^6)\Bigg] \cdot \mathbb{E}[\|\boldsymbol{\Phi}^* - \boldsymbol{\Phi}_t\|^2] \\
&+ (1+\beta_t/\alpha_{t+1})\Big((12\alpha_{t+1}^2\delta^2K^2 + 2L^2\alpha_{t+1}^3/\beta_t + 6K^2\delta^2\alpha_{t+1}^3/\beta_t)4\beta_t^2\delta^2 \\
&+ 6\alpha_{t+1}^2\delta^2K^2(1+B^2) + 2\alpha_{t+1}^2K^2L^2B^2 + 2L^2K^2B^2\alpha_{t+1}^3/\beta_t + \alpha_{t+1}^3/\beta_t(3K^2B^2+3K^2\delta^2)\Big) \\
&+ (1+\alpha_{t+1}/\beta_t)\cdot 4L^2\beta_t^2\delta^2\Bigg].
\end{aligned}
\tag{51}
$$

To simplify the notations, we define

$$
\begin{aligned}
D_1 := {}& (4\beta_t^2(L^2+L^4) + (7\beta_t/\alpha_t + 2\beta_t\alpha_tL^2 + 6\beta_t\alpha_t\delta^2)4\tau^2\beta_0^2/\alpha_0^2 \\
&+ (6\beta_t/\alpha_t + 6\beta_t\alpha_t\delta^2(1+L^2) + 4\beta_t\alpha_tL^2(3+4L^2)) + \beta_tL/\alpha_t) \\
&+ \frac{\beta_t}{\alpha_{t+1}}\Bigg[(1+\beta_t/\alpha_{t+1})\Bigg((12\alpha_{t+1}^2\delta^2K^2 + 6K^2\delta^2\alpha_{t+1}^3/\beta_t) \\
&+ (12\alpha_{t+1}^2\delta^2K^2 + 2L^2\alpha_{t+1}^3/\beta_t + 6K^2\delta^2\alpha_{t+1}^3/\beta_t)(4\beta_t(L^2+L^4))\Bigg) \\
&+ (1+\alpha_{t+1}/\beta_t)4\beta_t^2(L^4+L^6)\Bigg],
\end{aligned}
\tag{52}
$$

and

$$
\begin{aligned}
&D_2 := 4\beta_t^3/\alpha_{t+1}L^2 + \alpha_t^2L + 16\alpha_t^2L^2 + 6\alpha_t\alpha_t\delta^2 \\
&+ \Bigg[\Bigg((2\beta_t/\alpha_{t+1}) + (12\alpha_{t+1}^2\delta^2K^2 + 2L^2\alpha_{t+1}^3/\beta_t + 6K^2\delta^2\alpha_{t+1}^3/\beta_t)\frac{4\beta_t^2L^2}{N}\Bigg) + (1+\alpha_{t+1}/\beta_t)\frac{4\beta_t^2L^4}{N}\Bigg] \\
&+ \Bigg[\beta_t/\alpha_{t+1}\Bigg((1+2\beta_t/\alpha_{t+1} - 2\alpha_{t+1}K\omega) \\
&+ (12\alpha_{t+1}^2\delta^2K^2 + 2L^2\alpha_{t+1}^3/\beta_t + 6K^2\delta^2\alpha_{t+1}^3/\beta_t)\frac{4\beta_t^2L^2}{N}\Bigg) + (1+\alpha_{t+1}/\beta_t)\frac{4\beta_t^2L^4}{N}\Bigg].
\end{aligned}
\tag{53}
$$

Since D_1 is of higher orders of $o(\beta_t)$ and D_2 is of higher order of $o(\alpha_{t+1})$, we can let $D_1 \leq \omega\beta_t$ and $D_2 \leq K\omega\alpha_{t+1}$. Therefore, we have

$$
\begin{aligned}
&M(\{\boldsymbol{\theta}_{t+2}^i\}, \boldsymbol{\Phi}_{t+1}) \leq (1-\omega\beta_t)M(\{\boldsymbol{\theta}_{t+1}^i\}, \boldsymbol{\Phi}_t) \\
&+ (144\tau^2K^2L^2\delta^2 + 4L^4/N)\beta_t\alpha_{t+1}\left[\mathbb{E}[\|\boldsymbol{\Phi}_t - \boldsymbol{\Phi}^*\|^2] + \frac{1}{N}\mathbb{E}\left[\sum_{i=1}^N \left\|\boldsymbol{\theta}_{t+1}^i - y^i(\boldsymbol{\Phi}_t)\right\|^2\right]\right] \\
&+ 4\alpha_{t+1}\beta_tK^2(3\delta^2(1+B^2) + L^2B^2) + 2\alpha_{t+1}^2(3K^2B^2 + 3K^2\delta^2 + 2L^2K^2B^2) + 8\alpha_{t+1}\beta_t\delta^2 \\
&\leq (1-\omega\beta_t)M(\{\boldsymbol{\theta}_{t+1}^i\}, \boldsymbol{\Phi}_t) \\
&+ (144\tau^2K^2L^2\delta^2 + 4L^2/N)\beta_t\alpha_t\left[\mathbb{E}[\|\boldsymbol{\Phi}_t - \boldsymbol{\Phi}^*\|^2] + \frac{1}{N}\mathbb{E}\left[\sum_{i=1}^N \left\|\boldsymbol{\theta}_{t+1}^i - y^i(\boldsymbol{\Phi}_t)\right\|^2\right]\right] \\
&+ 4\alpha_t\beta_tK^2(3\delta^2(1+B^2) + L^2B^2) + 2\alpha_t^2(3K^2B^2 + 3K^2\delta^2 + 2L^2K^2B^2) + 8\alpha_t\beta_t\delta^2,
\end{aligned}
\tag{54}
$$

where the first inequality holds by omitting the higher order of learning rates, and the second inequality holds due to the decreasing learning rates of α_t.

We now set the proper decaying learning rates. Let $\alpha_t = \alpha_0/(t+2)^{5/6}$ and $\beta_t = \beta_0/(t+2)$. We then have

$$(t+2)^2 \cdot (1-\omega\beta_t) = (t+2)^2(1-\omega\beta_0)/(t+2) \leq (t+1)^2, \tag{55}$$

if $\omega\beta_o < 2$. In addition, we have the following inequalities

$$(t+2)^2 \cdot \alpha_t\beta_t \leq \alpha_0\beta_0(t+2)^{1/3},$$
$$(t+2)^2 \cdot \alpha_t^2 = \alpha_0^2(t+2)^2.$$

Hence, multiplying both sides with $(t+2)^2$, we have

$$\begin{aligned}&(t+2)^2 M(\{\boldsymbol{\theta}_{t+2}^i\}, \boldsymbol{\Phi}_{t+1}) \leq (t+1)^2 M(\{\boldsymbol{\theta}_{t+1}^i\}, \boldsymbol{\Phi}_t)\\ &+ (144\tau^2K^2L^2\delta^2 + 4L^2/N)\alpha_0\beta^0(t+2)^{1/3}\left[\mathbb{E}[\|\boldsymbol{\Phi}_t - \boldsymbol{\Phi}^*\|^2] + \frac{1}{N}\mathbb{E}\left[\sum_{i=1}^N \|\boldsymbol{\theta}_{t+1}^i - y^i(\boldsymbol{\Phi}_t)\|^2\right]\right]\\ &+ (4\alpha_0\beta_0K^2(3\delta^2(1+B^2) + L^2B^2) + 2\alpha_0^2(3K^2B^2 + 3K^2\delta^2 + 2L^2K^2B^2) + 8\alpha_0\beta_0\delta^2)(t+2)^{1/3}.\end{aligned}$$

Summing the above equation from $t = 0, \ldots, T$, we have

$$\begin{aligned}&(T+2)^2 M(\{\boldsymbol{\theta}_{t+2}^i\}, \boldsymbol{\Phi}_{t+1}) \leq M(\{\boldsymbol{\theta}_1^i\}, \boldsymbol{\Phi}_0)\\ &+ (144\tau^2K^2L^2\delta^2 + 4L^2/N)\alpha_0\beta^0(T+2)^{4/3}\left[\mathbb{E}[\|\boldsymbol{\Phi}_0 - \boldsymbol{\Phi}^*\|^2] + \frac{1}{N}\mathbb{E}\left[\sum_{i=1}^N \|\boldsymbol{\theta}_1^i - y^i(\boldsymbol{\Phi}_0)\|^2\right]\right]\\ &+ (4\alpha_0\beta_0K^2(3\delta^2(1+B^2) + L^2B^2) + 2\alpha_0^2(3K^2B^2 + 3K^2\delta^2 + 2L^2K^2B^2) + 8\alpha_0\beta_0\delta^2)(T+2)^{4/3}.\end{aligned}$$

Dividing both sides by $(T+2)^2$, we have

$$\begin{aligned}&M(\{\boldsymbol{\theta}_{t+2}^i\}, \boldsymbol{\Phi}_{t+1}) \leq \frac{M(\{\boldsymbol{\theta}_1^i\}, \boldsymbol{\Phi}_0)}{(T+2)^2}\\ &+ (144\tau^2K^2L^2\delta^2 + 4L^2/N)\alpha_0\beta_0(T+2)^{-2/3}\left[\mathbb{E}[\|\boldsymbol{\Phi}_0 - \boldsymbol{\Phi}^*\|^2] + \frac{1}{N}\mathbb{E}\left[\sum_{i=1}^N \|\boldsymbol{\theta}_1^i - y^i(\boldsymbol{\Phi}_0)\|^2\right]\right]\\ &+ (4\alpha_0\beta_0K^2(3\delta^2(1+B^2) + L^2B^2) + 2\alpha_0^2(3K^2B^2 + 3K^2\delta^2 + 2L^2K^2B^2) + 8\alpha_0\beta_0\delta^2)(T+2)^{-2/3}.\end{aligned}$$

This completes the proof.

F.2 Proof of Corollary 4.3

If $\alpha_0 = \beta_0 = o(N^{-1/3}K^{-1/2})$, we have

$$M(\{\boldsymbol{\theta}_{t+2}^i\}, \boldsymbol{\Phi}_{t+1}) \leq \mathcal{O}\left(\frac{1}{(T+2)^2} + \frac{1}{N^{2/3}(T+2)^{2/3}} + \frac{1}{K^2N^{5/3}(T+2)^{2/3}} + \frac{1}{K^2N^{2/3}(T+2)^{2/3}}\right),$$

which is dominated by $\mathcal{O}\left(\frac{1}{N^{2/3}(T+2)^{2/3}}\right)$ if $T^2 > N$.

G Additional experiment details

PFedDQN-Rep in CartPole Environment. We evaluate the performance PFedDQN-Rep (see Appendix B) in a modified CartPole environment Brockman et al. (2016). Similar to Jin et al. (2022), we change the length of pole to create different environments. Specifically, we consider 10 agents with varying pole length from 0.38 to 0.74 with a step size of 0.04. We compare PFedDQN-Rep with (i)

Table 2: Parameter setting

Parameter	Description
Input size	4
Hidden size	$128 \times 128 \times 128$
Output size	2
Activation function	ReLu
Number of episodes	500
Batch size	64
Discount factor	0.98
ϵ greedy parameter	0.01
Target update	30
Buffer size	10000
Minimal size	500
Learning rate	0.002, decays every 100 episodes

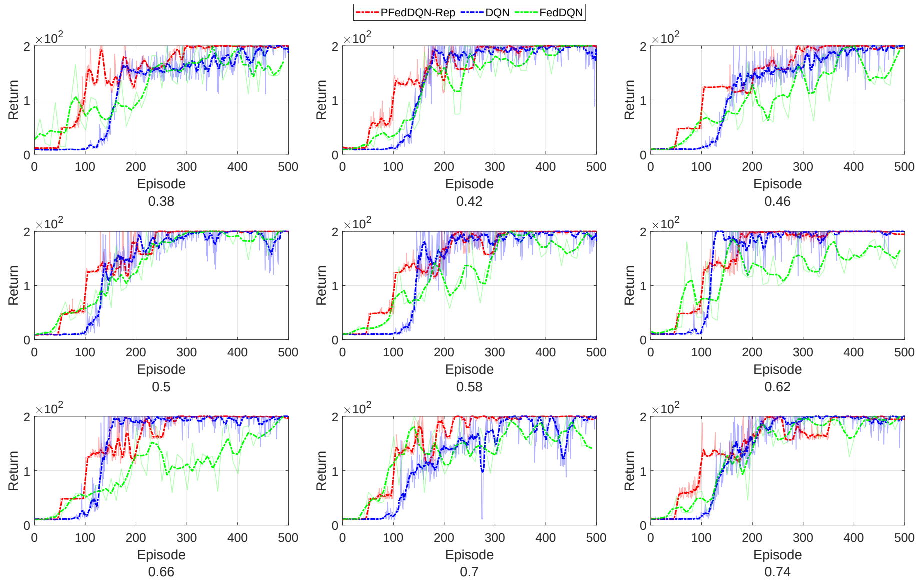


Figure 8: Comparison of control by DQN, FedDQN and PFedDQN-Rep in Cartpole Environments.

a conventional DQN that each agent learns its own environment independently; and (ii) a federated version DQN (FedDQN) that allows all agents to collaboratively learn a single policy (without personalization). We randomly choose one agent and present its performance in Figure 3(top)(a). The results of the other agents are presented in Figure 8. Again, we observe that our PFedDQN-Rep achieves the maximized return much faster than the conventional DQN due to leveraging shared representations among agents; and obtains larger reward than FedDQN, thanks to our personalized policy. We further evaluate the effectiveness of shared representation learned by PFedDQN-Rep when generalizes it to a new agent. As shown in Figure 3(top)(b), our PFedDQN-Rep generalizes quickly to the new environment. Detailed parameter settings can be found in Table 2.

PFedDQN-Rep in Acrobot Environment. We further evaluate FedDQN-Rep in a modified Acrobot environment Brockman et al. (2016). The pole length is adjusted with [-0.3, 0.3] with a step size of 0.06, and the pole mass with be adjusted accordingly Jin et al. (2022). The same two benchmarks are compared as in Figure 3(top). The parameter setting remains the same except number of episodes decreases to 100. Similar observations can be made from Figure 3(bottom) and Figure 9 as those for the Cartpole enviroments.

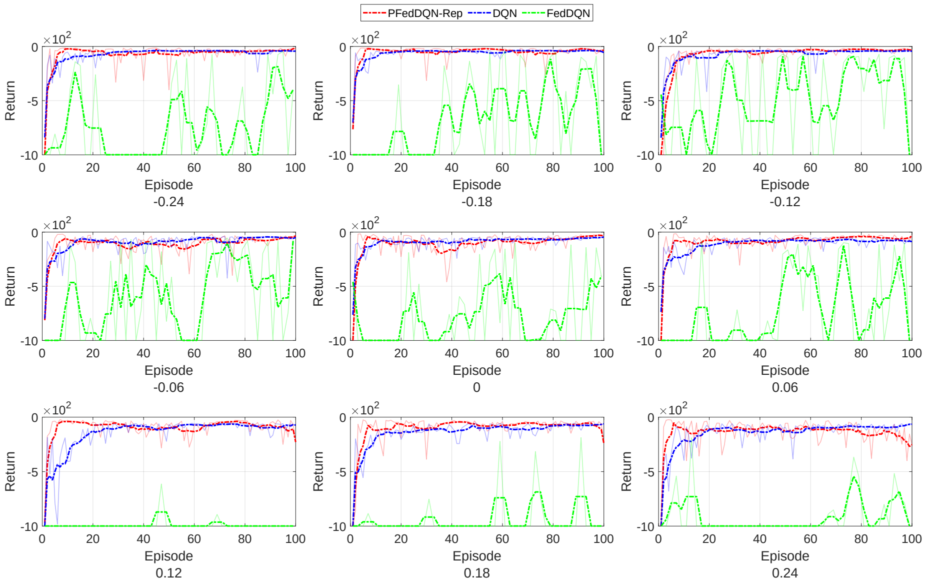


Figure 9: Comparison of control by DQN, FedDQN and PFedDQN-Rep in Acrobot Environments.