Federated Temporal Difference Learning with Linear Function Approximation under Environmental Heterogeneity

Han Wang, Aritra Mitra, Hamed Hassani, George J. Pappas, James Anderson *

Abstract

We initiate the study of federated reinforcement learning under environmental heterogeneity by considering a policy evaluation problem. Our setup involves N agents interacting with environments that share the same state and action space but differ in their reward functions and state transition kernels. Assuming agents can communicate via a central server, we ask: *Does exchanging information expedite the process of evaluating a common policy?* To answer this question, we provide the first comprehensive finite-time analysis of a federated temporal difference (TD) learning algorithm with linear function approximation, while accounting for Markovian sampling, heterogeneity in the agents' environments, and multiple local updates to save communication. Our analysis crucially relies on several novel ingredients: (i) deriving perturbation bounds on TD fixed points as a function of the heterogeneity in the agents' underlying Markov decision processes (MDPs); (ii) introducing a virtual MDP to closely approximate the dynamics of the federated TD algorithm; and (iii) using the virtual MDP to make explicit connections to federated optimization. Putting these pieces together, we rigorously prove that in a low-heterogeneity regime, exchanging model estimates leads to linear convergence speedups in the number of agents.

1 Introduction

In the popular federated learning (FL) paradigm [28, 38], a set of agents aim to find a common statistical model that explains their collective observations. The motivation to collaborate stems from the fact that if the underlying distributions generating the agents' observations are "similar", then each agent can end up learning a "better" model than if it otherwise used just its own data. This idea has been formalized by the canonical FL algorithm `FedAvg` (and its many variants) where agents communicate local models via a central server while keeping their raw data private. To achieve communication-efficiency - a key consideration in FL - the agents perform multiple local model-updates between two successive communication rounds. There is a rich literature that analyzes the performance of `FedAvg`, focusing primarily on the aspect of *statistical heterogeneity* that originates from differences in the agents' underlying data distributions [25, 62, 22, 43, 6, 40, 39]. Notably, the above works focus on supervised learning problems that are modeled within the framework of distributed optimization. However, for sequential decision-making with multiple agents

*Han Wang and James Anderson are with the Department of Electrical Engineering, Columbia University in the City of New York. Email: {hw2786, james.anderson}@columbia.edu. Aritra Mitra is with the Department of Electrical and Computer Engineering, North Carolina State University. Email: amitra2@ncsu.edu. H. Hassani and G. Pappas are with the Department of Electrical and Systems Engineering, University of Pennsylvania. Email: {pappasg, hassani}@seas.upenn.edu. This work was supported by the DoE under grant DE-SC0022234, NSF awards 2144634 & 2231350, NSF Award 1837253, NSF CAREER award CIF 1943064, and AFOSR-YIP under award FA9550-20-1-0111.

interacting with *potentially different environments*, little to nothing is known about the effect of heterogeneity. This is the gap we seek to fill with our work.

The recent survey paper by [45] describes a federated reinforcement learning (FRL) framework which incorporates some of the key ideas from FL in reinforcement learning (RL); applications of FRL in robotics [34], autonomous driving [7], and edge computing [60] are discussed in detail in this paper. As RL algorithms often require many samples to achieve acceptable accuracy, FRL aims to achieve *sample-efficiency* by leveraging information from multiple agents interacting with similar environments. Importantly, as in standard FL, the FRL framework requires agents to keep their personal experiences (e.g., rewards, states, and actions) private, and adhere to stringent communication constraints. While FRL is a promising idea, to model realistic scenarios, one needs to account for the crucial fact that different agents may interact with *non-identical* environments. Indeed, just as statistical heterogeneity is a major challenge in FL, *environmental heterogeneity* is identified as a key open challenge in FRL [45].

To tackle this challenge, we focus on a policy evaluation problem. Our setup involves N agents where each agent interacts with an environment modeled as a MDP. The agents' MDPs share the same state and action space but have different reward functions and state transition kernels, thereby capturing environmental heterogeneity. Each agent seeks to compute the discounted cumulative reward (value function) associated with a common policy μ. Notably, the value functions induced by μ may differ across environments. This leads to the central question we investigate: *Can an agent expedite the process of learning its own value function by leveraging information from potentially different MDPs?* This is a non-trivial question since the effect of combining data from non-identical MDPs is poorly understood.

A typical application of the above FRL setup is that of an autonomous driving system where vehicles in different geographical locations share local models capturing their learned experiences to train a shared model that benefits from the collective exploration data of all vehicles. Although the vehicles (agents) essentially have the same operations (e.g., steering, braking, accelerating, etc.), they can be exposed to different environments (e.g., road and weather conditions, routes, driving regulations etc.). This is precisely what contributes to environmental heterogeneity.

1.1 Our Contributions

We study a federated version of the temporal difference (TD) learning algorithm `TD`(0) [53]. The structure of this algorithm, which we call `FedTD`(0), is as follows. At each iteration, each agent plays an action according to the policy μ, observes a reward, and transitions to a new state based on its *own* MDP. It then uses `TD`(0) with linear function approximation to update a local model that approximates its own value function. To (potentially) benefit from other agents' data in a communication-efficient manner, each agent periodically synchronizes with a central server, and performs multiple local updates in between. Notably, as in FL, agents only exchange models but never their personal observations. We perform a comprehensive analysis of `FedTD`(0) under environmental heterogeneity, and make the following contributions:

Effect of heterogeneity on `TD`(0) fixed points. Towards understanding the behavior of `FedTD`(0), we start by asking: *How does heterogeneity in the transition kernels and reward functions of MDPs manifest into differences in the long-term behavior of `TD`(0) (with linear function approximation) on such MDPs?* Theorem 1 provides an answer by characterizing how perturbing a MDP perturbs the `TD`(0) fixed point for that MDP. To arrive at this result, we combine results from the perturbation theories of Markov chains and linear equations. Theorem 1 establishes the first

perturbation result for TD(0) fixed points, and complements results of a similar flavor in the RL literature such as the *Simulation Lemma* [23].

The Virtual MDP framework. In FL algorithms such as FedAvg, the average of the negative gradients of the agents' loss functions drives the iterates of FedAvg towards the minimizer of a global loss function. In our setting, there is no such global loss function. *So by averaging TD(0) update directions of different MDPs, where do we end up?* To answer this question, we construct a virtual MDP in Section 3.2, and characterize several important properties of this fictitious MDP that aid our subsequent analysis. Along the way, we derive a simple yet key result (Proposition 1) pertaining to convex combinations of Markov matrices associated with aperiodic and irreducible Markov chains; this result may be of independent interest.

Analysis under an i.i.d. assumption. To isolate the effect of heterogeneity and build intuition, we start by analyzing FedTD(0) under a standard i.i.d. assumption in the RL literature [9, 2, 11]. After T communication rounds with K local model-updating steps per round, we prove that FedTD(0) guarantees convergence at a rate of $\tilde{O}(1/NKT)$ to a neighborhood of each agent's optimal linear approximation parameter; see Theorem 2. The size of the neighborhood depends on the level of heterogeneity in the agents' MDPs. *The key implication of this result is that in a low-heterogeneity regime, each agent can enjoy an N-fold speed-up in convergence via collaboration.* To prove this result, we introduce a new analysis technique that combines the virtual MDP idea with the optimization interpretation of TD(0) dynamics in [2]. An important benefit of this technique is that it highlights the connections between the dynamics of FedTD(0) and standard FL algorithms, allowing one to leverage existing FL optimization proofs for federated RL.

Bias introduced by Heterogeneity. Our convergence result in Theorem 2 features a bias term due to heterogeneity that cannot be eliminated even by making the step-size arbitrarily small. *Is such a term unavoidable?* We explore this question in Theorem 3 by studying a "steady-state" deterministic version of FedTD(0). Even for this simple case, we prove that a bias term depending on a natural measure of heterogeneity shows up *inevitably* in the long-term dynamics of FedTD(0). Moreover, unlike the standard FL setting where the effect of heterogeneity manifests itself only when the number of local steps K is strictly greater than 1 [6], the bias term in Theorem 3 persists even when $K = 1$. This reveals a key difference between our setting and federated optimization.

Analysis for the Markovian setting. Our most significant contribution is to provide the *first analysis of a federated RL algorithm (FedTD(0)) that simultaneously accounts for linear function approximation, Markovian sampling, multiple local updates, and heterogeneity.* The effect of heterogeneity coupled with complex temporal correlations makes this setting challenging to analyze. Nonetheless, in Theorem 4, we prove that one can essentially recover the same guarantees as in the i.i.d. setting (Theorem 2). Our result complements the myriad of federated optimization results that account for heterogeneity [25, 62].

We now briefly discuss most directly related work; a detailed description is given in the Appendix.

Related Work. In [11, 35], the authors analyze multi-agent TD learning with linear function approximation over peer-to-peer networks. Neither approach accounts for local steps nor Markovian sampling. Very recently, the authors in [26] do study the effect of Markovian sampling for federated TD learning. However, all of the above papers consider a *homogeneous setting with identical* MDPs for all agents. The only paper we are aware of that performs any theoretical analysis of heterogeneity

in FRL is [21]. However, their analysis is limited to the much more simpler tabular setting with no function approximation.

2 Policy Evaluation in a Centralized Setting

Our RL setting is based on a Markov Decision Process (MDP) [54] defined by the tuple $\mathcal{M} = \langle \mathcal{S}, \mathcal{A}, \mathcal{R}, \mathcal{P}, \gamma \rangle$, where $\mathcal{S}$ is a finite state space of size n, $\mathcal{A}$ is a finite action space, $\mathcal{P}$ is a set of action-dependent Markov transition kernels, $\mathcal{R}$ is a reward function, and $\gamma \in (0,1)$ is the discount factor. We consider the problem of evaluating the value function V_μ of a given policy μ, where $\mu : \mathcal{S} \to \mathcal{A}$. The policy μ induces a Markov reward process (MRP) characterized by a transition matrix P_μ, and a reward function R_μ. Under the action of the policy μ at an initial state s, $P_\mu(s, s')$ is the probability of transitioning from state s to state s', and $R_\mu(s)$ is the expected instantaneous reward. The discounted expected cumulative reward obtained by playing policy μ starting from initial state s is:

$$V_\mu(s) = \mathbb{E}\left[\sum_{t=0}^{\infty} \gamma^t R_\mu(s_t) | s_0 = s\right],$$

where s_t is the state of the Markov chain at time t. From [57], we know that V_μ is the fixed point of the Bellman operator $T_\mu : \mathbb{R}^n \to \mathbb{R}^n$, i.e., $T_\mu V_\mu = V_\mu$, where for any $V \in \mathbb{R}^n$,

$$(T_\mu V)(s) = R_\mu(s) + \gamma \sum_{s' \in \mathcal{S}} P_\mu(s, s') V(s'), \ \forall s \in \mathcal{S}.$$

TD learning with linear function approximation. We consider the setting where the number of states is very large, making it practically infeasible to compute the value function V_μ directly. To mitigate the curse of dimensionality, a common approach [54] is to consider a low-dimensional linear function approximation of the value function V_μ. Let $\{\Phi_k\}_{k=1}^d$ be a set of d linearly independent basis vectors in $\mathbb{R}^n$, and $\Phi \in \mathbb{R}^{n \times d}$ be a matrix with these basis vectors as its columns, i.e., the k-th column of Φ is Φ_k. A parametric approximation $\hat{V}_\theta$ of V_μ in the span of $\{\Phi_k\}_{k=1}^d$ is then given by $\hat{V}_\theta = \Phi\theta$, where $\theta \in \mathbb{R}^d$ is a parameter vector to be learned. Notably, this is tractable since $d \ll n$. We denote the s-th row of Φ by $\phi(s) \in \mathbb{R}^d$, and refer to it as the fixed feature vector corresponding to state s. We write $\hat{V}_\theta(s) = \phi(s)^\top \theta$ and make the standard assumption [2] that $\|\phi(s)\|^2 \leq 1, \forall s \in \mathcal{S}$.

The objective is to find the best linear approximation of V_μ in the span of $\{\Phi_k\}_{k=1}^d$. More precisely, we seek a parameter vector θ^* that minimizes the distance between $\hat{V}_\theta$ and V_μ (in a suitable sense). When the underlying MDP is *unknown*, one of the most popular techniques to achieve this goal is the classical TD(0) algorithm. TD(0) starts from an initial guess $\theta_0 \in \mathbb{R}^d$. Subsequently, at the t-th iteration, upon playing the given policy μ, a new data tuple $O_t = (s_t, r_t = R_\mu(s_t), s_{t+1})$ comprising of the current state, the instantaneous reward, and the next state is observed. Let us define the TD(0) update direction as

$$g_t(\theta_t) \triangleq \left(r_t + \gamma \phi\left(s_{t+1}\right)^\top \theta_t - \phi\left(s_t\right)^\top \theta_t\right) \phi\left(s_t\right).$$

Using a step-size $\alpha_t \in (0,1)$, the parameter θ_t is then updated as $\theta_{t+1} = \theta_t + \alpha_t g_t(\theta_t)$. Under some mild technical assumptions, it was shown in [57] that the TD(0) iterates converge asymptotically almost surely to a vector θ^*, where θ^* is the unique solution of the projected Bellman equation $\Pi_D T_\mu(\Phi\theta^*) = \Phi\theta^*$. Here, D is a diagonal matrix with entries given by the elements of the stationary distribution π of the Markov matrix P_μ. Furthermore, $\Pi_D(\cdot)$ is the projection operator onto the subspace spanned by $\{\phi_k\}_{k=1}^d$ with respect to the inner product $\langle \cdot, \cdot \rangle_D$.[1]

[1] We will use $\|\cdot\|_D^2$ to denote the quadratic norm $x^T D x$ induced by the positive definite matrix D, and $\|\cdot\|$ to represent the standard Euclidean norm for vectors and ℓ_2 induced norm for matrices.

Objective. We study a multi-agent RL problem where agents interact with similar, but *non-identical* MDPs that share the same state and action space. All agents seek to evaluate the same policy. Our goal is to understand: *Can an agent evaluate the value function of its own MDP in a more sample-efficient way by leveraging data from other agents?* Answering this question is non-trivial since one needs to (i) model heterogeneity in the agents' MDPs; and (ii) understand the effects of such heterogeneity on the convergence of algorithms that combine information from non-identical MDPs. Existing FL analyses that study statistical heterogeneity in supervised learning fall short of resolving the above issues, since *our problem does not involve minimizing a static loss function.* In the next section, we will formally introduce our setup and the key ideas needed for our subsequent analysis.

3 Heterogeneous Federated RL

We consider a federated reinforcement learning setting comprising of N agents that interact with potentially different environments. Agent i's environment is characterized by the following MDP: $\mathcal{M}^{(i)} = \langle \mathcal{S}, \mathcal{A}, \mathcal{R}^{(i)}, \mathcal{P}^{(i)}, \gamma \rangle$. While all agents share the same state and action space, the reward functions and state transition kernels of their environments can differ. We focus on a policy evaluation problem where all agents seek to evaluate a common policy μ that induces N Markov reward processes characterized by the tuples $\{P_\mu^{(i)}, R_\mu^{(i)}\}_{i \in [N]}$.[2] Agent i aims to find a linearly parameterized approximation of its value function $V_\mu^{(i)}$. Trivially, agent i can do so without interacting with any other agent by employing the TD(0) algorithm. However, the key question we ask is: *By using data from other agents, can it achieve a desired level of approximation with fewer samples relative to when it acts alone?* Naturally, the answer to the above question depends on the level of heterogeneity in the agents' MDPs. Accordingly, inspired by notions of bounded heterogeneity in federated supervised learning [47], we make the following assumptions.

Assumption 1. ***(Markov Kernel Heterogeneity)*** *There exists an $\epsilon > 0$ such that for all agents $i, j \in [N]$, it holds that $|P^{(i)}(s, s') - P^{(j)}(s, s')| \leq \epsilon |P^{(i)}(s, s')|, \forall s, s' \in \mathcal{S}$. Here, for each $i \in [N]$, $P^{(i)}(s, s')$ represents the (s, s')-th element of the matrix $P^{(i)}$.*

Assumption 2. ***(Reward Heterogeneity)*** *There exists an $\epsilon_1 > 0$ such that for all $i, j \in [N]$, it holds that $\|R^{(i)} - R^{(j)}\| \leq \epsilon_1$.*

Clearly, smaller values of ϵ and ϵ_1 capture more similarity in the agents' MDPs. In line with the standard communication architecture in FL [28, 38], suppose all agents can exchange information via a central server. Via such communication, the standard FL task is to find one common model that explains the data of all agents. In a similar spirit, our goal is to find one common parameter θ such that $\hat{V}_\theta = \Phi\theta$ approximates each $V_\mu^{(i)}, i \in [N]$. There is a natural tension here. While using data from multiple agents can help find an approximate model *quickly*, such a model may not *accurately* capture the value function of *any* agent if the agents' MDPs are very dissimilar. *So does more data help or hurt?* It turns out that to answer the above question, we need to carefully understand how the structural heterogeneity assumptions on the MDPs (namely, Assumptions 1 and 2) manifest into differences in the long-term dynamics of TD(0) on these MDPs. In the sequel, we will comprehensively explore this topic.

[2] For simplicity of notation, we will henceforth drop the dependence of $P^{(i)}$ and $R^{(i)}$ on the policy μ.

3.1 Impact of Heterogeneity on TD fixed points

Intuitively, if the MRPs induced by a common policy for two different environments are similar, then the long-term behavior of TD(0) on these two MRPs should also be similar. In particular, the TD(0) fixed points of these MRPs should be close. As we shall see later, characterizing this "closeness" in TD(0) fixed points will play a key role in understanding how environmental heterogeneity affects the behavior of a federated TD algorithm. To proceed, we make the following standard assumption.

Assumption 3. *For each $i \in [N]$, the Markov chain induced by the policy μ, corresponding to the state transition matrix $P^{(i)}$, is aperiodic and irreducible.*

The above assumption implies the existence of a unique stationary distribution $\pi^{(i)}$ for each $i \in [N]$; let $D^{(i)}$ be a diagonal matrix with the entries of $\pi^{(i)}$ on its diagonal. For each agent i, we then use θ_i^* to denote the solution of the projected Bellman equation $\Pi_{D^{(i)}} T_\mu^{(i)}(\Phi\theta_i^*) = \Phi\theta_i^*$ for agent i. In words, θ_i^* is the best linear approximation of $V_\mu^{(i)}$ in the span of $\{\phi_k\}_{k=1}^d$. Based on the discussion in Section 2, we know that the iterates of TD(0) on agent i's MRP will converge asymptotically (almost surely) to θ_i^*. Our goal is to provide a bound on the gap $\|\theta_i^* - \theta_j^*\|$ as a function of the heterogeneity parameters ϵ and ϵ_1 appearing in Assumptions 1 and 2. The key observation we will exploit is that for each $i \in [N]$, θ_i^* is the unique solution of the linear equation $\bar{A}_i\theta_i^* = \bar{b}_i$, where $\bar{A}_i = \Phi^\top D^{(i)}(\Phi - \gamma P^{(i)}\Phi)$ and $\bar{b}_i = \Phi^\top D^{(i)} R^{(i)}$. For an agent $j \neq i$, viewing $\bar{A}_j$ and $\bar{b}_j$ as perturbed versions of $\bar{A}_i$ and $\bar{b}_i$, we can now appeal to results from the perturbation theory of linear equations [19, Chapter 5.8] to bound $\|\theta_i^* - \theta_j^*\|$. To that end, we first recall a result from the perturbation theory of Markov chains [42] which shows that under Assumption 1, the stationary distributions $\pi^{(i)}$ and $\pi^{(j)}$ are close for any pair $i, j \in [N]$.

Lemma 1. *Suppose Assumption 1 holds. Then, for any pair of agents $i, j \in [N]$, the stationary distributions $\pi^{(i)}$ and $\pi^{(j)}$ satisfy:*

$$\|\pi^{(i)} - \pi^{(j)}\|_1 \leq 2(n-1)\epsilon + \mathcal{O}(\epsilon^2). \tag{1}$$

We will now use the bound on $\|\pi^{(i)} - \pi^{(j)}\|_1$ in Lemma 1 to bound $\|\bar{A}_i - \bar{A}_j\|$ and $\|\bar{b}_i - \bar{b}_j\|$. To state our results, we make the standard assumption that for each $i \in [N]$, it holds that $|R^{(i)}(s)| \leq R_{\max}, \forall s \in \mathcal{S}$, i.e., the rewards are uniformly bounded. In [57], it was shown that $-\bar{A}_i$ is a negative definite matrix; thus, there exists some $\delta_1 > 0$ such that $\|\bar{A}_i\| \geq \delta_1$ holds for every agent $i \in [N]$. We also assume that there exists a constant $\delta_2 > 0$ such that $\|\bar{b}_i\| \geq \delta_2, \forall i \in [N]$. We have the following result on the perturbation of TD(0) fixed points.

Theorem 1. ***(Perturbation bounds on TD(0) fixed points)*** *For all $i, j \in [N]$, we have:*

(i) $\|\bar{A}_i - \bar{A}_j\| \leq A(\epsilon) \triangleq \gamma\sqrt{n}\epsilon + (1+\gamma)[2(n-1)\epsilon + \mathcal{O}(\epsilon^2)]$.

(ii) $\|\bar{b}_i - \bar{b}_j\| \leq b(\epsilon, \epsilon_1) \triangleq R_{\max}\left(2(n-1)\epsilon + \mathcal{O}(\epsilon^2)\right) + \mathcal{O}(\epsilon_1)$.

(iii) Suppose $\exists H > 0$ such that $\|\theta_i^*\| \leq H$, $\forall i \in [N]$. Let $\kappa(\bar{A}_i)$ denote the condition number of $\bar{A}_i$. Then:

$$\|\theta_i^* - \theta_j^*\| \leq \Gamma(\epsilon, \epsilon_1) \triangleq \max_{i \in [N]} \left\{ \frac{\kappa(\bar{A}_i) H}{1 - \kappa(\bar{A}_i)\frac{A(\epsilon)}{\delta_1}} \left(\frac{A(\epsilon)}{\delta_1} + \frac{b(\epsilon, \epsilon_1)}{\delta_2} \right) \right\}.$$

Discussion. Theorem 1 reveals how heterogeneity in the rewards and transition kernels of MDPs can be mapped to differences in the limiting behavior of TD(0) on such MDPs from a fixed-point perspective. It formalizes the intuition that if the level of heterogeneity - as captured by ϵ and ϵ_1 - is small, then so is the gap in the TD(0) limit points of the agents' MDPs. This result is novel, and complements similar perturbation results in the RL literature such as the *Simulation Lemma* [23].[3]

In what follows, we will introduce the key concept of a virtual MDP, and build on Theorem 1 to relate properties of this virtual MDP to those of the agents' individual MDPs.

3.2 Virtual Markov Decision Process

One of the main goals of our paper is to draw explicit parallels between federated optimization and FRL. Doing so would enable us to apply the rich set of ideas and techniques developed in standard FL to our setting. However, drawing such parallels requires some effort. In a standard FL setting, the goal is to typically minimize a global loss function $f(x) = (1/N)\sum_{i\in[N]} f_i(x)$ composed of the local loss functions of N agents; here, $f_i(x)$ is the local loss function of agent i. In FL, due to heterogeneity in the agents' loss functions, there is a "drift" effect [5, 22]: the local iterates of each agent i drift towards the minimizer of $f_i(x)$. However, when the heterogeneity is moderate, the average of the agents' iterates converges towards the minimizer of $f(x)$. To develop an analogous theory for FRL, we need to first answer: *When we average TD(0) update directions from different MDPs, where does the average TD(0) update direction lead us?* It is precisely to answer this question that we introduce the concept of a *virtual MDP*.

To model a virtual environment that captures the "average" of the agents' individual environments, we construct an MDP $\bar{\mathcal{M}} = \langle \mathcal{S}, \mathcal{A}, \bar{\mathcal{R}}, \bar{\mathcal{P}}, \gamma \rangle$, where $\bar{\mathcal{P}} = (1/N)\sum_{i=1}^N \mathcal{P}^{(i)}$, and $\bar{\mathcal{R}} = (1/N)\sum_{i=1}^N \mathcal{R}^{(i)}$. Note that the virtual MDP is a fictitious MDP that we construct solely for the purpose of analysis, and it may not coincide with any of the agents' MDPs, in general.

Properties of the Virtual MDP. When applied to $\bar{\mathcal{M}}$, let the policy μ that we seek to evaluate induce a virtual MRP characterized by the tuple $\{\bar{P}, \bar{R}\}$. It is easy to verify that $\bar{P} = (1/N)\sum_{i=1}^N P^{(i)}$, and $\bar{R} = (1/N)\sum_{i=1}^N R^{(i)}$. The following result shows how the virtual MRP inherits certain basic properties from the individual MRPs; the result is quite general and may be of independent interest.

Proposition 1. *Let $\{P^{(i)}\}_{i=1}^N$ be a set of Markov matrices associated with Markov chains that share the same states, and are each aperiodic and irreducible. Then, for any set of weights $\{w_i\}_{i=1}^N$ satisfying $w_i \geq 0, \forall i \in [N]$ and $\sum_{i\in[N]} w_i = 1$, the Markov chain corresponding to the matrix $\sum_{i\in[N]} w_i P^{(i)}$ is also aperiodic and irreducible.*

The above result immediately tells us that the Markov chain corresponding to $\bar{P}$ is aperiodic and irreducible. Thus, there exists an unique stationary distribution $\bar{\pi}$ of this Markov chain; let $\bar{D}$ be the corresponding diagonal matrix. As before, let us define $\bar{A} \triangleq \Phi^\top \bar{D}(\Phi - \gamma \bar{P}\Phi)$, $\bar{b} \triangleq \Phi^\top \bar{D}\bar{R}$, and use θ^* to denote the solution to the equation $\bar{A}\theta^* = \bar{b}$. Our next result is a consequence of Theorem 1, and characterizes the gap between θ_i^* and θ^*, for each $i \in [N]$.

Proposition 2. *Fix any $i \in [N]$. Using the same definitions as in Theorem 1, we have $\|\bar{A}_i - \bar{A}\| \leq A(\epsilon)$, $\|\bar{b}_i - \bar{b}\| \leq b(\epsilon, \epsilon_1)$ and $\|\theta_i^* - \theta^*\| \leq \Gamma(\epsilon, \epsilon_1)$.*

[3] The simulation lemma tells us that if two MDPs with the same state and action spaces are similar, then so are the value functions induced by a common policy on these MDPs.

Algorithm 1 Description of `FedTD`(0)

1: **Input:** Policy μ, local step-size α_l, global step-size $\alpha_g^{(t)}$ at t-th communication round
2: **Initialize:** $\bar{\theta}_0 = \theta_0$ and $s_{0,0}^{(i)} = s_0, \forall i \in [N]$
3: **for** each round $t = 0, \ldots, T-1$ **do**
4: **for** each agent $i \in [N]$ **do**
5: **for** $k = 0, \ldots, K-1$ **do**
6: Agent i initializes $\theta_{t,0}^{(i)} = \bar{\theta}_t$
7: Agent i plays $\mu(s_{t,k}^{(i)})$, observes tuple $O_{t,k}^{(i)} = (s_{t,k}^{(i)}, r_{t,k}^{(i)}, s_{t,k+1}^{(i)})$,
8: and updates local model as $\theta_{t,k+1}^{(i)} = \theta_{t,k}^{(i)} + \alpha_l g_i(\theta_{t,k}^{(i)})$,
9: where $g_i(\theta_{t,k}^{(i)}) \triangleq \left(r_{t,k}^{(i)} + \gamma\phi(s_{t,k+1}^{(i)})^\top \theta_{t,k}^{(i)} - \phi(s_{t,k}^{(i)})^\top \theta_{t,k}^{(i)}\right)\phi(s_{t,k}^{(i)})$
10: **end for**
11: send $\Delta_t^{(i)} = \theta_{t,K}^{(i)} - \bar{\theta}_t$ back to the server
12: **end for**
13: Server computes and broadcasts global model $\bar{\theta}_{t+1} = \Pi_{2,\mathcal{H}}\left(\bar{\theta}_t + \frac{\alpha_g^{(t)}}{N}\sum_{i\in[N]}\Delta_t^{(i)}\right)$
14: **end for**

We will later argue that the federated TD algorithm (to be introduced in Section 4) converges to a ball centered around the `TD`(0) fixed point θ^* of the virtual MRP. Proposition 2 is thus particularly important since it tells us that in a low-heterogeneity regime, by converging close to θ^*, we also converge close to the optimal parameter θ_i^* that minimizes the projected Bellman error for MDP $\mathcal{M}^{(i)}$. This justifies studying the convergence behavior of `FedTD`(0) on the virtual MRP. We end this section with a result which follows in part from Proposition 1.

Proposition 3. *For the virtual MRP, the following hold: (i)* $\lambda_{\max}(\Phi^\top \bar{D}\Phi) \leq 1$*; and (ii)* $\exists\ \bar{\omega} > 0$ *s.t.* $\lambda_{\min}(\Phi^\top \bar{D}\Phi) \geq \bar{\omega}$.

4 Federated TD Algorithm

In this section, we describe the `FedTD`(0) algorithm, a federated version of `TD`(0). We outline its steps in Algo. 1. The goal of `FedTD`(0) is to generate a model θ such that $\hat{V}_\theta$ is a good approximation of each agent i's value function $V_\mu^{(i)}$, corresponding to the policy μ. In line with both standard FL algorithms, and also works in MARL/FRL (in homogeneous settings) [11, 26], the agents keep their raw observations (i.e., their rewards, states, and actions) private, and only exchange local models.

`FedTD`(0) starts from a common initial model and a common starting state for all agents. Subsequently, in each round t, each agent $i \in [N]$ starts from a common model $\bar{\theta}_t$ and uses its local data to perform K local updates of the following form: at each local iteration k, each agent i takes action $\mu(s_{t,k}^{(i)})$ and observes a data tuple $O_{t,k}^{(i)} = \left(s_{t,k}^{(i)}, r_{t,k}^{(i)}, s_{t+1,k}^{(i)}\right)$ based on its *own* Markov reward process, i.e., $\{P^{(i)}, R^{(i)}\}$; we note here that *observations are independent across agents*. Using its data tuple $O_{t,k}^{(i)}$, each agent i updates its own local model $\theta_{t,k}^{(i)}$ along the direction $g_i(\theta_{t,k}^{(i)})$; see line 7.

Since each agent seeks to benefit from the samples acquired by the other agents, there is intermittent communication via the server. However, such communication needs to be limited as communication-efficiency is a key concern in FL. As such, the agents upload their local models' difference $\Delta_t^{(i)}$ to the server only once every K time-steps (line 11). On the server side, the model differences $\{\Delta_t^{(i)}\}$ are averaged, and a projection is carried out (line 13) to construct a global model

$\bar{\theta}_{t+1}$ that is then broadcast to all agents. Here, we use $\Pi_{2,\mathcal{H}}(\cdot)$ to denote the standard Euclidean projection on to a convex compact subset $\mathcal{H} \subset \mathbb{R}^d$ that is assumed to contain each $\theta_i^*, i \in [N]$, and also θ^*. Such a projection step on the server-side ensures that the global models do not blow up, and is common in the literature on stochastic approximation [3] and RL [2, 11]. Each agent then resumes its local updating process from this global model. We note that the structure of `FedTD`(0) mirrors that of `FedAvg` (and its many variants) where agents perform multiple local model-updates in isolation using their own data (to save communication), and synchronize periodically via a server. From another perspective, the `FedTD`(0) algorithm, which seeks to find the fixed point of the average of the TD update directions, can be grouped into the class of problems that seek to find fixed points using information from different sources [37]. However, there are significant differences in the dynamics of standard FL algorithms and `FedTD`(0), making it quite challenging to derive finite-time convergence results for the latter. We discuss some of these challenges below.

Challenges in Analysis. First, existing FL analyses are essentially distributed optimization proofs; although our setting bears a cosmetic connection to optimization, federated TD learning does not correspond to minimizing any fixed objective function. Second, unlike the FL setting where the data seen by each agent are drawn i.i.d. from some distribution, the data tuples observed by each agent in `FedTD`(0) are all part of one single Markovian trajectory. This creates complex time-correlations that are challenging to deal with even in a centralized setting with just one agent. Thus, we cannot directly appeal to standard FL proofs. Third, existing analyses in MARL/FRL that go beyond the simple tabular setting all end up assuming that every agent interacts with the *same* MDP, i.e., there is no heterogeneity effect at all to contend with in these works. Concretely, the analysis for `FedTD`(0) we provide in the subsequent sections is unique in that it simultaneously accounts for several key aspects: linear function approximation, Markovian sampling, multiple local updates, and heterogeneity in MDPs.

5 Analysis of the I.I.D. Setting

To isolate the effect of heterogeneity and provide key insights regarding our main proof ideas, we will analyze a simpler i.i.d. setting in this section. Specifically, we assume that for each agent $i \in [N]$, the data tuples $\{O_{t,k}^{(i)}\}$ are sampled i.i.d. from the stationary distribution $\pi^{(i)}$ of the Markov matrix $P^{(i)}$. Such an i.i.d assumption is common in the finite-time analysis of RL algorithms [9, 2, 11]. To proceed, for a fixed θ and for each $i \in [N]$, let us define $\bar{g}_i(\theta) \triangleq \mathbb{E}_{O_{t,k}^{(i)} \sim \pi^{(i)}}[g_i(\theta)]$ as the expected `TD`(0) update direction at iterate θ when the Markov tuple $O_{t,k}^{(i)}$ hits its stationary distribution $\pi^{(i)}$. We make the following standard bounded variance assumption [2]; similar assumptions are also made in FL analyses.

Assumption 4. $\mathbb{E}\|g_i(\theta) - \bar{g}_i(\theta)\|^2 \leq \sigma^2$ *holds for all agents* $i \in [N]$*, in each round* t *and local update* k*, and* $\forall \theta$.

Let H denote the radius of the set $\mathcal{H}$. Also, define $G \triangleq R_{\max} + 2H$ and $\nu = (1-\gamma)\bar{\omega}$, where $\bar{\omega}$ is as in Proposition 3. We can now state our first main result for `FedTD`(0).

Theorem 2. ***(I.I.D. Setting)*** *There exists a decreasing global step-size sequence* $\{\alpha_g^{(t)}\}$*, a fixed local step-size* α_l*, and a set of convex weights, such that a convex combination* $\tilde{\theta}_T$ *of the global models* $\{\bar{\theta}_t\}$ *satisfies the following for each* $i \in [N]$ *after* T *rounds:*

$$\mathbb{E}\left\|V_{\tilde{\theta}_T} - V_{\theta_i^*}\right\|_{\bar{D}}^2 \leq \tilde{\mathcal{O}}\Big(\frac{G^2}{K^2T^2} + \frac{\sigma^2}{\nu^2 NKT} + \frac{\sigma^2}{\nu^4 KT^2} + Q(\epsilon, \epsilon_1)\Big), \tag{2}$$

where $Q(\epsilon, \epsilon_1) = \tilde{\mathcal{O}}(\frac{B(\epsilon,\epsilon_1)G}{\nu} + \Gamma^2(\epsilon, \epsilon_1))$, $B(\epsilon, \epsilon_1) = H\left(\sqrt{n}\epsilon + 2(n-1)\epsilon + \mathcal{O}(\epsilon^2) + \mathcal{O}(\epsilon_1)\right)$, *and* $\Gamma(\epsilon, \epsilon_1)$ *is as defined in Theorem 1.*

There are several important messages conveyed by Theorem 2 that we now discuss.

Discussion. To parse Theorem 2, let us start by noting that the term $Q(\epsilon, \epsilon_1)$ in Eq. (2) captures the effect of heterogeneity; we will comment on this term later. When $T \gg N$, the dominant term among the first three terms in Eq. (2) is the $\sigma^2/(\nu^2 NKT)$ term. To appreciate the tightness of this term, we note that in a centralized setting (i.e., when $N = 1$), given access to KT samples, the convergence rate of `TD`(0) is $\sigma^2/(\nu^2 KT)$ [2]. Our analysis thus reveals that by communicating just T times in KT iterations, each agent i can reduce the noise variance σ^2 further by a factor of N, i.e., *achieve a linear speedup w.r.t. the number of agents.* In a low-heterogeneity regime, i.e., when $Q(\epsilon, \epsilon_1)$ is small, we note that by combining data from different MDPs, `FedTD`(0) guarantees *fast* convergence to a model that is a good approximation of each agent's value function; by fast, we imply a N-fold speedup over the rate each agent would have achieved had it not communicated at all. Thus with little communication, `FedTD`(0) quickly provides each agent with a good model that it can then fine-tune for personalization. Theorem 2 is the first result to provide such a guarantee in the context of MARL/FRL, and complements results of a similar flavor in FL [25, 62]. When all the MDPs are identical, $Q(\epsilon, \epsilon_1) = 0$. But when the MDPs are different, should we expect such a heterogeneity term?

To further understand the effect of heterogeneity, it suffices to get rid of all the randomness in our setting. As such, suppose we replace the random `TD`(0) direction $g_i(\theta_{t,k}^{(i)})$ of each agent i in Algo. 1 by its *steady-state* deterministic version $\bar{g}_i(\theta_{t,k}^{(i)}) = \bar{b}_i - \bar{A}_i\theta_{t,k}^{(i)}$, where $\bar{A}_i$ and $\bar{b}_i$ are as defined in Section 3.1. This leads to a deterministic version of `FedTD`(0) that we call *mean-path* `FedTD`(0). For simplicity, we assume that there is no projection step in *mean-path* `FedTD`(0). In our next result, we exploit the affine nature of the steady-state `TD`(0) directions to characterize the effect of heterogeneity in the limiting behavior of `FedTD`(0).

Theorem 3. ***(Heterogeneity Bias)*** *Suppose* $N = 2$ *and* $K = 1$. *Let the step-size* $\alpha = \alpha_g\alpha_l$ *be chosen such that* $I - \alpha\hat{A}$ *is Schur stable, where* $\hat{A} = \left(\bar{A}_1 + \bar{A}_2\right)/2$. *Define* $e_{i,t} \triangleq \bar{\theta}_t - \theta_i^*, i \in \{1, 2\}$. *The output of mean-path* ***`FedTD`****`(0)`* *then satisfies:*

$$\lim_{t\to\infty} e_{1,t} = \frac{1}{2}\hat{A}^{-1}\bar{A}_2(\theta_1^* - \theta_2^*); \quad \lim_{t\to\infty} e_{2,t} = \frac{1}{2}\hat{A}^{-1}\bar{A}_1(\theta_2^* - \theta_1^*). \tag{3}$$

Discussion: For the setting described in Theorem 3, the mean-path `FedTD`(0) updates follow the deterministic recursion $\bar{\theta}_{t+1} = (I - \alpha\hat{A})\bar{\theta}_t + \alpha\hat{b}$, where $\hat{b} = (1/2)(\bar{b}_1 + \bar{b}_2)$. This is a discrete-time linear time-invariant system (LTI). The dynamics of this system are stable if and only if the state transition matrix $(I - \alpha\hat{A})$ is Schur stable, justifying the choice of α in Theorem 3. The most important message conveyed by this result is that the gap between the limit point of mean-path `FedTD`(0) and the optimal parameter θ_i^* of either of the two MRPs bears a dependence on *the difference in the optimal parameters of the MRPs - a natural indicator of heterogeneity between the two MRPs.* Furthermore, this term has no dependence on the step-size α, i.e., the effect of the bias introduced by heterogeneity cannot be eliminated by making α arbitrarily small. Aligning with this observation, notice that the heterogeneity term $Q(\epsilon, \epsilon_1)$ in Eq. (2) is also step-size independent. The above discussion sheds some light on the fact that a term of the form $Q(\epsilon, \epsilon_1)$ is to be expected in Theorem 2. *Notably, the bias term in Eq. (3) persists even when the number of local steps is just one, i.e., even when the agents communicate with the server at all time steps.* This is a crucial difference with

the standard federated optimization setting where the effect of statistical heterogeneity manifests itself *only* when the number of local steps K is strictly larger than 1 [6, 22, 40].

We end this section with a proof sketch for Theorem 2.

Proof Sketch for Theorem 2. To make a connection to the existing FL optimization proofs, we start with a key observation made in [2]. In this paper, the authors showed that for each $i \in [N]$, the mean-path TD(0) direction $\bar{g}_i(\theta)$ acts like a pseudo-gradient and drives the iterates towards θ_i^*. Unfortunately, however, the average $(1/N)\sum_{i=1}^N \bar{g}_i(\theta)$ of the agents' mean-path TD(0) directions may not *exactly* correspond to the mean-path TD(0) direction of any MDP. Nonetheless, using Proposition 2, we prove the following key result that comes to our aid.

Lemma 2. ***(Expected pseudo-gradient heterogeneity)*** *For each $\theta \in \mathcal{H}$, we have:*

$$\left\|\bar{g}(\theta) - \frac{1}{N}\sum_{i=1}^N \bar{g}_i(\theta)\right\| \leq B(\epsilon, \epsilon_1), \tag{4}$$

where $B(\epsilon, \epsilon_1)$ is as in Theorem 2, and $\bar{g}(\theta)$ is the steady-state expected TD(0) direction of the virtual MDP.

Lemma 2 is crucial to our analysis as it shows that at least in the steady-state, the resulting FedTD(0) update direction can be closely approximated by the mean-path TD(0) direction of the virtual MDP. Furthermore, the latter acts like a pseudo-gradient pointing towards θ^* which is close to each θ_i^* based on Proposition 2. While this reasoning gives us hope, arriving at Eq. (2) requires a lot of work as we still need to (i) establish a linear-speedup in reducing the variance σ^2 in the noisy setting; and (ii) analyze a "client-drift" effect for our setting akin to what shows up in FL due to statistical heterogeneity and multiple local steps. In the Appendix, we provide a careful analysis that accounts for each of these issues.

6 Analysis of the Markovian Setting

Although the i.i.d. setting we discussed in Section 5 helped build a lot of intuition about the dynamics of FedTD(0), our main interest is in analyzing the setting where for each agent $i \in [N]$, the data tuples $\{O_{t,k}^{(i)}\}$ are all part of a *single Markovian trajectory* generated by $P^{(i)}$. The only assumption we will make is that these trajectories are independent across agents, i.e., the agents' observations are independent. Below, we briefly summarize some of the key difficulties that show up in the analysis for the Markovian setting, and that merit technical innovations on our part. To that end, let us write $g_i(\theta_{t,k}^{(i)})$ more explicitly as $g_i(\theta_{t,k}^{(i)}, O_{t,k}^{(i)})$; this will make certain statistical dependencies more transparent in our subsequent discussion.

Challenges in the Markovian analysis. First, our setting inherits all the difficulties in analyzing Markovian behavior from the centralized case [2]; in particular, for each $i \in [N]$, the parameter sequence $\{\theta_{t,k}^{(i)}\}$ and the data tuples $\{O_{t,k}^{(i)}\}$ are intricately coupled. Second, the synchronization step in FedTD(0) creates complex statistical dependencies between the local parameter of any given agent and the past observations of *all* other agents. Third, just as in the centralized case, we need to control the gradient bias $(1/NK)\sum_{i=1}^N \sum_{k=0}^{K-1} \left(g_i(\theta_{t,k}^{(i)}, O_{t,k}^{(i)}) - \bar{g}_i(\theta_{t,k}^{(i)})\right)$ and the gradient norm $\mathbb{E}\|(1/NK)\sum_{i=1}^N \sum_{k=0}^{K-1} g_i(\theta_{t,k}^{(i)})\|^2$. However to achieve the $\mathcal{O}(1/NKT)$-type rate, i.e., to prove linear speedup w.r.t. the number of agents N, we need to provide an analog of the variance reduction (i.e.,

the second term in Eq (2)) in the i.i.d. setting, which requires a much more delicate analysis relative to [2], since the observations of each agent i are correlated at different local steps. Indeed, naively bounding terms using the projection radius will not yield the linear speedup property. Finally, we need to control the "client-drift" effect (due to environmental heterogeneity) under the strong coupling between different random variables discussed above.

In our analysis, we will make use of the geometric mixing property of finite-state, aperiodic, and irreducible Markov chains [32]. Specifically, under Assumption 3, for each $i \in [N]$, there exists some $m_i \geq 1$ and $\rho_i \in (0,1)$, such that for all $t \geq 0$ and $0 \leq k \leq K-1$,

$$d_{TV}\left(\mathbb{P}\left(s_{t,k}^{(i)} = \cdot \mid s_{0,0}^{(i)} = s\right), \pi^{(i)}\right) \leq m_i \rho_i^{tK+k}, \forall s \in \mathcal{S},$$

holds, where we use $d_{TV}(P,Q)$ to denote the total-variation distance between two probability measures P and Q. For any $\bar{\epsilon} > 0$, let us define the mixing time for $P^{(i)}$ as $\tau_i^{\text{mix}}(\bar{\epsilon}) \triangleq \min\{t \in \mathbb{N}_0 \mid m_i \rho_i^t \leq \bar{\epsilon}\}$. Finally, let $\tau(\bar{\epsilon}) = \max_{i \in [N]} \tau_i^{\text{mix}}(\bar{\epsilon})$ represent the mixing time corresponding to the Markov chain that mixes the slowest. As one might expect, and as formalized in our main result below, it is this slowest-mixing Markov chain that dictates certain terms in the convergence rate of `FedTD`(0).

Theorem 4. *(**Markovian Setting**) There exists a decreasing global step-size sequence $\{\alpha_g^{(t)}\}$, a fixed local step-size α_l, and a set of convex weights, such that a convex combination $\tilde{\theta}_T$ of the global models $\{\bar{\theta}_t\}$ satisfies the following for each agent $i \in [N]$ after T rounds:*

$$\mathbb{E}\left\|V_{\tilde{\theta}_T} - V_{\theta_i^*}\right\|_{\bar{D}}^2 \leq \tilde{\mathcal{O}}\left(\frac{\tau^2 G^2}{K^2 T^2} + \frac{c_{quad}(\tau)}{\nu^2 NKT} + \frac{c_{lin}(\tau)}{\nu^4 KT^2} + Q(\epsilon, \epsilon_1)\right),$$

where $\tau = \lceil \frac{\tau^{\text{mix}}(\alpha_T^2)}{K} \rceil$, $\alpha_T = K\alpha_l \alpha_g^{(T)}$, $c_{quad}(\tau)$ and $c_{lin}(\tau)$ are quadratic and linear functions in τ, respectively, and $Q(\epsilon, \epsilon_1)$ is as defined in Theorem 2.

Discussion: Other than the effect of the mixing time τ which also shows up in a centralized setting [2], the rate in Theorem 4 mirrors that for the i.i.d. case in Theorem 2. *Theorem 4 is significant in that it marks the first comprehensive analysis of environmental heterogeneity in FRL under Markovian sampling.*

Proof Sketch for Theorem 4. As mentioned earlier, we cannot naively use a projection bound of the form $\mathbb{E}\|(1/NK)\sum_{i=1}^N \sum_{k=0}^{K-1} g_i(\theta_{t,k}^{(i)})\|^2 = \mathcal{O}(G^2)$ from the centralized analysis in [2] since the local models may not belong to the set $\mathcal{H}$. More importantly, going down that route will obscure the linear speedup effect. As such, we depart from the analysis techniques in [2, 50] by further decomposing the random TD direction of each agent i as $g_i(\theta_{t,k}^{(i)}) = b_i(O_{t,k}^{(i)}) - A_i(O_{t,k}^{(i)})\theta_{t,k}^{(i)}$. Since $A_i(O_{t,k}^{(i)})$ and $b_i(O_{t,k}^{(i)})$ only depend on the randomness from the Markov chain, and $O_{t,k}^{(i)}$ and $O_{t,k}^{(j)}$ are independent, we can show that the variances of $(1/NK)\sum_{i=1}^N \sum_{k=0}^{K-1} A_i(O_{t,k}^{(i)})$ and $(1/NK)\sum_{i=1}^N \sum_{k=0}^{K-1} b_i(O_{t,k}^{(i)})$ get scaled down by NK (up to higher order terms). Furthermore, to account for the fact that $A_i(O_{t,k}^{(i)})$ and $b_i(O_{t,k}^{(i)})$ differ across agents, we appeal to Lemma 2. Putting these pieces together in a careful manner yields the final rate in Theorem 4.

7 Conclusion

In this work, we have studied the problem of federated reinforcement learning under environmental heterogeneity and explored the question: *Can an agent expedite the process of learning its own*

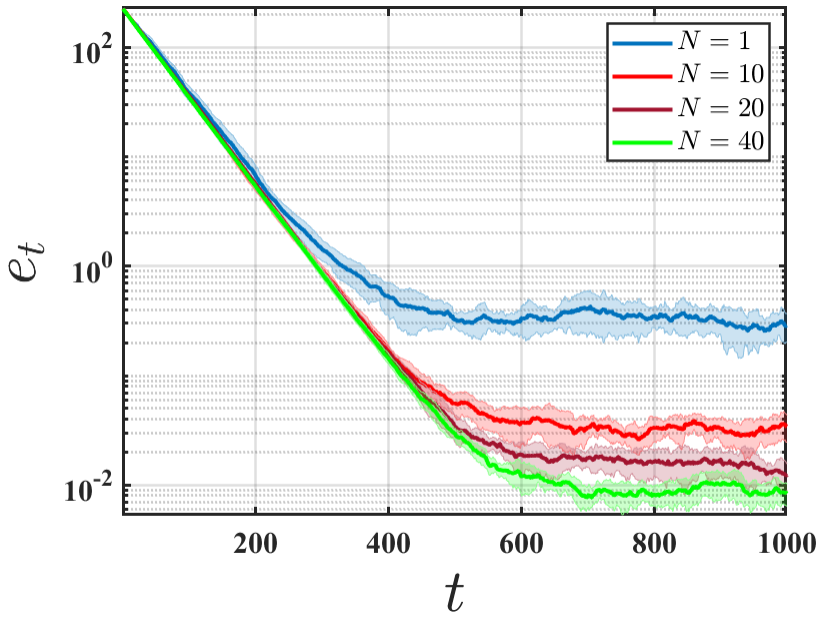

Figure 1: Performance of FedTD(0) under Markovian sampling with varying number of agents N. The MDP $\mathcal{M}^{(1)}$ of the first agent is randomly generated with a state space of size $n = 100$. The remaining MDPs are perturbations of $\mathcal{M}^{(1)}$ with the heterogeneity levels $\epsilon = 0.05$ and $\epsilon_1 = 0.1$. We evaluate the convergence in terms of the running error $e_t = \|\bar{\theta}_t - \theta_1^*\|^2$. Complying with theory, increasing N reduces this error. We choose the number of local steps as $K = 10$.

value function by using information from agents interacting with different MDPs? To answer this question, we studied the convergence of a federated TD(0) algorithm with linear function approximation, where N agents under different environments collaboratively evaluate a common policy. The main differences from the existing works are: (i) proposing a new definition of environmental heterogeneity; (ii) characterizing the effect of heterogeneity on TD(0) fixed points; (iii) introducing a virtual MDP to analyze the long-term behavior of the FedTD(0) algorithm; and (iv) making an explicit connection between federated reinforcement learning and federated supervised learning/optimization by leveraging the virtual MDP. With these elements, we proved that if the environmental heterogeneity between agents' environments is small, then FedTD(0) can achieve a linear speedup under both the i.i.d and the Markovian settings, and with multiple local updates.

A few interesting extensions to this work are as follows. First, it is natural to study federated variants of other RL algorithms beyond the TD(0) algorithm. Second, it would be interesting to investigate whether the personalization techniques in the traditional FL optimization literature can be applied to solve FedRL problems. Instead of learning a common value function/policy, can we design personalized value functions/policies that might perform better in high-heterogeneity regimes? We leave the exploration of this interesting question as future work.

References

[1] D. A. E. Acar, Y. Zhao, R. M. Navarro, M. Mattina, P. N. Whatmough, and V. Saligrama. Federated learning based on dynamic regularization. *arXiv preprint arXiv:2111.04263*, 2021.

[2] J. Bhandari, D. Russo, and R. Singal. A finite time analysis of temporal difference learning with linear function approximation. In *Conference on learning theory*, pages 1691–1692. PMLR, 2018.

[3] V. S. Borkar. *Stochastic approximation: a dynamical systems viewpoint*, volume 48. Springer, 2009.

[4] V. S. Borkar and S. P. Meyn. The ode method for convergence of stochastic approximation and reinforcement learning. *SIAM Journal on Control and Optimization*, 38(2):447–469, 2000.

[5] Z. Charles and J. Konečnỳ. On the outsized importance of learning rates in local update methods. *arXiv preprint arXiv:2007.00878*, 2020.

[6] Z. Charles and J. Konečnỳ. Convergence and Accuracy Trade-Offs in Federated Learning and Meta-Learning. In *International Conference on Artificial Intelligence and Statistics*, pages 2575–2583. PMLR, 2021.

[7] C. Chen, A. Seff, A. Kornhauser, and J. Xiao. Deepdriving: Learning affordance for direct perception in autonomous driving. In *Proceedings of the IEEE international conference on computer vision*, pages 2722–2730, 2015.

[8] L. Collins, H. Hassani, A. Mokhtari, and S. Shakkottai. Exploiting shared representations for personalized federated learning. In *International Conference on Machine Learning*, pages 2089–2099. PMLR, 2021.

[9] G. Dalal, B. Szörényi, G. Thoppe, and S. Mannor. Finite sample analyses for TD (0) with function approximation. In *Proceedings of the AAAI Conference on Artificial Intelligence*, volume 32, 2018.

[10] Y. Deng, M. M. Kamani, and M. Mahdavi. Adaptive personalized federated learning. *arXiv preprint arXiv:2003.13461*, 2020.

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

[12] A. Fallah, A. Mokhtari, and A. Ozdaglar. Personalized federated learning: A meta-learning approach. *arXiv preprint arXiv:2002.07948*, 2020.

[13] G. Frobenius, F. G. Frobenius, F. G. Frobenius, F. G. Frobenius, and G. Mathematician. Über matrizen aus nicht negativen elementen. 1912.

[14] A. Ghosh, J. Chung, D. Yin, and K. Ramchandran. An efficient framework for clustered federated learning. *Advances in Neural Information Processing Systems*, 33:19586–19597, 2020.

[15] E. Gorbunov, F. Hanzely, and P. Richtárik. Local SGD: Unified theory and new efficient methods. In *International Conference on Artificial Intelligence and Statistics*, pages 3556–3564. PMLR, 2021.

[16] F. Haddadpour, M. M. Kamani, M. Mahdavi, and V. Cadambe. Local SGD with periodic averaging: Tighter analysis and adaptive synchronization. In *Advances in Neural Information Processing Systems*, pages 11082–11094, 2019.

[17] F. Haddadpour and M. Mahdavi. On the convergence of local descent methods in federated learning. *arXiv preprint arXiv:1910.14425*, 2019.

[18] F. Hanzely, S. Hanzely, S. Horváth, and P. Richtárik. Lower bounds and optimal algorithms for personalized federated learning. *Advances in Neural Information Processing Systems*, 33:2304–2315, 2020.

[19] R. A. Horn and C. R. Johnson. *Matrix analysis*. Cambridge university press, 2012.

[20] X. Huang, D. Lee, E. Dobriban, and H. Hassani. Collaborative learning of discrete distributions under heterogeneity and communication constraints. In *Advances in Neural Information Processing Systems*, 2022.

[21] H. Jin, Y. Peng, W. Yang, S. Wang, and Z. Zhang. Federated Reinforcement Learning with Environment Heterogeneity. In *International Conference on Artificial Intelligence and Statistics*, pages 18–37. PMLR, 2022.

[22] S. P. Karimireddy, S. Kale, M. Mohri, S. Reddi, S. Stich, and A. T. Suresh. Scaffold: Stochastic controlled averaging for federated learning. In *International Conference on Machine Learning*, pages 5132–5143. PMLR, 2020.

[23] M. Kearns and S. Singh. Near-optimal reinforcement learning in polynomial time. *Machine learning*, 49(2):209–232, 2002.

[24] A. Khaled, K. Mishchenko, and P. Richtárik. First analysis of local GD on heterogeneous data. *arXiv preprint arXiv:1909.04715*, 2019.

[25] A. Khaled, K. Mishchenko, and P. Richtárik. Tighter theory for local SGD on identical and heterogeneous data. In *International Conference on Artificial Intelligence and Statistics*, pages 4519–4529. PMLR, 2020.

[26] S. Khodadadian, P. Sharma, G. Joshi, and S. T. Maguluri. Federated Reinforcement Learning: Linear Speedup Under Markovian Sampling. In *International Conference on Machine Learning*, pages 10997–11057. PMLR, 2022.

[27] A. Koloskova, N. Loizou, S. Boreiri, M. Jaggi, and S. U. Stich. A unified theory of decentralized SGD with changing topology and local updates. *arXiv preprint arXiv:2003.10422*, 2020.

[28] J. Konečnỳ, H. B. McMahan, D. Ramage, and P. Richtárik. Federated optimization: Distributed machine learning for on-device intelligence. *arXiv preprint arXiv:1610.02527*, 2016.

[29] N. Korda and P. La. On TD(0) with function approximation: Concentration bounds and a centered variant with exponential convergence. In *International conference on machine learning*, pages 626–634. PMLR, 2015.

[30] Y. Laguel, K. Pillutla, J. Malick, and Z. Harchaoui. A superquantile approach to federated learning with heterogeneous devices. In *2021 55th Annual Conference on Information Sciences and Systems (CISS)*, pages 1–6. IEEE, 2021.

[31] C. Lakshminarayanan and C. Szepesvári. Linear stochastic approximation: Constant step-size and iterate averaging. *arXiv preprint arXiv:1709.04073*, 2017.

[32] D. A. Levin and Y. Peres. *Markov chains and mixing times*, volume 107. American Mathematical Soc., 2017.

[33] X. Li, K. Huang, W. Yang, S. Wang, and Z. Zhang. On the convergence of fedavg on non-iid data. *arXiv preprint arXiv:1907.02189*, 2019.

[34] B. Liu, L. Wang, and M. Liu. Lifelong federated reinforcement learning: a learning architecture for navigation in cloud robotic systems. *IEEE Robotics and Automation Letters*, 4(4):4555–4562, 2019.

[35] R. Liu and A. Olshevsky. Distributed TD (0) with almost no communication. *arXiv preprint arXiv:2104.07855*, 2021.

[36] R. Liu and A. Olshevsky. Temporal difference learning as gradient splitting. In *International Conference on Machine Learning*, pages 6905–6913. PMLR, 2021.

[37] G. Malinovskiy, D. Kovalev, E. Gasanov, L. Condat, and P. Richtarik. From local SGD to local fixed-point methods for federated learning. In *International Conference on Machine Learning*, pages 6692–6701. PMLR, 2020.

[38] B. McMahan, E. Moore, D. Ramage, S. Hampson, and B. A. y Arcas. Communication-efficient learning of deep networks from decentralized data. In *Artificial Intelligence and Statistics*, pages 1273–1282. PMLR, 2017.

[39] K. Mishchenko, G. Malinovsky, S. Stich, and P. Richtárik. ProxSkip: Yes! Local Gradient Steps Provably Lead to Communication Acceleration! Finally! *arXiv preprint arXiv:2202.09357*, 2022.

[40] A. Mitra, R. Jaafar, G. J. Pappas, and H. Hassani. Linear convergence in federated learning: Tackling client heterogeneity and sparse gradients. *Advances in Neural Information Processing Systems*, 34:14606–14619, 2021.

[41] C. Narayanan and C. Szepesvári. Finite time bounds for temporal difference learning with function approximation: Problems with some "state-of-the-art" results. Technical report, Technical report, 2017.

[42] C. A. O'cinneide. Entrywise perturbation theory and error analysis for markov chains. *Numerische Mathematik*, 65(1):109–120, 1993.

[43] R. Pathak and M. J. Wainwright. FedSplit: An algorithmic framework for fast federated optimization. *arXiv preprint arXiv:2005.05238*, 2020.

[44] H. Pishro-Nik. Introduction to probability, statistics, and random processes. 2016.

[45] J. Qi, Q. Zhou, L. Lei, and K. Zheng. Federated reinforcement learning: techniques, applications, and open challenges. *arXiv preprint arXiv:2108.11887*, 2021.

[46] A. Reisizadeh, A. Mokhtari, H. Hassani, A. Jadbabaie, and R. Pedarsani. Fedpaq: A communication-efficient federated learning method with periodic averaging and quantization. In *International Conference on Artificial Intelligence and Statistics*, pages 2021–2031. PMLR, 2020.

[47] A. K. Sahu, T. Li, M. Sanjabi, M. Zaheer, A. Talwalkar, and V. Smith. On the convergence of federated optimization in heterogeneous networks. *arXiv preprint arXiv:1812.06127*, 3, 2018.

[48] F. Sattler, K.-R. Müller, and W. Samek. Clustered federated learning: Model-agnostic distributed multitask optimization under privacy constraints. *IEEE transactions on neural networks and learning systems*, 32(8):3710–3722, 2020.

[49] A. Spiridonoff, A. Olshevsky, and I. C. Paschalidis. Local SGD With a Communication Overhead Depending Only on the Number of Workers. *arXiv preprint arXiv:2006.02582*, 2020.

[50] R. Srikant and L. Ying. Finite-time error bounds for linear stochastic approximation and td learning. In *Conference on Learning Theory*, pages 2803–2830. PMLR, 2019.

[51] S. U. Stich. Local SGD converges fast and communicates little. *arXiv preprint arXiv:1805.09767*, 2018.

[52] L. Su, J. Xu, and P. Yang. Global convergence of federated learning for mixed regression. *arXiv preprint arXiv:2206.07279*, 2022.

[53] R. S. Sutton. Learning to predict by the methods of temporal differences. *Machine learning*, 3(1):9–44, 1988.

[54] R. S. Sutton, A. G. Barto, et al. Introduction to Reinforcement learning. 1998.

[55] C. T Dinh, N. Tran, and J. Nguyen. Personalized federated learning with moreau envelopes. *Advances in Neural Information Processing Systems*, 33:21394–21405, 2020.

[56] A. Z. Tan, H. Yu, L. Cui, and Q. Yang. Towards personalized federated learning. *IEEE Transactions on Neural Networks and Learning Systems*, 2022.

[57] J. N. Tsitsiklis and B. Van Roy. An analysis of temporal-difference learning with function approximation. In *IEEE Transactions on Automatic Control*, 1997.

[58] H. Wang, S. Marella, and J. Anderson. Fedadmm: A federated primal-dual algorithm allowing partial participation. In *2022 IEEE 61st Conference on Decision and Control (CDC)*, pages 287–294. IEEE, 2022.

[59] J. Wang and G. Joshi. Cooperative SGD: A unified framework for the design and analysis of communication-efficient SGD algorithms. *arXiv preprint arXiv:1808.07576*, 2018.

[60] X. Wang, Y. Han, C. Wang, Q. Zhao, X. Chen, and M. Chen. In-Edge AI: Intelligentizing mobile edge computing, caching and communication by federated learning. *IEEE Network*, 33(5):156–165, 2019.

[61] B. Woodworth, K. K. Patel, S. U. Stich, Z. Dai, B. Bullins, H. B. McMahan, O. Shamir, and N. Srebro. Is Local SGD Better than Minibatch SGD? *arXiv preprint arXiv:2002.07839*, 2020.

[62] B. E. Woodworth, K. K. Patel, and N. Srebro. Minibatch vs local SGD for heterogeneous distributed learning. *Advances in Neural Information Processing Systems*, 33:6281–6292, 2020.

Part I
Appendix

Table of Contents

A Outline

This appendix provides a detailed literature survey, supporting results, and full proofs for all theorems, lemmas, and propositions in the main text. A detailed survey of relevant works is provided in Section B. The proofs to Theorem 1, Propositions 1-3, and Lemma 2 are shown in sections C, D, and E respectively. In Section F, we provide some lemmas that are used in both the i.i.d. and Markovian sampling settings. In section G, we introduce some notations which are relevant to the proofs of the main theorems.

Our main result in the i.i.d. sampling regime is proven in Section H and involves several key sub-results involving (amongst other things) a variance reduction result, and bounding the "client-drift" term at each iteration. These results are provided in Section H.1 and the main result, Theorem 2 is proven in Section H.2.

The heterogeneity bias theorem, Theorem 3, is proven in Section I.

In Section 6, several key intermediate steps to proving Theorem 4 are given in subsections J.2.1-J.2.6, with the main result being proven in Section J.3. More simulation results are shown in Section K.

B Detailed Literature Survey

Federated Learning Algorithms. The literature on algorithmic developments in federated learning is vast; as such, we only cover some of the most relevant/representative works here. The most popularly used FL algorithm, `FedAvg`, was first introduced in [38]. Several works went on to provide a detailed theoretical analysis of `FedAvg` both in the homogeneous case when all clients minimize the same objective function [51, 59, 49, 46, 16, 61], and also in the more challenging heterogeneous setting [24, 25, 17, 33, 27]. In the latter scenario, it was soon realized that `FedAvg` suffers from a "client-drift" effect that hurts its convergence performance [5, 6, 22, 58].

Since then, a lot of effort has gone into improving the convergence guarantees of `FedAvg` via a variety of technical approaches: proximal methods in `FxedProx` [47]; operator-splitting in `FedSplit` [43]; variance-reduction in `Scaffold` [22] and `S-Local-SVRG` [15]; gradient-tracking in `FedLin` [40]; and dynamic regularization in [1]. While these methods improved upon `FedAvg` in various ways, they all fell short of providing any theoretical justification for performing multiple local updates under arbitrary statistical heterogeneity. Very recently, [39] introduced the `ProxSkip` algorithm, and showed that it can indeed lead to communication savings via multiple local steps, despite arbitrary heterogeneity.

Some other approaches to tackling heterogeneous statistical distributions in FL include personalization [10, 12, 55, 18, 56], clustering [14, 48, 52, 20], representation learning [8], and the use of quantiles [30].

Analysis of TD Learning Algorithms. The first work to provide a comprehensive asymptotic analysis of the temporal difference learning algorithm with value function approximation was [57]. In this work, the authors employed the ODE method [4] that is typically used to study asymptotic convergence rates of stochastic approximation algorithms. Providing finite-time bounds, however, turns out to be a much harder problem. Some early efforts in this direction were [29], [41], [9], and [31]. While these works were able to establish finite-time bounds for linear stochastic approximation algorithms (that subsume the TD learning algorithm), their analysis was limited to the i.i.d. sampling model. For the more challenging Markovian setting, finite-time rates have been recently derived using various perspectives: (i) by making explicit connections to optimization [2]; (ii) by taking a control-theoretic approach and studying the drift of a suitable Lyapunov function [50]; and (iii) by arguing that the mean-path temporal difference direction acts as a "gradient-splitting" of an appropriately chosen function [36]. Each of these interpretations provides interesting new insights into the dynamics of TD algorithms.

C Perturbation bounds for TD(0) fixed points

C.1 Proof of Theorem 1

In this section, we prove the perturbation bounds on TD(0) fixed points shown in Theorem 1. We start by observing that:

$$\begin{aligned}
\|\bar{A}_i - \bar{A}_j\| &= \|\Phi^\top D^{(i)}(\Phi - \gamma P^{(i)}\Phi) - \Phi^\top D^{(j)}(\Phi - \gamma P^{(j)}\Phi)\| \\
&\leq \|\Phi^\top D^{(i)}(\Phi - \gamma P^{(i)}\Phi) - \Phi^\top D^{(i)}(\Phi - \gamma P^{(j)}\Phi) + \Phi^\top D^{(i)}(\Phi - \gamma P^{(j)}\Phi) - \Phi^\top D^{(j)}(\Phi - \gamma P^{(j)}\Phi)\| \\
&\leq \|\Phi^\top D^{(i)}(\Phi - \gamma P^{(i)}\Phi) - \Phi^\top D^{(i)}(\Phi - \gamma P^{(j)}\Phi)\| \\
&+ \|\Phi^\top D^{(i)}(\Phi - \gamma P^{(j)}\Phi) - \Phi^\top D^{(j)}(\Phi - \gamma P^{(j)}\Phi)\| \\
&\overset{(a)}{\leq} \gamma\|\Phi\|^2\|D^{(i)}\|\|P^{(i)} - P^{(j)}\| + \|\Phi\|^2\|D^{(i)} - D^{(j)}\|\|(I - \gamma P^{(j)})\| \\
&\overset{(b)}{\leq} \gamma\sqrt{n}\epsilon + (1+\gamma)[2(n-1)\epsilon + \mathcal{O}(\epsilon^2)], \qquad (5)
\end{aligned}$$

where (a) follows from the triangle inequality. The first term in (b) uses the fact that $\|\Phi\| \leq 1$, $\|D^{(i)}\| \leq 1$, and

$$\|P^{(i)} - P^{(j)}\| \leq \sqrt{n}\|P^{(i)} - P^{(j)}\|_\infty \leq \epsilon\sqrt{n}\|P^{(i)}\|_\infty = \epsilon\sqrt{n},$$

where we use Assumption 1 in the second inequality. The second term in (b) uses the the facts that $\|I - \gamma P^{(j)}\| \leq 1 + \gamma$, $\|D^{(i)} - D^{(j)}\| \leq \|D^{(i)} - D^{(j)}\|_1 \leq \|\pi^{(i)} - \pi^{(j)}\|_1$, along with Lemma 1.

Next, we bound

$$\begin{aligned}
\|\bar{b}_i - \bar{b}_j\| &= \|\Phi D^{(i)} R^{(i)} - \Phi D^{(j)} R^{(j)}\| \\
&\leq \|\Phi D^{(i)} R^{(i)} - \Phi D^{(i)} R^{(j)}\| + \|\Phi D^{(i)} R^{(j)} - \Phi D^{(j)} R^{(j)}\| \\
&\leq \|\Phi\|\|D^{(i)}\|\|R^{(i)} - R^{(j)}\| + \|\Phi\|\|D^{(i)} - D^{(j)}\|\|R^{(j)}\| \\
&\leq \epsilon_1 + R_{\max}\left(2(n-1)\epsilon + \mathcal{O}(\epsilon^2)\right), \qquad (6)
\end{aligned}$$

where we use Assumption 2 in the last inequality and follow the same reasoning as we used to bound $\|\bar{A}_i - \bar{A}_j\|$ above.

We are now ready to bound the gap between fixed points as:

$$\frac{\|\theta_i^* - \theta_j^*\|}{\|\theta_i^*\|} \leq \frac{\kappa(\bar{A}_i)}{1 - \kappa(\bar{A}_i)\frac{\|\bar{A}_i - \bar{A}_j\|}{\|\bar{A}_i\|}}\left(\frac{\|\bar{A}_i - \bar{A}_j\|}{\|\bar{A}_i\|} + \frac{\|\bar{b}_i - \bar{b}_j\|}{\|\bar{b}_i\|}\right). \qquad (7)$$

Here, we leveraged the perturbation theory of linear equations in [19] Section 5.8. Finally, for any $\|\theta_i^*\| \leq H$, we have

$$\|\theta_i^* - \theta_j^*\| \leq \Gamma(\epsilon, \epsilon_1) \triangleq \frac{\kappa(\bar{A}_i)H}{1 - \kappa(\bar{A}_i)\frac{A(\epsilon)}{\delta_1}}\left(\frac{A(\epsilon)}{\delta_1} + \frac{b(\epsilon, \epsilon_1)}{\delta_2}\right),$$

where we used the fact that δ_1 and δ_2 are positive constants that lower bound $\|\bar{A}_i\|$ and $\|\bar{b}_i\|$, respectively.

D Properties of the Virtual Markov Decision Process

D.1 Proof of Proposition 1

Before we prove this proposition, we present the following fact from [44]: *a Markov matrix P is irreducible and aperiodic if and only if there exists a positive integer k such that every entry of the matrix P^k is strictly positive, i.e., $P^k_{s,s'} > 0$, for all $s, s' \in \mathcal{S}$.*

For every Markov matrix $P^{(i)}$, we know that there exists such an integer k_i according to the above fact and Assumption 3 in the paper. Then we define a set $J = \{i \in [N] | w_i > 0\}$. Since $\sum_{i=1}^N w_i = 1$, and $w_i \geq 0$ holds for all $i \in [N]$, we know that J is a non-empty set. If we define $\bar{k} = \min_{i\in[J]}\{k_i\}$ and $j = \arg\min_{i\in[J]}\{k_i\}$, then we have:

$$\left(\sum_{i\in[N]} w_i P^{(i)}\right)^{\bar{k}} = \underbrace{w_j^{\bar{k}}\left(P^{(j)}\right)^{\bar{k}}}_{\text{positive}} + \underbrace{\cdots\cdots}_{\text{nonnegative}}, \tag{8}$$

where each entry of $w_j^{\bar{k}}\left(P^{(j)}\right)^{\bar{k}}$ is strictly positive while the other matrices in the summation are non-negative. Thus, we can conclude that the Markov chain associated with the Markov matrix $\sum_{i\in[N]} w_i P^{(i)}$ is also irreducible and aperiodic.

D.2 Proof of Proposition 2

Following similar arguments as in Theorem 1, we bound $\|\bar{A}_i - \bar{A}\|$:

$$\begin{aligned}
\|\bar{A}_i - \bar{A}\| &= \|\Phi^\top D^{(i)}(\Phi - \gamma P^{(i)}\Phi) - \Phi^\top \bar{D}(\Phi - \gamma \bar{P}\Phi)\| \\
&\overset{(a)}{\leq} \gamma\|\Phi\|^2\|D^{(i)}\|\|P^{(i)} - \bar{P}\| + \|\Phi\|^2\|D^{(i)} - \bar{D}\|\|(I - \gamma\bar{P})\| \\
&\overset{(b)}{\leq} \gamma\sqrt{n}\epsilon + (1+\gamma)[2(n-1)\epsilon + \mathcal{O}(\epsilon^2)] = A(\epsilon),
\end{aligned} \tag{9}$$

where inequality (a) follows the same reasoning as (a) in Eq. (5), (b) uses the same fact as (b) in Eq. (5), and $\|P^{(i)} - \bar{P}\| \leq \frac{1}{N}\sum_{j=1}^N \|P^{(i)} - P^{(j)}\| \leq \epsilon\sqrt{n}$ and $\|D^{(i)} - \bar{D}\| \leq 2(n-1)\epsilon + \mathcal{O}(\epsilon^2)$.

Based on the above facts: (i) $\|\bar{R}\| \leq \frac{1}{N}\sum_{i=1}^N \|R^{(i)}\| \leq R_{\max}$, (ii) $\|R^{(i)} - \bar{R}\| \leq \frac{1}{N}\sum_{j=1}^N \|R^{(i)} - R^{(j)}\| \leq \epsilon_1$ and (iii) $\|D^{(i)} - \bar{D}\| \leq 2(n-1)\epsilon + \mathcal{O}(\epsilon^2)$, we finish the proof by showing that $\|\bar{b}_i - \bar{b}\| \leq b(\epsilon, \epsilon_1)$. To do so, we follow the same steps as Eq. (6), and prove the bound on $\|\theta_i^* - \theta^*\|$ by following the same analysis as Eq. (7).

D.3 Proof of Proposition 3

Since the virtual MDP is an average of the agents' MDPs, i.e., $\bar{P} = \frac{1}{N}\sum_{i=1}^N P^{(i)}$, the virtual Markov chain is irreducible and aperiodic from Proposition 1. The maximum eigenvalue of a symmetric positive-semidefinite matrix is a convex function. Then we have $\lambda_{\max}(\Phi^\top\bar{D}\Phi) \leq \sum_{s\in\mathcal{S}} \bar{\pi}(s)\lambda_{\max}\left(\phi(s)\phi(s)^\top\right) \leq \sum_{s\in\mathcal{S}}\bar{\pi}(s) = 1$.

To show that there exists $\omega > 0$ such that $\lambda_{\min}(\Phi^\top\bar{D}\Phi) \geq \omega > 0$, we will establish that $\Phi^\top\bar{D}\Phi$ is a positive-definite matrix. Since Φ is full-column rank, this amounts to showing that $\bar{D}$ is a positive definite matrix. From the definition of $\bar{D}$, establishing positive-definiteness of $\bar{D}$ is equivalent to arguing that every element of the stationary distribution vector $\bar{\pi}$ is strictly positive; here, $\bar{\pi}^\top\bar{P} = \bar{\pi}$. To that end, from Proposition 1, we know that the Markov chain associated with $\bar{P}$ is aperiodic and irreducible. From the Perron-Frobenius theorem [13], we conclude that indeed every entry of $\bar{\pi}$ is strictly positive. If we choose $\omega = \min_{s\in\mathcal{S}}\{\bar{\pi}(s)\} > 0$, we have $\lambda_{\min}(\Phi^\top\bar{D}\Phi) \geq \omega > 0$.

E Pseudo-gradient heterogeneity: Proof of Lemma 2

For each $\theta \in \mathcal{H}$, we have:

$$
\begin{aligned}
\Big\|\bar{g}(\theta) - \frac{1}{N}\sum_{i=1}^{N}\bar{g}_i(\theta)\Big\| &= \Big\|\Phi^T\bar{D}(\bar{T}_\mu\Phi\theta - \Phi\theta) - \frac{1}{N}\Big(\sum_{i=1}^{N}\Phi^T D^{(i)}(T_\mu^{(i)}\Phi\theta - \Phi\theta)\Big)\Big\| \\
&\overset{(a)}{\leq} \frac{1}{N}\sum_{i=1}^{N}\Big\|\Phi^T\bar{D}(\bar{T}_\mu\Phi\theta - \Phi\theta) - \Phi^T D^{(i)}(T_\mu^{(i)}\Phi\theta - \Phi\theta)\Big\| \\
&\overset{(b)}{\leq} \frac{1}{N}\sum_{i=1}^{N}\Big\|\bar{D}\Big[\frac{1}{N}\sum_{j=1}^{N}R^{(j)} + \gamma\frac{1}{N}\sum_{j=1}^{N}P^{(j)}\Phi\theta - \Phi\theta\Big] - D^{(i)}(T_\mu^{(i)}\Phi\theta - \Phi\theta)\Big\| \\
&\leq \frac{1}{N}\sum_{i=1}^{N}\Big\|\bar{D}\Big[\frac{1}{N}\sum_{j=1}^{N}R^{(j)} + \gamma\frac{1}{N}\sum_{j=1}^{N}P^{(j)}\Phi\theta - \Phi\theta\Big] - \bar{D}(T_\mu^{(i)}\Phi\theta - \Phi\theta) \\
&\quad + \bar{D}(T_\mu^{(i)}\Phi\theta - \Phi\theta) - D^{(i)}(T_\mu^{(i)}\Phi\theta - \Phi\theta)\Big\| \\
&\overset{(c)}{\leq} \frac{1}{N}\sum_{i=1}^{N}\Big\|\bar{D}\Big[\frac{1}{N}\sum_{j=1}^{N}R^{(j)} + \gamma\frac{1}{N}\sum_{j=1}^{N}P^{(j)}\Phi\theta - \Phi\theta\Big] - \bar{D}(T_\mu^{(i)}\Phi\theta - \Phi\theta)\Big\| \\
&+ \frac{1}{N}\sum_{i=1}^{N}\Big\|\bar{D}(T_\mu^{(i)}\Phi\theta - \Phi\theta) - D^{(i)}(T_\mu^{(i)}\Phi\theta - \Phi\theta)\Big\| \\
&\leq \frac{1}{N}\sum_{i=1}^{N}\Big\|\bar{D}\Big\|\Big\|\frac{1}{N}\sum_{j=1}^{N}R^{(j)} - R^{(i)}\Big\| + \gamma\Big\|\frac{1}{N}\sum_{j=1}^{N}P^{(j)} - P^{(i)}\Big\|\Big\|\Phi\theta\Big\| \\
&+ \frac{1}{N}\sum_{i=1}^{N}\Big\|\bar{D} - D^{(i)}\Big\|\Big\|T_\mu^{(i)}\Phi\theta - \Phi\theta\Big\| \\
&\overset{(d)}{\leq} \frac{1}{N}\sum_{i=1}^{N}\Big\|\frac{1}{N}\sum_{j=1}^{N}R^{(j)} - R^{(i)}\Big\|_2 + \gamma\Big\|\frac{1}{N}\sum_{j=1}^{N}P^{(j)} - P^{(i)}\Big\|\Big\|\Phi\theta\Big\| \\
&+ \frac{1}{N}\sum_{i=1}^{N}\Big\|\bar{D} - D^{(i)}\Big\|\Big\|T_\mu^{(i)}\Phi\theta - \Phi\theta\Big\| \\
&\overset{(e)}{\leq} \Big[\epsilon_1 + \gamma\sqrt{n}\epsilon\|\Phi\theta\| + \Big[2(n-1)\epsilon + \mathcal{O}(\epsilon^2)\Big]\|\Phi\theta\| \\
&\leq H\Big[\mathcal{O}(\epsilon_1) + \gamma\sqrt{n}\epsilon + 2(n-1)\epsilon + \mathcal{O}(\epsilon^2)\Big] = B(\epsilon, \epsilon_1). \qquad (10)
\end{aligned}
$$

Inequalities (a) and (c) follow from the triangle inequality, (b) is due to $\|\Phi\| \leq 1$; (d) is due to the fact that $\|\bar{D}\| \leq 1$; and (e) uses the following facts: (i) $\|R^{(i)} - \bar{R}\| \leq \epsilon_1$; (ii) $\|P^{(i)} - P^{(j)}\| \leq \sqrt{n}\|P^{(i)} - P^{(j)}\|_\infty \leq \epsilon\sqrt{n}\|P^{(i)}\|_\infty = \epsilon\sqrt{n}$, which, in turn, follows from the proof of Theorem 1; (iii) $\|D^{(i)} - \bar{D}\| \leq 2(n-1)\epsilon + \mathcal{O}(\epsilon^2)$, which, in turn, follows from the proof of Theorem 1 or Eq (5); and (iv) $\|\theta\| \leq H$ for any $\theta \in \mathcal{H}$.

F Auxiliary results used in the I.I.D. and Markovian settings

We make repeated use throughout the appendix (often without explicitly stating so) of the following inequalities:

- Given any two vectors $x, y \in \mathbb{R}^d$, for any $\beta > 0$, we have

$$\|x+y\|^2 \leq (1+\beta)\|x\|^2 + \left(1+\frac{1}{\beta}\right)\|y\|^2. \tag{11}$$

- Given any two vectors $x, y \in \mathbb{R}^d$, for any $\beta > 0$, we have

$$\langle x, y\rangle \leq \frac{\beta}{2}\|x\|^2 + \frac{1}{2\beta}\|y\|^2. \tag{12}$$

  This inequality goes by the name of Young's inequality.

- Given m vectors $x_1, \ldots, x_m \in \mathbb{R}^d$, the following is a simple application of Jensen's inequality:

$$\left\|\sum_{i=1}^m x_i\right\|^2 \leq m\sum_{i=1}^m \|x_i\|^2. \tag{13}$$

We prove the following result for the virtual MDP.

Lemma 3. *For any $\theta_1, \theta_2 \in \mathbb{R}^d$,*

$$(\theta_2-\theta_1)^\top [\bar{g}(\theta_1) - \bar{g}(\theta_2)] \geq (1-\gamma)\left\|\hat{V}_{\theta_1} - \hat{V}_{\theta_2}\right\|_{\bar{D}}^2. \tag{14}$$

Proof. Consider a stationary sequence of states with random initial state $s \sim \bar{\pi}$ and subsequent state s', which, conditioned on s, is drawn from $\bar{P}(\cdot \mid s)$. Define $\phi \triangleq \phi(s)$ and $\phi' \triangleq \phi(s')$. Define $\chi_1 \triangleq \hat{V}_{\theta_2}(s) - \hat{V}_{\theta_1}(s) = (\theta_2-\theta_1)^\top \phi$ and $\chi_2 \triangleq \hat{V}_{\theta_2}(s') - \hat{V}_{\theta_1}(s') = (\theta_2-\theta_1)^\top \phi'$. By stationarity, χ_1 and χ_2 are two correlated random variables with the same same marginal distribution. By definition, $\mathbb{E}[\chi_1^2] = \mathbb{E}[\chi_2^2] = \left\|\hat{V}_{\theta_2} - \hat{V}_{\theta_2}\right\|_{\bar{D}}^2$ since s, s' are drawn from $\bar{\pi}$. And we have,

$$\bar{g}(\theta_1) - \bar{g}(\theta_2) = \mathbb{E}\left[\phi\left(\gamma\phi' - \phi\right)^\top(\theta_1-\theta_2)\right] = \mathbb{E}\left[\phi\left(\chi_1 - \gamma\chi_2\right)\right].$$

Therefore,

$$\begin{aligned}
(\theta_2-\theta_1)^\top [\bar{g}(\theta_1) - \bar{g}(\theta_2)] &= \mathbb{E}\left[\chi_1\left(\chi_1 - \gamma\chi_2\right)\right] \\
&= \mathbb{E}\left[\chi_1^2\right] - \gamma\mathbb{E}\left[\chi_1\chi_2\right] \\
&\geq (1-\gamma)\mathbb{E}\left[\chi_1^2\right] \\
&= (1-\gamma)\left\|\hat{V}_{\theta_2} - \hat{V}_{\theta_2}\right\|_{\bar{D}}^2,
\end{aligned}$$

where we use the Cauchy-Schwartz inequality to conclude $\mathbb{E}[\chi_1\chi_2] \leq \sqrt{\mathbb{E}[\chi_1^2]}\sqrt{\mathbb{E}[\chi_2^2]} = \mathbb{E}[\chi_1^2]$. □

Lemma 4. *For any $\theta_1, \theta_2 \in \mathbb{R}^d$, we have*

$$\|\bar{g}(\theta_1) - \bar{g}(\theta_2)\| \leq 2\left\|\hat{V}_{\theta_1} - \hat{V}_{\theta_2}\right\|_{\bar{D}}. \tag{15}$$

Proof. Following the analysis of Lemma 3, we have

$$\|\bar{g}(\theta_1) - \bar{g}(\theta_2)\| = \|\mathbb{E}\left[\phi\left(\chi_1 - \gamma\chi_2\right)\right]\|$$

$$
\begin{aligned}
&\leq \sqrt{\mathbb{E}\left[\|\phi\|^2\right]}\sqrt{\mathbb{E}\left[\left(\chi_1-\gamma\chi_2\right)^2\right]} \\
&\leq \sqrt{\mathbb{E}\left[\chi_1^2\right]}+\gamma\sqrt{\mathbb{E}\left[\chi_2^2\right]} \\
&= (1+\gamma)\sqrt{\mathbb{E}\left[\chi_1^2\right]}, \qquad (16)
\end{aligned}
$$

where the second inequality is due to $\|\phi\| \leq 1$ and the final equality is due to $\mathbb{E}\left[\chi_1^2\right] = \mathbb{E}\left[\chi_2^2\right]$. We finish the proof by using the fact that $\mathbb{E}\left[\chi_1^2\right] = \left\|\hat{V}_{\theta_2} - \hat{V}_{\theta_2}\right\|_{\bar{D}}^2$ and $1+\gamma \leq 2$. □

With this Lemma, we next show that the steady-state TD(0) update direction $\bar{g}$ and $\bar{g}_i$ are 2-Lipschitz.

Lemma 5. *(2-Lipschitzness of steady-state TD(0) update direction) For any $\theta_1, \theta_2 \in \mathbb{R}^d$, we have*

$$\|\bar{g}(\theta_1) - \bar{g}(\theta_2)\| \leq 2\|\theta_1 - \theta_2\|. \qquad (17)$$

And for each agent $i \in [N]$, we have

$$\|\bar{g}_i(\theta_1) - \bar{g}_i(\theta_2)\| \leq 2\|\theta_1 - \theta_2\|. \qquad (18)$$

Proof. From Lemma 4, we can easily conclude that the steady-state TD(0) update direction $\bar{g}$ for the vitual MDP is 2-Lipschitz, i.e.,

$$\|\bar{g}(\theta_1) - \bar{g}(\theta_2)\| \leq 2\|\theta_1 - \theta_2\|, \qquad (19)$$

based on the fact that $\lambda_{\max}(\Phi^\top \bar{D}\Phi) \leq 1$. We can follow the same reasoning to prove Eq (18) since $\|\bar{g}_i(\theta_1) - \bar{g}_i(\theta_2)\| \leq 2\left\|\hat{V}_{\theta_1} - \hat{V}_{\theta_2}\right\|_{D_i}$ holds for each $i \in [N]$ from [2]. □

Next, we prove an analog of the Lipschitz property in Lemma 5 for the random TD(0) update direction of each agent i.

Lemma 6. *(2-Lipschitzness of random TD(0) update direction) For any $\theta_1, \theta_2 \in \mathbb{R}^d$ and $i \in [N]$, we have*

$$\|g_i(\theta_1) - g_i(\theta_2)\| \leq 2\|\theta_1 - \theta_2\|.$$

Proof. In this proof, we will use the fact that the random TD(0) update direction of agent i at the t-th communication round and k-th local update is an affine function of the parameter θ. In particular, we have $g_i(\theta) = b_i(O_{t,k}^{(i)}) - A_i(O_{t,k}^{(i)})\theta$, where $A_i(O_{t,k}^{(i)}) = \phi(s_{t,k}^{(i)})(\phi^\top(s_{t,k}^{(i)}) - \gamma\phi^\top(s_{t,k+1}^{(i)}))$ and $b_i(O_{t,k}^{(i)}) = r(s_{t,k}^{(i)})\phi(s_{t,k}^{(i)})$. Thus, we have

$$
\begin{aligned}
\|g_i(\theta_1) - g_i(\theta_2)\| &= \left\|A_i(O_{t,k}^{(i)})(\theta_1 - \theta_2)\right\| \\
&\leq \left\|A_i(O_{t,k}^{(i)})\right\|\|\theta_1 - \theta_2\| \\
&\leq \left(\left\|\phi\left(s_{t,k}^i\right)\right\|^2 + \gamma\left\|\phi\left(s_{t,k}^i\right)\right\|\left\|\phi\left(s_{t,k+1}^i\right)\right\|\right)\|\theta_1 - \theta_2\| \\
&\leq 2\|\theta_1 - \theta_2\|,
\end{aligned}
$$

where we used that $\|\phi(s)\| \leq 1, \forall s \in \mathcal{S}$ in the last step. □

G Notation

For our subsequent analysis, we will use $\mathcal{F}_k^t$ to denote the filtration that captures all the randomness up to the k-th local step in round t. We will also use $\mathcal{F}^t$ to represent the filtration capturing all the randomness up to the end of round $t-1$. With a slight abuse of notation, $\mathcal{F}_{-1}^t$ is to be interpreted as $\mathcal{F}^t$. Based on the description of `FedTD`(0), it should be apparent that for each $i \in [N]$, $\theta_{t,k}^{(i)}$ is $\mathcal{F}_{k-1}^t$-measurable and $\bar{\theta}_t$ is $\mathcal{F}^t$-measurable. Furthermore, we use $\mathbb{E}_t$ to represent the expectation conditioned on all the randomness up to the end of round $t-1$.

For simplicity, we define $\delta_t = \frac{1}{NK}\sum_{i=1}^{N}\sum_{k=0}^{K-1}\left\|\theta_{t,k}^{(i)} - \bar{\theta}_t\right\|$ and $\Delta_t = \frac{1}{NK}\sum_{i=1}^{N}\sum_{k=0}^{K-1}\left\|\theta_{t,k}^{(i)} - \bar{\theta}_t\right\|^2$. The latter term is referred to as the *drift term*. Note that $(\delta_t)^2 \leq \Delta_t$ holds for all t via Jensen's inequality. Unless specified otherwise, $\|\cdot\|$ denotes the Euclidean norm.

Step-size: Throughout the paper, we encounter three kinds of step-sizes: local step-size α_l, global step-size α_g, and the effective step-size α. Some of our results will rely on effective step-sizes that decay as a function of the communication round t; we will use $\{\alpha_t\}$ to represent such a decaying effective step-size sequence. While the local step-size α_ℓ will always be held constant, the decay in the effective step-size will be achieved by making the global step-size at the server decay with the communication round. Accordingly, we will use $\{\alpha_g^{(t)}\}$ to represent the decaying global step-size sequence at the server. In what follows, unless specified in the subscript, all the step-sizes appearing in the proofs refer to the effective step-size.

H Proof of the i.i.d. setting

H.1 Auxiliary lemmas for Theorem 2

H.1.1 Variance reduction

Lemma 7. *(Variance reduction in the i.i.d. setting). In the i.i.d. setting, under Assumption 4, at each round t, we have* $\mathbb{E}\left\|\frac{1}{NK}\sum_{i=1}^{N}\sum_{k=0}^{K-1}\left[g_i(\theta_{t,k}^{(i)})-\bar{g}_i(\theta_{t,k}^{(i)})\right]\right\|^2 \leq \frac{\sigma^2}{NK}$.

Proof. Define $Y_{t,k}^{(i)} \triangleq g_i(O_{t,k}^{(i)},\theta_{t,k}^{(i)})-\bar{g}_i(\theta_{t,k}^{(i)})$. Since $\{O_{t,k}^{(i)}\}$ is drawn i.i.d. over time from its stationary distribution $\pi^{(i)}$, we have $\mathbb{E}[Y_{t,k}^{(i)}]=\mathbb{E}\left[\mathbb{E}[Y_{t,k}^{(i)}\mid\theta_{t,k}^{(i)}]\right]=0$. As we mentioned before, for each $i\in[N]$, $\theta_{t,k}^{(i)}$ is $\mathcal{F}_{k-1}^{t}$-measurable. If we condition on $\mathcal{F}_{k-1}^{t}$, we know that $\theta_{t,k}^{(i)}$ and $\theta_{t,k}^{(j)}$ are deterministic and the only randomness in $Y_{t,k}^{(i)}$ and $Y_{t,k}^{(j)}$ come from $O_{t,k}^{(i)}$ and $O_{t,k}^{(j)}$, which are independent. Therefore, $Y_{t,k}^{(i)}$ and $Y_{t,k}^{(j)}$ are independent conditioned on $\mathcal{F}_{k-1}^{t}$.

For every $i\neq j\in[N]$, we have

$$\mathbb{E}\left[\left\langle Y_{t,k}^{(i)},Y_{t,k}^{(j)}\right\rangle\right]=\mathbb{E}\left[\mathbb{E}\left[\left\langle Y_{t,k}^{(i)},Y_{t,k}^{(j)}\right\rangle\mid\mathcal{F}_{k-1}^{t}\right]\right]\overset{(a)}{=}\mathbb{E}\left[\left\langle\mathbb{E}[Y_{t,k}^{(i)}\mid\mathcal{F}_{k-1}^{t}],\mathbb{E}[Y_{t,k}^{(j)}\mid\mathcal{F}_{k-1}^{t}]\right\rangle\right]=0, \tag{20}$$

where (a) follows from the fact that $Y_{t,k}^{(i)}$ and $Y_{t,k}^{(j)}$ are independent conditioned on $\mathcal{F}_{k-1}^{t}$. For every $k<l$ and $i,j\in[N]$,

$$\mathbb{E}\left[\left\langle Y_{t,k}^{(i)},Y_{t,l}^{(j)}\right\rangle\right]=\mathbb{E}\left[\mathbb{E}\left[\left\langle Y_{t,k}^{(i)},Y_{t,l}^{(j)}\right\rangle\,\middle|\,\mathcal{F}_{l-1}^{t}\right]\right]=\mathbb{E}\left[\left\langle Y_{t,k}^{(i)},\mathbb{E}[Y_{t,l}^{(j)}\mid\mathcal{F}_{l-1}^{t}]\right\rangle\right]=0. \tag{21}$$

Then,

$$\begin{aligned}
\mathbb{E}\left\|\frac{1}{NK}\sum_{i=1}^{N}\sum_{k=0}^{K-1}\left[g_i(\theta_{t,k}^{(i)})-\bar{g}_i(\theta_{t,k}^{(i)})\right]\right\|^2 &= \mathbb{E}\left\|\frac{1}{NK}\sum_{i=1}^{N}\sum_{k=0}^{K-1}Y_{t,k}^{(i)}\right\|^2 \\
&= \frac{1}{N^2K^2}\sum_{i=1}^{N}\sum_{k=0}^{K-1}\mathbb{E}\|Y_{t,k}^{(i)}\|^2+\frac{2}{N^2K^2}\underbrace{\sum_{i<j}\sum_{k=0}^{K-1}\mathbb{E}[\langle Y_{t,k}^{(i)},Y_{t,k}^{(j)}\rangle]}_{0} \\
&\quad+\frac{2}{N^2K^2}\sum_{i,j=1}^{N}\sum_{k<l}\underbrace{\mathbb{E}[\langle Y_{t,k}^{(i)},Y_{t,l}^{(j)}\rangle]}_{0} \\
&\leq \frac{\sigma^2}{NK},
\end{aligned}$$

where the second equality is due to Eq (20) and Eq (21) and the last inequality is due to Assumption 4. □

H.1.2 Per Round Progress

First, we characterize the error decrease at each iteration in the following lemma.

Lemma 8. *(Per Round Progress). If the local step-size α_l satisfies $\alpha_l\leq\frac{(1-\gamma)\bar{\omega}}{48K}$, then the updates of `FedTD(0)` with any global step-size α_g satisfy*

$$\mathbb{E}\|\bar{\theta}_{t+1}-\theta^*\|^2\leq(1+\zeta_1)\mathbb{E}\left\|\bar{\theta}_t-\theta^*\right\|^2+2\alpha\mathbb{E}\langle\bar{g}(\bar{\theta}_t),\bar{\theta}_t-\theta^*\rangle+6\alpha^2\mathbb{E}\left\|\bar{g}(\bar{\theta}_t)\right\|^2$$

$$+ 4\alpha^2\left(\frac{1}{\zeta_1}+6\right)\mathbb{E}[\Delta_t] + \frac{2\alpha^2\sigma^2}{NK} + 2\alpha B(\epsilon,\epsilon_1)G + 6\alpha^2 B^2(\epsilon,\epsilon_1), \tag{22}$$

where ζ_1 is any positive constant, and α is the effective step-size, i.e., $\alpha = K\alpha_l\alpha_g$.

Proof.

$$\begin{aligned}
\mathbb{E}\|\bar{\theta}_{t+1}-\theta^*\|^2 &= \mathbb{E}\Big\|\Pi_{2,\mathcal{H}}\left(\bar{\theta}_t + \frac{\alpha}{NK}\sum_{i=1}^{N}\sum_{k=0}^{K-1} g_i(\theta_{t,k}^{(i)}) - \theta^*\right)\Big\|^2 \quad \text{(updating rule)}\\
&\le \mathbb{E}\Big\|\bar{\theta}_t + \frac{\alpha}{NK}\sum_{i=1}^{N}\sum_{k=0}^{K-1} g_i(\theta_{t,k}^{(i)}) - \theta^*\Big\|^2 \quad \text{(projection is non-expansive)}\\
&= \mathbb{E}\Big\|\bar{\theta}_t-\theta^*\Big\|^2 + 2\mathbb{E}\langle\frac{\alpha}{NK}\sum_{i=1}^{N}\sum_{k=0}^{K-1} g_i(\theta_{t,k}^{(i)}), \bar{\theta}_t-\theta^*\rangle + \mathbb{E}\Big\|\frac{\alpha}{NK}\sum_{i=1}^{N}\sum_{k=0}^{K-1} g_i(\theta_{t,k}^{(i)})\Big\|^2\\
&= \mathbb{E}\Big\|\bar{\theta}_t-\theta^*\Big\|^2 + \frac{2\alpha}{NK}\underbrace{\sum_{i=1}^{N}\sum_{k=0}^{K-1}\mathbb{E}\langle g_i(\theta_{t,k}^{(i)}) - \bar{g}_i(\theta_{t,k}^{(i)}), \bar{\theta}_t-\theta^*\rangle}_{\mathcal{C}_1=0}\\
&+ \frac{2\alpha}{NK}\sum_{i=1}^{N}\sum_{k=0}^{K-1}\mathbb{E}\langle\bar{g}_i(\theta_{t,k}^{(i)}), \bar{\theta}_t-\theta^*\rangle + \mathbb{E}\Big\|\frac{\alpha}{NK}\sum_{i=1}^{N}\sum_{k=0}^{K-1} g_i(\theta_{t,k}^{(i)})\Big\|^2\\
&= \mathbb{E}\Big\|\bar{\theta}_t-\theta^*\Big\|^2 + \frac{2\alpha}{NK}\sum_{i=1}^{N}\sum_{k=0}^{K-1}\mathbb{E}\langle\bar{g}_i(\theta_{t,k}^{(i)}), \bar{\theta}_t-\theta^*\rangle + \mathbb{E}\Big\|\frac{\alpha}{NK}\sum_{i=1}^{N}\sum_{k=0}^{K-1} g_i(\theta_{t,k}^{(i)})\Big\|^2\\
&\le \mathbb{E}\Big\|\bar{\theta}_t-\theta^*\Big\|^2 + \frac{2\alpha}{NK}\sum_{i=1}^{N}\sum_{k=0}^{K-1}\mathbb{E}\langle\bar{g}_i(\theta_{t,k}^{(i)}), \bar{\theta}_t-\theta^*\rangle\\
&+ 2\mathbb{E}\Big\|\frac{\alpha}{NK}\sum_{i=1}^{N}\sum_{k=0}^{K-1}\left[g_i(\theta_{t,k}^{(i)}) - \bar{g}_i(\theta_{t,k}^{(i)})\right]\Big\|^2 + 2\mathbb{E}\Big\|\frac{\alpha}{NK}\sum_{i=1}^{N}\bar{g}_i(\theta_{t,k}^{(i)})\Big\|^2 \quad \text{(Young's inequality (12))}\\
&\overset{(a)}{\le} \mathbb{E}\Big\|\bar{\theta}_t-\theta^*\Big\|^2 + \frac{2\alpha}{NK}\sum_{i=1}^{N}\sum_{k=0}^{K-1}\mathbb{E}\langle\bar{g}_i(\theta_{t,k}^{(i)}), \bar{\theta}_t-\theta^*\rangle + \frac{2\sigma^2}{NK} + 2\mathbb{E}\Big\|\frac{\alpha}{NK}\sum_{i=1}^{N}\bar{g}_i(\theta_{t,k}^{(i)})\Big\|^2\\
&= \mathbb{E}\Big\|\bar{\theta}_t-\theta^*\Big\|^2 + \frac{2\alpha}{NK}\sum_{i=1}^{N}\sum_{k=0}^{K-1}\mathbb{E}\langle\bar{g}_i(\theta_{t,k}^{(i)}) - \bar{g}_i(\bar{\theta}_t) + \bar{g}_i(\bar{\theta}_t) - \bar{g}(\bar{\theta}_t) + \bar{g}(\bar{\theta}_t), \bar{\theta}_t-\theta^*\rangle \qquad (23)\\
&+ 2\mathbb{E}\Big\|\frac{\alpha}{NK}\sum_{i=1}^{N}\sum_{k=0}^{K-1}\bar{g}_i(\theta_{t,k}^{(i)})\Big\|^2 + \frac{2\alpha^2\sigma^2}{NK}\\
&\le \mathbb{E}\Big\|\bar{\theta}_t-\theta^*\Big\|^2 + \frac{2\alpha}{NK}\sum_{i=1}^{N}\sum_{k=0}^{K-1}\mathbb{E}\langle\bar{g}_i(\theta_{t,k}^{(i)}) - \bar{g}_i(\bar{\theta}_t), \bar{\theta}_t-\theta^*\rangle + \frac{2\alpha}{N}\sum_{i=1}^{N}\mathbb{E}\langle\bar{g}_i(\bar{\theta}_t) - \bar{g}(\bar{\theta}_t), \bar{\theta}_t-\theta^*\rangle\\
&+ 2\alpha\mathbb{E}\langle\bar{g}(\bar{\theta}_t), \bar{\theta}_t-\theta^*\rangle + 2\mathbb{E}\Big\|\frac{\alpha}{NK}\sum_{i=1}^{N}\sum_{k=0}^{K-1}\bar{g}_i(\theta_{t,k}^{(i)})\Big\|^2 + \frac{2\alpha^2\sigma^2}{NK}\\
&\le (1+\zeta_1)\mathbb{E}\Big\|\bar{\theta}_t-\theta^*\Big\|^2 + \frac{1}{\zeta_1}\mathbb{E}\Big\|\frac{\alpha}{NK}\sum_{i=1}^{N}\sum_{k=0}^{K-1}\left[\bar{g}_i(\theta_{t,k}^{(i)}) - \bar{g}_i(\bar{\theta}_t)\right]\Big\|^2 + 2\alpha B(\epsilon,\epsilon_1)G\\
&+ 2\alpha\mathbb{E}\langle\bar{g}(\bar{\theta}_t), \bar{\theta}_t-\theta^*\rangle + 2\mathbb{E}\Big\|\frac{\alpha}{NK}\sum_{i=1}^{N}\sum_{k=0}^{K-1}\bar{g}_i(\theta_{t,k}^{(i)})\Big\|^2 + \frac{2\alpha^2\sigma^2}{NK} \quad \text{(Eq (12) and Lemma 2)}
\end{aligned}$$

$$
\begin{aligned}
&\leq (1+\zeta_1)\mathbb{E}\left\|\bar{\theta}_t-\theta^*\right\|^2+\frac{4\alpha^2}{\zeta_1 NK}\sum_{i=1}^{N}\sum_{k=0}^{K-1}\mathbb{E}\left\|\theta_{t,k}^{(i)}-\bar{\theta}_t\right\|^2+2\alpha B(\epsilon,\epsilon_1)G\\
&+2\alpha\mathbb{E}\langle\bar{g}(\bar{\theta}_t),\bar{\theta}_t-\theta^*\rangle+2\mathbb{E}\left\|\frac{\alpha}{NK}\sum_{i=1}^{N}\sum_{k=0}^{K-1}\bar{g}_i(\theta_{t,k}^{(i)})\right\|^2+\frac{2\alpha^2\sigma^2}{NK}\quad\text{(2-Lipschitz of }\bar{g}_i\text{ in Lemma 5)}\\
&\leq (1+\zeta_1)\mathbb{E}\left\|\bar{\theta}_t-\theta^*\right\|^2+\frac{4\alpha^2}{\zeta_1}\mathbb{E}[\Delta_t]+\frac{2\alpha^2\sigma^2}{NK}+2\alpha B(\epsilon,\epsilon_1)G\\
&+2\alpha\mathbb{E}\langle\bar{g}(\bar{\theta}_t),\bar{\theta}_t-\theta^*\rangle+2\mathbb{E}\left\|\frac{\alpha}{NK}\sum_{i=1}^{N}\sum_{k=0}^{K-1}\left[\bar{g}_i(\theta_{t,k}^{(i)})-\bar{g}_i(\bar{\theta}_t)+\bar{g}_i(\bar{\theta}_t)-\bar{g}(\bar{\theta}_t)+\bar{g}(\bar{\theta}_t)\right]\right\|^2\\
&\leq (1+\zeta_1)\mathbb{E}\left\|\bar{\theta}_t-\theta^*\right\|^2+\frac{4\alpha^2}{\zeta_1}\mathbb{E}[\Delta_t]+\frac{2\alpha^2\sigma^2}{NK}+2\alpha B(\epsilon,\epsilon_1)G\\
&+2\alpha\mathbb{E}\langle\bar{g}(\bar{\theta}_t),\bar{\theta}_t-\theta^*\rangle+6\mathbb{E}\left\|\frac{\alpha}{NK}\sum_{i=1}^{N}\sum_{k=0}^{K-1}\bar{g}_i(\theta_{t,k}^{(i)})-\bar{g}_i(\bar{\theta}_t)\right\|^2\\
&+6\mathbb{E}\left\|\frac{\alpha}{N}\sum_{i=1}^{N}\left[\bar{g}_i(\bar{\theta}_t)-\bar{g}(\bar{\theta}_t)\right]\right\|^2+6\mathbb{E}\left\|\alpha\bar{g}(\bar{\theta}_t)\right\|^2\quad\text{(Eq (12) and Lemma 2)}\\
&\leq (1+\zeta_1)\mathbb{E}\left\|\bar{\theta}_t-\theta^*\right\|^2+\frac{4\alpha^2}{\zeta_1}\mathbb{E}[\Delta_t]+\frac{2\alpha^2\sigma^2}{NK}+2\alpha B(\epsilon,\epsilon_1)G\\
&+2\alpha\mathbb{E}\langle\bar{g}(\bar{\theta}_t),\bar{\theta}_t-\theta^*\rangle+24\alpha^2\mathbb{E}[\Delta_t]\quad\text{(2-Lipschitz of }\bar{g}_i\text{)}\\
&+6\alpha^2B^2(\epsilon,\epsilon_1)+6\alpha^2\mathbb{E}\left\|\bar{g}(\bar{\theta}_t)\right\|^2\quad\text{(Eq (12))}\\
&=(1+\zeta_1)\mathbb{E}\left\|\bar{\theta}_t-\theta^*\right\|^2+2\alpha\mathbb{E}\langle\bar{g}(\bar{\theta}_t),\bar{\theta}_t-\theta^*\rangle+6\alpha^2\mathbb{E}\left\|\bar{g}(\bar{\theta}_t)\right\|^2\\
&+4\alpha^2\left(\frac{1}{\zeta_1}+6\right)\mathbb{E}[\Delta_t]+\frac{2\alpha^2\sigma^2}{NK}+2\alpha B(\epsilon,\epsilon_1)G+6\alpha^2B^2(\epsilon,\epsilon_1),
\end{aligned}
\tag{24}
$$

where (a) is due to Lemma 7. Furthermore, the reason why $\mathcal{C}_1=0$ is as follows:

$$
\begin{aligned}
\mathcal{C}_1&=\sum_{i=1}^{N}\sum_{k=0}^{K-1}\mathbb{E}\langle g_i(\theta_{t,k}^{(i)})-\bar{g}_i(\theta_{t,k}^{(i)}),\bar{\theta}_t-\theta^*\rangle\\
&=\sum_{i=1}^{N}\sum_{k=0}^{K-2}\mathbb{E}\langle g_i(\theta_{t,k}^{(i)})-\bar{g}_i(\theta_{t,k}^{(i)}),\bar{\theta}_t-\theta^*\rangle+\sum_{i=1}^{N}\mathbb{E}\langle g_i(\theta_{t,K-1}^{(i)})-\bar{g}_i(\theta_{t,K-1}^{(i)}),\bar{\theta}_t-\theta^*\rangle\\
&=\sum_{i=1}^{N}\sum_{k=0}^{K-2}\mathbb{E}\langle g_i(\theta_{t,k}^{(i)})-\bar{g}_i(\theta_{t,k}^{(i)}),\bar{\theta}_t-\theta^*\rangle+\sum_{i=1}^{N}\mathbb{E}\left[\mathbb{E}\left[\langle g_i(\theta_{t,K-1}^{(i)})-\bar{g}_i(\theta_{t,K-1}^{(i)}),\bar{\theta}_t-\theta^*\rangle\mid\mathcal{F}_{K-1}^t\right]\right]\\
&=\sum_{i=1}^{N}\sum_{k=0}^{K-2}\mathbb{E}\langle g_i(\theta_{t,k}^{(i)})-\bar{g}_i(\theta_{t,k}^{(i)}),\bar{\theta}_t-\theta^*\rangle+\sum_{i=1}^{N}\mathbb{E}\left[\left\langle\bar{\theta}_t-\theta^*,\underbrace{\mathbb{E}\left[g_i(\theta_{t,k}^{(i)})-\bar{g}_i(\theta_{t,k}^{(i)})\mid\mathcal{F}_{K-1}^t\right]}_{0}\right\rangle\right]\\
&=\sum_{i=1}^{N}\sum_{k=0}^{K-2}\mathbb{E}\langle g_i(\theta_{t,k}^{(i)})-\bar{g}_i(\theta_{t,k}^{(i)}),\bar{\theta}_t-\theta^*\rangle.
\end{aligned}
$$

We can keep repeating this procedure by iteratively conditioning on $\mathcal{F}_{K-2}^t,\cdots,\mathcal{F}_1^t,\mathcal{F}_0^t$. □

H.1.3 Drift Term Analysis

We now turn to bounding the drift term Δ_t.

Lemma 9. *(Bounded Client Drift) The drift term Δ_t at the t-th round can be bounded as*

$$\mathbb{E}[\Delta_t] = \frac{1}{NK}\sum_{i=1}^{N}\sum_{k=0}^{K-1}\mathbb{E}\left\|\theta_{t,k}^{(i)} - \bar{\theta}_t\right\|^2 \leq 27(\sigma^2 + 3KB^2(\epsilon,\epsilon_1) + 2KG^2)\frac{\alpha^2}{K\alpha_g^2}, \tag{25}$$

provided the fixed local step-size α_l satisfies $\alpha_l \leq \min\frac{(1-\gamma)\bar{\omega}}{48K}$.

Proof.

$$\begin{aligned}
\mathbb{E}\left\|\theta_{t,k}^{(i)} - \bar{\theta}_t\right\|^2 &= \mathbb{E}\left\|\theta_{t,k-1}^{(i)} + \alpha_l g_i(\theta_{t,k-1}^{(i)}) - \bar{\theta}_t\right\|^2 \quad \text{(updating rule)} \\
&= \mathbb{E}\left\|\theta_{t,k-1}^{(i)} + \alpha_l \bar{g}_i(\theta_{t,k-1}^{(i)}) - \bar{\theta}_t + \alpha_l\left(g_i(\theta_{t,k-1}^{(i)}) - \bar{g}_i(\theta_{t,k-1}^{(i)})\right)\right\|^2 \\
&= \mathbb{E}\left\|\theta_{t,k-1}^{(i)} + \alpha_l \bar{g}_i(\theta_{t,k-1}^{(i)}) - \bar{\theta}_t\right\|^2 + \alpha_l^2\mathbb{E}\left\|g_i(\theta_{t,k-1}^{(i)}) - \bar{g}_i(\theta_{t,k-1}^{(i)})\right\|^2 \\
&+ \underbrace{2\alpha_l\,\mathbb{E}\left[\mathbb{E}\left\langle g_i(\theta_{t,k-1}^{(i)}) - \bar{g}_i(\theta_{t,k-1}^{(i)}), \theta_{t,k-1}^{(i)} + \alpha_l\bar{g}_i(\theta_{t,k-1}^{(i)}) - \bar{\theta}_t \;\middle|\; \mathcal{F}_{k-1}^t\right\rangle\right]}_{\mathcal{C}_2=0} \\
&\overset{(a)}{\leq} (1+\zeta_2)\mathbb{E}\left\|\theta_{t,k-1}^{(i)} + \alpha_l\bar{g}(\theta_{t,k-1}^{(i)}) - \bar{\theta}_t\right\|^2 + (1+\frac{1}{\zeta_2})\alpha_l^2\mathbb{E}\left\|\bar{g}(\theta_{t,k-1}^{(i)}) - \bar{g}_i(\theta_{t,k-1}^{(i)})\right\|^2 \\
&+ \alpha_l^2\mathbb{E}\left\|g_i(\theta_{t,k-1}^{(i)}) - \bar{g}_i(\theta_{t,k-1}^{(i)})\right\|^2 \\
&\overset{(b)}{\leq} (1+\zeta_2)(1+\zeta_3)\mathbb{E}\left\|\theta_{t,k-1}^{(i)} + \alpha_l\bar{g}(\theta_{t,k-1}^{(i)}) - \bar{\theta}_t - \alpha_l\bar{g}(\bar{\theta}_t)\right\|^2 + (1+\zeta_2)(1+\frac{1}{\zeta_3})\alpha_l^2\mathbb{E}\left\|\bar{g}(\bar{\theta}_t)\right\|^2 \\
&+ (1+\frac{1}{\zeta_2})\alpha_l^2\mathbb{E}\left\|\bar{g}(\theta_{t,k-1}^{(i)}) - \bar{g}(\bar{\theta}_t) + \bar{g}(\bar{\theta}_t) - \bar{g}_i(\bar{\theta}_t) + \bar{g}_i(\bar{\theta}_t) - \bar{g}_i(\theta_{t,k-1}^{(i)})\right\|^2 + \alpha_l^2\sigma^2 \\
&\overset{(c)}{\leq} (1+\zeta_2)(1+\zeta_3)\mathbb{E}\left\|\theta_{t,k-1}^{(i)} + \alpha_l\bar{g}(\theta_{t,k-1}^{(i)}) - \bar{\theta}_t - \alpha_l\bar{g}(\bar{\theta}_t)\right\|^2 + (1+\zeta_2)(1+\frac{1}{\zeta_3})\alpha_l^2\mathbb{E}\left\|\bar{g}(\bar{\theta}_t)\right\|^2 \\
&+ 3(1+\frac{1}{\zeta_2})\alpha_l^2\mathbb{E}\left\|\bar{g}(\theta_{t,k-1}^{(i)}) - \bar{g}(\bar{\theta}_t)\right\|^2 + 3(1+\frac{1}{\zeta_2})\alpha_l^2\mathbb{E}\left\|\bar{g}(\bar{\theta}_t) - \bar{g}_i(\bar{\theta}_t)\right\|^2 \\
&+ 3(1+\frac{1}{\zeta_2})\alpha_l^2\mathbb{E}\left\|\bar{g}_i(\bar{\theta}_t) - \bar{g}_i(\theta_{t,k-1}^{(i)})\right\|^2 + \alpha_l^2\sigma^2 \\
&\overset{(d)}{\leq} (1+\zeta_2)(1+\zeta_3)\left[1 - (2\alpha_l(1-\gamma) - 4\alpha_l^2)\bar{\omega}\right]\mathbb{E}\left\|\theta_{t,k-1}^{(i)} - \bar{\theta}_t\right\|^2 + (1+\zeta_2)(1+\frac{1}{\zeta_3})\alpha_l^2\mathbb{E}\left\|\bar{g}(\bar{\theta}_t)\right\|^2 \\
&+ 12(1+\frac{1}{\zeta_2})\alpha_l^2\mathbb{E}\left\|\theta_{t,k-1}^{(i)} - \bar{\theta}_t\right\|^2 + 3(1+\frac{1}{\zeta_3})\alpha_l^2B^2(\epsilon,\epsilon_1) + 12(1+\frac{1}{\zeta_3})\alpha_l^2\mathbb{E}\left\|\theta_{t,k-1}^{(i)} - \bar{\theta}_t\right\|^2 + \alpha_l^2\sigma^2 \\
&= (1+\zeta_2)(1+\zeta_3)\left[1 - (2\alpha_l(1-\gamma) - 4\alpha_l^2)\bar{\omega} + \frac{24(1+\frac{1}{\zeta_3})\alpha_l^2}{(1+\zeta_2)(1+\zeta_3)}\right]\mathbb{E}\left\|\theta_{t,k-1}^{(i)} - \bar{\theta}_t\right\|^2 \\
&+ \underbrace{(1+\zeta_2)(1+\frac{1}{\zeta_3})\alpha_l^2\mathbb{E}\left\|\bar{g}(\bar{\theta}_t)\right\|^2 + 3(1+\frac{1}{\zeta_3})\alpha_l^2B^2(\epsilon,\epsilon_1) + \alpha_l^2\sigma^2}_{\mathcal{D}_1},
\end{aligned}$$

where we used the inequality in Eq (11) with any positive constant ζ_2 for (a); for (b), we used Assumption 4 and the same reasoning as Eq (11) with any positive constant ζ_3; for (c), we used the inequality in Eq (13) to bound the third term; and for (d), we used Lemma 3 and Lemma 4 to bound the first term, the 2-Lipschitz property of $\bar{g}$, $\bar{g}_i$ (i.e., Lemma 5) in the third term and the fifth term, and the gradient heterogeneity bound from Lemma 2 in the fourth term. If we define

$\zeta_4 \triangleq (1+\zeta_2)(1+\zeta_3)\left[1-(2\alpha_l(1-\gamma)-4\alpha_l^2)\bar{\omega}+\frac{24(1+\frac{1}{\zeta_3})\alpha_l^2}{(1+\zeta_2)(1+\zeta_3)}\right]$ and define $\mathcal{D}_1$ as above, we have that

$$\mathbb{E}\left\|\theta_{t,k}^{(i)}-\bar{\theta}_t\right\|^2 \leq \zeta_4\mathbb{E}\left\|\theta_{t,k-1}^{(i)}-\bar{\theta}_t\right\|^2+\mathcal{D}_1. \tag{26}$$

Next, we set $\zeta_2=\zeta_3=\frac{1}{K-1}, K\geq 2$, and choose the local step-size α_l to satisfy

$$\frac{\alpha_l(1-\gamma)\bar{\omega}}{2}\geq 4\alpha_l^2\bar{\omega} \;\;\&\;\; \frac{\alpha_l(1-\gamma)\bar{\omega}}{2}\geq\frac{24(1+\frac{1}{\zeta_3})\alpha_l^2}{(1+\zeta_2)(1+\zeta_3)},$$

so that $\left[1-(2\alpha_l(1-\gamma)-4\alpha_l^2)\bar{\omega}+\frac{24(1+\frac{1}{\zeta_2})\alpha_l^2}{(1+\zeta_2)(1+\zeta_3)}\right]\leq 1-\alpha_l(1-\gamma)\bar{\omega}$. These inequalities hold when $\alpha_l\leq\min\frac{(1-\gamma)\bar{\omega}}{48K}$. Then, Eq (26) becomes

$$\mathbb{E}\left\|\theta_{t,k}^{(i)}-\bar{\theta}_t\right\|^2 \leq (1+\frac{3}{K-1})\left[1-\alpha_l(1-\gamma)\bar{\omega}\right]\mathbb{E}\left\|\theta_{t,k-1}^{(i)}-\bar{\theta}_t\right\|^2+\mathcal{D}_1.$$

If we unroll this recurrence above, using $\theta_{r,0}^{(i)}=\bar{\theta}_t$, we have that

$$\begin{aligned}
\mathbb{E}\left\|\theta_{t,k}^{(i)}-\bar{\theta}_t\right\|^2 &\leq \sum_{s=0}^{k-1}\mathcal{D}_1\left\{\Pi_{j=s+1}^{k-1}(1+\frac{3}{K-1})\left[1-\alpha(1-\gamma)\bar{\omega}\right]\right\}\\
&\overset{(e)}{\leq}\sum_{s=0}^{k-1}\left[\alpha_l^2\sigma^2+3K\alpha_l^2B^2(\epsilon,\epsilon_1),+2\alpha_l^2K\mathbb{E}\left\|\bar{g}(\bar{\theta}_t)\right\|^2\right]\times\Pi_{j=s+1}^{k-1}(1+\frac{3}{K-1})[1-\alpha_l(1-\gamma)\bar{\omega}]\\
&\leq\sum_{s=0}^{k-1}\left[\alpha_l^2\sigma^2+3\alpha_l^2KB^2(\epsilon,\epsilon_1)+2\alpha_l^2K\mathbb{E}\left\|\bar{g}(\bar{\theta}_t)\right\|^2\right](1+\frac{3}{K-1})^{K-1}\Pi_{j=s+1}^{k-1}[1-\alpha_l(1-\gamma)\bar{\omega}]\\
&\overset{(f)}{\leq} 27(\sigma^2+3KB^2(\epsilon,\epsilon_1)+2K\mathbb{E}\left\|\bar{g}(\bar{\theta}_t)\right\|^2)\sum_{s=0}^{k-1}\alpha_l^2\times\underbrace{\Pi_{j=s+1}^{k-1}[1-\alpha(1-\gamma)\bar{\omega}]}_{\leq 1}\\
&\leq 27(\sigma^2+3KB^2(\epsilon,\epsilon_1)+2KG^2)K\alpha_l^2 \quad \text{(constant local step-size)}
\end{aligned}$$

where we used the fact that $(1+\zeta_2)(1+\frac{1}{\zeta_3})\leq 2K$ for (e) and $(1+\frac{3}{K-1})^{K-1}\leq 27$ for (f). we finish the proof by substituting $\alpha_l=\frac{\alpha}{K\alpha_g}$. □

If we incorporate Eq (25) into Eq (22), we have that

$$\begin{aligned}
\mathbb{E}\left\|\bar{\theta}_{t+1}-\theta^*\right\|^2 &\leq (1+\zeta_1)\mathbb{E}\left\|\bar{\theta}_r-\theta^*\right\|^2+2\alpha\mathbb{E}\langle\bar{g}(\bar{\theta}_r),\bar{\theta}_r-\theta^*\rangle+6\alpha^2\mathbb{E}\left\|\bar{g}(\bar{\theta}_r)\right\|^2\\
&\quad+108\frac{\alpha^4}{K\alpha_g^2}(6+\frac{1}{\zeta_1})(\sigma^2+3KB^2(\epsilon,\epsilon_1)+2KG^2)+\frac{2\alpha^2\sigma^2}{NK}+2\alpha B(\epsilon,\epsilon_1)G+6\alpha^2B^2(\epsilon,\epsilon_1)
\end{aligned} \tag{27}$$

H.1.4 Parameter Selection

Lemma 10. *Define $\nu\triangleq(1-\gamma)\bar{\omega}$. If we choose any effective step-size $\alpha=K\alpha_g\alpha_l<\frac{(1-\gamma)\bar{\omega}}{96}$, any local step-size $\alpha_l\leq\min\frac{(1-\gamma)\bar{\omega}}{48K}$, and choose the constant $\zeta_1=\alpha\nu$, the updates of `FedTD(0)` satisfy*

$$\nu_1\mathbb{E}\left\|V_{\bar{\theta}_t}-V_{\theta^*}\right\|_{\bar{D}}^2\leq(\frac{1}{\alpha}-\nu_1)\mathbb{E}\left\|\bar{\theta}_t-\theta^*\right\|^2-\frac{1}{\alpha}\mathbb{E}\left\|\bar{\theta}_{t+1}-\theta^*\right\|^2+\underbrace{\frac{2\alpha\sigma^2}{NK}}_{O(\alpha^1)}$$

$$+\underbrace{\frac{1080\alpha^2}{K\alpha_g^2\nu}(\sigma^2+3KB^2(\epsilon,\epsilon_1)+2KG^2)}_{O(\alpha^2)}+\underbrace{2B(\epsilon,\epsilon_1)G+6\alpha B^2(\epsilon,\epsilon_1)}_{\textit{heterogeneity term}}, \tag{28}$$

where $\nu_1=\frac{\nu}{4}=\frac{(1-\gamma)\bar{\omega}}{4}$.

Proof. From Eq (27) and $\zeta_1=\alpha\nu$, we know

$$\begin{aligned}
\mathbb{E}\left\|\bar{\theta}_{t+1}-\theta^*\right\|^2 &\le (1+\zeta_1)\mathbb{E}\left\|\bar{\theta}_t-\theta^*\right\|^2+2\alpha\mathbb{E}\langle\bar{g}(\bar{\theta}_t),\bar{\theta}_t-\theta^*\rangle+6\alpha^2\mathbb{E}\left\|\bar{g}(\bar{\theta}_r)\right\|^2\\
&+108\frac{\alpha^4}{K\alpha_g^2}(6+\frac{1}{\zeta_1})(\sigma^2+3KB^2(\epsilon,\epsilon_1)+2KG^2)+\frac{2\alpha^2\sigma^2}{NK}+2\alpha B(\epsilon,\epsilon_1)G+6\alpha^2B^2(\epsilon,\epsilon_1)\\
&\le (1+\alpha\nu-2\alpha\nu)\mathbb{E}\left\|\bar{\theta}_t-\theta^*\right\|^2+24\alpha^2\mathbb{E}\left\|V_{\bar{\theta}_t}-V_{\theta^*}\right\|_{\bar{D}}^2+\frac{2\alpha^2\sigma^2}{NK}\quad\text{(Lemma 3 and 4)}\\
&+108\frac{\alpha^4}{K\alpha_g^2}(6+\frac{1}{\alpha\nu})(\sigma^2+3KB^2(\epsilon,\epsilon_1)+2KG^2)+2\alpha B(\epsilon,\epsilon_1)G+6\alpha^2B^2(\epsilon,\epsilon_1)\\
&\le (1-\frac{\alpha\nu}{2})\mathbb{E}\left\|\bar{\theta}_t-\theta^*\right\|^2-\frac{\alpha\nu}{2}\mathbb{E}\left\|\bar{\theta}_t-\theta^*\right\|^2+24\alpha^2\mathbb{E}\left\|V_{\bar{\theta}_t}-V_{\theta^*}\right\|_{\bar{D}}^2+\frac{2\alpha^2\sigma^2}{NK}\\
&+108\frac{\alpha^4}{K\alpha_g^2}(6+\frac{1}{\alpha\nu})(\sigma^2+3KB^2(\epsilon,\epsilon_1)+2KG^2)+2\alpha B(\epsilon,\epsilon_1)G+6\alpha^2B^2(\epsilon,\epsilon_1)\\
&\overset{(a)}{\le} (1-\frac{\alpha\nu}{2})\mathbb{E}\left\|\bar{\theta}_t-\theta^*\right\|^2-\frac{\alpha\nu}{2}\mathbb{E}\left\|V_{\bar{\theta}_t}-V_{\theta^*}\right\|_{\bar{D}}^2+\frac{\alpha\nu}{4}\mathbb{E}\left\|V_{\bar{\theta}_t}-V_{\theta^*}\right\|_{\bar{D}}^2+\frac{2\alpha^2\sigma^2}{NK}\\
&+108\frac{\alpha^4}{K\alpha_g^2}(6+\frac{1}{\alpha\nu})(\sigma^2+3KB^2(\epsilon,\epsilon_1)+2KG^2)+2\alpha B(\epsilon,\epsilon_1)G+6\alpha^2B^2(\epsilon,\epsilon_1),
\end{aligned}$$

where (a) comes from $\lambda_{\max}(\Phi^T\bar{D}\Phi)\le 1$ and $24\alpha^2\le 24\alpha\frac{(1-\gamma)\bar{w}}{96}=\frac{\alpha\nu}{4}$. Moving $\mathbb{E}\left\|V_{\bar{\theta}_t}-V_{\theta^*}\right\|_{\bar{D}}^2$ (on the right-hand side of (a)) to the left hand side of the above inequality yields:

$$\begin{aligned}
\frac{\alpha\nu}{4}\mathbb{E}\left\|V_{\bar{\theta}_t}-V_{\theta^*}\right\|_{\bar{D}}^2 &\le (1-\frac{\alpha\nu}{2})\mathbb{E}\left\|\bar{\theta}_t-\theta^*\right\|^2-\mathbb{E}\left\|\bar{\theta}_{t+1}-\theta^*\right\|^2+\frac{2\alpha^2\sigma^2}{NK}\\
&+108(\frac{6\alpha^4}{K\alpha_g^2}+\frac{\alpha^3}{K\alpha_g^2\nu})(\sigma^2+3KB^2(\epsilon,\epsilon_1)+2KG^2)+2\alpha B(\epsilon,\epsilon_1)G+6\alpha^2B^2(\epsilon,\epsilon_1).
\end{aligned}$$

Dividing by α on both sides of the inequality above and changing ν into ν_1, we have:

$$\begin{aligned}
\nu_1\mathbb{E}\left\|V_{\bar{\theta}_t}-V_{\theta^*}\right\|_{\bar{D}}^2 &\le (\frac{1}{\alpha}-\nu_1)\mathbb{E}\left\|\bar{\theta}_t-\theta^*\right\|^2-\frac{1}{\alpha}\mathbb{E}\left\|\bar{\theta}_{t+1}-\theta^*\right\|^2+\frac{2\alpha\sigma^2}{NK}\\
&+108(\frac{6\alpha^3}{K\alpha_g^2}+\frac{4\alpha^2}{K\alpha_g^2\nu_1})(\sigma^2+3KB^2(\epsilon,\epsilon_1)+2KG^2)+2B(\epsilon,\epsilon_1)G+6\alpha B^2(\epsilon,\epsilon_1)\\
&\le (\frac{1}{\alpha}-\nu_1)\mathbb{E}\left\|\bar{\theta}_t-\theta^*\right\|^2-\frac{1}{\alpha}\mathbb{E}\left\|\bar{\theta}_{t+1}-\theta^*\right\|^2+\underbrace{\frac{2\alpha\sigma^2}{NK}}_{O(\alpha^1)}\\
&+\underbrace{\frac{1080\alpha^2}{K\alpha_g^2\nu_1}(\sigma^2+3KB^2(\epsilon,\epsilon_1)+2KG^2)}_{O(\alpha^2)}+\underbrace{2B(\epsilon,\epsilon_1)G+6\alpha B^2(\epsilon,\epsilon_1)}_{\text{heterogeneity term}},
\end{aligned}$$

where we used the fact that $\alpha\le 1$ in the last inequality. □

With these lemmas, we are now ready to prove Theorem 2, which we restate for clarity.

H.2 Proof of Theorem 2

Given a fixed local step-size $\alpha_l = \frac{1}{2}\frac{(1-\gamma)\bar{\omega}}{48K}$, decreasing effective step-sizes $\alpha_t = \frac{8}{\nu(a+t+1)} = \frac{8}{(1-\gamma)\bar{\omega}(a+t+1)}$, decreasing global step-sizes $\alpha_g^{(t)} = \frac{\alpha_t}{K\alpha_l}$, and weights $w_t = (a+t)$, we have that

$$\mathbb{E}\left\|V_{\tilde{\theta}_T} - V_{\theta_i^*}\right\|_{\bar{D}}^2 \leq \tilde{\mathcal{O}}\left(\frac{G^2}{K^2T^2} + \frac{\sigma^2}{\nu^4KT^2} + \frac{\sigma^2}{\nu^2NKT} + \frac{B(\epsilon,\epsilon_1)G}{\nu} + \Gamma^2(\epsilon,\epsilon_1)\right) \tag{29}$$

holds for any agent $i \in [N]$.

Proof. We take the effective step-size $\alpha_t = \frac{8}{\nu(a+t+1)} = \frac{2}{\nu_1(a+t+1)}$ for $a > 0$. In addition, we define weights $w_t = (a+t)$ and define

$$\tilde{\theta}_T = \frac{1}{W}\sum_{t=1}^{T} w_t\bar{\theta}_t,$$

where $W = \sum_{t=1}^{T} w_t \geq \frac{1}{2}T(a+T)$. By convexity of positive definite quadratic forms ($\lambda_{\min}(\Phi^T\bar{D}\Phi) \geq \bar{\omega} > 0$), we have that

$$\begin{aligned}
\nu_1\mathbb{E}\left\|V_{\tilde{\theta}_T} - V_{\theta^*}\right\|_{\bar{D}}^2 &\leq \frac{\nu_1}{W}\sum_{t=1}^{T}(a+t)\mathbb{E}\left\|V_{\bar{\theta}_t} - V_{\theta^*}\right\|_{\bar{D}}^2 \\
&\overset{(28)}{\leq} \frac{\nu_1(a+1)(a+2)G^2}{2W} + \frac{1}{W}\sum_{t=1}^{T}\left[\frac{2(a+t)\alpha_t}{NK}\sigma^2\right] \\
&+ \frac{1}{W}\sum_{t=1}^{T}\left[\frac{1080(a+t)\alpha_t^2}{K\alpha_g^2\nu_1}(\sigma^2 + 3KB^2(\epsilon,\epsilon_1) + 2KG^2)\right] \\
&+ \frac{1}{W}\sum_{t=1}^{T}(a+t)\left[2B(\epsilon,\epsilon_1)G + 6\alpha_tB^2(\epsilon,\epsilon_1)\right] \\
&\leq \frac{\nu_1(a+1)(a+2)G^2}{2W} + \frac{2\sigma^2}{NKW}\sum_{t=1}^{T}(a+t)\alpha_t \\
&+ \frac{1080(\sigma^2 + 3KB^2(\epsilon,\epsilon_1) + 2KG^2)}{K\alpha_g^2\nu_1W}\sum_{t=1}^{T}(a+t)\alpha_t^2 + 2B(\epsilon,\epsilon_1)G + \frac{6B^2(\epsilon,\epsilon_1)}{W}\sum_{t=1}^{T}(a+t)\alpha_t \\
&\leq \frac{\nu_1(a+1)(a+2)G^2}{2W} + \frac{4\sigma^2}{\nu_1NKW}\cdot T \\
&+ \frac{4320(\sigma^2 + 3KB^2(\epsilon,\epsilon_1) + 2KG^2)}{K\alpha_g^2\nu_1^3W}\cdot(1 + \log(a+T)) + 2B(\epsilon,\epsilon_1)G + \frac{12B^2(\epsilon,\epsilon_1)}{\nu_1W}\cdot T,
\end{aligned}$$

where we used $\left\|V_{\bar{\theta}_0} - V_{\theta^*}\right\|_{\bar{D}}^2 \leq G^2$. Dividing by ν_1 on both sides, changing ν_1 into ν, and using $W \geq \frac{T(a+T)}{2}$, we have:

$$\mathbb{E}\left\|V_{\tilde{\theta}_T} - V_{\theta^*}\right\|_{\bar{D}}^2 \leq \tilde{\mathcal{O}}\left(\frac{G^2}{K^2T^2} + \frac{\sigma^2}{\nu^4KT^2} + \frac{\sigma^2}{\nu^2NKT} + \frac{B(\epsilon,\epsilon_1)G}{\nu}\right).$$

We finish the proof by using the following inequality: $\mathbb{E}\left\|V_{\tilde{\theta}_T} - V_{\theta_i^*}\right\|_{\bar{D}}^2 \leq 2\mathbb{E}\left\|V_{\tilde{\theta}_T} - V_{\theta^*}\right\|_{\bar{D}}^2 + 2\mathbb{E}\left\|V_{\theta_i^*} - V_{\theta^*}\right\|_{\bar{D}}^2$, in tandem with the third point in Theorem 1. □

I Heterogeneity bias: Proof of Theorem 3

In this section, we prove Theorem 3.

Proof of Theorem 3. As θ_1^* and θ_2^* are the TD(0) fixed points of agents 1 and 2, respectively, we have $\theta_1^* = \bar{A}_1^{-1}\bar{b}_1$ and $\theta_2^* = \bar{A}_2^{-1}\bar{b}_2$ from Section 3.1. The output of mean-path `FedTD`(0) with $k = 1$ and $\alpha = \alpha_g\alpha_l$ satisfies:

$$\begin{aligned}
&\bar{\theta}_{t+1} = \bar{\theta}_t + \alpha(-\hat{A}\bar{\theta}_t + \hat{b}) \\
&\implies \bar{\theta}_{t+1} - \theta_1^* = \bar{\theta}_t - \theta_1^* + \alpha(-\hat{A}(\bar{\theta}_t - \theta_1^* + \theta_1^*) + \hat{b}) \\
&\implies e_{1,t+1} = (I - \alpha\hat{A})e_{1,t} - \alpha\hat{A}\theta_1^* + \alpha\hat{b} \\
&\implies e_{1,t+1} = (I - \alpha\hat{A})e_{1,t} - \alpha\left(\frac{\bar{A}_1 + \bar{A}_2}{2}\right)\bar{A}_1^{-1}\bar{b}_1 + \alpha\frac{\bar{b}_1 + \bar{b}_2}{2} \\
&\implies e_{1,t+1} = (I - \alpha\hat{A})e_{1,t} - \alpha\frac{\bar{A}_2\bar{A}_1^{-1}\bar{b}_1}{2} + \alpha\frac{\bar{b}_2}{2} \\
&\implies e_{1,t+1} = (I - \alpha\hat{A})e_{1,t} - \frac{\alpha\bar{A}_2}{2}\left(\bar{A}_1^{-1}\bar{b}_1 - \bar{A}_2^{-1}\bar{b}_2\right) \\
&\implies e_{1,t+1} = \underbrace{(I - \alpha\hat{A})}_{\tilde{\mathcal{A}}}e_{1,t} + \underbrace{\frac{\alpha\bar{A}_2}{2}(\theta_2^* - \theta_1^*)}_{\tilde{\mathcal{Y}}}. \qquad (30)
\end{aligned}$$

Let us now note that $e_{1,t+1} = \tilde{\mathcal{A}}e_{1,t} + \tilde{\mathcal{Y}}$ can be viewed as a discrete-time linear time-invariant (LTI) system where α is chosen s.t. $\tilde{\mathcal{A}}$ is Schur stable, i.e., $|\lambda_{\max}(\tilde{\mathcal{A}})| < 1$. At the t-th iteration, we have:

$$e_{1,t} = \tilde{\mathcal{A}}^t e_{1,0} + \sum_{k=0}^{t-1}\tilde{\mathcal{A}}^k\tilde{\mathcal{Y}}.$$

As $t \to \infty$, the small gain theorem tells us that because $\rho(\tilde{\mathcal{A}}) < 1$ (where $\rho(\cdot)$ denotes the spectral radius), $\sum_{k=0}^{t-1}\tilde{\mathcal{A}}^k$ exists and is given by $(I - \tilde{\mathcal{A}})^{-1}$. We can then conclude that

$$\begin{aligned}
\lim_{t\to\infty} e_{1,t} &= (I - \tilde{\mathcal{A}})^{-1}\tilde{\mathcal{Y}} \\
&= \left(\alpha\hat{A}\right)^{-1}\frac{\alpha\bar{A}_2}{2}(\theta_1^* - \theta_2^*) \\
&= \frac{1}{2}\hat{A}^{-1}\bar{A}_2(\theta_1^* - \theta_2^*). \qquad (31)
\end{aligned}$$

The limiting expression for $e_{2,t}$ follows the same analysis.

J Proof of the Markovian setting

We now turn our attention to proving the main result of the paper, namely, Theorem 4.

J.1 Outline

As mentioned in the main body, one of the main obstacles to overcome in the analysis is that in general, $\mathbb{E}[(1/N)\sum_{i=1}^{N}\left(g_i(\theta_{t,k}^{(i)}, O_{t,k}^{(i)}) - \bar{g}_i(\theta_{t,k}^{(i)})\right)] \neq 0$. In order to show that a linear speedup is achievable, we first decompose the random TD direction of each agent i as $g_i(\theta_{t,k}^{(i)}) = b_i(O_{t,k}^{(i)}) - A_i(O_{t,k}^{(i)})\theta_{t,k}^{(i)}$ in subsection J.2.1 and show that the variances of $(1/NK)\sum_{i=1}^{N}\sum_{k=0}^{K-1} A_i(O_{t,k}^{(i)})$ and $(1/NK)\sum_{i=1}^{N}\sum_{k=0}^{K-1} b_i(O_{t,k}^{(i)})$ get scaled down by NK in subsection J.2.2. To decouple the randomness between the parameter $\theta_{t,k}^{(i)}$ and the observations $O_{t,k}^{(i)}$ using the method called *information theoretic control of coupling* in [2], we need to bound $\mathbb{E}\left[\left\|\bar{\theta}_t - \bar{\theta}_{t-\tau}\right\|^2\right]$ in subsection J.2.3. As the analysis in the i.i.d. setting and traditional FL, we characterize the drift term, per-iteration error decrease, and parameter selection in subsections J.2.4, J.2.5 and J.2.6, respectively. Finally, we prove Theorem 4 in subsection J.3.

Additional Notation: Under Assumption 3, for each MDP i, there exists some $m_i \geq 1$ and some $\rho_i \in (0,1)$, such that for all $t \geq 0$ and $0 \leq k \leq K-1$, it holds that

$$d_{TV}\left(\mathbb{P}\left(s_{t,k}^{(i)} = \cdot \mid s_{0,0}^{(i)} = s\right), \pi^{(i)}\right) \leq m_i \rho_i^{tK+k}, \forall s \in \mathcal{S}.$$

Furthermore, we define $\rho = \max_{i\in[N]}\{\rho_i\}$, $m = \max_{i\in[N]}\{m_i\}$.

J.2 Auxiliary lemmas for Theorem 4

J.2.1 Decomposition Form

The first step in our proof of Theorem 4 is to rewrite agent i's update direction of `FedTD`(0) as:

$$g_i(\theta_{t,k}^{(i)}) = -A_i(O_{t,k}^{(i)})\theta_{t,k}^{(i)} + b_i(O_{t,k}^{(i)})$$

where $A_i(O_{t,k}^{(i)}) = \phi(s_{t,k}^{(i)})(\phi^\top(s_{t,k}^{(i)}) - \gamma\phi^\top(s_{t,k+1}^{(i)}))$ and $b_i(O_{t,k}^{(i)}) = r(s_{t,k}^{(i)})\phi(s_{t,k}^{(i)})$. Note that the steady-state value of $\mathbb{E}[b_i(O_{t,k}^{(i)})]$ is not equal to 0. For convenience, we apply appropriate centering to rewrite g_i as:

$$g_i(\theta_{t,k}^{(i)}) = -A_i(O_{t,k}^{(i)})(\theta_{t,k}^{(i)} - \theta_i^*) + \underbrace{b_i(O_{t,k}^{(i)}) - A_i(O_{t,k}^{(i)})\theta_i^*}_{Z_i(O_{t,k}^{(i)})}. \tag{32}$$

Define $Z_i(O_{t,k}^{(i)}) \triangleq b_i(O_{t,k}^{(i)}) - A_i(O_{t,k}^{(i)})\theta_i^*$. As $\bar{g}_i(\theta) \triangleq \mathbb{E}_{O_{t,k}^{(i)}\sim\pi^{(i)}}\left[g_i(\theta)\right]$, we have:

$$\bar{g}_i(\theta_{t,k}^{(i)}) = -\bar{A}_i(\theta_{t,k}^{(i)} - \theta_i^*). \tag{33}$$

where $\bar{A}_i = \Phi^\top D^{(i)}(\Phi - \gamma P^{(i)}\Phi)$. Note that $\mathbb{E}_{O_{t,k}^{(i)}\sim\pi^{(i)}}\left[Z_i(O_{t,k}^{(i)})\right]$ equals to 0. Taking into account the definitions above, we establish the following lemmas:

Lemma 11. *(Uniform norm bound) There exist some constants $c_1, c_2, c_3 \geq 0$ such that $\left\|A_i\left(O_{t,k}^{(i)}\right)\right\| \leq c_1 := 1+\gamma$, $\left\|\bar{A}_i\right\| \leq c_2 := 1+\gamma$ and $\left\|Z_i\left(O_{t,k}^{(i)}\right)\right\| \leq c_3 := R_{\max} + c_1 H$ holds for all $i \in [N]$.*

Proof. Based on the definition and the fact that $\|\phi(s)\| \leq 1$, we have

$$\left\|A_i\left(O_{t,k}^{(i)}\right)\right\| = \left\|\phi(s_{t,k}^{(i)})(\phi^\top(s_{t,k}^{(i)}) - \gamma\phi^\top(s_{t,k+1}^{(i)}))\right\| \leq \left\|\phi(s_{t,k}^{(i)})\right\|\left\|\phi^\top(s_{t,k}^{(i)}) - \gamma\phi^\top(s_{t,k+1}^{(i)})\right\| \leq 1+\gamma.$$

Similarly, making use of the fact that $r(s) \leq R_{\max}$ for any $s \in \mathcal{S}$, we apply the same reasoning to conclude that

$$\left\|\bar{A}_i\right\| \leq 1+\gamma \;\;\&\;\; \left\|Z_i\left(O_{t,k}^{(i)}\right)\right\| \leq R_{\max} + c_1 H$$

. □

Lemma 12. *There exist some constants $L_1, L_2 \geq 0$ such that*

$$\left\|\bar{A}_i - \mathbb{E}\left[A_i\left(O_{t_2,k_2}^{(i)}\right) \mid \mathcal{F}_{k_1}^{t_1}\right]\right\| \leq L_1\rho^{(t_2-t_1)K+k_2-k_1} \;\;\&\;\; \left\|\bar{A}_i - \mathbb{E}_{t_1}\left[A_i\left(O_{t_2,k_2}^{(i)}\right)\right]\right\| \leq L_1\rho^{(t_2-t_1)K+k_2},$$
$$\left\|\mathbb{E}\left[Z_i\left(O_{t_2,k_2}^{(i)}\right) \mid \mathcal{F}_{k_1}^{t_1}\right]\right\| \leq L_2\rho^{(t_2-t_1)K+k_2-k_1} \;\;\&\;\; \left\|\mathbb{E}_{t_1}\left[Z_i\left(O_{t_2,k_2}^{(i)}\right)\right]\right\| \leq L_2\rho^{(t_2-t_1)K+k_2}$$

hold for any $i \in [N]$, $0 \leq k_1, k_2 \leq K-1$ and $t_2 \geq t_1 \geq 0$.

Proof. We have:

$$\begin{aligned}
\left\|\mathbb{E}\left[Z_i\left(O_{t_2,k_2}^{(i)}\right) \mid \mathcal{F}_{k_1}^{t_1}\right]\right\| &= \left\|\mathbb{E}\left[Z_i\left(O_{t_2,k_2}^{(i)}\right) \mid \mathcal{F}_{k_1}^{t_1}\right] - \mathbb{E}_{O_{t_2,k_2}^{(i)}\sim\pi^{(i)}}\left[Z_i\left(O_{t_2,k_2}^{(i)}\right) \mid \mathcal{F}_{k_1}^{t_1}\right]\right\| \\
&= \Bigg\| \sum_{s_{t_2,k_2}^{(i)}, s_{t_2+1,k_2+1}^{(i)}} \Big(\pi^{(i)}(s_{t_2,k_2}^{(i)})P(s_{t_2+1,k_2+1}^{(i)} \mid s_{t_2,k_2}^{(i)}) \\
&\quad - P(s_{t_2,k_2}^{(i)} = \cdot \mid s_{t_1,k_1}^{(i)})P(s_{t_2+1,k_2+1}^{(i)} \mid s_{t_2,k_2}^{(i)})\Big) Z_i(O_{t_2,k_2}^{(i)})\Bigg\| \\
&\leq \sum_{s_{t_2,k_2}^{(i)}} \left|\pi^{(i)}(s_{t_2,k_2}^{(i)}) - P(s_{t_2,k_2}^{(i)} = \cdot \mid s_{t_1,k_1}^{(i)})\right| \left\|Z_i(O_{t_2,k_2}^{(i)})\right\| \\
&\overset{(a)}{\leq} \sum_{s_{t_2,k_2}^{(i)}} \left|\pi^{(i)}(s_{t_2,k_2}^{(i)}) - P(s_{t_2,k_2}^{(i)} = \cdot \mid s_{t_1,k_1}^{(i)})\right| (R_{\max} + c_1 H) \\
&= 2(R_{\max} + c_1 H) d_{TV}\left(\mathbb{P}\left(s_{t_2,k_2}^{(i)} = \cdot \mid s_{t_1,k_1}^{(i)} = s\right), \pi^{(i)}\right) \\
&\leq 2(R_{\max} + c_1 H) m_i \rho_i^{(t_2-t_1)K+k_2-k_1}
\end{aligned}$$

where (a) is due to Lemma 11 and the last step follows from Assumption 3. We finish the proof by choosing $L_2 \triangleq \max_{i\in[N]}\{2(R_{\max} + c_1 H)m_i\} = 2c_3 m$. And we follow the same analysis to bound:

$$\begin{aligned}
\left\|\bar{A}_i - \mathbb{E}\left[A_i\left(O_{t_2,k_2}^{(i)}\right) \mid \mathcal{F}_{k_1}^{t_1}\right]\right\| &= \left\|\|\mathbb{E}\left[A_i\left(O_{t_2,k_2}^{(i)}\right) \mid \mathcal{F}_{k_1}^{t_1}\right] - \mathbb{E}_{O_{t_2,k_2}^{(i)}\sim\pi^{(i)}}\left[A_i\left(O_{t_2,k_2}^{(i)}\right) \mid \mathcal{F}_{k_1}^{t_1}\right]\right\| \\
&= \Bigg\| \sum_{s_{t_2,k_2}^{(i)}, s_{t_2+1,k_2+1}^{(i)}} \Big(\pi^{(i)}(s_{t_2,k_2}^{(i)})P(s_{t_2+1,k_2+1}^{(i)} \mid s_{t_2,k_2}^{(i)})
\end{aligned}$$

$$
\begin{aligned}
&-P(s^{(i)}_{t_2,k_2} = \cdot \mid s^{(i)}_{t_1,k_1}) P(s^{(i)}_{t_2+1,k_2+1} \mid s^{(i)}_{t_2,k_2})\Big) A_i(O^{(i)}_{t_2,k_2})\Big\| \\
&\leq \sum_{s^{(i)}_{t_2,k_2}} \left|\pi^{(i)}(s^{(i)}_{t_2,k_2}) - P(s^{(i)}_{t_2,k_2} = \cdot \mid s^{(i)}_{t_1,k_1})\right| \left\|A_i(O^{(i)}_{t_2,k_2})\right\| \\
&\overset{(b)}{\leq} 2c_1 d_{TV}\left(\mathbb{P}\left(s^{(i)}_{t_2,k_2} = \cdot \mid s^{(i)}_{t_1,k_1} = s\right), \pi^{(i)}\right) \\
&\leq 2c_1 m_i \rho_i^{(t_2-t_1)K+k_2-k_1}
\end{aligned}
$$

We finish the proof by choosing $L_1 \triangleq \max_{i\in[N]}\{2c_1 m_i\} = 2c_1 m$. We employ the same reasoning to prove the remaining three inequalities. □

J.2.2 Variance Reduction

We are now ready to present the variance reduction Lemma in the Markov setting. The following Lemma establishes an analog of the variance reduction Lemma 7 in the i.i.d. setting. Based on the assumption that trajectories are independent across agents, it is easy to understand that the variance of $(1/NK)\sum_{i=1}^{N}\sum_{k=0}^{K-1} A_i(O^{(i)}_{t,k})$ and $(1/NK)\sum_{i=1}^{N}\sum_{k=0}^{K-1} b_i(O^{(i)}_{t,k})$ can be scaled by the number of agents N. However, it is not obvious that the variances can be scaled by K (the number of local iterations), since the observations of each agent $O^{(i)}_{t,k_1}$ and $O^{(i)}_{t,k_2}$ are correlated at different local steps k_1, k_2. Due to the geometric mixing property of the Markov chain, the correlation between $O^{(i)}_{t,k_1}$ and $O^{(i)}_{t,k_2}$ will geometrically decay after the mixing time. Based on this fact, we show that the variances of $(1/NK)\sum_{i=1}^{N}\sum_{k=0}^{K-1} A_i(O^{(i)}_{t,k})$ and $(1/NK)\sum_{i=1}^{N}\sum_{k=0}^{K-1} b_i(O^{(i)}_{t,k})$ get scaled down by NK with an additional additive, higher order term dependent on the mixing time τ, which is formally stated as follows:

Lemma 13. *(Variance reduction in the Markovian setting) For any $0 < \tau < t$, there exists $d_1, d_2 > 0$ such that:*

$$
\mathbb{E}_{t-\tau}\left[\left\|\frac{1}{NK}\sum_{i=1}^{N}\sum_{k=0}^{K-1}\left[A_i(O^{(i)}_{t,k}) - \bar{A}_i\right]\right\|\right] \leq \frac{d_1}{\sqrt{NK}} + 2L_1\rho^{\tau K}, \tag{34}
$$

$$
\mathbb{E}_{t-\tau}\left[\left\|\frac{1}{NK}\sum_{i=1}^{N}\sum_{k=0}^{K-1}\left[A_i(O^{(i)}_{t,k}) - \bar{A}_i\right]\right\|^2\right] \leq \frac{d_1^2}{NK} + 4L_1^2\rho^{2\tau K}, \tag{35}
$$

$$
\mathbb{E}_{t-\tau}\left[\left\|\frac{1}{NK}\sum_{i=1}^{N}\sum_{k=0}^{K-1} Z_i(O^{(i)}_{t,k})\right\|\right] \leq \frac{d_2}{\sqrt{NK}} + 2L_2\rho^{\tau K}, \quad \textit{and} \tag{36}
$$

$$
\mathbb{E}_{t-\tau}\left[\left\|\frac{1}{NK}\sum_{i=1}^{N}\sum_{k=0}^{K-1} Z_i(O^{(i)}_{t,k})\right\|^2\right] \leq \frac{d_2^2}{NK} + 4L_2^2\rho^{2\tau K}, \tag{37}
$$

where $d_1 \triangleq \sqrt{(c_1+c_2)^2 + \frac{2(c_1+c_2)L_1\rho}{1-\rho}}$ and $d_2 \triangleq \sqrt{c_3^2 + \frac{2c_3L_2\rho}{1-\rho}}$.

Proof.

$$
\mathbb{E}_{t-\tau}\left[\left\|\frac{1}{NK}\sum_{i=1}^{N}\sum_{k=0}^{K-1} Z_i(O^{(i)}_{t,k})\right\|\right] = \mathbb{E}_{t-\tau}\left[\sqrt{\left(\frac{1}{NK}\sum_{i=1}^{N}\sum_{k=0}^{K-1} Z_i(O^{(i)}_{t,k})\right)^{\top}\left(\frac{1}{NK}\sum_{i=1}^{N}\sum_{k=0}^{K-1} Z_i(O^{(i)}_{t,k})\right)}\right]
$$

$$\overset{(a)}{\leq} \sqrt{\mathbb{E}_{t-\tau}\left[\left(\frac{1}{NK}\sum_{i=1}^{N}\sum_{k=0}^{K-1} Z_i(O_{t,k}^{(i)})\right)^\top \left(\frac{1}{NK}\sum_{i=1}^{N}\sum_{k=0}^{K-1} Z_i(O_{t,k}^{(i)})\right)\right]}$$

$$= \Bigg\{ \mathbb{E}_{t-\tau}\Bigg[\frac{1}{N^2K^2}\sum_{i=1}^{N}\sum_{k=0}^{K-1} Z_i(O_{t,k}^{(i)})^\top Z_i(O_{t,k}^{(i)}) + \underbrace{\frac{2}{N^2K^2}\sum_{i=1}^{N}\sum_{k<l} Z_i(O_{t,k}^{(i)})^\top Z_i(O_{t,l}^{(i)})}_{T_1}$$

$$+ \underbrace{\frac{2}{N^2K^2}\sum_{i<j}\sum_{k=0}^{K-1} Z_i(O_{t,k}^{(i)})^\top Z_j(O_{t,k}^{(j)})}_{T_2} + \underbrace{\frac{2}{N^2K^2}\sum_{i<j}\sum_{k<l} Z_i(O_{t,k}^{(i)})^\top Z_j(O_{t,l}^{(j)})}_{T_3}\Bigg]\Bigg\}^{\frac{1}{2}} \tag{38}$$

where (a) is due to the concavity of square root and Jensen's inequality. Furthermore, the term T_1 can be further bounded by:

$$\begin{aligned}
\mathbb{E}_{t-\tau}[T_1] &= \mathbb{E}_{t-\tau}\left[\frac{2}{N^2K^2}\sum_{i=1}^{N}\sum_{k<l} Z_i(O_{t,k}^{(i)})^\top Z_i(O_{t,l}^{(i)})\right] \\
&= \mathbb{E}_{t-\tau}\left[\frac{2}{N^2K^2}\sum_{i=1}^{N}\sum_{k<l} Z_i(O_{t,k}^{(i)})^\top \mathbb{E}\left[Z_i(O_{t,l}^{(i)}) \mid \mathcal{F}_k^t\right]\right] \\
&\leq \mathbb{E}_{t-\tau}\left[\frac{2}{N^2K^2}\sum_{i=1}^{N}\sum_{k<l} \left\|Z_i(O_{t,k}^{(i)})\right\| \left\|\mathbb{E}\left[Z_i(O_{t,l}^{(i)}) \mid \mathcal{F}_k^t\right]\right\|\right] \quad \text{(Cauchy–Schwarz inequality)} \\
&\leq \mathbb{E}_{t-\tau}\left[\frac{2}{N^2K^2}\sum_{i=1}^{N}\sum_{k<l} c_3L_2\rho^{(l-k)}\right] \quad \text{(Lemma 11 and 12)} \\
&\leq \mathbb{E}_{t-\tau}\left[\frac{2}{N^2K^2}\sum_{i=1}^{N}\sum_{k=0}^{K-1}\sum_{m=1}^{\infty} c_3L_2\rho^{m}\right] \\
&= \frac{2c_3L_2NK}{N^2K^2}\frac{\rho}{1-\rho} = \frac{2c_3L_2\rho}{NK(1-\rho)}.
\end{aligned}$$

And T_2 can be bounded by:

$$\begin{aligned}
\mathbb{E}_{t-\tau}[T_2] &= \frac{2}{N^2K^2}\sum_{i<j}\sum_{k=0}^{K-1} \mathbb{E}_{t-\tau}\left[Z_i(O_{t,k}^{(i)})\right]^\top \mathbb{E}_{t-\tau}\left[Z_j(O_{t,k}^{(j)})\right] \ (O_{t,k}^{(i)} \text{ and } O_{t,k}^{(j)} \text{ are independent}) \\
&\leq \frac{2}{N^2K^2}\sum_{i<j}\sum_{k=0}^{K-1} L_2^2\rho^{2\tau K+2k} \quad \text{(Lemma 12)} \\
&\leq \frac{2}{K}L_2^2\rho^{2\tau K}.
\end{aligned}$$

Meanwhile, T_3 can be bounded by:

$$\mathbb{E}_{t-\tau}[T_3] = \frac{2}{N^2K^2}\sum_{i<j}\sum_{k<l} \mathbb{E}_{t-\tau}\left[Z_i(O_{t,k}^{(i)})\right]^\top \mathbb{E}_{t-\tau}\left[Z_j(O_{t,l}^{(j)})\right] \ (O_{t,k}^{(i)} \text{ and } O_{t,l}^{(j)} \text{ are independent})$$

$$\leq \frac{2}{N^2K^2}\sum_{i<j}\sum_{k<l} L_2^2\rho^{2\tau K+k+l} \quad \text{(Lemma 12)}$$
$$\leq 2L_2^2\rho^{2\tau K}$$

Substituting the upper bound of T_1, T_2 and T_3 into Eq (38), we have:

$$\begin{aligned}
\mathbb{E}_{t-\tau}\left[\left\|\frac{1}{NK}\sum_{i=1}^{N}\sum_{k=0}^{K-1} Z_i(O_{t,k}^{(i)})\right\|\right] &\leq \left(\frac{1}{N^2K^2}\sum_{i=1}^{N}\sum_{k=0}^{K-1}\mathbb{E}_{t-\tau}\left[Z_i(O_{t,k}^{(i)})^\top Z_i(O_{t,k}^{(i)})\right]\right.\\
&\quad \left.+\frac{2c_3L_2\rho}{NK(1-\rho)}+\frac{2}{K}L_2^2\rho^{2\tau K}+2L_2^2\rho^{2\tau K}\right)^{\frac{1}{2}}\\
&\overset{(a)}{\leq} \sqrt{\frac{NK}{N^2K^2}c_3^2+\frac{2c_3L\rho}{NK(1-\rho)}+\frac{2}{K}L_2^2\rho^{2\tau K}+2L_2^2\rho^{2\tau K}}\\
&\leq \sqrt{\frac{1}{NK}\left(c_3^2+\frac{2c_3L_2\rho}{1-\rho}\right)+4L_2^2\rho^{2\tau K}} \quad (K\geq 1)\\
&\leq \sqrt{\frac{1}{NK}\left(c_3^2+\frac{2c_3L_2\rho}{1-\rho}\right)}+\sqrt{4L_2^2\rho^{2\tau K}}\\
&= \sqrt{\frac{1}{NK}\left(c_3^2+\frac{2c_3L_2\rho}{1-\rho}\right)}+2L_2\rho^{\tau K}.
\end{aligned}$$

where (a) used the fact that $\left\|Z_i\left(O_{t,k}^{(i)}\right)\right\| \leq c_3$ mentioned in Lemma 11. The proof of other inequalities follows the same reasoning. □

J.2.3 Bounding $\mathbb{E}\left[\left\|\bar{\theta}_t-\bar{\theta}_{t-\tau}\right\|^2\right]$

Lemma 14. *(Bounding $\|\theta_t-\theta_{t-\tau}\|^2$) Consider $\tau=\lceil\frac{\tau^{\text{mix}}(\alpha_T^2)}{K}\rceil$ and choose the effective step-size*

$$\alpha \leq \min\left\{\frac{1}{30c_4(\tau+1)},\frac{1}{96c_4^2\tau},1\right\}$$

where $c_4=3c_1$. For any $t\geq 2\tau$, we have the following bound:

$$\begin{aligned}
\mathbb{E}_{t-2\tau}\left[\left\|\bar{\theta}_t-\bar{\theta}_{t-\tau}\right\|^2\right] &\leq 8\alpha^2\tau^2c_4^2\mathbb{E}_{t-2\tau}\left[\left\|\bar{\theta}_t-\theta^*\right\|^2\right]+14\alpha^2\tau^2\frac{d_2^2}{NK}+\frac{52L_2^2\alpha^4\tau}{1-\rho^2}\\
&\quad +4\alpha^2c_4^2\tau\sum_{s=0}^{\tau}E_{t-2\tau}[\Delta_{t-s}]+3200\alpha^2c_4^2c_1^2\tau^3\Gamma^2(\epsilon,\epsilon_1)+4\alpha^2c_1^2\tau^2\Gamma^2(\epsilon,\epsilon_1). \quad (39)
\end{aligned}$$

Proof. For any $l\geq 2\tau$, we have

$$\begin{aligned}
\left\|\bar{\theta}_{l+1}-\bar{\theta}_l\right\|^2 &= \left\|\Pi_{2,\mathcal{H}}\left(\bar{\theta}_l+\frac{\alpha}{NK}\sum_{i=1}^{N}\sum_{k=0}^{K-1}g_i(\theta_{l,k}^{(i)})\right)-\bar{\theta}_l\right\|^2\\
&\leq \left\|\bar{\theta}_l+\frac{\alpha}{NK}\sum_{i=1}^{N}\sum_{k=0}^{K-1}g_i(\theta_{l,k}^{(i)})-\bar{\theta}_l\right\|^2\\
&= \alpha^2\left\|\frac{1}{NK}\sum_{i=1}^{N}\sum_{k=0}^{K-1}\left[-A_i(O_{l,k}^{(i)})\left(\theta_{l,k}^{(i)}-\theta_i^*\right)+Z_i(O_{l,k}^{(i)})\right]\right\|^2
\end{aligned}$$

$$
\begin{aligned}
&\leq 2\alpha^2\Big\|\frac{1}{NK}\sum_{i=1}^{N}\sum_{k=0}^{K-1}\Big[-A_i(O_{l,k}^{(i)})\Big(\theta_{l,k}^{(i)}-\theta^*\Big)+Z_i(O_{l,k}^{(i)})\Big]\Big\|^2\\
&+2\alpha^2\Big\|\frac{1}{NK}\sum_{i=1}^{N}\sum_{k=0}^{K-1}\Big[-A_i(O_{l,k}^{(i)})\Big(\theta^*-\theta_i^*\Big)\Big]\Big\|^2\\
&\overset{(a)}{\leq} 2\alpha^2\Big\|\frac{1}{NK}\sum_{i=1}^{N}\sum_{k=0}^{K-1}\Big[-A_i(O_{l,k}^{(i)})\Big(\theta_{l,k}^{(i)}-\theta^*\Big)+Z_i(O_{l,k}^{(i)})\Big]\Big\|^2+2\alpha^2c_1^2\Gamma^2(\epsilon,\epsilon_1)\\
&=6\alpha^2\Big\|\frac{1}{NK}\sum_{i=1}^{N}\sum_{k=0}^{K-1}A_i(O_{l,k}^{(i)})\Big(\theta_{l,k}^{(i)}-\bar{\theta}_l\Big)\Big\|^2+6\alpha^2\Big\|\frac{1}{NK}\sum_{i=1}^{N}\sum_{k=0}^{K-1}A_i(O_{l,k}^{(i)})\Big(\bar{\theta}_l-\theta^*\Big)\Big\|^2\\
&+6\alpha^2\Big\|\frac{1}{NK}\sum_{i=1}^{N}\sum_{k=0}^{K-1}Z_i(O_{l,k}^{(i)})\Big\|^2+2\alpha^2c_1^2\Gamma^2(\epsilon,\epsilon_1)\\
&\leq 6\alpha^2\left(\frac{c_1}{NK}\sum_{i=1}^{N}\sum_{k=0}^{K-1}\Big\|\theta_{l,k}^{(i)}-\bar{\theta}_l\Big\|\right)^2+6\alpha^2c_1^2\Big\|\bar{\theta}_l-\theta^*\Big\|^2\\
&+6\alpha^2\Big\|\frac{1}{NK}\sum_{i=1}^{N}\sum_{k=0}^{K-1}Z_i(O_{l,k}^{(i)})\Big\|^2+2\alpha^2c_1^2\Gamma^2(\epsilon,\epsilon_1),
\end{aligned}
\tag{40}
$$

where (a) comes from the upper bound of fixed points distance in Theorem 1 and the fact that $\Big\|A_i\left(O_{t,k}^{(i)}\right)\Big\|\leq c_1$ in Lemma 11. Taking square root on both sides of the inequality above, we get:

$$
\begin{aligned}
\Big\|\bar{\theta}_{l+1}-\bar{\theta}_l\Big\| &\leq 3\sqrt{\alpha^2\left(\frac{c_1}{NK}\sum_{i=1}^{N}\sum_{k=0}^{K-1}\Big\|\theta_{l,k}^{(i)}-\bar{\theta}_l\Big\|\right)^2}+3\sqrt{\alpha^2c_1^2\Big\|\bar{\theta}_l-\theta^*\Big\|^2}\\
&+3\sqrt{\alpha^2\Big\|\frac{1}{NK}\sum_{i=1}^{N}\sum_{k=0}^{K-1}Z_i(O_{l,k}^{(i)})\Big\|^2}+\sqrt{2\alpha^2c_1^2\Gamma^2(\epsilon,\epsilon_1)}\\
&\leq \frac{3\alpha c_1}{NK}\sum_{i=1}^{N}\sum_{k=0}^{K-1}\Big\|\theta_{l,k}^{(i)}-\bar{\theta}_l\Big\|+3\alpha c_1\Big\|\bar{\theta}_l-\theta^*\Big\|+3\alpha\Big\|\frac{1}{NK}\sum_{i=1}^{N}\sum_{k=0}^{K-1}Z_i(O_{l,k}^{(i)})\Big\|+2\alpha c_1\Gamma(\epsilon,\epsilon_1).
\end{aligned}
\tag{41}
$$

By using the fact that $\Big\|\bar{\theta}_{l+1}-\theta^*\Big\|\leq\Big\|\bar{\theta}_l-\theta^*\Big\|+\Big\|\bar{\theta}_{l+1}-\bar{\theta}_l\Big\|$, we have:

$$
\Big\|\bar{\theta}_{l+1}-\theta^*\Big\|\leq(1+3\alpha c_1)\Big\|\bar{\theta}_l-\theta^*\Big\|+\frac{3\alpha c_1}{NK}\sum_{i=1}^{N}\sum_{k=0}^{K-1}\Big\|\theta_{l,k}^{(i)}-\bar{\theta}_l\Big\|+3\alpha\Big\|\frac{1}{NK}\sum_{i=1}^{N}\sum_{k=0}^{K-1}Z_i(O_{l,k}^{(i)})\Big\|+2\alpha c_1\Gamma(\epsilon,\epsilon_1).
\tag{42}
$$

For simplicity, we define $c_4\triangleq 3c_1$ and $\delta_l\triangleq\frac{1}{NK}\sum_{i=1}^{N}\sum_{k=0}^{K-1}\Big\|\theta_{l,k}^{(i)}-\bar{\theta}_l\Big\|$. Taking the square on both sides of Eq (42), we have:

$$
\begin{aligned}
\Big\|\bar{\theta}_{l+1}-\theta^*\Big\|^2 &\leq (1+\alpha c_4)^2\Big\|\bar{\theta}_l-\theta^*\Big\|^2+\alpha^2c_4^2\delta_l^2+9\alpha^2\Big\|\frac{1}{NK}\sum_{i=1}^{N}\sum_{k=0}^{K-1}Z_i(O_{l,k}^{(i)})\Big\|^2+4\alpha^2c_1^2\Gamma^2(\epsilon,\epsilon_1)\\
&\underbrace{+6\alpha(1+\alpha c_4)\Big\|\bar{\theta}_l-\theta^*\Big\|\Big\|\frac{1}{NK}\sum_{i=1}^{N}\sum_{k=0}^{K-1}Z_i(O_{l,k}^{(i)})\Big\|}_{H_1}+\underbrace{2\alpha c_4(1+\alpha c_4)\Big\|\bar{\theta}_l-\theta^*\Big\|\delta_l}_{H_2}
\end{aligned}
$$

$$+\underbrace{6\alpha^2 c_4\delta_l\Big\|\frac{1}{NK}\sum_{i=1}^{N}\sum_{k=0}^{K-1}Z_i(O_{l,k}^{(i)})\Big\|}_{H_3}+\underbrace{4\alpha^2 c_1c_4\delta_l\Gamma(\epsilon,\epsilon_1)}_{H_4}+\underbrace{4\alpha c_1(1+\alpha c_4)\Big\|\bar{\theta}_l-\theta^*\Big\|\Gamma(\epsilon,\epsilon_1)}_{H_5}$$

$$+\underbrace{12\alpha^2 c_1\Big\|\frac{1}{NK}\sum_{i=1}^{N}\sum_{k=0}^{K-1}Z_i(O_{l,k}^{(i)})\Big\|\Gamma(\epsilon,\epsilon_1)}_{H_6}. \tag{43}$$

We can further bound H_1 as:

$$\begin{aligned}
H_1 &= 6\alpha(1+\alpha c_4)\Big\|\bar{\theta}_l-\theta^*\Big\|\Big\|\frac{1}{NK}\sum_{i=1}^{N}\sum_{k=0}^{K-1}Z_i(O_{l,k}^{(i)})\Big\| \\
&= 2\sqrt{3\alpha(1+\alpha c_4)}\Big\|\bar{\theta}_l-\theta^*\Big\|\cdot\sqrt{3\alpha(1+\alpha c_4)}\Big\|\frac{1}{NK}\sum_{i=1}^{N}\sum_{k=0}^{K-1}Z_i(O_{l,k}^{(i)})\Big\| \\
&\le 3\alpha(1+\alpha c_4)\Big\|\bar{\theta}_l-\theta^*\Big\|^2+3\alpha(1+\alpha c_4)\Big\|\frac{1}{NK}\sum_{i=1}^{N}\sum_{k=0}^{K-1}Z_i(O_{l,k}^{(i)})\Big\|^2 \\
&\le 6\alpha\Big\|\bar{\theta}_l-\theta^*\Big\|^2+6\alpha\Big\|\frac{1}{NK}\sum_{i=1}^{N}\sum_{k=0}^{K-1}Z_i(O_{l,k}^{(i)})\Big\|^2.
\end{aligned} \tag{44}$$

where we use the fact $1+\alpha c_4\le 2$ in the last inequality. Similary, we can bound H_2 as:

$$H_2 = 2\alpha c_4(1+\alpha c_4)\Big\|\bar{\theta}_l-\theta^*\Big\|\delta_l \le 2\alpha\Big\|\bar{\theta}_l-\theta^*\Big\|^2+2\alpha c_4^2\delta_l^2. \tag{45}$$

And we bound H_3 as:

$$H_3 = 6\alpha^2 c_4\delta_l\Big\|\frac{1}{NK}\sum_{i=1}^{N}\sum_{k=0}^{K-1}Z_i(O_{l,k}^{(i)})\Big\| \le 3\alpha^2\Big\|\frac{1}{NK}\sum_{i=1}^{N}\sum_{k=0}^{K-1}Z_i(O_{l,k}^{(i)})\Big\|^2+3\alpha^2c_4^2\delta_l^2. \tag{46}$$

For H_4, H_5, H_6, we have:

$$H_4 = 4\alpha^2c_1c_4\delta_l\Gamma(\epsilon,\epsilon_1) \le 2\alpha^2c_4^2\delta_l^2+2\alpha^2c_1^2\Gamma^2(\epsilon,\epsilon_1),$$

$$H_5 = 4\alpha c_1(1+\alpha c_4)\Big\|\bar{\theta}_l-\theta^*\Big\|\Gamma(\epsilon,\epsilon_1) \le 4\alpha\Big\|\bar{\theta}_l-\theta^*\Big\|^2+4\alpha c_1^2\Gamma^2(\epsilon,\epsilon_1),$$

$$H_6 = 12\alpha^2c_1\Big\|\frac{1}{NK}\sum_{i=1}^{N}\sum_{k=0}^{K-1}Z_i(O_{l,k}^{(i)})\Big\|\Gamma(\epsilon,\epsilon_1) \le 6\alpha^2\Big\|\frac{1}{NK}\sum_{i=1}^{N}\sum_{k=0}^{K-1}Z_i(O_{l,k}^{(i)})\Big\|^2+6\alpha^2c_1^2\Gamma^2(\epsilon,\epsilon_1),$$

Substituting the upper bound of $H_1, H_2, \ldots, H_6$ into Eq (43) and noting that $(1+\alpha c_4)^2 \le 1+3\alpha c_4$ because $\alpha c_4\le 1$, we have:

$$\begin{aligned}
\Big\|\bar{\theta}_{l+1}-\theta^*\Big\|^2 &\le (1+\alpha(3c_4+12))\Big\|\bar{\theta}_l-\theta^*\Big\|^2+(6\alpha^2+2\alpha)c_4^2\delta_l^2 \\
&\quad+(18\alpha^2+6\alpha)\Big\|\frac{1}{NK}\sum_{i=1}^{N}\sum_{k=0}^{K-1}Z_i(O_{l,k}^{(i)})\Big\|^2+(12\alpha^2+4\alpha)c_1^2\Gamma^2(\epsilon,\epsilon_1)
\end{aligned}$$

$$\leq (1+\alpha h_1)\left\|\bar{\theta}_l-\theta^*\right\|^2+8\alpha c_4^2\delta_l^2+24\alpha\left\|\frac{1}{NK}\sum_{i=1}^{N}\sum_{k=0}^{K-1}Z_i(O_{l,k}^{(i)})\right\|^2+16\alpha c_1^2\Gamma^2(\epsilon,\epsilon_1), \tag{47}$$

where we denote $h_1 \triangleq 3c_4+12$ for simplicity. For any $t-\tau \leq l \leq t$, conditioning on $\mathcal{F}_{t-2\tau}$ on both sides of the above inequality, we have:

$$\begin{aligned}
\mathbb{E}_{t-2\tau}\left\|\bar{\theta}_{l+1}-\theta^*\right\|^2 &\leq (1+\alpha h_1)\,\mathbb{E}_{t-2\tau}\left\|\bar{\theta}_l-\theta^*\right\|^2+24\alpha\mathbb{E}_{t-2\tau}\left\|\frac{1}{NK}\sum_{i=1}^{N}\sum_{k=0}^{K-1}Z_i(O_{l,k}^{(i)})\right\|^2\\
&\quad+8\alpha c_4^2\mathbb{E}_{t-2\tau}\left[\delta_l^2\right]+\alpha M_3(\epsilon,\epsilon_1)\\
&\leq (1+\alpha h_1)\,\mathbb{E}_{t-2\tau}\left\|\bar{\theta}_l-\theta^*\right\|^2+24\alpha\left[\frac{d_2^2}{NK}+4L_2^2\rho^{2(l-t+2\tau)K}\right] \quad \text{(Lemma 13)}\\
&\quad+8\alpha c_4^2\mathbb{E}_{t-2\tau}\left[\delta_l^2\right]+\alpha M_3(\epsilon,\epsilon_1)\\
&\overset{(a)}{\leq} (1+\alpha h_1)\,\mathbb{E}_{t-2\tau}\left\|\bar{\theta}_l-\theta^*\right\|^2+24\alpha\left[\frac{d_2^2}{NK}+4L_2^2\alpha^2\rho^{2(l-t+\tau)K}\right]\\
&\quad+8\alpha c_4^2\mathbb{E}_{t-2\tau}\left[\delta_l^2\right]+\alpha M_3(\epsilon,\epsilon_1)\\
&\leq (1+\alpha h_1)\,\mathbb{E}_{t-2\tau}\left\|\bar{\theta}_l-\theta^*\right\|^2+\alpha c_t(l)+8\alpha c_4^2\mathbb{E}_{t-2\tau}\left[\delta_l^2\right]+\alpha M_3(\epsilon,\epsilon_1),
\end{aligned} \tag{48}$$

where we denote $M_3(\epsilon,\epsilon_1) \triangleq 16c_1^2\Gamma^2(\epsilon,\epsilon_1)$ and $c_t(l)=24\left[\frac{d_2^2}{NK}+4L_2^2\alpha^2\rho^{2(l-t+\tau)K}\right]$ for simplicity. Inequality (a) is due to $\rho^{2\tau K}\leq\alpha_T^4\leq\alpha_t^2$. In the following steps, we try to map $\mathbb{E}_{t-2\tau}\left\|\bar{\theta}_{l+1}-\theta^*\right\|^2$ to $\mathbb{E}_{t-2\tau}\left\|\bar{\theta}_{t-\tau}-\theta^*\right\|^2$ for any $t-\tau\leq l\leq t$. By applying Eq (48) recursively, we have:

$$\begin{aligned}
\mathbb{E}_{t-2\tau}\left\|\bar{\theta}_{l+1}-\theta^*\right\|^2 &\leq (1+\alpha h_1)^{l+1-t+\tau}\,\mathbb{E}_{t-2\tau}\left\|\bar{\theta}_{t-\tau}-\theta^*\right\|^2+\alpha\sum_{k=t-\tau}^{l}(1+\alpha h_1)^{l-k}\left(c_t(k)+M_3(\epsilon,\epsilon_1)\right)\\
&\quad+8\alpha c_4^2\mathbb{E}_{t-2\tau}\left[\sum_{k=t-\tau}^{l}(1+\alpha h_1)^{l-k}\,\delta_k^2\right]\\
&\overset{(b)}{\leq} (1+\alpha h_1)^{\tau+1}\,\mathbb{E}_{t-2\tau}\left\|\bar{\theta}_{t-\tau}-\theta^*\right\|^2+\underbrace{\alpha\sum_{k=t-\tau}^{t}(1+\alpha h_1)^{l-k}\left(c_t(k)+M_3(\epsilon,\epsilon_1)\right)}_{H_7}\\
&\quad+8\alpha c_4^2\mathbb{E}_{t-2\tau}\underbrace{\left[\sum_{k=t-\tau}^{t}(1+\alpha h_1)^{l-k}\,\delta_k^2\right]}_{H_8}
\end{aligned} \tag{49}$$

where (b) is due to $l\leq t$. For H_7, we have:

$$\begin{aligned}
H_7 &\leq \sum_{k=t-\tau}^{t}(1+\alpha h_1)^{t-k}\left(c_t(k)+M_3(\epsilon,\epsilon_1)\right) \quad (l\leq t)\\
&= \sum_{k'=0}^{\tau}(1+\alpha h_1)^{\tau-k'}\left(c_t(k'+t-\tau)+M_3(\epsilon,\epsilon_1)\right) \quad (\text{ changing index } k \text{ into } k' \text{ with } k'=k+\tau-t)
\end{aligned}$$

$$
\begin{aligned}
&\overset{(a)}{\leq} 24 \sum_{k'=0}^{\tau} (1+\alpha h_1)^{\tau-k'} \left[\frac{d_2^2}{NK} + 4L_2^2\alpha^2\rho^{2k'K} + M_3(\epsilon,\epsilon_1) \right] \\
&= 24 \left[\left(\frac{d_2^2}{NK} + M_3(\epsilon,\epsilon_1) \right) \frac{(1+\alpha h_1)^{\tau+1}-1}{\alpha h_1} + 4L_2^2\alpha^2 (1+\alpha h_1)^{\tau} \sum_{k'=0}^{\tau} \left(\frac{\rho^{2K}}{1+\alpha h_1} \right)^{k'} \right] \\
&\leq 24 \left[\left(\frac{d_2^2}{NK} + M_3(\epsilon,\epsilon_1) \right) \frac{(1+\alpha h_1)^{\tau+1}-1}{\alpha h_1} + 4L_2^2\alpha^2 (1+\alpha h_1)^{\tau} \sum_{k'=0}^{\tau} \rho^{2k'K} \right] \quad (1+\alpha h_1 \geq 1) \\
&\leq 24 \left[\left(\frac{d_2^2}{NK} + M_3(\epsilon,\epsilon_1) \right) \frac{(1+\alpha h_1)^{\tau+1}-1}{\alpha h_1} + 4L_2^2\alpha^2 (1+\alpha h_1)^{\tau} \frac{1}{1-\rho^2} \right].
\end{aligned}
$$

where (a) is due to the definition of $c_t(k')$. Here we follow the analysis in [26]. Notice that for $x \leq \frac{\log 2}{\tau}$, we have $(1+x)^{\tau+1} \leq 1+2x(\tau+1)$. If $\alpha \leq \frac{1}{4h_1\tau} \leq \frac{\log 2}{h_1\tau}$ and $\alpha \leq \frac{1}{2h_1(\tau+1)}$, we have $(1+\alpha h_1)^{\tau+1} \leq 1+2\alpha h_1(\tau+1) \leq 2$ and $(1+\alpha h_1)^{\tau} \leq 1+2\alpha h_1\tau \leq 1+1/2 \leq 2$. Hence, we have

$$
H_7 \leq 24 \left[\left(\frac{d_2^2}{NK} + M_3(\epsilon,\epsilon_1) \right) 2(\tau+1) + \frac{8L_2^2\alpha^2}{1-\rho^2} \right].
$$

We apply the similar analysis to bound H_8 as:

$$
H_8 = \sum_{k=0}^{\tau} (1+\alpha h_1)^{\tau-k} \delta_{t-\tau+k}^2 \leq \sum_{k=0}^{\tau} (1+\alpha h_1)^{\tau} \delta_{t-\tau+k}^2 \leq \sum_{k=0}^{\tau} (1+2\alpha h_1\tau) \delta_{t-\tau+k}^2 \leq 2\sum_{k=0}^{\tau} \delta_{t-k}^2.
$$

Substituting the upper bound of H_7 and H_8 into Eq (49), we have:

$$
\begin{aligned}
\mathbb{E}_{t-2\tau}\left\|\bar{\theta}_{l+1}-\theta^*\right\|^2 \leq 2\mathbb{E}_{t-2\tau}\left\|\bar{\theta}_{t-\tau}-\theta^*\right\|^2 &+ 24\alpha \left[\left(\frac{d_2^2}{NK} + M_3(\epsilon,\epsilon_1) \right) 2(\tau+1) + \frac{8L_2^2\alpha^2}{1-\rho^2} \right] \\
&+ 16\alpha c_4^2 \sum_{k=0}^{\tau} \mathbb{E}_{t-2\tau}[\delta_{t-k}^2].
\end{aligned}
$$

Then it is straightforward to bound $\mathbb{E}_{t-2\tau}\left\|\bar{\theta}_{l}-\theta^*\right\|^2$ as:

$$
\begin{aligned}
\mathbb{E}_{t-2\tau}\left\|\bar{\theta}_{l}-\theta^*\right\|^2 \leq 2\mathbb{E}_{t-2\tau}\left\|\bar{\theta}_{t-\tau}-\theta^*\right\|^2 &+ 24\alpha \left[\left(\frac{d_2^2}{NK} + M_3(\epsilon,\epsilon_1) \right) 4\tau + \frac{8L_2^2\alpha^2}{1-\rho^2} \right] \\
&+ 16\alpha c_4^2 \sum_{k=0}^{\tau} \mathbb{E}_{t-2\tau}[\delta_{t-k}^2]. \qquad (50)
\end{aligned}
$$

Furthermore, based on the triangle inequality, we have:

$$
\begin{aligned}
\left\|\bar{\theta}_t-\bar{\theta}_{t-\tau}\right\|^2 &\leq \left(\sum_{s=t-\tau}^{t-1} \left\|\bar{\theta}_{s+1}-\bar{\theta}_s\right\| \right)^2 \leq \tau \sum_{s=t-\tau}^{t-1} \left\|\bar{\theta}_{s+1}-\bar{\theta}_s\right\|^2 \\
&\leq \tau \sum_{s=t-\tau}^{t-1} \left[\alpha^2c_4^2\left\|\bar{\theta}_s-\theta^*\right\|^2 + \alpha^2c_4^2\delta_s^2 + 6\alpha^2\left\| \frac{1}{NK}\sum_{i=1}^{N}\sum_{k=0}^{K-1} Z_i(O_{s,k}^{(i)}) \right\|^2 + 2\alpha^2c_1^2\Gamma^2(\epsilon,\epsilon_1) \right]
\end{aligned}
$$

where the last inequality is due to Eq (40) with $c_4 = 3c_1$. If we take the expectation on both sides, we have:

$$
\mathbb{E}_{t-2\tau}\left\|\bar{\theta}_t-\bar{\theta}_{t-\tau}\right\|^2 \leq \tau \sum_{s=t-\tau}^{t-1} \left[\alpha^2c_4^2\mathbb{E}_{t-2\tau}\left\|\bar{\theta}_s-\theta^*\right\|^2 + \alpha^2c_4^2\delta_s^2 \right.
$$

$$
\begin{aligned}
&+6\alpha^2\mathbb{E}_{t-2\tau}\Big\|\frac{1}{NK}\sum_{i=1}^{N}\sum_{k=0}^{K-1}Z_i(O_{s,k}^{(i)})\Big\|^2+2\alpha^2c_1^2\Gamma^2(\epsilon,\epsilon_1)\Big]\\
&\leq\tau\alpha^2c_4^2\sum_{s=t-\tau}^{t-1}\Bigg[2\mathbb{E}_{t-2\tau}\Big\|\bar{\theta}_{t-\tau}-\theta^*\Big\|^2+24\alpha\left[\left(\frac{d_2^2}{NK}+M_3(\epsilon,\epsilon_1)\right)4\tau+\frac{8L_2^2\alpha^2}{1-\rho^2}\right]\\
&+16\alpha c_4^2\sum_{k=0}^{\tau}\mathbb{E}_{t-2\tau}[\delta_{t-k}^2]\Bigg]\quad\text{(Eq (50))}\\
&+6\alpha^2\tau\sum_{s=t-\tau}^{t-1}\left(\frac{d_2^2}{NK}+4L_2^2\rho^{2(s-t+2\tau)K}\right)\quad\text{(Lemma 13)}\\
&+\alpha^2c_4^2\tau\sum_{s=t-\tau}^{t-1}\mathbb{E}_{t-2\tau}[\delta_s^2]+2\alpha^2c_1^2\tau^2\Gamma^2(\epsilon,\epsilon_1)\\
&\overset{(a)}{\leq}\tau^2\alpha^2c_4^2\left[2\mathbb{E}_{t-2\tau}\Big\|\bar{\theta}_{t-\tau}-\theta^*\Big\|^2+96\left(\frac{d_2^2}{NK}\alpha\tau+\frac{2L_2^2\alpha^3}{1-\rho^2}\right)\right]\\
&+6\alpha^2\tau\left[\frac{d_2^2}{NK}\tau+\frac{4L_2^2\alpha^2}{1-\rho^{2K}}\right]+\alpha^2c_4^2\tau(1+16\alpha\tau c_4^2)\sum_{s=0}^{\tau}E_{t-2\tau}[\delta_{t-s}^2]\\
&+96\alpha^2c_4^2\tau^3M_3(\epsilon,\epsilon_1)+2\alpha^2c_1^2\tau^2\Gamma^2(\epsilon,\epsilon_1)\\
&\overset{(b)}{\leq}2\tau^2\alpha^2c_4^2\mathbb{E}_{t-2\tau}\Big\|\bar{\theta}_{t-\tau}-\theta^*\Big\|^2+\frac{d_2^2}{NK}\alpha^2\tau^2\left(96\alpha\tau c_4^2+6\right)+\frac{12L_2^2\alpha^4\tau}{1-\rho^2}\left(16\alpha c_4^2\tau+2\right)\\
&+\alpha^2c_4^2\tau(1+16\alpha\tau c_4^2)\sum_{s=0}^{\tau}E_{t-2\tau}[\Delta_{t-s}]+96\alpha^2c_4^2\tau^3M_3(\epsilon,\epsilon_1)+2\alpha^2c_1^2\tau^2\Gamma^2(\epsilon,\epsilon_1)
\end{aligned}
\tag{51}
$$

Where we used the fact that $\rho^{2\tau K}\leq\alpha^2$ for (a) and (b), and that $\delta_t^2\leq\Delta_t$ (via Jensens' inequality) for all $t\geq 0$ in the last inequality. Let us choose α such that $96\alpha\tau c_4^2+6\leq 7$, $16\alpha c_4^2\tau+2\leq\frac{13}{6}$ and $1+16\alpha\tau c_4^2\leq 2$, this holds when

$$
\alpha\leq\min\left\{\frac{1}{96\tau c_4^2},\frac{1}{96c_4^2\tau},\frac{1}{16\tau c_4^2},1\right\}.
$$

Based on the fact that $\|\bar{\theta}_{t-\tau}-\theta^*\|^2\leq 2\|\bar{\theta}_t-\bar{\theta}_{t-\tau}\|^2+2\|\bar{\theta}_t-\theta^*\|^2$ and the requirement on α, we have

$$
\begin{aligned}
2\alpha^2\tau^2c_4^2\mathbb{E}_{t-2\tau}\|\bar{\theta}_{t-\tau}-\theta^*\|^2&\leq 4\alpha^2\tau^2c_4^2\mathbb{E}_{t-2\tau}\|\bar{\theta}_t-\bar{\theta}_{t-\tau}\|^2+4\alpha^2\tau^2c_4^2\mathbb{E}_{t-2\tau}\|\bar{\theta}_t-\theta^*\|^2\\
&\overset{(a)}{\leq}0.5\mathbb{E}_{t-2\tau}\|\bar{\theta}_t-\bar{\theta}_{t-\tau}\|^2+4\alpha^2\tau^2c_4^2\mathbb{E}_{t-2\tau}\|\bar{\theta}_t-\theta^*\|^2\\
&\overset{(b)}{\leq}\tau^2\alpha^2c_4^2\mathbb{E}_{t-2\tau}\Big\|\bar{\theta}_{t-\tau}-\theta^*\Big\|^2+\frac{7d_2^2}{2NK}\alpha^2\tau^2+\frac{13L_2^2\alpha^4\tau}{(1-\rho^2)}\\
&+\alpha^2c_4^2\tau\sum_{s=0}^{\tau}E_{t-2\tau}[\Delta_{t-s}]+48\alpha^2c_4^2\tau^3M_3(\epsilon,\epsilon_1)+\alpha^2c_1^2\tau^2\Gamma^2(\epsilon,\epsilon_1)\\
&+4\alpha^2\tau^2c_4^2\mathbb{E}_{t-2\tau}\|\bar{\theta}_t-\theta^*\|^2
\end{aligned}
\tag{52}
$$

where (a) is due to $4\alpha^2\tau^2c_4^2\leq 0.5$, and (b) is due to Eq (51) and the choice of α. Putting the term $\tau^2\alpha^2c_4^2\mathbb{E}_{t-2\tau}\Big\|\bar{\theta}_{t-\tau}-\theta^*\Big\|^2$ together by rearranging the terms, we have:

$$
\alpha^2\tau^2c_4^2\mathbb{E}_{t-2\tau}\|\bar{\theta}_{t-\tau}-\theta^*\|^2\leq\frac{7d_2^2}{2NK}\alpha^2\tau^2+\frac{13L_2^2\alpha^4\tau}{(1-\rho^2)}
$$

$$+ \alpha^2 c_4^2 \tau \sum_{s=0}^{\tau} E_{t-2\tau}[\Delta_{t-s}] + 48\alpha^2 c_4^2 \tau^3 M_3(\epsilon, \epsilon_1) + \alpha^2 c_1^2 \tau^2 \Gamma^2(\epsilon, \epsilon_1)$$
$$+ 4\alpha^2 \tau^2 c_4^2 \mathbb{E}_{t-2\tau} \|\bar{\theta}_t - \theta^*\|^2 \tag{53}$$

The proof is completed by substituting this inequality into Eq (51) and the definition of $M_3(\epsilon, \epsilon_1)$. Note that we require the effective step-size

$$\alpha \leq \min\left\{\frac{1}{4h_1\tau}, \frac{1}{2h_1(\tau+1)}, \frac{1}{96c_4^2\tau}, 1\right\}$$

in this proof, which holds when $\alpha \leq \min\left\{\frac{1}{30c_4(\tau+1)}, \frac{1}{96c_4^2\tau}, 1\right\}$ since $c_4 = 3c_1 \geq 1$.

□

J.2.4 Drift Term Analysis.

Now we bound the drift term as follows:

Lemma 15. *(Bounded Client Drift) If $\alpha_l \leq \frac{1}{2\sqrt{2}c_1(K-1)}$, the drift term satisfies*

$$\mathbb{E}[\Delta_t] = \frac{1}{NK}\sum_{i=1}^{N}\sum_{k=0}^{K-1}\mathbb{E}\left\|\theta_{t,k}^{(i)} - \bar{\theta}_t\right\|^2 \leq \frac{4\alpha^2}{K\alpha_g^2}\left[c_3^2 + \frac{2c_3L_2\rho}{1-\rho} + 8c_1^2(K-1)H^2\right]. \tag{54}$$

Proof.

$$\frac{1}{NK}\sum_{i=1}^{N}\sum_{k=0}^{K-1}\mathbb{E}\left\|\theta_{t,k}^{(i)} - \bar{\theta}_t\right\|^2 = \frac{1}{NK}\sum_{i=1}^{N}\sum_{k=0}^{K-1}\mathbb{E}\left\|\bar{\theta}_t + \alpha_l\sum_{s=0}^{k-1} g_i(\theta_{t,s}^{(i)}) - \bar{\theta}_t\right\|^2$$
$$= \alpha_l^2 \frac{1}{NK}\sum_{i=1}^{N}\sum_{k=0}^{K-1}\mathbb{E}\left\|\sum_{s=0}^{k-1} -A_i(O_{t,s}^{(i)})\left(\theta_{t,s}^{(i)} - \theta_i^*\right) + Z_i(O_{t,s}^{(i)})\right\|^2$$
$$\leq 2\alpha_l^2 \frac{1}{NK}\sum_{i=1}^{N}\sum_{k=0}^{K-1}\mathbb{E}\left\|\sum_{s=0}^{k-1} -A_i(O_{t,s}^{(i)})\left(\theta_{t,s}^{(i)} - \theta_i^*\right)\right\|^2 + 2\alpha_l^2 \frac{1}{NK}\sum_{i=1}^{N}\sum_{k=0}^{K-1}\mathbb{E}\left\|\sum_{s=0}^{k-1} Z_i(O_{t,s}^{(i)})\right\|^2$$
$$\leq 2\alpha_l^2 \frac{1}{NK}\sum_{i=1}^{N}\sum_{k=0}^{K-1} k\sum_{s=0}^{k-1}\mathbb{E}\left\|A_i(O_{t,s}^{(i)})\left(\theta_{t,s}^{(i)} - \theta_i^*\right)\right\|^2 + 2\alpha_l^2 \frac{1}{NK}\sum_{i=1}^{N}\sum_{k=0}^{K-1}\sum_{s=0}^{k-1}\mathbb{E}\left\|Z_i(O_{t,s}^{(i)})\right\|^2$$
$$+ 2\alpha_l^2 \frac{1}{NK}\sum_{i=1}^{N}\sum_{k=0}^{K-1}\sum_{\substack{s,s'=0\\ s\neq s'}}^{k-1}\mathbb{E}\left\langle Z_i(O_{t,s}^{(i)}), Z_i(O_{t,s'}^{(i)})\right\rangle$$
$$\leq 2\alpha_l^2 \frac{1}{NK}\sum_{i=1}^{N}\sum_{k=0}^{K-1} kc_1^2\sum_{s=0}^{k-1}\mathbb{E}\left\|\theta_{t,s}^{(i)} - \theta_i^*\right\|^2 + 2\alpha_l^2 \frac{1}{NK}\sum_{i=1}^{N}\sum_{k=0}^{K-1} kc_3^2 \quad \text{(Lemma 11)}$$
$$+ 2\alpha_l^2 \frac{1}{NK}\sum_{i=1}^{N}\sum_{k=0}^{K-1}\sum_{\substack{s,s'=0\\ s\neq s'}}^{k-1}\mathbb{E}\left[\mathbb{E}\left\langle Z_i(O_{t,s}^{(i)}), Z_i(O_{t,s'}^{(i)})\right\rangle \mid \mathcal{F}_s^t\right]$$
$$\leq 2\alpha_l^2 \frac{1}{NK}\sum_{i=1}^{N}\sum_{k=0}^{K-1} kc_1^2\sum_{s=0}^{k-1}\mathbb{E}\left\|\theta_{t,s}^{(i)} - \theta_i^*\right\|^2 + 2\alpha_l^2 \frac{1}{NK}\sum_{i=1}^{N}\sum_{k=0}^{K-1} kc_3^2$$
$$+ 2\alpha_l^2 \frac{1}{NK}\sum_{i=1}^{N}\sum_{k=0}^{K-1}\sum_{\substack{s,s'=0\\ s\neq s'}}^{k-1}\mathbb{E}\left[\left\langle Z_i(O_{t,s}^{(i)}), \mathbb{E}\left[Z_i(O_{t,s'}^{(i)}) \mid \mathcal{F}_s^t\right]\right\rangle\right]$$

$$
\begin{aligned}
&\leq 2\alpha_l^2 \frac{1}{NK}\sum_{i=1}^{N}\sum_{k=0}^{K-1} kc_1^2 \sum_{s=0}^{k-1}\mathbb{E}\left\|\theta_{t,s}^{(i)}-\theta_i^*\right\|^2 + 2\alpha_l^2\frac{1}{NK}\sum_{i=1}^{N}\sum_{k=0}^{K-1} kc_3^2 \\
&+ 4\alpha_l^2 \frac{1}{NK}\sum_{i=1}^{N}\sum_{k=0}^{K-1}\sum_{\substack{s,s'=0\\ s<s'}}^{k-1}\mathbb{E}\left[\left\|Z_i(O_{t,s}^{(i)})\right\|\left\|\mathbb{E}\left[Z_i(O_{t,s'}^{(i)})\mid \mathcal{F}_s^t\right]\right\|\right] \\
&\leq 4\alpha_l^2 \frac{1}{NK}\sum_{i=1}^{N}\sum_{k=0}^{K-1} kc_1^2 \sum_{s=0}^{k-1}\mathbb{E}\left\|\theta_{t,s}^{(i)}-\bar{\theta}_t\right\|^2 + 4\alpha_l^2\frac{1}{NK}\sum_{i=1}^{N}\sum_{k=0}^{K-1} kc_1^2\sum_{s=0}^{k-1}\mathbb{E}\left\|\bar{\theta}_t-\theta_i^*\right\|^2 \quad \text{(Eq (11))} \\
&+ 2\alpha_l^2 \frac{1}{NK}\sum_{i=1}^{N}\sum_{k=0}^{K-1} kc_3^2 + 4\alpha_l^2\frac{1}{NK}\sum_{i=1}^{N}\sum_{k=0}^{K-1}\sum_{\substack{s,s'=0\\ s<s'}}^{k-1} c_3L_2\rho^{s'-s} \quad \text{(Lemma 12)} \\
&\leq 4\alpha_l^2 \frac{1}{NK}\sum_{i=1}^{N}\sum_{k=0}^{K-1} kc_1^2 \sum_{s=0}^{k-1}\mathbb{E}\left\|\theta_{t,s}^{(i)}-\bar{\theta}_t\right\|^2 + 4\alpha_l^2\frac{1}{NK}\sum_{i=1}^{N}\sum_{k=0}^{K-1} 4kc_1^2(K-1)H^2 \\
&+ 2\alpha_l^2 \frac{1}{NK}\sum_{i=1}^{N}\sum_{k=0}^{K-1} kc_3^2 + 4\alpha_l^2\frac{1}{NK}\sum_{i=1}^{N}\sum_{k=0}^{K-1}\underbrace{\sum_{\substack{s,s'=0\\ s<s'}}^{k-1} c_3L_2\rho^{s'-s}}_{\mathcal{M}_1}
\end{aligned} \tag{55}
$$

where we used the property that $\bar{\theta}_t, \theta_i^* \in \mathcal{H}$ in the last inequality, i.e., $\|\bar{\theta}_t\| \leq H^2$ and $\|\theta_i^*\| \leq H^2$. We now bound $\mathcal{M}_1$ as:

$$
\sum_{\substack{s,s'=0\\ s<s'}}^{k-1} c_3L_2\rho^{s'-s} = c_3L_2\sum_{s=0}^{k-1}\sum_{s'=s+1}^{k-1}\rho^{s'-s} = c_3L_2\sum_{s=0}^{k-1}\frac{\rho-\rho^{s-s'}}{1-\rho} \leq c_3L_2\frac{\rho k}{1-\rho} \tag{56}
$$

Define $\mathcal{R}_K \triangleq \sum_{i=1}^{N}\sum_{k=0}^{K-1}\mathbb{E}\left\|\theta_{t,k}^{(i)}-\bar{\theta}_t\right\|^2$ and note that $\mathcal{R}_K$ is monotonically increasing in K. With this definition, if we plug in the upper bound of $\mathcal{M}_1$ into Eq (55), we have:

$$
\begin{aligned}
\mathcal{R}_K &\leq 4\alpha_l^2\sum_{i=1}^{N}\sum_{k=0}^{K-1} kc_1^2\sum_{s=0}^{k-1}\mathbb{E}\left\|\theta_{t,s}^{(i)}-\bar{\theta}_t\right\|^2 + 4\alpha_l^2\sum_{i=1}^{N}\sum_{k=0}^{K-1} 4kc_1^2(K-1)H^2 \\
&+ 2\alpha_l^2\sum_{i=1}^{N}\sum_{k=0}^{K-1} kc_3^2 + 4\alpha_l^2\sum_{i=1}^{N}\sum_{k=0}^{K-1} c_3L_2\frac{\rho k}{1-\rho} \\
&\leq 2\alpha_l^2(K-1)NK\left[c_3^2+\frac{2c_3L_2\rho}{1-\rho}+8c_1^2(K-1)H^2\right] + 4\alpha_l^2c_1^2(K-1)\sum_{k=1}^{K-1}\underbrace{\sum_{i=1}^{N}\sum_{s=0}^{k-1}\mathbb{E}\left\|\theta_{t,s}^{(i)}-\bar{\theta}_t\right\|^2}_{\mathcal{R}_k} \\
&= 2\alpha_l^2(K-1)NK\left[c_3^2+\frac{2c_3L_2\rho}{1-\rho}+8c_1^2(K-1)H^2\right] + 4\alpha_l^2c_1^2(K-1)\sum_{k=1}^{K-1}\mathcal{R}_k
\end{aligned} \tag{57}
$$

By the monotonicity of $\mathcal{R}_k$, we have

$$
\mathcal{R}_K \leq 2\alpha_l^2(K-1)NK\left[c_3^2+\frac{2c_3L_2\rho}{1-\rho}+8c_1^2(K-1)H^2\right] + 4\alpha_l^2c_1^2(K-1)^2\mathcal{R}_{K-1}
$$

Let us choose α_l such that $4\alpha_l^2c_1^2(K-1)^2 \leq \frac{1}{2}$, i.e., $\alpha_l \leq \frac{1}{2\sqrt{2}c_1(K-1)}$, the following recursion holds:

$$
\mathcal{R}_K \leq \frac{1}{2}\mathcal{R}_{K-1} + 2\alpha_l^2(K-1)NK\left[c_3^2+\frac{2c_3L_2\rho}{1-\rho}+8c_1^2(K-1)H^2\right] \tag{58}
$$

for all $k \in [K]$. Next, we unroll the recurrence, go back $K-1$ steps and use the fact that $\mathcal{R}_1 = 0$, we have:

$$\begin{aligned}\mathcal{R}_K &\leq \left\{\sum_{l=1}^{\infty}\left(\frac{1}{2}\right)^l\right\}\left(2\alpha_l^2(K-1)NK\left[c_3^2 + \frac{2c_3L_2\rho}{1-\rho} + 8c_1^2(K-1)H^2\right]\right) \\ &= 4\alpha_l^2(K-1)NK\left[c_3^2 + \frac{2c_3L_2\rho}{1-\rho} + 8c_1^2(K-1)H^2\right] \end{aligned} \tag{59}$$

We finish the proof by dividing NK on both sides and substituting $\alpha_l = \frac{\alpha}{K\alpha_g}$. □

J.2.5 Per Round Progress

Lemma 16. *(Per Round Progress). If the local step-size $\alpha_l \leq \frac{1}{2\sqrt{2}c_1(K-1)}$, and the effective step-size $\alpha = K\alpha_l\alpha_g$ satisfies:*

$$\alpha \leq \min\{\frac{\xi_1}{24(c_1+c_2)^2 + 24\xi_1^2 + 16}, 1, \frac{\xi_1(c_1+c_2)}{2L_1 + 8\tau^2c_4^2}, \frac{1}{30c_4(\tau+1)}, \frac{1}{96c_4^2\tau}, \mathcal{X}\},$$ [4]

where

$$\mathcal{X} = \frac{2B(\epsilon,\epsilon_1)G + 3\xi_1(c_1+c_2)\Gamma^2(\epsilon,\epsilon_1)}{4B^2(\epsilon,\epsilon_1) + 24(c_1+c_2)^2\Gamma^2(\epsilon,\epsilon_1) + 2L_1\Gamma(\epsilon,\epsilon_1)G + 6400c_1^2c_4^2\tau^3\Gamma^2(\epsilon,\epsilon_1) + 8c_1^2\tau^2\Gamma^2(\epsilon,\epsilon_1)},$$

and choose $\tau = \lceil \frac{\tau^{\text{mix}}(\alpha_T^2)}{K} \rceil$, then we have,

$$\begin{aligned}\mathbb{E}_{t-2\tau}\|\bar{\theta}_{t+1} - \theta^*\|^2 \leq & \underbrace{(1 + 32\alpha\xi_1(c_1+c_2))\,\mathbb{E}_{t-2\tau}\left\|\bar{\theta}_t - \theta^*\right\|^2 + 2\alpha\mathbb{E}_{t-2\tau}\left\langle \bar{g}(\bar{\theta}_t), \bar{\theta}_t - \theta^*\right\rangle + 4\alpha^2\mathbb{E}_{t-2\tau}\left\|\bar{g}(\bar{\theta}_t)\right\|^2}_{\textit{Expected progress for the virtual MDP}} \\ & + \underbrace{\frac{9+28\tau^2}{NK}\alpha^2d_2^2}_{\textit{Linear speedup}} + \underbrace{\alpha^3\left(36L_2^2 + \frac{108\tau}{1-\rho^2}L_2^2 + 4L_1G^2 + 2L_2G\right)}_{\textit{High order terms: } O(\alpha^3)} \\ & + \underbrace{\frac{4\alpha^3}{K\alpha_g^2}(\frac{14}{\xi_1} + 14\xi_1)(c_1+c_2)\left[c_3^2 + \frac{2c_3L_2\rho}{1-\rho} + 4c_1^2(K-1)H^2\right]}_{\textit{drift term}} \\ & + \underbrace{4\alpha B(\epsilon,\epsilon_1)G + 6\alpha\xi_1(c_1+c_2)\Gamma^2(\epsilon,\epsilon_1)}_{\textit{heterogeneity term}}. \end{aligned} \tag{60}$$

where ξ_1 is any universal positive constant.

Proof. According to the updating rule and the fact that the projection operator is non-expansive, we have:

$$\begin{aligned}\mathbb{E}_{t-\tau}\left\|\bar{\theta}_{t+1} - \theta^*\right\|^2 &= \mathbb{E}_{t-\tau}\left\|\Pi_{2,\mathcal{H}}\left(\bar{\theta}_t + \frac{\alpha}{NK}\sum_{i=1}^{N}\sum_{k=0}^{K-1} g_i(\theta_{t,k}^{(i)})\right) - \theta^*\right\|^2 \\ &\leq \mathbb{E}_{t-\tau}\left\|\bar{\theta}_t + \frac{\alpha}{NK}\sum_{i=1}^{N}\sum_{k=0}^{K-1} g_i(\theta_{t,k}^{(i)}) - \theta^*\right\|^2\end{aligned}$$

[4] This requirement is very easy to satisfy since the denominator in $\mathcal{X}$ is composed by the heterogeneity terms, which is quite small and thereby makes $\mathcal{X}$ large. Overall, the feasible set of the step-sizes is not empty.

$$
\begin{aligned}
=&\mathbb{E}_{t-\tau}\Big\|\bar{\theta}_t-\theta^*\Big\|^2+2\mathbb{E}_{t-\tau}\Big\langle\frac{\alpha}{NK}\sum_{i=1}^{N}\sum_{k=0}^{K-1}\bar{g}_i(\theta_{t,k}^{(i)}),\bar{\theta}_t-\theta^*\Big\rangle\\
&+2\mathbb{E}_{t-\tau}\Big\langle\frac{\alpha}{NK}\sum_{i=1}^{N}\sum_{k=0}^{K-1}[g_i(\theta_{t,k}^{(i)})-\bar{g}_i(\theta_{t,k}^{(i)})],\bar{\theta}_t-\theta^*\Big\rangle+\alpha^2\mathbb{E}_{t-\tau}\Big\|\frac{1}{NK}\sum_{i=1}^{N}\sum_{k=0}^{K-1}g_i(\theta_{t,k}^{(i)})\Big\|^2\\
\le&\mathbb{E}_{t-\tau}\underbrace{\left\{\Big\|\bar{\theta}_t-\theta^*\Big\|^2+2\Big\langle\frac{\alpha}{N}\sum_{i=1}^{N}\bar{g}_i(\bar{\theta}_t),\bar{\theta}_t-\theta^*\Big\rangle+2\Big\langle\frac{\alpha}{NK}\sum_{i=1}^{N}\sum_{k=0}^{K-1}\bar{g}_i(\theta_{t,k}^{(i)})-\bar{g}_i(\bar{\theta}_t),\bar{\theta}_t-\theta^*\Big\rangle\right\}}_{\mathcal{B}_1}\\
&+2\alpha\mathbb{E}_{t-\tau}\underbrace{\Big\langle\frac{1}{NK}\sum_{i=1}^{N}\sum_{k=0}^{K-1}[g_i(\theta_{t,k}^{(i)})-\bar{g}_i(\theta_{t,k}^{(i)})],\bar{\theta}_t-\theta^*\Big\rangle}_{\mathcal{B}_2}+\alpha^2\mathbb{E}_{t-\tau}\underbrace{\Big\|\frac{1}{NK}\sum_{i=1}^{N}\sum_{k=0}^{K-1}g_i(\theta_{t,k}^{(i)})\Big\|^2}_{\mathcal{B}_3} \qquad (61)
\end{aligned}
$$

We now begin to bound the gradient bias term $\mathcal{B}_2$ by decomposing this term into three terms:

$$
\begin{aligned}
&\Big\langle\frac{1}{NK}\sum_{i=1}^{N}\sum_{k=0}^{K-1}[g_i(\theta_{t,k}^{(i)})-\bar{g}_i(\theta_{t,k}^{(i)})],\bar{\theta}_t-\theta^*\Big\rangle\\
=&\underbrace{\Big\langle\frac{1}{NK}\sum_{i=1}^{N}\sum_{k=0}^{K-1}[g_i(\theta_{t,k}^{(i)})-\bar{g}_i(\theta_{t,k}^{(i)})],\bar{\theta}_t-\bar{\theta}_{t-\tau}\Big\rangle}_{\mathcal{B}_{21}}\\
&+\underbrace{\Big\langle\frac{1}{NK}\sum_{i=1}^{N}\sum_{k=0}^{K-1}[g_i(\theta_{t,k}^{(i)})-g_i(\theta_{t-\tau,k}^{(i)})-\bar{g}_i(\theta_{t,k}^{(i)})+\bar{g}_i(\theta_{t-\tau,k}^{(i)})],\bar{\theta}_{t-\tau}-\theta^*\Big\rangle}_{\mathcal{B}_{22}}\\
&+\underbrace{\Big\langle\frac{1}{NK}\sum_{i=1}^{N}\sum_{k=0}^{K-1}[g_i(\theta_{t-\tau,k}^{(i)})-\bar{g}_i(\theta_{t-\tau,k}^{(i)})],\bar{\theta}_{t-\tau}-\theta^*\Big\rangle}_{\mathcal{B}_{23}}. \qquad (62)
\end{aligned}
$$

Next, we bound $\mathbb{E}_{t-\tau}[\mathcal{B}_{21}]$ as:

$$
\begin{aligned}
&\mathbb{E}_{t-\tau}\Big\langle\frac{1}{NK}\sum_{i=1}^{N}\sum_{k=0}^{K-1}[g_i(\theta_{t,k}^{(i)})-\bar{g}_i(\theta_{t,k}^{(i)})],\bar{\theta}_t-\bar{\theta}_{t-\tau}\Big\rangle\le\mathbb{E}_{t-\tau}\Big\|\frac{1}{NK}\sum_{i=1}^{N}\sum_{k=0}^{K-1}g_i(\theta_{t,k}^{(i)})-\bar{g}_i(\theta_{t,k}^{(i)})\Big\|\Big\|\bar{\theta}_t-\bar{\theta}_{t-\tau}\Big\|\\
&\overset{(a)}{=}\mathbb{E}_{t-\tau}\left[\Big\|\frac{1}{NK}\sum_{i=1}^{N}\sum_{k=0}^{K-1}(-A_i(O_{t,k}^{(i)})+\bar{A}_i)(\theta_{t,k}^{(i)}-\theta_i^*)+Z_i(O_{t,k}^{(i)})\Big\|\Big\|\bar{\theta}_t-\bar{\theta}_{t-\tau}\Big\|\right]\\
&\le\mathbb{E}_{t-\tau}\left[\Big\|\frac{1}{NK}\sum_{i=1}^{N}\sum_{k=0}^{K-1}(A_i(O_{t,k}^{(i)})-\bar{A}_i)(\theta_{t,k}^{(i)}-\theta_i^*)\Big\|\Big\|\bar{\theta}_t-\bar{\theta}_{t-\tau}\Big\|\right]+\mathbb{E}_{t-\tau}\left[\Big\|\frac{1}{NK}\sum_{i=1}^{N}\sum_{k=0}^{K-1}Z_i(O_{t,k}^{(i)})\Big\|\Big\|\bar{\theta}_t-\bar{\theta}_{t-\tau}\Big\|\right]\\
&\le\frac{1}{NK}\sum_{i=1}^{N}\sum_{k=0}^{K-1}\mathbb{E}_{t-\tau}\left[\Big\|(A_i(O_{t,k}^{(i)})-\bar{A}_i)(\theta_{t,k}^{(i)}-\theta_i^*)\Big\|\Big\|\bar{\theta}_t-\bar{\theta}_{t-\tau}\Big\|\right]+\\
&\qquad\frac{\alpha}{2}\mathbb{E}_{t-\tau}\left[\Big\|\frac{1}{NK}\sum_{i=1}^{N}\sum_{k=0}^{K-1}Z_i(O_{t,k}^{(i)})\Big\|^2\right]+\frac{1}{2\alpha}\mathbb{E}_{t-\tau}\left[\Big\|\bar{\theta}_t-\bar{\theta}_{t-\tau}\Big\|^2\right]\\
&\overset{(b)}{\le}\frac{(c_1+c_2)}{NK}\sum_{i=1}^{N}\sum_{k=0}^{K-1}\mathbb{E}_{t-\tau}\Big\|\theta_{t,k}^{(i)}-\theta_i^*\Big\|\Big\|\bar{\theta}_t-\bar{\theta}_{t-\tau}\Big\|+\frac{\alpha}{2}\mathbb{E}_{t-\tau}\Big\|\frac{1}{NK}\sum_{i=1}^{N}\sum_{k=0}^{K-1}Z_i(O_{t,k}^{(i)})\Big\|^2+\frac{1}{2\alpha}\mathbb{E}_{t-\tau}\Big\|\bar{\theta}_t-\bar{\theta}_{t-\tau}\Big\|^2
\end{aligned}
$$

$$
\begin{aligned}
&\leq \frac{\xi_1(c_1+c_2)}{2NK}\sum_{i=1}^{N}\sum_{k=0}^{K-1}\mathbb{E}_{t-\tau}\left\|\theta_{t,k}^{(i)}-\theta_i^*\right\|^2+\frac{(c_1+c_2)}{2\xi_1 NK}\sum_{i=1}^{N}\sum_{k=0}^{K-1}\mathbb{E}_{t-\tau}\left\|\bar{\theta}_t-\bar{\theta}_{t-\tau}\right\|^2 \quad \text{(Young's inequality (12))}\\
&+\frac{\alpha}{2}\mathbb{E}_{t-\tau}\left\|\frac{1}{NK}\sum_{i=1}^{N}\sum_{k=0}^{K-1}Z_i(O_{t,k}^{(i)})\right\|^2+\frac{1}{2\alpha}\mathbb{E}_{t-\tau}\left\|\bar{\theta}_t-\bar{\theta}_{t-\tau}\right\|^2\\
&\overset{(c)}{\leq}\frac{3\xi_1(c_1+c_2)}{2NK}\sum_{i=1}^{N}\sum_{k=0}^{K-1}\mathbb{E}_{t-\tau}\left\|\theta_{t,k}^{(i)}-\bar{\theta}_t\right\|^2+\frac{3\xi_1(c_1+c_2)}{2NK}\sum_{i=1}^{N}\sum_{k=0}^{K-1}\mathbb{E}_{t-\tau}\left\|\bar{\theta}_t-\theta^*\right\|^2\\
&+\frac{3\xi_1(c_1+c_2)}{2NK}\sum_{i=1}^{N}\sum_{k=0}^{K-1}\mathbb{E}_{t-\tau}\left\|\theta^*-\theta_i^*\right\|^2\\
&+\frac{(c_1+c_2)}{2\xi_1 NK}\sum_{i=1}^{N}\sum_{k=0}^{K-1}\mathbb{E}_{t-\tau}\left\|\bar{\theta}_t-\bar{\theta}_{t-\tau}\right\|^2+\frac{\alpha}{2}\mathbb{E}_{t-\tau}\left\|\frac{1}{NK}\sum_{i=1}^{N}\sum_{k=0}^{K-1}Z_i(O_{t,k}^{(i)})\right\|^2+\frac{1}{2\alpha}\mathbb{E}_{t-\tau}\left\|\bar{\theta}_t-\bar{\theta}_{t-\tau}\right\|^2\\
&=\frac{3\xi_1(c_1+c_2)}{2NK}\sum_{i=1}^{N}\sum_{k=0}^{K-1}\mathbb{E}_{t-\tau}\left\|\theta_{t,k}^{(i)}-\bar{\theta}_t\right\|^2+\frac{3\xi_1(c_1+c_2)}{2}\mathbb{E}_{t-\tau}\left\|\bar{\theta}_t-\theta^*\right\|^2+\frac{3\xi_1(c_1+c_2)}{2}\Gamma^2(\epsilon,\epsilon_1)\\
&+\frac{(c_1+c_2)}{2\xi_1}\mathbb{E}_{t-\tau}\left\|\bar{\theta}_t-\bar{\theta}_{t-\tau}\right\|^2+\frac{\alpha}{2}\mathbb{E}_{t-\tau}\left\|\frac{1}{NK}\sum_{i=1}^{N}\sum_{k=0}^{K-1}Z_i(O_{t,k}^{(i)})\right\|^2+\frac{1}{2\alpha}\mathbb{E}_{t-\tau}\left\|\bar{\theta}_t-\bar{\theta}_{t-\tau}\right\|^2\\
&\overset{(d)}{\leq}\frac{3\xi_1(c_1+c_2)}{2NK}\sum_{i=1}^{N}\sum_{k=0}^{K-1}\mathbb{E}_{t-\tau}\left\|\theta_{t,k}^{(i)}-\bar{\theta}_t\right\|^2+\frac{3\xi_1(c_1+c_2)}{2}\mathbb{E}_{t-\tau}\left\|\bar{\theta}_t-\theta^*\right\|^2+\frac{3\xi_1(c_1+c_2)}{2}\Gamma^2(\epsilon,\epsilon_1)\\
&+\frac{c_1+c_2}{2\xi_1}\mathbb{E}_{t-\tau}\left\|\bar{\theta}_t-\bar{\theta}_{t-\tau}\right\|^2+\frac{\alpha}{2}\left[\frac{d_2^2}{NK}+4L_2^2\rho^{2\tau K}\right]+\frac{1}{2\alpha}\mathbb{E}_{t-\tau}\left\|\bar{\theta}_t-\bar{\theta}_{t-\tau}\right\|^2\\
&=\frac{3\xi_1(c_1+c_2)}{2NK}\sum_{i=1}^{N}\sum_{k=0}^{K-1}\mathbb{E}_{t-\tau}\left\|\theta_{t,k}^{(i)}-\bar{\theta}_t\right\|^2+\frac{3\xi_1(c_1+c_2)}{2}\mathbb{E}_{t-\tau}\left\|\bar{\theta}_t-\theta^*\right\|^2+\frac{3\xi_1(c_1+c_2)}{2}\Gamma^2(\epsilon,\epsilon_1)\\
&+\left(\frac{c_1+c_2}{2\xi_1}+\frac{1}{2\alpha}\right)\mathbb{E}_{t-\tau}\left\|\bar{\theta}_t-\bar{\theta}_{t-\tau}\right\|^2+\frac{\alpha}{2}\left[\frac{d_2^2}{NK}+\underbrace{4L_2^2\rho^{2\tau K}}_{\leq 4L_2^2\alpha^2}\right], \qquad (63)
\end{aligned}
$$

where (a) is due to $g_i(\theta_{t,k}^{(i)})=-A_i(O_{t,k}^{(i)})(\theta_{t,k}^{(i)}-\theta_i^*)+Z_i(O_{t,k}^{(i)})$, (b) is due to Lemma 12 (the upper bound of $A_i(O_{t,k}^{(i)})$ and $\bar{A}_i$), (c) is due to Eq (13) and (d) is due to Lemma 13.

And we bound $\mathcal{B}_{22}$ as:

$$
\begin{aligned}
\mathcal{B}_{22}&=\left\langle\frac{1}{NK}\sum_{i=1}^{N}\sum_{k=0}^{K-1}\left[g_i(\theta_{t,k}^{(i)})-g_i(\theta_{t-\tau,k}^{(i)})-\bar{g}_i(\theta_{t,k}^{(i)})+\bar{g}_i(\theta_{t-\tau,k}^{(i)})\right],\bar{\theta}_{t-\tau}-\theta^*\right\rangle\\
&\leq\frac{1}{NK}\sum_{i=1}^{N}\sum_{k=0}^{K-1}\left\|g_i(\theta_{t,k}^{(i)})-g_i(\theta_{t-\tau,k}^{(i)})-\bar{g}_i(\theta_{t,k}^{(i)})+\bar{g}_i(\theta_{t-\tau,k}^{(i)})\right\|\left\|\bar{\theta}_{t-\tau}-\theta^*\right\| \text{ (Cauchy-Schwarz inequality)}\\
&\leq\frac{1}{NK}\sum_{i=1}^{N}\sum_{k=0}^{K-1}\left[\left\|g_i(\theta_{t,k}^{(i)})-g_i(\theta_{t-\tau,k}^{(i)})\right\|+\left\|\bar{g}_i(\theta_{t,k}^{(i)})-\bar{g}_i(\theta_{t-\tau,k}^{(i)})\right\|\right]\left\|\bar{\theta}_{t-\tau}-\theta^*\right\|\\
&\overset{(a)}{\leq}\frac{1}{NK}\sum_{i=1}^{N}\sum_{k=0}^{K-1}\left[2\left\|\theta_{t,k}^{(i)}-\theta_{t-\tau,k}^{(i)}\right\|+2\left\|\theta_{t,k}^{(i)}-\theta_{t-\tau,k}^{(i)}\right\|\right]\left\|\bar{\theta}_{t-\tau}-\theta^*\right\|
\end{aligned}
$$

$$
\begin{aligned}
&\leq \frac{1}{NK}\sum_{i=1}^{N}\sum_{k=0}^{K-1}\left[4\left\|\theta_{t,k}^{(i)}-\bar{\theta}_t\right\|+4\left\|\bar{\theta}_t-\bar{\theta}_{t-\tau}\right\|+4\left\|\bar{\theta}_{t-\tau}-\theta_{t-\tau,k}^{(i)}\right\|\right]\left\|\bar{\theta}_{t-\tau}-\theta^*\right\| \text{ (Triangle inequality)}\\
&\leq 4\delta_t\left\|\bar{\theta}_{t-\tau}-\theta^*\right\|+4\left\|\bar{\theta}_t-\bar{\theta}_{t-\tau}\right\|\left\|\bar{\theta}_{t-\tau}-\theta^*\right\|+4\delta_{t-\tau}\left\|\bar{\theta}_{t-\tau}-\theta^*\right\|\\
&\overset{(b)}{\leq} \frac{2}{\xi_2}\Delta_t+\frac{2}{\xi_2}\Delta_{t-\tau}+(2\xi_2+4\xi_2)\left\|\bar{\theta}_{t-\tau}-\theta^*\right\|^2+\frac{2}{\xi_2}\left\|\bar{\theta}_t-\bar{\theta}_{t-\tau}\right\|^2\\
&\leq \frac{2}{\xi_2}\Delta_t+\frac{2}{\xi_2}\Delta_{t-\tau}+12\xi_2\left\|\bar{\theta}_t-\theta^*\right\|^2+(12\xi_2+\frac{2}{\xi_2})\left\|\bar{\theta}_t-\bar{\theta}_{t-\tau}\right\|^2 \quad \text{(Eq 13)} \qquad (64)
\end{aligned}
$$

where (a) is due to the 2-Lipschitz property of steady-state $\bar{g}$ (i.e., Lemma 5) and random direction g_i (i.e., Lemma 6), $\delta_t = \frac{1}{NK}\sum_{i=1}^{N}\sum_{k=0}^{K-1}\left\|\theta_{t,k}^{(i)}-\bar{\theta}_t\right\|$ and $\Delta_t \triangleq \frac{1}{NK}\sum_{i=1}^{N}\sum_{k=0}^{K-1}\mathbb{E}\left\|\theta_{t,k}^{(i)}-\bar{\theta}_t\right\|^2$, and (b) is due to Young's inequality (12) with constants ξ_2 and $\delta_t^2 \leq \Delta_t$.

Now, we bound $\mathcal{B}_{23}$ as:

$$
\begin{aligned}
\mathbb{E}_{t-\tau}[\mathcal{B}_{23}] &= \left\langle \frac{1}{NK}\sum_{i=1}^{N}\sum_{k=0}^{K-1}\mathbb{E}_{t-\tau}[g_i(\theta_{t-\tau,k}^{(i)})-\bar{g}_i(\theta_{t-\tau,k}^{(i)})], \bar{\theta}_{t-\tau}-\theta^*\right\rangle\\
&\leq \frac{1}{NK}\sum_{i=1}^{N}\sum_{k=0}^{K-1}\left\|\bar{\theta}_{t-\tau}-\theta^*\right\|\left\|\mathbb{E}_{t-\tau}\left[g_i(\theta_{t-\tau,k}^{(i)})-\bar{g}_i(\theta_{t-\tau,k}^{(i)})\right]\right\| \text{ (Cauchy-Schwarz inequality)}\\
&= \frac{1}{NK}\sum_{i=1}^{N}\sum_{k=0}^{K-1}\left\|\bar{\theta}_{t-\tau}-\theta^*\right\|\left\|\mathbb{E}_{t-\tau}\left[-A_i(O_{t,k}^{(i)})(\theta_{t-\tau,k}^{(i)}-\theta_i^*)+Z_i(O_{t,k}^{(i)})+\bar{A}_i(\theta_{t-\tau,k}^{(i)}-\theta_i^*)\right]\right\|\\
&\leq \frac{1}{NK}\sum_{i=1}^{N}\sum_{k=0}^{K-1}\left\|\bar{\theta}_{t-\tau}-\theta^*\right\|\left\{\left\|\mathbb{E}_{t-\tau}(A_i(O_{t,k}^{(i)})-\bar{A}_i)(\theta_{t-\tau,k}^{(i)}-\theta_i^*)\right\|+\left\|\mathbb{E}_{t-\tau}\left[Z_i(O_{t,k}^{(i)})\right]\right\|\right\}\\
&\overset{(a)}{\leq} \frac{1}{NK}\sum_{i=1}^{N}\sum_{k=0}^{K-1}\left\|\bar{\theta}_{t-\tau}-\theta^*\right\|\left\{L_1\rho^{\tau K+k}\left\|\theta_{t-\tau,k}^{(i)}-\theta_i^*\right\|+L_2\rho^{\tau K+k}\right\}\\
&\leq \frac{1}{NK}\sum_{i=1}^{N}\sum_{k=0}^{K-1}\left\|\bar{\theta}_{t-\tau}-\theta^*\right\|\left\{L_1\rho^{\tau K+k}\left[\left\|\theta_{t-\tau,k}^{(i)}-\bar{\theta}_{t-\tau}\right\|+\left\|\bar{\theta}_{t-\tau}-\theta^*\right\|+\left\|\theta^*-\theta_i^*\right\|\right]+L_2\rho^{\tau K+k}\right\}\\
&\overset{(b)}{\leq} \alpha^2 L_1\left\|\bar{\theta}_{t-\tau}-\theta^*\right\|\delta_{t-\tau}+\alpha^2 L_1\left\|\bar{\theta}_{t-\tau}-\theta^*\right\|^2+\alpha^2 L_1\Gamma(\epsilon,\epsilon_1)G+\alpha^2 L_2 G\\
&\leq \alpha^2 L_1\left\|\bar{\theta}_{t-\tau}-\theta^*\right\|^2+\alpha^2 L_1\Delta_{t-\tau}+\alpha^2 L_1\left\|\bar{\theta}_{t-\tau}-\theta^*\right\|^2+\alpha^2 L_1\Gamma(\epsilon,\epsilon_1)G+\alpha^2 L_2 G\\
&\overset{(c)}{\leq} 2\alpha^2 L_1 G^2+\alpha^2 L_2 G+\alpha^2 L_1\Delta_{t-\tau}+\alpha^2 L_1\Gamma(\epsilon,\epsilon_1)G, \qquad (65)
\end{aligned}
$$

where (a) is due to Lemma 12, (b) is due to the fact that $\bar{\theta}_{t-\tau}, \theta^* \in \mathcal{H}$, which radius is $H \leq \frac{G}{2}$, and $\tau = \lceil \frac{\log_\rho(\alpha_T^2)}{K} \rceil$ (i.e., $\rho^{\tau K} \leq \alpha^2$) and (c) is due to the fact that $\bar{\theta}_{t-\tau}, \theta^* \in \mathcal{H}$. Then, the term $\mathcal{B}_2$ can be bounded as:

$$
\begin{aligned}
\mathbb{E}_{t-\tau}[\mathcal{B}_2] &= \mathbb{E}_{t-\tau}[\mathcal{B}_{21}+\mathcal{B}_{22}+\mathcal{B}_{23}]\\
&\leq \frac{3\xi_1(c_1+c_2)}{2}\mathbb{E}_{t-\tau}[\Delta_t]+\frac{3\xi_1(c_1+c_2)}{2}\mathbb{E}_{t-\tau}\left\|\bar{\theta}_t-\theta^*\right\|^2+\frac{3\xi_1(c_1+c_2)}{2}\Gamma^2(\epsilon,\epsilon_1)\\
&+\left(\frac{c_1+c_2}{2\xi_1}+\frac{1}{2\alpha}\right)\mathbb{E}_{t-\tau}\left\|\bar{\theta}_t-\bar{\theta}_{t-\tau}\right\|^2+\frac{\alpha}{2}\left[\frac{d_2^2}{NK}+4L_2^2\rho^{2\tau K}\right]
\end{aligned}
$$

$$
\begin{aligned}
&+\frac{2}{\xi_2}\mathbb{E}_{t-\tau}[\Delta_t]+\frac{2}{\xi_2}\mathbb{E}_{t-\tau}[\Delta_{t-\tau}]+12\xi_2\mathbb{E}_{t-\tau}\left\|\bar{\theta}_t-\theta^*\right\|^2+(12\xi_2+\frac{2}{\xi_2})\mathbb{E}_{t-\tau}\left\|\bar{\theta}_t-\bar{\theta}_{t-\tau}\right\|^2\\
&+2\alpha^2L_1G^2+\alpha^2L_2G+\alpha^2L_1\mathbb{E}_{t-\tau}[\Delta_{t-\tau}]+\alpha^2L_1\Gamma(\epsilon,\epsilon_1)G\\
&\leq\left(\frac{3\xi_1(c_1+c_2)}{2}+12\xi_2\right)\mathbb{E}_{t-\tau}\left\|\bar{\theta}_t-\theta^*\right\|^2+\left(\frac{c_1+c_2}{2\xi_1}+\frac{1}{2\alpha}+12\xi_2+\frac{2}{\xi_2}\right)\mathbb{E}_{t-\tau}\left\|\bar{\theta}_t-\bar{\theta}_{t-\tau}\right\|^2\\
&+\left(\frac{3\xi_1(c_1+c_2)}{2}+\frac{2}{\xi_2}\right)\mathbb{E}_{t-\tau}[\Delta_t]+\left(\frac{2}{\xi_2}+\alpha^2L_1\right)\Delta_{t-\tau}+\frac{\alpha}{2}\left[\frac{d_2^2}{NK}+4L_2^2\alpha^2\right]\\
&+2\alpha^2L_1G^2+\alpha^2L_2G+\frac{3\xi_1(c_1+c_2)}{2}\Gamma^2(\epsilon,\epsilon_1)+\alpha^2L_1\Gamma(\epsilon,\epsilon_1)G \qquad (66)
\end{aligned}
$$

Next, we bound $\mathcal{B}_3$ as:

$$
\begin{aligned}
\mathbb{E}_{t-\tau}[\mathcal{B}_3]&=\mathbb{E}_{t-\tau}\left\|\frac{1}{NK}\sum_{i=1}^{N}\sum_{k=0}^{K-1}\left[g_i(\theta_{t,k}^{(i)})-\bar{g}_i(\theta_{t,k}^{(i)})+\bar{g}_i(\theta_{t,k}^{(i)})-\bar{g}_i(\bar{\theta}_t)+\bar{g}_i(\bar{\theta}_t)-\bar{g}(\bar{\theta}_t)+\bar{g}(\bar{\theta}_t)\right]\right\|^2\\
&\leq 4\mathbb{E}_{t-\tau}\left\|\frac{1}{NK}\sum_{i=1}^{N}\sum_{k=0}^{K-1}\left(g_i(\theta_{t,k}^{(i)})-\bar{g}_i(\theta_{t,k}^{(i)})\right)\right\|^2+4\mathbb{E}_{t-\tau}\left\|\frac{1}{NK}\sum_{i=1}^{N}\sum_{k=0}^{K-1}\left(\bar{g}_i(\theta_{t,k}^{(i)})-\bar{g}_i(\bar{\theta}_t)\right)\right\|^2\\
&+4\mathbb{E}_{t-\tau}\left\|\frac{1}{NK}\sum_{i=1}^{N}\sum_{k=0}^{K-1}\left(\bar{g}_i(\bar{\theta}_t)-\bar{g}(\bar{\theta}_t)\right)\right\|^2+4\mathbb{E}_{t-\tau}\left\|\frac{1}{NK}\sum_{i=1}^{N}\sum_{k=0}^{K-1}\bar{g}(\bar{\theta}_t)\right\|^2 \quad \text{(Eq 13)}\\
&=4\mathbb{E}_{t-\tau}\left\|\frac{1}{NK}\sum_{i=1}^{N}\sum_{k=0}^{K-1}\left[\left(\bar{A}_i-A_i(O_{t,k}^{(i)})\right)(\theta_{t,k}^{(i)}-\theta_i^*)+Z_i(O_{t,k}^{(i)})\right]\right\|^2\\
&+4\mathbb{E}_{t-\tau}\left\|\frac{1}{NK}\sum_{i=1}^{N}\sum_{k=0}^{K-1}\left(\bar{g}_i(\theta_{t,k}^{(i)})-\bar{g}_i(\bar{\theta}_t)\right)\right\|^2+\underbrace{4\mathbb{E}_{t-\tau}\left\|\frac{1}{N}\sum_{i=1}^{N}\left(\bar{g}_i(\bar{\theta}_t)-\bar{g}(\bar{\theta}_t)\right)\right\|^2}_{\text{Lemma 2}}+4\mathbb{E}_{t-\tau}\left\|\bar{g}(\bar{\theta}_t)\right\|^2\\
&\overset{(a)}{\leq}8\mathbb{E}_{t-\tau}\left\|\frac{1}{NK}\sum_{i=1}^{N}\sum_{k=0}^{K-1}\left(\bar{A}_i-A_i(O_{t,k}^{(i)})\right)(\theta_{t,k}^{(i)}-\theta_i^*)\right\|^2+8\mathbb{E}_{t-\tau}\left\|\frac{1}{NK}\sum_{i=1}^{N}\sum_{k=0}^{K-1}Z_i(O_{t,k}^{(i)})\right\|^2\\
&+16\frac{1}{NK}\sum_{i=1}^{N}\sum_{k=0}^{K-1}\mathbb{E}_{t-\tau}\left\|\theta_{t,k}^{(i)}-\bar{\theta}_t\right\|^2+4B^2(\epsilon,\epsilon_1)+4\mathbb{E}_{t-\tau}\left\|\bar{g}(\bar{\theta}_t)\right\|^2\\
&\leq\frac{8}{NK}\sum_{i=1}^{N}\sum_{k=0}^{K-1}\mathbb{E}_{t-\tau}\left\|\bar{A}_i-A_i(O_{t,k}^{(i)})\right\|^2\left\|\theta_{t,k}^{(i)}-\theta_i^*\right\|^2+8\mathbb{E}_{t-\tau}\left\|\frac{1}{NK}\sum_{i=1}^{N}\sum_{k=0}^{K-1}Z_i(O_{t,k}^{(i)})\right\|^2\\
&+16\mathbb{E}_{t-\tau}[\Delta_t]+4B^2(\epsilon,\epsilon_1)+4\mathbb{E}_{t-\tau}\left\|\bar{g}(\bar{\theta}_t)\right\|^2\\
&\leq\frac{8(c_1+c_2)^2}{NK}\sum_{i=1}^{N}\sum_{k=0}^{K-1}\mathbb{E}_{t-\tau}\left\|\theta_{t,k}^{(i)}-\theta_i^*\right\|^2+8\mathbb{E}_{t-\tau}\left\|\frac{1}{NK}\sum_{i=1}^{N}\sum_{k=0}^{K-1}Z_i(O_{t,k}^{(i)})\right\|^2\\
&+16\mathbb{E}_{t-\tau}[\Delta_t]+4B^2(\epsilon,\epsilon_1)+4\mathbb{E}_{t-\tau}\left\|\bar{g}(\bar{\theta}_t)\right\|^2\\
&\leq\frac{24(c_1+c_2)^2}{NK}\sum_{i=1}^{N}\sum_{k=0}^{K-1}\mathbb{E}_{t-\tau}\left\|\theta_{t,k}^{(i)}-\bar{\theta}_t\right\|^2+\frac{24(c_1+c_2)^2}{NK}\sum_{i=1}^{N}\sum_{k=0}^{K-1}\mathbb{E}_{t-\tau}\left\|\bar{\theta}_t-\theta^*\right\|^2\\
&+\frac{24(c_1+c_2)^2}{NK}\sum_{i=1}^{N}\sum_{k=0}^{K-1}\mathbb{E}_{t-\tau}\left\|\theta_i^*-\theta^*\right\|^2+8\mathbb{E}_{t-\tau}\left\|\frac{1}{NK}\sum_{i=1}^{N}\sum_{k=0}^{K-1}Z_i(O_{t,k}^{(i)})\right\|^2\\
&+16\mathbb{E}_{t-\tau}[\Delta_t]+4B^2(\epsilon,\epsilon_1)+4\mathbb{E}_{t-\tau}\left\|\bar{g}(\bar{\theta}_t)\right\|^2
\end{aligned}
$$

$$
\begin{aligned}
&= 24(c_1+c_2)^2\mathbb{E}_{t-\tau}[\Delta_t] + 24(c_1+c_2)^2\mathbb{E}_{t-\tau}\left\|\bar{\theta}_t-\theta^*\right\|^2 + 24(c_1+c_2)^2\Gamma^2(\epsilon,\epsilon_1)\\
&+ 8\mathbb{E}_{t-\tau}\Big\|\frac{1}{NK}\sum_{i=1}^{N}\sum_{k=0}^{K-1} Z_i(O_{t,k}^{(i)})\Big\|^2 + 16\mathbb{E}_{t-\tau}[\Delta_t] + 4B^2(\epsilon,\epsilon_1) + 4\mathbb{E}_{t-\tau}\left\|\bar{g}(\bar{\theta}_t)\right\|^2\\
&\overset{(b)}{\leq} (24(c_1+c_2)^2+16)\mathbb{E}_{t-\tau}[\Delta_t] + 8(\frac{d_2^2}{NK} + \underbrace{4L_2^2\rho^{\tau K}}_{\leq 4L_2^2\alpha^2}) + 24(c_1+c_2)^2\mathbb{E}_{t-\tau}\left\|\bar{\theta}_t-\theta^*\right\|^2\\
&+ 4\mathbb{E}_{t-\tau}\left\|\bar{g}(\bar{\theta}_t)\right\|^2 + 4B^2(\epsilon,\epsilon_1) + 24(c_1+c_2)^2\Gamma^2(\epsilon,\epsilon_1), \qquad (67)
\end{aligned}
$$

where (a) is due to 2-Lipschitz of $\bar{g}_i$ (i.e., Lemma 5) and the gradient heterogeneity (i.e., Lemma 2) and (b) is due to Lemma 13.

Next, we bound $\mathcal{B}_1$ as:

$$
\begin{aligned}
\mathbb{E}_{t-\tau}[\mathcal{B}_1] &= \mathbb{E}_{t-\tau}\left\|\bar{\theta}_t-\theta^*\right\|^2 + 2\mathbb{E}_{t-\tau}\Big\langle\frac{\alpha}{N}\sum_{i=1}^{N}\bar{g}_i(\bar{\theta}_t),\bar{\theta}_t-\theta^*\Big\rangle + 2\mathbb{E}_{t-\tau}\Big\langle\frac{\alpha}{NK}\sum_{i=1}^{N}\sum_{k=0}^{K-1}\bar{g}_i(\theta_{t,k}^{(i)})-\bar{g}_i(\bar{\theta}_t),\bar{\theta}_t-\theta^*\Big\rangle\\
&\leq \mathbb{E}_{t-\tau}\left\|\bar{\theta}_t-\theta^*\right\|^2 + 2\alpha\mathbb{E}_{t-\tau}\Big\langle\frac{1}{N}\sum_{i=1}^{N}\bar{g}_i(\bar{\theta}_t)-\bar{g}(\bar{\theta}_t),\bar{\theta}_t-\theta^*\Big\rangle + 2\alpha\mathbb{E}_{t-\tau}\Big\langle\bar{g}(\bar{\theta}_t),\bar{\theta}_t-\theta^*\Big\rangle\\
&+ 2\alpha\mathbb{E}_{t-\tau}\Big\langle\frac{1}{NK}\sum_{i=1}^{N}\sum_{k=0}^{K-1}\bar{g}_i(\theta_{t,k}^{(i)})-\bar{g}_i(\bar{\theta}_t),\bar{\theta}_t-\theta^*\Big\rangle\\
&\leq \mathbb{E}_{t-\tau}\left\|\bar{\theta}_t-\theta^*\right\|^2 + 2\alpha\mathbb{E}_{t-\tau}\Big\|\frac{1}{N}\sum_{i=1}^{N}\bar{g}_i(\bar{\theta}_t)-\bar{g}(\bar{\theta}_t)\Big\|\left\|\bar{\theta}_t-\theta^*\right\| + 2\alpha\mathbb{E}_{t-\tau}\Big\langle\bar{g}(\bar{\theta}_t),\bar{\theta}_t-\theta^*\Big\rangle\\
&+ \frac{\alpha}{\xi_3}\mathbb{E}_{t-\tau}\Big\|\frac{1}{NK}\sum_{i=1}^{N}\sum_{k=0}^{K-1}\left(\bar{g}_i(\theta_{t,k}^{(i)})-\bar{g}_i(\bar{\theta}_t)\right)\Big\|^2 + \alpha\xi_3\mathbb{E}_{t-\tau}\left\|\bar{\theta}_t-\theta^*\right\|^2\\
&\qquad \text{(Young's inequality Eq (12) with constant } \xi_3)\\
&\overset{(a)}{\leq} \mathbb{E}_{t-\tau}\left\|\bar{\theta}_t-\theta^*\right\|^2 + 2\alpha B(\epsilon,\epsilon_1)G + 2\alpha\mathbb{E}_{t-\tau}\Big\langle\bar{g}(\bar{\theta}_t),\bar{\theta}_t-\theta^*\Big\rangle\\
&+ \frac{\alpha}{\xi_3}\mathbb{E}_{t-\tau}\Big\|\frac{1}{NK}\sum_{i=1}^{N}\sum_{k=0}^{K-1}\left(\bar{g}_i(\theta_{t,k}^{(i)})-\bar{g}_i(\bar{\theta}_t)\right)\Big\|^2 + \alpha\xi_3\mathbb{E}_{t-\tau}\left\|\bar{\theta}_t-\theta^*\right\|^2\\
&\overset{(b)}{\leq} \mathbb{E}_{t-\tau}\left\|\bar{\theta}_t-\theta^*\right\|^2 + 2\alpha B(\epsilon,\epsilon_1)G + 2\alpha\mathbb{E}_{t-\tau}\Big\langle\bar{g}(\bar{\theta}_t),\bar{\theta}_t-\theta^*\Big\rangle + \frac{4\alpha}{\xi_3}\mathbb{E}_{t-\tau}[\Delta_t] + \alpha\xi_3\mathbb{E}_{t-\tau}\left\|\bar{\theta}_t-\theta^*\right\|^2,
\end{aligned}
\qquad (68)
$$

where (a) is due to the fact that $\bar{\theta}_t, \theta^* \in \mathcal{H}$ and the gradient heterogeneity; (b) is due to 2-Lipschitz property of function $\bar{g}$ in Lemma 5.

Incorporating the upper of $\mathcal{B}_1$ from Eq (68), $\mathcal{B}_2$ from Eq (66) and $\mathcal{B}_3$ from Eq (67) into Eq (61), we have:

$$
\begin{aligned}
\mathbb{E}_{t-\tau}\left\|\bar{\theta}_{t+1}-\theta^*\right\|^2 &\leq \mathbb{E}_{t-\tau}\left\|\bar{\theta}_t-\theta^*\right\|^2 + 2\alpha\mathbb{E}_{t-\tau}\Big\langle\bar{g}(\bar{\theta}_t),\bar{\theta}_t-\theta^*\Big\rangle + 4\alpha^2\mathbb{E}_{t-\tau}\left\|\bar{g}(\bar{\theta}_t)\right\|^2\\
&+ \left(\alpha\xi_3 + \alpha(3\xi_1(c_1+c_2)+24\xi_2) + 24\alpha^2(c_1+c_2)^2\right)\mathbb{E}_{t-\tau}\left\|\bar{\theta}_t-\theta^*\right\|^2\\
&+ \alpha\left(\frac{c_1+c_2}{\xi_1} + \frac{1}{\alpha} + 24\xi_2 + \frac{4}{\xi_2}\right)\mathbb{E}_{t-\tau}\left\|\bar{\theta}_t-\bar{\theta}_{t-\tau}\right\|^2
\end{aligned}
$$

$$
\begin{aligned}
&+\frac{9d_2^2}{NK}\alpha^2+36L_2^2\alpha^4+4\alpha^3L_1G^2+2\alpha^3L_2G+\left(\frac{4\alpha}{\xi_2}+2\alpha^3L_1\right)\Delta_{t-\tau}\\
&+\alpha\left(\frac{4}{\xi_3}+3\xi_1(c_1+c_2)+\frac{4}{\xi_2}+\alpha^2\left(24(c_1+c_2)^2+16\right)\right)\mathbb{E}_{t-\tau}[\Delta_t]\\
&+2\alpha B(\epsilon,\epsilon_1)G+4\alpha^2B^2(\epsilon,\epsilon_1)+24\alpha^2(c_1+c_2)^2\Gamma^2(\epsilon,\epsilon_1)\\
&+3\alpha\xi_1(c_1+c_2)\Gamma^2(\epsilon,\epsilon_1)+2\alpha^3L_1\Gamma(\epsilon,\epsilon_1)G \qquad (69)
\end{aligned}
$$

Conditioned on $\mathcal{F}_{t-2\tau}$ and using Lemma 14 to give an upper bound of $\mathbb{E}_{t-2\tau}\left\|\bar{\theta}_t-\bar{\theta}_{t-\tau}\right\|^2$, we have:

$$
\begin{aligned}
\mathbb{E}_{t-2\tau}\left\|\bar{\theta}_{t+1}-\theta^*\right\|^2 \le\ & \mathbb{E}_{t-2\tau}\left\|\bar{\theta}_t-\theta^*\right\|^2+2\alpha\mathbb{E}_{t-2\tau}\left\langle\bar{g}(\bar{\theta}_t),\bar{\theta}_t-\theta^*\right\rangle+4\alpha^2\mathbb{E}_{t-2\tau}\left\|\bar{g}(\bar{\theta}_t)\right\|^2\\
&+\underbrace{\left(\alpha\xi_3+\alpha(3\xi_1(c_1+c_2)+24\xi_2)+24\alpha^2(c_1+c_2)^2\right)}_{\mathcal{E}_1}\mathbb{E}_{t-2\tau}\left\|\bar{\theta}_t-\theta^*\right\|^2\\
&+\alpha\underbrace{\left(\frac{c_1+c_2}{\xi_1}+\frac{1}{\alpha}+24\xi_2+\frac{4}{\xi_2}\right)}_{\mathcal{E}_2}\Bigg\{8\alpha^2\tau^2c_4^2\mathbb{E}_{t-2\tau}\left[\left\|\bar{\theta}_t-\theta^*\right\|^2\right]+14\alpha^2\tau^2\frac{d_2^2}{NK}+\frac{52L_2^2\alpha^4\tau}{1-\rho^2}\\
&+4\alpha^2c_4^2\tau\sum_{s=0}^{\tau}E_{t-2\tau}[\Delta_{t-s}]+3200\alpha^2c_1^2c_4^2\tau^3\Gamma^2(\epsilon,\epsilon_1)+4\alpha^2c_1^2\tau^2\Gamma^2(\epsilon,\epsilon_1)\Bigg\}\\
&+\frac{9d_2^2}{NK}\alpha^2+36L_2^2\alpha^4+4\alpha^3L_1G^2+2\alpha^3L_2G+\left(\frac{4\alpha}{\xi_2}+2\alpha^3L_1\right)\mathbb{E}_{t-2\tau}[\Delta_{t-\tau}]\\
&+\alpha\underbrace{\left(\frac{4}{\xi_3}+3\xi_1(c_1+c_2)+\frac{4}{\xi_2}+\alpha^2\left(24(c_1+c_2)^2+16\right)\right)}_{\mathcal{E}_3}\mathbb{E}_{t-2\tau}[\Delta_t]\\
&+2\alpha B(\epsilon,\epsilon_1)G+4\alpha^2B^2(\epsilon,\epsilon_1)+24\alpha^2(c_1+c_2)^2\Gamma^2(\epsilon,\epsilon_1)\\
&+3\alpha\xi_1(c_1+c_2)\Gamma^2(\epsilon,\epsilon_1)+2\alpha^3L_1\Gamma(\epsilon,\epsilon_1)G \qquad (70)
\end{aligned}
$$

If we choose step-size α such that $\alpha\mathcal{E}_2=\alpha\left(\frac{c_1+c_2}{\xi_1}+\frac{1}{\alpha}+24\xi_2+\frac{4}{\xi_2}\right)\le 2$, $\xi_1=\xi_2=\xi_3$, $\mathcal{E}_1=\alpha\xi_3+\alpha(3\xi_1(c_1+c_2)+24\xi_2)+24\alpha^2(c_1+c_2)^2\le 28\alpha\xi_1(c_1+c_2)+24\alpha^2(c_1+c_2)^2\le 30\alpha\xi_1(c_1+c_2)$ $(c_1,c_2>1)$ and $\mathcal{E}_3=\frac{4}{\xi_3}+3\xi_1(c_1+c_2)+\frac{4}{\xi_2}+\alpha^2\left(24(c_1+c_2)^2+16\right)\le(\frac{9}{\xi_1}+9\xi_1)(c_1+c_2)$, i.e.,

$$
\begin{aligned}
&\alpha\le\frac{1}{\left(\frac{c_1+c_2}{\xi_1}+24\xi_2+\frac{4}{\xi_2}\right)}=\frac{\xi_1}{(c_1+c_2+24\xi_1^2+4)}\\
&\alpha\le\min\{\frac{\xi_1}{12(c_1+c_2)},1,\frac{(\frac{5}{\xi_1}+5\xi_1)(c_1+c_2)}{24(c_1+c_2)^2+16}\},
\end{aligned}
$$

which is sufficient to hold when $\alpha\le\min\{\frac{\xi_1}{24(c_1+c_2)^2+24\xi_1^2+16},1\}$, then we have:

$$
\begin{aligned}
\mathbb{E}_{t-2\tau}\left\|\bar{\theta}_{t+1}-\theta^*\right\|^2 \le\ & \mathbb{E}_{t-2\tau}\left\|\bar{\theta}_t-\theta^*\right\|^2+2\alpha\mathbb{E}_{t-2\tau}\left\langle\bar{g}(\bar{\theta}_t),\bar{\theta}_t-\theta^*\right\rangle+4\alpha^2\mathbb{E}_{t-2\tau}\left\|\bar{g}(\bar{\theta}_t)\right\|^2\\
&+30\alpha\xi_1(c_1+c_2)\mathbb{E}_{t-2\tau}\left\|\bar{\theta}_t-\theta^*\right\|^2\\
&+2\Bigg\{8\alpha^2\tau^2c_4^2\mathbb{E}_{t-2\tau}\left[\left\|\bar{\theta}_t-\theta^*\right\|^2\right]+14\alpha^2\tau^2\frac{d_2^2}{NK}+\frac{52L_2^2\alpha^4\tau}{1-\rho^2}
\end{aligned}
$$

$$
\begin{aligned}
&+4\alpha^2c_4^2\tau\sum_{s=0}^{\tau}E_{t-2\tau}[\Delta_{t-s}]+3200\alpha^2c_1^2c_4^2\tau^3\Gamma^2(\epsilon,\epsilon_1)+4\alpha^2c_1^2\tau^2\Gamma^2(\epsilon,\epsilon_1)\Big\}\\
&+\frac{9d_2^2}{NK}\alpha^2+36L_2^2\alpha^4+4\alpha^3L_1G^2+2\alpha^3L_2G+\left(\frac{4\alpha}{\xi_2}+2\alpha^3L_1\right)\mathbb{E}_{t-2\tau}[\Delta_{t-\tau}]\\
&+\alpha\left(\frac{4}{\xi_3}+3\xi_1(c_1+c_2)+\frac{4}{\xi_2}+\alpha^2\left(24(c_1+c_2)^2+16\right)\right)\mathbb{E}_{t-2\tau}[\Delta_t]\\
&+2\alpha B(\epsilon,\epsilon_1)G+4\alpha^2B^2(\epsilon,\epsilon_1)+24\alpha^2(c_1+c_2)^2\Gamma^2(\epsilon,\epsilon_1)\\
&+3\alpha\xi_1(c_1+c_2)\Gamma^2(\epsilon,\epsilon_1)+2\alpha^3L_1\Gamma(\epsilon,\epsilon_1)G\\
&\leq\mathbb{E}_{t-2\tau}\left\|\bar{\theta}_t-\theta^*\right\|^2+2\alpha\mathbb{E}_{t-2\tau}\left\langle\bar{g}(\bar{\theta}_t),\bar{\theta}_t-\theta^*\right\rangle+4\alpha^2\mathbb{E}_{t-2\tau}\left\|\bar{g}(\bar{\theta}_t)\right\|^2\\
&+\left(30\alpha\xi_1(c_1+c_2)+16\alpha^2\tau^2c_4^2\right)\mathbb{E}_{t-2\tau}\left\|\bar{\theta}_t-\theta^*\right\|^2\\
&+\frac{9+28\tau^2}{NK}\alpha^2d_2^2+36\left(1+\frac{3\tau}{1-\rho^2}\right)L_2^2\alpha^4+4\alpha^3L_1G^2+2\alpha^3L_2G\\
&+\left(\frac{4\alpha}{\xi_1}+2\alpha^3L_1\right)\mathbb{E}_{t-2\tau}[\Delta_{t-\tau}]+\alpha(\frac{9}{\xi_1}+9\xi_1)(c_1+c_2)\mathbb{E}_{t-2\tau}[\Delta_t]+8\alpha^2c_4^2\tau\sum_{s=0}^{\tau}E_{t-2\tau}[\Delta_{t-s}]\\
&+2\alpha B(\epsilon,\epsilon_1)G+4\alpha^2B^2(\epsilon,\epsilon_1)+24\alpha^2(c_1+c_2)^2\Gamma^2(\epsilon,\epsilon_1)\\
&+3\alpha\xi_1(c_1+c_2)\Gamma^2(\epsilon,\epsilon_1)+2\alpha^3L_1\Gamma(\epsilon,\epsilon_1)G\\
&+6400\alpha^2c_1^2c_4^2\tau^3\Gamma^2(\epsilon,\epsilon_1)+8\alpha^2c_1^2\tau^2\Gamma^2(\epsilon,\epsilon_1) \qquad (71)
\end{aligned}
$$

if we choose the step-size α such that the high order $O(\alpha^2)$ terms are dominanted by the first order terms $O(\alpha)$, i.e., $4\alpha^2B^2(\epsilon,\epsilon_1)+24\alpha^2(c_1+c_2)^2\Gamma^2(\epsilon,\epsilon_1)+2\alpha^3L_1\Gamma(\epsilon,\epsilon_1)G+6400\alpha^2c_1^2c_4^2\tau^3\Gamma^2(\epsilon,\epsilon_1)+8\alpha^2c_1^2\tau^2\Gamma^2(\epsilon,\epsilon_1)\leq 2\alpha B(\epsilon,\epsilon_1)G+3\alpha\xi_1(c_1+c_2)\Gamma^2(\epsilon,\epsilon_1)$, i.e.,

$$
\alpha\leq\min\{\frac{2B(\epsilon,\epsilon_1)G+3\xi_1(c_1+c_2)\Gamma^2(\epsilon,\epsilon_1)}{4B^2(\epsilon,\epsilon_1)+24(c_1+c_2)^2\Gamma^2(\epsilon,\epsilon_1)+2L_1\Gamma(\epsilon,\epsilon_1)G+6400c_1^2c_4^2\tau^3\Gamma^2(\epsilon,\epsilon_1)+8c_1^2\tau^2\Gamma^2(\epsilon,\epsilon_1)},1\}
$$

we have:

$$
\begin{aligned}
\mathbb{E}_{t-2\tau}\left\|\bar{\theta}_{t+1}-\theta^*\right\|^2\leq&\,\mathbb{E}_{t-2\tau}\left\|\bar{\theta}_t-\theta^*\right\|^2+2\alpha\mathbb{E}_{t-2\tau}\left\langle\bar{g}(\bar{\theta}_t),\bar{\theta}_t-\theta^*\right\rangle+4\alpha^2\mathbb{E}_{t-2\tau}\left\|\bar{g}(\bar{\theta}_t)\right\|^2\\
&+\left(30\alpha\xi_1(c_1+c_2)+16\alpha^2\tau^2c_4^2\right)\mathbb{E}_{t-2\tau}\left\|\bar{\theta}_t-\theta^*\right\|^2\\
&+\frac{9+28\tau^2}{NK}\alpha^2d_2^2+36\left(1+\frac{3\tau}{1-\rho^2}\right)L_2^2\alpha^4+4\alpha^3L_1G^2+2\alpha^3L_2G\\
&+\left(\frac{4\alpha}{\xi_1}+2\alpha^3L_1\right)\mathbb{E}_{t-2\tau}[\Delta_{t-\tau}]+\alpha(\frac{9}{\xi_1}+9\xi_1)(c_1+c_2)\mathbb{E}_{t-2\tau}[\Delta_t]+8\alpha^2c_4^2\tau\sum_{s=0}^{\tau}E_{t-2\tau}[\Delta_{t-s}]\\
&+4\alpha B(\epsilon,\epsilon_1)G+6\alpha\xi_1(c_1+c_2)\Gamma^2(\epsilon,\epsilon_1) \qquad (72)
\end{aligned}
$$

With Lemma (15), we have the upper bound of $\mathbb{E}_{t-2\tau}[\Delta_t]$, $\mathbb{E}_{t-2\tau}[\Delta_{t-\tau}]$ and $\tau\sum_{s=0}^{\tau}E_{t-2\tau}[\Delta_{t-s}]$. Then we have:

$$
\begin{aligned}
\mathbb{E}_{t-2\tau}\left\|\bar{\theta}_{t+1}-\theta^*\right\|^2\leq&\,\mathbb{E}_{t-2\tau}\left\|\bar{\theta}_t-\theta^*\right\|^2+2\alpha\mathbb{E}_{t-2\tau}\left\langle\bar{g}(\bar{\theta}_t),\bar{\theta}_t-\theta^*\right\rangle+4\alpha^2\mathbb{E}_{t-2\tau}\left\|\bar{g}(\bar{\theta}_t)\right\|^2\\
&+\underbrace{\left(30\alpha\xi_1(c_1+c_2)+16\alpha^2\tau^2c_4^2\right)}_{\mathcal{E}_4}\mathbb{E}_{t-2\tau}\left\|\bar{\theta}_t-\theta^*\right\|^2
\end{aligned}
$$

$$+\frac{9+28\tau^2}{NK}\alpha^2 d_2^2+\alpha^3\left(36L_2^2+\frac{108\tau}{1-\rho^2}L_2^2+4L_1G^2+2L_2G\right)$$
$$+\frac{4\alpha^2}{K\alpha_g^2}\underbrace{\left(\frac{4\alpha}{\xi_1}+2\alpha^3L_1+\alpha(\frac{9}{\xi_1}+9\xi_1)(c_1+c_2)+8\alpha^2c_4^2\tau^2\right)}_{\mathcal{E}_5}\left[c_3^2+\frac{2c_3L_2\rho}{1-\rho}+4c_1^2(K-1)H^2\right]$$
$$+4\alpha B(\epsilon,\epsilon_1)G+6\alpha\xi_1(c_1+c_2)\Gamma^2(\epsilon,\epsilon_1) \tag{73}$$

If we choose step-size such that $\mathcal{E}_4=30\alpha\xi_1(c_1+c_2)+16\alpha^2\tau^2c_4^2\leq 32\alpha\xi_1(c_1+c_2)$ and $\mathcal{E}_5=\frac{4\alpha}{\xi_1}+2\alpha^3L_1+\alpha(\frac{9}{\xi_1}+9\xi_1)(c_1+c_2)+8\alpha^2c_4^2\tau^2\leq\alpha(\frac{14}{\xi_1}+14\xi_1)(c_1+c_2)$, i.e.,

$$\alpha\leq\min\{\frac{\xi_1(c_1+c_2)}{8\tau^2c_4^2},1,\frac{(\frac{1}{\xi_1}+\xi_1)(c_1+c_2)}{2L_1+8c_4^2\tau^2}\},$$

which is sufficient to hold when $\alpha\leq\frac{\xi_1(c_1+c_2)}{2L_1+8\tau^2c_4^2}$, then we have:

$$\begin{aligned}E_{t-2\tau}\left\|\bar{\theta}_{t+1}-\theta^*\right\|^2\leq{}&\mathbb{E}_{t-2\tau}\left\|\bar{\theta}_t-\theta^*\right\|^2+2\alpha\mathbb{E}_{t-2\tau}\left\langle\bar{g}(\bar{\theta}_t),\bar{\theta}_t-\theta^*\right\rangle+4\alpha^2\mathbb{E}_{t-2\tau}\left\|\bar{g}(\bar{\theta}_t)\right\|^2\\&+32\alpha\xi_1(c_1+c_2)\mathbb{E}_{t-2\tau}\left\|\bar{\theta}_t-\theta^*\right\|^2\\&+\frac{9+28\tau^2}{NK}\alpha^2d_2^2+\alpha^3\left(36L_2^2+\frac{108\tau}{1-\rho^2}L_2^2+4L_1G^2+2L_2G\right)\\&+\frac{4\alpha^3}{K\alpha_g^2}(\frac{14}{\xi_1}+14\xi_1)(c_1+c_2)\left[c_3^2+\frac{2c_3L_2\rho}{1-\rho}+8c_1^2(K-1)H^2\right]\\&+4\alpha B(\epsilon,\epsilon_1)G+6\alpha\xi_1(c_1+c_2)\Gamma^2(\epsilon,\epsilon_1).\end{aligned} \tag{74}$$

□

J.2.6 Parameter Selection

With Lemma 16, we have:

$$\begin{aligned}\mathbb{E}_{t-2\tau}\left\|\bar{\theta}_{t+1}-\theta^*\right\|^2\leq{}&(1+32\alpha\xi_1(c_1+c_2))\,\mathbb{E}_{t-2\tau}\left\|\bar{\theta}_t-\theta^*\right\|^2+2\alpha\mathbb{E}_{t-2\tau}\left\langle\bar{g}(\bar{\theta}_t),\bar{\theta}_t-\theta^*\right\rangle+4\alpha^2\mathbb{E}_{t-2\tau}\left\|\bar{g}(\bar{\theta}_t)\right\|^2\\&+\frac{9+28\tau^2}{NK}\alpha^2d_2^2+\alpha^3\left(36L_2^2+\frac{108\tau}{1-\rho^2}L_2^2+4L_1G^2+2L_2G\right)\\&+\frac{4\alpha^3}{K\alpha_g^2}(\frac{14}{\xi_1}+14\xi_1)(c_1+c_2)\left[c_3^2+\frac{2c_3L_2\rho}{1-\rho}+8c_1^2(K-1)H^2\right]\\&+4\alpha B(\epsilon,\epsilon_1)G+6\alpha\xi_1(c_1+c_2)\Gamma^2(\epsilon,\epsilon_1).\end{aligned} \tag{75}$$

Proposition 4. *If α satisfies the requirement as Lemma 16, choose $\xi_1=\frac{(1-\gamma)\bar{\omega}}{32(c_1+c_2)}$ and $\tau=\lceil\frac{\tau^{\text{mix}}(\alpha_T^2)}{K}\rceil$, we have:*

$$\nu_1\mathbb{E}_{t-2\tau}\left\|V_{\bar{\theta}_t}-V_{\theta^*}\right\|_{\bar{D}}^2\leq(\frac{1}{\alpha}-\nu_1)\mathbb{E}_{t-2\tau}\left\|\bar{\theta}_t-\theta^*\right\|^2-\frac{1}{\alpha}\mathbb{E}_{t-2\tau}\left\|\bar{\theta}_{t+1}-\theta^*\right\|^2+\frac{9+28\tau^2}{NK}\alpha d_2^2$$

$$+\alpha^2\left(36L_2^2+\frac{108\tau}{1-\rho^2}L_2^2+4L_1G^2+2L_2G\right)$$
$$+\frac{\alpha^2c_6}{K}\left[c_3^2+\frac{2c_3L_2\rho}{1-\rho}+8c_1^2(K-1)H^2\right]+4B(\epsilon,\epsilon_1)G+\nu_1\Gamma^2(\epsilon,\epsilon_1) \tag{76}$$

where $\nu_1=\frac{\nu}{4}=\frac{(1-\gamma)\bar{\omega}}{4}$ *and* $c_6\triangleq\frac{4}{\alpha_g^2}(\frac{14}{\xi_1}+14\xi_1)(c_1+c_2)$.

Proof. Incorporating $\xi_1=\frac{(1-\gamma)\bar{\omega}}{32(c_1+c_2)}$, $c_6\triangleq\frac{4}{\alpha_g^2}(\frac{14}{\xi_1}+14\xi_1)(c_1+c_2)$ and $6\xi_1(c_1+c_2)\leq\nu_1$ into Eq (75), we have

$$\begin{aligned}
\mathbb{E}_{t-2\tau}\left\|\bar{\theta}_{t+1}-\theta^*\right\|^2 &\leq \mathbb{E}_{t-2\tau}\left\|\bar{\theta}_t-\theta^*\right\|^2+2\alpha\mathbb{E}_{t-2\tau}\left\langle\bar{g}(\bar{\theta}_t),\bar{\theta}_t-\theta^*\right\rangle+4\alpha^2\mathbb{E}_{t-2\tau}\left\|\bar{g}(\bar{\theta}_t)\right\|^2\\
&+\alpha(1-\gamma)\bar{\omega}\mathbb{E}_{t-2\tau}\left\|\bar{\theta}_t-\theta^*\right\|^2\\
&+\frac{9+28\tau^2}{NK}\alpha^2d_2^2+\alpha^3\left(36L_2^2+\frac{108\tau}{1-\rho^2}L_2^2+4L_1G^2+2L_2G\right)\\
&+\frac{\alpha^3c_6}{K}\left[c_3^2+\frac{2c_3L_2\rho}{1-\rho}+8c_1^2(K-1)H^2\right]\\
&+4\alpha B(\epsilon,\epsilon_1)G+\alpha\nu_1\Gamma^2(\epsilon,\epsilon_1)\\
&\overset{(a)}{\leq}\mathbb{E}_{t-2\tau}\left\|\bar{\theta}_t-\theta^*\right\|^2-2\alpha(1-\gamma)\bar{\omega}\mathbb{E}_{t-2\tau}\left\|\bar{\theta}_t-\theta^*\right\|^2+16\alpha^2\mathbb{E}_{t-2\tau}\left\|V_{\bar{\theta}_t}-V_{\theta^*}\right\|_{\bar{D}}^2\\
&+\alpha(1-\gamma)\bar{\omega}\mathbb{E}_{t-2\tau}\left\|\bar{\theta}_t-\theta^*\right\|^2\\
&+\frac{9+28\tau^2}{NK}\alpha^2d_2^2+\alpha^3\left(36L_2^2+\frac{108\tau}{1-\rho^2}L_2^2+4L_1G^2+2L_2G\right)\\
&+\frac{\alpha^3c_6}{K}\left[c_3^2+\frac{2c_3L_2\rho}{1-\rho}+8c_1^2(K-1)H^2\right]\\
&+4\alpha B(\epsilon,\epsilon_1)G+\alpha\nu_1\Gamma^2(\epsilon,\epsilon_1)\\
&=\mathbb{E}_{t-2\tau}\left\|\bar{\theta}_t-\theta^*\right\|^2-\frac{\alpha(1-\gamma)\bar{\omega}}{2}\mathbb{E}_{t-2\tau}\left\|\bar{\theta}_t-\theta^*\right\|^2-\frac{\alpha(1-\gamma)\bar{\omega}}{2}\mathbb{E}_{t-2\tau}\left\|\bar{\theta}_t-\theta^*\right\|^2+16\alpha^2\mathbb{E}_{t-2\tau}\left\|V_{\bar{\theta}_t}-V_{\theta^*}\right\|_{\bar{D}}^2\\
&+\frac{9+28\tau^2}{NK}\alpha^2d_2^2+\alpha^3\left(36L_2^2+\frac{108\tau}{1-\rho^2}L_2^2+4L_1G^2+2L_2G\right)\\
&+\frac{\alpha^3c_6}{K}\left[c_3^2+\frac{2c_3L_2\rho}{1-\rho}+8c_1^2(K-1)H^2\right]\\
&+4\alpha B(\epsilon,\epsilon_1)G+\alpha\nu_1\Gamma^2(\epsilon,\epsilon_1)\\
&\leq\mathbb{E}_{t-2\tau}\left\|\bar{\theta}_t-\theta^*\right\|^2-\frac{\alpha(1-\gamma)\bar{\omega}}{2}\mathbb{E}_{t-2\tau}\left\|\bar{\theta}_t-\theta^*\right\|^2-\frac{\alpha(1-\gamma)\bar{\omega}}{2}\mathbb{E}_{t-2\tau}\left\|V_{\bar{\theta}_t}-V_{\theta^*}\right\|_{\bar{D}}^2+16\alpha^2\mathbb{E}_{t-2\tau}\left\|V_{\bar{\theta}_t}-V_{\theta^*}\right\|_{\bar{D}}^2\\
&+\frac{9+28\tau^2}{NK}\alpha^2d_2^2+\alpha^3\left(36L_2^2+\frac{108\tau}{1-\rho^2}L_2^2+4L_1G^2+2L_2G\right)\\
&+\frac{\alpha^3c_6}{K}\left[c_3^2+\frac{2c_3L_2\rho}{1-\rho}+8c_1^2(K-1)H^2\right]\\
&+4\alpha B(\epsilon,\epsilon_1)G+\alpha\nu_1\Gamma^2(\epsilon,\epsilon_1)\\
&\overset{(b)}{\leq}\mathbb{E}_{t-2\tau}\left\|\bar{\theta}_t-\theta^*\right\|^2-\frac{\alpha(1-\gamma)\bar{\omega}}{2}\mathbb{E}_{t-2\tau}\left\|\bar{\theta}_t-\theta^*\right\|^2-\frac{\alpha(1-\gamma)\bar{\omega}}{4}\mathbb{E}_{t-2\tau}\left\|V_{\bar{\theta}_t}-V_{\theta^*}\right\|_{\bar{D}}^2\\
&+\frac{9+28\tau^2}{NK}\alpha^2d_2^2+\alpha^3\left(36L_2^2+\frac{108\tau}{1-\rho^2}L_2^2+4L_1G^2+2L_2G\right)
\end{aligned}$$

$$
\begin{aligned}
&+\frac{\alpha^3 c_6}{K}\left[c_3^2+\frac{2c_3L_2\rho}{1-\rho}+8c_1^2(K-1)H^2\right]\\
&+4\alpha B(\epsilon,\epsilon_1)G+\alpha\nu_1\Gamma^2(\epsilon,\epsilon_1)\\
&\leq(1-2\alpha\nu_1)\mathbb{E}_{t-2\tau}\left\|\bar{\theta}_t-\theta^*\right\|^2-\alpha\nu_1\mathbb{E}_{t-2\tau}\left\|V_{\bar{\theta}_t}-V_{\theta^*}\right\|_{\bar{D}}^2+\frac{9+28\tau^2}{NK}\alpha^2d_2^2\\
&+\alpha^3\left(36L_2^2+\frac{108\tau}{1-\rho^2}L_2^2+4L_1G^2+2L_2G\right)\\
&+\frac{\alpha^3 c_6}{K}\left[c_3^2+\frac{2c_3L_2\rho}{1-\rho}+8c_1^2(K-1)H^2\right]+4\alpha B(\epsilon,\epsilon_1)G+\alpha\nu_1\Gamma^2(\epsilon,\epsilon_1) \qquad (77)
\end{aligned}
$$

where (a) is due to Lemma 3 and the selection of parameter; (b) is due to $16\alpha^2\leq\frac{\alpha(1-\gamma)\bar{\omega}}{4}$. Rearranging the terms and using the fact $1-2\alpha\nu_1\leq1-\alpha\nu_1$, we have:

$$
\begin{aligned}
\alpha\nu_1\mathbb{E}_{t-2\tau}\left\|V_{\bar{\theta}_t}-V_{\theta^*}\right\|_{\bar{D}}^2\leq&(1-\alpha\nu_1)\mathbb{E}_{t-2\tau}\left\|\bar{\theta}_t-\theta^*\right\|^2-\mathbb{E}_{t-2\tau}\left\|\bar{\theta}_{t+1}-\theta^*\right\|^2+\frac{9+28\tau^2}{NK}\alpha^2d_2^2\\
&+\alpha^3\left(36L_2^2+\frac{108\tau}{1-\rho^2}L_2^2+4L_1G^2+2L_2G\right)\\
&+\frac{\alpha^3 c_6}{K}\left[c_3^2+\frac{2c_3L_2\rho}{1-\rho}+8c_1^2(K-1)H^2\right]+4\alpha B(\epsilon,\epsilon_1)G+\alpha\nu_1\Gamma^2(\epsilon,\epsilon_1) \qquad (78)
\end{aligned}
$$

Then we finish the proof by dividing α on both sides. □

With these Lemmas, we are now ready to prove Theorem 4.

J.3 Proof of Theorem 4.

Given a fixed local step-size $\alpha_l\leq\frac{1}{4\sqrt{2}c_1(K-1)}$, decreasing effective step-sizes $\alpha_t=\frac{8}{\nu(a+t+1)}=\frac{8}{(1-\gamma)\bar{\omega}(a+t+1)}$, decreasing global step-sizes $\alpha_g^{(t)}=\frac{\alpha_t}{K\alpha_l}$ and weights $w_t=(a+t)$, we have:

$$
\mathbb{E}\left\|V_{\tilde{\theta}_T}-V_{\theta_i^*}\right\|_{\bar{D}}^2\leq\tilde{\mathcal{O}}\left(\frac{\tau^2G^2}{K^2T^2}+\frac{c_{\text{quad}}(\tau)}{\nu^2NKT}+\frac{c_{\text{lin}}(\tau)}{\nu^4KT^2}+\frac{B(\epsilon,\epsilon_1)G}{\nu}+\Gamma^2(\epsilon,\epsilon_1)\right) \qquad (79)
$$

Proof. We take the step-size $\alpha_t=\frac{8}{\nu(a+t+1)}=\frac{2}{\nu_1(a+t+1)}$ for $a>0$. In addition, we define weights $w_t=(a+t)$ and define

$$
\tilde{\theta}_T=\frac{1}{W}\sum_{t=1}^{T}w_t\bar{\theta}_t
$$

where $W=\sum_{t=1}^{T}w_t\geq\frac{1}{2}T(a+T)$. By convexity of positive definite quadratic forms ($\lambda_{\min}(\Phi^T\bar{D}\Phi)\geq\bar{\omega}>0$), we have

$$
\begin{aligned}
\nu_1\mathbb{E}\left\|V_{\tilde{\theta}_T}-V_{\theta^*}\right\|_{\bar{D}}^2\leq&\frac{\nu_1}{W}\sum_{t=1}^{T}(a+t)\mathbb{E}\left\|V_{\bar{\theta}_t}-V_{\theta^*}\right\|_{\bar{D}}^2\\
\leq&\frac{\nu_1}{W}\sum_{t=1}^{2\tau-1}(a+t)\mathbb{E}\left\|V_{\bar{\theta}_t}-V_{\theta^*}\right\|_{\bar{D}}^2+\frac{\nu_1}{W}\sum_{t=2\tau}^{T}(a+t)\mathbb{E}\left\|V_{\bar{\theta}_t}-V_{\theta^*}\right\|_{\bar{D}}^2
\end{aligned}
$$

$$
\begin{aligned}
&\leq \nu_1 \frac{(2\tau-1)(a+2\tau-1)G^2}{W} + \frac{\nu_1}{W}\sum_{t=2\tau}^{T}(a+t)\mathbb{E}\left\|V_{\tilde{\theta}_t} - V_{\theta^*}\right\|_{\bar{D}}^2 \\
&\overset{(76)}{\leq} \nu_1 \frac{(2\tau-1)(a+2\tau-1)G^2}{W} + \frac{\nu_1(a+2\tau)(a+2\tau+1)G^2}{2W} \\
&+ \frac{1}{W}\sum_{t=2\tau}^{T}\left[\frac{(9+28\tau^2)d_2^2}{NK}(a+t)\alpha_t + (a+t)\alpha_t^2\left(36L_2^2 + \frac{108\tau}{1-\rho^2}L_2^2 + 4L_1G^2 + 2L_2G\right)\right] \\
&+ \frac{1}{W}\sum_{t=2\tau}^{T}\frac{(a+t)\alpha^2 c_6}{K}\left[c_3^2 + \frac{2c_3L_2\rho}{1-\rho} + 8c_1^2(K-1)H^2\right] \\
&+ \frac{1}{W}\sum_{t=2\tau}^{T}\left[4(a+t)B(\epsilon,\epsilon_1)G + (a+t)\nu_1\Gamma^2(\epsilon,\epsilon_1)\right] \qquad (80)
\end{aligned}
$$

where $\left\|V_{\tilde{\theta}_{2\tau}} - V_{\theta^*}\right\|_{\bar{D}}^2 \leq G^2$. We know that $\frac{1}{W}\sum_{t=2\tau}^{T}(a+t)\alpha_t^2 \leq \frac{1}{W}\sum_{t=1}^{T}(a+t)\frac{4}{\nu_1^2(a+t)^2} \leq \frac{8\log(a+T)}{\nu_1^2T^2}$ and that $\frac{1}{W}\sum_{t=2\tau}^{T}(a+t)\alpha_t \leq \frac{4}{\nu_1 T}$. Plugging in these inequalities into Eq (80), we have:

$$
\begin{aligned}
\nu_1\mathbb{E}\left\|V_{\tilde{\theta}} - V_{\theta^*}\right\|_{\bar{D}}^2 &\leq \frac{3\nu_1(a+2\tau)(a+2\tau+1)G^2}{2W} + \frac{4(9+28\tau^2)d_2^2}{\nu_1 NKT} \\
&+ \frac{8\log(a+T)}{\nu_1^2T^2}\left(36L_2^2 + \frac{108\tau}{1-\rho^2}L_2^2 + 4L_1G^2 + 2L_2G\right) \\
&+ \frac{8c_6\log(a+T)}{\nu_1^2T^2K}\left[c_3^2 + \frac{2c_3L_2\rho}{1-\rho} + 8c_1^2(K-1)H^2\right] \\
&+ 4B(\epsilon,\epsilon_1)G + \nu_1\Gamma^2(\epsilon,\epsilon_1) \\
&= \frac{3\nu_1(a+2\tau)(a+2\tau+1)G^2}{2W} + \frac{4(9+28\tau^2)d_2^2}{\nu_1 NKT} \\
&+ \frac{8\log(a+T)}{\nu_1^2T^2K}\underbrace{\left[K\left(36L_2^2 + \frac{108\tau}{1-\rho^2}L_2^2 + 4L_1G^2 + 2L_2G\right) + c_6\left(c_3^2 + \frac{2c_3L_2\rho}{1-\rho} + 8c_1^2(K-1)H^2\right)\right]}_{c_{\text{lin}}(\tau)} \\
&+ 4B(\epsilon,\epsilon_1)G + \nu_1\Gamma^2(\epsilon,\epsilon_1) \qquad (81)
\end{aligned}
$$

where $c_{\text{quad}}(\tau) = 4d_2^2(9+28\tau^2)$. Dividing ν_1 on the both sides, changing ν_1 into ν ($\nu = (1-\gamma)\bar{\omega}$) and noting that $c_6 = \frac{4}{\alpha_g^2}(\frac{14}{\xi_1} + 14\xi_1)(c_1+c_2) = \mathcal{O}(\frac{1}{\nu})$, we have:

$$
\mathbb{E}\left\|V_{\tilde{\theta}_T} - V_{\theta^*}\right\|_{\bar{D}}^2 \leq \tilde{\mathcal{O}}\left(\frac{\tau^2G^2}{K^2T^2} + \frac{c_{\text{quad}}(\tau)}{\nu^2NKT} + \frac{c_{\text{lin}}(\tau)}{\nu^4KT^2} + \frac{B(\epsilon,\epsilon_1)G}{\nu} + \Gamma^2(\epsilon,\epsilon_1)\right). \qquad (82)
$$

We finish the proof by using the inequality, $\mathbb{E}\left\|V_{\tilde{\theta}_T} - V_{\theta_i^*}\right\|_{\bar{D}}^2 \leq 2\mathbb{E}\left\|V_{\tilde{\theta}_T} - V_{\theta^*}\right\|_{\bar{D}}^2 + 2\mathbb{E}\left\|V_{\theta_i^*} - V_{\theta^*}\right\|_{\bar{D}}^2$ and combining with the third point in Theorem 1. □

K Additional Simulation Results

K.1 Simulation results for the I.I.D. setting

In this subsection, we provide numerical results for `FedTD`(0) under the i.i.d. sampling setting to verify the theoretical results of Theorem 2. In particular, the MDP $\mathcal{M}^{(1)}$ of the first agent is randomly generated with a state space of size $n = 100$. The remaining MDPs are perturbations of $\mathcal{M}^{(1)}$ with the heterogeneity levels $\epsilon = 0.1$ and $\epsilon_1 = 0.1$. The number of local steps is chosen as $K = 20$. We evaluate the convergence in terms of the running error $e_t = \|\bar{\theta}_t - \theta_1^*\|^2$. Each experiment is run 10 times. We plot the mean and standard deviation across the 10 runs in Figure 2.

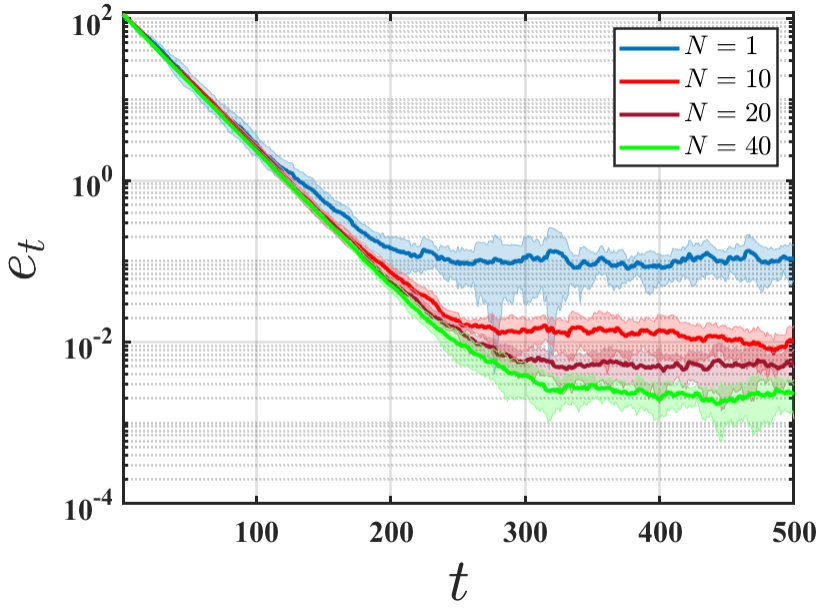

Figure 2: Performance of `FedTD`(0) with i.i.d. sampling with varying number of agents N. Solid lines denote the mean and shaded regions indicate the standard deviation over ten runs.

As shown in Fig 2, `FedTD`(0) converges for all choices of N. Larger values of N decreases the error, which is consistent with our theoretical analysis in Theorem 2.

K.2 Simulation results for the Markovian setting

In this subsection, we provide numerical results for `FedTD`(0) under the Markovian sampling setting to verify the theoretical results of Theorem 4. Here we generate all MDPs in the same way as the i.i.d setting and choose the number of local steps as $K = 20$. All the remaining parameters are kept the same as those in the subsection K.1.

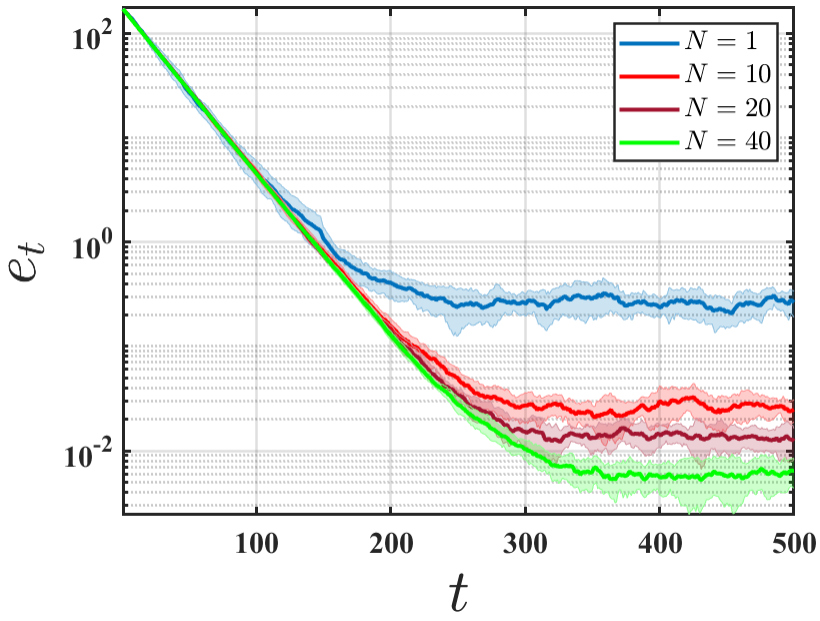

Figure 3: Performance of `FedTD`(0) with the Markovian sampling with varying number of agents N. Solid lines denote the mean and shaded regions indicate the standard deviation over ten runs.

As shown in Fig 3, `FedTD`(0) converges for all choices of N. Larger values of N decreases the error, which is consistent with our theoretical analysis in Theorem 4.