

# Federated TD Learning with Linear Function Approximation under Environmental Heterogeneity

**Han Wang**

*Department of Electrical Engineering  
Columbia University*

*hw2786@columbia.edu*

**Aritra Mitra**

*Department of Electrical and Computer Engineering  
North Carolina State University*

*amitra2@ncsu.edu*

**Hamed Hassani**

*Department of Electrical and Systems Engineering  
University of Pennsylvania*

*hassani@seas.upenn.edu*

**George J. Pappas**

*Department of Electrical and Systems Engineering  
University of Pennsylvania*

*pappasg@seas.upenn.edu*

**James Anderson**

*Department of Electrical Engineering  
Columbia University*

*james.anderson@columbia.edu*

**Reviewed on OpenReview:** <https://openreview.net/forum?id=hdQspgyFrk>

## 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 prove that in a low-heterogeneity regime, exchanging model estimates leads to linear convergence speedups in the number of agents. Our theoretical contribution is significant in that it is the first result of its kind in multi-agent/federated reinforcement learning that complements the numerous analogous results in heterogeneous federated optimization.

## 1 Introduction

In the popular federated learning (FL) paradigm (Konečný et al., 2016; McMahan et al., 2017), 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

Figure 1: (Left) Illustration of how FedTD works. Each agent performs  $K$  local TD update steps on its own MDP, and transmits its updated model to a server. The virtual MDP serves to approximate the dynamics of FedTD. The global model  $\theta_t$  at the server is used to construct a linearly parameterized approximation of the value function associated with a policy  $\pi$ . (Right) FedTD helps each agent converge quickly to a ball  $B(\theta^*; \epsilon)$  centered around the optimal parameter  $\theta^*$  of the virtual MDP. Here,  $\epsilon$  captures the heterogeneity in the agents' MDPs. Using the output  $\theta_T$  of FedTD, each agent can then re-tune based on its own data to converge exactly to its own optimal parameter  $\theta_i^*$ .

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 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 (Sahu et al., 2018; Khaled et al., 2019; 2020; Li et al., 2019; Koloskova et al., 2020; Woodworth et al., 2020b; Malinovskiy et al., 2020; Pathak & Wainwright, 2020; Wang et al., 2020; Karimireddy et al., 2020b; Acar et al., 2021; Gorbunov et al., 2021; Mitra et al., 2021; Mishchenko et al., 2022). 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 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 (Qi et al., 2021) describes a federated reinforcement learning (FRL) framework which incorporates some of the key ideas from FL into reinforcement learning (RL); applications of FRL in robotics (Liu et al., 2019), autonomous driving (Chen et al., 2015), and edge computing (Wang et al., 2019) 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 raw data (e.g., rewards, states, and actions) private, and adhere to stringent communication constraints.

**Motivation and Scope of this Work.** While FRL is a promising idea, in reality, it will rarely be the case that different agents end up interacting with exactly the same environment. Unfortunately, this is the running assumption in almost all multi-agent RL (MARL) and FRL works (Doan et al., 2019; Liu & Olshevsky, 2021a; Khodadadian et al., 2022; Shen et al., 2023). Departing from this somewhat unrealistic yet prevalent assumption, the main motivation of this paper is to build a systematic theoretical framework for reasoning about what to expect when one mixes information from non-identical Markov processes. The nature of this question is fundamental, and while we motivate its study from the perspective of FRL<sup>1</sup>, it can just as easily be connected to stochastic control and estimation problems where one seeks to fuse data generated from non-identical dynamical systems with noisy inputs (Wang et al., 2022b; Guo et al., 2023; Xin et al., 2023).

<sup>1</sup>Just as statistical heterogeneity is a major challenge in FL, environmental heterogeneity is identified as a key open challenge in FRL (Qi et al., 2021).

To initiate a principled study of heterogeneity in FRL, we focus on the simplest RL problem, namely policy evaluation. 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  $\pi$ . Notably, the value functions induced by  $\pi$  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? As we explain shortly, this is a non-trivial question to answer even for policy evaluation; hence, our focus on policy evaluation as a starting point. That said, recent works have shown that with minor modifications to the analysis of TD learning for policy evaluation (Srikant & Ying, 2019), one can analyze Q-learning for control (Chen et al., 2019). As such, we envision that the developments in this paper can be suitably extended to control algorithms like Q-learning as well.

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. The vehicles (agents) essentially have the same operations (e.g., steering, braking, accelerating, etc.), but can be exposed to different environments (e.g., road and weather conditions, routes, driving regulations etc.).

## 1.1 Our Contributions

We study a federated version of the temporal difference (TD) learning algorithm TD(0) (Sutton, 1988). 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  $\pi$ , 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 benefit from other agents' data in a communication-efficient manner, each agent periodically synchronizes with a central server, and performs multiple local model-updates in between - as depicted in Figure 1. 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:

1. 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 due to Kearns & Singh (2002). As such, Theorem 1 can serve as a tool of independent interest in RL.
2. 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 3.1, 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 appears to be new, and may be of independent interest.
3. Linear Speedup under Markovian Sampling and Heterogeneity. 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. In Theorem 2, we prove that after  $T$  communication rounds with  $K$  local model-updating steps per round, FedTD(0) guarantees convergence at a rate  $\mathcal{O}(1/NKT)$  to a neighborhood of each agent's optimal parameter. 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 linear speed-up in convergence via collaboration, and converge quickly to a vicinity of its own optimal parameter. One can view this as a coarse tuning phase. As is typically done in FL (Collins et al., 2022), each agent can use the solution of FedTD(0) to then re-tune (personalize) based on its own data. This is visually illustrated in Figure 1. Theorem 2 is significant in that it is the first result in FRL that complements the myriad of federated optimization results that account for the effects of heterogeneity (Sahu et al., 2018; Khaled et al., 2019; 2020; Li et al., 2019; Koloskova et al., 2020; Woodworth et al., 2020b; Malinovskiy et al., 2020; Pathak & Wainwright, 2020; Wang et al., 2020; Karimireddy et al., 2020b; Acar et al., 2021; Gorbunov et al., 2021; Mitra et al., 2021; Mishchenko et al., 2022).

4. **Novel Proof Framework.** One might be tempted to think that the proof of Theorem 2 is a simple combination of the standard FedAvg analysis with that of TD learning. We briefly explain here why this isn't quite the case, and defer a more elaborate explanation to Section 5.1. First, in the centralized TD analysis (Bhandari et al., 2018; Srikant & Ying, 2019), and in the existing analysis for MARL/FRL (Doan et al., 2019; Liu & Olshevsky, 2021a; Khodadadian et al., 2022) with identical MDPs, the dynamics of the update rules correspond to one single MDP. In our setup, the dynamics of FedTD(0) may not correspond to any MDP at all! Thus, we need new tools relative to existing RL analyses. Second, while existing FL analyses are essentially distributed optimization proofs, federated TD learning does not correspond to minimizing any fixed loss function. Moreover, unlike the i.i.d. data model in FL, the data tuples observed by each agent in FedTD(0) are part of a single Markovian trajectory. This creates complex time-correlations that are challenging to deal with even in a single-agent setting. Thus, we cannot directly employ FL proofs either. As such, we introduce a new analysis framework where we argue that the dynamics of FedTD(0) can be approximated by that of TD(0) on a virtual MDP, up to an error term that captures heterogeneity in the agents' MDPs. Carefully tracking how this error term propagates over time accounts for the effect of heterogeneity; establishing linear speedup under Markovian sampling and local steps requires much more work.
5. **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). This result sheds further light on the effect of heterogeneity in FRL.

## 1.2 Related Work

In what follows, we discuss the most relevant threads of literature.

1. **Finite-Time Analysis of TD Learning Algorithms.** In their seminal paper, Tsitsiklis & Van Roy (1997) provided an asymptotic convergence analysis of the temporal difference (TD) learning algorithm (Sutton, 1988; Sutton et al., 1998) with value function approximation, using tools from stochastic approximation theory. Several years later, the work by Korda & La (2015) provided finite-time rates for TD learning. However, the authors in Narayanan & Szepesvári (2017) noted some issues with the proofs in Korda & La (2015). Under the i.i.d. observation model described in Section 5, Dalal et al. (2018) and Lakshminarayanan & Szepesvári (2017) were able to resolve the issues in Korda & La (2015). Even so, a non-asymptotic convergence analysis for the challenging Markovian setting (that we consider in this paper) remained elusive till the work by Bhandari et al. (2018). While the authors in Bhandari et al. (2018) made some elegant connections between the dynamics of TD learning and gradient descent, an alternative proof technique using Stein's method was developed by Srikant & Ying (2019). Yet another interesting interpretation was provided by Liu & Olshevsky (2021b): they argued that the steady-state temporal difference direction acts as a "gradient-splitting" of an appropriately chosen function. Recently, a short proof of TD learning with linear function approximation and more general nonlinear contractive stochastic approximation schemes was provided by Mitra (2024) based on a novel inductive proof technique. While all the

above works provide upper-bounds for the task of policy evaluation, for minimax lower bounds, we refer the reader to the work of [Khamaru et al. \(2021\)](#).

2. Multi-Agent and Federated RL. In [Doan et al. \(2019\)](#) and [Liu & Olshevsky \(2021a\)](#), the authors analyze multi-agent TD learning with linear function approximation over peer-to-peer networks. Neither approach accounts for local steps or Markovian sampling. In [Shen et al. \(2023\)](#), the authors study a parallel version of asynchronous actor-critic algorithms, and establish a linear speedup result - albeit under an i.i.d. sampling assumption. Very recently, the authors in [Khodadadian et al. \(2022\)](#) and [Dal Fabbro et al. \(2023\)](#) studied the effect of Markovian sampling for federated TD learning. However, all of the above papers consider homogeneous setting with identical MDPs for all agents. In contrast, our work has to tackle the challenge of understanding the long-term effects of mixing TD update directions from non-identical MDPs. The only two papers we are aware of that perform any theoretical analysis of heterogeneity in FRL are [Jin et al. \(2022\)](#) and [Xie & Song \(2023\)](#). However, their analyses are limited to the much simpler tabular setting with no function approximation. In particular, the work of [Xie & Song \(2023\)](#) only comes with asymptotic results, i.e., they do not provide finite-time rates. Moreover, unlike us, neither [Jin et al. \(2022\)](#) nor [Xie & Song \(2023\)](#) provide any explicit linear speedup result. In conclusion, we are the first to establish a finite-time theory for FRL under function approximation, environmental heterogeneity, and Markovian sampling. Considering different settings, [Zhang et al. \(2024\)](#) proposed the FEDSARSA algorithm to solve the on-policy FRL problem and [Wang et al. \(2023\)](#) proposed FedLQR to solve the federated control design problem. A more detailed description of related work on federated learning is relegated to the Appendix.

## 2 Model and Problem Formulation

We consider a Markov Decision Process (MDP) ([Sutton et al., 1998](#)) defined by the tuple  $M = \langle S, A; R; P; \gamma \rangle$ , where  $S$  is a finite state space of size  $n$ ,  $A$  is a finite action space,  $P$  is a set of action-dependent Markov transition kernels,  $R$  is a reward function, and  $\gamma \in (0, 1)$  is the discount factor. We consider the problem of evaluating the value function  $V$  of a given policy  $\pi$ , where  $\pi : S \rightarrow A$ . The policy  $\pi$  induces a Markov reward process (MRP) characterized by a transition matrix  $P_\pi$ , and a reward function  $R_\pi$ . Under the action of the policy  $\pi$  at an initial state  $s$ ,  $P_\pi(s; s^0)$  is the probability of transitioning from state  $s$  to state  $s^0$ , and  $R_\pi(s)$  is the expected instantaneous reward. The discounted expected cumulative reward obtained by playing policy  $\pi$  starting from initial state  $s$  is:

$$V(s) = \mathbb{E} \sum_{t=0}^{\infty} \gamma^t R(s_t) | s_0 = s;$$

where  $s_t$  is the state of the Markov chain at time  $t$ . From [Tsitsiklis & Van Roy \(1997\)](#), we know that  $V$  is the fixed point of the policy-specific Bellman operator  $T : \mathbb{R}^n \rightarrow \mathbb{R}^n$ , i.e.,  $T V = V$ , where for any  $V \in \mathbb{R}^n$ ,

$$(T V)(s) = R(s) + \sum_{s^0 \in S} P_\pi(s; s^0) V(s^0); \quad \forall s \in S;$$

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  $V$  directly. To mitigate the curse of dimensionality, a common approach ([Sutton et al., 1998](#)) is to consider a low-dimensional linear function approximation of the value function  $V$ . 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  $\Phi$  is  $\phi_k$ . A parametric approximation  $\hat{V}$  of  $V$  in the span of  $\{\phi_k\}_{k=1}^d$  is then given by  $\hat{V} = \Phi \theta$ , 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  $\Phi$  by  $\phi(s) \in \mathbb{R}^d$ , and refer to it as the fixed feature vector corresponding to states. We write  $\hat{V}(s) = \phi(s)^\top \theta$  and make the standard assumption ([Bhandari et al., 2018](#)) that  $\|\phi(s)\|_2 \leq 1; \forall s \in S$ .

The objective is to find the best linear approximation of  $V$  in the span of  $\{\phi_k\}_{k=1}^d$ . More precisely, we seek a parameter vector  $\theta$  that minimizes the distance between  $\hat{V}$  and  $V$  (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  $\pi$ , a new data tuple  $O_t = (s_t; r_t = R(s_t); s_{t+1})$  comprising of the current state, the instantaneous reward, and the next state is observed. Let us define the TD(0) update direction as  $g_t(\theta_t) = r_t + \gamma V_\pi(s_{t+1}) - V_\pi(s_t)$ . Using a step-size  $\alpha \in (0, 1)$ , the parameter  $\theta_t$  is then updated as

$$\theta_{t+1} = \theta_t + \alpha g_t(\theta_t)$$

Under some mild technical assumptions, it was shown in Tsitsiklis & Van Roy (1997) that the TD(0) iterates converge asymptotically almost surely to a vector  $\theta^*$ , where  $\theta^*$  is the unique solution of the projected Bellman equation  $D T(\theta) = \theta$ . Here,  $D$  is a diagonal matrix with entries given by the elements of the stationary distribution  $\pi$  of the Markov matrix  $P$ . Furthermore,  $P_D(\cdot)$  is the projection operator onto the subspace spanned by  $\{g_{k=1}^d\}$  with respect to the inner product  $\langle \cdot, \cdot \rangle_D$ .<sup>2</sup>

**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? Existing FL analyses that study statistical heterogeneity in supervised learning/empirical risk minimization fall short of answering this question, since our problem does not involve minimizing a static loss function. As such, the question we have posed above is non-trivial, and requires several new ideas and tools. In the next section, we will start building these tools in a systematic manner by accomplishing the following goals.

- Goal 1. Formally defining what we mean by model heterogeneity in the agents' MDPs.
- Goal 2. Characterizing how such model heterogeneity translates to differences in the fixed points of the TD(0) algorithm when run on the agents' MDPs.
- Goal 3. Introducing the notion of a virtual MDP that will play a crucial role in reasoning about the long-term behavior of algorithms that combine information from non-identical MDPs.

### 3 Heterogeneous Federated RL

We consider a federated RL setting comprising of  $N$  agents that interact with potentially different environments. Agent  $i$ 's environment is characterized by the following MDP:  $M^{(i)} = (S; A; R^{(i)}; P^{(i)}; \gamma)$ . 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 that induces  $N$  Markov reward processes characterized by the tuple  $\{P^{(i)}; R^{(i)}\}_{i \in [N]}$ .<sup>3</sup> Agent  $i$  aims to find a linearly parameterized approximation of its own value function  $V^{(i)}$ . Trivially, agent  $i$  can do so without interacting with any other agent by simply running TD(0). However, the key question we ask pertains to the value of side-information: 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, we introduce the following definitions.

**Assumption 1. (Markov Kernel Heterogeneity)** There exists an  $\epsilon > 0$  such that for all agents  $i, j \in [N]$ , it holds that  $|P^{(i)}(s; s^0) - P^{(j)}(s; s^0)| \leq \epsilon$  for all  $s, s^0 \in S$ . Here, for each  $i \in [N]$ ,  $P^{(i)}(s; s^0)$  represents the  $(s; s^0)$ -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  $|R^{(i)}(s) - R^{(j)}(s)| \leq \epsilon_1$ .

Clearly, smaller values of  $\epsilon$  and  $\epsilon_1$  capture more similarity in the agents' MDPs. Suppose all agents can communicate via a central server. Via such communication, the standard FL task is to find one common

<sup>2</sup>We will use  $\|\cdot\|_D$  to denote the norm induced by the matrix  $D$ , and  $\|\cdot\|$  to represent the standard Euclidean norm for vectors and  $\|\cdot\|_2$  induced norm for matrices.

<sup>3</sup>We will henceforth drop the dependence of  $P^{(i)}$  and  $R^{(i)}$  on the policy  $\pi$ .

model that "ts" the data of all agents. In a similar spirit, our goal is to find a common parameter  $\hat{V}$  such that  $\hat{V} = \sum_{i=1}^n \alpha_i V^{(i)}$  approximates each  $V^{(i)}$ ;  $i \in [N]$ . The role of this common  $\hat{V}$  will be to quickly (i.e., by leveraging samples of all agents) provide a coarse model that the agents can then use as a warm-start to fine-tune based on personal data. There is a natural tension here. While federation can help converge faster to a coarse model, 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?

**Impact of Heterogeneity on TD fixed points.** 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. Since long-term dynamics are intimately tied to fixed points, we first set out to characterize the "closeness" in TD(0) fixed points across different MDPs. To proceed, we make the following standard assumption.

**Assumption 3.** For each  $i \in [N]$ , the Markov chain induced by the policy  $\pi_i$ , 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  $\hat{V}_i$  to denote the solution of the projected Bellman equation  $D^{(i)} T^{(i)}(\hat{V}_i) = \hat{V}_i$  for agent  $i$ . In words,  $\hat{V}_i$  is the best linear approximation of  $V^{(i)}$  in the span of  $\{ \mathbf{1} \}_{k=1}^d$ . From Section 2, we know that the iterates of TD(0) on agent  $i$ 's MRP will converge to  $\hat{V}_i$  asymptotically almost surely. Our goal is to bound the gap  $\| \hat{V}_i - \hat{V}_j \|_k$  as a function of the heterogeneity parameters  $\alpha_i$  and  $\beta_i$  appearing in Assumptions 1 and 2. The key observation we will exploit is that for each  $i \in [N]$ ,  $\hat{V}_i$  is the unique solution of the linear equation  $A_i \hat{V}_i = \mathbf{b}_i$ , where  $A_i = \sum_{s \in \mathcal{S}} D^{(i)}(P^{(i)} - I) \mathbf{1} \mathbf{1}^T$  and  $\mathbf{b}_i = \sum_{s \in \mathcal{S}} D^{(i)} R^{(i)}(s)$ . For an agent  $j \neq i$ , viewing  $A_j$  and  $\mathbf{b}_j$  as perturbed versions of  $A_i$  and  $\mathbf{b}_i$ , we can now appeal to results from the perturbation theory of linear equations (Horn & Johnson, 2012a, Chapter 5.8) to bound  $\| \hat{V}_i - \hat{V}_j \|_k$ . To that end, we first recall a result from the perturbation theory of Markov chains (O'cinneide, 1993) 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.** (Perturbation bound on Stationary Distributions) Suppose Assumption 1 holds. Then, for any pair of agents  $i, j \in [N]$ , the stationary distributions  $\pi^{(i)}$  and  $\pi^{(j)}$  satisfy:

$$\| \pi^{(i)} - \pi^{(j)} \|_1 \leq 2(n-1) \alpha_i + O(\alpha_i^2) \quad (1)$$

We will now use the above result to bound  $\| A_i - A_j \|_k$  and  $\| \mathbf{b}_i - \mathbf{b}_j \|_k$ : To state our results, we make the standard assumption that for each  $i \in [N]$ , it holds that  $\| jR^{(i)}(s) \| \leq R_{\max}$ ;  $s \in \mathcal{S}$ , i.e., the rewards are uniformly bounded. In (Tsitsiklis & Van Roy, 1997), it was shown that  $-A_i$  is a negative definite matrix; thus,  $\exists \alpha_i > 0$  such that  $\| A_i \|_k \leq \alpha_i$ ;  $i \in [N]$ . We also assume that  $\exists \beta_i > 0$  such that  $\| \mathbf{b}_i \|_k \leq \beta_i$ ;  $i \in [N]$ . In our first technical result, stated below, we provide a bound on the perturbation of TD fixed points.

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

$$1. \| A_i - A_j \|_k \leq \alpha_i \alpha_j + (1 + \alpha_j) 2(n-1) \alpha_i + O(\alpha_i^2) :$$

$$2. \| \mathbf{b}_i - \mathbf{b}_j \|_k \leq \beta_i \alpha_j + R_{\max} 2(n-1) \alpha_i + O(\alpha_i) :$$

3. Suppose  $\exists \theta > 0$  s.t.  $\| \mathbf{1} \|_k \leq \theta$ ,  $i \in [N]$ . Let  $\kappa(A_i)$  be the condition number of  $A_i$ . Then:

$$\| \hat{V}_i - \hat{V}_j \|_k \leq \alpha_j \alpha_i \left( \frac{\kappa(A_i) \theta}{1 - \frac{\kappa(A_i) \theta}{\alpha_i}} + \frac{\beta_j \alpha_i}{\alpha_i} + \frac{\beta_i \alpha_j}{\alpha_j} \right) :$$

**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  $\alpha_i$  and  $\beta_i$  - 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 (Kearns & Singh, 2002).<sup>4</sup>

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.

### 3.1 Virtual Markov Decision Process

In a standard FL setting, the goal is to typically minimize a global loss function  $f(x) = \frac{1}{N} \sum_{i \in [N]} f_i(x)$  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 (Charles & Konečný, 2020; Karimireddy et al., 2020b): the local iterates of each agent 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  $M = (S; A; R; P; \gamma)$ , where  $P = \frac{1}{N} \sum_{i=1}^N P^{(i)}$ ; and  $R = \frac{1}{N} \sum_{i=1}^N R^{(i)}$ . 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  $M$ , let the policy  $\pi$  that we seek to evaluate induce a virtual MRP characterized by the tuple  $(P; R; \gamma)$ . It is easy to see that  $P = \frac{1}{N} \sum_{i=1}^N P^{(i)}$ ; and  $R = \frac{1}{N} \sum_{i=1}^N R^{(i)}$ . 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.** (Convex combinations of Markov matrices) 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  $P$  is aperiodic and irreducible. Thus, there exists a unique stationary distribution  $\pi$  of this Markov chain; let  $D$  be the corresponding diagonal matrix. As before, let us define  $A = \gamma D(P - I)$ ,  $b = \gamma DR$ , and use  $\pi$  to denote the solution to the equation  $A\pi = b$ . Our next result is a consequence of Theorem 1, and characterizes the gap between  $\pi_i$  and  $\pi_j$ , for each  $i, j \in [N]$ .

**Proposition 2.** (Virtual MRP is "close" to Individual MRPs) Fix any  $i \in [N]$ . Using the same definitions as in Theorem 1, we have  $\|A_i - A\| \leq \gamma \max_{j \in [N]} \|P^{(j)} - P\|$  and  $\|b_i - b\| \leq \gamma \max_{j \in [N]} \|R^{(j)} - R\|$ .

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  $\pi$  of the virtual MRP. Proposition 2 is thus particularly important since it tells us that in a low-heterogeneity regime, by converging close to  $\pi$ , we also converge close to the optimal parameter  $\pi_i$  of each agent. This justifies studying the convergence behavior of FedTD(0) on the virtual MRP. Define  $\lambda = \gamma \max_{i \in [N]} \max_{j \in [N]} \|P^{(j)} - P^{(i)}\|$ . The smallest eigenvalue of this matrix will end up dictating the convergence rate of our proposed algorithm. We end this section with a result showing that this eigenvalue is bounded away from zero.

**Proposition 3.** For the virtual MRP, it holds that  $\max_{i \in [N]} \|\pi_i - \pi_j\| \leq \lambda$ , and  $\exists \epsilon > 0$  s.t.  $\min_{i \in [N]} \|\pi_i - \pi_j\| \geq \epsilon$ .

## 4 Federated TD Algorithm

In this section, we describe the FedTD(0) algorithm (outlined in Algorithm 1). The goal of FedTD(0) is to generate a model  $\hat{V}$  such that  $\hat{V}$  is a good approximation of each agent's value function  $V^{(i)}$ , corresponding

<sup>4</sup>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 FedTQ0

---

```

1: Input: Policy  $\pi$ , local step-size  $\eta$ , global step-size  $\eta_g^{(t)}$  that depends on communication round  $t$ 
2: Initialize:  $\theta_0 = \theta_0$  and  $s_{0;0}^{(i)} = s_0; \forall i \in [N]$ 
3: for each round  $t = 0; \dots; T - 1$  do
4:   for each agent  $i \in [N]$  do
5:     for  $k = 0; \dots; K - 1$  do with initial model  $\theta_{t;0}^{(i)} = \theta_t$ 
6:       Agent  $i$  plays  $(s_{t;k}^{(i)})$ , observes  $O_{t;k}^{(i)} = (S_{t;k}^{(i)}; r_{t;k}^{(i)}; s_{t;k+1}^{(i)})$ , and updates local model:
           
$$\theta_{t;k+1}^{(i)} = \theta_{t;k}^{(i)} + \eta g(\theta_{t;k}^{(i)}); \text{ where } g(\theta_{t;k}^{(i)}) = \mathbb{E}_{(s_{t;k+1}^{(i)})} \left[ \nabla_{\theta} P_{i \in [N]}(\theta_{t;k}^{(i)}; (s_{t;k+1}^{(i)})) \right]$$

7:     end for
8:     Agent  $i$  sends  $\theta_{t;K}^{(i)} - \theta_{t;0}^{(i)}$  back to the server
9:   end for
10:  Server broadcasts the following global model:  $\theta_{t+1} = \Pi_{\mathcal{H}} \left( \theta_t + \left( \eta_g^{(t)} = N \right) \sum_{i \in [N]} \theta_{t;K}^{(i)} \right)$ 
11: end for

```

---

to the policy  $\pi$ . In line with both standard FL algorithms, and also works in MARL/FRL (in homogeneous settings) (Doan et al., 2019; Khodadadian et al., 2022), the agents keep their raw observations (i.e., their rewards, states, and actions) private, and only exchange local models. In each round  $t$ , each agent  $i \in [N]$  starts from a common global model  $\theta_t$  and uses its local data to perform  $K$  local updates of the following form: at each local iteration  $k$ , agent  $i$  takes action  $(s_{t;k}^{(i)})$  and observes a data tuple  $O_{t;k}^{(i)}$  based on its own MRP, i.e.,  $f P^{(i)}; R^{(i)} g$ ; we note here that observations are independent across agents. Using its data tuple, agent  $i$  then updates its own local model  $\theta_{t;k}^{(i)}$  along the direction  $g(\theta_{t;k}^{(i)})$  in line 6. 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  $\theta_{t;K}^{(i)} - \theta_{t;0}^{(i)}$  to the server only once every  $K$  time-steps. The server averages these model differences and performs a projection to construct a global model  $\theta_{t+1}$  that is then broadcast to all agents (line 10). Here, we use  $\Pi_{\mathcal{H}}(\cdot)$  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  $\theta_0$ . Such a projection step ensures that the global models do not blow up, and is common in stochastic approximation (Borkar, 2009) and RL (Bhandari et al., 2018; Doan et al., 2019). Each agent then resumes its local updating process from this global model.

We note that the structure of FedTQ0 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. However, there are significant differences in the dynamics of standard FL algorithms and FedTQ0, making it quite challenging to derive finite-time convergence results for the latter. In the next section where we analyze FedTQ0, we will explain the nature of these challenges, and discuss how we overcome them.

## 5 Main Result and Analysis

To state our main convergence result for FedTQ0, we need to introduce a few objects. First, let  $H$  denote the radius of the set  $\mathcal{H}$  in line 10 of Algorithm 1. Also, define  $G = R_{\max} + 2H$  and  $\beta = (1 - \epsilon)!$ , where  $\epsilon$  is as in Proposition 3. In our analysis, we will make use of the geometric mixing property of finite-state, aperiodic, and irreducible Markov chains (Levin & Peres, 2017). Specifically, under Assumption 3, for each  $i \in [N]$ , there exists some  $m_i \geq 1$  and  $\delta_i \in (0, 1)$ , such that for all  $t \geq 0$  and  $0 \leq k \leq K - 1$ :

$$d_{TV} \left( P_{s_{t;k}^{(i)}} = \cdot \mid s_{0;0}^{(i)} = s_0; \cdot \right) \leq m_i \delta_i^{tK+k}; \forall s \in \mathcal{S}$$

Here, we use  $d_{TV}(P; Q)$  to denote the total-variation distance between two probability measures  $P$  and  $Q$ : For any  $\epsilon > 0$ , let us define the mixing time for  $P^{(i)}$  as  $\tau_i^{\text{mix}}(\epsilon) = \min\{t \geq 0 \mid \sum_{j=1}^n |m_j^{(i)}(t) - g_j| \leq \epsilon\}$ . Finally, let  $\tau(\epsilon) = \max_{i \in [N]} \tau_i^{\text{mix}}(\epsilon)$  represent the mixing time corresponding to the Markov chain that mixes the slowest. As one might expect, and as formalized by our main result below, it is this slowest-mixing Markov chain that dictates certain terms in the convergence rate of FedTD(0).

**Theorem 2. (Main Result)** There exists a decreasing global step-size sequence  $\eta_g^{(t)}$ , a fixed local step-size  $\eta_l$ , and a set of convex weights, such that a convex combination  $\bar{\pi}$  of the global models  $\pi_t$  satisfies the following for each agent  $i \in [N]$  after  $T$  rounds:

$$\mathbb{E} \|V_{i,T} - V_i^*\|_D^2 \leq \mathcal{O}\left(\frac{2G^2 + K^2}{K^2 T^2} + \frac{c_{\text{quad}}(\cdot)}{2NKT} + \frac{c_{\text{lin}}(\cdot)}{4KT^2} + Q(\cdot; \eta_l)\right); \quad (2)$$

where  $\eta_g^{(t)} = d \frac{\text{mix}(\frac{2}{t})}{K} e$ ,  $\tau = K \frac{1}{\eta_g^{(T)}}$ , and  $c_{\text{quad}}(\cdot)$  and  $c_{\text{lin}}(\cdot)$  are quadratic and linear functions in  $\tau$ , respectively. Moreover,  $B(\cdot; \eta_l) = H \frac{1}{n} + 2(n-1) + \mathcal{O}(\tau^2) + \mathcal{O}(\eta_l)$ ,  $(\cdot; \eta_l)$  is as defined in Theorem 1, and  $Q(\cdot; \eta_l) = \mathcal{O}\left(\frac{B(\cdot; \eta_l)G}{K} + \tau^2(\cdot; \eta_l)\right)$ .

The proof of the above result is deferred to Appendix I. We now discuss its implications.

**Discussion.** To parse Theorem 2, let us start by noting that the term  $Q(\cdot; \eta_l)$  in Eq. (2) captures the effect of heterogeneity; we will comment on this term later. When  $T \ll N$ , the dominant term among the first three terms in Eq. (2) is  $c_{\text{quad}}(\cdot) = \frac{2G^2 + K^2}{K^2 T^2}$ . 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  $\mathcal{O}\left(\frac{1}{KT}\right)$  (Bhandari et al., 2018). Our analysis thus reveals that by communicating just  $T$  times in  $KT$  iterations, each agent can achieve a linear speedup w.r.t. the number of agents  $n$  in a low-heterogeneity regime, i.e., when  $Q(\cdot; \eta_l)$  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 re-tune for personalization. Theorem 2 is significant in that it is the first result of its kind in MARL/FRL with heterogeneous environments, and complements the numerous analogous results in heterogeneous federated optimization (Sahu et al., 2018; Khaled et al., 2019; 2020; Li et al., 2019; Koloskova et al., 2020; Woodworth et al., 2020b; Malinovskiy et al., 2020; Pathak & Wainwright, 2020; Wang et al., 2020; Karimireddy et al., 2020b; Acar et al., 2021; Gorbunov et al., 2021; Mitra et al., 2021; Mishchenko et al., 2022).

When all the MDPs are identical,  $Q(\cdot; \eta_l) = 0$ . But when the MDPs are different, should we expect such a 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  $g_t^{(i)}$  of each agent  $i$  in Algorithm 1 by its steady-state deterministic version  $g_t^{(i)} = b_i - A_i g_{t,k}^{(i)}$ , where  $A_i$  and  $b_i$  are as in Section 3. We call the resulting deterministic algorithm mean-path FedTD(0). For simplicity, we skip the projection step. 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  $\eta = \frac{1}{t}$  be chosen such that  $I - \hat{A}$  is Schur stable, where  $\hat{A} = A_1 + A_2 = 2$ . Define  $e_{1,t}$ ,  $e_{2,t}$ ;  $i \in \{1, 2\}$ . The output of mean-path FedTD(0) then satisfies:

$$\lim_{t \uparrow \infty} e_{1,t} = \frac{1}{2} \hat{A}^{-1} A_2 (b_1 - b_2); \lim_{t \uparrow \infty} e_{2,t} = \frac{1}{2} \hat{A}^{-1} A_1 (b_2 - b_1); \quad (3)$$

**Discussion:** For the setting described in Theorem 3, the mean-path FedTD(0) updates follow the deterministic recursion  $e_{i,t+1} = (I - \hat{A}) e_{i,t} + \hat{b}_i$ , where  $\hat{b} = (1-2)(b_1 + b_2)$ . 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  $(I - \hat{A})$  is Schur stable, justifying the choice of  $\eta$  in Theorem 3. The main message conveyed by this result is that the gap between the limit point of mean-path FedTD(0) and the optimal parameter  $\pi_i$  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, i.e., the effect of the heterogeneity-induced bias cannot be eliminated by making  $\eta$  arbitrarily small. Aligning with this observation, notice that  $Q(\eta)$  in Eq. (2) is also step-size independent. The above discussion sheds some light on the fact that a term of the form  $Q(\eta)$  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 key difference with the standard FL setting where the effect of heterogeneity manifests itself only when the number of local steps  $K$  strictly exceeds 1 (Charles & Konečný, 2021; Karimireddy et al., 2020b; Mitra et al., 2021).

## 5.1 Main Technical Challenges and Overview of the Novel Ingredients in Our Analysis

**Challenges.** We summarize the major technical challenges that show up in the analysis of Theorem 2. First, the FedTD(0) update direction may not correspond to the TD(0) update direction of any MDP. This challenge is unique to our setting, and neither shows up in the centralized TD(0) analysis (Bhandari et al., 2018; Srikant & Ying, 2019), nor in the existing MARL/FRL analyses with homogeneous MDPs (Doan et al., 2019; Khodadadian et al., 2022). Second, unlike standard FL analyses that deal with i.i.d. observations for each agent, our setting is complicated by the fact that each agent's data is generated from a Markov chain. Moreover, for each agent  $i$ , the parameter sequence  $\{g_{t;k}^{(i)}\}$  and the data tuples  $\{O_{t;k}^{(i)}\}$  are intricately coupled. Third, 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. Fourth, controlling the gradient bias  $(1/NK) \sum_{i=1}^N \sum_{k=0}^{K-1} g_{t;k}^{(i)} \cdot O_{t;k}^{(i)} - g_{t;k}^{(i)}$  and the gradient norm  $E_k(1/NK) \sum_{i=1}^N \sum_{k=0}^{K-1} \|g_{t;k}^{(i)}\|^2$  requires a very delicate analysis when one seeks to establish the linear speedup property w.r.t. the number of agents  $N$ , i.e., the  $O(1/NKT)$ -type rate. In particular, naively bounding terms using the projection radius (as in the centralized analysis (Bhandari et al., 2018)) 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 the different random variables discussed above.

**Proof Sketch for Theorem 2.** Our first key innovation is to build on the results in Section 3 to show that the mean-path (steady-state) FedTD(0) update direction  $(1/N) \sum_{i=1}^N g_{t;k}^{(i)}$  is "close" to the mean-path TD(0) update direction  $g(\cdot) = b - A$  of the virtual MRP we constructed in Section 3.1; here,  $b, A$  are as defined in Section 3.1. Formally, we have the following result.

**Lemma 2.** (Steady-state Pseudo-Gradient Heterogeneity) For each  $\delta \in (0, 2H)$ , we have:

$$g_{t;k}(\cdot) - \frac{1}{N} \sum_{i=1}^N g_{t;k}^{(i)}(\cdot) \leq B(\delta); \quad (4)$$

where  $B(\delta)$  is as in Theorem 2, and  $g_{t;k}(\cdot)$  is the steady-state TD(0) direction of the virtual MRP.

From Bhandari et al. (2018), we know that  $g_{t;k}(\cdot)$  acts like a pseudo-gradient pointing towards the optimal model of the virtual MRP. Since based on Proposition 2, we know that  $\|g_{t;k}(\cdot) - g^*\| \leq \delta$ , Lemma 2 tells us that at least in the steady-state, the iterates of FedTD(0) will converge to a neighborhood of each agent's optimal model, where the size of the neighborhood depends on the level of heterogeneity. While this helps build intuition, all the valuable insights conveyed by Lemma 2 only pertain to the steady state dynamics of FedTD(0), i.e., all the statistical challenges we alluded to still need to be resolved. In particular, as mentioned earlier, we cannot naively use a projection bound of the form  $E_k(1/NK) \sum_{i=1}^N \sum_{k=0}^{K-1} \|g_{t;k}^{(i)}\|^2 = O(G^2)$  from the centralized analysis in Bhandari et al. (2018), since the local models may not belong to the set  $\mathcal{H}$ . Also, this will obscure the linear speedup effect. We overcome this difficulty by decomposing the random TD direction of each agent  $i$  as  $g_{t;k}^{(i)} = b_i(O_{t;k}^{(i)}) - A_i(O_{t;k}^{(i)})$ . 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 2. The detailed analysis, along with some simulations, are deferred to the Appendix.

(a) Simulations on the effect of the linear speedup      (b) Simulations on the effect of the heterogeneity level

Figure 2: Performance of FedTD(0) under Markovian sampling. (a) Performance of FedTD(0) for varying number of agents  $N$ . The MDP  $M^{(1)}$  of the first agent is randomly generated with a state space of size  $n = 100$ . The remaining MDPs are perturbations of  $M^{(1)}$  with the heterogeneity levels  $\beta = 0.05$  and  $\beta_1 = 0.1$ . We evaluate the convergence in terms of the running error  $\|e_t\| = k_t^{-1} k^2$ . (b) Performance of FedTD(0) for varying heterogeneity level, with a fixed number of agents  $N = 20$ . Complying with theory, increasing  $N$  reduces the error, and increasing the level of heterogeneity increases the size of the ball to which FedTD(0) converges. We choose the number of local steps  $k = 10$  in both plots.

## 6 Conclusion

In this work, we have studied the problem of federated reinforcement learning under environmental heterogeneity and explored the following question: Can an agent expedite the process of learning its own value function by using information from agents interacting with potentially 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, the FedTD(0) can achieve a linear speedup under both i.i.d and 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 used in the traditional FL optimization literature can be applied to solve federated RL 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.

### Acknowledgments

Anderson and Wang are partially supported by the NSF under awards 2144634 & 2231350 from the EPCN program.

Hamed Hassani is supported by The Institute for Learning-enabled Optimization at Scale (TILOS), under award number NSF-CCF-2112665.

## References

- Durmus Alp Emre Acar, Yue Zhao, Ramon Matas Navarro, Matthew Mattina, Paul N Whatmough, and Venkatesh Saligrama. Federated learning based on dynamic regularization. *arXiv preprint arXiv:2111.04263*, 2021.
- Jalaj Bhandari, Daniel Russo, and Raghav Singal. A finite time analysis of temporal difference learning with linear function approximation. In *Conference on learning theory* pp. 1691–1692. PMLR, 2018.
- Vivek S Borkar. *Stochastic approximation: A dynamical systems viewpoint* volume 48. Springer, 2009.
- Zachary Charles and Jakub Konečný. On the outsized importance of learning rates in local update methods. *arXiv preprint arXiv:2007.00878*, 2020.
- Zachary Charles and Jakub Konečný. Convergence and Accuracy Trade-Offs in Federated Learning and Meta-Learning. In *International Conference on Artificial Intelligence and Statistics*, pp. 2575–2583. PMLR, 2021.
- Chenyi Chen, Ari Seif, Alain Kornhauser, and Jianxiong Xiao. Deepdriving: Learning a ordinance for direct perception in autonomous driving. In *Proceedings of the IEEE international conference on computer vision* pp. 2722–2730, 2015.
- Zaiwei Chen, Sheng Zhang, Thanh T Doan, Siva Theja Maguluri, and John-Paul Clarke. Performance of Q-learning with linear function approximation: Stability and finite-time analysis. *arXiv preprint arXiv:1905.11425* pp. 4, 2019.
- Liam Collins, Hamed Hassani, Aryan Mokhtari, and Sanjay Shakkottai. Exploiting shared representations for personalized federated learning. In *International Conference on Machine Learning*, pp. 2089–2099. PMLR, 2021.
- Liam Collins, Hamed Hassani, Aryan Mokhtari, and Sanjay Shakkottai. FedAvg with fine tuning: Local updates lead to representation learning. *arXiv preprint arXiv:2205.13692*, 2022.
- Nicolò Dal Fabbro, Aritra Mitra, and George J Pappas. Federated TD learning over finite-rate erasure channels: Linear speedup under markovian sampling. *IEEE Control Systems Letters*, 2023.
- Gal Dalal, Balázs Szörényi, Gagan Thoppe, and Shie Mannor. Finite sample analyses for TD(0) with function approximation. In *Proceedings of the AAAI Conference on Artificial Intelligence*, volume 32, 2018.
- Yuyang Deng, Mohammad Mahdi Kamani, and Mehrdad Mahdavi. Adaptive personalized federated learning. *arXiv preprint arXiv:2003.13461*, 2020.
- Thanh Doan, Siva Maguluri, and Justin Romberg. Finite-time analysis of distributed TD(0) with linear function approximation on multi-agent reinforcement learning. In *International Conference on Machine Learning*, pp. 1626–1635. PMLR, 2019.
- Alireza Fallah, Aryan Mokhtari, and Asuman Ozdaglar. Personalized federated learning: A meta-learning approach. *arXiv preprint arXiv:2002.07948*, 2020.
- Georg Frobenius, Ferdinand Georg Frobenius, Ferdinand Georg Frobenius, Ferdinand Georg Frobenius, and Germany Mathematician. Über matrizen aus nicht negativen elementen. 1912.
- Avishek Ghosh, Jichan Chung, Dong Yin, and Kannan Ramchandran. An efficient framework for clustered federated learning. *Advances in Neural Information Processing Systems* 33:19586–19597, 2020.
- Eduard Gorbunov, Filip Hanzely, and Peter Richtárik. Local SGD: Unified theory and new efficient methods. In *International Conference on Artificial Intelligence and Statistics*, pp. 3556–3564. PMLR, 2021.



- Taosha Guo, Abed AlRahman Al Makdah, Vishaal Krishnan, and Fabio Pasqualetti. Imitation and transfer learning for LQG control. arXiv preprint arXiv:2303.09002, 2023.
- Farzin Haddadpour and Mehrdad Mahdavi. On the convergence of local descent methods in federated learning. arXiv preprint arXiv:1910.14425, 2019.
- Farzin Haddadpour, Mohammad Mahdi Kamani, Mehrdad Mahdavi, and Viveck Cadambe. Local SGD with periodic averaging: Tighter analysis and adaptive synchronization. In *Advances in Neural Information Processing Systems*, pp. 11082–11094, 2019.
- Filip Hanzely, Slavomír Hanzely, Samuel Horváth, and Peter Richtárik. Lower bounds and optimal algorithms for personalized federated learning. *Advances in Neural Information Processing Systems* 33:2304–2315, 2020.
- Roger A Horn and Charles R Johnson. *Matrix analysis*. Cambridge university press, 2012a.
- Roger A Horn and Charles R Johnson. *Matrix analysis*. Cambridge university press, 2012b.
- Hao Jin, Yang Peng, Wenhao Yang, Shusen Wang, and Zhihua Zhang. Federated Reinforcement Learning with Environment Heterogeneity. In *International Conference on Artificial Intelligence and Statistics*, pp. 18–37. PMLR, 2022.
- Sai Praneeth Karimireddy, Satyen Kale, Mehryar Mohri, Sashank Reddi, Sebastian Stich, and Ananda Theertha Suresh. Scalold: Stochastic controlled averaging for federated learning. In *International Conference on Machine Learning* pp. 5132–5143. PMLR, 2020a.
- Sai Praneeth Karimireddy, Satyen Kale, Mehryar Mohri, Sashank Reddi, Sebastian Stich, and Ananda Theertha Suresh. Scalold: Stochastic controlled averaging for federated learning. In *International Conference on Machine Learning* pp. 5132–5143. PMLR, 2020b.
- Michael Kearns and Satinder Singh. Near-optimal reinforcement learning in polynomial time. *Machine learning*, 49(2):209–232, 2002.
- Ahmed Khaled, Konstantin Mishchenko, and Peter Richtárik. First analysis of local GD on heterogeneous data. arXiv preprint arXiv:1909.04715, 2019.
- Ahmed Khaled, Konstantin Mishchenko, and Peter Richtárik. Tighter theory for local SGD on identical and heterogeneous data. In *International Conference on Artificial Intelligence and Statistics*, pp. 4519–4529. PMLR, 2020.
- Koulik Khamaru, Ashwin Pananjady, Feng Ruan, Martin J Wainwright, and Michael I Jordan. Is temporal difference learning optimal? An instance-dependent analysis. *SIAM Journal on Mathematics of Data Science* 3(4):1013–1040, 2021.
- Sajad Khodadadian, Pranay Sharma, Gauri Joshi, and Siva Theja Maguluri. Federated Reinforcement Learning: Linear Speedup Under Markovian Sampling. In *International Conference on Machine Learning*, pp. 10997–11057. PMLR, 2022.
- Anastasia Koloskova, Nicolas Loizou, Sadra Boreiri, Martin Jaggi, and Sebastian U Stich. A unified theory of decentralized SGD with changing topology and local updates. arXiv preprint arXiv:2003.10422, 2020.
- Jakub Konečný, H Brendan McMahan, Daniel Ramage, and Peter Richtárik. Federated optimization: Distributed machine learning for on-device intelligence. arXiv preprint arXiv:1610.02527, 2016.
- Nathaniel Korda and Prashanth La. On TD(0) with function approximation: Concentration bounds and a centered variant with exponential convergence. In *International conference on machine learning* pp. 626–634. PMLR, 2015.
- Yassine Laguel, Krishna Pillutla, Jérôme Malick, and Zaid Harchaoui. A superquantile approach to federated learning with heterogeneous devices. In *2021 55th Annual Conference on Information Sciences and Systems (CISS)*, pp. 1–6. IEEE, 2021.

- Chandrashekar Lakshminarayanan and Csaba Szepesvári. Linear stochastic approximation: Constant step-size and iterate averaging. arXiv preprint arXiv:1709.04073, 2017.
- David A Levin and Yuval Peres. Markov chains and mixing times, volume 107. American Mathematical Soc., 2017.
- Xiang Li, Kaixuan Huang, Wenhao Yang, Shusen Wang, and Zhihua Zhang. On the convergence of fedavg on non-iid data. arXiv preprint arXiv:1907.02189, 2019.
- Boyi Liu, Lujia Wang, and Ming Liu. Lifelong federated reinforcement learning: a learning architecture for navigation in cloud robotic systems. IEEE Robotics and Automation Letters, 4(4):4555–4562, 2019.
- Rui Liu and Alex Olshevsky. Distributed TD(0) with almost no communication. arXiv preprint arXiv:2104.07855, 2021a.
- Rui Liu and Alex Olshevsky. Temporal difference learning as gradient splitting. In International Conference on Machine Learning, pp. 6905–6913. PMLR, 2021b.
- Grigory Malinovskiy, Dmitry Kovalev, Elnur Gasanov, Laurent Condat, and Peter Richtárik. From Local SGD to Local Fixed-Point Methods for Federated Learning. In International Conference on Machine Learning, pp. 6692–6701. PMLR, 2020.
- Brendan McMahan, Eider Moore, Daniel Ramage, Seth Hampson, and Blaise Agüera y Arcas. Communication-efficient learning of deep networks from decentralized data. In Artificial Intelligence and Statistics, pp. 1273–1282. PMLR, 2017.
- Konstantin Mishchenko, Grigory Malinovsky, Sebastian Stich, and Peter Richtárik. ProxSkip: Yes! Local Gradient Steps Provably Lead to Communication Acceleration! Finally! arXiv preprint arXiv:2202.09357, 2022.
- Aritra Mitra. A simple finite-time analysis of TD learning with linear function approximation. arXiv preprint arXiv:2403.02476, 2024.
- Aritra Mitra, Rayana Jaafar, George J Pappas, and Hamed Hassani. Linear convergence in federated learning: Tackling client heterogeneity and sparse gradients. Advances in Neural Information Processing Systems 34: 14606–14619, 2021.
- C Narayanan and Csaba Szepesvári. Finite time bounds for temporal difference learning with function approximation: Problems with some state-of-the-art results. Technical report, Technical report, 2017.
- Colm Art O’cinneide. Entrywise perturbation theory and error analysis for markov chains. Numerische Mathematik, 65(1):109–120, 1993.
- Reese Pathak and Martin J Wainwright. FedSplit: An algorithmic framework for fast federated optimization. arXiv preprint arXiv:2005.05238, 2020.
- Hossein Pishro-Nik. Introduction to probability, statistics, and random processes. 2016.
- Jiaju Qi, Qihao Zhou, Lei Lei, and Kan Zheng. Federated reinforcement learning: techniques, applications, and open challenges. arXiv preprint arXiv:2108.11887, 2021.
- Amirhossein Reiszadeh, Aryan Mokhtari, Hamed Hassani, Ali Jadbabaie, and Ramtin Pedarsani. Fedpaq: A communication-efficient federated learning method with periodic averaging and quantization. In International Conference on Artificial Intelligence and Statistics, pp. 2021–2031. PMLR, 2020.
- Anit Kumar Sahu, Tian Li, Maziar Sanjabi, Manzil Zaheer, Ameet Talwalkar, and Virginia Smith. On the convergence of federated optimization in heterogeneous networks. arXiv preprint arXiv:1812.06127, 3, 2018.

- Felix Sattler, Klaus-Robert Müller, and Wojciech 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.
- Han Shen, Kaiqing Zhang, Mingyi Hong, and Tianyi Chen. Towards understanding asynchronous advantage actor-critic: Convergence and linear speedup. *IEEE Transactions on Signal Processing* 2023.
- Artin Spiridonov, Alex Olshevsky, and Ioannis Ch Paschalidis. Local SGD With a Communication Overhead Depending Only on the Number of Workers. *arXiv preprint arXiv:2006.02582*, 2020.
- Rayadurgam Srikant and Lei Ying. Finite-time error bounds for linear stochastic approximation and TD learning. In *Conference on Learning Theory* pp. 2803–2830. PMLR, 2019.
- Sebastian U Stich. Local SGD converges fast and communicates little. *arXiv preprint arXiv:1805.09767*, 2018.
- Lili Su, Jiaming Xu, and Pengkun Yang. Global convergence of federated learning for mixed regression. *arXiv preprint arXiv:2206.07279*, 2022.
- Richard S Sutton. Learning to predict by the methods of temporal differences. *Machine learning*, 3(1):9–44, 1988.
- Richard S Sutton, Andrew G Barto, et al. *Introduction to Reinforcement learning*. 1998.
- Canh T Dinh, Nguyen Tran, and Josh Nguyen. Personalized federated learning with moreau envelopes. *Advances in Neural Information Processing Systems* 33:21394–21405, 2020.
- Alysa Ziyang Tan, Han Yu, Lizhen Cui, and Qiang Yang. Towards personalized federated learning. *IEEE Transactions on Neural Networks and Learning Systems* 2022.
- John N Tsitsiklis and Benjamin Van Roy. An analysis of temporal-difference learning with function approximation. In *IEEE Transactions on Automatic Control*, 1997.
- Han Wang, Siddhartha Marella, and James Anderson. FedADMM: A federated primal-dual algorithm allowing partial participation. In *2022 IEEE 61st Conference on Decision and Control (CDC)* pp. 287–294. IEEE, 2022a.
- Han Wang, Leonardo F Toso, and James Anderson. FedSysID: A federated approach to sample-efficient system identification. *arXiv preprint arXiv:2211.14393*, 2022b.
- Han Wang, Leonardo F Toso, Aritra Mitra, and James Anderson. Model-free learning with heterogeneous dynamical systems: A federated LQR approach. *arXiv preprint arXiv:2308.11743*, 2023.
- Jianyu Wang and Gauri Joshi. Cooperative SGD: A unified framework for the design and analysis of communication-efficient SGD algorithms. *arXiv preprint arXiv:1808.07576*, 2018.
- Jianyu Wang, Qinghua Liu, Hao Liang, Gauri Joshi, and H Vincent Poor. Tackling the objective inconsistency problem in heterogeneous federated optimization. *Advances in Neural Information Processing Systems* 33, 2020.
- Xiaofei Wang, Yiwen Han, Chenyang Wang, Qiyang Zhao, Xu Chen, and Min Chen. In-Edge AI: Intelligentizing mobile edge computing, caching and communication by federated learning. *IEEE Network*, 33(5):156–165, 2019.
- Blake Woodworth, Kumar Kshitij Patel, Sebastian U Stich, Zhen Dai, Brian Bullins, H Brendan McMahan, Ohad Shamir, and Nathan Srebro. Is Local SGD Better than Minibatch SGD? *arXiv preprint arXiv:2002.07839*, 2020a.
- Blake E Woodworth, Kumar Kshitij Patel, and Nati Srebro. Minibatch vs local SGD for heterogeneous distributed learning. *Advances in Neural Information Processing Systems* 33:6281–6292, 2020b.

Zhijie Xie and Shenghui Song. Fedkl: Tackling data heterogeneity in federated reinforcement learning by penalizing kl divergence. *IEEE Journal on Selected Areas in Communications* 41(4):1227-1242, 2023.

Lei Xin, Lintao Ye, George Chiu, and Shreyas Sundaram. Learning dynamical systems by leveraging data from similar systems. *arXiv preprint arXiv:2302.04344*, 2023.

Chenyu Zhang, Han Wang, Aritra Mitra, and James Anderson. Finite-time analysis of on-policy heterogeneous federated reinforcement learning. *International Conference on Learning Representations* 2024.

## Contents

1	Introduction	1
1.1	Our Contributions	3
1.2	Related Work	4
2	Model and Problem Formulation	5
3	Heterogeneous Federated RL	6
3.1	Virtual Markov Decision Process	8
4	Federated TD Algorithm	8
5	Main Result and Analysis	9
5.1	Main Technical Challenges and Overview of the Novel Ingredients in Our Analysis	11
6	Conclusion	12
A	Additional Literature Survey	20
B	Perturbation bounds for TD(0) fixed points	21
B.1	Proof of Theorem 1	21
C	Properties of the Virtual Markov Decision Process	22
C.1	Proof of Proposition 1	22
C.2	Proof of Proposition 2	22
C.3	Proof of Proposition 3	22
D	Pseudo-gradient heterogeneity: Proof of Lemma 2	23
E	Auxiliary results used in the I.I.D. and Markovian settings	23
F	Notation	26
G	Warm-up: Analysis of FedTD under i.i.d. sampling	27
G.1	Auxiliary lemmas for Theorem 4	27
G.1.1	Variance reduction	27
G.1.2	Per Round Progress	28
G.1.3	Drift Term Analysis	30
G.1.4	Parameter Selection	32
G.2	Proof of Theorem 4	33



H Heterogeneity bias: Proof of Theorem 3	35
I Proof of the Markovian setting	36
I.1 Outline	36
I.2 Auxiliary lemmas for Theorem 2	36
I.2.1 Decomposition Form	36
I.2.2 Variance Reduction	38
I.2.3 Bounding $E \sum_{t=1}^T \sum_{i=1}^n \dots$	40
I.2.4 Drift Term Analysis	46
I.2.5 Per Round Progress	48
I.2.6 Parameter Selection	56
I.3 Proof of Theorem 2	58
J Simulation Results	60
J.1 Simulation results for the I.I.D. setting	60
J.2 Simulation results for the Markovian setting	61
J.3 Simulation on the effect of the heterogeneity level for the Markovian setting	62

## A Additional 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 [McMahan et al. \(2017\)](#). 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 ([Stich, 2018](#); [Wang & Joshi, 2018](#); [Spiridonov et al., 2020](#); [Reisizadeh et al., 2020](#); [Haddadpour et al., 2019](#); [Woodworth et al., 2020a](#)), and also in the more challenging heterogeneous setting ([Khaled et al., 2019](#); [2020](#); [Haddadpour & Mahdavi, 2019](#); [Li et al., 2019](#); [Koloskova et al., 2020](#)). In the latter scenario, it was soon realized that FedAvg suffers from a "client-drift" effect that hurts its convergence performance ([Charles & Konečný, 2020](#); [2021](#); [Karimireddy et al., 2020a](#)).

Since then, a lot of effort has gone into improving the convergence guarantees of FedAvg via a variety of technical approaches: proximal methods in FedProx ([Sahu et al., 2018](#)); operator-splitting in FedSplit ([Pathak & Wainwright, 2020](#)); variance-reduction in Scaffold ([Karimireddy et al., 2020a](#)) and S-Local-SVRG ([Gorbunov et al., 2021](#)); gradient-tracking in FedLin ([Mitra et al., 2021](#)); dynamic regularization in [Acar et al. \(2021\)](#); and ADMM in FedADMM ([Wang et al., 2022a](#)). 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, the authors in [Mishchenko et al. \(2022\)](#) introduced the ProxSkip algorithm, and showed that it can indeed lead to communication savings via local steps, despite arbitrary heterogeneity.

Some other approaches to tackling heterogeneous statistical distributions in FL include personalization ([Deng et al., 2020](#); [Fallah et al., 2020](#); [T Dinh et al., 2020](#); [Hanzely et al., 2020](#); [Tan et al., 2022](#)), clustering ([Ghosh et al., 2020](#); [Sattler et al., 2020](#); [Su et al., 2022](#)), representation learning ([Collins et al., 2021](#)), and the use of quantiles ([Laguel et al., 2021](#)).

## B Perturbation bounds for TD(0) fixed points

### B.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:

$$\begin{aligned}
 \|kA_i - A_j\| &= \|k > D^{(i)}(P^{(i)}) - > D^{(i)}(P^{(j)})\| \\
 &\leq \|k > D^{(i)}(P^{(i)}) - > D^{(i)}(P^{(j)})\| + \\
 &\quad \|> D^{(i)}(P^{(j)}) - > D^{(j)}(P^{(j)})\| \\
 &\leq \|k > D^{(i)}(P^{(i)}) - > D^{(i)}(P^{(j)})\| + \\
 &\quad \|k > D^{(i)}(P^{(j)}) - > D^{(j)}(P^{(j)})\| \\
 \text{(a)} \quad &\leq k^2 kD^{(i)}\|P^{(i)} - P^{(j)}\| + k^2 kD^{(i)}\|D^{(i)} - D^{(j)}\| \|P^{(j)}\| \\
 \text{(b)} \quad &\leq \frac{P_{\bar{n}}}{\bar{n}} + (1 + \epsilon)[2(n-1) + O(\epsilon^2)];
 \end{aligned} \tag{5}$$

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

$$\|kP^{(i)} - P^{(j)}\| \leq \frac{P_{\bar{n}}}{\bar{n}} \|kP^{(i)} - P^{(j)}\|_1 \leq \frac{P_{\bar{n}}}{\bar{n}} \|kP^{(i)} - P^{(j)}\|_1 = \frac{P_{\bar{n}}}{\bar{n}};$$

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

Next, we bound

$$\begin{aligned}
 \|kb_i - b_j\| &= \|kD^{(i)}R^{(i)} - D^{(j)}R^{(j)}\| \\
 &\leq \|kD^{(i)}R^{(i)} - D^{(i)}R^{(j)}\| + \|D^{(i)}R^{(j)} - D^{(j)}R^{(j)}\| \\
 &\leq k \|kD^{(i)} - D^{(i)}\| \|R^{(i)} - R^{(j)}\| + k \|kD^{(i)} - D^{(j)}\| \|R^{(j)}\| \\
 &\leq (1 + R_{\max}) [2(n-1) + O(\epsilon^2)];
 \end{aligned} \tag{6}$$

where we use Assumption 2 in the last inequality and follow the same reasoning as we used to bound  $\|kA_i - A_j\|$  above.

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

$$\frac{\|k_i - j_k\|}{\|k_i\|} \leq \frac{(A_i)}{1 - (A_i) \frac{kA_i - A_j}{kA_i}} \frac{\|kA_i - A_j\|}{\|kA_i\|} + \frac{\|kb_i - b_j\|}{\|kb_i\|}; \tag{7}$$

Here, we leveraged the perturbation theory of linear equations in (Horn & Johnson, 2012b) Section 5.8. Finally, for any  $\|k_i\| \leq H$ , we have

$$\|k_i - j_k\| \geq (\epsilon_1), \frac{(A_i)H}{1 - (A_i) \frac{A(\cdot)}{1}} \frac{A(\cdot)}{1} + \frac{b(\cdot; \epsilon_1)}{2};$$

where we used the fact that  $\epsilon_1$  and  $\epsilon_2$  are positive constants that lower bound  $\|kA_i\|$  and  $\|kb_i\|$ , respectively.

## C Properties of the Virtual Markov Decision Process

### C.1 Proof of Proposition 1

Before we prove this proposition, we present the following fact from (Pishro-Nik, 2016): 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.,  $P_{s,s^0}^k > 0$ ; for all  $s; s^0 \in S$ :

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  $J = \{i \in [N] \mid w_i > 0\}$ . 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  $k = \min_{i \in J} k_i$  and  $j = \arg \min_{i \in J} k_i$ , then we have:

$$\sum_{i \in [N]} w_i P^{(i)k} = \underbrace{w_j P^{(j)k}}_{\text{positive}} + \underbrace{\sum_{i \in [N] \setminus J} w_i P^{(i)k}}_{\text{nonnegative}}; \quad (8)$$

where each entry of  $w_j P^{(j)k}$  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.

### C.2 Proof of Proposition 2

Following similar arguments as in Theorem 1, we bound  $\|A_i - A_k\|$ :

$$\begin{aligned} \|A_i - A_k\| &= \|D^{(i)}(P^{(i)}) - D(P)\| \\ &\stackrel{(a)}{\leq} k^2 kD^{(i)}\|kP^{(i)} - P\| + k^2 kD^{(i)}\|D\| \|I - P\| \\ &\stackrel{(b)}{\leq} P \bar{n} + (1 + \dots) [2(n-1) + O(\epsilon^2)] = A(\epsilon); \end{aligned} \quad (9)$$

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

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

### C.3 Proof of Proposition 3

Since the virtual MDP is an average of the agents' MDPs, i.e.,  $P = \frac{1}{N} \sum_{i=1}^N P^{(i)}$ ; 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  $\max_{s \in S} (\lambda > D) = \max_{s \in S} (\lambda(s)) = \max_{s \in S} (\lambda(s)) > \max_{s \in S} (\lambda(s)) = 1$ :

To show that there exists  $\epsilon > 0$  such that  $\min_{s \in S} (\lambda > D) = \epsilon > 0$ , we will establish that  $\lambda > D$  is a positive-definite matrix. Since  $\lambda$  is full-column rank, this amounts to showing that  $D$  is a positive-definite matrix. From the definition of  $D$ , establishing positive-definiteness of  $D$  is equivalent to arguing that every element of the stationary distribution vector  $\lambda$  is strictly positive; here,  $\lambda > P = \lambda$ : To that end, from Proposition 1, we know that the Markov chain associated with  $P$  is aperiodic and irreducible. From the Perron-Frobenius theorem (Frobenius et al., 1912), we conclude that indeed every entry of  $\lambda$  is strictly positive. If we choose  $\epsilon = \min_{s \in S} \lambda(s) > 0$ ; we have  $\min_{s \in S} (\lambda > D) = \epsilon > 0$ :

## D Pseudo-gradient heterogeneity: Proof of Lemma 2

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

$$\begin{aligned}
g(\eta) - \frac{1}{N} \sum_{i=1}^N g_i(\eta) &= \left\langle \sum_{i=1}^N D(T^{(i)}), \eta - T^{(i)} \right\rangle - \frac{1}{N} \sum_{i=1}^N \left\langle D^{(i)}(T^{(i)}), \eta - T^{(i)} \right\rangle \\
&\stackrel{(a)}{\leq} \frac{1}{N} \sum_{i=1}^N \left\langle D(T^{(i)}), \eta - T^{(i)} \right\rangle - \frac{1}{N} \sum_{i=1}^N \left\langle D^{(i)}(T^{(i)}), \eta - T^{(i)} \right\rangle \\
&\stackrel{(b)}{\leq} \frac{1}{N} \sum_{i=1}^N \left\langle D \left( \frac{h}{N} \sum_{j=1}^N R^{(j)} + \frac{1}{N} \sum_{j=1}^N P^{(j)} \right), \eta - T^{(i)} \right\rangle - \frac{1}{N} \sum_{i=1}^N \left\langle D^{(i)}(T^{(i)}), \eta - T^{(i)} \right\rangle \\
&\quad + \frac{1}{N} \sum_{i=1}^N \left\langle D \left( \frac{h}{N} \sum_{j=1}^N R^{(j)} + \frac{1}{N} \sum_{j=1}^N P^{(j)} \right), T^{(i)} - T^{(i)} \right\rangle \\
&\stackrel{(c)}{\leq} \frac{1}{N} \sum_{i=1}^N \left\langle D \left( \frac{h}{N} \sum_{j=1}^N R^{(j)} + \frac{1}{N} \sum_{j=1}^N P^{(j)} \right), \eta - T^{(i)} \right\rangle \\
&\quad + \frac{1}{N} \sum_{i=1}^N \left\langle D(T^{(i)}), T^{(i)} - T^{(i)} \right\rangle \\
&\quad + \frac{1}{N} \sum_{i=1}^N \left\langle D \left( \frac{1}{N} \sum_{j=1}^N R^{(j)} - R^{(i)} + \frac{1}{N} \sum_{j=1}^N P^{(j)} - P^{(i)} \right), T^{(i)} \right\rangle \\
&\quad + \frac{1}{N} \sum_{i=1}^N \left\langle D \left( D^{(i)}(T^{(i)}) - D \left( \frac{h}{N} \sum_{j=1}^N R^{(j)} + \frac{1}{N} \sum_{j=1}^N P^{(j)} \right) \right), T^{(i)} \right\rangle \\
&\stackrel{(d)}{\leq} \frac{1}{N} \sum_{i=1}^N \left\langle \frac{1}{N} \sum_{j=1}^N R^{(j)} - R^{(i)} + \frac{1}{N} \sum_{j=1}^N P^{(j)} - P^{(i)}, T^{(i)} \right\rangle \\
&\quad + \frac{1}{N} \sum_{i=1}^N \left\langle D \left( D^{(i)}(T^{(i)}) - D \left( \frac{h}{N} \sum_{j=1}^N R^{(j)} + \frac{1}{N} \sum_{j=1}^N P^{(j)} \right) \right), T^{(i)} \right\rangle \\
&\stackrel{(e)}{\leq} \frac{h}{N} \sum_{i=1}^N \left\langle \frac{1}{N} \sum_{j=1}^N R^{(j)} - R^{(i)} + \frac{1}{N} \sum_{j=1}^N P^{(j)} - P^{(i)}, T^{(i)} \right\rangle \\
&\quad + \frac{h}{N} \sum_{i=1}^N \left\langle D \left( D^{(i)}(T^{(i)}) - D \left( \frac{h}{N} \sum_{j=1}^N R^{(j)} + \frac{1}{N} \sum_{j=1}^N P^{(j)} \right) \right), T^{(i)} \right\rangle \\
&= O\left(\frac{h}{N}\right) + \frac{h}{N} \sum_{i=1}^N \left( \frac{1}{N} \sum_{j=1}^N \left\langle R^{(j)} - R^{(i)}, T^{(i)} \right\rangle + \frac{1}{N} \sum_{j=1}^N \left\langle P^{(j)} - P^{(i)}, T^{(i)} \right\rangle \right) + O\left(\frac{h^2}{N}\right) = B\left(\frac{h}{N}; \eta\right); \tag{10}
\end{aligned}$$

Inequalities (a) and (c) follow from the triangle inequality, (b) is due to  $\|k\| \leq 1$ ; (d) is due to the fact that  $\|Dk\| \leq 1$ ; and (e) uses the following facts: (i)  $\|kR^{(i)} - R^{(i)}\| \leq \frac{1}{N}$ ; (ii)  $\|kP^{(i)} - P^{(i)}\| \leq \frac{1}{N}$ ; (iii)  $\|D^{(i)}(T^{(i)}) - D\left(\frac{h}{N} \sum_{j=1}^N R^{(j)} + \frac{1}{N} \sum_{j=1}^N P^{(j)}\right)\| \leq \frac{1}{N}$ , which, in turn, follows from the proof of Theorem 1 or Eq. (5); and (iv)  $\|k\| \leq H$  for any  $\eta \in \mathcal{H}$ :

## E 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  $\epsilon > 0$ , we have

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



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

$$\langle x, y \rangle \leq \frac{\epsilon}{2} \|x\|^2 + \frac{1}{2\epsilon} \|y\|^2 \quad (12)$$

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

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

$$\frac{1}{m} \sum_{i=1}^m \|x_i\|^2 \geq \left\| \frac{1}{m} \sum_{i=1}^m x_i \right\|^2 \quad (13)$$

We prove the following result for the virtual MDP.

Lemma 3. For any  $x_1, x_2 \in \mathbb{R}^d$ ,

$$\langle x_2 - x_1, g(x_1) - g(x_2) \rangle \leq (1 - \epsilon) \hat{V}_1 - \hat{V}_2 \quad (14)$$

Proof. Consider a stationary sequence of states with random initial states  $s_0$  and subsequent states  $s_t$ , which, conditioned on  $s_0$ , is drawn from  $P(\cdot | s_0)$ . Define  $\hat{V}_1(s) = \langle x_1, g(x_1) \rangle$  and  $\hat{V}_2(s) = \langle x_2, g(x_2) \rangle$ . Define  $\hat{V}_1(s^0) = \langle x_1, g(x_1) \rangle$  and  $\hat{V}_2(s^0) = \langle x_2, g(x_2) \rangle$ . By stationarity,  $\hat{V}_1$  and  $\hat{V}_2$  are two correlated random variables with the same same marginal distribution. By definition,  $E[\hat{V}_1] = E[\hat{V}_2] = \hat{V}_1 - \hat{V}_2$  since  $s_t$  are drawn from  $P(\cdot | s_0)$ . And we have,

$$\langle x_1, g(x_1) \rangle - \langle x_2, g(x_2) \rangle = E[\langle x_1, g(x_1) \rangle - \langle x_2, g(x_2) \rangle] = E[\langle x_1 - x_2, g(x_1) - g(x_2) \rangle]$$

Therefore,

$$\begin{aligned} \langle x_2 - x_1, g(x_1) - g(x_2) \rangle &= E[\langle x_1 - x_2, g(x_1) - g(x_2) \rangle] \\ &= E[\langle x_1 - x_2, g(x_1) \rangle] - E[\langle x_1 - x_2, g(x_2) \rangle] \\ &= (1 - \epsilon) E[\hat{V}_1] - E[\hat{V}_2] \\ &= (1 - \epsilon) \hat{V}_1 - \hat{V}_2 \end{aligned}$$

where we use the Cauchy-Schwartz inequality to conclude  $E[\langle x_1 - x_2, g(x_1) \rangle] \leq \sqrt{E[\|x_1 - x_2\|^2]} \sqrt{E[\|g(x_1)\|^2]} = E[\hat{V}_1]$ .  $\square$

Lemma 4. For any  $x_1, x_2 \in \mathbb{R}^d$ , we have

$$\|g(x_1) - g(x_2)\| \leq 2 \hat{V}_1 - \hat{V}_2 \quad (15)$$

Proof. Following the analysis of Lemma 3, we have

$$\begin{aligned} \|g(x_1) - g(x_2)\| &= \sqrt{E[\|g(x_1) - g(x_2)\|^2]} \\ &\leq \sqrt{E[\|g(x_1)\|^2] + E[\|g(x_2)\|^2]} \\ &= \sqrt{2E[\hat{V}_1]} \end{aligned}$$

where the second inequality is due to  $k \leq 1$  and the final equality is due to  $\mathbb{E} \|\hat{V}_1\|^2 = \mathbb{E} \|\hat{V}_2\|^2$ . We finish the proof by using the fact that  $\mathbb{E} \|\hat{V}_1\|^2 = \mathbb{E} \|\hat{V}_2\|^2 \leq \frac{2}{D}$  and  $1 + \frac{2}{D} \leq 2$ .  $\square$

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

**Lemma 5.** (2-Lipschitzness of steady-state TD(0) update direction) For any  $\mu_1, \mu_2 \in \mathbb{R}^d$ ; we have

$$\|g(\mu_1) - g(\mu_2)\| \leq 2\|\mu_1 - \mu_2\|. \quad (16)$$

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

$$\|g_i(\mu_1) - g_i(\mu_2)\| \leq 2\|\mu_1 - \mu_2\|. \quad (17)$$

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

$$\|g(\mu_1) - g(\mu_2)\| \leq 2\|\mu_1 - \mu_2\|; \quad (18)$$

based on the fact that  $\max_{\mu \in \mathbb{R}^d} \|\mu\| \leq 1$ . We can follow the same reasoning to prove Eq(17) since  $\|g_i(\mu_1) - g_i(\mu_2)\| \leq 2\|\hat{V}_1\| \|\hat{V}_2\|_{D_i}$  holds for each  $i \in [N]$  from [Bhandari et al. \(2018\)](#).  $\square$

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  $\mu_1, \mu_2 \in \mathbb{R}^d$  and  $i \in [N]$ ; we have

$$\|g_i(\mu_1) - g_i(\mu_2)\| \leq 2\|\mu_1 - \mu_2\|.$$

*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  $\mu$ . In particular, we have  $g_i(\mu) = b_i(O_{t,k}^{(i)}) - A_i(O_{t,k}^{(i)}) \mu$ , where  $A_i(O_{t,k}^{(i)}) = (s_{t,k}^{(i)})(s_{t,k}^{(i)} - s_{t,k+1}^{(i)})$  and  $b_i(O_{t,k}^{(i)}) = r(s_{t,k}^{(i)}) - (s_{t,k}^{(i)})$ . Thus, we have

$$\begin{aligned} \|g_i(\mu_1) - g_i(\mu_2)\| &= \|A_i(O_{t,k}^{(i)}) (\mu_1 - \mu_2)\| \\ &\leq \|A_i(O_{t,k}^{(i)})\| \|\mu_1 - \mu_2\| \\ &\leq \|s_{t,k}^{(i)}\|^2 + \|s_{t,k}^{(i)} - s_{t,k+1}^{(i)}\| \|\mu_1 - \mu_2\| \\ &\leq 2\|\mu_1 - \mu_2\|; \end{aligned}$$

where we used that  $\|s\| \leq 1; \|s\| \leq S$  in the last step.  $\square$

## F 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,  $\mathcal{F}_{t-1}^t$  is to be interpreted as  $\mathcal{F}^t$ . Based on the description of FedTD(0), it should be apparent that for each  $i \in [N]$ ,  $\mathcal{F}_{t,k}^{(i)}$  is  $\mathcal{F}_{t-1}^t$ -measurable and  $\mathcal{F}_k^t$  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  $\sigma_t = \frac{1}{NK} \sum_{i=1}^N \sum_{k=0}^{K-1} \mathcal{V}_{t,k}^{(i)}$  and  $\tau_t = \frac{1}{NK} \sum_{i=1}^N \sum_{k=0}^{K-1} \mathcal{V}_{t,k}^{(i)2}$ . The latter term is referred to as the drift term. Note that  $(\sigma_t)^2 \leq \tau_t$  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  $\eta$ , global step-size  $\eta_g$ , and the effective step-size  $\eta_e$ . Some of our results will rely on effective step-sizes that decay as a function of the communication round  $t$ ; we will use  $f_t g$  to represent such a decaying effective step-size sequence. While the local step-size  $\eta$  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  $f_t^{(g)} g$  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.

## G Warm-up: Analysis of FedTDunder i.i.d. sampling

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  $\{O_{t,k}^{(i)}\}_k$  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 (Dalal et al., 2018; Bhandari et al., 2018; Doan et al., 2019). To proceed, for a fixed  $i$  and for each  $k \in [N]$ , let us define  $g^{(i)}(\cdot)$ ,  $E_{O_{t,k}^{(i)}}[g^{(i)}(\cdot)]$  as the expected TD(0) update direction at iterate  $t$  when the Markov tuple  $O_{t,k}^{(i)}$  hits its stationary distribution  $\pi^{(i)}$ . We make the following standard bounded variance assumption (Bhandari et al., 2018); similar assumptions are also made in FL analyses.

**Assumption 4.**  $\|E_k g^{(i)}(\cdot) - g^{(i)}(\cdot)\|^2 \leq \sigma^2$  holds for all agents  $i \in [N]$ , in each round  $t$  and local update  $k$ , and  $\sigma \leq \delta$ :

Let  $H$  denote the radius of the set  $\mathcal{H}$ . Also, define  $G = R_{\max} + 2H$  and  $\beta = (1 - \delta)!$ , where  $!$  is as in Proposition 3. Our convergence result for FedTD(0) in the i.i.d. setting is as follows.

**Theorem 4.** (I.I.D. Setting) There exists a decreasing global step-size sequence  $\{\eta_t\}_t$ , a fixed local step-size  $\eta_l$ , and a set of convex weights, such that a convex combination  $\bar{w}_t$  of the global models  $\{w_{t,k}\}_k$  satisfies the following for each  $i \in [N]$  after  $T$  rounds:

$$E \|V_{\bar{w}_T} - V_{i,D}\|^2 \leq \frac{G^2}{K^2 T^2} + \frac{\sigma^2}{2NK T} + \frac{\sigma^2}{4KT^2} + Q(\delta; \eta_l); \quad (19)$$

where  $Q(\delta; \eta_l) = \frac{G^2 \beta^2}{2NK} + \sigma^2(\delta; \eta_l)$ ,  $B(\delta; \eta_l) = H \sqrt{\frac{\beta}{n}} + 2(n-1)\eta_l + O(\eta_l^2) + O(\eta_l)$ , and  $(\delta; \eta_l)$  is as defined in Theorem 1.

In what follows, we provide a detailed convergence analysis of the above result.

### G.1 Auxiliary lemmas for Theorem 4

#### G.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  $E \frac{1}{NK} \sum_{i=1}^N \sum_{k=0}^{K-1} \|g^{(i)}(O_{t,k}^{(i)}) - g^{(i)}(O_{t,k}^{(i)})\|^2 \leq \frac{\sigma^2}{NK}$ .

*Proof.* Define  $Y_{t,k}^{(i)} = g^{(i)}(O_{t,k}^{(i)}; O_{t,k}^{(i)}) - g^{(i)}(O_{t,k}^{(i)})$ : Since  $\{O_{t,k}^{(i)}\}_k$  is drawn i.i.d. over time from its stationary distribution  $\pi^{(i)}$ , we have  $E[Y_{t,k}^{(i)}] = E[E[Y_{t,k}^{(i)} | O_{t,k}^{(i)}]] = 0$ : As we mentioned before, for each  $i \in [N]$ ,  $O_{t,k}^{(i)}$  is  $\mathcal{F}_{k-1}^t$ -measurable. If we condition on  $\mathcal{F}_{k-1}^t$ , we know that  $O_{t,k}^{(i)}$  and  $O_{t,k}^{(j)}$  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

$$E \|Y_{t,k}^{(i)}; Y_{t,k}^{(j)}\|^2 = E E \|Y_{t,k}^{(i)}; Y_{t,k}^{(j)}\|^2 | \mathcal{F}_{k-1}^t \stackrel{(a)}{=} E \|E[Y_{t,k}^{(i)} | \mathcal{F}_{k-1}^t]; E[Y_{t,k}^{(j)} | \mathcal{F}_{k-1}^t]\|^2 = 0; \quad (20)$$

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  $i, j \in [N]$ ,

$$E \|Y_{t,k}^{(i)}; Y_{t,l}^{(j)}\|^2 = E E \|Y_{t,k}^{(i)}; Y_{t,l}^{(j)}\|^2 | \mathcal{F}_{l-1}^t = E \|Y_{t,k}^{(i)}; E[Y_{t,l}^{(j)} | \mathcal{F}_{l-1}^t]\|^2 = 0; \quad (21)$$

Then,

$$E \frac{1}{NK} \sum_{i=1}^N \sum_{k=0}^{K-1} \|g^{(i)}(O_{t,k}^{(i)}) - g^{(i)}(O_{t,k}^{(i)})\|^2$$

$$\begin{aligned}
 &= E \frac{1}{NK} \sum_{i=1}^N \sum_{k=0}^{K-1} Y_{t;k}^{(i)2} \\
 &= \frac{1}{N^2K^2} \sum_{i=1}^N \sum_{k=0}^{K-1} E k Y_{t;k}^{(i)2} + \frac{2}{N^2K^2} \sum_{\substack{i < j \\ k=0}}^{K-1} E [h Y_{t;k}^{(i)}; Y_{t;k}^{(j)}] \\
 &\quad + \frac{2}{N^2K^2} \sum_{i,j=1}^N \sum_{k < l}^{K-1} E [h Y_{t;k}^{(i)}; Y_{t;l}^{(j)}] \\
 &\quad \frac{2}{NK};
 \end{aligned}$$

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

### G.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  $\eta$  satisfies  $\eta \leq \frac{(1-\epsilon)^2}{48K}$ , then the updates of FedTQ(0) with any global step-size  $\alpha$  satisfy

$$\begin{aligned}
 E \|x_{t+1} - x^*\|^2 &\leq (1 + \epsilon) E \|x_t - x^*\|^2 + 2 E h \alpha (\epsilon; t) \|x_t - x^*\|^2 + 6 \alpha^2 E g(t)^2 \\
 &\quad + 4 \alpha^2 \frac{1}{N} + 6 \alpha E [x_t] + \frac{2 \alpha^2}{NK} + 2 B(\epsilon; \alpha) G + 6 \alpha^2 B^2(\epsilon; \alpha); \tag{22}
 \end{aligned}$$

where  $\epsilon$  is any positive constant, and  $\alpha$  is the effective step-size, i.e.,  $\alpha = K \eta$ .

Proof.

$$\begin{aligned}
 &E \|x_{t+1} - x^*\|^2 \\
 &= E \|x_t + \frac{1}{NK} \sum_{i=1}^N \sum_{k=0}^{K-1} g_i^{(i)}(t, k) - x^*\|^2 \quad (\text{updating rule}) \\
 &= E \|x_t + \frac{1}{NK} \sum_{i=1}^N \sum_{k=0}^{K-1} g_i^{(i)}(t, k) - x_t + x_t - x^*\|^2 \quad (\text{projection is non-expansive}) \\
 &= E \|x_t - x^*\|^2 + 2 E h \frac{1}{NK} \sum_{i=1}^N \sum_{k=0}^{K-1} g_i^{(i)}(t, k) \cdot (x_t - x^*) + E \frac{1}{NK} \sum_{i=1}^N \sum_{k=0}^{K-1} g_i^{(i)}(t, k)^2 \\
 &= E \|x_t - x^*\|^2 + \frac{2}{NK} \sum_{i=1}^N \sum_{k=0}^{K-1} E h g_i^{(i)}(t, k) \cdot (x_t - x^*) + E \frac{1}{NK} \sum_{i=1}^N \sum_{k=0}^{K-1} g_i^{(i)}(t, k)^2 \\
 &\quad + \frac{2}{NK} \sum_{i=1}^N \sum_{k=0}^{K-1} E h g_i^{(i)}(t, k) \cdot (x_t - x^*) + E \frac{1}{NK} \sum_{i=1}^N \sum_{k=0}^{K-1} g_i^{(i)}(t, k)^2 \\
 &= E \|x_t - x^*\|^2 + \frac{2}{NK} \sum_{i=1}^N \sum_{k=0}^{K-1} E h g_i^{(i)}(t, k) \cdot (x_t - x^*) + E \frac{1}{NK} \sum_{i=1}^N \sum_{k=0}^{K-1} g_i^{(i)}(t, k)^2 \\
 &\quad + 2 E \frac{1}{NK} \sum_{i=1}^N \sum_{k=0}^{K-1} h g_i^{(i)}(t, k) \cdot (x_t - x^*) + 2 E \frac{1}{NK} \sum_{i=1}^N \sum_{k=0}^{K-1} g_i^{(i)}(t, k)^2 \quad (\text{Young's inequality (12)})
 \end{aligned}$$

$$\begin{aligned}
& \text{(a) } E_t \left[ \sum_{i=1}^N \sum_{k=0}^{K-1} \frac{2}{NK} \text{Ehg} \left( \begin{matrix} (i) \\ t;k \end{matrix} \right); t \right]^2 + \frac{2}{NK} + 2E \sum_{i=1}^N \frac{1}{NK} g \left( \begin{matrix} (i) \\ t;k \end{matrix} \right)^2 \\
&= E_t \left[ \sum_{i=1}^N \sum_{k=0}^{K-1} \frac{2}{NK} \text{Ehg} \left( \begin{matrix} (i) \\ t;k \end{matrix} \right) g \left( \begin{matrix} (i) \\ t \end{matrix} \right) + g \left( \begin{matrix} (i) \\ t \end{matrix} \right) g \left( \begin{matrix} (i) \\ t \end{matrix} \right) + g \left( \begin{matrix} (i) \\ t \end{matrix} \right); t \right]^2 + \\
&+ 2E \sum_{i=1}^N \sum_{k=0}^{K-1} \frac{1}{NK} g \left( \begin{matrix} (i) \\ t;k \end{matrix} \right)^2 + \frac{2}{NK} \\
&E_t \left[ \sum_{i=1}^N \sum_{k=0}^{K-1} \frac{2}{NK} \text{Ehg} \left( \begin{matrix} (i) \\ t;k \end{matrix} \right) g \left( \begin{matrix} (i) \\ t \end{matrix} \right); t \right]^2 + \frac{2}{N} \sum_{i=1}^N \text{Ehg} \left( \begin{matrix} (i) \\ t \end{matrix} \right) g \left( \begin{matrix} (i) \\ t \end{matrix} \right); t \right]^2 + \\
&+ 2 \text{Ehg} \left( \begin{matrix} (i) \\ t \end{matrix} \right); t \right]^2 + 2E \sum_{i=1}^N \sum_{k=0}^{K-1} \frac{1}{NK} g \left( \begin{matrix} (i) \\ t;k \end{matrix} \right)^2 + \frac{2}{NK} \\
&(1 + \frac{1}{1}) E_t \left[ \sum_{i=1}^N \sum_{k=0}^{K-1} \frac{1}{NK} \text{Ehg} \left( \begin{matrix} (i) \\ t;k \end{matrix} \right) g \left( \begin{matrix} (i) \\ t \end{matrix} \right) \right]^2 + 2 B \left( \begin{matrix} (i) \\ t \end{matrix} \right) G \\
&+ 2 \text{Ehg} \left( \begin{matrix} (i) \\ t \end{matrix} \right); t \right]^2 + 2E \sum_{i=1}^N \sum_{k=0}^{K-1} \frac{1}{NK} g \left( \begin{matrix} (i) \\ t;k \end{matrix} \right)^2 + \frac{2}{NK} \quad (\text{Eq (12) and Lemma 2}) \\
&(1 + \frac{1}{1}) E_t \left[ \sum_{i=1}^N \sum_{k=0}^{K-1} \frac{4}{1NK} \text{Ehg} \left( \begin{matrix} (i) \\ t;k \end{matrix} \right) \right]^2 + 2 B \left( \begin{matrix} (i) \\ t \end{matrix} \right) G \\
&+ 2 \text{Ehg} \left( \begin{matrix} (i) \\ t \end{matrix} \right); t \right]^2 + 2E \sum_{i=1}^N \sum_{k=0}^{K-1} \frac{1}{NK} g \left( \begin{matrix} (i) \\ t;k \end{matrix} \right)^2 + \frac{2}{NK} \quad (2\text{-Lipschitz of } g_i \text{ in Lemma 5}) \\
&(1 + \frac{1}{1}) E_t \left[ \sum_{i=1}^N \sum_{k=0}^{K-1} \frac{4}{1} \text{Ehg} \left( \begin{matrix} (i) \\ t;k \end{matrix} \right) \right]^2 + \frac{2}{NK} + 2 B \left( \begin{matrix} (i) \\ t \end{matrix} \right) G \\
&+ 2 \text{Ehg} \left( \begin{matrix} (i) \\ t \end{matrix} \right); t \right]^2 + 2E \sum_{i=1}^N \sum_{k=0}^{K-1} \frac{1}{NK} \text{Ehg} \left( \begin{matrix} (i) \\ t;k \end{matrix} \right) g \left( \begin{matrix} (i) \\ t \end{matrix} \right) + g \left( \begin{matrix} (i) \\ t \end{matrix} \right) g \left( \begin{matrix} (i) \\ t \end{matrix} \right) + g \left( \begin{matrix} (i) \\ t \end{matrix} \right)^2 \\
&(1 + \frac{1}{1}) E_t \left[ \sum_{i=1}^N \sum_{k=0}^{K-1} \frac{4}{1} \text{Ehg} \left( \begin{matrix} (i) \\ t;k \end{matrix} \right) \right]^2 + \frac{2}{NK} + 2 B \left( \begin{matrix} (i) \\ t \end{matrix} \right) G \\
&+ 2 \text{Ehg} \left( \begin{matrix} (i) \\ t \end{matrix} \right); t \right]^2 + 6E \sum_{i=1}^N \sum_{k=0}^{K-1} \frac{1}{NK} g \left( \begin{matrix} (i) \\ t;k \end{matrix} \right) g \left( \begin{matrix} (i) \\ t \end{matrix} \right)^2 \\
&+ 6E \sum_{i=1}^N \frac{1}{N} g \left( \begin{matrix} (i) \\ t \end{matrix} \right) g \left( \begin{matrix} (i) \\ t \end{matrix} \right)^2 + 6E g \left( \begin{matrix} (i) \\ t \end{matrix} \right)^2 \quad (\text{Eq (12) and Lemma 2}) \\
&(1 + \frac{1}{1}) E_t \left[ \sum_{i=1}^N \sum_{k=0}^{K-1} \frac{4}{1} \text{Ehg} \left( \begin{matrix} (i) \\ t;k \end{matrix} \right) \right]^2 + \frac{2}{NK} + 2 B \left( \begin{matrix} (i) \\ t \end{matrix} \right) G \\
&+ 2 \text{Ehg} \left( \begin{matrix} (i) \\ t \end{matrix} \right); t \right]^2 + 24 \text{Ehg} \left( \begin{matrix} (i) \\ t \end{matrix} \right) \quad (2\text{-Lipschitz of } g_i) \\
&+ 6 \text{Ehg} \left( \begin{matrix} (i) \\ t \end{matrix} \right); t \right]^2 + 6 \text{Ehg} \left( \begin{matrix} (i) \\ t \end{matrix} \right) \quad (\text{Eq (12)}) \\
&= (1 + \frac{1}{1}) E_t \left[ \sum_{i=1}^N \sum_{k=0}^{K-1} \frac{2}{NK} \text{Ehg} \left( \begin{matrix} (i) \\ t;k \end{matrix} \right) \right]^2 + 2 \text{Ehg} \left( \begin{matrix} (i) \\ t \end{matrix} \right); t \right]^2 + 6 \text{Ehg} \left( \begin{matrix} (i) \\ t \end{matrix} \right) \\
&+ 4 \text{Ehg} \left( \begin{matrix} (i) \\ t \end{matrix} \right); t \right]^2 + 6 \text{Ehg} \left( \begin{matrix} (i) \\ t \end{matrix} \right) + \frac{2}{NK} + 2 B \left( \begin{matrix} (i) \\ t \end{matrix} \right) G + 6 \text{Ehg} \left( \begin{matrix} (i) \\ t \end{matrix} \right) \quad (23)
\end{aligned}$$

where (a) is due to Lemma 7. Furthermore, the reason why  $C_t = 0$  is as follows:

$$C_t = \sum_{i=1}^N \sum_{k=0}^{K-1} \text{Ehg} \left( \begin{matrix} (i) \\ t;k \end{matrix} \right) g \left( \begin{matrix} (i) \\ t;k \end{matrix} \right); t \right]^2 + \frac{2}{NK}$$

$$\begin{aligned}
 &= \sum_{i=1}^N \sum_{k=0}^{K-2} \mathbb{E} \mathbf{h}_{t,k}^{(i)} \mathbf{g}_{t,k}^{(i)}; t \quad i+ \\
 &\quad \sum_{i=1}^N \mathbb{E} \mathbf{h}_{t,K-1}^{(i)} \mathbf{g}_{t,K-1}^{(i)}; t \quad i \\
 &= \sum_{i=1}^N \sum_{k=0}^{K-2} \mathbb{E} \mathbf{h}_{t,k}^{(i)} \mathbf{g}_{t,k}^{(i)}; t \quad i+ \\
 &\quad \sum_{i=1}^N \mathbb{E} \mathbf{h}_{t,K-1}^{(i)} \mathbf{g}_{t,K-1}^{(i)}; t \quad i \quad \mathbb{E} \mathbf{F}_{K-1}^t \quad ii \\
 &= \sum_{i=1}^N \sum_{k=0}^{K-2} \mathbb{E} \mathbf{h}_{t,k}^{(i)} \mathbf{g}_{t,k}^{(i)}; t \quad i+ \\
 &\quad \sum_{i=1}^N \mathbb{E} \mathbf{h}_{t,K-1}^{(i)} \mathbf{g}_{t,K-1}^{(i)}; t \quad i \quad \mathbb{E} \mathbf{F}_{K-1}^t \quad iii \\
 &= \sum_{i=1}^N \sum_{k=0}^{K-2} \mathbb{E} \mathbf{h}_{t,k}^{(i)} \mathbf{g}_{t,k}^{(i)}; t \quad i:
 \end{aligned}$$

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

### G.1.3 Drift Term Analysis

We now turn to bounding the drift term  $\Delta_t$ .

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

$$\mathbb{E}[\Delta_t] = \frac{1}{NK} \sum_{i=1}^N \sum_{k=0}^{K-1} \mathbb{E} \|\mathbf{g}_{t,k}^{(i)}\|_2^2 \leq 27(\eta^2 + 3KB^2(\eta; \eta) + 2KG^2) \frac{\eta^2}{K^{\frac{2}{g}}}; \tag{24}$$

provided the fixed local step-size  $\eta$  satisfies  $\eta \leq \min\{\frac{1}{48K}\}$ :

Proof.

$$\begin{aligned}
 &\mathbb{E} \|\mathbf{g}_{t,k}^{(i)}\|_2^2 \\
 &= \mathbb{E} \|\mathbf{g}_{t,k-1}^{(i)} + \eta \mathbf{g}_{t,k-1}^{(i)}\|_2^2 \quad (\text{updating rule}) \\
 &= \mathbb{E} \|\mathbf{g}_{t,k-1}^{(i)}\|_2^2 + \eta^2 \mathbb{E} \|\mathbf{g}_{t,k-1}^{(i)}\|_2^2 + 2\eta \mathbb{E} \langle \mathbf{g}_{t,k-1}^{(i)}, \mathbf{g}_{t,k-1}^{(i)} \rangle \\
 &= \mathbb{E} \|\mathbf{g}_{t,k-1}^{(i)}\|_2^2 + \eta^2 \mathbb{E} \|\mathbf{g}_{t,k-1}^{(i)}\|_2^2 + 2\eta \mathbb{E} \langle \mathbf{g}_{t,k-1}^{(i)}, \mathbf{g}_{t,k-1}^{(i)} \rangle \\
 &\quad + 2\eta \mathbb{E} \langle \mathbf{g}_{t,k-1}^{(i)}, \mathbf{g}_{t,k-1}^{(i)} \rangle; \mathbb{E} \mathbf{F}_{k-1}^t \quad \mathbb{E} ii \\
 &\quad \underbrace{\mathbb{E} \langle \mathbf{g}_{t,k-1}^{(i)}, \mathbf{g}_{t,k-1}^{(i)} \rangle}_{C_2=0} \\
 &\stackrel{(a)}{=} (1 + 2\eta) \mathbb{E} \|\mathbf{g}_{t,k-1}^{(i)}\|_2^2 + (1 + \frac{1}{2}) \eta^2 \mathbb{E} \|\mathbf{g}_{t,k-1}^{(i)}\|_2^2 \\
 &\quad + \eta^2 \mathbb{E} \|\mathbf{g}_{t,k-1}^{(i)}\|_2^2 \\
 &\stackrel{(b)}{=} (1 + 2\eta)(1 + 3\eta) \mathbb{E} \|\mathbf{g}_{t,k-1}^{(i)}\|_2^2 + \eta^2 \mathbb{E} \|\mathbf{g}_{t,k-1}^{(i)}\|_2^2
 \end{aligned}$$

$$\begin{aligned}
 & + (1 + \alpha_2)(1 + \frac{1}{3}) \frac{1}{l} \mathbb{E} \|g(t)\|^2 \\
 & + (1 + \frac{1}{2}) \frac{1}{l} \mathbb{E} \|g(t_{k-1}^{(i)}) - g(t) + g(t) - g_i(t) + g_i(t) - g_i(t_{k-1}^{(i)})\|^2 + \frac{1}{l} \mathbb{E} \|g(t)\|^2 \\
 (c) \quad & (1 + \alpha_2)(1 + \alpha_3) \mathbb{E} \|g(t_{k-1}^{(i)}) - g(t) + g(t) - g_i(t) + g_i(t) - g_i(t_{k-1}^{(i)})\|^2 + (1 + \alpha_2)(1 + \frac{1}{3}) \frac{1}{l} \mathbb{E} \|g(t)\|^2 \\
 & + 3(1 + \frac{1}{2}) \frac{1}{l} \mathbb{E} \|g(t_{k-1}^{(i)}) - g(t)\|^2 + 3(1 + \frac{1}{2}) \frac{1}{l} \mathbb{E} \|g(t) - g_i(t)\|^2 \\
 & + 3(1 + \frac{1}{2}) \frac{1}{l} \mathbb{E} \|g(t) - g_i(t_{k-1}^{(i)})\|^2 + \frac{1}{l} \mathbb{E} \|g(t)\|^2 \\
 (d) \quad & (1 + \alpha_2)(1 + \alpha_3) \frac{1}{l} (2 - \alpha_1(1 - \alpha_2)) \frac{4}{l} \frac{1}{l} \mathbb{E} \|g(t_{k-1}^{(i)}) - g(t)\|^2 \\
 & + (1 + \alpha_2)(1 + \frac{1}{3}) \frac{1}{l} \mathbb{E} \|g(t)\|^2 \\
 & + 12(1 + \frac{1}{2}) \frac{1}{l} \mathbb{E} \|g(t_{k-1}^{(i)}) - g(t)\|^2 + 3(1 + \frac{1}{3}) \frac{1}{l} \mathbb{E} \|B^2(\cdot; \alpha_1)\|^2 \\
 & + 12(1 + \frac{1}{3}) \frac{1}{l} \mathbb{E} \|g(t_{k-1}^{(i)}) - g(t)\|^2 + \frac{1}{l} \mathbb{E} \|g(t)\|^2 \\
 = & (1 + \alpha_2)(1 + \alpha_3) \frac{1}{l} (2 - \alpha_1(1 - \alpha_2)) \frac{4}{l} \frac{1}{l} \mathbb{E} \|g(t_{k-1}^{(i)}) - g(t)\|^2 + \frac{24(1 + \frac{1}{3}) \frac{1}{l} \mathbb{E} \|g(t)\|^2}{(1 + \alpha_2)(1 + \alpha_3)} \# \\
 & + (1 + \alpha_2)(1 + \frac{1}{3}) \frac{1}{l} \mathbb{E} \|g(t)\|^2 + 3(1 + \frac{1}{3}) \frac{1}{l} \mathbb{E} \|B^2(\cdot; \alpha_1)\|^2 + \frac{1}{l} \mathbb{E} \|g(t)\|^2; \\
 & \underbrace{\hspace{10em}}_{D_1}
 \end{aligned}$$

where we used the inequality in Eq(11) with any positive constant  $\alpha_2$  for (a); for (b), we used Assumption 4 and the same reasoning as Eq(11) with any positive constant  $\alpha_3$ ; 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  $g, g_i$  (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  $\alpha_4 = (1 + \alpha_2)(1 + \alpha_3) \frac{1}{l} (2 - \alpha_1(1 - \alpha_2)) \frac{4}{l} \frac{1}{l} + \frac{24(1 + \frac{1}{3}) \frac{1}{l} \mathbb{E} \|g(t)\|^2}{(1 + \alpha_2)(1 + \alpha_3)}$  and define  $D_1$  as above, we have that

$$\mathbb{E} \|g(t_{k-1}^{(i)}) - g(t)\|^2 \leq \alpha_4 \mathbb{E} \|g(t_{k-1}^{(i)}) - g(t)\|^2 + D_1; \tag{25}$$

Next, we set  $\alpha_2 = \alpha_3 = \frac{1}{K-1}$ ;  $K \geq 2$ , and choose the local step-size  $\eta$  to satisfy

$$\frac{l(1 - \alpha_1)^l}{2} \frac{4}{l} \frac{1}{l} \mathbb{E} \|g(t)\|^2 \leq \frac{l(1 - \alpha_1)^l}{2} \frac{24(1 + \frac{1}{3}) \frac{1}{l} \mathbb{E} \|g(t)\|^2}{(1 + \alpha_2)(1 + \alpha_3)};$$

so that  $\frac{1}{l} (2 - \alpha_1(1 - \alpha_2)) \frac{4}{l} \frac{1}{l} \mathbb{E} \|g(t)\|^2 + \frac{24(1 + \frac{1}{3}) \frac{1}{l} \mathbb{E} \|g(t)\|^2}{(1 + \alpha_2)(1 + \alpha_3)} \leq \frac{l(1 - \alpha_1)^l}{2}$ . These inequalities hold when  $\eta \leq \min \frac{l(1 - \alpha_1)^l}{48K}$ : Then, Eq (25) becomes

$$\mathbb{E} \|g(t_{k-1}^{(i)}) - g(t)\|^2 \leq (1 + \frac{3}{K-1}) [1 - \frac{l(1 - \alpha_1)^l}{2}] \mathbb{E} \|g(t_{k-1}^{(i)}) - g(t)\|^2 + D_1;$$

If we unroll this recurrence above, using  $g_{r,0}^{(i)} = g(t)$ , we have that

$$\begin{aligned}
 \mathbb{E} \|g(t_k) - g(t)\|^2 & \leq \prod_{s=0}^{k-1} D_1 \prod_{j=s+1}^{k-1} (1 + \frac{3}{K-1}) [1 - \frac{l(1 - \alpha_1)^l}{2}] \\
 & \stackrel{(e)}{\leq} \prod_{s=0}^{k-1} \left[ \frac{1}{l} \mathbb{E} \|g(t)\|^2 + 3K \frac{1}{l} \mathbb{E} \|B^2(\cdot; \alpha_1)\|^2 + 2 \frac{1}{l} \mathbb{E} \|g(t)\|^2 \right]
 \end{aligned}$$



$$\begin{aligned} & \prod_{j=s+1}^k \left(1 + \frac{3}{K-1}\right) [1 - \rho(1 - \rho)^i] \\ & \sum_{s=0}^{k-1} \left( \rho^2 + 3 \rho^2 KB^2(\rho; \rho) + 2 \rho^2 KE g(\rho) \right) \rho^{2i} \\ & \left(1 + \frac{3}{K-1}\right)^{k-1} \prod_{j=s+1}^k [1 - \rho(1 - \rho)^i] \\ (f) \quad & 27 \left( \rho^2 + 3KB^2(\rho; \rho) + 2KE g(\rho) \right) \sum_{s=0}^{k-1} \rho^{2i} \left| \frac{\rho^{k-1} [1 - \rho(1 - \rho)^i]}{1 - \rho} \right| \\ & 27 \left( \rho^2 + 3KB^2(\rho; \rho) + 2KG^2 \right) K \rho^2 \quad (\text{constant local step-size } \rho) \end{aligned}$$

where we used the fact that  $(1 + \rho)(1 + \frac{1}{3}) \leq 2K$  for (e) and  $(1 + \frac{3}{K-1})^{k-1} \leq 27$  for (f). we finish the proof by substituting  $\rho = \frac{1}{K-9}$ .  $\square$

If we incorporate Eq (24) into Eq (22), we have that

$$\begin{aligned} E_{t+1} & \leq \left(1 + \frac{1}{3}\right) E_t + 2 E_t h(\rho; \rho) + 6 \rho^2 E_t g(\rho) + 108 \frac{\rho^4}{K-9} \left( \rho^2 + 3KB^2(\rho; \rho) + 2KG^2 \right) + \frac{2 \rho^2}{NK} + 2 B(\rho; \rho) G + 6 \rho^2 B^2(\rho; \rho); \quad (26) \\ & \quad + 108 \frac{\rho^4}{K-9} \left( \rho^2 + 3KB^2(\rho; \rho) + 2KG^2 \right) + \frac{2 \rho^2}{NK} + 2 B(\rho; \rho) G + 6 \rho^2 B^2(\rho; \rho); \quad (27) \end{aligned}$$

### G.1.4 Parameter Selection

Lemma 10. Define  $\rho = \frac{1}{K-9}$ . If we choose any effective step-size  $\rho = \frac{1}{K-9} < \frac{(1-\rho)^!}{96}$ ; any local step-size  $\rho \leq \min\left\{\frac{(1-\rho)^!}{48K}\right\}$ , and choose the constant  $\rho = \frac{1}{K-9}$ , the updates of FedTD(0) satisfy

$$\begin{aligned} \rho E_{t+1} & \leq V_t + V_D \left( \frac{1}{K-9} \right) E_t + \frac{1}{K-9} E_{t+1} + \frac{2 \rho^2}{NK} \\ & \quad + \frac{1080 \rho^2}{K-9} \left( \rho^2 + 3KB^2(\rho; \rho) + 2KG^2 \right) + \frac{2 B(\rho; \rho) G + 6 \rho^2 B^2(\rho; \rho)}{\left| \frac{\rho^{k-1} [1 - \rho(1 - \rho)^i]}{1 - \rho} \right|}; \quad (28) \end{aligned}$$

$O(\rho^2)$  heterogeneity term

where  $\rho = \frac{1}{4} = \frac{(1-\rho)^!}{4}$ .

Proof. From Eq (26) and  $\rho = \frac{1}{K-9}$ , we know

$$\begin{aligned} E_{t+1} & \leq \left(1 + \frac{1}{3}\right) E_t + 2 E_t h(\rho; \rho) + 6 \rho^2 E_t g(\rho) + 108 \frac{\rho^4}{K-9} \left( \rho^2 + 3KB^2(\rho; \rho) + 2KG^2 \right) + \frac{2 \rho^2}{NK} + 2 B(\rho; \rho) G + 6 \rho^2 B^2(\rho; \rho) \\ & \quad + 108 \frac{\rho^4}{K-9} \left( \rho^2 + 3KB^2(\rho; \rho) + 2KG^2 \right) + \frac{2 \rho^2}{NK} + 2 B(\rho; \rho) G + 6 \rho^2 B^2(\rho; \rho) \quad (\text{Lemma 3 and 4}) \\ & \quad + 108 \frac{\rho^4}{K-9} \left( \rho^2 + 3KB^2(\rho; \rho) + 2KG^2 \right) + 2 B(\rho; \rho) G + 6 \rho^2 B^2(\rho; \rho) \\ & \quad + \left(1 + \frac{1}{2}\right) E_t + \frac{1}{2} E_t + 24 \rho^2 E_t V_t + V_D + \frac{2 \rho^2}{NK} \end{aligned}$$

$$\begin{aligned}
 & + 108 \frac{4}{K \frac{2}{g}} (6 + \frac{1}{g}) ( \frac{2}{g} + 3KB^2(\cdot; \frac{1}{g}) + 2KG^2 ) + 2B(\cdot; \frac{1}{g})G + 6 \frac{2}{g} B^2(\cdot; \frac{1}{g}) \\
 (a) \quad & (1 - \frac{2}{g}) E_{t+1}^2 - \frac{2}{g} E_{t+1} V_{t+1}^2 + \frac{2}{4} E_{t+1} V_{t+1}^2 + \frac{2 \frac{2}{g}^2}{NK} \\
 & + 108 \frac{4}{K \frac{2}{g}} (6 + \frac{1}{g}) ( \frac{2}{g} + 3KB^2(\cdot; \frac{1}{g}) + 2KG^2 ) + 2B(\cdot; \frac{1}{g})G + 6 \frac{2}{g} B^2(\cdot; \frac{1}{g});
 \end{aligned}$$

where (a) comes from  $\max(\lambda^T D) \leq 1$  and  $24 \frac{2}{g} \frac{24 \frac{(1-\frac{2}{g})^w}{96}}{96} = \frac{2}{4}$ . Moving  $E_{t+1} V_{t+1}^2$  (on the right-hand side of (a)) to the left hand side of the above inequality yields:

$$\begin{aligned}
 & -\frac{2}{4} E_{t+1} V_{t+1}^2 + (1 - \frac{2}{g}) E_{t+1}^2 - E_{t+1}^2 + \frac{2 \frac{2}{g}^2}{NK} \\
 & + 108 ( \frac{6 \frac{4}{g}^3}{K \frac{2}{g}} + \frac{4 \frac{3}{g}}{K \frac{2}{g}} ) ( \frac{2}{g} + 3KB^2(\cdot; \frac{1}{g}) + 2KG^2 ) + 2B(\cdot; \frac{1}{g})G + 6 \frac{2}{g} B^2(\cdot; \frac{1}{g});
 \end{aligned}$$

Dividing by  $\frac{2}{4} E_{t+1} V_{t+1}^2$  on both sides of the inequality above and changing into  $\frac{1}{g}$ , we have:

$$\begin{aligned}
 & (1 - \frac{1}{g}) E_{t+1}^2 - E_{t+1}^2 + \frac{2 \frac{2}{g}^2}{NK} \\
 & + 108 ( \frac{6 \frac{3}{g}}{K \frac{2}{g}} + \frac{4 \frac{2}{g}}{K \frac{2}{g}} ) ( \frac{2}{g} + 3KB^2(\cdot; \frac{1}{g}) + 2KG^2 ) + 2B(\cdot; \frac{1}{g})G + 6 \frac{2}{g} B^2(\cdot; \frac{1}{g}) \\
 & (1 - \frac{1}{g}) E_{t+1}^2 - E_{t+1}^2 + \frac{2 \frac{2}{g}^2}{NK} \\
 & + \frac{1080 \frac{2}{g}}{K \frac{2}{g}} ( \frac{2}{g} + 3KB^2(\cdot; \frac{1}{g}) + 2KG^2 ) + \underbrace{2B(\cdot; \frac{1}{g})G + 6 \frac{2}{g} B^2(\cdot; \frac{1}{g})}_{\text{heterogeneity term}};
 \end{aligned}$$

where we used the fact that  $\frac{1}{g} \leq 1$  in the last inequality. □

With these lemmas, we are now ready to prove Theorem 4, which we restate for clarity.

### G.2 Proof of Theorem 4

Given a fixed local step-size  $\frac{1}{g} = \frac{1}{2} \frac{(1-\frac{1}{g})!}{48K}$ ; decreasing effective step-sizes  $\frac{2}{g} = \frac{8}{(a+t+1)} = \frac{8}{(1-\frac{1}{g})! (a+t+1)}$ , decreasing global step-sizes  $\frac{2}{g} = \frac{1}{K \frac{1}{g}}$ , and weights  $w_t = (a+t)$ , we have that

$$E_{t+1} V_{t+1}^2 - V_{t+1}^2 \leq \frac{G^2}{K^2 T^2} + \frac{2}{4KT^2} + \frac{2}{2NKT} + \frac{B(\cdot; \frac{1}{g})G}{2} + \frac{2}{g} B^2(\cdot; \frac{1}{g}) \tag{29}$$

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

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

$$\tilde{T} = \frac{1}{W} \sum_{t=1}^T w_t;$$

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

$$\frac{1}{g} E_{t+1} V_{t+1}^2 - V_{t+1}^2$$

$$\begin{aligned}
 & \frac{1}{W} \sum_{t=1}^T X^T (a+t) E V_{i_t} V_D^2 \\
 (28) \quad & \frac{1(a+1)(a+2)G^2}{2W} + \frac{1}{W} \sum_{t=1}^T \frac{2(a+t)}{NK} \\
 & + \frac{1}{W} \sum_{t=1}^T \frac{1080(a+t)}{K \frac{2}{g} W} (t^2 + 3KB^2(\cdot; \cdot) + 2KG^2) \\
 & + \frac{1}{W} \sum_{t=1}^T (a+t) 2B(\cdot; \cdot)G + 6 \sum_{t=1}^T B^2(\cdot; \cdot) \\
 & \frac{1(a+1)(a+2)G^2}{2W} + \frac{2}{NKW} \sum_{t=1}^T (a+t) t \\
 & + \frac{1080(t^2 + 3KB^2(\cdot; \cdot) + 2KG^2)}{K \frac{2}{g} W} \sum_{t=1}^T (a+t) t^2 + 2B(\cdot; \cdot)G + \frac{6B^2(\cdot; \cdot)}{W} \sum_{t=1}^T (a+t) t \\
 & \frac{1(a+1)(a+2)G^2}{2W} + \frac{4}{1NKW} T \\
 & + \frac{4320(t^2 + 3KB^2(\cdot; \cdot) + 2KG^2)}{K \frac{2}{g} W} (1 + \log(a+T)) + 2B(\cdot; \cdot)G + \frac{12B^2(\cdot; \cdot)}{1W} T;
 \end{aligned}$$

where we used  $V_0 V_D^2 G^2$ : Dividing by  $\frac{1}{2}$  on both sides, changing  $\frac{1}{2}$  into  $\frac{1}{4}$ , and using  $W \frac{T(a+T)}{2}$ , we have:

$$E V_{\tau} V_D^2 \leq \frac{G^2}{K^2 T^2} + \frac{2}{4KT^2} + \frac{2}{2NKT} + \frac{B(\cdot; \cdot)G}{1} :$$

We finish the proof by using the following inequality:  $E V_{\tau} V_D^2 \leq 2E V_{\tau} V_D^2 + 2E V_i V_D^2$ , in tandem with the third point in Theorem 1. □

## H Heterogeneity bias: Proof of Theorem 3

In this section, we prove Theorem 3.

**Proof of Theorem 3.** As  $x_1$  and  $x_2$  are the TD(0) fixed points of agents 1 and 2, respectively, we have  $x_1 = A_1^{-1}b_1$  and  $x_2 = A_2^{-1}b_2$ . The output of mean-path FedTD(0) with  $k = 1$  and  $\alpha = \alpha_1$  satisfies:

$$\begin{aligned}
 x_{1,t+1} &= x_{1,t} + \alpha_1 (A_1 x_{1,t} + b_1) \\
 \Rightarrow x_{1,t+1} - x_1 &= x_{1,t} - x_1 + \alpha_1 (A_1(x_{1,t} - x_1) + b_1 - A_1 x_1) \\
 \Rightarrow e_{1,t+1} &= (I - \alpha_1 A_1) e_{1,t} + \alpha_1 (b_1 - A_1 x_1) \\
 \Rightarrow e_{1,t+1} &= (I - \alpha_1 A_1) e_{1,t} + \frac{A_1 + A_2}{2} A_1^{-1} b_1 + \frac{b_1 + b_2}{2} \\
 \Rightarrow e_{1,t+1} &= (I - \alpha_1 A_1) e_{1,t} + \frac{A_2 A_1^{-1} b_1}{2} + \frac{b_2}{2} \\
 \Rightarrow e_{1,t+1} &= (I - \alpha_1 A_1) e_{1,t} + \frac{A_2}{2} A_1^{-1} b_1 + A_2^{-1} b_2 \\
 \Rightarrow e_{1,t+1} &= \underbrace{(I - \alpha_1 A_1)}_{\mathcal{A}} e_{1,t} + \underbrace{\frac{A_2}{2} A_1^{-1} b_1 + A_2^{-1} b_2}_{\Upsilon}
 \end{aligned} \tag{30}$$

Let us now note that  $e_{1,t+1} = \mathcal{A}e_{1,t} + \Upsilon$  can be viewed as a discrete-time linear time-invariant (LTI) system where  $\alpha_1$  is chosen s.t.  $\mathcal{A}$  is Schur stable, i.e.,  $\rho(\mathcal{A}) < 1$ . At the  $t$ -th iteration, we have:

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

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

$$\begin{aligned}
 \lim_{t \rightarrow \infty} e_{1,t} &= (I - \mathcal{A})^{-1} \Upsilon \\
 &= \mathcal{A}^{-1} \frac{A_2}{2} (x_1 - x_2) \\
 &= \frac{1}{2} \mathcal{A}^{-1} A_2 (x_1 - x_2)
 \end{aligned} \tag{31}$$

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

## I Proof of the Markovian setting

We now turn our attention to proving the main result of the paper, namely, Theorem 2.

### I.1 Outline

As mentioned in the main body, one of the main obstacles to overcome in the analysis is that in general,  $E[\sum_{i=1}^N \mathbf{g}_i(\mathbf{O}_{t;k}^{(i)}; \mathbf{O}_{t;k}^{(i)})] \neq 0$ . In order to show that a linear speedup is achievable, we first decompose the random TD direction of each agent as  $\mathbf{g}_i(\mathbf{O}_{t;k}^{(i)}) = \mathbf{b}_i(\mathbf{O}_{t;k}^{(i)}) + \mathbf{A}_i(\mathbf{O}_{t;k}^{(i)}) \mathbf{z}_i^{(i)}$  in subsection I.2.1 and show that the variances of  $\sum_{i=1}^N \mathbf{A}_i(\mathbf{O}_{t;k}^{(i)})$  and  $\sum_{i=1}^N \mathbf{b}_i(\mathbf{O}_{t;k}^{(i)})$  get scaled down by  $NK$  in subsection I.2.2. To decouple the randomness between the parameters  $\mathbf{z}_i^{(i)}$  and the observations  $\mathbf{O}_{t;k}^{(i)}$  using the method called information theoretic control of coupling in Bhandari et al. (2018), we need to bound  $E[\sum_{t=0}^T \sum_{k=0}^{K-1} \|\mathbf{z}_i^{(i)}\|^2]$  in subsection I.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 I.2.4, I.2.5 and I.2.6, respectively. Finally, we prove Theorem 2 in subsection I.3.

**Additional Notation:** Under Assumption 3, for each MDP  $i$ , there exists some  $m_i \in [1, 2]$  and some  $\beta_i \in (0, 1)$ , such that for all  $t \geq 0$  and  $0 \leq k < K-1$ , it holds that

$$d_{TV}(\mathcal{P}_{t;k}^{(i)} \circ \mathbf{s}_{0;0}^{(i)} = \mathbf{s}_{t;k}^{(i)}; \mathcal{P}_{t;k}^{(i)} \circ \mathbf{m}_i^{tK+k}; \mathcal{S}) \leq \beta_i$$

Furthermore, we define  $\beta = \max_{i \in [N]} \beta_i$ ,  $m = \max_{i \in [N]} m_i$ :

### I.2 Auxiliary lemmas for Theorem 2

#### I.2.1 Decomposition Form

The first step in our proof of Theorem 2 is to rewrite agent  $i$ 's update direction of FedTD(0) as:

$$\mathbf{g}_i(\mathbf{O}_{t;k}^{(i)}) = \mathbf{A}_i(\mathbf{O}_{t;k}^{(i)}) \mathbf{z}_i^{(i)} + \mathbf{b}_i(\mathbf{O}_{t;k}^{(i)})$$

where  $\mathbf{A}_i(\mathbf{O}_{t;k}^{(i)}) = (\mathbf{s}_{t;k}^{(i)} - \mathbf{s}_{t;k+1}^{(i)}) (\mathbf{s}_{t;k}^{(i)} - \mathbf{s}_{t;k+1}^{(i)})^\top$  and  $\mathbf{b}_i(\mathbf{O}_{t;k}^{(i)}) = r(\mathbf{s}_{t;k}^{(i)}) - \mathbf{s}_{t;k}^{(i)}$ . Note that the steady-state value of  $E[\mathbf{b}_i(\mathbf{O}_{t;k}^{(i)})]$  is not equal to 0. For convenience, we apply appropriate centering to rewrite  $\mathbf{g}_i$  as:

$$\mathbf{g}_i(\mathbf{O}_{t;k}^{(i)}) = \mathbf{A}_i(\mathbf{O}_{t;k}^{(i)}) (\mathbf{z}_i^{(i)} - \mathbf{z}_i^{(i)}(\mathbf{O}_{t;k}^{(i)})) + \mathbf{b}_i(\mathbf{O}_{t;k}^{(i)}) + \underbrace{\mathbf{A}_i(\mathbf{O}_{t;k}^{(i)}) \mathbf{z}_i^{(i)}(\mathbf{O}_{t;k}^{(i)})}_{\mathbf{Z}_i(\mathbf{O}_{t;k}^{(i)})}; \quad (32)$$

Define  $\mathbf{Z}_i(\mathbf{O}_{t;k}^{(i)}) = \mathbf{b}_i(\mathbf{O}_{t;k}^{(i)}) + \mathbf{A}_i(\mathbf{O}_{t;k}^{(i)}) \mathbf{z}_i^{(i)}(\mathbf{O}_{t;k}^{(i)})$ . As  $\mathbf{g}_i(\mathbf{O}_{t;k}^{(i)}) = \mathbf{A}_i(\mathbf{O}_{t;k}^{(i)}) (\mathbf{z}_i^{(i)} - \mathbf{z}_i^{(i)}(\mathbf{O}_{t;k}^{(i)})) + \mathbf{Z}_i(\mathbf{O}_{t;k}^{(i)})$ ; we have:

$$\mathbf{g}_i(\mathbf{O}_{t;k}^{(i)}) = \mathbf{A}_i(\mathbf{O}_{t;k}^{(i)}) (\mathbf{z}_i^{(i)} - \mathbf{z}_i^{(i)}(\mathbf{O}_{t;k}^{(i)})) + \mathbf{Z}_i(\mathbf{O}_{t;k}^{(i)}); \quad (33)$$

where  $\mathbf{A}_i = \mathbf{D}^{(i)}(\mathbf{P}^{(i)})$ . Note that  $E_{\mathcal{O}_{t;k}^{(i)}}[\mathbf{Z}_i(\mathbf{O}_{t;k}^{(i)})] = 0$ . 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  $\|\mathbf{A}_i(\mathbf{O}_{t;k}^{(i)})\| \leq c_1$ ,  $\|\mathbf{b}_i(\mathbf{O}_{t;k}^{(i)})\| \leq c_2$ , and  $\|\mathbf{Z}_i(\mathbf{O}_{t;k}^{(i)})\| \leq c_3$  holds for all  $i \in [N]$ :

Proof. Based on the definition and the fact that  $\|s_{t;k}^{(i)}\|_1 \leq 1$ , we have

$$A_i O_{t;k}^{(i)} = (s_{t;k}^{(i)})^\top (s_{t;k}^{(i)} - s_{t;k+1}^{(i)}) + (s_{t;k}^{(i)})^\top s_{t;k+1}^{(i)} \leq 1 + \dots$$

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

$$A_i \leq 1 + \dots \quad \& \quad Z_i O_{t;k}^{(i)} \leq R_{\max} + c_1 H$$

□

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

$$\begin{aligned} A_i &\leq E_{t_1} \sum_{k_1}^h A_i O_{t_2;k_2}^{(i)} \sum_{k_1}^i F_{k_1}^{t_1} \leq L_1 (t_2 - t_1) K + k_2 - k_1; \\ A_i &\leq E_{t_1} \sum_{k_2}^h A_i O_{t_2;k_2}^{(i)} \leq L_1 (t_2 - t_1) K + k_2; \\ E \sum_{k_2}^h Z_i O_{t_2;k_2}^{(i)} \sum_{k_1}^i F_{k_1}^{t_1} &\leq L_2 (t_2 - t_1) K + k_2 - k_1; \\ E_{t_1} \sum_{k_2}^h Z_i O_{t_2;k_2}^{(i)} &\leq L_2 (t_2 - t_1) K + k_2 \end{aligned}$$

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:

$$\begin{aligned} E \sum_{k_2}^h Z_i O_{t_2;k_2}^{(i)} \sum_{k_1}^i F_{k_1}^{t_1} &= E \sum_{k_2}^h Z_i O_{t_2;k_2}^{(i)} \sum_{k_1}^i F_{k_1}^{t_1} E_{O_{t_2;k_2}^{(i)}} \sum_{k_1}^h Z_i O_{t_2;k_2}^{(i)} \sum_{k_1}^i F_{k_1}^{t_1} \\ &= \sum_{s_{t_2;k_2}^{(i)}, s_{t_2+1;k_2+1}^{(i)}}^{(i)} (s_{t_2;k_2}^{(i)})^\top P(s_{t_2+1;k_2+1}^{(i)} | s_{t_2;k_2}^{(i)}) \\ &\quad P(s_{t_2;k_2}^{(i)} = \sum_{k_1} s_{t_1;k_1}^{(i)}) P(s_{t_2+1;k_2+1}^{(i)} | s_{t_2;k_2}^{(i)}) Z_i (O_{t_2;k_2}^{(i)}) \\ &\quad \times \sum_{s_{t_2;k_2}^{(i)}}^{(i)} (s_{t_2;k_2}^{(i)})^\top P(s_{t_2;k_2}^{(i)} = \sum_{k_1} s_{t_1;k_1}^{(i)}) Z_i (O_{t_2;k_2}^{(i)}) \\ (a) \quad &\sum_{s_{t_2;k_2}^{(i)}}^{(i)} (s_{t_2;k_2}^{(i)})^\top P(s_{t_2;k_2}^{(i)} = \sum_{k_1} s_{t_1;k_1}^{(i)}) (R_{\max} + c_1 H) \\ &= 2(R_{\max} + c_1 H) d_{TV} \sum_{s_{t_2;k_2}^{(i)} = \sum_{k_1} s_{t_1;k_1}^{(i)}}^{(i)} P(s_{t_2;k_2}^{(i)} = \sum_{k_1} s_{t_1;k_1}^{(i)}) \\ &\quad 2(R_{\max} + c_1 H) m_i^{(t_2 - t_1)K + k_2 - k_1} \end{aligned}$$

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

$$\begin{aligned} A_i &\leq E_{t_1} \sum_{k_2}^h A_i O_{t_2;k_2}^{(i)} \sum_{k_1}^i F_{k_1}^{t_1} = k E_{t_1} \sum_{k_2}^h A_i O_{t_2;k_2}^{(i)} \sum_{k_1}^i F_{k_1}^{t_1} E_{O_{t_2;k_2}^{(i)}} \sum_{k_1}^h A_i O_{t_2;k_2}^{(i)} \sum_{k_1}^i F_{k_1}^{t_1} \\ &= \sum_{s_{t_2;k_2}^{(i)}, s_{t_2+1;k_2+1}^{(i)}}^{(i)} (s_{t_2;k_2}^{(i)})^\top P(s_{t_2+1;k_2+1}^{(i)} | s_{t_2;k_2}^{(i)}) \\ &\quad P(s_{t_2;k_2}^{(i)} = \sum_{k_1} s_{t_1;k_1}^{(i)}) P(s_{t_2+1;k_2+1}^{(i)} | s_{t_2;k_2}^{(i)}) A_i (O_{t_2;k_2}^{(i)}) \\ &\quad \times \sum_{s_{t_2;k_2}^{(i)}}^{(i)} (s_{t_2;k_2}^{(i)})^\top P(s_{t_2;k_2}^{(i)} = \sum_{k_1} s_{t_1;k_1}^{(i)}) A_i (O_{t_2;k_2}^{(i)}) \end{aligned}$$

$$(b) \quad \mathbb{P} \left[ \sum_{i=1}^N \sum_{k=0}^{K-1} \left| \frac{1}{NK} \sum_{i=1}^N \sum_{k=0}^{K-1} A_i(O_{t;k}^{(i)}) - \mathbb{E} \left[ \frac{1}{NK} \sum_{i=1}^N \sum_{k=0}^{K-1} A_i(O_{t;k}^{(i)}) \right] \right|^2 \right] \leq \frac{d_1}{NK} + 2L_1^2 K;$$

We finish the proof by choosing  $L_1$ ,  $\max_{i \in [N]} \sum_{k=0}^{K-1} \mathbb{E} \left[ \sum_{i=1}^N \sum_{k=0}^{K-1} A_i(O_{t;k}^{(i)}) \right] = 2c_1 m$ . We employ the same reasoning to prove the remaining three inequalities.  $\square$

### 1.2.2 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  $(1/NK) \sum_{i=1}^N \sum_{k=0}^{K-1} A_i(O_{t;k}^{(i)})$  and  $(1/NK) \sum_{i=1}^N \sum_{k=0}^{K-1} b_i(O_{t;k}^{(i)})$  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  $O_{t;k_1}^{(i)}$  and  $O_{t;k_2}^{(i)}$  are correlated at different local steps  $k_1, k_2$ . Due to the geometric mixing property of the Markov chain, the correlation between  $O_{t;k_1}^{(i)}$  and  $O_{t;k_2}^{(i)}$  will geometrically decay after the mixing time. Based on this fact, we show that the variances of  $(1/NK) \sum_{i=1}^N \sum_{k=0}^{K-1} A_i(O_{t;k}^{(i)})$  and  $(1/NK) \sum_{i=1}^N \sum_{k=0}^{K-1} b_i(O_{t;k}^{(i)})$  get scaled down by  $NK$  with an additional additive, higher order term dependent on the mixing time, which is formally stated as follows:

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

$$\mathbb{E}_t \left[ \sum_{i=1}^N \sum_{k=0}^{K-1} \left| \frac{1}{NK} \sum_{i=1}^N \sum_{k=0}^{K-1} A_i(O_{t;k}^{(i)}) - \mathbb{E} \left[ \frac{1}{NK} \sum_{i=1}^N \sum_{k=0}^{K-1} A_i(O_{t;k}^{(i)}) \right] \right|^2 \right] \leq \frac{d_1}{NK} + 2L_1^2 K; \quad (34)$$

$$\mathbb{E}_t \left[ \sum_{i=1}^N \sum_{k=0}^{K-1} \left| \frac{1}{NK} \sum_{i=1}^N \sum_{k=0}^{K-1} A_i(O_{t;k}^{(i)}) - \mathbb{E} \left[ \frac{1}{NK} \sum_{i=1}^N \sum_{k=0}^{K-1} A_i(O_{t;k}^{(i)}) \right] \right|^2 \right] \leq \frac{d_1^2}{NK} + 4L_1^2 K; \quad \text{and} \quad (35)$$

$$\mathbb{E}_t \left[ \sum_{i=1}^N \sum_{k=0}^{K-1} \left| \frac{1}{NK} \sum_{i=1}^N \sum_{k=0}^{K-1} Z_i(O_{t;k}^{(i)}) - \mathbb{E} \left[ \frac{1}{NK} \sum_{i=1}^N \sum_{k=0}^{K-1} Z_i(O_{t;k}^{(i)}) \right] \right|^2 \right] \leq \frac{d_2}{NK} + 2L_2^2 K; \quad (36)$$

$$\mathbb{E}_t \left[ \sum_{i=1}^N \sum_{k=0}^{K-1} \left| \frac{1}{NK} \sum_{i=1}^N \sum_{k=0}^{K-1} Z_i(O_{t;k}^{(i)}) - \mathbb{E} \left[ \frac{1}{NK} \sum_{i=1}^N \sum_{k=0}^{K-1} Z_i(O_{t;k}^{(i)}) \right] \right|^2 \right] \leq \frac{d_2^2}{NK} + 4L_2^2 K; \quad (37)$$

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

Proof.

$$\begin{aligned} & \mathbb{E}_t \left[ \sum_{i=1}^N \sum_{k=0}^{K-1} \left| \frac{1}{NK} \sum_{i=1}^N \sum_{k=0}^{K-1} Z_i(O_{t;k}^{(i)}) - \mathbb{E} \left[ \frac{1}{NK} \sum_{i=1}^N \sum_{k=0}^{K-1} Z_i(O_{t;k}^{(i)}) \right] \right|^2 \right] \\ &= \mathbb{E}_t \left[ \frac{1}{NK} \sum_{i=1}^N \sum_{k=0}^{K-1} \left| \frac{1}{NK} \sum_{i=1}^N \sum_{k=0}^{K-1} Z_i(O_{t;k}^{(i)}) - \mathbb{E} \left[ \frac{1}{NK} \sum_{i=1}^N \sum_{k=0}^{K-1} Z_i(O_{t;k}^{(i)}) \right] \right|^2 \right] \\ &= \mathbb{E}_t \left[ \frac{1}{NK} \sum_{i=1}^N \sum_{k=0}^{K-1} \left| \frac{1}{NK} \sum_{i=1}^N \sum_{k=0}^{K-1} Z_i(O_{t;k}^{(i)}) - \mathbb{E} \left[ \frac{1}{NK} \sum_{i=1}^N \sum_{k=0}^{K-1} Z_i(O_{t;k}^{(i)}) \right] \right|^2 \right] \end{aligned}$$

(concavity of square root and Jensen's inequality)

$$\begin{aligned}
 &= \mathbb{E}_t \left[ \frac{1}{4N^2K^2} \sum_{i=1}^N \sum_{k=0}^{K-1} Z_i(O_{t;k}^{(i)}) > Z_i(O_{t;l}^{(i)}) + \frac{2}{N^2K^2} \sum_{i=1}^N \sum_{k<l} Z_i(O_{t;k}^{(i)}) > Z_i(O_{t;l}^{(i)}) \right] \\
 &+ \frac{2}{N^2K^2} \sum_{i<j} \sum_{k=0}^{K-1} Z_i(O_{t;k}^{(i)}) > Z_j(O_{t;k}^{(j)}) + \frac{2}{N^2K^2} \sum_{i<j} \sum_{k<l} Z_i(O_{t;k}^{(i)}) > Z_j(O_{t;l}^{(j)}) \tag{38}
 \end{aligned}$$

where  $T_1$  can be further bounded by:

$$\begin{aligned}
 \mathbb{E}_t [T_1] &= \mathbb{E}_t \left[ \frac{2}{N^2K^2} \sum_{i=1}^N \sum_{k<l} Z_i(O_{t;k}^{(i)}) > Z_i(O_{t;l}^{(i)}) \right] \\
 &= \mathbb{E}_t \left[ \frac{2}{N^2K^2} \sum_{i=1}^N \sum_{k<l} Z_i(O_{t;k}^{(i)}) > \mathbb{E} [Z_i(O_{t;l}^{(i)})] F_k^t \right] \\
 &= \mathbb{E}_t \left[ \frac{2}{N^2K^2} \sum_{i=1}^N \sum_{k<l} Z_i(O_{t;k}^{(i)}) \mathbb{E} [Z_i(O_{t;l}^{(i)})] F_k^t \right] \\
 &\quad \text{(Cauchy Schwarz inequality)} \\
 &= \mathbb{E}_t \left[ \frac{2}{N^2K^2} \sum_{i=1}^N \sum_{k<l} c_3 L_2^{(i,k)} \right] \quad \text{(Lemma 11 and 12)} \\
 &= \mathbb{E}_t \left[ \frac{2}{N^2K^2} \sum_{i=1}^N \sum_{k=0}^{K-1} \sum_{m=1}^m c_3 L_2^m \right] \\
 &= \frac{2c_3 L_2 N K}{N^2 K^2} = \frac{2c_3 L_2}{NK}.
 \end{aligned}$$

And  $T_2$  can be bounded by:

$$\begin{aligned}
 \mathbb{E}_t [T_2] &= \frac{2}{N^2K^2} \sum_{i<j} \sum_{k=0}^{K-1} \mathbb{E}_t [Z_i(O_{t;k}^{(i)}) > Z_j(O_{t;k}^{(j)})] \quad (O_{t;k}^{(i)} \text{ and } O_{t;k}^{(j)} \text{ are independent}) \\
 &= \frac{2}{N^2K^2} \sum_{i<j} \sum_{k=0}^{K-1} L_2^{2K+2k} \quad \text{(Lemma 12)} \\
 &= \frac{2}{K} L_2^{2K} :
 \end{aligned}$$

Meanwhile,  $T_3$  can be bounded by:

$$\begin{aligned}
 \mathbb{E}_t [T_3] &= \frac{2}{N^2K^2} \sum_{i<j} \sum_{k<l} \mathbb{E}_t [Z_i(O_{t;k}^{(i)}) > Z_j(O_{t;l}^{(j)})] \quad (O_{t;k}^{(i)} \text{ and } O_{t;l}^{(j)} \text{ are independent}) \\
 &= \frac{2}{N^2K^2} \sum_{i<j} \sum_{k<l} L_2^{2K+k+l} \quad \text{(Lemma 12)} \\
 &= 2L_2^{2K}
 \end{aligned}$$

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

$$\mathbb{E}_t \left[ \frac{1}{NK} \sum_{i=1}^N \sum_{k=0}^{K-1} Z_i(O_{t;k}^{(i)}) \right]$$



$$\begin{aligned}
 & \frac{1}{N^2 K^2} \sum_{i=1}^N \sum_{k=0}^{K-1} \mathbb{E}_t \left[ \sum_{h=1}^H Z_i(O_{t;k}^{(i)}) > Z_i(O_{t;k}^{(i)})^i + \frac{2c_3 L_2}{NK(1-\gamma)} + \frac{2}{K} L_2^2 \gamma^{2K} + 2L_2^2 \gamma^{2K} \right] \\
 (a) \quad & \frac{NK}{N^2 K^2} c_3^2 + \frac{2c_3 L_2}{NK(1-\gamma)} + \frac{2}{K} L_2^2 \gamma^{2K} + 2L_2^2 \gamma^{2K} \\
 & \frac{1}{NK} \left( c_3^2 + \frac{2c_3 L_2}{1-\gamma} + 4L_2^2 \gamma^{2K} \right) (K-1) \\
 & \frac{1}{NK} \left( c_3^2 + \frac{2c_3 L_2}{1-\gamma} + 4L_2^2 \gamma^{2K} \right) = \frac{1}{NK} \left( c_3^2 + \frac{2c_3 L_2}{1-\gamma} + 2L_2^2 \gamma^{2K} \right) :
 \end{aligned}$$

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

### I.2.3 Bounding $\mathbb{E} \left[ \sum_{t=2}^T \sum_{i=1}^N \sum_{k=0}^{K-1} \right]$

Lemma 14. (Bounding  $\sum_{t=2}^T \sum_{i=1}^N \sum_{k=0}^{K-1} \mathbb{E} \left[ \sum_{h=1}^H Z_i(O_{t;k}^{(i)}) > Z_i(O_{t;k}^{(i)})^i \right]$ ) Consider  $\gamma = d \frac{\text{mix}(\frac{\gamma}{T})}{K}$  and choose the effective step-size

$$\min \left\{ \frac{1}{30c_4(1-\gamma)}, \frac{1}{96c_4^2}, 1 \right\}$$

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

$$\begin{aligned}
 \mathbb{E}_t \left[ \sum_{i=1}^N \sum_{k=0}^{K-1} \sum_{h=1}^H Z_i(O_{t;k}^{(i)}) > Z_i(O_{t;k}^{(i)})^i \right] & \leq 8 \gamma^2 c_4^2 \mathbb{E}_t \left[ \sum_{i=1}^N \sum_{k=0}^{K-1} \sum_{h=1}^H Z_i(O_{t;k}^{(i)}) > Z_i(O_{t;k}^{(i)})^i \right] + 14 \gamma^2 \frac{d^2}{NK} + \frac{52L_2^2 \gamma^4}{1-\gamma^2} \\
 & + 4 \gamma^2 c_4^2 \sum_{s=0}^{t-2} \mathbb{E}_t \left[ \sum_{i=1}^N \sum_{k=0}^{K-1} \sum_{h=1}^H Z_i(O_{t-s;k}^{(i)}) > Z_i(O_{t-s;k}^{(i)})^i \right] + 3200 \gamma^2 c_4^2 \gamma^{2(t-2)} (1-\gamma) + 4 \gamma^2 c_1^2 \gamma^{2(t-2)} (1-\gamma) : \quad (39)
 \end{aligned}$$

Proof. For any  $t \geq 2$ , we have

$$\begin{aligned}
 & \sum_{i=1}^N \sum_{k=0}^{K-1} \sum_{h=1}^H Z_i(O_{t;k}^{(i)}) > Z_i(O_{t;k}^{(i)})^i \\
 & = \sum_{i=1}^N \sum_{k=0}^{K-1} \sum_{h=1}^H \left[ \frac{1}{NK} \sum_{i=1}^N \sum_{k=0}^{K-1} \sum_{h=1}^H g_i(O_{t;k}^{(i)}) \right] \\
 & = \sum_{i=1}^N \sum_{k=0}^{K-1} \sum_{h=1}^H \left[ \frac{1}{NK} \sum_{i=1}^N \sum_{k=0}^{K-1} \sum_{h=1}^H A_i(O_{t;k}^{(i)}) \right] + Z_i(O_{t;k}^{(i)})^i \\
 & = \sum_{i=1}^N \sum_{k=0}^{K-1} \sum_{h=1}^H \left[ \frac{1}{NK} \sum_{i=1}^N \sum_{k=0}^{K-1} \sum_{h=1}^H A_i(O_{t;k}^{(i)}) \right] + Z_i(O_{t;k}^{(i)})^i \\
 & \quad + 2 \sum_{i=1}^N \sum_{k=0}^{K-1} \sum_{h=1}^H \left[ \frac{1}{NK} \sum_{i=1}^N \sum_{k=0}^{K-1} \sum_{h=1}^H A_i(O_{t;k}^{(i)}) \right] + Z_i(O_{t;k}^{(i)})^i \\
 (a) \quad & = 2 \sum_{i=1}^N \sum_{k=0}^{K-1} \sum_{h=1}^H \left[ \frac{1}{NK} \sum_{i=1}^N \sum_{k=0}^{K-1} \sum_{h=1}^H A_i(O_{t;k}^{(i)}) \right] + Z_i(O_{t;k}^{(i)})^i + 2 \gamma^2 c_1^2 \gamma^{2(t-2)} (1-\gamma)
 \end{aligned}$$

$$\begin{aligned}
 &= 6 \sum_{i=1}^L \sum_{k=0}^{K-1} \frac{1}{NK} \mathcal{X}^i \mathcal{X}^{K-1-k} A_i(O_{l;k}^{(i)}) + 6 \sum_{i=1}^L \sum_{k=0}^{K-1} \frac{1}{NK} \mathcal{X}^i \mathcal{X}^{K-1-k} A_i(O_{l;k}^{(i)}) \\
 &+ 6 \sum_{i=1}^L \sum_{k=0}^{K-1} \frac{1}{NK} \mathcal{X}^i \mathcal{X}^{K-1-k} Z_i(O_{l;k}^{(i)}) + 2 \sum_{i=1}^L \sum_{k=0}^{K-1} c_1^2 \mathcal{X}^i \mathcal{X}^{K-1-k} \\
 &+ 6 \sum_{i=1}^L \sum_{k=0}^{K-1} \frac{c_1}{NK} \mathcal{X}^i \mathcal{X}^{K-1-k} Z_i(O_{l;k}^{(i)}) + 6 \sum_{i=1}^L \sum_{k=0}^{K-1} \frac{1}{NK} \mathcal{X}^i \mathcal{X}^{K-1-k} Z_i(O_{l;k}^{(i)}) + 2 \sum_{i=1}^L \sum_{k=0}^{K-1} c_1^2 \mathcal{X}^i \mathcal{X}^{K-1-k}; \quad (40)
 \end{aligned}$$

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

$$\begin{aligned}
 &3 \sum_{i=1}^{L+1} \sum_{k=0}^{K-1} \frac{1}{NK} \mathcal{X}^i \mathcal{X}^{K-1-k} A_i(O_{l;k}^{(i)}) + 3 \sum_{i=1}^L \sum_{k=0}^{K-1} c_1^2 \mathcal{X}^i \mathcal{X}^{K-1-k} \\
 &+ 3 \sum_{i=1}^L \sum_{k=0}^{K-1} \frac{1}{NK} \mathcal{X}^i \mathcal{X}^{K-1-k} Z_i(O_{l;k}^{(i)}) + 2 \sum_{i=1}^L \sum_{k=0}^{K-1} c_1^2 \mathcal{X}^i \mathcal{X}^{K-1-k} \\
 &\frac{3c_1}{NK} \sum_{i=1}^L \sum_{k=0}^{K-1} \mathcal{X}^i \mathcal{X}^{K-1-k} A_i(O_{l;k}^{(i)}) + 3c_1 \sum_{i=1}^L \sum_{k=0}^{K-1} \mathcal{X}^i \mathcal{X}^{K-1-k} + 3 \sum_{i=1}^L \sum_{k=0}^{K-1} \frac{1}{NK} \mathcal{X}^i \mathcal{X}^{K-1-k} Z_i(O_{l;k}^{(i)}) + 2c_1 \sum_{i=1}^L \sum_{k=0}^{K-1} \mathcal{X}^i \mathcal{X}^{K-1-k}; \quad (41)
 \end{aligned}$$

By using the fact that  $\sum_{i=1}^{L+1} \sum_{k=0}^{K-1} \mathcal{X}^i \mathcal{X}^{K-1-k} + \sum_{i=1}^L \sum_{k=0}^{K-1} \mathcal{X}^i \mathcal{X}^{K-1-k}$ , we have:

$$\begin{aligned}
 &\sum_{i=1}^{L+1} \sum_{k=0}^{K-1} (1 + 3c_1) \mathcal{X}^i \mathcal{X}^{K-1-k} + \frac{3c_1}{NK} \sum_{i=1}^L \sum_{k=0}^{K-1} \mathcal{X}^i \mathcal{X}^{K-1-k} A_i(O_{l;k}^{(i)}) \\
 &+ 3 \sum_{i=1}^L \sum_{k=0}^{K-1} \frac{1}{NK} \mathcal{X}^i \mathcal{X}^{K-1-k} Z_i(O_{l;k}^{(i)}) + 2c_1 \sum_{i=1}^L \sum_{k=0}^{K-1} \mathcal{X}^i \mathcal{X}^{K-1-k}; \quad (42)
 \end{aligned}$$

For simplicity, we define  $c_4$ ,  $3c_1$  and  $\mathcal{H}_1$ ,  $\frac{1}{NK} \sum_{i=1}^L \sum_{k=0}^{K-1} \mathcal{X}^i \mathcal{X}^{K-1-k} A_i(O_{l;k}^{(i)})$ . Taking the square on both sides of Eq (42), we have:

$$\begin{aligned}
 &\sum_{i=1}^{L+1} \sum_{k=0}^{K-1} (1 + c_4)^2 \mathcal{X}^i \mathcal{X}^{K-1-k} + 2 \sum_{i=1}^L \sum_{k=0}^{K-1} c_4^2 \mathcal{X}^i \mathcal{X}^{K-1-k} + 9 \sum_{i=1}^L \sum_{k=0}^{K-1} \frac{1}{NK} \mathcal{X}^i \mathcal{X}^{K-1-k} Z_i(O_{l;k}^{(i)}) + 4 \sum_{i=1}^L \sum_{k=0}^{K-1} c_1^2 \mathcal{X}^i \mathcal{X}^{K-1-k} \\
 &+ 6 \sum_{i=1}^L \sum_{k=0}^{K-1} (1 + c_4) \mathcal{X}^i \mathcal{X}^{K-1-k} \left\{ \frac{1}{NK} \sum_{i=1}^L \sum_{k=0}^{K-1} Z_i(O_{l;k}^{(i)}) \right\} + 2c_4(1 + c_4) \sum_{i=1}^L \sum_{k=0}^{K-1} \mathcal{X}^i \mathcal{X}^{K-1-k} \\
 &+ 6 \sum_{i=1}^L \sum_{k=0}^{K-1} c_4 \mathcal{X}^i \mathcal{X}^{K-1-k} \left\{ \frac{1}{NK} \sum_{i=1}^L \sum_{k=0}^{K-1} Z_i(O_{l;k}^{(i)}) \right\} + 4 \sum_{i=1}^L \sum_{k=0}^{K-1} c_1 c_4 \mathcal{X}^i \mathcal{X}^{K-1-k} \left\{ \frac{1}{NK} \sum_{i=1}^L \sum_{k=0}^{K-1} Z_i(O_{l;k}^{(i)}) \right\} \\
 &\quad \mathcal{H}_1 \quad \mathcal{H}_2 \quad \mathcal{H}_3 \quad \mathcal{H}_4
 \end{aligned}$$

$$+ 4 c_1(1 + c_4) \frac{1}{H_5} \left( \sum_{i=1}^N \sum_{k=0}^{K-1} Z_i(O_{l;k}^{(i)}) \right) + 12 c_1 \frac{1}{NK} \sum_{i=1}^N \sum_{k=0}^{K-1} Z_i(O_{l;k}^{(i)}) \quad (43)$$

We can further bound  $H_1$  as:

$$\begin{aligned} H_1 &= 6(1 + c_4) \frac{1}{NK} \sum_{i=1}^N \sum_{k=0}^{K-1} Z_i(O_{l;k}^{(i)}) \\ &= 2^p \frac{1}{3(1 + c_4)} \frac{1}{NK} \sum_{i=1}^N \sum_{k=0}^{K-1} Z_i(O_{l;k}^{(i)}) \\ &\quad + 3(1 + c_4) \frac{1}{NK} \sum_{i=1}^N \sum_{k=0}^{K-1} Z_i(O_{l;k}^{(i)})^2 \\ &\quad + 6 \frac{1}{NK} \sum_{i=1}^N \sum_{k=0}^{K-1} Z_i(O_{l;k}^{(i)})^2 \end{aligned} \quad (44)$$

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

$$H_2 = 2 c_4(1 + c_4) \frac{1}{NK} \sum_{i=1}^N \sum_{k=0}^{K-1} Z_i(O_{l;k}^{(i)})^2 + 2 c_4^2 \frac{1}{NK} \sum_{i=1}^N \sum_{k=0}^{K-1} Z_i(O_{l;k}^{(i)})^2 \quad (45)$$

And we bound  $H_3$  as:

$$H_3 = 6 c_4 \frac{1}{NK} \sum_{i=1}^N \sum_{k=0}^{K-1} Z_i(O_{l;k}^{(i)}) + 3 \frac{1}{NK} \sum_{i=1}^N \sum_{k=0}^{K-1} Z_i(O_{l;k}^{(i)})^2 + 3 c_4^2 \frac{1}{NK} \sum_{i=1}^N \sum_{k=0}^{K-1} Z_i(O_{l;k}^{(i)})^2 \quad (46)$$

For  $H_4; H_5; H_6$ , we have:

$$H_4 = 4 c_1 c_4 \frac{1}{NK} \sum_{i=1}^N \sum_{k=0}^{K-1} Z_i(O_{l;k}^{(i)}) + 2 c_4^2 \frac{1}{NK} \sum_{i=1}^N \sum_{k=0}^{K-1} Z_i(O_{l;k}^{(i)})^2 + 2 c_1^2 \frac{1}{NK} \sum_{i=1}^N \sum_{k=0}^{K-1} Z_i(O_{l;k}^{(i)})^2;$$

$$H_5 = 4 c_1(1 + c_4) \frac{1}{NK} \sum_{i=1}^N \sum_{k=0}^{K-1} Z_i(O_{l;k}^{(i)}) + 4 \frac{1}{NK} \sum_{i=1}^N \sum_{k=0}^{K-1} Z_i(O_{l;k}^{(i)})^2 + 4 c_1^2 \frac{1}{NK} \sum_{i=1}^N \sum_{k=0}^{K-1} Z_i(O_{l;k}^{(i)})^2;$$

$$H_6 = 12 c_1 \frac{1}{NK} \sum_{i=1}^N \sum_{k=0}^{K-1} Z_i(O_{l;k}^{(i)}) + 6 \frac{1}{NK} \sum_{i=1}^N \sum_{k=0}^{K-1} Z_i(O_{l;k}^{(i)})^2 + 6 c_1^2 \frac{1}{NK} \sum_{i=1}^N \sum_{k=0}^{K-1} Z_i(O_{l;k}^{(i)})^2;$$

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

$$\begin{aligned} & \frac{1}{NK} \sum_{i=1}^N \sum_{k=0}^{K-1} Z_i(O_{l;k}^{(i)})^2 \\ & (1 + (3c_4 + 12)) \frac{1}{NK} \sum_{i=1}^N \sum_{k=0}^{K-1} Z_i(O_{l;k}^{(i)})^2 + (6^2 + 2) c_4^2 \frac{1}{NK} \sum_{i=1}^N \sum_{k=0}^{K-1} Z_i(O_{l;k}^{(i)})^2 \\ & + (18^2 + 6) \frac{1}{NK} \sum_{i=1}^N \sum_{k=0}^{K-1} Z_i(O_{l;k}^{(i)})^2 + (12^2 + 4) c_1^2 \frac{1}{NK} \sum_{i=1}^N \sum_{k=0}^{K-1} Z_i(O_{l;k}^{(i)})^2 \\ & (1 + h_1) \frac{1}{NK} \sum_{i=1}^N \sum_{k=0}^{K-1} Z_i(O_{l;k}^{(i)})^2 + 8 c_4^2 \frac{1}{NK} \sum_{i=1}^N \sum_{k=0}^{K-1} Z_i(O_{l;k}^{(i)})^2 + 24 \frac{1}{NK} \sum_{i=1}^N \sum_{k=0}^{K-1} Z_i(O_{l;k}^{(i)})^2 + 16 c_1^2 \frac{1}{NK} \sum_{i=1}^N \sum_{k=0}^{K-1} Z_i(O_{l;k}^{(i)})^2 \end{aligned} \quad (47)$$

where we denote  $h_1 = 3c_4 + 12$  for simplicity. For any  $t \geq 1$ , conditioning on  $\mathcal{F}_{t-2}$  on both sides of the above inequality, we have:

$$E_{t-2} \left[ \frac{1}{NK} \sum_{i=1}^N \sum_{k=0}^{K-1} Z_i(O_{l;k}^{(i)})^2 \right]$$

$$\begin{aligned}
 & (1 + h_1) E_{t-2}^2 + 24 E_{t-2} \frac{1}{NK} \sum_{i=1}^N \sum_{k=0}^{K-1} Z_i(O_{i;k}^{(i)})^2 \\
 & + 8 c_4^2 E_{t-2}^2 + M_3(\cdot; \cdot) \\
 & (1 + h_1) E_{t-2}^2 + 24 \frac{d_2^2}{NK} + 4L_2^2 2^{2(l-t+2)K} \quad (\text{Lemma 13}) \\
 & + 8 c_4^2 E_{t-2}^2 + M_3(\cdot; \cdot) \\
 & \stackrel{(a)}{(1 + h_1) E_{t-2}^2 + 24 \frac{d_2^2}{NK} + 4L_2^2 2^{2(l-t+2)K}} \\
 & + 8 c_4^2 E_{t-2}^2 + M_3(\cdot; \cdot) \\
 & (1 + h_1) E_{t-2}^2 + c_t(l) + 8 c_4^2 E_{t-2}^2 + M_3(\cdot; \cdot); \tag{48}
 \end{aligned}$$

where we denote  $M_3(\cdot; \cdot)$ ,  $16c_4^2 2^{2(l-t+2)K}$  and  $c_t(l) = 24 \frac{d_2^2}{NK} + 4L_2^2 2^{2(l-t+2)K}$  for simplicity. Inequality (a) is due to  $2^K \frac{1}{t} \frac{1}{t}$ . In the following steps, we try to map  $E_{t-2}^2$  to  $E_{t-1}^2$  for any  $t \leq l$ . By applying Eq (48) recursively, we have:

$$\begin{aligned}
 & E_{t-2}^2 \\
 & (1 + h_1)^{l+1-t} E_{t-2}^2 + \sum_{k=t}^l (1 + h_1)^{l-k} (c_t(k) + M_3(\cdot; \cdot)) \\
 & + 8 c_4^2 E_{t-2}^2 \sum_{k=t}^l (1 + h_1)^{l-k} \\
 & \stackrel{(b)}{(1 + h_1)^{l+1-t} E_{t-2}^2 + \sum_{k=t}^l (1 + h_1)^{l-k} (c_t(k) + M_3(\cdot; \cdot))} \\
 & + 8 c_4^2 E_{t-2}^2 \sum_{k=t}^l (1 + h_1)^{l-k} \tag{49}
 \end{aligned}$$

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

$$\begin{aligned}
 & H_7 \sum_{k=t}^l (1 + h_1)^{l-k} (c_t(k) + M_3(\cdot; \cdot)) \\
 & = \sum_{k^0=0}^X (1 + h_1)^{k^0} (c_t(k^0 + t) + M_3(\cdot; \cdot)) \\
 & \quad (\text{changing index } k \text{ into } k^0 \text{ with } k^0 = k + t) \\
 & = 24 \sum_{k^0=0}^X (1 + h_1)^{k^0} \frac{d_2^2}{NK} + 4L_2^2 2^{2k^0K} + M_3(\cdot; \cdot) \\
 & \quad (\text{Substituting the definition of } c_t(k^0) \text{ inside}) \\
 & = 24 \frac{d_2^2}{NK} + M_3(\cdot; \cdot) \sum_{k^0=0}^X \frac{(1 + h_1)^{l+1-k^0}}{h_1} + 4L_2^2 2^{2(l-k^0)K} \sum_{k^0=0}^X \frac{2K}{1 + h_1} \\
 & = 24 \frac{d_2^2}{NK} + M_3(\cdot; \cdot) \frac{(1 + h_1)^{l+1}}{h_1} + 4L_2^2 2^{2(l+1)K} \sum_{k^0=0}^X \frac{2K}{(1 + h_1)^{k^0}}
 \end{aligned}$$

$$24 \frac{d_2^2}{NK} + M_3(\cdot; \cdot) \frac{(1+h_1)^{+1}}{h_1} + 4L_2^2 (1+h_1) \frac{1}{1} \frac{1}{2} : \quad \#$$

Here we follow the analysis in (Khodadadian et al., 2022). Notice that for  $\frac{\log 2}{h_1}$ , we have  $(1+x)^{+1} \leq 1+2x(1+x)$ : If  $\frac{1}{4h_1} \leq \frac{\log 2}{h_1}$  and  $\frac{1}{2h_1(1+x)}$ , we have  $(1+h_1)^{+1} \leq 1+2h_1(1+x)$  and  $(1+h_1) \leq 1+2h_1$ . Hence, we have

$$H_7 \leq 24 \frac{d_2^2}{NK} + M_3(\cdot; \cdot) (2(1+x) + \frac{8L_2^2}{1} \frac{1}{2}) :$$

We apply the similar analysis to bound  $H_8$  as:

$$H_8 = \sum_{k=0}^X (1+h_1)^k \frac{1}{t} + \sum_{k=0}^X (1+h_1)^k \frac{1}{t} + \sum_{k=0}^X (1+2h_1)^k \frac{1}{t} + 2 \sum_{k=0}^X \frac{1}{t} :$$

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

$$E_{t-1} \leq 2E_t + 24 \frac{d_2^2}{NK} + M_3(\cdot; \cdot) (2(1+x) + \frac{8L_2^2}{1} \frac{1}{2}) + 16c_4^2 \sum_{k=0}^X E_t \left[ \frac{1}{t} \right] :$$

Then it is straightforward to bound  $E_{t-1}$  as:

$$E_{t-1} \leq 2E_t + 24 \frac{d_2^2}{NK} + M_3(\cdot; \cdot) (4 + \frac{8L_2^2}{1} \frac{1}{2}) + 16c_4^2 \sum_{k=0}^X E_t \left[ \frac{1}{t} \right] : \quad (50)$$

Furthermore, based on the triangle inequality, we have:

$$\sum_{s=t}^X \sum_{k=0}^X \frac{1}{t} \leq \sum_{s=t}^X \sum_{k=0}^X \frac{1}{t} + \sum_{s=t}^X \sum_{k=0}^X \frac{1}{t} + 6 \sum_{i=1}^X \sum_{k=0}^X \frac{1}{NK} Z_i(O_{s;k}^{(i)}) + 2 \sum_{i=1}^X c_1^2 \frac{1}{t} : \quad \#$$

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

$$E_{t-1} \leq 2E_t + \sum_{s=t}^X \sum_{k=0}^X \frac{1}{t} + \sum_{s=t}^X \sum_{k=0}^X \frac{1}{t} + 6 \sum_{i=1}^X \sum_{k=0}^X \frac{1}{NK} Z_i(O_{s;k}^{(i)}) + 2 \sum_{i=1}^X c_1^2 \frac{1}{t} + \sum_{s=t}^X \sum_{k=0}^X \frac{1}{t} :$$

$$\begin{aligned}
 & + 24 \frac{d_2^2}{NK} + M_3(\cdot; \cdot) + 4 + \frac{8L_2^2}{1} \frac{2}{2} + 16 c_4^2 \sum_{k=0}^X E_{t-2} [ \binom{2}{t-k} ]^i \quad (\text{Eq (50)}) \\
 & + 6 \sum_{s=t}^X \frac{d_2^2}{NK} + 4L_2^2 2^{(s-t+2)K} \\
 & + 2c_4^2 \sum_{s=t}^X E_{t-2} [ \binom{2}{s} ] + 2 \sum_{s=t}^X c_1^2 2^{2(s-t)} \quad (\text{Lemma 13}) \\
 \text{(a)} \quad & 2 \sum_{s=t}^X c_4^2 2E_{t-2} + 96 \frac{d_2^2}{NK} + \frac{2L_2^2}{1} \frac{3}{2} + 6 \sum_{s=t}^X \frac{d_2^2}{NK} + \frac{4L_2^2}{1} \frac{2}{2K} \\
 & + 2c_4^2 (1 + 16 c_4^2) \sum_{s=0}^X E_{t-2} [ \binom{2}{t-s} ] + 96 \sum_{s=0}^X c_4^2 M_3(\cdot; \cdot) + 2 \sum_{s=0}^X c_1^2 2^{2(s-t)} \quad (\text{51}) \\
 \text{(b)} \quad & 2 \sum_{s=t}^X c_4^2 E_{t-2} + \frac{d_2^2}{NK} 2 \sum_{s=t}^X 96 c_4^2 + 6 + \frac{12L_2^2}{1} \frac{4}{2} 16 c_4^2 + 2 \\
 & + 2c_4^2 (1 + 16 c_4^2) \sum_{s=0}^X E_{t-2} [ \binom{2}{t-s} ] + 96 \sum_{s=0}^X c_4^2 M_3(\cdot; \cdot) + 2 \sum_{s=0}^X c_1^2 2^{2(s-t)} \quad (\text{51})
 \end{aligned}$$

where we used the fact that  $2^{2K} \geq 2$  for (a) and (b), and that  $\sum_{t=0}^n \binom{2}{t} \leq 2^n$  (via Jensens' inequality) for all  $t \geq 0$  in the last inequality. Let us choose  $n$  such that  $96 c_4^2 + 6 \geq 7, 16 c_4^2 + 2 \geq \frac{13}{6}$  and  $1 + 16 c_4^2 \geq 2$ , this holds when

$$\min \left\{ \frac{1}{96 c_4^2}, \frac{1}{96 c_4^2}, \frac{1}{16 c_4^2}, 1 \right\} :$$

Based on the fact that  $k \leq k^2, 2k \leq k^2 + 2k, k^2 \leq k^2$  and the requirement on  $n$ , we have

$$\begin{aligned}
 2 \sum_{s=t}^X c_4^2 E_{t-2} k & \leq k^2 + 4 \sum_{s=t}^X c_4^2 E_{t-2} k + k^2 + 4 \sum_{s=t}^X c_4^2 E_{t-2} k + k^2 \\
 & \leq 0.5 E_{t-2} k + k^2 + 4 \sum_{s=t}^X c_4^2 E_{t-2} k + k^2 + (4 \sum_{s=t}^X c_4^2 + 0.5) \\
 \text{(a)} \quad & 2 \sum_{s=t}^X c_4^2 E_{t-2} + \frac{7d_2^2}{2NK} 2 \sum_{s=t}^X + \frac{13L_2^2}{(1-2)} \frac{4}{2} \\
 & + 2c_4^2 \sum_{s=0}^X E_{t-2} [ \binom{2}{t-s} ] + 48 \sum_{s=0}^X c_4^2 M_3(\cdot; \cdot) + 2 \sum_{s=0}^X c_1^2 2^{2(s-t)} \\
 & + 4 \sum_{s=t}^X c_4^2 E_{t-2} k + k^2 \quad (\text{52})
 \end{aligned}$$

where (a) is due to Eq(51) and the choice of  $n$ . Putting the term  $2 \sum_{s=t}^X c_4^2 E_{t-2} k + k^2$  together by rearranging the terms, we have:

$$\begin{aligned}
 2 \sum_{s=t}^X c_4^2 E_{t-2} k + k^2 & \leq k^2 + \frac{7d_2^2}{2NK} 2 \sum_{s=t}^X + \frac{13L_2^2}{(1-2)} \frac{4}{2} \\
 & + 2c_4^2 \sum_{s=0}^X E_{t-2} [ \binom{2}{t-s} ] + 48 \sum_{s=0}^X c_4^2 M_3(\cdot; \cdot) + 2 \sum_{s=0}^X c_1^2 2^{2(s-t)} \\
 & + 4 \sum_{s=t}^X c_4^2 E_{t-2} k + k^2 \quad (\text{53})
 \end{aligned}$$

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

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

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

□

## I.2.4 Drift Term Analysis.

Now we bound the drift term as follows:

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

$$E[\Delta_t] = \frac{1}{NK} \sum_{i=1}^N \sum_{k=0}^{K-1} E_{t,k}^{(i)} \|\Delta_t\|^2 \leq \frac{4}{K} \frac{2}{9} c_3^2 + \frac{2c_3 L_2}{1} + 8c_1^2 (K-1)H^2 : \quad (54)$$

Proof.

$$\begin{aligned} \frac{1}{NK} \sum_{i=1}^N \sum_{k=0}^{K-1} E_{t,k}^{(i)} \|\Delta_t\|^2 &= \frac{1}{NK} \sum_{i=1}^N \sum_{k=0}^{K-1} E_{t,k}^{(i)} \|\Delta_t + \sum_{s=0}^{k-1} g_{t,s}^{(i)}\|^2 \\ &= \frac{2}{NK} \sum_{i=1}^N \sum_{k=0}^{K-1} E_{t,k}^{(i)} \sum_{s=0}^{k-1} \|A_i(O_{t,s}^{(i)}) - Z_i(O_{t,s}^{(i)})\|^2 \\ &\quad + 2 \sum_{i=1}^N \sum_{k=0}^{K-1} \frac{1}{NK} E_{t,k}^{(i)} \sum_{s=0}^{k-1} \|A_i(O_{t,s}^{(i)}) - Z_i(O_{t,s}^{(i)})\|^2 \\ &\quad + 2 \sum_{i=1}^N \sum_{k=0}^{K-1} \frac{1}{NK} \sum_{s=0}^{k-1} E_{t,k}^{(i)} \|A_i(O_{t,s}^{(i)}) - Z_i(O_{t,s}^{(i)})\|^2 \\ &\quad + 2 \sum_{i=1}^N \sum_{k=0}^{K-1} \frac{1}{NK} \sum_{s=0}^{k-1} E_{t,k}^{(i)} \|Z_i(O_{t,s}^{(i)})\|^2 \\ &\quad + 2 \sum_{i=1}^N \sum_{k=0}^{K-1} \frac{1}{NK} \sum_{s=0}^{k-1} E_{t,k}^{(i)} \|Z_i(O_{t,s}^{(i)}) - Z_i(O_{t,s^0}^{(i)})\|^2 \\ &\quad + 2 \sum_{i=1}^N \sum_{k=0}^{K-1} \frac{1}{NK} \sum_{s=0}^{k-1} E_{t,k}^{(i)} \|Z_i(O_{t,s^0}^{(i)})\|^2 \\ &\quad + 2 \sum_{i=1}^N \sum_{k=0}^{K-1} \frac{1}{NK} \sum_{s=0}^{k-1} E_{t,k}^{(i)} \|Z_i(O_{t,s}^{(i)}) - Z_i(O_{t,s^0}^{(i)})\|^2 \\ &\quad + 2 \sum_{i=1}^N \sum_{k=0}^{K-1} \frac{1}{NK} \sum_{s=0}^{k-1} E_{t,k}^{(i)} \|Z_i(O_{t,s^0}^{(i)})\|^2 \\ &\quad + 2 \sum_{i=1}^N \sum_{k=0}^{K-1} \frac{1}{NK} \sum_{s=0}^{k-1} E_{t,k}^{(i)} \|Z_i(O_{t,s}^{(i)}) - Z_i(O_{t,s^0}^{(i)})\|^2 \\ &\quad + 2 \sum_{i=1}^N \sum_{k=0}^{K-1} \frac{1}{NK} \sum_{s=0}^{k-1} E_{t,k}^{(i)} \|Z_i(O_{t,s^0}^{(i)})\|^2 \\ &\quad + 4 \sum_{i=1}^N \sum_{k=0}^{K-1} \frac{1}{NK} \sum_{s=0}^{k-1} E_{t,k}^{(i)} \|Z_i(O_{t,s}^{(i)}) - Z_i(O_{t,s^0}^{(i)})\|^2 \\ &\quad + 4 \sum_{i=1}^N \sum_{k=0}^{K-1} \frac{1}{NK} \sum_{s=0}^{k-1} E_{t,k}^{(i)} \|Z_i(O_{t,s^0}^{(i)})\|^2 \quad (\text{Eq (11)}) \end{aligned}$$

$$\begin{aligned}
 &+ 2 \sum_{i=1}^N \frac{1}{NK} \sum_{k=0}^{K-1} \sum_{\substack{s; s^0=0 \\ s < s^0}} c_3^2 + 4 \sum_{i=1}^N \frac{1}{NK} \sum_{k=0}^{K-1} \sum_{\substack{s; s^0=0 \\ s < s^0}} c_3 L_2^{s^0-s} \quad (\text{Lemma 12}) \\
 &4 \sum_{i=1}^N \frac{1}{NK} \sum_{k=0}^{K-1} \sum_{s=0}^{K-1-k} E_{t;s}^{(i)} t^2 + 4 \sum_{i=1}^N \frac{1}{NK} \sum_{k=0}^{K-1} 4kc_1^2(K-1)H^2 \\
 &+ 2 \sum_{i=1}^N \frac{1}{NK} \sum_{k=0}^{K-1} \sum_{\substack{s; s^0=0 \\ s < s^0}} c_3^2 + 4 \sum_{i=1}^N \frac{1}{NK} \sum_{k=0}^{K-1} \sum_{\substack{s; s^0=0 \\ s < s^0}} c_3 L_2^{s^0-s} \quad (55) \\
 &\quad \left\{ \frac{z}{M_1} \right\}
 \end{aligned}$$

where we used the property that  $t; i \geq 2H$  in the last inequality, i.e.,  $k_t k \leq H^2$  and  $k_i k \leq H^2$ . We now bound  $M_1$  as:

$$\sum_{\substack{s; s^0=0 \\ s < s^0}} c_3 L_2^{s^0-s} = \sum_{s=0}^{s^0-1} c_3 L_2^{s^0-s} = c_3 L_2 \sum_{s=0}^{s^0-1} L_2^{-s} = c_3 L_2 \frac{1-L_2^{-s^0}}{1-L_2^{-1}} \leq c_3 L_2 \frac{1}{1-L_2^{-1}} \leq \frac{c_3 L_2}{1-L_2^{-1}} \quad (56)$$

Define  $R_K = \sum_{i=1}^N \sum_{k=0}^{K-1} E_{t;k}^{(i)} t^2$  and note that  $R_K$  is monotonically increasing in  $K$ . With this definition, if we plug in the upper bound of  $M_1$  into Eq (55), we have:

$$\begin{aligned}
 R_K &4 \sum_{i=1}^N \sum_{k=0}^{K-1} \sum_{s=0}^{K-1-k} E_{t;s}^{(i)} t^2 + 4 \sum_{i=1}^N \sum_{k=0}^{K-1} 4kc_1^2(K-1)H^2 \\
 &+ 2 \sum_{i=1}^N \sum_{k=0}^{K-1} \sum_{\substack{s; s^0=0 \\ s < s^0}} c_3^2 + 4 \sum_{i=1}^N \sum_{k=0}^{K-1} \sum_{\substack{s; s^0=0 \\ s < s^0}} c_3 L_2 \frac{k}{1-L_2^{-1}} \\
 &= 2 \sum_{i=1}^N \sum_{k=0}^{K-1} \sum_{s=0}^{K-1-k} E_{t;s}^{(i)} t^2 + 4 \sum_{i=1}^N \sum_{k=0}^{K-1} \sum_{\substack{s; s^0=0 \\ s < s^0}} c_3^2 + \frac{2c_3 L_2}{1-L_2^{-1}} + 8c_1^2(K-1)H^2 \\
 &+ 4 \sum_{i=1}^N \sum_{k=0}^{K-1} \sum_{s=0}^{K-1-k} E_{t;s}^{(i)} t^2 \left\{ \frac{z}{R_k} \right\} \\
 &= 2 \sum_{i=1}^N \sum_{k=0}^{K-1} \sum_{s=0}^{K-1-k} E_{t;s}^{(i)} t^2 + \frac{2c_3 L_2}{1-L_2^{-1}} + 8c_1^2(K-1)H^2 + 4 \sum_{i=1}^N \sum_{k=0}^{K-1} c_1^2(K-1) R_k \quad (57)
 \end{aligned}$$

By the monotonicity of  $R_k$ , we have

$$R_K \geq 2 \sum_{i=1}^N \sum_{k=0}^{K-1} \sum_{s=0}^{K-1-k} E_{t;s}^{(i)} t^2 + \frac{2c_3 L_2}{1-L_2^{-1}} + 8c_1^2(K-1)H^2 + 4 \sum_{i=1}^N \sum_{k=0}^{K-1} c_1^2(K-1)^2 R_{K-1}$$

Let us choose  $\beta$  such that  $4 \sum_{i=1}^N \sum_{k=0}^{K-1} \sum_{s=0}^{K-1-k} E_{t;s}^{(i)} t^2 \geq \frac{1}{2}$ , i.e.,  $\beta = \frac{1}{2 \sum_{i=1}^N \sum_{k=0}^{K-1} \sum_{s=0}^{K-1-k} E_{t;s}^{(i)} t^2}$ , the following recursion holds:

$$R_K \geq \frac{1}{2} R_{K-1} + 2 \sum_{i=1}^N \sum_{k=0}^{K-1} \sum_{s=0}^{K-1-k} E_{t;s}^{(i)} t^2 + \frac{2c_3 L_2}{1-L_2^{-1}} + 8c_1^2(K-1)H^2 \quad (58)$$

for all  $k \geq 2$  [K]: Next, we unroll the recurrence, go back  $K-1$  steps and use the fact that  $R_1 = 0$ , we have:

$$\begin{aligned}
 R_K &\geq \left( \sum_{i=1}^N \sum_{k=0}^{K-1} \sum_{s=0}^{K-1-k} E_{t;s}^{(i)} t^2 \right)^{\frac{1}{2}} \left( 2 \sum_{i=1}^N \sum_{k=0}^{K-1} \sum_{s=0}^{K-1-k} E_{t;s}^{(i)} t^2 + \frac{2c_3 L_2}{1-L_2^{-1}} + 8c_1^2(K-1)H^2 \right) \\
 &= 4 \sum_{i=1}^N \sum_{k=0}^{K-1} \sum_{s=0}^{K-1-k} E_{t;s}^{(i)} t^2 + \frac{2c_3 L_2}{1-L_2^{-1}} + 8c_1^2(K-1)H^2 \quad (59)
 \end{aligned}$$

We finish the proof by dividing  $NK$  on both sides and substituting  $\beta = \frac{1}{K^g}$ . □



### I.2.5 Per Round Progress

Lemma 16. (Per Round Progress). If the local step-size  $\eta = \frac{1}{2^{p-1} 2c_1(K-1)}$ , and the effective step-size  $\eta = K^{-1} \eta$  satisfies:

$$\min\left\{\frac{1}{24(c_1 + c_2)^2 + 24\eta^2 + 16}; 1; \frac{1(c_1 + c_2)}{2L_1 + 8\eta^2 c_4^2}; \frac{1}{30c_4(K-1)}; \frac{1}{96c_4^2}; Xg\right\}^5$$

where

$$X = \frac{2B(\eta)G + 3\eta(c_1 + c_2)^2(\eta)}{4B^2(\eta) + 24(c_1 + c_2)^2(\eta) + 2L_1(\eta)G + 6400c_1^2c_4^2\eta^3 + 8c_1^2\eta^2(\eta)}$$

and choose  $\eta = d \frac{\text{mix}(\frac{2}{T})}{K} e$ , then we have,

$$\begin{aligned} E_{t-2k} - E_{t+1} &\leq \underbrace{k^2 \left( (1 + 32\eta(c_1 + c_2)) E_{t-2k} + 2E_{t-2k}^D g(t); t + 4\eta^2 E_{t-2k} g(t)^2 \right)}_{\text{Expected progress for the virtual MDP}} \\ &+ \underbrace{\frac{9 + 28\eta^2}{NK} d^2}_{\text{Linear speedup}} + \underbrace{36L_2^2 + \frac{108}{1} L_2^2 + 4L_1G^2 + 2L_2G}_{\text{High order terms: } O(\eta^3)} \\ &+ \underbrace{\frac{4}{K} \left( \frac{14}{9} + 14\eta(c_1 + c_2) \right) c_3^2 + \frac{2c_3L_2}{1} + 4c_1^2(K-1)H^2}_{\text{drift term}} \\ &+ 4 \underbrace{B(\eta)G + 6\eta(c_1 + c_2)^2(\eta)}_{\text{heterogeneity term}} \end{aligned} \quad (60)$$

where  $\eta$  is any universal positive constant.

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

$$\begin{aligned} E_t &\leq E_{t+1} \\ &= E_t + \frac{1}{NK} \sum_{i=1}^N \sum_{k=0}^{K-1} g_{t;k}^{(i)} \\ &= E_t + \frac{1}{NK} \sum_{i=1}^N \sum_{k=0}^{K-1} g_{t;k}^{(i)} + 2E_t^D \frac{1}{NK} \sum_{i=1}^N \sum_{k=0}^{K-1} g_{t;k}^{(i)} \\ &\quad + 2E_t^D \frac{1}{NK} \sum_{i=1}^N \sum_{k=0}^{K-1} g_{t;k}^{(i)} g_{t;k}^{(i)} \end{aligned}$$

<sup>5</sup>This requirement is very easy to satisfy since the denominator in  $X$  is composed by the heterogeneity terms, which is quite small and thereby makes  $X$  large. Overall, the feasible set of the step-sizes is not empty.

$$\begin{aligned}
 & + 2E_t \frac{1}{NK} \sum_{i=1}^N \sum_{k=0}^{K-1} g_{t;k}^{(i)2} \\
 & \left( E_t \frac{D}{N} \sum_{i=1}^N \sum_{k=0}^{K-1} g_{t;k}^{(i)} g_{t;k}^{(i)} + 2 \frac{E}{NK} \sum_{i=1}^N \sum_{k=0}^{K-1} g_{t;k}^{(i)} g_{t;k}^{(i)} \right) \\
 & \left| \frac{\quad}{B_1} \right\} \\
 & + 2 E_t \frac{D}{NK} \sum_{i=1}^N \sum_{k=0}^{K-1} g_{t;k}^{(i)} g_{t;k}^{(i)} + 2 E_t \frac{1}{NK} \sum_{i=1}^N \sum_{k=0}^{K-1} g_{t;k}^{(i)2} \quad (61) \\
 & \left| \frac{\quad}{B_2} \right\} \quad \left| \frac{\quad}{B_3} \right\}
 \end{aligned}$$

We now begin to bound the gradient bias term  $B_2$  by decomposing this term into three terms:

$$\begin{aligned}
 & \frac{D}{NK} \sum_{i=1}^N \sum_{k=0}^{K-1} g_{t;k}^{(i)} g_{t;k}^{(i)} + 2 \frac{E}{NK} \sum_{i=1}^N \sum_{k=0}^{K-1} g_{t;k}^{(i)} g_{t;k}^{(i)} \\
 & = \frac{D}{NK} \sum_{i=1}^N \sum_{k=0}^{K-1} g_{t;k}^{(i)} g_{t;k}^{(i)} + 2 \frac{E}{NK} \sum_{i=1}^N \sum_{k=0}^{K-1} g_{t;k}^{(i)} g_{t;k}^{(i)} \\
 & \left| \frac{\quad}{B_{21}} \right\} \\
 & + \frac{D}{NK} \sum_{i=1}^N \sum_{k=0}^{K-1} g_{t;k}^{(i)} g_{t;k}^{(i)} + 2 \frac{E}{NK} \sum_{i=1}^N \sum_{k=0}^{K-1} g_{t;k}^{(i)} g_{t;k}^{(i)} \\
 & \left| \frac{\quad}{B_{22}} \right\} \\
 & + \frac{D}{NK} \sum_{i=1}^N \sum_{k=0}^{K-1} g_{t;k}^{(i)} g_{t;k}^{(i)} + 2 \frac{E}{NK} \sum_{i=1}^N \sum_{k=0}^{K-1} g_{t;k}^{(i)} g_{t;k}^{(i)} \quad (62) \\
 & \left| \frac{\quad}{B_{23}} \right\}
 \end{aligned}$$

Next, we bound  $E_t [B_{21}]$  as:

$$\begin{aligned}
 & E_t \frac{D}{NK} \sum_{i=1}^N \sum_{k=0}^{K-1} g_{t;k}^{(i)} g_{t;k}^{(i)} + 2 \frac{E}{NK} \sum_{i=1}^N \sum_{k=0}^{K-1} g_{t;k}^{(i)} g_{t;k}^{(i)} \\
 & \stackrel{(a)}{=} E_t \frac{1}{NK} \sum_{i=1}^N \sum_{k=0}^{K-1} (A_i(O_{t;k}^{(i)}) + A_i)(O_{t;k}^{(i)}) + Z_i(O_{t;k}^{(i)}) \\
 & \quad \frac{1}{NK} \sum_{i=1}^N \sum_{k=0}^{K-1} (A_i(O_{t;k}^{(i)}) - A_i)(O_{t;k}^{(i)}) + Z_i(O_{t;k}^{(i)}) \\
 & \quad + E_t \frac{1}{NK} \sum_{i=1}^N \sum_{k=0}^{K-1} Z_i(O_{t;k}^{(i)}) \\
 & \quad \frac{1}{NK} \sum_{i=1}^N \sum_{k=0}^{K-1} (A_i(O_{t;k}^{(i)}) - A_i)(O_{t;k}^{(i)}) + Z_i(O_{t;k}^{(i)}) \\
 & \quad + \frac{1}{2} E_t \frac{1}{NK} \sum_{i=1}^N \sum_{k=0}^{K-1} Z_i(O_{t;k}^{(i)})^2 + \frac{1}{2} E_t \sum_{i=1}^N \sum_{k=0}^{K-1} \quad \\
 & \stackrel{(b)}{=} \frac{(C_1 + C_2)}{NK} \sum_{i=1}^N \sum_{k=0}^{K-1} E_t \sum_{i=1}^N \sum_{k=0}^{K-1} \quad
 \end{aligned}$$

$$\begin{aligned}
 & + \frac{1}{2} E_t \sum_{i=1}^N \sum_{k=0}^{K-1} Z_i(O_{t;k}^{(i)})^2 + \frac{1}{2} E_t \sum_{i=1}^N \sum_{k=0}^{K-1} \left( \frac{c_1 + c_2}{2NK} \sum_{i=1}^N \sum_{k=0}^{K-1} E_t \sum_{t;k}^{(i)} \right)^2 + \frac{(c_1 + c_2)}{2NK} \sum_{i=1}^N \sum_{k=0}^{K-1} E_t \sum_{t;k}^{(i)} \\
 & \quad \text{(Young's inequality (12) with constant } \frac{1}{2} \text{)} \\
 & + \frac{1}{2} E_t \sum_{i=1}^N \sum_{k=0}^{K-1} Z_i(O_{t;k}^{(i)})^2 + \frac{1}{2} E_t \sum_{i=1}^N \sum_{k=0}^{K-1} \left( \frac{3}{2NK} \sum_{i=1}^N \sum_{k=0}^{K-1} E_t \sum_{t;k}^{(i)} \right)^2 + \frac{3}{2NK} \sum_{i=1}^N \sum_{k=0}^{K-1} E_t \sum_{t;k}^{(i)} \\
 & \quad \text{(c) } \frac{3}{2NK} \sum_{i=1}^N \sum_{k=0}^{K-1} E_t \sum_{t;k}^{(i)} + \frac{3}{2NK} \sum_{i=1}^N \sum_{k=0}^{K-1} E_t \sum_{t;k}^{(i)} + \frac{(c_1 + c_2)}{2NK} \sum_{i=1}^N \sum_{k=0}^{K-1} E_t \sum_{t;k}^{(i)} \\
 & + \frac{3}{2NK} \sum_{i=1}^N \sum_{k=0}^{K-1} E_t \sum_{t;k}^{(i)} + \frac{(c_1 + c_2)}{2NK} \sum_{i=1}^N \sum_{k=0}^{K-1} E_t \sum_{t;k}^{(i)} \\
 & + \frac{1}{2} E_t \sum_{i=1}^N \sum_{k=0}^{K-1} Z_i(O_{t;k}^{(i)})^2 + \frac{1}{2} E_t \sum_{i=1}^N \sum_{k=0}^{K-1} \left( \frac{3}{2NK} \sum_{i=1}^N \sum_{k=0}^{K-1} E_t \sum_{t;k}^{(i)} \right)^2 \\
 & = \frac{3}{2NK} \sum_{i=1}^N \sum_{k=0}^{K-1} E_t \sum_{t;k}^{(i)} + \frac{3}{2} E_t \sum_{i=1}^N \sum_{k=0}^{K-1} \left( \sum_{t;k}^{(i)} \right)^2 \\
 & + \frac{3}{2} \sum_{i=1}^N \sum_{k=0}^{K-1} E_t \sum_{t;k}^{(i)} + \frac{(c_1 + c_2)}{2} E_t \sum_{i=1}^N \sum_{k=0}^{K-1} \left( \sum_{t;k}^{(i)} \right)^2 \\
 & + \frac{1}{2} E_t \sum_{i=1}^N \sum_{k=0}^{K-1} Z_i(O_{t;k}^{(i)})^2 + \frac{1}{2} E_t \sum_{i=1}^N \sum_{k=0}^{K-1} \left( \frac{3}{2NK} \sum_{i=1}^N \sum_{k=0}^{K-1} E_t \sum_{t;k}^{(i)} \right)^2 \\
 & \quad \text{(d) } \frac{3}{2NK} \sum_{i=1}^N \sum_{k=0}^{K-1} E_t \sum_{t;k}^{(i)} + \frac{3}{2} E_t \sum_{i=1}^N \sum_{k=0}^{K-1} \left( \sum_{t;k}^{(i)} \right)^2 \\
 & + \frac{3}{2} \sum_{i=1}^N \sum_{k=0}^{K-1} E_t \sum_{t;k}^{(i)} + \frac{c_1 + c_2}{2} E_t \sum_{i=1}^N \sum_{k=0}^{K-1} \left( \sum_{t;k}^{(i)} \right)^2 \\
 & + \frac{d_2^2}{2NK} + 4L_2^2 \sum_{i=1}^N \sum_{k=0}^{K-1} E_t \sum_{t;k}^{(i)} + \frac{1}{2} E_t \sum_{i=1}^N \sum_{k=0}^{K-1} \left( \sum_{t;k}^{(i)} \right)^2 \\
 & = \frac{3}{2NK} \sum_{i=1}^N \sum_{k=0}^{K-1} E_t \sum_{t;k}^{(i)} + \frac{3}{2} E_t \sum_{i=1}^N \sum_{k=0}^{K-1} \left( \sum_{t;k}^{(i)} \right)^2 \\
 & + \frac{3}{2} \sum_{i=1}^N \sum_{k=0}^{K-1} E_t \sum_{t;k}^{(i)} + \frac{c_1 + c_2}{2} E_t \sum_{i=1}^N \sum_{k=0}^{K-1} \left( \sum_{t;k}^{(i)} \right)^2 + \frac{6}{2} \frac{d_2^2}{NK} + 4 \frac{L_2^2 \sum_{i=1}^N \sum_{k=0}^{K-1} E_t \sum_{t;k}^{(i)}}{4L_2^2 \sum_{i=1}^N \sum_{k=0}^{K-1} E_t \sum_{t;k}^{(i)}}; \quad (63)
 \end{aligned}$$

where (a) is due to  $g_i(O_{t;k}^{(i)}) = A_i(O_{t;k}^{(i)})(\sum_{t;k}^{(i)} + Z_i(O_{t;k}^{(i)}))$ , (b) is due to Lemma 12 (the upper bound of  $A_i(O_{t;k}^{(i)})$  and  $A_i$ ), (c) is due to Eq (13) and (d) is due to Lemma 13.

And we bound  $B_{22}$  as:

$$\begin{aligned}
 B_{22} & = \frac{D}{NK} \sum_{i=1}^N \sum_{k=0}^{K-1} g_i(O_{t;k}^{(i)}) g_i(O_{t;k}^{(i)}) + g_i(O_{t;k}^{(i)}) + g_i(O_{t;k}^{(i)}); \quad E \\
 & \quad \frac{1}{NK} \sum_{i=1}^N \sum_{k=0}^{K-1} g_i(O_{t;k}^{(i)}) g_i(O_{t;k}^{(i)}) + g_i(O_{t;k}^{(i)}) + g_i(O_{t;k}^{(i)}) \\
 & \quad \text{(Cauchy-Schwarz inequality)}
 \end{aligned}$$

$$\begin{aligned}
 & \frac{1}{NK} \sum_{i=1}^N \sum_{k=0}^{K-1} \mathbb{E} \left[ \mathbf{g}_{t;k}^{(i)} \cdot \mathbf{g}_{t;k}^{(i)} \right] + \mathbb{E} \left[ \mathbf{g}_{t;k}^{(i)} \cdot \mathbf{g}_{t;k}^{(i)} \right] \\
 (a) \quad & \frac{1}{NK} \sum_{i=1}^N \sum_{k=0}^{K-1} \mathbb{E} \left[ 2 \|\mathbf{g}_{t;k}^{(i)}\|_{t;k}^{(i)} + 2 \|\mathbf{g}_{t;k}^{(i)}\|_{t;k}^{(i)} \right] \\
 & \frac{1}{NK} \sum_{i=1}^N \sum_{k=0}^{K-1} \mathbb{E} \left[ 4 \|\mathbf{g}_{t;k}^{(i)}\|_{t;k}^{(i)} + 4 \|\mathbf{g}_{t;k}^{(i)}\|_{t;k}^{(i)} + 4 \|\mathbf{g}_{t;k}^{(i)}\|_{t;k}^{(i)} \right] \\
 & \text{(Triangle inequality)} \\
 & \frac{2}{2} \|\mathbf{g}_{t;k}^{(i)}\|_{t;k}^{(i)} + \frac{2}{2} \|\mathbf{g}_{t;k}^{(i)}\|_{t;k}^{(i)} + (2 \|\mathbf{g}_{t;k}^{(i)}\|_{t;k}^{(i)} + 2 \|\mathbf{g}_{t;k}^{(i)}\|_{t;k}^{(i)}) \\
 & \text{(Young's inequality (12) with constants } \frac{1}{2} \text{ and } \frac{1}{2} \|\mathbf{g}_{t;k}^{(i)}\|_{t;k}^{(i)} \text{)} \\
 & \frac{2}{2} \|\mathbf{g}_{t;k}^{(i)}\|_{t;k}^{(i)} + \frac{2}{2} \|\mathbf{g}_{t;k}^{(i)}\|_{t;k}^{(i)} + 12 \|\mathbf{g}_{t;k}^{(i)}\|_{t;k}^{(i)} + (12 \|\mathbf{g}_{t;k}^{(i)}\|_{t;k}^{(i)} + \frac{2}{2} \|\mathbf{g}_{t;k}^{(i)}\|_{t;k}^{(i)}) \quad (\text{Eq 13}) \quad (64)
 \end{aligned}$$

where (a) is due to the 2-Lipschitz property of steady-state  $\mathbf{g}$  (i.e., Lemma 5) and random direction  $\mathbf{g}$  (i.e., Lemma 6),  $\|\mathbf{g}_{t;k}^{(i)}\|_{t;k}^{(i)} = \frac{1}{NK} \sum_{i=1}^N \sum_{k=0}^{K-1} \|\mathbf{g}_{t;k}^{(i)}\|_{t;k}^{(i)}$  and  $\|\mathbf{g}_{t;k}^{(i)}\|_{t;k}^{(i)} = \frac{1}{NK} \sum_{i=1}^N \sum_{k=0}^{K-1} \mathbb{E} \left[ \|\mathbf{g}_{t;k}^{(i)}\|_{t;k}^{(i)} \right]$ .

Now, we bound  $B_{23}$  as:

$$\begin{aligned}
 & \mathbb{E}_t [B_{23}] \\
 & = \frac{D}{NK} \sum_{i=1}^N \sum_{k=0}^{K-1} \mathbb{E}_t \left[ \mathbf{g}_{t;k}^{(i)} \cdot \mathbf{g}_{t;k}^{(i)} \right] \\
 & = \frac{1}{NK} \sum_{i=1}^N \sum_{k=0}^{K-1} \mathbb{E}_t \left[ \|\mathbf{g}_{t;k}^{(i)}\|_{t;k}^{(i)} \right] \quad (\text{Cauchy-Schwarz inequality}) \\
 & = \frac{1}{NK} \sum_{i=1}^N \sum_{k=0}^{K-1} \mathbb{E}_t \left[ \|\mathbf{A}_i(\mathbf{O}_{t;k}^{(i)})\|_{t;k}^{(i)} + \|\mathbf{Z}_i(\mathbf{O}_{t;k}^{(i)})\|_{t;k}^{(i)} + \|\mathbf{A}_i(\mathbf{O}_{t;k}^{(i)})\|_{t;k}^{(i)} \right] \\
 & = \frac{1}{NK} \sum_{i=1}^N \sum_{k=0}^{K-1} \mathbb{E}_t \left[ \|\mathbf{A}_i(\mathbf{O}_{t;k}^{(i)})\|_{t;k}^{(i)} + \|\mathbf{A}_i(\mathbf{O}_{t;k}^{(i)})\|_{t;k}^{(i)} + \|\mathbf{Z}_i(\mathbf{O}_{t;k}^{(i)})\|_{t;k}^{(i)} \right] \\
 (a) \quad & \frac{1}{NK} \sum_{i=1}^N \sum_{k=0}^{K-1} \mathbb{E}_t \left[ L_1 \|\mathbf{O}_{t;k}^{(i)}\|_{t;k}^{(i)} + L_2 \|\mathbf{O}_{t;k}^{(i)}\|_{t;k}^{(i)} \right] \\
 & = \frac{1}{NK} \sum_{i=1}^N \sum_{k=0}^{K-1} \mathbb{E}_t \left[ L_1 \|\mathbf{O}_{t;k}^{(i)}\|_{t;k}^{(i)} + L_2 \|\mathbf{O}_{t;k}^{(i)}\|_{t;k}^{(i)} \right] \\
 (b) \quad & 2L_1 \|\mathbf{O}_{t;k}^{(i)}\|_{t;k}^{(i)} + 2L_1 \|\mathbf{O}_{t;k}^{(i)}\|_{t;k}^{(i)} + 2L_1 \|\mathbf{O}_{t;k}^{(i)}\|_{t;k}^{(i)} + 2L_2 \|\mathbf{O}_{t;k}^{(i)}\|_{t;k}^{(i)} \\
 & = 2L_1 \|\mathbf{O}_{t;k}^{(i)}\|_{t;k}^{(i)} + 2L_1 \|\mathbf{O}_{t;k}^{(i)}\|_{t;k}^{(i)} + 2L_1 \|\mathbf{O}_{t;k}^{(i)}\|_{t;k}^{(i)} + 2L_2 \|\mathbf{O}_{t;k}^{(i)}\|_{t;k}^{(i)} \\
 (c) \quad & 2 \|\mathbf{O}_{t;k}^{(i)}\|_{t;k}^{(i)} + 2L_2 \|\mathbf{O}_{t;k}^{(i)}\|_{t;k}^{(i)} + 2L_1 \|\mathbf{O}_{t;k}^{(i)}\|_{t;k}^{(i)} + 2L_1 \|\mathbf{O}_{t;k}^{(i)}\|_{t;k}^{(i)} \quad (65)
 \end{aligned}$$

where (a) is due to Lemma 12, (b) is due to the fact that  $\|\mathbf{O}_{t;k}^{(i)}\|_{t;k}^{(i)} \leq 2H$ , which radius is  $H \leq \frac{G}{2}$ , and  $\|\mathbf{O}_{t;k}^{(i)}\|_{t;k}^{(i)} = d^{\frac{\log(\frac{2}{\epsilon})}{K}} e$  (i.e.,  $K \geq 2$ ) and (c) is due to the fact that  $\|\mathbf{O}_{t;k}^{(i)}\|_{t;k}^{(i)} \leq 2H$ . Then, the term  $B_2$  can be

bounded as:

$$\begin{aligned}
& E_t [B_2] \\
&= E_t [B_{21} + B_{22} + B_{23}] \\
&\quad \frac{3}{2} \frac{(c_1 + c_2)}{2} E_t [\dots] + \frac{3}{2} \frac{(c_1 + c_2)}{2} E_t [\dots]^2 + \frac{3}{2} \frac{(c_1 + c_2)}{2} E_t [\dots]^2 (\dots) \\
&\quad + \frac{c_1 + c_2}{2} \frac{1}{2} E_t [\dots]^2 + \frac{1}{2} \frac{d_2^2}{NK} + 4L_2^2 \dots K \\
&\quad + \frac{2}{2} E_t [\dots] + \frac{2}{2} E_t [\dots] + 12 \dots E_t [\dots]^2 + (12 \dots + \frac{2}{2}) E_t [\dots] \dots^2 \\
&\quad + 2^2 L_1 G^2 + \dots L_2 G + \dots L_1 E_t [\dots] + \dots L_1 (\dots) G \\
&\quad \frac{3}{2} \frac{(c_1 + c_2)}{2} + 12 \dots E_t [\dots]^2 + \frac{c_1 + c_2}{2} \frac{1}{2} + 12 \dots + \frac{2}{2} E_t [\dots] \dots^2 \\
&\quad + \frac{3}{2} \frac{(c_1 + c_2)}{2} + \frac{2}{2} E_t [\dots] + \frac{2}{2} \dots L_1 \dots + \frac{1}{2} \frac{d_2^2}{NK} + 4L_2^2 \dots^2 \\
&\quad + 2^2 L_1 G^2 + \dots L_2 G + \frac{3}{2} \frac{(c_1 + c_2)}{2} E_t [\dots]^2 (\dots) + \dots L_1 (\dots) G \tag{66}
\end{aligned}$$

Next, we bound  $B_3$  as:

$$\begin{aligned}
& E_t [B_3] \\
&= E_t \frac{1}{NK} \sum_{i=1}^N \sum_{k=0}^{K-1} g_{t,k}^{(i)} g_{t,k}^{(i)} + g_{t,k}^{(i)} g_{t,k}^{(i)} + g_{t,k}^{(i)} g_{t,k}^{(i)} + g_{t,k}^{(i)} g_{t,k}^{(i)} + g_{t,k}^{(i)} g_{t,k}^{(i)} + g_{t,k}^{(i)} g_{t,k}^{(i)} \\
&\quad 4E_t \frac{1}{NK} \sum_{i=1}^N \sum_{k=0}^{K-1} g_{t,k}^{(i)} g_{t,k}^{(i)} + 4E_t \frac{1}{NK} \sum_{i=1}^N \sum_{k=0}^{K-1} g_{t,k}^{(i)} g_{t,k}^{(i)} \\
&\quad + 4E_t \frac{1}{NK} \sum_{i=1}^N \sum_{k=0}^{K-1} g_{t,k}^{(i)} g_{t,k}^{(i)} + 4E_t \frac{1}{NK} \sum_{i=1}^N \sum_{k=0}^{K-1} g_{t,k}^{(i)} g_{t,k}^{(i)} \text{ (Eq 13)} \\
&= 4E_t \frac{1}{NK} \sum_{i=1}^N \sum_{k=0}^{K-1} A_i A_i(O_{t,k}^{(i)}) (O_{t,k}^{(i)}) + Z_i(O_{t,k}^{(i)}) + 4E_t g_{t,k}^{(i)} g_{t,k}^{(i)} \\
&\quad + 4E_t \frac{1}{NK} \sum_{i=1}^N \sum_{k=0}^{K-1} g_{t,k}^{(i)} g_{t,k}^{(i)} + 4E_t \frac{1}{N} \sum_{i=1}^N g_{t,k}^{(i)} g_{t,k}^{(i)} \\
&\quad \text{(a) } 8E_t \frac{1}{NK} \sum_{i=1}^N \sum_{k=0}^{K-1} A_i A_i(O_{t,k}^{(i)}) (O_{t,k}^{(i)}) + 8E_t \frac{1}{NK} \sum_{i=1}^N \sum_{k=0}^{K-1} Z_i(O_{t,k}^{(i)}) \\
&\quad + 16 \frac{1}{NK} \sum_{i=1}^N \sum_{k=0}^{K-1} E_t g_{t,k}^{(i)} g_{t,k}^{(i)} + 4B^2(\dots) + 4E_t g_{t,k}^{(i)} g_{t,k}^{(i)} \\
&\quad \frac{8}{NK} \sum_{i=1}^N \sum_{k=0}^{K-1} E_t A_i A_i(O_{t,k}^{(i)}) (O_{t,k}^{(i)}) + 8E_t \frac{1}{NK} \sum_{i=1}^N \sum_{k=0}^{K-1} Z_i(O_{t,k}^{(i)}) \\
&\quad + 16E_t [\dots] + 4B^2(\dots) + 4E_t g_{t,k}^{(i)} g_{t,k}^{(i)} \\
&\quad \frac{8(c_1 + c_2)^2}{NK} \sum_{i=1}^N \sum_{k=0}^{K-1} E_t g_{t,k}^{(i)} g_{t,k}^{(i)} + 8E_t \frac{1}{NK} \sum_{i=1}^N \sum_{k=0}^{K-1} Z_i(O_{t,k}^{(i)}) \\
&\quad + 16E_t [\dots] + 4B^2(\dots) + 4E_t g_{t,k}^{(i)} g_{t,k}^{(i)}
\end{aligned}$$

$$\begin{aligned}
& \frac{24(c_1 + c_2)^2}{NK} \sum_{i=1}^N \sum_{k=0}^{K-1} \mathbb{E}_t \left\| \sum_{t,k}^{(i)} \right\|_t^2 + \frac{24(c_1 + c_2)^2}{NK} \sum_{i=1}^N \sum_{k=0}^{K-1} \mathbb{E}_t \left\| \sum_{t,k} \right\|_t^2 \\
& + \frac{24(c_1 + c_2)^2}{NK} \sum_{i=1}^N \sum_{k=0}^{K-1} \mathbb{E}_t \left\| \sum_{t,k}^{(i)} \right\|_i^2 + 8 \mathbb{E}_t \left\| \sum_{i=1}^N \sum_{k=0}^{K-1} \frac{1}{NK} \sum_{t,k}^{(i)} Z_i(O_{t,k}^{(i)}) \right\|_t^2 \\
& + 16 \mathbb{E}_t \left\| \sum_{t,k} \right\|_t + 4 B^2(\cdot; \cdot) + 4 \mathbb{E}_t \left\| \sum_{t,k} g(\cdot) \right\|_t^2 \\
& = 24(c_1 + c_2)^2 \mathbb{E}_t \left\| \sum_{t,k} \right\|_t + 24(c_1 + c_2)^2 \mathbb{E}_t \left\| \sum_{t,k} \right\|_t^2 + 24(c_1 + c_2)^2 \mathbb{E}_t \left\| \sum_{t,k} \right\|_t^2 \\
& + 8 \mathbb{E}_t \left\| \sum_{i=1}^N \sum_{k=0}^{K-1} \frac{1}{NK} \sum_{t,k}^{(i)} Z_i(O_{t,k}^{(i)}) \right\|_t^2 + 16 \mathbb{E}_t \left\| \sum_{t,k} \right\|_t + 4 B^2(\cdot; \cdot) + 4 \mathbb{E}_t \left\| \sum_{t,k} g(\cdot) \right\|_t^2 \\
& \stackrel{(b)}{=} (24(c_1 + c_2)^2 + 16) \mathbb{E}_t \left\| \sum_{t,k} \right\|_t + 8 \left( \frac{d_2^2}{NK} + 4 \frac{L_2^2 K}{4L_2^2} \right) + 24(c_1 + c_2)^2 \mathbb{E}_t \left\| \sum_{t,k} \right\|_t^2 \\
& + 4 \mathbb{E}_t \left\| \sum_{t,k} g(\cdot) \right\|_t^2 + 4 B^2(\cdot; \cdot) + 24(c_1 + c_2)^2 \mathbb{E}_t \left\| \sum_{t,k} \right\|_t^2; \tag{67}
\end{aligned}$$

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

Next, we bound  $B_1$  as:

$$\begin{aligned}
& \mathbb{E}_t \left\| \sum_{t,k} [B_1] \right\|_t \\
& = \mathbb{E}_t \left\| \sum_{t,k} \right\|_t^2 + 2 \mathbb{E}_t \left\| \sum_{i=1}^D \frac{1}{N} \sum_{t,k} g(\cdot); \cdot \right\|_t \mathbb{E} \\
& + 2 \mathbb{E}_t \left\| \sum_{i=1}^D \frac{1}{NK} \sum_{i=1}^N \sum_{k=0}^{K-1} g(\cdot^{(i)}); \cdot \right\|_t \mathbb{E} \\
& \mathbb{E}_t \left\| \sum_{t,k} \right\|_t^2 + 2 \mathbb{E}_t \left\| \sum_{i=1}^D \frac{1}{N} \sum_{t,k} g(\cdot); \cdot \right\|_t \mathbb{E} + 2 \mathbb{E}_t \left\| \sum_{t,k} \right\|_t \mathbb{E} \\
& + 2 \mathbb{E}_t \left\| \sum_{i=1}^D \frac{1}{NK} \sum_{i=1}^N \sum_{k=0}^{K-1} g(\cdot^{(i)}); \cdot \right\|_t \mathbb{E} \\
& \mathbb{E}_t \left\| \sum_{t,k} \right\|_t^2 + 2 \mathbb{E}_t \left\| \sum_{i=1}^D \frac{1}{N} \sum_{t,k} g(\cdot); \cdot \right\|_t \mathbb{E} + 2 \mathbb{E}_t \left\| \sum_{t,k} \right\|_t \mathbb{E} \\
& + \frac{1}{3} \mathbb{E}_t \left\| \sum_{i=1}^D \frac{1}{NK} \sum_{i=1}^N \sum_{k=0}^{K-1} g(\cdot^{(i)}); \cdot \right\|_t^2 + \frac{1}{3} \mathbb{E}_t \left\| \sum_{t,k} \right\|_t^2 \\
& \quad \text{(Young's inequality Eq (12) with constant } \frac{1}{3} \text{)} \\
& \stackrel{(a)}{=} \mathbb{E}_t \left\| \sum_{t,k} \right\|_t^2 + 2 B(\cdot; \cdot) G + 2 \mathbb{E}_t \left\| \sum_{t,k} \right\|_t \mathbb{E} \\
& + \frac{1}{3} \mathbb{E}_t \left\| \sum_{i=1}^D \frac{1}{NK} \sum_{i=1}^N \sum_{k=0}^{K-1} g(\cdot^{(i)}); \cdot \right\|_t^2 + \frac{1}{3} \mathbb{E}_t \left\| \sum_{t,k} \right\|_t^2 \\
& \stackrel{(b)}{=} \mathbb{E}_t \left\| \sum_{t,k} \right\|_t^2 + 2 B(\cdot; \cdot) G + 2 \mathbb{E}_t \left\| \sum_{t,k} \right\|_t \mathbb{E} \\
& + \frac{4}{3} \mathbb{E}_t \left\| \sum_{t,k} \right\|_t + \frac{1}{3} \mathbb{E}_t \left\| \sum_{t,k} \right\|_t^2; \tag{68}
\end{aligned}$$

where (a) is due to the fact that  $\sum_{t,k} \leq 2H$  and the gradient heterogeneity; (b) is due to 2-Lipschitz property of function  $g$  in Lemma 5.

Incorporating the upper of  $B_1$  from Eq (68),  $B_2$  from Eq (66) and  $B_3$  from Eq (67) into Eq (61), we have:

$$\begin{aligned}
 E_{t+1} & \leq E_t + 2E_t \frac{D}{g(t)} + 4E_t^2 \frac{E}{g(t)^2} \\
 & + \frac{3 + (3_1(c_1 + c_2) + 24_2) + 24_2^2(c_1 + c_2)^2}{E_1} E_t^2 \\
 & + \frac{c_1 + c_2}{1} + \frac{1}{2} + 24_2 + \frac{4}{2} E_t^2 \\
 & + \frac{9d_2^2}{NK} + 36L_2^2 + 4_3L_1G^2 + 2_3L_2G + \frac{4}{2} + 2_3L_1 \\
 & + \frac{4}{3} + 3_1(c_1 + c_2) + \frac{4}{2} + 24(c_1 + c_2)^2 + 16 E_t [t] \\
 & + 2B(;_1)G + 4_2B^2(;_1) + 24_2^2(c_1 + c_2)^2_2(;_1) \\
 & + 3_1(c_1 + c_2)^2_2(;_1) + 2_3L_1(;_1)G
 \end{aligned} \tag{69}$$

Conditioned on  $F_{t-2}$  and using Lemma 14 to give an upper bound of  $E_{t-2}$ , we have:

$$\begin{aligned}
 E_{t-2} & \leq E_{t-2} + 2E_{t-2} \frac{D}{g(t)} + 4E_{t-2}^2 \frac{E}{g(t)^2} \\
 & + \frac{3 + (3_1(c_1 + c_2) + 24_2) + 24_2^2(c_1 + c_2)^2}{E_1} E_{t-2}^2 \\
 & + \frac{c_1 + c_2}{1} + \frac{1}{2} + 24_2 + \frac{4}{2} E_{t-2}^2 + 14_2^2 \frac{d_2^2}{NK} + \frac{52L_2^2}{1} \\
 & + 4_2^2 c_4^2 \sum_{s=0}^X E_{t-2} [t_s] + 3200_2^2 c_1^2 c_4^2_3_2(;_1) + 4_2^2 c_1^2_2_2(;_1) \\
 & + \frac{9d_2^2}{NK} + 36L_2^2 + 4_3L_1G^2 + 2_3L_2G + \frac{4}{2} + 2_3L_1 E_{t-2} [t] \\
 & + \frac{4}{3} + 3_1(c_1 + c_2) + \frac{4}{2} + 24(c_1 + c_2)^2 + 16 E_{t-2} [t] \\
 & + 2B(;_1)G + 4_2B^2(;_1) + 24_2^2(c_1 + c_2)^2_2(;_1) \\
 & + 3_1(c_1 + c_2)^2_2(;_1) + 2_3L_1(;_1)G
 \end{aligned} \tag{70}$$

If we choose step-size such that  $E_2 = \frac{c_1 + c_2}{1} + \frac{1}{2} + 24_2 + \frac{4}{2}$ ,  $\frac{1}{2} = \frac{1}{2} = \frac{1}{3}$ ,  $E_1 = \frac{3 + (3_1(c_1 + c_2) + 24_2) + 24_2^2(c_1 + c_2)^2}{28_1(c_1 + c_2) + 24_2^2(c_1 + c_2)^2} - \frac{30_1(c_1 + c_2)}{(c_1; c_2 > 1)}$  and  $E_3 = \frac{4}{3} + 3_1(c_1 + c_2) + \frac{4}{2} + 24(c_1 + c_2)^2 + 16 - \frac{(9_1 + 9_1)(c_1 + c_2)}{1}$ , i.e.,

$$\begin{aligned}
 \frac{1}{\frac{c_1 + c_2}{1} + 24_2 + \frac{4}{2}} & = \frac{1}{(c_1 + c_2 + 24_1 + 4)} \\
 \min_f \frac{1}{12(c_1 + c_2)}; 1; & \frac{(\frac{5}{1} + 5_1)(c_1 + c_2)}{24(c_1 + c_2)^2 + 16} g;
 \end{aligned}$$

which is sufficient to hold when  $\min_f \frac{1}{24(c_1 + c_2)^2 + 24_1 + 16}; 1g$ , then we have:

$$E_{t-2} \leq E_{t+1}$$

$$\begin{aligned}
 & E_{t-2}^2 + 2 E_{t-2} \frac{D}{g(t); t} E + 4^2 E_{t-2} g(t)^2 \\
 & + 30 {}_1(c_1 + c_2) E_{t-2}^2 \\
 & + 2 \cdot 8^2 \cdot 2^2 c_4^2 E_{t-2}^h + 14^2 \cdot 2^2 \frac{d_2^2}{NK} + \frac{52L_2^2}{1}^4 \\
 & + 4^2 c_4^2 \sum_{s=0}^X E_{t-2} [t_s] + 3200^2 c_1^2 c_4^2 {}^3 2(; 1) + 4^2 c_1^2 {}^2 2(; 1) \\
 & + \frac{9d_2^2}{NK} {}^2 + 36L_2^2 {}^4 + 4^3 L_1 G^2 + 2^3 L_2 G + \frac{4}{2} + 2^3 L_1 E_{t-2} [t] \\
 & + \frac{4}{3} + 3 {}_1(c_1 + c_2) + \frac{4}{2} + {}^2 24(c_1 + c_2)^2 + 16 E_{t-2} [t] \\
 & + 2 B (; 1) G + 4^2 B^2 (; 1) + 24^2 (c_1 + c_2)^2 {}^2 (; 1) \\
 & + 3 {}_1(c_1 + c_2) {}^2 (; 1) + 2^3 L_1 (; 1) G \\
 & E_{t-2}^2 + 2 E_{t-2} \frac{D}{g(t); t} E + 4^2 E_{t-2} g(t)^2 \\
 & + 30 {}_1(c_1 + c_2) + 16^2 \cdot 2^2 c_4^2 E_{t-2}^2 \\
 & + \frac{9+28^2}{NK} d_2^2 + 36 {}_1 + \frac{3}{1} \frac{3}{2} L_2^2 {}^4 + 4^3 L_1 G^2 + 2^3 L_2 G \\
 & + \frac{4}{1} + 2^3 L_1 E_{t-2} [t] + (\frac{9}{1} + 9 {}_1)(c_1 + c_2) E_{t-2} [t] + 8^2 c_4^2 \sum_{s=0}^X E_{t-2} [t_s] \\
 & + 2 B (; 1) G + 4^2 B^2 (; 1) + 24^2 (c_1 + c_2)^2 {}^2 (; 1) \\
 & + 3 {}_1(c_1 + c_2) {}^2 (; 1) + 2^3 L_1 (; 1) G \\
 & + 6400^2 c_1^2 c_4^2 {}^3 2(; 1) + 8^2 c_1^2 {}^2 2(; 1) \tag{71}
 \end{aligned}$$

if we choose the step-size such that the high order  $O(\cdot^2)$  terms are dominated by the first order terms  $O(\cdot)$ ; i.e.,  $4^2 B^2 (; 1) + 24^2 (c_1 + c_2)^2 {}^2 (; 1) + 2^3 L_1 (; 1) G + 6400^2 c_1^2 c_4^2 {}^3 2(; 1) + 8^2 c_1^2 {}^2 2(; 1) > 2 B (; 1) G + 3 {}_1(c_1 + c_2) {}^2 (; 1)$ ; i.e.,

$$\min_f \frac{2B (; 1)G + 3 {}_1(c_1 + c_2) {}^2 (; 1)}{4B^2 (; 1) + 24(c_1 + c_2)^2 {}^2 (; 1) + 2L_1 (; 1)G + 6400c_1^2 c_4^2 {}^3 2(; 1) + 8c_1^2 {}^2 2(; 1)}; 1g;$$

we have:

$$\begin{aligned}
 & E_{t-2}^2 \\
 & E_{t-2}^2 + 2 E_{t-2} \frac{D}{g(t); t} E + 4^2 E_{t-2} g(t)^2 \\
 & + 30 {}_1(c_1 + c_2) + 16^2 \cdot 2^2 c_4^2 E_{t-2}^2 \\
 & + \frac{9+28^2}{NK} d_2^2 + 36 {}_1 + \frac{3}{1} \frac{3}{2} L_2^2 {}^4 + 4^3 L_1 G^2 + 2^3 L_2 G \\
 & + \frac{4}{1} + 2^3 L_1 E_{t-2} [t] + (\frac{9}{1} + 9 {}_1)(c_1 + c_2) E_{t-2} [t] + 8^2 c_4^2 \sum_{s=0}^X E_{t-2} [t_s] \\
 & + 4 B (; 1) G + 6 {}_1(c_1 + c_2) {}^2 (; 1) \tag{72}
 \end{aligned}$$

With Lemma (15), we have the upper bound of  $E_{t-2} [t]$ ,  $E_{t-2} [t]$  and  $\sum_{s=0}^P E_{t-2} [t_s]$ : Then we have:

$$E_{t-2}^2$$



$$\begin{aligned}
 & E_{t-2}^2 + 2 E_{t-2}^D g(t); t + 4 E_{t-2}^E g(t)^2 \\
 & + \frac{30}{E_4} (c_1 + c_2) + 16 \frac{2c_4^2}{E_4} E_{t-2}^2 \\
 & + \frac{9+28}{NK} d_2^2 + 36L_2^2 + \frac{108}{1} L_2^2 + 4L_1G^2 + 2L_2G \\
 & + \frac{4}{K} \frac{2}{g} \left| \frac{4}{1} + 2 \frac{3L_1}{1} + \left( \frac{9}{1} + 9 \right) (c_1 + c_2) + 8 \frac{2c_4^2}{1} c_3^2 + \frac{2c_3L_2}{1} + 4c_1^2(K-1)H^2 \right. \\
 & \left. + 4B(\cdot; 1)G + 6 \frac{1}{1} (c_1 + c_2)^2(\cdot; 1) \right. \tag{73}
 \end{aligned}$$

If we choose step-size such that  $E_4 = 30 \frac{1}{1} (c_1 + c_2) + 16 \frac{2c_4^2}{1} c_3^2 + 32 \frac{1}{1} (c_1 + c_2)$  and  $E_5 = \frac{4}{1} + 2 \frac{3L_1}{1} + \left( \frac{9}{1} + 9 \right) (c_1 + c_2) + 8 \frac{2c_4^2}{1} c_3^2 + \left( \frac{14}{1} + 14 \right) (c_1 + c_2)$ ; i.e.,

$$\min_f \frac{1}{8} \frac{(c_1 + c_2)}{2c_4^2}; 1; \frac{(\frac{1}{1} + 1)(c_1 + c_2)}{2L_1 + 8c_4^2} g;$$

which is sufficient to hold when  $\frac{1}{2L_1 + 8c_4^2} (c_1 + c_2)$ , then we have:

$$\begin{aligned}
 & E_{t-2}^2 + 2 E_{t-2}^D g(t); t + 4 E_{t-2}^E g(t)^2 \\
 & + 32 \frac{1}{1} (c_1 + c_2) E_{t-2}^2 \\
 & + \frac{9+28}{NK} d_2^2 + 36L_2^2 + \frac{108}{1} L_2^2 + 4L_1G^2 + 2L_2G \\
 & + \frac{4}{K} \frac{3}{g} \left( \frac{14}{1} + 14 \right) (c_1 + c_2) c_3^2 + \frac{2c_3L_2}{1} + 8c_1^2(K-1)H^2 \\
 & + 4B(\cdot; 1)G + 6 \frac{1}{1} (c_1 + c_2)^2(\cdot; 1); \tag{74}
 \end{aligned}$$

□

### I.2.6 Parameter Selection

With Lemma 16, we have:

$$\begin{aligned}
 & E_{t-2}^2 + 2 E_{t-2}^D g(t); t + 4 E_{t-2}^E g(t)^2 \\
 & + (1 + 32 \frac{1}{1} (c_1 + c_2)) E_{t-2}^2 \\
 & + \frac{9+28}{NK} d_2^2 + 36L_2^2 + \frac{108}{1} L_2^2 + 4L_1G^2 + 2L_2G \\
 & + \frac{4}{K} \frac{3}{g} \left( \frac{14}{1} + 14 \right) (c_1 + c_2) c_3^2 + \frac{2c_3L_2}{1} + 8c_1^2(K-1)H^2 \\
 & + 4B(\cdot; 1)G + 6 \frac{1}{1} (c_1 + c_2)^2(\cdot; 1); \tag{75}
 \end{aligned}$$

Proposition 4. If  $\eta$  satisfies the requirement as Lemma 16, choose  $\eta = \frac{(1-\eta)^!}{32(c_1+c_2)}$  and  $\eta = d \frac{\text{mix}(\frac{2}{1})}{K} e$ , we have:

$$\frac{1}{1} E_{t-2}^2 \leq V_t \leq V_D \left( \frac{1}{1} - \eta \right) E_{t-2}^2 + \frac{1}{1} E_{t-2}^2 + \frac{9+28}{NK} d_2^2$$

$$\begin{aligned}
 & + 2 \cdot 36L_2^2 + \frac{108}{1} L_2^2 + 4L_1G^2 + 2L_2G \\
 & + \frac{2c_6}{K} c_3^2 + \frac{2c_3L_2}{1} + 8c_1^2(K-1)H^2 + 4B(\cdot; 1)G + \cdot^2(\cdot; 1) \quad (76)
 \end{aligned}$$

where  $\cdot = \frac{1}{4} = \frac{(1)}{4}!$  and  $c_6, \frac{4}{9}(\frac{14}{1} + 14 \cdot)(c_1 + c_2)$ .

Proof. Incorporating  $\cdot = \frac{(1)}{32(c_1+c_2)}$ ;  $c_6, \frac{4}{9}(\frac{14}{1} + 14 \cdot)(c_1 + c_2)$  and  $6 \cdot (c_1 + c_2) \cdot$  into Eq (75), we have

$$\begin{aligned}
 & E_{t-2} \cdot \cdot + 2 \cdot E_{t-2} \cdot \cdot + 2 \cdot E_{t-2} \cdot \cdot + 4 \cdot E_{t-2} \cdot \cdot \\
 & + (1) \cdot E_{t-2} \cdot \cdot \\
 & + \frac{9+28}{NK} \cdot d_2^2 + \cdot 36L_2^2 + \frac{108}{1} L_2^2 + 4L_1G^2 + 2L_2G \\
 & + \frac{3c_6}{K} c_3^2 + \frac{2c_3L_2}{1} + 8c_1^2(K-1)H^2 \\
 & + 4B(\cdot; 1)G + \cdot^2(\cdot; 1) \\
 (a) & E_{t-2} \cdot \cdot + 2 \cdot (1) \cdot E_{t-2} \cdot \cdot + 16 \cdot E_{t-2} \cdot \cdot \\
 & + (1) \cdot E_{t-2} \cdot \cdot \\
 & + \frac{9+28}{NK} \cdot d_2^2 + \cdot 36L_2^2 + \frac{108}{1} L_2^2 + 4L_1G^2 + 2L_2G \\
 & + \frac{3c_6}{K} c_3^2 + \frac{2c_3L_2}{1} + 8c_1^2(K-1)H^2 \\
 & + 4B(\cdot; 1)G + \cdot^2(\cdot; 1) \\
 & = E_{t-2} \cdot \cdot + \frac{(1)}{2} \cdot E_{t-2} \cdot \cdot + \frac{(1)}{2} \cdot E_{t-2} \cdot \cdot \\
 & + 16 \cdot E_{t-2} \cdot \cdot + \frac{9+28}{NK} \cdot d_2^2 + \cdot 36L_2^2 + \frac{108}{1} L_2^2 + 4L_1G^2 + 2L_2G \\
 & + \frac{3c_6}{K} c_3^2 + \frac{2c_3L_2}{1} + 8c_1^2(K-1)H^2 \\
 & + 4B(\cdot; 1)G + \cdot^2(\cdot; 1) \\
 & E_{t-2} \cdot \cdot + \frac{(1)}{2} \cdot E_{t-2} \cdot \cdot + \frac{(1)}{2} \cdot E_{t-2} \cdot \cdot \\
 & + 16 \cdot E_{t-2} \cdot \cdot + \frac{9+28}{NK} \cdot d_2^2 + \cdot 36L_2^2 + \frac{108}{1} L_2^2 + 4L_1G^2 + 2L_2G \\
 & + \frac{3c_6}{K} c_3^2 + \frac{2c_3L_2}{1} + 8c_1^2(K-1)H^2 \\
 & + 4B(\cdot; 1)G + \cdot^2(\cdot; 1) \\
 (b) & E_{t-2} \cdot \cdot + \frac{(1)}{2} \cdot E_{t-2} \cdot \cdot + \frac{(1)}{4} \cdot E_{t-2} \cdot \cdot \\
 & + \frac{9+28}{NK} \cdot d_2^2 + \cdot 36L_2^2 + \frac{108}{1} L_2^2 + 4L_1G^2 + 2L_2G \\
 & + \frac{3c_6}{K} c_3^2 + \frac{2c_3L_2}{1} + 8c_1^2(K-1)H^2 \\
 & + 4B(\cdot; 1)G + \cdot^2(\cdot; 1)
 \end{aligned}$$

$$\begin{aligned}
 & (1 - 2\epsilon)E_{t-2} - \epsilon \sum_{i=1}^2 V_i - \epsilon \sum_{i=1}^2 V_i^2 + \frac{9+28\epsilon^2}{NK} \epsilon^2 d_2^2 \\
 & + 36L_2^2 + \frac{108}{1-2\epsilon} L_2^2 + 4L_1 G^2 + 2L_2 G \\
 & + \frac{3c_6}{K} \epsilon_3^2 + \frac{2c_3 L_2}{1-2\epsilon} + 8c_1^2 (K-1)H^2 + 4B(\epsilon; 1)G + \epsilon^2(\epsilon; 1)
 \end{aligned} \tag{77}$$

where (a) is due to Lemma 3 and the selection of parameter; (b) is due to  $\epsilon^2 \frac{(1-\epsilon)!}{4}$ . Rearranging the terms and using the fact  $\frac{1}{1-2\epsilon} = \frac{1}{1-\epsilon} + \frac{\epsilon}{1-2\epsilon}$ , we have:

$$\begin{aligned}
 & \epsilon \sum_{i=1}^2 V_i - \epsilon \sum_{i=1}^2 V_i^2 + (1-\epsilon)E_{t-2} - \epsilon \sum_{i=1}^2 V_i - \epsilon \sum_{i=1}^2 V_i^2 + \frac{9+28\epsilon^2}{NK} \epsilon^2 d_2^2 \\
 & + 36L_2^2 + \frac{108}{1-2\epsilon} L_2^2 + 4L_1 G^2 + 2L_2 G \\
 & + \frac{3c_6}{K} \epsilon_3^2 + \frac{2c_3 L_2}{1-2\epsilon} + 8c_1^2 (K-1)H^2 + 4B(\epsilon; 1)G + \epsilon^2(\epsilon; 1)
 \end{aligned} \tag{78}$$

Then we finish the proof by dividing (78) on both sides. □

With these Lemmas, we are now ready to prove Theorem 2.

### I.3 Proof of Theorem 2.

Given a fixed local step-size  $\epsilon = \frac{1}{4^p \frac{1}{2c_1(K-1)}}$ , decreasing effective step-sizes  $\epsilon_t = \frac{8}{(a+t+1)} = \frac{8}{(t+1)! (a+t+1)}$ , decreasing global step-sizes  $\epsilon_g^{(t)} = \frac{1}{K^t}$  and weights  $w_t = (a+t)$ , we have:

$$E_{t-1} - \sum_{i=1}^2 V_i - \sum_{i=1}^2 V_i^2 \leq \frac{2G^2}{K^{2T^2}} + \frac{c_{quad}(\epsilon)}{2NK^T} + \frac{c_{in}(\epsilon)}{4KT^2} + \frac{B(\epsilon; 1)G}{K^T} + \epsilon^2(\epsilon; 1) \tag{79}$$

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

$$\tilde{X}_T = \frac{1}{W} \sum_{t=1}^T X^T w_t \epsilon_t$$

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

$$\begin{aligned}
 & \epsilon \sum_{i=1}^2 V_i - \sum_{i=1}^2 V_i^2 \\
 & \frac{1}{W} \sum_{t=1}^T X^T (a+t)E_{t-1} - \sum_{i=1}^2 V_i - \sum_{i=1}^2 V_i^2 \\
 & \frac{1}{W} \sum_{t=1}^T X^T (a+t)E_{t-1} - \sum_{i=1}^2 V_i - \sum_{i=1}^2 V_i^2 + \frac{1}{W} \sum_{t=2}^T X^T (a+t)E_{t-1} - \sum_{i=1}^2 V_i - \sum_{i=1}^2 V_i^2 \\
 & \frac{(2-1)(a+2-1)G^2}{W} + \frac{1}{W} \sum_{t=2}^T X^T (a+t)E_{t-1} - \sum_{i=1}^2 V_i - \sum_{i=1}^2 V_i^2 \\
 & \frac{(2-1)(a+2-1)G^2}{W} + \frac{1(a+2)(a+2+1)G^2}{2W}
 \end{aligned} \tag{76}$$

$$\begin{aligned}
 & + \frac{1}{W} \sum_{t=2}^T \frac{(9+28^{-2})d_2^2}{NK} (a+t)^{-1} + (a+t)^{-2} \left( 36L_2^2 + \frac{108}{1-2} L_2^2 + 4L_1G^2 + 2L_2G \right) \\
 & + \frac{1}{W} \sum_{t=2}^T \frac{(a+t)^{-2}c_6}{K} c_3^2 + \frac{2c_3L_2}{1} + 8c_1^2(K-1)H^2 \\
 & + \frac{1}{W} \sum_{t=2}^T 4(a+t)B(\cdot; \cdot)G + (a+t)^{-1} (\cdot; \cdot)^2 \tag{80}
 \end{aligned}$$

where  $V_2 = V_D^2$   $G^2$ : We know that  $\frac{1}{W} \sum_{t=2}^T (a+t)^{-2} = \frac{1}{W} \sum_{t=1}^T (a+t)^{-\frac{4}{2(a+t)^2}} = \frac{8 \log(a+T)}{2T^2}$  and that  $\frac{1}{W} \sum_{t=2}^T (a+t)^{-1} = \frac{4}{1T}$ : Plugging in these inequalities into Eq (80), we have:

$$\begin{aligned}
 & E V_2 = V_D^2 \\
 & \frac{3(a+2)(a+2+1)G^2}{2W} + \frac{4(9+28^{-2})d_2^2}{1NKT} \\
 & + \frac{8 \log(a+T)}{2T^2} \left( 36L_2^2 + \frac{108}{1-2} L_2^2 + 4L_1G^2 + 2L_2G \right) \\
 & + \frac{8c_6 \log(a+T)}{2T^2K} c_3^2 + \frac{2c_3L_2}{1} + 8c_1^2(K-1)H^2 \\
 & + 4B(\cdot; \cdot)G + (\cdot; \cdot)^2 \\
 & = \frac{3(a+2)(a+2+1)G^2}{2W} + \frac{4(9+28^{-2})d_2^2}{1NKT} \\
 & + \frac{8 \log(a+T)}{2T^2K} \left( K \left( 36L_2^2 + \frac{108}{1-2} L_2^2 + 4L_1G^2 + 2L_2G \right) + c_6 c_3^2 + \frac{2c_3L_2}{1} + 8c_1^2(K-1)H^2 \right) \\
 & + 4B(\cdot; \cdot)G + (\cdot; \cdot)^2 \tag{81}
 \end{aligned}$$

where  $c_{quad}(\cdot) = 4d_2^2(9+28^{-2})$ . Dividing  $(\cdot)$  on the both sides, changing  $(\cdot)$  into  $(\cdot = (1-\cdot)!)$  and noting that  $c_6 = \frac{4}{9} \left( \frac{14}{1} + 14 \cdot \right) (c_1 + c_2) = O(\frac{1}{\cdot})$ , we have:

$$E V_{\tau} = V_D^2 \leq \frac{2G^2}{K^2T^2} + \frac{c_{quad}(\cdot)}{2NKT} + \frac{c_{lin}(\cdot)}{4KT^2} + \frac{B(\cdot; \cdot)G}{1} + (\cdot; \cdot)^2 \tag{82}$$

We finish the proof by using the inequality,  $E V_{\tau} = V_D^2 \leq 2E V_{\tau} = V_D^2 + 2E V_{\tau} = V_D^2$  and combining with the third point in Theorem 1.  $\square$

## J Simulation Results

### J.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 4. 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 3.

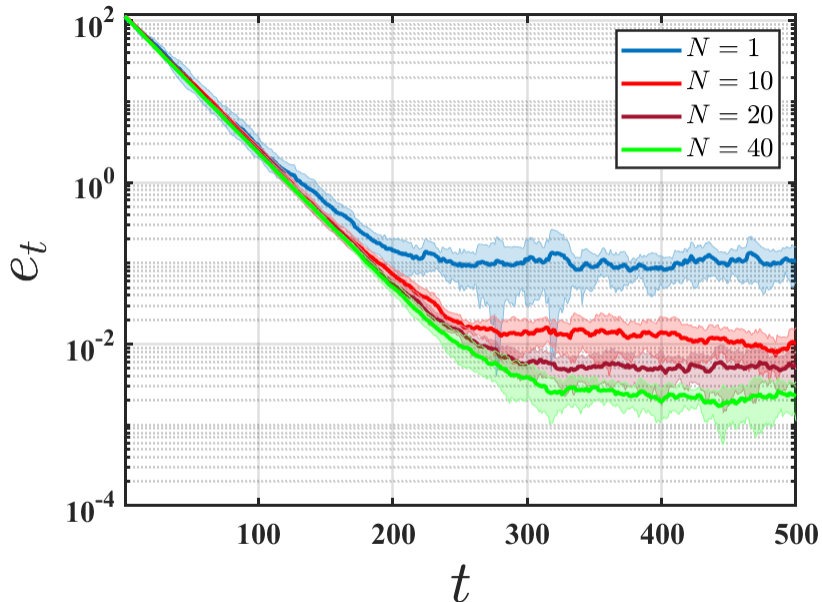


Figure 3: Performance of FedTD(0) with i.i.d. sampling with varying number of agents  $N$ . Solid lines denote the mean and shaded regions indicate the standard deviation over ten runs.

As shown in Fig 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.

## J.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 2. 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 J.1.

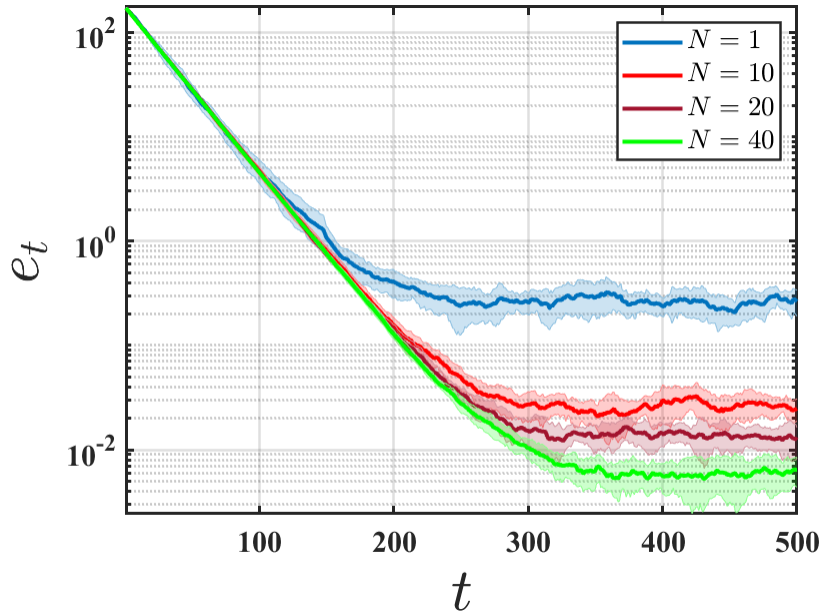


Figure 4: Performance of FedTD(0) with the Markovian sampling with varying number of agents  $N$ . Solid lines denote the mean and shaded regions indicate the standard deviation over ten runs.

As shown in Fig 4, 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.

