# Stochastic Approximation with Delayed Updates: Finite-Time Rates under Markovian Sampling

**Arman Adibi*** Princeton University
**Nicolò Dal Fabbro*** University of Pennsylvania
**Luca Schenato** University of Padova
**Sanjeev Kulkarni** Princeton University
**H. Vincent Poor** Princeton University
**George J. Pappas** University of Pennsylvania
**Hamed Hassani** University of Pennsylvania
**Aritra Mitra** NC State University

## Abstract

Motivated by applications in large-scale and multi-agent reinforcement learning, we study the non-asymptotic performance of stochastic approximation (SA) schemes with delayed updates under Markovian sampling. While the effect of delays has been extensively studied for optimization, the manner in which they interact with the underlying Markov process to shape the finite-time performance of SA remains poorly understood. In this context, our first main contribution is to show that under time-varying bounded delays, the delayed SA update rule guarantees exponentially fast convergence of the *last iterate* to a ball around the SA operator's fixed point. Notably, our bound is *tight* in its dependence on both the maximum delay $\tau_{max}$, and the mixing time $\tau_{mix}$. To achieve this tight bound, we develop a novel inductive proof technique that, unlike various existing delayed-optimization analyses, relies on establishing uniform boundedness of the iterates. As such, our proof may be of independent interest. Next, to mitigate the impact of the maximum delay on the convergence rate, we provide the first finite-time analysis of a delay-adaptive SA scheme under Markovian sampling. In particular, we show that the exponent of convergence of this scheme gets scaled down by $\tau_{avg}$, as opposed to $\tau_{max}$ for the vanilla delayed SA rule; here, $\tau_{avg}$ denotes the average delay across all iterations. Moreover, the adaptive scheme requires no prior knowledge of the delay sequence for step-size tuning. Our theoretical findings shed light on the finite-time effects of delays for a broad class of algorithms, including TD learning, Q-learning, and stochastic gradient descent under Markovian sampling.

## 1 INTRODUCTION

The goal of Stochastic Approximation (SA) theory, introduced by Robbins and Monro (1951), is to find the root (or fixed point) of an operator, given access to noisy observations. The general framework of SA finds applications in various fields like control, optimization, and reinforcement learning (RL). Recently, the surge of interest in distributed and asynchronous learning has motivated the study of variants of SA - e.g., stochastic gradient descent (SGD) - in the presence of delays. This leads to the study of iterative optimization algorithms where the gradients used for the iterative updates are computed at potentially stale iterates from the past (Feyzmahdavian et al., 2016; Koloskova et al., 2022). Notably, the literature on understanding the effects of delays in SA has focused almost exclusively on optimization.

In particular, while delayed versions of more general SA algorithms have been shown to perform well in practice in the context of asynchronous and multi-agent RL (Bouteiller et al., 2020), there is little to no theory to substantiate these empirical observations. Bridging the above gap is the main objective of this paper. However, this task is non-trivial since the transition from optimization to RL requires contending with one major challenge: in the latter, the noisy observations are typically generated from a Markov process. As such, unlike the i.i.d. sampling assumption in op-

*These authors contributed equally to this work.

Proceedings of the 27th International Conference on Artificial Intelligence and Statistics (AISTATS) 2024, Valencia, Spain. PMLR: Volume 238. 

timization, the observations in RL are *temporally correlated*. The interplay between such temporal correlations and delayed updates remains poorly understood in RL. Given this motivation, we provide the first comprehensive *finite-time* analysis of SA schemes under Markovian sampling and delayed updates. As we explain later in the paper, this entails the development of novel analysis techniques that may be of independent interest to both the optimization and RL communities.

### 1.1 Related Works

To position our work and contributions, we now provide a brief summary of relevant literature. A more elaborate survey is deferred to Appendix A.

• *SA under Markovian Sampling.* The asymptotic convergence of SA under correlated Markov randomness was thoroughly investigated for temporal difference (TD) learning in the pioneering work by Tsitsiklis and Vanroy (1997). However, *finite-time* analyses for TD learning and linear SA were only recently provided by Bhandari et al. (2018); Srikant and Ying (2019). These contributions were followed by finite-time analyses of more general nonlinear SA algorithms, such as Q-learning (Chen et al., 2023b; Qu and Wierman, 2020). More recent work has delved into the non-asymptotic convergence of decentralized SA (Zeng et al., 2022; Doan, 2023).

• *Delays and Asynchrony in Optimization.* The study of delays and asynchrony in distributed optimization dates back to the seminal work by Tsitsiklis et al. (1986). While the focus of this work was on asymptotic convergence, the current focus in ML has shifted towards finite-sample guarantees. As such, motivated by the growing popularity of distributed ML, a significant body of work has explored the *finite-time* effects of delays on the performance of stochastic optimization algorithms such as stochastic gradient descent (SGD), and variants thereof (Feyzmahdavian et al., 2016; Gurbuzbalaban et al., 2017; Dutta et al., 2018; Cohen et al., 2021; Koloskova et al., 2022; Glasgow and Wootters, 2022; Nguyen et al., 2022). We note here that stochastic optimization is a special case of SA, where the noise samples that enter into the gradients are generated in an i.i.d. manner, i.e., they exhibit no temporal correlations. However, for applications centered around asynchronous multi-agent RL (MARL) (Bouteiller et al., 2020; Chen et al., 2023a; Mnih et al., 2016), one needs to contend with two key challenges simultaneously: (i) the randomness in the observations comes from a *temporally correlated* Markov process, and (ii) the noisy SA operator is evaluated at potentially stale iterates. *We are unaware of any work that provides a finite-time analysis of SA schemes that feature both the above challenges.*

### 1.2 Contributions

In light of the above discussion, the main contributions of this work are summarized below.

1. **Novel Problem Formulation:** Our work provides the first comprehensive finite-time convergence analysis for delayed SA schemes under Markovian sampling.

2. **Tight Bound in the Constant Delay Case:** We start by analyzing the constant-delay setting in Section 3, and prove exponential convergence of the delayed SA update rule to a noise ball around the fixed point of the SA operator; see Theorem 1. Our analysis reveals that the exponent of convergence gets scaled down by $\max\{\tau, \tau_{mix}\}$, where $\tau$ represents the constant delay, and $\tau_{mix}$ represents the mixing time of the underlying Markov process. Notably, our result (i) complies with existing finite-time bounds for non-delayed SA schemes (Srikant and Ying, 2019); and (ii) exhibits a *tight* dependence on the delay $\tau$; see Section 3 for more discussion on this aspect.

3. **Tight Bound for Time-Varying Delays:** One of the main contributions of this work is to generalize our result on constant delays to *arbitrary time-varying* delays that are bounded. We provide such a result in Theorem 2 of Section 4, where we show that the exponent of convergence gets scaled down this time by $\max\{\tau_{max}, \tau_{mix}\}$. Here, $\tau_{max}$ is the maximum possible delay. An interesting observation stemming from our analysis is that for slowly mixing Markov chains where $\tau_{mix} > \tau_{max}$, the effect of delays gets subsumed by the natural mixing of the underlying Markov process. *As such, slowly mixing Markov chains tend to be more robust to delays.*

4. **Novel Proof Technique:** As we explain in detail in Section 4, achieving a tight dependence on $\tau_{max}$ in Theorem 2 requires a significantly different proof technique than those existing. In particular, our proof neither employs generating function techniques as in (Arjevani et al., 2020), nor the popular error-feedback framework as in (Stich and Karimireddy, 2020). Instead, it relies on a new inductive argument to establish uniform boundedness of the iterates generated by the delayed SA scheme. This allows us to treat the joint effect of delays and Markovian sampling as a bounded perturbation, thereby considerably simplifying the subsequent analysis. The resulting approach is novel, and may be of independent interest to both RL and optimization.

5. **Introduction and Analysis of a Delay-Adaptive Algorithm:** In Section 5, we propose and analyze an intuitive delay-adaptive SA scheme, where an update is made only when the staleness in the update direction falls below a carefully chosen threshold. In Theorem 3, we show that the convergence rate of this scheme depends on the average delay $\tau_{avg}$, as opposed to the maximum delay $\tau_{max}$. Furthermore, the step-size for this adaptive rule *requires no knowledge at all of the delay sequence.*

6. **Applications of Our Results:** As we explain in Section 2.2, our theoretical findings find applications in a variety of settings, including the effects of delays in TD learning, Q-learning, and SGD under Markovian sampling.

*Motivation and Scope.* One of the main practical motivations for the present study lies in multi-agent RL where asynchronous communication naturally leads to delays. That said, the scope of this paper is limited to the single-agent case since even this basic setting poses major technical challenges that remain completely unexplored. As such, by laying the foundations for this single-agent setting, our work opens up avenues for understanding and designing algorithms in more complex MARL environments as future work. We also note that this type of *single-agent* configuration has been the key enabler for the fundamental understanding of finite-time rates for SGD with delayed updates (under i.i.d. sampling) (Feyzmahdavian et al., 2014; Arjevani et al., 2020; Stich and Karimireddy, 2020).

## 2 PROBLEM FORMULATION

The objective of general SA is to solve a root finding problem of the following form:

$$\text{Find } \boldsymbol{\theta}^* \in \mathbb{R}^m \text{ such that } \bar{\mathbf{g}}(\boldsymbol{\theta}^*) = 0, \tag{1}$$

where, for a given approximation parameter $\boldsymbol{\theta} \in \mathbb{R}^m$, the deterministic function $\bar{\mathbf{g}}(\boldsymbol{\theta})$ is the expectation of a noisy operator $\mathbf{g}(\boldsymbol{\theta}, o_t)$ taken over a distribution $\pi$, and $\{o_t\}$ denotes a stochastic *observation process*, which is typically assumed to converge in distribution to $\pi$ (Borkar, 2009; Meyn, 2023).

### 2.1 SA under Markovian Sampling

In this paper, we consider SA under Markovian sampling, i.e., the observations $\{o_t\}$ *are temporally correlated and form a Markov chain.* We define

$$\bar{\mathbf{g}}(\boldsymbol{\theta}) \triangleq \mathbb{E}_{o_t \sim \pi}[\mathbf{g}(\boldsymbol{\theta}, o_t)], \tag{2}$$

where $\pi$ is the *stationary distribution* of the Markov chain $\{o_t\}$. SA consists in finding an approximate solution to (1) while having access only to sampled instances $\mathbf{g}(\boldsymbol{\theta}, o_t)$ of $\bar{\mathbf{g}}(\boldsymbol{\theta})$. The typical iterative SA update rule with a constant step size $\alpha$ is as follows (Srikant and Ying, 2019; Chen et al., 2022),

$$\boldsymbol{\theta}_{t+1} = \boldsymbol{\theta}_t + \alpha \mathbf{g}(\boldsymbol{\theta}_t, o_t). \tag{3}$$

The asymptotic convergence of SA update rules of the form in Eq. (3) under Markovian sampling was first investigated by Tsitsiklis and Vanroy (1997). On the other hand, *finite-time* convergence results under Markovian sampling have only recently been established for linear (Srikant and Ying, 2019; Bhandari et al., 2018) and non-linear (Chen et al., 2022) SA, requiring fairly involved analyses. Indeed, temporal correlations between the samples of $\{o_t\}$, which is also inherited by the iterate sequence $\{\boldsymbol{\theta}_t\}$, prevents the use of techniques commonly adopted for the finite-time study of SA under i.i.d. sampling, triggering the need for a more elaborate analysis.

### 2.2 Exemplar Applications

In this subsection, we provide some examples of the considered SA setting under Markovian sampling.

**TD learning.** TD learning with linear function approximation is an SA algorithm that can be used to learn an approximation of the value function $V_\mu(s)$ of a Markov decision process (MDP) associated to a given policy $\mu$. In this example, we denote by $s \in \mathcal{S}$ the MDP state from a finite $n$-dimensional state space $\mathcal{S}$. The algorithm works by iteratively updating a linear function approximation parameter $\boldsymbol{\theta} \in \mathbb{R}^m$ (with $m < n$) of the (approximated) value function using the following update rule:

$$\begin{aligned} &\boldsymbol{\theta}_{t+1} = \boldsymbol{\theta}_t + \alpha(R_t + \gamma \hat{V}(s_{t+1}; \boldsymbol{\theta}_t) - \hat{V}(s_t; \boldsymbol{\theta}_t))\boldsymbol{\phi}(s_t), \\ &\hat{V}(s; \boldsymbol{\theta}) = \boldsymbol{\theta}^\top \boldsymbol{\phi}(s), \end{aligned} \tag{4}$$

where $\alpha$ is the learning rate and $0 < \gamma < 1$ a discount factor. At iteration $t$, $\boldsymbol{\theta}_t$ is the parameter vector of the linear function approximation, $s_t$ is the state of the MDP that is part of a single Markovian trajectory, $R_t$ is the received reward, $s_{t+1}$ is the next state, $\hat{V}(s_t; \boldsymbol{\theta}_t)$ is the estimated value of the state $s_t$, and $\boldsymbol{\phi}(s) \in \mathbb{R}^m$ is a feature vector that maps the state $s$ to a vector in $\mathbb{R}^m$. The noise in the update rule is introduced by the random reward $R_t$, and the random next state $s_{t+1}$ whose distribution is dictated by the transition probabilities of the underlying Markov chain. Specifically, the update in (4) is an instance of SA with

$$\mathbf{g}(\boldsymbol{\theta}_t, o_t) = (R_t + \gamma \hat{V}(s_{t+1}; \boldsymbol{\theta}_t) - \hat{V}(s_t; \boldsymbol{\theta}_t))\boldsymbol{\phi}(s_t), \tag{5}$$

where $o_t = (R_t, s_t, s_{t+1})$ is the correlated sampling process inducing the randomness. For more formal details on TD learning with linear function approximation, see Srikant and Ying (2019); Bhandari et al. (2018).

**Q-learning.** Q-learning is a reinforcement learning algorithm that can be used to learn an optimal policy for a Markov decision process (MDP). In Q-learning, the goal is to learn a Q-function that maps state-action pairs to their corresponding expected rewards. When the state and action spaces are large, one typically resorts to a parameterized approximation of the Q-function for each state-action pair. The special case of linear function approximation takes the form: $\hat{Q}(s,a;\boldsymbol{\theta}) = \boldsymbol{\theta}^\top \boldsymbol{\phi}(s,a)$, where $\boldsymbol{\phi}(s,a)$ is a fixed feature vector for the state-action pair $(s,a) \in \mathcal{S} \times \mathcal{A}$, and $\mathcal{S}$ and $\mathcal{A}$ represent the (finite) state and action spaces, respectively. The Q-learning update rule with linear function approximation is another instance of SA that takes the following form:

$$\boldsymbol{\theta}_{t+1} = \boldsymbol{\theta}_t + \alpha g(\boldsymbol{\theta}_t, o_t),$$
$$g(\boldsymbol{\theta}_t, o_t) = (R_t + \gamma \max_{a' \in \mathcal{A}} \hat{Q}(s_{t+1}, a'; \boldsymbol{\theta}_t) - \hat{Q}(s_{t+1}, a_t; \boldsymbol{\theta}_t))\boldsymbol{\phi}(s_t, a_t).$$

The notation in the above rule is consistent with what we used to describe the TD update rule in (4). The information used in each time-step is encapsulated in $o_t = (R_t, s_t, a_t, s_{t+1})$. For more details, we refer the reader to (Chen et al., 2022).

**SGD with Markovian noise.** Stochastic Gradient Descent (SGD) is a widely used optimization method for minimizing a function $F(\boldsymbol{\theta}) = \mathbb{E}_{o\sim\pi}[f(\boldsymbol{\theta}, o)]$ with respect to a parameter vector $\boldsymbol{\theta} \in \mathbb{R}^m$ and a distribution $\pi$. SGD operates by updating the parameter vector iteratively in the direction of the function's noisy negative gradient. In this context, SGD can be regarded as an instance of SA with the update rule defined as in (3), with $\mathbf{g}(\boldsymbol{\theta}_t, o_t) = -\nabla f(\boldsymbol{\theta}_t, o_t)$. In online, dynamic settings, the samples $\{o_t\}$ used in SGD can exhibit temporal correlations. In this respect, our results are applicable to the SGD framework in which the samples $\{o_t\}$ are correlated and form a Markov chain - a setting studied for example by Doan (2022) and Even (2023).

### 2.3 SA with delayed updates

In many real-world applications, the SA operator $\mathbf{g}(\cdot)$ is only available when computed with delayed iterates and/or observations. The main **objective** of this paper is to provide a unified framework to analyse the finite-time convergence of SA under the joint effects of Markovian sampling and delayed updates. We proceed by formally introducing the setting. We consider the following stochastic recursion with delayed updates:

$$\boldsymbol{\theta}_{t+1} = \boldsymbol{\theta}_t + \alpha \mathbf{g}(\boldsymbol{\theta}_{t-\tau_t}, o_{t-\tau_t}), \quad 0 \le \tau_t \le t, \tag{6}$$

where $\alpha$ is a constant step size and $\tau_t$ is the delay with which the operator $\mathbf{g}(\cdot)$ is available to be used at iteration $t$. This specific update rule is motivated by many scenarios of practical interest. For instance, in distributed machine learning and reinforcement learning, it is often the case that the agents' updates are performed in an asynchronous manner (Bouteiller et al., 2020; Chen et al., 2023a), leading naturally to update rules of the form (6).

Update rules of the form (6) have been recently studied in the context of SA but with i.i.d. observations (see for example the works by Koloskova et al. (2022) and Nguyen et al. (2022) for SGD updates with delays). However, to the best of our knowledge, nothing is known about the finite-time convergence behaviour of such update rules under Markovian observations. Compared with i.i.d. settings, the Markovian setting introduces major technical challenges, including dealing with (i) the use of a delayed operator $\mathbf{g}(\boldsymbol{\theta}_{t-\tau_t}, o_{t-\tau_t})$, and (ii) sequences of correlated observation samples $\{o_t\}$, which, in turn, induce temporal correlations in the iterates $\{\boldsymbol{\theta}_t\}$. The interplay between such delays and temporal correlations requires a notably careful analysis, one which we provide as a main contribution of this paper. The key features and challenges of the analysis relative to previous works and other settings are provided with more details in sections 3 and 4.

### 2.4 Assumptions and Definitions

We proceed by describing a few assumptions needed for our analysis. First, we make the following natural assumption on the underlying Markov chain $\{o_t\}$ (Bhandari et al., 2018; Srikant and Ying, 2019; Chen et al., 2022).

**Assumption 1.** *The Markov chain $\{o_t\}$ is aperiodic and irreducible.*

Next, we state three further assumptions that are common in the analysis of SA algorithms.

**Assumption 2.** *Problem (1) admits a solution $\boldsymbol{\theta}^*$, and $\exists\, \mu > 0$ such that for all $\boldsymbol{\theta} \in \mathbb{R}^m$, we have*

$$\langle \boldsymbol{\theta} - \boldsymbol{\theta}^*, \bar{\mathbf{g}}(\boldsymbol{\theta}) - \bar{\mathbf{g}}(\boldsymbol{\theta}^*) \rangle \le -\mu \|\boldsymbol{\theta} - \boldsymbol{\theta}^*\|^2. \tag{7}$$

**Assumption 3.** *There exists $L > 0$ such that for any $\boldsymbol{\theta}_1, \boldsymbol{\theta}_2 \in \mathbb{R}^m$ and $o \in \{o_t\}$, we have*

$$\|\mathbf{g}(\boldsymbol{\theta}_1, o) - \mathbf{g}(\boldsymbol{\theta}_2, o)\| \le L\|\boldsymbol{\theta}_1 - \boldsymbol{\theta}_2\|. \tag{8}$$

*Furthermore, there exists $\sigma > 0$ such that for any $\boldsymbol{\theta} \in \mathbb{R}^m$, we have*

$$\|\mathbf{g}(\boldsymbol{\theta}, o)\| \le L(\|\boldsymbol{\theta}\| + \sigma). \tag{9}$$

Assumption 2 is a strong monotone property of the map $-\bar{\mathbf{g}}(\boldsymbol{\theta})$ that guarantees that the iterates generated by a "mean-path" (steady-state) version of Eq. (1), $\boldsymbol{\theta}_{t+1} = \boldsymbol{\theta}_t + \alpha\bar{\mathbf{g}}(\boldsymbol{\theta}_t)$, converge exponentially fast to $\boldsymbol{\theta}^*$. Assumption 3 states that $\mathbf{g}(\boldsymbol{\theta}, o_t)$ is globally uniformly (w.r.t. $o_t$) Lipschitz in the parameter $\boldsymbol{\theta}$.

Finally, we introduce an assumption on the time-varying delay sequence $\{\tau_t\}$.

**Assumption 4.** *There exists an integer $\tau_{max} \geq 0$ such that $\tau_t \leq \tau_{max}, \ \forall t \geq 0$.*

**Remark 1.** *Assumption 2 holds for TD learning (Lemma 1 and Lemma 3 in Bhandari et al. (2018)) with linear function approximation, variants of Q-learning with linear function approximation (Chen et al., 2022), and for strongly convex functions in the context of optimization. Similarly, Assumption 3 holds for TD learning (Bhandari et al., 2018) and Q-learning with linear function approximation (Chen et al., 2022), and is typically made in the analysis of SGD (Doan, 2022).*

We now introduce the following notion of *mixing time* $\tau_{mix}$, that plays a crucial role in our analysis, as in the analysis of all existing finite-time convergence studies of SA under Markovian sampling.

**Definition 1.** *Let $\tau_{mix}$ be such that, for every $\boldsymbol{\theta} \in \mathbb{R}^m$ and $o_0$, we have*

$$\|\mathbb{E}\left[\mathbf{g}(\boldsymbol{\theta}, o_t)|o_0\right] - \bar{\mathbf{g}}(\boldsymbol{\theta})\| \leq \alpha\left(\|\boldsymbol{\theta}\|+1\right), \forall t \geq \tau_{mix}. \quad (10)$$

**Remark 2.** *Note that Assumption 1 implies that the Markov chain $\{o_t\}$ mixes at a geometric rate. This, in turn, implies the existence of some $K \geq 1$ such that $\tau_{mix}$ in Definition 1 satisfies $\tau_{mix} \leq K\log(\frac{1}{\alpha})$. In words, this means that for a fixed $\boldsymbol{\theta}$, if we want the noisy operator $\mathbf{g}(\boldsymbol{\theta}, o_t)$ to be $\alpha$-close (relative to $\boldsymbol{\theta}$) to the expected operator $\bar{\mathbf{g}}(\boldsymbol{\theta})$, then the amount of time we need to wait for this to happen scales logarithmically in the precision $\alpha$.*

## 3 WARM UP: STOCHASTIC APPROXIMATION WITH CONSTANT DELAYS

In this section, we present the first finite-time convergence analysis of a SA scheme with a constant delay under Markovian sampling. With respect to the SA scheme with delayed updates introduced in (6), we fix $\tau_t = \tau$, where $\tau$ represents the constant delay. This leads to the following update rule:

***SA with Constant Delay:***

$$\boldsymbol{\theta}_{t+1} = \begin{cases} \boldsymbol{\theta}_0 & \text{if } \ 0 \leq t < \tau, \\ \boldsymbol{\theta}_t + \alpha\mathbf{g}(\boldsymbol{\theta}_{t-\tau}, o_{t-\tau}) & \text{if } \ t \geq \tau. \end{cases} \quad (11)$$

We set $\boldsymbol{\theta}_{t+1}$ to $\boldsymbol{\theta}_0$ for the first $\tau$ time-steps to simplify the exposition of the analysis; such a choice is not necessary, and one can use other natural initial conditions here without affecting the final form of our result. We now state our first main result that provides a finite-time convergence bound for the update rule in (11).

**Theorem 1.** *Suppose Assumptions 1-3 hold. Let $w_t \triangleq (1-0.5\alpha\mu)^{-(t+1)}$ and $W_T = \sum_{t=0}^{T} w_t$. Let $\boldsymbol{\theta}_{out}$ be an iterate chosen randomly from $\{\boldsymbol{\theta}_t\}_{t=0}^{T}$, such that $\boldsymbol{\theta}_{out} = \boldsymbol{\theta}_t$ with probability $\frac{w_t}{W_T}$. Define $r_{out} \triangleq \|\boldsymbol{\theta}_{out} - \boldsymbol{\theta}^*\|$ and $\bar{\tau} \triangleq \max\{\tau, \tau_{mix}\}$. There exists a universal constant $C_1 \geq 2$, such that, for $\alpha \leq \frac{\mu}{C_1 L^2 \bar{\tau}}$, the following holds for the rule in (11) $\forall T \geq 0$:*

$$\mathbb{E}\left[r_{out}^2\right] \leq C_\alpha \exp\left(-0.5\alpha\mu T\right) + O(\sigma^2)\frac{\alpha L^2 \bar{\tau}}{\mu}, \quad (12)$$

*with $C_\alpha = O\left(\frac{1}{\alpha\mu} + \frac{\bar{\tau}\sigma^2}{\mu}\right)$. Setting $\alpha = \frac{\mu}{C_1 L^2 \bar{\tau}}$, we get*

$$\mathbb{E}\left[r_{out}^2\right] \leq C_{\bar{\tau}} \exp\left(-0.5\frac{\mu^2}{C_1 L^2 \bar{\tau}} T\right) + O(\sigma^2), \quad (13)$$

*with $C_{\bar{\tau}} = O\left(\frac{2C_1 L^2 \bar{\tau}}{\mu^2} + \frac{\sigma^2 \bar{\tau}}{\mu}\right)$.*

**Main Takeaways:** We now outline the key takeaways from the above Theorem. From (13), we note that the delayed SA scheme in (11) guarantees exponentially fast convergence (in mean-squared sense) to a ball around the fixed point $\boldsymbol{\theta}^*$, where the size of the ball is proportional to the noise-variance level $\sigma^2$. For SA problems satisfying Assumptions 1-3, a convergence bound of this flavor is consistent with what is observed in the non-delayed setting as well under Markovian sampling, when one employs a constant step-size. See, for instance, the bounds for TD learning in (Srikant and Ying, 2019; Bhandari et al., 2018), and for Q-learning in (Chen et al., 2022).

*The Effect of Delays.* A close look at (13) reveals that the exponent of convergence in the first term of (13) scales inversely with $\bar{\tau} = \max\{\tau, \tau_{mix}\}$. Hence, if $\tau \geq \tau_{mix}$, the constant delay $\tau$ can slacken the rate of exponential convergence to the noise ball. Is such an effect of the delay $\tau$ inevitable, or just an artifact of our analysis? We elaborate on this point below.

*On the Tightness of our Bounds.* It turns out that the inverse scaling of the exponent with $\tau$ shows up even for SGD with constant delays under i.i.d. sampling - a special case of the general SA scheme we study here. Furthermore, this dependency has been shown to be tight for SGD in (Arjevani et al., 2020). As such, our rate in Theorem 1 is tight in its dependence on the delay $\tau$. We also note that the inverse scaling of the exponent with the mixing time $\tau_{mix}$ is consistent with prior work (Srikant and Ying, 2019; Bhandari et al., 2018),

Table 1: Summary of results. Here, we do not show problem specific constants other than $\tau_{mix}$ and $\tau_{max}$.

| Algorithm | Variance Bound | Bias Bound |
|---|---|---|
| Constant Delay (11) | $O(\sigma^2)$ | $O\left(\exp\left(\frac{-T}{\max\{\tau,\tau_{mix}\}}\right)\right)$ |
| Time-Varying Delays (14) | $O(\sigma^2)$ | $O\left(\exp\left(\frac{-T}{\max\{\tau_{mix},\tau_{max}\}}\right)\right)$ |
| Time-Varying Delays Delay-Adaptive Update (23) | $O(\sigma^2)$ | $O\left(\exp\left(\frac{-T}{\tau_{mix}.\tau_{avg}}\right)\right)$ |

and is, in fact, *unavoidable* (Nagaraj et al., 2020). To sum up, in Theorem 1 we provide the *first finite-time convergence bound for a SA scheme with updates subject to constant delays under Markovian sampling*, and establish a convergence rate that has *tight* dependencies on both the delay $\tau$ and the mixing time $\tau_{mix}$.

*Slowly-mixing Markov Chains are Robust to Delays.* An interesting conclusion from our analysis is that if the underlying Markov chain mixes slowly, i.e., if $\tau_{mix}$ is large, and in particular, if $\tau_{mix} > \tau$, then the convergence rate will be dictated by $\tau_{mix}$, since then $\bar{\tau} = \max\{\tau, \tau_{mix}\} = \tau_{mix}$. In other words, the effect of the delays will be dominated by the natural mixing effect of the underlying Markov process. This observation is novel to our setting and analysis, and absent in optimization under delays with i.i.d. sampling.

*Comments on the Analysis.* Our proof for Theorem 1 - provided in Appendix B.1 - is partially inspired by the analysis in (Stich and Karimireddy, 2020). However, the key technical challenge relative to (Stich and Karimireddy, 2020) arises from the need to simultaneously contend with delays and temporal correlations introduced by Markovian sampling; notably, the existing literature on optimization under delays (Stich and Karimireddy, 2020; Dutta et al., 2018; Koloskova et al., 2022) *does not* need to deal with the latter aspect. This dictates the need for new ingredients in the analysis that we outline in Appendix B.1. However, as we explain in Appendix B.2, the technique we employ to prove Theorem 1 falls apart in the face of time-varying delays. In the next section, we will explain in some detail how to overcome this challenge.

## 4 STOCHASTIC APPROXIMATION WITH TIME-VARYING DELAYS

The goal of this section is to analyze (6) in its full generality by accounting for Markovian sampling and arbitrary time-varying delays that are only required to be bounded, as per Assumption 4. In particular, we will study the following SA update rule:

***SA with Time-Varying Delays:***

$$\boldsymbol{\theta}_{t+1} = \boldsymbol{\theta}_t + \alpha \mathbf{g}(\boldsymbol{\theta}_{t-\tau_t}, o_{t-\tau_t}), \tau_t \leq \min\{t, \tau_{max}\}. \tag{14}$$

The main contribution of this paper - as stated below - is a finite-time convergence result for the above rule.

**Theorem 2.** *Suppose Assumptions 1-4 hold. Let* $r_t \triangleq \|\boldsymbol{\theta}_t - \boldsymbol{\theta}^*\|$, *and* $\bar{\tau} \triangleq \max\{\tau_{mix}, \tau_{max}\}$. *There exists a universal constant* $C \geq 2$, *such that, for* $\alpha \leq \frac{\mu}{CL^2\bar{\tau}}$, *the iterates generated by the update rule (14) satisfy the following* $\forall T \geq 3\bar{\tau}$*:*

$$\mathbb{E}\left[r_T^2\right] \leq \left(\exp\left(-2\alpha\mu T\right) + \frac{\alpha L^2(\tau_{mix} + \tau_{max})}{\mu}\right) O(\sigma^2). \tag{15}$$

*Setting* $\alpha = \frac{\mu}{CL^2\bar{\tau}}$, *we get*

$$\mathbb{E}\left[r_T^2\right] \leq \left(\exp\left(-\frac{2\mu^2 T}{CL^2\bar{\tau}}\right) + 1\right) O(\sigma^2). \tag{16}$$

**Main Takeaways:** Comparing (16) with (13), we note that our bound in Theorem 2 for time-varying delays mirrors that for the constant delay setting in Theorem 1, with $\tau_{\max}$ appearing in the exponent in (16) exactly in the same manner as $\tau$ shows up in (13). Thus, all the conclusions we drew in Section 3 after stating Theorem 1 carry over to the setting we study here as well. In particular, since the constant-delay model is a special case of the arbitrary-delay model (with bounded delays), and we have already argued the tightness of our bound for the constant-delay setting, the dependencies of our bound (in (16)) on the maximum delay $\tau_{max}$, and the mixing time $\tau_{mix}$, are optimal. At this stage, one might contemplate the need for studying the constant-delay case in Section 3, given that our results in this section clearly subsume the former. To explain our rationale, it is instructive to consider what is already known about delays in optimization under i.i.d. sampling.

Some of the early works analyzing the effect of time-varying bounded delays on gradient descent and variants thereof, show that for smooth, strongly-convex objectives, the iterates generated by the delayed rule converge exponentially fast (Assran et al., 2020; Feyzmahdavian et al., 2016; Gurbuzbalaban et al., 2017). However, the exponent of convergence in these works is *sub-optimal* in that it scales inversely with $\tau_{\max}^2$, as opposed to $\tau_{\max}$ in our bound in Theorem 2. More recently, under a constant delay $\tau$, the authors in (Arjevani et al., 2020) established the optimal dependence on $\tau$ in the exponent for quadratics; the same result

was generalized to smooth, strongly-convex functions in (Stich and Karimireddy, 2020). While our analysis for the constant-delay case in Theorem 1 borrows some ideas from that in (Stich and Karimireddy, 2020), such a technique appears to be insufficient for achieving the tight dependence on $\tau_{\max}$ in Theorem 2.

The proof of Theorem 2 departs from previous proof techniques for optimization under delays, avoiding the need for using generating functions as in (Arjevani et al., 2020), or the error-feedback framework with carefully crafted weighted averages of iterates as in (Stich and Karimireddy, 2020). The purpose of studying the constant-delay case first is precisely to highlight the critical points that make it difficult to adapt the aforementioned existing proof techniques to suit our needs - we provide such a discussion in Appendix B.2. Relative to the above works that *only provide bounds for a constant delay under i.i.d. sampling*, our proof of Theorem 2 adopts a conceptually simpler route: it relies on a novel inductive argument to demonstrate the uniform boundedness of iterates. Surprisingly, this simpler approach - which we explain in some detail in the next section - provides a tight bound for the more challenging general case of time-varying delays under Markovian sampling.

### 4.1 Overview of Our Proof Technique

In this section, we begin by outlining the main challenges that make it difficult to adapt existing optimization and RL proofs to our setting. We then explain the key novel technical ingredients that allow us to overcome such challenges. To proceed, let us define the error (at time $t$) introduced by the delay as follows:

$$\mathbf{e}_t \triangleq \mathbf{g}(\boldsymbol{\theta}_t, o_t) - \mathbf{g}(\boldsymbol{\theta}_{t-\tau_t}, o_{t-\tau_t}). \tag{17}$$

Using the above in (14) yields

$$\boldsymbol{\theta}_{t+1} = \boldsymbol{\theta}_t + \alpha \mathbf{g}(\boldsymbol{\theta}_t, o_t) - \alpha \mathbf{e}_t. \tag{18}$$

We examine $\|\boldsymbol{\theta}_{t+1} - \boldsymbol{\theta}^*\|^2$ using (18), which leads to

$$\|\boldsymbol{\theta}_{t+1} - \boldsymbol{\theta}^*\|^2 = J_{t,1} + \alpha^2 J_{t,2} - 2\alpha J_{t,3}, \text{ with} \tag{19}$$

$$\begin{aligned} J_{t,1} &\triangleq \|\boldsymbol{\theta}_t - \boldsymbol{\theta}^* + \alpha \mathbf{g}(\boldsymbol{\theta}_t, o_t)\|^2, \\ J_{t,2} &\triangleq \|\mathbf{e}_t\|^2, \\ J_{t,3} &\triangleq \langle \mathbf{e}_t, \boldsymbol{\theta}_t - \boldsymbol{\theta}^* \rangle + \alpha \langle \mathbf{e}_t, \mathbf{g}(\boldsymbol{\theta}_t, o_t) \rangle. \end{aligned} \tag{20}$$

Note that the presence of $J_{t,2}$ and $J_{t,3}$ in (19) is a consequence of the delay, and does not show up in the analysis of vanilla non-delayed SA schemes. Our convergence analysis is built upon providing bounds in expectation on each of the three terms defined in (20). We now elaborate on why this task is non-trivial.

• **Challenge 1: Bounding the Drift.** Bounding $J_{t,1}$ in (20) requires exploiting the geometric mixing property of the underlying Markov chain in (10). To do so, the standard technique in SA schemes with Markovian sampling is to condition on the state of the system sufficiently into the past. This creates the need to bound a drift term of the form $\|\boldsymbol{\theta}_t - \boldsymbol{\theta}_{t-\tau_{mix}}\|$. In the absence of delays, there are two main approaches to handling this drift term. In (Bhandari et al., 2018), the authors are able to provide a uniform bound on this term by assuming a projection step. However, such a projection step may be computationally expensive, and requires knowledge of the radius of the ball that contains $\boldsymbol{\theta}^*$. In (Srikant and Ying, 2019), the authors were able to bypass the need for a projection step via a finer analysis; in particular, Lemma 3 in (Srikant and Ying, 2019) provides a bound of the following form:

$$\|\boldsymbol{\theta}_t - \boldsymbol{\theta}_{t-\tau_{mix}}\| \leq O(\alpha \tau_{mix})(\|\boldsymbol{\theta}_t\| + \sigma), \forall t \geq \tau_{mix}. \tag{21}$$

The key feature of the above result is that it relates the drift to the *current iterate* $\boldsymbol{\theta}_t$. However, for the delayed SA scheme we study in (14), a bound of the form in (21) no longer applies. The reason for this follows from the fact that the presence of delays in our setting causes the drift $\|\boldsymbol{\theta}_t - \boldsymbol{\theta}_{t-\tau_{mix}}\|$ *to be a function of not just the current iterate, but several other iterates from the past.* Thus, we need a new technical result that is able to explicitly relate the drift to all past iterates over a certain time window. We provide such a result in Lemma 1.

• **Challenge 2: Handling Delay-induced Errors and Temporal Correlations.** Bounding $J_{t,3}$ in (20) requires handling the term $\langle \mathbf{e}_t, \boldsymbol{\theta}_t - \boldsymbol{\theta}^* \rangle$. This step is much more challenging compared to the i.i.d. sampling setting considered in the optimization literature with delays (Zhou et al., 2018; Koloskova et al., 2022; Arjovsky et al., 2017; Cohen et al., 2021). The difficulty here arises due to the statistical correlation among the terms $\boldsymbol{e}_t = \mathbf{g}(\boldsymbol{\theta}_t, o_t) - \mathbf{g}(\boldsymbol{\theta}_{t-\tau_t}, o_{t-\tau_t})$ and $\boldsymbol{\theta}_t - \boldsymbol{\theta}^*$. In particular, there are two issues to contend with: (i) due to the correlated nature of the Markovian samples, $\mathbb{E}[\mathbf{g}(\boldsymbol{\theta}_t, o_t)] \neq \bar{\mathbf{g}}(\boldsymbol{\theta}_t)$, and (ii) the observation $o_{t-\tau_t}$ influences the iterate $\boldsymbol{\theta}_t$. In view of the above issues, we depart from the standard routes employed in the optimization literature to handle similar delay-induced error terms, and employ a careful mixing time-argument instead to bound $\langle \mathbf{e}_t, \boldsymbol{\theta}_t - \boldsymbol{\theta}^* \rangle$.

• **Challenge 3: Time-Varying Delays.** As we alluded to earlier in Section 3, the extension from the constant-delay setting to the time-varying setting appears to be quite non-trivial. In Appendix B.2, we consider a few potential candidate proof strategies for the setting with time-varying delays; these include: (i) a natural extension of the technique we use to

prove Theorem 1, and (ii) an adaptation of the proof in (Feyzmahdavian et al., 2014). Unfortunately, while these strategies do end up providing rates, the exponent of convergence exhibits a *sub-optimal* inverse scaling with $\tau_{max}^2$.

### 4.2 Auxiliary Lemmas

The proof of Theorem 2 hinges on three new technical results. We outline them below. Our first result provides bounds in expectation on drift terms of the form $\|\boldsymbol{\theta}_t - \boldsymbol{\theta}_{t-\tau}\|^2$. To state this result, we define $\tau' \triangleq 2\tau_{max} + \tau_{mix}$, and $r_{t,2} \triangleq \max_{t-\tau' \le l \le t} \mathbb{E}\left[r_l^2\right]$.

**Lemma 1.** *For any $t \ge \tau_{mix}$, we have*

$$(i)\ \mathbb{E}\left[\|\boldsymbol{\theta}_t - \boldsymbol{\theta}_{t-\tau_{mix}}\|^2\right] \le 2\alpha^2\tau_{mix}^2 L^2(2r_{t,2} + 3\sigma^2).$$

*Similarly, for any $t \ge 0$ and $\tau_t \le t$,*

$$(ii)\ \mathbb{E}\left[\|\boldsymbol{\theta}_t - \boldsymbol{\theta}_{t-\tau_t}\|^2\right] \le 2\alpha^2\tau_{max}^2 L^2(2r_{t,2} + 3\sigma^2).$$

Unlike the analogous drift bounds in the RL literature (Srikant and Ying, 2019), Lemma 1 bounds the drift as a function of the maximum iterate-suboptimality $r_{t,2}$ over a horizon whose length scales as $O(\tau_{mix} + \tau_{max})$. Exploiting Lemma 1, we can provide bounds on $\mathbb{E}\left[J_{t,1}\right]$, $\mathbb{E}\left[J_{t,2}\right]$, and $\mathbb{E}\left[J_{t,3}\right]$ in terms of $r_{t,2}$. This is the subject of the next result.

**Lemma 2.** *Let $t \ge \tau' = 2\tau_{max} + \tau_{mix}$. Then*

$$\begin{aligned}
(i)\quad & \mathbb{E}\left[J_{t,1}\right] \le (1-2\alpha\mu)\mathbb{E}\left[r_t^2\right] + O(\alpha^2\tau_{mix}L^2 r_{t,2}) \\
& \qquad + O(\alpha^2\tau_{mix}L^2\sigma^2). \\
(ii)\quad & \mathbb{E}\left[J_{t,2}\right] \le O(L^2(2r_{t,2} + 3\sigma^2)). \\
(iii)\quad & \mathbb{E}\left[-J_{t,3}\right] \le O(\alpha L^2(\tau_{mix} + \tau_{max})(r_{t,2} + \sigma^2)).
\end{aligned}$$

The most challenging part of proving the above lemma is part $(iii)$, where mixing time arguments need to be carefully applied to deal with temporal correlations. Plugging the above bounds in (19) yields:

$$\mathbb{E}\left[r_{t+1}^2\right] \le (1-2\alpha\mu)\mathbb{E}\left[r_t^2\right] + O(\alpha^2L^2\bar{\tau})(r_{t,2} + \sigma^2), \tag{22}$$

where $\bar{\tau} \triangleq \max\{\tau_{mix}, \tau_{max}\}$. The above inequality suggests that the one-step progress can be captured by a contractive term, and an additive perturbation that jointly captures the effects of delays and Markovian sampling. The crucial step now is to handle this perturbation term, while achieving the optimal dependence on $\tau_{max}$. Our key technical innovation here is to use a novel inductive argument to prove that the iterates generated by (14) remain uniformly bounded in expectation. We have the following result.

**Lemma 3.** *Suppose $\alpha \le \frac{\mu}{C\bar{\tau}L^2}$. There exists a universal constant $C \ge 1$ such that for all $t \ge 0$, it holds that $\mathbb{E}\left[r_t^2\right] \le O(\sigma^2)$.*

The above result immediately implies that the perturbation term appearing in (22) can be *uniformly* bounded by $O(\alpha^2L^2\bar{\tau}\sigma^2)$. From this point onward, the analysis is straightforward. It is worth emphasizing here that the *idea of establishing a uniform bound on the perturbations induced by delays and Markovian sampling - without a projection step - is the main departure from all existing analyses.*

## 5 DELAY-ADAPTIVE STOCHASTIC APPROXIMATION

In the previous section, we analyzed the vanilla delayed SA rule under time-varying delays. Notably, the resulting convergence rate was dependent on the maximum delay, $\tau_{max}$, and selecting the proper step size required a priori knowledge of $\tau_{max}$. In practice, the worst-case delay $\tau_{max}$ may be very large, leading to slow convergence of (14); furthermore, $\tau_{max}$ may be unknown. Here, we introduce a new *delay-adaptive* update rule whose convergence rate only depends on the *average* delay, $\tau_{avg} = \frac{1}{T}\sum_{t=1}^{T}\tau_t$, for the $T$ iterations over which the algorithm is executed. Moreover, the proposed *delay-adaptive* algorithm does not require any knowledge of the delay sequence at all for tuning the step size. Our proposed rule is as follows.

***Delay-adaptive SA:***

$$\boldsymbol{\theta}_{t+1} = \begin{cases} \boldsymbol{\theta}_t + \alpha\mathbf{g}(\boldsymbol{\theta}_{t-\tau_t}, o_{t-\tau_t}) & \|\boldsymbol{\theta}_t - \boldsymbol{\theta}_{t-\tau_t}\| \le \epsilon, \\ \boldsymbol{\theta}_t & \text{otherwise.} \end{cases} \tag{23}$$

The rationale behind the above rule is quite simple: it makes an update only if the pseudo-gradient $\mathbf{g}(\boldsymbol{\theta}_{t-\tau_t}, o_{t-\tau_t})$ available at time $t$ is not too stale in the sense that it is evaluated at an iterate that is at most $\epsilon$ away from the current iterate $\boldsymbol{\theta}_t$. Unlike the vanilla SA update rule, we can control the effect of delays on the convergence rate by carefully picking the threshold $\epsilon$. This is reflected in the next result.

**Theorem 3.** *Suppose Assumptions 1-4 hold. Let $r_t \triangleq \|\boldsymbol{\theta}_t - \boldsymbol{\theta}^*\|$. There exists a universal constant $C_2 \ge 1$ such that for $\alpha \le \frac{\mu}{C_2L^2\tau_{mix}}$, and $\epsilon = \alpha$, the iterates of (23) satisfy the following for $T \ge \tau_{mix}$:*

$$\mathbb{E}\left[r_T^2\right] \le \left(\exp\left(\frac{-\alpha\mu T}{4(\tau_{avg}+1)}\right) + \frac{L^2\alpha\tau_{mix}}{\mu}\right)O(\sigma^2).$$

*Additionally, if we set $\alpha = \frac{\mu}{C_2L^2\tau_{mix}}$, we obtain*

$$\mathbb{E}\left[r_T^2\right] \le \left(\exp\left(\frac{-\mu^2 T}{4C_2L^2\tau_{mix}(\tau_{avg}+1)}\right) + 1\right)O(\sigma^2).$$

**Discussion and Insights.** Relative to Theorem 2, there are two main takeaways from the above result:

(i) the choice of the step-size $\alpha$ requires no information about the delay sequence, and (ii) the exponent of convergence gets scaled down by $\tau_{avg}$, and not $\tau_{max}$. This result is significant because *it is the first to provide a finite-time result for a delay-adaptive SA scheme under Markovian sampling.*

We now provide some intuition as to what makes the above result possible. The vanilla SA rule in (14) always makes an update, *no matter how delayed the pseudo-gradients are.* As a result, to counter the effect of potentially large delays, one needs to necessarily use a *conservative* step-size $\alpha$ that scales inversely with $\tau_{max}$. This is precisely what ends up slackening the final convergence rate. In sharp contrast, our proposed delay adaptive SA rule in (23), by design, rejects overly stale pseudo-gradients. This allows us to analyze (23) as a variant of the non-delayed SA rule in (3) with at most $O(\epsilon)$ error. To see this, observe that $\|\boldsymbol{\theta}_t - \boldsymbol{\theta}_{t-\tau_t}\| \leq \epsilon$ implies $\|\mathbf{g}(\boldsymbol{\theta}_t, o) - \mathbf{g}(\boldsymbol{\theta}_{t-\tau_t}, o)\| \leq L\epsilon$. Thus, every time we update, we move in the correct direction up to only a small amount of error. We formalize this key insight in Appendix B.3, where we provide the full proof of Theorem 3. In Appendix C, we also provide simulation results comparing non-delayed SA with the vanilla delayed SA update (6), and with the delay-adaptive algorithm (23).

## 6 CONCLUSIONS AND FUTURE WORK

We studied the interplay between delays and Markovian sampling in the context of general stochastic approximation with a contractive operator. Our analysis allowed for arbitrary time-varying (potentially random) delays that are uniformly bounded. Leveraging a novel inductive proof technique, we provided the first non-asymptotic convergence result for this setting, obtaining a rate that has a *tight* dependence on both the maximum delay $\tau_{max}$ and the mixing time of the underlying Markov chain, $\tau_{mix}$. Furthermore, we proposed the first delay-adaptive SA scheme that features two distinct advantages relative to a vanilla delayed SA protocol: (i) the rates for the former depend on the average delay $\tau_{avg}$ as opposed to the maximum delay $\tau_{max}$; and (ii) implementing the delay-adaptive rule requires no knowledge whatsoever of the delay sequence. The insights and novel analysis techniques from our work pave the way for various interesting future research avenues. We discuss some of them below.

1. The most natural next step would be to consider the study of asynchronous multi-agent RL algorithms where delays are inevitable. In this context, a recent line of work on federated RL has revealed the benefits of cooperation among the agents by establishing *linear speedups* in the sample complexity of learning, despite Markovian sampling (Khodadadian et al., 2022; Dal Fabbro et al., 2023; Wang et al., 2023; Zhang et al., 2024). Can we continue to expect such speedups in the presence of delays? What if we have stragglers in the system that slow down the pace of computation? Do the ideas (Dutta et al., 2018; Reisizadeh et al., 2022) that apply in the context of distributed optimization/supervised learning to tackle stragglers carry over to the MARL setting? This remains an open direction worth exploring.

2. At a high level, the findings in this paper contribute to a robustness theory for iterative RL algorithms. While we specifically focused on robustness to delays, we believe that the tools that were developed in the process should apply to the study of robustness to other types of structured perturbations. For instance, in a recent paper, Mitra et al. (2023) showed that TD learning protocols (with linear function approximation) can be just as robust to extreme compression/sparsification as their SGD counterparts, e.g., the `SignSGD` algorithm (Bernstein et al., 2018). We conjecture that the inductive proof used to study delays here can simplify and sharpen the results in Mitra et al. (2023). We plan to verify this conjecture as part of future work.

3. Yet another avenue is to consider RL settings more general than the ones studied here. In particular, all our results relied on the strong monotone property in Assumption 2. Deriving tight rates under arbitrary delays in the absence of such an assumption will likely require some work. One could also seek to generalize our results to settings that involve nonlinear function approximation schemes (e.g., neural nets) as in Tian et al. (2023); Cayci et al. (2023), or to two-time-scale SA protocols that capture actor-critic methods.

4. Finally, delays in physical systems often adhere to some structure. For instance, one can imagine such delays being generated randomly from some distribution. If so, if the delays are generated in an i.i.d. manner, one should be able to compute estimates of the mean and higher-order moments (if they exist) of the delay distribution with high probability, given enough samples. Can we exploit such information to design algorithms with better bounds than in this paper? Intuition dictates that this should be possible. However, to our knowledge, this remains a relatively unexplored direction.

## Acknowledgements

This work was supported, in part, by the Italian Ministry of Education, University and Research through the PRIN Project under Grant 2017NS9FEY. It was also partially supported by NSF Award 1837253 and by the ARL grant DCIST CRA W911NF-17-2-0181. The research was also supported, in part, by NSF Grant ECCS-2335876 and NSF CAREER Grant 1943064.

## References

Agarwal, A. and Duchi, J. C. (2011). Distributed delayed stochastic optimization. *Advances in neural information processing systems*, 24.

Arjevani, Y., Shamir, O., and Srebro, N. (2020). A tight convergence analysis for stochastic gradient descent with delayed updates. In *Algorithmic Learning Theory*, pages 111–132. PMLR.

Arjovsky, M., Chintala, S., and Bottou, L. (2017). Wasserstein generative adversarial networks. In *International conference on machine learning*, pages 214–223. PMLR.

Assran, M., Aytekin, A., Feyzmahdavian, H. R., Johansson, M., and Rabbat, M. G. (2020). Advances in asynchronous parallel and distributed optimization. *Proceedings of the IEEE*, 108(11):2013–2031.

Bernstein, J., Wang, Y.-X., Azizzadenesheli, K., and Anandkumar, A. (2018). signsgd: Compressed optimisation for non-convex problems. In *International Conference on Machine Learning*, pages 560–569. PMLR.

Bertsekas, D. P. and Tsitsiklis, J. N. (1989). Convergence rate and termination of asynchronous iterative algorithms. In *Proceedings of the 3rd International Conference on Supercomputing*, pages 461–470.

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

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

Bouteiller, Y., Ramstedt, S., Beltrame, G., Pal, C., and Binas, J. (2020). Reinforcement learning with random delays. In *International conference on learning representations*.

Cayci, S., Satpathi, S., He, N., and Srikant, R. (2023). Sample complexity and overparameterization bounds for temporal difference learning with neural network approximation. *IEEE Transactions on Automatic Control*.

Chen, M., Bai, Y., Poor, H. V., and Wang, M. (2023a). Efficient RL with impaired observability: Learning to act with delayed and missing state observations. *arXiv preprint arXiv:2306.01243*.

Chen, Z., Maguluri, S. T., and Zubeldia, M. (2023b). Concentration of contractive stochastic approximation: Additive and multiplicative noise. *arXiv preprint arXiv:2303.15740*.

Chen, Z., Zhang, S., Doan, T. T., Clarke, J.-P., and Maguluri, S. T. (2022). Finite-sample analysis of nonlinear stochastic approximation with applications in reinforcement learning. *Automatica*, 146:110623.

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

Dal Fabbro, N., Mitra, A., and Pappas, G. J. (2023). Federated td learning over finite-rate erasure channels: Linear speedup under markovian sampling. *IEEE Control Systems Letters*.

Doan, T. T. (2022). Finite-time analysis of markov gradient descent. *IEEE Transactions on Automatic Control*.

Doan, T. T. (2023). Finite-time convergence rates of distributed local stochastic approximation. *Automatica*, 158:111294.

Dutta, S., Joshi, G., Ghosh, S., Dube, P., and Nagpurkar, P. (2018). Slow and stale gradients can win the race: Error-runtime trade-offs in distributed sgd. In *International conference on artificial intelligence and statistics*, pages 803–812. PMLR.

Even, M. (2023). Stochastic gradient descent under markovian sampling schemes. *arXiv preprint arXiv:2302.14428*.

Feyzmahdavian, H. R., Aytekin, A., and Johansson, M. (2014). A delayed proximal gradient method with linear convergence rate. In *2014 IEEE international workshop on machine learning for signal processing (MLSP)*, pages 1–6. IEEE.

Feyzmahdavian, H. R., Aytekin, A., and Johansson, M. (2016). An asynchronous mini-batch algorithm for regularized stochastic optimization. *IEEE Transactions on Automatic Control*, 61(12):3740–3754.

Glasgow, M. R. and Wootters, M. (2022). Asynchronous distributed optimization with stochastic delays. In *International Conference on Artificial Intelligence and Statistics*, pages 9247–9279. PMLR.

Gurbuzbalaban, M., Ozdaglar, A., and Parrilo, P. A. (2017). On the convergence rate of incremental aggregated gradient algorithms. *SIAM Journal on Optimization*, 27(2):1035–1048.

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

Koloskova, A., Stich, S. U., and Jaggi, M. (2022). Sharper convergence guarantees for asynchronous sgd for distributed and federated learning. *Advances in Neural Information Processing Systems*, 35:17202–17215.

Meyn, S. (2023). Stability of q-learning through design and optimism. *arXiv preprint arXiv:2307.02632*.

Mitra, A., Pappas, G. J., and Hassani, H. (2023). Temporal difference learning with compressed updates: Error-feedback meets reinforcement learning. *arXiv preprint arXiv:2301.00944*.

Mnih, V., Badia, A. P., Mirza, M., Graves, A., Lillicrap, T., Harley, T., Silver, D., and Kavukcuoglu, K. (2016). Asynchronous methods for deep reinforcement learning. In *International conference on machine learning*, pages 1928–1937. PMLR.

Nagaraj, D., Wu, X., Bresler, G., Jain, P., and Netrapalli, P. (2020). Least squares regression with markovian data: Fundamental limits and algorithms. *Advances in neural information processing systems*, 33:16666–16676.

Nguyen, J., Malik, K., Zhan, H., Yousefpour, A., Rabbat, M., Malek, M., and Huba, D. (2022). Federated learning with buffered asynchronous aggregation. In *International Conference on Artificial Intelligence and Statistics*, pages 3581–3607. PMLR.

Pike-Burke, C., Agrawal, S., Szepesvari, C., and Grunewalder, S. (2018). Bandits with delayed, aggregated anonymous feedback. In *International Conference on Machine Learning*, pages 4105–4113. PMLR.

Qu, G. and Wierman, A. (2020). Finite-time analysis of asynchronous stochastic approximation and $q$-learning. In *Conference on Learning Theory*, pages 3185–3205. PMLR.

Reisizadeh, A., Tziotis, I., Hassani, H., Mokhtari, A., and Pedarsani, R. (2022). Straggler-resilient federated learning: Leveraging the interplay between statistical accuracy and system heterogeneity. *IEEE Journal on Selected Areas in Information Theory*, 3(2):197–205.

Robbins, H. and Monro, S. (1951). A stochastic approximation method. *The Annals of Mathematical Statistics*, 22(3):400–407.

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

Stich, S. U. and Karimireddy, S. P. (2020). The error-feedback framework: Better rates for sgd with delayed gradients and compressed updates. *The Journal of Machine Learning Research*, 21(1):9613–9648.

Thune, T. S., Cesa-Bianchi, N., and Seldin, Y. (2019). Nonstochastic multiarmed bandits with unrestricted delays. *Advances in Neural Information Processing Systems*, 32.

Tian, H., Paschalidis, I. C., and Olshevsky, A. (2023). On the performance of temporal difference learning with neural networks. *arXiv preprint arXiv:2312.05397*.

Tsitsiklis, J., Bertsekas, D., and Athans, M. (1986). Distributed asynchronous deterministic and stochastic gradient optimization algorithms. *IEEE Transactions on Automatic Control*, 31(9):803–812.

Tsitsiklis, J. and Vanroy, B. (1997). An analysis of temporal-difference learning with function approximation. *IEEE Transactions on Automatic Control*, 42(5):674–690.

Vernade, C., Carpentier, A., Lattimore, T., Zappella, G., Ermis, B., and Brueckner, M. (2020). Linear bandits with stochastic delayed feedback. In *International Conference on Machine Learning*, pages 9712–9721. PMLR.

Wang, H., Mitra, A., Hassani, H., Pappas, G. J., and Anderson, J. (2023). Federated temporal difference learning with linear function approximation under environmental heterogeneity. *arXiv:2302.02212*.

Zeng, S., Doan, T. T., and Romberg, J. (2022). Finite-time convergence rates of decentralized stochastic approximation with applications in multi-agent and multi-task learning. *IEEE Transactions on Automatic Control*.

Zhang, C., Wang, H., Mitra, A., and Anderson, J. (2024). Finite-time analysis of on-policy heterogeneous federated reinforcement learning. *arXiv preprint arXiv:2401.15273*.

Zhou, Z., Mertikopoulos, P., Bambos, N., Glynn, P., Ye, Y., Li, L.-J., and Fei-Fei, L. (2018). Distributed asynchronous optimization with unbounded delays: How slow can you go? In *International Conference on Machine Learning*, pages 5970–5979. PMLR.

## Checklist

1. For all models and algorithms presented, check if you include:
   (a) A clear description of the mathematical setting, assumptions, algorithm, and/or model. Yes
   (b) An analysis of the properties and complexity (time, space, sample size) of any algorithm. Yes
   (c) (Optional) Anonymized source code, with specification of all dependencies, including external libraries. Not Applicable

2. For any theoretical claim, check if you include:
   (a) Statements of the full set of assumptions of all theoretical results. Yes
   (b) Complete proofs of all theoretical results. Yes
   (c) Clear explanations of any assumptions. Yes

3. For all figures and tables that present empirical results, check if you include:
   (a) The code, data, and instructions needed to reproduce the main experimental results (either in the supplemental material or as a URL). Not Applicable
   (b) All the training details (e.g., data splits, hyperparameters, how they were chosen). Not Applicable
   (c) A clear definition of the specific measure or statistics and error bars (e.g., with respect to the random seed after running experiments multiple times). Not Applicable
   (d) A description of the computing infrastructure used. (e.g., type of GPUs, internal cluster, or cloud provider). Not Applicable

4. If you are using existing assets (e.g., code, data, models) or curating/releasing new assets, check if you include:
   (a) Citations of the creator If your work uses existing assets. Not Applicable
   (b) The license information of the assets, if applicable. Not Applicable
   (c) New assets either in the supplemental material or as a URL, if applicable. Not Applicable
   (d) Information about consent from data providers/curators. Not Applicable
   (e) Discussion of sensible content if applicable, e.g., personally identifiable information or offensive content. Not Applicable

5. If you used crowdsourcing or conducted research with human subjects, check if you include:
   (a) The full text of instructions given to participants and screenshots.Not Applicable
   (b) Descriptions of potential participant risks, with links to Institutional Review Board (IRB) approvals if applicable. Not Applicable
   (c) The estimated hourly wage paid to participants and the total amount spent on participant compensation. Not Applicable

# SUPPLEMENTARY MATERIALS

## A Related Work

In this section, we review selected works related to the existing literature on delays in optimization, bandits, and reinforcement learning (RL).

### A.1 Delays in Optimization

The study of delays and asynchrony in optimization has been a topic of interest since the seminal work by Bertsekas and Tsitsiklis (1989), which investigates convergence rates of asynchronous iterative algorithms in parallel or distributed computing systems. Subsequently, many researchers have explored the effects of delay and asynchrony on various learning and optimization methods. We summarize some of the significant works in this area below.

Agarwal and Duchi (2011) focus on distributed delayed stochastic optimization, specifically gradient-based optimization algorithms that rely on delayed stochastic gradient information. They analyze the convergence of such algorithms and propose procedures to overcome communication bottlenecks and synchronization requirements. Their work demonstrates that delays are asymptotically negligible, achieving order-optimal convergence results for smooth stochastic problems in distributed optimization settings.

Stich and Karimireddy (2020) introduce an error-feedback framework, which examines stochastic gradient descent (SGD) with delayed updates on smooth quasi-convex and non-convex functions. They derive non-asymptotic convergence rates and show that the delay only linearly slows down the higher-order deterministic term, while the stochastic term remains unaffected. This result illustrates the robustness of SGD to delayed stochastic gradient updates, improving upon previous rates for different forms of delayed gradients. Notably, this work provides the best-known rate for SGD with i.i.d. noise. It is worth mentioning that most existing literature has focused on bounds depending only on the maximum delay. However, the recent works of Cohen et al. (2021) and Koloskova et al. (2022) have explored convergence rates that depend on the average delay sequence.

The aforementioned studies on delays in optimization contribute to understanding the impact of delays and asynchrony in various optimization algorithms. They provide insights into the convergence properties and shed light on the robustness of these methods to different forms of delay. Nevertheless, there is still a gap in the literature regarding the finite-time convergence rates of delayed stochastic approximation schemes under Markovian sampling/noise.

### A.2 Delays in Bandits

There has been significant research efforts on the impact of delays in bandits. Some of the key works in this area are summarized in the following.

Non-stochastic multi-armed bandits with unrestricted delays were studied in Thune et al. (2019). The authors prove that the "delayed" Exp3 algorithm achieves the $\mathcal{O}(\sqrt{(KT+D)\ln K})$ regret bound for variable but bounded delays. They also introduce a new algorithm that handles delays without prior knowledge of the total delay, achieving the same regret bound. The paper provides insights into the regret bounds for bandit problems with delays.

The challenges of stochastic linear bandits with delayed feedback, where the feedback is randomly delayed and delays are only partially observable, were addressed in Vernade et al. (2020). The authors propose computationally efficient algorithms, OTFLinUCB and OTFLinTS, capable of integrating new information as it becomes available and handling permanently censored feedback. The authors prove optimal regret bounds for the proposed algorithms and validate their findings through experiments on simulated and real data.

Another paper investigates a variant of the stochastic $K$-armed bandit problem called "bandits with delayed, aggregated anonymous feedback" (Pike-Burke et al., 2018). In this setting, the player observes only the sum of a number of previously generated rewards that arrive in each round, and the information of which arm led to a particular reward is lost. The authors provide an algorithm that achieves the same worst-case regret as in the non-anonymous problem when the delays are bounded.

The above papers demonstrate that it is possible to design algorithms that can achieve good performance in the presence of delays in bandits. However, the performance metric of interest in these papers is typically cumulative regret. It is important to note here that the conclusions drawn for such a regret metric do not necessarily have implications for the sample-complexity bounds we care about in this work in the context of delayed stochastic approximation.

### A.3 Delays in RL

Until recently, the field of reinforcement learning had not thoroughly explored the impact of delays. In what follows, we highlight some key research works in this area.

Bouteiller et al. (2020) conducted a study on reinforcement learning with random delays, specifically focusing on environments with delays in actions and observations. They introduced the Delay-Correcting Actor-Critic (DCAC) algorithm, which incorporates off-policy multi-step value estimation to accommodate delays. Through theoretical analysis and practical experiments using a delay-augmented version of the MuJoCo continuous control benchmark, the authors demonstrated that DCAC outperforms other algorithms in delayed environments. In this work, however, no finite-time convergence analysis is provided for the analyzed algorithms.

Mnih et al. (2016) introduced asynchronous methods for deep reinforcement learning. They presented a lightweight framework that utilizes asynchronous gradient descent to optimize deep neural network controllers. The authors showed that parallel actor-learners have a stabilizing effect on training and achieve superior performance compared to state-of-the-art methods in domains such as Atari games and continuous motor control problems. Note that, however, although large measurement campaigns are conducted, no finite-time convergence analysis is performed in this work.

Chen et al. (2023a) studied the problem of policy learning in environments with delayed or missing observations. They showed that it is possible to learn a near-optimal policy in this setting, even though the agent does not have access to the most recent state of the system. They established near-optimal regret bounds for this case. Note that this work focuses on regret analysis, while ours is focused on the impact of the joint effect of delays and Markovian sampling on the finite-time rate of convergence to the SA solution, which we study through the lenses of analyzing the interplay between maximum/average delay and mixing time on the convergence rates we provide.

# B Proofs of the Theorems

In this Appendix, we provide the proofs for the theoretical results stated in the paper. We start by recalling some implications of the Assumptions of Section 2 in the following.

### Preliminaries

First, recall that from Assumption 2 we have, $\forall \boldsymbol{\theta} \in \mathbb{R}^d$:

$$\langle \boldsymbol{\theta}^* - \boldsymbol{\theta}, \bar{\mathbf{g}}(\boldsymbol{\theta}) \rangle \geq \mu \|\boldsymbol{\theta}^* - \boldsymbol{\theta}\|^2. \tag{24}$$

Throughout the proof, we will often invoke the mixing property (see Definition 1), which implies that, for a fixed $\boldsymbol{\theta}$, the following is true:

$$\|\mathbb{E}\left[\mathbf{g}(\boldsymbol{\theta}, o_t) | o_{t-\tau_{mix}}\right] - \bar{\mathbf{g}}(\boldsymbol{\theta})\| \leq \alpha \left(\|\boldsymbol{\theta}\| + \sigma\right). \tag{25}$$

We will also use the fact that the SA update directions and their steady-state versions are $L$-Lipschitz (Assumption 3), i.e., $\forall o \in \{o_t\}_{t\in\mathbb{N}}$, and $\forall \boldsymbol{\theta}, \boldsymbol{\theta}' \in \mathbb{R}^d$, we have:

$$\begin{aligned} &\|\bar{\mathbf{g}}(\boldsymbol{\theta}) - \bar{\mathbf{g}}(\boldsymbol{\theta}')\| \leq L\|\boldsymbol{\theta} - \boldsymbol{\theta}'\|, \text{ and} \\ &\|\mathbf{g}(\boldsymbol{\theta}, o_t) - \mathbf{g}(\boldsymbol{\theta}', o_t)\| \leq L\|\boldsymbol{\theta} - \boldsymbol{\theta}'\|. \end{aligned} \tag{26}$$

We further have

$$\|\mathbf{g}(\boldsymbol{\theta}, o)\| \leq L(\|\boldsymbol{\theta}\| + \sigma), \forall o \in \{o_t\}_{t\in\mathbb{N}}, \forall \boldsymbol{\theta} \in \mathbb{R}^d. \tag{27}$$

Given that $(x+y)^2 \leq 2(x^2+y^2), \forall x, y \in \mathbb{R}$, we will often use the following inequality:

$$\|\mathbf{g}(\boldsymbol{\theta}, o_t)\|^2 \leq L^2(\|\boldsymbol{\theta}\| + \sigma)^2 \leq 2L^2(\|\boldsymbol{\theta}\|^2 + \sigma^2). \tag{28}$$

Without loss of generality, we assume that

$$L \geq 1, \qquad \sigma \geq \max\{\|\boldsymbol{\theta}_0\|, \|\boldsymbol{\theta}^*\|\}, \qquad \mu < 1. \tag{29}$$

We will often use the fact that, for any $x, y \in \mathbb{R}$, we have

$$xy \leq \frac{1}{2}(x^2 + y^2). \tag{30}$$

In addition, we will often use the fact that, for $t \geq 2$, $a_i \in \mathbb{R}, i = 0, ..., t-1$, it holds

$$\left(\sum_{i=0}^{t-1} a_i\right)^2 \leq t \sum_{i=0}^{t-1} a_i^2 \tag{31}$$

## B.1 Proof of Theorem 1 and Related Lemmas

First, we recall the definition of the SA recursion with constant delay:

$$\boldsymbol{\theta}_{t+1} = \begin{cases} \boldsymbol{\theta}_0 & \text{if } \ 0 \leq t < \tau \\ \boldsymbol{\theta}_t + \alpha \mathbf{g}(\boldsymbol{\theta}_{t-\tau}, o_{t-\tau}). & \text{if } \ t \geq \tau \end{cases} \tag{32}$$

For analysis purposes, we define a virtual iterate, $\tilde{\boldsymbol{\theta}}_t$. This virtual iterate is updated with the SA update direction without delays, and it is defined as follows:

$$\tilde{\boldsymbol{\theta}}_{t+1} = \tilde{\boldsymbol{\theta}}_t + \alpha \mathbf{g}(\boldsymbol{\theta}_t, o_t), \ \tilde{\boldsymbol{\theta}}_0 = \boldsymbol{\theta}_0. \tag{33}$$

We also introduce the related error term $\mathbf{d}_t$, which is the gap between the virtual iterate and the actual iterate.

$$\tilde{\boldsymbol{\theta}}_t = \boldsymbol{\theta}_t + \mathbf{d}_t, \quad \text{with } \mathbf{d}_0 = \mathbf{0}. \tag{34}$$

From the definition of $\tilde{\boldsymbol{\theta}}_t$, we can write the following recursions for $\mathbf{d}_t$, for $t \geq 0$:

$$\mathbf{d}_{t+1} = \mathbf{d}_t + \alpha(\mathbf{g}(\boldsymbol{\theta}_t, o_t) - \mathbf{g}(\boldsymbol{\theta}_{t-\tau}, o_{t-\tau})). \tag{35}$$

We define $\mathbf{g}(\boldsymbol{\theta}_l, o_l) = \boldsymbol{\theta}_l = \mathbf{d}_l = \mathbf{0}$ for $l < 0$. We also define $\tilde{r}_t = \|\tilde{\boldsymbol{\theta}}_t - \boldsymbol{\theta}^*\|$. For convenience, we define $\tilde{r}_t = 0$ for $t < 0$.

### B.1.1 Auxiliary Lemmas

Here, we present the main Lemmas needed to prove Theorem 1. We start with three bounds on $\|\mathbf{d}_t\|$, $\|\mathbf{d}_t\|^2$ and $\sum_{t=0}^{T} w_t\|\mathbf{d}_t\|^2$, as follows.

**Lemma 4.** *The following three inequalities hold:*

$$(i) \qquad \|\mathbf{d}_t\| \leq \alpha\tau L\sigma + \alpha L \sum_{l=t-\tau}^{t-1} \|\boldsymbol{\theta}_l\|, \tag{36}$$

$$(ii) \qquad \|\mathbf{d}_t\|^2 \leq 2\alpha^2\tau^2 L^2\sigma^2 + 2\alpha^2\tau L^2 \sum_{l=t-\tau}^{t-1} \|\boldsymbol{\theta}_l\|^2, \tag{37}$$

$$(iii) \quad \sum_{t=0}^{T} w_t\|\mathbf{d}_t\|^2 \leq 4W_T\alpha^2\tau^2 L^2\sigma^2 + 16\alpha^2\tau^2 L^2 \sum_{t=0}^{T} w_t\|\tilde{\boldsymbol{\theta}}_t\|^2, \tag{38}$$

*where (iii) holds for* $\alpha \leq \frac{1}{4\tau L}$.

Part (*iii*) of this Lemma is key to obtain the bound in (104). In the next Lemma, we provide bounds on the terms $\|\tilde{\boldsymbol{\theta}}_t - \tilde{\boldsymbol{\theta}}_{t-\tau_{mix}}\|$ and $\|\tilde{\boldsymbol{\theta}}_t - \tilde{\boldsymbol{\theta}}_{t-\tau_{mix}}\|^2$.

**Lemma 5.** *For any* $t \geq \tau_{mix}$, *we have*

$$(i) \qquad \|\tilde{\boldsymbol{\theta}}_{t-\tau_{mix}} - \tilde{\boldsymbol{\theta}}_t\| \leq L\alpha\sigma\tau_{mix} + L\alpha \sum_{l=t-\tau_{mix}}^{t-1} \|\boldsymbol{\theta}_l\|. \tag{39}$$

$$(ii) \qquad \|\tilde{\boldsymbol{\theta}}_{t-\tau_{mix}} - \tilde{\boldsymbol{\theta}}_t\|^2 \leq 2L^2\alpha^2\tau_{mix}^2\sigma^2 + 2L^2\alpha^2\tau_{mix} \sum_{l=t-\tau_{mix}}^{t-1} \|\boldsymbol{\theta}_l\|^2. \tag{40}$$

Note that this Lemma is a variation of Lemma 3 in Srikant and Ying (2019), which is key to invoke mixing time arguments to get finite-time convergence bounds in existing non-delayed SA analysis. Let us define

$$\begin{aligned} n_t &\triangleq \|\mathbf{g}(\boldsymbol{\theta}_t, o_t)\|^2, \\ m_t &\triangleq \langle \mathbf{g}(\boldsymbol{\theta}_t, o_t) - \mathbf{g}(\tilde{\boldsymbol{\theta}}_t, o_t), \tilde{\boldsymbol{\theta}}_t - \boldsymbol{\theta}^\star \rangle, \\ h_t &\triangleq \langle \tilde{\boldsymbol{\theta}}_t - \boldsymbol{\theta}^*, \mathbf{g}(\tilde{\boldsymbol{\theta}}_t, o_t) - \bar{\mathbf{g}}(\tilde{\boldsymbol{\theta}}_t) \rangle. \end{aligned} \tag{41}$$

To obtain a bound in the form (96), we need to bound $\mathbb{E}[h_t]$ properly, for which, in turn, we need Lemma 5. Furthermore, note that, in contrast to Srikant and Ying (2019), the bound is obtained for the sequence of *virtual iterates*. In the next lemma, we provide bounds for the three key terms of the bound in (90), i.e., $\|\mathbf{g}(\boldsymbol{\theta}_t, o_t)\|^2$, $m_t$, and $\mathbb{E}[h_t]$.

**Lemma 6.** *For all* $t \geq 0$, *we have*

$$(i) \quad n_t \leq 4L^2\|\mathbf{d}_t\|^2 + 8L^2\tilde{r}_t^2 + 10L^2\sigma^2, \tag{42}$$

$$(ii) \quad m_t \leq 6\alpha\tau L^2\sigma^2 + 3\alpha\tau L^2\tilde{r}_t^2 + 2\alpha L^2 \sum_{l=t-\tau}^{t-1} \left(\|\mathbf{d}_l\|^2 + 2\tilde{r}_l^2\right), \tag{43}$$

$$(iii) \quad \mathbb{E}[h_t] \leq \begin{cases} 111\sigma^2 L^2, \ \textit{for } 0 \leq t \leq \tau_{mix} \\ 4\alpha\tau_{mix}L^2\left(8\sigma^2 + 3\mathbb{E}\left[\tilde{r}_t^2\right]\right) + 8\alpha L^2 \sum_{l=t-\tau_{mix}}^{t-1} \mathbb{E}\left[\|\mathbf{d}_l\|^2 + 2\tilde{r}_l^2\right], \ \textit{for } t \geq \tau_{mix} \end{cases} \tag{44}$$

*where (iii) holds for* $\alpha \leq \frac{1}{36L^2\tau_{mix}}$.

The proof of this last Lemma relies on the bound on $\|\mathbf{d}_t\|$ established in Lemma 4. The proof of (*iii*) relies on the mixing properties of the Markov chain $\{o_t\}$ and on the bounds on $\|\boldsymbol{\theta}_t - \boldsymbol{\theta}_{t-\tau_{mix}}\|$ and $\|\boldsymbol{\theta}_t - \boldsymbol{\theta}_{t-\tau_{mix}}\|^2$ established in Lemma 5. Part (*iii*) is the key and most challenging part of the proof, which allows us to get to the bound in (96). Using this last Lemma, in combination with Lemma 4, we are able to get the bound in (12). The conclusion of the proof is enabled by using $\mathbb{E}\left[r_t^2\right] \leq 2\mathbb{E}\left[\tilde{r}_t^2\right] + 2\mathbb{E}\left[\|\mathbf{d}_t\|^2\right]$ and some further manipulations.

### B.1.2 Proofs of Auxiliary Lemmas

We first state and prove the following lemma, which we will use later in the proof of Theorem 1.

**Lemma 7.** *For $w_t \triangleq (1-0.5\mu\alpha)^{-(t+1)}$ with $\alpha \leq \frac{\mu}{C\bar{\tau}}$, $C \geq 2$, the following inequality holds for $0 \leq i \leq 2\bar{\tau}$, and for any $t$,*

$$w_t \leq 2w_{t-i}. \tag{45}$$

*Proof.*

$$\begin{aligned}
w_t &= w_{t-i}\left(1-\frac{\mu\alpha}{2}\right)^{-i} \\
&\overset{(a)}{\leq} w_{t-i}\left(1-\frac{\mu^2}{2C\bar{\tau}}\right)^{-i} \\
&\overset{(b)}{\leq} w_{t-i}\left(1-\frac{\mu^2}{2C\bar{\tau}}\right)^{-\bar{\tau}} \\
&\overset{(c)}{\leq} w_{t-i}\left(1-\frac{1}{4\bar{\tau}}\right)^{-\bar{\tau}} \\
&\overset{(d)}{\leq} w_{t-i}\left(1+\frac{1}{2\bar{\tau}}\right)^{\bar{\tau}} \\
&\overset{(e)}{\leq} w_{t-i}\exp\left(\frac{1}{2}\right) \\
&\leq 2w_{t-i}.
\end{aligned} \tag{46}$$

In $(a)$, we used the bound on $\alpha$; in $(b)$, we used the bound on $i$; in $(c)$, we used $\mu < 1$ and $C \geq 2$; in $(d)$, we used

$$(1-x)^{-1} \leq (1+2x) \text{ for } \quad 0 \leq x \leq \frac{1}{2}, \tag{47}$$

and for $(e)$, we used $(1+x)^k \leq \exp(xk)$ for $k \geq 0$. □

Note that we defined $\mathbf{g}(\boldsymbol{\theta}_i, o_i) = \mathbf{0}$ for $i < 0$, $\boldsymbol{\theta}_t = \mathbf{0}$ for $t < 0$, and $\mathbf{d}_t = \mathbf{0}$ for $t < 0$. First, note that, starting from the definition of $\mathbf{d}_t$ in (35),

$$\begin{aligned}
\mathbf{d}_{t+1} &= \mathbf{d}_t + \alpha\left(\mathbf{g}(\boldsymbol{\theta}_t, o_t) - \mathbf{g}(\boldsymbol{\theta}_{t-\tau}, o_{t-\tau})\right) \\
&= \mathbf{d}_{t-1} + \alpha\left(\mathbf{g}(\boldsymbol{\theta}_{t-1}, o_{t-1}) - \mathbf{g}(\boldsymbol{\theta}_{t-1-\tau}, o_{t-1-\tau})\right) \\
&\qquad + \alpha\left(\mathbf{g}(\boldsymbol{\theta}_t, o_t) - \mathbf{g}(\boldsymbol{\theta}_{t-\tau}, o_{t-\tau})\right) \\
&= \mathbf{d}_0 \quad + \alpha\sum_{l=0}^{t}\left(\mathbf{g}(\boldsymbol{\theta}_l, o_l) - \mathbf{g}(\boldsymbol{\theta}_{l-\tau}, o_{l-\tau})\right) \\
&\overset{(*)}{=} \mathbf{0} \quad + \alpha\sum_{l=t-\tau+1}^{t}\mathbf{g}(\boldsymbol{\theta}_l, o_l),
\end{aligned} \tag{48}$$

where $(*)$ follows because the overlapping terms in the sum cancel out. So, we obtain, for all $t \geq 0$,

$$\mathbf{d}_t = \alpha\sum_{l=t-\tau}^{t-1}\mathbf{g}(\boldsymbol{\theta}_l, o_l). \tag{49}$$

We can now prove Lemma 4, which is key to proving Theorem 1.

**Proof of Lemma 4 - (i), (ii).** From (49), using the triangle inequality and the bound on the update direc-

tion (27), we get, recalling that $\sigma \geq \|\boldsymbol{\theta}_0\|$,

$$\begin{aligned} \|\mathbf{d}_t\| &= \|\alpha \sum_{l=t-\tau}^{t-1} \mathbf{g}(\boldsymbol{\theta}_l, o_l)\| \\ &\overset{(27)}{\leq} \alpha L \sum_{l=t-\tau}^{t-1} (\|\boldsymbol{\theta}_l\|+\sigma) \\ &\leq \alpha\tau L\sigma + \alpha L \sum_{l=t-\tau}^{t-1} \|\boldsymbol{\theta}_l\|, \end{aligned} \tag{50}$$

which proves (i). We now prove (ii). Using the triangle inequality and (31),

$$\begin{aligned} \|\mathbf{d}_t\|^2 &= \|\alpha \sum_{l=t-\tau}^{t-1} \mathbf{g}(\boldsymbol{\theta}_l, o_l)\|^2 \\ &\overset{(31)}{\leq} \alpha^2\tau \sum_{l=t-\tau}^{t-1} \|\mathbf{g}(\boldsymbol{\theta}_l, o_l)\|^2. \end{aligned} \tag{51}$$

Now, using the upper bound on the squared gradient norm (28),

$$\begin{aligned} \|\mathbf{d}_t\|^2 &\leq \alpha^2\tau \sum_{l=t-\tau}^{t-1} \|\mathbf{g}(\boldsymbol{\theta}_l, o_l)\|^2 \\ &\leq 2\alpha^2\tau L^2 \sum_{l=t-\tau}^{t-1} (\|\boldsymbol{\theta}_l\|^2+\sigma^2) \\ &\leq 2\alpha^2\tau^2 L^2\sigma^2 + 2\alpha^2\tau L^2 \sum_{l=t-\tau}^{t-1} \|\boldsymbol{\theta}_l\|^2, \end{aligned} \tag{52}$$

which concludes the proof. □

Using the above inequalities, we can now prove part (*iii*) of Lemma 4.

**Proof of Lemma 4 - (iii).** First, recall that, from Lemma 4, we have

$$\|\mathbf{d}_t\|^2 \leq 2\alpha^2\tau^2 L^2\sigma^2 + 2\alpha^2\tau L^2 \sum_{l=t-\tau}^{t-1} \|\boldsymbol{\theta}_l\|^2. \tag{53}$$

Based on Lemma 7, for $0 \leq i \leq 2\bar{\tau}$, we have $w_t \leq 2w_{t-i}$ (see (46)). Using (53),

$$\begin{aligned}
\sum_{t=0}^{T} w_t \|\mathbf{d}_t\|^2 &\leq \sum_{t=0}^{T} w_t \left( 2\alpha^2\tau^2 L^2\sigma^2 + 2\alpha^2\tau L^2 \sum_{l=t-\tau}^{t-1} \|\boldsymbol{\theta}_l\|^2 \right) \\
&\leq 2W_T\alpha^2\tau^2L^2\sigma^2 + 2\alpha^2\tau L^2 \sum_{t=0}^{T} w_t \sum_{l=t-\tau}^{t-1} \|\boldsymbol{\theta}_l\|^2 \\
&\overset{(*)}{\leq} 2W_T\alpha^2\tau^2L^2\sigma^2 + 4\alpha^2\tau L^2 \sum_{t=0}^{T} \sum_{l=t-\tau}^{t-1} w_l \|\boldsymbol{\theta}_l\|^2 \\
&\overset{(**)}{\leq} 2W_T\alpha^2\tau^2L^2\sigma^2 + 4\alpha^2\tau^2 L^2 \sum_{t=0}^{T} w_t \|\boldsymbol{\theta}_t\|^2 \\
&\leq 2W_T\alpha^2\tau^2L^2\sigma^2 + 8\alpha^2\tau^2 L^2 \sum_{t=0}^{T} w_t \left( \|\tilde{\boldsymbol{\theta}}_t\|^2 + \|\mathbf{d}_t\|^2 \right) \\
&\leq 2W_T\alpha^2\tau^2L^2\sigma^2 + 8\alpha^2\tau^2 L^2 \sum_{t=0}^{T} w_t \|\tilde{\boldsymbol{\theta}}_t\|^2 + \frac{1}{2} \sum_{t=0}^{T} w_t \|\mathbf{d}_t\|^2,
\end{aligned} \tag{54}$$

where for $(*)$ we used the fact that $w_t \leq 2w_l$ for $t - 2\bar{\tau} \leq l \leq t-1$, and for $(**)$ we used the fact that each element $w_l\|\boldsymbol{\theta}_l\|^2$ appears at most $\tau$ times in the sum, for $l = 0, ..., T-1$ (note that, by definition, $\boldsymbol{\theta}_l = 0$ for $l < 0$). In the last inequality, we used $\alpha \leq \frac{1}{4\tau L}$. We can conclude getting

$$\sum_{t=0}^{T} w_t \|\mathbf{d}_t\|^2 \leq 4W_T\alpha^2\tau^2L^2\sigma^2 + 16\alpha^2\tau^2L^2 \sum_{t=0}^{T} w_t \|\tilde{\boldsymbol{\theta}}_t\|^2. \tag{55}$$

□

We now prove Lemma 5, that provides a bound on the norm of the gap $\|\tilde{\boldsymbol{\theta}}_{t-\tau_{mix}} - \tilde{\boldsymbol{\theta}}_t\|$ and its squared version $\|\tilde{\boldsymbol{\theta}}_{t-\tau_{mix}} - \tilde{\boldsymbol{\theta}}_t\|^2$.

**Proof of Lemma 5.** Inequality (i) of the Lemma can be easily proved by applying the definition of the recursion (33):

$$\begin{aligned}
\|\tilde{\boldsymbol{\theta}}_{t-\tau_{mix}} - \tilde{\boldsymbol{\theta}}_t\| &\leq \sum_{l=t-\tau_{mix}}^{t-1} \|\tilde{\boldsymbol{\theta}}_{l+1} - \tilde{\boldsymbol{\theta}}_l\| \\
&\leq \alpha \sum_{l=t-\tau_{mix}}^{t-1} \|\mathbf{g}(\boldsymbol{\theta}_l, o_l)\| \\
&\leq L\alpha \sum_{l=t-\tau_{mix}}^{t-1} (\|\boldsymbol{\theta}_l\| + \sigma) \\
&= L\alpha\sigma\tau_{mix} + L\alpha \sum_{l=t-\tau_{mix}}^{t-1} \|\boldsymbol{\theta}_l\|.
\end{aligned} \tag{56}$$

Similarly, for inequality (ii), note that, squaring equation (56),

$$\|\tilde{\boldsymbol{\theta}}_{t-\tau_{mix}} - \tilde{\boldsymbol{\theta}}_t\|^2 \leq 2L^2\alpha^2\tau_{mix}^2\sigma^2 + 2L^2\alpha^2\tau_{mix} \sum_{l=t-\tau_{mix}}^{t-1} \|\boldsymbol{\theta}_l\|^2. \tag{57}$$

□

We now prove Lemma 6, which provide bounds for $\|\mathbf{g}(\boldsymbol{\theta}_t, o_t)\|^2$, $m_t$, and $\mathbb{E}[h_t]$.

**Proof of Lemma 6 - (i).** From (28), we have $\|\mathbf{g}(\boldsymbol{\theta}_t, o_t)\|^2 \leq 2L^2(\|\boldsymbol{\theta}_t\|^2+\sigma^2)$, and so

$$\begin{aligned} n_t = \|\mathbf{g}(\boldsymbol{\theta}_t, o_t)\|^2 &\leq 2L^2(\|\boldsymbol{\theta}_t\|^2+\sigma^2) \\ &\leq 2L^2\|\boldsymbol{\theta}_t - \tilde{\boldsymbol{\theta}}_t + \tilde{\boldsymbol{\theta}}_t\|^2+2L^2\sigma^2 \\ &\leq 4L^2\|\mathbf{d}_t\|^2+4L^2\|\tilde{\boldsymbol{\theta}}_t\|^2+2L^2\sigma^2 \\ &\leq 4L^2\|\mathbf{d}_t\|^2+4L^2\|\tilde{\boldsymbol{\theta}}_t - \boldsymbol{\theta}^* + \boldsymbol{\theta}^*\|^2+2L^2\sigma^2 \\ &\leq 4L^2\|\mathbf{d}_t\|^2+8L^2\tilde{r}_t^2 + 8L^2\|\boldsymbol{\theta}^*\|^2+2L^2\sigma^2 \\ &\leq 4L^2\|\mathbf{d}_t\|^2+8L^2\tilde{r}_t^2 + 10L^2\sigma^2, \end{aligned} \tag{58}$$

where we used $\|\boldsymbol{\theta}^*\| \leq \sigma$. This concludes the proof. □

**Proof of Lemma 6 - (ii).** By the Cauchy-Schwarz inequality, Lipschitz continuity of $\mathbf{g}(\boldsymbol{\theta}, o_t)$ in $\boldsymbol{\theta}$ (see (26)), and from the definition of $\mathbf{d}_t$, we get

$$\begin{aligned} m_t &= \langle \mathbf{g}(\boldsymbol{\theta}_t, o_t) - \mathbf{g}(\tilde{\boldsymbol{\theta}}_t, o_t), \tilde{\boldsymbol{\theta}}_t - \boldsymbol{\theta}^\star \rangle \\ &\leq \|\mathbf{g}(\boldsymbol{\theta}_t, o_t) - \mathbf{g}(\tilde{\boldsymbol{\theta}}_t, o_t)\| \|\tilde{\boldsymbol{\theta}}_t - \boldsymbol{\theta}^*\| \\ &\leq L\|\tilde{\boldsymbol{\theta}}_t - \boldsymbol{\theta}_t\| \|\tilde{\boldsymbol{\theta}}_t - \boldsymbol{\theta}^*\| \\ &= L\|\mathbf{d}_t\|\tilde{r}_t. \end{aligned} \tag{59}$$

Applying Lemma 4 to bound $\|\mathbf{d}_t\|$, we get

$$\begin{aligned} m_t &\leq L\left(\alpha\tau L\sigma + \alpha L \sum_{l=t-\tau}^{t-1} \|\boldsymbol{\theta}_l\|\right)\tilde{r}_t \\ &= \alpha\tau L^2\sigma\tilde{r}_t + \alpha L^2 \sum_{l=t-\tau}^{t-1} \|\boldsymbol{\theta}_l\|\tilde{r}_t \\ &\overset{(30)}{\leq} 2\alpha\tau L^2\sigma^2 + 2\alpha\tau L^2\tilde{r}_t^2 + \alpha L^2 \sum_{l=t-\tau}^{t-1} \left(\|\boldsymbol{\theta}_l\|^2+\tilde{r}_t^2\right) \\ &= 2\alpha\tau L^2\sigma^2 + 3\alpha\tau L^2\tilde{r}_t^2 + \alpha L^2 \sum_{l=t-\tau}^{t-1} \|\boldsymbol{\theta}_l\|^2 \\ &\overset{(31)}{\leq} 2\alpha\tau L^2\sigma^2 + 3\alpha\tau L^2\tilde{r}_t^2 + 2\alpha L^2 \sum_{l=t-\tau}^{t-1} \left(\|\mathbf{d}_l\|^2+\|\tilde{\boldsymbol{\theta}}_l\|^2\right) \\ &\leq 6\alpha\tau L^2\sigma^2 + 3\alpha\tau L^2\tilde{r}_t^2 + 2\alpha L^2 \sum_{l=t-\tau}^{t-1} \|\mathbf{d}_l\|^2+4\alpha L^2 \sum_{l=t-\tau}^{t-1} \tilde{r}_l^2. \end{aligned} \tag{60}$$

□

Next, we provide the proof of Lemma 6, which, in turn, provides a bound for $\mathbb{E}[h_t]$ - the term related to Markovian sampling whose analysis requires special care and mixing time arguments.

**Proof of Lemma 6 - (iii).** We start with the case $0 \leq t \leq \tau_{mix}$. Note that, using (28),

$$\begin{aligned} h_t &= \langle \tilde{\boldsymbol{\theta}}_t - \boldsymbol{\theta}^*, \mathbf{g}(\tilde{\boldsymbol{\theta}}_t, o_t) - \bar{\mathbf{g}}(\tilde{\boldsymbol{\theta}}_t) \rangle \\ &\leq \|\tilde{\boldsymbol{\theta}}_t - \boldsymbol{\theta}^*\| \|\mathbf{g}(\tilde{\boldsymbol{\theta}}_t, o_t) - \bar{\mathbf{g}}(\tilde{\boldsymbol{\theta}}_t)\| \\ &\overset{(30)}{\leq} \frac{1}{2}\tilde{r}_t^2 + \frac{1}{2}\|\mathbf{g}(\tilde{\boldsymbol{\theta}}_t, o_t) - \bar{\mathbf{g}}(\tilde{\boldsymbol{\theta}}_t)\|^2 \\ &\overset{(31)}{\leq} \frac{1}{2}\tilde{r}_t^2 + \|\mathbf{g}(\tilde{\boldsymbol{\theta}}_t, o_t)\|^2 + \|\bar{\mathbf{g}}(\tilde{\boldsymbol{\theta}}_t)\|^2 \\ &\overset{(28)}{\leq} \frac{\tilde{r}_t^2}{2} + 2L^2\|\tilde{\boldsymbol{\theta}}_t\|^2 + 2L^2\sigma^2 + 2L^2\|\tilde{\boldsymbol{\theta}}_t\|^2 + 2L^2\sigma^2 \\ &\leq \frac{\tilde{r}_t^2}{2} + 8L^2\tilde{r}_t^2 + 12L^2\sigma^2 \\ &\leq 9L^2\tilde{r}_t^2 + 12L^2\sigma^2. \end{aligned} \tag{61}$$

Recall that

$$\boldsymbol{\theta}_{t+1} = \begin{cases} \boldsymbol{\theta}_0 & \text{if } 0 \leq t < \tau \\ \boldsymbol{\theta}_t + \alpha \mathbf{g}(\boldsymbol{\theta}_{t-\tau}, o_{t-\tau}) & \text{if } t \geq \tau \end{cases}, \tag{62}$$

from which we can write, for $t \geq \tau$,

$$\begin{aligned} r_{t+1}^2 &= r_t^2 + 2\alpha\langle \boldsymbol{\theta}_t - \boldsymbol{\theta}^*, \mathbf{g}(\boldsymbol{\theta}_{t-\tau}, o_{t-\tau}) \rangle + \alpha^2 \|\mathbf{g}(\boldsymbol{\theta}_{t-\tau}, o_{t-\tau})\|^2 \\ &\overset{(30)}{\leq} r_t^2 + \alpha r_t^2 + \alpha\|\mathbf{g}(\boldsymbol{\theta}_{t-\tau}, o_{t-\tau})\|^2 + \alpha^2\|\mathbf{g}(\boldsymbol{\theta}_{t-\tau}, o_{t-\tau})\|^2 \\ &\overset{\alpha<1}{\leq} (1+\alpha) r_t^2 + 2\alpha\|\mathbf{g}(\boldsymbol{\theta}_{t-\tau}, o_{t-\tau})\|^2, \end{aligned} \tag{63}$$

and note that, for $t < \tau$, $r_{t+1}^2 = r_t^2$, and hence (63) holds true for all $t \geq 0$. Now note that

$$\|\mathbf{g}(\boldsymbol{\theta}_{t-\tau}, o_{t-\tau})\|^2 \leq 2L^2(\|\boldsymbol{\theta}_{t-\tau}\|^2 + \sigma^2) \leq 4L^2 r_{t-\tau}^2 + 6L^2\sigma^2. \tag{64}$$

Therefore, we can write

$$r_{t+1}^2 \leq (1+\alpha) r_t^2 + 8\alpha L^2 r_{t-\tau}^2 + 12\alpha L^2\sigma^2. \tag{65}$$

Now, we show that, for $k < \tau_{mix}$,

$$r_k^2 \leq \rho^k r_0^2 + \epsilon_k, \tag{66}$$

with $\epsilon_k = \rho\epsilon_{k-1} + \beta$, $\epsilon_0 = 0$, where $\rho = 1 + \alpha + 8\alpha L^2 > 1$, and $\beta = 12\alpha L^2\sigma^2$. We show it by induction. The base case $k = 0$ is trivially true. Now suppose that inequality (66) is true up to some $k \geq 0$, thus

$$r_s^2 \leq \rho^s r_0^2 + \epsilon_s, \quad \forall s \leq k. \tag{67}$$

We can get, noting that, for all $k$, $0 \leq \epsilon_k \leq \epsilon_{k+1}$,

$$\begin{aligned} r_{k+1}^2 &\leq (1+\alpha) r_k^2 + 8\alpha L^2 r_{k-\tau} + 12\alpha L^2\sigma^2 \\ &\leq (1+\alpha)(\rho^k r_0^2 + \epsilon_k) + 8\alpha L^2(\rho^k r_0^2 + \epsilon_k) + 12\alpha L^2\sigma^2 \\ &= (1+\alpha+8\alpha L^2)\rho^k r_0^2 + (1+\alpha+8\alpha L^2)\epsilon_k + 12\alpha L^2\sigma^2 \\ &= \rho^{k+1} r_0^2 + \rho\epsilon_k + \beta \\ &= \rho^{k+1} r_0^2 + \epsilon_{k+1}, \end{aligned} \tag{68}$$

which concludes the induction proof of (66). Now note that, given that $L \geq 1$, $\rho \leq 1 + 9\alpha L^2$, and, for $\alpha \leq \frac{1}{36L^2\tau_{mix}}$,

$$\rho^k \leq (1+9\alpha L^2)^k \leq (1+9\alpha L^2)^{\tau_{mix}} \leq e^{9\alpha L^2\tau_{mix}} \leq e^{0.25} \leq 2. \tag{69}$$

Also note that, for all $k \leq \tau_{mix}$,

$$\epsilon_k = \beta\sum_{j=0}^{k-1}(1+9\alpha L^2)^j \leq \beta\sum_{j=0}^{\tau_{mix}-1}(1+9\alpha L^2)^{\tau_{mix}} \leq 2\beta\tau_{mix}, \tag{70}$$

and we can get, for all $k \leq \tau_{mix}$, noting that $r_0^2 \leq 4\sigma^2$,

$$r_k^2 \leq 2r_0^2 + 2\beta\tau_{mix} = 2r_0^2 + 24\alpha L^2\sigma^2\tau_{mix} \leq 9\sigma^2. \tag{71}$$

Now note that, similarly to the calculations performed above, for $t < \tau_{mix}$,

$$\begin{aligned}\tilde{r}_{t+1}^2 &= \tilde{r}_t^2 + 2\alpha\langle \boldsymbol{\theta}_t - \boldsymbol{\theta}^*, \mathbf{g}(\boldsymbol{\theta}_t, o_t)\rangle + \alpha^2\|\mathbf{g}(\boldsymbol{\theta}_t, o_t)\|^2 \\ &\leq (1+\alpha)\tilde{r}_t^2 + 2\alpha\|\mathbf{g}(\boldsymbol{\theta}_t, o_t)\|^2 \\ &\overset{(28)}{\leq} (1+\alpha)\tilde{r}_t^2 + 4\alpha L^2(\|\boldsymbol{\theta}_t\|^2 + \sigma^2) \\ &\leq (1+\alpha)\tilde{r}_t^2 + 4\alpha L^2(2r_t^2 + 3\sigma^2).\end{aligned} \tag{72}$$

Using the bound established in (71), we can get

$$\begin{aligned}\tilde{r}_{t+1}^2 &\leq (1+\alpha)\tilde{r}_t^2 + 8\alpha L^2 r_t^2 + 12\alpha L^2\sigma^2 \\ &\leq (1+\alpha)\tilde{r}_t^2 + 84\alpha L^2\sigma^2.\end{aligned} \tag{73}$$

From this, we can proceed as follows:

$$\begin{aligned}\tilde{r}_{t+1}^2 &\leq (1+\alpha)\tilde{r}_t^2 + 84\alpha L^2\sigma^2 \\ &\leq (1+\alpha)^2\tilde{r}_{t-1}^2 + (1+\alpha)84\alpha L^2\sigma^2 + 84\alpha L^2\sigma^2 \\ &\leq (1+\alpha)^{t+1}\tilde{r}_0^2 + 84\alpha L^2\sigma^2\sum_{j=0}^{t}(1+\alpha)^j.\end{aligned} \tag{74}$$

So, for $0 \leq t < \tau_{mix}$,

$$\tilde{r}_{t+1}^2 \leq (1+\alpha)^{\tau_{mix}}\tilde{r}_0^2 + 84\alpha L^2\sigma^2\sum_{j=0}^{\tau_{mix}}(1+\alpha)^j. \tag{75}$$

Now, given that $L \geq 1$, note that, for $\alpha\tau_{mix} \leq \frac{1}{36L^2}$ and $j = 0, ..., \tau_{mix} - 1$, we have $(1+\alpha)^j \leq (1+\alpha)^{\tau_{mix}} \leq e^{\alpha\tau_{mix}} \leq e^{0.25} \leq 2$. Thus, we get

$$\tilde{r}_t^2 \leq 2\tilde{r}_0^2 + 84\alpha L^2\sigma^2\tau_{mix} \leq 11\sigma^2. \tag{76}$$

Finally,

$$\begin{aligned}h_t &\leq 9L^2\tilde{r}_t^2 + 12L^2\sigma^2 \\ &\leq 9L^2(11\sigma^2) + 12L^2\sigma^2 \\ &\leq 111L^2\sigma^2.\end{aligned} \tag{77}$$

We now analyze the case in which $t \geq \tau_{mix}$. Adding and subtracting $\tilde{\boldsymbol{\theta}}_{t-\tau_{mix}}$ in the left hand side of the inner product, we have

$$\begin{aligned}h_t &= \langle \tilde{\boldsymbol{\theta}}_t - \boldsymbol{\theta}^*, \mathbf{g}(\tilde{\boldsymbol{\theta}}_t, o_t) - \bar{\mathbf{g}}(\tilde{\boldsymbol{\theta}}_t)\rangle \\ &= \underbrace{\langle \tilde{\boldsymbol{\theta}}_t - \tilde{\boldsymbol{\theta}}_{t-\tau_{mix}}, \mathbf{g}(\tilde{\boldsymbol{\theta}}_t, o_t) - \bar{\mathbf{g}}(\tilde{\boldsymbol{\theta}}_t)\rangle}_{T_1} + \underbrace{\langle \tilde{\boldsymbol{\theta}}_{t-\tau_{mix}} - \boldsymbol{\theta}^*, \mathbf{g}(\tilde{\boldsymbol{\theta}}_t, o_t) - \bar{\mathbf{g}}(\tilde{\boldsymbol{\theta}}_t)\rangle}_{T_2},\end{aligned} \tag{78}$$

where, using (27), Cauchy-Schwarz inequality and Lemma 5,

$$
\begin{aligned}
T_1 &\leq \|\tilde{\boldsymbol{\theta}}_t - \tilde{\boldsymbol{\theta}}_{t-\tau_{mix}}\|(\|\mathbf{g}(\tilde{\boldsymbol{\theta}}_t, o_t)\|+\|\bar{\mathbf{g}}(\tilde{\boldsymbol{\theta}}_t)\|) \\
&\overset{(27)}{\leq} \|\tilde{\boldsymbol{\theta}}_t - \tilde{\boldsymbol{\theta}}_{t-\tau_{mix}}\| 2L(\|\tilde{\boldsymbol{\theta}}_t\|+\sigma) \\
&\leq 2\alpha L^2 \left(\sigma\tau_{mix} + \sum_{l=t-\tau_{mix}}^{t-1} \|\boldsymbol{\theta}_l\|\right)(\|\tilde{\boldsymbol{\theta}}_t\|+\sigma) \\
&\leq 2\alpha L^2 \sigma\tau_{mix}(\|\tilde{\boldsymbol{\theta}}_t\|+\sigma) + 2\alpha L^2 \sum_{l=t-\tau_{mix}}^{t-1} \|\boldsymbol{\theta}_l\|(\|\tilde{\boldsymbol{\theta}}_t\|+\sigma) \\
&\overset{(30)}{\leq} 2\alpha L^2\sigma^2\tau_{mix} + 2\alpha L^2\tau_{mix}\sigma\|\tilde{\boldsymbol{\theta}}_t\| \\
&+ 2\alpha L^2 \sum_{l=t-\tau_{mix}}^{t-1} \left(\frac{1}{2}\|\boldsymbol{\theta}_l\|^2+\frac{1}{2}(\|\tilde{\boldsymbol{\theta}}_t\|+\sigma)^2\right) \\
&\overset{(31)}{\leq} 2\alpha L^2\sigma^2\tau_{mix} + \alpha L^2\tau_{mix}\sigma^2 + \alpha L^2\tau_{mix}\|\tilde{\boldsymbol{\theta}}_t\|^2 \\
&+ 2\alpha L^2 \sum_{l=t-\tau_{mix}}^{t-1} \left(\frac{1}{2}\|\boldsymbol{\theta}_l\|^2+\|\tilde{\boldsymbol{\theta}}_t\|^2+\sigma^2\right) \\
&\leq 11\alpha L^2\sigma^2\tau_{mix} + 6\alpha L^2\tau_{mix}\tilde{r}_t^2 + \alpha L^2 \sum_{l=t-\tau_{mix}}^{t-1} \|\boldsymbol{\theta}_l\|^2.
\end{aligned} \tag{79}
$$

So, taking the expectation,

$$
\mathbb{E}\left[T_1\right] \leq 11\alpha L^2\sigma^2\tau_{mix} + 6\alpha L^2\tau_{mix}\mathbb{E}\left[\tilde{r}_t^2\right] + \alpha L^2 \sum_{l=t-\tau_{mix}}^{t-1} \mathbb{E}\left[\|\boldsymbol{\theta}_l\|^2\right]. \tag{80}
$$

Now, we focus on $T_2$. Note that, adding and subtracting $\mathbf{g}(\tilde{\boldsymbol{\theta}}_{t-\tau_{mix}}, o_t)$ and $\bar{\mathbf{g}}(\tilde{\boldsymbol{\theta}}_{t-\tau_{mix}})$ to the right hand side of the inner product, we can write

$$
\begin{aligned}
T_2 &= \langle \tilde{\boldsymbol{\theta}}_{t-\tau_{mix}} - \boldsymbol{\theta}^*, \mathbf{g}(\tilde{\boldsymbol{\theta}}_t, o_t) - \bar{\mathbf{g}}(\tilde{\boldsymbol{\theta}}_t)\rangle \\
&= \bar{T}_1 + \bar{T}_2 + \bar{T}_3
\end{aligned} \tag{81}
$$

with

$$
\begin{aligned}
\bar{T}_1 &= \langle \tilde{\boldsymbol{\theta}}_{t-\tau_{mix}} - \boldsymbol{\theta}^*, \mathbf{g}(\tilde{\boldsymbol{\theta}}_{t-\tau_{mix}}, o_t) - \bar{\mathbf{g}}(\tilde{\boldsymbol{\theta}}_{t-\tau_{mix}})\rangle \\
\bar{T}_2 &= \langle \tilde{\boldsymbol{\theta}}_{t-\tau_{mix}} - \boldsymbol{\theta}^*, \mathbf{g}(\tilde{\boldsymbol{\theta}}_t, o_t) - \mathbf{g}(\tilde{\boldsymbol{\theta}}_{t-\tau_{mix}}, o_t)\rangle \\
\bar{T}_3 &= \langle \tilde{\boldsymbol{\theta}}_{t-\tau_{mix}} - \boldsymbol{\theta}^*, \bar{\mathbf{g}}(\tilde{\boldsymbol{\theta}}_{t-\tau_{mix}}) - \bar{\mathbf{g}}(\tilde{\boldsymbol{\theta}}_t)\rangle.
\end{aligned} \tag{82}
$$

We first bound $\bar{T}_2$ and $\bar{T}_3$. Note that, using the Lipschitz property of the TD update direction (26) and Lemma 5,

$$\begin{aligned}
\bar{T}_2 &\leq \|\tilde{\boldsymbol{\theta}}_{t-\tau_{mix}} - \boldsymbol{\theta}^*\| \|\mathbf{g}(\tilde{\boldsymbol{\theta}}_t, o_t) - \mathbf{g}(\tilde{\boldsymbol{\theta}}_{t-\tau_{mix}}, o_t)\| \\
&\leq L\|\tilde{\boldsymbol{\theta}}_{t-\tau_{mix}} - \boldsymbol{\theta}^*\| \|\tilde{\boldsymbol{\theta}}_{t-\tau_{mix}} - \tilde{\boldsymbol{\theta}}_t\| \\
&\leq L\|\tilde{\boldsymbol{\theta}}_{t-\tau_{mix}} - \tilde{\boldsymbol{\theta}}_t + \tilde{\boldsymbol{\theta}}_t - \boldsymbol{\theta}^*\| \|\tilde{\boldsymbol{\theta}}_{t-\tau_{mix}} - \tilde{\boldsymbol{\theta}}_t\| \\
&\leq L\tilde{r}_t\|\tilde{\boldsymbol{\theta}}_{t-\tau_{mix}} - \tilde{\boldsymbol{\theta}}_t\| + L\|\tilde{\boldsymbol{\theta}}_{t-\tau_{mix}} - \tilde{\boldsymbol{\theta}}_t\|^2 \\
&\leq L^2\alpha\left(\sigma\tau_{mix} + \sum_{l=t-\tau_{mix}}^{t-1}\|\boldsymbol{\theta}_l\|\right)\tilde{r}_t + L\left(2L^2\alpha^2\tau_{mix}^2\sigma^2 + 2L^2\alpha^2\tau_{mix}\sum_{l=t-\tau_{mix}}^{t-1}\|\boldsymbol{\theta}_l\|^2\right) \\
&= L^2\alpha\tau_{mix}\frac{1}{2}\left(\sigma^2 + \tilde{r}_t^2\right) + \frac{1}{2}L^2\alpha\sum_{l=t-\tau_{mix}}^{t-1}\left(\|\boldsymbol{\theta}_l\|^2 + \tilde{r}_t^2\right) \\
&+ 2L^3\alpha^2\tau_{mix}^2\sigma^2 + 2L^3\alpha^2\tau_{mix}\sum_{l=t-\tau_{mix}}^{t-1}\|\boldsymbol{\theta}_l\|^2 \\
&\leq \alpha\tau_{mix}L^2\sigma^2 + \alpha\tau_{mix}L^2\tilde{r}_t^2 + \alpha L^2\sum_{l=t-\tau_{mix}}^{t-1}\|\boldsymbol{\theta}_l\|^2,
\end{aligned} \tag{83}$$

where in the last inequality we used $\alpha \leq \frac{1}{8\tau_{mix}L}$. Taking the expectation,

$$\mathbb{E}\left[\bar{T}_2\right] \leq \alpha\tau_{mix}L^2\sigma^2 + \alpha\tau_{mix}L^2\mathbb{E}\left[\tilde{r}_t^2\right] + \alpha L^2\sum_{l=t-\tau_{mix}}^{t-1}\mathbb{E}\left[\|\boldsymbol{\theta}_l\|^2\right]. \tag{84}$$

With the same calculations, we can get

$$\mathbb{E}\left[\bar{T}_3\right] \leq \alpha\tau_{mix}L^2\sigma^2 + \alpha\tau_{mix}L^2\mathbb{E}\left[\tilde{r}_t^2\right] + \alpha L^2\sum_{l=t-\tau_{mix}}^{t-1}\mathbb{E}\left[\|\boldsymbol{\theta}_l\|^2\right]. \tag{85}$$

We now proceed to bound $\bar{T}_1$.

$$\begin{aligned}
\mathbb{E}\left[\bar{T}_1\right] &= \mathbb{E}\left[\langle\tilde{\boldsymbol{\theta}}_{t-\tau_{mix}} - \boldsymbol{\theta}^*, \mathbf{g}(\tilde{\boldsymbol{\theta}}_{t-\tau_{mix}}, o_t) - \bar{\mathbf{g}}(\tilde{\boldsymbol{\theta}}_{t-\tau_{mix}})\rangle\right] \\
&= \mathbb{E}\left[\langle\tilde{\boldsymbol{\theta}}_{t-\tau_{mix}} - \boldsymbol{\theta}^*, \mathbb{E}\left[\mathbf{g}(\tilde{\boldsymbol{\theta}}_{t-\tau_{mix}}, o_t)|o_{t-\tau_{mix}}, \tilde{\boldsymbol{\theta}}_{t-\tau_{mix}}\right] - \bar{\mathbf{g}}(\tilde{\boldsymbol{\theta}}_{t-\tau_{mix}})\rangle\right] \\
&\leq \mathbb{E}\left[\|\tilde{\boldsymbol{\theta}}_{t-\tau_{mix}} - \boldsymbol{\theta}^*\| \|\mathbb{E}\left[\mathbf{g}(\tilde{\boldsymbol{\theta}}_{t-\tau_{mix}}, o_t)|o_{t-\tau_{mix}}, \tilde{\boldsymbol{\theta}}_{t-\tau_{mix}}\right] - \bar{\mathbf{g}}(\tilde{\boldsymbol{\theta}}_{t-\tau_{mix}})\|\right] \\
&\overset{(*)}{\leq} \alpha\mathbb{E}\left[\|\tilde{\boldsymbol{\theta}}_{t-\tau_{mix}} - \boldsymbol{\theta}^*\|(\|\tilde{\boldsymbol{\theta}}_{t-\tau_{mix}}\| + \sigma)\right] \\
&\leq \alpha\mathbb{E}\left[\|\tilde{\boldsymbol{\theta}}_{t-\tau_{mix}} - \boldsymbol{\theta}^*\|(\|\tilde{\boldsymbol{\theta}}_{t-\tau_{mix}} - \boldsymbol{\theta}^*\| + 2\sigma)\right] \\
&\leq \alpha\mathbb{E}\left[\frac{1}{2}\|\tilde{\boldsymbol{\theta}}_{t-\tau_{mix}} - \boldsymbol{\theta}^*\|^2 + \frac{1}{2}(\|\tilde{\boldsymbol{\theta}}_{t-\tau_{mix}} - \boldsymbol{\theta}^*\| + 2\sigma)^2\right] \\
&\leq \alpha\mathbb{E}\left[\frac{1}{2}\|\tilde{\boldsymbol{\theta}}_{t-\tau_{mix}} - \boldsymbol{\theta}^*\|^2 + \|\tilde{\boldsymbol{\theta}}_{t-\tau_{mix}} - \boldsymbol{\theta}^*\|^2 + 2\sigma^2\right] \\
&\leq 2\alpha\mathbb{E}\left[\|\tilde{\boldsymbol{\theta}}_{t-\tau_{mix}} - \boldsymbol{\theta}^*\|^2 + \sigma^2\right] \\
&\leq 2\alpha\mathbb{E}\left[2\|\tilde{\boldsymbol{\theta}}_t - \boldsymbol{\theta}^*\|^2 + 2\|\tilde{\boldsymbol{\theta}}_t - \tilde{\boldsymbol{\theta}}_{t-\tau_{mix}}\|^2 + \sigma^2\right] \\
&\leq 2\alpha\mathbb{E}\left[2\tilde{r}_t^2 + 2(2L^2\alpha^2\tau_{mix}^2\sigma^2 + 2L^2\alpha^2\tau_{mix}\sum_{l=t-\tau_{mix}}^{t-1}\|\boldsymbol{\theta}_l\|^2) + \sigma^2\right] \\
&\leq 4\alpha\mathbb{E}\left[\tilde{r}_t^2\right] + 3\alpha\sigma^2 + \alpha\sum_{l=t-\tau_{mix}}^{t-1}\mathbb{E}\left[\|\boldsymbol{\theta}_l\|^2\right],
\end{aligned} \tag{86}$$

where for $(*)$ we used Definition 1 of mixing time and the fact that $\sigma \geq 1$, and in the last inequality we used $\alpha \leq \frac{1}{8\tau_{mix}L}$. So, we get

$$\begin{aligned}\mathbb{E}\left[T_2\right] &= \mathbb{E}\left[\bar{T}_1\right] + \mathbb{E}\left[\bar{T}_2\right] + \mathbb{E}\left[\bar{T}_3\right] \\ &\leq 6\alpha\tau_{mix}L^2\mathbb{E}\left[\tilde{r}_t^2\right] + 5\alpha\tau_{mix}L^2\sigma^2 + 3\alpha L^2 \sum_{l=t-\tau_{mix}}^{t-1} \mathbb{E}\left[\|\boldsymbol{\theta}_l\|^2\right].\end{aligned} \tag{87}$$

Finally, we get

$$\begin{aligned}\mathbb{E}\left[h_t\right] &= \mathbb{E}\left[T_1\right] + \mathbb{E}\left[T_2\right] \\ &\leq 16\alpha\tau_{mix}L^2\sigma^2 + 12\alpha\tau_{mix}L^2\mathbb{E}\left[\tilde{r}_t^2\right] + 4\alpha L^2 \sum_{l=t-\tau_{mix}}^{t-1} \mathbb{E}\left[\|\boldsymbol{\theta}_l\|^2\right] \\ &\leq 16\alpha\tau_{mix}L^2\sigma^2 + 12\alpha\tau_{mix}L^2\mathbb{E}\left[\tilde{r}_t^2\right] \\ &+ 8\alpha L^2 \sum_{l=t-\tau_{mix}}^{t-1} \left(\mathbb{E}\left[\|\mathbf{d}_l\|^2\right] + \mathbb{E}\left[\|\tilde{\boldsymbol{\theta}}_l\|^2\right]\right) \\ &\leq 32\alpha\tau_{mix}L^2\sigma^2 + 12\alpha\tau_{mix}L^2\mathbb{E}\left[\tilde{r}_t^2\right] + 8\alpha L^2 \sum_{l=t-\tau_{mix}}^{t-1} \mathbb{E}\left[\|\mathbf{d}_l\|^2 + 2\tilde{r}_l^2\right].\end{aligned} \tag{88}$$

□

### B.1.3 Proof of Theorem 1

First, we have

$$\tilde{r}_{t+1}^2 = \tilde{r}_t^2 + 2\alpha\langle \mathbf{g}(\boldsymbol{\theta}_t, o_t), \tilde{\boldsymbol{\theta}}_t - \boldsymbol{\theta}^\star\rangle + \alpha^2\|\mathbf{g}(\boldsymbol{\theta}_t, o_t)\|^2. \tag{89}$$

Then, using (24), i.e., $\langle \bar{\mathbf{g}}(\tilde{\boldsymbol{\theta}}_t), \tilde{\boldsymbol{\theta}}_t - \boldsymbol{\theta}^\star\rangle \leq -\mu\tilde{r}_t^2$, we have

$$\begin{aligned}\tilde{r}_{t+1}^2 &= \tilde{r}_t^2 + 2\alpha\langle \mathbf{g}(\boldsymbol{\theta}_t, o_t), \tilde{\boldsymbol{\theta}}_t - \boldsymbol{\theta}^*\rangle + \alpha^2\|\mathbf{g}(\boldsymbol{\theta}_t, o_t)\|^2 \\ &= \tilde{r}_t^2 + 2\alpha\langle \bar{\mathbf{g}}(\tilde{\boldsymbol{\theta}}_t), \tilde{\boldsymbol{\theta}}_t - \boldsymbol{\theta}^\star\rangle + 2\alpha h_t + 2\alpha m_t + \alpha^2\|\mathbf{g}(\boldsymbol{\theta}_t, o_t)\|^2 \\ &\leq (1 - 2\alpha\mu)\tilde{r}_t^2 + 2\alpha h_t + 2\alpha m_t + \alpha^2 n_t.\end{aligned} \tag{90}$$

We now apply the inequalities obtained in Lemma 6 to bound $\mathbb{E}\left[h_t\right]$, $m_t$, and $n_t$. Recall that $\bar{\tau} = \max\{\tau, \tau_{mix}\}$. Note that, from Lemma 6 - (iii), we can write $\mathbb{E}\left[h_t\right] \leq \bar{h}_t$, defining

$$\bar{h}_t = \begin{cases} B & \text{if } \ 0 \leq t < \tau_{mix} \\ q_t & \text{if } \ t \geq \tau_{mix} \end{cases}, \tag{91}$$

with $B = 111\sigma^2$, and

$$q_t = \alpha\tau_{mix}L^2\left(32\sigma^2 + 12\mathbb{E}\left[\tilde{r}_t^2\right]\right) + 8\alpha L^2 \sum_{l=t-\tau_{mix}}^{t-1} \mathbb{E}\left[\|\mathbf{d}_l\|^2 + 2\tilde{r}_l^2\right]. \tag{92}$$

As a consequence, we can write, for every $t \geq 0$,

$$\mathbb{E}\left[h_t\right] \leq q_t + \bar{B}_t, \tag{93}$$

where, in turn,

$$\bar{B}_t = \begin{cases} B & \text{if } \ 0 \leq t < \tau_{mix} \\ 0 & \text{otherwise} \end{cases}. \tag{94}$$

Also, recall that, from Lemma 6, we have

$$\begin{aligned} n_t &\leq 4L^2\|\mathbf{d}_t\|^2 + 8L^2\tilde{r}_t^2 + 10L^2\sigma^2, \\ m_t &\leq 6\alpha\tau L^2\sigma^2 + 3\alpha\tau L^2\tilde{r}_t^2 + 2\alpha L^2 \sum_{l=t-\tau}^{t-1} \left(\|\mathbf{d}_l\|^2 + 2\tilde{r}_l^2\right). \end{aligned} \tag{95}$$

Combining these inequalities together, we have, for $t \geq 0$,

$$\begin{aligned}\mathbb{E}\left[\tilde{r}_{t+1}^2\right] &\leq (1-2\alpha\mu)\mathbb{E}\left[\tilde{r}_t^2\right] + 2\alpha\mathbb{E}\left[h_t\right] + 2\alpha\mathbb{E}\left[m_t\right] + \alpha^2\mathbb{E}\left[\|\mathbf{g}(\boldsymbol{\theta}_t, o_t)\|^2\right] \\ &\leq (1-2\alpha\mu)\mathbb{E}\left[\tilde{r}_t^2\right] + 2\alpha^2\tau_{mix}L^2\left(32\sigma^2 + 12\mathbb{E}\left[\tilde{r}_t^2\right]\right) \\ &\quad + 16\alpha^2L^2\sum_{l=t-\tau_{mix}}^{t-1}\mathbb{E}\left[\|\mathbf{d}_l\|^2 + 2\tilde{r}_l^2\right] \\ &\quad + 12\alpha^2\tau L^2\sigma^2 + 6\alpha^2\tau L^2\mathbb{E}\left[\tilde{r}_t^2\right] + 4\alpha^2L^2\sum_{l=t-\tau}^{t-1}\mathbb{E}\left[\|\mathbf{d}_l\|^2 + 2\tilde{r}_l^2\right] \\ &\quad + 4\alpha^2L^2\mathbb{E}\left[\|\mathbf{d}_t\|^2 + 2\tilde{r}_t^2\right] + 10\alpha^2L^2\sigma^2 + 2\alpha\bar{B}_t.\end{aligned} \tag{96}$$

Combining terms, we can get

$$\begin{aligned}\mathbb{E}\left[\tilde{r}_{t+1}^2\right] &\leq (1-2\alpha\mu + 48\alpha^2L^2\bar{\tau})\mathbb{E}\left[\tilde{r}_t^2\right] + 128\alpha^2L^2\bar{\tau}\sigma^2 \\ &\quad + 4\alpha^2L^2\mathbb{E}\left[\|\mathbf{d}_t\|^2\right] + 20\alpha^2L^2\sum_{l=t-\bar{\tau}}^{t-1}\mathbb{E}\left[\|\mathbf{d}_l\|^2 + 2\tilde{r}_l^2\right] + 2\alpha\bar{B}_t,\end{aligned} \tag{97}$$

where we have used $\tau + \tau_{mix} \leq 2\bar{\tau}$. Now, using the fact that $r_t^2 \leq 2\tilde{r}_t^2 + 2\|\mathbf{d}_t\|^2$, which implies $-\tilde{r}_t^2 \leq -\frac{r_t^2}{2} + \|\mathbf{d}_t\|^2$, we have

$$\begin{aligned}(1-2\alpha\mu + 48\alpha^2L^2\bar{\tau})\mathbb{E}\left[\tilde{r}_t^2\right] &= (1-\alpha\mu + 48\alpha^2L^2\bar{\tau})\mathbb{E}\left[\tilde{r}_t^2\right] - \alpha\mu\mathbb{E}\left[\tilde{r}_t^2\right] \\ &\leq (1-\alpha\mu + 48\alpha^2L^2\bar{\tau})\mathbb{E}\left[\tilde{r}_t^2\right] - \alpha\mu\frac{\mathbb{E}\left[r_t^2\right]}{2} + \alpha\mathbb{E}\left[\|\mathbf{d}_t\|^2\right],\end{aligned} \tag{98}$$

and, using $\mu \leq 1$, we can re-write (97) as

$$\begin{aligned}\mathbb{E}\left[\tilde{r}_{t+1}^2\right] &\leq (1-\alpha\mu + 48\alpha^2L^2\bar{\tau})\mathbb{E}\left[\tilde{r}_t^2\right] - \alpha\mu\frac{\mathbb{E}\left[r_t^2\right]}{2} + 128\alpha^2L^2\bar{\tau}\sigma^2 \\ &\quad + \alpha(1+4\alpha L^2)\mathbb{E}\left[\|\mathbf{d}_t\|^2\right] + 20\alpha^2L^2\sum_{l=t-\bar{\tau}}^{t-1}\mathbb{E}\left[\|\mathbf{d}_l\|^2 + 2\tilde{r}_l^2\right] + 2\alpha\bar{B}_t.\end{aligned} \tag{99}$$

Multiplying both sides by $w_t$, we have

$$\begin{aligned}w_t\mathbb{E}\left[\tilde{r}_{t+1}^2\right] &\leq (1-\alpha\mu + 48\alpha^2L^2\bar{\tau})w_t\mathbb{E}\left[\tilde{r}_t^2\right] - \alpha\mu\frac{w_t\mathbb{E}\left[r_t^2\right]}{2} + 128w_t\alpha^2L^2\bar{\tau}\sigma^2 \\ &\quad + \alpha(1+4\alpha L^2)w_t\mathbb{E}\left[\|\mathbf{d}_t\|^2\right] + 20\alpha^2L^2w_t\sum_{l=t-\bar{\tau}}^{t-1}\mathbb{E}\left[\|\mathbf{d}_l\|^2 + 2\tilde{r}_l^2\right] + 2\alpha w_t\bar{B}_t.\end{aligned} \tag{100}$$

By summing over $t = 0, ..., T$, we get, with $W_T = \sum_{t=0}^T w_t$:

$$\begin{aligned}\sum_{t=0}^T w_t\mathbb{E}\left[\tilde{r}_{t+1}^2\right] &\leq (1-\alpha\mu + 48\alpha^2L^2\bar{\tau})\sum_{t=0}^T w_t\mathbb{E}\left[\tilde{r}_t^2\right] - \frac{\alpha\mu}{2}\sum_{t=0}^T w_t\mathbb{E}\left[r_t^2\right] \\ &\quad + 128W_T\alpha^2L^2\bar{\tau}\sigma^2 + \alpha(1+4\alpha L^2)\underbrace{\sum_{t=0}^T w_t\mathbb{E}\left[\|\mathbf{d}_t\|^2\right]}_{p_1} \\ &\quad + 20\alpha^2L^2\underbrace{\sum_{t=0}^T w_t\sum_{l=t-\bar{\tau}}^{t-1}\mathbb{E}\left[\|\mathbf{d}_l\|^2 + 2\tilde{r}_l^2\right]}_{p_2} + 2W_{\tau_{mix}-1}\alpha B.\end{aligned} \tag{101}$$

Note that, from Lemma 4 - (iii), we have, picking $\alpha \leq \frac{1}{72\tau L^2}$,

$$\begin{aligned} p_1 = \sum_{t=0}^{T} w_t \mathbb{E}\left[\|\mathbf{d}_t\|^2\right] &\leq 4W_T\alpha^2\tau^2L^2\sigma^2 + 16\alpha^2\tau^2L^2\sum_{t=0}^{T} w_t\mathbb{E}\left[\|\tilde{\boldsymbol{\theta}}_t\|^2\right] \\ &\leq 36W_T\alpha^2\tau^2L^2\sigma^2 + 32\alpha^2\tau^2L^2\sum_{t=0}^{T} w_t\mathbb{E}\left[\tilde{r}_t^2\right] \\ &\leq \frac{\alpha\tau W_T\sigma^2}{2} + \frac{\alpha\tau}{2}\sum_{t=0}^{T} w_t\mathbb{E}\left[\tilde{r}_t^2\right]. \end{aligned} \tag{102}$$

Furthermore, using the fact that $w_t \leq 2w_l$ for $l = t-\bar{\tau}, ..., t-1$, the above bound on $p_1$, and picking $\alpha \leq \frac{1}{72\tau L^2}$, we can bound $p_2$ as follows:

$$\begin{aligned} p_2 &= \sum_{t=0}^{T} w_t \sum_{l=t-\bar{\tau}}^{t-1} \mathbb{E}\left[\|\mathbf{d}_l\|^2 + 2\tilde{r}_l^2\right] \\ &\overset{(a)}{\leq} 2\sum_{t=0}^{T}\sum_{l=t-\bar{\tau}}^{t-1} w_l\mathbb{E}\left[\|\mathbf{d}_l\|^2 + 2\tilde{r}_l^2\right] \\ &\overset{(b)}{\leq} 2\bar{\tau}\sum_{t=0}^{T} w_t\mathbb{E}\left[\|\mathbf{d}_t\|^2 + 2\tilde{r}_t^2\right]. \\ &\leq 2\bar{\tau}\sum_{t=0}^{T} w_t\mathbb{E}\left[\|\mathbf{d}_t\|^2\right] + 4\bar{\tau}\sum_{t=0}^{T} w_t\mathbb{E}\left[\tilde{r}_t^2\right] \\ &\overset{(c)}{\leq} 2\bar{\tau}\left(\frac{\alpha\tau W_T\sigma^2}{2} + \frac{\alpha\tau}{2}\sum_{t=0}^{T} w_t\mathbb{E}\left[\tilde{r}_t^2\right]\right) + 4\bar{\tau}\sum_{t=0}^{T} w_t\mathbb{E}\left[\tilde{r}_t^2\right] \\ &\leq 5\bar{\tau}\sum_{t=0}^{T} w_t\mathbb{E}\left[\tilde{r}_t^2\right] + W_T\sigma^2\bar{\tau}, \end{aligned} \tag{103}$$

where for $(a)$ we used Lemma 7, for $(b)$ we used the fact that each element $w_l\|\boldsymbol{\theta}_l\|^2$ appears at most $\tau$ times in the sum, for $l = 0, ..., T-1$ (note that, by definition, $\|\mathbf{d}_l\| = \tilde{r}_l = 0$ for $l < 0$), and for $(c)$, we used the bound on $p_1$. In the last inequality we simply used $\alpha\tau \leq 1$. Plugging the two bounds on $p_1$ and $p_2$ in (101), we get

$$\begin{aligned} \sum_{t=0}^{T} w_t\mathbb{E}\left[\tilde{r}_{t+1}^2\right] \leq (1 - \alpha\mu + 150\alpha^2L^2\bar{\tau})\sum_{t=0}^{T} w_t\mathbb{E}\left[\tilde{r}_t^2\right] - \frac{\alpha\mu}{2}\sum_{t=0}^{T} w_t\mathbb{E}\left[r_t^2\right] \\ + 150W_T\alpha^2L^2\bar{\tau}\sigma^2 + 2W_{\tau_{mix}-1}\alpha B. \end{aligned} \tag{104}$$

Now, note that for $\alpha \leq \frac{\mu}{100L^2\bar{\tau}}$, it holds that $(1 - 2\alpha\mu + 150\alpha^2L^2\bar{\tau}) \leq (1 - 0.5\alpha\mu)$. We can then re-write (104) as

$$\begin{aligned} \sum_{t=0}^{T} w_t\mathbb{E}\left[\tilde{r}_{t+1}^2\right] \leq (1 - 0.5\alpha\mu)\sum_{t=0}^{T} w_t\mathbb{E}\left[\tilde{r}_t^2\right] - \frac{\alpha\mu}{2}\sum_{t=0}^{T} w_t\mathbb{E}\left[r_t^2\right] \\ + 150W_T\alpha^2L^2\bar{\tau}\sigma^2 + 2W_{\tau_{mix}-1}\alpha B. \end{aligned} \tag{105}$$

Now, dividing by $W_T$ both sides of (105), bringing $\sum_{t=0}^{T} w_t\mathbb{E}\left[\tilde{r}_{t+1}^2\right]$ to the right hand side of the inequality and $-\frac{\alpha\mu}{2}\sum_{t=0}^{T} w_t\mathbb{E}\left[r_t^2\right]$ to the left side, we get

$$\begin{aligned} \frac{\alpha\mu}{2}\sum_{t=0}^{T}\frac{w_t}{W_T}\mathbb{E}\left[r_t^2\right] \leq \frac{1}{W_T}\sum_{t=0}^{T}\left(w_t(1 - 0.5\alpha\mu)\mathbb{E}\left[\tilde{r}_t^2\right] - w_t\mathbb{E}\left[\tilde{r}_{t+1}^2\right]\right) \\ + 150\alpha^2L^2\bar{\tau}\sigma^2 + \frac{2W_{\tau_{mix}-1}\alpha B}{W_T}. \end{aligned} \tag{106}$$

Now, recalling that $w_t = (1-0.5\alpha\mu)^{-(t+1)}$, note that $w_t(1-0.5\alpha\mu) = w_{t-1}$, and we can get, noting that $w_{-1} = 1$,

$$\begin{aligned}\sum_{t=0}^{T}\left(w_t(1-0.5\alpha\mu)\mathbb{E}\left[\tilde{r}_t^2\right] - w_t\mathbb{E}\left[\tilde{r}_{t+1}^2\right]\right) &= \sum_{t=0}^{T}\left(w_{t-1}\mathbb{E}\left[\tilde{r}_t^2\right] - w_t\mathbb{E}\left[\tilde{r}_{t+1}^2\right]\right) \\ &\leq \mathbb{E}\left[\tilde{r}_0^2\right] - w_T\mathbb{E}\left[\tilde{r}_{T+1}^2\right] \leq \tilde{r}_0^2.\end{aligned} \tag{107}$$

Hence, we can write (106) as

$$\frac{\alpha\mu}{2}\sum_{t=0}^{T}\frac{w_t}{W_T}\mathbb{E}\left[r_t^2\right] \leq \frac{\tilde{r}_0^2}{W_T} + 150\alpha^2L^2\bar{\tau}\sigma^2 + \frac{2W_{\tau_{mix}-1}\alpha B}{W_T}. \tag{108}$$

Now note that

$$W_{\tau_{mix}-1} = \sum_{t=0}^{\tau_{mix}-1} w_t = \sum_{t=0}^{\tau_{mix}-1}(1-0.5\alpha\mu)^{-(t+1)} \leq \sum_{t=0}^{\tau_{mix}-1}(1+\alpha\mu)^{t+1} \leq 2\tau_{mix} \tag{109}$$

and that

$$\frac{1}{W_T} \leq \frac{1}{w_T} = (1-0.5\alpha\mu)^{T+1}, \tag{110}$$

from which we can obtain, re-arranging the different terms in (108),

$$\begin{aligned}\frac{1}{W_T}\sum_{t=0}^{T}w_t\mathbb{E}\left[r_{t+1}^2\right] &\leq (1-0.5\alpha\mu)^{T+1}\tilde{r}_0^2\left(\frac{2}{\alpha\mu} + \frac{4\bar{\tau}B}{\mu}\right) + 300\frac{\alpha L^2\bar{\tau}\sigma^2}{\mu} \\ &= C_\alpha(1-0.5\alpha\mu)^{T+1}\tilde{r}_0^2 + C_2\frac{\alpha L^2\bar{\tau}\sigma^2}{\mu},\end{aligned} \tag{111}$$

where we define $C_\alpha = \left(\frac{2}{\alpha\mu} + \frac{4\bar{\tau}B}{\mu}\right)$ and $C_2 = 300$. By plugging the maximum value for the step size $\alpha = \frac{\mu}{150L^2\bar{\tau}}$, we can get, defining $C_{\bar{\tau}} = \frac{\bar{\tau}}{\mu}\left(\frac{2C_1L^2}{\mu} + 4B\right)$,

$$\frac{1}{W_T}\sum_{t=0}^{T}w_t\mathbb{E}\left[r_{t+1}^2\right] \leq C_2(1-0.5\alpha\mu)^{T+1}\tilde{r}_0^2 + 2\sigma^2. \tag{112}$$

Indeed, for $\alpha = \frac{\mu}{150L^2\bar{\tau}}$, it holds $C_2\frac{\alpha L^2\bar{\tau}}{\mu} = 2$. Finally, by definition of $\boldsymbol{\theta}_{out}$ in Theorem 1, note that we have

$$\mathbb{E}\left[\|\boldsymbol{\theta}_{out} - \boldsymbol{\theta}^*\|^2\right] = \frac{1}{W_T}\sum_{t=0}^{T}w_t\mathbb{E}\left[r_t^2\right], \tag{113}$$

and we can conclude the proof of Theorem 1. □

### B.2 Proof of Theorem 2 and Related Lemmas

Let $r_t \triangleq \|\boldsymbol{\theta}_t - \boldsymbol{\theta}^*\|$. Define $\tau' = 2\tau_{max} + \tau_{mix}$, and recall

$$r_{t,2} \triangleq \max_{t-\tau' \leq l \leq t} \mathbb{E}\left[r_l^2\right]. \tag{114}$$

As we mentioned in Section 3, the proof technique of Theorem 1 cannot be directly extended to a time-varying delay setting. To see why this is the case, please take a look back at (49) in the proof of Theorem 1, and note that the specific way in which we can write $\mathbf{d}_t$, which is crucial in several key steps of the proof, *relies on the fact that the delay is constant*. Accordingly, we do not see how we could generalize that type of identity to a time-varying delay setting in a way that leads to the convergence rate we are aiming to obtain. In the rest of this section, we provide the details and the proofs of the auxiliary lemmas introduced in Section 4, and the proof of Theorem 2, whose structure departs completely from the proof of Theorem 1 and whose outline has been illustrated in Section 4.

#### B.2.1 Proofs of Auxiliary Lemmas

We start by proving Lemma 1, i.e., the bounds on terms of the form $\|\boldsymbol{\theta}_t - \boldsymbol{\theta}_{t-\tau}\|^2$, for some $0 \leq \tau \leq t$.

**Proof of Lemma 1.** To prove (i), note that we can get

$$\begin{aligned}
\|\boldsymbol{\theta}_t - \boldsymbol{\theta}_{t-\tau_{mix}}\|^2 &\leq \left(\sum_{l=t-\tau_{mix}}^{t-1} \|\boldsymbol{\theta}_{l+1} - \boldsymbol{\theta}_l\|\right)^2 \\
&\overset{(31)}{\leq} \tau_{mix} \sum_{l=t-\tau_{mix}}^{t-1} \|\boldsymbol{\theta}_{l+1} - \boldsymbol{\theta}_l\|^2 \\
&= \tau_{mix}\alpha^2 \sum_{l=t-\tau_{mix}}^{t-1} \|\mathbf{g}(\boldsymbol{\theta}_{l-\tau_l}, o_{l-\tau_l})\|^2 \\
&\overset{(28)}{\leq} 2\alpha^2\tau_{mix}L^2 \sum_{l=t-\tau_{mix}}^{t-1} (\|\boldsymbol{\theta}_{l-\tau_l}\|^2 + \sigma^2) \\
&\leq 2\alpha^2\tau_{mix}L^2 \sum_{l=t-\tau_{mix}}^{t-1} (2r_{l-\tau_l}^2 + 3\sigma^2).
\end{aligned} \tag{115}$$

Taking the expectation on both sides of the inequality, we get

$$\begin{aligned}
\mathbb{E}\left[\|\boldsymbol{\theta}_t - \boldsymbol{\theta}_{t-\tau_{mix}}\|^2\right] &\leq 2\alpha^2\tau_{mix}L^2 \sum_{l=t-\tau_{mix}}^{t-1} (2\mathbb{E}\left[r_{l-\tau_l}^2\right] + 3\sigma^2) \\
&\leq 2\alpha^2\tau_{mix}L^2 \sum_{l=t-\tau_{mix}}^{t-1} (2 \max_{t-\tau_{mix}-\tau_{max} \leq j \leq t} \mathbb{E}\left[r_j^2\right] + 3\sigma^2) \\
&\leq 4\tau_{mix}^2\alpha^2L^2r_{t,2} + 6\alpha^2\tau_{mix}^2L^2\sigma^2 \\
&= 2\alpha^2\tau_{mix}^2L^2(2r_{t,2} + 3\sigma^2).
\end{aligned} \tag{116}$$

With analogous computations, we can get part (ii) of the Lemma, i.e.

$$\mathbb{E}\left[\|\boldsymbol{\theta}_t - \boldsymbol{\theta}_{t-\tau_t}\|^2\right] \leq 2\alpha^2\tau_{max}^2L^2(2r_{t,2} + 3\sigma^2). \tag{117}$$

□

Recall the definition of $\mathbf{e}_t$:

$$\mathbf{e}_t \triangleq \mathbf{g}(\boldsymbol{\theta}_t, o_t) - \mathbf{g}(\boldsymbol{\theta}_{t-\tau_t}, o_{t-\tau_t}). \tag{118}$$

As illustrated in the outline of the analysis in Section 4, for the purpose of analysis, we write the update rule as follows:

$$\boldsymbol{\theta}_{t+1} = \boldsymbol{\theta}_t + \alpha \mathbf{g}(\boldsymbol{\theta}_t, o_t) - \alpha \mathbf{e}_t, \tag{119}$$

from which we can write

$$\|\boldsymbol{\theta}_{t+1} - \boldsymbol{\theta}^*\|^2 = J_{t,1} + \alpha^2 J_{t,2} - 2\alpha J_{t,3}, \tag{120}$$

with

$$\begin{aligned} J_{t,1} &\triangleq \|\boldsymbol{\theta}_t - \boldsymbol{\theta}^* + \alpha \mathbf{g}(\boldsymbol{\theta}_t, o_t)\|^2 \\ J_{t,2} &\triangleq \|\mathbf{e}_t\|^2 \\ J_{t,3} &\triangleq \langle \mathbf{e}_t, \boldsymbol{\theta}_t - \boldsymbol{\theta}^* + \alpha \mathbf{g}(\boldsymbol{\theta}_t, o_t)\rangle. \end{aligned} \tag{121}$$

**Proof of Lemma 2 - (i).** Note that

$$\begin{aligned} J_{t,1} = \|\boldsymbol{\theta}_t - \boldsymbol{\theta}^* + \alpha \mathbf{g}(\boldsymbol{\theta}_t, o_t)\|^2 = r_t^2 + 2\alpha \underbrace{\langle \boldsymbol{\theta}_t - \boldsymbol{\theta}^*, \mathbf{g}(\boldsymbol{\theta}_t, o_t)\rangle}_{J_{t,11}} \\ + \alpha^2 \underbrace{\|\mathbf{g}(\boldsymbol{\theta}_t, o_t)\|^2}_{J_{t,12}}. \end{aligned} \tag{122}$$

We also observe that

$$\begin{aligned} \mathbb{E}\left[J_{t,12}\right] &= \mathbb{E}\left[\|\mathbf{g}(\boldsymbol{\theta}_t, o_t)\|^2\right] \\ &\leq \mathbb{E}\left[2L^2\left(\|\boldsymbol{\theta}_t\|^2 + \sigma^2\right)\right] \\ &\leq 2L^2\left(2\mathbb{E}\left[r_t^2\right] + 3\sigma^2\right) \\ &\leq 2L^2\left(2r_{t,2} + 3\sigma^2\right). \end{aligned} \tag{123}$$

Now note that, using (24),

$$\begin{aligned} J_{t,11} = \langle \boldsymbol{\theta}_t - \boldsymbol{\theta}^*, \mathbf{g}(\boldsymbol{\theta}_t, o_t)\rangle &= -\langle \boldsymbol{\theta}^* - \boldsymbol{\theta}_t, \bar{\mathbf{g}}(\boldsymbol{\theta}_t)\rangle \\ &+ \langle \boldsymbol{\theta}_t - \boldsymbol{\theta}^*, \mathbf{g}(\boldsymbol{\theta}_t, o_t) - \bar{\mathbf{g}}(\boldsymbol{\theta}_t)\rangle \\ &\leq -\mu r_t^2 + \underbrace{\langle \boldsymbol{\theta}_t - \boldsymbol{\theta}^*, \mathbf{g}(\boldsymbol{\theta}_t, o_t) - \bar{\mathbf{g}}(\boldsymbol{\theta}_t)\rangle}_{T_1'}, \end{aligned} \tag{124}$$

where we now omit the dependence on the iterate $t$ in the terms we bound, for notational convenience. Now, note that

$$T_1' = \underbrace{\langle \boldsymbol{\theta}_t - \boldsymbol{\theta}_{t-\tau_{mix}}, \mathbf{g}(\boldsymbol{\theta}_t, o_t) - \bar{\mathbf{g}}(\boldsymbol{\theta}_t)\rangle}_{T_{11}'} + \underbrace{\langle \boldsymbol{\theta}_{t-\tau_{mix}} - \boldsymbol{\theta}^*, \mathbf{g}(\boldsymbol{\theta}_t, o_t) - \bar{\mathbf{g}}(\boldsymbol{\theta}_t)\rangle}_{T_{12}'}, \tag{125}$$

where, using the Cauchy-Schwarz inequality and the triangle inequality,

$$\begin{aligned} T_{11}' &\leq \|\boldsymbol{\theta}_t - \boldsymbol{\theta}_{t-\tau_{mix}}\|(\|\mathbf{g}(\boldsymbol{\theta}_t, o_t)\| + \|\bar{\mathbf{g}}(\boldsymbol{\theta}_t)\|) \\ &\overset{(27)}{\leq} 2L(\|\boldsymbol{\theta}_t - \boldsymbol{\theta}_{t-\tau_{mix}}\|(\|\boldsymbol{\theta}_t\| + \sigma)) \\ &\overset{(*)}{\leq} L\left(\frac{1}{\alpha\tau_{mix}L}\|\boldsymbol{\theta}_t - \boldsymbol{\theta}_{t-\tau_{mix}}\|^2 + \alpha\tau_{mix}L(\|\boldsymbol{\theta}_t\| + \sigma)^2\right) \\ &\overset{(31)}{\leq} L\left(\frac{1}{\alpha\tau_{mix}L}\|\boldsymbol{\theta}_t - \boldsymbol{\theta}_{t-\tau_{mix}}\|^2 + 2\alpha\tau_{mix}L(\|\boldsymbol{\theta}_t\|^2 + \sigma^2)\right), \end{aligned} \tag{126}$$

where for $(*)$ we used the fact that, from (30), we have

$$ab = \left(\frac{1}{\sqrt{c}}a\right)(\sqrt{c}b) \leq \frac{1}{2c}a^2 + \frac{cb^2}{2}, \tag{127}$$

specifically with $c = \alpha\tau_{mix}L$. Taking the expectation on both sides and applying (ii) of Lemma 1, we get

$$\begin{aligned} \mathbb{E}\left[T_{11}'\right] &\leq L\left(\frac{1}{2\alpha\tau_{mix}L}\mathbb{E}\left[\|\boldsymbol{\theta}_t - \boldsymbol{\theta}_{t-\tau_{mix}}\|^2\right] + \alpha\tau_{mix}L(2\mathbb{E}\left[r_t^2\right] + 3\sigma^2)\right) \\ &\leq L\left(\frac{2\alpha^2\tau_{mix}^2L^2}{2\alpha\tau_{mix}L}(2r_{t,2} + 3\sigma^2) + \alpha\tau_{mix}L(2r_{t,2} + 3\sigma^2)\right) \\ &= 4\alpha\tau_{mix}L^2r_{t,2} + 6\alpha\tau_{mix}L^2\sigma^2. \end{aligned} \tag{128}$$

Now, we proceed to bound $\mathbb{E}[T'_{12}]$. Note that

$$\begin{aligned} T'_{12} &= \langle \boldsymbol{\theta}_{t-\tau_{mix}} - \boldsymbol{\theta}^*, \mathbf{g}(\boldsymbol{\theta}_t, o_t) - \bar{\mathbf{g}}(\boldsymbol{\theta}_t)\rangle \\ &= \bar{T}_1 + \bar{T}_2 + \bar{T}_3, \end{aligned} \tag{129}$$

with

$$\begin{aligned} \bar{T}_1 &= \langle \boldsymbol{\theta}_{t-\tau_{mix}} - \boldsymbol{\theta}^*, \mathbf{g}(\boldsymbol{\theta}_{t-\tau_{mix}}, o_t) - \bar{\mathbf{g}}(\boldsymbol{\theta}_{t-\tau})\rangle \\ \bar{T}_2 &= \langle \boldsymbol{\theta}_{t-\tau_{mix}} - \boldsymbol{\theta}^*, \mathbf{g}(\boldsymbol{\theta}_t, o_t) - \mathbf{g}(\boldsymbol{\theta}_{t-\tau_{mix}}, o_t)\rangle \\ \bar{T}_3 &= \langle \boldsymbol{\theta}_{t-\tau_{mix}} - \boldsymbol{\theta}^*, \bar{\mathbf{g}}(\boldsymbol{\theta}_{t-\tau_{mix}}) - \bar{\mathbf{g}}(\boldsymbol{\theta}_t)\rangle. \end{aligned} \tag{130}$$

We first bound $\bar{T}_2$ and $\bar{T}_3$. Note that, using the Lipschitz property of the TD update direction (26), and calculations similar to the ones used to bound $\mathbb{E}[T'_{11}]$, we get

$$\begin{aligned} \bar{T}_2 &\leq \|\boldsymbol{\theta}_{t-\tau_{mix}} - \boldsymbol{\theta}^*\| \|\mathbf{g}(\boldsymbol{\theta}_t, o_t) - \mathbf{g}(\boldsymbol{\theta}_{t-\tau_{mix}}, o_t)\| \\ &\leq L\|\boldsymbol{\theta}_{t-\tau_{mix}} - \boldsymbol{\theta}^*\| \|\boldsymbol{\theta}_{t-\tau_{mix}} - \boldsymbol{\theta}_t\| \\ &\overset{(127)}{\leq} \frac{L^2\alpha\tau_{mix}}{2} r^2_{t-\tau_{mix}} + \frac{\|\boldsymbol{\theta}_t - \boldsymbol{\theta}_{t-\tau_{mix}}\|^2}{2\alpha\tau_{mix}}. \end{aligned} \tag{131}$$

Taking the expectation and applying (ii) of Lemma 1, we can get

$$\begin{aligned} \mathbb{E}\left[\bar{T}_2\right] &\leq \frac{\alpha\tau_{mix}L^2 r_{t,2}}{2} + 2\alpha\tau_{mix}L^2 r_{t,2} + 3\alpha\tau_{mix}L^2\sigma^2 \\ &\leq 3\alpha\tau_{mix}L^2(r_{t,2} + \sigma^2). \end{aligned} \tag{132}$$

With the same calculations, we can get

$$\mathbb{E}\left[\bar{T}_3\right] \leq 3\alpha\tau_{mix}L^2(r^2_{t,2} + \sigma^2). \tag{133}$$

We now proceed to bound $\bar{T}_1$.

$$\begin{aligned} \mathbb{E}\left[\bar{T}_1\right] &= \mathbb{E}\left[\langle \boldsymbol{\theta}_{t-\tau_{mix}} - \boldsymbol{\theta}^*, \mathbf{g}(\boldsymbol{\theta}_{t-\tau_{mix}}, o_t) - \bar{\mathbf{g}}(\boldsymbol{\theta}_{t-\tau_{mix}})\rangle\right] \\ &= \mathbb{E}\left[\langle \boldsymbol{\theta}_{t-\tau_{mix}} - \boldsymbol{\theta}^*, \mathbb{E}\left[\mathbf{g}(\boldsymbol{\theta}_{t-\tau_{mix}}, o_t)|o_{t-\tau}, \boldsymbol{\theta}_{t-\tau_{mix}}\right] - \bar{\mathbf{g}}(\boldsymbol{\theta}_{t-\tau_{mix}})\rangle\right] \\ &\leq \mathbb{E}\left[\|\boldsymbol{\theta}_{t-\tau_{mix}} - \boldsymbol{\theta}^*\| \|\mathbb{E}\left[\mathbf{g}(\boldsymbol{\theta}_{t-\tau_{mix}}, o_t)|o_{t-\tau_{mix}}, \boldsymbol{\theta}_{t-\tau_{mix}}\right] - \bar{\mathbf{g}}(\boldsymbol{\theta}_{t-\tau_{mix}})\|\right] \\ &\overset{(*)}{\leq} \alpha\mathbb{E}\left[\|\boldsymbol{\theta}_{t-\tau_{mix}} - \boldsymbol{\theta}^*\|(\|\boldsymbol{\theta}_{t-\tau_{mix}}\| + \sigma)\right] \\ &\leq \alpha\mathbb{E}\left[\|\boldsymbol{\theta}_{t-\tau_{mix}} - \boldsymbol{\theta}^*\|(\|\boldsymbol{\theta}_{t-\tau_{mix}} - \boldsymbol{\theta}^*\| + 2\sigma)\right] \\ &\leq \alpha\mathbb{E}\left[\frac{1}{2}\left(r^2_{t-\tau_{mix}} + 2r^2_{t-\tau_{mix}} + 4\sigma^2\right)\right] \\ &\leq 2\alpha(r_{t,2} + \sigma^2), \end{aligned} \tag{134}$$

where for $(*)$ we used Definition 1 of the mixing time and the fact that $\sigma \geq 1$. So, putting the above bounds together, we get

$$\mathbb{E}\left[T'_{12}\right] = \mathbb{E}\left[\bar{T}_1\right] + \mathbb{E}\left[\bar{T}_2\right] + \mathbb{E}\left[\bar{T}_3\right] \leq 8\alpha\tau_{mix}L^2(r^2_{t,2} + \sigma^2). \tag{135}$$

This then implies

$$\begin{aligned} \mathbb{E}\left[T'_1\right] &= \mathbb{E}\left[T'_{11}\right] + \mathbb{E}\left[T'_{12}\right] \\ &\leq 4\alpha\tau_{mix}L^2 r_{t,2} + 6\alpha\tau_{mix}L^2\sigma^2 + 8\alpha\tau_{mix}L^2(r_{t,2} + \sigma^2) \\ &\leq 12\alpha\tau_{mix}L^2 r_{t,2} + 14\alpha\tau_{mix}L^2\sigma^2. \end{aligned} \tag{136}$$

We also have

$$\mathbb{E}\left[J_{t,11}\right] \leq -\mu\mathbb{E}\left[r^2_t\right] + \mathbb{E}\left[T'_1\right]. \tag{137}$$

Hence,

$$\begin{aligned} \mathbb{E}\left[J_{t,1}\right] &= \mathbb{E}\left[r^2_t\right] + 2\alpha\mathbb{E}\left[J_{t,11}\right] + \alpha^2\mathbb{E}\left[J_{t,12}\right] \\ &\leq (1 - 2\alpha\mu)\mathbb{E}\left[r^2_t\right] + 28\alpha^2\tau_{mix}L^2 r_{t,2} + 34\alpha^2\tau_{mix}L^2\sigma^2, \end{aligned} \tag{138}$$

which concludes the proof of the Lemma. □

**Proof of Lemma 2 - (ii).** Note that

$$\begin{aligned} J_{t,2} = \|\mathbf{e}_t\|^2 &= \|\mathbf{g}(\boldsymbol{\theta}_t, o_t) - \mathbf{g}(\boldsymbol{\theta}_{t-\tau_t}, o_{t-\tau_t})\|^2 \\ &\overset{(31)}{\leq} 2\left(\|\mathbf{g}(\boldsymbol{\theta}_t, o_t)\|^2 + \|\mathbf{g}(\boldsymbol{\theta}_{t-\tau_t}, o_{t-\tau_t})\|^2\right) \\ &\overset{(28)}{\leq} 2\left(2L^2(\|\boldsymbol{\theta}_t\|^2+\sigma^2) + 2L^2(\|\boldsymbol{\theta}_{t-\tau_t}\|^2+\sigma^2)\right) \\ &\leq 4L^2(2r_t^2 + 3\sigma^2 + 2r_{t-\tau_t} + 3\sigma^2). \end{aligned} \tag{139}$$

Taking the expectation, we conclude getting

$$\mathbb{E}\left[J_{t,2}\right] = \mathbb{E}\left[\|\mathbf{e}_t\|^2\right] \leq 8L^2(2r_{t,2} + 3\sigma^2). \tag{140}$$

**Proof of Lemma 2 - (iii).** In the following, we drop the dependence on the iteration $t$ in most of the terms we bound. We write

$$\begin{aligned} -J_{t,3} &= \langle \mathbf{e}_t, \boldsymbol{\theta}_t - \boldsymbol{\theta}^* + \alpha \mathbf{g}(\boldsymbol{\theta}_t, o_t)\rangle \\ &= \underbrace{\langle -\mathbf{e}_t, \boldsymbol{\theta}_t - \boldsymbol{\theta}^*\rangle}_{\Delta} + \underbrace{\alpha\langle -\mathbf{e}_t, \mathbf{g}(\boldsymbol{\theta}_t, o_t)\rangle}_{\bar{\Delta}}. \end{aligned} \tag{141}$$

Note that

$$\begin{aligned} \bar{\Delta} = \alpha\langle -\mathbf{e}_t, \mathbf{g}(\boldsymbol{\theta}_t, o_t)\rangle &\leq \alpha\|\mathbf{e}_t\|\|\mathbf{g}(\boldsymbol{\theta}_t, o_t)\| \\ &\leq \frac{\alpha}{2}\left(\|\mathbf{e}_t\|^2 + \|\mathbf{g}(\boldsymbol{\theta}_t, o_t)\|^2\right). \end{aligned} \tag{142}$$

Using (140) and (28) to bound $\mathbb{E}\left[\|\mathbf{e}_t\|^2\right]$ and $\mathbb{E}\left[\|\mathbf{g}(\boldsymbol{\theta}_t, o_t)\|^2\right]$, respectively, we get

$$\begin{aligned} \mathbb{E}\left[\bar{\Delta}\right] &\leq \frac{\alpha}{2}\left(8L^2\left(2r_{t,2} + 3\sigma^2\right) + 2L^2\left(2r_{t,2} + 3\sigma^2\right)\right) \\ &= 10\alpha L^2 r_{t,2} + 15\alpha L^2\sigma^2. \end{aligned} \tag{143}$$

We now proceed to bound $\Delta$ as follows:

$$\begin{aligned} \Delta = \langle -\mathbf{e}_t, \boldsymbol{\theta}_t - \boldsymbol{\theta}^*\rangle &= \langle -\mathbf{g}(\boldsymbol{\theta}_t, o_t) + \mathbf{g}(\boldsymbol{\theta}_{t-\tau_t}, o_{t-\tau_t}), \boldsymbol{\theta}_t - \boldsymbol{\theta}^*\rangle \\ &= \underbrace{\langle -\mathbf{g}(\boldsymbol{\theta}_t, o_t) + \mathbf{g}(\boldsymbol{\theta}_t, o_{t-\tau_t}), \boldsymbol{\theta}_t - \boldsymbol{\theta}^*\rangle}_{\Delta_1} \\ &+ \underbrace{\langle -\mathbf{g}(\boldsymbol{\theta}_t, o_{t-\tau_t}) + \mathbf{g}(\boldsymbol{\theta}_{t-\tau_t}, o_{t-\tau_t}), \boldsymbol{\theta}_t - \boldsymbol{\theta}^*\rangle}_{\Delta_2}. \end{aligned} \tag{144}$$

Note that, thanks to the Lipschitz property of the TD direction and with calculations analogous to the ones performed to obtain the bound on $\mathbb{E}\left[\bar{T}_2\right]$ (see (131) and (132)), we get

$$\mathbb{E}\left[\Delta_2\right] \leq \mathbb{E}\left[L\|\boldsymbol{\theta}_t - \boldsymbol{\theta}_{t-\tau_t}\| r_t\right] \leq 3\alpha\tau_{mix}L^2(r_{t,2} + \sigma^2). \tag{145}$$

We now bound $\Delta_1$:

$$\begin{aligned} \Delta_1 = &\underbrace{\langle -\mathbf{g}(\boldsymbol{\theta}_t, o_t) + \mathbf{g}(\boldsymbol{\theta}_{t-\tau_{mix}}, o_t), \boldsymbol{\theta}_t, -\boldsymbol{\theta}^*\rangle}_{\Delta_{11}} \\ &+ \underbrace{\langle -\mathbf{g}(\boldsymbol{\theta}_{t-\tau_{mix}}, o_t) + \mathbf{g}(\boldsymbol{\theta}_t, o_{t-\tau_t}), \boldsymbol{\theta}_t - \boldsymbol{\theta}^*\rangle}_{\Delta_{12}}. \end{aligned} \tag{146}$$

With calculations analogous to the ones performed to obtain the bound on $\mathbb{E}\left[\bar{T}_2\right]$ (see (131) and (132)), we get

$$\mathbb{E}\left[\Delta_{11}\right] \leq \mathbb{E}\left[L\|\boldsymbol{\theta}_t - \boldsymbol{\theta}_{t-\tau_t}\| r_t\right] \leq 3\alpha\tau_{mix}L^2(r_{t,2} + \sigma^2). \tag{147}$$

We now proceed to bound $\Delta_{12}$:

$$\begin{aligned} \Delta_{12} = &\underbrace{\langle -\mathbf{g}(\boldsymbol{\theta}_{t-\tau_{mix}}, o_t) + \bar{\mathbf{g}}(\boldsymbol{\theta}_{t-\tau_{mix}}), \boldsymbol{\theta}_t - \boldsymbol{\theta}^*\rangle}_{\Delta_1'} \\ &+ \underbrace{\langle -\bar{\mathbf{g}}(\boldsymbol{\theta}_{t-\tau_{mix}}) + \mathbf{g}(\boldsymbol{\theta}_t, o_{t-\tau_t}), \boldsymbol{\theta}_t - \boldsymbol{\theta}^*\rangle}_{\Delta_2'}. \end{aligned} \tag{148}$$

We have

$$\begin{aligned}\Delta_1' = &\underbrace{\langle -\mathbf{g}(\boldsymbol{\theta}_{t-\tau_{mix}}, o_t) + \bar{\mathbf{g}}(\boldsymbol{\theta}_{t-\tau_{mix}}), \boldsymbol{\theta}_{t-\tau_{mix}} - \boldsymbol{\theta}^* \rangle}_{\Delta_{11}'} \\ &+ \underbrace{\langle -\mathbf{g}(\boldsymbol{\theta}_{t-\tau_{mix}}, o_t) + \bar{\mathbf{g}}(\boldsymbol{\theta}_{t-\tau_{mix}}), \boldsymbol{\theta}_t - \boldsymbol{\theta}_{t-\tau_{mix}} \rangle}_{\Delta_{12}'}.\end{aligned} \tag{149}$$

Note that

$$\begin{aligned}\Delta_{12}' &\leq \|\mathbf{g}(\boldsymbol{\theta}_{t-\tau_{mix}}, o_t) - \bar{\mathbf{g}}(\boldsymbol{\theta}_{t-\tau_{mix}})\| \|\boldsymbol{\theta}_t - \boldsymbol{\theta}_{t-\tau_{mix}}\| \\ &\leq (\|\mathbf{g}(\boldsymbol{\theta}_{t-\tau_{mix}}, o_t)\| + \|\bar{\mathbf{g}}(\boldsymbol{\theta}_{t-\tau_{mix}})\|) \|\boldsymbol{\theta}_t - \boldsymbol{\theta}_{t-\tau_{mix}}\| \\ &\overset{(27)}{\leq} 2L (\|\boldsymbol{\theta}_{t-\tau_{mix}}\| + \sigma) \|\boldsymbol{\theta}_t - \boldsymbol{\theta}_{t-\tau_{mix}}\| \\ &\leq 2L \left(r_{t-\tau_{mix}}^2 + 2\sigma\right) \|\boldsymbol{\theta}_t - \boldsymbol{\theta}_{t-\tau_{mix}}\| \\ &\leq 2\alpha L^2 \tau_{mix} (r_{t-\tau_{mix}}^2 + 2\sigma^2) + \frac{1}{2\alpha\tau_{mix}} \|\boldsymbol{\theta}_t - \boldsymbol{\theta}_{t-\tau_{mix}}\|^2.\end{aligned} \tag{150}$$

Taking expectation on both sides and applying Lemma 1, we get

$$\begin{aligned}\mathbb{E}\left[\Delta_{12}'\right] &\leq 2\alpha\tau_{mix} L^2 \left(\mathbb{E}\left[r_{t-\tau_{mix}}^2\right] + 2\sigma^2\right) + \frac{1}{2\alpha\tau_{mix}} \mathbb{E}\left[\|\boldsymbol{\theta}_t - \boldsymbol{\theta}_{t-\tau_{mix}}\|^2\right] \\ &\leq 2\alpha\tau_{mix} L^2 \left(r_{t,2} + 2\sigma^2\right) + \alpha\tau_{mix} L^2 \left(2r_{t,2} + 3\sigma^2\right) \\ &= 4\alpha\tau_{mix} L^2 r_{t,2} + 7\alpha\tau_{mix} L^2 \sigma^2.\end{aligned} \tag{151}$$

Now note that $\Delta_{11}'$ can be bounded using the same calculations used for $\bar{T}_1$ in (134):

$$\mathbb{E}\left[\Delta_{11}'\right] \leq 2\alpha(r_{t,2} + \sigma^2). \tag{152}$$

Next note that

$$\begin{aligned}\Delta_2' &= \langle -\bar{\mathbf{g}}(\boldsymbol{\theta}_{t-\tau_{mix}}) + \mathbf{g}(\boldsymbol{\theta}_t, o_{t-\tau_t}), \boldsymbol{\theta}_t - \boldsymbol{\theta}^* \rangle \\ &= \underbrace{\langle -\bar{\mathbf{g}}(\boldsymbol{\theta}_{t-\tau_{mix}}) + \mathbf{g}(\boldsymbol{\theta}_{t-\tau_{mix}}, o_{t-\tau_t}), \boldsymbol{\theta}_t - \boldsymbol{\theta}^* \rangle}_{\Delta_{21}'} \\ &+ \underbrace{\langle -\mathbf{g}(\boldsymbol{\theta}_{t-\tau_{mix}}, o_{t-\tau_t}) + \mathbf{g}(\boldsymbol{\theta}_t, o_{t-\tau_t}), \boldsymbol{\theta}_t - \boldsymbol{\theta}^* \rangle}_{\Delta_{22}'}.\end{aligned} \tag{153}$$

To bound $\mathbb{E}\left[\Delta_{22}'\right]$, we can proceed with calculations analogous to the ones performed to obtain the bound on $\mathbb{E}\left[\bar{T}_2\right]$ (see (131) and (132)), getting

$$\mathbb{E}\left[\Delta_{22}'\right] \leq 3\alpha\tau_{mix} L^2 (r_{t,2} + \sigma^2). \tag{154}$$

Now, we write

$$\begin{aligned}\Delta_{21}' = &\underbrace{\langle -\bar{\mathbf{g}}(\boldsymbol{\theta}_{t-\tau_{mix}}) + \bar{\mathbf{g}}(\boldsymbol{\theta}_{t-\tau_{mix}-\tau_t}), \boldsymbol{\theta}_t - \boldsymbol{\theta}^* \rangle}_{\bar{\Delta}_1} \\ &+ \underbrace{\langle -\bar{\mathbf{g}}(\boldsymbol{\theta}_{t-\tau_{mix}-\tau_t}) + \mathbf{g}(\boldsymbol{\theta}_{t-\tau_{mix}}, o_{t-\tau_t}), \boldsymbol{\theta}_t - \boldsymbol{\theta}^* \rangle}_{\bar{\Delta}_2}.\end{aligned} \tag{155}$$

We see that, as before, we can bound $\mathbb{E}\left[\bar{\Delta}_1\right]$ with the same procedure we used to bound $\mathbb{E}\left[\Delta_{22}'\right]$:

$$\mathbb{E}\left[\bar{\Delta}_1\right] \leq 3\alpha\tau_{max} L^2 (r_{t,2} + \sigma^2). \tag{156}$$

We write

$$\begin{aligned}\bar{\Delta}_2 = &\underbrace{\langle -\bar{\mathbf{g}}(\boldsymbol{\theta}_{t-\tau_{mix}-\tau_t}) + \mathbf{g}(\boldsymbol{\theta}_{t-\tau_{mix}-\tau_t}, o_{t-\tau_t}), \boldsymbol{\theta}_t - \boldsymbol{\theta}^* \rangle}_{\bar{\Delta}_{21}} \\ &+ \underbrace{\langle -\mathbf{g}(\boldsymbol{\theta}_{t-\tau_{mix}-\tau_t}, o_{t-\tau_t}) + \mathbf{g}(\boldsymbol{\theta}_{t-\tau_{mix}}, o_{t-\tau_t}), \boldsymbol{\theta}_t - \boldsymbol{\theta}^* \rangle}_{\bar{\Delta}_{22}}.\end{aligned} \tag{157}$$

Now note that $\mathbb{E}\left[\bar{\Delta}_{22}\right]$ can be bounded with calculations analogous to the ones performed to obtain the bound on $\bar{\Delta}_1$:

$$\begin{aligned}\mathbb{E}\left[\bar{\Delta}_{22}\right] &\leq \mathbb{E}\left[L(\|\boldsymbol{\theta}_{t-\tau_{mix}-\tau_t} - \boldsymbol{\theta}_{t-\tau_{mix}}\|\|\boldsymbol{\theta}_t - \boldsymbol{\theta}^*\|)\right] \\ &\leq 3\alpha\tau_{max}L^2(r_{t,2} + \sigma^2).\end{aligned} \tag{158}$$

Next, observe that

$$\begin{aligned}\bar{\Delta}_{21} = &\underbrace{\left\langle -\bar{\mathbf{g}}(\boldsymbol{\theta}_{t-\tau_{mix}-\tau_t}) + \mathbf{g}(\boldsymbol{\theta}_{t-\tau_{mix}-\tau_t}, o_{t-\tau_t}), \boldsymbol{\theta}_{t-\tau_{mix}-\tau_t} - \boldsymbol{\theta}^*\right\rangle}_{\bar{\Delta}_{211}} \\ &+ \underbrace{\left\langle -\bar{\mathbf{g}}(\boldsymbol{\theta}_{t-\tau_{mix}-\tau_t}) + \mathbf{g}(\boldsymbol{\theta}_{t-\tau_{mix}-\tau_t}, o_{t-\tau_t}), \boldsymbol{\theta}_t - \boldsymbol{\theta}_{t-\tau_{mix}-\tau_t}\right\rangle}_{\bar{\Delta}_{212}}.\end{aligned} \tag{159}$$

With calculations analogous to the ones performed to obtain the bound on $\mathbb{E}\left[\Delta'_{12}\right]$ (see (151)), we get

$$\begin{aligned}\mathbb{E}\left[\bar{\Delta}_{212}\right] &\leq \mathbb{E}\left[\|\boldsymbol{\theta}_t - \boldsymbol{\theta}_{t-\tau_{mix}-\tau_t}\|L\left(\|\bar{\mathbf{g}}(\boldsymbol{\theta}_{t-\tau_{mix}-\tau_t})\| + \|\mathbf{g}(\boldsymbol{\theta}_{t-\tau_{mix}-\tau_t}, o_{t-\tau_t})\|\right)\right] \\ &\leq 4\alpha(\tau_{mix} + \tau_{max})L^2 r_{t,2} + 7\alpha(\tau_{mix} + \tau_{max})L^2\sigma^2.\end{aligned} \tag{160}$$

Finally, note that $\bar{\Delta}_{211}$ can be bounded using the same calculations used for $\bar{T}_1$ in (134), yielding

$$\mathbb{E}\left[\bar{\Delta}_{211}\right] \leq 2\alpha(r_{t,2} + \sigma^2). \tag{161}$$

So, $\mathbb{E}\left[T_3\right]$ can be upper-bounded by a sum of terms that are upper-bounded by either $O(\alpha)(r_{t,2}^2 + \sigma^2)$, $O(\alpha\tau_{max})(r_{t,2}^2 + \sigma^2)$, $O(\alpha\tau_{mix})(r_{t,2}^2 + \sigma^2)$, or $O(\alpha)(\tau_{mix} + \tau_{max})(r_{t,2}^2 + \sigma^2)$. Putting all the terms together, we can get

$$\mathbb{E}\left[-J_{t,3}\right] \leq 28\alpha L^2(\tau_{mix} + \tau_{max})(r_{t,2} + \sigma^2), \tag{162}$$

which concludes the proof. □

Now, recall the definition of the update rule for delayed SA with time-varying delay under Assumption 4:

$$\textbf{\textit{Delayed SA:}} \qquad \boldsymbol{\theta}_{t+1} = \boldsymbol{\theta}_t + \alpha\mathbf{g}(\boldsymbol{\theta}_{t-\tau_t}, o_{t-\tau_t}), \quad \tau_t \leq \min\{t, \tau_{max}\}. \tag{163}$$

Consider the mean squared error term $\mathbb{E}\left[r_t^2\right] = \mathbb{E}\left[\|\boldsymbol{\theta}_t - \boldsymbol{\theta}^*\|^2\right]$, and its expression derived in (120). Let us define $\tau' = 2\tau_{max} + \tau_{mix}$. The bounds on $\mathbb{E}\left[J_{t,1}\right]$, $\mathbb{E}\left[J_{t,2}\right]$, and $\mathbb{E}\left[J_{t,3}\right]$ provided in the previous section are such that the update rule (163) satisfies the following $\forall t \geq \tau'$:

$$\begin{aligned}\mathbb{E}\left[r_{t+1}^2\right] &= \mathbb{E}\left[\|\boldsymbol{\theta}_{t+1} - \boldsymbol{\theta}^*\|^2\right] \\ &= \mathbb{E}\left[J_{t,1}\right] + \alpha^2\mathbb{E}\left[J_{t,2}\right] - 2\alpha\mathbb{E}\left[J_{t,3}\right] \\ &\leq (1 - 2\alpha\mu)\mathbb{E}\left[r_t^2\right] + 98\alpha^2L^2(\tau_{mix} + \tau_{max})(r_{t,2} + \sigma^2),\end{aligned} \tag{164}$$

with

$$r_{t,2} = \max_{t-\tau' \leq l \leq t} \mathbb{E}\left[r_l^2\right].$$

As mentioned in the outline of the analysis in Section 4, the final part of the proof of Theorem 2 is based on a crucial argument that shows that, for a sufficiently small step size, the iterates generated by (163) are uniformly bounded, which is shown in Lemma 3. Note that, starting from the above inequality (164), one could apply a technique similar to Feyzmahdavian et al. (2014) to handle the delayed optimality gaps in the bound. However, *this technique would provide a suboptimal convergence rate with a $\tau_{max}^2$ dependency on the maximum delay.* To prove Lemma 3, that enables the obtainment of the tight linear dependence on $\tau_{max}$, we first provide the following result, which proves the base case of the induction proof on which the proof of Lemma 3 relies on.

**Lemma 8.** *Consider the update rule in (163) and let $B = 9\sigma^2$. For $0 \leq t \leq \tau' = 2\tau_{max} + \tau_{mix}$ and $\alpha \leq \frac{1}{24L^2\tau'}$, we have*

$$\mathbb{E}\left[r_t^2\right] \leq B, \quad 0 \leq t \leq \tau'. \tag{165}$$

*Proof.* Note that

$$\begin{aligned} r_{t+1}^2 &= r_t^2 + 2\alpha\langle \boldsymbol{\theta}_t - \boldsymbol{\theta}^*, \mathbf{g}(\boldsymbol{\theta}_{t-\tau_t}, o_{t-\tau_t})\rangle + \alpha^2\|\mathbf{g}(\boldsymbol{\theta}_{t-\tau_t}, o_{t-\tau_t})\|^2 \\ &\overset{(30)}{\leq} r_t^2 + \alpha r_t^2 + \alpha\|\mathbf{g}(\boldsymbol{\theta}_{t-\tau_t}, o_{t-\tau_t})\|^2 + \alpha^2\|\mathbf{g}(\boldsymbol{\theta}_{t-\tau_t}, o_{t-\tau_t})\|^2 \\ &\leq (1+\alpha)r_t^2 + 2\alpha\|\mathbf{g}(\boldsymbol{\theta}_{t-\tau_t}, o_{t-\tau_t})\|^2 \\ &\overset{(28)}{\leq} (1+\alpha)r_t^2 + 4\alpha L^2(\|\boldsymbol{\theta}_{t-\tau_t}\|^2 + \sigma^2) \\ &\leq (1+\alpha)r_t^2 + 4\alpha L^2(2r_{t-\tau_t}^2 + 3\sigma^2) \\ &= (1+\alpha)r_t^2 + 8\alpha L^2 r_{t-\tau_t}^2 + 12\alpha L^2\sigma^2. \end{aligned} \tag{166}$$

Taking the expectation on both sides, we get

$$\mathbb{E}\left[r_{t+1}^2\right] \leq (1+\alpha)\mathbb{E}\left[r_t^2\right] + 8\alpha L^2\mathbb{E}\left[r_{t-\tau_t}^2\right] + 12\alpha L^2\sigma^2. \tag{167}$$

Hence, we get an inequality of the following form:

$$V_{t+1} \leq pV_t + qV_{t-\tau_t} + \beta, \quad 0 \leq \tau_t \leq \min\{t, \tau_{max}\}, \tag{168}$$

with $V_t = \mathbb{E}\left[r_t^2\right]$, $p = 1+\alpha$, $q = 8\alpha L^2\tau_{max}$, and $\beta = 12\alpha L^2\sigma^2$. We define $\rho = p + q$, noting that $\rho > 1$. We now prove by induction that, for all $t \geq 0$,

$$V_t \leq \rho^t V_0 + \epsilon_t, \tag{169}$$

where

$$\epsilon_t = \begin{cases} \rho\epsilon_{t-1} + \beta \text{ for } t \geq 1 \\ 0 \text{ for } t = 0 \end{cases} \tag{170}$$

The base case is trivially satisfied, because $V_0 \leq V_0$. As the induction hypothesis, suppose that (169) is true for $0 \leq s \leq k$, for some $k \geq 0$, so that

$$V_s \leq \rho^s V_0 + \epsilon_s, \quad 0 \leq s \leq k. \tag{171}$$

Now, we check the property for $k+1$, using (168), and noting that $\epsilon_k$ is an increasing sequence:

$$\begin{aligned} V_{k+1} &\leq pV_k + qV_{k-\tau_k} + \beta, \\ &\leq p(\rho^k V_0 + \epsilon_k) + q(\rho^{k-\tau_k}V_0 + \epsilon_{k-\tau_k}) + \beta \\ &\leq p(\rho^k V_0 + \epsilon_k) + q(\rho^k V_0 + \epsilon_k) + \beta \\ &\leq (p+q)\rho^k V_0 + (p+q)\epsilon_k + \beta \\ &= \rho^{k+1}V_0 + \epsilon_{k+1}. \end{aligned} \tag{172}$$

From this, we can conclude the proof of (169). Now, note that

$$\epsilon_t = \beta\sum_{j=0}^{t-1}\rho^j, \tag{173}$$

and so we can write, for $0 \leq t \leq \tau'$,

$$\rho^t \leq \rho^{\tau'} \leq (1+\alpha)^{\tau'} \leq e^{\alpha\tau'} \leq e^{0.25} \leq 2, \tag{174}$$

using the fact that $\alpha \leq \frac{1}{4\tau'}$. Hence, using the above results, we can get for $0 \leq t \leq \tau'$:

$$\begin{aligned} \mathbb{E}\left[r_t^2\right] \leq \rho^t r_0^2 + \epsilon_t \leq 2r_0^2 + \beta\sum_{j=0}^{\tau'-1}\rho^j \leq 2r_0^2 + 2\beta\tau' = 2r_0^2 + 2(12\alpha L^2\sigma^2)\tau' \\ \leq 2r_0^2 + \sigma^2 \leq 9\sigma^2, \end{aligned} \tag{175}$$

where we used the fact that $\alpha \leq \frac{1}{24L^2\tau'}$ and that $r_0^2 = \|\boldsymbol{\theta}_0 - \boldsymbol{\theta}^*\|^2 \leq 2\|\boldsymbol{\theta}_0\|^2 + 2\|\boldsymbol{\theta}^*\|^2 \leq 4\sigma^2$.

We have just shown that, for $0 \leq t \leq \tau'$, it holds

$$\mathbb{E}\left[r_t^2\right] \leq 9\sigma^2. \tag{176}$$

□

Now, we provide the proof of Lemma 3, which relies on Lemma 8.

**Proof of Lemma 3.** We know from Lemma 8 that for $t = 0, ..., \tau'$, with $\tau' = 2\tau_{max} + \tau_{mix}$, and $\alpha \leq \frac{1}{24L^2\tau'}$, we have

$$\mathbb{E}\left[r_t^2\right] \leq B. \tag{177}$$

We now proceed by induction to show that the bound holds true also for any $t \geq \tau'$, thus for all $t \geq 0$. We use (177) as the base case for the induction proof. Fix any $t \geq \tau'$, and as the induction hypothesis, assume that the property is true for all $s$ satisfying $\tau' \leq s \leq t$, i.e.,

$$\mathbb{E}\left[r_s^2\right] \leq B \quad \forall \tau' \leq s \leq t. \tag{178}$$

Now, from (164) we can write

$$\mathbb{E}\left[r_{t+1}^2\right] \leq (1 - 2\alpha\mu)\mathbb{E}\left[r_t^2\right] + 98\alpha^2 L^2(\tau_{mix} + \tau_{max})(r_{t,2} + \sigma^2). \tag{179}$$

Observe that from the induction hypothesis and the induction base case, the following is true:

$$r_{t,2} = \max_{t-\tau' \leq l \leq t} \mathbb{E}\left[r_l^2\right] \leq B. \tag{180}$$

Hence, we can write, recalling that $B = 9\sigma^2 \geq \sigma^2$,

$$\begin{aligned}
\mathbb{E}\left[r_{t+1}^2\right] &\leq (1 - 2\alpha\mu)\mathbb{E}\left[r_t^2\right] + 98\alpha^2 L^2(\tau_{mix} + \tau_{max})(r_{t,2} + \sigma^2) \\
&\leq (1 - 2\alpha\mu)B + 98\alpha^2 L^2(\tau_{mix} + \tau_{max})(B + \sigma^2) \\
&\leq (1 - 2\alpha\mu)B + 2B98\alpha^2 L^2(\tau_{mix} + \tau_{max}) \\
&\leq (1 - 2\alpha\mu + 196\alpha^2 L^2(\tau_{mix} + \tau_{max}))B.
\end{aligned} \tag{181}$$

Thus, for $\alpha \leq \frac{\mu}{196L^2\bar{\tau}} \leq \frac{\mu}{98L^2\tau'} \leq \frac{\mu}{98L^2(\tau_{mix}+\tau_{max})}$, we get $1 - 2\alpha\mu + 196\alpha^2 L^2(\tau_{mix} + \tau_{max}) \leq 1$, implying

$$\mathbb{E}\left[r_{t+1}^2\right] \leq B. \tag{182}$$

This concludes the proof. □

Using this last Lemma in conjunction with (164), we can now prove Theorem 2.

### B.2.2 Proof of Theorem 2

Note that, from (164), we can write, for $T \geq \tau' = 2\tau_{max} + \tau_{mix}$,

$$\begin{aligned}
\mathbb{E}\left[r_{t+1}^2\right] &\leq (1 - 2\alpha\mu)\mathbb{E}\left[r_t^2\right] + 98\alpha^2 L^2(\tau_{mix} + \tau_{max})(r_{t,2} + \sigma^2) \\
&\overset{(*)}{\leq} (1 - 2\alpha\mu)\mathbb{E}\left[r_t^2\right] + 2B98\alpha^2 L^2(\tau_{mix} + \tau_{max}),
\end{aligned} \tag{183}$$

where for $(*)$ we used the fact that, for $\alpha \leq \frac{1}{196L^2\bar{\tau}}$, it holds $r_{t,2} \leq B = 9\sigma^2$, as established by Lemma 3. Iterating the inequality, we get

$$\begin{aligned}
\mathbb{E}\left[r_{t+1}^2\right] &\leq (1 - 2\alpha\mu)^{t+1-\tau'} r_{\tau'}^2 + 98L^2\alpha^2(\tau_{mix} + \tau_{max})2B\sum_{j=0}^{\infty}(1 - 2\alpha\mu)^j \\
&\leq (1 - 2\alpha\mu)^{t+1-\tau'} r_{\tau'}^2 + \frac{98L^2\alpha(\tau_{mix} + \tau_{max})B}{\mu} \\
&\leq (1 - 2\alpha\mu)^{t+1-\tau'} B + \frac{98L^2\alpha(\tau_{mix} + \tau_{max})B}{\mu} \\
&\overset{(*)}{\leq} (1 - 2\alpha\mu)^{t+1} 2B + \frac{98L^2\alpha(\tau_{mix} + \tau_{max})B}{\mu} \\
&\leq e^{-2\alpha\mu(t+1)} 2B + \frac{98L^2\alpha(\tau_{mix} + \tau_{max})B}{\mu},
\end{aligned} \tag{184}$$

where $(*)$ follows because

$$(1-2\alpha\mu)^{-\tau'} \leq e^{2\alpha\mu\tau'} \leq e^{0.25} \leq 2, \tag{185}$$

and by noting that $\alpha\mu \leq \alpha \leq \frac{1}{196L^2\bar{\tau}} \leq \frac{1}{8\tau'}$. Hence, for $\alpha \leq \frac{1}{CL^2\bar{\tau}}$, with $C = 196$, we get the result. Setting $\alpha = \frac{1}{CL^2\bar{\tau}}$, with $C \geq 196$ and $C' = 98$, we can also get

$$\mathbb{E}\left[r_T^2\right] \leq \exp\left(-\frac{2\mu^2 T}{CL^2\bar{\tau}}\right) 2B + \frac{C'B}{C}. \tag{186}$$

□

### B.3 Proof of Theorem 3

For a precision threshold $\epsilon > 0$ to be specified soon, recall that the update rule of delay-adaptive SA takes the form:

$$\boldsymbol{\theta}_{t+1} = \begin{cases} \boldsymbol{\theta}_t + \alpha \left(\mathbf{g}(\boldsymbol{\theta}_{t-\tau_t}, o_{t-\tau_t})\right) & \text{if } \|\boldsymbol{\theta}_t - \boldsymbol{\theta}_{t-\tau_t}\| \leq \epsilon \\ \boldsymbol{\theta}_t & \text{otherwise.} \end{cases} \tag{187}$$

Let us define

$$\beta \triangleq \epsilon + \sigma, \quad \text{and} \quad c \triangleq L\beta,$$

where $L > 0$ is the Lipschitz constant defined in Assumption 3.

We use the following way of writing the update rule:

$$\boldsymbol{\theta}_{t+1} = \boldsymbol{\theta}_t + \alpha \mathbf{g}(\boldsymbol{\theta}_t, o_t) - \alpha \mathbf{e}_t, \tag{188}$$

where $\mathbf{e}_t = \mathbf{g}(\boldsymbol{\theta}_t, o_t) - \mathbf{g}(\boldsymbol{\theta}_{t-\tau_t}, o_{t-\tau_t})$. We also define the following indicator function $I_t$:

$$I_t = \begin{cases} 1 & \text{if } \|\boldsymbol{\theta}_t - \boldsymbol{\theta}_{t-\tau_t}\| \leq \epsilon \\ 0 & \text{otherwise.} \end{cases} \tag{189}$$

#### B.3.1 Outline of the Analysis

Recall that $I_t = 1$ if $\|\boldsymbol{\theta}_t - \boldsymbol{\theta}_{t-\tau_t}\| \leq \epsilon$ and $I_t = 0$ otherwise. Then, we can express $\|\boldsymbol{\theta}_{t+1} - \boldsymbol{\theta}^*\|^2$ as

$$\|\boldsymbol{\theta}_{t+1} - \boldsymbol{\theta}^*\|^2 = I_t \left(K_{t,1} + \alpha^2 K_{t,2} - 2\alpha K_{t,3}\right) + (1 - I_t)\|\boldsymbol{\theta}_t - \boldsymbol{\theta}^*\|^2, \tag{190}$$

with

$$\begin{aligned} K_{t,1} &\triangleq \|\boldsymbol{\theta}_t - \boldsymbol{\theta}^* + \alpha \mathbf{g}(\boldsymbol{\theta}_t, o_t)\|^2, \\ K_{t,2} &\triangleq \|\mathbf{e}_t\|^2, \\ K_{t,3} &\triangleq \langle \mathbf{e}_t, \boldsymbol{\theta}_t - \boldsymbol{\theta}^* + \alpha \mathbf{g}(\boldsymbol{\theta}_t, o_t)\rangle. \end{aligned} \tag{191}$$

In order to analyze the convergence of the delay-adaptive SA update rule, we first derive bounds for $K_{t,1}$, $K_{t,2}$, and $-K_{t,3}$ for iterations in which $I_t = 1$. Then, we establish a lower bound on the number of iterations in which $I_t = 1$. By using these two results, we are able to obtain the finite-time convergence rate of Theorem 3.

#### B.3.2 Auxiliary Lemmas

Similarly to Section 4, we provide three lemmas that are instrumental to proving the main result of this section, i.e., Theorem 3. As in the case of the vanilla delayed SA update, we start by providing a result that provides a bound on the norm of $\boldsymbol{\theta}_t - \boldsymbol{\theta}_{t-\tau}$. Note that here, unlike the non-adaptive vanilla case of Section 4, we are able to get a bound on $\|\boldsymbol{\theta}_t - \boldsymbol{\theta}_{t-\tau}\|$ which is a function only of the *current* iterate $\boldsymbol{\theta}_t$. This aspect significantly simplifies the proof. Note that this step, which plays a crucial role in enabling us to remove the dependency of the rate on the maximum delay, is provided by the specific adaptive update rule, which only uses SA update directions $\mathbf{g}(\boldsymbol{\theta}_{t-\tau_t}, o_{t-\tau_t})$ when the distance between the delayed iterate $\boldsymbol{\theta}_{t-\tau_t}$ - at which the direction was computed - and the current approximation parameter $\boldsymbol{\theta}_t$ is smaller than the threshold $\epsilon$, with $\epsilon$ set to $\alpha$. In the proof of Theorem 3 and related lemmas, without loss of generality, we also use negative indices for iterates and observations. In particular, we define, for $j < 0$, $\boldsymbol{\theta}_j = \boldsymbol{\theta}_0$ and $o_j = o_0$.

**Lemma 9.** *For any $\tau \geq 1$ and $t \geq 0$, we have*

$$\|\boldsymbol{\theta}_t - \boldsymbol{\theta}_{t-\tau}\| \leq 4L\alpha\tau(r_t + 2\beta), \tag{192}$$

*where $\beta = L\epsilon + L\sigma$.*

Using Lemma 9, we can bound $K_{t,1}$, $K_{t,2}$, and $K_{t,3}$ in Lemma 10.

**Lemma 10.** *For $t \geq \tau_{mix}$, if $I_t = 1$, we have, for some absolute constants $C_{11}, C_{12} \geq 1$,*

$$\begin{aligned}
&(i) \quad \mathbb{E}\left[K_{t,1}\right] \leq \left(1 - 2\alpha\mu + C_{11}L^2(\alpha^2\tau_{mix})\right)\mathbb{E}\left[r_t^2\right] + C_{11}L^2(\alpha^2\tau_{mix})\beta^2 \\
&(ii) \quad \mathbb{E}\left[K_{t,2}\right] \leq 2L^2\tau_{mix}(\mathbb{E}\left[r_t^2\right] + \beta^2) \\
&(iii) \quad \mathbb{E}\left[-K_{t,3}\right] \leq C_{12}L^2\alpha\tau_{mix}(\mathbb{E}\left[r_t^2\right] + \beta^2).
\end{aligned}$$

With the help of the previous lemma, we can rewrite Equation (190) as follows, when $I_t = 1$:

$$\begin{aligned}
\mathbb{E}\left[r_{t+1}^2\right] &= \mathbb{E}\left[\|\boldsymbol{\theta}_{t+1} - \boldsymbol{\theta}^*\|^2\right] \\
&= \mathbb{E}\left[K_{t,1}\right] + \alpha^2\mathbb{E}\left[K_{t,2}\right] - 2\alpha\mathbb{E}\left[K_{t,3}\right] \\
&\leq \left(1 - 2\alpha\mu + C_{11}L^2(\alpha^2\tau_{mix})\right)\mathbb{E}\left[r_t^2\right] + C_1L^2(\alpha^2\tau_{mix})\beta^2.
\end{aligned} \tag{193}$$

To finish the proof, we only need to determine the number of times we update the parameter in $T$ iterations. This can be achieved through the following lemma, which we borrow from Cohen et al. (2021). By utilizing this lemma, we can complete the proof and obtain the bound presented in Theorem 3.

**Lemma 11.** *Let $\tau_{avg}$ be the average delay, i.e., $\tau_{avg} = \frac{1}{T}\sum_{t=1}^{T}\tau_t$. The number of updates that the delay-adaptive SA algorithm makes is at least $\frac{T}{4\tau_{avg}+4}$.*

In the remainder of the section, we first provide the proofs for all the auxiliary lemmas, and then conclude with the proof of Theorem 3.

### B.3.3 Proofs of Auxiliary Lemmas

**Proof of Lemma 9.** Let $t^{'} = t - \tau_t$, then

$$\begin{aligned}
\|\boldsymbol{\theta}_{t+1} - \boldsymbol{\theta}_t\| &\leq \alpha I_t\|\mathbf{g}(\boldsymbol{\theta}_{t'}, o_{t'})\| \\
&\leq \alpha L I_t(\|\boldsymbol{\theta}_{t'}\| + \sigma) \\
&\leq \alpha L I_t(\|\boldsymbol{\theta}_t\| + \|\boldsymbol{\theta}_t - \boldsymbol{\theta}_{t'}\| + \sigma) \\
&\leq \alpha L I_t(\|\boldsymbol{\theta}_t\| + \epsilon + \sigma) \\
&\leq \alpha L\left(\|\boldsymbol{\theta}_t\| + \beta\right),
\end{aligned} \tag{194}$$

from which we get

$$\|\boldsymbol{\theta}_{t+1}\| \leq (1 + \alpha L)\|\boldsymbol{\theta}_t\| + \alpha L\beta. \tag{195}$$

In what follows, we will recursively use the above inequality in tandem with the following facts: (i) If $I_t = 0$, then $\boldsymbol{\theta}_t = \boldsymbol{\theta}_{t-1}$ and our next steps hold trivially, and that (ii) $\boldsymbol{\theta}_j = \boldsymbol{\theta}_0$ for $j < 0$. We then have that $\forall t \geq 0$:

$$\begin{aligned}
\|\boldsymbol{\theta}_t\| &\leq (1 + \alpha L)\|\boldsymbol{\theta}_{t-1}\| + \alpha L\beta \\
&\leq (1 + \alpha L)^2\|\boldsymbol{\theta}_{t-2}\| + \alpha L\beta(1 + \alpha L) + \alpha L\beta \\
&\leq (1 + \alpha L)^{\tau}\|\boldsymbol{\theta}_{t-\tau}\| + \alpha L\beta\sum_{j=0}^{\tau-1}(1 + \alpha L)^j \\
&\leq 2\|\boldsymbol{\theta}_{t-\tau}\| + 2\alpha L\beta\tau,
\end{aligned} \tag{196}$$

where we used $\alpha\tau L \leq \frac{1}{4}$ and the fact that, for $0 \leq j \leq \tau$, $(1+x)^j \leq (1+x)^\tau \leq e^{x\tau} \leq 2$ for $x\tau \leq \frac{1}{4}$.
Now, using the above inequality on $\|\boldsymbol{\theta}_t\|$,

$$\begin{aligned}\|\boldsymbol{\theta}_t - \boldsymbol{\theta}_{t-\tau}\| &\leq \sum_{j=t-\tau}^{t-1} \|\boldsymbol{\theta}_{j+1} - \boldsymbol{\theta}_j\| \\ &\overset{(194)}{\leq} \sum_{j=t-\tau}^{t-1} \alpha L(\|\boldsymbol{\theta}_j\|+\beta) \\ &\overset{(196)}{\leq} \alpha L \sum_{j=t-\tau}^{t-1} (2\|\boldsymbol{\theta}_{t-\tau}\|+2\alpha L\beta\tau) + \alpha\tau L\beta \\ &\leq \alpha\tau L(2\|\boldsymbol{\theta}_{t-\tau}\|+2L\alpha\beta\tau) + \alpha L\beta\tau \\ &\leq 2\alpha L\tau\|\boldsymbol{\theta}_{t-\tau}\|+2L\alpha\beta\tau,\end{aligned} \tag{197}$$

where in the last inequality we used the fact that $\alpha\tau \leq \frac{1}{4L}$. Moreover,

$$\begin{aligned}\|\boldsymbol{\theta}_t - \boldsymbol{\theta}_{t-\tau}\| &\leq 2\alpha\tau L\|\boldsymbol{\theta}_{t-\tau}\|+2L\alpha\beta\tau \\ &\leq 2\alpha\tau L\|\boldsymbol{\theta}_t\|+2L\alpha\beta\tau + 2\alpha\tau L\|\boldsymbol{\theta}_t - \boldsymbol{\theta}_{t-\tau}\| \\ &\leq 2\alpha\tau L\|\boldsymbol{\theta}_t\|+2L\alpha\beta\tau + \frac{1}{2}\|\boldsymbol{\theta}_t - \boldsymbol{\theta}_{t-\tau}\|,\end{aligned} \tag{198}$$

which results in

$$\|\boldsymbol{\theta}_t - \boldsymbol{\theta}_{t-\tau}\| \leq 4\alpha\tau L\|\boldsymbol{\theta}_t\|+4L\alpha\beta\tau. \tag{199}$$

Recalling $c = L\beta$, we can write

$$\begin{aligned}\|\boldsymbol{\theta}_t - \boldsymbol{\theta}_{t-\tau}\| &\leq 4\alpha\tau L\|\boldsymbol{\theta}_t\|+4\alpha c\tau \\ &\leq 4\alpha\tau L(r_t + 2\beta).\end{aligned} \tag{200}$$

Also note that (recall $c = L\beta$):

$$\begin{aligned}\|\boldsymbol{\theta}_t - \boldsymbol{\theta}_{t-\tau}\|^2 &\leq (4\alpha\tau L\|\boldsymbol{\theta}_t\|+4\alpha c\tau)^2 \\ &\leq (4\alpha\tau L r_t + 8\alpha c\tau)^2 \\ &\leq 16(\alpha\tau L r_t + 2\alpha c\tau)^2 \\ &\overset{(31)}{\leq} 32(\alpha^2\tau^2L^2r_t^2 + 4\alpha^2\tau^2c^2) \\ &\leq 32\alpha^2\tau^2L^2(r_t^2 + 4\beta^2).\end{aligned} \tag{201}$$

□

Note that for an iteration $t$ in which an update is not made, i.e., $I_t = 0$, we have

$$\|\boldsymbol{\theta}_{t+1} - \boldsymbol{\theta}^*\|^2 = \|\boldsymbol{\theta}_t - \boldsymbol{\theta}^*\|^2. \tag{202}$$

If $I_t = 1$, we have:

$$\|\boldsymbol{\theta}_{t+1} - \boldsymbol{\theta}^*\|^2 = \underbrace{\|\boldsymbol{\theta}_t + \alpha\mathbf{g}(\boldsymbol{\theta}_t, o_t) - \boldsymbol{\theta}^*\|^2}_{K_{t,1}} + \underbrace{\alpha^2\|\mathbf{e}_t\|^2}_{K_{t,2}} \underbrace{-2\alpha\langle\mathbf{e}_t, \boldsymbol{\theta}_t - \boldsymbol{\theta}^* + \alpha\mathbf{g}(\boldsymbol{\theta}_t, o_t)\rangle}_{K_{t,3}}. \tag{203}$$

**Proof of Lemma 10 - (i).** In the following, we drop the dependence on the iteration $t$ in the terms we bound. We write

$$\begin{aligned}K_{t,1} = \|\boldsymbol{\theta}_t - \boldsymbol{\theta}^* + \alpha\mathbf{g}(\boldsymbol{\theta}_t, o_t)\|^2 = r_t^2 + 2\alpha \underbrace{\langle\boldsymbol{\theta}_t - \boldsymbol{\theta}^*, \mathbf{g}(\boldsymbol{\theta}_t, o_t)\rangle}_{I_1} \\ + \alpha^2 \underbrace{\|\mathbf{g}(\boldsymbol{\theta}_t, o_t)\|^2}_{I_2}.\end{aligned} \tag{204}$$

Note that

$$\mathbb{E}\left[I_2\right] = \mathbb{E}\left[\|\mathbf{g}(\boldsymbol{\theta}_t, o_t)\|^2\right] \overset{(28)}{\leq} 2L^2\left(\mathbb{E}\left[\|\boldsymbol{\theta}_t\|^2\right] + \sigma^2\right) \quad \leq 2L^2\left(2\mathbb{E}\left[r_t^2\right] + 3\sigma^2\right). \tag{205}$$

Next, note that

$$\begin{aligned} I_1 &= \langle \boldsymbol{\theta}_t - \boldsymbol{\theta}^*, \mathbf{g}(\boldsymbol{\theta}_t, o_t)\rangle = -\langle \boldsymbol{\theta}^* - \boldsymbol{\theta}_t, \bar{\mathbf{g}}(\boldsymbol{\theta}_t)\rangle \\ &\quad + \langle \boldsymbol{\theta}_t - \boldsymbol{\theta}^*, \mathbf{g}(\boldsymbol{\theta}_t, o_t) - \bar{\mathbf{g}}(\boldsymbol{\theta}_t)\rangle \\ &\leq -\omega(1-\gamma) r_t^2 + \underbrace{\langle \boldsymbol{\theta}_t - \boldsymbol{\theta}^*, \mathbf{g}(\boldsymbol{\theta}_t, o_t) - \bar{\mathbf{g}}(\boldsymbol{\theta}_t)\rangle}_{T_1'}. \end{aligned} \tag{206}$$

Now,

$$T_1' = \underbrace{\langle \boldsymbol{\theta}_t - \boldsymbol{\theta}_{t-\tau_{mix}}, \mathbf{g}(\boldsymbol{\theta}_t, o_t) - \bar{\mathbf{g}}(\boldsymbol{\theta}_t)\rangle}_{T_{11}'} + \underbrace{\langle \boldsymbol{\theta}_{t-\tau_{mix}} - \boldsymbol{\theta}^*, \mathbf{g}(\boldsymbol{\theta}_t, o_t) - \bar{\mathbf{g}}(\boldsymbol{\theta}_t)\rangle}_{T_{12}'}. \tag{207}$$

Using the fact that $\beta \leq \sigma$ and $L \geq 1$, we can bound $T_{11}'$ as follows:

$$\begin{aligned} T_{11}' &\leq \|\boldsymbol{\theta}_t - \boldsymbol{\theta}_{t-\tau_{mix}}\| \|\mathbf{g}(\boldsymbol{\theta}_t, o_t) - \bar{\mathbf{g}}(\boldsymbol{\theta}_t)\| \\ &\overset{(27)}{\leq} \|\boldsymbol{\theta}_t - \boldsymbol{\theta}_{t-\tau_{mix}}\| 2L(\|\boldsymbol{\theta}_t\| + \sigma) \\ &\overset{(201)}{\leq} (4\alpha\tau_{mix} L\|\boldsymbol{\theta}_t\| + 4\alpha\tau_{mix} c) 2L(\|\boldsymbol{\theta}_t\| + \sigma) \\ &\leq 8\alpha\tau_{mix} L^2\|\boldsymbol{\theta}_t\|^2 + 8\alpha c\tau_{mix} L\sigma + (8\alpha\tau_{mix} L^2\sigma + 8\alpha\tau_{mix} Lc)\|\boldsymbol{\theta}_t\| \\ &\leq 8\alpha\tau_{mix} L^2\|\boldsymbol{\theta}_t\|^2 + 8\alpha L^2\beta\tau_{mix}\sigma + (8\alpha\tau_{mix} L^2\sigma + 8\alpha\tau_{mix} L^2\beta)\|\boldsymbol{\theta}_t\| \\ &\leq 8\alpha\tau_{mix} L^2\|\boldsymbol{\theta}_t\|^2 + 8\alpha L^2\tau_{mix}\beta^2 + 32\alpha\tau_{mix} L^2\beta^2 + 32\alpha\tau_{mix} L^2\|\boldsymbol{\theta}_t\|^2 \\ &= 40\alpha\tau_{mix} L^2\|\boldsymbol{\theta}_t\|^2 + 40\alpha\tau_{mix} L^2\beta^2 \\ &\leq 80\alpha\tau_{mix} L^2 r_t^2 + 120\alpha\tau_{mix} L^2\beta^2 \\ &\leq 40L^2\alpha\tau_{mix}(2r_t^2 + 3\beta^2). \end{aligned} \tag{208}$$

Taking expectations on both sides of the above display, we obtain

$$\mathbb{E}\left[T_{11}'\right] \leq 40L^2\alpha\tau_{mix}(2\mathbb{E}\left[r_t^2\right] + 3\beta^2). \tag{209}$$

Now, observe that

$$\begin{aligned} T_{12}' &= \langle \boldsymbol{\theta}_{t-\tau_{mix}} - \boldsymbol{\theta}^*, \mathbf{g}(\boldsymbol{\theta}_t, o_t) - \bar{\mathbf{g}}(\boldsymbol{\theta}_t)\rangle \\ &= \bar{T}_1 + \bar{T}_2 + \bar{T}_3, \end{aligned} \tag{210}$$

with

$$\begin{aligned} \bar{T}_1 &= \langle \boldsymbol{\theta}_{t-\tau_{mix}} - \boldsymbol{\theta}^*, \mathbf{g}(\boldsymbol{\theta}_{t-\tau_{mix}}, o_t) - \bar{\mathbf{g}}(\boldsymbol{\theta}_{t-\tau_{mix}})\rangle \\ \bar{T}_2 &= \langle \boldsymbol{\theta}_{t-\tau_{mix}} - \boldsymbol{\theta}^*, \mathbf{g}(\boldsymbol{\theta}_t, o_t) - g(\boldsymbol{\theta}_{t-\tau_{mix}}, o_t)\rangle \\ \bar{T}_3 &= \langle \boldsymbol{\theta}_{t-\tau_{mix}} - \boldsymbol{\theta}^*, \bar{\mathbf{g}}(\boldsymbol{\theta}_{t-\tau_{mix}}) - \bar{\mathbf{g}}(\boldsymbol{\theta}_t)\rangle. \end{aligned} \tag{211}$$

Note that, from (201), we have, recalling $c = L\beta$,

$$\|\boldsymbol{\theta}_t - \boldsymbol{\theta}_{t-\tau_{mix}}\|^2 \leq 32\alpha^2\tau_{mix}^2 L^2\left(r_t^2 + 4\beta^2\right). \tag{212}$$

We now proceed to bound $\bar{T}_2$ and $\bar{T}_3$. Note that, using the Lipschitz property (Assumption 3) and the fact that

$\alpha \leq \frac{1}{12\tau_{mix}L^2}$, we obtain

$$\begin{aligned}
\bar{T}_2 &\leq \|\boldsymbol{\theta}_{t-\tau_{mix}} - \boldsymbol{\theta}^*\|\|\mathbf{g}(\boldsymbol{\theta}_t, o_t) - \mathbf{g}(\boldsymbol{\theta}_{t-\tau_{mix}}, o_t)\| \\
&\leq L\|\boldsymbol{\theta}_{t-\tau_{mix}} - \boldsymbol{\theta}^*\|\|\boldsymbol{\theta}_{t-\tau_{mix}} - \boldsymbol{\theta}_t\| \\
&\overset{(30)}{\leq} \frac{\alpha\tau_{mix}L^2}{2}\|\boldsymbol{\theta}_{t-\tau_{mix}} - \boldsymbol{\theta}^*\|^2 + \frac{1}{2\alpha\tau_{mix}}\|\boldsymbol{\theta}_{t-\tau_{mix}} - \boldsymbol{\theta}_t\|^2 \\
&\leq \alpha\tau_{mix}L^2\|\boldsymbol{\theta}_t - \boldsymbol{\theta}^*\|^2 + \alpha\tau_{mix}L^2\|\boldsymbol{\theta}_{t-\tau_{mix}} - \boldsymbol{\theta}_t\|^2 + \frac{1}{2\alpha\tau_{mix}}\|\boldsymbol{\theta}_{t-\tau_{mix}} - \boldsymbol{\theta}_t\|^2 \\
&\leq \alpha\tau_{mix}L^2 r_t^2 + \alpha\tau_{mix}L^2\left(32\alpha^2\tau_{mix}^2L^2\left(r_t^2 + 4\beta^2\right)\right) \\
&+ \frac{1}{2\alpha\tau_{mix}}\left(32\alpha^2\tau_{mix}^2L^2\left(r_t^2 + 4\beta^2\right)\right) \\
&\leq \alpha\tau_{mix}L^2 r_t^2 + \alpha\tau_{mix}L^2(r_t^2 + \beta^2) + 16\alpha\tau_{mix}L^2 r_t^2 + 64\alpha\tau_{mix}L^2\beta^2 \\
&\leq \alpha\tau_{mix}L^2(18r_t^2 + 65\beta^2).
\end{aligned} \tag{213}$$

Taking the expectation, we get

$$\mathbb{E}\left[\bar{T}_2\right] \leq \alpha\tau_{mix}L^2(18\mathbb{E}\left[r_t^2\right] + 65\beta^2). \tag{214}$$

With the same calculations, we can get

$$\mathbb{E}\left[\bar{T}_3\right] \leq \alpha\tau_{mix}L^2(18\mathbb{E}\left[r_t^2\right] + 65\beta^2). \tag{215}$$

We now proceed to bound $\mathbb{E}\left[\bar{T}_1\right]$. Using the fact that $\alpha\tau_{mix} \leq \frac{1}{12L^2}$, we can write

$$\begin{aligned}
\mathbb{E}\left[\bar{T}_1\right] &= \mathbb{E}\left[\langle\boldsymbol{\theta}_{t-\tau_{mix}} - \boldsymbol{\theta}^*, \mathbf{g}(\boldsymbol{\theta}_{t-\tau_{mix}}, o_t) - \bar{\mathbf{g}}(\boldsymbol{\theta}_{t-\tau_{mix}})\rangle\right] \\
&= \mathbb{E}\left[\langle\boldsymbol{\theta}_{t-\tau_{mix}} - \boldsymbol{\theta}^*, \mathbb{E}\left[\mathbf{g}(\boldsymbol{\theta}_{t-\tau_{mix}}, o_t)|o_{t-\tau_{mix}}, \boldsymbol{\theta}_{t-\tau_{mix}}\right] - \bar{\mathbf{g}}(\boldsymbol{\theta}_{t-\tau_{mix}})\rangle\right] \\
&\leq \mathbb{E}\left[\|\boldsymbol{\theta}_{t-\tau_{mix}} - \boldsymbol{\theta}^*\|\|\mathbb{E}\left[\mathbf{g}(\boldsymbol{\theta}_{t-\tau_{mix}}, o_t)|o_{t-\tau_{mix}}, \boldsymbol{\theta}_{t-\tau_{mix}}\right] - \bar{\mathbf{g}}(\boldsymbol{\theta}_{t-\tau_{mix}})\|\right] \\
&\overset{(25)}{\leq} \alpha\mathbb{E}\left[\|\boldsymbol{\theta}_{t-\tau_{mix}} - \boldsymbol{\theta}^*\|(\|\boldsymbol{\theta}_{t-\tau_{mix}}\| + \sigma)\right] \\
&\overset{(30)}{\leq} \frac{\alpha}{2}\mathbb{E}\left[\|\boldsymbol{\theta}_{t-\tau_{mix}} - \boldsymbol{\theta}^*\|^2 + \left(\|\boldsymbol{\theta}_{t-\tau_{mix}} - \boldsymbol{\theta}^*\| + 2\sigma\right)^2\right] \\
&\leq \frac{\alpha}{2}\mathbb{E}\left[2r_t^2 + 2\|\boldsymbol{\theta}_t - \boldsymbol{\theta}_{t-\tau_{mix}}\|^2 + 2\|\boldsymbol{\theta}_{t-\tau_{mix}} - \boldsymbol{\theta}^*\|^2 + 8\sigma^2\right] \\
&\leq \frac{\alpha}{2}\mathbb{E}\left[2r_t^2 + 2\|\boldsymbol{\theta}_t - \boldsymbol{\theta}_{t-\tau_{mix}}\|^2 + 2\left(2\|\boldsymbol{\theta}_t - \boldsymbol{\theta}_{t-\tau_{mix}}\|^2 + 2\|\boldsymbol{\theta}_t - \boldsymbol{\theta}^*\|^2\right) + 8\sigma^2\right] \\
&\leq 3\alpha\mathbb{E}\left[r_t^2\right] + 3\alpha\mathbb{E}\left[\|\boldsymbol{\theta}_t - \boldsymbol{\theta}_{t-\tau_{mix}}\|^2\right] + 4\alpha\sigma^2 \\
&\leq 3\alpha\mathbb{E}\left[r_t^2\right] + 96\alpha^3\tau_{mix}^2\left(\mathbb{E}\left[r_t^2\right] + 4c^2\right) + 4\alpha\sigma^2 \\
&\leq 4\alpha\mathbb{E}\left[r_t^2\right] + 8\alpha c^2.
\end{aligned} \tag{216}$$

So, we obtain

$$\begin{aligned}
\mathbb{E}\left[T_{12}'\right] &= \mathbb{E}\left[\bar{T}_1\right] + \mathbb{E}\left[\bar{T}_2\right] + \mathbb{E}\left[\bar{T}_3\right] \\
&\leq 4\alpha\mathbb{E}\left[r_t^2\right] + 8\alpha c^2 + 2\left(\alpha\tau_{mix}L^2\left(18\mathbb{E}\left[r_t^2\right] + 65\beta^2\right)\right) \\
&\leq 46\alpha\tau_{mix}L^2\left(\mathbb{E}\left[r_t^2\right] + 3\beta^2\right).
\end{aligned} \tag{217}$$

Combining the bounds on $\mathbb{E}\left[T_{11}'\right]$ and $\mathbb{E}\left[T_{12}'\right]$, we obtain

$$\begin{aligned}
\mathbb{E}\left[T_1'\right] &= \mathbb{E}\left[T_{11}'\right] + \mathbb{E}\left[T_{12}'\right] \\
&\leq 40\alpha\tau_{mix}L^2(2\mathbb{E}\left[r_t^2\right] + 3\beta^2) + 46\alpha\tau_{mix}L^2\left(\mathbb{E}\left[r_t^2\right] + 3\beta^2\right) \\
&\leq 130\alpha\tau_{mix}L^2(\mathbb{E}\left[r_t^2\right] + 2\beta^2).
\end{aligned} \tag{218}$$

This immediately yields:

$$\begin{aligned}
\mathbb{E}\left[I_1\right] &\leq -\omega(1-\gamma)\mathbb{E}\left[r_t^2\right] + \mathbb{E}\left[T_1'\right] \\
&\leq -\omega(1-\gamma)\mathbb{E}\left[r_t^2\right] + 130\alpha\tau_{mix}L^2(\mathbb{E}\left[r_t^2\right] + 2\beta^2).
\end{aligned} \tag{219}$$

Hence,

$$\begin{aligned}\mathbb{E}\left[K_{t,1}\right] &= \mathbb{E}\left[r_t^2\right] + 2\alpha\mathbb{E}\left[I_1\right] + \alpha^2\mathbb{E}\left[I_2\right] \\ &\leq \left(1 - 2\alpha\mu + 264\alpha^2\tau_{mix}L^2\right)\mathbb{E}\left[r_t^2\right] + 526\alpha^2\tau_{mix}L^2\beta^2.\end{aligned} \tag{220}$$

□

**Proof of Lemma 10 - (ii).** Using the notation $t^{'} = t - \tau_t$, we get

$$\begin{aligned}\|\mathbf{e}_t\| &\leq \|\mathbf{g}(\boldsymbol{\theta}_{t'}, o_{t'}) - \mathbf{g}(\boldsymbol{\theta}_t, o_{t'})\| + \|\mathbf{g}(\boldsymbol{\theta}_t, o_{t'}) - \mathbf{g}(\boldsymbol{\theta}_t, o_t)\| \\ &\leq L\|\boldsymbol{\theta}_t - \boldsymbol{\theta}_{t'}\| + 2L(\|\boldsymbol{\theta}_t\| + \sigma) \\ &\leq 2L\|\boldsymbol{\theta}_t\| + L\epsilon + 2L\sigma \\ &\leq 2L\|\boldsymbol{\theta}_t\| + 2L(\epsilon + \sigma) \\ &= 2(L\|\boldsymbol{\theta}_t\| + c).\end{aligned} \tag{221}$$

Note that, using the above inequality, we have

$$\|\mathbf{e}_t\| \leq L\|\boldsymbol{\theta}_t\| + 2c. \tag{222}$$

Using this inequality, we can get

$$\begin{aligned}\mathbb{E}\left[\alpha^2 K_{t,2}\right] &\leq \alpha^2\mathbb{E}\left[\|\mathbf{e}_t\|^2\right] \\ &\leq \alpha^2\mathbb{E}\left[\left(L\|\boldsymbol{\theta}_t\| + 2c\right)^2\right] \\ &\leq \alpha^2\mathbb{E}\left[\left(2L^2\|\boldsymbol{\theta}_t\|^2 + 8c^2\right)\right] \\ &\leq 4\alpha^2L^2\mathbb{E}\left[r_t^2\right] + 12\alpha^2c^2.\end{aligned} \tag{223}$$

□

**Proof of Lemma 10 - (iii).** We write

$$\begin{aligned}-K_{t,3} &= \langle -\mathbf{e}_t, \boldsymbol{\theta}_t - \boldsymbol{\theta}^* + \alpha\mathbf{g}(\boldsymbol{\theta}_t, o_t)\rangle \\ &= \underbrace{\langle -\mathbf{e}_t, \boldsymbol{\theta}_t - \boldsymbol{\theta}^*\rangle}_{\Delta} + \underbrace{\alpha\langle -\mathbf{e}_t, \mathbf{g}(\boldsymbol{\theta}_t, o_t)\rangle}_{\bar{\Delta}}.\end{aligned} \tag{224}$$

Note that using (223), we have

$$\|\mathbf{e}_t\|^2 \leq 4L^2r_t^2 + 12c^2. \tag{225}$$

Using this fact, we can get

$$\begin{aligned}\bar{\Delta} &= \alpha\langle -\mathbf{e}_t, \mathbf{g}(\boldsymbol{\theta}_t, o_t)\rangle \\ &\leq \alpha\|\mathbf{e}_t\|\|\mathbf{g}(\boldsymbol{\theta}_t, o_t)\| \\ &\overset{(30)}{\leq} \frac{\alpha}{2}\left(\|\mathbf{e}_t\|^2 + \|\mathbf{g}(\boldsymbol{\theta}_t, o_t)\|^2\right) \\ &\overset{(28)}{\leq} \frac{\alpha}{2}\left(4L^2r_t^2 + 12c^2 + 2L^2(2r_t^2 + 3\sigma^2)\right) \\ &\leq \frac{\alpha}{2}(8L^2r_t^2 + 18c^2) \\ &\leq 4\alpha L^2r_t^2 + 9\alpha c^2,\end{aligned} \tag{226}$$

implying that

$$\mathbb{E}\left[\bar{\Delta}\right] \leq \alpha L^2(4\mathbb{E}\left[r_t^2\right] + 9\beta^2). \tag{227}$$

We now proceed to bound $\Delta$ as follows:

$$\begin{aligned}\Delta &= \langle -\mathbf{e}_t, \boldsymbol{\theta}_t - \boldsymbol{\theta}^*\rangle = \langle -\mathbf{g}(\boldsymbol{\theta}_t, o_t) + \mathbf{g}(\boldsymbol{\theta}_{t-\tau_t}, o_{t-\tau_t}), \boldsymbol{\theta}_t - \boldsymbol{\theta}^*\rangle \\ &= \underbrace{\langle -\mathbf{g}(\boldsymbol{\theta}_t, o_t) + \bar{\mathbf{g}}(\boldsymbol{\theta}_t) - \bar{\mathbf{g}}(\boldsymbol{\theta}_t) + \mathbf{g}(\boldsymbol{\theta}_t, o_{t-\tau_t}), \boldsymbol{\theta}_t - \boldsymbol{\theta}^*\rangle}_{\Delta_1} \\ &+ \underbrace{\langle -\mathbf{g}(\boldsymbol{\theta}_t, o_{t-\tau_t}) + \mathbf{g}(\boldsymbol{\theta}_{t-\tau_t}, o_{t-\tau_t}), \boldsymbol{\theta}_t - \boldsymbol{\theta}^*\rangle}_{\Delta_2}.\end{aligned} \tag{228}$$

Next, noting that $\|\mathbf{g}(\boldsymbol{\theta}_t, o_{t-\tau_t}) - \mathbf{g}(\boldsymbol{\theta}_{t-\tau_t}, o_{t-\tau_t})\| \leq L\|\boldsymbol{\theta}_t - \boldsymbol{\theta}_{t-\tau_t}\| \leq L\epsilon$, we have

$$\Delta_2 \leq L\epsilon r_t \leq \frac{\alpha L}{2}\left(r_t^2 + \left(\frac{\epsilon}{\alpha}\right)^2\right). \tag{229}$$

Now, using the fact that $\epsilon \leq \alpha$ and $1 \leq \sigma \leq \beta$, we have

$$\Delta_2 \leq \frac{\alpha L}{2}(r_t^2 + \beta^2), \tag{230}$$

and

$$\Delta_1 \leq \underbrace{\langle -\mathbf{g}(\boldsymbol{\theta}_t, o_t) + \bar{\mathbf{g}}(\boldsymbol{\theta}_t), \boldsymbol{\theta}_t - \boldsymbol{\theta}^*\rangle}_{\Delta_{11}} + \underbrace{\langle -\bar{\mathbf{g}}(\boldsymbol{\theta}_t) + \mathbf{g}(\boldsymbol{\theta}_t, o_{t-\tau_t}), \boldsymbol{\theta}_t - \boldsymbol{\theta}^*\rangle}_{\Delta_{12}}. \tag{231}$$

Note that $\Delta_{11} = T_1'$ for $T_1'$ defined above (see (206)), and therefore can be bounded accordingly as in (218):

$$\mathbb{E}\left[\Delta_{11}\right] \leq 130\alpha\tau_{mix}L^2(\mathbb{E}\left[r_t^2\right] + 2\beta^2). \tag{232}$$

We now bound $\mathbb{E}\left[\Delta_{12}\right]$ as follows:

$$\begin{aligned}
\Delta_{12} &= \langle -\bar{\mathbf{g}}(\boldsymbol{\theta}_t) + \mathbf{g}(\boldsymbol{\theta}_t, o_{t-\tau_t}), \boldsymbol{\theta}_t - \boldsymbol{\theta}^*\rangle \\
&= \underbrace{\langle -\bar{\mathbf{g}}(\boldsymbol{\theta}_{t-\tau_t}) + \mathbf{g}(\boldsymbol{\theta}_t, o_{t-\tau_t}), \boldsymbol{\theta}_t - \boldsymbol{\theta}^*\rangle}_{\Delta_1'} \\
&+ \underbrace{\langle -\bar{\mathbf{g}}(\boldsymbol{\theta}_t) + \bar{\mathbf{g}}(\boldsymbol{\theta}_{t-\tau_t}), \boldsymbol{\theta}_t - \boldsymbol{\theta}^*\rangle}_{\Delta_2'}.
\end{aligned} \tag{233}$$

Note that using the fact that $\epsilon \leq \alpha$, we have

$$\begin{aligned}
\Delta_2' &\leq L\|\boldsymbol{\theta}_t - \boldsymbol{\theta}_{t-\tau_t}\|\|\boldsymbol{\theta}_t - \boldsymbol{\theta}^*\| \\
&\leq L(\epsilon r_t) \\
&\leq L\left(\alpha r_t^2 + \frac{\epsilon^2}{\alpha}\right) \\
&\leq L\alpha(r_t^2 + \beta^2).
\end{aligned} \tag{234}$$

We now bound $\mathbb{E}\left[\Delta_1'\right]$ as follows:

$$\begin{aligned}
\Delta_1' &= \langle -\bar{\mathbf{g}}(\boldsymbol{\theta}_{t-\tau_t}) + \mathbf{g}(\boldsymbol{\theta}_t, o_{t-\tau_t}), \boldsymbol{\theta}_t - \boldsymbol{\theta}^*\rangle \\
&= \underbrace{\langle -\bar{\mathbf{g}}(\boldsymbol{\theta}_{t-\tau_t}) + \mathbf{g}(\boldsymbol{\theta}_{t-\tau_t}, o_{t-\tau_t}), \boldsymbol{\theta}_t - \boldsymbol{\theta}^*\rangle}_{\Delta_{11}'} \\
&+ \underbrace{\langle -\mathbf{g}(\boldsymbol{\theta}_{t-\tau_t}, o_{t-\tau_t}) + \mathbf{g}(\boldsymbol{\theta}_t, o_{t-\tau_t}), \boldsymbol{\theta}_t - \boldsymbol{\theta}^*\rangle}_{\Delta_{12}'}.
\end{aligned} \tag{235}$$

Note that, using Lipschitzness (Assumption 3) and the same calculations used to arrive at (234), one can obtain

$$\begin{aligned}
\Delta_{12}' &\leq \|\mathbf{g}(\boldsymbol{\theta}_{t-\tau_t}, o_{t-\tau_t}) - \mathbf{g}(\boldsymbol{\theta}_t, o_{t-\tau_t})\|\|\boldsymbol{\theta}_t - \boldsymbol{\theta}^*\| \\
&\leq L\|\boldsymbol{\theta}_{t-\tau_t} - \boldsymbol{\theta}_t\|\|\boldsymbol{\theta}_t - \boldsymbol{\theta}^*\| \\
&\leq L\alpha(r_t^2 + \beta^2).
\end{aligned} \tag{236}$$

Let us also observe that

$$\begin{aligned}
\Delta_{11}' &= \langle -\bar{\mathbf{g}}(\boldsymbol{\theta}_{t-\tau_t}) + \mathbf{g}(\boldsymbol{\theta}_{t-\tau_t}, o_{t-\tau_t}), \boldsymbol{\theta}_t - \boldsymbol{\theta}^*\rangle \\
&= \underbrace{\langle -\bar{\mathbf{g}}(\boldsymbol{\theta}_{t-\tau_t}) + \bar{\mathbf{g}}(\boldsymbol{\theta}_{t-\tau_t-\tau_{mix}}), \boldsymbol{\theta}_t - \boldsymbol{\theta}^*\rangle}_{\bar{\Delta}_1} \\
&+ \underbrace{\langle -\bar{\mathbf{g}}(\boldsymbol{\theta}_{t-\tau_t-\tau_{mix}}) + \mathbf{g}(\boldsymbol{\theta}_{t-\tau_t}, o_{t-\tau_t}), \boldsymbol{\theta}_t - \boldsymbol{\theta}^*\rangle}_{\bar{\Delta}_2}.
\end{aligned} \tag{237}$$

Now, using Lemma 9 and the fact that $\epsilon \leq \beta$, we have

$$\begin{aligned}\bar{\Delta}_1 &\leq L\left(\|\boldsymbol{\theta}_{t-\tau_t} - \boldsymbol{\theta}_{t-\tau_t-\tau_{mix}}\|\right)\|\boldsymbol{\theta}_t - \boldsymbol{\theta}^*\| \\ &\leq 4\alpha\tau_{mix}L^2(\|\boldsymbol{\theta}_{t-\tau_t}\|+\beta)r_t \\ &\leq 4\alpha\tau_{mix}L^2(\|\boldsymbol{\theta}_t\|+\epsilon+\beta)r_t \\ &\leq 4\alpha\tau_{mix}L^2(r_t+3\beta)r_t \\ &= 4\alpha\tau_{mix}L^2(r_t^2+3\beta r_t) \\ &\leq 4\alpha\tau_{mix}L^2(3r_t^2+2\beta^2).\end{aligned} \tag{238}$$

We also have that

$$\begin{aligned}\bar{\Delta}_2 = &\underbrace{\langle -\bar{\mathbf{g}}(\boldsymbol{\theta}_{t-\tau_t-\tau_{mix}}) + \mathbf{g}(\boldsymbol{\theta}_{t-\tau_t-\tau_{mix}}, o_{t-\tau_t}), \boldsymbol{\theta}_t - \boldsymbol{\theta}^*\rangle}_{\bar{\Delta}_{21}} \\ &+ \underbrace{\langle -\mathbf{g}(\boldsymbol{\theta}_{t-\tau_t-\tau_{mix}}, o_{t-\tau_t}) + \mathbf{g}(\boldsymbol{\theta}_{t-\tau_t}, o_{t-\tau_t}), \boldsymbol{\theta}_t - \boldsymbol{\theta}^*\rangle}_{\bar{\Delta}_{22}}.\end{aligned} \tag{239}$$

Observe that with calculations analogous to the ones done to arrive at (238), we can obtain

$$\mathbb{E}\left[\bar{\Delta}_{22}\right] \leq 4\alpha\tau_{mix}L^2(3\mathbb{E}\left[r_t^2\right] + 2\beta^2). \tag{240}$$

Now, we write

$$\begin{aligned}\bar{\Delta}_{21} = &\underbrace{\langle -\bar{\mathbf{g}}(\boldsymbol{\theta}_{t-\tau_t-\tau_{mix}}) + \mathbf{g}(\boldsymbol{\theta}_{t-\tau_t-\tau_{mix}}, o_{t-\tau_t}), \boldsymbol{\theta}_{t-\tau_t-\tau_{mix}} - \boldsymbol{\theta}^*\rangle}_{\bar{\Delta}_{211}} \\ &+ \underbrace{\langle -\bar{\mathbf{g}}(\boldsymbol{\theta}_{t-\tau_t-\tau_{mix}}) + \mathbf{g}(\boldsymbol{\theta}_{t-\tau_t-\tau_{mix}}, o_{t-\tau_t}), \boldsymbol{\theta}_t - \boldsymbol{\theta}_{t-\tau_t-\tau_{mix}}\rangle}_{\bar{\Delta}_{212}}.\end{aligned} \tag{241}$$

We first bound $\bar{\Delta}_{212}$. Using Lemma 9 and (27),

$$\begin{aligned}\bar{\Delta}_{212} &\leq \|\bar{\mathbf{g}}(\boldsymbol{\theta}_{t-\tau_t-\tau_{mix}}) - \mathbf{g}(\boldsymbol{\theta}_{t-\tau_t-\tau_{mix}}, o_{t-\tau_t})\|\|\boldsymbol{\theta}_t - \boldsymbol{\theta}_{t-\tau_t-\tau_{mix}}\| \\ &\leq 2L(\|\boldsymbol{\theta}_{t-\tau_t-\tau_{mix}}\|+\sigma)\|\boldsymbol{\theta}_t - \boldsymbol{\theta}_{t-\tau_t-\tau_{mix}}\| \\ &\leq 2L(\|\boldsymbol{\theta}_t\|+\|\boldsymbol{\theta}_t - \boldsymbol{\theta}_{t-\tau_t-\tau_{mix}}\|+\sigma)\|\boldsymbol{\theta}_t - \boldsymbol{\theta}_{t-\tau_t-\tau_{mix}}\| \\ &\leq 2L(\|\boldsymbol{\theta}_t\|+\|\boldsymbol{\theta}_t - \boldsymbol{\theta}_{t-\tau_t-\tau_{mix}}\|+\sigma)\|\boldsymbol{\theta}_t - \boldsymbol{\theta}_{t-\tau_t-\tau_{mix}}\|.\end{aligned} \tag{242}$$

Furthermore, using (200) from Lemma 9 and the fact that $\epsilon \leq \alpha$, we have

$$\begin{aligned}\|\boldsymbol{\theta}_t - \boldsymbol{\theta}_{t-\tau_t-\tau_{mix}}\| &\leq \|\boldsymbol{\theta}_t - \boldsymbol{\theta}_{t-\tau_t}\|+\|\boldsymbol{\theta}_{t-\tau_t} - \boldsymbol{\theta}_{t-\tau_t-\tau_{mix}}\| \\ &\leq \epsilon + \|\boldsymbol{\theta}_{t-\tau_t} - \boldsymbol{\theta}_{t-\tau_t-\tau_{mix}}\| \\ &\leq \epsilon + 4\alpha\tau_{mix}L(r_{t-\tau_t}+2\beta) \\ &\leq \alpha + 4\alpha\tau_{mix}L(r_t+\epsilon+2\beta) \\ &\leq 13\alpha\tau_{mix}L(r_t+\beta).\end{aligned} \tag{243}$$

Hence, given that $\sigma \leq \beta$ and $\alpha\tau_{mix} \leq \frac{1}{26L^2}$,

$$\begin{aligned}\bar{\Delta}_{212} &\leq 2L(\|\boldsymbol{\theta}_t\|+\|\boldsymbol{\theta}_t - \boldsymbol{\theta}_{t-\tau_t-\tau_{mix}}\|+\sigma)\|\boldsymbol{\theta}_t - \boldsymbol{\theta}_{t-\tau_t-\tau_{mix}}\| \\ &\overset{(30)}{\leq} \frac{2L}{26\alpha\tau_{mix}L}\|\boldsymbol{\theta}_t - \boldsymbol{\theta}_{t-\tau_{mix}-\tau_t}\|^2 + \frac{13\alpha\tau_{mix}L^2}{2}(\|\boldsymbol{\theta}_t\|+\|\boldsymbol{\theta}_t - \boldsymbol{\theta}_{t-\tau_{mix}-\tau_t}\|+\beta)^2 \\ &\leq 26\alpha\tau_{mix}L^2(r_t^2+\beta^2) + \frac{13\alpha\tau_{mix}L^2}{2}(2(\|\boldsymbol{\theta}_t\|+\|\boldsymbol{\theta}_t - \boldsymbol{\theta}_{t-\tau_{mix}-\tau_t}\|)^2 + 2\sigma^2) \\ &\leq 26\alpha\tau_{mix}L^2(r_t^2+\beta^2) + 13\alpha\tau_{mix}L^2(2\|\boldsymbol{\theta}_t\|^2+2\|\boldsymbol{\theta}_t - \boldsymbol{\theta}_{t-\tau_{mix}-\tau_t}\|^2+\sigma^2) \\ &\leq 26\alpha\tau_{mix}L^2(r_t^2+\beta^2) + 13\alpha\tau_{mix}L^2(4r_t^2+4\beta^2+2(13^2)\alpha^2\tau_{mix}^2L^2 2(r_t^2+\beta^2)+\sigma^2) \\ &\leq 92\alpha\tau_{mix}L^2(r_t^2+\beta^2).\end{aligned} \tag{244}$$

To conclude, we bound $\mathbb{E}\left[\bar{\Delta}_{211}\right]$. We use (25), the fact that $\alpha\tau_{mix} \leq 1$ and (243). For notational convenience, define $t' \triangleq t - \tau_t - \tau_{mix}$. Using manipulations similar to the ones performed above, we can get

$$\begin{aligned}
\mathbb{E}\left[\bar{\Delta}_{211}\right] &= \mathbb{E}\left[\langle \bar{\mathbf{g}}(\boldsymbol{\theta}_{t'}) - \mathbf{g}(\boldsymbol{\theta}_{t'}, o_{t-\tau_t}), \boldsymbol{\theta}_{t'} - \boldsymbol{\theta}^* \rangle\right] \\
&= \mathbb{E}\left[\langle \bar{\mathbf{g}}(\boldsymbol{\theta}_{t'}) - \mathbb{E}\left[\mathbf{g}(\boldsymbol{\theta}_{t'}, o_{t-\tau_t}) | \boldsymbol{\theta}_{t'}, o_{t'}\right], \boldsymbol{\theta}_{t'} - \boldsymbol{\theta}^* \rangle\right] \\
&\leq \mathbb{E}\left[\|\bar{\mathbf{g}}(\boldsymbol{\theta}_{t'}) - \mathbb{E}\left[\mathbf{g}(\boldsymbol{\theta}_{t'}, o_{t-\tau_t})\|\right] \|\boldsymbol{\theta}_{t'} - \boldsymbol{\theta}^*\|\right] \\
&\leq \mathbb{E}\left[(\alpha)(\|\boldsymbol{\theta}_{t'}\| + \sigma)(\|\boldsymbol{\theta}_{t'} - \boldsymbol{\theta}_t\| + \|\boldsymbol{\theta}_t - \boldsymbol{\theta}^*\|)\right] \\
&\leq \alpha\mathbb{E}\left[(r_t + \|\boldsymbol{\theta}_{t'} - \boldsymbol{\theta}_t\| + \sigma)(r_t + \|\boldsymbol{\theta}_{t'} - \boldsymbol{\theta}_t\| + \beta)\right] \\
&\leq 196L^2\alpha(\mathbb{E}\left[r_t^2\right] + \beta^2).
\end{aligned} \tag{245}$$

Putting together the above bounds, we can conclude that

$$\mathbb{E}\left[-K_{t,3}\right] \leq C_{12}L^2\alpha\tau_{mix}(\mathbb{E}\left[r_t^2\right] + \beta^2), \tag{246}$$

where $C_{12} = 440$. □

**Proof of Lemma 11.** The proof follow the same arguments as in Cohen et al. (2021), we provide details only for completeness. Recall the definition of the average delay up to time $T$, $\tau_{avg} \triangleq \frac{1}{T}\sum_{t=1}^{T}\tau_t$. Consider $U_{2\tau_{avg}}$, the number of steps $t$ for which the delay $\tau_t$ is at least $2\tau_{avg}$. We must have $U_{2\tau_{avg}} \leq \frac{T}{2}$ (otherwise the total sum of delays exceeds $\tau_{avg}T$, contradicting the definition of $\tau_{avg}$). On the other hand, let $k$ be the number of updates that the algorithm makes. Let $t_1 < t_2 < \ldots < t_k$ be the steps in which an update is made. Denote $t_0 = 0$ and $t_{k+1} = T$. Now, fix $i$ and consider the steps at times $s_n = t_i + n$ for $n \in [1, 2, \ldots, t_{i+1} - t_i - 1]$. In all those steps no update takes place and $\boldsymbol{\theta}_{s_n} = \boldsymbol{\theta}_{t_i}$. We must have $\tau_{s_n} > n$ for all $n$ (otherwise $\boldsymbol{\theta}_t = \boldsymbol{\theta}_{t-t_{\tau_t}}$ for $t = s_n$ and an update occurs). In particular, we have that $\tau_{s_n} \geq 2\tau_{avg}$ in at least $t_{i+1} - t_i - 1 - 2\tau_{avg}$ steps. Formally,

$$\begin{aligned}
&\#\text{ of steps in } [t_i, t_{i+1}] \text{ with delay bigger or equal to } 2\tau_{avg} \\
&\qquad \geq \max\{0, t_{i+1} - t_i - 1 - 2\tau_{avg}\} \\
&\qquad \geq t_{i+1} - t_i - 1 - 2\tau_{avg}.
\end{aligned} \tag{247}$$

Hence,

$$\begin{aligned}
U_{2\tau_{avg}} &\geq \sum_{i=0}^{k-1}(t_{i+1} - t_i - 1 - 2\tau_{avg}) \\
&= T - k(1 + 2\tau_{avg}).
\end{aligned}$$

Finally, it follows that $T - k(1 + 2\tau_{avg}) \leq \frac{T}{2}$ which implies $k \geq \frac{T}{4(\tau_{avg}+1)}$. □

### B.3.4 Initial Condition

Now, we provide a bound on the initial condition $r_t^2$ for $t \leq \tau_{mix}$. Note that

$$\begin{aligned}
r_{t+1}^2 &= r_t^2 + 2\alpha\langle \boldsymbol{\theta}_t - \boldsymbol{\theta}^*, \mathbf{g}(\boldsymbol{\theta}_{t-\tau_t}, o_{t-\tau_t})\rangle + \alpha^2\|\mathbf{g}(\boldsymbol{\theta}_{t-\tau_t}, o_{t-\tau_t})\|^2 \\
&\overset{(30)}{\leq} r_t^2 + \alpha r_t^2 + \alpha\|\mathbf{g}(\boldsymbol{\theta}_{t-\tau_t}, o_{t-\tau_t})\|^2 + \alpha^2\|\mathbf{g}(\boldsymbol{\theta}_{t-\tau_t}, o_{t-\tau_t})\|^2 \\
&\leq (1+\alpha)r_t^2 + 2\alpha\|\mathbf{g}(\boldsymbol{\theta}_{t-\tau_t}, o_{t-\tau_t})\|^2 \\
&\overset{(28)}{\leq} (1+\alpha)r_t^2 + 4\alpha L^2(\|\boldsymbol{\theta}_{t-\tau_t}\|^2 + \sigma^2) \\
&\leq (1+\alpha)r_t^2 + 4\alpha L^2(2\epsilon^2 + 2\|\boldsymbol{\theta}_t\|^2 + \sigma^2) \\
&\leq (1 + 17\alpha L^2)r_t^2 + 28\alpha L^2\sigma^2 \\
&\leq (1 + 17\alpha L^2)^{t+1}r_0^2 + 28\alpha L^2\sigma^2\sum_{j=0}^{t}(1 + \alpha L^2)^j.
\end{aligned} \tag{248}$$

Using $\alpha \leq \frac{1}{68\tau_{mix}L^2}$ and $(1+x) \leq e^x \leq 2$ for $x \leq 0.25$, we have:

$$\begin{aligned} r_{\tau_{mix}}^2 &\leq (1+17\alpha L^2)^{\tau_{mix}}\sigma^2 + 28\alpha L^2\sigma^2\tau_{mix}(1+17\alpha L^2)^{\tau_{mix}} \\ &\leq e^{17\alpha L^2\tau_{mix}}\sigma^2 + 28\alpha L^2\sigma^2\tau_{mix}e^{17\alpha L^2\tau_{mix}} \\ &\leq 2\sigma^2 + 56\alpha L^2\sigma^2\tau_{mix} \\ &\leq 3\sigma^2. \end{aligned} \tag{249}$$

### B.3.5 Proof of Theorem 3

Putting together the results of Lemma 10, we have, for $I_t = 1$:

$$\begin{aligned} \mathbb{E}\left[\|\boldsymbol{\theta}_{t+1} - \boldsymbol{\theta}^*\|^2\right] &= \mathbb{E}\left[K_{t,1}\right] + \mathbb{E}\left[\alpha^2 K_{t,2}\right] - \mathbb{E}\left[2\alpha K_{t,3}\right] \\ &\leq \left(1 - 2\alpha\mu + 264\alpha^2\tau_{mix}L^2\right)\mathbb{E}\left[r_t^2\right] + 526\alpha^2\tau_{mix}L^2\beta^2 \\ &\quad + 4\alpha^2L^2\mathbb{E}\left[r_t^2\right] + 12\alpha^2c^2 + 2C_{12}L^2\alpha^2\tau_{mix}(\mathbb{E}\left[r_t^2\right] + \beta^2) \qquad (250) \\ &\leq \left(1 - 2\alpha\mu + 1152\alpha^2\tau_{mix}L^2\right)\mathbb{E}\left[r_t^2\right] + 1418L^2\alpha^2\tau_{mix}\beta^2. \qquad (251) \end{aligned}$$

Assume $t_0^{'}, t_1^{'}, t_2^{'}, \dots t_k^{'} \in [T]$ are iterations in which updates happen after iteration $\tau_{max} + \tau_{mix}$, so that $\tau_{mix} + \tau_{max} \leq t_0^{'} \leq t_1^{'} \leq \cdots \leq t_k^{'}$. Then, we have $r_j = r_{t_i^{'}}$ for all $j \in [t_i^{'}, t_{i+1}^{'})$. We can write, for all $j \in [t_{i+1}^{'}, t_{i+2}^{'})$:

$$\begin{aligned} \mathbb{E}\left[r_j^2\right] &= \mathbb{E}\left[r_{t_{i+1}^{'}}^2\right] \\ &= \mathbb{E}\left[\|\boldsymbol{\theta}_{t_{i+1}^{'}-1} - \boldsymbol{\theta} + \alpha\mathbf{g}(\boldsymbol{\theta}_{t_{i+1}^{'}-1}, o_{t_{i+1}^{'}-1})\|^2\right] \\ &+ \alpha^2\mathbb{E}\left[\|\mathbf{e}_{t_{i+1}^{'}-1}\|^2\right] \\ &- 2\alpha\mathbb{E}\left[\langle \mathbf{e}_{t_{i+1}^{'}-1}\boldsymbol{\theta}_{t_{i+1}^{'}-1} - \boldsymbol{\theta}^* + \alpha\mathbf{g}(\boldsymbol{\theta}_{t_{i+1}^{'}-1}, o_{t_{i+1}^{'}-1})\rangle\right] \\ &= \mathbb{E}\left[\|\boldsymbol{\theta}_{t_i^{'}} - \boldsymbol{\theta} + \alpha\mathbf{g}(\boldsymbol{\theta}_{t_i^{'}}, o_{t_i^{'}})\|^2\right] + \alpha^2\mathbb{E}\left[\|\mathbf{e}_{t_i^{'}}\|^2\right] \\ &- 2\alpha\mathbb{E}\left[\langle \mathbf{e}_{t_i^{'}}\boldsymbol{\theta}_{t_i^{'}} - \boldsymbol{\theta}^* + \alpha\mathbf{g}(\boldsymbol{\theta}_{t_i^{'}}, o_{t_i^{'}})\rangle\right] \\ &\leq \left(1 - 2\alpha\mu + 1152\alpha^2\tau_{mix}L^2\right)\mathbb{E}\left[r_{t_i^{'}}^2\right] \\ &+ 1418L^2\alpha^2\tau_{mix}\beta^2. \end{aligned} \tag{252}$$

By choosing the step size to be $\alpha \leq \frac{\mu}{1152L^2\tau_{mix}}$, we have for all $j \in [t_{i+1}^{'}, t_{i+2}^{'})$:

$$\begin{aligned} \mathbb{E}\left[r_j^2\right] &\leq \left(1 - 2\alpha\mu + 1152\alpha^2\tau_{mix}L^2\right)\mathbb{E}\left[r_{t_i^{'}}^2\right] + 1418L^2\alpha^2\tau_{mix}\beta^2 \\ &\leq \underbrace{(1-\alpha\mu)}_{\rho}\mathbb{E}\left[r_{t_i^{'}}^2\right] + 1418L^2\alpha^2\tau_{mix}\beta^2. \end{aligned} \tag{253}$$

By recursively using the above equation, we have

$$\mathbb{E}\left[r_T^2\right] \leq \rho^k\mathbb{E}\left[r_{t_0^{'}}^2\right] + \gamma, \tag{254}$$

with $\rho = (1-\alpha\mu)$ and $\gamma \triangleq 1418L^2\alpha\tau_{mix}\beta^2\mu^{-1}$. Based on Lemma 11, $k \geq \frac{T}{4\tau_{avg}+4} - \tau_{mix}$, and based on (249), we know $\mathbb{E}\left[r_{t_0^{'}}^2\right] = \mathbb{E}\left[r_{\tau_{mix}}^2\right] \leq 3\sigma^2$. This yields:

$$\begin{aligned} \mathbb{E}\left[r_T^2\right] \leq \rho^k\mathbb{E}\left[r_{t_0^{'}}^2\right] + \gamma &\leq \rho^k 3\sigma^2 + \gamma \leq \exp\left(-\alpha\mu\frac{T}{4\tau_{avg}+4} + \alpha\mu\tau_{mix}\right)3\sigma^2 + 1418\frac{L^2\alpha\tau_{mix}(\epsilon+\sigma)^2}{\mu} \\ &\leq \exp\left(-\alpha\mu\frac{T}{4\tau_{avg}+4}\right)6\sigma^2 + 1418\frac{L^2\alpha\tau_{mix}(\epsilon+\sigma)^2}{\mu}, \end{aligned} \tag{255}$$

where we used $\exp\left(\frac{\mu^2}{1152L^2}\right) \leq \exp\left(\frac{1}{1152L^2}\right) \leq 2$.

Additionally, if we set $\epsilon = \alpha$, $\alpha = \alpha_0 = \frac{\mu}{1152L^2\tau_{mix}}$, we obtain, defining $C_2 \triangleq 1152$,

$$\mathbb{E}\left[r_T^2\right] \leq \left(\exp\left(\frac{-\mu^2 T}{4C_2L^2\tau_{mix}(\tau_{avg}+1)}\right)6+8\right)\sigma^2. \tag{256}$$

This completes the proof.

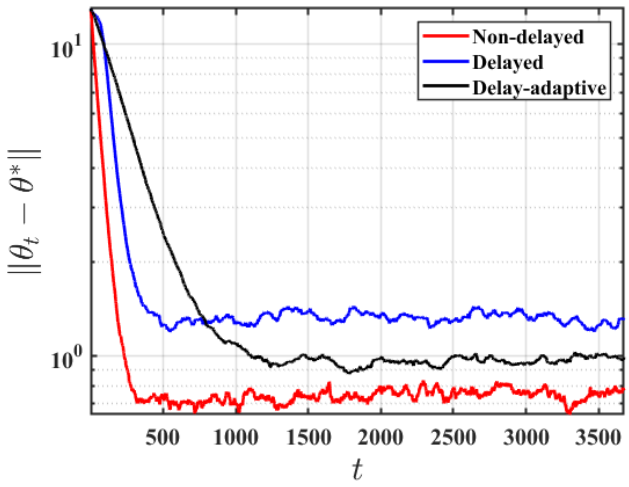


Figure 1: Simulation performance of a TD(0) learning algorithm under delayed updates and Markovian sampling. We compare three different algorithms: a non-delayed TD learning algorithm, a vanilla-delayed algorithm (equivalent to update rule (6)), and a delay-adaptive algorithm (equivalent to update rule (23)).

## C Simulation Results

In this section, we show numerical results to validate our theoretical analysis. We simulate a TD(0) learning algorithm for policy evaluation via linear function approximation, with a setting similar to the one adopted by Dal Fabbro et al. (2023) (see technical report) and Mitra et al. (2023) for TD learning under communication constraints. We generate a Markov Decision Process with a state space $\mathcal{S}$ of dimension $|\mathcal{S}|= 20$, resorting to value function approximation with a feature space spanned by $d = 10$ orthonormal basis vectors. We set the discount factor of the problem to $\gamma = 0.5$. We generate the delay sequence by picking a random delay $\tau_t$ at each iteration $t$ as a uniform random variable in the range $[1, \tau_{max}]$. We set the maximum delay to $\tau_{max} = 200$. We compare a non-delayed version of TD learning ("non delayed") with the delayed scenario, for which we implement both the vanilla approach (update rule in (6)) and the proposed delay-adaptive strategy (update rule in (23)). For the delay-adaptive algorithm, as required by the theory (Theorem 3), we set $\epsilon = \alpha$. To validate our theoretical results, we compare the results obtained with the different algorithms when the step size is the same for all algorithms. Specifically, we set $\alpha = 0.35$. We show the results in Fig. 1. As expected based on the results on Theorem 2 and 3, we see that both the vanilla-delayed and the delay-adaptive algorithms converge to bigger noise balls around the best linear approximation parameter $\boldsymbol{\theta}^*$. At the same time, we see that the rate of convergence of the delay-adaptive algorithm (for the same choice of step-size) is worse than the vanilla-delayed one, while the convergence ball is smaller (for the delay-adaptive). This is in line with the theoretical results. Indeed, if we substitute the same step size in the bounds of Theorem 2 and 3, we get a worse rate for the delay-adaptive algorithm but a better convergence ball. Note that the main advantage of the delay-adaptive algorithm is the fact that, unlike the vanilla-delayed algorithm, the choice of step size does not inversely depend on the maximum delay in the delay sequence. Thus, the delay-adaptive algorithm allows one to be less conservative in the step-size choice and agnostic of the maximum delay.