Bias and Extrapolation in Markovian Linear Stochastic Approximation with Constant Stepsizes

DONGYAN (LUCY) HUO, Cornell University, USA YUDONG CHEN, University of Wisconsin-Madison, USA QIAOMIN XIE, University of Wisconsin-Madison, USA

We consider Linear Stochastic Approximation (LSA) with constant stepsize and Markovian data. Viewing the joint process of the data and LSA iterate as a time-homogeneous Markov chain, we prove its convergence to a unique limiting and stationary distribution in Wasserstein distance and establish non-asymptotic, geometric convergence rates. Furthermore, we show that the bias vector of this limit admits an infinite series expansion with respect to the stepsize. Consequently, the bias is proportional to the stepsize up to higher order terms. This result stands in contrast with LSA under i.i.d. data, for which the bias vanishes. In the reversible chain setting, we provide a general characterization of the relationship between the bias and the mixing time of the Markovian data, establishing that they are roughly proportional to each other.

While Polyak-Ruppert tail-averaging reduces the variance of the LSA iterates, it does not affect the bias. The above characterization allows us to show that the bias can be reduced using Richardson-Romberg extrapolation with $m \geq 2$ stepsizes, which eliminates the m-1 leading terms in the bias expansion. This extrapolation scheme leads to an exponentially smaller bias and an improved mean squared error, both in theory and empirically. Our results immediately apply to the Temporal Difference learning algorithm with linear function approximation, Markovian data, and constant stepsizes.

CCS Concepts: • Theory of computation \rightarrow Stochastic approximation.

Additional Key Words and Phrases: linear stochastic approximation, Markov chain, weak convergence, Wasserstein metric, asymptotic bias, Richardson-Romberg extrapolation

ACM Reference Format:

1 INTRODUCTION

In this paper, we consider the following Linear Stochastic Approximation (LSA) iteration driven by Markovian noise:

$$\theta_{k+1} = \theta_k + \alpha (A(x_k)\theta_k + b(x_k)), \quad k = 0, 1, 2, ...,$$

where $(x_k)_{k\geq 0}$ is a Markov chain representing the underlying data stream, A and b are deterministic functions, and $\alpha>0$ is a constant stepsize. LSA is an iterative data-driven procedure for

Authors' addresses: Dongyan (Lucy) Huo, Cornell University, School of Operations Research and Information Engineering, Ithaca, New York, USA, dh622@cornell.edu; Yudong Chen, University of Wisconsin-Madison, Department of Computer Sciences, Madison, Wisconsin, USA, yudong.chen@wisc.edu; Qiaomin Xie, University of Wisconsin-Madison, Department of Industrial and Systems Engineering, Madison, Wisconsin, USA, qiaomin.xie@wisc.edu.

Permission to make digital or hard copies of all or part of this work for personal or classroom use is granted without fee provided that copies are not made or distributed for profit or commercial advantage and that copies bear this notice and the full citation on the first page. Copyrights for components of this work owned by others than ACM must be honored. Abstracting with credit is permitted. To copy otherwise, or republish, to post on servers or to redistribute to lists, requires prior specific permission and/or a fee. Request permissions from permissions@acm.org.

© 2023 Association for Computing Machinery.

XXXX-XXXX/2023/6-ART \$15.00

https://doi.org/10.1145/nnnnnnn.nnnnnnn

approximating the solution θ^* to the linear fixed point equation $\bar{A}\theta^* + \bar{b} = 0$, where $\bar{A} := \sum_i \pi_i A(i)$, $\bar{b} := \sum_i \pi_i b(i)$, and π is the stationary distribution of the chain $(x_k)_{k \ge 0}$.

Stochastic Approximation (SA), which uses recursive stochastic updates to solve fixed-point equations, is a fundamental algorithmic paradigm in many areas, such as stochastic control and filtering [KY03, Bor08], approximate dynamic programming and reinforcement learning (RL) [Ber19, SB18]. For example, the celebrated Temporal Difference (TD) learning algorithm [Sut88] in RL, potentially equipped with linear function approximation, is a special case of LSA and an important algorithm primitive in RL. Variants of TD algorithm such as $TD(\lambda)$ and Gradient TD, as well as Stochastic Gradient Descent for linear-quadratic estimation, can also be written as LSA [LS18].

Classical work on SA and LSA focuses on the setting with diminishing stepsizes, which allows for asymptotic convergence of θ_k to θ^* [RM51, Blu54, BM00]. Due to its simplicity and fast convergence, SA with constant stepsizes has attracted attention in a growing line of recent work, which establishes non-asymptotic results valid for finite values of k [LS18, SY19, CMSS21b, BRS21]. These results provide $upper\ bounds$ on the mean-squared error (MSE) $\mathbb{E}\|\theta_k-\theta^*\|^2$, and such bounds typically consist of the sum of two terms: a finite-time "bias" term¹ that decays with k, and a steady-state MSE upper bound that is independent of k.

In this work, we study LSA with constant stepsizes in the lens of Markov chain theory. We provide a more precise characterization of the MSE in terms of the decomposition

$$\mathbb{E}\|\theta_k - \theta^*\|^2 \asymp \underbrace{\|\mathbb{E}\theta_k - \mathbb{E}\theta_\infty^{(\alpha)}\|^2}_{\text{optimization error}} + \underbrace{\|\mathbb{E}\theta_\infty^{(\alpha)} - \theta^*\|^2}_{\text{asymptotic bias}^2} + \underbrace{\text{Var}(\theta_k)}_{\text{variance}},$$

where the random variable $\theta_{\infty}^{(\alpha)}$ denotes the limit (as $k \to \infty$) of the LSA iterate θ_k with stepsize α . Our main results characterize the behavior of the three terms above.

Convergence and optimization error. With a constant stepsize α , the process $(x_k, \theta_k)_{k \geq 0}$ is a time-homogeneous Markov chain. We show that under appropriate conditions, the sequence of (x_k, θ_k) converges to a unique limiting random variable $(x_\infty, \theta_\infty^{(\alpha)})$ in distribution and in W_2 , the Wasserstein distance of order 2, regardless of the initial distribution. Moreover, the distribution of $(x_\infty, \theta_\infty^{(\alpha)})$ corresponds to the unique stationary distribution of the chain $(x_k, \theta_k)_{k \geq 0}$. We further provide non-asymptotic bounds on the distributional distance between θ_k and $\theta_\infty^{(\alpha)}$ in W_2 , which in turn upper bounds the optimization error $\|\mathbb{E}\theta_k - \mathbb{E}\theta_\infty^{(\alpha)}\|$. Both bounds decay exponentially in k thanks to the use of a constant stepsize. We emphasize that the existence of the limit θ_∞ and the convergence rate cannot be deduced from the existing upper bound on the MSE $\mathbb{E}\|\theta_k - \theta^*\|^2$, which does not vanish as $k \to \infty$.

Variance and asymptotic bias. By the law of large numbers, the variance $\text{Var}(\theta_k)$ can be eliminated by averaging the LSA iterates. For example, the Polyak-Ruppert tail-averaged iterate $\bar{\theta}_k := \frac{1}{k/2} \sum_{t=k/2}^{k-1} \theta_t$ has variance of order O(1/k). Consequently, for large k, the MSE of $\bar{\theta}_k$ is dominated by the asymptotic bias, i.e., $\mathbb{E}\|\bar{\theta}_k - \theta_*\|^2 \approx \|\mathbb{E}\bar{\theta}_\infty - \theta^*\|^2 = \|\mathbb{E}\theta_\infty^{(\alpha)} - \theta^*\|^2$. Our second main result establishes that the asymptotic bias is proportional to the stepsize α (up to a second order term):

$$\mathbb{E}\theta_{\infty}^{(\alpha)} - \theta^* = \alpha B^{(1)} + O(\alpha^2),\tag{1.1}$$

where $B^{(1)}$ is a vector independent of α and admits an explicit expression in terms of A, b and the transition kernel P of the data Markov chain $(x_k)_{k\geq 0}$. Crucially, equation (1.1) is an *equality* rather than an upper bound. The asymptotic bias is not affected by averaging the LSA iterates.

¹Not to be confused with the *asymptotic* bias discussed below.

Bias expansion and extrapolation. The equality (1.1) implies that bias can be reduced using a simple and clever technique called Richardson-Romberg (RR) extrapolation: run LSA with two stepsizes α and 2α , compute the respective averaged iterates $\bar{\theta}_k^{(\alpha)}$ and $\bar{\theta}_k^{(2\alpha)}$, and output their linear combination $\bar{\theta}_k^{(\alpha)} = 2\bar{\theta}_k^{(\alpha)} - \bar{\theta}_k^{(2\alpha)}$. Doing so cancels out the leading term in the bias characterization (1.1) and results in an order-wise smaller bias $\mathbb{E}\widetilde{\theta}_\infty^{(\alpha)} - \theta^* = O(\alpha^2)$.

In fact, the bias characterization (1.1) extends to higher orders. We establish that the bias admits the following *infinite series expansion*:

$$\mathbb{E}\theta_{\infty}^{(\alpha)} - \theta^* = \alpha B^{(1)} + \alpha^2 B^{(2)} + \alpha^3 B^{(3)} + \cdots, \tag{1.2}$$

where the $B^{(i)}$'s are independent of α . Consequently, RR extrapolation can be executed with any $m \ge 2$ stepsizes to eliminate the m-1 leading terms in equation (1.2), reducing the asymptotic bias to a high order term $O(\alpha^m)$.

When put together, the above results show that the combination of Constant Stepsize, Averaging, and Extrapolation allows one to approach the *best of three worlds*: (a) using a constant stepsize leads to fast, geometric-in-k convergence for the optimization error, (b) tail-averaging eliminates the variance at an (optimal) 1/k rate, and (c) RR extrapolation order-wise reduces the asymptotic bias. We highlight that the m iterate sequences used in RR extrapolation can be computed in parallel, using the same data stream $(x_k)_{k\geq 0}$. Therefore, compared with standard LSA, the above-combined procedure is data efficient (in terms of the sample complexity k for achieving a given MSE), does not require sophisticated tuning of stepsize, and incurs a minimal increase in computational cost.

The results above should be contrasted with the setting of LSA with *i.i.d.* data, where the x_k 's are sampled independently from the distribution π . In this setting, it has been shown (sometimes implicitly) in existing work that the asymptotic bias is zero [LS18, MLW⁺20]. Such a result should not be surprising, as similar results are well known in the literature on stochastic gradient descent (SGD) for optimizing quadratic functions given i.i.d. data, for which the SGD update is linear. It is perhaps surprising that using Markovian data leads to a non-zero asymptotic bias, even when the LSA iteration is linear in θ_k . In Figure 1, we provide the dependency graphs for LSA with i.i.d. data and Markovian data. In the Markovian setting, the correlation between the x_k 's leads to additional correlation among the iterate θ_k 's; in particular, the iterate sequence $(\theta_k)_{k\geq 0}$ is no longer a Markov chain by itself. As such, θ_{k+1} has an implicit, nonlinear dependence on θ_k through (x_{k-1}, x_k) . This non-linearity is the source of the asymptotic bias.

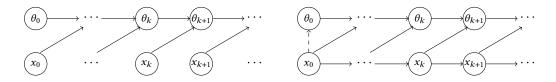


Fig. 1. Dependency Graphs of LSA. Left: i.i.d. data. Right: Markovian data.

Bias and mixing time. We generalize and quantify the observations above by relating the asymptotic bias to the mixing time of the underlying Markov chain $(x_k)_{k\geq 0}$ and the second largest eigenvalue modulus (SLEM) $|\lambda_2|$ of the transition kernel. We show that the leading coefficient $B^{(1)}$ in the expansion (1.2) has norm upper bounded by $O(\frac{|\lambda_2|}{1-|\lambda_2|})$, at least in the setting where the chain $(x_k)_{k\geq 0}$ is reversible. It is well known that the mixing time of $(x_k)_{k\geq 0}$ can be tightly upper and lower bounded by $|\lambda_2|$ [LP17]. Consequently, the faster the underlying chain $(x_k)_{k\geq 0}$ mixes, the

smaller the asymptotic bias is. As a special case, LSA with i.i.d. data has zero mixing time and $|\lambda_2|$, hence zero bias.

All our results can be immediately specialized to the TD algorithm in RL with linear function approximation and Markovian data. For both LSA and TD, we provide numerical results that corroborate the prediction of our theory and demonstrate the benefit of using constant stepsizes, tail averaging and RR extrapolation.

Paper Organization: In Section 2, we review existing results related to our work. We formalize the problem and assumptions in Section 3, and present our main results in Section 4. In Section 5, we provide numerical results for both LSA and TD. We outline the proofs of the main results in Section 6. The paper is concluded in Section 7 with a discussion of future directions.

2 RELATED WORK

In this section, we review existing results that are most related to our work.

2.1 Classical Results on Stochastic Approximation

The study of stochastic approximation can be traced back to the work of Robbins and Monro [RM51]. Under suitable assumptions, Robbins and Monro [RM51] prove that the SA algorithm asymptotically converges in L^2 , and Blum [Blu54] shows that the convergence holds almost surely. Subsequent works [Rup88, Pol90] propose the technique of iterate averaging, now known as the Polyak-Ruppert (PR) averaging, which improves the convergence rates. A Central Limit Theorem (CLT) for asymptotic normality of the averaged iterates is established in [PJ92]. Borkar and Meyn [BM00] introduce the Ordinary Differential Equation (ODE) technique for analyzing SA algorithms, and they show that under certain conditions, the SA iterates approximate the solution of a suitable ODE. Utilizing the ODE technique, recent work [BCD+21] establishes a functional CLT for SA driven by Markovian noise.

The asymptotic theory of SA is well-developed and covered in several excellent textbooks [KY03, Bor08, BMP12]. Convergence results in classical SA work typically assume that the stepsize α_k at iteration k satisfies: $\sum_{k=1}^{\infty} \alpha_k = \infty$ and $\sum_{k=1}^{\infty} \alpha_k^2 < \infty$, This assumption implies that the stepsize sequence is diminishing but square-summable. Our work, on the other hand, focuses on the setting of constant stepsizes, i.e., $\alpha_k = \alpha$ for all $k \ge 0$.

2.2 SA and SGD with Constant Stepsizes

Using constant stepsizes has been a popular choice in practice due to fast convergence and easy implementation. Recent years have witnessed a growing body of work on the constant stepsize setting of SA and the closely related Stochastic Gradient Descent (SGD) algorithm.

A majority of work in this line studies SA and SGD under the i.i.d. noise assumption, and some provide finite-time bounds. The work in [LS18] analyzes LSA and establishes finite-time upper and lower bounds on the MSE. The work [MLW $^+$ 20] provides refined results, establishing tight bounds with the optimal dependence on problem-specific constants as well as a CLT for the averaged iterates with a characterization of the exact asymptotic covariance matrix. A line of recent work makes use of new results on random matrix products to analyze LSA: the work [DMN $^+$ 21] establishes tight concentration bounds of LSA, and the paper [DMNS22] extends these bounds to LSA with iterate averaging.

The work in [DDB20] studies constant stepsize SGD for strongly convex and smooth functions. By connecting SGD to classical Markov chain analysis, they establish that the iterates converge to a unique stationary distribution. This result is generalized to non-convex and non-smooth functions

with quadratic growth in the work [YBVE21]. The paper [YBVE21] establishes asymptotic normality of the averaged SGD iterates, and subsequent work [CMM22] studies the limit of the stationary distribution as stepsize goes to zero. These results are established under the i.i.d. noise setting.

More recent work studies constant-stepsize SA under Markovian noise. The work [SY19] provides finite-time bounds on the MSE of LSA. The work [MPWB21] considers LSA with PR averaging and establishes instance-dependent MSE upper bounds with tight dimension dependence. Some papers [SY19, DMNS22] also provide instance-dependent bounds on higher moments of LSA iterates. Going beyond LSA, the work [CMSS20, CMSS21b] considers general SA with contractive mapping and provides finite-time convergence results.

A portion of our results are similar in spirit to [DDB20, Proposition 2] and [DMN+21, Theorem 3], in that we both study LSA and SGD with constant stepsizes in the lens of time-homogeneous Markov chains. A crucial difference is that we consider the Markovian data setting whereas they consider i.i.d. data. Arising naturally in stochastic control and RL problems, the Markovian setting leads to non-zero asymptotic bias and new analytical challenges, which are not present in the i.i.d. setting. Our analysis for the Markovian setting involves more delicate coupling arguments and builds on the Lyapunov function techniques from [SY19]. Along the way, we obtain a refinement of the MSE bounds from the work [SY19]. We discuss these analytical challenges and improvements in greater detail after stating our theorems; see Sections 4 and 6.

2.3 Applications in Reinforcement Learning and TD Learning

Many iterative algorithms in RL aim to solve for the fixed point of Bellman equations and can be viewed as special cases of SA [SB18, Ber19]. For example, the TD algorithms [Sut88] with linear function approximation, including TD(0) and more generally TD(λ), are LSA procedures. Our results can be specialized to TD learning and hence are related to existing work in this line.

Classical results on TD Learning, similarly to those on SA, focus on asymptotic convergence under diminishing stepsizes [Sut88, Day92, DS94, TVR97]. More recent works provide finite-time results. The work [DSTM18] is among the first to provide MSE and concentration bounds for linear TD learning in its original form without any centering or projection steps, and their analysis assumes diminishing stepsize and i.i.d. noise. The work [BRS21] presents finite-time analysis of TD(0) under both i.i.d. and Markovian noise, with both diminishing and constant stepsizes. Their results require adding a projection step in TD(0) to ensure boundedness. The Lyapunov analysis in [SY19] on LSA with constant stepsize, when specialized to TD(0), removes this projection step and proves similar upper bounds on the MSE. The recent work in [CMSS21a, CMSS21b] uses Lyapunov theory to study the tabular TD and obtains finite sample convergence guarantees. The paper [KPR+21] provides sharp, instance-dependent ℓ_{∞} error bounds for the tabular TD algorithm with i.i.d. data.

Q-learning [WD92], a standard algorithm in RL for estimating the optimal action-value function, can be viewed as a (nonlinear) SA procedure with general contractive mappings. Classical works in Q-learning [Tsi94, Sze97, EDMB03] establish asymptotic convergence under diminishing stepsizes. More recent works characterize finite-time convergence rates. The work [CMSS21b] establishes finite-time MSE bounds for tabular Q-learning with constant stepsizes and Markovian noise. The work [CBD22] presents high probability bounds for Q-learning with diminishing stepsizes. While related, results on Q-learning are not directly comparable to ours, as Q-learning involves a nonlinear update.

3 SET-UP AND ASSUMPTIONS

In this section, we formally set up the problem and the assumptions/notations used in the sequel.

3.1 Problem Set-up

3Let $(x_k)_{k\geq 0}$ be a Markov chain on the state space X with stationary distribution π . Consider the following linear stochastic approximation iteration

$$\theta_{k+1}^{(\alpha)} = \theta_k^{(\alpha)} + \alpha \left(A(x_k) \theta_k^{(\alpha)} + b(x_k) \right), \quad k = 0, 1, \dots,$$
(3.1)

where $A: X \to \mathbb{R}^{d \times d}$ and $b: X \to \mathbb{R}^d$ are fixed functions, and $\alpha > 0$ is a constant stepsize. In what follows, we omit the superscript in $\theta_k^{(\alpha)}$ when the dependence on α is clear from the context. The initial distribution of θ_0 is arbitrary and may depend on x_0 , as illustrated by the dotted line between x_0 and θ_0 in the dependency graph in the right pane of Figure 1. We assume that θ_0 is independent of $(x_k)_{k\geq 1}$ given x_0 .

The iterative procedure (3.1) is a common approach for approximating the target vector $\theta^* \in \mathbb{R}^d$, defined as the solution to the steady-state equation

$$\mathbb{E}_{\pi}[A(x)]\theta + \mathbb{E}_{\pi}[b(x)] = 0, \tag{3.2}$$

where $\mathbb{E}_{\pi}[\cdot]$ denotes the expectation with respect to $x \sim \pi$, and π is the stationary distribution of the Markov chain (x_k) . Our general goal is to characterize the relationship between the finite-time iterate θ_k and the target solution θ^* .

The stochastic process $(\theta_k)_{k\geq 0}$ of the LSA iterates is not a Markov chain itself. In particular, given θ_k , the random variables θ_{k+1} and θ_{k-1} are correlated through the underlying Markov process (x_0, x_1, \ldots, x_k) . However, it is easy to see that the joint process $(x_k, \theta_k)_{k\geq 0}$ is a Markov chain on the state space $X \times \mathbb{R}^d$, and this chain is time-homogeneous as the stepsize α is independent of k. Moreover, the following independence property holds:

$$(\theta_0, x_0, \theta_1, x_1, \dots, \theta_k) \perp (x_{k+1}, x_{k+2}, \dots) \mid x_k, \quad \forall k \ge 1.$$
 (3.3)

Consequently, we have $\theta_k \perp \!\!\! \perp x_{k+1} \mid x_k$ for all $k \geq 1$. The above facts, which we use repeatedly in the sequel, can be proved by direct calculation. Alternatively, one may verify that the joint distribution of $(x_k, \theta_k)_{k \geq 0}$ obeys the Markov property with respect to the directed acyclic graph in the right pane of Figure 1, hence the aforementioned (in)dependence properties follow from standard results on directed graphical models [CDLS99, Corollary 5.11 and Theorem 5.14].

Part of our analysis makes use of the notion of time reversal of Markov chains. Let $P=(p_{ij})$ denote the probability transition kernel of the Markov chain $(x_k)_{k\geq 0}$. If P is irreducible, then the chain $(x_k)_{k\geq 0}$ starting from the stationary distribution and running backward in time is also a Markov chain, whose transition kernel $\hat{P}=(\hat{p}_{ij})$ is given by the equation

$$\pi_i \hat{p}_{ii} = \pi_i p_{ij}$$
 for all i, j .

Moreover, \hat{P} is also irreducible and its stationary distribution is the same as that of P [Nor97, Theorem 1.9.1]. The Markov chain $(x_k)_{k\geq 0}$ is called *reversible* if $\hat{P} = P$.

3.2 Assumptions

Below we state the assumptions needed for our main theorems.

Assumption 1. $(x_k)_{k\geq 0}$ is an irreducible and aperiodic Markov chain on a finite state space $X = \{1, 2, ..., n\}$ with stationary distribution π . In addition, the distribution of the initial state x_0 is π .

Under the first part of Assumption 1, the Markov chain $(x_k)_{k\geq 0}$ is positive recurrent with a unique stationary distribution π , and the chain converges to π from any initial distribution of x_0 . Moreover, the chain has a geometric mixing rate [LP17, Theorem 4.9]. That is, there exist constants $\rho \in [0, 1)$ and c > 0 such that

$$\max_{x \in \mathcal{X}} \left\| (p_x^k)^\top - \pi \right\|_{\text{TV}} \le c \rho^k, \tag{3.4}$$

where $(p_x^k)^{\top}$ denotes the x-th row of the k-step transition matrix P^k and $\|\cdot\|_{\text{TV}}$ is the total variation norm.

We impose the additional stationarity assumption $x_0 \sim \pi$ so as to simplify several mathematical expressions. This assumption is not essential and can be relaxed by applying our analysis after the chain $(x_k)_{k\geq 0}$ has approximately mixed, which happens quickly owing to the geometric mixing property (3.4). Note that the same stationarity assumption is also used in several previous works [BRS21, MPWB21]. We believe the assumption of a finite state space $\mathcal X$ can also be relaxed and replaced by an appropriate mixing condition (such as those from [SY19, MPWB21]); we do not pursue this direction in this paper. The irreducible and aperiodic assumptions are also standard in the literature (e.g., [SY19, BRS21, CMSS21b]) and ensure a well-defined limiting stationary distribution, which is the object of focus in this paper.

Our next two assumptions are similar to those used in the work [SY19, DMNS22]. Below we use $\|\cdot\|$ to denote the Euclidean ℓ_2 -norm for vectors and the spectral norm for matrices (i.e., the largest singular value of matrices).

Assumption 2. We assume that

$$A_{\max} := \max_{x \in \mathcal{X}} \|A(x)\| \leq 1 \quad \text{and} \quad b_{\max} := \max_{x \in \mathcal{X}} \|b(x)\| < \infty.$$

Under Assumptions 1 and 2, we know that the following two limits exist,

$$\lim_{k \to \infty} \mathbb{E}[A(x_k)] = \bar{A} := \mathbb{E}_{\pi}[A(x)],$$

$$\lim_{k \to \infty} \mathbb{E}[b(x_k)] = \bar{b} := \mathbb{E}_{\pi}[b(x)],$$
(3.5)

and there hold the bounds

$$\|\bar{A}\| \le A_{\max} \le 1$$
 and $\|\bar{b}\| \le b_{\max}$.

Playing an important role in our analysis is the mixing time of the Markov chain $(x_k)_{k\geq 0}$ with respect to the functions $A(\cdot)$ and $b(\cdot)$, defined as follows.

Definition 3.1. For $\epsilon \in (0,1)$, define $\tau_{\epsilon} \geq 1$ to be the ϵ -mixing time of $(x_k)_{k\geq 0}$ in the sense that

$$\left\| \mathbb{E}[A(x_k)|x_0 = x] - \bar{A} \right\| \le \epsilon \cdot A_{\max}, \quad \forall x \in \mathcal{X}, \forall k \ge \tau_{\epsilon}, \tag{3.6}$$

$$\left\| \mathbb{E}[b(x_k)|x_0 = x] - \bar{b} \right\| \le \epsilon \cdot b_{\max}, \quad \forall x \in \mathcal{X}, \forall k \ge \tau_{\epsilon}.$$
 (3.7)

Under Assumptions 1 and 2, the ϵ -mixing time satisfies $\tau_{\epsilon} \leq K \log \frac{1}{\epsilon}$ for all $\epsilon \in (0, 1)$, where the number $K \geq 1$ is independent of ϵ . This fact can be seen from the following inequality

$$\left\|\mathbb{E}[b(x_k)|x_0=x]-\bar{b}\right\| = \left\|\sum_{i\in Y}(p_{xi}^k - \pi_i)b(i)\right\| \leq b_{\max}\cdot \left(2\max_{x\in X}\|p_x^k - \pi\|_{\text{TV}}\right) \leq 2cb_{\max}\cdot \rho^k,$$

where the last step follows from the geometric mixing bound (3.4); a similar argument applies to $A(x_k)$.

In the sequel, unless specified otherwise, we always choose $\epsilon = \alpha$ and write $\tau \equiv \tau_{\alpha}$.

Assumption 3. The matrix \bar{A} is Hurwitz, i.e., all eigenvalues have strictly negative real parts. Therefore, there exists a symmetric positive definite matrix Γ such that

$$\bar{A}^{\mathsf{T}}\Gamma + \Gamma\bar{A} = -I,\tag{3.8}$$

where I is the d-by-d identity matrix.

The Hurwitz Assumption 3 is standard in the study of the stability of dynamical systems. Under this assumption, the matrix \bar{A} is invertible and hence $s_{\min}(\bar{A}) > 0$, where $s_{\min}(\bar{A})$ denotes the smallest singular value of \bar{A} . Denote by γ_{\min} and γ_{\max} the minimum and maximum eigenvalues of the matrix Γ , respectively. By assumption we have $\gamma_{\max} \geq \gamma_{\min} > 0$ and

$$\gamma_{\min} \|v\|^2 \le v^{\mathsf{T}} \Gamma v \le \gamma_{\max} \|v\|^2, \quad \forall v \in \mathbb{R}^d.$$
(3.9)

3.3 Notations

In the sequel, we generally adopt the following notational convention. Upper case letters (e.g., M) denote matrices and lower case letters (e.g., u) denote vectors or scalars; these quantities may be deterministic or random. The lowercase letter c and its derivatives c', c_0 , etc. denote universal numerical constants, whose values may change from line to line.

Slightly deviating from the above convention, we use $B \equiv B(A, b, P)$ and its derivatives B', B_0 , etc. to denote quantities (vectors or matrices) that depend only on A, b and P, but independent of the stepsize α and the iteration index k. We use $C \equiv C(A, b, \pi)$ to denote a quantity that depends only on A, b and π , but is independent of α , k and other properties of P (i.e., its mixing time). While one may overcast $C(A, b, \pi)$ with B(A, b, P), we distinguish these two types of quantities whenever possible. This distinction allows us to isolate the impact of π and its corresponding Perron-Frobenius eigenvalue 1 from the remaining eigenvalues/vectors of P, the transition kernel of the underlying Markov chain $(x_k)_{k\geq 0}$. Note that these remaining eigenvalues/vectors, including the eigenvalue with the second largest modulus (SLEM), determine the mixing time of P. The above convention plays an important role in our results concerning the relationship to mixing time.

As we are primarily interested in the scaling relationship of various quantities with respect to the stepsize α and the number of iteration k, we make use of the following big-O notation: for a given function f of α and k, we write $h = O(f(\alpha, k))$ if it holds that

$$||h|| \le B(A, b, P) \cdot f(\alpha, k)$$

for some B(A, b, P) independent of α and k, where $\|\cdot\|$ denotes the Euclidean ℓ_2 -norm for vectors and the spectral norm for matrices (i.e., the largest singular value). For example, $h = O(\alpha/k)$ means $\|h\| \le B(A, b, P) \cdot \alpha/k$.

For a random variable z, let $\mathcal{L}(z)$ denote the law/distribution of z. Let $\mathcal{P}_2(\mathbb{R}^d)$ be the space of square-integrable distributions on \mathbb{R}^d , and similarly let $\mathcal{P}_2(X \times \mathbb{R}^d)$ be the set of probability measures \bar{v} on the product space $X \times \mathbb{R}^d$ with the property that the marginal of \bar{v} on \mathbb{R}^d is square-integrable. Let I_m and 1_m denote the $m \times m$ identity matrix and the m-dimensional all-one vector, respectively. We omit the subscript when the dimension m is clear from the context.

4 MAIN RESULTS

In this section, we present our main results. In Section 4.1, we prove that the LSA iterates $(x_k, \theta_k)_{k\geq 0}$, viewed as a time-homogeneous Markov chain, converge to a unique limiting distribution, and we provide its explicit convergence rates. In Sections 4.2 and 4.3, we further characterize the above limit and its relationship with the stepsize and mixing time. We explore the implications of these results for PR tail averaging, RR extrapolation and linear TD(0) Learning in Sections 4.4 and 4.5.

4.1 Convergence to Limit Distribution

Our convergence results are based on the Wasserstein distance of order 2; see [Vil09, Chapter 6] for an introduction to this topic. The Wasserstein-2 distance between two probability measures μ

and ν in $\mathcal{P}_2(\mathbb{R}^d)$ is defined as

$$W_{2}(\mu, \nu) = \inf_{\xi \in \Pi(\mu, \nu)} \left(\int_{\mathbb{R}^{d}} \|u - v\|^{2} d\xi(u, v) \right)^{1/2}$$

= $\inf \left\{ \left(\mathbb{E}[\|\theta - \theta'\|^{2}] \right)^{1/2} : \mathcal{L}(\theta) = \mu, \mathcal{L}(\theta') = \nu \right\},$

where $\Pi(\mu, \nu)$ denotes the set of all couplings between μ and ν , i.e., the collection of joint distributions in $\mathcal{P}_2(\mathbb{R}^d \times \mathbb{R}^d)$ with marginal distributions μ and ν . To study the joint process $(x_k, \theta_k)_{k \geq 0}$, we extend the above Wasserstein-2 distance to the space $\mathcal{P}_2(X \times \mathbb{R}^d)$. Let $d_0: X \times X \to \mathbb{R}_+$ denote the discrete metric on X, that is, $d_0(x, x') := \mathbb{1}\{x \neq x'\}$. Define the following metric \bar{d} on the product space $X \times \mathbb{R}^d$:

$$\bar{d}((x,\theta),(x',\theta')) := \sqrt{d_0(x,x') + \|\theta - \theta'\|^2}.$$

For a pair of distributions $\bar{\mu}$ and $\bar{\nu}$ in $\mathcal{P}_2(X \times \mathbb{R}^d)$, we consider the following Wasserstein-2 distance w.r.t. the metric \bar{d} :

$$\bar{W}_{2}(\bar{\mu}, \bar{\nu}) = \inf \left\{ \left(\mathbb{E} \left[\bar{d}(z, z')^{2} \right] \right)^{1/2} : \mathcal{L}(z) = \bar{\mu}, \ \mathcal{L}(z') = \bar{\nu} \right\} \\
= \inf \left\{ \left(\mathbb{E} \left[d_{0}(x, x') + \|\theta - \theta'\|^{2} \right] \right)^{1/2} : \mathcal{L}((x, \theta)) = \bar{\mu}, \ \mathcal{L}((x', \theta')) = \bar{\nu} \right\}.$$
(4.1)

The simple relationship below follows directly from the definition:

$$W_2(\mathcal{L}(\theta), \mathcal{L}(\theta')) \leq \bar{W}_2(\mathcal{L}(x, \theta), \mathcal{L}(x', \theta')).$$

Also note that convergence in W_2 or \overline{W}_2 implies the usual convergence in distribution plus the convergence of the first two moments [Vil09, Definition 6.8, Theorem 6.9].

Our first theorem establishes the convergence of the Markov chain $(x_k, \theta_k)_{k \ge 0}$ in \bar{W}_2 to a unique stationary distribution and characterizes the convergence rate.

Theorem 4.1. Suppose that Assumptions 1, 2 and 3 hold, and the stepsize α satisfies

$$\alpha \tau_{\alpha} < \frac{0.05}{95 \gamma_{\text{max}}}.\tag{4.2}$$

(1) Under all initial distributions of θ_0 , the sequence of random variables $(x_k, \theta_k)_{k \geq 0}$ converges in \overline{W}_2 to a unique limit $(x_\infty, \theta_\infty) \sim \overline{\mu}$. Moreover, it holds that

$$\operatorname{Var}(\theta_{\infty}) \leq \frac{\gamma_{\max}}{0.9\gamma_{\min}} \cdot \alpha \tau_{\alpha} \cdot \kappa,$$

where

$$\kappa := 640\gamma_{\text{max}} \cdot s_{\text{min}}^{-2}(\bar{A}) \cdot b_{\text{max}}^2. \tag{4.3}$$

- (2) $\bar{\mu}$ is the unique stationary distribution of the Markov chain $(x_k, \theta_k)_{k>0}$.
- (3) Let $\mu := \mathcal{L}(\theta_{\infty})$ be the second marginal of $\bar{\mu}$. For all $k \geq \tau_{\alpha}$, it holds that

$$W_{2}^{2}(\mathcal{L}(\theta_{k}), \mu) \leq \bar{W}_{2}^{2}(\mathcal{L}(x_{k}, \theta_{k}), \bar{\mu}) \leq 16 \frac{\gamma_{\max}}{\gamma_{\min}} \left(\mathbb{E}[\|\theta_{0}\|^{2}] + \mathbb{E}[\|\theta_{\infty}\|^{2}] \right) \cdot \left(1 - \frac{0.9\alpha}{\gamma_{\max}} \right)^{k}. \tag{4.4}$$

We outline the proof of Theorem 4.1 in Section 6.2, deferring the complete proof to Appendix A.2. Theorem 4.1 states that the Markov chain $(x_k, \theta_k)_{k \geq 0}$ converges geometrically to a unique limit $(x_\infty, \theta_\infty) \sim \bar{\mu}$. Note that the limiting distribution $\bar{\mu}$ is in general not a product distribution of its marginals π and μ . This convergence result is valid under the stepsize condition (4.2), stated as an upper bound on the product $\alpha \tau_\alpha$. Since $\tau_\alpha \leq K \log \frac{1}{\alpha}$ for some constant $K \geq 1$ independent of α (see Section 3.2), the condition (4.2) is satisfied for sufficiently small α .

We remark on the techniques for proving Theorem 4.1. To establish the convergence of a Markov chain and the existence of stationary distribution, a standard approach is to show that the chain is positive recurrent by verifying irreducibility and Lyapunov drift conditions. This approach has been developed for Markov chains on general state spaces [MT09] and is adopted in the prior work [YBVE21, BCD+21, LM22]. However, it is not clear how to implement this approach for the LSA iteration (3.1). For example, suppose that the stepsize α and the functions A and b take on rational values. If the initial θ_0 is rational, then θ_k only takes rational values for all $k \geq 0$. If θ_0 is irrational, then θ_k remains irrational. As such, it seems challenging to certify ψ -irreducibility and recurrence for the Markov chain $(x_k, \theta_k)_{k \ge 0}$ on the state space $X \times \mathbb{R}^d$. Instead, we prove weak convergence to a limiting distribution through the convergence in the Wasserstein distance, which can be bounded via coupling arguments. The Wasserstein distance is also used in works [DDB20, DMN+21] to study SGD and LSA under the i.i.d. data assumption; this assumption is heavily relied upon in their analysis. Moreover, their analysis requires the contraction property $W_2^2(\mathcal{L}(\theta_{k+1}), \mu) < W_2^2(\mathcal{L}(\theta_k), \mu)$. Establishing this property in our Markovian setting is difficult if not impossible. Our proof makes use of alternative and substantially more delicate coupling arguments.

Convergence of θ_k in the \overline{W}_2 metric implies convergence of the first two moments. Explicit convergence rates can be deduced from Theorem 4.1, as recorded in the following corollary.

COROLLARY 4.2. Under the setting of Theorem 4.1, for all $k \geq \tau_{\alpha}$ we have

$$\|\mathbb{E}[\theta_k] - \mathbb{E}[\theta_\infty]\| \le C \cdot \left(1 - \frac{0.9\alpha}{\gamma_{\text{max}}}\right)^{k/2} \tag{4.5}$$

and

$$\left\| \mathbb{E} \left[\theta_k \theta_k^{\mathsf{T}} \right] - \mathbb{E} \left[\theta_{\infty} \theta_{\infty}^{\mathsf{T}} \right] \right\| \le C' \cdot \left(1 - \frac{0.9\alpha}{\gamma_{\max}} \right)^{k/2} \tag{4.6}$$

for some $C \equiv C(A, b, \pi)$ and $C' \equiv C'(A, b, \pi)$ that are independent of α and k.

The proof of Corollary 4.2 is given in Appendix A.3.

Theorem 4.1 and Corollary 4.2 can be contrasted with convergence results for LSA with a diminishing stepsize sequence, under which the iterates typically converge almost surely to the true solution θ^* , with a sublinear convergence rate [Blu54, BRS21, CMSS21b]. For example, when using a diminishing stepsize sequence of the form $\alpha_k = a/(b+k)^c$ with $c \in (0,1]$ and appropriate values of a,b, the LSA iterates θ_k converge to θ^* in MSE at a rate $O(1/k^c)$ [CMSS21b, Theorem 2.1]. In comparison, constant-stepsize LSA converges weakly and may have a nonzero bias $\mathbb{E}\theta_\infty - \theta^*$; see Theorem 4.3 in the next sub-section. On the other hand, using a constant stepsize leads to a geometric convergence rate of the form $O((1-v)^k)$ and hence faster initial convergence, as demonstrated in our numerical experiments in Section 5.

In practice, when using a diminishing stepsize sequence $\alpha_k = a/(b+k)^c$, the values of a, b and c have a significant impact on the convergence behavior and need to be tuned carefully. For example, if c=1 and a is overly small, then the O(1/k) rate will be lost and the convergence performance will degrade substantially [CMSS21b]. In comparison, LSA with constant stepsizes has only one tuning parameter and often enjoys a more robust performance [LS18].

4.2 Bias Expansion

Theorem 4.1 above establishes the convergence of $\theta_k^{(\alpha)}$ to a limit $\theta_\infty^{(\alpha)}$. Our next theorem provides a characterization of $\mathbb{E}[\theta_\infty^{(\alpha)}] - \theta^*$, the asymptotic bias, as a function of the stepsize α .

THEOREM 4.3. Suppose that Assumptions 1, 2 and 3 hold, and α satisfies equation (4.2). The following hold for some vectors $B^{(i)} \equiv B^{(i)}(A, b, P)$, i = 1, 2, ... that are independent of α .

(1) For each m = 1, 2, ..., we have the expansion

$$\mathbb{E}[\theta_{\infty}^{(\alpha)}] = \theta^* + \sum_{i=1}^m \alpha^i B^{(i)} + O(\alpha^{m+1}). \tag{4.7}$$

(2) For α in a neighborhood of 0, we have the infinite series expansion

$$\mathbb{E}[\theta_{\infty}^{(\alpha)}] = \theta^* + \sum_{i=1}^{\infty} \alpha^i B^{(i)}. \tag{4.8}$$

We outline the proof of Theorem 4.3 in Section 6.3, with the complete proof given in Appendix A.4. Theorem 4.3 is akin to a Taylor series expansion of $\mathbb{E}[\theta_{\infty}^{(\alpha)}]$ with respect to α . The existence of such an expansion is non-trivial: $\theta_{\infty}^{(\alpha)}$ is undefined at $\alpha=0$, and it is not clear a priori whether $\mathbb{E}[\theta_{\infty}^{(\alpha)}]$ is a differentiable and analytic function of α . We emphasize that equations (4.7) and (4.8) are equalities, hence the bias is non-zero whenever $B^{(i)} \neq 0$. In particular, averaging the LSA iterates θ_k does not affect this bias and only reduces the variance.

The proof of Theorem 4.3 is based on the following idea. As discussed in Section 1, the asymptotic bias arises due to the implicit nonlinear dependence of θ_{k+1} on θ_k as both of them depend on the state x_k of the underlying Markov chain. If θ_k were independent of x_k , the bias would be zero. This observation suggests that the bias is determined by the strength of dependence between θ_k and x_k , which can be quantified by the variation of the conditional expectation $\mathbb{E}[\theta_k|x_k=i]$ as a function of $i \in X$. Therefore, our analysis is based on understanding this conditional expectation in steady state, namely $\mathbb{E}[\theta_{\infty}|x_{\infty}=i]$. We characterize this quantity using the Basic Adjoint Relationship (BAR) [Har85, HW87, DD11] for the steady state with a specific choice of test functions.

Our proof is non-asymptotic in nature. See equation (A.43) for an explicit upper bound on α under which part 2 of Thoerem 4.3 holds. If desired, one can read off from the proof an explicit expression or upper bound for the coefficients $B^{(i)}$ of the bias expansions; we provide such an explicit bound for $B^{(1)}$ in Section 4.3 for reversible P. Importantly, regardless of the functional form of $B^{(i)}$, Richardson-Romberg extrapolation can be used to cancel the lower order terms of α in the expansions (4.7) and (4.8), which reduces the bias to a higher order term of α . These results are formally presented in Section 4.4.

4.3 Bias and Mixing Time

As mentioned, the bias $\mathbb{E}[\theta_{\infty}] - \theta^*$ arises due to the Markovian correlation in the data $(x_k)_{k\geq 0}$. If the chain $(x_k)_{k\geq 0}$ mixes slowly, the correlation is strong and intuitively leads to a large bias. Our next result formalizes this intuition, quantifying the relationship between the bias and the mixing time of the chain $(x_k)_{k\geq 0}$ when this chain is reversible.

Recall that $P=(p_{ij})$ is the transition probability matrix of the chain $(x_k)_{k\geq 0}$. By Perron-Frobenius Theorem, the eigenvalues of P can be ordered in such a way that $1=\lambda_1>|\lambda_2|\geq |\lambda_3|\geq \ldots$, where $|\lambda_2|$ is called the second largest eigenvalue modulus (SLEM). The SLEM $|\lambda_2|$ is closely related to the ϵ -mixing time τ_{ϵ} . For example, it holds that

$$\frac{|\lambda_2|}{1-|\lambda_2|} \cdot K' \log(1/\epsilon) \le \tau_\epsilon \le \frac{1}{1-|\lambda_2|} \cdot K'' \log(1/\epsilon) \tag{4.9}$$

for some K' and K'' independent of ϵ [LP17, Theorem 12.4, Theorem 12.5]. In the extreme case where the x_k 's are i.i.d. with distribution π , we have $P = 1\pi^{\top}$ and hence $\lambda_2 = 0$.

The theorem below provides upper bounds on the coefficients $B^{(1)}$ and $B^{(2)}$ in the bias expansions (4.7) in terms of $|\lambda_2|$.

Theorem 4.4. Suppose that Assumptions 1, 2 and 3 hold, α satisfies equation (4.2), and the Markov chain $(x_k)_{k\geq 0}$ is reversible. It holds that

$$\mathbb{E}\left[\theta_{\infty}\right] - \theta^* = \alpha B + \alpha^2 B',\tag{4.10}$$

for some vectors $B \equiv B(A, b, P)$ and $B' \equiv B'(A, b, P, \alpha)$ that satisfy

$$||B|| \le c \cdot \frac{|\lambda_2|}{1 - |\lambda_2|} \cdot \frac{n}{\pi_{\min}} \cdot s_{\min}^{-2}(\bar{A}) \cdot b_{\max} \qquad and$$

$$(4.11)$$

$$||B'|| \le c' \cdot \frac{|\lambda_2|^2}{(1-|\lambda_2|)^2} \cdot \frac{n^{3/2}}{\pi_{\max} \pi_{\min}^2} \cdot \frac{(1+\gamma_{\max})^2}{\gamma_{\min}} \cdot s_{\min}^{-2}(\bar{A}) \cdot b_{\max}, \tag{4.12}$$

where c and c' are some universal constants.

The salient feature of Theorem 4.4 is that $||B|| \le C \frac{|\lambda_2|}{1-|\lambda_2|}$ and $||B'|| \le C' \left(\frac{|\lambda_2|}{1-|\lambda_2|}\right)^2$ for some C and C' that only depend on A, b and π . Consequently, the bias admits the bound $||\mathbb{E}\left[\theta_{\infty}\right] - \theta^*|| \le 2C \cdot \alpha \frac{|\lambda_2|}{1-|\lambda_2|}$ for sufficiently small stepsize α . In light of the relationship (4.9), we see that the bias is small when the underlying Markov chain $(x_k)_{k>0}$ mixes fast.

As a special case of Theorem 4.4, the SLEM $|\lambda_2|$ and the asymptotic bias are zero when the data $(x_k)_{k\geq 0}$ are i.i.d. This zero-bias property is implicit in the results in [LS18, Theorem 1] and [MLW⁺20, Theorem 1], which are dedicated to LSA in the i.i.d. setting. As such, Theorem 4.4 can be viewed as a generalization of these results to nonzero values of λ_2 .

In addition to the bootstrapping technique used in proving Theorem 4.3, the proof of Theorem 4.4 makes use of the eigendecomposition of a reversible transition probability matrix P. The detailed proof can be found in Appendix A.5. Nevertheless, we believe results in Theorem 4.4 can be extended to the non-reversible setting.

4.4 Implications for Averaging and Extrapolation

In this subsection, we exploit the results above to study the performance of LSA in conjunction with Polyak-Ruppert/tail averaging and Richardson-Romberg extrapolation. We focus on exploiting the convergence rate bounds in Theorem 4.1 and the bias expansion with m=1 in Theorem 4.3, namely

$$\mathbb{E}[\theta_{\infty}^{(\alpha)}] = \theta^* + \alpha B^{(1)} + O(\alpha^2).$$

Using Theorems 4.1 and 4.3, we characterize the MSE in terms of its decomposition into the optimization error, squared bias and variance, and study how these three quantities interplay with the use of constant stepsizes, averaging and extrapolation.

Our main focus is on the dependence on the stepsize α and iteration count k, and we follow the notation convention in Section 3.3. Throughout this sub-section, $B \equiv B(A,b,P)$ and $B' \equiv B'(A,b,P)$ denote vectors or matrices that are independent of α and k, and their values may change from line to line. We also use the big-O notation that hides factors independent of α and k.

²Theorem 4.4 provides explicit upper bounds on C and C'. We do not focus on optimizing these bounds in terms of their scaling with A, b, and π , and we believe tighter bounds can be obtained by more careful analysis.

4.4.1 Polyak-Ruppert Averaging. Polyak-Ruppert averaging [Rup88, PJ92] is a classical approach for reducing the variance and accelerating the convergence of stochastic approximation. Here we consider the tail-averaging variant of PR averaging [JKK+18]. Define the tail-averaged iterates

$$\bar{\theta}_{k_0,k} := \frac{1}{k - k_0} \sum_{t = k_0}^{k - 1} \theta_t, \quad k = k_0 + 1, k_0 + 2, \dots,$$

where $k_0 \ge 0$ is a user-specified burn-in period.

The following corollary provides non-asymptotic characterization for the first two moments of $\bar{\theta}_{k_0,k}$. The proof can be found in Appendix A.6.

COROLLARY 4.5. Under the setting of Theorem 4.1, the tail-averaged iterates satisfy the following bounds for all $k > k_0 \ge \tau_{\alpha}$:

$$\mathbb{E}[\bar{\theta}_{k_0,k}] - \theta^* = \alpha B + O\left(\alpha^2 + \frac{1}{\alpha(k - k_0)} \exp\left(-\frac{\alpha k_0}{4\gamma_{\text{max}}}\right)\right) \quad and \tag{4.13}$$

$$\mathbb{E}\left[\left(\bar{\theta}_{k_0,k} - \theta^*\right)\left(\bar{\theta}_{k_0,k} - \theta^*\right)^{\top}\right] = \alpha^2 B' + O\left(\alpha^3 + \frac{\sqrt{\tau_\alpha/\alpha}}{k - k_0} + \frac{1}{\alpha^2(k - k_0)^2}\exp\left(-\frac{\alpha k_0}{4\gamma_{\max}}\right)\right). \tag{4.14}$$

To parse the above results, let us fix $k_0 = k/2$ and take the trace of both sides of equation (4.14), which gives the following bound on the MSE:

$$\mathbb{E}\left[\left\|\bar{\theta}_{k/2,k} - \theta^*\right\|^2\right] = \underbrace{\alpha^2 B'' + O(\alpha^3)}_{\substack{T_1: \text{ asymptotic} \\ \text{ squared bias}}} + \underbrace{O\left(\frac{\sqrt{\tau_\alpha/\alpha}}{k}\right)}_{\substack{T_2: \text{ variance}}} + \underbrace{O\left(\frac{1}{\alpha^2 k^2} \exp\left(-\frac{\alpha k}{8\gamma_{\text{max}}}\right)\right)}_{\substack{T_3: \text{ optimization error}}}.$$
 (4.15)

The three terms on the right-hand side above have the following interpretation. The term $T_1 = \|\mathbb{E}\bar{\theta}_{\infty/2,\infty} - \theta^*\|^2 = \|\mathbb{E}\theta_{\infty} - \theta^*\|^2$ is the asymptotic squared bias, which is not affected by averaging. The term T_2 roughly corresponds to the variance $\operatorname{Var}(\bar{\theta}_{k/2,k})$, which enjoys a 1/k decay rate due to averaging. The term T_3 corresponds to the optimization error $\|\mathbb{E}\bar{\theta}_{k/2,k} - \bar{\theta}_{\infty/2,\infty}\|^2$, which decays geometrically in k thanks to using a constant stepsize α and only averaging the last k/2 iterates. Note that for large k, the squared bias (term T_1) becomes the dominating term in the MSE bound (4.15).

We remark in passing that by setting $k_0 = k - 1$ in Corollary 4.5 and observing that $\bar{\theta}_{k-1,k} = \theta_{k-1}$ (also see Remark 2), we obtain the following characterization for the MSE of the raw LSA iterate θ_k :

$$\mathbb{E}\left[\|\theta_k - \theta^*\|^2\right] = \alpha^2 B'' + O\left(\alpha \tau_\alpha\right) + O\left(e^{-\alpha k/(8\gamma_{\text{max}})}\right).$$

This result is consistent with existing MSE upper bounds in [SY19, BRS21]. The power of our result lies in that it is an equality (up to higher order terms) rather than merely an upper bound, and that it decouples the contributions from the squared bias $\alpha^2 B''$ and variance. This decoupling is crucial in understanding the effect of tail-averaging (in Corollary 4.5) and RR extrapolation (in Corollary 4.6 to follow).

4.4.2 Richardson-Romberg Extrapolation. We next show that one can use the RR extrapolation technique [SB02] to reduce the bias to a higher order term of α . Let $\bar{\theta}_{k_0,k}^{(\alpha)}$ and $\bar{\theta}_{k_0,k}^{(2\alpha)}$ denote the tail-averaged iterates computed using two stepsizes α and 2α using the same data stream $(x_k)_{k\geq 0}$. The corresponding RR extrapolated iterates are given by

$$\widetilde{\theta}_{k_0,k}^{(\alpha)} = 2\bar{\theta}_{k_0,k}^{(\alpha)} - \bar{\theta}_{k_0,k}^{(2\alpha)}.$$

With $k_0, k \to \infty$, Theorems 4.1 and 4.3 ensure that $\widetilde{\theta}_{k_0,k}^{(\alpha)}$ converges to $2\theta_{\infty}^{(\alpha)} - \theta_{\infty}^{(2\alpha)}$, which has bias

$$2\big(\mathbb{E}\theta_{\infty}^{(\alpha)}-\theta^*\big)-\big(\mathbb{E}\theta_{\infty}^{(2\alpha)}-\theta^*\big)=2\big(\alpha B^{(1)}+O(\alpha^2)\big)-\big(2\alpha B^{(1)}+O(4\alpha^2)\big)=O(\alpha^2).$$

We see that the extrapolation cancels out the first-order term of α , hence reducing the bias by a factor of α .

The following corollary formalizes the above argument and provides non-asymptotic characterization for the first two moments of $\widetilde{\theta}_{k_0,k}^{(\alpha)}$. The proof can be found in Appendix A.7.

COROLLARY 4.6. Under the setting of Theorem 4.1, the RR extrapolated iterates with stepsizes α and 2α satisfy the following bounds for all $k > k_0 \ge \tau_{\alpha}$:

$$\mathbb{E}\left[\widetilde{\theta}_{k_0,k}^{(\alpha)}\right] - \theta^* = O(\alpha^2) + O\left(\frac{1}{\alpha(k - k_0)} \exp\left(-\frac{\alpha k_0}{4\gamma_{\max}}\right)\right)$$

and

$$\mathbb{E}\left[\left(\widetilde{\theta}_{k_{0},k}^{(\alpha)} - \theta^{*}\right)\left(\widetilde{\theta}_{k_{0},k}^{(\alpha)} - \theta^{*}\right)^{\top}\right] = \underbrace{O\left(\alpha^{4}\right)}_{\text{asymptotic squared bias}} + \underbrace{O\left(\frac{\sqrt{\tau_{\alpha}/\alpha}}{k - k_{0}}\right)}_{\text{variance}} + \underbrace{O\left(\frac{1}{\alpha^{2}(k - k_{0})^{2}} \exp\left(-\frac{\alpha k_{0}}{4\gamma_{\text{max}}}\right)\right)}_{\text{optimization error}}.$$
(4.16)

Comparing the bound (4.16) with (4.14), we see that RR extrapolation reduces the squared bias by a factor of α^2 while retaining the 1/k and $\exp(-k)$ convergence rates for the variance and optimization error, respectively.

Thanks to the infinite series expansion in Theorem 4.3, RR extrapolation can in fact be applied to more than two stepsizes to further reduce the bias. Let $\mathcal{A} = \{\alpha_1, \alpha_2, \ldots, \alpha_m\}$ be a set of $m \geq 2$ distinct stepsizes and $\alpha = \max_{1 \leq i \leq m} \alpha_i$. Let $(h_1, h_2, \ldots, h_m) \in \mathbb{R}^m$ be the solution to the following linear equation system:

$$\sum_{i=1}^{m} h_i = 1; \qquad \sum_{i=1}^{m} h_i \alpha_i^t = 0, \quad t = 1, 2, \dots, m - 1.$$
 (4.17)

Note that the solution is unique since the coefficient matrix of the system is a Vandermonde matrix. Then, RR extrapolated iterates with stepsizes in \mathcal{A} and the burn-in period k_0 is given by

$$\widetilde{\theta}_{k_0,k}^{\mathcal{A}} = \sum_{i=1}^{m} h_i \cdot \bar{\theta}_{k_0,k}^{(\alpha_i)}. \tag{4.18}$$

This procedure eliminates the first m-1 terms in the bias expansion (4.8), thereby reducing the bias to

$$\mathbb{E}\left[\widetilde{\theta}_{k_0,\infty}^{\mathcal{A}}\right] - \theta^* = \sum_{i=1}^m h_i \cdot \left(\mathbb{E}\left[\theta_{\infty}^{(\alpha_i)}\right] - \theta^*\right) = O(\alpha^m).$$

It is possible to derive non-asymptotic bounds similar to Corollary 4.6 — we omit the details. In Section 5, we numerically verify the efficacy of this high-order RR extrapolation approach.

4.5 Implications for TD Learning

TD(0) is an iterative algorithm in RL for evaluating a given policy for a Markov Decision Process (MDP), or equivalently for computing the value function of a Markov Reward Process (MRP) [Ber19, SB18]. Potentially equipped with function approximation, TD(0) is a special case of LSA. Consequently, all the results in the previous sub-sections can be specialized to TD(0), as we show below.

Consider an MRP (S, P^S, r, γ) , where $S = \{1, \dots, n^S\}$ is a finite state space, $P^S = (p_{ij}^S) \in [0, 1]^{n^S \times n^S}$ is the transition probability matrix, $r : S \to [-r_{\max}, r_{\max}]$ is a deterministic reward function, and $y \in [0, 1)$ is the discount factor. We assume that P^S is irreducible and aperiodic with a unique stationary distribution π^S . The value function $V : S \to \mathbb{R}$ of this MRP is defined as $V(s) = \mathbb{E}\left[\sum_{t=0}^{\infty} \gamma^t r(s_t)|s_0 = s\right]$, where $(s_k)_{k\geq 0}$ is a Markov chain on S with P^S as the transition matrix. When the state space S is large, it is common to assume that the value function can be approximated by a linear function as $V(s) \approx \phi(s)^{\top}\theta$, where $\phi(s) \in \mathbb{R}^d$ denotes a known feature vector for state s and θ is an unknown weight vector. If we view V as a vector in \mathbb{R}^{n^S} and let $\Phi \in \mathbb{R}^{n^S \times d}$ be the feature matrix with $\phi(s)^{\top}$ as the s-th row, then the above model can be written compactly as $V \approx \Phi\theta$. As is standard [BRS21, SY19], we assume that $n^S \geq d$, the matrix Φ has full column rank, and the feature matrix is re-scaled such that $\phi_{\max} := \sup_{s \in S} \|\phi(s)\| \leq \frac{1}{\sqrt{1+V}}$.

Given a single Markovian data stream $(s_k)_{k\geq 0}$, the TD(0) algorithm with linear function approximation is defined by the update rule

$$\theta_{k+1} = \theta_k + \alpha \left[r(s_k) + \gamma \phi(s_{k+1})^\top \theta_k - \phi(s_k)^\top \theta_k \right] \phi(s_k). \tag{4.19}$$

We assume that the chain $(s_k)_{k\geq 0}$ is initialized at $s_0 \sim \pi^S$. TD(0) aims to approximately solve the so-called projected Bellman equation, whose solution θ^* is given by $\theta^* = -(\Phi^T D(\gamma P - I)\Phi)^{-1}\Phi^T Dr$, where $D = \operatorname{diag}(\pi^S) \in [0, 1]^{n^S \times n^S}$ is a diagonal matrix and $r \in \mathbb{R}^{n^S}$ is the reward function viewed as a vector.

Remark 1. A special case of the above setting is when $d = n^S$ and the feature vectors are one-hot, i.e., $\phi(s)$ is the s-th standard basis vector in \mathbb{R}^{n^S} . In this case, the update (4.19) becomes the (asynchronous) tabular TD(0) algorithm, and the target solution θ^* is the value function V itself.

It is easy to see that the TD(0) update (4.19) is a special case of the LSA update (3.1) with

$$x_k = (s_k, s_{k+1}), \quad A(x_k) = \phi(s_k) (\gamma \phi(s_{k+1}) - \phi(s_k))^\top, \quad b(x_k) = r(s_k) \phi(s_k),$$

and $X = \{(i, j) \in S \times S : p_{ij}^S > 0\}$. Below we verify that TD(0) satisfies the required assumptions.

- Assumption 1: The irreducibility and aperiodicity of the chain $(s_k)_{k\geq 0}$ imply that the augmented chain $(x_k)_{k\geq 0}=((s_k,s_{k+1}))_{k\geq 0}$ also satisfies the same properties. Moreover, by assumption the chain $(s_k)_{k\geq 0}$ starts in the steady state, hence so does the chain $(x_k)_{k\geq 0}$.
- Assumption 2: We have

$$A_{\max} = \sup_{i,j \in \mathcal{S}} \|\phi(i)(\gamma\phi(j) - \phi(i))^{\top}\| \le (1+\gamma)\phi_{\max}^2 \quad \text{and} \quad b_{\max} = \sup_{i \in \mathcal{S}} \|r(i)\phi(i)\| \le r_{\max}\phi_{\max}.$$

Therefore, Assumption 2 holds since the features are assumed normalized as $\phi_{\text{max}} \leq \frac{1}{\sqrt{1+\nu}}$.

• Assumption 3: The unique stationary distribution π of the Markov chain $(x_k)_{k\geq 0}$ is given by

$$\pi_x = \pi_i^{\mathcal{S}} p_{ij}^{\mathcal{S}}, \quad \text{for } x = (i, j) \in \mathcal{X}$$

$$\tag{4.20}$$

By direct calculation we have $\bar{A} = \sum_{x \in \mathcal{X}} \pi_x A(x) = \Phi^{\top} D(\gamma P - I) \Phi$. It is known that the matrix \bar{A} is negative definite (i.e., $x^{\top} \bar{A} x < 0$, $\forall x \neq 0$) and hence Hurwitz [TVR97].

Consequently, all the results in Sections 4.1–4.4 apply to TD(0) with linear function approximation, Markovian data, and constant stepsizes.

We emphasize that the above results hold for TD(0) driven by a single Markovian data stream, where nonzero asymptotic bias generally exists. In comparison, many existing non-asymptotic

results on TD(0) consider the i.i.d. data setting; see, e.g., the work in [DSTM18, BRS21, KPR⁺21, DMN⁺21]. TD(0) under this setting corresponds to the update

$$\theta_{k+1} = \theta_k + \alpha \left[r(s_k) + \gamma \phi(s_k^{\text{next}})^\top \theta_k - \phi(s_k)^\top \theta_k \right] \phi(s_k),$$

where the data $x_k = (s_k, s_k^{\text{next}})$ is independent across k and has the distribution $\mathbb{P}(s_k = i, s_k^{\text{next}} = j) = \pi_i^S p_{ij}^S$; equivalently, we have $x_k \stackrel{\text{i.i.d.}}{\sim} \pi$, where π is given in equation (4.20). In this setting, Theorem 4.4 implies that TD(0) with a constant stepsize has no asymptotic bias, i.e., $\mathbb{E}[\theta_{\infty}] = \theta^*$.

We note in passing that the $TD(\lambda)$ algorithm [Sut88, TVR97], a generalization of TD(0), can also be cast as an LSA procedure; see, e.g., the work in [SY19, MPWB21] for a formal reduction. Our LSA results can be used to characterize $TD(\lambda)$ as well. We omit the details.

5 NUMERICAL EXPERIMENTS

In this section, we provide numerical experimental results for both the LSA iteration (3.1) and the TD(0) iteration (4.19) with linear function approximation.

5.1 Experiments for LSA

We consider the LSA update (3.1) in dimension d=4 with n=8 states. We generate transition probability P and the functions A and b randomly; see Appendix C.1 for the details. Given P, we generate a single trajectory of the Markov chain $(x_k)_{k=1}^K$ of length $K=10^8$, and run the LSA iteration with initialization $\theta_0^{(\alpha)}=0$ and stepsizes $\alpha \in \{0.2,0.4,0.8\}$.

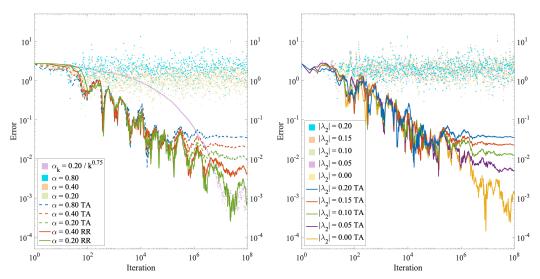
In Figure 2(a), we plot the error $\|\theta_k^{(\alpha)} - \theta^*\|$ for the raw LSA iterates $\theta_k^{(\alpha)}$, the error for the tail-averaged (TA) iterates $\bar{\theta}_{k/2,k}^{(\alpha)}$, and the error for the RR extrapolated iterates $\bar{\theta}_k^{(\alpha)}$ with stepsizes α and 2α . For comparison, we also include the errors for LSA with a diminishing stepsize $\alpha_k = 0.2/k^{0.75}$.

We see that with constant stepsizes, the raw LSA iterates oscillate, consistent with the distributional convergence result in Theorem 4.1, whereas the tail averaged iterates converge to a limit. We also observe that the converged error, which corresponds to the asymptotic bias, is smaller when using a smaller stepsize. In particular, this error can be seen to be proportional to the stepsize (note the equal spacing in the log scale between the three TA lines), as predicted by Theorem 4.3. Finally, RR extrapolation with two stepsizes further reduces the bias, as can be seen by comparing, e.g., the dashed red line (TA with $\alpha=0.4$) and the solid red line (RR with $\alpha=0.4$ and 0.8). This observation is consistent with Corollary 4.6. We also observe that the iterates using a diminishing stepsize converge with a vanishing error, though its initial convergence is substantially slower than that of the tail-averaged iterates with constant stepsizes. This agrees with the theoretical prediction of the sublinear convergence rate of the former and the geometric convergence rate of the latter.

We next investigate the relationship between the error and the mixing time. Given the transition probability matrix P generated above and its stationary distribution π , we construct another transition probability matrix parameterized by $\beta \in [0, 1]$ as follows:

$$P^{(\beta)} = \beta \cdot P + (1 - \beta) \cdot 1\pi^{\top}.$$

Note that $P^{(1)}=P$, and that $P^{(\beta)}$ has the same stationary distribution π as P for any β . As β decreases from 1 to 0, the SLEM $|\lambda_2|$ of $P^{(\beta)}$ decreases and approaches 0, and hence the mixing time decreases as well. For different values of β , we run the LSA iteration (3.1) with $P^{(\beta)}$ as the transition probability matrix of the underlying Markov chain $(x_k)_{k\geq 0}$. In Figure 2(b), we plot the corresponding errors of the tail-averaged iterates. We see that a smaller $|\lambda_2|$ leads to a smaller final error. Moreover, when $\lambda_2=0$, which corresponds to the i.i.d. data setting, it can be observed that the error is converging to zero, which indicates a zero asymptotic bias. These observations are consistent with Theorem 4.4 on the relationship between the asymptotic bias and mixing time.



- (a) The errors of the raw LSA iterates, tail-averaged (TA) (b) The errors of the raw LSA iterates and tail-averaged (TA) α.
- iterates and RR extrapolated iterates with different stepsizes iterates under different SLEM $|\lambda_2|$. The stepsize α is fixed at

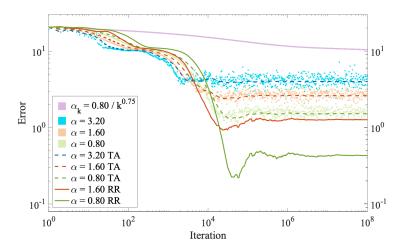


Fig. 2. Experiment results for LSA

Fig. 3. The errors of the raw TD(0), tail-averaged (TA) and RR extrapolated iterates with different stepsizes α .

Experiments for TD(0) with Linear Function Approximation

We perform a similar set of experiments as in the previous sub-section on the TD(0) algorithm. In particular, we consider the "Problematic MDP" from the classical work [KP00, LP03], and use TD(0) with linear function approximation to estimate the value function of a given policy. See Appendix C.2 for the details of the MDP, the policy, and the choice of the feature vectors.

In Figure 3, we plot the errors of the raw TD(0) iterates, tail-averaged iterates and RR extrapolated iterates with different stepsizes α . The results are qualitatively similar to those in Figure 2(a). In

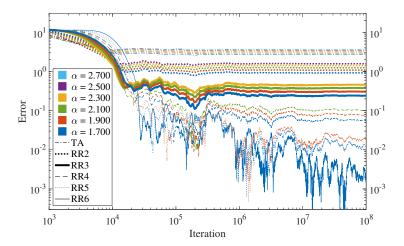


Fig. 4. Comparison between tail-averaging (TA) and RR extrapolation with m stepsizes, for $m=2,\ldots,6$. The setting for each line in the plot is given by its line style (representing the number of stepsizes used in RR) and line color (representing the smallest stepsize involved). For example, the dash-dotted green line corresponds to TA with stepsize $\alpha=2.1$, and the dashed red line corresponds to RR with four stepsizes $\alpha\in\{1.9,\,2.1,\,2.3,\,2.5\}$.

addition, we observe that the TA iterates with a larger stepsize have faster initial convergence. This observation is consistent with the predicted relationship between the optimization error and stepsize in Corollary 4.5.

Lastly, we investigate the benefit of RR extrapolation with more than 2 stepsizes, using the scheme outlined in equations (4.17) and (4.18). Specifically, we compare the errors of the tail-averaged iterates and the RR extrapolated iterates with $m \in \{2, 3, ..., 6\}$ stepsizes. The results are shown in Figure 4. Note that we use 6 large stepsizes (of similar magnitudes), resulting in fast initial convergence. We see that increasing the number of stepsizes in RR extrapolation further reduces the final errors by a significant margin. In particular, the error of RR extrapolation with 6 stepsizes is smaller by 3 orders of magnitude when compared to TA with the same stepsizes. We emphasize that this error reduction is obtained almost for free, as we can run the six TD(0) iterations in parallel using the same data stream.

6 PROOF OUTLINE

In this section, we outline the proofs for Theorem 4.1 (convergence of LSA) and Theorem 4.3 (bias expansion). The proofs make use of a pilot result Proposition 6.1, stated in Section 6.1, which serves as the first step of subsequent analysis. The complete proofs of these results and other main theorems/corollaries are given in the appendix.

6.1 A Pilot Result

We have the following non-asymptotic upper bound on the MSE $\mathbb{E}[\|\theta_k - \theta^*\|^2]$.

Proposition 6.1. Under Assumptions 1, 2 and 3, if α satisfies equation (4.2), then the following bound holds for all $k \geq \tau$,

$$\mathbb{E}[\|\theta_k - \theta^*\|^2] \leq 8 \frac{\gamma_{\max}}{\gamma_{\min}} \left(1 - \frac{0.9\alpha}{\gamma_{\max}}\right)^k \left(\mathbb{E}[\|\theta_0 - \theta^*\|^2] + s_{\min}^{-2}(\bar{A})b_{\max}^2\right) + \frac{\gamma_{\max}}{0.9\gamma_{\min}} \cdot \alpha\tau \cdot \kappa,$$

with κ defined in equation (4.3).

Proposition 6.1 is a moderate improvement of [SY19, Theorem 7]. When $b_{\max}=0$ (which means $b(x)=0, \forall x\in\mathcal{X}$), Proposition 6.1 guarantees that $\kappa=0$, in which case θ_k converges in mean squared to θ^* as $k\to\infty$. This fact plays an important role in proving our main Theorem 4.1 in the setting with a general b and nonzero b_{\max} . In particular, the proof of Theorem 4.1 makes use of a coupling argument that constructs another process with $b_{\max}=0$. In comparison, the bound in [SY19, Theorem 7] gives a non-zero value of κ even when $b_{\max}=0$ and hence is insufficient for executing the coupling argument. Moreover, the stepsize condition (4.2) required by Proposition 6.1 (and by all our other results) does not involve b_{\max} , which correctly reflects the translation invariance of the LSA update (3.1). The stepsize condition in [SY19, Theorem 7], on the other hand, has a superfluous dependence on b_{\max} .

The proof of Proposition 6.1 is similar to that of [SY19, Theorem 7] with a more refined analysis. For completeness, we provide the proof in Appendix A.1. One key refinement in our proof is to avoid invoking inequalities of the form $2u \le 1 + u^2$, and to use instead $2u \le \beta^2 u^2 + 1/\beta^2$ with a judicious choice of β that respects the translation invariance of the LSA update (3.1).

6.2 Proof Outline of Theorem 4.1

In this sub-section, we sketch the main ideas in the proof of Theorem 4.1. The complete proof can be found in Appendix A.2.

The proof consists of bounding Wasserstein distances of the form $\bar{W}_2(\mathcal{L}(x_k,\theta_k),\mathcal{L}(x_{k+1},\theta_{k+1}))$ and $\bar{W}_2(\mathcal{L}(x_t,\theta_t),\mathcal{L}(x_\infty,\theta_\infty))$. Since the Wasserstein distance is defined by the optimal coupling, it can be upper bounded by constructing a particular coupling. With this strategy in mind, we consider coupling two Markov chains $(x_k^{[1]},\theta_k^{[1]})_{k\geq 0}$ and $(x_k^{[2]},\theta_k^{[2]})_{k\geq 0}$, which are two copies of LSA iteration (3.1). We make use of two types of coupling in the proof.

The first type of coupling is constructed by letting the two Markov chains above share the same underlying data stream $(x_k)_{k\geq 0}$, i.e., letting $x_k^{[1]}=x_k^{[2]}=x_k$ for all $k\geq 0$. Explicitly, the iterates $\theta_{k+1}^{[1]}$ and $\theta_{k+2}^{[2]}$ are given by the update

$$\theta_{k+1}^{[1]} = \theta_k^{[1]} + \alpha (A(x_k)\theta_k^{[1]} + b(x_k)),
\theta_{k+1}^{[2]} = \theta_k^{[2]} + \alpha (A(x_k)\theta_k^{[2]} + b(x_k)),$$

$$k = 0, 1, ...$$

Taking the difference of the two equations above, we see that the difference $\omega_k := \theta_k^{[1]} - \theta_k^{[2]}$ satisfies the following recursion

$$\omega_{k+1} = (I + \alpha A(x_k)) \cdot \omega_k, \qquad k = 0, 1, \dots$$

Our key observation is that the above recursion is a special case of the LSA iteration (3.1) with ω_k as the variable and $b_{\max} = \sup_{x \in \mathcal{X}} \|b(x)\| = 0$. Consequently, the pilot result in Proposition 6.1 can be invoked to obtain the following geometric convergence bound for ω_k :

$$\mathbb{E}[\|\omega_k\|^2] \le C(A, b, \pi) \left(1 - \frac{0.9\alpha}{\gamma_{\text{max}}}\right)^k \mathbb{E}[\|\omega_0\|^2].$$

We then judiciously choose the conditional distribution of $\theta_0^{[2]}$ given $(x_k, \theta_0^{[1]}, \theta_0^{[2]})$ such that $(x_k, \theta_k^{[2]}) \stackrel{\mathrm{d}}{=} (x_{k+1}, \theta_{k+1}^{[1]})$ for all $k \geq 0$, where $\stackrel{\mathrm{d}}{=}$ denotes equality in distribution. It follows from the above geometric convergence bound that

$$\bar{W}_{2}^{2}(\mathcal{L}(x_{k},\theta_{k}^{[1]}),\mathcal{L}(x_{k+1},\theta_{k+1}^{[1]})) \leq \mathbb{E}[\|\theta_{k}^{[1]} - \theta_{k}^{[2]}\|^{2}] \to 0 \quad \text{as } k \to \infty.$$

As such, $(x_k, \theta_k^{[1]})_{k \geq 0}$ is a Cauchy sequence and hence converges to a unique limit $(x_\infty, \theta_\infty)$ with the limiting distribution $\bar{\mu} := \mathcal{L}((x_\infty, \theta_\infty))$. This proves Part 1 of Theorem 4.1.

We next show that $\bar{\mu}$ is the invariant distribution of the Markov chain $(x_k, \theta_k)_{k \ge 0}$. This invariance property would follow easily if one could establish the one-step contraction property

$$\bar{W}_{2}\left(\mathcal{L}\left(x_{1}^{[1]}, \theta_{1}^{[1]}\right), \mathcal{L}\left(x_{1}^{[2]}, \theta_{1}^{[2]}\right)\right) \leq \rho \cdot \bar{W}_{2}\left(\mathcal{L}\left(x_{0}^{[1]}, \theta_{0}^{[1]}\right), \mathcal{L}\left(x_{0}^{[2]}, \theta_{0}^{[2]}\right)\right)$$

for some $\rho \in [0, 1)$. In fact, this is the approach taken in [DDB20] for analyzing SGD under i.i.d. noise. For our Markovian data setting, however, establishing one-step contraction is challenging if not impossible. Thankfully, to prove invariance of $\bar{\mu}$, it suffices to have the following weaker property

$$\bar{W}_{2}^{2}\left(\mathcal{L}(x_{1}^{[1]}, \theta_{1}^{[1]}), \mathcal{L}(x_{1}^{[2]}, \theta_{1}^{[2]})\right) \\
\leq \rho_{1} \cdot \bar{W}_{2}^{2}\left(\mathcal{L}(x_{0}^{[1]}, \theta_{0}^{[1]}), \mathcal{L}(x_{0}^{[2]}, \theta_{0}^{[2]})\right) + \sqrt{\rho_{2} \cdot \bar{W}_{2}^{2}\left(\mathcal{L}(x_{0}^{[1]}, \theta_{0}^{[1]}), \mathcal{L}(x_{0}^{[2]}, \theta_{0}^{[2]})\right)}, \tag{6.1}$$

where $\mathcal{L}(x_0^{[1]}, \theta_0^{[1]}) = \bar{\mu}$ and the quantities ρ_1 and ρ_2 are finite and independent of $\mathcal{L}(x_0^{[2]}, \theta_0^{[2]})$. We establish the property (6.1) by using a second type of coupling between $(x_k^{[1]}, \theta_k^{[1]})_{k\geq 0}$ and $(x_k^{[2]}, \theta_k^{[2]})_{k\geq 0}$, such that

$$x_{k+1}^{[1]} = x_{k+1}^{[2]}$$
 if $x_k^{[1]} = x_k^{[2]}$, $\forall k \ge 0$.

That is, the two underlying Markov chains $(x_k^{[1]})_{k\geq 0}$ and $(x_k^{[2]})_{k\geq 0}$ evolve separately until they reach the same state, after which they coalesce and follow the same trajectory. Given the property (6.1), for any $k\geq 0$, if we set $\mathcal{L}(x_0^{[1]},\theta_0^{[1]})=\mathcal{L}(x_0,\theta_0)=\bar{\mu}$ and $\mathcal{L}(x_0^{[2]},\theta_0^{[2]})=\mathcal{L}(x_k,\theta_k)$, then

$$\bar{W}_{2}^{2}(\mathcal{L}(x_{1},\theta_{1}),\mathcal{L}(x_{k+1},\theta_{k+1})) \leq \rho_{1} \cdot \bar{W}_{2}^{2}(\bar{\mu},\mathcal{L}(x_{k},\theta_{k})) + \sqrt{\rho_{2} \cdot \bar{W}_{2}^{2}(\bar{\mu},\mathcal{L}(x_{k},\theta_{k}))}.$$

It follows from the triangle inequality of Wasserstein distance that

$$\bar{W}_{2}(\mathcal{L}(x_{1},\theta_{1}),\bar{\mu}) \leq \bar{W}_{2}(\mathcal{L}(x_{1},\theta_{1}),\mathcal{L}(x_{k+1},\theta_{k+1})) + W_{2}(\mathcal{L}(x_{k+1},\theta_{k+1}),\bar{\mu})
\leq \sqrt{\rho_{1} \cdot \bar{W}_{2}^{2}(\bar{\mu},\mathcal{L}(x_{k},\theta_{k})) + \sqrt{\rho_{2} \cdot \bar{W}_{2}^{2}(\bar{\mu},\mathcal{L}(x_{k},\theta_{k}))} + W_{2}(\mathcal{L}(x_{k+1},\theta_{k+1}),\bar{\mu})
\longrightarrow 0 \quad \text{as } k \to \infty,$$

which establishes the invariance of $\bar{\mu}$ and proves Part 2 of Theorem 4.1.

Finally, the non-asymptotic bound in Part 3 of Theorem 4.1 follows from the non-asymptotic bound on ω_k and invariance property of $\bar{\mu}$ established above.

6.3 Proof Outline of Theorem 4.3

In this sub-section, we outline the proof of Theorem 4.3. The complete proof can be found in Appendix A.4.

As discussed after Theorem 4.3, our proof centers around the condition expectations $\mathbb{E}\left[\theta_{\infty}|x_{\infty}=i\right]$, $i \in \mathcal{X}$. To characterize these quantities, we make use of the Basic Adjoint Relationship

$$\mathbb{E}\left[f(x_{\infty},\theta_{\infty})\right] = \mathbb{E}\left[f(x_{\infty+1},\theta_{\infty+1})\right],$$

where the test function f is chosen to be $f^{(i)}(x, \theta) := \theta \cdot \mathbb{1}\{x = i\}$ for each $i \in X$. This choice allows us to establish the following recursive relationship:

$$\mathbb{E}\left[\theta_{\infty}|x_{\infty}=i\right] = \left(\sum_{s \in \mathcal{X}} \hat{p}_{is} \mathbb{E}\left[\theta_{\infty}|x_{\infty}=s\right]\right) + \alpha \left(\sum_{s \in \mathcal{X}} \hat{p}_{is} \left(A(s) \mathbb{E}\left[\theta_{\infty}|x_{\infty}=s\right] + b(s)\right)\right), \quad \forall i \in \mathcal{X}. \quad (6.2)$$

By choosing state $1 \in X$ as the reference state and defining $\delta_i := \mathbb{E} \left[\theta_{\infty} | x_{\infty} = i \right] - \mathbb{E} \left[\theta_{\infty} | x_{\infty} = 1 \right]$, we obtain from (6.2) that

$$\delta_{i} - \left(\sum_{s \in \mathcal{X}} \hat{p}_{is} \delta_{s}\right) = \alpha \left(\sum_{s \in \mathcal{X}} \hat{p}_{is}(A(s) \mathbb{E}\left[\theta_{\infty} | x_{\infty} = s\right] + b(s))\right), \quad \forall i \in \mathcal{X}.$$

$$(6.3)$$

Equation (6.3), together with the bound $\mathbb{E}\left[\theta_{\infty}\right] = O(1)$ (which follows from Theorem 4.1), imply the coarse bound $\vec{\delta} = \begin{bmatrix} \delta_1^\top \delta_2^\top & \cdots & \delta_n^\top \end{bmatrix}^\top = O(\alpha)$.

Moreover, owing to the construction of δ_i , we are able to obtain

$$\mathbb{E}\left[\theta_{\infty}|x_{\infty}=1\right] = \theta^* - \bar{A}^{-1} \sum_{s \in \mathcal{X}} \pi_s A(s) \delta_s. \tag{6.4}$$

Substituting (6.4) into (6.3) and simplifying the expression, we are able to establish the following self-expressing equation for δ :

$$\vec{\delta} = \alpha \Upsilon_1 + \alpha \Xi \vec{\delta},\tag{6.5}$$

where the vector Υ_1 and matrix Ξ are independent of α . With the expression (6.5), we can bootstrap from the coarse bound $\vec{\delta} = O(\alpha)$ to obtain

$$\vec{\delta} = \alpha \Upsilon_1 + \Xi O(\alpha^2).$$

Continuing this bootstrapping argument for *m* steps, we derive the expansion

$$\vec{\delta} = \sum_{i=1}^{m} \alpha^{i} \Upsilon_{i} + \Xi^{m} O(\alpha^{m+1}). \tag{6.6}$$

Lastly, we convert the expansion (6.6) for $\vec{\delta}$ into an expansion for $\mathbb{E}[\theta_{\infty}]$. To this end, we note that the expression (6.4) implies that

$$\mathbb{E}[\theta_{\infty}] = \sum_{s \in \mathcal{X}} \pi_s (\mathbb{E}[\theta_{\infty} | x_{\infty} = 1] + \delta_s) = \theta^* + \sum_{s \in \mathcal{X}} \pi_s (-\bar{A}^{-1}A(s) + I) \delta_s,$$

Combining the above equation with the expansion (6.6) proves the desired expansion for $\mathbb{E}[\theta_{\infty}]$ given in Theorem 4.3:

$$\mathbb{E}[\theta_{\infty}] = \theta^* + \sum_{i=1}^{m} \alpha^i B^{(i)} + O(\alpha^{m+1}). \tag{6.7}$$

Moreover, if the stepsize α is sufficiently small to ensure that the matrix $\alpha\Xi$ in equation (6.5) has a spectral radius less than one, then one can take $m \to \infty$ in the expansions (6.6) and (6.7).

CONCLUSION

In this paper, we study linear stochastic approximation with constant stepsizes and Markovian data. We analyze the convergence rates to a limiting distribution and identify the existence of asymptotic bias. We characterize the bias as a function of the stepsize and mixing time, and rigorously establish the benefit of Richardson-Romberg extrapolation. Our results provide a refined characterization of linear stochastic approximation, identifying the effect of stepsize, averaging, and extrapolation on the optimization error, variance, and bias.

Based on our work, immediate next steps include tightening the dimension dependence in our bounds and relaxing the reversibility assumption. Below we remark on two possible further extensions and the associated challenges.

Infinite State Space. Our current results hold when the underlying Markov chain $(x_k)_{k\geq 0}$ is supported on a finite state space. Extension to (countably or uncountably) infinite state spaces is of immediate interest. Such an extension is straightforward for the convergence guarantees in Theorem 4.1, which can be shown to hold under the same assumptions as in [SY19], namely, geometric mixing of the underlying Markov chain. Generalizing the bias characterization in Theorem 4.3 is more non-trivial. Recall that a key step in our proof involves the linear equation (6.5). When the state space is infinite, this equation becomes an infinite-dimensional one, and it requires additional technical work to establish the solvability of the equation.

Nonlinear Stochastic Approximation. It is interesting to see if our results can be extended to more general stochastic approximation procedures of the form

$$\theta_{k+1} = \theta_k + \alpha (f(\theta_k, x_k) - \theta_k),$$

where f is a nonlinear function of θ_k satisfying certain contractive properties [CMSS20, CMSS21b, CBD22]. Contractive SA covers many important algorithms in RL, such as the TD algorithm with neural networks and the Q-learning algorithm. Under appropriate conditions, we expect that similar weak convergence and bias characterization results hold for contractive SA, where the convergence rate and bias may depend on the contraction and smoothness parameters. In particular, it may be possible to extend Theorem 4.1 to contractive SA by leveraging the Lyapunov theory and MSE bounds in [CMSS21b]. Extending Theorem 4.3 may involve manipulating a nonlinear Basic Adjoint Relationship, for which certain linearization argument may be useful.

Besides the aforementioned, other interesting future directions include: (a) study higher moments of the LSA iterates and provide high probability bounds; (b) exploit our results to formulate an algorithm that guides the choice and scheduling of the stepsize.

Acknowledgements

Y. Chen is supported in part by NSF CAREER Award CCF-2047910 and grant CCF-1704828. Q. Xie is supported in part by NSF grant CNS-1955997 and a J.P. Morgan Faculty Research Award. We would like to thank Jim Dai for inspiring discussion.

REFERENCES

- [BCD⁺21] Vivek S. Borkar, Shuhang Chen, Adithya Devraj, Ioannis Kontoyiannis, and Sean P. Meyn. The ODE method for asymptotic statistics in stochastic approximation and reinforcement learning, 2021.
 - [Ber19] Dimitri P. Bertsekas. Reinforcement learning and Optimal Control. Athena Scientific, Belmont, Massachusetts, USA, 2019.
 - [Blu54] Julius R. Blum. Approximation methods which converge with probability one. *The Annals of Mathematical Statistics*, 25(2):382 386, 1954.
 - [BM00] Vivek S. Borkar and Sean P. Meyn. The O.D.E. method for convergence of stochastic approximation and reinforcement learning. SIAM Journal on Control and Optimization, 38(2):447–469, Jan 2000.
- [BMP12] Albert Benveniste, Michel Metivier, and Pierre Priouret. Adaptive Algorithms and Stochastic Approximations. Springer Berlin Heidelberg, 1st edition, 2012.
- [Bor08] Vivek S. Borkar. Stochastic Approximation: A Dynamical Systems Viewpoint. Hindustan Book Agency Gurgaon, 2008.
- [Bré20] Pierre Brémaud. Markov Chains: Gibbs Fields, Monte Carlo Simulation, and Queues. Springer International Publishing, Cham, Switzerland, 2nd edition, 2020.
- [BRS21] Jalaj Bhandari, Daniel Russo, and Raghav Singal. A finite time analysis of temporal difference learning with linear function approximation. *Operations Research*, 69(3):950–973, May 2021.
- [CBD22] Siddharth Chandak, Vivek S. Borkar, and Parth Dodhia. Concentration of contractive stochastic approximation and reinforcement learning. *Stochastic Systems*, Jul 2022.
- [CDLS99] Robert G. Cowell, A. Philip Dawid, Steffen L. Lauritzen, and David J. Spiegelhalter. Probabilistic Networks and Expert Systems: Exact Computational Methods for Bayesian Networks. Information Science and Statistics. Springer Publishing Company, Incorporated, New York, NY, USA, 1st edition, 1999.
- [CMM22] Zaiwei Chen, Shancong Mou, and Siva Theja Maguluri. Stationary behavior of constant stepsize sgd type algorithms: An asymptotic characterization. Proceedings of the ACM on Measurement and Analysis of Computing Systems, 6(1), 02 2022.

- [CMSS20] Zaiwei Chen, Siva Theja Maguluri, Sanjay Shakkottai, and Karthikeyan Shanmugam. Finite-sample analysis of contractive stochastic approximation using smooth convex envelopes. In H. Larochelle, M. Ranzato, R. Hadsell, M.F. Balcan, and H. Lin, editors, Advances in Neural Information Processing Systems, volume 33, pages 8223–8234. Curran Associates, Inc., 2020.
- [CMSS21a] Zaiwei Chen, Siva Theja Maguluri, Sanjay Shakkottai, and Karthikeyan Shanmugam. Finite-sample analysis of off-policy TD-learning via generalized Bellman operators. In M. Ranzato, A. Beygelzimer, Y. Dauphin, P.S. Liang, and J. Wortman Vaughan, editors, Advances in Neural Information Processing Systems, volume 34, pages 21440–21452. Curran Associates, Inc., 2021.
- [CMSS21b] Zaiwei Chen, Siva Theja Maguluri, Sanjay Shakkottai, and Karthikeyan Shanmugam. A Lyapunov theory for finite-sample guarantees of asynchronous Q-Learning and TD-Learning variants, 2021.
 - [Day92] Peter Dayan. The convergence of $TD(\lambda)$ for general λ . Machine Learning, 8(3):341–362, May 1992.
 - [DD11] J. G. Dai and Antonius B. Dieker. Nonnegativity of solutions to the Basic Adjoint Relationship for some diffusion processes. *Queueing Systems*, 68(3):295, Jul 2011.
 - [DDB20] Aymeric Dieuleveut, Alain Durmus, and Francis Bach. Bridging the gap between constant step size stochastic gradient descent and Markov chains. *The Annals of Statistics*, 48(3):1348 1382, 2020.
- [DMN+21] Alain Durmus, Eric Moulines, Alexey Naumov, Sergey Samsonov, Kevin Scaman, and Hoi-To Wai. Tight high probability bounds for linear stochastic approximation with fixed stepsize. In M. Ranzato, A. Beygelzimer, Y. Dauphin, P.S. Liang, and J. Wortman Vaughan, editors, Advances in Neural Information Processing Systems, volume 34, pages 30063–30074. Curran Associates, Inc., 2021.
- [DMNS22] Alain Durmus, Eric Moulines, Alexey Naumov, and Sergey Samsonov. Finite-time high-probability bounds for Polyak-Ruppert averaged iterates of linear stochastic approximation, 2022.
 - [DS94] Peter Dayan and Terrence J. Sejnowski. $\text{TD}(\lambda)$ converges with probability 1. *Machine Learning*, 14(3):295–301, Mar 1994.
- [DSTM18] Gal Dalal, Balázs Szörényi, Gugan Thoppe, and Shie Mannor. Finite sample analyses for TD(0) with function approximation. In Proceedings of the Thirty-Second AAAI Conference on Artificial Intelligence and Thirtieth Innovative Applications of Artificial Intelligence Conference and Eighth AAAI Symposium on Educational Advances in Artificial Intelligence, AAAI'18/IAAI'18/EAAI'18, New Orleans, Louisiana, USA, Apr 2018. AAAI Press.
- [EDMB03] Eyal Even-Dar, Yishay Mansour, and Peter Bartlett. Learning rates for Q-Learning. *Journal of Machine Learning Research*, 5(1):1–25, Dec 2003.
 - [Har85] J. Michael Harrison. Brownian motion and Stochastic Flow Systems. Wiley, New York, NY, USA, 1985.
 - [HW87] J. Michael Harrison and Ruth J. Williams. Multidimensional reflected Brownian motions having exponential stationary distributions. *The Annals of Probability*, 15(1):115–137, Jan 1987.
- [JKK⁺18] Prateek Jain, Sham M. Kakade, Rahul Kidambi, Praneeth Netrapalli, and Aaron Sidford. Parallelizing stochastic gradient descent for least squares regression: Mini-batching, averaging, and model misspecification. *The Journal of Machine Learning Research*, 18(223):1–42, 2018.
 - [KP00] Daphne Koller and Ronald Parr. Policy iteration for factored MDPs. In Proceedings of the Sixteenth Conference on Uncertainty in Artificial Intelligence, UAI'00, page 326–334, San Francisco, CA, USA, 2000. Morgan Kaufmann Publishers Inc.
- [KPR+21] Koulik Khamaru, Ashwin Pananjady, Feng Ruan, Martin J. Wainwright, and Michael I. Jordan. Is temporal difference learning optimal? an instance-dependent analysis. SIAM Journal on Mathematics of Data Science, 3(4):1013–1040, Jan 2021.
 - [KY03] Harold J. Kushner and G. George Yin. Stochastic Approximation and Recursive Algorithms and Applications. Stochastic Modelling and Applied Probability. Springer, New York, NY, USA, 2nd edition, 2003.
 - [LM22] Caio Kalil Lauand and Sean P. Meyn. Markovian foundations for quasi-stochastic approximation with applications to extremum seeking control, 2022.
 - [LP03] Michail G. Lagoudakis and Ronald Parr. Least-squares policy iteration. The Journal of Machine Learning Research, 4:1107–1149, Dec 2003.
 - [LP17] David A. Levin and Yuval Peres. Markov Chains and Mixing Times. American Mathematical Society, Providence, Rhode Island, USA, 2nd edition, 2017.
 - [LS18] Chandrashekar Lakshminarayanan and Csaba Szepesvári. Linear stochastic approximation: How far does constant step-size and iterate averaging go? In Amos Storkey and Fernando Perez-Cruz, editors, Proceedings of the Twenty-First International Conference on Artificial Intelligence and Statistics, volume 84 of Proceedings of Machine Learning Research, pages 1347–1355. PMLR, 09–11 Apr 2018.
- [Mey00] Carl D. Meyer. Matrix Analysis and Applied Linear Algebra. Society for Industrial and Applied Mathematics, Philadelphia, 2000.
- [MLW⁺20] Wenlong Mou, Chris Junchi Li, Martin J. Wainwright, Peter L. Bartlett, and Michael I. Jordan. On linear stochastic approximation: Fine-grained Polyak-Ruppert and non-asymptotic concentration. In Jacob Abernethy and

- Shivani Agarwal, editors, Proceedings of Thirty Third Conference on Learning Theory, volume 125 of Proceedings of Machine Learning Research, pages 2947–2997. PMLR, 09–12 Jul 2020.
- [MPWB21] Wenlong Mou, Ashwin Pananjady, Martin J. Wainwright, and Peter L. Bartlett. Optimal and instance-dependent guarantees for Markovian linear stochastic approximation, 2021.
 - [MT09] Sean P. Meyn and Richard L. Tweedie. Markov Chains and Stochastic Stability. Cambridge Mathematical Library. Cambridge University Press, Cambridge, 2nd edition, 2009.
 - [Nor97] James R. Norris. Markov Chains. Cambridge Series in Statistical and Probabilistic Mathematics. Cambridge University Press, Cambridge, 1997.
 - [PJ92] Boris T. Polyak and Anatoli B. Juditsky. Acceleration of stochastic approximation by averaging. SIAM Journal on Control and Optimization, 30(4):838–855, Jul 1992.
 - [Pol90] Boris T. Polyak. New stochastic approximation type procedures. Automation and Remote Control, 51(7):98–107, Jul 1990.
 - [RM51] Herbert Robbins and Sutton Monro. A stochastic approximation method. The Annals of Mathematical Statistics, 22(3):400 – 407, 1951.
 - [Rup88] David Ruppert. Efficient estimations from a slowly convergent Robbins-Monro process. Technical report, Cornell University, February 1988.
 - [SB02] Josef Stoer and Roland Bulirsch. Introduction to Numerical Analysis. Springer, New York, NY, USA, 3rd edition, 2002.
 - [SB18] Richard S. Sutton and Andrew G. Barto. Reinforcement Learning: An Introduction. A Bradford Book, Cambridge, MA, USA, 2018.
 - [Sha74] Eliezer Shapiro. On the Lyapunov matrix equation. IEEE Transactions on Automatic Control, 19(5):594-596, 1974.
 - [Sut88] Richard S. Sutton. Learning to predict by the methods of temporal differences. *Machine Learning*, 3(1):9–44, Aug 1988.
 - [SY19] Rayadurgam Srikant and Lei Ying. Finite-time error bounds for linear stochastic approximation and TD learning. In Alina Beygelzimer and Daniel Hsu, editors, Proceedings of the Thirty-Second Conference on Learning Theory, volume 99 of Proceedings of Machine Learning Research, pages 2803–2830. PMLR, 25–28 Jun 2019.
 - [Sze97] Csaba Szepesvári. The asymptotic convergence-rate of Q-Learning. In M. Jordan, M. Kearns, and S. Solla, editors, Advances in Neural Information Processing Systems, volume 10. MIT Press, 1997.
 - [Tsi94] John N. Tsitsiklis. Asynchronous stochastic approximation and Q-Learning. Machine Learning, 16(3):185–202, Sep 1994.
 - [TVR97] John N. Tsitsiklis and Benjamin Van Roy. An analysis of temporal-difference learning with function approximation. *IEEE Transactions on Automatic Control*, 42(5):674–690, 1997.
 - [Vil09] Cédric Villani. Optimal Transport: Old and New. Grundlehren der mathematischen Wissenschaften. Springer Berlin Heidelberg, 2009.
 - [WD92] Christopher J. C. H. Watkins and Peter Dayan. Q-Learning. Machine Learning, 8(3):279-292, May 1992.
- [YBVE21] Lu Yu, Krishnakumar Balasubramanian, Stanislav Volgushev, and Murat A. Erdogdu. An analysis of constant step size SGD in the non-convex regime: Asymptotic normality and bias. In M. Ranzato, A. Beygelzimer, Y. Dauphin, P.S. Liang, and J. Wortman Vaughan, editors, Advances in Neural Information Processing Systems, volume 34, pages 4234–4248. Curran Associates, Inc., 2021.

Appendices

A PROOFS

In this section, we provide the proofs for our pilot result in Section 6 and our main results in Section 4.

Recall that $\tau \equiv \tau_{\alpha}$ is the α -mixing time defined in Section 3.2. In the sequel, we frequently make use of the following fact: when the stepsize α satisfies the condition (4.2), we have $\alpha \tau \leq \frac{1}{4}$. This fact follows from combining the condition (4.2) with the lower bound

$$\gamma_{\max} \ge \gamma_{\min} \stackrel{\text{(i)}}{\ge} \frac{1}{2s_1(\bar{A})} \stackrel{\text{(ii)}}{\ge} \frac{1}{2},\tag{A.1}$$

where the inequality (i) is given in the paper [Sha74], and the inequality (ii) holds under Assumption 2.

A.1 Proof of Proposition 6.1

In this sub-section, we prove our pilot result in Proposition 6.1, which provides an upper bound on the finite-time MSE $\mathbb{E}[\|\theta_k - \theta^*\|^2]$.

We argue that it suffices to prove Proposition 6.1 in the special case where \bar{b} in (3.5) is assumed to be 0. Consider the LSA update rule stated in equation (3.1) with a general \bar{b} . We can center this update rule by subtracting θ^* from both sides of (3.1), which gives

$$\theta_{k+1} - \theta^* = \theta_k - \theta^* + \alpha \left[A(x_k)(\theta_k - \theta^*) + b(x_k) + A(x_k)\theta^* \right]. \tag{A.2}$$

Setting $\theta'_k := \theta_k - \theta^*$ and $b'(x_k) := b(x_k) + A(x_k)\theta^*$, we rewrite equation (A.2) as

$$\theta'_{k+1} = \theta'_k + \alpha \left[A(x_k) \theta'_k + b'(x_k) \right]. \tag{A.3}$$

Equation (A.3) is an LSA update in the variable (θ'_k) and satisfies

$$\bar{b}' := \lim_{k \to \infty} \mathbb{E}[b'(x_k)]$$

$$= \lim_{k \to \infty} \mathbb{E}[b(x_k)] + \mathbb{E}[A(x_k)]\theta^*$$

$$= \bar{b} + \bar{A}\theta^* = 0.$$

where the last equality holds since θ^* is defined as the solution to $\mathbb{E}_{\pi}[A(x)]\theta + \mathbb{E}_{\pi}[b(x)] = 0$.

Let $b'_{\text{max}} := \sup_{x \in X} ||b'(x)||$. The convergence rate of the new LSA update (A.3) is given in the following proposition, which is a centered version of Proposition 6.1.

PROPOSITION A.1. Under Assumptions 1, 2 and 3, if α satisfies equation (4.2), then the update (A.3) with $\bar{b}' = 0$ satisfies for all $k \geq \tau$,

$$\mathbb{E}[\|\theta_k'\|^2] \leq \frac{\gamma_{\max}}{\gamma_{\min}} \left(1 - \frac{0.9\alpha}{\gamma_{\max}}\right)^{k-\tau} \left(4\mathbb{E}[\|\theta_0'\|^2] + (b_{\max}')^2\right) + \frac{\gamma_{\max}}{0.9\gamma_{\min}} \cdot \alpha\tau \left(160\gamma_{\max}(b_{\max}')^2\right).$$

We prove the above proposition in Appendix A.1.1. Taking Proposition A.1 as given, we now complete the proof of the general Proposition 6.1.

PROOF OF PROPOSITION 6.1. By definition of b', we have

$$||b'(x)|| = ||b(x) + A(x)\theta^*|| \le ||b(x)|| + ||A(x)|| ||\theta^*||, \quad \forall x \in \mathcal{X},$$

whence

$$b'_{\max} \le b_{\max} + A_{\max} \|\theta^*\|$$

 $\le (1 + A_{\max}/s_{\min}(\bar{A}))b_{\max} \le 2s_{\min}^{-1}(\bar{A})b_{\max}.$

Applying Proposition A.1 with $\theta'_k = \theta_k - \theta^*$ and the above bound on b'_{max} , we obtain that for all $k \geq \tau$,

$$\mathbb{E}[\|\theta_{k} - \theta^{*}\|^{2}] \leq 4 \frac{\gamma_{\max}}{\gamma_{\min}} \left(1 - \frac{0.9\alpha}{\gamma_{\max}}\right)^{k-\tau} \left(\mathbb{E}[\|\theta_{0} - \theta^{*}\|^{2}] + s_{\min}^{-2}(\bar{A})b_{\max}^{2}\right) + \frac{\gamma_{\max}}{0.9\gamma_{\min}} \cdot \alpha\tau \left(640\gamma_{\max}s_{\min}^{-2}(\bar{A})b_{\max}^{2}\right).$$

We can simplify the above expression using the following simple bound, whose proof is postponed to the end of this sub-sub-section.

Claim 1. We have
$$\left(1 - \frac{0.9\alpha}{\gamma_{\text{max}}}\right)^{-\tau} \leq 2$$
.

Using the bound in Claim 1 and the definition of κ in equation (4.3), we obtain that for all $k \geq \tau$,

$$\mathbb{E}[\|\theta_k - \theta^*\|^2] \leq 8 \frac{\gamma_{\max}}{\gamma_{\min}} \left(1 - \frac{0.9\alpha}{\gamma_{\max}}\right)^k \left(\mathbb{E}[\|\theta_0 - \theta^*\|^2] + s_{\min}^{-2}(\bar{A})b_{\max}^2\right) + \frac{\gamma_{\max}}{0.9\gamma_{\min}} \cdot \alpha\tau \cdot \kappa.$$

As such, we have completed the proof of Proposition 6.1.

PROOF OF CLAIM 1. Observe that

$$\frac{0.9\alpha}{\gamma_{\text{max}}} \stackrel{\text{(i)}}{\leq} \frac{0.9\alpha\tau}{\gamma_{\text{max}}} \stackrel{\text{(ii)}}{\leq} 2\alpha\tau \stackrel{\text{(iii)}}{\leq} \frac{1}{2},\tag{A.4}$$

where step (i) holds since $\tau \ge 1$, step (ii) follows from the bound (A.1), and step (iii) holds since $\alpha \tau \le \frac{1}{4}$ under the stepsize condition (4.2). To proceed, we use the Bernoulli inequality

$$(1+x)^t \ge 1+xt$$
 for $x \ge -1, t \ge 1$,

which is equivalent to

$$(1-x)^{-t} \le (1-xt)^{-1}$$
 for $0 < x < 1, t \ge 1$.

In light of equation (A.4), the Bernoulli inequality holds with $x=\frac{0.9\alpha}{\gamma_{\rm max}}$ and $t=\tau$, hence

$$\left(1 - \frac{0.9\alpha}{\gamma_{\text{max}}}\right)^{-\tau} \le \frac{1}{1 - \frac{0.9\alpha\tau}{\gamma_{\text{max}}}} \le 2,$$

where the last step follows from (A.4). We have completed the proof of Claim 1.

A.1.1 Proof of Proposition A.1. We need the following technical lemmas.

LEMMA A.2. Given any $t \ge 1$, if $\alpha \cdot t \le \frac{1}{4}$, then the following inequalities hold for all $k \ge t$,

$$\|\theta_k - \theta_{k-t}\| \le 2\alpha t \|\theta_{k-t}\| + 2\alpha t b_{\text{max}},$$
 (A.5)

$$\|\theta_k - \theta_{k-t}\| \le 4\alpha t \|\theta_k\| + 4\alpha t b_{\text{max}},\tag{A.6}$$

$$\|\theta_k - \theta_{k-t}\|^2 \le 32\alpha^2 t^2 \|\theta_k\|^2 + 32\alpha^2 t^2 b_{\text{max}}^2.$$
(A.7)

LEMMA A.3. The following inequality holds for any $k \geq 0$,

$$\left| (\theta_{k+1} - \theta_k)^{\mathsf{T}} \Gamma(\theta_{k+1} - \theta_k) \right| \le 2\alpha^2 \gamma_{\max} \|\theta_k\|^2 + 2\alpha^2 \gamma_{\max} b_{\max}^2.$$

Lemma A.4. The following inequality holds for all $k \ge \tau$, with α chosen sufficiently small such that $\alpha \tau \le \frac{1}{4}$,

$$\mathbb{E}\left[\theta_k^{\top} \Gamma(A(x_k) - \bar{A})\theta_k\right] \leq \kappa_1 \mathbb{E}[\|\theta_k\|^2] + \kappa_2,$$

where

$$\kappa_1 = 88\alpha\tau\gamma_{\text{max}} \quad and \quad \kappa_2 = 64\alpha\tau\gamma_{\text{max}}b_{\text{max}}^2$$

LEMMA A.5. The following inequality holds for all $k \ge \tau$, with α chosen sufficiently small such that $\alpha \tau \le \frac{1}{4}$,

$$\mathbb{E}\left[\theta_k^\top \Gamma(b(x_k) - \bar{b})\right] \leq \widetilde{\kappa}_1 \mathbb{E}[\|\theta_k\|^2] + \widetilde{\kappa}_2,$$

where

$$\widetilde{\kappa}_1 = 5\alpha\tau\gamma_{\max}$$
 and $\widetilde{\kappa}_2 = 15\alpha\tau\gamma_{\max}b_{\max}^2$.

The proofs of the technical lemmas above are delayed to Appendix A.1.2. Note that all lemmas above hold for the LSA update (3.1) with general \bar{b} . Below we shall apply these lemmas to the centered LSA update (A.3) for θ'_k with $\bar{b}' = 0$ to prove Proposition A.1.

PROOF OF PROPOSITION A.1. Consider the following drift:

$$\begin{split} & \mathbb{E}[\theta_{k+1}^{\prime}^{\top}\Gamma\theta_{k+1}^{\prime} - \theta_{k}^{\prime}^{\top}\Gamma\theta_{k}^{\prime}] = 2\mathbb{E}[\theta_{k}^{\prime}^{\top}\Gamma(\theta_{k+1}^{\prime} - \theta_{k}^{\prime})] + \mathbb{E}[(\theta_{k+1}^{\prime} - \theta_{k}^{\prime})^{\top}\Gamma(\theta_{k+1}^{\prime} - \theta_{k}^{\prime})] \\ & = 2\alpha\underbrace{\mathbb{E}[\theta_{k}^{\prime}^{\top}\Gamma(A(x_{k}) - \bar{A})\theta_{k}^{\prime}]}_{T_{1}} + 2\alpha\underbrace{\mathbb{E}[\theta_{k}^{\prime}^{\top}\Gamma b^{\prime}(x_{k})]}_{T_{2}} + 2\alpha\underbrace{\mathbb{E}[\theta_{k}^{\prime}^{\top}\Gamma \bar{A}\theta_{k}^{\prime}]}_{T_{3}} + \underbrace{\mathbb{E}[(\theta_{k+1}^{\prime} - \theta_{k}^{\prime})^{\top}\Gamma(\theta_{k+1}^{\prime} - \theta_{k}^{\prime})]}_{T_{4}}. \end{split}$$

We can bound T_1 using Lemma A.4, T_2 using Lemma A.5, and T_4 using Lemma A.3. For T_3 , we note that by the property of the Lyapunov equation in Assumption 3,

$$2\alpha \mathbb{E}[\theta_k'^{\top} \Gamma \bar{A} \theta_k'] = \alpha \mathbb{E}[\theta_k'^{\top} \underbrace{(\bar{A}^{\top} \Gamma + \Gamma \bar{A})}_{=-I} \theta_k'] = -\alpha \mathbb{E}[\|\theta_k'\|^2].$$

Combining the above bounds, we derive that

$$\begin{split} & \mathbb{E}[\theta_{k+1}^{\prime}^{\top} \Gamma \theta_{k+1}^{\prime} - \theta_{k}^{\prime}^{\top} \Gamma \theta_{k}^{\prime}] \\ &= T_{1} + T_{2} + T_{3} + T_{4} \\ &\leq 2\alpha \left(\kappa_{1} \mathbb{E}[\|\theta_{k}^{\prime}\|^{2}] + \kappa_{2} \right) + 2\alpha \left(\tilde{\kappa}_{1} \mathbb{E}[\|\theta_{k}^{\prime}\|^{2}] + \tilde{\kappa}_{2} \right) - \alpha \mathbb{E}[\|\theta_{k}^{\prime}\|^{2}] + \left(2\gamma_{\max}\alpha^{2} \mathbb{E}[\|\theta_{k}^{\prime}\|^{2}] + 2\alpha^{2}\gamma_{\max}(b_{\max}^{\prime})^{2} \right) \\ &= -\alpha (1 - 2(\kappa_{1} + \tilde{\kappa}_{1} + \alpha\gamma_{\max})) \mathbb{E}[\|\theta_{k}^{\prime}\|^{2}] + 2\alpha (\kappa_{2} + \tilde{\kappa}_{2} + \alpha\gamma_{\max}(b_{\max}^{\prime})^{2}). \end{split}$$

We simplify the above bound by noting that

$$\kappa_1 + \tilde{\kappa}_1 + \alpha \gamma_{\text{max}} = 88\alpha \tau \gamma_{\text{max}} + 5\alpha \tau \gamma_{\text{max}} + \alpha \gamma_{\text{max}}$$

$$\leq 95\alpha \tau \gamma_{\text{max}},$$

and

$$\kappa_2 + \tilde{\kappa}_2 + \alpha \gamma_{\text{max}} (b'_{\text{max}})^2$$

$$= 64\alpha \tau \gamma_{\text{max}} (b'_{\text{max}})^2 + 15\alpha \tau \gamma_{\text{max}} (b'_{\text{max}})^2 + \alpha \gamma_{\text{max}} (b'_{\text{max}})^2$$

$$\leq 80\alpha \tau \gamma_{\text{max}} (b'_{\text{max}})^2.$$

Furthermore, when α is chosen according to (4.2), we obtain that for all $k \geq \tau$,

$$\begin{split} \mathbb{E}[\boldsymbol{\theta}_{k+1}^{\prime}^{\top}\boldsymbol{\Gamma}\boldsymbol{\theta}_{k+1}^{\prime} - \boldsymbol{\theta}_{k}^{\prime}^{\top}\boldsymbol{\Gamma}\boldsymbol{\theta}_{k}^{\prime}] &\leq -0.9\alpha\mathbb{E}[\|\boldsymbol{\theta}_{k}^{\prime}\|^{2}] + 160\alpha^{2}\tau\gamma_{\max}(b_{\max}^{\prime})^{2} \\ &\leq -\frac{0.9\alpha}{\gamma_{\max}}\mathbb{E}[\boldsymbol{\theta}_{k}^{\prime}^{\top}\boldsymbol{\Gamma}\boldsymbol{\theta}_{k}^{\prime}] + 160\alpha^{2}\tau\gamma_{\max}(b_{\max}^{\prime})^{2}, \end{split}$$

or equivalently

$$\mathbb{E}[\theta_{k+1}^{\prime}^{\top}\Gamma\theta_{k+1}^{\prime}] \leq \left(1 - \frac{0.9\alpha}{\gamma_{\max}}\right) \mathbb{E}[\theta_{k}^{\prime}^{\top}\Gamma\theta_{k}^{\prime}] + 160\alpha^{2}\tau\gamma_{\max}(b_{\max}^{\prime})^{2}.$$

Next, we recursively apply the above inequality to obtain

$$\begin{split} \mathbb{E}[\boldsymbol{\theta}_{k}^{\prime\top}\boldsymbol{\Gamma}\boldsymbol{\theta}_{k}^{\prime}] &\leq \left(1 - \frac{0.9\alpha}{\gamma_{\max}}\right)^{k-\tau} \mathbb{E}[\boldsymbol{\theta}_{\tau}^{\prime\top}\boldsymbol{\Gamma}\boldsymbol{\theta}_{\tau}^{\prime}] + \sum_{t=0}^{(k-\tau)-1} \left(1 - \frac{0.9\alpha}{\gamma_{\max}}\right)^{t} \cdot \left(160\alpha^{2}\tau\gamma_{\max}(b_{\max}^{\prime})^{2}\right) \\ &\leq \left(1 - \frac{0.9\alpha}{\gamma_{\max}}\right)^{k-\tau} \mathbb{E}[\boldsymbol{\theta}_{\tau}^{\prime\top}\boldsymbol{\Gamma}\boldsymbol{\theta}_{\tau}^{\prime}] + \frac{\gamma_{\max}}{0.9} \cdot \left(160\alpha\tau\gamma_{\max}(b_{\max}^{\prime})^{2}\right). \end{split}$$

We then apply the properties in (3.9) to the above inequality and obtain the following bounds in terms of $\|\theta'_k\|^2$, for $k \ge \tau$,

$$\begin{split} \mathbb{E}[\|\theta_k'\|^2] &\leq \frac{1}{\gamma_{\min}} \mathbb{E}[\theta_k'^{\top} \Gamma \theta_k'] \\ &\leq \frac{\gamma_{\max}}{\gamma_{\min}} \left(1 - \frac{0.9\alpha}{\gamma_{\max}} \right)^{k-\tau} \mathbb{E}[\|\theta_\tau'\|^2] + \frac{\gamma_{\max}}{0.9\gamma_{\min}} \cdot \alpha \tau \left(160\gamma_{\max}(b_{\max}')^2 \right). \end{split}$$

Lastly, we have

$$\begin{aligned} \|\theta_{\tau}'\|_{2}^{2} &\leq \left(\|\theta_{\tau}' - \theta_{0}'\| + \|\theta_{0}'\|\right)^{2} \\ &\stackrel{\text{(i)}}{\leq} \left((1 + 2\alpha\tau)\|\theta_{0}'\| + 2\alpha\tau b_{\max}'\right)^{2} \\ &\stackrel{\text{(ii)}}{\leq} \left(1.5\|\theta_{0}'\| + 0.5b_{\max}'\right)^{2} \leq 4\|\theta_{0}'\|^{2} + (b_{\max}')^{2}, \end{aligned}$$

where in step (i) we make use of Lemma A.2 to bound $\|\theta'_{\tau} - \theta'_{0}\|$ with $\|\theta'_{0}\|$, and step (ii) holds for α is chosen according to (4.2) such that $\alpha \tau < \frac{1}{4}$. Therefore, we have

$$\mathbb{E}[\|\boldsymbol{\theta}_k'\|^2] \leq \frac{\gamma_{\max}}{\gamma_{\min}} \left(1 - \frac{0.9\alpha}{\gamma_{\max}}\right)^{k-\tau} \left(4\mathbb{E}[\|\boldsymbol{\theta}_0'\|^2] + (b_{\max}')^2\right) + \frac{\gamma_{\max}}{0.9\gamma_{\min}} \cdot \alpha\tau \left(160\gamma_{\max}(b_{\max}')^2\right).$$

This concludes the proof for Proposition A.1.

A.1.2 Proof of Technical Lemmas. We provide the proofs of the technical lemmas stated at the beginning of the previous sub-sub-section.

PROOF OF LEMMA A.2. We first note the following,

$$\begin{aligned} \|\theta_{k+1}\| &= \|\theta_k + \alpha(A(x_k)\theta_k + b(x_k))\| \\ &\leq \|I + \alpha A(x_k)\| \|\theta_k\| + \alpha \|b(x_k)\| \\ &\leq (1 + \alpha A_{\max}) \|\theta_k\| + \alpha b_{\max} \leq (1 + \alpha) \|\theta_k\| + \alpha b_{\max}. \end{aligned}$$

As such, for $k - t < i \le k$, we have

$$\|\theta_{i}\| \leq (1+\alpha)^{i-(k-t)} \|\theta_{k-t}\| + \alpha b_{\max} \sum_{j=0}^{(i-1)-(k-t)} (1+\alpha)^{j}$$

$$\leq (1+\alpha)^{t} \|\theta_{k-t}\| + \alpha b_{\max} \sum_{j=0}^{t-1} (1+\alpha)^{j}$$

$$= (1+\alpha)^{t} \|\theta_{k-t}\| + b_{\max} ((1+\alpha)^{t} - 1)$$

$$\stackrel{(i)}{\leq} (1+2\alpha t) \|\theta_{k-t}\| + 2\alpha t b_{\max}, \tag{A.8}$$

where step (i) holds true as long as $\alpha t \leq \log 2$, and since we have restricted $\alpha t \leq \frac{1}{4} \leq \log 2$, the inequality carries through.

It follows that

$$\|\theta_{k} - \theta_{k-t}\| = \left\| \sum_{i=k-t}^{k-1} \theta_{i+1} - \theta_{i} \right\| \leq \sum_{i=k-t}^{k-1} \|\theta_{i+1} - \theta_{i}\| = \alpha \sum_{i=k-t}^{k-1} \|A(x_{k})\theta_{i} + b(x_{k})\|$$

$$\leq \alpha \sum_{i=k-t}^{k-1} (\|A(x_{k})\| \|\theta_{i}\| + \|b(x_{k})\|)$$

$$\leq \alpha A_{\max} \left(\sum_{i=k-t}^{k-1} \|\theta_{i}\| \right) + \alpha t b_{\max} \qquad \text{by Assumption 2}$$

$$\leq \alpha A_{\max} \left(\sum_{i=k-t}^{k-1} (1 + 2\alpha t) \|\theta_{k-t}\| + 2\alpha t b_{\max} \right) + \alpha t b_{\max} \qquad \text{by (A.8)}$$

$$= (1 + 2\alpha t) (\alpha t A_{\max} \|\theta_{k-t}\| + \alpha t b_{\max})$$

$$\stackrel{\text{(ii)}}{\leq} 2\alpha t (A_{\max} \|\theta_{k-t}\| + b_{\max}) < 2\alpha t \|\theta_{k-t}\| + 2\alpha t b_{\max},$$

where step (ii) holds since $2\alpha t < 1$. As such, we have established (A.5).

Now that with (A.5), it is easy to see

$$\|\theta_k - \theta_{k-t}\| \le 2\alpha t \|\theta_{k-t}\| + 2\alpha t b_{\max}$$

$$\le 2\alpha t (\|\theta_k - \theta_{k-t}\| + \|\theta_k\|) + 2\alpha t b_{\max}.$$

Reorganizing the above inequality, we have

$$(1 - 2\alpha t)\|\theta_k - \theta_{k-t}\| \le 2\alpha t\|\theta_k\| + 2\alpha t b_{\max}.$$

Together with the assumption $\alpha t \leq \frac{1}{4}$, we now obtain (A.6),

$$\|\theta_k - \theta_{k-t}\| \le 4\alpha t \|\theta_k\| + 4\alpha t b_{\max}$$

Lastly, we have

$$\begin{aligned} \|\theta_k - \theta_{k-t}\|^2 &\le (4\alpha t \|\theta_k\| + 4\alpha t b_{\max})^2 \\ &\le 2(4\alpha t \|\theta_k\|)^2 + 2(4\alpha t b_{\max})^2 \\ &= 32\alpha^2 t^2 \|\theta_k\|^2 + 32\alpha^2 t^2 b_{\max}^2, \end{aligned}$$

which establishes (A.7).

Proof of Lemma A.3. We have

$$\begin{split} \left| (\theta_{k+1} - \theta_k)^\top \Gamma(\theta_{k+1} - \theta_k) \right| &\leq \gamma_{\max} \|\theta_{k+1} - \theta_k\|^2 \\ &= \gamma_{\max} \alpha^2 \|A(x_k) \theta_k + b(x_k)\|^2 \\ &\leq \gamma_{\max} \alpha^2 \left(\|A(x_k) \| \|\theta_k\| + \|b(x_k)\| \right)^2 \\ &\leq \gamma_{\max} \alpha^2 \left(A_{\max} \|\theta_k\| + b_{\max} \right)^2 \\ &\leq 2\gamma_{\max} \alpha^2 \|\theta_k\|^2 + 2\alpha^2 \gamma_{\max} b_{\max}^2. \end{split}$$

This completes the proof of Lemma A.3.

PROOF OF LEMMA A.4. As explained in Section 3.2, (x_k) enjoys a geometric mixing rate. Thus, it holds as claimed that for a sufficiently small α ,

$$\alpha \tau_{\alpha} \le K \alpha \log \frac{1}{\alpha} \le \frac{1}{4}.$$

Let us decompose the quantity of interest as

$$\mathbb{E}\left[\theta_{k}^{\top}\Gamma(A(x_{k})-\bar{A})\theta_{k}\right]$$

$$=\mathbb{E}\left[(\theta_{k}-\theta_{k-\tau}+\theta_{k-\tau})^{\top}\Gamma(A(x_{k})-\bar{A})(\theta_{k}-\theta_{k-\tau}+\theta_{k-\tau})\right]$$

$$=\mathbb{E}\left[(\theta_{k}-\theta_{k-\tau})^{\top}\Gamma(A(x_{k})-\bar{A})(\theta_{k}-\theta_{k-\tau})\right] + \mathbb{E}\left[\theta_{k-\tau}^{\top}\Gamma(A(x_{k})-\bar{A})\theta_{k-\tau}\right]$$

$$+\mathbb{E}\left[(\theta_{k}-\theta_{k-\tau})^{\top}\Gamma(A(x_{k})-\bar{A})\theta_{k-\tau}\right] + \mathbb{E}\left[\theta_{k-\tau}^{\top}\Gamma(A(x_{k})-\bar{A})(\theta_{k}-\theta_{k-\tau})\right].$$

We now bound each of the RHS terms respectively.

For T_1 , we have

$$T_{1} = \mathbb{E}\left[(\theta_{k} - \theta_{k-\tau})^{\mathsf{T}}\Gamma(A(x_{k}) - \bar{A})(\theta_{k} - \theta_{k-\tau})\right]$$

$$\stackrel{\text{(i)}}{\leq} 2\gamma_{\max}\mathbb{E}\left[\|\theta_{k} - \theta_{k-\tau}\|^{2}\right]$$

$$\stackrel{\text{(ii)}}{\leq} 2\gamma_{\max}\left(32\alpha^{2}\tau^{2}\mathbb{E}[\|\theta_{k}\|^{2}] + 32\alpha^{2}\tau^{2}b_{\max}^{2}\right)$$

$$\leq 64\gamma_{\max}\alpha^{2}\tau^{2}\mathbb{E}[\|\theta_{k}\|^{2}] + 64\gamma_{\max}\alpha^{2}\tau^{2}b_{\max}^{2},$$

where (i) holds true for both $A(x_k)$ and \bar{A} have norm less than $A_{\max} (\leq 1)$ in Assumption 2 and Γ is symmetric positive definite with γ_{\max} as the largest eigenvalue in Assumption 3, and (ii) is obtained by making use of (A.7) of Lemma A.2.

For T_2 , we have

$$\begin{split} T_2 &= \mathbb{E}\left[\theta_{k-\tau}^{\top} \Gamma(A(x_k) - \bar{A})\theta_{k-\tau}\right] \\ &= \mathbb{E}\left[\mathbb{E}\left[\theta_{k-\tau}^{\top} \Gamma(A(x_k) - \bar{A})\theta_{k-\tau} \middle| \theta_{k-\tau}, x_{k-\tau}\right]\right] \\ &= \mathbb{E}\left[\theta_{k-\tau}^{\top} \Gamma \mathbb{E}\left[A(x_k) - \bar{A} \middle| \theta_{k-\tau}, x_{k-\tau}\right]\theta_{k-\tau}\right] \\ &\stackrel{\text{(iii)}}{=} \mathbb{E}\left[\theta_{k-\tau}^{\top} \Gamma \mathbb{E}\left[A(x_k) - \bar{A} \middle| x_{k-\tau}\right]\theta_{k-\tau}\right]. \end{split}$$

where step (iii) holds true due to conditional independence of $x_k \perp \!\!\! \perp \theta_{k-\tau} | x_{k-\tau}$, as demonstrated earlier in equation (3.3). Since Γ has largest eigenvalue γ_{\max} by Assumption 3 and $\tau \equiv \tau_{\alpha}$ is the α -mixing time, it follows that

$$T_{2} \leq \alpha \gamma_{\max} \mathbb{E} \left[\|\theta_{k-\tau}\|^{2} \right] = \alpha \gamma_{\max} \mathbb{E} \left[\|(\theta_{k} - \theta_{k-\tau}) + \theta_{k}\|^{2} \right]$$

$$\leq \alpha \gamma_{\max} \mathbb{E} \left[\left(\|\theta_{k} - \theta_{k-\tau}\| + \|\theta_{k}\| \right)^{2} \right]$$

$$\leq \alpha \gamma_{\max} \mathbb{E} \left[\left(4\alpha \tau \|\theta_{k}\| + 4\alpha \tau b_{\max} + \|\theta_{k}\| \right)^{2} \right] \quad \text{by (A.6)}$$

$$\stackrel{\text{(iv)}}{\leq} \alpha \gamma_{\max} \cdot 2 \left((1 + 4\alpha \tau)^{2} \mathbb{E} [\|\theta_{k}\|^{2}] + 16\alpha^{2} \tau^{2} b_{\max}^{2} \right)$$

$$\leq 8\alpha \tau \gamma_{\max} \mathbb{E} [\|\theta_{k}\|^{2}] + 32\alpha^{3} \tau^{2} \gamma_{\max} b_{\max}^{2},$$

where (iv) follows from the inequality $(a+b)^2 \le 2(a^2+b^2)$, and the last step holds since $\alpha \tau \le \frac{1}{4}$ and $\tau \ge 1$.

For T_3 , we have

$$\begin{split} T_{3} &= \mathbb{E}\left[(\theta_{k} - \theta_{k-\tau})^{\top}\Gamma(A(x_{k}) - \bar{A})\theta_{k-\tau}\right] \\ &\leq 2\gamma_{\max}\mathbb{E}\left[\|\theta_{k} - \theta_{k-\tau}\| \cdot \|\theta_{k-\tau}\|\right] \\ &\leq 2\gamma_{\max}\mathbb{E}\left[\|\theta_{k} - \theta_{k-\tau}\| \cdot (\|\theta_{k} - \theta_{k-\tau}\| + \|\theta_{k}\|)\right] \\ &\leq 2\gamma_{\max}\mathbb{E}\left[(4\alpha\tau\|\theta_{k}\| + 4\alpha\tau b_{\max})(4\alpha\tau\|\theta_{k}\| + 4\alpha\tau b_{\max} + \|\theta_{k}\|)\right] \quad \text{by (A.6)} \\ &= 2\gamma_{\max}\mathbb{E}\left[4\alpha\tau(1 + 4\alpha\tau)\|\theta_{k}\|^{2} + 4\alpha\tau b_{\max}(1 + 8\alpha\tau)\|\theta_{k}\| + 16\alpha^{2}\tau^{2}b_{\max}^{2}\right] \\ &= 8\alpha\tau(1 + 4\alpha\tau)\gamma_{\max}\mathbb{E}\left[\|\theta_{k}\|^{2}\right] + 8\alpha\tau(1 + 8\alpha\tau)\gamma_{\max}b_{\max}\mathbb{E}\left[\|\theta_{k}\|\right] + 32\alpha^{2}\tau^{2}\gamma_{\max}b_{\max}^{2} \\ &\leq 8\alpha\tau(1 + 4\alpha\tau)\gamma_{\max}\mathbb{E}\left[\|\theta_{k}\|^{2}\right] + 4\alpha\tau(1 + 8\alpha\tau)\gamma_{\max}\left(b_{\max}^{2} + \mathbb{E}[\|\theta_{k}\|^{2}]\right) + 32\alpha^{2}\tau^{2}\gamma_{\max}b_{\max}^{2} \\ &= 4\alpha\tau\gamma_{\max}\left(2(1 + 4\alpha\tau) + (1 + 8\alpha\tau)\right)\mathbb{E}[\|\theta_{k}\|^{2}] + 4\alpha\tau\gamma_{\max}\left((1 + 8\alpha\tau) + 8\alpha\tau\right)b_{\max}^{2} \\ &\leq 32\alpha\tau\gamma_{\max}\mathbb{E}[\|\theta_{k}\|^{2}] + 20\alpha\tau\gamma_{\max}b_{\max}^{2}, \end{split}$$

where (v) utilizes the inequality $2b_{\max}\mathbb{E}[\|\theta_k\|] \le b_{\max}^2 + \mathbb{E}[\|\theta_k\|^2]$, and (vi) holds with $\alpha\tau \le \frac{1}{4}$. Similarly, for T_4 , we have for $\alpha\tau \le \frac{1}{4}$,

$$T_4 = \mathbb{E}\left[\theta_{k-\tau}^{\top} \Gamma(A(x_k) - \bar{A})(\theta_k - \theta_{k-\tau})\right]$$

$$\leq 32\alpha\tau\gamma_{\max} \mathbb{E}[\|\theta_k\|^2] + 20\alpha\tau\gamma_{\max}b_{\max}^2.$$

Combining the bounds for T_1 – T_4 , we obtain that

$$\begin{split} &\mathbb{E}\left[(\theta_{k} - \theta_{k-\tau} + \theta_{k-\tau})^{\top} \Gamma(A(x_{k}) - \bar{A})(\theta_{k} - \theta_{k-\tau} + \theta_{k-\tau})\right] \\ = & T_{1} + T_{2} + T_{3} + T_{4} \\ \leq & \left(64\gamma_{\max}\alpha^{2}\tau^{2}\mathbb{E}[\|\theta_{k}\|^{2}] + 64\gamma_{\max}\alpha^{2}\tau^{2}b_{\max}^{2}\right) + \left(8\alpha\tau\gamma_{\max}\mathbb{E}[\|\theta_{k}\|^{2}] + 32\alpha^{3}\tau^{2}\gamma_{\max}b_{\max}^{2}\right) \\ & + 2\left(32\alpha\tau\gamma_{\max}\mathbb{E}[\|\theta_{k}\|^{2}] + 20\alpha\tau\gamma_{\max}b_{\max}^{2}\right) \\ = & \left(64\gamma_{\max}\alpha^{2}\tau^{2} + 8\alpha\tau\gamma_{\max} + 64\alpha\tau\gamma_{\max}\right)\mathbb{E}[\|\theta_{k}\|^{2}] \\ & + \left(64\gamma_{\max}\alpha^{2}\tau^{2}b_{\max}^{2} + 32\alpha^{3}\tau^{2}\gamma_{\max}b_{\max}^{2} + 40\alpha\tau\gamma_{\max}b_{\max}^{2}\right) \\ \leq & 88\alpha\tau\gamma_{\max}\mathbb{E}[\|\theta_{k}\|^{2}] + 64\alpha\tau\gamma_{\max}b_{\max}^{2}, \end{split}$$

where the last step holds with $\alpha \le 1$ and $\alpha \tau \le \frac{1}{4}$. This completes the proof of Lemma A.4.

PROOF OF LEMMA A.5. We make use of the following decomposition:

$$\mathbb{E}\left[\theta_{k}^{\mathsf{T}}\Gamma(b(x_{k})-\bar{b})|\theta_{k-\tau},x_{k-\tau}\right]$$

$$=\mathbb{E}\left[(\theta_{k}-\theta_{k-\tau}+\theta_{k-\tau})^{\mathsf{T}}\Gamma(b(x_{k})-\bar{b})|\theta_{k-\tau},x_{k-\tau}\right]$$

$$=\mathbb{E}\left[\theta_{k-\tau}^{\mathsf{T}}\Gamma(b(x_{k})-\bar{b})|\theta_{k-\tau},x_{k-\tau}\right] + \mathbb{E}\left[(\theta_{k}-\theta_{k-\tau})^{\mathsf{T}}\Gamma(b(x_{k})-\bar{b})|\theta_{k-\tau},x_{k-\tau}\right].$$

We separately bound T_1 and T_2 .

Starting with T_1 , we have

$$\mathbb{E}\left[\theta_{k-\tau}^{\top}\Gamma(b(x_k) - \bar{b})|\theta_{k-\tau}, x_{k-\tau}\right] = \theta_{k-\tau}^{\top}\Gamma\mathbb{E}\left[(b(x_k) - \bar{b})|\theta_{k-\tau}, x_{k-\tau}\right] \\ \leq \alpha\gamma_{\max}b_{\max}\|\theta_{k-\tau}\|.$$

, Vol. 1, No. 1, Article . Publication date: June 2023.

Next, for T_2 , we have

$$\mathbb{E}\left[(\theta_k - \theta_{k-\tau})^{\top} \Gamma(b(x_k) - \bar{b}) | \theta_{k-\tau}, x_{k-\tau}\right] \leq 2b_{\max} \gamma_{\max} \mathbb{E}\left[\|\theta_k - \theta_{k-\tau}\| | \theta_{k-\tau}, x_{k-\tau}\right]$$

$$\leq 2b_{\max} \gamma_{\max} \left(2\alpha\tau \|\theta_{k-\tau}\| + 2\alpha\tau b_{\max}\right),$$

where we use (A.5) to obtain the second inequality.

Combining the two terms, we have

$$\mathbb{E}\left[\theta_{k}^{\mathsf{T}}\Gamma(b(x_{k}) - \bar{b})|\theta_{k-\tau}, x_{k-\tau}\right]$$

$$\leq \alpha \gamma_{\max} b_{\max} \|\theta_{k-\tau}\| + 2b_{\max} \gamma_{\max} \left(2\alpha\tau \|\theta_{k-\tau}\| + 2\alpha\tau b_{\max}\right)$$

$$= (\alpha \gamma_{\max} b_{\max} + 4\alpha\tau \gamma_{\max} b_{\max}) \|\theta_{k-\tau}\| + 4\alpha\tau \gamma_{\max} b_{\max}^{2}$$

$$= \alpha \gamma_{\max} b_{\max} \left(1 + 4\tau\right) \|\theta_{k-\tau}\| + 4\alpha\tau \gamma_{\max} b_{\max}^{2}$$

$$\leq \alpha \gamma_{\max} b_{\max} \left(1 + 4\tau\right) \left(\mathbb{E}[\|\theta_{k} - \theta_{k-\tau}\| \|\theta_{k-\tau}, x_{k-\tau}] + \mathbb{E}[\|\theta_{k}\| \|\theta_{k-\tau}, x_{k-\tau}]\right) + 4\alpha\tau \gamma_{\max} b_{\max}^{2}$$

$$\stackrel{(i)}{\leq} \alpha \gamma_{\max} b_{\max} \left(1 + 4\tau\right) \left(\left(1 + 4\alpha\tau\right) \mathbb{E}[\|\theta_{k}\| \|\theta_{k-\tau}, x_{k-\tau}] + 4\alpha\tau b_{\max}\right) + 4\alpha\tau \gamma_{\max} b_{\max}^{2}$$

$$\leq 10\alpha\tau \gamma_{\max} b_{\max} \mathbb{E}[\|\theta_{k}\| \|\theta_{k-\tau}, x_{k-\tau}] + 9\alpha\tau \gamma_{\max} b_{\max}^{2},$$

where we use (A.6) to obtain (i), and $\alpha \tau \leq \frac{1}{4}$, $\alpha \leq 1$ and $\tau \geq 1$ to obtain the last inequality.

Together with the inequality $2b_{\max}\|\theta_k\| \le b_{\max}^2 + \|\theta_k\|^2$, the above inequality further simplifies to

$$\mathbb{E}\left[\theta_k^{\top}\Gamma(b(x_k) - \bar{b})|\theta_{k-\tau}, x_{k-\tau}\right] \leq 5\alpha\tau\gamma_{\max}(b_{\max}^2 + \mathbb{E}\left[\|\theta_k\|^2|\theta_{k-\tau}, x_{k-\tau}\right]) + 9\alpha\tau\gamma_{\max}b_{\max}^2$$
$$\leq 5\alpha\tau\gamma_{\max}\mathbb{E}\left[\|\theta_k\|^2|\theta_{k-\tau}, x_{k-\tau}\right] + 15\alpha\tau\gamma_{\max}b_{\max}^2.$$

Lastly, we take expectations on both sides of the last display equation to obtain

$$\mathbb{E}\left[\theta_k^\top \Gamma(b(x_k) - \bar{b})\right] \leq 5\alpha\tau\gamma_{\max}\mathbb{E}[\|\theta_k\|^2] + 15\alpha\tau\gamma_{\max}b_{\max}^2.$$

This completes the proof of Lemma A.5.

A.2 Proof of Theorem 4.1

In this sub-section, we prove Theorem 4.1 on the convergence of LSA to a limit.

A.2.1 Coupling and Geometric Convergence. Recall that $(x_k)_{k\geq 0}$ is the underlying Markov chain that drives the LSA iteration (3.1). We consider a pair of coupled Markov chains, $(x_k, \theta_k^{[1]})_{k\geq 0}$ and $(x_k, \theta_k^{[2]})_{k\geq 0}$, defined as

$$\theta_{k+1}^{[1]} = \theta_k^{[1]} + \alpha \left(A(x_k) \theta_k^{[1]} + b(x_k) \right),
\theta_{k+1}^{[2]} = \theta_k^{[2]} + \alpha \left(A(x_k) \theta_k^{[2]} + b(x_k) \right),$$

$$k = 0, 1, \dots$$
(A.9)

Note that $(\theta_k^{[1]})_{k\geq 0}$ and $(\theta_k^{[2]})_{k\geq 0}$ are two sample paths of the LSA iteration (3.1), coupled by sharing the underlying process $(x_k)_{k\geq 0}$. We assume that the initial iterates $\theta_0^{[1]}$ and $\theta_0^{[2]}$ may depend on each other and on x_0 , but are independent of $(x_k)_{k\geq 1}$ given x_0 .

It follows from the definition that

$$\theta_{k+1}^{[1]} - \theta_{k+1}^{[2]} = (I + \alpha A(x_k)) \cdot (\theta_k^{[1]} - \theta_k^{[2]}), \qquad k = 0, 1, \dots$$

If we define the shorthand $\omega_k := \theta_k^{[1]} - \theta_k^{[2]}$, then the above equation becomes

$$\omega_{k+1} = (I + \alpha A(x_k)) \cdot \omega_k, \qquad k = 0, 1, \dots$$
(A.10)

, Vol. 1, No. 1, Article . Publication date: June 2023.

Our key observation is that equation (A.10) is a special case of the LSA iteration (3.1) with ω_k as the variable and $b_{\text{max}} = \sup_{x \in \mathcal{X}} \|b(x)\| = 0$. Applying Proposition 6.1 to this LSA iteration, we obtain the following finite-time geometric bound.

COROLLARY A.6. Suppose that α satisfies (4.2). Then, for all $k \geq \tau$, we have

$$\begin{split} W_2^2\Big(\mathcal{L}(\theta_k^{[1]}), \mathcal{L}(\theta_k^{[2]})\Big) &\overset{\text{(i)}}{\leq} \bar{W}_2^2\Big(\mathcal{L}(x_k, \theta_k^{[1]}), \mathcal{L}(x_k, \theta_k^{[2]})\Big) \\ &\overset{\text{(ii)}}{\leq} \mathbb{E}\Big[\big\|\theta_k^{[1]} - \theta_k^{[2]}\big\|^2\Big] \\ &\overset{\text{(iii)}}{\leq} 8 \frac{\gamma_{\max}}{\gamma_{\min}} \left(1 - \frac{0.9\alpha}{\gamma_{\max}}\right)^k \mathbb{E}\Big[\big\|\theta_0^{[1]} - \theta_0^{[2]}\big\|^2\Big]. \end{split}$$

PROOF OF COROLLARY A.6. The inequality (i) follows from the definition of W_2 and \overline{W}_2 . The inequality (ii) holds since the Wasserstein distance is defined by an infimum as in equation (4.1). Inequality (iii) follows from applying Proposition 6.1 with $b_{\text{max}} = 0$ to the LSA iteration (A.10).

With Corollary A.6, we are ready to prove Theorem 4.1 on the convergence of the Markov chain $(x_k, \theta_k)_{k \ge 0}$. Theorem 4.1 has three parts, whose proofs are given in the next three sub-sub-sections.

A.2.2 Part 1: Existence of Limiting Distribution. Note that Corollary A.6 is valid under any joint distribution of initial iterates $(x_0, \theta_0^{[1]}, \theta_0^{[2]})$. Arbitrarily fix the distribution of $(x_0, \theta_0^{[1]})$. Given $(x_0, \theta_0^{[1]})$, we shall judiciously choose the conditional distribution of $\theta_0^{[2]}$ in a way that ensures $(x_k, \theta_k^{[2]}) \stackrel{d}{=} (x_{k+1}, \theta_{k+1}^{[1]})$ for all $k \ge 0$, where $\stackrel{d}{=}$ denotes equality in distribution. Specifically, recall that \hat{P} is the transition probability matrix for the time-reversed Markov chain of $(x_k)_{k\ge 0}$ and that the initial distribution of x_0 is assumed to be the stationary distribution π ; see Sections 3.1 and 3.2. Given x_0 , let x_{-1} be sampled from $\hat{P}(\cdot|x_0)$. Let $\theta_{-1}^{[2]}$ be a random variable which satisfies $\theta_{-1}^{[2]} \stackrel{d}{=} \theta_0^{[1]}$ and is independent of $(x_k)_{k\ge -1}$. Finally, set $\theta_0^{[2]}$ as

$$\theta_0^{[2]} = \theta_{-1}^{[2]} + \alpha \left(A(x_{-1})\theta_{-1}^{[2]} + b(x_{-1}) \right). \tag{A.11}$$

We argue that this initialization has the desired property.

Claim 2. Under the initialization (A.11), we have $(x_k, \theta_k^{[2]}) \stackrel{d}{=} (x_{k+1}, \theta_{k+1}^{[1]})$ for all $k \ge 0$.

PROOF OF CLAIM 2. From standard results on time-reversed Markov chains, we have $(x_k)_{k\geq -1} \stackrel{d}{=} (x_k)_{k\geq 0}$. Since by construction $\theta_{-1}^{[2]} \stackrel{d}{=} \theta_0^{[1]}$ and $\theta_{-1}^{[2]}$ is independent of $(x_k)_{k\geq -1}$, the claim follows from comparing the update rules for $(\theta_k^{[1]})_{k\geq 0}$ and $(\theta_k^{[2]})_{k\geq -1}$ given in equations (A.9) and (A.11).

Using the above claim, we have for all $k \ge \tau$,

$$\begin{split} \bar{W}_{2}^{2} \Big(\mathcal{L} \big(x_{k}, \theta_{k}^{[1]} \big), \mathcal{L} \big(x_{k+1}, \theta_{k+1}^{[1]} \big) \Big) &= \bar{W}_{2}^{2} \Big(\mathcal{L} \big(x_{k}, \theta_{k}^{[1]} \big), \mathcal{L} \big(x_{k}, \theta_{k}^{[2]} \big) \Big) \\ &\leq 8 \frac{\gamma_{\max}}{\gamma_{\min}} \left(1 - \frac{0.9\alpha}{\gamma_{\max}} \right)^{k} \mathbb{E} \big[\| \theta_{0}^{[1]} - \theta_{0}^{[2]} \|^{2} \big], \end{split}$$

, Vol. 1, No. 1, Article . Publication date: June 2023.

where in the second step above we use Corollary A.6. It follows that

$$\sum_{k=0}^{\infty} \bar{W}_{2}^{2} \left(\mathcal{L}(x_{k}, \theta_{k}^{[1]}), \mathcal{L}(x_{k+1}, \theta_{k+1}^{[1]}) \right) \\
\leq \sum_{k=0}^{\tau-1} \bar{W}_{2}^{2} \left(\mathcal{L}(x_{k}, \theta_{k}^{[1]}), \mathcal{L}(x_{k+1}, \theta_{k+1}^{[1]}) \right) + 8 \frac{\gamma_{\max}}{\gamma_{\min}} \sum_{k=\tau}^{\infty} \left(1 - \frac{0.9\alpha}{\gamma_{\max}} \right)^{k} \mathbb{E}[\|\theta_{0}^{[1]} - \theta_{0}^{[2]}\|^{2}] \\
< \infty,$$

where the last step holds since $\frac{0.9\alpha}{\gamma_{\max}} \in (0,1)$ under the assumption (4.2). The inequality above means that $(\mathcal{L}(x_k,\theta_k^{[1]}))_{k\geq 0}$ is a Cauchy sequence in the metric \bar{W}_2 . Since the space $\mathcal{P}_2(X\times\mathbb{R}^d)$ endowed with \bar{W}_2 is a Polish space [Vil09, Theorem 6.18], every Cauchy sequence converges. Furthermore, convergence in Wasserstein distance implies weak convergence [Vil09, Theorem 6.9]. We conclude that the sequence $(\mathcal{L}(x_k,\theta_k^{[1]}))_{k\geq 0}$ converges weakly to a limit $\bar{\mu}\in\mathcal{P}_2(X\times\mathbb{R}^d)$.

We next show that the limit $\bar{\mu}$ is independent of the initial distribution of $\theta_0^{[1]}$. Suppose that another sequence $(x_k, \tilde{\theta}_k^{[1]})_{k \geq 0}$ with a different initial distribution converges to a limit $\tilde{\mu}$, then

$$\bar{W}_{2}(\bar{\mu}, \tilde{\mu}) \leq \bar{W}_{2}\left(\bar{\mu}, \mathcal{L}(x_{k}, \theta_{k}^{[1]})\right) + \bar{W}_{2}\left(\mathcal{L}(x_{k}, \theta_{k}^{[1]}), \mathcal{L}(x_{k}, \tilde{\theta}_{k}^{[1]})\right) + \bar{W}_{2}\left(\mathcal{L}(x_{k}, \tilde{\theta}_{k}^{[1]}), \tilde{\mu}\right) \xrightarrow{k \to \infty} 0, \text{ (A.12)}$$

where the last step holds since $\bar{W}_2(\mathcal{L}(x_k, \theta_k^{[1]}), \mathcal{L}(x_k, \tilde{\theta}_k^{[1]})) \xrightarrow{k \to \infty} 0$ by Corollary A.6. Therefore, we have $\bar{W}_2(\bar{\mu}, \tilde{\mu}) = 0$ and hence the limit $\bar{\mu}$ is unique.

Finally, the bound on $\text{Var }\theta_{\infty}$ follows from the lemma below, for which we recall that the number κ is defined in Proposition 6.1.

LEMMA A.7. Under Assumptions 1, 2 and 3, and when α is chosen according to (4.2), we have

$$\operatorname{Var}(\theta_{\infty}) \le \mathbb{E}[\|\theta_{\infty} - \theta^*\|^2] \le \frac{\gamma_{\max}}{0.9\gamma_{\min}} \cdot \alpha \tau \kappa \tag{A.13}$$

and

$$\left(\mathbb{E}[\|\theta_{\infty}\|]\right)^{2} \le \mathbb{E}[\|\theta_{\infty}\|^{2}] \le C(A, b, \pi) \tag{A.14}$$

for some $C(A, b, \pi)$ that is independent of α .

Proof of Lemma A.7. We have shown that the sequence $(\theta_k)_{k\geq 0}$ converges weakly to θ_∞ in $\mathcal{P}_2(\mathbb{R}^d)$. It is known that weak convergence in $\mathcal{P}_2(\mathbb{R}^d)$ is equivalent to convergence in distribution and the convergence of the first two moments [Vil09, Definition 6.8]. Consequently, we have

$$\mathbb{E}[\|\theta_{\infty} - \theta^*\|^2] = \lim_{k \to \infty} \mathbb{E}[\|\theta_k - \theta^*\|^2]. \tag{A.15}$$

Proposition 6.1 ensures that

$$\mathbb{E}[\|\theta_k - \theta^*\|^2] \leq 8 \frac{\gamma_{\max}}{\gamma_{\min}} \left(1 - \frac{0.9\alpha}{\gamma_{\max}}\right)^k \left(\mathbb{E}[\|\theta_0 - \theta^*\|^2] + s_{\min}^{-1}(\bar{A})b_{\max}\right) + \frac{\gamma_{\max}}{0.9\gamma_{\min}} \cdot \alpha\tau \cdot \kappa.$$

Taking $k \to \infty$ and combining with equation (A.15) gives

$$\mathbb{E}[\|\theta_{\infty} - \theta^*\|^2] \le \frac{\gamma_{\text{max}}}{0.9\gamma_{\text{min}}} \cdot \alpha\tau\kappa \stackrel{(i)}{\le} \frac{\gamma_{\text{max}}}{0.9\gamma_{\text{min}}} \cdot \frac{1}{4}\kappa,$$

where step (i) holds since $\alpha \tau \leq \frac{1}{4}$. Equation (A.13) follows from the above inequality together with the fact that θ^* is a deterministic quantity.

Furthermore, we have

$$(\mathbb{E}[\|\theta_{\infty}\|])^{2} \leq \mathbb{E}[\|\theta_{\infty}\|^{2}]$$

$$\leq \mathbb{E}[(\|\theta_{\infty} - \theta^{*}\| + \|\theta^{*}\|)^{2}]$$

$$\leq 2\mathbb{E}[\|\theta_{\infty} - \theta^{*}\|^{2}] + 2\|\theta^{*}\|^{2} \leq \frac{\gamma_{\max}}{\gamma_{\min}} \cdot \kappa + 2\|\theta^{*}\|^{2}. \tag{A.16}$$

Equation (A.14) then follows from noting that γ_{\max} , γ_{\min} , κ and θ^* only depend on A, b and π . \Box

We have proved part 1 of Theorem 4.1.

A.2.3 Part 2: Invariance. We next show that $\bar{\mu}$ is the unique invariant distribution. Suppose that the initial distribution of (x_0, θ_0) is $\bar{\mu}$. By the triangle inequality of Wasserstein distance, we have

$$\bar{W}_2(\mathcal{L}(x_1, \theta_1), \bar{\mu}) \le \bar{W}_2(\mathcal{L}(x_1, \theta_1), \mathcal{L}(x_{k+1}, \theta_{k+1})) + W_2(\mathcal{L}(x_{k+1}, \theta_{k+1}), \bar{\mu}). \tag{A.17}$$

We proceed by noting the following lemma, whose proof is given at the end of this sub-sub-section.

LEMMA A.8. Let $(x_k, \theta_k)_{\geq 0}$ and $(x'_k, \theta'_k)_{k\geq 0}$ be two copies of the LSA trajectory (3.1), where $\mathcal{L}(x_0, \theta_0) = \bar{\mu}$ and $\mathcal{L}(x'_0, \theta'_0) \in \mathcal{P}_2(X \times \mathbb{R}^d)$ is arbitrary. Under Assumptions 1, 2 and 3, and when α is chosen according to equation (4.2), we have

$$\bar{W}_{2}^{2}(\mathcal{L}(x_{1},\theta_{1}),\mathcal{L}(x_{1}',\theta_{1}')) \leq \rho_{1} \cdot \bar{W}_{2}^{2}(\mathcal{L}(x_{0},\theta_{0}),\mathcal{L}(x_{0}',\theta_{0}')) + \sqrt{\rho_{2} \cdot \bar{W}_{2}^{2}(\mathcal{L}(x_{0},\theta_{0}),\mathcal{L}(x_{0}',\theta_{0}'))}, \tag{A.18}$$

where the quantities $\rho_1 := 1 + 2(1 + \alpha)^2 + 16\alpha^2 b_{\max}^2 < \infty$ and $\rho_2 := 16\alpha^2 \cdot \mathbb{E}_{\theta_0 \sim \mu} \left[\|\theta_0\|^4 \right] < \infty$ are independent of $\mathcal{L}(x_0', \theta_0')$. In particular, for any $k \geq 0$, if we set $\mathcal{L}(x_0', \theta_0') = \mathcal{L}(x_k, \theta_k)$, then

$$\bar{W}_{2}^{2}(\mathcal{L}(x_{1},\theta_{1}),\mathcal{L}(x_{k+1},\theta_{k+1})) \leq \rho_{1} \cdot \bar{W}_{2}^{2}(\bar{\mu},\mathcal{L}(x_{k},\theta_{k})) + \sqrt{\rho_{2} \cdot \bar{W}_{2}^{2}(\bar{\mu},\mathcal{L}(x_{k},\theta_{k}))}. \tag{A.19}$$

Applying Lemma A.8 to bound the first term on the RHS of equation (A.17), we obtain that

$$\bar{W}_{2}(\mathcal{L}(x_{1},\theta_{1}),\bar{\mu}) \leq \sqrt{\rho_{1} \cdot \bar{W}_{2}(\bar{\mu},\mathcal{L}(x_{k},\theta_{k})) + \sqrt{\rho_{2} \cdot \bar{W}_{2}^{2}(\bar{\mu},\mathcal{L}(x_{k},\theta_{k}))}} + \bar{W}_{2}(\mathcal{L}(x_{k+1},\theta_{k+1}),\bar{\mu})$$

$$\stackrel{k \to \infty}{\longrightarrow} 0,$$

where the last step follows from the weak convergence result established in the last sub-sub-section. We therefore conclude that $W_2(\mathcal{L}(x_1, \theta_1), \bar{\mu}) = 0$ and hence $\bar{\mu}$ is an invariant distribution of the Markov chain $(x_k, \theta_k)_{k \geq 0}$. The uniqueness of the invariant distribution follows from a similar argument as in equation (A.12). We have proved part 2 of Theorem 4.1.

PROOF OF LEMMA A.8. We choose a coupling between the two processes $(x_k, \theta_k)_{k\geq 0}$ and $(x'_k, \theta'_k)_{k\geq 0}$ such that

$$\bar{W}_{2}^{2}(\mathcal{L}(x_{0},\theta_{0}),\mathcal{L}(x_{0}',\theta_{0}')) = \mathbb{E}\left[d_{0}(x_{0},x_{0}') + \|\theta_{0} - \theta_{0}'\|^{2}\right] \quad \text{and}$$
(A.20)

$$x_{k+1} = x'_{k+1}$$
 if $x_k = x'_k$, $\forall k \ge 0$. (A.21)

The existence of a coupling satisfying equation (A.20) at step k=0 is a standard result in optimal transport [Vil09, Theorem 4.1]. We can ensure equation (A.21) by further coupling the two processes for the subsequent steps $k \ge 1$, such that the two underlying Markov chains $(x_k)_{k\ge 0}$ and $(x_k')_{k\ge 0}$ evolve separately (subject to the above coupling at step k=0) until they reach the same state, after which they coalesce and follow the same trajectory.

To prove Lemma A.8, we begin by observing that

$$\bar{W}_{2}^{2}(\mathcal{L}(x_{1},\theta_{1}),\mathcal{L}(x_{1}',\theta_{1}')) \leq \mathbb{E}\left[d_{0}(x_{1},x_{1}') + \|\theta_{1} - \theta_{1}'\|^{2}\right]. \tag{A.22}$$

thanks to the definition (4.1) of \bar{W}_2 using an infimum. Recalling the definition of the discrete metric $d_0(x_0',x_0):=\mathbb{1}\left\{x_0'\neq x_0\right\}$, we have the identities

$$A(x_0) = A(x'_0) + d_0(x'_0, x_0) \cdot (A(x_0) - A(x'_0)) \quad \text{and} \quad b(x_0) = b(x'_0) + d_0(x'_0, x_0) \cdot (b(x_0) - b(x'_0)).$$

The update rule (3.1) together with the above identities implies that

$$\theta_{1} - \theta'_{1} = \theta_{0} + \alpha (A(x_{0})\theta_{0} + b(x_{0})) - \theta'_{0} - \alpha (A(x'_{0})\theta'_{k} + b(x'_{0}))$$

$$= (I + \alpha A(x'_{0})) \cdot (\theta_{0} - \theta'_{k}) + \alpha d_{0}(x'_{0}, x_{0}) \cdot [(A(x_{0}) - A(x'_{0}))\theta_{0} + b(x_{0}) - b(x'_{0})],$$

whence

$$\begin{aligned} \left\| \theta_{1} - \theta_{1}' \right\| &\leq \left\| I + \alpha A(x_{0}') \right\| \cdot \left\| \theta_{0} - \theta_{0}' \right\| + \alpha d_{0}(x_{0}', x_{0}) \cdot \left\| \left(A(x_{0}) - A(x_{0}') \right) \theta_{0} + b(x_{0}) - b(x_{0}') \right\| \\ &\leq (1 + \alpha) \left\| \theta_{0} - \theta_{L}' \right\| + \alpha d_{0}(x_{0}', x_{0}) \cdot 2(\left\| \theta_{0} \right\| + b_{\max}), \end{aligned}$$

where the last step follows from the boundedness Assumption 2. Also note that $d_0(x_1, x_1') \le d_0(x_0, x_0')$ thanks to the coupling in equation (A.21). Combining the above inequalities gives

$$\mathbb{E}\left[d_{0}(x_{1}, x_{1}') + \left\|\theta_{1} - \theta_{1}'\right\|^{2}\right]$$

$$\leq \mathbb{E}\left[d_{0}(x_{0}, x_{0}')\right] + 2(1 + \alpha)^{2} \cdot \mathbb{E}\left[\left\|\theta_{0} - \theta_{0}'\right\|^{2}\right] + 2\alpha^{2} \cdot \mathbb{E}\left[d_{0}(x_{0}', x_{0}) \cdot 8(\|\theta_{0}\|^{2} + b_{\max}^{2})\right]. \tag{A.23}$$

By Cauchy-Schwarz's inequality, we have

$$\mathbb{E}\left[d_0(x_0', x_0) \cdot \|\theta_0\|^2\right] \le \sqrt{\mathbb{E}\left[d_0(x_0', x_0)\right]} \sqrt{\mathbb{E}_{\theta_0 \sim \mu}\left[\|\theta_0\|^4\right]}.$$
(A.24)

Moreover, we claim that

$$\mathbb{E}_{\theta_0 \sim \mu} \left[\|\theta_0\|^4 \right] = \mathbb{E} \left[\|\theta_\infty\|^4 \right] < \infty. \tag{A.25}$$

This claim follows from a moderate tightening of the result in [SY19, Theorem 9], which provides sufficient conditions for the existence of higher moments of θ_{∞} . In Appendix B, we explain how to tighten their result to show that the 4th moment exists under our stepsize condition (4.2).

Combining equations (A.23) and (A.24) and recalling the values of ρ_1 and ρ_2 given in the statement of the lemma, we obtain that

$$\mathbb{E}\left[d_{0}(x_{1}, x_{1}') + \|\theta_{1} - \theta_{1}'\|^{2}\right]$$

$$\leq \rho_{1} \cdot \mathbb{E}\left[d_{0}(x_{0}, x_{0}') + \|\theta_{0} - \theta_{k}'\|^{2}\right] + \sqrt{\rho_{2} \cdot \mathbb{E}\left[d_{0}(x_{0}, x_{0}') + \|\theta_{0} - \theta_{0}'\|^{2}\right]}$$

$$= \rho_{1} \cdot \bar{W}_{2}^{2}\left(\mathcal{L}(x_{0}, \theta_{0}), \mathcal{L}(x_{0}', \theta_{0}')\right) + \sqrt{\rho_{2} \cdot \bar{W}_{2}^{2}\left(\mathcal{L}(x_{0}, \theta_{0}), \mathcal{L}(x_{0}', \theta_{0}')\right)}, \tag{A.26}$$

where the last step from our choice of coupling in equation (A.20). Combining equations (A.22) and (A.26) proves the first equation (A.18) in Lemma A.8. The second equation (A.19) is then immediate. \Box

A.2.4 Part 3: Convergence Rate. We have established that the joint sequence $(\mathcal{L}(x_k, \theta_k^{[1]}))_{k\geq 0}$ converges weakly to the invariant distribution $\bar{\mu} \in \mathcal{P}_2(X \times \mathbb{R}^d)$. Consequently, $(\mathcal{L}(\theta_k^{[1]}))_{k\geq 0}$ converges weakly to $\mu \in \mathcal{P}_2(\mathbb{R}^d)$, where μ is the marginal distribution of $\bar{\mu}$ over \mathbb{R}^d . We now characterize the convergence rate.

Again consider the coupled processes defined in equation (A.9). Suppose that the initial distribution of $(x_0, \theta_0^{[2]})$ is the invariant distribution $\bar{\mu}$, hence $\mathcal{L}(x_k, \theta_k^{[2]}) = \bar{\mu}$ and $\mathcal{L}(\theta_k^{[2]}) = \mu$ for all $k \geq 0$. Applying Corollary A.6, we have for all $k \geq \tau$,

$$\begin{split} W_{2}^{2}(\mathcal{L}(\theta_{k}^{[1]}), \mu) &= W_{2}^{2}(\mathcal{L}(\theta_{k}^{[1]}), \mathcal{L}(\theta_{k}^{[2]})) \\ &\leq \bar{W}_{2}^{2}(\mathcal{L}(x_{k}, \theta_{k}^{[1]}), \mathcal{L}(x_{k}, \theta_{k}^{[2]})) \\ &\leq 8 \frac{\gamma_{\max}}{\gamma_{\min}} \left(1 - \frac{0.9\alpha}{\gamma_{\max}}\right)^{k} \mathbb{E}[\|\theta_{0}^{[1]} - \theta_{0}^{[2]}\|^{2}] \\ &\leq 16 \frac{\gamma_{\max}}{\gamma_{\min}} \left(1 - \frac{0.9\alpha}{\gamma_{\max}}\right)^{k} \cdot \left(\mathbb{E}[\|\theta_{0}^{[1]}\|^{2}] + \mathbb{E}[\|\theta_{\infty}^{[1]}\|^{2}]\right), \end{split}$$

where the last step holds since the chain $(x_k, \theta_k^{[2]})_{k\geq 0}$ is at stationarity and hence $\mathbb{E}\|\theta_0^{[2]}\|^2 = \mathbb{E}\|\theta_\infty^{[1]}\|^2$. This proves equation (4.4) in part 3 of the theorem.

We have completed the proof of Theorem 4.1.

A.3 Proof of Corollary 4.2

PROOF OF COROLLARY 4.2. By Lemma A.7, we have $\mathbb{E}[\|\theta_{\infty}\|^2] = O(1)$. Combining this bound with equation (4.4) in Theorem 4.1, we obtain that for $k \geq \tau$,

$$W_2^2(\mathcal{L}(\theta_k), \mu) \le C(A, b, \pi) \cdot \left(1 - \frac{0.9\alpha}{\gamma_{\text{max}}}\right)^k.$$

By [Vil09, Theorem 4.1], there exists a coupling between θ_k and θ_{∞} such that

$$W_2^2(\mathcal{L}(\theta_k), \mu) = \mathbb{E}[\|\theta_k - \theta_\infty\|^2].$$

Utilizing the above bounds and applying Jensen's inequality twice, we obtain that

$$\begin{split} \|\mathbb{E}[\theta_k - \theta_{\infty}]\|^2 &\leq \left(\mathbb{E}[\|\theta_k - \theta_{\infty}\|]\right)^2 \\ &\leq \mathbb{E}\left[\|\theta_k - \theta_{\infty}\|^2\right] \\ &\leq C(A, b, \pi) \cdot \left(1 - \frac{0.9\alpha}{\gamma_{\max}}\right)^k. \end{split}$$

It thus follows that for $k \geq \tau$,

$$\|\mathbb{E}[\theta_k] - \mathbb{E}[\theta_\infty]\| \le \mathbb{E}[\|\theta_k - \theta_\infty\|] \le C(A, b, \pi) \cdot \left(1 - \frac{0.9\alpha}{\gamma_{\max}}\right)^{k/2},$$

which establishes the convergence rate for the first moment in equation (4.5). Turning to the second moment, we observe that

$$\begin{split} & \left\| \mathbb{E} \left[\theta_{k} \theta_{k}^{\mathsf{T}} \right] - \mathbb{E} \left[\theta_{\infty} \theta_{\infty}^{\mathsf{T}} \right] \right\| \\ &= \left\| \mathbb{E} \left[(\theta_{k} - \theta_{\infty} + \theta_{\infty}) (\theta_{k} - \theta_{\infty} + \theta_{\infty})^{\mathsf{T}} \right] - \mathbb{E} \left[\theta_{\infty} \theta_{\infty}^{\mathsf{T}} \right] \right\| \\ &= \left\| \mathbb{E} \left[(\theta_{k} - \theta_{\infty}) (\theta_{k} - \theta_{\infty})^{\mathsf{T}} \right] + \mathbb{E} \left[\theta_{\infty} (\theta_{k} - \theta_{\infty})^{\mathsf{T}} \right] + \mathbb{E} \left[(\theta_{k} - \theta_{\infty}) \theta_{\infty}^{\mathsf{T}} \right] \right\| \\ &\leq \left\| \mathbb{E} \left[(\theta_{k} - \theta_{\infty}) (\theta_{k} - \theta_{\infty})^{\mathsf{T}} \right] \right\| + \left\| \mathbb{E} \left[\theta_{\infty} (\theta_{k} - \theta_{\infty})^{\mathsf{T}} \right] \right\| + \left\| \mathbb{E} \left[(\theta_{k} - \theta_{\infty}) \theta_{\infty}^{\mathsf{T}} \right] \right\| \\ &\leq \mathbb{E} \left[\left\| (\theta_{k} - \theta_{\infty}) (\theta_{k} - \theta_{\infty})^{\mathsf{T}} \right\| \right] + \mathbb{E} \left[\left\| (\theta_{k} - \theta_{\infty}) \theta_{\infty}^{\mathsf{T}} \right\| \right] \\ &= \mathbb{E} \left[\left\| \theta_{k} - \theta_{\infty} \right\|^{2} \right] + 2 \mathbb{E} \left[\left\| \theta_{\infty}^{\mathsf{T}} (\theta_{k} - \theta_{\infty}) \right\| \right] \\ &\leq \mathbb{E} \left[\left\| \theta_{k} - \theta_{\infty} \right\|^{2} \right] + 2 (\mathbb{E} \left[\left\| \theta_{k} - \theta_{\infty} \right\|^{2} \right] \mathbb{E} \left[\left\| \theta_{\infty} \right\|^{2} \right] \right)^{1/2}, \end{split} \tag{A.27}$$

where the last inequality (A.27) holds true by Cauchy-Schwarz inequality. On the other hand, we have already established that for $k \ge \tau$,

$$\mathbb{E}[\|\theta_k - \theta_\infty\|^2] \le C(A, b, \pi) \left(1 - \frac{0.9\alpha}{\gamma_{\text{max}}}\right)^k \quad \text{and} \quad \mathbb{E}[\|\theta_\infty\|^2] \le C'(A, b, \pi).$$

Substituting the above bounds into the right-hand side of inequality (A.27), we obtain

$$\left\| \mathbb{E} \left[\theta_k \theta_k^{\top} \right] - \mathbb{E} \left[\theta_{\infty} \theta_{\infty}^{\top} \right] \right\| \leq C(A, b, \pi) \cdot \left(1 - \frac{0.9\alpha}{\gamma_{\max}} \right)^{k/2},$$

thereby completing the proof for equation (4.6) in Corollary 4.2.

A.4 Proof of Theorem 4.3

In this sub-section, we prove Theorem 4.3 on characterizing the asymptotic bias of LSA. The proof is divided into four steps, which are given in Appendices A.4.1–A.4.4 to follow.

A.4.1 Step 1: Basic Adjoint Relationship. Following the strategy discussed after Theorem 4.3, we begin by deriving a recursive relationship for the following quantities

$$z_i := \mathbb{E}\left[\theta_{\infty}|x_{\infty}=i\right], \quad i \in \mathcal{X}.$$

To put our derivation in context, we present it in the language of Basic Adjoint Relationship (BAR). Recall that $(x_k)_{k\geq 0}$ is a time-homogeneous Markov chain with transition kernel $P=(p_{ij})$ and unique stationary distribution π . Theorem 4.1 shows that the Markov chain $(x_k,\theta_k)_{k\geq 0}$ also has a unique stationary distribution $\bar{\mu}$, and (x_k,θ_k) converges in distribution to a limit $(x_\infty,\theta_\infty)\sim\bar{\mu}$, where $\theta_\infty\sim\mu$ and $x_\infty\sim\pi$. Given (x_∞,θ_∞) , let $x_{\infty+1}$ be the random variable with conditional distribution $\mathbb{P}(x_{\infty+1}=j|x_\infty=i)=p_{ij}$, and $\theta_{\infty+1}=\theta_\infty+\alpha\left(A(x_\infty)\theta_\infty+b(x_\infty)\right)$; that is, $(x_{\infty+1},\theta_{\infty+1})$ is the state following (x_∞,θ_∞) .

Denote by Q the transition kernel of $(x_k, \theta_k)_{k \ge 0}$. Since $\bar{\mu}$ is invariant for Q, it satisfies the BAR:

$$\bar{\mu}(I-Q)f=0$$

for any test function $f: \mathcal{X} \times \mathbb{R}^d \to \mathbb{R}^d$ that satisfies $||f(x, \theta)|| \le C(1 + ||\theta||^2), \forall (x, \theta)$ for some $C \in \mathbb{R}$ [Vil09, Definition 6.8 and Theorem 6.9]. The above BAR can be written equivalently as

$$\mathbb{E}\left[f(x_{\infty}, \theta_{\infty})\right] = \bar{\mu}f = \bar{\mu}Qf = \mathbb{E}\left[f(x_{\infty+1}, \theta_{\infty+1})\right]. \tag{A.28}$$

It is known that equation (A.28) with a sufficiently large class of test functions f completely characterizes the invariant distribution $\bar{\mu}$ [Har85, HW87, DD11].

It suffices for our purpose to consider the test functions $f^{(i)}$, $i \in X$ defined as

$$f^{(i)}(x,\theta) = \theta \cdot \mathbb{1}\{x = i\}.$$

Substituting $f = f^{(i)}$ into the BAR (A.28) gives

$$\mathbb{E}\left[\theta_{\infty} \cdot \mathbb{1}\left\{x_{\infty} = i\right\}\right] = \mathbb{E}\left[\theta_{\infty+1} \cdot \mathbb{1}\left\{x_{\infty+1} = i\right\}\right]. \tag{A.29}$$

Let us calculate the left and right-hand sides above. Recall that $\hat{P} = (\hat{p}_{ij})$ is the transition kernel of the time-reversal of the Markov chain $(x_k)_{k>0}$; see Section 3.1. We have

$$\mathbb{E}\left[\theta_{\infty} \cdot \mathbb{1}\left\{x_{\infty} = i\right\}\right] = \pi_{i} \cdot \mathbb{E}\left[\theta_{\infty} \middle| x_{\infty} = i\right]$$
$$= \pi_{i} z_{i},$$

and

$$\mathbb{E}\left[\theta_{\infty+1} \cdot \mathbb{I}\left\{x_{\infty+1} = i\right\}\right] = \pi_{i} \cdot \mathbb{E}\left[\theta_{\infty+1} | x_{\infty+1} = i\right]$$

$$= \pi_{i} \cdot \mathbb{E}\left[\theta_{\infty} + \alpha \left(A(x_{\infty})\theta_{\infty} + b(x_{\infty})\right) | x_{\infty+1} = i\right]$$

$$= \pi_{i} \cdot \sum_{s \in \mathcal{X}} \hat{p}_{is} \mathbb{E}\left[\theta_{\infty} + \alpha \left(A(s)\theta_{\infty} + b(s)\right) | x_{\infty} = s, x_{\infty+1} = i\right]$$

$$\stackrel{(i)}{=} \pi_{i} \cdot \sum_{s \in \mathcal{X}} \hat{p}_{is} \mathbb{E}\left[\theta_{\infty} + \alpha \left(A(s)\theta_{\infty} + b(s)\right) | x_{\infty} = s\right]$$

$$= \pi_{i} \cdot \sum_{s \in \mathcal{X}} \hat{p}_{is} \left[z_{s} + \alpha \left(A(s)z_{s} + b(s)\right)\right],$$

where step (i) holds since $\theta_{\infty} \perp \!\!\! \perp x_{\infty+1} | x_{\infty}$ as explained in equation (3.3). Plugging back into the equation (A.29) and noting that $\pi_i > 0$, $\forall i \in \mathcal{X}$, we obtain the recursive relationship:

$$z_{i} = \left(\sum_{s \in \mathcal{X}} \hat{p}_{is} z_{s}\right) + \alpha \left(\sum_{s \in \mathcal{X}} \hat{p}_{is} \left(A(s) z_{s} + b(s)\right)\right), \quad \forall i \in \mathcal{X}.$$
(A.30)

A.4.2 Step 2: Setting up System of Δ . Define the difference $\delta_i := z_i - z_1$ for each $i \in X$, where we arbitrarily take state $1 \in X$ as the reference state. Subtracting z_1 from both sides of equation (A.30) and rearranging terms, we obtain

$$\delta_{i} - \left(\sum_{s \in \mathcal{X}} \hat{p}_{is} \delta_{s}\right) = \alpha \left(\sum_{s \in \mathcal{X}} \hat{p}_{is} (A(s) z_{s} + b(s))\right), \quad \forall i \in \mathcal{X}.$$
(A.31)

We consolidate the variables z_i and δ_i , $i \in \mathcal{X}$ into the matrices Z and Δ , defined as

$$Z = \begin{bmatrix} z_1^\top \\ z_2^\top \\ \vdots \\ z_n^\top \end{bmatrix} \in \mathbb{R}^{n \times d} \quad \text{and} \quad \Delta = \begin{bmatrix} \delta_1^\top \\ \delta_2^\top \\ \vdots \\ \delta_n^\top \end{bmatrix} \in \mathbb{R}^{n \times d}.$$

Let $\vec{\delta} := \text{vec}(\Delta^{\top}) = \begin{bmatrix} \delta_1^{\top} & \delta_2^{\top} & \cdots & \delta_n^{\top} \end{bmatrix}^{\top} \in \mathbb{R}^{nd}$ denote the vectorization of Δ , and similarly $\vec{z} := \text{vec}(Z^{\top}) \in \mathbb{R}^{nd}$ the vectorization of Z. With the above notations, equation (A.31) can be written compactly as

$$((I_n - \hat{P}) \otimes I_d) \vec{\delta} = \alpha (\Psi \vec{z} + \psi), \qquad (A.32)$$

for some matrix $\Psi \equiv \Psi(A, \hat{P}) \in \mathbb{R}^{nd \times nd}$ and vector $\psi \equiv \psi(b, \hat{P}) \in \mathbb{R}^{nd}$, where \otimes denotes the Kronecker product of matrices. Note that the above RHS is a linear function of \vec{z} .

The system (A.32) consists of nd equations. Since $\delta_1 = 0$ by definition, the first d equations are redundant. The remaining (n-1)d equations is given by

$$\left((I_{n-1} - \hat{P}^{(-1)}) \otimes I_d \right) \vec{\delta}^{(-1)} = \alpha \left(\Psi^{(-1)} \vec{z} + \psi^{(-1)} \right), \tag{A.33}$$

where $\hat{P}^{(-1)} \in \mathbb{R}^{(n-1)\times(n-1)}$ is obtained from \hat{P} with the first row and column removed, and $\vec{\delta}^{(-1)} \in \mathbb{R}^{(n-1)d}$ (resp., $\Psi^{(-1)} \in \mathbb{R}^{(n-1)d\times nd}$ and $\psi^{(-1)} \in \mathbb{R}^{(n-1)d}$) is obtained from $\vec{\delta}$ (resp., Ψ and ψ) with the first d rows removed. Note that $\vec{\delta}^{(-1)} = \begin{bmatrix} \delta_2^\top & \cdots & \delta_n^\top \end{bmatrix}^\top$.

We make the following claim.

CLAIM 3. The spectral radius of the matrix $\hat{P}^{(-1)}$, denoted as $\rho(\hat{P}^{(-1)})$, satisfies $\rho(\hat{P}^{(-1)}) < 1$.

PROOF OF CLAIM 3. We have the element-wise inequality

$$\tilde{P}^{(-1)} := \begin{bmatrix} 0 & 0 \\ 0 & \hat{P}^{(-1)} \end{bmatrix} \leq \hat{P}.$$

Wielandt's theorem [Mey00, Chapter 8.3] ensures that $\rho(\tilde{P}^{(-1)}) \leq \rho(\hat{P}) = 1$; moreover, if $\rho(\tilde{P}^{(-1)}) = \rho(\hat{P}) = 1$, then there exists a number ς and a non-singular diagonal matrix D such that $M = e^{i\varsigma}D\hat{P}D^{-1}$, which implies the first row and column of \hat{P} are zero. This however contradicts the irreducibility of \hat{P} , a consequence of the irreducibility assumption of P. We therefore conclude that $\rho(\tilde{P}^{(-1)}) < 1$, which in turn implies that $\rho(\hat{P}^{(-1)}) < 1$.

It follows from the above claim that the matrix $I_{n-1} - \hat{P}^{(-1)}$ is invertible. Consequently, equation (A.33) implies that

$$\vec{\delta}^{(-1)} = \alpha \cdot \left(\left(I_{n-1} - \hat{P}^{(-1)} \right) \otimes I_d \right)^{-1} \cdot \left(\Psi^{(-1)} \vec{z} + \psi^{(-1)} \right)$$

$$= \alpha \cdot \left(\left(I_{n-1} - \hat{P}^{(-1)} \right)^{-1} \otimes I_d \right) \cdot \left(\Psi^{(-1)} \vec{z} + \psi^{(-1)} \right).$$
(A.34)

Equation (A.34) concatenated with $\delta_1 = 0$ can be rewritten compactly as

$$\vec{\delta} = \alpha (B(A, b, P) \cdot \vec{z} + B'(A, b, P)) \tag{A.35}$$

for some matrix $B(A, b, P) \in \mathbb{R}^{nd \times nd}$ and vector $B'(A, b, P) \in \mathbb{R}^{nd}$.

A.4.3 Step 3: Establishing $\Delta = O(\alpha)$. It is tempting to think that the right-hand side of equation (A.35) is linear in α . This is however not the case since \vec{z} , which is a function of θ_{∞} , also depends on α . In what follows, we show that $\vec{\delta} = O(\alpha)$, which is done by first establishing the bounds $\mathbb{E}[\|\theta_{\infty}\|] = O(1)$ and $\vec{z} = O(1)$.

The bound $\mathbb{E}[\|\theta_{\infty}\|] = O(1)$ is established in equation (A.14) in Lemma A.7. We next show that Z = O(1). An explicit bound is given in the following lemma.

LEMMA A.9. Under Assumptions 1, 2 and 3, and when α is chosen according to (4.2), we have

$$||z_i||_2 \leq C_1(A, b, \pi), \quad \forall i \in [n],$$

where

$$C_1(A, b, \pi) = \frac{2}{\pi_{\min}} \cdot \frac{\gamma_{\max}}{\gamma_{\min}} \cdot \left(\sqrt{\kappa} + \|\theta^*\|\right). \tag{A.36}$$

PROOF OF LEMMA A.9. By definition, it holds for each $i \in X$ that

$$z_i = \mathbb{E}\left[\theta_{\infty}|x_{\infty}=i\right] = \frac{\mathbb{E}\left[\theta_{\infty}\mathbb{1}\left\{x_{\infty}=i\right\}\right]}{\pi_i}.$$

It then follows that

$$\begin{aligned} ||z_i|| &\leq \frac{\mathbb{E}[||\theta_{\infty}||]}{\pi_i} \\ &\leq \frac{1}{\pi_{\min}} \cdot \sqrt{\frac{\gamma_{\max}}{\gamma_{\min}} \cdot \kappa + 2||\theta^*||^2} \\ &\leq \frac{2}{\pi_{\min}} \cdot \frac{\gamma_{\max}}{\gamma_{\min}} \cdot \left(\sqrt{\kappa} + ||\theta^*||\right), \end{aligned}$$

where in step (i) we use the bound (A.16) and $\pi_i > 0$, $\forall i \in \mathcal{X}$ by Assumption 1. This completes the proof of Lemma A.9.

Finally, applying Lemma A.9 to upper bound the RHS of equation (A.35), we conclude that

$$\|\vec{\delta}\| \le \alpha \cdot B''(A, b, P) \tag{A.37}$$

for some number B''(A, b, P) that is independent of α .

A.4.4 Step 4: Bootstrapping. We rewrite equation (A.30) as

$$-z_i + \left(\sum_{s=1}^n \hat{p}_{is}z_s\right) + \alpha \left(\sum_{s=1}^n \hat{p}_{is}A(s)z_s\right) = -\alpha \sum_{s=1}^n \hat{p}_{is}b(s), \quad \forall i = 1, \dots, n.$$

We multiply both sides above by π_i and sum over i = 1, ..., n. Simplifying the resulting equation using the identity $\sum_i \pi_i \hat{p}_{is} = \pi_s$, $\forall s$, we obtain

$$\alpha\left(\sum_{s=1}^{n} \pi_s A(s) z_s\right) = -\alpha \sum_{s=1}^{n} \pi_s b(s) = -\alpha \bar{b}.$$

Cancelling $\alpha > 0$ on both sides and substituting $z_i = z_1 + \delta_i$ gives

$$\sum_{s=1}^{n} \pi_{j} A(s)(z_{1} + \delta_{s}) = \bar{A}z_{1} + \sum_{s=1}^{n} \pi_{j} A(s)\delta_{j} = -\bar{b}.$$

Solving the above equation for z_1 and recalling that $\theta^* = -\bar{A}^{-1}\bar{b}$, we obtain

$$z_1 = \theta^* - \bar{A}^{-1} \sum_{s=1}^n \pi_s A(s) \delta_s, \tag{A.38}$$

and consequently

$$z_i = z_1 + \delta_i = \theta^* - \bar{A}^{-1} \sum_{i=1}^{n} \pi_s A(s) \delta_s + \delta_i, \quad \forall i = 1, ..., n.$$
 (A.39)

The above equation can be written in matrix form as

$$\vec{z} = 1_n \otimes \theta^* + C(A, b, \pi) \vec{\delta} \tag{A.40}$$

for some matrix $C(A, b, \pi)$ that is independent α .

Note that equation (A.40) expresses \vec{z} as a linear function of $\vec{\delta}$. On the other hand, equation (A.35) expresses $\vec{\delta}$ as a linear function of \vec{z} . Plugging equation (A.40) into equation (A.35), we obtain the following self-expressing equation for Δ :

$$\vec{\delta} = \alpha \left[B(A, b, P) \cdot \left(1_n \otimes \theta^* + C(A, b, \pi) \vec{\delta} \right) + B'(A, b, P) \right]$$

$$= \alpha \Upsilon_1 + \alpha \Xi \vec{\delta}, \tag{A.41}$$

where the quantities

$$\Upsilon_1 \equiv \Upsilon_1(A, b, P) := B(A, b, P) \cdot (1_n \otimes \theta^*) + B'(A, b, P)$$
 and
$$\Xi \equiv \Xi(A, b, P) := B(A, b, P)C(A, b, \pi)$$

are independent of α , and we use the fact that θ^* is a function of A, b, π . Before proceeding, we record the following explicit upper bound on the spectral norm of Ξ . The proof is given in Appendix A.4.5.

LEMMA A.10. We have

$$\|\Xi\| \le 2n^{3/2} \cdot \iota \cdot (\hat{p}_{\min}^{\iota})^{-1} \cdot s_{\min}^{-1}(\bar{A}),$$

where

$$\iota = \inf\{t : \hat{p}_{ij}^t > 0, \ \forall i, j = 1, \dots, n\}, \quad \hat{p}_{\min}^t = \min_{i=2} \hat{p}_{i1}^k,$$

and \hat{p}_{ij}^t is the (i, j)-th entry of matrix \hat{P}^t .

Using self-expressing equation (A.41) for $\vec{\delta}$, we can bootstrap from the bound $\vec{\delta} = O(\alpha)$ in equation (A.37) to obtain increasingly precise expansions of $\vec{\delta}$ in terms of α . We perform this bootstrapping procedure using induction. Plugging the bound $\Delta = O(\alpha)$ into equation (A.41) gives

$$\vec{\delta} = \alpha \Upsilon_1 + \Xi O(\alpha^2),$$

which establishes our base case. Next, define the vectors $\Upsilon_i := \Xi^{i-1}\Upsilon_1$, $i=1,\ldots$ As our induction hypothesis, suppose that

$$\vec{\delta} = \sum_{i=1}^{m} \alpha^{i} \Upsilon_{i} + \Xi^{m} O(\alpha^{m+1}), \tag{A.42}$$

for an integer $m \ge 1$. Plugging the above expression into equation (A.41) gives

$$\vec{\delta} = \alpha \Upsilon_1 + \alpha \Xi \left(\sum_{i=1}^m \alpha^i \Upsilon_i + \Xi^m O(\alpha^{m+1}) \right) \stackrel{\text{(i)}}{=} \sum_{i=1}^{m+1} \alpha^i \Upsilon_i + \Xi^{m+1} O(\alpha^{m+2}),$$

where step (i) holds since $\Xi \Upsilon_i = \Upsilon_{i+1}$ by definition. The above induction argument establishes that the expansion (A.42) holds for all integer $m \ge 1$.

If it further holds that

$$\alpha < \frac{1}{2\iota} n^{-\frac{3}{2}} \cdot \hat{p}_{\min}^{\iota} s_{\min}(\bar{A}), \tag{A.43}$$

then in light of Lemma A.10, we have $\alpha < \|\Xi\|^{-1}$, which implies $\rho(\alpha\Xi) < 1$, and hence $\|\Xi^m \alpha^{m+1}\| \to 0$ as $m \to \infty$. In this case, taking $m \to \infty$ in equation (A.42) gives the infinite series expansion

$$\vec{\delta} = \sum_{i=1}^{\infty} \alpha^{i} \Upsilon_{i}. \tag{A.44}$$

Finally, we show that the above expansion for $\vec{\delta}$ implies the desired expansion for $\mathbb{E}[\theta_{\infty}]$. By definition of $\{z_i\}$ and $\{\delta_i\}$, we have the expression

$$\mathbb{E}[\theta_{\infty}] = \sum_{s=1}^{n} \pi_{s}(z_{1} + \delta_{s})$$

$$= z_{1} + \sum_{s=1}^{n} \pi_{s}\delta_{s}$$

$$\stackrel{(i)}{=} \theta^{*} + \sum_{s=1}^{n} \pi_{s} \left(-\bar{A}^{-1}A(s) + I\right)\delta_{s},$$
(A.45)

where step (i) follows from the expression (A.38) for z_1 . The above equation can be written as

$$\mathbb{E}[\theta_{\infty}] = \theta^* + C'(A, b, \pi)\vec{\delta} \tag{A.46}$$

for some matrix $C'(A, b, \pi)$. Combining with the expansion (A.42), we obtain

$$\mathbb{E}[\theta_{\infty}] = \theta^* + \sum_{i=1}^m \alpha^i B^{(i)} + C'(A, b, \pi) \Xi^m O(\alpha^{m+1}),$$

where we define the vectors $B^{(i)} \equiv B^{(i)}(A, b, P) = C'(A, b, \pi)\Upsilon_i$ for i = 1, 2, ... Furthermore, combining equation (A.46) with the infinite series expansion (A.44), which is valid under $\alpha < \|\Xi\|^{-1}$,

we have

$$\mathbb{E}[\theta_{\infty}] = \theta^* + \sum_{i=1}^{\infty} \alpha^i B^{(i)}.$$

This completes the proof of Theorem 4.3.

A.4.5 Proof of Lemma A.10. In the proof below, we use $||v||_2$ to denote the vector ℓ_2 norm, making it clear that it is the norm of a vector, while the spectral norm of a matrix M is denoted by ||M|| without the subscript as stated above.

PROOF OF LEMMA A.10. As $\Xi(A, b, P) = B(A, b, P)C(A, b, \pi)$, by sub-multiplicativity of the matrix spectral norm, we have

$$\|\Xi\| \le \|B(A, b, P)\| \cdot \|C(A, b, \pi)\|.$$

Hence, we first try to bound ||B(A, b, P)|| and $||C(A, b, \pi)||$ respectively.

We start with bounding ||B(A, b, P)||. We first attempt to write out B(A, b, P) explicitly. By observing (A.31) and (A.32), we have

$$\begin{bmatrix} (1-\hat{p}_{11})I_d & -\hat{p}_{12}I_d & \cdots & -\hat{p}_{1n}I_d \\ -\hat{p}_{21}I_d & (1-\hat{p}_{22})I_d & \cdots & -\hat{p}_{2n}I_d \\ \vdots & & & & \\ -\hat{p}_{n1}I_d & -\hat{p}_{n2}I_d & \cdots & (1-\hat{p}_{nn})I_d \end{bmatrix} \begin{bmatrix} \delta_1 \\ \delta_2 \\ \vdots \\ \delta_n \end{bmatrix} = \alpha \begin{bmatrix} \hat{p}_{11}A(1) & \hat{p}_{12}A(2) & \cdots & \hat{p}_{1n}A(n) \\ \hat{p}_{21}A(1) & \hat{p}_{22}A(2) & \cdots & \hat{p}_{2n}A(n) \\ \vdots & & & & \\ \hat{p}_{n1}A(1) & \hat{p}_{n2}A(2) & \cdots & \hat{p}_{nn}A(n) \end{bmatrix} \begin{bmatrix} z_1 \\ z_2 \\ \vdots \\ z_n \end{bmatrix} + \psi \\ .$$

Since δ_1 =0 by construction, and $I - \hat{P}^{(-1)}$ is invertible by Claim 3, the above equation becomes

$$\begin{bmatrix} \delta_1 \\ \delta_2 \\ \vdots \\ \delta_n \end{bmatrix} = \alpha \left(\begin{bmatrix} 0 & & & & \\ (I - \hat{P}^{(-1)})^{-1} \otimes I_d \end{pmatrix} \begin{bmatrix} \hat{p}_{21}A(1) & \hat{p}_{22}A(2) & \cdots & \hat{p}_{2n}A(n) \\ \vdots & & & \\ \hat{p}_{n1}A(1) & \hat{p}_{n2}A(2) & \cdots & \hat{p}_{nn}A(n) \end{bmatrix} \begin{bmatrix} z_1 \\ z_2 \\ \vdots \\ z_n \end{bmatrix} + \begin{bmatrix} 0 \\ \psi^{(-1)} \end{bmatrix} \right).$$

As such, by equation (A.35), we are able to write B(A, b, P) explicitly as

$$B(A,b,P) = \begin{bmatrix} 0 & & & \\ (I-\hat{P}^{(-1)})^{-1} \otimes I_d \end{pmatrix} \begin{bmatrix} \hat{p}_{21}A(1) & \hat{p}_{22}A(2) & \cdots & \hat{p}_{2n}A(n) \\ \vdots & & & \\ \hat{p}_{n1}A(1) & \hat{p}_{n2}A(2) & \cdots & \hat{p}_{nn}A(n) \end{bmatrix}.$$

Then, it is easy to see that

$$||B(A,b,P)||^{2} = \left\| \left((I - \hat{P}^{(-1)})^{-1} \otimes I_{d} \right) \begin{bmatrix} \hat{p}_{21}A(1) & \hat{p}_{22}A(2) & \cdots & \hat{p}_{2n}A(n) \\ \vdots & & & \\ \hat{p}_{n1}A(1) & \hat{p}_{n2}A(2) & \cdots & \hat{p}_{nn}A(n) \end{bmatrix} \right\|^{2}$$

$$\leq \left\| (I - \hat{P}^{(-1)})^{-1} \otimes I_{d} \right\|^{2} \left\| \begin{bmatrix} \hat{p}_{21}A(1) & \hat{p}_{22}A(2) & \cdots & \hat{p}_{2n}A(n) \\ \vdots & & & \\ \hat{p}_{n1}A(1) & \hat{p}_{n2}A(2) & \cdots & \hat{p}_{nn}A(n) \end{bmatrix} \right\|^{2}$$

$$\stackrel{(i)}{\leq} \left\| (I - \hat{P}^{(-1)})^{-1} \right\|^{2} \left\| \begin{bmatrix} \hat{p}_{21}A(1) & \hat{p}_{22}A(2) & \cdots & \hat{p}_{2n}A(n) \\ \vdots & & & \\ \hat{p}_{n1}A(1) & \hat{p}_{n2}A(2) & \cdots & \hat{p}_{nn}A(n) \end{bmatrix} \right\|^{2}, \quad (A.47)$$

where (i) holds for

$$\left\| (I - \hat{P}^{(-1)})^{-1} \otimes I_d \right\|^2 = \left\| (I - \hat{P}^{(-1)})^{-1} \right\|^2 \cdot \|I_d\|^2 = \left\| (I - \hat{P}^{(-1)})^{-1} \right\|^2.$$

Hence, to bound ||B(A, b, P)||, we need to obtain an upper bound for each of the two norms on the right-hand side of the inequality (A.47).

We start with bounding the operator norm of the second term of (A.47). We note that for any $i=2,\ldots,n$ and any $v=[v_1^\top v_2^\top \cdots v_n^\top] \in \mathbb{R}^{nd}$, we have

$$\begin{split} \left\| \begin{bmatrix} \hat{p}_{i1}A(1) & \hat{p}_{i2}A(2) & \cdots & \hat{p}_{in}A(n) \end{bmatrix} \begin{bmatrix} v_1 \\ v_2 \\ \vdots \\ v_n \end{bmatrix} \right\|_2^2 &= \left\| \sum_{j=1}^n \hat{p}_{ij}A(j)v_j \right\|_2^2 \\ & \stackrel{\text{(ii)}}{\leq} \sum_{j=1}^n \hat{p}_{ij} \|A(j)v_j\|_2^2 \leq \sum_{j=1}^n \hat{p}_{ij}A_{\max}^2 \|v_j\|_2^2 \\ & \stackrel{\text{(iii)}}{\leq} \sum_{j=1}^n \|v_j\|_2^2 = \|v\|_2^2, \end{split}$$

where we simply apply Jensen's inequality to obtain (ii), and relax p_{ij} and A_{max} to 1 to have (iii). From the above inequalities, we know that

$$\left\| \begin{bmatrix} \hat{p}_{i1}A(1) & \hat{p}_{i2}A(2) & \cdots & \hat{p}_{in}A(n) \end{bmatrix} \right\|^2 \leq 1.$$

Hence, we have the following upper bound,

$$\left\| \begin{bmatrix} \hat{p}_{21}A(1) & \hat{p}_{22}A(2) & \cdots & \hat{p}_{2n}A(n) \\ \vdots & & & \\ \hat{p}_{n1}A(1) & \hat{p}_{n2}A(2) & \cdots & \hat{p}_{nn}A(n) \end{bmatrix} \right\|^{2} \le (n-1) \left(\max_{i=2,\dots,n} \left\| \begin{bmatrix} \hat{p}_{i1}A(1) & \hat{p}_{i2}A(2) & \cdots & \hat{p}_{in}A(n) \end{bmatrix} \right\|^{2} \right)$$

$$\le n-1$$
(A.48)

We bound the operator norm of the first term of (A.47) with the following claim, whose proof is postponed to the end of this sub-sub-section.

CLAIM 4. We have

$$\|(I - \hat{P}^{(-1)})^{-1}\|_{2} \le \sqrt{n} \cdot \iota(\hat{p}_{\min}^{l})^{-1},$$
 (A.49)

where $\iota \geq 0$ such that $\hat{p}_{ij}^{\iota} > 0 \ \forall i,j \ and \ \hat{p}_{\min}^{k} = \min_{i=2,...,n} \hat{p}_{i1}^{k}$.

Combining (A.49) from Claim 4 and (A.48), we obtain

$$||B(A, b, P)|| \le \left(\sqrt{n} \cdot \iota(\hat{p}_{\min}^{\iota})^{-1}\right) \cdot \sqrt{n-1}$$

$$\le n \cdot \iota(\hat{p}_{\min}^{\iota})^{-1}. \tag{A.50}$$

Next, we proceed to obtain an upper bound for $||C(A, b, \pi)||$. Taking a similar approach as bounding ||B(A, b, P)||, we start with giving the explicit formulation of $C(A, b, \pi)$. From (A.39) and

(A.40), we have

$$C(A, b, \pi) = I_{nd} - \begin{bmatrix} \pi_1 \bar{A}^{-1} A(1) & \pi_2 \bar{A}^{-1} A(2) & \cdots & \pi_n \bar{A}^{-1} A(n) \\ \vdots & & & \\ \pi_1 \bar{A}^{-1} A(1) & \pi_2 \bar{A}^{-1} A(2) & \cdots & \pi_n \bar{A}^{-1} A(n) \end{bmatrix}$$
$$= I_{nd} - 1_n \otimes \begin{bmatrix} \pi_1 \bar{A}^{-1} A(1) & \pi_2 \bar{A}^{-1} A(2) & \cdots & \pi_n \bar{A}^{-1} A(n) \end{bmatrix}.$$

Then, an application of the triangle inequality gives us the following upper bound,

$$||C(A, b, \pi)||^{2} = ||I_{nd} - 1_{n} \otimes [\pi_{1}\bar{A}^{-1}A(1) \quad \pi_{2}\bar{A}^{-1}A(2) \quad \cdots \quad \pi_{n}\bar{A}^{-1}A(n)]||^{2}$$

$$\leq (||I_{nd}|| + ||1_{n}||_{2} \cdot ||[\pi_{1}\bar{A}^{-1}A(1) \quad \pi_{2}\bar{A}^{-1}A(2) \quad \cdots \quad \pi_{n}\bar{A}^{-1}A(n)]||)^{2}. \quad (A.51)$$

We observe that for any $v \in \mathbb{R}^{nd}$,

$$\begin{aligned} \left\| \begin{bmatrix} \pi_{1}\bar{A}^{-1}A(1) & \pi_{2}\bar{A}^{-1}A(2) & \cdots & \pi_{n}\bar{A}^{-1}A(n) \end{bmatrix} \begin{bmatrix} v_{1} \\ v_{2} \\ \vdots \\ v_{n} \end{bmatrix} \right\|_{2}^{2} = \left\| \sum_{i=1}^{n} \pi_{i}\bar{A}^{-1}A(i)v_{i} \right\|_{2}^{2} \\ & \stackrel{\text{(iv)}}{\leq} \sum_{i=1}^{n} \pi_{i} \|\bar{A}^{-1}\|^{2} \|A(i)\|^{2} \|v_{i}\|_{2}^{2} \\ & \stackrel{\text{(v)}}{\leq} \left(s_{\min}^{-2}(\bar{A})A_{\max}^{2} \right) \sum_{i=1}^{n} \|v_{i}\|_{2}^{2} \\ & = \left(s_{\min}^{-2}(\bar{A})A_{\max}^{2} \right) \|v\|_{2}^{2}, \end{aligned}$$

where we apply Jensen's inequality to obtain (iv), and we simply relax π_i to 1 to arrive at (v). Hence, we have the following bound,

$$\left\| \begin{bmatrix} \pi_1 \bar{A}^{-1} A(1) & \pi_2 \bar{A}^{-1} A(2) & \cdots & \pi_n \bar{A}^{-1} A(n) \end{bmatrix} \right\| \le s_{\min}^{-1}(\bar{A}) A_{\max} \le s_{\min}^{-1}.$$

We now substitute the above norm upper bound back into (A.51), and we obtain

$$||C(A, b, \pi)|| \le (1 + \sqrt{n}s_{\min}^{-1}(\bar{A})) \le 2\sqrt{n} \cdot s_{\min}^{-1}(\bar{A}),$$
 (A.52)

where the last inequality holds for $n \ge 1$ and $s_{\min}(\bar{A}) < 1$.

Combining (A.50) and (A.52), we are able to conclude with the claimed upper bound for $\|\Xi\|$,

$$\begin{split} \|\Xi\| &\leq \|B(A, b, P)\| \|C(A, b, \pi)\| \\ &\leq \left(n \cdot \iota(\hat{p}_{\min}^{\iota})^{-1}\right) \cdot \left(2\sqrt{n} \cdot s_{\min}^{-1}(\bar{A})\right) \\ &= 2n^{3/2} s_{\min}^{-1}(\bar{A}) \cdot \iota(\hat{p}_{\min}^{\iota})^{-1}, \end{split}$$

thereby completing the proof of Lemma A.10.

Proof of Claim 4. To bound $||(I - \hat{P}^{(-1)})^{-1}||$, we first recall the definition of $\hat{P}^{(-1)}$,

$$\widetilde{P}^{(-1)} = \begin{bmatrix} 0 & 0 \\ 0 & \widehat{P}^{(-1)} \end{bmatrix} \le \widehat{P}.$$

By non-negativity of transition matrix \hat{P} , it is clear that $(\tilde{P}^{(-1)})^k \leq (\hat{P})^k$ for all $k \geq 0$ and hence,

$$(\widetilde{P}^{(-1)})^k = \begin{bmatrix} 0 & 0 \\ 0 & (\widehat{P}^{(-1)})^k \end{bmatrix} \leq (\widehat{P})^k.$$

Given this inequality, it is easy to see that for any i = 2, ..., n,

$$\sum_{i=2}^{n} (\widetilde{p}^{(-1)})_{ij}^{k} \le \sum_{i=2}^{n} \hat{p}_{ij}^{k} < 1,$$

where $(\widetilde{p}^{(-1)})_{ij}^k$ is the (i, j)-th entry of matrix $(\widetilde{P}^{(-1)})^k$, which corresponds to the (i - 1, j - 1)-th entry of matrix $(\widehat{P}^{(-1)})^k$, and as defined earlier, \widehat{p}_{ij}^t is the (i, j)-th entry of matrix \widehat{P}^t .

Next, by the irreducibility and aperiodicity of \hat{P} , there exists some $\iota \geq 0$, such that $\hat{p}^{\iota}_{ij} > 0$ for all i, j. Denote $\hat{p}^k_{\min} = \min_{i=2,...,n} \hat{p}^k_{i1}$, then for any i=2,...,n, and any $k \geq 0$, we have

$$\sum_{i=2}^{n} (\tilde{p}^{(-1)})_{ij}^{k} \le 1 - \hat{p}_{\min}^{k}.$$

When we set $t = \iota$, we therefore obtain

$$\|(\hat{P}^{(-1)})^{\iota}\|_{\infty} \le (1 - \hat{p}_{\min}^{\iota}) < 1.$$

Then, as previously shown in Claim 3 that $\rho(\hat{P}^{(-1)}) < 1$, we can apply Neumann series and obtain

$$\left\| (I - \hat{P}^{(-1)})^{-1} \right\|_{\infty} = \left\| \sum_{k=0}^{\infty} (\hat{P}^{(-1)})^{k} \right\|_{\infty}$$

$$\leq \sum_{k=0}^{\infty} \left\| (\hat{P}^{(-1)})^{k} \right\|_{\infty}$$

$$= \sum_{m=0}^{\infty} \sum_{l=0}^{t-1} \left\| (\hat{P}^{(-1)})^{m_{l}+l} \right\|_{\infty}. \tag{A.53}$$

By the sub-multiplicative property of the $\|\cdot\|_{\infty}$ norm, we further have

$$\left\| (\hat{P}^{(-1)})^{m\iota + l} \right\|_{\infty} \le \left\| (\hat{P}^{(-1)})^{\iota} \right\|_{\infty}^{m} \left\| (\hat{P}^{(-1)})^{l} \right\|_{\infty} \le (1 - \hat{p}_{\min}^{\iota})^{m}.$$

We then substitute it back into (A.53), and obtain

$$\begin{split} \left\| (I - \hat{P}^{(-1)})^{-1} \right\|_{\infty} &\leq \sum_{m=0}^{\infty} \iota \cdot (1 - \hat{p}_{\min}^{\iota})^{m} \\ &= \iota \cdot \sum_{m=0}^{\infty} (1 - \hat{p}_{\min}^{\iota})^{m} = \iota \cdot (\hat{p}_{\min}^{\iota})^{-1}. \end{split}$$

Making use of equivalence of norm, we therefore obtain the following bound on the spectral norm,

$$\left\| (I - \hat{P}^{(-1)})^{-1} \right\|_{2} \le \sqrt{n-1} \cdot \left\| (I - \hat{P}^{(-1)})^{-1} \right\|_{\infty} \le \iota \sqrt{n} \cdot (\hat{p}_{\min}^{\iota})^{-1}.$$

and complete the proof of Claim 4.

A.5 Proof of Theorem 4.4

In this section, we prove Theorem 4.4 on the relationship between the bias and the SLEM of the underlying reversible Markov chain.

A.5.1 Eigendecomposition for reversible Markov chain. Our proof makes use of the Perron-Frobenius theory and eigendecomposition of reversible Markov chains. We briefly review this topic, following [Bré20, Chapter 6, Section 2]. Suppose that the transition kernel $P = (p_{ij}) \in \mathbb{R}^{n \times n}$ is irreducible and reversible w.r.t. the stationary distribution $\pi \in \mathbb{R}^n$, i.e.,

$$\pi_i p_{ij} = \pi_j p_{ij}, \qquad \forall i, j \in [n], \tag{A.54}$$

where π has strictly positive components. Define the diagonal matrix $D = \operatorname{diag}(\pi) \in \mathbb{R}^{n \times n}$. Let $\ell^2(\pi)$ be the real vector space \mathbb{R}^n endowed with the weighted inner product $\langle x, y \rangle_{\pi} = \sum_i x_i y_i \pi_i = x^{\top} Dy$ and weighted norm $\|x\|_{\pi} = \sqrt{\langle x, x \rangle_{\pi}}$. Reversibility is equivalent to self-adjointness in $\ell^2(\pi)$, i.e., $\langle Px, y \rangle_{\pi} = \langle x, Py \rangle_{\pi}$, $\forall x, y$. Similarly define the vector space $\ell^2(\frac{1}{\pi})$.

Under the reversibility condition (A.54), the eigenvalues of P are real and can be ordered as $1=\lambda_1>|\lambda_2|\geq\ldots\geq|\lambda_n|$. Moreover, the corresponding right and and left eigenvectors $(u_i)_{i=1}^n$ and $(v_i)_{i=1}^n$ can be written as $u_i=D^{-\frac{1}{2}}w_i$ and $v_i=D^{\frac{1}{2}}w_i$ for some orthonormal vectors w_1,\ldots,w_n . Moreover, $u_1=1$ and $v_1=\pi$. These vectors satisfy $v_i=Du_i$ and

$$\langle u_i, v_j \rangle = \delta_{ij}, \quad \langle u_i, u_j \rangle_{\pi} = \delta_{ij} \quad \text{and} \quad \langle v_i, v_j \rangle_{\frac{1}{n}} = \delta_{ij}, \quad \forall i, j$$

where $\delta_{ij} := \mathbb{1}\{i=j\}$ is the Dirac delta. Consequently, $\{u_i\}$ and $\{v_j\}$ are, respectively, an orthonormal basis of the spaces $\ell^2(\pi)$ and $\ell^2(\frac{1}{\pi})$. Each vector $y \in \mathbb{R}^n$ can be expressed as $y = \sum_{i=1}^n \beta_i u_i = \sum_{j=1}^n \beta_j' v_j$, for some $\beta = (\beta_i)$ and $\beta' = (\beta_j)$ satisfying $\|\beta\|_2^2 = \|y\|_{\pi}^2$ and $\|\beta'\|_2^2 = \|y\|_{\frac{1}{2}}^2$.

The matrix P admits the eigendecomposition

$$P = U\Lambda V^{\top} = 1\pi^{\top} + \sum_{i=2}^{n} \lambda_i u_i v_i^{\top},$$

where $\Lambda = \operatorname{diag}(\lambda_1, \dots, \lambda_n)$ is a diagonal matrix, and $U \in \mathbb{R}^{n \times n}$ and $V \in \mathbb{R}^{n \times n}$ are matrices with columns $\{u_i\}$ and $\{w_i\}$, respectively. Moreover, if we let $W \in \mathbb{R}^{n \times n}$ be the matrix with columns $\{w_i\}$, then it holds that $WW^{\top} = W^{\top}W = I$, $U = D^{-\frac{1}{2}}W$ and $V = D^{\frac{1}{2}}W$.

We now proceed with the proof of Theorem 4.4, which is divided into three steps given in the next three sub-sub-sections. The proof shares the same notations and a similar high-level strategy as in that of Theorem 4.3. In particular, we characterize the quantities $z_i := \mathbb{E}\left[\theta_{\infty}|x_{\infty}=i\right] \in \mathbb{R}^d$ and $\delta_i := z_i - z_1 \in \mathbb{R}^d$, and track their dependence on the SLEM λ_2 . To simplify notation, we assume below that $\lambda_2 \geq 0$ and hence $|\lambda_2| = \lambda_2$. The proof for the general case of $\lambda_2 < 0$ is similar.

A.5.2 Step 1: Setting up System of Δ . In the proof of Theorem 4.3, we establish that

$$\delta_i = \left(\sum_{s \in [n]} p_{is} \delta_s\right) + \alpha \left(\sum_{s \in [n]} p_{is} \left(A(s) z_s + b(s)\right)\right), \quad \forall i \in [n].$$

Note that under the reversible Markov chain assumption, we have $\hat{P} = P$. Also, recall that the choice of state 1 as the reference state is arbitrary, so we may assume WLOG that $\pi_1 = \pi_{\max} := \max_{i \in [n]} \pi_i$. Let $G := P - 1\pi^\top \in \mathbb{R}^{n \times n}$ denote the gap between the one-step transition kernel and its mixed

version; explicitly, $q_{is} = p_{is} - \pi_s$. The above equation can be rewritten as

$$\delta_{i} = \left(\sum_{s \in [n]} (\pi_{s} + p_{is} - \pi_{s}) \delta_{s}\right) + \alpha \left(\sum_{s \in [n]} (\pi_{s} + p_{is} - \pi_{s}) (A(s)z_{s} + b(s))\right)$$

$$= \sum_{s} \pi_{s} \delta_{s} + \alpha \sum_{s} \pi_{s} (A(s)z_{s} + b(s)) + \sum_{s} g_{is} (\delta_{s} + \alpha A(s)z_{s} + \alpha b(s)), \quad \forall i \in [n],$$

$$= T$$

where $T \in \mathbb{R}^d$ is independent of i . But $\delta_1 = 0$, hence

$$T = -\sum_s g_{1s} \left(\delta_s + \alpha A(s) z_s + \alpha b(s) \right).$$

Combining the last two display equations to cancel out ω and rearranging terms, we obtain

$$\delta_i - \sum_{s} (g_{is} - g_{1s}) \, \delta_s = \sum_{s} (g_{is} - g_{1s}) \left(\alpha A(s) z_s + \alpha b(s) \right), \quad \forall i \in [n].$$

Let $Y \in \mathbb{R}^{n \times d}$ be the matrix whose s-th row is $y_s^{\top} := \alpha (A(s)z_s + b(s))^{\top} \in \mathbb{R}^d$, and recall that $\Delta \in \mathbb{R}^{n \times d}$ has rows $\{\delta_i^{\top}\}$. The above equation can be written compactly as

$$\left[I - \left(G - 1g^{\mathsf{T}}\right)\right] \Delta = \left(G - 1g^{\mathsf{T}}\right) Y,\tag{A.55}$$

where g^{T} is the first row of G.

A.5.3 Step 2: Establishing $\Delta = O(\alpha \frac{\lambda_2}{1-\lambda_2})$. We proceed by bounding the two sides of (A.55). Using the eigendecomposition of P, we have

$$G = P - 1\pi^{\top} = \sum_{i=2}^{n} \lambda_{i} u_{i} v_{i}^{\top},$$

$$1g^{\top} = \sum_{i=2}^{n} \lambda_{i} u_{i}(1) 1 v_{i}^{\top}.$$
(A.56)

Let $\delta(k) \in \mathbb{R}^n$ denote the k-th column of Δ and y(k) denote the k-th column of Y. Using the expressions in (A.56), we establish the following two lemmas, which give a lower bound of the LHS of (A.55) and an upper bound of its RHS. The proofs of these two lemmas are given in Appendix A.5.5.

LEMMA A.11. We have

$$\|[I - (G - 1g^{\mathsf{T}})]x\|_{\pi}^2 \ge \frac{(1 - \lambda_2)^2 \pi_{\max}}{2} \|x\|_{\pi}^2, \quad \forall x \in \mathbb{R}^n.$$

Consequently, for each $k \in [d]$,

$$\|[I - (G - 1g^{\mathsf{T}})] \delta(k)\|_{\pi}^{2} \ge \frac{(1 - \lambda_{2})^{2} \pi_{\max}}{2} \|\delta(k)\|_{\pi}^{2}.$$

Lemma A.12. For each $k \in [d]$, we have

$$\|(G - 1g^{\mathsf{T}}) y(k)\|_{\pi}^{2} \le 2\lambda_{2}^{2} (1 + \pi_{\max}^{-1}) \|y(k)\|_{\pi}^{2}.$$

Applying these two lemmas to the two sides of (A.55), we obtain that

$$\frac{(1-\lambda_2)^2 \pi_{\max}}{2} \|\delta(k)\|_{\pi}^2 \le 2\lambda_2^2 \left(1 + \pi_{\max}^{-1}\right) \|y(k)\|_{\pi}^2, \quad \forall k \in [d]. \tag{A.57}$$

In the sequel, we use the following equivalence relationship between the ℓ^2 and $\ell^2(\pi)$ norms:

$$\sqrt{\pi_{\min}} \|x\|_2 \le \|x\|_{\pi} \le \sqrt{\pi_{\max}} \|x\|_2 \le \|x\|_2$$
,

where $\pi_{\min} := \min_{i \in [N]} \pi_i > 0$. With the above convention and relationship, the bound (A.57) implies that

$$(1 - \lambda_2)^2 \|\delta(k)\|_2^2 \le 8\pi_{\max}^{-1} \lambda_2^2 \|y(k)\|_2^2, \quad \forall k \in [d],$$

whence

$$(1 - \lambda_2)^2 \|\Delta\|_F^2 \le 8\pi_{\max}^{-1} \lambda_2^2 \|Y\|_F^2. \tag{A.58}$$

Recall that the matrix $Y \in \mathbb{R}^{n \times d}$ has rows $y_s^{\top} = \alpha (A(s)z_s + b(s))^{\top}$, which satisfy

$$\|y_s\|_2 \le \alpha \left(A_{\max} \|z_s\|_2 + b_{\max}\right), \quad \forall s \in [n].$$

As shown in Lemma A.9, that

$$||z_s||_2 \le c \cdot C_1(A, b, \pi),$$

so it follows that

$$||Y||_F^2 = \sum_{i=1}^n ||y_i||_2^2 \le n \cdot \alpha^2 (A_{\text{max}} \cdot c \cdot C_1(A, b, \pi) + b_{\text{max}})^2,$$

and

$$||Y||_F \le \alpha \sqrt{n} \cdot (A_{\text{max}} \cdot c \cdot C_1(A, b, \pi) + b_{\text{max}})$$

Combining with (A.58), we get

$$\begin{split} \|\Delta\|_{F} &\leq \sqrt{8\pi_{\max}^{-1}} \cdot \frac{\lambda_{2}}{1 - \lambda_{2}} \|Y\|_{F} \\ &\leq \left(\sqrt{8\pi_{\max}^{-1}} \cdot \frac{\lambda_{2}}{1 - \lambda_{2}}\right) \cdot \left(\alpha\sqrt{n} \cdot (A_{\max} \cdot c \cdot C_{1}(A, b, \pi) + b_{\max})\right) \\ &\leq c \cdot \alpha\sqrt{n} \cdot \frac{1}{\pi_{\max}} \cdot \frac{\lambda_{2}}{1 - \lambda_{2}} \cdot (A_{\max} \cdot C_{1}(A, b, \pi) + b_{\max}) \\ &\stackrel{(i)}{\leq} c \cdot \alpha\sqrt{n} \cdot \frac{1}{\pi_{\max}} \cdot \frac{\lambda_{2}}{1 - \lambda_{2}} \cdot \left(A_{\max} \cdot \frac{2}{\pi_{\min}} \cdot \frac{\gamma_{\max}}{\gamma_{\min}} \cdot (\sqrt{\kappa} + \|\theta^{*}\|) + b_{\max}\right) \\ &\stackrel{(ii)}{\leq} c \cdot \alpha\sqrt{n} \cdot \frac{1}{\pi_{\max}} \cdot \frac{\lambda_{2}}{1 - \lambda_{2}} \cdot \left(\frac{2}{\pi_{\min}} \cdot \frac{\gamma_{\max}}{\gamma_{\min}} \cdot (\sqrt{640\gamma_{\max}s_{\min}^{-2}(\bar{A})b_{\max}^{2}} + s_{\min}^{-1}(\bar{A})b_{\max}) + b_{\max}\right) \\ &\leq c' \cdot \alpha\sqrt{n} \cdot \frac{\lambda_{2}}{1 - \lambda_{2}} \cdot \frac{1}{\pi_{\max}\pi_{\min}} \cdot \frac{\gamma_{\max}}{\gamma_{\min}} \left(\gamma_{\max} + 1\right) s_{\min}^{-1}(\bar{A})b_{\max}, \end{split} \tag{A.59}$$

where we substitute the definition of C_1 in (A.36) to obtain (i), and substitute the definition of κ in (4.3) to obtain (ii).

As shown in (A.59), we are able to establish that

$$\Delta = O\left(\alpha \frac{\lambda_2}{1 - \lambda_2}\right).$$

A.5.4 Step 3: Characterizing the Bias. We can bootstrap from the upper bound (A.59) to obtain a more precise relationship between Δ , α and λ_2 . This is done in the following lemma, whose proof is given in Appendix A.5.5. Note that the proof provides explicit formulas for $B_3(A, b, P)$ and $B_4(A, b, P, \alpha)$; see equation (A.68).

Lemma A.13. We have

$$\Delta = \alpha B_3(A, b, P) + \alpha^2 B_4(A, b, P, \alpha), \tag{A.60}$$

for some vectors $B_3(A, b, P)$ and $B_4(A, b, P, \alpha)$ satisfying

$$\begin{split} \|B_3(A,b,P)\|_2 &\leq c\sqrt{n} \cdot \frac{\lambda_2}{1-\lambda_2} \cdot \left(\frac{1}{\pi_{\min}} \cdot s_{\min}^{-1}(\bar{A}) \cdot b_{\max}\right) \quad and \\ \|B_4(A,b,P,\alpha)\|_2 &\leq c'n \cdot \left(\frac{\lambda_2}{1-\lambda_2}\right)^2 \cdot \left(\frac{1}{\pi_{\max}\pi_{\min}^2} \cdot \frac{\gamma_{\max}}{\gamma_{\min}} \left(\gamma_{\max} + 1\right) s_{\min}^{-1}(\bar{A}) b_{\max}\right), \end{split}$$

where c and c' are universal constants.

We first take the above lemma for granted and complete the proof of Theorem 4.4. The key step of the proof is to relate Δ to the bias $\mathbb{E}[\theta_{\infty}] - \theta^*$. To see the relationship, we first recall equations (A.38) and (A.45), restated below:

$$z_1 = \theta^* - \bar{A}^{-1} \sum_{j=1}^n \pi_j A(j) \delta_j,$$
$$\mathbb{E}[\theta_{\infty}^{\top}] = \pi^{\top} Z = \pi^{\top} (\Delta + \mathbf{1}_n \cdot z_1^{\top}) = z_1^{\top} + \pi^{\top} \Delta.$$

Therefore, substituting Z_1 into the right-hand side of $\mathbb{E}[\theta_{\infty}^{\mathsf{T}}]$, we have

$$\mathbb{E}[\theta_{\infty}] - \theta^* = -\bar{A}^{-1} \sum_{j=1}^n \pi_j A(j) \delta_j + \Delta^\top \pi$$

$$= \sum_{j=1}^n \pi_j (I - \bar{A}^{-1} A(j)) \delta_j$$

$$= \left[\pi_1 (I - \bar{A}^{-1} A(1)) \quad \cdots \quad \pi_n (I - \bar{A}^{-1} A(n)) \right] \cdot \vec{\delta}$$

$$\stackrel{\text{(i)}}{=} \alpha \cdot B(A, b, P) + \alpha^2 \cdot B'(A, b, P, \alpha),$$

where step (i) holds for some appropriate vectors B(A, b, P) and B'(A, b, P) in light of the expression (A.60) in Lemma A.13. This proves equation 4.10 in Theorem 4.4.

Moreover, applying the upper bounds in Lemma A.13, we obtain that

$$\begin{split} \|B(A,b,P)\|_2 &\overset{\text{(ii)}}{\leq} c \left(1 + \frac{A_{\max}}{s_{\min}(\bar{A})}\right) \cdot \|B_3\|_F \\ &\leq c' \left(1 + \frac{A_{\max}}{s_{\min}(\bar{A})}\right) \cdot n \cdot \frac{1}{\pi_{\min}} \cdot s_{\min}^{-1}(\bar{A}) \cdot b_{\max} \\ &\leq c'' \frac{\lambda_2}{1 - \lambda_2} \cdot \frac{n}{\pi_{\min}} \cdot s_{\min}^{-2}(\bar{A}) \cdot b_{\max}, \end{split}$$

where step (ii) follows from a loose bound of the term $\pi_i(I - \bar{A}^{-1}A(i))$.

, Vol. 1, No. 1, Article . Publication date: June 2023.

Similarly, we have

$$\begin{split} \|B'(A,b,P,\alpha)\|_2 &\leq c \left(1 + \frac{A_{\max}}{s_{\min}(\bar{A})}\right) \cdot \|B_4\|_F \\ &\leq c' \left(1 + \frac{A_{\max}}{s_{\min}(\bar{A})}\right) \cdot \left(c'n^{3/2} \cdot \left(\frac{\lambda_2}{1-\lambda_2}\right)^2 \cdot \left(\frac{1}{\pi_{\max}\pi_{\min}^2} \cdot \frac{\gamma_{\max}}{\gamma_{\min}} \left(\gamma_{\max} + 1\right) s_{\min}^{-1}(\bar{A}) b_{\max}\right)\right) \\ &\leq c' \left(1 + \frac{A_{\max}}{s_{\min}(\bar{A})}\right) \cdot \left(c'n^{3/2} \cdot \left(\frac{\lambda_2}{1-\lambda_2}\right)^2 \cdot \left(\frac{1}{\pi_{\max}\pi_{\min}^2} \cdot \frac{\gamma_{\max}}{\gamma_{\min}} \left(\gamma_{\max} + 1\right) s_{\min}^{-1}(\bar{A}) b_{\max}\right)\right) \\ &\leq c'' \cdot \left(\frac{\lambda_2}{1-\lambda_2}\right)^2 \cdot \left(\frac{n^{3/2}}{\pi_{\max}\pi_{\min}^2} \cdot \frac{(1+\gamma_{\max})^2}{\gamma_{\min}} \cdot s_{\min}^{-2}(\bar{A}) b_{\max}\right). \end{split}$$

We have established the bounds (4.11) and (4.12) in Theorem 4.4 and completed the proof thereof.

A.5.5 Proofs of Technical Lemmas. We first present the following simple upper bound, which is needed in the proof of Lemma A.11 and A.12.

LEMMA A.14. It holds that

$$\left(\sum_{i=2}^n \lambda_i \beta_i u_i(1)\right)^2 \le \frac{\lambda_2^2}{\pi_{\max}} \left(\sum_{i=2}^n \beta_i^2\right).$$

PROOF. We have

$$\left(\sum_{i=2}^{n} \lambda_{i} \beta_{i} u_{i}(1)\right)^{2} \leq \lambda_{2}^{2} \left(\sum_{i=2}^{n} |\beta_{i}| |u_{i}(1)|\right)^{2} \qquad |\lambda_{2}| \geq |\lambda_{3}| \geq \ldots \geq |\lambda_{n}|$$

$$\leq \lambda_{2}^{2} \left(\sum_{i=2}^{n} \beta_{i}^{2}\right) \left(\sum_{i=2}^{n} u_{i}(1)^{2}\right). \qquad \text{Cauchy-Schwarz}$$

Note that

$$\begin{split} \sum_{i=2}^{n} u_{i}(1)^{2} &= \sum_{i=2}^{n} e_{1}^{\top} u_{i} u_{i}^{\top} e_{1} \\ &\leq \sum_{i=1}^{n} e_{1}^{\top} u_{i} u_{i}^{\top} e_{1} \\ &= e_{1}^{\top} U U^{\top} e_{1} \\ &= e_{1}^{\top} D^{-\frac{1}{2}} W W^{\top} D^{-\frac{1}{2}} e_{1} & U = D^{-\frac{1}{2}} W W^{\top} e_{1} \\ &= e_{1}^{\top} D^{-1} e_{1} & W W^{\top} = I \\ &= \frac{1}{\pi_{1}}. \end{split}$$

Combining the above bounds and recalling that $\pi_1 = \pi_{\text{max}}$, we complete the proof of Lemma A.14.

PROOF OF LEMMA A.11. Fix an arbitrary vector $x \in \mathbb{R}^n$. Since $\{u_i\}$ is an orthonormal basis of $\ell^2(\pi)$, we can $x = \sum_{j=1}^n \beta_j u_j$ for some $\beta \in \mathbb{R}^n$ satisfying $\|\beta\|_2^2 = \|x\|_{\pi}^2$. Using the expressions of G

and $1q^{T}$ in equation (A.56), we have

$$\begin{split} \left[I - \left(G - 1g^{\top}\right)\right] x &= \left[I - \sum_{i=2}^{n} \lambda_{i} \left(u_{i} - u_{i}(1)1\right) v_{i}^{\top}\right] \left(\sum_{j=1}^{n} \beta_{j} u_{j}\right) \\ &= \sum_{i=1}^{n} \beta_{i} u_{j} - \sum_{i=2}^{n} \lambda_{i} \beta_{i} \left(u_{i} - u_{i}(1)1\right) \\ &= \left(\beta_{1} - \sum_{j=2}^{n} \lambda_{j} \beta_{j} u_{i}(1)\right) 1 + \sum_{i=2}^{n} \beta_{i} (1 - \lambda_{i}) u_{i}. \end{split} \qquad u_{1} = 1 \end{split}$$

Since the vectors $\{1, u_2, \dots, u_n\}$ are orthonormal in $\ell^2(\pi)$, we have

$$\left\| \left[I - \left(G - 1g^{\mathsf{T}} \right) \right] x \right\|_{\pi}^{2} = \left(\beta_{1} - \sum_{i=2}^{n} \lambda_{j} \beta_{j} u_{i}(1) \right)^{2} + \sum_{i=2}^{n} \beta_{i}^{2} (1 - \lambda_{i})^{2}.$$
 (A.61)

Consider two cases:

• If $\beta_1^2 \le \pi_{\max}^{-1} \sum_{i=2}^n \beta_i^2$, then $\|\beta\|_2^2 \le (1 + \pi_{\max}^{-1}) \sum_{i=2}^n \beta_i^2$. Combining with (A.61) gives

$$\begin{aligned} \left\| \left[I - \left(G - 1g^{\top} \right) \right] x \right\|_{\pi}^{2} &\geq (1 - \lambda_{2})^{2} \sum_{i=2}^{n} \beta_{i}^{2} & |\lambda_{2}| \geq |\lambda_{3}| \geq \dots \geq |\lambda_{n}| \\ &\geq (1 - \lambda_{2})^{2} \frac{1}{1 + \pi_{\max}^{-1}} \|\beta\|_{2}^{2} \\ &= \frac{(1 - \lambda_{2})^{2}}{1 + \pi_{\max}^{-1}} \|x\|_{\pi}^{2}. & \|\beta\|_{2}^{2} = \|x\|_{\pi}^{2} \end{aligned}$$

• If $\beta_1^2 > \pi_{\max}^{-1} \sum_{i=2}^n \beta_i^2$, then $\|\beta\|_2^2 < \beta_1^2 + \pi_{\max} \beta_1^2 \le (1 + \pi_{\max}^{-1})\beta_1^2$. It follows that

$$\left| \beta_{1} - \sum_{j=2}^{n} \lambda_{j} \beta_{j} u_{i}(1) \right| \geq |\beta_{1}| - \lambda_{2} \sqrt{\pi_{\max}^{-1} \sum_{i=2}^{n} \beta_{i}^{2}}$$
 Lemma A.14
$$> |\beta_{1}| - \lambda_{2} |\beta_{1}|$$

$$= (1 - \lambda_{2}) |\beta_{1}|$$

$$\geq (1 - \lambda_{2}) \cdot \frac{1}{\sqrt{1 + \pi_{\max}^{-1}}} \|\beta\|_{2}.$$

Combining with (A.61) gives

$$\left\| \left[I - \left(G - 1g^{\mathsf{T}} \right) \right] x \right\|_{\pi}^{2} \ge \frac{(1 - \lambda_{2})^{2}}{1 + \pi_{-1}^{-1}} \left\| \beta \right\|_{2}^{2} = \frac{(1 - \lambda_{2})^{2}}{1 + \pi_{-1}^{-1}} \left\| x \right\|_{\pi}^{2}.$$

Therefore, in both cases we have

$$\|[I - (G - 1g^{\mathsf{T}})] x\|_{\pi}^{2} \ge \frac{(1 - \lambda_{2})^{2} \pi_{\max}}{2} \|x\|_{\pi}^{2},$$

where we use the fact that $\pi_{\max}^{-1} \ge 1$. This proves the first part of Lemma A.11. Taking x to be each column of Δ , we obtain that

$$\|[I - (G - 1g^{\mathsf{T}})] \delta(k)\|_{\pi}^{2} \ge \frac{(1 - \lambda_{2})^{2} \pi_{\max}}{2} \|\delta(k)\|_{\pi}^{2}, \quad \forall k \in [d],$$

which proves the second part of Lemma A.11.

PROOF OF LEMMA A.12. Fix an arbitrary index $k \in [d]$ and recall that y(k) is the k-th column of Y. Using the expression of G and $1g^{\mathsf{T}}$ in equation (A.56), we have

$$(G - 1g^{\mathsf{T}}) y(k) = \sum_{i=2}^{n} \lambda_i (u_i - u_i(1)1) v_i^{\mathsf{T}} y(k).$$
 (A.62)

Since $\{u_i\}$ is an orthogonal basis of $\ell^2(\pi)$, we can write $y(k) = \sum_{j=1}^n \beta_j u_j$ for some $\beta \in \mathbb{R}^n$ satisfying $\|\beta\|_2^2 = \|y(k)\|_{\pi}^2$. Plugging into (A.62) and using the property $v_i^{\mathsf{T}} u_j = \delta_{ij}$, we obtain

$$(G - 1g^{\top}) y_k = \sum_{i=2}^n \sum_{j=1}^n \lambda_i \beta_j (u_i - u_i(1)1) v_i^{\top} u_j$$
$$= \sum_{i=2}^n \lambda_i \beta_i (u_i - u_i(1)1)$$
$$= \sum_{i=2}^n \lambda_i \beta_i u_i - \left(\sum_{i=2}^N \lambda_i \beta_i u_i(1)\right) 1.$$

It follows that

$$\begin{split} \left\| \left(G - 1g^{\mathsf{T}} \right) y_{k} \right\|_{\pi}^{2} &\leq 2 \left\| \sum_{i=2}^{n} \lambda_{i} \beta_{i} u_{i} \right\|_{\pi}^{2} + 2 \left(\sum_{i=2}^{n} \lambda_{i} \beta_{i} u_{i}(1) \right)^{2} \|1\|_{\pi}^{2} \quad \text{triangle inequality, } (a+b)^{2} \leq 2a^{2} + 2b^{2} \\ &= 2 \left(\sum_{i=2}^{n} \lambda_{i}^{2} \beta_{i}^{2} \right) + 2 \left(\sum_{i=2}^{n} \lambda_{i} \beta_{i} u_{i}(1) \right)^{2} \qquad \qquad \ell^{2}(\pi) \text{ orthonormality of } \{u_{i}\} \\ &\leq 2\lambda_{2}^{2} \left(\sum_{i=2}^{n} \beta_{i}^{2} \right) + 2 \cdot \frac{\lambda_{2}^{2}}{\pi_{\text{max}}} \left(\sum_{i=2}^{n} \beta_{i}^{2} \right) \qquad \qquad |\lambda_{2}| \geq |\lambda_{3}| \geq \dots \geq |\lambda_{n}|, \text{ Lemma A.14} \\ &\leq 2\lambda_{2}^{2} \left(1 + \pi_{\text{max}}^{-1} \right) \|y(k)\|_{\pi}^{2}, \qquad \qquad \|\beta\|_{2}^{2} = \|y(k)\|_{\pi}^{2} \end{split}$$

which completes the proof of Lemma A.12.

PROOF OF LEMMA A.13. By Lemma A.11, we know that

$$\begin{split} \pi_{\max} \left\| I - (G - 1g^{\top}) x \right\|_{2}^{2} &\geq \left\| I - (G - 1g^{\top}) x \right\|_{\pi}^{2} \\ &\geq \frac{(1 - \lambda_{2})^{2} \pi_{\max}}{2} \left\| x \right\|_{\pi}^{2} \\ &\geq \frac{(1 - \lambda_{2})^{2} \pi_{\max}}{2} \cdot \pi_{\min} \| x \|_{2}^{2}. \end{split}$$

Canceling out π_{max} on both sides, we obtain

$$||I - (G - 1g^{\mathsf{T}})x||_2^2 \ge \frac{(1 - \lambda_2)^2 \pi_{\min}}{2} ||x||_2^2,$$
 (A.63)

which implies that $s_{\min}(I-(G-1g^{\top})) \ge \frac{(1-\lambda_2)^2\pi_{\min}}{2}$. By Assumption 1, it is clear that $\pi_{\min} > 0$ and $|\lambda_2| < 1$. Hence, $I-(G-1g^{\top})$ is invertible. As such, we can rewrite (A.55) as

$$\Delta = [I - (G - 1g^{\mathsf{T}})]^{-1} (G - 1g^{\mathsf{T}}) Y. \tag{A.64}$$

Then, as $Y \in \mathbb{R}^{n \times d}$ has rows $y_s^{\top} = \alpha (A(s)z_s + b(s))^{\top}$, we substitute it into (A.64), and obtain

$$\Delta = \left[I - \left(G - 1g^{\mathsf{T}} \right) \right]^{-1} \left(G - 1g^{\mathsf{T}} \right) \cdot \alpha \left(\mathbf{M} + \mathbf{b} \right), \tag{A.65}$$

where

$$\mathbf{M} = \begin{bmatrix} (A(1)z_1)^{\top} \\ \vdots \\ (A(n)z_n)^{\top} \end{bmatrix}, \quad \text{and} \quad \mathbf{b} = \begin{bmatrix} b_1^{\top} \\ \vdots \\ b_n^{\top} \end{bmatrix}.$$

Next, we recall that shown in Appendix A.4.4 that

$$z_s = \theta^* - \bar{A}^{-1} \sum_{j=1}^n \pi_j A(j) \delta_j + \delta_s = \theta^* + C^{(s)}(A,b,\pi) \vec{\delta},$$

where

$$C^{(s)}(A, b, \pi) = \begin{bmatrix} \pi_1 A(1) & \pi_2 A(2) & \cdots & \pi_n A(n) \end{bmatrix} + e_s^{\top} \otimes I_d \in \mathbb{R}^{d \times nd}.$$

We then substitute the above definition of z_s into **M** and have

$$\mathbf{M} = \begin{bmatrix} (A(1) \cdot (\theta^* + C^{(1)}(A, b, \pi)\vec{\delta}))^{\top} \\ \vdots \\ (A(n) \cdot (\theta^* + C^{(n)}(A, b, \pi)\vec{\delta}))^{\top} \end{bmatrix} = \begin{bmatrix} (A(1) \cdot \theta^*)^{\top} \\ \vdots \\ (A(n) \cdot \theta^*)^{\top} \end{bmatrix} + \begin{bmatrix} (A(1) \cdot C^{(1)}(A, b, \pi)\vec{\delta})^{\top} \\ \vdots \\ (A(n) \cdot C^{(n)}(A, b, \pi)\vec{\delta})^{\top} \end{bmatrix}. \quad (A.66)$$

We set $B_1(A, b, P) = [I - (G - 1g^{\top})]^{-1} (G - 1g^{\top})$. Together with (A.66), (A.65) becomes

$$\Delta = \alpha B_1(A, b, P) \left(\begin{bmatrix} (A(1) \cdot \theta^* + b_1)^{\mathsf{T}} \\ \vdots \\ (A(n) \cdot \theta^* + b_n)^{\mathsf{T}} \end{bmatrix} + \begin{bmatrix} (A(1) \cdot C^{(1)}(A, b, \pi)\vec{\delta})^{\mathsf{T}} \\ \vdots \\ (A(n) \cdot C^{(n)}(A, b, \pi)\vec{\delta})^{\mathsf{T}} \end{bmatrix} \right). \tag{A.67}$$

We further denote

$$C_4(A,b,\pi) = \begin{bmatrix} (A(1) \cdot \theta^* + b_1)^\top \\ \vdots \\ (A(n) \cdot \theta^* + b_n)^\top \end{bmatrix}, \quad \text{and} \quad B_2(A,b,P,\alpha) = \begin{bmatrix} (A(1) \cdot C^{(1)}(A,b,\pi)\vec{\delta})^\top \\ \vdots \\ (A(n) \cdot C^{(n)}(A,b,\pi)\vec{\delta})^\top \end{bmatrix},$$

and

$$B_3(A, b, P) = B_1(A, b, P) \cdot C_4(A, b, \pi), \text{ and } B_4(A, b, P, \alpha) = B_1(A, b, P) \cdot \frac{B_2(A, b, P, \alpha)}{\alpha}.$$
 (A.68)

As such, Δ in (A.67) can be represented as

$$\Delta = \alpha B_3(A, b, P) + \alpha^2 B_4(A, b, P, \alpha).$$

Therefore, it remains to bound $||B_3(A, b, P)||$ and $||B_4(A, b, P, \alpha)||$.

As B_1 is a present in both B_3 and B_4 , we start with obtaining an upper bound of $||B_1(A, b, P)||$,

$$||B_{1}(A, b, P)||_{2} = ||\left[I - \left(G - 1g^{\mathsf{T}}\right)\right]^{-1} \left(G - 1g^{\mathsf{T}}\right)||_{2}$$

$$\leq ||\left[I - \left(G - 1g^{\mathsf{T}}\right)\right]^{-1}||_{2}||\left(G - 1g^{\mathsf{T}}\right)||_{2}.$$

By (A.63), we first have

$$\left\| \left[I - \left(G - 1g^{\mathsf{T}} \right) \right]^{-1} \right\|_{2} \le \sqrt{\frac{2}{(1 - \lambda_{2})^{2} \pi_{\min}}}$$

Then we know from Lemma A.12,

$$\begin{split} \pi_{\min} \| (G - 1g^{\top}) y(k) \|_{2}^{2} &\leq \| (G - 1g^{\top}) y(k) \|_{\pi}^{2} \\ &\leq 2 \lambda_{2}^{2} (1 + \pi_{\max}^{-1}) \| y(k) \|_{\pi}^{2} \leq 2 \lambda_{2}^{2} (1 + \pi_{\max}) \| y(k) \|_{2}^{2}. \end{split}$$

The inequality above implies the following upper bound on the spectral norm of $||G - 1g^{\top}||_2$,

$$||G - 1g^{\top}||_2 \le \sqrt{\frac{2\lambda_2^2(1 + \pi_{\max})}{\pi_{\min}}}.$$

Hence, we obtain the following bound on $||B_1||$,

$$||B_1(A,b,P)||_2 \le \sqrt{\frac{2}{(1-\lambda_2)^2 \pi_{\min}}} \cdot \sqrt{\frac{2\lambda_2^2 (1+\pi_{\max})}{\pi_{\min}}} \le 4 \cdot \frac{\lambda_2}{1-\lambda_2} \cdot \frac{1}{\pi_{\min}}.$$
 (A.69)

Then, to bound $||B_3||$, we need an upper bound for $||C_4||$. We observe that

$$||C_4(A, b, \pi)||_2^2 \le ||C_4(A, b, \pi)||_F^2 = \sum_{i=1}^n ||A(i)\theta^* + b_i||^2 \le n \left(\frac{A_{\text{max}}}{s_{\text{min}}(\bar{A})} + 1\right)^2 b_{\text{max}}^2.$$
(A.70)

Therefore, (A.69) and (A.70) together give us the following upper bound on $||B_3||$,

$$\begin{split} \|B_{3}(A,b,P)\|_{2} &\leq \|B_{1}(A,b,P)\| \|C_{4}(A,b,\pi)\| \\ &\leq \left(4 \cdot \frac{\lambda_{2}}{1-\lambda_{2}} \cdot \frac{1}{\pi_{\min}}\right) \cdot \left(\sqrt{n} \left(\frac{A_{\max}}{s_{\min}(\bar{A})} + 1\right) b_{\max}\right) \\ &\leq 8\sqrt{n} \cdot \frac{\lambda_{2}}{(1-\lambda_{2})\pi_{\min}} \cdot \frac{A_{\max}}{s_{\min}(\bar{A})} \cdot b_{\max} \\ &\leq c\sqrt{n} \cdot \frac{\lambda_{2}}{(1-\lambda_{2})} \cdot \frac{1}{\pi_{\min}} \cdot s_{\min}^{-1}(\bar{A}) \cdot b_{\max}. \end{split}$$

Lastly, we proceed to bound $||B_4(A, b, P, \alpha)||$, and the key is to understand $||B_2(A, b, P, \alpha)||$. We start with the Frobenius norm of B_2 , and we observe

$$\begin{split} \|B_{2}(A,b,P,\alpha)\|_{F}^{2} &= \sum_{i=1}^{n} \|A(i) \cdot C^{(i)}(A,b,\pi) \operatorname{vec}(\Delta^{\top})\|_{2}^{2} = \sum_{i=1}^{n} \left\|A(i) \left(\left(\sum_{j=1}^{n} \pi_{j} A(j) \delta_{j}\right) + \delta_{i}\right)\right)\right\|^{2} \\ &\leq \sum_{i=1}^{n} A_{\max}^{2} \|\left(\sum_{j=1}^{n} \pi_{j} A(j) \delta_{j}\right) + \delta_{i}\right)\|^{2} \leq 2 \sum_{i=1}^{n} A_{\max}^{2} \left(\left\|\sum_{j=1}^{n} \pi_{j} A(j) \delta_{j}\right\|^{2} + \|\delta_{i}\|^{2}\right) \\ &\leq 2 \sum_{i=1}^{n} A_{\max}^{2} \left(\sum_{j=1}^{n} \pi_{j} \|A(j) \delta_{j}\|^{2} + \|\delta_{i}\|^{2}\right) \\ &\leq 2 \sum_{i=1}^{n} A_{\max}^{2} (A_{\max}^{2} + 1) \|\Delta\|_{F}^{2} \\ &\leq 2 \sum_{i=1}^{n} A_{\max}^{2} (A_{\max}^{2} + 1) \|\Delta\|_{F}^{2} \\ &\leq 2 \sum_{i=1}^{n} A_{\max}^{2} \left(\sum_{j=1}^{n} \pi_{j} \|A(j) \delta_{j}\|^{2} + \|\delta_{i}\|^{2}\right) \\ &\leq 2 \sum_{i=1}^{n} A_{\max}^{2} \left(A_{\max}^{2} + 1\right) \|\Delta\|_{F}^{2} \\ &\leq 2 \sum_{i=1}^{n} A_{\max}^{2} \left(A_{\max}^{2} + 1\right) \|\Delta\|_{F}^{2} \\ &\leq 2 \sum_{i=1}^{n} A_{\max}^{2} \left(A_{\max}^{2} + 1\right) \|A(j) \delta_{j}\|^{2} + \|\delta_{i}\|^{2} \\ &\leq 2 \sum_{i=1}^{n} A_{\max}^{2} \left(A_{\max}^{2} + 1\right) \|\Delta\|_{F}^{2} \\ &\leq 2 \sum_{i=1}^{n} A_{\max}^{2} \left(A_{\max}^{2} + 1\right) \|A(j) \delta_{j}\|^{2} + \|\delta_{i}\|^{2} \\ &\leq 2 \sum_{i=1}^{n} A_{\max}^{2} \left(A_{\max}^{2} + 1\right) \|A(j) \delta_{j}\|^{2} + \|\delta_{i}\|^{2} \\ &\leq 2 \sum_{i=1}^{n} A_{\max}^{2} \left(A_{\max}^{2} + 1\right) \|A(j) \delta_{j}\|^{2} + \|\delta_{i}\|^{2} \\ &\leq 2 \sum_{i=1}^{n} A_{\max}^{2} \left(A_{\max}^{2} + 1\right) \|A(j) \delta_{j}\|^{2} + \|\delta_{i}\|^{2} \\ &\leq 2 \sum_{i=1}^{n} A_{\max}^{2} \left(A_{\max}^{2} + 1\right) \|A(j) \delta_{j}\|^{2} + \|\delta_{i}\|^{2} \\ &\leq 2 \sum_{i=1}^{n} A_{\max}^{2} \left(A_{\max}^{2} + 1\right) \|A(j) \delta_{j}\|^{2} + \|\delta_{i}\|^{2} \\ &\leq 2 \sum_{i=1}^{n} A_{\max}^{2} \left(A_{\max}^{2} + 1\right) \|A(j) \delta_{j}\|^{2} + \|\delta_{i}\|^{2} \\ &\leq 2 \sum_{i=1}^{n} A_{\max}^{2} \left(A_{\max}^{2} + 1\right) \|A(j) \delta_{j}\|^{2} + \|\delta_{i}\|^{2} \\ &\leq 2 \sum_{i=1}^{n} A_{\max}^{2} \left(A_{\max}^{2} + 1\right) \|A(j) \delta_{j}\|^{2} + \|\delta_{i}\|^{2} \\ &\leq 2 \sum_{i=1}^{n} A_{\max}^{2} \left(A_{\max}^{2} + 1\right) \|A(j) \delta_{j}\|^{2} + \|\delta_{i}\|^{2} \\ &\leq 2 \sum_{i=1}^{n} A_{\max}^{2} \left(A_{\max}^{2} + 1\right) \|A(j) \delta_{j}\|^{2} + \|\delta_{i}\|^{2} \\ &\leq 2 \sum_{i=1}^{n} A_{\min}^{2} \left(A_{\min}^{2} + 1\right) \|A(j) \delta_{j}\|^{2} + \|\delta_{i}\|^{2} \\ &\leq 2 \sum_{i=1}^{n} A_{\min}^{2} \left(A_{\min}^{2} + 1\right) \|A(j) \delta_{i}\|^{2} \\ &\leq 2 \sum_{i=1}^{n} A_{\min}^{2} \left(A_{\min}^{2} + 1\right) \|A(j) \delta_{i}\|^{2} \\ &\leq 2 \sum_{i=1}^{n} A_{\min}^{2} \left(A_{\min}^{2} + 1\right) \|A(j) \delta_{i}\|^{2} \\ &\leq 2 \sum_{i=1}^{n} A_{\min}^{2} \left(A_{\min}^{2} + 1\right) \|A(j) \delta_{i}\|^{2}$$

where we make use of the upper bound of $\|\Delta\|_F^2$ in (A.59) to achieve the inequality (i).

Therefore, we conclude that

$$\begin{split} \|B_4(A,b,P,\alpha)\|_2 &\leq \frac{1}{\alpha} \|B_1(A,b,P)\|_2 \|B_2(A,b,P,\alpha)\|_2 \\ &\leq \frac{1}{\alpha} \cdot \left(4 \cdot \frac{\lambda_2}{1-\lambda_2} \cdot \frac{1}{\pi_{\min}}\right) \cdot \left(c \cdot n \cdot \alpha \cdot \left(\frac{\lambda_2}{1-\lambda_2} \cdot \frac{1}{\pi_{\max}\pi_{\min}} \cdot \frac{\gamma_{\max}}{\gamma_{\min}} \left(\gamma_{\max}+1\right) s_{\min}^{-1}(\bar{A}) b_{\max}\right)\right) \\ &\leq c' n \cdot \left(\frac{\lambda_2}{1-\lambda_2}\right)^2 \cdot \left(\frac{1}{\pi_{\max}\pi_{\min}^2} \cdot \frac{\gamma_{\max}}{\gamma_{\min}} \left(\gamma_{\max}+1\right) s_{\min}^{-1}(\bar{A}) b_{\max}\right). \end{split}$$

As such, we have completed the proof of Lemma A.13.

A.6 Proof of Corollary 4.5

We prove the first and second moment bounds in Corollary 4.5.

A.6.1 First Moment. We first have

$$\mathbb{E}[\bar{\theta}_{k_0,k}] - \theta^* = (\mathbb{E}[\theta_\infty] - \theta^*) + \frac{1}{k - k_0} \underbrace{\sum_{t=k_0}^{k-1} \mathbb{E}[\theta_t - \theta_\infty]}_{T_t}.$$

To bound T_1 , we recall (4.5): for $k \ge \tau$,

$$\|\mathbb{E}[\theta_k] - \mathbb{E}[\theta_\infty]\| \le C(A, b, \pi) \cdot \left(1 - \frac{0.9\alpha}{\gamma_{\max}}\right)^{k/2}.$$

As the burn-in period satisfies $k_0 \ge \tau$, we have the following bound,

$$||T_1|| = \left\| \sum_{t=k_0}^{k-1} \mathbb{E}[\theta_t - \theta_\infty] \right\| \le \sum_{t=k_0}^{k-1} ||\mathbb{E}[\theta_t] - \mathbb{E}[\theta_\infty]||$$

$$\le C(A, b, \pi) \left(1 - \frac{0.9\alpha}{\gamma_{\text{max}}} \right)^{k_0/2} \frac{\gamma_{\text{max}}}{0.9\alpha}$$

$$\le C'(A, b, \pi) \cdot \frac{1}{\alpha} \cdot \exp\left(-\frac{\alpha k_0}{4\gamma_{\text{max}}} \right). \tag{A.71}$$

Together with (4.7), we obtain that

$$\mathbb{E}[\bar{\theta}_{k_0,k}] - \theta^* = \alpha B(A,b,P) + O(\alpha^2) + O\left(\frac{1}{\alpha(k-k_0)} \exp\left(-\frac{\alpha k_0}{4\gamma_{\max}}\right)\right),$$

thereby establishing equation (4.13).

, Vol. 1, No. 1, Article . Publication date: June 2023.

A.6.2 Second Moment. Before we move on to obtain a bound for $\mathbb{E}\left[\left(\bar{\theta}_{k_0,k}-\theta^*\right)\left(\bar{\theta}_{k_0,k}-\theta^*\right)^{\top}\right]$, the PR-averaged second moment, we first try to understand $\mathbb{E}\left[\left(\bar{\theta}_{k_0,k}-\mathbb{E}[\theta_{\infty}]\right)(\bar{\theta}_{k_0,k}-\mathbb{E}[\theta_{\infty}])^{\top}\right]$,

$$\mathbb{E}\left[\left(\bar{\theta}_{k_{0},k} - \mathbb{E}[\theta_{\infty}]\right)\left(\bar{\theta}_{k_{0},k} - \mathbb{E}[\theta_{\infty}]\right)^{\top}\right]$$

$$= \frac{1}{(k-k_{0})^{2}} \mathbb{E}\left[\left(\sum_{t=k_{0}}^{k-1} \theta_{k_{0},k} - \mathbb{E}[\theta_{\infty}]\right)\left(\sum_{t=k_{0}}^{k-1} \theta_{k_{0},k} - \mathbb{E}[\theta_{\infty}]\right)^{\top}\right]$$

$$= \underbrace{\frac{1}{(k-k_{0})^{2}} \sum_{t=k_{0}}^{k-1} \mathbb{E}\left[\left(\theta_{t} - \mathbb{E}[\theta_{\infty}]\right)\left(\theta_{t} - \mathbb{E}[\theta_{\infty}]\right)^{\top}\right]}_{T_{1}}$$

$$+ \underbrace{\frac{1}{(k-k_{0})^{2}} \sum_{t=k_{0}}^{k-1} \sum_{l=t+1}^{k-1} \left(\mathbb{E}\left[\left(\theta_{t} - \mathbb{E}[\theta_{\infty}]\right)\left(\theta_{l} - \mathbb{E}[\theta_{\infty}]\right)^{\top}\right] + \mathbb{E}\left[\left(\theta_{l} - \mathbb{E}[\theta_{\infty}]\right)\left(\theta_{t} - \mathbb{E}[\theta_{\infty}]\right)^{\top}\right]}_{T_{2}}$$

Below we control T_1 and T_2 respectively.

For T_1 , we start with the following decomposition,

$$\mathbb{E}\left[\left(\theta_{t} - \mathbb{E}[\theta_{\infty}]\right)\left(\theta_{t} - \mathbb{E}[\theta_{\infty}]\right)^{\top}\right] \\
= \mathbb{E}\left[\theta_{t}\theta_{t}^{\top} - \theta_{t}\mathbb{E}[\theta_{\infty}^{\top}] - \mathbb{E}[\theta_{\infty}]\theta_{t}^{\top} + \mathbb{E}[\theta_{\infty}]\mathbb{E}[\theta_{\infty}^{\top}]\right] \\
= \mathbb{E}[\theta_{t}\theta_{t}^{\top}] - \mathbb{E}[\theta_{t}]\mathbb{E}[\theta_{\infty}^{\top}] - \mathbb{E}[\theta_{\infty}]\mathbb{E}[\theta_{t}^{\top}] + \mathbb{E}[\theta_{\infty}]\mathbb{E}[\theta_{\infty}^{\top}] \\
= \left(\mathbb{E}[\theta_{t}\theta_{t}^{\top}] - \mathbb{E}[\theta_{\infty}\theta_{\infty}^{\top}]\right) + \left(\mathbb{E}[\theta_{\infty}\theta_{\infty}^{\top}] - \mathbb{E}[\theta_{\infty}]\mathbb{E}[\theta_{\infty}^{\top}]\right) - \left(\mathbb{E}[\theta_{t}]\mathbb{E}[\theta_{\infty}^{\top}] + \mathbb{E}[\theta_{\infty}]\mathbb{E}[\theta_{t}^{\top}] - 2\mathbb{E}[\theta_{\infty}]\mathbb{E}[\theta_{\infty}^{\top}]\right) \\
= \left(\mathbb{E}[\theta_{t}\theta_{t}^{\top}] - \mathbb{E}[\theta_{\infty}\theta_{\infty}^{\top}]\right) + \operatorname{Var}(\theta_{\infty}) - \mathbb{E}[\theta_{t} - \theta_{\infty}]\mathbb{E}[\theta_{\infty}^{\top}] - \mathbb{E}[\theta_{\infty}]\mathbb{E}[(\theta_{t} - \theta_{\infty})^{\top}]. \tag{A.72}$$

By Corollary 4.2 and Lemma A.7, the following bounds hold for $t \geq \tau$:

$$\mathbb{E}[\|\theta_{t} - \theta_{\infty}\|] \leq C(A, b, \pi) \cdot \left(1 - \frac{0.9\alpha}{\gamma_{\text{max}}}\right)^{t/2}$$

$$\|\mathbb{E}\left[\theta_{t}\theta_{t}^{\top}\right] - \mathbb{E}\left[\theta_{\infty}\theta_{\infty}^{\top}\right]\| \leq C'(A, b, \pi) \cdot \left(1 - \frac{0.9\alpha}{\gamma_{\text{max}}}\right)^{t/2}$$

$$\mathbb{E}[\|\theta_{\infty}\|] \leq C''(A, b, \pi),$$

$$\operatorname{Var}(\theta_{\infty}) \leq C'''(A, b, \pi) \cdot \alpha\tau.$$
(A.74)

Plugging these bounds into equation (A.72), we obtain that

$$\mathbb{E}\bigg[\left(\theta_t - \mathbb{E}[\theta_{\infty}]\right)\left(\theta_t - \mathbb{E}[\theta_{\infty}]\right)^{\top}\bigg] = O\left(\left(1 - \frac{0.9\alpha}{\gamma_{\max}}\right)^{t/2} + \alpha\tau\right).$$

Hence, we have the following bound for T_1 ,

$$\frac{1}{(k-k_0)^2} \sum_{t=k_0}^{k-1} \mathbb{E} \left[(\theta_t - \mathbb{E}[\theta_\infty]) (\theta_t - \mathbb{E}[\theta_\infty])^\top \right] = \frac{1}{(k-k_0)^2} \sum_{t=k_0}^{k-1} O\left(\left(1 - \frac{0.9\alpha}{\gamma_{\text{max}}} \right)^{t/2} + \alpha \tau \right) \\
= O\left(\frac{1}{(k-k_0)^2} \sum_{t=k_0}^{\infty} \left(1 - \frac{0.9\alpha}{\gamma_{\text{max}}} \right)^{t/2} \right) + O\left(\frac{\alpha \tau}{k-k_0} \right) \\
= O\left(\frac{1}{(k-k_0)^2} \cdot \frac{2\gamma_{\text{max}}}{0.9\alpha} \cdot \left(1 - \frac{0.9\alpha}{\gamma_{\text{max}}} \right)^{k_0/2} \right) + O\left(\frac{\alpha \tau}{k-k_0} \right) \\
= O\left(\frac{1}{\alpha(k-k_0)^2} \cdot \exp\left(-\frac{\alpha k_0}{4\gamma_{\text{max}}} \right) + \frac{\alpha \tau}{k-k_0} \right).$$

For T_2 , we observe that for l > t, we have

$$\begin{split} \mathbb{E}\left[\left(\theta_{t} - \mathbb{E}[\theta_{\infty}]\right)\left(\theta_{l} - \mathbb{E}[\theta_{\infty}]\right)^{\top}\right] &= \mathbb{E}\left[\mathbb{E}\left[\left(\theta_{t} - \mathbb{E}[\theta_{\infty}]\right)\left(\theta_{l} - \mathbb{E}[\theta_{\infty}]\right)^{\top}\middle|\theta_{t}\right]\right] \\ &= \mathbb{E}\left[\left(\theta_{t} - \mathbb{E}[\theta_{\infty}]\right)\mathbb{E}\left[\theta_{l} - \mathbb{E}[\theta_{\infty}]\middle|\theta_{t}\right]^{\top}\right] \\ &= \mathbb{E}\left[\left(\theta_{t} - \mathbb{E}[\theta_{\infty}]\right)\left(\mathbb{E}[\theta_{l}\middle|\theta_{t}\right] - \mathbb{E}[\theta_{\infty}]\right)^{\top}\right]. \end{split}$$

For any $x \in \mathbb{R}^d$, it holds that

$$\left\| \mathbb{E}[\theta_l \middle| \theta_t = x] - \mathbb{E}[\theta_{\infty}] \right\| = \left\| \mathbb{E}[\theta_{l-t} \middle| \theta_0 = x] - \mathbb{E}[\theta_{\infty}] \right\| \stackrel{\text{(ii)}}{\leq} C(A, b, \pi) \cdot \left(1 - \frac{0.9\alpha}{\gamma_{\max}}\right)^{(l-t)/2},$$

where (ii) follows from the first-moment bound in equation (4.5), which is valid for any initial value θ_0 . Hence, when l > t, we have the following inequality,

$$\begin{split} & \mathbb{E}\left[\left\|(\theta_{t} - \mathbb{E}[\theta_{\infty}])\left(\mathbb{E}[\theta_{l}|\theta_{t}] - \mathbb{E}[\theta_{\infty}]\right)^{\top}\right\|\right] \\ = & \mathbb{E}\left[\left\|\theta_{t} - \mathbb{E}[\theta_{\infty}]\right\|\left\|\mathbb{E}[\theta_{l}|\theta_{t}] - \mathbb{E}[\theta_{\infty}]\right\|\right] \\ \leq & \mathbb{E}\left[\left\|\theta_{t} - \mathbb{E}[\theta_{\infty}]\right\|\right] \cdot \left(C(A, b, \pi) \cdot \left(1 - \frac{0.9\alpha}{\gamma_{\max}}\right)^{(l-t)/2}\right) \\ \leq & \left(\mathbb{E}\left[\left\|\theta_{t} - \theta_{\infty}\right\|\right] + 2\mathbb{E}\left[\left\|\theta_{\infty}\right\|\right]\right) \cdot \left(C(A, b, \pi) \cdot \left(1 - \frac{0.9\alpha}{\gamma_{\max}}\right)^{(l-t)/2}\right) \\ \leq & \left(C'(A, b, \pi) \cdot \left(1 - \frac{0.9\alpha}{\gamma_{\max}}\right)^{t/2} + C''(A, b, \pi) \cdot (\alpha\tau_{\alpha})^{1/2}\right) \cdot \left(C(A, b, \pi) \cdot \left(1 - \frac{0.9\alpha}{\gamma_{\max}}\right)^{(l-t)/2}\right) \\ = & C(A, b, \pi) \left(1 - \frac{0.9\alpha}{\gamma_{\max}}\right)^{l/2} + C'(A, b, \pi) \cdot (\alpha\tau_{\alpha})^{1/2} \cdot \left(1 - \frac{0.9\alpha}{\gamma_{\max}}\right)^{(l-t)/2}, \end{split}$$

where in step (iii) we use (A.73) to bound $\mathbb{E}[\|\theta_t - \theta_\infty\|]$ and (A.74) to bound $\mathbb{E}[\|\theta_\infty\|]$, for $\mathbb{E}[\|\theta_\infty\|] \le (\text{Var}(\theta_\infty))^{1/2}$.

, Vol. 1, No. 1, Article . Publication date: June 2023.

Therefore, we have

$$\frac{1}{(k-k_0)^2} \sum_{t=k_0}^{k-1} \sum_{l=t+1}^{k-1} O\left(\left(1 - \frac{0.9\alpha}{\gamma_{\text{max}}}\right)^{l/2}\right) \le \frac{1}{(k-k_0)^2} \sum_{t=k_0}^{\infty} \sum_{l=t+1}^{\infty} O\left(\left(1 - \frac{0.9\alpha}{\gamma_{\text{max}}}\right)^{l/2}\right) \\
\le \frac{1}{(k-k_0)^2} \left(\frac{2\gamma_{\text{max}}}{0.9\alpha}\right)^2 \left(1 - \frac{0.9\alpha}{\gamma_{\text{max}}}\right)^{k_0/2}.$$
(A.75)

and

$$\frac{1}{(k-k_0)^2} \sum_{t=k_0}^{k-1} \sum_{l=t+1}^{k-1} O\left(1 - \frac{0.9\alpha}{\gamma_{\text{max}}}\right)^{(l-t)/2} \le \frac{1}{(k-k_0)^2} \sum_{t=k_0}^{k-1} \sum_{l=t+1}^{\infty} O\left(1 - \frac{0.9\alpha}{\gamma_{\text{max}}}\right)^{(l-t)/2}$$

$$\le \frac{1}{k-k_0} \left(\frac{2\gamma_{\text{max}}}{0.9\alpha}\right). \tag{A.76}$$

Hence, we obtain the following upper bound for T_2 ,

$$\begin{split} T_2 &= \frac{1}{(k-k_0)^2} \sum_{t=k_0}^{k-1} \sum_{l=t+1}^{k-1} O\left(\left(1 - \frac{0.9\alpha}{\gamma_{\text{max}}}\right)^{l/2} + (\alpha \tau_\alpha)^{1/2} \cdot \left(1 - \frac{0.9\alpha}{\gamma_{\text{max}}}\right)^{(l-t)/2} \right) \\ &= O\left(\left(\frac{2\gamma_{\text{max}}}{0.9\alpha(k-k_0)}\right)^2 \left(1 - \frac{0.9\alpha}{\gamma_{\text{max}}}\right)^{k_0/2} \right) + O\left((\alpha \tau_\alpha)^{1/2} \cdot \frac{2\gamma_{\text{max}}}{0.9\alpha(k-k_0)}\right) \\ &= O\left(\frac{1}{\alpha^2(k-k_0)^2} \exp\left(-\frac{\alpha k_0}{4\gamma_{\text{max}}}\right) + \frac{\sqrt{\tau_\alpha/\alpha}}{k-k_0}\right). \end{split}$$

Combining the above bounds for T_1 and T_2 , we obtain

$$\mathbb{E}\left[\left(\bar{\theta}_{k_{0},k} - \mathbb{E}[\theta_{\infty}]\right)\left(\bar{\theta}_{k_{0},k} - \mathbb{E}[\theta_{\infty}]\right)^{\top}\right]$$

$$\leq O\left(\frac{1}{\alpha(k-k_{0})^{2}} \exp\left(-\frac{\alpha k_{0}}{4\gamma_{\max}}\right) + \frac{\alpha \tau}{k-k_{0}}\right) + O\left(\frac{1}{\alpha^{2}(k-k_{0})^{2}} \exp\left(-\frac{\alpha k_{0}}{4\gamma_{\max}}\right) + \frac{\sqrt{\tau_{\alpha}/\alpha}}{k-k_{0}}\right)$$

$$\leq O\left(\frac{\alpha \tau}{k-k_{0}} + \frac{\sqrt{\tau_{\alpha}/\alpha}}{k-k_{0}} + \frac{1}{\alpha^{2}(k-k_{0})^{2}} \exp\left(-\frac{\alpha k_{0}}{4\gamma_{\max}}\right)\right).$$
(A.78)

We are now ready to bound the second moment of the tail-averaged iterate. We make use of the following decomposition:

$$\mathbb{E}\left[\left(\bar{\theta}_{k_{0},k} - \theta^{*}\right)\left(\bar{\theta}_{k_{0},k} - \theta^{*}\right)^{\top}\right]$$

$$=\mathbb{E}\left[\left(\bar{\theta}_{k_{0},k} - \mathbb{E}[\theta_{\infty}] + \mathbb{E}[\theta_{\infty}] - \theta^{*}\right)\left(\bar{\theta}_{k_{0},k} - \mathbb{E}[\theta_{\infty}] + \mathbb{E}[\theta_{\infty}] - \theta^{*}\right)^{\top}\right]$$

$$=\mathbb{E}\left[\left(\bar{\theta}_{k_{0},k} - \mathbb{E}[\theta_{\infty}]\right)\left(\bar{\theta}_{k_{0},k} - \mathbb{E}[\theta_{\infty}]\right)^{\top}\right] + \mathbb{E}\left[\left(\mathbb{E}[\theta_{\infty}] - \theta^{*}\right)\left(\bar{\theta}_{k_{0},k} - \mathbb{E}[\theta_{\infty}]\right)^{\top}\right]$$

$$+ \mathbb{E}\left[\left(\bar{\theta}_{k_{0},k} - \mathbb{E}[\theta_{\infty}]\right)\left(\mathbb{E}[\theta_{\infty}] - \theta^{*}\right)^{\top}\right] + \mathbb{E}\left[\left(\mathbb{E}[\theta_{\infty}] - \theta^{*}\right)\left(\mathbb{E}[\theta_{\infty}] - \theta^{*}\right)^{\top}\right].$$

We can bound the first term on the right-hand side above using equation (A.78). For the other terms on the RHS above, we have

$$\mathbb{E}\left[\left(\bar{\theta}_{k_0,k} - \mathbb{E}[\theta_{\infty}]\right) \left(\mathbb{E}[\theta_{\infty}] - \theta^*\right)^{\top}\right] = \frac{1}{k - k_0} \left(\sum_{t = k_0}^{k - 1} \mathbb{E}\left[\theta_t - \theta_{\infty}\right]\right) \left(\mathbb{E}[\theta_{\infty}] - \theta^*\right)^{\top}$$

$$\stackrel{\text{(iv)}}{=} O\left(\frac{1}{\alpha(k - k_0)} \exp\left(-\frac{\alpha k_0}{4\gamma_{\max}}\right)\right) \left(\alpha B(A, b, P) + O(\alpha^2)\right)$$

$$= O\left(\frac{1}{k - k_0} \exp\left(-\frac{\alpha k_0}{4\gamma_{\max}}\right)\right),$$

where step (iv) is due to equations (A.71) and (4.7), and

$$\mathbb{E}\left[\left(\mathbb{E}[\theta_{\infty}] - \theta^*\right) \left(\mathbb{E}[\theta_{\infty}] - \theta^*\right)^{\top}\right] = \left(\mathbb{E}[\theta_{\infty}] - \theta^*\right) \left(\mathbb{E}[\theta_{\infty}] - \theta^*\right)^{\top}$$

$$\stackrel{(v)}{=} (\alpha B(A, b, P) + O(\alpha^2))(\alpha B(A, b, P) + O(\alpha^2))^{\top}$$

$$= \alpha^2 B'(A, b, P) + O(\alpha^3),$$

where step (v) holds by equation (4.7). Combining all the pieces, we obtain

$$\begin{split} \mathbb{E}\left[\left(\bar{\theta}_{k}-\theta^{*}\right)\left(\bar{\theta}_{k}-\theta^{*}\right)^{\top}\right] = &\alpha^{2}B'(A,b,P) + O(\alpha^{3}) \\ &+O\left(\frac{\alpha\tau_{\alpha}}{k-k_{0}}\right) + O\left(\frac{\sqrt{\tau_{\alpha}/\alpha}}{k-k_{0}}\right) + O\left(\frac{1}{\alpha^{2}(k-k_{0})^{2}}\exp\left(-\frac{\alpha k_{0}}{4\gamma_{\max}}\right)\right). \end{split}$$

Lastly, we note that $\frac{\alpha \tau_{\alpha}}{k - k_0} = O\left(\frac{\sqrt{\tau_{\alpha}/\alpha}}{k - k_0}\right)$, as $\alpha \le \alpha \tau_{\alpha} \le 1$ in light of equations (4.2) and (A.1). As such, we have established the desired equation (4.14),

$$\mathbb{E}\left[\left(\bar{\theta}_{k_0,k} - \theta^*\right)\left(\bar{\theta}_{k_0,k} - \theta^*\right)^{\top}\right] = \alpha^2 B' + O\left(\alpha^3 + \frac{\sqrt{\tau_\alpha/\alpha}}{k - k_0} + \frac{1}{\alpha^2(k - k_0)^2}\exp\left(-\frac{\alpha k_0}{4\gamma_{\max}}\right)\right),$$

and completed the proof of Corollary 4.5.

Remark 2. The above bounds can be improved when $k_0 = k - 1$, which corresponds to the setting without tail averaging and $\bar{\theta}_{k_0,k} = \theta_{k-1}$. In particular, in equation (A.75) and (A.76) above, we upper bound a finite sum by an infinite sum. One can bypass this step when $k_0 = k - 1$, in which case the final second moment bound (4.14) can be improved to

$$\mathbb{E}\left[\left(\theta_{k} - \theta^{*}\right)\left(\theta_{k} - \theta^{*}\right)^{\top}\right] = \alpha^{2}B' + O\left(\alpha^{3} + \alpha\tau_{\alpha} + \frac{1}{\alpha^{2}}\exp\left(-\frac{\alpha k}{4\gamma_{\max}}\right)\right)$$
$$= \alpha^{2}B'' + O\left(\alpha\tau_{\alpha}\right) + O(e^{-\alpha k/(8\gamma_{\max})}).$$

A.7 Proof of Corollary 4.6

We prove the first and second moment bounds in Corollary 4.6.

A.7.1 First Moment. We have

$$\begin{split} \mathbb{E}[\widetilde{\theta}_{k_0,k}^{(\alpha)}] - \theta^* &= \left(2\bar{\theta}_{k_0,k}^{(\alpha)} - \bar{\theta}_{k_0,k}^{(2\alpha)}\right) - \theta^* \\ &= 2\left(\bar{\theta}_{k_0,k}^{(\alpha)} - \theta^*\right) - \left(\bar{\theta}_{k_0,k}^{(2\alpha)} - \theta^*\right) \\ &\stackrel{\text{(i)}}{=} 2\left(\alpha B(A,b,P) + O(\alpha^2) + O\left(\frac{1}{\alpha(k-k_0)}\exp\left(-\frac{\alpha k_0}{4\gamma_{\max}}\right)\right)\right) \\ &- \left(2\alpha B(A,b,P) + O(\alpha^2) + O\left(\frac{1}{\alpha(k-k_0)}\exp\left(-\frac{\alpha k_0}{2\gamma_{\max}}\right)\right)\right) \\ &= O(\alpha^2) + O\left(\frac{1}{\alpha(k-k_0)}\exp\left(-\frac{\alpha k_0}{4\gamma_{\max}}\right)\right), \end{split}$$

where (i) holds following from equation (4.13).

A.7.2 Second Moment. Introduce the following short hands:

$$\begin{split} u_1 &:= \bar{\theta}_{k_0,k}^{(\alpha)} - \mathbb{E}\left[\theta_{\infty}^{(\alpha)}\right], \quad u_2 := \bar{\theta}_{k_0,k}^{(2\alpha)} - \mathbb{E}\left[\theta_{\infty}^{(2\alpha)}\right] \\ \text{and} \quad v := 2\mathbb{E}\left[\theta_{\infty}^{(\alpha)}\right] - \mathbb{E}\left[\theta_{\infty}^{(2\alpha)}\right] + \theta^*. \end{split}$$

With these notations, we write $\widetilde{\theta}_{k_0,k} - \theta^* = 2u_1 - u_2 + v$ and observe the bound

$$\left\| \mathbb{E} \left[\left(\tilde{\theta}_{k_0, k} - \theta^* \right) \left(\tilde{\theta}_{k_0, k} - \theta^* \right)^{\mathsf{T}} \right] \right\| = \left\| \mathbb{E} \left[\left(2u_1 - u_2 + v \right) \left(2u_1 - u_2 + v \right)^{\mathsf{T}} \right] \right\|$$

$$\leq \mathbb{E} \left[\left\| 2u_1 - u_2 + v \right\|^2 \right]$$

$$\leq \mathbb{E} \left\| 2u_1 \right\|^2 + 3\mathbb{E} \left\| u_2 \right\|^2 + 3 \left\| v \right\|^2.$$

By equation (A.78) we have

$$\mathbb{E} \|u_1\|^2 = \operatorname{Tr} \mathbb{E} \left[u_1 u_1^\top \right] = O\left(\frac{\sqrt{\tau_\alpha/\alpha}}{k - k_0} \right) + O\left(\frac{1}{\alpha^2 (k - k_0)^2} \exp\left(-\frac{\alpha k_0}{4 \gamma_{\max}} \right) \right)$$

and similarly,

$$\mathbb{E} \|u_2\|^2 = O\left(\frac{\sqrt{\tau_{2\alpha}/\alpha}}{k-k_0}\right) + O\left(\frac{1}{\alpha^2(k-k_0)^2} \exp\left(-\frac{\alpha k_0}{2\gamma_{\max}}\right)\right).$$

Furthermore, by equation (4.7) we have $||v||^2 = O(\alpha^4)$.

Combining these bounds and noting that $\tau_{2\alpha} \leq \tau_{\alpha}$, we obtain

$$\mathbb{E}\left[\left(\tilde{\theta}_{k-k_0} - \theta^*\right)\left(\tilde{\theta}_{k-k_0} - \theta^*\right)^{\mathsf{T}}\right] = O\left(\frac{\sqrt{\tau_\alpha/\alpha}}{k-k_0}\right) + O\left(\frac{1}{\alpha^2(k-k_0)^2}\exp\left(-\frac{\alpha k_0}{4\gamma_{\max}}\right)\right) + O\left(\alpha^4\right).$$

We have completed the proof of Corollary 4.6.

B EXISTENCE OF HIGHER MOMENTS

The result in [SY19, Theorem 9] provides a sufficient condition for the existence of the m-th moment of the LSA iterates θ_k . Their condition turns out to be more restrictive than necessary. By tightening several intermediate steps in their proof, we can establish the following Proposition B.1, which gives a more relaxed condition. In Appendix B.1 to follow, we explain how to modify the proof of [SY19, Theorem 9] to prove Proposition B.1.

Proposition B.1. Assume the stepsize α satisfies the condition (4.2). Then, for each positive integer m obeying

$$m \cdot \alpha \tau < \frac{1}{4\sqrt{\gamma_{\text{max}}}} \left(\frac{1}{\sqrt{\gamma_{\text{min}}}} + 1\right)^{-1},$$
 (B.1)

it holds for all $k \ge k_m$ that

$$\mathbb{E}[\|\theta_k\|^{2m}] \le (2m-1)!!(c\alpha\tau)^m,$$

where

$$k_m = m\tau + \frac{\tilde{c}}{\alpha} \left(\log \frac{1}{\alpha} \right) \sum_{t=1}^m \frac{1}{t},$$

and both c and \tilde{c} are constants independent of α and m.

In the proof of Theorem 4.1, we make use of the existence of the 4th moment. Taking m = 2 in Proposition B.1, we see that the condition (B.1) becomes

$$\alpha \tau < \frac{1}{8\sqrt{\gamma_{\max}}} \left(\frac{1}{\sqrt{\gamma_{\min}}} + 1\right)^{-1}.$$

Recall our stepsize condition (4.2): $\alpha \tau \leq \frac{0.05}{95 \gamma_{\text{max}}}$. Using the inequality $\gamma_{\text{max}} \geq \gamma_{\text{min}} \geq \frac{1}{2}$ established in equation (A.1), we have

$$\frac{0.05}{95\gamma_{\text{max}}} = \frac{0.05}{95\sqrt{\gamma_{\text{max}}}} \cdot \frac{1}{\sqrt{\gamma_{\text{max}}}} \le \frac{0.1}{95\sqrt{\gamma_{\text{max}}}} \le \frac{1}{32\sqrt{\gamma_{\text{max}}}} < \frac{1}{8\sqrt{\gamma_{\text{max}}}} \left(\frac{1}{\sqrt{\gamma_{\text{min}}}} + 1\right)^{-1}.$$

Therefore, the condition (4.2) implies that the condition (B.1) holds with m=2, which in turn ensures the existence of a finite 4th moment and proves the claim in equation (A.25).

B.1 Proof of Proposition **B.1**

The proof is similar to that of [SY19, Theorem 9]. We only point out the differences. In the proof of [SY19, Theorem 9], the key constraint on $\alpha\tau$ and m that ensures a finite m-th moment arises when bounding $\mathbb{E}[\|\Psi_0\|^{2m}]$, where $\Psi_k = \Gamma^{1/2}\theta_{k+k_{m-1}}$; see [SY19, Appendix D.4]. Below we provide a refinement of the arguments therein.

We start with the following decomposition

$$\|\Psi_{0}\|^{2m} - \|\Psi_{k}\|^{2m} = \sum_{t=0}^{2m-1} \left(\|\Psi_{0}\|^{2m} - t \|\Psi_{k}\|^{t} - \|\Psi_{0}\|^{2m-(t+1)} \|\Psi_{k}\|^{t+1} \right)$$

$$= \sum_{t=0}^{2m-1} \|\Psi_{0}\|^{2m-(t+1)} \|\Psi_{k}\|^{t} \left(\|\Psi_{0}\| - \|\Psi_{k}\| \right). \tag{B.2}$$

Note that

$$\begin{split} \|\Psi_0\| - \|\Psi_k\| &\leq \|\Psi_k - \Psi_0\| \\ &\leq \sqrt{\gamma_{\max}} \|\theta_k - \theta_0\| \\ &\stackrel{\text{(i)}}{\leq} 2\alpha k \sqrt{\gamma_{\max}} (\|\theta_0\| + b_{\max}) \\ &\leq 2\alpha k \sqrt{\gamma_{\max}} \left(\frac{1}{\sqrt{\gamma_{\min}}} \|\Psi_0\| + b_{\max}\right), \end{split}$$

where we make use of Lemma A.2 to obtain the inequality (i). Hence, for the t-th summand on the right-hand side of equation (B.2), we have

$$\begin{split} &\|\Psi_{0}\|^{2m-(t+1)}\|\Psi_{k}\|^{t}\left(\|\Psi_{0}\|-\|\Psi_{k}\|\right) \\ &\leq 2\alpha k \sqrt{\gamma_{\max}}\|\Psi_{0}\|^{2m-(t+1)}\|\Psi_{k}\|^{t}\left(\frac{1}{\sqrt{\gamma_{\min}}}\|\Psi_{0}\|+b_{\max}\right) \\ &\leq 2\alpha k \sqrt{\gamma_{\max}}\left(\frac{1}{\sqrt{\gamma_{\min}}}\|\Psi_{0}\|^{2m-t}\|\Psi_{k}\|^{t}+b_{\max}\|\Psi_{0}\|^{2m-(t+1)}\|\Psi_{k}\|^{t}\right) \\ &\leq 2\alpha k \sqrt{\gamma_{\max}}\left(\frac{1}{\sqrt{\gamma_{\min}}}(\|\Psi_{0}\|^{2m}+\|\Psi_{k}\|^{2m})+b_{\max}(\|\Psi_{0}\|^{2m-1}+\|\Psi_{k}\|^{2m-1})\right). \end{split}$$

We further note the following bound:

$$\frac{1}{\sqrt{\gamma_{\min}}} \|\Psi_{0}\|^{2m} + b_{\max} \|\Psi_{0}\|^{2m-1} = \|\Psi_{0}\|^{2(m-1)} \left(\frac{1}{\sqrt{\gamma_{\min}}} \|\Psi_{0}\|^{2} + b_{\max} \|\Psi_{0}\| \right) \\
\leq \|\Psi_{0}\|^{2(m-1)} \left(\frac{1}{\sqrt{\gamma_{\min}}} \|\Psi_{0}\|^{2} + (b_{\max}^{2} + \|\Psi_{0}\|^{2}) \right) \\
= \|\Psi_{0}\|^{2(m-1)} \left(\left(\frac{1}{\sqrt{\gamma_{\min}}} + 1 \right) \|\Psi_{0}\|^{2} + b_{\max}^{2} \right) \\
= \left(\frac{1}{\sqrt{\gamma_{\min}}} + 1 \right) \|\Psi_{0}\|^{2m} + b_{\max}^{2} \|\Psi_{0}\|^{2(m-1)}. \tag{B.3}$$

Similarly, we have

$$\frac{1}{\sqrt{\gamma_{\min}}} \|\Psi_k\|^{2m} + b_{\max} \|\Psi_k\|^{2m-1} \le \left(\frac{1}{\sqrt{\gamma_{\min}}} + 1\right) \|\Psi_k\|^{2m} + b_{\max}^2 \|\Psi_k\|^{2(m-1)}. \tag{B.4}$$

Combining equations (B.3) and (B.4), the t-th summand on the right-hand side of (B.2) admits the following upper bound:

$$\begin{split} &\|\Psi_0\|^{2m-(t+1)}\|\Psi_k\|^t \left(\|\Psi_0\| - \|\Psi_k\|\right) \\ &\leq & 2\alpha k \sqrt{\gamma_{\max}} \left(\left(\frac{1}{\sqrt{\gamma_{\min}}} + 1\right) \left(\|\Psi_0\|^{2m} + \|\Psi_k\|^{2m}\right) + b_{\max}^2 (\|\Psi_0\|^{2(m-1)} + \|\Psi_k\|^{2(m-1)}\right) \right) \\ &\leq & 2\alpha k \sqrt{\gamma_{\max}} \left(\left(\frac{1}{\sqrt{\gamma_{\min}}} + 1\right) \left(\|\Psi_0\|^{2m} + \|\Psi_k\|^{2m}\right) + b_{\max}^2 (\|\Psi_0\|^{2(m-1)} + \|\Psi_k\|^{2(m-1)}\right) \right). \end{split}$$

Substituting the above bound back into equation (B.2), we have

$$\begin{split} \|\Psi_0\|^{2m} - \|\Psi_k\|^{2m} & \leq 4m\alpha k \sqrt{\gamma_{\max}} \left(\left(\frac{1}{\sqrt{\gamma_{\min}}} + 1 \right) (\|\Psi_0\|^{2m} + \|\Psi_k\|^{2m}) + b_{\max}^2 (\|\Psi_0\|^{2(m-1)} + \|\Psi_k\|^{2(m-1)}) \right). \\ \text{Set } C & \equiv C(A,b,\pi) = 4\sqrt{\gamma_{\max}} \left(\frac{1}{\sqrt{\gamma_{\min}}} + 1 \right) \text{ and } C' \equiv C'(A,b,\pi) = \sqrt{\gamma_{\max}} b_{\max}^2. \text{ We have the inequalities} \\ \|\Psi_0\|^{2m} - \|\Psi_k\|^{2m} & \leq m\alpha k C (\|\Psi_0\|^{2m} + \|\Psi_k\|^{2m}) + m\alpha k C' (\|\Psi_0\|^{2(m-1)} + \|\Psi_k\|^{2(m-1)}), \\ (1 - m\alpha k C) \|\Psi_0\|^{2n} & \leq (1 + m\alpha k C) \|\Psi_k\|^{2m} + m\alpha k C' (\|\Psi_0\|^{2(m-1)} + \|\Psi_k\|^{2(m-1)}), \\ \|\Psi_0\|^{2m} & \leq \frac{1 + m\alpha k C}{1 + m\alpha k C} \|\Psi_k\|^{2m} + \frac{n\alpha k C'}{1 + m\alpha k C} (\|\Psi_0\|^{2(m-1)} + \|\Psi_k\|^{2(m-1)}). \end{split}$$

Therefore, the constraint on m arises as we set $\tau = k$ and require $m\alpha\tau C < 1$. Hence, to ensure a finite m-th moment, we require $m\alpha\tau < \frac{1}{C}$, which corresponds to the condition (B.1) in the statement of Proposition B.1.

C DETAILS FOR NUMERICAL EXPERIMENTS

In this section, we provide the details for the setup of the numerical experiments in Section 5.

C.1 Setup for LSA Experiments

For the experiments on LSA, we generate the transition probability matrix P and functions A and b randomly as follows.

We first illustrate the steps we take to generate the transition matrix P. For a given $n (= |\mathcal{X}|)$, we start with a random matrix $M^{(P)} \in [0,1]^{n \times n}$ with entries $m_{ij}^{(P)} \stackrel{i.i.d.}{\sim} U[0,1]$, and normalize it to obtain a stochastic matrix $\hat{M}^{(P)} = \left(\hat{m}_{ij}^{(P)}\right)$ with $\hat{m}_{ij}^{(P)} = \frac{m_{ij}^{(P)}}{\sum_{k=1}^n m_{ik}^{(P)}}$. We then examine the period and reducibility of the stochastic matrix $\hat{M}^{(P)}$ to ensure that it is aperiodic and irreducible as required in Assumption 1. If $\hat{M}^{(P)}$ is not aperiodic or irreducible, we then repeat the above procedure until we obtain one, and set $P := \hat{M}^{(P)}$. Now with P generated, we compute the stationary distribution π .

Next, we proceed to generate A(x) for $x \in X$. As we also need $\bar{A} = \mathbb{E}_{\pi}[A(x)]$ Hurwitz as required in Assumption 3, we start with generating the Hurwitz matrix \bar{A} and then add noise to obtain the respective A(x). We first generate a random matrix $M^{(A)} \in \mathbb{R}^{d \times d}$ with $m_{ij}^{(A)}$ i.i.d. $\mathcal{N}(0,1)$, and examine the eigenvalues $\lambda_i(M^{(A)})$, as Hurwitz matrix has eigenvalues all with strictly negative real parts. If $\operatorname{Re}(\lambda_i(M^{(A)})) < 0$ for all $i = 1, \ldots, d$, then $M^{(A)}$ is Hurwitz and we set it as $\bar{A} := M^{(A)}$. Otherwise, we adjust $M^{(A)}$ to obtain a Hurwitz matrix, $\bar{A} := M^{(A)} - 2 \max(\operatorname{Re}(\lambda_i(M^{(A)}))) \cdot I_d$. With \bar{A} generated, we add a noise matrix $E(x) \in [-1,1]^{d \times d}$ to \bar{A} to obtain A(x), i.e., $A(x) = \bar{A} + E(x)$. As $\mathbb{E}_{\pi}[E(x)] = 0$, we only generate E(x) with $e(x)_{ij} \stackrel{i.i.d.}{\sim} U[-1,1]$ for $x = 1,\ldots,n-1$, and set $A(n) = \bar{A} - \sum_{x=1}^{n-1} \pi_x E(x)$. Lastly, to align with our assumption, we normalize A(x) by the following procedure,

$$A(x) \leftarrow A(x)/\max_{x} ||A(x)||, \quad \bar{A} \leftarrow \bar{A}/\max_{x} ||A(x)||,$$

to ensure that $A_{\text{max}} := 1$.

Lastly, we generate $b(x) \in \mathbb{R}^d$ with $b(x)_i \overset{i.i.d.}{\sim} [-1,1]$ and obtain $\bar{b} = \sum_x \pi_x b(x)$ and $b_{\max} = \max_x \|b(x)\|$.

C.2 Setup for TD(0) Experiments

We consider the TD(0) algorithm applied to the so-called "problematic MDP" considered in the work [KP00, LP03]. This MDP involves $n^S = 4$ states, $S = \{1, 2, 3, 4\}$, arranged from left to right. At each state, there are two available actions, "Left" (L) and "Right" (R). When the action L is chosen, with probability 0.9 the state transitions to the left (or stays at the same position if the current state is the leftmost state 1), and with probability 0.1 the state transitions in the opposite direction (or stay at the same position if the current state is the rightmost state 4). The dynamics under the action R is defined symmetrically. The reward function is given by r(1) = 0, r(2) = 1, r(3) = 3, r(4) = 0, with a discount factor $\gamma = 0.9$. We consider evaluating the policy that takes the actions R, R, L, and L at states 1, 2, 3, 4, respectively (this policy is the optimal policy for this MDP). The induced MRP is illustrated in Figure 5.

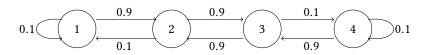


Fig. 5. The Problematic MDP under "RRLL" Policy.

We apply TD(0) with linear function approximation to the above MRP. For each state $s \in \{1, 2, 3, 4\}$, the corresponding d = 3 dimensional feature vector is given by

$$\phi(s) = (1, s, s^2)^{\mathsf{T}},$$

which is used in the work [KP00]. We then normalize each row of the feature matrix $\Phi \in \mathbb{R}^{n^S \times d}$ to have unit ℓ_2 ; explicitly, we set

$$\phi(s)_i \leftarrow \frac{\phi(s)_i}{\sum_{s=1}^4 \phi(s)_i}, \quad i = 1, 2, 3, 4.$$

Note that one may ensure the condition $\max_{s \in \mathcal{S}} \|\phi(s)\| \le \frac{1}{\sqrt{1+\gamma}}$ required by our theory by further rescaling the entire matrix Φ . In our experiments, we ignore this rescaling step, as it is equivalent to simply rescaling the stepsize and iterates.