Optimal and instance-dependent guarantees for Markovian linear stochastic approximation

Wenlong Mou[◇] Ashwin Pananjady[⋆]
Martin J. Wainwright[◇,†] Peter L. Bartlett[◇,†]

Department of Electrical Engineering and Computer Sciences[◇]
Department of Statistics[†]
UC Berkeley

Schools of Industrial & Systems Engineering, and
Electrical & Computer Engineering[⋆]
Georgia Tech

Abstract

We study stochastic approximation procedures for approximately solving a d-dimensional linear fixed point equation based on observing a trajectory of length n from an ergodic Markov chain. We first exhibit a non-asymptotic bound of the order $t_{\mathrm{mix}}\frac{d}{n}$ on the squared error of the last iterate of a standard scheme, where t_{mix} is a mixing time. We then prove a non-asymptotic instance-dependent bound on a suitably averaged sequence of iterates, with a leading term that matches the local asymptotic minimax limit, including sharp dependence on the parameters (d, t_{mix}) in the higher order terms. We complement these upper bounds with a non-asymptotic minimax lower bound that establishes the instance-optimality of the averaged SA estimator. We derive corollaries of these results for policy evaluation with Markov noise—covering the TD(λ) family of algorithms for all $\lambda \in [0, 1)$—and linear autoregressive models. Our instance-dependent characterizations open the door to the design of fine-grained model selection procedures for hyperparameter tuning (e.g., choosing the value of λ when running the TD(λ) algorithm).

1 Introduction

Linear Z-estimation problems—in which we are interested in computing the fixed point of a linear system of equations—are widely used in many application domains, including reinforcement learning and approximate dynamic programming [Ber19, Sze10], stochastic control and filtering [BMP12, Bor09, KY03], and time-series analysis [Ham20]. In many of these applications, the data-generating mechanism is modeled using an underlying Markov chain. The resulting dependency among the observations presents challenges for algorithm design as well as statistical analysis. In this paper, our goal to is provide an instance-dependent statistical analysis—one that captures the difficulty of the particular Z-estimation problem at hand—and to develop computationally efficient algorithms that match these fundamental limits.

A linear Z-estimation problem in $\mathbb{R}^d$ is specified by a fixed point equation of the form

$$\theta = \bar{L}\theta + \bar{b}, \tag{1}$$

where the matrix $\bar{L} \in \mathbb{R}^{d\times d}$ and the vector $\bar{b} \in \mathbb{R}^d$ are parameters of the problem. In settings of interest in this paper, the problem parameters $(\bar{L}, \bar{b})$ are unknown, and we observe only a sequence $(L_t, b_t)_{t\geq 1}$ of noisy observations, generated according to a Markov process in the following manner. The Markov process generates a sequence $(s_t)_{t\geq 0}$ of states taking values in some underlying state space $\mathbb{X}$. This chain is assumed to be ergodic, with a unique stationary

distribution ξ. The observed pair (L_t, b_t) at each time t depends on the current state s_t, and moreover, their expectations under the stationary distribution ξ are equal to their population-level counterparts $(\bar{L}, \bar{b})$.

This general formulation includes a number of special cases of interest. In the simplest setting, at each time t, we observe a matrix-vector pair of the form $L_{t+1} = \boldsymbol{L}(s_t)$ and $b_{t+1} = \boldsymbol{b}(s_t)$, where $\boldsymbol{L} : \mathbb{X} \to \mathbb{R}^{d \times d}$ and $\boldsymbol{b} : \mathbb{X} \to \mathbb{R}^d$ are deterministic mappings such that

$$\mathbb{E}_\xi\big[\boldsymbol{L}(s)\big] = \bar{L}, \quad \text{and} \quad \mathbb{E}_\xi\big[\boldsymbol{b}(s)\big] = \bar{b}. \tag{2a}$$

Many applications involve additional sources of randomness beyond that naturally associated with the Markov chain itself. In order to accommodate this possibility, we can consider observations of the form

$$L_{t+1} = \boldsymbol{L}_{t+1}(s_t), \quad \text{and} \quad b_{t+1} = \boldsymbol{b}_{t+1}(s_t). \tag{2b}$$

Here the mappings $\boldsymbol{L}_{t+1}$ and $\boldsymbol{b}_{t+1}$ are now allowed to be i.i.d. random, independent of s_t, but are required to be related to the deterministic mappings $\boldsymbol{L}$ and $\boldsymbol{b}$ via the relation

$$\mathbb{E}\big[\boldsymbol{L}_{t+1}(s)\big] = \boldsymbol{L}(s), \quad \mathbb{E}\big[\boldsymbol{b}_{t+1}(s)\big] = \boldsymbol{b}(s), \quad \text{for all } s \in \mathbb{X}. \tag{2c}$$

By the tower property of conditional expectation, equations (2a) and (2c) imply that $\boldsymbol{L}_{t+1}(s_t)$ and $\boldsymbol{b}_{t+1}(s_t)$ are unbiased estimates of $\bar{L}$ and $\bar{b}$, respectively.[1]

Stochastic approximation methods, dating back to the seminal work of Robbins and Monro [RM51], are standard iterative procedures for using data to approximately compute θ. These algorithms proceed in a streaming fashion: upon receiving each data point, an incremental update is made and the (averaged or) final iterate is returned in a single pass. In this way, each iteration of stochastic approximation incurs only mild computational and storage costs. Given these attractive computational properties, it is natural to ask if there are SA methods that also enjoy optimal statistical performance.

In this paper, we analyze the SA procedure based on the updates

$$\theta_{t+1} := (1 - \eta)\theta_t + \eta(L_{t+1}\theta_t - b_{t+1}), \quad \text{for } t = 0, 1, \ldots \tag{3a}$$

$$\widehat{\theta}_n := \frac{1}{n - n_0} \sum_{t=n_0}^{n-1} \theta_t \quad \text{for } n = n_0 + 1, n_0 + 2, \ldots. \tag{3b}$$

Equation (3a) describes a standard stochastic approximation update with constant stepsize $\eta > 0$, whereas equation (3b) corresponds to an application of the Polyak–Ruppert averaging procedure [PJ92, Rup88] to the iterates, with burn-in period n_0. When each matrix observation L_{t+1} has a constant rank independent of the dimension d—as is the case for temporal difference learning methods in reinforcement learning (see Section 2.2)—the SA method (3) can be implemented with $\mathcal{O}(d)$ computational and storage cost per iteration.

There is an extensive body of past work on stochastic approximation methods with Markov data. Here we provide an overview of the literature most germane to our contributions, and defer a more detailed review to Section 1.2. Asymptotic convergence of SA procedures with Markovian data can be established using either the ODE method [Bor09] or the Poisson equation method [BMP12]. The paper [TVR97] analyzes the asymptotic convergence of SA in the specific context of temporal difference methods in reinforcement learning. Although

[1] However, equation (2c) does not require the observations to be conditionally unbiased.

asymptotic guarantees provide helpful guidance, it is often most useful to have non-asymptotic guarantees that account both for limited sample size and scale of modern problems, and for these reasons, non-asymptotic analysis of Markovian SA procedures has attracted much recent attention.

Assuming a mixing time bound on the Markov chain, a projected variant of linear SA was analyzed by Bhandari et al. [BRS18], who established non-asymptotic rates that are near-optimal in their dependence on the sample size n. Srikant and Ying [SY19] analyzed the standard SA scheme without the projection step used in [BRS18], and obtained the same convergence rate in both mean-squared error and higher moments. Under an appropriate Lyapunov function assumption on the Markov chain, Durmus et al. [DMN+21] proved finite-time bounds for linear SA using stability properties of random matrix products. Variants and special cases of SA procedures with Markov data have also been studied, including two-time-scale algorithms [KMN+20], gradient-based optimization under Markov data [DNPR20], and estimation in auto-regressive models [BJN+20, JKNN21].

Despite this encouraging progress to date, two important questions still remain open, and form the focus of this paper:

- **Sample complexity in high dimensions:** The primary goal of non-asymptotic analysis is to provide guarantees on the estimation error that have an explicit dependence on the problem at hand, and that hold true for a reasonable range of values of the sample size n. For instance, suppose the linear Z-estimation problem in $\mathbb{R}^d$ is driven by an underlying Markov chain of mixing time t_{mix}. Then under natural noise assumptions, one should expect the estimation error to scale as $\mathcal{O}(t_{\mathrm{mix}} d/n)$, with this being the dominant term whenever $n \gtrsim t_{\mathrm{mix}} d$. However, existing analyses of linear SA do not provide such tight dimension-dependence. Using the notation of this paper, the sample size bounds in the papers [SY19, BRS18] rely on a uniform upper bound on the operator norm of the stochastic matrix $L_{t+1}(s_t)$; this quantity scales linearly with dimension d in many applications. Consequently, the resulting bounds on the MSE have a sub-optimal dependence on dimension, which is unsatisfactory for high-dimensional problems. Similarly, the bounds in the papers [DMN+21, KLL20, CMSS21] also exhibit a sub-optimal dependence on dimension. To the best of our knowledge, the question of whether linear SA succeeds under the minimal conditions on sample size—in particular, with n mildly larger than $d \cdot t_{\mathrm{mix}}$—remains open.
- **Instance-dependent optimality:** While many estimators may exhibit near-optimal statistical performance in the globally minimax (i.e., worst-case) sense, some of them perform significantly better than others when applied to practical problem instances. This phenomenon motivates the study of local (i.e., instance-dependent) performance in the non-asymptotic regime. Such results have recently been established for linear Z-estimation in the i.i.d. setting [PW21, LWC+20, KPR+21, MPW20]. The latter two papers listed provide non-asymptotic analogs of classical theory on local asymptotic minimaxity (c.f. [vdV00]), which establishes lower bounds by looking at the worst-case family of instances in a local neighborhood of a given problem. In the Markov setting, two questions naturally arise: (1) What does it mean for an estimator to be locally optimal in a non-asymptotic sense? (2) Does the linear SA estimator (3) match the local lower bound for every problem instance?

1.1 Contributions and organization

The primary goal of this paper is to resolve these challenges, and provide a sharp analysis of (averaged) linear SA algorithms. These answers are not merely of theoretical interest: they also provide important guidance for practice, such as in choosing algorithm parameters such

as the burn-in period and stepsize. In more detail:

- We perform a fine-grained analysis of linear SA and produce an upper bound on its statistical error that transparently tracks the dependence on problem-specific complexity as well as step-size. Furthermore, our bound holds true provided $n \gtrsim t_{\text{mix}} \cdot d$, establishing that the algorithm does indeed attain a sharp sample complexity guarantee in high dimensions.
- In a complementary direction to our upper bounds, we show a local minimax lower bound with an appropriately defined notion of local neighborhood of Markov chains. This lower bound certifies the statistical optimality of the linear SA estimator, again in an instance-dependent sense.
- We derive consequences of our general analysis for temporal difference methods in reinforcement learning, demonstrating a key problem-dependent quantity in matching upper and lower bounds.

One technical aspect of our analysis is noteworthy. En route to establishing bounds with sharp dimension dependence, we introduce a careful "bootstrapping" argument: starting with a loose bound, we progressively refine it via the repeated application of certain self-bounding inequalities. We suspect that this method may be of independent interest in providing sharp analyses of other stochastic approximation methods.

The remainder of this paper is organized as follows. We complete this section by introducing notation to be used throughout the paper, and then providing a more detailed discussion of related work. In Section 2, we provide the basic problem set-up, discuss the underlying assumptions, and give some illustrative examples. Section 3 is devoted to the presentation of our main results, which include upper bounds on the estimation error of stochastic approximation procedures, along with local minimax lower bounds that apply to any estimator. In Section 4, we develop some consequences of these results for specific models, including policy evaluation in reinforcement learning and estimation in autoregressive models. Sections 5, 6, and 7 are devoted to the proofs of Proposition 1, Theorem 1, and Theorem 2, respectively. We conclude with a discussion in Section 8. The proof of some auxiliary results and corollaries are postponed to the appendix.

Notation: We let $(\mathbb{X}, \rho)$ denote a metric space. For any $x \in \mathbb{X}$, we use δ_x to denote the distribution that places all its mass on $\{x\}$. Given a random variable X, we use the notation $\mathcal{L}(X)$ to denote its probability distribution. For a pair (π, μ) of probability distributions on $\mathbb{X}$, let $\Gamma(\pi, \mu)$ denote the space of all possible couplings of μ and π. For any $p \geq 1$, the Wasserstein-p distance between π and μ is given by

$$\mathcal{W}_p(\pi, \mu) := \left\{ \inf_{\gamma \in \Gamma(\pi,\mu)} \int_{\mathbb{X}\times\mathbb{X}} \rho(x,y)^p d\gamma(x,y) \right\}^{1/p}, \tag{4}$$

and the total variation distance between π and μ by

$$d_{\text{TV}}(\pi, \mu) := \sup_{A \subseteq \mathbb{X}} |\pi(A) - \mu(A)| .$$

Our analysis also involves various other divergences between probability measures. For any pair of probability distributions P and Q on the same space, we use $P \ll Q$ to denote the fact that P is absolute continuous with respect to Q, and use $\frac{dP}{dQ}$ to indicate the Radon-Nikodym

derivative. Given $P \ll Q$, we define:

$$\begin{aligned} \text{KL Divergence:} \quad & D_{\mathrm{KL}}(P \,\|\, Q) := \mathbb{E}_P\big[\log \tfrac{dP}{dQ}(X)\big], \\ \chi^2 \text{ divergence:} \quad & \chi^2\,(P \,||\, Q) := \mathbb{E}_P\big[\tfrac{dP}{dQ}(X) - 1\big], \\ \text{Max divergence:} \quad & D_\infty(P||Q) := \sup_{x \in \mathrm{supp}(Q)} \Big|\log \tfrac{dP}{dQ}(x)\Big|. \end{aligned}$$

Given any matrix $A = (a_{ij}) \in \mathbb{R}^{n \times m}$, its vectorization is obtained by concatenating its columns—viz. $\mathrm{vec}(A) := \begin{bmatrix} a_{11} & a_{2,1} & \cdots & a_{n1} & a_{12} & \cdots & a_{n2} & \cdots & a_{1m} & \cdots & a_{nm} \end{bmatrix}^\top \in \mathbb{R}^{nm}$. We use $\{e_j\}_{j=1}^d$ to denote the standard basis vectors in the Euclidean space $\mathbb{R}^d$, i.e., e_j is a vector with a 1 in the j-th coordinate and zeros elsewhere. For two matrices $A \in \mathbb{R}^{d_1 \times d_2}$ and $B \in \mathbb{R}^{d_3 \times d_4}$, we use $A \otimes B$ to denote their Kronecker product, a $d_1 d_3 \times d_2 d_4$ real matrix. For symmetric matrices $A, B \in \mathbb{R}^{d \times d}$, the notation $A \preceq B$ means that $B - A$ is a positive semi-definite matrix, whereas $A \prec B$ indicates that $B - A$ is positive definite. We use $\lambda_{\max}(A)$ and $\lambda_{\min}(A)$ to denote the largest and smallest eigenvalue of the matrix A, respectively. We use the following notation for matrix norms: for any matrix $A \in \mathbb{R}^{d_1 \times d_2}$, we use the notation $\|A\|_{\mathrm{op}}$, $\|A\|_F$ and $\|A\|_{\mathrm{nuc}}$ to denote its operator norm, Frobenius norm and nuclear norm, respectively.

Finally, throughout the paper, we use $\mathcal{F}_t := \sigma\big((b_i, L_i, s_i)_{i \le t}\big)$ to denote the natural filtration induced by the Markov observations.

1.2 Additional related work

This paper analyzes stochastic approximation algorithms based on Markov data, and has consequences for reinforcement learning. So as to put our results into context, we now provide more background on past work in these areas.

1.2.1 Statistical estimation based on Markov data

There is a large body of past work on statistical estimation based on observing a single trajectory of a Markov chain; for example, see Billingsley [Bil61] for an overview of some classical results. For the problem of functional estimation under the stationary distribution, the asymptotic efficiency of plug-in estimators[2] has been established for discrete-state Markov chains [Pen91, GW95] and Itô diffusion processes [Kut97]. In this paper, we provide non-asymptotic bounds, both upper and lower, that depend on a certain instance-dependent functional that also appears in an asymptotic analysis. More recent work has seen non-asymptotic results for statistical estimation with Markovian data, including the estimation of transition kernels [WK21, LWZ18], mixing times [HKL+19], the parameters of Gaussian hidden Markov models [YBW17], as well for certain testing problems [DDG18]. These papers can be roughly divided into two categories. Papers in the first category focus on estimating parameters for each individual state of the Markov chain (e.g., transition kernels), and thus require sample sizes that scale with the complexity of the state space (e.g., its cardinality in the discrete case). By contrast, papers in the second category are concerned with estimating properties of the Markov chain (e.g., the expectation of a functional under the stationary distribution), and the sample complexity of such problems need not depend on the size of the state space. Our paper falls within the second category.

[2] These papers refer to such methods as "empirical" estimators.

1.2.2 Stochastic approximation methods

The use of recursive stochastic procedures for solving fixed point equations dates back to the seminal work of Robbins and Monro [RM51]; see the reference books [Bor09, BMP12, KY03] for more background. By averaging the iterates of the SA procedure, it is known that one can obtain both an improved convergence rate and central limit behavior [PJ92, Rup88]. A variety of stochastic approximation procedures now serve as the workhorse for modern large-scale machine learning and statistical inference [NJLS09, BCN18], and many algorithmic techniques are known to accelerate their convergence [GL12, JZ13, LMWJ20]. In particular, non-asymptotic bounds matching the optimal Gaussian limit have been established in a variety of settings [MB11, GP17, DDB20, MLW+20, MPW20].

While the instance-dependent nature of this line of investigation aligns with the objective of our work, prior work either assumes an i.i.d. observation model or imposes a martingale difference assumption on the noise.[3] The first study of SA procedures without a martingale difference assumption was initiated by Kushner and Clark [KC78], who give a general criteria for convergence, as well as Ljung [Lju77a, Lju77b], who analyzed linear problems motivated by control and filtering. The work [MP84] analyzed general SA problems for controlled Markov processes by applying the Kushner–Clark lemma. In addition to this classical work, stochastic approximation in the Markov setting has attracted much recent attention. Central limit theorems [For15] and non-asymptotic convergence rates [KMMW19] have been established for controlled Markov processes. In addition to the papers discussed in Section 1, several recent works have considered particular aspects of SA with Markov data, including two-time-scale variants [DNPR20, KB18], observation skipping schemes for bias reduction [KLL20], Lyapunov function-based analysis under general norms [CMSS21], and proving guarantees under weaker ergodicity conditions [DDA21].

1.2.3 Application to RL problems

Markovian observations arise naturally in the context of stochastic control and reinforcement learning (RL). See the book [BMP12] for a historical survey of algorithms for stochastic control and filtering with Markovian stochastic approximation, and the books [Ber19, Sze10] for more background on the RL setting. In RL problems, SA algorithms are typically used to solve Bellman equations, a class of linear or non-linear fixed-point equations. In policy evaluation problems, temporal difference (TD) methods [Sut88] use linear stochastic approximation to estimate the value function of a given policy, with asymptotic convergence guarantees [DS94, TVR97, Boy02] and non-asymptotic bounds [BRS18, KPR+21, MPW20]. In the non-linear case, the Q-learning algorithm [WD92] is a stochastic approximation method that estimates the Q-function of a Markov decision process from data. There is a long line of past work on this algorithm, including convergence guarantees [Tsi94, Sze98, EDM03], results on linear function approximation for optimal stopping problems [TVR99, BRS18], and non-asymptotic rates under general norms in both the i.i.d. setting [Wai19a, Bor21] as well as the Markovian setting [CMSS21]. A class of variants of TD and Q-learning are also studied in literature, including actor-critic methods [KT00], SARSA [RN94], and methods that employ variance-reduction [SWW+18, KPR+21, Wai19b, KXWJ21]. A concurrent preprint to this manuscript [LLP21] proves lower bounds on the oracle complexity of policy evaluation with access to temporal difference operators, and develops an acceleration scheme with variance

[3] In the linear equation setup, the martingale difference noise assumes that $\mathbb{E}[L_{t+1} \mid \mathcal{F}_t] = \bar{L}$ and $\mathbb{E}[b_{t+1} \mid \mathcal{F}_t] = \bar{b}$, which does not cover the Markov case.

reduction to achieve these lower bounds while retaining the optimal sample complexity.

It should be noted that an important feature of reinforcement learning is function approximation, i.e., using a given function class (e.g. a linear subspace) to approximate the solution to the Bellman equation of interest. This method enables estimation with a sample size depending on the intrinsic complexity of the function class, instead of the cardinality of state-action space. On the other hand, an approximation error is induced by projecting the Bellman equation onto this function class. This trade-off is central to the class of TD algorithms, as studied in a line of past work [TVR97, YB10, Ber11, MS08, MPW20]. Prior work by a subset of the authors [MPW20] focuses on the i.i.d. setting, and shows that projected linear equations have a non-standard tradeoff between approximation and estimation errors. The current paper is complementary in nature, building on this work by analyzing the more challenging setting of Markov observations. Among the concrete consequences of this paper are an instance-optimal analysis of TD algorithms in the Markov setting with linear function approximation. This analysis provides the basis for a principled choice of the parameter λ in the broader class of TD(λ) algorithms.

2 Problem set-up

Recall from our earlier set-up (cf. equation (1)) that we are interested in solving a fixed point equation of the form $\theta = \bar{L}\theta + \bar{b}$, based on noisy observations of the pair $(\bar{L}, \bar{b})$, as defined by the Markov observation model (2). We require that the matrix $\bar{L}$ satisfies the conditions

$$\kappa := \frac{1}{2}\lambda_{\max}\big(\bar{L} + \bar{L}^{\top}\big) < 1, \quad \text{and} \quad \|\bar{L}\|_{\text{op}} \leq \gamma_{\max}. \tag{5}$$

2.1 Assumptions

We now introduce and discuss the remaining four assumptions that underlie our analysis.

2.1.1 Conditions on Markov chain

We first describe the conditions imposed on the underlying Markov chain in our observation model. Let $\{s_t\}_{t \geq 0}$ denote a trajectory drawn from a Markov chain with transition kernel P. We assume that this chain has a unique stationary distribution ξ, and impose the following mixing condition in Wasserstein-1 distance:

Assumption 1. *There exists a natural number t_{mix} and a universal constant $c_0 \geq 1$ such that for any $x, y \in \mathbb{X}$, we have the bounds*

$$\mathcal{W}_{1,\rho}(\delta_x P^{t_{\text{mix}}}, \delta_y P^{t_{\text{mix}}}) \overset{(a)}{\leq} \frac{1}{2}\rho(x,y), \quad \textit{and} \quad \mathcal{W}_{1,\rho}(\delta_x P^{t}, \delta_y P^{t}) \overset{(b)}{\leq} c_0 \rho(x,y), \quad \textit{for all } t = 0, 1, 2, \ldots. \tag{6}$$

We assume throughout that the chain is initialized with a sample $s_0 \sim \xi$ from the stationary distribution. Given that our mixing time bound guarantees exponential decay of the Wasserstein distance, this condition is mild: it can be removed by waiting $\mathcal{O}(t_{\text{mix}})$ iterations for the process to mix.

2.1.2 Tail conditions on noise

In our observation model, the "noise" terms correspond to the differences $\boldsymbol{L}_{t+1}(s_t) - \boldsymbol{L}(s_t)$ and $\boldsymbol{L}(s_t) - \bar{L}$, along with analogous quantities for the vector b. Our second assumption imposes conditions on these noise variables. We consider separate conditions on these martingale $\boldsymbol{L}_{t+1}(s_t) - \boldsymbol{L}(s_t)$ and Markov $\boldsymbol{L}(s_t) - \bar{L}$ parts of the noise, as well as the b-noise analogues.

Assumption 2. *There exists an even integer $\bar{p} \in [2, +\infty]$ and non-negative constants σ_L and σ_b, such that for any positive even integer $p \leq \bar{p}$, scalar $t \geq 0$, vector $u \in \mathbb{S}^{d-1}$, and index $j \in \{1, \ldots, d\}$, we have*

$$\mathbb{E}\big[\langle e_j, \big(\boldsymbol{L}_{t+1}(s_t) - \boldsymbol{L}(s_t)\big)u\rangle^p \mid \mathcal{F}_t\big] \leq p!\sigma_L^p, \quad \text{and} \quad \mathbb{E}_{s\sim\xi}\big[\mathbb{E}\big[\langle e_j, \boldsymbol{b}_{t+1}(s) - b(s)\rangle^p \mid s\big]\big] \leq p!\sigma_b^p, \tag{7a}$$

as well as

$$\mathbb{E}_{s\sim\xi}\big[\langle e_j, \big(\boldsymbol{L}(s) - \bar{L}\big)u\rangle^p\big] \leq p!\sigma_L^p, \quad \text{and} \quad \mathbb{E}_{s\sim\xi}\big[\langle e_j, \boldsymbol{b}(s_t) - \bar{b}\rangle^p\big] \leq p!\sigma_b^p. \tag{7b}$$

Note that this assumption is mildest for $\bar{p} = 2$, and strongest for $\bar{p} = \infty$. In the latter case, when $\bar{p} = \infty$, the assumption requires L_{t+1} and b_{t+1} to be sub-exponential random variables in the standard coordinate directions (since $\log(p!) \leq p\log(p/2)$ by concavity of the log function). This condition covers, for instance, the case where L_{t+1} is the outer product of sub-Gaussian random vectors, as in temporal difference learning methods. In addition to accommodating this case, Assumption 2 also covers the heavier-tailed setting in which only finitely many moments exist. In particular, when $\bar{p} = 2$, the second moment assumption coincides with the assumption made in the paper [MPW20].

An important quantity in our analysis is the *effective noise level* given by

$$\bar{\sigma} := \sup_{p\in[2,\bar{p}]} \sup_{j\in[d]} p^{-1}\big(\mathbb{E}\big[\langle e_j, (\boldsymbol{L}_{t+1}(s_t) - \bar{L})\bar{\theta} + (\boldsymbol{b}_{t+1}(s_t) - \bar{b})\rangle^p\big]\big)^{1/p}.$$

Note that under Assumption 2, we have the upper bound $\bar{\sigma} \leq \sigma_L\|\bar{\theta}\|_2 + \sigma_b$.

2.1.3 Metric space conditions

For most of our analysis, we impose the following condition:

Assumption 3. *The metric space $(\mathbb{X}, \rho)$ has diameter at most one.*

Note that our assumption of unit diameter is arbitrary; boundedness suffices. In order to accommodate the general case, it suffices to rescale the parameters σ_L and σ_b.

When applying our theory to unbounded spaces (e.g., $\mathbb{X} = \mathbb{R}^d$), we use a truncation argument to show that there is an event over a reduced state space on which this condition holds with probability tending exponentially to 1. (See Appendix A for the details of this argument.)

2.1.4 Lipschitz condition

Finally, we place a Lipschitz assumption—under the metric ρ—on the mapping from the metric space $\mathbb{X}$ to the stochastic operators. Given the Markov chain setup in the metric space $(\mathbb{X}, \rho)$, it is alluring to assume dimension-free Lipschitz bound on the mappings $(\boldsymbol{L}_t, \boldsymbol{b}_t)$. However, as the space $\mathbb{X}$ has diameter bounded by 1, such Lipschitz constants typically depend

on dimension for practical problems. Concretely, view the $\bar{L}$-scale parameters $(\kappa, \gamma_{\max})$ as constants and assume that the observations $\boldsymbol{L}_{t+1}(s_t)$ each have rank at most r. We then have

$$\mathbb{E}\big[\|\boldsymbol{L}_{t+1}(s_t)\|_{\mathrm{op}}\big] \geq \frac{\mathbb{E}\big[\|\boldsymbol{L}_{t+1}(s_t)\|_{\mathrm{nuc}}\big]}{r} \geq \frac{\mathrm{trace}\big(\mathbb{E}\big[\boldsymbol{L}_{t+1}(s_t)\big]\big)}{r} = \frac{\mathrm{trace}(\bar{L})}{r}. \tag{8}$$

The trace $\mathrm{trace}(\bar{L})$ typically scale as $\Theta(d)$, even in the "easy case" when $\bar{L}$ is a constant multiple of identity matrix.

So the Lipschitz constant for the mapping $\boldsymbol{L}_t : \mathbb{X} \to \mathbb{R}^{d\times d}$ is at least $\Omega(d)$. On the other hand, as a d-dimensional standard Gaussian random variable has norm $\Omega(\sqrt{d})$ with high probability, it is natural to assume the Lipschitz constant for the vector-valued mapping $\boldsymbol{b}_t : \mathbb{X} \to \mathbb{R}^d$ to be of order at least $\Omega(\sqrt{d})$. We therefore make the following assumption:

Assumption 4. *There exist constants $\sigma_L, \sigma_b > 0$ such that, almost surely for any $x, y \in \mathbb{X}$, we have*

$$\|\boldsymbol{L}_t(x) - \boldsymbol{L}_t(y)\|_{op} \leq \sigma_L d \cdot \rho(x, y) \quad \textit{and} \quad \|\boldsymbol{b}_t(x) - \boldsymbol{b}_t(y)\|_2 \leq \sigma_b \sqrt{d} \cdot \rho(x, y) \tag{9}$$

for all $t = 1, 2, \ldots$.

Note that in Assumption 4, we explicitly rescale the RHS of the inequalities with factors that depend on the problem dimension d, so that the pair (σ_L, σ_b) should indeed be viewed as dimension-free. The notation (σ_L, σ_b) is actually overloaded in Assumptions 2 and 4. In practice, we can take the maximum of the bounds in the two assumptions. Besides, as shown in Appendix A, for certain natural problem classes, Assumption 2 indeed implies Assumption 4 with discrete metric, up to logarithmic factors.

2.2 Some illustrative examples

Our assumptions cover a broad range of ergodic Markov chains, and the fixed-point equation (1) associated with their stationary distribution naturally arises from several problems. In this section, we describe a few concrete examples of our general setup. We first discuss the class of Markov chains satisfying our assumptions, and then describe the linear Z-estimators associated with such problems.

2.2.1 Examples of Markov chains

By varying our choice of the metric ρ, we recover several important classes of Markov chains that satisfy Assumptions 1 and 3.

- Consider a Markov chain defined on a countable state space $\mathbb{X}$, and consider the discrete metric $\rho(x, y) := \mathbf{1}_{x\neq y}$. In this context, Assumption 1 corresponds to mixing time bound in total variation—viz.

  $$d_{\mathrm{TV}}(\delta_x P^{t_{\mathrm{mix}}}, \delta_y P^{t_{\mathrm{mix}}}) \leq \tfrac{1}{2} \qquad \text{for all pairs } x, y \in \mathbb{X}.$$

  This mixing condition is satisfied for some finite t_{mix} when the Markov chain is irreducible, aperiodic and positive recurrent. Moreover, this metric space has unit diameter, so that Assumption 3 holds as well.
- As another example, consider the state space $\mathbb{X} = \mathbb{B}(0, 1) \subseteq \mathbb{R}^d$ equipped with the Euclidean metric $\rho(x, y) = \|x - y\|_2$. We can define a Markov chain on this space via the random evolution $X_{k+1} = \mathcal{T}_{k+1}(X_k)$, where the random non-linear operators $\{\mathcal{T}_k\}_{k\geq 1} \subseteq \mathbb{X}^{\mathbb{X}}$ are

drawn i.i.d. from some distribution. We assume that the expected operator operator $\bar{\mathcal{T}} := \mathbb{E}[\mathcal{T}_1]$ satisfies the contraction condition $\|\bar{\mathcal{T}}(x) - \bar{\mathcal{T}}(y)\|_2 \leq \gamma\|x - y\|_2$ with some $\gamma < 1$. Assuming the stochastic operator $\mathcal{T}$ to be Lipschitz and to satisfy a second moment bound, this dynamical system satisfies the Wasserstein contraction condition under the Euclidean metric.

2.2.2 Examples of linear Z-estimators

We now describe some interesting examples of linear Z-estimators, to which we will return in later sections.

Example 1 (Approximate policy evaluation)**.** We begin by considering the TD(0) algorithm for approximate estimation of value functions. This problem arises in the context of Markov reward processes (MRPs), which are Markov chains that are augmented with a reward function $r : \mathbb{X} \to \mathbb{R}$. A trajectory from a Markov reward process is a sequence $\{(s_t, R_t)\}_{t\geq 0}$, where $\{s_t\}_{t\geq 0}$ is the Markov trajectory of states, and R_t is a random reward, corresponding to a conditionally unbiased estimate (given s_t) of the reward function value $r(s_t)$. Given a discount factor $\gamma \in [0, 1)$, the expected discount reward defines the *value function* $V^*(s) = \mathbb{E}\big[\sum_{t=0}^{\infty} \gamma^t R_t \mid s_0 = s\big]$.

This value function is connected to linear Z-estimators via the Bellman principle. Let P denote the transition operator of the Markov chain, and let ξ denote the stationary distribution. Note that the P maps the space $\mathbb{L}^2(\mathbb{X}, \xi)$ to itself. With this notation, the value function V^* is known to be the unique fixed point of the *Bellman evaluation equation*

$$V = \gamma PV + r. \tag{10}$$

In general, this equation is non-trivial to solve, especially given a limited trajectory length. In practice, it is standard to compute approximate solutions using linear basis expansions, [BB96, TVR97], and this approach underlies the family of TD algorithms.

Let $\{\phi_j\}_{j=1}^d$ be a collection of linearly independent real-valued functions defined on the state space, and consider the linear subspace $\mathbb{S}$ of all functions of the form $V_\theta(s) = \sum_{j=1}^d \theta_j \phi_j(s)$. This subspace defines the *projected Bellman equation*

$$\bar{V} = \Pi_{\mathbb{S}}\big(\gamma P\bar{V} + r\big), \tag{11}$$

where $\Pi_{\mathbb{S}}$ is the orthogonal projection operator under $\mathbb{L}^2(\mathbb{X}, \xi)$.

By definition, the projected fixed point $\bar{V}$ can be written in the form $\bar{V}(s) = \sum_{j=1}^d \bar{\theta}_j \phi_j(s)$ for some vector $\bar{\theta} \in \mathbb{R}^d$. Some simple calculations show that this parameter vector must satisfy the linear system

$$\Sigma_0\bar{\theta} = \gamma\Sigma_1\bar{\theta} + \mathbb{E}_{s\sim\xi}\big[R_0(s)\phi(s)\big], \tag{12}$$

where $\Sigma_0 = \mathbb{E}_{s\sim\xi}\big[\phi(s)\phi(s)^\top\big]$ is the second-moment matrix of $\phi(s)$ under the stationary distribution, and $\Sigma_1 = \mathbb{E}[\phi(s)\phi(s^+)^\top]$ is the cross-moment operator of the Markov chain. In defining this cross-moment, the expectation is taken over $s \sim \xi$ and $s^+ \sim P(s, \cdot)$.

This problem can be viewed within our framework by considering a Markov chain on the augmented state space $\omega_t = (s_t, s_{t+1})$. Equation (12) defines a fixed point equation under the stationary distribution of this Markov chain. Define the minimum and maximum eigenvalues $\mu := \lambda_{\min}(\Sigma_0)$ and $\beta := \lambda_{\max}(\Sigma_0)$, along with the observation functions

$$\boldsymbol{b}_{t+1}(\omega_t) = \tfrac{1}{\beta} R_t(s_t)\phi(s_t), \quad \text{and} \quad \boldsymbol{L}_{t+1}(\omega_t) = I_d - \tfrac{1}{\beta}\big[\phi(s_t)\phi(s_t)^\top - \gamma\phi(s_t)\phi(s_{t+1})^\top\big]. \tag{13}$$

With these choices, the stochastic approximation procedure (3) is the widely used TD(0) algorithm. On the other hand, for a stationary Markov chain $(s_t)_{t\in\mathbb{Z}}$, the fixed-point equation $\bar{\theta} = \mathbb{E}\left[\boldsymbol{L}_{t+1}(\omega_t)\right] \cdot \bar{\theta} + \mathbb{E}\left[\boldsymbol{b}_{t+1}(\omega_t)\right]$ is equivalent to Eq (12). Note that though the expression for the mappings $\boldsymbol{b}_{t+1}$ and $\boldsymbol{L}_{t+1}$ depends on unknown parameter β, they can be absorbed into the stepsize choice, and the algorithm works well without such knowledge.

Typically, the Euclidean norm $\|\phi(s)\|_2$ of the feature vectors scales as $\sqrt{d}$, and under the stationary distribution ξ, the variance of any coordinate of $\phi(s)$ is of constant order. Under these conditions, the cross-moment matrix Σ_1 has operator norm of constant order. On the other hand, as for the random observations, we have the scalings $\|\!|L_{t+1}|\!\|_{\mathrm{op}} = \mathcal{O}(d)$ and $\|b_{t+1}\|_2 = \mathcal{O}(\sqrt{d})$, so that Assumptions 2 and 4 are satisfied. ♣

In the context of TD, it is natural to consider a *sieve estimator*. Given a collection of basis functions $\{\phi_j\}_{j=1}^{\infty}$, we can define the nested family $\mathbb{S}_1 \subset \mathbb{S}_2 \subset \cdots$, where $\mathbb{S}_d$ denotes the span of the sub-collection $\{\phi_j\}_{j=1}^{d}$. Here the choice of the sieve parameter d is key: larger values reduce the approximation error at the expense of increasing the estimation error. We discuss how this can be done in Section 4.

Another extension of the TD(0) algorithm—one that becomes feasible under the Markovian observation model—is the TD(λ) family of procedures. A fundamental question is how well the solution of the projected fixed-point equation (11) approximates the true value function V^*. The paper [MPW20] analyzes this quantity, and provides matching upper and lower bounds in the i.i.d. setting. However, the Markovian observation model actually allows this approximation error to be reduced, albeit at the cost of increased estimation error, as discussed in our next example.

Example 2 (Policy evaluation with TD(λ))**.** The family of TD(λ) algorithms is motivated by the following observation: since the value function V^* is the fixed point of Eq (10), it is also the fixed point of the composition of itself. Concretely, for any $k \geq 1$, we have:

$$V^* = (\gamma P)^k V^* + \sum_{j=0}^{k-1} (\gamma P)^j r.$$

For any $\lambda \in [0, 1)$, we take the weighted average of the above (infinite) collection of equations using exponentially-decaying weight $(1, \lambda, \lambda^2, \cdots)$, and obtain the following equation.

$$V = (1-\lambda) \sum_{k=0}^{\infty} \lambda^k (\gamma P)^{k+1} V + \sum_{k=0}^{\infty} \lambda^k (\gamma P)^k r. \tag{14a}$$

The solution V^* to the equation (10) also solves Eq (14a).

Following the same route as TD(0), for a given subspace $\mathbb{S}$ of functions, we seek a solution $\bar{V}^{(\lambda)}$ to the projected fixed equation equation

$$\bar{V}^{(\lambda)} = (1-\lambda) \sum_{k=0}^{\infty} \lambda^k \Pi_{\mathbb{S}} (\gamma P)^{k+1} \bar{V}^{(\lambda)} + \sum_{k=0}^{\infty} \lambda^k \Pi_{\mathbb{S}} (\gamma P)^k r, \tag{14b}$$

in which the operator P has been replaced by the projection $\Pi_{\mathbb{S}} P$. Although the fixed points of equation (14a) and the Bellman equation (10) coincide, the projected version (14b) has a different set of fixed points.

Since the value function $\bar{V}^{(\lambda)}$ lies in the linear space $\mathbb{S}$, it has a representation of the form $\bar{V}^{(\lambda)}(s) = \sum_{j=1}^{d} \bar{\theta}_j^{(\lambda)} \phi_j(s)$ for some coefficient vector $\bar{\theta}^{(\lambda)} \in \mathbb{R}^d$. From equation (14b), this vector must satisfy a linear system of the form

$$\Big[\sum_{k=0}^{\infty}(\lambda\gamma)^k \Sigma_k\Big]\bar{\theta}^{(\lambda)} = \Big[\sum_{k=0}^{\infty}(\lambda\gamma)^k \gamma\Sigma_{k+1}\Big]\bar{\theta}^{(\lambda)} + \sum_{k=0}^{\infty}(\lambda\gamma)^k \mathbb{E}\big[R_0(s_0)\phi(s_{-k})\big], \tag{15}$$

where $\{s_k\}_{k=-\infty}^{\infty}$ is a stationary Markov chain following the transition kernel P, and we define $\Sigma_k = \mathbb{E}[\phi(s_{-k})\phi(s_0)^\top]$ for each integer k. As it should, when we set $\lambda = 0$, equation (15) reduces to the TD(0) update from equation (12).

In order to use stochastic approximation methods to solve this equation, we consider an augmented Markov process $(s_{t+1}, s_t, g_t)_{t\in\mathbb{Z}}$ in the space $\mathbb{X}^2 \times \mathbb{R}^d$, which evolves as

$$s_{t+1} \sim P(s_t, \cdot), \quad \text{and} \quad g_t = \phi(s_t) + \gamma\lambda g_{t-1}. \tag{16a}$$

If feature vectors $\phi(s_t)$ lie in a compact set almost surely, we have $g_t = \sum_{k=0}^{+\infty}(\gamma\lambda)^k\phi(s_{t-k})$. Let $\widetilde{\xi}$ be the stationary distribution of this augmented Markov chain.[4] In terms of an element $\omega = (s, s^+, g)$ drawn according this stationary distribution, the fixed-point equation (14b) admits the succinct representation

$$\mathbb{E}_{\widetilde{\xi}}\big[g\phi(s)^\top\big]\bar{\theta}^{(\lambda)} = \gamma\mathbb{E}_{\widetilde{\xi}}\big[g\phi(s^+)^\top\big]\bar{\theta}^{(\lambda)} + \mathbb{E}_{\widetilde{\xi}}\big[R_0(s)g\big]. \tag{16b}$$

By choosing the observation functions

$$\boldsymbol{L}_{t+1}(\omega_t) = I_d - \nu\cdot\big(g_t\phi(s_t)^\top - \gamma g_t\phi(s_{t+1})^\top\big), \quad \boldsymbol{b}_{t+1}(\omega_t) = \nu\cdot R_t(s_t)\phi(s_t), \tag{16c}$$

for a scalar $\nu > 0$, this algorithm is a special case of our general set-up. In particular, by substituting the infinite-sum expression for the random variable g_t into Eq (16b), we obtain the projected linear equation (15) under the low-dimensional representation. See Section 4 for a more detailed verification of the assumptions needed to apply our main results for this problem. ♣

For our last example, we turn to a different class of problems involving vector autoregressive (VAR) models for time series [Lö5].

Example 3 (Parameter estimation in autoregressive models)**.** An m-dimensional VAR model of order k describes the evolution of a random vector X_t as a k^{th}-order Markov process. The model is specified by a collection of $m \times m$ matrices $\{A_j^*\}_{j=1}^k$, and the random vector evolves according to the recursion

$$X_{t+1} = \sum_{j=1}^{k} A_j^* X_{t-j+1} + \varepsilon_{t+1}, \tag{17}$$

where the noise sequence $\big(\varepsilon_t\big)_{t\geq 0}$ is i.i.d. and zero-mean. and supported on a bounded set.

Considering the $(k+1)$-fold tuple $\omega_t = (X_{t+1}, X_t, \cdots, X_{t-k+1})$, the process $\big(\omega_t\big)_{t\geq 0}$ is Markovian. Under appropriate stability assumptions on the model parameter, the process mixes rapidly under the $(k+1)m$-dimensional Euclidean metric. Let $\widetilde{\xi}$ denote its stationary distribution, and suppose for convenience that the chain is observed at stationarity.

[4] Such a stationary distribution exists and is unique under suitable assumptions. See 4.2 for details.

In order to estimate the model parameters, we consider the following set of Yule–Walker estimation equations:

$$\mathbb{E}\big[X_{t+1}X_{t-\ell}^\top\big] = A_1^*\mathbb{E}\big[X_tX_{t-\ell}^\top\big] + A_2^*\mathbb{E}\big[X_{t-1}X_{t-\ell}^\top\big] + \cdots + A_k^*\mathbb{E}\big[X_{t-k+1}X_{t-\ell}^\top\big], \tag{18}$$

for $\ell = 0, 1, \cdots, k-1$.

These equations form a km^2-dimensional linear system for estimating km^2-dimensional parameters. Note that the parameters live in the space of matrix sequences, and so we slightly abuse our notation for simplicity: L denotes a linear operator from $\mathbb{R}^{k\times m\times m}$ to itself, and b is an element in $\mathbb{R}^{k\times m\times m}$. At the sample level, for any collection $A := \{A_j\}_{j=1}^k \in \mathbb{R}^{k\times m\times m}$ of system matrices, the stochastic observations are given by

$$\big[\boldsymbol{b}_{t+1}(\omega_t)\big]_\ell = \nu\, X_{t+1}X_{t-\ell}^\top \quad \text{for } \ell = 0, 1, \ldots, k-1, \text{ and}$$
$$\big(\boldsymbol{L}_{t+1}(\omega_t)\big)[A]_\ell = A_\ell - \nu \sum_{j=0}^{k-1} A_j X_{t-j}X_{t-\ell}^\top, \quad \text{for } \ell = 0, 1, \ldots, k-1.$$

Once again, the parameter ν is a scaling constant needed to fit into the fixed-point equation framework, and is absorbed into the stepsize choice of the algorithm. ♣

3 Main results

We now turn to the statement of our main results, beginning with our upper bounds in Section 3.1, followed by lower bounds in Section 3.2.

3.1 Instance-dependent upper bounds

In this section, we begin by stating some upper bounds (Theorem 1) on the behavior of the Polyak–Ruppert averaged SA scheme (3b). These bounds are instance-dependent, in the sense that they are specified in terms of an explicit function of the operator $\bar{L}$ and the fixed point $\bar{\theta}$. We then state a second result (Proposition 1) on the non-averaged iterates, which plays a key role in proving Theorem 1.

3.1.1 Instance-dependent bounds on the averaged iterates

For any state $s \in \mathbb{X}$, define the functions

$$\varepsilon_{\mathrm{MG}}(s) := (\boldsymbol{b}_1(s) - \boldsymbol{b}(s)) + (\boldsymbol{L}_1(s) - \boldsymbol{L}(s))\bar{\theta}, \quad \text{and} \quad \varepsilon_{\mathrm{Mkv}}(s) := \boldsymbol{b}(s) + \boldsymbol{L}(s)\bar{\theta} - \bar{\theta}.$$

Note that for a fixed state s, the quantity $\varepsilon_{\mathrm{MG}}(s)$ depends on the random variables $\boldsymbol{b}_1(s)$ and $\boldsymbol{L}_1(s)$, and so is a random vector, whereas by contrast, the quantity $\varepsilon_{\mathrm{Mkv}}(s)$ is deterministic. Letting $(\widetilde{s}_t)_{t=-\infty}^{\infty}$ be a stationary Markov chain under the transition kernel P, we then define the matrices

$$\Sigma_{\mathrm{MG}}^* := \mathbb{E}_\xi\big[\operatorname{cov}\big(\varepsilon_{\mathrm{MG}}(s) \mid s\big)\big], \quad \text{and} \quad \Sigma_{\mathrm{Mkv}}^* := \sum_{t=-\infty}^{\infty} \mathbb{E}\big[\varepsilon_{\mathrm{Mkv}}(\widetilde{s}_t)\varepsilon_{\mathrm{Mkv}}(\widetilde{s}_0)^\top\big]. \tag{19}$$

Overall, the performance of our algorithm depends on the *matrix sum* $\Sigma^* := \Sigma_{\mathrm{MG}}^* + \Sigma_{\mathrm{Mkv}}^*$, as well as the *effective noise variance* $\bar{\sigma}^2 := \sigma_L^2\|\bar{\theta}\|_2^2 + \sigma_b^2$. In terms of these quantities, we have the following guarantee:

Theorem 1. *Under Assumptions 1–3, suppose that we set the stepsize η and burn-in parameter n_0 as $\eta = \big(c(\sigma_L^2 d + \gamma_{\max}^2)(1-\kappa)n^2 t_{\mathrm{mix}}\big)^{-1/3}$ and $n_0 = \frac{1}{2}n$, where c is a suitably chosen universal constant. Then for any sample size n satisfying $\frac{n}{\log^2 n} \geq \frac{2t_{\mathrm{mix}}(\sigma_L^2 d + \gamma_{\max}^2)}{(1-\kappa)^2}\log(c_0 d)$, the Polyak–Ruppert estimate (3b) has MSE bounded as*

$$\mathbb{E}\big[\|\widehat{\theta}_n - \bar{\theta}\|_2^2\big] \leq \frac{c'}{n}\mathrm{Tr}\big((I-\bar{L})^{-1}(\Sigma^*_{\mathrm{MG}} + \Sigma^*_{\mathrm{Mkv}})(I-\bar{L})^{-\top}\big) + c'\big(\tfrac{\bar{\sigma}^2 d t_{\mathrm{mix}}}{(1-\kappa)^2 n}\big)^{4/3}\log^2 n. \tag{20}$$

See Section 6 for the proof of this theorem.

A few remarks are in order. First, and as shown in the next section, the first term $n^{-1}\mathrm{Tr}\big((I-\bar{L})^{-1}\Sigma^*(I-\bar{L})^{-1}\big)$ is optimal for the Markovian stochastic approximation problem in an instance-dependent sense. This term appears in existing central limit results for Markovian stochastic approximation [For15], whereas our bound captures this dependence in a non-asymptotic manner.

The first term can always be upper bounded by $c'\frac{\bar{\sigma}^2}{(1-\kappa)^2 n}t_{\mathrm{mix}} d \cdot \log^2(c_0 d)$.[5] On the other hand, disregarding dependence on (σ_L, σ_b) and logarithmic factors in the sample size, the second term in the bound scales as $\mathcal{O}\big(\big(\frac{t_{\mathrm{mix}} d}{(1-\kappa)^2 n}\big)^{4/3}\big)$. Consequently, up to polylogarithmic factors, we have

$$\mathbb{E}\big[\|\widehat{\theta}_n - \bar{\theta}\|_2^2\big] \lesssim \frac{\bar{\sigma}^2 t_{\mathrm{mix}} d}{(1-\kappa)^2 n}. \tag{21}$$

Thus, at least in a worst-case sense, the second term is always dominated by the first term.

We note that Theorem 1 makes two types of tail assumptions on the random observations: Assumption 2 with $\bar{p} = 2$ requires dimension-free second moment bounds in any coordinate direction, whereas the Lipschitz condition (Assumption 4) together with Assumption 3 (boundedness of the domain) imply a (dimension-dependent) uniform upper bound on the noise. The two assumptions play very different roles in the analysis of high-dimensional problems. As we will see in Proposition 3, such assumptions are naturally satisfied in the context of sieve estimators, for which dimension d of the problem is selected adaptively based on sample size n.

Finally, we also note that the requirement on the sample size n is nearly optimal, since we require $n = \widetilde{\Omega}\big(\frac{t_{\mathrm{mix}} d}{(1-\kappa)^2}\big)$ to make the estimation error (21) less than a constant (by seeing σ_L and $\gamma_{\max}$ as constants). Up to an additional $\mathcal{O}(t_{\mathrm{mix}})$ factor, the sample size requirement in Theorem 1 also matches that of linear stochastic approximation in the i.i.d. setting [LS18, MLW+20, MPW20]. This additional $\mathcal{O}(t_{\mathrm{mix}})$ factor is unavoidable, which can be seen from the following reduction from the Markov to the i.i.d.setting. Consider a problem instance in the i.i.d. setup, given by a probability distribution $\mathbb{P}$ over $\mathbb{R}^{d\times d} \times \mathbb{R}^d$. Defining the state (L_t, b_t), consider a lazy Markov chain that remains at the same state with probability $1 - \frac{1}{t_{\mathrm{mix}}}$, and jumps to an independent state drawn from $\mathbb{P}$ with probability $\frac{1}{t_{\mathrm{mix}}}$. A Markov trajectory of size n in this lazy Markov chain is approximately equivalent to $\mathcal{O}(n/t_{\mathrm{mix}})$ samples in the i.i.d. model, and results in a multiplicative blow-up of $\mathcal{O}(t_{\mathrm{mix}})$ in the sample complexity requirement for the Markov case.

[5] This can be easily seen from exponential decay of the correlation; in particular, see equation (74) in the proof of the theorem.

3.1.2 Bounds on the non-averaged iterates

The proof of Theorem 1 involves first analyzing the non-averaged iterates. Since the upper bound established in this step is of independent interest, we state and discuss it here:

Proposition 1. *Under Assumptions 1—3, there are universal positive constants* (c_0, c_1) *such that for any integer* $p \in \{1\} \cup [\log n, \bar{p}/2]$*, scalar* $\tau \geq 2pt_{\mathrm{mix}} \log(c_0 d)$*, and positive stepsize* $\eta \in \big(0, \frac{1-\kappa}{2cp^3(\sigma_L^2 d + \gamma_{\max}^2)\tau}\big]$*, we have*

$$\big(\mathbb{E}\|\theta_t - \bar{\theta}\|_2^{2p}\big)^{1/p} \leq e^{-\frac{1}{2}\eta(1-\kappa)t}\big(\mathbb{E}\|\theta_0 - \bar{\theta}\|_2^{2p}\big)^{1/p} + \frac{cp^3\eta}{1-\kappa}\bar{\sigma}^2\tau d \tag{22}$$

for all $t = 1, \ldots, n$*.*

See Section 5 for the proof of this proposition.

Note that the guarantees on the unaveraged iterates in Proposition 1—unlike those of Theorem 1 for the averaged iterates—do not match the optimal instance-dependent behavior. This is to be expected, since at least asymptotically, the unaveraged sequence converges to a Gaussian random vector with covariance specified by the solution of a Riccati equation. (For details, see Section 4.5.3 of the book [BMP12]). This covariance term need not match the optimal statistical error.

On the other hand, by choosing $\eta \asymp \frac{\log n}{(1-\kappa)n}$, the bound in Proposition 1 matches the worst-case bound in equation (21), up to log factors. We also note that in Proposition 1, the exponent p can take values in two ranges: regardless of the value of $\bar{p} \in [2, \infty]$, one can always take $p = 1$ and obtain an upper bound on the mean-squared error $\mathbb{E}\big[\|\theta_t - \bar{\theta}\|_2^2\big]$. This bound only requires Assumption 2 to hold true with $\bar{p} \geq 2$, which covers many important examples (see Section 4). On the other hand, when Assumption 2 is satisfied with $\bar{p} \geq 2\log n$ and a stronger moment assumption is imposed, one can obtain a p-th moment bound for any $p \geq [2\log n, \bar{p}]$. This bound can be readily converted into a high-probability bound for the last iterate of stochastic approximation. It is worth noting that we study these two cases separately, using slightly different proof techniques.

It is worthwhile making some comparisons between Proposition 1 and existing results on the unaveraged forms of Markovian stochastic approximation. As we have noted in our examples, in many cases, the quantities $(\sigma_L, \sigma_b, \bar{\sigma})$ do not depend on the dimension, in which case the error bound in Proposition 1 grows linearly with dimension d. In comparison, in terms of our notation, the error bounds in the papers [BRS18, SY19] both exhibit quadratic dependency on the quantity $\frac{\max_{s\in\mathbb{X}} \|\boldsymbol{L}_t(s)\|_{\mathrm{op}}}{1-\kappa}$. As we noted previously in equation (8), this quantity scales linearly in dimension when the observations have a constant rank (independent of dimension), so that (even after optimal parameter tuning), the bounds from these parameters scale at least proportionally to $\frac{d^2}{n}$. This scaling should be contrasted with the $\mathcal{O}(d/n)$ guarantees from our bounds. On the other hand, the analysis in the paper [DMN+21] involves a different mixing assumption, and so is not directly comparable to our results. However, it is worth noting that their bound $\|\theta_t - \bar{\theta}\|_2$ also has an explicit $\mathcal{O}\left(d/\sqrt{n}\right)$ term (cf. equation (32) in their paper), showing that the MSE bound grows quadratically with dimension.

3.2 Local minimax lower bounds

Thus far, we established instance-dependent upper bounds for the averaged SA scheme with Markov noise. It is natural to wonder whether these bounds can be improved. Answering this question requires the development of local minimax lower bounds, which we describe in this section.

3.2.1 Set-up and local neighborhoods

We begin with the set-up and the definition of local neighborhoods for our lower bounds. Let P be an irreducible Markov transition kernel on a finite state space $\mathbb{X}$ with associated stationary measure ξ_P. Consider the solution $\bar{\theta}(P)$ to the following fixed-point equation

$$\bar{\theta}(P) = \mathbb{E}_{\xi_P}\big[\boldsymbol{L}(s)\big] \cdot \bar{\theta}(P) + \mathbb{E}_{\xi_P}\big[\boldsymbol{b}(s)\big]. \tag{23}$$

where the maps $\boldsymbol{b}$ and $\boldsymbol{L}$ are known to the estimator, whereas the Markov transition kernel is unknown. For some fixed P_0 with stationary measure ξ_0, we would like to lower bound the number of observations required to estimate $\bar{\theta}(P_0)$ to a given accuracy. In order to obtain such a lower bound, we consider the fixed point problem (23) over a local neighborhood[6] of the pair (P_0, ξ_0). We assume that the estimator is based on a Markov trajectory $\{s_t\}_{t=0}^n$, with initial state s_0 drawn according to the original[7] stationary distribution ξ_0, and successive states evolving according to the transition kernel P.

In order to quantify the complexity of estimation localized around the Markov transition kernel P_0, we define the following two notions of local neighborhood:

$$\mathfrak{N}_{\text{Prob}}\big(P_0, \varepsilon\big) := \big\{P : \sum_{x \in \mathbb{X}} \xi_0(x) \cdot \chi^2\left(P(x, \cdot) \;||\; P_0(x, \cdot)\right) \leq \varepsilon^2\big\}, \tag{24a}$$

$$\mathfrak{N}_{\text{Est}}\big(P_0, \varepsilon\big) := \big\{P : \|\bar{\theta}(P) - \bar{\theta}(P_0)\|_2 \leq \varepsilon\big\}. \tag{24b}$$

The two notions of neighborhood focus on different types of locality restrictions on the model class: the local problem class $\mathfrak{N}_{\text{Prob}}$ contains all the Markov transition kernels that are "globally close" to a given kernel P_0, measured by a weighted χ^2 divergence. It is worth noting that this weighted χ^2 divergence has an operational interpretation. Suppose we draw $x \sim \xi_0$, and then draw the next state $y \sim P_0(x, \cdot)$ accordingly the original Markov kernel P_0, as well as $y' \sim P(x, \cdot)$ under the kernel P. Then the weighted χ^2 divergence is the χ^2 divergence between the joint laws of (x, y) and (x, y').

On the other hand, the local class $\mathfrak{N}_{\text{Est}}$ contains Markov transition kernels P such that the solution $\bar{\theta}(P)$ to the fixed-point equation (23) lies in a local neighborhood of the given solution $\bar{\theta}(P_0)$, measured by the Euclidean distance. This problem class captures the complexity specifically for solving the fixed-point equation, without the need to estimate the entire transition kernel. In particular, it is easy to construct a Markov kernel P such that the solution $\bar{\theta}(P)$ is very close to $\bar{\theta}(P_0)$, but the distance between the transition kernels P and P_0 (e.g. measured in weighted χ^2 divergence) is arbitrarily large.

3.2.2 Instance-dependent lower bound

Our lower bound is proved on the smallest worst-case risk attainable over the intersection of $\mathfrak{N}_{\text{Prob}}$ and $\mathfrak{N}_{\text{Est}}$. We use the shorthand notation $\bar{L}^{(0)} := \mathbb{E}_{\xi_0}\big[\boldsymbol{L}(s)\big]$. Also recall the covariance matrix $\Sigma^*_{\text{Mkv}} = \sum_{t=-\infty}^{\infty} \mathbb{E}\big[\varepsilon_{\text{Mkv}}(\widetilde{s}_t)\varepsilon_{\text{Mkv}}(\widetilde{s}_0)^\top\big]$, as previously defined in equation (19), for a stationary trajectory $(\widetilde{s}_t)_{t \in \mathbb{Z}}$ under the transition kernel P_0. Our bound depends on the *local radius*

$$\varepsilon_n = n^{-1/2} \sqrt{\text{trace}\left((I - \bar{L}^{(0)})^{-1} \Sigma^*_{\text{Mkv}} (I - \bar{L}^{(0)})^{-\top}\right)}, \tag{25}$$

[6] Doing so is necessary to rule out trivial estimators, and the possibility of super-efficiency.

[7] In our construction, both kernels P_0 and P are rapidly mixing and their stationary measure are sufficiently close in TV distance that the choice of initial distribution does not affect the result. Drawing $s_0 \sim \xi_0$ is made for theoretical convenience.

which is the contribution of Markovian noise to the upper bound stated in Theorem 1.

We are now ready to state our lower bound. Recall that we have assumed that the kernel P_0 is irreducible and aperiodic. We also assume the mixing condition (Assumption 1) holds with the discrete metric $\rho(x,y) = \mathbf{1}_{\{x \neq y\}}$ and mixing time t_{mix}, and that $\mathrm{supp}\big(P_0(s,\cdot)\big) \geq 2$ for all $s \in \mathbb{X}$.

Theorem 2. *Under the assumptions stated above, there exist universal positive constants (c, c_1, c_2) such that for any sample size n lower bounded as*

$$n \geq \frac{c t_{\mathrm{mix}}^2 \sigma_L^2 d^2 \log^2 d}{(1-\kappa)^2}, \quad \text{and} \quad n^2 \varepsilon_n^2 \geq \frac{2c(1+\sigma_L^2)\bar{\sigma}^2 t_{\mathrm{mix}}^4 d^2}{(1-\kappa)^4} \log^6 \big(\tfrac{d}{\min_s \xi_0(s)}\big), \tag{26a}$$

we have the minimax lower bound

$$\inf_{\widehat{\theta}_n} \sup_{P \in \mathfrak{N}'} \mathbb{E}\big[\|\widehat{\theta}_n - \bar{\theta}(P)\|_2^2\big] \geq c_2 \varepsilon_n^2, \tag{26b}$$

where $\mathfrak{N}' := \mathfrak{N}_{\mathrm{Prob}}\big(P_0, c_1\sqrt{\tfrac{d}{n}}\big) \cap \mathfrak{N}_{\mathrm{Est}}(P_0, c_1\varepsilon_n)$.

See Section 7 for the proof of this theorem.

A few remarks are in order. First, note that the minimax lower bound is with respect to the problem class $\mathfrak{N}_{\mathrm{Prob}}\big(P_0, c_1\sqrt{\tfrac{d}{n}}\big) \cap \mathfrak{N}_{\mathrm{Est}}(P_0, c_1\varepsilon_n)$, which requires both the transition kernel P and the solution $\bar{\theta}(P)$ to be close to the given problem instance $(P_0, \bar{\theta}(P_0))$. The size of the weighted χ^2 neighborhood scales with the standard parametric rate $\sqrt{d/n}$, as desired in such problems. On the other hand, the size of the neighborhood around $\bar{\theta}(P_0)$ is proportional to the local radius ε_n that appears in the lower bound. Operationally, this result indicates that even if the estimator knows in advance that $\bar{\theta}(P)$ lies in the ball $\mathbb{B}(\bar{\theta}(P_0), c_1\varepsilon_n)$, one cannot do much better than simply outputting an arbitrary point in this ball without looking at the data.

Second, it should be noted that quantity ε_n^2 matches (up to a constant factor) the optimal mean-squared error given by the local asymptotic minimax theorem [vdV00, GW95]. In contrast to such asymptotic theory, however, Theorem 2 applies when n is finite, and does not impose any regularity assumptions on the estimator. Furthermore, the radius ε_n that is used to define the local neighborhood $\mathfrak{N}_{\mathrm{Est}}(P_0, \varepsilon_n)$ is optimal in the following sense. On the one hand, since the plug-in estimator is asymptotically normal [GW95], for any decreasing sequence ε_n' such that $\varepsilon_n' > \varepsilon_n$ and $\varepsilon_n' \to 0^+$, the minimax risk within the neighborhood $\mathfrak{N}_{\mathrm{Est}}(P_0, \varepsilon_n')$ behaves asymptotically as ε_n^2 up to constant factors. On the other hand, for any decreasing sequence ε_n' such that $\varepsilon_n' < \varepsilon_n$, the minimax risk in the neighborhood $\mathfrak{N}_{\mathrm{Est}}(P_0, \varepsilon_n')$ is at most ε_n'. In the latter case, the neighborhood is so small that it provides more information than the data provides.

Theorem 2 matches the Markov noise term in Theorem 1, establishing its optimality when the martingale part of the noise vanishes, i.e., $\boldsymbol{L}_t(s) = \boldsymbol{L}(s)$ and $\boldsymbol{b}_t(s) = \boldsymbol{b}(s)$. The lower bound does not capture the martingale part of the noise because we assume that the functions $L : \mathbb{X} \to \mathbb{R}^{d \times d}$ and $b : \mathbb{X} \to \mathbb{R}^d$ are known to the estimator. In the setting where these functions are also observed only through noisy i.i.d. data (L_t, b_t), Theorem 3 of the paper [MPW20] implies a lower bound of the form $c_2 n^{-1} \operatorname{trace}\big((I - \bar{L}^{(0)})^{-1} \Sigma_{\mathrm{MG}}^* (I - \bar{L}^{(0)})^{-\top}\big)$. Combining it with Theorem 2 implies a minimax lower bound involving the term $c_2' n^{-1} \operatorname{trace}\big((I - \bar{L}^{(0)})^{-1} (\Sigma_{\mathrm{Mkv}}^* + \Sigma_{\mathrm{MG}}^*)(I - \bar{L}^{(0)})^{-\top}\big)$ in a properly defined local neighborhood,

thus establishing the optimality of Theorem 1. Finally, we note that Theorem 2 requires the sample size to be at least $t_{\text{mix}}^2 d^2$, which is more stringent than the $\mathcal{O}(t_{\text{mix}} d)$ requirement in the upper bound. While Theorem 1 holds true with a linear sample-size $n = \mathcal{O}(d)$, it is only shown to be instance-optimal for larger $n = \Omega(d^2)$. This mismatch is due to the fact that small perturbations of the Markov transition kernel in certain directions can destroy its fast mixing property. That being said, Theorem 2 is still a finite-sample result, with polynomial dependency on the quantities $\big(t_{\text{mix}}, d, \frac{1}{1-\kappa}\big)$, and poly-logarithmic dependency on the smallest stationary probability.

4 Some consequences for specific problems

In this section, we specialize our analysis to the examples described in Section 2.2, namely approximate policy evaluation using TD algorithms, and estimation in autoregressive time series models. By verifying the conditions needed to apply Theorem 1 and Proposition 1, we obtain some more concrete corollaries of our general theory.

4.1 TD(0) method

Recall the TD(0) algorithm for policy evaluation, as previously described in Example 1. We are interested in estimating the solution V^* of the Bellman equation (10), and an approximation scheme is employed using the basis functions $(\phi_j)_{j=1}^d$. Using the shorthand $\langle \theta, \phi(s) \rangle = \sum_{j=1}^d \theta_j \phi_j(s)$ for the Euclidean inner product in $\mathbb{R}^d$, with observation model $(\boldsymbol{L}_{t+1}(\omega_t), \boldsymbol{b}_{t+1}(\omega_t))$ defined in Eq (13), the averaged SA procedure (3) is given by:

$$\theta_{t+1} \overset{(a)}{=} \theta_t - \eta\big\{\langle \phi(s_t) - \gamma\phi(s_{t+1}), \theta_t \rangle - R_{t+1}(s_t)\big\}\phi(s_t), \quad \text{and} \quad \widehat{\theta}_n \overset{(b)}{=} \frac{1}{n-n_0}\sum_{t=n_0}^{n-1} \theta_t. \tag{27}$$

To be clear, the update (27)(a) is the standard TD(0) algorithm with stepsize η, whereas the addition of the averaging step (27)(b) yields the Polyak–Ruppert averaged version of the scheme. Note that we re-scale the stepsize η by a factor of β for notational convenience. In the following subsections, we derive corollaries of our general theory for the averaged scheme under different mixing conditions on the underlying Markov chain.

4.1.1 Markov chains with mixing in total variation distance

We first assume that the Markov chain satisfies a mixing condition (cf. Assumption 1) in the discrete metric: i.e., after t_{mix} steps, we have $d_{\text{TV}}(\delta_s P^{t_{\text{mix}}}, \delta_{s'} P^{t_{\text{mix}}}) \leq \frac{1}{2}$ for any pair $s, s' \in \mathbb{X}$. Let ξ denote the stationary distribution of the Markov chain that generates the trajectory $\{s_t\}_{t \geq 0}$, and let P denote its transition kernel. Note that the augmented state vector $\omega_t = (s_t, s_{t+1})$ evolves according to a Markov process with mixing time $t_{\text{mix}} + 1$. Moreover, the stationary distribution of the pair $\omega = (s, s^+)$ has the form $s \sim \xi$, $s^+ \sim P(\cdot \mid s)$. We denote the stationary covariance of the feature vectors as $B := \mathbb{E}_{s\sim\xi}\big[\phi(s)\phi(s)^\top\big]$, and also define the minimum and maximum eigenvalues $\mu := \lambda_{\min}(B)$ and $\beta := \lambda_{\max}(B)$. We assume that

$$\|B^{-1/2}\phi(s)\|_2 \overset{(a)}{\leq} \varsigma\sqrt{d} \quad \text{and} \quad |R_t(s)| \overset{(b)}{\leq} \varsigma \quad \text{for all } s \in \mathbb{X}, \text{ and} \tag{28a}$$

$$\mathbb{E}_\xi\big[\langle B^{-1/2}\phi(s), u\rangle^4\big] \leq \varsigma^4 \quad \text{for all } u \in \mathbb{S}^{d-1}. \tag{28b}$$

In order to state our result, we define the following quantities:

$$M := \gamma B^{-1/2} \cdot \mathbb{E}_{s\sim\xi, s^+\sim P(s,\cdot)}\big[\phi(s)\phi(s^+)^\top\big] \cdot B^{-1/2},$$
$$\varepsilon_{\mathrm{Mkv}}(s, s^+) := B^{-1/2}\big(\phi(s)^\top\bar{\theta} - \gamma\phi(s^+)^\top\bar{\theta} - r(s)\big)\phi(s), \quad \varepsilon_{\mathrm{MG}}(s) := B^{-1/2}(R(s) - r(s))\phi(s)$$

We also define the following covariance matrices according to Eq (19):

$$\Sigma^*_{\mathrm{Mkv}} := \sum_{t=-\infty}^{\infty} \mathbb{E}\big[\varepsilon_{\mathrm{Mkv}}(s_t, s_{t+1})\varepsilon_{\mathrm{Mkv}}(s_0, s_1)^\top\big],$$
$$\Sigma^*_{\mathrm{MG}} := \mathbb{E}_{s\sim\xi}\big[\mathbb{E}\big[\varepsilon_{\mathrm{MG}}(s)\varepsilon_{\mathrm{MG}}(s)^\top \mid s\big]\big].$$

Finally, we define the quantity

$$\bar{\sigma}^2 := \varsigma^2 \cdot \sqrt{\mathbb{E}\big[\big(\phi(s_t)^\top\bar{\theta} - \gamma\phi(s_{t+1})\bar{\theta} - R_t(s_t)\big)^4\big]}, \tag{29}$$

and let $\kappa := \frac{1}{2}\lambda_{\max}(M + M^\top)$. It is easy to see that $\kappa \le \gamma < 1$. Assuming that $\mu > 0$, we are then ready to state our main result for the TD(0) method.

Corollary 1. *Under the setup above, take the stepsize η and burn-in period n_0 as*

$$\eta = \tfrac{1}{c\beta((\varsigma^4+1)d(1-\kappa)n^2 t_{\mathrm{mix}})^{1/3}}, \quad \text{and} \quad n_0 = \tfrac{1}{2}n, \tag{30}$$

and suppose that $\frac{n}{\log^3 n} \ge \frac{2t_{\mathrm{mix}}(\varsigma^4+1)d\beta^2}{(1-\kappa)^2\mu^2}$. The estimator $\widehat{V}_n := \widehat{\theta}_n\phi$ obtained from the Polyak–Ruppert procedure (27) satisfies the bound

$$\mathbb{E}\big[\|\widehat{V}_n - \bar{V}\|^2_{\mathbb{L}^2(\mathbb{X},\xi)}\big] \le \tfrac{c}{n}\mathrm{Tr}\big\{(I_d - M)^{-1}(\Sigma^*_{\mathrm{Mkv}} + \Sigma^*_{\mathrm{MG}})(I_d - M)^{-\top}\big\} + c\big(\tfrac{\beta^2\bar{\sigma}^2 dt_{\mathrm{mix}}}{\mu^2(1-\kappa)^2 n}\big)^{4/3}\log^2 n, \tag{31}$$

where $\bar{V}$ is the solution to the projected fixed-point equation (11) and $c > 0$ is a universal constant.

See Appendix E.1 for the proof of this corollary.

A few remarks are in order. First, we measure the estimation error in the canonical $\|\cdot\|_{\mathbb{L}^2(\mathbb{X},\xi)}$ norm, instead of the Euclidean distance in $\mathbb{R}^d$. Consequently, the proof of this corollary actually uses a generalized version of Theorem 1 proved for weighted ℓ^2 norms. On the other hand, we note that the error bound (31) is with respect to the solution $\bar{V}$ to the projected fixed-point equation. In the well-specified case where $V^* \in \mathbb{S}$, this solution coincides with the value function V^*. In general, the approximation error needs to be taken into account, and was the focus of the paper [MPW20]. In conjunction with this result, Corollary 1 implies the error bound

$$\mathbb{E}\big[\|\widehat{V}_n - V^*\|^2_{\mathbb{L}^2(\mathbb{X},\xi)}\big] \le c\big[1 + \lambda_{\max}\big((I_d - M)^{-1}(\gamma^2 I_d - MM^\top)(I_d - M)^{-\top}\big)\big] \inf_{V\in\mathbb{S}} \|V - V^*\|^2_{\mathbb{L}^2(\mathbb{X},\xi)}$$
$$+ \tfrac{c}{n}\mathrm{Tr}\big\{(I_d - M)^{-1}(\Sigma^*_{\mathrm{Mkv}} + \Sigma^*_{\mathrm{MG}}))(I_d - M)^{-\top}\big\} + c\big(\tfrac{\beta^2\bar{\sigma}^2 dt_{\mathrm{mix}}}{\mu^2(1-\kappa)^2 n}\big)^{4/3}\log^2 n. \tag{32}$$

In Section 4.2 to follow, we provide a general recipe to trade off approximation and estimation errors to choose the value of λ in the class of TD(λ) algorithms. Before that, we discuss two extensions of Corollary 1.

4.1.2 Markov chains with mixing in Wasserstein metric

Note that for Corollary 1, the mixing time condition is imposed with total variation distance. When the state space $\mathbb{X}$ is continuous, e.g., the set $\mathbb{X}$ is a subset of $\mathbb{R}^m$, mixing in Wasserstein distance could capture the geometry of the underlying metric better. In this section, we extend our analysis to such settings, highlighting the dimension dependency in the sample complexity.

Concretely, we consider a Markov chain $(s_t)_{t\geq 0}$ on a compact domain $\mathbb{X} \subseteq \mathbb{R}^m$, and a feature mapping $\phi : \mathbb{X} \to \mathbb{R}^d$. We assume that the Markov chain admits a unique stationary measure ξ, and the mixing time assumption holds in Wasserstein-1 distance, so that $\mathcal{W}_1\big(\delta_x P^{t_{\mathrm{mix}}}, \delta_y P^{t_{\mathrm{mix}}}\big) \leq \frac{1}{2}\|x-y\|_2$ for all $x, y \in \mathbb{X}$. For the sake of normalization, we assume that $\mathbb{X} \subseteq \mathbb{B}(0,1)$ and $\phi(0) = 0$. On the feature mapping ϕ, we assume the following:

$$\exists \mu, \beta > 0, \quad \mu I_d \preceq B := \mathbb{E}_{s\sim\xi}\big[\phi(s)\phi(s)^\top\big] \preceq \beta I_d, \tag{33a}$$

$$\forall x, y \in \mathbb{X}, \quad \|B^{-1/2}\big(\phi(x) - \phi(y)\big)\|_2 \leq \varsigma\sqrt{d}\|x-y\|_2, \tag{33b}$$

$$\forall u \in \mathbb{S}^{d-1}, \quad \mathbb{E}_{s\sim\xi}\big[\langle u,\, B^{-1/2}\phi(s)\rangle^4\big] \leq \varsigma^4, \tag{33c}$$

$$\forall s, s' \in \mathbb{X},\ t \geq 1, \quad |R_t(s) - R_t(s')| \leq \varsigma\|s - s'\|_2,\ |R_t(s)| \leq \varsigma \quad \text{a.s.} \tag{33d}$$

Here, we regard the parameters (ς, μ, β) as dimension-independent positive constants. Since the state space $\mathbb{X}$ has diameter bounded by 2, the feature mapping ϕ satisfying equation (33a) necessarily has Lipschitz constant of order $\mathcal{O}\left(\sqrt{d}\right)$. For a simple example, take the state x itself as the feature vector (after appropriate re-scaling), which corresponds to the case of $m = d$ and $\phi(x) = \sqrt{d} \cdot x$.

With this set-up, we have the following guarantee:

Corollary 2. *Assuming the conditions in equation (33), taking stepsize and burn-in period as equation (30), for the Polyak–Ruppert averaged stochastic approximation procedure (27), the bound (31) holds.*

See Appendix E.2 for the proof.

Corollary 2 shows that the same instance-dependent bound holds true for a continuous state space setting. Such a bound is useful for many applications, one of which is the case of quadratic value functions, where the dimension satisfies the relation $d = m^2$ the mapping ϕ takes the form $\phi : x \mapsto m \cdot xx^\top$. Assuming that the process $(s_t)_{t\geq 0}$ is supported in a unit ball $\mathbb{B}(0,1)$ and has well-conditioned stationary covariance, it is easy to verify that Assumptions (33) are satisfied with dimension-free constants (ς, μ, β). This example is particularly useful for policy evaluation in Linear Quadratic Regulators (LQR). Nevertheless, our results hold more generally for any random dynamical system that is rapidly mixing in the $\mathcal{W}_1$ distance.

4.1.3 Analysis of a sieve estimator

The optimal dimension dependency in Theorem 1 allows us to obtain optimal estimators for various classes of non-parametric problems, in which the dimension is a parameter to be chosen. In particular, sieve methods are a class of non-parametric estimators based on nested sequences of finite-dimensional approximations. In this section, we analyze the behavior of a stochastic approximation sieve estimator in the Markovian setting. The optimal dimension

dependence in our theorem recovers the minimax optimal rates for estimation, while our instance-dependent bounds help in capturing more refined structure in the problem instance.

Concretely, assuming that the Hilbert space $\mathbb{L}^2(\mathbb{X}, \xi)$ is separable, let $(\phi_j)_{j=1}^{\infty}$ be a set of (not necessarily orthogonal) basis functions. We consider the case where the mixing condition holds true with total variation distance[8]. The following assumptions are imposed on the basis functions:

$$\forall j \in \mathbb{N}^+, \quad \sup_{x \in \mathbb{X}} |\phi_j(x)| \leq \varsigma, \tag{34a}$$

$$\forall d \in \mathbb{N}^+, \quad \mu I_d \leq \big[\mathbb{E}_{s\sim\xi}\big(\phi_j(s)\phi_\ell(s)\big)\big]_{j,\ell\in[d]} \leq \beta I_d, \tag{34b}$$

$$\forall t \geq 1, \quad \sup_{x \in \mathbb{X}} |R_t(x)| \leq \varsigma. \tag{34c}$$

The first assumption is standard in nonparametric regression, and satisfied by many useful basis functions such as the Fourier basis and Walsh-Hadamard basis. The second assumption relaxes the orthogonality requirement on the bases, by only requiring the Gram matrix to be well-conditioned.

We define the noise level $\bar{\sigma}$ using the second moment:

$$\bar{\sigma}^2 := \varsigma^2 \cdot \sqrt{\mathbb{E}\big[\big(\bar{V}(s_t) - \gamma \bar{V}(s_{t+1}) - R_t(s_t)\big)^2\big]}. \tag{35}$$

Once again, we run the averaged stochastic approximation procedure (27) on this problem. A crucial point of departure from the parametric models discussed above is that the number of basis functions d_n in sieve estimators is chosen based on the problem structure and sample size. Let $\mathbb{S}(d_n) := \mathrm{span}(\phi_1, \phi_2, \cdots, \phi_{d_n})$ denote the subspace spanned by the first d_n basis functions. The following result is a direct corollary of our theorem, and covers the case of fixed d_n; we discuss the trade-off between approximation and estimation error in the choice of d_n presently.

Corollary 3. *Assuming the conditions in equation (34), take the stepsize and burn-in period as in equation (30). Assuming that $\mu, \beta, \varsigma \asymp 1$, the Polyak–Ruppert averaged stochastic approximation procedure (27) satisfies the bound (31) with $d = d_n$.*

See Appendix E.3 for the proof.

Recall that by taking into account the approximation error, the error for estimating the true value function V^* takes the following form:

$$\begin{aligned}\mathbb{E}\big[\|\widehat{V}_n - V^*\|^2_{\mathbb{L}^2(\mathbb{X},\xi)}\big] \leq c\big[1 + \lambda_{\max}\big((I-M)^{-1}(\gamma^2 I_d - MM^\top)(I-M)^{-\top}\big)\big] \inf_{V\in\mathbb{S}} \|V - V^*\|^2_{\mathbb{L}^2(\mathbb{X},\xi)} \\ + \tfrac{c}{n}\mathrm{Tr}\big((I-M)^{-1}(\Sigma^*_{\mathrm{Mkv}} + \Sigma^*_{\mathrm{MG}})(I-M)^{-\top}\big) + c\big(\tfrac{\bar{\sigma}^2 t_{\mathrm{mix}} d_n}{(1-\kappa)^2 n}\big)^{4/3} \log^2 n.\end{aligned}$$

Let $\{\psi_j\}_{j=1}^{+\infty}$ be an orthonormal basis of $\mathbb{L}^2(\mathbb{X}, \xi)$ such that $\mathrm{span}(\psi_1, \cdots, \psi_d) = \mathrm{span}(\phi_1, \cdots, \phi_d)$ for any $d \geq 1$. (For instance, one can let $\{\psi_j\}_{j=1}^{+\infty}$ be the Gram-Schmidt orthonormalization of the original basis functions). Given a non-increasing sequence $\{\alpha_j\}_{j=1}^{\infty}$ of positive reals such that $\lim_{j\to+\infty} \alpha_j = 0$, we first let $\mathcal{H}_0$ be a linear subspace of $\mathbb{L}^2(\mathbb{X}, \xi)$, consisting of all the

[8]By following the approach in the previous subsection, the analysis can also be extended to the case of mixing in Wasserstein distance.

finite linear combination of basis vectors $\{\psi_j\}_{j=1}^{+\infty}$, equipped with the following inner product:

$$\forall u, v \in \mathcal{H}_0, \quad \langle u, v\rangle_{\mathcal{H}_0} := \sum_{j=1}^{\infty} \alpha_j^{-1} \cdot \langle u, \psi_j\rangle \cdot \langle v, \psi_j\rangle.$$

Note that the summation shown above is actually finite, since both both sequences $(\langle u, \psi_j\rangle)_{j=1}^{+\infty}$, $(\langle v, \psi_j\rangle)_{j=1}^{+\infty}$ only have finite non-zero entries. We then define the inner product space $(\mathcal{H}, \langle\cdot, \cdot\rangle_{\mathcal{H}})$ as the completion of $(\mathcal{H}_0, \langle\cdot, \cdot\rangle_{\mathcal{H}_0})$. It is easy to see that $\mathcal{H}$ is a Hilbert space, and a linear subspace of $\mathbb{L}^2(\mathbb{X}, \xi)$.

For any $V^* \in \mathcal{H}$, the estimation error is at most (in the worst-case)

$$\mathbb{E}\big[\|\widehat{V}_n - V^*\|^2_{\mathbb{L}^2(\mathbb{X},\xi)}\big] \leq \tfrac{c}{1-\gamma} \cdot \alpha_{d_n} \|V^*\|^2_{\mathcal{H}} + \tfrac{c\bar{\sigma}^2 d_n t_{\mathrm{mix}}}{(1-\gamma)^2 n}. \tag{36}$$

For example, when the eigenvalues of Hilbert space $\mathcal{H}$ decay as $\alpha_j \asymp j^{-2s}$ for some $s > 0$, the estimator achieves a rate of $\mathcal{O}\big((t_{\mathrm{mix}}/n)^{\frac{2s}{2s+1}}\big)$, which matches the minimax optimal rate proved by Duan et al. [DWW21] in the i.i.d. setting, but with a multiplicative correction to the effective sample size by a factor t_{mix} to accommodate Markovian observations. Furthermore, since one can estimate the quantities $(M, \Sigma^*_{\mathrm{Mkv}}, \Sigma^*_{\mathrm{MG}})$ in the bound (31) using $\mathcal{O}(d)$ samples, instance-dependent model selection can in principle be conducted. Bounds of the form (36) thus open the door to asking important questions of this type.

4.2 TD(λ) methods

Now we turn to stochastic approximation methods for the TD(λ) projected fixed-point equation (14b), with some given $\lambda \in [0, 1)$. With observation model $(\boldsymbol{L}_{t+1}(\omega_t), \boldsymbol{b}_{t+1}(\omega_t))$ given by Eq (16c), the averaged SA procedure (3) can be written as

$$\theta_{t+1} = \theta_t - \eta\Big\{\langle \phi(s_t) - \gamma\phi(s_{t+1})^\top, \theta_t\rangle - R_t(s_t)\Big\} g_t, \quad \text{where} \tag{37a}$$

$$g_t = \gamma\lambda g_{t-1} + \phi(s_t) \quad \text{and,} \tag{37b}$$

$$\widehat{\theta}_n = \tfrac{1}{n-n_0} \sum_{t=n_0}^{n-1} \theta_t. \tag{37c}$$

The update on g_t is the so-called "eligibility trace" in the TD(λ) algorithm. As before, we assume the two bounds in equation (28a), and assume that the mixing time condition 1 holds true for the chain $(s_t)_{t\geq 1}$, with discrete metric and mixing time t_{mix}. We consider the augmented Markov chain $\omega_t := \big(s_t, s_{t+1}, \frac{1-\gamma\lambda}{\varsigma\sqrt{\beta d}} g_t\big) \in \mathbb{X}^2 \times \mathbb{B}(0, 1)$ and begin by establishing mixing conditions on this augmented chain.

Proposition 2. *Under the setup above, consider the metric*

$$\rho\big((s_1, s_2, h), (s_1', s_2', h')\big) := \tfrac{1}{4}\big(\mathbf{1}_{s_1 \neq s_1'} + \mathbf{1}_{s_2 \neq s_2'} + \|h - h'\|_2\big). \tag{38a}$$

Taking $\tau = 4\big(t_{\mathrm{mix}} + \frac{1}{1-\gamma\lambda}\big)$, *the augmented chain* $\big\{\omega_t = (s_t, s_{t+1}, \frac{1-\gamma\lambda}{\varsigma\sqrt{\beta d}} g_t)\big\}_{t\geq 0}$ *satisfies the mixing bound*

$$\mathcal{W}_{1,\rho}\big(\mathcal{L}(\omega_\tau), \mathcal{L}(\omega_\tau')\big) \leq \frac{1}{2}\rho\big(\omega_0, \omega_0'\big) \tag{38b}$$

for two chains $(\omega_t)_{t\geq 0}$ *and* $(\omega_t')_{t\geq 0}$ *starting from* ω_0 *and* ω_0', *respectively. In particular, the stationary distribution* $\widetilde{\xi}$ *of the chain* $(\omega_t)_{t\geq 0}$ *exists and is unique.*

See Appendix F.1 for the proof of this proposition.

Taking this proposition as given, we are now ready to present our main corollary for TD(λ) procedures. We consider the following instantiation of quantities in Theorem 1:

The projected linear operator $(1-\lambda)\sum_{k=0}^{+\infty}\lambda^k(\gamma\Pi_{\mathbb{S}}P)^{k+1}$ in the equation (14b) can be represented in the orthonormal basis of the subspace $\mathbb{S}$ as

$$\begin{aligned} M_\lambda &:= I_d - B^{-1/2}\mathbb{E}_{(s,s^+,\frac{1-\gamma\lambda}{\varsigma\sqrt{\beta d}}g)\sim\widetilde{\xi}}\big[g\phi(s)^\top - \gamma g\phi(s^+)^\top\big]B^{-1/2} \\ &= (1-\lambda)B^{-1/2}\sum_{t=0}^{\infty}\lambda^t\gamma^{t+1}\mathbb{E}\big[\phi(s_0)\phi(s_{t+1})\big]B^{-1/2}. \end{aligned}$$

The Markovian and martingale part of the noise (in the low-dimensional subspace $\mathbb{S}$) takes the following form:

$$\begin{aligned} \varepsilon_{\mathrm{Mkv},\lambda}\big(s,s^+,\frac{1-\gamma\lambda}{\varsigma\sqrt{\beta d}}g\big) &= B^{-1/2}\big(\phi(s)^\top\bar{\theta} - \gamma\phi(s^+)\bar{\theta} - r(s)\big)g, \\ \varepsilon_{\mathrm{MG},\lambda}\big(s,s^+,\frac{1-\gamma\lambda}{\varsigma\sqrt{\beta d}}g\big) &= B^{-1/2}(R_0(s)-r(s))g \end{aligned}$$

Finally, we define the covariance matrices $\Sigma^*_{\mathrm{Mkv},\lambda}$ and $\Sigma^*_{\mathrm{MG},\lambda}$ according to Eq (19):

$$\begin{aligned} \Sigma^*_{\mathrm{Mkv},\lambda} &:= \sum_{t=-\infty}^{\infty}\mathbb{E}\big[\varepsilon_{\mathrm{Mkv},\lambda}\big(s_t,s_{t+1},\frac{1-\gamma\lambda}{\varsigma\sqrt{\beta d}}g_t\big)\varepsilon_{\mathrm{Mkv},\lambda}\big(s_0,s_1,\frac{1-\gamma\lambda}{\varsigma\sqrt{\beta d}}g_0\big)^\top\big], \\ \Sigma^*_{\mathrm{MG},\lambda} &:= \mathbb{E}_{s\sim\xi}\big[\mathbb{E}\big[\varepsilon_{\mathrm{MG},\lambda}(s)\varepsilon_{\mathrm{MG},\lambda}(s)^\top \mid s\big]\big]. \end{aligned}$$

As before, we let $\beta := \lambda_{\max}(B)$, $\mu := \lambda_{\min}(B)$ and $\kappa_\lambda := \frac{1}{2}\lambda_{\max}(M_\lambda + M_\lambda^\top)$, and define the quantity $\bar{\sigma}$ according to equation (29). Note that a straightforward calculation reveals that $\kappa_\lambda \le \frac{(1-\lambda)\gamma}{1-\lambda\gamma} < 1$. Assuming that $\mu > 0$, we are then ready to state our main result for TD(λ) methods.

Corollary 4. *Under the setup above, take the stepsize and burn-in period as*

$$\eta = \frac{(1-\gamma\lambda)^{2/3}}{c\beta\big((\varsigma^4+1)d(1-\kappa_\lambda)n^2\big(t_{\mathrm{mix}}+\frac{1}{1-\gamma\lambda}\big)\big)^{1/3}}, \quad \text{and} \quad n_0 = \tfrac{1}{2}n, \tag{39a}$$

and suppose that $\frac{n}{\log^3 n} \ge \frac{2(t_{\mathrm{mix}}+\frac{1}{1-\gamma\lambda})\,(\varsigma^4 d+1)\beta^2}{(1-\kappa_\lambda)^2(1-\gamma\lambda)^2\mu^2}$. *Then the value function estimate* $\widehat{V}_n(s) := \langle\widehat{\theta}_n, \phi(s)\rangle$ *obtained from the Polyak–Ruppert procedure* (37) *has MSE bounded as*

$$\begin{aligned} \mathbb{E}\big[\|\widehat{V}_n - \bar{V}^{(\lambda)}\|^2_{\mathbb{L}^2(\mathbb{X},\xi)}\big] \le cn^{-1}\mathrm{Tr}\big((I_d-M_\lambda)^{-1}(\Sigma^*_{\mathrm{Mkv}}+\Sigma^*_{\mathrm{MG}})(I_d-M_\lambda)^{-\top}\big) \\ + c\big(\tfrac{\beta^2\bar{\sigma}^2 d\left(t_{\mathrm{mix}}+\frac{1}{1-\gamma\lambda}\right)}{\mu^2(1-\kappa_\lambda)^2(1-\gamma\lambda)^2 n}\big)^{4/3}\log^2 n, \end{aligned} \tag{39b}$$

where $\bar{V}^{(\lambda)}$ *is the solution to the projected fixed-point equation* (11).

See Appendix F.2 for the proof of this corollary.

A few remarks are in order. First, using the same argument as in Corollaries 2 and 3, one can extend the results for TD(λ) to the cases of continuous state spaces with Wasserstein mixing, as well as to nonparametric sieve estimators. As is well-known, different choices of the tuning parameter λ interpolate the "temporal difference" method, in which we aim at solving the Bellman equation, and the "Monte Carlo" method, in which the value function is estimated directly by averaging the rollout of a Markovian trajectory. For example, on the one hand, letting $\lambda = 0$ recovers the instance-dependent upper bound for TD(0) method in Corollary 1. On the other hand, by taking $\lambda = \gamma$, we have $\kappa_\lambda \le \frac{\gamma}{1+\gamma} \le \frac{1}{2}$, and the dependence on the discount factor γ appears only through the variance of the noise, instead of through the conditioning of the matrix M_λ. In the next section, we sketch a recipe for the instance-dependent selection of λ that also takes the approximation error into account.

4.2.1 Using instance-dependent results to select λ

Recall that the TD(λ) algorithm aims at estimating the solution $\bar{V}^{(\lambda)}$ to the projected fixed-point equation (14b). The linear operator in the unprojected fixed-point equation (14a) satisfies the norm bound

$$\|\!|(1-\lambda)\sum_{k=0}^{\infty}\lambda^k\gamma^{k+1}P^{k+1}\|\!|_{\mathbb{L}^2(\mathbb{X},\xi)\to\mathbb{L}^2(\mathbb{X},\xi)} \le (1-\lambda)\sum_{k=0}^{\infty}\lambda^k\gamma^{k+1} = \tfrac{(1-\lambda)\gamma}{1-\lambda\gamma}.$$

Consequently, invoking Theorem 1 of the paper [MPW20], the approximation error satisfies the bound

$$\|\bar{V}^{(\lambda)} - V^*\|^2_{\mathbb{L}^2(\mathbb{X},\xi)} \le \alpha\big(M_\lambda, \tfrac{(1-\lambda)\gamma}{1-\lambda\gamma}\big)\cdot \inf_{V\in\mathbb{S}}\|V - V^*\|^2_{\mathbb{L}^2(\mathbb{X},\xi)},$$

where $\alpha(M,z) := 1+\lambda_{\max}\big((I_d-M)^{-1}(z^2I_d - MM^\top)(I_d-M)^{-\top}\big)$ is the approximation factor. Combining with Corollary 4, we obtain the following bound on the distance to the true value function:

$$\begin{aligned}\mathbb{E}\big[\|\widehat{V}_n - V^*\|^2_{\mathbb{L}^2(\mathbb{X},\xi)}\big] \le c\alpha\big(M_\lambda, \tfrac{(1-\lambda)\gamma}{1-\lambda\gamma}\big)\cdot\inf_{V\in\mathbb{S}}\|V-V^*\|^2_{\mathbb{L}^2(\mathbb{X},\xi)} + c\Big(\tfrac{\beta^2\bar{\sigma}^2 d\big(t_{\mathrm{mix}}+\frac{1}{1-\gamma\lambda}\big)}{\mu^2(1-\kappa_\lambda)^2(1-\gamma\lambda)^2 n}\Big)^{4/3}\log^2 n \\ + \tfrac{c}{n}\mathrm{Tr}\big((I_d-M_\lambda)^{-1}(\Sigma^*_{\mathrm{Mkv}}+\Sigma^*_{\mathrm{MG}})(I_d-M_\lambda)^{-\top}\big) \quad (40)\end{aligned}$$

for a universal constant $c > 0$.

It can be seen that $\alpha\big(M_\lambda, \frac{(1-\lambda)\gamma}{1-\lambda\gamma}\big) \le c'\frac{1-\lambda\gamma}{1-\gamma}$ for a universal constant. We also recall that $\kappa_\lambda \le \frac{(1-\lambda)\gamma}{1-\lambda\gamma}$. If we take the parameters (μ,β,ς) to be of constant order, in the worst case, the upper bound (40) takes the simplified form

$$\mathbb{E}\big[\|\widehat{V}_n - V^*\|^2_{\mathbb{L}^2(\mathbb{X},\xi)}\big] \le c\frac{1-\lambda\gamma}{1-\gamma}\inf_{V\in\mathbb{S}}\|V-V^*\|^2_{\mathbb{L}^2(\mathbb{X},\xi)} + c\frac{\big(t_{\mathrm{mix}}+\frac{1}{1-\gamma\lambda}\big)d}{(1-\gamma)^3 n}.$$

From such an upper bound, it may appear that the optimal choice of λ is always $\lambda = \gamma \wedge (1 - 1/t_{\mathrm{mix}})$, so that the approximation factor is minimized and the variance remains controlled. However, this choice could be overly conservative, since the actual variance with small λ can be significantly smaller, with the feature vectors still having bounded one-step cross-correlation. Choosing the parameter λ close to 1 cannot take advantage of small one-step correlation. On the other hand, a fine-grained bound of the form (40) can be used to perform instance-dependent model selection, as follows:

- Construct a uniform finite grid $0 = \lambda_1 < \lambda_2 < \cdots < \lambda_m = \gamma$ for possible values of λ.
- For each $\ell \in [m]$, compute the TD(λ_ℓ) estimator, and construct empirical plug-in estimates $\big(\widehat{M}_{\lambda,n}, \widehat{\Sigma^*}_{\mathrm{Mkv},\lambda,n}, \widehat{\Sigma^*}_{\mathrm{MG},\lambda,n}\big)$ for the matrices $\big(M_\lambda, \Sigma^*_{\mathrm{Mkv},\lambda}, \Sigma^*_{\mathrm{MG},\lambda}\big)$ by replacing the expectations by empirical averages. Similarly replace $\bar{\theta}^{(\lambda)}$ by $\widehat{\theta}_n$.
- Estimate the approximation factor $\alpha\big(M_\lambda, \frac{(1-\lambda)\gamma}{1-\lambda\gamma}\big)$ and the covariance $(I_d - M_\lambda)^{-1}(\Sigma^*_{\mathrm{Mkv}} + \Sigma^*_{\mathrm{MG}})(I_d - M_\lambda)^{-\top}$ by plugging in the estimated matrices described above, for each $\lambda = \lambda_\ell$ with $\ell \in [m]$. Based on prior knowledge about the scale of the optimal approximation error $\inf_{V \in \mathbb{S}} \|V - V^*\|^2_{\mathbb{L}^2(\mathbb{X},\xi)}$, select λ_ℓ in the grid that minimizes our estimate of the total error according to equation (40).

Note that the procedure above is simply a sketch; a formal proof of correctness would show bounds that are uniform over all m estimators. It is an important direction of future work to provide sharp non-asymptotic analysis of such a model selection procedure.

4.3 Autoregressive models

Next, we turn to Example 3, the multivariate auto-regressive model. We study the stochastic approximation procedure in which, for any $i \in [k]$, we have

$$A^{(i)}_{t+1} = A^{(i)}_t - \eta\big(\sum_{j=0}^{k-1} A^{(j)}_t X_{t-j} X^\top_{t+1-i} - X_{t+1} X^\top_{t+1-i}\big), \quad \text{and} \quad \widehat{A}^{(i)}_n = \frac{1}{n-n_0}\sum_{t=n_0}^{n-1} A^{(i)}_t.$$

The first step in our analysis is to establish necessary and sufficient conditions for the existence and uniqueness of the stationary distribution of the process (17). The following $km \times km$ matrix plays a crucial role in this context:

$$R_* = \begin{bmatrix} A^*_1 & A^*_2 & \cdots & & A^*_k \\ I_m & 0 & \cdots & & 0 \\ 0 & I_m & 0 & \cdots & 0 \\ 0 & & \ddots & & 0 \\ 0 & \cdots & 0 & I_m & 0 \end{bmatrix}.$$

In the noiseless case, the stability of the linear dynamical system is equivalent to the following *Lyapunov stability condition* (see e.g. [Nem01], Section 3.3):

$$\exists P_* \succ 0, Q_* \succ 0, \quad \text{such that } R_*^\top P_* R_* = P_* - Q_*. \tag{41}$$

Clearly we have $P_* \succ Q_*$. We let $\beta := \lambda_{\max}(P_*)$ and $\mu := \lambda_{\min}(Q_*)$. Based on stability theory for discrete-time linear systems [BD09], condition (41) is necessary for the stationary distribution to exist. In the following proposition, we show that this condition is also sufficient, with a concrete mixing time bound.

Proposition 3. *Under the Lyapunov stability condition (41) and assuming that the noise has bounded first moment $\mathbb{E}\big[\|\varepsilon_t\|_2\big] < \infty$, the stationary distribution $\widetilde{\xi}$ for the sliding window $\omega_t = (X_{t+1}, X_t, \cdots, X_{t-k+1})$ of the auto-regressive process (17) exists and is unique. Furthermore, the mixing assumption 1 is satisfied with Wasserstein distance in $\mathbb{R}^{(k+1)m}$ and a mixing time bound $t_{\mathrm{mix}} = ck + c\frac{\beta}{\mu}\big(1 + \log\frac{\beta}{\mu}\big)$.*

See Section G.1 for the proof of this claim.

In addition to this mixing guarantee, we also make the following assumptions on the noise:

$$\mathbb{E}\big[\varepsilon_t\big] = 0, \quad \sup_{u\in\mathbb{S}^{d-1}} \mathbb{E}\big[\langle u,\, \varepsilon_t\rangle^4\big] \le \varsigma^4, \quad \text{and} \quad \|\varepsilon_t\|_2 \le \varsigma\sqrt{m},\ \text{a.s.} \tag{42}$$

We are now in a position to consider the problem of parameter estimation using stochastic approximation. Consider the vectorized version of the parameter $\theta = \mathrm{vec}\big(\big[A^{(1)}; A^{(2)}; \cdots; A^{(k)}\big]\big) \in \mathbb{R}^{km^2}$. The population-level Yule–Walker estimation equation (18) can be written as

$$\Big(\underbrace{\big[\Gamma_{j-i}\big]_{i,j\in[k]}}_{H^*} \otimes I_m\Big)\theta = \mathrm{vec}\big(\big[\Gamma_1; \Gamma_2; \cdots; \Gamma_k\big]\big), \tag{43}$$

where $\Gamma_i := \mathbb{E}\big[X_i X_0^\top\big] \in \mathbb{R}^{m\times m}$, for $i \in \mathbb{Z}$. We assume that

$$\frac{1}{2}\big(H^* + (H^*)^\top\big) \succeq h^* I_{km}, \quad \text{for some } h^* > 0.$$

In order to state the main corollary of Theorem 1 to auto-regressive models, the following quantities are relevant:

$$\varepsilon_{\mathrm{Mkv}}(\omega_t) := \mathrm{vec}\Big(\Big(\sum_{j=0}^{k-1} A_*^{(j)} X_{t-j} - X_{t+1}\Big) \cdot \begin{bmatrix} X_{t-1}^\top & X_{t-2}^\top & \cdots X_{t-k}^\top \end{bmatrix}\Big)$$

$$\Sigma^*_{\mathrm{Mkv}} := \sum_{t=-\infty}^{\infty} \mathbb{E}\big[\varepsilon_{\mathrm{Mkv}}(\omega_t)\varepsilon_{\mathrm{Mkv}}(\omega_0)^\top\big].$$

Corollary 5. *Under the setup above, take the stepsize and burn-in period as*

$$\eta = \frac{1}{c\Big(n^2\big(\frac{\beta}{\mu}\log\frac{\beta}{\mu}\big)(h^*)^2\varsigma^4 k^3 m^2 \beta^8/\mu^8\Big)^{1/3}}, \quad \textit{and} \quad n_0 = \tfrac{1}{2}n, \tag{44a}$$

and suppose that $\frac{n}{\log^3 n} \ge \big(k + \frac{\beta}{\mu}\log\frac{\beta}{\mu}\big)\varsigma^4 k^3 m^2 \frac{\beta^8}{\mu^8 (h^*)^2}$. *Then the Polyak–Ruppert estimator* $(\widehat{A}_n^{(j)})_{j\in[k]}$ *satisfies*

$$\sum_{j=1}^{k} \mathbb{E}\big[\|\!|\widehat{A}_n^{(j)} - A_j^*|\!\|_F^2\big] \le \frac{c}{n}\mathrm{Tr}\big(\big(H^*\otimes I_m\big)^{-1}\Sigma_{\mathrm{Mkv}}\big(H^*\otimes I_m\big)^{-1}\big) + \Big\{\frac{km^2\cdot\lambda_{max}\Big(\mathbb{E}\big[\varepsilon_{\mathrm{Mkv}}(s_0)\varepsilon_{\mathrm{Mkv}}(s_0)^\top\big]\Big)}{(h^*)^2 n}\big(k + \frac{\beta}{\mu}\log\frac{\beta}{\mu}\big)\Big\}^{4/3}\log^2 n. \tag{44b}$$

A few remarks are in order. First, the leading-order term in the bound (44b) matches the variance of asymptotic efficient estimators for AR(m) models, up to a constant factor (see [BD09], Section 8). This simply follows from the fact that the plug-in Yule-Walker estimator is asymptotically efficient for auto-regressive models. On the other hand, Corollary 5 is completely non-asymptotic, holding true for any reasonably large sample size. Note that the sample complexity lower bound exhibits an $\mathcal{O}\big(\beta^9/\mu^9\big)$ dependency on the conditioning β/μ of the Lyapunov stability certificate (P_*, Q_*). In particular, a term linear in β/μ arises from the mixing time $\frac{\beta}{\mu}\log\frac{\beta}{\mu}$, and all other factors are from the almost-sure bounds on $\|X_t\|_2$ and moment bound $\sup_{u\in\mathbb{S}^{m-1}}\langle u, X_t\rangle^4$. If we instead assumed these quantities were bounded explicitly as in some prior work [JKNN21], the factor $\beta^8\varsigma^4 k^2/\mu^8$ in the sample size requirement and stepsize choice can be replaced by such a bound.

5 Proof of Proposition 1

We begin by proving the bound on the last iterate claimed in Proposition 1. Define the error term $\Delta_t := \theta_t - \bar{\theta}$, as well as the noise terms

$$Z_{t+1} := L_{t+1} - \boldsymbol{L}(s_t), \qquad \zeta_{t+1} := (L_{t+1} - \boldsymbol{L}(s_t))\bar{\theta} + (b_{t+1} - \boldsymbol{b}(s_t)), \tag{45a}$$

$$N_t := \boldsymbol{L}(s_t) - \bar{L}, \qquad \nu_t := (\boldsymbol{L}(s_t) - \bar{L})\bar{\theta} + (\boldsymbol{b}(s_t) - \bar{b}). \tag{45b}$$

Using this notation, we have the recursion

$$\Delta_{t+1} = (I - \eta(I - \bar{L}))\Delta_t + \eta\big(N_t + Z_{t+1}\big)\Delta_t + \eta(\nu_t + \zeta_{t+1}). \tag{46}$$

Taking squared norms on both sides yields the bound $\|\Delta_{t+1}\|_2^2 \leq \sum_{i=1}^4 T_i$, where

$$T_1 := \|(I - \eta(I - \bar{L}))\Delta_t\|_2^2, \qquad T_3 := 2\eta\langle (I - \eta(I - \bar{L}))\Delta_t, \big(Z_{t+1}\Delta_t + \zeta_{t+1}\big)\rangle,$$
$$T_2 := 2\eta\langle (I - \eta(I - \bar{L}))\Delta_t, N_t\Delta_t + \nu_t\rangle, \quad \text{and} \quad T_4 := 4\eta^2\big(\|N_t\Delta_t\|_2^2 + \|Z_{t+1}\Delta_t\|_2^2 + \|\zeta_{t+1}\|_2^2 + \|\nu_t\|_2^2\big).$$

Beginning with the term T_1, expanding the square and then invoking the condition (5) yields

$$T_1 = \|\Delta_t\|^2 - 2\eta\langle \Delta_t, (I - \bar{L})\Delta_t\rangle + \eta^2\|(I - \bar{L})\Delta_t\|^2 \leq \big(1 - 2\eta(1 - \kappa) + 2\eta^2(1 + \gamma_{\max}^2)\big)\|\Delta_t\|^2.$$

As for the cross terms involved in T_2 and T_3, we note that

$$\begin{aligned}
2\langle (I - \bar{L})\Delta_t, N_t\Delta_t\rangle &\leq \|(I - \bar{L})\Delta_t\|_2^2 + \|N_t\Delta_t\|_2^2 \;\leq\; 2(1 + \gamma_{\max}^2)\|\Delta_t\|_2^2 + \|N_t\Delta_t\|_2^2, \\
2\langle (I - \bar{L})\Delta_t, \nu_t\rangle &\leq \|(I - \bar{L})\Delta_t\|_2^2 + \|\nu_t\|_2^2 \;\leq\; 2(1 + \gamma_{\max}^2)\|\Delta_t\|_2^2 + \|\nu_t\|_2^2, \\
2\langle (I - \bar{L})\Delta_t, Z_{t+1}\Delta_t\rangle &\leq \|(I - \bar{L})\Delta_t\|_2^2 + \|Z_{t+1}\Delta_t\|_2^2 \;\leq\; 2(1 + \gamma_{\max}^2)\|\Delta_t\|_2^2 + \|Z_{t+1}\Delta_t\|_2^2, \\
2\langle (I - \bar{L})\Delta_t, \zeta_{t+1}\rangle &\leq \|(I - \bar{L})\Delta_t\|_2^2 + \|\zeta_{t+1}\|_2^2 \;\leq\; 2(1 + \gamma_{\max}^2)\|\Delta_t\|_2^2 + \|\zeta_{t+1}\|_2^2.
\end{aligned}$$

We collect the above bounds on the sum $\sum_{i=1}^4 T_i$ and use the stepsize bound $\eta \leq \frac{1-\kappa}{12(1+\gamma_{\max}^2)}$, which results in the recursive inequality

$$\begin{aligned}
\|\Delta_{t+1}\|_2^2 \leq \big(1 - \eta(1 - \kappa)\big)\|\Delta_t\|_2^2 &+ 2\eta \underbrace{\big(\langle \Delta_t, N_t\Delta_t\rangle + \langle \Delta_t, \nu_t\rangle\big)}_{:=H_1(t)} \\
&+ 2\eta \underbrace{\big(\langle \Delta_t, Z_{t+1}\Delta_t\rangle + \langle \Delta_t, \zeta_{t+1}\rangle\big)}_{:=H_2(t)} + 8\eta^2 \underbrace{\big(\|N_t\Delta_t\|_2^2 + \|Z_{t+1}\Delta_t\|_2^2 + \|\zeta_{t+1}\|_2^2 + \|\nu_t\|_2^2\big)}_{:=H_3(t)}.
\end{aligned}$$

Multiplying both sides by $e^{\eta(1-\kappa)(t+1)}$ and using the fact that $\big(1 - \eta(1 - \kappa)\big) \leq e^{-\eta(1-\kappa)}$, we have

$$e^{\eta(1-\kappa)(t+1)}\|\Delta_{t+1}\|_2^2 \leq e^{\eta(1-\kappa)t}\|\Delta_t\|_2^2 + 2\eta e^{\eta(1-\kappa)(t+1)}\big(H_1(t) + H_2(t)\big) + 8\eta^2 e^{\eta(1-\kappa)(t+1)}H_3(t).$$

Unrolling this expression yields

$$e^{\eta(1-\kappa)n}\|\Delta_n\|_2^2 \leq \|\Delta_0\|_2^2 + 2\eta\sum_{t=0}^{n-1} e^{\eta(1-\kappa)(t+1)}\big(H_1(t) + H_2(t)\big) + 8\eta^2\sum_{t=0}^{n-1} e^{\eta(1-\kappa)(t+1)}H_3(t), \tag{47}$$

which is the key recursion underlying our analysis.

5.1 Analyzing the recursion (47)

Note that the running sum $M_2(n) := \sum_{t=0}^{n-1} e^{\eta(1-\kappa)t} H_2(t)$ is, by construction, a martingale adapted to the filtration $(F_t)_{t\geq 0}$. In contrast, the analogous quantity defined in terms of the process H_1 is *not* an adapted martingale. In order to circumvent this obstacle, our proof is based on introducing a *surrogate version* $\widetilde{H}_1$ of the process H_1, such that the running sum

$$\widetilde{M}_1(n) := \sum_{t=0}^{n-1} e^{\eta(1-\kappa)(t+\tau)} \widetilde{H}_1(t+\tau)$$

can be decomposed as a sum of τ martingales. See the proof of Lemma 1 for the details of the construction of $\widetilde{H}_1$. This decomposition allows us to apply standard maximal inequalities for martingales. Of course, we also need the bound the moments of the differences $\widetilde{H}_1(t) - H_1(t)$; see Lemma 1 for the bound that we provide on this difference.

We prove the MSE bounds and higher-moment bounds using slightly different analysis tools. In order to study the mean-squared error (the case $p = 1$), we note that both $\widetilde{M}_1(t)$ and $H_2(t)$ have zero expectation for any $t \geq 0$. Taking expectations on both sides of equation (47), we obtain the bound

$$e^{\eta(1-\kappa)n}\mathbb{E}\big[\|\Delta_n\|_2^2\big] \leq \|\Delta_0\|_2^2 + 2\eta \sum_{t=0}^{n-1} e^{\eta(1-\kappa)(t+1)}\mathbb{E}\big[\,\Big|H_1(t) - \widetilde{H}_1(t)\Big|\,\big] + 8\eta^2 \sum_{t=0}^{n-1} e^{\eta(1-\kappa)(t+1)}\mathbb{E}\big[H_3(t)\big]. \tag{48}$$

For higher moments, our analysis of the recursion (47) is based on a Lyapunov function Φ_n and auxiliary function Λ_n given by

$$\Phi_n := \big(\mathbb{E}\big[\sup_{0\leq t\leq n} e^{\eta(1-\kappa)tp}\|\Delta_t\|_2^{2p}\big]\big)^{1/p}, \quad \text{and} \quad \Lambda_n = \max_{t\in\{0,1,\ldots,n\}} e^{-\frac{\eta(1-\kappa)t}{2}}\Phi_t.$$

By applying Minkowski's inequality to the recursion (47), we obtain the upper bound

$$\begin{aligned}\Phi_n \leq \Phi_0 + 4\eta\big(\mathbb{E}\sup_{0\leq t\leq n}|\widetilde{M}_1(t)|^p\big)^{1/p} + 4\eta\big(\mathbb{E}\big(\sum_{t=0}^{n-1} e^{\eta(1-\kappa)t}|H_1(t) - \widetilde{H}_1(t)|\big)^p\big)^{1/p} \\ + 4\eta\big(\mathbb{E}\sup_{0\leq t\leq n}|M_2(t)|^p\big)^{1/p} + 16\eta^2\big(\mathbb{E}\big(\sum_{t=0}^{n-1} e^{\eta(1-\kappa)t}H_3(t)\big)^p\big)^{1/p}.\end{aligned} \tag{49}$$

In order to complete the proof, we need to control each of the terms on the right-hand side. The following auxiliary results provide the needed control; in all cases, the quantities (c, c_0) etc. denote universal constants; the number n in the following lemmas is seen as a general iteration index, instead of the total sample size in the final statement of the theorem.

Our first auxiliary result guarantees the existence of the surrogate variables $\widetilde{H}_1(t)$ with desirable properties:

Lemma 1. *There is a surrogate version $\{\widetilde{H}_1(t)\}_{t\geq 0}$ of the process $\{H_1(t)\}_{t\geq 0}$ such that $\mathbb{E}\big[\widetilde{H}(t)\big] = 0$ for any $t \geq 0$, and for any integer $p \in [1, \bar{p}/2]$, scalar $\tau \geq cpt_{\text{mix}}\log(c_0 t_{\text{mix}} d)$ and stepsize $\eta \leq \frac{1}{ct_{\text{mix}}(\gamma_{\max} + p\sigma_L d)}$, we have the following bounds for any $n > 0$:*

$$\big(\mathbb{E}\big[\,\Big|H_1(n) - \widetilde{H}_1(n)\Big|^p\,\big]\big)^{1/p} \leq c\eta p^2\tau\big((d\sigma_L^2 + \gamma_{\max}^2)\cdot\big(\mathbb{E}\|\Delta_{n-\tau\vee 0}\|_2^{2p}\big)^{\frac{1}{p}} + \bar{\sigma}^2 d\big), \tag{50a}$$

and for any $p \geq 2$, we have that

$$\big(\mathbb{E} \sup_{0\leq t\leq n} |\widetilde{M}_1(t)|^p\big)^{1/p} \leq \frac{cp^{3/2}}{\sqrt{\eta(1-\kappa)}}\big(\sigma_L\sqrt{d}\Phi_n + \bar{\sigma}\sqrt{e^{\eta(1-\kappa)n}\Phi_n d}\big). \tag{50b}$$

See Section 5.2 for the proof of this claim. We note that it is especially challenging to prove the bound (50a).

Our second auxiliary result is a more straightforward bound on a martingale supremum:

Lemma 2. *The process M_2 is a martingale adapted to the filtration $(\mathcal{F}_t)_{t\geq 0}$. Furthermore, for each $p \in [1, \bar{p}/2]$, $\tau \geq 2pt_{\text{mix}}\log(c_0 d)$ and $\eta \leq \frac{1}{c(\gamma_{\max}+\sigma_L d)\tau}$, for any $n > 0$, we have that*

$$\big(\mathbb{E} \sup_{0\leq t\leq n} |M_2(t)|^p\big)^{1/p} \leq \frac{cp^{3/2}\tau^{1/2}}{\sqrt{\eta(1-\kappa)}}\big(\sigma_L\sqrt{d}\Phi_n + \bar{\sigma}\sqrt{e^{\eta(1-\kappa)n}\Phi_n d}\big). \tag{51}$$

See Section 5.3 for the proof of this claim.

Finally, our third auxiliary result provides control on the process $H_3(t)$:

Lemma 3. *There is a universal constant c such that given $\tau \geq cpt_{\text{mix}}\log(c_0 t_{\text{mix}} d)$ and stepsize $\eta \leq \frac{1}{ct_{\text{mix}}(\gamma_{\max}+\sigma_L d)}$, for any $p \in [1, \bar{p}/2]$, we have*

$$\big(\mathbb{E}\big[H_3(t)^p\big]\big)^{1/p} \leq c\big(p^2\sigma_L^2 d + \gamma_{\max}^2\big)\big(\mathbb{E}\big[\|\Delta_{t-\tau\vee 0}\|_2^{2p}\big]\big)^{1/p} + cp^2\bar{\sigma}^2 d. \tag{52}$$

See Section 5.4 for the proof of this claim.

We now use these three lemmas to complete the proof of Proposition 1. We prove the case of $\bar{p} = 2$ and $\bar{p} \geq \log n$ separately.

Proof in the case of $\bar{p} = 2$: By Lemma 1 with $\tau = ct_{\text{mix}}\log(c_0 t_{\text{mix}} d)$ and Cauchy–Schwarz inequality, we have that

$$\begin{aligned}\mathbb{E}\big[\sum_{t=0}^{n-1} e^{\eta(1-\kappa)t}|\widetilde{H}_1(t) - H_1(t)|\big] &\leq c\eta\tau\sum_{t=0}^{n-1} e^{\eta(1-\kappa)t}\big((\sigma_L^2 d + \gamma_{\max}^2)\mathbb{E}\big[\|\Delta_{t-\tau\vee 0}\|_2^2\big] + \bar{\sigma}^2 d\big)\\ &\leq \frac{c\tau\bar{\sigma}^2 d}{1-\kappa}e^{\eta(1-\kappa)n} + ce\eta\tau(\sigma_L^2 d + \gamma_{\max}^2)\sum_{t=0}^{n-1} e^{\eta(1-\kappa)t}\mathbb{E}\big[\|\Delta_t\|_2^2\big].\end{aligned}$$

Similarly, by applying Lemma 3 to the last term of equation (48), we obtain the bound

$$\sum_{t=0}^{n-1} e^{\eta(1-\kappa)(t+1)}\mathbb{E}\big[H_3(t)\big] \leq \frac{c\bar{\sigma}^2 d}{(1-\kappa)\eta}e^{\eta(1-\kappa)n} + ce(\sigma_L^2 d + \gamma_{\max}^2)\sum_{t=0}^{n-1} e^{\eta(1-\kappa)t}\mathbb{E}\big[\|\Delta_t\|_2^2\big].$$

Combining them with the decomposition (48), we find that $e^{\eta(1-\kappa)n}\mathbb{E}\big[\|\Delta_n\|_2^2\big]$ is upper bounded by

$$\|\Delta_0\|_2^2 + c\frac{\eta\tau\bar{\sigma}^2 d}{1-\kappa}e^{\eta(1-\kappa)n} + c\eta^2\tau(\sigma_L^2 d + \gamma_{\max}^2)\sum_{t=0}^{n-1} e^{\eta(1-\kappa)t}\mathbb{E}\big[\|\Delta_t\|_2^2\big] \qquad \text{for any } n = 1, 2, \cdots. \tag{53}$$

In order to exploit this recursive upper bound, we define the partial sum sequence $S_n := \sum_{t=0}^{n} e^{\eta(1-\kappa)t}\mathbb{E}\big[\|\Delta_t\|_2^2\big]$. Equation (48) implies that

$$\begin{aligned}
S_n &\le S_0 + c\frac{\eta\tau\bar{\sigma}^2 d}{1-\kappa}e^{\eta(1-\kappa)n} + \big(1 + c\eta^2\tau(\sigma_L^2 d + \gamma_{\max}^2)\big)S_{n-1} \\
&\le S_0 \cdot \sum_{t=0}^{n} e^{c\eta^2\tau(\sigma_L^2 d+\gamma_{\max}^2)t} + c\frac{\eta\tau\bar{\sigma}^2 d}{1-\kappa} \cdot \sum_{t=0}^{n} e^{c\eta^2\tau(\sigma_L^2 d+\gamma_{\max}^2)t+\eta(1-\kappa)(n-t)} \\
&\le \frac{3}{(1-\kappa)\eta}e^{\eta(1-\kappa)n/3}S_0 + \frac{3c\tau\bar{\sigma}^2 d}{(1-\kappa)^2}e^{\eta(1-\kappa)n}.
\end{aligned}$$

Substituting back into the recursion (53) yields

$$\begin{aligned}
\mathbb{E}\big[\|\Delta_n\|_2^2\big] &\le \frac{6}{(1-\kappa)\eta}e^{-\eta(1-\kappa)n/3}\|\Delta_0\|_2^2 + c\frac{\eta\tau\bar{\sigma}^2 d}{1-\kappa} + c\eta^2\tau(\sigma_L^2 d + \gamma_{\max}^2)\cdot\frac{2c\tau\bar{\sigma}^2 d}{(1-\kappa)^2} \\
&\le e^{-\eta(1-\kappa)n/2}\|\Delta_0\|_2^2 + c'\frac{\eta\tau\bar{\sigma}^2 d}{1-\kappa},
\end{aligned}$$

which completes the proof of the MSE bound.

Proof in the case of $\bar{p} \ge \log n$: Now we turn to prove the p-th moment bound under Assumption 2 with $\bar{p} \ge \log n$. Recall that we analyze the growth of the Lyapunov function Φ_n, and we start from the decomposition (49).

The first term in equation (49) is simply $\|\Delta_0\|_2^2$, and the second term is controlled using equation (50b) in Lemma 1. In order to bound the third term, we apply Hölder's inequality, and obtain the bound

$$\mathbb{E}\Big(\sum_{t=0}^{n-1} e^{\eta(1-\kappa)t}\Big|H_1(t) - \widetilde{H}_1(t)\Big|\Big)^p \le \Big(\sum_{t=0}^{n-1} e^{\frac{\eta(1-\kappa)pt}{2(p-1)}}\Big)^{p-1}\cdot\sum_{t=0}^{n-1} e^{\frac{\eta p(1-\kappa)t}{2}}\mathbb{E}\Big[\Big|H_1(t) - \widetilde{H}_1(t)\Big|^p\Big].$$

By equation (50a) in Lemma 1, this quantity is at most

$$(\eta(1-\kappa))^{1-p}e^{\frac{\eta(1-\kappa)pn}{2}}\sum_{t=0}^{n-1} e^{\frac{\eta p(1-\kappa)t}{2}}\big(c\tau\big(p^2\sigma_L^2 d + \gamma_{\max}^2\big)\big(\mathbb{E}\big[\|\Delta_{t-\tau\vee 0}\|_2^{2p}\big]\big)^{1/p} + c\tau p^2\bar{\sigma}^2 d\big)^p.$$

We then obtain the inequality:

$$\begin{aligned}
&\Big(\mathbb{E}\Big(\sum_{t=0}^{n-1} e^{\eta(1-\kappa)t}\Big|H_1(t) - \widetilde{H}_1(t)\Big|\Big)^p\Big)^{1/p} \\
&\le cp^2\frac{e^{\eta(1-\kappa)n}}{\eta(1-\kappa)}\bar{\sigma}^2\tau d + c\big(p^2\sigma_L^2 d + \gamma_{\max}^2\big)\tau\frac{e^{\frac{1}{2}\eta(1-\kappa)n}}{\eta(1-\kappa)}\Big(\sum_{t=0}^{n-1} e^{\frac{1}{2}\eta p(1-\kappa)t}\mathbb{E}\big[\|\Delta_t\|_2^{2p}\big]\Big)^{1/p} \\
&\le cp^2\frac{e^{\eta(1-\kappa)n}}{\eta(1-\kappa)}\bar{\sigma}^2\tau d + c\big(p^2\sigma_L^2 d + \gamma_{\max}^2\big)\tau\frac{e^{\frac{1}{2}\eta(1-\kappa)n}}{\eta(1-\kappa)}\Big(\sum_{t=0}^{n-1} e^{-\frac{1}{2}\eta p(1-\kappa)t}\Phi_t^p\Big)^{1/p} \\
&\le cp^2\frac{e^{\eta(1-\kappa)n}}{\eta(1-\kappa)}\bar{\sigma}^2\tau d + c\big(p^2\sigma_L^2 d + \gamma_{\max}^2\big)\tau\frac{e^{\frac{1}{2}\eta(1-\kappa)n}}{\eta(1-\kappa)}n^{1/p}\Lambda_n.
\end{aligned}$$

Similarly, the fourth term on the right hand side is controlled using Lemma 2, and the bounds for the last term are based on Lemma 3 and the same strategy as above. Concretely, combining Hölder's inequality with the bound (52) yields

$$\mathbb{E}\big(\sum_{t=0}^{n-1} e^{\eta(1-\kappa)t} H_3(t)\big)^p \le \big(\sum_{t=0}^{n-1} e^{\frac{\eta(1-\kappa)pt}{2(p-1)}}\big)^{p-1} \cdot \sum_{t=0}^{n-1} e^{\frac{\eta p(1-\kappa)t}{2}} \mathbb{E}[H_3(t)^p].$$

This quantity is at most

$$(\eta(1-\kappa))^{1-p} e^{\frac{\eta(1-\kappa)pn}{2}} \sum_{t=0}^{n-1} e^{\frac{\eta p(1-\kappa)t}{2}} \big(c\big(p^2\sigma_L^2 d + \gamma_{\max}^2\big)\big(\mathbb{E}\big[\|\Delta_{t-\tau\vee 0}\|_2^{2p}\big]\big)^{1/p} + cp^2\bar{\sigma}^2 d\big)^p.$$

Noting that each term satisfies the inequality $e^{\frac{\eta p(1-\kappa)t}{2}}\big(\mathbb{E}\big[\|\Delta_{t-\tau\vee 0}\|_2^{2p}\big]\big)^{1/p} \le \Lambda_n$ for $t \in [0,n]$. We conclude that the moment $\big(\mathbb{E}\big(\sum_{t=0}^{n-1} e^{\eta(1-\kappa)t} H_3(t)\big)^p\big)^{1/p}$ is upper bounded by

$$cp^2 \frac{e^{\eta(1-\kappa)n}}{\eta(1-\kappa)}\bar{\sigma}^2 d + c\big(p^2\sigma_L^2 d + \gamma_{\max}^2\big)\frac{e^{\frac{1}{2}\eta(1-\kappa)n}}{\eta(1-\kappa)} n^{1/p}\Lambda_n.$$

Collecting the above bounds and substituting into the decomposition (49), we note that

$$\begin{aligned}
\Phi_n &\le \Phi_0 + c\sqrt{\frac{p^3\eta}{1-\kappa}}\big(\sigma_L\sqrt{d}\Phi_n + \bar{\sigma}\sqrt{e^{\eta(1-\kappa)n}\Phi_n d}\big) \\
&\quad + cp^2\frac{e^{\eta(1-\kappa)n}}{\eta(1-\kappa)}\bar{\sigma}^2\tau d + \big(p^2\sigma_L^2 d + \gamma_{\max}^2\big)\frac{e^{\frac{1}{2}\eta(1-\kappa)n}}{\eta(1-\kappa)}\tau n^{1/p}\Lambda_n \\
&\le \Phi_0 + 4c\sigma_L\sqrt{\frac{p^3\tau\eta d}{1-\kappa}}\Phi_n + \frac{1}{4}\Phi_n + c\eta\frac{\bar{\sigma}^2 p^3 d\tau}{1-\kappa}\cdot e^{\eta(1-\kappa)n} \\
&\quad + cp^2\eta\frac{e^{\eta(1-\kappa)n}}{1-\kappa}\bar{\sigma}^2\tau d + c\eta\big(p^2\sigma_L^2 d + \gamma_{\max}^2\big)\frac{e^{\frac{1}{2}\eta(1-\kappa)n}}{1-\kappa}\tau\Lambda_n
\end{aligned}$$

In the last step, we apply Young's inequality to the term $\sqrt{e^{\eta(1-\kappa)n}\Phi_n d}$, and use the condition $p \ge \log n$ to the last term so that $n^{1/p} \le e$.

Taking the stepsize $\eta \le \frac{1-\kappa}{64c^2\sigma_L^2\tau d p^3}$, we arrive at the following bound valid for any $n \in [1, e^p]$:

$$e^{-\frac{\eta(1-\kappa)n}{2}}\Phi_n \le 2\Phi_0 + cp^3\eta\frac{e^{\frac{1}{2}\eta(1-\kappa)n}}{1-\kappa}\bar{\sigma}^2\tau d + c\eta\frac{p^2\sigma_L^2 d + \gamma_{\max}^2}{1-\kappa}\tau\Lambda_n.$$

Note that the right-hand-side of above expression is monotonic increasing in the index n. For any integer pair (t,n) such that $0 < t \le n \le e^p$, we have the inequality:

$$\begin{aligned}
e^{-\frac{\eta(1-\kappa)t}{2}}\Phi_t &\le 2\Phi_0 + cp^3\eta\frac{e^{\frac{1}{2}\eta(1-\kappa)t}}{1-\kappa}\bar{\sigma}^2\tau d + c\eta\frac{p^2\sigma_L^2 d + \gamma_{\max}^2}{1-\kappa}\tau\Lambda_t \\
&\le 2\Phi_0 + cp^3\eta\frac{e^{\frac{1}{2}\eta(1-\kappa)n}}{1-\kappa}\bar{\sigma}^2\tau d + c\eta\frac{p^2\sigma_L^2 d + \gamma_{\max}^2}{1-\kappa}\tau\Lambda_n.
\end{aligned}$$

Given the value of n fixed and taking supremum over $t \in \{0,1,2,\cdots,n\}$ in the left-hand-side, we arrive at the conclusion:

$$\Lambda_n = \sup_{t\in\{0,1,\cdots,n\}} e^{-\frac{\eta(1-\kappa)t}{2}}\Phi_t \le 2\Phi_0 + cp^3\eta\frac{e^{\frac{1}{2}\eta(1-\kappa)n}}{1-\kappa}\bar{\sigma}^2\tau d + c\eta\frac{p^2\sigma_L^2 d + \gamma_{\max}^2}{1-\kappa}\tau\Lambda_n.$$

Given the stepsize $\eta \le \frac{1-\kappa}{2c(p^3\sigma_L^2 d+\gamma_{\max}^2)\tau}$, we arrive at the bound

$$\big(\mathbb{E}\|\Delta_t\|_2^p\big)^{1/p} \le e^{-\frac{1}{2}\eta(1-\kappa)n}\Lambda_n \le e^{-\frac{1}{2}\eta(1-\kappa)n}\big(\mathbb{E}\|\Delta_0\|_2^p\big)^{1/p} + \frac{cp^3\eta}{1-\kappa}\bar{\sigma}^2\tau d,$$

which completes the proof of the theorem.

It remains to prove our three auxiliary lemmas.

5.2 Proof of Lemma 1

We break the proof into three steps. In the first step, given in Section 5.2.1, we construct the surrogate process, whereas the remaining two steps are devoted to the proving the bounds (50b) and (50a), as detailed in Sections 5.2.2 and 5.2.3 respectively.

5.2.1 Construction of the surrogate process

We first claim that for any $t = 1, 2, \ldots$ and any $\tau \in \{0, \ldots, t\}$, there is a random variable $\widetilde{s}_t \in \mathbb{X}$ such that $\widetilde{s}_t \mid \mathcal{F}_{t-\tau} \sim \xi$, and

$$\big(\mathbb{E}\big[\rho(s_t, \widetilde{s}_t)^p \mid \mathcal{F}_{t-\tau}\big]\big)^{1/p} \le c_0 \exp\big(-\tfrac{\tau}{2t_{\mathrm{mix}}p}\big) \quad \text{for each } p \ge 2. \tag{54}$$

Here c_0 is a universal constant.
Our construction is based on the following bound on the Wasserstein distance:

Lemma 4. *Under Assumptions 1 and 3, the Wasserstein distance is upper bounded as*

$$\mathcal{W}_{1,\rho}\big(\delta_x P^\tau, \xi\big) \le c_0 \exp\big\{\lfloor \tfrac{\tau}{t_{\mathrm{mix}}} \rfloor\big\},$$

valid for any $x \in \mathbb{X}$ and $\tau \ge 0$.

See Appendix B.1 for the proof of this claim.

We now use Lemma 4 to construct the desired process. We begin by constructing a coupling conditionally on the σ-field $\mathcal{F}_{t-\tau}$: let $\widetilde{s}_t$ be a state whose conditional law is ξ, satisfying the identity:

$$\mathbb{E}\big[\rho(s_t, \widetilde{s}_t) \mid \mathcal{F}_{t-\tau}\big] = \mathcal{W}_{1,\rho}\big(\mathcal{L}(s_t \mid \mathcal{F}_{t-\tau}), \xi\big). \tag{55}$$

The existence of such $\widetilde{s}_t$ is guaranteed by the definition of Wasserstein distance. We now bound the relevant quantities based on this construction.

Combining the identity (55) with Lemma 4 yields $\mathbb{E}\big[\rho(s_t, \widetilde{s}_t) \mid \mathcal{F}_{t-\tau}\big] \le c_0 \cdot 2^{-\lfloor \frac{\tau}{t_{\mathrm{mix}}} \rfloor}$. Applying Cauchy–Schwarz inequality and invoking Assumption 3, we find that

$$\begin{aligned}
\big(\mathbb{E}\big[\rho(s_t, \widetilde{s}_t)^p \mid \mathcal{F}_{t-\tau}\big]\big)^{1/p} &\le \big(\mathbb{E}\big[\rho(s_t, \widetilde{s}_t) \mid \mathcal{F}_{t-\tau}\big]\big)^{\frac{1}{2p}} \cdot \big(\mathbb{E}\big[\rho(s_t, \widetilde{s}_t)^{2p-1} \mid \mathcal{F}_{t-\tau}\big]\big)^{\frac{1}{2p}} \\
&\le \big(\mathbb{E}\big[\rho(s_t, \widetilde{s}_t) \mid \mathcal{F}_{t-\tau}\big]\big)^{\frac{1}{2p}} \\
&\le c_0 \cdot 2^{1-\frac{\tau}{2t_{\mathrm{mix}}p}},
\end{aligned} \tag{56}$$

which establishes the claim.

We now use the sequence of random variable $\widetilde{s}_t$ just constructed to define the extended filtration $\widetilde{\mathcal{F}}_t := \sigma\big((s_k)_{0\le k\le t}, (\widetilde{s}_k)_{0\le k\le t}, \big((L_k, b_k)\big)_{0\le k\le t}\big)$, as well as the surrogate quantities

$$\widetilde{\nu}_t := \big(\boldsymbol{L}(\widetilde{s}_t) - \bar{L}\big)\bar{\theta} + \big(\boldsymbol{b}(\widetilde{s}_t) - \bar{b}\big), \quad \text{and} \quad \widetilde{H}_1(t) := \langle \Delta_{(t-\tau)\vee 0}, \widetilde{\nu}_t\rangle + \langle \Delta_{(t-\tau)\vee 0}, \big(\boldsymbol{L}(\widetilde{s}_t) - \bar{L}\big)\Delta_{(t-\tau)\vee 0}\rangle.$$

Note that by definition, we have $\mathbb{E}\big[\widetilde{H}_1(t) \mid \widetilde{\mathcal{F}}_{(t-\tau)\vee 0}\big] = 0$ for each $t = 0, 1, 2, \ldots$.

5.2.2 Proof of the bound (50b)

We first perform a decomposition on the process $\widetilde{M}_1$. In particular, for $\ell \in \{0, 1, \cdots, \tau-1\}$, we define the stochastic process $\widetilde{M}_1^{(\ell)}(n) := \sum_{t=0}^{n-1} e^{\eta(1-\kappa)(t+\tau)} \widetilde{H}_1(t+\tau)\mathbf{1}_{\{t \bmod \tau = \ell\}}$. Clearly, we have $\widetilde{M}_1(n) = \sum_{\ell=0}^{\tau-1} \widetilde{M}_1^{(\ell)}(n)$ for any $n \geq 0$. Furthermore, we note for any $t \geq 0$, we have the relations:

$$\mathbb{E}\big[\widetilde{H}_1(t+\tau) \mid \widetilde{\mathcal{F}}_t\big] = 0, \quad \text{and} \quad \widetilde{H}_1(t) \in \widetilde{\mathcal{F}}_t.$$

So for each $\ell \in [0, \tau-1]$, the process $\widetilde{M}_1^{(\ell)}$ is a martingale adapted to the filtration $\big(\widetilde{\mathcal{F}}_t\big)_{t\geq 0}$.

By the BDG inequality, we have the maximal inequality $\big(\mathbb{E}\sup_{0\leq t\leq n}|\widetilde{M}_1^{(\ell)}(t)|^p\big)^{1/p} \leq cp\big(\mathbb{E}\big([\widetilde{M}_1^{(\ell)}]_n\big)^{p/2}\big)^{1/p}$, valid for all $\ell = 0, 1, \ldots, \tau-1$. Similarly, for the quadratic variation term $[\widetilde{M}_1^{(\ell)}]_n$, we have that

$$\begin{aligned}
\mathbb{E}\big[\big([\widetilde{M}_1^{(\ell)}]_n\big)^{p/2}\big] &= \mathbb{E}\big[\big(\sum_{k=0}^{\lfloor\frac{n-1}{\tau}\rfloor} e^{\eta(1-\kappa)(k\tau+\tau+\ell)}\|\widetilde{H}_1(k\tau+\ell)\|_2^2\big)^{p/2}\big] \\
&\leq \big(\sum_{k=0}^{\lfloor\frac{n-1}{\tau}\rfloor} e^{\eta(1-\kappa)p(k\tau+\tau+\ell)}\mathbb{E}\big[\|\widetilde{H}_1(k\tau+\ell)\|_2^p\big]\big)\cdot\big(\sum_{t=0}^{n-1} e^{-\frac{p^2}{2p-4}\tau\eta(1-\kappa)t}\big)^{\frac{p-2}{2}},
\end{aligned}$$

which is at most

$$\big(\eta\tau(1-\kappa)\big)^{-\frac{p}{2}+1}\sum_{t=\tau}^{n-1} e^{\eta(1-\kappa)tp}\big(\mathbb{E}\big[\,\big|2\langle\Delta_{t-\tau},\,(\bar{L}(\widetilde{s}_t)-\bar{L})\Delta_{t-\tau}\rangle\big|^p\,\big] + \mathbb{E}\big[\,|2\langle\widetilde{\nu}_t,\,\Delta_{t-\tau}\rangle|^p\,\big]\big)\mathbf{1}_{\{t\bmod\tau=\ell\}}.$$

Invoking the tail condition in Assumption 2 under the stationary distribution, we have that

$$\begin{aligned}
\mathbb{E}\big[\,\big|2\langle\Delta_{t-\tau},\,(\bar{L}(\widetilde{s}_t)-\bar{L})\Delta_{t-\tau}\rangle\big|^p \mid \mathcal{F}_{t-\tau}\big] &\leq \big(p\sigma_L\sqrt{d}\cdot\|\Delta_{t-\tau}\|_2^2\big)^p, \quad \text{and} \\
\mathbb{E}\big[\,|\langle\widetilde{\nu}_t,\,\Delta_{t-\tau}\rangle|^p \mid \mathcal{F}_{t-\tau}\big] &\leq \big(p\bar{\sigma}\sqrt{d}\cdot\|\Delta_{t-\tau}\|_2\big)^p.
\end{aligned}$$

Substituting into the moment bounds for $[\widetilde{M}_1^{(\ell)}]_n$ and combining the results for $\ell = 0, 1, \cdots, \tau-1$ using Minkowski's inequality, we arrive at the bound

$$\begin{aligned}
&\big(\mathbb{E}\sup_{0\leq t\leq n}|\widetilde{M}_1(t)|^p\big)^{1/p} \\
&\leq \sum_{\ell=0}^{\tau-1}\big(\mathbb{E}\sup_{0\leq t\leq n}|\widetilde{M}_1^{(\ell)}(t)|^p\big)^{1/p} \\
&\leq \frac{\tau\cdot n^{\frac{1}{p}}\sqrt{p}}{\big(\eta\tau(1-\kappa)\big)^{\frac{1}{2}+\frac{1}{p}}}\big\{p\sigma_L\sqrt{d}\cdot\max_{0\leq t\leq n}\big[e^{\eta(1-\kappa)t}\big(\mathbb{E}\|\Delta_t\|_2^{2p}\big)^{1/p}\big] + e^{\frac{\eta(1-\kappa)n}{2}}p\bar{\sigma}\sqrt{d}\max_{0\leq t\leq n}\big[e^{\eta(1-\kappa)t/2}\big(\mathbb{E}\|\Delta_t\|_2^p\big)^{1/p}\big]\big\} \\
&\leq \sqrt{\frac{\tau p}{\eta(1-\kappa)}}\big(p\sigma_L\sqrt{d}\Phi_n + p\bar{\sigma}\sqrt{e^{\eta(1-\kappa)n}\Phi_n d}\big),
\end{aligned}$$

which completes the proof of this lemma.

5.2.3 Proof of the bound (50a)

By Minkowski's inequality, we can upper bound the error as $\big(\mathbb{E}\big[(H_1(t) - \widetilde{H}_1(t))^p\big]\big)^{1/p} \leq \sum_{k=1}^{6} J_k$, where

$$J_1 := \big(\mathbb{E}\big[\langle \Delta_{t-\tau},\, \nu_t - \widetilde{\nu}_t\rangle^p\big]\big)^{1/p}, \qquad J_2 := \big(\mathbb{E}\big[\langle \Delta_t - \Delta_{t-\tau},\, \nu_t\rangle^p\big]\big)^{1/p}$$
$$J_3 := \big(\mathbb{E}\big[\langle \Delta_{t-\tau},\, \big(L(\widetilde{s}_t) - \boldsymbol{L}(s_t)\big)\Delta_{t-\tau}\rangle^p\big]\big)^{1/p}, \qquad J_4 := \big(\mathbb{E}\big[\langle \Delta_t - \Delta_{t-\tau},\, N_t\Delta_{t-\tau}\rangle^p\big]\big)^{1/p}$$
$$J_5 := \big(\mathbb{E}\big[\langle \Delta_t,\, N_t(\Delta_t - \Delta_{t-\tau})\rangle^p\big]\big)^{1/p} \qquad J_6 := \big(\mathbb{E}\big[\langle \Delta_t - \Delta_{t-\tau},\, N_t(\Delta_t - \Delta_{t-\tau})\rangle^p\big]\big)^{1/p}$$

The terms J_1 and J_3 can be controlled using the bound on $\rho(s_t, \widetilde{s}_t)$ and the Lipschitz condition (4); doing so yields the bound

$$J_1 \leq \bar{\sigma} d\big(\mathbb{E}\big[\|\Delta_{t-\tau}\|_2^p \cdot \mathbb{E}\big[\rho(s_t, \widetilde{s}_t)^p \mid \mathcal{F}_{t-\tau}\big]\big]\big)^{1/p} \leq 2c_0\bar{\sigma} d\big(\mathbb{E}\|\Delta_{t-\tau}\|_2^p\big)^{1/p} \cdot 2^{-\frac{\tau}{2pt_{\mathrm{mix}}}}, \quad \text{and}$$
$$J_3 \leq \sigma_L d\big(\mathbb{E}\big[\|\Delta_{t-\tau}\|_2^{2p} \cdot \mathbb{E}\big[\rho(s_t, \widetilde{s}_t)^p \mid \mathcal{F}_{t-\tau}\big]\big]\big)^{1/p} \leq 2c_0\sigma_L d\big(\mathbb{E}\|\Delta_{t-\tau}\|_2^{2p}\big)^{1/p} \cdot 2^{-\frac{\tau}{2pt_{\mathrm{mix}}}}.$$

Given the time lag parameter $\tau \geq cpt_{\mathrm{mix}} \log(c_0 t_{\mathrm{mix}} d) \geq 2pt_{\mathrm{mix}} \log\big(\frac{d}{\eta}\big)$, we have the bound

$$J_1 \leq \eta\bar{\sigma}\sqrt{d}\big(\mathbb{E}\|\Delta_{t-\tau}\|_2^p\big)^{1/p}, \quad \text{and} \quad J_3 \leq \eta\eta\sigma_L\sqrt{d}\big(\mathbb{E}\|\Delta_{t-\tau}\|_2^{2p}\big)^{1/p}. \tag{57}$$

Turning to the J_2 term, applying the Cauchy–Schwarz inequality yields

$$J_2 \leq \big(\mathbb{E}\|\Delta_t - \Delta_{t-\tau}\|_2^{2p}\big)^{\frac{1}{2p}} \cdot \big(\mathbb{E}\|\nu_t\|_2^{2p}\big)^{\frac{1}{2p}} \overset{(i)}{\leq} \big(\mathbb{E}\|\Delta_t - \Delta_{t-\tau}\|_2^{2p}\big)^{\frac{1}{2p}} \cdot p\bar{\sigma}\sqrt{d}. \tag{58}$$

where step (i) follows from Assumption 2.

The terms J_4 and J_5 can be controlled via once again replacing s_t with its surrogate $\widetilde{s}_t$. First, by Cauchy–Schwarz inequality, we note that

$$J_4 \leq \big(\mathbb{E}\|\Delta_t - \Delta_{t-\tau}\|_2^{2p}\big)^{\frac{1}{2p}} \cdot \big(\mathbb{E}\|N_t\Delta_{t-\tau}\|_2^{2p}\big)^{\frac{1}{2p}}, \quad J_5 \leq \big(\mathbb{E}\|\Delta_t - \Delta_{t-\tau}\|_2^{2p}\big)^{\frac{1}{2p}} \cdot \big(\mathbb{E}\|N_t^\top\Delta_{t-\tau}\|_2^{2p}\big)^{\frac{1}{2p}}.$$

Using the decomposition $N_t = (\boldsymbol{L}(\widetilde{s}_t) - \bar{L}) + (\boldsymbol{L}(s_t) - \boldsymbol{L}(\widetilde{s}_t))$, we note that

$$\big(\mathbb{E}\|N_t\Delta_{t-\tau}\|_2^{2p}\big)^{\frac{1}{2p}} \leq \big(\mathbb{E}\|(\boldsymbol{L}(\widetilde{s}_t) - \bar{L})\Delta_{t-\tau}\|_2^{2p}\big)^{\frac{1}{2p}} + \big(\mathbb{E}\|(\boldsymbol{L}(s_t) - \boldsymbol{L}(\widetilde{s}_t))\Delta_{t-\tau}\|_2^{2p}\big)^{\frac{1}{2p}}.$$

We bound the conditional expectations of the quantities above. The first term can be controlled via Assumption 2:

$$\mathbb{E}\big[\|(\boldsymbol{L}(\widetilde{s}_t) - \bar{L})\Delta_{t-\tau}\|_2^{2p} \mid \mathcal{F}_{t-\tau}\big] \leq (\sigma_L p\sqrt{d})^{2p}\|\Delta_{t-\tau}\|_2^{2p},$$

and the second term is controlled using the Lipschitz condition 4:

$$\begin{aligned}\mathbb{E}\big[\|(\boldsymbol{L}(s_t) - \boldsymbol{L}(\widetilde{s}_t))\Delta_{t-\tau}\|_2^{2p} \mid \mathcal{F}_{t-\tau}\big] &\leq (\sigma_L d)^{2p} \cdot \mathbb{E}\big[\rho(s_t, \widetilde{s}_t)^{2p} \mid \mathcal{F}_{t-\tau}\big] \cdot \|\Delta_{t-\tau}\|_2^{2p} \\ &\leq (\sigma_L d)^{2p} \cdot c_0 \cdot 2^{1-\frac{\tau}{t_{\mathrm{mix}}}} \cdot \|\Delta_{t-\tau}\|_2^{2p}.\end{aligned}$$

Consequently, taking $\tau \geq 2t_{\mathrm{mix}} p \log(c_0 d)$, we have the bounds

$$\big(\mathbb{E}\|N_t\Delta_{t-\tau}\|_2^{2p}\big)^{\frac{1}{2p}} \leq \sigma_L p\sqrt{d} \cdot \big(\mathbb{E}\|\Delta_{t-\tau}\|_2^{2p}\big)^{\frac{1}{2p}}, \quad \text{and} \quad \big(\mathbb{E}\|N_t^\top\Delta_{t-\tau}\|_2^{2p}\big)^{\frac{1}{2p}} \leq \sigma_L p\sqrt{d} \cdot \big(\mathbb{E}\|\Delta_{t-\tau}\|_2^{2p}\big)^{\frac{1}{2p}}.$$

Putting together the pieces, we arrive at the bound

$$J_4 + J_5 \leq 2\big(\mathbb{E}\|\Delta_t - \Delta_{t-\tau}\|_2^{2p}\big)^{\frac{1}{2p}} \cdot \sigma_L p\sqrt{d} \cdot \big(\mathbb{E}\|\Delta_{t-\tau}\|_2^{2p}\big)^{\frac{1}{2p}}. \tag{59}$$

By the Lipschitz condition (4) and the assumed boundedness (3) of the metric space, the term J_6 admits the simple upper bound

$$J_6 \leq \big(\mathbb{E}\big[\|\!|N_t|\!\|_{\mathrm{op}}^p \|\Delta_t - \Delta_{t-\tau}\|_2^{2p}\big]\big)^{\frac{1}{p}} \leq \sigma_L d \big(\mathbb{E}\|\Delta_t - \Delta_{t-\tau}\|_2^{2p}\big)^{\frac{1}{p}} \tag{60}$$

From all of these bounds, we see that the remaining crucial piece is to bound $\mathbb{E}\|\Delta_t - \Delta_{t-\tau}\|_2^{2p}$. In order to do so, we require the following two helper lemmas

Lemma 5. *Given $p \geq 2$ and $\ell > 0$, the iterates (3a) with stepsize $\eta \leq \big(6(\gamma_{\max} + \sigma_L d)\ell\big)^{-1}$ satisfy the bound*

$$\big(\mathbb{E}\big[\|\Delta_{t+\ell} - \Delta_t\|_2^p\big]\big)^{1/p} \leq e\eta\ell(\gamma_{\max} + \sigma_L d)\big(\mathbb{E}\big[\|\Delta_t\|_2^p\big]\big)^{1/p} + 3\eta p\ell\sqrt{d}\bar{\sigma}, \tag{61a}$$

and consequently,

$$\frac{1}{2}\big(\mathbb{E}\big[\|\Delta_t\|_2^p\big]\big)^{1/p} - 6\eta p\ell\sqrt{d}\bar{\sigma} \leq \big(\mathbb{E}\big[\|\Delta_{t+\ell}\|_2^p\big]\big)^{1/p} \leq e\big(\mathbb{E}\big[\|\Delta_t\|_2^p\big]\big)^{1/p} + 6\eta p\ell\sqrt{d}\bar{\sigma}. \tag{61b}$$

See Appendix B.2 for the proof of this claim.

Our second auxiliary result is of a bootstrap nature: it is based on assuming that for some given an integer $p \geq 2$, fix any integer $\tau \geq 2t_{\mathrm{mix}} p \log(c_0 d)$, there exist positive scalars $\omega_p, \beta_p > 0$ such that

$$\big(\mathbb{E}\big[\|\Delta_{t+\ell} - \Delta_t\|_2^p\big]\big)^{1/p} \leq \eta\omega_p \cdot \big(\mathbb{E}\big[\|\Delta_t\|_2^p\big]\big)^{1/p} + \eta\beta_p\bar{\sigma} \tag{62}$$

for any $t \geq 0$, $\eta \leq \frac{1}{48(\gamma_{\max}+\sigma_L d)\tau}$ and $\ell \in [0, \tau]$. We then have the following guarantee:

Lemma 6. *When the condition (62) holds, then, for any $t \geq 0$, $\eta \leq \frac{1}{48(\gamma_{\max}+\sigma_L d)\tau}$, and $\ell \in [0, \tau]$, we have*

$$\begin{aligned}\big(\mathbb{E}\big[\|\Delta_{t+\ell} - \Delta_t\|_2^p\big]\big)^{1/p} \leq \eta\big(12\big(p\sqrt{d}\sigma_L + \gamma_{\max}\big)\ell + \frac{\omega_p}{2}\big)\big(\big(\mathbb{E}\|\Delta_t\|_2^p\big)^{1/p} + \eta p(\tau+\ell)\sqrt{d}\bar{\sigma}\big) \\ + \eta\big(2p\ell\sqrt{d} + \frac{1}{2}\beta_p\big)\bar{\sigma}.\end{aligned} \tag{63}$$

See Appendix B.3 for the proof of this claim.

We now complete the proof of the bound (50a) by using a bootstrapping argument in order to obtain a sharp bound on $\mathbb{E}\|\Delta_t - \Delta_{t-\tau}\|_2^p$. Let $\omega_p^{(0)} := e\tau(\gamma_{\max} + \sigma_L d)$ and $\beta_p^{(0)} := p\tau\sqrt{d}$, and define the following recursion:

$$\begin{cases}\omega_p^{(i+1)} = \frac{1}{2}\omega_p^{(i)} + 12\big(p\sqrt{d}\sigma_L + \gamma_{\max}\big)\tau, \\ \beta_p^{(i+1)} = \frac{1}{2}\beta_p^{(i)} + 2p\tau\sqrt{d} + 2\eta\big(12\big(p\sqrt{d}\sigma_L + \gamma_{\max}\big)\tau + \frac{1}{2}\omega_p^{(i)}\big)p\tau\sqrt{d}.\end{cases}$$

It can be seen that as $i \to \infty$, the sequence $(\omega_p^{(i)}, \beta_p^{(i)})$ converges to a unique limit (ω_p^*, β_p^*); this limit is the unique fixed point of the iterates defined above.

By Lemma 6, if the iterates satisfy the bound (62) with constants $\big(\omega_p^{(i)}, \beta_p^{(i)}\big)$, then it also satisfy the bound with constants $\big(\omega_p^{(i+1)}, \beta_p^{(i+1)}\big)$. By Lemma 5, the iterates satisfy bound with constants $\big(\omega_p^{(0)}, \beta_p^{(0)}\big)$. An induction argument then yields the bound for any $\big(\omega_p^{(i)}, \beta_p^{(i)}\big)$. In particular, the bound is satisfied by the fixed point $\big(\omega_p^*, \beta_p^*\big)$.

Solving directly for the fixed-point equation, we find that

$$\omega_p^* = 24\big(p\sqrt{d}\sigma_L + \gamma_{\max}\big)\tau, \quad \text{and} \quad \beta_p^* = 4p\tau\sqrt{d} + 96\eta\big(p\sqrt{d}\sigma_L + \gamma_{\max}\big)p\tau^2\sqrt{d}.$$

Taking the stepsize $\eta \leq \frac{1}{48(\gamma_{\max}+p\sigma_L d)\tau}$, we arrive at the bound

$$\big(\mathbb{E}\big[\|\Delta_{t+\ell} - \Delta_t\|_2^p\big]\big)^{1/p} \leq 24\eta\tau\big(p\sqrt{d}\sigma_L + \gamma_{\max}\big)\big(\mathbb{E}\|\Delta_t\|_2^p\big)^{1/p} + 6\eta p\tau\sqrt{d}\bar{\sigma}, \tag{64}$$

for any $t \geq 0$ and $\ell \in [0, \tau]$.

Collecting the bounds (57), (58), (59), (60) and (64) and taking the stepsize $\eta \leq \frac{1}{c(\gamma_{\max}+p\sigma_L d)\tau}$, we arrive at the bound

$$\big(\mathbb{E}\big[(H_1(t) - \widetilde{H}_1(t))^p\big]\big)^{1/p} \leq c\eta p^2\tau\big((d\sigma_L^2 + \gamma_{\max}^2)\cdot\big(\mathbb{E}\|\Delta_{t-\tau}\|_2^{2p}\big)^{\frac{1}{p}} + \bar{\sigma}^2 d\big),$$

thereby completing the proof of the bound (50a).

5.3 Proof of Lemma 2

By the BDG inequality, we have the bound $\big(\mathbb{E}\sup_{0\leq t\leq n}|M_2(t)|^p\big)^{1/p} \leq cp\big(\mathbb{E}\big([M_2]_n\big)^{p/2}\big)^{1/p}$, valid for all $\ell = 0, 1, \ldots, \tau - 1$.

As for the quadratic variation $[M_2]_n$, applying Hölder's inequality yields

$$\begin{aligned}
\mathbb{E}\big[\big([M_2]_n\big)^{p/2}\big] &= \mathbb{E}\big[\big(\sum_{t=0}^{n-1} e^{\eta(1-\kappa)t}\|H_2(t)\|_2^2\big)^{p/2}\big] \\
&\leq \big(\sum_{t=0}^{n-1} e^{\eta(1-\kappa)tp}\mathbb{E}\big[\|H_2(t)\|_2^p\big]\big)\cdot\big(\sum_{t=0}^{n-1} e^{-\frac{p^2}{2p-4}\eta(1-\kappa)t}\big)^{\frac{p-2}{2}} \\
&\leq \big(\eta(1-\kappa)\big)^{-\frac{p}{2}+1}\sum_{t=0}^{n-1} e^{\eta(1-\kappa)tp}\big(\mathbb{E}\big[\,|2\langle\Delta_t,\, Z_{t+1}\Delta_t\rangle|^p\,\big] + \mathbb{E}\big[\,|2\langle\zeta_{t+1},\, \Delta_t\rangle|^p\,\big]\big).
\end{aligned}$$

For the moment terms above, we invoke Assumption 2, and obtain the following bounds:

$$\mathbb{E}\big[\,|\langle\Delta_t,\, Z_{t+1}\Delta_t\rangle|^p \mid \mathcal{F}_t\big] \leq \|\Delta_t\|_2^p\cdot\mathbb{E}\big[\big(\sum_{j=1}^d\langle e_j,\, Z_{t+1}\Delta_t\rangle^2\big)^{p/2} \mid \mathcal{F}_t\big] \leq \big(p\sigma_L\sqrt{d}\cdot\|\Delta_t\|_2^2\big)^p,$$

$$\mathbb{E}\big[\,|\langle\zeta_{t+1},\, \Delta_t\rangle|^p \mid \mathcal{F}_t\big] \leq \|\Delta_t\|_2^p\cdot\mathbb{E}\big[\big(\sum_{j=1}^d\langle e_j,\, \zeta_{t+1}\rangle^2\big)^{p/2} \mid \mathcal{F}_t\big] \leq \big(p\bar{\sigma}\sqrt{d}\cdot\|\Delta_t\|_2\big)^p.$$

Substituting into the bound above, we find that

$$\begin{aligned}
&\big(\mathbb{E}\big[\big([M_2]_n\big)^{p/2}\big]\big)^{1/p} \\
&\leq \tfrac{(\eta(1-\kappa))^{-\frac{1}{p}}\cdot n^{\frac{1}{p}}}{\sqrt{\eta(1-\kappa)}}\big\{p\sigma_L\sqrt{d}\cdot\max_{0\leq t\leq n}\big[e^{\eta(1-\kappa)t}\big(\mathbb{E}\|\Delta_t\|_2^{2p}\big)^{1/p}\big] + e^{\frac{\eta(1-\kappa)n}{2}}p\bar{\sigma}\sqrt{d}\max_{0\leq t\leq n}\big[e^{\eta(1-\kappa)t/2}\big(\mathbb{E}\|\Delta_t\|_2^p\big)^{1/p}\big]\big\} \\
&\leq \tfrac{1}{\sqrt{\eta(1-\kappa)}}\big(p\sigma_L\sqrt{d}\Phi_n + p\bar{\sigma}\sqrt{e^{\eta(1-\kappa)n}\Phi_n d}\big).
\end{aligned}$$

5.4 Proof of Lemma 3

Recall the definitions (45a) and (45b). By Minkowski's inequality, we have the upper bound

$$\big(\mathbb{E}\big[H_3(t)^p\big]\big)^{1/p} \leq \big(\mathbb{E}\|N_t\Delta_t\|_2^{2p}\big)^{1/p} + \big(\mathbb{E}\|Z_{t+1}\Delta_t\|_2^{2p}\big)^{1/p} + \big(\mathbb{E}\|\zeta_{t+1}\|_2^{2p}\big)^{1/p} + \big(\mathbb{E}\|\nu_t\|_2^{2p}\big)^{1/p}, \tag{65}$$

For the martingale part of the noise, we note that Assumption 2 implies that

$$\big(\mathbb{E}\|Z_{t+1}\Delta_t\|_2^{2p} \mid \mathcal{F}_t\big)^{1/p} \leq p^2\sigma_L^2 d \cdot \|\Delta_t\|_2^2, \quad \text{and} \quad \big(\mathbb{E}\|\zeta_{t+1}\|_2^{2p}\big)^{1/p} \leq p^2\bar{\sigma}^2 d.$$

For the additive Markov noise, applying Assumption 2 yields the bound $\big(\mathbb{E}\|\nu_t\|_2^{2p}\big)^{1/p} \leq p^2\bar{\sigma}^2 d$.

For the Markov part of the multiplicative noise, we make use of the construction given in Section 5.2.1, where we showed that for a given $\tau > 0$, there exists a random variable $\widetilde{s}_t$ such that $\widetilde{s}_t \mid \mathcal{F}_{t-\tau} \sim \xi$, and $\mathbb{E}\big[\rho^p(s_t, \widetilde{s}_t) \mid \mathcal{F}_{t-\tau}\big] \leq c_0 \cdot 2^{1-\frac{\tau}{t_{\text{mix}}}}$. Observe the decomposition

$$N_t\Delta_t = \big(\boldsymbol{L}(s_t) - L(\widetilde{s}_t)\big)\Delta_{t-\tau} + \big(L(\widetilde{s}_t) - \bar{L}\big)\Delta_{t-\tau} + N_t\big(\Delta_t - \Delta_{t-\tau}\big).$$

Using the Lipschitz condition (4), we have that

$$\mathbb{E}\big[\|\big(\boldsymbol{L}(s_t) - L(\widetilde{s}_t)\big)\Delta_{t-\tau}\|_2^{2p} \mid \mathcal{F}_{t-\tau}\big] \leq c_0 \cdot 2^{1-\frac{\tau}{t_{\text{mix}}}}\big(\sigma_L d\|\Delta_{t-\tau}\|_2\big)^{2p}.$$

For any $\tau \geq 2pt_{\text{mix}}\log d$, we have the bound

$$\big(\mathbb{E}\big[\|\big(\boldsymbol{L}(s_t) - L(\widetilde{s}_t)\big)\Delta_{t-\tau}\|_2^{2p}\big]\big)^{1/p} \leq p^2\sigma_L^2 d \cdot \big(\mathbb{E}\|\Delta_t\|_2^{2p}\big)^{1/p}.$$

By the moment bounds (2) on the stationary distribution, we have

$$\mathbb{E}\big[\|\big(\boldsymbol{L}(\widetilde{s}_t) - \bar{L}\big)\Delta_{t-\tau}\|_2^{2p} \mid \mathcal{F}_{t-\tau}\big] \leq \big(2p\sigma_L\sqrt{d}\|\Delta_{t-\tau}\|_2\big)^{2p}.$$

For the last term, we use the Lipschitz condition 4 as well as the boundedness condition 3 of metric space. In conjunction with the inequality (64), for $\tau \geq 2pt_{\text{mix}}\log(c_0 d)$ and stepsize $\eta \leq \frac{1}{48\tau(\sigma_L d + \gamma_{\max})}$, we arrive at the bound

$$\begin{aligned}
\big(\mathbb{E}\big[\|N_t(\Delta_t - \Delta_{t-\tau})\|_2^{2p}\big]\big)^{1/p} &\leq \sigma_L^2 d^2 \cdot \big(\mathbb{E}\big[\|\Delta_t - \Delta_{t-\tau}\|_2^{2p}\big]\big)^{1/p} \\
&\leq c\eta^2\sigma_L^2 d^2\tau^2\big(p^2\sigma_L^2 d + \gamma_{\max}^2\big)\big(\mathbb{E}\big[\|\Delta_{t-\tau}\|_2^{2p}\big]\big)^{1/p} + c\eta^2 p^2\sigma_L^2\bar{\sigma}^2 d^3\tau^2 \\
&\leq c\big(p^2\sigma_L^2 d + \gamma_{\max}^2\big)\big(\mathbb{E}\big[\|\Delta_{t-\tau}\|_2^{2p}\big]\big)^{1/p} + cp^2\bar{\sigma}^2 d,
\end{aligned}$$

for a universal constant $c > 0$.

Collecting the bounds above and substituting into our initial bound (65), we find that

$$\big(\mathbb{E}\big[H_3(t)^p\big]\big)^{1/p} \leq c\big(p^2\sigma_L^2 d + \gamma_{\max}^2\big)\big(\mathbb{E}\big[\|\Delta_{t-\tau}\|_2^{2p}\big]\big)^{1/p} + cp^2\bar{\sigma}^2 d,$$

as claimed.

6 Proof of Theorem 1

From the defining equations (3a) and (3b), we have the telescoping relation

$$\frac{\theta_n-\theta_{n_0}}{\eta(n-n_0)} = \frac{1}{n-n_0}\sum_{t=n_0}^{n-1}\big(\theta_t - L_{t+1}\theta_t - b_{t+1}\big) = (I-\bar{L})(\widehat{\theta}_n-\bar{\theta}) + \frac{1}{n-n_0}\Psi_{n_0,n} + \frac{1}{n-n_0}\Upsilon_{n_0,n} \quad (66)$$

where $\Psi_{n_0,n} = \sum_{t=n_0}^{n-1}\big(L_{t+1}\theta_t + b_{t+1} - \mathbb{E}\big[L_{t+1}\theta_t + b_{t+1}|\mathcal{F}_t\big]\big)$ and $\Upsilon_{n_0,n} := \frac{1}{n-n_0}\sum_{t=n_0}^{n-1}\big(\boldsymbol{L}(s_t)\theta_t + \boldsymbol{b}(s_t) - \bar{L}\theta_t - \bar{b}\big)$. Some algebra yields

$$\widehat{\theta}_n - \bar{\theta} = \frac{(I-\bar{L})^{-1}\big(\theta_n-\theta_{n_0}\big)}{\eta(n-n_0)} - \frac{(I-\bar{L})^{-1}\Psi_{n_0,n}}{n-n_0} - \frac{(I-\bar{L})^{-1}\Upsilon_{n_0,n}}{n-n_0} =: I_1 + I_2 + I_3 \quad (67)$$

From the triangle inequality, it suffices to bound the norms of I_1, I_2 and I_3.

In the following, we prove a slightly stronger claim, which gives bounds on an arbitrary quadratic loss functional. In particular, given a matrix $Q \succ 0$, we seek bounds on the Q-norm $\|\widehat{\theta}_n - \bar{\theta}\|_Q := \sqrt{(\widehat{\theta}_n-\bar{\theta})^\top Q(\widehat{\theta}_n-\bar{\theta})}$.

6.1 Bounding the three terms

We now bound each term in the decomposition (67) in turn.

6.1.1 Bounding the term I_1

The bound for term I_1 follows directly from Proposition 1. In particular, given a sample size $n \geq \frac{8}{\eta(1-\kappa)}\log\big(\|\theta_0-\bar{\theta}\|_2 d/\eta\big)$ and burn-in period $n_0 = n/2$, we have

$$\mathbb{E}\big[\|\theta_n-\bar{\theta}\|_2^2\big] \leq \frac{c\eta}{1-\kappa}\bar{\sigma}^2\tau d, \quad \text{and} \quad \mathbb{E}\big[\|\theta_{n_0}-\bar{\theta}\|_2^2\big] \leq \frac{c\eta}{1-\kappa}\bar{\sigma}^2\tau d.$$

Noting that $\|(I-\bar{L})^{-1}\|_{\mathrm{op}} \leq (1-\kappa)^{-1}$, we conclude that

$$\mathbb{E}\big[\|I_1\|_Q^2\big] \leq \lambda_{\max}(Q)\mathbb{E}\big[\|I_1\|_2^2\big] \leq \lambda_{\max}(Q)\cdot\tfrac{c\bar{\sigma}^2\tau d}{\eta(1-\kappa)^3 n^2}. \quad (68)$$

6.1.2 Bounding the term I_2

For the term I_2, note that the process $(\Psi_t)_{t\geq n_0}$ is a martingale adapted to the natural filtration. Its second moment equals the quadratic variation:

$$\mathbb{E}\big[\|I_2\|_Q^2\big] = \frac{4}{n^2}\mathbb{E}\big[[Q^{1/2}(I-\bar{L})^{-1}\Psi]_{n_0,n}\big] = \frac{4}{n^2}\sum_{t=n_0}^{n-1}\mathbb{E}\big[\|(I-\bar{L})^{-1}\big((L_{t+1}-\boldsymbol{L}(s_t))\theta_t + b_{t+1} - \boldsymbol{b}(s_t)\big)\|_Q^2\big].$$

By the Cauchy–Schwarz inequality, we have the bound

$$\begin{aligned}
\mathbb{E}\big[\|I_2\|_Q^2\big] &\leq \tfrac{8}{n^2}\sum_{t=n_0}^{n-1}\mathbb{E}\big[\|(I-\bar{L})^{-1}\zeta_{t+1}\|_Q^2\big] + \tfrac{8}{n^2}\sum_{t=n_0}^{n-1}\mathbb{E}\big[\|(I-\bar{L})^{-1}Z_{t+1}\Delta_t\|_Q^2\big] \\
&\leq \tfrac{16}{n}\mathrm{Tr}\big(Q(I-\bar{L})^{-1}\Sigma_{\mathrm{MG}}^*(I-\bar{L})^{-\top}\big) + \tfrac{16\sigma_L^2\lambda_{\max}(Q)d}{(1-\kappa)^2n^2}\sum_{t=n_0}^{n-1}\mathbb{E}\big[\|\Delta_t\|_2^2\big] \\
&\leq \tfrac{16}{n}\mathrm{Tr}\big((I-\bar{L})^{-1}\Sigma_{\mathrm{MG}}^*(I-\bar{L})^{-\top}\big) + \lambda_{\max}(Q)\cdot\tfrac{16\sigma_L^2 d}{(1-\kappa)^2 n}\cdot\tfrac{c\eta d\tau}{1-\kappa}\bar{\sigma}^2. \quad (69)
\end{aligned}$$

6.1.3 Bounding the term I_3

Applying the Cauchy-Schwarz inequality yields

$$\mathbb{E}\big[\|(I-\bar{L})^{-1}\Upsilon_{n_0,n}\|_2^2\big] \leq 2\mathbb{E}\big[\|\sum_{t=n_0}^{n-1}(I-\bar{L})^{-1}\nu_t\|_2^2\big] + 2\mathbb{E}\big[\|\sum_{t=n_0}^{n-1}(I-\bar{L})^{-1}N_t\Delta_t\|_2^2\big]. \tag{70}$$

We make use of the two auxiliary lemmas in order to control the terms in the decomposition (70).

Lemma 7. *Under the setup above, for a sample size n satisfying the bound $\frac{n}{\log n} \geq 2t_{\mathrm{mix}}\log(c_0 d)$, there exists a universal constant $c>0$ such that*

$$\mathbb{E}\big[\|\sum_{t=n_0}^{n-1}(I-\bar{L})^{-1}\nu_t\|_Q^2\big] \leq (n-n_0)\cdot \mathrm{Tr}\big(Q(I-\bar{L})^{-1}\Sigma_{\mathrm{Mkv}}^*(I-\bar{L})^{-\top}\big) + \lambda_{max}(Q)\cdot\frac{ct_{\mathrm{mix}}^2\bar{\sigma}^2 d}{(1-\kappa)^2}\log^2(c_0 d).$$

See Section 6.2.1 for the proof of this claim.

Lemma 8. *Under the above conditions, there exists a universal constant $c>0$s such that for any scalar $\tau \geq 3t_{\mathrm{mix}}\log^2(c_0 dn)$, stepsize $\eta \in \big(0, \frac{1-\kappa}{c\tau(\sigma_L^2 d+\gamma_{\max}^2)}\big]$ and burn-in time $n_0 \geq \tau + \frac{2}{(1-\kappa)\eta}\log(nd)$, we have $\mathbb{E}\big[\|\sum_{t=n_0}^{n-1}N_t\Delta_t\|_2^2\big] \leq c\eta^2 n^2\tau^2 d^2\sigma_L^2\bar{\sigma}^2$.*

See Section 6.2.2 for the proof of this claim.

We now exploit the preceding two lemmas to upper bound the term I_3. We have

$$\begin{aligned}\mathbb{E}\big[\|I_3\|_Q^2\big] &\leq \tfrac{2}{(n-n_0)^2}\mathbb{E}\big[\|\sum_{t=n_0}^{n-1}(I-\bar{L})^{-1}\nu_t\|_Q^2\big] + \tfrac{2}{(n-n_0)^2}\mathbb{E}\big[\|\sum_{t=n_0}^{n-1}(I-\bar{L})^{-1}N_t\Delta_t\|_Q^2\big] \\ &\leq \tfrac{8\mathrm{Tr}\left(Q(I-\bar{L})^{-1}\Sigma_{\mathrm{Mkv}}^*(I-\bar{L})^{-\top}\right)}{n} + \lambda_{\max}(Q)\cdot\tfrac{ct_{\mathrm{mix}}^2\bar{\sigma}^2 d}{(1-\kappa)^2 n^2}\log^2(c_0 d) + \lambda_{\max}(Q)\cdot\tfrac{c\eta^2\tau^2 d^2\sigma_L^2\bar{\sigma}^2}{(1-\kappa)^2}.\end{aligned} \tag{71}$$

Collecting the bounds (68), (69), and (71), we find that

$$\begin{aligned}\mathbb{E}\big[\|\widehat{\theta}_n-\bar{\theta}\|_Q^2\big] \leq \tfrac{c}{n}&\mathrm{Tr}\big(Q(I-\bar{L})^{-1}(\Sigma_{\mathrm{MG}}^*+\Sigma_{\mathrm{Mkv}}^*)(I-\bar{L})^{-\top}\big) \\ &+ \lambda_{\max}(Q)\cdot\big[\frac{c\bar{\sigma}^2 t_{\mathrm{mix}}d}{\eta(1-\kappa)^3 n^2} + \frac{16\sigma_L^2 d}{(1-\kappa)^2 n}\cdot\frac{c\eta d t_{\mathrm{mix}}}{1-\kappa}\bar{\sigma}^2\big] \\ &+ \lambda_{\max}(Q)\cdot\big[\frac{ct_{\mathrm{mix}}^2\bar{\sigma}^2 d}{(1-\kappa)^2 n^2}\log^2(c_0 dn) + \frac{c\eta^2 t_{\mathrm{mix}}^2 d^2\sigma_L^2\bar{\sigma}^2}{(1-\kappa)^2}\big].\end{aligned}$$

For a sample size n lower bounded as $\frac{n}{\log^2 n} \geq \frac{2t_{\mathrm{mix}}(\sigma_L^2 d+\gamma_{\max}^2)}{(1-\kappa)^2}\log(c_0 d)$, we can take the optimal stepsize $\eta = \big[c\big((1-\kappa)n^2 t_{\mathrm{mix}}(\sigma_L^2 d+\gamma_{\max}^2)\big)\big]^{-1/3}$. With this choice, we have

$$\mathbb{E}\big[\|\widehat{\theta}_n-\bar{\theta}\|_Q^2\big] \leq \tfrac{c}{n}\mathrm{Tr}\big(Q(I-\bar{L})^{-1}(\Sigma_{\mathrm{MG}}^*+\Sigma_{\mathrm{Mkv}}^*)(I-\bar{L})^{-\top}\big) + c\lambda_{\max}(Q)\cdot\big(\frac{\sigma_L^2 d t_{\mathrm{mix}}}{(1-\kappa)^2 n}\big)^{4/3}\log^2 n. \tag{72}$$

Setting $Q := I_d$ completes the proof.

6.2 Proof of auxiliary results

In this section, we prove the two auxiliary results used in the proof of Theorem 1: namely, Lemma 7 and Lemma 8.

6.2.1 Proof of Lemma 7

Given an integer $k \geq 0$, we define the k-step correlation under the stationary Markov chain as

$$\mu_k := \mathbb{E}_{s\sim\xi, s'\sim P^k\delta_s}\big[\langle Q^{1/2}(I-\bar{L})^{-1}\nu(s),\, Q^{1/2}(I-\bar{L})^{-1}\nu(s')\rangle\big].$$

Clearly, we have $\mu_0 \geq 0$, and by Cauchy–Schwarz inequality, for any $k \geq 0$, there is:

$$|\mu_k| \leq \sqrt{\mathbb{E}_{s\sim\xi}\|(I-\bar{L})^{-1}\nu(s)\|_Q^2}\cdot\sqrt{\mathbb{E}_{s'\sim\xi}\|(I-\bar{L})^{-1}\nu(s')\|_Q^2} = \mu_0.$$

The desired quantity can be written as $\mathrm{Tr}\big(Q^{1/2}(I-\bar{L})^{-1}\Sigma^*_{\mathrm{Mkv}}(I-\bar{L})^{-\top}Q^{1/2}\big) = \mu_0 + 2\sum_{k=1}^{+\infty}\mu_k$. Expanding the squared norm yields

$$\begin{aligned}\mathbb{E}\big[\|\sum_{t=n_0}^{n-1} Q^{1/2}(I-\bar{L})^{-1}\nu_t\|_2^2\big] &= \sum_{n_0\leq t_1,t_2\leq n-1}\mathbb{E}\big[\langle Q^{1/2}(I-\bar{L})^{-1}\nu(s_{t_1}),\, Q^{1/2}(I-\bar{L})^{-1}\nu(s_{t_2})\rangle\big]\\ &= (n-n_0)\mu_0 + 2\sum_{k=1}^{n-n_0-1}(n-n_0-k)\mu_k.\end{aligned}$$

We claim that the cross-correlations μ_k satisfy the bound

$$|\mu_k| \leq c_0\frac{\bar{\sigma}^2\|Q\|_{\mathrm{op}}d^2}{(1-\kappa)^2}\cdot 2^{1-\frac{k}{2t_{\mathrm{mix}}}}. \tag{73}$$

We return to prove this fact momentarily. Taking it as given, this inequality, in conjunction with the bound $|\mu_k| \leq \mu_0$, can be employed to bound the tail sums needed for the proof. We have

$$\Big|\sum_{k=1}^{n-n_0-1} k\mu_k\Big| \leq \sum_{k=1}^{\tau}\tau|\mu_k| + \sum_{k=\tau+1}^{\infty}k|\mu_k| \leq \tau^2\mu_0 + 2c_0\frac{\bar{\sigma}^2\|Q\|_{\mathrm{op}}d^2}{(1-\kappa)^2}\sum_{k=\tau+1}^{\infty}k\cdot 2^{-\frac{k}{2t_{\mathrm{mix}}}}.$$

With the choice $\tau := 2t_{\mathrm{mix}}\log(c_0 d)$, simplifying yields

$$\begin{aligned}\Big|\sum_{k=1}^{n-n_0-1} k\mu_k\Big| &\leq \frac{\tau^2\bar{\sigma}^2 d\|Q\|_{\mathrm{op}}}{(1-\kappa)^2} + 2c_0\frac{\bar{\sigma}^2 d^2\|Q\|_{\mathrm{op}}}{(1-\kappa)^2}\cdot 2t_{\mathrm{mix}}\big(\tau+1+2t_{\mathrm{mix}}\big)\cdot 2^{-\frac{\tau+1}{t_{\mathrm{mix}}}}\\ &\leq \frac{2\tau^2\bar{\sigma}^2 d}{(1-\kappa)^2}\|Q\|_{\mathrm{op}},\end{aligned}$$

and for n satisfying $\frac{n}{\log n} \geq 2\log(c_0 d t_{\mathrm{mix}})$, we have:

$$\sum_{k=n-n_0}^{\infty}|\mu_k| \leq 2c_0\frac{\bar{\sigma}^2 d^2\|Q\|_{\mathrm{op}}}{(1-\kappa)^2}\sum_{k=\frac{1}{2}n}^{\infty}\cdot 2^{-\frac{k}{2t_{\mathrm{mix}}}} \leq 2c_0\frac{\bar{\sigma}^2 d^2\|Q\|_{\mathrm{op}}}{(1-\kappa)^2}\cdot 2^{-\frac{n}{2t_{\mathrm{mix}}}} \leq 2c_0\frac{\bar{\sigma}^2 d}{(1-\kappa)^2 n^2}\|Q\|_{\mathrm{op}}.$$

Putting together these bounds yields

$$\begin{aligned}\mathbb{E}\big[\|\sum_{t=n_0}^{n-1}(I-\bar{L})^{-1}\nu_t\|_Q^2\big] &= (n-n_0)\big(\mu_0+2\sum_{k=1}^{\infty}\mu_k\big)-2(n-n_0)\sum_{k=n-n_0}^{\infty}\mu_k-2\sum_{k=1}^{n-n_0-1}k\mu_k\\ &\le (n-n_0)\cdot\mathrm{Tr}\big((I-\bar{L})^{-1}\Sigma_{\mathrm{Mkv}}^{*}(I-\bar{L})^{-1}\big)+\frac{3\tau^2\bar{\sigma}^2 d}{(1-\kappa)^2}\|\!|Q|\!\|_{\mathrm{op}},\end{aligned}$$

which completes the proof of the lemma.

Proof of equation (73) Let $s_0\sim\xi$ and $(s_t)_{t\ge 0}$ be a stationary Markov chain starting from s_0. By the construction given in Section 5.2.1, there exists a random variable $\widetilde{s}_k$, such that $\widetilde{s}_k$ is independent of s_0, $\widetilde{s}_k\sim\xi$, and such that $\mathbb{E}\big[\rho(s_k,\widetilde{s}_k)\mid s_0\big]\le c_0\cdot 2^{1-\frac{k}{t_{\mathrm{mix}}}}$. We then obtain the bound

$$\begin{aligned}|\mu_k| &= \Big|\mathbb{E}\big[\langle Q^{1/2}(I-\bar{L})^{-1}\nu(s_0),\, Q^{1/2}(I-\bar{L})^{-1}\nu(s_k)\rangle\big]\Big|\\ &\le \Big|\mathbb{E}\big[\langle Q^{1/2}(I-\bar{L})^{-1}\nu(s_0),\, \mathbb{E}\big[Q^{1/2}(I-\bar{L})^{-1}\nu(\widetilde{s}_k)\mid s_0\big]\rangle\big]\Big|\\ &\quad+\Big|\mathbb{E}\big[Q^{1/2}\langle(I-\bar{L})^{-1}\nu(s_0),\, \mathbb{E}\big[Q^{1/2}(I-\bar{L})^{-1}\big(\nu(s_k)-\nu(\widetilde{s}_k)\big)\mid s_0\big]\rangle\big]\Big|\\ &\le 0+\sqrt{\mathbb{E}\big[\|Q^{1/2}(I-\bar{L})^{-1}\nu(s_0)\|_2^2\big]}\cdot\sqrt{\mathbb{E}\big[\|Q^{1/2}(I-\bar{L})^{-1}\big(\nu(s_k)-\nu(\widetilde{s}_k)\big)\|_2^2\big]}\\ &\le \sqrt{\mu_0}\cdot\frac{1}{1-\kappa}\sqrt{\mathbb{E}\big[\rho(s_k,\widetilde{s}_k)^2\cdot(\sigma_L\|\bar{\theta}\|_2+\sigma_b)^2d^2\big]}\\ &\le c_0\frac{\bar{\sigma}d}{1-\kappa}\sqrt{\mu_0}\cdot 2^{1-\frac{k}{2t_{\mathrm{mix}}}}.\end{aligned}\tag{74}$$

On the other hand, applying the moment condition (2) yields $\mu_0\le\frac{1}{(1-\kappa)^2}\cdot\mathbb{E}\big[\|\nu(s_0)\|_Q^2\big]\le\frac{\bar{\sigma}^2 d}{(1-\kappa)^2}\|\!|Q|\!\|_{\mathrm{op}}$. Substituting this bound into our previous inequality (74) completes the proof.

6.2.2 Proof of Lemma 8

The proof of this claim relies on a bootstrap argument: we bound the summation of interest by a more complicated summation that involves product of noise matrices. Recursively applying the result for $m=\log d$ times yields the desired bound.

Lemma 9. *Given any integer $m\ge 0$, deterministic sequence $0=k_0<k_1<\cdots<k_m<n_0$, and scalar $\tau\ge 3mt_{\mathrm{mix}}p\log(c_0dn)$, we have the second moment bound*

$$\begin{aligned}&\mathbb{E}\big[\|\sum_{t=n_0}^{n-1}\big(\prod_{j=0}^{m}N_{t-k_j}\big)\Delta_{t-k_m}\|_2^2\big]\\ &\le 2n^2d^{2m}\sigma_L^{2m+2}\cdot\frac{c\eta}{1-\kappa}dt_{\mathrm{mix}}\bar{\sigma}^2+4\eta^2\tau\sum_{k_{m+1}=k_m+1}^{k_m+\tau}\mathbb{E}\big[\|\sum_{t=n_0}^{n}\big\{\prod_{j=0}^{m+1}N_{t-k_j}\Delta_{t-k_{m+1}}\big\}\|_2^2\big]\\ &\quad+4\eta^2\tau\sum_{k_{m+1}=k_m+1}^{k_m+\tau}\mathbb{E}\big[\|\sum_{t=n_0}^{n}\big\{\prod_{j=0}^{m}N_{t-k_j}\big(\nu_{t-k_{m+1}}+\zeta_{t-k_{m+1}+1}\big)\big\}\|_2^2\big],\end{aligned}\tag{75a}$$

and in the special case $m = 0$, we have

$$\mathbb{E}\big[\|\sum_{t=n_0}^{n-1} N_t\Delta_t\|_2^2\big] \le c\sigma_L^2 d\cdot\big(n\tau + n^2\eta^2\sigma_L^2 d\tau^2\big)\frac{c\eta}{1-\kappa}dt_{\mathrm{mix}}\bar{\sigma}^2 + 4\eta^2\tau\sum_{k_1=1}^{\tau}\mathbb{E}\big[\|\sum_{t=n_0}^{n} N_t N_{t-k_1}\Delta_{t-k_1}\|_2^2\big]$$
$$+ 4\eta^2\tau\sum_{k_1=1}^{\tau}\mathbb{E}\big[\|\sum_{t=n_0}^{n} N_t\big(\nu_{t-k_1}+\zeta_{t-k_1+1}\big)\|_2^2\big]. \tag{75b}$$

See Appendix C.1 for the proof of this lemma.

The following lemma controls the last term of the bound (75a):

Lemma 10. *Under the setup above, there exists a universal constant $c > 0$, such that for any integer $m > 0$ and deterministic sequence $0 = k_0 < k_1 < \cdots < k_m < n_0$, we have:*

$$\mathbb{E}\big[\|\sum_{t=n_0}^{n-1}\big(\prod_{j=0}^{m-1} N_{t-k_j}\big)\big(\nu_{t-k_m}+\zeta_{t-k_m+1}\big)\|_2^2\big] \le c\big(n^2 + nd(k_m + t_{\mathrm{mix}}\log(c_0 d))\big)\sigma_L^{2m} d^{2m}\bar{\sigma}^2.$$

See Appendix C.2 for the proof of this lemma.

Taking these lemmas as given, we now proceed with the proof of Lemma 8. Given the scalar $\tau := 3t_{\mathrm{mix}}\log^2(c_0 dn)$, we define

$$\mathfrak{H}_m := \sup_{0=k_0<k_1<\cdots<k_m\le\tau}\mathbb{E}\big[\|\sum_{t=n_0}^{n-1}\big(\prod_{j=0}^{m} N_{t-k_j}\big)\Delta_{t-k_m}\|_2^2\big]$$

for $m = 0, 1, 2, \cdots, \log d$. By equation (75b) and Lemma 10, we have the bound

$$\mathfrak{H}_0 \le c\sigma_L^2 d\cdot\big(n\tau + n^2\eta^2\sigma_L^2 d\tau^2\big)\frac{c\eta}{1-\kappa}dt_{\mathrm{mix}}\bar{\sigma}^2 + 4\eta^2\tau^2\mathfrak{H}_1 + 4c\eta^2\tau^2\big(n^2 + nd(\tau + t_{\mathrm{mix}}\log(c_0 d))\big)\sigma_L^2 d^2\bar{\sigma}^2$$
$$\le 4\eta^2\tau^2\mathfrak{H}_1 + c'\eta^2 n^2\tau^2 d^2\sigma_L^2\bar{\sigma}^2.$$

In deriving the last inequality, we used the inequalities $\eta \le \frac{1-\kappa}{\sigma_L^2 d\tau}$ and $n \ge \frac{1}{(1-\kappa)\eta}$.

By equation (75a) and Lemma 10, we have the recursive relation

$$\mathfrak{H}_m \le 4\eta^2\tau^2\mathfrak{H}_{m+1} + cn^2 d^{2m+1}\tau\sigma_L^{2m+2}\cdot\frac{\eta\log^3 n}{1-\kappa}\bar{\sigma}^2 + c\eta^2\tau^2 n^2\sigma_L^{2m+2}d^{2m+2}\bar{\sigma}^2$$
$$\le 4\eta^2\tau^2\mathfrak{H}_{m+1} + cn^2\sigma_L^{2m}d^{2m}\bar{\sigma}^2\cdot\log^3 n.$$

Recursively applying these bounds yields

$$\mathfrak{H}_0 \le (4\eta^2\tau^2)^m\mathfrak{H}_m + c\eta^2 n^2\tau^2 d^2\sigma_L^2\bar{\sigma}^2 + c\cdot\sum_{q=1}^{m-1}(4\eta^2\tau^2)^q n^2\sigma_L^{2q}d^{2q}\bar{\sigma}^2 \ \le (4\eta^2\tau^2)^m\mathfrak{H}_m + 3c\eta^2 n^2\tau^2 d^2\sigma_L^2\bar{\sigma}^2.$$

In order to control the term $\mathfrak{H}_m$, we employ the coarse bound

$$\mathbb{E}\big[\|\sum_{t=n_0}^{n-1}\big(\prod_{j=0}^{m} N_{t-k_j}\big)\Delta_{t-k_m}\|_2^2\big] \le n\sum_{t=n_0}^{n-1}\mathbb{E}\big[\|\big(\prod_{j=0}^{m} N_{t-k_j}\big)\Delta_{t-k_m}\|_2^2\big] \le n^2(\sigma_L d)^{2m+2}\cdot\frac{c\eta t_{\mathrm{mix}}d\bar{\sigma}^2}{1-\kappa}.$$

Taking the supremum and noting that $\eta \le \frac{1-\kappa}{\sigma_L^2 d\tau}$ leads to $\mathfrak{H}_m \le cn^2\sigma_L^{2m}d^{2m+2}\bar{\sigma}^2$. Consequently, we have established that $\mathfrak{H}_0 \le 3c\eta^2 n^2\tau^2 d^2\sigma_L^2\bar{\sigma}^2\big[1 + \big(2\eta\tau\sigma_L d\big)^{\frac{2m+2}{2m}}\big]$. Taking $m = \lceil\log d\rceil$ and $\eta \le \frac{1}{6\tau\sigma_L d}$, we have $(2\eta\tau\sigma_L d^{\frac{2m+2}{2m}})^{2m} < 1$, and thus $\mathfrak{H}_0 \le 6c\eta^2 n^2\tau^2 d^2\sigma_L^2\bar{\sigma}^2\log^3 n$, which completes the proof of this lemma.

7 Proof of Theorem 2

Our strategy is to prove a Bayes risk lower bound. We construct a prior distribution over transition kernels by perturbing the base matrix P_0 appropriately. We then apply the Bayesian Cramér–Rao lower bound to obtain our result.

Let us describe the construction in more detail. For each $s \in \mathbb{X}$, suppose we have a perturbation vector $h_s \in \mathbb{R}^{\mathbb{X}}$. Use these to define the perturbed transition kernel

$$P_h(x,y) := \frac{P_0(x,y)\, e^{h_x(y)}}{\sum_{z\in\mathbb{X}} P_0(x,z) e^{h_x(z)}} \qquad \text{for each } x, y \in \mathbb{X}.$$

Note that by construction, for any $x \in \mathbb{X}$ and any $h_x \in \mathbb{R}^{\mathbb{X}}$, we have $\operatorname{supp}\big(P_h(x,\cdot)\big) = \operatorname{supp}\big(P_0(x,\cdot)\big)$. Since P_0 is irreducible and aperiodic, so is P_h. Therefore, the stationary distribution ξ_h of P_h exists and is unique. When the perturbation is small enough, a quantitative perturbation principle can be obtained, which we collect in Lemma 11 below.

It remains to specify how the perturbation vectors are generated. We parameterize h with a linear transformation, writing $h = Qw$ for a linear operator Q to be specified shortly, and a random vector $w \in \mathbb{R}^d$ drawn from a distribution ρ. In particular, given a collection of vectors $\{q_x(y)\}_{x,y\in\mathbb{X}} \subseteq \mathbb{R}^d$, we consider the linear transformation $Q : \mathbb{R}^d \to \mathbb{R}^{\mathbb{X}\times\mathbb{X}}$ given by $w \mapsto \big[\langle w,\, q_x(y)\rangle\big]_{x,y\in\mathbb{X}}$.

Next we specify the prior ρ, and along with some associated notation. Define the subspace $\mathbb{H}_h := \big\{f \in \mathbb{R}^{\mathbb{X}} : \mathbb{E}_{\xi_h}[f(s)] = 0\big\}$, and note that P_h maps $\mathbb{H}_h$ to itself. Furthermore, since P_h is irreducible and aperiodic, the mapping $(I - P_h)$ is invertible on $\mathbb{H}_h$. Consequently, for any function $f : \mathbb{X} \to \mathbb{R}$, the following Green function operator is well-defined:

$$\mathcal{A}_h f := (I - P_h)^{-1}\big|_{\mathbb{H}_h} \cdot \big(f - \mathbb{E}_{\xi_h}[f]\big) \in \mathbb{R}^{\mathbb{X}}.$$

We also define an operator $\mathcal{P}_h$ on the space of real-valued functions on $\mathbb{X}$ as follows:

$$\mathcal{P}_h f(x) := \mathbb{E}_{Y\sim P_h(x,\cdot)}[f(Y)].$$

Importantly, $\mathcal{P}_h$ is an operator mapping functions to functions, and distinct from the matrix P_h. It is straightforward to see that the operator $\mathcal{P}_h$ commutes with the operator $\mathcal{A}_h$, for any perturbation matrix h. Finally, for any $h \in \mathbb{R}^{\mathbb{X}\times\mathbb{X}}$ and for all $x \in \mathbb{X}$, we define

$$\boldsymbol{g}_h(x) = \big(I_d - \mathbb{E}_{\xi_h}[\boldsymbol{L}(s)]\big)^{-1}\big(\mathcal{A}_h \boldsymbol{L}(x)\cdot\bar{\theta}(P_h) + \mathcal{A}_h \boldsymbol{b}(x)\big). \tag{76}$$

Since the proof works under the perturbed probability transition kernel P_h, it is useful to study the effect of small perturbation on its stationary distribution. The following lemma provides non-asymptotic bounds on the mixing time of perturbed Markov chain and its stationary distribution ξ_h, which will be useful throughout the proof.

Lemma 11. *Under the setup above, suppose that* $h_{\max} := \max_{x\in\mathbb{X}} \|h_x\|_\infty < \frac{1}{128 t_{\mathrm{mix}}}$. *Then the perturbed transition kernel satisfies the following.*

- *The Markov transition kernel P_h satisfies the mixing condition (Assumption 1) with the discrete metric and mixing time $4t_{\mathrm{mix}}$.*
- *The stationary distribution ξ_h satisfies the bound*

$$\max_{s\in\mathbb{X}} \Big\{ \log\tfrac{\xi_0(s)}{\xi_h(s)},\ \ \log\tfrac{\xi_h(s)}{\xi_0(s)}\Big\} \le t_{\mathrm{mix}}\big(2 + \log h_{\max}^{-1} + \log\tfrac{1}{\min_x \xi_0(x)}\big) h_{\max}.$$

See Section 7.1 for the proof of this lemma.

With this notation in hand, we are ready to construct the prior distribution on w. We begin with the following one-dimensional density function, taken from Tsybakov [Tsy08]:

$$\mu(t) := \cos^2\left(\tfrac{\pi t}{2}\right) \cdot \mathbf{1}_{t\in[-1,1]}. \tag{77a}$$

Also, define the positive-definite matrix $\Lambda := \mathbb{E}_{X\sim\xi_0}\left[\operatorname{cov}_{Y\sim P_0(X,\cdot)}\left(\boldsymbol{g}_0(Y) \mid X\right)\right]$, and let $\Lambda = UDU^\top$ denote its eigen-decomposition. For a random variable $\psi \sim \mu^{\otimes d}$, define the perturbation parameter

$$w = \tfrac{1}{\sqrt{n}} U D^{-1/2}\psi, \tag{77b}$$

and let its density denote the prior distribution ρ. Note that for any $w \in \operatorname{supp}(\rho)$, we have

$$\|\Lambda w\|_2 = \|UD^{1/2}\psi\|_2 = \|D^{1/2}\psi\|_2 \le \sqrt{\operatorname{trace}(D)/n} = \sqrt{\operatorname{trace}(\Lambda)/n}. \tag{77c}$$

The final ingredient in our construction is to specify the linear transformation Q. For each $x, y \in \mathbb{X}$, we set

$$q_x(y) := \boldsymbol{g}_0(y) - \mathbb{E}_{s'\sim P_0(x,\cdot)}\left[\boldsymbol{g}_0(s')\right], \tag{77d}$$

where the Green function $\boldsymbol{g}$ is defined in equation (76). Recall that $h = Qw$ for $w \sim \rho$. This specifies our prior over transition kernels, and concludes the construction.

Next, we state the version of the Bayesian Cramér–Rao bound that we use. Before stating the result, it is useful to introduce the general setup and basic notation for parametric models. Given a family $\mathcal{P}_\Theta = \left(\mathbb{P}_\eta : \eta \in \Theta\right)$ of probability distributions of sample $X \in \mathbb{X}$, parameterized by $\eta \in \Theta$, where Θ is an open subset of $\mathbb{R}^d$. Assume that each element in this family is absolute continuous with respect to a base measure λ over $\mathbb{X}$, and denote the Radon–Nikodym derivative by $p_\eta := \frac{d\mathbb{P}_\eta}{d\lambda}$. Assuming differentiability and integrability of relevant quantities, for any $\eta \in \Theta$, we define the Fisher information matrix $I(\eta)$ as

$$I(\eta) := \mathbb{E}_{X\sim\mathbb{P}_\eta}\left[\nabla_\eta \log p_\eta(X) \nabla_\eta \log p_\eta(X)^\top\right] \in \mathbb{R}^{d\times d}.$$

Now we are ready to state the Bayesian Cramér–Rao lower bound.

Proposition 4 (Theorem 1 of [GL95], special case)**.** *Under the setup above, given a prior distribution ρ with continuously differentiable density and bounded support contained within Θ, let $T : \operatorname{supp}(\rho) \mapsto \mathbb{R}^d$ denote a locally continuously differentiable functional. Then for any estimator $\widehat{T}$ based on observing X, we have*

$$\mathop{\mathbb{E}}_{\eta\sim\rho}\mathop{\mathbb{E}}_{X\sim p_\eta}\|\widehat{T}(X) - T(\eta)\|_2^2 \ge \frac{\left(\int \operatorname{trace}\left(\frac{\partial T}{\partial \eta}(\eta)\right)\rho(\eta)d\eta\right)^2}{\int \operatorname{trace}\left(I(\eta)\right)\rho(\eta)d\eta + \int \|\nabla \log \rho(\eta)\|_2^2 \rho(\eta)d\eta}. \tag{78}$$

In order to complete the proof, we provide non-asymptotic estimates on the three quantities involved in the right-hand-side of Proposition 4. These require a few technical lemmas, whose proofs can be found at the end of the section.

Bounds on the term $\operatorname{trace}\left(\nabla_w \bar{\theta}\right)$**:** We state two technical lemmas that are helpful in bounding this quantity. The first computes the Jacobian matrix of the desired functional $\bar{\theta}(h)$ with respect to the parameter w.

Lemma 12. *Under the given set-up, for any $w \in \mathbb{R}^d$, we have*

$$\nabla_w \bar{\theta}(P_h) = \mathbb{E}_{X\sim\xi_h}\Big[\operatorname{cov}_{Y\sim P_h(X,\cdot)}\big\{\boldsymbol{g}_h(Y) - \mathcal{P}_h\boldsymbol{g}_h(X), \boldsymbol{g}_0(Y) - \mathcal{P}_0\boldsymbol{g}_0(X) \mid X\big\}\Big]. \tag{79}$$

See Section 7.2 for the proof of this lemma. Next, we control the RHS of equation (79) by replacing $\boldsymbol{g}_h$ with $\boldsymbol{g}_0$.

Lemma 13. *Under the given set-up and for a sample size lower bounded as $n \geq \frac{ct_{\text{mix}}^2\sigma_L^2 d^2 \log^2 d}{(1-\kappa)^2}$ and $\max_{x\in\mathbb{X}} \|h_x\|_\infty \leq \frac{1}{128 t_{\text{mix}}}$, we have*

$$\mathbb{E}_{Z\sim\xi_h}\big[\|\boldsymbol{g}_h(Z) - \boldsymbol{g}_0(Z)\|_2^2\big] \leq \tfrac{c(1+\sigma_L^2)\bar{\sigma}^2 t_{\text{mix}}^4 d^2}{(1-\kappa)^4 n} \log^6 \tfrac{d}{\min_x \xi_0(x)}.$$

Furthermore, for any w in the support of ρ, we have

$$\|\bar{\theta}(P_h) - \bar{\theta}(P_0)\|_2 \leq \tfrac{3}{2}\sqrt{\operatorname{trace}(\Lambda)/n} + \sqrt{\tfrac{c(1+\sigma_L^2)\bar{\sigma}^2 t_{\text{mix}}^4 d^3}{(1-\kappa)^4 n^2} \log^6 \tfrac{d}{\min_x \xi_0(x)}}.$$

See Section 7.3 for the proof of this lemma.

Combining these two lemmas yields

$$\begin{aligned}
&\operatorname{trace}\big(\nabla_w \bar{\theta}\big) \\
&\geq \mathbb{E}_{X\sim\xi_h}\big[\operatorname{var}_{Y\sim P_h(X,\cdot)}\big(\boldsymbol{g}_0(Y) - \mathcal{P}_0\boldsymbol{g}_0(X) \mid X\big)\big] \\
&\quad - \mathbb{E}_{X\sim\xi_h}\Big[\sqrt{\operatorname{var}_{Y\sim P_h(X,\cdot)}\big(\boldsymbol{g}_0(Y) - \mathcal{P}_0\boldsymbol{g}_0(X) \mid X\big)}\Big] \cdot \sqrt{\mathbb{E}_{Z\sim\xi_h}\big[\|\boldsymbol{g}_h(Z) - \boldsymbol{g}_0(Z)\|_2^2\big]} \\
&\geq \operatorname{trace}\big(\Lambda\big) - \sqrt{\operatorname{trace}\big(\Lambda\big)} \cdot \tfrac{c(1+\sigma_L)\bar{\sigma} t_{\text{mix}}^2 d}{(1-\kappa)^2\sqrt{n}} \log^3 \tfrac{d}{\min_x \xi_0(x)}.
\end{aligned}$$

Now given a sample size lower bounded as $n \geq \frac{ct_{\text{mix}}^2\sigma_L^2 d^2 \log^2 d}{(1-\kappa)^2} + \frac{2c(1+\sigma_L^2)\bar{\sigma}^2 t_{\text{mix}}^4 d^2}{(1-\kappa)^4 \operatorname{trace}(\Lambda)} \log^6 \frac{d}{\min_x \xi_0(x)}$, we can conclude that

$$\operatorname{trace}\big(\nabla_w \bar{\theta}\big) \geq \tfrac{1}{2}\operatorname{trace}(\Lambda) \qquad \text{for any } w \text{ in the support of } \rho. \tag{80}$$

Bounds on the Fisher information $I^{(n)}(w)$: We now state an upper bound on the Fisher information of the observed trajectory:

Lemma 14. *Under the given set-up, for any $w \in \mathbb{R}^d$, if $h_{\max} := \max_x \|h\|_\infty$ satisfies the inequality $h_{\max}^{-1} \geq ct_{\text{mix}}\big(\log h_{\max}^{-1} + \log(\min \xi_0)^{-1}\big)$, we have*

$$I^{(n)}(w) := \mathbb{E}_h\big[\nabla_w \log \mathbb{P}_h\big(s_0^n\big) \nabla_w \log \mathbb{P}_h\big(s_0^n\big)^\top\big] \preceq \tfrac{3n}{2}\mathbb{E}_{X\sim\xi_h}\big[\operatorname{cov}_{Y\sim P_h(X,\cdot)}\big(q_X(Y) \mid X\big)\big].$$

See Section 7.4 for the proof of this lemma.

In order to apply the preceding lemma, we must verify the condition on $h_{\max}$ for our setting. Under our construction, we have $\max_{x\in\mathbb{X}} \|h_x\|_\infty = \max_{x,y\in\mathbb{X}} \langle \boldsymbol{g}_0(y) - \mathcal{P}_0\boldsymbol{g}_0(x), w\rangle$. Note that Assumption 2 and Lemma 17 in the Appendix together imply the following bound for any $\delta > 0$:

$$\xi_0\big(s : |\langle \boldsymbol{g}_0(s), w\rangle| \leq \tfrac{c\bar{\sigma} t_{\text{mix}} \|w\|_2}{1-\kappa} \cdot \log^3 \tfrac{d}{\delta}\big) > 1 - \delta.$$

Taking $\delta := \frac{1}{2}\min_{s\in\mathbb{X}} \xi_0(s) > 0$, we have the uniform bound

$$\max_{s\in\mathbb{X}} |\langle \boldsymbol{g}_0(s),\, w\rangle| \le \tfrac{c\bar{\sigma} t_{\mathrm{mix}}\|w\|_2}{1-\kappa} \log^3\big(d/\min_s \xi_0(s)\big).$$

Note that $\mathcal{P}_0$ is a probability transition kernel, for any $s \in \mathbb{X}$, the vector $\mathcal{P}_0\boldsymbol{g}_0(s)$ lies in the convex hull of $\big(\boldsymbol{g}_0(s')\big)_{s'\in\mathbb{X}}$. So we have the bound $\max_{s\in\mathbb{X}} |\langle \mathcal{P}_0\boldsymbol{g}_0(s),\, w\rangle| \le \max_{s\in\mathbb{X}} |\langle \boldsymbol{g}_0(s),\, w\rangle| \le \frac{c\bar{\sigma} t_{\mathrm{mix}}\|w\|_2}{1-\kappa} \log^3\big(d/\min_s \xi_0(s)\big)$. Putting them together leads to the bound

$$\max_{x\in\mathbb{X}} \|h_x\|_\infty \le 2c\bar{\sigma} t_{\mathrm{mix}}\|w\|_2 \log^3\big(d/\min_s \xi_0(s)\big).$$

Now given a sample size

$$n \ge c t_{\mathrm{mix}}^3 \bar{\sigma}^2 \cdot \mathrm{trace}(\Lambda) \cdot \log^3 \tfrac{d}{\min_s \xi_0(s)}, \tag{81}$$

we have that $\max_x \|h_x\|_\infty < \frac{1}{128 t_{\mathrm{mix}}}$. This satisfies the condition in Lemma 11 in the appendix. Applying this lemma, we see that the condition

$$h_{\mathrm{max}}^{-1} \ge c t_{\mathrm{mix}}\big(\log h_{\mathrm{max}}^{-1} + \log(\min \xi_0)^{-1}$$

is satisfied, so that Lemma 14 guarantees that

$$\begin{aligned}
\mathrm{trace}\left(I^{(n)}(w)\right) &\preceq \tfrac{3n}{2}\mathbb{E}_{X\sim\xi_h}\big[\,\mathrm{var}_{Y\sim P_h(X,\cdot)}\left(\boldsymbol{g}_0(Y) - \mathcal{P}_0\boldsymbol{g}_0(X) \mid X\right)\big] \\
&\preceq \left(\tfrac{3}{2}\right)^3 n \cdot \mathbb{E}_{X\sim\xi_0}\big[\,\mathrm{var}_{Y\sim P_0(X,\cdot)}\left(\boldsymbol{g}_0(Y) \mid X\right)\big] \\
&= \tfrac{27n}{8}\,\mathrm{trace}\left(\Lambda\right).
\end{aligned} \tag{82}$$

The last inequality follows because $\xi_h \preceq \frac{3}{2}\xi_0$ $P_h(x,\cdot) \preceq \frac{3}{2}P_0(x,\cdot)$ for all $x \in \mathbb{X}$.

Bounds on the prior Fisher information: From Lemma 10 in the paper [MPW20], the density ρ of w has Fisher information

$$I(\rho) = U D^{1/2} I\big(\mu^{\otimes d}\big) D^{1/2} U^\top = n\pi\Lambda. \tag{83}$$

Consequently, we have $\int \|\nabla \log \rho(w)\|_2^2 \rho(w)\, dw\, \mathrm{trace}\left(I(\rho)\right) = n\pi \cdot \mathrm{trace}(\Lambda)$.

Putting together the pieces: Combining the bounds (80), (82), and (83) and applying Proposition 4, we obtain the lower bound

$$\inf_{\widehat{\theta}_n} \int_{\mathbb{R}^d} \mathbb{E}_{X_1^n\sim\mathbb{P}_{Qw}}\big[\|\widehat{\theta}_n - \bar{\theta}(P_{Qw})\|_2^2\big]\rho(dw) \ge \tfrac{1}{4(5+\pi)n}\,\mathrm{trace}(\Lambda). \tag{84}$$

It remains to relate the matrix Λ to the local complexity ε_n in the theorem. In order to do so, we require the following lemma.

Lemma 15. *Under the setup above, for any function $f : \mathbb{X} \to \mathbb{R}$ such that $\mathbb{E}_{\xi_0}[f(s)] = 0$, we have $\mathbb{E}_{X\sim\xi_0, Y\sim P_0(X,\cdot)}\big[\big(\mathcal{A}_0 f(Y) - \mathcal{P}_0\mathcal{A}_0 f(X)\big)^2\big] = \sum_{k=-\infty}^{\infty} \mathbb{E}\big[f(s_0)f(s_k)\big]$, where $(s_k)_{k\in\mathbb{Z}}$ is a stationary Markov chain following P_0.*

See Section 7.5 for the proof of this lemma.

Applying Lemma 15 with $f_j(s) = \langle (I_d - \bar{L}^{(0)})^{-1}\big(\boldsymbol{L}(s)\bar{\theta}(P_0) + \boldsymbol{b}(s)\big),\, e_j\rangle$ for $j = 1, 2, \cdots, d$ respectively, we arrive at the chain of equalities

$$\begin{aligned}\mathrm{trace}(\Lambda) &= \sum_{j=1}^{d} \mathbb{E}_{X\sim\xi_0, Y\sim P_0(X,\cdot)}\big[\big(\mathcal{A}_0 f_j(Y) - \mathcal{P}_0\mathcal{A}_0 f_j(X)\big)^2\big] \\ &= \sum_{j=1}^{d}\sum_{k=-\infty}^{\infty} \mathbb{E}\big[f_j(s_0) f_j(s_k)\big] = \mathrm{trace}\big((I - \bar{L}^{(0)})^{-1}\Sigma^*_{\mathrm{Mkv}}(I - \bar{L}^{(0)})^{-\top}\big) = n\varepsilon_n^2.\end{aligned}$$

Thus, the right-hand-side of equation (84) is exactly $\frac{\varepsilon_n^2}{4(5+\pi)}$.

It remains to bound the size of the neighborhood. Given a sample size n satisfying the bound (81), Lemma 13 implies that $\|\bar{\theta}(P_h) - \bar{\theta}(P_0)\|_2 \leq \sqrt{\frac{\mathrm{trace}(\Lambda)}{n}}$. Consequently, for any w on the support of ρ, we have $P_{Qw} \in \mathfrak{N}_{\mathrm{Est}}(P_0, 2\varepsilon_n)$.

On the other hand, for any $w \in \mathrm{supp}(\rho)$ and any $x \in \mathbb{X}$ and perturbation $h = Qw$, we have

$$\begin{aligned}\chi^2\left(P_h(x,\cdot) \,||\, P_0(x,\cdot)\right) &= \mathbb{E}_{Y\sim P_0(x,\cdot)}\big[\big(\tfrac{P_h(x,Y)}{P_0(x,Y)} - 1\big)^2\big] \\ &= \mathbb{E}_{Y\sim P_0(x,\cdot)}\big[\big(e^{h_x(Y)} - 1\big)^2\big] \cdot \Big(\sum_{z\in\mathbb{X}} P_0(x,z) e^{h_x(z)}\Big)^{-2} \\ &\overset{(i)}{\leq} \mathbb{E}_{Y\sim P_0(x,\cdot)}\big[\big(e^{h_x(Y)} - 1\big)^2\big] \\ &\overset{(ii)}{\leq} e\cdot \mathbb{E}_{Y\sim P_0(x,\cdot)}\big[h_x(Y)^2\big],\end{aligned}$$

where step (i) follows by using Jensen's inequality to assert that

$$\sum_{z\in\mathbb{X}} P_0(x,z) e^{h_x(z)} \geq e^{\sum_{z\in\mathbb{X}} P_0(x,z) h_x(z)} = 1,$$

and step (ii) follows from the inequality $|e^x - 1| \leq e\cdot|x|$, valid for $x \in [-1, 1]$.

Accordingly, the average χ^2-divergence admits the bound

$$\begin{aligned}\sum_{x\in\mathbb{X}} \xi_0(x)\chi^2\left(P_h(x,\cdot) \,||\, P_0(x,\cdot)\right) &\leq e\cdot \mathbb{E}_{X\sim\xi_0, Y\sim P_0(X,\cdot)}\big[\langle w,\, \boldsymbol{g}_0(Y) - \mathcal{P}_0\boldsymbol{g}_0(X)\rangle^2\big] \\ &\leq e\cdot w^\top \Lambda w \leq \tfrac{ed}{n}.\end{aligned}$$

For any w on the support of ρ, we thus have $P_{Qw} \in \mathfrak{N}_{\mathrm{Prob}}(P_0, e\sqrt{\frac{d}{n}})$, as claimed. The Bayes risk lower bound (84) then implies the desired minimax lower bound.

7.1 Proof of Lemma 11

The proof relies on a total variation distance bound on the transition kernel. In particular, for each $s \in \mathbb{X}$, we have

$$\begin{aligned}d_{\mathrm{TV}}\big(P_0(x,\cdot), P_h(x,\cdot)\big) \leq \sqrt{\tfrac{1}{2}\chi^2\left(P_0(x,\cdot) \,||\, P_h(x,\cdot)\right)} &= \sqrt{\tfrac{1}{2}\sum_{y\in\mathbb{X}} P_0(x,y)\cdot\big(\tfrac{P_h(x,y)}{P_0(x,y)} - 1\big)^2} \\ &\leq \sqrt{\tfrac{1}{2}\big(e^{\|h_x\|_\infty} - 1\big)^2} \leq e\cdot\max_{x\in\mathbb{X}}\|h_x\|_\infty. \quad (85)\end{aligned}$$

The last inequality follows from the fact that $\|h_x\|_\infty < 1$.

Next, we turn to proofs of the two claims. We first prove the mixing time bound. Note that the non-expansive condition (6)(b) is automatically satisfied with $c_0 = 1$ for total variation distance (by a naïve coupling). Given a fixed pair $x, y \in \mathbb{X}$, invoking Lemma 4 with $\tau = 4t_{\mathrm{mix}}$ yields the existence of a joint distribution over the random sequence $\{x_k\}_{0\le k\le\tau}$ and $\{y_k\}_{0\le k\le\tau}$, such that $\{x_k\}$ and $\{y_k\}$ follows the Markov chain P_0, starting from $x_0 = x$ and $y_0 = y$, respectively. Furthermore, we have the bound $\mathbb{P}\big(x_\tau \neq y_\tau\big) \le \frac{1}{4}$.

Now we construct a coupling between the original chain and perturbed chain. Taking the initial point $\widetilde{x}_0 = x$, we iteratively construct the sequence $\{\widetilde{x}_k\}_{0\le k\le\tau}$ as follows: given $\widetilde{x}_k$ and x_k, we construct the conditional distribution of $\widetilde{x}_{k+1}$ as follows:

- If $x_k = \widetilde{x}_k$, we let $\mathbb{P}\big(\widetilde{x}_{k+1} \neq x_{k+1} \mid x_k, \widetilde{x}_k\big) = d_{\mathrm{TV}}\big(P_0(x_k,\cdot), P_h(x_k,\cdot)\big)$.
- If $x_k \neq \widetilde{x}_k$, we simply take $\widetilde{x}_{k+1}$ and x_{k+1} to be conditionally independent, following their respective transition kernels.

We construct the sequence $\{\widetilde{y}_k\}_{0\le k\le\tau}$ in a similar fashion.

By the union bound, it follows that

$$\begin{aligned}\mathbb{P}\big(x_\tau \neq \widetilde{x}_\tau\big) \le \sum_{k=0}^{\tau-1} \mathbb{E}\big[\mathbb{P}\big(x_{k+1} \neq \widetilde{x}_{k+1} \mid x_k = \widetilde{x}_k\big)\big] &= \sum_{k=0}^{\tau-1} \mathbb{E}\big[d_{\mathrm{TV}}\big(P_0(x_k,\cdot), P_h(x_k,\cdot)\big)\big] \\ &\le 4et_{\mathrm{mix}} \cdot \max_{x\in\mathbb{X}} \|h_x\|_\infty < \tfrac{1}{8}.\end{aligned}$$

In the last step, we have used the total variation distance bound (85).

Similarly, the process $\{\widetilde{y}_k\}$ satisfies the bound $\mathbb{P}\big(y_\tau \neq \widetilde{y}_\tau\big) < \frac{1}{8}$. Putting together the pieces, we conclude that

$$\begin{aligned}d_{\mathrm{TV}}\big(\delta_x P_h^\tau, \delta_y P_h^\tau\big) \le \mathbb{P}\big(\widetilde{x}_\tau \neq \widetilde{y}_\tau\big) &\le \mathbb{P}\big(\widetilde{x}_\tau \neq x_\tau\big) + \mathbb{P}\big(x_\tau \neq y_\tau\big) + \mathbb{P}\big(y_\tau \neq \widetilde{y}_\tau\big) \\ &< \tfrac{1}{8} + \tfrac{1}{4} + \tfrac{1}{8} = \tfrac{1}{2},\end{aligned}$$

which shows that the perturbed chain P_h satisfies the condition (6)(a) with mixing time $\tau = 4t_{\mathrm{mix}}$.

Next, we prove the perturbation result for the stationary distribution. Given any fixed initial distribution π_0, note that for any deterministic sequence $(x_0, x_2, \cdots, x_n)$, we have the following expression for the Radon-Nikodym derivative:

$$\frac{d\mathbb{P}_h\big(x_0,x_1,\cdots,x_n\big)}{d\mathbb{P}_0\big(x_0,x_1,\cdots,x_n\big)} = \prod_{k=0}^{n-1} \frac{P_h(x_k,x_{k+1})}{P_0(x_k,x_{k+1})} = \prod_{k=0}^{n-1} \frac{e^{h_{x_k}(x_{k+1})}}{\sum_{y\in\mathbb{X}} e^{h_{x_k}(y)} P(x_k,y)}.$$

We then have the max-divergence bound

$$D_\infty\left(\mathbb{P}_h\big(x_0^n\big) \;||\; \mathbb{P}_0\big(x_0^n\big)\right) := \sup_{x_0^n\in\mathbb{X}^n} \left|\log \frac{d\mathbb{P}_h\big(x_0,x_1,\cdots,x_n\big)}{d\mathbb{P}_0\big(x_0,x_1,\cdots,x_n\big)}\right| \le n \cdot \max_x \|h_x\|_\infty.$$

Taking the marginal distribution, we see that the bound $D_\infty\left(\pi_0 P_h^n \;||\; \pi_0 P_0^n\right) \le n \cdot h_{\max}$ holds for any initial distribution π_0 and any $n > 0$.

To obtain the desired claim, we take π_0 to be the stationary distribution ξ_h of the chain P_h, and let $n = t_{\mathrm{mix}} \log\big(\frac{2}{h_{\max}\cdot \min_x \xi_0(x)}\big)$. Note that $\pi_0 P_h^n = \xi_h$ in such case. On the other hand, by Lemma 4, the total variation distance can be upper bounded as $d_{\mathrm{TV}}\big(\pi_0 P_0^n, \xi_0\big) \le$

$2^{1-\frac{n}{t_{\mathrm{mix}}}} \le h_{\max} \cdot \min_{x\in\mathbb{X}} \xi_0(x)$. Note that this bound is smaller than $\min_{x\in\mathbb{X}} \xi_0(x)$; it translates to the max-divergence bound

$$D_\infty\left(\pi_0 P_0^n \;||\; \xi_0\right) = \max_{x\in\mathbb{X}} \left|\log \tfrac{\pi_0 P_0^n(x)}{\xi_0(x)}\right| \le \max_{x\in\mathbb{X}} \left|\tfrac{\pi_0 P_0^n(x)}{\xi_0(x)} - 1\right| \le \tfrac{d_{\mathrm{TV}}\left(\pi_0 P_0^n, \xi_0\right)}{\min_{x\in\mathbb{X}} \xi_0(x)} \le h_{\max}.$$

Finally, applying the triangle inequality yields

$$\begin{aligned} D_\infty\left(\xi_h \;||\; \xi_0\right) &\le D_\infty\left(\pi_0 P_h^n \;||\; \pi_0 P_0^n\right) + D_\infty\left(\pi_0 P_0^n \;||\; \xi_0\right) \\ &\le (n+1) h_{\max} \;\le\; t_{\mathrm{mix}}\big(2 + \log h_{\max}^{-1} + \log \tfrac{1}{\min_x \xi_0(x)}\big) h_{\max}, \end{aligned}$$

which proves the second claim.

7.2 Proof of Lemma 12

We first consider the functional $h \mapsto \bar{\theta}(P_h) := \big(I - \mathbb{E}_{\xi_h}[\boldsymbol{L}(s)]\big)^{-1} \mathbb{E}_{\xi_h}[\boldsymbol{b}(s)]$. Note that the stationary distribution ξ_h satisfies the identity $\xi_h P_h = \xi_h$. Taking derivatives, we obtain the following equality for all $x, y \in \mathbb{X}$:

$$\frac{\partial \xi_h}{\partial h_x(y)} \cdot (I - P_h) = \xi_h \cdot \frac{\partial P_h}{\partial h_x(y)} = \xi_h(x) P_h(x,y) \cdot \big[\mathbf{1}_{z=y} - P_h(x,z)\big]_{z\in\mathbb{X}}.$$

Note that the linear operator $(I - P_h)$ is invertible on the subspace $\mathbb{H}_h$. For any $f \in \mathbb{H}_h$, we have

$$\frac{\partial}{\partial h_x(y)} \mathbb{E}_{\xi_h}\big[f(s)\big] = \sum_{z\in\mathbb{X}} \frac{\partial \xi_h(z)}{\partial h_x(y)} \cdot f(s) = \xi_h(x) P_h(x,y) \cdot \big[\mathbf{1}_{z=y} - P_h(x,z)\big]_{z\in\mathbb{X}} \cdot \big(I - P_h\big)^{-1}\big|_{\mathbb{H}_h} \cdot f.$$

In the above expression, the notation $\big(I - P_h\big)^{-1}\big|_{\mathbb{H}_h}$ denotes the inverse of the operator $I - P_h$ within the subspace $\mathbb{H}_h$, a bounded linear operator on this space. Note that the derivative is invariant under translation. For any $f \in \mathbb{R}^{\mathbb{X}}$, define the auxiliary function $\widetilde{f} := f - \mathbb{E}_{\xi_h}[f]$, and write

$$\begin{aligned} \frac{\partial}{\partial h_x(y)} \mathbb{E}_{\xi_h}\big[f(s)\big] &= \frac{\partial}{\partial h_x(y)} \mathbb{E}_{\xi_h}\big[\widetilde{f}(s)\big] = \xi_h(x) P_h(x,y) \cdot \big[\mathbf{1}_{z=y} - P_h(x,z)\big]_{z\in\mathbb{X}} \cdot \big(I - P_h\big)^{-1}\big|_{\mathbb{H}_h} \cdot \widetilde{f} \\ &= \xi_h(x) P_h(x,y) \cdot \big[\mathbf{1}_{z=y} - P_h(x,z)\big]_{z\in\mathbb{X}} \cdot \big(I - P_h\big)^{-1}\big|_{\mathbb{H}_h} \cdot \big(f - \mathbb{E}_{\xi_h}[f]\big) \\ &= \xi_h(x) P_h(x,y) \cdot \big(\mathcal{A}_h f(y) - \sum_{z\in\mathbb{X}} P_h(x,z) \mathcal{A}_h f(z)\big). \end{aligned} \tag{86}$$

On the other hand, we can express the desired functional $\bar{\theta}(P_h)$ in the form above. In particular, setting $\bar{L}^{(h)} := \mathbb{E}_{\xi_h}\big[\boldsymbol{L}(s)\big]$ and $\bar{b}^{(h)} := \mathbb{E}_{\xi_h}\big[\boldsymbol{b}(s)\big]$, we see that for any $x, y \in \mathbb{X}$, we have

$$\begin{aligned} \frac{\partial \bar{\theta}(P_h)}{\partial h_x(y)} &= \big(I - \bar{L}^{(h)}\big)^{-1} \frac{\partial \bar{L}^{(h)}}{\partial h_x(y)} \big(I - \bar{L}^{(h)}\big)^{-1} \bar{b}^{(h)} + \big(I - \bar{L}^{(h)}\big)^{-1} \frac{\partial \bar{b}^{(h)}}{\partial h_x(y)} \\ &= \big(I - \bar{L}^{(h)}\big)^{-1} \big(\big(\frac{\partial}{\partial h_x(y)} \mathbb{E}_{\xi_h}\big[\boldsymbol{L}(s)\big]\big) \cdot \bar{\theta}(P_h) + \frac{\partial}{\partial h_x(y)} \mathbb{E}_{\xi_h}\big[\boldsymbol{b}(s)\big]\big). \end{aligned}$$

Following the formula (86), we conclude that

$$\begin{aligned} \frac{\partial \bar{\theta}(P_h)}{\partial h_x(y)} &= \xi_h(x) P_h(x,y) \big(I - \bar{L}^{(h)}\big)^{-1} \big[\mathcal{A}_h\big(\boldsymbol{L}(y)\bar{\theta}(P_h) + \boldsymbol{b}(y)\big)\big] \\ &\quad - \xi_h(x) P_h(x,y) \sum_{z\in\mathbb{X}} P_h(x,z) \big(I - \bar{L}^{(h)}\big)^{-1} \big[\mathcal{A}_h\big(\boldsymbol{L}(z)\bar{\theta}(P_h) + \boldsymbol{b}(z)\big)\big]. \end{aligned} \tag{87}$$

Recall the shorthand notation from before, where for each $s \in \mathbb{X}$, we defined

$$\boldsymbol{g}_h(s) = \big(I - \bar{L}^{(h)}\big)^{-1}\big[\mathcal{A}_h\big(\boldsymbol{L}(s)\bar{\theta}(P_h) + \boldsymbol{b}(s)\big)\big].$$

Given $w \in \mathbb{R}^d$, if we parameterize the perturbation as $h = Qw$, the chain rule yields

$$\begin{aligned}
\nabla_w \bar{\theta}(P_h) &= Q^\top \cdot \nabla_h \bar{\theta}(P_h) \\
&= \sum_{x\in\mathbb{X}} \xi_h(x)\Big(\sum_{y\in\mathbb{X}} P_h(x,y)\boldsymbol{g}(y)q_x(y)^\top - \Big(\sum_{y\in\mathbb{X}} P_h(x,y)\boldsymbol{g}(y)\Big)\Big(\sum_{y\in\mathbb{X}} P_h(x,y)\boldsymbol{g}_h(y)q_x(y)\Big)^\top\Big) \\
&= \mathbb{E}_{X\sim\xi_h}\big[\operatorname{cov}_{Y\sim P_h(X,\cdot)}\big(\boldsymbol{g}_h(Y) - \mathcal{P}_h\boldsymbol{g}_h(X), q_X(Y) \mid X\big)\big],
\end{aligned}$$

as claimed. □

7.3 Proof of Lemma 13

The following technical lemma is used throughout the proof, and proved in Appendix D.1.

Lemma 16. *Given a perturbation vector w satisfying $\|w\|_2 \le \frac{1-\kappa}{2ct_{\mathrm{mix}}\sigma_L\sqrt{d\cdot\|\Lambda\|_{op}}\log d}$, for $h = Qw$, the matrix $I - \bar{L}^{(h)}$ is invertible, with $\big\|\big(I-\bar{L}^{(h)}\big)^{-1}\big\|_{op} \le \frac{2}{1-\kappa}$.*

Before proceeding with the proof, we note two direct consequences of Lemma 17 from Appendix D.2. First, by taking $f(x) := \langle e_j, \boldsymbol{L}(x)u\rangle$ and $f(x) := \langle e_j, \boldsymbol{b}(x)\rangle$, applying the tail assumption 2 and the boundedness assumption 4, we have the following second moment estimate for any $u \in \mathbb{S}^{d-1}$ and $j \in [d]$:

$$\mathbb{E}_{X\sim\xi_h}\big[\langle e_j, \mathcal{A}_h\boldsymbol{L}(X)u\rangle^2\big] \le ct_{\mathrm{mix}}^2\sigma_L^2\log^2 d, \quad \text{and} \quad \mathbb{E}_{X\sim\xi_h}\big[\langle e_j, \mathcal{A}_h\boldsymbol{b}(X)\rangle^2\big] \le ct_{\mathrm{mix}}^2\sigma_b^2\log^2 d. \tag{88}$$

Second, by taking $f_j(x) := \langle e_j, \boldsymbol{L}(x)\bar{\theta}(P_h) + \boldsymbol{b}(x)\rangle$, for any integer $p \ge 1$ and $K > 0$, Markov's inequality yields the bound

$$\mathbb{P}_{X\sim\xi_h}\Big[\mathcal{A}_h f_j(X) \ge K\Big] \le K^{-2p}\mathbb{E}_{X\sim\xi_h}\big[\mathcal{A}_h f_j(X)^{2p}\big] \le \Big(\tfrac{cp^2 t_{\mathrm{mix}}(\sigma_L\|\bar{\theta}\|_2+\sigma_b)\log d}{K}\Big)^{2p}.$$

By taking $K = 2cp^2 t_{\mathrm{mix}}(\sigma_L\|\bar{\theta}\|_2 + \sigma_b)\log d$ and $p = -2\log\min_{x\in\mathbb{X}}\xi_0(x)$, we find that

$$\mathbb{P}_{X\sim\xi_h}\Big[\mathcal{A}_h f_j(X) \ge 8ct_{\mathrm{mix}}(\sigma_L\|\bar{\theta}\|_2 + \sigma_b)\log^3\big(\tfrac{d}{\min_{x\in\mathbb{X}}\xi_0(x)}\big)\Big] < \frac{1}{2}\min_{x\in\mathbb{X}}\xi_0(x) \le \min_{x\in\mathbb{X}}\xi_h(x),$$

Since ξ_h is a discrete measure, this high-probability bound implies a deterministic bound

$$\mathcal{A}_h f_j(x) \le 8ct_{\mathrm{mix}}(\sigma_L\|\bar{\theta}\|_2 + \sigma_b)\log^3\big(\tfrac{d}{\min_{x'\in\mathbb{X}}\xi_0(x')}\big) \qquad \text{for all } x \in \mathbb{X}.$$

Combining the estimates for all j coordinates yields the bound

$$\max_{x\in\mathbb{X}}\|\boldsymbol{g}_h(x)\|_2 \le \tfrac{1}{1-\kappa}\max_{x\in\mathbb{X}}\|\mathcal{A}_h\big[f_j(x)\big]_{j\in[d]}\|_2 \le \tfrac{ct_{\mathrm{mix}}(\sigma_L\|\bar{\theta}\|_2+\sigma_b)\sqrt{d}}{1-\kappa}\log^3\big(\tfrac{d}{\min_{x\in\mathbb{X}}\xi_0(x)}\big). \tag{89}$$

Given the two lemmas and facts derived above, we now proceed to the proof of Lemma 13. Taking derivatives on both sides of equation (76), we obtain

$$\begin{aligned}
\nabla_w\boldsymbol{g}_h(z) &= \big(I_d - \bar{L}^{(h)}\big)^{-1}\cdot\mathcal{A}_h\boldsymbol{L}(z)\cdot\nabla_w\bar{\theta}(P_h) \\
&\quad + \big(I_d - \bar{L}^{(h)}\big)^{-1}\cdot\big(\nabla_w\mathcal{A}_h\big)\big(\boldsymbol{L}(z)\bar{\theta}(P_h) + \boldsymbol{b}(z)\big) \\
&\quad - \big(I_d - \bar{L}^{(h)}\big)^{-1}\nabla_w\big(\bar{L}^{(h)}\big)\big(I_d - \bar{L}^{(h)}\big)^{-1}\big(\mathcal{A}_h\boldsymbol{L}(z)\cdot\bar{\theta}(P_h) + \mathcal{A}_h\boldsymbol{b}(z)\big) \\
&=: J_1(h,z) + J_2(h,z) + J_3(h,z).
\end{aligned}$$

We then have the integral relation

$$\boldsymbol{g}_h(z) - \boldsymbol{g}_0(z) = \int_0^1 \nabla_w \boldsymbol{g}_{sh}(z) \cdot w \; ds = \int_0^1 \big(J_1(sh, z) + J_2(sh, z) + J_3(sh, z)\big) \cdot w \; ds.$$

It thus suffices to prove individual upper bounds on the terms $J_1(sh, z) \cdot w$, $J_2(sh, z) \cdot w$ and $J_3(sh, z) \cdot w$.

Bounds on the term $J_1(sh, z) \cdot w$: Invoking Lemma 12, we have

$$\nabla_w \bar{\theta}(P_h) = \mathbb{E}_{X \sim \xi_h, Y \sim P_h(X, \cdot)} \big[\big(\boldsymbol{g}_h(Y) - \mathcal{P}_h \boldsymbol{g}_h(X)\big) \big(\boldsymbol{g}_0(Y) - \mathcal{P}_0 \boldsymbol{g}_0(X)\big)^\top \big].$$

Consequently,

$$\begin{aligned}
&\|\nabla_w \bar{\theta}(P_h) w\|_2 \\
&\leq \|\operatorname{cov}_{X \sim \xi_h, Y \sim P_h(X, \cdot)} \big(\boldsymbol{g}_0(Y) - \mathcal{P}_0 \boldsymbol{g}_0(X)\big) \cdot w\|_2 \\
&\quad + \|\mathbb{E}_{X \sim \xi_h} \big[\operatorname{cov}_{Y \sim P_h(X, \cdot)} \big(\boldsymbol{g}_h(Y) - \boldsymbol{g}_0(Y) - \mathcal{P}_0 \boldsymbol{g}_0(X) + \mathcal{P}_0 \boldsymbol{g}_0(X), \boldsymbol{g}_0(Y) - \mathcal{P}_0 \boldsymbol{g}_0(X)\big) \big] \cdot w\|_2.
\end{aligned}$$

For perturbation matrix h satisfying the condition $\max_{x \in \mathbb{X}} \|h_x\|_\infty \leq \frac{1}{128 t_{\mathrm{mix}}}$, Lemma 11 implies the sandwich relations

$$\tfrac{1}{2} \xi_0 \preceq \xi_h \preceq \tfrac{3}{2} \xi_0, \quad \text{and} \quad \tfrac{1}{2} P_0(x) \preceq P_h(x, \cdot) \preceq \tfrac{3}{2} P_0(x), \quad \text{for all } x \in \mathbb{X}.$$

For the first term in above decomposition, we have

$$\begin{aligned}
\|\operatorname{cov}_{X \sim \xi_h, Y \sim P_h(X, \cdot)} \big(\boldsymbol{g}_0(Y) - \mathcal{P}_0 \boldsymbol{g}_0(X)\big) \cdot w\|_2 &\leq \tfrac{3}{2} \|\operatorname{cov}_{X \sim \xi_0, Y \sim P_0(X, \cdot)} \big(\boldsymbol{g}_0(Y) - \mathcal{P}_0 \boldsymbol{g}_0(X)\big) \cdot w\|_2 \\
&= \tfrac{3}{2} \|\Lambda w\|_2 \leq \tfrac{3}{2} \sqrt{\operatorname{trace}(\Lambda)/n},
\end{aligned}$$

where the last inequality is due to the bound (77c).

For the second term in the decomposition, we have

$$\begin{aligned}
&\|\mathbb{E}_{X \sim \xi_h} \big[\operatorname{cov}_{Y \sim P_h(X, \cdot)} \big(\boldsymbol{g}_h(Y) - \boldsymbol{g}_0(Y) - \mathcal{P}_0 \boldsymbol{g}_0(X) + \mathcal{P}_0 \boldsymbol{g}_0(X), \boldsymbol{g}_0(Y) - \mathcal{P}_0 \boldsymbol{g}_0(X)\big) \big] \cdot w\|_2 \\
&= \sup_{v \in \mathbb{S}^{d-1}} \mathbb{E}_{X \sim \xi_h, Y \sim P_h(X, \cdot)} \big[\big(\boldsymbol{g}_h(Y) - \boldsymbol{g}_0(Y) - \mathcal{P}_0 \boldsymbol{g}_0(X) + \mathcal{P}_0 \boldsymbol{g}_0(X)\big)^\top v \cdot \big(\boldsymbol{g}_0(Y) - \mathcal{P}_0 \boldsymbol{g}_0(X)\big)^\top w \big] \\
&\leq \sup_{v \in \mathbb{S}^{d-1}} \sqrt{\mathbb{E}_{X \sim \xi_h} \big[\langle \big(\boldsymbol{g}_h(X) - \boldsymbol{g}_0(X)\big), v \rangle^2 \big]} \cdot \sqrt{\mathbb{E}_{X \sim \xi_h, Y \sim P_h(X, \cdot)} \big[\big(\big(\boldsymbol{g}_0(Y) - \mathcal{P}_0 \boldsymbol{g}_0(X)\big)^\top w \big)^2 \big]} \\
&\leq \tfrac{3}{2} \sqrt{w^\top \Lambda w} \sqrt{\mathbb{E}_{X \sim \xi_h} \|\boldsymbol{g}_h(X) - \boldsymbol{g}_0(X)\|_2^2}.
\end{aligned}$$

By equation (77b), on the support of the prior density, we have the bound $w^\top \Lambda w = n^{-1} \psi^\top D^{-1/2} U^\top \Lambda U D^{-1/2} \psi \leq \frac{d}{n}$. Consequently, we have the upper bound

$$\|\nabla_w \bar{\theta}(P_h) w\|_2 \leq \tfrac{3}{2} \sqrt{\tfrac{\operatorname{trace}(\Lambda)}{n}} + \tfrac{3}{2} \cdot \sqrt{\tfrac{d}{n} \cdot \mathbb{E}_{X \sim \xi_h} \|\boldsymbol{g}_h(X) - \boldsymbol{g}_0(X)\|_2^2}. \tag{90}$$

Collecting the bounds above and invoking equation (88) and Lemma 16, we obtain the following bound on the desired term:

$$\begin{aligned}
&\mathbb{E}_{Y \sim \xi_h} \big[\|J_1(\ell h, Y) w\|_2^2 \big] \\
&\leq \big\| \big(I_d - \bar{L}^{(\ell h)}\big)^{-1} \big\|_{\mathrm{op}}^2 \cdot \mathbb{E}_{Y \sim \xi_h} \big[\|\mathcal{A}_{\ell h} \boldsymbol{L}(Y) \cdot \nabla_w \bar{\theta}(P_{\ell h}) w\|_2^2 \big] \\
&\leq \tfrac{4}{(1-\kappa)^2} \cdot \tfrac{3}{2} \mathbb{E}_{Y \sim \xi_{\ell h}} \big[\|\mathcal{A}_{\ell h} \boldsymbol{L}(Y) \cdot \nabla_w \bar{\theta}(P_{\ell h}) w\|_2^2 \big] \\
&\leq \tfrac{6}{(1-\kappa)^2} \cdot c t_{\mathrm{mix}}^2 \sigma_L^2 d \log^2 d \cdot \|\nabla_w \bar{\theta}(P_{\ell h}) w\|_2^2 \\
&\leq \tfrac{c t_{\mathrm{mix}}^2 \sigma_L^2 d \log^2 d}{(1-\kappa)^2} \cdot \tfrac{\operatorname{trace}(\Lambda)}{n} + \tfrac{c t_{\mathrm{mix}}^2 \sigma_L^2 d^2 \log^2 d}{(1-\kappa)^2 n} \sup_{0 \leq \ell \leq 1} \mathbb{E}_{X \sim \xi_{\ell h}} \|\boldsymbol{g}_{\ell h}(X) - \boldsymbol{g}_0(X)\|_2^2.
\end{aligned}$$

Bounds on the term $J_2(sh, z) \cdot w$: For any function $\mathbb{X} \to \mathbb{R}^d$ and $x, y \in \mathbb{X}$, we note that

$$\begin{aligned}\tfrac{\partial}{\partial h_x(y)}\mathcal{A}_h f &= -(I - \mathcal{P}_h)^{-1}|_{\mathbb{H}_h} \cdot \tfrac{\partial \mathcal{P}_h}{\partial h_x(y)} \cdot (I - \mathcal{P}_h)^{-1}|_{\mathbb{H}_h} f \\ &= -\mathcal{A}_h \cdot \big[\mathbf{1}_{s=x} P_h(x, y) \cdot \big(\mathbf{1}_{s'=y} - P_h(x, s')\big)\big]_{s,s' \in \mathbb{X}} \cdot \mathcal{A} f \\ &= -\mathcal{A}_h \cdot \big[\mathbf{1}_{s=x} P_h(x, y) \cdot \big(\mathcal{A}_h f(y) - \sum_{s'} P_h(x, s') \mathcal{A}_h f(s')\big)\big]_{s \in \mathbb{X}}.\end{aligned}$$

We can then derive the formula for derivative with respect to the parameter w, as

$$\begin{aligned}&\big(\nabla_w \mathcal{A}_h\big) f(z) = \sum_{x,y \in \mathbb{X}} \big(\frac{\partial}{\partial h_x(y)} \mathcal{A}_h f(z)\big) \cdot q_x(y)^\top \\ &= -\sum_{x,y \in \mathbb{X}} P_h(x, y) \mathcal{A}_h \mathbf{1}_x(z) \cdot \big(\mathcal{A}_h f(y) - \mathcal{P}_h \mathcal{A}_h f(x)\big) \cdot \big(\boldsymbol{g}_0(y) - \mathcal{P}_0 \boldsymbol{g}_0(x)\big)^\top \\ &= -\sum_{x,y \in \mathbb{X}} \sum_{t=0}^{\infty} \big(P_h^t(z, x) - \xi_h(x)\big) P_h(x, y) \big(\mathcal{A}_h f(y) - \mathcal{P}_h \mathcal{A}_h f(x)\big) \big(\boldsymbol{g}_0(y) - \mathcal{P}_0 \boldsymbol{g}_0(x)\big)^\top.\end{aligned}$$

Substituting $f(z) = \boldsymbol{L}(z)\bar{\theta}(P_h) + \boldsymbol{b}(z)$, we note that $\mathcal{A}_h f = \boldsymbol{g}_h$, and consequently,

$$\begin{aligned}&\big(\nabla_w \mathcal{A}_h\big) \big(\boldsymbol{L}(z)\bar{\theta}(P_h) + \boldsymbol{b}(z)\big) \\ &= \sum_{t=0}^{\infty} \big(\mathbb{E}_{X \sim P_h^t(z,\cdot), Y \sim P_h(X,\cdot)} \big[\big(\boldsymbol{g}_h(Y) - \mathcal{P}_h \boldsymbol{g}_h(X)\big) \big(\boldsymbol{g}_0(Y) - \mathcal{P}_0 \boldsymbol{g}_0(X)\big)^\top\big] \\ &\qquad - \mathbb{E}_{X \sim \xi_h, Y \sim P_h(X,\cdot)} \big[\big(\boldsymbol{g}_h(Y) - \mathcal{P}_h \boldsymbol{g}_h(X)\big) \big(\boldsymbol{g}_0(Y) - \mathcal{P}_0 \boldsymbol{g}_0(X)\big)^\top\big]\big) \\ &=: \sum_{t=0}^{\infty} D_t(z).\end{aligned}$$

Next, we estimate the difference term above in two different ways, depending on the value of t. On the one hand, note that

$$\begin{aligned}&\mathbb{E}_{Z \sim \xi_h} \|\mathbb{E}_{X \sim P_h^t(Z,\cdot), Y \sim P_h(X,\cdot)} \big[\big(\boldsymbol{g}_h(Y) - \mathcal{P}_h \boldsymbol{g}_h(X)\big) \big(\boldsymbol{g}_0(Y) - \mathcal{P}_0 \boldsymbol{g}_0(X)\big)^\top\big] w\|_2^2 \\ &\le \sup_{x,y \in \mathbb{X}} \|\boldsymbol{g}_h(y) - \mathcal{P}_h \boldsymbol{g}_h(x)\|_2^2 \cdot \mathbb{E}_{X \sim \xi_h, Y \sim P_h(X,\cdot)} \big[\langle w,\, \boldsymbol{g}_0(Y) - \mathcal{P}_0 \boldsymbol{g}_0(X)\rangle^2\big] \\ &\le 4 \sup_{x \in \mathbb{X}} \|\boldsymbol{g}_h(x)\|_2^2 \cdot \mathbb{E}_{X \sim \xi_h, Y \sim P_h(X,\cdot)} \big[\langle w,\, \boldsymbol{g}_0(Y) - \mathcal{P}_0 \boldsymbol{g}_0(X)\rangle^2\big],\end{aligned}$$

where the bound for the factor $\sup_{x \in \mathbb{X}} \|\boldsymbol{g}_h(x)\|_2^2$ follows from equation (89). For the latter term in the display above, we note that

$$\begin{aligned}\mathbb{E}_{X \sim \xi_h, Y \sim P_h(X,\cdot)} \big[\langle w,\, \boldsymbol{g}_0(Y) - \mathcal{P}_0 \boldsymbol{g}_0(X)\rangle^2\big] &\le 2\mathbb{E}_{X \sim \xi_0, Y \sim P_0(X,\cdot)} \big[\langle w,\, \boldsymbol{g}_0(Y) - \mathcal{P}_0 \boldsymbol{g}_0(X)\rangle^2\big] \\ &\le 2 w^\top \Lambda w = \tfrac{2d}{n}.\end{aligned}$$

Putting together the pieces yields the first estimate

$$\mathbb{E}_{Z \sim \xi_h} \big[\|D_t(Z) w\|_2^2\big] \le \tfrac{c t_{\mathrm{mix}}^2 \bar{\sigma}^2 d^2}{(1-\kappa)^2 n} \log^6 \big(\tfrac{d}{\min_{x \in \mathbb{X}} \xi_0(x)}\big).$$

On the other hand, given $z \in \mathbb{X}$ and the Markov chain $(s_t)_{t \ge 0}$ starting from $s_0 = z$, for any $t > 0$, there exists a random state $\widetilde{s}_t$ such that $\widetilde{s}_t \sim \xi_h$, and we have $\mathbb{P}\big(\widetilde{s}_t \neq s_t\big) \le 2^{\lfloor \frac{t}{t_{\mathrm{mix}}} \rfloor}$.

Define a random variable $\widetilde{s}_{t+1}$ by setting $\widetilde{s}_{t+1} = s_{t+1}$ whenever $s_t = \widetilde{s}_t$, and drawing $\widetilde{s}_{t+1} \sim P(\widetilde{s}_t, \cdot)$ otherwise. From this construction, we have

$$\begin{aligned}
\|D_t(z)w\|_2 &\le \sup_{u\in\mathbb{S}^{d-1}} \Big\{ \mathbb{E}\big[u^\top \big(\boldsymbol{g}_h(s_{t+1}) - \mathcal{P}_h \boldsymbol{g}_h(s_t)\big) \cdot w^\top \big(\boldsymbol{g}_0(s_{t+1}) - \mathcal{P}_0 \boldsymbol{g}_0(s_t)\big) \mid z\big] \\
&\qquad\qquad - \mathbb{E}\big[u^\top \big(\boldsymbol{g}_h(\widetilde{s}_{t+1}) - \mathcal{P}_h \boldsymbol{g}_h(\widetilde{s}_t)\big) \cdot w^\top \big(\boldsymbol{g}_0(\widetilde{s}_{t+1}) - \mathcal{P}_0 \boldsymbol{g}_0(\widetilde{s}_t)\big) \mid z\big] \Big\} \\
&\le \sup_{u\in\mathbb{S}^{d-1}} \mathbb{E}\big[u^\top \big(\boldsymbol{g}_h(s_{t+1}) - \mathcal{P}_h \boldsymbol{g}_h(s_t)\big) \cdot w^\top \big(\boldsymbol{g}_0(s_{t+1}) - \mathcal{P}_0 \boldsymbol{g}_0(s_t)\big) \mathbf{1}_{s_t \neq \widetilde{s}_t} \mid z\big] \\
&\qquad + \sup_{u\in\mathbb{S}^{d-1}} \mathbb{E}\big[u^\top \big(\boldsymbol{g}_h(\widetilde{s}_{t+1}) - \mathcal{P}_h \boldsymbol{g}_h(\widetilde{s}_t)\big) \cdot w^\top \big(\boldsymbol{g}_0(\widetilde{s}_{t+1}) - \mathcal{P}_0 \boldsymbol{g}_0(\widetilde{s}_t)\big) \mathbf{1}_{s_t \neq \widetilde{s}_t} \mid z\big].
\end{aligned}$$

Applying the Cauchy–Schwarz inequality twice yields

$$\begin{aligned}
&\mathbb{E}_{Z\sim\xi_h}\big[\|D_t(Z)w\|_2^2\big] \\
&\le \mathbb{E}\big[\|\boldsymbol{g}_h(s_{t+1}) - \mathcal{P}_h \boldsymbol{g}_h(s_t)\|_2^4\big]^{1/2} \cdot \mathbb{E}\big[w^\top \big(\boldsymbol{g}_0(s_{t+1}) - \mathcal{P}_0 \boldsymbol{g}_0(s_t)\big)^8\big]^{1/4} \cdot \mathbb{E}\big[\mathbf{1}_{s_t \neq \widetilde{s}_t}\big]^{1/4} \\
&\quad + \mathbb{E}\big[\|\boldsymbol{g}_h(\widetilde{s}_{t+1}) - \mathcal{P}_h \boldsymbol{g}_h(s_t)\|_2^4\big]^{1/2} \cdot \mathbb{E}\big[w^\top \big(\boldsymbol{g}_0(\widetilde{s}_{t+1}) - \mathcal{P}_0 \boldsymbol{g}_0(\widetilde{s}_t)\big)^8\big]^{1/4} \cdot \mathbb{E}\big[\mathbf{1}_{s_t \neq \widetilde{s}_t}\big]^{1/4} \\
&\le \frac{c t_{\mathrm{mix}}^4}{(1-\kappa)^4} \bar{\sigma}^4 d \|w\|_2^2 \cdot \log^6 d \cdot 2^{1 - \frac{t}{4 t_{\mathrm{mix}}}},
\end{aligned}$$

corresponding to the second estimate.

Finally, setting $\tau = c t_{\mathrm{mix}} \log \frac{t_{\mathrm{mix}} d}{1-\kappa}$ yields

$$\begin{aligned}
\mathbb{E}_{Z\sim\xi_h}\big[\|\sum_{t=0}^{\infty} D_t(Z)w\|_2^2\big] &\le \big(\sum_{t=0}^{\infty} e^{-\frac{t}{\tau}}\big) \cdot \big(\sum_{t=0}^{\infty} e^{\frac{t}{\tau}} \mathbb{E}_{Z\sim\xi_h}\big[\|D_t(Z)w\|_2^2\big]\big) \\
&\le \tfrac{c t_{\mathrm{mix}}^4 \bar{\sigma}^2 d^2}{(1-\kappa)^2 n} \log^6\big(\tfrac{d}{\min_{x\in\mathbb{X}} \xi_0(x)}\big),
\end{aligned}$$

so that

$$\mathbb{E}_{Z\sim\xi_h}\big[\|J_2(\ell h, Z)w\|_2^2\big] \le \tfrac{c t_{\mathrm{mix}}^4 \bar{\sigma}^2 d^2}{(1-\kappa)^4 n} \log^6\big(\tfrac{d}{\min_{x\in\mathbb{X}} \xi_0(x)}\big).$$

Bounds on the term $J_3(sh, z) \cdot w$: By equation (86), for any vector $u \in \mathbb{S}^{d-1}$, we have

$$\nabla_w\big(\bar{L}^{(h)} u\big) = \sum_{x,y\in\mathbb{X}} \xi_h(x) P_h(x,y) \big(\mathcal{A}_h \bar{L}^{(h)}(y) - \sum_{z\in\mathbb{X}} P_h(x,z) \mathcal{A}_h \bar{L}^{(h)}(z)\big) u \cdot q_x(y)^\top.$$

For any $z \in \mathbb{X}$, we obtain

$$\begin{aligned}
&\|\nabla_w\big(\bar{L}^{(h)}\big)\boldsymbol{g}_h(z)w\|_2 \\
&= \sup_{u\in\mathbb{S}^{d-1}} \mathbb{E}_{X\sim\xi_h, Y\sim P_h(X,\cdot)}\big[u^\top \big(\mathcal{A}_h \bar{L}^{(h)}(Y) - \mathcal{P}_h \mathcal{A}_h \bar{L}^{(h)}(X)\big) \boldsymbol{g}_h(z) q_X(Y)^\top w\big] \\
&\le \sup_{u\in\mathbb{S}^{d-1}} \sqrt{\mathbb{E}\big(u^\top \big(\mathcal{A}_h \bar{L}^{(h)}(Y) - \mathcal{P}_h \mathcal{A}_h \bar{L}^{(h)}(X)\big) \boldsymbol{g}_h(z)\big)^2} \cdot \sqrt{\mathbb{E}\big[\big(q_X(Y)^\top w\big)^2\big]} \\
&\le c t_{\mathrm{mix}} \sigma_L \|\boldsymbol{g}_h(z)\|_2 \log d \cdot \sqrt{\tfrac{d}{n}},
\end{aligned}$$

where the final inequality is due to equation (88). Combining with Lemma 16, we have the bound

$$\begin{aligned}
\mathbb{E}_{Z\sim\xi_h}\big[\|J_3(\ell h, Z)w\|_2^2\big] &\le \tfrac{c d^2}{(1-\kappa)^2 n} \cdot t_{\mathrm{mix}}^2 \sigma_L^2 \log^2 d \cdot \mathbb{E}_{Z\sim\xi_h}\big[\|\boldsymbol{g}_h(Z)\|_2^2\big] \\
&\le \tfrac{c \sigma_L^2 \bar{\sigma}^2 t_{\mathrm{mix}}^4 d^2}{(1-\kappa)^4 n} \log^2 d.
\end{aligned}$$

Finishing the proof. Collecting the bounds for J_1, J_2 and J_3 and for $n \geq \frac{ct_{\mathrm{mix}}^2 \sigma_L^2 d^2 \log^2 d}{(1-\kappa)^2}$, we have

$$\sup_{0\leq \ell \leq 1} \mathbb{E}_{Z\sim\xi_h}\big[\|\boldsymbol{g}_{\ell h}(Z) - \boldsymbol{g}_0(Z)\|_2^2\big] \leq \tfrac{c(1+\sigma_L^2)\bar{\sigma}^2 t_{\mathrm{mix}}^4 d^2}{(1-\kappa)^4 n} \log^6\big(\tfrac{d}{\min_x \xi_0(x)}\big) + \tfrac{1}{2} \sup_{0\leq \ell \leq 1} \mathbb{E}_{Z\sim\xi_h}\big[\|\boldsymbol{g}_{\ell h}(Z) - \boldsymbol{g}_0(Z)\|_2^2\big],$$

which completes the proof of the first claim of the lemma.

For the second claim, we combine the first claim with equation (90) and obtain

$$\|\nabla_w \bar{\theta}(P_h) w\|_2 \leq \tfrac{3}{2}\sqrt{\tfrac{\mathrm{trace}(\Lambda)}{n}} + \sqrt{\tfrac{c(1+\sigma_L^2)\bar{\sigma}^2 t_{\mathrm{mix}}^4 d^3}{(1-\kappa)^4 n^2} \log^6\big(\tfrac{d}{\min_x \xi_0(x)}\big)}.$$

Taking the integral yields

$$\|\bar{\theta}(P_h) - \bar{\theta}(P_0)\|_2 \leq \int_0^1 \|\nabla_w \bar{\theta}(P_{\ell h}) w\|_2 d\ell \leq \tfrac{3}{2}\sqrt{\tfrac{\mathrm{trace}(\Lambda)}{n}} + \sqrt{\tfrac{c(1+\sigma_L^2)\bar{\sigma}^2 t_{\mathrm{mix}}^4 d^3}{(1-\kappa)^4 n^2} \log^6\big(\tfrac{d}{\min_x \xi_0(x)}\big)},$$

which proves the second claim.

7.4 Proof of Lemma 14

We first compute the Fisher information with respect to the perturbation vector h, and then transform this via chain rule into a formula that holds with respect to the parameter w. We are interested in the matrix $I^{(n)}(h) := \mathbb{E}_h\big[\nabla_h \log \mathbb{P}_h(s_0^n) \nabla_h \log \mathbb{P}_h(s_0^n)^\top\big]$. When the Markov chain P_h is run under the initial distribution ξ_0, the joint distribution of the observed trajectory $(s_t)_{t=0}^n$ can be factorized as $\mathbb{P}_h\big(s_0, s_1, \cdots, s_n\big) = \xi_0(s_0) \cdot \prod_{t=1}^n P_h(s_{t-1}, s_t)$.

Let us now study the Fisher information matrix. For any pair $x, y \in \mathbb{X}$ with $P(x,y) > 0$, performing some algebra yields the expression

$$\tfrac{\partial}{\partial h_x(y)} \log \mathbb{P}_h\big(s_0, s_1, \cdots, s_n\big) = \sum_{t=1}^n \mathbf{1}_{s_{t-1}=x}\big(\mathbf{1}_{s_t=y} - P_h(x,y)\big).$$

Consider the natural filtration $\mathcal{F}_t := \sigma(s_0, s_1, \cdots, s_t)$. Note that under the transition kernel P_h, we have the identity

$$\mathbb{E}_h\big[\mathbf{1}_{s_{t-1}=x}\big(\mathbf{1}_{s_t=y} - P_h(x,y)\big) \mid \mathcal{F}_{t-1}\big] = \mathbf{1}_{s_{t-1}=x} \cdot \big(\mathbb{E}_h\big[\mathbf{1}_{s_t=y} \mid s_{t-1} = x\big] - P_h(x,y)\big) = 0.$$

Therefore, the process $\{\nabla_h \log \mathbb{P}_h(s_0, s_1, \cdots, s_n)\}_{n\geq 0}$ is a martingale adapted to the filtration $\{\mathcal{F}_t\}_{t\geq 0}$. Its second moment is given by

$$S = \mathbb{E}\big[\nabla_h \log \mathbb{P}_h(s_0^n) \cdot \nabla_h^\top \log \mathbb{P}_h(s_0^n)\big] = \sum_{t=1}^n \mathbb{E}\big[\nabla_h \log P_h(s_{t-1}, s_t) \cdot \nabla_h^\top \log P_h(s_{t-1}, s_t)\big].$$

We find that

$$\begin{aligned} S &= \big[\mathbf{1}_{x_1=x_2} \cdot \sum_{t=1}^n \mathbb{E}\big[\mathbf{1}_{x_1=s_{t-1}} \cdot \big(\mathbf{1}_{s_t=y_1} - P_h(x_1,y_1)\big)\big] \cdot \big(\mathbf{1}_{s_t=y_2} - P_h(x_2,y_2)\big)\big]_{(x_1,y_1),(x_2,y_2)} \\ &= \sum_{t=1}^n \mathrm{diag}\big(\big\{\mathbb{P}_h\big(s_{t-1}=x\big) \cdot P_h(x,y)\big\}_{(x,y)}\big) - \sum_{t=1}^n \big[\mathbb{P}_h(s_{t-1}=x) \cdot P_h(x,y_1) \cdot P_h(x,y_2)\big]_{(x,y_1),(x,y_2)}. \end{aligned}$$

Consequently, the Fisher information matrix is a block diagonal matrix $I^{(n)}(h) = \mathrm{diag}\big(\{I_x^{(n)}(h)\}_{x\in\mathbb{X}}\big)$, where each block matrix $I_x^{(n)}(h) \in \mathbb{R}^{\mathbb{X}\times\mathbb{X}}$ takes the form

$$I_x^{(n)}(h) = \sum_{t=1}^{n} \mathbb{P}_h\big(s_{t-1} = x\big) \cdot \big[\mathrm{diag}\big(\{P_h(x,y)\}_{y\in\mathbb{X}}\big) - \big[P_h(x,y)\big]_{y\in\mathbb{X}}\big[P_h(x,y)\big]_{y\in\mathbb{X}}^{\top}\big].$$

By Lemma 11, for $h_{\max}$ satisfying the inequality $h_{\max}^{-1} \geq ct_{\mathrm{mix}}\big(\log h_{\max}^{-1} + \log(\min \xi_0)^{-1}\big)$ for some constant $c > 0$, we have the bound $\frac{1}{2}\xi_h \preceq \xi_0 \preceq \frac{3}{2}\xi_h$, and hence $\frac{1}{2}P_h^k\xi_h \preceq P_h^k\xi_0 \preceq \frac{3}{2}P_h^k\xi_h$ for each $k = 0, 1, 2, \ldots$. From this sandwiching, we find that

$$\begin{aligned} I_x^{(n)}(h) &\preceq \tfrac{3}{2}\sum_{t=1}^{n} P_h^{t-1}\xi_h(x) \cdot \big[\mathrm{diag}\big(\{P_h(x,y)\}_{y\in\mathbb{X}}\big) - \big[P_h(x,y)\big]_{y\in\mathbb{X}}\big[P_h(x,y)\big]_{y\in\mathbb{X}}^{\top}\big] \\ &= \tfrac{3n}{2}\xi_h(x)\big[\mathrm{diag}\big(\{P_h(x,y)\}_{y\in\mathbb{X}}\big) - \big[P_h(x,y)\big]_{y\in\mathbb{X}}\big[P_h(x,y)\big]_{y\in\mathbb{X}}^{\top}\big]. \end{aligned}$$

Turning to the Fisher information, we compute

$$\begin{aligned} I^{(n)}(w) = Q^{\top}I^{(n)}(h)Q &\preceq \tfrac{3n}{2}\sum_{x\in\mathbb{X}} \xi_h(x)\big(\sum_{y\in\mathbb{X}} P_h(x,y)q_x(y)q_x(y)^{\top} - \big(\sum_{y\in\mathbb{X}} P_h(x,y)q_x(y)\big)\big(\sum_{y\in\mathbb{X}} P_h(x,y)q_x(y)\big)^{\top}\big) \\ &= \tfrac{3n}{2}\mathbb{E}_{X\sim\xi_h}\big[\mathbb{E}_{Y\sim P_h(X,\cdot)}\big[q_X(Y)q_X(Y)^{\top}\big] - \mathbb{E}_{Y\sim P_h(X,\cdot)}\big[q_X(Y)\big]\cdot\mathbb{E}_{Y\sim P_h(X,\cdot)}\big[q_X(Y)\big]^{\top}\big] \\ &= \tfrac{3n}{2}\mathbb{E}_{X\sim\xi_h}\big[\,\mathrm{cov}_{P_h(X,\cdot)}\big(q_X(Y) \mid X\big)\big]. \end{aligned}$$

7.5 Proof of Lemma 15

For each $k \in \mathbb{Z}$, by the definition of the Green function, we note that

$$f(s_k) = \mathcal{A}_0 f(s_k) - \mathbb{E}\big[\mathcal{A}_0 f(s_{k+1}) \mid s_k\big] = \mathcal{A}_0 f(s_k) - \mathcal{P}_0\mathcal{A}_0 f(s_k). \tag{91}$$

By stationarity, we have

$$\sum_{k=-\infty}^{\infty} \mathbb{E}\big[f(s_k)f(s_0)\big] = \mathbb{E}[f^2(s_0)] + 2\sum_{k=1}^{\infty}\mathbb{E}\big[f(s_k)f(s_0)\big] \overset{(i)}{=} -\mathbb{E}[f(s_0)^2] + 2\mathbb{E}\big[f(s_0)\cdot\sum_{k=0}^{\infty}\mathbb{E}\big[f(s_k) \mid s_0\big]\big]$$

where step (i) makes use of the dominated convergence theorem, in particular by noting that $\big|\mathbb{E}\big[f(s_k) \mid s_0\big]\big| \leq \|f\|_{\infty}\cdot 2^{1-k/t_{\mathrm{mix}}}$ from Lemma 4. Consequently, we can write

$$\begin{aligned} \sum_{k=-\infty}^{\infty} \mathbb{E}\big[f(s_k)f(s_0)\big] &= -\mathbb{E}[f^2(s_0)] + 2\mathbb{E}\big[f(s_0)\cdot\mathcal{A}_0 f(s_0)\big] \\ &\overset{(ii)}{=} -\mathbb{E}\big[\big(\mathcal{A}_0 f(s_0) - \mathcal{P}_0\mathcal{A}_0 f(s_0)\big)^2\big] + 2\mathbb{E}\big[\big(\mathcal{A}_0 f(s_0) - \mathcal{P}_0\mathcal{A}_0 f(s_0)\big)\cdot\mathcal{A}_0 f(s_0)\big] \\ &= \mathbb{E}\big[\big(\mathcal{A}_0 f(s_0)\big)^2\big] - \mathbb{E}\big[\big(\mathcal{P}_0\mathcal{A}_0 f(s_0)\big)^2\big], \end{aligned}$$

where step (ii) follows from equation (91).

With $\mathbb{E}$ denoting expectation over $X \sim \xi_0, Y \sim P_0(X,\cdot)$, we have

$$\begin{aligned} \mathbb{E}\big[\big(\mathcal{A}_0 f(Y) - \mathcal{P}_0\mathcal{A}_0 f(X)\big)^2\big] &= \mathbb{E}\big[\big(\mathcal{A}_0 f(s_1) - \mathcal{P}_0\mathcal{A}_0 f(s_0)\big)^2\big] \\ &= \mathbb{E}\big[\big(\mathcal{A}_0 f(s_1)\big)^2\big] + \mathbb{E}\big[\big(\mathcal{P}_0\mathcal{A}_0 f(s_0)\big)^2\big] - 2\mathbb{E}\big[\big(\mathcal{A}_0 f(s_1)\big)\cdot\big(\mathcal{P}_0\mathcal{A}_0 f(s_0)\big)\big] \\ &= \mathbb{E}\big[\big(\mathcal{A}_0 f(s_0)\big)^2\big] + \mathbb{E}\big[\big(\mathcal{P}_0\mathcal{A}_0 f(s_0)\big)^2\big] - 2\mathbb{E}\big[\mathbb{E}\big[\mathcal{A}_0 f(s_1) \mid s_0\big]\cdot\big(\mathcal{P}_0\mathcal{A}_0 f(s_0)\big)\big] \\ &= \mathbb{E}\big[\big(\mathcal{A}_0 f(s_0)\big)^2\big] - \mathbb{E}\big[\big(\mathcal{P}_0\mathcal{A}_0 f(s_0)\big)^2\big], \end{aligned}$$

and combining the pieces completes the proof of this lemma. □

8 Discussion

In this paper, we established sharp instance-optimal guarantees for linear stochastic approximation (SA) procedures based on Markovian data. Under ergodicity along with natural tail conditions, we proved non-asymptotic upper bounds on the squared error of both the last iterate of a standard SA scheme, as well as the Polyak–Ruppert averaged sequence. The results highlight two important aspects: an optimal sample complexity of $O(t_{\mathrm{mix}}d)$ for problems in dimension d with mixing time t_{mix}; and an instance-dependent error upper bound for the averaged estimator with carefully chosen stepsize. Complementary to the upper bound, we also showed a non-asymptotic local minimax lower bound over a small neighborhood of a given Markov chain instance, certifying the statistical optimality of the proposed estimators. Our proof of the upper bounds uses a bootstrapping argument of possibly independent interest.

Throughout the paper, we have introduced novel techniques of analysis and motivated several open questions. In the following, we collect a few interesting future directions:

- **Nonlinear stochastic approximation and controlled dynamics:** Our paper focuses on linear Z-equations where the underlying Markov chain does not involve a control. Though this setting already covers many important examples (as described in Section 2.2), its applicability to practical problems is still relatively restricted. To set up a general framework, one could consider a *controlled Markov chain* $(s_t)_{t\geq 0}$ where the transition is given by $s_{t+1} \sim P(\cdot|s_t, \theta_t)$. For any $\theta \in \mathbb{R}^d$, let ξ_θ be the stationary distribution of the Markov chain $P(\cdot|\cdot, \theta)$ induced by the control θ. Given a non-linear operator $H : \mathbb{X} \times \mathbb{R}^d \to \mathbb{R}^d$, suppose that we wish to solve the equation $\mathbb{E}_{s\sim\xi(\theta)}\big[H(\theta; s)\big] = 0$; see Benveniste et al. [BMP12] for a summary of classical asymptotic theory for such problems. The analysis tools introduced in this paper provide an avenue by which one could obtain optimal sample complexity bounds (especially in terms of dimension dependency) and instance-dependent guarantees for such problems.
- **Online statistical inference:** By carefully choosing the burn-in period, one can show that the Polyak–Ruppert estimator $\widehat{\theta}_n$ is asymptotically normal and locally minimax optimal. In particular, under suitable conditions, the following limiting result holds true (see the paper [For15] for details):

$$\sqrt{n}(\widehat{\theta}_n - \bar{\theta}) \xrightarrow{d} \mathcal{N}\big((I_d - \bar{L})^{-1}(\Sigma^*_{\mathrm{MG}} + \Sigma^*_{\mathrm{Mkv}})(I_d - \bar{L})^{-\top}\big). \tag{92}$$

  In order to construct confidence intervals for the solution $\bar{\theta}$ with streaming data, it suffices to estimate the asymptotic covariance in equation (92). In the i.i.d. setting, online procedures have been developed to estimate such covariances, with non-asymptotic error guarantees [CLTZ20]. The problem becomes more subtle in the Markovian setting, as the matrix Σ^*_{Mkv} involves auto-correlations of the noise process. It is an important open direction to construct online estimators of this matrix to enable inference in a streaming fashion.
- **Model selection and optimal methods for policy evaluation** The policy evaluation problem involves manual choice of two important parameters: the feature vector dimension d and the resolvent parameter λ in TD(λ). In Section 4.1.3 and 4.2, we provide optimal instance-dependent guarantees on both the approximation factor and the estimation error, for a fixed choice of d and λ. An important direction of future research is to select such parameters adaptively based on data, possibly under a streaming computational model. Ideally, we want the risk of such estimator to attain the infimum of the right hand side of equation (39b), over $\lambda \in (0, 1)$ and $d \in \mathbb{N}_+$. A possible candidate approach towards

such a model selection problem is the celebrated Lepskii method for adaptive bandwidth selection [Lep91].

.

Acknowledgments

We gratefully acknowledge the support of the NSF through grants DMS-2023505 and of the ONR through MURI award N000142112431 to PLB. This work was also supported in part by NSF-DMS grant 2015454, NSFFODSI grant 202350, and DOD-ONR Office of Naval Research N00014-21-1-2842 to MJW. This work was also supported by NSF-IIS grant 1909365 to PLB and MJW. AP was supported in part by the National Science Foundation grant CCF-2107455, and is thankful to the Simons Institute for the Theory of Computing for their hospitality when part of this work was performed.

References

[BB96] Steven J Bradtke and Andrew G Barto, *Linear least-squares algorithms for temporal difference learning*, Machine Learning **22** (1996), no. 1-3, 33–57. (Cited on page 10.)

[BCN18] Léon Bottou, Frank E Curtis, and Jorge Nocedal, *Optimization methods for large-scale machine learning*, SIAM Review **60** (2018), no. 2, 223–311. (Cited on page 6.)

[BD09] Peter J Brockwell and Richard A Davis, *Time series: theory and methods*, Springer Science & Business Media, 2009. (Cited on pages 25 and 26.)

[Ber11] Dimitri P Bertsekas, *Temporal difference methods for general projected equations*, IEEE Transactions on Automatic Control **56** (2011), no. 9, 2128–2139. (Cited on page 7.)

[Ber19] ———, *Reinforcement learning and optimal control*, Athena Scientific Belmont, MA, 2019. (Cited on pages 1 and 6.)

[Bil61] Patrick Billingsley, *Statistical methods in Markov chains*, The Annals of Mathematical Statistics (1961), 12–40. (Cited on page 5.)

[BJN+20] Guy Bresler, Prateek Jain, Dheeraj Nagaraj, Praneeth Netrapalli, and Xian Wu, *Least squares regression with Markovian data: Fundamental limits and algorithms*, arXiv preprint arXiv:2006.08916 (2020). (Cited on page 3.)

[BMP12] Albert Benveniste, Michel Métivier, and Pierre Priouret, *Adaptive algorithms and stochastic approximations*, vol. 22, Springer Science & Business Media, 2012. (Cited on pages 1, 2, 6, 15, and 56.)

[Bor09] Vivek S Borkar, *Stochastic approximation: a dynamical systems viewpoint*, vol. 48, Springer, 2009. (Cited on pages 1, 2, and 6.)

[Bor21] ———, *A concentration bound for contractive stochastic approximation*, Systems & Control Letters **153** (2021), 104947. (Cited on page 6.)

[Boy02] Justin A Boyan, *Technical update: Least-squares temporal difference learning*, Machine Learning **49** (2002), no. 2-3, 233–246. (Cited on page 6.)

[BRS18] Jalaj Bhandari, Daniel Russo, and Raghav Singal, *A finite time analysis of temporal difference learning with linear function approximation*, arXiv preprint arXiv:1806.02450 (2018). (Cited on pages 3, 6, and 15.)

[CLTZ20] Xi Chen, Jason D Lee, Xin T Tong, and Yichen Zhang, *Statistical inference for model parameters in stochastic gradient descent*, The Annals of Statistics **48** (2020), no. 1, 251–273. (Cited on page 56.)

[CMSS21] 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*, arXiv preprint arXiv:2102.01567 (2021). (Cited on pages 3 and 6.)

[DDA21] Vianney Debavelaere, Stanley Durrleman, and Stéphanie Allassonnière, *On the convergence of stochastic approximations under a subgeometric ergodic Markov dynamic*, Electronic Journal of Statistics **15** (2021), no. 1, 1583–1609. (Cited on page 6.)

[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** (2020), no. 3, 1348–1382. (Cited on page 6.)

[DDG18] Constantinos Daskalakis, Nishanth Dikkala, and Nick Gravin, *Testing symmetric Markov chains from a single trajectory*, Conference On Learning Theory, PMLR, 2018, pp. 385–409. (Cited on page 5.)

[DMN+21] Alain Durmus, Eric Moulines, Alexey Naumov, Sergey Samsonov, and Hoi-To Wai, *On the stability of random matrix product with Markovian noise: Application to linear stochastic approximation and TD learning*, arXiv preprint arXiv:2102.00185 (2021). (Cited on pages 3 and 15.)

[DNPR20] Thinh T Doan, Lam M Nguyen, Nhan H Pham, and Justin Romberg, *Finite-time analysis of stochastic gradient descent under Markov randomness*, arXiv preprint arXiv:2003.10973 (2020). (Cited on pages 3 and 6.)

[DS94] Peter Dayan and Terrence J Sejnowski, *TD(λ) converges with probability 1*, Machine Learning **14** (1994), no. 3, 295–301. (Cited on page 6.)

[DWW21] Yaqi Duan, Mengdi Wang, and Martin J Wainwright, *Optimal policy evaluation using kernel-based temporal difference methods*, arXiv preprint arXiv:2109.12002 (2021). (Cited on page 22.)

[EDM03] E. Even-Dar and Y. Mansour, *Learning rates for Q-learning*, Journal of Machine Learning Research **5** (2003), 1–25. (Cited on page 6.)

[For15] Gersende Fort, *Central limit theorems for stochastic approximation with controlled Markov chain dynamics*, ESAIM: Probability and Statistics **19** (2015), 60–80. (Cited on pages 6, 14, and 56.)

[GL95] Richard D Gill and Boris Y Levit, *Applications of the van Trees inequality: a Bayesian Cramér-Rao bound*, Bernoulli **1** (1995), no. 1-2, 59–79. (Cited on page 44.)

[GL12] Saeed Ghadimi and Guanghui Lan, *Optimal stochastic approximation algorithms for strongly convex stochastic composite optimization I: A generic algorithmic framework*, SIAM Journal on Optimization **22** (2012), no. 4, 1469–1492. (Cited on page 6.)

[GP17] Sébastien Gadat and Fabien Panloup, *Optimal non-asymptotic bound of the Ruppert–Polyak averaging without strong convexity*, arXiv preprint arXiv:1709.03342 (2017). (Cited on page 6.)

[GW95] Priscilla E Greenwood and Wolfgang Wefelmeyer, *Efficiency of empirical estimators for Markov chains*, The Annals of Statistics (1995), 132–143. (Cited on pages 5 and 17.)

[Ham20] James Douglas Hamilton, *Time series analysis*, Princeton university press, 2020. (Cited on page 1.)

[HKL+19] Daniel Hsu, Aryeh Kontorovich, David A Levin, Yuval Peres, Csaba Szepesvári, and Geoffrey Wolfer, *Mixing time estimation in reversible Markov chains from a single sample path*, The Annals of Applied Probability **29** (2019), no. 4, 2439–2480. (Cited on page 5.)

[JKNN21] Prateek Jain, Suhas S Kowshik, Dheeraj Nagaraj, and Praneeth Netrapalli, *Streaming linear system identification with reverse experience replay*, arXiv preprint arXiv:2103.05896 (2021). (Cited on pages 3 and 26.)

[JZ13] Rie Johnson and Tong Zhang, *Accelerating stochastic gradient descent using predictive variance reduction*, Advances in neural information processing systems **26** (2013), 315–323. (Cited on page 6.)

[KB18] Prasenjit Karmakar and Shalabh Bhatnagar, *Two time-scale stochastic approximation with controlled Markov noise and off-policy temporal-difference learning*, Mathematics of Operations Research **43** (2018), no. 1, 130–151. (Cited on page 6.)

[KC78] Harold J. Kushner and Dean S. Clark, *Stochastic approximation methods for constrained and unconstrained systems*, Applied Mathematical Sciences, vol. 26, Springer-Verlag, New York-Berlin, 1978. MR 499560 (Cited on page 6.)

[KLL20] Georgios Kotsalis, Guanghui Lan, and Tianjiao Li, *Simple and optimal methods for stochastic variational inequalities, II: Markovian noise and policy evaluation in reinforcement learning*, arXiv preprint arXiv:2011.08434 (2020). (Cited on pages 3 and 6.)

[KMMW19] Belhal Karimi, Blazej Miasojedow, Eric Moulines, and Hoi-To Wai, *Non-asymptotic analysis of biased stochastic approximation scheme*, Conference on Learning Theory, PMLR, 2019, pp. 1944–1974. (Cited on page 6.)

[KMN+20] Maxim Kaledin, Eric Moulines, Alexey Naumov, Vladislav Tadic, and Hoi-To Wai, *Finite time analysis of linear two-timescale stochastic approximation with Markovian noise*, Conference on Learning Theory, PMLR, 2020, pp. 2144–2203. (Cited on page 3.)

[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** (2021), no. 4, 1013–1040. (Cited on pages 3 and 6.)

[KT00] Vijay R Konda and John N Tsitsiklis, *Actor-critic algorithms*, Advances in Neural Information Processing Systems, 2000, pp. 1008–1014. (Cited on page 6.)

[Kut97] Yury A Kutoyants, *Efficiency of the empirical distribution for ergodic diffusion*, Bernoulli (1997), 445–456. (Cited on page 5.)

[KXWJ21] Koulik Khamaru, Eric Xia, Martin J Wainwright, and Michael I Jordan, *Instance-optimality in optimal value estimation: Adaptivity via variance-reduced Q-learning*, arXiv preprint arXiv:2106.14352 (2021). (Cited on page 6.)

[KY03] Harold Kushner and G George Yin, *Stochastic approximation and recursive algorithms and applications*, vol. 35, Springer Science & Business Media, 2003. (Cited on pages 1 and 6.)

[LÖ5] H. Lütkepohl, *New introduction to multiple time series analysis*, Springer, New York, 2005. (Cited on page 12.)

[Lep91] OV Lepskii, *On a problem of adaptive estimation in Gaussian white noise*, Theory of Probability & Its Applications **35** (1991), no. 3, 454–466. (Cited on page 57.)

[Lju77a] Lennart Ljung, *Analysis of recursive stochastic algorithms*, IEEE Transactions on Automatic Control **22** (1977), no. 4, 551–575. (Cited on page 6.)

[Lju77b] ———, *On positive real transfer functions and the convergence of some recursive schemes*, IEEE Transactions on Automatic Control **22** (1977), no. 4, 539–551. (Cited on page 6.)

[LLP21] Tianjiao Li, Guanghui Lan, and Ashwin Pananjady, *Accelerated and instance-optimal policy evaluation with linear function approximation*, preprint (2021). (Cited on page 6.)

[LMWJ20] Chris Junchi Li, Wenlong Mou, Martin J Wainwright, and Michael I Jordan, *ROOT-SGD: Sharp nonasymptotics and asymptotic efficiency in a single algorithm*, arXiv preprint arXiv:2008.12690 (2020). (Cited on page 6.)

[LS18] Chandrashekar Lakshminarayanan and Csaba Szepesvári, *Linear stochastic approximation: How far does constant step-size and iterate averaging go?*, International Conference on Artificial Intelligence and Statistics, 2018, pp. 1347–1355. (Cited on page 14.)

[LWC+20] Gen Li, Yuting Wei, Yuejie Chi, Yuantao Gu, and Yuxin Chen, *Breaking the sample size barrier in model-based reinforcement learning with a generative model*, Advances in Neural Information Processing Systems, vol. 33, Curran Associates, Inc., 2020, pp. 12861–12872. (Cited on page 3.)

[LWZ18] Xudong Li, Mengdi Wang, and Anru Zhang, *Estimation of Markov chain via rank-constrained likelihood*, International Conference on Machine Learning, PMLR, 2018, pp. 3033–3042. (Cited on page 5.)

[MB11] Éric Moulines and Francis R Bach, *Non-asymptotic analysis of stochastic approximation algorithms for machine learning*, Advances in Neural Information Processing Systems, 2011, pp. 451–459. (Cited on page 6.)

[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*, Proceedings of Thirty Third Conference on Learning Theory, vol. 125, 2020, pp. 2947–2997. (Cited on pages 6 and 14.)

[MP84] Michel Metivier and Pierre Priouret, *Applications of a Kushner and Clark lemma to general classes of stochastic algorithms*, IEEE Transactions on Information Theory **30** (1984), no. 2, 140–151. (Cited on page 6.)

[MPW20] Wenlong Mou, Ashwin Pananjady, and Martin J Wainwright, *Optimal oracle inequalities for solving projected fixed-point equations*, arXiv preprint arXiv:2012.05299 (2020). (Cited on pages 3, 6, 7, 8, 11, 14, 17, 19, 24, and 46.)

[MS08] Rémi Munos and Csaba Szepesvári, *Finite-time bounds for fitted value iteration*, Journal of Machine Learning Research **9** (2008), no. May, 815–857. (Cited on page 7.)

[Nem01] Arkadi Nemirovski, *Lectures on modern convex optimization*, Society for Industrial and Applied Mathematics (SIAM, Citeseer, 2001. (Cited on page 25.)

[NJLS09] Arkadi Nemirovski, Anatoli Juditsky, Guanghui Lan, and Alexander Shapiro, *Robust stochastic approximation approach to stochastic programming*, SIAM Journal on Optimization **19** (2009), no. 4, 1574–1609. (Cited on page 6.)

[Pen91] Spiridon Penev, *Efficient estimation of the stationary distribution for exponentially ergodic Markov chains*, Journal of Statistical Planning and Inference **27** (1991), no. 1, 105–123. (Cited on page 5.)

[PJ92] Boris T Polyak and Anatoli B Juditsky, *Acceleration of stochastic approximation by averaging*, SIAM Journal on Control and Optimization **30** (1992), no. 4, 838–855. (Cited on pages 2 and 6.)

[PW21] Ashwin Pananjady and Martin J. Wainwright, *Instance-dependent ℓ_∞-bounds for policy evaluation in tabular reinforcement learning*, IEEE Transactions on Information Theory **67** (2021), no. 1, 566–585. (Cited on page 3.)

[RM51] Herbert Robbins and Sutton Monro, *A stochastic approximation method*, The Annals of Mathematical Statistics (1951), 400–407. (Cited on pages 2 and 6.)

[RN94] Gavin A Rummery and Mahesan Niranjan, *On-line Q-learning using connectionist systems*, Tech. report, Cambridge University Engineering Department, 1994. (Cited on page 6.)

[Rup88] David Ruppert, *Efficient estimations from a slowly convergent Robbins-Monro process*, Tech. report, Cornell University Operations Research and Industrial Engineering, 1988. (Cited on pages 2 and 6.)

[Sut88] Richard S Sutton, *Learning to predict by the methods of temporal differences*, Machine Learning **3** (1988), no. 1, 9–44. (Cited on page 6.)

[SWW+18] Aaron Sidford, Mengdi Wang, Xian Wu, Lin F Yang, and Yinyu Ye, *Near-optimal time and sample complexities for solving Markov decision processes with a generative model*, Proceedings of the 32nd International Conference on Neural Information Processing Systems, 2018, pp. 5192–5202. (Cited on page 6.)

[SY19] Rayadurgam Srikant and Lei Ying, *Finite-time error bounds for linear stochastic approximation and TD learning*, Conference on Learning Theory, PMLR, 2019, pp. 2803–2830. (Cited on pages 3 and 15.)

[Sze98] Csaba Szepesvári, *The asymptotic convergence-rate of Q-learning*, Advances in neural information processing systems (1998), 1064–1070. (Cited on page 6.)

[Sze10] Csaba Szepesvári, *Algorithms for reinforcement learning*, Morgan & Claypool Publishers, 2010. (Cited on pages 1 and 6.)

[Tsi94] J. N. Tsitsiklis, *Asynchronous stochastic approximation and Q-learning*, Machine Learning **16** (1994), 185–202. (Cited on page 6.)

[Tsy08] Alexandre B Tsybakov, *Introduction to nonparametric estimation*, Springer Science & Business Media, 2008. (Cited on page 44.)

[TVR97] John N Tsitsiklis and Benjamin Van Roy, *Analysis of temporal-diffference learning with function approximation*, Advances in Neural Information Processing Systems, 1997, pp. 1075–1081. (Cited on pages 2, 6, 7, and 10.)

[TVR99] ———, *Optimal stopping of Markov processes: Hilbert space theory, approximation algorithms, and an application to pricing high-dimensional financial derivatives*, IEEE Transactions on Automatic Control **44** (1999), no. 10, 1840–1851. (Cited on page 6.)

[vdV00] Aad W van der Vaart, *Asymptotic statistics*, vol. 3, Cambridge university press, 2000. (Cited on pages 3 and 17.)

[Wai19a] Martin J Wainwright, *Stochastic approximation with cone-contractive operators: Sharper ℓ_∞-bounds for Q-learning*, arXiv preprint arXiv:1905.06265 (2019). (Cited on page 6.)

[Wai19b] ———, *Variance-reduced Q-learning is minimax optimal*, arXiv preprint arXiv:1906.04697 (2019). (Cited on page 6.)

[WD92] Christopher JCH Watkins and Peter Dayan, *Q-learning*, Machine Learning **8** (1992), no. 3-4, 279–292. (Cited on page 6.)

[WK21] Geoffrey Wolfer and Aryeh Kontorovich, *Statistical estimation of ergodic Markov chain kernel over discrete state space*, Bernoulli **27** (2021), no. 1, 532–553. (Cited on page 5.)

[YB10] Huizhen Yu and Dimitri P Bertsekas, *Error bounds for approximations from projected linear equations*, Mathematics of Operations Research **35** (2010), no. 2, 306–329. (Cited on page 7.)

[YBW17] Fanny Yang, Sivaraman Balakrishnan, and Martin J Wainwright, *Statistical and computational guarantees for the Baum–Welch algorithm*, The Journal of Machine Learning Research **18** (2017), no. 1, 4528–4580. (Cited on page 5.)

A Auxiliary truncation results related to the assumptions

In this section, we present two auxiliary results on the relations between assumptions 2, 3, and 4. These results are based on truncation arguments.

A.1 Assumption 2 (almost) implies assumption 4 under discrete metric

For the discrete metric $\rho(x,y) := \mathbf{1}_{x\neq y}$, the Lipschitz assumption 4 is equivalent to the following uniform upper bounds:

$$\|\boldsymbol{L}_{t+1}(s_t) - \bar{L}\|_{\mathrm{op}} \leq \sigma_L d \quad \text{and} \quad \|\boldsymbol{b}_{t+1}(s_t) - \bar{b}\|_2 \leq \sigma_b\sqrt{d}.$$

The following proposition provides uniform high-probability upper bounds on such quantities based on the moment assumption:

Proposition 5. *Under Assumption 2 with $\bar{p} = +\infty$, there exists a universal constant $c > 0$, such that for any $\delta > 0$, the following bounds hold true uniformly over $t = 1, 2, \cdots, n$, with probability $1-\delta$:*

$$\|\boldsymbol{L}_{t+1}(s_t) - \bar{L}\|_{op} \leq cd \cdot \sigma_L \log\frac{nd}{\delta} \quad \textit{and} \quad \|\boldsymbol{b}_{t+1}(s_t) - \bar{b}\|_2 \leq c\sqrt{d} \cdot \sigma_b \log\frac{nd}{\delta}. \tag{93}$$

We prove this proposition at the end of this section.

When the random observations (L_{t+1}, b_{t+1}) are not almost-surely bounded, but satisfies the moment assumption 2 with $\bar{p} = +\infty$, we can apply our theorems on the event that Eq (93) holds true, and the main theorems hold true conditionally on such an event, with constants (σ_L, σ_b) inflated with a factor $\log(nd/\delta)$.

Proof of Proposition 5: For a given $t \in [n]$, we note that:

$$\|L_{t+1} - \bar{L}\|_{\mathrm{op}}^2 \leq \|L_{t+1} - \bar{L}\|_F^2 = \sum_{j,\ell=1}^{d} \left[e_j^\top (L_{t+1} - \bar{L}) e_\ell\right]^2.$$

For each pair $j, \ell \in [d]$, Assumption 2 implies that:

$$\mathbb{P}\left(\left|e_j^\top (L_{t+1}(s_t) - \bar{L}) e_\ell\right| \geq c\sigma_L \log(nd/\delta)\right) \leq \frac{\delta}{2d^2 n}$$

Taking union bound over all the coordinate pairs (j, ℓ) and substituting into above expansion, we have that:

$$\mathbb{P}\left(\|L_{t+1} - \bar{L}\|_{\mathrm{op}} \geq cd \cdot \sigma_L \log(nd/\delta)\right) \leq \delta/(2n).$$

Similarly, for the vector-valued observations b_{t+1}, we have the following bounds with probability $1 - \delta/n$:

$$\|b_{t+1} - \bar{b}\|_2^2 \leq \sum_{j=1}^{d} \left(e_j^\top (b_{t+1} - \bar{b})\right)^2 \leq c\sigma_b^2 d \cdot \log^2(nd/\delta).$$

Taking union bound over $t = 1, 2, \cdots, n$, we complete the proof of this proposition.

A.2 On the stationary tail and boundedness assumption 3

Note that in many applications, the Markov chain $(s_t)_{t\geq 0}$ lives in an unbounded state space. However, as long as the stationary distribution ξ of P is sufficiently light-tailed, a simple

truncation argument applies, which we illustrate for completeness. Concretely, suppose that there exists a constant $\sigma_\rho > 0$, such that the following bound holds true for any $p \geq 2$:

$$\mathbb{E}_{s\sim\xi}\big[\rho(s, s_0)^p\big] \leq p! \cdot \sigma_\rho^p. \tag{94}$$

Given a stationary Markovian trajectory $\{s_t\}_{t=1}^n$, consider the event

$$\mathscr{E}_{n,\delta} = \big\{\forall t \in [1, n], \rho(s_0, s_t) \leq 2\sigma_\rho \log \tfrac{n}{\delta}\big\}.$$

By the tail assumption (94) and a union bound, it directly follows that $\mathbb{P}\big(\mathscr{E}_{n,\delta}\big) \geq 1 - \delta$. Consider a truncated Markov transition kernel P' defined as

$$\forall x \in \mathbb{X}, Z \subseteq \mathbb{X}, \quad P'(x, Z) := P\big(x, Z \cap \mathbb{B}\big(0, 2\sigma_\rho \log(n/\delta)\big)\big) + P\big(x, \mathbb{B}\big(0, 2\sigma_\rho \log(n/\delta)\big)^c\big)\mathbf{1}_{s_0\in Z}.$$

In words, the Markov chain P' attempts to make the transition from s_t to s_{t+1} according to the transition kernel P'. If the state s_{t+1} lies in the ball $\mathbb{B}\big(0, 2\sigma_\rho \log(n/\delta)\big)^c$, we keep it as is; otherwise, we let the next-step transition be deterministically s_0.

Given a trajectory $\{s'_t\}_{t=1}^n$ of the Markov chain P', there exists a coupling such that

$$\mathbb{P}\big(\{s_t\}_{t=1}^n \neq \{s'_t\}_{t=1}^n\big) \leq \mathbb{P}\big(\mathscr{E}_{n,\delta}^c\big) \leq \delta.$$

One can then proceed by working on the high probability event $\mathscr{E}_{n,\delta}$, where the Markov chain has a effective diameter of $O\big(\sigma_\rho \log \frac{n}{\delta}\big)$.

B Auxiliary results underlying Proposition 1

This appendix is devoted to the proofs of auxiliary lemmas that are used in the proof of Proposition 1.

B.1 Proof of Lemma 4

Throughout the proof, we let $x \in \mathbb{X}$ be an arbitrary but fixed state. Note that any positive integer τ can be represented as $\tau = kt_{\mathrm{mix}} + q$ with $k \in \mathbb{N}_+$ and $0 \leq q \leq t_{\mathrm{mix}} - 1$. We show the desired claim by induction over $k \geq 0$.

Base case: When $k = 0$, Assumption 3 implies that

$$\mathcal{W}_{1,\rho}(\delta_x P^\tau, \xi) \leq \sup_{s,s'} \rho(s, s') \leq 1 \leq c_0,$$

so that the base case ($k = 0$) holds for our induction proof.

Induction step: At step k of the argument, the induction hypothesis ensures that

$$\mathcal{W}_{1,\rho}\big(\delta_x P^{kt_{\mathrm{mix}}+q}, \xi\big) \leq c_0 \cdot 2^{-k}, \quad \text{for } q = 0, 1, \cdots, t_{\mathrm{mix}} - 1. \tag{95}$$

We now need to show that the result holds for any $\tau = (k+1)t_{\mathrm{mix}}+q$, where $q \in \{0, 1, \ldots, t_{\mathrm{mix}}-1\}$ is arbitrary. We do so via a coupling argument. Take a random initial state $y \sim \xi$, and consider two processes $\{s_t\}_{t\geq 0}$ and $\{s'_t\}_{t\geq 0}$ starting from x and y, respectively. Their joint distribution is defined as follows: choose the coupling between the law of $s_{kt_{\mathrm{mix}}+q}$ and $s'_{kt_{\mathrm{mix}}+q}$ to satisfy the identity $\mathbb{E}\big[\rho(s_{kt_{\mathrm{mix}}+q}, s'_{kt_{\mathrm{mix}}+q})\big] = \mathcal{W}_{1,\rho}\big(\delta_x P^{kt_{\mathrm{mix}}+q}, \xi\big)$. Conditionally on

$(s_{kt_{\text{mix}}+q}, s'_{kt_{\text{mix}}+q})$, Assumption 1 guarantees the existence of a coupling between $\delta_{s_{kt_{\text{mix}}+q}} P^{t_{\text{mix}}}$ and $s'_{kt_{\text{mix}}+q} P^{t_{\text{mix}}}$ such that

$$\mathbb{E}\big[\rho\big(s_{(k+1)t_{\text{mix}}+q}, s'_{(k+1)t_{\text{mix}}+q}\big) \mid (s_{kt_{\text{mix}}+q}, s'_{kt_{\text{mix}}+q})\big] \leq \tfrac{1}{2}\rho(s_{kt_{\text{mix}}+q}, s'_{kt_{\text{mix}}+q}).$$

Taking expectation on both sides and substituting with equation (95), we find that

$$\mathcal{W}_{1,\rho}\big(\delta_x P^{(k+1)t_{\text{mix}}+q}, \xi\big) \leq \mathbb{E}\big[\rho\big(s_{(k+1)t_{\text{mix}}+q}, s'_{(k+1)t_{\text{mix}}+q}\big)\big] \leq c_0 \cdot 2^{-(k+1)},$$

which completes the proof of the induction step.

B.2 Proof of Lemma 5

Our proof is based on the following intermediate claim

$$\big(\mathbb{E}\big[\|\Delta_{t+\ell}\|_2^p\big]\big)^{1/p} \leq e\big(\mathbb{E}\big[\|\Delta_t\|_2^p\big]\big)^{1/p} + 6\eta p\ell\sqrt{d}\big(\sigma_L\|\bar{\theta}\|_2 + \sigma_b\big). \tag{96}$$

This bound, which we return to prove at the end of this section, is a weaker form of the claim in the lemma.

We now use the bound (96) to prove the lemma. Applying Minkowski's inequality to the recursive relation (46), we find that for any $p \geq 2$, the p^{th} moment is upper bounded as

$$\big(\mathbb{E}\big[\|\Delta_{t+\ell+1} - \Delta_t\|_2^p\big]\big)^{1/p} \leq \big(\mathbb{E}\big[\|\Delta_{t+\ell} - \Delta_t\|_2^p\big]\big)^{1/p} + \eta\big(\mathbb{E}\big[\|L_{t+\ell+1}\Delta_{t+\ell}\|_2^p\big]\big)^{1/p} + \eta\big(\mathbb{E}\big[\|\nu_{t+\ell} + \zeta_{t+\ell+1}\|_2^p\big]\big)^{1/p}.$$

For the martingale part of the noise, we take the decomposition $L_{t+\ell+1} = \boldsymbol{L}(s_{t+\ell}) + Z_{t+\ell+1}$. By Assumption 2 and Hölder's inequality, we have the bounds

$$\mathbb{E}\big[\|Z_{t+\ell+1}\Delta_{t+\ell}\|_2^p \mid \mathcal{F}_t\big] \leq d^{\frac{p}{2}} \sum_{j=1}^{d} \mathbb{E}\big[\langle e_j,\, Z_{t+\ell+1}\Delta_{t+\ell}\rangle^p \mid \mathcal{F}_t\big] \leq \big(p\sigma_L\sqrt{d}\big)^p \mathbb{E}\big[\|\Delta_{t+\ell}\|_2^p \mid \mathcal{F}_t\big], \quad \text{and}$$

$$\mathbb{E}\big[\|\zeta_{t+\ell+1}\|_2^p \mid \mathcal{F}_t\big] \leq d^{\frac{p}{2}} \sum_{j=1}^{d} \mathbb{E}\big[\langle e_j,\, \zeta_{t+\ell+1}\rangle^p \mid \mathcal{F}_t\big] \leq (p\sqrt{d})^p \cdot \big(\sigma_L\|\bar{\theta}\|_2 + \sigma_b\big)^p.$$

Similarly, for the Markov part of the noise, we have:

$$\mathbb{E}\big[\|\nu_{t+\ell+1}\|_2^p\big] \leq (p\sqrt{d})^p \cdot \big(\sigma_L\|\bar{\theta}\|_2 + \sigma_b\big)^p.$$

On the other hand, the Lipschitz condition (4) and the boundedness condition (3) of the metric space imply that

$$\|L_{t+\ell+1}(s) - \bar{L}\|_{\text{op}} \leq \sigma_L d, \quad \text{and} \quad \|\boldsymbol{b}(s) - \bar{b}\|_2 \leq \sigma_b\sqrt{d} \quad \text{for all } s \in \mathbb{X}.$$

Substituting into the decomposition above, we arrive at the bounds

$$\big(\mathbb{E}\big[\|L_{t+\ell+1}\Delta_{t+\ell}\|_2^p\big]\big)^{1/p} \leq (\gamma_{\max} + \sigma_L p\sqrt{d} + \sigma_L d)\big(\mathbb{E}\big[\|\Delta_{t+\ell}\|_2^p\big]\big)^{1/p}, \quad \text{and}$$
$$\big(\mathbb{E}\big[\|\nu_{t+\ell} + \zeta_{t+\ell+1}\|_2^p\big]\big)^{1/p} \leq 2p\sqrt{d}\big(\sigma_L\|\bar{\theta}\|_2 + \sigma_b\big).$$

Applying equation (96) yields

$$\big(\mathbb{E}\big[\|\Delta_{t+\ell+1} - \Delta_t\|_2^p\big]\big)^{1/p} \leq \big(\mathbb{E}\big[\|\Delta_{t+\ell} - \Delta_t\|_2^p\big]\big)^{1/p} + e\eta(\gamma_{\max} + \sigma_L d)\big(\mathbb{E}\big[\|\Delta_t\|_2^p\big]\big)^{1/p} + 2(1 + 6\eta\ell)\eta p\sqrt{d}\big(\sigma_L\|\bar{\theta}\|_2 + \sigma_b\big).$$

Solving this recursion leads to the bound

$$\big(\mathbb{E}\big[\|\Delta_{t+\ell}-\Delta_t\|_2^p\big]\big)^{1/p} \le e\eta\ell(\gamma_{\max}+\sigma_L d)\big(\mathbb{E}\big[\|\Delta_t\|_2^p\big]\big)^{1/p} + 3\eta p\ell\sqrt{d}\big(\sigma_L\|\bar{\theta}\|_2+\sigma_b\big),$$

which establishes the first claim.

Since the stepsize is upper bounded as $\eta \le \big(2e\eta\ell(\gamma_{\max}+\sigma_L d)\big)^{-1}$, we have the lower bound

$$\begin{aligned}\big(\mathbb{E}\big[\|\Delta_{t+\ell}\|_2^p\big]\big)^{1/p} &\ge \big(\mathbb{E}\big[\|\Delta_t\|_2^p\big]\big)^{1/p} - \big(\mathbb{E}\big[\|\Delta_{t+\ell}-\Delta_t\|_2^p\big]\big)^{1/p} \\ &\ge \tfrac{1}{2}\big(\mathbb{E}\big[\|\Delta_t\|_2^p\big]\big)^{1/p} - 3\eta p\ell\sqrt{d}\big(\sigma_L\|\bar{\theta}\|_2+\sigma_b\big),\end{aligned}$$

which, in conjunction with the bound (96), establishes the second claim.

Proof of equation (96): Applying Minkowski's inequality to the recursive relation (46) yields (for any $p \ge 2$) a bound on the p^{th} conditional moment:

$$\big(\mathbb{E}\big[\|\Delta_{t+\ell+1}\|_2^p\big]\big)^{1/p} \le \big(\mathbb{E}\big[\|(I-\eta L_{t+\ell+1})\Delta_{t+\ell}\|_2^p\big]\big)^{1/p} + \eta\big(\mathbb{E}\big[\|\nu_{t+\ell}+\zeta_{t+\ell+1}\|_2^p\big]\big)^{1/p}. \tag{97}$$

Our next step is to bound the two terms above.

Substituting into the recursive relation (97), and applying Minkowski's inequality, we find that the moment $\big(\mathbb{E}\big[\|\Delta_{t+\ell+1}\|_2^p\big]\big)^{1/p}$ is upper bounded by

$$(1+\eta\gamma_{\max})\big(\mathbb{E}\big[\|\Delta_{t+\ell}\|_2^p\big]\big)^{1/p} + \eta\sigma_L d\big(\mathbb{E}\big[\|\Delta_{t+\ell}\|_2^p\big]\big)^{1/p} + 2\eta p\sqrt{d}\big(\sigma_L\|\bar{\theta}\|_2+\sigma_b\big).$$

Solving this recursive inequality leads to

$$\big(\mathbb{E}\big[\|\Delta_{t+\ell}\|_2^p\big]\big)^{1/p} \le \exp\big(\eta\ell(\gamma_{\max}+\sigma_L d)\big)\Big(\big(\mathbb{E}\big[\|\Delta_t\|_2^p\big]\big)^{1/p} + 2\eta p\ell\sqrt{d}\big(\sigma_L\|\bar{\theta}\|_2+\sigma_b\big)\Big).$$

For any stepsize $\eta \in \big(0, \frac{1}{(\gamma_{\max}+\sigma_L d)\ell}\big]$, we have

$$\big(\mathbb{E}\big[\|\Delta_{t+\ell}\|_2^p\big]\big)^{1/p} \le e\big(\mathbb{E}\big[\|\Delta_t\|_2^p\big]\big)^{1/p} + 6\eta p\ell\sqrt{d}\big(\sigma_L\|\bar{\theta}\|_2+\sigma_b\big),$$

which establishes the claim.

B.3 Proof of Lemma 6

For notational simplicity, we extend the process $(\Delta_t)_{t\ge 0}$ to the entire set $\mathbb{Z}$ of integers, in particular by defining $\Delta_t := \Delta_0$ for negative integer t. Note that under our assumption, Lemma 5 and the assumed bound (62) both hold true for the extended process, with index set $t \in \mathbb{Z}$. Moreover, as in the proof of Lemma 5, for each $p \ge 2$, we have the moment bound

$$\begin{aligned}\big(\mathbb{E}\big[\|\Delta_{t+\ell+1}-\Delta_t\|_2^p\big]\big)^{1/p} \le \big(\mathbb{E}\big[\|\Delta_{t+\ell}-\Delta_t\|_2^p\big]\big)^{1/p} &+ \eta\big(\mathbb{E}\big[\|L_{t+\ell+1}\Delta_{t+\ell}\|_2^p\big]\big)^{1/p} \\ &+ \eta\big(\mathbb{E}\big[\|\nu_{t+\ell}+\zeta_{t+\ell+1}\|_2^p\big]\big)^{1/p}.\end{aligned}$$

Our next step is to exploit the coarse bound (62) so as to obtain upper bounds on the second term $\big(\mathbb{E}\big[\|L_{t+\ell+1}\Delta_{t+\ell}\|_2^p\big]\big)^{1/p}$. Given the time lag $\tau > 0$, we take the decomposition $\Delta_{t+\ell} = \Delta_{t+\ell-\tau} + (\Delta_{t+\ell}-\Delta_{t+\ell-\tau})$, and by Minkowski's inequality, we have that

$$\big(\mathbb{E}\big[\|L_{t+\ell+1}\Delta_{t+\ell}\|_2^p\big]\big)^{1/p} \le \big(\mathbb{E}\big[\|L_{t+\ell+1}\Delta_{t+\ell-\tau}\|_2^p\big]\big)^{1/p} + \big(\mathbb{E}\big[\|L_{t+\ell+1}(\Delta_{t+\ell}-\Delta_{t+\ell-\tau})\|_2^p\big]\big)^{1/p}. \tag{98}$$

The latter term of the bound (98) can be controlled through Assumption 4:

$$\|L_{t+\ell+1}(s_{t+\ell})(\Delta_{t+\ell}-\Delta_{t+\ell-\tau})\|_2 \le (\gamma_{\max}+\sigma_L d)\|\Delta_{t+\ell}-\Delta_{t+\ell-\tau}\|_2, \quad \text{a.s.}$$

The distance $\|\Delta_{t+\ell}-\Delta_{t+\ell-\tau}\|_2$ is controlled via the coarse bound (62). Putting together the pieces, we find that

$$\big(\mathbb{E}\big[\|L_{t+\ell+1}(\Delta_{t+\ell}-\Delta_{t+\ell-\tau})\|_2^p\big]\big)^{1/p} \le \eta\big(\gamma_{\max}+\sigma_L d\big)\cdot\Big(\omega_p\big(\mathbb{E}\big[\|\Delta_{t+\ell-\tau}\|_2^p\big]\big)^{1/p}+\beta_p\bar{\sigma}\Big). \quad (99)$$

In order to bound the former term $\big(\mathbb{E}\big[\|L_{t+\ell+1}\Delta_{t+\ell-\tau}\|_2^p\big]\big)^{1/p}$ in the bound (98), we invoke Lemma 4, and obtain a random variable $\widetilde{s}_{t+\ell}$, such that

$$\widetilde{s}_{t+\ell} \mid \mathcal{F}_{t+\ell-\tau} \sim \xi, \quad \text{and} \quad \big(\mathbb{E}\big[\rho(s_{t+\ell},\widetilde{s}_{t+\ell-\tau})^p \mid \mathcal{F}_{t+\ell-\tau}\big]\big)^{1/p} \le c_0\cdot 2^{1-\frac{\tau}{2t_{\text{mix}}p}}. \quad (100)$$

By Assumption 2, we have the bounds

$$\mathbb{E}\big[\|Z_{t+\ell+1}\Delta_{t+\ell-\tau}\|_2^p \mid \mathcal{F}_{t+\ell-\tau}\big] \le (p\sqrt{d}\sigma_L)^p\|\Delta_{t+\ell-\tau}\|_2^p, \quad \text{and} \quad (101a)$$

$$\mathbb{E}\big[\|\big(\boldsymbol{L}(\widetilde{s}_{t+\ell-\tau})-\bar{L}\big)\cdot\Delta_{t+\ell-\tau}\|_2^p \mid \mathcal{F}_{t+\ell-\tau}\big] \le (p\sqrt{d}\sigma_L)^p\|\Delta_{t+\ell-\tau}\|_2^p. \quad (101b)$$

Invoking the moment bound (100) and using the Lipschitz condition (4), we find that

$$\begin{aligned}\mathbb{E}\big[\|\big(\boldsymbol{L}(\widetilde{s}_{t+\ell-\tau})-\boldsymbol{L}(s_{t+\ell-\tau})\big)\cdot\Delta_{t+\ell-\tau}\|_2^p \mid \mathcal{F}_{t+\ell-\tau}\big] &\le \mathbb{E}\big[\|\boldsymbol{L}(\widetilde{s}_{t+\ell-\tau})-\boldsymbol{L}(s_{t+\ell-\tau})\|_{\text{op}}^p \mid \mathcal{F}_{t+\ell-\tau}\big]\cdot\|\Delta_{t+\ell-\tau}\|_2^p \\ &\le \big(\sigma_L c_0 d\cdot 2^{1-\frac{\tau}{2t_{\text{mix}}p}}\|\Delta_{t+\ell-\tau}\|_2\big)^p. \quad (101c)\end{aligned}$$

Finally, we have the operator norm bound

$$\|\bar{L}\Delta_{t+\ell-\tau}\|_2 \le \gamma_{\max}\|\Delta_{t+\ell-\tau}\|_2. \quad (101d)$$

Collecting the results from equations (101)(a)—(d), we arrive at the bound

$$\big(\mathbb{E}\big[\|L_{t+\ell+1}\Delta_{t+\ell-\tau}\|_2^p \mid \mathcal{F}_{t+\ell-\tau}\big]\big)^{1/p} \le \big(2p\sqrt{d}\sigma_L+\gamma_{\max}+\sigma_L c_0 d\cdot 2^{1-\frac{\tau}{2t_{\text{mix}}p}}\big)\|\Delta_{t+\ell-\tau}\|_2. \quad (102)$$

According to Lemma 5, given a stepsize bounded as $\eta \le \big(6(\gamma_{\max}+\sigma_L d)\tau\big)^{-1}$, we have

$$\big(\mathbb{E}\|\Delta_{t+\ell-\tau}\|_2^p\big)^{1/p} \le 2\big(\mathbb{E}\|\Delta_{t+\ell}\|_2^p\big)^{1/p}+12\eta p\tau\sqrt{d}\big(\sigma_L\|\bar{\theta}\|_2+\sigma_b\big).$$

Collecting the bounds (99) and (102), and substituting into the decomposition (98), for $\tau \ge 2t_{\text{mix}}p\log(c_0 d)$, we arrive at the inequality:

$$\begin{aligned}\big(\mathbb{E}\big[\|L_{t+\ell+1}\Delta_{t+\ell}\|_2^p\big]\big)^{1/p} \le 2\Big(\big(p\sqrt{d}\sigma_L+\gamma_{\max}\big)+\eta\omega_p\big(\gamma_{\max}+\sigma_L d\big)\Big)\cdot\Big(\big(\mathbb{E}\|\Delta_{t+\ell}\|_2^p\big)^{1/p}+\eta p\tau\sqrt{d}\bar{\sigma}\Big) \\ +\eta\big(\gamma_{\max}+\sigma_L d\big)\beta_p\bar{\sigma}.\end{aligned}$$

By following the derivation in the proof of Lemma 5, we can show that the third term is upper bounded as

$$\big(\mathbb{E}\big[\|\nu_{t+\ell}+\zeta_{t+\ell+1}\|_2^p\big]\big)^{1/p} \le 2p\sqrt{d}(\sigma_L\|\bar{\theta}\|_2+\sigma_b).$$

Substituting back into the original decomposition, we find that the difference in moments $D := \big(\mathbb{E}\big[\|\Delta_{t+\ell+1} - \Delta_t\|_2^p\big]\big)^{1/p} - \big(\mathbb{E}\big[\|\Delta_{t+\ell} - \Delta_t\|_2^p\big]\big)^{1/p}$ is bounded as

$$D \le 2\eta\Big\{\big(p\sqrt{d}\sigma_L + \gamma_{\max}\big) + \eta\omega_p\big(\gamma_{\max} + \sigma_L d\big)\Big\} \cdot \Big(\big(\mathbb{E}\|\Delta_{t+\ell}\|_2^p\big)^{1/p} + \eta p\tau\sqrt{d}\bar{\sigma}\Big) + \big(2\eta p\sqrt{d} + \eta^2\big(\gamma_{\max} + \sigma_L d\big)\beta_p\big)$$

Lemma 5 implies that $\big(\mathbb{E}\big[\|\Delta_{t+\ell}\|_2^p\big]\big)^{1/p} \le e\big(\mathbb{E}\big[\|\Delta_t\|_2^p\big]\big)^{1/p} + 6\eta p\ell\sqrt{d}\bar{\sigma}$ and solving the recursion, we arrive at the bound

$$\begin{aligned}
&\big(\mathbb{E}\big[\|\Delta_{t+\ell} - \Delta_t\|_2^p\big]\big)^{1/p} \\
&\le 12\eta\ell\big(\big(p\sqrt{d}\sigma_L + \gamma_{\max}\big) + \eta\omega_p\big(\gamma_{\max} + \sigma_L d\big)\big) \cdot \Big(\big(\mathbb{E}\|\Delta_t\|_2^p\big)^{1/p} + \eta p(\tau+\ell)\sqrt{d}\bar{\sigma}\Big) \\
&\quad + \big(2\eta p\sqrt{d} + \eta^2\big(\gamma_{\max} + \sigma_L d\big)\beta_p\big)\ell\bar{\sigma} \\
&\le \eta\big(12\big(p\sqrt{d}\sigma_L + \gamma_{\max}\big)\ell + \tfrac{\omega_p}{2}\big)\Big(\big(\mathbb{E}\|\Delta_t\|_2^p\big)^{1/p} + \eta p(\tau+\ell)\sqrt{d}\bar{\sigma}\Big) + \eta\big(2p\ell\sqrt{d} + \tfrac{1}{2}\beta_p\big)\bar{\sigma},
\end{aligned}$$

for any $\tau \ge 2t_{\mathrm{mix}} p\log(c_0 d)$ and stepsize choice $\eta \le \frac{c}{48(\gamma_{\max}+\sigma_L d)}$.

C Auxiliary results underlying Theorem 1

In this appendix, we prove two auxiliary lemmas that were used in the proof of Theorem 1.

C.1 Proof of Lemma 9

According to Lemma 4, given $\tau > 0$ fixed, for any $t \ge \tau + k_m$, there exists a random variable $\widetilde{s}_{t-k_m}$ such that $\widetilde{s}_{t-k_m} \mid \mathcal{F}_{t-k_m-\tau} \sim \xi$, and $\mathbb{E}\big[\rho(s_{t-k_m}, \widetilde{s}_{t-k_m}) \mid \mathcal{F}_{t-\tau-k_m}\big] \le c_0 \cdot 2^{1-\frac{\tau}{t_{\mathrm{mix}}}}$. By Assumption 1, conditionally on the pair of states $(s_{t-k_m}, \widetilde{s}_{t-k_m})$, we have the following bound for $j \in [m]$:

$$\mathcal{W}_{\rho,1}\big(P^{k_j-k_{j-1}}\delta_{s_{t-k_j}}, P^{k_j-k_{j-1}}\delta_{\widetilde{s}_{t-k_j}}\big) \le c_0 \cdot \rho\big(s_{t-k_j}, \widetilde{s}_{t-k_j}\big), \quad \text{a.s.}$$

Consequently, there exists a sequence of random variables $(\widetilde{s}_{t-k_j})_{0\le j\le m-1}$, such that the following relations hold true for $j = 1, 2, \cdots, m$:

$$\begin{aligned}
&\widetilde{s}_{t-k_{j-1}} \mid \mathcal{F}_{t-k_m} \sim P^{k_j-k_{j-1}}\delta_{\widetilde{s}_{t-k_j}}, \quad \text{and} \\
&\mathbb{E}\big[\rho\big(\widetilde{s}_{t-k_{j-1}}, s_{t-k_{j-1}}\big) \mid \mathcal{F}_{t+k-\ell}\big] \le c_0^{m+1-j} \cdot \rho\big(s_{t-k_m}, \widetilde{s}_{t-k_m}\big).
\end{aligned}$$

Based on above construction, we consider the following decomposition:

$$\begin{aligned}
\big(\prod_{j=0}^m N_{t-k_j}\big)\Delta_{t-k_m} &= \big(\prod_{j=0}^m N(s_{t-k_j}) - \prod_{j=0}^m N(\widetilde{s}_{t-k_j})\big)\Delta_{t-k_m-\tau} + \big(\prod_{j=0}^m N(\widetilde{s}_{t-k_j})\big)\cdot\Delta_{t-k_m-\tau} \\
&\quad + \big(\prod_{j=0}^m N(s_{t-k_j})\big)\cdot\big(\Delta_{t-k_m} - \Delta_{t-\tau-k_m}\big) := Q_1(t) + Q_2(t) + Q_3(t).
\end{aligned} \tag{103}$$

In the following, we bound the moments for the summation of the three terms above, respectively. For the first term, we note the telescoping equation:

$$\prod_{j=0}^{m} N(s_{t-k_j}) - \prod_{j=0}^{m} N(\widetilde{s}_{t-k_j}) = \sum_{q=0}^{m} \big(\prod_{j=0}^{q-1} N(s_{t-k_j}) \big) \cdot \big(\boldsymbol{L}(s_{t-k_q}) - \boldsymbol{L}(\widetilde{s}_{t-k_q}) \big) \cdot \big(\prod_{j=q+1}^{m} N(\widetilde{s}_{t-k_j}) \big).$$

Note that each matrix in the product has operator norm uniformly bounded by $\sigma_L d$. We can then use the Lipschitz condition 4 as well as the bound on the distance $\rho(s_{t-k_q}, \widetilde{s}_{t-k_q})$, and obtain the bound

$$\begin{aligned}
&\mathbb{E}\big[\| \prod_{j=0}^{m} N(s_{t-k_j}) - \prod_{j=0}^{m} N(\widetilde{s}_{t-k_j}) \|_{\mathrm{op}}^2 \mid \mathcal{F}_{t-k_m-\tau} \big] \\
&\leq (m+1) \cdot (\sigma_L d)^m \sum_{q=0}^{m} \mathbb{E}\big[\| \boldsymbol{L}(s_{t-k_q}) - \boldsymbol{L}(\widetilde{s}_{t-k_q}) \|_{\mathrm{op}}^2 \mid \mathcal{F}_{t-k_m-\tau} \big] \;\leq\; (m+1)^2 (c_0 \sigma_L d)^{m+1} \cdot 2^{-\frac{\tau}{t_{\mathrm{mix}}}}.
\end{aligned}$$

Applying the bound on $\|\Delta_{t-\tau}\|_2$ in Proposition 1 and taking $\tau \geq 3 m t_{\mathrm{mix}} p \log(c_0 d n)$, we find that

$$\begin{aligned}
\mathbb{E}\big[\|Q_1(t)\|_2^2 \big] &\leq \mathbb{E}\big[\mathbb{E}\big[\| \prod_{j=0}^{m} N(s_{t-k_j}) - \prod_{j=0}^{m} N(\widetilde{s}_{t-k_j}) \|_{\mathrm{op}}^2 \mid \mathcal{F}_{t-k_m-\tau} \big] \cdot \|\Delta_{t-\tau-k_m}\|_2^2 \big] \\
&\leq (m+1)^2 (c_0 \sigma_L d)^{m+1} \cdot 2^{-\frac{\tau}{t_{\mathrm{mix}}}} c \bar{\sigma}^2 \tfrac{\eta \tau d \log^2 n}{1-\kappa} \;\leq\; \tfrac{\sigma_L^{m+1}}{n^2} \bar{\sigma}^2. \qquad (104)
\end{aligned}$$

Now we turn to bounding the term $Q_2(t)$. First, we note that

$$\begin{aligned}
\mathbb{E}\big[\|Q_2(t)\|_2^2 \big] &\leq \mathbb{E}\big[\| \prod_{j=0}^{m-1} N(\widetilde{s}_{t-k_j}) \|_{\mathrm{op}}^2 \cdot \|N(\widetilde{s}_{t-k_m}) \Delta_{t-k_m-\tau}\|_2^2 \big] \\
&\leq (\sigma_L d)^{2m} \mathbb{E}\big[\|N(\widetilde{s}_{t-k_m}) \Delta_{t-k_m-\tau}\|_2^2 \big] \leq (\sigma_L d)^{2m} \cdot \sigma_L^2 d \cdot \mathbb{E}\big[\|\Delta_{t-k_m-\tau}\|_2^2 \big].
\end{aligned}$$

By Proposition 1, for $t \geq n_0$ and $n_0 \geq 2(\tau + k_m)$, we have: $\mathbb{E}\big[\|\Delta_{t-k_m-\tau}\|_2^2 \big] \leq \frac{c\eta}{1-\kappa} t_{\mathrm{mix}} d \bar{\sigma}^2$. If $m = 0$, we have that $\mathbb{E}\big[N(\widetilde{s}_{t+\tau}) \mid \mathcal{F}_t \big] = 0$ almost surely for each $t \geq n_0$. For $m \geq 1$, the conditional unbiasedness does not hold true, but we still have the following upper bound on the bias

$$\begin{aligned}
\| \mathbb{E}\big[\prod_{j=0}^{m} N(\widetilde{s}_{t+k_m+\tau-k_j}) \mid \mathcal{F}_t \big] \|_{\mathrm{op}} &= \sup_{u,v \in \mathbb{S}^{d-1}} \mathbb{E}\big[\langle u, \prod_{j=0}^{m} N(\widetilde{s}_{t+k_m+\tau-k_j}) v \rangle \big] \\
&\leq \sup_{u,v \in \mathbb{S}^{d-1}} \mathbb{E}\big[\|N(\widetilde{s}_{t+k_m+\tau})^\top u\|_2 \cdot \| \prod_{j=1}^{m-1} N(\widetilde{s}_{t+k_m+\tau-k_j}) \|_{\mathrm{op}} \cdot \|N(\widetilde{s}_{t+\tau}) v\|_2 \big] \\
&\leq (\sigma_L d)^{m-1} \sup_{u,v \in \mathbb{S}^{d-1}} \sqrt{\mathbb{E}\|N(\widetilde{s}_{t+k_m+\tau})^\top u\|_2^2 \cdot \mathbb{E}\|N(\widetilde{s}_{t+\tau}) v\|_2^2} \\
&\leq (\sigma_L d)^{m-1} \cdot \sigma_L^2 d.
\end{aligned}$$

Denote $Y_t := \prod_{j=0}^m N(s_{t-k_j})$ and $\widetilde{Y}_t := \prod_{j=0}^m N(\widetilde{s}_{t-k_j})$ for any $t \geq k_m$. We have the expansion:

$$\begin{aligned}\mathbb{E}\big[\|\sum_{t=n_0}^{n-1} Q_2(t)\|_2^2\big] &\leq 2\mathbb{E}\big[\|\sum_{t=n_0}^{n-1} \mathbb{E}[\widetilde{Y}_t]\cdot\Delta_{t-k_m-\tau}\|_2^2\big] + 2\mathbb{E}\big[\|\sum_{t=n_0}^{n-1}(\widetilde{Y}_t - \mathbb{E}[\widetilde{Y}_t])\cdot\Delta_{t-k_m-\tau}\|_2^2\big] \\ &\leq 2n\big(d^m\sigma_L^{m+1}\big)^2 \sum_{t=n_0}^{n}\mathbb{E}\|\Delta_{t-k_m-\tau}\|_2^2 \\ &\quad + 2\sum_{n_0\leq s,t\leq n-1}\mathbb{E}\big[\langle(\widetilde{Y}_t - \mathbb{E}[\widetilde{Y}_t])\cdot\Delta_{t-k_m-\tau},\,(\widetilde{Y}_s - \mathbb{E}[\widetilde{Y}_s])\cdot\Delta_{s-k_m-\tau}\rangle\big].\end{aligned}$$

Note that in the special case of $m = 0$, we have $\mathbb{E}[\widetilde{Y}_t] = 0$ so that the bound holds without the first term on the RHS.

For $t > s + \tau + k_m$, we have the relations

$$\mathbb{E}\big[(\widetilde{Y}_t - \mathbb{E}[\widetilde{Y}_t])\cdot\Delta_{t-k_m-\tau} \mid \widetilde{\mathcal{F}}_{t-k_m-\tau}\big] = 0, \quad \text{and} \quad (\widetilde{Y}_s - \mathbb{E}[\widetilde{Y}_s])\cdot\Delta_{s-k_m-\tau} \in \widetilde{\mathcal{F}}_{t-k_m-\tau},$$

meaning that the product term vanishes when $|s-t| > \tau + k_m$. Therefore, we arrive at the bound

$$\mathbb{E}\big[\|\sum_{t=n_0}^{n-1} Q_2(t)\|_2^2\big] \leq \begin{cases}\big(2n^2\big(d^m\sigma_L^{m+1}\big)^2 + 4n(k_m+\tau)\cdot(\sigma_L d)^{2m}\cdot\sigma_L^2 d\big)\cdot\frac{c\eta}{1-\kappa}dt_{\mathrm{mix}}\bar{\sigma}^2 & m \geq 1, \\ 4n\tau\sigma_L^2 d\cdot\frac{c\eta}{1-\kappa}dt_{\mathrm{mix}}\bar{\sigma}^2 & m = 0.\end{cases} \tag{105}$$

Now we turn to the last term in the decomposition (103). We start with the decomposition:

$$\Delta_t - \Delta_{t-\tau} = \eta\sum_{\ell=1}^{\tau}\big(L_{t-\ell+1}(s_{t-\ell})\Delta_{t-\ell} + \nu_{t-\ell} + \zeta_{t-\ell+1}\big).$$

We therefore have the following decomposition:

$$\begin{aligned}\mathbb{E}\big[\|\sum_{t=n_0}^{n-1} Q_3(t)\|_2^2\big] &\leq 4\eta^2\mathbb{E}\big[\|\sum_{t=n_0}^{n}\big\{Y_t\cdot\big(\sum_{\ell=1}^{\tau} Z_{t-k_m-\ell+1}\Delta_{t-k_m-\ell}\big)\big\}\|_2^2\big] + 4\eta^2\mathbb{E}\big[\|\sum_{t=n_0}^{n}\big\{Y_t\cdot\big(\bar{L}\sum_{\ell=1}^{\tau}\Delta_{t-k_m-\ell}\big)\big\}\|_2^2\big] \\ &\quad + 4\eta^2\mathbb{E}\big[\|\sum_{t=n_0}^{n}\big\{Y_t\cdot\big(\sum_{\ell=1}^{\tau} N_{t-k_m-\ell}\Delta_{t-k_m-\ell}\big)\big\}\|_2^2\big] \\ &\quad + 4\eta^2\mathbb{E}\big[\|\sum_{t=n_0}^{n}\big\{Y_t\cdot\big(\sum_{\ell=1}^{\tau}(\nu_{t-k_m-\ell} + \zeta_{t-k_m-\ell+1})\big)\big\}\|_2^2\big]\end{aligned}$$

For the martingale component of the noise, note that each term $\prod_{j=0}^m N(s_{t-k_j})\cdot Z_{t-\ell+1}(s_{t-\ell})$ has zero conditional mean conditioned on $\mathcal{F}_{t-\ell}$. We have that

$$\begin{aligned}\mathbb{E}\big[\|\sum_{t=n_0}^{n} Y_t Z_{t-k_m-\ell+1}(s_{t-k_m-\ell})\Delta_{t-k_m-\ell}\|_2^2\big] &= \sum_{t=n_0}^{n-1}\mathbb{E}\big[\|Y_t Z_{t-k_m-\ell+1}(s_{t-k_m-\ell})\Delta_{t-k_m-\ell}\|_2^2\big] \\ \leq (\sigma_L d)^{2(m+1)}\sum_{t=n_0}^{n-1}\mathbb{E}\big[\|Z_{t-k_m-\ell+1}(s_{t-k_m-\ell})\Delta_{t-k_m-\ell}\|_2^2\big] &\leq \sigma_L^{2m+4}d^{2m+3}n\cdot\frac{c\eta}{1-\kappa}dt_{\mathrm{mix}}\bar{\sigma}^2.\end{aligned}$$

From the Lipschitz condition (4) and the boundedness condition (3) on the metric space, it follows that $\|\!|Y_t|\!\|_{\mathrm{op}} \le (\sigma_L d)^{m+1}$ almost surely. Using this fact, the second term can be bounded as

$$\begin{aligned}\mathbb{E}\big[\|\sum_{t=n_0}^{n}\big\{Y_t\cdot\big(\bar{L}\sum_{\ell=1}^{\tau}\Delta_{t-k_m-\ell}\big)\big\}\|_2^2\big] &\le n\tau(\sigma_L d)^{2m+2}\gamma_{\max}^2\sum_{t=n_0}^{n-1}\sum_{\ell=1}^{\tau}\mathbb{E}\|\Delta_{t-k_m-\ell}\|_2^2\\ &\le n^2\tau^2(\sigma_L d)^{2m+2}\gamma_{\max}^2\cdot\tfrac{c\eta}{1-\kappa}dt_{\mathrm{mix}}\bar{\sigma}^2.\end{aligned}$$

Collecting equations (104) and (105) as well as the above bounds for Q_3, we arrive at the upper bound $\mathbb{E}\big[\|\sum_{t=n_0}^{n-1}\big(\prod_{j=0}^{m}N_{t-k_j}\big)\Delta_{t-k_m}\|_2^2\big] \le \sum_{j=1}^{3}T_j$, where

$$\begin{aligned}T_1 &:= n^2d^{2m}\sigma_L^{2m+2}\big(1+\eta^2\tau^2\gamma_{\max}^2d^2\sigma_L^2+\eta^2\tau^2d^3\sigma_L^2/n\big)\cdot\tfrac{c\eta}{1-\kappa}dt_{\mathrm{mix}}\bar{\sigma}^2\\ T_2 &:= 4\eta^2\mathbb{E}\big[\|\sum_{t=n_0}^{n}\big\{Y_t\big(\sum_{\ell=1}^{\tau}N_{t-k_m-\ell}\Delta_{t-k_m-\ell}\big)\big\}\|_2^2\big],\quad\text{and}\\ T_3 &:= 4\eta^2\mathbb{E}\big[\|\sum_{t=n_0}^{n}\big\{Y_t\big(\sum_{\ell=1}^{\tau}(\nu_{t-k_m-\ell}+\zeta_{t-k_m-\ell+1})\big)\big\}\|_2^2\big].\end{aligned}$$

In the special case of $m=0$, we have:

$$\begin{aligned}\mathbb{E}\big[\|\sum_{t=n_0}^{n-1}N_t\Delta_t\|_2^2\big] \le c\sigma_L^2d\cdot\big(n\tau+n^2\eta^2\sigma_L^2d\tau^2\big)\frac{c\eta}{1-\kappa}dt_{\mathrm{mix}}\bar{\sigma}^2+4\eta^2\tau\sum_{k_1=1}^{\tau}\mathbb{E}\big[\|\sum_{t=n_0}^{n}N_tN_{t-k_1}\Delta_{t-k_1}\|_2^2\big]\\ +4\eta^2\tau\sum_{k_1=1}^{\tau}\mathbb{E}\big[\|\sum_{t=n_0}^{n}N_t\big(\nu_{t-k_1}+\zeta_{t-k_1+1}\big)\|_2^2\big].\end{aligned}$$

which completes the proof of this lemma.

C.2 Proof of Lemma 10

We study the bias and variance of the summation separately. For the bias term, we have:

$$\begin{aligned}\|\mathbb{E}\big[\big(\prod_{j=0}^{m-1}N_{t-k_j}\big)\big(\nu_{t-k_m}+\zeta_{t-k_m+1}\big)\big]\|_2 &= \sup_{z\in\mathbb{S}^{d-1}}\mathbb{E}\big[\langle\big(\prod_{j=0}^{m-1}N_{t-k_j}\big)\big(\nu_{t-k_m}+\zeta_{t-k_m+1}\big),\,z\rangle\big]\\ &\overset{(i)}{\le}\sup_{z\in\mathbb{S}^{d-1}}\sqrt{\mathbb{E}\|N_t^\top z\|_2^2}\cdot\big[\mathbb{E}\|\big(\prod_{j=1}^{m-1}N_{t-k_j}\big)\big(\nu_{t-k}+\zeta_{t-k+1}\big)\|_2^2\big]^{1/2}\\ &\overset{(ii)}{\le}\sigma_L\sqrt{d}\cdot(\sigma_Ld)^{m-1}\cdot2\bar{\sigma}\sqrt{d}=2(\sigma_Ld)^m\bar{\sigma},\end{aligned}\tag{106}$$

where step (i) uses the Cauchy–Schwarz inequality, and step (ii) follows by invoking the moment assumption 2 as well as the Lipschitz assumption 4.

For $t\in[k_m,n]$, we define

$$\lambda_t := \big(\prod_{j=0}^{m-1}N_{t-k_j}\big)\big(\nu_{t-k_m}+\zeta_{t-k_m+1}\big)-\mathbb{E}\big[\big(\prod_{j=0}^{m-1}N_{t-k_j}\big)\big(\nu_{t-k_m}+\zeta_{t-k_m+1}\big)\big].$$

We have

$$\mathbb{E}\big[\|\lambda_t\|_2^2\big] \;\le\; \mathbb{E}\big[\big(\prod_{j=0}^{m-1}\|N_{t-k_j}\|_{\mathrm{op}}^2\big)\cdot\|\nu_{t-k_m}+\zeta_{t-k_m+1}\|_2^2\big] \le (\sigma_L d)^{2m}\cdot\mathbb{E}\big[\|\nu_{t-k}+\zeta_{t-k+1}\|_2^2\big]$$
$$\le d^{2m+1}\sigma_L^{2m}\bar{\sigma}^2.$$

For integers $t\ge 0$ and $\ell\ge k_m$, by Lemma 4, there exists a random variable $\widetilde{s}_{t+\ell-k_m}$, such that $\widetilde{s}_{t+\ell-k_m}\mid\mathcal{F}_t\sim\xi$, and that $\mathbb{E}\big[\rho(s_{t+\ell-k_m},\widetilde{s}_{t+\ell-k_m})\mid\mathcal{F}_t\big]\le c_0\cdot 2^{1-\frac{\ell-k_m}{t_{\mathrm{mix}}}}$. By Assumption 1, conditionally on the pair of states $(s_{t+\ell-k_m},\widetilde{s}_{t+\ell-k_m})$, we have the following bound for $j\in[m]$:

$$\mathcal{W}_{\rho,1}\big(P^{k_j-k_{j-1}}\delta_{s_{t+\ell-k_j}},P^{k_j-k_{j-1}}\delta_{\widetilde{s}_{t+\ell-k_j}}\big)\le c_0\cdot\rho\big(s_{t+\ell-k_j},\widetilde{s}_{t+\ell-k_j}\big),\quad\text{a.s.}$$

Consequently, there exists a sequence of random variables $(\widetilde{s}_{t+\ell-k_j})_{0\le j\le m-1}$, such that the following relations hold true for $j=1,2,\cdots,m$:

$$\widetilde{s}_{t+\ell-k_{j-1}}\mid\mathcal{F}_{t+\ell-k_m}\sim P^{k_j-k_{j-1}}\delta_{\widetilde{s}_{t+\ell-k_j}},\quad\text{and}$$
$$\mathbb{E}\big[\rho\big(\widetilde{s}_{t+\ell-k_{j-1}},s_{t+\ell-k_{j-1}}\big)\mid\mathcal{F}_{t+\ell-k_m}\big]\le c_0^{m+1-j}\cdot\rho\big(s_{t+\ell-k_m},\widetilde{s}_{t+\ell-k_m}\big).$$

Given the random variables constructed above, we can then construct the proxy random variable for $\lambda_{t+\ell}$:

$$\widetilde{\lambda}_{t+\ell}:=\big(\prod_{j=0}^{m-1}N(\widetilde{s}_{t+\ell-k_j})\big)\big(\nu(\widetilde{s}_{t+\ell-k_m})+\zeta_{t+\ell-k_m+1}(\widetilde{s}_{t+\ell-k_m})\big)-\mathbb{E}\big[\big(\prod_{j=0}^{m-1}N_{t-k_j}\big)\big(\nu_{t-k_m}+\zeta_{t-k_m+1}\big)\big].$$

By stationarity, we have $\mathbb{E}\big[\widetilde{\lambda}_{t+\ell}\mid\mathcal{F}_t\big]=0$ almost surely. In order to bound the difference, we note the telescope relation: $\widetilde{\lambda}_{t+\ell}-\lambda_{t+\ell}=\sum_{q=0}^{m-1}E_q^{(mix)}+\bar{E}^{(mix)}$, where

$$E_q^{(mix)}:=\big(\prod_{j=0}^{q-1}N(s_{t+\ell-k_j})\big)\big(\bar{L}(\widetilde{s}_{t+\ell-k_q})-\boldsymbol{L}(s_{t+\ell-k_q})\big)\big(\prod_{j=q+1}^{m-1}N(\widetilde{s}_{t+\ell-k_j})\big)\big(\nu(\widetilde{s}_{t+\ell-k_m})+\zeta_{t+\ell-k_m+1}(\widetilde{s}_{t+\ell-k_m})\big),$$

and $\bar{E}^{(mix)}:=\prod_{j=0}^{m-1}N(s_{t+\ell-k_j})\cdot\big(\nu(\widetilde{s}_{t+\ell-k_m})+\zeta_{t+\ell-k_m+1}(\widetilde{s}_{t+\ell-k_m})-\nu(s_{t+\ell-k_m})+\zeta_{t+\ell-k_m+1}(s_{t+\ell-k_m})\big)$.

Using the Wasserstein distance bounds and Lipschitz condition 4, we find the conditional expectation $A=\mathbb{E}\big[\|E_q^{(mix)}\|_2\mid\mathcal{F}_t\big]$ is bounded as

$$\begin{aligned}A&\le(\sigma_L d)^{m-1}\mathbb{E}\big[\|\boldsymbol{L}(s_{t+\ell-k_q})-\boldsymbol{L}(\widetilde{s}_{t+\ell-k_q})\|_{\mathrm{op}}\cdot\|\nu(\widetilde{s}_{t+\ell-k})+\zeta_{t+\ell-k+1}(\widetilde{s}_{t+\ell-k})\|_2\mid\widetilde{\mathcal{F}}_t\big]\\&\le(\sigma_L d)^m\sqrt{\mathbb{E}[\rho(s_{t+\ell-k_q},\widetilde{s}_{t+\ell-k_q})^2\mid\widetilde{\mathcal{F}}_t]}\cdot\sqrt{\mathbb{E}[\|\nu(\widetilde{s}_{t+\ell-k})+\zeta_{t+\ell-k+1}(\widetilde{s}_{t+\ell-k})\|_2^2\mid\widetilde{\mathcal{F}}_t]}\\&\le(\sigma_L d)^m c_0\cdot 2^{1-\frac{\ell-k_q}{2t_{\mathrm{mix}}}}\cdot 2d\bar{\sigma},\end{aligned}$$

and the conditional expectation $B=\mathbb{E}\big[\|\bar{E}^{(mix)}\|_2\mid\mathcal{F}_t\big]$ is bounded as

$$\begin{aligned}B&\le(\sigma_L d)^m\big(\sqrt{\mathbb{E}\big[\|\zeta_{t+\ell-k+1}(s_{t+\ell-k})-\zeta_{t+\ell-k+1}(\widetilde{s}_{t+\ell-k})\|_2^2\mid\mathcal{F}_t\big]}+\sqrt{\mathbb{E}\big[\|\nu(s_{t+\ell-k})-\nu(\widetilde{s}_{t+\ell-k})\|_2^2\mid\mathcal{F}_t\big]}\big)\\&\le(\sigma_L d)^m d\bar{\sigma}c_0\cdot 2^{1-\frac{\ell-k_m}{2t_{\mathrm{mix}}}}.\end{aligned}$$

Consequently, we can bound the cross term as

$$\begin{aligned}\mathbb{E}\big[\langle\lambda_t,\,\lambda_{t+\ell}\rangle\big] &= \mathbb{E}\big[\langle\lambda_t,\,\mathbb{E}\big[\widetilde{\lambda}_{t+\ell}\mid\mathcal{F}_t\big]\rangle\big] + \mathbb{E}\big[\langle\lambda_t,\,\mathbb{E}\big[\lambda_{t+\ell}-\widetilde{\lambda}_{t+\ell}\mid\mathcal{F}_t\big]\rangle\big]\\ &\le 0 + \mathbb{E}\big[\|\lambda_t\|_2\cdot\mathbb{E}\big[\|\lambda_{t+\ell}-\widetilde{\lambda}_{t+\ell}\|_2\mid\mathcal{F}_t\big]\big]\\ &\le 12c_0 d^{m+1}\sigma_L^m\bar{\sigma}\cdot 2^{-\frac{\ell-k}{2t_{\mathrm{mix}}}}\cdot\sqrt{\mathbb{E}\|\lambda_t\|_2^2}\\ &\le 12c_0 d^{2m+2}\sigma_L^{2m}\bar{\sigma}^2\cdot 2^{-\frac{\ell-k}{2t_{\mathrm{mix}}}}.\end{aligned}$$

Taking $\tau = 16t_{\mathrm{mix}}\log(c_0 d)$, we can control the cross terms in two different ways:

$$\mathbb{E}\big[\langle\lambda_t,\,\lambda_{t+\ell}\rangle\big] \le \begin{cases}\sqrt{\mathbb{E}\|\lambda_t\|_2^2}\cdot\sqrt{\mathbb{E}\|\lambda_{t+\ell}\|_2^2}\le d^{2m+1}\sigma_L^{2m}\bar{\sigma}^2, & 0\le\ell\le k_m+\tau,\\ 12c_0 d^{2m+2}\sigma_L^{2m}\bar{\sigma}^2\cdot 2^{-\frac{\ell-k}{2t_{\mathrm{mix}}}}\le d^{2m}\sigma_L^{2m}\bar{\sigma}^2 & \ell\ge k_m+\tau.\end{cases}$$

Summing them up these terms yields

$$\mathbb{E}\big[\|\sum_{t=n_0}^{n-1}\lambda_t\|_2^2\big] = \sum_{t=n_0}^{n-1}\mathbb{E}\|\lambda_t\|_2^2 + 2\sum_{n_0\le t_1<t_2\le n-1}\mathbb{E}\big[\langle\lambda_{t_1},\,\lambda_{t_2}\rangle\big] \le (k+\tau+1)nd^{2m+1}\sigma_L^{2m}\bar{\sigma}^2 + n^2 d^{2m}\sigma_L^{2m}\bar{\sigma}^2.$$

Combining with the bound (106), we find that

$$\begin{aligned}\mathbb{E}\big[\|\sum_{t=n_0}^{n-1}\big(\prod_{j=0}^{m-1}N_{t-k_j}\big)\big(\nu_{t-k_m}+\zeta_{t-k_m+1}\big)\|_2^2\big] &= \|\sum_{t=n_0}^{n-1}\mathbb{E}\big[\big(\prod_{j=0}^{m-1}N_{t-k_j}\big)\big(\nu_{t-k_m}+\zeta_{t-k_m+1}\big)\big]\|_2^2 + \mathbb{E}\big[\|\sum_{t=n_0}^{n-1}\lambda_t\|_2^2\big]\\ &\le c\big(n^2+(k_m+\tau)nd\big)\sigma_L^{2m}d^{2m}\bar{\sigma}^2,\end{aligned}$$

for a universal constant $c>0$.

D Auxiliary results underlying Theorem 2

In this section, we collect statements and proofs of some technical lemmas used in proving the lower bound in Theorem 2. Recall that the lower bound is defined by a local neighborhood of a given Markov chain with transition kernel P_0. Given a perturbation matrix $h = Qw$, the problem instance $\big(\bar{L}^{(h)}, \bar{b}^{(h)}\big)$ is the expectation of $(\boldsymbol{L}(s), \boldsymbol{b}(s))$ under the stationary distribution ξ_h of P_h. Recall the definition (76) of the Green function $\boldsymbol{g}_h$.

D.1 Proof of Lemma 16

By following the derivation of equation (86), we find that

$$\tfrac{\partial}{\partial h_x(y)}\bar{L}^{(h)} = \xi_h(x)P_h(x,y)\Big\{\mathcal{A}_h\boldsymbol{L}(y) - \sum_{z\in\mathbb{X}}P_h(x,z)\mathcal{A}\boldsymbol{L}(z)\Big\}.$$

Consequently, for any $u\in\mathbb{S}^{d-1}$, we have the bound

$$\begin{aligned}\|\nabla_w\big(\bar{L}^{(h)}u\big)\|_{\mathrm{op}} &\le \sup_{z,v\in\mathbb{S}^{d-1}}\sqrt{\mathbb{E}_{Y\sim\xi_h}\big[\big(z^\top\mathcal{A}_h\boldsymbol{L}(Y)u\big)^2\big]}\cdot\sqrt{\mathbb{E}_{X\sim\xi_h,Y\sim P_h(X,\cdot)}\big[\big((\boldsymbol{g}_0(Y)-\mathcal{P}_0\boldsymbol{g}_0(X))^\top v\big)^2\big]}\\ &\le \sup_{v\in\mathbb{S}^{d-1}}\sqrt{\mathbb{E}_{Y\sim\xi_h}\big[\|\mathcal{A}_h\boldsymbol{L}(Y)u\|_2^2\big]}\cdot\frac{3}{2}\sqrt{\mathbb{E}_{X\sim\xi_0,Y\sim P_0(X,\cdot)}\big[\big((\boldsymbol{g}_0(Y)-\mathcal{P}_0\boldsymbol{g}_0(X))^\top v\big)^2\big]}\\ &\le ct_{\mathrm{mix}}\sigma_L\sqrt{d\cdot\|\Lambda\|_{\mathrm{op}}}\log d.\end{aligned}$$

We thus obtain

$$\begin{aligned}\|\bar{L}^{(h)} - \bar{L}^{(0)}\|_{\mathrm{op}} &\le \sup_{u\in\mathbb{S}^{d-1}} \|(\bar{L}^{(h)} - \bar{L}^{(0)})u\|_2 \le \int_0^1 \sup_{u\in\mathbb{S}^{d-1}} \|\nabla_w(\bar{L}^{(sQw)}u)\cdot w\|_2 ds \\ &\le c t_{\mathrm{mix}}\sigma_L\sqrt{d\cdot \mathrm{trace}(\Lambda)}\log d\cdot\|w\|_2.\end{aligned}$$

Now given a perturbation vector satisfying the bound $\|w\|_2 \le \frac{1-\kappa}{2ct_{\mathrm{mix}}\sigma_L\sqrt{d\cdot\|\Lambda\|_{\mathrm{op}}}\log d}$, we have the following bound for any $u\in\mathbb{S}^{d-1}$:

$$\|(I-\bar{L}^{(h)})u\|_2 \ge \|(I-\bar{L}^{(0)})u\|_2 - \|(\bar{L}^{(h)}-\bar{L}^{(0)})u\|_2 \ge (1-\kappa) - \|\bar{L}^{(h)}-\bar{L}^{(0)}\|_{\mathrm{op}} \ge \tfrac{1-\kappa}{2},$$

which implies that $\|I-\bar{L}^{(h)})^{-1}\|_{\mathrm{op}} \le \frac{2}{1-\kappa}$, as claimed.

D.2 A useful moment bound

Finally, we state and prove a moment bound that is useful in multiple proofs. Recall that the operator $\mathcal{P}_h$ is a the perturbed probability transition kernel under perturbation matrix h, and the operator $\mathcal{A}_h$ is the Green function operator associated with this transition kernel.

Lemma 17. *Consider a bounded function $f:\mathbb{X}\to\mathbb{R}$, and a perturbation vector h satisfying the condition in Lemma 11. There there exists a universal constant $c>0$, such that for any integer $p\ge 1$*

$$\big(\mathbb{E}_{X\sim\xi_h}\big[\big(\mathcal{A}_h f(X)\big)^{2p}\big]\big)^{\frac{1}{2p}} \le c\,p\,t_{\mathrm{mix}}\Big[\mathbb{E}_{X\sim\xi_h}\big[f(X)^{2p}\big]\Big]^{\frac{1}{2p}}\log\Big\{\frac{\|f\|_\infty^{2p}}{\mathbb{E}_{X\sim\xi_h}\big[f(X)^{2p}\big]}\Big\}$$

The proof is similar to that of Lemma 7. For any function $f:\mathbb{X}\to\mathbb{R}$ such that $\mathbb{E}_{\xi_h}[f(X)]=0$, we first observe that $\mathcal{A}_h f(s)=\sum_{k=0}^\infty \mathcal{P}_h^k f(s)$ for all $s\in\mathbb{X}$. Note that Lemma 11 guarantees that the perturbed chain satisfies Assumption 1 with mixing time $4t_{\mathrm{mix}}$. By Lemma 4 and the coupling definition of total variation distance, for each $t\ge 0$, there exists a random variable $\widetilde{s}_t$ such that $\widetilde{s}_t \mid s_0\sim\xi_h$, and $\mathbb{P}\big(\widetilde{s}_t\neq s_t\mid s\big)\le 2^{-\lfloor\frac{t}{4t_{\mathrm{mix}}}\rfloor}$.

By construction, the state $\widetilde{s}_t$ is independent of s. Consequently, we have the equivalence $\mathcal{A}_h f(s)=\sum_{k=0}^\infty \mathbb{E}\big[f(s_k)-f(\widetilde{s}_k)\mid s\big]$, and for any $\alpha>0$,

$$\begin{aligned}\mathbb{E}_{s\sim\xi_h}\big[\big(\mathcal{A}_h f(s)\big)^{2p}\big] &\le \Big(\sum_{k=0}^\infty e^{2p\alpha t}\mathbb{E}\big(\mathbb{E}\big[f(s_k)-f(\widetilde{s}_k)\mid s\big]\big)^{2p}\Big)\cdot\Big(\sum_{k=0}^\infty e^{-\frac{2p}{2p-1}\alpha k}\Big)^{2p-1} \\ &\le \alpha^{1-2p}\sum_{k=0}^\infty e^{2p\alpha k}\mathbb{E}\big[\,|f(s_k)-f(\widetilde{s}_k)|^{2p}\,\big].\end{aligned}$$

We bound the moment of $f(s_k)-f(\widetilde{s}_k)$ for different values of k in two ways. On the one hand, Young's inequality directly leads to the following naive bound

$$\mathbb{E}\big[\,|f(s_k)-f(\widetilde{s}_k)|^{2p}\,\big]\le 2^{2p-1}\big(\mathbb{E}\big[f(s_k)^{2p}\big]+\mathbb{E}\big[f(\widetilde{s}_k)^{2p}\big]\big)=2^{2p}\mathbb{E}_{s\sim\xi_h}\big[f(s)^{2p}\big].$$

On the other hand, for any bounded function f, we have

$$\mathbb{E}\big[\,|f(s_k)-f(\widetilde{s}_k)|^{2p}\,\big]\le\|f\|_\infty^{2p}\cdot\mathbb{P}\big(s_k\neq\widetilde{s}_k\big)\le\|f\|_\infty^{2p}\cdot 2^{1-\frac{k}{4t_{\mathrm{mix}}}}.$$

Combining the two estimates yields the bound

$$\mathbb{E}\big[(\mathcal{A}_h f(X))^{2p}\big]\le\alpha^{1-2p}\Big\{2^{2p}\cdot e^{2p\alpha\tau}\tau\mathbb{E}_{s\sim\xi_h}\big[f(s)^{2p}\big]+\|f\|_\infty^{2p}\sum_{k=\tau+1}^\infty e^{2p\alpha k}\cdot 2^{1-\frac{k}{4t_{\mathrm{mix}}}}\Big\},$$

valid for any $\alpha>0$ and $\tau>0$. Setting $\tau=c\,t_{\mathrm{mix}}\log\frac{\|f\|_\infty^{2p}}{\mathbb{E}[f(X)^{2p}]}$ and $\alpha=\frac{1}{16\tau p}$ yields the claim.

E Proofs for TD(0)

We stated three corollaries applicable to this method, and in this section, we prove each of them in turn.

E.1 Proof of Corollary 1

The bulk of the proof involves verifying the conditions needed to apply Proposition 1 and Theorem 1, but some additional care is needed in order to deal with non-orthonormal basis functions $(\phi_j)_{j\in[d]}$. First, we note that the SA procedure (27) can be equivalently written as

$$\theta_{t+1} = (1-\eta\beta)\theta_t + \eta\beta \boldsymbol{L}_{t+1}(\omega_t)\theta_t - \eta\beta \boldsymbol{b}_{t+1}(\omega_t), \tag{107}$$

where $\boldsymbol{L}_{t+1}(\omega_t) := \big(I_d - \beta^{-1}\phi(s_t)\phi(s_t)^\top + \gamma\beta^{-1}\phi(s_t)\phi(s_{t+1})^\top\big)$, and $\boldsymbol{b}_{t+1}(\omega_t) := \beta^{-1}R_t(s_t)\phi(s_t)$. This is an SA scheme with stepsize $\eta\beta$.

For any matrix $A \in \mathbb{R}^{d\times d}$, define $\kappa(A) := \frac{1}{2}\lambda_{\max}\big(A + A^\top\big)$. We verify the eigenvalue condition (5) by noting that

$$\begin{aligned}\tfrac{1}{2}\lambda_{\max}\big(\bar{L} + \bar{L}^\top\big) &= 1 - \tfrac{1}{\beta}\,\kappa\big(\gamma\mathbb{E}_{s\sim\xi, s^+\sim P(s,\cdot)}\big[\phi(s)\phi(s^+)^\top\big] - \mathbb{E}_\xi\big[\phi(s)\phi(s)^\top\big]\big) \\ &= 1 - \tfrac{1}{\beta}\lambda_{\max}\big(B^{1/2}\big(I_d - \tfrac{M+M^\top}{2}\big)B^{1/2}\big) \;=\; 1 - \tfrac{\mu}{\beta}(1-\kappa) < 1,\end{aligned}$$

and

$$\|\bar{L}\|_{\mathrm{op}} \le 1 + \tfrac{1}{\beta}\big(\|\mathbb{E}_{s\sim\xi, s^+\sim P(s,\cdot)}\big[\phi(s)\phi(s^+)^\top\big]\|_{\mathrm{op}} + \|\mathbb{E}_\xi\big[\phi(s)\phi(s)^\top\big]\|_{\mathrm{op}}\big) \le 3.$$

For the two-step sliding-window Markov chain $\omega_t = (s_t, s_{t+1})$, Assumption 1 holds with mixing time $(t_{\mathrm{mix}}+1)$ in the discrete metric, and the metric space has diameter at most 1. It remains to verify the boundedness and moment assumptions.

In order to verify Assumption 4, we note that the bounds (28a) imply that

$$\begin{aligned}&\|\boldsymbol{L}_{t+1}(s_t)\|_{\mathrm{op}} \le 1 + \tfrac{1}{\beta}\big(\|\phi(s_t)\phi(s_{t+1})\|_{\mathrm{op}} + \|\phi(s_t)\phi(s_t)^\top\|_{\mathrm{op}}\big) \le (1+\varsigma^2)d, \quad \text{and} \\ &\|\boldsymbol{b}_{t+1}(s_t)\|_2 \le \tfrac{1}{\beta}|R_t(s_t)|\cdot\|\phi(s_t)\|_2 \le \varsigma^2\sqrt{d/\beta}.\end{aligned}$$

Turning to the moment assumption, given any vector $u \in \mathbb{S}^{d-1}$ and coordinate vector e_j, we have the bounds

$$\begin{aligned}\mathbb{E}_{s\sim\xi, s^+\sim P(s,\cdot)}\big[\big(e_j^\top\phi(s)\phi(s^+)^\top u\big)^2\big] &\le \sqrt{\mathbb{E}_{s\sim\xi}\big[\big(e_j^\top\phi(s)\big)^4\big]}\cdot\sqrt{\mathbb{E}_{s\sim\xi}\big[\big(u^\top\phi(s)\big)^4\big]} \le \beta^2\varsigma^4, \\ \mathbb{E}_{s\sim\xi}\big[\big(e_j^\top\phi(s)\phi(s)^\top u\big)^2\big] &\le \sqrt{\mathbb{E}_{s\sim\xi}\big[\big(e_j^\top\phi(s)\big)^4\big]}\cdot\sqrt{\mathbb{E}_{s\sim\xi}\big[\big(u^\top\phi(s)\big)^4\big]} \le \beta^2\varsigma^4, \quad \text{and} \\ \mathbb{E}_{s\sim\xi}\big[\big(e_j^\top R_t(s)\phi(s)\big)^2\big] &\le \varsigma^2\mathbb{E}_{s\sim\xi}\big[\big(e_j^\top\phi(s)\big)^2\big] \le \beta\varsigma^4.\end{aligned}$$

Finally, the quantity $\bar{\sigma}$ from equation (29) is bounded as

$$\begin{aligned}&\max_{j\in[d]}\mathbb{E}\big[\langle e_j,\, (\boldsymbol{L}_{t+1}(\omega_t) - \bar{L})\bar{\theta} + (\boldsymbol{b}_{t+1}(\omega_t) - b)\rangle^2\big] \\ &\qquad\le \max_{j\in[d]}\sqrt{\mathbb{E}\big[\langle e_j,\, \phi(s_t)\rangle^4\big]}\cdot\sqrt{\big(\mathbb{E}\big[\phi(s_t)^\top\bar{\theta} - \gamma\phi(s_{t+1})^\top\bar{\theta} - R_t(s_t)\big)^4\big]} \le \bar{\sigma}^2.\end{aligned}$$

Invoking equation (72) with the test matrix $Q := B$ and substituting with the representation $V(s) = \langle\theta,\, \phi(s)\rangle$ yields the claim.

E.2 Proof of Corollary 2

We prove this corollary by verifying the assumptions used in our main theorem. Assumption 2 directly follows from (33c) and the boundedness of reward; Assumption 1 is exactly the $\mathcal{W}_1$ mixing time bound imposed on the Markov chain. In order to verify that $\boldsymbol{L}(s, s^+) = I_d - \beta^{-1}\big(\phi(s)\phi(s)^\top - \gamma\phi(s)\phi(s^+)^\top\big)$ satisfies Assumption 4, we first note that

$$\|\boldsymbol{L}(s_1, s_1^+) - \boldsymbol{L}(s_2, s_2^+)\|_{\mathrm{op}} \le \tfrac{1}{\beta}\|\phi(s_1)\phi(s_1)^\top - \phi(s_2)\phi(s_2)^\top\|_{\mathrm{op}} + \tfrac{\gamma}{\beta}\|\phi(s_1)\phi(s_1^+)^\top - \phi(s_2)\phi(s_2^+)^\top\|_{\mathrm{op}}.$$

By adding and subtracting terms, we have the bound

$$\begin{aligned}\|\phi(s_1)\phi(s_1)^\top - \phi(s_2)\phi(s_2)^\top\|_{\mathrm{op}} &\le \big\{\|\phi(s_1)\|_2 + \|\phi(s_2)\|_2\big\}\,\|\phi(s_1) - \phi(s_2)\|_2 \\ &\overset{(i)}{\le} 2\varsigma^2\beta d\|s_1 - s_2\|_2,\end{aligned}$$

The step (i) follows from the Lipschitz condition (33b) and boundedness of the metric space $\mathbb{X}$. More precisely, we have $\|\phi(s_1) - \phi(s_2)\|_2 \le \varsigma\sqrt{\beta d}\|s_1 - s_2\|_2$ and $\|\phi(s_1)\|_2 = \|\phi(s_1) - \phi(0)\|_2 \le \varsigma\sqrt{\beta d}$. A similar argument yields that

$$\|\phi(s_1)\phi(s_1^+)^\top - \phi(s_2)\phi(s_2^+)^\top\|_{\mathrm{op}} \le \varsigma^2 d\big(\|s_1^+ - s_2^+\|_2 + \|s_1 - s_2\|_2\big).$$

Putting together the pieces, we have shown that the mapping $L : \mathbb{X} \to \mathbb{R}^{d\times d}$ is $3\varsigma^2 d$-Lipschitz with respect to the metric $\rho\big((s_1, s_1^+), (s_2, s_2^+)\big) = \|s_1 - s_2\|_2 + \|s_1^+ - s_2^+\|_2$.

Similarly, for the vector observation $\boldsymbol{b}_t(s) = R_t(s)\phi(s)$, we note that for any $s_1, s_2 \in \mathbb{X}$,

$$\begin{aligned}\|\boldsymbol{b}_t(s_1) - \boldsymbol{b}_t(s_2)\|_2 &\le |R_t(s_1) - R_t(s_2)| \cdot \|\phi(s_1)\|_2 + |R_t(s_2)| \cdot \|\phi(s_1) - \phi(s_2)\|_2 \\ &\le 2\varsigma\sqrt{d/\beta}\|\phi(s_1) - \phi(s_2)\|_2,\end{aligned}$$

which shows that $b : \mathbb{X} \to \mathbb{R}^{d/\beta}$ is $2\varsigma^2\sqrt{d}$-Lipschitz. Having verified the assumptions, we complete the proof by following the same steps as in the proof as Corollary 1.

E.3 Proof of Corollary 3

In order to verify that Assumption 4 holds with respect to the discrete metric, note that for any $d_n \ge 1$, we have $\|\boldsymbol{b}_t(s)\|_2 \le \frac{\varsigma}{\beta}\sqrt{\sum_{j=1}^{d_n}\phi_j^2(s)} \le \frac{\varsigma^2}{\beta}\sqrt{d_n}$, and

$$\|\boldsymbol{L}(s_1, s_2)\|_{\mathrm{op}} \le 1 + \tfrac{1}{\beta}\sum_{j=1}^{d_n}\phi_j^2(s_1) + \tfrac{1}{\beta}\sqrt{\sum_{j=1}^{d_n}\phi_j^2(s_1)}\cdot\sqrt{\sum_{j=1}^{d_n}\phi_j^2(s_2)} \le \tfrac{1+\varsigma^2}{\beta}d_n.$$

Turning to the moment condition, let $\mathbb{E}$ denote expectation over a pair $s \sim \xi$ and $s^+ \sim P(s, \cdot)$. Then for any vector $u \in \mathbb{S}^{d_n-1}$ and index $j \in [d_n]$, we have

$$\begin{aligned}\mathbb{E}\big[\langle e_j,\, \boldsymbol{L}(s, s^+)u\rangle^2\big] &\le 3 + \tfrac{3}{\beta^2}\mathbb{E}\Big[\big(\langle e_j,\, \phi(s)\rangle\,\langle\phi(s^+),\, u\rangle\big)^2\Big] + \tfrac{3}{\beta^2}\mathbb{E}\Big[\big(\langle e_j,\, \phi(s)\rangle\,\langle\phi(s),\, u\rangle\big)^2\Big] \\ &\le 3 + \tfrac{6}{\beta^2}\|\phi_j\|_\infty^2 \cdot \mathbb{E}\big[\langle\phi(s),\, u\rangle^2\big] \\ &\le 3 + \tfrac{6}{\beta}\varsigma^2.\end{aligned}$$

For each $t = 1, 2, \ldots$, we also have $\mathbb{E}\big[\langle e_j,\, \boldsymbol{b}_{t+1}(s_t)\rangle^2\big] \le \frac{1}{\beta^2}\|R_t\|_\infty^2 \cdot \mathbb{E}_{s\sim\xi}\big[\phi_j(s)^2\big] \le \frac{\varsigma^2}{\beta}$, which is an order-one quantity. Following the same steps as in the proof as Corollary 1 then yields the claim.

F Proofs for TD(λ)

We first prove Proposition 2—the mixing time result—and then use it to establish Corollary 4.

F.1 Proof of Proposition 2

We prove the claim via a coupling argument. Consider two initial states $\omega_0 = (s_0, s_1, h_0)$ and $\omega_0' = (s_0', s_1', h_1')$. By Assumption 1 (mixing time) for the original chain in total variation distance, there exists a coupling between a chains $(s_t)_{t\geq 1}$ and $(s_t')_{t\geq 1}$ starting from s_1 and s_1' respectively, such that $\mathbb{P}\big(s_{(k+1)t_{\mathrm{mix}}+1} \neq s'_{(k+1)t_{\mathrm{mix}}+1} \mid \{s_t, s_t'\}_{t=1}^{kt_{\mathrm{mix}}+1}\big) \leq \frac{1}{2}$. Furthermore, whenever $s_t = s_t'$ for some $t \geq 1$, the two processes are always identical from then on. Let $(g_t)_{t\geq 0}$ and $(g_t')_{t\geq 0}$ be the eligibility trace process (37b) associated to $(s_t)_{t\geq 0}$ and $(s_t')_{t\geq 0}$, respectively, and let $h_t = \frac{1-\lambda\gamma}{\varsigma\sqrt{\beta d}} g_t$ and $h_t' = \frac{1-\lambda\gamma}{\varsigma\sqrt{\beta d}} g_t'$.

Under this coupling, we note that $\mathbb{P}\big(s_{3t_{\mathrm{mix}}+1} \neq s'_{3t_{\mathrm{mix}}+1}\big) \leq \frac{1}{8}$. Conditioning on the event $\mathscr{E} := \big\{s_{3t_{\mathrm{mix}}+1} = s'_{3t_{\mathrm{mix}}+1}\big\}$, for any $t \geq 3t_{\mathrm{mix}} + 1$, we have

$$\|h_{t+1} - h'_{t+1}\|_2 = \gamma\lambda \|h_t - h_t'\|_2 = \cdots = (\gamma\lambda)^{t-3t_{\mathrm{mix}}-1} \|h_{3t_{\mathrm{mix}}+1} - h'_{3t_{\mathrm{mix}}+1}\|_2. \tag{108}$$

We split the remainder of the proof into two cases.

Case I: $s_1 \neq s_1'$**:** The coupling bound implies that $\mathbb{P}(\mathscr{E}) \geq \frac{7}{8}$. On the event $\mathscr{E}$, for $\tau \geq 3t_{\mathrm{mix}} + 1 + \frac{4}{1-\gamma\lambda}$, we have the bound $\|h_{t+1} - h'_{t+1}\|_2 \leq \frac{1}{16}\|h_{3t_{\mathrm{mix}}+1} - h'_{3t_{\mathrm{mix}}+1}\|_2 \leq \frac{1}{8}$ almost surely. Under this coupling, we may write

$$\begin{aligned}
\mathbb{E}\big[\rho\big((s_\tau, s_{\tau+1}, h_\tau), (s'_\tau, s'_{\tau+1}, h_\tau)\big)\big] &= \tfrac{1}{4}\big(\mathbb{P}\big(s_\tau \neq s'_\tau\big) + \mathbb{P}\big(s_{\tau+1} \neq s'_{\tau+1}\big) + \mathbb{E}\big[\|h_\tau - h'_\tau\|_2\big]\big) \\
&\leq \tfrac{3}{4}\mathbb{P}(\mathscr{E}^c) + \tfrac{1}{4}\mathbb{E}\big[\|h_\tau - h'_\tau\|_2 \mid \mathscr{E}\big] \\
&\leq \tfrac{1}{8} = \tfrac{1}{2} \cdot \tfrac{1}{4}\mathbf{1}_{s_1 \neq s_1'} \leq \tfrac{1}{2}\rho\big((s_0, s_1, h_0), (s_0', s_1', h_0)\big),
\end{aligned}$$

which proves the Wasserstein contraction in this case.

Case II: $s_1 = s_1'$ In this case, the coupling construction ensures that $s_t = s_t'$ for any $t \geq 1$. Invoking the bound (108) then yields

$$\mathbb{E}\big[\rho\big((s_\tau, s_{\tau+1}, h_\tau), (s'_\tau, s'_{\tau+1}, h_\tau)\big)\big] = \tfrac{1}{4}\mathbb{E}\big[\|h_\tau - h'_\tau\|_2\big] \leq \tfrac{1}{8}\|h_0 - h_0'\|_2 \leq \tfrac{1}{2}\rho(\omega_0, \omega_0'),$$

which establishes contraction in this case. Combining the two cases proves the proposition.

F.2 Proof of Corollary 4

We note that the SA procedure (37a) can be written as

$$\theta_{t+1} = (1 - \eta\beta)\theta_t + \eta\beta \boldsymbol{L}_{t+1}(\omega_t)\theta_t - \eta\beta \boldsymbol{b}_{t+1}(\omega_t),$$

where $\boldsymbol{L}_{t+1}(\omega_t) = \big(I_d - \frac{1}{\beta} g_t \phi(s_t)^\top + \gamma \frac{1}{\beta} g_t \phi(s_{t+1})^\top\big)$ and $\boldsymbol{b}_{t+1}(\omega_t) = \frac{1}{\beta} R_t(s_t) g_t$.

Recalling that $M_\lambda = (1-\lambda)\gamma \sum_{t=0}^\infty \lambda^t \gamma^{t+1} B^{-1/2} \mathbb{E}\big[\phi(s_0)\phi(s_{t+1})^\top\big] B^{-1/2}$, we first study the eigenvalues of the symmetrized version of M_λ, and relate these back to those of $\bar{L} = \mathbb{E}_{\widetilde{\xi}}\big[\boldsymbol{L}_{t+1}(\omega_t)\big]$. Note that by the Cauchy–Schwarz inequality, for any vector $u \in \mathbb{S}^{d-1}$, we have

$$u^\top B^{-1/2} \mathbb{E}\big[\phi(s_0)\phi(s_t)^\top\big] B^{-1/2} u \leq \sqrt{\mathbb{E}\big[\big(u^\top B^{-1/2}\phi(s_0)\big)^2\big] \cdot \mathbb{E}\big[\big(u^\top B^{-1/2}\phi(s_t)\big)^2\big]} = 1.$$

We therefore have the bound $\frac{1}{2}\lambda_{\min}(M_\lambda + M_\lambda^\top) \le (1-\lambda)\gamma\sum_{t=0}^\infty(\gamma\lambda)^t = \frac{(1-\lambda)\gamma}{1-\lambda\gamma}$. As in the proof of Corollary 1, we can deduce that

$$\tfrac{1}{2}\lambda_{\max}\big(\bar{L}+\bar{L}^\top\big) = \tfrac{1}{\beta}\lambda_{\max}\big(B^{1/2}\big(\tfrac{M_\lambda+M_\lambda^\top}{2}\big)B^{1/2}\big) \ge \tfrac{(1-\lambda)\gamma}{1-\lambda\gamma}.$$

Next, we verify Assumption 2 on the noise moments. By the update rule (37b), under a stationary trajectory, we have the expression $g_t = \sum_{k=0}^\infty(\gamma\lambda)^k\phi(s_{t-k})$. For any $u \in \mathbb{S}^{d-1}$, invoking Hölder's inequality yields

$$\mathbb{E}\big[\langle g_t,\, u\rangle^4\big] \le \big(\sum_{k=0}^\infty(\gamma\lambda)^k\big)^3 \cdot \sum_{k=0}^\infty(\gamma\lambda)^k\mathbb{E}\big[\langle u,\, \phi(s_{t-k})\rangle^4\big] \le \beta^2\big(\tfrac{\varsigma}{1-\gamma\lambda}\big)^4.$$

In other words, for all standard basis vectors e_j, we have

$$\begin{aligned}
\mathbb{E}\big[\langle e_j,\, \boldsymbol{L}_{t+1}(\omega_t)u\rangle^2\big] &\le 1 + \tfrac{2}{\beta^2}\sqrt{\mathbb{E}\big[\langle e_j,\, \phi(s_t)\rangle^4\big]}\cdot\sqrt{\mathbb{E}\big[\langle g_t,\, u\rangle^4\big]} \le 1 + 2\tfrac{\varsigma^4}{(1-\gamma\lambda)^2},\\
\mathbb{E}\big[\langle e_j,\, \boldsymbol{b}_{t+1}(\omega_t)u\rangle^2\big] &\le \tfrac{\varsigma^2}{\beta^2}\mathbb{E}\big[\langle g_t,\, e_j\rangle^2\big] \le \tfrac{\varsigma^4}{\beta(1-\gamma\lambda)^2}.
\end{aligned}$$

It remains to verify Assumption 4. Note that for any pair $\omega = (s, s_+, h)$ and $\omega' = (s', s'_+, h')$, the operator norm $T := \|\boldsymbol{L}_{t+1}(\omega) - \boldsymbol{L}_{t+1}(\omega')\|_{\mathrm{op}}$ is almost surely upper bounded as

$$\begin{aligned}
T &\le \tfrac{\varsigma\sqrt{d/\beta}}{1-\lambda\gamma}\cdot\big(\|h^\top\phi(s) - (h')^\top\phi(s')\|_{\mathrm{op}} + \|h^\top\phi(s_+) - (h')^\top\phi(s'_+)\|_{\mathrm{op}}\big)\\
&\le \tfrac{\varsigma\sqrt{d/\beta}}{1-\lambda\gamma}\cdot\big(\|(h-h')^\top\phi(s')\|_{\mathrm{op}} + \|h^\top(\phi(s') - \phi(s))\|_{\mathrm{op}} + \|(h-h')^\top\phi(s'_+)\|_{\mathrm{op}} + \|h^\top(\phi(s'_+) - \phi(s_+))\|_{\mathrm{op}}\big)\\
&\le \tfrac{2\varsigma^2 d}{1-\lambda\gamma}\big(\mathbf{1}_{s\neq s'} + \mathbf{1}_{s_+\neq s'_+} + \|h-h'\|_2\big) = \tfrac{8\varsigma^2 d}{1-\lambda\gamma}\rho(\omega,\omega').
\end{aligned}$$

Finally, we note that the quantity $\bar{\sigma}$ defined in equation (29) satisfies the bound

$$\begin{aligned}
&\sup_{j\in[d]}\mathbb{E}\big[\langle e_j,\, (\boldsymbol{L}_{t+1}(\omega_t) - \bar{L})\bar{\theta} + (\boldsymbol{b}_{t+1}(\omega_t) - b)\rangle^2\big]\\
&\le \sup_{j\in[d]}\sqrt{\mathbb{E}\big[\langle e_j,\, g_t\rangle^4\big]}\cdot\sqrt{\big(\mathbb{E}\big[\phi(s_t)^\top\bar{\theta} - \gamma\phi(s_{t+1})^\top\bar{\theta} - R_t(s_t)\big)^4\big]} \le \tfrac{\bar{\sigma}^2}{(1-\gamma\lambda)^2}.
\end{aligned}$$

Invoking equation (72), with the test matrix $Q := B$ and substituting the expression $V(s) = \langle\theta,\, \phi(s)\rangle$ yields the claim.

G Proofs for vector autoregressive estimation

In this section, we present proofs of results on vector autoregressive models, as introduced in Example 3.

G.1 Proof of Proposition 3

We prove the claim by a direct construction of the coupling. Given two initial points $\omega_0 = \big[X_1^\top, X_0^\top, \cdots, X_{-k+1}^\top\big]^\top$ and $\omega'_0 = \big[X_1'^\top, X_0'^\top, \cdots, X_{-k+1}'^\top\big]^\top$, we consider a pair of stochastic processes $(X_t)_{t\ge1}$ and $(X'_t)_{t\ge1}$ starting from ω_0 and ω', respectively, driven by the same noise process $(\varepsilon_t)_{t\ge0}$. Introduce the shorthand $Y_{t+1} = \big[X_{t+1} \;\; \ldots \;\; X_{t-k+2}\big]^\top$ (note that Y_{t+1} is a sliding window with length one unit shorter than ω_t). We have:

$$\begin{aligned}
\|Y_{t+1} - Y'_{t+1}\|_{P_*}^2 \;=\; \|R_*\,(Y_t - Y'_t)\|_{P_*}^2 &= \|Y_t - Y'_t\|_{P_*}^2 - \|Y_t - Y'_t\|_{Q_*}^2\\
&\le \big(1 - \tfrac{\mu}{\beta}\big)\|Y_t - Y'_t\|_{P_*}^2.
\end{aligned}$$

Consequently, the augmented processes $\omega_t = (X_{t+1}, X_t, \cdots, X_{t-k+1})$ and $\omega'_t = (X'_{t+1}, X'_t, \cdots, X'_{t-k+1})$ satisfy the bound

$$\begin{aligned}\|\omega_t - \omega'_t\|_2 \le \|Y_{t+1} - Y'_{t+1}\|_2 + \|Y_t - Y'_t\|_2 &\le \frac{1}{\sqrt{\lambda_{\min}(P_*)}}\big(\|Y_{t+1} - Y'_{t+1}\|_{P_*} + \|Y_t - Y'_t\|_{P_*}\big) \\ &\le 2\sqrt{\frac{\lambda_{\max}(P_*)}{\lambda_{\min}(P_*)}}\big(1 - \tfrac{\mu}{2\beta}\big)^t \|\omega_0 - \omega'_0\|_2\end{aligned}$$

Note that since $P_* \succeq Q_*$, we have $\lambda_{\min}(P_*) \ge \lambda_{\min}(Q_*) = \mu$. Taking $t_{\mix} = c\frac{\beta}{\mu}\big(1 + \log\frac{\beta}{\mu}\big)$ yields the contraction bound $\|\omega_{t_{\mix}} - \omega'_{t_{\mix}}\|_2 \le \frac{1}{2}\|\omega_0 - \omega'_0\|_2$. Taking expectations on both sides completes the proof.

G.2 Proof of Corollary 5

We begin by showing norm bounds and moment bounds on the process $(X_t)_{t\ge 0}$. By definition (17) of the process and stability, the block vector $Y_t := \begin{bmatrix} X_t & X_{t-1} & \cdots & X_{t-k+1}\end{bmatrix}^\top$ satisfies the recursion $Y_t = \sum_{i=0}^\infty R_*^i \varepsilon_{t-i} e_1$, where e_1 is the standard block basis vector equal to identify on the first block. We therefore have the bound

$$\|X_t\|_2 \le \tfrac{1}{\mu}\|Y_t\|_{P_*} \le \sum_{i=0}^\infty \|R_*^i \varepsilon_{t-i} e_1\|_{P_*} \le \tfrac{1}{\mu}\sum_{i=0}^\infty \big(1 - \tfrac{\mu}{\beta}\big)^i \|\varepsilon_{t-i} e_1\|_{P_*} \le \tfrac{\beta^2}{\mu^2}\varsigma\sqrt{m}.$$

Moreover, for each $u \in \mathbb{S}^{m-1}$, we have

$$\begin{aligned}\mathbb{E}\big[\langle X_t,\, u\rangle^4\big] &\le \big(\sum_{i=0}^\infty e^{-\frac{i\mu}{6\beta}}\big)^3 \cdot \sum_{i=0}^\infty e^{\frac{i\mu}{2\beta}}\, \mathbb{E}\big[\langle R_*^i \varepsilon_{t-i} e_1,\, u e_1\rangle^4\big] \\ &\le c\big(\beta/\mu\big)^3 \cdot \sum_{i=0}^\infty e^{\frac{i\mu}{2\beta}} \cdot \tfrac{\beta^4}{\mu^4} \cdot e^{-\frac{i\mu}{\beta}} \varsigma^4 \;\le\; c'\big(\tfrac{\beta^2\varsigma}{\mu^2}\big)^4.\end{aligned}$$

Next, we proceed with verifying the assumptions used in Theorem 1. Letting $\nu := 1/\|\!|H^*|\!\|_{\op}$, the stochastic approximation procedure can be rewritten as

$$\theta_{t+1} = (1 - \tfrac{\eta}{\nu})\theta_t + \tfrac{\eta}{\nu}\big(\theta_t - \nu\big(\big[X_{t-j}X_{t+1-i}^\top\big]_{i,j\in[m]} \otimes I_m\big)\theta_t + \nu \cdot \vec{}\big(\begin{bmatrix} X_{t+1}X_t^\top & \cdots & X_{t+1}X_{t-k+1}^\top\end{bmatrix}\big)\big).$$

Observe that the matrix $\bar{L} := I_{km^2} - \nu H^* \otimes I_m$ satisfies the eigenvalue bound

$$\tfrac{1}{2}\lambda_{\max}(\bar{L} + \bar{L}^\top) \le 1 - \tfrac{\nu}{2}\lambda_{\min}(H^* + (H^*)^\top) \le 1 - \nu h^*.$$

On the other hand, the empirical observations satisfy the almost-sure bounds

$$\begin{aligned}\|\!|\boldsymbol{L}_{t+1}(\omega_t) - \bar{L}|\!\|_{\op} &\le \nu \cdot \|\!|\big[X_{t-j}X_{t+1-i}^\top\big]_{i,j\in[m]}|\!\|_{\op} \le \nu \cdot \tfrac{\beta^4}{\mu^4}\varsigma^2 mk \text{ and} \\ \|\!|\boldsymbol{b}_{t+1}(\omega_t) - \bar{b}|\!\|_{\op} &\le \nu \cdot \|\!|\begin{bmatrix} X_{t+1}X_t^\top & \cdots & X_{t+1}X_{t-k+1}^\top\end{bmatrix}|\!\|_F \le \nu \cdot \tfrac{\beta^4}{\mu^4}\varsigma^2 m\sqrt{k}.\end{aligned}$$

For two collections of matrices $\mathcal{U} = \big(U^{(j)}\big)_{j=1}^k$ and $\mathcal{V} = \big(V^{(j)}\big)_{j=1}^k \subseteq \mathbb{R}^{m\times m}$ such that $\sum_{j=1}^k \|\!|U^{(j)}|\!\|_F^2 = \sum_{j=1}^k \|\!|V^{(j)}|\!\|_F = 1$, the corresponding moment can be bounded as

$$\mathbb{E}\big[\langle \vec{}(\mathcal{U}),\, \big(\boldsymbol{L}_{t+1}(\omega_t) - \bar{L}\big)\vec{}(\mathcal{V})\rangle^2\big] \le \nu^2 \mathbb{E}\big[\big(\sum_{\ell=0}^{k-1}\langle U^{(\ell)},\, \sum_{j=0}^{k-1} V^{(j)} X_{t-j}X_{t-\ell}^\top\rangle_F\big)^2\big],$$

which is in turn at most

$$\nu^2 k^2 \sum_{\ell=0}^{k-1}\sum_{j=0}^{k-1} \sqrt{\mathbb{E}\big[X_{t-\ell}^{\otimes 4}\big]\big[(U^{(\ell)})^\top U^{(\ell)}, (U^{(\ell)})^\top U^{(\ell)}\big]} \cdot \sqrt{\mathbb{E}\big[X_{t-j}^{\otimes 4}\big]\big[(V^{(j)})^\top V^{(j)}, (V^{(j)})^\top V^{(j)}\big]}.$$

In order to bound this last quantity, we let $(U^{(\ell)})^\top U^{(\ell)} = \sum_{i=1}^m \lambda_i^2 u_i u_i^\top$ be its singular value decomposition, and note that

$$\begin{aligned}\mathbb{E}\big[X_{t-\ell}^{\otimes 4}\big]\big[(U^{(\ell)})^\top U^{(\ell)}, (U^{(\ell)})^\top U^{(\ell)}\big] &= \mathbb{E}\big[X_{t-\ell}^{\otimes 4}\big]\big[\sum_{i=1}^m \lambda_i^2 u_i u_i^\top, \sum_{i=1}^m \lambda_i^2 u_i u_i^\top\big] \\ &= \sum_{i,i'} \mathbb{E}\big[X_{t-\ell}^{\otimes 4}\big][u_i, u_i, u_{i'}, u_{i'}] \cdot \lambda_i^2 \lambda_{i'}^2 \le c'\big(\tfrac{\beta^2\varsigma}{\mu^2}\big)^4 \big(\sum_i \lambda_i^2\big)^2 = c'\big(\tfrac{\beta^2\varsigma}{\mu^2}\big)^4 |\!|\!| U^{(\ell)} |\!|\!|_F^2.\end{aligned}$$

Putting together the pieces, we have

$$\mathbb{E}\big[\langle \mathrm{vec}(\mathcal{U}),\, \big(\boldsymbol{L}_{t+1}(\omega_t) - \bar{L}\big)\mathrm{vec}(\mathcal{V})\rangle^2\big] \le \nu^2 k^2 c'\big(\tfrac{\beta^2\varsigma}{\mu^2}\big)^4 \cdot \sum_{\ell=0}^{k-1}\sum_{j=0}^{k-1} |\!|\!| U^{(\ell)} |\!|\!|_F^2 |\!|\!| V^{(j)} |\!|\!|_F^2 \le c\big(\nu \cdot \tfrac{\beta^4 k \varsigma^2}{\mu^4}\big)^2.$$

Similarly, we can prove analogous moment bounds on $\boldsymbol{b}_{t+1}(\omega_t)$. In particular, for indices $\ell \in [k]$ and $i, j \in [m]$, we consider the coordinate direction of the (i,j) entry in the ℓ-th matrix to deduce that

$$\begin{aligned}\mathbb{E}\big[\langle e_{\ell,i,j},\, (\boldsymbol{b}_{t+1}(\omega_t) - \bar{b})\rangle^2\big] &\le \nu^2 \mathbb{E}\big[\langle e_i e_j^\top,\, X_{t+1} X_{t-\ell+1}\rangle^2\big] \\ &\le \nu^2 \sqrt{\mathbb{E}\big[\langle e_j^\top,\, X_{t+1}\rangle^4\big]} \cdot \sqrt{\mathbb{E}\big[\langle e_i^\top,\, X_{t-\ell+1}\rangle^4\big]} \le c'\big(\nu \cdot \tfrac{\beta^2\varsigma}{\mu^2}\big)^4.\end{aligned}$$

Applying Theorem 1 completes the proof of this corollary.