Math. Stat. Learn. 7 (2024), 41–153
DOI 10.4171/MSL/44

Optimal and instance-dependent guarantees for Markovian linear stochastic approximation

Wenlong Mou, Ashwin Pananjady, Martin J. Wainwright, and Peter L. Bartlett

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—arise in many application domains, including reinforcement learning and approximate dynamic programming [4,73], stochastic control and filtering [2,7,44], and time-series analysis [32]. 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 is to 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.

Mathematics Subject Classification 2020: 62L20 (primary); 60J22, 62C20, 62M05, 93E35 (secondary).
Keywords: Markov chains, stochastic approximation, reinforcement learning, temporal difference methods, instance-dependent optimality.

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

$$\theta = \bar{L}\theta + \bar{b}, \tag{1.1}$$

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+1}, b_{t+1}) 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

$$\mathbb{E}_\xi\big[\boldsymbol{L}(s)\big] = \bar{L} \quad \text{and} \quad \mathbb{E}_\xi\big[\boldsymbol{b}(s)\big] = \bar{b}. \tag{1.2a}$$

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

$$L_{t+1} = \boldsymbol{L}_{t+1}(s_t) \quad \text{and} \quad b_{t+1} = \boldsymbol{b}_{t+1}(s_t). \tag{1.2b}$$

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

$$\mathbb{E}\big[\boldsymbol{L}_{t+1}(s)\big] = \boldsymbol{L}(s), \quad \mathbb{E}\big[\boldsymbol{b}_{t+1}(s)\big] = \boldsymbol{b}(s) \quad \text{for all } s \in \mathbb{X}. \tag{1.2c}$$

By the tower property of conditional expectation, for a stationary Markov chain, equations (1.2a) and (1.2c) 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] The random operator observed at each iteration is therefore given by $\theta \mapsto \boldsymbol{L}_{t+1}(s_t)\theta + \boldsymbol{b}_{t+1}(s_t)$. This is a natural generalization of "random field noise" [21,87] to the Markovian setting: instead of observing i.i.d. random fields at iteration, we observe random functional of a Markov chain's states.

Stochastic approximation (SA) methods, dating back to the seminal work of Robbins and Monro [66], are standard iterative procedures for using data to approximately

[1] However, equation (1.2c) does not require the observations to be conditionally unbiased.

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. To motivate the SA updates, we could start by considering a stochastic version of the fixed-point iteration to solve equation (1.1):

$$\theta_{t+1} = L_{t+1}\theta_t + b_{t+1}.$$

With the randomness of the observations (L_{t+1}, b_{t+1}), the iterates will fluctuate at a constant order and may not converge. In order to stabilize the stochastic fixed-point iteration, a stepsize $\eta \in (0, 1)$ may be introduced, leading to the canonical SA updates.

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

$$\theta_{t+1} := (1-\eta)\theta_t + \eta(L_{t+1}\theta_t + b_{t+1}) \quad \text{for } t = 0, 1, \ldots, \tag{1.3a}$$

$$\widehat{\theta}_n := \frac{1}{n-n_0}\sum_{t=n_0}^{n-1}\theta_t \quad \text{for } n = n_0+1, n_0+2, \ldots. \tag{1.3b}$$

Equation (1.3a) describes a standard stochastic approximation update with constant stepsize $\eta > 0$, whereas equation (1.3b) corresponds to an application of the Polyak–Ruppert averaging procedure [64, 68] 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 (1.3) 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 Appendix A. Asymptotic convergence of SA procedures with Markovian data can be established using either the ODE method [7] or the Poisson equation method [2]. Tsitsiklis and Van Roy [75] analyze the asymptotic convergence of SA in the specific context of temporal difference methods in reinforcement learning. Although asymptotic guarantees provide helpful guidance, it is often most useful to have non-asymptotic guarantees that account for both 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 in the paper [5], leading to non-asymptotic rates that are near-optimal in their dependence on the sample size n. Srikant and Ying [70] analyzed the standard SA scheme without the projection step used in the paper [5] 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. [24] 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 [36], gradient-based optimization under Markov data [22], and estimation in autoregressive models [13, 34].

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

- *Sample complexity with optimal dimension dependence:* 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 that 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 an effective sample size of the order n/t_{mix} so that the mean-squared error should scale as $\mathcal{O}(t_{\text{mix}} d/n)$, with this being the dominant term whenever $n \gtrsim t_{\text{mix}} d$. Such an error bound is particularly important for sieve estimators, where the problem dimension d is adaptively chosen based on the sample size n. However, existing analyses of linear SA do not provide such tight dimension dependence. Using the notation of equation (1.3a), the estimation error bounds in the papers [5, 70] 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 problems with growing dimensions. Similarly, the bounds in the papers [17, 24, 42] 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 $t_{\text{mix}} \cdot d$—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 [39, 50, 58, 62]. The latter two papers listed provide non-asymptotic analogs of classical theory on local asymptotic minimaxity (cf. [78]), 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 (1.3) 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. Arguably, our results are not merely of theoretical interest; they also provide important guidance for practice, such as in choosing algorithm parameters including the burn-in period and stepsize. In more detail, we consider the following.

- We perform a fine-grained analysis of linear SA and produce an upper bound on its statistical error that explicitly tracks the dependence on problem-specific complexity as well as stepsize. Furthermore, our bound holds true provided $n \gtrsim t_{\mathrm{mix}} \cdot d$, establishing the fact that the algorithm does indeed attain a sharp sample complexity guarantee with optimal dimension dependence.
- 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 setup, 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 and 6 are devoted to the proofs of Proposition 1 and Theorem 1, respectively. We conclude with a discussion in Section 7. The proof of Theorem 2 and some auxiliary results, as well as some 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

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

and the total variation distance between π and μ is given by

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

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

$$\text{KL divergence:} \quad D_{\mathrm{KL}}(P \parallel Q) := \mathbb{E}_P\left[\log \frac{dP}{dQ}(X)\right],$$

$$\chi^2 \text{ divergence:} \quad \chi^2(P \,||\, Q) := \mathbb{E}_P\left[\frac{dP}{dQ}(X) - 1\right],$$

$$\text{Max divergence:} \quad D_\infty(P||Q) := \sup_{x \in \mathrm{supp}(Q)} \left|\log \frac{dP}{dQ}(x)\right|.$$

Given any matrix $A = (a_{ij}) \in \mathbb{R}^{n\times m}$, its vectorization is obtained by concatenating its columns—viz.

$$\mathrm{vec}(A) := \begin{bmatrix} a_{11} & a_{21} & \cdots & a_{n1} & a_{12} & \cdots & a_{n2} & \cdots & a_{1m} & \cdots & a_{nm} \end{bmatrix}^\top \in \mathbb{R}^{nm}.$$

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 $d_1d_3 \times d_2d_4$ 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 eigenvalues 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

$$\mathcal{F}_t := \sigma((b_i, L_i, s_i)_{i\leq t})$$

to denote the natural filtration induced by the Markovian observations.

2. Problem setup

Recall from our earlier setup (cf. equation (1.1)) 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 (1.2). We require that the matrix $\bar{L}$ satisfies the conditions

$$\kappa := \frac{1}{2}\lambda_{\max}\big(\bar{L} + \bar{L}^\top\big) < 1 \quad \text{and} \quad \|\bar{L}\|_{\text{op}} \leq \gamma_{\max}. \tag{2.1}$$

This condition is used throughout the paper.

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 exist a natural number t_{mix} and a universal constant $c_0 > 0$ such that, for any $x, y \in \mathbb{X}$, we have*

$$\mathcal{W}_{1,\rho}(\delta_x P^{t_{\text{mix}}}, \delta_y P^{t_{\text{mix}}}) \overset{(a)}{\leq} \frac{1}{2}\rho(x, y) \quad \textit{and} \quad \mathcal{W}_{1,\rho}(\delta_x P^{t}, \delta_y P^{t}) \overset{(b)}{\leq} c_0\rho(x, y) \tag{2.2}$$

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

It is known that such a condition implies rapid mixing (see [48, Section 4.5]). For most parts of the paper, we assume 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 for $\mathcal{O}(t_{\text{mix}})$ iterations for the process to mix. By making this intuition rigorous, we will also present a slightly weaker error bound under arbitrary initial distribution (see Corollary 1).

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 (i.e., $\boldsymbol{L}_{t+1}(s_t) - \boldsymbol{L}(s_t)$ and $\boldsymbol{b}_{t+1}(s_t) - \boldsymbol{b}(s_t)$) and Markov (i.e., $\boldsymbol{L}(s_t) - \bar{L}$ and $\boldsymbol{b}(s_t) - \bar{b}$) parts of the noise.

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

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

as well as

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

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 $\log(p!) \le p\log(p/2)$ 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 [58].

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

$$\bar{\sigma} := \sup_{p\in[2,\bar{p}]} \sup_{j\in[d]} \sup_{t\ge 0} p^{-1}\Big(\mathbb{E}\big[\langle e_j, (\boldsymbol{L}_{t+1}(s_t) - \bar{L})\bar{\theta} + (\boldsymbol{b}_{t+1}(s_t) - \bar{b})\rangle^p\big]\Big)^{1/p}. \quad (2.3)$$

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

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 B for the details of this argument.) To unify the notation, we always assume the distance to be of constant order with high probability, which results in constant diameter of the truncated space. In high-dimensional Euclidean spaces, the distance between two generic random vectors can easily become dimension dependent. In such cases, we rescale the space to make it a constant. The rescaling could lead to dimension-dependent Lipschitz constants, which is captured in Assumption 4 to follow.

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 tempting to assume a dimension-free Lipschitz bound on the mappings $(\boldsymbol{L}_t, \boldsymbol{b}_t)$. However, 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

$$\mathbb{E}\big[\|\boldsymbol{L}_{t+1}(s_t)\|_{\mathrm{op}}\big] \geq \frac{\mathbb{E}\big[\|\boldsymbol{L}_{t+1}(s_t)\|_{\mathrm{nuc}}\big]}{r} \geq \frac{\operatorname{trace}\big(\mathbb{E}\big[\boldsymbol{L}_{t+1}(s_t)\big]\big)}{r} = \frac{\operatorname{trace}(\bar{L})}{r}. \quad (2.4)$$

Note that the term $\operatorname{trace}(\bar{L})$ typically scales as $\Theta(d)$, even in the "easy case" when $\bar{L}$ is a constant multiple of identity matrix.

Consequently, the Lipschitz constant for the mapping

$$\boldsymbol{L}_t : \mathbb{X} \to \mathbb{R}^{d \times d}$$

grows at least linearly in dimension d. On the other hand, as a d-dimensional standard Gaussian random variable has norm $\sqrt{d} - \widetilde{\mathcal{O}}(1)$ 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*

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

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

Note that, in Assumption 4, we have explicitly scaled the right-hand side 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. It is also worth noting that the notation (σ_L, σ_b) is overloaded, since we can take the maximum of the bounds in Assumptions 2 and 4. As shown in Appendix B, 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 (1.1) 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.

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

  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 $\overline{\mathcal{T}} := \mathbb{E}[\mathcal{T}_1]$ satisfies the contraction condition $\|\overline{\mathcal{T}}(x) - \overline{\mathcal{T}}(y)\|_2 \leq \gamma \|x - y\|_2$ 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 temporal difference (TD) 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*

$$V^*(s) = \mathbb{E}\left[\sum_{t=0}^{\infty} \gamma^t R_t \mid s_0 = s \right].$$

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*

$$V = \gamma P V + r. \tag{2.5}$$

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 [12, 75], and this approach underlies the family of TD algorithms with linear function approximation.

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*

$$\bar{V} = \Pi_{\mathbb{S}}\big(\gamma P \bar{V} + r\big), \tag{2.6}$$

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 $\bar{\theta} \in \mathbb{R}^d$. In defining the vector-valued mapping $\phi = [\phi_j]_{j=1}^d$, some simple calculations show that this parameter vector must satisfy the linear system

$$\Sigma_0 \bar{\theta} = \gamma \Sigma_1 \bar{\theta} + \mathbb{E}_{s\sim\xi}\big[R_0(s)\phi(s)\big], \tag{2.7}$$

where $\Sigma_0 = \mathbb{E}_{s\sim\xi}[\phi(s)\phi(s)^\top]$ 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 (2.7) 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

$$\begin{aligned} \boldsymbol{b}_{t+1}(\omega_t) &= \frac{1}{\beta} R_t(s_t)\phi(s_t), \\ \boldsymbol{L}_{t+1}(\omega_t) &= I_d - \frac{1}{\beta}\big[\phi(s_t)\phi(s_t)^\top - \gamma\phi(s_t)\phi(s_{t+1})^\top\big]. \end{aligned} \tag{2.8}$$

With these choices, the stochastic approximation procedure (1.3) 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 $\bar{\theta} = \mathbb{E}[\boldsymbol{L}_{t+1}(\omega_t)] \cdot \bar{\theta} + \mathbb{E}[\boldsymbol{b}_{t+1}(\omega_t)]$ is equivalent to equation (2.7). 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 (2.6) approximates the true value function V^*. Prior work by a subset of the current authors [58] 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 equation (2.5), 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 $(1, \lambda, \lambda^2, \ldots)$ and obtain the following equation:

$$V = (1-\lambda) \sum_{k=0}^{\infty} \lambda^k (\gamma P)^{k+1} V + \sum_{k=0}^{\infty} \lambda^k (\gamma P)^k r. \tag{2.9a}$$

The solution V^* to equation (2.5) also solves equation (2.9a).

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

$$\bar{V}^{(\lambda)} = (1-\lambda) \sum_{k=0}^{\infty} \lambda^k \Pi_{\mathbb{S}} (\gamma P)^{k+1} \bar{V}^{(\lambda)} + \sum_{k=0}^{\infty} \lambda^k \Pi_{\mathbb{S}} (\gamma P)^k r, \tag{2.9b}$$

in which the operator P has been replaced by the projection $\Pi_{\mathbb{S}} P$. Although the fixed points of equation (2.9a) and the Bellman equation (2.5) coincide, the projected version (2.9b) has a different set of fixed points.

Since the value function $\overline{V}^{(\lambda)}$ lies in the linear space $\mathbb{S}$, it has a representation of the form $\overline{V}^{(\lambda)}(s) = \sum_{j=1}^{d} \overline{\theta}_j^{(\lambda)} \phi_j(s)$ for some coefficient vector $\overline{\theta}^{(\lambda)} \in \mathbb{R}^d$. From equation (2.9b), this vector must satisfy a linear system of the form

$$\left[\sum_{k=0}^{\infty} (\lambda\gamma)^k \Sigma_k\right] \overline{\theta}^{(\lambda)} = \left[\sum_{k=0}^{\infty} (\lambda\gamma)^k \gamma \Sigma_{k+1}\right] \overline{\theta}^{(\lambda)} + \sum_{k=0}^{\infty} (\lambda\gamma)^k \mathbb{E}\big[R_0(s_0)\phi(s_{-k})\big], \tag{2.10}$$

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 (2.10) reduces to the TD(0) update from equation (2.7).

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

$$s_{t+1} \sim P(s_t, \cdot) \quad \text{and} \quad g_t = \phi(s_t) + \gamma\lambda g_{t-1}. \tag{2.11a}$$

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.[2] In terms of an element $\omega = (s, s^+, g)$ drawn according this stationary distribution, the fixed-point equation (2.9b) admits the succinct representation

$$\mathbb{E}_{\widetilde{\xi}}\big[g\phi(s)^\top\big]\overline{\theta}^{(\lambda)} = \gamma\mathbb{E}_{\widetilde{\xi}}\big[g\phi(s^+)^\top\big]\overline{\theta}^{(\lambda)} + \mathbb{E}_{\widetilde{\xi}}\big[R_0(s)g\big]. \tag{2.11b}$$

By choosing the observation functions

$$\boldsymbol{L}_{t+1}(\omega_t) = I_d - \nu \cdot \big(g_t\phi(s_t)^\top - \gamma g_t \phi(s_{t+1})^\top\big), \quad \boldsymbol{b}_{t+1}(\omega_t) = \nu \cdot R_t(s_t)\phi(s_t), \tag{2.11c}$$

for a scalar $\nu > 0$, this algorithm is a special case of our general setup. In particular, by substituting the infinite-sum expression for the random variable g_t into equation (2.11b), we obtain the projected linear equation (2.10) 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. ∎

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

For our last example, we turn to a different class of problems involving vector autoregressive (VAR) models for time series [55].

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 kth-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

$$X_{t+1} = \sum_{j=1}^{k} A_j^* X_{t-j+1} + \varepsilon_{t+1}, \tag{2.12}$$

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

Considering the $(k+1)$-fold tuple $\omega_t = (X_{t+1}, X_t, \ldots, X_{t-k+1})$, the process $(\omega_t)_{t\geq 0}$ 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.

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

$$\begin{aligned}&\mathbb{E}\big[X_{t+1}X_{t-\ell}^\top\big] \\ &\quad = A_1^*\mathbb{E}\big[X_t X_{t-\ell}^\top\big] + A_2^*\mathbb{E}\big[X_{t-1}X_{t-\ell}^\top\big] + \cdots + A_k^*\mathbb{E}\big[X_{t-k+1}X_{t-\ell}^\top\big]\end{aligned} \tag{2.13}$$

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

These equations form a km^2-dimensional linear system of equations 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

$$\big[\boldsymbol{b}_{t+1}(\omega_t)\big]_\ell = \nu\, X_{t+1}X_{t-\ell}^\top \quad \text{for } \ell = 0, 1, \ldots, k-1$$

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 (1.3b). 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, it is a random vector, whereas by contrast, the quantity $\varepsilon_{\mathrm{Mkv}}(s)$ is deterministic. Letting $(\tilde{s}_t)_{t=-\infty}^{\infty}$ be a stationary Markov chain under the transition kernel P, we then define the matrices

$$\Sigma^*_{\mathrm{MG}} := \mathbb{E}_\xi\big[\operatorname{cov}\big(\varepsilon_{\mathrm{MG}}(s) \mid s\big)\big] \quad \text{and} \quad \Sigma^*_{\mathrm{Mkv}} := \sum_{t=-\infty}^{\infty} \mathbb{E}\big[\varepsilon_{\mathrm{Mkv}}(\tilde{s}_t)\varepsilon_{\mathrm{Mkv}}(\tilde{s}_0)^\top\big]. \tag{3.1}$$

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 $\bar{\sigma}^2$ defined in equation (2.3). 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 $\eta = (c(\sigma_L^2 d + \gamma_{\max}^2)(1-\kappa)n^2 t_{\mathrm{mix}})^{-1/3}$ and $n_0 = \frac{1}{2}n$, where c is a suitably chosen universal constant. There exist universal constants $c_1, c_2 > 0$ such that, 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 (1.3b) has MSE bounded as*

$$\mathbb{E}\big[\|\widehat{\theta}_n - \bar{\theta}\|_2^2\big] \leq \frac{c_1}{n}\operatorname{Tr}\big((I-\bar{L})^{-1}(\Sigma^*_{\mathrm{MG}} + \Sigma^*_{\mathrm{Mkv}})(I-\bar{L})^{-\top}\big) + c_2\left(\frac{\sigma_L^2 d\, t_{\mathrm{mix}}}{(1-\kappa)^2 n}\right)^{4/3}\bar{\sigma}^2\log^2 n. \tag{3.2}$$

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 $n^{-1}\operatorname{Tr}((I-\bar{L})^{-1}\Sigma^*(I-\bar{L})^{-1})$ 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 [26], and our bound captures this dependence in a non-asymptotic manner up to a universal constant. It is worth noting that, when the Markov chain is uniformly geometrically ergodic, a central limit theorem for the averaged iterate $\widehat{\theta}_n$ directly follows from classical Markovian CLT (see [57, Chapter 17]).

The first term in the bound (3.2) can always be further upper bounded[3] by

$$c_1 \frac{\bar{\sigma}^2}{(1-\kappa)^2 n} t_{\mathrm{mix}} d \cdot \log^2(c_0 d).$$

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 $\mathcal{O}((\frac{t_{\mathrm{mix}} d}{(1-\kappa)^2 n})^{4/3})$. Consequently, up to polylogarithmic factors, we have

$$\mathbb{E}\big[\|\widehat{\theta}_n - \bar{\theta}\|_2^2\big] \lesssim \frac{\bar{\sigma}^2 t_{\mathrm{mix}} d}{(1-\kappa)^2 n}. \tag{3.3}$$

Thus, at least in a worst-case sense, the second term is always dominated by the first term, and our instance-dependent analysis also recovers a worst-case optimal statistical rate for linear Z-estimation with Markovian data. It is also worth noting that the second term in equation (3.2) decays with sample size at $n^{-4/3}$ rate, faster than the $O(n^{-1})$ leading-order term. For sufficiently large sample size n, this term is dominated by the first term, and the behavior of the estimator $\widehat{\theta}_n$ is governed by the instance-optimal quantity. It should be noted that the $n^{-4/3}$-rate of the second-order term—indicating how fast the exact instance-optimal behavior kicks in—may not be optimal. Indeed, it decays more slowly than the n^{-2} second-order asymptotic efficiency in regular parametric models [29, 65], and we conjecture that such a second-order term is unavoidable for stochastic approximation. That being said, the sub-optimality is only a second-order phenomenon, and the main message of Theorem 1 is unaffected: with a reasonable sample size, the Polyak–Ruppert estimator is instance-optimal, up to constant factors.

Note that Theorem 1 involves inexplicit universal constants (c, c_1, c_2). Since our theory focuses on optimal instance-dependent quantities and sample complexities up to universal constant factors, we do not try to optimize these constants. Our proof

[3]This can be easily seen from exponential decay of the correlation; in particular, see equation (6.8) in the proof of the theorem.

gives an upper bound with $c_1 = 16$.[4] That being said, for the tail-averaging procedure described above, we can make the constant c_1 arbitrarily close to 2, as we are using the latter half of the data in Polyak–Ruppert averaging. With a more careful choice of the burn-in period, e.g., $n_0 \asymp \frac{\log n}{\eta(1-\kappa)}$, the constant c_1 in equation (3.2) can be made arbitrarily close to 1. The proof is straightforward—the leading-order term in equation (3.2) comes from the variance of sum of a functional on the Markov chain state space, which can be computed directly.

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) implies a (dimension-dependent) uniform upper bound on the noise. The two assumptions play very different roles when the dimension dependence is taken into account. As we will see in Corollary 4, 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 (3.3) 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 [46, 58, 79]. 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 blowup of $\mathcal{O}(t_{\mathrm{mix}})$ in the sample complexity requirement for the Markov case.

Starting from non-stationary s_0. Note that Theorem 1 is shown under an initial state satisfying $s_0 \sim \xi$. Such an assumption may not be always available in practice. However, in Corollary 1 to follow, we will show that, under minor modification, our conclusion easily extends to non-stationary initial distributions.

For non-stationary initial distributions, we wait for a cold-start period

$$n_c := n/4$$

[4]The universal constants c_2 in the high-order term depend on the constant pre-factor in Proposition 1 to follow. We will not track their explicit values for simplicity.

to start running the stochastic approximation iterate (1.3a); i.e., we take the stepsize sequence as

$$\eta_t := \begin{cases} 0, & t \in \{0, 1, \ldots, n_c - 1\}, \\ \eta, & t \geq n_c. \end{cases}$$

The rest of the SA procedure, including the average step (1.3b), remains the same as before. For notational simplicity, we let $\theta_0 = 0$.

Corollary 1. *Under the same setup and parameter choice as in Theorem 1, assume furthermore that Assumption 1 is satisfied with $c_0 = 1$ and set $n_c = n/4$. There exists an event $\mathcal{E}$ such that $\mathbb{P}(\mathcal{E}) \geq 1 - e^{-n/(16 t_{\mathrm{mix}})}$ and*

$$\begin{aligned} \mathbb{E}\big[\|\widehat{\theta}_n - \bar{\theta}\|_2^2 \mathbf{1}_{\mathcal{E}}\big] \leq \frac{c'}{n} \operatorname{Tr}\big((I - \bar{L})^{-1}(\Sigma^*_{\mathrm{MG}} + \Sigma^*_{\mathrm{Mkv}})(I - \bar{L})^{-\top}\big) \\ + c'\bigg(\frac{\sigma_L^2 d t_{\mathrm{mix}}}{(1-\kappa)^2 n}\bigg)^{4/3} \bar{\sigma}^2 \log^2 n \\ + (\sigma_b^2 + \sigma_L^2 \|\bar{\theta}\|_2^2) \cdot \exp\bigg(-\frac{n}{32 t_{\mathrm{mix}}}\bigg). \end{aligned}$$

See Section B.3 for the proof of this corollary. A few remarks are in order. Compared to the MSE bound in Theorem 1, the bound in Corollary 1 exhibits two differences: we need to exclude an extreme event $\mathcal{E}^c$ that occurs with exponentially small probability, and the right-hand side of the bound involves an additional, exponentially decaying term. Roughly speaking, the high-probability event $\mathcal{E}$ corresponds to the Markov chain states being close to a coupled chain. In the regime $n \gg t_{\mathrm{mix}}$ (which is implied by the sample size requirement in Theorem 1), both the probability of the extreme event and the additional term are very small, and the guarantees under a non-stationary initial distribution behave qualitatively similar to the stationary case. For technical reasons, Corollary 1 requires a slightly stronger condition on the Markov chain—the transition kernel needs to be non-expansive under the metric ρ; that is, we require that $c_0 = 1$ in Assumption 1. However, we note that such a non-expansive property holds for a wide range of applications: it is automatically satisfied in the case of $\rho(x, y) = \mathbf{1}_{x \neq y}$ and $\mathcal{W}_{1,\rho} = d_{\mathrm{TV}}$. For general metric spaces, the mixing time bound in Assumption 1 (a) is usually established by showing that the transition kernel P is a contraction, i.e., $c_0 < 1$, which implies non-expansiveness (see, e.g., [10]). Finally, we note that the higher moment bounds for the last iterate established in Proposition 1 can also be extended to the case of non-stationary initial distributions, yielding similar results.

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* $\tau \geq 2pt_{\text{mix}} \log(c_0 d/\eta)$, *and positive stepsize* $\eta \in (0, \frac{1-\kappa}{2cp^3(\sigma_L^2 d + \gamma_{\max}^2)\tau}]$, *we have*

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

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 [2].) 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 (3.3), 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 $\mathbb{E}[\|\theta_t - \overline{\theta}\|_2^2]$. 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.

Let us now make 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 $(\sigma_L, \sigma_b, \overline{\sigma})$ 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 [5, 70] both exhibit quadratic dependency on the quantity $\frac{\max_{s\in\mathbb{X}} \|\boldsymbol{L}_t(s)\|_{\text{op}}}{1-\kappa}$. As we noted previously in equation (2.4), 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 a complementary note, the analysis in [24] involves a different mixing assumption, and so, it is not directly comparable to our results. However, it is worth noting that their bound $\|\theta_t - \overline{\theta}\|_2$ also has an explicit $\mathcal{O}(d/\sqrt{n})$ 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 stochastic approximation 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. Setup and local neighborhoods. We begin with the setup 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 $\overline{\theta}(P)$ to the fixed-point equation

$$\overline{\theta}(P) = \mathbb{E}_{\xi_P}\big[\boldsymbol{L}(s)\big] \cdot \overline{\theta}(P) + \mathbb{E}_{\xi_P}\big[\boldsymbol{b}(s)\big], \tag{3.4}$$

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 $\overline{\theta}(P_0)$ to a given accuracy. In order to obtain such a lower bound, we consider the fixed-point problem (3.4) over a local neighborhood[5] 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[6] 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:

$$\mathfrak{N}_{\text{Prob}}(P_0, \varepsilon) := \Big\{P : \sum_{x \in \mathbb{X}} \xi_0(x) \cdot \chi^2(P(x, \cdot) \,||\, P_0(x, \cdot)) \le \varepsilon^2\Big\},$$
$$\mathfrak{N}_{\text{Est}}(P_0, \varepsilon) := \big\{P : \|\overline{\theta}(P) - \overline{\theta}(P_0)\|_2 \le \varepsilon\big\}.$$

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 that we draw $x \sim \xi_0$ and then draw the next state $y \sim P_0(x, \cdot)$ according to 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').

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

[6] In our construction, both kernels P_0 and P are rapidly mixing and their stationary measures 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.

On the other hand, the local class $\mathfrak{N}_{\mathrm{Est}}$ contains Markov transition kernels P such that the solution $\overline{\theta}(P)$ to the fixed-point equation (3.4) lies in a local neighborhood of the given solution $\overline{\theta}(P_0)$, 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 $\overline{\theta}(P)$ is very close to $\overline{\theta}(P_0)$, 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 $\mathfrak{N}_{\mathrm{Prob}}$ and $\mathfrak{N}_{\mathrm{Est}}$. We use the shorthand notation $\overline{L}^{(0)} := \mathbb{E}_{\xi_0}[\boldsymbol{L}(s)]$. Also, recall the covariance matrix Σ^*_{Mkv} defined in equation (3.1) for a stationary trajectory $(\tilde{s}_t)_{t\in\mathbb{Z}}$ under the transition kernel P_0. Our bound depends on the *local radius*

$$\varepsilon_n = n^{-1/2}\sqrt{\operatorname{trace}\left((I-\overline{L}^{(0)})^{-1}\Sigma^*_{\mathrm{Mkv}}(I-\overline{L}^{(0)})^{-\top}\right)},$$

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 that 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 $\operatorname{supp}(P_0(s,\cdot)) \geq 2$ 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*

$$n \geq \frac{ct_{\mathrm{mix}}^2\sigma_L^2 d^2\log^2 d}{(1-\kappa)^2} \quad \text{and} \quad n^2\varepsilon_n^2 \geq \frac{2c(1+\sigma_L^2)\overline{\sigma}^2 t_{\mathrm{mix}}^4 d^2}{(1-\kappa)^4}\log^6\left(\frac{d}{\min_s \xi_0(s)}\right),$$

we have the minimax lower bound

$$\inf_{\widehat{\theta}_n}\sup_{P\in\mathfrak{N}'}\mathbb{E}\left[\|\widehat{\theta}_n - \overline{\theta}(P)\|_2^2\right] \geq c_2\varepsilon_n^2,$$

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

See Appendix E 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 $\mathfrak{N}_{\mathrm{Prob}}(P_0, c_1\sqrt{\frac{d}{n}}) \cap \mathfrak{N}_{\mathrm{Est}}(P_0, c_1\varepsilon_n)$, which requires both the transition kernel P and the solution $\overline{\theta}(P)$ to be close to the given problem instance $(P_0, \overline{\theta}(P_0))$. 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 $\overline{\theta}(P_0)$ 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 $\overline{\theta}(P)$ lies in the ball $\mathbb{B}(\overline{\theta}(P_0), c_1\varepsilon_n)$, one cannot do much better than simply outputting an arbitrary point in this ball without looking at the data. Let $\mathfrak{N}_{\text{global}}$ be the set of Markov chain fixed-point equation problem instances satisfying Assumptions 1–4. Following the discussion in Section 3.1.1, a worst-case lazy Markov chain trajectory of length n yields an effective i.i.d. sample size of order n/t_{mix}. Note that, in the i.i.d. settings, the fixed-point equation problem considered in this paper covers linear regression (see [58]). Following the well-known minimax lower bound for linear regression (see [80, Section 15.3]), we can obtain a global minimax lower bound using this reduction

$$\inf_{\widehat{\theta}_n} \sup_{P \in \mathfrak{N}_{\text{global}}} \mathbb{E}\big[\|\widehat{\theta}_n - \overline{\theta}(P)\|_2^2\big] \gtrsim \frac{t_{\text{mix}} d}{(1-\kappa)^2 n},$$

which is also achieved by Theorem 1. Compared to this global minimax lower bound, the local minimax formulation in Theorem 2 provides a more fine-grained characterization of the minimax risk landscape across different problem instances: there could be many different estimators that achieve the global minimax lower bound (for example, Proposition 1 shows that the last iterate is near-optimal up to logarithmic factors), but the Polyak–Rupert averaged estimator is optimally adaptive to the complexity associated to any problem instance, characterized by the quantity ε_n^2.

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 [31, 78]. 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 $\mathfrak{N}_{\text{Est}}(P_0, \varepsilon_n)$ is optimal in the following sense. On the one hand, since the plug-in estimator is asymptotically normal [31], for any decreasing sequence ε_n' such that $\varepsilon_n' > \varepsilon_n$ and $\varepsilon_n' \to 0^+$, the minimax risk within the neighborhood $\mathfrak{N}_{\text{Est}}(P_0, \varepsilon_n')$ 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 $\mathfrak{N}_{\text{Est}}(P_0, \varepsilon_n')$ is at most ε_n'. In the latter case, the neighborhood is so small that it provides more information than the data provides.

Third, note that Theorem 2 involves inexplicit universal constants (c_1, c_2). We do not optimize these constants in our proof, and our proof gives a bound with $c_1 = 1$ and $c_2 = \frac{1}{4(5+\pi)}$. Note that the local asymptotic minimax lower bound for this problem [31] implies that

$$\limsup_{c_1 \to +\infty} \lim_{n \to +\infty} \inf_{\widehat{\theta}_n} \sup_{P \in \mathfrak{N}_{\text{Prob}}(P_0, c_1/\sqrt{n})} \mathbb{E}\big[\|\widehat{\theta}_n - \overline{\theta}(P)\|_2^2\big] \geq \varepsilon_n^2,$$

which suggests that the constants can be sharpened. Indeed, using more careful arguments, the non-asymptotic lower bound exhibits a similar nature: in Theorem 2, if we

take the constant c_1 in the size of the neighborhood sufficiently large, the pre-factor c_2 in the minimax lower bound can be made close to 1, exactly matching the asymptotic lower bound and the refined upper bound (see discussion following Theorem 1 for details). In doing so, we can re-scale the prior distribution in the proof of Theorem 2 with c_1, and the leading-order term in the Bayesian Cramér–Rao lower bound will come with a pre-factor 1. For brevity, we do not dive into details of this argument.

Fourth, 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 in the paper [58] implies a lower bound of the form $c_2 n^{-1} \operatorname{trace}((I - \bar{L}^{(0)})^{-1} \Sigma^*_{\mathrm{MG}} (I - \bar{L}^{(0)})^{-\top})$. Combining it with Theorem 2 implies a minimax lower bound involving the term $c_2' n^{-1} \operatorname{trace}((I - \bar{L}^{(0)})^{-1} (\Sigma^*_{\mathrm{Mkv}} + \Sigma^*_{\mathrm{MG}})(I - \bar{L}^{(0)})^{-\top})$ in a properly defined local neighborhood, thus establishing the optimality of Theorem 1. At the same time, we note that Theorem 2 requires the sample size to be at least $t_{\mathrm{mix}}^2 d^2$, which is more stringent than the $\mathcal{O}(t_{\mathrm{mix}} d)$ 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 $(t_{\mathrm{mix}}, d, \frac{1}{1-\kappa})$ and poly-logarithmic dependency on the quantity $\min_s \xi_0(s)$.

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 (2.5) when 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 equation (2.8), the averaged SA

procedure (1.3) is given by

$$\begin{aligned}\theta_{t+1} &\overset{(a)}{=} \theta_t - \eta\big\{\langle\phi(s_t) - \gamma\phi(s_{t+1}),\, \theta_t\rangle - R_{t+1}(s_t)\big\}\phi(s_t),\\ \widehat{\theta}_n &\overset{(b)}{=} \frac{1}{n-n_0}\sum_{t=n_0}^{n-1}\theta_t.\end{aligned} \tag{4.1}$$

To be clear, the update (4.1)(a) is the standard TD(0) algorithm with stepsize η, whereas the addition of the averaging step (4.1)(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 $d_{\mathrm{TV}}(\delta_s P^{t_{\mathrm{mix}}}, \delta_{s'} P^{t_{\mathrm{mix}}}) \leq \frac{1}{2}$ 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 $t_{\mathrm{mix}} + 1$. 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

$$B := \mathbb{E}_{s\sim\xi}[\phi(s)\phi(s)^\top]$$

and also define the minimum and maximum eigenvalues $\mu := \lambda_{\min}(B)$ and $\beta := \lambda_{\max}(B)$. We assume that

$$\|B^{-1/2}\phi(s)\|_2 \overset{(a)}{\leq} \varsigma\sqrt{d} \quad \text{and} \quad |R_t(s)| \overset{(b)}{\leq} \varsigma \quad \text{for all } s \in \mathbb{X}, \tag{4.2a}$$

$$\mathbb{E}_\xi\big[\langle B^{-1/2}\phi(s),\, u\rangle^4\big] \leq \varsigma^4 \quad \text{for all } u \in \mathbb{S}^{d-1}. \tag{4.2b}$$

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

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

We also define the following covariance matrices according to equation (3.1):

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

Finally, we define the quantity

$$\bar{\sigma}^2 := \varsigma^2 \cdot \sqrt{\mathbb{E}\big[\big(\phi(s_t)^\top \bar{\theta} - \gamma \phi(s_{t+1})\bar{\theta} - R_t(s_t)\big)^4\big]}, \tag{4.3}$$

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

Corollary 2. *Under the setup above, take the stepsize η and burn-in period n_0 as*

$$\eta = \frac{1}{c\beta((\varsigma^4+1)d(1-\kappa)n^2 t_{\mathrm{mix}})^{1/3}} \quad \text{and} \quad n_0 = \frac{1}{2}n, \tag{4.4}$$

and suppose that $\frac{n}{\log^3 n} \geq \frac{2t_{\mathrm{mix}}(\varsigma^4+1)d\beta^2}{(1-\kappa)^2\mu^2}$. The estimator

$$\widehat{V}_n := \widehat{\theta}_n \phi$$

obtained from the Polyak–Ruppert procedure (4.1) satisfies the bound

$$\begin{aligned}\mathbb{E}\big[\|\widehat{V}_n - \bar{V}\|^2_{\mathbb{L}^2(\mathbb{X},\xi)}\big] \leq \frac{c}{n}\operatorname{Tr}\big\{(I_d - M)^{-1}(\Sigma^*_{\mathrm{Mkv}} + \Sigma^*_{\mathrm{MG}})(I_d - M)^{-\top}\big\} \\ + c\left(\frac{\beta^2\varsigma^4 d t_{\mathrm{mix}}}{\mu^2(1-\kappa)^2 n}\right)^{4/3}\bar{\sigma}^2\log^2 n,\end{aligned} \tag{4.5}$$

where $\bar{V}$ is the solution to the projected fixed-point equation (2.6) and $c > 0$ is a universal constant.

See Appendix F.1.1 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 (4.5) 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 this was the focus of our prior paper [58]. In conjunction with this result, Corollary 2 implies the error bound

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

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 2.

4.1.2. Markov chains with mixing in Wasserstein metric. Note that, for Corollary 2, 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 $\mathcal{W}_1(\delta_x P^{t_{\text{mix}}}, \delta_y P^{t_{\text{mix}}}) \leq \frac{1}{2}\|x - y\|_2$ 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:

$$\exists \mu, \beta > 0, \quad \mu I_d \preceq B := \mathbb{E}_{s\sim\xi}\big[\phi(s)\phi(s)^\top\big] \preceq \beta I_d, \tag{4.6a}$$

$$\forall x, y \in \mathbb{X}, \quad \|B^{-1/2}\big(\phi(x) - \phi(y)\big)\|_2 \leq \varsigma\sqrt{d}\,\|x - y\|_2, \tag{4.6b}$$

$$\forall u \in \mathbb{S}^{d-1}, \quad \mathbb{E}_{s\sim\xi}\big[\langle u,\, B^{-1/2}\phi(s)\rangle^4\big] \leq \varsigma^4, \tag{4.6c}$$

$$\forall s, s' \in \mathbb{X},\ t \geq 1, \quad |R_t(s) - R_t(s')| \leq \varsigma\|s - s'\|_2, \quad |R_t(s)| \leq \varsigma \quad \text{a.s.} \tag{4.6d}$$

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 (4.6a) necessarily has Lipschitz constant of order $\mathcal{O}(\sqrt{d})$. 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 setup, we have the following guarantee.

Corollary 3. *Assuming the conditions in equation* (4.6)*, taking stepsize and burn-in period as equation* (4.4)*, for the Polyak–Ruppert averaged stochastic approximation procedure* (4.1)*, the bound* (4.5) *holds.*

See Appendix F.1.2 for the proof.

Corollary 3 shows that the same instance-dependent bound holds true for a continuous state space setting. Such a bound is useful for many applications; for example, in the case of quadratic value functions on a subset of $\mathbb{R}^m$, the feature mapping takes the form

$$x \mapsto \phi(x) := mxx^T$$

so that the dimension $d = m^2$. 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 (4.6) are satisfied with dimension-free constants (ς, μ, β). This example is particularly useful for policy evaluation in Linear Quadratic Regulators (LQR) and more generally for other stochastic dynamical systems.

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[7]. The following assumptions are imposed on the basis functions:

$$\forall j \in \mathbb{N}^+, \quad \sup_{x \in \mathbb{X}} |\phi_j(x)| \leq \varsigma, \tag{4.7a}$$

$$\forall d \in \mathbb{N}^+, \quad \mu I_d \leq \left[\mathbb{E}_{s \sim \xi}\big(\phi_j(s)\phi_\ell(s)\big)\right]_{j,\ell \in [d]} \leq \beta I_d, \tag{4.7b}$$

$$\forall t \geq 1, \quad \sup_{x \in \mathbb{X}} |R_t(x)| \leq \varsigma. \tag{4.7c}$$

The first assumption is standard in non-parametric 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 $\overline{\sigma}$ using the second moment:

$$\overline{\sigma}^2 := \varsigma^2 \cdot \sqrt{\mathbb{E}\left[\left(\bar{V}(s_t) - \gamma \bar{V}(s_{t+1}) - R_t(s_t)\right)^2\right]}. \tag{4.8}$$

Once again, we run the averaged stochastic approximation procedure (4.1) 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 $\mathbb{S}(d_n) := \operatorname{span}(\phi_1, \phi_2, \ldots, \phi_{d_n})$ 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 4. *Assuming the conditions in equation* (4.7)*, take the stepsize and burn-in period as in equation* (4.4)*. Assuming that* $\mu, \beta, \varsigma \asymp 1$*, the Polyak–Ruppert averaged stochastic approximation procedure* (4.1) *satisfies the bound* (4.5) *with* $d = d_n$*.*

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

See Appendix F.1.3 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:

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

Let $\{\psi_j\}_{j=1}^{+\infty}$ be an orthonormal basis of $\mathbb{L}^2(\mathbb{X},\xi)$ such that $\operatorname{span}(\psi_1,\dots,\psi_d) = \operatorname{span}(\phi_1,\dots,\phi_d)$ for any $d \ge 1$. (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 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 in the equation above is actually finite, since both sequences $(\langle u, \psi_j\rangle)_{j=1}^{+\infty}$, $(\langle v, \psi_j\rangle)_{j=1}^{+\infty}$ only have a finite number of 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)

$$\mathbb{E}\big[\|\widehat{V}_n - V^*\|^2_{\mathbb{L}^2(\mathbb{X},\xi)}\big] \le \frac{c}{1-\gamma}\cdot\alpha_{d_n}\|V^*\|^2_{\mathcal{H}} + \frac{c\bar{\sigma}^2 d_n t_{\mathrm{mix}}}{(1-\gamma)^2 n}. \tag{4.9}$$

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 $\mathcal{O}((t_{\mathrm{mix}}/n)^{\frac{2s}{2s+1}})$, which matches the minimax optimal rate proved by [23] 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 (4.5) using $\mathcal{O}(d)$ samples, instance-dependent model selection can in principle be conducted. Bounds of the form (4.9) 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 (2.9b), with some given discount factor $\lambda \in [0,1)$. With observation

model $(\boldsymbol{L}_{t+1}(\omega_t), \boldsymbol{b}_{t+1}(\omega_t))$ given by equation (2.11c), the averaged SA procedure (1.3) can be written as

$$\theta_{t+1} = \theta_t - \eta\big\{\langle \phi(s_t) - \gamma\phi(s_{t+1})^\top, \theta_t\rangle - R_t(s_t)\big\} g_t, \tag{4.10a}$$

where

$$g_t = \gamma\lambda g_{t-1} + \phi(s_t) \tag{4.10b}$$

and

$$\widehat{\theta}_n = \frac{1}{n-n_0}\sum_{t=n_0}^{n-1}\theta_t. \tag{4.10c}$$

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 (4.2a), and assume that the mixing time condition in Assumption 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

$$\omega_t := \left(s_t, s_{t+1}, \frac{1-\gamma\lambda}{\varsigma\sqrt{\beta d}} g_t\right) \in \mathbb{X}^2 \times \mathbb{B}(0,1)$$

and begin by establishing mixing conditions on this augmented chain.

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

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

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

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

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.2.1 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 (2.9b) 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 form

$$\varepsilon_{\mathrm{Mkv},\lambda}\left(s, s^+, \frac{1-\gamma\lambda}{\varsigma\sqrt{\beta d}}g\right) = B^{-1/2}\big(\phi(s)^\top\overline{\theta} - \gamma\phi(s^+)\overline{\theta} - r(s)\big)g,$$
$$\varepsilon_{\mathrm{MG},\lambda}\left(s, s^+, \frac{1-\gamma\lambda}{\varsigma\sqrt{\beta d}}g\right) = B^{-1/2}(R_0(s) - r(s))g.$$

Finally, we define the covariance matrices $\Sigma^*_{\mathrm{Mkv},\lambda}$ and $\Sigma^*_{\mathrm{MG},\lambda}$ according to equation (3.1):

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

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 $\overline{\sigma}$ according to equation (4.3). 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 5. *Under the setup above, take the stepsize and burn-in period as*

$$\eta = \frac{(1-\gamma\lambda)^{2/3}}{c\beta\big((\varsigma^4+1)d(1-\kappa_\lambda)n^2\big(t_{\mathrm{mix}} + \frac{1}{1-\gamma\lambda}\big)\big)^{1/3}} \quad \text{and} \quad n_0 = \frac{1}{2}n, \tag{4.12a}$$

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* (4.10) *has MSE bounded as*

$$\mathbb{E}\big[\|\widehat{V}_n - \overline{V}^{(\lambda)}\|^2_{\mathbb{L}^2(\mathbb{X},\xi)}\big] \le cn^{-1}\operatorname{Tr}\big((I_d - M_\lambda)^{-1}(\Sigma^*_{\mathrm{Mkv}} + \Sigma^*_{\mathrm{MG}})(I_d - M_\lambda)^{-\top}\big)$$
$$+ c\left(\frac{\beta^2\varsigma^4 d\left(t_{\mathrm{mix}} + \frac{1}{1-\gamma\lambda}\right)}{\mu^2(1-\kappa_\lambda)^2(1-\gamma\lambda)^2 n}\right)^{4/3}\overline{\sigma}^2\log^2 n, \tag{4.12b}$$

where $\overline{V}^{(\lambda)}$ *is the solution to the projected fixed-point equation* (2.6).

See Appendix F.2.2 for the proof of this corollary.

A few remarks are in order. First, using the same argument as in Corollaries 3 and 4, one can extend the results for TD(λ) to the cases of continuous state spaces with Wasserstein mixing, as well as to non-parametric 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 roll-out 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 2. On the other hand, by taking $\lambda = \gamma$, we have $\kappa_\lambda \leq \frac{\gamma}{1+\gamma} \leq \frac{1}{2}$, 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 (2.9b). The linear operator in the unprojected fixed-point equation (2.9a) satisfies the norm bound

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

Consequently, invoking Theorem 1 in the paper [58], the approximation error satisfies the bound

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

where $\alpha(M, z) := 1 + \lambda_{\max}((I_d - M)^{-1}(z^2 I_d - MM^\top)(I_d - M)^{-\top})$ is the approximation factor. Combining with Corollary 5, we obtain the following bound on the distance to the true value function:

$$\begin{aligned} \mathbb{E}\big[\|\widehat{V}_n - V^*\|^2_{\mathbb{L}^2(\mathbb{X},\xi)}\big] \leq\ & c\alpha\left(M_\lambda, \frac{(1-\lambda)\gamma}{1-\lambda\gamma}\right) \cdot \inf_{V \in \mathbb{S}} \|V - V^*\|^2_{\mathbb{L}^2(\mathbb{X},\xi)} \\ & + \frac{c}{n} \operatorname{Tr}\left((I_d - M_\lambda)^{-1}(\Sigma^*_{\mathrm{Mkv}} + \Sigma^*_{\mathrm{MG}})(I_d - M_\lambda)^{-\top}\right) \\ & + c\left(\frac{\beta^2 \varsigma^4 d\left(t_{\mathrm{mix}} + \frac{1}{1-\gamma\lambda}\right)}{\mu^2(1-\kappa_\lambda)^2(1-\gamma\lambda)^2 n}\right)^{4/3} \bar{\sigma}^2 \log^2 n \end{aligned} \tag{4.13}$$

for a universal constant $c > 0$.

It can be seen that $\alpha(M_\lambda, \frac{(1-\lambda)\gamma}{1-\lambda\gamma}) \leq c' \frac{1-\lambda\gamma}{1-\gamma}$ for a universal constant. We also recall that $\kappa_\lambda \leq \frac{(1-\lambda)\gamma}{1-\lambda\gamma}$. If we take the parameters (μ, β, ς) to be of constant order, in the worst case, the upper bound (4.13) takes the simplified form

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

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 (4.13) 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 $(\widehat{M}_{\lambda,n}, \widehat{\Sigma^*}_{\mathrm{Mkv},\lambda,n}, \widehat{\Sigma^*}_{\mathrm{MG},\lambda,n})$ for the matrices $(M_\lambda, \Sigma^*_{\mathrm{Mkv},\lambda}, \Sigma^*_{\mathrm{MG},\lambda})$ by replacing the expectations by empirical averages. Similarly, replace $\overline{\theta}^{(\lambda)}$ by $\widehat{\theta}_n$.
- Estimate the approximation factor $\alpha(M_\lambda, \frac{(1-\lambda)\gamma}{1-\lambda\gamma})$ 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 (4.13).

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 autoregressive model. We study the stochastic approximation procedure in which, for any $i \in [k]$, we have

$$A^{(i)}_{t+1} = A^{(i)}_t - \eta\left(\sum_{j=0}^{k-1} A^{(j)}_t X_{t-j} X^\top_{t+1-i} - X_{t+1} X^\top_{t+1-i}\right)$$

and

$$\widehat{A}^{(i)}_n = \frac{1}{n - n_0} \sum_{t=n_0}^{n-1} A^{(i)}_t.$$

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 (2.12). 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., [1, Section 3.3]):

$$\exists P_* \succ 0, Q_* \succ 0 \quad \text{such that } R_*^\top P_* R_* = P_* - Q_*. \tag{4.14}$$

Clearly, we have $P_* \succ Q_*$. We let $\beta := \lambda_{\max}(P_*)$ and $\mu := \lambda_{\min}(Q_*)$. Based on the stability theory for discrete-time linear systems [14], condition (4.14) 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* (4.14) *and assuming that the noise has bounded first moment* $\mathbb{E}[\|\varepsilon_t\|_2] < \infty$, *the stationary distribution* $\widetilde{\xi}$ *for the sliding window* $\omega_t = (X_{t+1}, X_t, \ldots, X_{t-k+1})$ *of the autoregressive process* (2.12) *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* $t_{\text{mix}} = ck + c\frac{\beta}{\mu}(1 + \log\frac{\beta}{\mu})$.

See Appendix F.3.1 for the proof of this claim.

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

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

We are now in a position to consider the problem of parameter estimation using stochastic approximation. Consider the vectorized version of the parameter

$$\theta = \operatorname{vec}\big(\big[A^{(1)}; A^{(2)}; \ldots; A^{(k)}\big]\big) \in \mathbb{R}^{km^2}.$$

The population-level Yule–Walker estimation equation (2.13) can be written as

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

where $\Gamma_i := \mathbb{E}[X_i X_0^\top] \in \mathbb{R}^{m\times m}$ 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 for autoregressive models, the following quantities are relevant:

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

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

Let $\overline{\sigma}$ be defined according to equation (2.3). We have the following corollary for autoregressive models.

Corollary 6. *Under the setup above, take the stepsize and burn-in period as*

$$\eta = \frac{1}{c\big(n^2\big(\frac{\beta}{\mu}\log\frac{\beta}{\mu}\big)(h^*)^2\varsigma^4 k^3 m^2 \beta^8/\mu^8\big)^{1/3}} \quad \text{and} \quad n_0 = \frac{1}{2}n, \tag{4.15a}$$

and suppose that

$$\frac{n}{\log^3 n} \geq \left(k + \frac{\beta}{\mu}\log\frac{\beta}{\mu}\right)\varsigma^4 k^3 m^2 \frac{\beta^8}{\mu^8 (h^*)^2}.$$

Then, the Polyak–Ruppert estimator $(\widehat{A}_n^{(j)})_{j\in[k]}$ *satisfies*

$$\sum_{j=1}^{k} \mathbb{E}\big[\|\widehat{A}_n^{(j)} - A_j^*\|_F^2\big] \leq \frac{c}{n}\,\mathrm{Tr}\left(\big(H^* \otimes I_m\big)^{-1}\Sigma_{\mathrm{Mkv}}\big(H^* \otimes I_m\big)^{-1}\right)$$
$$+ \left\{\frac{k m^2 \varsigma^2 \cdot \lambda_{\max}\big(\mathbb{E}\big[\varepsilon_{\mathrm{Mkv}}(s_0)\varepsilon_{\mathrm{Mkv}}(s_0)^\top\big]\big)}{(h^*)^2 n}\left(k + \frac{\beta}{\mu}\log\frac{\beta}{\mu}\right)\right\}^{4/3} \overline{\sigma}^2 \log^2 n. \tag{4.15b}$$

A few remarks are in order. First, the leading-order term in the bound (4.15b) matches the variance of asymptotic efficient estimators for AR(m) models up to a constant factor (see [14, Section 8]). This simply follows from the fact that the plug-in Yule–Walker estimator is asymptotically efficient for autoregressive models. On the other hand, Corollary 6 is completely non-asymptotic, holding true for any reasonably large sample size. Note that the sample complexity lower bound exhibits an $\mathcal{O}(\beta^9/\mu^9)$ dependency on the conditioning β/μ of the Lyapunov stability certificate (P_*, Q_*). The contributions arise from 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 had made other assumptions on these uniform or moment bounds as in some past work [34], these would have reflected in our result instead of the factor $\beta^8\varsigma^4 k^2/\mu^8$.

5. Proof of Proposition 1

This section is devoted to proving the bound on the last iterate claimed in Proposition 1. We begin in Section 5.1 by deriving a key recursion that underlies the analysis. In Section 5.2, we provide a high-level overview of the proof structure, and the remaining subsections deal with the technical arguments.

5.1. An initial recursion

Define the error term $\Delta_t := \theta_t - \overline{\theta}$, as well as the noise terms

$$Z_{t+1} := L_{t+1} - \boldsymbol{L}(s_t), \quad \zeta_{t+1} := (L_{t+1} - \boldsymbol{L}(s_t))\overline{\theta} + (b_{t+1} - \boldsymbol{b}(s_t)), \tag{5.1a}$$

$$N_t := \boldsymbol{L}(s_t) - \bar{L}, \qquad \nu_t := (\boldsymbol{L}(s_t) - \bar{L})\overline{\theta} + (\boldsymbol{b}(s_t) - \bar{b}). \tag{5.1b}$$

Using this notation, we have the recursion

$$\Delta_{t+1} = (I - \eta(I - \bar{L}))\Delta_t + \eta(N_t + Z_{t+1})\Delta_t + \eta(\nu_t + \zeta_{t+1}). \tag{5.2}$$

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

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

Beginning with the term T_1, expanding the square and then invoking the condition (2.1) yields

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

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

$$\begin{aligned}\|\Delta_{t+1}\|_2^2 \leq \big(1-\eta(1-\kappa)\big)\|\Delta_t\|_2^2 &+ 2\eta \underbrace{\big(\langle \Delta_t,\, N_t\Delta_t\rangle + \langle \Delta_t,\, \nu_t\rangle\big)}_{:=H_1(t)} \\ &+ 2\eta \underbrace{\big(\langle \Delta_t,\, Z_{t+1}\Delta_t\rangle + \langle \Delta_t,\, \zeta_{t+1}\rangle\big)}_{:=H_2(t)} \\ &+ 8\eta^2 \underbrace{\big(\|N_t\Delta_t\|_2^2 + \|Z_{t+1}\Delta_t\|_2^2 + \|\zeta_{t+1}\|_2^2 + \|\nu_t\|_2^2\big)}_{:=H_3(t)}. \end{aligned} \tag{5.3}$$

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

$$(1-\eta(1-\kappa)) \leq e^{-\eta(1-\kappa)},$$

we have

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

Unrolling this expression yields

$$\begin{aligned}e^{\eta(1-\kappa)n}\|\Delta_n\|_2^2 \leq \|\Delta_0\|_2^2 &+ 2\eta\sum_{t=0}^{n-1} e^{\eta(1-\kappa)(t+1)}\big(H_1(t)+H_2(t)\big) \\ &+ 8\eta^2\sum_{t=0}^{n-1} e^{\eta(1-\kappa)(t+1)}H_3(t),\end{aligned} \tag{5.4}$$

which is the key recursion underlying our analysis.

5.2. High-level overview of the proof strategy

Before diving into the remainder of the proof, let us provide a brief overview of our strategy, highlighting the key technical challenges and our solutions to them.

For simplicity, let us give intuition for the analysis under mean-squared error. In order to analyze the recursive error expansion (5.3), we need to bound the terms $\mathbb{E}[H_1(t)]$, $\mathbb{E}[H_2(t)]$, and $\mathbb{E}[H_3(t)]$, respectively. For the martingale noise part, we have $\mathbb{E}[H_2(t)] = 0$. As for the term $H_3(t)$, following Assumption 2, we have that

$$\mathbb{E}[\|\nu_t\|_2^2 + \|\zeta_{t+1}\|_2^2] \lesssim d \quad \text{and} \quad \mathbb{E}[\|Z_{t+1}\Delta_t\|_2^2] \lesssim d \cdot \mathbb{E}[\|\Delta_t\|_2^2].$$

These bounds are similar to the analysis under i.i.d. setup (see the paper [58]). If other terms were not present, we could unroll this recursion and obtain a last-iterate error

bound of $O(\eta d)$, as long as $\eta \ll d^{-1}$. The technical challenges arise, however, with the interaction between Markovian noises and the error Δ_t. In particular, we observe the following facts:

- Since Δ_t and (ν_t, N_t) are inter-dependent, the term $H_1(t)$ does not have zero expectation. If we simply bound it using Assumption 4, for any stepsize $\eta > 0$, the error recursion will diverge as t grows.
- Assumption 4 implies that

$$\mathbb{E}[\|N_t \Delta_t\|_2^2] \lesssim d^2 \cdot \mathbb{E}[\|\Delta_t\|_2^2].$$

  In order to unroll recursion using this bound and obtain convergent result, we need the stepsize $\eta \lesssim d^{-2}$. This will lead to a sub-optimal sample complexity, since we need at least $n \gtrsim \eta^{-1}$ steps.

In tackling the aforementioned difficulties, our first proof technique makes use of the rapid mixing nature of the underlying Markov chain—the Markov chain state after $O(t_{\mathrm{mix}})$ steps is nearly independent of the current iterate. We elaborate on the key ideas as follows.

Multi-step looking-back for the cross terms. Let $\tau \asymp t_{\mathrm{mix}} \log(d/\eta)$. The dependence between $\Delta_{t-\tau}$ and s_t is weak, and consequently, we can show that

$$\begin{aligned} \left|\mathbb{E}\left[\langle N_t \Delta_{t-\tau},\ \Delta_{t-\tau}\rangle\right]\right| &\lesssim \eta d \cdot \mathbb{E}[\|\Delta_{t-\tau}\|_2^2], \\ \left|\mathbb{E}\left[\langle \nu_t,\ \Delta_{t-\tau}\rangle\right]\right| &\lesssim \eta d + \eta d \cdot \mathbb{E}[\|\Delta_{t-\tau}\|_2^2], \end{aligned}$$

and

$$\mathbb{E}\left[\|N_t \Delta_{t-\tau}\|_2^2\right] \lesssim d \cdot \mathbb{E}[\|\Delta_{t-\tau}\|_2^2].$$

In showing these bounds, we construct an auxiliary process $(\tilde{s}_{t-\tau+\ell})_{\ell \geq 0}$, which starts from $\tilde{s}_{t-\tau} \sim \xi$ independent of the data and moves according to the optimal coupling that achieves the Wasserstein mixing. With the value τ given above, we can ensure that $\mathcal{W}_{1,\rho}(s_t, \tilde{s}_t) \lesssim \eta/d$. We can then apply bounds under independent $\tilde{s}_t$ and $\Delta_{t-\tau}$ and bound the residual using Wasserstein distance and the Lipschitz assumption 4. See the proof of Lemma 1 for details.

However, this does not complete the analysis, as we originally need to bound the cross terms between (ν_t, N_t) and Δ_t instead of the τ-step looking-back version $\Delta_{t-\tau}$. In order to convert the above estimates to a useful bound for analyzing the recursion (5.11), we need a stability estimate, i.e., an upper bound on $\mathbb{E}[\|\theta_t - \theta_{t-\tau}\|_2^2]$. This is the major technical challenge we face in order to obtain the sharp dimension dependence. In tackling this challenge, we introduce a novel bootstrapping argument, which may be of independent interest.

Bootstrapping arguments for stability bounds. Expanding the recursion (1.3) for τ steps yields

$$\theta_{t+\tau} - \theta_t = \eta \cdot \sum_{\ell=1}^{\tau} \{(L_{t+\ell} - I_d)\theta_{t+\ell-1} + b_{t+\ell}\}.$$

If we use triangle inequality and Assumptions 2, 4 to bound the difference, some calculations will lead to a coarse bound (see Lemma 5):

$$\sqrt{\mathbb{E}\big[\|\theta_{t+\tau} - \theta_t\|_2^2\big]} \lesssim \eta\tau d \sqrt{\mathbb{E}[\|\Delta_t\|_2^2]} + \eta\tau\sqrt{d}. \tag{5.5}$$

However, if we directly substitute this bound into the arguments above, we will need the stepsize η to satisfy $\eta \lesssim d^{-2}$ in order to make the iterates stable. As we have discussed above, this will cost us a sub-optimal sample complexity of $n \gtrsim d^2$. In order to make the arguments work with a larger stepsize, we need the pre-factor in the first term of the right-hand side of equation (5.5) to be scaling as $O(\tau\eta\sqrt{d})$. To achieve this goal, we start with the bound (5.5) and gradually improve it using a bootstrapping lemma. In Lemma 6 to follow, we show a bootstrapping result: as long as we have the bound

$$\sqrt{\mathbb{E}\big[\|\theta_{t+\tau} - \theta_t\|_2^2\big]} \lesssim \eta\tau\omega \sqrt{\mathbb{E}[\|\Delta_t\|_2^2]} + \eta\tau\beta,$$

we can establish the improved bound

$$\sqrt{\mathbb{E}\big[\|\theta_{t+\tau} - \theta_t\|_2^2\big]} \lesssim \eta\tau\Big(\frac{\omega}{2} + \sqrt{d}\Big)\sqrt{\mathbb{E}[\|\Delta_t\|_2^2]} + \eta\tau\Big(\frac{\beta}{2} + \eta\sqrt{d}\cdot\omega + \sqrt{d}\Big).$$

Once again, the proof of this lemma relies on the multi-step looking-back arguments explained above: when analyzing the iterate (1.3), we can gain the near-independent by replacing θ_t with $\theta_{t-\tau}$, at an additional cost depending on the stability bound $\mathbb{E}[\|\theta_t - \theta_{t-\tau}\|_2^2]$. By repeatedly applying this lemma, we obtain a sequence of pairs (ω, β), which converges to the fixed point

$$\omega \asymp \sqrt{d} \quad \text{and} \quad \beta \asymp \sqrt{d},$$

which yield the desirable stability bound.

Completing the proof by solving the recursion. The improved stability bound allows us to establish sharp bounds on the terms $\mathbb{E}[H_1(t)]$ and $\mathbb{E}[H_3(t)]$. These bounds involve not only the current iterate error Δ_t, but also the looking-back iterate error $\Delta_{t-\tau}$. In order to analyze this type of recursion, we multiply the inequality with an exponentially growing factor $e^{\eta(1-\kappa)t}$ and telescope the summation. Solving it directly yields the MSE bound. As for higher-order moments, we apply martingale concentration inequalities to the martingale noise $H_2(t)$ and the martingales created from the auxiliary processes in analyzing $H_1(t)$. The recursive inequalities in this case can be solved using techniques similar to our prior work [79].

5.3. Analyzing the recursion (5.4)

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 (5.4), we obtain the bound

$$\begin{aligned} e^{\eta(1-\kappa)n}\mathbb{E}\big[\|\Delta_n\|_2^2\big] \leq \|\Delta_0\|_2^2 + 2\eta \sum_{t=0}^{n-1} e^{\eta(1-\kappa)(t+1)}\mathbb{E}\big[\big|H_1(t) - \widetilde{H}_1(t)\big|\big] \\ + 8\eta^2 \sum_{t=0}^{n-1} e^{\eta(1-\kappa)(t+1)}\mathbb{E}\big[H_3(t)\big]. \end{aligned} \tag{5.6}$$

For higher moments, our analysis of the recursion (5.4) 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 (5.4), we obtain the upper bound

$$\begin{aligned} \Phi_n \leq \Phi_0 + 4\eta\big(\mathbb{E}\sup_{0\leq t\leq n}|\widetilde{M}_1(t)|^p\big)^{1/p} + 4\eta\left(\mathbb{E}\left(\sum_{t=0}^{n-1} e^{\eta(1-\kappa)t}|H_1(t)-\widetilde{H}_1(t)|\right)^p\right)^{1/p} \\ + 4\eta\big(\mathbb{E}\sup_{0\leq t\leq n}|M_2(t)|^p\big)^{1/p} + 16\eta^2\left(\mathbb{E}\left(\sum_{t=0}^{n-1} e^{\eta(1-\kappa)t}H_3(t)\right)^p\right)^{1/p}. \end{aligned} \tag{5.7}$$

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 $\mathbb{E}[\widetilde{H}(t)] = 0$ for any $t \geq 0$, and for any integer $p \in [1, \bar{p}/2]$, scalar*

$$\tau \geq cpt_{\text{mix}} \log(c_0 t_{\text{mix}} d/\eta)$$

and stepsize $\eta \leq \frac{1}{ct_{\text{mix}}(\gamma_{\max}+p\sigma_L d)}$, we have the following bounds for any $n > 0$:

$$\big(\mathbb{E}\big[\big|H_1(n) - \widetilde{H}_1(n)\big|^p\big]\big)^{1/p} \leq c\eta p^2\tau\big((d\sigma_L^2 + \gamma_{\max}^2)\cdot\big(\mathbb{E}\|\Delta_{n-\tau\vee 0}\|_2^{2p}\big)^{\frac{1}{p}} + \bar{\sigma}^2 d\big), \tag{5.8a}$$

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

$$\big(\mathbb{E}\sup_{0\leq t\leq n}|\widetilde{M}_1(t)|^p\big)^{1/p} \leq \frac{cp^{3/2}}{\sqrt{\eta(1-\kappa)}}\big(\sigma_L\sqrt{d}\,\Phi_n + \bar{\sigma}\sqrt{e^{\eta(1-\kappa)n}\Phi_n d}\big). \tag{5.8b}$$

See Section 5.4 for the proof of this claim. We note that it is especially challenging to prove the bound (5.8a).

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*

$$\big(\mathbb{E}\sup_{0\leq t\leq n}|M_2(t)|^p\big)^{1/p} \leq \frac{cp^{3/2}\tau^{1/2}}{\sqrt{\eta(1-\kappa)}}\big(\sigma_L\sqrt{d}\,\Phi_n + \bar{\sigma}\sqrt{e^{\eta(1-\kappa)n}\Phi_n d}\big). \tag{5.9}$$

See Section 5.5 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 $\tau \geq cpt_{\text{mix}}\log(c_0 t_{\text{mix}} d/\eta)$ and stepsize $\eta \leq \frac{1}{ct_{\text{mix}}(\gamma_{\max}+\sigma_L d)}$, for any $p \in [1, \bar{p}/2]$, we have*

$$\big(\mathbb{E}\big[H_3(t)^p\big]\big)^{1/p} \leq c\big(p^2\sigma_L^2 d + \gamma_{\max}^2\big)\big(\mathbb{E}\big[\|\Delta_{t-\tau\vee 0}\|_2^{2p}\big]\big)^{1/p} + cp^2\bar{\sigma}^2 d. \tag{5.10}$$

See Section 5.6 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 $\tau = ct_{\text{mix}} \log(c_0 t_{\text{mix}} d/\eta)$ and Cauchy–Schwarz inequality, we have that

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

Similarly, by applying Lemma 3 to the last term of equation (5.6), we obtain the bound

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

Combining them with the decomposition (5.6), for any $n = 1, 2, \ldots$, we find that $e^{\eta(1-\kappa)n}\mathbb{E}[\|\Delta_n\|_2^2]$ is upper bounded by

$$\|\Delta_0\|_2^2 + c\frac{\eta\tau\bar{\sigma}^2 d}{1-\kappa} e^{\eta(1-\kappa)n} + c\eta^2\tau(\sigma_L^2 d + \gamma_{\max}^2) \sum_{t=0}^{n-1} e^{\eta(1-\kappa)t} \mathbb{E}\big[\|\Delta_t\|_2^2\big]. \quad (5.11)$$

In order to exploit this recursive upper bound, we define the partial sum sequence $S_n := \sum_{t=0}^{n} e^{\eta(1-\kappa)t}\mathbb{E}[\|\Delta_t\|_2^2]$. Equation (5.6) 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 (5.11) yields

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

which completes the proof of the MSE bound.

Proof in the case of $\bar{p} \geq \log n$. Now, we turn to prove the p-th moment bound under Assumption 2 with

$$\bar{p} \geq \log n.$$

Recall that we analyze the growth of the Lyapunov function Φ_n, and we start from the decomposition (5.7).

The first term in equation (5.7) is simply $\|\Delta_0\|_2^2$, and the second term is controlled using equation (5.8b) in Lemma 1. In order to bound the third term, we apply Hölder's inequality and obtain the bound

$$\begin{aligned}
&\mathbb{E}\left(\sum_{t=0}^{n-1} e^{\eta(1-\kappa)t}\left|H_1(t) - \widetilde{H}_1(t)\right|\right)^p \\
&\quad \leq \left(\sum_{t=0}^{n-1} e^{\frac{\eta(1-\kappa)pt}{2(p-1)}}\right)^{p-1} \cdot \sum_{t=0}^{n-1} e^{\frac{\eta p(1-\kappa)t}{2}} \mathbb{E}\left|H_1(t) - \widetilde{H}_1(t)\right|^p.
\end{aligned}$$

By equation (5.8a) in Lemma 1, this quantity is at most

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

We then obtain the inequality

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

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 (5.10) yields

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

This quantity is at most

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

Noting that each term satisfies the inequality

$$e^{\frac{\eta p(1-\kappa)t}{2}} (\mathbb{E}[\|\Delta_{t-\tau\vee 0}\|_2^{2p}])^{1/p} \le \Lambda_n$$

for $t \in [0, n]$. We conclude that the moment $(\mathbb{E}(\sum_{t=0}^{n-1} e^{\eta(1-\kappa)t} H_3(t))^p)^{1/p}$ is upper bounded by

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

Collecting the above bounds and substituting into the decomposition (5.7), we note that

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

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 $n \in [1, e^p]$:

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

Note that the right-hand side of the 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

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

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

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

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

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

which completes the proof of the theorem.

It remains to prove our three auxiliary lemmas.

5.4. Proof of Lemma 1

We break the proof into three steps. In the first step, given in Section 5.4.1, we construct the surrogate process, whereas the remaining two steps are devoted to the proving the bounds (5.8b) and (5.8a), as detailed in Sections 5.4.2 and 5.4.3, respectively.

5.4.1. 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 $\tilde{s}_t \in \mathbb{X}$ such that $\tilde{s}_t \mid \mathcal{F}_{t-\tau} \sim \xi$, and

$$\big(\mathbb{E}\big[\rho(s_t, \tilde{s}_t)^p \mid \mathcal{F}_{t-\tau}\big]\big)^{1/p} \le c_0 \exp\Big(-\frac{\tau}{2t_{\text{mix}}\, p}\Big) \quad \text{for each } p \ge 2. \tag{5.12}$$

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*

$$\mathcal{W}_{1,\rho}\big(\delta_x P^\tau, \xi\big) \le c_0 2^{-\lfloor \frac{\tau}{t_{\text{mix}}} \rfloor},$$

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

See Appendix C.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 $\tilde{s}_t$ be a state whose conditional law is ξ, satisfying the identity

$$\mathbb{E}\big[\rho(s_t, \tilde{s}_t) \mid \mathcal{F}_{t-\tau}\big] = \mathcal{W}_{1,\rho}\big(\mathcal{L}(s_t \mid \mathcal{F}_{t-\tau}), \xi\big). \tag{5.13}$$

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

Combining the identity (5.13) with Lemma 4 yields $\mathbb{E}[\rho(s_t, \tilde{s}_t) \mid \mathcal{F}_{t-\tau}] \leq c_0 \cdot 2^{-\lfloor \frac{\tau}{t_{\mathrm{mix}}} \rfloor}$. Applying Cauchy–Schwarz inequality and invoking Assumption 3, we find that

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

which establishes the claim.

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

$$\tilde{\nu}_t := \big(\boldsymbol{L}(\tilde{s}_t) - \bar{L}\big)\bar{\theta} + \big(\boldsymbol{b}(\tilde{s}_t) - \bar{b}\big)$$

and

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

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

5.4.2. Proof of the bound (5.8b). We first perform a decomposition on the process $\widetilde{M}_1$. In particular, for $\ell \in \{0, 1, \ldots, \tau - 1\}$, 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 that, 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 $(\widetilde{\mathcal{F}}_t)_{t \geq 0}$.

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

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

which is at most

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

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

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

and

$$\mathbb{E}\big[\big|\langle \tilde{\nu}_t,\, \Delta_{t-\tau}\rangle\big|^p \mid \mathcal{F}_{t-\tau}\big] \le \big(p\bar{\sigma}\sqrt{d}\cdot\|\Delta_{t-\tau}\|_2\big)^p.$$

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

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

which completes the proof of this lemma.

5.4.3. Proof of the bound (5.8a). By Minkowski's inequality, we can upper bound the error as $(\mathbb{E}[(H_1(t) - \tilde{H}_1(t))^p])^{1/p} \le \sum_{k=1}^6 J_k$, where

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

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

$$\begin{aligned} J_1 &\le \overline{\sigma} d \left(\mathbb{E}\left[\|\Delta_{t-\tau}\|_2^p \cdot \mathbb{E}\left[\rho(s_t, \tilde{s}_t)^p \mid \mathcal{F}_{t-\tau} \right] \right] \right)^{1/p} \\ &\le 2^{1 - \frac{\tau}{2pt_{\mathrm{mix}}}} c_0 \overline{\sigma} d \left(\mathbb{E} \|\Delta_{t-\tau}\|_2^p \right)^{1/p}, \\ J_3 &\le \sigma_L d \left(\mathbb{E}\left[\|\Delta_{t-\tau}\|_2^{2p} \cdot \mathbb{E}\left[\rho(s_t, \tilde{s}_t)^p \mid \mathcal{F}_{t-\tau} \right] \right] \right)^{1/p} \\ &\le 2^{1 - \frac{\tau}{2pt_{\mathrm{mix}}}} c_0 \sigma_L d \left(\mathbb{E} \|\Delta_{t-\tau}\|_2^{2p} \right)^{1/p}. \end{aligned}$$

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

$$J_1 \le \eta \overline{\sigma} \sqrt{d} \left(\mathbb{E} \|\Delta_{t-\tau}\|_2^p \right)^{1/p} \quad \text{and} \quad J_3 \le \eta \sigma_L \sqrt{d} \left(\mathbb{E} \|\Delta_{t-\tau}\|_2^{2p} \right)^{1/p}. \tag{5.14}$$

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

$$\begin{aligned} J_2 &\le \left(\mathbb{E} \|\Delta_t - \Delta_{t-\tau}\|_2^{2p} \right)^{\frac{1}{2p}} \cdot \left(\mathbb{E} \|v_t\|_2^{2p} \right)^{\frac{1}{2p}} \\ &\overset{(i)}{\le} \left(\mathbb{E} \|\Delta_t - \Delta_{t-\tau}\|_2^{2p} \right)^{\frac{1}{2p}} \cdot p\overline{\sigma} \sqrt{d}, \end{aligned} \tag{5.15}$$

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 $\tilde{s}_t$. First, by Cauchy–Schwarz inequality, we note that

$$\begin{aligned} J_4 &\le \left(\mathbb{E} \|\Delta_t - \Delta_{t-\tau}\|_2^{2p} \right)^{\frac{1}{2p}} \cdot \left(\mathbb{E} \|N_t \Delta_{t-\tau}\|_2^{2p} \right)^{\frac{1}{2p}}, \\ J_5 &\le \left(\mathbb{E} \|\Delta_t - \Delta_{t-\tau}\|_2^{2p} \right)^{\frac{1}{2p}} \cdot \left(\mathbb{E} \|N_t^\top \Delta_{t-\tau}\|_2^{2p} \right)^{\frac{1}{2p}}. \end{aligned}$$

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

$$\begin{aligned} \left(\mathbb{E} \|N_t \Delta_{t-\tau}\|_2^{2p} \right)^{\frac{1}{2p}} &\le \left(\mathbb{E} \|(\boldsymbol{L}(\tilde{s}_t) - \overline{L}) \Delta_{t-\tau}\|_2^{2p} \right)^{\frac{1}{2p}} \\ &\quad + \left(\mathbb{E} \|(\boldsymbol{L}(s_t) - \boldsymbol{L}(\tilde{s}_t)) \Delta_{t-\tau}\|_2^{2p} \right)^{\frac{1}{2p}}. \end{aligned}$$

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

$$\mathbb{E}\left[\|(\boldsymbol{L}(\tilde{s}_t) - \overline{L}) \Delta_{t-\tau}\|_2^{2p} \mid \mathcal{F}_{t-\tau} \right] \le (\sigma_L p \sqrt{d})^{2p} \|\Delta_{t-\tau}\|_2^{2p},$$

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

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

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

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

and

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

Putting together the pieces, we arrive at the bound

$$J_4 + J_5 \leq 2\big(\mathbb{E}\|\Delta_t - \Delta_{t-\tau}\|_2^{2p}\big)^{\frac{1}{2p}} \cdot \sigma_L\, p\sqrt{d} \cdot \big(\mathbb{E}\|\Delta_{t-\tau}\|_2^{2p}\big)^{\frac{1}{2p}}. \tag{5.16}$$

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

$$J_6 \leq \big(\mathbb{E}\big[\|\!|N_t|\!\|_{\text{op}}^p \|\Delta_t - \Delta_{t-\tau}\|_2^{2p}\big]\big)^{\frac{1}{p}} \leq \sigma_L d\big(\mathbb{E}\|\Delta_t - \Delta_{t-\tau}\|_2^{2p}\big)^{\frac{1}{p}}. \tag{5.17}$$

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 (1.3a) with stepsize $\eta \leq (6(\gamma_{\max} + \sigma_L d)\ell)^{-1}$ satisfy the bound*

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

and consequently,

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

See Appendix C.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 $p \geq 2$, fixing any integer $\tau \geq 2t_{\text{mix}}\, p \log(c_0 d)$, there exist positive scalars $\omega_p, \beta_p > 0$ such that

$$\big(\mathbb{E}\big[\|\Delta_{t+\ell} - \Delta_t\|_2^p\big]\big)^{1/p} \leq \eta\omega_p \cdot \big(\mathbb{E}\big[\|\Delta_t\|_2^p\big]\big)^{1/p} + \eta\beta_p\overline{\sigma} \tag{5.19}$$

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 (5.19) 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*

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

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

We now complete the proof of the bound (5.8a) by using a bootstrapping argument in order to obtain a sharp bound on $\mathbb{E}\|\Delta_t - \Delta_{t-\tau}\|_2^p$. Let

$$\omega_p^{(0)} := e\tau(\gamma_{\max} + \sigma_L d) \quad \text{and} \quad \beta_p^{(0)} := p\tau\sqrt{d},$$

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

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

$$\big(\mathbb{E}\big[\|\Delta_{t+\ell} - \Delta_t\|_2^p\big]\big)^{1/p} \le 24\eta\tau\big(p\sqrt{d}\sigma_L + \gamma_{\max}\big)\big(\mathbb{E}\|\Delta_t\|_2^p\big)^{1/p} + 6\eta p\tau\sqrt{d}\bar{\sigma} \tag{5.20}$$

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

Collecting the bounds (5.14), (5.15), (5.16), (5.17), and (5.20) and taking the stepsize $\eta \le \frac{1}{c(\gamma_{\max} + p\sigma_L d)\tau}$, we arrive at the bound

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

thereby completing the proof of the bound (5.8a).

5.5. Proof of Lemma 2

By the BDG inequality, we have the bound

$$(\mathbb{E}\sup_{0\le t\le n}|M_2(t)|^p)^{1/p} \le cp(\mathbb{E}([M_2]_n)^{p/2})^{1/p},$$

valid for all $\ell = 0, 1, \ldots, \tau - 1$.

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

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

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

$$\begin{aligned}
\mathbb{E}\big[\big|\langle\Delta_t,\, Z_{t+1}\Delta_t\rangle\big|^p \mid \mathcal{F}_t\big] &\le \|\Delta_t\|_2^p\cdot\mathbb{E}\left[\left(\sum_{j=1}^d\langle e_j,\, Z_{t+1}\Delta_t\rangle^2\right)^{p/2}\mid\mathcal{F}_t\right] \\
&\le \big(p\sigma_L\sqrt{d}\cdot\|\Delta_t\|_2^2\big)^p, \\
\mathbb{E}\big[\big|\langle\zeta_{t+1},\, \Delta_t\rangle\big|^p \mid \mathcal{F}_t\big] &\le \|\Delta_t\|_2^p\cdot\mathbb{E}\left[\left(\sum_{j=1}^d\langle e_j,\, \zeta_{t+1}\rangle^2\right)^{p/2}\mid\mathcal{F}_t\right] \\
&\le \big(p\bar{\sigma}\sqrt{d}\cdot\|\Delta_t\|_2\big)^p.
\end{aligned}$$

Substituting into the bound above, we find that

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

5.6. Proof of Lemma 3

Recall the definitions (5.1a) and (5.1b). By Minkowski's inequality, we have the upper bound

$$\begin{aligned}
\big(\mathbb{E}\big[H_3(t)^p\big]\big)^{1/p} &\le \big(\mathbb{E}\|N_t\Delta_t\|_2^{2p}\big)^{1/p} + \big(\mathbb{E}\|Z_{t+1}\Delta_t\|_2^{2p}\big)^{1/p} \\
&\quad + \big(\mathbb{E}\|\zeta_{t+1}\|_2^{2p}\big)^{1/p} + \big(\mathbb{E}\|\nu_t\|_2^{2p}\big)^{1/p}. \qquad (5.21)
\end{aligned}$$

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

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

For the additive Markov noise, applying Assumption 2 yields the bound

$$(\mathbb{E}\|\nu_t\|_2^{2p})^{1/p} \le p^2\overline{\sigma}^2 d.$$

For the Markov part of the multiplicative noise, we make use of the construction given in Section 5.4.1, where we showed that, for a given $\tau > 0$, there exists a random variable $\tilde{s}_t$ such that $\tilde{s}_t \mid \mathcal{F}_{t-\tau} \sim \xi$, and $\mathbb{E}[\rho^p(s_t, \tilde{s}_t) \mid \mathcal{F}_{t-\tau}] \le c_0 \cdot 2^{1-\frac{\tau}{t_{\text{mix}}}}$. Observe the decomposition

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

Using the Lipschitz condition 4, we have that

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

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

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

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

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

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 (5.20), for

$$\tau \ge 2pt_{\text{mix}}\log(c_0 d)$$

and stepsize $\eta \le \frac{1}{48\tau(\sigma_L\sqrt{d}+\gamma_{\max})}$, we arrive at the bound

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

for a universal constant $c > 0$.

Collecting the bounds above and substituting into our initial bound (5.21), we find that

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

as claimed.

6. Proof of Theorem 1

From the defining equations (1.3a) and (1.3b), we have the telescoping relation

$$\begin{aligned}\frac{\theta_n - \theta_{n_0}}{\eta(n-n_0)} &= \frac{1}{n-n_0}\sum_{t=n_0}^{n-1}\big(\theta_t - L_{t+1}\theta_t - b_{t+1}\big)\\ &= (I-\bar{L})(\hat{\theta}_n - \bar{\theta}) + \frac{1}{n-n_0}\Psi_{n_0,n} + \frac{1}{n-n_0}\Upsilon_{n_0,n},\end{aligned}$$

where $\Psi_{n_0,n} = \sum_{t=n_0}^{n-1}(L_{t+1}\theta_t + b_{t+1} - \mathbb{E}[L_{t+1}\theta_t + b_{t+1}|\mathcal{F}_t])$ and

$$\Upsilon_{n_0,n} := \sum_{t=n_0}^{n-1}(\boldsymbol{L}(s_t)\theta_t + \boldsymbol{b}(s_t) - \bar{L}\theta_t - \bar{b}).$$

Some algebra yields

$$\begin{aligned}\hat{\theta}_n - \bar{\theta} &= \frac{(I-\bar{L})^{-1}\big(\theta_n - \theta_{n_0}\big)}{\eta(n-n_0)} - \frac{(I-\bar{L})^{-1}\Psi_{n_0,n}}{n-n_0} - \frac{(I-\bar{L})^{-1}\Upsilon_{n_0,n}}{n-n_0}\\ &=: I_1 + I_2 + I_3. \end{aligned} \tag{6.1}$$

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

$$\|\hat{\theta}_n - \bar{\theta}\|_Q := \sqrt{(\hat{\theta}_n - \bar{\theta})^\top Q(\hat{\theta}_n - \bar{\theta})}.$$

6.1. Bounding the three terms

We now bound each term in the decomposition (6.1) 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

$$n \geq \frac{8}{\eta(1-\kappa)}\log(\|\theta_0 - \bar{\theta}\|_2 d/\eta)$$

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

$$\mathbb{E}\big[\|I_1\|_Q^2\big] \leq \lambda_{\max}(Q)\mathbb{E}\big[\|I_1\|_2^2\big] \leq \lambda_{\max}(Q)\cdot\frac{c\bar{\sigma}^2\tau d}{\eta(1-\kappa)^3 n^2}. \tag{6.2}$$

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:

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

By the Cauchy–Schwarz inequality, we have the bound

$$\begin{aligned}&\mathbb{E}\big[\|I_2\|_Q^2\big] \\ &\leq \frac{8}{n^2}\sum_{t=n_0}^{n-1}\mathbb{E}\big[\|(I-\bar{L})^{-1}\zeta_{t+1}\|_Q^2\big] + \frac{8}{n^2}\sum_{t=n_0}^{n-1}\mathbb{E}\big[\|(I-\bar{L})^{-1}Z_{t+1}\Delta_t\|_Q^2\big] \\ &\leq \frac{16}{n}\operatorname{Tr}\big(Q(I-\bar{L})^{-1}\Sigma_{\mathrm{MG}}^*(I-\bar{L})^{-\top}\big) + \frac{16\sigma_L^2\lambda_{\max}(Q)d}{(1-\kappa)^2n^2}\sum_{t=n_0}^{n-1}\mathbb{E}\big[\|\Delta_t\|_2^2\big] \\ &\leq \frac{16}{n}\operatorname{Tr}\big((I-\bar{L})^{-1}\Sigma_{\mathrm{MG}}^*(I-\bar{L})^{-\top}\big) + \lambda_{\max}(Q)\cdot\frac{16\sigma_L^2 d}{(1-\kappa)^2 n}\cdot\frac{c\eta d\tau}{1-\kappa}\bar{\sigma}^2. \qquad (6.3)\end{aligned}$$

6.1.3. Bounding the term I_3. Applying the Cauchy–Schwarz inequality yields

$$\begin{aligned}\mathbb{E}\big[\|(I-\bar{L})^{-1}\Upsilon_{n_0,n}\|_2^2\big] \leq\ & 2\mathbb{E}\left[\left\|\sum_{t=n_0}^{n-1}(I-\bar{L})^{-1}\nu_t\right\|_2^2\right] \\ &+ 2\mathbb{E}\left[\left\|\sum_{t=n_0}^{n-1}(I-\bar{L})^{-1}N_t\Delta_t\right\|_2^2\right]. \qquad (6.4)\end{aligned}$$

We make use of the two auxiliary lemmas in order to control the terms in the decomposition (6.4).

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

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

See Section 6.2.1 for the proof of this claim.

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

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

$$\begin{aligned}
\mathbb{E}\big[\|I_3\|_Q^2\big] &\leq \frac{2}{(n-n_0)^2}\mathbb{E}\left[\left\|\sum_{t=n_0}^{n-1}(I-\bar{L})^{-1}\nu_t\right\|_Q^2\right] \\
&\quad + \frac{2}{(n-n_0)^2}\mathbb{E}\left[\left\|\sum_{t=n_0}^{n-1}(I-\bar{L})^{-1}N_t\Delta_t\right\|_Q^2\right] \\
&\leq \frac{8\,\mathrm{Tr}\left(Q(I-\bar{L})^{-1}\Sigma^*_{\text{Mkv}}(I-\bar{L})^{-\top}\right)}{n} \\
&\quad + \lambda_{\max}(Q)\left\{\frac{ct_{\text{mix}}^2\overline{\sigma}^2 d}{(1-\kappa)^2 n^2}\log^2(c_0 d) + \frac{c\eta^2\tau^2 d^2\sigma_L^2\overline{\sigma}^2}{(1-\kappa)^2}\right\}. \qquad (6.5)
\end{aligned}$$

Collecting the bounds (6.2), (6.3), and (6.5), we find that

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

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

$$\begin{aligned}
\mathbb{E}\big[\|\widehat{\theta}_n - \overline{\theta}\|_Q^2\big] &\leq \frac{c}{n}\,\mathrm{Tr}\left(Q(I-\bar{L})^{-1}(\Sigma^*_{\text{MG}} + \Sigma^*_{\text{Mkv}})(I-\bar{L})^{-\top}\right) \\
&\quad + c\lambda_{\max}(Q)\cdot\left(\frac{\sigma_L^2 d t_{\text{mix}}}{(1-\kappa)^2 n}\right)^{4/3}\overline{\sigma}^2\log^2 n. \qquad (6.6)
\end{aligned}$$

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

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

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

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

The desired quantity can be written as

$$\mathrm{Tr}(Q^{1/2}(I-\bar{L})^{-1}\Sigma^*_{\mathrm{Mkv}}(I-\bar{L})^{-\top}Q^{1/2}) = \mu_0 + 2\sum_{k=1}^{+\infty}\mu_k.$$

Expanding the squared norm yields

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

We claim that the cross-correlations μ_k satisfy the bound

$$|\mu_k| \leq c_0\frac{\bar{\sigma}^2\|Q\|_{\mathrm{op}}d^2}{(1-\kappa)^2}\cdot 2^{1-\frac{k}{2t_{\mathrm{mix}}}}. \tag{6.7}$$

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

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

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

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

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

$$\begin{aligned}\sum_{k=n-n_0}^{\infty} |\mu_k| &\leq 2c_0 \frac{\overline{\sigma}^2 d^2 \|\!| Q |\!\|_{\text{op}}}{(1-\kappa)^2} \sum_{k=\frac{1}{2}n}^{\infty} \cdot 2^{-\frac{k}{2t_{\text{mix}}}} \\ &\leq 2c_0 \frac{\overline{\sigma}^2 d^2 \|\!| Q |\!\|_{\text{op}}}{(1-\kappa)^2} \cdot 2^{-\frac{n}{2t_{\text{mix}}}} \leq 2c_0 \frac{\overline{\sigma}^2 d}{(1-\kappa)^2 n^2} \|\!| Q |\!\|_{\text{op}}.\end{aligned}$$

Putting together these bounds yields

$$\begin{aligned}&\mathbb{E}\left[\left\|\sum_{t=n_0}^{n-1} (I-\bar{L})^{-1} v_t\right\|_Q^2\right] \\ &\quad= (n-n_0)\left(\mu_0 + 2\sum_{k=1}^{\infty}\mu_k\right) - 2(n-n_0)\sum_{k=n-n_0}^{\infty}\mu_k - 2\sum_{k=1}^{n-n_0-1} k\mu_k \\ &\quad\leq (n-n_0)\cdot \operatorname{Tr}\left((I-\bar{L})^{-1}\Sigma^*_{\text{Mkv}}(I-\bar{L})^{-1}\right) + \frac{3\tau^2\overline{\sigma}^2 d}{(1-\kappa)^2}\|\!| Q |\!\|_{\text{op}},\end{aligned}$$

which completes the proof of the lemma.

Proof of equation (6.7). Let $s_0 \sim \xi$ and $(s_t)_{t\geq 0}$ be a stationary Markov chain starting from s_0. By the construction given in Section 5.4.1, there exists a random variable $\tilde{s}_k$ such that $\tilde{s}_k$ is independent of s_0, $\tilde{s}_k \sim \xi$, and such that

$$\mathbb{E}[\rho(s_k, \tilde{s}_k) \mid s_0] \leq c_0 \cdot 2^{1-\frac{k}{t_{\text{mix}}}}.$$

We then obtain the bound

$$\begin{aligned}|\mu_k| &= \left|\mathbb{E}\left[\langle Q^{1/2}(I-\bar{L})^{-1}v(s_0),\, Q^{1/2}(I-\bar{L})^{-1}v(s_k)\rangle\right]\right| \\ &\leq \left|\mathbb{E}\left[\left\langle Q^{1/2}(I-\bar{L})^{-1}v(s_0),\, \mathbb{E}\left[Q^{1/2}(I-\bar{L})^{-1}v(\tilde{s}_k) \mid s_0\right]\right\rangle\right]\right| \\ &\quad + \left|\mathbb{E}\left[Q^{1/2}\left\langle (I-\bar{L})^{-1}v(s_0),\, \mathbb{E}\left[Q^{1/2}(I-\bar{L})^{-1}(v(s_k)-v(\tilde{s}_k)) \mid s_0\right]\right\rangle\right]\right| \\ &\leq 0 + \sqrt{\mathbb{E}\left[\|Q^{1/2}(I-\bar{L})^{-1}v(s_0)\|_2^2\right]}\cdot\sqrt{\mathbb{E}\left[\|Q^{1/2}(I-\bar{L})^{-1}(v(s_k)-v(\tilde{s}_k))\|_2^2\right]} \\ &\leq \sqrt{\mu_0 \|\!| Q |\!\|_{\text{op}}}\cdot\frac{\sqrt{\|\!| Q |\!\|_{\text{op}}}}{1-\kappa}\sqrt{\mathbb{E}\left[\rho(s_k,\tilde{s}_k)^2\cdot\overline{\sigma}^2 d^2\right]} \\ &\leq c_0\frac{\overline{\sigma} d}{1-\kappa}\sqrt{\mu_0}\cdot 2^{1-\frac{k}{2t_{\text{mix}}}}. \qquad (6.8)\end{aligned}$$

On the other hand, applying the moment condition (2) yields

$$\mu_0 \leq \frac{1}{(1-\kappa)^2}\cdot\mathbb{E}[\|v(s_0)\|_Q^2] \leq \frac{\overline{\sigma}^2 d}{(1-\kappa)^2}\|\!| Q |\!\|_{\text{op}}.$$

Substituting this bound into our previous inequality (6.8) 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 products of noise matrices. Recursively applying the result for $m = \log d$ times yields the desired bound.

Lemma 9. *Given any integer $m \geq 0$, deterministic sequence $0 = k_0 < k_1 < \cdots < k_m < n_0$, and scalar $\tau \geq 3mt_{\mathrm{mix}}\, p \log(c_0 d\, n)$, we have the second moment bound*

$$\begin{aligned}
&\mathbb{E}\left[\left\|\sum_{t=n_0}^{n-1}\left(\prod_{j=0}^{m} N_{t-k_j}\right)\Delta_{t-k_m}\right\|_2^2\right] \\
&\quad\leq 2n^2 d^{2m}\sigma_L^{2m+2}\cdot\frac{c\eta}{1-\kappa}dt_{\mathrm{mix}}\bar{\sigma}^2 \\
&\qquad+4\eta^2\tau\sum_{k_{m+1}=k_m+1}^{k_m+\tau}\mathbb{E}\left[\left\|\sum_{t=n_0}^{n}\left\{\prod_{j=0}^{m+1} N_{t-k_j}\Delta_{t-k_{m+1}}\right\}\right\|_2^2\right] \\
&\qquad+4\eta^2\tau\sum_{k_{m+1}=k_m+1}^{k_m+\tau}\mathbb{E}\left[\left\|\sum_{t=n_0}^{n}\left\{\prod_{j=0}^{m} N_{t-k_j}\left(v_{t-k_{m+1}}+\zeta_{t-k_{m+1}+1}\right)\right\}\right\|_2^2\right], \qquad (6.9a)
\end{aligned}$$

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

$$\begin{aligned}
\mathbb{E}\left[\left\|\sum_{t=n_0}^{n-1} N_t\Delta_t\right\|_2^2\right] &\leq c\sigma_L^2 d\cdot\left(n\tau+n^2\eta^2\sigma_L^2 d\tau^2\right)\frac{c\eta}{1-\kappa}dt_{\mathrm{mix}}\bar{\sigma}^2 \\
&\quad+4\eta^2\tau\sum_{k_1=1}^{\tau}\mathbb{E}\left[\left\|\sum_{t=n_0}^{n} N_t N_{t-k_1}\Delta_{t-k_1}\right\|_2^2\right] \\
&\quad+4\eta^2\tau\sum_{k_1=1}^{\tau}\mathbb{E}\left[\left\|\sum_{t=n_0}^{n} N_t\left(v_{t-k_1}+\zeta_{t-k_1+1}\right)\right\|_2^2\right]. \qquad (6.9b)
\end{aligned}$$

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

The following lemma controls the last term of the bound (6.9a).

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*

$$\begin{aligned}
&\mathbb{E}\left[\left\|\sum_{t=n_0}^{n-1}\left(\prod_{j=0}^{m-1} N_{t-k_j}\right)\left(v_{t-k_m}+\zeta_{t-k_m+1}\right)\right\|_2^2\right] \\
&\quad\leq c\left(n^2+nd(k_m+t_{\mathrm{mix}}\log(c_0 d))\right)\sigma_L^{2m} d^{2m}\bar{\sigma}^2.
\end{aligned}$$

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

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

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

for $m = 0, 1, 2, \ldots, \log d$. By equation (6.9b) and Lemma 10, we have the bound

$$\begin{aligned}\mathfrak{S}_0 &\le c\sigma_L^2 d\cdot\big(n\tau + n^2\eta^2\sigma_L^2 d\tau^2\big)\frac{c\eta}{1-\kappa}dt_{\text{mix}}\overline{\sigma}^2 \\ &\quad + 4\eta^2\tau^2\mathfrak{S}_1 + 4c\eta^2\tau^2\big(n^2 + nd(\tau + t_{\text{mix}}\log(c_0 d))\big)\sigma_L^2 d^2\overline{\sigma}^2 \\ &\le 4\eta^2\tau^2\mathfrak{S}_1 + c'\eta^2 n^2\tau^2 d^2\sigma_L^2\overline{\sigma}^2.\end{aligned}$$

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 (6.9a) and Lemma 10, we have the recursive relation

$$\begin{aligned}\mathfrak{S}_m &\le 4\eta^2\tau^2\mathfrak{S}_{m+1} + cn^2 d^{2m+1}\tau\sigma_L^{2m+2}\cdot\frac{\eta\log^3 n}{1-\kappa}\overline{\sigma}^2 + c\eta^2\tau^2 n^2\sigma_L^{2m+2}d^{2m+2}\overline{\sigma}^2 \\ &\le 4\eta^2\tau^2\mathfrak{S}_{m+1} + cn^2\sigma_L^{2m}d^{2m}\overline{\sigma}^2\cdot\log^3 n.\end{aligned}$$

Recursively applying these bounds yields

$$\begin{aligned}\mathfrak{S}_0 &\le (4\eta^2\tau^2)^m\mathfrak{S}_m + c\eta^2 n^2\tau^2 d^2\sigma_L^2\overline{\sigma}^2 + c\cdot\sum_{q=1}^{m-1}(4\eta^2\tau^2)^q n^2\sigma_L^{2q}d^{2q}\overline{\sigma}^2 \\ &\le (4\eta^2\tau^2)^m\mathfrak{S}_m + 3c\eta^2 n^2\tau^2 d^2\sigma_L^2\overline{\sigma}^2.\end{aligned}$$

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

$$\begin{aligned}\mathbb{E}\left[\left\|\sum_{t=n_0}^{n-1}\left(\prod_{j=0}^{m}N_{t-k_j}\right)\Delta_{t-k_m}\right\|_2^2\right] &\le n\sum_{t=n_0}^{n-1}\mathbb{E}\left[\left\|\left(\prod_{j=0}^{m}N_{t-k_j}\right)\Delta_{t-k_m}\right\|_2^2\right] \\ &\le n^2(\sigma_L d)^{2m+2}\cdot\frac{c\eta t_{\text{mix}}d\overline{\sigma}^2}{1-\kappa}.\end{aligned}$$

Taking the supremum and noting that $\eta \le \frac{1-\kappa}{\sigma_L^2 d\tau}$ leads to $\mathfrak{S}_m \le cn^2\sigma_L^{2m}d^{2m+2}\overline{\sigma}^2$. Consequently, we have established that

$$\mathfrak{S}_0 \le 3c\eta^2 n^2\tau^2 d^2\sigma_L^2\overline{\sigma}^2[1 + (2\eta\tau\sigma_L d)^{\frac{2m+2}{2m}}].$$

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,

$$\mathfrak{S}_0 \le 6c\eta^2 n^2\tau^2 d^2\sigma_L^2\overline{\sigma}^2\log^3 n,$$

which completes the proof of this lemma.

7. 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 and 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.

- *Non-linear 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 $\mathbb{E}_{s\sim\xi(\theta)}[H(\theta; s)] = 0$; see the book [2] 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. In particular, the multi-step looking-back technique and bootstrapping stability bounds introduced in Proposition 1 could be extended to non-linear operators, and it would be very interesting to see how Markovian SA achieves optimal dependence on (t_{mix}, d) in general. On the other hand, the proof of Theorem 1 is specialized to linear operators, as it explicitly involves bounding product of random matrices (see Lemma 9). Obtaining sharp and instance-dependent results for the non-linear SA may require novel proof techniques and is an important direction of future work.
- *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, we have the following

limiting result (see the paper [26] for details):

$$\sqrt{n}(\widehat{\theta}_n - \overline{\theta}) \xrightarrow{d} \mathcal{N}\big((I_d - \overline{L})^{-1}(\Sigma^*_{\mathrm{MG}} + \Sigma^*_{\mathrm{Mkv}})(I_d - \overline{L})^{-\top}\big). \tag{7.1}$$

In order to construct confidence intervals for the solution $\overline{\theta}$ with streaming data, it suffices to estimate the asymptotic covariance in equation (7.1). In the i.i.d. setting, online procedures have been developed to estimate such covariances, with non-asymptotic error guarantees [16]. 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 Sections 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 (4.12b) 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 [47].

A. 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.

A.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 [6] for an overview of some classical results. For the problem of functional estimation under the stationary distribution, the asymptotic efficiency of plug-in estimators[8] has been established for discrete-state Markov chains [31,63] and Itô diffusion processes [45]. 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

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

non-asymptotic results for statistical estimation with Markovian data, including the estimation of transition kernels [52, 84], mixing times [33], and the parameters of Gaussian hidden Markov models [85], as well for certain testing problems [18]. 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.

A.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 [66]; see the reference books [2,7,44] 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 [64,68]. A variety of stochastic approximation procedures now serve as the workhorse for modern large-scale machine learning and statistical inference [9,61], and many algorithmic techniques are known to accelerate their convergence [28,35,49]. In particular, non-asymptotic bounds matching the optimal Gaussian limit have been established in a variety of settings [21,27,58,59,79].

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.[9] The first study of SA procedures without a martingale difference assumption was initiated by [43], who give a general criteria for convergence, as well as [53,54], who analyzed linear problems motivated by control and filtering. The paper [56] 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. The paper [15] provides finite-sample error bounds on the averaged iterate of Markovian linear stochastic approximation, with an optimal leading-order term. Central limit theorems [26] and non-asymptotic convergence rates [37] 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 [22, 38], observation skipping schemes for

[9] 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.

bias reduction [42], Lyapunov function-based analysis under general norms [17], and proving guarantees under weaker ergodicity conditions [20].

A.3. Application to RL problems

Markovian observations arise naturally in the context of stochastic control and reinforcement learning (RL). See [2] for a historical survey of algorithms for stochastic control and filtering with Markovian stochastic approximation and the books [4, 73] 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 [71] use linear stochastic approximation to estimate the value function of a given policy, with asymptotic convergence guarantees [11, 19, 75] and non-asymptotic bounds [5, 39, 58]. In the non-linear case, the Q-learning algorithm [83] 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 [25, 72, 74], results on linear function approximation for optimal stopping problems [5, 76], and non-asymptotic rates under general norms in both the i.i.d. setting [8, 81] and the Markovian setting [17]. A class of variants of TD and Q-learning are also studied in the literature, including actor-critic methods [41], SARSA [67], and methods that employ variance reduction [39, 40, 69, 82]. A concurrent preprint to this manuscript [51] proves lower bounds on the oracle complexity of policy evaluation with access to temporal difference operators and develops an acceleration scheme with variance 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 [3, 58, 60, 75, 86]. Prior work by a subset of the current authors [58] 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.

B. 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.

B.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 4. *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 $t = 1, 2, \ldots, n$, with probability $1 - \delta$:*

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

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 satisfy the moment assumption 2 with $\bar{p} = +\infty$, we can apply our theorems on the event that equation (B.1) 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 4. 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 $1 - \delta/n$:

$$\|b_{t+1} - \bar{b}\|_2^2 \le \sum_{j=1}^{d} \left(e_j^\top (b_{t+1} - \bar{b})\right)^2 \le c\sigma_b^2 d \cdot \log^2(nd/\delta).$$

Taking union bound over

$$t = 1, 2, \ldots, n,$$

we complete the proof of this proposition.

B.2. On the stationary tail and boundedness assumption 3

Note that, in many applications, the Markov chain $(s_t)_{t\ge 0}$ 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 $p \ge 2$:

$$\mathbb{E}_{s\sim\xi}\left[\rho(s, s_0)^p\right] \le p! \cdot \sigma_\rho^p. \tag{B.2}$$

Given a stationary Markovian trajectory $\{s_t\}_{t=1}^n$, consider the event

$$\mathcal{E}_{n,\delta} = \left\{\forall t \in [1, n],\ \rho(s_0, s_t) \le 2\sigma_\rho \log \frac{n}{\delta}\right\}.$$

By the tail assumption (B.2) and a union bound, it directly follows that

$$\mathbb{P}(\mathcal{E}_{n,\delta}) \ge 1 - \delta.$$

Consider a truncated Markov transition kernel P' defined as

$$P'(x, Z) := P\left(x, Z \cap \mathbb{B}\left(0, 2\sigma_\rho \log(n/\delta)\right)\right) + P\left(x, \mathbb{B}\left(0, 2\sigma_\rho \log(n/\delta)\right)^c\right)\mathbf{1}_{s_0\in Z}$$

for any $x \in \mathbb{X}$ and $Z \subseteq \mathbb{X}$.

In words, the Markov chain P' attempts to make the transition from s_t to s_{t+1} according to the original Markov transition kernel P'. If the state s_{t+1} lies in the ball $\mathbb{B}(0, 2\sigma_\rho \log(n/\delta))^c$, 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

$$\mathbb{P}\left(\{s_t\}_{t=1}^n \ne \{s'_t\}_{t=1}^n\right) \le \mathbb{P}\left(\mathcal{E}_{n,\delta}^c\right) \le \delta.$$

One can then proceed by working on the high-probability event $\mathcal{E}_{n,\delta}$, where the Markov chain has an effective diameter of $O(\sigma_\rho \log \frac{n}{\delta})$.

B.3. Proof of Corollary 1

Suppose that $s_0 \sim \pi_0$, by Lemma 4 and convexity of the Wasserstein distance, we have

$$\mathcal{W}_{1,\rho}\big(\pi_0 P^{n_c}, \xi\big) \le 2^{-\lfloor n_c / t_{\mathrm{mix}} \rfloor} \le 2\exp\Big(-\frac{n}{8t_{\mathrm{mix}}}\Big).$$

Let $(\tilde{s}_t)_{t\ge 0}$ be a stationary chain with $\tilde{s}_0 \sim \xi$ independent of s_0. There exists a coupling between the paths such that

$$\mathbb{E}\big[\rho(s_{n_c}, \tilde{s}_{n_c})\big] \le 2\exp\Big(-\frac{n}{8t_{\mathrm{mix}}}\Big).$$

Applying Assumption 1 (b) conditionally on $(s_{n_c}, \tilde{s}_{n_c})$, since

$$c_0 = 1,$$

there exists a coupling between the next-step transitions such that

$$\mathbb{E}\big[\rho(s_{n_c+1}, \tilde{s}_{n_c+1}) \mid (s_{n_c}, \tilde{s}_{n_c})\big] \le \rho(s_{n_c}, \tilde{s}_{n_c}).$$

Similarly, we can inductively construct the coupling between s_{n_c+i+1} and $\tilde{s}_{n_c+i+1}$ conditionally on the pair $(s_{n_c+i}, \tilde{s}_{n_c+i})$ for $i = 1, 2, \ldots$. Putting them together, we obtain a coupling between the two paths such that $(\rho(s_{n_c+i+1}, \tilde{s}_{n_c+i+1}))_{i\ge 0}$ is a super-martingale. By Markov inequality, for each $t \ge n_c$, we have

$$\mathbb{P}\big(\rho(s_t, \tilde{s}_t) \ge e^{-\frac{n}{16t_{\mathrm{mix}}}}\big) \le 2\exp\Big(-\frac{n}{8t_{\mathrm{mix}}}\Big).$$

Define the event

$$\mathcal{E} := \big\{\rho(s_t, \tilde{s}_t) \le e^{-\frac{n}{16t_{\mathrm{mix}}}} : t = n_c, n_c+1, \ldots, n\big\}.$$

By union bound, we have

$$\mathbb{P}\big(\mathcal{E}\big) \ge 1 - 2n\exp\Big(-\frac{n}{8t_{\mathrm{mix}}}\Big) \ge 1 - \exp\Big(-\frac{n}{16t_{\mathrm{mix}}}\Big).$$

Define the error scalar

$$\delta_n := e^{-\frac{n}{16t_{\mathrm{mix}}}},$$

and let $(\theta_t)_{t\ge n_c}$, $(\widetilde{\theta}_t)_{t\ge n_c}$ be the iterate sequences generated by the Markov chains $(s_t)_{t\ge n_c}$ and $(\tilde{s}_t)_{t\ge n_c}$, respectively. For any $t \ge n_c$, we note that

$$\theta_{t+1} = \big((1-\eta)I_d + \eta \boldsymbol{L}_{t+1}(s_t)\big)\theta_t + \eta \boldsymbol{b}_{t+1}(s_t),$$
$$\widetilde{\theta}_{t+1} = \big((1-\eta)I_d + \eta \boldsymbol{L}_{t+1}(\tilde{s}_t)\big)\widetilde{\theta}_t + \eta \boldsymbol{b}_{t+1}(\tilde{s}_t).$$

Taking their difference and applying triangle inequality, on the event $\mathcal{E}$, we have the almost sure upper bound on the one-step error

$$\begin{aligned}
&\|\widetilde{\theta}_{t+1} - \theta_{t+1}\|_2 \\
&\le \|\!|(1-\eta) I_d + \eta \boldsymbol{L}_{t+1}(s_t)|\!\|_{\mathrm{op}} \cdot \|\widetilde{\theta}_t - \theta_t\|_2 + \eta \|\!|\boldsymbol{L}_{t+1}(\tilde{s}_t) - \boldsymbol{L}_{t+1}(s_t)|\!\|_{\mathrm{op}} \cdot \|\widetilde{\theta}_t\|_2 \\
&\quad + \eta \|\boldsymbol{b}_{t+1}(\tilde{s}_t) - \boldsymbol{b}_{t+1}(s_t)\|_2 \\
&\le (1 + \eta(\gamma_{\max} + d\sigma_L))\|\widetilde{\theta}_t - \theta_t\|_2 + \eta\delta_n \cdot \big\{\sigma_L d \|\widetilde{\theta}_t\|_2 + \sigma_b\sqrt{d}\big\},
\end{aligned}$$

where, in the last step, we use the Lipschitz assumption 4.

Solving this recursion yields the uniform upper bound for $t \in \{n_c, n_c+1, \ldots, n\}$:

$$\|\widetilde{\theta}_t - \theta_t\|_2 \le \eta\delta_n \exp\big(\eta(\gamma_{\max} + d\sigma_L)n\big) \cdot \sum_{t=n_c}^{n} \big\{\sigma_L d \|\widetilde{\theta}_t\|_2 + \sigma_b\sqrt{d}\big\},$$

holding with probability 1 on the event $\mathcal{E}$.

Given a stepsize satisfying $\eta \le \frac{1}{32 t_{\mathrm{mix}}(\gamma_{\max} + d\sigma_L)}$, we have

$$\delta_n \exp\big(\eta(\gamma_{\max} + d\sigma_L)n\big) \le \exp\Big(-\frac{n}{32 t_{\mathrm{mix}}}\Big).$$

For the summation term, we apply Cauchy–Schwarz inequality and obtain the MSE bound

$$\begin{aligned}
\mathbb{E}\left\{\sum_{t=n_c}^{n} \sigma_L d \|\theta_t\|_2 + \sigma_b\sqrt{d}\right\}^2 &\le 2n^2 d^2\big(\sigma_b^2 + \sigma_L^2 \|\bar{\theta}\|_2^2\big) + 4n\sigma_L^2 d^2 \sum_{t=n_c}^{n} \mathbb{E}\big[\|\widetilde{\theta}_t - \bar{\theta}\|_2^2\big] \\
&\overset{(i)}{\le} n^2 d^2 \Big(2\sigma_b^2 + 6\sigma_L^2 \|\bar{\theta}\|_2^2 + \frac{4c\eta t_{\mathrm{mix}} \eta}{1-\kappa} \bar{\sigma}^2 \log n\Big) \\
&\le 12 n^3 d^2 \big(\sigma_b^2 + \sigma_L^2 \|\bar{\theta}\|_2^2\big),
\end{aligned}$$

where, in step (i), we apply Proposition 1 to the iterate sequence $(\widetilde{\theta}_t)_{t \ge n_c}$.

Putting them together, we conclude that

$$\mathbb{E}\big[\|\widetilde{\theta}_t - \theta_t\|_2^2 \mathbf{1}_{\mathcal{E}}\big] \le 12 n^3 d^2 \big(\sigma_b^2 + \sigma_L^2 \|\bar{\theta}\|_2^2\big) \exp\Big(-\frac{n}{16 t_{\mathrm{mix}}}\Big).$$

Let $\widehat{\theta}'_n := \frac{1}{n-n_0} \sum_{t=n_0}^{n-1} \widetilde{\theta}_t$; applying Cauchy–Schwarz inequality, we have

$$\begin{aligned}
\mathbb{E}\big[\|\widehat{\theta}'_n - \widehat{\theta}_n\|_2^2 \mathbf{1}_{\mathcal{E}}\big] &\le \frac{4}{n} \sum_{t=n_0}^{n} \mathbb{E}\big[\|\widetilde{\theta}_t - \theta_t\|_2^2 \mathbf{1}_{\mathcal{E}}\big] \\
&\le 12 n^3 d^2 \big(\sigma_b^2 + \sigma_L^2 \|\bar{\theta}\|_2^2\big) \exp\Big(-\frac{n}{16 t_{\mathrm{mix}}}\Big) \\
&\le e^{-\frac{n}{32 t_{\mathrm{mix}}}} \big(\sigma_b^2 + \sigma_L^2 \|\bar{\theta}\|_2^2\big)
\end{aligned}$$

for a sample size satisfying $\frac{n}{\log n} \geq 2400 t_{\text{mix}} \log d$. Invoking Theorem 1 on the estimator $\widehat{\theta}'_n$ completes the proof of this corollary.

C. Auxiliary results underlying Proposition 1

This appendix is devoted to the proofs of auxiliary lemmas that are used in the proof of Proposition 1.

C.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 $\tau = k t_{\text{mix}} + q$ with $k \in \mathbb{N}_+$ and $0 \leq q \leq t_{\text{mix}} - 1$. 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

$$\mathcal{W}_{1,\rho}\big(\delta_x P^{k t_{\text{mix}}+q}, \xi\big) \leq c_0 \cdot 2^{-k} \quad \text{for } q = 0, 1, \ldots, t_{\text{mix}} - 1. \tag{C.1}$$

We now need to show that the result holds for any $\tau = (k+1)t_{\text{mix}} + q$, where $q \in \{0, 1, \ldots, t_{\text{mix}} - 1\}$ 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 $s_{k t_{\text{mix}}+q}$ and $s'_{k t_{\text{mix}}+q}$ to satisfy the identity

$$\mathbb{E}[\rho(s_{k t_{\text{mix}}+q}, s'_{k t_{\text{mix}}+q})] = \mathcal{W}_{1,\rho}(\delta_x P^{k t_{\text{mix}}+q}, \xi).$$

Conditionally on $(s_{k t_{\text{mix}}+q}, s'_{k t_{\text{mix}}+q})$, Assumption 1 guarantees the existence of a coupling between $\delta_{s_{k t_{\text{mix}}+q}} P^{t_{\text{mix}}}$ and $s'_{k t_{\text{mix}}+q} P^{t_{\text{mix}}}$ such that

$$\mathbb{E}\big[\rho(s_{(k+1)t_{\text{mix}}+q}, s'_{(k+1)t_{\text{mix}}+q}) \mid (s_{k t_{\text{mix}}+q}, s'_{k t_{\text{mix}}+q})\big] \leq \frac{1}{2}\rho(s_{k t_{\text{mix}}+q}, s'_{k t_{\text{mix}}+q}).$$

Taking expectation on both sides and substituting with equation (C.1), 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.

C.2. Proof of Lemma 5

Our proof is based on the following intermediate claim:

$$\big(\mathbb{E}\big[\|\Delta_{t+\ell}\|_2^p\big]\big)^{1/p} \le e\big(\mathbb{E}\big[\|\Delta_t\|_2^p\big]\big)^{1/p} + 6\eta p \ell \overline{\sigma}\sqrt{d}. \tag{C.2}$$

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 (C.2) to prove the lemma. Applying Minkowski's inequality to the recursive relation (5.2), we find that, for any $p \ge 2$, the p-th moment is upper bounded as

$$\begin{aligned}\big(\mathbb{E}\big[\|\Delta_{t+\ell+1} - \Delta_t\|_2^p\big]\big)^{1/p} \le \big(\mathbb{E}\big[\|\Delta_{t+\ell} - \Delta_t\|_2^p\big]\big)^{1/p} + \eta\big(\mathbb{E}\big[\|L_{t+\ell+1}\Delta_{t+\ell}\|_2^p\big]\big)^{1/p} \\ + \eta\big(\mathbb{E}\big[\|\nu_{t+\ell} + \zeta_{t+\ell+1}\|_2^p\big]\big)^{1/p}.\end{aligned}$$

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

$$\begin{aligned}\mathbb{E}\big[\|Z_{t+\ell+1}\Delta_{t+\ell}\|_2^p \mid \mathcal{F}_t\big] &\le d^{\frac{p}{2}} \sum_{j=1}^d \mathbb{E}\big[\langle e_j,\, Z_{t+\ell+1}\Delta_{t+\ell}\rangle^p \mid \mathcal{F}_t\big] \\ &\le \big(p\sigma_L\sqrt{d}\big)^p \mathbb{E}\big[\|\Delta_{t+\ell}\|_2^p \mid \mathcal{F}_t\big]\end{aligned}$$

and

$$\mathbb{E}\big[\|\zeta_{t+\ell+1}\|_2^p\big] \le d^{\frac{p}{2}} \sum_{j=1}^d \mathbb{E}\big[\langle e_j,\, \zeta_{t+\ell+1}\rangle^p\big] \le (p\sqrt{d})^p \cdot \overline{\sigma}^p.$$

Similarly, for the Markov part of the noise, we have

$$\mathbb{E}\big[\|\nu_{t+\ell+1}\|_2^p\big] \le (p\sqrt{d})^p \cdot \overline{\sigma}^p.$$

On the other hand, the Lipschitz condition 4 and the boundedness condition (3) of the metric space imply that

$$\|\!|L_{t+\ell+1}(s) - \overline{L}|\!\|_{\mathrm{op}} \le \sigma_L d \quad \text{for all } s \in \mathbb{X}.$$

Substituting into the decomposition above, we arrive at the bounds

$$\big(\mathbb{E}\big[\|L_{t+\ell+1}\Delta_{t+\ell}\|_2^p\big]\big)^{1/p} \le (\gamma_{\max} + \sigma_L p\sqrt{d} + \sigma_L d)\big(\mathbb{E}\big[\|\Delta_{t+\ell}\|_2^p\big]\big)^{1/p}$$

and

$$\big(\mathbb{E}\big[\|\nu_{t+\ell} + \zeta_{t+\ell+1}\|_2^p\big]\big)^{1/p} \le 2p\overline{\sigma}\sqrt{d}.$$

Applying equation (C.2) yields

$$\begin{aligned}\big(\mathbb{E}\big[\|\Delta_{t+\ell+1}-\Delta_t\|_2^p\big]\big)^{1/p} &\le \big(\mathbb{E}\big[\|\Delta_{t+\ell}-\Delta_t\|_2^p\big]\big)^{1/p} \\ &\quad + e\eta(\gamma_{\max}+\sigma_L d)\big(\mathbb{E}\big[\|\Delta_t\|_2^p\big]\big)^{1/p} \\ &\quad + 2(1+6\eta\ell)\eta p\bar{\sigma}\sqrt{d},\end{aligned}$$

where the second inequality comes from the definition of $\bar{\sigma}$.

Solving this recursion leads to the bound

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

which establishes the first claim.

Since the stepsize is upper bounded as $\eta \le (2e\eta\ell(\gamma_{\max}+\sigma_L d))^{-1}$, we have the lower bound

$$\begin{aligned}\big(\mathbb{E}\big[\|\Delta_{t+\ell}\|_2^p\big]\big)^{1/p} &\ge \big(\mathbb{E}\big[\|\Delta_t\|_2^p\big]\big)^{1/p} - \big(\mathbb{E}\big[\|\Delta_{t+\ell}-\Delta_t\|_2^p\big]\big)^{1/p} \\ &\ge \frac{1}{2}\big(\mathbb{E}\big[\|\Delta_t\|_2^p\big]\big)^{1/p} - 3\eta p\ell\bar{\sigma}\sqrt{d},\end{aligned}$$

which, in conjunction with the bound (C.2), establishes the second claim.

Proof of equation (C.2)**.** Applying Minkowski's inequality to the recursive relation (5.2) yields (for any $p \ge 2$) a bound on the p-th conditional moment:

$$\begin{aligned}\big(\mathbb{E}\big[\|\Delta_{t+\ell+1}\|_2^p\big]\big)^{1/p} &\le \big(\mathbb{E}\big[\|(I-\eta L_{t+\ell+1})\Delta_{t+\ell}\|_2^p\big]\big)^{1/p} \\ &\quad + \eta\big(\mathbb{E}\big[\|v_{t+\ell}+\zeta_{t+\ell+1}\|_2^p\big]\big)^{1/p}. \qquad \text{(C.3)}\end{aligned}$$

Our next step is to bound the two terms above.

Substituting into the recursive relation (C.3), and applying Minkowski's inequality, we find that the moment $(\mathbb{E}[\|\Delta_{t+\ell+1}\|_2^p])^{1/p}$ is upper bounded by

$$(1+\eta\gamma_{\max})\big(\mathbb{E}\big[\|\Delta_{t+\ell}\|_2^p\big]\big)^{1/p} + \eta\sigma_L d\big(\mathbb{E}\big[\|\Delta_{t+\ell}\|_2^p\big]\big)^{1/p} + 2\eta p\sqrt{d}\bar{\sigma}.$$

Solving this recursive inequality leads to

$$\big(\mathbb{E}\big[\|\Delta_{t+\ell}\|_2^p\big]\big)^{1/p} \le \exp\big(\eta\ell(\gamma_{\max}+\sigma_L d)\big)\big(\mathbb{E}\big[\|\Delta_t\|_2^p\big]\big)^{1/p} + 2\eta p\ell\sqrt{d}\bar{\sigma}.$$

For any stepsize $\eta \in (0, \frac{1}{(\gamma_{\max}+\sigma_L d)\ell}]$, we have

$$\big(\mathbb{E}\big[\|\Delta_{t+\ell}\|_2^p\big]\big)^{1/p} \le e\big(\mathbb{E}\big[\|\Delta_t\|_2^p\big]\big)^{1/p} + 6\eta p\ell\sqrt{d}\bar{\sigma},$$

which establishes the claim.

C.3. Proof of Lemma 6

For notational simplicity, we extend the process $(\Delta_t)_{t\geq 0}$ 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 (5.19) both hold true for the extended process, with index set $t \in \mathbb{Z}$. Moreover, as in the proof of Lemma 5, for each $p \geq 2$, we have the moment bound

$$\begin{aligned}\big(\mathbb{E}\big[\|\Delta_{t+\ell+1} - \Delta_t\|_2^p\big]\big)^{1/p} &\leq \big(\mathbb{E}\big[\|\Delta_{t+\ell} - \Delta_t\|_2^p\big]\big)^{1/p} + \eta\big(\mathbb{E}\big[\|L_{t+\ell+1}\Delta_{t+\ell}\|_2^p\big]\big)^{1/p} \\ &\quad + \eta\big(\mathbb{E}\big[\|\nu_{t+\ell} + \zeta_{t+\ell+1}\|_2^p\big]\big)^{1/p}.\end{aligned}$$

Our next step is to exploit the coarse bound (5.19) so as to obtain upper bounds on the second term $(\mathbb{E}[\|L_{t+\ell+1}\Delta_{t+\ell}\|_2^p])^{1/p}$. 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

$$\begin{aligned}&\big(\mathbb{E}\big[\|L_{t+\ell+1}\Delta_{t+\ell}\|_2^p\big]\big)^{1/p} \\ &\leq \big(\mathbb{E}\big[\|L_{t+\ell+1}\Delta_{t+\ell-\tau}\|_2^p\big]\big)^{1/p} + \big(\mathbb{E}\big[\|L_{t+\ell+1}(\Delta_{t+\ell} - \Delta_{t+\ell-\tau})\|_2^p\big]\big)^{1/p}. \quad \text{(C.4)}\end{aligned}$$

The latter term of the bound (C.4) can be controlled through Assumption 4:

$$\|L_{t+\ell+1}(s_{t+\ell})(\Delta_{t+\ell} - \Delta_{t+\ell-\tau})\|_2 \leq (\gamma_{\max} + \sigma_L d)\|\Delta_{t+\ell} - \Delta_{t+\ell-\tau}\|_2, \quad \text{a.s.}$$

The distance $\|\Delta_{t+\ell} - \Delta_{t+\ell-\tau}\|_2$ is controlled via the coarse bound (5.19). Putting together the pieces, we find that

$$\begin{aligned}&\big(\mathbb{E}\big[\|L_{t+\ell+1}(\Delta_{t+\ell} - \Delta_{t+\ell-\tau})\|_2^p\big]\big)^{1/p} \\ &\quad \leq \eta\big(\gamma_{\max} + \sigma_L d\big) \cdot \big(\omega_p\big(\mathbb{E}\big[\|\Delta_{t+\ell-\tau}\|_2^p\big]\big)^{1/p} + \beta_p\bar{\sigma}\big). \quad \text{(C.5)}\end{aligned}$$

In order to bound the former term $(\mathbb{E}[\|L_{t+\ell+1}\Delta_{t+\ell-\tau}\|_2^p])^{1/p}$ in the bound (C.4), we invoke Lemma 4 and obtain a random variable $\tilde{s}_{t+\ell}$ such that

$$\tilde{s}_{t+\ell} \mid \mathcal{F}_{t+\ell-\tau} \sim \xi \quad \text{and} \quad \big(\mathbb{E}\big[\rho(s_{t+\ell}, \tilde{s}_{t+\ell-\tau})^p \mid \mathcal{F}_{t+\ell-\tau}\big]\big)^{1/p} \leq c_0 \cdot 2^{1-\frac{\tau}{2t_{\text{mix}}p}}. \quad \text{(C.6)}$$

By Assumption 2, we have the bounds

$$\mathbb{E}\big[\|Z_{t+\ell+1}\Delta_{t+\ell-\tau}\|_2^p \mid \mathcal{F}_{t+\ell-\tau}\big] \leq (p\sqrt{d}\sigma_L)^p\|\Delta_{t+\ell-\tau}\|_2^p \quad \text{(C.7a)}$$

and

$$\mathbb{E}\big[\|(\boldsymbol{L}(\tilde{s}_{t+\ell-\tau}) - \bar{L}) \cdot \Delta_{t+\ell-\tau}\|_2^p \mid \mathcal{F}_{t+\ell-\tau}\big] \leq (p\sqrt{d}\sigma_L)^p\|\Delta_{t+\ell-\tau}\|_2^p. \quad \text{(C.7b)}$$

Invoking the moment bound (C.6) and using the Lipschitz condition 4, we find that

$$\begin{aligned}
&\mathbb{E}\big[\|(\boldsymbol{L}(\tilde{s}_{t+\ell-\tau}) - \boldsymbol{L}(s_{t+\ell-\tau}))\cdot\Delta_{t+\ell-\tau}\|_2^p \mid \mathcal{F}_{t+\ell-\tau}\big] \\
&\quad\le \mathbb{E}\big[\|\boldsymbol{L}(\tilde{s}_{t+\ell-\tau}) - \boldsymbol{L}(s_{t+\ell-\tau})\|_{\mathrm{op}}^p \mid \mathcal{F}_{t+\ell-\tau}\big]\cdot\|\Delta_{t+\ell-\tau}\|_2^p \\
&\quad\le \big(\sigma_L c_0 d\cdot 2^{1-\frac{\tau}{2t_{\mathrm{mix}}p}}\|\Delta_{t+\ell-\tau}\|_2\big)^p. \qquad \text{(C.7c)}
\end{aligned}$$

Finally, we have the operator norm bound

$$\|\bar{L}\Delta_{t+\ell-\tau}\|_2 \le \gamma_{\max}\|\Delta_{t+\ell-\tau}\|_2. \qquad \text{(C.7d)}$$

Collecting the results from equations (C.7) (a)–(d), we arrive at the bound

$$\begin{aligned}
&\big(\mathbb{E}\big[\|L_{t+\ell+1}\Delta_{t+\ell-\tau}\|_2^p \mid \mathcal{F}_{t+\ell-\tau}\big]\big)^{1/p} \\
&\quad\le \big(2p\sqrt{d}\sigma_L + \gamma_{\max} + \sigma_L c_0 d\cdot 2^{1-\frac{\tau}{2t_{\mathrm{mix}}p}}\big)\|\Delta_{t+\ell-\tau}\|_2. \qquad \text{(C.8)}
\end{aligned}$$

According to Lemma 5, given a stepsize bounded as

$$\eta \le (6(\gamma_{\max} + \sigma_L d)\tau)^{-1},$$

we have

$$\big(\mathbb{E}\|\Delta_{t+\ell-\tau}\|_2^p\big)^{1/p} \le 2\big(\mathbb{E}\|\Delta_{t+\ell}\|_2^p\big)^{1/p} + 12\eta p\tau\bar{\sigma}\sqrt{d}.$$

Collecting the bounds (C.5) and (C.8), and substituting into the decomposition (C.4), for $\tau \ge 2t_{\mathrm{mix}}p\log(c_0 d)$, we arrive at the inequality

$$\begin{aligned}
&\big(\mathbb{E}\big[\|L_{t+\ell+1}\Delta_{t+\ell}\|_2^p\big]\big)^{1/p} \\
&\le 2\big(\big(p\sqrt{d}\sigma_L + \gamma_{\max}\big) + \eta\omega_p\big(\gamma_{\max} + \sigma_L d\big)\big)\cdot\Big(\big(\mathbb{E}\|\Delta_{t+\ell}\|_2^p\big)^{1/p} + \eta p\tau\sqrt{d}\bar{\sigma}\Big) \\
&\quad + \eta\big(\gamma_{\max} + \sigma_L d\big)\beta_p\bar{\sigma}.
\end{aligned}$$

By following the derivation in the proof of Lemma 5, we can show that the third term is upper bounded as

$$\big(\mathbb{E}\big[\|v_{t+\ell} + \zeta_{t+\ell+1}\|_2^p\big]\big)^{1/p} \le 2p\bar{\sigma}\sqrt{d}.$$

Substituting back into the original decomposition, we find that the difference in moments

$$D := (\mathbb{E}[\|\Delta_{t+\ell+1} - \Delta_t\|_2^p])^{1/p} - (\mathbb{E}[\|\Delta_{t+\ell} - \Delta_t\|_2^p])^{1/p}$$

is bounded as

$$\begin{aligned}
D &\le 2\eta\big\{\big(p\sqrt{d}\sigma_L + \gamma_{\max}\big) + \eta\omega_p\big(\gamma_{\max} + \sigma_L d\big)\big\}\cdot\Big(\big(\mathbb{E}\|\Delta_{t+\ell}\|_2^p\big)^{1/p} + \eta p\tau\sqrt{d}\bar{\sigma}\Big) \\
&\quad + \big(2\eta p\sqrt{d} + \eta^2\big(\gamma_{\max} + \sigma_L d\big)\beta_p\big).
\end{aligned}$$

Lemma 5 implies that

$$(\mathbb{E}[\|\Delta_{t+\ell}\|_2^p])^{1/p} \le e(\mathbb{E}[\|\Delta_t\|_2^p])^{1/p} + 6\eta p\ell\sqrt{d}\bar{\sigma},$$

and solving the recursion, we arrive at the bound

$$\begin{aligned}
&\big(\mathbb{E}\big[\|\Delta_{t+\ell} - \Delta_t\|_2^p\big]\big)^{1/p} \\
&\le 12\eta\ell\big(\big(p\sqrt{d}\sigma_L + \gamma_{\max}\big) + \eta\omega_p\big(\gamma_{\max} + \sigma_L d\big)\big) \cdot \Big(\big(\mathbb{E}\|\Delta_t\|_2^p\big)^{1/p} + \eta p(\tau+\ell)\sqrt{d}\bar{\sigma}\Big) \\
&\quad + \big(2\eta p\sqrt{d} + \eta^2\big(\gamma_{\max} + \sigma_L d\big)\beta_p\big)\ell\bar{\sigma} \\
&\le \eta\Big(12\big(p\sqrt{d}\sigma_L + \gamma_{\max}\big)\ell + \frac{\omega_p}{2}\Big)\Big(\big(\mathbb{E}\|\Delta_t\|_2^p\big)^{1/p} + \eta p(\tau+\ell)\sqrt{d}\bar{\sigma}\Big) \\
&\quad + \eta\left(2p\ell\sqrt{d} + \frac{1}{2}\beta_p\right)\bar{\sigma}
\end{aligned}$$

for any $\tau \ge 2t_{\mathrm{mix}}\, p\log(c_0 d)$ and stepsize choice

$$\eta \le \frac{c}{48(\gamma_{\max} + \sigma_L d)}.$$

D. Auxiliary results underlying Theorem 1

In this appendix, we prove two auxiliary lemmas that were used in the proof of Theorem 1.

D.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 $\tilde{s}_{t-k_m}$ such that $\tilde{s}_{t-k_m} \mid \mathcal{F}_{t-k_m-\tau} \sim \xi$, and

$$\mathbb{E}[\rho(s_{t-k_m}, \tilde{s}_{t-k_m}) \mid \mathcal{F}_{t-\tau-k_m}] \le c_0 \cdot 2^{1-\frac{\tau}{t_{\mathrm{mix}}}}.$$

By Assumption 1, conditionally on the pair of states $(s_{t-k_m}, \tilde{s}_{t-k_m})$, we have the following bound for $j \in [m]$:

$$\mathcal{W}_{\rho,1}\big(P^{k_j-k_{j-1}}\delta_{s_{t-k_j}}, P^{k_j-k_{j-1}}\delta_{\tilde{s}_{t-k_j}}\big) \le c_0 \cdot \rho\big(s_{t-k_j}, \tilde{s}_{t-k_j}\big), \quad \text{a.s.}$$

Consequently, there exists a sequence of random variables $(\tilde{s}_{t-k_j})_{0\le j\le m-1}$ such that the following relations hold true for $j = 1, 2, \ldots, m$:

$$\tilde{s}_{t-k_{j-1}} \mid \mathcal{F}_{t-k_m} \sim P^{k_j-k_{j-1}}\delta_{\tilde{s}_{t-k_j}}$$

and

$$\mathbb{E}\big[\rho\big(\tilde{s}_{t-k_{j-1}}, s_{t-k_{j-1}}\big) \mid \mathcal{F}_{t+k-\ell}\big] \le c_0^{m+1-j} \cdot \rho\big(s_{t-k_m}, \tilde{s}_{t-k_m}\big).$$

Based on above construction, we consider the following decomposition:

$$\begin{aligned}\left(\prod_{j=0}^{m} N_{t-k_j}\right)\Delta_{t-k_m} &= \left(\prod_{j=0}^{m} N(s_{t-k_j}) - \prod_{j=0}^{m} N(\tilde{s}_{t-k_j})\right)\Delta_{t-k_m-\tau} \\ &\quad + \left(\prod_{j=0}^{m} N(\tilde{s}_{t-k_j})\right)\cdot \Delta_{t-k_m-\tau} \\ &\quad + \left(\prod_{j=0}^{m} N(s_{t-k_j})\right)\cdot \left(\Delta_{t-k_m} - \Delta_{t-\tau-k_m}\right) \\ &:= Q_1(t) + Q_2(t) + Q_3(t). \end{aligned} \tag{D.1}$$

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

$$\begin{aligned}&\prod_{j=0}^{m} N(s_{t-k_j}) - \prod_{j=0}^{m} N(\tilde{s}_{t-k_j}) \\ &\quad = \sum_{q=0}^{m}\left(\prod_{j=0}^{q-1} N(s_{t-k_j})\right)\cdot \left(\boldsymbol{L}(s_{t-k_q}) - \boldsymbol{L}(\tilde{s}_{t-k_q})\right)\cdot \left(\prod_{j=q+1}^{m} N(\tilde{s}_{t-k_j})\right).\end{aligned}$$

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 $\rho(s_{t-k_q}, \tilde{s}_{t-k_q})$ and obtain the bound

$$\begin{aligned}&\mathbb{E}\left[\left\|\prod_{j=0}^{m} N(s_{t-k_j}) - \prod_{j=0}^{m} N(\tilde{s}_{t-k_j})\right\|_{\mathrm{op}}^2 \mid \mathcal{F}_{t-k_m-\tau}\right] \\ &\quad \le (m+1)\cdot(\sigma_L d)^m \sum_{q=0}^{m} \mathbb{E}\left[\|\boldsymbol{L}(s_{t-k_q}) - \boldsymbol{L}(\tilde{s}_{t-k_q})\|_{\mathrm{op}}^2 \mid \mathcal{F}_{t-k_m-\tau}\right] \\ &\quad \le (m+1)^2 (c_0\sigma_L d)^{m+1}\cdot 2^{-\frac{\tau}{t_{\mathrm{mix}}}}.\end{aligned}$$

Applying the bound on $\|\Delta_{t-\tau}\|_2$ in Proposition 1 and taking $\tau \ge 3m t_{\mathrm{mix}} p \log(c_0 d n)$, we find that

$$\begin{aligned}&\mathbb{E}\left[\|Q_1(t)\|_2^2\right] \\ &\quad \le \mathbb{E}\left[\mathbb{E}\left[\left\|\prod_{j=0}^{m} N(s_{t-k_j}) - \prod_{j=0}^{m} N(\tilde{s}_{t-k_j})\right\|_{\mathrm{op}}^2 \mid \mathcal{F}_{t-k_m-\tau}\right]\cdot \|\Delta_{t-\tau-k_m}\|_2^2\right] \\ &\quad \le (m+1)^2 (c_0\sigma_L d)^{m+1}\cdot 2^{-\frac{\tau}{t_{\mathrm{mix}}}} c\bar{\sigma}^2 \frac{\eta\tau d\,\log^2 n}{1-\kappa} \;\le\; \frac{\sigma_L^{m+1}}{n^2}\bar{\sigma}^2.\end{aligned} \tag{D.2}$$

Now, we turn to bounding the term $Q_2(t)$. First, we note that

$$\begin{aligned}\mathbb{E}\big[\|Q_2(t)\|_2^2\big] &\le \mathbb{E}\left[\left\|\prod_{j=0}^{m-1} N(\tilde{s}_{t-k_j})\right\|_{\mathrm{op}}^2 \cdot \|N(\tilde{s}_{t-k_m})\Delta_{t-k_m-\tau}\|_2^2\right] \\ &\le (\sigma_L d)^{2m}\mathbb{E}\big[\|N(\tilde{s}_{t-k_m})\Delta_{t-k_m-\tau}\|_2^2\big] \\ &\le (\sigma_L d)^{2m}\cdot\sigma_L^2 d\cdot\mathbb{E}\big[\|\Delta_{t-k_m-\tau}\|_2^2\big].\end{aligned}$$

By Proposition 1, for $t \ge n_0$ and $n_0 \ge 2(\tau + k_m)$, we have

$$\mathbb{E}[\|\Delta_{t-k_m-\tau}\|_2^2] \le \frac{c\eta}{1-\kappa} t_{\mathrm{mix}} d\bar{\sigma}^2.$$

If $m = 0$, we have that $\mathbb{E}[N(\tilde{s}_{t+\tau}) \mid \mathcal{F}_t] = 0$ almost surely for each $t \ge n_0$. For $m \ge 1$, the conditional unbiasedness does not hold true, but we still have the following upper bound on the bias:

$$\begin{aligned}&\left\|\mathbb{E}\left[\prod_{j=0}^{m} N(\tilde{s}_{t+k_m+\tau-k_j}) \mid \mathcal{F}_t\right]\right\|_{\mathrm{op}} \\ &= \sup_{u,v\in\mathbb{S}^{d-1}} \mathbb{E}\left[\left\langle u, \prod_{j=0}^{m} N(\tilde{s}_{t+k_m+\tau-k_j})v\right\rangle\right] \\ &\le \sup_{u,v\in\mathbb{S}^{d-1}} \mathbb{E}\left[\|N(\tilde{s}_{t+k_m+\tau})^\top u\|_2 \cdot \left\|\prod_{j=1}^{m-1} N(\tilde{s}_{t+k_m+\tau-k_j})\right\|_{\mathrm{op}} \cdot \|N(\tilde{s}_{t+\tau})v\|_2\right] \\ &\le (\sigma_L d)^{m-1} \sup_{u,v\in\mathbb{S}^{d-1}} \sqrt{\mathbb{E}\|N(\tilde{s}_{t+k_m+\tau})^\top u\|_2^2 \cdot \mathbb{E}\|N(\tilde{s}_{t+\tau})v\|_2^2} \\ &\le (\sigma_L d)^{m-1}\cdot\sigma_L^2 d.\end{aligned}$$

Denote $Y_t := \prod_{j=0}^m N(s_{t-k_j})$ and $\widetilde{Y}_t := \prod_{j=0}^m N(\tilde{s}_{t-k_j})$ for any $t \ge k_m$. We have the expansion

$$\begin{aligned}&\mathbb{E}\left[\left\|\sum_{t=n_0}^{n-1} Q_2(t)\right\|_2^2\right] \\ &\le 2\mathbb{E}\left[\left\|\sum_{t=n_0}^{n-1} \mathbb{E}[\widetilde{Y}_t]\cdot\Delta_{t-k_m-\tau}\right\|_2^2\right] + 2\mathbb{E}\left[\left\|\sum_{t=n_0}^{n-1} (\widetilde{Y}_t - \mathbb{E}[\widetilde{Y}_t])\cdot\Delta_{t-k_m-\tau}\right\|_2^2\right] \\ &\le 2n(d^m\sigma_L^{m+1})^2 \sum_{t=n_0}^{n} \mathbb{E}\|\Delta_{t-k_m-\tau}\|_2^2 \\ &\quad + 2\sum_{n_0\le s,t\le n-1} \mathbb{E}\big[\langle(\widetilde{Y}_t - \mathbb{E}[\widetilde{Y}_t])\cdot\Delta_{t-k_m-\tau}, (\widetilde{Y}_s - \mathbb{E}[\widetilde{Y}_s])\cdot\Delta_{s-k_m-\tau}\rangle\big].\end{aligned}$$

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 right-hand side.

For $t > s + \tau + k_m$, we have the relations

$$\mathbb{E}\big[(\widetilde{Y}_t - \mathbb{E}[\widetilde{Y}_t]) \cdot \Delta_{t-k_m-\tau} \mid \widetilde{\mathcal{F}}_{t-k_m-\tau}\big] = 0 \ \text{ and } \ (\widetilde{Y}_s - \mathbb{E}[\widetilde{Y}_s]) \cdot \Delta_{s-k_m-\tau} \in \widetilde{\mathcal{F}}_{t-k_m-\tau},$$

meaning that the product term vanishes when $|s - t| > \tau + k_m$. Therefore, we arrive at the bound

$$\begin{aligned}
&\mathbb{E}\left[\left\|\sum_{t=n_0}^{n-1} Q_2(t)\right\|_2^2\right] \\
&\leq \begin{cases} \left(2n^2\left(d^m\sigma_L^{m+1}\right)^2 + 4n(k_m+\tau)\cdot(\sigma_L d)^{2m}\cdot\sigma_L^2 d\right)\cdot\frac{c\eta}{1-\kappa}dt_{\text{mix}}\overline{\sigma}^2, & m \geq 1, \\ 4n\tau\sigma_L^2 d\cdot\frac{c\eta}{1-\kappa}dt_{\text{mix}}\overline{\sigma}^2, & m = 0. \end{cases}
\end{aligned} \tag{D.3}$$

Now, we turn to the last term in the decomposition (D.1). 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:

$$\begin{aligned}
\mathbb{E}\left[\left\|\sum_{t=n_0}^{n-1} Q_3(t)\right\|_2^2\right] \leq\ & 4\eta^2\mathbb{E}\left[\left\|\sum_{t=n_0}^{n}\left\{Y_t\cdot\left(\sum_{\ell=1}^{\tau} Z_{t-k_m-\ell+1}\Delta_{t-k_m-\ell}\right)\right\}\right\|_2^2\right] \\
&+ 4\eta^2\mathbb{E}\left[\left\|\sum_{t=n_0}^{n}\left\{Y_t\cdot\left(\bar{L}\sum_{\ell=1}^{\tau}\Delta_{t-k_m-\ell}\right)\right\}\right\|_2^2\right] \\
&+ 4\eta^2\mathbb{E}\left[\left\|\sum_{t=n_0}^{n}\left\{Y_t\cdot\left(\sum_{\ell=1}^{\tau} N_{t-k_m-\ell}\Delta_{t-k_m-\ell}\right)\right\}\right\|_2^2\right] \\
&+ 4\eta^2\mathbb{E}\left[\left\|\sum_{t=n_0}^{n}\left\{Y_t\cdot\left(\sum_{\ell=1}^{\tau}(\nu_{t-k_m-\ell}+\zeta_{t-k_m-\ell+1})\right)\right\}\right\|_2^2\right].
\end{aligned}$$

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

$$\begin{aligned}
&\mathbb{E}\left[\left\|\sum_{t=n_0}^{n} Y_t Z_{t-k_m-\ell+1}(s_{t-k_m-\ell})\Delta_{t-k_m-\ell}\right\|_2^2\right] \\
&\quad= \sum_{t=n_0}^{n-1}\mathbb{E}\big[\|Y_t Z_{t-k_m-\ell+1}(s_{t-k_m-\ell})\Delta_{t-k_m-\ell}\|_2^2\big]
\end{aligned}$$

$$\leq (\sigma_L d)^{2(m+1)} \sum_{t=n_0}^{n-1} \mathbb{E}\big[\|Z_{t-k_m-\ell+1}(s_{t-k_m-\ell})\Delta_{t-k_m-\ell}\|_2^2\big]$$
$$\leq \sigma_L^{2m+4} d^{2m+3} n \cdot \frac{c\eta}{1-\kappa} d t_{\text{mix}} \overline{\sigma}^2.$$

From the Lipschitz condition 4 and the boundedness condition (3) on the metric space, it follows that $\|\!|Y_t|\!\|_{\text{op}} \leq (\sigma_L d)^{m+1}$ almost surely. Using this fact, the second term can be bounded as

$$\mathbb{E}\left[\left\|\sum_{t=n_0}^{n}\left\{Y_t \cdot \left(\bar{L}\sum_{\ell=1}^{\tau}\Delta_{t-k_m-\ell}\right)\right\}\right\|_2^2\right] \leq n\tau(\sigma_L d)^{2m+2}\gamma_{\max}^2 \sum_{t=n_0}^{n-1}\sum_{\ell=1}^{\tau}\mathbb{E}\|\Delta_{t-k_m-\ell}\|_2^2$$
$$\leq n^2\tau^2(\sigma_L d)^{2m+2}\gamma_{\max}^2 \cdot \frac{c\eta}{1-\kappa} d t_{\text{mix}}\overline{\sigma}^2.$$

Collecting equations (D.2) and (D.3) as well as the above bounds for Q_3, we arrive at the upper bound

$$\mathbb{E}\left[\left\|\sum_{t=n_0}^{n-1}\left(\prod_{j=0}^{m} N_{t-k_j}\right)\Delta_{t-k_m}\right\|_2^2\right] \leq \sum_{j=1}^{3} T_j,$$

where

$$T_1 := n^2 d^{2m}\sigma_L^{2m+2}\big(1+\eta^2\tau^2\gamma_{\max}^2 d^2\sigma_L^2 + \eta^2\tau^2 d^3\sigma_L^2/n\big)\cdot\frac{c\eta}{1-\kappa} d t_{\text{mix}}\overline{\sigma}^2,$$
$$T_2 := 4\eta^2\mathbb{E}\left[\left\|\sum_{t=n_0}^{n}\left\{Y_t\left(\sum_{\ell=1}^{\tau} N_{t-k_m-\ell}\Delta_{t-k_m-\ell}\right)\right\}\right\|_2^2\right],$$
$$T_3 := 4\eta^2\mathbb{E}\left[\left\|\sum_{t=n_0}^{n}\left\{Y_t\left(\sum_{\ell=1}^{\tau}(v_{t-k_m-\ell}+\zeta_{t-k_m-\ell+1})\right)\right\}\right\|_2^2\right].$$

In the special case of $m = 0$, we have

$$\mathbb{E}\left[\left\|\sum_{t=n_0}^{n-1} N_t\Delta_t\right\|_2^2\right] \leq c\sigma_L^2 d\cdot\big(n\tau + n^2\eta^2\sigma_L^2 d\tau^2\big)\frac{c\eta}{1-\kappa} d t_{\text{mix}}\overline{\sigma}^2$$
$$+ 4\eta^2\tau\sum_{k_1=1}^{\tau}\mathbb{E}\left[\left\|\sum_{t=n_0}^{n} N_t N_{t-k_1}\Delta_{t-k_1}\right\|_2^2\right]$$
$$+ 4\eta^2\tau\sum_{k_1=1}^{\tau}\mathbb{E}\left[\left\|\sum_{t=n_0}^{n} N_t\big(v_{t-k_1}+\zeta_{t-k_1+1}\big)\right\|_2^2\right],$$

which completes the proof of this lemma.

D.2. Proof of Lemma 10

We study the bias and variance of the summation separately. For the bias term, we have

$$\begin{aligned}
&\left\| \mathbb{E}\left[\left(\prod_{j=0}^{m-1} N_{t-k_j} \right) \left(v_{t-k_m} + \zeta_{t-k_m+1} \right) \right] \right\|_2 \\
&\quad = \sup_{z \in \mathbb{S}^{d-1}} \mathbb{E}\left[\left\langle \left(\prod_{j=0}^{m-1} N_{t-k_j} \right) \left(v_{t-k_m} + \zeta_{t-k_m+1} \right), z \right\rangle \right] \\
&\quad \overset{(i)}{\leq} \sup_{z \in \mathbb{S}^{d-1}} \sqrt{\mathbb{E}\| N_t^\top z \|_2^2} \cdot \left[\mathbb{E}\left\| \left(\prod_{j=1}^{m-1} N_{t-k_j} \right) \left(v_{t-k} + \zeta_{t-k+1} \right) \right\|_2^2 \right]^{1/2} \\
&\quad \overset{(ii)}{\leq} \sigma_L \sqrt{d} \cdot (\sigma_L d)^{m-1} \cdot 2\bar{\sigma}\sqrt{d} = 2(\sigma_L d)^m \bar{\sigma}, \qquad \text{(D.4)}
\end{aligned}$$

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

$$\lambda_t := \left(\prod_{j=0}^{m-1} N_{t-k_j} \right) \left(v_{t-k_m} + \zeta_{t-k_m+1} \right) - \mathbb{E}\left[\left(\prod_{j=0}^{m-1} N_{t-k_j} \right) \left(v_{t-k_m} + \zeta_{t-k_m+1} \right) \right].$$

We have

$$\begin{aligned}
\mathbb{E}\left[\| \lambda_t \|_2^2 \right] &\leq \mathbb{E}\left[\left(\prod_{j=0}^{m-1} \| N_{t-k_j} \|_{\mathrm{op}}^2 \right) \cdot \| v_{t-k_m} + \zeta_{t-k_m+1} \|_2^2 \right] \\
&\leq (\sigma_L d)^{2m} \cdot \mathbb{E}\left[\| v_{t-k} + \zeta_{t-k+1} \|_2^2 \right] \leq d^{2m+1} \sigma_L^{2m} \bar{\sigma}^2.
\end{aligned}$$

For integers $t \geq 0$ and $\ell \geq k_m$, by Lemma 4, there exists a random variable $\tilde{s}_{t+\ell-k_m}$ such that $\tilde{s}_{t+\ell-k_m} \mid \mathcal{F}_t \sim \xi$, and that $\mathbb{E}[\rho(s_{t+\ell-k_m}, \tilde{s}_{t+\ell-k_m}) \mid \mathcal{F}_t] \leq c_0 \cdot 2^{1-\frac{\ell-k_m}{t_{\mathrm{mix}}}}$. By Assumption 1, conditionally on the pair of states $(s_{t+\ell-k_m}, \tilde{s}_{t+\ell-k_m})$, we have the following bound for $j \in [m]$:

$$\mathcal{W}_{\rho,1}\left(P^{k_j - k_{j-1}} \delta_{s_{t+\ell-k_j}}, P^{k_j - k_{j-1}} \delta_{\tilde{s}_{t+\ell-k_j}} \right) \leq c_0 \cdot \rho\left(s_{t+\ell-k_j}, \tilde{s}_{t+\ell-k_j} \right), \quad \text{a.s.}$$

Consequently, there exists a sequence of random variables $(\tilde{s}_{t+\ell-k_j})_{0 \leq j \leq m-1}$ such that the following relations hold true for $j = 1, 2, \ldots, m$:

$$\tilde{s}_{t+\ell-k_{j-1}} \mid \mathcal{F}_{t+\ell-k_m} \sim P^{k_j - k_{j-1}} \delta_{\tilde{s}_{t+\ell-k_j}}$$

and

$$\mathbb{E}\left[\rho\left(\tilde{s}_{t+\ell-k_{j-1}}, s_{t+\ell-k_{j-1}} \right) \mid \mathcal{F}_{t+\ell-k_m} \right] \leq c_0^{m+1-j} \cdot \rho\left(s_{t+\ell-k_m}, \tilde{s}_{t+\ell-k_m} \right).$$

Given the random variables constructed above, we can then construct the proxy random variable for $\lambda_{t+\ell}$:

$$\widetilde{\lambda}_{t+\ell} := \left(\prod_{j=0}^{m-1} N(\tilde{s}_{t+\ell-k_j})\right)\left(\nu(\tilde{s}_{t+\ell-k_m}) + \zeta_{t+\ell-k_m+1}(\tilde{s}_{t+\ell-k_m})\right) - \mathbb{E}\left[\left(\prod_{j=0}^{m-1} N_{t-k_j}\right)\left(\nu_{t-k_m} + \zeta_{t-k_m+1}\right)\right].$$

By stationarity, we have $\mathbb{E}[\widetilde{\lambda}_{t+\ell} \mid \widetilde{\mathcal{F}}_t] = 0$ almost surely. In order to bound the difference, we note the telescope relation $\widetilde{\lambda}_{t+\ell} - \lambda_{t+\ell} = \sum_{q=0}^{m-1} E_q^{(\mathrm{mix})} + \bar{E}^{(\mathrm{mix})}$, where

$$E_q^{(\mathrm{mix})} := \left(\prod_{j=0}^{q-1} N(s_{t+\ell-k_j})\right)\left(\bar{L}(\tilde{s}_{t+\ell-k_q}) - \boldsymbol{L}(s_{t+\ell-k_q})\right) \times \left(\prod_{j=q+1}^{m-1} N(\tilde{s}_{t+\ell-k_j})\right)\left(\nu(\tilde{s}_{t+\ell-k_m}) + \zeta_{t+\ell-k_m+1}(\tilde{s}_{t+\ell-k_m})\right)$$

and

$$\bar{E}^{(\mathrm{mix})} := \prod_{j=0}^{m-1} N(s_{t+\ell-k_j}) \cdot \left(\nu(\tilde{s}_{t+\ell-k_m}) + \zeta_{t+\ell-k_m+1}(\tilde{s}_{t+\ell-k_m}) - \nu(s_{t+\ell-k_m}) + \zeta_{t+\ell-k_m+1}(s_{t+\ell-k_m})\right).$$

Using the Wasserstein distance bounds and Lipschitz condition 4, we find that the conditional expectation $A = \mathbb{E}[\|E_q^{(\mathrm{mix})}\|_2 \mid \mathcal{F}_t]$ is bounded as

$$\begin{aligned} A &\le (\sigma_L d)^{m-1}\mathbb{E}\big[\|\boldsymbol{L}(s_{t+\ell-k_q}) - \boldsymbol{L}(\tilde{s}_{t+\ell-k_q})\|_{\mathrm{op}} \\ &\qquad \cdot \|\nu(\tilde{s}_{t+\ell-k}) + \zeta_{t+\ell-k+1}(\tilde{s}_{t+\ell-k})\|_2 \mid \widetilde{\mathcal{F}}_t\big] \\ &\le (\sigma_L d)^m \sqrt{\mathbb{E}[\rho(s_{t+\ell-k_q}, \tilde{s}_{t+\ell-k_q})^2 \mid \widetilde{\mathcal{F}}_t]} \\ &\quad \cdot \sqrt{\mathbb{E}[\|\nu(\tilde{s}_{t+\ell-k}) + \zeta_{t+\ell-k+1}(\tilde{s}_{t+\ell-k})\|_2^2 \mid \widetilde{\mathcal{F}}_t]} \\ &\le (\sigma_L d)^m c_0 \cdot 2^{1-\frac{\ell-k_q}{2t_{\mathrm{mix}}}} \cdot 2d\bar{\sigma}, \end{aligned}$$

and the conditional expectation $B = \mathbb{E}[\|\bar{E}^{(\mathrm{mix})}\|_2 \mid \mathcal{F}_t]$ is bounded as

$$\begin{aligned} B &\le (\sigma_L d)^m \Big(\sqrt{\mathbb{E}\big[\|\zeta_{t+\ell-k+1}(s_{t+\ell-k}) - \zeta_{t+\ell-k+1}(\tilde{s}_{t+\ell-k})\|_2^2 \mid \mathcal{F}_t\big]} \\ &\qquad + \sqrt{\mathbb{E}\big[\|\nu(s_{t+\ell-k}) - \nu(\tilde{s}_{t+\ell-k})\|_2^2 \mid \mathcal{F}_t\big]}\Big) \\ &\le (\sigma_L d)^m d\bar{\sigma} c_0 \cdot 2^{1-\frac{\ell-k_m}{2t_{\mathrm{mix}}}}. \end{aligned}$$

Consequently, we can bound the cross term as

$$\begin{aligned}\mathbb{E}\big[\langle \lambda_t,\, \lambda_{t+\ell}\rangle\big] &= \mathbb{E}\big[\big\langle \lambda_t,\, \mathbb{E}\big[\widetilde{\lambda}_{t+\ell} \mid \mathcal{F}_t\big]\big\rangle\big] + \mathbb{E}\big[\big\langle \lambda_t,\, \mathbb{E}\big[\lambda_{t+\ell} - \widetilde{\lambda}_{t+\ell} \mid \mathcal{F}_t\big]\big\rangle\big] \\ &\leq 0 + \mathbb{E}\big[\|\lambda_t\|_2 \cdot \mathbb{E}\big[\|\lambda_{t+\ell} - \widetilde{\lambda}_{t+\ell}\|_2 \mid \mathcal{F}_t\big]\big] \\ &\leq 12 c_0 d^{m+1} \sigma_L^m \overline{\sigma} \cdot 2^{-\frac{\ell-k}{2t_{\mathrm{mix}}}} \cdot \sqrt{\mathbb{E}\|\lambda_t\|_2^2} \\ &\leq 12 c_0 d^{2m+2} \sigma_L^{2m} \overline{\sigma}^2 \cdot 2^{-\frac{\ell-k}{2t_{\mathrm{mix}}}}.\end{aligned}$$

Taking $\tau = 16 t_{\mathrm{mix}} \log(c_0 d)$, we can control the cross terms in two different ways

$$\mathbb{E}\big[\langle \lambda_t,\, \lambda_{t+\ell}\rangle\big] \leq \begin{cases} \sqrt{\mathbb{E}\|\lambda_t\|_2^2} \cdot \sqrt{\mathbb{E}\|\lambda_{t+\ell}\|_2^2} \leq d^{2m+1} \sigma_L^{2m} \overline{\sigma}^2, & 0 \leq \ell \leq k_m + \tau, \\ 12 c_0 d^{2m+2} \sigma_L^{2m} \overline{\sigma}^2 \cdot 2^{-\frac{\ell-k}{2t_{\mathrm{mix}}}} \leq d^{2m} \sigma_L^{2m} \overline{\sigma}^2, & \ell \geq k_m + \tau. \end{cases}$$

Summing up these terms yields

$$\begin{aligned}\mathbb{E}\left[\left\|\sum_{t=n_0}^{n-1} \lambda_t\right\|_2^2\right] &= \sum_{t=n_0}^{n-1} \mathbb{E}\|\lambda_t\|_2^2 + 2 \sum_{n_0 \leq t_1 < t_2 \leq n-1} \mathbb{E}\big[\langle \lambda_{t_1},\, \lambda_{t_2}\rangle\big] \\ &\leq (k + \tau + 1) n d^{2m+1} \sigma_L^{2m} \overline{\sigma}^2 + n^2 d^{2m} \sigma_L^{2m} \overline{\sigma}^2.\end{aligned}$$

Combining with the bound (D.4), we find that

$$\begin{aligned}&\mathbb{E}\left[\left\|\sum_{t=n_0}^{n-1} \left(\prod_{j=0}^{m-1} N_{t-k_j}\right)\big(v_{t-k_m} + \zeta_{t-k_m+1}\big)\right\|_2^2\right] \\ &\quad= \left\|\sum_{t=n_0}^{n-1} \mathbb{E}\left[\left(\prod_{j=0}^{m-1} N_{t-k_j}\right)\big(v_{t-k_m} + \zeta_{t-k_m+1}\big)\right]\right\|_2^2 + \mathbb{E}\left[\left\|\sum_{t=n_0}^{n-1} \lambda_t\right\|_2^2\right] \\ &\quad\leq c\big(n^2 + (k_m + \tau) n d\big) \sigma_L^{2m} d^{2m} \overline{\sigma}^2\end{aligned}$$

for a universal constant $c > 0$.

E. 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 that we have a perturbation vector $h_s \in \mathbb{R}^{\mathbb{X}}$. Use these to define the perturbed transition kernel

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

Note that by construction, for any $x \in \mathbb{X}$ and any $h_x \in \mathbb{R}^{\mathbb{X}}$, we have $\mathrm{supp}(P_h(x,\cdot)) = \mathrm{supp}(P_0(x,\cdot))$. 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 $w \mapsto [\langle w, q_x(y)\rangle]_{x,y\in\mathbb{X}}$.

Next, we specify the prior ρ, 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. Indeed, if we denote $\mathcal{L}_h := \mathcal{P}_h - I$ as the generator. The green function $\mathcal{A}_h f$ solves the Poisson equation $-\mathcal{L}_h u = f - \mathbb{E}_{\xi_h}[f(s)]$.

Finally, for any $h \in \mathbb{R}^{\mathbb{X}\times\mathbb{X}}$ and for all $x \in \mathbb{X}$, we define

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

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*

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

See Section E.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 [77]:

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

Also, define the positive-definite matrix

$$\Lambda := \mathbb{E}_{X\sim\xi_0}[\mathrm{cov}_{Y\sim P_0(X,\cdot)}(\boldsymbol{g}_0(Y)\mid X)],$$

and let $\Lambda = UDU^\top$ denote its eigen-decomposition. For a random variable $\psi\sim\mu^{\otimes d}$, define the perturbation parameter

$$w = \frac{1}{\sqrt{n}}UD^{-1/2}\psi, \tag{E.2b}$$

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

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

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

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

where the Green function $\boldsymbol{g}$ is defined in equation (E.1). 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 $\mathcal{P}_\Theta = (\mathbb{P}_\eta : \eta\in\Theta)$ 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 absolutely 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

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

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

Proposition 5 ([30, Theorem 1], special case). *Under the setup above, given a prior distribution ρ with continuously differentiable density and bounded support contained within Θ, let $T : \mathrm{supp}(\rho) \mapsto \mathbb{R}^d$ denote a locally continuously differentiable functional. Then, for any estimator $\widehat{T}$ based on observing X, we have*

$$\mathop{\mathbb{E}}_{\eta\sim\rho}\mathop{\mathbb{E}}_{X\sim p_\eta}\|\widehat{T}(X)-T(\eta)\|_2^2 \geq \frac{\big(\int \mathrm{trace}\,\big(\frac{\partial T}{\partial \eta}(\eta)\big)\rho(\eta)d\eta\big)^2}{\int \mathrm{trace}\,\big(I(\eta)\big)\rho(\eta)d\eta + \int \|\nabla \log \rho(\eta)\|_2^2\rho(\eta)d\eta}.$$

In order to complete the proof, we provide non-asymptotic estimates on the three quantities involved in the right-hand side of Proposition 5. These require a few technical lemmas, whose proofs can be found at the end of the section.

Bounds on the term trace$(\nabla_{\boldsymbol{w}}\overline{\boldsymbol{\theta}})$. We state two technical lemmas that are helpful in bounding this quantity. The first computes the Jacobian matrix of the desired functional $\overline{\theta}(h)$ with respect to the parameter w.

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

$$\begin{aligned}&\nabla_w \overline{\theta}(P_h)\\ &= \mathbb{E}_{X\sim\xi_h}\big[\mathrm{cov}_{Y\sim P_h(X,\cdot)}\big\{\boldsymbol{g}_h(Y)-\mathcal{P}_h\boldsymbol{g}_h(X), \boldsymbol{g}_0(Y)-\mathcal{P}_0\boldsymbol{g}_0(X)\mid X\big\}\big].\end{aligned} \tag{E.3}$$

See Section E.2 for the proof of this lemma. Next, we control the right-hand side of equation (E.3) by replacing $\boldsymbol{g}_h$ with $\boldsymbol{g}_0$.

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

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

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

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

See Section E.3 for the proof of this lemma.

Combining these two lemmas yields

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

Now, given a sample size lower bounded as

$$n \geq \frac{ct_{\text{mix}}^2\sigma_L^2 d^2 \log^2 d}{(1-\kappa)^2} + \frac{2c(1+\sigma_L^2)\overline{\sigma}^2 t_{\text{mix}}^4 d^2}{(1-\kappa)^4 \operatorname{trace}(\Lambda)} \log^6 \frac{d}{\min_x \xi_0(x)},$$

we can conclude that

$$\operatorname{trace}\left(\nabla_w \bar{\theta}\right) \geq \frac{1}{2}\operatorname{trace}(\Lambda) \quad \text{for any } w \text{ in the support of } \rho. \tag{E.4}$$

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

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

See Section E.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 Section E.7 together imply the following bound for any $\delta > 0$:

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

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

$$\max_{s\in\mathbb{X}} \left|\langle \boldsymbol{g}_0(s),\, w\rangle\right| \leq \frac{c\overline{\sigma}t_{\text{mix}}\|w\|_2}{1-\kappa} \log^3\left(d/\min_s \xi_0(s)\right).$$

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 $(\boldsymbol{g}_0(s'))_{s'\in\mathbb{X}}$. So, we have the bound $\max_{s\in\mathbb{X}}\left|\langle \mathcal{P}_0\boldsymbol{g}_0(s),\, w\rangle\right| \leq \max_{s\in\mathbb{X}}\left|\langle \boldsymbol{g}_0(s),\, w\rangle\right| \leq \frac{c\overline{\sigma}t_{\text{mix}}\|w\|_2}{1-\kappa}\log^3(d/\min_s \xi_0(s))$. Putting them together leads to the bound

$$\max_{x\in\mathbb{X}} \|h_x\|_\infty \leq 2c\overline{\sigma}t_{\text{mix}}\|w\|_2 \log^3\left(d/\min_s \xi_0(s)\right).$$

Now, given a sample size

$$n \geq ct_{\text{mix}}^3\overline{\sigma}^2 \cdot \operatorname{trace}(\Lambda) \cdot \log^3 \frac{d}{\min_s \xi_0(s)}, \tag{E.5}$$

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

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

is satisfied so that Lemma 14 guarantees that

$$\begin{aligned} \mathrm{trace}\left(I^{(n)}(w)\right) &\preceq \frac{3n}{2} \mathbb{E}_{X \sim \xi_h} \big[\mathrm{var}_{Y \sim P_h(X,\cdot)} \left(\boldsymbol{g}_0(Y) - \mathcal{P}_0 \boldsymbol{g}_0(X) \mid X \right) \big] \\ &\preceq \left(\frac{3}{2}\right)^3 n \cdot \mathbb{E}_{X \sim \xi_0} \big[\mathrm{var}_{Y \sim P_0(X,\cdot)} \left(\boldsymbol{g}_0(Y) \mid X \right) \big] \\ &= \frac{27n}{8} \mathrm{trace}\left(\Lambda\right). \end{aligned} \tag{E.6}$$

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 [58, Lemma 10], the density ρ of w has Fisher information

$$I(\rho) = U D^{1/2} I\left(\mu^{\otimes d}\right) D^{1/2} U^\top = n\pi\Lambda. \tag{E.7}$$

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

Putting together the pieces. Combining the bounds (E.4), (E.6), and (E.7) and applying Proposition 5, we obtain the lower bound

$$\inf_{\widehat{\theta}_n} \int_{\mathbb{R}^d} \mathbb{E}_{X_1^n \sim \mathbb{P}_{Qw}} \big[\|\widehat{\theta}_n - \overline{\theta}(P_{Qw})\|_2^2 \big] \rho(dw) \geq \frac{1}{4(5+\pi)n} \mathrm{trace}(\Lambda). \tag{E.8}$$

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 $\mathbb{E}_{X \sim \xi_0, Y \sim P_0(X,\cdot)}[(\mathcal{A}_0 f(Y) - \mathcal{P}_0 \mathcal{A}_0 f(X))^2] = \sum_{k=-\infty}^{\infty} \mathbb{E}[f(s_0) f(s_k)]$, where $(s_k)_{k \in \mathbb{Z}}$ is a stationary Markov chain following P_0.*

See Section E.5 for the proof of this lemma.

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

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

Thus, the right-hand side of equation (E.8) 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 (E.5), 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

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

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

$$\left|e^x - 1\right| \leq e \cdot |x|,$$

valid for $x \in [-1, 1]$.

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

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

For any w on the support of ρ, we thus have

$$P_{Qw} \in \mathfrak{N}_{\mathrm{Prob}}\left(P_0, e\sqrt{\frac{d}{n}}\right),$$

as claimed. The Bayes risk lower bound (E.8) then implies the desired minimax lower bound.

E.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

$$\begin{aligned} d_{\mathrm{TV}}\big(P_0(x,\cdot), P_h(x,\cdot)\big) &\le \sqrt{\frac{1}{2}\chi^2(P_0(x,\cdot) \,||\, P_h(x,\cdot))} \\ &= \sqrt{\frac{1}{2}\sum_{y\in\mathbb{X}} P_0(x,y)\cdot\left(\frac{P_h(x,y)}{P_0(x,y)}-1\right)^2} \\ &\overset{(i)}{\le} \sqrt{\frac{1}{2}\big(e^{\|h_x\|_\infty}-1\big)^2} \overset{(ii)}{\le} e\cdot\max_{x\in\mathbb{X}}\|h_x\|_\infty. \end{aligned} \tag{E.9}$$

In step (i), we use the fact that

$$\frac{P_h(x,y)}{P_0(x,y)} = \frac{e^{h_x(y)}}{\sum_{z\in\mathbb{X}} P_0(x,y)e^{h_x(z)}} \in [e^{-\|h_x\|_\infty}, e^{\|h_x\|_\infty}],$$

and in step (ii), we use the fact that $\|h_x\|_\infty < 1$.

Next, we turn to the proofs of the two claims. We first prove the mixing time bound. Note that the non-expansive condition (2.2) (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 $\{y_k\}$ follow the Markov chain P_0, starting from $x_0 = x$ and $y_0 = y$, respectively. Furthermore, we have the bound $\mathbb{P}(x_\tau \ne y_\tau) \le \frac{1}{4}$.

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

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

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

By the union bound, it follows that

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

In the last step, we have used the total variation distance bound (E.9).

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

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

which shows that the perturbed chain P_h satisfies the condition (2.2) (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 $(x_0, x_2, \ldots, x_n)$, we have the following expression for the Radon–Nikodym derivative:

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

We then have the max-divergence bound

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

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

To obtain the desired claim, we take the initial distribution to be the stationary distribution ξ_h of the chain P_h, and let $n = t_{\mathrm{mix}} \log(\frac{2}{h_{\max} \cdot \min_x \xi_0(x)})$. Note that $\xi_h P_h^n = \xi_h$ in such case. On the other hand, by Lemma 4, the total variation distance can be upper bounded as $d_{\mathrm{TV}}(\xi_h P_0^n, \xi_0) \leq 2^{1 - \frac{n}{t_{\mathrm{mix}}}} \leq h_{\max} \cdot \min_{x\in\mathbb{X}} \xi_0(x)$. Therefore, for any $x \in \mathbb{X}$, we have

$$\left| \frac{\xi_h P_0^n(x)}{\xi_0(x)} - 1 \right| \leq \frac{d_{\mathrm{TV}}\big(\xi_h P_0^n, \xi_0\big)}{\min_{x\in\mathbb{X}} \xi_0(x)} \leq h_{\max} < \frac{1}{2}.$$

Invoking the inequality $|\log z| \leq 2|z - 1|$ for $|z| \leq 1/2$, we can translate the bound into a max-divergence bound

$$D_\infty\big(\xi_h P_0^n \,||\, \xi_0\big) = \max_{x\in\mathbb{X}} \left| \log \frac{\xi_h P_0^n(x)}{\xi_0(x)} \right| \leq 2h_{\max}.$$

Finally, applying the triangle inequality yields

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

which proves the second claim.

E.2. Proof of Lemma 12

We first consider the functional $h \mapsto \bar{\theta}(P_h) := (I - \mathbb{E}_{\xi_h}[\boldsymbol{L}(s)])^{-1}\mathbb{E}_{\xi_h}[\boldsymbol{b}(s)]$. 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

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

In the above expression, the notation $(I - P_h)^{-1}|_{\mathbb{H}_h}$ 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 $\tilde{f} := f - \mathbb{E}_{\xi_h}[f]$, and write

$$\begin{aligned}&\frac{\partial}{\partial h_x(y)}\mathbb{E}_{\xi_h}\big[f(s)\big] \\ &= \frac{\partial}{\partial h_x(y)}\mathbb{E}_{\xi_h}\big[\tilde{f}(s)\big] = \xi_h(x) P_h(x, y) \cdot \big[\mathbf{1}_{z=y} - P_h(x, z)\big]_{z \in \mathbb{X}} \cdot \big(I - P_h\big)^{-1}\big|_{\mathbb{H}_h} \cdot \tilde{f} \\ &= \xi_h(x) P_h(x, y) \cdot \big[\mathbf{1}_{z=y} - P_h(x, z)\big]_{z \in \mathbb{X}} \cdot \big(I - P_h\big)^{-1}\big|_{\mathbb{H}_h} \cdot \big(f - \mathbb{E}_{\xi_h}[f]\big) \\ &= \xi_h(x) P_h(x, y) \cdot \Big(\mathcal{A}_h f(y) - \sum_{z \in \mathbb{X}} P_h(x, z) \mathcal{A}_h f(z)\Big). \qquad \text{(E.10)}\end{aligned}$$

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

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

Following the formula (E.10), we conclude that

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

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

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

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

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

as claimed. ∎

E.3. Proof of Lemma 13

The following technical lemma is used throughout the proof and proved in Section E.6.

Lemma 16. *Given a perturbation vector w satisfying*

$$\|w\|_2 \le \frac{1-\kappa}{2ct_{\mathrm{mix}}\sigma_L\sqrt{d\cdot\|\!|\Lambda|\!\|_{\mathrm{op}}}\log d},$$

for $h = Qw$, the matrix $I - \overline{L}^{(h)}$ is invertible, with

$$\|\!|(I - \overline{L}^{(h)})^{-1}|\!\|_{\mathrm{op}} \le \frac{2}{1-\kappa}.$$

Before proceeding with the proof, we note two direct consequences of Lemma 17 from Section E.7. 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 $j \in [d]$:

$$\mathbb{E}_{X\sim\xi_h}\big[\langle e_j, \mathcal{A}_h\boldsymbol{L}(X)u\rangle^2\big] \le ct_{\mathrm{mix}}^2\sigma_L^2\log^2 d \tag{E.11a}$$

and

$$\mathbb{E}_{X\sim\xi_h}\big[\langle e_j, \mathcal{A}_h\boldsymbol{b}(X)\rangle^2\big] \le ct_{\mathrm{mix}}^2\sigma_b^2\log^2 d. \tag{E.11b}$$

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

$$\mathbb{P}_{X\sim\xi_h}\big[\mathcal{A}_h f_j(X) \ge K\big] \le K^{-2p}\mathbb{E}_{X\sim\xi_h}\big[\mathcal{A}_h f_j(X)^{2p}\big] \le \left(\frac{cp^2t_{\mathrm{mix}}\overline{\sigma}\log d}{K}\right)^{2p}.$$

By taking $K = 2cp^2 t_{\mathrm{mix}}\overline{\sigma}\log d$ and $p = -2\log\min_{x\in\mathbb{X}}\xi_0(x)$, we find that

$$\mathbb{P}_{X\sim\xi_h}\left[\mathcal{A}_h f_j(X) \geq 8ct_{\mathrm{mix}}\overline{\sigma}\log^3\left(\frac{d}{\min_{x\in\mathbb{X}}\xi_0(x)}\right)\right] < \frac{1}{2}\min_{x\in\mathbb{X}}\xi_0(x) \leq \min_{x\in\mathbb{X}}\xi_h(x).$$

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

$$\mathcal{A}_h f_j(x) \leq 8ct_{\mathrm{mix}}\overline{\sigma}\log^3\left(\frac{d}{\min_{x'\in\mathbb{X}}\xi_0(x')}\right) \quad \text{for all } x \in \mathbb{X}.$$

Combining the estimates for all j coordinates yields the bound

$$\begin{aligned}\max_{x\in\mathbb{X}}\|\boldsymbol{g}_h(x)\|_2 &\leq \frac{1}{1-\kappa}\max_{x\in\mathbb{X}}\|\mathcal{A}_h\big[f_j(x)\big]_{j\in[d]}\|_2 \\ &\leq \frac{ct_{\mathrm{mix}}\overline{\sigma}\sqrt{d}}{1-\kappa}\log^3\left(\frac{d}{\min_{x\in\mathbb{X}}\xi_0(x)}\right). \end{aligned} \tag{E.12}$$

Given the two lemmas and facts derived above, we now proceed to the proof of Lemma 13. Taking derivatives on both sides of equation (E.1), we obtain

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

We then have the integral relation

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

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

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

Consequently, for $X \sim \xi_h$ and $Y \sim P_h(X,\cdot)$, we have the error decomposition

$$\begin{aligned}\|\nabla_w\overline{\theta}(P_h)w\|_2 &= \big\|\mathbb{E}\big[\mathrm{cov}\big(\boldsymbol{g}_h(Y), \boldsymbol{g}_0(Y)\mid X\big)\big]w\big\|_2 \\ &\leq \big\|\mathbb{E}\big[\mathrm{cov}\big(\boldsymbol{g}_0(Y)\mid X\big)\big]w\big\|_2 \\ &\quad + \big\|\mathbb{E}\big[\mathrm{cov}\big(\boldsymbol{g}_h(Y) - \boldsymbol{g}_0(Y), \boldsymbol{g}_0(Y)\mid X\big)\big]w\big\|_2.\end{aligned}$$

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

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

For the first term in above decomposition, we have

$$\begin{aligned}
\big\|\mathbb{E}\big[\mathrm{cov}\big(\boldsymbol{g}_0(Y) \mid X\big)\big]w\big\|_2 &\le \big\|\mathbb{E}\big[(\boldsymbol{g}_0(Y) - \mathcal{P}_0\boldsymbol{g}_0(X))(\boldsymbol{g}_0(Y) - \mathcal{P}_0\boldsymbol{g}_0(X))^\top\big]w\big\|_2 \\
&\le \frac{9}{4}\big\|\mathrm{cov}_{X\sim\xi_0, Y\sim P_0(X,\cdot)}\big(\boldsymbol{g}_0(Y) - \mathcal{P}_0\boldsymbol{g}_0(X)\big)\cdot w\big\|_2 \\
&= \frac{9}{4}\big\|\Lambda w\big\|_2 \le \frac{9}{4}\sqrt{\mathrm{trace}(\Lambda)/n},
\end{aligned}$$

where the last inequality is due to the bound (E.2c).

For the second term in the decomposition, for $X \sim \xi_h$ and $Y \sim P_h(X,\cdot)$, we have

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

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

$$\|\nabla_w\overline{\theta}(P_h)w\|_2 \le \frac{9}{4}\sqrt{\frac{\mathrm{trace}(\Lambda)}{n}} + \frac{3}{2}\cdot\sqrt{\frac{d}{n}\cdot\mathbb{E}_{X\sim\xi_h}\|\boldsymbol{g}_h(X) - \boldsymbol{g}_0(X)\|_2^2}. \tag{E.13}$$

Collecting the bounds above and invoking equation (E.11) and Lemma 16, we obtain the following bound on the desired term:

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

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

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

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

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

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

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

where the bound for the factor $\sup_{x\in\mathbb{X}} \|\boldsymbol{g}_h(x)\|_2^2$ follows from equation (E.12). For the latter term in the display above, we note that

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

Putting together the pieces yields the first estimate

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

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 $\tilde{s}_t$ such that $\tilde{s}_t \sim \xi_h$, and we have $\mathbb{P}(\tilde{s}_t \ne s_t) \le 2^{\lfloor \frac{t}{t_{\mathrm{mix}}}\rfloor}$. Define a random variable $\tilde{s}_{t+1}$ by setting $\tilde{s}_{t+1} = s_{t+1}$ whenever $s_t = \tilde{s}_t$, and drawing $\tilde{s}_{t+1} \sim P(\tilde{s}_t, \cdot)$ otherwise. From this construction, we have

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

Applying the Cauchy–Schwarz inequality twice yields

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

corresponding to the second estimate.

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

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

so that

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

Bounds on the term $J_3(sh, z)\cdot w$. By equation (E.10), 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

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

where the final inequality is due to equation (E.11). Combining with Lemma 16, we have the bound

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

Finishing the proof. Collecting the bounds for J_1, J_2, and J_3 and for

$$n \ge \frac{ct_{\mathrm{mix}}^2\sigma_L^2 d^2\log^2 d}{(1-\kappa)^2},$$

we have

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

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

For the second claim, we combine the first claim with equation (E.13) and obtain

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

Taking the integral yields

$$\begin{aligned}\|\bar{\theta}(P_h) - \bar{\theta}(P_0)\|_2 &\le \int_0^1 \|\nabla_w \bar{\theta}(P_{\ell h}) w\|_2 d\ell \\ &\le \frac{3}{2}\sqrt{\frac{\operatorname{trace}(\Lambda)}{n}} + \sqrt{\frac{c(1+\sigma_L^2)\bar{\sigma}^2 t_{\mathrm{mix}}^4 d^3}{(1-\kappa)^4 n^2}\log^6\left(\frac{d}{\min_x \xi_0(x)}\right)},\end{aligned}$$

which proves the second claim.

E.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

$$I^{(n)}(h) := \mathbb{E}_h[\nabla_h \log \mathbb{P}_h(s_0^n) \nabla_h \log \mathbb{P}_h(s_0^n)^\top].$$

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

$$\mathbb{P}_h(s_0, s_1, \ldots, s_n) = \xi_0(s_0) \cdot \prod_{t=1}^n P_h(s_{t-1}, s_t).$$

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

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

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

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

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

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

We find that

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

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

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

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

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

Turning to the Fisher information, we compute

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

E.5. Proof of Lemma 15

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

$$f(s_k) = \mathcal{A}_0 f(s_k) - \mathbb{E}\big[\mathcal{A}_0 f(s_{k+1}) \mid s_k\big] = \mathcal{A}_0 f(s_k) - \mathcal{P}_0 \mathcal{A}_0 f(s_k). \tag{E.14}$$

By stationarity, we have

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

where step (i) makes use of the dominated convergence theorem, in particular by noting that $\big|\mathbb{E}[f(s_k) \mid s_0]\big| \le \|f\|_\infty \cdot 2^{1-k/t_{\text{mix}}}$ 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 (E.14).

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. ∎

E.6. Proof of Lemma 16

By following the derivation of equation (E.10), we find that

$$\frac{\partial}{\partial h_x(y)} \overline{L}^{(h)} = \xi_h(x) P_h(x, y) \Big\{ \mathcal{A}_h \boldsymbol{L}(y) - \sum_{z \in \mathbb{X}} P_h(x, z) \mathcal{A}\boldsymbol{L}(z) \Big\}.$$

Consequently, for any $u \in \mathbb{S}^{d-1}$, we have the bound

$$\begin{aligned}
&\left|\!\left|\!\left| \nabla_w \big(\bar{L}^{(h)} u\big) \right|\!\right|\!\right|_{\mathrm{op}} \\
&\quad \le \sup_{z,v \in \mathbb{S}^{d-1}} \sqrt{\mathbb{E}_{Y\sim\xi_h}\big[\big(z^\top \mathcal{A}_h \boldsymbol{L}(Y) u\big)^2\big]} \\
&\qquad \cdot \sqrt{\mathbb{E}_{X\sim\xi_h, Y\sim P_h(X,\cdot)}\big[\big((\boldsymbol{g}_0(Y) - \mathcal{P}_0 \boldsymbol{g}_0(X))^\top v\big)^2\big]} \\
&\quad \le \sup_{v \in \mathbb{S}^{d-1}} \sqrt{\mathbb{E}_{Y\sim\xi_h}\big[\|\mathcal{A}_h \boldsymbol{L}(Y) u\|_2^2\big]} \\
&\qquad \cdot \frac{3}{2} \sqrt{\mathbb{E}_{X\sim\xi_0, Y\sim P_0(X,\cdot)}\big[\big((\boldsymbol{g}_0(Y) - \mathcal{P}_0 \boldsymbol{g}_0(X))^\top v\big)^2\big]} \\
&\quad \le c t_{\mathrm{mix}} \sigma_L \sqrt{d \cdot |\!|\!| \Lambda |\!|\!|_{\mathrm{op}}} \log d.
\end{aligned}$$

We thus obtain

$$\begin{aligned}
|\!|\!| \bar{L}^{(h)} - \bar{L}^{(0)} |\!|\!|_{\mathrm{op}} &\le \sup_{u \in \mathbb{S}^{d-1}} \big\| \big(\bar{L}^{(h)} - \bar{L}^{(0)}\big) u \big\|_2 \\
&\le \int_0^1 \sup_{u \in \mathbb{S}^{d-1}} \big\| \nabla_w \big(\bar{L}^{(sQw)} u\big) \cdot w \big\|_2 ds \\
&\le c t_{\mathrm{mix}} \sigma_L \sqrt{d \cdot \mathrm{trace}(\Lambda)} \log d \cdot \|w\|_2.
\end{aligned}$$

Now, given a perturbation vector satisfying the bound $\|w\|_2 \le \frac{1-\kappa}{2 c t_{\mathrm{mix}} \sigma_L \sqrt{d \cdot |\!|\!| \Lambda |\!|\!|_{\mathrm{op}}} \log d}$, we have the following bound for any $u \in \mathbb{S}^{d-1}$:

$$\begin{aligned}
\|(I - \bar{L}^{(h)}) u\|_2 &\ge \|(I - \bar{L}^{(0)}) u\|_2 - \|(\bar{L}^{(h)} - \bar{L}^{(0)}) u\|_2 \\
&\ge (1 - \kappa) - |\!|\!| \bar{L}^{(h)} - \bar{L}^{(0)} |\!|\!|_{\mathrm{op}} \ge \frac{1-\kappa}{2},
\end{aligned}$$

which implies that $|\!|\!| (I - \bar{L}^{(h)})^{-1} |\!|\!|_{\mathrm{op}} \le \frac{2}{1-\kappa}$, as claimed.

E.7. 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 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 exists a universal constant $c > 0$ such that for any integer $p \ge 1$*

$$\big(\mathbb{E}_{X\sim\xi_h}\big[\big(\mathcal{A}_h f(X)\big)^{2p}\big]\big)^{\frac{1}{2p}} \le c\, p\, t_{\mathrm{mix}} \big[\mathbb{E}_{X\sim\xi_h}\big[f(X)^{2p}\big]\big]^{\frac{1}{2p}} \log\left\{\frac{\|f\|_\infty^{2p}}{\mathbb{E}_{X\sim\xi_h}\big[f(X)^{2p}\big]}\right\}.$$

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 $t \geq 0$, there exists a random variable $\tilde{s}_t$ such that $\tilde{s}_t \mid s_0 \sim \xi_h$, and $\mathbb{P}(\tilde{s}_t \neq s_t \mid s) \leq 2^{-\lfloor \frac{t}{4t_{\mathrm{mix}}} \rfloor}$.

By construction, the state $\tilde{s}_t$ is independent of s. Consequently, we have the equivalence $\mathcal{A}_h f(s) = \sum_{k=0}^{\infty} \mathbb{E}[f(s_k) - f(\tilde{s}_k) \mid s]$, and for any $\alpha > 0$,

$$\begin{aligned}
\mathbb{E}_{s\sim\xi_h}\big[\big(\mathcal{A}_h f(s)\big)^{2p}\big] &\leq \left(\sum_{k=0}^{\infty} e^{2p\alpha t}\mathbb{E}\big(\mathbb{E}\big[f(s_k) - f(\tilde{s}_k) \mid s\big]\big)^{2p}\right) \\
&\quad \cdot \left(\sum_{k=0}^{\infty} e^{-\frac{2p}{2p-1}\alpha k}\right)^{2p-1} \\
&\leq \alpha^{1-2p} \sum_{k=0}^{\infty} e^{2p\alpha k}\mathbb{E}\big[\big|f(s_k) - f(\tilde{s}_k)\big|^{2p}\big].
\end{aligned}$$

We bound the moment of $f(s_k) - f(\tilde{s}_k)$ for different values of k in two ways. On the one hand, Young's inequality directly leads to the following naive bound:

$$\mathbb{E}\big[\big|f(s_k) - f(\tilde{s}_k)\big|^{2p}\big] \leq 2^{2p-1}\big(\mathbb{E}\big[f(s_k)^{2p}\big] + \mathbb{E}\big[f(\tilde{s}_k)^{2p}\big]\big) = 2^{2p}\mathbb{E}_{s\sim\xi_h}\big[f(s)^{2p}\big].$$

On the other hand, for any bounded function f, we have

$$\mathbb{E}\big[\big|f(s_k) - f(\tilde{s}_k)\big|^{2p}\big] \leq \|f\|_\infty^{2p} \cdot \mathbb{P}\big(s_k \neq \tilde{s}_k\big) \leq \|f\|_\infty^{2p} \cdot 2^{1-\frac{k}{4t_{\mathrm{mix}}}}.$$

Combining the two estimates yields the bound

$$\begin{aligned}
&\mathbb{E}\big[(\mathcal{A}_h f(X))^{2p}\big] \\
&\quad \leq \alpha^{1-2p}\left\{2^{2p} \cdot e^{2p\alpha\tau}\tau\mathbb{E}_{s\sim\xi_h}\big[f(s)^{2p}\big] + \|f\|_\infty^{2p} \sum_{k=\tau+1}^{\infty} e^{2p\alpha k} \cdot 2^{1-\frac{k}{4t_{\mathrm{mix}}}}\right\},
\end{aligned}$$

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.

F. Proofs for the examples

We collect the proofs of the consequences to specific examples in this section.

F.1. Proofs for TD(0)

We stated three corollaries applicable to this method, and in this section, we prove each of them in turn.

F.1.1. Proof of Corollary 2. 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 (4.1) can be equivalently written as

$$\theta_{t+1} = (1-\eta\beta)\theta_t + \eta\beta \boldsymbol{L}_{t+1}(\omega_t)\theta_t - \eta\beta \boldsymbol{b}_{t+1}(\omega_t),$$

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 $\kappa(A) := \frac{1}{2}\lambda_{\max}(A + A^\top)$. We verify the eigenvalue condition (2.1) by noting that

$$\begin{aligned}\frac{1}{2}\lambda_{\max}\big(\bar{L} + \bar{L}^\top\big) &= 1 - \frac{1}{\beta}\kappa\big(\gamma\mathbb{E}_{s\sim\xi,s^+\sim P(s,\cdot)}\big[\phi(s)\phi(s^+)^\top\big] - \mathbb{E}_\xi\big[\phi(s)\phi(s)^\top\big]\big)\\ &= 1 - \frac{1}{\beta}\lambda_{\max}\left(B^{1/2}\left(I_d - \frac{M+M^\top}{2}\right)B^{1/2}\right)\\ &= 1 - \frac{\mu}{\beta}(1-\kappa) < 1,\end{aligned}$$

and

$$\|\bar{L}\|_{\mathrm{op}} \le 1 + \frac{1}{\beta}\big(\big\|\mathbb{E}_{s\sim\xi,s^+\sim P(s,\cdot)}\big[\phi(s)\phi(s^+)^\top\big]\big\|_{\mathrm{op}} + \big\|\mathbb{E}_\xi\big[\phi(s)\phi(s)^\top\big]\big\|_{\mathrm{op}}\big) \le 3.$$

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 (4.2a) imply that

$$\|\boldsymbol{L}_{t+1}(s_t)\|_{\mathrm{op}} \le 1 + \frac{1}{\beta}\big(\|\phi(s_t)\phi(s_{t+1})\|_{\mathrm{op}} + \|\phi(s_t)\phi(s_t)^\top\|_{\mathrm{op}}\big) \le (1+\varsigma^2)d$$

and

$$\|\boldsymbol{b}_{t+1}(s_t)\|_2 \le \frac{1}{\beta}|R_t(s_t)| \cdot \|\phi(s_t)\|_2 \le \varsigma^2\sqrt{d/\beta}.$$

Turning to the moment assumption, given any vector $u \in \mathbb{S}^{d-1}$ and coordinate vector e_j, we have the bounds

$$\begin{aligned}\mathbb{E}_{s\sim\xi, s^+\sim P(s,\cdot)}\big[\big(e_j^\top \phi(s)\phi(s^+)^\top u\big)^2\big] &\le \sqrt{\mathbb{E}_{s\sim\xi}\big[\big(e_j^\top \phi(s)\big)^4\big]}\cdot\sqrt{\mathbb{E}_{s\sim\xi}\big[\big(u^\top \phi(s)\big)^4\big]}\\ &\le \beta^2\varsigma^4,\\ \mathbb{E}_{s\sim\xi}\big[\big(e_j^\top \phi(s)\phi(s)^\top u\big)^2\big] &\le \sqrt{\mathbb{E}_{s\sim\xi}\big[\big(e_j^\top \phi(s)\big)^4\big]}\cdot\sqrt{\mathbb{E}_{s\sim\xi}\big[\big(u^\top \phi(s)\big)^4\big]}\\ &\le \beta^2\varsigma^4,\\ \mathbb{E}_{s\sim\xi}\big[\big(e_j^\top R_t(s)\phi(s)\big)^2\big] &\le \varsigma^2\mathbb{E}_{s\sim\xi}\big[\big(e_j^\top \phi(s)\big)^2\big] \le \beta\varsigma^4.\end{aligned}$$

Finally, the quantity $\overline{\sigma}$ from equation (4.3) is bounded as

$$\begin{aligned}&\max_{j\in[d]}\mathbb{E}\big[\langle e_j,\, (\boldsymbol{L}_{t+1}(\omega_t)-\overline{L})\overline{\theta}+(\boldsymbol{b}_{t+1}(\omega_t)-b)\rangle^2\big]\\ &\quad\le \max_{j\in[d]}\sqrt{\mathbb{E}\big[\langle e_j,\,\phi(s_t)\rangle^4\big]}\cdot\sqrt{\mathbb{E}\big[\big(\phi(s_t)^\top\overline{\theta}-\gamma\phi(s_{t+1})^\top\overline{\theta}-R_t(s_t)\big)^4\big]}\le\overline{\sigma}^2.\end{aligned}$$

Invoking equation (6.6) with the test matrix $Q := B$ and substituting with the representation $V(s) = \langle\theta,\,\phi(s)\rangle$ yield the claim.

F.1.2. Proof of Corollary 3. We prove this corollary by verifying the assumptions used in our main theorem. Assumption 2 directly follows from (4.6c) 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

$$\boldsymbol{L}(s,s^+) = I_d - \beta^{-1}(\phi(s)\phi(s)^\top - \gamma\phi(s)\phi(s^+)^\top)$$

satisfies Assumption 4, we first note that

$$\begin{aligned}&\|\boldsymbol{L}(s_1,s_1^+)-\boldsymbol{L}(s_2,s_2^+)\|_{\mathrm{op}}\\ &\le\frac{1}{\beta}\|\phi(s_1)\phi(s_1)^\top-\phi(s_2)\phi(s_2)^\top\|_{\mathrm{op}}+\frac{\gamma}{\beta}\|\phi(s_1)\phi(s_1^+)^\top-\phi(s_2)\phi(s_2^+)^\top\|_{\mathrm{op}}.\end{aligned}$$

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 (4.6b) 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 $\boldsymbol{L} : \mathbb{X} \to \mathbb{R}^{d \times d}$ is $3\varsigma^2 d$-Lipschitz with respect to the metric

$$\rho((s_1, s_1^+), (s_2, s_2^+)) = \|s_1 - s_2\|_2 + \|s_1^+ - s_2^+\|_2.$$

Similarly, for the vector observation $\boldsymbol{b}_t(s) = R_t(s)\phi(s)$, we note that, for any $s_1, s_2 \in \mathbb{X}$,

$$\begin{aligned}\|\boldsymbol{b}_t(s_1) - \boldsymbol{b}_t(s_2)\|_2 &\leq \big|R_t(s_1) - R_t(s_2)\big| \cdot \|\phi(s_1)\|_2 + \big|R_t(s_2)\big| \cdot \|\phi(s_1) - \phi(s_2)\|_2 \\ &\leq 2\varsigma\sqrt{d/\beta}\|\phi(s_1) - \phi(s_2)\|_2,\end{aligned}$$

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 2.

F.1.3. Proof of Corollary 4. In order to verify that Assumption 4 holds with respect to the discrete metric, note that, for any $d_n \geq 1$, we have

$$\|\boldsymbol{b}_t(s)\|_2 \leq \frac{\varsigma}{\beta}\sqrt{\sum_{j=1}^{d_n}\phi_j^2(s)} \leq \frac{\varsigma^2}{\beta}\sqrt{d_n}$$

and

$$\|\boldsymbol{L}(s_1, s_2)\|_{\mathrm{op}} \leq 1 + \frac{1}{\beta}\sum_{j=1}^{d_n}\phi_j^2(s_1) + \frac{1}{\beta}\sqrt{\sum_{j=1}^{d_n}\phi_j^2(s_1)} \cdot \sqrt{\sum_{j=1}^{d_n}\phi_j^2(s_2)} \leq \frac{1+\varsigma^2}{\beta}d_n.$$

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

$$\begin{aligned}&\mathbb{E}\big[\langle e_j, \boldsymbol{L}(s, s^+)u\rangle^2\big] \\ &\quad\leq 3 + \frac{3}{\beta^2}\mathbb{E}\big[\big(\langle e_j, \phi(s)\rangle\langle\phi(s^+), u\rangle\big)^2\big] + \frac{3}{\beta^2}\mathbb{E}\big[\big(\langle e_j, \phi(s)\rangle\langle\phi(s), u\rangle\big)^2\big] \\ &\quad\leq 3 + \frac{6}{\beta^2}\|\phi_j\|_\infty^2 \cdot \mathbb{E}\big[\langle\phi(s), u\rangle^2\big] \leq 3 + \frac{6}{\beta}\varsigma^2.\end{aligned}$$

For each $t = 1, 2, \ldots$, we also have $\mathbb{E}[\langle e_j, \boldsymbol{b}_{t+1}(s_t)\rangle^2] \leq \frac{1}{\beta^2}\|R_t\|_\infty^2 \cdot \mathbb{E}_{s\sim\xi}[\phi_j(s)^2] \leq \frac{\varsigma^2}{\beta}$, which is an order-one quantity. Following the same steps as in the proof as Corollary 2 then yields the claim.

F.2. Proofs for TD(λ)

We first prove Proposition 2—the mixing time result—and then use it to establish Corollary 5.

F.2.1. 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

$$\mathbb{P}(s_{(k+1)t_{\mathrm{mix}}+1} \neq s'_{(k+1)t_{\mathrm{mix}}+1} \mid \{s_t, s_t'\}_{t=1}^{kt_{\mathrm{mix}}+1}) \leq \frac{1}{2}.$$

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 (4.10b) 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

$$\mathbb{P}(s_{3t_{\mathrm{mix}}+1} \neq s'_{3t_{\mathrm{mix}}+1}) \leq \frac{1}{8}.$$

Conditioning on the event $\mathcal{E} := \{s_{3t_{\mathrm{mix}}+1} = s'_{3t_{\mathrm{mix}}+1}\}$, for any $t \geq 3t_{\mathrm{mix}} + 1$, we have

$$\|h_{t+1} - h'_{t+1}\|_2 = \gamma\lambda \|h_t - h_t'\|_2 = \cdots = (\gamma\lambda)^{t-3t_{\mathrm{mix}}-1} \|h_{3t_{\mathrm{mix}}+1} - h'_{3t_{\mathrm{mix}}+1}\|_2. \quad \text{(F.1)}$$

We split the remainder of the proof into two cases.

Case I: $s_1 \neq s_1'$. The coupling bound implies that $\mathbb{P}(\mathcal{E}) \geq \frac{7}{8}$. On the event $\mathcal{E}$, 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

$$\begin{aligned}
&\mathbb{E}\big[\rho\big((s_\tau, s_{\tau+1}, h_\tau), (s_\tau', s_{\tau+1}', h_\tau)\big)\big] \\
&\quad = \frac{1}{4}\big(\mathbb{P}\big(s_\tau \neq s_\tau'\big) + \mathbb{P}\big(s_{\tau+1} \neq s_{\tau+1}'\big) + \mathbb{E}\big[\|h_\tau - h_\tau'\|_2\big]\big) \\
&\quad \leq \frac{3}{4}\mathbb{P}(\mathcal{E}^c) + \frac{1}{4}\mathbb{E}\big[\|h_\tau - h_\tau'\|_2 \mid \mathcal{E}\big] \\
&\quad \leq \frac{1}{8} = \frac{1}{2} \cdot \frac{1}{4} \mathbf{1}_{s_1 \neq s_1'} \leq \frac{1}{2}\rho\big((s_0, s_1, h_0), (s_0', s_1', h_0)\big),
\end{aligned}$$

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 (F.1) then yields

$$\begin{aligned}
&\mathbb{E}\big[\rho\big((s_\tau, s_{\tau+1}, h_\tau), (s_\tau', s_{\tau+1}', h_\tau)\big)\big] \\
&\quad = \frac{1}{4}\mathbb{E}\big[\|h_\tau - h_\tau'\|_2\big] \leq \frac{1}{8}\|h_0 - h_0'\|_2 \leq \frac{1}{2}\rho(\omega_0, \omega_0'),
\end{aligned}$$

which establishes contraction in this case. Combining the two cases proves the proposition.

F.2.2. Proof of Corollary 5. We note that the SA procedure (4.10a) 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 $\boldsymbol{L}_{t+1}(\omega_t) = (I_d - \frac{1}{\beta} g_t \phi(s_t)^\top + \gamma \frac{1}{\beta} g_t \phi(s_{t+1})^\top)$ and $\boldsymbol{b}_{t+1}(\omega_t) = \frac{1}{\beta} R_t(s_t) g_t$.

Recalling that

$$M_\lambda = (1-\lambda)\gamma \sum_{t=0}^{\infty} \lambda^t \gamma^{t+1} B^{-1/2} \mathbb{E}[\phi(s_0)\phi(s_{t+1})^\top] B^{-1/2},$$

we first study the eigenvalues of the symmetrized version of M_λ and relate these back to those of $\bar{L} = \mathbb{E}_{\tilde{\xi}}[\boldsymbol{L}_{t+1}(\omega_t)]$. Note that by the Cauchy–Schwarz inequality, for any vector $u \in \mathbb{S}^{d-1}$, we have

$$\begin{aligned} &u^\top B^{-1/2} \mathbb{E}\big[\phi(s_0)\phi(s_t)^\top\big] B^{-1/2} u \\ &\quad \le \sqrt{\mathbb{E}\big[\big(u^\top B^{-1/2}\phi(s_0)\big)^2\big] \cdot \mathbb{E}\big[\big(u^\top B^{-1/2}\phi(s_t)\big)^2\big]} = 1. \end{aligned}$$

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 2, we can deduce that

$$\frac{1}{2}\lambda_{\max}\big(\bar{L} + \bar{L}^\top\big) = \frac{1}{\beta}\lambda_{\max}\left(B^{1/2}\left(\frac{M_\lambda + M_\lambda^\top}{2}\right)B^{1/2}\right) \ge \frac{(1-\lambda)\gamma}{1-\lambda\gamma}.$$

Next, we verify Assumption 2 on the noise moments. By the update rule (4.10b), 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

$$\mathbb{E}\big[\langle g_t, u\rangle^4\big] \le \left(\sum_{k=0}^{\infty} (\gamma\lambda)^k\right)^3 \cdot \sum_{k=0}^{\infty} (\gamma\lambda)^k \mathbb{E}\big[\langle u, \phi(s_{t-k})\rangle^4\big] \le \beta^2 \left(\frac{\varsigma}{1-\gamma\lambda}\right)^4.$$

In other words, for all standard basis vectors e_j, we have

$$\begin{aligned} \mathbb{E}\big[\langle e_j, \boldsymbol{L}_{t+1}(\omega_t)u\rangle^2\big] &\le 1 + \frac{2}{\beta^2}\sqrt{\mathbb{E}\big[\langle e_j, \phi(s_t)\rangle^4\big]} \cdot \sqrt{\mathbb{E}\big[\langle g_t, u\rangle^4\big]} \\ &\le 1 + 2\frac{\varsigma^4}{(1-\gamma\lambda)^2}, \\ \mathbb{E}\big[\langle e_j, \boldsymbol{b}_{t+1}(\omega_t)u\rangle^2\big] &\le \frac{\varsigma^2}{\beta^2}\mathbb{E}\big[\langle g_t, e_j\rangle^2\big] \le \frac{\varsigma^4}{\beta(1-\gamma\lambda)^2}. \end{aligned}$$

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

$$\begin{aligned} T &\le \frac{\varsigma\sqrt{d/\beta}}{1-\lambda\gamma}\cdot\left(\|h^\top\phi(s)-(h')^\top\phi(s')\|_{\mathrm{op}}+\|h^\top\phi(s_+)-(h')^\top\phi(s'_+)\|_{\mathrm{op}}\right)\\ &\le \frac{\varsigma\sqrt{d/\beta}}{1-\lambda\gamma}\cdot\left(\|(h-h')^\top\phi(s')\|_{\mathrm{op}}+\|h^\top(\phi(s')-\phi(s))\|_{\mathrm{op}}\right)\\ &\quad+\frac{\varsigma\sqrt{d/\beta}}{1-\lambda\gamma}\cdot\left(\|(h-h')^\top\phi(s'_+)\|_{\mathrm{op}}+\|h^\top(\phi(s'_+)-\phi(s_+))\|_{\mathrm{op}}\right)\\ &\le \frac{2\varsigma^2 d}{1-\lambda\gamma}\left(\mathbf{1}_{s\neq s'}+\mathbf{1}_{s_+\neq s'_+}+\|h-h'\|_2\right)=\frac{8\varsigma^2 d}{1-\lambda\gamma}\rho(\omega,\omega'). \end{aligned}$$

Finally, we note that the quantity $\overline{\sigma}$ defined in equation (4.3) satisfies the bound

$$\begin{aligned} &\sup_{j\in[d]}\mathbb{E}\left[\langle e_j,\,(\boldsymbol{L}_{t+1}(\omega_t)-\overline{L})\overline{\theta}+(\boldsymbol{b}_{t+1}(\omega_t)-b)\rangle^2\right]\\ &\le \sup_{j\in[d]}\sqrt{\mathbb{E}\left[\langle e_j,\,g_t\rangle^4\right]}\cdot\sqrt{\mathbb{E}\left[\left(\phi(s_t)^\top\overline{\theta}-\gamma\phi(s_{t+1})^\top\overline{\theta}-R_t(s_t)\right)^4\right]}\le\frac{\overline{\sigma}^2}{(1-\gamma\lambda)^2}. \end{aligned}$$

Invoking equation (6.6), with the test matrix $Q := B$, and substituting the expression

$$V(s)=\langle\theta,\,\phi(s)\rangle$$

yield the claim.

F.3. Proofs for vector autoregressive estimation

In this section, we present proofs of results on vector autoregressive models, as introduced in Example 3.

F.3.1. Proof of Proposition 3. We prove the claim by a direct construction of the coupling. Given two initial points

$$\omega_0=\left[X_1^\top,X_0^\top,\ldots,X_{-k+1}^\top\right]^\top\quad\text{and}\quad\omega'_0=\left[X_1'^\top,X_0'^\top,\ldots,X_{-k+1}'^\top\right]^\top,$$

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

$$Y_{t+1}=\begin{bmatrix}X_{t+1} & \cdots & X_{t-k+2}\end{bmatrix}^\top.$$

(Note that Y_{t+1} is a sliding window with length one unit shorter than ω_t.) We have

$$\begin{aligned} \|Y_{t+1}-Y'_{t+1}\|_{P_*}^2&=\|R_*(Y_t-Y'_t)\|_{P_*}^2=\|Y_t-Y'_t\|_{P_*}^2-\|Y_t-Y'_t\|_{Q_*}^2\\ &\le\left(1-\frac{\mu}{\beta}\right)\|Y_t-Y'_t\|_{P_*}^2. \end{aligned}$$

Consequently, the augmented processes $\omega_t = (X_{t+1}, X_t, \ldots, X_{t-k+1})$ and $\omega'_t = (X'_{t+1}, X'_t, \ldots, X'_{t-k+1})$ satisfy the bound

$$\begin{aligned}\|\omega_t - \omega'_t\|_2 &\le \|Y_{t+1} - Y'_{t+1}\|_2 + \|Y_t - Y'_t\|_2 \\ &\le \frac{1}{\sqrt{\lambda_{\min}(P_*)}}\big(\|Y_{t+1} - Y'_{t+1}\|_{P_*} + \|Y_t - Y'_t\|_{P_*}\big) \\ &\le 2\sqrt{\frac{\lambda_{\max}(P_*)}{\lambda_{\min}(P_*)}}\Big(1 - \frac{\mu}{2\beta}\Big)^t \|\omega_0 - \omega'_0\|_2.\end{aligned}$$

Note that since $P_* \succeq Q_*$, we have $\lambda_{\min}(P_*) \ge \lambda_{\min}(Q_*) = \mu$. Taking

$$t_{\mathrm{mix}} = c\frac{\beta}{\mu}\left(1 + \log\frac{\beta}{\mu}\right)$$

yields the contraction bound $\|\omega_{t_{\mathrm{mix}}} - \omega'_{t_{\mathrm{mix}}}\|_2 \le \frac{1}{2}\|\omega_0 - \omega'_0\|_2$. Taking expectations on both sides completes the proof.

F.3.2. Proof of Corollary 6. We begin by showing norm bounds and moment bounds on the process $(X_t)_{t\ge 0}$. By definition (2.12) of the process and stability, the block vector $Y_t := [\,X_t \;\; X_{t-1} \;\cdots\; X_{t-k+1}\,]^\top$ 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

$$\|X_t\|_2 \le \frac{1}{\mu}\|Y_t\|_{P_*} \le \sum_{i=0}^{\infty}\|R_*^i \varepsilon_{t-i} e_1\|_{P_*} \le \frac{1}{\mu}\sum_{i=0}^{\infty}\Big(1 - \frac{\mu}{\beta}\Big)^i \|\varepsilon_{t-i} e_1\|_{P_*} \le \frac{\beta^2}{\mu^2}\varsigma\sqrt{m}.$$

Moreover, for each $u \in \mathbb{S}^{m-1}$, we have

$$\begin{aligned}\mathbb{E}\big[\langle X_t, u\rangle^4\big] &\le \left(\sum_{i=0}^{\infty} e^{-\frac{i\mu}{6\beta}}\right)^3 \cdot \sum_{i=0}^{\infty} e^{\frac{i\mu}{2\beta}}\mathbb{E}\big[\langle R_*^i \varepsilon_{t-i} e_1, u e_1\rangle^4\big] \\ &\le c(\beta/\mu)^3 \cdot \sum_{i=0}^{\infty} e^{\frac{i\mu}{2\beta}} \cdot \frac{\beta^4}{\mu^4} \cdot e^{-\frac{i\mu}{\beta}}\varsigma^4 \le c'\left(\frac{\beta^2\varsigma}{\mu^2}\right)^4.\end{aligned}$$

Next, we proceed with verifying the assumptions used in Theorem 1. Letting $\nu := 1/\|H^*\|_{\mathrm{op}}$, the stochastic approximation procedure can be rewritten as

$$\begin{aligned}\theta_{t+1} = \Big(1 - \frac{\eta}{\nu}\Big)\theta_t + \frac{\eta}{\nu}\Big(\theta_t - \nu\big(\big[X_{t-j}X_{t+1-i}^\top\big]_{i,j\in[m]} \otimes I_m\big)\theta_t \\ + \nu \cdot \mathrm{vec}\left(\begin{bmatrix} X_{t+1}X_t^\top & \cdots & X_{t+1}X_{t-k+1}^\top \end{bmatrix}\right)\Big).\end{aligned}$$

Observe that the matrix $\bar{L} := I_{km^2} - \nu H^* \otimes I_m$ satisfies the eigenvalue bound

$$\frac{1}{2}\lambda_{\max}(\bar{L} + \bar{L}^\top) \le 1 - \frac{\nu}{2}\lambda_{\min}(H^* + (H^*)^\top) \le 1 - \nu h^*.$$

On the other hand, the empirical observations satisfy the almost sure bounds

$$\|\boldsymbol{L}_{t+1}(\omega_t) - \bar{L}\|_{\mathrm{op}} \le \nu \cdot \left\|\left[X_{t-j} X_{t+1-i}^\top\right]_{i,j\in[m]}\right\|_{\mathrm{op}} \le \nu \cdot \frac{\beta^4}{\mu^4}\varsigma^2 mk$$

and

$$\|\boldsymbol{b}_{t+1}(\omega_t) - \bar{b}\|_{\mathrm{op}} \le \nu \cdot \left\| \begin{bmatrix} X_{t+1}X_t^\top & \cdots & X_{t+1}X_{t-k+1}^\top \end{bmatrix} \right\|_F \le \nu \cdot \frac{\beta^4}{\mu^4}\varsigma^2 m\sqrt{k}.$$

For two collections of matrices $\mathcal{U} = (U^{(j)})_{j=1}^k$ and $\mathcal{V} = (V^{(j)})_{j=1}^k \subseteq \mathbb{R}^{m\times m}$ 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

$$\begin{aligned}
&\mathbb{E}\left[\left\langle \mathrm{vec}(\mathcal{U}),\, \left(\boldsymbol{L}_{t+1}(\omega_t) - \bar{L}\right)\mathrm{vec}(\mathcal{V})\right\rangle^2\right] \\
&\quad\le \nu^2 \mathbb{E}\left[\left(\sum_{\ell=0}^{k-1}\left\langle U^{(\ell)}, \sum_{j=0}^{k-1} V^{(j)} X_{t-j} X_{t-\ell}^\top\right\rangle_F\right)^2\right],
\end{aligned}$$

which is in turn at most

$$\begin{aligned}
&\nu^2 k^2 \sum_{\ell=0}^{k-1}\sum_{j=0}^{k-1} \sqrt{\mathbb{E}\left[X_{t-\ell}^{\otimes 4}\right]\left[(U^{(\ell)})^\top U^{(\ell)}, (U^{(\ell)})^\top U^{(\ell)}\right]} \\
&\qquad\times \sqrt{\mathbb{E}\left[X_{t-j}^{\otimes 4}\right]\left[(V^{(j)})^\top V^{(j)}, (V^{(j)})^\top V^{(j)}\right]}.
\end{aligned}$$

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

$$\begin{aligned}
\mathbb{E}\left[X_{t-\ell}^{\otimes 4}\right]\left[(U^{(\ell)})^\top U^{(\ell)}, (U^{(\ell)})^\top U^{(\ell)}\right] &= \mathbb{E}\left[X_{t-\ell}^{\otimes 4}\right]\left[\sum_{i=1}^m \lambda_i^2 u_i u_i^\top, \sum_{i=1}^m \lambda_i^2 u_i u_i^\top\right] \\
&= \sum_{i,i'} \mathbb{E}\left[X_{t-\ell}^{\otimes 4}\right][u_i, u_i, u_{i'}, u_{i'}] \cdot \lambda_i^2 \lambda_{i'}^2 \\
&\le c'\left(\frac{\beta^2\varsigma}{\mu^2}\right)^4 \left(\sum_i \lambda_i^2\right)^2 \\
&= c'\left(\frac{\beta^2\varsigma}{\mu^2}\right)^4 \|U^{(\ell)}\|_F^2.
\end{aligned}$$

Putting together the pieces, we have

$$\begin{aligned}
&\mathbb{E}\left[\left\langle \mathrm{vec}(\mathcal{U}),\, \left(\boldsymbol{L}_{t+1}(\omega_t) - \bar{L}\right)\mathrm{vec}(\mathcal{V})\right\rangle^2\right] \\
&\quad\le \nu^2 k^2 c'\left(\frac{\beta^2\varsigma}{\mu^2}\right)^4 \cdot \sum_{\ell=0}^{k-1}\sum_{j=0}^{k-1} \|U^{(\ell)}\|_F^2 \|V^{(j)}\|_F^2 \le c\left(\nu \cdot \frac{\beta^4 k\varsigma^2}{\mu^4}\right)^2.
\end{aligned}$$

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

$$\begin{aligned}\mathbb{E}\big[\langle e_{\ell,i,j},\, (\boldsymbol{b}_{t+1}(\omega_t) - \bar{b})\rangle^2\big] &\le \nu^2 \mathbb{E}\big[\langle e_i e_j^\top,\, X_{t+1} X_{t-\ell+1}\rangle^2\big] \\ &\le \nu^2 \sqrt{\mathbb{E}\big[\langle e_j^\top,\, X_{t+1}\rangle^4\big]} \cdot \sqrt{\mathbb{E}\big[\langle e_i^\top,\, X_{t-\ell+1}\rangle^4\big]} \\ &\le c'\big(\nu \cdot \frac{\beta^2 \varsigma}{\mu^2}\big)^4.\end{aligned}$$

Applying Theorem 1 completes the proof of this corollary.

Acknowledgments. The authors thank Yaqi Duan for helpful discussions.

Funding. AP was partially supported by the National Science Foundation through grants nos. 2107455 and 2210734 and by awards/gifts from Adobe, Amazon, and Mathworks. MJW and PLB were partially supported by the NSF through grants nos. IIS-1909365 and DMS-2023505. PLB was partially supported by the ONR through MURI award N000142112431. This work was partially supported by NSF grant CCF-1955450, ONR grant N00014-21-1-2842, and NSF grant DMS-2311072 to MJW.

References

[1] A. Ben-Tal and A. Nemirovski, *Lectures on modern convex optimization*. MPS/SIAM Series on Optimization, Society for Industrial and Applied Mathematics (SIAM), Philadelphia, PA; Mathematical Programming Society (MPS), Philadelphia, PA, 2001 Zbl 0986.90032 MR 1857264

[2] A. Benveniste, M. Métivier, and P. Priouret, *Adaptive algorithms and stochastic approximations*. Appl. Math. (New York) 22, Springer, Berlin, 1990 Zbl 0752.93073 MR 1082341

[3] D. P. Bertsekas, Temporal difference methods for general projected equations. *IEEE Trans. Automat. Control* **56** (2011), no. 9, 2128–2139 Zbl 1368.90155 MR 2865769

[4] D. P. Bertsekas, *Reinforcement learning and optimal control*. Athena Sci. Optim. Comput. Ser., Athena Scientific, Belmont, MA, 2019 MR 4315932

[5] J. Bhandari, D. Russo, and R. Singal, A finite time analysis of temporal difference learning with linear function approximation. *Oper. Res.* **69** (2021), no. 3, 950–973 Zbl 1472.90150 MR 4280425

[6] P. Billingsley, Statistical methods in Markov chains. *Ann. Math. Statist.* **32** (1961), 12–40 Zbl 0104.12802 MR 0123420

[7] V. S. Borkar, *Stochastic approximation: A dynamical systems viewpoint*. Cambridge University Press, Cambridge; Hindustan Book Agency, New Delhi, 2008 Zbl 1181.62119 MR 2442439

[8] V. S. Borkar, A concentration bound for contractive stochastic approximation. *Systems Control Lett.* **153** (2021), article no. 104947 Zbl 1475.93106 MR 4253964

[9] L. Bottou, F. E. Curtis, and J. Nocedal, Optimization methods for large-scale machine learning. *SIAM Rev.* **60** (2018), no. 2, 223–311 Zbl 1397.65085 MR 3797719

[10] N. Bou-Rabee, A. Eberle, and R. Zimmer, Coupling and convergence for Hamiltonian Monte Carlo. *Ann. Appl. Probab.* **30** (2020), no. 3, 1209–1250 MR 4133372

[11] J. A. Boyan, Technical update: Least-squares temporal difference learning. *Mach. Learn.* **49** (2002), no. 2-3, 233–246

[12] S. J. Bradtke and A. G. Barto, Linear least-squares algorithms for temporal difference learning. *Mach. Learn.* **22** (1996), no. 1-3, 33–57 Zbl 0845.68091

[13] G. Bresler, P. Jain, D. Nagaraj, P. Netrapalli, and X. Wu, Least squares regression with Markovian data: Fundamental limits and algorithms. 2020, arXiv:2006.08916

[14] P. J. Brockwell and R. A. Davis, *Time series: Theory and methods*. Springer Ser. Statist., Springer, New York, 2006 Zbl 1169.62074 MR 2839251

[15] S. Chen, A. Devraj, A. Busic, and S. Meyn, Explicit mean-square error bounds for Monte–Carlo and linear stochastic approximation. 2020, arXiv:2002.02584

[16] X. Chen, J. D. Lee, X. T. Tong, and Y. Zhang, Statistical inference for model parameters in stochastic gradient descent. *Ann. Statist.* **48** (2020), no. 1, 251–273 Zbl 1440.62287 MR 4065161

[17] Z. Chen, S. T. Maguluri, S. Shakkottai, and K. Shanmugam, A lyapunov theory for finite-sample guarantees of asynchronous Q-learning and TD-learning variants. [v1] 2021, [v4] 2023, arXiv:2102.01567v4

[18] C. Daskalakis, N. Dikkala, and N. Gravin, Testing symmetric Markov chains from a single trajectory. [v1] 2017, [v2] 2017, arXiv:1704.06850v2

[19] P. Dayan and T. J. Sejnowski, TD(λ) converges with probability 1. *Mach. Learn.* **14** (1994), no. 3, 295–301

[20] V. Debavelaere, S. Durrleman, and S. Allassonnière, On the convergence of stochastic approximations under a subgeometric ergodic Markov dynamic. *Electron. J. Stat.* **15** (2021), no. 1, 1583–1609 Zbl 1466.60136 MR 4255313

[21] A. Dieuleveut, A. Durmus, and F. Bach, Bridging the gap between constant step size stochastic gradient descent and Markov chains. *Ann. Statist.* **48** (2020), no. 3, 1348–1382 Zbl 1454.62242 MR 4124326

[22] T. T. Doan, L. M. Nguyen, N. H. Pham, and J. Romberg, Finite-Time analysis of stochastic gradient descent under Markov randomness. [v1] 2020, [v2] 2020, arXiv:2003.10973v2

[23] Y. Duan, M. Wang, and M. J. Wainwright, Optimal policy evaluation using kernel-based temporal difference methods. 2021, arXiv:2109.12002

[24] A. Durmus, E. Moulines, A. Naumov, S. Samsonov, and H.-T. Wai, On the stability of random matrix product with Markovian noise: Application to linear stochastic approximation and TD learning. 2021, arXiv:2102.00185

[25] E. Even-Dar and Y. Mansour, Learning rates for Q-learning. *J. Mach. Learn. Res.* **5** (2003/04), 1–25 Zbl 1222.68196 MR 2247972

[26] G. Fort, Central limit theorems for stochastic approximation with controlled Markov chain dynamics. *ESAIM Probab. Stat.* **19** (2015), 60–80 Zbl 1333.60029 MR 3374869

[27] S. Gadat and F. Panloup, Optimal non-asymptotic bound of the Ruppert–Polyak averaging without strong convexity. 2017, arXiv:1709.03342

[28] S. Ghadimi and G. Lan, Optimal stochastic approximation algorithms for strongly convex stochastic composite optimization I: A generic algorithmic framework. *SIAM J. Optim.* **22** (2012), no. 4, 1469–1492 Zbl 1301.62077 MR 3023780

[29] J. K. Ghosh, B. K. Sinha, and H. S. Wieand, Second order efficiency of the MLE with respect to any bounded bowl-shaped loss function. *Ann. Statist.* **8** (1980), no. 3, 506–521 Zbl 0436.62031 MR 0568717

[30] R. D. Gill and B. Y. Levit, Applications of the Van Trees inequality: a Bayesian Cramér–Rao bound. *Bernoulli* **1** (1995), no. 1-2, 59–79 Zbl 0830.62035 MR 1354456

[31] P. E. Greenwood and W. Wefelmeyer, Efficiency of empirical estimators for Markov chains. *Ann. Statist.* **23** (1995), no. 1, 132–143 Zbl 0822.62067 MR 1331660

[32] J. D. Hamilton, *Time series analysis*. Princeton University Press, Princeton, NJ, 1994 Zbl 0831.62061 MR 1278033

[33] D. Hsu, A. Kontorovich, D. A. Levin, Y. Peres, C. Szepesvári, and G. Wolfer, Mixing time estimation in reversible Markov chains from a single sample path. *Ann. Appl. Probab.* **29** (2019), no. 4, 2439–2480 Zbl 1466.60143 MR 3983341

[34] P. Jain, S. S. Kowshik, D. Nagaraj, and P. Netrapalli, Streaming linear system identification with reverse experience replay. [v1] 2021, [v3] 2021, arXiv:2103.05896v3

[35] R. Johnson and T. Zhang, Accelerating stochastic gradient descent using predictive variance reduction. In *Advances in neural information processing systems*, pp. 315–323, 26, 2013

[36] M. Kaledin, E. Moulines, A. Naumov, V. Tadic, and H.-T. Wai, Finite time analysis of linear two-timescale stochastic approximation with Markovian noise. 2020, arXiv:2002.01268

[37] B. Karimi, B. Miasojedow, E. Moulines, and H.-T. Wai, Non-asymptotic analysis of biased stochastic approximation scheme. [v1] 2019, [v4] 2019, arXiv:1902.00629v4

[38] P. Karmakar and S. Bhatnagar, Two time-scale stochastic approximation with controlled Markov noise and off-policy temporal-difference learning. *Math. Oper. Res.* **43** (2018), no. 1, 130–151 Zbl 1434.62174 MR 3774637

[39] K. Khamaru, A. Pananjady, F. Ruan, M. J. Wainwright, and M. I. Jordan, Is temporal difference learning optimal? An instance-dependent analysis. *SIAM J. Math. Data Sci.* **3** (2021), no. 4, 1013–1040 Zbl 07419556 MR 4320891

[40] K. Khamaru, E. Xia, M. J. Wainwright, and M. I. Jordan, Instance-optimality in optimal value estimation: Adaptivity via variance-reduced Q-learning. 2021, arXiv:2106.14352

[41] V. R. Konda and J. N. Tsitsiklis, Actor-critic algorithms. In *Advances in neural information processing systems*, pp. 1008–1014, 2000

[42] G. Kotsalis, G. Lan, and T. Li, Simple and optimal methods for stochastic variational inequalities, II: Markovian noise and policy evaluation in reinforcement learning. *SIAM J. Optim.* **32** (2022), no. 2, 1120–1155 Zbl 1493.90205 MR 4429422

[43] H. J. Kushner and D. S. Clark, *Stochastic approximation methods for constrained and unconstrained systems*. Appl. Math. Sci. 26, Springer, New York, 1978 Zbl 0381.60004 MR 0499560

[44] H. J. Kushner and G. G. Yin, *Stochastic approximation and recursive algorithms and applications*. 2nd edn., Appl. Math. (New York) 35, Springer, New York, 2003 Zbl 1026.62084 MR 1993642

[45] Y. A. Kutoyants, Efficiency of the empirical distribution for ergodic diffusion. *Bernoulli* **3** (1997), no. 4, 445–456 Zbl 0910.62079 MR 1483698

[46] C. Lakshminarayanan and C. Szepesvári, Linear stochastic approximation: How far does constant step-size and iterate averaging go? In *International conference on artificial intelligence and statistics*, pp. 1347–1355, 2018

[47] O. V. Lepskiĭ, A problem of adaptive estimation in Gaussian white noise. *Theory Probab. Appl.* **35** (1990), no. 3, 459–470 Zbl 0745.62083 MR 1091202

[48] D. A. Levin and Y. Peres, *Markov chains and mixing times*. American Mathematical Society, Providence, RI, 2017 Zbl 1390.60001 MR 3726904

[49] C. J. Li, W. Mou, M. J. Wainwright, and M. I. Jordan, ROOT-SGD: Sharp nonasymptotics and asymptotic efficiency in a single algorithm. [v1] 2020, [v2] 2023, arXiv:2008.12690v2

[50] G. Li, Y. Wei, Y. Chi, Y. Gu, and Y. Chen, Breaking the sample size barrier in model-based reinforcement learning with a generative model. [v1] 2020, [v8] 2023, arXiv:2005.12900v8

[51] T. Li, G. Lan, and A. Pananjady, Accelerated and instance-optimal policy evaluation with linear function approximation. *SIAM J. Math. Data Sci.* **5** (2023), no. 1, 174–200 Zbl 07669892 MR 4562584

[52] X. Li, M. Wang, and A. Zhang, Estimation of Markov chain via rank-constrained likelihood. [v1] 2018, [v2] 2018, arXiv:1804.00795v2

[53] L. Ljung, Analysis of recursive stochastic algorithms. *IEEE Trans. Automatic Control* **22** (1977), no. 4, 551–575 Zbl 0362.93031 MR 0465458

[54] L. Ljung, On positive real transfer functions and the convergence of some recursive schemes. *IEEE Trans. Automatic Control* **22** (1977), no. 4, 539–551 Zbl 0361.93063 MR 0456829

[55] H. Lütkepohl, *New introduction to multiple time series analysis*. Springer, Berlin, 2005 Zbl 1072.62075 MR 2172368

[56] M. Métivier and P. Priouret, Applications of a Kushner and Clark lemma to general classes of stochastic algorithms. *IEEE Trans. Inform. Theory* **30** (1984), no. 2, part 1, 140–151 Zbl 0546.62056 MR 0807052

[57] S. Meyn and R. L. Tweedie, *Markov chains and stochastic stability*. Springer Science & Business Media, 2009

[58] W. Mou, A. Pananjady, and M. J. Wainwright, Optimal oracle inequalities for solving projected fixed-point equations. 2020, arXiv:2012.05299

[59] E. Moulines and F. R. Bach, Non-asymptotic analysis of stochastic approximation algorithms for machine learning. In *Advances in neural information processing systems*, pp. 451–459, 2011

[60] R. Munos and C. Szepesvári, Finite-time bounds for fitted value iteration. *J. Mach. Learn. Res.* **9** (2008), 815–857 Zbl 1225.68203 MR 2417255

[61] A. Nemirovski, A. Juditsky, G. Lan, and A. Shapiro, Robust stochastic approximation approach to stochastic programming. *SIAM J. Optim.* **19** (2008), no. 4, 1574–1609 Zbl 1189.90109 MR 2486041

[62] A. Pananjady and M. J. Wainwright, Instance-dependent ℓ_∞-bounds for policy evaluation in tabular reinforcement learning. *IEEE Trans. Inform. Theory* **67** (2021), no. 1, 566–585 Zbl 1473.62082 MR 4231973

[63] S. Penev, Efficient estimation of the stationary distribution for exponentially ergodic Markov chains. *J. Statist. Plann. Inference* **27** (1991), no. 1, 105–123 Zbl 0727.62079 MR 1089356

[64] B. T. Polyak and A. B. Juditsky, Acceleration of stochastic approximation by averaging. *SIAM J. Control Optim.* **30** (1992), no. 4, 838–855 Zbl 0762.62022 MR 1167814

[65] C. R. Rao, First and second order asymptotic efficiencies of estimators. (With discussion). *Ann. Fac. Sci. Univ. Clermont-Ferrand* **8** (1962), 33–40 MR 0293765

[66] H. Robbins and S. Monro, A stochastic approximation method. *Ann. Math. Statistics* **22** (1951), 400–407 Zbl 0054.05901 MR 0042668

[67] G. A. Rummery and M. Niranjan, On-line Q-learning using connectionist systems. Tech. rep., Cambridge University Engineering Department, 1994

[68] D. Ruppert, Efficient estimations from a slowly convergent Robbins–Monro process. Tech. rep., Cornell University Operations Research and Industrial Engineering, 1988

[69] A. Sidford, M. Wang, X. Wu, L. F. Yang, and Y. Ye, Near-optimal time and sample complexities for solving Markov decision processes with a generative model. In *Proceedings of the 32nd international conference on neural information processing systems*, pp. 5192–5202, 2018

[70] R. Srikant and L. Ying, Finite-time error bounds for linear stochastic approximation and TD learning. In *Conference on learning theory, PMLR*, pp. 2803–2830, 2019

[71] R. S. Sutton, Learning to predict by the methods of temporal differences. *Mach. Learn.* **3** (1988), no. 1, 9–44

[72] C. Szepesvári, The asymptotic convergence-rate of Q-learning. In *Advances in neural information processing systems*, pp. 1064–1070, 1998

[73] C. Szepesvári, *Algorithms for reinforcement learning*. Synthesis Lectures on Artificial Intelligence and Machine Learning 9, Morgan & Claypool, San Rafael, CA, 2010 Zbl 1205.68320

[74] J. N. Tsitsiklis, Asynchronous stochastic approximation and Q-learning. *Mach. Learn.* **16** (1994), 185–202 Zbl 0820.68105

[75] J. N. Tsitsiklis and B. Van Roy, An analysis of temporal-difference learning with function approximation. *IEEE Trans. Automat. Control* **42** (1997), no. 5, 674–690 Zbl 0914.93075 MR 1454208

[76] J. N. Tsitsiklis and B. Van Roy, Optimal stopping of Markov processes: Hilbert space theory, approximation algorithms, and an application to pricing high-dimensional financial derivatives. *IEEE Trans. Automat. Control* **44** (1999), no. 10, 1840–1851 Zbl 0958.60042 MR 1716061

[77] A. B. Tsybakov, *Introduction to nonparametric estimation*. Springer Science & Business Media, 2008

[78] A. W. van der Vaart, *Asymptotic statistics*. Camb. Ser. Stat. Probab. Math. 3, Cambridge University Press, Cambridge, 1998 Zbl 0910.62001 MR 1652247

[79] C. J. L. W. Mou, M. J. Wainwright, P. L. Bartlett, and M. I. Jordan, On linear stochastic approximation: Fine-grained Polyak–Ruppert and non-asymptotic concentration. In *Proceedings of thirty third conference on learning theory*, pp. 2947–2997, 125, 2020

[80] M. J. Wainwright, *High-dimensional statistics: A non-asymptotic viewpoint*. Camb. Ser. Stat. Probab. Math. 48, Cambridge University Press, Cambridge, 2019 Zbl 1457.62011 MR 3967104

[81] M. J. Wainwright, Stochastic approximation with cone-contractive operators: Sharp ℓ_∞-bounds for Q-learning. [v1] 2019, [v2] 2019, arXiv:1905.06265v2

[82] M. J. Wainwright, Variance-reduced Q-learning is minimax optimal. [v1] 2019, [v2] 2019, arXiv:1906.04697v2

[83] C. J. Watkins and P. Dayan, Q-learning. *Mach. Learn.* **8** (1992), no. 3-4, 279–292 Zbl 0773.68062

[84] G. Wolfer and A. Kontorovich, Statistical estimation of ergodic Markov chain kernel over discrete state space. *Bernoulli* **27** (2021), no. 1, 532–553 Zbl 1472.62133 MR 4177379

[85] F. Yang, S. Balakrishnan, and M. J. Wainwright, Statistical and computational guarantees for the Baum–Welch algorithm. *J. Mach. Learn. Res.* **18** (2017), article no. 125 Zbl 1442.62192 MR 3763759

[86] H. Yu and D. P. Bertsekas, Error bounds for approximations from projected linear equations. *Math. Oper. Res.* **35** (2010), no. 2, 306–329 Zbl 1218.90211 MR 2674722

[87] L. Yu, K. Balasubramanian, S. Volgushev, and M. A. Erdogdu, An analysis of constant step size SGD in the non-convex regime: Asymptotic normality and bias. [v1] 2020, [v2] 2020, arXiv:2006.07904v2

Received 10 August 2022; revised 31 August 2023.

Wenlong Mou
Department of Statistical Sciences and Vector Institute for Artificial Intelligence, University of Toronto, 700 University Avenue, Toronto, M5G 1X6, Canada; wmou.work@gmail.com

Ashwin Pananjady
School of ISyE and School of ECE, Georgia Institute of Technology, 765 Ferst Dr. NW, Atlanta, GA 30332, USA; ashwinpm@gatech.edu

Martin J. Wainwright
Department of EECS, Department of Mathematics, Laboratory for Information and Decision Systems and Statistics and Data Science Center, Massachusetts Institute of Technology, 32 Vassar Street, Cambridge, MA 02139, USA; wainwrigwork@gmail.com

Peter L. Bartlett
Department of EECS and Department of Statistics, University of California, 367 Evans Hall, Berkeley, CA 94720; Google DeepMind, 1600 Amphitheatre Parkway, Mountain View, CA 94043, USA; peter@berkeley.edu