Almost Sure Convergence Rates of Stochastic Approximation and Reinforcement Learning via a Poisson-Moreau Drift

Xinyu Liu
University of Virginia
xinyuliu@virginia.edu

Zixuan Xie
University of Virginia
xie.zixuan@email.virginia.edu

Shangtong Zhang
University of Virginia
shangtong@virginia.edu

Abstract

Establishing almost sure convergence rates for stochastic approximation and reinforcement learning under Markovian noise is a fundamental theoretical challenge. We make progress towards this challenge for a class of stochastic approximation algorithms whose expected updates are contractive, a setting that arises in many reinforcement learning algorithms such as Q-learning and linear temporal difference learning. Specifically, for a power-law learning rate $\mathcal{O}(n^{-\eta})$ with $\eta \in (1/2, 1)$, we obtain an almost sure convergence rate arbitrarily close to $o(n^{1-2\eta})$. For the harmonic learning rate $\alpha_n = \alpha/(n+1)$, the almost sure squared-error rate is $o(n^{-\zeta})$ for every $\zeta < \min\{1, 2(1-\kappa)\alpha\}$, where κ is the contraction factor. Thus exponents arbitrarily close to one are available when $2(1-\kappa)\alpha \geq 1$. This approaches the n^{-1} polynomial exponent of classical law-of-the-iterated-logarithm benchmarks. Key to our analysis is a novel Lyapunov drift construction that applies a Poisson-equation based correction for Markovian noise to the well-established Moreau-envelope smoothing for the contractive mapping.

1 Introduction

Introduced by Robbins and Monro [1951], stochastic approximation (SA) remains a foundational framework for learning from continuous streams of noisy observations. The canonical recursion,

$$\theta_{n+1} = \theta_n + \alpha_n F(\theta_n, Y_n),$$

where $\theta_n \in \mathbb{R}^d$ is the active estimate and α_n is a decaying learning rate, is remarkably versatile. It natively captures stochastic gradient descent (SGD) algorithms [Robbins and Monro, 1951] as well as core reinforcement learning (RL, Sutton and Barto [2018]) algorithms like Q-learning [Watkins and Dayan, 1992] and linear temporal difference (TD) learning [Sutton, 1988]. Over the decades, a rich classical theory has been developed to answer the basic consistency question: does the sequence $\{\theta_n\}$ converge? While convergence can be analyzed in probability or in L^p, our focus is strictly on almost sure convergence. This mode is particularly vital because it guarantees the limiting behavior of almost every individual sample path, offering a stronger, trajectory-wise assurance than average-case or tail-probability metrics.

Historically, almost sure convergence is established via the ODE method, see, e.g., Benveniste et al. [1990], Kushner and Yin [2003], Borkar [2009], Borkar et al. [2025], Liu et al. [2025a]. While powerful, these techniques typically fall short of quantifying exactly how fast a realized trajectory

40th Conference on Neural Information Processing Systems (NeurIPS 2026).

approaches its limit, i.e., they establish only asymptotic almost sure convergence without a rate. An alternative path leverages the almost-supermartingale framework pioneered by Robbins and Siegmund [1971]. While their original work did not immediately extract convergence rates, a recent wave of literature has successfully adapted the almost-supermartingale approach to derive explicit almost sure rates under various conditions [Liu and Yuan, 2024, Karandikar and Vidyasagar, 2024, Qian et al., 2024, Liu et al., 2025b]. Beyond supermartingale arguments, researchers have derived almost sure rates by imposing stricter structural assumptions. In the realm of linear SA, for example, Chong et al. [1999] and Tadic [2004] handle Markovian sequences by exploiting the product-of-matrices dynamics unique to affine updates, whereas Kouritzin and Sadeghi [2015] achieve similar linear results assuming independent noise. When moving to general nonlinear operators, however, the literature generally bottlenecks at the noise assumptions. Koval and Schwabe [2003] and Vidyasagar [2023] establish fundamental convergence rate guarantees, yet their analyses heavily rely on the noise $\{Y_n\}$ being fully independent. Karandikar and Vidyasagar [2024] attempt to relax this to allow non-i.i.d. sequences; however, their framework mandates that the conditional bias decays fast enough to remain summable, an assumption that fundamentally breaks down under standard Markovian noise where the mixing rate is constant. Within the specific domain of RL, notable early milestones include Szepesvári [1997] for Q-learning and Tadić [2002] for linear TD. Yet, Szepesvári [1997] relies heavily on count-based learning rates, making it an essentially i.i.d. sampling protocol.

To summarize, although there are many results on almost sure convergence rates for SA and RL, if we consider two desiderata that are characteristic of modern RL algorithms, namely

(i) the noise $\{Y_n\}$ forms a Markov chain, and

(ii) the expectation of F is not necessarily linear and is contractive only in some possibly non-Euclidean norm such as ℓ_∞,

the only existing result that handles both, to the best of our knowledge, is the recent work Qian et al. [2024], which establish almost sure convergence rates for arbitrary-norm contractive stochastic approximations with Markovian noise via a skeleton-iterates technique. However, Qian et al. [2024] cover only a limited range of learning rates and their convergence rates are slower than ours. Tables 1 & 2 contextualize the strength of our results with the existing literature about almost sure convergence rates. In short, our key contribution is **sharper almost sure convergence rates for general stochastic approximations under both (i) and (ii).** Our results apply to general contractive Markovian SA under uniform Poisson regularity.

	Algorithm type (i.e., F)	Markovian $\{Y_n\}$
Pelletier [1998]	SGD	
Godichon-Baggioni [2016]	SGD	
Godichon-Baggioni [2019]	Averaged SGD	
Sebbouh et al. [2021]	SGD / Stochastic Heavy Ball	
Liu and Yuan [2022]	SGD	
Liang et al. [2023]	Mirror descent	
Liu and Yuan [2024]	SGD	
Weissmann et al. [2025]	SGD	
Szepesvári [1997]	Q-learning	
Tadić [2002]	Linear TD	✓
Chong et al. [1999]	Linear	✓
Tadic [2004]	Linear	✓
Kouritzin and Sadeghi [2015]	Linear	
Koval and Schwabe [2003]	General	
Vidyasagar [2023]	General	
Karandikar and Vidyasagar [2024]	General	
Qian et al. [2024]	General	✓
This work	General	✓

Table 1: Comparison of works that establish almost sure convergence rates for stochastic approximation algorithms. By "General", we mean a setting that allows for nonlinear F and arbitrary-norm contractivity.

learning rate regime	Qian et al. [2024]	**This work**
$1/2 < \eta \leq 2/3$	N/A	a.c. to $o(n^{-(2\eta-1)})$
$2/3 < \eta < 1$	a.c. to $o(n^{-(\frac{3}{2}\eta-1)})$	a.c. to $o(n^{-(2\eta-1)})$
$\eta = 1$	$o\left(\frac{1}{\exp\left(\zeta_1 \log^{1/(1+\zeta_2)} n\right)}\right), \zeta_1 \in (0,1), 0 < \zeta_2$	a.c. to $o(n^{-1})$ †

Table 2: Comparison with Qian et al. [2024] for learning rates of $\alpha_n = \mathcal{O}(n^{-\eta})$. The entries describe squared-error rates. By "a.c.", we mean "arbitrarily close". For example, "a.c. to $o(n^{-\zeta})$" means that the rate holds for $o(n^{-\zeta+\epsilon})$ for arbitrary $\epsilon > 0$. †For the harmonic learning rate $\alpha_n = \alpha/(n+1)$, the displayed "a.c. to $o(n^{-1})$" rate requires $2(1-\kappa)\alpha \geq 1$; in general, Theorem 1 permits every $\zeta < \min\{1, 2(1-\kappa)\alpha\}$. To understand the rate of Qian et al. [2024] for $\eta = 1$, consider the extreme unattainable case $\zeta_1 = 1$ and $\zeta_2 = 0$, then the rate degrades to $o(n^{-1})$. For any $\zeta_1 < 1$ and $\zeta_2 > 0$, the rate is subpolynomial, i.e., it decays slower than $n^{-\zeta}$ for any $\zeta > 0$ because $\lim_{n\to\infty} n^{\zeta}/\exp\left(\zeta_1 \log^{1/(1+\zeta_2)} n\right) = \infty$ for any $\zeta > 0$.

From a technical side, we use Moreau-envelope smoothing to address (ii), which is a well established technique from Chen et al. [2020, 2024]. The key technical novelty is that we apply a Poisson-equation based correction to the Moreau envelop based Lyapunov function. This correction allows us to handle the Markovian noise in (i) by obtaining almost-supermartingales and applying the recent advances in almost-supermartingale convergence rates from Liu and Yuan [2024], generating sharper almost sure convergence rates than Qian et al. [2024]. Our construction of the Poisson-Moreau drift is, to the best of our knowledge, novel.

2 Main Results

We consider the stochastic approximation recursion initialized from a deterministic $\theta_0 \in \mathbb{R}^d$

$$\theta_{n+1} = \theta_n + \alpha_n F(\theta_n, Y_n). \tag{SA}$$

Let $(\mathsf{Y}, \mathcal{Y})$ be a measurable state space and let P be a Markov kernel on Y. The process $\{Y_n\}_{n\geq 0}$ is a Markov chain whose transitions are governed by P. Let $\{\mathcal{F}_n\}_{n\geq 0}$ be the filtration defined as the σ-algebra generated by the history of the Markov chain, i.e., $\mathcal{F}_n = \sigma(Y_0, Y_1, \ldots, Y_n)$. By construction, θ_n is $\mathcal{F}_{n-1}$-measurable and Y_n is $\mathcal{F}_n$-measurable. The process satisfies

$$\Pr(Y_{n+1} \in A | \mathcal{F}_n) = P(Y_n, A) \quad \text{a.s. for every } A \in \mathcal{Y}.$$

Throughout the paper, $\mathbb{E}_n[\cdot] \doteq \mathbb{E}[\cdot|\mathcal{F}_n]$.

Assumption 2.1. *The Markov kernel P admits a unique invariant probability measure π.*

Under Assumption 2.1, define the expected update $f(\theta) \doteq \int_{\mathsf{Y}} F(\theta, y)\pi(dy)$. We consider the setting where the expected update is the residual of some contractive mapping, i.e., we assume $f(\theta) = T(\theta) - \theta$ for some $T : \mathbb{R}^d \to \mathbb{R}^d$ that is contractive under a norm $\|\cdot\|$. Or formally,

Assumption 2.2. *Let $T(\theta) \doteq \theta + f(\theta)$. Then there exists $\kappa \in (0,1)$ and a norm $\|\cdot\|$ such that*

$$\|T(\theta) - T(\theta')\| \leq \kappa \|\theta - \theta'\|, \quad \forall \theta, \theta' \in \mathbb{R}^d. \tag{1}$$

This assumption allows the contraction norm $\|\cdot\|$ to be non-Euclidean, for example ℓ_∞. We use θ_* to denote the unique fixed point of T, i.e., $f(\theta_*) = 0$. Such contraction driven stochastic approximations arise in many RL algorithms, such as Q-learning and TD. Notably, it also arises in linear stochastic approximation with Hurwitz matrix, i.e., when $f(\theta) = A\theta + b$ with A being a Hurwitz matrix. See Lemma 6.1 of Chen et al. [2025] for more details on how to verify the contraction condition for such linear stochastic approximation. This inclusion of linear stochastic approximation with Hurwitz matrix is important because it covers many RL algorithms such as linear TD and gradient TD [Sutton et al., 2008, 2009]. We further assume that F is Lipschitz continuous.

Assumption 2.3. *There exists some constant L_F such that, for $\forall \theta, \theta' \in \mathbb{R}^d$ and $\forall y \in \mathsf{Y}$,*

$$\|F(\theta, y) - F(\theta', y)\| \leq L_F \|\theta - \theta'\|, \quad \|F(\theta_*, y)\| \leq L_F.$$

We then make a mild assumption on the Markovian noise, assuming that it admits a Poisson equation representation.

Assumption 2.4. *There is a measurable function $H : \mathbb{R}^d \times \mathsf{Y} \to \mathbb{R}^d$ such that for $\forall \theta \in \mathbb{R}^d$ and $\forall y \in \mathsf{Y}$,*

$$H_\theta(y) - (PH_\theta)(y) = F(\theta, y) - f(\theta). \tag{2}$$

Here $H_\theta(y)$ is a shorthand for $H(\theta, y)$ and

$$(PH_\theta)(y) \doteq \int_{\mathsf{Y}} H_\theta(y')P(y, dy'), \quad y \in \mathsf{Y}.$$

We further assume there exists some constant L_H such that, for $\forall \theta, \theta' \in \mathbb{R}^d$ and $\forall y \in \mathsf{Y}$,

$$\|H_\theta(y) - H_{\theta'}(y)\| \le L_H\|\theta - \theta'\|, \quad \|H_{\theta_*}(y)\| \le L_H.$$

Remark 1. *Assumption 2.4 is standard in the stochastic approximation literature (see, e.g., Benveniste et al. [1990], Kushner and Yin [2003], Borkar et al. [2025]). The algorithm does not compute H_θ; verifying existence and the stated bounds suffices. This assumption is relatively mild. In particular, for any finite state space $\mathcal{Y}$ where the Markov chain $\{Y_n\}$ is irreducible and aperiodic, the assumption is automatically satisfied for an arbitrary function F. Under such finite-state conditions and Assumption 2.3 , the solution H_θ can be explicitly constructed via the fundamental matrix (see, e.g., Section 8.2.3 of Puterman [2014]) and satisfies the required boundedness at θ_* and Lipschitz continuity in θ . For generic state spaces, various structural conditions suffice to establish this property.*

For example, uniform geometric ergodicity, $\|P^t g\|_\infty \le C_{\text{mix}}\mu^t\|g\|_\infty$ for bounded centered g, gives

$$H_\theta(y) = \textstyle\sum_{t=0}^{\infty} P^t\big(F(\theta, \cdot) - f(\theta)\big)(y), \qquad L_H = \mathcal{O}\Big(\frac{C_{\text{mix}}L_F}{1-\mu}\Big),$$

where $\mu \in (0, 1)$ and $\|g\|_\infty = \sup_y \|g(y)\|$. For a fixed contraction norm and factor, the mixing rate μ does not affect the exponent of n in Theorem 1.

For instance, the assumption holds if $\{Y_n\}$ is a positive Harris chain[1] that satisfies the Lyapunov drift condition (V4)[2] with a constant function $V(y) \equiv 1$[3]. Under these conditions, Theorem 2.3 of Glynn and Meyn [1996] guarantees the assumption through the boundedness of the corresponding fundamental kernel. On unbounded state spaces, weighted Poisson bounds must imply our uniform bounds. Elaborating on the complete measure-theoretic prerequisites for these bounds would deviate from the primary focus of this paper. Consequently, we directly state Assumption 2.4 to maintain focus on our main convergence results.

We now state our choice of learning rates and the main convergence result.

Assumption 2.5. $\alpha_n = \frac{\alpha}{(n+1)^\eta}$, *where $\alpha > 0$ and $\frac{1}{2} < \eta \le 1$.*

Theorem 1. *Let Assumptions 2.1 - 2.5 hold. The iterates $\{\theta_n\}$ generated by (SA) satisfy*

$$\lim_{n\to\infty}(n+1)^\zeta\|\theta_n - \theta_*\|^2 = 0 \quad a.s.$$

Here, if $\frac{1}{2} < \eta < 1$, the above holds for any $\zeta \in (0, 2\eta - 1)$; if $\eta = 1$, the above holds for any $\zeta \in (0, \min\{1, 2(1-\kappa)\alpha\})$.

The proof is given in Section B.4. Here, the restriction $\eta > 1/2$ ensures $\sum_n \alpha_n^2 < \infty$, as required to control the noise residuals in our almost-supermartingale proof.

3 Poisson-Moreau Drift and Proof of Theorem 1

In this section, we illustrate the key technical novelty of this paper. Write $e_n \doteq \theta_n - \theta_*$. A common approach to analyze the convergence of e_n is to construct a drift function $V : \mathbb{R}^d \to \mathbb{R}_+$ and establish the recursion between $V(e_{n+1})$ and $V(e_n)$. If the norm in Assumption 2.2 is Euclidean, then the natural choice is $V(x) = \frac{1}{2}\|x\|^2$, see, e.g., Bhandari et al. [2018], Srikant and Ying [2019]. To

[1] The formal definitions of positive and Harris chains are provided on pages 204 and 235 of Meyn and Tweedie [2012].

[2] See page 386 of Meyn and Tweedie [2012] for the precise definition of the (V4) condition.

[3] We note that finite irreducible and aperiodic chains trivially satisfy this uniform drift condition.

handle a possibly non-Euclidean norm, Moreau-envelope smoothing is a standard technique. To our knowledge, the first to use Moreau envelopes in the context of stochastic approximation is Chen et al. [2020]. And it has since been widely used in the literature on arbitrary norm contractive stochastic approximation, see, e.g., Chen et al. [2021], Zhang et al. [2022], Chen et al. [2024], Qian et al. [2024], Liu et al. [2025c], Chen et al. [2025]. Below we provide a brief background on this technique. We do note that these are not our contributions and we provide them here only for completeness. That being said, we will shortly articulate why existing developments on Moreau envelopes are not sufficient for our purposes and how we need to further develop the technique by adding a Poisson-equation based correction to the Moreau envelope.

3.1 Background on Moreau Envelopes

Denote the Euclidean norm as $\|\cdot\|_2$. Thanks to norm equivalence in finite dimensions, we can choose constants $\ell, u > 0$ such that

$$\ell\|x\|_2 \le \|x\| \le u\|x\|_2, \quad x \in \mathbb{R}^d.$$

For $\xi > 0$, define the Moreau envelope of $x \mapsto \frac{1}{2}\|x\|^2$ by

$$M_\xi(x) \doteq \inf_{u\in\mathbb{R}^d} \left\{ \tfrac{1}{2}\|u\|^2 + \tfrac{1}{2\xi}\|x-u\|_2^2 \right\}.$$

The following are standard facts that we will use, with $\langle\cdot,\cdot\rangle$ denoting the Euclidean inner product.

Lemma 1 (Proposition A.1 and Section A.2 of Chen et al. [2021])**.** *For every $\xi > 0$, the function M_ξ is continuously differentiable and $1/\xi$-smooth with respect to $\|\cdot\|_2$:*

$$\|\nabla M_\xi(x) - \nabla M_\xi(y)\|_2 \le \tfrac{1}{\xi}\|x-y\|_2, \quad x, y \in \mathbb{R}^d. \tag{3}$$

Moreover, there exists a norm $\|\cdot\|_{m,\xi}$ such that $M_\xi(x) = \frac{1}{2}\|x\|_{m,\xi}^2$. With $\ell_\xi \doteq \sqrt{1+\xi\ell^2}, u_\xi \doteq \sqrt{1+\xi u^2}$, we have

$$\ell_\xi\|x\|_{m,\xi} \le \|x\| \le u_\xi\|x\|_{m,\xi}, \quad x \in \mathbb{R}^d. \tag{4}$$

Finally, for all $x, y \in \mathbb{R}^d$,

$$|\langle\nabla M_\xi(x), y\rangle| \le \|x\|_{m,\xi}\|y\|_{m,\xi}, \quad \langle\nabla M_\xi(x), x\rangle \ge \|x\|_{m,\xi}^2. \tag{5}$$

We choose $\xi > 0$ small enough that the contraction survives the smoothing step:

$$\kappa\tfrac{u_\xi}{\ell_\xi} < 1, \tag{6}$$

and define the resulting smooth drift rate by

$$\mu_\xi \doteq 2\left(1 - \kappa\tfrac{u_\xi}{\ell_\xi}\right) > 0.$$

Such a choice is possible because $u_\xi/\ell_\xi \to 1$ as $\xi \downarrow 0$.

Lemma 2. *Under Assumption 2.2 and the choice (6), for every $\theta \in \mathbb{R}^d$,*

$$\langle\nabla M_\xi(\theta-\theta_*), f(\theta)\rangle \le -\mu_\xi M_\xi(\theta-\theta_*).$$

The proof is in Appendix B.1. This lemma is how the Moreau envelope is typically used: it converts the contraction of T in $\|\cdot\|$ into a differentiable one-step drift for the smooth energy M_ξ.

3.2 Limitations of Prior Moreau-Based Analyses

In prior works, the drift function is typically taken to be $V(x) = M_\xi(x)$ directly [Chen et al., 2020, 2025]. This is sufficient when $\{Y_n\}$ are i.i.d. But when $\{Y_n\}$ is a Markov chain, additional terms appear and require additional control. There are two typical approaches to control these terms resulting from Markovian noise. The first is based on Srikant and Ying [2019], which constructs an auxiliary stationary Markov chain to compare with, using the mixing property of the Markov chain to show that the bias terms are small on average, see, e.g., Chen et al. [2021], Zhang et al. [2022],

Chen et al. [2024]. Essentially, this approach will establish the recursion between $\mathbb{E}[M_\xi(e_{n+1})]$ and $\mathbb{E}[M_\xi(e_n)]$. This is sufficient for obtaining convergence rates in expectation. But it does not give pathwise control of the bias terms and thus is not sufficient for obtaining almost sure rates. The second approach is the skeleton iterates approach from Qian et al. [2024]. Instead of working with the original timescale $\{1, 2, \dots\}$, Qian et al. [2024] carefully choose a subsequence of time points $\{n_k\}$ and investigate $\{\theta_{n_k}\}$ instead of $\{\theta_n\}$. This allows to construct the recursion between $\mathbb{E}_{n_k}\big[M_\xi(e_{n_{k+1}})\big]$ and $M_\xi(e_{n_k})$. With this recursion, Qian et al. [2024] show that $\{M_\xi(e_{n_k})\}$ is almost a supermartingale and apply the almost supermartingale convergence results from Liu and Yuan [2024] to obtain almost sure convergence rates for $\{M_\xi(e_{n_k})\}$, which is then propagated back to $\{M_\xi(e_n)\}$. The foundation of this approach is recently formally verified by Zhang [2026]. However, as summarized in Table 2, the skeleton-iterate result requires $\eta > 2/3$ and yields squared-error exponents $\zeta < 3\eta/2 - 1$ for $2/3 < \eta < 1$, whereas our result covers $\eta > 1/2$ and gives $\zeta < 2\eta - 1$. Its harmonic guarantee is subpolynomial . In this paper, we provide a third approach that applies a Poisson-equation based correction to the Moreau drift directly.

3.3 Poisson Correction

The key technical novelty of this work is the following drift

$$V_n^\xi \doteq M_\xi(e_n) + \alpha_n \langle \nabla M_\xi(e_n), H_{\theta_n}(Y_n)\rangle + K\alpha_n^2. \tag{7}$$

Here K is a positive constant to be chosen large enough. The middle term is the Poisson correction. It has the same order as one update step. So it does not change the asymptotic size of the drift but cancels a leading Markovian term in the conditional drift. For i.i.d. sampling, the Poisson correction is unnecessary, and the drift reduces to $M_\xi(e_n) + K\alpha_n^2$. The last term is a technical correction to control the residual terms that appear after the Poisson correction. We then show that V_n^ξ is almost a nonnegative supermartingale.

Lemma 3. *Under Assumptions 2.1, 2.3, and 2.4, fix $\xi > 0$, there exists some constant $C_\xi < \infty$ such that*

$$\begin{aligned}\mathbb{E}_n[M_\xi(e_{n+1})] \le & M_\xi(e_n) + \alpha_n \langle \nabla M_\xi(e_n), f(\theta_n)\rangle \\ & + \alpha_n \mathbb{E}_n\big[\langle \nabla M_\xi(e_n), H_{\theta_n}(Y_n)\rangle - \langle \nabla M_\xi(e_{n+1}), H_{\theta_{n+1}}(Y_{n+1})\rangle\big] \\ & + C_\xi \alpha_n^2 \big(\|e_n\|_{m,\xi}^2 + 1\big).\end{aligned}$$

The proof is in Appendix B.2.

Lemma 4. *Under Assumptions 2.1 - 2.5 and the choice (6), there exists $K_\xi < \infty$ such that, for every $K \ge K_\xi$, there exist constants $C_{\xi,K} < \infty$ and $N_{\xi,K} < \infty$ such that, for every $n \ge N_{\xi,K}$,*

$$V_n^\xi \ge \frac{1}{4}\|e_n\|_{m,\xi}^2 \ge 0, \tag{8}$$

$$\mathbb{E}_n[V_{n+1}^\xi] \le \big(1 - \mu_\xi \alpha_n + C_{\xi,K} r_n\big) V_n^\xi + C_{\xi,K} r_n, \quad \text{a.s.}, \tag{9}$$

where $r_n \doteq \alpha_n^2 + |\alpha_{n+1} - \alpha_n|$.

The proof is in Appendix B.3. As established in Lemma 4, the shifted energy V_n^ξ satisfies an almost-supermartingale drift inequality. Therefore, we can directly invoke almost-supermartingale convergence results to obtain the almost sure convergence rates for V_n^ξ, which then translate to the original error $\|\theta_n - \theta_*\|$. The detailed proof of Theorem 1 is deferred to Appendix B.4.

3.4 Limitations of Prior Poisson-Based Analyses

The Poisson equation is a standard tool to decompose Markovian noise (cf. RHS of (2)) into a martingale difference (cf. LHS of (2)) plus a telescoping remainder, enabling ODE-based convergence proofs [Benveniste et al., 1990, Kushner and Yin, 2003, Borkar et al., 2025]. More recently, the Poisson equation also appears in convergence rates analysis of RL algorithms. Chandak and Borkar [2026] use the Poisson equation to handle Markovian noise in deriving concentration bounds for linear TD. Blaser and Zhang [2026] develop novel Poisson-equation-based bounds on noise terms for nonexpansive SA with Markovian noise, proving almost sure convergence of tabular average-reward TD. Chandak et al.

[2025] combine Moreau envelopes with Poisson equation solutions to obtain finite-time mean-square bounds for two-timescale SA with arbitrary norm contractions and Markovian noise. What we perceive as common across all these classical and recent works is telescoping. Namely, obtaining Lemma 3 through Poisson equation decomposition is entirely a standard approach and is not novel at all. Specifically, Lemma 3 can be viewed as a one-timescale specialization of the two-timescale recursion Lemma 3(a) of Chandak et al. [2025]. Lemma 3 is interesting because we can further decompose the Poisson difference term $\alpha_n\langle\nabla M_\xi(e_n), H_{\theta_n}(Y_n)\rangle - \alpha_n\langle\nabla M_\xi(e_{n+1}), H_{\theta_{n+1}}(Y_{n+1})\rangle$ into a telescoping term $\alpha_n\langle\nabla M_\xi(e_n), H_{\theta_n}(Y_n)\rangle - \alpha_{n+1}\langle\nabla M_\xi(e_{n+1}), H_{\theta_{n+1}}(Y_{n+1})\rangle$ plus a residual term $(\alpha_{n+1}-\alpha_n)\langle\nabla M_\xi(e_{n+1}), H_{\theta_{n+1}}(Y_{n+1})\rangle$. The telescoping term will then cancel when we perform summation over n. However, this nice telescoping is not directly available because the terms are inside a conditional expectation indexed by n and thus cannot be directly telescoped. The typical approach to handle this in all the above works is to take total expectations. The n-indexed conditional expectation is then removed , and we can telescope $\alpha_n\mathbb{E}[\langle\nabla M_\xi(e_n), H_{\theta_n}(Y_n)\rangle] - \alpha_{n+1}\mathbb{E}\big[\langle\nabla M_\xi(e_{n+1}), H_{\theta_{n+1}}(Y_{n+1})\rangle\big]$ directly. This is sufficient for obtaining convergence rates in expectation, but it does not give pathwise control of the bias terms and thus is not sufficient for obtaining almost sure rates. Our key novelty is instead to directly incorporate the Poisson correction into the drift function via the novel construction of the Poisson-Moreau drift V_n^ξ in (7), allowing us to obtain an almost-supermartingale structure that preserves the conditional expectation. What is particularly important is the selection of K such that our drift is nonnegative (cf. (8)), which is a key requirement for applying almost-supermartingale convergence results.

3.5 A Side Product

As a byproduct of our analysis, we can take the unconditional expectation on both sides of (9). This generates a convergence rate in L^2 for all $\eta \in (0, 1]$.

Corollary 1. *Let the assumptions of Lemma 4 hold but allow any $0 < \eta \leq 1$ in the learning rate. Then the iterates $\{\theta_n\}$ generated by (SA) satisfy*

1. *If $0 < \eta < 1$, then $\mathbb{E}\Big[\|\theta_n - \theta_*\|^2\Big] = \mathcal{O}(n^{-\eta})$.*
2. *If $\eta = 1$, then*

$$\mathbb{E}\Big[\|\theta_n - \theta_*\|^2\Big] = \begin{cases} \mathcal{O}(n^{-\min\{\mu_\xi\alpha, 1\}}), & \text{if } \mu_\xi\alpha \neq 1, \\ \mathcal{O}(\frac{\log n}{n}), & \text{if } \mu_\xi\alpha = 1. \end{cases}$$

The proof is in Appendix B.5.

Remark 2. *We find this corollary interesting because it does not require square-summability of the learning rate (i.e., $0 < \eta \leq 1/2$ is allowed). This is not new because Lauand and Meyn [2024] already obtain L^p convergence rates for any p and $\eta \in (0, 1)$. We nevertheless include this corollary because our approach is different from that of Lauand and Meyn [2024] and it is conceptually interesting to see that the exact same phenomenon follows naturally from a short Poisson-Moreau drift argument in the arbitrary-norm contractive setting. It is worth noting that the L^2 convergence rates in Chandak et al. [2025] only hold for $\eta \in (1/2, 1]$, according to their theorem statement. It is not clear to us whether their analysis can eventually cover the full range $\eta \in (0, 1]$.*

4 Applications

Although our theoretical framework applies broadly to contractive RL algorithms, including Q-learning and linear TD, we in this section focus on a finite-window generalized TD template with bounded linear features and generalized importance sampling factors ; tabular TD is a special case . This choice is motivated by the fact that almost sure convergence rates for several classical RL algorithms have already been studied using algorithm-specific arguments, whereas pathwise rates for the single-trajectory version of this generalized TD family do not appear to be available. The recursion we study is the unified TD-learning template considered by Chen et al. [2025]. By choosing the generalized importance sampling factors c and ρ, this template recovers on-policy n-step TD [Sutton and Barto, 2018], off-policy n-step TD [Precup et al., 2000, Sutton and Barto, 2018], $Q^\pi(\lambda)$ [Harutyunyan et al., 2016], Tree-Backup [Precup et al., 2000], Retrace [Munos et al., 2016], and Q-trace [Khodadadian et al., 2021]. We therefore do not claim novelty in the definition of the

recursion or in the structural Hurwitz conditions for this TD family. Our contribution is instead to combine these structural properties with the Poisson-Moreau drift to obtain pathwise almost sure rates under Markovian single-trajectory sampling. This is complementary to Chen et al. [2025], who establish maximal high-probability concentration bounds for the same generalized TD template under i.i.d. sampling, but do not provide almost sure convergence rates for the Markovian single-trajectory recursion.

Consider an infinite-horizon discounted Markov decision process (MDP, Bellman [1957]) with a finite state space $\mathcal{S}$, a finite action space $\mathcal{A}$, a discount factor $\gamma \in (0,1)$, a transition function $p : \mathcal{S} \times \mathcal{S} \times \mathcal{A} \to [0,1]$, and a reward function $r : \mathcal{S} \times \mathcal{A} \to \mathbb{R}$. Let π be the target policy and π_b be the behavior policy. We assume strict behavior coverage,

$$\pi_b(a|s) > 0, \quad \forall (s,a) \in \mathcal{S} \times \mathcal{A}, \tag{10}$$

and the trajectory is generated by $A_t \sim \pi_b(\cdot|S_t)$, $S_{t+1} \sim p(\cdot|S_t, A_t)$. We assume the behavior state-action chain $(S_t, A_t)_{t\geq 0}$ is irreducible and aperiodic, with stationary distribution

$$d_b(s,a) \doteq \kappa_b(s)\pi_b(a|s), \quad d_{\min} \doteq \min_{(s,a)\in\mathcal{S}\times\mathcal{A}} d_b(s,a) > 0.$$

Fix a horizon $N \geq 1$ and define the Markovian data window

$$Y_n \doteq (S_n, A_n, S_{n+1}, A_{n+1}, \ldots, S_{n+N}, A_{n+N}).$$

The window chain $(Y_n)_{n\geq 0}$ is finite-state and has a unique invariant distribution on its recurrent class. Denote it by ϖ. Let $\phi : \mathcal{S} \times \mathcal{A} \to \mathbb{R}^d$ be a bounded feature map. The tabular case is recovered by taking $d = |\mathcal{S}||\mathcal{A}|$ and $\phi(s,a) = e_{(s,a)}$ with $e_{(s,a)}$ being the standard basis vector corresponding to the state-action pair (s,a). Let $c, \rho : \mathcal{S} \times \mathcal{A} \to \mathbb{R}_+$ be bounded generalized importance sampling factors. For a window $y = (s_0, a_0, \ldots, s_N, a_N)$, write $\phi_i \doteq \phi(s_i, a_i)$, $c_i \doteq c(s_i, a_i)$, and $\rho_i \doteq \rho(s_i, a_i)$. For $w \in \mathbb{R}^d$, define

$$\delta_i(w; y) \doteq r(s_i, a_i) + \gamma\rho_{i+1}\phi_{i+1}^\top w - \phi_i^\top w, \quad i = 0, \ldots, N-1,$$

with the convention that an empty product is one, and define the generalized TD increment

$$g_{c,\rho}(w, y) \doteq \phi_0 \sum_{i=0}^{N-1} \gamma^i \Big(\prod_{j=1}^{i} c_j\Big)\delta_i(w; y).$$

The single-trajectory generalized TD recursion is

$$w_{n+1} = w_n + a_n g_{c,\rho}(w_n, Y_n), \quad a_n = \tfrac{a}{(n+1)^\eta}. \tag{11}$$

Equivalently, it is a linear stochastic approximation. Define

$$A_{c,\rho}(y) \doteq \phi_0 \sum_{i=0}^{N-1} \gamma^i \Big(\prod_{j=1}^{i} c_j\Big)(\gamma\rho_{i+1}\phi_{i+1} - \phi_i)^\top, \tag{12}$$

$$b_{c,\rho}(y) \doteq -\phi_0 \sum_{i=0}^{N-1} \gamma^i \Big(\prod_{j=1}^{i} c_j\Big) r(s_i, a_i). \tag{13}$$

Then

$$g_{c,\rho}(w, y) = A_{c,\rho}(y)w - b_{c,\rho}(y).$$

Let $\bar{A}_{c,\rho} \doteq \mathbb{E}_\varpi[A_{c,\rho}(Y)]$, $\bar{b}_{c,\rho} \doteq \mathbb{E}_\varpi[b_{c,\rho}(Y)]$. The corresponding mean root is $w^*_{c,\rho} \doteq \bar{A}^{-1}_{c,\rho}\bar{b}_{c,\rho}$, whenever $\bar{A}_{c,\rho}$ is nonsingular.

The only algorithm-specific structural input is the following standard Hurwitz condition for generalized TD.

Condition 4.1 (Generalized TD Hurwitz condition)**.** *The matrix $\bar{A}_{c,\rho}$ is Hurwitz.*

Under Condition 4.1, there exist $\beta_{c,\rho} > 0$, a norm $\|\cdot\|_{\mathrm{GTD}}$, and a constant $\kappa_{\mathrm{GTD}} \in (0,1)$ such that

$$T_{\mathrm{GTD}}(w) \doteq w + \beta_{c,\rho}\big(\bar{A}_{c,\rho}w - \bar{b}_{c,\rho}\big)$$

is a contraction:

$$\|T_{\mathrm{GTD}}(w) - T_{\mathrm{GTD}}(w')\|_{\mathrm{GTD}} \leq \kappa_{\mathrm{GTD}}\|w - w'\|_{\mathrm{GTD}}, \quad w, w' \in \mathbb{R}^d. \tag{14}$$

This follows from the standard Lyapunov reduction for Hurwitz linear systems.

Table 3: Examples of generalized importance sampling factors covered by the recursion (11). Here $\varrho(s,a) \doteq \pi(a|s)/\pi_b(a|s)$. The theorem below applies whenever the associated Hurwitz condition holds.

Algorithm	$c(s,a)$	$\rho(s,a)$
On-policy n-step TD	1	1
Off-policy n-step TD	$\varrho(s,a)$	$\varrho(s,a)$
Q-trace	$\min\{\bar{c}, \varrho(s,a)\}$	$\min\{\bar{\rho}, \varrho(s,a)\}$
Retrace(λ)	$\lambda \min\{1, \varrho(s,a)\}$	$\varrho(s,a)$
Tree-Backup(λ)	$\lambda\pi(a|s)$	$\varrho(s,a)$
$Q^\pi(\lambda)$	λ	$\varrho(s,a)$

In particular, let $P_{c,\rho} \succ 0$ be the unique solution of

$$\bar{A}_{c,\rho}^\top P_{c,\rho} + P_{c,\rho}\bar{A}_{c,\rho} = -I, \qquad \|x\|_{\mathrm{GTD}} \doteq \sqrt{x^\top P_{c,\rho} x}.$$

This is equivalent to the Euclidean norm, with constants $\sqrt{\lambda_{\min}(P_{c,\rho})}$ and $\sqrt{\lambda_{\max}(P_{c,\rho})}$. Choose $\beta_{c,\rho} > 0$ sufficiently small and set

$$\kappa_{\mathrm{GTD}} \doteq \left\| P_{c,\rho}^{1/2}(I + \beta_{c,\rho}\bar{A}_{c,\rho})P_{c,\rho}^{-1/2} \right\|_2 < 1.$$

Here $\|\cdot\|_2$ is the Euclidean operator norm, and $P_{c,\rho}$, $\beta_{c,\rho}$, and κ_{GTD} are used only in the analysis .

Condition 4.1 is already verified for important members of the generalized TD family. In the on-policy case $\pi_b = \pi$ with $c = \rho = 1$, Chen et al. [2025] show that $\bar{A}_{c,\rho}$ is Hurwitz. In the off-policy case, their generalized TD analysis gives sufficient conditions on c, ρ and the horizon N that imply the same Hurwitz property.

To apply Theorem 1 to the actual recursion (11), set

$$\theta_n = w_n, \quad \theta_* = w_{c,\rho}^*, \quad F_{\mathrm{GTD}}(w,y) \doteq \beta_{c,\rho}\big(A_{c,\rho}(y)w - b_{c,\rho}(y)\big), \quad \alpha_n \doteq \tfrac{a_n}{\beta_{c,\rho}}. \tag{15}$$

Then (11) is exactly in the form (SA), and its averaged map is T_{GTD}.

Proposition 1. *Let (10) and Condition 4.1 hold. Let the behavior state-action chain be irreducible and aperiodic. Then the single-trajectory generalized TD recursion (11), written in the form (15), satisfies Assumptions 2.1, 2.2, 2.3, and 2.4 with contraction norm $\|\cdot\|_{\mathrm{GTD}}$ and contraction factor κ_{GTD}.*

The proof is in Appendix C.

Combining Proposition 1 with Theorem 1 gives the following pathwise rate.

Theorem 2 (Almost sure rate for single-trajectory generalized TD)**.** *Under the assumptions of Proposition 1, let $a_n = a/(n+1)^\eta$ with $a > 0$ and $1/2 < \eta \le 1$. Then*

$$(n+1)^\zeta \|w_n - w_{c,\rho}^*\|_{\mathrm{GTD}}^2 \longrightarrow 0 \quad \text{a.s.} \tag{16}$$

If $1/2 < \eta < 1$, then (16) holds for every $\zeta \in (0, 2\eta - 1)$. If $\eta = 1$, then (16) holds for every

$$\zeta \in \left(0, \min\left\{1, 2(1-\kappa_{\mathrm{GTD}})\tfrac{a}{\beta_{c,\rho}}\right\}\right).$$

Here, for harmonic learning rates, the conclusion of Theorem 2 holds for every $\zeta < 1$ whenever $2(1-\kappa_{\mathrm{GTD}})a/\beta_{c,\rho} \ge 1$. No projection or bounded-iterate assumption is required. Appendix D provides a numerical illustration of our theoretical results.

5 Related Work

Convergence Modes. Almost sure convergence rates are in general challenging to obtain. This can be seen from the fact that a rich literature on RL has been devoted to establishing almost sure convergence of various algorithms [Watkins, 1989, Watkins and Dayan, 1992, Jaakkola et al., 1993,

Tsitsiklis, 1994, Tsitsiklis and Van Roy, 1997, Tsitsiklis and Roy, 1999, Konda and Tsitsiklis, 1999, Abounadi et al., 2001, Sutton et al., 2008, 2009, Maei, 2011, Yu, 2015, 2017, Zhang et al., 2020, 2021, Wan et al., 2021, Qian and Zhang, 2025, Wang and Zhang, 2026, Blaser et al., 2026], but only a few works have established almost sure convergence rates (Table 1). For classical discounted linear TD, Tadić [2002] establishes the almost sure squared-error rate $\mathcal{O}((\log\log n)/n)$ under specialized assumptions and harmonic learning rates. This is typically the optimal almost sure squared-error benchmark in classical linear SA and TD settings. Our theorem covers nonlinear mean fields and arbitrary contraction norms under uniform Poisson regularity; if $2(1-\kappa)\alpha \geq 1$, harmonic learning rates yield $o(n^{-\zeta})$ almost surely for every $\zeta < 1$, approaching the same n^{-1} polynomial exponent.

SA and RL have also been extensively studied through the lenses of other convergence modes. Notable works on L^2 convergence include Polyak and Juditsky [1992], Moulines and Bach [2011], Bhandari et al. [2018], Srikant and Ying [2019], Zou et al. [2019], Mou et al. [2020], Chen et al. [2021], Liu and Olshevsky [2021], Zhang and Whiteson [2022], Liu et al. [2025c], Xie et al. [2025] , Mitra [2025], Lee and Orabona [2025] . Notable works on L^p convergence include Godichon-Baggioni [2016, 2019], Gupta et al. [2019], Srikant and Ying [2019], Qu and Wierman [2020], Durmus et al. [2021], Jentzen et al. [2021], Wang et al. [2021], Chen et al. [2022], Liu et al. [2025b]. Notable works on high probability concentration include Korda and La [2015], Dalal et al. [2018], Shah and Xie [2018], Thoppe and Borkar [2019], Wainwright [2019], Borkar [2021], Li et al. [2020], Mou et al. [2020], Qu and Wierman [2020], Patil et al. [2023], Li et al. [2024a,b], Samsonov et al. [2024b], Durmus et al. [2025], Chandak et al. [2025], Chen et al. [2025], Khodadadian and Zubeldia [2025], Chandak and Borkar [2026]. Notable works on distributional convergence, particularly, (functional) central limit theorem (CLT), include Ruppert [1988], Benveniste et al. [1990], Polyak and Juditsky [1992], Borkar [2009], Li et al. [2023], Hu et al. [2024], Samsonov et al. [2024a], Zhang and Xie [2024], Borkar et al. [2025], Srikant [2025], Liu [2026], Wu et al. [2026]. This work on almost sure convergence rates provides a precise pathwise characterization and is complementary to these works on other convergence modes.

Poisson Equation. The foundational theory (e.g., existence of solutions, growth bounds, square integrability) is developed by Glynn and Iglehart [1989] and Glynn and Meyn [1996], who also use these foundations to derive functional CLTs and perturbation results for Markov processes. Makowski and Shwartz [1994] provided a probabilistic treatment of existence, uniqueness, growth estimates, and parametric dependence for the Poisson equation on countable state spaces. In addition to the use of Poisson equations in SA and RL surveyed in Section 3.4, in the MCMC literature, Poisson equations are also used to construct control variates for variance reduction [Dellaportas and Kontoyiannis, 2012] and unbiased estimators via coupling [Douc et al., 2026].

6 Conclusion

Through the introduction of a Poisson-corrected Moreau drift, this work establishes sharper almost sure convergence rates for contractive stochastic approximation under Markovian noise. A few questions remain open. First, Corollary 1 gives L^2 rates for $\eta \in (0, 1]$, but Theorem 1 only gives almost sure rates for $\eta \in (1/2, 1]$. It is not clear to us whether our methodology can yield almost sure rates for the full range $\eta \in (0, 1]$. Second, although our almost sure rates are sharper than Qian et al. [2024], they are still not tight. It is not clear to us whether we can fully match the $O(n^{-1}\log\log n)$ rate given by the law of the iterated logarithm. Third, in addition to almost sure rates, Qian et al. [2024] also establish high probability concentration bounds with exponentially decaying tails for contractive SA with Markovian noise. It is not clear to us whether our Poisson-Moreau framework can be refined to yield complementary concentration bounds with exponential tails.

Acknowledgments and Disclosure of Funding

This work is supported in part by the US National Science Foundation under the awards III-2128019, SLES-2331904, and CAREER-2442098, the Commonwealth Cyber Initiative's Central Virginia Node under the award VV-1Q26-001, VV-1Q27-007, and VV-1Q27005, 4VA at UVA, a Cisco Faculty Research Award, and a Google Research Award.

References

Jinane Abounadi, Dimitri Bertsekas, and Vivek S Borkar. Learning algorithms for markov decision processes with average cost. SIAM Journal on Control and Optimization, 2001.

Richard Bellman. A markovian decision process. Journal of Mathematics and Mechanics, 1957.

Albert Benveniste, Michel Métivier, and Pierre Priouret. Adaptive Algorithms and Stochastic Approximations. Springer, 1990.

Jalaj Bhandari, Daniel Russo, and Raghav Singal. A finite time analysis of temporal difference learning with linear function approximation. In Proceedings of the Conference on Learning Theory, 2018.

Ethan Blaser and Shangtong Zhang. Asymptotic and finite sample analysis of nonexpansive stochastic approximations with markovian noise. In Proceedings of the AAAI Conference on Artificial Intelligence, 2026.

Ethan Blaser, Jiuqi Wang, and Shangtong Zhang. Almost sure convergence of differential temporal difference learning for average reward markov decision processes. In Proceedings of the International Conference on Artificial Intelligence and Statistics, 2026.

Vivek Borkar, Shuhang Chen, Adithya Devraj, Ioannis Kontoyiannis, and Sean Meyn. The ODE method for asymptotic statistics in stochastic approximation and reinforcement learning. The Annals of Applied Probability, 2025.

Vivek S Borkar. Stochastic approximation: a dynamical systems viewpoint. Springer, 2009.

Vivek S Borkar. A concentration bound for contractive stochastic approximation. Systems & Control Letters, 2021.

Siddharth Chandak and Vivek S Borkar. A concentration bound for TD (0) with function approximation. Stochastic Systems, 2026.

Siddharth Chandak, Shaan Ul Haque, and Nicholas Bambos. Finite-time bounds for two-time-scale stochastic approximation with arbitrary norm contractions and markovian noise. In IEEE Conference on Decision and Control, 2025.

Zaiwei Chen, Siva Theja Maguluri, Sanjay Shakkottai, and Karthikeyan Shanmugam. Finite-sample analysis of contractive stochastic approximation using smooth convex envelopes. In Advances in Neural Information Processing Systems, 2020.

Zaiwei Chen, Siva Theja Maguluri, Sanjay Shakkottai, and Karthikeyan Shanmugam. A Lyapunov theory for finite-sample guarantees of asynchronous Q-learning and TD-learning variants. arXiv Preprint, 2021.

Zaiwei Chen, Sheng Zhang, Thinh T Doan, John-Paul Clarke, and Siva Theja Maguluri. Finite-sample analysis of nonlinear stochastic approximation with applications in reinforcement learning. Automatica, 2022.

Zaiwei Chen, Siva T Maguluri, Sanjay Shakkottai, and Karthikeyan Shanmugam. A Lyapunov theory for finite-sample guarantees of markovian stochastic approximation. Operations Research, 2024.

Zaiwei Chen, Siva Theja Maguluri, and Martin Zubeldia. Concentration of contractive stochastic approximation: Additive and multiplicative noise. The Annals of Applied Probability, 2025.

Edwin KP Chong, I-Jeng Wang, and Sanjeev R Kulkarni. Noise conditions for prespecified convergence rates of stochastic approximation algorithms. IEEE Transactions on Information Theory, 1999.

Gal Dalal, Balázs Szörényi, Gugan Thoppe, and Shie Mannor. Finite sample analyses for TD(0) with function approximation. In Proceedings of the AAAI Conference on Artificial Intelligence, 2018.

Petros Dellaportas and Ioannis Kontoyiannis. Control variates for estimation based on reversible markov chain monte carlo samplers. Journal of the Royal Statistical Society Series B: Statistical Methodology, 2012.

Randal Douc, Pierre E Jacob, Anthony Lee, and Dootika Vats. Solving the poisson equation using coupled markov chains. The Annals of Statistics, 2026.

Alain Durmus, Eric Moulines, Alexey Naumov, Sergey Samsonov, and Hoi-To Wai. On the stability of random matrix product with markovian noise: Application to linear stochastic approximation and td learning. In Proceedings of the Conference on Learning Theory, 2021.

Alain Durmus, Eric Moulines, Alexey Naumov, and Sergey Samsonov. Finite-time high-probability bounds for polyak–ruppert averaged iterates of linear stochastic approximation. Mathematics of Operations Research, 2025.

Peter W Glynn and Donald L Iglehart. Importance sampling for stochastic simulations. Management Science, 1989.

Peter W Glynn and Sean P Meyn. A liapounov bound for solutions of the poisson equation. The Annals of Probability, 1996.

Antoine Godichon-Baggioni. Estimating the geometric median in hilbert spaces with stochastic gradient algorithms: Lp and almost sure rates of convergence. Journal of Multivariate Analysis, 2016.

Antoine Godichon-Baggioni. Lp and almost sure rates of convergence of averaged stochastic gradient algorithms: locally strongly convex objective. ESAIM: Probability and Statistics, 2019.

Harsh Gupta, Rayadurgam Srikant, and Lei Ying. Finite-time performance bounds and adaptive learning rate selection for two time-scale reinforcement learning. In Advances in Neural Information Processing Systems, 2019.

Anna Harutyunyan, Marc G Bellemare, Tom Stepleton, and Rémi Munos. Q(λ) with off-policy corrections. In Proceedings of the International Conference on Algorithmic Learning Theory, 2016.

Jie Hu, Vishwaraj Doshi, and Do Young Eun. Central limit theorem for two-timescale stochastic approximation with markovian noise: Theory and applications. In Proceedings of the International Conference on Artificial Intelligence and Statistics, 2024.

Tommi Jaakkola, Michael Jordan, and Satinder Singh. Convergence of stochastic iterative dynamic programming algorithms. In Advances in Neural Information Processing Systems, 1993.

Arnulf Jentzen, Benno Kuckuck, Ariel Neufeld, and Philippe von Wurstemberger. Strong error analysis for stochastic gradient descent optimization algorithms. IMA Journal of Numerical Analysis, 2021.

Rajeeva Laxman Karandikar and Mathukumalli Vidyasagar. Convergence rates for stochastic approximation: Biased noise with unbounded variance, and applications. Journal of Optimization Theory and Applications, 2024.

Sajad Khodadadian and Martin Zubeldia. A general-purpose theorem for high-probability bounds of stochastic approximation with polyak averaging. In Advances in Neural Information Processing Systems, 2025.

Sajad Khodadadian, Zaiwei Chen, and Siva Theja Maguluri. Finite-sample analysis of off-policy natural actor-critic algorithm. In Proceedings of the International Conference on Machine Learning, 2021.

Vijay R. Konda and John N. Tsitsiklis. Actor-critic algorithms. In Advances in Neural Information Processing Systems, 1999.

Nathaniel Korda and Prashanth La. On TD (0) with function approximation: Concentration bounds and a centered variant with exponential convergence. In Proceedings of the International Conference on Machine Learning, 2015.

Michael A Kouritzin and Samira Sadeghi. Convergence rates and decoupling in linear stochastic approximation algorithms. SIAM Journal on Control and Optimization, 2015.

Valery Koval and Rainer Schwabe. A law of the iterated logarithm for stochastic approximation procedures in d-dimensional euclidean space. Stochastic Processes and Their Applications, 2003.

Harold Kushner and G George Yin. Stochastic approximation and recursive algorithms and applications. Springer Science & Business Media, 2003.

Caio Kalil Lauand and Sean Meyn. Revisiting step-size assumptions in stochastic approximation. arXiv Preprint, 2024.

Wei-Cheng Lee and Francesco Orabona. A finite-time analysis of TD learning with linear function approximation without projections nor strong convexity. arXiv Preprint, 2025.

Gen Li, Yuting Wei, Yuejie Chi, Yuantao Gu, and Yuxin Chen. Sample complexity of asynchronous Q-learning: Sharper analysis and variance reduction. In Advances in Neural Information Processing Systems, 2020.

Gen Li, Changxiao Cai, Yuxin Chen, Yuting Wei, and Yuejie Chi. Is Q-learning minimax optimal? a tight sample complexity analysis. Operations Research, 2024a.

Gen Li, Weichen Wu, Yuejie Chi, Cong Ma, Alessandro Rinaldo, and Yuting Wei. High-probability sample complexities for policy evaluation with linear function approximation. IEEE Transactions on Information Theory, 2024b.

Xiang Li, Wenhao Yang, Jiadong Liang, Zhihua Zhang, and Michael I Jordan. A statistical analysis of polyak-ruppert averaged q-learning. In Proceedings of the International Conference on Artificial Intelligence and Statistics, 2023.

Yuqing Liang, Dongpo Xu, Naimin Zhang, and Danilo P Mandic. Almost sure convergence of stochastic composite objective mirror descent for non-convex non-smooth optimization. Optimization Letters, 2023.

Jun Liu and Ye Yuan. On almost sure convergence rates of stochastic gradient methods. In Proceedings of the Conference on Learning Theory, 2022.

Jun Liu and Ye Yuan. Almost sure convergence rates analysis and saddle avoidance of stochastic gradient methods. Journal of Machine Learning Research, 2024.

Rui Liu and Alex Olshevsky. Temporal difference learning as gradient splitting. In Proceedings of the International Conference on Machine Learning, 2021.

Shuze Liu, Shuhang Chen, and Shangtong Zhang. The ODE method for stochastic approximation and reinforcement learning with markovian noise. Journal of Machine Learning Research, 2025a.

Xingtu Liu. Central limit theorems for asynchronous averaged q-learning. In Learning for Dynamics and Control Conference, 2026.

Xinyu Liu, Zixuan Xie, and Shangtong Zhang. Extensions of robbins-siegmund theorem with applications in reinforcement learning. arXiv Preprint, 2025b.

Xinyu Liu, Zixuan Xie, and Shangtong Zhang. Linear Q-learning does not diverge in L^2: Convergence rates to a bounded set. In Proceedings of the International Conference on Machine Learning, 2025c.

Hamid Reza Maei. Gradient temporal-difference learning algorithms. PhD thesis, University of Alberta, 2011.

Armand M Makowski and Adam Shwartz. On the poisson equation for countable markov chains: Existence of solutions and parameter dependence by probabilistic methods. Technical report, 1994.

Sean P Meyn and Richard L Tweedie. Markov chains and stochastic stability. Springer Science & Business Media, 2012.

Aritra Mitra. A simple finite-time analysis of TD learning with linear function approximation. IEEE Transactions on Automatic Control, 2025.

Wenlong Mou, Chris Junchi Li, Martin J Wainwright, Peter L Bartlett, and Michael I Jordan. On linear stochastic approximation: Fine-grained polyak-ruppert and non-asymptotic concentration. In Proceedings of the Conference on Learning Theory, 2020.

Eric Moulines and Francis Bach. Non-asymptotic analysis of stochastic approximation algorithms for machine learning. In Advances in Neural Information Processing Systems, 2011.

Rémi Munos, Tom Stepleton, Anna Harutyunyan, and Marc Bellemare. Safe and efficient off-policy reinforcement learning. In Advances in Neural Information Processing Systems, 2016.

Gandharv Patil, Prashanth L. A., Dheeraj Nagaraj, and Doina Precup. Finite time analysis of temporal difference learning with linear function approximation: Tail averaging and regularisation. In Proceedings of the International Conference on Artificial Intelligence and Statistics, 2023.

Mariane Pelletier. On the almost sure asymptotic behaviour of stochastic algorithms. Stochastic Processes and Their Applications, 1998.

Boris T Polyak and Anatoli B Juditsky. Acceleration of stochastic approximation by averaging. SIAM Journal on Control and Optimization, 1992.

Doina Precup, Richard S. Sutton, and Satinder P. Singh. Eligibility traces for off-policy policy evaluation. In Proceedings of the International Conference on Machine Learning, 2000.

Martin L Puterman. Markov decision processes: discrete stochastic dynamic programming. John Wiley & Sons, 2014.

Xiaochi Qian and Shangtong Zhang. Revisiting a design choice in gradient temporal difference learning. In Proceedings of the International Conference on Learning Representations, 2025.

Xiaochi Qian, Zixuan Xie, Xinyu Liu, and Shangtong Zhang. Almost sure convergence rates and concentration of stochastic approximation and reinforcement learning with markovian noise. arXiv Preprint, 2024.

Guannan Qu and Adam Wierman. Finite-time analysis of asynchronous stochastic approximation and q-learning. In Proceedings of the Conference on Learning Theory, 2020.

Herbert Robbins and Sutton Monro. A stochastic approximation method. The Annals of Mathematical Statistics, 1951.

Herbert Robbins and David Siegmund. A convergence theorem for non negative almost supermartingales and some applications. Optimizing Methods in Statistics, 1971.

David Ruppert. Efficient estimations from a slowly convergent robbins-monro process. Technical Report, 1988.

Sergey Samsonov, Eric Moulines, Qi-Man Shao, Zhuo-Song Zhang, and Alexey Naumov. Gaussian approximation and multiplier bootstrap for polyak-ruppert averaged linear stochastic approximation with applications to td learning. In Advances in Neural Information Processing Systems, 2024a.

Sergey Samsonov, Daniil Tiapkin, Alexey Naumov, and Eric Moulines. Improved high-probability bounds for the temporal difference learning algorithm via exponential stability. In Proceedings of the Conference on Learning Theory, 2024b.

Othmane Sebbouh, Robert M Gower, and Aaron Defazio. Almost sure convergence rates for stochastic gradient descent and stochastic heavy ball. In Proceedings of the Conference on Learning Theory, 2021.

Devavrat Shah and Qiaomin Xie. Q-learning with nearest neighbors. In Advances in Neural Information Processing Systems, 2018.

Rayadurgam Srikant. Rates of convergence in the central limit theorem for markov chains, with an application to td learning. Mathematics of Operations Research, 2025.

Rayadurgam Srikant and Lei Ying. Finite-time error bounds for linear stochastic approximation and td learning. In Proceedings of the Conference on Learning Theory, 2019.

Richard S. Sutton. Learning to predict by the methods of temporal differences. Machine Learning, 1988.

Richard S Sutton and Andrew G Barto. Reinforcement Learning: An Introduction (2nd Edition). MIT Press, 2018.

Richard S. Sutton, Csaba Szepesvári, and Hamid Reza Maei. A convergent o(n) temporal-difference algorithm for off-policy learning with linear function approximation. In Advances in Neural Information Processing Systems, 2008.

Richard S. Sutton, Hamid Reza Maei, Doina Precup, Shalabh Bhatnagar, David Silver, Csaba Szepesvári, and Eric Wiewiora. Fast gradient-descent methods for temporal-difference learning with linear function approximation. In Proceedings of the International Conference on Machine Learning, 2009.

Csaba Szepesvári. The asymptotic convergence-rate of Q-learning. In Advances in Neural Information Processing Systems, 1997.

Vladislav B Tadić. On the almost sure rate of convergence of temporal-difference learning algorithms. IFAC Proceedings Volumes, 2002.

Vladislav B Tadic. On the almost sure rate of convergence of linear stochastic approximation algorithms. IEEE Transactions on Information Theory, 2004.

Gugan Thoppe and Vivek Borkar. A concentration bound for stochastic approximation via alekseev's formula. Stochastic Systems, 2019.

John N Tsitsiklis. Asynchronous stochastic approximation and Q-learning. Machine Learning, 1994.

John N. Tsitsiklis and Benjamin Van Roy. Average cost temporal-difference learning. Automatica, 1999.

John N. Tsitsiklis and Benjamin Van Roy. An analysis of temporal-difference learning with function approximation. IEEE Transactions on Automatic Control, 1997.

M Vidyasagar. Convergence of stochastic approximation via martingale and converse lyapunov methods. Mathematics of Control, Signals, and Systems, 2023.

Martin J Wainwright. Stochastic approximation with cone-contractive operators: Sharp l_∞-bounds for q-learning. arXiv Preprint, 2019.

Yi Wan, Abhishek Naik, and Richard S. Sutton. Learning and planning in average-reward markov decision processes. In Proceedings of the International Conference on Machine Learning, 2021.

Hongjian Wang, Mert Gurbuzbalaban, Lingjiong Zhu, Umut Simsekli, and Murat A Erdogdu. Convergence rates of stochastic gradient descent under infinite noise variance. In Advances in Neural Information Processing Systems, 2021.

Jiuqi Wang and Shangtong Zhang. Almost sure convergence of linear temporal difference learning with arbitrary features. Journal of Machine Learning Research, 2026.

Christopher JCH Watkins and Peter Dayan. Q-learning. Machine Learning, 1992.

Christopher John Cornish Hellaby Watkins. Learning from delayed rewards. PhD thesis, King's College, Cambridge, 1989.

Simon Weissmann, Sara Klein, Waïss Azizian, and Leif Döring. Almost sure convergence of stochastic gradient methods under gradient domination. Transactions on Machine Learning Research, 2025.

Weichen Wu, Gen Li, Yuting Wei, and Alessandro Rinaldo. Statistical inference for policy evaluation with temporal difference learning. Reinforcement Learning Journal, 2026.

Zixuan Xie, Xinyu Liu, Rohan Chandra, and Shangtong Zhang. Finite sample analysis of linear temporal difference learning with arbitrary features. In Advances in Neural Information Processing Systems, 2025.

Huizhen Yu. On convergence of emphatic temporal-difference learning. In Proceedings of the Conference on Learning Theory, 2015.

Huizhen Yu. On convergence of some gradient-based temporal-differences algorithms for off-policy learning. arXiv Preprint, 2017.

Shangtong Zhang. Towards formalizing reinforcement learning theory: A Robbins-Siegmund approach. Reinforcement Learning Journal, 2026.

Shangtong Zhang and Shimon Whiteson. Truncated emphatic temporal difference methods for prediction and control. Journal of Machine Learning Research, 2022.

Shangtong Zhang, Bo Liu, Hengshuai Yao, and Shimon Whiteson. Provably convergent two-timescale off-policy actor-critic with function approximation. In Proceedings of the International Conference on Machine Learning, 2020.

Shangtong Zhang, Yi Wan, Richard S. Sutton, and Shimon Whiteson. Average-reward off-policy policy evaluation with function approximation. In Proceedings of the International Conference on Machine Learning, 2021.

Shangtong Zhang, Remi Tachet, and Romain Laroche. Global optimality and finite sample analysis of softmax off-policy actor critic under state distribution mismatch. Journal of Machine Learning Research, 2022.

Yixuan Zhang and Qiaomin Xie. Constant stepsize Q-learning: Distributional convergence, bias and extrapolation. Reinforcement Learning Journal, 2024.

Shaofeng Zou, Tengyu Xu, and Yingbin Liang. Finite-sample analysis for SARSA with linear function approximation. In Advances in Neural Information Processing Systems, 2019.

A Auxiliary Lemmas

Lemma 5 (Discrete Gronwall Inequality, Lemma 8 in Section 11.2 of Borkar [2009])**.** *For nonnegative real sequences* $\{x_n, n \geq 0\}$ *and* $\{a_n, n \geq 0\}$ *and scalar* $L \geq 0$*, it holds*

$$x_{n+1} \leq C + L\sum_{i=0}^{n} a_i x_i \quad \forall n \implies x_{n+1} \leq (C + x_0)\exp(L\sum_{i=0}^{n} a_i) \quad \forall n.$$

Lemma 6. *Let* $\{X_n\}_{n\geq N}$ *be nonnegative, adapted, and integrable. Suppose that*

$$\mathbb{E}[X_{n+1}|\mathcal{F}_n] \leq (1-\beta_n)X_n + b_n \quad \textit{a.s. for every } n \geq N, \tag{17}$$

where $0 \leq \beta_n \leq 1$ *and* $b_n \geq 0$ *are deterministic. Let* $(q_n)_{n\geq N}$ *be a positive deterministic sequence and define*

$$\gamma_n \doteq 1 - \frac{q_{n+1}}{q_n}(1-\beta_n). \tag{18}$$

Assume that, after increasing N *if necessary,*

$$0 \leq \gamma_n \leq 1, \quad \sum_{n=N}^{\infty} \gamma_n = \infty, \quad \sum_{n=N}^{\infty} q_{n+1}b_n < \infty.$$

Then

$$q_n X_n \longrightarrow 0 \quad \textit{a.s.}$$

Proof. Multiplying (17) by q_{n+1} and using (18), we obtain

$$\begin{aligned}\mathbb{E}[q_{n+1}X_{n+1}|\mathcal{F}_n] &\leq q_{n+1}(1-\beta_n)X_n + q_{n+1}b_n \\ &= (1-\gamma_n)q_nX_n + q_{n+1}b_n \\ &= q_nX_n - \gamma_n q_n X_n + q_{n+1}b_n.\end{aligned}$$

Since $\sum_n q_{n+1}b_n < \infty$, the Robbins–Siegmund almost-supermartingale theorem [Robbins and Siegmund, 1971] implies that q_nX_n converges almost surely to a finite random variable and

$$\sum_{n=N}^{\infty} \gamma_n q_n X_n < \infty \quad \text{a.s.}$$

If the almost sure limit of q_nX_n were positive on an event of positive probability, then on that event q_nX_n would be bounded below by a positive constant for all sufficiently large n, contradicting $\sum_n \gamma_n = \infty$. Hence the limit is zero almost surely. □

Lemma 7. *Let*

$$\alpha_n = \frac{\alpha}{(n+1)^\eta}, \quad \alpha > 0, \quad \frac{1}{2} < \eta \leq 1.$$

Let $\{X_n\}$ *be nonnegative, adapted, and integrable. Suppose that, on a deterministic tail,*

$$\mathbb{E}[X_{n+1}|\mathcal{F}_n] \leq \big(1 - c_0\alpha_n + C_0 r_n\big)X_n + C_0 r_n \tag{19}$$

for constants $c_0 > 0$ *and* $C_0 < \infty$*. Then the following statements hold.*

1. *If* $\frac{1}{2} < \eta < 1$*, then* $(n+1)^\zeta X_n \to 0$ *almost surely for every* $0 \leq \zeta < 2\eta - 1$.
2. *If* $\eta = 1$*, then* $(n+1)^\zeta X_n \to 0$ *almost surely for every* $0 \leq \zeta < \min\{1, c_0\alpha\}$.

Proof. First,

$$|\alpha_{n+1} - \alpha_n| = \alpha\big((n+1)^{-\eta} - (n+2)^{-\eta}\big) \leq \alpha\eta(n+1)^{-\eta-1} \leq \frac{\eta}{\alpha}\alpha_n^2,$$

where the last inequality uses $\eta \leq 1$. Thus $r_n \leq C\alpha_n^2$.

Fix any $c \in (0, c_0)$. Since $\alpha_n \to 0$, (19) implies, after increasing the deterministic tail index,

$$\mathbb{E}[X_{n+1}|\mathcal{F}_n] \leq (1 - c\alpha_n)X_n + C\alpha_n^2, \quad 0 \leq c\alpha_n \leq 1. \tag{20}$$

Assume first that $\frac{1}{2} < \eta < 1$, and let $0 \le \zeta < 2\eta - 1$. Apply Lemma 6 to (20) with $q_n = (n+1)^\zeta$, $\beta_n = c\alpha_n$, and $b_n = C\alpha_n^2$. The weighted error term is summable because

$$\sum_n q_{n+1} b_n \le C \sum_n (n+2)^\zeta (n+1)^{-2\eta} < \infty.$$

Moreover,

$$\gamma_n = 1 - \frac{q_{n+1}}{q_n}(1-\beta_n) = \frac{c\alpha}{(n+1)^\eta} + O((n+1)^{-1}).$$

Since $\eta < 1$, the positive $(n+1)^{-\eta}$ term dominates the $(n+1)^{-1}$ term. Thus, on a further deterministic tail, $0 \le \gamma_n \le 1$ and $\sum_n \gamma_n = \infty$. Lemma 6 gives $(n+1)^\zeta X_n \to 0$ almost surely.

Now assume $\eta = 1$, and let $0 \le \zeta < \min\{1, c_0\alpha\}$. Choose $c \in (0, c_0)$ so that $\zeta < c\alpha$. With the same choices of q_n, β_n, and b_n, the weighted error term is summable because $\zeta < 1$. Also,

$$\gamma_n = 1 - \left(1 + \frac{1}{n+1}\right)^\zeta \left(1 - \frac{c\alpha}{n+1}\right) = \frac{c\alpha - \zeta}{n+1} + O\left(\frac{1}{(n+1)^2}\right).$$

Since $c\alpha - \zeta > 0$, the coefficients γ_n are eventually in $[0,1]$ and satisfy $\sum_n \gamma_n = \infty$. Lemma 6 gives $(n+1)^\zeta X_n \to 0$ almost surely. □

Lemma 8. *Under Assumptions 2.3 and 2.4, fix $\xi > 0$. There exists $K_\xi < \infty$ such that, for every $K \ge K_\xi$, there exist some constants $C'_{\xi,K}$ and $N_{\xi,K}$ such that, for every $n \ge N_{\xi,K}$, we have*

$$V_n^\xi \ge \frac{1}{4}\|e_n\|_{m,\xi}^2, \tag{21}$$

$$\alpha_n \left|V_n^\xi - M_\xi(e_n)\right| \le C'_{\xi,K} r_n \left(\|e_n\|_{m,\xi}^2 + 1\right). \tag{22}$$

Proof. By the norm comparison from Lemma 1, we have $\|\cdot\|_{m,\xi} \le \ell_\xi^{-1}\|\cdot\|$ and $\|\cdot\| \le u_\xi \|\cdot\|_{m,\xi}$. We then obtain

$$\begin{aligned}
\|H_{\theta_n}(Y_n)\|_{m,\xi} &\le \ell_\xi^{-1}\|H_{\theta_n}(Y_n)\| \\
&\le \ell_\xi^{-1}\Big(\|H_{\theta_n}(Y_n) - H_{\theta_*}(Y_n)\| + \|H_{\theta_*}(Y_n)\|\Big) \\
&\le \ell_\xi^{-1} L_H\big(\|e_n\| + 1\big) && \text{(By Assumption 2.4)} \\
&\le \ell_\xi^{-1} L_H\big(u_\xi \|e_n\|_{m,\xi} + 1\big) \\
&\le L_H \ell_\xi^{-1}(u_\xi + 1)\big(\|e_n\|_{m,\xi} + 1\big).
\end{aligned}$$

Therefore,

$$\begin{aligned}
|\langle \nabla M_\xi(e_n), H_{\theta_n}(Y_n)\rangle| &\le \|e_n\|_{m,\xi}\|H_{\theta_n}(Y_n)\|_{m,\xi} && \text{(By (5))} \\
&\le L_H \ell_\xi^{-1}(u_\xi + 1)\|e_n\|_{m,\xi}\big(\|e_n\|_{m,\xi} + 1\big). && (23)
\end{aligned}$$

Since α_n is decreasing, choose a deterministic index $N_\xi^{(1)}$ such that, for every $n \ge N_\xi^{(1)}$,

$$\alpha_n \le 1, \quad L_H \ell_\xi^{-1}(u_\xi + 1)\alpha_n \le \frac{1}{8}.$$

Then, for every $n \ge N_\xi^{(1)}$,

$$\begin{aligned}
\alpha_n |\langle \nabla M_\xi(e_n), H_{\theta_n}(Y_n)\rangle| &\le L_H \ell_\xi^{-1}(u_\xi + 1)\alpha_n \|e_n\|_{m,\xi}^2 + L_H \ell_\xi^{-1}(u_\xi + 1)\alpha_n \|e_n\|_{m,\xi} \\
&\le \frac{1}{8}\|e_n\|_{m,\xi}^2 + L_H \ell_\xi^{-1}(u_\xi + 1)\alpha_n \|e_n\|_{m,\xi} \\
&\le \frac{1}{8}\|e_n\|_{m,\xi}^2 + \frac{1}{8}\|e_n\|_{m,\xi}^2 + 2L_H^2 \ell_\xi^{-2}(u_\xi + 1)^2 \alpha_n^2 \\
&\qquad \text{(Cauchy–Schwarz inequality)} \\
&= \frac{1}{4}\|e_n\|_{m,\xi}^2 + 2L_H^2 \ell_\xi^{-2}(u_\xi + 1)^2 \alpha_n^2.
\end{aligned}$$

Set

$$K_\xi \doteq 2L_H^2 \ell_\xi^{-2}(u_\xi+1)^2.$$

Fix $K \geq K_\xi$ and set $N_{\xi,K} \doteq N_\xi^{(1)}$. For every $n \geq N_{\xi,K}$,

$$\begin{aligned}
V_n^\xi =& M_\xi(e_n) + \alpha_n \langle \nabla M_\xi(e_n), H_{\theta_n}(Y_n)\rangle + K\alpha_n^2 \\
\geq& M_\xi(e_n) - \alpha_n |\langle \nabla M_\xi(e_n), H_{\theta_n}(Y_n)\rangle| + K\alpha_n^2 \\
=& \frac{1}{2}\|e_n\|_{m,\xi}^2 - \alpha_n |\langle \nabla M_\xi(e_n), H_{\theta_n}(Y_n)\rangle| + K\alpha_n^2 \\
\geq& \frac{1}{2}\|e_n\|_{m,\xi}^2 - \frac{1}{4}\|e_n\|_{m,\xi}^2 - K_\xi\alpha_n^2 + K\alpha_n^2 \\
\geq& \frac{1}{4}\|e_n\|_{m,\xi}^2.
\end{aligned}$$

This proves (21).

It remains to prove (22). From (23), we have

$$\begin{aligned}
|\langle \nabla M_\xi(e_n), H_{\theta_n}(Y_n)\rangle| \leq& L_H \ell_\xi^{-1}(u_\xi+1)\big(\|e_n\|_{m,\xi}^2 + \|e_n\|_{m,\xi}\big) \\
\leq& 2L_H \ell_\xi^{-1}(u_\xi+1)\big(\|e_n\|_{m,\xi}^2 + 1\big).
\end{aligned}$$

For every $n \geq N_{\xi,K}$,

$$\begin{aligned}
\alpha_n\big|V_n^\xi - M_\xi(e_n)\big| =& \alpha_n\big|\alpha_n\langle \nabla M_\xi(e_n), H_{\theta_n}(Y_n)\rangle + K\alpha_n^2\big| \\
\leq& \alpha_n^2 |\langle \nabla M_\xi(e_n), H_{\theta_n}(Y_n)\rangle| + K\alpha_n^3 \\
\leq& 2L_H\ell_\xi^{-1}(u_\xi+1)\alpha_n^2\big(\|e_n\|_{m,\xi}^2+1\big) + K\alpha_n^3 \\
\leq& \big(2L_H\ell_\xi^{-1}(u_\xi+1)+K\big)\alpha_n^2\big(\|e_n\|_{m,\xi}^2+1\big)
\end{aligned}$$

Since $r_n \geq \alpha_n^2$, denote $C'_{\xi,K} \doteq 2L_H\ell_\xi^{-1}(u_\xi+1)+K$ then completes the proof. □

B Proofs of Section 3

B.1 Proof of Lemma 2

Proof. Since $f(\theta) = T(\theta) - \theta$ and $T(\theta_*) = \theta_*$,

$$\langle \nabla M_\xi(\theta-\theta_*), f(\theta)\rangle = \langle \nabla M_\xi(\theta-\theta_*), T(\theta)-\theta_*\rangle - \langle \nabla M_\xi(\theta-\theta_*), \theta-\theta_*\rangle.$$

The Moreau gradient inequality (5), the norm comparison (4), and the contraction property (1) give

$$\begin{aligned}
\langle \nabla M_\xi(\theta-\theta_*), T(\theta)-\theta_*\rangle \leq& \|\theta-\theta_*\|_{m,\xi}\|T(\theta)-\theta_*\|_{m,\xi} \\
\leq& \ell_\xi^{-1}\|\theta-\theta_*\|_{m,\xi}\|T(\theta)-\theta_*\| \\
\leq& \kappa\frac{u_\xi}{\ell_\xi}\|\theta-\theta_*\|_{m,\xi}^2.
\end{aligned}$$

Again by (5),

$$\langle \nabla M_\xi(\theta-\theta_*), \theta-\theta_*\rangle \geq \|\theta-\theta_*\|_{m,\xi}^2.$$

Therefore,

$$\begin{aligned}
\langle \nabla M_\xi(\theta-\theta_*), f(\theta)\rangle \leq& -\left(1-\kappa\frac{u_\xi}{\ell_\xi}\right)\|\theta-\theta_*\|_{m,\xi}^2 \\
=& -\mu_\xi M_\xi(\theta-\theta_*).
\end{aligned}$$

□

B.2 Proof of Lemma 3

Proof. Let $\bar{\alpha} = \sup_{n\geq 0} \alpha_n < \infty$. We first collect the bounds used in the proof.

$$\begin{aligned}
\|F(\theta_n, Y_n)\| \leq& \|F(\theta_n, Y_n) - F(\theta_*, Y_n)\| + \|F(\theta_*, Y_n)\| \\
\leq& L_F\big(\|e_n\| + 1\big) && \text{(By Assumption 2.3)} \\
\leq& L_F(u_\xi + 1)\big(\|e_n\|_{m,\xi} + 1\big), \\
\|F(\theta_n, Y_n)\|_2 \leq& \ell^{-1}\|F(\theta_n, Y_n)\| \leq L_F\ell^{-1}(u_\xi + 1)\big(\|e_n\|_{m,\xi} + 1\big), && (24) \\
\|F(\theta_n, Y_n)\|_{m,\xi} \leq& \ell_\xi^{-1}\|F(\theta_n, Y_n)\| \leq L_F\ell_\xi^{-1}(u_\xi + 1)\big(\|e_n\|_{m,\xi} + 1\big). && (25)
\end{aligned}$$

Similarly, by Assumption 2.4, for every $y \in \mathsf{Y}$,

$$\begin{aligned}
\|H_{\theta_n}(y)\| \leq& \|H_{\theta_n}(y) - H_{\theta_*}(y)\| + \|H_{\theta_*}(y)\| \\
\leq& L_H\big(\|e_n\| + 1\big) && \text{(By Assumption 2.4)} \\
\leq& L_H(u_\xi + 1)\big(\|e_n\|_{m,\xi} + 1\big), \\
\|H_{\theta_n}(y)\|_2 \leq& \ell^{-1}\|H_{\theta_n}(y)\| \\
\leq& L_H\ell^{-1}(u_\xi + 1)\big(\|e_n\|_{m,\xi} + 1\big). && (26)
\end{aligned}$$

Since $e_{n+1} = e_n + \alpha_n F(\theta_n, Y_n)$,

$$\begin{aligned}
\|e_{n+1}\|_{m,\xi} =& \|e_n + \alpha_n F(\theta_n, Y_n)\|_{m,\xi} \\
\leq& \|e_n\|_{m,\xi} + \alpha_n\|F(\theta_n, Y_n)\|_{m,\xi} \\
\leq& \|e_n\|_{m,\xi} + \alpha_n L_F\ell_\xi^{-1}(u_\xi + 1)\big(\|e_n\|_{m,\xi} + 1\big) && \text{(By (25))} \\
\leq& \Big(1 + \bar{\alpha}L_F\ell_\xi^{-1}(u_\xi + 1)\Big)\big(\|e_n\|_{m,\xi} + 1\big). && (27)
\end{aligned}$$

For every $x \in \mathbb{R}^d$,

$$\begin{aligned}
\|\nabla M_\xi(x)\|_2 =& \sup_{\|z\|_2\leq 1} |\langle \nabla M_\xi(x), z\rangle| \\
\leq& \sup_{\|z\|_2\leq 1} \|x\|_{m,\xi}\|z\|_{m,\xi} && \text{(By (5))} \\
\leq& u\ell_\xi^{-1}\|x\|_{m,\xi}. && (28)
\end{aligned}$$

Combining (28) with (27) gives

$$\|\nabla M_\xi(e_{n+1})\|_2 \leq u\ell_\xi^{-1}\Big(1 + \bar{\alpha}L_F\ell_\xi^{-1}(u_\xi + 1)\Big)\big(\|e_n\|_{m,\xi} + 1\big). \tag{29}$$

The smoothness of M_ξ and the Lipschitzness of H_θ also imply

$$\begin{aligned}
\|\nabla M_\xi(e_{n+1}) - \nabla M_\xi(e_n)\|_2 \leq& \frac{1}{\xi}\|e_{n+1} - e_n\|_2 && \text{(By (3))} \\
=& \frac{\alpha_n}{\xi}\|F(\theta_n, Y_n)\|_2 \\
\leq& \frac{\alpha_n L_F(u_\xi + 1)}{\xi\ell}\big(\|e_n\|_{m,\xi} + 1\big), && (30) \\
\big\|H_{\theta_{n+1}}(y) - H_{\theta_n}(y)\big\|_2 \leq& \ell^{-1}\big\|H_{\theta_{n+1}}(y) - H_{\theta_n}(y)\big\| \\
\leq& L_H\ell^{-1}\|\theta_{n+1} - \theta_n\| && \text{(By Assumption 2.4)} \\
=& \alpha_n L_H\ell^{-1}\|F(\theta_n, Y_n)\| \\
\leq& \alpha_n L_F L_H\ell^{-1}(u_\xi + 1)\big(\|e_n\|_{m,\xi} + 1\big). && (31)
\end{aligned}$$

We now use the smoothness of M_ξ. Since $e_{n+1} - e_n = \alpha_n F(\theta_n, Y_n)$,

$$M_\xi(e_{n+1}) \leq M_\xi(e_n) + \alpha_n\langle \nabla M_\xi(e_n), F(\theta_n, Y_n)\rangle + \frac{\alpha_n^2}{2\xi}\|F(\theta_n, Y_n)\|_2^2 \qquad \text{(By (3))}$$

$$
\begin{aligned}
&\le M_\xi(e_n) + \alpha_n \langle \nabla M_\xi(e_n), F(\theta_n, Y_n)\rangle \\
&\quad + \frac{L_F^2 (u_\xi+1)^2}{2\xi\ell^2}\alpha_n^2 \big(\|e_n\|_{m,\xi}+1\big)^2 && \text{(By (24))} \\
&\le M_\xi(e_n) + \alpha_n \langle \nabla M_\xi(e_n), F(\theta_n, Y_n)\rangle \\
&\quad + \frac{L_F^2 (u_\xi+1)^2}{\xi\ell^2}\alpha_n^2 \big(\|e_n\|_{m,\xi}^2+1\big). && (32)
\end{aligned}
$$

The Poisson equation gives

$$
\begin{aligned}
\langle \nabla M_\xi(e_n), F(\theta_n, Y_n)\rangle =& \langle \nabla M_\xi(e_n), f(\theta_n)\rangle + \langle \nabla M_\xi(e_n), H_{\theta_n}(Y_n)\rangle \\
&- \langle \nabla M_\xi(e_n), (PH_{\theta_n})(Y_n)\rangle. && \text{(By (2))}
\end{aligned}
$$

Since $\nabla M_\xi(e_n)$, θ_n, and Y_n are $\mathcal{F}_n$-measurable,

$$
\langle \nabla M_\xi(e_n), (PH_{\theta_n})(Y_n)\rangle = \mathbb{E}_n[\langle \nabla M_\xi(e_n), H_{\theta_n}(Y_{n+1})\rangle].
$$

Combining with (32) and taking conditional expectation, we get

$$
\begin{aligned}
\mathbb{E}_n[M_\xi(e_{n+1})] \le& M_\xi(e_n) + \alpha_n \langle \nabla M_\xi(e_n), f(\theta_n)\rangle \\
&+ \alpha_n \mathbb{E}_n[\langle \nabla M_\xi(e_n), H_{\theta_n}(Y_n)\rangle - \langle \nabla M_\xi(e_n), H_{\theta_n}(Y_{n+1})\rangle] \\
&+ \frac{L_F^2 (u_\xi+1)^2}{\xi\ell^2}\alpha_n^2 \big(\|e_n\|_{m,\xi}^2+1\big). && (33)
\end{aligned}
$$

It remains to replace the one-step Poisson scalar by the genuine next-time scalar. By (26), (29), (30), and (31),

$$
\begin{aligned}
&\big|\langle \nabla M_\xi(e_{n+1}), H_{\theta_{n+1}}(Y_{n+1})\rangle - \langle \nabla M_\xi(e_n), H_{\theta_n}(Y_{n+1})\rangle\big| \\
\le& \|\nabla M_\xi(e_{n+1}) - \nabla M_\xi(e_n)\|_2 \|H_{\theta_n}(Y_{n+1})\|_2 \\
&+ \|\nabla M_\xi(e_{n+1})\|_2 \big\|H_{\theta_{n+1}}(Y_{n+1}) - H_{\theta_n}(Y_{n+1})\big\|_2 && \text{(Cauchy–Schwarz inequality)} \\
\le& \frac{\alpha_n L_F L_H (u_\xi+1)^2}{\xi\ell^2}\big(\|e_n\|_{m,\xi}+1\big)^2 \\
&+ \alpha_n L_F L_H \frac{u(u_\xi+1)}{\ell_\xi \ell}\Big(1+\bar\alpha L_F \ell_\xi^{-1}(u_\xi+1)\Big)\big(\|e_n\|_{m,\xi}+1\big)^2 \\
\le& 2\alpha_n L_F L_H \left[\frac{(u_\xi+1)^2}{\xi\ell^2} + \frac{u(u_\xi+1)}{\ell_\xi\ell}\Big(1+\bar\alpha L_F\ell_\xi^{-1}(u_\xi+1)\Big)\right]\big(\|e_n\|_{m,\xi}^2+1\big).
\end{aligned}
$$

Therefore,

$$
\begin{aligned}
&\alpha_n \mathbb{E}_n[\langle \nabla M_\xi(e_n), H_{\theta_n}(Y_n)\rangle - \langle \nabla M_\xi(e_n), H_{\theta_n}(Y_{n+1})\rangle] \\
=& \alpha_n \mathbb{E}_n\big[\langle \nabla M_\xi(e_n), H_{\theta_n}(Y_n)\rangle - \langle \nabla M_\xi(e_{n+1}), H_{\theta_{n+1}}(Y_{n+1})\rangle\big] \\
&+ \alpha_n \mathbb{E}_n\big[\langle \nabla M_\xi(e_{n+1}), H_{\theta_{n+1}}(Y_{n+1})\rangle - \langle \nabla M_\xi(e_n), H_{\theta_n}(Y_{n+1})\rangle\big] \\
\le& \alpha_n \mathbb{E}_n\big[\langle \nabla M_\xi(e_n), H_{\theta_n}(Y_n)\rangle - \langle \nabla M_\xi(e_{n+1}), H_{\theta_{n+1}}(Y_{n+1})\rangle\big] \\
&+ 2\alpha_n^2 L_F L_H \left[\frac{(u_\xi+1)^2}{\xi\ell^2} + \frac{u(u_\xi+1)}{\ell_\xi\ell}\Big(1+\bar\alpha L_F\ell_\xi^{-1}(u_\xi+1)\Big)\right]\big(\|e_n\|_{m,\xi}^2+1\big).
\end{aligned}
$$

Substituting into (33) and denote

$$
C_\xi \doteq \frac{L_F^2 (u_\xi+1)^2}{\xi\ell^2} + 2L_F L_H \left[\frac{(u_\xi+1)^2}{\xi\ell^2} + \frac{u(u_\xi+1)}{\ell_\xi\ell}\Big(1+\bar\alpha L_F\ell_\xi^{-1}(u_\xi+1)\Big)\right],
$$

we then obtain

$$
\begin{aligned}
\mathbb{E}_n[M_\xi(e_{n+1})] \le& M_\xi(e_n) + \alpha_n \langle \nabla M_\xi(e_n), f(\theta_n)\rangle \\
&+ \alpha_n \mathbb{E}_n\big[\langle \nabla M_\xi(e_n), H_{\theta_n}(Y_n)\rangle - \langle \nabla M_\xi(e_{n+1}), H_{\theta_{n+1}}(Y_{n+1})\rangle\big] \\
&+ C_\xi \alpha_n^2 \big(\|e_n\|_{m,\xi}^2+1\big),
\end{aligned}
$$

which proves the claim. □

B.3 Proof of Lemma 4

Proof. Choose K_ξ as in Lemma 8, and fix $K \geq K_\xi$. Let $C'_{\xi,K}$ be the constant in (22). Choose a deterministic tail $N_{\xi,K}$ such that, for every $n \geq N_{\xi,K}$, (21), (22), and $\alpha_n \leq 1$ all hold.

The lower bound (8) follows directly from (21). In particular, on this deterministic tail,

$$\|e_n\|_{m,\xi}^2 \leq 4V_n^\xi, \quad V_n^\xi \geq 0. \tag{34}$$

It remains to prove (9). Recall we have $\|x\|_{m,\xi} \leq \ell_\xi^{-1}\|x\|$ and $\|x\| \leq u_\xi \|x\|_{m,\xi}$. For every $n \geq N_{\xi,K}$, we then derive

$$\begin{aligned}
\|F(\theta_n, Y_n)\|_{m,\xi} \leq& \ell_\xi^{-1}\|F(\theta_n, Y_n)\| \\
\leq& \ell_\xi^{-1}\Big(\|F(\theta_n, Y_n) - F(\theta_*, Y_n)\| + \|F(\theta_*, Y_n)\|\Big) \\
\leq& \ell_\xi^{-1} L_F\big(\|e_n\| + 1\big) && \text{(By Assumption 2.3)} \\
\leq& L_F \ell_\xi^{-1}(u_\xi + 1)\big(\|e_n\|_{m,\xi} + 1\big).
\end{aligned}$$

Since $e_{n+1} = e_n + \alpha_n F(\theta_n, Y_n)$ and $\alpha_n \leq 1$ on the tail,

$$\begin{aligned}
\|e_{n+1}\|_{m,\xi} \leq& \|e_n\|_{m,\xi} + \alpha_n \|F(\theta_n, Y_n)\|_{m,\xi} \\
\leq& \Big(1 + L_F \ell_\xi^{-1}(u_\xi + 1)\Big)\big(\|e_n\|_{m,\xi} + 1\big).
\end{aligned} \tag{35}$$

Similarly, we have

$$\begin{aligned}
\big\|H_{\theta_{n+1}}(Y_{n+1})\big\|_{m,\xi} \leq& \ell_\xi^{-1}\big\|H_{\theta_{n+1}}(Y_{n+1})\big\| \\
\leq& \ell_\xi^{-1} L_H\big(\|e_{n+1}\| + 1\big) && \text{(By Assumption 2.4)} \\
\leq& L_H \ell_\xi^{-1}(u_\xi + 1)\big(\|e_{n+1}\|_{m,\xi} + 1\big) \\
\leq& L_H \ell_\xi^{-1}(u_\xi + 1)\Big(2 + L_F \ell_\xi^{-1}(u_\xi + 1)\Big)\big(\|e_n\|_{m,\xi} + 1\big).
\end{aligned} \tag{36}$$

Combining (35) and (36), we then get

$$\begin{aligned}
\big|\langle \nabla M_\xi(e_{n+1}), H_{\theta_{n+1}}(Y_{n+1})\rangle\big| \leq& \|e_{n+1}\|_{m,\xi}\big\|H_{\theta_{n+1}}(Y_{n+1})\big\|_{m,\xi} && \text{(By (5))} \\
\leq& C'_\xi\big(\|e_n\|_{m,\xi}^2 + 1\big),
\end{aligned} \tag{37}$$

where $C'_\xi \doteq 2L_H \ell_\xi^{-1}(u_\xi + 1)\Big(1 + L_F \ell_\xi^{-1}(u_\xi + 1)\Big)\Big(2 + L_F \ell_\xi^{-1}(u_\xi + 1)\Big)$.

Now apply Lemma 3. Since $\langle \nabla M_\xi(e_n), H_{\theta_n}(Y_n)\rangle$ is $\mathcal{F}_n$-measurable,

$$\begin{aligned}
\mathbb{E}_n\big[V_{n+1}^\xi\big] =& \mathbb{E}_n\big[M_\xi(e_{n+1})\big] + \alpha_{n+1}\mathbb{E}_n\big[\langle \nabla M_\xi(e_{n+1}), H_{\theta_{n+1}}(Y_{n+1})\rangle\big] + K\alpha_{n+1}^2 \\
\leq& M_\xi(e_n) + \alpha_n\langle \nabla M_\xi(e_n), f(\theta_n)\rangle + \alpha_n\langle \nabla M_\xi(e_n), H_{\theta_n}(Y_n)\rangle \\
& - \alpha_n\mathbb{E}_n\big[\langle \nabla M_\xi(e_{n+1}), H_{\theta_{n+1}}(Y_{n+1})\rangle\big] + C_\xi\alpha_n^2\big(\|e_n\|_{m,\xi}^2 + 1\big) \\
& + \alpha_{n+1}\mathbb{E}_n\big[\langle \nabla M_\xi(e_{n+1}), H_{\theta_{n+1}}(Y_{n+1})\rangle\big] + K\alpha_{n+1}^2 && \text{(By Lemma 3)} \\
=& V_n^\xi + \alpha_n\langle \nabla M_\xi(e_n), f(\theta_n)\rangle + (\alpha_{n+1} - \alpha_n)\mathbb{E}_n\big[\langle \nabla M_\xi(e_{n+1}), H_{\theta_{n+1}}(Y_{n+1})\rangle\big] \\
& + C_\xi\alpha_n^2\big(\|e_n\|_{m,\xi}^2 + 1\big) + K\big(\alpha_{n+1}^2 - \alpha_n^2\big) && \text{(By (7))} \\
\leq& V_n^\xi + \alpha_n\langle \nabla M_\xi(e_n), f(\theta_n)\rangle + |\alpha_{n+1} - \alpha_n|\mathbb{E}_n\big[\big|\langle \nabla M_\xi(e_{n+1}), H_{\theta_{n+1}}(Y_{n+1})\rangle\big|\big] \\
& + C_\xi\alpha_n^2\big(\|e_n\|_{m,\xi}^2 + 1\big),
\end{aligned} \tag{38}$$

where the last step uses that α_n is nonincreasing and $K \geq 0$. Taking conditional expectation in (37) and substituting into (38) yields

$$\mathbb{E}_n\big[V_{n+1}^\xi\big] \leq V_n^\xi + \alpha_n\langle \nabla M_\xi(e_n), f(\theta_n)\rangle + \Big(C_\xi + C'_\xi\Big) r_n\big(\|e_n\|_{m,\xi}^2 + 1\big), \tag{39}$$

where $r_n = \alpha_n^2 + |\alpha_{n+1} - \alpha_n|$.

By Lemma 2,

$$\langle \nabla M_\xi(e_n), f(\theta_n)\rangle \leq -\mu_\xi M_\xi(e_n). \tag{40}$$

Combining (39) and (40), we obtain

$$\mathbb{E}_n\big[V_{n+1}^\xi\big] \leq V_n^\xi - \mu_\xi\alpha_n M_\xi(e_n) + \Big(C_\xi + C'_\xi\Big) r_n\big(\|e_n\|_{m,\xi}^2 + 1\big). \tag{41}$$

The comparison bound (22) gives

$$\begin{aligned} -\mu_\xi\alpha_n M_\xi(e_n) &= -\mu_\xi\alpha_n V_n^\xi + \mu_\xi\alpha_n\big(V_n^\xi - M_\xi(e_n)\big) \\ &\leq -\mu_\xi\alpha_n V_n^\xi + \mu_\xi C'_{\xi,K} r_n\big(\|e_n\|_{m,\xi}^2 + 1\big). \end{aligned}$$

Substituting into (41) then gives

$$\begin{aligned} \mathbb{E}_n\big[V_{n+1}^\xi\big] &\leq \big(1-\mu_\xi\alpha_n\big)V_n^\xi + (C_\xi + C'_\xi + \mu_\xi C'_{\xi,K}) r_n\big(\|e_n\|_{m,\xi}^2 + 1\big) \\ &\leq \big(1-\mu_\xi\alpha_n\big)V_n^\xi + (C_\xi + C'_\xi + \mu_\xi C'_{\xi,K}) r_n\big(4V_n^\xi + 1\big). \qquad \text{(By (34))} \end{aligned}$$

Let $C_{\xi,K} \doteq 4(C_\xi + C'_\xi + \mu_\xi C'_{\xi,K})$ then completes the proof. □

B.4 Proof of Theorem 1

Proof. Fix ζ in the range claimed by the theorem. If $\frac{1}{2} < \eta < 1$, choose any $\xi > 0$ satisfying (6). If $\eta = 1$, then $\zeta < \min\{1, 2(1-\kappa)\alpha\}$. Since

$$\mu_\xi = 2\left(1 - \kappa\frac{u_\xi}{\ell_\xi}\right) \longrightarrow 2(1-\kappa) \quad \text{as } \xi \downarrow 0,$$

we may choose $\xi > 0$ small enough such that

$$\kappa\frac{u_\xi}{\ell_\xi} < 1, \quad \zeta < \min\{1, \mu_\xi\alpha\}. \tag{42}$$

Let K_ξ be the constant from Lemma 4, and fix $K \geq K_\xi$. Then there exist constants $C_{\xi,K} < \infty$ and $N_{\xi,K} < \infty$ such that, for every $n \geq N_{\xi,K}$,

$$\begin{aligned} V_n^\xi &\geq \frac{1}{4}\|e_n\|_{m,\xi}^2 \geq 0, \\ \mathbb{E}_n\big[V_{n+1}^\xi\big] &\leq \big(1 - \mu_\xi\alpha_n + C_{\xi,K} r_n\big)V_n^\xi + C_{\xi,K} r_n. \end{aligned}$$

Thus Lemma 7 applies on this deterministic tail with $X_n = V_n^\xi$, $c_0 = \mu_\xi$, and $C_0 = C_{\xi,K}$. Hence

$$(n+1)^\zeta V_n^\xi \longrightarrow 0 \quad \text{a.s.}$$

Indeed, when $\frac{1}{2} < \eta < 1$, this follows from $\zeta < 2\eta - 1$. When $\eta = 1$, this follows from (42).

It remains to translate the rate for V_n^ξ into the rate for the original error. On the same deterministic tail,

$$(n+1)^\zeta\|\theta_n - \theta_*\|^2 = (n+1)^\zeta\|e_n\|^2 \leq u_\xi^2(n+1)^\zeta\|e_n\|_{m,\xi}^2 \leq 4u_\xi^2(n+1)^\zeta V_n^\xi.$$

Thus $(n+1)^\zeta\|\theta_n - \theta_*\|^2 \to 0$ almost surely. Since changing finitely many initial terms does not affect the almost sure limit, the claimed convergence holds for the full sequence. □

B.5 Proof of Corollary 1

Proof. The proof of Lemma 4 does not use square summability of the learning rates. Thus, for any fixed $0 < \eta \leq 1$, the same argument gives the following deterministic-tail bounds. Fix $K \geq K_\xi$. There exist constants $C_{\xi,K} < \infty$ and a deterministic index N_1 such that, for every $n \geq N_1$,

$$V_n^\xi \geq \frac{1}{4}\|e_n\|_{m,\xi}^2 \geq 0, \tag{43}$$

$$\mathbb{E}_n\left[V_{n+1}^{\xi}\right] \leq \left(1-\mu_\xi\alpha_n + C_{\xi,K}r_n\right)V_n^{\xi} + C_{\xi,K}r_n. \tag{44}$$

We first handle the finite initial segment before this deterministic tail. For every $m \geq 0$,

$$\begin{aligned}\|e_{m+1}\| + 1 \leq& \|e_m\| + 1 + \alpha_m\|F(\theta_m, Y_m)\| \\ \leq& \|e_m\| + 1 + L_F\alpha_m\left(\|e_m\| + 1\right) && \text{(By Assumption 2.3)} \\ =& \left(1 + L_F\alpha_m\right)\left(\|e_m\| + 1\right).\end{aligned}$$

Iterating to any fixed time m gives

$$\|e_m\| + 1 \leq \left(\|e_0\| + 1\right)\prod_{i=0}^{m-1}\left(1 + L_F\alpha_i\right) < \infty. \tag{45}$$

Hence, by (5), Assumption 2.4, and norm comparison,

$$\begin{aligned}\left|V_m^{\xi}\right| \leq& \frac{1}{2}\|e_m\|_{m,\xi}^2 + \alpha_m\|e_m\|_{m,\xi}\|H_{\theta_m}(Y_m)\|_{m,\xi} + K\alpha_m^2 \\ \leq& \frac{1}{2}\ell_\xi^{-2}\|e_m\|^2 + L_H\ell_\xi^{-2}\alpha_m\|e_m\|\left(\|e_m\| + 1\right) + K\alpha_m^2 < \infty.\end{aligned}$$

Consequently, for every deterministic index m, $\mathbb{E}\left[\left|V_m^{\xi}\right|\right] < \infty$.

For $0 < \eta \leq 1$, the mean-value theorem gives

$$\begin{aligned}|\alpha_{n+1} - \alpha_n| =& \alpha\Big((n+1)^{-\eta} - (n+2)^{-\eta}\Big) \\ \leq& \alpha\eta(n+1)^{-\eta-1} \\ \leq& \frac{\eta}{\alpha}\alpha_n^2,\end{aligned}$$

where the last step uses $\eta \leq 1$. Therefore, $r_n \leq C_r\alpha_n^2$, where $C_r \doteq 1 + \frac{\eta}{\alpha}$. Taking total expectation in (44) then gives, for every $n \geq N_1$,

$$\mathbb{E}\left[V_{n+1}^{\xi}\right] \leq \left(1 - \mu_\xi\alpha_n + C_{\xi,K}C_r\alpha_n^2\right)\mathbb{E}\left[V_n^{\xi}\right] + C_{\xi,K}C_r\alpha_n^2. \tag{46}$$

We first consider $0 < \eta < 1$. Since $\alpha_n = \alpha(n+1)^{-\eta}$, choose a deterministic index N_2 such that

$$C_{\xi,K}C_r\alpha_n^2 \leq \frac{\mu_\xi\alpha}{2}(n+1)^{-\eta}$$

for every $n \geq N_2$. Hence (46) implies

$$\mathbb{E}\left[V_{n+1}^{\xi}\right] \leq \left(1 - \frac{\mu_\xi\alpha}{2(n+1)^{\eta}}\right)\mathbb{E}\left[V_n^{\xi}\right] + \frac{C_{\xi,K}C_r\alpha^2}{(n+1)^{2\eta}} \tag{47}$$

for every $n \geq \max\{N_1, N_2\}$. Choose a deterministic index N_3 such that, for every $n \geq N_3$,

$$(n+2)^{-\eta} \geq (n+1)^{-\eta} - \frac{\mu_\xi\alpha}{4}(n+1)^{-2\eta}. \tag{48}$$

This is possible because $\eta < 1$. Set

$$N \doteq \max\{N_1, N_2, N_3\}.$$

Choose $B < \infty$ large enough such that

$$\mathbb{E}\left[V_N^{\xi}\right] \leq B(N+1)^{-\eta}, \quad \frac{B\mu_\xi\alpha}{4} \geq C_{\xi,K}C_r\alpha^2.$$

Then induction gives, for every $n \geq N$,

$$\mathbb{E}\left[V_n^{\xi}\right] \leq B(n+1)^{-\eta}. \tag{49}$$

Indeed, if the bound holds at time n, then

$$\mathbb{E}\left[V_{n+1}^{\xi}\right] \leq B(n+1)^{-\eta} - \frac{B\mu_\xi\alpha}{2}(n+1)^{-2\eta} + C_{\xi,K}C_r\alpha^2(n+1)^{-2\eta} \qquad \text{(By (47))}$$

$$\leq B(n+1)^{-\eta} - \frac{B\mu_\xi\alpha}{4}(n+1)^{-2\eta}$$
$$\leq B(n+2)^{-\eta}. \qquad \text{(By (48))}$$

Thus $\mathbb{E}\left[V_n^\xi\right] = \mathcal{O}(n^{-\eta})$ when $0 < \eta < 1$.

Now consider $\eta = 1$. In this case (46) gives, for every $n \geq N_1$,

$$\mathbb{E}\left[V_{n+1}^\xi\right] \leq \left(1 - \frac{\mu_\xi\alpha}{n+1} + \frac{C_{\xi,K}C_r\alpha^2}{(n+1)^2}\right)\mathbb{E}\left[V_n^\xi\right] + \frac{C_{\xi,K}C_r\alpha^2}{(n+1)^2}. \tag{50}$$

Choose a deterministic index N_4 such that, for every $j \geq N_4$,

$$1 - \frac{\mu_\xi\alpha}{j+1} + \frac{C_{\xi,K}C_r\alpha^2}{(j+1)^2} \geq 0.$$

Set

$$N \doteq \max\{N_1, N_4\}.$$

For $n > i \geq N-1$,

$$\begin{aligned}
&\prod_{j=i+1}^{n-1}\left(1 - \frac{\mu_\xi\alpha}{j+1} + \frac{C_{\xi,K}C_r\alpha^2}{(j+1)^2}\right) \\
&\leq \exp\left(-\mu_\xi\alpha\sum_{j=i+1}^{n-1}\frac{1}{j+1} + C_{\xi,K}C_r\alpha^2\sum_{j=i+1}^{n-1}\frac{1}{(j+1)^2}\right) \\
&\leq \exp\left(C_{\xi,K}C_r\alpha^2\sum_{j=0}^{\infty}\frac{1}{(j+1)^2}\right)\left(\frac{i+2}{n+1}\right)^{\mu_\xi\alpha} \\
&\leq \exp\left(2C_{\xi,K}C_r\alpha^2\right)\left(\frac{i+2}{n+1}\right)^{\mu_\xi\alpha}.
\end{aligned}$$

Iterating (50) from the deterministic time N yields

$$\begin{aligned}
\mathbb{E}\left[V_n^\xi\right] \leq& \mathbb{E}\left[V_N^\xi\right]\prod_{j=N}^{n-1}\left(1 - \frac{\mu_\xi\alpha}{j+1} + \frac{C_{\xi,K}C_r\alpha^2}{(j+1)^2}\right) \\
&+ C_{\xi,K}C_r\alpha^2\sum_{i=N}^{n-1}\frac{1}{(i+1)^2}\prod_{j=i+1}^{n-1}\left(1 - \frac{\mu_\xi\alpha}{j+1} + \frac{C_{\xi,K}C_r\alpha^2}{(j+1)^2}\right) \\
\leq& \exp\left(2C_{\xi,K}C_r\alpha^2\right)\mathbb{E}\left[V_N^\xi\right]\left(\frac{N+1}{n+1}\right)^{\mu_\xi\alpha} \\
&+ C_{\xi,K}C_r\alpha^2\exp\left(2C_{\xi,K}C_r\alpha^2\right)\sum_{i=N}^{n-1}\frac{1}{(i+1)^2}\left(\frac{i+2}{n+1}\right)^{\mu_\xi\alpha}.
\end{aligned} \tag{51}$$

The sum in (51) is $\mathcal{O}(n^{-\mu_\xi\alpha})$ when $\mu_\xi\alpha < 1$, $\mathcal{O}(\log n/n)$ when $\mu_\xi\alpha = 1$, and $\mathcal{O}(n^{-1})$ when $\mu_\xi\alpha > 1$. Therefore,

$$\mathbb{E}\left[V_n^\xi\right] = \begin{cases}\mathcal{O}(n^{-\min\{\mu_\xi\alpha,1\}}), & \text{if } \mu_\xi\alpha \neq 1, \\ \mathcal{O}(\log n/n), & \text{if } \mu_\xi\alpha = 1.\end{cases} \tag{52}$$

Finally, for every $n \geq N$, (43) and (4) give

$$\|\theta_n - \theta_*\|^2 = \|e_n\|^2 \leq u_\xi^2\|e_n\|_{m,\xi}^2 \leq 4u_\xi^2 V_n^\xi.$$

Taking expectations and using (49) and (52) gives the stated mean-square rates on the deterministic tail. The finitely many indices $n < N$ have finite second moments by (45), so they do not affect the asymptotic $\mathcal{O}(\cdot)$ rates. □

C Proof of Proposition 1

Proof. First, Assumption 2.1 holds. The process $Y_n = (S_n, A_n, \dots, S_{n+N}, A_{n+N})$ is a finite-state Markov chain induced by the behavior state-action chain. Since the latter is irreducible and aperiodic, the window chain has a unique invariant distribution ϖ on its recurrent class. All quantities in (12)–(13) are bounded on this finite state space, so $F_{\mathrm{GTD}}(w, \cdot)$ is ϖ-integrable for every w.

Second, Assumption 2.2 holds by Condition 4.1 and the construction of T_{GTD}. Indeed,

$$f_{\mathrm{GTD}}(w) \doteq \mathbb{E}_{\varpi}[F_{\mathrm{GTD}}(w, Y)] = \beta_{c,\rho}\big(\bar{A}_{c,\rho} w - \bar{b}_{c,\rho}\big),$$

and hence $w + f_{\mathrm{GTD}}(w) = T_{\mathrm{GTD}}(w)$, which is a κ_{GTD}-contraction in $\|\cdot\|_{\mathrm{GTD}}$ by (14).

Third, Assumption 2.3 holds. Because N is fixed, rewards and features are bounded, and c, ρ are bounded, the matrices $A_{c,\rho}(y)$ and vectors $b_{c,\rho}(y)$ are uniformly bounded over all windows y. Therefore F_{GTD} is affine in w with uniformly bounded coefficients. In particular, after changing constants to account for norm equivalence in finite dimension, there is $L < \infty$ such that for all $w, w' \in \mathbb{R}^d$ and all windows y,

$$\|F_{\mathrm{GTD}}(w, y) - F_{\mathrm{GTD}}(w', y)\|_{\mathrm{GTD}} \le L\|w - w'\|_{\mathrm{GTD}},$$
$$\big\|F_{\mathrm{GTD}}(w^*_{c,\rho}, y)\big\|_{\mathrm{GTD}} \le L.$$

Finally, Assumption 2.4 holds by the finite-state Poisson-equation argument. Let P_Y be the transition matrix of the window chain. For each fixed w, the centered function $F_{\mathrm{GTD}}(w, \cdot) - f_{\mathrm{GTD}}(w)$ has mean zero under ϖ, and the Poisson equation admits the solution

$$H_w(y) \doteq \sum_{t=0}^{\infty} P_Y^t \big(F_{\mathrm{GTD}}(w, \cdot) - f_{\mathrm{GTD}}(w)\big)(y), \tag{53}$$

where the series converges componentwise on the recurrent class by finite-state geometric ergodicity. Since F_{GTD} is affine in w with uniformly bounded coefficients, and the linear operator in (53) is bounded on centered functions over the finite state space, the Poisson solution inherits the required Lipschitz and growth bounds:

$$\|H_w(y) - H_{w'}(y)\|_{\mathrm{GTD}} \le L_H\|w - w'\|_{\mathrm{GTD}}, \quad \Big\|H_{w^*_{c,\rho}}(y)\Big\|_{\mathrm{GTD}} \le L_H$$

for a finite constant L_H. This verifies Assumption 2.4. The identification $\alpha_n = a_n/\beta_{c,\rho}$ in (15) gives the learning-rate form required by Assumption 2.5, with stepsize constant $a/\beta_{c,\rho}$. □

D Numerical Experiment

As Lauand and Meyn [2024] provide numerical experiments for linear Markovian SA with power-law stepsizes $\alpha_n \propto n^{-\eta}$, $0 < \eta < 1$, we focus here on the harmonic case $\eta = 1$.

We run linear TD(0) on a 100-state nonreversible Markov reward process with discount $\gamma = 0.7$. Let $\Pi_1, \dots, \Pi_4$ be independent uniformly sampled permutation matrices (environment seed 123), and set

$$P = 0.05J + 0.95(0.1I + 0.9Q), \qquad Q = \tfrac{1}{4}\sum_{j=1}^{4} \Pi_j, \qquad J = \tfrac{1}{100}\mathbf{1}\mathbf{1}^\top.$$

Then $D = I/100$ is stationary and $\left\|P - P^\top\right\|_F = 6.02$. The seven features are the constant feature and the sine/cosine Fourier modes of frequencies $k = 1, 2, 3$, with each nonconstant feature multiplied by $\sqrt{2}$, so $\Phi^\top D \Phi = I$ up to machine precision. For $s = 0, \dots, 99$, rewards are

$$r(s) = 0.4 + 0.35 \sin u_s + 0.2 \cos(2u_s) + 0.05 \sin(3u_s), \qquad u_s = 2\pi s/100.$$

With $A = \Phi^\top D(\gamma P - I)\Phi$, we compute the exact TD root $w^\star = -A^{-1}\Phi^\top D r$ and solve $A^\top G + GA = -I$. The contraction scaling $\beta = 0.1$ gives $\kappa = 0.97$.

For each $a \in \{1/3, 2/3, 1, 4/3, 1.6\}$, we run 30 independent single trajectories from uniform initial states and $w_0 = 0$, using stepsizes $a/(n+1)$ at updates $n = 1, \dots, 10^6$. NumPy's default generator uses seeds 2000–2004 for these five cases, respectively. We record 400 logarithmically spaced points in $[10^2, 10^6]$ and measure

$$E_n = (w_n - w^\star)^\top G (w_n - w^\star), \qquad \zeta_{\max} = \min\left\{1, \frac{2(1-\kappa)a}{\beta}\right\}.$$

Theorem 1 permits every $\zeta < \zeta_{\max}$. Each fitted exponent $\widehat{\zeta}$ is the negative least-squares slope of $\log E_n$ against $\log n$ at the recorded points in $[10^5, 10^6]$. Figure 1 shows the error trajectories and the fitted exponents, whose medians track the admissible endpoint as a varies.

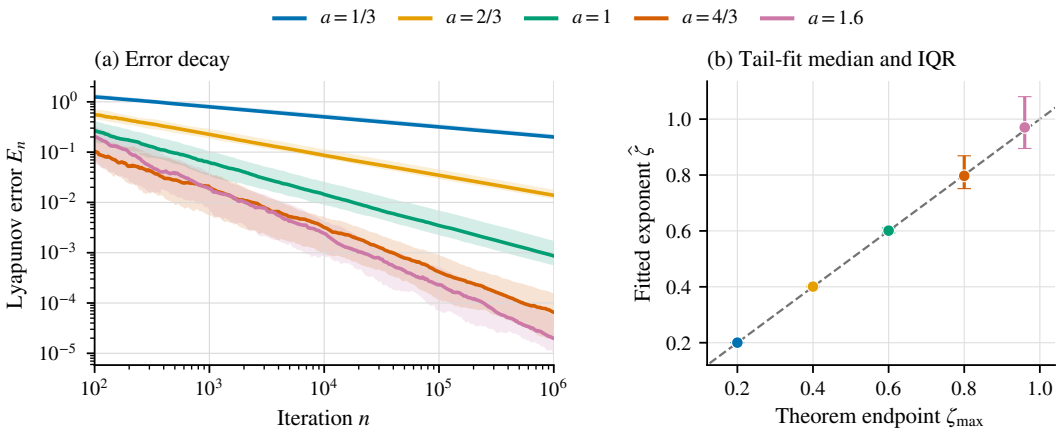


Figure 1: Harmonic-learning-rate experiment, with 30 paths per stepsize. (a) Solid curves and shading show pointwise medians and interquartile ranges of E_n. (b) Points and error bars show medians and interquartile ranges of the individually fitted exponents over $[10^5, 10^6]$; the dashed line is the identity reference.

NeurIPS Paper Checklist

1. **Claims**

   Question: Do the main claims made in the abstract and introduction accurately reflect the paper's contributions and scope?

   Answer: [Yes]

   Justification: The abstract and introduction state that the paper studies almost sure convergence rates for stochastic approximation with Markovian noise under contractive expected updates, including possibly non-Euclidean contraction norms. The main technical and rate claims are formalized in Theorem 1 under Assumptions 2.1–2.5, and the reinforcement-learning application claims are formalized in Proposition 1 and Theorem 2. The application section also clarifies that the generalized TD recursion and its Hurwitz structural conditions are not new contributions; the contribution is to combine these structural properties with the proposed Poisson–Moreau drift to obtain pathwise almost sure rates under single-trajectory Markovian sampling.

   Guidelines:

   - The answer [N/A] means that the abstract and introduction do not include the claims made in the paper.
   - The abstract and/or introduction should clearly state the claims made, including the contributions made in the paper and important assumptions and limitations. A [No] or [N/A] answer to this question will not be perceived well by the reviewers.
   - The claims made should match theoretical and experimental results, and reflect how much the results can be expected to generalize to other settings.
   - It is fine to include aspirational goals as motivation as long as it is clear that these goals are not attained by the paper.

2. **Limitations**

   Question: Does the paper discuss the limitations of the work performed by the authors?

   Answer: [Yes]

   Justification: Sections 2 and 4 state the assumptions limiting the scope of the results. The conclusion identifies extending the almost-sure learning-rate range, attaining sharp LIL-order rates, and deriving concentration bounds as open problems.

   Guidelines:

   - The answer [N/A] means that the paper has no limitation while the answer [No] means that the paper has limitations, but those are not discussed in the paper.
   - The authors are encouraged to create a separate "Limitations" section in their paper.
   - The paper should point out any strong assumptions and how robust the results are to violations of these assumptions (e.g., independence assumptions, noiseless settings, model well-specification, asymptotic approximations only holding locally). The authors should reflect on how these assumptions might be violated in practice and what the implications would be.
   - The authors should reflect on the scope of the claims made, e.g., if the approach was only tested on a few datasets or with a few runs. In general, empirical results often depend on implicit assumptions, which should be articulated.
   - The authors should reflect on the factors that influence the performance of the approach. For example, a facial recognition algorithm may perform poorly when image resolution is low or images are taken in low lighting. Or a speech-to-text system might not be used reliably to provide closed captions for online lectures because it fails to handle technical jargon.
   - The authors should discuss the computational efficiency of the proposed algorithms and how they scale with dataset size.
   - If applicable, the authors should discuss possible limitations of their approach to address problems of privacy and fairness.

- While the authors might fear that complete honesty about limitations might be used by reviewers as grounds for rejection, a worse outcome might be that reviewers discover limitations that aren't acknowledged in the paper. The authors should use their best judgment and recognize that individual actions in favor of transparency play an important role in developing norms that preserve the integrity of the community. Reviewers will be specifically instructed to not penalize honesty concerning limitations.

3. **Theory assumptions and proofs**

   Question: For each theoretical result, does the paper provide the full set of assumptions and a complete (and correct) proof?

   Answer: [Yes]

   Justification: The paper explicitly states the assumptions used for the general stochastic approximation result in Assumptions 2.1–2.5, and Theorem 1 is proved in Appendix B.4 using the supporting lemmas developed in Section 3 and Appendix B. The generalized TD application states its algorithm-specific assumptions in Section 4, verifies the abstract assumptions in Proposition 1 with proof in Appendix C, and then derives Theorem 2 by applying Theorem 1. Auxiliary results used in the proof, including the Moreau-envelope facts, the Poisson–Moreau drift recursion, and the almost-supermartingale rate lemma, are stated and proved or properly referenced in the appendix.

   Guidelines:

   - The answer [N/A] means that the paper does not include theoretical results.
   - All the theorems, formulas, and proofs in the paper should be numbered and cross-referenced.
   - All assumptions should be clearly stated or referenced in the statement of any theorems.
   - The proofs can either appear in the main paper or the supplemental material, but if they appear in the supplemental material, the authors are encouraged to provide a short proof sketch to provide intuition.
   - Inversely, any informal proof provided in the core of the paper should be complemented by formal proofs provided in appendix or supplemental material.
   - Theorems and Lemmas that the proof relies upon should be properly referenced.

4. **Experimental result reproducibility**

   Question: Does the paper fully disclose all the information needed to reproduce the main experimental results of the paper to the extent that it affects the main claims and/or conclusions of the paper (regardless of whether the code and data are provided or not)?

   Answer: [Yes]

   Justification: Appendix D specifies the synthetic Markov process, features, rewards, initialization, learning rates, random seeds, trajectory lengths, and exponent-fitting procedure.

   Guidelines:

   - The answer [N/A] means that the paper does not include experiments.
   - If the paper includes experiments, a [No] answer to this question will not be perceived well by the reviewers: Making the paper reproducible is important, regardless of whether the code and data are provided or not.
   - If the contribution is a dataset and/or model, the authors should describe the steps taken to make their results reproducible or verifiable.
   - Depending on the contribution, reproducibility can be accomplished in various ways. For example, if the contribution is a novel architecture, describing the architecture fully might suffice, or if the contribution is a specific model and empirical evaluation, it may be necessary to either make it possible for others to replicate the model with the same dataset, or provide access to the model. In general. releasing code and data is often one good way to accomplish this, but reproducibility can also be provided via detailed instructions for how to replicate the results, access to a hosted model (e.g., in the case of a large language model), releasing of a model checkpoint, or other means that are appropriate to the research performed.

- While NeurIPS does not require releasing code, the conference does require all submissions to provide some reasonable avenue for reproducibility, which may depend on the nature of the contribution. For example
  (a) If the contribution is primarily a new algorithm, the paper should make it clear how to reproduce that algorithm.
  (b) If the contribution is primarily a new model architecture, the paper should describe the architecture clearly and fully.
  (c) If the contribution is a new model (e.g., a large language model), then there should either be a way to access this model for reproducing the results or a way to reproduce the model (e.g., with an open-source dataset or instructions for how to construct the dataset).
  (d) We recognize that reproducibility may be tricky in some cases, in which case authors are welcome to describe the particular way they provide for reproducibility. In the case of closed-source models, it may be that access to the model is limited in some way (e.g., to registered users), but it should be possible for other researchers to have some path to reproducing or verifying the results.

5. **Open access to data and code**

   Question: Does the paper provide open access to the data and code, with sufficient instructions to faithfully reproduce the main experimental results, as described in supplemental material?

   Answer: [No]

   Justification: The experimental code and recorded trajectories are not released. Appendix D provides the synthetic environment construction and the simulation and evaluation protocol.

   Guidelines:

   - The answer [N/A] means that paper does not include experiments requiring code.
   - Please see the NeurIPS code and data submission guidelines (https://neurips.cc/public/guides/CodeSubmissionPolicy) for more details.
   - While we encourage the release of code and data, we understand that this might not be possible, so [No] is an acceptable answer. Papers cannot be rejected simply for not including code, unless this is central to the contribution (e.g., for a new open-source benchmark).
   - The instructions should contain the exact command and environment needed to run to reproduce the results. See the NeurIPS code and data submission guidelines (https://neurips.cc/public/guides/CodeSubmissionPolicy) for more details.
   - The authors should provide instructions on data access and preparation, including how to access the raw data, preprocessed data, intermediate data, and generated data, etc.
   - The authors should provide scripts to reproduce all experimental results for the new proposed method and baselines. If only a subset of experiments are reproducible, they should state which ones are omitted from the script and why.
   - At submission time, to preserve anonymity, the authors should release anonymized versions (if applicable).
   - Providing as much information as possible in supplemental material (appended to the paper) is recommended, but including URLs to data and code is permitted.

6. **Experimental setting/details**

   Question: Does the paper specify all the training and test details (e.g., data splits, hyperparameters, how they were chosen, type of optimizer) necessary to understand the results?

   Answer: [Yes]

   Justification: Appendix D provides the stepsize constants, number and length of trajectories, recording grid, and regression window used for the numerical illustration.

   Guidelines:

   - The answer [N/A] means that the paper does not include experiments.
   - The experimental setting should be presented in the core of the paper to a level of detail that is necessary to appreciate the results and make sense of them.

- The full details can be provided either with the code, in appendix, or as supplemental material.

7. **Experiment statistical significance**

   Question: Does the paper report error bars suitably and correctly defined or other appropriate information about the statistical significance of the experiments?

   Answer: [Yes]

   Justification: Figure 1 reports medians and empirical 25th–75th percentile intervals over 30 independent trajectories per stepsize, capturing variability from the sampled Markovian paths.

   Guidelines:

   - The answer [N/A] means that the paper does not include experiments.
   - The authors should answer [Yes] if the results are accompanied by error bars, confidence intervals, or statistical significance tests, at least for the experiments that support the main claims of the paper.
   - The factors of variability that the error bars are capturing should be clearly stated (for example, train/test split, initialization, random drawing of some parameter, or overall run with given experimental conditions).
   - The method for calculating the error bars should be explained (closed form formula, call to a library function, bootstrap, etc.)
   - The assumptions made should be given (e.g., Normally distributed errors).
   - It should be clear whether the error bar is the standard deviation or the standard error of the mean.
   - It is OK to report 1-sigma error bars, but one should state it. The authors should preferably report a 2-sigma error bar than state that they have a 96% CI, if the hypothesis of Normality of errors is not verified.
   - For asymmetric distributions, the authors should be careful not to show in tables or figures symmetric error bars that would yield results that are out of range (e.g., negative error rates).
   - If error bars are reported in tables or plots, the authors should explain in the text how they were calculated and reference the corresponding figures or tables in the text.

8. **Experiments compute resources**

   Question: For each experiment, does the paper provide sufficient information on the computer resources (type of compute workers, memory, time of execution) needed to reproduce the experiments?

   Answer: [No]

   Justification: Appendix D specifies the simulation scale, but hardware specifications and execution times are not reported.

   Guidelines:

   - The answer [N/A] means that the paper does not include experiments.
   - The paper should indicate the type of compute workers CPU or GPU, internal cluster, or cloud provider, including relevant memory and storage.
   - The paper should provide the amount of compute required for each of the individual experimental runs as well as estimate the total compute.
   - The paper should disclose whether the full research project required more compute than the experiments reported in the paper (e.g., preliminary or failed experiments that didn't make it into the paper).

9. **Code of ethics**

   Question: Does the research conducted in the paper conform, in every respect, with the NeurIPS Code of Ethics https://neurips.cc/public/EthicsGuidelines?

   Answer: [Yes]

Justification: This work consists of theoretical convergence analysis and a synthetic numerical illustration; it involves no human subjects, external data collection, released models, or deployment, and conforms to the NeurIPS Code of Ethics.

Guidelines:

- The answer [N/A] means that the authors have not reviewed the NeurIPS Code of Ethics.
- If the authors answer [No], they should explain the special circumstances that require a deviation from the Code of Ethics.
- The authors should make sure to preserve anonymity (e.g., if there is a special consideration due to laws or regulations in their jurisdiction).

10. **Broader impacts**

Question: Does the paper discuss both potential positive societal impacts and negative societal impacts of the work performed?

Answer: [N/A]

Justification: [N/A]

Guidelines:

- The answer [N/A] means that there is no societal impact of the work performed.
- If the authors answer [N/A] or [No], they should explain why their work has no societal impact or why the paper does not address societal impact.
- Examples of negative societal impacts include potential malicious or unintended uses (e.g., disinformation, generating fake profiles, surveillance), fairness considerations (e.g., deployment of technologies that could make decisions that unfairly impact specific groups), privacy considerations, and security considerations.
- The conference expects that many papers will be foundational research and not tied to particular applications, let alone deployments. However, if there is a direct path to any negative applications, the authors should point it out. For example, it is legitimate to point out that an improvement in the quality of generative models could be used to generate Deepfakes for disinformation. On the other hand, it is not needed to point out that a generic algorithm for optimizing neural networks could enable people to train models that generate Deepfakes faster.
- The authors should consider possible harms that could arise when the technology is being used as intended and functioning correctly, harms that could arise when the technology is being used as intended but gives incorrect results, and harms following from (intentional or unintentional) misuse of the technology.
- If there are negative societal impacts, the authors could also discuss possible mitigation strategies (e.g., gated release of models, providing defenses in addition to attacks, mechanisms for monitoring misuse, mechanisms to monitor how a system learns from feedback over time, improving the efficiency and accessibility of ML).

11. **Safeguards**

Question: Does the paper describe safeguards that have been put in place for responsible release of data or models that have a high risk for misuse (e.g., pre-trained language models, image generators, or scraped datasets)?

Answer: [N/A]

Justification: [N/A]

Guidelines:

- The answer [N/A] means that the paper poses no such risks.
- Released models that have a high risk for misuse or dual-use should be released with necessary safeguards to allow for controlled use of the model, for example by requiring that users adhere to usage guidelines or restrictions to access the model or implementing safety filters.
- Datasets that have been scraped from the Internet could pose safety risks. The authors should describe how they avoided releasing unsafe images.

- We recognize that providing effective safeguards is challenging, and many papers do not require this, but we encourage authors to take this into account and make a best faith effort.

12. **Licenses for existing assets**

    Question: Are the creators or original owners of assets (e.g., code, data, models), used in the paper, properly credited and are the license and terms of use explicitly mentioned and properly respected?

    Answer: [N/A]

    Justification: [N/A]

    Guidelines:

    - The answer [N/A] means that the paper does not use existing assets.
    - The authors should cite the original paper that produced the code package or dataset.
    - The authors should state which version of the asset is used and, if possible, include a URL.
    - The name of the license (e.g., CC-BY 4.0) should be included for each asset.
    - For scraped data from a particular source (e.g., website), the copyright and terms of service of that source should be provided.
    - If assets are released, the license, copyright information, and terms of use in the package should be provided. For popular datasets, `paperswithcode.com/datasets` has curated licenses for some datasets. Their licensing guide can help determine the license of a dataset.
    - For existing datasets that are re-packaged, both the original license and the license of the derived asset (if it has changed) should be provided.
    - If this information is not available online, the authors are encouraged to reach out to the asset's creators.

13. **New assets**

    Question: Are new assets introduced in the paper well documented and is the documentation provided alongside the assets?

    Answer: [N/A]

    Justification: [N/A]

    Guidelines:

    - The answer [N/A] means that the paper does not release new assets.
    - Researchers should communicate the details of the dataset/code/model as part of their submissions via structured templates. This includes details about training, license, limitations, etc.
    - The paper should discuss whether and how consent was obtained from people whose asset is used.
    - At submission time, remember to anonymize your assets (if applicable). You can either create an anonymized URL or include an anonymized zip file.

14. **Crowdsourcing and research with human subjects**

    Question: For crowdsourcing experiments and research with human subjects, does the paper include the full text of instructions given to participants and screenshots, if applicable, as well as details about compensation (if any)?

    Answer: [N/A]

    Justification: [N/A]

    Guidelines:

    - The answer [N/A] means that the paper does not involve crowdsourcing nor research with human subjects.
    - Including this information in the supplemental material is fine, but if the main contribution of the paper involves human subjects, then as much detail as possible should be included in the main paper.

- According to the NeurIPS Code of Ethics, workers involved in data collection, curation, or other labor should be paid at least the minimum wage in the country of the data collector.

15. **Institutional review board (IRB) approvals or equivalent for research with human subjects**

    Question: Does the paper describe potential risks incurred by study participants, whether such risks were disclosed to the subjects, and whether Institutional Review Board (IRB) approvals (or an equivalent approval/review based on the requirements of your country or institution) were obtained?

    Answer: [N/A]

    Justification: [N/A]

    Guidelines:

    - The answer [N/A] means that the paper does not involve crowdsourcing nor research with human subjects.
    - Depending on the country in which research is conducted, IRB approval (or equivalent) may be required for any human subjects research. If you obtained IRB approval, you should clearly state this in the paper.
    - We recognize that the procedures for this may vary significantly between institutions and locations, and we expect authors to adhere to the NeurIPS Code of Ethics and the guidelines for their institution.
    - For initial submissions, do not include any information that would break anonymity (if applicable), such as the institution conducting the review.

16. **Declaration of LLM usage**

    Question: Does the paper describe the usage of LLMs if it is an important, original, or non-standard component of the core methods in this research? Note that if the LLM is used only for writing, editing, or formatting purposes and does not impact the core methodology, scientific rigor, or originality of the research, declaration is not required.

    Answer: [N/A]

    Justification: [N/A]

    Guidelines:

    - The answer [N/A] means that the core method development in this research does not involve LLMs as any important, original, or non-standard components.
    - Please refer to our LLM policy in the NeurIPS handbook for what should or should not be described.