Journal of Machine Learning Research 26 (2025) 1-76 Submitted 1/24; Revised 10/24; Published 2/25

The ODE Method for Stochastic Approximation and Reinforcement Learning with Markovian Noise

Shuze Daniel Liu SHUZELIU@VIRGINIA.EDU
Department of Computer Science
University of Virginia
Charlottesville, VA, 22903, USA

Shuhang Chen SHUHANG@SCALEDFOUNDATIONS.AI
Scaled Foundations

Shangtong Zhang SHANGTONG@VIRGINIA.EDU
Department of Computer Science
University of Virginia
Charlottesville, VA, 22903, USA

Editor: Martha White

Abstract

Stochastic approximation is a class of algorithms that update a vector iteratively, incrementally, and stochastically, including, e.g., stochastic gradient descent and temporal difference learning. One fundamental challenge in analyzing a stochastic approximation algorithm is to establish its stability, i.e., to show that the stochastic vector iterates are bounded almost surely. In this paper, we extend the celebrated Borkar-Meyn theorem for stability from the Martingale difference noise setting to the Markovian noise setting, which greatly improves its applicability in reinforcement learning, especially in those off-policy reinforcement learning algorithms with linear function approximation and eligibility traces. Central to our analysis is the diminishing asymptotic rate of change of a few functions, which is implied by both a form of the strong law of large numbers and a form of the law of the iterated logarithm.

Keywords: stochastic approximation, reinforcement learning, stability, almost sure convergence, eligibility trace

1. Introduction

Stochastic approximation (Robbins and Monro, 1951; Benveniste et al., 1990; Kushner and Yin, 2003; Borkar, 2009) is a class of algorithms that update a vector iteratively, incrementally, and stochastically. Successful examples include stochastic gradient descent (Kiefer and Wolfowitz, 1952) and temporal difference learning (Sutton, 1988). Given an initial $x_0 \in \mathbb{R}^d$, stochastic approximation algorithms typically generate a sequence of vectors $\{x_n\}$ recursively as

$$x_{n+1} = x_n + \alpha(n)H(x_n, Y_{n+1}) \quad n = 0, 1, \dots \tag{1}$$

Here $\{\alpha(n)\}_{n=0}^{\infty}$ is a sequence of deterministic learning rates, $\{Y_n\}_{n=1}^{\infty}$ is a sequence of random noise in a general space $\mathcal{Y}$ (not necessarily compact), and $H : \mathbb{R}^d \times \mathcal{Y} \to \mathbb{R}^d$ is a function that maps the current iterate x_n and noise Y_{n+1} to the actual incremental update.

One way to analyze the asymptotic behavior of $\{x_n\}$ is to regard $\{x_n\}$ as Euler's discretization of the ODE

$$\frac{\mathrm{d}x(t)}{\mathrm{d}t} = h(x(t)), \tag{2}$$

where $h(x) \doteq \mathbb{E}[H(x, y)]$ is the expected updates (the expectation will be rigorously defined shortly). Then the asymptotic behavior of the discrete and stochastic iterates $\{x_n\}$ can be characterized by continuous and deterministic trajectories of the ODE (2). To establish this connection between the two, however, requires to establish the stability of $\{x_n\}$ first (Kushner and Yin, 2003; Borkar, 2009). In other words, one needs to first show that

$$\sup_n \|x_n\| < \infty \quad \text{a.s.},$$

which is in general challenging. Once the stability is confirmed, the convergence of $\{x_n\}$ follows easily (Kushner and Yin, 2003; Borkar, 2009). The seminal Borkar-Meyn theorem (Borkar and Meyn, 2000) establishes the desired stability assuming the global asymptotic stability of the following ODE

$$\frac{\mathrm{d}x(t)}{\mathrm{d}t} = h_\infty(x),$$

where $h_\infty(x) \doteq \frac{h(cx)}{c}$. Despite the celebrated success of the Borkar-Meyn theorem (see, e.g., Abounadi et al. (2001); Maei (2011)), one major limit is that the Borkar-Meyn theorem requires $\{Y_n\}$ to be i.i.d. noise. As a result, $\{H(x_n, Y_{n+1}) - h(x_n)\}_{n=0}^\infty$ is then a Martingale difference sequence and the Martingale convergence theorem applies under certain conditions. However, in many Reinforcement Learning (RL, Sutton and Barto (2018)) problems, $\{Y_n\}$ is a Markov chain and is not i.i.d. *Our main contribution is to extend the Borkar-Meyn theorem to the Markovian noise setting with verifiable assumptions.* The extension to Markovian noise has been previously explored by Ramaswamy and Bhatnagar (2018); Borkar et al. (2021). However, their assumptions are way more restrictive than ours so their results are not applicable in many important RL algorithms, particularly, off-policy RL algorithms with eligibility traces (Yu, 2012, 2015, 2017). See Section 5 for more discussion on this class of RL algorithms.

In Ramaswamy and Bhatnagar (2018), it is assumed that the Differential Inclusion (DI)

$$\frac{\mathrm{d}x(t)}{\mathrm{d}t} \in \overline{\mathrm{co}}\{H_\infty(x(t), y) | y \in \mathcal{Y}\}$$

is stable, where $\overline{\mathrm{co}}(\cdot)$ denotes the convex hull and $H_\infty(x, y) \doteq \lim_{c\to\infty} \frac{H(cx,y)}{c}$. To demonstrate the challenge in verifying this assumption, we consider a special linear case where $H(x, y) = A(y)x + b(y)$ for some matrix-valued function $A(y)$ and vector-valued function $b(y)$. Then one sufficient and commonly used condition (Molchanov and Pyatnitskiy, 1989) for this DI to be stable is that the $A(y)$ is uniformly negative definite, i.e., there exists some strictly positive η such that $x^\top A(y)x \leq -\eta\|x\|^2 \, \forall x \in \mathbb{R}^d, y \in \mathcal{Y}$. However, in many RL algorithms (e.g., Sutton (1988); Sutton et al. (2008, 2009, 2016), as well as the off-policy RL algorithms with eligibility traces in Section 5), we can at most say that $\mathbb{E}[A(y)]$ is negative definite. The individual matrix $A(y)$ does not have any special property. Intuitively, Ramaswamy and Bhatnagar (2018) assume that the function $H_\infty(x, y)$ behaves well almost surely, significantly limiting its application in RL. In fact, we are not aware of any application of Ramaswamy and Bhatnagar (2018) in standard RL algorithms. By contrast, we only need $h_\infty(x)$ to behave well, i.e., we only need $H_\infty(x, y)$ to behave well in expectation. Ramaswamy and Bhatnagar (2018) also assume $\mathcal{Y}$ to be compact. Unfortunately, in many important RL algorithms mentioned above, neither DI's stability nor the compactness holds.

In Borkar et al. (2021), it is assumed that a V4 Laypunov drift condition holds for $\{Y_n\}$ and the eighth moment of some function is bounded. Unfortunately, in many important RL algorithms (see, e.g., those in Section 5), neither assumption holds. We instead establish the stability via examining the *asymptotic rate of change* of certain functions, inspired by Kushner and Yin (2003). When V4 does not hold, a form of the strong law of large numbers and a form of the law of the iterated logarithm can be used to establish the desired asymptotic rate of change. When V4 does hold, we only need the second moment, instead of the eighth moment, to be bounded to establish the desired asymptotic rate of change.

We demonstrate in Section 5 the wide applicability of our results in RL, especially in off-policy RL algorithms with linear function approximation and eligibility traces, where the Markovian noise $\{Y_n\}$ can easily grow *unbounded almost surely* and have *unbounded second moment*. The key idea of our approach is to apply the Arzela-Ascoli theorem to the scaled iterates. Then the Moore-Osgood theorem computes a double limit, confirming that the scaled iterates converge to the corresponding limiting ODEs along a carefully chosen *subsequence*. This subsequence view is an important technical innovation of our approach. By contrast, previous works concerning the Borkar-Meyn theorem (Borkar and Meyn, 2000; Bhatnagar, 2011; Lakshminarayanan and Bhatnagar, 2017; Ramaswamy and Bhatnagar, 2017, 2018; Borkar et al., 2021) all seek to establish the convergence along the entire sequence to invoke a proof by contradiction argument to establish the desired stability. This subsequence view is essential for our approach because the Arzela-Ascoli theorem can only guarantee the existence of a convergent subsequence. As a result, we need a variant of the standard proof by contradiction argument to establish the desired stability.

2. Main Results

Assumption 1 *The Markov chain $\{Y_n\}$ has a unique invariant probability measure (i.e., stationary distribution), denoted by $d_\mathcal{Y}$.*

Technically speaking, the uniqueness and even the existence of the invariant probability measure can be relaxed, as long as the average of certain functions exists. We are, however, not aware of any applications where such relaxation is a must. We, therefore, use Assumption 1 to ease presentation and refer the reader to A1.3 in Chapter 6 of Kushner and Yin (2003) as an example of such relaxation. In light of the update (1), we use the convention that $\{Y_n\}$ starts from $n = 1$.

Assumption 2 *The learning rates $\{\alpha(n)\}$ are positive, decreasing, and satisfy*

$$\sum_{i=0}^{\infty} \alpha(i) = \infty, \ \lim_{n\to\infty} \alpha(n) = 0, \text{ and } \frac{\alpha(n) - \alpha(n+1)}{\alpha(n)} = \mathcal{O}\left(\alpha(n)\right). \tag{3}$$

Remark 1 *For any* $\alpha(n) = \frac{B_1}{(n+B_2)^\beta}$ *with* $\beta \in (0.5, 1]$, *it can be easily computed that*

$$\frac{\alpha(n) - \alpha(n+1)}{\alpha(n)} = \mathcal{O}\left(\frac{\beta}{n}\right) = \mathcal{O}\left(\alpha(n)\right).$$

Next, we make a few assumptions about the function H. For any $c \in [1, \infty)$, define

$$H_c(x, y) \doteq \frac{H(cx, y)}{c}. \tag{4}$$

The function H_c is the rescaled version of the function H and will be used to construct rescaled iterates, which are key techniques in proving the Borkar-Meyn theorem (see, e.g., Borkar and Meyn (2000); Borkar (2009)). Similar to Borkar and Meyn (2000); Borkar (2009), we need the limit of H_c to exist in a certain sense when $c \to \infty$.

Assumption 3 *There exists a measurable function* $H_\infty(x, y)$, *a function* $\kappa : \mathbb{R} \to \mathbb{R}$ *(independent of* x, y*), and a measurable function* $b(x, y)$ *such that for any* x, y,

$$\begin{aligned} H_c(x, y) - H_\infty(x, y) =& \kappa(c) b(x, y), \\ \lim_{c \to \infty} \kappa(c) =& 0, \end{aligned} \tag{5}$$

Moreover, there exists a measurable function $L_b(y)$ *such that* $\forall x, x', y$,

$$\left\|b(x, y) - b(x', y)\right\| \leq L_b(y) \left\|x - x'\right\|. \tag{6}$$

And the expectation $L_b \doteq \mathbb{E}_{y \sim d_\mathcal{Y}}\left[L_b(y)\right]$ *is well-defined and finite.*

Assumption 3 provides details on how H_c converges to H_∞ when $c \to \infty$. We note that in many RL applications, see, e.g., Section 5, the function $b(x, y)$ actually does not depend on x so (6) trivially holds. We consider $b(x, y)$ as a function of both x and y for generality. Next, we assume Lipschitz continuity of the functions H_c, which guarantees the existence and uniqueness of the solutions to the corresponding ODEs.

Assumption 4 *There exists a measurable function* $L(y)$ *such that for any* x, x', y,

$$\left\|H(x, y) - H(x', y)\right\| \leq L(y) \left\|x - x'\right\|, \tag{7}$$

$$\left\|H_\infty(x, y) - H_\infty(x', y)\right\| \leq L(y) \left\|x - x'\right\|. \tag{8}$$

Moreover, the following expectations are well-defined and finite for any x:

$$\begin{aligned} h(x) &\doteq \mathbb{E}_{y \sim d_\mathcal{Y}}[H(x, y)], \\ h_\infty(x) &\doteq \mathbb{E}_{y \sim d_\mathcal{Y}}[H_\infty(x, y)], \\ L &\doteq \mathbb{E}_{y \sim d_\mathcal{Y}}[L(y)]. \end{aligned}$$

Apparently, the function $x \mapsto H_c(x, y)$ shares the same Lipschitz constant $L(y)$ as the function $x \mapsto H(x, y)$. Similar to (4), we define

$$h_c(x) \doteq \frac{h(cx)}{c}.$$

The following assumption is the central assumption in the original proof of the Borkar-Meyn theorem.

Assumption 5 *(Assumption A5 in Chapter 3 of Borkar (2009)) As $c \to \infty$, $h_c(x)$ converges to $h_\infty(x)$ uniformly in x on any compact subsets of $\mathbb{R}^d$. The ODE*

$$\frac{dx(t)}{dt} = h_\infty(x(t)) \tag{ODE@∞}$$

has 0 as its globally asymptotically stable equilibrium.

We refer the reader to Dai (1995); Dai and Meyn (1995); Borkar and Meyn (2000); Borkar (2009); Fort et al. (2008); Meyn (2008, 2022) for the root and history of (ODE@∞).

Assumption 6 *Let g denote any of the following functions:*

$$y \mapsto H(x,y) \quad (\forall x), \tag{9}$$
$$y \mapsto L_b(y), \tag{10}$$
$$y \mapsto L(y). \tag{11}$$

Then for any initial condition Y_1, it holds that

$$\lim_{n\to\infty} \alpha(n) \sum_{i=1}^n \left(g(Y_i) - \mathbb{E}_{y\sim d_\mathcal{Y}}\left[g(y)\right]\right) = 0 \quad a.s. \tag{12}$$

Remark 2 *Consider $\alpha(n) = \frac{B_1}{(n+B_2)^\beta}$ as an example again. For $\beta = 1$, (12) is implied by the following Law of Large Numbers (LLN)*

$$\lim_{n\to\infty} \frac{1}{n} \sum_{i=1}^n \left(g(Y_i) - \mathbb{E}_{y\sim d_\mathcal{Y}}\left[g(y)\right]\right) = 0 \quad a.s. \tag{LLN}$$

For $\beta \in (0.5, 1]$, (12) is implied by the following Law of the Iterated Logarithm (LIL)

$$\left\| \sum_{i=1}^n \left(g(Y_n) - \mathbb{E}_{y\sim d_\mathcal{Y}}\left[g(y)\right]\right) \right\| \leq \zeta \sqrt{n \log\log n} \quad a.s., \tag{LIL}$$

where ζ is a sample path dependent finite constant.

Remark 3 *If the Markov chain $\{Y_n\}$ is positive[1] Harris[2], then (LLN) holds for any function g whenever $\mathbb{E}[\|g(y)\|] < \infty$ (Theorem 17.0.1 (i) of Meyn and Tweedie (2012)). If $\{Y_n\}$ is further V-uniformly ergodic[3], then (LIL) holds (Theorem 17.0.1 (iii) and (iv) of Meyn and Tweedie (2012)). For the special case where $\mathcal{Y}$ is finite, (LLN) holds when the Markov chain is irreducible and (LIL) holds when it is further aperiodic.*

Remark 4 *We note that (LLN) is stronger than Doob's strong law of large numbers on stationary processes (see, e.g., Theorem 17.1.2 of Meyn and Tweedie (2012), referred to as Doob's LLN hereafter). Doob's LLN concludes (at most) that (LLN) holds for any $Y_1 \in \mathcal{Y}_g$,*

[1] See page 235 of Meyn and Tweedie (2012) for the definition of positive chains.
[2] See page 204 of Meyn and Tweedie (2012) for the definition of Harris chains.
[3] See page 387 of Meyn and Tweedie (2012) for the definition of V-uniform ergodicity.

where $\mathcal{Y}_g$ is an unknown, probably g-dependent set such that $d_\mathcal{Y}(\mathcal{Y}_g) = 1$. If we use only Doob's LLN, all the "almost surely" statements in the paper must be replaced by "$\mathcal{Y}_*$-almost surely", where $\mathcal{Y}_* \doteq \bigcap_g \mathcal{Y}_g$. This means that all the statements hold only when $Y_1 \in \mathcal{Y}_*$. However, since the g functions in Assumption 6 depend on x, this $\mathcal{Y}_*$ is an intersection of possibly uncountably many sets $\{\mathcal{Y}_g\}$. It is possible that in some applications $\mathcal{Y}_*$ turns out to be a set of interest, where (LLN) can indeed be relaxed to Doob's LLN. But in general, characterizing $\mathcal{Y}_*$ is pretty challenging.

Remark 5 *The Markov chain $\{Y_n\}$ we consider in our RL applications in Section 5 is a general space Markov chain but is not positive Harris. Fortunately, Yu (2012, 2015, 2017) have established that (LLN) holds for those chains. Whether (LIL) holds for those chains remains open.*

To better contrast our work with Borkar et al. (2021), in the following, we provide an alternative to Assumption 6.

Assumption 6′ *The learning rates $\{\alpha(n)\}$ further satisfy $\sum_{n=0}^{\infty} \alpha(n)^2 < \infty$. The Markov chain $\{Y_n\}$ is ψ-irreducible[4]. The Lyapunov drift condition (V4) holds for the Markov chain $\{Y_n\}$.[5] In other words, there exists a Lyapunov function $v : \mathcal{Y} \to [1, \infty]$ such that for any $y \in \mathcal{Y}$,*

$$\mathbb{E}\left[v(Y_{n+1}) - v(Y_n) | Y_n = y\right] \le -\delta v(y) + \tau \mathbb{I}_C(y). \tag{V4}$$

Here $\delta > 0, \tau < \infty$ are constants, C is a small set[6], and $\mathbb{I}$ is the indicator function. Moreover, let g be any of the functions $H(0, y), L_b(y)$, and $L(y)$. Then $g \in \mathcal{L}^2_{v,\infty}$[7].

Assumption 6′ uses the idea of Borkar et al. (2021) but is weaker than its counterparts. See more detailed comparisons in Section 3.

Remark 6 *Assumption 6′ is listed here mostly for better comparison with Borkar et al. (2021). We are not aware of any RL application where Assumption 6′ holds but Assumption 6 does not hold. Instead, in the RL applications in Section 5, Assumption 6 holds but Assumption 6′ does not. That being said, the applicability of Assumptions 6 and 6′ outside RL is beyond the scope of this work.*

Having listed all the assumptions, our main theorem confirms the stability of $\{x_n\}$.

Theorem 7 *Let Assumptions 1 - 5 hold. Let Assumption 6 or 6′ hold. Then the iterates $\{x_n\}$ generated by (1) are stable, i.e.,*

$$\sup_n \|x_n\| < \infty \quad a.s.$$

Its proof is in Section 4. Once the stability is established, the convergence follows easily.

[4] See page 91 of Meyn and Tweedie (2012) for the definition of ψ-irreducibility.

[5] See page 371 of Meyn and Tweedie (2012) for in-depth discussion about (V4).

[6] See page 109 of Meyn and Tweedie (2012) for the definition of small sets.

[7] g belongs to $\mathcal{L}^p_{v,\infty}$ if and only if $\sup_{y\in\mathcal{Y}} \frac{\|g(y)\|_p^p}{v(y)} < \infty$, where v is the Lyapunov function in (V4).

Corollary 8 *Let Assumptions 1 - 5 hold. Let Assumption 6 or 6′ hold. Then the iterates $\{x_n\}$ generated by (1) converge almost surely to a (sample path dependent) bounded invariant set[8] of the ODE[9]*

$$\frac{\mathrm{d}x(t)}{\mathrm{d}t} = h(x(t)). \tag{13}$$

Arguments used in proving Corollary 8 are similar but much simpler than the counterparts in the proof of Theorem 7. We include a proof of Corollary 8 in Appendix B.9 with the details of those similar but simpler lemmas omitted to avoid verbatim repetition.

It is worth mentioning that it is easy to extend our results to more general updates

$$x_{n+1} = x_n + \alpha(n) \left(H(x_n, Y_{n+1}) + M_{n+1} + \epsilon_n \right),$$

where M_{n+1} is a Martingale difference sequence and ϵ_n is another additive noise. Similarly, it would require the asymptotic rate of change of $\{M_{n+1}\}$ and $\{\epsilon_n\}$ to diminish. We refer the reader to Kushner and Yin (2003) for more details. Since our main contribution is the stability under the Markovian noise $\{Y_{n+1}\}$, we use the simpler update rule (1) for improving clarity.

3. Related Work

General H. In this paper, the function H can be a general function and we do not make any linearity assumptions. We first compare our results with existing works applicable to general H and Markovian noise $\{Y_n\}$. Since convergence follows easily from stability, we focus on comparison in terms of establishing stability. Notably, the related stability results in Borkar and Meyn (2000); Borkar (2009) are superceded by Borkar et al. (2021). We, therefore, discuss only Borkar et al. (2021); Kushner and Yin (2003); Benveniste et al. (1990).

Compared with Borkar et al. (2021), our improvements lie in two aspects. First, central to Borkar et al. (2021) are (i) a V4 Laypunov drift condition, (ii) an aperiodicity assumption of $\{Y_n\}$, and (iii) a boundedness assumption $L(y) \in \mathcal{L}^8_{v,\infty}$. By contrast, our Assumption 6′ only requires $L(y) \in \mathcal{L}^2_{v,\infty}$ and does not need aperiodicity. Second, we further provide an approach that establishes the stability based on Assumption 6 without using (V4), aperiodicity, and the boundedness in $\mathcal{L}^8_{v,\infty}$. As noted in Remark 6, Assumption 6 is more applicable than Assumption 6′ in RL.

Compared with Kushner and Yin (2003), our main improvement is that we prove stability under the asymptotic rate of change conditions. By contrast, Kushner and Yin (2003) mostly use stability as a priori and are concerned with the convergence of projected algorithms in the form of

$$x_{n+1} = \Pi \left(x_n + \alpha(n) H(x_n, Y_{n+1}) \right),$$

[8] A set X is an invariant set of the ODE (13) if and only if for every $x \in X$, there exists a solution $x(t)$ to the ODE (13) such that $x(0) = x$ and $x(t) \in X$ for all $t \in (-\infty, \infty)$. If the ODE (13) is globally asymptotically stable, the only bounded invariant set is the singleton $\{x_*\}$, where x_* denotes the unique globally asymptotically stable equilibrium. We refer the reader to page 105 of Kushner and Yin (2003) for more details.

[9] By $\{x_n\}$ converges to a set X, we mean $\lim_{n\to\infty} \inf_{x\in X} \|x_n - x\| = 0$.

where Π is a projection to some compact set to ensure stability of $\{x_n\}$. As a result, the corresponding ODE (cf. Corollary 8) becomes

$$\frac{\mathrm{d}x(t)}{\mathrm{d}t} = h(x(t)) + \xi(t),$$

where $\xi(t)$ is a reflection term resulting from the projection Π. We refer the reader to Section 5.2 of Kushner and Yin (2003) for more details regarding this reflection term. Analyzing these reflection terms typically requires strong domain knowledge, see, e.g., Yu (2015); Zhang et al. (2021b), and Section 5.4 of Borkar (2009).

We argue that this work combines the best of both Borkar and Meyn (2000) and Kushner and Yin (2003), i.e., the ODE@∞ technique for establishing stability from Borkar and Meyn (2000) and the asymptotic rate of change technique for averaging out the Markovian noise $\{Y_n\}$. As a result, our results are more general than both Borkar et al. (2021) and Kushner and Yin (2003) in the aforementioned sense.

Compared with Benveniste et al. (1990), our main improvement is that despite the proof under Assumption $6'$ essentially using Poisson's equation[10], the proof under Assumption 6 does not need Poisson's equation at all. Notably, Benveniste et al. (1990) assume Poisson's equation directly without specifying sufficient conditions to establish Poisson's equation. Moreover, to establish stability, Benveniste et al. (1990) require a Lyapunov function for the ODE (13) that is always greater than or equal to $\alpha\|\cdot\|^2$ for some $\alpha > 0$ (Condition (ii) of Theorem 17 in Benveniste et al. (1990)). By contrast, our Assumption 5 does not put any restriction on the possible Lyapunov functions. We also note that Borkar et al. (2021) is also based on an error representation similar to Benveniste et al. (1990) enabled by Poisson's equation.

Linear H. If we further assume that the function $H(x, y)$ has a linear form, i.e.,

$$H(x, y) = A(y)x + b(y),$$

there are several other results regarding the stability (and thus convergence), e.g., Konda and Tsitsiklis (1999); Tadic (2001); Yu (2015) and Proposition 4.8 of Bertsekas and Tsitsiklis (1996). They, however, all require that the matrix $A \doteq \mathbb{E}_{y\sim d_\mathcal{Y}}[A(y)]$ is negative definite[11]. But contrast, our Assumption 5 only requires A to be Hurwitz[12] (see, e.g., Theorem 4.5 of Khalil (2002)), which is a weaker condition.[13] In Section 5, we provide a concrete RL algorithm where the corresponding A matrix is Hurwitz but not negative definite.

[10] Let g be a function defined on $\mathcal{Y}$. The Poisson's equation holds for g if there exists a finite function $\hat{g}$ such that $\hat{g}(y) = g(y) - \mathbb{E}_{y\sim d_\mathcal{Y}}[g(y)] + \int_\mathcal{Y} P(y, y')\hat{g}(y')dy'$ holds for any $y \in \mathcal{Y}$, where P denotes the transition kernel of $\{Y_n\}$. The drift condition (V4), together with some other mild conditions, is sufficient to ensure the existence of Poisson's equation. We refer the reader to Theorem 17.4.2 of Meyn and Tweedie (2012) for more details.

[11] A real matrix A, not necessarily symmetric, is negative definite if and only if all the eigenvalues of the symmetric matrix $A + A^\top$ is strictly negative.

[12] A real matrix A is Hurwitz if and only if the real parts of all its eigenvalues are strictly negative.

[13] All negative definite matrices are Hurwitz, but many Hurwitz matrices are not negative definite. See Chapter 2 of Horn and Johnson (1991) for more details.

Local clock. Another approach to deal with Markovian noise $\{Y_n\}$ is to apply results in asynchronous schemes. We refer the reader to Chapter 7 of Borkar (2009) for details. The major limitation is that it requires count-based learning rates. At the n-th iteration, instead of using $\alpha(n)$, where n can be regarded as a "global lock", the asynchronous schemes use $\alpha(\varpi(n, Y_{n+1}))$ as the learning rate, where $\varpi(n, y)$ counts the number of visits to the state y until time n and can be regarded as a "local clock". The asynchronous schemes also have other assumptions regarding the local clock. Successful examples include Abounadi et al. (2001); Wan et al. (2021). However, we are not aware of any successful applications of such count-based learning rates in RL with function approximation, where an RL algorithm typically only has access to some feature $\phi(Y_n)$ instead of Y_n directly. Unless ϕ is a one-to-one mapping, there will be no way to count the state visitation.

Other type of noise. The Borkar-Meyn theorem applies to only Martingale difference noise, which is, later on, relaxed to allow more types of noise, e.g., Bhatnagar (2011); Ramaswamy and Bhatnagar (2017). However, none of those extensions applies to general Markovian noise.

4. Proof of Theorem 7

This section is dedicated to proving Theorem 7. Overall, we prove by contradiction. Section 4.1 sets up notations and establishes the desired diminishing asymptotic rate of change of a few functions. Section 4.2 establishes the desired equicontinuity. Section 4.3 assumes the opposite and thus identifies a subsequence of interest. Section 4.4 analyzes the property of the subsequence, helping the *reductio ad absurdum* in Section 4.5. Lemmas in this section are derived on an arbitrary sample path $\{x_0, \{Y_i\}_{i=1}^{\infty}\}$ such that the assumptions in Section 2 hold. Thus, we omit "$a.s.$" on the lemma statements for simplicity.

4.1 Diminishing Asymptotic Rate of Change

We divide the non-negative real axis $[0, \infty)$ into segments of length $\{\alpha(i)\}_{i=0,1,\dots}$. Those segments are then grouped into larger intervals $\{[T_n, T_{n+1})\}_{n=0,1,\dots}$. The sequence $\{T_n\}$ has the property that $T_{n+1} - T_n \approx T$ for some fixed T and as n tends to ∞, the error in this approximation diminishes. Precisely speaking, we define

$$
\begin{aligned}
t(0) &\doteq 0, \\
t(n) &\doteq \sum_{i=0}^{n-1} \alpha(i) \quad n = 1, 2, \dots \quad .
\end{aligned}
$$

For any $T > 0$, define

$$
m(T) = \max\{i | T \geq t(i)\} \tag{14}
$$

to be the largest i that has $t(i)$ smaller or equal to T. Intuitively, $t(m(T))$ is "just" left to T in the real axis. Then $t(m(T))$ has the follow properties:

$$
t(m(T)) \leq T < t(m(T) + 1) = t(m(T)) + \alpha(m(T)), \tag{15}
$$

$$
t(m(T)) > T - \alpha(m(T)). \tag{16}
$$

Define

$$\begin{aligned}
&T_0 = 0, \\
&T_{n+1} = t(m(T_n + T) + 1).
\end{aligned} \tag{17}$$

Intuitively, T_{n+1} is "just" right to $T_n + T$ in the real axis. For proving Theorem 7, it suffices to work with solutions of ODEs in only $[0, \infty)$. But for Corollary 8, it is necessary to consider solutions of ODEs in $(-\infty, \infty)$. To this end, we define

$$\begin{aligned}
\alpha(i) =& 0 \quad \forall i < 0, \\
m(t) =& 0 \quad \forall t \leq 0,
\end{aligned} \tag{18}$$

for simplifying notations. For any given function f with domain $\mathcal{Y}$, its asymptotic rate of change is defined as

$$\limsup_n \sup_{-\tau \leq t_1 \leq t_2 \leq \tau} \left\| \sum_{i=m(t(n)+t_1)}^{m(t(n)+t_2)-1} \alpha(i)[f(Y_{i+1}) - \mathbb{E}_{y \sim d_{\mathcal{Y}}}[f(y)]] \right\|.$$

The asymptotic rate of change characterizes the asymptotic regularity of the sequence $\{f(Y_n)\}$ and is a powerful tool to study stochastic approximation iterates. We refer the reader to Sections 5.3.2 and 6.2 of Kushner and Yin (2003) for an in-depth exposition of this tool. In the following, we demonstrate that the asymptotic rate of change is 0 for the functions in Assumption 6.

Lemma 9 *Let Assumptions 1, 2, and 4 hold. Let Assumption 6 or 6′ hold. Then the asymptotic rate of change of the functions (9), (10), and (11) is 0, i.e., for any fixed $\tau > 0$ and x, it holds that*

$$\limsup_n \sup_{-\tau \leq t_1 \leq t_2 \leq \tau} \left\| \sum_{i=m(t(n)+t_1)}^{m(t(n)+t_2)-1} \alpha(i) \left[H(x, Y_{i+1}) - h(x)\right] \right\| = 0 \quad a.s.,$$

$$\limsup_n \sup_{-\tau \leq t_1 \leq t_2 \leq \tau} \left\| \sum_{i=m(t(n)+t_1)}^{m(t(n)+t_2)-1} \alpha(i)[L_b(Y_{i+1}) - L_b] \right\| = 0 \quad a.s., \tag{19}$$

$$\limsup_n \sup_{-\tau \leq t_1 \leq t_2 \leq \tau} \left\| \sum_{i=m(t(n)+t_1)}^{m(t(n)+t_2)-1} \alpha(i)[L(Y_{i+1}) - L] \right\| = 0 \quad a.s. \tag{20}$$

Its proof is in Appendix D.1. Furthermore, the convergence of H_c to H_∞ in Assumption 3 demonstrates a similar pattern.

Lemma 10 *Let Assumptions 1, 2, 3, and 4 hold. Let Assumption 6 or 6′ hold. It then holds that*

$$\lim_{c \to \infty} \sup_{x \in \mathcal{B}} \sup_n \sup_{t \in [0,T]} \left\| \sum_{i=m(T_n)}^{m(T_n+t)-1} \alpha(i) \left[H_c(x, Y_{i+1}) - H_\infty(x, Y_{i+1})\right] \right\| = 0 \quad a.s.,$$

where $\mathcal{B}$ denote an arbitrary compact set of $\mathbb{R}^d$.

Its proof is in Appendix B.1.

4.2 Equicontinuity of Scaled Iterates

Fix a sample path $\{x_0, \{Y_n\}\}$. Let $\bar{x}(t)$ be the piecewise constant interpolation [14] of x_n at points $\{t(n)\}_{n=0,1,\dots}$, i.e.,

$$\bar{x}(t) \doteq \begin{cases} x_0 & t \in [0, t(1)) \\ x_1 & t \in [t(1), t(2)) \\ x_2 & t \in [t(2), t(3)) \\ \vdots \end{cases}$$

Using (14) to simplify it, we get

$$\bar{x}(t) \doteq x_{m(t)}. \tag{21}$$

Notably, $\bar{x}(t)$ is right continuous and has left limits. By (1), $\forall n \geq 0$, we have

$$\bar{x}(t(n+1)) = \bar{x}(t(n)) + \alpha(n) H(\bar{x}(t(n)), Y_{n+1}).$$

Now we scale $\bar{x}(t)$ in each segment $[T_n, T_{n+1})$.

Definition 11 *$\forall n \in \mathbb{N}, t \in [0, T)$, define*

$$\hat{x}(T_n + t) \doteq \frac{\bar{x}(T_n + t)}{r_n} \tag{22}$$

where

$$r_n \doteq \max\{1, \|\bar{x}(T_n)\|\}. \tag{23}$$

This implies

$$\forall n \in \mathbb{N}, \|\hat{x}(T_n)\| \leq 1. \tag{24}$$

Moreover[15], $\forall n \in \mathbb{N}, t \in [0, T)$,

$$\begin{aligned} \hat{x}(T_n + t) &= \frac{\bar{x}(T_n) + \sum_{i=m(T_n)}^{m(T_n+t)-1} \alpha(i) H(\bar{x}(t(i)), Y_{i+1})}{r_n}. \\ &= \hat{x}(T_n) + \sum_{i=m(T_n)}^{m(T_n+t)-1} \alpha(i) H_{r_n}(\hat{x}(t(i)), Y_{i+1}). \end{aligned}$$

The function $t \mapsto \hat{x}(T_n + t)$ is the scaled version of $\bar{x}(t)$ (by r_n) in the interval $[T_n, T_{n+1})$. Its domain is $[0, T_{n+1} - T_n)$. In most of the rest of this work, we will restrict it to $[0, T)$, such that the sequence of functions $\{t \mapsto \hat{x}(T_n + t)\}_{n=0,1,\dots}$ have the same domain $[0, T)$, which is crucial in applying the Arzela-Ascoli Theorem. The excess part $[T, T_{n+1} - T_n)$ diminishes asymptotically (cf. Lemma 35) and thus can be easily processed when necessary. Notably, $\hat{x}(T_n + t)$ can be regarded as the Euler's discretization of $z_n(t)$ defined below.

[14] It also works if we consider a piecewise linear interpolation following Borkar (2009). The piecewise linear interpolation, however, will significantly complicate the presentation. We, therefore, follow Kushner and Yin (2003) to use piecewise constant interpolation.

[15] In this paper, we use the convention that $\sum_{k=i}^{j} \alpha(k) = 0$ when $j < i$

Definition 12 *$\forall n \in \mathbb{N}, t \in [0, T)$, define $z_n(t)$ as the solution of the ODE*

$$\frac{dz_n(t)}{dt} = h_{r_n}(z_n(t)) \tag{25}$$

with initial condition

$$z_n(0) = \hat{x}(T_n). \tag{26}$$

Apparently, $z_n(t)$ can also be written as

$$z_n(t) = \hat{x}(T_n) + \int_0^t h_{r_n}(z_n(s))ds. \tag{27}$$

Ideally, we would like to see that the error of Euler's discretization diminishes asymptotically. Precisely speaking, the discretization error is defined as

$$f_n(t) \doteq \hat{x}(T_n + t) - z_n(t) \tag{28}$$

and we would like that $f_n(t)$ diminishes to 0 as $n \to \infty$ in a certain sense. To this end, we study the following three sequences of functions

$$\{t \mapsto \hat{x}(T_n + t)\}_{n=0}^{\infty}, \{z_n(t)\}_{n=0}^{\infty}, \{f_n(t)\}_{n=0}^{\infty}. \tag{29}$$

In particular, we show that they are all equicontinuous in the extended sense. To understand equicontinuity in the extended sense, we first give the definition of equicontinuity.

Definition 13 *A sequence of functions $\left\{g_n : [0, T) \to \mathbb{R}^K\right\}$ is equicontinuous on $[0, T)$ if $\sup_n \|g_n(0)\| < \infty$ and $\forall \epsilon > 0$, $\exists \delta > 0$ such that*

$$\sup_n \sup_{0 \leq |t_1 - t_2| \leq \delta, 0 \leq t_1 \leq t_2 < T} \|g_n(t_1) - g_n(t_2)\| \leq \epsilon.$$

One example of equicontinuity is a sequence of bounded Lipschitz continuous functions with a common Lipschitz constant. Obviously, if $\{g_n\}$ is equicontinuous, each g_n must be continuous. However, the functions of interest in this work, i.e., $\hat{x}(T_n + t), f_n(t)$, are not continuous so equicontinuity would not apply. We, therefore, introduce the following equicontinuity in the extended sense[16] akin to Kushner and Yin (2003).

Definition 14 *A sequence of functions $\left\{g_n : [0, T) \to \mathbb{R}^K\right\}$ is equicontinuous in the extended sense on $[0, T)$ if $\sup_n \|g_n(0)\| < \infty$ and $\forall \epsilon > 0$, $\exists \delta > 0$ such that*

$$\limsup_n \sup_{0 \leq |t_1 - t_2| \leq \delta, 0 \leq t_1 \leq t_2 < T} \|g_n(t_1) - g_n(t_2)\| \leq \epsilon.$$

[16] We must use this equicontinuity in the extended sense because we have chosen to use piecewise constant instead of piecewise linear interpolation. For piecewise linear interpolation, the standard equicontinuity is enough. However, as also argued in Kushner and Yin (2003), piecewise linear interpolation complicates the presentation much more than the equicontinuity in the extended sense.

Notably, Kushner and Yin (2003) show that $\left\{t \in (-\infty, \infty) \mapsto \bar{x}(t(n) + t) \in \mathbb{R}^d\right\}_{n=0}^{\infty}$ is equicontinuous in the extended sense with *a priori* that

$$\sup_n \|x_n\| < \infty.$$

We do *not* have this *a priori*. Instead, we prove *a posteriori* that

$$\sup_{n \geq 0, t \in [0,T)} \|\hat{x}(T_n + t)\| < \infty$$

and show that $\left\{t \in [0, T) \mapsto \hat{x}(T_n + t) \in \mathbb{R}^d\right\}_{n=0}^{\infty}$ is equicontinuous in the extended sense. We remark that our function $t \mapsto \hat{x}(T_n + t)$ actually belongs to the J_1 Skorokhod topology (Skorokhod, 1956; Billingsley, 1999; Kern, 2023), although we will not work on this topology explicitly. Nevertheless, the following lemmas establish the desired equicontinuity, where Lemma 9 plays a key role.

Lemma 15 *The three sequences of functions* $\{\hat{x}(T_n + t)\}, \{z_n(t)\}$, *and* $\{f_n(t)\}$ *are all equicontinuous in the extended sense on* $t \in [0, T)$.

Its proof is in appendix B.2.

4.3 A Convergent Subsequence

According to the Arzela-Ascoli theorem in the extended sense (Theorem A.4), a sequence of equicontinuous functions always has a subsequence of functions that uniformly converge to a continuous limit. In the following, we use this to identify a particular subsequence of interest.

We observe the following inequality

$$\forall n, \quad \left\|x_{m(T_n)}\right\| = \|\bar{x}(T_n)\| \leq r_n. \tag{30}$$

Thus, to prove Theorem 7, we first show

$$\sup_n r_n < \infty,$$

and which is implied by

$$\limsup_n r_n < \infty. \tag{31}$$

In the following, we aim to show (31) by contradiction. We first assume the opposite, i.e., $\limsup_n r_n = \infty$. Based on this assumption and applying Gronwall's inequality a few times, we can find a particular subsequence of interest, along which all the three sequences of functions in (29) converge uniformly.

Lemma 16 *Suppose* $\limsup_n r_n = \infty$. *Then there exists a subsequence* $\{n_k\}_{k=0}^{\infty} \subseteq \{0, 1, 2, \dots\}$ *that has the following properties:*

$$\lim_{k \to \infty} r_{n_k} = \infty,$$
$$r_{n_k + 1} > r_{n_k} \quad \forall k. \tag{32}$$

Moreover, there exist some continuous functions $f^{\lim}(t)$ *and* $\hat{x}^{\lim}(t)$ *such that* $\forall t \in [0, T)$*,*

$$\begin{aligned} \lim_{k\to\infty} f_{n_k}(t) =& f^{\lim}(t), \\ \lim_{k\to\infty} \hat{x}(T_{n_k} + t) =& \hat{x}^{\lim}(t), \end{aligned} \tag{33}$$

where both convergences are uniform in t *on* $[0, T)$*. Furthermore, let* $z^{\lim}(t)$ *denote the unique solution to the* (ODE@∞) *with the initial condition*

$$z^{\lim}(0) = \hat{x}^{\lim}(0),$$

in other words,

$$z^{\lim}(t) = \hat{x}^{\lim}(0) + \int_0^t h_\infty(z^{\lim}(s))ds. \tag{34}$$

Then $\forall t \in [0, T)$*, we have*

$$\lim_{k\to\infty} z_{n_k}(t) = z^{\lim}(t),$$

where the convergence is uniform in t *on* $[0, T)$*.*

Its proof is in Appendix B.3. We use the subsequence $\{n_k\}$ intensively in the remaining proofs.

4.4 Diminishing Discretization Error

Recall that $f_n(t)$ denotes the discretization error of $\hat{x}(T_n + t)$ of $z_n(t)$. We now proceed to prove that this discretization error diminishes along $\{n_k\}$. We note that we are able to improve over Borkar et al. (2021) because we only require the discretization error to diminish along the subsequence $\{n_k\}$, while Borkar et al. (2021) aim to show that the discretization error diminishes along the entire sequence $\{n\}$, which is unnecessary given (32).

In particular, we aim to prove that

$$\lim_{k\to\infty} \|f_{n_k}(t)\| = \left\|f^{\lim}(t)\right\| = 0.$$

This means $\hat{x}(T_{n_k} + t)$ is close to $z_{n_k}(t)$ as $k \to \infty$. For any $t \in [0, T)$, we have

$$\begin{aligned} &\lim_{k\to\infty} \|f_{n_k}(t)\| \\ =& \lim_{k\to\infty} \left\| \hat{x}(T_{n_k}) + \sum_{i=m(T_{n_k})}^{m(T_{n_k}+t)-1} \alpha(i) H_{r_{n_k}}(\hat{x}(t(i)), Y_{i+1}) - z_{n_k}(t) \right\| && \text{(by (28))} \\ =& \lim_{k\to\infty} \left\| \sum_{i=m(T_{n_k})}^{m(T_{n_k}+t)-1} \alpha(i) H_{r_{n_k}}(\hat{x}(t(i)), Y_{i+1}) - \int_0^t h_{r_{n_k}}(z_{n_k}(s))ds \right\| && \text{(by (27))} \\ \leq& \lim_{k\to\infty} \left\| \sum_{i=m(T_{n_k})}^{m(T_{n_k}+t)-1} \alpha(i) H_{r_{n_k}}(\hat{x}(t(i)), Y_{i+1}) - \int_0^t h_{r_{n_k}}(\hat{x}^{\lim}(s))ds \right\| \end{aligned}$$

$$+ \lim_{k\to\infty} \left\| \int_0^t h_{r_{n_k}}(\hat{x}^{\lim}(s))ds - \int_0^t h_{r_{n_k}}(z_{n_k}(s))ds \right\|. \tag{35}$$

We now prove that the first term in the RHS of (35) is 0. Precisely speaking, we aim to prove $\forall t \in [0, T)$,

$$\lim_{k\to\infty} \left\| \sum_{i=m(T_{n_k})}^{m(T_{n_k}+t)-1} \alpha(i) H_{r_{n_k}}(\hat{x}(t(i)), Y_{i+1}) - \int_0^t h_{r_{n_k}}(\hat{x}^{\lim}(s))ds \right\| = 0. \tag{36}$$

To compute the limit above, we first fix any $t \in [0, T)$ and compute the following stronger double limit, which implies the existence of the above limit (cf. Lemma 48).

$$\lim_{\substack{j\to\infty \\ k\to\infty}} \left\| \sum_{i=m(T_{n_k})}^{m(T_{n_k}+t)-1} \alpha(i) H_{r_{n_j}}(\hat{x}(t(i)), Y_{i+1}) - \int_0^t h_{r_{n_j}}(\hat{x}^{\lim}(s))ds \right\|. \tag{37}$$

To compute this double limit, we use the Moore-Osgood theorem (Theorem A.5) to make it iterated limits. To invoke the Moore-Osgood theorem, we first prove the uniform convergence in k when $j \to \infty$.

Lemma 17 *$\forall t \in [0, T)$,*

$$\begin{aligned} &\lim_{j\to\infty} \left\| \sum_{i=m(T_{n_k})}^{m(T_{n_k}+t)-1} \alpha(i) H_{r_{n_j}}(\hat{x}(t(i)), Y_{i+1}) - \int_0^t h_{r_{n_j}}(\hat{x}^{\lim}(s))ds \right\| \\ =&\left\| \sum_{i=m(T_{n_k})}^{m(T_{n_k}+t)-1} \alpha(i) H_{\infty}(\hat{x}(t(i)), Y_{i+1}) - \int_0^t h_{\infty}(\hat{x}^{\lim}(s))ds \right\| \end{aligned}$$

uniformly in k.

Its proof is in Appendix B.4, where Lemma 10 plays a key role. Next, we prove, for each j, the convergence with $k \to \infty$.

Lemma 18 *$\forall t \in [0, T)$, $\forall j$,*

$$\lim_{k\to\infty} \left\| \sum_{i=m(T_{n_k})}^{m(T_{n_k}+t)-1} \alpha(i) H_{r_{n_j}}(\hat{x}(t(i)), Y_{i+1}) - \int_0^t h_{r_{n_j}}(\hat{x}^{\lim}(s))ds \right\| = 0.$$

The proof of Lemma 18 follows the proof sketch of a similar problem on page 168 of Kushner and Yin (2003) with some minor changes and is the central averaging technique of Kushner and Yin (2003). We expect a reader familiar with Kushner and Yin (2003) should have belief in its correctness. We anyway still include all the details in the Appendix D.2 for completeness. We are now ready to compute the limit in (36).

Lemma 19 $\forall t \in [0, T)$,

$$\lim_{k\to\infty} \left\| \sum_{i=m(T_{n_k})}^{m(T_{n_k}+t)-1} \alpha(i) H_{r_{n_k}}(\hat{x}(t(i)), Y_{i+1}) - \int_0^t h_{r_{n_k}}(\hat{x}^{\lim}(s)) ds \right\| = 0.$$

Proof It follows immediately from Lemmas 17 & 18, the Moore-Osgood theorem, and Lemma 48. ■

Lemma 19 confirms that the first term in the RHS of (35) is 0. Moreover, it also enables us to rewrite $\hat{x}^{\lim}(t)$ from a summation form to an integral form.

$$\begin{aligned} &\hat{x}^{\lim}(t) \\ =&\lim_{k\to\infty} \hat{x}(T_{n_k}) + \sum_{i=m(T_{n_k})}^{m(T_{n_k}+t)-1} \alpha(i) H_{r_{n_k}}(\hat{x}(t(i)), Y_{i+1}) \\ =&\lim_{k\to\infty} \hat{x}(T_{n_k}) + \int_0^t h_{r_{n_k}}(\hat{x}^{\lim}(s)) ds. && \text{(by Lemma 19)} \quad (38) \end{aligned}$$

This, together with a few Gronwall's inequality arguments, confirms that the discretization error indeed diminishes along $\{n_k\}$.

Lemma 20 $\forall t \in [0, T)$,

$$\lim_{k\to\infty} \|f_{n_k}(t)\| = 0.$$

Its proof is in Appendix B.6.

4.5 Identifying Contradiction and Completing Proof

Having made sure that the error of the discretization $\hat{x}(T_n + t)$ of $z_n(t)$ diminishes along $\{n_k\}$, we now study the behavior $\hat{x}(T_{n_k} + t)$ through $z_{n_k}(t)$ and identify a contradiction. The underlying idea is identical to Borkar (2009). However, the execution is different so we cannot use the arguments from Borkar (2009) directly. Namely, to use the arguments in Chapter 3 of Borkar (2009) directly, we have to prove that the discretization error diminishes along the entire sequence. This is impossible for us because the Arzela-Ascoli theorem only guarantees convergence along the subsequence $\{n_k\}$. Nevertheless, after carefully choosing the subsequence in Lemma 16, we are still able to execute the contradiction idea as documented below.

Lemma 21 *Suppose* $\limsup_n r_n = \infty$. *Then there exists a* k_0 *such that*

$$r_{n_{k_0}+1} \leq r_{n_{k_0}}.$$

Its proof is in Appendix B.7. This lemma constructs a contradiction to (32). This means the proposition $\limsup_n r_n = \infty$ is impossible. This completes the proof of

$$\sup_n r_n < \infty. \tag{39}$$

By decomposition,

$$\begin{aligned}&\sup_n \|x_n\| \\ =&\sup_n \sup_{i\in\{i|m(T_n)\leq m(T_n)+i<m(T_{n+1})\}} \left\|x_{m(T_n)+i}\right\| - \left\|x_{m(T_n)}\right\| + \left\|x_{m(T_n)}\right\| \\ \leq&\sup_n \sup_{i\in\{i|m(T_n)\leq m(T_n)+i<m(T_{n+1})\}} \left\|x_{m(T_n)+i}\right\| - \left\|x_{m(T_n)}\right\| + \sup_n r_n. \qquad \text{(by (30))} \quad (40)\end{aligned}$$

We show the first term above is also bounded.

Lemma 22

$$\sup_n \sup_{i\in\{i|m(T_n)\leq m(T_n)+i<m(T_{n+1})\}} \left\|x_{m(T_n)+i}\right\| - \left\|x_{m(T_n)}\right\| < \infty.$$

Its proof is in Appendix B.8. Thus, (39), (40) and Lemma 22 conclude Theorem 7.

5. Applications in Reinforcement Learning

In this section, we discuss broad applications of Corollary 8 in RL. In particular, we both demonstrate state-of-the-art analysis in Section 5.3 and greatly simplify existing analysis in Section 5.4. We first introduce notations and lay out the background of RL.

All vectors are column vectors. For a vector $d \in \mathbb{R}^N$ with strictly positive entries, we use $\|x\|_d$ to denote the d-weighted ℓ_2 norm, i.e., $\|x\|_d \doteq \sqrt{\sum_{i=1}^N d_i x_i^2}$. We also abuse $\|\cdot\|_d$ to denote the corresponding induced matrix norm. We use $\|\cdot\|$ to denote a general norm that respects sub-multiplicity. We use vectors and functions interchangeably when it does not confuse. For example, for some $g : \mathcal{S} \to \mathbb{R}$, we also interpret g as a vector in $\mathbb{R}^{|\mathcal{S}|}$. We use $\Pi_{\Phi,d}$ to denote a projection operator that projects a vector to the column space of a matrix Φ, assuming Φ has a full column rank. In other words,

$$\Pi_{\Phi,d} v = \Phi \arg\min_\theta \|\Phi\theta - v\|_d^2.$$

When it is clear from the context, we write $\Pi_{\Phi,d}$ as Π_d for simplifying presentation.

We consider an MDP with a finite state space[17] $\mathcal{S}$, a finite action space $\mathcal{A}$, a reward function $r : \mathcal{S} \times \mathcal{A} \to \mathbb{R}$, a transition function $p : \mathcal{S} \times \mathcal{S} \times \mathcal{A} \to [0, 1]$, an initial distribution $p_0 : \mathcal{S} \to [0, 1]$, and a discount factor $\gamma \in [0, 1)$. At time step 0, an initial state S_0 is sampled from p_0. At time t, given the state S_t, the agent samples an action $A_t \sim \pi(\cdot|S_t)$, where $\pi : \mathcal{A} \times \mathcal{S} \to [0, 1]$ is the policy being followed by the agent. A reward $R_{t+1} \doteq r(S_t, A_t)$ is then emitted and the agent proceeds to a successor state $S_{t+1} \sim p(\cdot|S_t, A_t)$. The return at time t is defined as $G_t \doteq \sum_{i=1}^\infty \gamma^{i-1} R_{t+i}$, using which we define the state-value function $v_\pi(s)$ and action-value function $q_\pi(s)$ as

$$\begin{aligned} v_\pi(s) \doteq& \mathbb{E}_{\pi,p}\left[G_t|S_t = s\right], \\ q_\pi(s, a) \doteq& \mathbb{E}_{\pi,p}\left[G_t|S_t = s, A_t = a\right]. \end{aligned}$$

[17] It is worth mentioning that even if the MDP problem itself is finite, the Markov chains used to analyze many RL algorithms still evolve in an uncountable and unbounded space. This will be seen shortly.

The value function v_π is the unique fixed point of the Bellman operator

$$\mathcal{T}_\pi v \doteq r_\pi + \gamma P_\pi v,$$

where $r_\pi \in \mathbb{R}^{|\mathcal{S}|}$ is the reward vector induced by the policy π, i.e., $r_\pi(s) \doteq \sum_a \pi(a|s) r(s,a)$, and $P_\pi \in \mathbb{R}^{|\mathcal{S}|\times|\mathcal{S}|}$ is the transition matrix induced by the policy π, i.e., $P_\pi(s,s') \doteq \pi(a|s)p(s'|s,a)$. With a $\lambda \in [0,1]$, we can rewrite $v_\pi = \mathcal{T}_\pi v_\pi$ using the identity $v_\pi = (1-\lambda)v_\pi + \lambda \mathcal{T}_\pi v_\pi$ as

$$\begin{aligned}
v_\pi =& r_\pi + \gamma P_\pi((1-\lambda)v_\pi + \lambda \mathcal{T}_\pi v_\pi) \\
=& r_\pi + \gamma(1-\lambda)P_\pi v_\pi + \gamma\lambda P_\pi(r_\pi + \gamma P_\pi v_\pi) \\
=& r_\pi + \gamma\lambda P_\pi r_\pi + \gamma(1-\lambda)P_\pi v_\pi + \gamma^2\lambda P_\pi^2((1-\lambda)v_\pi + \lambda\mathcal{T}_\pi v_\pi) \\
=& \dots \\
=& \sum_{i=0}^\infty (\gamma\lambda P_\pi)^i r_\pi + (1-\lambda)\sum_{i=1}^\infty \lambda^{i-1}\gamma^i P_\pi^i v_\pi, \\
=& (I - \gamma\lambda P_\pi)^{-1} r_\pi + (1-\lambda)\gamma(I-\gamma\lambda P_\pi)^{-1}P_\pi v_\pi.
\end{aligned}$$

This suggests that we define a λ-Bellman operator as

$$\mathcal{T}_{\pi,\lambda} v \doteq r_{\pi,\lambda} + \gamma P_{\pi,\lambda} v,$$

where $r_{\pi,\lambda} \doteq (I-\gamma\lambda P_\pi)^{-1} r_\pi, P_{\pi,\lambda} \doteq (1-\lambda)(I-\gamma\lambda P_\pi)^{-1}P_\pi$. It is then easy to see that when $\lambda = 0$, $\mathcal{T}_{\pi,\lambda}$ reduces to $\mathcal{T}_\pi$. When $\lambda = 1$, $\mathcal{T}_{\pi,\lambda}$ reduces to a constant function that always output $(I-\gamma P_\pi)^{-1} r_\pi$. It is proved that $\mathcal{T}_{\pi,\lambda}$ is a $\frac{\gamma(1-\lambda)}{1-\gamma\lambda}$-contraction w.r.t. $\|\cdot\|_{d_\pi}$ (see, e.g., Lemma 6.6 of (Bertsekas and Tsitsiklis, 1996)), where we use $d_\pi \in \mathbb{R}^{|\mathcal{S}|}$ to denote the stationary distribution of the Markov chain induced by π. Obviously, v_π is the unique fixed point of $\mathcal{T}_{\pi,\lambda}$.

One fundamental task in RL is prediction, i.e., to estimate v_π, for which temporal difference (TD, Sutton (1988)) learning is the most powerful method. In particular, Sutton (1988) considers a linear architecture. Let $\phi : \mathcal{S} \to \mathbb{R}^K$ be the feature function that maps a state to a K-dimensional feature. Linear TD(λ) (Sutton, 1988) aims to find a $\theta \in \mathbb{R}^K$ such that $\phi(s)^\top\theta$ is close to $v_\pi(s)$ for every $s \in \mathcal{S}$. To this end, linear TD(λ) updates θ recursively as

$$\begin{aligned}
e_t =& \lambda\gamma e_{t-1} + \phi_t, \\
\theta_{t+1} =& \theta_t + \alpha_t\left(R_{t+1} + \gamma\phi_{t+1}^\top\theta_t - \phi_t^\top\theta_t\right)e_t,
\end{aligned} \tag{41}$$

where we have used $\phi_t \doteq \phi(S_t)$ as shorthand and $e_t \in \mathbb{R}^K$ is the *eligiblity trace* with an arbitrary initial e_{-1}. We use $\Phi \in \mathbb{R}^{|\mathcal{S}|\times K}$ to denote the feature matrix, each row of which is $\phi(s)^\top$. It is proved (Tsitsiklis and Roy, 1996) that, under some conditions, $\{\theta_t\}$ converges to the unique zero of $J_{\text{on}}(\theta) \doteq \|\Pi_{d_\pi}\mathcal{T}_{\pi,\lambda}\Phi\theta - \Phi\theta\|^2_{d_\pi}$. This $J_{\text{on}}(\theta)$ is referred to as the on-policy mean squared projected Bellman error (MSPBE).

In many scenarios, due to the concerns of data efficiency (Lin, 1992; Sutton et al., 2011) or safety (Dulac-Arnold et al., 2019), we would like to estimate v_π but select actions using a different policy, called μ. This is off-policy learning, where π is called the target policy and μ is called the behaivor policy. In the rest of this section, we always consider the off-policy setting, i.e., the action A_t is sampled from $\mu(\cdot|S_t)$. Correspondingly, off-policy linear TD(λ) updates θ recursively as

$$\begin{aligned} e_t =& \lambda\gamma\rho_{t-1}e_{t-1} + \phi_t, \\ \theta_{t+1} =& \theta_t + \alpha_t\rho_t\left(R_{t+1} + \gamma\phi_{t+1}^\top\theta_t - \phi_t^\top\theta_t\right)e_t, \end{aligned} \tag{42}$$

where $\rho_t \doteq \rho(S_t, A_t) \doteq \frac{\pi(A_t|S_t)}{\mu(A_t|S_t)}$ is the importance sampling ratio to account for the discrepancy in action selection between π and μ. Obviously, if $\pi = \mu$, then (42) reduces to (41). Let $d_\mu \in \mathbb{R}^{|\mathcal{S}|}$ be the stationary distribution of the Markov chain induced by μ. If $\{\theta_t\}$ in (42) converged, it would converge to the unique zero of

$$J_{\text{off}}(\theta) \doteq \left\|\Pi_{d_\mu}\mathcal{T}_{\pi,\lambda}\Phi\theta - \Phi\theta\right\|_{d_\mu}^2,$$

which is the off-policy MSPBE.

5.1 Eligibility Trace

The eligibility trace is one of the most fundamental ingredients in RL and is deeply rooted in RL since the very beginning of RL (Klopf, 1972; Sutton, 1978; Barto and Sutton, 1981a,b; Barto et al., 1983; Sutton, 1984). The eligibility trace in (41) is called the accumulating trace, first introduced in Barto and Sutton (1981a). Later on, this trace is also used in control by Rummery and Niranjan (1994). Its off-policy version in (42) is introduced by Precup et al. (2000, 2001) and further developed by Bertsekas and Yu (2009); Yu (2012). Other forms of traces include the Dutch trace introduced by Seijen and Sutton (2014) and the followon trace introduced by Sutton et al. (2016). In short, traces are usually used to accelerate credit assignment, which is a fundamental challenge in RL. Intuitively, traces are able to achieve this goal because they function as memory of the past. Empirically, RL algorithms with traces usually outperform those without traces (Sutton and Barto, 2018). Traces are also important in establishing the equivalence between backward and forward views of RL algorithms (Sutton et al., 2014).

Despite the superiority of traces in multiple aspects, they usually complicate the analysis of RL algorithms. Without any trace, to analyze an RL algorithm it is usually sufficient to consider the Markov chain $\{(S_t, A_t)\}$. Under a finite MDP assumption, this augmented Markov chain is still finite. Once trace is introduced, we, however, must consider the Markov chain $\{(S_t, A_t, e_t)\}$, see, e.g., Tsitsiklis and Roy (1996). This augmented Markov chain now immediately evolves in an uncountable space $\mathcal{S} \times \mathcal{A} \times \mathbb{R}^d$. In the on-policy case (cf. (41)), this is still managable. It is clear from (41) that e_t remains bounded almost surely. So the augmented Markov chain evolves in a compact space. In the off-policy case (cf. (42)), the trace e_t can easily be unbounded almost surely due to the importance sampling ratio ρ_{t-1} (Yu, 2012). The augmented Markov chain then evolves in an *unbounded and uncountable* space. Even worse, sometimes the second moment of e_t can also be unbounded (Yu, 2012), further complicating the analysis. Despite that e_t is demonstrated to obey a form of the strong law of large numbers (Yu, 2012), there does not exist a general tool to make use of this in convergence analysis before this work. In other words, this work is the first to provide a general tool to analyze the stability (and thus convergence) of RL algorithms with off-policy traces.

5.2 The Deadly Triad

Despite the aforementioned superiority of off-policy learning in safety and data efficiency, it complicates RL algorithms in at least two aspects. The first is that it makes traces extremely hard to analyze, as demonstrated in the section above. Second, it makes the RL algorithm behaves poorly in expectation. In other words, even if there is no noise (cf. replacing $H(x_n, Y_{n+1})$ with $h(x_n)$), the RL algorithm can still behave poorly. A concrete example is that, for a general λ, the iterates $\{\theta_t\}$ in (42) can possibly diverge to infinity, as documented in Baird (1995); Tsitsiklis and Roy (1996); Sutton and Barto (2018). This is the notorious *deadly triad*, which refers to the instability of an RL algorithm when it combines bootstrapping, function approximation, and off-policy learning simultaneously while maintaining a constant $\mathcal{O}(K)$ computational complexity each step.

The deadly triad has been one of the central challenges of RL in the past three decades and numerous works have been done in this topic (Precup et al., 2000, 2001; Sutton et al., 2008, 2009; Maei et al., 2009, 2010; Maei and Sutton, 2010; Maei, 2011; Sutton et al., 2011; Yu, 2012; Mahadevan et al., 2014; Liu et al., 2015; Yu, 2015; White and White, 2016; Mahmood et al., 2017; Yu, 2017; Wang et al., 2017; Touati et al., 2018; Liu et al., 2018; Zhang et al., 2020b; Nachum et al., 2019; Xu et al., 2019; Zhang et al., 2021a, 2020c; Ghiassian et al., 2020; Wang and Zou, 2020; Zhang et al., 2020a; Guan et al., 2021; Zhang et al., 2021b; Zhang and Whiteson, 2022; Qian and Zhang, 2025; Liu et al., 2025). We refer the reader to Chapter 11 of Sutton and Barto (2018) and Zhang (2022) for more detailed exposition.

Among all those works, gradient temporal difference learning (GTD, Sutton et al. (2008)) and emphatic temporal difference learning (ETD, Sutton et al. (2016)) are the two most important solutions to the deadly triad in terms of policy evaluation. GTD and ETD are also important building blocks for other algorithms. They can be used in convergent off-policy actor-critic algorithms for control, see, e.g., Imani et al. (2018); Maei (2018); Zhang et al. (2020b); Xu et al. (2021); Graves et al. (2023). They can also be used to learn value functions w.r.t. some augmented reward function to construct behavior policies for efficient and unbiased Monte Carlo policy evaluation, see, e.g., Liu and Zhang (2024); Liu et al. (2024b); Chen et al. (2024); Liu et al. (2024a). But surprisingly, the convergence analysis of their ultimate form with eligibility trace, i.e., GTD(λ) and ETD(λ), is still not fully settled down. In the next, we shall analyze GTD(λ) and ETD(λ) in the sequel. Throughout the rest of Section 5, we make the following assumptions.

Assumption 5.1 *Both $\mathcal{S}$ and $\mathcal{A}$ are finite. The Markov chain $\{S_t\}$ induced by the behavior policy μ is irreducible. And $\mu(a|s) > 0$ for all s, a.*

We note again that in light of Section 5.1, even if the MDP itself is finite, the augmented Markov chain used to analyze GTD(λ) and ETD(λ) still evolves in an unbounded and uncountable space. The analysis is, therefore, very challenging. Assumption 5.1 is a standard assumption in off-policy RL to ensure enough exploration, see, e.g., Precup et al. (2001); Sutton et al. (2016). The condition $\mu(a|s) > 0$ can be easily relaxed to $\pi(a|s) > 0 \implies \mu(a|s) > 0$, at the price of complicating the presentation.

Assumption 5.2 *The learning rates $\{\alpha_t\}$ have the form $\alpha_t = \frac{B_1}{t+B_2}$.*

Assumption 5.2 is also used in existing works, see, e.g., Yu (2012, 2015, 2017).

Assumption 5.3 *The feature matrix Φ has a full column rank.*

Assumption 5.3 is a standard assumption in RL with linear function approximation to ensure the existence and uniqueness of the solution, see, e.g., Tsitsiklis and Roy (1996).

5.3 Gradient Temporal Difference Learning

The idea of GTD is to perform stochastic gradient descent on $J_{\text{off}}(\theta)$ directly and use a weight duplication trick or Fenchel's duality to address a double sampling issue in estimating $\nabla J_{\text{off}}(\theta)$. We refer the reader to Sutton et al. (2009); Liu et al. (2015) for detailed derivation. GTD has many different variants, see, e.g., Sutton et al. (2008, 2009); Maei (2011); Yu (2017); Zhang et al. (2021a); Qian and Zhang (2025). In this paper, we present and analyze the following arguably most representative one, referred to as GTD(λ) for simplicity.[18] In particular, GTD(λ) employs an additional weight vector $\nu \in \mathbb{R}^K$ and update θ and ν simultaneously in a recursive way as

$$\begin{aligned} e_t =& \lambda\gamma\rho_{t-1}e_{t-1} + \phi_t, \\ \delta_t =& R_{t+1} + \gamma\phi_{t+1}^\top\theta_t - \phi_t^\top\theta_t, \\ \nu_{t+1} =& \nu_t + \alpha_t\left(\rho_t\delta_t e_t - \phi_t\phi_t^\top\nu_t\right), \\ \theta_{t+1} =& \theta_t + \alpha_t\rho_t(\phi_t - \gamma\phi_{t+1})e_t^\top\nu_t. \end{aligned} \tag{43}$$

This additional weight vector results from the weight duplication or Fenchel's duality. To analyze (43), we first express the update to ν and θ in a compact form as

$$\begin{bmatrix}\nu_{t+1}\\ \theta_{t+1}\end{bmatrix} = \begin{bmatrix}\nu_t\\ \theta_t\end{bmatrix} + \alpha_t\left(\begin{bmatrix} -\phi_t\phi_t^\top & \rho_t e_t(\gamma\phi_{t+1} - \phi_t)^\top \\ -(\gamma\phi_{t+1} - \phi_t)\rho_t e_t^\top & 0\end{bmatrix}\begin{bmatrix}\nu_t\\ \theta_t\end{bmatrix} + \begin{bmatrix}\rho_t R_{t+1}e_t\\ 0\end{bmatrix}\right).$$

To further simplify it, we define an augmented Markov chain $\{Y_t\}$ as

$$Y_{t+1} \doteq (S_t, A_t, S_{t+1}, e_t), \quad t = 0, 1, \dots.$$

We also define shorthands

$$\begin{aligned} x \doteq& \begin{bmatrix}\nu\\ \theta\end{bmatrix}, x_t \doteq \begin{bmatrix}\nu_t\\ \theta_t\end{bmatrix}, \\ y \doteq& (s, a, s', e), \\ A(y) \doteq& \rho(s,a)e(\gamma\phi(s') - \phi(s))^\top, \\ b(y) \doteq& \rho(s,a)r(s,a)e, \\ C(y) \doteq& \phi(s)\phi(s)^\top, \\ H(x,y) \doteq& \begin{bmatrix} -C(y) & A(y) \\ -A(y)^\top & 0\end{bmatrix}x + \begin{bmatrix}b(y)\\ 0\end{bmatrix}. \end{aligned}$$

[18] This is the GTDa in Yu (2017) and is the GTD2 in Sutton et al. (2009) with eligibility trace.

Then GTD(λ) can be expressed as

$$x_{t+1} = x_t + \alpha_t H(x_t, Y_{t+1}),$$

which reduces to the form of (1). We now proceed to prove the almost sure convergence of $\{x_t\}$ using Corollary 8. Apparently, $\{Y_t\}$ evolves in the state space

$$\mathcal{Y} \doteq \mathcal{S} \times \mathcal{A} \times \mathcal{S} \times \mathbb{R}^K.$$

Despite that both $\mathcal{S}$ and $\mathcal{A}$ are finite, $\mathcal{Y}$ can still be unbounded and uncountable. It is shown in Proposition 3.1 of Yu (2012) that as long as there is a cycle in $\{S_t\}$, e_t is unbounded almost surely in arguably almost all natural problems. Nevertheless, Yu (2012) shows that $\{Y_t\}$ has the following property.

Lemma 23 *(Theorems 3.2 & 3.3 of Yu (2012)) Let Assumption 5.1 hold. Then*

(i) $\{Y_t\}$ has a unique invariant probability measure, referred to as $d_\mathcal{Y}$.

(ii) For any matrix/vector-valued function $g(s, a, s', e)$ on $\mathcal{Y}$ which is Lipschitz continuous in e with a Lipschitz constant L_g, i.e.,

$$\left\|g(s, a, s', e) - g(s, a, s', e')\right\| \leq L_g \left\|e - e'\right\|, \quad \forall s, a, s', e, e',$$

the expectation $\mathbb{E}_{y\sim d_\mathcal{Y}}[g(y)]$ exists and is finite, and the (LLN) holds for the g function.

Yu (2012) also shows that

$$\begin{aligned} A \doteq& \mathbb{E}_{y\sim d_\mathcal{Y}}[A(y)] = \Phi^\top D_\mu (\gamma P_{\pi,\lambda} - I)\Phi, \\ b \doteq& \mathbb{E}_{y\sim d_\mathcal{Y}}[b(y)] = \Phi^\top D_\mu r_{\pi,\lambda}, \\ C \doteq& \mathbb{E}_{y\sim d_\mathcal{Y}}[C(y)] = \Phi^\top D_\mu \Phi, \end{aligned}$$

where we use D_μ to denote the diagonal matrix whose diagonal entry is d_μ.

Theorem 24 *Let Assumptions 5.1 - 5.3 hold. Assume A is nonsingular. Then the iterates $\{\theta_t\}$ generated by GTD(λ) (43) satisfy*

$$\lim_{t\to\infty} \theta_t = -A^{-1}b \quad a.s.$$

Its proof is in Appendix B.10. It can be shown easily that $-A^{-1}b$ is the unique zero of $J_{\text{off}}(\theta)$, see, e.g., Sutton et al. (2009). Notably, Theorem 24 is the first almost sure convergence analysis of GTD with eligibility trace without adding additional bias terms. Most existing convergence analyses of GTD (see, e.g., Sutton et al. (2008, 2009); Maei (2011); Liu et al. (2015); Wang et al. (2017); Qian and Zhang (2025)) do not have eligibility trace. To our knowledge, the only previous analysis of GTD with eligibility trace is Yu (2017), which, however, relies on additional projection operators or regularization to ensure the stability and unavoidably introduces bias into the final limiting point. As a result, Yu (2017) cannot establish the almost sure convergence of GTD(λ) to the unique zero of $J_{\text{off}}(\theta)$. Yu (2017) also introduces extensions to λ. Instead of being a constant, it can be a state-dependent function $\lambda : \mathcal{S} \to [0, 1]$. The almost sure convergence of GTD(λ) with a state-dependent λ function follows similarly. We present the simplest constant λ case for clarity. Yu (2017) also introduces history-dependent λ function, which we leave for future work.

5.4 Emphatic Temporal Difference Learning

The idea of ETD is to reweight the off-policy linear TD update (42) by an additional factor. Similar to GTD, ETD also has many different variants, see, e.g., Yu (2015); Sutton et al. (2016); Hallak et al. (2016); Zhang et al. (2020b); Zhang and Whiteson (2022); Guan et al. (2021). Variants of ETD have also been applied in deep RL, see, e.g., Jiang et al. (2021, 2022); Mathieu et al. (2023). In this section, we consider the original ETD(λ) in Yu (2015); Sutton et al. (2016). ETD(λ) updates θ recursively in the following way

$$\begin{aligned} F_t =& \gamma \rho_{t-1} F_{t-1} + i(S_t), \\ M_t =& \lambda i(S_t) + (1-\lambda) F_t, \\ e_t =& \lambda \gamma \rho_{t-1} e_{t-1} + M_t \phi_t, \\ \theta_{t+1} =& \theta_t + \alpha_t \rho_t \left(R_{t+1} + \gamma \phi_{t+1}^\top \theta_t - \phi_t^\top \theta_t \right) e_t, \end{aligned} \tag{44}$$

where $i : \mathcal{S} \to (0, \infty)$ is an arbitrary "interest" function (Sutton et al., 2016), specifying user's preference for different states, despite that in most applications, $i(s)$ is a constant function which is always 1. See Zhang et al. (2019) for an example where the interest function is not trivially 1. Comparing the eligibility trace e_t in (44) with that in (42), one can find that there is an additional scalar multiplier M_t proceeding ϕ_t. This M_t is called "emphasis" (Sutton et al., 2016), which is the accumulation of F_t, called "followon trace" (Sutton et al., 2016). We refer the reader to Sutton et al. (2016) for the intuition behind ETD. Nevertheless, Yu (2015) proves that, under mild conditions, $\{\theta_t\}$ in (44) converges almost surely to the unique zero of

$$J_{\text{emphatic}}(\theta) = \|\Pi_m \mathcal{T}_{\pi,\lambda} \Phi\theta - \Phi\theta\|_m^2,$$

where $m \doteq (I - \gamma P_{\pi,\lambda}^\top)^{-1} D_\mu i$. We remark that the zero of $J_{\text{emphatic}}(\theta)$ has better theoretical guarantees than the zero of $J_{\text{off}}(\theta)$ in terms of the approximation error for v_π (Hallak et al., 2016). ETD, however, usually suffers from a larger variance than GTD (Sutton and Barto, 2018).

To analyze ETD(λ), Yu (2015) considers the following augmented Markov chain

$$Y_{t+1} = (S_t, A_t, S_{t+1}, e_t, F_t).$$

Again, $\{Y_t\}$ behaves poorly in that (e_t, F_t) can be unbounded almost surely and its variance can grow to infinity as time progresses. We refer the reader to Remark A.1 in Yu (2015) for an in-depth discussion regarding this poor behavior. Nevertheless, Yu (2015) shows that $\{Y_t\}$ has the following property.

Lemma 25 *(Theorems 3.2 & 3.3 of Yu (2015)) Let Assumption 5.1 hold. Then*

(i) $\{Y_t\}$ *has a unique invariant probability measure, referred to as* $d_\mathcal{Y}$.

(ii) For any matrix / vector-valued function $g(s, a, s', e, f)$ *on* $\mathcal{Y}$ *which is Lipschitz continuous in* (e, f) *with a Lipschitz constant* L_g*, i.e.,*

$$\left\| g(s,a,s',e,f) - g(s,a,s',e',f') \right\| \leq L_g \left\| e - e' \right\|, \quad \forall s,a,s',e,e',f,f',$$

the expectation $\mathbb{E}_{y \sim d_\mathcal{Y}}[g(y)]$ *exists and is finite, and the* (LLN) *holds for the function* g.

We now discuss how Yu (2015) establishes the almost sure convergence of $\{\theta_t\}$. First, we define shorthands

$$\begin{aligned} y \doteq& (s, a, s', e, f), \\ A(y) =& \rho(s, a) e (\gamma \phi(s') - \phi(s))^\top, \\ b(y) =& \rho(s, a) r(s, a) e, \\ H(\theta, y) =& A(y)\theta + b(y). \end{aligned}$$

Then the ETD(λ) update can be expressed as

$$\theta_{t+1} = \theta_t + \alpha_t H(\theta_t, Y_{t+1}).$$

Yu (2015) also shows that

$$\begin{aligned} A \doteq& \mathbb{E}_{y \sim d_{\mathcal{Y}}} [A(y)] = \Phi^\top D_m (\gamma P_{\pi,\lambda} - I)\Phi, \\ b \doteq& \mathbb{E}_{y \sim d_{\mathcal{Y}}} [b(y)] = \Phi^\top D_m r_{\pi,\lambda}, \end{aligned}$$

and $-A^{-1}b$ is the unique zero of $J_{\text{emphatic}}(\theta)$. Despite that A is negative definite (see, e.g., Section 4 of Sutton et al. (2016)) and the corresponding ODE@∞ is, therefore, globally asymptotically stable, Yu (2015) is not able to establish the stability of $\{\theta_t\}$ directly, simply because the results in the stochastic approximation community are not ready yet. See Section 3 for a comprehensive review. As a workaround, Yu (2015) analyzes a constrained variant of ETD(λ) first:

$$\theta'_{t+1} = \Pi\left(\theta'_t + \alpha_t H(\theta'_t, Y_{t+1})\right),$$

where Π is a projection to a centered ball of properly chosen radius w.r.t. ℓ_2 norm. Yu (2015) then proves that the difference between $\{\theta_t\}$ and $\{\theta'_t\}$ diminishes almost surely and therefore establishes the convergence of $\{\theta_t\}$ indirectly. To establish the convergence of $\{\theta'_t\}$, Yu (2015) invokes Theorem 1.1 in Chapter 6 of Kushner and Yin (2003). Now with our Corollary 8, the same arguments Yu (2015) use to invoke Kushner and Yin (2003) can lead to the convergence of $\{\theta_t\}$ directly. Our contribution is, therefore, a greatly simplified almost sure convergence analysis of ETD(λ). In particular, we have

Theorem 26 *Let Assumptions 5.1 - 5.3 hold. Then the iterates $\{\theta_t\}$ generated by ETD(λ) (44) satisfy*

$$\lim_{t \to \infty} \theta_t = -A^{-1}b \quad a.s.$$

The proof of Theorem 26 is a verbatim repetition of the proof of Theorem 24 in Appendix B.10 after noticing that A is negative definite and Lemma 25 and is thus omitted. Notably, this proof does not involve the comparison between $\{\theta_t\}$ and $\{\theta'_t\}$.

We remark that the comparison technique between $\{\theta_t\}$ and $\{\theta'_t\}$ used by Yu (2015) heavily relies on the fact that A is negative definite (see Lemma 4.1 of Yu (2015)). But in GTD(λ), the corresponding matrix is $\begin{bmatrix} -C & A \\ -A^\top & 0 \end{bmatrix}$, which is Hurwitz but not negative definite. In fact, it is only negative semidefinite. As a result, the comparison technique in Yu (2015) does not apply to GTD(λ).

6. Conclusion

In this work, we develop a novel stability result of stochastic approximations, extending the celebrated Borkar-Meyn theorem from the Martingale difference noise setting to the Markovian noise setting. Our result is built on the diminishing asymptotic rate of change of a few functions, which is implied by both a form of the strong law of larger numbers and a form of the law of the iterated logarithm. We demonstrate the wide applicability of our results in RL, generating state-of-the-art analysis for important RL algorithms in breaking the notorious deadly triad. There are many possible directions for future work. One direction is to characterize the behavior of the iterates in (1) in more aspects. For example, it is possible to establish a (functional) central limit theorem following Borkar et al. (2021). It is also possible to establish an almost sure convergence rate, a high probability concentration bound, and an L^p convergence rate following Qian et al. (2024). Another direction is to weaken the required assumptions further. In the context of RL, Assumption 5 is typically obtained by assuming h is related to some contraction operator and the feature matrix Φ has a full column rank. It is possible to weaken h to nonexpansive operators following Blaser and Zhang (2024). It is also possible to allow Φ to have arbitrary ranks following Wang and Zhang (2024).

Acknowledgments

The authors thank Vagul Mahadevan for the detailed proofreading. The authors thank the anonymous reviewers for their highly constructive comments. This work is supported in part by the US National Science Foundation under grants III-2128019 and SLES-2331904.

Table of Contents

1 Introduction **1**

2 Main Results **3**

3 Related Work **7**

4 Proof of Theorem 7 **9**
- 4.1 Diminishing Asymptotic Rate of Change . . . 9
- 4.2 Equicontinuity of Scaled Iterates . . . 11
- 4.3 A Convergent Subsequence . . . 13
- 4.4 Diminishing Discretization Error . . . 14
- 4.5 Identifying Contradiction and Completing Proof . . . 16

5 Applications in Reinforcement Learning **17**
- 5.1 Eligibility Trace . . . 19
- 5.2 The Deadly Triad . . . 20
- 5.3 Gradient Temporal Difference Learning . . . 21
- 5.4 Emphatic Temporal Difference Learning . . . 23

6 Conclusion **25**

A Mathematical Background **27**

B Technical Proofs **30**
- B.1 Proof of Lemma 10 . . . 30
- B.2 Proof of Lemma 15 . . . 31
- B.3 Proof of Lemma 16 . . . 33
- B.4 Proof of Lemma 17 . . . 35
- B.5 Proof of Lemma 19 . . . 36
- B.6 Proof of Lemma 20 . . . 37
- B.7 Proof of Lemma 21 . . . 38
- B.8 Proof of Lemma 22 . . . 40
- B.9 Proof of Corollary 8 . . . 40
- B.10 Proof of Theorem 24 . . . 44

C Auxiliary Lemmas **46**

D Proofs for Completeness **61**
- D.1 Proof of Lemma 9 . . . 61
- D.2 Proof of Lemma 18 . . . 64
- D.3 Proof of Lemma 53 . . . 65
- D.4 Proof of Lemma 54 . . . 65
- D.5 Proof of Lemma 55 . . . 68

Appendix A. Mathematical Background

Theorem A.1 (Gronwall Inequality) *(Lemma 6 in Section 11.2 in Borkar (2009)) For a continuous function $u(\cdot) \geq 0$ and scalars $C, K, T \geq 0$,*

$$u(t) \leq C + K \int_0^t u(s)ds \quad \forall t \in [0, T]$$

implies

$$u(t) \leq Ce^{tK}, \forall t \in [0, T].$$

Theorem A.2 (Gronwall Inequality in the Reverse Time) *For a continuous function $u(\cdot) \geq 0$ and scalars $C, K, T \geq 0$,*

$$u(t) \leq C + K \int_t^0 u(s)ds \quad \forall t \in [-T, 0] \tag{45}$$

implies

$$u(t) \leq Ce^{-tK}, \forall t \in [-T, 0].$$

Proof $\forall s \in [0, T]$, define

$$v(s) \doteq e^{sK} K \int_s^0 u(r)dr. \tag{46}$$

Taking the derivative of $v(s)$,

$$\begin{aligned} v'(s) &= -e^{sK} K u(s) + e^{sK} K^2 \int_s^0 u(r)dr \\ &= e^{sK} K \left[-u(s) + K \int_s^0 u(r)dr \right] && \text{(by (45))} \\ &\geq -Ce^{sK} K. \end{aligned}$$

Thus,

$$v(t) = v(0) - \int_t^0 v'(s)ds \leq v(0) + \int_t^0 Ce^{sK} K ds = KC \int_t^0 e^{sK} ds.$$

By (46),

$$\begin{aligned} K \int_t^0 u(s)ds &= v(t)e^{-tK} \\ &\leq KC \int_t^0 e^{sK} ds e^{-tK} \\ &\leq KC \int_t^0 e^{(s-t)K} ds \end{aligned}$$

$$
\begin{aligned}
&= KC[\frac{1}{k}e^{(0-t)K} - \frac{1}{k}e^{(t-t)K}] \\
&= -C + Ce^{-tK}.
\end{aligned}
$$

Thus,

$$
u(t) \leq C + K\int_t^0 u(s)ds \leq Ce^{-tK}.
$$

■

Theorem A.3 (Discrete Gronwall Inequality) *(Lemma 8 in Section 11.2 in Borkar (2009)) For non-negative sequences $\{x_n, n \geq 0\}$ and $\{a_n, n \geq 0\}$ and scalars $C, L \geq 0$,*

$$
x_{n+1} \leq C + L\sum_{i=0}^{n} a_i x_i \quad \forall n
$$

implies

$$
x_{n+1} \leq Ce^{L\sum_{i=0}^{n} a_i} \quad \forall n.
$$

Theorem A.4 (The Arzela-Ascoli Theorem in the Extended Sense on $[0, T)$) *Let $\{t \in [0, T) \mapsto g_n(t)\}$ be equicontinuous in the extended sense. Then, there exists a subsequence $\{g_{n_k}(t)\}$ that converges to some continuous limit $g^{\lim}(t)$, uniformly in t on $[0, T)$.*

The proof of the Arzela-Ascoli Theorem can be found in any standard analysis textbook, see, e.g., Royden and Fitzpatrick (1968); Dunford and Schwartz (1988). The proof of the Arzela-Ascoli Theorem in the extended sense is virtually the same. The difference is that in the standard Arzela-Ascoli Theorem, one uses the compactness to find a finite subcover. But in the extended one, $[0, T)$ is not compact. However, finding a finite cover for this specific set $[0, T)$ is indeed trivial. We anyway still include the full proof below for completeness.
Proof Fix an arbitrary $\epsilon > 0$, by Definition 14, $\exists \delta > 0$ such that

$$
\limsup_n \sup_{0 \leq |t_1 - t_2| \leq \delta, 0 \leq t_1 \leq t_2 < T} \|g_n(t_1) - g_n(t_2)\| \leq \epsilon. \tag{47}
$$

This means by the definition of equicontinuity in the extended sense, when n is large enough, for any $0 \leq |t_1 - t_2| \leq \delta$, the function values $g_n(t_1)$ and $g_n(t_2)$ are also close. To conveniently utilize this property, we divide $[0, T)$ into a set of disjoint intervals and each interval has a length δ such that the t in each interval is close. In particular, we define

$$
\begin{aligned}
N &\doteq \max\{i \mid i\delta < T, i \in \mathbb{Z}\}, \\
I_i &\doteq [i\delta, (i+1)\delta), \quad i = 0, 1, \dots, N.
\end{aligned}
$$

The set of intervals $\{I_i\}_{i=0}^N$ covers the domain $[0, T)$,

$$
[0, T) \subseteq \bigcup_{i=0}^{N} I_i.
$$

We now show $g_n(t)$ is uniformly bounded uniformly on the set of dividing points $\{i\delta\}_{i=0}^N$. In particular, we have for any $i \in \{0, 1, \ldots, N\}$,

$$\begin{aligned}
&\limsup_n \|g_n(i\delta)\| \\
\leq &\limsup_n \|g_n(i\delta) - g_n((i-1)\delta)\| \\
&+ \limsup_n \|g_n((i-1)\delta) - g_n((i-2)\delta)\| \\
&+ \ldots \\
&+ \limsup_n \|g_n(\delta) - g_n(0)\| \\
&+ \limsup_n \|g_n(0)\| \\
\leq &(N+1)\epsilon + \limsup_n \|g_n(0)\| && \text{(by (47))} \\
\leq &(N+1)\epsilon + \sup_n \|g_n(0)\| \\
< &\infty. && (\sup_n \|g_n(0)\| < \infty \text{ in Definition 14})
\end{aligned}$$

This implies

$$\sup_{i \in \{0,1,\ldots,N\}, n \geq 0} \|g_n(i\delta)\| < \infty.$$

By the Bolzano-Weierstrass theorem, there exists a subsequence of functions $\left\{g_{n_{0,k}}\right\}$ in $\{g_n\}$ such that $\left\{g_{n_{0,k}}(0 \cdot \delta)\right\}$ converges. Repeating the same argument for the sequence of points $\left\{g_{n_{0,k}}(1 \cdot \delta)\right\}$, there exists a subsequence $\left\{g_{n_{1,k}}\right\}$ of $\left\{g_{n_{0,k}}\right\}$ such that $\left\{g_{n_{1,k}}(1 \cdot \delta)\right\}$ converges. Repeating this process, because N is finite, there exists a subsequence $\{g_{n_k}\}$ that converges at all dividing points $t \in \{i\delta\}_{i=0}^N$. Due to the finiteness of N, $\exists k_0$, such that $\forall i \in \{0, 1, \ldots, N\}$, $\forall k_1 \geq k_0, \forall k_2 \geq k_0$, we have

$$\left\|g_{n_{k_1}}(i\delta) - g_{n_{k_2}}(i\delta)\right\| \leq \epsilon. \tag{48}$$

By (47), $\exists k_1$ such that $\forall k \geq k_1$,

$$\sup_{0 \leq |t_1 - t_2| \leq \delta, \, 0 \leq t_1 \leq t_2 < T} \|g_{n_k}(t_1) - g_{n_k}(t_2)\| \leq 2\epsilon. \tag{49}$$

Thus, $\forall t \in [0, T)$, $\forall k \geq \max\{k_0, k_1\}, \forall k' \geq \max\{k_0, k_1\}$,

$$\begin{aligned}
&\left\|g_{n_k}(t) - g_{n_{k'}}(t)\right\| \\
\leq &\|g_{n_k}(t) - g_{n_k}(\lfloor t/\delta \rfloor \cdot \delta)\| + \left\|g_{n_k}(\lfloor t/\delta \rfloor \cdot \delta) - g_{n_{k'}}(\lfloor t/\delta \rfloor \cdot \delta)\right\| \\
&+ \left\|g_{n_{k'}}(\lfloor t/\delta \rfloor \cdot \delta) - g_{n_{k'}}(t)\right\| \\
\leq &2\epsilon + \left\|g_{n_k}(\lfloor t/\delta \rfloor \cdot \delta) - g_{n_{k'}}(\lfloor t/\delta \rfloor \cdot \delta)\right\| + 2\epsilon && \text{(by (49))} \\
\leq &2\epsilon + \epsilon + 2\epsilon && \text{(by (48))} \\
= &5\epsilon.
\end{aligned}$$

This shows that the sequence $\{g_{n_k}\}$ is uniformly Cauchy and therefore uniformly converges to a continuous function.

■

Theorem A.5 (Moore-Osgood Theorem for Interchanging Limits) *If* $\lim_{n\to\infty} a_{n,m} = b_m$ *uniformly in* m *and* $\lim_{m\to\infty} a_{n,m} = c_n$ *for each large* n*, then both* $\lim_{m\to\infty} b_m$ *and* $\lim_{n\to\infty} c_n$ *exists and are equal to the double limit, i.e.,*

$$\lim_{m\to\infty}\lim_{n\to\infty} a_{n,m} = \lim_{n\to\infty}\lim_{m\to\infty} a_{n,m} = \lim_{\substack{n\to\infty\\ m\to\infty}} a_{n,m}.$$

Appendix B. Technical Proofs

B.1 Proof of Lemma 10

Proof Let Assumptions 1, 2, 3, and 4 hold. Fix an arbitrary sample path $\{x_0, \{Y_i\}_{i=1}^\infty\}$. Use $\mathcal{B}$ to denote an arbitrary compact set of x.

$$\begin{aligned}
&\lim_{c\to\infty}\sup_{x\in\mathcal{B}}\sup_n\sup_{t\in[0,T]}\left\|\sum_{i=m(T_n)}^{m(T_n+t)-1}\alpha(i)\left[H_c(x,Y_{i+1}) - H_\infty(x,Y_{i+1})\right]\right\| \\
=&\lim_{c\to\infty}\sup_{x\in\mathcal{B}}\sup_n\sup_{t\in[0,T]}\left\|\sum_{i=m(T_n)}^{m(T_n+t)-1}\alpha(i)\kappa(c)b(x,Y_{i+1})\right\| && \text{(by (5))}\\
=&\lim_{c\to\infty}\kappa(c)\sup_{x\in\mathcal{B}}\sup_n\sup_{t\in[0,T]}\left\|\sum_{i=m(T_n)}^{m(T_n+t)-1}\alpha(i)b(x,Y_{i+1})\right\| \\
=&0\sup_{x\in\mathcal{B}}\sup_n\sup_{t\in[0,T]}\left\|\sum_{i=m(T_n)}^{m(T_n+t)-1}\alpha(i)b(x,Y_{i+1})\right\| && (50)
\end{aligned}$$

We now show that the function

$$x \mapsto \sup_n\sup_{t\in[0,T]}\left\|\sum_{i=m(T_n)}^{m(T_n+t)-1}\alpha(i)b(x,Y_{i+1})\right\| \tag{51}$$

is Lipschitz continuous. $\forall x, x'$,

$$\begin{aligned}
&\left|\sup_n\sup_{t\in[0,T]}\left\|\sum_{i=m(T_n)}^{m(T_n+t)-1}\alpha(i)b(x,Y_{i+1})\right\| - \sup_n\sup_{t\in[0,T]}\left\|\sum_{i=m(T_n)}^{m(T_n+t)-1}\alpha(i)b(x',Y_{i+1})\right\|\right| \\
\leq&\sup_n\sup_{t\in[0,T]}\left|\left\|\sum_{i=m(T_n)}^{m(T_n+t)-1}\alpha(i)b(x,Y_{i+1})\right\| - \left\|\sum_{i=m(T_n)}^{m(T_n+t)-1}\alpha(i)b(x',Y_{i+1})\right\|\right| \\
&\text{(by } |\sup_x f(x) - \sup_x g(x)| \leq \sup_x |f(x) - g(x)|)
\end{aligned}$$

$$\leq \sup_n \sup_{t\in[0,T]} \left\| \sum_{i=m(T_n)}^{m(T_n+t)-1} \alpha(i) b(x, Y_{i+1}) - \sum_{i=m(T_n)}^{m(T_n+t)-1} \alpha(i) b(x', Y_{i+1}) \right\|$$

$$\leq \sup_n \sup_{t\in[0,T]} \sum_{i=m(T_n)}^{m(T_n+t)-1} \alpha(i) \left\| b(x, Y_{i+1}) - b(x', Y_{i+1}) \right\|$$

$$\leq \sup_n \sup_{t\in[0,T]} \left(\sum_{i=m(T_n)}^{m(T_n+t)-1} \alpha(i) L_b(Y_{i+1}) \right) \left\| x - x' \right\| \qquad \text{(by (6))}$$

Additionally, let Assumption 6 or 6′ hold. By Lemma 9 and (93),

$$\sup_n \sup_{t\in[0,T]} \left(\sum_{i=m(T_n)}^{m(T_n+t)-1} \alpha(i) L_b(Y_{i+1}) \right) < \infty$$

can be viewed as the Lipschitz constant. Thus, (51) is a continuous function. Since $\mathcal{B}$ is compact, the extreme value theorems asserts that the supremum of (51) in $\mathcal{B}$ is attainable at some $x_{\mathcal{B}}$ and is finite. This means the RHS of (50) is 0,

$$\lim_{c\to\infty} \sup_{x\in\mathcal{B}} \sup_n \sup_{t\in[0,T]} \left\| \sum_{i=m(T_n)}^{m(T_n+t)-1} \alpha(i) \left[H_c(x, Y_{i+1}) - H_\infty(x, Y_{i+1}) \right] \right\| = 0.$$

■

B.2 Proof of Lemma 15

Proof By (24),

$$\sup_n \| \hat{x}(T_n + 0) \| \leq 1.$$

$\forall \xi > 0$, by (88), $\exists \delta_0$, such that $\forall 0 < \delta \leq \delta_0$,

$$\sup_{c\geq 1} \limsup_n \sup_{0\leq t_2 - t_1 \leq \delta} \left\| \sum_{i=m(T_n+t_1)}^{m(T_n+t_2)-1} \alpha(i) H_c(0, Y_{i+1}) \right\| \leq \xi. \tag{52}$$

By (92), $\exists \delta_1$, such that $\forall 0 < \delta \leq \delta_1$,

$$\limsup_n \sup_{0\leq t_2 - t_1 \leq \delta} \sum_{i=m(T_n+t_1)}^{m(T_n+t_2)-1} \alpha(i) L(Y_{i+1}) \leq \xi. \tag{53}$$

Without loss of generality, let $t_1 \leq t_2$. Then$\forall \delta \leq \min\{\delta_0, \delta_1\}$, we have

$$\limsup_n \sup_{0\leq t_2 - t_1 \leq \delta} \| \hat{x}(T_n + t_1) - \hat{x}(T_n + t_2) \|$$

$$
\begin{aligned}
&= \limsup_n \sup_{0\le t_2-t_1\le\delta} \left\| \sum_{i=m(T_n+t_1)}^{m(T_n+t_2)-1} \alpha(i) H_{r_n}(\hat{x}(t(i)), Y_{i+1}) \right\| \\
&\le \limsup_n \sup_{0\le t_2-t_1\le\delta} \left\| \sum_{i=m(T_n+t_1)}^{m(T_n+t_2)-1} \alpha(i) H_{r_n}(\hat{x}(t(i)), Y_{i+1}) \right\| - \left\| \sum_{i=m(T_n+t_1)}^{m(T_n+t_2)-1} \alpha(i) H_{r_n}(0, Y_{i+1}) \right\| \\
&\quad + \limsup_n \sup_{0\le t_2-t_1\le\delta} \left\| \sum_{i=m(T_n+t_1)}^{m(T_n+t_2)-1} \alpha(i) H_{r_n}(0, Y_{i+1}) \right\| \\
&\le \limsup_n \sup_{0\le t_2-t_1\le\delta} \left\| \sum_{i=m(T_n+t_1)}^{m(T_n+t_2)-1} \alpha(i) H_{r_n}(\hat{x}(t(i)), Y_{i+1}) \right\| - \left\| \sum_{i=m(T_n+t_1)}^{m(T_n+t_2)-1} \alpha(i) H_{r_n}(0, Y_{i+1}) \right\| \\
&\quad + \sup_{c\ge 1} \limsup_n \sup_{0\le t_2-t_1\le\delta} \left\| \sum_{i=m(T_n+t_1)}^{m(T_n+t_2)-1} \alpha(i) H_{c}(0, Y_{i+1}) \right\| \\
&\le \limsup_n \sup_{0\le t_2-t_1\le\delta} \left\| \sum_{i=m(T_n+t_1)}^{m(T_n+t_2)-1} \alpha(i) H_{r_n}(\hat{x}(t(i)), Y_{i+1}) \right\| - \left\| \sum_{i=m(T_n+t_1)}^{m(T_n+t_2)-1} \alpha(i) H_{r_n}(0, Y_{i+1}) \right\| \\
&\quad + \xi && \text{(by (52))} \\
&\le \limsup_n \sup_{0\le t_2-t_1\le\delta} \left\| \sum_{i=m(T_n+t_1)}^{m(T_n+t_2)-1} \alpha(i) H_{r_n}(\hat{x}(t(i)), Y_{i+1}) - \sum_{i=m(T_n+t_1)}^{m(T_n+t_2)-1} \alpha(i) H_{r_n}(0, Y_{i+1}) \right\| \\
&\quad + \xi \\
&\le \limsup_n \sup_{0\le t_2-t_1\le\delta} \sum_{i=m(T_n+t_1)}^{m(T_n+t_2)-1} \alpha(i) \| H_{r_n}(\hat{x}(t(i)), Y_{i+1}) - H_{r_n}(0, Y_{i+1}) \| + \xi \\
&\le \limsup_n \sup_{0\le t_2-t_1\le\delta} \sum_{i=m(T_n+t_1)}^{m(T_n+t_2)-1} \alpha(i) L(Y_{i+1}) \|\hat{x}(t(i))\| + \xi \\
&\le C_{\hat{x}} \limsup_n \sup_{0\le t_2-t_1\le\delta} \sum_{i=m(T_n+t_1)}^{m(T_n+t_2)-1} \alpha(i) L(Y_{i+1}) + \xi && \text{(by Lemma 40)} \\
&\le C_{\hat{x}} \xi + \xi, && \text{(by (53))}
\end{aligned}
$$

which implies that $\{\hat{x}(T_n + t)\}$ is equicontinuous in the extended sense.
For $\{z_n(t)\}$, by (24) and (26), we have

$$\sup_n \|z_n(0)\| \le 1.$$

Without loss of generality, let $t_1 \le t_2$. Then $\forall \delta > 0$, we have

$$
\begin{aligned}
&\sup_n \sup_{0\le |t_1-t_2|\le\delta, 0\le t_1\le t_2 < T} \|z_n(t_1) - z_n(t_2)\| \\
= &\sup_n \sup_{0\le |t_1-t_2|\le\delta, 0\le t_1\le t_2 < T} \left\| \int_{t_1}^{t_2} h_{r_n}(z_n(s)) ds \right\|
\end{aligned}
$$

$$
\begin{aligned}
&= \sup_n \sup_{0 \le |t_1 - t_2| \le \delta, 0 \le t_1 \le t_2 < T} \left\| \int_{t_1}^{t_2} \left[h_{r_n}(z_n(s)) - h_{r_n}(0) \right] ds + \int_{t_1}^{t_2} h_{r_n}(0) ds \right\| \\
&\le \sup_n \sup_{0 \le |t_1 - t_2| \le \delta, 0 \le t_1 \le t_2 < T} \int_{t_1}^{t_2} \| h_{r_n}(z_n(s)) - h_{r_n}(0) \| ds + \sup_n \sup_{0 \le |t_1 - t_2| \le \delta, 0 \le t_1 \le t_2 < T} \int_{t_1}^{t_2} \| h_{r_n}(0) \| ds \\
&\le \sup_n \sup_{0 \le |t_1 - t_2| \le \delta, 0 \le t_1 \le t_2 < T} \int_{t_1}^{t_2} L \| z_n(s) \| ds + \sup_n \sup_{0 \le |t_1 - t_2| \le \delta, 0 \le t_1 \le t_2 < T} \int_{t_1}^{t_2} \| h_{r_n}(0) \| ds \\
& \hspace{28em} \text{(by Lemma 36)} \\
&\le \delta L C_{\hat{x}} + \sup_n \sup_{0 \le |t_1 - t_2| \le \delta, 0 \le t_1 \le t_2 < T} \int_{t_1}^{t_2} \| h_{r_n}(0) \| ds \qquad \text{(by Lemma 41)} \\
&\le \delta (L C_{\hat{x}} + C_H), \qquad \text{(by (95))}
\end{aligned}
$$

which implies that $\{z_n(t)\}$ is equicontinuous.
For $\{f_n(t)\}$, we have

$$
\sup_n f_n(0) = \sup_n \hat{x}(T_n) - z_n(0) = \sup_n \hat{x}(T_n) - \hat{x}(T_n) = 0 < \infty.
$$

Because $\{\hat{x}(T_n + t)\}$ and $\{z_n(t)\}$ are equicontinuous, $\forall \epsilon > 0$, $\exists \delta$ such that

$$
\begin{aligned}
\limsup_n \sup_{0 \le t_2 - t_1 \le \delta} \| \hat{x}(T_n + t_1) - \hat{x}(T_n + t_2) \| \le \frac{\epsilon}{2}, \\
\sup_n \sup_{0 \le t_2 - t_1 \le \delta} \| z_n(t_1) - z_n(t_2) \| \le \frac{\epsilon}{2}.
\end{aligned}
$$

Without loss of generality let $t_1 \le t_2$. Then $\forall \epsilon$, $\exists \delta$ such that

$$
\begin{aligned}
& \limsup_n \sup_{0 \le t_2 - t_1 \le \delta} \| f_n(t_1) - f_n(t_2) \| \\
=& \limsup_n \sup_{0 \le t_2 - t_1 \le \delta} \| \hat{x}(T_n + t_1) - \hat{x}(T_n + t_2) - (z_n(t_1) - z_n(t_2)) \| \\
\le& \limsup_n \sup_{0 \le t_2 - t_1 \le \delta} \| \hat{x}(T_n + t_1) - \hat{x}(T_n + t_2) \| + \limsup_n \sup_{0 \le t_2 - t_1 \le \delta} \| z_n(t_1) - z_n(t_2) \| \\
\le& \limsup_n \sup_{0 \le t_2 - t_1 \le \delta} \| \hat{x}(T_n + t_1) - \hat{x}(T_n + t_2) \| + \sup_n \sup_{0 \le t_2 - t_1 \le \delta} \| z_n(t_1) - z_n(t_2) \| \\
\le& \epsilon,
\end{aligned}
$$

which implies that $\{f_n\}$ is equicontinuous in the extended sense. ■

B.3 Proof of Lemma 16

Proof We can construct a subsequence $\left\{ r_{n_{1,k}} \right\}$ that diverges to infinity and satisfies $\forall k$, $\forall n < n_{1,k}$,

$$
r_n < r_{n_{1,k}}. \tag{54}
$$

For example, we can define

$$\begin{aligned} & n_{1,0} \doteq 1 \\ & n_{1,k} \doteq \min \left\{ n \mid n > n_{1,k-1}, r_n > r_{n_{1,k-1}} + 1 \right\}. \end{aligned} \tag{55}$$

Because $\limsup_n r_n = \infty$, we know $\forall k > 0, \left\{ n \mid n > n_{1,k-1}, r_n > r_{n_{1,k-1}} + 1 \right\} \neq \emptyset$. Because $\forall k > 0,\ r_{n_{1,k}} - r_{n_{1,k-1}} > 1$,

$$\lim_{k \to \infty} r_{n_{1,k}} = \infty. \tag{56}$$

Because (55) defines $n_{1,k}$ to be the first index that is large enough after $n_{1,k-1}$, (54) holds. Otherwise $n_{1,k}$ would not be the first. Define a sequence $\{n_{2,k}\}$ as

$$n_{2,k} \doteq n_{1,k} - 1 \quad \forall k. \tag{57}$$

We make two observations. First, $n_{2,k}$ and $n_{1,k}$ are neighbors so $r_{n_{2,k}}$ and $r_{n_{1,k}}$ correspond to $\bar{x}(T_n)$ and $\bar{x}(T_{n+1})$ for some n. Second, by Lemma 42, the increment of $\bar{x}(t)$ in $[T_n, T_{n+1})$ is bounded in the following sense $\forall n$,

$$\|\bar{x}(T_{n+1})\| \leq (\|\bar{x}(T_n)\| C_H + C_H)\, e^{C_H} + \|\bar{x}(T_n)\|$$

where C_H is a positive constant. This means that if $r_{n_{2,k}}$ is not large enough, $r_{n_{1,k}}$ will not be large enough either. We can then prove by contradiction in Lemma 43 that

$$\limsup_k r_{n_{2,k}} = \infty.$$

Thus, using the similar method as (55), we can construct a subsequence $\{n_{3,k}\}$ from $\{n_{2,k}\}$ such that

$$\lim_k r_{n_{3,k}} = \infty.$$

Moreover, since $\{n_{3,k} + 1\}$ is a subsequence of $\{n_{1,k}\}$, (54) implies that

$$r_{n_{3,k}} < r_{n_{3,k}+1}.$$

Since $\{f_n\}$ is equicontinuous in the extended sense, $\left\{ f_{n_{3,k}} \right\}_{k=0,1,\ldots}$ is also equicontinuous in the extended sense. By the Arzela-Ascoli Theorem (Theorem A.4), it has a uniformly convergent subsequence, referred to as $\left\{ f_{n_{4,k}} \right\}$. Because the sequence $\left\{ \hat{x}(T_{n_{4,k}} + t) \right\}$ is also equicontinuous in the extended sense, it has a uniformly convergent subsequence $\{\hat{x}(T_{n_k} + t)\}$. To summarize,

$$\{n_k\} \subseteq \{n_{4,k}\} \subseteq \{n_{3,k}\} \subseteq \{n_{2,k}\} \subseteq \{n_{1,k} - 1\} \subseteq \mathbb{N}. \tag{58}$$

We construct $\{n_k\}$ in this way because it then inherits all uniform convergence properties. Precisely speaking, by the Arzela-Ascoli theorem in Appendix A.4, we have the following corollary.

Corollary 27 *There exist some continuous functions $f^{\lim}(t)$ and $\hat{x}^{\lim}(t)$ such that $\forall t \in [0, T)$,*

$$\begin{aligned}\lim_{k\to\infty} f_{n_k}(t) =& f^{\lim}(t),\\ \lim_{k\to\infty} \hat{x}(T_{n_k}+t) =& \hat{x}^{\lim}(t).\end{aligned}$$

Moreover, the convergence is uniform in t on $[0, T)$.

In terms of the three sequences of functions in (29), Corollary 27 has identified that two of them converge along $\{n_k\}$. Lemma 47 further confirms that $z^{\lim}$ is the limit of $\{z_{n_k}\}$. That is $\forall t \in [0, T)$,

$$\lim_{k\to\infty} z_{n_k}(t) = z^{\lim}(t).$$

Moreover, the convergence is uniform in t on $[0, T)$. By (58), we have

$$\begin{aligned}\lim_{k\to\infty} r_{n_k} =& \infty,\\ \lim_{k\to\infty} r_{n_k+1} =& \infty,\end{aligned} \tag{59}$$

which completes the proof. ■

B.4 Proof of Lemma 17

Proof $\forall j$, $\forall k$, $\forall t \in [0, T)$,

$$\begin{aligned}&\Bigg|\left\|\sum_{i=m(T_{n_k})}^{m(T_{n_k}+t)-1} \alpha(i) H_{r_{n_j}}(\hat{x}(t(i)), Y_{i+1}) - \int_0^t h_{r_{n_j}}(\hat{x}^{\lim}(s))ds\right\|\\ &-\left\|\sum_{i=m(T_{n_k})}^{m(T_{n_k}+t)-1} \alpha(i) H_\infty(\hat{x}(t(i)), Y_{i+1}) - \int_0^t h_\infty(\hat{x}^{\lim}(s))ds\right\|\Bigg|\\ \le&\Bigg\|\sum_{i=m(T_{n_k})}^{m(T_{n_k}+t)-1} \alpha(i) H_{r_{n_j}}(\hat{x}(t(i)), Y_{i+1}) - \int_0^t h_{r_{n_j}}(\hat{x}^{\lim}(s))ds\\ &-\sum_{i=m(T_{n_k})}^{m(T_{n_k}+t)-1} \alpha(i) H_\infty(\hat{x}(t(i)), Y_{i+1}) + \int_0^t h_\infty(\hat{x}^{\lim}(s))ds\Bigg\| \qquad (\text{by } |\|a\| - \|b\|| \le \|a-b\|)\\ \le&\left\|\sum_{i=m(T_{n_k})}^{m(T_{n_k}+t)-1} \alpha(i)(H_{r_{n_j}}(\hat{x}(t(i)), Y_{i+1}) - H_\infty(\hat{x}(t(i)), Y_{i+1}))\right\| + \left\|\int_0^t h_{r_{n_j}}(\hat{x}^{\lim}(s)) - h_\infty(\hat{x}^{\lim}(s))ds\right\|\end{aligned}$$

$$\leq \left\| \sum_{i=m(T_{n_k})}^{m(T_{n_k}+t)-1} \alpha(i)(H_{r_{n_j}}(\hat{x}(t(i)), Y_{i+1}) - H_\infty(\hat{x}(t(i)), Y_{i+1})) \right\| + \int_0^t \left\| h_{r_{n_j}}(\hat{x}^{\lim}(s)) - h_\infty(\hat{x}^{\lim}(s)) \right\| ds \tag{60}$$

By Lemma 40, $\hat{x}(t(i))$ is in a compact set $\mathcal{B}_{\hat{x}}$. By Lemma 10, for the compact set $\mathcal{B}_{\hat{x}}$, $\forall \epsilon > 0$, $\exists j_1$ such that $\forall j \geq j_1$, $\forall k$, $\forall x \in \mathcal{B}$, $\forall t \in [0, T)$,

$$\left\| \sum_{i=m(T_{n_k})}^{m(T_{n_k}+t)-1} \alpha(i) \left[H_{r_{n_j}}(x, Y_{i+1}) - H_\infty(x, Y_{i+1}) \right] \right\| \leq \epsilon. \tag{61}$$

Similar to the proof of Lemma 46, we have

$$\lim_{j\to\infty} h_{r_{n_j}}(\hat{x}(T_k + t)) = h_\infty(\hat{x}(T_k + t)) \tag{62}$$

uniformly in k and $t \in [0, T)$. By (62), $\forall \epsilon > 0$, $\exists j_2$ such that $\forall j > j_2$, $\forall k$, $\forall t \in [0, T)$,

$$\left\| h_{r_{n_j}}(\hat{x}(T_k + t)) - h_\infty(\hat{x}(T_k + t)) \right\| \leq \epsilon. \tag{63}$$

Define $j_0 \doteq \max\{j_1, j_2\}$. $\forall j \geq j_0$, $\forall k$, $\forall t \in [0, T)$,

$$\begin{aligned}
&\Bigg| \left\| \sum_{i=m(T_{n_k})}^{m(T_{n_k}+t)-1} \alpha(i) H_{r_{n_j}}(\hat{x}(t(i)), Y_{i+1}) - \int_0^t h_{r_{n_j}}(\hat{x}^{\lim}(s)) ds \right\| \\
&- \left\| \sum_{i=m(T_{n_k})}^{m(T_{n_k}+t)-1} \alpha(i) H_\infty(\hat{x}(t(i)), Y_{i+1}) - \int_0^t h_\infty(\hat{x}^{\lim}(s)) ds \right\| \Bigg| \\
\leq& \left\| \sum_{i=m(T_{n_k})}^{m(T_{n_k}+t)-1} \alpha(i)(H_{r_{n_j}}(\hat{x}(t(i)), Y_{i+1}) - H_\infty(\hat{x}(t(i)), Y_{i+1})) \right\| + T\epsilon && \text{(by (60), (63))} \\
\leq& \epsilon + T\epsilon && \text{(by (60), (61))} \\
\leq& (T + 1)\epsilon.
\end{aligned}$$

This completes the proof of uniform convergence.

■

B.5 Proof of Lemma 19

Proof

$$\lim_{\substack{j\to\infty \\ k\to\infty}} \left\| \sum_{i=m(T_{n_k})}^{m(T_{n_k}+t)-1} \alpha(i) H_{r_{n_j}}(\hat{x}(t(i)), Y_{i+1}) - \int_0^t h_{r_{n_j}}(\hat{x}^{\lim}(s)) ds \right\|$$

$$= \lim_{j\to\infty}\lim_{k\to\infty}\left\|\sum_{i=m(T_{n_k})}^{m(T_{n_k}+t)-1}\alpha(i)H_{r_{n_j}}(\hat{x}(t(i)),Y_{i+1}) - \int_0^t h_{r_{n_j}}(\hat{x}^{\lim}(s))ds\right\|$$

(by Lemma 17, 18, and Moore-Osgood Theorem for interchanging limits in Theorem A.5)

$$= \lim_{j\to\infty} 0 \qquad \text{(by Lemma 18)}$$

$$=0. \tag{64}$$

■

B.6 Proof of Lemma 20

Proof We now proceed to investigate the property of $f_{n_k}(t)$. $\forall t \in [0,T)$,

$$\begin{aligned}
&\lim_{k\to\infty}\|f_{n_k}(t)\| \\
\leq &\lim_{k\to\infty}\left\|\sum_{i=m(T_{n_k})}^{m(T_{n_k}+t)-1}\alpha(i)H_{r_{n_k}}(\hat{x}(t(i)),Y_{i+1}) - \int_0^t h_{r_{n_k}}(\hat{x}^{\lim}(s))ds\right\| \\
&+ \lim_{k\to\infty}\left\|\int_0^t h_{r_{n_k}}(\hat{x}^{\lim}(s))ds - \int_0^t h_{r_{n_k}}(z_{n_k}(s))ds\right\| && \text{(by (35))} \\
= &\lim_{k\to\infty}\left\|\int_0^t h_{r_{n_k}}(\hat{x}^{\lim}(s))ds - \int_0^t h_{r_{n_k}}(z_{n_k}(s))ds\right\| && \text{(by (64))} \\
= &\left\|\int_0^t h_\infty(\hat{x}^{\lim}(s))ds - \int_0^t h_\infty(z^{\lim}(s))ds\right\|. && \text{(by Lemma 49 and Lemma 50)} \quad (65)
\end{aligned}$$

We now show the relationship between $\hat{x}^{\lim}(t)$ and $z^{\lim}(t)$.

$$\begin{aligned}
&\left\|\hat{x}^{\lim}(t) - z^{\lim}(t)\right\| && (66) \\
= &\left\|\lim_{k\to\infty}\left[\hat{x}(T_{n_k}) + \int_0^t h_{r_{n_k}}(\hat{x}^{\lim}(s))ds\right] - \left[\hat{x}^{\lim}(0) + \int_0^t h_\infty(z^{\lim}(s))ds\right]\right\| \\
& && \text{(by (34) and (38))} \\
= &\left\|\hat{x}^{\lim}(0) + \int_0^t h_\infty(\hat{x}^{\lim}(s))ds - \left[\hat{x}^{\lim}(0) + \int_0^t h_\infty(z^{\lim}(s))ds\right]\right\| && \text{(by Lemma 49)} \\
= &\left\|\int_0^t h_\infty(\hat{x}^{\lim}(s))ds - \int_0^t h_\infty(z^{\lim}(s))ds\right\| && (67) \\
\leq &\int_0^t L\left\|\hat{x}^{\lim}(s) - z^{\lim}(s)\right\|ds && \text{(by Lemma 36)} \\
\leq &0. && \text{(by Gronwall inequality in Theorem A.1)}
\end{aligned}$$

Thus,

$$\left\|\lim_{k\to\infty} f_{n_k}(t)\right\|$$

$$\leq \left\| \int_0^t h_\infty(\hat{x}^{\lim}(s))ds - \int_0^t h_\infty(z^{\lim}(s))ds \right\| \qquad \text{(by (65))}$$
$$= \left\| \hat{x}^{\lim}(t) - z^{\lim}(t) \right\| \qquad \text{(by (67))}$$
$$\leq 0. \qquad \text{(by (66))}$$

■

B.7 Proof of Lemma 21

Proof According to (25), to study $\{z_{n_k}(t)\}$, it is instrumental to study the following ODE

$$\frac{\mathrm{d}\phi_c(t)}{\mathrm{d}t} = h_c(\phi_c(t))$$

for some $c \geq 1$. Let $\phi_{c,x}(t)$ denote the unique solution of the ODE above with the initial condition $\phi_{c,x}(0) = x$. Intuitively, as $c \to \infty$, the above ODE approaches the (ODE@∞). Since any trajectory of (ODE@∞) will diminish to 0 (Assumption 5), $\phi_{c,x}(t)$ should also diminish to some extent for sufficiently large c. Precisely speaking, we have the following lemma.

Lemma 28 *(Corollary 3.3 in Borkar (2009)) There exist $c_1 > 0$ and $\tau > 0$ such that for all initial conditions x with $\|x\| \leq 1$, we have*

$$\|\phi_{c,x}(t)\| \leq \frac{1}{4}$$

for $t \in [\tau, \tau + 1]$ and $c \geq c_1$.

Here the $\frac{1}{4}$ is entirely arbitrary. Now we fix any $c_0 \geq \max\{c_1, 1\}$ and set $T = \tau$. Then Lemma 28 confirms that $z_{n_k}(t)$ will diminish to some extent as t approaches T for sufficiently large k, so does $\hat{x}(T_{n_k} + t)$. We, however, recall that $\hat{x}(T_{n_k} + t)$ and $\bar{x}(T_{n_k} + t)$ are well defined on $[0, T_{n+1} - T_n)$ and we restrict them to $[0, T)$ for applying the Arzela-Ascoli theorem. Lemma 52 processes the excess part $[T, T_{n+1} - T_n)$, by showing that $\bar{x}(T_{n_k} + t)$ cannot grow too much in the excess part. By Lemma 52,

$$\lim_{k \to \infty} \frac{\|\bar{x}(T_{n_k+1})\| - \lim_{t \to T^-} \|\bar{x}(T_{n_k} + t)\|}{\|\bar{x}(T_{n_k})\|} = 0. \tag{68}$$

We are now in the position to identify the contradiction. By (59), $\exists k_1$ such that $\forall k \geq k_1$,

$$r_{n_k+1} > (c_0 C_H + C_H)\, e^{C_H} + c_0 > c_0 > 1. \tag{69}$$

By Lemma 20, $\exists k_2$ such that $\forall k \geq k_2$,

$$\lim_{t \to T^-} \|f_{n_k}(t)\| = \lim_{t \to T^-} \|\hat{x}(T_{n_k} + t) - z_{n_k}(t)\| \leq \frac{1}{4}. \tag{70}$$

By (68), $\exists k_3$ such that $\forall k \geq k_3$,

$$\frac{\|\bar{x}(T_{n_k+1})\| - \lim_{t \to T^-} \|\bar{x}(T_{n_k} + t)\|}{\|\bar{x}(T_{n_k})\|} \leq \frac{1}{4}. \tag{71}$$

By (59), $\exists k_4$ such that $\forall k \geq k_4$,

$$r_{n_k} > c_0.$$

Define $k_0 \doteq \max\{k_1, k_2, k_3, k_4\}$. Because $r_{n_{k_0}} > c_0$, by Lemma 28 and (25), we have

$$\lim_{t \to T^-} \left\| z_{n_{k_0}}(t) \right\| \leq \frac{1}{4}. \tag{72}$$

We have

$$\begin{aligned}
&\lim_{t \to T^-} \left\| \hat{x}(T_{n_{k_0}} + t) \right\| \\
\leq &\lim_{t \to T^-} \left\| \hat{x}(T_{n_{k_0}} + t) - z_{n_{k_0}}(t) \right\| + \left\| z_{n_{k_0}}(t) \right\| \\
\leq &\frac{1}{2}. && \text{(by (70) and (72))}
\end{aligned} \tag{73}$$

This implies

$$\begin{aligned}
&\frac{\left\| \bar{x}(T_{n_{k_0}+1}) \right\|}{\left\| \bar{x}(T_{n_{k_0}}) \right\|} \\
= &\frac{\left\| \bar{x}(T_{n_{k_0}+1}) \right\| - \lim_{t \to T^-} \left\| \bar{x}(T_{n_{k_0}} + t) \right\|}{\left\| \bar{x}(T_{n_{k_0}}) \right\|} + \frac{\lim_{t \to T^-} \left\| \bar{x}(T_{n_{k_0}} + t) \right\|}{\left\| \bar{x}(T_{n_{k_0}}) \right\|} \\
\leq &\frac{1}{4} + \frac{\lim_{t \to T^-} \left\| \bar{x}(T_{n_{k_0}} + t) \right\|}{\left\| \bar{x}(T_{n_{k_0}}) \right\|} && \text{(by (71))} \\
= &\frac{1}{4} + \frac{\lim_{t \to T^-} \left\| \hat{x}(T_{n_{k_0}} + t) \right\|}{\left\| \hat{x}(T_{n_{k_0}}) \right\|} && \text{(by (22))} \\
= &\frac{1}{4} + \lim_{t \to T^-} \left\| \hat{x}(T_{n_{k_0}} + t) \right\| && \left(\left\| \hat{x}(T_{n_{k_0}}) \right\| = 1 \text{ because of } r_{n_{k_0}} > c_0 > 1 \text{ and (22)}\right) \\
\leq &\frac{3}{4}. && \text{(by (73))}
\end{aligned} \tag{74}$$

Now, we can derive the following inequality.

$$\begin{aligned}
r_{n_{k_0}+1} &= \left\| \bar{x}(T_{n_{k_0}+1}) \right\| && \text{(by (69))} \\
&\leq \frac{3}{4} \left\| \bar{x}(T_{n_{k_0}}) \right\| && \text{(by (74))}
\end{aligned}$$

$$
\begin{aligned}
&\leq \left\|\bar{x}(T_{n_{k_0}})\right\| \\
&\leq r_{n_{k_0}}, && \text{(by } r_{n_{k_0}} > c_0 > 1 \text{ and (23))}
\end{aligned}
$$

which completes the proof. ■

B.8 Proof of Lemma 22

Proof

$$
\begin{aligned}
&\sup_n \sup_{i \in \{i | m(T_n) \leq m(T_n)+i < m(T_{n+1})\}} \left\|x_{m(T_n)+i}\right\| - \left\|x_{m(T_n)}\right\| \\
\leq &\sup_n \sup_{i \in \{i | m(T_n) \leq m(T_n)+i < m(T_{n+1})\}} \left\|x_{m(T_n)+i} - x_{m(T_n)}\right\| \\
= &\sup_n \sup_{i \in \{i | m(T_n) \leq m(T_n)+i < m(T_{n+1})\}} \left\|\bar{x}(t(m(T_n)+i)) - \bar{x}(T_n)\right\| \\
= &\sup_n \sup_{t \in [T_n, T_{n+1})} \left\|\bar{x}(T_n + t) - \bar{x}(T_n)\right\| && \text{(by (21))} \\
\leq &\sup_n \sup_{t \in [T_n, T_{n+1})} \left[\|\bar{x}(T_n)\| C_H + C_H\right] e^{C_H} && \text{(by (99))} \\
\leq &\sup_n \sup_{t \in [T_n, T_{n+1})} \left[r_n C_H + C_H\right] e^{C_H} && \text{(by (23))} \\
= &\sup_n [r_n C_H + C_H] e^{C_H} \\
< &\infty. && \text{(by (39))}
\end{aligned}
$$

■

B.9 Proof of Corollary 8

This proof follows the idea of the proof of Theorem 2.1 in Chapter 5 of Kushner and Yin (2003).

Proof Let Assumptions 1 - 5 hold. Let Assumption 6 or 6′ hold. To prove convergence results on $t \in (-\infty, \infty)$ in Corollary 8, we fix an arbitrary sample path $\{x_0, \{Y_i\}_{i=1}^{\infty}\}$. The stability results from Theorem 7 hold. To prove properties on $t \in (-\infty, \infty)$, we first fix an arbitrary $\tau > 0$ and show properties on $\forall t \in [-\tau, \tau]$.

Definition 29 *$\forall n \in \mathbb{N}$, define $\bar{z}_n(t)$ as the solution to the ODE (13) in $(-\infty, \infty)$ with an initial condition*

$$
\bar{z}_n(0) = \bar{x}(t(n)).
$$

Apparently, $\bar{z}_n(t)$ can also be written as

$$
\bar{z}_n(t) = \bar{x}(t(n)) + \int_0^t h(\bar{z}_n(s)) ds, \quad \forall t \in (-\infty, \infty). \tag{75}
$$

The major difference between the $\{\bar{z}_n(t)\}$ here and the $\{z_n(t)\}$ in (25) is that all $\{\bar{z}_n(t)\}$ here are solutions to one same ODE (13), just with different initial conditions, but $\{z_n(t)\}$ is for different ODEs with different initial conditions and rescale factors r_n and is written as

$$z_n(t) = \hat{x}(T_n) + \int_0^t h_{r_n}(z_n(s))ds. \qquad \text{(Restatement of (27))}$$

Ideally, we would like to see that the error of Euler's discretization diminishes asymptotically. With (18) and (21), $\forall \tau > 0$, $\forall t \in [-\tau, \tau]$,

$$\bar{x}(t(n)+t) = x_{m(t(n)+t)} = \begin{cases} \bar{x}(t(n)) + \sum_{i=n}^{m(t(n)+t)-1} \alpha(i) H(\bar{x}(t(i)), Y_{i+1}) & \text{if } t \geq 0 \\ \bar{x}(t(n)) - \sum_{i=m(t(n)+t)}^{n-1} \alpha(i) H(\bar{x}(t(i)), Y_{i+1}) & \text{if } t < 0. \end{cases} \tag{76}$$

Notably, the property (18) that $\forall t < 0, m(t) = 0$ in (76) ensures $\bar{x}(t(n)+t)$ is well-defined when $t(n)+t < 0$. Precisely speaking, $\forall \tau > 0$, $\forall t \in [-\tau, \tau]$, the discretization error is defined as

$$\bar{f}_n(t) \doteq \bar{x}(t(n)+t) - \bar{z}_n(t). \tag{77}$$

and we would like $\bar{f}_n(t)$ diminishes to 0 as $n \to \infty$ in certain sense. To this end, we study the following three sequences of functions

$$\{\bar{x}(t(n)+t)\}_{n=0}^{\infty}, \{\bar{z}_n(t)\}_{n=0}^{\infty}, \{\bar{f}_n(t)\}_{n=0}^{\infty}.$$

Equicontinuity in the extended sense on domain $(-\infty, \infty)$ is defined as following (Section 4.2.1 in Kushner and Yin (2003)).

Definition 30 *A sequence of functions* $\{g_n : (-\infty, \infty) \to \mathbb{R}^K\}$ *is equicontinuous in the extended sense on* $(-\infty, \infty)$ *if* $\sup_n \|g_n(0)\| < \infty$ *and* $\forall \tau > 0$, $\forall \epsilon > 0$, $\exists \delta > 0$ *such that*

$$\limsup_n \sup_{0 \leq |t_1 - t_2| \leq \delta, |t_1| \leq \tau, |t_2| \leq \tau} \|g_n(t_1) - g_n(t_2)\| \leq \epsilon.$$

We show $\{\bar{x}(t(n)+t)\}$, $\{\bar{z}_n(t)\}$ and $\{\bar{f}_n(t)\}$ are all equicontinuous in the extended sense.

Lemma 31 *The three sequences of functions* $\{\bar{x}(t(n)+t)\}_{n=0}^{\infty}$, $\{\bar{z}_n(t)\}_{n=0}^{\infty}$, *and* $\{\bar{f}_n(t)\}_{n=0}^{\infty}$ *are all equicontinuous in the extended sense on* $t \in (-\infty, \infty)$.

To prove those lemmas, we need the Gronwall inequality in the reverse time in Appendix A.2. Compared to lemmas in the main text which have domain $t \in [0, T)$, lemmas in this section have similar proofs because we first fix an arbitrary τ and prove properties on the domain $t \in [-\tau, \tau]$. We omit proofs for Lemma 31 because they are ditto to proofs of Lemma 15. Similar to Lemma 16, we now construct a particular subsequence of interest.

Lemma 32 *There exists a subsequence* $\{n_k\}_{k=0}^{\infty} \subseteq \{0, 1, 2, \dots\}$ *and some continuous functions* $\bar{f}^{\lim}(t)$ *and* $\bar{x}^{\lim}(t)$ *such that* $\forall \tau$, $\forall t \in [-\tau, \tau]$,

$$\lim_{k \to \infty} \bar{f}_{n_k}(t) = \bar{f}^{\lim}(t),$$

$$\lim_{k\to\infty} \bar{x}(T_{n_k}+t) = \bar{x}^{\lim}(t),$$

where both convergences are uniform in t on $[-\tau,\tau]$. Furthermore, let $\bar{z}^{\lim}(t)$ denote the unique solution to the ODE (13) with the initial condition

$$\bar{z}^{\lim}(0) = \bar{x}^{\lim}(0),$$

in other words,

$$\bar{z}^{\lim}(t) = \bar{x}^{\lim}(0) + \int_0^t h(\bar{z}^{\lim}(s))ds.$$

Then $\forall\tau$, $\forall t \in [-\tau,\tau]$, we have

$$\lim_{k\to\infty} \bar{z}_{n_k}(t) = \bar{z}^{\lim}(t),$$

where the convergence is uniform in t on $[-\tau,\tau]$.

Its proof is ditto to the proof of Lemma 16 and is omitted. We use the subsequence $\{n_k\}$ intensively in the remaining proofs. Recall that $\bar{f}_n(t)$ denotes the discretization error between $\bar{x}(t(n)+t)$ and $\bar{z}_n(t)$. We now proceed to prove that this discretization error diminishes along $\{n_k\}$. In particular, we aim to prove that $\forall\tau$, $\forall t \in [-\tau,\tau]$,

$$\lim_{k\to\infty} \left\|\bar{f}_{n_k}(t)\right\| = \left\|\bar{f}^{\lim}(t)\right\| = 0.$$

This means $\bar{x}(t(n_k)+t)$ is close to $\bar{z}_{n_k}(t)$ as $k\to\infty$. For $t\in(0,\tau]$, the proof for this part is the same as the proof we have done in Section 4.4. Thus, we only discuss the proof for $t\in[-\tau,0]$. $\forall\tau$, $\forall t\in[-\tau,0]$,

$$\begin{aligned}
&\lim_{k\to\infty} \left\|\bar{f}_{n_k}(t)\right\| \\
=&\lim_{k\to\infty} \left\|\bar{x}(t(n_k)) - \sum_{i=m(t(n_k)+t)}^{n_k-1} \alpha(i)H(\bar{x}(t(i)),Y_{i+1}) - \bar{z}_{n_k}(t)\right\| && \text{(by (76) and (77))} \\
=&\lim_{k\to\infty} \left\|-\sum_{i=m(t(n_k)+t)}^{n_k-1} \alpha(i)H(\bar{x}(t(i)),Y_{i+1}) - \int_0^t h(\bar{z}_{n_k}(s))ds\right\| && \text{(by (75))} \\
\leq&\lim_{k\to\infty} \left\|-\sum_{i=m(t(n_k)+t)}^{n_k-1} \alpha(i)H(\bar{x}(t(i)),Y_{i+1}) - \int_0^t h(\bar{x}^{\lim}(s))ds\right\| \\
&+\lim_{k\to\infty} \left\|\int_0^t h(\bar{x}^{\lim}(s))ds - \int_0^t h(\bar{z}_{n_k}(s))ds\right\|. && (78)
\end{aligned}$$

We now prove that the first term in the RHS of (78) is 0.

Lemma 33 *$\forall \tau$, $\forall t \in [-\tau, 0]$,*

$$\lim_{k\to\infty} \left\| - \sum_{i=m(t(n_k)+t)}^{n_k-1} \alpha(i) H(\bar{x}(t(i)), Y_{i+1}) - \int_0^t h(\bar{x}^{\lim}(s)) ds \right\| = 0.$$

Its proof is ditto to the proof of Lemma 18 and is omitted. This convergence is also simpler than (36) because here we have only a single (H, h). But in (36), we have a sequence $\{(H_{n_k}, h_{n_k})\}$, for which we have to split it to a double limit (37) and then invoke the Moore-Osgood theorem to reduce it to the single (H, h) case.

Lemma 33 confirms that the first term in the RHS of (78) is 0. Moreover, it also enables us to rewrite $\bar{x}^{\lim}(t)$ from a summation form to an integral form. $\forall \tau$, $\forall t \in [-\tau, 0]$

$$\begin{aligned} &\bar{x}^{\lim}(t) \\ =& \lim_{k\to\infty} \bar{x}(t(n_k)) - \sum_{i=m(t(n_k)+t)}^{n_k-1} \alpha(i) H(\bar{x}(t(i)), Y_{i+1}) \\ =& \lim_{k\to\infty} \bar{x}(t(n_k)) + \int_0^t h(\bar{x}^{\lim}(s)) ds. && \text{(by Lemma 33)} \end{aligned}$$

Thus, we can show the following diminishing discretization error.

Lemma 34 *$\forall \tau$, $\forall t \in [-\tau, \tau]$,*

$$\lim_{k\to\infty} \left\| \bar{f}_{n_k}(t) \right\| = 0.$$

Moreover, the convergence is uniform in t on $[-\tau, \tau]$.

Its proof is ditto to the proof of Lemma 20 and is omitted. This immediately implies that for any $t \in (-\infty, \infty)$

$$\lim_{k\to\infty} \bar{x}(t(n_k) + t) = \bar{z}^{\lim}(t). \tag{79}$$

Theorem 7 then yields that

$$\sup_{t\in(-\infty,\infty)} \left\| \bar{z}^{\lim}(t) \right\| < \infty.$$

Let X be the limit set of $\{x_n\}$, i.e., X consists of all the limits of all the convergent subsequences of $\{x_n\}$. By Theorem 7, $\sup_n \|x_n\| < \infty$, so X is bounded and nonempty. We now prove X is an invariant set of the ODE (13). For any $x \in X$, there exists a subsequence $\{x_{n_k}\}$ such that

$$\lim_{k\to\infty} x_{n_k} = x.$$

Since $\{\bar{x}(t(n_k) + t)\}$ is equicontinuous in the extended sense, following the way we arrive at (79), we can construct a subsequence $\{n'_k\} \subseteq \{n_k\}$ such that

$$\lim_{k\to\infty} \bar{x}(t(n'_k) + t) = z^{\lim}(t), \tag{80}$$

where $z^{\lim}(t)$ is a solution to the ODE (13) and $z^{\lim}(0) = x$. The remaining is to show that $z^{\lim}(t)$ lies entirely in X. For any $t \in (-\infty, \infty)$, by the piecewise constant nature of $\bar{x}$ in (76), the above limit (80) implies that there exists a subsequence of $\{x_n\}$ that converges to $z^{\lim}(t)$, indicating $z^{\lim}(t) \in X$ by the definition of the limit set. We now have proved $\forall x \in X$, there exists a solution $z^{\lim}(t)$ to the ODE (13) such that $z^{\lim}(0) = x$ and $\forall t \in (-\infty, \infty), z^{\lim}(t) \in X$. This means X is an invariant set, by definition. In particular, X is a bounded invariant set.

We now prove that $\{x_n\}$ converges to X. Let $\{x_{n_k}\}$ be any convergent subsequence of $\{x_n\}$ with its limit denoted by x. We must have $x \in X$ by the definition of the limit set. So we have proved that all convergent subsequences of $\{x_n\}$ converge to a point in the bounded invariant set X. If $\{x_n\}$ does not converge to X, there must exists a subsequence $\left\{x_{n'_k}\right\}$ such that $\left\{x_{n'_k}\right\}$ is always away from X by some small $\epsilon_0 > 0$, i.e., $\forall k$,

$$\inf_{x \in X} \left\| x_{n'_k} - x \right\| \geq \epsilon_0. \tag{81}$$

But $\left\{x_{n'_k}\right\}$ is bounded so it must have a convergent subsequence, which, by the definition of the limit set, converges to some point in X. This contradicts (81). So we must have $\{x_n\}$ converges to X, which is a bounded invariant set of the ODE (13). This completes the proof. ■

B.10 Proof of Theorem 24

Proof For simplicity, we define

$$A' \doteq \begin{bmatrix} -C & A \\ -A^\top & 0 \end{bmatrix},$$
$$b' \doteq \begin{bmatrix} b \\ 0 \end{bmatrix}.$$

We first invoke Corollary 8 to show that

$$\lim_{t \to \infty} x_t = -A'^{-1} b' \quad \text{a.s.}$$

Assumption 1 follows immediately from Lemma 23.

Assumption 2 follows immediately from Assumption 5.2.

For Assumption 3, define

$$H_\infty(x, y) \doteq \begin{bmatrix} -C(y) & A(y) \\ -A(y)^\top & 0 \end{bmatrix} x.$$

Then we have

$$H_c(x, y) - H_\infty(x, y) = \frac{1}{c} \begin{bmatrix} b(y) \\ 0 \end{bmatrix}.$$

After noticing

$$\left\|b((s,a,s',e)) - b((s,a,s',e'))\right\| = \rho(s,a)|r(s,a)|\left\|e-e'\right\|, \quad \forall s,a,s',e,e',$$

Assumption 3 follows immediately from Lemma 23.

For Assumption 4, it can be easily verified that both $H(x,y)$ and $H_\infty(x,y)$ are Lipschitz continuous in x for each y with the Lipschitz constant being

$$L(y) \doteq \left\|\begin{bmatrix} -C(y) & A(y) \\ -A(y)^\top & 0 \end{bmatrix}\right\|.$$

Since $A(y), b(y), C(y)$ are Lipschitz continuous in e for each (s,a,s'), Lemma 23 implies that

$$\begin{aligned} h(x) =& A'x + b', \\ h_\infty(x) =& A'x, \\ L =& \left\|A'\right\|. \end{aligned}$$

Assumption 4 then follows.

For Assumption 5, we have

$$\|h_c(x) - h_\infty(x)\| \leq \frac{\|b'\|}{c},$$

the uniform convergence of h_c to h_∞ follows immediately. Proving that A' is Hurwitz is a standard exercise using the field of values of A'. We refer the reader to Section 5 of Sutton et al. (2009) for details and omit the proof. This immediately implies the globally asymptotically stability of the following two ODEs

$$\frac{\mathrm{d}x(t)}{\mathrm{d}t} = A'x(t) + b', \quad \frac{\mathrm{d}x(t)}{\mathrm{d}t} = A'x(t).$$

The unique globally asymptotically equilibrium of the former is $-A'^{-1}b'$. That of the latter is 0. Assumption 5 then follows.

Assumption 6 follows immediately from Lemma 23 and Assumption 5.2.

Corollary 8 then implies that

$$\lim_{t\to\infty} x_t = -A'^{-1}b' \quad \text{a.s.}$$

Block matrix inversion immediately shows that the lower half of $A'^{-1}b'$ is $A^{-1}b$, yielding

$$\lim_{t\to\infty} \theta_t = -A^{-1}b \quad \text{a.s.},$$

which completes the proof. ∎

Appendix C. Auxiliary Lemmas

Lemma 35

$$\begin{aligned} \forall n, \, T_{n+1} - T_n \geq & T, \\ \lim_{n\to\infty} T_{n+1} - T_n = & T. \end{aligned}$$

Moreover, $\forall \tau > 0, t_1, t_2$ *such that* $-\tau \leq t_1 \leq t_2 \leq \tau$*, we have*

$$\lim_{n\to\infty} \sum_{i=m(t(n)+t_1)}^{m(t(n)+t_2)-1} \alpha(i) = t_2 - t_1. \tag{82}$$

Proof $\forall n$,

$$\begin{aligned} & T_{n+1} - T_n \\ =& t(m(T_n + T) + 1) - T_n && \text{(by (17))} \\ \geq & T_n + T - T_n && \text{(by (15))} \\ \geq & T. \end{aligned}$$

Thus,

$$\lim_{n\to\infty} T_{n+1} - T_n \geq T.$$

With

$$\begin{aligned} & \lim_{n\to\infty} T_{n+1} - T_n \\ =& \lim_{n\to\infty} t(m(T_n + T) + 1) - T_n \\ =& \lim_{n\to\infty} t(m(T_n + T)) + \alpha(m(T_n + T)) - T_n \\ \leq & \lim_{n\to\infty} T_n + T + \alpha(m(T_n + T)) - T_n && \text{(by (15))} \\ =& T, \end{aligned}$$

by the squeeze theorem, we have $\lim_{n\to\infty} T_{n+1} - T_n = T$.

To prove (82), $\forall \tau$, $\forall -\tau \leq t_1 \leq t_2 \leq \tau$, it suffices to only consider large n such that $t(n) - \tau \geq 0$. We have

$$\begin{aligned} & \lim_{n\to\infty} \sum_{i=m(t(n)+t_1)}^{m(t(n)+t_2)-1} \alpha(i) \\ =& \lim_{n\to\infty} t(m(t(n) + t_2)) - t(m(t(n) + t_1)) \\ \leq & \lim_{n\to\infty} t(n) + t_2 - t(m(t(n) + t_1)) && \text{(by (15))} \\ \leq & \lim_{n\to\infty} t(n) + t_2 - (t(n) + t_1 - \alpha(m(t(n) + t_1))) && \text{(by (16))} \\ =& t_2 - t_1 + \lim_{n\to\infty} \alpha(m(t(n) + t_1)) \end{aligned}$$

$$=t_2 - t_1 \qquad \text{(by (3))}$$

and

$$\begin{aligned}
&\lim_{n\to\infty} \sum_{i=m(t(n)+t_1)}^{m(t(n)+t_2)-1} \alpha(i) \\
=& \lim_{n\to\infty} t(m(t(n)+t_2)) - t(m(t(n)+t_1)) \\
\geq& \lim_{n\to\infty} t(n) + t_2 - \alpha(m(t(n)+t_2)) - t(m(t(n)+t_1)) && \text{(by (16))} \\
\geq& \lim_{n\to\infty} t(n) + t_2 - \alpha(m(t(n)+t_2)) - (t(n)+t_1) && \text{(by (15))} \\
=& \lim_{n\to\infty} t_2 - t_1 - \alpha(m(t(n)+t_2)) \\
=& t_2 - t_1. && \text{(by (3))}
\end{aligned}$$

By the squeeze theorem, we have

$$\lim_n \sum_{i=m(t(n)+t_1)}^{m(t(n)+t_2)-1} \alpha(i) = t_2 - t_1.$$

■

Lemma 36 *For any $x, x', c \geq 1$, including $c = \infty$,*

$$\left\|H_c(x,y) - H_c(x',y)\right\| \leq L(y)\left\|x - x'\right\|, \tag{83}$$
$$\left\|h_c(x) - h_c(x')\right\| \leq L\left\|x - x'\right\|. \tag{84}$$

Proof To prove (83), we first consider $1 \leq c < \infty$,

$$\begin{aligned}
&\left\|H_c(x,y) - H_c(x',y)\right\| \\
=&\left\|\frac{H(cx,y)}{c} - \frac{H(cx',y)}{c}\right\| && \text{(by (4))} \\
\leq&\frac{\|H(cx,y) - H(cx',y)\|}{c} \\
\leq& L(y)\frac{\|cx - cx'\|}{c} && \text{(by (7))} \\
=& L(y)\left\|x - x'\right\|.
\end{aligned}$$

By (8),

$$\left\|H_\infty(x,y) - H_\infty(x',y)\right\| \leq L(y)\left\|x - x'\right\|.$$

To prove (84), $\forall x$, $\forall x'$, $\forall c \geq 1$ including $c = \infty$,

$$\left\|h_c(x) - h_c(x')\right\|$$

$$
\begin{aligned}
&= \left\| \mathbb{E}_{y \sim d_{\mathcal{Y}}} \left[H_c(x, y) - H_c(x', y) \right] \right\| \\
&\leq \mathbb{E}_{y \sim d_{\mathcal{Y}}} \left[\left\| H_c(x, y) - H_c(x', y) \right\| \right] \\
&\leq \mathbb{E}_{y \sim d_{\mathcal{Y}}} \left[L(y) \left\| x - x' \right\| \right] \\
&\leq L \left\| x - x' \right\|.
\end{aligned}
$$

■

Lemma 37 $\forall x$,

$$\sup_{c \geq 1} \|h_c(0)\| < \infty, \tag{85}$$

$$\sup_{c \geq 1} \limsup_{n} \sup_{0 \leq t_1 \leq t_2 \leq T_{n+1} - T_n} \left\| \sum_{i=m(T_n+t_1)}^{m(T_n+t_2)-1} \alpha(i) \left[H_c(x, Y_{i+1}) - h_c(x) \right] \right\| = 0 \quad a.s., \tag{86}$$

$$\sup_{c \geq 1} \sup_{n} \sup_{0 \leq t_1 \leq t_2 \leq T_{n+1} - T_n} \left\| \sum_{i=m(T_n+t_1)}^{m(T_n+t_2)-1} \alpha(i) H_c(0, Y_{i+1}) \right\| < \infty \quad a.s., \tag{87}$$

$$\lim_{\delta \to 0^+} \sup_{c \geq 1} \limsup_{n} \sup_{0 \leq t_2 - t_1 \leq \delta} \left\| \sum_{i=m(T_n+t_1)}^{m(T_n+t_2)-1} \alpha(i) H_c(0, Y_{i+1}) \right\| = 0 \quad a.s. \tag{88}$$

Proof

Proof of (85):

$$\sup_{c \geq 1} \|h_c(0)\| = \sup_{c \geq 1} \left\| \frac{h(0)}{c} \right\| \leq \sup_{c \geq 1} \|h(0)\| = \|h(0)\| < \infty.$$

Proof of (86): $\forall x$,

$$
\begin{aligned}
&\sup_{c \geq 1} \limsup_{n} \sup_{0 \leq t_1 \leq t_2 \leq T_{n+1} - T_n} \left\| \sum_{i=m(T_n+t_1)}^{m(T_n+t_2)-1} \alpha(i) \left[H_c(x, Y_{i+1}) - h_c(x) \right] \right\| \\
=& \sup_{c \geq 1} \limsup_{n} \sup_{0 \leq t_1 \leq t_2 \leq T_{n+1} - T_n} \left\| \sum_{i=m(T_n+t_1)}^{m(T_n+t_2)-1} \alpha(i) \left[\frac{H(cx, Y_{i+1})}{c} - \frac{h(cx)}{c} \right] \right\| \\
=& \sup_{c \geq 1} \frac{1}{c} \limsup_{n} \sup_{0 \leq t_1 \leq t_2 \leq T_{n+1} - T_n} \left\| \sum_{i=m(T_n+t_1)}^{m(T_n+t_2)-1} \alpha(i) \left[H(cx, Y_{i+1}) - h(cx) \right] \right\| \\
\leq& \sup_{c \geq 1} \frac{1}{c} \limsup_{n} \sup_{0 \leq t_1 \leq t_2 \leq T + \sup_j \alpha(j)} \left\| \sum_{i=m(T_n+t_1)}^{m(T_n+t_2)-1} \alpha(i) \left[H(cx, Y_{i+1}) - h(cx) \right] \right\| \\
& \qquad (\forall n, T_{n+1} - T_n \leq T + \textstyle\sup_j \alpha(j))
\end{aligned}
$$

$$
\begin{aligned}
=&\sup_{c\geq 1}\frac{1}{c}\cdot 0 && \text{(by Lemma 9)}\\
=&0. && (89)
\end{aligned}
$$

Proof of (87):

$$
\begin{aligned}
&\limsup_n \sup_{0\leq t_1\leq t_2\leq T_{n+1}-T_n}\left\|\sum_{i=m(T_n+t_1)}^{m(T_n+t_2)-1}\alpha(i)H(0,Y_{i+1})\right\|\\
=&\limsup_n \sup_{0\leq t_1\leq t_2\leq T_{n+1}-T_n}\left\|\sum_{i=m(T_n+t_1)}^{m(T_n+t_2)-1}\alpha(i)[H(0,Y_{i+1})-h(0)+h(0)]\right\|\\
\leq&\limsup_n \sup_{0\leq t_1\leq t_2\leq T_{n+1}-T_n}\left\|\sum_{i=m(T_n+t_1)}^{m(T_n+t_2)-1}\alpha(i)[H(0,Y_{i+1})-h(0)]\right\|\\
&+\limsup_n \sup_{0\leq t_1\leq t_2\leq T_{n+1}-T_n}\left\|\sum_{i=m(T_n+t_1)}^{m(T_n+t_2)-1}\alpha(i)h(0)\right\|\\
\leq&\limsup_n \sup_{0\leq t_1\leq t_2\leq T+\sup_j\alpha(j)}\left\|\sum_{i=m(T_n+t_1)}^{m(T_n+t_2)-1}\alpha(i)[H(0,Y_{i+1})-h(0)]\right\|\\
&+\limsup_n \sup_{0\leq t_1\leq t_2\leq T+\sup_j\alpha(j)}\left\|\sum_{i=m(T_n+t_1)}^{m(T_n+t_2)-1}\alpha(i)h(0)\right\| && (\forall n, T_{n+1}-T_n\leq T+\textstyle\sup_j\alpha(j))\\
=&\limsup_n \sup_{0\leq t_1\leq t_2\leq T+\sup_j\alpha(j)}\left\|\sum_{i=m(T_n+t_1)}^{m(T_n+t_2)-1}\alpha(i)h(0)\right\| && \text{(by Lemma 9)}\\
=&\|h(0)\|\limsup_n \sup_{0\leq t_1\leq t_2\leq T+\sup_j\alpha(j)}\sum_{i=m(T_n+t_1)}^{m(T_n+t_2)-1}\alpha(i)\\
=&\|h(0)\|(T+\sup_j\alpha(j)) && \text{(by Lemma 35)}\\
<&\infty. && (90)
\end{aligned}
$$

We now consider c in the above bounds. We first get

$$
\begin{aligned}
&\sup_{c\geq 1}\sup_n \sup_{0\leq t_1\leq t_2\leq T_{n+1}-T_n}\left\|\sum_{i=m(T_n+t_1)}^{m(T_n+t_2)-1}\alpha(i)H_c(0,Y_{i+1})\right\|\\
=&\sup_{c\geq 1}\sup_n \sup_{0\leq t_1\leq t_2\leq T_{n+1}-T_n}\left\|\sum_{i=m(T_n+t_1)}^{m(T_n+t_2)-1}\alpha(i)\frac{H(0,Y_{i+1})}{c}\right\| && \text{(by (4))}\\
=&\sup_n \sup_{0\leq t_1\leq t_2\leq T_{n+1}-T_n}\left\|\sum_{i=m(T_n+t_1)}^{m(T_n+t_2)-1}\alpha(i)H(0,Y_{i+1})\right\| && \text{(by } c\geq 1)
\end{aligned}
$$

$$<\infty. \qquad \text{(by (90))}$$

Proof of (88):

$$\begin{aligned}
&\lim_{\delta\to 0^+} \sup_{c\geq 1} \limsup_{n} \sup_{0\leq t_2-t_1\leq\delta} \left\| \sum_{i=m(T_n+t_1)}^{m(T_n+t_2)-1} \alpha(i) H_c(0, Y_{i+1}) \right\| \\
\leq & \lim_{\delta\to 0^+} \sup_{c\geq 1} \limsup_{n} \sup_{0\leq t_2-t_1\leq\delta} \left\| \sum_{i=m(T_n+t_1)}^{m(T_n+t_2)-1} \alpha(i) \left[H_c(0, Y_{i+1}) - h_c(0) \right] \right\| \\
& + \lim_{\delta\to 0^+} \sup_{c\geq 1} \limsup_{n} \sup_{0\leq t_2-t_1\leq\delta} \left\| \sum_{i=m(T_n+t_1)}^{m(T_n+t_2)-1} \alpha(i) h_c(0) \right\| \\
\leq & 0 + \lim_{\delta\to 0^+} \sup_{c\geq 1} \limsup_{n} \sup_{0\leq t_2-t_1\leq\delta} \left\| \sum_{i=m(T_n+t_1)}^{m(T_n+t_2)-1} \alpha(i) h_c(0) \right\| && \text{(by (89))} \\
\leq & 0 + \lim_{\delta\to 0^+} \sup_{c\geq 1} \limsup_{n} \sup_{0\leq t_2-t_1\leq\delta} \left\| \sum_{i=m(T_n+t_1)}^{m(T_n+t_2)-1} \alpha(i) \frac{h(0)}{c} \right\| \\
\leq & 0 + \|h(0)\| \lim_{\delta\to 0^+} \sup_{c\geq 1} \frac{1}{c} \limsup_{n} \sup_{0\leq t_2-t_1\leq\delta} \sum_{i=m(T_n+t_1)}^{m(T_n+t_2)-1} \alpha(i) \\
\leq & \|h(0)\| \lim_{\delta\to 0^+} \sup_{c\geq 1} \frac{1}{c}\delta && \text{(by (82))} \\
= & \|h(0)\| \lim_{\delta\to 0^+} \delta \\
= & 0.
\end{aligned}$$

■

Lemma 38

$$\sup_{n} \sup_{0\leq t_1\leq t_2\leq T_{n+1}-T_n} \left(\sum_{i=m(T_n+t_1)}^{m(T_n+t_2)-1} \alpha(i) L(Y_{i+1}) \right) < \infty \quad a.s., \tag{91}$$

$$\lim_{\delta\to 0^+} \limsup_{n} \sup_{0\leq t_2-t_1\leq\delta} \left(\sum_{i=m(T_n+t_1)}^{m(T_n+t_2)-1} \alpha(i) L(Y_{i+1}) \right) = 0 \quad a.s., \tag{92}$$

$$\sup_{n} \sup_{0\leq t_1\leq t_2\leq T_{n+1}-T_n} \left(\sum_{i=m(T_n+t_1)}^{m(T_n+t_2)-1} \alpha(i) L_b(Y_{i+1}) \right) < \infty \quad a.s. \tag{93}$$

Its proof is similar to the proof of Lemma 37 and is thus omitted.

Lemma 39 *Fix a sample path $\{x_0, \{Y_i\}_{i=1}^\infty\}$, there exists a constant C_H such that*

$$LT \leq C_H, \tag{94}$$

$$\sup_{c\geq 1} \|h_c(0)\| \leq \frac{C_H}{T}, \tag{95}$$

$$\sup_{c\geq 1}\sup_{n}\sup_{0\leq t_1\leq t_2\leq T_{n+1}-T_n} \left\| \sum_{i=m(T_n+t_1)}^{m(T_n+t_2)-1} \alpha(i) H_c(0, Y_{i+1}) \right\| \leq C_H, \tag{96}$$

$$\sup_{n}\sup_{0\leq t_1\leq t_2\leq T_{n+1}-T_n} \sum_{i=m(T_n+t_1)}^{m(T_n+t_2)-1} \alpha(i) L(Y_{i+1}) \leq C_H. \tag{97}$$

Moreover, for the presentation convenience, we denote

$$C_{\hat{x}} \doteq [1 + C_H]\, e^{C_H}. \tag{98}$$

Proof Fix a sample path $\{x_0, \{Y_i\}_{i=1}^\infty\}$,

$$LT < \infty, \qquad (L \text{ and } T \text{ are constants})$$

$$\sup_{c\geq 1} \|h_c(0)\| T < \infty, \qquad (\text{by } (85))$$

$$\sup_{c\geq 1}\sup_{n}\sup_{0\leq t_1\leq t_2\leq T_{n+1}-T_n} \left\| \sum_{i=m(T_n+t_1)}^{m(T_n+t_2)-1} \alpha(i) H_c(0, Y_{i+1}) \right\| < \infty, \qquad (\text{by } (87))$$

$$\sup_{n}\sup_{0\leq t_1\leq t_2\leq T_{n+1}-T_n} \sum_{i=m(T_n+t_1)}^{m(T_n+t_2)-1} \alpha(i) L(Y_{i+1}) < \infty. \qquad (\text{by } (91))$$

Thus, there exists a constant C_H such that

$$LT \leq C_H$$

$$\sup_{c\geq 1} \|h_c(0)\| \leq \frac{C_H}{T},$$

$$\sup_{c\geq 1}\sup_{n}\sup_{0\leq t_1\leq t_2\leq T_{n+1}-T_n} \left\| \sum_{i=m(T_n+t_1)}^{m(T_n+t_2)-1} \alpha(i) H_c(0, Y_{i+1}) \right\| \leq C_H,$$

$$\sup_{n}\sup_{0\leq t_1\leq t_2\leq T_{n+1}-T_n} \sum_{i=m(T_n+t_1)}^{m(T_n+t_2)-1} \alpha(i) L(Y_{i+1}) \leq C_H.$$

■

Lemma 40 $\sup_{n,t\in[0,T)} \|\hat{x}(T_n + t)\| \leq C_{\hat{x}}$.

Proof $\forall n \in \mathbb{N}, t \in [0, T)$,

$$
\begin{aligned}
&\|\hat{x}(T_n+t)\| \\
=&\left\|\hat{x}(T_n) + \sum_{i=m(T_n)}^{m(T_n+t)-1} \alpha(i) H_{r_n}(\hat{x}(t(i)), Y_{i+1})\right\| \\
\leq&\|\hat{x}(T_n)\| + \left\|\sum_{i=m(T_n)}^{m(T_n+t)-1} \alpha(i) H_{r_n}(\hat{x}(t(i)), Y_{i+1})\right\| \\
=&\|\hat{x}(T_n)\| + \left\|\sum_{i=m(T_n)}^{m(T_n+t)-1} \alpha(i) \left[H_{r_n}(\hat{x}(t(i)), Y_{i+1}) - H_{r_n}(0, Y_{i+1})\right] + \sum_{i=m(T_n)}^{m(T_n+t)-1} \alpha(i) H_{r_n}(0, Y_{i+1})\right\| \\
\leq&\|\hat{x}(T_n)\| + \sum_{i=m(T_n)}^{m(T_n+t)-1} \alpha(i) \|H_{r_n}(\hat{x}(t(i)), Y_{i+1}) - H_{r_n}(0, Y_{i+1})\| + \left\|\sum_{i=m(T_n)}^{m(T_n+t)-1} \alpha(i) H_{r_n}(0, Y_{i+1})\right\| \\
\leq&\|\hat{x}(T_n)\| + \sum_{i=m(T_n)}^{m(T_n+t)-1} \alpha(i) L(Y_{i+1}) \|\hat{x}(t(i))\| + \left\|\sum_{i=m(T_n)}^{m(T_n+t)-1} \alpha(i) H_{r_n}(0, Y_{i+1})\right\| \\
\leq&\|\hat{x}(T_n)\| + \sum_{i=m(T_n)}^{m(T_n+t)-1} \alpha(i) L(Y_{i+1}) \|\hat{x}(t(i))\| + C_H && \text{(by (96))} \\
\leq&1 + \sum_{i=m(T_n)}^{m(T_n+t)-1} \alpha(i) L(Y_{i+1}) \|\hat{x}(t(i))\| + C_H && \text{(by (24))} \\
\leq&\left[1 + C_H\right] e^{\sum_{i=m(T_n)}^{m(T_n+t)-1} \alpha(i) L(Y_{i+1})} \\
&\qquad \text{(by } \hat{x}(T_n+t) = \hat{x}(t(m(T_n+t))) \text{ and discrete Gronwall inequality in Theorem A.3)} \\
\leq&\left[1 + C_H\right] e^{C_H} && \text{(by (97))} \\
=&C_{\hat{x}}. && \text{(by (98))}
\end{aligned}
$$

■

Lemma 41 $\sup_{n, t \in [0, T)} \|z_n(t)\| \leq C_{\hat{x}}$.

Proof $\forall n, t \in [0, T)$,

$$
\begin{aligned}
&\|z_n(t)\| \\
=&\left\|z_n(0) + \int_0^t h_{r_n}(z_n(s)) ds\right\| \\
\leq&\|z_n(0)\| + \left\|\int_0^t h_{r_n}(z_n(s)) ds\right\|
\end{aligned}
$$

$$
\begin{aligned}
\leq & \|z_n(0)\| + \int_0^t \|h_{r_n}(z_n(s)) - h_{r_n}(0)\| ds + \int_0^t \|h_{r_n}(0)\| ds \\
\leq & \|z_n(0)\| + \int_0^t L\|z_n(s)\| ds + \int_0^t \|h_{r_n}(0)\| ds && \text{(by Lemma 36)} \\
\leq & \|z_n(0)\| + \int_0^t L\|z_n(s)\| ds + T\|h_{r_n}(0)\| \\
\leq & \|z_n(0)\| + \int_0^t L\|z_n(s)\| ds + T\frac{C_H}{T} && \text{(by (95))} \\
\leq & 1 + \int_0^t L\|z_n(s)\| ds + C_H && \text{(by (24), (26))} \\
\leq & \left[1 + C_H\right] e^{LT} && \text{(by Gronwall inequality in Theorem A.1)} \\
\leq & \left[1 + C_H\right] e^{C_H} && \text{(by (94))} \\
= & C_{\hat{x}} && \text{(by (98))}
\end{aligned}
$$

■

Lemma 42 $\forall n$,

$$
\|\bar{x}(T_{n+1})\| \leq \left(\|\bar{x}(T_n)\| C_H + C_H\right) e^{C_H} + \|\bar{x}(T_n)\|
$$

where C_H is a positive constant defined in Lemma 39.

Proof We first show the difference between $\bar{x}(T_{n+1})$ and $\bar{x}(T_n)$ by the following derivations. $\forall n$, $\forall t \in [0, T_{n+1} - T_n]$,

$$
\begin{aligned}
& \|\bar{x}(T_n + t) - \bar{x}(T_n)\| \\
= & \|\bar{x}(t(m(T_n + t))) - \bar{x}(T_n)\| \\
= & \left\| \bar{x}(T_n) + \sum_{i=m(T_n)}^{m(T_n+t)-1} \alpha(i) H(\bar{x}(t(i)), Y_{i+1}) - \bar{x}(T_n) \right\| \\
= & \left\| \sum_{i=m(T_n)}^{m(T_n+t)-1} \alpha(i) H(\bar{x}(t(i)), Y_{i+1}) \right\| \\
\leq & \sum_{i=m(T_n)}^{m(T_n+t)-1} \alpha(i) \|H(\bar{x}(t(i)), Y_{i+1}) - H(\bar{x}(T_n), Y_{i+1})\| + \left\| \sum_{i=m(T_n)}^{m(T_n+t)-1} \alpha(i) H(\bar{x}(T_n), Y_{i+1}) \right\| \\
\leq & \sum_{i=m(T_n)}^{m(T_n+t)-1} \alpha(i) L(Y_{i+1}) \|\bar{x}(t(i)) - \bar{x}(T_n)\| + \left\| \sum_{i=m(T_n)}^{m(T_n+t)-1} \alpha(i) H(\bar{x}(T_n), Y_{i+1}) \right\| \\
\leq & \sum_{i=m(T_n)}^{m(T_n+t)-1} \alpha(i) L(Y_{i+1}) \|\bar{x}(t(i)) - \bar{x}(T_n)\| + \sum_{i=m(T_n)}^{m(T_n+t)-1} \alpha(i) \|H(\bar{x}(T_n), Y_{i+1}) - H(0, Y_{i+1})\|
\end{aligned}
$$

$$
\begin{aligned}
&+ \left\| \sum_{i=m(T_n)}^{m(T_n+t)-1} \alpha(i) H(0, Y_{i+1}) \right\| \\
&\le \sum_{i=m(T_n)}^{m(T_n+t)-1} \alpha(i) L(Y_{i+1}) \|\bar{x}(t(i)) - \bar{x}(T_n)\| + \sum_{i=m(T_n)}^{m(T_n+t)-1} \alpha(i) L(Y_{i+1}) \|\bar{x}(T_n)\| \\
&+ \left\| \sum_{i=m(T_n)}^{m(T_n+t)-1} \alpha(i) H(0, Y_{i+1}) \right\| && \text{(by Assumption 4)} \\
&= \sum_{i=m(T_n)}^{m(T_n+t)-1} \alpha(i) L(Y_{i+1}) \|\bar{x}(t(i)) - \bar{x}(T_n)\| + \|\bar{x}(T_n)\| \sum_{i=m(T_n)}^{m(T_n+t)-1} \alpha(i) L(Y_{i+1}) \\
&+ \left\| \sum_{i=m(T_n)}^{m(T_n+t)-1} \alpha(i) H(0, Y_{i+1}) \right\| \\
&\le \sum_{i=m(T_n)}^{m(T_n+t)-1} \alpha(i) L(Y_{i+1}) \|\bar{x}(t(i)) - \bar{x}(T_n)\| + \|\bar{x}(T_n)\| C_H + \left\| \sum_{i=m(T_n)}^{m(T_n+t)-1} \alpha(i) H(0, Y_{i+1}) \right\| \\
& && \text{(by (97))} \\
&\le \sum_{i=m(T_n)}^{m(T_n+t)-1} \alpha(i) L(Y_{i+1}) \|\bar{x}(t(i)) - \bar{x}(T_n)\| + [\|\bar{x}(T_n)\| C_H + C_H] && \text{(by (96))} \\
&\le [\|\bar{x}(T_n)\| C_H + C_H] \, e^{\sum_{i=m(T_n)}^{m(T_n+t)-1} \alpha(i) L(Y_{i+1})} \\
& && \text{(by discrete Gronwall inequality in Theorem A.3)} \\
&\le [\|\bar{x}(T_n)\| C_H + C_H] \, e^{C_H} && \text{(by (97))} \qquad (99)
\end{aligned}
$$

■

Lemma 43

$$\limsup_k r_{n_{2,k}} = \infty.$$

Proof We use proof by contradiction. Suppose

$$\limsup_k r_{n_{2,k}} = C_r < \infty$$

where C_r is a constant. $\forall \epsilon > 0$, $\exists k_0$ such that $\forall k \ge k_0$,

$$r_{n_{2,k}} \le C_r + \epsilon.$$

By Lemma 42, $\forall k \ge k_0$,

$$r_{n_{1,k}} = \max\left\{ \left\|\bar{x}(T_{n_{1,k}})\right\|, 1 \right\} \qquad \text{(by (23))}$$

$$
\begin{aligned}
&= \max\left\{\left\|\bar{x}(T_{n_{2,k}+1})\right\|, 1\right\} && \text{(by (57))}\\
&\le \left\|\bar{x}(T_{n_{2,k}+1})\right\| + 1\\
&\le \left(\left\|\bar{x}(T_{n_{2,k}})\right\| C_H + C_H\right) e^{C_H} + \left\|\bar{x}(T_{n_{2,k}})\right\| + 1\\
&\le \left(r_{n_{2,k}} C_H + C_H\right) e^{C_H} + r_{n_{2,k}} + 1\\
&\le \left[(C_r + \epsilon) C_H + C_H\right] e^{C_H} + (C_r + \epsilon) + 1\\
&< \infty.
\end{aligned}
$$

This contradicts (56). Thus,

$$
\limsup_k r_{n_{2,k}} = \infty.
$$

■

Lemma 44 $\sup_{n, t\in[0,T)} \|h_{r_n}(z_n(t))\| < \infty$.

Proof $\forall n, \forall t \in [0, T)$,

$$
\begin{aligned}
&\|h_{r_n}(z_n(t))\|\\
\le& \|h_{r_n}(z_n(t)) - h_{r_n}(0)\| + \|h_{r_n}(0)\|\\
\le& L\|z_n(t)\| + \|h_{r_n}(0)\| && \text{(by Lemma 36)}\\
\le& LC_{\hat{x}} + \|h_{r_n}(0)\| && \text{(by Lemma 41)}\\
\le& LC_{\hat{x}} + \frac{C_H}{T}. && \text{(by (23) and (95))}
\end{aligned}
$$

Thus, because $C_{\hat{x}}, C_H$ are independent of n, t, $\sup_{n, t\in[0,T)} \|h_{r_n}(z_n(t))\| < \infty$.

■

Lemma 45 $\sup_{t\in[0,T)} \left\|z^{\lim}(t)\right\| \le C_{\hat{x}}$.

Proof $\forall t \in [0, T)$,

$$
\begin{aligned}
&\left\|z^{\lim}(t)\right\|\\
=& \left\|z^{\lim}(0) + \int_0^t h_\infty(z^{\lim}(s)) ds\right\|\\
\le& \left\|z^{\lim}(0)\right\| + \left\|\int_0^t h_\infty(z^{\lim}(s)) ds\right\|\\
=& \left\|z^{\lim}(0)\right\| + \left\|\int_0^t \left[h_\infty(z^{\lim}(s)) - h_\infty(0)\right] ds + \int_0^t h_\infty(0) ds\right\|\\
\le& \left\|z^{\lim}(0)\right\| + \int_0^t \left\|h_\infty(z^{\lim}(s)) - h_\infty(0)\right\| ds + \int_0^t \|h_\infty(0)\| ds
\end{aligned}
$$

$$
\begin{aligned}
&\leq \left\| z^{\lim}(0) \right\| + \int_0^t L \left\| z^{\lim}(s) \right\| ds + \int_0^t \left\| h_\infty(0) \right\| ds && \text{(by Lemma 36)} \\
&\leq 1 + \int_0^t L \left\| z^{\lim}(s) \right\| ds + \int_0^t \left\| h_\infty(0) \right\| ds && \text{(by (24), (26))} \\
&\leq 1 + \int_0^t L \left\| z^{\lim}(s) \right\| ds + T \left\| h_\infty(0) \right\| \\
&\leq 1 + \int_0^t L \left\| z^{\lim}(s) \right\| ds + C_H && \text{(by Assumption 5 and (95))} \\
&\leq [1 + C_H] \, e^{\int_0^t L ds} && \text{(by Gronwall inequality in Theorem A.1)} \\
&\leq [1 + C_H] \, e^{LT} \\
&\leq C_{\hat{x}}. && \text{(by (94), (98))}
\end{aligned}
$$

■

Lemma 46 $\lim_{k\to\infty} h_{r_{n_k}}(z^{\lim}(t)) = h_\infty(z^{\lim}(t))$ *uniformly in* $t \in [0, T)$.

Proof By Assumption 5, $\lim_{k\to\infty} h_{r_{n_k}}(v) = h_\infty(v)$ uniformly in a compact set $\left\{ v | v \in \mathbb{R}^d, \|v\| \leq C_x \right\}$. By Lemma 45, $\left\{ z^{\lim}(t) | t \in [0, T) \right\} \subseteq \left\{ v | v \in \mathbb{R}^d, \|v\| \leq C_x \right\}$. Therefore, $\lim_{k\to\infty} h_{r_{n_k}}(z^{\lim}(t)) = h_\infty(z^{\lim}(t))$ uniformly in $\left\{ z^{\lim}(t) | t \in [0, T) \right\}$ and on $t \in [0, T)$. ■

Lemma 47 $\forall t \in [0, T)$, *we have*

$$
\lim_{k\to\infty} z_{n_k}(t) = z^{\lim}(t).
$$

Moreover, the convergence is uniform in t *on* $[0, T)$.

Proof By (33), $\forall \delta > 0$, there exists a k_1 such that $\forall k \geq k_1$, $\forall t \in [0, T)$,

$$
\left\| \hat{x}(T_{n_k} + t) - \hat{x}^{\lim}(t) \right\| \leq \delta. \tag{100}
$$

By Lemma 46, there exists a k_2 such that $\forall k \geq k_2$, $\forall t \in [0, T)$,

$$
\left\| h_{r_{n_k}}(z^{\lim}(t)) - h_\infty(z^{\lim}(t)) \right\| \leq \delta. \tag{101}
$$

$\forall k \geq \max\{k_1, k_2\}$, $\forall t \in [0, T)$

$$
\begin{aligned}
&\left\| z_{n_k}(t) - z^{\lim}(t) \right\| \\
=&\left\| \hat{x}(T_{n_k}) + \int_0^t h_{r_{n_k}}(z_{n_k}(s)) ds - \hat{x}^{\lim}(0) - \int_0^t h_\infty(z^{\lim}(s)) ds \right\| \\
\leq&\left\| \hat{x}(T_{n_k}) - \hat{x}^{\lim}(0) \right\| + \left\| \int_0^t h_{r_{n_k}}(z_{n_k}(s)) ds - \int_0^t h_\infty(z^{\lim}(s)) ds \right\|
\end{aligned}
$$

$$
\begin{aligned}
&\leq \delta + \left\| \int_0^t h_{r_{n_k}}(z_{n_k}(s)) - h_\infty(z^{\lim}(s)) ds \right\| && \text{(by (100))} \\
&\leq \delta + \int_0^t \left\| h_{r_{n_k}}(z_{n_k}(s)) - h_{r_{n_k}}(z^{\lim}(s)) \right\| ds + \int_0^t \left\| h_{r_{n_k}}(z^{\lim}(s)) - h_\infty(z^{\lim}(s)) \right\| ds \\
&\leq \delta + L \int_0^t \left\| z_{n_k}(s) - z^{\lim}(s) \right\| ds + \int_0^t \left\| h_{r_{n_k}}(z^{\lim}(s)) - h_\infty(z^{\lim}(s)) \right\| ds && \text{(by Lemma 36)} \\
&\leq \delta + t\delta + L \int_0^t \left\| z_{n_k}(s) - z^{\lim}(s) \right\| ds && \text{(by (101))} \\
&\leq (\delta + t\delta) e^{Lt} && \text{(by Gronwall inequality in Theorem A.1)} \\
&\leq (\delta + T\delta) e^{LT},
\end{aligned}
$$

which completes the proof. ■

Lemma 48 *For any function $f : \mathbb{R} \times \mathbb{R} \to \mathbb{R}$, if $\lim_{\substack{a \to \infty \\ b \to \infty}} f(a, b) = L$ then $\lim_{c \to \infty} f(c, c) = L$ where L is a constant.*

Proof By definition, $\forall \epsilon > 0, \exists a_0, b_0$ such that $\forall a > a_0, b > b_0$, $\|f(a, b) - L\| < \epsilon$. Thus, $\forall \epsilon > 0, \exists c_0 = \max\{a_0, b_0\}$ such that $\forall c > c_0$, $\|f(c, c) - L\| < \epsilon$. ■

Lemma 49 $\forall t \in [0, T)$,

$$
\lim_{k \to \infty} \int_0^t h_{r_{n_k}}(\hat{x}^{\lim}(s)) ds = \int_0^t h_\infty(\hat{x}^{\lim}(s)) ds.
$$

Proof From Lemma 40, it is easy to see that

$$
\sup_{t \in [0, T)} \left\| \hat{x}^{\lim}(t) \right\| < \infty,
$$

which, similar to Lemma 44, implies that

$$
\sup_{k, t \in [0, T)} \left\| h_{r_{n_k}} \left(\hat{x}^{\lim}(t) \right) \right\| < \infty.
$$

By the dominated convergence theorem, $\forall t \in [0, T)$,

$$
\lim_{k \to \infty} \int_0^t h_{r_{n_k}}(\hat{x}^{\lim}(s)) ds = \int_0^t \lim_{k \to \infty} h_{r_{n_k}}(\hat{x}^{\lim}(s)) ds = \int_0^t h_\infty(\hat{x}^{\lim}(s)) ds,
$$

which completes the proof. ■

Lemma 50 $\forall t \in [0, T)$,

$$
\lim_{k \to \infty} \int_0^t h_{r_{n_k}}(z_{n_k}(s)) ds = \int_0^t h_\infty(z^{\lim}(s)) ds.
$$

Proof $\forall \epsilon > 0$, by Lemma 46, $\exists k_0$ such that $\forall k \geq k_0$, $\forall t \in [0, T)$,

$$\left\| h_{r_{n_k}}(z^{\lim}(s)) - h_\infty(z^{\lim}(s)) \right\| \leq \epsilon. \tag{102}$$

By Lemma 47, $\exists k_1$ such that $\forall k \geq k_1$, $\forall t \in [0, T)$,

$$\left\| z_{n_k}(t) - z^{\lim}(t) \right\| \leq \epsilon. \tag{103}$$

Thus, $\forall k \geq \max\{k_0, k_1\}$, $\forall t \in [0, T)$,

$$\begin{aligned}
&\left\| \int_0^t h_{r_{n_k}}(z_{n_k}(s)) ds - \int_0^t h_\infty(z^{\lim}(s)) ds \right\| \\
\leq & \left\| \int_0^t h_{r_{n_k}}(z_{n_k}(s)) ds - \int_0^t h_{r_{n_k}}(z^{\lim}(s)) ds \right\| + \left\| \int_0^t h_{r_{n_k}}(z^{\lim}(s)) ds - \int_0^t h_\infty(z^{\lim}(s)) ds \right\| \\
\leq & \int_0^t \left\| h_{r_{n_k}}(z_{n_k}(s)) - h_{r_{n_k}}(z^{\lim}(s)) \right\| ds + \int_0^t \left\| h_{r_{n_k}}(z^{\lim}(s)) - h_\infty(z^{\lim}(s)) \right\| ds \\
\leq & \int_0^t \left\| h_{r_{n_k}}(z_{n_k}(s)) - h_{r_{n_k}}(z^{\lim}(s)) \right\| ds + T\epsilon && \text{(by (102))} \\
\leq & \int_0^t L \left\| z_{n_k}(s) - z^{\lim}(s) \right\| ds + T\epsilon && \text{(by Lemma 36)} \\
\leq & LT\epsilon + T\epsilon. && \text{(by (103))}
\end{aligned}$$

Thus, $\forall t \in [0, T)$,

$$\lim_{k \to \infty} \int_0^t h_{r_{n_k}}(z_{n_k}(s)) ds = \int_0^t h_\infty(z^{\lim}(s)) ds.$$

■

Lemma 51

$$\lim_n \lim_{t \to T^-} \left\| \sum_{i=m(T_n+t)}^{m(T_{n+1})-1} \alpha(i) L(Y_{i+1}) \right\| = 0, \tag{104}$$

$$\lim_n \lim_{t \to T^-} \left\| \sum_{i=m(T_n+t)}^{m(T_{n+1})-1} \alpha(i) H(0, Y_{i+1}) \right\| = 0. \tag{105}$$

Proof

$$\begin{aligned}
&\limsup_n \lim_{t \to T^-} \left\| \sum_{i=m(T_n+t)}^{m(T_{n+1})-1} \alpha(i) L(Y_{i+1}) \right\| \\
=& \limsup_n \lim_{t \to T^-} \left\| \sum_{i=m(T_n+t)}^{m(T_{n+1})-1} \alpha(i) [L(Y_{i+1}) - L] + \sum_{i=m(T_n+t)}^{m(T_{n+1})-1} \alpha(i) L \right\|
\end{aligned}$$

$$
\begin{aligned}
&\leq \limsup_n \lim_{t\to T^-} \left\| \sum_{i=m(T_n+t)}^{m(T_{n+1})-1} \alpha(i)[L(Y_{i+1}) - L] \right\| + \limsup_n \lim_{t\to T^-} \left\| \sum_{i=m(T_n+t)}^{m(T_{n+1})-1} \alpha(i)L \right\| \\
&\leq \limsup_n \lim_{t\to T^-} \left\| \sum_{i=m(T_n+t)}^{m(T_{n+1})-1} \alpha(i)[L(Y_{i+1}) - L] \right\| + L \limsup_n \alpha(m(T_{n+1}) - 1) \\
&\leq \limsup_n \lim_{t\to T^-} \left\| \sum_{i=m(T_n+t)}^{m(T_{n+1})-1} \alpha(i)[L(Y_{i+1}) - L] \right\| + 0 && \text{(by (3))} \\
&\leq \limsup_n \sup_{0\leq t_1\leq t_2\leq T+\sup_j \alpha(j)} \left\| \sum_{i=m(T_n+t_1)}^{m(T_n+t_2)-1} \alpha(i)[L(Y_{i+1}) - L] \right\| \\
&=0. && \text{(by (20))}
\end{aligned}
$$

This implies

$$
\lim_n \lim_{t\to T^-} \left\| \sum_{i=m(T_n+t)}^{m(T_{n+1})-1} \alpha(i)L(Y_{i+1}) \right\| = 0.
$$

Following a similar proof, we have

$$
\lim_n \lim_{t\to T^-} \left\| \sum_{i=m(T_n+t)}^{m(T_{n+1})-1} \alpha(i)H(0, Y_{i+1}) \right\| = 0.
$$

■

Lemma 52 $\lim_{k\to\infty} \frac{\left\|\bar{x}(T_{n_k+1})\right\| - \lim_{t\to T^-} \left\|\bar{x}(T_{n_k}+t)\right\|}{\left\|\bar{x}(T_{n_k})\right\|} = 0.$

Proof We first analyze the numerator. $\forall k$,

$$
\begin{aligned}
&\left| \|\bar{x}(T_{n_k+1})\| - \lim_{t\to T^-} \|\bar{x}(T_{n_k} + t)\| \right| \\
=& \lim_{t\to T^-} \left| \|\bar{x}(T_{n_k+1})\| - \|\bar{x}(T_{n_k} + t)\| \right| \\
\leq& \lim_{t\to T^-} \|\bar{x}(T_{n_k+1}) - \bar{x}(T_{n_k} + t)\| \\
=& \lim_{t\to T^-} \left\| \bar{x}(T_{n_k}) + \sum_{i=m(T_{n_k})}^{m(T_{n_k+1})-1} \alpha(i)H(\bar{x}(t(i)), Y_{i+1}) - \bar{x}(T_{n_k}) - \sum_{i=m(T_{n_k})}^{m(T_{n_k}+t)-1} \alpha(i)H(\bar{x}(t(i)), Y_{i+1}) \right\| \\
=& \lim_{t\to T^-} \left\| \sum_{i=m(T_{n_k}+t)}^{m(T_{n_k+1})-1} \alpha(i)H(\bar{x}(t(i)), Y_{i+1}) \right\|
\end{aligned}
$$

$$
\begin{aligned}
&\leq \lim_{t\to T^-}\left\|\sum_{i=m(T_{n_k}+t)}^{m(T_{n_k+1})-1}\alpha(i)\left[H(\bar{x}(t(i)),Y_{i+1})-H(0,Y_{i+1})\right]\right\|+\left\|\sum_{i=m(T_{n_k}+t)}^{m(T_{n_k+1})-1}\alpha(i)H(0,Y_{i+1})\right\|\\
&\leq \lim_{t\to T^-}\sum_{i=m(T_{n_k}+t)}^{m(T_{n_k+1})-1}\alpha(i)L(Y_{i+1})\|\bar{x}(t(i)\|+\left\|\sum_{i=m(T_{n_k}+t)}^{m(T_{n_k+1})-1}\alpha(i)H(0,Y_{i+1})\right\|\\
&=\|\bar{x}(t(m(T_{n_k+1})-1)\|\left[\lim_{t\to T^-}\sum_{i=m(T_{n_k}+t)}^{m(T_{n_k+1})-1}\alpha(i)L(Y_{i+1})\right]+\lim_{t\to T^-}\left\|\sum_{i=m(T_{n_k}+t)}^{m(T_{n_k+1})-1}\alpha(i)H(0,Y_{i+1})\right\|\\
&\hspace{20em}(\forall k,\lim_{t\to T^-}m(T_{n_k}+t)=m(T_{n_k+1})-1)\\
&\leq\left(\left[\|\bar{x}(T_{n_k})\|C_H+C_H\right]e^{C_H}+\|\bar{x}(T_{n_k})\|\right)\left[\lim_{t\to T^-}\sum_{i=m(T_{n_k}+t)}^{m(T_{n_k+1})-1}\alpha(i)L(Y_{i+1})\right]\\
&\quad+\lim_{t\to T^-}\left\|\sum_{i=m(T_{n_k}+t)}^{m(T_{n_k+1})-1}\alpha(i)H(0,Y_{i+1})\right\|. \qquad\text{(by (99))}
\end{aligned}
$$

By (59), we have

$$
\lim_{k\to\infty}\|\bar{x}(T_{n_k})\|=\lim_{k\to\infty}r_{n_k}=\infty. \tag{106}
$$

Thus,

$$
\begin{aligned}
&\lim_{k\to\infty}\left|\frac{\|\bar{x}(T_{n_k+1})\|-\lim_{t\to T^-}\|\bar{x}(T_{n_k}+t)\|}{\|\bar{x}(T_{n_k})\|}\right|\\
&=\lim_{k\to\infty}\frac{|\|\bar{x}(T_{n_k+1})\|-\lim_{t\to T^-}\|\bar{x}(T_{n_k}+t)\||}{\|\bar{x}(T_{n_k})\|}\\
&=\lim_{k\to\infty}\frac{\left(\left[\|\bar{x}(T_{n_k})\|C_H+C_H\right]e^{C_H}+\|\bar{x}(T_{n_k})\|\right)\left[\lim_{t\to T^-}\sum_{i=m(T_{n_k}+t)}^{m(T_{n_k+1})-1}\alpha(i)L(Y_{i+1})\right]}{\|\bar{x}(T_{n_k})\|}\\
&\quad+\lim_{k\to\infty}\frac{\lim_{t\to T^-}\left\|\sum_{i=m(T_{n_k}+t)}^{m(T_{n_k+1})-1}\alpha(i)H(0,Y_{i+1})\right\|}{\|\bar{x}(T_{n_k})\|}\\
&\leq\left(C_He^{C_H}+1\right)\left[\lim_{k\to\infty}\lim_{t\to T^-}\sum_{i=m(T_{n_k}+t)}^{m(T_{n_k+1})-1}\alpha(i)L(Y_{i+1})\right]+\lim_{k\to\infty}\frac{\lim_{t\to T^-}\left\|\sum_{i=m(T_{n_k}+t)}^{m(T_{n_k+1})-1}\alpha(i)H(0,Y_{i+1})\right\|}{\|\bar{x}(T_{n_k})\|}\\
&\hspace{30em}\text{(by (106))}\\
&\leq\left(C_He^{C_H}+1\right)\cdot 0+0 \qquad\text{(by (104) and (105))}\\
&=0.
\end{aligned}
$$

■

Appendix D. Proofs for Completeness

Proofs in this section have used ideas and sketches from Kushner and Yin (2003) but are self-contained and complete.

D.1 Proof of Lemma 9

Proof
Case 1: Let Assumptions 1, 2, 4, and 6 hold.
Fixed an arbitrary $\tau > 0$. For an arbitrary $x, t \in (-\infty, \infty)$, define

$$\begin{aligned}
\psi(i) &\doteq H(x, Y_{i+1}) - h(x), \\
S(n) &\doteq \sum_{i=0}^{n-1} \psi(i), \\
\Psi(t) &\doteq \sum_{i=0}^{m(t)-1} \alpha(i)\psi(i).
\end{aligned}$$

Here, we use (18) so that $\forall t < 0, m(t) = 0$ and the convention that $\sum_{k=i}^{j} \alpha(k) = 0$ when $j < i$. Fix a sample path $\{x_0, \{Y_i\}_{i=1}^{\infty}\}$ where Assumptions 1, 2, 4, & 6 hold. Assumption 6 implies that

$$\lim_{n\to\infty} \alpha(n) S(n+1) = 0.$$

Use subscript j to denote the jth dimension of a vector, we then have

$$\limsup_{n\to\infty} \sup_{-\tau\le t\le\tau} |\alpha(m(t(n)+t))S(m(t(n)+t)+1)_j| = 0. \tag{107}$$

Moreover, for $\forall t \in [-\tau, \tau]$, we have

$$\begin{aligned}
\Psi(t) =& \sum_{i=0}^{m(t)-1} \alpha(i)\psi(i) \\
=& \sum_{i=0}^{m(t)-1} \alpha(i) \left[\sum_{j=0}^{i} \psi(j) - \sum_{j=0}^{i-1} \psi(j)\right] \\
=& \sum_{i=0}^{m(t)-1} \alpha(i) \sum_{j=0}^{i} \psi(j) - \sum_{i=0}^{m(t)-1} \alpha(i) \sum_{j=0}^{i-1} \psi(j) \\
=& \sum_{i=0}^{m(t)-1} \alpha(i) \sum_{j=0}^{i} \psi(j) - \sum_{i=0}^{m(t)-2} \alpha(i+1) \sum_{j=0}^{i} \psi(j) \\
=& \alpha(m(t)-1) \sum_{i=0}^{m(t)-1} \psi(i) + \sum_{i=0}^{m(t)-2} [\alpha(i) - \alpha(i+1)] \sum_{j=0}^{i} \psi(j) \\
=& \alpha(m(t)-1) \sum_{i=0}^{m(t)-1} \psi(i) + \sum_{i=0}^{m(t)-2} S(i+1)[\alpha(i) - \alpha(i+1)]
\end{aligned}$$

$$=\alpha(m(t)-1)S(m(t))+\sum_{i=0}^{m(t)-2}S(i+1)\frac{\alpha(i)-\alpha(i+1)}{\alpha(i)}\alpha(i). \tag{108}$$

Thus, for any dimension j,

$$\begin{aligned}
&\limsup_{n\to\infty}\sup_{-\tau\le t_1\le t_2\le\tau}\left|\sum_{i=m(t(n)+t_1)}^{m(t(n)+t_2)-1}\alpha(i)(H(x,Y_{i+1})_j-h(x)_j)\right|\\
=&\limsup_{n\to\infty}\sup_{-\tau\le t_1\le t_2\le\tau}|\Psi(t(n)+t_2)_j-\Psi(t(n)+t_1)_j|\\
\le&\limsup_{n\to\infty}\sup_{-\tau\le t_1\le t_2\le\tau}|\alpha(m(t(n)+t_2)-1)S(m(t(n)+t_2))_j|+|\alpha(m(t(n)+t_1)-1)S(m(t(n)+t_1))_j|\\
&+\left|\sum_{i=m(t(n)+t_1)-1}^{m(t(n)+t_2)-2}S(i+1)_j\frac{\alpha(i)-\alpha(i+1)}{\alpha(i)}\alpha(i)\right| \qquad \text{(by (108))}\\
=&\limsup_{n\to\infty}\sup_{-\tau\le t_1\le t_2\le\tau}\left|\sum_{i=m(t(n)+t_1)-1}^{m(t(n)+t_2)-2}S(i+1)_j\frac{\alpha(i)-\alpha(i+1)}{\alpha(i)}\alpha(i)\right| \qquad \text{(by (107))}\\
\le&\limsup_{n\to\infty}\sup_{-\tau\le t_1\le t_2\le\tau}\sum_{i=m(t(n)+t_1)-1}^{m(t(n)+t_2)-2}\left|S(i+1)_j\frac{\alpha(i)-\alpha(i+1)}{\alpha(i)}\alpha(i)\right|\\
\le&\limsup_{n\to\infty}\sup_{-\tau\le t_1\le t_2\le\tau}\sum_{i=m(t(n)+t_1)-1}^{m(t(n)+t_2)-2}|\alpha(i)S(i+1)_j|\left|\frac{\alpha(i)-\alpha(i+1)}{\alpha(i)}\right|\\
\le&\limsup_{n\to\infty}\sup_{-\tau\le t_1\le t_2\le\tau}\left(\sup_{m(t(n)+t_1)-1\le i\le m(t(n)+t_2)-2}|\alpha(i)S(i+1)_j|\right)\sum_{i=m(t(n)+t_1)-1}^{m(t(n)+t_2)-2}\left|\frac{\alpha(i)-\alpha(i+1)}{\alpha(i)}\right|\\
\le&\limsup_{n\to\infty}\sup_{-\tau\le t_1\le t_2\le\tau}\left(\sup_{m(t(n)+t_1)-1\le i\le m(t(n)+t_2)-2}|\alpha(i)S(i+1)_j|\right)C_\alpha\sum_{i=m(t(n)+t_1)-1}^{m(t(n)+t_2)-2}\alpha(i)\\
&\qquad\text{(by Assumption 2, } C_\alpha \text{ is a constant from the big } \mathcal{O} \text{ notation–}\frac{\alpha(n)-\alpha(n+1)}{\alpha(n)}=\mathcal{O}\left(\alpha(n)\right)\text{)}\\
=&\limsup_{n\to\infty}\left[\sup_{-\tau\le t_1\le t_2\le\tau}\left(\sup_{m(t(n)+t_1)-1\le i\le m(t(n)+t_2)-2}|\alpha(i)S(i+1)_j|\right)\right.\\
&\left.\cdot C_\alpha\left(\sum_{i=m(t(n)+t_1)}^{m(t(n)+t_2)-1}\alpha(i)+\alpha(m(t(n)+t_1)-1)\right)\right]\\
=&\limsup_{n\to\infty}\sup_{-\tau\le t_1\le t_2\le\tau}\left(\sup_{m(t(n)+t_1)-1\le i\le m(t(n)+t_2)-2}|\alpha(i)S(i+1)_j|\right)C_\alpha\left(t_2-t_1+\alpha(m(t(n)+t_1)-1)\right)\\
&\qquad\text{(by (82))}
\end{aligned}$$

$$\begin{aligned}
&\leq \limsup_{n\to\infty} \left(\sup_{m(t(n)-\tau)-1\leq i} |\alpha(i)S(i+1)_j| \right) C_\alpha \left(t_2 - t_1 + \alpha(m(t(n)+t_1)-1)\right) \\
&\leq 2C_\alpha\tau \limsup_{n\to\infty} \left(\sup_{m(t(n)-\tau)-1\leq i} |\alpha(i)S(i+1)_j| \right) \\
&\leq 2C_\alpha\tau \limsup_{n\to\infty} \left(\sup_{n\leq i} |\alpha(i)S(i+1)_j| \right) \\
&= 0. && \text{(by (107))}
\end{aligned}$$

Thus, $\forall \tau > 0$, $\forall x$,

$$\limsup_{n} \sup_{-\tau\leq t_1\leq t_2\leq\tau} \left\| \sum_{i=m(t(n)+t_1)}^{m(t(n)+t_2)-1} \alpha(i)\left[H(x, Y_{i+1}) - h(x)\right] \right\| = 0 \quad a.s.$$

The proofs for (19) and (20) follow the same logic and thus are omitted.
Case 2: Let Assumptions 1, 2, 4, and 6′ hold.
By Assumption 4 and the equivalence between norms, we have

$$\|H(x,y)\|_2 \leq C\left(\|H(0,y)\|_2 + L(y)\|x\|_2\right)$$

for some constant C independent of x, y. So for any x,

$$\sup_y \frac{\|H(x,y)\|_2^2}{v(y)} \leq \sup_y \frac{2C^2\|H(0,y)\|_2^2 + 2C^2L(y)^2\|x\|_2^2}{v(y)} < \infty.$$

In other words, for any x,

$$y \mapsto H(x,y) \in \mathcal{L}^2_{v,\infty}.$$

Similarly, we have for any x,

$$y \mapsto L_b(y) \in \mathcal{L}^2_{v,\infty}.$$

Let g denote any of the following functions:

$$\begin{aligned}
y \mapsto & H(x,y) \quad (\forall x), \\
y \mapsto & L_b(y) \quad (\forall x), \\
y \mapsto & L(y).
\end{aligned}$$

We now always have $g \in \mathcal{L}^2_{v,\infty}$. Proposition 6 of Borkar et al. (2021) then confirms that

$$\sum_{i=0}^{\infty} \alpha(i)\left(g(Y_{i+1}) - \mathbb{E}_{y\sim d_{\mathcal{Y}}}\left[g(y)\right]\right)$$

converges almost surely to a square-integrable random variable. Lemma 9 then follows immediately from the Cauchy convergence test.

■

D.2 Proof of Lemma 18

To prove Lemma 18, we first decompose it into three terms. Then, we prove the convergence of each term in Lemmas 53, 54, & 55. Finally, we restate Lemma 18 and connect everything.

For each t, let $\{\Delta_l\}_{l=1}^{\infty}$ be a strictly decreasing sequence of real numbers such that $\lim_{l\to\infty}\Delta_l = 0$ and $\forall l$, $\frac{t}{\Delta_l} - 1 \in \mathbb{N}$, e.g., $\Delta_l \doteq \frac{t}{l+1}$. Because $\forall l$,

$$\sum_{i=m(T_{n_k})}^{m(T_{n_k}+t)-1} \alpha(i) H_{r_{n_j}}(\hat{x}(t(i)), Y_{i+1}) = \sum_{a=0}^{\frac{t}{\Delta_l}-1} \sum_{i=m(T_{n_k}+a\Delta_l)}^{m(T_{n_k}+a\Delta_l+\Delta_l)-1} \alpha(i) H_{r_{n_j}}(\hat{x}(t(i)), Y_{i+1}),$$

we have

$$\lim_{k\to\infty} \left\| \sum_{i=m(T_{n_k})}^{m(T_{n_k}+t)-1} \alpha(i) H_{r_{n_j}}(\hat{x}(t(i)), Y_{i+1}) - \int_0^t h_{r_{n_j}}(\hat{x}^{\lim}(s)) ds \right\| \tag{109}$$

$$= \lim_{l\to\infty}\lim_{k\to\infty} \left\| \sum_{a=0}^{\frac{t}{\Delta_l}-1} \sum_{i=m(T_{n_k}+a\Delta_l)}^{m(T_{n_k}+a\Delta_l+\Delta_l)-1} \alpha(i) H_{r_{n_j}}(\hat{x}(t(i)), Y_{i+1}) - \int_0^t h_{r_{n_j}}(\hat{x}^{\lim}(s)) ds \right\|$$

$$\le \lim_{l\to\infty}\lim_{k\to\infty} \left\| \sum_{a=0}^{\frac{t}{\Delta_l}-1} \sum_{i=m(T_{n_k}+a\Delta_l)}^{m(T_{n_k}+a\Delta_l+\Delta_l)-1} \alpha(i) h_{r_{n_j}}(\hat{x}^{\lim}(a\Delta_l)) - \int_0^t h_{r_{n_j}}(\hat{x}^{\lim}(s)) ds \right\| \tag{110}$$

$$+ \lim_{l\to\infty}\lim_{k\to\infty} \left\| \sum_{a=0}^{\frac{t}{\Delta_l}-1} \sum_{i=m(T_{n_k}+a\Delta_l)}^{m(T_{n_k}+a\Delta_l+\Delta_l)-1} \alpha(i) \left(H_{r_{n_j}}(\hat{x}(t(i)), Y_{i+1}) - H_{r_{n_j}}(\hat{x}^{\lim}(a\Delta_l), Y_{i+1}) \right) \right\| \tag{111}$$

$$+ \lim_{l\to\infty}\lim_{k\to\infty} \left\| \sum_{a=0}^{\frac{t}{\Delta_l}-1} \sum_{i=m(T_{n_k}+a\Delta_l)}^{m(T_{n_k}+a\Delta_l+\Delta_l)-1} \alpha(i) \left(H_{r_{n_j}}(\hat{x}^{\lim}(a\Delta_l), Y_{i+1}) - h_{r_{n_j}}(\hat{x}^{\lim}(a\Delta_l)) \right) \right\|. \tag{112}$$

Now, we show the limit of (110), (111), and (112) are 0 in Lemmas 53, 54, and 55 with proofs in Appendix D.3, D.4, and D.5.

Lemma 53 $\forall j, \forall t \in [0, T)$,

$$\lim_{l\to\infty}\lim_{k\to\infty} \left\| \sum_{a=0}^{\frac{t}{\Delta_l}-1} \sum_{i=m(T_{n_k}+a\Delta_l)}^{m(T_{n_k}+a\Delta_l+\Delta_l)-1} \alpha(i) h_{r_{n_j}}(\hat{x}^{\lim}(a\Delta_l)) - \int_0^t h_{r_{n_j}}(\hat{x}^{\lim}(s)) \right\| = 0.$$

Lemma 54 $\forall j, \forall t \in [0, T)$,

$$\lim_{l\to\infty}\lim_{k\to\infty} \left\| \sum_{a=0}^{\frac{t}{\Delta_l}-1} \sum_{i=m(T_{n_k}+a\Delta_l)}^{m(T_{n_k}+a\Delta_l+\Delta_l)-1} \alpha(i) \left(H_{r_{n_j}}(\hat{x}(t(i)), Y_{i+1}) - H_{r_{n_j}}(\hat{x}^{\lim}(a\Delta_l), Y_{i+1}) \right) \right\| = 0.$$

Lemma 55 $\forall j, \forall t \in [0, T)$,

$$\lim_{l\to\infty}\lim_{k\to\infty}\left\|\sum_{a=0}^{\frac{t}{\Delta_l}-1}\sum_{i=m(T_{n_k}+a\Delta_l)}^{m(T_{n_k}+a\Delta_l+\Delta_l)-1}\alpha(i)\left(H_{r_{n_j}}(\hat{x}^{\lim}(a\Delta_l), Y_{i+1}) - h_{r_{n_j}}(\hat{x}^{\lim}(a\Delta_l))\right)\right\| = 0.$$

Plugging Lemmas 53, 54, and 55 back to (109) completes the proof of Lemma 18.

D.3 Proof of Lemma 53

Proof $\forall j, \forall t \in [0, T)$,

$$\begin{aligned}
&\lim_{l\to\infty}\lim_{k\to\infty}\sum_{a=0}^{\frac{t}{\Delta_l}-1}\sum_{i=m(T_{n_k}+a\Delta_l)}^{m(T_{n_k}+a\Delta_l+\Delta_l)-1}\alpha(i)h_{r_{n_j}}(\hat{x}^{\lim}(a\Delta_l))\\
=&\lim_{l\to\infty}\sum_{a=0}^{\frac{t}{\Delta_l}-1}h_{r_{n_j}}(\hat{x}^{\lim}(a\Delta_l))\lim_{k\to\infty}\sum_{i=m(T_{n_k}+a\Delta_l)}^{m(T_{n_k}+a\Delta_l+\Delta_l)-1}\alpha(i)\\
=&\lim_{l\to\infty}\sum_{a=0}^{\frac{t}{\Delta_l}-1}h_{r_{n_j}}(\hat{x}^{\lim}(a\Delta_l))\Delta_l && \text{(by (82))}\\
=&\int_0^t h_{r_{n_j}}(\hat{x}^{\lim}(s))ds. && \text{(by definition of integral)}
\end{aligned}$$

Thus, $\forall j, \forall t \in [0, T)$,

$$\begin{aligned}
&\lim_{l\to\infty}\lim_{k\to\infty}\left\|\sum_{a=0}^{\frac{t}{\Delta_l}-1}\sum_{i=m(T_{n_k}+a\Delta_l)}^{m(T_{n_k}+a\Delta_l+\Delta_l)-1}\alpha(i)h_{r_{n_j}}(\hat{x}^{\lim}(a\Delta_l)) - \int_0^t h_{r_{n_j}}(\hat{x}^{\lim}(s))\right\|\\
=&\left\|\lim_{l\to\infty}\lim_{k\to\infty}\sum_{a=0}^{\frac{t}{\Delta_l}-1}\sum_{i=m(T_{n_k}+a\Delta_l)}^{m(T_{n_k}+a\Delta_l+\Delta_l)-1}\alpha(i)h_{r_{n_j}}(\hat{x}^{\lim}(a\Delta_l)) - \int_0^t h_{r_{n_j}}(\hat{x}^{\lim}(s))\right\|\\
=&\left\|\int_0^t h_{r_{n_j}}(\hat{x}^{\lim}(s)) - \int_0^t h_{r_{n_j}}(\hat{x}^{\lim}(s))\right\|\\
=&0.
\end{aligned}$$

■

D.4 Proof of Lemma 54

Proof $\forall j, \forall t \in [0, T), \forall l$

$$\lim_{k\to\infty}\left\|\sum_{a=0}^{\frac{t}{\Delta_l}-1}\sum_{i=m(T_{n_k}+a\Delta_l)}^{m(T_{n_k}+a\Delta_l+\Delta_l)-1}\alpha(i)\left(H_{r_{n_j}}(\hat{x}(t(i)), Y_{i+1}) - H_{r_{n_j}}(\hat{x}^{\lim}(a\Delta_l), Y_{i+1})\right)\right\|$$

$$\leq \lim_{k\to\infty} \sum_{a=0}^{\frac{t}{\Delta_l}-1} \sum_{i=m(T_{n_k}+a\Delta_l)}^{m(T_{n_k}+a\Delta_l+\Delta_l)-1} \alpha(i) \left\| H_{r_{n_j}}(\hat{x}(t(i)), Y_{i+1}) - H_{r_{n_j}}(\hat{x}^{\lim}(a\Delta_l), Y_{i+1}) \right\|$$

$$\leq \lim_{k\to\infty} \sum_{a=0}^{\frac{t}{\Delta_l}-1} \sum_{i=m(T_{n_k}+a\Delta_l)}^{m(T_{n_k}+a\Delta_l+\Delta_l)-1} \alpha(i) L(Y_{i+1}) \left\| \hat{x}(t(i)) - \hat{x}^{\lim}(a\Delta_l) \right\| \qquad \text{(by Assumption 4)}$$

$$\leq \lim_{k\to\infty} \left[\sup_{0\leq a\leq \frac{t}{\Delta_l}-1} \sup_{m(T_{n_k}+a\Delta_l)\leq i\leq m(T_{n_k}+a\Delta_l+\Delta_l)-1} \left\| \hat{x}(t(i)) - \hat{x}^{\lim}(a\Delta_l) \right\| \right] \sum_{a=0}^{\frac{t}{\Delta_l}-1} \sum_{i=m(T_{n_k}+a\Delta_l)}^{m(T_{n_k}+a\Delta_l+\Delta_l)-1} \alpha(i) L(Y_{i+1})$$

$$= \lim_{k\to\infty} \left[\sup_{0\leq a\leq \frac{t}{\Delta_l}-1} \sup_{m(T_{n_k}+a\Delta_l)\leq i\leq m(T_{n_k}+a\Delta_l+\Delta_l)-1} \left\| \hat{x}(t(i)) - \hat{x}^{\lim}(a\Delta_l) \right\| \right] \sum_{i=m(T_{n_k})}^{m(T_{n_k}+t)-1} \alpha(i) L(Y_{i+1}). \tag{113}$$

We show the limit of the following term.

$$\lim_{k\to\infty} \left[\sup_{0\leq a\leq \frac{t}{\Delta_l}-1} \sup_{m(T_{n_k}+a\Delta_l)\leq i\leq m(T_{n_k}+a\Delta_l+\Delta_l)-1} \left\| \hat{x}(t(i)) - \hat{x}^{\lim}(a\Delta_l) \right\| \right]$$

$$= \lim_{k\to\infty} \left[\sup_{0\leq a\leq \frac{t}{\Delta_l}-1} \sup_{t(m(T_{n_k}+a\Delta_l))\leq t(i)\leq t(m(T_{n_k}+a\Delta_l+\Delta_l)-1)} \left\| \hat{x}(t(i)) - \hat{x}^{\lim}(a\Delta_l) \right\| \right]$$

$$\leq \lim_{k\to\infty} \left[\sup_{0\leq a\leq \frac{t}{\Delta_l}-1} \sup_{t(m(T_{n_k}+a\Delta_l))\leq \tau\leq t(m(T_{n_k}+a\Delta_l+\Delta_l)-1)} \left\| \hat{x}(\tau) - \hat{x}^{\lim}(a\Delta_l) \right\| \right]$$

$$= \lim_{k\to\infty} \left[\sup_{0\leq a\leq \frac{t}{\Delta_l}-1} \sup_{T_{n_k}+a\Delta_l\leq \tau\leq t(m(T_{n_k}+a\Delta_l+\Delta_l)-1)} \left\| \hat{x}(\tau) - \hat{x}^{\lim}(a\Delta_l) \right\| \right]$$

($\hat{x}$ is a constant function on interval $[t(m(T_{n_k}+a\Delta_l)), T_{n_k}+a\Delta_l]$ by (21) and (22))

$$\leq \lim_{k\to\infty} \left[\sup_{0\leq a\leq \frac{t}{\Delta_l}-1} \sup_{T_{n_k}+a\Delta_l\leq \tau< T_{n_k}+a\Delta_l+\Delta_l} \left\| \hat{x}(\tau) - \hat{x}^{\lim}(a\Delta_l) \right\| \right] \qquad \text{(by (16))}$$

$$= \lim_{k\to\infty} \left[\sup_{0\leq a\leq \frac{t}{\Delta_l}-1} \sup_{0\leq \tau<\Delta_l} \left\| \hat{x}(T_{n_k}+a\Delta_l+\tau) - \hat{x}^{\lim}(a\Delta_l) \right\| \right]. \tag{114}$$

By (33), $\forall \delta > 0$, $\exists k_0$ such that $\forall k \geq k_0$, $\forall t \in [0, T)$,

$$\left\| \hat{x}(T_{n_k}+t) - \hat{x}^{\lim}(t) \right\| \leq \delta.$$

$\forall t \in [0, T), \forall l, \forall a, \forall k \geq k_0,$

$$\left| \sup_{0 \leq a \leq \frac{t}{\Delta_l} - 1} \sup_{0 \leq \tau < \Delta_l} \left\| \hat{x}(T_{n_k} + a\Delta_l + \tau) - \hat{x}^{\lim}(a\Delta_l) \right\| - \sup_{0 \leq a \leq \frac{t}{\Delta_l} - 1} \sup_{0 \leq \tau < \Delta_l} \left\| \hat{x}^{\lim}(a\Delta_l + \tau) - \hat{x}^{\lim}(a\Delta_l) \right\| \right|$$

$$\begin{aligned}
&\leq \sup_{0 \leq a \leq \frac{t}{\Delta_l} - 1} \sup_{0 \leq \tau < \Delta_l} \left| \left\| \hat{x}(T_{n_k} + a\Delta_l + \tau) - \hat{x}^{\lim}(a\Delta_l) \right\| - \left\| \hat{x}^{\lim}(a\Delta_l + \tau) - \hat{x}^{\lim}(a\Delta_l) \right\| \right| \\
&\qquad\qquad (\text{by } |\sup_x f(x) - \sup_x g(x)| \leq \sup_x |f(x) - g(x)|) \\
&\leq \sup_{0 \leq a \leq \frac{t}{\Delta_l} - 1} \sup_{0 \leq \tau < \Delta_l} \left\| \hat{x}(T_{n_k} + a\Delta_l + \tau) - \hat{x}^{\lim}(a\Delta_l) - \hat{x}^{\lim}(a\Delta_l + \tau) + \hat{x}^{\lim}(a\Delta_l) \right\| \\
&\leq \sup_{0 \leq a \leq \frac{t}{\Delta_l} - 1} \sup_{0 \leq \tau < \Delta_l} \left\| \hat{x}(T_{n_k} + a\Delta_l + \tau) - \hat{x}^{\lim}(a\Delta_l + \tau) \right\| \\
&\leq \sup_{0 \leq a \leq \frac{t}{\Delta_l} - 1} \sup_{0 \leq \tau < \Delta_l} \delta \\
&\leq \delta.
\end{aligned}$$

Thus, $\forall t \in [0, T), \forall l, \forall a,$

$$\lim_{k \to \infty} \sup_{0 \leq a \leq \frac{t}{\Delta_l} - 1} \sup_{0 \leq \tau < \Delta_l} \left\| \hat{x}(T_{n_k} + a\Delta_l + \tau) - \hat{x}^{\lim}(a\Delta_l) \right\| = \sup_{0 \leq a \leq \frac{t}{\Delta_l} - 1} \sup_{0 \leq \tau < \Delta_l} \left\| \hat{x}^{\lim}(a\Delta_l + \tau) - \hat{x}^{\lim}(a\Delta_l) \right\|.$$

Therefore,

$$\begin{aligned}
&\lim_{k \to \infty} \left[\sup_{0 \leq a \leq \frac{t}{\Delta_l} - 1} \sup_{m(T_{n_k} + a\Delta_l) \leq i \leq m(T_{n_k} + a\Delta_l + \Delta_l) - 1} \left\| \hat{x}(t(i)) - \hat{x}^{\lim}(a\Delta_l) \right\| \right] \\
&= \lim_{k \to \infty} \sup_{0 \leq a \leq \frac{t}{\Delta_l} - 1} \sup_{0 \leq \tau < \Delta_l} \left\| \hat{x}(T_{n_k} + a\Delta_l + \tau) - \hat{x}^{\lim}(a\Delta_l) \right\| && (\text{by } (114)) \\
&= \sup_{0 \leq a \leq \frac{t}{\Delta_l} - 1} \sup_{0 \leq \tau < \Delta_l} \left\| \hat{x}^{\lim}(a\Delta_l + \tau) - \hat{x}^{\lim}(a\Delta_l) \right\|. && (115)
\end{aligned}$$

$\forall j, \forall t \in [0, T), \forall l,$

$$\begin{aligned}
&\lim_{k \to \infty} \left\| \sum_{a=0}^{\frac{t}{\Delta_l} - 1} \sum_{i=m(T_{n_k} + a\Delta_l)}^{m(T_{n_k} + a\Delta_l + \Delta_l) - 1} \alpha(i) \left(H_{r_{n_j}}(\hat{x}(t(i)), Y_{i+1}) - H_{r_{n_j}}(\hat{x}^{\lim}(a\Delta_l), Y_{i+1}) \right) \right\| \\
&\leq \lim_{k \to \infty} \left[\sup_{0 \leq a \leq \frac{t}{\Delta_l} - 1} \sup_{m(T_{n_k} + a\Delta_l) \leq i \leq m(T_{n_k} + a\Delta_l + \Delta_l) - 1} \left\| \hat{x}(t(i)) - \hat{x}^{\lim}(a\Delta_l) \right\| \right] \sum_{i=m(T_{n_k})}^{m(T_{n_k} + t) - 1} \alpha(i) L(Y_{i+1}) \\
&\qquad\qquad (\text{by } (113))
\end{aligned}$$

$$
\begin{aligned}
&\leq \lim_{k\to\infty}\left[\sup_{0\leq a\leq\frac{t}{\Delta_l}-1}\ \sup_{m(T_{n_k}+a\Delta_l)\leq i\leq m(T_{n_k}+a\Delta_l+\Delta_l)-1}\left\|\hat{x}(t(i))-\hat{x}^{\lim}(a\Delta_l)\right\|\right]\\
&\quad\limsup_{k\to\infty}\left[\sum_{i=m(T_{n_k})}^{m(T_{n_k}+t)-1}\alpha(i)L(Y_{i+1})\right]\\
&\leq \lim_{k\to\infty}\left[\sup_{0\leq a\leq\frac{t}{\Delta_l}-1}\ \sup_{m(T_{n_k}+a\Delta_l)\leq i\leq m(T_{n_k}+a\Delta_l+\Delta_l)-1}\left\|\hat{x}(t(i))-\hat{x}^{\lim}(a\Delta_l)\right\|\right]C_H \qquad \text{(by (97))}\\
&=C_H\sup_{0\leq a\leq\frac{t}{\Delta_l}-1}\ \sup_{0\leq\tau<\Delta_l}\left\|\hat{x}^{\lim}(a\Delta_l+\tau)-\hat{x}^{\lim}(a\Delta_l)\right\|. \qquad \text{(by (115))} \qquad (116)
\end{aligned}
$$

By Corollary 27, $\hat{x}^{\lim}$ is continuous and $[0,t]$ is a compact set, $\forall\epsilon>0,\exists\eta$ such that

$$
\sup_{0\leq|t_1-t_2|\leq\eta,t_1\in[0,t],t_2\in[0,t]}\left\|\hat{x}^{\lim}(t_1)-\hat{x}^{\lim}(t_2)\right\|\leq\epsilon. \tag{117}
$$

Thus, $\forall\epsilon>0$, $\exists l_0$ such that $\forall l\geq l_0,\Delta_l\leq\eta$ and we will have

$$
0\leq\sup_{0\leq a\leq\frac{t}{\Delta_l}-1}\ \sup_{0\leq\tau<\Delta_l}\left\|\hat{x}^{\lim}(a\Delta_l+\tau)-\hat{x}^{\lim}(a\Delta_l)\right\|\leq\epsilon. \qquad \text{(by (117))}
$$

Therefore, $\forall t$,

$$
\lim_{l\to\infty}\left[\sup_{0\leq a\leq\frac{t}{\Delta_l}-1}\ \sup_{0\leq\tau<\Delta_l}\left\|\hat{x}^{\lim}(a\Delta_l+\tau)-\hat{x}^{\lim}(a\Delta_l)\right\|\right]=0. \tag{118}
$$

This concludes $\forall j,\forall t\in[0,T)$,

$$
\begin{aligned}
&\lim_{l\to\infty}\lim_{k\to\infty}\left\|\sum_{a=0}^{\frac{t}{\Delta_l}-1}\sum_{i=m(T_{n_k}+a\Delta_l)}^{m(T_{n_k}+a\Delta_l+\Delta_l)-1}\alpha(i)\left(H_{r_{n_j}}(\hat{x}(t(i)),Y_{i+1})-H_{r_{n_j}}(\hat{x}^{\lim}(a\Delta_l),Y_{i+1})\right)\right\|\\
&=\lim_{l\to\infty}C_H\sup_{0\leq a\leq\frac{t}{\Delta_l}-1}\ \sup_{0\leq\tau<\Delta_l}\left\|\hat{x}^{\lim}(a\Delta_l+\tau)-\hat{x}^{\lim}(a\Delta_l)\right\| \qquad \text{(by (116))}\\
&=C_H\cdot 0 \qquad \text{(by (118))}\\
&=0.
\end{aligned}
$$

■

D.5 Proof of Lemma 55

Proof By (86), $\forall j,\forall a,\forall l$,

$$
\lim_{k\to\infty}\left\|\sum_{i=m(T_{n_k}+a\Delta_l)}^{m(T_{n_k}+a\Delta_l+\Delta_l)-1}\alpha(i)\left[H_{r_{n_j}}(\hat{x}^{\lim}(a\Delta_l),Y_{i+1})-h_{r_{n_j}}(\hat{x}^{\lim}(a\Delta_l))\right]\right\|=0. \tag{119}
$$

Thus,$\forall j, \forall t \in [0, T)$,

$$
\begin{aligned}
&\lim_{l\to\infty}\lim_{k\to\infty}\left\|\sum_{a=0}^{\frac{t}{\Delta_l}-1}\sum_{i=m(T_{n_k}+a\Delta_l)}^{m(T_{n_k}+a\Delta_l+\Delta_l)-1}\alpha(i)\left[H_{r_{n_j}}(\hat{x}^{\lim}(a\Delta_l),Y_{i+1})-h_{r_{n_j}}(\hat{x}^{\lim}(a\Delta_l))\right]\right\| \\
\le&\lim_{l\to\infty}\sum_{a=0}^{\frac{t}{\Delta_l}-1}\lim_{k\to\infty}\left\|\sum_{i=m(T_{n_k}+a\Delta_l)}^{m(T_{n_k}+a\Delta_l+\Delta_l)-1}\alpha(i)\left[H_{r_{n_j}}(\hat{x}^{\lim}(a\Delta_l),Y_{i+1})-h_{r_{n_j}}(\hat{x}^{\lim}(a\Delta_l))\right]\right\| \\
=&\lim_{l\to\infty}\sum_{a=0}^{\frac{t}{\Delta_l}-1}0 \qquad \text{(by (119))} \\
=&0.
\end{aligned}
$$

■

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.

Leemon C. Baird. Residual algorithms: Reinforcement learning with function approximation. In *Proceedings of the International Conference on Machine Learning*, 1995.

Andrew G Barto and Richard S Sutton. Goal seeking components for adaptive intelligence: An initial assessment. air force wright aeronautical laboratories. Technical report, Avionics Laboratory Technical Report AFWAL-TR-81-1070, Wright-Patterson AFB, 1981a.

Andrew G Barto and Richard S Sutton. Landmark learning: An illustration of associative search. *Biological Cybernetics*, 1981b.

Andrew G Barto, Richard S Sutton, and Charles W Anderson. Neuronlike adaptive elements that can solve difficult learning control problems. *IEEE Transactions on Systems, Man, and Cybernetics*, 1983.

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

Dimitri P Bertsekas and John N Tsitsiklis. *Neuro-Dynamic Programming*. Athena Scientific Belmont, MA, 1996.

Dimitri P Bertsekas and Huizhen Yu. Projected equation methods for approximate solution of large linear systems. *Journal of Computational and Applied Mathematics*, 2009.

Shalabh Bhatnagar. The borkar-meyn theorem for asynchronous stochastic approximations. *Systems & Control Letters*, 2011.

Patrick Billingsley. *Convergence of probability measures*. Wiley Series in Probability and Statistics: Probability and Statistics, 1999.

Ethan Blaser and Shangtong Zhang. Asymptotic and finite sample analysis of nonexpansive stochastic approximations with markovian noise. *arXiv preprint arXiv:2409.19546*, 2024.

Vivek Borkar, Shuhang Chen, Adithya Devraj, Ioannis Kontoyiannis, and Sean Meyn. The ode method for asymptotic statistics in stochastic approximation and reinforcement learning. *arXiv preprint arXiv:2110.14427*, 2021.

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

Vivek S Borkar and Sean P Meyn. The ode method for convergence of stochastic approximation and reinforcement learning. *SIAM Journal on Control and Optimization*, 2000.

Claire Chen, Shuze Liu, and Shangtong Zhang. Efficient policy evaluation with safety constraint for reinforcement learning. *arXiv preprint arXiv:2410.05655*, 2024.

Jim G Dai. On positive harris recurrence of multiclass queueing networks: a unified approach via fluid limit models. *The Annals of Applied Probability*, 1995.

Jim G Dai and Sean P Meyn. Stability and convergence of moments for multiclass queueing networks via fluid limit models. *IEEE Transactions on Automatic Control*, 1995.

Gabriel Dulac-Arnold, Daniel Mankowitz, and Todd Hester. Challenges of real-world reinforcement learning. *arXiv preprint arXiv:1904.12901*, 2019.

Nelson Dunford and Jacob T Schwartz. *Linear operators, part 1: general theory.* John Wiley & Sons, 1988.

Gersende Fort, Sean Meyn, Eric Moulines, and Pierre Priouret. The ODE method for stability of skip-free Markov chains with applications to MCMC. *The Annals of Applied Probability*, 2008.

Sina Ghiassian, Andrew Patterson, Shivam Garg, Dhawal Gupta, Adam White, and Martha White. Gradient temporal-difference learning with regularized corrections. In *International Conference on Machine Learning*, 2020.

Eric Graves, Ehsan Imani, Raksha Kumaraswamy, and Martha White. Off-policy actor-critic with emphatic weightings. *Journal of Machine Learning Research*, 2023.

Ziwei Guan, Tengyu Xu, and Yingbin Liang. Per-etd: A polynomially efficient emphatic temporal difference learning method. *arXiv preprint arXiv:2110.06906*, 2021.

Assaf Hallak, Aviv Tamar, Rémi Munos, and Shie Mannor. Generalized emphatic temporal difference learning: Bias-variance analysis. In *Proceedings of the AAAI Conference on Artificial Intelligence*, 2016.

Roger A. Horn and Charles R. Johnson. *Topics in Matrix Analysis*. Cambridge University Press, 1991.

Ehsan Imani, Eric Graves, and Martha White. An off-policy policy gradient theorem using emphatic weightings. In *Advances in Neural Information Processing Systems*, 2018.

Ray Jiang, Tom Zahavy, Adam White, Zhongwen Xu, Matteo Hessel, Charles Blundell, and Hado van Hasselt. Emphatic algorithms for deep reinforcement learning. In *Proceedings of the International Conference on Machine Learning*, 2021.

Ray Jiang, Shangtong Zhang, Veronica Chelu, Adam White, and Hado van Hasselt. Learning expected emphatic traces for deep RL. In *Proceedings of the AAAI Conference on Artificial Intelligence*, 2022.

Julian Kern. Skorokhod topologies: What they are and why we should care. *Mathematische Semesterberichte*, 2023.

Hassan K. Khalil. *Nonlinear Systems.* Prentice Hall, 2002.

J. Kiefer and J. Wolfowitz. Stochastic estimation of the maximum of a regression function. *Annals of Mathematical Statistics*, 1952.

A Harry Klopf. *Brain function and adaptive systems: a heterostatic theory*. Air Force Cambridge Research Laboratories, Air Force Systems Command, United States Air Force, 1972.

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

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

Chandrashekar Lakshminarayanan and Shalabh Bhatnagar. A stability criterion for two timescale stochastic approximation schemes. *Automatica*, 2017.

Long Ji Lin. Self-improving reactive agents based on reinforcement learning, planning and teaching. *Machine Learning*, 1992.

Bo Liu, Ji Liu, Mohammad Ghavamzadeh, Sridhar Mahadevan, and Marek Petrik. Finite-sample analysis of proximal gradient TD algorithms. In *Proceedings of the Conference on Uncertainty in Artificial Intelligence*, 2015.

Qiang Liu, Lihong Li, Ziyang Tang, and Dengyong Zhou. Breaking the curse of horizon: Infinite-horizon off-policy estimation. In *Advances in Neural Information Processing Systems*, 2018.

Shuze Liu and Shangtong Zhang. Efficient policy evaluation with offline data informed behavior policy design. In *Proceedings of the International Conference on Machine Learning*, 2024.

Shuze Liu, Claire Chen, and Shangtong Zhang. Doubly optimal policy evaluation for reinforcement learning. *arXiv preprint arXiv:2410.02226*, 2024a.

Shuze Liu, Yuxin Chen, and Shangtong Zhang. Efficient multi-policy evaluation for reinforcement learning. *arXiv preprint arXiv:2408.08706*, 2024b.

Xinyu Liu, Zixuan Xie, and Shangtong Zhang. Linear Q-learning does not diverge: Convergence rates to a bounded set. *arXiv preprint arXiv:2501.19254*, 2025.

Hamid Maei, Csaba Szepesvari, Shalabh Bhatnagar, Doina Precup, David Silver, and Richard S Sutton. Convergent temporal-difference learning with arbitrary smooth function approximation. *Advances in Neural Information Processing Systems*, 2009.

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

Hamid Reza Maei. Convergent actor-critic algorithms under off-policy training and function approximation. *arXiv preprint arXiv:1802.07842*, 2018.

Hamid Reza Maei and Richard S Sutton. GQ (λ): A general gradient algorithm for temporal-difference prediction learning with eligibility traces. In *Proceedings of the Conference on Artificial General Intelligence*, 2010.

Hamid Reza Maei, Csaba Szepesvári, Shalabh Bhatnagar, and Richard S. Sutton. Toward off-policy learning control with function approximation. In *Proceedings of the International Conference on Machine Learning*, 2010.

Sridhar Mahadevan, Bo Liu, Philip S. Thomas, William Dabney, Stephen Giguere, Nicholas Jacek, Ian Gemp, and Ji Liu. Proximal reinforcement learning: A new theory of sequential decision making in primal-dual spaces. *arXiv preprint arXiv:1405.6757*, 2014.

Ashique Rupam Mahmood, Huizhen Yu, and Richard S Sutton. Multi-step off-policy learning without importance sampling ratios. *arXiv preprint arXiv:1702.03006*, 2017.

Michaël Mathieu, Sherjil Ozair, Srivatsan Srinivasan, Caglar Gulcehre, Shangtong Zhang, Ray Jiang, Tom Le Paine, Richard Powell, Konrad Żołna, Julian Schrittwieser, et al. Alphastar unplugged: Large-scale offline reinforcement learning. *arXiv preprint arXiv:2308.03526*, 2023.

Sean Meyn. *Control techniques for complex networks.* Cambridge University Press, 2008.

Sean Meyn. *Control systems and reinforcement learning.* Cambridge University Press, 2022.

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

Alexander P Molchanov and Ye S Pyatnitskiy. Criteria of asymptotic stability of differential and difference inclusions encountered in control theory. *Systems & Control Letters*, 1989.

Ofir Nachum, Yinlam Chow, Bo Dai, and Lihong Li. Dualdice: Behavior-agnostic estimation of discounted stationary distribution corrections. In *Advances in Neural Information Processing Systems*, 2019.

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.

Doina Precup, Richard S. Sutton, and Sanjoy Dasgupta. Off-policy temporal difference learning with function approximation. In *Proceedings of the International Conference on Machine Learning*, 2001.

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 arXiv:2411.13711*, 2024.

Arunselvan Ramaswamy and Shalabh Bhatnagar. A generalization of the borkar-meyn theorem for stochastic recursive inclusions. *Mathematics of Operations Research*, 2017.

Arunselvan Ramaswamy and Shalabh Bhatnagar. Stability of stochastic approximations with "controlled markov" noise and temporal difference learning. *IEEE Transactions on Automatic Control*, 2018.

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

Halsey Lawrence Royden and Patrick Fitzpatrick. *Real analysis*. Macmillan New York, 1968.

Gavin A Rummery and Mahesan Niranjan. *On-line Q-learning using connectionist systems*. University of Cambridge, Department of Engineering Cambridge, UK, 1994.

Harm Seijen and Rich Sutton. True online TD(λ). In *Proceedings of the International Conference on Machine Learning*, 2014.

Anatolii V. Skorokhod. Limit theorems for stochastic processes. *Theory of Probability & Its Applicationss*, 1956.

Richard Sutton, Ashique Rupam Mahmood, Doina Precup, and Hado Hasselt. A new Q(λ) with interim forward view and monte carlo equivalence. In *International Conference on Machine Learning*, 2014.

Richard S Sutton. Single channel theory: A neuronal theory of learning. *Brain Theory Newsletter*, 1978.

Richard S. Sutton. *Temporal credit assignment in reinforcement learning*. University of Massachusetts Amherst, 1984.

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.

Richard S. Sutton, Joseph Modayil, Michael Delp, Thomas Degris, Patrick M. Pilarski, Adam White, and Doina Precup. Horde: a scalable real-time architecture for learning knowledge from unsupervised sensorimotor interaction. In *Proceedings of the International Conference on Autonomous Agents and Multiagent Systems*, 2011.

Richard S. Sutton, Ashique Rupam Mahmood, and Martha White. An emphatic approach to the problem of off-policy temporal-difference learning. *Journal of Machine Learning Research*, 2016.

Vladislav Tadic. On the convergence of temporal-difference learning with linear function approximation. *Machine Learning*, 2001.

Ahmed Touati, Pierre-Luc Bacon, Doina Precup, and Pascal Vincent. Convergent tree backup and retrace with function approximation. In *Proceedings of the International Conference on Machine Learning*, 2018.

John N. Tsitsiklis and Benjamin Van Roy. Analysis of temporal-diffference learning with function approximation. In *IEEE Transactions on Automatic Control*, 1996.

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.

Jiuqi Wang and Shangtong Zhang. Almost sure convergence of linear temporal difference learning with arbitrary features. *arXiv preprint arXiv:2409.12135*, 2024.

Yue Wang and Shaofeng Zou. Finite-sample analysis of greedy-gq with linear function approximation under markovian noise. In *Proceedings of the Conference on Uncertainty in Artificial Intelligence*, 2020.

Yue Wang, Wei Chen, Yuting Liu, Zhiming Ma, and Tie-Yan Liu. Finite sample analysis of the GTD policy evaluation algorithms in markov setting. In *Advances in Neural Information Processing Systems*, 2017.

Adam White and Martha White. Investigating practical linear temporal difference learning. *arXiv preprint arXiv:1602.08771*, 2016.

Tengyu Xu, Shaofeng Zou, and Yingbin Liang. Two time-scale off-policy td learning: Non-asymptotic analysis over markovian samples. *Advances in Neural Information Processing Systems*, 32, 2019.

Tengyu Xu, Zhuoran Yang, Zhaoran Wang, and Yingbin Liang. Doubly robust off-policy actor-critic: Convergence and optimality. *arXiv preprint arXiv:2102.11866*, 2021.

Huizhen Yu. Least squares temporal difference methods: An analysis under general conditions. *SIAM Journal on Control and Optimization*, 2012.

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 arXiv:1712.09652*, 2017.

Shangtong Zhang. *Breaking the deadly triad in reinforcement learning.* PhD thesis, University of Oxford, 2022.

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

Shangtong Zhang, Wendelin Boehmer, and Shimon Whiteson. Generalized off-policy actor-critic. In *Advances in Neural Information Processing Systems*, 2019.

Shangtong Zhang, Bo Liu, and Shimon Whiteson. GradientDICE: Rethinking generalized offline estimation of stationary values. In *Proceedings of the International Conference on Machine Learning*, 2020a.

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*, 2020b.

Shangtong Zhang, Vivek Veeriah, and Shimon Whiteson. Learning retrospective knowledge with reverse reinforcement learning. In *Advances in Neural Information Processing Systems*, 2020c.

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*, 2021a.

Shangtong Zhang, Hengshuai Yao, and Shimon Whiteson. Breaking the deadly triad with a target network. In *Proceedings of the International Conference on Machine Learning*, 2021b.