SIAM J. Control Optim.
Vol. 64, No. 3, pp. 1456–1481

ASYNCHRONOUS STOCHASTIC APPROXIMATION WITH APPLICATIONS TO AVERAGE-REWARD REINFORCEMENT LEARNING*

HUIZHEN YU†, YI WAN‡, and RICHARD S. SUTTON‡

Abstract. This paper investigates the stability and convergence properties of asynchronous stochastic approximation (SA) algorithms, with a focus on extensions relevant to average-reward reinforcement learning. We first extend a stability proof method of Borkar and Meyn to accommodate more general noise conditions than previously considered, thereby yielding broader convergence guarantees for asynchronous SA. To sharpen the convergence analysis, we further examine the shadowing properties of asynchronous SA, building on a dynamical systems approach of Hirsch and Benaïm. These results provide a theoretical foundation for a class of relative value iteration-based reinforcement learning algorithms—developed and analyzed in a companion paper—for solving average-reward Markov and semi-Markov decision processes.

Key words. asynchronous stochastic approximation, stability and convergence, shadowing properties, average-reward reinforcement learning, Markov and semi-Markov decision processes

MSC codes. 62L20, 90C40, 93E20

DOI. 10.1137/25M1769806

1. Introduction. Asynchronous stochastic approximation (SA) methods play a central role in reinforcement learning (RL), particularly in model-free algorithms for solving Markov and semi-Markov decision processes (MDPs and SMDPs). These processes are versatile models for sequential decision-making under uncertainty in discrete or continuous time, with applications in robotics, finance, operations research, and beyond [20]. This work is motivated by the need to advance RL techniques for average-reward MDPs and SMDPs, where the objective is to optimize sustained long-term performance—a setting in which existing asynchronous SA theory remains inadequate. We focus on establishing general stability and convergence results for asynchronous SA under broad noise conditions arising in average-reward RL. These theoretical developments lay the foundation for a companion paper [33] that analyzes and designs average-reward RL algorithms.

The general asynchronous SA framework we consider is based on the seminal works of Borkar [9, 10] and Borkar and Meyn [12]. The algorithms operate in a finite-dimensional space $\mathbb{R}^d$ and, given an initial vector $x_0 \in \mathbb{R}^d$, iteratively compute $x_n \in \mathbb{R}^d$ for $n \geq 1$ using an asynchronous scheme. This asynchrony involves selective updates to individual components at each iteration. Specifically, at the start of iteration $n \geq 0$, a nonempty subset $Y_n \subset \mathcal{I} := \{1, 2, \ldots, d\}$ is randomly selected according to some mechanism. The ith component $x_n(i)$ of x_n is then updated as

$$x_{n+1}(i) = x_n(i) + \beta_{n,i} \left(h_i(x_n) + \omega_{n+1}(i) \right) \quad \text{if } i \in Y_n, \tag{1.1}$$

*Received by the editors June 18, 2025; accepted for publication (in revised form) December 1, 2025; published electronically May 20, 2026.
https://doi.org/10.1137/25M1769806
Funding: This work was funded by DeepMind, Alberta Machine Intelligence Institute (Amii), and by NSERC grants RGPIN-2024-04939 and DGECR-2024-00312.
†Corresponding author. Department of Computing Science, University of Alberta, Edmonton, Canada (janey.hzyu@gmail.com).
‡Department of Computing Science, University of Alberta, Edmonton, Canada (wan6@ualberta.ca, rsutton@ualberta.ca).

or remains unchanged, $x_{n+1}(i) = x_n(i)$ if $i \notin Y_n$. This process involves a diminishing random step size $\beta_{n,i}$, a Lipschitz continuous function $h : \mathbb{R}^d \to \mathbb{R}^d$ expressed as $h = (h_1, \ldots, h_d)$, and a random noise term $\omega_{n+1} = (\omega_{n+1}(1), \ldots, \omega_{n+1}(d)) \in \mathbb{R}^d$. As will be detailed later, the step sizes and the choices of the sets Y_n satisfy conditions similar to those introduced by Borkar [9, 10], while the function h meets a stability criterion introduced by Borkar and Meyn [12]. This criterion, expressed via a scaling limit of h, imposes conditions on the solutions to the ordinary differential equation (ODE) $\dot{x} = h(x)$ when they are far from the origin, thereby ensuring their stability and other properties.

Building on the above framework, this paper makes two main contributions: first, a stability result for asynchronous SA under more general noise conditions than previously considered; and second, a convergence analysis based on shadowing properties. The former not only addresses key gaps in the literature on average-reward RL but also strengthens the theoretical foundation for extending such methods to broader settings. The latter sharpens existing convergence results in asynchronous SA and, in turn, yields stronger convergence guarantees for RL algorithms that build on this theory. We now review the relevant background and explain how these results build upon and extend prior work.

1.1. Stability: Prior work and our result. Stability, referring to the boundedness of the iterates $\{x_n\}$, is crucial for convergence analysis of SA algorithms. Various approaches exist for ensuring stability (see [11, Chap. 4], [18]). We focus on the Borkar–Meyn stability criterion due to its generality and its suitability for average-reward RL, where the mappings underlying the algorithms generally lack contraction or nonexpansion properties—unlike in discounted or total reward settings (see, e.g., [2, 25, 31] and [6, Chaps. 4 and 5]). An alternative approach for stability, also applied in RL, is to relate the original algorithm (1.1) to a hypothetical, stable counterpart, such as a scaled iteration [2] or projected iterates [21]. However, linking the original iterates to the hypothetical ones is often challenging, involving nonexpansive mappings [2, Lem. 2.1] or requiring bounded differences between original and hypothetical iterates [21, Assump. S5, sect. IV.A]—conditions difficult to satisfy in average-reward RL. In contrast, the Borkar–Meyn stability criterion does not rely on such conditions.

Our first main result is a stability proof for asynchronous SA in the Borkar–Meyn framework under general conditions on the noise terms $\{\omega_n\}$, which arise naturally in average-reward RL for SMDPs. Specifically, the noise is decomposed into a centered component and a biased component as $\omega_{n+1} = M_{n+1} + \epsilon_{n+1}$, where $\{M_{n+1}\}$ is a martingale difference sequence subject to conditional variance conditions, and ϵ_{n+1} satisfies $\|\epsilon_{n+1}\| \le \delta_{n+1}(1 + \|x_n\|)$ with $\delta_{n+1} \to 0$ almost surely (a.s.) as $n \to \infty$ (see Assumption 2.2). In average-reward SMDP applications, the biased term ϵ_{n+1} arises because the function h depends on expected holding times (i.e., expected durations of state transitions), which are unknown SMDP parameters that RL algorithms can only estimate with increasing accuracy over time. The conditions on $\{M_{n+1}\}$ considered here are also weaker than those typically assumed, accommodating RL settings without a priori lower bounds on expected holding times (see section 2.3 and [33]).

To the best of our knowledge, these general noise conditions have not previously been treated in the stability analysis of asynchronous SA using the Borkar–Meyn stability criterion (although conditions of this type are standard in the synchronous case [11, 18]). Borkar and Meyn provided a stability proof for the *synchronous* setting [12], but the stability of *asynchronous* SA was asserted in their theorem [12, Thm. 2.5] without an explicit proof. Moreover, their noise conditions are stronger than ours, requiring M_{n+1} to be generated from x_n and independent and identically distributed (i.i.d.) random disturbances via a function that is uniformly Lipschitz continuous in x_n, as detailed in Remark 2.2. (It should be mentioned, however, that their theorem [12, Thm. 2.5] pertains to a general distributed computing framework with communication delays, which we do not consider.) Within the Borkar–Meyn framework, Bhatnagar [8, Thm. 1] provided a stability proof for asynchronous SA (with bounded communication delays and $\epsilon_{n+1} = 0$) under the condition that $\|M_{n+1}\| \leq K(1 + \|x_n\|)$ for all $n \geq 0$ for some deterministic constant K. This condition is more restrictive than the standard condition on martingale-difference noises, which itself can be too strong for SMDPs (see section 2.3).

One prominent class of RL algorithms where such SA results play a central role involves learning methods based on *relative value iteration* (RVI) [22, 23, 30] for solving average-reward problems. In a seminal paper [1], Abounadi, Bertsekas, and Borkar first applied the Borkar–Meyn stability criterion to develop an asynchronous stochastic RVI algorithm, RVI Q-learning, for finite-space MDPs under an average-reward criterion. Their work has since been extended by Wan, Naik, and Sutton [27, 28] to a broader algorithmic framework and to hierarchical control in average-reward MDPs, as well as by the authors [29] to relax the MDP/SMDP model conditions from unichain to weakly communicating assumptions. These developments in RL all rely on asynchronous SA theory—alongside domain-specific analyses involving the structure of MDPs/SMDPs and the RVI approach—to establish convergence, making the stability of the SA process a foundational issue.

However, the theoretical analyses of RVI Q-learning in the prior studies [1, 27, 28] inherit limitations stemming from gaps in the underlying asynchronous SA theory regarding stability. Although these studies showed that the Borkar–Meyn stability criterion is satisfied by the functions h (cf. (1.1)) associated with their respective algorithms, for the stability of these asynchronous algorithms, the original study of RVI Q-learning [1] relied on an unproven assertion in [12, Thm. 2.5], while the later works [27, 28] argued incorrectly. Bhatnagar's stability result [8, Thm. 1] partially addresses some of these issues, under the assumption of bounded random rewards per stage. However, it does not apply to one algorithm from [27] for hierarchical control, which aims to solve an associated average-reward SMDP where the noise conditions required in [8, Thm. 1] are too restrictive.

Our stability result for asynchronous SA presented in this paper (see Theorem 2.1) not only (a) resolves the open stability question in existing RVI Q-learning algorithms from [1, 27, 28], but also (b) solidifies the groundwork for further extensions of RVI Q-learning. Both points (a) and (b) were already demonstrated in our recent work [29] mentioned earlier, where we applied the SA results established here. In the companion paper [33], which builds on this work, we further support point (b) by developing a new, generalized RVI Q-learning algorithm for SMDPs.

As for the means by which we obtain our stability result, the main strategy—beyond several technical subtleties—is to use stopping-time techniques to construct auxiliary processes with desirable properties, enabling the application of Borkar and Meyn's reasoning developed in the context of synchronous SA [12]. This approach allows us to handle various types of noise terms in a unified way, while retaining the structure of their original stability proof.

1.2. Shadowing: Prior work and our result. Once stability is ensured, a judicious choice of interpolation for the iterates $\{x_n\}$ allows us to apply existing SA theory and conclude that $\{x_n\}$ converges a.s. to a compact, connected subset of $\mathbb{R}^d$ with invariance properties with respect to (w.r.t.) the ODE $\dot{x} = h(x)$ (see Theorem 2.2). This limit set is typically larger than the ω-limit set of a single ODE solution [3, 5].

An approach developed by Hirsch and Benaïm narrows this limit set to an ω-limit set of the ODE by using the concept of 'shadowing' from dynamical systems. In the SA context, shadowing refers to the existence of an ODE solution whose forward trajectory asymptotically coincides with an interpolated trajectory of $\{x_n\}$. The notion of shadowing was first introduced and studied by Bowen [14] for perturbed orbits in dynamical systems. Hirsch [17] later proved a shadowing theorem, which has since found significant applications in SA by enabling sharper characterizations of the limit set; see Benaïm and Hirsch [5] and Benaim [3, 4].

These prior studies focused primarily on synchronous SA algorithms. In this work, we build on their results to analyze the shadowing properties of asynchronous SA under additional conditions on the biased noise terms, step sizes, and asynchrony (see Theorem 2.3 and section 4). We tailor the convergence theorem to a setting relevant to average-reward RL applications, where the focus is on convergence to equilibrium points, but our proof arguments apply more broadly. This constitutes our second main contribution and, to the best of our knowledge, provides the first shadowing analysis for asynchronous SA in the Borkar–Meyn framework.

1.3. Summary of contributions and paper organization. To summarize the main contributions of this paper:

(i) We extend Borkar and Meyn's stability method [12] to address more general noise conditions by employing stopping-time techniques. This result (Theorem 2.1), combined with existing SA theory, leads to broader convergence guarantees for asynchronous SA (Theorem 2.2).

(ii) We further sharpen the convergence analysis for asynchronous SA algorithms by analyzing their shadowing properties (see section 4), building on a dynamical systems approach of Hirsch and Benaïm [3, 4, 5, 17]. This yields enhanced convergence results for asynchronous SA (Theorem 2.3), providing sufficient conditions to ensure convergence to a unique (path-dependent) equilibrium within a compact and connected set of equilibria of the associated ODEs.

We apply these results in [33] to develop and analyze a generalized RVI Q-learning algorithm for SMDPs. An important future research direction is to extend this work to distributed computation frameworks that account for communication delays.

This paper is organized as follows. Section 2 introduces the asynchronous SA framework, presents our stability and convergence results, and discusses their applications to average-reward RL. The stability proof and an outlined basic convergence proof are given in section 3, while the shadowing properties are analyzed in section 4. Section 5 concludes with a brief summary.

Notation: In this paper, $\|\cdot\|$ denotes either the Euclidean norm $\|\cdot\|_2$ or the infinity norm $\|\cdot\|_\infty$, depending on the context. Additionally, $\mathbb{1}\{E\}$ denotes the indicator function of an event E, and $\mathbb{E}[\,\cdot\,]$ denotes expectation. For $a, b \in \mathbb{R}$, $a \vee b := \max\{a, b\}$ and $a \wedge b := \min\{a, b\}$. For convergence of functions, we write $\xrightarrow{p}$ for pointwise convergence and $\xrightarrow{u.c.}$ for uniform convergence on compact subsets of the domain.

2. Asynchronous SA: Stability and convergence.

We start with a detailed description of the asynchronous SA framework outlined earlier in (1.1).

2.1. Algorithmic framework. Let $\alpha_n > 0$, $n \geq 0$, be a given positive sequence of diminishing step sizes. Consider an asynchronous SA algorithm of the following form: at iteration $n \geq 0$, choose a subset $Y_n \neq \varnothing$ of $\mathcal{I}$. For $i \notin Y_n$, let $x_{n+1}(i) = x_n(i)$;

and for $i \in Y_n$, with $\nu(n,i) := \sum_{k=0}^{n-1} \mathbb{1}\{i \in Y_k\}$ denoting the cumulative number of updates to the ith component prior to iteration n, let

$$x_{n+1}(i) = x_n(i) + \alpha_{\nu(n,i)}\big(h_i(x_n) + M_{n+1}(i) + \epsilon_{n+1}(i)\big), \qquad i \in Y_n. \tag{2.1}$$

The algorithm is associated with an increasing family $\{\mathcal{F}_n\}_{n\geq 0}$ of σ-fields, where $\mathcal{F}_n \supset \sigma(x_m, Y_m, M_m, \epsilon_m; m \leq n)$. The following conditions will be assumed throughout.

Assumption 2.1 (conditions on function h).

(i) h is Lipschitz continuous: for some $L_h \geq 0$, $\|h(x) - h(y)\| \leq L_h \|x - y\|$ for all $x, y \in \mathbb{R}^d$.

(ii) Define $h_c(x) := h(cx)/c$ for $c \geq 1$. As $c \uparrow \infty$, $h_c \overset{u.c.}{\to} h_\infty : \mathbb{R}^d \to \mathbb{R}^d$.

(iii) The ODE $\dot{x}(t) = h_\infty(x(t))$ has the origin as its unique globally asymptotically stable equilibrium.

Assumption 2.2 (conditions on noise terms M_n, ϵ_n). For all $n \geq 0$, we have

(i) $\mathbb{E}[\|M_{n+1}\|] < \infty$, $\mathbb{E}[M_{n+1} \mid \mathcal{F}_n] = 0$, and $\mathbb{E}[\|M_{n+1}\|^2 \mid \mathcal{F}_n] \leq K_n(1 + \|x_n\|^2)$ a.s. for some $\mathcal{F}_n$-measurable $K_n \geq 0$ with $\sup_n K_n < \infty$ a.s;

(ii) $\|\epsilon_{n+1}\| \leq \delta_{n+1}(1 + \|x_n\|)$, where δ_{n+1} is $\mathcal{F}_{n+1}$-measurable and $\delta_n \overset{n\to\infty}{\to} 0$ a.s.

Assumption 2.3 (step size conditions).

(i) $\sum_n \alpha_n = \infty$, $\sum_n \alpha_n^2 < \infty$, and $\alpha_{n+1} \leq \alpha_n$ for all n sufficiently large.

(ii) For $x \in (0,1)$, $\sup_n \frac{\alpha_{[xn]}}{\alpha_n} < \infty$, where $[xn]$ denotes the integral part of xn.

(iii) For $x \in (0,1)$, as $n \to \infty$, $\frac{\sum_{k=0}^{[yn]} \alpha_k}{\sum_{k=0}^{n} \alpha_k} \to 1$ uniformly in $y \in [x, 1]$.

For $x > 0$ and $n \geq 0$, define $N(n,x) := \min\{m > n : \sum_{k=n}^{m} \alpha_k \geq x\}$.

Assumption 2.4 (asynchronous update conditions).

(i) For some deterministic $\Delta > 0$, $\liminf_{n\to\infty} \nu(n,i)/n \geq \Delta$ a.s. for all $i \in \mathcal{I}$.

(ii) For each $x > 0$, the limit $\lim_{n\to\infty} \frac{\sum_{k=\nu(n,i)}^{\nu(N(n,x),i)} \alpha_k}{\sum_{k=\nu(n,j)}^{\nu(N(n,x),j)} \alpha_k}$ exists a.s. for all $i, j \in \mathcal{I}$.

Remark 2.1. Assumption 2.1 on h is the Borkar–Meyn stability criterion [12]. Note that the functions h_c and h_∞ are also Lipschitz continuous with modulus L_h, and $h_\infty(0) = 0$. The required uniform convergence condition $h_c \overset{u.c.}{\to} h_\infty$ is equivalent to $h_c \overset{p}{\to} h_\infty$. For the ODE $\dot{x}(t) = h(x(t))$, this assumption not only implies the boundedness of every solution trajectory $x(t)$ for $t \geq 0$, but also guarantees the existence of at least one equilibrium point, as we will show in Lemma 2.1.

Remark 2.2. Assumption 2.2(i) relaxes the standard condition on the martingale difference noise terms $\{M_n\}$, which uses a deterministic constant K instead of the K_n's in the conditional variance bounds. We will discuss in section 2.3 the necessity of each condition in the context of our RL applications.

The noise terms for asynchronous SA in Borkar [9] and Borkar and Meyn [12] satisfy Assumption 2.2 but are more specific: $\epsilon_{n+1} = 0$ and $M_{n+1} = F(x_n, \zeta_{n+1})$, where $\{\zeta_n\}$ are exogenous and i.i.d. random variables, and F is a function uniformly Lipschitz in its first argument. (Stability of the algorithm is assumed in [9] and asserted in [12, Thm. 2.5] without explicit proof.) In our arXiv preprint [32], we provide an alternative stability proof for this specific form of martingale-difference noises, which is slightly simpler than our stability proof under the more general Assumption 2.2.

Remark 2.3. Assumptions 2.3 and 2.4 regarding step sizes and asynchrony are nearly identical to those used in [1] for RVI Q-learning. They accommodate commonly used step sizes, such as $1/n$ or $1/(n \ln n)$, and typical RL scenarios, where $\mathcal{I}$ is the space of state-action pairs and the sets Y_n are selected based on Markov chains on $\mathcal{I}$ induced by RL agents (see [29, Ex. 3] for an example). These conditions, with some minor variations in Assumption 2.4(ii), were originally introduced in a broader asynchronous SA context by Borkar [9, 10], specifically for the step size structure $\alpha_{\nu(n,i)}$. Their purpose is to impose partial asynchrony so that the asymptotic behavior of the asynchronous algorithm aligns, on average, with that of a synchronous counterpart, thereby facilitating analysis. (This is evident from the detailed analysis in [9, 10], which shows that the limits in Assumption 2.4(ii) must in fact equal 1; see also our Lemmas 3.2, 3.4, 4.1(i).)

This partial asynchrony is crucial for our average-reward RL applications. While Q-learning achieves stability and convergence in fully asynchronous schemes for both discounted and certain undiscounted total-reward MDPs [25, 31], these results do not extend to the average-reward Q-learning algorithms of interest, as their associated mappings are generally neither contractive nor nonexpansive.

2.2. Results. We now present our stability and convergence theorems for algorithm (2.1). The stability theorem, our first main result in this section, parallels Borkar [11, Thm. 4.1] for synchronous algorithms. Its proof, given in section 3, extends the Borkar–Meyn stability argument through a stopping-time construction, as noted earlier in the introduction.

THEOREM 2.1. *For algorithm* (2.1) *under Assumptions* 2.1–2.4, $\{x_n\}$ *is bounded a.s.*

Combining Theorem 2.1 with established SA theory [9, 11] leads to our convergence theorem. This theorem characterizes the asymptotic behavior of both the individual iterates $\{x_n\}$ and a continuous trajectory formed by them, which we introduce first.

Let $\mathcal{C}((-\infty,\infty);\mathbb{R}^d)$ (resp., $\mathcal{C}([0,\infty);\mathbb{R}^d)$) denote the space of all $\mathbb{R}^d$-valued continuous functions on $\mathbb{R}$ (resp., $\mathbb{R}_+$), equipped with a metric such that convergence in this space corresponds to uniform convergence on compact intervals. These spaces are complete. By the Arzelá–Ascoli theorem, a family of functions in either space is relatively compact (i.e., has compact closure) if and only if these functions are equicontinuous and pointwise bounded (see [11, App. A.1] or [18, Chap. 4.2.1]).

Define a linearly interpolated trajectory $\bar{x}(t)$ from $\{x_n\}$ with aggregated step sizes $\tilde{\alpha}_n = \sum_{i\in Y_n} \alpha_{\nu(n,i)}$, $n \geq 0$, as the elapsed times between consecutive iterates. Specifically, for $n \geq 0$, let $\tilde{t}(n) := \sum_{k=0}^{n-1} \tilde{\alpha}_k$ with $\tilde{t}(0) := 0$, and define $\bar{x}(\tilde{t}(n)) := x_n$ and

$$\bar{x}(t) := x_n + \frac{t - \tilde{t}(n)}{\tilde{t}(n+1) - \tilde{t}(n)} (x_{n+1} - x_n), \quad t \in [\tilde{t}(n), \tilde{t}(n+1)]. \tag{2.2}$$

We refer to the temporal coordinate of $\bar{x}(t)$ as the "ODE time." For the results below, we extend $\bar{x}(\cdot)$ to a function in $\mathcal{C}((-\infty,\infty);\mathbb{R}^d)$ by setting $\bar{x}(\cdot) \equiv x_0$ on $(-\infty, 0)$.

THEOREM 2.2. *For algorithm* (2.1) *under Assumptions* 2.1–2.4*, a.s.,*

(i) *the sequence* $\{x_n\}$ *converges to a (possibly sample path-dependent) compact, connected, internally chain transitive,*[1] *invariant set* D *of the ODE* $\dot{x}(t) = h(x(t))$*;*

[1] Recall from [4, sect. 5.1, p. 21] and [11, Chap. 2.1, p. 16] that a compact invariant set D is called *internally chain transitive* if for any $x, y \in D$ and any $\epsilon, T > 0$, there exists a finite sequence of points in D starting at x and ending in y: $x, x_0, x_1, \ldots, x_{n-1}, x_n = y$, such that $\|x - x_0\| < \epsilon$ and, for each $0 \leq i < n$, the ODE trajectory starting from x_i comes within ϵ distance of x_{i+1} at a time greater than T. Internally chain transitive sets and related concepts, introduced by Bowen [14] and Conley [15], are fundamental in dynamical systems theory. Benaim [3] was the first to use them to characterize the asymptotic behavior of SA algorithms. Although our proofs do not explicitly involve these sets, they enter through existing convergence theory and provide the theoretical underpinning.

(ii) *with* $\bar{x}(\cdot)$ *defined as above, the family* $\{\bar{x}(t+\cdot)\}_{t\in\mathbb{R}}$ *is relatively compact in* $\mathcal{C}((-\infty,\infty);\mathbb{R}^d)$*, and any limit point of* $\bar{x}(t+\cdot)$ *as* $t\to\infty$ *is a solution of the ODE* $\dot{x}(t)=\frac{1}{d}h(x(t))$ *that remains entirely in* D*.*

The proof of Theorem 2.2 will be outlined in section 3. Part (i) of Theorem 2.2 parallels [11, Thm. 2.1] for synchronous algorithms, while part (ii) is similar to [11, Thm. 6.1] for asynchronous algorithms, given stability. A key distinction from [11, Thm. 6.1] is our choice of the step sizes that define the ODE time in $\bar{x}(\cdot)$. The step size choice in [11, Thm. 6.1] appears less advantageous *under Assumptions* 2.3 *and* 2.4, as it not only results in multiple limiting ODEs but also complicates their characterization (see [32, Rem. 3.2] for a detailed discussion). Here, we obtain a unique limiting ODE (Theorem 2.2(ii)), which, beyond technical convenience, facilitates further analysis of the shadowing properties of $\bar{x}(\cdot)$, as will be discussed below.

We now focus on a scenario relevant to RL, where the goal is for the algorithms to converge to the equilibrium set of the ODE $\dot{x}(t)=h(x(t))$, defined as $E_h:=\{x\in\mathbb{R}^d \mid h(x)=0\}$. Before proceeding, we note an implication of Assumption 2.1 for E_h.

Lemma 2.1. *If* h *satisfies Assumption* 2.1*, then* E_h *is nonempty, compact, and contained in some compact, connected, globally asymptotically stable*[2] *set of the ODE* $\dot{x}(t)=h(x(t))$*.*

Proof. By [11, Corollary 4.1], there exist $\bar{c}>0$ and $T>0$ such that for any solution $x(\cdot)$ of the ODE $\dot{x}(t)=h(x(t))$, if $\|x(t)\|\geq\bar{c}$, then $\|x(t+T)\|<\|x(t)\|/4$. Thus, for any initial condition $x(0)$, there exists a sequence of times $t_n\uparrow\infty$ such that $x(t_n)\in B_{\bar{c}}:=\{y\in\mathbb{R}^d \mid \|y\|\leq\bar{c}\}$. The compact set $B_{\bar{c}}$ is hence a global weak attractor (by the definition of such an attractor; see [7, Chap. V.1]) and contains at least one equilibrium point by [7, Thm. V.3.9]. This shows that $B_{\bar{c}}\supset E_h\neq\varnothing$. Since h is Lipschitz continuous, E_h is compact. For the desired globally asymptotically stable set containing E_h, we can take the first positive prolongation set of $B_{\bar{c}}$.[3] ∎

As will be discussed in section 2.3, for average-reward RL in MDPs/SMDPs without unichain model restrictions, E_h is generally not a singleton but rather consists of infinitely many connected equilibrium points. Corollary 2.1 below applies Theorem 2.2 to this setting, using the fact that if E_h is globally asymptotically stable, it must contain all compact invariant sets.

Corollary 2.1. *Let Assumptions* 2.1–2.4 *hold, and let* E_h *be globally asymptotically stable for the ODE* $\dot{x}(t)=h(x(t))$*. Then the following hold a.s. for algorithm* (2.1)*:*

(i) *The sequence* $\{x_n\}$ *converges to a compact connected subset of* E_h*.*

(ii) *For any* $\delta>0$ *and any convergent subsequence* $\{x_{n_k}\}$*, as* $k\to\infty$*,*

$$\tau_{\delta,k}:=\min\{|s|:\|\bar{x}(t_{n_k}+s)-x^*\|>\delta,\ s\in\mathbb{R}\}\to\infty,$$

where $\bar{x}(\cdot)$ *is the continuous trajectory defined above,* $t_{n_k}=\tilde{t}(n_k)$ *is the ODE-time when* x_{n_k} *is generated, and* $x^*\in E_h$ *is the point to which* $\{x_{n_k}\}$ *converges.*

[2]Recall that a compact set D is globally asymptotically stable if every solution $x(t)$ of the ODE converges to D as $t\to\infty$, and D is Lyapunov stable (see [18, Chap. 4.2.2]).

[3]Details: As a compact global weak attractor, $B_{\bar{c}}$ lies in some compact globally asymptotically stable set, the smallest being its first positive prolongation set $D^+(B_{\bar{c}})$ [7, Thm. V.1.25]. Explicitly, $D^+(B_{\bar{c}})=\cup_{x\in B_{\bar{c}}}D^+(x)$, where $D^+(x)=\{y\in\mathbb{R}^d \mid \exists\{x_n\}\subset\mathbb{R}^d$ and $\{t_n\}\subset\mathbb{R}_+$ s.t. $x_n\to x,\ \phi(t_n;x_n)\to y$ as $n\to\infty\}$ with $\phi(t;\bar{x})$ denoting the ODE solution with $x(0)=\bar{x}$. Since each $D^+(x)$, $x\in B_{\bar{c}}$, is connected [7, Thm. II.4.4] and $B_{\bar{c}}$ is connected, $D^+(B_{\bar{c}})$ is connected.

Corollary 2.1(ii) shows that algorithm (2.1) spends increasing ODE-time in arbitrarily small neighborhoods around its iterates' limit points in E_h, with each visit near a limit point lasting longer and tending to infinity. Thus, when E_h contains nonisolated equilibria, the algorithm's behavior may resemble convergence to a single point, even if it does not truly converge.

We now analyze this case further[4] by presenting sufficient conditions for the asynchronous algorithm (2.1) to converge to a unique point in E_h, a result also relevant to our RL applications. Our approach builds on Hirsch [17], Benaïm and Hirsch [5], and Benaim [3, 4]. Roughly speaking, the idea is to ensure that $\bar{x}(\cdot)$ asymptotically tracks a unique solution trajectory of the limiting ODE, a concept known as *shadowing* in the cited literature. One way to achieve this is by ensuring sufficiently rapid tracking of the ODE solutions. In the asynchronous SA setting, we decompose the "tracking error" into two components—one from stochastic noise and one from asynchrony—and introduce conditions to control both.

We formalize these conditions as follows.

Assumption 2.5 (additional condition on noise term ϵ_n). There exists a deterministic constant $\mu_\delta < 0$ such that $\limsup_{n\to\infty} \frac{\ln(\delta_{n+1})}{\sum_{k=0}^{n} \alpha_k} \le \mu_\delta$ a.s., where $\{\delta_n\}$ are the random variables involved in Assumption 2.2(ii) for $\{\epsilon_n\}$.

Consider two specific forms of step sizes: $\alpha_n = \frac{1}{An}$ (class 1) and $\alpha_n = \frac{1}{An \ln n}$ (class 2), where $A > 0$ is a scaling parameter (with α_n set to $\frac{1}{A}$ if the denominator is zero). For class-1 step sizes, we impose an additional condition on asynchony.

Assumption 2.6 (additional asynchrony condition). For class-1 step sizes, the asynchronous update schedules are such that a.s. for all $i \in \mathcal{I}$, $\nu(n,i)/n \to p_i$ as $n \to \infty$ for some (sample path-dependent) $p_i \in (0,1]$. Moreover, there exists a deterministic constant $\gamma > 0$ such that $\limsup_{n\to\infty} n^\gamma \left|\nu(n,i)/n - p_i\right| < \infty$ a.s.

The following convergence theorem is our second main result in this section; its proof is given in section 4. All conditions can be met in our RL applications (see section 2.3). Let L_h be the Lipschitz constant of h under $\|\cdot\|_\infty$—this norm is chosen because, in our RL applications, an effective bound on L_h under this norm can be obtained with minimal or no model knowledge (see section 2.3).

THEOREM 2.3. *Consider algorithm* (2.1) *with class*-1 *step sizes, where* $\frac{A}{2} > L_h$ *or class*-2 *step sizes, where* $A > L_h$. *Let Assumptions* 2.1–2.4, 2.5 *with* $\mu_\delta < -L_h$ *hold, and for class*-1 *step sizes, also assume Assumption* 2.6 *with* $\gamma A > L_h$. *For the ODE* $\dot{x}(t) = h(x(t))$, *suppose that the set* E_h *is globally asymptotically stable and that every solution* $x(t)$ *converges to a unique point in* E_h *as* $t \to \infty$. *Then the sequence* $\{x_n\}$ *from algorithm* (2.1) *converges a.s. to a point in* E_h *that depends on the sample path.*

We now make a few general remarks on the additional conditions and the proof of Theorem 2.3, before discussing the average-reward RL application (section 2.3).

Remark 2.4 (on the proof). Following the general approach of Hirsch and Benaïm, most arguments in our proof of Theorem 2.3 also apply to analyzing shadowing properties of asynchronous SA algorithms in more general settings. In particular, much of the analysis is independent of E_h and its associated conditions—these are only used to specialize the conclusions. Moreover, rather than relying on the specific forms of the step sizes $\{\alpha_n\}$, the proof primarily depends on the quantity $\ell(\{\alpha_n\}) := \limsup_{n\to\infty}$

[4]Any compact, connected subset of E_h is internally chain transitive by definition, so Corollary 2.1(i) cannot be further improved based on Theorem 2.2(i).

$\frac{\ln(\alpha_n)}{\sum_{k=0}^{n}\alpha_k}$, similar to Benaim's analysis [3, 4] for synchronous SA. Class-1 and class-2 step sizes have $\ell(\{\alpha_n\}) = -A$ and $-\infty$, respectively, with their specific forms used only to obtain explicit estimates of the tracking error due to asynchrony (see Lemma 4.5 in section 4.3).

Remark 2.5 (on the additional ODE conditions). Borkar and Soumyanath [13] showed that if $h(x) = F(x) - x$ with $F : \mathbb{R}^d \to \mathbb{R}^d$ nonexpansive in the p-norm ($p \in (1, +\infty]$) and with nonempty fixed-point set E_h, then the solution $x(t)$ of $\dot{x}(t) = h(x(t))$ converges to a point in E_h with distance to E_h nonincreasing. It follows that if h also satisfies the Borkar–Meyn stability criterion, the additional ODE conditions in Theorem 2.3 hold. This identifies another class of problems where Theorem 2.3 applies, besides average-reward RL. In this setting, Theorem 2.3 sharpens the previous result [9, Thm. 3.2] in the absence of communication delays.

Remark 2.6 (on the additional step size/asynchrony conditions).

(a) For class-2 step sizes, their rapid decrease eliminates the need for the additional asynchrony condition (Assumption 2.6), as Assumption 2.4 alone suffices to control the tracking error from asynchrony (see the proof of Lemma 4.5). In the synchronous case, the scaling parameter A for class-2 step sizes can be chosen freely, provided that Assumption 2.5 holds with $\mu_\delta < -L_h$. This is consistent with our asynchronous-case proof and aligns with prior shadowing results [3, 4] for synchronous SA.

(b) For class-1 step sizes, one way to satisfy the additional asynchrony condition Assumption 2.6 is to select components for updating x_n in a way that eventually follows an irreducible Markov chain on $\mathcal{I}$. The law of the iterated logarithm then ensures that Assumption 2.6 holds for any $0 < \gamma < \frac{1}{2}$, allowing the required condition $\gamma A > L_h$ to be met by choosing the step size scaling parameter $A > 2L_h$.

2.3. Applications to average-reward RL. This subsection briefly overviews the average-reward RL applications studied in our companion paper [33] and recent work [29]. There, E_h corresponds to a compact, connected subset of solutions to the average-reward optimality equation for a weakly communicating[5] SMDP or MDP on finite state and action spaces. A fundamental result in MDP theory [24] shows that the full solution set is homeomorphic to an unbounded convex polyhedron, while E_h is homeomorphic to a compact section of this polyhedron—generally a nonsingleton set (unlike in the unichain case).

We apply Corollary 2.1 and Theorem 2.3 to analyze RVI Q-learning, treating several existing algorithms in [29] and developing a more general version in [33]. Corollary 2.1 is used to establish convergence to a compact, connected subset of solutions to the optimality equation, while Theorem 2.3 is used to ensure convergence to a unique, sample path-dependent solution (see [33, sections. 3.3 and 4]).

To apply these SA results, we prove that Assumptions 2.1 and 2.2 on the function h and the noise, along with the ODE and E_h conditions, hold based on the structure of weakly communicating MDPs/SMDPs and the algorithmic design of RVI Q-learning. (This part of the analysis builds on [1, 13] and is nontrivial, though it does not invoke the SA results themselves—even though those results shaped the algorithmic design.)

Regarding the noise conditions: for SMDPs, the standard condition on the martingale difference terms $\{M_n\}$ suffices when a lower bound on the expected holding

[5] I.e., excluding states that are transient under all policies, the remaining states form a single closed communicating class under some policy; see, e.g., [20, Chap. 8.3].

times is known a priori; otherwise, the more general condition, Assumption 2.2(i) is needed. As noted earlier, Assumption 2.2(ii) on the biased noise terms $\{\epsilon_n\}$ is required because the function h depends on the expected holding times, which are estimated from data with increasing accuracy by the RL algorithm. (See [33, Lem. 4.2] for exactly how Assumption 2.2 is used in our RL context.) For MDPs, which are SMDPs with unit holding times, all $\epsilon_n = 0$, and the standard condition on $\{M_n\}$ suffices.

The remaining assumptions required by our SA results can be satisfied through appropriate choices of step sizes and asynchronous update schedules. In Remarks 2.3 and 2.6, we already addressed some of these in an application-independent way. Here, we focus on the application-specific aspects. For RL in average-reward SMDPs, since the biased noise terms $\{\epsilon_n\}$ arise from estimated expected holding times, the additional noise condition Assumption 2.5 with $\mu_\delta < -L_h$ can be translated into algorithmic requirements on the estimation procedure and ensured accordingly (see [33, Thm. 3.2 and Lem. 4.8, used in its proof]). In the special case of MDPs, where all $\epsilon_n = 0$, Assumption 2.5 holds automatically with $\mu_\delta = -\infty$. An effective upper bound on L_h can be obtained with minimal model knowledge—specifically, a lower bound on the minimum expected holding time—and used in place of the exact value of L_h when setting the threshold for the step size scaling parameter A (see [33, Thm. 3.2 and its proof]). In summary, all conditions of the preceding SA results can be satisfied in our RL setting through suitable algorithmic choices, with minimal or no model knowledge (see [33] for details).

3. Proofs for section 2.

This section provides the proofs of all results in section 2, except for Theorem 2.3, whose proof appears in ection 4. Specifically, we establish the basic stability and convergence theorems—Theorems 2.1, 2.2, and Corollary 2.1—for the asynchronous SA algorithm (2.1). Assumptions 2.1–2.4 remain in effect throughout. To avoid repetition, we draw heavily from Borkar [9, 11], focusing on elements essential to the asynchronous algorithm. In this section, we use $\|\cdot\|$ to denote the Euclidean norm $\|\cdot\|_2$, consistent with [11]; norm equivalence allows us to use the stated conditions without modification.

We use an ODE-based approach from [9], working with linearly interpolated trajectories formed from the iterates $\{x_n\}$ and connecting these to solutions of *nonautonomous* ODEs of the form $\dot{x}(t) = \lambda(t)g(x(t))$, where $\lambda(t)$ arises from asynchrony. Depending on the context, g may be h, h_c, etc. We construct these trajectories differently for the stability and convergence analyses, leading to different $\lambda(\cdot)$ functions. In section. 3.1, we first define and analyze these time-dependent components $\lambda(\cdot)$ and their asymptotic properties, which are crucial for the subsequent proofs (sections. 3.2–3.3).

3.1. Preliminary analysis.

For the stability proof, we use the deterministic step size sequence $\{\alpha_n\}$ as the elapsed times between consecutive iterates to define a continuous trajectory $\bar{x}(\cdot)$. Once stability is established, we switch to a random step size sequence in the convergence analysis, both for technical convenience and sharper results. Using random step sizes in the stability analysis seems nonviable under our noise conditions.

Specifically, for the stability proof, we define a linearly interpolated trajectory $\bar{x}(t)$ as follows: Let $t(0) := 0$ and $t(n) := \sum_{k=0}^{n-1} \alpha_k$ for $n \geq 1$. Let

$$\bar{x}(t) := x_n + \frac{t - t(n)}{t(n+1) - t(n)} (x_{n+1} - x_n), \quad t \in [t(n), t(n+1)],\ n \geq 0. \tag{3.1}$$

To define $\lambda(\cdot)$, we first rewrite algorithm (2.1) explicitly in terms of $\{\alpha_n\}$ as

$$x_{n+1}(i) = x_n(i) + \alpha_n \, b(n,i) \left(h_i(x_n) + M_{n+1}(i) + \epsilon_{n+1}(i) \right), \quad i \in \mathcal{I}, \tag{3.2}$$

where $b(n,i) := \frac{\alpha_{\nu(n,i)}}{\alpha_n} \mathbb{1}\{i \in Y_n\}$. We show that $b(n,i)$ is eventually bounded by a deterministic constant, which we then use to define $\lambda(\cdot)$ and its space Υ:

LEMMA 3.1. *For some deterministic constant $C \geq 1$, it holds a.s. that for all sufficiently large (sample path-dependent) n, $\max_{i\in\mathcal{I}} b(n,i) \leq C$ and $\sum_{i\in\mathcal{I}} b(n,i) \geq 1$.*

Proof. As discussed in [9, p. 842], Assumption 2.3(ii), together with $\{\alpha_n\}$ being eventually nonincreasing (Assumption 2.3(i)), implies that $\sup_n \sup_{y\in[x,1]} \frac{\alpha_{[yn]}}{\alpha_n} < \infty$ for $x \in (0,1)$. By Assumption 2.4(i), for n sufficiently large, $\min_{i\in\mathcal{I}} \nu(n,i)/n \geq \Delta/2$ a.s. Thus, for the finite deterministic constant $C := \sup_n \sup_{y\in[\Delta/2,1]} \frac{\alpha_{[yn]}}{\alpha_n} \geq 1$, we have $\max_{i\in\mathcal{I}} b(n,i) \leq \max_{i\in\mathcal{I}} \frac{\alpha_{\nu(n,i)}}{\alpha_n} \leq C$ for n sufficiently large, a.s. Since $\{\alpha_n\}$ is eventually nonincreasing and the sets $Y_n \neq \varnothing$, Assumption 2.4(i) also implies that for n sufficiently large, $\sum_{i\in\mathcal{I}} b(n,i) = \sum_{i\in Y_n} \frac{\alpha_{\nu(n,i)}}{\alpha_n} \geq 1$ a.s. ◻

Let C be the constant given in Lemma 3.1. Let Υ comprise all Borel-measurable functions that map $t \geq 0$ to a $d \times d$ diagonal matrix with nonnegative diagonal entries bounded by C. Two such functions are regarded as the same element if they are equal almost everywhere (a.e.) w.r.t. the Lebesgue measure. Similarly to [9], we equip Υ with the coarsest topology that makes the mappings $\psi_{t,f} : \lambda' \in \Upsilon \mapsto \int_0^t \lambda'(s) f(s)\, ds$ continuous for all $t > 0$ and $f \in L_2([0,t];\mathbb{R}^d)$ (the space of all $\mathbb{R}^d$-valued square-integrable functions on $[0,t]$). With this topology, Υ is compact and metrizable by the Banach–Alaoglu theorem and the separability of the Hilbert spaces $L_2([0,t];\mathbb{R}^d)$, $t > 0$ (see [11, Chaps. 6.2, A.2]). We regard Υ as a compact metric space with a compatible metric. Note that any sequence in Υ contains a convergent subsequence.

We now define $\lambda(t)$ as a diagonal matrix-valued, piecewise constant function by

$$\lambda(t) := \mathrm{diag}\big(b(n,1)\wedge C,\, b(n,2)\wedge C,\, \ldots,\, b(n,d)\wedge C\big), \quad t \in [t(n), t(n+1)),\ n \geq 0. \tag{3.3}$$

For $t \geq 0$, $\lambda(t+\cdot)$ on $[0,\infty)$ is an element of Υ. Let I denote the identity matrix. The next lemma characterizes the limit points of $\lambda(t+\cdot)$ as $t \to \infty$. Its proof (given in our arXiv preprint [32]) is similar to that of [9, 10, Thm. 3.2], invoking Assumption 2.4(ii) specifically to apply L'Hôpital's rule and using Lemma 3.1 to lower bound the limiting functions.

LEMMA 3.2. *A.s., for any sequence $t_n \geq 0$ with $t_n \uparrow \infty$, all limit points of the sequence $\{\lambda(t_n+\cdot)\}_{n\geq 0}$ in Υ have the form $\lambda^*(t) = \rho(t) I$, where $\rho(\cdot)$ is a real-valued Borel-measurable function satisfying $\frac{1}{d} \leq \rho(t) \leq C$ for all $t \geq 0$.*

As mentioned, our convergence proof uses a different setup. Specifically, we work with the trajectory $\bar{x}(t)$ defined by (2.2), which places the iterate x_n at the temporal coordinate $\tilde{t}(n) = \sum_{k=0}^{n-1} \tilde{\alpha}_k$, using aggregated random step sizes $\tilde{\alpha}_k = \sum_{i\in Y_k} \alpha_{\nu(k,i)}$. As we show below, this interpolation scheme leads to a simpler limiting behavior of the associated function $\lambda(\cdot)$, denoted by $\tilde{\lambda}(\cdot)$ in this context.

Let us write algorithm (2.1) equivalently in terms of $\{\tilde{\alpha}_n\}$ as

$$x_{n+1}(i) = x_n(i) + \tilde{\alpha}_n \, \tilde{b}(n,i) \left(h_i(x_n) + M_{n+1}(i) + \epsilon_{n+1}(i) \right), \qquad i \in \mathcal{I}, \tag{3.4}$$

where $\tilde{b}(n,i) := \frac{\alpha_{\nu(n,i)}}{\tilde{\alpha}_n} \mathbb{1}\{i \in Y_n\}$ and thus $\sum_{i\in\mathcal{I}} \tilde{b}(n,i) = 1$. Define $\tilde{\lambda}(\cdot)$ as a diagonal matrix-valued, piecewise constant trajectory by

$$\tilde{\lambda}(t) := \mathrm{diag}\big(\tilde{b}(n,1),\, \tilde{b}(n,2),\, \ldots,\, \tilde{b}(n,d)\big), \qquad t \in [\tilde{t}(n), \tilde{t}(n+1)),\ n \geq 0. \tag{3.5}$$

We view $\tilde{\lambda}(\cdot)$ as an element in the space $\tilde{\Upsilon}$ which comprises all Borel-measurable functions that map $t\geq 0$ to a $d\times d$ diagonal matrix with nonnegative diagonal entries summing to 1. Regarded as a subset of Υ with the relative topology, $\tilde{\Upsilon}$ is a compact metric space.

Let us show that as $t\to\infty$, $\tilde{\lambda}(t+\cdot)$ has a unique limit point in $\tilde{\Upsilon}$, given by the constant function $\bar{\lambda}(\cdot)\equiv\frac{1}{d}I$. Our proof actually shows more: it also establishes the equivalence of two conditions on asynchrony used in the literature, [1, Assum. 2.4] and the condition introduced earlier in [9, 10].

Define $\tilde{N}(n,x) := \min\left\{m>n : \sum_{k=n}^{m}\sum_{i\in Y_k}\alpha_{\nu(k,i)}\geq x\right\}$ for $x>0$.

LEMMA 3.3. *Given Assumptions* 2.3 *and* 2.4(i), *Assumption* 2.4(ii) *is equivalent to*

$$\text{(3.6)}\qquad \textit{for each } x>0,\quad \lim_{n\to\infty}\frac{\sum_{k=\nu(n,i)}^{\nu(\tilde{N}(n,x),i)}\alpha_k}{\sum_{k=\nu(n,j)}^{\nu(\tilde{N}(n,x),j)}\alpha_k}=1 \quad a.s. \quad \forall\, i,j\in\mathcal{I}.$$

Proof. First, assume Assumption 2.4(ii). Consider a sample path for which Assumption 2.4 and Lemma 3.2 hold. Fix $x>0$. By the definition of $\tilde{N}(n,x)$ and Assumption 2.4(i), $\sum_{i\in\mathcal{I}}\sum_{k=\nu(n,i)}^{\nu(\tilde{N}(n,x),i)}\alpha_k \approx \sum_{k=n}^{\tilde{N}(n,x)}\sum_{i\in Y_k}\alpha_{\nu(k,i)} \overset{n\to\infty}{\to} x$, where we define "$\approx$" to denote equality up to an $o(1)$ term. Thus, (3.6) is equivalent to $\lim_{n\to\infty}\sum_{k=\nu(n,i)}^{\nu(\tilde{N}(n,x),i)}\alpha_k=\frac{x}{d}$ $\forall i\in\mathcal{I}$. To prove this, it suffices to show that for any increasing sequence $\{n_\ell\}_{\ell\geq 1}$ of natural numbers, there is a subsequence $\{n'_\ell\}_{\ell\geq 1}$ along which $\sum_{k=\nu(n'_\ell,i)}^{\nu(\tilde{N}(n'_\ell,x),i)}\alpha_k \overset{\ell\to\infty}{\to}\frac{x}{d}$ $\forall i\in\mathcal{I}$. To this end, with $t_{n_\ell}:=t(n_\ell)$, consider a convergent subsequence of $\{\lambda(t_{n_\ell}+\cdot)\}_{\ell\geq 1}$ in Υ with limit point $\lambda^*(\cdot)=\rho(\cdot)I$ (Lemma 3.2). Denote this convergent subsequence again by $\{\lambda(t_{n_\ell}+\cdot)\}_{\ell\geq 1}$ to simplify notation. Thus, $\lambda(t_{n_\ell}+\cdot)\to\lambda^*$ and we need to prove $\sum_{k=\nu(n_\ell,i)}^{\nu(\tilde{N}(n_\ell,x),i)}\alpha_k\overset{\ell\to\infty}{\to}\frac{x}{d}$ $\forall i\in\mathcal{I}$.

Choose $\epsilon\in(0,x)$. Since $\rho(s)\in[1/d,C]$ for all $s\geq 0$ by Lemma 3.2, the two equations, $\int_0^{\underline{\tau}}\rho(s)ds=\frac{x-\epsilon}{d}$ and $\int_0^{\bar{\tau}}\rho(s)ds=\frac{x+\epsilon}{d}$ uniquely define two constants $\underline{\tau}>0$ and $\bar{\tau}>0$. Then, since $\lambda(t_{n_\ell}+\cdot)\to\lambda^*$, we have that for all $i\in\mathcal{I}$, as $\ell\to\infty$,

$$\int_0^{\underline{\tau}}\lambda_{ii}(t_{n_\ell}+s)ds\to\frac{x-\epsilon}{d},\qquad \int_0^{\bar{\tau}}\lambda_{ii}(t_{n_\ell}+s)ds\to\frac{x+\epsilon}{d}.$$

By Lemma 3.1 and the definition of λ [see (3.3)], this implies

$$\text{(3.7)}\qquad \underline{c}_\ell(i):=\sum_{k=\nu(n_\ell,i)}^{\nu(N(n_\ell,\underline{\tau}),i)}\alpha_k\to\frac{x-\epsilon}{d},\qquad \bar{c}_\ell(i):=\sum_{k=\nu(n_\ell,i)}^{\nu(N(n_\ell,\bar{\tau}),i)}\alpha_k\to\frac{x+\epsilon}{d}.$$

Hence $\sum_{k=n_\ell}^{N(n_\ell,\underline{\tau})}\sum_{i\in Y_k}\alpha_{\nu(k,i)}\approx\sum_{i\in\mathcal{I}}\underline{c}_\ell(i)\to x-\epsilon$ and $\sum_{k=n_\ell}^{N(n_\ell,\bar{\tau})}\sum_{i\in Y_k}\alpha_{\nu(k,i)}\approx\sum_{i\in\mathcal{I}}\bar{c}_\ell(i)\to x+\epsilon$. From these relations and the definition of $\tilde{N}(n,x)$, it follows that for all ℓ sufficiently large, $N(n_\ell,\underline{\tau})<\tilde{N}(n_\ell,x)<N(n_\ell,\bar{\tau})$ and, consequently, $\underline{c}_\ell(i)\leq\sum_{k=\nu(n_\ell,i)}^{\nu(\tilde{N}(n_\ell,x),i)}\alpha_k\leq\bar{c}_\ell(i)$ $\forall i\in\mathcal{I}$. This together with (3.7) and the arbitrariness of ϵ implies that $\sum_{k=\nu(n_\ell,i)}^{\nu(\tilde{N}(n_\ell,x),i)}\alpha_k\overset{\ell\to\infty}{\to}\frac{x}{d}$ $\forall i\in\mathcal{I}$, proving that Assumption 2.4(ii) entails the stated condition (3.6).

Conversely, to show that Assumption 2.4(ii) is implied by condition (3.6), we argue similarly to the preceding proof but with the roles of $\tilde{N}(n,\cdot)$ and $N(n,\cdot)$ reversed. Instead of Lemma 3.2, we use Lemma 3.4 (an implication of condition (3.6), presented below), and we also use the compactness of the space Υ. (See [32] for details.) ∎

From (3.6) and Assumptions 2.3, 2.4(i), Lemma 3.4 follows by [9, 10, proof of Thm. 3.2].

LEMMA 3.4. *As $t \to \infty$, $\tilde{\lambda}(t+\cdot)$ converges a.s. in $\tilde{\Upsilon}$ to $\bar{\lambda}(\cdot) \equiv \frac{1}{d} I$.*

This lemma implies a unique limiting ODE associated with the interpolated trajectory $\bar{x}(\cdot)$ in (2.2) (see section 3.3). As noted earlier, this uniqueness is not only technically convenient for the convergence analysis but also facilitates further investigation of the shadowing properties of the asynchronous SA algorithm (2.1) (see section 4).

3.2. Stability proof. We now proceed to prove Theorem 2.1.

3.2.1. Relating scaled iterates to ODE solutions. Consider algorithm (2.1) in its equivalent form (3.2) and the continuous trajectory $\bar{x}(t)$ defined in (3.1). Following [12] and [11, Chap. 4.2], we work with a scaled trajectory $\hat{x}(\cdot)$ derived from $\bar{x}(\cdot)$ as follows: divide the time axis into intervals of about length T, and on each interval, scale $\bar{x}(\cdot)$ so that the value at the start lies within the unit ball.

Specifically, let $T > 0$; we will choose a specific value for T later in the proof. Recall each iterate x_m is positioned at time $t(m) = \sum_{k=0}^{m-1} \alpha_k$ with $t(0) = 0$, as defined previously. With $m(0) := 0$ and $T_0 := 0$, define recursively, for $n \geq 0$,

$$m(n+1) := \min\{m : t(m) \geq T_n + T\}, \quad T_{n+1} := t\big(m(n+1)\big). \tag{3.8}$$

This divides $[0, \infty)$ into intervals $[T_n, T_{n+1})$, $n \geq 0$, with $T_{n+1} - T_n \to T$ as $n \to \infty$. To simplify expressions, we assume $\sup_n \alpha_n \leq 1$ in this proof, so that each interval is at most $T + 1$ in length. We then define a piecewise linear function $\hat{x}(\cdot)$ by scaling $\bar{x}(t)$ as follows: for each $n \geq 0$, with $r(n) := \|x_{m(n)}\| \vee 1$,

$$\hat{x}(t) := \bar{x}(t)/r(n) \quad \text{for } t \in [T_n, T_{n+1}). \tag{3.9}$$

As $\hat{x}(\cdot)$ can have "jumps" at times $T_1, T_2, \ldots$, to analyze the behavior of $\hat{x}(t)$ on the semiclosed interval $[T_n, T_{n+1})$, we introduce, for notational convenience, a "copy" denoted by $\hat{x}^n(t)$ defined on the closed interval $[T_n, T_{n+1}]$:

$$\hat{x}^n(t) := \hat{x}(t) \ \text{ for } t \in [T_n, T_{n+1}), \qquad \hat{x}^n(T_{n+1}) := \hat{x}(T_{n+1}^-) := \lim_{t \uparrow T_{n+1}} \hat{x}(t). \tag{3.10}$$

Let $x^n : [T_n, T_{n+1}] \to \mathbb{R}^d$ be the unique solution of the ODE defined by the scaled function $h_{r(n)}$ and the trajectory $\lambda(\cdot)$ given in (3.3), with initial condition $\hat{x}(T_n)$:

$$\dot{x}(t) = \lambda(t) h_{r(n)}(x(t)), \ \ t \in [T_n, T_{n+1}] \ \text{ with } x^n(T_n) = \hat{x}(T_n) = x_{m(n)}/r(n). \tag{3.11}$$

We aim to show that as $n \to \infty$, $\hat{x}^n(\cdot)$ "tracks" the ODE solution $x^n(\cdot)$.

A key intermediate result is proving $\sup_t \|\hat{x}(t)\| < \infty$. For synchronous SA (under stronger noise conditions than ours), this is shown in [11, Chap. 4.2] in several steps, starting with $\sup_t \mathbb{E}[\|\hat{x}(t)\|^2] < \infty$, proved through the bound in [11, Lem. 4.3] that for some constants $\bar{K}_1, \bar{K}_2$ independent of n,

$$\mathbb{E}\left[\|\hat{x}^n(t(k+1))\|^2\right]^{\frac{1}{2}} \leq e^{\bar{K}_1(T+1)}(1 + \bar{K}_2(T+1)), \quad m(n) \leq k < m(n+1). \tag{3.12}$$

In the asynchronous case here, we will take a similar approach. However, (3.12) need not hold for $\{\hat{x}^n(\cdot)\}$ in our case. By (3.2) and the definition of $\hat{x}(\cdot)$, we have that for k with $m(n) \leq k < m(n+1)$,

$$\hat{x}^n(t(k+1)) = \hat{x}^n(t(k)) + \alpha_k \Lambda_k h_{r(n)}(\hat{x}^n(t(k))) + \alpha_k \Lambda_k \hat{M}_{k+1} + \alpha_k \Lambda_k \hat{\epsilon}_{k+1}, \tag{3.13}$$

where $\hat{M}_{k+1} := M_{k+1}/r(n)$, $\hat{\epsilon}_{k+1} := \epsilon_{k+1}/r(n)$, and

$$\Lambda_k := \mathrm{diag}\big(b(k,1), b(k,2), \ldots, b(k,d)\big).$$

Since $r(n) \geq 1$ and $\hat{x}^n(t(k)) = x_k/r(n)$, by Assumption 2.2, a.s.,

$$\mathbb{E}[\|\hat{M}_{k+1}\|] < \infty,\ \mathbb{E}[\hat{M}_{k+1} \mid \mathcal{F}_k] = 0,\ \mathbb{E}[\|\hat{M}_{k+1}\|^2 \mid \mathcal{F}_k] \leq K_k(1 + \|\hat{x}^n(t(k))\|^2), \tag{3.14}$$

$$\|\hat{\epsilon}_{k+1}\| \leq \delta_{k+1}(1 + \|\hat{x}^n(t(k))\|). \tag{3.15}$$

While the scalars $\{K_k\}_{k\geq 0}$ in (3.14) and the entries of the diagonal matrices $\{\Lambda_k\}_{k\geq 0}$ are bounded a.s. by Assumption 2.2(i) and Lemma 3.1, respectively, they need not be bounded by a deterministic constant. While $\delta_{k+1} \to 0$ a.s. by Assumption 2.2(ii), there is no requirement on the conditional variance of δ_{k+1}. These factors prevent us from directly applying the arguments of [11, Chap. 4.2], which rely on the relation (3.12), to prove the desired boundedness of the scaled trajectory $\hat{x}(\cdot)$.

To work around this issue, we now use stopping techniques to construct "better-behaved" auxiliary processes $\tilde{x}^n(t)$ on $[T_n, T_{n+1}]$ for $n \geq 0$. Later we will relate them to $\hat{x}^n(\cdot)$ to establish $\sup_t \|\hat{x}(t)\| < \infty$.

Let C be the constant given by Lemma 3.1, and fix $\bar{a} > 0$. The following construction applies to each positive integer $N \geq 1$. For notational simplicity, however, we will temporarily suppress the indication of this dependence on N in the constructed processes $\tilde{x}^n(\cdot)$, $n \geq 0$, until we relate them to the original processes $\hat{x}^n(\cdot)$.

Let $N \geq 1$. For $n \geq 0$, define $k_n := k_{n,1} \wedge k_{n,2}$, where

$$k_{n,1} := \min\Big\{k \,\Big|\, K_k > N \text{ or } \max_{i\in\mathcal{I}} b(k,i) > C;\, m(n) \leq k < m(n+1)\Big\}, \tag{3.16}$$

$$k_{n,2} := \min\big\{k \,\big|\, \delta_k > \bar{a},\ m(n)+1 \leq k \leq m(n+1)\big\} \tag{3.17}$$

with $k_{n,1} := \infty$ and $k_{n,2} := \infty$ if the sets in their respective defining equations are empty. Then define $\tilde{x}^n(\cdot)$ on $[T_n, T_{n+1}]$ as follows. Let $\tilde{x}^n(T_n) := \hat{x}(T_n)$. For $m(n) \leq k < m(n+1)$, let

$$\tilde{x}^n(t(k+1)) = \tilde{x}^n(t(k)) + \alpha_k \tilde{\Lambda}_k h_{r(n)}(\tilde{x}^n(t(k))) + \alpha_k \tilde{M}_{k+1} + \alpha_k \tilde{\epsilon}_{k+1}, \tag{3.18}$$

where $\tilde{\Lambda}_k := \mathbb{1}\{k < k_n\}\Lambda_k$, $\tilde{M}_{k+1} := \tilde{\Lambda}_k \hat{M}_{k+1}$, and

$$\tilde{\epsilon}_{k+1} := \tilde{\Lambda}_k \cdot \mathbb{1}\{k+1 < k_{n,2}\}\hat{\epsilon}_{k+1}. \tag{3.19}$$

Finally, on the interval $(t(k), t(k+1))$, let $\tilde{x}^n(t)$ be the linear interpolation between $\tilde{x}^n(t(k))$ and $\tilde{x}^n(t(k+1))$. As can be seen, in each time interval $[T_n, T_{n+1}]$,

$$\tilde{x}^n(t) = \tilde{x}^n(t(k_n)) \quad \text{if } t \geq t(k_n) \text{ (where } t(\infty) := \infty),$$

$$\tilde{x}^n(t) = \hat{x}^n(t) \quad \text{for } t \leq t(k_n) \quad \text{if } k_n = k_{n,1} < k_{n,2}, \tag{3.20}$$

$$\tilde{x}^n(t) = \hat{x}^n(t) \quad \text{for } t \leq t(k_n - 1) \quad \text{if } k_n = k_{n,2} \leq k_{n,1}. \tag{3.21}$$

By definition $k_{n,1}, k_{n,2}$, and k_n are stopping times w.r.t. $\{\mathcal{F}_k\}$, so $\mathbb{1}\{k < k_n\}$ is $\mathcal{F}_k$-measurable and $\mathbb{1}\{k+1 < k_{n,2}\}$ is $\mathcal{F}_{k+1}$-measurable. Hence $\tilde{\Lambda}_k$ is $\mathcal{F}_k$-measurable, whereas $\tilde{M}_{k+1}$ and $\tilde{\epsilon}_{k+1}$ are $\mathcal{F}_{k+1}$-measurable. As the entries of the diagonal matrices Λ_k, $m(n) \leq k < k_n$, are all bounded by C, the matrix norm of $\tilde{\Lambda}_k = \mathbb{1}\{k < k_n\}\Lambda_k$ can be bounded by a deterministic constant $\bar{C}$:

$$\|\tilde{\Lambda}_k\| \leq \bar{C} \qquad \forall\, m(n) \leq k < m(n+1). \tag{3.22}$$

Moreover, by the construction of $\tilde{x}^n$ in (3.18)–(3.19) and Assumption 2.2 (cf. (3.14)–(3.15)), we have the following lemma.

LEMMA 3.5. *For $n \geq 0$ and all k with $m(n) \leq k < m(n+1)$, $\mathbb{E}[\|\tilde{M}_{k+1}\|] < \infty$, $\mathbb{E}[\tilde{M}_{k+1} \mid \mathcal{F}_k] = 0$ a.s., and*

$$\mathbb{E}[\|\tilde{M}_{k+1}\|^2 \mid \mathcal{F}_k] \leq \bar{C}^2 N(1 + \|\tilde{x}^n(t(k))\|^2) \quad a.s., \tag{3.23}$$

$$\|\tilde{\epsilon}_{k+1}\| \leq \bar{C}(\bar{a} \wedge \delta_{k+1})(1 + \|\tilde{x}^n(t(k))\|) \quad a.s. \tag{3.24}$$

Using (3.18), (3.22), and Lemma 3.5, we can derive a bound on $\mathbb{E}[\|\tilde{x}^n(t(k+1))\|^2]^{\frac{1}{2}}$ analogous to the bound (3.12), similarly to [11, Lem. 4.3]. From this key bound, we can then successively derive the four assertions in the lemma below by applying a convergence theorem for square-integrable martingales [19, Prop. VII-2-3(c)] to $\sum_{k=0}^{m-1} \alpha_k \tilde{M}_{k+1}$ and the arguments of [11, Chap. 4.2] to the auxiliary processes $\{\tilde{x}^n(\cdot)\}$.

LEMMA 3.6. *The following hold for $\{\tilde{x}^n(\cdot)\}$:*
(i) $\sup_{n\geq 0} \sup_{t \in [T_n, T_{n+1}]} \mathbb{E}[\|\tilde{x}^n(t)\|^2] < \infty$.
(ii) $\tilde{\zeta}_m := \sum_{k=0}^{m-1} \alpha_k \tilde{M}_{k+1}$ *converges a.s. in* $\mathbb{R}^d$ *as* $m \to \infty$.
(iii) $\sup_{n\geq 0} \sup_{t \in [T_n, T_{n+1}]} \|\tilde{x}^n(t)\| < \infty$ *a.s.*
(iv) $\lim_{n\to\infty} \sup_{t \in [T_n, t(k_n) \wedge T_{n+1}]} \|\tilde{x}^n(t) - x^n(t)\| = 0$ *a.s.*

Lemma 3.6 shows that the $\tilde{x}^n(t)$'s are bounded and asymptotically track the ODE solutions $x^n(t)$ (see (3.11)), albeit over the subintervals $[T_n, t(k_n) \wedge T_{n+1}]$ of $[T_n, T_{n+1}]$. We now use these auxiliary processes to establish the boundedness of the original scaled trajectories $\hat{x}^n(\cdot)$ and their relationship to the ODE solutions $x^n(\cdot)$.

LEMMA 3.7. *The following hold for $\{\hat{x}^n(\cdot)\}$:*
(i) $\sup_{n\geq 0} \sup_{t \in [T_n, T_{n+1}]} \|\hat{x}^n(t)\| < \infty$ *a.s.;*
(ii) $\lim_{n\to\infty} \sup_{t \in [T_n, T_{n+1}]} \|\hat{x}^n(t) - x^n(t)\| = 0$ *a.s.*

Proof. For each $N \geq 1$, denote the auxiliary processes constructed above by $\tilde{x}^n_N(\cdot)$ and their associated stopping times by $k_{N,n}$, $n \geq 0$. By Lemma 3.6, a.s., for all $N \geq 1$,

$$\sup_{n\geq 0} \sup_{t \in [T_n, T_{n+1}]} \|\tilde{x}^n_N(t)\| < \infty, \qquad \lim_{n\to\infty} \sup_{t \in [T_n, t(k_{N,n}) \wedge T_{n+1}]} \|\tilde{x}^n_N(t) - x^n(t)\| = 0. \tag{3.25}$$

Consider the set Ω' of sample paths for which (3.25) holds for all $N \geq 1$, $\sup_n K_n < \infty$, and, for all sufficiently large n, $\delta_n \leq \bar{a}$ and $\max_{i \in \mathcal{I}} b(n,i) \leq C$. This set Ω' has probability 1 by Lemma 3.6, Assumption 2.2, and Lemma 3.1. For each sample path $\omega \in \Omega'$, we have $\sup_n K_n < N(\omega)$ for some positive integer $N(\omega)$ and $k_{N(\omega),n} = \infty$ for all n sufficiently large; consequently, for all n sufficiently large, $t(k_{N(\omega),n}) \wedge T_{n+1} = T_{n+1}$ and $\hat{x}^n(\cdot)$ coincides with $\tilde{x}^n_{N(\omega)}(\cdot)$ by (3.20)–(3.21). Combining this with (3.25) yields the conclusions stated in parts (i), (ii). ∎

3.2.2. Stability in scaling limits of corresponding ODEs and proof completion. The ODE solution $x^n(\cdot)$ is determined by $h_{r(n)}$, $\lambda(T_n + \cdot)$, and an initial condition within the unit ball of $\mathbb{R}^d$ (see (3.11)). If $r(n)$ becomes sufficiently large, the initial condition lies on $\mathbb{S}_1 := \{x \in \mathbb{R}^d \mid \|x\| = 1\}$, and $h_{r(n)}$ approaches the function h_∞ due to Assumption 2.1(ii). By Lemma 3.2, a.s., any limit point of $\{\lambda(T_n + \cdot)\}_{n\geq 0}$ in Υ has the form $\lambda^*(t) = \rho(t) I$ with $\frac{1}{d} \leq \rho(t) \leq C$ for all $t \geq 0$.

This leads us to consider *an arbitrary function $\lambda^* \in \Upsilon$ of this form,* the set of which we denote by Υ^*, and the associated limiting ODE is

$$\dot{x}(t) = \lambda^*(t)\, h_\infty(x(t)) \quad \text{with} \quad x(0) \in \mathbb{S}_1. \tag{3.26}$$

Below, we examine stability properties for such ODEs and their "nearby" ODEs, by generalizing [11, Lem. 4.2, Cor. 4.1] from the synchronous context with a single limiting ODE to the asynchronous context with multiple limiting ODEs.

For $x \in \mathbb{S}_1$ and $\lambda^* \in \Upsilon^*$, let $\phi_\infty^{\lambda^*}(t;x)$ denote the unique solution of (3.26) with initial condition $\phi_\infty^{\lambda^*}(0;x) = x$. For $c \geq 1$ and $\lambda' \in \Upsilon$, let $\phi_{c,\lambda'}(t;x)$ denote the unique solution of $\dot{x}(t) = \lambda'(t)\, h_c(x(t))$ with $\phi_{c,\lambda'}(0;x) = x$. (Existence/uniqueness of these ODE solutions follow from standard Carathéodory-type results [26].)

Lemma 3.8. *There exists $T > 0$ such that for all $\lambda^* \in \Upsilon^*$ and initial conditions $x \in \mathbb{S}_1$, $\|\phi_\infty^{\lambda^*}(t;x)\| < 1/8$ for all $t \geq T$.*

Proof. For $\lambda^* = \rho(\cdot)I \in \Upsilon^*$, $\phi_\infty^{\lambda^*}(\cdot;x)$ is related by time scaling to $\phi_\infty^o(\cdot;x)$, the solution of the ODE $\dot{x}(t) = h_\infty(x(t))$ with initial condition $\phi_\infty^o(0;x) = x$; in particular, $\phi_\infty^{\lambda^*}(t;x) = \phi_\infty^o(\tau(t);x)$, where $\tau(t) = \int_0^t \rho(s)ds$. Under Assumption 2.1, there exists $T^o > 0$ such that for all $x \in \mathbb{S}_1$, $\|\phi_\infty^o(t;x)\| < 1/8$ for all $t \geq T^o$ [11, Lem. 4.1]. Since $\rho(t) \geq \frac{1}{d}$, this means that for all $t \geq T^o d$, $\|\phi_\infty^{\lambda^*}(t;x)\| = \|\phi_\infty^o(\tau(t);x)\| < 1/8$. ◻

The next three lemmas extend the stability property of the limiting ODEs (3.26), as given in the preceding lemma, to nearby ODEs within a certain time horizon.

Let L_h be the Lipschitz modulus of h (Assumption 2.1(i)). Consider the set H of all Lipschitz continuous functions $\bar{h} : \mathbb{R}^d \to \mathbb{R}^d$ with a Lipschitz modulus no greater than L_h and $\|\bar{h}(0)\| \leq \|h(0)\|$. Endow H with the topology of uniform convergence on compacts and a compatible metric, rendering H a compact metric space by the Arzelá–Ascoli theorem. The next lemma extends a similar result [9, Lem. 3.1(b)], which deals with a fixed function h rather than the set H, and can be proved using similar arguments:

Lemma 3.9. *Consider the function Ψ that maps each $(\bar{\lambda}, \bar{h}, x) \in \Upsilon \times H \times \mathbb{R}^d$ to the unique solution of the ODE $\dot{x}(t) = \bar{\lambda}(t)\bar{h}(x(t))$ on $[0,\infty)$ with initial condition $x(0) = x$. Then Ψ is continuous from $\Upsilon \times H \times \mathbb{R}^d$ into $\mathcal{C}([0,\infty);\mathbb{R}^d)$.*

Lemma 3.10. *There exist $\bar{T} > 0$, $\bar{c} \geq 1$, and a neighborhood $D(\Upsilon^*)$ of Υ^* in Υ such that for all $\lambda' \in D(\Upsilon^*)$ and initial conditions $x \in \mathbb{S}_1$, $\|\phi_{c,\lambda'}(t;x)\| < 1/4$ for all $t \in [\bar{T}, \bar{T}+1]$ and $c \geq \bar{c}$.*

Proof. Let $\bar{T}$ be the time T given by Lemma 3.8. As the function Ψ is continuous on $\Upsilon \times H \times \mathbb{R}^d$ (Lemma 3.9), it is uniformly continuous on the compact set $\Upsilon \times H \times \mathbb{S}_1$. Consequently, there exist a neighborhood $D(h_\infty)$ of h_∞ in H and, for some sufficiently small $\epsilon > 0$, an ϵ-neighborhood $D(\Upsilon^*)$ of Υ^* in Υ such that

$$\sup_{t \in [0,\bar{T}+1]} \left|\Psi(\lambda', h', x)(t) - \Psi(\lambda^*_{\lambda'}, h_\infty, x)(t)\right| \leq 1/8 \quad \forall x \in \mathbb{S}_1,\, h' \in D(h_\infty),\, \lambda' \in D(\Upsilon^*), \tag{3.27}$$

where $\lambda^*_{\lambda'}$ is any $\lambda^* \in \Upsilon^*$ within distance ϵ of λ'. Since $h_c \in D(h_\infty)$ for all c sufficiently large (Assumption 2.1(ii)), we obtain the desired conclusion by (3.27) and Lemma 3.8. ◻

Below, let $\bar{T} > 0$ and $\bar{c} \geq 1$ be given by Lemma 3.10, *and use $T := \bar{T} + 1/2$ in defining the sequence of times, $T_n, n \geq 0$, for the scaled trajectory $\hat{x}(\cdot)$ and the solutions $x^n(\cdot)$* introduced previously in section 3.2.1 [see (3.8), (3.9), and (3.11)].

Lemma 3.11. *Almost surely, there exists a sample path-dependent integer $\bar{n} \geq 0$ such that for all $n \geq \bar{n}$, $\lambda'_n := \lambda(T_n + \cdot) \in \Upsilon$ satisfies that $\|\phi_{c,\lambda'_n}(t;x)\| < 1/4$ for all $t \in [\bar{T}, \bar{T}+1]$, $c \geq \bar{c}$, and $x \in \mathbb{S}_1$.*

Proof. Consider a sample path for which Lemma 3.2 holds. Let G be the set of all limit points of $\{\lambda(T_n + \cdot)\}_{n \geq 0}$ in Υ. Since G is a closed subset of the compact metric space Υ, G is compact. By Lemma 3.2, $G \subset \Upsilon^* \subset D(\Upsilon^*)$, where $D(\Upsilon^*)$ is the neighborhood of Υ^* given by Lemma 3.10. Then, as $\lambda(T_n + \cdot) \overset{n\to\infty}{\to} G$, it follows that for some finite integer $\bar{n}$, $\lambda'_n = \lambda(T_n + \cdot) \in D(\Upsilon^*)$ for all $n \geq \bar{n}$. By Lemma 3.10, this implies the desired conclusion. ∎

Finally, noting that $x^n(T_n + \cdot) = \phi_{r(n),\lambda'_n}(\cdot\,; \hat{x}(T_n))$ on $[0, T_{n+1} - T_n]$, we can deduce from Lemmas 3.7 and 3.11 that the unscaled $\bar{x}(\cdot)$ has these properties a.s.:

(i) $\sup_t \|\bar{x}(t)\| < \infty$ if $\sup_n \|\bar{x}(T_n)\| < \infty$;

(ii) $\|\bar{x}(T_{n+1})\| \leq \bar{c}K^*$ if $\|\bar{x}(T_n)\| < \bar{c}$, where $K^* := \sup_t \|\hat{x}(t)\| < \infty$ (Lemma 3.7(i));

(iii) for all sufficiently large n, $\|\bar{x}(T_{n+1})\| \leq \frac{1}{2}\|\bar{x}(T_n)\|$ if $\|\bar{x}(T_n)\| \geq \bar{c}$.

These properties imply the a.s. boundedness of $\{x_n\}$, as shown in the proof of [11, Thm. 4.1]. (Details can be found in our preprint on arXiv [32].) This establishes Theorem 2.1.

3.3. Convergence given stability. Given the a.s. boundedness of the iterates $\{x_n\}$ (Theorem 2.1), we prove the convergence results (Theorem 2.2 and Corollary 2.1) by analyzing the trajectory $\bar{x}(\cdot)$ defined instead by (2.2), where the ODE-time is determined by the aggregated step sizes $\tilde{\alpha}_n = \sum_{i\in Y_n} \alpha_{\nu(n,i)}$ for $n \geq 0$, as recalled. For $\tilde{\lambda}(\cdot)$ defined by (3.5), we show similarly to [11, Lem. 2.1] that asymptotically, $\bar{x}(\cdot)$ tracks the solutions of the nonautonomous ODE

$$\dot{x}(t) = \tilde{\lambda}(t)h(x(t)). \tag{3.28}$$

By Lemma 3.4, the limiting ODE is unique and autonomous:

$$\dot{x}(t) = \frac{1}{d}h(x(t)), \tag{3.29}$$

which allows us to apply [9, Lem. 3.1(b)] (cf. Lemma 3.9) to conclude $\bar{x}(\cdot)$ tracks the solutions of the limiting ODE (3.29). We then obtain Theorem 2.2(i) by essentially the same reasoning as in [11, Thm. 2.1] for synchronous SA. This analysis also yields Theorem 2.2(ii) by standard arguments, and Corollary 2.1 follows directly from Theorem 2.2. Proofs are omitted here but provided in our arXiv preprint [32].

4. Further convergence analysis via shadowing properties. In this section, we prove Theorem 2.3, which establishes the convergence of the asynchronous SA algorithm (2.1) to a unique equilibrium point in E_h. The proof is based on an analysis of the shadowing properties of the trajectories of the algorithm (2.1), building on prior work by Hirsch and Benaim.

To keep the notation concise, we will not repeat a.s. in every instance below. When working with sample path properties, it should be understood that *we are considering a sample path on which all previously established a.s. properties (e.g., stability and convergence) hold, along with the a.s. conditions required by Theorem* 2.3. *Any properties proven to hold a.s. in the analysis below will apply to this path from the point they are established.*

4.1. Main argument and proof structure for Theorem 2.3. For algorithm (2.1), we consider the interpolated trajectory $\bar{x}(t)$ defined by (2.2) and analyze its behavior at discrete moments $t = j \in \{0, 1, \ldots\}$. For each $j \geq 0$, let $\tilde{x}^j(\cdot)$ and $x^j(\cdot)$ denote the unique solutions on the interval $[j, j+1]$ of the ODEs (3.28) and (3.29), respectively, with initial conditions $\tilde{x}^j(j) = x^j(j) = \bar{x}(j)$.

Part (i) of the next proposition follows from Hirsch's proof of his shadowing theorem [17] (see also Benaim [3, sect. 5] and [4, sect. 8.2]) and is the key argument in the proof of Theorem 2.3. We include the proof of part (i) for completeness and clarity, as we use the Lipschitz continuity of the mappings involved here in place of the 'expansion rates' of smooth (C^1) flows/semiflows considered in [3, 4, 17]. Recall that L_h is the Lipschitz constant of h w.r.t. $\|\cdot\|_\infty$. (Below, we take $\|\cdot\| = \|\cdot\|_\infty$.)

PROPOSITION 4.1. *If* $\limsup_{j\to\infty} \frac{1}{j}\ln(\|\bar{x}(j+1) - x^j(j+1)\|) < -L_h/d$, *then*

(i) *there exists* $z_* \in \mathbb{R}^d$ *such that, for some* $\delta \in (0, e^{-L_h/d})$ *and a path-dependent constant* c, *the inequality* $\|\bar{x}(j) - z(j)\| \le c\delta^j$ *holds for all* $j \ge 0$, *where* $z(\cdot)$ *is the solution of the limiting ODE* (3.29), $\dot{x}(t) = \frac{1}{d}h(x(t))$, *with initial condition* $z(0) = z_*$;

(ii) *the sequence* $\{x_n\}$ *from algorithm* (2.1) *converges to some point in* E_h.

Proof. For the limiting ODE (3.29), let $F : \mathbb{R}^d \to \mathbb{R}^d$ be the map that takes an initial condition $y(0) = y$ to the ODE solution at time 1. Then F^{-1} maps an initial condition y to the ODE solution at time -1 and by Gronwall's inequality,

$$\|F^{-1}(y) - F^{-1}(y')\| \le e^{L_h/d}\|y - y'\| \quad \forall\, y, y' \in \mathbb{R}^d. \tag{4.1}$$

For $j \ge 0$, denote $y_j = \bar{x}(j)$. The condition of the proposition implies that for some $\beta > L_h/d$ and sufficiently large $\hat{j}$, we have

$$\|y_{j+1} - F(y_j)\| < e^{-\beta j} < e^{-L_h j/d} \quad \forall\, j \ge \hat{j}. \tag{4.2}$$

Choose $\delta \in (0,1)$ with $e^{-\beta} < \delta < e^{-L_h/d}$. Let $B_j := B(y_j, \delta^j)$ denote the closed ball centered at y_j with radius δ^j. For $j \ge \hat{j}$ and $z \in B(y_{j+1}, \delta^{j+1})$, by (4.1) and (4.2),

$$\begin{aligned}\|F^{-1}(z) - y_j\| &= \|F^{-1}(z) - F^{-1}(F(y_j))\| \le e^{L_h/d}\|z - F(y_j)\| \\ &\le e^{L_h/d}(\|z - y_{j+1}\| + \|y_{j+1} - F(y_j)\|) \le e^{L_h/d}(\delta^{j+1} + e^{-\beta j}),\end{aligned}$$

which, in view of the choice of δ, implies that $\|F^{-1}(z) - y_j\| \le \delta^j$ for all j sufficiently large. Thus, there exists $\bar{j} \ge \hat{j}$ such that $F^{-1}(B_{j+1}) \subset B_j$ for all $j \ge \bar{j}$. Consequently,

$$F^{-i}(B_{\bar{j}+i}) \subset F^{-i+1}(B_{\bar{j}+i-1}) \subset \cdots \subset F^{-1}(B_{\bar{j}+1}) \subset B_{\bar{j}} \qquad \forall\, i \ge 0. \tag{4.3}$$

The intersection of these sets, $\cap_{i=0}^\infty F^{-i}(B_{\bar{j}+i})$, contains a single point $\bar{z}$ (since these sets form a nested sequence of nonempty compact sets, with shrinking diameters as implied by (4.1) and the choice of δ).

From $\bar{z} \in \cap_{i=0}^\infty F^{-i}(B_{\bar{j}+i})$ we obtain a desired shadowing property:

$$\|y_{\bar{j}+i} - F^i(\bar{z})\| \le \delta^{\bar{j}+i} \to 0, \quad \text{as } i \to \infty. \tag{4.4}$$

Expressing (4.4) in terms of $\bar{x}(\cdot)$ and the solution $z(t)$ to the ODE (3.29) with initial condition $z(0) = F^{-\bar{j}}(\bar{z}) =: z_*$, we have that, for some (path-dependent) constant c, $\|\bar{x}(j) - z(j)\| \le c\,\delta^j \to 0$ as $j \to \infty$, which proves part (i).

By the condition of Theorem 2.3 on the ODE (3.29), $z(j) \to z^*$ for some $z^* \in E_h$. Thus, the preceding inequality gives $\bar{x}(j) \to z^*$. Then, by Corollary 2.1(ii), the sequence $\{x_n\}$ must also converge to z^*, proving part (ii). ∎

Clearly, Theorem 2.3 follows immediately if we show that the condition in Proposition 4.1 holds. To this end, decompose the tracking error of the trajectory $\bar{x}(\cdot)$ as

$$\|\bar{x}(j+1) - x^j(j+1)\| \le \|\bar{x}(j+1) - \tilde{x}^j(j+1)\| + \|\tilde{x}^j(j+1) - x^j(j+1)\|, \tag{4.5}$$

separating the first component, attributable to stochastic noise and representing the error in tracking the solutions of the nonautonomous ODE (3.28), from the second component, which represents the error due to asynchrony. In the next two subsections, we analyze each component separately and prove that, under the conditions of Theorem 2.3,

$$\limsup_{j\to\infty} \frac{1}{j} \ln \left(\|\bar{x}(j+1) - \tilde{x}^j(j+1)\| \right) < -L_h/d, \tag{4.6}$$

$$\limsup_{j\to\infty} \frac{1}{j} \ln \left(\|\tilde{x}^j(j+1) - x^j(j+1)\| \right) < -L_h/d. \tag{4.7}$$

This will establish that the condition in Proposition 4.1 is met and hence Theorem 2.3 holds.

Define $\mathcal{K}(t) := \max\{n \geq 0 \mid \tilde{t}(n) \leq t\}$. Note that $\mathcal{K}(t)$ is a stopping time, since $\tilde{t}(n) = \sum_{k=0}^{n-1} \tilde{\alpha}_k$ (as defined before (2.2)) is $\mathcal{F}_{n-1}$-measurable. Let

$$\mathcal{K}_1[j] := \{k \mid \mathcal{K}(j) \leq k \leq \mathcal{K}(j+1)\}.$$

Recall from Remark 2.4 the quantity $\ell(\{\alpha_n\}) = \limsup_{n\to\infty} \frac{\ln(\alpha_n)}{\sum_{k=0}^{n} \alpha_k}$ associated with the step size sequence $\{\alpha_n\}$. For the class-1 and class-2 step sizes defined in section 2.2, we have $\ell(\{\alpha_n\}) = -A$ and $-\infty$, respectively, where A is the step size scaling parameter. Most proofs below rely only on the quantity $\ell(\{\alpha_n\})$ and the nonincreasing property of the step sizes, without assuming specific step size forms—except near the end of the proof of (4.7) (in Lemma 4.5, section 4.3). We first prove (4.6).

4.2. Proof for (4.6). Recall the equivalent form (3.4) of algorithm (2.1) in vector notation:

$$x_{n+1} = x_n + \tilde{\alpha}_n \tilde{\Lambda}_n \left(h(x_n) + M_{n+1} + \epsilon_{n+1} \right), \tag{4.8}$$

where $\tilde{\Lambda}_n$ is a $d \times d$-diagonal matrix given by $\tilde{\Lambda}_n := \mathrm{diag}\big(\tilde{b}(n,1), \tilde{b}(n,2), \ldots, \tilde{b}(n,d)\big)$ with diagonal entries $\tilde{b}(n,i) = \frac{\alpha_{\nu(n,i)}}{\tilde{\alpha}_n} \mathbb{1}\{i \in Y_n\} \in [0,1]$. (Note that $\sum_n \tilde{\alpha}_n = \infty$ and $\sum_n \tilde{\alpha}_n^2 < \infty$ a.s. by Assumptions 2.3(i) and 2.4(i).)

By (4.8) and the definitions of $\bar{x}(\cdot)$ and $\tilde{x}^j(\cdot)$, we can bound $\|\bar{x}(j+1) - \tilde{x}^j(j+1)\|$ using similar derivations as in the proof of [11, Lemma 2.1], yielding that for all $j \geq 0$,

$$\|\bar{x}(j+1) - \tilde{x}^j(j+1)\| \leq c_0 \sup_{k \in \mathcal{K}_1[j]} \tilde{\alpha}_k + c_1 D_1(j) + c_2 D_2(j). \tag{4.9}$$

Here, the c_i's are constants that depend on the sample path but not on j, and

$$D_1(j) := \sup_{k \in \mathcal{K}_1[j]} \left\| \sum_{i=\mathcal{K}(j)}^{k} \tilde{\alpha}_i \tilde{\Lambda}_i \epsilon_{i+1} \right\|, \qquad D_2(j) := \sup_{k \in \mathcal{K}_1[j]} \left\| \sum_{i=\mathcal{K}(j)}^{k} \tilde{\alpha}_i \tilde{\Lambda}_i M_{i+1} \right\|. \tag{4.10}$$

Before proceeding to bound the three terms on the right-hand side (r.h.s.) of (4.9), we derive some limiting properties of the step sizes that will be needed in our analysis.

LEMMA 4.1. *The following hold a.s.:*
(i) *For all $i' \in \mathcal{I}$, $\lim_{j\to\infty} \frac{1}{j} \sum_{k=0}^{\mathcal{K}(j)} \alpha_{\nu(k,i')} \mathbb{1}\{i' \in Y_k\} = \lim_{j\to\infty} \frac{1}{j} \sum_{k=0}^{\mathcal{K}(j)} \alpha_k = \frac{1}{d}$.*
(ii) $\limsup_{j\to\infty} \frac{1}{j} \ln(\alpha_{\mathcal{K}(j)}) \leq \frac{\ell(\{\alpha_n\})}{d}$.
(iii) $\limsup_{j\to\infty} \frac{1}{j} \ln(\sum_{i' \in \mathcal{I}} \alpha_{\nu(\mathcal{K}(j),i')}) \leq \frac{\ell(\{\alpha_n\})}{d}$.

Proof. Part (i) follows from Assumptions 2.3(iii), 2.4(i) and the definition of the ODE-time j via the step sizes $\{\tilde{\alpha}_n\}$. Specifically, recall the constant $\Delta \in (0,1]$

from Assumption 2.4(i), and let $\Delta' \in (0,\Delta)$. By Assumption 2.3(iii), as $j \to \infty$, $\frac{\sum_{k=0}^{[y\mathcal{K}(j)]} \alpha_k}{\sum_{k=0}^{\mathcal{K}(j)} \alpha_k} \to 1$ uniformly in $y \in [\Delta', 1]$. Combined with Assumption 2.4(i), this yields $\frac{\sum_{k=0}^{\mathcal{K}(j)} \alpha_{\nu(k,i')} \mathbb{1}\{i' \in Y_k\}}{\sum_{k=0}^{\mathcal{K}(j)} \alpha_k} \to 1$ for all $i' \in \mathcal{I}$. Since $\lim_{j\to\infty} \left| j - \sum_{i'\in\mathcal{I}} \sum_{k=0}^{\mathcal{K}(j)} \alpha_{\nu(k,i')} \mathbb{1}\{i' \in Y_k\}\right| = 0$, part (i) follows. By part (i), $\limsup_{j\to\infty} \frac{1}{j} \ln(\alpha_{\mathcal{K}(j)}) = \limsup_{j\to\infty} \frac{\ln(\alpha_{\mathcal{K}(j)})}{d\sum_{k=0}^{\mathcal{K}(j)} \alpha_k} \le \frac{\ell(\{\alpha_n\})}{d}$, proving part (ii).

By Assumption 2.3(ii), we have that for some constant c, $\frac{\alpha_{[\Delta'\mathcal{K}(j)]}}{\alpha_{\mathcal{K}(j)}} \le c$ for all $j \ge 0$. Since $\{\alpha_n\}$ is nonincreasing, we get from this inequality and Assumption 2.4(i) that for all sufficiently large j, $\sum_{i'\in\mathcal{I}} \alpha_{\nu(\mathcal{K}(j),i')} \le \sum_{i'\in\mathcal{I}} \alpha_{[\Delta'\mathcal{K}(j)]} \le dc\alpha_{\mathcal{K}(j)}$. Part (iii) now follows from part (ii). ∎

Bounding the first two terms on the r.h.s. of (4.9) is straightforward.

LEMMA 4.2. *A.s., with μ_δ as given in Assumption* 2.5*,*

$$\limsup_{j\to\infty} \frac{1}{j} \ln\Big(\sup_{k\in\mathcal{K}_1[j]} \tilde{\alpha}_k\Big) \le \frac{\ell(\{\alpha_n\})}{d}, \qquad \limsup_{j\to\infty} \frac{1}{j} \ln\big(D_1(j)\big) \le \frac{\mu_\delta}{d}. \tag{4.11}$$

Proof. Since $\sup_{k\in\mathcal{K}_1[j]} \tilde{\alpha}_k \le \sum_{i'\in\mathcal{I}} \alpha_{\nu(\mathcal{K}(j),i')}$ (due to $\{\alpha_k\}$ being nonincreasing), Lemma 4.1(iii) gives the first bound in (4.11). From Assumption 2.2(ii) on $\{\epsilon_k\}$, the a.s. boundedness of $\{x_n\}$ (Theorem 2.1), and the fact that $\sum_{i=\mathcal{K}(j)}^{\mathcal{K}(j+1)} \tilde{\alpha}_i \le 1 + 2d\sup_n \alpha_n$, it follows that $D_1(j) \le c\sup_{k\in\mathcal{K}_1[j]} \delta_{k+1}$ for some (sample path-dependent) constant c. Then, by Assumption 2.5 on $\{\delta_k\}$ and Lemma 4.1(i), we obtain $\limsup_{j\to\infty} \frac{1}{j} \ln\big(D_1(j)\big) = \limsup_{j\to\infty} \frac{\ln\big(D_1(j)\big)}{d\sum_{k=0}^{\mathcal{K}(j)} \alpha_k} \le \frac{\mu_\delta}{d}$, proving the second bound in (4.11). ∎

We now bound the third term $D_2(j)$ in (4.9) similarly to Benaïm [4, proofs of Props. 4.2, 8.3], with additional effort to account for the random step sizes $\{\tilde{\alpha}_i\}$ and more general noise terms $\{M_i\}$ in our setting.

LEMMA 4.3. *A.s.,* $\limsup_{j\to\infty} \frac{1}{j} \ln(D_2(j)) \le \frac{\ell(\{\alpha_n\})}{2d}$.

Proof. For integers $N \ge 1$, define stopping times τ_N and auxiliary variables $M_i^{(N)}$,

$$\tau_N := \min\{i \ge 0 : \|x_i\| > N \text{ or } K_i > N\}, \qquad M_{i+1}^{(N)} := \mathbb{1}\{i < \tau_N\} M_{i+1}, \quad i \ge 0,$$

where $\{K_i\}$ is the a.s. bounded random sequence in Assumption 2.2(i). By Assumption 2.2(i), for each N, $\{M_i^{(N)}\}_{i\ge1}$ is a martingale-difference sequence satisfying $\sup_i \mathbb{E}[\|M_{i+1}^{(N)}\|_2^2 \mid \mathcal{F}_i] \le c_N$ for some constant c_N depending on N. Since $\{x_i\}$ is bounded a.s. by Theorem 2.1, a.s., $\{M_i\}$ coincides with $\{M_i^{(N)}\}$ for some sample path-dependent value of N. Thus, to prove the lemma, it suffices to prove that a.s. for *each* N,

$$\limsup_{j\to\infty} \frac{1}{j} \ln\big(\tilde{D}_2(j)\big) \le \frac{\ell(\{\alpha_n\})}{2d}, \quad \text{where } \tilde{D}_2(j) := \sup_{k\in\mathcal{K}_1[j]} \left\| \sum_{i=\mathcal{K}(j)}^{k} \tilde{\alpha}_i \tilde{\Lambda}_i M_{i+1}^{(N)} \right\|. \tag{4.12}$$

To this end, fix N and let $T = 1 + d\sup_n a_n$. For each n, define

$$\zeta_{n,k} := \sum_{i=n}^{k-1} \tilde{\alpha}_i \cdot \mathbb{1}\{\tilde{t}(i) \le \tilde{t}(n) + T\} \cdot \tilde{\Lambda}_i M_{i+1}^{(N)} \quad \text{for } k \ge n \quad \text{with } \zeta_{n,n} := 0.$$

Then $\{\zeta_{n,k}\}_{k\geq n}$ is a square-integrable martingale w.r.t. $\{\mathcal{F}_k\}_{k\geq n}$, and

$$\lim_{\bar{k}\to\infty} \mathbb{E}\Big[\sup_{k\leq \bar{k}} \|\zeta_{n,k}\|_2^2 \,\Big|\, \mathcal{F}_n\Big] = \mathbb{E}\left[\sup_{k:\, \tilde{t}(n)\leq \tilde{t}(k)\leq \tilde{t}(n)+T} \Big\| \sum_{i=n}^{k} \tilde{\alpha}_i \tilde{\Lambda}_i M_{i+1}^{(N)} \Big\|_2^2 \,\Bigg|\, \mathcal{F}_n\right] \tag{4.13}$$

by the monotone convergence theorem. For $\bar{k} > n$, applying Doob's and Burkholder's inequalities [16, Thms. 2.2, 2.10], we obtain that for some constant c_0 (independent of the specific problem data),

(4.14)

$$\begin{aligned}
\mathbb{E}\Big[\sup_{k\leq \bar{k}} \|\zeta_{n,k}\|_2^2 \mid \mathcal{F}_n\Big] &\leq c_0\, \mathbb{E}\left[\sum_{i=n}^{\bar{k}-1} \Big\| \tilde{\alpha}_i \, \mathbb{1}\{\tilde{t}(i) \leq \tilde{t}(n)+T\}\, \tilde{\Lambda}_i M_{i+1}^{(N)} \Big\|_2^2 \,\Big|\, \mathcal{F}_n\right] \\
&\leq c_0\, \mathbb{E}\left[\sum_{i=n}^{\bar{k}-1} \tilde{\alpha}_i^2 \, \mathbb{1}\{\tilde{t}(i) \leq \tilde{t}(n)+T\}\, \|\tilde{\Lambda}_i\|_2^2 \cdot \mathbb{E}\big[\|M_{i+1}^{(N)}\|_2^2 \mid \mathcal{F}_i\big] \,\Big|\, \mathcal{F}_n\right] \\
&\leq c_0 c_1 c_N\, \mathbb{E}\left[\sum_{i=n}^{\bar{k}-1} \tilde{\alpha}_i^2 \, \mathbb{1}\{\tilde{t}(i) \leq \tilde{t}(n)+T\} \,\Big|\, \mathcal{F}_n\right] \\
&\leq c_2\, \mathbb{E}\left[\left(\sum_{i=n}^{\mathcal{K}(\tilde{t}(n)+T)} \tilde{\alpha}_i\right) \cdot \sup_{n\leq i\leq \mathcal{K}(\tilde{t}(n)+T)} \tilde{\alpha}_i \,\Big|\, \mathcal{F}_n\right] \leq c_3 \sum_{i'\in\mathcal{I}} \alpha_{\nu(n,i')}. \qquad (4.15)
\end{aligned}$$

Here, c_1 is a deterministic bound on $\sup_{i\geq 0} \|\tilde{\Lambda}_i\|_2^2$. To derive (4.14), we also used the bound $\sup_{i\geq 0} \mathbb{E}\big[\|M_{i+1}^{(N)}\|_2^2 \mid \mathcal{F}_i\big] \leq c_N$. In (4.15), c_2 and c_3 are deterministic constants independent of n and $\bar{k}$, specifically, $c_2 = c_0 c_1 c_N$ and $c_3 = c_2(T + d\sup_i \alpha_i)$. In the last inequality in (4.15), we used the fact that $\sum_{i=n}^{\mathcal{K}(\tilde{t}(n)+T)} \tilde{\alpha}_i \leq T + d\sup_i \alpha_i$ and the step sizes are nonincreasing.

Combining (4.13) and (4.15), we obtain

$$\mathbb{E}\left[\sup_{k:\, \tilde{t}(n)\leq \tilde{t}(k)\leq \tilde{t}(n)+T} \Big\| \sum_{i=n}^{k} \tilde{\alpha}_i \tilde{\Lambda}_i M_{i+1}^{(N)} \Big\|_2^2 \,\Bigg|\, \mathcal{F}_n\right] \leq c_3 \sum_{i'\in\mathcal{I}} \alpha_{\nu(n,i')} \qquad \forall\, n \geq 0.$$

Then by [19, Prop. II-1-3], the above inequality holds for any stopping time τ (w.r.t. $\{\mathcal{F}_n\}$), with τ in place of n and with $\mathcal{F}_\tau$ in place of $\mathcal{F}_n$. Consequently, for the stopping times $\mathcal{K}(j), j \geq 1$, we obtain the bounds

$$\mathbb{E}\left[\sup_{k\in\mathcal{K}_1[j]} \Big\| \sum_{i=\mathcal{K}(j)}^{k} \tilde{\alpha}_i \tilde{\Lambda}_i M_{i+1}^{(N)} \Big\|_2^2 \,\Big|\, \mathcal{F}_{\mathcal{K}(j)}\right] \leq c_3 \sum_{i'\in\mathcal{I}} \alpha_{\nu(\mathcal{K}(j),i')} \quad \forall\, j \geq 1, \tag{4.16}$$

where we have also used the fact that $\mathcal{K}_1[j] \subset \{k : \tilde{t}(\mathcal{K}(j)) \leq \tilde{t}(k) \leq \tilde{t}(\mathcal{K}(j)) + T\}$.

Finally, to prove (4.12), consider any $\beta > 0$ with $\beta + \frac{\ell(\{\alpha_n\})}{2d} < 0$. Since $\|x\|_\infty \leq \|x\|_2$, applying (4.16), Markov's inequality, and Lemma 4.1(iii), we obtain

$$\sum_{j=1}^{\infty} \Pr\left(\tilde{D}_2(j) \geq e^{-\beta j} \,\Big|\, \mathcal{F}_{\mathcal{K}(j)}\right) \leq c_3 \sum_{j=1}^{\infty} e^{2\beta j} \cdot \left(\sum_{i'\in\mathcal{I}} \alpha_{\nu(\mathcal{K}(j),i')}\right) < \infty. \tag{4.17}$$

By the extended Borel–Cantelli lemma [16, Cor. 2.3], this implies that, a.s., $\tilde{D}_2(j) < e^{-\beta j}$ for all sufficiently large j.[6] Letting $\beta \uparrow -\frac{\ell(\{\alpha_n\})}{2d}$ then proves (4.12). $\square$

By (4.9)–(4.10) and Lemmas 4.2 and 4.3, inequality (4.6) holds a.s., namely,

$$\limsup_{j\to\infty} \frac{1}{j} \ln\left(\|\bar{x}(j+1) - \tilde{x}^j(j+1)\|\right) < -L_h/d \quad a.s.,$$

if $\max\{\frac{\ell(\{\alpha_n\})}{2}, \mu_\delta\} < -L_h$. This condition becomes $\max\{\frac{-A}{2}, \mu_\delta\} < -L_h$ for class-1 step sizes and $\mu_\delta < -L_h$ for class-2 step sizes. Since these are the conditions required in Theorem 2.3, it follows that (4.6) holds.

4.3. Proof for (4.7). We now proceed to establish (4.7). Since $\tilde{x}^j(\cdot)$ and $x^j(\cdot)$ are solutions of the ODEs (3.28) and (3.29), respectively, with $\tilde{x}^j(j) = x^j(j) = \bar{x}(j)$, we have that for $t \in [j, j+1]$,

$$\begin{aligned}\tilde{x}^j(t) - x^j(t) &= \int_j^t \tilde{\lambda}(\tau) h(\tilde{x}^j(\tau)) d\tau - \int_j^t \frac{1}{d} h(x^j(\tau)) d\tau \\ &= \int_j^t \left(\tilde{\lambda}(\tau) - \frac{1}{d} I\right) h(x^j(\tau)) d\tau + \int_j^t \tilde{\lambda}(\tau)\left(h(\tilde{x}^j(\tau)) - h(x^j(\tau))\right) d\tau.\end{aligned}$$

As the trajectory $\tilde{\lambda}(\cdot)$ is bounded by definition (3.5) and h is Lipschitz continuous, applying Gronwall's inequality yields that for some constant c independent of j,

$$\|\tilde{x}^j(j+1) - x^j(j+1)\| \leq c \sup_{s\in[0,1]} \left\| \int_0^s \left(\tilde{\lambda}(j+\tau) - \frac{1}{d} I\right) h(x^j(j+\tau)) d\tau \right\|. \tag{4.18}$$

For each $i \in \mathcal{I}$, let $F_i^j(\cdot)$ denote the distribution function of the measure with density $\tilde{\lambda}_{ii}(j+\cdot)$ on $[0,1]$; that is, $F_i^j(s) = \int_0^s \tilde{\lambda}_{ii}(j+\tau) d\tau$, $s \in [0,1]$.

LEMMA 4.4. *A.s., for some (sample path-dependent) constant c,*

$$\|\tilde{x}^j(j+1) - x^j(j+1)\| \leq c \sup_{i\in\mathcal{I}} \sup_{s\in[0,1]} \left| F_i^j(s) - \frac{s}{d} \right| \qquad \forall j \geq 0 \tag{4.19}$$

Proof. First, observe that on $[0,1]$, the family of functions $h(x^j(j+\cdot))$ for $j \geq 0$ is uniformly bounded and uniformly Lipschitz continuous with a common Lipschitz constant that depends on the sample path. (This follows from the Lipschitz continuity of h, the definition of the ODE solutions $x^j(\cdot)$, and the boundedness of the iterates $\{x_n\}$.) Now, for a fixed $j \geq 0$ and for each $i \in \mathcal{I}$, consider the ith component of the integral term in (4.18), $\psi(s) := \int_0^s (\tilde{\lambda}_{ii}(j+\tau) - \frac{1}{d}) g_i(\tau) d\tau$, where $g_i(\tau) := h_i(x^j(j+\tau))$. By the integration by parts formula for the absolutely continuous functions $g_i(\tau)$ and $F_i^j(\tau) - \frac{\tau}{d}$, we have

$$\psi(s) = g_i(s)\left(F_i^j(s) - \frac{s}{d}\right) - \int_0^s \left(F_i^j(\tau) - \frac{\tau}{d}\right) g_i'(\tau) d\tau, \qquad s \in [0,1], \tag{4.20}$$

where the derivative $g_i'(\tau)$ is defined a.e. and bounded by the Lipschitz constant of g_i. In view of the uniform boundedness and uniform Lipschitz continuity of the family

[6] Details: To apply the extended Borel–Cantelli lemma [16, Cor. 2.3], take an integer $m \geq 1 + d\sup_n \alpha_n$ and consider separately each subsequence $\{\tilde{D}_2(i+km)\}_{k\geq 0}$ for $i = 1, \ldots, m$. Note that $\tilde{D}_2(j)$ is $\mathcal{F}_{\mathcal{K}(j+1)+1}$-measurable by definition, while $\mathcal{K}(j+1) + 1 \leq \mathcal{K}(j+m)$. Therefore, $\tilde{D}_2(j)$ is $\mathcal{F}_{\mathcal{K}(j+m)}$-measurable. Consequently, for each $i \leq m$, [16, Cor. 2.3] applies to the subsequence $\{\tilde{D}_2(i+km)\}_{k\geq 0}$, and (4.17) implies $\tilde{D}_2(i+km) < e^{-\beta(i+km)}$ for all sufficiently large k.

of functions $\{h(x^j(j+\cdot))\}_{j\geq 0}$ on $[0,1]$, the expression (4.20) implies that for some (path-dependent) constant c independent of j, $|\psi(s)| \leq c \sup_{\tau\in[0,s]} |F_i^j(\tau) - \frac{\tau}{d}|$. This, together with inequality (4.18), leads to the desired bound (4.19) for all j. ∎

Lemma 3.4 implies that for each $i \in \mathcal{I}$, as $j \to \infty$, the measure on $[0,1]$ with density $\tilde{\lambda}_{ii}(j+\cdot)$ converges setwise to $\frac{1}{d}$ times the Lebesgue measure. Therefore, the r.h.s. of (4.19) converges to 0 as $j \to \infty$. The next lemma addresses the speed of this convergence—sufficiently rapid convergence is needed for inequality (4.7) to hold.

LEMMA 4.5. *A.s.,*

$$\limsup_{j\to\infty} \frac{1}{j} \ln\left(\sup_{s\in[0,1]} \left| F_i^j(s) - \frac{s}{d} \right| \right) \leq \frac{\ell}{d} \qquad \forall\, i \in \mathcal{I}, \tag{4.21}$$

where $\ell = -(\gamma \wedge 1)A$ for class-1 stepsizes (with γ being the constant in Assumption 2.6(i)), and $\ell = -A$ for class-2 step sizes (where A is the step size scaling factor in both cases).

Proof. By definition the distribution functions F_i^j satisfy that for $s \in [0,1]$,

$$F_i^j(s) = \int_0^s \tilde{\lambda}_{ii}(j+\tau)d\tau, \qquad \sum_{i\in\mathcal{I}} F_i^j(s) = \int_0^s \sum_{i\in\mathcal{I}} \tilde{\lambda}_{ii}(j+\tau)d\tau = s \tag{4.22}$$

(since $\sum_{i\in\mathcal{I}} \tilde{\lambda}_{ii}(\cdot) \equiv 1$ by the definition of $\tilde{\lambda}(\cdot)$; see (3.5)). Hence, for each $\bar{i} \in \mathcal{I}$,

$$F_{\bar{i}}^j(s) - \frac{s}{d} = F_{\bar{i}}^j(s) - \frac{1}{d}\sum_{i\in\mathcal{I}} F_i^j(s) = F_{\bar{i}}^j(s) - G^j(s) - \frac{1}{d}\sum_{i\in\mathcal{I}} \big(F_i^j(s) - G^j(s)\big), \tag{4.23}$$

where $G^j(s) := \int_{\mathcal{K}(j)}^{\mathcal{K}(j+s)} g_\alpha(y)dy$ with $g_\alpha(y) = \frac{1}{Ay}$ for class-1 $\{\alpha_n\}$ and $g_\alpha(y) = \frac{1}{Ay\ln y}$ for class-2 $\{\alpha_n\}$. From (4.23) we obtain

$$\left| F_{\bar{i}}^j(s) - \frac{s}{d} \right| \leq 2 \sup_{i\in\mathcal{I}} \left| F_i^j(s) - G^j(s) \right|, \quad s \in [0,1]. \tag{4.24}$$

For large j, we have $F_i^j(s) \approx \sum_{k=\nu(\mathcal{K}(j),i)}^{\nu(\mathcal{K}(j+s),i)} \alpha_k \approx \int_{\nu(\mathcal{K}(j),i)}^{\nu(\mathcal{K}(j+s),i)} g_\alpha(y)dy$. Specifically, the difference in the second approximation is bounded by $\alpha_{\nu(\mathcal{K}(j),i)}$. The difference in the first approximation, based on the first relation in (4.22) and the definitions of the ODE-time and $\tilde{\lambda}(\cdot)$ (see (3.5)), can be bounded by

$$\left| F_i^j(s) - \sum_{k=\nu(\mathcal{K}(j),i)}^{\nu(\mathcal{K}(j+s),i)} \alpha_k \right| \leq \alpha_{\nu(\mathcal{K}(j),i)} + \alpha_{\nu(\mathcal{K}(j+s),i)}.$$

Since $\{\alpha_n\}$ is nonincreasing, $\alpha_{\nu(\mathcal{K}(j+s),i)} \leq \alpha_{\nu(\mathcal{K}(j),i)}$ and, moreover, by Assumptions 2.3(ii) and 2.4(i), for sufficiently large j, $\alpha_{\nu(\mathcal{K}(j),i)} \leq c\,\alpha_{\mathcal{K}(j)}$ for some constant c independent of j. It then follows that for sufficiently large j, for all $i \in \mathcal{I}$ and $s \in [0,1]$,

$$\left| F_i^j(s) - G^j(s) \right| \leq c\,\alpha_{\mathcal{K}(j)} + \left| \underbrace{\int_{\nu(\mathcal{K}(j),i)}^{\nu(\mathcal{K}(j+s),i)} g_\alpha(y)dy - \int_{\mathcal{K}(j)}^{\mathcal{K}(j+s)} g_\alpha(y)dy}_{e_i(j,s)} \right|, \tag{4.25}$$

where c is a constant independent of j.

To establish the desired inequality (4.21), we now bound the r.h.s. of (4.25) separately for each class of step sizes considered.

Case 1: $g_\alpha(y) = \frac{1}{Ay}$. We have $e_i(j,s) = \frac{1}{A}\ln(\frac{\nu(\mathcal{K}(j+s),i)}{\mathcal{K}(j+s)}) - \frac{1}{A}\ln(\frac{\nu(\mathcal{K}(j),i)}{\mathcal{K}(j)})$. Under Assumption 2.6, a direct calculation shows that for sufficiently large j, $|e_i(j,s)| \leq c\mathcal{K}(j)^{-\gamma}$ for some (path-dependent) constant c independent of j and s. We also have $\alpha_{\mathcal{K}(j)} = \frac{1}{A\mathcal{K}(j)}$ and $\limsup_{j\to\infty}\frac{1}{j}\ln(\alpha_{\mathcal{K}(j)}) \leq \frac{\ell(\{\alpha_n\})}{d} = \frac{-A}{d}$ by Lemma 4.1(ii). Combining these relations with (4.24) and (4.25) yields that

$$\limsup_{j\to\infty}\frac{1}{j}\ln\left(\sup_{s\in[0,1]}\left|F_i^j(s) - \frac{s}{d}\right|\right) \leq \frac{-(\gamma\wedge 1)A}{d} \qquad \forall i\in\mathcal{I},$$

proving inequality (4.21) with $\ell = -(\gamma\wedge 1)A$.

Case 2: $g_\alpha(y) = \frac{1}{Ay\ln y}$. Then $e_i(j,s) = \frac{1}{A}\ln(\frac{\ln(\nu(\mathcal{K}(j+s),i))}{\ln(\mathcal{K}(j+s))}) - \frac{1}{A}\ln(\frac{\ln(\nu(\mathcal{K}(j),i))}{\ln(\mathcal{K}(j))})$. Expressing the two numerators as $\ln(\frac{\nu(\mathcal{K}(j+s),i)}{\mathcal{K}(j+s)}) + \ln(\mathcal{K}(j+s))$ and $\ln(\frac{\nu(\mathcal{K}(j),i)}{\mathcal{K}(j)}) + \ln(\mathcal{K}(j))$, respectively, and using Assumption 2.4(i), a direct calculation shows that for sufficiently large j, we have $|e_i(j,s)| \leq \frac{c}{\ln(\mathcal{K}(j))}$ for some (path-dependent) constant c independent of j and s. We also have $\lim_{j\to\infty}\frac{d\ln\ln(\mathcal{K}(j))}{Aj} = 1$ by Lemma 4.1(i) and the definition of step sizes in this case, while $\limsup_{j\to\infty}\frac{1}{j}\ln(\alpha_{\mathcal{K}(j)}) \leq \frac{\ell(\{\alpha_n\})}{d} = -\infty$ by Lemma 4.1(ii). Combining these relations with (4.24) and (4.25) yields that

$$\limsup_{j\to\infty}\frac{1}{j}\ln\left(\sup_{s\in[0,1]}\left|F_i^j(s) - \frac{s}{d}\right|\right) \leq \lim_{j\to\infty}\frac{-\ln\ln\left(\mathcal{K}(j)\right)}{\frac{d}{A}\ln\ln\left(\mathcal{K}(j)\right)} = \frac{-A}{d} \qquad \forall i\in\mathcal{I},$$

establishing the desired inequality (4.21) with $\ell = -A$. $\square$

By Lemmas 4.4 and 4.5, we conclude that inequality (4.7) holds a.s., namely,

$$\limsup_{j\to\infty}\frac{1}{j}\ln\left(\|\tilde{x}^j(j+1) - x^j(j+1)\|\right) < -L_h/d \quad a.s.,$$

if $(\gamma\wedge 1)A > L_h$ when class-1 step sizes are used and $A > L_h$ when class-2 step sizes are used. As these are the conditions required in Theorem 2.3, we obtain (4.7), thereby completing the proof of Theorem 2.3.

5. Discussion. In this paper, we established the stability and convergence of a class of asynchronous SA algorithms under more general noise conditions than previously considered. Our stability analysis extends a method of Borkar and Meyn, resolving open questions about the stability of such algorithms in average-reward RL. To sharpen the convergence analysis, we developed new characterizations of the shadowing properties of asynchronous SA by building on a dynamical systems approach of Hirsch and Benaïm. These results provide a theoretical foundation for the analysis of an important class of RL algorithms based on RVI (see [29, 33] for related developments).

Although our focus has been on partially asynchronous update schemes relevant to average-reward RL, certain techniques—particularly the construction of auxiliary scaled processes using stopping arguments—may prove useful for broader classes of asynchronous schemes, including those discussed in [9], given suitable choices of the function h. Future work may explore these possibilities, as well as potential extensions to distributed computation frameworks with communication delays.

Acknowledgments. We thank Prof. Eugene Feinberg for helpful discussion on average-reward SMDPs and Dr. Martha Steenstrup for critical feedback on parts of our earlier draft. We also thank an anonymous reviewer for valuable comments, particularly suggesting the potential for sharper convergence results, which motivated our study of shadowing properties and led us to strengthen the analysis. In preparing this paper, we used OpenAI ChatGPT (GPT-4 and GPT-5) to refine the writing style.

REFERENCES

[1] J. Abounadi, D. P. Bertsekas, and V. S. Borkar, *Learning algorithms for Markov decision processes with average cost*, SIAM J. Control Optim., 40 (2001), pp. 681–698, https://doi.org/10.1137/S0363012999361974.

[2] J. Abounadi, D. P. Bertsekas, and V. S. Borkar, *Stochastic approximation for nonexpansive maps: Applications to Q-learning algorithms*, SIAM J. Control Optim., 41 (2002), pp. 1–22, https://doi.org/10.1137/S0363012998346621.

[3] M. Benaim, *A dynamical systems approach to stochastic approximations*, SIAM J. Control Optim., 34 (1996), pp. 437–472, https://doi.org/10.1137/S0363012993253534.

[4] M. Benaïm, *Dynamics of stochastic approximation algorithms*, in Séminaire de Probabilities, Lecture Notes in Mathematics 1769, Springer, New York, 1999, pp. 1–69.

[5] M. Benaïm and M. W. Hirsch, *Asymptotic pseudotrajectories and chain recurrent flows, with applications*, J. Dynam. Differential Equations, 8 (1996), pp. 141–176, https://doi.org/10.1007/BF02218617.

[6] D. P. Bertsekas and J. N. Tsitsiklis, *Neural-Dynamic Programming*, Athena Scientific, Belmont, MA, 1996.

[7] N. P. Bhatia and G. P. Szegö, *Stability Theory of Dynamical Systems*, Springer, Berlin, 2002.

[8] S. Bhatnagar, *The Borkar–Meyn theorem for asynchronous stochastic approximations*, Systems Control Lett., 60 (2011), pp. 472–478, https://doi.org/10.1016/j.sysconle.2011.04.002.

[9] V. S. Borkar, *Asynchronous stochastic approximations*, SIAM J. Control Optim., 36 (1998), pp. 840–851, https://doi.org/10.1137/S0363012995282784.

[10] V. S. Borkar, *Erratum: Asynchronous stochastic approximations*, SIAM J. Control Optim., 38 (2000), pp. 662–663, https://doi.org/10.1137/S0363012998345913.

[11] V. S. Borkar, *Stochastic Approximations: A Dynamical Systems Viewpoint*, 2nd ed., Springer, Singapore, 2023.

[12] V. S. Borkar and S. Meyn, *The O.D.E. method for convergence of stochastic approximation and reinforcement learning*, SIAM J. Control Optim., 38 (2000), pp. 447–469, https://doi.org/10.1137/S0363012997331639.

[13] V. S. Borkar and K. Soumyanath, *A new analog parallel scheme for fixed point computation, Part* I*: Theory*, IEEE Trans. Circuits Systems. I. Fund. Theory Appl., 44 (1997), pp. 351–355, https://doi.org/10.1109/81.563625.

[14] R. Bowen, *ω-limit sets of Axiom A diffeomorphisms*, J. Differential Equations, 18 (1975), pp. 333–339, https://doi.org/10.1016/0022-0396(75)90065-0.

[15] C. C. Conley, *Isolated Invariant Sets and the Morse Index*, CBMS Regional Conference Series in Mathematics 38, AMS, Providence, RI, 1978.

[16] P. Hall and C. C. Heyde, *Martingale Limit Theory and Its Application*, Academic Press, New York, 1980.

[17] M. W. Hirsch, *Asymptotic phase, shadowing and reaction-diffusion systems*, in Differential Equations, Dynamical Systems and Control Science, Lecture Notes in Pure Appl. Math. 152, Marcel Dekker, New York, 1994, pp. 87–99.

[18] H. J. Kushner and G. G. Yin, *Stochastic Approximation and Recursive Algorithms and Applications*, 2nd ed., Springer, New York, 2003.

[19] J. Neveu, *Discrete Parameter Martingales*, North-Holland, Amsterdam, 1975.

[20] M. L. Puterman, *Markov Decision Processes: Discrete Stochastic Dynamic Programming*, .Wiley., New York, 1994.

[21] A. Ramaswamy, S. Bhatnagar, and D. E. Quevedo, *Asynchronous stochastic approximations with asymptotically biased errors and deep multiagent learning*, IEEE Trans. Automat. Control, 66 (2021), pp. 3969–3983, https://doi.org/10.1109/TAC.2020.3026269.

[22] P. J. Schweitzer, *Iterative solution of the functional equations of undiscounted Markov renewal programming*, J. Math. Anal. Appl., 34 (1971), pp. 495–501, https://doi.org/10.1016/0022-247X(71)90094-1.

[23] P. J. Schweitzer and A. Federgruen, *The asymptotic behavior of undiscounted value iteration in Markov decision problems*, Math. Oper. Res., 2 (1977), pp. 360–381, https://doi.org/10.1287/moor.2.4.360.

[24] P. J. Schweitzer and A. Federgruen, *The functional equations of undiscounted Markov renewal programming*, Math. Oper. Res., 3 (1978), pp. 308–321, https://doi.org/10.1287/moor.3.4.308.

[25] J. Tsitsiklis, *Asynchronous stochastic approximation and Q-learning*, Mach. Learn., 16 (1994), pp. 195–202, https://doi.org/10.1023/A:1022689125041.

[26] W. Walter, *Ordinary Differential Equations*, Springer, New York, 1998.

[27] Y. Wan, A. Naik, and R. S. Sutton, *Average-reward learning and planning with options*, in Neural Inform. Proc. Syst., Curran Associates, Red Hook, NY, 2021, pp. 22758–22769.

[28] Y. Wan, A. Naik, and R. S. Sutton, *Learning and planning in average-reward Markov decision processes*, Proc. Mach. Learn. Res. (PMLR), 139 (2021), pp. 10653–10662.

[29] Y. Wan, H. Yu, and R. S. Sutton, *On Convergence of Average-Reward Q-learning in Weakly Communicating Markov Decision Processes*, preprint, arXiv:2408.16262, 2024.

[30] D. J. White, *Dynamic programming, Markov chains, and the method of successive approximations*, J. Math. Anal. Appl., 6 (1963), pp. 373–376, https://doi.org/10.1016/0022-247X(63)90017-9.

[31] H. Yu and D. P. Bertsekas, *On boundedness of Q-learning iterates for stochastic shortest path problems*, Math. Oper. Res., 38 (2013), pp. 209–227, https://doi.org/10.1287/moor.1120.0562.

[32] H. Yu, Y. Wan, and R. S. Sutton, *Asynchronous Stochastic Approximation with Applications to Average-Reward Reinforcement Learning*, preprint, arXiv:2409.03915, 2025, (extended report version of this paper).

[33] H. Yu, Y. Wan, and R. S. Sutton, *Average-Reward Reinforcement Learning in Semi-Markov Decision Processes via Relative Value Iteration*, preprint, arXiv:2512.06218, 2025.