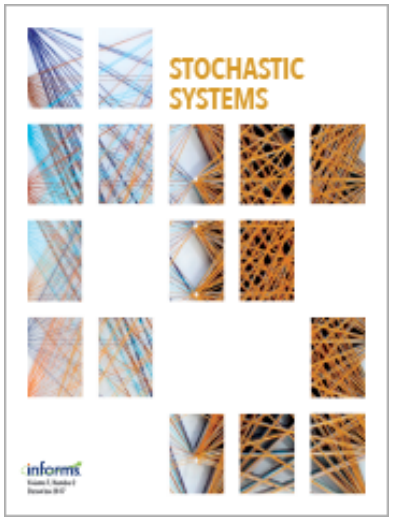
This article was downloaded by: [216.73.217.172] On: 15 November 2026, At: 06:26
Publisher: Institute for Operations Research and the Management Sciences (INFORMS)
INFORMS is located in Maryland, USA



Stochastic Systems

Publication details, including instructions for authors and subscription information:
http://pubsonline.informs.org

Asynchronous Stochastic Approximation with Differential Inclusions

Steven Perkins, David S. Leslie

To cite this article:
Steven Perkins, David S. Leslie (2012) Asynchronous Stochastic Approximation with Differential Inclusions. Stochastic Systems 2(2):409-446. https://doi.org/10.1287/11-SSY056

This work is licensed under a Creative Commons Attribution 4.0 International License. You are free to copy, distribute, transmit and adapt this work, but you must attribute this work as "*Stochastic Systems*. Copyright 2012 The Author(s). https://doi.org/10.1287/11-SSY056, used under a Creative Commons Attribution License: http://creativecommons.org/licenses/by/4.0/."

Copyright © 2012, The author(s)

Please scroll down for article—it is on subsequent pages

With 12,500 members from nearly 90 countries, INFORMS is the largest international association of operations research (O.R.) and analytics professionals and students. INFORMS provides unique networking and learning opportunities for individual professionals, and organizations of all types and sizes, to better understand and use O.R. and analytics tools and methods to transform strategic visions and achieve better outcomes.
For more information on INFORMS, its publications, membership, or meetings visit http://www.informs.org

Stochastic Systems
2012, Vol. 2, No. 2, 409–446
DOI: 10.1214/11-SSY056

Downloaded from informs.org by [216.73.217.172] on 15 November 2026, at 06:26 . For personal use only, all rights reserved.

ASYNCHRONOUS STOCHASTIC APPROXIMATION WITH DIFFERENTIAL INCLUSIONS

By Steven Perkins and David S. Leslie

University of Bristol

The asymptotic pseudo-trajectory approach to stochastic approximation of Benaïm, Hofbauer and Sorin is extended for asynchronous stochastic approximations with a set-valued mean field. The asynchronicity of the process is incorporated into the mean field to produce convergence results which remain similar to those of an equivalent synchronous process. In addition, this allows many of the restrictive assumptions previously associated with asynchronous stochastic approximation to be removed. The framework is extended for a coupled asynchronous stochastic approximation process with set-valued mean fields. Two-timescales arguments are used here in a similar manner to the original work in this area by Borkar. The applicability of this approach is demonstrated through learning in a Markov decision process.

1. Introduction. Many learning algorithms include a stochastic updating schedule, often based on a Markov chain. Studying the performance of these processes can be carried out using the asynchronous stochastic approximation framework. However, the previous work in this area has focused on continuous, single-valued updates as discussed in the literature (see for example [9, 14, 17, 18, 24]). Furthermore some of the assumptions which are typically used are challenging to verify. In this work we expand the asymptotic pseudo-trajectory approach of Benaïm, Hofbauer and Sorin [5] to asynchronous stochastic approximations with set-valued mean fields. We incorporate the asynchronicity into the mean field to give a differential inclusion which will characterise the limiting behaviour of the associated learning process.

Consider an iterative process $\{x_n\}_{n\in\mathbb{N}}$ where $x_n \in \mathbb{R}^K$ and denote the i^{th} component of x_n as $x_n(i)$ where $i \in I$ and $K = |I|$ is finite. A typical stochastic approximation (SA) is of the form

$$x_{n+1} = x_n + \alpha(n+1)\left[F(x_n) + V_{n+1} + d_{n+1}\right], \tag{1.1}$$

Received December 2011.

Keywords and phrases: Asynchronous stochastic approximation, differential inclusion, two-timescales.

where $\{\alpha(n)\}_{n\in\mathbb{N}}$ is a positive, decreasing sequence, $\{V_n\}_{n\in\mathbb{N}}$ is a zero-mean martingale difference process, $\{d_n\}_{n\in\mathbb{N}}$ is a bounded sequence which converges to zero and $F(\cdot) : \mathbb{R}^K \to \mathbb{R}^K$ is a Lipschitz continuous mean field. Standard arguments (e.g. [4]) are then used to show that the limiting behaviour of the iterative process in (1.1) can be studied through the ordinary differential equation (ODE)

$$\frac{dx}{dt} = F(x). \tag{1.2}$$

Commonly known as the ODE method of stochastic approximation, originally proposed by Ljung [20], this technique has been extended by numerous authors, for example Benaïm [3], Benaïm, Hofbauer and Sorin [5], Borkar [11], Kushner and Clark [16] and Kushner and Yin [17, 18]. In particular Benaïm, Hofbauer and Sorin [5] have developed the approach so that under some weak criteria $\{x_n\}_{n\in\mathbb{N}}$ can be updated via a set-valued mean field, $F(\cdot)$. This allows for the limiting behaviour to be studied using the associated differential inclusion.

Standard stochastic approximations are not always applicable; an example which we examine in this paper is when learning action values in a Markov decision process (MDP) and this is also discussed by Konda and Tsitsiklis [15], Tsitsiklis [24] and Singh et al. [22]. In a MDP updates are made to a single random component at each iteration. Therefore we have a stochastic, asynchronous updating pattern, where a subset of an iterative process similar to (1.1) can be updated many times before the remaining components are selected for a single update. Based on this idea extensions to the standard theory have been examined such as those by Kushner and Yin [17, 18]. Here however we follow the extension to asynchronous stochastic approximation provided by Borkar [9] and Konda and Borkar [14]. They show that when the iterative updates have a Lipschitz continuous mean field then, similarly to a standard stochastic approximation, the limiting behaviour can be studied via the associated differential equation,

$$\frac{dx}{dt} = M(t)F(x), \tag{1.3}$$

where $M(t)$ is a $K \times K$ diagonal matrix and the diagonal elements of $M(t)$ lie in the set $[0, 1]$ for all $t > 0$. This early work on asynchronous stochastic approximations has certain restrictions which limit its usability. In particular, many of the assumptions made in the work of Borkar [9] and Konda and Borkar [14] are given in implicit form and are difficult to verify in specific situations.

Downloaded from informs.org by [216.73.217.172] on 15 November 2026, at 06:26 . For personal use only, all rights reserved.

Downloaded from informs.org by [216.73.217.172] on 15 November 2026, at 06:26 . For personal use only, all rights reserved.

As with the initial results for a standard stochastic approximation the results of Borkar [9] are limited to the case when the mean field, $F(\cdot)$, is a Lipschitz continuous function. The subsequent work by Benaïm, Hofbauer and Sorin [5] on set-valued mean fields leaves the natural question of whether similar results are possible for asynchronous stochastic approximations when a set-valued mean field is used. In addition, the ODE in (1.3) is non-autonomous and the scaling matrix $M(\cdot)$ is not explicitly defined. This makes analysis of the limiting behaviour more difficult to study, although some methods for verifying global convergence are outlined by Borkar [11].

Borkar [8] originally extended the stochastic approximation framework to two-timescales. Since then Leslie and Collins [19] extended this idea to multiple timescales and Konda and Borkar [14] provide a first venture into the two-timescale asynchronous stochastic approximation. However, all of these only consider stochastic approximations when the mean field is Lipschitz continuous.

The aim of this work is to combine and generalise the results by Borkar [8, 9], Konda and Borkar [14] and Benaïm, Hofbauer and Sorin [5] to create a framework for single and two-timescale asynchronous stochastic approximations which is straightforward to use in practical applications. In this paper we show that, under a set of verifiable assumptions, the diagonal elements of $M(t)$ lie in the closed set $[\varepsilon, 1]$, for some $\varepsilon > 0$. The set $[\varepsilon, 1]$ can be combined with the mean field $F(\cdot)$ to form a set-valued mean field, $\bar{F}(\cdot)$. Under some straightforward assumptions a stochastic approximation recursion with mean field $\bar{F}(\cdot)$ can have it's limiting behaviour studied via the associated differential inclusion using the results of Benaïm, Hofbauer and Sorin [5]. A natural benefit of using the differential inclusion framework is that $F(\cdot)$ can be set-valued as this does not alter the analysis.

This paper is organised in the following manner: Section 2 reviews some previous results on stochastic approximation with differential inclusions and asynchronous stochastic approximations. In Section 3 we focus on the single-timescale asynchronous stochastic approximation. We state the main theorem before presenting the weak convergence results required for the proof. Section 4 examines the extension to a two-timescale asynchronous stochastic approximation process. Large parts of this section follow directly from the results in Section 3. In Section 5 we present an example of a learning algorithm for discounted reward Markov decision processes and obtain convergence results by applying the method shown in Section 4. This illustrates the ease in which this framework can be used. Finally, the paper concludes with a summary of the work. Throughout this paper many of the proofs are omitted from the main flow of text and are instead presented in an appendix.

Downloaded from informs.org by [216.73.217.172] on 15 November 2026, at 06:26 . For personal use only, all rights reserved.

2. Background. Throughout this paper we use two main ideas from the stochastic approximation literature. The first relates to the work by Benaïm, Hofbauer and Sorin [5] on stochastic approximation with differential inclusions, and the second concerns the asynchronous stochastic approximation framework introduced by Borkar [9]. We take the opportunity to review the pertinent features of their work in this section.

In what is to follow we use the standard concept of set multiplication: if A is a set of $K \times K$ matrices and $B \subset \mathbb{R}^K$

$$A \cdot B = \left\{ a \cdot b;\, a \in A, b \in B \right\} \subset \mathbb{R}^K.$$

Note that if A and B are closed, convex sets then $A \cdot B$ is also closed and convex. This definition is still used if either or both of the sets A and B are single valued, however, in this case we often drop the '$\cdot$' notation for convenience.

2.1. *Stochastic approximation with differential inclusions.* We begin by outlining the current convergence results for stochastic approximations with set-valued maps proved by Benaïm, Hofbauer and Sorin [5]. These results are heavily used in Section 3, most notably to prove our main result. Firstly, it is important to consider when a solution to a differential inclusion will exist and so we take criteria from the original work on stochastic approximations with differential inclusions by Benaïm et al. [5], although they are more widely used in the differential inclusions literature.

DEFINITION 2.1. Call $F(\cdot) : X \to Y$ a *Marchaud map*[1] if it satisfies the following;

(i) $F(\cdot)$ is a closed set-valued map. That is,

$$\text{Graph}(F) = \{(x, y); y \in F(x)\},$$

is a closed set. Equivalently, $F(\cdot)$ is an upper semi-continuous set-valued map.

(ii) For all $x \in X$, $F(x)$ is a non-empty, compact, convex subset of Y.

(iii) There exists a $c > 0$ such that for all $x \in X$,

$$\sup_{z \in F(x)} \|z\| \leq c\,(1 + \|x\|)\,.$$

[1]The term "Marchaud map" is not used by Benaïm et al. [5], but this is a term used to describe the conditions in Definition 2.1 (see for example [1, pg. 62]).

Downloaded from informs.org by [216.73.217.172] on 15 November 2026, at 06:26 . For personal use only, all rights reserved.

Take $F(\cdot) : \mathbb{R}^K \to \mathbb{R}^K$ as a Marchaud map; a typical differential inclusion is in the form,

$$\frac{dx}{dt} \in F\big(x\big), \tag{2.1}$$

and a *solution* to (2.1) is an absolutely continuous mapping $x : \mathbb{R} \to \mathbb{R}^K$ such that $x(0) = x$ and for almost every $t > 0$,

$$\frac{dx(t)}{dt} \in F\big(x(t)\big).$$

The *flow* induced by (2.1) is defined by,

$$\Phi_t(x) = \{x(t); x(\cdot) \text{ is a solution to (2.1) with } x(0) = x\},$$

DEFINITION 2.2 (Benaïm et al. [5]). A continuous function $y : \mathbb{R}^+ \to \mathbb{R}^K$ is an *asymptotic pseudo-trajectory* to Φ if

$$\lim_{t\to\infty} \sup_{s\in[0,T]} d\Big(y\big(t+s\big), \Phi_s\big(y(t)\big)\Big) = 0,$$

for any $T > 0$ and where $d(\cdot,\cdot)$ is a distance measure on $\mathbb{R}^K$.

Properties of flows induced by differential inclusions, and their asymptotic pseudo-trajectories are discussed in detail by Benaïm, Hofbauer and Sorin [5]. Most important for the work here is that an asymptotic pseudo-trajectory to (2.1) will behave in a similar manner to the solutions of the differential inclusion and hence the limiting behaviour will be closely related.

We conclude this section by considering a standard iterative process in the form of (1.1) where the mean field $F(\cdot) : \mathbb{R}^K \to \mathbb{R}^K$ is a Marchaud map. The following theorem states that under four assumptions a linear interpolation of $\{x_n\}_{n\in\mathbb{N}}$ (a function defined precisely in Section 3.1) is an asymptotic pseudo-trajectory to the differential inclusion (2.1). Hence the limiting behaviour of $\{x_n\}_{n\in\mathbb{N}}$ can be studied via this differential inclusion.

THEOREM 2.3. *Consider the stochastic approximation process*

$$x_{n+1} \in x_n + \alpha(n+1)\left[F(x_n) + V_{n+1} + d_{n+1}\right], \tag{2.2}$$

under the assumptions;

(i) For all $T > 0$

$$\lim_{n\to\infty} \sup_k \left\{ \left\| \sum_{i=n}^{k-1} \alpha(i+1) V_{i+1} \right\| ; k = n+1, \ldots, m(\tau_n + T) \right\} = 0 \tag{2.3}$$

where $\tau_0 = 0$, $\tau_n = \sum_{i=1}^{n} \alpha(i)$ *and* $m(t) = \sup\{k \geq 0; t \geq \tau_k\}$,

(ii) $\sup_n \|x_n\| = \mathcal{X} < \infty$,
(iii) $F(\cdot) : \mathbb{R}^K \to \mathbb{R}^K$ *is a Marchaud map,*
(iv) $d_n \to 0$ *as* $n \to \infty$ *and* $\sup_n \|d_n\| = d < \infty$.

Then a linear interpolation of the iterative process $\{x_n\}_{n\in\mathbb{N}}$ *given by* (2.2) *is an asymptotic pseudo-trajectory of the differential inclusion* (2.1).

This is a slight modification to a result stated by Benaïm, Hofbauer and Sorin [5, Proposition 1.3] to include the $\{d_n\}_{n\in\mathbb{N}}$ terms. It is trivial to verify that this will not alter any of the asymptotic results of the original work.

2.2. *Asynchronous stochastic approximations.* Now we fully introduce the asynchronous stochastic approximation notation used. A typical asynchronous stochastic approximation such as those studied by Borkar [9] fits the following framework. Let I be the set of all components of the iterative process. For example if $x_n \in \mathbb{R}^K$ then $I = \{1, \ldots, K\}$. If 2^I is the power set of all possible updating combinations in I then let $\bar{I}_n \in 2^I$ be the components of the iterative process $\{x_n\}_{n\in\mathbb{N}}$ updated at iteration n. Using a counter for each $i \in I$,

$$\nu_n(i) := \sum_{k=1}^{n} \mathbb{I}_{\{i\in\bar{I}_k\}},$$

we consider processes in which no component, i, in the asynchronous process needs to know the global counter, n, merely its own counter, $\nu_n(i)$. Let $F_i(x) := \{y_i\,;\, y \in F(x)\}$ and $x_n(i)$, $V_n(i)$ and $d_n(i)$ be the i^{th} component of x_n, V_n and d_n respectively, for $i = 1, \ldots, K$. We directly extend the notation used in (2.2) for an asynchronous stochastic approximation; let $F(\cdot)$ be a Marchaud map, then for $i = 1, \ldots, K$ let

$$x_{n+1}(i) \in x_n(i) + \alpha\big(\nu_{n+1}(i)\big)\mathbb{I}_{\{i\in\bar{I}_{n+1}\}}\big[F_i(x_n) + V_{n+1}(i) + d_{n+1}(i)\big]. \tag{2.4}$$

Define the *asynchronous step sizes*, $\bar{\alpha}_n$, and the *relative step sizes*, $\mu_n(i)$, to be

$$\bar{\alpha}_n := \max_{i\in\bar{I}_n} \alpha\big(\nu_n(i)\big), \quad \mu_n(i) := \frac{\alpha\big(\nu_n(i)\big)}{\bar{\alpha}_n}\mathbb{I}_{\{i\in\bar{I}_n\}}.$$

The asynchronous step sizes, $\bar{\alpha}_n$, are random step sizes (in contrast with the deterministic $\alpha(n)$ terms) whilst the relative step sizes, $\mu_n(i)$, are zero whenever the i^{th} component of the iterative process is not updated. Clearly $\mu_n(i) \in [0, 1]$. By letting M_n be the $K \times K$ diagonal matrix of the $\mu_n(i)$ terms we can express the previous asynchronous stochastic approximation (2.4) in the more concise form

$$x_{n+1} - x_n - \bar{\alpha}_{n+1}M_{n+1}\left[V_{n+1} + d_{n+1}\right] \in \bar{\alpha}_{n+1}M_{n+1} \cdot F(x_n). \tag{2.5}$$

Downloaded from informs.org by [216.73.217.172] on 15 November 2026, at 06:26 . For personal use only, all rights reserved.

Downloaded from informs.org by [216.73.217.172] on 15 November 2026, at 06:26 . For personal use only, all rights reserved.

This is a more familiar form for a stochastic approximation with a set-valued mean field. (2.5) differs from (2.2) only because the step sizes in (2.5) are random and the addition of the M_{n+1} coefficient. Instead of treating M_{n+1} as a coefficient of the step sizes we combine it with the mean field. Convergence of the error term, d_{n+1}, will be unaffected and, under a set of assumptions in Section 3.2, the noise term $M_{n+1}V_{n+1}$ will still satisfy the Kushner and Clark noise condition in (2.3). Combining the M_{n+1} term and the mean field term $F(\cdot)$ into a single set provides an intuitive method of rephrasing the stochastic approximation and leads to a set-valued mean field.

Proceeding with this intuition is not immediately straightforward since M_n is time varying and can be zero infinitely often, in which case the mean field could be zero even when the original update term, $F(\cdot)$, is not. This would mean the limiting behaviour of the differential inclusion in the asynchronous stochastic approximation could be different to the synchronous case, where ultimately we wish to say that the two behave in the same manner in the limit. To avoid this scenario we follow Borkar [9] and consider the weak limit of the interpolations of $\{M_n\}_{n\in\mathbb{N}}$, which will always be bounded away from zero under some verifiable assumptions, given in Section 3.2.

3. Asynchronous SA with differential inclusions. We begin by presenting the main result of this paper which concerns the limiting behaviour of the asynchronous stochastic approximation in (2.5), before outlining the results required to prove this in the remainder of the section.

3.1. *Main result.* Assume that $F(\cdot) : \mathbb{R}^K \to \mathbb{R}^K$ is a Marchaud map and for all n define $f_n \in F(x_n)$ by its component parts, $f_n(i)$, $i = 1, \ldots, K$, such that

$$f_n(i)\mu_{n+1}(i) := \frac{x_{n+1}(i) - x_n(i)}{\bar{\alpha}_{n+1}} - \mu_{n+1}(i)\left[V_{n+1}(i) + d_{n+1}(i)\right]. \tag{3.1}$$

Notice that if $\mu_{n+1}(i) = 0$ then we can select any $f_n(i) \in F_i(x_n)$. Then we can write the iterative process in (2.5) as

$$x_{n+1} = x_n + \bar{\alpha}_{n+1}M_{n+1}\Big[f_n + V_{n+1} + d_{n+1}\Big]. \tag{3.2}$$

For some fixed $\varepsilon > 0$, let $\tilde{M}_n$ be a series of $K \times K$ diagonal matrices with entries in the set $[\varepsilon, 1]$, for all n, to be defined in Section 3.3. We can again rewrite the iterative process in (3.2) as

$$x_{n+1} = x_n + \bar{\alpha}_{n+1}\Big[\tilde{M}_{n+1}f_n + \big(M_{n+1} - \tilde{M}_{n+1}\big)f_n + M_{n+1}V_{n+1} + M_{n+1}d_{n+1}\Big].$$

Now by letting $\bar{V}_{n+1} = f_n\big(M_{n+1} - \tilde{M}_{n+1}\big) + M_{n+1}V_{n+1}$ and $\bar{d}_{n+1} = M_{n+1}d_{n+1}$ we get,

$$x_{n+1} = x_n + \bar{\alpha}_{n+1}\Big[\tilde{M}_{n+1}f_n + \bar{V}_{n+1} + \bar{d}_{n+1}\Big]. \tag{3.3}$$

For general $k, \delta > 0$ let

$$\Omega_k^\delta := \big\{\mathrm{diag}(\omega_1, ..., \omega_k); \omega_i \in [\delta, 1], \forall i = 1, ..., k\big\}, \tag{3.4}$$

and define

$$\bar{F}(x) := \Omega_K^\varepsilon \cdot F(x). \tag{3.5}$$

If $F(\cdot)$ is Lipschitz continuous direct comparisons can be made between the mean field, $\bar{F}(x)$, and the analogous mean field $M(t)F(x)$ from (1.3) which is used by Borkar [9] and Konda and Borkar [14]. This provides the key insight into the new approach we take. Under the assumptions used in Section 3.2 the equivalent $M(t)$ values almost surely lie in Ω_K^ε. By combining this with $F(x)$ we produce a differential inclusion which is more straightforward to study than a non-autonomous differential equation and naturally fits the stochastic approximation framework of Benaïm, Hofbauer and Sorin [5]. In addition, this idea naturally lends itself to examining a similar process for a set-valued mean field as we proceed to do in this paper.

Equation (3.3) can be expressed in the form of a stochastic approximation with a set-valued mean field as in [5]:

$$x_{n+1} - x_n - \bar{\alpha}_{n+1}\bar{V}_{n+1} - \bar{\alpha}_{n+1}\bar{d}_{n+1} \in \bar{\alpha}_{n+1}\bar{F}(x_n). \tag{3.6}$$

Let $\bar{\tau}_0 := 0$, $\bar{\tau}_n := \sum_{k=1}^n \bar{\alpha}_k$ be the timescale for the asynchronous updates. To allow this process to be analysed in continuous time consider an interpolated version of the stochastic approximation (3.6) so that this process can be considered in continuous time,

$$\bar{x}(\bar{\tau}_n + s) = x_n + s\frac{x_{n+1} - x_n}{\bar{\alpha}_{n+1}}, \quad s \in [0, \bar{\alpha}_{n+1}). \tag{3.7}$$

Under the assumptions (A1)-(A5), presented in Section 3.2, we show in Section 3.3 that a sequence, $\{\tilde{M}_n\}_{n\in\mathbb{N}}$, can be defined such that $\{\bar{\alpha}_n\}_{n\in\mathbb{N}}$ and $\{\bar{V}_n\}_{n\in\mathbb{N}}$ satisfy the Kushner and Clark noise condition in (2.3). By invoking Theorem 2.3 we obtain our main result, which is proved in Section 3.4.

THEOREM 3.1. *Under the assumptions (A1)-(A5), with probability 1, $\bar{x}(t)$ is an asymptotic pseudo-trajectory to the differential inclusion,*

$$\frac{dx}{dt} \in \bar{F}(x), \tag{3.8}$$

where $\bar{F}(\cdot)$ is as described in (3.5).

Downloaded from informs.org by [216.73.217.172] on 15 November 2026, at 06:26 . For personal use only, all rights reserved.

Downloaded from informs.org by [216.73.217.172] on 15 November 2026, at 06:26 . For personal use only, all rights reserved.

In Theorem 3.1 we do not explicitly explain how a suitable $\varepsilon > 0$ is chosen to create $\bar{F}(\cdot)$. We note that $\varepsilon = A_\eta^{-1}\eta$ is sufficient, where A_η is as defined in (A2)(b) and η is taken from Lemma A.1. The reason this choice is appropriate can be seen in the proof of Lemma 3.4, although often the exact value of $\varepsilon > 0$ doesn't need to be known, as is the case in Corollary 3.2 below.

Directly from Theorem 3.1 and [5, Proposition 3.27] we get the key result concerning the convergence of an asynchronous stochastic approximation process.

COROLLARY 3.2. *If for all $\varepsilon > 0$ there is a global attractor, $A \subset C$, for the differential inclusion (3.8), and assumptions (A1)-(A5) are satisfied, then the iterative process (2.5) will almost surely converge to A.*

3.2. *Assumptions.* Throughout this section we study the convergence properties of the iterative process (2.5). We make reference to the following assumptions, (A1)-(A5), all of which are either standard requirements for a stochastic approximation or can be verified prior to running the asynchronous stochastic approximation process. This is in contrast with the previous work on asynchronous stochastic approximations by Borkar [9] and Konda and Borkar [14].

(A1) (a) For a compact set, $C \subset \mathbb{R}^K$, $x_n \in C$ for all n.

(b) $\{d_n\}_{n\in\mathbb{N}}$ is a bounded sequence such that $d_n \to 0$ as $n \to \infty$.

(A2) Let $\alpha(n)$ satisfy the following criteria;

(a) $\sum_n \alpha(n) = \infty$ and $\alpha(n) \to 0$ as $n \to \infty$.

(b) For $x \in (0,1)$, $\sup_n \alpha([xn])/\alpha(n) < A_x < \infty$, where $[\cdots]$, means the "integer part of". In addition, for all n, $\alpha(n) \geq \alpha(n+1)$.

(A3) $F(\cdot) : C \to C$ is a Marchaud map.

(A1)(a) is a slight strengthening of the standard stochastic approximation boundedness assumption; however this is still a relatively mild condition. Methods to ensure that it is satisfied are discussed elsewhere, for example [13, 14] or [24]. A basic restriction is placed on $\{d_n\}_{n\in\mathbb{N}}$ in (A1)(b); in this form the sequence does not affect the asymptotic behaviour of the process. (A2)(a) is a standard assumption required for stochastic approximation, and (A2)(b) is a mild technical condition required to deal with the asynchronicity, which is also used by Borkar [9]. We have dropped the additional restriction on the step-sizes used by Borkar which severely restricts the possible choices of $\{\alpha(n)\}_{n\in\mathbb{N}}$. (A3) ensures that we can use the convergence

Downloaded from informs.org by [216.73.217.172] on 15 November 2026, at 06:26 . For personal use only, all rights reserved.

results presented in Theorem 2.3 and is a standard assumption for stochastic approximations with a set-valued mean field.

Define $\bar{I} \subset 2^I$ as the set of all the possible combinations which have positive probability of occurring. As an example, if every element of I gets updated and it is known that $\bar{I}_n$ is a singleton for each n, then $\bar{I} = I$.

Let $\mathcal{F}_n$ be a sigma algebra containing all the information up to and including the n^{th} iteration. That is $\mathcal{F}_n := \sigma(\{\bar{I}_m\}_m, \{x_m\}_m, \{\nu_m(i)\}_{i,m}; \forall m \leq n, i = 1, \ldots, K)$.

(A4) (a) For all $x \in C$, $\mathcal{I}_n, \mathcal{I}_{n+1} \in \bar{I}$,

$$\mathbb{P}\Big(\bar{I}_{n+1} = \mathcal{I}_{n+1} | \mathcal{F}_n\Big) = \mathbb{P}\Big(\bar{I}_{n+1} = \mathcal{I}_{n+1} | \bar{I}_n = \mathcal{I}_n, x_n = x\Big).$$

Let

$$P_{(\mathcal{I}_n, \mathcal{I}_{n+1})}(x) := \mathbb{P}\Big(\bar{I}_{n+1} = \mathcal{I}_{n+1} | \bar{I}_n = \mathcal{I}_n, x_n = x\Big).$$

(b) For all $x \in C$ the transition probabilities $P_{(\mathcal{I}_n, \mathcal{I}_{n+1})}(x)$ form an aperiodic, irreducible Markov chain over $\bar{I}$ and for all $i \in I$ there exists an $\mathcal{I} \in \bar{I}$ such that $i \in \mathcal{I}$.

(c) The map $x \mapsto P_{(\mathcal{I}_n, \mathcal{I}_{n+1})}(x)$ is Lipschitz continuous.

(A4)(a) assumes that the transitions between the updated elements in $\bar{I}$ are part of a controlled Markov chain. (A4)(b) is a straightforward assumption on this controlled Markov chain which can be verified prior to implementation which allows us to negate the need for some of the original technical assumptions made by Konda and Borkar [14]. In this previous work Konda and Borkar assume that every state is updated at a comparable rate in the limit which cannot directly be verified prior to running the process. (A4)(c) is a condition which is required later to use a result from Ma et al. [21] on the convergence of stochastic approximation with Markovian Noise.

(A5) Assume that for all n, V_{n+1} and $\bar{I}_{n+1}$ are uncorrelated given $\mathcal{F}_n$ and either

(a) For some $q \geq 2$

$$\sum_n \alpha(n)^{1+q/2} < \infty, \quad \text{and} \quad \sup_n \mathbb{E}\big(\|V_n\|^q\big) < \infty;$$

or

(b) $V_n(i)$ is independent of $V_n(j)$ for $i \neq j$. Let $\langle a, b \rangle = \sum_k a_k b_k$. Then assume there exists a positive Γ such that for all $\theta \in \mathbb{R}^K$,

$$\mathbb{E}\Big[\exp\Big\{\langle \theta, V_{n+1} \rangle\Big\} | \mathcal{F}_n\Big] \leq \exp\Big\{\frac{\Gamma}{2}\|\theta\|^2\Big\};$$

and

$$\sum_n e^{-c/\alpha(n)} < \infty,$$

for each $c > 0$.

An assumption similar to (A5) is used by Benaïm, Hofbauer and Sorin [5] to verify a condition for noise convergence and is similar to that used by Kushner and Clark [16]. We say that (A5) holds if *either* (A5)(a) or (A5)(b) is true. This assumption is only used to show the noise term still satisfies the Kushner-Clark condition given in Lemma 3.3; the proof is presented via two lemmas in Appendix A.2.

LEMMA 3.3. *Assume that (A2)(b), (A4) and (A5) hold. Then with probability 1, for all $T > 0$,*

$$\lim_{n\to\infty} \sup_k \left\{ \left\| \sum_{i=n}^{k-1} \bar{\alpha}_{i+1} M_{i+1} V_{i+1} \right\|; k = n+1, \ldots, \bar{m}(\bar{\tau}_n + T) \right\} = 0$$

where $\bar{\tau}_0 := 0$, $\bar{\tau}_n := \sum_{k=1}^n \bar{\alpha}_k$ and $\bar{m}(t) := \sup\{k \geq 0; t \geq \bar{\tau}_k\}$.

Note that if Lemma 3.3 can be verified directly without (A5) then this assumption is redundant, and hence we only require (A1)-(A4) and Lemma 3.3 to hold. This approach is used in Section 4.

3.3. *Weak convergence of asynchronous updates.* As discussed in the introduction, the key issue with asynchronous stochastic approximations is how to handle the interaction of the relative step sizes and the mean field. It is important to be able to bound the limit of the relative step sizes, $\mu_n(i)$, away from zero for all $i \in I$ in order to produce an asynchronous stochastic approximation mean field, $\bar{F}(\cdot)$, which will behave similarly to the synchronous mean field, $F(\cdot)$. However the relative step size of a state is zero whenever that state is not updated, hence it is not immediately clear that this is even possible. Despite this it is sufficient that for any $T > 0$ an 'average' of $\mu_n(i)$ over length T in the continuous time interpolation converges to a value which is bounded away from zero. In this section we prove that under (A2)(b) and (A4) this is indeed the case.

For $T > 0$ the space $L^2([0,T])$ is the set of measurable functions $h(\cdot) : [0,T] \to \mathbb{R}$ such that,

$$\Big(\int_0^T |h(s)|^2 \, \mathrm{d}s \Big)^{\frac{1}{2}} < \infty.$$

Downloaded from informs.org by [216.73.217.172] on 15 November 2026, at 06:26 . For personal use only, all rights reserved.

Downloaded from informs.org by [216.73.217.172] on 15 November 2026, at 06:26 . For personal use only, all rights reserved.

Following the method used in [11] and [14], define $\mathcal{U}$ to be the space of maps $u(\cdot) : \mathbb{R} \to [0,1]$ with the coarsest topology which for all $T > 0$ leaves continuous the map,

$$u(\cdot) \to \int_0^T h(s)u(s)\,\mathrm{d}s$$

for all $h(\cdot) \in L^2([0,T])$. Hence $\mathcal{U}$ is a space of $[0,1]$ trajectories. This means that for any map defined on $\mathcal{U}$ convergence to a limit point will be in the weak sense, along a subsequence. That is, a sequence of maps $\{\varphi_n(\cdot)\}_{n\in\mathbb{N}}$ such that $\varphi_n(\cdot) \in \mathcal{U}$ for all n is said to possess a limit point $\tilde{\varphi}(\cdot) \in \mathcal{U}$ if for fixed $T > 0$ there exists a subsequence $k(n)$, such that for any $h(\cdot) \in L^2([0,T])$,

$$\int_0^T h(s)\varphi_{k(n)}(s)\,\mathrm{d}s \to \int_0^T h(s)\tilde{\varphi}(s)\,\mathrm{d}s, \quad \text{as } n \to \infty. \tag{3.9}$$

Many authors provide a more detailed discussion on weak convergence; for example [7], [11, Appendix A] or [12].

Now we extend the relative step sizes, μ_n, to continuous time; for all $i \in I$ let $u_i(t) = \mu_{n+1}(i)$ for $t \in [\bar{\tau}_n, \bar{\tau}_{n+1})$ and let $u(\cdot) = \big(u_1(\cdot), \ldots, u_K(\cdot)\big)$. For all $i \in I$ and $t > 0$ define $\bar{u}_i^n(t) := u_i(t + \bar{\tau}_n) \in \mathcal{U}$.

LEMMA 3.4. *Under (A1)(a), (A2)(b) and (A4), for all $i \in I$ and for any $T > 0$, $\{\bar{u}_i^n(\cdot)\}_{n\in\mathbb{N}}$ converges along a subsequence to a limit point $\tilde{u}_i(\cdot)$ such that for some $\varepsilon > 0$ and any $0 < v \le T$,*

$$\int_0^v \tilde{u}_i(s)\,\mathrm{d}s \ge v\varepsilon, \quad \textit{a.s.}. \tag{3.10}$$

PROOF. See Appendix A.3. □

COROLLARY 3.5. *For any $T > 0$ let $\tilde{u}(\cdot)$ be a limit point of $\{\bar{u}^n(\cdot)\}_{n\in\mathbb{N}}$ in $\mathcal{U}$. Then under (A1)(a), (A2)(b) and (A4) there exists an $\varepsilon > 0$ such that for all $i \in I$ and any $t, v > 0$ such that $t + v \le T$,*

$$\int_t^{t+v} \tilde{u}_i(s)\,\mathrm{d}s \ge v\varepsilon, \quad \textit{a.s.}.$$

PROOF. See Appendix A.3. □

We now expand upon the discussions in Sections 2.2 and 3.1 on producing a sequence of matrices $\{\tilde{M}_n\}_{n\in\mathbb{N}}$. In order to use the differential inclusions framework described in Section 2.1 we need to define a sequence of diagonal matrices, $\{\tilde{M}_n\}_{n\in\mathbb{N}}$ with diagonal entries which are always in the set $[\varepsilon, 1]$,

Downloaded from informs.org by [216.73.217.172] on 15 November 2026, at 06:26 . For personal use only, all rights reserved.

for some $\varepsilon > 0$, and such that the terms converge (weakly) to the same limit as the terms of $\{M_n\}_{n\in\mathbb{N}}$. Recall that M_n is a diagonal matrix containing the $\mu_n(i)$ terms.

Fix $\varepsilon > 0$ taken from Lemma 3.4 and define a new function $v(\cdot) : \mathbb{R} \to \mathbb{R}^K$ such that

$$v_i(t) := \max\{u_i(t), \varepsilon\}.$$

For all $t > 0$ let $\bar{v}_i^n(t) := v_i(t+\bar{\tau}_n)$. Corollary 3.5 shows that, with respect to the topology of $\mathcal{U}$, in the limit $u_i(t) \in [\varepsilon, 1]$ for almost every t and similarly $v_i(t) \in [\varepsilon, 1]$ for all t. From this it is clear $u_i(t)$ and $v_i(t)$ have the same limit point in $\mathcal{U}$. That is, if $\tilde{u}(t)$ is a limit point of $\{\bar{u}^n(\cdot)\}_{n\in\mathbb{N}}$ then it is also a limit point of $\{\bar{v}^n(\cdot)\}_{n\in\mathbb{N}}$. Hence for any $T > 0$ there exists a subsequence $k(n)$ such that for $h(\cdot) \in L^2([0,T])$,

$$\int_0^T \tilde{u}_i(t)h(t)\,\mathrm{d}t = \lim_{n\to\infty}\int_0^T \bar{u}_i^{k(n)}(t)h(t)\,\mathrm{d}t = \lim_{n\to\infty}\int_0^T \bar{v}_i^{k(n)}(t)h(t)\,\mathrm{d}t.$$

However, the key interest here is in the convergence of $u_i(\cdot)$ and $v_i(\cdot)$. Following the reasoning of Borkar [9] and Konda and Borkar [14], it does not matter whether $\bar{u}_i^n(\cdot)$ and $\bar{v}_i^n(\cdot)$ converge directly or via a subsequence as this does not affect the convergence of the continuous processes $u_i(\cdot)$ and $v_i(\cdot)$. Hence we can say that $u_i(\bar{\tau}_n + \cdot), v_i(\bar{\tau}_n + \cdot)$ converge weakly to a limit point $\tilde{u}_i(\cdot)$, or equivalently, if $h(\cdot)$ is a bounded, continuous function then for all $i = 1, \ldots, K$,

$$\lim_{n\to\infty}\left\|\int_{\bar{\tau}_n}^{\bar{\tau}_n+v}\big[v_i(s) - u_i(s)\big]h(s)\,\mathrm{d}s\right\| = 0. \tag{3.11}$$

Define $\tilde{M}(t)$ to be the $K \times K$ diagonal matrix of the $v_i(t)$ terms and let $\tilde{M}_n := \tilde{M}(\bar{\tau}_n)$.

LEMMA 3.6. *Almost surely under assumptions (A1)(a), (A2) and (A4),*

$$\lim_{n\to\infty}\sup_k\left\{\left\|\sum_{i=n}^{k-1}\bar{\alpha}_{i+1}f_i\big(M_{i+1} - \tilde{M}_{i+1}\big)\right\|; k = n+1, \ldots \bar{m}(\bar{\tau}_n + T)\right\} = 0.$$

PROOF. See Appendix A.4. The proof relies on (3.11). □

3.4. *Proof of Theorem 3.1.* We must verify that the four conditions of Theorem 2.3 hold for the stochastic approximation process in (3.6) to ascertain that $\bar{x}(t)$ is an asymptotic pseudo-trajectory of (3.8).

Downloaded from informs.org by [216.73.217.172] on 15 November 2026, at 06:26 . For personal use only, all rights reserved.

Fix $T > 0$. Then,

$$\sup_k \left\{ \left\| \sum_{i=n}^{k-1} \bar{\alpha}_{i+1} \bar{V}_{i+1} \right\| ; k = n+1, \dots \bar{m}(\bar{\tau}_n + T) \right\}$$

$$\leq \sup_k \left\{ \left\| \sum_{i=n}^{k-1} \bar{\alpha}_{i+1} M_{i+1} V_{i+1} \right\| ; k = n+1, \dots \bar{m}(\bar{\tau}_n + T) \right\} \tag{3.12}$$

$$+ \sup_k \left\{ \left\| \sum_{i=n}^{k-1} \bar{\alpha}_{i+1} f_i\big(M_{i+1} - \tilde{M}_{i+1}\big) \right\| ; k = n+1, \dots \bar{m}(\bar{\tau}_n + T) \right\}. \tag{3.13}$$

Using Lemma 3.3 and Lemma 3.6 immediately gives that (3.12) and (3.13) converge to zero a.s., and hence this verifies that property (i) holds. Assumption (A1)(a) directly gives that (ii) holds. Lastly, it is straightforward to verify that, under (A1)-(A5), $\bar{F}(\cdot)$ is a Marchaud map which verifies condition (iii), and (A1)(b) is equivalent to (iv). □

4. Two-timescale asynchronous stochastic approximation. A useful extension of standard stochastic approximations is to two timescales. This concept was originally introduced by Borkar [8] and has later been used by Leslie and Collins [19] for multiple timescales and Konda and Borkar [14] for two-timescales asynchronous stochastic approximation. If we have a coupled pair of stochastic approximations where one system can be seen to update more aggressively than the other then, in the limit, the aggressive process is always fully adjusted to the value of the other process. This is all controlled through the user's choice of step sizes in the stochastic approximation. The main result in this section is Corollary 4.8, which comes from combining Theorem 3.1 with the previous work of Konda and Borkar [14].

4.1. *Notation.* In what is to follow we consider the extension of Theorem 3.1 to the two-timescales setting, with updates $\{x_n\}_{n\in\mathbb{N}}$ and $\{y_n\}_{n\in\mathbb{N}}$ on different timescales. Let I be the set of individual elements of the x process as in Section 3, and define J similarly for the y process. Let $K = |I|$ and $L = |J|$ so that for all n, $x_n \in \mathbb{R}^K$ and $y_n \in \mathbb{R}^L$. As in Section 3 let $\bar{I} \subset 2^I$ be the set containing all combinations of elements in I which have a positive probability of being part of the asynchronous update, and define $\bar{J} \subset 2^J$ in the same manner for the y process. At iteration n let $\bar{I}_n \in \bar{I}$ and $\bar{J}_n \in \bar{J}$ be the updated components of each timescale respectively. Let each component of the two processes have a counter for the number of times it has been

Downloaded from informs.org by [216.73.217.172] on 15 November 2026, at 06:26 . For personal use only, all rights reserved.

selected to be updated defined by,

$$\nu_n(i) := \sum_{k=1}^{n} \mathbb{I}_{\{i\in\bar{I}_k\}}, \quad \phi_n(j) := \sum_{k=1}^{n} \mathbb{I}_{\{j\in\bar{J}_k\}}.$$

Here $\nu_n(i)$ is as in Section 3 and $\phi_n(j)$ has an analogous definition for the $\{y_n\}_{n\in\mathbb{N}}$ process. Let $\{V_n\}_{n\in\mathbb{N}}$, $\{U_n\}_{n\in\mathbb{N}}$ be martingale difference processes defined on $\mathbb{R}^K$ and $\mathbb{R}^L$ respectively, and $\{d_n\}_{n\in\mathbb{N}}$, $\{e_n\}_{n\in\mathbb{N}} \to 0$ as $n \to \infty$ also defined on $\mathbb{R}^K$ and $\mathbb{R}^L$ respectively. Let $V_n(i), d_n(i) \in \mathbb{R}$ be component i of V_n and d_n, and similarly let $U_n(j), e_n(j) \in \mathbb{R}$ be component j of U_n and e_n. As in the previous sections $\{\alpha(n)\}_{n\in\mathbb{N}}$, and now $\{\gamma(n)\}_{n\in\mathbb{N}}$, are positive, deceasing sequences of step sizes. Similar restrictions to those in (A2) will be placed on $\{\alpha(n)\}_{n\in\mathbb{N}}$, $\{\gamma(n)\}_{n\in\mathbb{N}}$ with an additional requirement for the two-timescale arguments to be valid; this will be made precise in Section 4.2. Finally, $F(\cdot,\cdot) : \mathbb{R}^K \times \mathbb{R}^L \to \mathbb{R}^K$ and $G(\cdot,\cdot) : \mathbb{R}^K \times \mathbb{R}^L \to \mathbb{R}^L$ are set-valued maps and for $i = 1\ldots, K$, $j = 1, \ldots, L$

$$F_i(x,y) := \{z_i\,;\, z \in F(x,y)\},$$
$$G_j(x,y) := \{z_j\,;\, z \in G(x,y)\}.$$

For all $i = 1\ldots, K$ and $j = 1, \ldots, L$ consider the following coupled process,

$$\begin{aligned} x_{n+1}(i) - x_n(i) &\\ &\in \alpha\big(\nu_{n+1}(i)\big)\mathbb{I}_{\{i\in\bar{I}_{n+1}\}}\big[F_i(x_n,y_n) + V_{n+1}(i) + d_{n+1}(i)\big],\\ y_{n+1}(j) - y_n(j) &\\ &\in \gamma\big(\phi_{n+1}(j)\big)\mathbb{I}_{\{j\in\bar{J}_{n+1}\}}\big[G_j(x_n,y_n) + U_{n+1}(j) + e_{n+1}(j)\big]. \end{aligned} \tag{4.1}$$

Notice that the only change to the first process from Sections 2 and 3 is that the mean field now depends on y_n as well as x_n. It follows that the asynchronous and relative step sizes retain the same form. Recall these definitions and extend them for the $\{y_n\}_{n\in\mathbb{N}}$ process:

$$\bar{\alpha}_n := \max_{i\in\bar{I}_n} \alpha\big(\nu_n(i)\big), \quad \mu_n(i) := \frac{\alpha\big(\nu_n(i)\big)}{\bar{\alpha}_n}\mathbb{I}_{\{i\in\bar{I}_n\}},$$
$$\bar{\gamma}_n := \max_{j\in\bar{J}_n} \gamma\big(\phi_n(j)\big), \quad \sigma_n(j) := \frac{\gamma\big(\phi_n(j)\big)}{\bar{\gamma}_n}\mathbb{I}_{\{j\in\bar{J}_n\}}.$$

As in Section 2.2 let M_n be the $K \times K$ diagonal matrix of the $\mu_n(i)$ terms and similarly let N_n be the $L \times L$ diagonal matrix of the $\sigma_n(j)$ terms. The coupled stochastic process (4.1) can be written more concisely as,

$$\begin{aligned} x_{n+1} - x_n - \bar{\alpha}_{n+1}M_{n+1}\big[V_{n+1} + d_{n+1}\big] &\in \bar{\alpha}_{n+1}M_{n+1}\cdot F(x_n,y_n),\\ y_{n+1} - y_n - \bar{\gamma}_{n+1}N_{n+1}\big[U_{n+1} + e_{n+1}\big] &\in \bar{\gamma}_{n+1}N_{n+1}\cdot G(x_n,y_n). \end{aligned} \tag{4.2}$$

Finally, define the two timescales; let $\bar{\tau}_0 := 0$, $\bar{\tau}_k := \sum_{i=1}^{k} \bar{\alpha}_i$, $\bar{\rho}_0 := 0$, $\bar{\rho}_k := \sum_{i=1}^{k} \bar{\gamma}_i$. The division of time on the 'slow' timescale is given by the increments $\{\bar{\tau}_n\}_{n\in\mathbb{N}}$ and similarly for $\{\bar{\rho}_n\}_{n\in\mathbb{N}}$ on the 'fast' timescale. In a similar manner to the previous sections let $\bar{m}_\alpha(t) := \sup\{k \geq 0; t \geq \bar{\tau}_k\}$, and $\bar{m}_\gamma(t) := \sup\{k \geq 0; t \geq \bar{\rho}_k\}$.

4.2. *Assumptions.* We state the assumptions (B1)-(B6) used for the convergence results of the two-timescale algorithm (4.2). These are exactly analogous to (A1)-(A5) and are simply extended to accommodate the two-timescales framework. The exceptions to this are (B2)(c) and (B6) and the slight adaptations to (B3), which are in line with those used by Borkar [8] and Konda and Borkar [14]. In (B4) we have produced a single combined Markov chain instead of one for each of the $\{x_n\}_{n\in\mathbb{N}}$ and $\{y_n\}_{n\in\mathbb{N}}$ processes to present a clearer assumption.

(B1) (a) For compact sets, $C \subset \mathbb{R}^K$, $D \subset \mathbb{R}^L$, $x_n \in C$, $y_n \in D$ for all n.

(b) $\{d_n\}_{n\in\mathbb{N}}$ and $\{e_n\}_{n\in\mathbb{N}}$ are bounded sequences such that $d_n, e_n \to 0$ as $n \to \infty$.

(B2) The following must be true for $a(n) = \alpha(n)$ and $a(n) = \gamma(n)$;

(a) $\sum_n a(n) = \infty$ and $a(n) \to 0$ as $n \to \infty$.

(b) For $x \in (0,1)$, $\sup_n a([xn])/a(n) < A_x < \infty$. In addition, for all n, $a(n) \geq a(n+1)$.

(c) $\frac{\alpha(n)}{\gamma(n)} \to 0$.

(B3) (a) $F(\cdot,\cdot) : C \times D \to C$ is upper semi-continuous and for all $y \in D$ $F(\cdot, y) : C \to C$ is a Marchaud map.

(b) $G(\cdot,\cdot) : C \times D \to D$ is a Marchaud map.

The first and second assumptions are direct extensions of (A1) and (A2) to two-timescales with the addition of (B2)(c) which is a standard two-timescale assumption used by Borkar [8]. Condition (B3)(a) is similar to (A3) for the 'slow' timescale, however this must hold for all values of the 'fast' timescale. (B3)(b) is a similar condition for the 'fast' timescale.

Define $\bar{H} \subset \bar{I} \times \bar{J}$ such that if $\mathcal{I} \in \bar{I}$ and $\mathcal{J} \in \bar{J}$ then $(\mathcal{I}, \mathcal{J}) \in \bar{H}$ if and only if $\mathcal{I}$ and $\mathcal{J}$ have a positive probability of occurring simultaneously (at the same iteration). This means that $\bar{H}$ is the combination of elements across $\bar{I} \times \bar{J}$ which have positive probability of being updated at any particular iteration. At iteration n $\bar{H}_n \in \bar{H}$ is taken to be the updated components across I and J. In addition, let $z_n = (x_n, y_n) \in C \times D$ and $\mathcal{F}_n$ be a sigma algebra containing all the information up to and including the n^{th} iteration. That is

Downloaded from informs.org by [216.73.217.172] on 15 November 2026, at 06:26 . For personal use only, all rights reserved.

$\mathcal{F}_n = \sigma(\{\bar{H}_m\}_m, \{z_m\}_m, \{\nu_m(i)\}_{i,m}, \{\phi_m(j)\}_{j,m}; \forall m \leq n, i = 1, \ldots, K, j = 1, \ldots, L)$.

(B4) (a) For all $z \in C \times D$ and $\mathcal{H}_n, \mathcal{H}_{n+1} \in \bar{H}$,

$$\mathbb{P}\Big(\bar{H}_{n+1} = \mathcal{H}_{n+1} | \mathcal{F}_n\Big) = \mathbb{P}\Big(\bar{H}_{n+1} = \mathcal{H}_{n+1} | \bar{H}_n = \mathcal{H}_n, z_n = z\Big).$$

Let

$$Q_{(\mathcal{H}_n, \mathcal{H}_{n+1})}(z) := \mathbb{P}\Big(\bar{H}_{n+1} = \mathcal{H}_{n+1} | \bar{H}_n = \mathcal{H}_n, z_n = z\Big).$$

(b) For all $z \in C \times D$ the transition probabilities $Q_{(\mathcal{H}_n, \mathcal{H}_{n+1})}(z)$ form aperiodic, irreducible Markov chains over $\bar{H}$ and for all $i \in I$, $j \in J$ there exists a $\mathcal{H}, \mathcal{H}' \in \bar{H}$, such that $i \in \mathcal{H}$, $j \in \mathcal{H}'$.

(c) the map $z \mapsto Q_{(\mathcal{H}_n, \mathcal{H}_{n+1})}(z)$, is Lipschitz continuous.

(B5) Separately for both

$$(a(n), W_n, \bar{Z}_n) = (\alpha(n), V_n, \bar{I}_n) \quad \text{and} \quad (a(n), W_n, \bar{Z}_n) = (\gamma(n), U_n, \bar{J}_n)$$

assume that for all n, W_{n+1} and $\bar{Z}_{n+1}$ are uncorrelated given $\mathcal{F}_n$ and either

(a) For some $q \geq 2$

$$\sum_n a(n)^{1+q/2} < \infty, \quad \text{and} \quad \sup_n \mathbb{E}\big(\|W_n\|^q\big) < \infty;$$

or

(b) $W_n(i)$ is independent of $W_n(j)$ for $i \neq j$. Let $\langle a, b \rangle = \sum_k a_k b_k$. Then assume there exists a positive Γ such that for all $\theta \in \mathbb{R}^K / \mathbb{R}^L$ (depending on the dimension of W_n),

$$\mathbb{E}\Big[\exp\Big\{\langle \theta, W_{n+1} \rangle\Big\} | \mathcal{F}_n\Big] \leq \exp\Big\{\frac{\Gamma}{2}\|\theta\|^2\Big\};$$

and

$$\sum_n e^{-c/a(n)} < \infty,$$

for each $c > 0$.

(B4) and (B5) are straightforward extensions to (A4) and (A5) where in (B4) we have chosen to create a combined Markov chain over both $\bar{I}$ and $\bar{J}$ to present a clearer assumption. As a result of (B4), Lemma 3.4 gives that every element of $\bar{H}$ (and hence every element of I and J) is updated some minimum proportion of the time in the limit (see Section

Downloaded from informs.org by [216.73.217.172] on 15 November 2026, at 06:26 . For personal use only, all rights reserved.

3.3). Let $\varepsilon > 0$ be this minimum proportion. As with (A5), we say that (B5) holds if for $(\alpha(n), V_n, \bar{I}_n)$ either (B5)(a) or (B5)(b) holds and similarly for $(\gamma(n), U_n, \bar{J}_n)$ either (B5)(a) or (B5)(b) holds. Define $\bar{F}(x,y) := \Omega_K^\varepsilon \cdot F(x,y)$ and $\bar{G}(x,y) := \Omega_L^\varepsilon \cdot G(x,y)$ analogously to the definition of $\bar{F}(\cdot)$ in (3.5), where Ω_K^ε, Ω_L^ε are as defined in (3.4).

(B6) For all $x \in C$ and every $\varepsilon > 0$ the differential inclusion,

$$\frac{dy}{dt} \in \bar{G}(x,y),$$

has a unique global attractor, $\Lambda(x)$, where $\Lambda(\cdot) : \mathbb{R}^K \to \mathbb{R}^L$ is bounded, continuous and single-valued for all $x \in C$.

The final assumption, (B6), is the asynchronous equivalent to the 'fast' timescale convergence criteria used by Borkar [8] and we use it throughout this section. However, at the end of Section 4.4 we provide an alternative assumption, (B6′), which allows the 'fast' timescale to converge instead to a globally attracting set, as opposed to a continuous single-valued function.

4.3. *Convergence of the 'fast' timescale.* Many of the proofs in this section use arguments which are identical to the corresponding results of Section 3; where this is the case we do not go into detail and instead direct the reader to the appropriate result(s) to identify the method used.

Firstly, we require an additional lemma used by Konda and Borkar [14] which shows that the key two-timescales arguments made by Borkar [8] will continue to hold in the asynchronous case.

LEMMA 4.1. *Under (B2) and (B4)* $\bar{\alpha}_n, \bar{\gamma}_n \to 0$ *and* $\frac{\bar{\alpha}_n}{\bar{\gamma}_n} \to 0$ *almost surely.*

PROOF. The proof of this result is identical to [14, Lemma 4.6]. The requisite assumptions are encapsulated in (B2) and the result of Lemma A.1, in Appendix A.1. □

Again, we follow the method of Borkar [11] when examining the set-valued updates. Define f_n and g_n in the following manner; for $i = 1, \ldots, K$, $j = 1, \ldots, L$ let

$$f_n(i)\mu_{n+1}(i) := \frac{x_{n+1}(i) - x_n(i)}{\bar{\alpha}_{n+1}} - \mu_{n+1}(i)\big[V_{n+1}(i) + d_{n+1}(i)\big],$$
$$g_n(j)\sigma_{n+1}(j) := \frac{y_{n+1}(j) - y_n(j)}{\bar{\gamma}_{n+1}} - \sigma_{n+1}(j)\big[U_{n+1}(j) + e_{n+1}(j)\big].$$

Downloaded from informs.org by [216.73.217.172] on 15 November 2026, at 06:26 . For personal use only, all rights reserved.

Downloaded from informs.org by [216.73.217.172] on 15 November 2026, at 06:26 . For personal use only, all rights reserved.

As in Section 3.1, if $\mu_{n+1}(i) = 0$ then we can select any $f_n(i) \in F_i(x_n, y_n)$ and similarly if $\sigma_{n+1}(j) = 0$ then we can select any $g_n(j) \in G_j(x_n, y_n)$. f_n and g_n represent the realised values of $F(x_n, y_n)$ and $G(x_n, y_n)$ respectively. We express (4.2) as,

$$\begin{aligned} x_{n+1} &= x_n + \bar{\alpha}_{n+1} M_{n+1}\big[f_n + V_{n+1} + d_{n+1}\big], \\ y_{n+1} &= y_n + \bar{\gamma}_{n+1} N_{n+1}\big[g_n + U_{n+1} + e_{n+1}\big]. \end{aligned} \tag{4.3}$$

Let $\underline{0}^K$ be a zero vector of length K, and define

$$z_n := \begin{pmatrix} x_n \\ y_n \end{pmatrix}, \quad \Gamma_n := \begin{pmatrix} M_n \\ N_n \end{pmatrix}, \quad \zeta_n := \begin{pmatrix} \frac{\bar{\alpha}_n}{\bar{\gamma}_n} V_n \\ U_n \end{pmatrix},$$
$$\kappa_{n+1} := \begin{pmatrix} \frac{\bar{\alpha}_n}{\bar{\gamma}_n}(f_n + d_{n+1}) \\ e_{n+1} \end{pmatrix}, \qquad \Psi(z_n) := \begin{pmatrix} \underline{0}^K \\ G(z_n) \end{pmatrix}.$$

Then we can express the coupled process in (4.3) as a single iterative process,

$$z_{n+1} - z_n - \bar{\gamma}_{n+1}\Gamma_{n+1}\big[\zeta_{n+1} + \kappa_{n+1}\big] \in \bar{\gamma}_{n+1}\Gamma_{n+1} \cdot \Psi(z_n). \tag{4.4}$$

This is in the same form as equation (2.5) which is examined throughout Section 3. The limiting behaviour of (4.4) can therefore be studied using the same method as in Section 3. Although assumptions (A1)-(A4) are embedded in (B1)-(B6), we negate the need for (A5) under (B1)-(B6) by immediately proving the corresponding result in Lemma 3.3 for (4.4).

LEMMA 4.2. *Under (B2)(b),(B4) and (B5), with probability 1,*

$$\lim_{n\to\infty} \sup_k \left\{ \left\| \sum_{j=n}^{k-1} \bar{\gamma}_{j+1}\Gamma_{j+1}\zeta_{j+1} \right\| ; k = n+1, \ldots, \bar{m}_\gamma(\bar{\rho}_n + T) \right\} = 0$$

PROOF. Firstly, let $\bar{\xi}_n^\alpha(T) := \sup\{k \geq 0; t \geq \bar{\tau}_{n+k}\}$ and $\bar{\xi}_n^\gamma(T) := \sup\{k \geq 0; t \geq \bar{\rho}_{n+k}\}$. By Lemma 4.1, $\bar{\gamma}_n > \bar{\alpha}_n$ in the limit and hence $\bar{\xi}_n^\alpha(t) \geq \bar{\xi}_n^\gamma(t)$. From this, for limiting values of n,

$$\bar{m}_\alpha(\bar{\tau}_n + T) = n + \bar{\xi}_n^\alpha(T) \geq n + \bar{\xi}_n^\gamma(T) = \bar{m}_\gamma(\bar{\rho}_n + T). \tag{4.5}$$

Now,

$$\begin{aligned} \left\| \sum_{j=n}^{k-1} \bar{\gamma}_{j+1}\Gamma_{j+1}\zeta_{j+1} \right\| \leq & \left\| \sum_{j=n}^{k-1} \bar{\alpha}_{j+1} M_{j+1} V_{j+1} \right\| \\ & + \left\| \sum_{j=n}^{k-1} \bar{\gamma}_{j+1} N_{j+1} U_{j+1} \right\|. \end{aligned}$$

Using that the assumptions for Lemma 3.3 are contained in (B2)(b), (B4) and (B5), the first term converges to zero for $k = n+1, \ldots, \bar{m}_\alpha(\bar{\tau}_n + T)$ and with the same arguments, the second term converges to zero for $k = n+1, \ldots, \bar{m}_\gamma(\bar{\rho}_n + T)$. Combining this with (4.5) proves the result for $k = n+1, \ldots, \bar{m}_\gamma(\bar{\rho}_n + T)$ as required. □

Now we note that assumption (B4) allows us to use identical arguments to those in Lemma 3.4. Recall the definition of Ω_k^δ from (3.4). Following the arguments of Section 3.3, there exists an $\varepsilon > 0$ such that we can define a $\tilde{\Gamma}_n \in \Omega_{K+L}^\varepsilon$ which will weakly converge to the same limit as Γ_n. Let $\bar{z}(t)$ be the linear interpolation of (4.4) in the same manner as the single timescale case in (3.7). Fix $\varepsilon > 0$ as above and let $\bar{\Psi}(\cdot) := \Omega_{K+L}^\varepsilon \cdot \Psi(\cdot)$ so that $\bar{\Psi}(\cdot) : \mathbb{R}^{K+L} \to \mathbb{R}^{K+L}$.

LEMMA 4.3. *Under assumptions (B1)-(B6), with probability 1, $\bar{z}(t) \in \mathbb{R}^{K+L}$ is an asymptotic pseudo-trajectory of the differential inclusion,*

$$\frac{dz}{\mathrm{d}t} \in \bar{\Psi}(z). \tag{4.6}$$

PROOF. Assumptions (A1)-(A4) for the stochastic approximation process in (4.4) are contained within (B1)-(B6). This, combined with Lemma 4.2 allows us to use Theorem 3.1 to give the result. □

Let the linear interpolations of the two-timescales in (4.3) be denoted by $\bar{x}(t)$ and $\bar{y}(t)$ respectively analogously to the single timescale case in (3.7).

COROLLARY 4.4. *Under assumptions (B1)-(B6), with probability 1, the interpolated process*

$$\big(\bar{x}(t), \bar{y}(t)\big) \to \Big\{\big(x, \Lambda(x)\big); x \in C\Big\} \text{ as } t \to \infty.$$

PROOF. Immediate from Lemma 4.3 and (B6) using the same arguments as Borkar [8]. □

4.4. *Convergence of the 'slow' timescale.* Now since we have a function $F(\cdot,\cdot)$ which depends on two variables, but we are treating one of these as fully calibrated to the other, we have a slightly different framework to that of Benaim, Hofbauer and Sorin [5]. Therefore we present a slight variation on their perturbed solution [5, Definition (II)]. Despite this we are still able to show that this is an asymptotic pseudo-trajectory to the desired differential inclusion. Hence the same convergence results still apply. To reduce notation,

Downloaded from informs.org by [216.73.217.172] on 15 November 2026, at 06:26 . For personal use only, all rights reserved.

define $F^{\Lambda}(\cdot) : \mathbb{R}^K \to \mathbb{R}^K$ as $F^{\Lambda}(x) := F\big(x, \Lambda(x)\big)$. Note that under (B3)(a) and (B6) $F^{\Lambda}(\cdot)$ is a Marchaud map. Following our previous notation, let

$$\bar{F}^{\Lambda}(x) := \Omega_K^{\varepsilon} \cdot F^{\Lambda}(x).$$

DEFINITION 4.5. A continuous function $z : \mathbb{R}^+ \to C$ is a *jointly perturbed solution* to the differential inclusion,

$$\frac{dx}{dt} \in \bar{F}^{\Lambda}(x), \tag{4.7}$$

if

(i) z is absolutely continuous.

(ii) $\frac{dz(t)}{dt} - V(t) - d(t) \in \bar{F}^{\Lambda}_{\delta(t)}\big(z(t)\big)$, for almost every $t > 0$, and for some bounded $d(t), \delta(t) \to 0$, where

$$\bar{F}^{\Lambda}_{\delta}(x) = \{f \in C\,; \|x' - x\| \le \delta, y' \in D, \\ \|y' - \Lambda(x')\| \le \delta, \inf_{f' \in \bar{F}(x', y')} \|f' - f\| \le \delta\}.$$

(iii) $t \to V(t)$ is a locally integrable function such that for all $T > 0$

$$\lim_{t \to \infty} \sup_{0 \le v \le T} \left\| \int_t^{t+v} V(s)\, \mathrm{d}s \right\| = 0.$$

The key difference between a jointly perturbed solution and a perturbed solution is that in the original work of Benaim, Hofbauer and Sorin [5] (and as in Section 3) the mean field depends on a single variable. In contrast to this, here the mean field depends on two variables. Hence in part (ii) of Definition 4.5 we must allow for perturbations in both variables simultaneously instead of perturbing just the one.

LEMMA 4.6. *Under assumptions (B1)-(B6) a jointly perturbed solution of* (4.7) *is an asymptotic pseudo-trajectory to the flow induced by* (4.7).

PROOF. The proof is identical to the proof of [5, Theorem 4.2] which establishes that a perturbed solution is an asymptotic pseudo-trajectory. □

Define $\tilde{M}(t)$ and $\tilde{M}_n$ in an identical manner to the same terms in Section 3.4. Corollary 4.4 allows us to consider the updates on the 'slow' timescale of the coupled process in (4.3) given by,

$$x_{n+1} - x_n - \bar{\alpha}_{n+1}\left[\bar{V}_{n+1} + \bar{d}_{n+1}\right] \in \bar{\alpha}_{n+1}\bar{F}(x_n, y_n), \tag{4.8}$$

Downloaded from informs.org by [216.73.217.172] on 15 November 2026, at 06:26 . For personal use only, all rights reserved.

Downloaded from informs.org by [216.73.217.172] on 15 November 2026, at 06:26 . For personal use only, all rights reserved.

where, as in Section 3.4, $\bar{V}_{n+1} = M_{n+1}V_{n+1} + [M_{n+1} - \tilde{M}_{n+1}]f_n$ and $\bar{d}_{n+1} = M_{n+1}d_{n+1}$. As with the single timescale framework let $\bar{x}(t)$ be a linear interpolation of (4.8) in the same manner as (3.7). That is,

$$\bar{x}(\bar{\tau}_n + s) = x_n + s\frac{x_{n+1} - x_n}{\bar{\alpha}_{n+1}}, \quad s \in [0, \bar{\alpha}_{n+1}) .$$

THEOREM 4.7. *Under assumptions (B1)-(B6), with probability 1, $\bar{x}(\cdot)$ is an asymptotic pseudo-trajectory to the differential inclusion* (4.7).

PROOF. We show that $\bar{x}(\cdot)$ is a jointly perturbed solution of (4.7) and hence by Lemma 4.6 it is an asymptotic pseudo-trajectory of (4.7).

The proof is almost identical to the proof of [5, Proposition 1.3]. The differences come with our choice of $\delta(t) = \|\bar{x}(t) - x_{m(t)}\| + \|y_{m(t)} - \Lambda(x_{m(t)})\|$, the first term of which will converge to zero as in the original proof and the second of which converges to zero as $\|y_n - \Lambda(x_n)\| \to 0$ almost surely as a result of the convergence of the 'fast' timescale. Clearly from Definition 4.5 part (ii), $F(x_{m(t)}, y_{m(t)}) \subset F^{\Lambda}_{\delta(t)}(\bar{x}(t))$. The rest of the proof completes as in [5, Proposition 1.3]. □

COROLLARY 4.8. *If for all $\varepsilon > 0$ there is a global attractor, $A \subset C$, for the differential inclusion* (4.7) *and assumptions (B1)-(B6) are satisfied, then the $\{x_n\}_{n\in\mathbb{N}}$ component of the two-timescale iterative process* (4.2) *will almost surely converge to A.*

PROOF. Immediate by combining Corollary 4.4 and Theorem 4.7 with [5, Proposition 3.27]. □

REMARK 1. It should be clear that the methods in this chapter can be applied to a standard two-timescale stochastic approximation (which doesn't use asynchronous updates) with set-valued mean fields. In this case (B4) is trivially satisfied, (B2)(b) can be removed and the use of the sets Ω^{ε}_K, Ω^{ε}_L can be replaced by a single $K \times K$ and $L \times L$ identity matrix respectively. This means that $\bar{F}(x, y) = F(x, y)$ and similarly $\bar{G}(x, y) = G(x, y)$. This naturally extends the original two-timescale approach of Borkar [8] to allow set-valued mean fields.

REMARK 2. This framework allows for the 'fast' timescale to converge to a set of limit points. Assumption (B6) can be replaced with the following

(B6′) For all $x \in C$ the differential inclusion,

$$\frac{dy}{dt} \in \bar{G}(x, y)$$

has a global attractor, $\Lambda(x)$ where $\Lambda(\cdot) : \mathbb{R}^K \to \mathbb{R}^L$ is an upper semi-continuous set valued map, such that for all $x \in C$, $\Lambda(x)$ is compact, convex and non-empty. In addition, for all $x \in C$, $F(x, \Lambda(x))$ is a convex set-valued map, i.e., for all $\lambda, \lambda' \in \Lambda(x)$,

$$\alpha F(x,\lambda) + (1-\alpha)F(x,\lambda') \subset F\big(x, \alpha\lambda + (1-\alpha)\lambda'\big), \qquad \text{for all } \alpha \in [0,1].$$

Under (B6′) $F^{\Lambda}(x)$ will still be a Marchaud map and, although the details of the arguments in Section 4.4 change slightly, the method remains almost identical.

5. An application: Learning in a Markov decision process. In this section we provide an example where our two-timescale asynchronous stochastic approximation approach is needed. The algorithm is an actor-critic style algorithm based upon estimating rewards and playing an ε-greedy best response to these estimates. Similarly to Konda and Borkar [14] we make assumptions on the Markov decision process (MDP) and the coupled learning algorithm separately. These correspond to (B1)-(B6).

Firstly, we begin by outlining a suitable infinite horizon, discounted reward MDP described by the tuple $\langle \mathcal{S}, \mathcal{A}, P, r, \beta \rangle$. $\mathcal{S}$ is the state space of the MDP, $\mathcal{A}$ is the set of actions of the decision-maker (agent), and $\mathcal{A}(s)$ is the set of actions available to the agent in state $s \in \mathcal{S}$. P represents the form of the stochastic transitions and in what is to follow we take $P_{ss'}(a)$ to denote probability of transitioning from state $s \in \mathcal{S}$ to $s' \in \mathcal{S}$ when the agent has selected action $a \in \mathcal{A}(s)$. The expected reward to the decision maker for selecting action $a \in \mathcal{A}(s)$ in state $s \in \mathcal{S}$ is denoted by $r(s,a)$, and $\beta \in (0,1)$ is the discount factor.

Let $s_n \in \mathcal{S}$ and $a_n \in \mathcal{A}(s_n)$ be the state and the action, respectively, selected by the decision-maker at iteration n. Assume that at every iteration the agent observes the state of the process and a noisy version of the reward received from the action they have chosen, denoted by R_n. If $\mathcal{F}_n := \{a_m, s_m, R_{m-1}; m = 1, \ldots, n\}$ then $\mathbb{E}[R_n | \mathcal{F}_n] = r(s_n, a_n)$, and we assume that R_n has a finite variance. Let $K := |\mathcal{S}|$ and $\Delta\big(\mathcal{A}(s)\big)$ represent the set of probability distributions over $\mathcal{A}(s)$. Then we denote the combination of K probability distributions as $\Delta := \Delta\big(\mathcal{A}(1)\big) \times \ldots \times \Delta\big(\mathcal{A}(K)\big)$. A strategy for state $s \in \mathcal{S}$ is denoted by $\pi(s) \in \Delta\big(\mathcal{A}(s)\big)$ and let $\pi := \big(\pi(1), \ldots, \pi(K)\big) \in \Delta$ be a strategy over all states. $\pi(s,a)$ is defined as the probability that action a is taken in state s. The agent wishes to find a strategy, π, to maximise their expected discounted reward,

$$\mathbb{E}\Big[\sum_{n=1}^{\infty} \beta^n r(s_n, a_n)\Big].$$

Downloaded from informs.org by [216.73.217.172] on 15 November 2026, at 06:26 . For personal use only, all rights reserved.

Downloaded from informs.org by [216.73.217.172] on 15 November 2026, at 06:26 . For personal use only, all rights reserved.

Define $V^\pi(s)$ for all $s \in \mathcal{S}$ in the following manner,

$$V^\pi(s) = \sum_{a \in \mathcal{A}(s)} \pi(s,a) \left[r(s,a) + \beta \sum_{s' \in S} P_{ss'}(a) V^\pi(s) \right]. \tag{5.1}$$

For all $s \in S$ there exists a maximum value of $V^\pi(s)$ for all $\pi \in \Delta$ [2]. Let $\tilde{\pi} \in \Delta$ be a strategy such that $V^{\tilde{\pi}}(s)$ is maximal in (5.1) for all $s \in \mathcal{S}$. $\tilde{\pi}$ is known as an *optimal strategy* for the MDP. Again we use a subscript n on $\pi(s,a)$, $\pi(s)$ and π to represent the strategy of the player at iteration n.

To reduce the complexity of the notation we assume that every state has the same number of actions which is denoted by $\Sigma := |\mathcal{A}(s)|$ for any $s \in \mathcal{S}$ and we let $m := K\Sigma$. Note that having a different number of actions in each state does not affect the validity of this approach but does make the notation more cumbersome.

We begin by placing restrictions on the learning rates and the MDP. In this algorithm we select learning rates $\{\alpha(n)\}_{n\in\mathbb{N}}$ and $\{\gamma(n)\}_{n\in\mathbb{N}}$ which satisfy (B2) and (B5)(a) and use two asynchronous counters,

$$\nu_n(s) := \sum_{i=1}^{n} \mathbb{I}_{\{s_i = s\}}, \quad \phi_n(s,a) := \sum_{i=1}^{n} \mathbb{I}_{\{(s_i,a_i)=(s,a)\}}.$$

In addition, we assume that for every $s \in \mathcal{S}$ there exists an $a(s) \in \mathcal{A}(s)$ such that the set of transition probabilities, $\{P_{ss'}(a(s))\}_{s,s'}$, form an aperiodic, irreducible Markov chain. Moreover, at every iteration when the MDP is in the state s we enforce that every action $a \in \mathcal{A}(s)$ is played with a non-zero minimum probability and that this holds for every $s \in \mathcal{S}$. Therefore for all $s \in \mathcal{S}$, $a \in \mathcal{A}(s)$ and $n \geq 0$, $\pi_n(s,a) \geq \varepsilon$ for some $\varepsilon > 0$.

Finally, before directly analysing the algorithm we present a method for verifying the global convergence of a standard differential inclusion in the form of (2.1).

Definition 5.1. Let $A \subset \mathbb{R}^K$. A continuous function $W : \mathbb{R}^K \mapsto \mathbb{R}$ is a *Lyapunov function* for the differential inclusion (2.1) if it satisfies the following criteria.

(i) $W(y) < W(x)$ for all $x \in \mathbb{R}^K \backslash A$, $y \in \Phi_t(x)$, $t > 0$.
(ii) $W(y) \leq W(x)$ for all $x \in A$, $y \in \Phi_t(x)$, $t > 0$.

Finding a Lyapunov function proves the global convergence of set-valued dynamical systems in the form of (2.1), a concept which is fully described by Benaïm, Hofbauer and Sorin [5].

Downloaded from informs.org by [216.73.217.172] on 15 November 2026, at 06:26 . For personal use only, all rights reserved.

Because we include the asynchronicity with the mean field the associated differential inclusions associated with our algorithm will be in the form,

$$\dot{x} \in \Omega_k^\delta \cdot \Big[h(x) - x\Big], \tag{5.2}$$

where $h(\cdot)$ is set-valued. Let $h'(x) := \Omega_k^\delta \cdot [h(x) - x]$. We state a slight modification to a result of Benaïm, Hofbauer and Sorin [6] to allow us to easily prove the convergence of differential inclusions in the form of (5.2). We do not state the proof of this as it is a straightforward extensions to [6, Theorem 3.10]. Let $\nabla_x W(x)$ denote the partial derivative of $W(x)$ with respect to x.

THEOREM 5.2. *Assuming that $h'(x)$ is a Marchaud map and there exists a positive definite function $W \in C^1(\mathbb{R}^k, \mathbb{R})$ such that if $A = \{W(x) = 0; x \in C\}$ and $x \in C \backslash A$ then for any $\omega \in \Omega_k^\delta$, $x' \in h(x)$,*

$$\big\langle \nabla_x W(x), \omega(x' - x) \big\rangle \leq 0.$$

Then $W(\cdot)$ is a Lyapunov function for (5.2) with attracting set A.

Using Theorem 5.2 we show the global convergence of the following algorithm.

5.1. *The algorithm.* The algorithm we present in (5.3) and (5.4) cannot be studied in the framework of Konda and Borkar [14] due to the Lipschitz continuous restriction they place on the mean fields of the coupled stochastic approximations. In this work we have relaxed this condition allowing the study of processes which are based on the best response. Firstly, if $\{Q(s,a)\}_{s,a}$ is the set of action values for a MDP, let $b_s(Q) := \arg\max_{a \in \mathcal{A}(s)} \{Q(s,a)\}$ be the best response set to $\{Q(s,a)\}_{s,a}$ for state $s \in \mathcal{S}$. Use the following coupled algorithm to estimate action values and the optimal strategy for all $s \in \mathcal{S}$ and $a \in \mathcal{A}(s)$,

$$\begin{aligned} Q_{n+1}(s,a) = Q_n(s,a) &+ \gamma\big(\phi_{n+1}(s,a)\big)\mathbb{I}_{\{(s,a)=(s_{n+1},a_{n+1})\}} \\ &\times \Big[R_{n+1} + \beta V_n(s_{n+2}) - Q_n(s,a)\Big], \end{aligned} \tag{5.3}$$

$$\pi_{n+1}(s) = \pi_n(s) + \mu(\nu_{n+1}(s))\mathbb{I}_{\{s=s_{n+1}\}}\Big[b_s(Q_{n+1}) - \pi_n(s)\Big], \tag{5.4}$$

where $V_n(s) = \sum_{a \in \mathcal{A}(s)} \pi_n(s,a) Q_n(s,a)$. The action a_{n+1} is selected using an ε-greedy version of the strategy π_n. For all $s \in \mathcal{S}$ and $a \in \mathcal{A}(s)$ let

$$\pi_n^\varepsilon(s,a) := \pi_n(s,a)(1 - \Sigma\varepsilon) + \varepsilon.$$

Downloaded from informs.org by [216.73.217.172] on 15 November 2026, at 06:26 . For personal use only, all rights reserved.

Then $\mathbb{P}(a_{n+1} = a | s_{n+1} = s, \mathcal{F}_n) = \pi_n^\varepsilon(s,a)$. We must verify that (B1)-(B6) hold for this algorithm. We do not directly verify that (B1) holds for this algorithm, but as pointed out in Section 3.2 methods to do so are discussed elsewhere. Furthermore the choice of learning parameters verifies (B2) and the choice of mean field in (5.3) and (5.4) immediately give that (B3) and (B5) hold.

For this algorithm we have $\bar{J} = \{(s,a); s \in \mathcal{S}, a \in \mathcal{A}(s)\}$ and $\bar{I} = \{s; s \in \mathcal{S}\}$. This gives that $\bar{H} = \big\{\big((s,a),s\big); s \in \mathcal{S}, a \in \mathcal{A}(s)\big\}$ and for simplicity we write,

$$\bar{H} = \big\{(s,a); s \in \mathcal{S}, a \in \mathcal{A}(s)\big\}.$$

Following the notation of Section 4 we have that $z_n = (Q_n, \pi_n)$. With $\mathcal{H}_n, \mathcal{H}_{n+1} \in \bar{H}$ such that $\mathcal{H}_n = (s,a)$, $\mathcal{H}_{n+1} = (s',a')$, then $Q_{(\mathcal{H}_n,\mathcal{H}_{n+1})}(z) = \pi_n^\varepsilon(s',a')P_{ss'}(a)$. This shows that (B4)(a) is satisfied. Again using the notation of Section 4 the set of transition probabilities are denoted,

$$\Big\{Q_{(\mathcal{H}_n,\mathcal{H}_{n+1})}(z_n)\Big\}_{\mathcal{H}_n,\mathcal{H}_{n+1},z_n} = \Big\{\pi_n^\varepsilon(s',a')P_{ss'}(a)\Big\}_{s,s',a,a'}.$$

By assumption on $\{P_{ss'}(a)\}_{s,s',a}$ we have that (B4)(b) is satisfied. Since $\pi_n^\varepsilon(s',a')P_{ss'}(a)$ is a continuous function of π_n^ε, which similarly is a continuous function of $\pi_n \in z_n$, (B4)(c) is satisfied.

A consequence of (B4) from Appendix A.1 is that in the limit every state of the MDP is visited a minimum proportion of the time, η, for some $\eta > 0$. Similarly, by placing the restriction that every action is selected with at least probability ε for some $\varepsilon > 0$ then every state-action pair is used a minimum proportion of the time, η', for some $\eta' > 0$. Using the approach of Section 4 we do not explicitly need to know the values η and η' as we verify convergence for every $\eta, \eta' > 0$.

Finally, we need to verify (B6). Define

$$Q^\pi(s,a) := r(s,a) + \beta \sum_{s' \in \mathcal{S}} P_{ss'}(a) V^\pi(s),$$

and let $Q^\pi(s)$ be the Σ-vector containing $Q^\pi(s,a)$ for all $a \in \mathcal{A}(s)$. Let $h(\cdot,\cdot) : \Delta \times \mathbb{R}^m \to \mathbb{R}^m$ be defined such that,

$$h_{s,a}(\pi, Q) := r(s,a) + \beta \sum_{s' \in \mathcal{S}} P_{ss'}(a) V(s'),$$

where $V(s) = \sum_{a \in \mathcal{A}(s)} \pi(s,a) Q(s,a)$. Let $h_s(\pi, Q)$ be the Σ-vector of the $h_{s,a}(\pi, Q)$ terms which means that $h(\pi, Q^\pi) = Q^\pi$. For fixed $\pi \in \Delta$ consider the differential inclusion

$$\dot{Q}_s(t) \in \Omega_\Sigma^{\eta'} \cdot \Big[h_s\big(\pi, Q_s(t)\big) - Q_s(t)\Big], \quad \text{for all } s \in \mathcal{S}, \tag{5.5}$$

LEMMA 5.3. *$Q^{\pi}(s)$ is the unique global attractor of (5.5).*

PROOF. $h_s(\pi, Q_s(t))$ is a contraction mapping [23, 24]. Hence for any fixed $\omega \in \Omega_{\Sigma}^{\eta'}$, $Q_s(t) \to Q^{\pi}(s)$. Combining this with the note by Borkar [11, Chapter 7.4] proves the claim. □

From this it follows that the values in $\{Q_n(s,a)\}_{s,a}$ converge to the true action values for the strategy π. These values are Lipschitz continuous in π, which ensures (B6) holds. Hence, Theorem 4.7 holds and the linear interpolation of the iterative process in (5.4) is an asymptotic pseudo-trajectory to the differential inclusion

$$\dot{\pi}_s(t) \in \Omega^{\eta} \cdot \left[b_s(Q^{\pi(t)}) - \pi_s(t) \right], \quad \text{for all } s \in \mathcal{S}, \tag{5.6}$$

for some $\eta > 0$. For a particular action $a \in \mathcal{A}(s)$, let $\pi_{s,a}(t)$ and $\dot{\pi}_{s,a}(t)$ represent the individual components of $\pi_s(t)$ and $\dot{\pi}_s(t)$ respectively, whilst $\pi(t)$ and $\dot{\pi}(t)$ are the $K \times \Sigma$ matrices containing all of the $\pi_{s,a}(t)$ and $\dot{\pi}_{s,a}(t)$ elements.

With assumptions (B1)-(B6) satisfied all that remains is to show that the differential inclusion (5.6) has a global attractor. Corollary 4.8 will then provide the convergence result of the coupled algorithm in (5.3) and (5.4). We note that (5.6) is in the form of (5.2) and hence we use Theorem 5.2 to prove the global convergence. Recall $\nabla_{\pi} V(s)$ denotes taking the partial derivative of $V(s)$ with respect to π.

LEMMA 5.4. *For every $s \in \mathcal{S}$, any fixed $\omega_s \in \Omega^{\eta}$, $\underline{\pi} \in \Delta$ and $\underline{\tilde{\pi}}_s \in b_s(Q^{\underline{\pi}})$, let $\underline{\dot{\pi}}_s := \omega_s \left[\underline{\tilde{\pi}}_s - \underline{\pi}_s \right]$ and $\underline{\dot{\pi}}$ be the vector of all $\underline{\dot{\pi}}_s$ terms. Then for every $s \in \mathcal{S}$,*

$$\left\langle \nabla_{\underline{\pi}} V^{\underline{\pi}}(s), \underline{\dot{\pi}} \right\rangle \geq 0.$$

PROOF. For all $s \in \mathcal{S}$, $a \in \mathcal{A}(s)$ let

$$K_{s,a}(\pi) := r(s,a) + \beta \sum_{s' \in S} P_{ss'}(a) V^{\pi}(s') - V^{\pi}(s),$$

and $K_s(\pi)$ be the Σ-vector of these terms for all $a \in \mathcal{A}(s)$. Consider,

$$\begin{aligned} \langle \underline{\dot{\pi}}_s, K_s(\underline{\pi}) \rangle = \omega_s & \left[\sum_{a \in \mathcal{A}(s)} \underline{\tilde{\pi}}_s(a) Q^{\underline{\pi}}(s,a) - \sum_{a \in \mathcal{A}(s)} \underline{\pi}_s(a) Q^{\underline{\pi}}(s,a) \right] \\ & - \omega_s \left[\sum_{a \in \mathcal{A}(s)} \underline{\tilde{\pi}}_s(a) V^{\underline{\pi}}(s) - \sum_{a \in \mathcal{A}(s)} \underline{\pi}_s(a) V^{\underline{\pi}}(s) \right]. \end{aligned}$$

Downloaded from informs.org by [216.73.217.172] on 15 November 2026, at 06:26 . For personal use only, all rights reserved.

The second term here is zero since $\sum_{a\in\mathcal{A}(s)}\rho(s,a)V^{\pi}(s) = V^{\pi}(s)$ for any $\rho(s)\in\Delta(\mathcal{A}(s))$. The first term is clearly positive by the definition of the best response. Hence

$$\langle \dot{\underline{\pi}}_s, K_s(\underline{\pi})\rangle \geq 0. \tag{5.7}$$

A slight modification to a result of Konda and Borkar [14, Lemma 5.4] states that for a vector field on Δ, G, which depends on π if

$$\langle G_s(\pi), K_s(\pi)\rangle \geq 0,$$

then

$$\langle \nabla_{\pi}V^{\pi}(s), G(\pi)\rangle \geq \langle G_s(\pi), K_s(\pi)\rangle \geq 0.$$

Combining this with (5.7) with $G_s(\pi) = \dot{\underline{\pi}}_s$ concludes the proof. □

Now we use Theorem 5.2 and Lemma 5.4 to produce a Lyapunov function for (5.6) and hence prove the global convergence of the algorithm given by (5.3) and (5.4). Let $A := \{\pi; \pi \text{ an optimal strategy}\}$.

LEMMA 5.5. *Fix $\tilde{\pi}$ as an optimal strategy. Then*

$$W(\pi) = \sum_{s\in\mathcal{S}}\left[V^{\tilde{\pi}}(s) - V^{\pi}(s)\right],$$

is a Lyapunov function for the differential inclusion (5.6) and A is a global attractor for (5.6).

PROOF. We prove the claim by applying Theorem 5.2 to (5.6). Clearly $W(\cdot)$ is positive semi-definite since for all $s\in S$, and any $\pi\in\Delta$, $V^{\tilde{\pi}}(s)\geq V^{\pi}(s)$, with equality for $\pi\in A$.

To prove the condition of Theorem 5.2 we note that for any fixed $t>0$,

$$\left\langle \nabla_{\pi(t)}W\big(\pi(t)\big), \dot{\pi}(t)\right\rangle = -\sum_{s\in\mathcal{S}}\left\langle \nabla_{\underline{\pi}}V^{\underline{\pi}}(s), \dot{\underline{\pi}}\right\rangle,$$

for some strategy $\underline{\pi}\in\Delta$, and for each $s\in\mathcal{S}$, a fixed $\omega_s\in\Omega^{\eta}$ and $\tilde{\underline{\pi}}_s\in b_s(Q^{\underline{\pi}})$ such that for all $s\in\mathcal{S}$,

$$\dot{\underline{\pi}}_s = \omega_s\left[\tilde{\underline{\pi}}_s - \underline{\pi}_s\right].$$

Using Lemma 5.4 immediately gives that $\left\langle \nabla_{\pi(t)}W\big(\pi(t)\big), \dot{\pi}(t)\right\rangle < 0$ for all $\pi(t)\in\Delta_k\backslash A$ and for any $t>0$. Applying Theorem 5.2 proves the claim. □

COROLLARY 5.6. *The coupled process (Q_n,π_n) from (5.3) and (5.4) converges to the limit $(Q^{\tilde{\pi}},\tilde{\pi})$, where $\tilde{\pi}$ is an optimal strategy and $\{Q^{\tilde{\pi}}(s,a)\}_{s,a}$ is the set of associated action values.*

Downloaded from informs.org by [216.73.217.172] on 15 November 2026, at 06:26 . For personal use only, all rights reserved.

PROOF. Lemma 5.5 shows that a Lyapunov function exists for the differential inclusion (5.6); this with Corollary 4.8 proves the claim. □

6. Summary. We have combined the work of Benaïm, Hofbauer and Sorin [5] on differential inclusions with the work of Konda and Borkar [14] and Borkar [9] on asynchronous stochastic approximation in order to provide a framework for asynchronous stochastic approximation with a set-valued mean field. This enables us to modify the previous work on asynchronous stochastic approximation to use a set of assumptions which are straightforward to verify *a priori*.

Furthermore we extended the work of Konda and Borkar on asynchronous two-timescale stochastic approximation using this new framework. By allowing the mean fields to be updated using set-valued functions we provide a new result in two-timescale stochastic approximations which clearly applies to asynchronous and synchronous stochastic approximations.

This approach provides a clear framework for single or multiple timescale asynchronous stochastic approximations with clear assumptions which differ little from those used in regular stochastic approximation. Where previously the additional and difficult to verify assumptions required for asynchronous stochastic approximations could be perceived as a reason to avoid their use, this framework removes many of these issues.

In Section 5 we provided an example of a coupled learning algorithm for a discounted reward Markov decision process. We analysed the limiting behaviour using the results of Section 4. The algorithm uses a set-valued mean field based upon a best response style of actor-critic learning. We used the results of Section 4 to show convergence to an optimal strategy under a straightforward set of assumptions. This algorithm demonstrates the main value of the approach to asynchronous stochastic approximation in this paper.

APPENDIX A: POSTPONED PROOFS

A.1. Minimum update proportion. For the asynchronous stochastic approximations in (2.5) we are interested in understanding how different components of x_n get selected to be updated. The previous work on this makes the direct assumption that in the limit all the elements of $\bar{I}$ are updated in an equally spaced manner in the limit and some minimum proportion of the iterations [14]; this assumption is difficult to verify prior to running the process. In this work we use results on Markov chains via assumption (A4) as an alternative which can be checked *a priori* if we know the transition probabilities of the state process.

Downloaded from informs.org by [216.73.217.172] on 15 November 2026, at 06:26 . For personal use only, all rights reserved.

Downloaded from informs.org by [216.73.217.172] on 15 November 2026, at 06:26 . For personal use only, all rights reserved.

LEMMA A.1. *Under (A1)(a) and (A4), there exists* $\eta > 0$ *such that* $\forall i \in I$,

$$\liminf_{n\to\infty} \frac{\nu_n(i)}{n} \geq \eta, \quad a.s..$$

PROOF. The values of $\bar{I}_n \in \bar{I}$ form a controlled Markov chain where $\bar{I}_{n+1}$ depends on the current updated component, $\bar{I}_n$, and the value of the iterative process, x_n. For $x \in C$ let $\pi_x(\cdot)$ be a stationary distribution for this Markov chain given by the transition probabilities $P_x(\cdot,\cdot)$; standard theory gives that, under (A4), a unique $\pi_x(\cdot)$ exists which is continuous as a function of x, and for some $\delta_x > 0$, $\pi_x(\mathcal{I}) > \delta_x$ for all $\mathcal{I} \in \bar{I}$. Let $\eta = \min_{x\in C} \delta_x$, which exists and is positive since C is compact. Then for all $\mathcal{I} \in \bar{I}$, $x \in C$ we have that

$$\pi_x(\mathcal{I}) > \delta_x \geq \eta. \tag{A.1}$$

For $\mathcal{I} \in \bar{I}$ define

$$w_n(\mathcal{I}) = \sum_{k=1}^{n} \mathbb{I}_{\{\bar{I}_k = \mathcal{I}\}},$$

And let $\bar{w}(\mathcal{I}) = w_n(\mathcal{I})/n$.

$$\begin{aligned} \bar{w}_n(\mathcal{I}) &= \frac{w_{n-1}(\mathcal{I})}{n} + \frac{1}{n}\mathbb{I}_{\{\bar{I}_n=\mathcal{I}\}} \\ &= \bar{w}_{n-1}(\mathcal{I}) + \frac{1}{n}\Big(\mathbb{I}_{\{\bar{I}_n=\mathcal{I}\}} - \bar{w}_{n-1}(\mathcal{I})\Big) \end{aligned} \tag{A.2}$$

This is in the form of a stochastic approximation with controlled Markovian noise as in [10].

Let $\bar{w}(t)$ be the linear interpolation of the $\{\bar{w}_n\}_{n\in\mathbb{N}}$ process. Using [10, Corollary 3.1] the limiting behaviour of the interpolated process, $\bar{w}(t)$, will be an asymptotic pseudo-trajectory to a differential equation,

$$\frac{dw(t)}{dt} = \pi_{x(t)} - w(t), \tag{A.3}$$

for a suitable process $x(t) \in C$ based upon $\{x_n\}_{n\in\mathbb{N}}$. We know that $\pi_x(\mathcal{I}) \geq \eta > 0$ for all n and hence the dynamics of $w(t)$ can be expressed as,

$$\frac{dw(t)}{dt} \in \Omega^{\eta}_{|\bar{I}|} - w(t),$$

where $\Omega^{\eta}_{|\bar{I}|}$ is as defined in (3.4). This implies that any limit point of $w(t)$ will be in $\Omega^{\eta}_{|\bar{I}|}$ and hence $\liminf_{n\to\infty} \frac{w_n(\mathcal{I})}{n} \geq \eta$ a.s. $\forall \mathcal{I} \in \bar{I}$.

For $i \in I$, define $\mathcal{I}(i) := \{\mathcal{I} \in \bar{I}; i \in \mathcal{I}\}$ then $\nu_n(i) = \sum_{\mathcal{I}\in\mathcal{I}(i)} w_n(\mathcal{I})$, and so for some $\mathcal{I} \in \bar{I}$,

$$\lim \inf_{n\to\infty} \frac{\nu_n(i)}{n} \geq \lim \inf_{n\to\infty} \frac{w_n(\mathcal{I})}{n} \geq \eta, \quad \text{a.s..} \qquad \square$$

A.2. Noise conditions for an asynchronous stochastic approximation. Let $\tau_0 = 0$, $\tau_n = \sum_{k=1}^n \alpha(k)$ and recall $\bar{\tau}_0 = 0$, $\bar{\tau}_n = \sum_{k=1}^n \bar{\alpha}_k$. Denote the *asynchronous noise* term $\hat{V}_n(i) = \frac{\alpha(\nu_n(i))}{\alpha(n)} \mathbb{I}_{\{i\in\bar{I}_n\}} V_n(i)$, and let $\hat{V}_n$ be the K-vector of these terms.

LEMMA A.2. *If $\{V_n\}_{n\in\mathbb{N}}$ is a martingale difference noise process and assuming (A2)(b) and (A4) hold;*

(i) If $\sup_n \mathbb{E}\Big[\|V_n\|^q\Big] < \infty$ for some $q > 0$ then almost surely,

$$\sup_n \mathbb{E}\Big[\|\hat{V}_n\|^q\Big] < \infty.$$

(ii) Suppose $V_{n+1}(i)$ is uncorrelated with $\bar{I}_{n+1}$ given $\mathcal{F}_n$ and $V_{n+1}(i)$ is independent of $V_{n+1}(j)$ for $i \neq j$. Let $\langle a, b\rangle = \sum_k a_k b_k$. If there exists a $\Gamma > 0$ such that for all $\theta \in \mathbb{R}^K$,

$$\mathbb{E}\Big[\exp\Big\{\langle\theta, V_{n+1}\rangle\Big\}|\mathcal{F}_n\Big] \leq \exp\Big\{\frac{\Gamma}{2}\|\theta\|^2\Big\}$$

then almost surely there exists a $\bar{\Gamma} > 0$ such that for all $\theta \in \mathbb{R}^K$,

$$\mathbb{E}\Big[\exp\Big\{\langle\theta, \hat{V}_{n+1}\rangle\Big\}|\mathcal{F}_n\Big] \leq \exp\Big\{\frac{\bar{\Gamma}}{2}\|\theta\|^2\Big\}.$$

PROOF. (i) Combine η from Lemma A.1 with (A2)(b) to give $A_\eta \geq 1$. Then $\|\hat{V}_n(i)\| \leq \|A_\eta V_n(i)\|$ using (A2)(b). From this it is immediate that

$$\mathbb{E}\Big[\|\hat{V}_n(j)\|^q\Big] \leq A_\eta^q \mathbb{E}\Big[\|V_n\|^q\Big] < \infty.$$

(ii) Let $\check{V}_n(i) := \mathbb{I}_{\{i\in\bar{I}_n\}} V_n(i)$. Clearly $\|\check{V}_n\| \leq \|V_n\|$. In addition let e_i be a K dimensional basis vector with a 1 in the i^{th} term and 0 everywhere else. From the assumption in Lemma A.2 that $V_{n+1}(i)$ is uncorrelated with $\bar{I}_{n+1}$ given $\mathcal{F}_n$,

$$\mathbb{E}\Big[\exp\Big\{\big\langle\theta_i e_i, \check{V}_{n+1}\big\rangle\Big\}|\mathcal{F}_n\Big] \leq \exp\Big\{\frac{\Gamma}{2}\theta_i^2\Big\}.$$

Downloaded from informs.org by [216.73.217.172] on 15 November 2026, at 06:26 . For personal use only, all rights reserved.

Downloaded from informs.org by [216.73.217.172] on 15 November 2026, at 06:26 . For personal use only, all rights reserved.

Using this gives the following,

$$\begin{aligned}\mathbb{E}\Big[\exp\Big\{\big\langle\theta,\check{V}_{n+1}\big\rangle\Big\}|\mathcal{F}_n\Big] &= \mathbb{E}\Big[\exp\Big\{\big\langle\sum_{i=1}^{K}\theta_i e_i,\check{V}_{n+1}\big\rangle\Big\}|\mathcal{F}_n\Big],\\ &= \prod_{i=1}^{K}\mathbb{E}\Big[\exp\Big\{\big\langle\theta_i e_i,\check{V}_{n+1}\big\rangle\Big\}|\mathcal{F}_n\Big],\\ &\le \prod_{i=1}^{K}\exp\Big\{\frac{\Gamma}{2}\theta_i^2\Big\},\\ &= \exp\Big\{\frac{\Gamma}{2}\|\theta\|^2\Big\}.\end{aligned}$$

Notice that $\frac{\alpha(\nu_{n+1}(i))}{\alpha(n+1)}\mathbb{I}_{\{i\in\bar{I}_{n+1}\}} = \frac{\alpha(\nu_n(i)+1)}{\alpha(n+1)}\mathbb{I}_{\{i\in\bar{I}_{n+1}\}}$ and $\frac{\alpha(\nu_n(i)+1)}{\alpha(n+1)}$ is deterministic given $\mathcal{F}_n$.

$$\begin{aligned}\mathbb{E}\Big[\exp\Big\{\big\langle\theta,\hat{V}_{n+1}\big\rangle\Big\}|\mathcal{F}_n\Big] &= \mathbb{E}\Big[\exp\Big\{\big\langle\theta,\sum_{i=1}^{K}\frac{\alpha(\nu_{n+1}(i))}{\alpha(n+1)}\check{V}_{n+1}(i)e_i\big\rangle\Big\}|\mathcal{F}_n\Big],\\ &= \mathbb{E}\Big[\prod_{i=1}^{K}\exp\Big\{\frac{\alpha(\nu_n(i)+1)}{\alpha(n+1)}\big\langle\theta,\check{V}_{n+1}(i)e_i\big\rangle\Big\}|\mathcal{F}_n\Big],\end{aligned}$$

Now using the independence of the individual V_{n+1} terms and letting $\bar{\theta}_n(i) := \frac{\alpha(\nu_n(i)+1)}{\alpha(n+1)}\theta$.

$$\begin{aligned}\mathbb{E}\Big[\exp\Big\{\big\langle\theta,\hat{V}_{n+1}\big\rangle\Big\}|\mathcal{F}_n\Big] &= \prod_{i=1}^{K}\mathbb{E}\Big[\exp\Big\{\big\langle\bar{\theta}_n(i),\check{V}_{n+1}(i)e_i\big\rangle\Big\}|\mathcal{F}_n\Big],\\ &= \prod_{i=1}^{K}\mathbb{E}\Big[\exp\Big\{\big\langle e_i\bar{\theta}_n(i),\check{V}_{n+1}\big\rangle\Big\}|\mathcal{F}_n\Big].\end{aligned}$$

Note that $\|\bar{\theta}_n(i)\|^2 \le \|A_\eta\theta\|^2$, with A_η taken from (A2)(b). Finally, this gives,

$$\begin{aligned}\mathbb{E}\Big[\exp\Big\{\big\langle\theta,\hat{V}_{n+1}\big\rangle\Big\}|\mathcal{F}_n\Big] &\le \prod_{i=1}^{K}\Big(\exp\Big\{\frac{\Gamma}{2}\|\bar{\theta}_n(i)\|^2\Big\}\Big),\\ &\le \prod_{i=1}^{K}\Big(\exp\Big\{\frac{\Gamma A_\eta^2}{2}\|\theta\|^2\Big\}\Big).\end{aligned}$$

Letting $\bar{\Gamma} := K\Gamma A_\eta^2$, which is constant, completes the proof of (ii).

□

Downloaded from informs.org by [216.73.217.172] on 15 November 2026, at 06:26 . For personal use only, all rights reserved.

PROOF OF LEMMA 3.3. Firstly, define $\xi_n(t) = \sup\{k \geq 0; t \geq \tau_{n+k}\}$ and let $\bar{\xi}_n(t) = \sup\{k \geq 0; t \geq \bar{\tau}_{n+k}\}$, and notice $\xi_n(t) \geq \bar{\xi}_n(t)$ since $\bar{\alpha}_n \geq \alpha(n)$ from (A2)(b). Therefore,

$$m(\tau_n + T) = n + \xi_n(T) \geq n + \bar{\xi}_n(T) = \bar{m}(\bar{\tau}_n + T).$$

Using this in the following

$$\begin{aligned}
&\sup_k \left\{ \left\| \sum_{j=n}^{k-1} \bar{\alpha}_{j+1} M_{j+1} V_{j+1} \right\|; k = n+1, \ldots, \bar{m}(\bar{\tau}_n + T) \right\}, \\
&= \sup_k \left\{ \left\| \sum_{j=n}^{k-1} \alpha(j+1) \hat{V}_{j+1} \right\|; k = n+1, \ldots, \bar{m}(\bar{\tau}_n + T) \right\}, \\
&\leq \sup_k \left\{ \left\| \sum_{j=n}^{k-1} \alpha(j+1) \hat{V}_{j+1} \right\|; k = n+1, \ldots, m(\tau_n + T) \right\}.
\end{aligned}$$

Now, using step sizes $\{\alpha(n)\}_{n\in\mathbb{N}}$ and martingale difference process $\{\hat{V}_n\}_{n\in\mathbb{N}}$, Lemma A.2 together with [5, Proposition 1.4] gives that this final term converges to zero as $n \to \infty$. □

A.3. Weak convergence of the relative step sizes. Firstly, we will need a result which follows from (A2)(b) and Lemma A.1,

LEMMA A.3. *Under assumptions (A1)(a), (A2)(b) and (A4)*

$$\liminf_{n\to\infty} \left(\frac{\alpha(n)}{\bar{\alpha}_n} \right) \geq A_\eta^{-1}, \quad a.s.. \tag{A.4}$$

PROOF.

$$\begin{aligned}
\liminf_{n\to\infty} \left(\frac{\alpha(n)}{\bar{\alpha}_n} \right) &= \liminf_{n\to\infty} \left(\frac{\alpha(n)}{\alpha(\nu_n(i))} \right), \\
&\geq \liminf_{n\to\infty} \left(\frac{\alpha(n)}{\alpha(n\eta)} \right), \quad \text{a.s.}, \\
&\geq A_\eta^{-1}.
\end{aligned}$$

Where the first step must hold for some $i \in I$, the second step follows from Lemma A.1 and the last step is directly from (A2)(b). □

Downloaded from informs.org by [216.73.217.172] on 15 November 2026, at 06:26 . For personal use only, all rights reserved.

PROOF OF LEMMA 3.4. Since $L^2([0,T])$ is a Hilbert space it is relatively compact and relatively sequentially compact using the Banach-Alaoglu Theorem [11, Appendix A], which guarantees that the sequence $\{\bar{u}_i^n(\cdot)\}_{n\in\mathbb{N}}$ has a weakly convergent subsequence with a limit point in $L^2([0,T])$. Hence, there exists a limit point, $\tilde{u}_i(\cdot)$, of this subsequence of $\{\bar{u}_i^n(\cdot)\}_{n\in\mathbb{N}}$. It remains to show that this limit point does satisfy (3.10). For a fixed $T>0$, $\tilde{u}_i(\cdot)$ must satisfy (3.9) for all $h(\cdot)\in L^2([0,T])$. Hence by showing that for an arbitrary fixed T and a single $h(\cdot)$ any limit point is bounded below this is enough to prove the claim. Fix $T>0$, select $0<v\leq T$, and take $h(t)=1$ for all $t\in[0,v]$ and zero otherwise. Let $\{k(n)\}_{n\in\mathbb{N}}$ be a subsequence of the natural numbers under which $\{\bar{u}_i^{k(n)}(\cdot)\}_{n\in\mathbb{N}}$ converges.

$$\begin{aligned}
\int_0^v \tilde{u}_i(s)\,\mathrm{d}s &= \lim_{n\to\infty}\int_0^v u_i(\bar{\tau}_{k(n)}+s)\,\mathrm{d}s\\
&= \lim_{n\to\infty}\int_{\bar{\tau}_{k(n)}}^{\bar{\tau}_{k(n)}+v} u_i(s)\,\mathrm{d}s\\
&\geq \lim_{n\to\infty}\sum_{j=k(n)}^{\bar{m}(\bar{\tau}_{k(n)}+v)-1}\bar{\alpha}_{j+1}\mu_{j+1}(i),\\
&= \lim_{n\to\infty}\sum_{j=k(n)}^{\bar{m}(\bar{\tau}_{k(n)}+v)-1}\alpha(j+1)\mathbb{I}_{\{i\in\bar{I}_{j+1}\}}.
\end{aligned}$$

The following part of the proof uses a slight modification to the result by Ma et al. [21, Theorem 2.2] combined with the stochastic approximation form of the updates w_n from Lemma A.1 given in (A.2). The modification comes because the transition probabilities for the Markov chain on $\bar{I}$ depends on x_n instead of w_n. This requires only a straightforward modification to the proofs in Sections 4 and 5 of [21]. Using this modification of [21, Theorem 2.2] gives,

$$\begin{aligned}
&\lim_{n\to\infty}\sum_{j=k(n)}^{\bar{m}(\bar{\tau}_{k(n)}+v)-1}\alpha(j+1)\mathbb{I}_{\{i\in\bar{I}_{j+1}\}}\\
&\qquad= \lim_{n\to\infty}\sum_{\mathcal{I}\in\mathcal{I}(i)}\sum_{j=k(n)}^{\bar{m}(\bar{\tau}_{k(n)}+v)-1}\alpha(j+1)\mathbb{I}_{\{\bar{I}_{j+1}=\mathcal{I}\}},\\
&\qquad= \lim_{n\to\infty}\sum_{\mathcal{I}\in\mathcal{I}(i)}\sum_{j=k(n)}^{\bar{m}(\bar{\tau}_{k(n)}+v)-1}\alpha(j+1)\pi_{x_j}(\mathcal{I}).
\end{aligned}$$

Downloaded from informs.org by [216.73.217.172] on 15 November 2026, at 06:26 . For personal use only, all rights reserved.

Now we combine the above with (A.1) to give

$$\lim_{n\to\infty}\sum_{\mathcal{I}\in\mathcal{I}(i)}\sum_{j=k(n)}^{\bar{m}(\bar{\tau}_{k(n)}+v)-1}\alpha(j+1)\pi_{x_j}(\mathcal{I})$$
$$\geq \lim_{n\to\infty}\sum_{\mathcal{I}\in\mathcal{I}(i)}\sum_{j=k(n)}^{\bar{m}(\bar{\tau}_{k(n)}+v)-1}\alpha(j+1)\eta,$$
$$\geq \lim_{n\to\infty}\sum_{j=k(n)}^{\bar{m}(\bar{\tau}_{k(n)}+v)-1}\bar{\alpha}_{j+1}\frac{\alpha(j+1)}{\bar{\alpha}_{j+1}}\eta.$$

Now using Lemma A.3,

$$\lim_{n\to\infty}\sum_{j=k(n)}^{\bar{m}(\bar{\tau}_{k(n)}+v)-1}\bar{\alpha}_{j+1}\frac{\alpha(j+1)}{\bar{\alpha}_{j+1}}\eta$$
$$\geq \lim_{n\to\infty}\sum_{j=k(n)}^{\bar{m}(\bar{\tau}_{k(n)}+v)-1}\bar{\alpha}_{j+1}A_\eta^{-1}\eta,\quad \text{a.s.},$$

We convert the sum back to an integral to give,

$$\lim_{n\to\infty}\sum_{j=k(n)}^{\bar{m}(\bar{\tau}_{k(n)}+v)-1}\bar{\alpha}_{j+1}A_\eta^{-1}\eta$$
$$\geq \lim_{n\to\infty}\int_0^v A_\eta^{-1}\eta\,\mathrm{d}s - \lim_{n\to\infty}\bar{\alpha}_{\bar{m}(\bar{\tau}_{k(n)}+v)+1}A_\eta^{-1}\eta,$$
$$= \int_0^v A_\eta^{-1}\eta\,\mathrm{d}s.$$

Taking $\varepsilon = A_\eta^{-1}\eta$ completes the proof. □

We shall now prove the important corollary to Lemma 3.4.

Proof of Corollary 3.5. Begin by noting that

$$\int_t^{t+v}\tilde{u}_i(s)\,\mathrm{d}s = \lim_{n\to\infty}\int_t^{t+v}u_i^{k(n)}(s)\,\mathrm{d}s,$$
$$= \lim_{n\to\infty}\int_{t+\bar{\tau}_{k(n)}}^{t+\bar{\tau}_{k(n)}+v}u_i(s)\,\mathrm{d}s.$$

Downloaded from informs.org by [216.73.217.172] on 15 November 2026, at 06:26 . For personal use only, all rights reserved.

Now let $\bar{n} = \bar{m}(t + \bar{\tau}_{k(n)})$ and note that $u_i(t) \in [0, 1]$.

$$\lim_{n\to\infty} \int_{t+\bar{\tau}_{k(n)}}^{t+\bar{\tau}_{k(n)}+v} u_i(s)\,\mathrm{d}s \geq \lim_{\bar{n}\to\infty} \left[\int_{\bar{\tau}_{\bar{n}}}^{\bar{\tau}_{\bar{n}}+v} u_i(s)\,\mathrm{d}s - \bar{\alpha}_{\bar{n}+1} \right],$$
$$= \lim_{\bar{n}\to\infty} \int_0^v u_i^{\bar{n}}(s)\,\mathrm{d}s.$$

Now, $\mathcal{U}$ is a compact metric space [11, Appendix A] and hence is weakly countably (limit point) compact. This means the infinite subset $\{u_i^{\bar{n}}(\cdot)\}_{n\in\mathbb{N}}$ has a limit in $\mathcal{U}$, and this limit point must satisfy Lemma 3.4. This gives,

$$\int_{\bar{t}}^{\bar{t}+v} \tilde{u}_i(s)\,\mathrm{d}s \geq \lim_{\bar{n}\to\infty} \int_0^v u_i^{\bar{n}}(s)\,\mathrm{d}s \geq v\varepsilon, \quad \text{a.s.},$$

and since this statement is true for all $t, v > 0$, the statement follows. □

A.4. Proof of Lemma 3.6. Extend our notation to continuous time by defining $M(t)$ as the $K \times K$ diagonal matrix of the $(u_1(t), \ldots, u_K(t))$ terms and recall that $\tilde{M}(t)$ is the $K \times K$ diagonal matrix of the $(v_1(t), \ldots, v_K(t))$ terms.

Let $h(\cdot)$ be a bounded continuous function on $[\bar{\tau}_n, \bar{\tau}_k]$ such that $h(\bar{\tau}_j) = f_j$ for all $j = n, \ldots, k$. This will mean that $h(\cdot) \in L^2([0, T])$. Throughout this paper we have considered the continuous interval $[0, \infty)$ divided into segments of length $\bar{\alpha}_n$, and hence we can approximate the sum from Lemma 3.6 as an integral,

$$\left\| \sum_{i=n}^{k-1} \bar{\alpha}_{i+1} f_i \big(M_{i+1} - \tilde{M}_{i+1}\big) \right\| = \left\| \int_{\bar{\tau}_n}^{\bar{\tau}_{k-1}} h(\bar{\tau}_{\bar{m}(t)}) \big(M_{\bar{m}(t)+1} - \tilde{M}_{\bar{m}(t)+1}\big)\,\mathrm{d}t \right\|.$$

Now we note that $u_i(t)$ and $v_i(t)$ are extensions of $\{\mu_n(i)\}_{n\in\mathbb{N}}$ to continuous time which are constant on intervals $[\bar{\tau}_n, \bar{\tau}_{n+1})$, and that $M(t)$ and $\tilde{M}(t)$ are just matrices containing $u_i(t)$ and $v_i(t)$ respectively. This will mean that $M(t) = M_{\bar{m}(t)+1}$ and $\tilde{M}(t) = \tilde{M}_{\bar{m}(t)+1}$ and hence we have,

$$\left\| \int_{\bar{\tau}_n}^{\bar{\tau}_{k-1}} h(\bar{\tau}_{\bar{m}(t)}) \big(M_{\bar{m}(t)+1} - \tilde{M}_{\bar{m}(t)+1}\big)\,\mathrm{d}t \right\|$$
$$= \left\| \int_{\bar{\tau}_n}^{\bar{\tau}_{k-1}} h(\bar{\tau}_{\bar{m}(t)}) \big(M(t) - \tilde{M}(t)\big)\,\mathrm{d}t \right\| \tag{A.5}$$

In order to prove the convergence to zero of this term we look to use (3.11), but a further expansion is required so that (A.5) is in the correct form.

$$\left\| \int_{\bar{\tau}_n}^{\bar{\tau}_{k-1}} h(\bar{\tau}_{\bar{m}(t)}) \big(M(t) - \tilde{M}(t) \big) \, \mathrm{d}t \right\|$$

$$\leq \left\| \int_{\bar{\tau}_n}^{\bar{\tau}_{k-1}} h(t) \big(M(t) - \tilde{M}(t) \big) \, \mathrm{d}t \right\| \tag{A.6}$$

$$+ \left\| \int_{\bar{\tau}_n}^{\bar{\tau}_{k-1}} M(t) \big[h(\bar{\tau}_{\bar{m}(t)}) - h(t) \big] \, \mathrm{d}t \right\| \tag{A.7}$$

$$+ \left\| \int_{\bar{\tau}_n}^{\bar{\tau}_{k-1}} \tilde{M}(t) \big[h(t) - h(\bar{\tau}_{\bar{m}(t)}) \big] \, \mathrm{d}t \right\|. \tag{A.8}$$

Now, to prove the claim of Lemma 3.6 we will show that each of these terms will converge to zero. Firstly, (A.6) converges to zero almost surely by (3.11).

Both (A.7) and (A.8) are dealt with using the same technique, which we demonstrate for (A.8).

$$\left\| \int_{\bar{\tau}_n}^{\bar{\tau}_{k-1}} \tilde{M}(t) \big[h(t) - h(\bar{\tau}_{\bar{m}(t)}) \big] \, \mathrm{d}t \right\|$$

$$\leq \left\| \int_{\bar{\tau}_n}^{\bar{\tau}_{k-1}} \tilde{M}(t) h(t) \, \mathrm{d}t - \sum_{i=n}^{k-1} \bar{\alpha}_{i+1} h(\bar{\tau}_i) \tilde{M}(\bar{\tau}_i) \right\|$$

$$+ \left\| \sum_{i=n}^{k-1} \bar{\alpha}_{i+1} h(\bar{\tau}_i) \tilde{M}(\bar{\tau}_i) - \int_{\bar{\tau}_n}^{\bar{\tau}_{k-1}} \tilde{M}(t) h(\bar{\tau}_{\bar{m}(t)}) \, \mathrm{d}t \right\|,$$

in both of these terms the sum is a Riemann approximation to the integral using the left hand points of the partition $[\bar{\tau}_n, \bar{\tau}_{n+1}]$ for all n. As $n \to \infty$ the width of the partition tends to zero and hence the above terms tend to zero. This proves that (A.7) and (A.8) will converge to zero. □

REFERENCES

[1] J. P. Aubin. *Viability Theory*. Springer, 2009.

[2] R. Bellman. On the theory of dynamic programming. *Proceedings of the National Academy of Sciences*, 38:716–719, 1952. MR0050856

[3] M. Benaïm. A dynamical system approach to stochastic approximations. *SIAM J. Control and Optimization*, 34:437–472, 1996. MR1377706

[4] M. Benaïm and M. W. Hirsch. Mixed equilibria and dynamical systems arising from fictitious play in perturbed games. *Games and Economic Behaviour*, 29:36–72, 1999. MR1729309

[5] M. Benaïm, J. Hofbauer, and S. Sorin. Stochastic approximations and differential inclusions. *SIAM J. on Control and Optimization*, 44:328–348, 2003. MR2177159

[6] M. Benaïm, J. Hofbauer, and S. Sorin. Stochastic approximations and differential inclusions, part II: Applications. *Mathematics of Operations Research*, 31:673–695, 2006. MR2281223

[7] P. Billingsley. *Convergence of probability measures*. John Wiley & Sons Inc., New York, second edition, 1999. MR1700749

Downloaded from informs.org by [216.73.217.172] on 15 November 2026, at 06:26 . For personal use only, all rights reserved.

[8] V. S. Borkar. Stochastic approximation with two time scales. *System & Control Letters*, 29:291–294, 1997. MR1432654

[9] V. S. Borkar. Asynchronous stochastic approximations. *SIAM J. on Control and Optimization*, 36:840–851, 1998. MR1613869

[10] V. S. Borkar. Stochastic approximation with 'controlled Markov' noise. *System & Control Letters*, 55:139–145, 2006. MR2187843

[11] V. S. Borkar. *Stochastic Approximation: A Dynamical Systems Viewpoint*. Cambridge University Press, 2008. MR2442439

[12] L. Debnath and P. Mikusiński. *Hilbert spaces with applications*. Elsevier Academic Press, 2005.

[13] S. Kamal. Stabilization of stochastic approximation by step size adaptation. Technical report, 2010.

[14] V. R. Konda and V. S. Borkar. Actor-critic-type learning algorithms for markov decision processes. *SIAM J. on Control and Optimization*, 38:94–123, 2000. MR1740605

[15] V. R. Konda and J. N. Tsitsiklis. On actor–critic algorithms. *SIAM J. on Control and Optimization*, 42:1143–1166, 2003. MR2044789

[16] H. J. Kushner and D. S. Clark. *Stochastic approximation methods for constrained and unconstrained systems*. Springer-Verlag, 1978. MR0499560

[17] H. J. Kushner and G. Yin. Asymptotic properties of distributed and communicating stochastic approximation algorithms. *SIAM J. on Control and Optimization*, 25:1266–1290, 1987. MR0905045

[18] H. J. Kushner and G. Yin. Stochastic approximation algorithms for parallel and distributed processing. *Stochastics*, 22:219–250, 1987. MR0918081

[19] D. S. Leslie and E. J. Collins. Convergent multiple-timescales reinforcement learning algorithms in normal form games. *Annals of Applied Probability*, 13:1231–1251, 2003. MR2023875

[20] L. Ljung. Analysis of recursive stochastic algorithms. *IEEE Transactions on Automatic Control*, 22:551–575, 1977. MR0465458

[21] D. J. Ma, A. M. Makowski, and A. Shwartz. Stochastic approximations for finite-state Markov chains. *Stochastic Processes and their Applications*, 35:27–45, 1990. MR1062581

[22] S. Singh, T. Jaakkola, M. L. Littman, and C. Szepesvári. Convergence results for single-step on-policy reinforcement-learning algorithms. *Machine Learning*, 38:287–308, 2000.

[23] C. Szepesvári and M. L. Littman. A unified analysis of value-function-based reinforcement learning algorithms. *Neural Computation*, 11:2017–2060, 1999.

[24] J. N. Tsitsiklis. Asynchronous stochastic approximation and Q-learning. *Machine Learning*, 16:185–202, 1994.

School of Mathematics
University of Bristol
University Walk
Clifton
Bristol
BS8 1TW
E-mail: Steven.Perkins.09@bristol.ac.uk
David.Leslie@bristol.ac.uk

Downloaded from informs.org by [216.73.217.172] on 15 November 2026, at 06:26 . For personal use only, all rights reserved.