Stochastic approximations of differential inclusions: almost sure boundedness and asymptotic convergence ⋆

Andrew R. Teel * **Rafal Goebel** ** **Ricardo G. Sanfelice** ***

* *UC Santa Barbara, USA (e-mail: teel@ucsb.edu)*
** *Loyola University Chicago, USA (e-mail: rgoebel1@luc.edu)*
*** *UC Santa Cruz, USA (e-mail: ricardo@ucsc.edu)*

Abstract: For a stochastic approximation of a differential inclusion, results are given on almost sure boundedness of the approximate solutions and on their convergence to the chain recurrent part of the global attractor. The former involve bounding the stochastic approximation's variance using a Lyapunov-like function. The novelty of the latter is in the use of Lyapunov-like characterizations of Morse decompositions of the attractor.

Keywords: Differential inclusions, stochastic approximation, Lyapunov methods, martingale methods, difference inclusions.

1. INTRODUCTION

This work analyzes the behavior of a stochastic approximation of a differential inclusion. The stochastic approximation is similar to the differential inclusion's Euler approximation with nonsummable, but square summable, step sizes. The mapping used in the approximation is a stochastic version of the differential inclusion's right-hand side. Namely, the expected values of its measurable selections belong to the right-hand side of the differential inclusion.

We give conditions on the variance of the stochastic approximation to guarantee 1) almost sure boundedness of every solution of the stochastic approximation, and 2) almost sure convergence to the chain recurrent part of the global attractor for the differential inclusion when a global attractor exists. The proofs are modular, exploiting Lyapunov functions and the main result of Robbins and Siegmund (1971), as in Vidyasagar (2023), and a converse Lyapunov theorem for differential inclusions and Morse decompositions by Li and Wang (2012), and basic facts that relate Morse decompositions of an attractor to the chain recurrent part of that attractor, as in Conley (1978), (Li, 2007, Theorem 6.12(1)), and (Goebel, 2023, Proposition 11(a)).

Stochastic approximations of differential inclusions, under various assumptions, have been considered previously in the literature, some of which is summarized in (Borkar, 2023, Chapter 5). Seminal results for general differential inclusions satisfying basic conditions appear in Benaïm et al. (2005). In that work, almost sure boundedness of the solutions to the stochastic approximation is assumed and convergence results are established, including convergence to the chain recurrent part of the differential inclusion's global attractor. The approach taken is to establish that the interpolation of the values of a solution of the stochastic approximation corresponds to an approximate solution of the differential inclusion; i.e., the "ODE method" of stochastic approximation is extended to differential inclusions. Nguyen and Yin (2023) also assume almost sure boundedness and then establish convergence by taking an approach that is similar to Benaïm et al. (2005) under more general assumptions. The work by Faure and Roth (2010) takes an approach similar to that in Benaïm et al. (2005) and provides conditions for convergence with positive probability. Results also exist for differential inclusions corresponding to generalized gradient flow for Lipschitz cost functions satisfying certain regularity properties; see Davis et al. (2020) for example.

Regarding almost sure boundedness of the solutions to a stochastic approximation of a differential inclusion, Ramaswamy and Bhatnagar (2017) take an approach using a scaling limit, as originally pursued by Borkar and Meyn (2000) in the context of stochastic approximations of differential equations. In Ramaswamy and Bhatnagar (2022), almost sure boundedness is established via an approach that relies on Lyapunov functions, though in a manner that is different from what it proposed here. In both Ramaswamy and Bhatnagar (2017) and Ramaswamy and Bhatnagar (2022), the results of Benaïm et al. (2005) are used for convergence conclusions.

Our approach is inspired by techniques for stochastic approximations of Lipschitz differential equations in Vidyasagar (2023). In particular, we follow the example of Vidyasagar (2023) to decouple boundedness conditions from convergence conditions, we exploit Lyapunov-like functions, which here are continuously differentiable with a globally Lipschitz gradient, and we rely on a fundamental lemma from Robbins and Siegmund (1971) for the analy-

⋆ Research supported in part by NSF Grants no. CNS-2039054 and CNS-2111688, by AFOSR Grants nos. FA9550-23-1-0145, FA9550-23-1-0313, and FA9550-23-1-0678, by AFRL Grant nos. FA8651-22-1-0017 and FA8651-23-1-0004, by ARO Grant no. W911NF-20-1-0253, and by DoD Grant no. W911NF-23-1-0158.

sis. Lyapunov functions with bounded Hessians are used in Vidyasagar (2023) and also in Kamal (2010) and Kamal (2012) (see also (Borkar, 2023, Sections 4.4 and 4.6)), the latter in the context of permitting stochastic step sizes for stochastic approximation of differential equations.

Compared to Vidyasagar (2023), our almost sure boundedness results apply to differential inclusions that satisfy basic conditions, and we relax the required Lyapunov-like conditions on the variance. In Vidyasagar (2023), the variance is bounded by an affine function of the state's squared distance to an equilibrium point and the Lyapunov function is upper and lower bounded by linear functions of this squared distance. In our result, the Lyapunov function must have compact sublevel sets and bound the variance only when the value of the Lyapunov-like function is large. This condition is similar to an assumption made in Kamal (2012) (see also (Borkar, 2023, Section 4.4)). It is also similar to results for boundedness in stochastic approximations of ODEs given in (Kushner and Yin, 1997, Theorem 4.3); see also (Benaïm, 1999, Theorem 7.6).

Once almost sure boundedness is secured, we use a completely separate Lyapunov analysis for almost sure convergence, and we assume a weak condition on the variance, which resembles a variance condition used in Kamal (2012) (see also (Borkar, 2023, Section 4.4)). Under this weaker condition, we show that the solutions of the stochastic approximation converge almost surely to the chain recurrent part of the global attractor, when such an attractor exists. The global attractor here is, equivalently, the smallest globally asymptotically stable set and the Omega-limit of every large enough bounded set. This convergence statement recovers the convergence result of Benaïm et al. (2005) (see also (Nguyen and Yin, 2023, Theorem 2.1)), but with a more modular, direct proof. An important intermediate step is to use a Lyapunov analysis to establish convergence to any Morse decomposition of the attractor.

2. A STOCHASTIC DIFFERENCE INCLUSION

Consider the stochastic difference inclusion

$$x^+ - x \in h^+ \widehat{F}(x, y^+) \tag{1}$$

where $x, x^+ \in \mathbb{R}^n$, $y^+ \in \mathbb{R}^m$, $\widehat{F} : \mathbb{R}^n \times \mathbb{R}^m \rightrightarrows \mathbb{R}^n$ is a set-valued mapping, and h^+ and y^+ are placeholders for, respectively, a sequence $\{h_k\}_{k=1}^\infty$ of (deterministic) positive real numbers that is square summable, i.e.,

$$\sum_{k=1}^{\infty} h_k^2 < \infty, \tag{2}$$

and a sequence $\{\mathbf{y}_k\}_{k=1}^\infty$ of random variables defined on a probability space $(\Omega, \mathcal{F}, \mathbb{P})$ taking values in a set $Y \subset \mathbb{R}^m$. Let $\{\mathcal{F}_k\}_{k=0}^\infty$ be the natural filtration of $\{\mathbf{y}_k\}_{k=1}^\infty$.

A solution of (1) is a sequence $\{\mathbf{x}_k\}_{k=0}^\infty$ such that, for all $k \in \mathbb{Z}_{\geq 0}$, $\mathbf{x}_k$ is $\mathcal{F}_k$-measurable and, for almost all $\omega \in \Omega$ and all $k \in \mathbb{Z}_{\geq 0}$,

$$\mathbf{x}_{k+1}(\omega) - \mathbf{x}_k(\omega) \in h_{k+1} \widehat{F}\left(\mathbf{x}_k(\omega), \mathbf{y}_{k+1}(\omega)\right).$$

To guarantee the existence of solutions to (1) — which is implicitly assumed in what follows — one could request that $\widehat{F}(x, y)$ be not empty for each $(x, y) \in \mathbb{R}^n \times Y$ and that the mapping $y \mapsto \mathrm{graph}(\widehat{F}(\cdot, y))$ be measurable with closed values; see (Teel, 2015, Lemma 3).

We are interested in using a Lyapunov-based approach to establish when the solutions of (1) are almost surely bounded and, when solutions are almost surely bounded, to characterize the set to which sample paths converge.

The next lemma, which is the main result of Robbins and Siegmund (1971), plays a fundamental role in our analysis:

Lemma 1. Let $(\Omega, \mathcal{F}, \mathbb{P})$ be a probability space and $\{\mathcal{F}_k\}_{k=0}^\infty$ be a filtration of $\mathcal{F}$. For each $k \in \mathbb{Z}_{\geq 0}$, let z_k, β_k, ξ_k and ζ_k be non-negative $\mathcal{F}_k$-measurable random variables such that, for almost all $\omega \in \Omega$,

$$\mathbb{E}\left[z_{k+1} | \mathcal{F}_k\right](\omega) \leq z_k(\omega)\left(1 + \beta_k(\omega)\right) + \xi_k(\omega) - \zeta_k(\omega). \tag{3}$$

Then, for $\omega \in \Omega$ satisfying

$$\sum_{k=0}^{\infty} \xi_k(\omega) < \infty \quad \& \quad \sum_{k=0}^{\infty} \beta_k(\omega) < \infty,$$

$\lim_{k\to\infty} z_k(\omega)$ exists and is finite, and $\sum_{k=0}^\infty \zeta_k(\omega) < \infty$.

3. LYAPUNOV CONDITIONS FOR ALMOST SURE BOUNDEDNESS

In this section, we give sufficient conditions for almost sure boundedness of solutions. Our assumption is inspired by that in Vidyasagar (2023):

Assumption 1. There exist a C^1 function $V : \mathbb{R}^n \to \mathbb{R}_{\geq 0}$ with a globally Lipschitz gradient and bounded sublevel sets, and nonnegative real numbers Δ and λ such that, for every solution $\mathbf{x}$ of (1), $k \in \mathbb{Z}_{\geq 0}$, and $\widehat{\mathbf{f}} \in \widehat{F}(\mathbf{x}_k, \mathbf{y}_{k+1})$ that is $\mathcal{F}_{k+1}$-measurable, and almost all $\omega \in \Omega$:
$V(\mathbf{x}_k(\omega)) \geq \Delta$ implies

$$\dot{V}_k(\omega) := \langle \nabla V(\mathbf{x}_k(\omega)), \mathbb{E}\left[\widehat{\mathbf{f}} \,|\, \mathcal{F}_k\right](\omega)\rangle \leq 0, \tag{4a}$$

$$\mathbb{E}\left[|\widehat{\mathbf{f}}|_2^2 \mid \mathcal{F}_k\right](\omega) \leq V(\mathbf{x}_k(\omega)), \tag{4b}$$

and $V(\mathbf{x}_k(\omega)) \leq \Delta$ implies

$$\mathbb{E}\left[|\widehat{\mathbf{f}}|_2^2 \mid \mathcal{F}_k\right](\omega) \leq \lambda. \tag{5}$$

□

Assumption 1 is similar to the conditions in (Vidyasagar, 2023, Theorem 5). The Lyapunov function there is C^2 with a bounded Hessian, rather than just C^1 with a globally Lipschitz gradient. The properties assumed there impose more structure on the Lyapunov function employed: it is upper and lower bounded by positive real numbers times the squared distance of x to some equilibrium point. However, the essence of that condition is the same. In terms of the type of Lyapunov function considered, the property assumed in (Vidyasagar, 2023, Theorem 5) corresponds to $\Delta = 0$ for the bounds in (4), but where the right-hand side of (4b) has the form $c_1 + c_2 V(x)$ where $c_1, c_2 > 0$. Here, we can take $c_1 = 0$ and $c_2 = 1$ without loss of generality by redefining V to be $c_1 + c_2 V$, which changes Δ to $c_1 + c_2\Delta$ and does not affect (4a). Assumption 1 also has a connection to (Swenson et al., 2022, Assumption B.2), which is used to establish almost sure boundedness in the context of distributed stochastic gradient descent; there, effectively, $V(x) = cx^T x$ for some $c > 0$.

Inspired by the "expected smoothness" condition of (Khaled and Richtárik, 2020, Assumption 2), which has also inspired conditions in Liu and Yuan (2024) and

Karandikar and Vidyasagar (2024) related to convergence rates in stochastic approximations, we might consider replacing the right-hand side of (4b) by $V(\mathbf{x}_k(\omega)) - \alpha \dot{V}_k(\omega)$ for some $\alpha > 0$. However, this seemingly more general bound does not provide any extra generality when V is C^1 with a globally Lipschitz gradient. To see this, first note that, for all x, $|\nabla V(x)|^2 \le 2LV(x)$ where L is a Lipschitz constant of ∇V. Indeed, for each $\sigma > 0$, (see (6) below)

$$0 \le V(x - \sigma \nabla V(x)) \le V(x) - \sigma |\nabla V(x)|^2 + \sigma^2 \frac{L}{2} |\nabla V(x)|^2.$$

Taking $\sigma = 1/L$ gives the claimed bound. Now, define

$$\Xi := \mathbb{E}\left[\widehat{\mathbf{f}} \mid \mathcal{F}_k \right](\omega), \qquad \Upsilon := \mathbb{E}\left[|\widehat{\mathbf{f}}|_2^2 \mid \mathcal{F}_k \right](\omega)$$

and note that $|\Xi|^2 \le \Upsilon$. Then, from the definition of $\dot{V}_k(\omega)$,

$$-\alpha \dot{V}_k(\omega) \le \frac{1}{2}\left(\alpha^2 |\nabla V(\mathbf{x}_k)|^2 + |\Xi|^2 \right) \le \alpha^2 L V(\mathbf{x}_k) + \frac{1}{2}\Upsilon.$$

Thus, if $\Upsilon \le V(\mathbf{x}_k) - \alpha \dot{V}_k(\omega)$ then $\Upsilon \le 2\left(1 + \alpha^2 L\right) V(\mathbf{x}_k)$. The constant multiplying V can then be absorbed into V.

While the parameter Δ in Assumption 1 appears to be a convenient degree of freedom, we can take $\Delta = 0$ without loss of generality, as the following lemma establishes.

Lemma 2. If Assumption 1 holds for $\Delta > 0$ then it holds with $\widehat{\Delta} = 0$ in place of Δ and $\widehat{V} := \rho(V)$ in place of V where $\rho : \mathbb{R}_{\ge 0} \to \mathbb{R}_{\ge 0}$ is any C^2 function that is nondecreasing, constant on $[0, \Delta]$, and satisfies $\rho(s) \ge \max\{\lambda, s\}$ for all $s \ge 0$, and $\rho(s) = s$ for large s.

Proof. Since, according to Assumption 1, V is C^1 with a globally Lipschitz gradient and its sublevel sets are compact, and since ρ is C^2 and equal to the identity for large values of its argument, $\widehat{V}$ is C^1 with a globally Lipschitz gradient and has compact sublevel sets.

Since ρ is C^2, nondecreasing, and constant on $[0, \Delta]$ and (4a) holds for $V(\mathbf{x}_k(\omega)) \ge \Delta$, it follows that (4a) holds with $\widehat{V}$ in place of V and $\widehat{\Delta} = 0$ in place of Δ.

Finally, because $\widehat{V}(x) \ge \max\{\lambda, V(x)\}$ for all $x \in \mathbb{R}^n$, the combination of (4b) and (5) gives that (4b) holds with $\widehat{V}$ in place of V and $\widehat{\Delta} = 0$ in place of Δ. ∎

The following theorem generalizes (Vidyasagar, 2023, Theorem 5) to differential inclusions while relaxing the Lyapunov conditions somewhat. Using a technique similar to that used to prove (Vidyasagar, 2023, Theorem 5), it establishes that Assumption 1 is sufficient to guarantee that the solutions of (1) are bounded almost surely.

Theorem 1. If the condition (2) and Assumption 1 hold then each solution of (1) is bounded almost surely.

Proof. According to Lemma 2, we can assume $\Delta = 0$ in Assumption 1. Note that, for all $(x, \eta) \in \mathbb{R}^n \times \mathbb{R}^n$,

$$\begin{aligned} V(x+\eta) &= V(x) + \int_0^1 \langle \nabla V(x + t\eta), \eta \rangle dt \\ &\le V(x) + \langle \nabla V(x), \eta \rangle + \frac{1}{2} L |\eta|_2^2 \end{aligned} \qquad (6)$$

where L is a Lipschitz constant for ∇V. Setting $x := \mathbf{x}_k(\omega)$ and $\eta := \mathbf{x}_{k+1}(\omega) - \mathbf{x}_k(\omega) \in h_{k+1} \widehat{F}(\mathbf{x}_k(\omega), \mathbf{y}_{k+1}(\omega))$ and taking conditional expectations in (6), it follows using (4), and suppressing the dependence on ω, that

$$\mathbb{E}\left[V(\mathbf{x}_{k+1}) \,|\, \mathcal{F}_k \right] \le V(\mathbf{x}_k) + h_{k+1}^2 \frac{L}{2} V(\mathbf{x}_k).$$

It then follows from Lemma 1 with $(z_k, \beta_k, \xi_k, \zeta_k) := \left(V(\mathbf{x}_k), h_{k+1}^2 \frac{L}{2}, 0, 0\right)$ for all $k \in \mathbb{Z}_{\ge 0}$ that, for almost all $\omega \in \Omega$, $\lim_{k \to \infty} V(\mathbf{x}_k(\omega))$ exists and is finite. Almost sure boundedness of $\mathbf{x}$ then follows from the condition in Assumption 1 that the sublevel sets of V are bounded. ∎

4. CONVERGENCE PROPERTIES OF DIFFERENTIAL INCLUSIONS

4.1 Omega-limit sets

To prepare for characterizing the asymptotic behavior of the solutions to (1), we review convergence properties of the solutions to

$$\dot{x} \in F(x) \qquad (7)$$

where the set-valued mapping $F : \mathbb{R}^n \rightrightarrows \mathbb{R}^n$ satisfies the following *basic conditions*:

Assumption 2. (Basic conditions) The set-valued mapping $F : \mathbb{R}^n \rightrightarrows \mathbb{R}^n$ is outer semi-continuous (i.e., its graph is closed) and locally bounded with non-empty, convex values. □

For each set $K \subset \mathbb{R}^n$, we let Omega(K) denote[1] the Omega-limit set from K for (7). With some abuse of notation, Omega(x) represents Omega$(\{x\})$.

Assumption 3. For the system (7), for every $i \in \mathbb{Z}_{\ge 0}$ the infinite horizon reachable set $\mathcal{R}_{t \ge 0}(i\mathbb{B})$ is bounded and the sequence $\{\text{Omega}(i\mathbb{B})\}_{i=1}^{\infty}$ of Omega-limit sets of (7) is bounded. □

Proposition 1 makes a straightforward observation about Omega-limit sets, useful in characterizing the convergence properties of the stochastic approximation algorithm.

Proposition 1. Under Assumptions 2-3, the set

$$\mathcal{A} := \lim_{i \to \infty} \text{Omega}(i\mathbb{B})$$

is nonempty, compact, weakly backward invariant, and globally asymptotically stable.

The second item of the following lemma, which gives conditions under which Assumption 3 holds, should be compared to (4a) in Assumption 1.

Lemma 3. If there exists a C^1 function $V : \mathbb{R}^n \to \mathbb{R}_{\ge 0}$ such that:

(1) $\{x \in \mathbb{R}^n : V(x) \le c\}$ is bounded for each $c > 0$,
(2) there exists $\Delta \ge 0$ such that
$$V(x) \ge \Delta \implies \langle \nabla V(x), f \rangle \le 0 \quad \forall f \in F(x),$$
(3) there exists $c^* > 0$ such that if $c > 0$ and $x(\cdot)$ is a complete solution of (7) satisfying $V(x(t)) = c$ for all $t \ge 0$ then $c \le c^*$,

then Assumption 3 holds.

Proof. Let $i \in \mathbb{Z}_{\ge 0}$ and define $c_i := \sup_{x \in i\mathbb{B}} V(x)$; this value is finite since $i\mathbb{B}$ is compact and V is continuous. It then follows from the first two assumptions of the lemma

[1] Typically, the symbol Ω is used for an Omega-limit set; however, here Ω represents the sample space of the probability space $(\Omega, \mathcal{F}, \mathbb{P})$.

that each maximal solution x of (7) starting in $i\mathbb{B}$ is complete, bounded, and satisfies $V(x(t)) \leq \max\{\Delta, c_i\}$ for all $t \geq 0$. Moreover, due to LaSalle's invariance principle and the third assumption of the lemma, it follows that $\limsup_{t\to\infty} V(x(t)) \leq \max\{\Delta, c^*\}$, which is an ultimate bound that is independent of i. These observations establish the result. ■

4.2 Morse decompositions

Let $\mathcal{A}$ be a compact, weakly backward invariant, globally asymptotically stable (and thus strongly forward invariant) set for the differential inclusion (7). A compact set $A \subset \mathcal{A}$ is called an attractor in $\mathcal{A}$ if there exists an open neighborhood $O \subset \mathbb{R}^n$ of A such that $A = \Omega(O \cap \mathcal{A})$. For an attractor A in $\mathcal{A}$, define

$$A^* := \{x \in \mathcal{A} : \mathrm{Omega}(x)\backslash A \neq \emptyset\}.$$

The pair (A, A^*) is called an attractor-repellor pair in $\mathcal{A}$.

An ordered collection $\mathcal{M} := \{M_1, \ldots, M_\ell\}$ of compact subsets of $\mathcal{A}$ is called a *Morse decomposition* of $\mathcal{A}$ if there exists an increasing sequence

$$\emptyset = A_0 \subset A_1 \subset \cdots \subset A_\ell = \mathcal{A}$$

of attractors in $\mathcal{A}$ such that

$$M_k = A_k \cap A^*_{k-1} \quad \forall k \in \{1, \ldots, \ell\}.$$

For an exposition, in the setting of unique solutions, examples, and alternative definitions, see the survey Ayala et al. (2006). Since there is at most countably many attractors in $\mathcal{A}$, there is at most countably many Morse decompositions. (See (Conley, 1978, Chapter II, 6.4.A) or the proof of (Li, 2007, Theorem 6.13).)

For each Morse decomposition of $\mathcal{A}$, let $\mathcal{D}_\mathcal{M} := \bigcup_{i=1}^{\ell} M_i$. Each decomposition admits a Lyapunov-like characterization, in terms of a smooth "Morse-Lyapunov" function:

Proposition 2. (Li and Wang, 2012, Theorem 1.1) Suppose Assumption 2 holds. Let $\mathcal{A}$ be a compact, weakly backward invariant, globally asymptotically stable set for (7) and let $\mathcal{M} = \{M_1, \ldots, M_\ell\}$ be a Morse decomposition of $\mathcal{A}$. Then there exist a smooth, radially unbounded function $U : \mathbb{R}^n \to \mathbb{R}_{\geq 0}$ and a continuous function $W : \mathbb{R}^n \to \mathbb{R}_{\geq 0}$ that is positive definite with respect to $\mathcal{D}_\mathcal{M}$ such that U is constant on each Morse set,

$$U(M_1) < U(M_2) < \cdots < U(M_\ell)$$

and

$$\langle \nabla U(x), f\rangle \leq -W(x) \qquad \forall x \in \mathbb{R}^n, f \in F(x).$$

The set

$$\mathcal{R}(\mathcal{A}) := \bigcap \{\mathcal{D}_\mathcal{M} \mid \mathcal{M} \text{ is a Morse decomp. of } \mathcal{A}\} \tag{8}$$

is the so-called *chain-recurrent part of* $\mathcal{A}$, consisting of points that are *chain recurrent*. The definitions go back to Conley (1978), for single-valued flows, and were generalized to a variety of settings without uniqueness of solutions; see the recent Goebel (2023) and the references therein. That (8) is a valid description of $\mathcal{R}(\mathcal{A})$ in the setting of this paper can be shown like (Li, 2007, Theorem 6.12(1)) was shown, but relying on (Goebel, 2023, Proposition 11(a)) rather than (Li, 2007, Theorem 6.11).

For an example of the chain-recurrent set for a differential inclusion, consider a locally Lipschitz $f : \mathbb{R}^n \to \mathbb{R}$ that is nonpathological. This concept goes back to Valadier (1988) and was popularized for control purposes, including Lyapunov analysis, by Shevitz and Paden (1994); see Bacciotti and Ceragioli (2006) for details and further references. Nonpathological functions satisfy an appropriate chain rule, when composed with an absolutely continuous solution to (7); see (Bacciotti and Ceragioli, 2006, Proposition 1). Then, thanks to the recent (Davis et al., 2020, Lemma 5.2), f is decreasing along every solution to $\dot{x} \in -\partial f(x)$ unless the solution is at a critical point of f, i.e., a point x such that $0 \in \partial f(x)$. Here, ∂f is the Clarke generalized gradient of f. From this, and subject to some compactness assumptions, it can be deduced that the set of chain-recurrent points for $\dot{x} \in -\partial f(x)$ equals the set of critical points of f.

5. CONVERGENCE PROPERTIES OF THE STOCHASTIC APPROXIMATION

In this section, we assume that the solutions of the stochastic difference inclusion (1) are almost surely bounded and we impose the following conditions that guarantee almost sure convergence to the chain-recurrent part of the differential inclusion's Omega-limit set under Assumption 3:

Assumption 4. The continuous, nondecreasing function $\gamma : \mathbb{R}_{\geq 0} \to \mathbb{R}_{\geq 0}$ and the set-valued mapping $F : \mathbb{R}^n \rightrightarrows \mathbb{R}^n$ are such that, for each solution $\mathbf{x}$ of (1) and each $k \in \mathbb{Z}_{\geq 0}$, if $\widehat{\mathbf{f}} \in \widehat{F}(\mathbf{x}_k, \mathbf{y}_{k+1})$ is $\mathcal{F}_{k+1}$-measurable then, almost surely,

$$\mathbb{E}\left[\widehat{\mathbf{f}} \mid \mathcal{F}_k\right](\omega) \in F(\mathbf{x}_k(\omega)) \tag{9a}$$

$$\mathbb{E}\left[|\widehat{\mathbf{f}}|^2 \mid \mathcal{F}_k\right](\omega) \leq \gamma(|\mathbf{x}_k(\omega)|). \tag{9b}$$

□

We note that (9b) is similar to (Davis et al., 2020, Assumption C, item 3), which appears in the study of stochastic approximations of subgradient dynamics. It is possible to allow $\widehat{F}$ to depend on h_{k+1} by changing the containment on the right-hand side of (9a) to be $F(\mathbf{x}_k(\omega)) + h_{k+1}\gamma(|\mathbf{x}_k(\omega)|)\mathbb{B}$; for more details, see the results for stochastic approximation of hybrid systems in Teel et al. (2025).

The main result of this section is the following:

Theorem 2. Suppose:

1) Assumptions 2-4 hold,
2) each solution of (1) is almost surely bounded,
3) the condition (2) holds,
4) and $\sum_{k=1}^{\infty} h_k = \infty$.

Let $\mathcal{M} := \{M_1, \ldots, M_\ell\}$ be a Morse decomposition of $\mathcal{A} := \lim_{i\to\infty} \mathrm{Omega}(i\mathbb{B})$. Then every solution $\mathbf{x}$ to (1) satisfies $\lim_{k\to\infty} |\mathbf{x}_k(\omega)|_{\mathcal{D}_\mathcal{M}} = 0$ almost surely.

The proof technique we use is noteworthy because, unlike in Benaïm et al. (2005) and Nguyen and Yin (2023), no attempt is made to relate the solutions of (1) to the solutions of (7). Instead, the converse Morse-Lyapunov theorem from Li and Wang (2012) for (7), quoted as Proposition 2 earlier, is invoked and the resulting function is used directly to establish the properties of the solutions to (1) stated in Theorem 2.

We do adopt a preliminary step from Benaïm et al. (2005), presented in Lemma 4 below. In preparation for that step, for each $n \in \mathbb{Z}_{\geq 0}$ and each $t \in \mathbb{R}_{\geq 0}$, let

$$\tau_n := \sum_{k=0}^{n-1} h_{k+1}, \quad m(t) := \sup\{n \in \mathbb{Z}_{\geq 0} : \tau_n \leq t\}.$$

In what follows, we suppress the argument $\omega \in \Omega$ of random variables, for notational clarity and to save space.

Lemma 4. Suppose (2) and Assumption 4 hold, and let $\mathbf{x}$ be an almost surely bounded solution of (1). For each $k \in \mathbb{Z}_{\geq 0}$, define

$$\widehat{\mathbf{f}}_{k+1} := (\mathbf{x}_{k+1} - \mathbf{x}_k)/h_{k+1} \in \widehat{F}(\mathbf{x}_k, \mathbf{y}_{k+1}) \tag{10a}$$

$$\mathbf{f}_k := \mathbb{E}\left[\widehat{\mathbf{f}}_{k+1} \,|\, \mathcal{F}_k\right] \in F(\mathbf{x}_k). \tag{10b}$$

For each $T > 0$ the following property holds almost surely:

$$\lim_{n\to\infty} \sup_{n \leq k \leq m(\tau_n+T)} \left| \mathbf{x}_k - \mathbf{x}_n - \sum_{i=n}^{k-1} h_{i+1}\mathbf{f}_i \right| = 0. \tag{11}$$

Proof. Define $\mathbf{s}_i : \Omega \to \mathbb{Z}_{\geq 0} \cup \{\infty\}$ as

$$\mathbf{s}_i := \inf\{k \in \mathbb{Z}_{\geq 0} : |\mathbf{x}_k| > i\}$$

and note that, for each $k \in \mathbb{Z}_{\geq 0}$,

$$\{\mathbf{s}_i \leq k\} = \cup_{\ell=0}^{k} \{\mathbf{x}_\ell| > i\} \in \mathcal{F}_k.$$

It follows that, for each $k \in \mathbb{Z}_{\geq 0}$, the function

$$\mathbf{1}(k < \mathbf{s}_i) := \begin{cases} 1 & k < \mathbf{s}_i \\ 0 & \text{otherwise} \end{cases}$$

is $\mathcal{F}_k$-measurable. For each $k \in \mathbb{Z}_{\geq 0}$, define

$$\mathbf{x}_k^{\mathbf{s}_i} := \mathbf{x}_{min\{k,\mathbf{s}_i\}} = \mathbf{1}(k < \mathbf{s}_i)\mathbf{x}_k + \sum_{j=0}^{k} \mathbf{1}(\mathbf{s}_i = j)\mathbf{x}_j$$

which is also $\mathcal{F}_k$-measurable. Using the definition of $\widehat{\mathbf{f}}_{k+1}$ in (10a), for each $k \in \mathbb{Z}_{\geq 0}$, define

$$\begin{aligned}
\widehat{\mathbf{f}}^{\mathbf{s}_i}_{k+1} &:= (\mathbf{x}^{\mathbf{s}_i}_{k+1} - \mathbf{x}^{\mathbf{s}_i}_k)/h_{k+1} = \mathbf{1}(k < \mathbf{s}_i)\widehat{\mathbf{f}}_{k+1} \\
\mathbf{f}^{\mathbf{s}_i}_k &:= \mathbb{E}\left[\widehat{\mathbf{f}}^{\mathbf{s}_i}_{k+1} \,|\, \mathcal{F}_k\right] \in \mathbf{1}(k < \mathbf{s}_i)F(\mathbf{x}_k) \\
\mathbf{u}^{\mathbf{s}_i}_{k+1} &:= \widehat{\mathbf{f}}^{\mathbf{s}_i}_{k+1} - \mathbf{f}^{\mathbf{s}_i}_k.
\end{aligned}$$

Note that

$$\begin{aligned}
\mathbb{E}\left[\mathbf{u}^{\mathbf{s}_i}_{k+1} \,|\, \mathcal{F}_k\right] &= \mathbb{E}\left[\widehat{\mathbf{f}}^{\mathbf{s}_i}_{k+1} \,|\, \mathcal{F}_k\right] - \mathbf{f}^{\mathbf{s}_i}_k = 0 \\
\mathbb{E}\left[|\mathbf{u}^{\mathbf{s}_i}_{k+1}|^2 \,|\, \mathcal{F}_k\right] &\leq \mathbb{E}\left[|\widehat{\mathbf{f}}^{\mathbf{s}_i}_{k+1}|^2 \,|\, \mathcal{F}_k\right] \\
&= \mathbf{1}(k < \mathbf{s}_i)\mathbb{E}\left[|\widehat{\mathbf{f}}_{k+1}|^2 \,|\, \mathcal{F}_k\right] \\
&\leq \mathbf{1}(k < \mathbf{s}_i)\gamma(|\mathbf{x}_k|) \leq \gamma(i).
\end{aligned}$$

Then, with (2), we use (Benaïm et al., 2005, Proposition 1.4) with $q = 2$ to conclude (11) for the stopped process $\mathbf{x}^{\mathbf{s}_i}$ for each $i \in \mathbb{Z}_{\geq 0}$. Finally, since the solution $\mathbf{x}$ is almost surely bounded, for almost every $\omega \in \Omega$, there exists $i^* \in \mathbb{Z}_{\geq 0}$ such that $\mathbf{s}_i(\omega) = \infty$ for all $i \geq i^*$. Thus, we then get that (11) holds almost surely. ■

Proof (of Theorem 2). Under the theorem's assumptions, Proposition 2 applies, and yields (U, W). Let $\kappa \in \mathcal{K}$ be a smooth concave function such that, for some $L > 0$,

$$\begin{aligned}
\kappa'(s) &> 0 \qquad \forall s > 0 \\
|\kappa'(U(x))\nabla^2 U(x)| &\leq L \quad \forall x \in \mathbb{R}^n.
\end{aligned}$$

For all $x \in \mathbb{R}^n$, define

$$V(x) := \kappa(U(x)), \qquad Y(x) := \kappa'(U(x))W(x).$$

It follows that V is constant on each Morse set, Y is positive definite with respect to $\mathcal{D}_{\mathcal{M}}$, V is smooth with a globally Lipschitz ∇V, with constant L, and

$$\langle \nabla V(x), f\rangle \leq -Y(x) \qquad \forall x \in \mathbb{R}^n, f \in F(x). \tag{12}$$

As in the proof of Theorem 1, for all $(x, \eta) \in \mathbb{R}^n \times \mathbb{R}^n$, (6) holds and, with $x := \mathbf{x}_k$ and $\eta := \mathbf{x}_{k+1} - \mathbf{x}_k$ and taking conditional expectations, it follows from (9) and (12) that

$$\mathbb{E}\left[V(\mathbf{x}_{k+1}) \mid \mathcal{F}_k\right] \leq V(\mathbf{x}_k) - h_{k+1}Y(\mathbf{x}_k) + h^2_{k+1}\frac{L}{2}\gamma(|\mathbf{x}_k|).$$

For all $k \in \mathbb{Z}_{\geq 0}$, define

$$(z_k, \beta_k, \xi_k, \zeta_k) := \left(V(\mathbf{x}_k), 0, h^2_{k+1}\frac{L}{2}\gamma\left(|\mathbf{x}_k|\right), h_{k+1}Y(\mathbf{x}_k)\right).$$

For almost all $\omega \in \Omega$,

$$\sum_{k=0}^{\infty} \xi_k(\omega) \leq \frac{L}{2}\gamma\left(\sup_{k\in\mathbb{Z}_{\geq 0}} |\mathbf{x}_k(\omega)|\right) \sum_{k=0}^{\infty} h^2_{k+1} < \infty$$

where the final bound follows from (2) and the almost sure boundedness of the sequence $\{\mathbf{x}_k\}_{k=0}^{\infty}$. It then follows from Lemma 1 that, for almost all $\omega \in \Omega$, $\lim_{k\to\infty} V(\mathbf{x}_k(\omega))$ exists and is finite and

$$\sum_{k=0}^{\infty} \zeta_k(\omega) = \sum_{k=0}^{\infty} h_{k+1}Y(\mathbf{x}_k(\omega)) < \infty. \tag{13}$$

Let $\widehat{\Omega} \subset \Omega$ with $\mathbb{P}(\widehat{\Omega}) = 1$ be such that, for all $\omega \in \widehat{\Omega}$, $\mathbf{x}(\omega)$ is bounded and (13) and (11) hold. For the remainder of the proof, fix a solution $\mathbf{x}$, fix $\omega \in \widehat{\Omega}$, and suppress the dependence of $\mathbf{x}$ and $\mathbf{f}$ on ω.

If $\limsup_{k\to\infty} Y(\mathbf{x}_k) > 0$ then there exists an accumulation point $x^* \in \mathbb{R}^n$ for the sequence $\{\mathbf{x}_k\}_{k=0}^{\infty}$ such that $\varepsilon := Y(x^*) > 0$. Continuity of Y yields $\delta > 0$ such that

$$Y(x) \geq 0.5\varepsilon \qquad \forall x \in \{x^*\} + \delta\mathbb{B}. \tag{14}$$

The boundedness of the sequence $\{\mathbf{x}_k\}_{k=0}^{\infty}$ and the local boundedness of F yield $T > 0$ sufficiently small such that

$$T|f| \leq \delta/3 \qquad \forall k \in \mathbb{Z}_{\geq 0}, \forall f \in F(\mathbf{x}_k). \tag{15}$$

Using (11), for this T let $n^* \in \mathbb{Z}_{\geq 0}$ be sufficiently large such that, for all $n \in \mathbb{Z}_{\geq n^*}$,

$$\sup_{n\leq k\leq m(\tau_n+T)} \left|\mathbf{x}_k - \mathbf{x}_n - \sum_{i=n}^{k-1} h_{i+1}\mathbf{f}_i\right| \leq \delta/3. \tag{16}$$

Let $\{n_\ell\}_{\ell=1}^{\infty}$ be an unbounded sequence of positive integers such that, for all $\ell \in \mathbb{Z}_{\geq 1}$,

$$n_{\ell+1} \geq \max\{n^*, m(\tau_{n_\ell} + T)\} \tag{17a}$$

$$\mathbf{x}_{n_\ell} \in \{x^*\} + (\delta/3)\mathbb{B}. \tag{17b}$$

Then, using (15), (16), and (17b),

$$\begin{aligned}
&\sup_{n\leq k\leq m(\tau_n+T)} |\mathbf{x}_k - x^*| \\
&= \sup_{n\leq k\leq m(\tau_n+T)} \left|\mathbf{x}_k - \mathbf{x}_{n_\ell} - \sum_{i=n_\ell}^{k-1} h_{i+1}\mathbf{f}_i \right. \\
&\qquad\qquad \left. + \sum_{i=n_\ell}^{k-1} h_{i+1}\mathbf{f}_i + \mathbf{x}_{n_\ell} - x^*\right| \leq \delta.
\end{aligned}$$

Let $\ell^* \in \mathbb{Z}_{\geq 1}$ be sufficiently large such that

$$\sum_{k=n_\ell}^{m(\tau_{n_\ell}+T)-1} h_{k+1} \in [0.5T, T] \qquad \forall \ell \in \mathbb{Z}_{\geq \ell^*}.$$

This, and (14), yield that for $\ell \in \mathbb{Z}_{\geq \ell^*}$,

$$\sum_{k=n_\ell}^{m(\tau_{n_\ell}+T)-1} h_{k+1} Y(\mathbf{x}_k) \geq 0.5\varepsilon \sum_{k=n_\ell}^{m(\tau_{n_\ell}+T)-1} h_{k+1} \geq 0.25\varepsilon T.$$

When combined with (17a), this contradicts (13). Thus $\lim_{k\to\infty} Y(\mathbf{x}_k) = 0$ and $\mathbf{x}$ converges to $\mathcal{D}_{\mathcal{M}}$. ■

Since Theorem 2 establishes convergence to $\mathcal{D}_{\mathcal{M}}$ for an arbitrary Morse decomposition $\mathcal{M}$, of which there are countably many, using (8) the conclusion can be refined to align with the conclusion of Benaïm et al. (2005) and Nguyen and Yin (2023).

Corollary 1. Under the assumptions of Theorem 2, for almost every $\omega \in \Omega$, $\mathbf{x}_k(\omega)$ converges to the chain-recurrent part of $\mathcal{A}$, i.e., $\lim_{k\to\infty} |\mathbf{x}_k(\omega)|_{\mathcal{R}(\mathcal{A})} = 0$.

Alternatively, a recent result in Goebel and Teel (2025) on the existence of a smooth Lyapunov function that is decreasing outside of $\mathcal{R}(\mathcal{A})$ permits obtaining Corollary 1 directly using the technique used to prove Theorem 2.

REFERENCES

Ayala, J., Corbin, P., Mc Conville, K., Colonius, F., Kliemann, W., and Peters, J. (2006). Morse decomposition, attractors and chain recurrence. *Proyecciones*, 25(1), 79–110.

Bacciotti, A. and Ceragioli, F. (2006). Nonpathological Lyapunov functions and discontinuous Carathéodory systems. *Automatica J. IFAC*, 42(3), 453–458.

Benaïm, M. (1999). Dynamics of stochastic approximation algorithms. In J. Azéma, M. Émery, M. Ledoux, and M. Yor (eds.), *Séminaire de Probabilités XXXIII*, 1–68. Springer Berlin Heidelberg, Berlin, Heidelberg.

Benaïm, M., Hofbauer, J., and Sorin, S. (2005). Stochastic approximations and differential inclusions. *SIAM J. Control Optim.*, 44(1), 328–348.

Borkar, V.S. and Meyn, S.P. (2000). The O.D.E. method for convergence of stochastic approximation and reinforcement learning. *SIAM J. Control Optim.*, 38(2), 447–469.

Borkar, V. (2023). *Stochastic Approximation: A Dynamical Systems Viewpoint, 2nd Edition.* Springer and Hindustan Book Agency.

Conley, C. (1978). *Isolated invariant sets and the Morse index*, volume 38 of *CBMS Regional Conference Series in Mathematics.* Amer. Math. Society, Providence, R.I.

Davis, D., Drusvyatskiy, D., Kakade, S., and Lee, J.D. (2020). Stochastic subgradient method converges on tame functions. *Found. Comput. Math.*, 20, 119–154.

Faure, M. and Roth, G. (2010). Stochastic approximations of set-valued dynamical systems: Convergence with positive probability to an attractor. *Math. Oper. Res.*, 35(3), 624–640.

Goebel, R. (2023). A direct proof of Conley's decomposition for well-posed hybrid inclusions. *Systems Control Lett.*, 180.

Goebel, R.K. and Teel, A.R. (2025). A smooth complete/total/Conley Lyapunov function for hybrid inclusions on $\mathbb{R}^n$. Submitted.

Kamal, S. (2010). On the convergence, lock-in probability, and sample complexity of stochastic approximation. *SIAM J. on Control and Opt.*, 48(8), 5178–5192.

Kamal, S. (2012). Stabilization of stochastic approximation by step size adaptation. *Systems Control Lett.*, 61(4), 543–548.

Karandikar, R. and Vidyasagar, M. (2024). Convergence rates for stochastic approximation: Biased noise with unbounded variance, and applications. *J. Optim. Theory Appl.*, 203, 2412–2450.

Khaled, A. and Richtárik, P. (2020). Better theory for SGD in the nonconvex world. URL `https://arxiv.org/abs/2002.03329`.

Kushner, H.J. and Yin, G.G. (1997). *Stochastic Approximation Algorithms and Applications.* Springer-Verlag, New York.

Li, D. (2007). Morse decompositions for general dynamical systems and differential inclusions with applications to control systems. *SIAM J. Control Optim.*, 46(1), 35–60.

Li, D. and Wang, Y. (2012). Smooth Morse–Lyapunov functions of strong attractors for differential inclusions. *SIAM J. Control Optim.*, 50(1), 368–387.

Liu, J. and Yuan, Y. (2024). Almost sure convergence rates analysis and saddle avoidance of stochastic gradient methods. *J. Mach. Learn. Res.*, 25(271), 1–40.

Nguyen, N. and Yin, G. (2023). Stochastic approximation with discontinuous dynamics, differential inclusions, and applications. *Ann. Appl. Probab.*, 33(1), 780–823.

Ramaswamy, A. and Bhatnagar, S. (2017). A generalization of the Borkar-Meyn theorem for stochastic recursive inclusions. *Math. Oper. Res.*, 42(3), 648–661.

Ramaswamy, A. and Bhatnagar, S. (2022). Analyzing approximate value iteration algorithms. *Math. Oper. Res.*, 47(3), 2138–2159.

Robbins, H. and Siegmund, D. (1971). A convergence theorem for nonnegative almost supermartingales and some applications. In J.S. Rustagi (ed.), *Optimizing Methods in Statistics*, 233–257. Academic Press.

Shevitz, D. and Paden, B. (1994). Lyapunov stability theory of nonsmooth systems. *IEEE Trans. Automat. Control*, 39(9), 1910–1914.

Swenson, B., Murray, R., Pook, H., and Kar, S. (2022). Distributed stochastic gradient descent: Nonconvexity, nonsmoothness, and convergence to local minima. *J. Mach. Learn. Res.*, 23, 1–62.

Teel, A.R. (2015). A recurrence principle for stochastic difference inclusions. *IEEE Trans. Automat. Control*, 60(2), 420–435.

Teel, A.R., Sanfelice, R.G., and Goebel, R.K. (2025). Stochastic approximation of hybrid systems: boundedness and asymptotic behavior. Submitted.

Valadier, M. (1988). Lignes de descente de fonctions lipschitziennes non pathologiques. In *Seminaire d'Analyse Convexe, Vol. 18 (Montpellier, 1988)*, Exp. No. 9, 10. Univ. Sci. Tech. Languedoc, Montpellier.

Vidyasagar, M. (2023). Convergence of stochastic approximation via martingale and converse Lyapunov methods. *Math. Control Signals Syst.*, 35, 351–374.