Stochastic approximation of hybrid systems: boundedness and asymptotic behavior

Andrew R. Teel[a], Ricardo G. Sanfelice[b], Rafal K. Goebel[c]

[a] *University of California Santa Barbara, Santa Barbara, 93106, CA, USA*
[b] *University of California Santa Cruz, Santa Cruz, 95064, CA, USA*
[c] *Loyola University Chicago, Chicago, 60660, IL, USA*

Abstract

This paper provides a framework for analyzing numerical simulations of a hybrid system when the system's flow map is approximated stochastically. In this way, the paper makes one of the first substantive connections between hybrid systems theory and the vast literature on stochastic approximations of non-stochastic dynamical systems. The results include conditions for almost sure boundedness and a characterization of the asymptotic behavior of solutions to the stochastic approximation. Namely, under certain conditions on the expected value and variance of the noisy flow map, and with appropriate step sizes, the sample paths of the solutions of the stochastic approximation are shown to have desired boundedness properties and converge to the chain-recurrent part of the global attractor, when one exists, of the underlying hybrid system. The results are established using martingale methods, Morse decompositions of attractors, and Morse-Lyapunov functions. Since the hybrid system's flow map and its approximation may be set-valued, stochastic approximations of differential inclusions are covered as a special case. Several examples and simulations are provided to illustrate hybrid systems, their stochastic approximations, and their convergence properties.

Keywords: hybrid systems, stochastic approximation, Lyapunov techniques, martingale methods, Morse decomposition, global attractors.

1. Introduction

In this paper, we build on stochastic approximation theory for differential equations and inclusions to develop stochastic approximation results for hybrid systems. Stochastic approximation of differential equations has a long and significant history. It was initiated with the stochastic root-finding algorithm of [1], followed by the idea of finding the maximum of a univariate function using noisy finite differences [2], with multi-dimensional versions offered shortly thereafter [3]. A particularly concise explanation for the efficacy of the root-finding algorithm of [1] can be found in [4], whose methods anticipated the main theorem of [5] on the behavior of almost supermartingales. The so-called ODE method for the analysis of stochastic algorithms with vanishing step sizes was established in [6], with noteworthy further developments presented in [7] and [8]. Important extensions of stochastic approximation theory to general differential inclusions appeared in [9], including a tight characterization of the convergence properties of the solutions of the stochastic approximation to a chain-recurrent set determined by the underlying differential inclusion. Additional stochastic approximation results for differential inclusions appear in [10]. A recent summary of stochastic approximation results can be found in the book [11], which includes results for stochastic approximation of both differential equations and differential inclusions. Especially noteworthy in [11] is a thorough discussion of the condition provided in [12] for almost sure boundedness of the solutions of a stochastic approximation and a comparison of this condition to other conditions for boundedness that have appeared in the literature.

Beyond the academic interest in extending stochastic approximation theory to a much broader class of systems than has been considered previously, the extension to hybrid systems allows rigorous statements to be made for the first time about, for example, the use of stochastic gradient descent in hybrid optimization algorithms that employ both continuous-time gradient flow and resets. Broadly speaking, hybrid dynamical systems combine features typically attributed to continuous-time dynamics with features typically attributed to discrete-time dynamics. Hybrid inclusions model such systems by combining a differential inclusion that models the flow, a difference inclusion that models the discrete dynamics (jumps, switches, resets, and more), and constraints on the resulting motions. This class of systems has been studied extensively over the last two decades; see [13], [14] for example, and [15] for a hybrid control framework built with a hybrid inclusion as a foundation.

A symbolic representation of a hybrid inclusion is

$$x \in C \qquad \dot{x} \in F(x) \tag{1a}$$

$$x \in D \qquad x^+ \in G(x). \tag{1b}$$

Here, $\dot{x}$ suggests the usual time derivative of a finite-dimensional state x while x^+ suggests the state after an iteration of discrete dynamics, in short, after a jump. Thanks to the adopted notion of a solution to (1), presented in detail in Section 2, (1) includes, as special cases, a differential inclusion $\dot{x} \in F(x)$ or just a differential equation $\dot{x} = F(x)$ if F is a function and not a set-valued mapping, a difference inclusion $x^+ \in G(x)$ or just a difference equation $x^+ = G(x)$, and many dynamical systems (switched, impulsive, reset, etc.) where these two kinds of dynamics interact.

A stochastic approximation, or simulator, of (1) considered in this paper replaces the differential inclusion $\dot{x} \in F(x)$ with discrete and stochastic dynamics, resulting in the symbolic representation

$$x \in C \qquad x^+ - x \in h^+ \widehat{F}(x, y^+, h^+) \tag{2a}$$

$$x \in D \qquad x^+ \in G(x). \tag{2b}$$

Above, h^+ is a placeholder for a sequence of positive real numbers $\{h_k\}_{k=1}^{\infty}$, called the "step sizes", and y^+ is a placeholder for a sequence of random variables $\{\mathbf{y}_k\}_{k=1}^{\infty}$ called the "noise". Solutions to (2) are rigorously introduced in Section 4.

Roughly speaking, the connection between (2) and (1) comes through an assumption that the expected value of $\widehat{F}(x, \mathbf{y}_{k+1}, h_{k+1})$ is approximately contained in $F(x)$ for all $x \in C$ and all $k \in \mathbb{Z}_{\geq 0}$. If $\widehat{F}(x, y^+, h^+)$ in (2) is formally replaced by $F(x)$, a non-stochastic simulator of (1) results, as studied in [16]. In that work, it is shown that the behavior of (non-stochastic) simulators can be related to the behavior of (1). Namely, under the assumption of small step sizes, the solutions of non-stochastic simulators are shown to be close to the solutions of (1) when considering bounded time domains.

As far as the authors are aware, stochastic approximation of hybrid systems that employ a stochastic approximation of the flow map has been studied only in [17] and [18]. In those papers, the focus was also on establishing conditions under which the sample paths of the solutions of stochastic approximation (2) are close to the solutions of the system (1) on bounded time domains when the step sizes are sufficiently small.

In contrast, the aim of this paper is to give conditions for almost sure boundedness and show that, when almost sure boundedness is assured, the asymptotic behavior of the sample paths of the solutions to the stochastic approximation (2) mimics the asymptotic behavior of the solutions of (1). An additional goal is to establish these results while being tutorial in nature, regarding hybrid systems and their convergence properties, for those not familiar with hybrid systems and/or Morse decompositions and chain-recurrent sets, and also to be tutorial about the most relevant concepts from probability theory for those familiar with hybrid systems but not stochastic systems.

The main results do not require the step sizes to be uniformly small. Rather, for the boundedness results, the sequence of step sizes is assumed to be square summable; for the convergence results, the sequence of step sizes is assumed to be nonsummable. These assumptions are common in the stochastic approximation literature and they also match those made in [19] to establish asymptotic convergence properties for non-stochastic simulators of hybrid systems.

The techniques used and the results obtained in this paper can be thought of as an extension to the hybrid setting of those in [20], developed there for a stochastic approximation of a differential inclusion. The contributions in [20] were inspired by the recent results on boundedness and convergence for a stochastic approximation of a differential equation given in [21] as well as other results on stochastic approximation of differential inclusions, as summarized in [11, Chapter 5] as well as in [10]. In order to generalize the results in [20] to hybrid systems and their stochastic approximation, two results from the discrete-time and continuous-time literature, respectively, are extended to the hybrid setting:

(i) [5, Theorem 1], on the behavior of a class of almost supermartingales, and

(ii) [22, Theorem 1], which gives a converse Lyapunov-like result for a Morse decomposition of an attractor for a differential inclusion.

The paper is organized as follows: In Section 2 the solution concept for the hybrid system (1) is defined and illustrated with examples. In Section 3, a review of relevant probability concepts is provided. In Section 4, the solution concept for the stochastic approximation (2) is given. Section 5 establishes the first of the two needed result hinted at above, Theorem 5.3. The theorem is then used to establish conditions for almost sure boundedness of the solutions to the stochastic approximation (2) in Section 6. The main result there is Theorem 6.9. Section 7 presents Morse decompositions and states the second result hinted at above, Theorem 7.3. The proof of it is in Section 10. Theorem 5.3 and the Lyapunov-like function from Theorem 7.3

are used to show convergence of the sample paths of the stochastic solutions to Morse sets, in Section 8. The main result there is Theorem 8.8. The Morse sets are then related to chain-recurrent parts of the attractor via the results in [23] to obtain the final convergence result, Corollary 8.9. This result can be viewed as an extension of the stochastic approximation for differential inclusion results in [9] to hybrid systems though the proof technique, which uses the two items listed above, is quite different. Simulation-based examples are in Section 9.

Notation: In what follows, $\mathbb{Z}_{\geq j} := \{j, j+1, j+2, \dots\}$, $\mathbb{N} := \mathbb{Z}_{\geq 1}$, $\mathbb{R}_{\geq 0} := [0, \infty)$, and other similar notation is self-explanatory. Further, $\mathbb{B} := \{x \in \mathbb{R}^n : |x| \leq 1\}$ is the closed unit ball in $\mathbb{R}^n$ with the Euclidean norm $|\cdot|$, centered at 0, so that $r\mathbb{B}$ is the ball of radius r. Sometimes the notation (x_1, x_2), where $x_1 \in \mathbb{R}^{n_1}$ and $x_2 \in \mathbb{R}^{n_2}$, is used to denote the vector $(x_1^T\ x_2^T)^T$. Set-valued analysis concepts used below follow [24]; see also [25] for an introduction.

2. Solutions to hybrid inclusions, and more

This section reviews the solution concept and some consequences of the hybrid basic conditions, in Assumption 2.1 given below, for the hybrid inclusion (1). For a more complete exposition of (1) that includes a discussion of other modeling approaches to hybrid dynamics, see the tutorial article [13]. For technical details behind this tutorial, see the book [14]. Throughout this section, the hybrid system (1) is subject to the following *hybrid basic conditions*:

Assumption 2.1. *The data (C, F, D, G) of the system (1) satisfies the* hybrid basic conditions*, i.e.:*

1. *$C, D \subset \mathbb{R}^n$ are closed;*
2. *$F, G : \mathbb{R}^n \rightrightarrows \mathbb{R}^n$ are set-valued mappings that are outer semicontinuous and locally bounded;*
3. *the values $F(x)$ are nonempty and convex for all $x \in C$, while the values of $G(x)$ are nonempty for all $x \in D$.*

The double-arrow notation in $M : \mathbb{R}^n \rightrightarrows \mathbb{R}^n$ means that for every $x \in \mathbb{R}^n$, $M(x)$ is a subset of $\mathbb{R}^n$. A set-valued map $M : \mathbb{R}^n \rightrightarrows \mathbb{R}^n$ is *outer* *semicontinuous* if, for every convergent sequence $\{(x_i, y_i)\}_{i=1}^{\infty}$ such that $y_i \in M(x_i)$ for each $i \in \mathbb{N}$, one has

$$\lim_{i\to\infty} y_i \in M\left(\lim_{i\to\infty} x_i\right).$$

It is immediate that this property is equivalent to the graph of M, namely the set $\operatorname{graph}(M) := \{(x, y) \in \mathbb{R}^n \times \mathbb{R}^n : y \in M(x)\}$, being closed. The map M is *locally bounded* if for every bounded set $S \subset \mathbb{R}^n$, the set $M(S) := \bigcup_{x\in S} M(x)$ is bounded. The combination of items 2 and 3 in Assumption 2.1 ensures that the values of F and G are compact. If compactness of the values is assured a priori, then the combination of outer semicontinuity and local boundedness is equivalent to "upper semicontinuity," which is another notion of continuity of a set-valued mapping.

The hybrid basic conditions in Assumption 2.1

(a) reduce, when C and D are closed, to F, or G, being continuous if F, or G, is a function;

(b) arise naturally when the effect of vanishing perturbations on a less-regular hybrid inclusion is accounted for [26], for reasons similar to those that led to Filippov [27] and Krasovskii [28] solutions to an irregular differential equation;

(c) have facilitated the development of robust asymptotic stability theory for hybrid inclusions [14], including necessary and sufficient Lyapunov conditions, invariance techniques, robustness to perturbations, and more, which turns out to be the foundation for the stochastic approximation results given herein. Indeed, the stochastic approximation (2) can be thought of as a perturbation of the hybrid inclusion (1), and thus results that address how perturbed solutions to (1) behave have a bearing on how the solutions to the approximation (2) behave.

Solutions to (1) are parameterized by pairs (t, j) belonging to particular subsets of $\mathbb{R}^2$. Parameterization by t only, usually done when treating hybrid systems as impulsive differential equations [29] and when studying switching systems [30], does not allow for satisfactory modeling of discrete dynamics. The use of pairs (t, j) also facilitates favorable structural properties of the sets of solutions to (1) when the hybrid inclusion is subject to Assumption 2.1.

A set $E \subset \mathbb{R}^2$ is a *compact hybrid time domain* if

$$E = ([t_0, t_1], 0) \cup ([t_1, t_2], 1) \cup \ldots \cup ([t_{J-1}, t_J], J-1) \tag{3}$$

where $J \in \mathbb{N}$ and $0 = t_0 \le t_1 \le t_2 \le \cdots \le t_J$ form a finite sequence of real numbers. A set $E \subset \mathbb{R}^2$ is a *hybrid time domain* if it is the union of a nondecreasing sequence $E_1 \subset E_2 \subset E_3 \subset \ldots$ of compact hybrid time domains. Thus, a hybrid time domain differs from (3) by allowing for an infinite number of intervals or, if the number is finite, by allowing for the last interval to be open, and possibly unbounded, to the right.

A function $\phi : \mathrm{dom}\, \phi \to \mathbb{R}^n$ is a *solution to* (1) if dom ϕ is a hybrid time domain, $\phi(0,0) \in C \cup D$, and

Flows: for every $([t, t'], j) \subset \mathrm{dom}\, \phi$ with $t < t'$, $t \mapsto \phi(t, j)$ is absolutely continuous on $[t, t']$, $\phi(t, j) \in C$ for all $t \in [t, t']$, and $\dot{\phi}(t, j) \in F(\phi(t, j))$ for almost all $t \in [t, t']$;

Jumps: for every $(t, j), (t, j+1) \in \mathrm{dom}\, \phi$, $\phi(t, j) \in D$ and $\phi(t, j+1) \in G(\phi(t, j))$.

A solution ϕ is *maximal* if it cannot be extended and *complete* if dom ϕ is unbounded. As in [14], in this work hybrid time domains begin at $(0, 0)$, so maximality and completeness is in the forward-in-hybrid-time sense. Generalizing to "negative" hybrid times is not necessary here, nor is defining some notion of backward-in-hybrid-time completeness.

In classical differential equations, many results are facilitated by continuous dependence of solutions on initial conditions or the fact that the limit of a uniformly convergent sequence of solutions is a solution. Here, different solutions to (1) may be defined on different hybrid time domains, and the appropriate versions of the mentioned desired properties cannot rely on uniform distance or uniform convergence. Graphical convergence and graphical distance have been successfully used for this purpose [14]. Roughly, a key result that facilitates what was outlined in item (c) above states that, under the hybrid basic conditions in Assumption 2.1, the (graphical) limit of a (graphically) convergent sequence of solutions to (1) is a solution to (1).

For a solution ϕ to (1), its *omega-limit* omega(ϕ) is the set defined by

$$\mathrm{omega}(\phi) := \left\{ x \in \mathbb{R}^n \,:\, \exists (t_i, j_i) \in \mathrm{dom}\, \phi \text{ s.t. } t_i + j_i \to \infty,\, \phi(t_i, j_i) \to x \right\}. \tag{4}$$

These omega-limit sets will be used in Section 7 to define Morse decompositions for global attractors of hybrid systems (1). If the solution ϕ is complete and bounded, then omega(ϕ) is nonempty, compact, weakly forward invariant (i.e., for every $x \in$ omega(ϕ) there exists a complete solution to (1) from x that remains in omega(ϕ)) and weakly backward invariant, in an appropriate sense. These properties lead to extensions of the LaSalle's Invariance Principle to hybrid inclusions [31, 14].

Occasionally, $\mathcal{H}$ may be used to represent the hybrid system (1). The notation $\mathcal{S}^{\mathcal{H}}$ represents the set of all maximal solutions to (1), $\mathcal{S}^{\mathcal{H}}(x)$ is the set of those $\phi \in \mathcal{S}^{\mathcal{H}}$ for which $\phi(0,0) = x$, and for a set $K \subset \mathbb{R}^n$, $\mathcal{S}^{\mathcal{H}}(K) := \bigcup_{x \in K} \mathcal{S}^{\mathcal{H}}(x)$. Then, for a set $K \subset \mathbb{R}^n$, the *Omega-limit from K for $\mathcal{H}$* is the set defined by

$$\mathrm{Omega}(K) := \left\{ x \in \mathbb{R}^n \; : \; \begin{array}{c} \exists\, \phi_i \in \mathcal{S}^{\mathcal{H}}(K),\; (t_i, j_i) \in \mathrm{dom}\; \phi_i \\ \text{s.t. } t_i + j_i \to \infty,\; \phi_i(t_i, j_i) \to x \end{array} \right\}. \tag{5}$$

This set can be, and often is, different from $\bigcup_{\phi \in \mathcal{S}^{\mathcal{H}}(K)}$ omega(ϕ), but also has favorable invariance properties. In particular, if Omega(K) is a nonempty subset of the interior of K, it is a strongly forward —in the sense that all solutions from it remain in it—and weakly backward invariant attractor, in the sense to be specified and used in Section 7. These Omega-limit sets will be used in Section 7 to define a global attractor for a hybrid system (1).

The following assumption is needed to ensure that the discussion of asymptotic properties of solutions to (1) is not vacuous and that the existence of attractors, as presented in Section 7, can be ensured.

Assumption 2.2. *There exists at least one complete solution to the hybrid system* (1) *and every maximal solution to* (1) *is complete.*

While Assumption 2.2 implies that $C \cup D \neq \emptyset$, it does not imply that $C \cup D = \mathbb{R}^n$. This flexibility, among other things, facilitates modeling in the format (1) of hybrid automata and other hybrid systems with logical modes; see, for example, [14, Section 1.4], or Example 2.5 below.

Under the basic conditions in Assumption 2.1, sufficient conditions for the existence of solutions from every $x \in C \cup D$ and for the completeness of every maximal solution to (1) are:

(i) $G(x) \subset C \cup D$ for every $x \in D$;

(ii) for every $x \in C \setminus D$, there exist $T > 0$ and a solution to the continuous-time component (1a) of (1) on $[0, T]$, which is implied by the following condition: there exists a neighborhood U of x such that, for every $x' \in U \cap C$,

$$F(x') \cap T_C(x') \neq \emptyset,$$

where $T_C(x')$ is the tangent cone to C at x', see [24, Definition 6.1] or [14, Definition 5.12];

(iii) solutions to the continuous-time component (1a) of (1) do not blow up in finite time, i.e., there is no $\phi : [0, T) \to \mathbb{R}^n$ such that $\phi(t) \in C$ for all $t \in [0, T)$, $\dot{\phi}(t) \in F(\phi(t))$ for almost all $t \in [0, T)$, and such that $\|\phi(t)\| \to \infty$ as $t \nearrow T$.

This result and a proof can be found in [14, Proposition 6.10].

The section concludes with examples of hybrid systems modeled as (1). Each example satisfies Assumption 2.1, and each example is revisited later in the paper, for the purpose of stochastic approximation.

Example 2.3 (Academic example). *Consider the hybrid system on $\mathbb{R}$, given by*

$$\begin{aligned} C &= (-\infty, 5] & F(x) &= x(x-1)(x-2)(x-4) \\ D &= [1.5, 2.5] \cup [5, \infty) & G([1.5, 2.5]) &= \{3\}, \ \ G([5, \infty)) = \{0\} \end{aligned} \tag{6}$$

and represented in Figure 1. The directions of flow are represented by the solid arrows and the jumps are represented by the dotted arrows. The flow outside of C is indicated but it is not relevant. Note that F and G are functions, but solutions from some initial conditions are not unique anyway, thanks to the overlap of the sets C and D. For example, one solution ϕ from 1.1 *has* $\operatorname{dom} \phi = ([0, \infty), 0)$ *and flows asymptotically to* 2. *Another solution ψ from* 1.1 *has* $\operatorname{dom} \psi = ([0, T], 0) \cup ([T, \infty), 1)$ *and flows from* 1.1 *to* 1.7, *jumps to* 3, *and then flows asymptotically to* 2. *From any initial condition x_0 in* $(2, 2.5]$ *there exists a periodic, in the hybrid sense, solution that jumps to* 3, *flows to x_0, jumps back to* 3, *etc. For such a solution, its omega-limit is* $[x_0, 3]$. *Note also that* $\mathrm{Omega}([1, 3]) = [1, 3]$ *while* $\bigcup_{\phi \in \mathcal{S}^{\mathcal{H}}([1,3])} \mathrm{omega}(\phi) = \{1\} \cup [2, 3]$, *and that for any arbitrarily large but bounded set K,* $\mathrm{Omega}(K) = [0, 5]$.

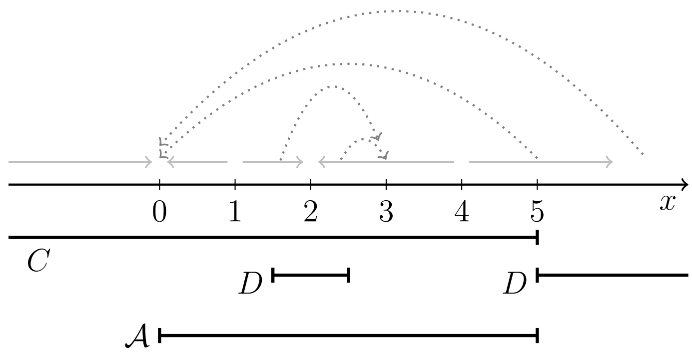

Figure 1: A graphical representation of the hybrid dynamics in (6).

The Bouncing Ball dynamical system has been motivational and illustrative for some developments in the dynamical and control systems literature. In its usual formulation, it features *Zeno behavior*: the occurrence of infinitely many jumps in finite (Zeno) time. In the control literature, in the early [32], it illustrated how solutions to hybrid systems can be simulated past their Zeno times; in [33], it was cast and studied in the hybrid inclusion (1) format; it illustrated hybrid dynamics with multiple Zeno times [34]; etc. The Bouncing Ball remains a curiosity when simulated, as the recent [35] shows and analyzes. For many further, related works, see the references in these few cited papers.

Example 2.4 (Bouncing Ball). *A hybrid model of a bouncing ball has state $x = (h, \dot{h}) \in \mathbb{R}^2$ where h is the height of the ball above the floor and $\dot{h}$ is its velocity, and the following data:*

$$C := \mathbb{R}_{\geq 0} \times \mathbb{R} \tag{7a}$$
$$F(x) := (x_2, -\gamma) \qquad \forall x \in C \tag{7b}$$
$$D := \{0\} \times \mathbb{R}_{\leq 0} \tag{7c}$$
$$G(x) := (0, -\delta x_2) \qquad \forall x \in D \tag{7d}$$

where $\gamma > 0$ and $\delta \in (0, 1)$.

The hybrid basic conditions in Assumption 2.1 impose that the flow set C and jump set D for the Bouncing Ball model are closed. The jump set D being closed induces a solution at the origin that jumps endlessly without ever flowing. This solution is the graphical limit of any sequence of solutions starting closer and closer to the origin in $C \cup D$; these latter solutions experience some small amount of flow between jumps. As mentioned earlier, that the limit of a graphically convergent sequence of solutions is another solution is an important, useful feature of the hybrid basic conditions imposed in Assumption 2.1.

Example 2.5 (Combinatorial optimization). *Consider the optimization problem:*

$$\min_{z \in \mathbb{R}^m, q \in \mathcal{Q}} f_q(z), \tag{8}$$

where $\mathcal{Q} \subset \mathbb{Z}$ is a finite index set and, for each $q \in \mathcal{Q}$, $f_q : \mathbb{R}^m \to \mathbb{R}$ is differentiable. Consider a hybrid optimization algorithm that alternates between continuous-time gradient flow to minimize $z \mapsto f_q(z)$ for fixed $q \in \mathcal{Q}$ and discrete updates to minimize $q \mapsto f_q(z)$ for fixed $z \in \mathbb{R}^m$. To make sure that the number of discrete updates is finite in every finite time window, suppose the algorithm limits the frequency of the updates of q by an average dwell-time constraint

$$N([s,t]) \leq \delta(t-s) + N_0 \tag{9}$$

where $\delta > 0$, $N_0 \in \mathbb{N}$ and $N([s,t])$ denotes the number of updates of q in the time interval $[s,t]$ with $t \geq s \geq 0$. In light of [36, Proposition 1.1], such a class of algorithms is covered by a hybrid system with state $x := (z,q,\tau) \in \mathbb{R}^{m+2}$ and data

$$C := \mathbb{R}^m \times \mathcal{Q} \times [0, N_0] \tag{10a}$$

$$F(x) := \begin{bmatrix} -\nabla f_q(z) \\ 0 \\ [\varepsilon, \delta] \end{bmatrix} \tag{10b}$$

$$D := \mathbb{R}^m \times \mathcal{Q} \times [1, N_0] \tag{10c}$$

$$G(x) := \begin{bmatrix} z \\ \{q \in \mathcal{Q} : f_q(z) \leq f_p(z) \quad \forall p \in \mathcal{Q}\} \\ \tau - 1 \end{bmatrix} \tag{10d}$$

where $\varepsilon \in (0, \delta)$ appears in (10b) to guarantee that at least one jump occurs every N_0/ε time units. It is not difficult to verify that Assumption 2.2 holds for the system with this data.

The average dwell-time mechanism in Example 2.5 allows up to N_0 jumps over arbitrarily small time intervals. This is seen by considering the situation where τ is initialized to the value N_0 and then the jump map is applied N_0 times in the arbitrarily small time interval until τ exits the interval $[1, N_0]$. The data satisfies the hybrid basic conditions in Assumption 2.1. Therefore, in the absence of finite escape times, the graphical limit of every graphically convergent sequence of solutions is the graph of a solution (see [14, Chapter 6]). This means that the average dwell-time mechanism also allows for up to N_0 jumps over an interval of zero length. This is also seen by initializing τ at N_0 and taking N_0 jumps without any flow time in between. Such behavior is one possible motivation for the use of hybrid time domains, which keep track of both the amount of flow time t and the number of jumps j that have occurred while reaching the current state.

Example 2.6 (Combinatorial optimization on a torus). *Consider the optimization problem in Example 2.5, where z is the state of a control system that evolves on the torus $\mathbb{S}^1 \times \mathbb{S}^1 \subset \mathbb{R}^4$, where $\mathbb{S}^1 \subset \mathbb{R}^2$ denotes the unit circle, with input $u \in \mathbb{R}^2$, i.e.,*

$$z \in \mathbb{S}^1 \times \mathbb{S}^1 \quad \dot{z} = R(u)z, \;\; R(u) := \mathrm{diag}(u_1 R_0, u_2 R_0), R_0 := \begin{bmatrix} 0 & 1 \\ -1 & 0 \end{bmatrix} \tag{11}$$

so that the z variable has inertia and the optimization problem adds the constraint $z \in \mathbb{S}^1 \times \mathbb{S}^1$. Consider the class of optimization algorithms modeled by the hybrid system with state $x = (z, u, q, \tau) \in \mathbb{R}^8$ and data

$$C := \mathbb{S}^1 \times \mathbb{S}^1 \times \mathbb{R}^2 \times \mathcal{Q} \times [0, N_0] \tag{12a}$$

$$F(x) := \begin{bmatrix} R(u)z \\ -u - \begin{bmatrix} \frac{\partial f_q(z)}{\partial z_1} R_0 z_1 \\ \frac{\partial f_q(z)}{\partial z_2} R_0 z_2 \end{bmatrix} \\ 0 \\ [\varepsilon, \delta] \end{bmatrix} \tag{12b}$$

$$D := \mathbb{S}^1 \times \mathbb{S}^1 \times \mathbb{R}^2 \times \mathcal{Q} \times [1, N_0] \tag{12c}$$

$$G(x) := \begin{bmatrix} z \\ u \\ \{q \in \mathcal{Q} : f_q(z) \le f_p(z) \quad \forall p \in \mathcal{Q}\} \\ \tau - 1 \end{bmatrix}. \tag{12d}$$

The z dynamics render the torus forward invariant. The (z, u) dynamics are similar to gradient flow of Example 2.5 but with inertia; namely, that part of the flow dynamics works to minimize the function $(z, u) \mapsto 0.5u^T u + f_q(z)$ rather than the function $z \mapsto f_q(z)$. The rest of the dynamics implements the average dwell-time constraint on jumps that update q to minimize $s \mapsto f_s(z)$, like in Example 2.5.

3. A short review of relevant probability concepts

Relevant concepts from probability theory and stochastic processes are reviewed here. For more details, see [37] and [24, Chapter 14], for example.

The triple $(\Omega, \mathcal{F}, \mathbb{P})$ is used to denote a probability space, where Ω is a sample space, $\mathcal{F}$ is a σ-field of subsets of Ω, and $\mathbb{P}$ is a probability measure on $(\Omega, \mathcal{F})$. A property depending on $\omega \in \Omega$ is said to hold *almost surely*, or *for almost all* $\omega \in \Omega$, if there exists a set $\widehat{\Omega} \subset \Omega$ satisfying $\mathbb{P}(\widehat{\Omega}) = 1$ such that the property holds for all $\omega \in \widehat{\Omega}$.

A sequence $\{\mathcal{F}_k\}_{k=0}^{\infty}$ of sub-σ-fields of $\mathcal{F}$ is a *filtration* of $(\Omega, \mathcal{F})$ if

$$\mathcal{F}_k \subset \mathcal{F}_{k+1} \quad \forall k \in \mathbb{Z}_{\geq 0}.$$

Let $\{\mathbf{y}_k\}_{k=1}^{\infty}$ be a sequence of random variables taking values in $\mathbb{R}^m$. That is, for each $k \in \mathbb{Z}_{\geq 1}$, $\mathbf{y}_k : \Omega \to \mathbb{R}^m$ is measurable, i.e., for each set B in the Borel σ-field on $\mathbb{R}^m$,

$$\mathbf{y}_k^{-1}(B) := \{\omega \in \Omega : \mathbf{y}_k(\omega) \in B\} \in \mathcal{F}.$$

The sequence of random variables $\{\mathbf{y}_k\}_{k=1}^{\infty}$ is said to be adapted to the filtration $\{\mathcal{F}_k\}_{k=0}^{\infty}$ if, for each $k \in \mathbb{Z}_{\geq 1}$, $\mathbf{y}_k$ is $\mathcal{F}_k$-measurable, i.e., $\mathbf{y}_k^{-1}(B) \in \mathcal{F}_k$ for each set B in the Borel σ-field on $\mathbb{R}^m$.

The natural filtration of the sequence of random variables $\{\mathbf{y}_k\}_{k=1}^{\infty}$ is the sequence $\{\mathcal{F}_k\}_{k=0}^{\infty}$ where $\mathcal{F}_k$ is the σ-field generated by $\{\mathbf{y}_i\}_{i=1}^{k}$, i.e., the smallest σ-field that makes all y_i, $i \in \{1, \ldots, k\}$, measurable, and $\mathcal{F}_0 := \{\emptyset, \Omega\}$. A sequence of random variables is adapted to its natural filtration.

A conditional expectation of a random variable $\mathbf{y} : \Omega \to \mathbb{R}^m$ with finite expectation given a sub-σ-field $\mathcal{H}$ is any $\mathcal{H}$-measurable function denoted $\mathbb{E}[\,\mathbf{y} \,|\, \mathcal{H}\,] : \Omega \to \mathbb{R}^m$ satisfying

$$\int_H \mathbb{E}[\,\mathbf{y} \,|\, \mathcal{H}\,]\, d\mathbb{P} = \int_H \mathbf{y} d\mathbb{P} \qquad \forall H \in \mathcal{H}.$$

When $\mathbf{y}, \mathbf{z} : \Omega \to \mathbb{R}$ are random variables such that $\mathbf{y}$ is measurable with respect to the sub-σ-field $\mathcal{H}$ of $\mathcal{F}$ then, almost surely,

$$\mathbb{E}\left[\, \mathbf{y}\mathbf{z} \,|\, \mathcal{H} \,\right] = \mathbf{y}\, \mathbb{E}\left[\, \mathbf{z} \,|\, \mathcal{H} \,\right].$$

Note that here, and in many other instances in the paper, the explicit dependence of random variables on ω is suppressed for notational convenience.

A set-valued mapping $M : \Omega \rightrightarrows \mathbb{R}^m$ is said to be $\mathcal{F}$-measurable (or simply measurable when the σ-field $\mathcal{F}$ is clear) if, for each open set $\mathcal{O} \subset \mathbb{R}^m$,

$$M^{-1}(\mathcal{O}) := \{\omega \in \Omega : M(\omega) \cap \mathcal{O} \neq \emptyset\} \in \mathcal{F}.$$

4. Solutions to stochastic approximation

For the stochastic approximation (2), a solution concept is used that is similar to those used in [17] and [18]. Several examples of stochastic approximations appear later in Section 6.

A solution of the stochastic approximation (2) comprises sample paths, each of which has a domain that is discrete. A set $E \subset \mathbb{R}^2$ is a *compact hybrid discrete time domain* if

$$E = \Big(\{k_0, \ldots, k_1\}, 0\Big) \cup \Big(\{k_1, \ldots, k_2\}, 1\Big) \cup \cdots \cup \Big(\{k_{J-1}, \ldots, k_J\}, J-1\Big)$$

where $J \in \mathbb{N}$ and $0 = k_0 \leq k_1 \leq \cdots \leq k_J$ form a finite sequence of integers. It is a *hybrid discrete time domain* if it is the union of a nondecreasing sequence of compact discrete time domains. It is a *hybrid discrete time domain that is unbounded in the t direction* if the set of all k's defining it is unbounded. Since k's play a role that is similar to the role of t in a hybrid time domain, the use of the "t direction" is justified.

Let $\{h_k\}_{i=1}^{\infty}$ be a sequence of positive numbers. As discussed in Section 3, let $(\Omega, \mathcal{F}, \mathbb{P})$ be a probability space, let $\{\mathbf{y}_k\}_{k=1}^{\infty}$ be a sequence of random variables defined on that probability space and let $\{\mathcal{F}_k\}_{k=0}^{\infty}$ be the natural filtration of $\{\mathbf{y}_k\}_{k=1}^{\infty}$. Let $\mathcal{X}$ denote the set of set-valued mappings from $\mathbb{R}^2$ to $\mathbb{R}^n$ with a closed and nonempty graph.

A mapping $\mathbf{x} : \Omega \to \mathcal{X}$ is said to be a *candidate solution* of (2) if, for each $\omega \in \Omega$, the domain of $\mathbf{x}(\omega)$ is a hybrid discrete time domain and, for each $k \in \mathbb{Z}_{\geq 0}$, the set-valued mapping

$$\omega \mapsto \text{graph}\,(\mathbf{x}(\omega)) \cap (\{k\} \times \mathbb{Z}_{\geq 0} \times \mathbb{R}^n) \tag{13}$$

is $\mathcal{F}_k$-measurable.

To save on notation, we suppress the dependence of $\mathbf{x}$ on ω when referring to the value of $\mathbf{x}$ at a particular time; that is, we avoid writing $\mathbf{x}(\omega)(t,j)$.

A candidate solution of (2) is a *solution* of (2) if, for almost all $\omega \in \Omega$,

1. $(k,j),(k+1,j) \in \operatorname{dom} \mathbf{x}(\omega)$ implies:
   $\mathbf{x}(k,j) \in C$ and $\mathbf{x}(k+1,j) - \mathbf{x}(k,j) \in h_{k+1}\widehat{F}\left(\mathbf{x}(k,j), \mathbf{y}_{k+1}, h_{k+1}\right)$
2. $(k,j),(k,j+1) \in \operatorname{dom} \mathbf{x}(\omega)$ implies:
   $\mathbf{x}(k,j) \in D$ and $\mathbf{x}(k,j+1) \in G(\mathbf{x}(k,j))$.

A solution is *complete* if, for almost all $\omega \in \Omega$, the domain of $\mathbf{x}(\omega)$ is unbounded. A solution is *complete in the t direction* if, for almost all $\omega \in \Omega$, the domain of $\mathbf{x}(\omega)$ is unbounded in the t direction. Conditions that guarantee the existence of complete, and complete in the t direction, solutions to (2) are given next. The key condition of the following proposition is motivated by [24, Corollary 14.14], which says that if $\mathbf{x} : \Omega \to \mathbb{R}^n$ is measurable and the mapping $\omega \mapsto \operatorname{graph}(M(\cdot,\omega)) \subset \mathbb{R}^m \times \mathbb{R}^n$ is measurable with closed values then the (set-valued) mapping $\omega \mapsto M(\mathbf{x}(\omega),\omega)$ is measurable with closed values. This measurability, in turn, permits a measurable selection [24, Corollary 14.6] that can be used to construct solutions.

Proposition 4.1. *Suppose the data $(C, \widehat{F}, D, G)$, the step sizes $\{h_k\}_{k=1}^{\infty}$, and noise $\{\mathbf{y}_k\}_{k=1}^{\infty}$ satisfy the following conditions:*

1. *$x \mapsto G(x) \cap (C \cup D)$ is outer semicontinuous with nonempty values on D.*
2. *for almost all $\omega \in \Omega$, all $k \in \mathbb{Z}_{\geq 0}$, and all $x \in C$,*
   $\Big(\{x\} + h_{k+1}\widehat{F}(x, \mathbf{y}_{k+1}(\omega), h_{k+1})\Big) \cap (C \cup D)$ is nonempty, and
3. *for each $k \in \mathbb{Z}_{\geq 0}$, the mapping*
   $$\omega \mapsto \Big\{(x,z) \in \mathbb{R}^n \times \mathbb{R}^n : z \in \Big(x + h_{k+1}\widehat{F}(x, \mathbf{y}_{k+1}(\omega), h_{k+1})\Big) \cap (C \cup D)\Big\}$$
   is $\mathcal{F}_{k+1}$-measurable with closed values.

Then, for any initial condition in $C \cup D$, there exists a complete solution to (2)*. If, in addition, there are no complete solutions to* (2b)*, equivalently* (1b)*, then each complete solution is complete in the t direction.*

Proof. It is enough to prove that, for each $\mathcal{F}_k$-measurable mapping $\mathbf{z}_c : \Omega \to C$, the mapping

$$\omega \mapsto \left(\mathbf{z}_c(\omega) + h_{k+1}\widehat{F}(\mathbf{z}_c(\omega), \mathbf{y}_{k+1}(\omega), h_{k+1})\right) \cap (C \cup D) \tag{14}$$

admits an $\mathcal{F}_{k+1}$-measurable selection defined for almost all $\omega \in \Omega$ and, for each $\mathcal{F}_k$-measurable mapping $\mathbf{z}_d : \Omega \to D$, the mapping

$$\omega \mapsto G(\mathbf{z}_d(\omega)) \cap (C \cup D) \tag{15}$$

admits an $\mathcal{F}_k$-measurable selection defined for almost all $\omega \in \Omega$. By assumption, both mappings are nonempty for almost all $\omega \in \Omega$. Thus, according to [24, Corollary 14.6], it is enough to verify that the mapping in (14) is $\mathcal{F}_{k+1}$-measurable and that the mapping in (15) is $\mathcal{F}_k$-measurable, each with closed values. Both of these properties follow from the assumptions of the proposition together with [24, Corollary 14.14]. $\square$

We now consider stochastic approximations of some of the examples introduced earlier.

Example 4.2 (Bouncing Ball, stochastically approximated). *Consider a stochastic approximation of the Bouncing Ball hybrid system from Example 2.4, with*

$$\widehat{F}(x, y^+, h^+) := F(x) - (0, y^+) \tag{16}$$

and exponentially distributed noise $\{\mathbf{y}_k\}_{k=1}^{\infty}$. *To guarantee the existence of complete solutions we are compelled to inflate the set* D *to the set* $\widehat{D}$ *given as*

$$\widehat{D} := \mathbb{R}_{\leq 0} \times \mathbb{R}_{\leq 0}. \tag{17}$$

This choice is made because the jumps from the flow set have the form

$$x_1^+ - x_1 = h_{k+1}x_2 \tag{18a}$$

$$x_2^+ - x_2 = h_{k+1}\left(-\gamma - y^+\right) \tag{18b}$$

with $y^+ \geq 0$, *resulting in* $(x_1^+, x_2^+) \in C \cup \widehat{D}$. *Note that this inflation of the jump set does not change the solutions of the underlying hybrid system, other than for initial conditions* $x(0,0) \in \widehat{D} \backslash D$, *which can jump to* $g(x(0,0))$ *and thereafter remain in* $C \cup D$.

Example 4.3 (Combinatorial optim. with stochastic gradients). *Consider the optimization problem from Example 2.5 where, for each $q \in \mathcal{Q}$, $f_q : \mathbb{R}^m \to \mathbb{R}$ is given by*

$$f_q(z) := \frac{1}{N_q} \sum_{i=1}^{N_q} f_{i,q}(z) \tag{19}$$

where N_q is a positive integer and, for each $q \in \mathcal{Q}$, and $i \in \{1, \ldots, N_q\}$, $f_{i,q}$ is differentiable. For the sake of the stochastic approximation of (C, F, D, G) in (10), to anticipate simulated flows that leave the interval $[0, N_0]$, the jump set component corresponding to τ can be inflated from $[1, N_0]$ to $[1, N_0 + \delta\overline{h}]$, where $\overline{h}$ is an upper bound on the step sizes, and the jump map for τ^+ from $\tau - 1$ to $\min\{\tau, N_0\} - 1$, without changing the nature of the solutions. Let the stochastic approximation of F be given as

$$\widehat{F}(x, y^+) := \begin{bmatrix} -\nabla f_{(y^+ \bmod N_q), q}(z) \\ 0 \\ [\varepsilon, \delta] \end{bmatrix} \tag{20}$$

where the noise sequence $\{\mathbf{y}_k\}_{k=1}^{\infty}$ is iid and uniformly distributed on $\{1, \ldots, N\}$ where N is the least common multiple of $\{N_q\}_{q \in \mathcal{Q}}$; that is, $\widehat{F}$ implements a stochastic gradient algorithm where, at each step and for each $q \in \mathcal{Q}$, the probability of implementing a gradient step in the direction $-\nabla f_{i,q}(z)$ for $i \in \{1, \ldots, N_q\}$ is $1/N_q$.

Example 4.4 (CO on a torus with stochastic finite differences). *Consider a stochastic approximation of the combinatorial optimization (abbreviated to CO in the title of this example) dynamics on a torus given in Example 2.6. The jump set and jump dynamics can be inflated as in Example 4.3. Let the stochastic approximation of F be given as*

$$\widehat{F}(x, y^+, h^+) := \tag{21}$$
$$\begin{bmatrix} \dfrac{1}{h^+}\Big(\exp\Big(R(u)h^+\Big) - I\Big)\, z \\ -u - \dfrac{y^+}{h^+}\Big(f_q\Big(\exp\big(R(y^+h^+)\, z\Big) - f_q\Big(\exp\Big(-R(y^+h^+)\Big)\, z\Big)\Big) \\ 0 \\ [\varepsilon, \delta] \end{bmatrix}$$

where the noise sequence $\{\mathbf{y}_k\}_{k=1}^{\infty}$ is iid and uniformly distributed on the 2-point set $\{(1,0)\} \cup \{(0,1)\} \subset \mathbb{R}^2$. The mapping $\widehat{F}$ exactly simulates the z component of the solutions of the hybrid system (12) during flows, thereby keeping the state on the torus. On the other hand, it provides a "noisy" Euler approximation for the flows of u, approximating a gradient through stochastic finite differences. In this way, the mapping $\widehat{F}$ can be interpreted as providing a model of a sample-and-hold, stochastic, hybrid extremum-seeking control on the torus. For other results on (non-stochastic) hybrid extremum seeking, see [38] for example.

5. A hybrid version of [5, Theorem 1]

In [5], the authors provide a now widely-cited result on convergence properties for what they call "almost" supermartingales. Their result generalizes Doob's classical convergence results for supermartingales [39, Chapter VII]; see also [40, Chapter XI] for a more recent treatment. Doob's result corresponds, in the lemma below, to the special case where $\mathbf{w}_i$, $\mathbf{v}_i$ and $\mathbf{u}_i$ are identically zero. That case generalizes to the stochastic setting the classical monotone convergence theorem. The generalization provided in [5] has been used extensively in the analysis of stochastic gradient descent and other aspects of optimization-based, large-scale machine learning; for context, see the review [41, Inset 4.1]. The main result of [5], which must be extended to adapted mappings defined on hybrid discrete time domains that are complete in the t direction, is the following:

Lemma 5.1. *Let $\{\mathcal{F}_i\}_{i=0}^{\infty}$ be a filtration of the probability space $(\Omega, \mathcal{F}, \mathbb{P})$. For each $i \in \mathbb{Z}_{\geq 0}$, let the random variables $\mathbf{z}_i, \mathbf{w}_i, \mathbf{v}_i$ and $\mathbf{u}_i$ be non-negative, $\mathcal{F}_i$-measurable, and such that, for almost all $\omega \in \Omega$,*

$$\mathbb{E}\left[\mathbf{z}_{i+1} \,|\, \mathcal{F}_i\right] \leq \mathbf{z}_i \left(1 + \mathbf{w}_i\right) + \mathbf{v}_i - \mathbf{u}_i. \tag{22}$$

Then, for almost all $\omega \in \Omega_s$, where

$$\Omega_s := \left\{ \omega \in \Omega : \sum_{i=0}^{\infty} \mathbf{w}_i < \infty, \sum_{i=0}^{\infty} \mathbf{v}_i < \infty \right\}, \tag{23}$$

we have that $\lim_{i\to\infty} \mathbf{z}_i$ exists and is finite and $\sum_{i=0}^{\infty} \mathbf{u}_i < \infty$.

This result is now extended to the hybrid setting. Following the notation introduced earlier, let $\mathcal{X}$ denote the set of set-valued mappings from $\mathbb{R}^2$ to $\mathbb{R}$ with a closed and nonempty graph. Also, let $\mathcal{T}$ be the set of all mappings $\mathbf{z} : \Omega \to \mathcal{X}$ such that, for almost all $\omega \in \Omega$, dom $\mathbf{z}(\omega)$ is a hybrid discrete time domain that is unbounded in the t direction, $\operatorname{rge} \mathbf{z}(\omega) \subset \mathbb{R}_{\geq 0}$, and for each $k \in \mathbb{Z}_{\geq 0}$, the set-valued mapping

$$\omega \mapsto \operatorname{graph}(\mathbf{z}(\omega)) \cap \Big(\{k\} \times \mathbb{Z}_{\geq 0} \times \mathbb{R} \Big) \tag{24}$$

is $\mathcal{F}_k$-measurable. Given $\mathbf{z} \in \mathcal{T}$, for each $k \in \mathbb{Z}_{\geq 0}$, define

$$\underline{\jmath}_k(\omega) := \min_{(k,j) \in \operatorname{dom}\, \mathbf{z}(\omega)} j, \qquad \bar{\jmath}_k(\omega) := \max_{(k,j) \in \operatorname{dom}\, \mathbf{z}(\omega)} j. \tag{25}$$

The maximum used to specify $\bar{\jmath}_k(\omega)$ is well defined since dom $\mathbf{z}(\omega)$ is unbounded in the t direction and thus, due to the nature of a hybrid discrete time domain, for a given k the set $\{j \in \mathbb{Z}_{\geq 0} : (k,j) \in \operatorname{dom} \mathbf{z}(\omega)\}$ is finite. Also note that, due to the definition of a hybrid discrete time domain, for each $k \in \mathbb{Z}_{\geq 0}$ and all $\omega \in \Omega$, $\underline{\jmath}_{k+1}(\omega) = \bar{\jmath}_k(\omega)$. The next lemma asserts that, for a hybrid signal $\mathbf{z} \in \mathcal{T}$, if for each $k \in \mathbb{Z}_{\geq 0}$ we focus on the values of $\mathbf{z}(k, \cdot)$ at either $\bar{\jmath}_k$ or $\underline{\jmath}_k$, we get adapted sequences.

Lemma 5.2. *Let $\mathbf{z} \in \mathcal{T}$ and let $\bar{\jmath}_k$ and $\underline{\jmath}_k$ be generated from $\mathbf{z}$ as in (25). Then $k \mapsto \mathbf{z}(k, \bar{\jmath}_k)$ and $k \mapsto \mathbf{z}(k, \underline{\jmath}_k)$ are adapted to the filtration $\{\mathcal{F}_k\}_{k=0}^{\infty}$.*

Proof. First, it follows from [24, Exercise 14.18] with $m = 0$, $d = 1$, $n = 2$, $T := \Omega$, S the mapping in (24), x playing the role of (k,j), and w playing the role of z that the mapping

$$\omega \mapsto (\operatorname{dom} \mathbf{z}(\omega)) \cap (\{k\} \times \mathbb{Z}_{\geq 0}) =: \Psi_k(\omega) \tag{26}$$

is $\mathcal{F}_k$-measurable. Then $\mathcal{F}_k$-measurability of $\underline{\jmath}_k$ in (25) follows from [24, Theorem 14.3(j)] with $x = (k, 0)$. $\mathcal{F}_k$-measurability of $\bar{\jmath}_k$ in (25) follows like in [42, Ch. 1, Example 1.3(i)] by noting that $\bar{\jmath}_k(\omega) = \sup\{|\psi - (k,0)| : \psi \in \Psi_k(\omega)\}$ is measurable since $\{\omega \in \Omega : \bar{\jmath}_k(\omega) > \alpha\}$ equals the set of $\omega \in \Omega$ such that $\Psi_k(\omega)$ intersects an open set being the complement of the closed ball of radius α centered at $(k, 0)$. Finally, the result of the lemma follows from the assumption that the mapping in (24) is measurable together with [24, Corollary 14.14] with $M(\omega, \cdot) := \mathbf{z}(k, \cdot)(\omega)$. $\square$

Here is the needed generalization of Lemma 5.1. It closely resembles that lemma but must account both for jumps that arise from approximating the flows (like those considered in Lemma 5.1) and jumps that are an inherent part of the hybrid system, while being cognizant of the nature of the time domains on which sample paths are defined. In the theorem statement, multiple mappings, each belonging to $\mathcal{T}$, are said to be *compatible* if their time domains agree almost surely.

Theorem 5.3. *Let $\{\mathcal{F}_i\}_{i=0}^{\infty}$ be a filtration of the probability space $(\Omega, \mathcal{F}, \mathbb{P})$. Let $\mathbf{z}, \mathbf{w}, \mathbf{v}$, $\mathbf{u}_c$ and $\mathbf{u}_d$ each belong to $\mathcal{T}$ and be compatible, and let $\bar{\jmath}_k$ and $\underline{\jmath}_k$ be generated from $\mathbf{z}$ as in (25). Suppose that, for each $k \in \mathbb{Z}_{\geq 0}$ and almost all $\omega \in \Omega$,*

$$\mathbb{E}\left[\mathbf{z}\left(k+1, \underline{\jmath}_{k+1}\right) \mid \mathcal{F}_k \right] \leq \mathbf{z}(k, \bar{\jmath}_k)\left(1 + \mathbf{w}(k, \bar{\jmath}_k)\right) + \mathbf{v}(k, \bar{\jmath}_k) - \mathbf{u}_c(k, \bar{\jmath}_k) \tag{27}$$

and

$$\mathbf{z}(k, j+1) \leq \mathbf{z}(k, j) - \mathbf{u}_d(k, j) \quad \forall j \in \left\{\underline{\jmath}_k, \ldots, \bar{\jmath}_k - 1\right\}. \tag{28}$$

Then, for almost all $\omega \in \Omega_s$, where

$$\Omega_s := \left\{ \omega \in \Omega : \quad \sum_{k=0}^{\infty} \mathbf{w}(k, \bar{\jmath}_k) < \infty, \quad \sum_{k=0}^{\infty} \mathbf{v}(k, \bar{\jmath}_k) < \infty \right\}, \tag{29}$$

we have that $\lim_{k+j \to \infty, (k,j) \in \mathrm{dom}\, \mathbf{z}} \mathbf{z}(k, j)$ exists and is finite, and

$$\sum_{k=0}^{\infty} \left(\mathbf{u}_c(k, \bar{\jmath}_k) + \sum_{j=\underline{\jmath}_k}^{\bar{\jmath}_k - 1} \mathbf{u}_d(k, j) \right) < \infty. \tag{30}$$

Proof. For each $k \in \mathbb{Z}_{\geq 0}$, define

$$\mathbf{s}_d(k) := \sum_{j=\underline{\jmath}_k}^{\bar{\jmath}_k - 1} \mathbf{u}_d(k, j), \tag{31}$$

with the convention that $\mathbf{s}_d(k) = 0$ when $\underline{\jmath}_k = \bar{\jmath}_k$. Like in the proof of Lemma 5.2, the sequence $\{\mathbf{s}_d(k)\}_{k=1}^{\infty}$ is adapted to the filtration $\{\mathcal{F}_k\}_{k=0}^{\infty}$.

For each $i \in \mathbb{Z}_{\geq 0}$, define

$$k(i) := \lfloor \frac{i}{2} \rfloor \tag{32a}$$

$$\mathbf{z}_i := \begin{cases} \mathbf{z}(k(i), \underline{\jmath}_{k(i)}) & i \text{ even} \\ \mathbf{z}(k(i), \bar{\jmath}_{k(i)}) & i \text{ odd.} \end{cases} \tag{32b}$$

$$\mathbf{w}_i := \begin{cases} 0 & i \text{ even} \\ \mathbf{w}(k(i), \bar{\jmath}_{k(i)}) & i \text{ odd.} \end{cases} \tag{32c}$$

$$\mathbf{v}_i := \begin{cases} 0 & i \text{ even} \\ \mathbf{v}(k(i), \bar{\jmath}_{k(i)}) & i \text{ odd.} \end{cases} \tag{32d}$$

$$\mathbf{u}_i := \begin{cases} \mathbf{s}_d(k(i)) & i \text{ even} \\ \mathbf{u}_c(k(i), \bar{\jmath}_{k(i)}) & i \text{ odd.} \end{cases} \tag{32e}$$

$$\mathcal{F}_i := \mathcal{F}_{k(i)}. \tag{32f}$$

It follows from Lemma 5.2 that the sequence $\{\mathbf{z}_i\}_{i=0}^{\infty}$ is adapted to the filtration $\{\mathcal{F}_i\}_{i=0}^{\infty}$. Moreover, it is not difficult to see that, for each $i \in \mathbb{Z}_{\geq 0}$ and almost all $\omega \in \Omega$,

$$\mathbb{E}\left[\mathbf{z}_{i+1} \,|\, \mathcal{F}_i\right] \leq \mathbf{z}_i \left(1 + \mathbf{w}_i\right) + \mathbf{v}_i - \mathbf{u}_i. \tag{33}$$

Next, note that the set Ω_s defined in (29) satisfies

$$\Omega_s = \left\{ \omega \in \Omega : \sum_{i=0}^{\infty} \mathbf{w}_i < \infty, \quad \sum_{i=0}^{\infty} \mathbf{v}_i < \infty \right\} \tag{34}$$

and the condition (30) is equivalent to the condition $\sum_{i=0}^{\infty} \mathbf{u}_i < \infty$. Thus, Lemma 5.1 gives the condition (30) and that $\lim_{i\to\infty} \mathbf{z}_i$ exists and is finite for almost all $\omega \in \Omega_s$. For such $\omega \in \Omega_s$, define

$$\mathbf{c} := \lim_{i\to\infty} \mathbf{z}_i. \tag{35}$$

Then inequality (28) and the definition of $\mathbf{z}_i$ in (32b) gives that

$$\begin{aligned} \mathbf{c} = \lim_{k\to\infty} \mathbf{z}_{2k+1} &\leq \liminf_{k\to\infty} \left(\min_{j \in \left\{\underline{\jmath}_k, \ldots, \bar{\jmath}_k\right\}} \mathbf{z}(k, j) \right) \\ &\leq \limsup_{k\to\infty} \left(\max_{j \in \left\{\underline{\jmath}_k, \ldots, \bar{\jmath}_k\right\}} \mathbf{z}(k, j) \right) \leq \lim_{k\to\infty} \mathbf{z}_{2k} = \mathbf{c} \end{aligned} \tag{36}$$

which establishes that $\lim_{k+j\to\infty,(k,j)\in\mathrm{dom}\,\mathbf{z}} \mathbf{z}(k,j)$ exists and is finite for almost all $\omega \in \Omega_s$. $\square$

6. Almost sure boundedness of solutions to (2)

In this section, conditions are pursued to guarantee that the solutions of the stochastic approximation (2) are almost surely bounded. The mean and variance of selections from $\widehat{F}$ are assumed to be bounded in a manner that resembles the bounds in [20, Assumption 1]. In contrast to [20, Assumption 1], there is no need for the bounding function V that appears in the assumption to be defined and have a Lipschitz gradient on all of $\mathbb{R}^n$. Instead, we ask for this latter property on line segments connecting $\mathbf{x}(k,j)$ to $\mathbf{x}(k+1,j)$ when $(k,j),(k+1,j) \in \mathrm{dom}\ \mathbf{x}(\omega)$, which is captured by the following set:

$$\begin{aligned} C^{[0,1]} := \{ & z \in \mathbb{R}^n : z = x + t\eta,\ x \in C,\ k \in \mathbb{Z}_{\geq 0},\ \omega \in \Omega,\ t \in [0,1], \\ & x + \eta \in \Big(x + h_{k+1}\widehat{F}(x, \mathbf{y}_{k+1}(\omega), h_{k+1}) \Big) \cap (C \cup D) \Big\}. \end{aligned} \tag{37}$$

Assumption 6.1. *There exists a continuous function* $V : \mathrm{dom}\ V \to \mathbb{R}$*, with* $C^{[0,1]} \cup D \subset \mathrm{dom}\ V$ *that is continuously differentiable on an open neighborhood of* $C^{[0,1]}$ *with a Lipschitz gradient on* $C^{[0,1]}$ *such that:*

1. *the set* $\{x \in C \cup D : V(x) \leq c\}$ *is bounded for each* $c > 0$*,*
2. *there exist nonnegative real numbers* Δ *and* λ *such that, for every solution* $\mathbf{x}$ *of (2) that is complete in the* t *direction,* $k \in \mathbb{Z}_{\geq 0}$*, and* $\widehat{\mathbf{f}} \in \widehat{F}(\mathbf{x}(k,\bar{\jmath}_k), \mathbf{y}_{k+1}, h_{k+1})$ *that is* $\mathcal{F}_{k+1}$*-measurable and such that* $\mathbf{x}(k,\bar{\jmath}_k) + h_{k+1}\widehat{\mathbf{f}} \in C \cup D$ *almost surely, and almost all* $\omega \in \Omega$ *(with the ensuing dependence on* ω *suppressed):*
   (a) $V(\mathbf{x}(k,\bar{\jmath}_k))) \geq \Delta$ *implies*
      i. *if* $\mathbf{x}(k,\bar{\jmath}_k) \in C$ *then:*

$$\nabla V(\mathbf{x}(k,\bar{\jmath}_k)) \cdot \mathbb{E}\left[\widehat{\mathbf{f}} \,|\, \mathcal{F}_k\right] \leq h_{k+1}\lambda V(\mathbf{x}(k,\bar{\jmath}_k)), \tag{38a}$$

$$\mathbb{E}\left[|\widehat{\mathbf{f}}|^2 \,|\, \mathcal{F}_k\right] \leq V(\mathbf{x}(k,\bar{\jmath}_k)) \tag{38b}$$

      ii. *if* $\mathbf{x}(k,\bar{\jmath}_k) \in D$ *then:*

$$V(g) \leq V(\mathbf{x}(k,\bar{\jmath}_k)) \quad \forall g \in G(\mathbf{x}(k,\bar{\jmath}_k)); \tag{39}$$

(b) $V(\mathbf{x}(k,\bar{\jmath}_k)) < \Delta$ *implies*

i. *if* $\mathbf{x}(k,\bar{\jmath}_k) \in C$ *then:*

$$\mathbb{E}\left[|\widehat{\mathbf{f}}|^2 \,|\, \mathcal{F}_k\right] \leq \lambda \tag{40}$$

ii. *if* $\mathbf{x}(k,\bar{\jmath}_k) \in D$ *then:*

$$V(g) \leq \lambda \qquad \forall g \in G(\mathbf{x}(k,\bar{\jmath}_k)). \tag{41}$$

Remark 6.2. *Observe that if* $\nabla V(x) \cdot f \leq 0$ *for all* $f \in F(x)$ *then (38a) holds when*

$$\mathbb{E}\left[\widehat{\mathbf{f}} \,|\, \mathcal{F}_k\right] \subset [\varepsilon, \Delta]\, F(\mathbf{x}(k,\bar{\jmath}_k)) + h_{k+1}\beta(\mathbf{x}(k,\bar{\jmath}_k))\mathbb{B} \tag{42}$$

(compare with (71a) in the next section) for some $\varepsilon > 0$ *and some continuous function* $\beta : \mathbb{R}^n \to \mathbb{R}_{\geq 0}$ *such that, for some* $\lambda \geq 0$*,*

$$V(x) \geq \Delta, x \in C \quad \Longrightarrow \quad |\nabla V(x)|\beta(x) \leq \lambda V(x). \tag{43}$$

We now illustrate Assumption 6.1 using several examples, including the examples considered earlier, such as the Bouncing Ball and the combinatorial optimization problem.

Example 6.3 (Resetting to a compact set). *Consider Assumption 6.1 in the special case where*

$$C \text{ is compact}, \quad G(D) \subset C. \tag{44}$$

This case occurs in resetting algorithms that ensure almost sure boundedness for stochastic approximation of differential equations or inclusions, as discussed in [11, §4.5 and 5.5], or [43, §IV] for example. Indeed, suppose Omega($i\mathbb{B}$) *for* $\dot{x} \in F(x)$ *is known to belong to some compact set* $C_1 \subset \mathbb{R}^n$ *for all large* i *and the solutions of* $\dot{x} \in F(x)$ *starting in* C_1 *do not reach the boundary of a larger compact set* $C_2 \subset \mathbb{R}^n$*. In this situation, for the data of the stochastic approximation (2) of the differential inclusion* $\dot{x} \in F(x)$*, it is reasonable to take* $C = C_2$*,* $D = \overline{\mathbb{R}^n \backslash C_2}$*,* $G(x) \subset C_1$ *for all* $x \in D$*, and* $\widehat{F}$ *to have finite conditional variance on the compact set* C*, at least if the goal is to achieve almost sure boundedness. The explanation for this minimal requirement on* $\widehat{F}$ *is given next.*

When (44) holds, to satisfy Assumption 6.1 we can take $V(x) := x^T x$ for example, which is smooth and has a Lipschitz gradient, and then take $\Delta > 0$ sufficiently large so that the set $C \cap \{x \in \operatorname{dom} V : V(x) \geq \Delta\}$ is empty. When this condition holds, in Assumption 6.1 there is nothing to check for condition 2(a)i and condition 2(a)ii holds since $G(D) \subset C$ implies that $g \in G(x)$ gives $V(g) \leq \Delta \leq V(x)$. Moreover, condition 2(b)ii holds since $G(D) \subset C$, C is compact, and V is continuous. Thus, the only condition that must be checked is condition 2(b)i, i.e., that $\widehat{F}$ has finite conditional variance on the compact set C.

In order to relate the asymptotic properties of the stochastic approximation to the asymptotic properties of the differential inclusion $\dot{x} \in F(x)$, which is discussed in Section 8, $\widehat{F}$ should satisfy a condition like (42). However, as emphasized above, this condition is not needed for almost sure boundedness.

Example 6.4 (Bouncing Ball, stochastically approximated). *Consider the stochastic approximation of the Bouncing Ball hybrid system from Examples 2.4 and 4.2, with $\widehat{F}$ as in* (16) *and where the noise $\{\mathbf{y}_k\}_{k=1}^{\infty}$ is iid with an exponential distribution having variance σ. It is a straightforward exercise to verify that the conditions of Assumption 6.1 hold with $\Delta = 0$, $\lambda = 0$, and*

$$V(x) := x_2^2 + 2\gamma x_1 + \sigma + \gamma^2 \tag{45}$$

when $D = \{0\} \times \mathbb{R}_{\geq 0}$. However, with D replaced by

$$\widehat{D} := \mathbb{R}_{\leq 0} \times \mathbb{R}_{\leq 0}, \tag{46}$$

as considered in Example 4.2, V is modified to be

$$\widetilde{V}(x) := x_2^2 + 2\gamma\rho(x_1) + \sigma + \gamma^2 \tag{47}$$

where $\rho : \mathbb{R} \to [-\varepsilon, \infty)$ is a $\mathcal{C}^2$, nondecreasing function with $\rho(s) = s$ for $s \geq 0$. Note that $V(x) = \widetilde{V}(x)$ for all $x \in C \cup D$. Next, we pick $\Delta > 0$ sufficiently large so that $\widetilde{V}(x) \geq \Delta$ and $x \in \widehat{D} \backslash D$ implies that

$$(1 - \delta^2) x_2^2 \geq 2\gamma\varepsilon. \tag{48}$$

Then, for $x \in \widehat{D} \backslash D$, and $\widetilde{V}(x) \geq \Delta$,

$$\begin{aligned}\widetilde{V}(g(x)) &= \delta^2 x_2^2 + \sigma + \gamma^2 \\ &\leq (\delta^2 - 1)x_2^2 + 2\gamma\varepsilon + x_2^2 - 2\gamma\varepsilon + \sigma + \gamma^2 \\ &\leq (\delta^2 - 1)x_2^2 + 2\gamma\varepsilon + \widetilde{V}(x) \\ &\leq \widetilde{V}(x).\end{aligned}$$

It follows that Assumption 6.1 also holds with $\widehat{D}$ in place of D by using $\widetilde{V}$ and the chosen Δ.

Example 6.5 (Combinatorial optim. with stochastic gradients). *Consider the optimization problem from Example 2.5 and its stochastic approximation from Example 4.3 where, for each $q \in \mathcal{Q}$, $f_q : \mathbb{R}^m \to \mathbb{R}$ is given by* (19) *and is radially unbounded with a globally Lipschitz continuous gradient. We also assume that, in (19), for each $q \in \mathcal{Q}$, and $i \in \{1, \ldots, N_q\}$, $f_{i,q}$ is continuously differentiable and that there exist $\kappa > 0$ and $\theta > 0$ such that*

$$|\nabla f_{i,q}(z)|^2 \leq \kappa f_q(z) + \theta \qquad \forall (z, q, i) \in \mathbb{R}^m \times \mathcal{Q} \in \{1, \ldots, N_q\}\,. \tag{49}$$

Observe, using the distribution for $\mathbf{y}_{k+1}$ and (49), that for all $x \in C$ and all $k \in \mathbb{Z}_{\geq 0}$ we have

$$\begin{aligned}\mathbb{E}\left[\left|\nabla f_{(\mathbf{y}_{k+1} \bmod N_q),q}(z)\right|^2 \mid \mathcal{F}_k\right] &= \frac{1}{N_q}\sum_{i=1}^{N_q} |\nabla f_{i,q}(z)|^2 \\ &\leq \frac{1}{N_q}\sum_{i=1}^{N_q} (\kappa f_q(z) + \theta) = \kappa f_q(z) + \theta.\end{aligned} \tag{50}$$

Now take

$$V(x) := \kappa f_q(z) + \theta \qquad \forall (z, q) \in \mathbb{R}^m \times \mathcal{Q} \tag{51}$$

and, for the sake of Assumption 6.1, extend the definition of V to $\mathbb{R}^m \times \mathcal{U}$ where $\mathcal{U}$ is an open neighborhood of $\mathcal{Q}$ small enough so that the projection $p(\cdot)$ from each point in $\mathcal{U}$ to $\mathcal{Q}$ is unique, so that the extended function is given by $V(z, p(q), \tau)$. Due to the assumptions on $z \mapsto f_q(z)$ for each $q \in \mathcal{Q}$, the fact that $\mathcal{Q}$ is finite, and the structure of $\widehat{F}$, it follows that V is continuous *on its domain, $C^{[0,1]} \cup D \subset \text{dom } V$, and V is continuously differentiable on an open neighborhood of $C^{[0,1]}$ with a globally Lipschitz gradient on $C^{[0,1]}$. In addition, the first item of Assumption 6.1 holds.*

Take $\Delta = \lambda = 0$. It follows from (50) and the definition of V that (38b) holds. Also, using Remark 6.2 and the fact that

$$\mathbb{E}\left[-\nabla f_{(\mathbf{y}_{k+1} \text{mod} N_q)}, q(z) \,|\, \mathcal{F}_k\right] = -\nabla f_q(z), \tag{52}$$

which follows from the distribution of $\mathbf{y}_{k+1}$, we have that (38a) holds. Finally, from the definition of G, it follows that (39) holds.

Example 6.6 (CO on a torus with stochastic finite differences). *Consider the stochastic approximation model in Example 4.4 for the combinatorial optimization on a torus model in Example 2.6.*

Note that, from the mean-value theorem, there exists $\vartheta > 0$ such that

$$\begin{aligned} &f_q\left(\exp\left(R(y^+h^+)\,z\right) - f_q\left(\exp\left(-R(y^+h^+)\right)z\right) \in \\ &\nabla f_q(z + h^+\vartheta\mathbb{B}) \cdot \left(\exp\left(R\left(y^+h^+\right)z\right) - \exp\left(-R\left(y^+h^+\right)z\right)\right) \end{aligned} \tag{53}$$

and

$$\exp\left(R\left(y^+h^+\right)z\right) - \exp\left(-R\left(y^+h^+\right)z\right) = 2R\left(y^+h^+\right)z + \mathcal{O}((h^+)^2) \tag{54}$$

so that

$$\begin{aligned} f_q\left(\exp\left(R(y^+h^+)\,z\right) - f_q\left(\exp\left(-R(y^+h^+)\right)z\right) = \\ \nabla f_q(z) \cdot 2R(y^+h^+)z + \mathcal{O}((h^+)^2). \end{aligned} \tag{55}$$

It then follows that there exists $\beta > 0$ such that, for every solution $\mathbf{x}$ of (2) that is complete in the t direction, $k \in \mathbb{Z}_{\geq 0}$, and $\widehat{\mathbf{f}} \in \widehat{F}(\mathbf{x}(k, \overline{\jmath}_k), \mathbf{y}_{k+1}, h_{k+1})$ that is $\mathcal{F}_{k+1}$-measurable, and almost all $\omega \in \Omega$ (with the ensuing dependence on ω suppressed):

$$\mathbb{E}\left[\widehat{\mathbf{f}} \,|\, \mathcal{F}_k\right] \subset F(\mathbf{x}(k, \overline{\jmath}_k)) + h_{k+1}\beta\mathbb{B}, \tag{56a}$$

$$\mathbb{E}\left[|\widehat{\mathbf{f}}|^2 \,|\, \mathcal{F}_k\right] \leq 2\mathbf{u}^T(k, \overline{\jmath}_k)\mathbf{u}(k, \overline{\jmath}_k) + \beta. \tag{56b}$$

At this point, using the ideas in Remark 6.2, it is not difficult to establish that Assumption 6.1 holds with $\Delta = \lambda = 0$ by taking V of the form

$$V(x) := \kappa \left(\frac{1}{2} u^T u + f_q(z) \right) + \vartheta \tag{57}$$

with $\kappa > 0$ and $\vartheta > 0$ sufficiently large.

In Assumption 6.1, which the previous examples have been shown to satisfy, the bound in (38a) is a relaxation of the bound in [20, (4a)] to account for the fact that $\widehat{F}$ is allowed to depend on h here. Nevertheless, the observation in [20, Lemma 2] still applies. Namely,

Lemma 6.7. *If Assumption 6.1 holds with $\Delta > 0$ then it holds with $\widehat{\Delta} = 0$ in place of Δ and $\widehat{V} := \rho(V)$ in place of V, where $\rho : \mathbb{R} \to \mathbb{R}_{\geq 0}$ is any $\mathcal{C}^2$ function that is constant on $(-\infty, \Delta]$, satisfies $\rho'(s) \in [0, 1]$ and $\rho(s) \geq \max\{\lambda, s\}$ for all $s \geq 0$, and satisfies $\rho(s) = s$ for all large s.*

Proof. The function $\widehat{V}$ is $\mathcal{C}^1$ because V is $\mathcal{C}^1$ and ρ is $\mathcal{C}^2$. Because ρ'' is nonzero only on a bounded set, it is bounded. This, and a version of the Mean Value Theorem, implies that $\nabla \widehat{V}$ is Lipschitz continuous. Because $\widehat{V}(x) \geq \max\{\lambda, V(x)\}$ for all $x \in \text{dom } V = \text{dom } \widehat{V}$, item 1 of Assumption 6.1 for V implies item 1 of Assumption 6.1 for $\widehat{V}$.

Since ρ is $\mathcal{C}^2$, $\rho'(s) \in [0, 1]$ for all $s \geq 0$, ρ constant on $[0, \Delta]$ and (38a) holds for $V(x) \geq \Delta$ and $x \in C$, it follows that (38a) holds for all $x \in C$ with $\widehat{V}$ in place of V on both the left-hand side and right-hand side of (38a).

Finally, because $\widehat{V}(x) \geq \max\{\lambda, V(x)\}$ for all $x \in \text{dom } V$, the combination of (38b) and (40) together with the combination of (39) and (41) give that (38b) holds for all $x \in C$ and (39) holds for all $x \in D$, with $\widehat{V}$ in place of V. $\square$

The last assumption imposed before stating the result on almost sure boundedness is the following:

Assumption 6.8. *The step sizes $\{h_k\}_{k=1}^{\infty}$ are square summable: $\sum_{k=1}^{\infty} h_k^2 < \infty$.*

With Assumptions 6.1 and 6.8 in hand, we have our main theorem of this section, which extends [20, Theorem 1] from the setting of stochastic difference inclusions to the setting of stochastic hybrid simulators. It applies to Example 6.3 (Projections to compact sets), Example 6.4 (Bouncing Ball, stochastically approximated), Example 6.5 (Combinatorial optimization with stochastic gradients), and Example 6.6 (Combinatorial optimization on a torus with stochastic finite differences), since Assumption 6.1 was shown to hold for each of these examples. The proof is analogous to the proof of [20, Theorem 1].

Theorem 6.9. *If Assumptions 6.1 and 6.8 hold, then each solution to* (2) *that is complete in the t direction is bounded almost surely.*

Proof. According to Lemma 6.7, we can assume that Assumption 6.1 holds with $\Delta = 0$. Letting L be a Lipschitz constant for ∇V on $C^{[0,1]}$, which was defined in (37), for all (x, η) such that $x + t\eta \in C^{[0,1]}$ for all $t \in [0, 1]$, we have that

$$\begin{aligned} V(x+\eta) &= V(x) + \int_0^1 \nabla V(x+t\eta)\cdot \eta \, dt \\ &\leq V(x) + \nabla V(x)\cdot\eta + \frac{1}{2}L|\eta|^2. \end{aligned} \tag{58}$$

Recall the notation $\bar{\jmath}_k$, $\underline{\jmath}_k$ introduced by (25). Let $x := \mathbf{x}(k, \bar{\jmath}_k)$ and let η be $\mathcal{F}_{k+1}$-measurable and such that

$$x + \eta = \mathbf{x}(k+1, \underline{\jmath}_{k+1}) \in \Big(x + h_{k+1}\widehat{F}(\mathbf{x}(k, \bar{\jmath}_k), \mathbf{y}_{k+1}, h_{k+1})\Big) \cap (C \cup D). \tag{59}$$

As mentioned above the statement of Assumption 6.1, this implies that $x + t\eta \in C^{[0,1]}$ for all $t \in [0, 1]$. Taking conditional expectations and using Assumption 6.1, we get

$$\begin{aligned} &\mathbb{E}\Big[V\Big(\mathbf{x}(k+1, \underline{\jmath}_{k+1})\Big) \mid \mathcal{F}_k\Big] \\ &\qquad \leq V(\mathbf{x}(k, \bar{\jmath}_k)) + h_{k+1}^2 \lambda V(\mathbf{x}(k, \bar{\jmath}_k)) + h_{k+1}^2 \frac{L}{2} V(\mathbf{x}(k, \bar{\jmath}_k)). \end{aligned} \tag{60}$$

Moreover, it follows from (39) in Assumption 6.1 that

$$V(\mathbf{x}(k, j+1)) \leq V(\mathbf{x}(k, j)) \qquad \forall j \in \Big\{\underline{\jmath}_k, \ldots, \bar{\jmath}_k - 1\Big\} \tag{61}$$

Then, we apply Theorem 5.3 with, for each $k \in \mathbb{Z}_{\geq 0}$ and $j \in \left\{\underline{\jmath}_k, \ldots, \bar{\jmath}_k\right\}$, the definitions

$$\mathbf{z}(k,j) := V(\mathbf{x}(k,j)), \tag{62a}$$

$$\mathbf{w}(k,j) := h_{k+1}^2 \left(\lambda + \frac{L}{2}\right), \tag{62b}$$

$$\mathbf{v}(k,j) = \mathbf{u}_c(k,j) = \mathbf{u}_d(k,j) := 0, \tag{62c}$$

and use Assumption 6.8 to get that, for almost all $\omega \in \Omega$,

$$\lim_{k+j\to\infty,(k,j)\in\mathrm{dom}\ x} V(\mathbf{x}(k,j))$$

exists and is finite. Almost sure boundedness of $\mathbf{x}$ now follows from item 1 of Assumption 6.1. $\square$

7. Morse decompositions and Lyapunov functions

This section, after some background material, presents the Morse and Conley decompositions of an attractor for (1), and gives a Lyapunov-like characterization for the Morse decomposition. The main definitions, given there for single-valued flows, go back to the seminal work by Conley [44]. For an overview, see the survey [45]. They were generalized to a variety of settings without uniqueness of solutions, from multi-valued discrete-time dynamics [46], differential inclusions [47], [22], and, more recently, hybrid inclusions like (1) in [23]. The main point behind the decompositions is that they identify in the finest detail where the transient behavior and where the asymptotic behavior occur.

Throughout this section, suppose that:

(i) The hybrid system (1) satisfies Assumption 2.1 and Assumption 2.2.

(ii) $\mathcal{A} \subset \mathbb{R}^n$ is a nonempty, compact global attractor for (1), in the sense that for all large enough $i \in \mathbb{N}$, $\mathcal{A} = \mathrm{Omega}(i\mathbb{B})$.

Here, and below, $\mathrm{Omega}(K)$ is the Omega-limit from the set $K \subset \mathbb{R}^n$ as defined earlier in (5). By [14, Proposition 6.26, Proposition 7.5], $\mathcal{A}$ is *asymptotically stable* for (1), that is: a) it is *Lyapunov stable*: for every $\varepsilon > 0$ there exists $\delta > 0$ such that every $\phi \in \mathcal{S}^{\mathcal{H}}(\mathcal{A} + \delta\mathbb{B})$ satisfies $\phi(t,j) \in \mathcal{A} + \varepsilon\mathbb{B}$ for all $(t, j) \in \text{dom}\, \phi$; and b) it is *attractive*: there exists $\delta > 0$ such that every $\phi \in \mathcal{S}^{\mathcal{H}}(\mathcal{A}+\delta\mathbb{B})$ converges to $\mathcal{A}$. The *basin of attraction* of the asymptotically stable $\mathcal{A}$, denoted $\mathcal{B}(\mathcal{A})$, is the set of all $x \in \mathbb{R}^n$ for which every $\phi \in \mathcal{S}^{\mathcal{H}}(x)$ converges to $\mathcal{A}$. By definition, $\mathcal{B}(\mathcal{A})$ contains $\mathbb{R}^n \setminus (C \cup D)$ and equals $\mathbb{R}^n$ — so that $\mathcal{A}$ is *globally asymptotically stable* — by b) above. The set $\mathcal{A}$, or any other asymptotically stable compact set $A \subset \mathbb{R}^n$, admits a characterization through a smooth Lyapunov function as indicated in the following, which is (a part of) [36, Theorem 3.13].

Lemma 7.1. *Let A be a nonempty compact asymptotically stable set. Then its basin of attraction $\mathcal{B}$ is open, contains A, and there exists a C^∞ function $V : \mathcal{B} \to \mathbb{R}_{\geq 0}$ such that $V(x) = 0$ if and only if $x \in A$ and $V(x_i) \to \infty$ if $|x_i| \to \infty$ or x_i converge to a boundary point of $\mathcal{B}$, and*

$$\begin{aligned} \nabla V(x) \cdot f &\leq -V(x) \quad \forall x \in C \cap \mathcal{B},\ f \in F(x), \\ V(g) &\leq \tfrac{1}{e} V(x) \quad \forall x \in D \cap \mathcal{B},\ g \in G(x). \end{aligned} \tag{63}$$

Above, the second inequality can be replaced by $V(g) \leq \gamma V(x)$ for any $\gamma \in (0, 1)$, but $\gamma = 1/e$, with e being the Euler's constant, is used to match the exponential decay rate of V during flow.

Besides asymptotic stability, and the implied strong forward invariance, $\mathcal{A}$ enjoys the property of weak backward invariance, also by [48, Proposition 6.26]. This, and Assumption 2.2, ensure that $\mathcal{A} \subset C \cup D$, and thus from every $x \in \mathcal{A}$ there exist solutions to (1) and the maximal ones are complete. Additionally, weak backward invariance ensures that $\text{Omega}(\mathcal{A}) = \mathcal{A}$.

Let $\mathcal{H}|_{\mathcal{A}} := (C \cap \mathcal{A}, F, D \cap \mathcal{A}, G')$, where for every $x \in \mathbb{R}^n$, $G'(x) := G(x) \cap \mathcal{A}$. Because $\mathcal{A}$ is strongly forward invariant, $G(x) \subset \mathcal{A}$ for every $x \in D \cap \mathcal{A}$, so that $G'(x) \neq \emptyset$ if $x \in D \cap \mathcal{A}$. It is then straightforward that $\mathcal{H}|_{\mathcal{A}}$ satisfies Assumption 2.1. The mentioned invariance properties of $\mathcal{A}$ ensure that $\mathcal{H}|_{\mathcal{A}}$ satisfies Assumption 2.2.

A compact set $A \subset \mathcal{A}$ is an *attractor in $\mathcal{A}$* if there exists a neighborhood U of A such that $\text{Omega}(U) = A$ for the dynamics given by $\mathcal{H}|_{\mathcal{A}}$. The *repeller* associated to an attractor A is the set

$$A^* := \left\{ x \in \mathcal{A} \, : \, \exists \phi \in \mathcal{S}^{\mathcal{H}|_{\mathcal{A}}}(x) \text{ s.t. } \text{omega}(\phi) \not\subset A \, \right\},$$

where $\mathcal{S}^{\mathcal{H}|_{\mathcal{A}}}(x)$ is the set of maximal, hence complete, solutions to $\mathcal{H}|_{\mathcal{A}}$ from x, and $\text{omega}(\phi)$ is the *omega-limit* of ϕ, as defined earlier in (4). Because

$\mathcal{A}$ is bounded and strongly forward invariant, and because each bounded solution ϕ converges to omega(ϕ), which is commonly used in invariance-based arguments, omega(ϕ) $\not\subset A$ above is equivalent to saying that ϕ does not converge to A. Like in the original idea of [44, Chapter II, Section 6.4, A], with details for the case of hybrid inclusions in [23, Lemma 9], there exists at most countably many attractors in $\mathcal{A}$.

By [14, Proposition 6.26, Proposition 7.5], a nonempty attractor A in $\mathcal{A}$ is asymptotically stable for $\mathcal{H}|_{\mathcal{A}}$. Furthermore, similarly to the original idea in [44, Chapter II, Section 5.3, D], by [19, Proposition 2.4] and because $\mathcal{A}$ is an attractor, A is an attractor for $\mathcal{H}$ itself. For the dynamics given by $\mathcal{H}|_{\mathcal{A}}$, $A^* = \mathbb{R}^n \setminus \mathcal{B}(A)$, in part because $\mathbb{R}^n \setminus \mathcal{A} \subset \mathcal{B}(A)$. For the dynamics $\mathcal{H}$, $A^* = \mathcal{A} \setminus \mathcal{B}(A)$. The set $\mathcal{A} \cap \mathcal{B}(A)$ is the same, whether the basin is for $\mathcal{H}$ or $\mathcal{H}|_{\mathcal{A}}$.

A *Morse decomposition* of $\mathcal{A}$ is an ordered collection

$$\mathcal{D} = \{M_1, M_2, \ldots, M_L\}$$

of compact subsets of $\mathcal{A}$ such that there exists an increasing sequence of attractors

$$\emptyset = A_0 \subset A_1 \subset \cdots \subset A_L = \mathcal{A} \tag{64}$$

in $\mathcal{A}$, such that, for each $l = 1, 2, \ldots, L$,

$$M_l = A_l \cap A^*_{l-1}. \tag{65}$$

Note that, by forward completeness and strong forward invariance of $\mathcal{A}$ for $\mathcal{H}$, $A^*_0 = \mathcal{A}$ and $A^*_L = \emptyset$. Since the number of attractors in $\mathcal{A}$ is either finite or infinite and countable, the number of different Morse decompositions of $\mathcal{A}$ is either finite or infinite and countable.

Each set in a Morse decomposition $\mathcal{D}$ is a *Morse set*. For any Morse decomposition $\mathcal{D}$ of $\mathcal{A}$, let

$$\mathcal{M}(\mathcal{D}) := M_1 \cup M_2 \cup \cdots \cup M_L.$$

It is elementary, as suggested in the original [44], that

$$\mathcal{M}(\mathcal{D}) = \bigcap_{l=0}^{L} A_l \cup A^*_l. \tag{66}$$

Indeed, the verification involves only (64), (65), elementary set operations, and no dynamics.

Example 7.2 (Academic example and Morse decomposition). *Recall the hybrid system on* $\mathbb{R}$ *from Example 2.3, illustrated in Figure 1. The global attractor* $\mathcal{A}$ *is* $[0,5]$*. One Morse decomposition and the resulting Morse sets are:*

$$\begin{array}{lll} A_0 = \emptyset & A_0^* = \mathcal{A} & \\ A_1 = \{0\} & A_1^* = [1,4] & M_1 = A_1 \cap A_0^* = \{0\} \\ A_2 = \{0\} \cup [2,3] & A_2^* = \{1,4\} & M_2 = A_2 \cap A_1^* = [2,3] \\ A_3 = \{0\} \cup [2,5] & A_3^* = \{1\} & M_3 = A_3 \cap A_2^* = \{4\} \\ A_4 = \mathcal{A} & A_4^* = \emptyset & M_4 = A_4 \cap A_3^* = \{1\} \end{array}$$

One can verify that (66) *holds. Not all attractors in* $\mathcal{A}$ *appear above, for example* $[2,3]$ *is an attractor.*

Just as an asymptotically stable set admits a (smooth) Lyapunov function, a Morse decomposition admits a (smooth) "Morse-Lyapunov function," as described in the main result of this section:

Theorem 7.3. *Let* $\mathcal{M}$ *be a Morse decomposition of* $\mathcal{A}$*. Then, there exists a smooth and radially unbounded function* $V : \mathbb{R}^n \to \mathbb{R}_{\geq 0}$ *that is constant on each Morse set and*

$$V(M_1) < V(M_2) < \cdots < V(M_L),$$

and a continuous function $w : \mathbb{R}^n \to \mathbb{R}_{\geq 0}$ *that is positive definite with respect to* $\mathcal{M}(\mathcal{D})$*, such that*

$$\nabla V(x) \cdot f \leq -w(x) \quad \forall x \in C, f \in F(x),$$

$$V(g) \leq V(x) - w(x) \quad \forall x \in D, g \in G(x).$$

The function V from Theorem 7.3 is used in the next section to establish the convergence properties not of solutions to (1), but of the solutions to the stochastic approximation (2). The proof of Theorem 7.3 is postponed to Section 10. A result similar to Theorem 7.3 appeared in the context of "multi-stability" — a concept very similar to the Morse decomposition, but not related to it by the authors — in a hybrid system in [49]. The proof in

[49] relies on earlier work, including results for continuous-time systems, and is not easy to track.

The set

$$\mathcal{R}(\mathcal{A}) := \bigcap \{\mathcal{M}(\mathcal{D}) \,:\, \mathcal{D} \text{ is a Morse decomposition of } \mathcal{A}\} \tag{67}$$

is the so-called *chain-recurrent part of* $\mathcal{A}$, consisting of points that are *chain recurrent*. The decomposition of $\mathcal{A}$ into $\mathcal{R}(\mathcal{A})$ and its complement, on which the dynamics is "gradient-like", is known as Conley's decomposition and has been referred to as "the fundamental theorem of dynamical systems"; see the expository [50]. "Gradient-like" means that there exists a continuous function that decreases along every solution in $\mathcal{A} \setminus \mathcal{R}(\mathcal{A})$. The main message behind the decomposition is that transient behaviors occur where the dynamics is gradient-like, i.e., on $\mathcal{A} \setminus \mathcal{R}(\mathcal{A})$, while asymptotic behaviors are represented by different components of $\mathcal{R}(\mathcal{A})$, which can be equilibria, periodic orbits, continua of equilibria, and more. The decomposition is often established by showing the existence of a Lyapunov-like function, usually called a total or a complete Lyapunov function. For (1), the appropriate result is [23, Theorem 2], which shows the existence of a continuous function $V : \mathcal{A} \to [0,1]$ such that $V(\phi(t,j)) < V(\phi(0,0))$ for every $\phi \in \mathcal{S}^{\mathcal{H}}(\mathcal{A} \setminus \mathcal{R}(\mathcal{A}))$ and every $(t,j) \in \operatorname{dom} \phi$ with $t + j > 0$, that is constant on each (appropriately defined) component of $\mathcal{R}(\mathcal{A})$, and has some further properties. See also the main result of [19]. That (67) is a valid description of $\mathcal{R}(\mathcal{A})$ follows from (66) and from [23, Proposition 11(a)], which says that $\mathcal{R}(\mathcal{A})$ is the intersection of $A \cup A^*$ over all attractors A in $\mathcal{A}$.

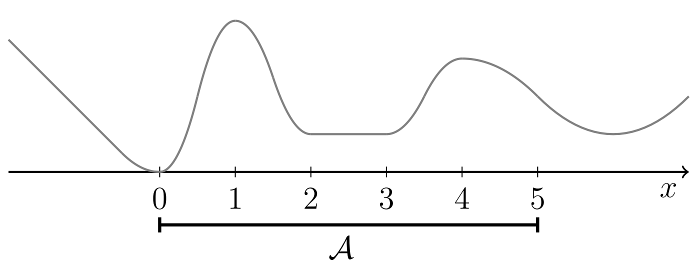

Figure 2: A Morse-Lyapunov function for (6), as described in Theorem 7.3, which, in this case, is also a total Lyapunov function.

For the example hybrid system (6), the chain-recurrent part turns out to be the union of the exhibited Morse sets:

$$\mathcal{R}(\mathcal{A}) = \{0\} \cup [2,3] \cup \{4\} \cup \{1\}.$$

Each of the three singletons, $\{0\}$, $\{4\}$, and $\{1\}$ is a chain-recurrent point because it is an equilibrium, while $[2,3]$ is a component of the chain-recurrent set $\mathcal{R}(\mathcal{A})$ because it is the image of a periodic (in the hybrid sense) solution. For this example, there is a finite number of components and, consequently, a smooth total Lyapunov function can be fairly easily built directly, see Figure 2. In general, there is at most countably many attractors possibly leading to infinitely many components of $\mathcal{R}(\mathcal{A})$. An example of this is provided by the differential equation

$$\dot{x} = f(x) := -x^2 \sin(1/x)$$

on $\mathbb{R}$, with the understanding that $f(0) = 0$, for which $[-1/\pi, 1/\pi]$ is the global attractor and in it, there are infinitely many equilibria, and — among them — infinitely many attractors in $\mathcal{A}$. For an academic example of this in a hybrid setting, see [23, equations (1), (2), and (3)].

For an example of the chain-recurrent set for a differential inclusion, consider the (generalized) gradient flow

$$\dot{x} \in -\partial f(x), \tag{68}$$

where $\partial f : \mathbb{R}^n \rightrightarrows \mathbb{R}^n$ is Clarke subdifferential / generalized gradient of a locally Lipschitz continuous function $f : \mathbb{R}^n \to \mathbb{R}$ that is nonpathological. This set-valued mapping has nonempty, compact, and convex values everywhere and is outer semicontinuous and locally bounded. A convenient formula (which is not its original definition) for the generalized gradient is

$$\partial f(x) = \operatorname{con}\left\{ \lim_{i\to\infty} \nabla f(x_i) \,|\, x_i \in O \setminus S,\ x_i \to x \right\},$$

where $O \subset \mathbb{R}^n$ is the set of all points at which f is differentiable (the complement of O has measure 0) and $S \subset \mathbb{R}^n$ is any set of measure 0, so that $O \setminus S$ is dense in $\mathbb{R}^n$. See [51, Theorem 2.5.1] or [24, Theorem 9.61]. The concept of nonpathological functions goes back to [52] and was popularized for control purposes, including Lyapunov analysis, by [53]; see [54] for details and further references. Nonpathological functions satisfy an appropriate chain rule, when composed with an absolutely continuous solution to a differential equation or inclusion; see [54, Proposition 1]. This, and the recent [55, Lemma 5.2] imply that any solution $\phi : [0,T] \to \mathbb{R}^n$ to (68) satisfies

$$f(\phi(T)) - f(\phi(0)) = -\int_0^T \|\dot{\phi}(s)\|^2 \, ds = -\int_0^T \operatorname{dist}^2(0, \partial f(\phi(s))) \, ds. \tag{69}$$

The example below generalizes a result of [56] that was given for a differentiable f and the resulting gradient flow. The argument uses similar ideas, but is in fact shorter here, because it is direct. Below, x is a *critical point* of f if $0 \in \partial f(x)$ and $r \in \mathbb{R}$ is a *regular value* of f if there is no critical point x such that $r = f(x)$. The final ingredient needed for the example is the original definition of the chain-recurrent set for a differential equation, or inclusion, here stated for (68). Following [44], $\mathcal{R}(\mathcal{A})$ is defined as the set of all points x such that, for every $\tau, \varepsilon > 0$, there exists a (τ, ε)-*chain* from x to x, i.e., there exist a finite sequence of points $x = x_0, x_1, \ldots, x_k = x$ in $\mathcal{A}$ and solutions $\phi_0, \phi_1, \ldots, \phi_{k-1}$ to (68) such that $\phi_i(0) = x_i$ and $x_{i+1} \in \phi_i(T_i) + \varepsilon\mathbb{B}$ for some $T_i \geq \tau$ and for $i = 0, 1, \ldots, k-1$. An alternative characterization of the chain recurrent set was already given in (67) in the setting of hybrid inclusions. Of course, this characterization applies here too, but is not used.

Example 7.4 ((Generalized) gradient flow and chain recurrence). *Suppose that a locally Lipschitz $f : \mathbb{R}^n \to \mathbb{R}$ is nonpathological and the set of regular values of f is dense in $\mathbb{R}$. Suppose that a nonempty and compact set $\mathcal{A} \subset \mathbb{R}^n$ is an attractor for (68). Then:*

$$x \in \mathcal{R}(\mathcal{A}) \text{ if and only if } x \text{ is a critical point of } f. \text{ [1]}$$

Clearly, if $0 \in \partial f(x)$ then a constant solution from x to x, on an arbitrarily long time interval, yields the needed (τ, ε)-chain, and $x \in \mathcal{R}(\mathcal{A})$. Pick $x \in \mathcal{A}$ such that $0 \notin \partial f(x)$. Let $S_1(x) \subset \mathcal{A}$ be the reachable set from x in time 1. Then $\max f(S_1) < f(x)$. Let r be a regular value of f such that $\max f(S_1) < r < f(x)$. Let $K := \{a \in \mathcal{A} : f(a) \leq r\}$ and $S_{[1,2]}(K)$ be the reachable set from K in time between 1 and 2. It is claimed that there exists $\varepsilon > 0$ such that $S_{[1,2]}(K) + \varepsilon\mathbb{B} \subset K$. If the claim is true, then there is no $(2, \varepsilon)$-chain from x to x, because every solution on $[0, T]$ with $T \geq 2$ can be concatenated from a solution on $[0, 1]$ and a solution, or solutions, on $[1, 2]$. Then $x \notin \mathcal{R}(\mathcal{A})$. To see the claim, suppose, to the contrary, that there exist $x_i \in K$ and solutions $\phi_i : [0, T_i] \to \mathbb{R}^n$, $T_i \in [1, 2]$, such that $f(y_i) \to r$, where $y_i = \phi_i(T_i)$. Without loss of generality, T_i converge to some $T \in [1, 2]$, y_i converge to some $y \in K$ with $f(y) = r$, and ϕ_i converge graphically to a solution $\phi : [0, T] \to \mathbb{R}$.

[1] In [20], we informally predicted this conclusion but didn't include the assumption on regular values.

Because $f(y_i) < f(x_i) \leq r$, $f(x_i) \to r$. Then, using the Cauchy-Schwartz inequality,

$$0 \quad \leq \quad \int_0^T \|\dot{\phi}_i(s)\| \leq \sqrt{T}\sqrt{\int_0^T \|\dot{\phi}_i(s)\|^2\, ds} = \sqrt{T}\sqrt{f(x_i) - f(y_i)} \to 0$$

so that the arc-lengths of ϕ_i converge to 0, so that $x_i \to y$. Then $\phi(T) = \phi(0) = y$, which is impossible in light of (69) *because $0 \notin \partial f(y)$. Indeed, $0 \in \partial f(y)$ and outer semicontinuity of ∂f ensure that there exists $\delta > 0$ such that* $\operatorname{dist}(0, \partial f(\phi(s))) \geq \delta$ *for all small enough $s > 0$, so that $f(\phi(T)) < f(\phi(0))$. This ends the argument.*

If $f : \mathbb{R}^n \to \mathbb{R}$ is $\mathcal{C}^n$ then, by the Sard's Theorem, the set of regular values of f is dense in $\mathbb{R}$. This conclusion may fail if f is "only" $\mathcal{C}^{n-1}$, and then, there may exist chain-recurrent points y for $\dot{x} = -\nabla f(x)$ for which $\nabla f(y) \neq 0$. See [56, Section 4], and note that every differentiable f is nonpathological, so being nonpathological does not imply the extra assumption in Proposition 7.4. For examples of quite strange behavior of solutions to (68), see [57]. Among other things, (68) may have a periodic solution whose image does not contain any critical points.

Arguments similar to those in Example 7.4, and a somewhat tedious case by case analysis of different hybrid behaviors, let one justify the conclusions in the following example:

Example 7.5 (Combinatorial optimization and chain recurrence). *Consider the combinatorial optimization model of Example 2.5 in the special case where $N_0 = 1$ and $\varepsilon = \delta > 0$. Suppose that $\mathcal{A} \subset \mathbb{R}^{m+2}$ is a global attractor. Let $Z \subset \mathbb{R}^{m+2}$ be the set given by*

$$\left\{ x = (z, q, \tau) \in \mathcal{A} \times \mathcal{Q} \times [0,1] \;\middle|\; \begin{array}{l} \exists\, p \in Q \text{ s.t. } \nabla f_p(z) = 0 \text{ and} \\ p, q \in \arg\min_{s \in \mathcal{Q}} s \mapsto f_s(z) \end{array} \right\}. \tag{70}$$

Then $Z \subset \mathcal{R}(\mathcal{A})$. If, for each $q \in \mathcal{Q}$, ∇f_q is Lipschitz continuous, which was an additional assumption made in Example 6.5, and for each $q \in \mathcal{Q}$, the set of critical values of f_q is nowhere dense, then $Z = \mathcal{R}(\mathcal{A})$. The relevance of Lipschitz continuity is that it implies that solutions to $\dot{z} = -\nabla f_q(z)$ are unique. Both assumptions, on Lipschitz continuity and on critical values, hold if each f_q is $\mathcal{C}^n$.

8. Asymptotic behavior of the solutions to (2)

In this section, we *assume* almost sure boundedness of solutions to (2) and weaken the condition on the variance of $\widehat{F}$ to arrive at conditions for and characterizations of convergence of the solutions of the stochastic approximation (2). The characterization of convergence is in terms of the asymptotic behavior of the solutions of the hybrid system (1). The following assumption, which weakens the variance condition in Assumption 6.1, is analogous to [20, Assumption 4]:

Assumption 8.1. *The continuous, nondecreasing function $\gamma : \mathbb{R}_{\geq 0} \to \mathbb{R}_{\geq 0}$, the set-valued mapping $F : \mathbb{R}^n \to \mathbb{R}^n$ in (1a), $\varepsilon \in (0, 1]$, and $\Delta \in [1, \infty)$ are such that, for each solution $\mathbf{x}$ of (2) that is complete in the t direction and each $k \in \mathbb{Z}_{\geq 0}$, if $\widehat{\mathbf{f}} \in \widehat{F}(\mathbf{x}(k, \bar{\jmath}_k), \mathbf{y}_{k+1}, h_{k+1})$ is $\mathcal{F}_{k+1}$-measurable and such that $\mathbf{x}(k, \bar{\jmath}_k) + h_{k+1}\widehat{\mathbf{f}} \in C \cup D$ almost surely then, almost surely*

$$\mathbb{E}\left[\widehat{\mathbf{f}} \,|\, \mathcal{F}_k\right] \subset [\varepsilon, \Delta]\, F(\mathbf{x}(k, \bar{\jmath}_k)) + h_{k+1}\gamma(|\mathbf{x}(k, \bar{\jmath}_k)|)\mathbb{B} \tag{71a}$$

$$\mathbb{E}\left[|\widehat{\mathbf{f}}|^2 \,|\, \mathcal{F}_k\right] \leq \gamma(|\mathbf{x}(k, \bar{\jmath}_k)|). \tag{71b}$$

Remark 8.2. *Note that when $\widehat{F}$, the noise sequences $\{\mathbf{y}_k\}_{k=1}^{\infty}$, and the step-size sequence $\{h_k\}_{k=1}^{\infty}$ satisfy the variance conditions given in Assumption 6.1 then they also satisfy the variance condition given in (71b) of Assumption 8.1. Hence, the data specified in Examples 6.3-6.6 discussed in Section 6 satisfy (71b). The containment (71a) supplies the condition required for the stochastic approximation in Example 6.3 to inherit the asymptotic convergence properties of the solutions to the differential inclusion. Moreover, it is easy to see that the data in Examples 4.2-4.4 satisfy (71a).*

A possible interest in the interval $[\varepsilon, \Delta]$ that appears on the right-hand of (71a) comes from recognizing that $\widehat{F}$ depends on h_{k+1} and this fact provides the freedom to limit the effective size of the step sizes. For example, let $\rho > 0$ and consider

$$\widehat{F}(x, y^+, h^+) = \frac{\rho}{\max\{\rho, h^+\}} \widetilde{F}(x, y^+) \tag{72}$$

where, for each solution $\mathbf{x}$ of (2) that is complete in the t direction, each $k \in \mathbb{Z}_{\geq 0}$, if $\widetilde{\mathbf{f}} \in \widetilde{F}(\mathbf{x}(k, \bar{\jmath}_k), \mathbf{y}_{k+1})$ is $\mathcal{F}_{k+1}$-measurable then

$$\mathbb{E}\left[\widetilde{\mathbf{f}} \,|\, \mathcal{F}_k\right] \in F(\mathbf{x}(k, \bar{\jmath}_k)). \tag{73}$$

This choice for $\widehat{F}$ is such that, when it is multiplied by the step-size placeholder h^+, we get an effective step size of $\min\{\rho, h^+\}$. Moreover,

$$\mathbb{E}\left[\frac{\rho}{\max\{\rho, h_{k+1}\}} \widetilde{\mathbf{f}} \;\middle|\; \mathcal{F}_k \right] \subset [\varepsilon, 1] F(\mathbf{x}(k, \bar{\jmath}_k)) \tag{74}$$

where

$$\varepsilon := \frac{\rho}{\max\left\{\rho, \sup_{k \in \mathbb{Z}_{\geq 0}} h_{k+1}\right\}}. \tag{75}$$

An important consequence of the variance condition (71b) in Assumption 8.1 is contained in the following lemma, which resembles [20, Lemma 4]. It uses the definitions

$$\tau_n := \sum_{k=0}^{n-1} h_{k+1}, \quad m(t) := \sup\{n \in \mathbb{Z}_{\geq 0} : \tau_n \leq t\}. \tag{76}$$

Lemma 8.3. *Suppose Assumptions 6.8 and 8.1 hold, and let* $\mathbf{x}$ *be a solution of* (2) *that is complete in the* t *direction and bounded, almost surely. For each* $k \in \mathbb{Z}_{\geq 0}$*, define (recall* $\underline{\jmath}_{k+1} = \bar{\jmath}_k$*)*

$$\widehat{\mathbf{f}}_{k+1} := \frac{\mathbf{x}(k+1, \underline{\jmath}_{k+1}) - \mathbf{x}(k, \bar{\jmath}_k)}{h_{k+1}} \in \widehat{F}(\mathbf{x}(k, \bar{\jmath}_k), \mathbf{y}_{k+1}, h_{k+1}), \tag{77a}$$

$$\mathbf{f}_k := \mathbb{E}\left[\widehat{\mathbf{f}}_{k+1} \,|\, \mathcal{F}_k \right]. \tag{77b}$$

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

$$\lim_{n \to \infty} \sup_{n+1 \leq k \leq m(\tau_n + T)} \left| \sum_{i=n}^{k-1} h_{i+1}(\widehat{\mathbf{f}}_{i+1} - \mathbf{f}_i) \right| = 0. \tag{78}$$

Proof. The proof follows that of [20, Lemma 4] with appropriate modifications to account for hybrid time domains. Let $\mathbf{s}_i : \Omega \to \mathbb{Z}_{\geq 0} \cup \{\infty\}$ be defined as

$$\mathbf{s}_i := \inf\{k \in \mathbb{Z}_{\geq 0} : |\mathbf{x}(k, \bar{\jmath}_k)| > i\}.$$

Then, for each $k \in \mathbb{Z}_{\geq 0}$,

$$\{\mathbf{s}_i \leq k\} = \bigcup_{\ell=0}^{k} \{|\mathbf{x}(\ell, \bar{\jmath}_\ell)| > i\} \in \mathcal{F}_k.$$

A consequence is that, for each $k \in \mathbb{Z}_{\geq 0}$, the indicator 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 $k \in \mathbb{Z}_{\geq 0}$, let $\widehat{\mathbf{f}}_{k+1}$ come from (77a) and define

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

Then

$$\mathbb{E}\left[\mathbf{u}^{\mathbf{s}_i}_{k+1} \,|\, \mathcal{F}_k \right] = \mathbb{E}\left[\widehat{\mathbf{f}}^{\mathbf{s}_i}_{k+1} \,|\, \mathcal{F}_k \right] - \mathbf{f}^{\mathbf{s}_i}_{k} = 0 \tag{79a}$$

$$\begin{aligned} \mathbb{E}\left[|\mathbf{u}^{\mathbf{s}_i}_{k+1}|^2 \,|\, \mathcal{F}_k \right] &\leq \mathbb{E}\left[|\widehat{\mathbf{f}}^{\mathbf{s}_i}_{k+1}|^2 \,|\, \mathcal{F}_k \right] \\ &= \mathbf{1}(k < \mathbf{s}_i)\mathbb{E}\left[|\widehat{\mathbf{f}}_{k+1}|^2 \,|\, \mathcal{F}_k \right] \\ &\leq \mathbf{1}(k < \mathbf{s}_i)\gamma(|\mathbf{x}_k|) \leq \gamma(i) \end{aligned} \tag{79b}$$

where (71b) and the definition of $\mathbf{s}_i$ have been used, respectively, for the last two inequalities in (79b). Using Assumption 6.8 and [9, Proposition 1.4] with $q = 2$, which is based on [58, Theorem 9], it follows that (78) holds with $\widehat{\mathbf{f}}^{\mathbf{s}_i}_{k+1}$ and $\mathbf{f}^{\mathbf{s}_i}_{k}$ in place of $\widehat{\mathbf{f}}_{k+1}$ and $\mathbf{f}_k$. Then, using that the solution $\mathbf{x}$ is bounded almost surely, for almost every $\omega \in \Omega$, there exists $i^* \in \mathbb{Z}_{\geq 0}$ such that $\mathbf{s}_i(\omega) = \infty$ for all $i \geq i^*$. Therefore, (78) holds almost surely. □

Our final assumption recasts [20, Assumption 3] in the setting of hybrid systems. It asks that the asymptotic behavior of the hybrid system (1) be contained in a compact set. The (infinite horizon) *reachable set* for (1), or for $\mathcal{H}$, from a set $K \subset \mathbb{R}^n$ is

$$\mathrm{Reach}^{\mathcal{H}}(K) := \left\{ \phi(t,j) \,:\, \phi \in \mathcal{S}^{\mathcal{H}}(K),\, (t,j) \in \mathrm{dom}\, \phi \right\}.$$

Assumption 8.4. *For the system (1), for every $i \in \mathbb{Z}_{\geq 0}$ the infinite horizon reachable set $\mathrm{Reach}^{\mathcal{H}}(i\mathbb{B})$ is bounded and the sequence $\{\mathrm{Omega}(i\mathbb{B})\}_{i=1}^{\infty}$ of Omega-limit sets of (1) is bounded.*

Remark 8.5. *The hybrid system involving resetting to a compact set, as specified in Example 6.3, satisfies Assumption 8.4, by construction. That the Bouncing Ball model in Example 2.4 satisfies Assumption 8.4 can be verified using the Lyapunov function candidate* $V(x) := 0.5x_2^2 + \gamma x_1$ *and the invariance principle for hybrid system [31]. That the combinatorial optimization on a torus model in Example 2.6 satisfies Assumption 8.4 can be verified using the Lyapunov function candidate* $V(x) := 0.5u^T u + f_q(z)$ *and the invariance principle for hybrid systems [31]. For the hybrid system in Example 2.5, Assumption 8.4 holds when, for each* $q \in \mathcal{Q}$, f_q *is radially unbounded and the zeros of* ∇f_q *belong to a compact set; under this condition, Assumption 8.4 can be verified using the Lyapunov function candidate* $V(x) := f_q(z)$ *and the invariance principle for hybrid systems. Finally, note that any hybrid system (1) that admits a globally asymptotically stable set satisfies Assumption 8.4.*

The proof of the following proposition uses standard ideas. In the setting of differential inclusions, such a result was given in [20, Proposition 1].

Proposition 8.6. *Under Assumptions 2.1, 2.2, and 8.4, the set*

$$\mathcal{A} := \lim_{i \to \infty} \mathrm{Omega}(i\mathbb{B}) \tag{80}$$

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

Proof. The limit $\lim_{i\to\infty} \mathrm{Omega}(i\mathbb{B}) =: \mathcal{A}$ exists, because the sequence $\{i\mathbb{B}\}_{i=1}^{\infty}$ is increasing, and so the sequence $\{\mathrm{Omega}(i\mathbb{B})\}_{i=1}^{\infty}$ is nondecreasing. Nonemptiness follows from the existence of at least one complete solution that is bounded, which in turn follows from Assumption 2.2 and Assumption 8.4. In fact, $\mathcal{A}$ is a nonempty, globally asymptotically stable compact set for (1). Indeed, since the sequence $\{\mathrm{Omega}(i\mathbb{B})\}_{i=1}^{\infty}$ is bounded, there exists a compact set $K \subset \mathbb{R}^n$ such that $\mathrm{Omega}(i\mathbb{B}) \subset \operatorname{int} K$ for all i. Take any large enough i so that $K \subset i\mathbb{B}$. Then $\mathrm{Omega}(i\mathbb{B}) \subset \operatorname{int} i\mathbb{B}$, and thus $\mathrm{Omega}(i\mathbb{B})$ is a nonempty, compact, weakly backward invariant, and asymptotically stable set with its basin of attraction containing $i\mathbb{B}$. These conclusions come from [14, Proposition 6.26, Proposition 7.5]. Because K is a compact neighborhood of $\mathrm{Omega}(i\mathbb{B})$ and is contained in the basin of attraction, there exists $T > 0$ such that the set of points reachable from $i\mathbb{B}$ at times greater than or equal to T is contained in K. This ensures that $\mathrm{Omega}(i\mathbb{B}) = \mathrm{Omega}(K)$. Thus, for all large enough i, the sequence $\{\mathrm{Omega}(i\mathbb{B})\}_{i=1}^{\infty}$ is constant, with

$\text{Omega}(i\mathbb{B}) = \text{Omega}(K)$, and so $\mathcal{A} = \text{Omega}(K)$. The basin of attraction of $\mathcal{A}$ contains $i\mathbb{B}$ for all large enough i, so that the basin equals $\mathbb{R}^n$. ◻

The following theorem, which can be viewed as the main technical contribution of this paper, extends [20, Theorem 2] from differential inclusions to hybrid inclusions. It provides mild conditions under which a solution that is bounded and complete in the t direction converges to every Morse decomposition of the global attractor. In addition to all of the previous assumptions except Assumption 6.1, it imposes the following condition on the step sizes:

Assumption 8.7. *The step sizes $\{h_k\}_{k=1}^{\infty}$ are not summable:* $\displaystyle\sum_{k=1}^{\infty} h_k = \infty$.

Theorem 8.8. *Let Assumptions 2.1, 2.2, 6.8, 8.1, 8.4, and 8.7 hold, let $\mathcal{D} = \{M_1, \ldots, M_\ell\}$ be a Morse decomposition of $\mathcal{A} := \lim_{i\to\infty} \text{Omega}(i\mathbb{B})$, and recall that $\mathcal{M}(\mathcal{D}) = \cup_{i=1}^{\ell} M_i$. If $\mathbf{x}$ is a solution to (2) that is complete in the t direction and almost surely bounded then $\mathbf{x}(k, j)$ converges, as $k+j \to \infty$ with $(k, j) \in \text{dom }\mathbf{x}$, to $\mathcal{M}(\mathcal{D})$ almost surely.*

Proof. The proof follows the lines of the proof of [20, Theorem 2], with modifications to account for two different types of jumps and hybrid time domains. Because of Assumptions 2.1, 2.2, and 8.4, Proposition 8.6 holds. In turn, Theorem 7.3 can be applied to give functions $U, w : \mathbb{R}^n \to \mathbb{R}_{\geq 0}$ that parallel V and w of that theorem. Let $\kappa : \mathbb{R}_{\geq 0} \to \mathbb{R}_{\geq 0}$ be a smooth, concave function such that

$$\kappa'(s) > 0 \qquad \forall s > 0, \tag{81a}$$
$$|\kappa'(U(x))\nabla^2 U(x)| \leq L \qquad \forall x \in C. \tag{81b}$$

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

$$V(x) := \kappa(U(x)), \tag{82a}$$
$$Y_c(x) := \kappa'(U(x))w(x), \tag{82b}$$
$$Y_d(x) := \kappa(U(x)) - \kappa(U(x) - w(x)), \tag{82c}$$
$$Y(x) := \min\{Y_c(x), Y_d(x)\}. \tag{82d}$$

It follows from Theorem 7.3 and the properties of κ that V is constant with a distinct value on each Morse set and smooth with a Lipschitz gradient having

Lipschitz constant L, Y is continuous and positive definite with respect to $\mathcal{M}_{\mathcal{D}}$, and

$$\begin{aligned} \nabla V(x) \cdot f &\leq -Y(x) \qquad \forall x \in C, f \in F(x) \qquad &(83a)\\ V(g) - V(x) &\leq -Y(x) \qquad \forall x \in D, g \in G(x). \qquad &(83b) \end{aligned}$$

Like in the proof of Theorem 6.9, we use the inequality (58), we let $x := \mathbf{x}(k, \overline{\jmath}_k)$ and we let η be $\mathcal{F}_{k+1}$-measurable such that

$$x + \eta = \mathbf{x}(k+1, \underline{\jmath}_{k+1}) \in \Big(\mathbf{x}(k, \overline{\jmath}_k) + h_{k+1}\widehat{F}(\mathbf{x}(k, \overline{\jmath}_k), \mathbf{y}_{k+1}, h_{k+1})\Big) \cap (C \cup D) \tag{84}$$

and then take conditional expectations, and use Assumption 8.1 to get, for all $k \in \mathbb{Z}_{\geq 0}$,

$$\begin{aligned} &\mathbb{E}\left[V\left(\mathbf{x}(k+1, \underline{\jmath}_{k+1})\right) \mid \mathcal{F}_k\right] \\ &\qquad \leq V(\mathbf{x}(k, \overline{\jmath}_k)) - h_{k+1}\varepsilon Y(\mathbf{x}(k, \overline{\jmath}_k)) + h_{k+1}^2\left(1 + \frac{L}{2}\right)\gamma(|\mathbf{x}(k, \overline{\jmath}_k)|) \end{aligned} \tag{85}$$

and

$$V\left(\mathbf{x}(k, j+1)\right) \leq V(\mathbf{x}(k, j)) - Y(\mathbf{x}(k, j)) \qquad \forall j \in \left\{\underline{\jmath}_k, \ldots, \overline{\jmath}_k - 1\right\}. \tag{86}$$

In preparation for applying Theorem 5.3, for each $k \in \mathbb{Z}_{\geq 0}$ and each $j \in \left\{\underline{\jmath}_k, \ldots, \overline{\jmath}_k\right\}$, define

$$\begin{aligned} \mathbf{z}(k, j) &:= V(\mathbf{x}(k, j)), \qquad &(87a)\\ \mathbf{w}(k, j) &:= 0, \qquad &(87b)\\ \mathbf{v}(k, j) &:= h_{k+1}^2\left(1 + \frac{L}{2}\right)\gamma(|\mathbf{x}(k, j)|), \qquad &(87c)\\ \mathbf{u}_c(k, j) &= h_{k+1}\varepsilon Y(\mathbf{x}(k, j)), \qquad &(87d)\\ \mathbf{u}_d(k, j) &:= Y(\mathbf{x}(k, j)). \qquad &(87e) \end{aligned}$$

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

$$\sum_{k=0}^{\infty} \mathbf{v}(k, \overline{\jmath}_k) \leq \gamma\left(\sup_{k \in \mathbb{Z}_{\geq 0}} |\mathbf{x}(k, \overline{\jmath}_k)|\right) \sum_{k=0}^{\infty} h_{k+1}^2\left(1 + \frac{L}{2}\right) < \infty, \tag{88}$$

where the finiteness uses the assumption that $\mathbf{x}$ is bounded almost surely and Assumption 6.8, which says that the sequence of step sizes is square summable. Due to Theorem 5.3 with the definitions in (87), for almost all $\omega \in \Omega_i$

$$\lim_{k+j\to\infty,(k,j)\in\operatorname{dom}\mathbf{x}} V(\mathbf{x}(k,j))$$

exists and is finite and

$$\sum_{(k,j)\in\operatorname{dom}\mathbf{x}} \min(1,\varepsilon h_{k+1}) Y(\mathbf{x}(k,j)) \leq \sum_{k=0}^{\infty}\left(\mathbf{u}_c(k,\overline{\jmath}_k) + \sum_{j=\underline{\jmath}_k}^{\overline{\jmath}_k-1} \mathbf{u}_d(k,j)\right) < \infty. \tag{89}$$

Finally we establish that

$$\lim_{k+j\to\infty,(k,j)\in\operatorname{dom}\mathbf{x}} Y(\mathbf{x}(k,j)) = 0 \tag{90}$$

for almost all $\omega \in \Omega$ which gives the desired result since Y is continuous and positive definite with respect to $\mathcal{M}_{\mathcal{D}}$ and $\mathbf{x}$ is bounded almost surely.

Take $\widehat{\Omega} \subset \Omega$ to satisfy $\mathbb{P}(\widehat{\Omega}) = 1$ and be such that, for all $\omega \in \widehat{\Omega}$, $\mathbf{x}(\omega)$ is bounded and (89) and (78) hold. Fix $\omega \in \widehat{\Omega}$.

Suppose $\limsup_{(k,j)\in\operatorname{dom}\mathbf{x},k+j\to\infty} Y(\mathbf{x}(k,j)) > 0$. It follows that there is an accumulation point $x^* \in \mathbb{R}^n$ for the solution $\mathbf{x}$ satisfying $\varepsilon := Y(x^*) > 0$. In fact, x^* must be an accumulation point of a sequence of the form $\left\{\mathbf{x}(k_\ell,\overline{\jmath}_{k_\ell})\right\}_{\ell=1}^{\infty}$ where $k_\ell \to \infty$ as $\ell \to \infty$ and $\overline{\jmath}_{k_\ell} = \underline{\jmath}_{k_\ell}$ for all $\ell \in \mathbb{Z}_{\geq 1}$ due to (89) and the definition of $\mathbf{u}_d$ in (87e). Since Y is continuous, there exists $\delta > 0$ such that

$$Y(x) \geq 0.5\varepsilon \qquad \forall x \in \{x^*\} + \delta\mathbb{B}. \tag{91}$$

Since $\mathbf{x}$ is bounded and F is locally bounded, with the definition of $\mathbf{f}_k$ in (77b) and the condition (71a), there exists $T > 0$ sufficiently small so that

$$T|\mathbf{f}_k| \leq \delta/3 \qquad \forall k \in \mathbb{Z}_{\geq 0}. \tag{92}$$

Using Lemma 8.3, in particular the definitions in (77) and the limit in (78), let $n^* \in \mathbb{Z}_{\geq 0}$ be large enough so that, for all $n \in \mathbb{Z}_{\geq n^*}$,

$$\sup_{n+1\leq k\leq m(\tau_n+T)} \left|\sum_{i=n}^{k-1} h_{i+1}\left(\widehat{\mathbf{f}}_{i+1} - \mathbf{f}_i\right)\right| \leq \delta/3. \tag{93}$$

Next, let $\{n_\ell\}_{\ell=1}^{\infty}$ be a sequence of positive integers that is unbounded and satisfies, for all $\ell \in \mathbb{Z}_{\geq 1}$,

$$n_{\ell+1} \geq \max\{n^*, m(\tau_{n_\ell} + T)\} \tag{94a}$$
$$\mathbf{x}_{n_\ell} := \mathbf{x}(n_\ell, \bar{\jmath}_{n_\ell}) \in \{x^*\} + (\delta/3)\mathbb{B}. \tag{94b}$$

Let $\ell_1 \in \mathbb{Z}_{\geq 1}$ be sufficiently large such that if $(k, j), (k, j+1) \in \operatorname{dom} \mathbf{x}$ and $(k, j) \succeq (n_{\ell_1}, \bar{\jmath}_{n_{\ell_1}})$ then $Y(\mathbf{x}(k, j)) \leq 0.25\varepsilon$. Such an ℓ_1 exists due to (89) and the definition of $\mathbf{u}_d$ in (87e). Let $\ell_2^* \in \mathbb{Z}_{\geq \ell_1^*}$ be large enough so that

$$\sum_{k=n_\ell}^{m(\tau_{n_\ell}+T)-1} h_{k+1} \in [0.5T, T] \qquad \forall \ell \in \mathbb{Z}_{\geq \ell_2^*}. \tag{95}$$

Let $s \in \{n_\ell, \ldots, m(\tau_{n_\ell} + T)\}$ be such that

$$(k, \bar{\jmath}_{n_\ell}) \in \operatorname{dom} \mathbf{x} \qquad \forall k \in \{n_\ell, \ldots, s\} \tag{96}$$

and define $\mathbf{x}_k := \mathbf{x}(k, \bar{\jmath}_{n_\ell})$ for all $k \in \{n_\ell, \ldots, s\}$. Note that, for each such k,

$$\mathbf{x}_k - \mathbf{x}_{n_\ell} = \sum_{i=n_\ell}^{k-1} h_{i+1}\widehat{f}_{i+1}. \tag{97}$$

Then, using (92), (93), and (94b),

$$\begin{aligned} &\sup_{n_\ell \leq k \leq s} |\mathbf{x}_k - x^*| \\ &= \sup_{n_\ell \leq k \leq s} \left| \mathbf{x}_k - \mathbf{x}_{n_\ell} - \sum_{i=n_\ell}^{k-1} h_{i+1}\mathbf{f}_i + \sum_{i=n_\ell}^{k-1} h_{i+1}\mathbf{f}_i + \mathbf{x}_{n_\ell} - x^* \right| \\ &\leq \sup_{n_\ell \leq k \leq m(\tau_{n_\ell}+T)} \left| \sum_{i=n_\ell}^{k-1} h_{i+1}(\widehat{\mathbf{f}}_{i+1} - \mathbf{f}_i) + \sum_{i=n_\ell}^{k-1} h_{i+1}\mathbf{f}_i + \mathbf{x}_{n_\ell} - x^* \right| \leq \delta. \end{aligned} \tag{98}$$

This bound, (91), and the definition of ℓ_1, yield that we can take $s = m(\tau_{n_\ell} + T)$ in (96). Moreover, from the definition of ℓ_2^*, for $\ell \in \mathbb{Z}_{\geq \ell_2^*}$,

$$\sum_{k=n_\ell}^{m(\tau_{n_\ell}+T)-1} h_{k+1} Y(\mathbf{x}(k, \bar{\jmath}_{n_\ell})) \geq 0.5\varepsilon \sum_{k=n_\ell}^{m(\tau_{n_\ell}+T)-1} h_{k+1} \geq 0.25\varepsilon T. \tag{99}$$

When combined with (94a), summing this lower bound over ℓ contradicts (89). Thus (90) holds, which implies that $\mathbf{x}$ converges to $\mathcal{M}_{\mathcal{D}}$. $\square$

With Theorem 8.8 in hand, the fact that the chain recurrent part of $\mathcal{A}$ has representation (67), and the fact that the number of Morse decompositions is, at most, countable, we arrive at the following corollary that is related to results in [9] and [10] that apply to differential inclusions.

Corollary 8.9. *Under the assumptions of Theorem 8.8, every solution* $\mathbf{x}$ *to (2) that is complete in the* t *direction converges almost surely to the chain-recurrent part of* $\mathcal{A} := \lim_{i\to\infty} \mathrm{Omega}(i\mathbb{B})$.

An alternative approach to establishing the above corollary is to invoke the recent converse result in [19], which provides a total Lyapunov function, called there a Conley-Lyapunov function, that characterizes the chain recurrent part of $\mathcal{A}$ (recall (67)). That function can then be used directly, as in the proof of Theorem 8.8, in lieu of using the Morse-Lyapunov function from Theorem 7.3."

Example 8.10 (Academic example, simulated with noise). *Consider the academic system introduced in Example 2.3 with data* (C, F, D, G)*, and revisited in Example 7.2, and suppose that, for all* $x \in C$,

$$\widehat{F}(x, y^+) := F(x) + y^+ \tag{100}$$

where y^+ *is a placeholder for a sequence of random variables* $\{\mathbf{y}_{k+1}\}_{k=0}^{\infty}$ *each with mean zero and the same compact support* $[-\alpha, \alpha]$ *where* $\alpha > 0$. *It is not difficult to see that (71a) holds with* $\varepsilon = 1$ *and the function* γ *identically equal to zero. In addition, (71b) holds with any continuous, nondecreasing function* $\gamma : \mathbb{R}_{\geq 0} \to \mathbb{R}_{\geq 0}$ *satisfying*

$$|F(x)|^2 + \alpha^2 \leq \gamma(|x|) \quad \forall x \in C. \tag{101}$$

It is not possible to satisfy the variance condition (38b) in Assumption 6.1 since the variance of $\widehat{F}(x, y^+)$ *is lower bounded by* $|F(x)|^2$ *and* V *is required to have a Lipschitz gradient, so the bound (38b) fails for* $x \in C$ *with large enough absolute value. Thus, it is impossible to verify the conditions of Assumption 6.1 for almost sure boundedness. However, Theorem 8.8 and Corollary 8.9 do not require Assumption 6.1. Instead, these results contain statements about solutions that have been established to be almost surely bounded, by* *whatever means. We can establish the almost sure boundedness property for any bounded sequence of step sizes* $\{h_{k+1}\}_{k=0}^{\infty}$ *as follows. Let* $\beta > 0$ *be such that* $h_{k+1} \in [0, \beta]$ *for all* $k \in \mathbb{Z}_{\geq 0}$. *Let* $\nu > 0$ *be such that* $x \leq -\nu$ *implies* $F(x) \geq \alpha$. *Then, for each* $k \in \mathbb{Z}_{\geq 0}$

$$x \leq x + h_{k+1}\left(F(x) - \alpha\right) \leq x + h_{k+1}\left(F(x) + \mathbf{y}_{k+1}\right) \leq x + \beta\left(F(x) + \alpha\right). \tag{102}$$

These bounds imply that, after a jump due to $x \in C \cap (-\infty, -\nu]$, x *lands in the interval*

$$\mathcal{I}_1(x) := [x, x + \beta\left(F(x) + \alpha\right)]. \tag{103}$$

Let

$$\mu_1 := \min_{x \in [-\nu, 5]} (x - \beta\left(|F(x)| + \alpha\right)) < 0 \tag{104a}$$

$$\mu_2 := \max_{x \in [-\nu, 5]} (x + \beta\left(|F(x)| + \alpha\right)) > 0 \tag{104b}$$

For jumps from $x \in C \cap [-\nu, \infty) = [-\nu, 5]$, x *lands in the interval*

$$\mathcal{I}_2 := [\mu_1, \mu_2] \supset [x - \beta\left(|F(x)| + \alpha\right), x + \beta\left(|F(x)| + \alpha\right)]. \tag{105}$$

Finally, for jumps from $x \in D$, x *lands in*

$$\mathcal{I}_3 := [0, 3] \supset G(D) = \{0, 3\}. \tag{106}$$

Given that $C = (-\infty, 5]$, *it follows each sample path evolves in the set*

$$\mathcal{I}_1(x(0,0)) \cup \mathcal{I}_2 \cup \mathcal{I}_3 \cup \left(\bigcup_{x \in \mathcal{I}_2 \cup \mathcal{I}_3} \mathcal{I}_1(x) \right). \tag{107}$$

This fact implies that every solution is bounded almost surely.

Since $C \cup D = \mathbb{R}$, *complete solutions exist. Since* $G(D) \cap D$ *is empty, every complete solution is complete in the* t *direction. Hence, we conclude from Corollary 8.9 that every complete solution of (2) converges to the chain recurrent part of* $[0, 5]$ *when the step sizes satisfy Assumptions 6.8 and 8.7.*

9. Simulations and examples

This section provides simulations for two examples considered earlier: the Bouncing Ball in Example 2.4 that is considered further in Example 6.4, and the academic Example 2.3 that is considridred further in Example 7.2.

9.1. Implementation of the Stochastic Hybrid System Simulator

We introduce an algorithm to compute solutions to the stochastic hybrid simulator (2). The inputs to the algorithm are as follows:

- The flow set C, the jump set D, and the jump map G of the hybrid system $\mathcal{H}$ in (1).
- The simulation horizon $(N, J) \in \mathbb{Z}^2_{\geq 0}$.
- The deterministic sequence $\{h_k\}_{k=1}^N$ of positive step sizes.
- The sequence of random variables $\{\mathbf{y}_k\}_{k=1}^N$ defined on a probability space $(\Omega, \mathcal{F}, \mathbb{P})$, where $\mathbf{y}_k : \Omega \to Y \subset \mathbb{R}^m$ for each $k \in \{1, 2, \ldots, N\}$.
- The map $\widehat{F}$ used to approximate the flows of $\mathcal{H}$.
- The initial condition x_0.

Roughly speaking, the algorithm uses this information to compute a sample path of a solution to (2) by evaluating the flow condition $x \in C$ and the jump condition $x \in D$, and according to the result of this evaluation, by appropriately generating the new value of the sample path via (2a) or (2b). Details are provided next.

9.1.1. Algorithm

Our algorithm to generate sample paths of solutions to (2) is given in Algorithm 1.

For the given inputs, the algorithm begins from the system's initial state with zero values for the counter k and the jump counter j. If the state reaches the jump set D at a point not in C, a discrete update is applied using the jump map G, and the jump count is incremented—steps 7-12. This part of the algorithm implements (2b). The state is updated using the approximation of the flows when it is C and not in D—steps 14-18–in this way, implementing (2a). When the state is in both C and D, the algorithm chooses between jumping and flowing in a way that ensures the solution is adapted to the filtration $\{\mathcal{F}_k\}_{k=0}^\infty$; i.e., the mapping in (13) is $\mathcal{F}_k$-measurable for each $k \in \mathbb{Z}_{\geq 0}$. This can be achieved by, for example, always choosing to jump when in the jump set or always choosing to flow when in the flow set. The simulation continues until either N iterations in the k direction or J jumps in the j direction occur, or the state does not belong to $C \cup D$.

Algorithm 1 Algorithm for the computation of a sample path of (2).

1: **Inputs:** C, D, and G of $\mathcal{H}$; simulation horizon (N, J); step sizes $\{h_k\}_{k=1}^N$; $\omega \in \Omega$ and noise $\{y_k = \mathbf{y}_k(\omega)\}_{k=1}^N$; $\widehat{F}$; initial condition x_0.
2: **Initialize Variables:**
3: Set $k = 0$, $j = 0$, $x = x_0$.
4: Create empty arrays to store the sample path.
5: **Compute sample path:**
6: **while** $k \leq N$ and $j \leq J$ **do**
7:     **if** $x \notin C$ and $x \in D$ **then** ▷ Trigger jumps
8:         Update the new value of the state using (2b):
9:         $x \leftarrow G(x)$ ▷ Update using the jump map
10:         $k \leftarrow k$ ▷ Keep step index constant
11:         $j \leftarrow j + 1$ ▷ Increment jump count
12:         Store (k, j, x)
13:     **end if**
14:     **if** $x \in C$ and $x \notin D$ **then** ▷ Approximate flows
15:         Compute the new value of the state using (2a):
16:         $x \leftarrow x + h_{k+1}\widehat{F}(x, y_{k+1}, h_{k+1})$
17:         Store (k, j, x)
18:     **end if**
19:     **if** $x \in C$ and $x \in D$ **then** ▷ Jumps and flow possible
20:         Pick between triggering jumps (steps 8-12) or allowing flow (steps 15-17), while ensuring that the solution is adapted.
21:     **end if**
22:     **if** $x \notin C$ and $x \notin D$ **then** ▷ State left the flow and jump set
23:         **Break.**
24:     **end if**
25: **end while**
26: **Return** sample path $(k, j, x(k, j))$.

9.1.2. Software Implementation

Similar to the Hybrid Equations Toolbox [59], we implement Algorithm 1 in Matlab for the case when $\widehat{F}$ and G are single-valued maps. Sample paths are computed using four Matlab functions encoding the data of (2) and a Matlab script that computes the sample path. The Matlab functions are defined as follows:

i) The flow set C is implemented in the Matlab function `C.m`. The argument of this function is the state of the system. The function returns 1 if the state belongs to the set C, and 0 otherwise.
ii) The map $\widehat{F}$ is implemented by the Matlab function `Fhat.m`. This function takes as inputs the current value of the state, the noise y_k, and the sequence h_k. The function returns the value of $\hat{F}$ for the given inputs.
iii) The jump set D is encoded by the Matlab function `D.m`. As for the flow set C, the argument of this function is the state of the system. The function returns 1 if the state belongs to D, and 0 otherwise.
iv) The jump map is defined in the Matlab function `G.m`, which evaluates G at the given state.

The Matlab script `HySAsolver.m` uses these functions to compute a sample path. The syntax to perform a simulation using `HySAsolver.m` is as follows:

```
[t j x] = HySAsolver(@C,@Fhat,@D,@G,x0,N,J,rule,options);
```

The first four arguments of `HySAsolver` are the data of (2). The argument `x0` is the initial condition, which is a $n \times 1$ vector. The parameters N and J define the simulation horizon. The scalar parameter `rule` determines how the simulator continues from a state that is in both C and D. If `rule`=1, then jumps are always triggered from such points. If `rule`=2, then flows are triggered from such points. When `rule`=3 is chosen, the simulator selects flow or jump consistent with ensuring that the solution is adapted to the filtration $\{\mathcal{F}_k\}_{k=0}^{\infty}$. The parameter `options` selects the sequence h_k and the noise signal y_k to use in the simulation.

`HySAsolver.m` returns the computed sample path along with its discrete hybrid time domain. The Matlab script `run.m` is provided to configure the simulation parameters and run `HySAsolver`. This script is also used to plot sample paths.[2]

[2]Code available at `https://github.com/HybridSystemsLab/HySAsolver`

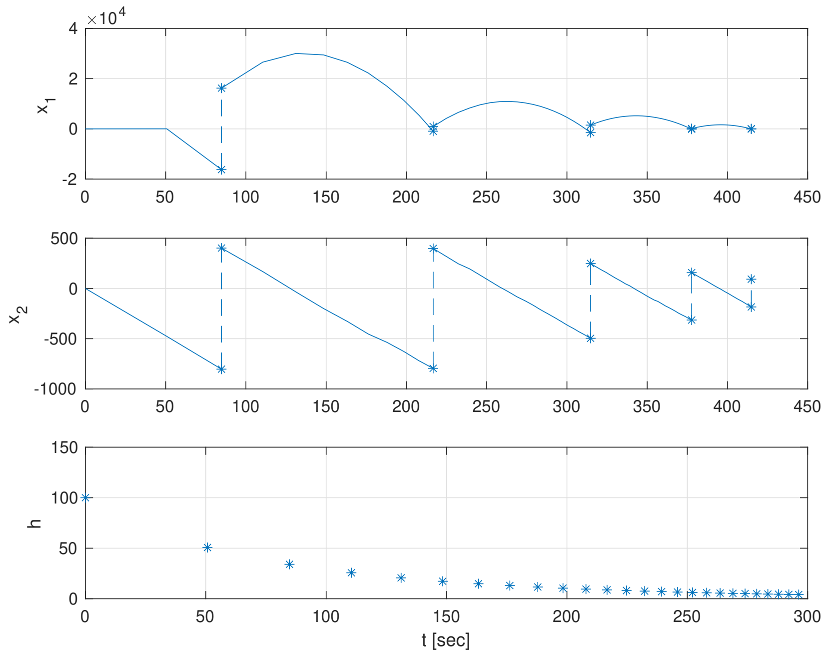

Figure 3: Position, velocity, and sequence h_k for the Bouncing Ball system revisited in Example 9.1. The values of h_k are denoted by $*$ in the bottom plot.

9.2. Examples

Next, two of the examples introduced earlier are simulated by the proposed algorithm.

Example 9.1 (Bouncing Ball revisited). *We compute sample paths of solutions to* (2) *associated with the Bouncing Ball system in Example 6.4. To guarantee existence of complete solutions, the model with the jump set $\hat{D}$ given therein is used. The gravity parameter is set to $\gamma = 9.8m/s^2$ and the restitution coefficient to $\delta = 0.8$. The sequence of step sizes*[3] *used is $h_k := \frac{a}{k^b}$*

[3]The choice $a > 0$ and $b \in (1/2, 1]$ assures that h_k is square summable but not summable, as required for convergence in Theorem 8.8. Picking a large leads to large initial steps in the sequence h_k. The value of b regulates the rate of decay to zero of h_k since h_k can be written as ah^{-b}. For instance, a choice of $b \in (1/2, 1]$ closer to zero leads to a slower rate than choosing b close to one. In practice, note that small step sizes would

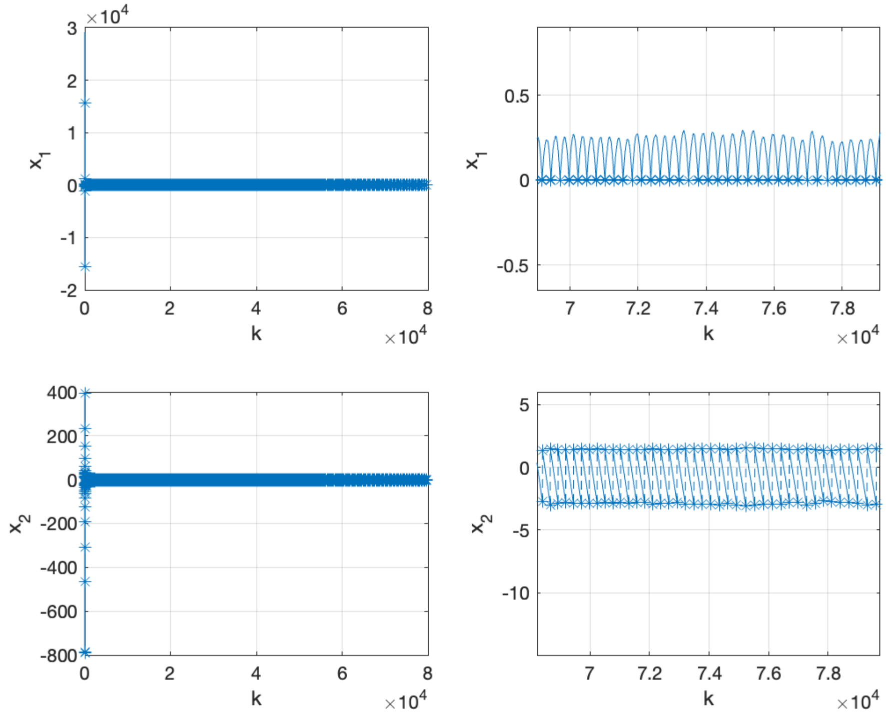

Figure 4: Position and velocity for the Bouncing Ball system revisited in Example 9.1 for a long simulation horizon. Each data point in the (discretized) solution component is denoted by $*$.

with $a = 100$ *and* $b = 0.98$*. The noise signal* y_k *is generated using an exponential distribution with variance equal to* $\frac{1}{2}$*. The map* $\widehat{F}$ *is given in (16) as* $\widehat{F}(x, y^+, h^+) = F(x) - (0, y^+)$*. Figure 3 shows a sample path to this system from* $x(0, 0) = (1, 0)$*, which is in the interior of the flow set, for a simulation horizon given by* $N = 20000$ *and* $J = 5$*. The simulation stops after* J *jumps are computed—the simulation performs* 1208 *iterations of* (2a)*. The position and velocity components of the sample path are shown with respect to the partial sums of* h_k*, which relate* k *to ordinary time* t*. After two itera-*

lead to an effective simulation horizon that might be too small to be useful, due to h_k approaching a value close to zero after just a few steps.

tions, these components of the computed sample path resemble trajectories of the bouncing ball. The figure also shows the first 300 terms of the sequence h_k.

Figure 4 shows a sample path obtained for a longer simulation horizon with $N = 5 \times 10^5$ *and* $J = 1000$*, now as a function of* k*. The zoomed plots show that the state at the tail of the sample path. By performing simulations with increasing* N*, it is confirmed that the supremum of the tail of each component of the sample paths decreases as* N *increases. As guaranteed by Theorem 8.8, the sample path converges to the system's global attractor, which is the origin.*

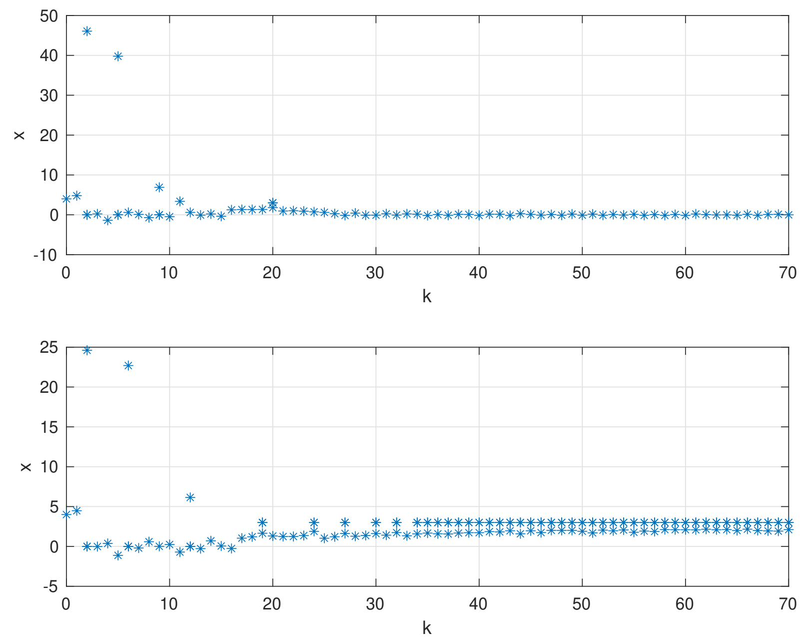

Figure 5: Two sample paths for the academic Example 7.2, revisited in Example 9.2, converging to different components of the chain-recurrent set in $\mathcal{A}$.

Example 9.2 (Academic example re-revisited). *For the scalar hybrid dynamics in the academic Example 2.3 and its stochastic simulator in Example 8.10, we compute sample paths of solutions to* (2)*. Similar to the* *previous example, the sequence of step sizes used is $h_k := \frac{a}{k^b}$ with $a = 2$ and $b = 0.6$ and the map $\widehat{F}$ is given in (100) as $\widehat{F}(x, y^+, h^+) = F(x) + y^+$. The noise signal $\mathbf{y}_k$ is generated using a uniform distribution with compact support $[-1, 1]$. For the simulation horizon $N = 100$ and $J = 50$, Figure 5 shows two sample paths from the same initial condition and parameters asymptotically converging to two different subsets of the global attractor $\mathcal{A} = [0, 5]$. The sample path in the top plot approaches the set $M_1 = \{0\}$. On the other hand, the bottom plot shows convergence to the interval $M_2 = [2, 3]$. These sample paths confirm the findings in Example 7.2, where a Morse decomposition of the academic Example 2.3 was presented, and the results of Theorem 8.8 and Corollary 8.9.*

10. A proof of the Morse-Lyapunov function existence

Before proving Theorem 7.3, a preliminary result is needed, on a Lyapunov-like characterization of each attractor in a Morse decomposition. The conclusions of item (b) of Proposition 10.1, in a narrower setting of differential inclusions, were stated and proven in [22, Theorem 3.1]. The proof of item (b) follows the ideas from the proof of [22, Theorem 3.1]. The proof here is, however, simpler despite the more general setting of hybrid inclusions. In [22], the authors approximated the differential inclusion using a Lipschitz continuous one and had a sequence of approximations of the asymptotically stable sets to deal with. Thanks to Lemma 7.1, such steps are not necessary below.

Proposition 10.1. *Let $A \subset \mathbb{R}^n$ be a nonempty asymptotically stable set and $\mathcal{B}$ its basin of attraction.*

(a) If $\mathcal{B} = \mathbb{R}^n$, then there exists a C^∞ and radially unbounded function $W : \mathbb{R}^n \to \mathbb{R}_{\geq 0}$ with $W(x) = 0$ if and only if $x \in \mathcal{A}$ such that

$$\begin{aligned} \nabla W(x) \cdot f &\leq -W(x) && \forall x \in C,\ f \in F(x), \\ W(g) &\leq W(x) - \left(1 - \tfrac{1}{e}\right) W(x) && \forall x \in D,\ g \in G(x). \end{aligned} \tag{108}$$

(b) If $\mathcal{B} \neq \mathbb{R}^n$, then there exists a C^∞ function $W : \mathbb{R}^n \to [0, 1]$ with $W(x) = 0$ if and only if $x \in A$ and with $W(x) = 1$ if and only if

$x \notin \mathcal{B}(A)$, and a continuous function $w : \mathbb{R}^n \to \mathbb{R}_{\geq 0}$ with $w(x) = 0$ if and only if $x \in A$ or $x \notin \mathcal{B}(A)$, such that

$$\begin{aligned} \nabla W(x) \cdot f &\leq -w(x) & \forall x \in C,\, f \in F(x), \\ W(g) &\leq W(x) - w(x) & \forall x \in D,\, g \in G(x). \end{aligned} \tag{109}$$

Proof. Item (a) follows from taking the V from Lemma 7.1, noting that $\mathcal{B} = \mathbb{R}^n$, and setting $W = V$. Item (b) requires some effort. Let a smooth $V : \mathcal{B} \to \mathbb{R}_{\geq 0}$ come from Lemma 7.1. Extend V from $\mathcal{B}$ to $\mathbb{R}^n$ by setting $V(x) = \infty$ if $x \notin \mathcal{B}$. For $n \in \mathbb{N}$, let $V_n : \mathbb{R}^n \to \mathbb{R}_{\geq 0}$ be defined by

$$V_n(x) := \min\{V(x), n\},$$

so that $V_n(x) = V(x)$ on $V_{\leq n}$ and $V_n(x) = n$ on $V_{\geq n}$, including on $\mathbb{R}^n \setminus \mathcal{B}$. Here, and below, $V_{<n} = \{x \in \mathbb{R}^n \,|\, V(x) < n\}$, $V_{\leq n}$, $V_{\geq n}$ are the sublevel and superlevel sets of (the extended) V. Let

$$L(s) = \operatorname{sign}(s) e^{-1/s^2},$$

with the convention that $L(0) = 0$. This function is $\mathcal{C}^\infty$, with $L^{(k)}(0) = 0$ for every $k \in \mathbb{N}$, and $L'(s) > 0$ for all $s \neq 0$. Let $W_n : \mathcal{B} \to [0, L(n)]$ be

$$W_n(x) := L(V_n(x) - n) + L(n).$$

Then $W_n(x) = 0$ iff $x \in A$ and $W_n(x) = L(n)$ iff $x \in V_{\geq n}$.

For every $x \in C \cap V_{<n}$ and $f \in F(x)$, one has

$$\begin{aligned} \nabla W_n(x) \cdot f &= L'(V_n(x) - n) \nabla V_n(x) \cdot f = L'(V_n(x) - n) \nabla V(x) \cdot f \\ &\leq -L'(V_n(x) - n) V(x) = -w_n^c(x), \end{aligned}$$

where $w_n^c : \mathbb{R}^n \to \mathbb{R}_{\geq 0}$ is defined by $w_n^c(x) = L'(V_n(x) - n) V(x)$ for $x \in V_{<x}$ and $w_n^c(x) = 0$ for $x \in V_{\geq n}$. This w_n^c is continuous on $\mathbb{R}^n$ and the inequality above holds for $x \in C \cap V_{\geq n}$ too, because there, $\nabla W_n(x) = 0$ and $w_n^c(x) = 0$.

For every $x \in D \cap V_{\leq n}$ and $g \in G(x)$, one has

$$\begin{aligned} W_n(g) - W_n(x) &= L(V_n(g) - n) - L(V_n(x) - n) \\ &\leq L(e^{-1} V_n(x) - n) - L(V_n(x) - n) \leq -w_n^d(x), \end{aligned}$$

where $w_n^d : \mathbb{R}^n \to \mathbb{R}_{\geq 0}$ is any continuous function such that $0 < w_n^d(x) \leq L(V_n(x) - n) - L(e^{-1} V_n(x) - n)$ for $x \in V_{<n}$ and $w_n^d(x) = 0$ for $x \in V_{\geq n}$. For

such a function, the inequalities above hold for $x \in D \cap V_{\geq n}$ too, because there, $W_n(x) = L(n)$, $W_n(g) \leq L(n)$, and $w_n^d(x) = 0$.

Let $w_n : \mathbb{R}^n \to \mathbb{R}_{\geq 0}$ be the continuous function given by $w_n(x) = \min\{w_n^c(x), w_n^d(x)\}$ at each $x \in \mathbb{R}^n$. It is positive on $V_{<n}$ and 0 on $V_{\geq n}$, and because $V_{<n} \subset V_{\leq n}$ and the latter is compact, w_n is bounded. Let the bound be $b_n > 0$.

Because $W_n(x) = L(n)$ on $V_{\geq n}$, and $V_{\leq n}$ is compact, every partial derivative of W_n, of any order, is bounded on $\mathbb{R}^n$. For $n \in \mathbb{N}$, let $a_n > 0$ be such that, for all $x \in \mathbb{R}^n$,

$$\|\nabla W_n(x)\| \leq a_n, \|\nabla^2 W_n(x)\| \leq a_n, \dots \|\nabla^{(n)} W_n(x)\| \leq a_n.$$

Let $c_n = a_n + b_n + L(n) < a_n + b_n + 1$ and

$$\gamma := \left(\sum_{n=1}^{\infty} \frac{1}{2^n c_n} L(n) \right)^{-1}.$$

Define $W : \mathbb{R}^n \to [0, 1]$, $w : \mathbb{R}^n \to \mathbb{R}_{\geq 0}$ by setting, for each $x \in \mathbb{R}^n$,

$$W(x) := \gamma \sum_{n=1}^{\infty} \frac{1}{2^n c_n} W_n(x), \quad w(x) := \gamma \sum_{n=1}^{\infty} \frac{1}{2^n c_n} w_n(x).$$

The choice of coefficients ensures that both series converge uniformly.

If $x \in A$ then each $W_n(x) = 0$, so $W(x) = 0$. If $x \notin \mathcal{B}(A)$, then each $W_n(x) = L(n)$, so $W(x) = 1$. If $x \in \mathcal{B}(A) \setminus A$, then $x \in V_{<n}$ for some n, so $W_n(x) \in (0, L(n))$ and thus $W(x) \in (0, 1)$. Similarly, w has the properties requested in item (b), and, clearly, the inequalities (109) hold.

Finally, for each $x \in \mathbb{R}^n$, and each $k \in \mathbb{N}$,

$$\sum_{n=k}^{\infty} \frac{1}{2^n c_n} \left\| \nabla^{(k)} W_n(x) \right\| \leq \sum_{n=k}^{\infty} \frac{1}{2^n c_n} a_n < \sum_{n=k}^{\infty} \frac{1}{2^n},$$

so the series

$$\sum_{n=1}^{\infty} \frac{1}{2^n c_n} \nabla^{(k)} W_n(x)$$

is uniformly convergent. Consequently, W is C^∞. □

With the building blocks prepared in Proposition 10.1, we build the Morse-Lyapunov function promised in Theorem 7.3.

Proof.(of Theorem 7.3) Let $A = A_L = \mathcal{A}$, let W be the function from item (a) of Proposition 10.1, and set $V_L = W$ and $w_L := (1 - 1/e)W$. For $l = 1, 2, \ldots, L-1$, consider $A = A_l$, let W and w come from item (b) of Proposition 10.1, and and let $V_l = W$ and $w_l = w$. Let

$$V(x) := \sum_{l=1}^{L} V_l(x), \quad w(x) := \sum_{l=1}^{L} w_l(x).$$

The Lyapunov inequalities in Proposition 10.1 add up to the Lyapunov inequalities in the theorem.

Take $x \in \mathcal{M}(\mathcal{D})$. Then $x \in M_l = A_l \cap A^*_{l-1}$ for some $l \in \{1, 2, \ldots, L\}$. Because $x \in A_l$, one has $x \in A_k$ for $k \in \{l, l+1, \ldots, L\}$, so that $V_k(x) = 0$ for these k. Because $x \in A^*_{l-1}$, one has $x \in A^*_k$ for $k \in \{1, 2, \ldots, l-1\}$, so that $V_k(x) = 1$ for these k. Then $V(x) = l - 1$.

If $x \in \mathcal{M}(\mathcal{D})$ then, by (66), each $w_l(x) = 0$, so $w(x) = 0$. Take $x \notin \mathcal{M}(\mathcal{D})$. If $x \notin \mathcal{A}$, then $w_L(x) > 0$. If $x \in \mathcal{A}$, then by (67), there exists $l \in \{1, 2, \ldots, L\}$ with $x \notin A_l \cup A^*_l$, so that $w_l(x) > 0$. Either way, $w(x) > 0$, so that w is positive away from $\mathcal{M}(\mathcal{D})$. $\square$

11. Acknowledgments

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.

References

[1] H. Robbins, S. Monro, A stochastic approximation method, The Annals of Mathematical Statistics (1951) 400–407.

[2] J. Kiefer, J. Wolfowitz, Stochastic estimation of the maximum of a regression function, The Annals of Mathematical Statistics (1952) 462–466.

[3] J. R. Blum, Multidimensional stochastic approximation methods, The Annals of Mathematical Statistics (1954) 737–744.

[4] E. G. Gladyšev, On stochastic approximation, Teor. Verojatnost. i Primenen. 10 (1965) 297–300, (English translation: [Theory of Probability and Its Applications 10:2 (1965), 275–278]).

[5] H. Robbins, D. Siegmund, A convergence theorem for nonnegative almost supermartingales and some applications, in: J. S. Rustagi (Ed.), Optimizing Methods in Statistics, Academic Press, 1971, pp. 233–257.

[6] L. Ljung, Analysis of recursive stochastic algorithms, IEEE Transactions on Automatic Control 22 (4) (1977) 551–575.

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

[8] H. J. Kushner, G. Yin, Stochastic Approximation and Recursive Algorithms and Applications, Springer-Verlag, 2003.

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

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

[11] V. Borkar, Stochastic Approximation: A Dynamical Systems Viewpoint, 2nd Edition, Springer and Hindustan Book Agency, 2023.

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

[13] R. Goebel, R. G. Sanfelice, A. R. Teel, Hybrid dynamical systems. robust stability and control for systems that combine continuous-time and discrete-time dynamics, IEEE Control Systems Magazine 29 (2009) 28–93.

[14] R. Goebel, R. G. Sanfelice, A. R. Teel, Hybrid Dynamical Systems, Princeton University Press, 2012.

[15] R. G. Sanfelice, Hybrid Feedback Control, Princeton University Press, 2021.

[16] R. G. Sanfelice, A. R. Teel, Dynamical properties of hybrid systems simulators, Automatica 46 (2) (2010) 239–248. doi:https://doi.org/10.1016/j.automatica.2009.09.026.

[17] A. R. Teel, R. Goebel, Stochastic approximations of hybrid systems with continuous flow maps, IFAC-PapersOnLine 58 (11) (2024) 141–146, 8th IFAC Conference on Analysis and Design of Hybrid Systems ADHS 2024. doi:https://doi.org/10.1016/j.ifacol.2024.07.438.

[18] A. R. Teel, R. Goebel, R. G. Sanfelice, M. F. Crisafulli, Stochastic approximation results for hybrid inclusions, in: Proceedings of the 63rd IEEE Conference on Decision and Control, 2024.

[19] R. Goebel, A. Teel, A smooth Conley-Lyapunov function for hybrid inclusions on $\mathbb{R}^n$, Systems & Control Letters 204 (106204) (October 2025).

[20] A. R. Teel, R. Goebel, R. G. Sanfelice, Stochastic approximations of differential inclusions: almost sure boundedness and asymptotic convergence, in: IFAC Symposium on Nonlinear Control Systems (NOLCOS), 2025.

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

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

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

[24] R. Rockafellar, R. J.-B. Wets, Variational Analysis, Springer, 1998.

[25] R. Goebel, Set-valued, Convex, and Nonsmooth Analysis in Dynamics and Control. An Introduction., SIAM, 2024.

[26] R. Sanfelice, R. Goebel, A. Teel, Generalized solutions to hybrid dynamical systems, ESAIM Control Optim. Calc. Var. 14 (2008) 699–724.

[27] A. Filippov, Diffrential Equations with Discontinuous Righthand Sides, Kluwer, 1988.

[28] N. Krasovskii, Game-Theoretic Problems of Capture, Nauka, Moscow, 1970, in Russian.

[29] W. Haddad, V. Chellaboina, S. Nersesov, Impulsive and Hybrid Dynamical Systems: Stability, Dissipativity, and Control, Princeton Univ. Press, 2006.

[30] D. Liberzon, Switching in Systems and Control, Systems and Control: Foundations and Applications, Birkhauser, 2003.

[31] R. G. Sanfelice, R. Goebel, A. R. Teel, Invariance principles for hybrid systems with connections to detectability and asymptotic stability, IEEE Transactions on Automatic Control 52 (12) (2007) 2282–2297. doi:10.1109/TAC.2007.910684.

[32] K. Johansson, M. Egerstedt, J. Lygeros, S. Sastry, On the regularization of Zeno hybrid automata, Systems & Control Letters 38 (1999) 141–150.

[33] Y. Or, A. R. Teel, Zeno stability of the set-valued bouncing ball, IEEE Trans. Automat. Control 56 (2) (2011) 447–452. doi:10.1109/TAC.2010.2090411.
URL https://doi.org/10.1109/TAC.2010.2090411

[34] S. Dashkovskiy, P. Feketa, Asymptotic properties of Zeno solutions, Nonlinear Anal. Hybrid Syst. 30 (2018) 256–265.

[35] K. Schindler, R. I. Leine, Paradoxical simulation results of chaos-like chattering in the bouncing ball system, Phys. D 419 (2021) Paper No. 132854, 10. doi:10.1016/j.physd.2021.132854.
URL https://doi.org/10.1016/j.physd.2021.132854

[36] C. Cai, A. R. Teel, R. Goebel, Smooth Lyapunov functions for hybrid systems. Part II: (Pre-)asymptotically stable compact sets, IEEE Trans. Automat. Control 53 (3) (2008) 734–748.

[37] B. Fristedt, L. Gray, A modern approach to probability theory, Birkhauser, 1997.

[38] J. I. Poveda, A. R. Teel, A framework for a class of hybrid extremum seeking controllers with dynamic inclusions, Automatica 76 (2017) 113–126. doi:https://doi.org/10.1016/j.automatica.2016.10.029.

[39] J. L. Doob, Stochastic Processes, New York: Wiley, 1953.

[40] J. Doob, Measure Theory, Springer, 1994.

[41] L. Bottou, F. E. Curtis, J. Nocedal, Optimization methods for large-scale machine learning, SIAM Review 60 (2) (2018) 223–311. doi:10.1137/16M1080173.

[42] I. Molchanov, Theory of Random Sets, Springer, Berlin, 2005.

[43] V. G. Yaji, S. Bhatnagar, Analysis of stochastic approximation schemes with set-valued maps in the absence of a stability guarantee and their stabilization, IEEE Transactions on Automatic Control 65 (3) (2020) 1100–1115. doi:10.1109/TAC.2019.2916688.

[44] C. Conley, Isolated invariant sets and the Morse index, Vol. 38 of CBMS Regional Conference Series in Mathematics, Amer. Math. Society, Providence, R.I., 1978.

[45] J. Ayala, P. Corbin, K. Mc Conville, F. Colonius, W. Kliemann, J. Peters, Morse decomposition, attractors and chain recurrence, Proyecciones 25 (1) (2006) 79–110. doi:10.4067/S0716-09172006000100006. URL https://doi.org/10.4067/S0716-09172006000100006

[46] E. Akin, The general topology of dynamical systems, American Mathematical Soc., 1993.

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

[48] R. Goebel, A. R. Teel, Solutions to hybrid inclusions via set and graphical convergence with stability theory applications, Automatica 42 (2006) 573–587.

[49] P. Forni, D. Angeli, Smooth Lyapunov functions for multistable hybrid systems on manifolds, in: 2017 IEEE 56th Annual Conference on Decision and Control (CDC), 2017, pp. 5481–5486.

[50] D. E. Norton, The fundamental theorem of dynamical systems, Comment. Math. Univ. Carolin. 36 (3) (1995) 585–597.

[51] F. Clarke, Optimization and Nonsmooth Analysis, Wiley, 1983.

[52] M. Valadier, Lignes de descente de fonctions lipschitziennes non pathologiques, in: Seminaire d'Analyse Convexe, Vol. 18 (Montpellier, 1988), Univ. Sci. Tech. Languedoc, Montpellier, 1988, pp. Exp. No. 9, 10.

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

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

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

[56] M. Hurley, Chain recurrence, semiflows, and gradients, J. Dynam. Differential Equations 7 (3) (1995) 437–456.

[57] A. Daniilidis, D. Drusvyatskiy, Pathological subgradient dynamics, SIAM J. Optim. 30 (2) (2020) 1327–1338.

[58] D. Burkholder, Martingale transforms, The Annals of Mathematical Statistics 37 (6) (1966) 1494–1504.

[59] R. G. Sanfelice, D. A. Copp, P. Nanez, A toolbox for simulation of hybrid systems in Matlab/Simulink: Hybrid Equations (HyEQ) Toolbox, in: Proceedings of Hybrid Systems: Computation and Control Conference, 2013, p. 101–106.
URL `https://hybrid.soe.ucsc.edu/files/preprints/74.pdf`