Stochastic approximation results for hybrid inclusions

Andrew R. Teel, Rafal K. Goebel, Ricardo G. Sanfelice and Max F. Crisafulli

***Abstract*— A stochastic simulator that approximates the behavior of a hybrid system given in terms of a hybrid inclusion is considered. A hybrid inclusion combines constrained differential and difference inclusions to model the continuous (flows) and discrete (jumps) dynamics, respectively. The simulator is a stochastic discrete-time system that employs a set-valued right-hand side when approximating flows. Under mild conditions on the data defining the simulator and the hybrid system, together with a non-uniform averaging condition, it is shown that almost every sample path of each solution generated by the stochastic simulator is close to a solution of the original hybrid system on compact time domains when the step size sequence is sufficiently small and converges to zero but is not summable. A probabilistic characterization is also provided. An example is provided to illustrate the interest of the proposed framework.**

I. Introduction

The stochastic approximation idea introduced by Robbins and Monro in [1] provides a method to estimate the root of a function. The idea is to iteratively update an estimate based on noisy observations of the function. The Robbins-Monro algorithm provides a framework for these updates, wherein the step size diminishes over time to ensure convergence to the root of the function. This idea has been applied to optimization problems [2], [3], [4], [5], and initial value problems for differential equations [6], [7], [8], [9], [10] and differential inclusions [7], [11], [12], [13], [14].

In this paper, an algorithm for approximating solutions to a hybrid dynamical system is studied. Hybrid systems have state variables that evolve continuously or discretely, based on the values of the state [15], [16], [17], [18], [19], [20]. Stochastic approximation of a hybrid system cast as a hybrid automaton has been considered in [21] and [22]. In this work, we model hybrid systems as hybrid inclusions [20]. A hybrid inclusion with state $x \in \mathbb{R}^n$ is represented by

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

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

This representation suggests that when in the flow set $C \subset \mathbb{R}^n$, solutions to (1) may "flow" according to the differential inclusion $\dot{x} \in F(x)$, where F is the flow map. When in the jump set $D \subset \mathbb{R}^n$, solutions may "jump" according to the difference inclusion $x^+ \in G(x)$, where G is the jump map.

Building on [23], which pertains to hybrid inclusions with a Lipschitz continuous flow map, this paper proposes a stochastic approximation result for the general case of a set-valued flow map F. The stochastic approximation is given by a discrete-time system that employs the data $(C, \widehat{F}, D, G)$ where $\widehat{F}$ involves a random process and approximates F through its average or expected value. In [23], convergence to the average is assumed to be uniform in the starting time; that assumption is dropped here. When the step sizes are sufficiently small, the proposed algorithm generates solutions with sample paths that are close to solutions of (1) on compact time domains. These properties are established under mild conditions on the data defining the hybrid system.

Our results are comparable to results for differential inclusions in [11], [12], [13], and [14], but we focus on finite (hybrid) time horizons. The previous results on stochastic approximation of hybrid systems in [21], [22] address stochastically approximating jump conditions, with a focus on avoiding Zeno solutions, rather than stochastically approximating the flow map. Part of our interest in stochastic approximations of F comes from envisioning hybrid extremum seeking algorithms that use stochasticity to approximate the generalized gradient of a nonsmooth readout map of a dynamical system. See the example in Section IV.

Notation: $\mathbb{R}_{\geq 0}$ ($\mathbb{Z}_{\geq 0}$) is the nonnegative real numbers (integers). For $\ell \in \mathbb{Z}_{\geq 0}$, $\mathbb{Z}_{\geq \ell}$ is the integers that are greater than or equal to ℓ. $\mathbb{B} \subset \mathbb{R}^n$ is the closed unit ball centered at the origin; given $\rho > 0$, $\rho\mathbb{B}$ is the closed ball of radius ρ centered at the origin. $\mathbb{B}^\circ$ denotes the open unit ball. $\lceil \cdot \rceil$ is the ceiling function. Given a set $S \subset \mathbb{R}^n$ and $\rho > 0$, $S + \rho\mathbb{B}$ denotes a ball of radius ρ around the set S, i.e., $S + \rho\mathbb{B} := \{z \in \mathbb{R}^n : z = x + v, x \in S, v \in \rho\mathbb{B}\}$. We adopt the set-valued terminology and notation of [24].

II. Stochastic Approximation of a Hybrid Inclusion: Setting and Assumptions

We consider the stochastic hybrid system simulator

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

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

for (1), where $x \in \mathbb{R}^n$ is the state and σ^+ is a placeholder for a deterministic sequence $\{\sigma_k\}_{k=1}^\infty$ of small positive numbers, i.e., the "step sizes", and y^+ is a placeholder for a random process $\{\mathbf{y}_k\}_{k=1}^\infty$ that is defined on a probability space $(\Omega, \mathcal{F}, \mathbb{P})$ and $\mathbf{y}_k : \Omega \to Y \subset \mathbb{R}^m$ for each $k \in \mathbb{Z}_{\geq 1}$, where Y is closed. We let $\{\mathcal{F}_j\}_{j=0}^\infty$ be the natural filtration of this

Andrew R. Teel and Max F. Crisafulli are with the ECE Department, University of California, Santa Barbara, CA 93106-9560, USA. (teel@ucsb.edu, maxcrisafulli@ucsb.edu). Andrew R. Teel's research supported in part by AFOSR FA9550-21-1-0452. Rafal K. Goebel is with the Department of Mathematics and Statistics, Loyola University Chicago, 1032 W. Sheridan Rd., Chicago, IL, 60660, USA. (rgoebel1@luc.edu). Ricardo G. Sanfelice is with the Electrical and Computer Engineering Department, University of California, Santa Cruz, CA 95064. (ricardo@ucsc.edu). Ricardo G. Sanfelice's research supported by NSF Grants no. CNS-2039054 and CNS-2111688, by AFOSR Grants nos. FA9550-19-1-0169, FA9550-20-1-0238, FA9550-23-1-0145, and FA9550-23-1-0313, 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.

random process; see, for example, [25, Section 11.3]. As in [23], for simplicity, the data $(C, \widehat{F}, D, G)$ does not depend on $\{\sigma_k\}_{k=1}^{\infty}$.

Simulator solutions in [26] are discrete-time arcs defined on discrete time domains in $\mathbb{R}^2$ satisfying the constraints implied by (2). We adapt those notions here. What follows also generalizes the definitions in [23] to the case of time-varying step sizes. Given $\{\sigma_k\}_{k=1}^{\infty}$, let

$$\sigma_0 := 0, \qquad \widetilde{\sigma}_k := \sum_{i=0}^{k} \sigma_i \quad \forall k \in \mathbb{Z}_{\geq 0}. \tag{3}$$

A *hybrid compact discrete time domain with step sizes* $\{\sigma_k\}_{k=1}^{\infty}$ is a set $E \subset \mathbb{R}_{\geq 0} \times \mathbb{Z}_{\geq 0}$ such that

$$E = \bigcup_{j=0}^{J} \bigcup_{k=K_j}^{K_{j+1}} (\widetilde{\sigma}_k, j) \tag{4}$$

for some $J \in \mathbb{Z}_{\geq 0}$ and finite sequence of integers $0 = K_0 \leq K_1 \leq \cdots \leq K_{J+1}$. It is a *hybrid discrete time domain with step sizes* $\{\sigma_k\}_{k=1}^{\infty}$ if it is the union of a nondecreasing sequence of compact discrete time domains with step sizes $\{\sigma_k\}_{k=1}^{\infty}$. A *hybrid discrete arc with step sizes* $\{\sigma_k\}_{k=1}^{\infty}$ is a function $x : \operatorname{dom} x \to \mathbb{R}^n$ such that $\operatorname{dom} x$ is a hybrid discrete time domain with step sizes $\{\sigma_k\}_{k=1}^{\infty}$. Let $\mathcal{X}$ be the set of set-valued mappings from $\mathbb{R}^2$ to $\mathbb{R}^n$ with a closed and nonempty graph. (Such mappings are outer semicontinuous [24, Theorem 5.7].) Given $\{\mathbf{y}_k\}_{k=1}^{\infty}$, a mapping $\mathbf{x} : \Omega \to \mathcal{X}$ is a *candidate solution of the simulator* (2) with step sizes $\{\sigma_k\}_{k=1}^{\infty}$ if, for all $\omega \in \Omega$, the *sample path* $\mathbf{x}(\omega)$ satisfies the following conditions:

1) $\mathbf{x}(\omega)$ is a hybrid discrete arc with step sizes $\{\sigma_k\}_{k=1}^{\infty}$;
2) $(\widetilde{\sigma}_k, j), (\widetilde{\sigma}_{k+1}, j) \in \operatorname{dom} \mathbf{x}(\omega)$ implies

$$\mathbf{x}(\omega)(\widetilde{\sigma}_k, j) \in C, \quad \& \tag{5a}$$

$$\mathbf{x}(\omega)(\widetilde{\sigma}_{k+1}, j) - \mathbf{x}(\omega)(\widetilde{\sigma}_k, j) \in \sigma_{k+1} \widehat{F}(\mathbf{x}(\omega)(\widetilde{\sigma}_k, j), \mathbf{y}_{k+1}(\omega)). \tag{5b}$$

3) $(\widetilde{\sigma}_k, j), (\widetilde{\sigma}_k, j+1) \in \operatorname{dom} \mathbf{x}(\omega)$ implies

$$\mathbf{x}(\omega)(\widetilde{\sigma}_k, j) \in D, \quad \& \tag{6a}$$

$$\mathbf{x}(\omega)(\widetilde{\sigma}_k, j+1) \in G(\mathbf{x}(\omega)(\widetilde{\sigma}_k, j)). \tag{6b}$$

A candidate solution of (2) is a *solution of* (2) if it is adapted to the filtration $\{\mathcal{F}_k\}_{k=0}^{\infty}$: for each $k \in \mathbb{Z}_{\geq 0}$,

$$\omega \mapsto \operatorname{graph}(\mathbf{x}(\omega)) \cap ([0, \widetilde{\sigma}_k] \times \mathbb{Z}_{\geq 0} \times \mathbb{R}^n) \tag{7}$$

is $\mathcal{F}_k$-measurable. The collection of such solutions is denoted $\mathcal{S}_\sigma$ and we write $\mathbf{x} \in \mathcal{S}_\sigma$ for such a solution. We study the behavior of the sample paths for small but not summable step sizes $\{\sigma_k\}_{k=1}^{\infty}$. For example, we are interested in conditions for the step sizes $\{\sigma_k\}_{k=1}^{\infty}$ under which almost every sample path of a solution of the simulator (2) is close to a solution of the hybrid system (1) when the sample path is restricted to a compact time domain. We use $\mathcal{S}$ to denote the set of maximal solutions to (1). We assume the following for the data (C, F, D, G) in (1):

Assumption 1: The sets C and D are compact, the mapping $F : \mathbb{R}^n \rightrightarrows \mathbb{R}^n$ is outer semicontinuous and locally bounded with nonempty convex values on C, and the mapping $G : \mathbb{R}^n \rightrightarrows \mathbb{R}^n$ is outer semicontinuous and locally bounded with nonempty values on D. ■

Remark 1: We assume compactness of C and D for simplicity and to convey the main ideas. It is simpler, but tedious, to assume instead that the continuous-time system $x \in C$, $\dot{x} \in F(x)$ has no finite time blow up and that initial conditions start in a given compact set. ■

Assumption 2: The mapping $\widehat{F} : \mathbb{R}^n \times \mathbb{R}^m \rightrightarrows \mathbb{R}^n$ satisfies the following properties:

1) its values are nonempty and closed on $C \times Y$;
2) the graphical mapping $y \mapsto \operatorname{graph}(\widehat{F}(\cdot, y)) \subset \mathbb{R}^n \times \mathbb{R}^n$ is measurable with closed values;
3) it is uniformly bounded on $C \times Y$, i.e., $\beta := \sup_{(x,y) \in C \times Y} |\widehat{F}(x, y)| < \infty$. ■

Remark 2: The last item can be relaxed to almost sure boundedness, i.e., for almost every $\omega \in \Omega$, there exists β such that $|\widehat{F}(x, \mathbf{y}_{k+1}(\omega))| \leq \beta$ for all $(x, k) \in C \times \mathbb{Z}_{\geq 0}$. ■

We now impose a condition that yields closeness of sample paths of (2) to solutions of (1). To save on notation, for each $(x, N, i, \ell, \omega) \in C \times \mathbb{Z}_{\geq 1} \times \mathbb{Z}_{\geq 1} \times \mathbb{Z}_{\geq 1} \times \Omega$, define

$$S^N_{x,i,\ell,\omega} := \frac{\displaystyle\sum_{k=\ell}^{\ell+N-1} \sigma_{k+1} \widehat{F}\left(\{x\} + i^{-1}\mathbb{B}, \mathbf{y}_{k+1}(\omega)\right)}{\displaystyle\sum_{k=\ell}^{\ell+N-1} \sigma_{k+1}}. \tag{8}$$

Note that $S^N_{x,i,\ell,\omega}$ depends on the sequence of step sizes.

Assumption 3: For every sequence of deterministic, nonsummable step sizes $\{\sigma_k\}_{k=1}^{\infty}$ and almost every $\omega \in \Omega$, the following condition holds:

(C1) For each $\varepsilon > 0$ there exists $\delta > 0$ and for each $\ell \in \mathbb{Z}_{\geq 0}$ there exists $N^* \in \mathbb{Z}_{\geq 1}$ such that,

$$S^N_{x,\lceil 1/\delta \rceil,\ell,\omega} \subset \overline{\operatorname{co}} F(x + \varepsilon\mathbb{B}) + \varepsilon\mathbb{B}. \tag{9}$$

for each $x \in C$ and $N \in \mathbb{Z}_{\geq N^*}$. ■

Remark 3: In contrast to [23, Assumption 3], we do not assume that the averaging condition is uniform in the starting index ℓ of the summation in (8) nor do we assume that the step sizes are constant. ■

The following is a sufficient condition for Assumption 3.

Proposition 1: Under Assumptions 1-2, Condition (C1) in Assumption 3 holds if, for every deterministic, nonsummable sequence of step sizes $\{\sigma_k\}_{k=1}^{\infty}$ and almost every $\omega \in \Omega$,

$$\limsup_{i \to \infty} \left(\bigcup_{\ell \in \mathbb{Z}_{\geq 0}} \limsup_{N \to \infty} S^N_{x,i,\ell,\omega} \right) \subset F(x) \quad \forall x \in C. \tag{10}$$

III. Main results

A. Generating nonsummable bounds on step sizes

Given $\omega \in \Omega$ such that Condition (C1) of Assumption 3 holds and given a seed $\rho^* > 0$ and an integer $J > 0$, we construct a nonsummable sequence of step sizes $\{\sigma_k^*\}_{k=1}^\infty$. Recall the definition of β in Assumption 2.3. Let Assumption 3 with $\varepsilon = \rho^*/2$ generate $\delta \in (0, \rho^*]$ and $N^*(\ell)$ for each $\ell \in \mathbb{Z}_{\geq 0}$. For each $p \in \mathbb{Z}_{\geq 0}$, define

$$\ell_p := \sum_{j=0}^{p-1} N^*(\ell_j); \tag{11}$$

in particular, $\ell_0 = 0$. For $p \in \mathbb{Z}_{\geq 0}$ and

$$k \in \{\ell_p, \ldots, \ell_{p+1} - 1\} =: \Upsilon_p \tag{12}$$

define

$$\sigma_{k+1}^* := \frac{\delta}{2N^*(\ell_p)\max\{\beta, J+1\}}. \tag{13}$$

Note that, for all $p \in \mathbb{Z}_{\geq 0}$,

$$\sum_{k \in \Upsilon_p} \sigma_{k+1}^* = \frac{\delta}{2\max\{\beta, J+1\}}. \tag{14}$$

Hence, $\{\sigma_k^*\}_{k=1}^\infty$ is not summable.

The family of all nonsummable sequences of step sizes $\{\sigma_k^*\}_{k=1}^\infty$ generated by $\rho^* > 0$ and J as above is denoted $\mathcal{TS}(\rho^*, J)$. For any sequence of positive real numbers $\{\sigma_k^*\}_{k=1}^\infty$, $\mathcal{TS}_{\leq \{\sigma_k^*\}_{k=1}^\infty}$ is the family of all nonsummable sequences of step sizes $\{\sigma_k\}_{k=1}^\infty$ that are "pointwise bounded" by $\{\sigma_k^*\}_{k=1}^\infty$, i.e., for every $k \in \mathbb{Z}_{\geq 1}$, $0 < \sigma_k \leq \sigma_k^*$.

B. Almost sure closeness on compact time domains

Our first result extends [23, Theorem 1]. It establishes closeness of each sample path of each simulator solution to a solution of the hybrid system (1) when truncating the domain of the sample path to a compact domain. Given a hybrid or hybrid discrete arc ϕ, for each $T > 0$ and $J \in \mathbb{Z}_{\geq 0}$,

$$\text{graph}_{\preceq(T,J)}(\phi) := \text{graph}\,(\phi) \cap \Big([0,T] \times \{0,1,\ldots,J\} \times \mathbb{R}^n \Big).$$

Theorem 1: If Assumptions 1-3 hold then, for almost every $\omega \in \Omega$, the following property holds: for each $T > 0$, $J \in \mathbb{Z}_{\geq 0}$, and $\varepsilon > 0$ there exists $\rho^* > 0$ such that, with $\{\sigma_k^*\}_{k=1}^\infty \in \mathcal{TS}(\rho^*, J)$, for each $\{\sigma_k\}_{k=1}^\infty \in \mathcal{TS}_{\leq \{\sigma_k^*\}_{k=1}^\infty}$ and for every $\mathbf{x} \in \mathcal{S}_\sigma$, there exists $\phi \in \mathcal{S}$ so that

$$\text{graph}_{\preceq(T,J)}\left(\mathbf{x}(\omega)\right) \subset \text{graph}(\phi) + \varepsilon\mathbb{B}. \tag{15}$$

Proof. Let $\Omega_1 \subset \Omega$ be such that $\mathbb{P}(\Omega_1) = 1$ and the condition in Assumption 3 holds. Fix $\omega \in \Omega_1$, $T > 0$, $J \in \mathbb{Z}_{\geq 0}$ and $\varepsilon > 0$. Pick $\rho \in (0, \varepsilon/2]$ so that, with the definitions

$$C_\rho := C + \rho\mathbb{B} \tag{16a}$$
$$F_\rho(x) := \overline{\text{co}}\,(F(x + \rho\mathbb{B})) + \rho\mathbb{B} \tag{16b}$$
$$D_\rho := D + \rho\mathbb{B} \tag{16c}$$
$$G_\rho(x) := G(x + \rho\mathbb{B}) + \rho\mathbb{B}\ , \tag{16d}$$

for each solution ψ of

$$x \in C_\rho \qquad \dot{x} \in F_\rho(x) \tag{17a}$$
$$x \in D_\rho \qquad x^+ \in G_\rho(x) \tag{17b}$$

there exists, by Assumption 1 and [20, Prop. 6.34], a solution ϕ of (1) i.e., $\phi \in \mathcal{S}$, such that

$$\text{graph}_{\preceq(T,J)}(\psi) \subset \text{graph}(\phi) + \frac{\varepsilon}{2}\mathbb{B}. \tag{18}$$

Let $\beta \geq 1$ come from Assumption 2.3. Let $\rho^* := \rho$. Let $\{\sigma_k^*\}_{k=1}^\infty \in \mathcal{TS}(\rho^*, J)$ and let $\{\sigma_k\}_{k=1}^\infty \in \mathcal{TS}_{\leq \{\sigma_k^*\}_{k=1}^\infty}$. Note that the process of constructing $\{\sigma_k^*\}_{i=1}^\infty \in \mathcal{TS}(\rho^*, J)$ generates $\delta \in (0, \rho^*]$ and $N^*(\ell)$ so that the condition (C1) in Assumption 3 holds. Let τ be the right-hand side of (14):

$$\tau := \frac{\delta}{2\max\{\beta, J+1\}}. \tag{19}$$

With this notation, thanks to (14), and with Υ_p as in (12),

$$\sum_{k \in \Upsilon_p} \sigma_{k+1} \leq \tau \qquad \forall p \in \mathbb{Z}_{\geq 0}. \tag{20}$$

Let $\mathbf{x} \in \mathcal{S}_\sigma$ so that, in particular, (5)-(6) hold for the given ω. For ease of notation, let χ be the truncation of $\mathbf{x}(\omega)$ to the domain $\text{dom}\,\mathbf{x}(\omega) \cap ([0,T] \times \{0,\ldots,J\})$. Using (3), let

$$\text{dom}\,\chi = \bigcup_{j=0}^{J'} \bigcup_{k=K_j}^{K_{j+1}} (\widetilde{\sigma}_k, j), \tag{21}$$

which is similar to (4). Note that $J' \in \{0, \ldots, J\}$.

Now we claim that there is a solution ψ of (17) such that

a) $\text{dom}\,\psi = \bigcup_{j=0}^{J'} ([t_j, t_{j+1}] \times \{j\})$ where $0 = t_0 \leq t_1 \leq \cdots \leq t_{J'+1}$;
b) $[\widetilde{\sigma}_{K_j}, \widetilde{\sigma}_{K_{j+1}}] \subset [t_j, t_{j+1}] + (j+1)\tau\mathbb{B} \subset [t_j, t_{j+1}] + \frac{\delta}{2}\mathbb{B}$;
c) $\text{graph}(\chi) \subset \text{graph}(\psi) + \frac{\varepsilon}{2}\mathbb{B}$.

The overall idea is this: Recall the definition of Υ_p in (12). For each $j \in \{0, \ldots, J'\}$, the Υ_p that fit in the $(j+1)$th horizontal segment of $\text{dom}\,\chi$, i.e., the Υ_p with $p \in \Phi_j$, where

$$\Phi_j := \{p \in \mathbb{Z}_{\geq 0} : \Upsilon_p \subset \{K_j, \ldots, K_{j+1} - 1\}\}, \tag{22}$$

give rise to a solution to the continuous-time dynamics in (17) defined on the horizontal segments of $\text{dom}\,\psi$. For this, Assumption 3 is essential. Each Υ_p that doesn't fit in any of the horizontal segments of $\text{dom}\,\chi$ parameterizes perhaps some "flow" of χ described by (5) and definitely some jumps of χ, i.e., the evolution of χ described by (6). Each such Υ_p gives rise to a solution of the discrete-time dynamics in (17); the potential "flow" of χ captured by Υ_p is essentially ignored. Putting these solutions to (17) together results in ψ.

For convenience, let

$$q_j := \text{card}(\Phi_j) \tag{23}$$

and order the set Φ_j as $(p_{1,j}, \ldots, p_{q_j,j})$. The promised time intervals defining $\text{dom}\,\psi$ are then defined as follows:

$$t_0 := 0 \tag{24a}$$
$$\tau_{i,j} := t_j + \sum_{k \in \bigcup_{s=1}^i \Upsilon_{p_{s,j}}} \sigma_{k+1} \qquad \forall i \in \{0, \ldots, q_j\} \tag{24b}$$
$$t_{j+1} := \tau_{q_j,j}. \tag{24c}$$

We note that, for each $i \in \{0, \dots, q_j - 1\}$,

$$\tau_{i+1,j} - \tau_{i,j} = \sum_{k \in \Upsilon_{p_{i,j}}} \sigma_{k+1}. \tag{25}$$

Also, applying (20) iteratively for $s \in \{0, \dots, j\}$, the definitions in (24) ensure that b) above is satisfied.

We now focus on evolution on the "horizontal segment" in $\operatorname{dom}\chi$ indexed by j, i.e., the $(j+1)$th "horizontal segment". More precisely, we focus on the part of that horizontal segment that is paratemerized by the index sets Υ_p that fit in it, i.e., by the Υ_p with $p \in \Phi_j$. For $i \in \{0, \dots, q_j\}$, define

$$\xi_i := \chi(\tau_{i,j}, j), \tag{26}$$

and label the intermediate values of χ in each Υ_p contained in the j-th horizontal piece of $\operatorname{dom}\chi$ as follows:

$$\xi_{k,i,j} := \chi(\widetilde{\sigma}_k, j) \qquad \forall k \in \Upsilon_{p_{i,j}}. \tag{27}$$

Then, for all $i \in \{0, \dots, q_j - 1\}$,

$$\xi_{i+1} - \xi_i \in \sum_{k \in \Upsilon_{p_{i,j}}} \sigma_{k+1} \widehat{F}(\xi_{k,i,j}, \mathbf{y}_{k+1}(\omega)). \tag{28}$$

Moreover, for each $k \in \Upsilon_{p_{i,j}}$,

$$\xi_{k,i,j} - \xi_i \in \sum_{s \in \Upsilon_{p_{i,j}} \cap \{0,\dots,k-1\}} \sigma_{s+1} \widehat{F}(\xi_{s,i,j}, \mathbf{y}_{k+1}(\omega)) \tag{29}$$

so that, using Assumption 2.3, (19)-(20) and the positivity of the step sizes,

$$|\xi_{k,i,j} - \xi_i| \le \tau\beta \le \frac{\delta}{2} \le \frac{\rho}{2} \le \frac{\varepsilon}{4} \qquad \forall k \in \Upsilon_{p_{i,j}}. \tag{30}$$

It follows from (28) and (30) that

$$\xi_{i+1} - \xi_i \in \sum_{k \in \Upsilon_{p_{i,j}}} \sigma_{k+1} \widehat{F}\left(\xi_i + \frac{\delta}{2}\mathbb{B}, \mathbf{y}_{k+1}(\omega)\right). \tag{31}$$

Moreover, by the definition of a solution,

$$\xi_i \in C \qquad \forall i \in \{0, \dots, q_j - 1\} \tag{32}$$

while, by (31) with $i = q_j - 1$, Assumption 2.3, (19), (20),

$$\xi_{q_j} \in C_{\frac{\delta}{2}} \subset C_{\frac{\rho}{2}}. \tag{33}$$

Next, using Assumption 3, and the construction of δ, we have for all $i \in \{0, \dots, q_j - 1\}$,

$$\xi_{i+1} - \xi_i \in \sum_{k \in \Upsilon_{p_{i,j}}} \sigma_{k+1} \left(\overline{\text{co}} F\left(\xi_i + \frac{\rho}{2}\mathbb{B}\right) + \frac{\rho}{2}\mathbb{B}\right). \tag{34}$$

Define $\zeta : [t_j, t_{j+1}] \times \{j\} \to \mathbb{R}^n$ by linearly interpolating the values of $\{\xi_i\}_{i=0}^{q_j}$: for $t \in [\tau_{ij}, \tau_{i+1,j})$, we have

$$\zeta(t, j) := \frac{\tau_{i+1,j} - t}{\tau_{i+1,j} - \tau_{i,j}} \xi_i + \frac{t - \tau_{i,j}}{\tau_{i+1,j} - \tau_{i,j}} \xi_{i+1}. \tag{35}$$

Then, for all $i \in \{0, \dots, q_j - 1\}$ and all $t \in [t_{ij}, t_{i+1,j})$,

$$|t - \tau_{ij}| \le \tau \le \frac{\delta}{2} \le \frac{\rho}{2} \le \frac{\varepsilon}{4} \tag{36a}$$

$$|\zeta(t, j) - \xi_i| \le |\xi_{i+1} - \xi_i| \le \tau\beta \le \frac{\delta}{2} \le \frac{\rho}{2} \le \frac{\varepsilon}{4}. \tag{36b}$$

The inequalities in (30) and (36) give

$$\text{graph}(\chi(\cdot, j)) \cap ([t_j, t_{j+1}] \times \mathbb{R}^n) \subset \text{graph}(\zeta(\cdot, j)) + \frac{\varepsilon}{2}\mathbb{B}. \tag{37}$$

From (34)-(35) together with (25), we also have

$$\begin{aligned} \dot{\zeta}(t, j) &\in \overline{\text{co}} F\left(\xi_i + \frac{\rho}{2}\mathbb{B}\right) + \frac{\rho}{2}\mathbb{B} \\ &\subset \overline{\text{co}} F\left(\zeta(t, j) + \rho\mathbb{B}\right) + \rho\mathbb{B}. \end{aligned} \tag{38}$$

Now, we focus on the evolution of χ indexed by those Υ_p that don't fit in any of the horizontal segments of $\operatorname{dom}\chi$. For $j \le J'$ and with $\Upsilon_{p_{q_j,j}}$ as above, let

$$\Upsilon_{p_{q_j,j}+1} =: \{s_0, s_0 + 1, \dots, S - 1\}.$$

In particular, $s_0 = m_j + 1$, where m_j is the last integer in $\Upsilon_{p_{q_j,j}}$, and S is the first integer in $\Upsilon_{p_{q_j,j}+2}$.

Let's partition $\{s_0, s_0 + 1, \dots, S\}$ into nonempty pieces:

- $I_0 := \{s_0, s_0 + 1, \dots, s_1\} \subset \{s_0, s_0 + 1, \dots, S\}$ is such that $(\widetilde{\sigma}_s, j) \in \operatorname{dom}\chi$ for all $s \in I_0$;
- $I_1 := \{s_1, s_1 + 1, \dots, s_2\} \subset \{s_0, s_0 + 1, \dots, S\}$ is such that $(\widetilde{\sigma}_s, j + 1) \in \operatorname{dom}\chi$ for all $s \in I_1$;
- $\dots$
- $I_\iota := \{s_\iota, s_\iota + 1, \dots, s_{\iota+1}\} \subset \{s_0, s_0 + 1, \dots, S\}$ is such that $(\widetilde{\sigma}_s, j + \iota) \in \operatorname{dom}\chi$ for all $s \in I_\iota$;

and either $s_{\iota+1} = S$ or $\operatorname{dom}\chi$ ends at $(\widetilde{\sigma}_{s_{\iota+1}}, j + \iota)$. Essentially, $I_0, \dots, I_\iota$ correspond to different horizontal segments in the portion of $\operatorname{dom}\chi$ that is parameterized by $\{s_0, s_0 + 1, \dots, S\}$. It could be that $\iota = 0$, and that some or all of these horizontal segments consist of one point each. (This is not exactly a partition, as the sets share endpoints.)

For $s, s' \in \{s_0, \dots, S\}$, and thanks to (14),

$$|\widetilde{\sigma}_s - \widetilde{\sigma}_{s'}| \le \tau. \tag{39}$$

For $i = 0, 1, \dots, \iota$ and $s, s' \in I_i$, because of (19), (20),

$$|\chi(\widetilde{\sigma}_s, j + i)) - \chi(\widetilde{\sigma}_{s'}, j + i))| \le \delta/2. \tag{40}$$

For $i = 0, 1, \dots, \iota - 1$,

$$\chi(\widetilde{\sigma}_{s_{i+1}}, j + i)) \in D,$$

$$\chi(\widetilde{\sigma}_{s_{i+1}}, j + i + 1)) \in G(\chi(\widetilde{\sigma}_{s_{i+1}}, j + i))),$$

and thus

$$\chi(\widetilde{\sigma}_{s_i}, j + i)) \in D + \frac{\delta}{2}\mathbb{B},$$

$$\chi(\widetilde{\sigma}_{s_{i+1}}, j + i + 1)) \in G\left(\chi(\widetilde{\sigma}_{s_i}, j + i)) + \frac{\delta}{2}\mathbb{B}\right),$$

$$\chi(\widetilde{\sigma}_{s_{i+2}}, j + i + 1)) \in G\left(\chi(\widetilde{\sigma}_{s_i}, j + i)) + \frac{\delta}{2}\mathbb{B}\right) + \frac{\delta}{2}\mathbb{B}.$$

Thus, the sequence $\chi(\widetilde{\sigma}_{s_0}, j))$, $\chi(\widetilde{\sigma}_{s_2}, j+1))$, $\dots \chi(\widetilde{\sigma}_{s_{\iota+1}}, j + \iota))$, solves the discrete dynamics in (17). Let

$$\Psi := \{(\widetilde{\sigma}_{s_0}, j + i, \chi(\widetilde{\sigma}_{s_i}, j + i)) \,|\, i \in \{0, 1, \dots, \iota\}\}$$

which is, essentially, the graph of the just-mentioned sequence, placed in the (t, j, x) space where the graphs of solutions to hybrid systems (1), (17) exist. Then, Ψ is the graph of

a solution to (17). By (39) and (40), the portion of the graph of χ, namely, $\{(\widetilde{\sigma}_s, j(s), \chi(\widetilde{\sigma}_s, j(s))) \,|\, s \in \{s_0, \ldots, S\}\}$, where $j(s)$ is such that $(s, j(s)) \in \operatorname{dom} \chi$, is a subset of

$$\Psi + \left([0, \tau] \times \{0\} \times \frac{\delta}{2} \mathbb{B} \right). \tag{41}$$

These observations with (37) and (38) establish item c). ■

C. Convergence in probability on compact time domains

The result below claims convergence in probability, on finite-time intervals, to a solution of the hybrid system (1). It is a generalization of [23, Theorem 2] to the setting considered here.

Theorem 2: Under Assumptions 1-3, for each $T > 0$, $J \in \mathbb{Z}_{\geq 0}$, $\varepsilon > 0$, and $\varrho > 0$ there exists $\rho^* > 0$ such that, with $\{\sigma_k^*\}_{k=1}^{\infty} \in \mathcal{TS}(\rho^*, J)$, for each $\{\sigma_k\}_{k=1}^{\infty} \in \mathcal{TS}_{\leq \{\sigma_k^*\}_{k=1}^{\infty}}$ and each $\mathbf{x} \in \mathcal{S}_\sigma$,

$$1 - \varrho \leq \mathbb{P}\Big(\omega \in \Omega : \exists \phi \in \mathcal{S} \text{ s.t.} \\ \operatorname{graph}_{\preceq (T,J)}(\mathbf{x}(\omega)) \subset \operatorname{graph}(\phi) + \varepsilon \mathbb{B}^\circ \Big). \tag{42}$$

IV. Example

We consider extremum-seeking involving a constraint. Violating the constraint induces a mode switch that severely degrades performance as viewed through the cost function. Recovery requires a hysteretic excursion through the phase space. The algorithm aims to keep the system near its optimal behavior for as much time as possible and to recover as quickly as possible when the constraint is violated.

For simplicity, we ignore transient behavior that appears with the change of set points produced by the algorithm. When the transients are fast, which is typical in extremum-seeking problems, this situation can be addressed via multi-time-scale hybrid systems, as in [27] for example.

Consider a fully-actuated system with state $z \in \mathbb{R}^2$, so that $\dot{z} = u$, subject to mode switching, with the mode state $q \in \{0, 1\}$. When $q = 1$, z is constrained to the half plane

$$C_{1,\theta} := \{ z \in \mathbb{R}^2 : z_1 + z_2 \leq \theta \} \tag{43}$$

where $\theta \geq 10$ is an uncertain, and perhaps time-varying or noisy, parameter. When $q = 0$, z is constrained to

$$C_0 := \{ z \in \mathbb{R}^2 : z_1 + z_2 \geq -10 \}. \tag{44}$$

A transition from $q = 1$ to $q = 0$ happens when there is an attempt to leave $C_{1,\theta}$. Similarly, a transition from $q = 0$ to $q = 1$ happens when there is an attempt to leave C_0. Thus, the control system corresponds to the hybrid dynamics

$$(z, q) \in C := (C_0 \times \{0\}) \cup (C_{1,\theta} \times \{1\}) \tag{45a}$$

$$\begin{cases} \dot{z} = u_q \\ \dot{q} = 0 \end{cases} \tag{45b}$$

$$(z, q) \in D := (\overline{\mathbb{R}^2 \backslash C_0} \times \{0\}) \cup (\overline{\mathbb{R}^2 \backslash C_{1,\theta}} \times \{1\}) \tag{45c}$$

$$\begin{cases} z^+ = z \\ q^+ = 1 - q. \end{cases} \tag{45d}$$

The flow set C and jump set D can be made compact by intersecting the sets above with a large compact ball around the origin. The aim of the control system is to drive q to 1 and z to the solution of the optimization problem

$$z^* := \operatorname{argmin}_{z \in C_{1,\theta}} \phi(z) \tag{46}$$

where ϕ is some convex function whose global minimum exists but does not belong to $C_{1,\theta}$. This must be achieved using only stochastic approximations of $\nabla \phi$, and with the uncertainty in θ. The control algorithm may attempt to drive z outside of $C_{1,\theta}$, causing a mode transition from $q = 1$ to $q = 0$. This may happen due to the uncertainty in θ and because $z^* \notin C_{1,\theta}$. When this happens, the system should return to $q = 1$ by driving z to the boundary of C_0. Simultaneously, to anticipate upcoming mode switches, the algorithm maintains an estimate of θ, denoted $\widehat{\theta}$, and updates it each time a mode transition occurs. In mode 0, the algorithm uses a constant controller u_0 to drive z to the boundary of C_0 and force a transition from mode 0 to 1. We use $u_0 := -\begin{bmatrix} 1 & 1 \end{bmatrix}^T$, which can be implemented with sample and hold to correspond to the theory of the paper.

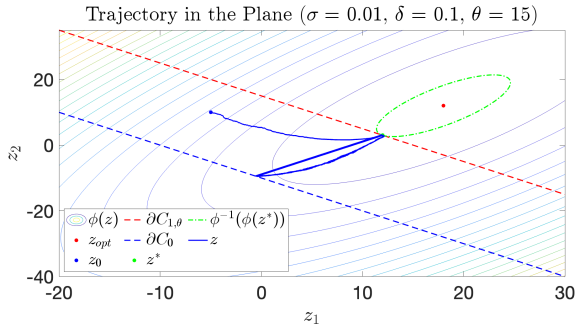

Fig. 1. *The evolution of z relative to z^*. The value z_{opt} refers to the minimizer for the unconstrained optimization problem.*

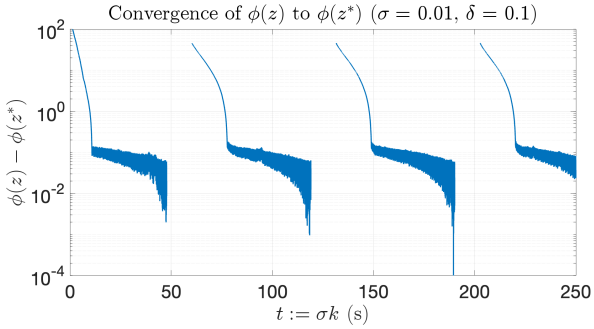

Fig. 2. *Evolution of $\phi(z) - \phi(z^*)$; σ denotes the step size of the iterations in mode 1 while k denotes the iteration number. δ is a parameter that determines the separation of the evaluations when forming the finite differences that approximate the gradient.*

In mode 1, the algorithm uses a (partial) stochastic approximation of the gradient of the (non-smooth) function

$$z \mapsto \phi(z) + K |z|_{C_{1,\hat{\theta}-\varepsilon}} \tag{47}$$

where $K > 0$ is large, $\varepsilon > 0$ is small, and $|z|_{C_{1,\hat{\theta}-\varepsilon}}$ denotes the distance of z to the set $C_{1,\hat{\theta}-\varepsilon}$. To allow for θ changing with time, $\hat{\theta}$ contains slow positive drift during flows that

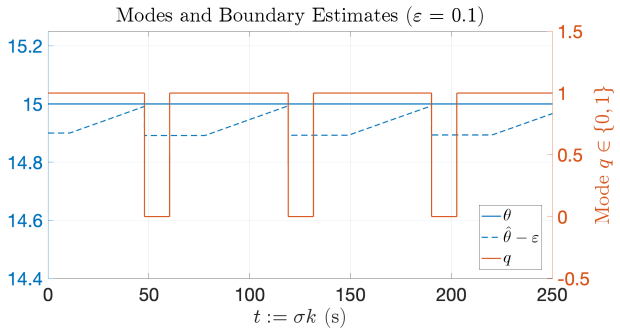

Fig. 3. *Evolution of mode variable q, oscillating between mode* 1 *(normal operation) and mode* 0 *(performance breakdown), and boundary parameter estimate $\hat{\theta}$ of θ (taken to be* 15*) minus the safety factor ε.*

occur when z is in the set $C_{1,\theta} \backslash C_{1,\hat{\theta}-\varepsilon}$. The control objective is to make the state (z, q) repeatedly return to a neighborhood of the optimal point given in (46) while in mode 1.

Figures 1-3 show simulations for the function

$$\phi(z) := z^T P z + b^T z + c, \qquad c := -0.0014$$
$$P := \begin{bmatrix} 0.3439 & -0.1818 \\ -0.1818 & 0.1709 \end{bmatrix}, \; b := \begin{bmatrix} -8.0191 \\ 2.4424 \end{bmatrix}$$

where the gradient is partially approximated with random finite differences of the form

$$z^+ - z \in -\frac{\sigma}{2\delta} v^+ \tanh\left(\phi(z + \delta v^+) - \phi(z - \delta v^+)\right) - \sigma \partial \left(K |z|_{C_{1,\hat{\theta}-\varepsilon}}\right). \quad (48)$$

This difference inclusion aligns with a particular stochastic, sample and hold implementation for $u_1(z, v^+)$. In (48), v^+ is a placeholder for a sequence of iid random variables uniformly distributed over the set $Y := \{e_1, e_2\}$, i.e., the standard basis in $\mathbb{R}^2$, and $\partial(\cdot)$ is the (set-valued) generalized gradient. The simulations use $\sigma = 0.01$, $\delta = 0.1$, $K = 5$, and $\epsilon = 0.1$. The 'tanh' function is used to respect a "velocity" constraint for the updates of z. By the Mean Value Theorem,

$$\phi(z + \delta e_i) - \phi(z - \delta e_i) \in 2\delta e_i^T \nabla \phi(z + \delta \mathbb{B}). \quad (49)$$

The flow map of the target hybrid system is set-valued due to the nonsmoothness of the penalty function $z \mapsto |z|_{C_{1,\hat{\theta}-\varepsilon}}$, and due to the ball that appears in (49). In particular, the target flow map for the z dynamics while in mode 1 is

$$F(z) := 0.5 \frac{1}{2\delta} \left(e_1^T \tanh\left(2\delta e_1^T \nabla \phi(z + \delta \mathbb{B})\right) + e_2^T \tanh(2\delta e_2^T \nabla \phi(z + \delta \mathbb{B})\right) - \partial \left(K |z|_{C_{1,\hat{\theta}-\varepsilon}}\right).$$

Figure 1 shows the evolution of z. The straight part of the hysteretic cycle shows z evolving from the boundary of $C_{1,\theta}$ to the boundary of C_0 when $q = 0$. The curved part corresponds to gradient descent implemented with partial finite differences. Figure 2 shows $\phi(z) - \phi(z^*)$; when $q = 0$, corresponding to performance degradation, the value of $\phi(z)$ is meaningless and is not shown. Figure 3 shows the evolution of the mode state q and the estimate $\hat{\theta}$ (of θ, which is taken to be 15 in the simulations) minus the safety factor ε. The plots indicate success of the proposed stochastic-approximation-based control algorithm in inducing recurrence of a neighborhood of the desired optimal point.

References

[1] H. Robbins and S. Monro, "A stochastic approximation method," *The Annals of Mathematical Statistics*, pp. 400–407, 1951.

[2] J. Kiefer and J. Wolfowitz, "Stochastic estimation of the maximum of a regression function," *The Annals of Math. Stat.*, pp. 462–466, 1952.

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

[4] Y. M. Ermoliev, "On the stochastic quasi-gradient method and stochastic quasi-Feyer sequences," *Kibernetika*, vol. 2, pp. 72–83, 1969.

[5] Y. Ermoliev, "Stochastic quasigradient methods and their application to system optimization," *Stochastics: An International Journal of Probability and Stochastic Processes*, vol. 9, no. 1-2, pp. 1–36, 1983.

[6] L. Ljung, "Analysis of recursive stochastic algorithms," *IEEE Transactions on Automatic Control*, vol. 22, no. 4, pp. 551–575, 1977.

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

[8] H. J. Kushner and A. Shwartz, "An invariant measure approach to the convergence of stochastic approximations with state dependent noise," *SIAM J. Control Optim.*, vol. 22, p. 13 – 27, 1984.

[9] H. J. Kushner, G., and Yin, "Asymptotic properties of distributed and communicating stochastic approximation algorithms," *SIAM Journal on Control and Optimization*, vol. 25, no. 5, pp. 1266–1290, 1987.

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

[11] M. Benaïm, J. Hofbauer, and S. Sorin, "Stochastic approximations and differential inclusions," *SIAM J. on Control and Opt.*, vol. 44, no. 1, pp. 328–348, 2005.

[12] A. Ramaswamy and S. Bhatnagar, "A generalization of the Borkar-Meyn theorem for stochastic recursive inclusions," *Mathematics of Operations Research*, vol. 42, no. 3, pp. 648–661, 2017.

[13] V. G. Yaji and S. Bhatnagar, "Analysis of stochastic approximation schemes with set-valued maps in the absence of a stability guarantee and their stabilization," *IEEE Trans. on Auto. Control*, vol. 65, no. 3, pp. 1100–1115, 2020.

[14] N. Nguyen and G. Yin, "Stochastic approximation with discontinuous dynamics, differential inclusions, and applications," *The Annals of Applied Probability*, vol. 33, no. 1, pp. 780–823, 2023.

[15] L. Tavernini, "Differential automata and their discrete simulators," *Nonlinear Anal.*, vol. 11, no. 6, pp. 665–683, 1987.

[16] A. van der Schaft and H. Schumacher, *An Introduction to Hybrid Dynamical Systems*. Springer, 2000.

[17] J.-P. Aubin, J. Lygeros, M. Quincampoix, S. Sastry, and N. Seube, "Impulse differential inclusions: a viability approach to hybrid systems," *IEEE Trans. on Auto. Control*, vol. 47, no. 1, pp. 2–20, 2002.

[18] J. Lygeros, K. Johansson, S. Simić, J. Zhang, and S. S. Sastry, "Dynamical properties of hybrid automata," *IEEE Transactions on Automatic Control*, vol. 48, no. 1, pp. 2–17, 2003.

[19] R. Goebel and A. R. Teel, "Solutions to hybrid inclusions via set and graphical convergence with stability theory applications," *Automatica*, vol. 42, pp. 573–587, 2006.

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

[21] A. Abate, A. Ames, and S. Sastry, "Stochastic approximations of hybrid systems," in *Proc. of the ACC*, 2005, pp. 1557–1562.

[22] A. Abate, A. D. Ames, and S. Sastry, "Error bounds based stochastic approximations and simulations of hybrid dynamical systems," in *Proceedings of the Amer. Control Conf.*, 2006, pp. 4742–4747.

[23] A. R. Teel and R. Goebel, "Stochastic approximations of hybrid systems with continuous flow maps," in *Proceedings of the 8th IFAC Conference on Analysis and Design of Hybrid Systems*, 2024.

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

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

[26] R. G. Sanfelice and A. R. Teel, "Dynamical properties of hybrid systems simulators," *Automatica*, vol. 46, no. 2, pp. 239–248, 2010.

[27] ——, "On singular perturbations due to fast actuators in hybrid control systems," *Automatica*, vol. 47, no. 4, pp. 692–701, 2011.