# Distributed Tracking of Correlated Equilibria in Regime Switching Noncooperative Games

Omid Namvar Gharehshiran, *Student Member, IEEE*, Vikram Krishnamurthy, *Fellow, IEEE*, and George Yin, *Fellow, IEEE*

***Abstract*—We consider a class of regime-switching noncooperative repeated games where agents exchange information over a graph. The parameters of the game (number of agents, payoffs, information exchange graph) evolve randomly over time according to a Markov chain. We present a regret-based stochastic approximation algorithm with constant step-size that prescribes how individual agents update their randomized strategies over time. We show that, if the Markov chain jump changes on the same timescale as the adaptation rate of the stochastic approximation algorithm and agents independently follow this algorithm, their collective behavior is agile in tracking the time-varying convex polytope of correlated equilibria. The analysis is carried out using weak convergence methods and Lyapunov stability of switched Markovian differential inclusions.**

***Index Terms*—Coordination in decision making, correlated equilibrium, Lyapunov stability, Markov chain, noncooperative game, regime-switching differential inclusion, stochastic approximation, tracking.**

## I. Introduction

CONSIDER a multi-agent noncooperative repeated game, where individual agents choose their actions from a randomized policy. Each agent refines its randomized policy over time using a stochastic approximation algorithm to optimize its regret function. It is well known (see [1]–[3]) that if the stochastic approximation algorithm deployed by each agent has decreasing step-size, the collective behavior (joint empirical frequency of actions taken by all agents) converges with probability one to a convex polytope that is the set of *correlated equilibria* (CE) [4]. Such emergent collective behavior can be viewed as coordination in decision making, since individual agents act autonomously, yet eventually all agents pick their actions from a common convex polytope of joint strategies. The CE has several advantages compared to Nash equilibria (NE) in such learning environments: 1) structural simplicity: CE forms a convex polytope, whereas the NE are isolated points at the extrema of this polytope [5]; 2) coordination capability: The CE directly takes into account the ability of agents to coordinate their actions. This coordination leads to potentially higher payoffs than if agents take their actions independently as required by NE [4]; and 3) CE is realistic in learning since the environment naturally correlates agents' actions. In contrast, NE assumes agents act independently which is rarely true in learning.

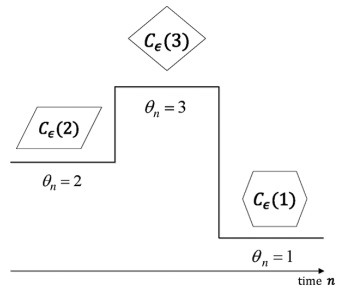


Fig. 1. Tracking convex polytope of correlated $\epsilon$-equilibria $\mathcal{C}_\epsilon(\theta_n)$ randomly evolving according to a Markov process $\theta_n$. Theorem 3.1 shows that such tracking is attainable via the proposed regret-based stochastic approximation algorithm.

*Main Results:* In this paper, we assume that the parameters of the game (number of agents, payoffs) evolve randomly over time. Therefore, the polytope comprising of the CE evolves over time. Our objective is to devise a stochastic approximation algorithms that prescribes how individual agents update their randomized strategies over time such that their collective behavior now tracks the time-evolving polytope of CE. Such tracking problems lie at the heart of applications of stochastic approximation algorithms [6]. This paper has two main results:

First, agents can exchange information with neighbors over an undirected graph. Such neighborhood structure incorporates a broad range of topologies. Each agent shares his picked action profile only with neighbors, however, is oblivious to the existence of non-neighbors. Agents take actions repeatedly and receive: 1) local payoffs: due to local interaction (e.g., performing collaborative tasks) within neighborhood, 2) global payoffs: due to strategic global interaction with non-neighbors. Parameters of the game model (e.g., payoff function, connectivity graph, etc.) may evolve with time randomly. Given the game model and the information exchange structure, *how can the agents pick decisions so that the collective behavior tracks the time-evolving CE set of the game*? Section III-A presents a regret-based stochastic approximation algorithm that exploits this information exchange structure and prescribes such individual strategies to the agents.

The second main result concerns performance analysis of the proposed algorithm. We seek to answer: When individual agents deploy the regret-based stochastic approximation algorithm, is the collective behavior agile in tracking the time-varying CE set of the randomly evolving game? It is well known that: 1) If the underlying parameters change too drastically, there is no chance one can track the time-varying properties via an adaptive stochastic approximation algorithm. (Such a phenomenon is known as trackability; see [7] for related discussions.) 2) If the parameters evolve on a slower timescale as compared to the stochastic approximation algorithm, they remain constant in the

Manuscript received May 23, 2012; revised February 01, 2013 and March 21, 2013; accepted March 27, 2013. Date of publication April 03, 2013; date of current version September 18, 2013. This work was supported by the NSERC Strategic grant and the Canada Research Chairs program. Recommended by associate editor H. S. Chang.

O. Namvar Gharehshiran and V. Krishnamurthy are with the Department of Electrical and Computer Engineering, University of British Columbia, Vancouver V6T 1Z4, Canada (e-mail: omidn@ece.ubc.ca; vikramk@ece.ubc.ca).

G. Yin is with the Department of Mathematics, Wayne State University, Detroit, MI 48202 USA (e-mail: gyin@math.wayne.edu).

Color versions of one or more of the figures in this paper are available online at http://ieeexplore.ieee.org.

Digital Object Identifier 10.1109/TAC.2013.2256684

fast timescale and their variation can be asymptotically ignored. The analysis in this paper differs as we show that, if the random evolution of the game is modelled by an unknown Markov chain that evolves on the *same* timescale as the stochastic approximation algorithm, tracking of the CE set can be achieved. That is, if this Markov chain has transition matrix $I + \mu Q$, where $\mu$ is the constant step-size of the stochastic approximation algorithm, then the collective behavior of all agents converges weakly to set of *correlated $\epsilon$-equilibria* ($\epsilon$-CE) [4]. In contrast to the standard treatment of stochastic approximation algorithms, the limiting system will be shown to follow a Markovian switched dynamical system in such cases.

Note that the stochastic approximation algorithm deployed by each agent does not assume a Markovian structure for the time evolution of the game. The Markovian assumption is only used in our weak convergence analysis to ascertain if the collective behavior is sufficiently agile to track the time-varying CE set. Tracking time-varying equilibria, when the underlying dynamics are unknown, is essential in applications such as sensor networks [8], [9] and cognitive radio [10], [11], where network entities operate in a time-varying (non-stationary) environment. The Markov switching assumption is also well-founded in such applications. Suppose $\theta_n$ denotes the number of alive sensors at time $n$ and that sensors run out of battery according to a geometric distribution. Then, $\theta_n$ evolves according to a slow Markov chain.

Our performance analysis is based on [12] and proceeds as follows: First, by a combined use of updated treatment on stochastic approximation [6] and Markov switched systems [13], [14], Theorem 4.1 in Section IV-A shows that the limit system for the discrete time iterates of Algorithm 1 is a randomly switching system of interconnected differential inclusions modulated by a continuous time Markov chain. This is in contrast to the usual approach of stochastic approximation, where the limit is a deterministic ordinary differential equation (ODE). (Differential inclusions arise naturally in game-theoretic learning, since the strategies of other agents are unknown.) We use weak convergence methods [6] to carry out the analysis. We then proceed with the stability analysis of the limit system. Using Lyapunov function methods for randomly switched systems [15], [16], it will be proved that the limit dynamical system is asymptotically stable almost surely. To complete the analysis, Section IV-D shows that tracking the set of global attractors of the derived individual limit systems by all agents provides the necessary and sufficient condition for the collective behavior of agents to track the convex polytope of $\epsilon$-CE.

*Literature:* Regret-matching as a strategy of play in long-run interactions has been introduced in [1], [2]. In [1], it is proved that when all agents share stage actions and follow the proposed regret-based adaptive procedure, the collective behavior converges to the CE. In [2], the authors assumed that agents do *not* observe others' actions and proposed a reinforcement learning procedure that converges to the polytope of $\epsilon$-CE in static games. While there are several papers that present procedures converging to CE in static environments [1], [2], [17], fewer consider learning in dynamic settings. This paper is inspired by [1], [2] and focuses for the first time on learning algorithms when agents observe the actions of some (but not all) of other agents and the underlying game evolves with time. Empirical numerical studies in Section V further illustrates that performance gains can be achieved by exploiting the information disclosed in the vicinity of agents, e.g., in static games, an order of magnitude faster convergence to $\epsilon$-CE polytope can be achieved, as compared to the reinforcement learning procedure in [2].

*Organization:* The rest of the paper is organized as follows: The Markov switched game with neighborhood structure is formulated in Section II. In Section III, the regret-based stochastic approximation algorithm is presented, the notion of $\epsilon$-CE is defined, and the main theorem of the paper characterizing the emergent behavior is given. Section IV presents the detailed proof of the main theorem. Finally, a numerical example is provided in Section V followed by the concluding remarks in Section VI.

## II. Markov Switched Game With Neighborhood Structure

In this section, we introduce a class of noncooperative games with regime switching modulated by a finite-state Markov chain. As mentioned above, the Markovian variation will only be used to analyze the agility of the regret-matching algorithm—it does not appear in the game-theoretic tracking algorithm. Consider the finite repeated noncooperative game

$$\mathcal{G} = \left(\mathcal{K}, (\mathcal{A}^k)_{k\in\mathcal{K}}, \mathcal{C}, \theta, P^\rho, (U^k)_{k\in\mathcal{K}}, (\boldsymbol{\sigma}^k)_{k\in\mathcal{K}}\right) \tag{1}$$

where each component is described as follows:

1) *Set of Agents*: $\mathcal{K} = \{1, 2, \ldots, K\}$. Individual agents are denoted by $k \in \mathcal{K}$.

2) *Action Set*: $\mathcal{A}^k = \{1, 2, \ldots, A^k\}$ denotes the set of actions for each agent $k \in \mathcal{K}$, where $|\mathcal{A}^k| = A^k$. Individual agent $k$'s action is denoted by $a^k \in \mathcal{A}^k$.

3) *Connectivity Graph $\mathcal{C}$*: Each agent $k$ exchanges information only with the neighbors $\mathcal{N}_k$ that are determined by the connectivity graph $\mathcal{C} = (\mathcal{V}, \mathcal{E})$. Here, $\mathcal{C}$ is an undirected graph, where $\mathcal{V} = \mathcal{K}$ is the set of agents. We make the *neighborhood monitoring* assumption (as opposed to the standard *perfect monitoring* [18]):

$$(k, k') \in \mathcal{E} \Leftrightarrow \begin{array}{c} k \text{ knows } a_n^{k'} \text{ and } k' \text{ knows } a_n^k \\ \text{at the end of period } n. \end{array} \tag{2}$$

The *open* and *closed neighborhoods* for each agent $k$ are then defined by $\mathcal{N}_k = \{k' \in \mathcal{V} \setminus \{k\}; (k, k') \in \mathcal{E}\}$ and $\mathcal{N}_k^c = \mathcal{N}_k \cup \{k\}$, respectively. Let further $\mathcal{S}_k = \mathcal{K} \setminus \mathcal{N}_k^c$ denote the set of non-neighbors of agent $k$.

4) *Markov Chain $\theta_n$ and Transition Dynamics $P^\rho$*: The game model may evolve with time due to i) agents joining/leaving the game, ii) changes in agents' payoffs, and iii) changes in neighborhoods. Suppose all time-variant parameters are finite-state and absorbed to a vector indexed by $\theta$. The game evolves according to a discrete time finite-state Markov chain $\theta_n$ with state space $\mathcal{M}_\theta = \{1, \ldots, \Theta\}$ and transition matrix

$$P^\rho = I + \rho Q. \tag{3}$$

Here, $\rho > 0$ is a small parameter, $I$ denotes the $\Theta \times \Theta$ identity matrix, and $Q = (q_{ij}) \in \mathbb{R}^{\Theta\times\Theta}$ is the generator of a continuous time Markov chain satisfying

$$q_{ij} \geq 0 \text{ for } i \neq j, \quad |q_{ij}| \leq 1, \ \forall i, j \in \mathcal{M}_\theta \text{[1]}$$

and

$$\sum_{j=1}^{\Theta} q_{ij} = 0 \quad \text{for all } i = 1, \ldots, \Theta. \tag{4}$$

[1]Note that this is no loss of generality. Given an arbitrary non-truly absorbing Markov chain with transition probability matrix $P^{\overline{\rho}} = I + \overline{\rho}\overline{Q}$, one can form $Q = (1/q_{\max}) \cdot \overline{Q}$, where $q_{\max} = \max_{i\in\mathcal{M}_s} |\overline{q}_{ii}|$. To ensure that the two Markov chains generated by $Q$ and $\overline{Q}$ evolve on the same timescale, $\rho = \overline{\rho} \cdot q_{\max}$.

Choosing $\rho$ small enough ensures that $p_{ij}^{\rho} = \delta_{ij} + \rho q_{ij} > 0$ in the entries of (3), where $\delta_{ij}$ denotes the Kronecker $\delta$ function satisfying $\delta_{ij} = 1$ if $i = j$ and 0 otherwise.

*Remark 2.1:* Agents do *not* observe the Markov chain $\theta_n$ and are oblivious to its dynamics (3).

5) *Payoff Function*: $U^k : \mathcal{A}^{\mathcal{K}} \times \mathcal{M}_\theta \to \mathbb{R}$ is bounded and denotes agent $k$'s payoff. Here, $\mathcal{A}^{\mathcal{K}} = \times_{k\in\mathcal{K}}\mathcal{A}^k$ represents the set of $K$-tuple of *action profiles*. A generic element of $\mathcal{A}^{\mathcal{K}}$ is denoted by $\boldsymbol{a} = (a^1, \ldots, a^K)$ and, following the common notation in game theory, can be rearranged as $(a^k, \boldsymbol{a}^{-k})$ for any agent $k$, where $\boldsymbol{a}^{-k} \in \times_{k'\in\mathcal{K}\setminus\{k\}}\mathcal{A}^{k'}$. We further rearrange $\boldsymbol{a}^{-k}$ as $(\boldsymbol{a}^{\mathcal{N}_k}, \boldsymbol{a}^{\mathcal{S}_k})$, where $\boldsymbol{a}^{\mathcal{N}_k} \in \mathcal{A}^{\mathcal{N}_k} = \times_{k'\in\mathcal{N}_k}\mathcal{A}^{k'}$ denotes the action profile of agent $k$'s neighbors and $\boldsymbol{a}^{\mathcal{S}_k} \in \mathcal{A}^{\mathcal{S}_k} = \times_{k'\in\mathcal{S}_k}\mathcal{A}^{k'}$ denotes the action profile of agent $k$'s non-neighbors, respectively.

The payoff is comprised of two terms: 1) *local payoff*, due to local interaction (e.g., performing tasks) within neighborhoods; 2) *global payoff*, due to strategic global interaction with agents outside neighborhood. Formally,

$$U^k(a^k, \mathbf{a}^{-k}; \theta) = U_l^k(a^k, \boldsymbol{a}^{\mathbf{a}_k}; \theta) + U_g^k(a^k, \boldsymbol{a}^{S_k}; \theta). \quad (5)$$

We assume $U_l^k(a^k, \boldsymbol{a}^{\mathcal{N}_k}; \theta) = 0$ if $\mathcal{N}_k = \emptyset$.

We work with discrete time $n = 1, 2, \ldots$ Each agent $k$ takes an action $a_n^k$ at time $n$ and receives a payoff $U_n^k(a_n^k)$. Agents know their local payoff function $U_l^k(\cdot; \theta)$ at each time $n$; Hence, taking action $a_n^k$ and knowing $\boldsymbol{a}_n^{\mathcal{N}_k}$, are capable of evaluating their local stage payoff. Agents do *not* know the global payoff function $U_g^k(\cdot; \theta)$, however, can compute the *realized* global payoffs by

$$U_{g,n}^k\left(a_n^k\right) = U_n^k\left(a_n^k\right) - U_l^k\left(a_n^k, \boldsymbol{a}_n^{\mathcal{N}_k}; \theta_n\right). \quad (6)$$

Note that $U_{g,n}^k(\cdot) = U_g^k(\cdot, \boldsymbol{a}_n^{\mathcal{S}_k}; \theta_n)$ is a time-variant function (partly due to the time-varying actions of non-neighbors $\boldsymbol{a}_n^{\mathcal{S}_k}$ and partly due to the Markov chain $\theta_n$). Even if agents knew $U_g^k(\cdot; \theta)$, they could not directly compute global payoffs as they do not acquire $\boldsymbol{a}_n^{\mathcal{S}_k}$.

6) *Strategy* $\boldsymbol{\sigma}^k$: Agent $k$ selects action according to a randomized strategy $\boldsymbol{\sigma}^k$ updated according to the stochastic approximation algorithm in Section III. A randomized strategy $\boldsymbol{\sigma}^k$ for agent $k$ belongs to the simplex $\Delta\mathcal{A}^k = \{\boldsymbol{p}^k \in \mathbb{R}^{A^k}; p^k(a) \geq 0, \sum_{a\in\mathcal{A}^k} p^k(a) = 1\}$ and is a map

$$\boldsymbol{\sigma}_{n+1}^k : \bigcup_n \mathcal{H}_n^k \to \Delta\mathcal{A}^k \quad (7)$$

where $\mathcal{H}_n^k = \left(\times_{k'\in\mathcal{N}_k^c}\mathcal{A}^{k'}, U^k\right)^n$ denotes the space of all possible joint moves of agent $k$, his neighbors $\mathcal{N}_k$ and possible payoffs to agent $k$ up to period $n$. The $\sigma$-algebra generated by $\boldsymbol{h}_n^k \in \mathcal{H}_n^k$ will be denoted by $\mathcal{F}_n^k$.

*Remark 2.2:* 1) Because of the transition matrix given in (3), $\theta_n$ changes slowly in time. Hence, although the game is evolving with time, it is piecewise constant. Simply put, $\theta_n$ takes a constant value $\imath$ for a random duration and jumps to $\jmath \neq \imath$, at a random time. For $\rho = 0$ in (3), the game $\mathcal{G}$ in (1) represents a static game which is an extension of that considered in [1], [2].

2) With a slight abuse of notation, $U^k(\cdot; \theta)$ will be used to denote the multilinear extension of the payoff function to the set of mixed strategies, henceforth, simply referred to as "strategies."

3) *Separable Approximation of Payoffs*: Consider the abstract setting where agents know their *overall* payoff function $U^k(a^k, \boldsymbol{a}^{-k}; \theta)$ and observe the actions of some (but not all) agents. One can apply "separable approximation" techniques to approximate the payoff function as

$$U^k(a^k, \boldsymbol{a}^{-k}; \theta) \approx U_{\text{in}}^k(a^k, \boldsymbol{a}^{\mathcal{N}_k}; \theta) + U_{\text{out}}^k(a^k, \boldsymbol{a}^{\mathcal{S}_k}; \theta). \quad (8)$$

Thus, the framework of this paper applies to a wider range of games where agents receive incremental information other than perceived payoffs.

## III. Tracking Correlated $\epsilon$-Equilibria: Algorithm and Main Results

This section presents the stochastic approximation algorithm that each agent employs to update strategies. We provide a precise definition for the collective behavior of agents and define the polytope of $\epsilon$-CE. Theorem 3.1 then proves that if every agent follows the proposed algorithm independently, the collective behavior tracks the slow Markovian switched polytope of $\epsilon$-CE. The convergence analysis in Section IV focuses on the case where $\rho = \mu$. That is, the game evolves on the same timescale as the adaptation rate of the stochastic approximation learning algorithm. (We do not consider $\rho = \mathcal{O}(\mu^2)$ as the analysis is similar to the case $\rho = 0$. That is, when $\theta_n$ evolves on a slower timescale, then on a fast timescale it can be regarded as a constant and its variation can be asymptotically ignored.)

### A. Regret-Based Stochastic Approximation Algorithm

Suppose the game $\mathcal{G}$, defined in (1), is played in discrete time $n = 1, 2, \ldots$ Define $|x|^+ = \max\{0, x\}$ for scalar $x$ and $|\boldsymbol{x}|^+ = (|x_1|^+, \ldots, |x_n|^+)$ for vector $\boldsymbol{x} \in \mathbb{R}^n$. Further let $I_{\{\cdot\}}$ denote the indicator function. The following algorithm is performed independently by each agent $k \in \mathcal{K}$.

**Algorithm 1***: Regret-based Learning with Neighborhood Information Exchange:*

***Step 0***) **Initialization**: Set the exploration factor $0 < \delta < 1$, the adaptation rate $0 < \mu \ll 1$, and $\xi^k > A^k \cdot |U_{\max}^k - U_{\min}^k|$, where $U_{\max}^k$ and $U_{\min}^k$ denote the upper and lower bounds on the payoff function for agent $k$, respectively. Initialize $\boldsymbol{\psi}_0^k = (1/A^k) \cdot \mathbb{1}_{A^k}$, where $\mathbb{1}_{A^k} = [1, \cdots, 1]_{A^k\times 1}$, $\boldsymbol{\alpha}_0^k = \mathbf{0}_{A^k\times A^k}$, and $\boldsymbol{\beta}_0^k = \mathbf{0}_{A^k\times A^k}$.

For $n = 1, 2, \ldots$, repeat:

***Step 1***) **Strategy Update and Action Selection**: Select action $a_n^k \sim \boldsymbol{\psi}_n^k$: (see (9) at the bottom of the page).

***Step 2***) **Local Information Exchange**: i) Broadcast picked action $a_n^k$ to neighbors $\mathcal{N}_k$; ii) Receive neighbors' action profile $\boldsymbol{a}_n^{\mathcal{N}_k}$.

***Step 3***) **Regret Update**: For all $i, j \in \mathcal{A}^k$:

$$\psi_n^k(i) = \mathbb{P}\left[a_n^k = i | a_{n-1}^k\right] = \begin{cases} (1-\delta)\min\left\{\frac{1}{\xi^k}\left|\alpha_n^k\left(a_{n-1}^k, i\right) + \beta_n^k\left(a_{n-1}^k, i\right)\right|^+, \frac{1}{A^k}\right\} + \frac{\delta}{A^k}, & i \neq a_{n-1}^k \\ 1 - \sum_{j\in\mathcal{A}^k\setminus\{i\}} \psi_n^k(j), & i = a_{n-1}^k \end{cases} \quad (9)$$

*Step 3.1*: Local-Regret Update (see (10) at the bottom of the page).
*Step 3.2*: Global-Regret Update (see (11) at the bottom of the page). In (11), $U^k_{g,n}(a^k_n)$ denotes the global payoff *realized* at time $n$ [computed according to (6)].

***Step 4***) **Recursion**: Set $n \leftarrow n+1$ and go to Step 1.

*Remark 3.1:* While the dynamics of $\theta_n$, $\rho$ and $Q$, in (3), will be used in our tracking analysis in Section IV-A, it does not explicitly enter implementation of the algorithm. Agents are indeed oblivious to the fact that their payoffs are dependent on $\theta_n$, nor do they observe $\theta_n$.

Algorithm 1 can be expressed in compact form as a Markov modulated stochastic approximation algorithm of the form

$$\begin{aligned} X_{n+1} &= X_n + \mu\,(U_1(\Phi_n, \Psi_n; \theta_n) - X_n) \\ Y_{n+1} &= Y_n + \mu\,(U_2(\Phi_n, \Xi_n; \theta_n) - Y_n) \\ \mathbb{P}(\Phi_n = j | \Phi_{n-1} = i) &= P_{ij}(X_n, Y_n). \end{aligned} \tag{12}$$

Here, $X_n, Y_n \in \mathbb{R}^r$, where $r = A^k \times A^k$, represent $\boldsymbol{\alpha}^k_n$ and $\boldsymbol{\beta}^k_n$, respectively, and $\theta_n$ is an exogenous Markov chain with transition probability matrix (3). In addition, $\Phi_n$ represents agent $k$'s action $a^k_n$, $\Psi_n$, and $\Xi_n$ represent the joint action profile of his neighbors $\boldsymbol{a}^{\mathcal{N}_k}_n$ and non-neighbors $\boldsymbol{a}^{\mathcal{S}_k}_n$, respectively. Note that in (12) $\Phi_n$ is a Markov chain that is (conditionally) independent of $\Psi_n, \Xi_n$. (However, $\Psi_n, \Xi_n$ may be correlated.) More precisely, given the history $\boldsymbol{h}_{n-1} = \{(\Phi_\tau, \Psi_\tau, \Xi_\tau)\}^{n-1}_{\tau=1}$,

$$\begin{aligned} &\mathbb{P}(\Phi_n = i, \Psi_n = j, \Xi_n = k | \boldsymbol{h}_{n-1}) \\ &\quad = \mathbb{P}(\Phi_n = i | \boldsymbol{h}_{n-1}) \mathbb{P}(\Psi_n = j, \Xi_n = k | \boldsymbol{h}_{n-1}) \\ &\quad = P_{\Phi_{n-1} i}(X_n, Y_n) \mathbb{P}(\Psi_n = j, \Xi_n = k | \boldsymbol{h}_{n-1}). \end{aligned} \tag{13}$$

Equation (12) simply describes a stochastic approximation algorithm that updates $X_n$ and $Y_n$. These then affect the transition matrix of a Markov chain $\Phi_n$ whose sample path is in turn fed back into the stochastic approximation algorithm. This compact formulation shows how Algorithm 1 compares with standard stochastic approximation algorithms [6] and will be used in Appendix A to obtain the regime-switching limit differential inclusion associated with Algorithm 1.

### B. Collective Behavior and Correlated $\epsilon$-Equilibrium

Consider game $\mathcal{G}$ defined in (1). The collective behavior $\overline{\boldsymbol{z}}_n$ is defined as the *discounted empirical frequency of joint play* of all agents up to period $n$. Formally,

$$\overline{\boldsymbol{z}}_n = \mu \sum_{\tau \le n} (1-\mu)^{n-\tau} \boldsymbol{e}_{\boldsymbol{a}_\tau} \tag{14}$$

where $\boldsymbol{e}_{\boldsymbol{a}_\tau}$ denotes the $(\prod_{k\in\mathcal{K}} A^k)$-dimensional unit vector with the element corresponding to $\boldsymbol{a}_\tau$ being equal to 1.

*Remark 3.2:* The collective behavior $\overline{\boldsymbol{z}}_n$ is a system "diagnostic" and is only used for the tracking analysis of Algorithm 1—it does not need to be updated by agents. In multi-agent systems, e.g., sensor networks, a network controller can monitor $\overline{\boldsymbol{z}}_n$ and use it to adjust agents' function.

Let us now define the set of correlated $\epsilon$-equilibria $\mathcal{C}_\epsilon(\theta)$ for the game $\mathcal{G}$.

*Definition 3.1 (Correlated $\epsilon$-Equilibrium [4]):* Let $\boldsymbol{\pi}_\theta$ denote a joint distribution on the joint action space $\mathcal{A}^{\mathcal{K}} = \times_{k\in\mathcal{K}} \mathcal{A}^k$, where $\pi_\theta(\boldsymbol{a}) \ge 0$ for all $\boldsymbol{a} \in \mathcal{A}^{\mathcal{K}}$ and $\sum_{\boldsymbol{a}\in\mathcal{A}^{\mathcal{K}}} \pi_\theta(\boldsymbol{a}) = 1$. The set of *correlated $\epsilon$-equilibria* for each state $\theta \in \mathcal{M}_\theta$, denoted by $\mathcal{C}_\epsilon(\theta)$, is the convex polytope

$$\begin{aligned} \mathcal{C}_\epsilon(\theta) = \Bigg\{ \boldsymbol{\pi}_\theta : &\sum_{\boldsymbol{a}^{-k}\in\mathcal{A}^{-k}} \pi^k_\theta(i, \boldsymbol{a}^{-k}) \big[U^k(j, \boldsymbol{a}^{-k}; \theta) - U^k(i, \boldsymbol{a}^{-k}; \theta)\big] \\ &\le \epsilon,\ \forall i,\ j \in \mathcal{A}^k,\ k \in \mathcal{K} \Bigg\}. \end{aligned} \tag{15}$$

For $\epsilon = 0$ in (15), $\mathcal{C}_0(\theta)$ is called the set of *correlated equilibria*.[2]

In (15), $\pi^k_\theta(i, \boldsymbol{a}^{-k})$ denotes the probability of agent $k$ playing $i$ and the rest playing $\boldsymbol{a}^{-k}$. Note that $\mathcal{C}_\epsilon(\theta)$ jump changes according to $\theta_n$ as the game evolves over time. We will prove that Algorithm 1 generates trajectories of $\overline{\boldsymbol{z}}_n$ that tracks slow Markovian switched $\mathcal{C}_\epsilon(\theta_n)$.

*Interpretation of Correlated Equilibrium as Global Coordination in Decision Making*: Suppose, before the game is played, each agent $k$ receives a private signal $\overline{a}^k \in \mathcal{A}^k$, given by the profile $\overline{\boldsymbol{a}} = (\overline{a}^1, \ldots, \overline{a}^K)$ drawn according to a commonly known distribution $\boldsymbol{\pi}$. A correlated equilibrium arises if all agents $k \in \mathcal{K}$ realize that $\overline{a}^k$ is a best-response to the random estimated play of others, provided that others, as well, follow their private signals. That is, no agent can benefit by deviating from a correlated equilibrium strategy. Therefore, reaching a correlated equilibrium can be viewed as all agents being coordinated in their choice of actions.

### C. Discussion of Algorithm 1

*1) Intuition on Strategy Update:* Each element $\alpha^k_n(i,j)$, $i, j \in \mathcal{A}^k$, gives the *weighted time-averaged local-regret* (losses in payoffs) had the agent selected action $j$ every time he played action $i$ in the past:

$$\begin{aligned} \alpha^k_n(i,j) = \mu \sum_{\tau=1}^{n} (1-\mu)^{n-\tau} \\ \times \left[U^k_l\left(j, \boldsymbol{a}^{\mathcal{N}_k}_\tau; \theta_\tau\right) - U^k_l\left(a^k_\tau, \boldsymbol{a}^{\mathcal{N}_k}_\tau; \theta_\tau\right)\right] I_{\{a^k_\tau = i\}}. \end{aligned} \tag{16}$$

Since agents know $U^k_l(\cdot; \theta_\tau)$ and observe $\boldsymbol{a}^{\mathcal{N}_k}_\tau$, they perform the thought experiment to compute $U^k_l(j, \boldsymbol{a}^{\mathcal{N}_k}_\tau; \theta_\tau)$ for alternative actions $j \in \mathcal{A}^k$, hence, compute the weighted average of

[2]Nash equilibrium corresponds to the special case where agents act independently. That is, $\boldsymbol{\pi}$ is a product measure: $\pi(a^1, \cdots, a^K) = \Pi_{k=1}^K \pi^k(a^k)$. Every Nash equilibrium is thus a correlated equilibrium. The set of correlated equilibria is nonempty, closed and contains the convex hull of Nash equilibria.

$$\alpha^k_n(i,j) = \alpha^k_{n-1}(i,j) + \mu\left[\left[U^k_l\left(j, \boldsymbol{a}^{\mathcal{N}_k}_n; \theta_n\right] - U^k_l\left[a^k_n, \boldsymbol{a}^{\mathcal{N}_k}_n; \theta_n\right]\right] \times I_{\{a^k_n = i\}} - \alpha^k_{n-1}(i,j)\right] \tag{10}$$

$$\beta^k_n(i,j) = \beta^k_{n-1}(i,j) + \mu\left[\left[\frac{\psi^k_n(i)}{\psi^k_n(j)} U^k_{g,n}\left(a^k_n\right) I_{\{a^k_n = j\}} - U^k_{g,n}\left(a^k_n\right) I_{\{a^k_n = i\}}\right] - \beta^k_{n-1}(i,j)\right] \tag{11}$$

local-regrets as in (16). Agents, however, do not receive the action profile of non-neighbors $\boldsymbol{a}_\tau^{S_k}$, hence, are unable to evaluate global payoffs for alternate actions. We thus formulate an *unbiased* estimator $\boldsymbol{\beta}_n^k$ of the *weighted time-averaged global-regrets* based only on $\{U_{g,\tau}^k\}$ [2]. Each element $\beta_n^k(i,j)$, $i,j \in \mathcal{A}^k$, is formulated as

$$\beta_n^k(i,j) = \mu \sum_{\tau=1}^{n} (1-\mu)^{n-\tau} \times \left[ \frac{\psi_\tau^k(i)}{\psi_\tau^k(j)} U_{g,\tau}^k\left(a_\tau^k\right) I_{\{a_\tau^k=j\}} - U_{g,\tau}^k\left(a_\tau^k\right) I_{\{a_\tau^k=i\}} \right]. \quad (17)$$

Positive overall-regrets $\alpha_n^k(i,j)+\beta_n^k(i,j)$ imply the opportunity to gain by switching from action $i$ to $j$ in future. Therefore, switching probabilities $\boldsymbol{\psi}_n^k$ are based on $|\alpha_n^k(i,j)+\beta_n^k(i,j)|^+$. $\delta$ is the exploration factor which is essential as agents continuously learn their global payoff functions. Taking the minimum with $1/A^k$ and $\xi^k$ in (9) also guarantees that $\sum_{a^k \in \mathcal{A}^k} \psi_n^k(a^k) = 1$. The rate of convergence is closely related to $\xi^k$—higher $\xi^k$ leads to slower convergence.

*2) Static Games:* Algorithm 1 can be modified, by replacing constant step-size $\mu$ with decreasing step-size $\mu_n = 1/n$ in (10) and (11), to achieve *almost sure convergence* to fixed polytope of $\epsilon$-CE in static games ($\rho = 0$ in (3)). We show in Section V that such algorithm can lead to an order of magnitude faster convergence to $\mathcal{C}_\epsilon$ as compared to [2].

*3) The Case of Observing $\theta_n$:* The main assumption in this paper is that $\theta_n$ is *not* observed by agents, yet agents jointly track the time-varying polytope of $\epsilon$-CE; see Theorem 3.1. Assuming agents observe $\theta_n$ leads to the less challenging case where agents form $(\boldsymbol{\alpha}_n^k(\theta), \boldsymbol{\beta}_n^k(\theta))$ and run Algorithm 1 (with decreasing step-size $\mu_n = 1/n$) separately for each $\theta \in \mathcal{M}_\theta$. In such setting, it can be shown that agents' collective behavior almost surely converges to $\mathcal{C}_\epsilon(\theta)$ for each $\theta \in \mathcal{M}_\theta$. However, this is not the focus of this paper.

### D. Main Result: From Individual to Collective Behavior

We now characterize the collective behavior emerging by each agent individually following Algorithm 1. To proceed, we pose the following condition:

$$\text{The process } \theta_n \text{ is slow in the sense that } \rho = \mu. \quad (18)$$

That is, the parameters of the game evolve according to an unknown Markov chain evolving at the same timescale as Algorithm 1. The following theorem is the main result of the paper and asserts that, if (18) holds, the collective behavior is agile and can track the time-varying polytope of $\epsilon$-CE. In what follows, $\Rightarrow$ denotes weak convergence (a generalization of convergence in distribution to a function space; see [6] for an exposition).

*Theorem 3.1:* Consider the game $\mathcal{G}$, defined in (1), and suppose (18) holds. Define the continuous time interpolated sequence of iterates:

$$\overline{\boldsymbol{z}}^\mu(t) = \overline{\boldsymbol{z}}_n \quad \text{for} \quad t \in [n\mu, (n+1)\mu]$$

where $\overline{\boldsymbol{z}}_n$ is defined in (14). Suppose further $\overline{\boldsymbol{z}}_0$ is independent of $\mu$ and let $q(\mu)$ be any sequence of real numbers satisfying $q(\mu) \to \infty$ as $\mu \to 0$. For each $\epsilon > 0$, there exists $\widehat{\delta}(\epsilon)$ such that if every agent $k$ employs Algorithm 1 with $\delta < \widehat{\delta}(\epsilon)$ in Step 1, as $\mu \to 0$, $\overline{\boldsymbol{z}}^\mu(\cdot + q(\mu))$ converges weakly to the regime switching polytope of $\mathcal{C}_\epsilon(\theta(\cdot))$ in the sense that

$$d\left(\overline{\boldsymbol{z}}^\mu\left(\cdot + q(\mu)\right), \mathcal{C}_\epsilon\left(\theta(\cdot)\right)\right) = \inf_{\boldsymbol{z} \in \mathcal{C}_\epsilon(\theta(\cdot))} |\overline{\boldsymbol{z}}^\mu\left(\cdot + q(\mu)\right) - \boldsymbol{z}(\cdot)| \Rightarrow 0 \quad (19)$$

where $\theta(\cdot)$ is a continuous time Markov chain with generator $Q$; see (3).

*Proof:* For detailed proof see Section IV. ■

*Interpretation of Theorem 3.1*: The proof uses martingale averaging techniques to show that the limiting behavior converges weakly to a switched Markovian differential inclusion. Then the stability of the differential inclusion is established. From the game-theoretic point of view, Theorem 3.1 says that non-fully-rational local behavior of individual agents (due to utilizing a better-reply rather than a best-reply strategy) leads to sophisticated globally rational behavior, where all agents pick actions from the set of $\epsilon$-CE. Note that Theorem 3.1 deals with tracking the set $\mathcal{C}_\epsilon(\theta(\cdot))$, rather than a point in $\mathcal{C}_\epsilon(\theta(\cdot))$.

*Remark 3.3:* If $\overline{\boldsymbol{z}}_0 = \overline{\boldsymbol{z}}^\mu(0)$, we require that $z^\mu(0)$ converges to $\overline{\boldsymbol{z}}_0$ weakly. However, for simplicity, we choose $\overline{\boldsymbol{z}}_0$ to be independent of $\mu$.

*Remark 3.4 (Generalized Adaptive Strategies):* The adaptive strategy (9) can be generalized as in [3, Sec. 10.3]. Define a continuously differentiable and nonnegative potential function $V(\boldsymbol{\alpha}^k, \boldsymbol{\beta}^k) : \mathbb{R}^r \times \mathbb{R}^r \to \mathbb{R}$, where $r$ denotes $A^k \times A^k$, such that $V(\boldsymbol{\alpha}^k, \boldsymbol{\beta}^k) = 0$ if and only if $\boldsymbol{\alpha}^k + \boldsymbol{\beta}^k \in (\mathbb{R}^r)^-$. In addition, $\nabla V(\boldsymbol{\alpha}^k, \boldsymbol{\beta}^k) \geq \mathbf{0}$ and $\langle \nabla V(\boldsymbol{\alpha}^k, \boldsymbol{\beta}^k), \boldsymbol{\alpha}^k + \boldsymbol{\beta}^k \rangle_F > 0$ if $\boldsymbol{\alpha}^k + \boldsymbol{\beta}^k \notin (\mathbb{R}^r)^-$, where $\nabla V = [\nabla_{ij} V]_{i,j \in \mathcal{A}^k}$ such that $\nabla_{ij} V = (\partial V / \partial \alpha^k(i,j)) + (\partial V / \partial \beta^k(i,j))$ and $\langle \cdot, \cdot \rangle_F$ denotes the *Frobenius inner product*. Then, if any strategy of the form

$$\boldsymbol{\sigma}_{n+1}^k = (1-\delta)\boldsymbol{\varphi}_n^k + \frac{\delta}{A^k} \cdot \mathbb{1}_{A^k} \quad (20)$$

where

$$\sum_{j \in \mathcal{A}^k \setminus \{i\}} \varphi_n^k(j) \nabla_{ji} V\big(\boldsymbol{\alpha}_n^k, \boldsymbol{\beta}_n^k\big) = \varphi_n^k(i) \sum_{j \in \mathcal{A}^k \setminus \{i\}} \nabla_{ij} V\big(\boldsymbol{\alpha}_n^k, \boldsymbol{\beta}_n^k\big)$$

is employed by all agents in lieu of (9), Theorem 3.1 still holds.

## IV. Proof of Theorem 3.1: Tracking Regime-Switching Correlated $\epsilon$-Equilibrium

This section presents the proof of the main result and is organized into four subsections: First, Section IV-A shows that the limit system associated with the discrete time iterates $(\boldsymbol{\alpha}_n^k, \boldsymbol{\beta}_n^k)$ is a switched Markovian system of interconnected differential inclusions. Next, Section IV-B investigates the stability of the limit dynamical system and characterizes its global attractors. Section IV-C then shows asymptotic stability of the interpolated process associated with $(\boldsymbol{\alpha}_n^k, \boldsymbol{\beta}_n^k)$. Finally, Section IV-D shows that the collective behavior emergent from such limit individual dynamics is attracted to the set of correlated $\epsilon$-equilibria.

### A. Weak Convergence to Markovian Switching Differential Inclusions

We use weak convergence methods to carry out the analysis. Before proceeding further, let us recall some definitions and notation. Let $Z_n$ and $Z$ be $\mathbb{R}^r$-valued random vectors. We say $Z_n$ converges weakly to $Z$ ($Z_n \Rightarrow Z$) if for any bounded and continuous function $f(\cdot)$, $Ef(Z_n) \to Ef(Z)$ as $n \to \infty$. The

sequence $\{Z_n\}$ is tight if for each $\eta > 0$ there is a compact set $K_\eta$ such that $P(Z_n \in K_\eta) \geq 1 - \eta$ for all $n$. The definitions of weak convergence and tightness extend to random elements in more general metric spaces. On a complete separable metric space, tightness is equivalent to relative compactness, which is known as Prohorov's Theorem [19]. By virtue of this theorem, we can extract convergent subsequences when tightness is verified. In what follows, we use a martingale problem formulation to establish the desired weak convergence. This usually requires to first prove tightness. The limit process is then characterized using certain operator related to the limit process. We refer the reader to [6, Ch. 7] for further details on weak convergence and related matters.

As is widely used in the analysis of stochastic approximation algorithms, we start by defining the continuous time interpolations of the discrete time pair process $(\boldsymbol{\alpha}_n^k, \boldsymbol{\beta}_n^k)$. We consider the piecewise constant continuous time interpolation of $(\boldsymbol{\alpha}_n^k, \boldsymbol{\beta}_n^k)$ defined as $(\boldsymbol{\alpha}^{k,\mu}(t), \boldsymbol{\beta}^{k,\mu}(t)) = (\boldsymbol{\alpha}_n^k, \boldsymbol{\beta}_n^k)$ for $t \in [n\mu, (n+1)\mu)$. Similarly, the continuous time interpolation for the Markov chain $\theta_n$ is defined as $\theta^\mu(t) = \theta_n$ for $t \in [n\mu, (n+1)\mu)$.

Before proceeding further, a few definitions are in order. Define

$$\boldsymbol{\sigma}^k = (1-\delta)\boldsymbol{\varphi}^k + \frac{\delta}{A^k} \cdot \mathbb{1}_{A^k} \tag{21}$$

where $\boldsymbol{\varphi}^k$ is an invariant measure of $\boldsymbol{\psi}^k$, given in (9), when $\delta = 0$, i.e.,

$$\sum_{j \in \mathcal{A}^k \setminus \{i\}} \varphi^k(j) \left|\alpha^k(j,i) + \beta^k(j,i)\right|^+ = \varphi^k(i) \sum_{j \in \mathcal{A}^k \setminus \{i\}} \left|\alpha^k(i,j) + \beta^k(i,j)\right|^+. \tag{22}$$

Equation (22) together with : i) $\boldsymbol{\varphi}^k \geq \mathbf{0}$, and ii) $\sum_{i \in \mathcal{A}^k} \varphi^k(i) = 1$ forms a linear programming feasibility problem (null objective function). Existence of $\boldsymbol{\varphi}^k$ can be established using strong duality; see [20, Sec. 5.8.2] for details. Let further $\boldsymbol{\nu}^k$ denote a probability measure over the joint action space $\mathcal{A}^{-k}$. The expected local payoff to agent $k$ can then be defined as

$$U_l^k(a^k, \boldsymbol{\nu}^k; \theta) = \sum_{\boldsymbol{a}^{\mathcal{N}_k}, \boldsymbol{a}^{\mathcal{S}_k}} U_l^k(a^k, \boldsymbol{a}^{\mathcal{N}_k}; \theta) \nu^k(\boldsymbol{a}^{\mathcal{N}_k}, \boldsymbol{a}^{\mathcal{S}_k}). \tag{23}$$

Alternatively, one can define the marginal distribution of $\boldsymbol{a}^{\mathcal{N}_k}$ by $n^k(\boldsymbol{a}^{\mathcal{N}_k}) = \sum_{\boldsymbol{a}^{\mathcal{S}_k}} \nu^k(\boldsymbol{a}^{\mathcal{N}_k}, \boldsymbol{a}^{\mathcal{S}_k})$, and write

$$U_l^k(a^k, \boldsymbol{n}^k; \theta) = \sum_{\boldsymbol{a}^{\mathcal{N}_k}} U_l^k(a^k, \boldsymbol{a}^{\mathcal{N}_k}; \theta) n^k(\boldsymbol{a}^{\mathcal{N}_k}). \tag{24}$$

We denote by $\Delta\mathcal{A}^{\mathcal{N}_k}$ the simplex of probability measures over $\mathcal{A}^{\mathcal{N}_k}$. Therefore, $\boldsymbol{n}^k \in \Delta\mathcal{A}^{\mathcal{N}_k}$.

Associated with Algorithm 1, define the switched Markovian system of interconnected differential inclusions:

$$\begin{cases} \frac{d\boldsymbol{\alpha}^k}{dt} \in \mathcal{L}^k\left(\boldsymbol{\alpha}^k, \boldsymbol{\beta}^k; \theta(t)\right) - \boldsymbol{\alpha}^k \\ \frac{d\boldsymbol{\beta}^k}{dt} = G^k\left(\boldsymbol{\alpha}^k, \boldsymbol{\beta}^k; \theta(t)\right) - \boldsymbol{\beta}^k \end{cases} \tag{25}$$

where $\theta(t)$ denotes a continuous time Markov chain with generator $Q$. Theorem 4.1 below determines the specific forms for the matrix-valued set $\mathcal{L}^k(\boldsymbol{\alpha}^k, \boldsymbol{\beta}^k; \theta(t))$ and matrix $G^k(\boldsymbol{\alpha}^k, \boldsymbol{\beta}^k; \theta(t))$ and shows that the limit system associated with the stochastic approximation iterates $(\boldsymbol{\alpha}_n^k, \boldsymbol{\beta}_n^k)$ can be represented by (25). In what follows, we use $r$ to denote $A^k \times A^k$ and use $D([0,\infty) : \mathbb{R}^r \times \mathcal{M}_\theta)$ to denote the space of functions that are defined in $[0,\infty)$ taking values in $\mathbb{R}^r \times \mathcal{M}_\theta$, and that are right continuous and have left limits with Skorohod topology (see [6, p. 228]).

*Theorem 4.1:* Use $\rho = \mu$ in (3). Then, as $\mu \to 0$, the interpolated process $((\boldsymbol{\alpha}^{k,\mu}(\cdot), \boldsymbol{\beta}^{k,\mu}(\cdot)), \theta^\mu(\cdot))$ is tight in $D([0,\infty) : \mathbb{R}^{2r} \times \mathcal{M}_\theta)$ and converges weakly to $((\boldsymbol{\alpha}^k(\cdot), \boldsymbol{\beta}^k(\cdot)), \theta(\cdot))$ that is a solution of the Markovian switched differential inclusion (25). Moreover, $\mathcal{L}^k$ and $G^k$ in (25) are given by

$$\mathcal{L}_{ij}^k\left(\boldsymbol{\alpha}^k, \boldsymbol{\beta}^k; \theta(t)\right) = \left\{ \left[U_l^k\left(j, \boldsymbol{n}^k; \theta(t)\right) - U_l^k\left(i, \boldsymbol{n}^k; \theta(t)\right)\right] \times \sigma^k(i); \boldsymbol{n}^k \in \Delta\mathcal{A}^{\mathcal{N}_k} \right\} \tag{26}$$

$$G_{ij}^k\left(\boldsymbol{\alpha}^k, \boldsymbol{\beta}^k; \theta(t)\right) = \left[U_{g,t}^k\left(j; \theta(t)\right) - U_{g,t}^k\left(i; \theta(t)\right)\right] \sigma^k(i). \tag{27}$$

In (27), $U_{g,t}^k(\cdot;\cdot)$ is the interpolated process of the global payoffs accrued from the game.

*Proof:* The proof uses stochastic averaging theory based on [6]; see Appendix A for the detailed proof in a general setting. ■

*Discussion of Theorem 4.1:* The interconnection of the dynamics in (25) is hidden in $\boldsymbol{\sigma}^k$, defined in (21), which is a function of $(\boldsymbol{\alpha}^k, \boldsymbol{\beta}^k)$; see (26) and (27). Note that, although agents observe $\boldsymbol{a}^{\mathcal{N}_k}$, they are oblivious to the strategies $\boldsymbol{n}^k$ from which $\boldsymbol{a}^{\mathcal{N}_k}$ has been drawn. Different $\boldsymbol{n}^k \in \Delta\mathcal{A}^{\mathcal{N}_k}$ form different trajectories of $\boldsymbol{\alpha}^k(t)$; thus, $\boldsymbol{\alpha}^{k,\mu}(t)$ converges weakly to the trajectories of a regime-switching differential inclusion rather than a deterministic ODE (that is typically used to model asymptotics of stochastic approximation algorithms). The same argument holds for $\boldsymbol{\beta}^{k,\mu}(t)$ except that the limit dynamics follows a regime-switching ODE. This is due to agents being oblivious to the facts: i) non-neighbors exist, and ii) global payoffs are affected by non-neighbors' actions. However, they realize the time-dependency of $U_{g,n}^k(\cdot)$ as taking the same action at various times results in different payoffs.

*Remark 4.1:* 1) *Differential Inclusions* are generalizations of ODEs and are of the form

$$\frac{d}{dt}X \in \mathcal{F}(X) \tag{28}$$

where $X \in \mathbb{R}^r$ and $\mathcal{F} : \mathbb{R}^r \to \mathbb{R}^r$ is a Marchaud map [21]. That is: i) the graph and domain of $\mathcal{F}$ are nonempty and closed, ii) the values $\mathcal{F}(X)$ are convex, and iii) the growth of $\mathcal{F}$ is linear, i.e., there exists $\eta > 0$ such that for every $X \in \mathbb{R}^r$

$$\sup_{Y \in \mathcal{F}(X)} \|Y\| \leq \eta\left(1 + \|X\|\right). \tag{29}$$

Here, $\|\cdot\|$ denotes any norm on $\mathbb{R}^r$. Since $d\boldsymbol{\alpha}^k/dt$ in (25) belongs to a compact convex set, the linear growth condition (29) is trivially satisfied.

2) When $\theta_n$ evolves on a slower timescale, e.g., $\rho = \mathcal{O}(\mu^2)$, it remains constant in the fast timescale (i.e., the regret-based learning algorithm). Therefore, the system of differential inclusions (25) will be deterministic.

*The Case of Observing $\boldsymbol{a}^{\mathcal{S}_k}$:* Suppose agent $k$ knows the function $U_g^k(\cdot;\theta_n)$ and, in contrast to the neighborhood monitoring assumption in Section II, the non-neighbors' joint action profile $\boldsymbol{a}_n^{\mathcal{S}_k}$ is revealed to agent $k$. In lieu of (11), agent $k$ can then use a stochastic approximation algorithm similar to (10) to update global-regrets. Suppose agent $k$ employs a strategy $\widetilde{\boldsymbol{\psi}}_n^k$

of the form (9) with the exception that $\boldsymbol{\beta}_n^k$ is replaced with $\widetilde{\boldsymbol{\beta}}_n^k$ recursively updated according to

$$\widetilde{\beta}_n^k(i,j)=\widetilde{\beta}_{n-1}^k(i,j)+\mu\Big(\left[U_g^k(j,\boldsymbol{a}_n^{\mathcal{S}_k};\theta_n)-U_g^k(a_n^k,\boldsymbol{a}_n^{\mathcal{S}_k};\theta_n)\right]\times I_{\{a_n^k=i\}}-\widetilde{\beta}_{n-1}^k(i,j)\Big). \quad (30)$$

Then, the proof of Theorem 4.1 (see Appendix A) shows the limit system associated with the iterates $(\boldsymbol{\alpha}_n^k,\widetilde{\boldsymbol{\beta}}_n^k)$ is

$$\begin{cases}\frac{d\boldsymbol{\alpha}^k}{dt}\in\mathcal{L}^k\big(\boldsymbol{\alpha}^k,\widetilde{\boldsymbol{\beta}}^k;\theta(t)\big)-\boldsymbol{\alpha}^k\\ \frac{d\widetilde{\boldsymbol{\beta}}^k}{dt}\in\mathcal{G}^k\big(\boldsymbol{\alpha}^k,\widetilde{\boldsymbol{\beta}}^k;\theta(t)\big)-\widetilde{\boldsymbol{\beta}}^k\end{cases} \quad (31)$$

where $\mathcal{L}^k$ and $\mathcal{G}^k$ are given by

$$\begin{aligned}\mathcal{L}_{ij}^k\big(\boldsymbol{\alpha}^k,\widetilde{\boldsymbol{\beta}}^k;\theta(t)\big)&=\big\{\left[U_l^k(j,\boldsymbol{n}^k;\theta(t))-U_l^k(i,\boldsymbol{n}^k;\theta(t))\right]\\&\quad\times\widetilde{\sigma}^k(i);\boldsymbol{n}^k\in\Delta\mathcal{A}^{\mathcal{N}_k}\big\}\\ \mathcal{G}_{ij}^k\big(\boldsymbol{\alpha}^k,\widetilde{\boldsymbol{\beta}}^k;\theta(t)\big)&=\big\{\left[U_g^k\left(j,\boldsymbol{s}^k;\theta(t)\right)-U_g^k\left(i,\boldsymbol{s}^k;\theta(t)\right)\right]\\&\quad\times\widetilde{\sigma}^k(i);\boldsymbol{s}^k\in\Delta\mathcal{A}^{\mathcal{S}_k}\big\}.\end{aligned} \quad (32)$$

In (32), $s^k(\boldsymbol{a}^{\mathcal{S}_k})=\sum_{\boldsymbol{a}^{\mathcal{N}_k}}\nu^k(\boldsymbol{a}^{\mathcal{N}_k},\boldsymbol{a}^{\mathcal{S}_k})$ and $\widetilde{\boldsymbol{\sigma}}^k=(1-\delta)\widetilde{\boldsymbol{\varphi}}^k+(\delta/A^k)\cdot\mathbb{1}_{A^k}$, where $\widetilde{\boldsymbol{\varphi}}^k$ is an invariant measure of $\widetilde{\boldsymbol{\psi}}^k$ (given by (9) after replacing $\boldsymbol{\beta}^k$ with $\widetilde{\boldsymbol{\beta}}^k$) when $\delta=0$.

Note that $U_{g,n}^k(i)=U_g^k(i,\boldsymbol{a}_n^{\mathcal{S}_k};\theta_n)$ and knowing $\boldsymbol{a}_n^{\mathcal{S}_k}$ only affects the ability to compute it. Therefore, $U_{g,n}^k(i)\in\{U_g^k(i,\boldsymbol{a}_n^{\mathcal{S}_k};\theta_n);\boldsymbol{a}_n^{\mathcal{S}_k}\in\mathcal{A}^{\mathcal{S}_k}\}$ and (31) represents the same continuous time process as (25). That is, both $(\boldsymbol{\alpha}_n^k,\boldsymbol{\beta}_n^k)$ and $(\boldsymbol{\alpha}_n^k,\widetilde{\boldsymbol{\beta}}_n^k)$ are stochastic approximations of the same differential inclusion (31). This property allows us to invoke Theorem 4.2 below, which proves that the $\boldsymbol{\psi}$-strategy induces the same asymptotic behavior for the pair processes $\{(\boldsymbol{\alpha}_n^k,\boldsymbol{\beta}_n^k)\}$ and $\{(\boldsymbol{\alpha}_n^k,\widetilde{\boldsymbol{\beta}}_n^k)\}$.

*Theorem 4.2:* Under a $\boldsymbol{\psi}$-strategy [see (9)], the limit sets of the pair processes $\{(\boldsymbol{\alpha}_n^k,\boldsymbol{\beta}_n^k)\}$, defined in (10), (11), and $\{(\boldsymbol{\alpha}_n^k,\widetilde{\boldsymbol{\beta}}_n^k)\}$, defined in (10), (30), coincide.

*Proof:* For a detailed proof, see Appendix C. ■

The above theorem simply asserts that realizing stage global payoffs $U_{g,\tau}^k(a_\tau^k)$ provides sufficient information to construct an unbiased estimator of $\{\widetilde{\boldsymbol{\beta}}_n^k\}$; hence, the two processes $\{(\boldsymbol{\alpha}_n^k,\boldsymbol{\beta}_n^k)\}$ and $\{(\boldsymbol{\alpha}_n^k,\widetilde{\boldsymbol{\beta}}_n^k)\}$ exhibit the same asymptotic behavior. This result will be used in Section IV-D to prove convergence of global behavior to the $\epsilon$-CE polytope.

### *B. Stability Analysis of Markovian Switching Differential Inclusions*

Section IV-A dealt with the first stage of the proof of Theorem 3.1, namely, showing that the limit behavior of the discrete time iterates $(\boldsymbol{\alpha}_n^k,\boldsymbol{\beta}_n^k)$ associated with Algorithm 1 is a switched Markovian system of interconnected differential inclusions (25). This subsection deals with the second stage of the proof of Theorem 3.1, namely, the stability analysis of the switched Markovian system of differential inclusions (25) and characterizing the set of global attractors. We start by defining stability of switched dynamical systems; see [16] and [14, Ch. 9]. In what follows, $d(\cdot,\cdot)$ denotes the usual distance function.

*Definition 4.1:* Consider the switched system

$$\begin{aligned}\dot{X}(t)&=f\left(X(t),\theta(t)\right)\\ X(0)=\ X_0,\theta(0)=&\theta_0,X(t)\in\mathbb{R}^r,\theta(t)\in\mathcal{M}_\theta\end{aligned} \quad (33)$$

where $\theta(t)$ is the switching signal, and $f:\mathbb{R}^r\times\mathcal{M}_\theta\to\mathbb{R}^r$ is locally Lipschitz for all $\theta\in\mathcal{M}_\theta$. A closed and bounded set $H\subset\mathbb{R}^r\times\mathcal{M}_\theta$ is

1) stable in probability if for any $\varepsilon>0$ and $\overline{\gamma}>0$, there is a $\gamma>0$ such that
$$\mathbb{P}\left(\sup_{t\geq0}d\left(\left(X(t),\theta(t)\right),H\right)<\overline{\gamma}\right)\geq1-\varepsilon$$
whenever $d((X_0,\theta_0),H)<\gamma$;
2) asymptotically stable in probability if it is stable in probability and
$$\mathbb{P}\left(\lim_{t\to\infty}d\left(\left(X(t),\theta(t)\right),H\right)=0\right)\to1$$
as $d((X_0,\theta_0),H)\to0$;
3) asymptotically stable almost surely if
$$\lim_{t\to\infty}d\left(\left(X(t),\theta(t)\right),H\right)=0\quad\text{a.s.}$$

With a slight abuse of notation, denote by $\boldsymbol{\alpha}^k(t)$ and $\boldsymbol{\beta}^k(t)$ the local- and global-regret matrices rearranged as $r\times1$ vectors, where $r$ denotes $A^k\times A^k$. Let

$$\mathcal{R}=\left\{(\boldsymbol{\alpha}^k,\boldsymbol{\beta}^k)\in\mathbb{R}^r\times\mathbb{R}^r;|\boldsymbol{\alpha}^k+\boldsymbol{\beta}^k|^+\leq\epsilon\cdot\mathbb{1}_r\right\}. \quad (34)$$

In the following theorem, we break down our analysis of the stability of the switching differential inclusions into two steps; First, we examine the stability of (25) for any fixed switching signal $\theta(t)=\theta$ to show stability of each subsystem. The set of global attractors is shown to comprise $\mathcal{R}$ for all $\theta\in\mathcal{M}_\theta$. The slow switching condition then allows us to apply the method of multiple Lyapunov functions [22, Ch. 3] to analyze stability of the switched system.

*Theorem 4.3:* Consider the system of differential inclusions described by (21)–(25). Let $\mathcal{R}_\theta=\mathcal{R}\times\mathcal{M}_\theta$, $(\boldsymbol{\alpha}^k(0),\boldsymbol{\beta}^k(0))=(\boldsymbol{\alpha}_0,\boldsymbol{\beta}_0)$ and $\theta(0)=\theta_0$. For each $\epsilon>0$, there exists $\widehat{\delta}(\epsilon)$ such that, if $\delta<\widehat{\delta}(\epsilon)$ in (21), the following results hold:

1) If $\theta(t)=\theta$ is fixed, the non-switching system is globally asymptotically stable for each $\theta\in\mathcal{M}_\theta$, i.e.,
$$\lim_{t\to\infty}d\left(\left(\boldsymbol{\alpha}^k(t),\boldsymbol{\beta}^k(t)\right),\mathcal{R}\right)=0. \quad (35)$$
2) The Markovian switching system is globally asymptotically stable almost surely. In particular:
   a) $P(\forall\epsilon>0,\exists\gamma(\epsilon)>0\text{ s.t. }d((\boldsymbol{\alpha}_0,\boldsymbol{\beta}_0),\mathcal{R})\leq\gamma(\epsilon)\Rightarrow\sup_{t\geq0}d((\boldsymbol{\alpha}^k(t),\boldsymbol{\beta}^k(t)),\mathcal{R})<\epsilon)=1$;
   b) $P\Big(\forall\overline{\gamma},\epsilon'>0,\exists T(\overline{\gamma},\epsilon')\geq0\text{ s.t. }d((\boldsymbol{\alpha}_0,\boldsymbol{\beta}_0),\mathcal{R})\leq\overline{\gamma}\Rightarrow\sup_{t\geq T(\overline{\gamma},\epsilon')}d((\boldsymbol{\alpha}^k(t),\boldsymbol{\beta}^k(t)),\mathcal{R})<\epsilon'\Big)=1$.

*Proof:* For detailed proof, see Appendix B. ■

The above theorem states that the set of global attractors of the switching system (25) is the same as that for all non-switching systems, $\theta(t)=\theta\in\mathcal{M}_\theta$ in (25), and constitutes $\mathcal{R}$. This sets the stage for Section IV-D where $\mathcal{R}$ (in regret space) is proved to represent the $\epsilon$-CE polytope (in strategy space).

### C. Asymptotic Stability of the Interpolated Process

In Theorem 4.1, we considered $\mu$ small and $n$ large, but $\mu n$ remained bounded. This gives a limit differential inclusion for the sequence of interest as $\mu \to 0$. Here, we study asymptotic stability and establish that the limit points of the switched differential inclusion and the stochastic approximation algorithm coincide as $t \to \infty$. We thus consider the case where $\mu \to 0$ and $n \to \infty$, however, $\mu n \to \infty$ now. However, instead of considering a two-stage limit by first letting $\mu \to 0$ and then letting $t \to \infty$, we study $(\boldsymbol{\alpha}^{k,\mu}(t+q(\mu)), \boldsymbol{\beta}^{k,\mu}(t+q(\mu)))$ and require $q(\mu) \to \infty$ as $\mu \to 0$. The following corollary concerns asymptotic stability of the interpolated process.

*Corollary 4.1:* Denote by $\{q(\mu)\}$ any sequence of real numbers satisfying $q(\mu) \to \infty$ as $\mu \to 0$. Assume that $\{\alpha_n^k, \beta_n^k : \mu > 0, n < \infty\}$ is tight or bounded in probability. Then, for each $\epsilon \geq 0$, there exists $\widehat{\delta}(\epsilon) \geq 0$ such that if $\delta \leq \widehat{\delta}(\epsilon)$ in (9)

$$\big(\alpha^{k,\mu}(\cdot + q(\mu)), \beta^{k,\mu}(\cdot + q(\mu))\big) \Rightarrow \mathcal{R}, \quad \text{as } \mu \to 0 \tag{36}$$

where $\mathcal{R}$ is defined in (34).

*Sketch of the Proof:* We only give an outline of the proof, which essentially follows from Theorems 4.1 and 4.3. Define $\widehat{\alpha}^{k,\mu}(\cdot) = \alpha^{k,\mu}(\cdot + q(\mu))$ and $\widehat{\beta}^{k,\mu}(\cdot) = \beta^{k,\mu}(\cdot + q(\mu))$. Then, it can be shown that $(\widehat{\alpha}^{k,\mu}(\cdot), \widehat{\beta}^{k,\mu}(\cdot))$ is tight. For any $T_1 < \infty$, take a weakly convergent subsequence of $\{\widehat{\alpha}^{k,\mu}(\cdot), \widehat{\alpha}^{k,\mu}(\cdot - T_1), \widehat{\beta}^{k,\mu}(\cdot), \widehat{\beta}^{k,\mu}(\cdot - T_1)\}$. Denote the limit by $(\widehat{\alpha}^k(\cdot), \widehat{\alpha}^k_{T_1}(\cdot), \widehat{\beta}^k(\cdot), \widehat{\beta}^k_{T_1}(\cdot))$. Note that $\widehat{\alpha}^k(0) = \widehat{\alpha}^k_{T_1}(T_1)$ and $\widehat{\beta}^k(0) = \widehat{\beta}^k_{T_1}(T_1)$. The value of $\{\widehat{\alpha}^k_{T_1}(0), \widehat{\beta}^k_{T_1}(0)\}$ may be unknown, but the set of all possible values of $\{\widehat{\alpha}^k_{T_1}(0), \widehat{\beta}^k_{T_1}(0)\}$ (over all $T_1$ and convergent subsequences) belong to a tight set. Using this and the stability condition and Theorems 4.1, for any $\eta > 0$ there is a $T_\eta < \infty$ such that for all $T_1 > T_\eta$, $d((\widehat{\alpha}^k_{T_1}(T_1), \widehat{\beta}^k_{T_1}(T_1)), \mathcal{R}) \geq 1 - \eta$. This implies that $d((\widehat{\alpha}^k(0), \widehat{\beta}^k(0)), \mathcal{R}) \geq 1 - \eta$. This is the desired result that we wish to establish.

### D. Characterizing the Limit Set as the Correlated $\epsilon$-Equilibria Set

In the final stage of the proof, we characterize the limit set of the differential inclusions as the set of correlated $\epsilon$-equilibria. The key point in the proof is to show that tracking the set of global attractors of (25) individually by all agents provides the necessary and sufficient condition for the collective behavior to track the $\epsilon$-CE polytope.

Associated with (30), define the continuous time interpolated sequence of iterates $\widetilde{\boldsymbol{\beta}}^{k,\mu}(t) = \widetilde{\boldsymbol{\beta}}^k_n$ for $t \in [n\mu, n\mu + \mu)$. Recall further $\overline{\boldsymbol{z}}^\mu(t) = \overline{\boldsymbol{z}}_n$ for $t \in [n\mu, n\mu + \mu)$; see (14). Then, for a fixed-$\theta$ process, i.e., $\theta_\tau = \theta$ for $\tau \geq 0$, we have

$$\begin{aligned}&\lim_{t\to\infty} \alpha^{k,\mu}(i,j)(t) + \widetilde{\beta}^{k,\mu}(i,j)(t)\\ &= \lim_{t\to\infty} \sum_{\boldsymbol{a}^{-k}} \overline{z}^{k,\mu}(i,\boldsymbol{a}^{-k})(t)\big[U^k(j,\boldsymbol{a}^{-k};\theta) - U^k(i,\boldsymbol{a}^{-k};\theta)\big].\end{aligned} \tag{37}$$

In (37), $\overline{z}^{k,\mu}(i,\boldsymbol{a}^{-k})(t)$ denotes the interpolated empirical frequency sequence associated with agent $k$ picking action $i$ and the rest $\boldsymbol{a}^{-k}$. On any convergent subsequence $\{\overline{\boldsymbol{z}}_{n'}\}_{n' \geq 0} \to \boldsymbol{\pi}_\theta$, with slight abuse of notation, let $\overline{\boldsymbol{z}}^\mu(t) = \overline{\boldsymbol{z}}_{n'}$ and $(\boldsymbol{\alpha}^{k,\mu}(t), \widetilde{\boldsymbol{\beta}}^{k,\mu}(t)) = (\boldsymbol{\alpha}^k_{n'}, \widetilde{\boldsymbol{\beta}}^k_{n'})$ for $t \in [n'\mu, n'\mu + \mu)$. This yields

$$\begin{aligned}&\lim_{t\to\infty} \alpha^{k,\mu}(i,j)(t) + \widetilde{\beta}^k_n(i,j)(t)\\ &\qquad = \sum_{\boldsymbol{a}^{-k}} \pi^k_\theta(i,\boldsymbol{a}^{-k})\big[U^k(j,\boldsymbol{a}^{-k};\theta) - U^k(i,\boldsymbol{a}^{-k};\theta)\big].\end{aligned} \tag{38}$$

Finally, comparing (38) with Definition 3.1, we conclude that, for each $\theta \in \mathcal{M}_\theta$, $\overline{\boldsymbol{z}}^\mu(t)$ converges to the $\epsilon$-CE set $\mathcal{C}_\epsilon(\theta)$ if and only if, for all $k \in \mathcal{K}$,

$$\lim_{t\to\infty} \Big|\boldsymbol{\alpha}^{k,\mu}(t) + \widetilde{\boldsymbol{\beta}}^{k,\mu}(t)\Big|^+ \leq \epsilon \cdot \mathbb{1}_r. \tag{39}$$

Combining (39) with (34) and Theorem 4.2 results in

$$\overline{\boldsymbol{z}}^\mu(t) \to \mathcal{C}_\epsilon(\theta) \;\; \textit{iff} \;\; \Big(\boldsymbol{\alpha}^{k,\mu}(t), \boldsymbol{\beta}^{k,\mu}(t)\Big) \to \mathcal{R}, \text{ for all } k \in \mathcal{K}. \tag{40}$$

Finally, combining (40) with Corollary 4.1 proves Theorem 3.1.

TABLE I
SENSORS' LOCAL PAYOFFS IN NORMAL FORM

Local: $(U_l^1, U_l^2, U_l^3; 1)$

| | $S_2:0$ | $S_2:1$ |
|---|---|---|
| $S_1:0$ | $(-1,3,1)$ | $(2,-1,3)$ |
| $S_1:1$ | $(1,-1,3)$ | $(1,4,1)$ |

$S_3:0$

Local: $(U_l^1, U_l^2, U_l^3; 2)$

| | $S_2:0$ | $S_2:1$ |
|---|---|---|
| $S_1:0$ | $(1,-1,3)$ | $(0,3,1)$ |
| $S_1:1$ | $(3,3,1)$ | $(-1,0,3)$ |

$S_3:1$

TABLE II
SENSORS' GLOBAL PAYOFFS IN NORMAL FORM

Global: $(U_g^1, U_g^2, U_g^3; 1)$

| | $S_2:0$ | $S_2:1$ |
|---|---|---|
| $S_1:0$ | $(-1,3,1)$ | $(2,-1,3)$ |
| $S_1:1$ | $(1,-1,3)$ | $(1,4,1)$ |

$S_3:0$

| | $S_2:0$ | $S_2:1$ |
|---|---|---|
| $S_1:0$ | $(1,-1,3)$ | $(0,3,1)$ |
| $S_1:1$ | $(3,3,1)$ | $(-1,0,3)$ |

$S_3:1$

Global: $(U_g^1, U_g^2, U_g^3; 2)$

| | $S_2:0$ | $S_2:1$ |
|---|---|---|
| $S_1:0$ | $(0,6,2)$ | $(2,-4,6)$ |
| $S_1:1$ | $(4,0,8)$ | $(-2,6,6)$ |

$S_3:0$

| | $S_2:0$ | $S_2:1$ |
|---|---|---|
| $S_1:0$ | $(2,-2,4)$ | $(-4,4,10)$ |
| $S_1:1$ | $(0,2,10)$ | $(4,-2,4)$ |

$S_3:1$

## V. NUMERICAL STUDY

This section illustrates the performance of Algorithm 1 with a numerical example. Consider game $\mathcal{G}$, defined in (1), and suppose $\mathcal{K} = \{1,2,3\}$, $\mathcal{A}^k = \{\text{ON}, \text{OFF}\}$. The connectivity graph is characterized by $\mathcal{V} = \mathcal{K}$ and $\mathcal{E} = \{(1,2),(2,1)\}$. The Markov chain $\theta_n$ is defined by the state space $\mathcal{M}_\theta = \{1,2\}$ and $P^\rho = I + \rho Q$, where $Q = \begin{bmatrix} -0.5 & 0.5 \\ 0.5 & -0.5 \end{bmatrix}$. Agents' local and global payoffs are given in Tables I and II, respectively. In our simulations, $\mu = 0.001$ and $\delta = 0.1$ in Algorithm 1, and $\lambda = 1$ in Algorithm 2 below.

As benchmarks to compare the performance of Algorithm 1, we introduce two algorithms: 1) *regret-based reinforcement learning* [2]; and 2) *Unobserved conditional smooth fictitious play*. To the best of our knowledge, these are the only algorithms proved to converge to the polytope of $\epsilon$-CE in scenarios where not all agents exchange action information. The following two

algorithms exploit perceived payoffs to extract information of other agents' behavior.

1) *Regret-Based Reinforcement Learning* [2]: Agents only observe the stream of payoffs, hence, is compliant to our setting. In [2], it is proved that, for decreasing step-size $\mu_n = 1/n$, any limit point of $(1/n)\sum_{\tau\le n} \boldsymbol{e}_{\boldsymbol{a}_\tau}$ is a correlated $\epsilon$-equilibrium of the stage game. For constant step-size $\mu_n = \mu$, similar techniques as in [6] can be used to establish weak convergence of the discounted empirical frequency $\overline{\boldsymbol{z}}_n$ [see (14)] to the set of correlated $\epsilon$-equilibria.

2) *Unobserved Conditional Smooth Fictitious Play*: Standard fictitious play requires each agent to record and update the empirical frequency of actions taken by other agents through time and best-respond to it; see [17] for an extensive treatment. Here, since agents do not observe the actions of non-neighbors, we present a variant which relies only on the realizations of the agents' payoffs. The algorithm is summarized below:

*Algorithm 2: Unobserved Conditional Smooth Fictitious Play*

0) **Initialization**: Set $\lambda$. Initialize $\overline{\boldsymbol{U}}_0^k = \mathbf{0}_{A^k\times A^k}$.

For $n = 1, 2, \ldots$, repeat:

1) **Strategy Update and Action Selection**: Agent $k$ picks action $a_n^k \sim \boldsymbol{\psi}_n^k$:

$$\psi_n^k = \frac{1}{\sum_i \exp\left(\frac{1}{\lambda}\overline{U}_{n-1}^k\left(a_{n-1}^k, i\right)\right)} \times\left[\exp\left(\frac{1}{\lambda}\overline{U}_{n-1}^k\left(a_{n-1}^k,1\right)\right),\ldots,\exp\left(\frac{1}{\lambda}\overline{U}_{n-1}^k\left(a_{n-1}^k,A^k\right)\right)\right]. \quad (41)$$

2) *Average Utility Update*: For all $j \in \mathcal{A}^k$:

$$\overline{U}_n^k(i,j) = \overline{U}_{n-1}^k(i,j) + \mu\left[\frac{\psi_n^k(i)}{\psi_n^k(j)} U_n^k\left(a_n^k\right) I_{\{a_n^k = i\}} - \overline{U}_{n-1}^k(i,j)\right] \quad (42)$$

where $U_n^k(a_n^k)$ denotes the *realized* payoff by agent $k$ at time $n$.

4) *Recursion*: Set $n \leftarrow n+1$ and go to Step 1.

In [23], it is proved that the conditional smooth fictitious play is $\epsilon$-*conditionally consistent*. The authors in [17] also prove that $\epsilon$-conditional consistency is equivalent to reaching $\epsilon$-CE when $|\mathcal{A}^k| = 2$ for all $k \in \mathcal{K}$.

The regret indicates how close the system is to the polytope of CE. Fig. 2(a) demonstrates how the distance to the CE polytope decreases with $n$ in static game $\mathcal{G}$ (i.e., $\rho = 0$ in (3)) when $\theta_n = 1$, $n \ge 1$. The distance diminishes for both Algorithms as the learning proceeds. However, comparing the slopes of the lines in Fig. 2(a) ($m_1 = -0.182$ for Algorithm 1 and $m_2 = -0.346$ for Algorithm 2), shows that exploiting information disclosed within neighborhoods provides an order of magnitude faster convergence to the polytope of $\epsilon$-CE.

For demonstration purposes, the Markov chain $\theta_n$ is assumed to change state from $\theta = 1$ to $\theta = 2$ at $n = 250$ in Fig. 2(b). Each point of the graphs is an average over 100 independent runs of the algorithms. Fig. 2(b) illustrates the sample paths of the average payoffs of agents. As shown, all algorithms respond to the jump of the Markov chain. However, Algorithm 1 outperforms the others as it approaches faster to the expected values in correlated equilibrium.

Fig. 3 illustrates the proportion of time that each tracking algorithm spends out of the polytope of $\epsilon$-CE (a measure of the expected time taken to adapt to the Markov chain jumps). Fig. 3(a) shows the proportion of the time spent out of $\epsilon$-CE when $\rho = 0.001$. As can be seen, both algorithms spend more time in $\mathcal{C}_\epsilon(\theta_n)$ as time goes by. However, Algorithm 1 exhibits superior performance for both small and large $n$. Fig. 3(b) also illustrates the time spent out of $\epsilon$-CE for different values of $\rho$ when $\mu = 0.001$. As expected, the performance of both algorithms degrade as the speed of the Markovian jumps increases. This is typical in stochastic approximation algorithms since dynamics of the Markov chain is not accounted for in the algorithm.

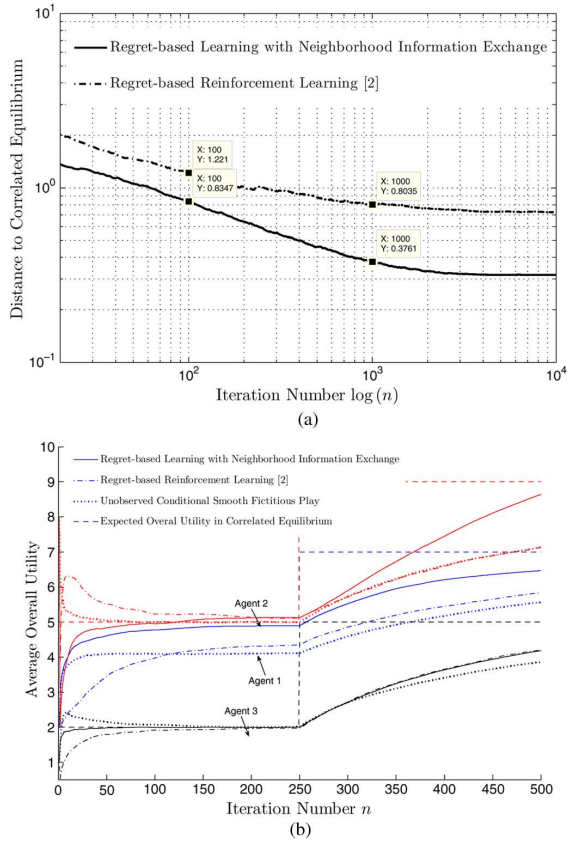


Fig. 2. (a) Distance to CE versus iteration number when $\rho = 0$ (static game), and $\theta_n = 1$, $n \ge 1$. (b) Average overall payoff in Markovian switching game ($\rho \neq 0$). The blue, red, and black lines illustrate the sample paths of average payoffs for agents 1, 2, and 3, respectively. The dashed lines represent the expected payoffs in CE.

## VI. Discussion

We considered a class of regime-switching noncooperative games modulated by a discrete time Markov chain where agents exchange action information over a graph. We presented a regret-based stochastic approximation algorithm with constant step-size that prescribes how individual agents update their randomized strategies over time. It was shown that, if all agents deploy this algorithm, their collective behavior properly tracks the time-varying convex polytope of $\epsilon$-CE for such games. Our analysis comprised two steps; First, we proved that the discrete time iterates of the stochastic approximation algorithm weakly converge to the set of global attractors of a regime-switching system of differential inclusions modulated by a continuous time Markov chain. We then proved that there exists equivalence between convergence to the global attractors of such limit dynamics in the regret space and convergence of agents' collective behavior to the polytope of $\epsilon$-CE. We demonstrated, via extensive simulations, that the proposed tracking algorithm benefits from the excess information shared within neighborhoods to ensure superior performance as compared to

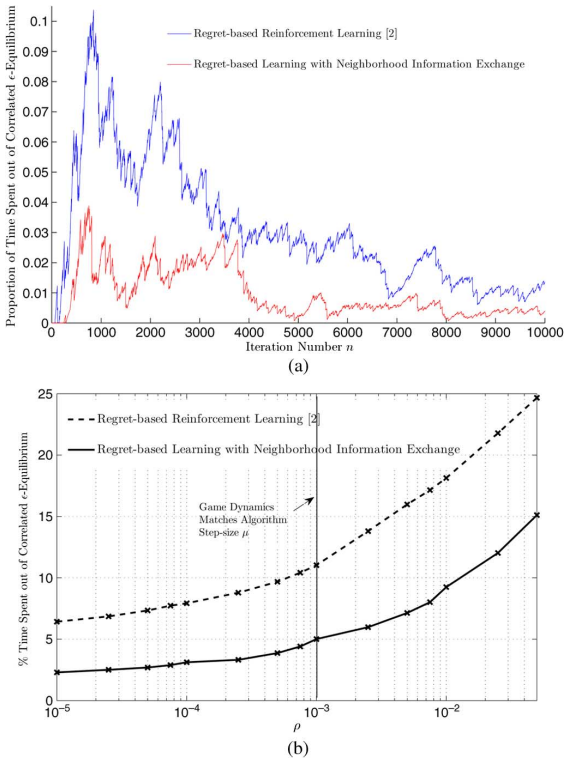


Fig. 3. Proportion of time spent out of $\epsilon$-CE: (a) versus $n$ for $\rho = 10^{-3}$, (b) versus $\rho$.

similar existing algorithms. Such a tracking algorithm can be employed in engineering applications such as smart sensor [8] and cognitive radio networks [10] to achieve coordination in a distributed fashion.

## Appendix A
## Proof of Theorem 4.1

Here, we consider a general setting where the limit system of the pair of processes $(X_n, Y_n)$ constitutes a system of interconnected differential inclusions. Theorem 4.1 can then be considered as a special case where the limit dynamics associated with $Y_n$ forms an ODE; hence, the results of this section will directly apply.

We first consider the pair $X_n, Y_n \in \mathbb{R}^r$ defined by

$$\begin{aligned} X_{n+1} &= X_n + \mu\left(U_1(\Phi_n, \Psi_n) - X_n\right) \\ Y_{n+1} &= Y_n + \mu\left(U_2(\Phi_n, \Xi_n) - Y_n\right) \\ \Phi_n, \Psi_n, \Xi_n &\in \mathcal{M} = \{1, \ldots, m\} \end{aligned} \tag{43}$$

without the Markovian switching process $\theta_n$. We will then extend the results to the switching case. For simplicity, assume that the initial data $X_0$ and $Y_0$ are independent of $\mu$. (With $\mu$-dependent initial data, we will need to assume that $(X_0^\mu, Y_0^\mu)$ converges weakly to $(X_0, Y_0)$.) Suppose that $\{X_n, Y_n, \Phi_{n-1}\}$ is a Markov chain such that the transition matrix is given by

$$\mathbb{P}(\Phi_n = j | \Phi_{n-1} = i, X_n = x, Y_n = y) = P_{ij}(x, y). \tag{44}$$

Write $P(x, y) = (P_{ij}(x, y)) \in \mathbb{R}^{m \times m}$. That is, the transition matrix is $(x, y)$ dependent. Such a case is referred to as state-dependent noise in [6, p. 185]. To proceed, we make the following assumptions:

C1) The transition matrix $P(\cdot, \cdot)$ satisfies:
  (a) it is continuous in both variables,
  (b) for each $(x, y)$, $P(x, y)$ is irreducible and aperiodic.

C2) $\{\Psi_n\}$ and $\{\Xi_n\}$ are sequences of bounded real-valued random variables that are independent of $\{\Phi_n\}$ such that for each $i = 1, 2$ and $j \in \mathcal{M}$, there is a set $S_i^j$

$$\begin{aligned} d\left(\frac{1}{n}\sum_{\ell=m}^{n+m-1} E_m U_1(j, \Psi_\ell), S_1^j\right) &\to 0 \text{ in probability} \\ d\left(\frac{1}{n}\sum_{\ell=m}^{n+m-1} E_m U_2(j, \Xi_\ell), S_2^j\right) &\to 0 \text{ in probability} \end{aligned} \tag{45}$$

where $d(\cdot, \cdot)$ denotes the usual distance function and $E_m$ denotes conditioning on the $\sigma$-algebra generated by $\{X_j, Y_j, \Phi_{j-1}, \Psi_{j-1}, \Xi_{j-1} : j \le m\}$.

C3) The sets

$$\begin{aligned} S_1(x, y) &= \sum_{j=1}^m \varphi_j(x, y) S_1^j \\ S_2(x, y) &= \sum_{j=1}^m \varphi_j(x, y) S_2^j \end{aligned} \tag{46}$$

are closed, convex, and upper semi-continuous (see [6, pp. 108-109]).

*Remark A.1:* We comment on the conditions briefly. C1) (a) indicates that the transition function is continuous in $(x, y)$; C1) (b) yields that the Markov chain is ergodic in the sense that there is a unique stationary distribution $\boldsymbol{\varphi}(x, y) = (\varphi_1(x, y), \ldots, \varphi_m(x, y))$ such that $[P(x, y)]^n \to \mathbb{1}\boldsymbol{\varphi}(x, y)$ a matrix with identical rows consisting of the stationary distribution as $n \to \infty$. The convergence in fact takes place exponentially fast. C2) assumes that $\{\Psi_n\}$ and $\{\Xi_n\}$ are real-valued bounded random variables. They are independent of $\Phi_n$ but may be dependent of each other. Here, the setup is more general. In the problem that we are interested, in fact, they are finite-valued processes taking values (without loss of generality) in $\mathcal{M}$ as well.

*Remark A.2:* We interpret the conditions in the setup of Algorithm 1. In C2), $S_1^j$ and $S_2^j$ are the convex hulls corresponding to the mean of local- and global-regrets associated with various sample paths of neighbors and non-neighbors action profiles, respectively, conditioned on agent $k$ picking $a_\ell^k = j$. More precisely, $S_1^j$ is a set-valued $A^k \times A^k$ matrix with all zero elements except the $j$th row such that the $ij$th elements, $i \in \mathcal{A}^k$, is given by $\{U_l^k(i, \boldsymbol{n}^k) - U_l^k(j, \boldsymbol{n}^k); \boldsymbol{n}^k \in \Delta\mathcal{A}^{\mathcal{N}_k}\}$. In C3), $S_1(x, y)$ and $S_2(x, y)$ are the convex hulls corresponding to the mean (no conditioning on $a_\ell^k = j$) of local- and global-regrets associated with various sample paths of neighbors and non-neighbors action profiles, respectively. More precisely, $S_1(x, y)$ is the convex combination of the sets $S_1^j$ such that the $ij$th element is given by $\{[U_l^k(i, \boldsymbol{n}^k) - U_l^k(j, \boldsymbol{n}^k)]\varphi_j(x, y); \boldsymbol{n}^k \in \Delta\mathcal{A}^{\mathcal{N}_k}\}$. $S_2^j$ and $S_2(x, y)$ can be defined similarly.

To proceed, define $X^\mu(t) = X_n$ and $Y^\mu(t) = Y_n$ on $t \in [n\mu, n\mu + \mu)$.

*Theorem A.1:* Assume that C1)–C3) hold. Then, $(X^\mu(\cdot), Y^\mu(\cdot))$ is tight in $D([0, \infty) : \mathbb{R}^{2r})$, and any weakly convergent sequence has limit $(X(\cdot), Y(\cdot))$ such that

$$\begin{aligned} \dot{X}(t) &\in S_1\left(X(t), Y(t)\right) - X(t) \\ \dot{Y}(t) &\in S_2\left(X(t), Y(t)\right) - Y(t) \end{aligned} \tag{47}$$

where $S_1(\cdot)$ and $S_2(\cdot)$ are defined in (46).

*Proof:* We first prove the tightness. Consider (43) for the sequence of $\mathbb{R}^{2r}$-valued vectors $\{(X_k, Y_k)^T\}$. Noting the boundedness of $\{U_1(\Phi_k, \Psi_k)\}$ and $\{U_2(\Phi_k, \Xi_k)\}$, and using Hölder's inequality and Gronwall's inequality, for any $0 < T_1 < \infty$, we obtain

$$\sup_{k \le T_1/\mu} \mathbb{E} \left| \begin{pmatrix} X_k \\ Y_k \end{pmatrix} \right|^2 < \infty \tag{48}$$

where in the above and hereafter $t/\mu$ is understood to be the integer parts of $t/\mu$ for each $t > 0$. Next, considering $(X^\mu(\cdot), Y^\mu(\cdot))$, the $\mathbb{R}^{2r}$-valued process, for any $t, s > 0$, $\delta > 0$, and $s < \delta$, it is fairly easy to verify that

$$\begin{pmatrix} X^\mu(t+s) \\ Y^\mu(t+s) \end{pmatrix} - \begin{pmatrix} X^\mu(t) \\ Y^\mu(t) \end{pmatrix} = \begin{pmatrix} \mu \sum_{k=t/\mu}^{(t+s)/\mu - 1} [U_1(\Phi_k, \Psi_k) - X_k] \\ \mu \sum_{k=t/\mu}^{(t+s)/\mu - 1} [U_2(\Xi_k, \Psi_k) - Y_k] \end{pmatrix}. \tag{49}$$

As a result,

$$\lim_{\delta \to 0} \limsup_{\mu \to 0} \left\{ \mathbb{E} \left[ \sup_{0 \le s \le \delta} \mathbb{E}_t^\mu \left| \begin{pmatrix} X^\mu(t+s) \\ Y^\mu(t+s) \end{pmatrix} - \begin{pmatrix} X^\mu(t) \\ Y^\mu(t) \end{pmatrix} \right|^2 \right] \right\} = 0. \tag{50}$$

Thus, $(X^\mu(\cdot), Y^\mu(\cdot))$ is tight in $D([0, \infty) : \mathbb{R}^{2r})$; see [24, Th. 3, p. 47] or [6, Ch. 7].

By Prohorov theorem [6], one can extract a convergent subsequence. For notational simplicity, we still denote the subsequence by $(X^\mu(\cdot), Y^\mu(\cdot))$ with limit $(X(\cdot), Y(\cdot))$. Thus, $(X^\mu(\cdot), Y^\mu(\cdot))$ converges weakly to $(X(\cdot), Y(\cdot))$. By Skorohod representation theorem [6], with a slight abuse of notation, one can assume $(X^\mu(\cdot), Y^\mu(\cdot)) \to (X(\cdot), Y(\cdot))$ in the sense of w.p.1 and the convergence is uniform on any finite interval. Using the martingale averaging methods, we need to characterize the limit process. Normally, one uses a smooth function $f(\cdot)$ with compact support to carry out the analysis. Here, for simplicity, we suppress the dependence of $f(\cdot)$ and proceed directly; see [25] for a similar argument. Choose a sequence of integers $\{n_\mu\}$ such that $n_\mu \to \infty$ as $\mu \to 0$, but $\delta_\mu = \mu n_\mu \to 0$. Let us focus on the recursion of $X_n$; the recursion of $Y_n$ can be handled similarly. Note

$$\begin{aligned} & X^\mu(t+s) - X^\mu(t) \\ &= \sum_{\ell : \ell\delta_\mu = t}^{t+s} \delta_\mu \sum_{j=1}^{m} \frac{1}{n_\mu} \sum_{k=\ell n_\mu}^{\ell n_\mu + n_\mu - 1} U_1(j, \Psi_k) I_{\{\Phi_k = j\}} \\ &\quad - \sum_{\ell : \ell\delta_\mu = t}^{t+s} \delta_\mu \frac{1}{n_\mu} \sum_{k=\ell n_\mu}^{\ell n_\mu + n_\mu - 1} X_k \end{aligned} \tag{51}$$

where $\sum_{\ell : \ell\delta_\mu = t}^{t+s}$ denotes the sum over $\ell$ in the range $t \le \ell\delta_\mu \le t + s$.

We shall use the techniques in [6, Ch. 8] to prove the desired result. Let $h(\cdot)$ be any bounded and continuous function, $t, s > 0$, $\kappa_0$ be an arbitrary positive integer, and $t_\iota \le t$ for all $\iota \le \kappa_0$. It is readily seen that by the weak convergence and the Skorohod representation (without changing notation), as $\mu \to 0$,

$$\begin{aligned} & \mathbb{E} h\left(X^\mu(t_\iota), Y^\mu(t_\iota) : \iota \le \kappa_0\right) \left[X^\mu(t+s) - X^\mu(t)\right] \\ & \quad \to \mathbb{E} h\left(X(t_\iota), Y(t_\iota) : \iota \le \kappa_0\right) \left[X(t+s) - X(t)\right]. \end{aligned} \tag{52}$$

By using the technique of stochastic approximation (see, e.g., [6, Ch. 8]), we also have

$$\begin{aligned} & \mathbb{E} h\left(X^\mu(t_\iota), Y^\mu(t_\iota) : \iota \le \kappa_0\right) \sum_{\ell : \ell\delta_\mu = t}^{t+s} \delta_\mu \frac{1}{n_\mu} \sum_{k=\ell n_\mu}^{\ell n_\mu + n_\mu - 1} X_k \\ & \to \mathbb{E} h\left(X(t_\iota), Y(t_\iota) : \iota \le \kappa_0\right) \left[\int_t^{t+s} X(u) du\right] \text{ as } \mu \to 0. \end{aligned} \tag{53}$$

Moreover, by the independence of $\Phi_n$ and $\Psi_n$, we have

$$\begin{aligned} & \lim_{\mu \to 0} \mathbb{E} h(X^\mu(t_\iota), Y^\mu(t_\iota) : \iota \le \kappa_0) \\ & \quad \times \left[ \sum_{\ell : \ell\delta_\mu = t}^{t+s} \delta_\mu \sum_{j=1}^{m} \frac{1}{n_\mu} \sum_{k=\ell n_\mu}^{\ell n_\mu + n_\mu - 1} U_1(j, \Psi_k) I_{\{\Phi_k = j\}} \right] \\ & = \lim_{\mu \to 0} \mathbb{E} h(X^\mu(t_\iota), Y^\mu(t_\iota) : \iota \le \kappa_0) \\ & \quad \times \left[ \sum_{\ell : \ell\delta_\mu = t}^{t+s} \delta_\mu \sum_{j=1}^{m} \frac{1}{n_\mu} \sum_{k=\ell n_\mu}^{\ell n_\mu + n_\mu - 1} E_{\ell n_\mu} U_1(j, \Psi_k) E_{\ell n_\mu} I_{\{\Phi_k = j\}} \right] \\ & = \lim_{\mu \to 0} \mathbb{E} h(X^\mu(t_\iota), Y^\mu(t_\iota) : \iota \le \kappa_0) \\ & \quad \times \Bigg[ \sum_{\ell : \ell\delta_\mu = t}^{t+s} \delta_\mu \sum_{j=1}^{m} \frac{1}{n_\mu} \sum_{k=\ell n_\mu}^{\ell n_\mu + n_\mu - 1} \sum_{i_0=1}^{m} I_{\{\Phi_{\ell n_\mu} = i_0\}} \\ & \qquad \times \Big[ E_{\ell n_\mu} U_1(j, \Psi_k) \varphi_j\big(X_{\ell n_\mu}, Y_{\ell n_\mu}\big) + E_{\ell n_\mu} U_1(j, \Psi_k) \\ & \qquad\quad \times \Big[ P_{i_0 j}^{\{k - \ell n_\mu\}}\big(X_{\ell n_\mu}, Y_{\ell n_\mu}\big) - \varphi_j\left(X_{\ell n_\mu}, Y_{\ell n_\mu}\right) \Big] \Big] \Bigg]. \end{aligned} \tag{54}$$

In (54), $P_{i_0 j}^{\{k - \ell n_\mu\}}(x, y)$ denotes the $(k - \ell n_\mu)$-step transition probability. Except the indicator function on the left-hand side of (55), the rest of the terms in the summand are independent of $i_0$; therefore,

$$\begin{aligned} & \sum_{i_0=1}^{m} I_{\{\Phi_{\ell n_\mu} = i_0\}} E_{\ell n_\mu} U_1(j, \Psi_k) \varphi_j\left(X_{\ell n_\mu}, Y_{\ell n_\mu}\right) \\ & \qquad = E_{\ell n_\mu} U_1(j, \Psi_k) \varphi_j\left(X_{\ell n_\mu}, Y_{\ell n_\mu}\right). \end{aligned} \tag{55}$$

Using C1), regardless of the choice of $i_0$, as $\mu \to 0$ and $k - \ell n_\mu \to \infty$, for each fixed $(x, y)$, $[P_{i_0 j}^{\{k - \ell n_\mu\}}(x, y) - \varphi_j(x, y)] \to 0$ exponentially fast. Thus,

$$\begin{aligned} & \mathbb{E} h\left(X^\mu(t_\iota), Y^\mu(t_\iota) : \iota \le \kappa_0\right) \\ & \quad \times \Bigg[ \sum_{\ell : \ell\delta_\mu = t}^{t+s} \delta_\mu \sum_{j=1}^{m} \frac{1}{n_\mu} \sum_{k=\ell n_\mu}^{\ell n_\mu + n_\mu - 1} \sum_{i_0=1}^{m} I_{\{\Phi_{\ell n_\mu} = i_0\}} E_{\ell n_\mu} U_1(j, \Psi_k) \\ & \qquad \times \Big[ P_{i_0 j}^{\{k - \ell n_\mu\}}\left(X_{\ell n_\mu}, Y_{\ell n_\mu}\right) - \varphi_j\left(X_{\ell n_\mu}, Y_{\ell n_\mu}\right) \Big] \Bigg] \\ & \quad \to 0 \text{ in probability as } \mu \to 0. \end{aligned} \tag{56}$$

Using C2),

$$\lim_{\mu \to 0} d\left( \frac{1}{n_\mu} \sum_{k=\ell n_\mu}^{\ell n_\mu + n_\mu - 1} E_{\ell n_\mu} U_1(j, \Psi_k), S_1^j \right) = 0 \tag{57}$$

where $d(\cdot,\cdot)$ is the distance function. Combining the estimates obtained thus far, we obtain

$$\dot{X}(t) \in \sum_{j=1}^{m} \varphi_j\left(X(t), Y(t)\right) S_1^j - X(t) \tag{58}$$

as desired. The limit of $Y^\mu(\cdot)$ can be derived similarly. The details are omitted. ∎

Next, we consider the regime-switching case. The algorithm of interest [see (18)] becomes

$$\begin{aligned} X_{n+1} &= X_n + \mu\left(U_1(\Phi_n, \Psi_n, \theta_n) - X_n\right) \\ Y_{n+1} &= Y_n + \mu\left(U_2(\Phi_n, \Xi_n, \theta_n) - Y_n\right) \\ \Phi_n, \Psi_n, \Xi_n &\in \mathcal{M} = \{1, \ldots, m\} \\ \theta_n &\in \mathcal{M}_\theta = \{1, \ldots, \Theta\} \end{aligned} \tag{59}$$

where $\theta_n$ is a discrete-time Markov chain. Define $X^\mu(\cdot)$ and $Y^\mu(\cdot)$ as before and $\theta^\mu(t) = \theta_n$ for $t \in [n\mu, n\mu + \mu)$. Assume that the one-step transition probability matrix for $\theta_n$ is given by $I + \mu Q$, where $Q$ is the generator of a continuous-time Markov chain. Now, $\theta^\mu(\cdot)$ becomes part of the state.

*Theorem A.2:* Assume that C1)–C3) hold with the modification that in C2), for each $i_0 \in \mathcal{M}_\theta$, as $\mu \to 0$

$$\begin{aligned} &d\left(\frac{1}{n}\sum_{\ell=m}^{n+m-1} E_m U_1(j, \Psi_\ell, i_0), S_1^j(i_0)\right) \to 0 \text{ in probability} \\ &d\left(\frac{1}{n}\sum_{\ell=m}^{n+m-1} E_m U_2(j, \Xi_\ell, i_0), S_2^j(i_0)\right) \to 0 \text{ in probability.} \end{aligned} \tag{60}$$

Assume further $\theta_n$ is independent of $\Phi_n$, $\Psi_n$, and $\Xi_n$. Then $((X^\mu(\cdot), Y^\mu(\cdot)), \theta^\mu(\cdot))$ is tight in $D([0,\infty) : \mathbb{R}^{2r} \times \mathcal{M}_\theta)$ and any weakly convergent sequence has limit $((X(\cdot), Y(\cdot)), \theta(\cdot))$ such that

$$\begin{aligned} \dot{X}(t) &\in S_1\left(X(t), Y(t), \theta(t)\right) - X(t) \\ \dot{Y}(t) &\in S_2\left(X(t), Y(t), \theta(t)\right) - Y(t) \end{aligned} \tag{61}$$

where

$$S_i(x, y, \theta) = \sum_{j=1}^{m} \varphi_j(x, y) S_i^j(\theta) \text{ for each } i = 1, 2 \tag{62}$$

and $\theta(\cdot)$ is a continuous-time Markov chain with generator $Q$ and state space $\mathcal{M}_\theta$.

We shall only outline the steps involved in the proof and point out the main new feature. The proof involves three steps:

Step 1: [Tightness of $((X^\mu\cdot, Y^\mu\cdot), \theta^\mu\cdot)$ in $D([0,\infty) : \mathbb{R}^{2r} \times \mathcal{M}_\theta)$.] This is similar to the proof of Theorem A.1. We merely note that by [26, Prop. 4.4], $\theta^\mu(\cdot)$ is tight and $\theta^\mu(\cdot) \Rightarrow \theta(\cdot)$ such that $\theta(\cdot)$ is a continuous-time Markov chain with generator $Q$.

Step 2.1: [Characterization of the limit process part (i).] For each $i_0 \in \mathcal{M}_\theta$, for any $f(\cdot,\cdot,i_0) : \mathbb{R}^r \times \mathbb{R}^r \mapsto \mathbb{R}$ with $f(\cdot,\cdot,i_0) \in C_0^1$ ($C^1$ function with compact support). Choose $t$, $s$, and $n_\mu$ as in the proof of Theorem A.1. Then,

$$\begin{aligned} &f\left(X^\mu(t+s), Y^\mu(t+s), \theta^\mu(t+s)\right) \\ &\quad - f\left(X^\mu(t), Y^\mu(t), \theta^\mu(t)\right) \\ &\quad = \sum_{\ell:\ell\delta_\mu=t}^{t+s} \left[f\left(X_{\ell n_\mu + n_\mu}, Y_{\ell n_\mu + n_\mu}, \theta_{\ell n_\mu + n_\mu}\right)\right. \\ &\qquad\qquad \left. - f\left(X_{\ell n_\mu}, Y_{\ell n_\mu}, \theta_{\ell n_\mu}\right)\right] \\ &\quad = \sum_{\ell:\ell\delta_\mu=t}^{t+s} \left[f\left(X_{\ell n_\mu + n_\mu}, Y_{\ell n_\mu + n_\mu}, \theta_{\ell n_\mu + n_\mu}\right)\right. \\ &\qquad\qquad \left. - f\left(X_{\ell n_\mu}, Y_{\ell n_\mu}, \theta_{\ell n_\mu + n_\mu}\right)\right] \\ &\qquad + \sum_{\ell:\ell\delta_\mu=t}^{t+s} \left[f\left(X_{\ell n_\mu}, Y_{\ell n_\mu}, \theta_{\ell n_\mu + n_\mu}\right)\right. \\ &\qquad\qquad \left. - f\left(X_{\ell n_\mu}, Y_{\ell n_\mu}, \theta_{\ell n_\mu}\right)\right]. \end{aligned} \tag{63}$$

First, we consider the last term in (63). Use $\kappa_0$, $\iota$, $h(\cdot)$, and $t_\iota$ as in the proof of Theorem A.1. Then, (see (64), shown at the bottom of the page), where

$$Qf(x, y, \cdot)(i) = \sum_{j \in \mathcal{M}_\theta} q_{ij} f(x, y, j). \tag{65}$$

Step 2.2: [Characterization of the limit process part (ii).] As for the next to the last term in (63), we can show that [see (66), shown at the bottom of the next page], where $f_x$ and $f_y$ denote the derivatives with respect to $x$ and $y$, respectively, and $f'$ denotes the transpose of $f$. To work out the limit

$$\begin{aligned} &\lim_{\mu\to 0} \mathbb{E} h\left(X^\mu(t_\iota), Y^\mu(t_\iota), \theta^\mu(t_\iota); \iota \le \kappa_0\right) \\ &\qquad \times \left[\sum_{\ell:\ell\delta_\mu=t}^{t+s} \delta_\mu \sum_{j=1}^{m} \frac{1}{n_\mu} \sum_{k=\ell n_\mu}^{\ell n_\mu + n_\mu - 1} U_1(j, \Psi_k, \theta_k) I_{\{\Phi_k = j\}}\right] \\ &\qquad = \lim_{\mu\to 0} \mathbb{E} h\left(X^\mu(t_\iota), Y^\mu(t_\iota), \theta^\mu(t_\iota); \iota \le \kappa_0\right) \\ &\qquad \times \left[\sum_{\ell:\ell\delta_\mu=t}^{t+s} \delta_\mu \sum_{i_1=1}^{\Theta} \sum_{i_0=1}^{\Theta} \sum_{j=1}^{m} \frac{1}{n_\mu} \sum_{k=\ell n_\mu}^{\ell n_\mu + n_\mu - 1} \mathbb{E}_{\ell n_\mu} U_1(j, \Psi_k, i_1)\right. \\ &\qquad\qquad \left. \times \mathbb{E}_{\ell n_\mu}\left[I_{\{\theta_k = i_1\}} | I_{\left\{\theta_{\ell n_\mu} = i_0\right\}}\right] \mathbb{E}_{\ell n_\mu} I_{\{\Phi_k = j\}}\right]. \end{aligned} \tag{67}$$

$$\begin{aligned} &\lim_{\mu\to 0} \mathbb{E} h\left(X^\mu(t_\iota), Y^\mu(t_\iota), \theta^\mu(t_\iota) : \iota \le \kappa_0\right) \left[\sum_{\ell:\ell\delta_\mu=t}^{t+s} \left[f\left(X_{\ell n_\mu}, Y_{\ell n_\mu}, \theta_{\ell n_\mu + \mu}\right) - f\left(X_{\ell n_\mu}, Y_{\ell n_\mu}, \theta_{\ell n_\mu}\right)\right]\right] \\ &\quad = \lim_{\mu\to 0} \mathbb{E} h\left(X^\mu(t_\iota), Y^\mu(t_\iota), \theta^\mu(t_\iota) : \iota \le \kappa_0\right) \left[\sum_{\ell:\ell\delta_\mu=t}^{t+s} \sum_{i_0=1}^{\Theta} \sum_{j_0=1}^{\Theta} \sum_{k=\ell n_\mu}^{\ell n_\mu + n_\mu - 1} \left[f\left(X_{\ell n_\mu}, Y_{\ell n_\mu}, j_0\right)\right.\right. \\ &\qquad\qquad \left.\left. \times P(\theta_{k+1} = j_0 | \theta_k = i_0) - f\left(X_{\ell n_\mu}, Y_{\ell n_\mu}, i_0\right)\right] I_{\{\theta_k = i_0\}}\right] \\ &\quad = \mathbb{E} h\left(X(t_\iota), Y(t_\iota), \theta(t_\iota) : \iota \le \kappa_0\right) \left[\int_t^{t+s} Qf\left(X(u), Y(u), \theta(u)\right) du\right] \end{aligned} \tag{64}$$

We will concentrate on the term involving the Markov chain $\theta_k$. Note that for large $k$ with $\ell n_\mu \le k \le \ell n_\mu + n_\mu$ and $k - \ell n_\mu \to \infty$, by [26, Prop. 4.4], for some $\widehat{k}_0 > 0$

$$\begin{aligned}(I+\mu Q)^{k-\ell n_\mu} &= Z\left((k-\ell n_\mu)\mu\right) \\ &\quad + O(\mu + \exp\left(-\widehat{k}_0(k-\ell n_\mu)\right) \quad (68)\\ \frac{dZ(t)}{dt} &= Z(t)Q,\ Z(0) = I.\end{aligned}$$

For $\ell n_\mu \le k \le \ell n_\mu + n_\mu$, letting $\mu \ell n_\mu \to u$ yields that $(k - \ell n_\mu)\mu \to 0$ as $\mu \to 0$. For such $k$, $Z((k-\ell n_\mu)\mu) \to I$. By the boundedness of $U_1(\cdot)$, it then follows that, as $\mu \to 0$

$$\begin{aligned}&\frac{1}{n_\mu}\sum_{k=\ell n_\mu}^{\ell n_\mu + n_\mu - 1} \left|\mathbb{E}_{\ell n_\mu} U_1(j,\Psi_k,i_1)\right| \left|\mathbb{E}_{\ell n_\mu} I_{\{\Phi_k = j\}}\right| \\ &\quad \times \left|\mathbb{E}_{\ell n_\mu}\left[I_{\{\theta_k = i_1\}} | I_{\left\{\theta_{\ell n_\mu} = i_0\right\}}\right] - I_{\left\{\theta_{\ell n_\mu} = i_0\right\}}\right| \to 0. \quad (69)\end{aligned}$$

Thus, the limit in the first line of (67) can be replaced by (see the second equation at the bottom of page). Note that $\theta_{\ell n_\mu} = \theta^\mu(\mu \ell n_\mu)$. By the weak convergence of $\theta^\mu(\cdot)$ to $\theta(\cdot)$, the Skorohod representation, and using $\mu \ell n_\mu \to u$, we can proceed to show that

$$\begin{aligned}&d\Bigg(\frac{1}{n_\mu}\sum_{i_0 \in \mathcal{M}_\theta}\sum_{k=\ell n_\mu}^{\ell n_\mu + n_\mu - 1} E_{\ell n_\mu} U_1(j,\Psi_k,i_0) \\ &\quad I_{\{\theta^\mu(\mu \ell n_\mu) = i_0\}}, S_1^j(i_0) I_{\{\theta(u) = i_0\}}\Bigg) \to 0 \text{ in probability.} \quad (70)\end{aligned}$$

We can also obtain (53). Combining the estimates obtained thus far yields

$$\dot{X}(t) \in \sum_{j=1}^m \varphi_j\left(X(t), Y(t)\right) S_1^j\left(\theta(t)\right) - X(t). \quad (71)$$

The limit of $Y^\mu(\cdot)$ can be derived similarly.

Step 3: Combining Steps 1 and 2, the desired result follows.

## Appendix B
## Proof of Theorem 4.3

Consider the regime-switching limit system (25). We start by showing that such continuous time process is asymptotically stable for each subsystem $\theta \in \mathcal{M}_\theta$. Define the Lyapunov function:

$$V(\boldsymbol{\alpha}^k, \boldsymbol{\beta}^k) = \left[d(\boldsymbol{\alpha}^k + \boldsymbol{\beta}^k, \mathbb{R}^-)\right]^2 \quad (72)$$

$$\begin{aligned}&\lim_{\mu\to 0}\mathbb{E}h\left(X^\mu(t_\iota), Y^\mu(t_\iota), \theta^\mu(t_\iota); \iota \le \kappa_0\right)\left[\sum_{\ell:\ell\delta_\mu = t}^{t+s}\left[f\left(X_{\ell n_\mu + n_\mu}, Y_{\ell n_\mu + n_\mu}, \theta_{\ell n_\mu + n_\mu}\right) - f\left(X_{\ell n_\mu}, Y_{\ell n_\mu}, \theta_{\ell n_\mu + n_\mu}\right)\right]\right] \\ &= \lim_{\mu\to 0}\mathbb{E}h\left(X^\mu(t_\iota), Y^\mu(t_\iota), \theta^\mu(t_\iota); \iota \le \kappa_0\right)\left[\sum_{\ell:\ell\delta_\mu = t}^{t+s}\left[f\left(X_{\ell n_\mu + n_\mu}, Y_{\ell n_\mu + n_\mu}, \theta_{\ell n_\mu}\right) - f\left(X_{\ell n_\mu}, Y_{\ell n_\mu}, \theta_{\ell n_\mu}\right)\right]\right] \\ &= \lim_{\mu\to 0}\mathbb{E}h\left(X^\mu(t_\iota), Y^\mu(t_\iota), \theta^\mu(t_\iota); \iota \le \kappa_0\right)\left[\sum_{\ell:\ell\delta_\mu = t}^{t+s} f_x'\left(X_{\ell n_\mu}, Y_{\ell n_\mu}, \theta_{\ell n_\mu}\right)\left[X_{\ell n_\mu + n_\mu} - X_{\ell n_\mu}\right]\right. \\ &\qquad \left. + f_y'\left(X_{\ell n_\mu}, Y_{\ell n_\mu}, \theta_{\ell n_\mu}\right)\left[Y_{\ell n_\mu + n_\mu} - Y_{\ell n_\mu}\right]\right] \\ &= \lim_{\mu\to 0}\mathbb{E}h\left(X^\mu(t_\iota), Y^\mu(t_\iota), \theta^\mu(t_\iota); \iota \le \kappa_0\right)\left\{\sum_{\ell:\ell\delta_\mu = t}^{t+s}\delta_\mu f_x'\left(X_{\ell n_\mu}, Y_{\ell n_\mu}, \theta_{\ell n_\mu}\right)\right. \\ &\qquad \times\left[\frac{1}{n_\mu}\sum_{j=1}^m\sum_{k=\ell n_\mu}^{\ell n_\mu + n_\mu - 1} U_1(j,\Psi_k,\theta_k) I_{\{\Phi_k = j\}} - \frac{1}{n_\mu}\sum_{k=\ell n_\mu}^{\ell n_\mu + n_\mu - 1} X_k\right] \\ &\qquad + \sum_{\ell:\ell\delta_\mu = t}^{t+s}\delta_\mu f_y'\left(X_{\ell n_\mu}, Y_{\ell n_\mu}, \theta_{\ell n_\mu}\right) \\ &\qquad \left.\times\left[\frac{1}{n_\mu}\sum_{j=1}^m\sum_{k=\ell n_\mu}^{\ell n_\mu + n_\mu - 1} U_2(j,\Xi_k,\theta_k) I_{\{\Phi_k = j\}} - \frac{1}{n_\mu}\sum_{k=\ell n_\mu}^{\ell n_\mu + n_\mu - 1} Y_k\right]\right\} \quad (66)\end{aligned}$$

$$\lim_{\mu\to 0}\mathbb{E}h\left(X^\mu(t_\iota), Y^\mu(t_\iota), \theta^\mu(t_\iota); \iota \le \kappa_0\right)\left[\sum_{\ell:\ell\delta_\mu = t}^{t+s}\delta_\mu\sum_{i_0\in\mathcal{M}_\theta}\sum_{j=1}^m\frac{1}{n_\mu}\sum_{k=\ell n_\mu}^{\ell n_\mu + n_\mu - 1} E_{\ell n_\mu} U_1(j,\Psi_k,i_0)\varphi_j\left(X_{\ell n_\mu}, Y_{\ell n_\mu}\right) I_{\left\{\theta_{\ell n_\mu} = i_0\right\}}\right]$$

where $\mathbb{R}^-$ denotes the nonpositive orthant in $\mathbb{R}^r$ space. Then,

$$\begin{aligned}\frac{d}{dt}V(\boldsymbol{\alpha}^k,\boldsymbol{\beta}^k)&=\left\langle\nabla_{\boldsymbol{\alpha}^k}V(\boldsymbol{\alpha}^k,\boldsymbol{\beta}^k),\frac{d}{dt}\boldsymbol{\alpha}^k\right\rangle_F\\&\quad+\left\langle\nabla_{\boldsymbol{\beta}^k}V(\boldsymbol{\alpha}^k,\boldsymbol{\beta}^k),\frac{d}{dt}\boldsymbol{\beta}^k\right\rangle_F\\&=2\sum_{i,j\in\mathcal{A}^k}\left|\alpha^k(i,j)+\beta^k(i,j)\right|^+\\&\quad\times\left[\frac{d}{dt}\alpha^k(i,j)+\frac{d}{dt}\beta^k(i,j)\right].\end{aligned}\tag{73}$$

Here, $\langle\cdot,\cdot\rangle_F$ denotes the *Frobenius inner product*: $\langle A,B\rangle_F=\sum_i\sum_j a_{ij}b_{ij}$, where $A$ and $B$ are matrices of the same size. Now, substituting (25)–(27) into (73) yields: (see (74), shown at the bottom of the page).

Subsequently, applying $\sigma^k(i)=(1-\delta)\varphi^k(i)+(\delta/A^k)$ in (74) results in (75), shown at the bottom of the page. Now, the following lemma proves that the first term in the right-hand side of (75) equals zero.

*Lemma B.1:* Let $\boldsymbol{\varphi}^k$ denote the vector of stationary distribution of

$$\psi^k(i)=\begin{cases}\frac{1}{\xi^k}\left|\alpha^k(a^k,i)+\beta^k(a^k,i)\right|^+, & i\neq a^k\\ 1-\sum_{\substack{j\in\mathcal{A}^k\\ j\neq i}}\psi^k(j), & i=a^k\end{cases}\tag{76}$$

corresponding to $\delta=0$ in (9). Then,

$$\begin{aligned}&\sum_{i,j\in\mathcal{A}^k}\varphi^k(i)\left|\alpha^k(i,j)+\beta^k(i,j)\right|^+\\&\times\left[U_l^k(j,\boldsymbol{s}^k;\theta)+U_{g,t}^k(j;\theta)-U_l^k(i,\boldsymbol{s}^k;\theta)-U_{g,t}^k(i;\theta)\right]=0.\end{aligned}\tag{77}$$

*Proof:* Given that $\boldsymbol{\varphi}^k$ is the stationary distribution of (76), for all $i\in\mathcal{A}^k$

$$\begin{aligned}\varphi^k(i)&=\varphi^k(i)\left[1-\sum_{j\in\mathcal{A}^k\setminus\{i\}}\frac{1}{\xi^k}\left|\alpha^k(j,i)+\beta^k(j,i)\right|^+\right]\\&\quad+\sum_{j\in\mathcal{A}^k\setminus\{i\}}\varphi^k(j)\frac{1}{\xi^k}\left|\alpha^k(i,j)+\beta^k(i,j)\right|^+.\end{aligned}\tag{78}$$

Simplifying (78) yields

$$\begin{aligned}&\sum_{j\in\mathcal{A}^k\setminus\{i\}}\varphi^k(j)\cdot\left|\alpha^k(j,i)+\beta^k(j,i)\right|^+\\&\quad=\varphi^k(i)\left[\sum_{j\in\mathcal{A}^k\setminus\{i\}}\left|\alpha^k(i,j)+\beta^k(i,j)\right|^+\right],\quad\forall i\in\mathcal{A}^k.\end{aligned}\tag{79}$$

We now proceed to the prove (77):

$$\begin{aligned}&\sum_{i,j\in\mathcal{A}^k}\varphi^k(i)\left|\alpha^k(i,j)+\beta^k(i,j)\right|^+\\&\quad\times\Big[\left[U_l^k(j,\boldsymbol{s}^k;\theta)+U_{g,t}^k(j;\theta)\right]\\&\qquad-\left[U_l^k(i,\boldsymbol{s}^k;\theta)+U_{g,t}^k(i;\theta)\right]\Big]\\&=\sum_{i,j\in\mathcal{A}^k}\varphi^k(i)\left|\alpha^k(i,j)+\beta^k(i,j)\right|^+\\&\qquad\times\left[U_l^k(j,\boldsymbol{s}^k;\theta)+U_{g,t}^k(j;\theta)\right]\\&\quad-\sum_{i,j\in\mathcal{A}^k}\varphi^k(i)\left|\alpha^k(i,j)+\beta^k(i,j\right|^+\\&\qquad\times\left[U_l^k(i,\boldsymbol{s}^k;\theta)+U_{g,t}^k(i;\theta)\right].\end{aligned}$$

Subsequently, applying $\sum_i\sum_j a_{ij}=\sum_j\sum_i a_{ji}$ into the first term in the right-hand side yields

$$\begin{aligned}&\sum_{j\in\mathcal{A}^k}\sum_{i\in\mathcal{A}^k}\varphi^k(j)\left|\alpha^k(j,i)+\beta^k(j,i)\right|^+\\&\qquad\times\left[U_l^k(i,\boldsymbol{s}^k;\theta)+U_{g,t}^k(i;\theta)\right]\\&\quad-\sum_{i,j\in\mathcal{A}^k}\varphi^k(i)\left|\alpha^k(i,j)+\beta^k(i,j)\right|^+\\&\qquad\times\left[U_l^k(i,\boldsymbol{s}^k;\theta)+U_{g,t}^k(i;\theta)\right]\\&=\sum_{i\in\mathcal{A}^k}\left[U_l^k(i,\boldsymbol{s}^k;\theta)+U_{g,t}^k(i;\theta)\right]\\&\qquad\times\left[\sum_{j\in\mathcal{A}^k}\varphi^k(j)\left|\alpha^k(j,i)+\beta^k(j,i)\right|^+\right.\\&\qquad\left.-\varphi^k(i)\sum_{j\in\mathcal{A}^k}\left|\alpha^k(i,j)+\beta^k(i,j)\right|^+\right].\end{aligned}\tag{80}$$

$$\begin{aligned}\frac{d}{dt}V(\boldsymbol{\alpha}^k,\boldsymbol{\beta}^k)&=2\sum_{i,j\in\mathcal{A}^k}\left|\alpha^k(i,j)+\beta^k(i,j)\right|^+\Big[\left[U_l^k(j,\boldsymbol{s}^k;\theta)-U_l^k(i,\boldsymbol{s}^k;\theta)\right]\sigma^k(i)-\alpha^k(i,j)+\left[U_{g,t}^k(j;\theta)-U_{g,t}^k(i;\theta)\right]\sigma^k(i)-\beta^k(i,j)\Big]\\&=2\sum_{i,j\in\mathcal{A}^k}\left|\alpha^k(i,j)+\beta^k(i,j)\right|^+\Big[\left[U_l^k\left(j,\boldsymbol{s}^k;\theta(t)\right)+U_{g,t}^k(j;\theta)-U_l^k\left(i,\boldsymbol{s}^k;\theta\right)-U_{g,t}^k(i;\theta)\right]\sigma^k(i)-\left[\alpha^k(i,j)+\beta^k(i,j)\right]\Big]\end{aligned}\tag{74}$$

$$\begin{aligned}\frac{d}{dt}V(\boldsymbol{\alpha}^k,\boldsymbol{\beta}^k)=&2(1-\delta)\left[\sum_{i,j\in\mathcal{A}^k}\varphi^k(i)\left|\alpha^k(i,j)+\beta^k(i,j)\right|^+\Big[\left[U_l^k\left(j,\boldsymbol{s}^k;\theta\right)+U_{g,t}^k(j;\theta)\right]-\left[U_l^k(i,\boldsymbol{s}^k;\theta)+U_{g,t}^k(i;\theta)\right]\Big]\right]\\&+\frac{2\delta}{A^k}\left[\sum_{i,j\in\mathcal{A}^k}\left|\alpha^k(i,j)+\beta^k(i,j)\right|^+\Big[\left[U_l^k(j,\boldsymbol{s}^k;\theta)+U_{g,t}^k(j;\theta)\right]-\left[U_l^k(i,\boldsymbol{s}^k;\theta)+U_{g,t}^k(i;\theta)\right]\Big]\right]\\&-2\sum_{i,j\in\mathcal{A}^k}\left|\alpha^k(i,j)+\beta^k(i,j)\right|^+\left[\alpha^k(i,j)+\beta^k(i,j)\right]\end{aligned}\tag{75}$$

Finally, knowing $|\alpha^k(i,i)+\beta^k(i,i)|^+ = 0$, $i \in \mathcal{A}^k$, and substituting (79) into (80) completes the proof. ■

By Lemma B.1, the first term in (75) is zero. Since the range of $U_l^k(\cdot, \boldsymbol{s}^k;\theta)+U_{g,t}^k(\cdot;\theta)$ is bounded for each $\theta \in \mathcal{M}_\theta$, we have (see (81), shown at the bottom of the page) for some constant $C(\theta)$. Then, since

$$\sum_{i,j\in\mathcal{A}^k} \left|\alpha^k(i,j)+\beta^k(i,j)\right|^+ \left[\alpha^k(i,j)+\beta^k(i,j)\right] = \sum_{i,j\in\mathcal{A}^k} \left[\left|\alpha^k(i,j)+\beta^k(i,j)\right|^+\right]^2 \quad (82)$$

one can rewrite (81) as

$$\frac{d}{dt}V(\boldsymbol{\alpha}^k,\boldsymbol{\beta}^k) \le \frac{2C(\theta)\delta}{A^k}\sum_{i\in\mathcal{A}^k}\sum_{j\in\mathcal{A}^k}\left|\alpha^k(i,j)+\beta^k(i,j)\right|^+ - 2V(\boldsymbol{\alpha}^k,\boldsymbol{\beta}^k). \quad (83)$$

For all $\epsilon > 0$, $|\alpha^k(i,j)+\beta^k(i,j)|^+ > \epsilon$ for all $i,j \in \mathcal{A}^k$ implies $\delta(\theta)$ can be chosen small enough such that

$$\frac{d}{dt}V(\boldsymbol{\alpha}^k,\boldsymbol{\beta}^k) \le -V(\boldsymbol{\alpha}^k,\boldsymbol{\beta}^k). \quad (84)$$

Therefore, each subsystem, corresponding to each $\theta \in \mathcal{M}_\theta$ in (25), is globally asymptotically stable and

$$\lim_{t\to\infty} d\left((\boldsymbol{\alpha}^k(t),\boldsymbol{\beta}^k(t)\right),\mathcal{R}) = 0. \quad (85)$$

Next, we proceed to study the stability in the regime-switching case. It is straight forward to use the method of multiple Lyapunov functions to extend [16, Corollary 12] to prove global asymptotic stability w.p.1 for the switching system (33).

*Theorem B.1 ([16], Corollary 12):* Consider the system (33) in Definition 4.1, where $\theta(t)$ is the state of a continuous time Markov chain with generator $Q$. Define $\overline{q} := \max_{\theta\in\mathcal{M}_\theta}|q_{\theta\theta}|$ and $\widetilde{q} := \max_{\theta,\theta'\in\mathcal{M}_\theta} q_{\theta\theta'}$. Suppose there exist continuously differentiable functions $V_\theta : \mathbb{R}^n \to \mathbb{R}^+$, $\theta \in \mathcal{M}_\theta$, strictly increasing functions $a_1, a_2 : \mathbb{R}^+ \to \mathbb{R}^+$ with $a_1(0) = a_2(0) = 0$ and $a_1(t), a_2(t) \to \infty$ as $t \to \infty$, a real number $v > 1$ such that the following hold:

1) $a_1(d(X,H)) \le V_\theta(X) \le a_2(d(X,H))$, $\forall X \in \mathbb{R}^r$, $\theta \in \mathcal{M}_\theta$;
2) $(\partial V_\theta/\partial X) f_\theta(X) \le -\lambda V_\theta(X)$, $\forall X \in \mathbb{R}^r$, $\forall \theta \in \mathcal{M}_\theta$;
3) $V_\theta(X) \le v V_{\theta'}(X)$, $\forall X \in \mathbb{R}^r$, $\theta, \theta' \in \mathcal{M}_\theta$;
4) $(\lambda+\widetilde{q})/\overline{q} > v$.

Then, (33) is globally asymptotically stable almost surely.

In light of (84), Hypothesis 2) in Theorem B.1 is trivially satisfied. Further, since the Lyapunov functions are the same for all subsystems $\theta \in \mathcal{M}_\theta$, existence of $v > 1$ in Hypothesis 3) is automatically guaranteed. Given that $\lambda = 1$ in hypothesis 2) (see (84)), it remains to ensure that the generator of Markov chain $Q$ satisfies hypothesis 4), i.e., $1+\widetilde{q} > \overline{q}$. This is trivially satisfied since, by (3), $|q_{\theta\theta'}| \le 1$, for all $\theta, \theta' \in \mathcal{M}_\theta$ and the proof is complete.

## Appendix C
## Proof of Theorem 4.2

Recall from Section IV-D that $\{(\boldsymbol{\alpha}_n^k, \widetilde{\boldsymbol{\beta}}_n^k)\}$ and $\{(\boldsymbol{\alpha}_n^k, \boldsymbol{\beta}_n^k)\}$ form discrete time stochastic approximation iterates of the same differential inclusion (31). Assuming that agent $k$ adopts a strategy $\boldsymbol{\psi}^k = \boldsymbol{\psi}(\boldsymbol{\alpha}^k, \boldsymbol{\beta}^k)$ of the form (9), we look at the coupled systems $(\boldsymbol{\alpha}^k(t), \boldsymbol{\beta}^k(t))$ and $(\boldsymbol{\alpha}^k(t), \widetilde{\boldsymbol{\beta}}^k(t))$. Note that we assume both systems apply the same strategy $\boldsymbol{\psi}^k$, which is a function of $\boldsymbol{\beta}^k$ (not $\widetilde{\boldsymbol{\beta}}^k$). (This is in contrast to (31), where a strategy $\widetilde{\boldsymbol{\psi}}^k$ was used.)

Let $D_t = \|\widetilde{\boldsymbol{\beta}}^k(t) - \boldsymbol{\beta}^k(t)\|$. Then, for all $0 \le \lambda \le 1$

$$\begin{aligned} D_{t+\lambda} &= \left\|\widetilde{\boldsymbol{\beta}}^k(t+\lambda) - \boldsymbol{\beta}^k(t+\lambda)\right\| \\ &= \left\|\widetilde{\boldsymbol{\beta}}^k(t) + \lambda\cdot\frac{d}{dt}\widetilde{\boldsymbol{\beta}}^k(t) - \boldsymbol{\beta}^k(t) - \lambda\cdot\frac{d}{dt}\boldsymbol{\beta}^k(t) + o(\lambda)\right\| \\ &\le D_t + \lambda\left\|\frac{d}{dt}\widetilde{\boldsymbol{\beta}}^k(t) - \frac{d}{dt}\boldsymbol{\beta}^k(t)\right\| + o(\lambda). \end{aligned} \quad (86)$$

Using the results from Section IV-A, the second term in the right-hand side of (86) becomes

$$\frac{d}{dt}\widetilde{\boldsymbol{\beta}}^k(t) - \frac{d}{dt}\boldsymbol{\beta}^k(t) = -\left(\widetilde{\boldsymbol{\beta}}^k(t) - \boldsymbol{\beta}^k(t)\right). \quad (87)$$

Applying (87) into (86) results in

$$D_{t+\lambda} \le (1-\lambda)D_t + o(\lambda). \quad (88)$$

Therefore,

$$\frac{d}{dt}D_t \le -D_t \quad (89)$$

for almost every $t$, from which it follows:

$$D_{-t}e^{-t} \ge D(0) \quad (90)$$

for all $t \ge 0$. Note that $D_t$ is bounded since both $\widetilde{\boldsymbol{\beta}}^k(t)$ and $\boldsymbol{\beta}^k(t)$ are bounded. Finally, since $\lim_{t\to\infty} D_{-t}e^{-t} = 0$ and $D_t \ge 0$, one can conclude from (90) that $D_t = D_0 = 0$. Hence, the limit sets coincide.

$$\begin{aligned} \frac{d}{dt}V(\boldsymbol{\alpha}^k,\boldsymbol{\beta}^k) &= 2\sum_{i,j\in\mathcal{A}^k}\left|\alpha^k(i,j)+\beta^k(i,j)\right|^+ \\ &\quad\times\left[\frac{\delta}{A^k}\left[U_l^k(j,\boldsymbol{s}^k;\theta)+U_{g,t}^k(j;\theta) - \left[U_l^k(i,\boldsymbol{s}^k;\theta)+U_{g,t}^k(i;\theta)\right]\right] - \left[\alpha^k(i,j)+\beta^k(i,j)\right]\right] \\ &\le 2\sum_{i\in\mathcal{A}^k}\sum_{j\in\mathcal{A}^k}\left|\alpha^k(i,j)+\beta^k(i,j)\right|^+\left[\frac{\delta}{A^k}\cdot C(\theta) - \left[\alpha^k(i,j)+\beta^k(i,j)\right]\right] \end{aligned} \quad (81)$$

## REFERENCES


[1] S. Hart and A. Mas-Colell, "A simple adaptive procedure leading to correlated equilibrium," *Econometrica*, vol. 68, no. 5, pp. 1127–1150, Sept. 2000.

[2] S. Hart and A. Mas-Colell, "A reinforcement procedure leading to correlated equilibrium," Economic Essays: A Festschrift for Werner Hildenbrand pp. 181–200.

[3] S. Hart, "Adaptive heuristics," *Econometrica*, vol. 73, no. 5, pp. 1401–1430, 2005.

[4] R. J. Aumann, "Correlated equilibrium as an expression of Bayesian rationality," *Econometrica*, vol. 55, no. 1, pp. 1–18, Jan. 1987.

[5] R. Nau, S. G. Canovas, and P. Hansen, "On the geometry of nash equilibria and correlated equilibria," *Int. J. Game Theory*, vol. 32, no. 4, pp. 443–453, 2004.

[6] H. J. Kushner and G. Yin, *Stochastic Approximation Algorithms and Applications*, 2nd ed. New York: Springer-Verlag, 2003.

[7] A. Benveniste, M. Metivier, and P. Prioret, *Adaptive Algorithms and Stochastic Approximations*. New York, NY, USA: Springer-Verlag, 1990.

[8] V. Krishnamurthy, M. Maskery, and G. Yin, "Decentralized adaptive filtering algorithms for sensor activation in an unattended ground sensor network," *IEEE Trans. Signal Process.*, vol. 56, no. 12, pp. 6086–6101, Dec. 2008.

[9] O. N. Gharehshiran and V. Krishnamurthy, "Coalition formation for bearings-only localization in sensor networks-a cooperative game approach," *IEEE Trans. Signal Process.*, vol. 58, no. 8, pp. 4322–4338, Aug. 2010.

[10] M. Maskery, V. Krishnamurthy, and Q. Zhao, "Decentralized dynamic spectrum access for cognitive radios: Cooperative design of a noncooperative game," *IEEE Trans. Commun.*, vol. 57, no. 2, pp. 459–469, Feb. 2009.

[11] O. N. Gharehshiran, A. Attar, and V. Krishnamurthy, "Collaborative subchannel allocation in cognitive LTE femto-cells: A cooperative gametheoretic approach," *IEEE Trans. Commun.*, vol. 61, no. 1, pp. 325–334, Jan. 2013.

[12] M. Benaïm, J. Hofbauer, and S. Sorin, "Stochastic approximations and differential inclusions; Part II: Applications," *Math. Oper. Res.*, vol. 31, no. 3, pp. 673–695, 2006.

[13] G. Yin and Q. Zhang, *Continuous-Time Markov Chains and Applications: A Singular Perturbation Approach*. New York, NY, USA: Springer-Verlag, 1998.

[14] G. Yin and C. Zhu, *Hybrid Switching Diffusions: Properties and Applications*. New York, NY, USA: Springer-Verlag, 2009, vol. 63.

[15] D. Chatterjee and D. Liberzon, "Stability analysis of deterministic and stochastic switched systems via a comparison principle and multiple Lyapunov functions," *SIAM J. Control Optim.*, vol. 45, no. 1, pp. 174–206, 2007.

[16] D. Chatterjee and D. Liberzon, "On stability of randomly switched nonlinear systems," *IEEE Trans. Autom. Control*, vol. 52, no. 12, pp. 2390–2394, Dec. 2007.

[17] D. Fudenberg and D. K. Levine, *The Theory of Learning in Games*. Cambridge, MA, USA: MIT Press, 1998, vol. 2.

[18] S. Hart and A. Mas-Colell, "A general class of adaptive strategies," *J. Econ. Theory*, vol. 98, no. 1, pp. 26–54, May 2001.

[19] P. Billingsley, *Convergence of Probability Measures*. New York, NY, USA: Wiley, 1968.

[20] S. Boyd and L. Vandenberghe, *Convex Optimization*. New York, NY, USA: Cambridge Univ. Press, 2004.

[21] J. P. Aubin, *Dynamic Economic Theory: A Viability Approach*. : Springer-Verlag, 1997, vol. 174.

[22] D. Liberzon, *Switching in Systems and Control*. New York, NY, USA: Springer, 2003.

[23] D. Fudenberg and D. K. Levine, "Conditional universal consistency," *Games Econ. Behav.*, vol. 29, no. 1–2, pp. 104–130, Oct. 1999.

[24] H. J. Kushner, *Approximation and Weak Convergence Methods for Random Processes With Applications to Stochastic Systems Theory*. Cambridge, MA, USA: MIT press, 1984, vol. 6.

[25] H. Kushner and A. Shwartz, "An invariant measure approach to the convergence of stochastic approximations with state dependent noise," *SIAM J. Control Optim.*, vol. 22, no. 1, pp. 13–27, 1984.

[26] G. Yin, V. Krishnamurthy, and C. Ion, "Regime switching stochastic approximation algorithms with application to adaptive discrete stochastic optimization," *SIAM J. Optim.*, vol. 14, no. 4, pp. 1187–1215, 2004.


**Omid Namvar Gharehshiran** (S'09) received the B.Sc. degree from Sharif University of Technology, Tehran, Iran, in 2007 and the M.A.Sc. degree from the University of British Columbia, Vancouver, BC, Canada, in 2010. He is currently working towards the Ph.D. degree at the University of British Columbia.

He is a member of the Statistical Signal Processing Laboratory, University of British Columbia. His research interests span stochastic optimization, games and learning theory.

Mr. Namvar Gharehshiran was the recipient of the UBC Four Year Doctoral Fellowship in 2010.

**Vikram Krishnamurthy** (F'05) was born in 1966. He received the bachelors degree from the University of Auckland, Auckland, New Zealand, in 1988 and Ph.D. degree from the Australian National University, Acton, in 1992.

He currently is a Professor and Canada Research Chair in the Department of Electrical Engineering, University of British Columbia, Vancouver, BC, Canada. His current research interests include game theory, stochastic control in sensor networks, and modeling of biological ion channels and biosensors.

Dr. Krishnamurthy served as Editor-in-Chief of IEEE JOURNAL SELECTED TOPICS IN SIGNAL PROCESSING and as Distinguished lecturer for the IEEE signal processing society. He has also served as associate editor for several journals including IEEE TRANSACTIONS AUTOMATIC CONTROL and IEEE TRANSACTIONS ON SIGNAL PROCESSING.

**George Yin** (F'02) received the B.S. degree in mathematics from the University of Delaware, Newark, DE, USA, in 1983 and the M.S. degree in electrical engineering and Ph.D. degree in applied mathematics from Brown University, Providence, RI, USA, in 1987.

He joined Wayne State University, Detroit, MI, USA, in 1987 and became a Professor in 1996. His research interests include stochastic systems theory, stochastic approximation, control, and applications.

Prof. Yin severed on many technical and conference committees; he cochaired the 2011 SIAM Conrol Conference, the 1996 AMS-SIAM Summer Seminar, and the 2003 AMS-IMS-SIAM Summer Research Conference, and co-organized 2005 IMA Workshop on Wireless Communications. He has served as vice chair and program director of SIAM Activity Group on control and systems theory; he also served as chair of the SIAM W.T. and Idalia Reid Prize Selection Committee, the SIAG/Control and Systems Theory Prize Selection Committee, and the SIAM SICON Best Paper Prize Committee. He is an associate editor of *SIAM Journal on Control and Optimization*, and is on the editorial board of a number of other journals. He was an associate editor of *Automatica* and IEEE TRANSACTIONS ON AUTOMATIC CONTROL. He is president of Wayne State University's Academy of Scholars.