On Asynchronous Implementations of Fictitious Play for Distributed Learning

BRIAN SWENSON[†*], SOUMMYA KAR[†] AND JOÃO XAVIER[⋆]

***Abstract*—The classical Fictitious Play (FP) algorithm is defined within a framework of synchronous repeated play. In practice, the global synchronization assumed in classical FP can be difficult to achieve in large-scale multi-agent settings. The paper considers *FP with asynchronous updates*—a variant of FP in which players are permitted to be either "active" or "idle" in each stage of the repeated play process. The FP process with asynchronous updates is shown to be a generalization of classical FP. Analytical convergence results are given for the asynchronous variant of FP. Furthermore, the paper studies an asynchronous continuous-time embedding of FP. The continuous-time embedded FP process may be implemented in a real-world setting where no global clock is available. Sufficient conditions for convergence of the continuous-time embedded process are provided as a consequence of the convergence analysis for FP with asynchronous updates. Example implementations that attain the sufficient condition are presented.**

I. Introduction

Many game-theoretic learning algorithms (e.g., [1]–[5]) operate within the learning framework of *synchronous repeated play*. In this framework, a set of players repeatedly face off in some fixed game. Each iteration of the repeated interaction, players may adaptively change their strategies according to some set of behavior rules. Ideally, the behavior rules should be designed in such a manner as to ensure that the induced dynamical system converges to the set of Nash equilibria (NE) (or some other desirable equilibrium set), thus allowing players to "learn" equilibrium strategies.

While convergence results for such algorithms are theoretically encouraging, they can be of limited practical value in real-world scenarios where global synchronization may be difficult to achieve.

Fictitious Play (FP) [6]–[8] is a highly prototypical game-theoretic learning algorithm that is traditionally defined within the framework of synchronous repeated play. In the FP algorithm, each agent tracks the empirical action history of opponents and chooses her next-iteration action according to a myopic best response rule. The algorithm has been shown to achieve NE learning in several classes of games, including the class of multiplayer potential games [9], [10].

Our main contribution is the proposal of *FP with asynchronous updates*—a variant of FP in which players may be active in some rounds and idle in others. We provide a theoretical convergence analysis and give sufficient conditions under which the asynchronous algorithm can be shown to achieve NE learning.

In order to demonstrate how the convergence result for FP with asynchronous updating can be applied to real-world applications we present an (asynchronous) continuous-time embedding of FP. Using the general asynchronous FP result discussed above, we derive a simple sufficient condition that ensures agents achieve NE learning in the continuous-time embedded process. The condition requires the timing of agents actions attain a fairly weak notion of synchronization: if $N_i(t)$ is a process counting the number of actions taken by agent i up to time t, then the ratio $N_i(t)/N_j(t)$ is required to converge to 1 for all agent pairs.

In order to demonstrate how the condition can be achieved in practice we present two simple examples of action timing rules that achieve the sufficient condition. In the first example, agents choose the timing of their actions according to a stochastic process. In the second example, a group of agents with disparate clock skew rates is assumed to be capable of periodically (with respect to their local clocks) observing some aggregate statistic. Agents adaptively choose the timing of their actions according to some deterministic function of their observation.

The remainder of the paper is organized as follows. Section II sets up the notation to be used in the subsequent development. Section III introduces the synchronous repeated play framework and introduces classical FP as an algorithm operating within this framework. Section IV presents our asynchronous repeated play framework, presents our asynchronous generalization of FP, and proves our main convergence result. Section V presents the continuous-time embedding of FP, states convergence results, and provides example implementations.

II. Preliminaries

A. Notation

A game in normal form is given by the triple $\Gamma = (\mathcal{N}, \{u_i, Y_i\}_{i,\in\mathcal{N}})$ where $\mathcal{N} = \{1,\ldots,N_p\}$ denotes the set of players, Y_i denotes the finite set of actions available to player i, and $u_i : \prod_{i\in N} Y_i \to \mathbb{R}$ denotes the utility function of player i. Denote by $Y := \prod_{i\in N} Y_i$ the joint action space.

In order to guarantee the existence of Nash equilibria, it is necessary to consider the mixed-extension of Γ in which players are permitted to play probabilistic strategies. Let $m_i := |Y_i|$ be the cardinality of the action space of player i, and let $\Delta_i := \{p \in \mathbb{R}^{m_i} : \sum_{k=1}^{m_i} p(k) = 1,\ p(k) \geq 0,\ \forall k\}$ denote the set of mixed strategies available to player i—note that a mixed

The work was partially supported by the FCT projects FCT [UID/EEA/5009/2013] and FCT [UID/EEA/50009/2013] through the Carnegie Mellon/Portugal Program managed by ICTI from FCT and by FCT Grant CMU-PT/SIA/0026/2009 and was partially supported by NSF grant ECCS-1306128.

[†]Department of Electrical and Computer Engineering, Carnegie Mellon University, Pittsburgh, PA 15213, USA (brianswe@andrew.cmu.edu and soummyak@andrew.cmu.edu).

[⋆]Institute for Systems and Robotics (ISR/IST), LARSyS, Instituto Superior Técnico, University of Lisbon (jxavier@isr.ist.utl.pt).

978-1-4673-8576-3/15/$31.00 ©2015 IEEE

strategy is a probability distribution over the action space of player i. Denote by $\Delta := \prod_{i\in N} \Delta_i$ the set of joint mixed strategies. When convenient, we represent a mixed strategy $p \in \Delta$ by $p = (p_i, p_{-i})$, where p_i denotes the marginal strategy of player i and p_{-i} is a $(n-1)$-tuple containing the marginal strategies of the other players.

In the context of mixed strategies, we often wish to retain the notion of playing a single deterministic action. For this purpose, let $A_i := \{e_1, \ldots, e_{m_i}\} \subset \Delta_i$ denote the set of "pure strategies" of player i, where e_j is the j-th canonical vector containing a 1 at position j and zeros otherwise. Note that there is a one-to-one correspondence between a player's action set Y_i and the player's set of pure strategies $A_i \subset \Delta_i$.

The mixed utility function of player i is given by

$$U_i(p) := \sum_{y\in Y} u_i(y) p_1(y) \ldots p_{N_p}(y)$$

where $U_i : \Delta \to \mathbb{R}$. Note that the mixed utility $U_i(p)$ may be interpreted as the expected utility of $u_i(y)$ given that players' (marginal) mixed strategies p_i are independent.

The set of Nash equilibria is given by $NE := \{p \in \Delta : U_i(p_i, p_{-i}) \geq U_i(p'_i, p_{-i}),\ \forall p'_i \in \Delta_i,\ \forall i \in N\}$. The distance of a distribution $p \in \Delta$ from a set $S \subset \Delta$ is given by $d(p, S) = \inf\{\|p - p'\| : p' \in S\}$. Throughout the paper $\|\cdot\|$ denotes the standard $\mathcal{L}_2$ Euclidean norm unless otherwise specified.

Throughout, we assume there exists a probability space $(\Omega, \mathcal{F}, \mathbb{P})$ rich enough to carry out the construction of the various random variables required in the paper. For a random object X defined on a measurable space $(\Omega, \mathcal{F})$, let $\sigma(X)$ denote the σ-algebra generated by X [11]. As a matter of convention, all equalities and inequalities involving random objects are to be interpreted almost surely (a.s.) with respect to the underlying probability measure unless otherwise stated.

III. Classical Fictitious Play

A. Classical (Synchronous) Repeated Play Learning

Let a normal form game Γ be fixed. Let players repeatedly face off in Γ in discrete stages $n \in \{1, 2, \ldots\}$. Let $a_i(n) \in A_i$ denote the action played by player i in stage n of the repeated play, and let the N_p-tuple $a(n) = (a_1(n), \ldots, a_{N_p}(n))$ denote the joint action in stage n.

Note that in a classical (synchronous) repeated play learning, each player is assumed to be given the opportunity to revise their action choice $a_i(n)$ each iteration of the repeated play process.

Let the empirical history distribution (or empirical distribution) of player i be given by[1]

$$q_i(n) := \frac{1}{n}\sum_{s=1}^{n} a_i(s),$$

and let the joint empirical distribution be given by the n-tuple $q(n) := (q_1(n), \ldots, q_{N_p}(n))$.

In Section IV we will consider a generalization of this classical repeated play learning framework which may be used to model asynchronous behavior within a discrete-time framework.

B. Classical Fictitious Play Process

Classical FP operates within the classical (synchronous) repeated play framework given in Section III-A. A sequence of actions $\{a(n)\}_{n\geq 1}$ is said to be a *fictitious play process* if for all $i \in \mathcal{N}$ and $n \geq 1$,[2]

$$a_i(n+1) \in \arg\max_{\alpha_i \in A_i} U_i(\alpha_i, q_{-i}(n)).$$

Intuitively speaking, this describes a process where each player (naively) assumes that her opponents are playing according to stationary independent strategies. Following this intuition, the player assumes that $q_{-i}(n)$ accurately represents the mixed strategy of her opponents and chooses a next-stage action in order to myopically optimize her next-stage utility.

In this paper we wish to demonstrate that FP retains its fundamental convergence properties when implemented in an asynchronous manner. In order to show this, we will leverage certain robustness properties of FP. In particular, throughout the paper we assume that the game Γ satisfies the following assumption:

A. 1. *The game* Γ *is such that if* $\epsilon_n \to 0$ *and* $U_i(a_i(t+1), q_{-i}(n)) \geq \max_{\alpha_i\in A_i} U_i(\alpha_i, q_{-i}(n)) - \epsilon_n,\ \forall i,$ *for all* $n \geq 1$*, then* $\lim_{n\to\infty} d(q(n), NE) = 0$.

If a game Γ satisfies the above assumption, then we say FP is *best-response robust* on Γ. Several classes of games have been shown to satisfy **A.1**, including two-player zero-sum games, generic $2 \times m$ games, and the class of multiplayer games known as potential games [12]–[14].

IV. Fictitious Play with Asynchronous Updates

In Section IV-A we introduce the generalized repeated play framework used to model asynchronous interactions. In Section IV we present our (discrete-time) model of FP with Asynchronous Updates.

Several practical models of asynchrony can be reduced to the framework considered here. For example, in Section V we consider the implementation of FP in a continuous-time setting where players do not have access to a global clock. We show that the technical analysis of such a process reduces to studying the (discrete-time) FP with Asynchronous Updates presented in this section.

A. Asynchronous Repeated Play Learning

In order to model asynchrony in repeated play learning, we consider an extension of the classical repeated play framework of Section III-A in which players may be "active" in some rounds and "idle" in others.

Let $n \in \mathbb{N}$, and let $\{X_i(n)\}_{n\geq 1}$, be a sequence of (deterministic or random) variables $X_i(n) \in \{0, 1\}$ indicating the rounds in which player i is active. Let $N_i(n)$ count the number

[1] Note that each $a_i(n) \in A_i$ is a Dirac distribution, and thus the empirical distribution $q_i(n)$ is a normalized histogram of the action choices of player i.

[2] The initial action $a(1)$ may be chosen arbitrarily.

of rounds in which player i has been active up to and including time n; i.e.,

$$N_i(n) := \sum_{s=1}^{n} X_i(s).$$

Let $a_i(n)$ represent the action chosen by player i in round n. Let the empirical distribution of player i be defined in this setting as

$$q_i(n) := \frac{1}{N_i(n)} \sum_{s=1}^{n} a_i(s) X_i(s).$$

B. Fictitious Play with Asynchronous Updates

Within the generalized repeated-play framework given above, we say a sequence of actions $\{a(n)\}_{n\geq 1}$ is a *FP process with asynchronous updates* (or asynchronous FP process) if[3]

$$a_i(n+1) \in \begin{cases} \arg\max_{\alpha_i \in A_i} U_i(\alpha_i, q_{-i}(n)) & \text{if } X_i(n+1) = 1, \\ a_i(n) & \text{otherwise.} \end{cases} \tag{1}$$

This models a scenario in which each player i may update her action in round $(n+1)$ according to traditional best-response dynamics only if $X_i(n) = 1$; otherwise, the action of player i persists from the previous round.[4]

Note that classical FP of Section III-B may be seen as a special case within this framework with $X_i(n) = 1,\ \forall i, n$.

Combined with **A.1**, the following assumption is sufficient (to be shown) to ensure that the FP process defined in (1) leads to NE learning:

A. 2. *(i) For each i there holds $\lim_{n\to\infty} N_i(n) = \infty$; (ii) for all i, j there holds, $\lim_{n\to\infty} \frac{N_i(n)}{N_j(n)} = 1$.*

Part (i) in the above assumption ensures that players are active in infinitely many rounds. Part (ii) in the above assumption ensures that the number of actions taken by each player remain relatively close; effectively (ii) ensures that players obtain a weak form of synchronization.

The following theorem shows that, under the above assumption, FP with asynchronous updates achieves NE learning in any game satisfying the robustness property **A.1**.

Theorem 1. *Let Γ be a finite normal form game satisfying **A.1**. Let the action sequence $\{a(n)\}_{n\geq 1}$ be determined according to a FP process with asynchronous updates and assume **A.2** holds. Then players learn NE strategies in the sense that $\lim_{n\to\infty} d(q(n), NE) = 0$.*

Sections IV-C–IV-E are devoted to proving Theorem 1. Sections IV-C–IV-D introduce notation and properties that are useful in the proof, and Section IV-E presents a proof of the theorem. We note that the proof of Theorem 1 below is inspired by the development in [15].

C. Some Additional Definitions

In order to prove Theorem 1 we will study an underlying (synchronous) FP process that is embedded in the asynchronous FP process defined in (1).

In particular, for $s \in \mathbb{N}_+$ define the following terms: $\tau_i(s) := \sup\{n \in \mathbb{N}_+ : N_i(n) \leq s\}$, $\tilde{a}_i(s) := a_i(\tau_i(s))$, $\tilde{a}(s) := (\tilde{a}_1(s), \ldots, \tilde{a}_{N_p}(s))$, $\tilde{q}_i(s) := q_i(\tau_i(s))$, $\tilde{q}(s) := (\tilde{q}_1(s), \ldots, \tilde{q}_{N_p}(s))$, $\hat{q}^i_j(s) := q_j(\tau_i(s+1)-1)$, $\hat{q}^i(s) := (\hat{q}^i_1(s), \ldots, \hat{q}^i_{N_p}(s))$. In words, the term $\tau_i(s)$ denotes the round number when player i is active for the s-th time. The terms marked with a $\sim$ correspond to the embedded (synchronous) FP process that we will study in the proof of Theorem 1.

When studying the embedded (synchronous) FP process $\{\tilde{a}(s)\}_{s\geq 1}$, it will be important to characterize the terms to which players are best responding. With this in mind, note that per (1), the action at time $\tau_i(s+1)$ is chosen as $a_i(\tau_i(s+1)) \in \arg\max_{\alpha_i \in A_i} U_i(\alpha_i, q_{-i}(\tau_i(s+1)-1))$. Thus, by construction, the $(s+1)$-th action of player i in the embedded (synchronous) FP process is chosen as

$$\tilde{a}_i(s+1) \in \arg\max_{\alpha_i \in A_i} U_i(\alpha_i, \hat{q}^i_{-i}(s)).$$

In the embedded (synchronous) FP process, the term $\tilde{q}_j(s)$ may be thought of as the "true" empirical distribution of player j, and the term $\hat{q}^i_j(s)$ may be thought of as an estimate which player i maintains of $\tilde{q}_j(s)$, and the term $\hat{q}^i(s)$ (note the superscript) may be thought of as player i's estimate of the joint empirical distribution $\tilde{q}(s)$ at the time of player i's $(s+1)$-th best response.

D. Some Useful Properties

Note that for $i \in N$ and $s \in \{1, 2, \ldots\}$,

$$N_i(\tau_i(s)) = s, \tag{2}$$

and for $i \in N$ and $t \in \{1, 2, \ldots\}$

$$X_i(n) = 1 \implies \tau_i(N_i(n)) = n. \tag{3}$$

Furthermore, note that $X_i(n) = 0$ implies that $N_i(n) = N_i(n-1)$, and in particular,

$$X_i(n) = 0 \implies q_i(n) = q_i(n-1). \tag{4}$$

These facts are readily verified by conferring with the definitions of τ_i, N_i, and X_i.

E. Proof of Theorem 1

Proof. As a first step, we wish to show that $\lim_{s\to\infty} d(\tilde{q}(s),\ NE) = 0$. We accomplish this by showing that there exists a sequence $\{\epsilon_s\}_{s\geq 1}$ such that $\lim_{s\to\infty} \epsilon_s = 0$ and

$$U_i(a_i(s+1), \tilde{q}_{-i}(s)) \geq \max_{\alpha_i \in A_i} U_i(\alpha_i, \tilde{q}_{-i}(s)) - \epsilon_s,\ \forall s \geq 1. \tag{5}$$

By assumption **A.1**, it will then follow that $\lim_{s\to\infty} d(\tilde{q}(s),\ NE) = 0$.

To that end, for $i \in \mathbb{N}$ define $v_i : \Delta_{-i} \to \mathbb{R}$ by $v(q_{-i}) := \max_{\alpha_i \in A_i} U_i(\alpha_i, q_{-i})$, and note that by (1),

[3] Let $X_i(1) = 1,\ \forall i$ and let the initial action $a_i(1)$ be chosen arbitrarily for all i.

[4] We chose to state the dynamics of (1) in their current form in order to more clearly draw the parallel between FP with asynchronous updates of this Section and the continuous-time embedded FP process of Section V. However, it is to be noted that when $X_i(n+1) = 0$ the action $a_i(n+1)$ may in fact be arbitrary (or even null, denoting a player is inactive and chooses not to play any action)—the main convergence result (Theorem 1) still holds.

$U_i(a_i(\tau_i(s+1)), q_{-i}(\tau_i(s+1)-1)) = v_i(q_{-i}(\tau_i(s+1)-1))$, or equivalently by the definitions of $\tilde{a}(s)$ and $\hat{q}^i(s)$,

$$U_i(\tilde{a}_i(s+1)), \hat{q}^i_{-i}(s)) = v_i(\hat{q}^i_{-i}(s)). \tag{6}$$

By Lemma 1 in the appendix (together with Lemma 7.3 of [15]) there holds $\lim_{s\to\infty} \|\hat{q}^i(s) - \tilde{q}(s)\| = 0$. It follows that $\lim_{s\to\infty} \|\hat{q}^i_{-i}(s) - \tilde{q}_{-i}(s)\| = 0$, which by the Lipschitz continuity of $U_i(\cdot)$ implies that $\lim_{s\to\infty} |U_i(\alpha_i, \hat{q}^i_{-i}(s)) - U_i(\alpha_i, \tilde{q}_{-i}(s))| = 0,\ \forall \alpha_i \in A_i, \forall i$, and $\lim_{s\to\infty} |v_i(\hat{q}^i_{-i}(s)) - v_i(\tilde{q}_{-i}(s))| = 0, \forall i$. By (6) we see that $\lim\limits_{s\to\infty} |U_i(\tilde{a}_i(s+1)), \tilde{q}_{-i}(s)) - v_i(\tilde{q}_{-i}(s))| = 0,\ \forall i$; i.e., there exists a sequence $\{\epsilon_s\}_{s\geq 1}$ such that $\epsilon_s \to 0$ and (5) holds. It follows by **A.1** that

$$\lim_{s\to\infty} d(\tilde{q}(s),\ NE) = 0. \tag{7}$$

We now proceed to show that $\lim_{n\to\infty} d(q(n),\ NE) = 0$. Let $\varepsilon > 0$ be given. By Lemma 1 (see appendix), for each $i \in \mathcal{N}$ there exists a time $S_i > 0$ such that $\forall s \geq S_i$, $\|q(\tau_i(s)) - \tilde{q}(s)\| < \frac{\varepsilon}{2}$. Let $S^{'} = \max_i\{S_i\}$. By (7) there exists a time $S^{''}$ such that $\forall s \geq S^{''}$, $d(\tilde{q}(s),\ NE) < \frac{\varepsilon}{2}$. Let $S = \max\{S^{'}, S^{''}\}$. Then

$$d(q(\tau_i(s)),\ NE) < \varepsilon,\ \forall i,\ \forall s \geq S. \tag{8}$$

Let $T = \max_i\{\tau_i(S)\}$. Note that for some i, $q(T) = q(\tau_i(S))$, and therefore by (8),

$$d(q(T),\ NE) < \varepsilon. \tag{9}$$

Also note that for any $n_0 > T$, it holds that $N_i(n_0) \geq S$ (since $N_i(\tau_i(S)) = S$, and $N_i(n)$ is non-decreasing in n), and moreover

$$\begin{aligned} &X_i(n_0) = 1 \text{ for some } i \implies q(n_0) = q(\tau_i(N_i(n_0))),\\ &X_i(n_0) = 0 \text{ for all } i \implies q(n_0) = q(n_0 - 1), \end{aligned} \tag{10}$$

where the first implication holds with $N_i(n_0) \geq S$. In the above, the first line follows from (3), and the second line follows from (4). Consider $n \geq T$. If for some i, $X_i(n) = 1$, then by (10) and (8), $d(q(n),\ NE) = d(q(\tau_i(N_i(n))),\ NE) < \varepsilon$. Otherwise, if $X_i(n) = 0\ \forall i$, then $q(n) = q(n-1)$.

Iterate this argument m times until either (i) $X_i(n-m) = 1$ for some i, or (ii), $t - m = T$. In the case of (i), $d(q(n),\ NE) = d(q(n-m),\ NE) = d(q(\tau_i(N_i(n-m))),\ NE) < \varepsilon$, where the inequality again follows from (8) and the fact that $n - m > T \implies N_i(n-m) \geq S$. In the case of (ii), $d(q(n),\ NE) = d(q(T),\ NE) < \varepsilon$, where the inequality follows from (9). Since $\varepsilon > 0$ was chosen arbitrarily, it follows that $\lim\limits_{n\to\infty} d(q(n),\ NE) = 0$. □

V. Continuous-Time Embedding of Fictitious Play

In classical FP it is supposed that agents take actions in a synchronous manner. In terms of real-world implementation, this is tantamount to assuming that agents have access to a global clock. In a large-scale distributed setting, this may be an impractical assumption.

In this section we consider an implementation of the FP algorithm in a setting where agents may not have access to a global clock. It will be shown that such algorithms are closely related to the discrete-time asynchronous FP process studied in Section IV. Indeed, sufficient conditions for the convergence of the continuous-time algorithms studied in this section will be derived as a consequence of Theorem 1.

As in the models of repeated play learning discussed in Sections III-A and IV-A, we suppose each player executes a (countable) sequence of actions $\{a_i(n)\}_{n\geq 1}$. Furthermore, we assume that each action is taken at some instant in real time $t \in [0,\infty)$ as measured by some universal clock.[5] In particular, for each player i, let $\{\tau_i(n)\}_{n=1}^{\infty} \subset [0,\infty)$ be an increasing sequence where $\tau_i(n)$ indicates the time (as measured by the universal clock) at which player i chooses an action for the n-th time. Let $a_i(n)$ denote the n-th action taken by player i; i.e., the action taken by player i at time $\tau_i(n)$. For $t \in [0,\infty)$, let $N_i(t) = \sup\{n:\ \tau_i(n) \leq t\}$ denote the number of actions taken by player i by time t. For $t \in [0,\infty)$, we define the empirical distribution of player i in this settings as

$$q_i(t) := \frac{1}{N_i(t)} \sum_{k=1}^{N_i(t)} a_i(k).$$

In particular, for $t \in [0,\infty)$, let $q_i(t_-) := \lim_{\tilde{t}\uparrow t} q_i(\tilde{t})$.

In this context, we say the sequence $\{a_i(n)\}_{n\geq 1}$ is an asynchronous FP action process if for $n \geq 1$ each player i chooses their stage n action according to the rule:[6]

$$a_i(n) \in \arg\max_{\alpha_i \in A_i} u_i(\alpha_i, q_{-i}(\tau_i(n)_-)).$$

We call the sequence $\{\tau_i(n)\}_{n\geq 1}$ the action-timing process for player i, and we refer to any method used to generate $\{\tau_i(n)\}_{n\geq 1}$ (whether deterministic or stochastic) as an action timing rule. Together, we refer to the joint sequence $\{\tau_i(n), a_i(n)\}_{i\in\mathcal{N}, n\geq 1}$ as a continuous-time embedded FP process.

The following assumption provides a sufficient condition on the action-timing process in order to ensure convergence of the continuous-time embedded FP process. The assumption is essentially a restatement of **A.2**, but in a continuous-time setting.

A. 3. *(i) For each i there holds $\lim_{t\to\infty} N_i(t) = \infty$, (ii) for each i, j there holds $\lim_{t\to\infty} N_i(t)/N_j(t) = 1$.*

Part (i) of the above assumption may be satisfied, for instance, as long as the clock skew of each agent stays bounded (with respect to the universal clock), and each agent takes actions infinitely often with respect to their local clock. In order to ensure (ii) is satisfied, slightly more care is needed, as demonstrated by the specific application scenarios below.

The following theorem demonstrates that if the action-timing sequence is chosen to satisfy **A.3**, then the continuous-time embedding of FP will converge to the set of NE.

Theorem 2. *Let Γ be a game satisfying **A.1**. Suppose that $\{a_i(n), \tau_i(n)\}_{i\in\mathcal{N},\ n\geq 1}$ is a continuous-time embedding of FP*

[5] We use the term "universal clock" to refer to some reference clock by which we can compare the timing of actions taken by individual players. However, the universal clock is merely an artifice for analyzing the process, and we do not suppose that players have any particular knowledge concerning it.

[6] Let $\tau_i(1) = 0$ for all i, and let the initial action $a_i(1)$ be chosen arbitrarily for all i.

*satisfying **A.3**. Then players learn NE strategies in the sense that* $\lim_{t\to\infty} d(q(t), NE) = 0$.

The proof of Theorem 2 follows readily from Theorem 1.

In the following two subsections, we give two simple examples of action-timing rules that illustrate different methods for achieving **A.3** (and hence NE learning in the continuous-time embedded FP process).

A. Independent Poisson Clocks

Let $w_i(n) = \tau_i(n+1) - \tau_i(n)$ denote the stage n "waiting time" for player i. Suppose that for each player i and $n \geq 1$, $w_i(n)$ is an independent random variable with distribution $w_i(n) \sim exp(\lambda)$, where $\lambda > 0$ is some parameter that is common among all i. In this case, the action-timing process $\{\tau_i(n)\}_{n\geq 1}$ is said to be a homogenous Poisson process.

The following theorem shows that if the action-timing process is randomly generated in this manner, then players will achieve NE learning.

Theorem 3. *Let* Γ *be a normal-form game satisfying **A.1**. Suppose that players are engaged in a continuous-time embedded asynchronous FP process and the action-timing sequences* $\{\tau_i(n)\}_{n\geq 1}$ *are generated as independent homogenous Poisson processes with common parameter* λ *for all* i*. Then players learn NE strategies in the sense that* $\lim_{t\to\infty} d(q(t), NE) = 0$*, almost surely.*

Proof. By Theorem 1 it is sufficient to show that $\lim_{t\to\infty} N_i(t) = \infty$, $\forall i$, and $\lim_{t\to\infty} \frac{N_i(t)}{N_j(t)} = 1$ for all i, j.

First, note that for any i and $n \geq 1$, $w_i(n) < \infty$ almost surely. Hence, $\tau_i(n) = \sum_{k=1}^n w_i(k) < \infty$ for all i, almost surely. Equivalently, for any $M > 0$, almost surely there exists a (random) time $T > 0$ such that $N_i(t) \geq M$ for all $t \geq T$. Hence, $\lim_{t\to\infty} N_i(t) = \infty$, almost surely.

Now we show that $\lim_{t\to\infty} \frac{N_i(t)}{N_j(t)} = 1$ for all i, j. Let $\tau(1) := \min_i \tau_i(1)$ and let $\mathcal{T}_1 := \{\tau_i(n)\}_{i\in\mathcal{N}, n\geq 1} \backslash \tau(1)$. For $n \geq 2$, let $\tau(n) := \min \mathcal{T}_{n-1}$ and let $\mathcal{T}_n := \mathcal{T}_{n-1}\backslash\tau(n)$. For $n \geq 1$, $i \in \mathcal{N}$, define $X_i(n) \in \{0,1\}$ to be an indicator variable with $X_i(n) = 1$ if $\tau(n) \in \{\tau_i(k)\}_{k\geq 1}$ and $X_i(n) = 0$ otherwise.

Let $\mathcal{F}_0 := \emptyset$ and for $n \geq 1$, let $\mathcal{F}_n := \sigma(\{\tau(k)\}_{k=1}^n)$. For $n \geq 1$ let $\xi_i(n) := \mathbb{P}(X_i(n) = 1 | \, \mathcal{F}_{n-1})$.

Since for each i, $\{\tau_i(n)\}_{n\geq 1}$ is a Poisson process with common parameter λ, there holds $\xi_i(n) = \frac{1}{N_p}$ for all i and n.[7] By Levi's extension of the Borel-Cantelli Lemma (see [11] p.124) there holds

$$\lim_{n\to\infty} \frac{\sum_{k=1}^n X_i(n)}{\sum_{k=1}^n \xi_i(n)} = 1, \text{ a.s.} \tag{11}$$

Note that for each i, $\sum_{k=1}^n X_i(k) = N_i(\tau(n))$ and $\sum_{k=1}^n \xi_i(n) = \frac{n}{N_p}$. Thus by (11),

$$\lim_{n\to\infty} \frac{N_i(\tau(n))}{N_j(\tau(n))} = \lim_{n\to\infty} \frac{N_i(\tau(n))}{n/N_p} \frac{n/N_p}{N_j(\tau(n))} = 1, \text{ a.s., } \forall i, j.$$

Finally, note that $\lim_{n\to\infty} \tau(n) = \infty$ a.s., and for each i $N_i(t)$ is constant on $[0,\infty)\backslash\{\tau(n)\}_{n\geq 1}$. Thus, $\lim_{t\to\infty} \frac{N_i(t)}{N_j(t)} = 1$, almost surely. □

B. Adaptive Clock Rates

In this section we consider a scenario in which each player chooses the timing of her actions (deterministically) according to a personal clock with a skew rate that may be different among players.

Let $w_i(n) = \tau_i(n+1) - \tau_i(n)$ again denote the stage n "waiting time" for player i. For each i, let $w_{i,0}$ denote a base waiting time for player i. The base waiting time of player i may be interpreted as the amount of time which expires according to the universal clock during one unit of time as measured by player i's personal clock. In this manner, the disparity in the $w_{i,0}$ reflects disparate skew rates among players' personal clocks.

Let $N_{min}(t) := \min_i N_i(t)$. At time t, we suppose that player i has knowledge of $N_{min}(s)$ at the time instances $s \in \{kw_{i,0} : \ k \in \mathbb{N}_+, kw_{i,0} \leq t\}$. For each i, let $B_i \in \mathbb{R}$ be a number satisfying $B_i > \max_i w_{i,0}$.

Suppose that player i adaptively chooses her stage n waiting time according to the rule:

$$\begin{aligned} w_i(n) = \min\big\{kw_{i,0} : \ k \in \mathbb{N}_+, \, N_{min}(\tau_i(n) + kw_{i,0}) \\ \geq N_i(\tau_i(n)) - B_i\big\} \end{aligned} \tag{12}$$

In words, this rule may be described as follows: Player i periodically observes $N_{min}(t)$. If $N_i(t) - N_{min}(t) \leq B_i$ then player i takes a new action. If $N_i(t) - N_{min}(t) > B_i$ then player i waits for $N_{min}(t)$ to increase sufficiently (satisfying $N_i(t) - N_{min}(t) \leq B_i$) before taking a new action.

Theorem 4. *Let* Γ *be a normal-form game satisfying **A.1**. Suppose that players are engaged in a continuous-time embedded asynchronous FP process in which the action-timing sequence* $\{\tau_i(n)\}_{n\geq 1}$ *is generated according to the adaptive rule* (12)*. Then players learn NE strategies in the sense that* $\lim_{t\to\infty} d(q(t), NE) = 0$.

Proof. By Theorem 1, it is sufficient to show that $\lim_{t\to\infty} N_i(t) = \infty$ for some (and hence all) i, and that $\lim_{t\to\infty} \frac{N_i(t)}{N_j(t)} = 1$.

Note that for $i^* \in \arg\max_i w_{i,0}$, there holds $N_{i^*}(t) = \lfloor \frac{t}{w_{i^*,0}} \rfloor + 1$, and hence $\lim_{t\to\infty} N_{i^*}(t) = \infty$. Furthermore, by construction, $|N_i(t) - N_{i^*}(t)| \leq 2\max_i B_i$ for all i and for all $t \geq 0$. Hence, $\lim_{t\to\infty} \frac{N_i(t)}{N_j(t)} = 1$, for all i, j. □

VI. Conclusions

Classical FP implicitly assumes agents are capable of synchronizing the timing of their actions. This assumption may be difficult to satisfy in large-scale multi-agent settings where no global clock is available. We introduced an asynchronous generalization of FP and provided sufficient conditions for convergence of the algorithm to the set of NE. We also studied a class of asynchronous continuous-time implementations of FP that may be practical to implement in real-world scenarios.

[7] Recall that N_p denotes the number of players.

Appendix

Lemma 1. *Let $i, j \in N$, let $\tau_i(s)$ and $\tilde{q}_j(s)$ be defined as in Section IV-C, and assume* **A.2** *holds. Then* $\lim_{s\to\infty} \|q_j(\tau_i(s)) - \tilde{q}_j(s)\| = 0$.

Proof. Note that for any $n \in \mathbb{N}$, $q_j(n) = q_j(\tau_j(N_j(n))) = \tilde{q}_j(N_j(n))$, where the first equality follows from [15] Lemma 7.4, and the second equality follows from the definition of $\tilde{q}_j(s)$. Note also that for $s \in \mathbb{N}_+$, $\|\tilde{q}_j(s+1) - \tilde{q}_j(s)\| \leq \frac{1}{s} M_j$, where $M_j := \max_{p',p''\in\Delta_j} \|p' - p''\|$, and more generally, for $s_1, s_2 \in \mathbb{N}$,

$$\|\tilde{q}_j(s_1) - \tilde{q}_j(s_2)\| \leq \sum_{s=\min(s_1,s2)}^{\max(s_1,s_2)-1} \|\tilde{q}_j(s+1) - \tilde{q}_j(s)\| \leq \frac{|s_2 - s_1|}{\min(s_1, s_2)} M_j.$$

Hence,

$$\begin{aligned}
&\|q_j(\tau_i(s)) - \tilde{q}_j(s)\| = \|\tilde{q}_j(N_j(\tau_i(s))) - \tilde{q}_j(s)\| \\
&= \|\tilde{q}_j(N_j(\tau_i(s))) - \tilde{q}_j(N_i(\tau_i(s)))\| \\
&\leq \frac{|N_j(\tau_i(s)) - N_i(\tau_i(s))|}{\min(N_i(\tau_i(s)), N_j(\tau_i(s)))} M_j,
\end{aligned}$$

where the second equality follows from the fact that $N_i(\tau_i(s)) = s$ (see (2)). Thus, it suffices to show that

$$\lim_{s\to\infty} \frac{|N_j(\tau_i(s)) - N_i(\tau_i(s))|}{\min(N_i(\tau_i(s)), N_j(\tau_i(s)))} = 0.$$

But, by **A.2**, for any i, j there holds: $0 = \lim_{n\to\infty} \frac{N_i(n)}{N_j(n)} - 1 = \lim_{s\to\infty} \frac{N_i(\tau_i(s))}{N_j(\tau_i(s))} - 1 = \lim_{s\to\infty} \frac{N_i(\tau_i(s)) - N_j(\tau_i(s))}{N_j(\tau_i(s))}$, where the second equality follows from the fact that (again by **A.2**) $\lim_{s\to\infty} \tau_i(s) = \infty$. □

References

[1] J. R. Marden, G. Arslan, and J. S. Shamma, "Joint strategy fictitious play with inertia for potential games," *IEEE Trans. Automat. Contr.*, vol. 54, no. 2, pp. 208–220, 2009.

[2] J. S. Shamma and G. Arslan, "Dynamic fictitious play, dynamic gradient play, and distributed convergence to Nash equilibria," *IEEE Trans. Automat. Contr.*, vol. 50, no. 3, pp. 312–327, 2005.

[3] S. Hart and A. Mas-Colell, *A reinforcement procedure leading to correlated equilibrium*. Springer, 2001.

[4] D. P. Foster and H. P. Young, "Regret testing: A simple payoff-based procedure for learning Nash equilibrium," *University of Pennsylvania and Johns Hopkins University (mimeo)*, 2003.

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

[6] G. W. Brown, *"Iterative Solutions of Games by Fictitious Play" In Activity Analysis of Production and Allocation*, T. Coopmans, Ed. New York: Wiley, 1951.

[7] D. Fudenberg and D. K. Levine, *The Theory of Learning in Games*. MIT press, 1998, vol. 2.

[8] H. P. Young, *Strategic learning and its limits*. Oxford University Press, 2004, vol. 2002.

[9] D. Monderer and L. Shapley, "Potential games," *Games and Econ. Behav.*, vol. 14, no. 1, pp. 124–143, 1996.

[10] D. Monderer and L. S. Shapley, "Fictitious play property for games with identical interests," *Journal of Economic Theory*, vol. 68, no. 1, pp. 258–265, 1996.

[11] D. Williams, *Probability with Martingales*. Cambridge University Press, 1991.

[12] D. S. Leslie and E. J. Collins, "Generalised weakened fictitious play," *Games and Econ. Behav.*, vol. 56, no. 2, pp. 285–298, 2006.

[13] B. Swenson, S. Kar, and J. Xavier, "On robustness properties in empirical centroid fictitious play," 2015, IEEE Conference on Decision and Control, to appear.

[14] B. Swenson, S. Kar, J. Xavier, and D. Leslie, "Robustness properties in fictitious-play-type algorithms," 2015, in preparation.

[15] B. Swenson, S. Kar, and J. Xavier, "From weak learning to strong learning in fictitious play type algorithms," May 2015, submitted for journal publication, http://arxiv.org/pdf/1504.04920v1.pdf.