Game-Theoretic Learning In A Distributed-Information Setting: Distributed Convergence To Mean-Centric Equilibria

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

***Abstract*—The paper considers distributed learning in large-scale games via fictitious-play type algorithms. Given a preassigned communication graph structure for information exchange among the players, this paper studies a distributed implementation of the Empirical Centroid Fictitious Play (ECFP) algorithm that is well-suited to large-scale games in terms of complexity and memory requirements. It is shown that the distributed algorithm converges to an equilibrium set denoted as the mean-centric equilibria (MCE) for a reasonably large class of games.**

I. Introduction

This paper is concerned with learning equilibria in large games via repeated play type learning algorithms. In particular, our interest is in algorithms that are practical for implementation in large games in terms of computation and communication requirements.

Fictitious Play (FP) is one of the most well-studied and prototypical repeated play learning algorithms [1]–[3]. Intuitively speaking, in FP players form a naive belief that opponents are using time-invariant strategies, and choose their next-stage action as a myopic best response to this belief. Though FP does not enable players to learn an equilibrium in all games [4]–[6], it has been proven to converge to equilibrium in certain large-scale games of interest, including potential games [7]–[9].

Learning in FP generally takes place in the sense that the empirical distribution of players' action histories converges to Nash equilibrium (NE) with respect to the (naive) belief to which each player has been best responding—this form of learning is known as convergence in empirical distribution, or convergence in beliefs.

Though FP has theoretically promising learning results, the practical demands of the algorithm (intense computation and communication requirements) can make it difficult to implement in large games.

Empirical Centroid Fictitious Play (ECFP) [10], [11] has been proposed as an adaptation of FP that is more practical for large-scale implementation. In ECFP it is assumed that players have knowledge of only the centroid of all players' empirical history distributions (as opposed to FP, where players are assumed capable of tracking the empirical history distribution of *all* individual opponents). In the spirit of FP, ECFP may be intuitively interpreted as an algorithm in which players form a naive belief about the strategies of opponents—based, in this case, on the player's knowledge of the centroid empirical history distribution—and choose their next-stage action as a best response to this belief.

ECFP has been shown to enable players to learn a NE in terms of the (less traditional) notion of learning—convergence in *average* empirical distribution [10], [11] (see Section II-C).

In order to study learning properties of ECFP in terms of the more traditional notion of learning—convergence in empirical distribution, or convergence in beliefs—the concept of mean-centric equilibrium (MCE) was introduced [12]. Analogous to the notion of learning considered in FP, it has been shown that in ECFP players learn a MCE in the sense that the empirical distribution of players' action histories converges to MCE with respect to the (naive) belief to which the player has been best responding.

ECFP, by itself addresses only the computational issues of FP in large games. In order to address the demanding communication requirements, the distributed-information learning framework has been proposed [10]. In this framework, players engaged in repeated play are assumed to be equipped with a preassigned communication graph. Players are permitted to communicate with a subset of local neighboring players once per iteration of the repeated play.

The main contribution of the present work is to study learning properties of ECFP in the distributed-information setting; in particular, we show that previous (centralized) results regarding convergence in beliefs to the set of MCE [12] can be extended to the distributed-information setting.

The remainder of the paper is orgainzed as follows. In Section II, the notation to be used in the subsequent development is introduced, and the concepts of repeated-play learning and classical FP are presented. Section III presents the (centralized) ECFP algorithm. Section IV presents the distributed-information framework. Section V presents the distributed-information implementation of ECFP and presents the formal convergence results for the same. Finally, Section VI concludes the paper.

II. Preliminaries

A. Game Theoretic Setup

A normal form game is given by the triple $\Gamma = (N, (Y_i)_{i\in N}, (u_i(\cdot))_{i\in N})$, where $N = \{1,\ldots,n\}$ represents the set of players, Y_i—a finite set of cardinality m_i—denotes the action space of player i and $u_i(\cdot) : \prod_{i=1}^n Y_i \rightarrow \mathbb{R}$ represents the utility function of player i.

The set of mixed strategies for player i is given by $\Delta_i = \{p \in \mathbb{R}^{m_i} : \sum_{k=1}^{m_i} p(k) = 1,\ p(k) \geq 0\ \forall k = 1,\ldots,m_i\}$, the m_i-simplex. A mixed strategy $p_i \in \Delta_i$ may be thought of as a probability distribution from which player i samples

This work was partially supported by NSF Grant ECCS-1306128. The work of J. Xavier and B. Swenson is supported by the FCT funding PEst-OE/EEI/LA0009/2013

[†]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), Instituto Superior Tecnico (IST), University of Lisbon, Portugal (jxavier@isr.ist.utl.pt).

to choose an action. The set of joint mixed strategies is given by $\Delta^n = \prod_{i=1}^n \Delta_i$. A joint mixed strategy is represented by the n-tuple $(p_1, p_2, \ldots, p_n)$, where $p_i \in \Delta_i$ represents the marginal strategy of player i, and it is implicity assumed that players' strategies are independent.

A pure strategy is a degenerate mixed strategy which places probability one on a single action in Y_i. We denote the set of pure strategies by $A_i = \{e_1, e_2, \ldots e_{m_i}\}$ where m_i is the number of actions available to player i, and e_j is the jth canonical vector in $\mathbb{R}^{m_i}$. The set of joint pure strategies is given by $A^n = \prod_{i=1}^n A_i$.

The mixed utility function for player i is given by the function $U_i(\cdot) : \Delta^n \to \mathbb{R}$, such that,

$$U_i(p_1, \ldots, p_n) := \sum_{y \in Y} u_i(y) p_1(y_1) \ldots p_n(y_n). \tag{1}$$

Note that $U_i(\cdot)$ may be interpreted as the expected value of $u_i(y)$ given that the players' mixed strategies are statistically independent. For convenience the notation $U_i(p)$ will often be written as $U_i(p_i, p_{-i})$, where $p_i \in \Delta_i$ is the mixed strategy for player i, and p_{-i} indicates the joint mixed strategy for all players other than i.

The set of Nash equilibria of Γ is given by

$$NE := \{p \in \Delta^n : U_i(p) \geq U_i(g_i, p_{-i}) \ \forall g_i \in \Delta_i, \ \forall i\}, \tag{2}$$

and the subset of consensus equilibria is given by

$$C := \{p \in NE \ : \ p_1 = p_2 = \cdots = p_n\}. \tag{3}$$

The distance of a distribution $p \in \Delta^n$ from a set $S \subset \Delta^n$ is given by $d(p, S) = \inf\{\|p - p'\| : p' \in S\}$. Throughout the paper $\|\cdot\|$ denotes the standard $\mathcal{L}_2$ Euclidean norm.

B. Repeated Play Learning

In a repeated-play learning algorithm, players repeatedly face off in a fixed game Γ. An algorithm designer's objective is to design the behavior rules of individual players in such a way as to ensure that the players eventually learn a Nash equilibrium of Γ through the repeated interaction.

Let $\{a_i(t)\}_{t=1}^\infty$ be a sequence of actions for player i, where $a_i(t) \in A_i$. Let $\{a(t)\}_{t=1}^\infty$ be the associated sequence of *joint* actions $a(t) = (a_1(t), \ldots, a_n(t)) \in A^n$. Note that $a_i(t) \in \mathbb{R}^{m_i}$.

Let $q_i(t)$ be the normalized histogram (empirical distribution) of the actions of player i up to time t, i.e., $q_i(t) := \frac{1}{t}\sum_{s=1}^t a_i(s)$. Similarly, let $q(t) := \frac{1}{t}\sum_{s=1}^t a(s)$ be the *joint* empirical distribution corresponding to the joint actions of the players up to time t.

In what follows, we often make the following assumption:

A. 1. *All players use the same strategy space.*

Under this assumption, let $\bar{q}(t) := (q_1(t) + q_2(t) + \cdots + q_n(t))/n$. Note that $\bar{q}(t) \in \mathbb{R}^m$, where m denotes the cardinality of the action spaces (assumed identical) of the individual players. We refer to $\bar{q}(t)$ as the empirical centroid distribution, or the average empirical distribution. Let $\bar{q}^n(t) := (\bar{q}(t), \bar{q}(t), \ldots, \bar{q}(t)) \in \Delta^n$ denote the mixed strategy where all players use the empirical average as their individual strategy. Finally, let the set of mean-centric equilibria of Γ be given by

$$MCE := \{p \in \Delta^n : \ U_i(p_i, \bar{p}_{-i}) \geq U_i(p'_i, \bar{p}_{-i}), \ \forall p'_i \in \Delta_i, \ \forall i\}, \tag{4}$$

where $\bar{p}_{-i} := (\bar{p}, \ldots, \bar{p}) \in \prod_{j \neq i} \Delta_j$ and $\bar{p} := 1/n \sum_{j \in N} p_j$ are well defined under **A.1**.

C. Notions of Learning

Let E be some equilibrium set. We say a repeated play learning algorithm leads players to learn equilibria E in terms of *convergence in average empirical distribution* if $d(\bar{q}^n(t), E) \to 0$ as $t \to \infty$ [10]. This is the manner in which players in ECFP learn consensus equilibria C, see (3) (see also Section III).

We say a repeated play learning algorithm leads players to learn equilibria E in terms of *convergence in empirical distribution* or *convergence in beliefs* if $d(q(t), E) \to 0$ as $t \to \infty$ [2], [3]. This is the manner in which classical FP leads players to learn elements of NE, see (2) (see also Section II-D), and the manner in which ECFP leads players to learn elements of MCE, see (4) (see also Section III).

D. Classical FP

Suppose players are engaged in repeated play; a sequence of actions $\{a(t)\}_{t \geq 1}$ is called a *fictitious play process* if for all $i \in N$ and $t \geq 2$,[1]

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

Intuitively speaking, this describes a process where player i forms a *belief* that opponents are playing according to statistically independent time-invariant mixed strategies, and chooses a next-stage action as a best response to this belief. In particular, each player i chooses a next-stage action that is a best response given the *belief* that $q_{-i}(t)$ accurately represents the mixed strategy of opponents, where $q_{-i} = (q_1(t), \ldots, q_{i-1}(t), q_{i+1}(t), \ldots, q_n(t)) \in \prod_{j \neq i} \Delta_j$.

The following fundamental result states that players engaged in a FP process learn a NE if the underlying game Γ is a potential game.

Theorem 1. *Let Γ be a potential game, and let $\{a(t)\}_{t \geq 1}$ be an FP process on Γ. Then players learn a NE of Γ in the sense that $d(q(t), NE) \to 0$ as $t \to \infty$.*

Note that the learning occurs in the form of convergence in empirical distribution. This form of learning is often also referred to as convergence in beliefs, since each player's empirical distribution $q_i(t)$ converges to equilibrium (per (2)) with respect to the *belief* that opponents are playing the mixed strategy given by $q_{-i}(t)$. This will be analogous to the manner in which players engaged in ECFP learn a MCE, as discussed in the following section.

[1] In all repeated play learning algorithms discussed in this paper, the action $a(1)$ may be chosen arbitrarily.

III. Empirical Centroid Fictitious Play

In ECFP, introduced in [10], [11], it is assumed that player i has knowledge of the empirical centroid $\bar{q}(t)$, but does not have knowledge of any player's individual empirical distribution. Such an assumption may be practical in large-scale games where it is difficult to track the empirical distribution of all individual players. The ECFP algorithm may be intuitively understood as follows. Players form a *belief* that the centroid $\bar{q}(t)$ accurately represents the mixed strategy in use by each opponent. Each player chooses their next-stage action as a best response according to this belief.

Formally, we say a sequence of action $\{a(t)\}_{t\geq 1}$, $a(t) \in \Delta^n$, is an ECFP process if, for $t \geq 2$

$$a_i(t) \in \arg\max_{\alpha_i \in A_i} U_i(\alpha_i, \bar{q}_{-i}(t)), \tag{5}$$

where $\bar{q}_{-i}(t) = (\bar{q}(t), \ldots, \bar{q}(t)) \in \prod_{j\neq i} \Delta_j$ is an $(n-1)$-tuple containing repeated copies of the centroid $\bar{q}(t)$. In [10], the ECFP algorithm was studied under the following assumption:

A. 2. *The players' utility functions are identical and permutation invariant. That is, for any $i, j \in N$, $u_i(y) = u_j(y)$, and $u([y']_i, [y'']_j, y_{-(i,j)}) = u([y'']_i, [y']_j, y_{-(i,j)})$, where, for any player $k \in N$, the notation $[y']_i$ indicates the action $y' \in Y_k$ being played by player k, and $y_{-(i,j)}$ denotes the set of actions being played by all players other than i and j.*

It was shown that players in an ECFP process learn a Nash consensus equilibrium in the sense that the process converges to the set of consensus equilibria C in terms of convergence in average empirical distribution. The result is summarized in the following theorem.

Theorem 2. *Let $\{a(t)\}_{t\geq 1}$ be an ECFP process such that **A.1-A.2** hold. Then $d(\bar{q}^n(t), C) \to 0$ as $t \to \infty$.*

While the above result does provide a useful characterization of a form of learning in ECFP, it does not characterize the asymptotic behavior of $q(t)$, as is the case in the typical repeated play learning—i.e., convergence in empirical distribution (see Section II-C).

The behavior of ECFP in terms of convergence in empirical distribution was studied in [12]. In studying this form of convergence, assumption **A.2** can be relaxed in favor of the following weaker assumption.

A. 3. *The players' utility functions can be decomposed as $u_i(y) = f_i(y_i) + \phi(y)$ where $f_i(y_i)$ depends only on the action of player i and $\phi(y)$ is a permutation-invariant function, identical for all players.*

The following Theorem (cf. [12], Theorem 2) studies learning in ECFP in terms of convergence in empirical distribution.

Theorem 3. *Let $\{a(t)\}_{t\geq 1}$ be an ECFP process such that assumptions **A.1** and **A.3** hold. Then $d(q(t), MCE) \to 0$ as $t \to \infty$.*

To understand the relative significance of this form of learning, recall that in ECFP players choose their next-stage action as a best response given the *belief* that $\bar{q}_{-i}(t)$ accurately represents the mixed strategies in use by opponents. In converging to MCE, a player's empirical distribution $q_i(t)$ converges to equilibrium (see (4)) with respect to this belief. This result is comparable with the notion of convergence in beliefs from classical FP (see Section II-D).

In the above result it is assumed that players have instantaneous access to $\bar{q}(t)$. Such an assumption may not be feasible in a large-scale setting. In order to study ECFP in an environment with more realistic assumptions on inter-agent communication, the following section introduces the notion of a distributed-information framework.

IV. Distributed-Information Framework

In the classical formulation of FP and the centralized formulation of ECFP, it is assumed that players have perfect knowledge of all information necessary to compute a best response. In the distributed-information framework (introduced in [10]), we consider a more realistic scenario where players have limited knowledge about the action histories of opponents, but may communicate with neighboring players via a preassigned communication structure. Formally we assume:

A. 4. *Players are endowed with a preassigned communication graph $G = (V, E)$, where the vertices V represent the players and the edge set E consists of communication links (bidirectional) between pairs of players than can communicate directly. The graph G is connected.*

A. 5. *Players directly observe only their own actions.*

A. 6. *A player may exchange information with immediate neighbors, as defined by G, at most once for each iteration or round of the repeated play.*

V. Distributed-Information Implementation of ECFP

Due to restrictions on communication and direct observation, in the distributed-information setting it is not possible for players to have perfect knowledge of the empirical centroid, $\bar{q}(t)$. Instead, we suppose that each player forms some estimate of $\bar{q}(t)$; let $\hat{q}_i(t)$ be the estimate that player i maintains of $\bar{q}(t)$. In general, we will not explicitly specify the manner in which players form their estimates $\hat{q}_i(t)$. For example, players may form their estimates according to a gossip-type algorithm [13] or according to a consensus-type algorithm [14]. Rather than explicitly specifying an information dissemination protocol used to estimate $\bar{q}(t)$, we impose only the general assumption that

A. 7. $\|\hat{q}_i(t) - \bar{q}(t)\| = O\left(\frac{\log(t)}{t}\right)$.

We will show that—regardless of the specifics of the information-dissemination protocol in use—so long as **A.7** is met, then any distributed implementation of ECFP will achieve the desired learning result. In Section V we discuss an example implementation that meets this assumption.

In a distributed ECFP process, players choose a best response according to the same basic format of the ECFP best response rule (5), but use $\hat{q}_i(t)$ as a surrogate for $\bar{q}(t)$ in (5). We say a sequence of actions $\{a(t)\}_{t\geq 1}$ is a distributed

ECFP process if players form their estimates $\hat{q}_i(t)$ in a manner conforming to **A.4**-**A.6** and if the next-stage action is chosen according to the rule

$$a_i(t+1) \in \arg\max_{\alpha_i \in A_i} U_i(\alpha_i, \hat{q}_{-i}(t)), \tag{6}$$

where, analogous to (5), $\hat{q}_{-i}(t) = (\hat{q}_i(t), \ldots, \hat{q}_i(t)) \in \prod_{j \neq i} \Delta_j$ is the $(n-1)$ tuple containing repeated copies of the player's estimate of $\bar{q}(t)$.

A. Main Result

The following theorem gives the main convergence result for distributed ECFP. In particular, it states that in distributed ECFP, players learn a consensus equilibrium in terms of convergence of the average empirical distribution (a result known from previous work, [10]), and players learn a MCE in terms of convergence in empirical distribution (a novel contribution of this paper).

Theorem 4. *Let $\{a(t)\}_{t \geq 1}$ be a distributed ECFP process.*
*(i) Assume **A.1**,**A.2**,**A.4**-**A.7** hold. Then the agents achieve asymptotic strategy learning in the sense that $d((\hat{q}_i(t))^n, C) \to 0$ as $t \to \infty$.*
*(ii) Assume **A.1**,**A.3**-**A.7** hold. Then the agents achieve asymptotic strategy learning in the sense that $d(q(t), MCE) \to 0$ as $t \to \infty$.*

Proof. The proof of (i) follows from [10], Theorem 1, and **A.7**. In order to prove (ii), let $\bar{a}(t) = \frac{1}{n}\sum_{i=1}^n a_i(t)$, and let $\bar{a}^n(t) \in \Delta^n$ be the n-tuple $(\bar{a}(t), \ldots, \bar{a}(t))$. Note that for $t \geq 1$

$$\bar{q}^n(t+1) = \bar{q}^n(t) + \frac{1}{t+1}\left(\bar{a}^n(t+1) - \bar{q}^n(t)\right). \tag{7}$$

Define the Φ_i to be the multilinear extension of ϕ_i and F_i to be the extension of f_i where ϕ_i and f_i are as defined in **A.3**, and the multilinear extensions are formed in the same manner as (1). Using (7) we write

$$\Phi(\bar{q}^n(t+1)) = \Phi\left(\bar{q}^n(t) + \frac{1}{t+1}\left(\bar{a}^n(t+1) - \bar{q}^n(t)\right)\right).$$

Applying the multilinearity of $\Phi(\cdot)$, we obtain

$$\Phi(\bar{q}^n(t+1)) = \Phi(\bar{q}^n(t)) + \frac{1}{t+1}\sum_{i=1}^n \Phi\left(\bar{a}_i(t+1), \bar{q}_{-i}(t)\right) - \frac{1}{t+1}\sum_{i=1}^n \Phi(\bar{q}_i(t), \bar{q}_{-i}(t)) + \zeta(t+1).$$

where we have explicitly written the first order terms of the expansion and collected the remaining terms in $\zeta(t+1)$. Note that the number of second order terms in the above expansion is finite and the terms are uniformly bounded since $\sup_{p \in \Delta^n} |\Phi(p)| < \infty$. Hence, there exists a positive constant B (independent of t) large enough such that $|\zeta(t+1| \leq B(t+1)^{-2}$ for all t. Thus,

$$\Phi(\bar{q}^n(t+1)) \geq \Phi(\bar{q}^n(t)) + \frac{1}{t+1}\sum_{i=1}^n \Phi\left(\bar{a}_i(t+1), \bar{q}_{-i}(t)\right) - \frac{1}{t+1}\sum_{i=1}^n \Phi(\bar{q}_i(t), \bar{q}_{-i}(t)) - \frac{B}{(t+1)^2}. \tag{8}$$

The permutation invariance and multilinearity of $\Phi(\cdot)$ permits a rearranging of terms. We use the notation $[a_j(t)]_i$ to indicate the action of player j at time t being played by player i. Observe that,

$$\sum_{i=1}^n \Phi(\bar{a}_i(t+1), \bar{q}_{-i}(t)) = \sum_{i=1}^n \Phi\left(\left[\frac{1}{n}\sum_{j=1}^n a_j(t+1)\right]_i, \bar{q}_{-i}(t)\right)$$

$$= \sum_{i=1}^n \frac{1}{n}\sum_{j=1}^n \Phi\left([a_j(t+1)]_i, \bar{q}_{-i}(t)\right)$$
$$= \sum_{i=1}^n \frac{1}{n}\sum_{j=1}^n \Phi\left([a_j(t+1)]_j, \bar{q}_{-j}(t)\right) = \sum_{j=1}^n \Phi\left(a_j(t+1), \bar{q}_{-j}(t)\right).$$

By similar reasoning it also holds that $\sum\limits_{i=1}^n \Phi\left(\bar{q}_i(t), \bar{q}_{-i}(t)\right) = \sum\limits_{j=1}^n \Phi\left(q_j(t), \bar{q}_{-j}(t)\right)$. Thus (8) can be expressed as,

$$\Phi\left(\bar{q}^n(t+1)\right) - \Phi\left(\bar{q}^n(t)\right) + \frac{B}{(t+1)^2} \geq \frac{1}{t+1}\sum_{i=1}^n \Phi\left(a_i(t+1), \bar{q}_{-i}(t)\right) - \frac{1}{t+1}\sum_{i=1}^n \Phi(q_i(t), \bar{q}_{-i}(t)). \tag{9}$$

Using (7) and the linearity of $F_i(\cdot)$ in Δ_i, note that

$$F_i(q_i(t+1)) - F_i(q_i(t)) = \frac{1}{t+1}\left(F_i(a_i(t+1)) - F_i(q_i(t))\right). \tag{10}$$

Combining (9) and (10) we get,

$$\begin{aligned}
&\sum_{i=1}^n \left[F_i(q_i(t+1)) - F_i(q_i(t))\right] \\
&\quad + \Phi\left(\bar{q}^n(t+1)\right) - \Phi\left(\bar{q}^n(t)\right) + \frac{B}{(t+1)^2} \\
\geq& \frac{1}{t+1}\sum_{i=1}^n F_i(a_i(t+1)) + \Phi\left(a_i(t+1), \bar{q}_{-i}(t)\right) \\
&- \frac{1}{t+1}\sum_{i=1}^n F_i(q_i(t)) + \Phi(q_i(t), \bar{q}_{-i}(t)) \\
=& \frac{1}{t+1}\sum_{i=1}^n \left[U_i(a_i(t+1), \bar{q}_{-i}(t)) - U_i(q_i(t), \bar{q}_{-i}(t))\right],
\end{aligned} \tag{11}$$

where the equality follows from the definition of $u_i(\cdot)$ in **A.3**. Under **A.1**, let Δ be the common mixed strategy space, and let $v_i^m(\cdot) : \Delta \to \mathbb{R}$,

$$v_i^m(f) := \max_{\alpha_i \in A_i} U_i(\alpha_i, f_{-i}) - U_i(f^n),$$

where $f_{-i} = (f, \ldots, f)$ is an $(n-1)$-tuple and $f^n = (f, \ldots, f)$ is an n-tuple, and let

$$L_i(t+1) := v_i^m\left(\hat{q}_i(t)\right) - U_i\left(a_i(t+1), \bar{q}_{-i}(t)\right).$$

Substituting $L_i(t+1)$ into (11) gives

$$\begin{aligned}
&\sum_{i=1}^n \left[F_i(q_i(t+1)) - F_i(q_i(t))\right] + \Phi\left(\bar{q}^n(t+1)\right) \\
&\quad - \Phi\left(\bar{q}^n(t)\right) + \frac{B}{(t+1)^2} + \frac{1}{t+1}\sum_{i=1}^n L_i(t+1) \\
\geq& \frac{1}{t+1}\sum_{i=1}^n \left(v_i^m\left(\hat{q}_i(t)\right) - U_i\left(q_i(t), \bar{q}_{-i}(t)\right)\right) = \frac{\alpha_{t+1}}{t+1},
\end{aligned} \tag{12}$$

where $\alpha_{t+1} := \sum_{i=1}^n \left(v_i^m\left(\hat{q}_i(t)\right) - U_i\left(q_i(t), \bar{q}_{-i}(t)\right)\right)$. Note that $U_i(\cdot)$ is multilinear and therefore locally Lipschitz continuous. Assumption **A.7** implies that $\{\hat{q}_i(t)\}_{t\geq 1}$ is contained in a compact subset of $\mathbb{R}^m$. Therefore, there exists a positive constant K (independent of t), such that $|U_i(a_i(t+1), \hat{q}_{-i}(t)) - U_i(a_i(t+1), \bar{q}_{-i}(t))| \leq K\|(a_i(t+1), \hat{q}_{-i}(t)) - (a_i(t+1), \bar{q}_{-i}(t))\|$, for all t. By assumption **A.7**, $\|\hat{q}_{-i}(t) - \bar{q}_{-i}(t)\| = O(\frac{\log t}{t})$, and hence $|U_i(a_i(t+1), \hat{q}_{-i}(t)) - U_i(a_i(t+1), \bar{q}_{-i}(t))| = O(\frac{\log t}{t})$. Note that by definition of a distributed ECFP process (6), $a_i(t+1) \in \arg\max_{\alpha_i \in A_i} U_i(\alpha_i, \hat{q}_{-i}(t))$, (and hence $v_i^m(\hat{q}_i(t)) = U_i(a_i(t+1), \hat{q}_{-i}(t))$) and the preceding is equivalent to

$$|v_i^m(\hat{q}_{-i}(t)) - U_i(a_i(t+1), \bar{q}_{-i}(t))| = O(\frac{\log t}{t}),$$

or equivalently, $L_i(t) = O(\frac{\log t}{t})$. In particular, note that $\sup_{T\geq 2} \sum_{t=2}^T \frac{L_i(t)}{t} < \infty$ is bounded above. Summing over $1 \leq t \leq T$ in (12),

$$\sum_{i=1}^n [F_i(q_i(T+1)) - F_i(q_i(1))] + \Phi(\bar{q}^n(T+1)) - \Phi(\bar{q}^n(1))$$
$$+ \sum_{t=1}^T \frac{B}{(t+1)^2} + \sum_{t=1}^T \sum_{i=1}^n \frac{L_i(t+1)}{t+1} \geq \sum_{t=1}^T \frac{\alpha_{t+1}}{t+1}.$$

Note that $\sum_{t=1}^T \frac{B}{(t+1)^2}$ is summable; therefore all terms on the left hand side are uniformly bounded above for all $T \geq 1$, and it follows that $\sum_{t=2}^T \frac{\alpha_t}{t} < \overline{B}$ is bounded above by some $\overline{B} \in \mathbb{R}$, for all $T \geq 2$.

Let $\beta_{t+1} := \sum_{i=1}^n \left[v_i^m(\bar{q}(t)) - U_i\left(q_i(t), \bar{q}_{-i}(t)\right)\right]$, and note that, by definition of $v_i^m(\cdot)$, $\beta_t \geq 0$ for all t. By [10] Lemma 4, there holds $|v_i^m(\hat{q}_i(t)) - v_i^m(\bar{q}(t))| = O\left(\frac{\log t}{t}\right)$. Thus, $|\alpha_t - \beta_t| = O\left(\frac{\log t}{t}\right)$, and hence by [10] Lemma 5, $\sum_{t=2}^T \frac{\beta_t}{t} < \infty$ converges as $T \to \infty$. By [10] Lemma 3 it follows that

$$\lim_{T\to\infty} \frac{\beta_2 + \beta_3 + \ldots + \beta_T}{T} = 0.$$

Subsequently, following reasoning analogous to [10] Lemma 6, we obtain for every $\varepsilon > 0$,

$$\lim_{T\to\infty} \frac{\#\{1 \leq t \leq T : q(t) \notin M_\varepsilon\}}{T} = 0.$$

Following reasoning analogous to [10] Lemma 7, this is equivalent to

$$\lim_{T\to\infty} \frac{\#\{1\leq t\leq T : q(t) \notin B_\delta(M)\}}{T} = 0, \ \forall \delta > 0.$$

Finally, following reasoning analogous to [10] Lemma 8, we obtain $d(q(t), MCE) \to 0$ as $t \to \infty$. □

B. Example Implementation

A detailed example implementation of ECFP in the distributed-information setting is studied in [10] Section VI. In the example implementation, players form their estimates $\hat{q}_i(t)$ by means of a dynamic consensus protocol (see [10], eqns. (11), (12)). It is shown that the example implementation meets **A.7** ([10], Lemma 2), and hence the results of Theorem 4 apply.

VI. Conclusions

The classical FP learning algorithm is not practical in large games due to demanding computation and communication requirements. The ECFP algorithm is an adaptation of FP that can mitigate computational issues. In order to address communication problems associated with FP in large games, a distributed-information setting is considered—a framework in which interagent communication is restricted a local subset of neighboring players.

We studied ECFP in a distributed-information setting, and in particular, we studied learning properties of the algorithm in terms of the typical notion of learning—convergence in empirical distribution. It was shown that in the distributed-information implementation of ECFP, players learn a mean-centric equilibrium (MCE)—a scalable equilibrium concept for large games—in terms of convergence in empirical distribution. Future work may include characterizing the efficiency of the set of MCE, and relaxing assumption **A.7** to allow for arbitrary decay rates.

References

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

[2] D. Fudenberg and D. K. Levine, *The theory of learning in games*. MIT press, 1998, vol. 2.

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

[4] L. S. Shapley, "Some topics in two-person games," *Advances in game theory*, vol. 52, pp. 1–29, 1964.

[5] J. S. Jordan, "Three problems in learning mixed-strategy Nash equilibria," *Games and Economic Behavior*, vol. 5, no. 3, pp. 368–386, Jul. 1993.

[6] D. P. Foster and H. P. Young, "On the nonconvergence of fictitious play in coordination games," *Games and Economic Behavior*, vol. 25, no. 1, pp. 79–96, Oct. 1998.

[7] 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, Jan. 1996.

[8] ——, "Potential Games," *Games and Economic Behavior*, vol. 14, no. 1, pp. 124–143, May 1996.

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

[10] B. Swenson, S. Kar, and J. Xavier, "Empirical centroid fictitious play: An approach for distributed learning in multi-agent games," *arXiv preprint arXiv:1304.4577*, 2014.

[11] ——, "Distributed learning in large-scale multi-agent games: A modified fictitious play approach," in *46th Asilomar Conference on Signals, Systems, and Computers*, Pacifc Grove, CA, USA, Nov. 4 - 7 2012, pp. 1490 – 1495.

[12] ——, "Mean-centric equilibriu: an equilibrium concept for learning in large-scale games," in *Global Conference on Signal and Information Processing (GlobalSIP), 2013 IEEE*. IEEE, Dec. 2013, pp. 571–574.

[13] A. G. Dimakis, S. Kar, J. M. Moura, M. G. Rabbat, and A. Scaglione, "Gossip algorithms for distributed signal processing," *Proceedings of the IEEE*, vol. 98, no. 11, pp. 1847–1864, Nov. 2010.

[14] S. Rajagopalan and D. Shah, "Distributed averaging in dynamic networks," *IEEE jornal of selected topics in signal processing*, vol. 5, no. 4, pp. 845–854, Aug. 2011.