Dynamic Fictitious Play, Dynamic Gradient Play, and Distributed Convergence to Nash Equilibria

Jeff S. Shamma and Gurdal Arslan

***Abstract*— We consider a continuous-time form of repeated matrix games in which player strategies evolve in reaction to opponent actions. Players observe each other's actions, but do not have access to other player utilities. Strategy evolution may be of the best response sort, as in fictitious play, or a gradient update. Such mechanisms are known to not necessarily converge. We introduce a form of "dynamic" fictitious and gradient play strategy update mechanisms. These mechanisms use derivative action in processing opponent actions and, in some cases, can lead to behavior converging to Nash equilibria in previously nonconvergent situations. We analyze convergence in the case of exact and approximate derivative measurements of the dynamic update mechanisms. In the ideal case of exact derivative measurements, we show that convergence to Nash equilibrium can always be achieved. In the case of approximate derivative measurements, we derive a characterization of local convergence that shows how the dynamic update mechanisms can converge if the traditional static counterparts do not. We primarily discuss two player games, but also outline extensions to multiplayer games. We illustrate these methods with convergent simulations of the well known Shapley and Jordan counterexamples.**

I. Overview

This paper considers a continuous-time form of a repeated game in which players continually update strategies in response to observations of opponent actions but without knowledge of opponent intentions. The primary objective is to understand how interacting players could converge to a Nash equilibrium, i.e., a set of strategies for which no player has a unilateral incentive to change.

The motivational setup is as follows. There are two players, each with a finite set of possible actions. Every time the game is played, each player selects an action according to a probability distribution that represents that player's *strategy*. The reward to each player, called the player's *utility*, depends on the actions taken by both players. While each player knows its own utility, these utilities are *not* shared between players.

Suppose that one player always used the same probability distribution to generate its action, i.e., the player maintained a constant strategy. Then the other player could, over time, via repeated play, learn this distribution by keeping a running average of opponent actions. Such running averages are called *empirical frequencies*. By playing the optimized best response to the observed empirical frequencies, the optimizing player will eventually converge to its own optimal response to the fixed strategy opponent.

Now if *both* players presumed that the other player is using a constant strategy, their strategy update mechanisms become *intertwined*. One such process is called *fictitious play* (FP). In this setting, players play the optimized best response to an opponent's empirical frequencies presuming (incorrectly) that the empirical frequency is representative of a constant probability distribution. The repeated game would be in equilibrium if the empirical frequencies converged. Since each player is employing the best response to observed behaviors, the game being in equilibrium would coincide with the players using strategies that are at a Nash equilibrium.

The procedure of FP was introduced in 1951 [9], [37] as a mechanism to compute Nash equilibria. There is a substantial body of literature on the topic. Related lines of research are discussed in the monographs [21], [**?**], [43] and the recent overview of [24].

Of particular concern is whether repeated play will indeed converge to a Nash equilibrium. A brief timeline of results that establish convergence of FP is as follows: 1951, two-player/zero-sum games [37]; 1961, two-player/two-move games [35]; 1993, noisy two-player/two-move games with a unique Nash equilibrium [20]; 1996, multiplayer games with identical player utilities [36]; 1999, noisy two-player/two-move games with countable Nash equilibria [6]; and 2003, two-player games where one player has only two moves [8].

It turns out that empirical frequencies need *not* converge. A counterexample due to Shapley in 1964 has two players with three moves each [40]. A 1993 counterexample due to Jordan has three players with two moves each [28].

In both the Shapley and Jordan counterexamples, the game under consideration admits a unique Nash equilibrium that is *completely mixed*, i.e., all moves have a positive probability of being played.

The concept of mixed Nash equilibria has received some scrutiny regarding its justification. The paper [38] raises various questions regarding finding an appropriate interpretation. Another concern is how a completely mixed Nash equilibrium could emerge as the outcome of a dynamic learning process among interacting players. The text [23, p. 22] states "game theory lacks a general and convincing argument that a Nash outcome will occur".

Indeed, there is a collection of negative results concerning the possibility of completely mixed equilibria emerging as a result of interactive behavior. The paper [11] shows that a broad class of strategy adjustment mechanisms (different from

Accepted for publication in *IEEE Transaction on Automatic Control*. Submitted October 13, 2003; revised October 13, 2004. Conference version submitted March 2003 to 42$^{\text{nd}}$ *IEEE Conference on Decision and Control* and appeared in conference proceedings, December 2003.

Research supported by AFOSR/MURI grant #F49620-01-1-0361 and summer support by the UF Graduate Engineering Research Center.

J.S. Shamma is with the Mechanical and Aerospace Engineering Department, University of California, Los Angeles 90095 USA (e-mail: shamma@ucla.edu).

G. Arslan is with the Electrical Engineering Department, University of Hawaii, Manoa 96822 USA (e-mail: gurdal@hawaii.edu).

FP) cannot converge to a mixed equilibrium. The paper [31] shows that "almost all" games in which players have more than two moves cannot converge to completely mixed equilibria in best response FP. Nonconvergence issues are also discussed in [12], [19], [26], [42].

In particular, the paper [26] shows that a generalized version of the Jordan game will not exhibit convergence for *any* strategy adjustment mechanism—not just a best response mechanism—provided that players do not share their utility functions and update mechanisms are static functions of empirical frequencies.

There are methods akin to parallel random search that are able to find a neighborhood of a Nash equilibrium. Relevant works are [17], [18], [27]. These methods conceptually differ from FP in that the driving mechanism is distributed randomized search in the strategy space as opposed to gradual strategic adjustment.

Contrary to the case of Nash equilibria, there are methods [16], [22], [25] that are guaranteed for all games to converge to the larger set of so-called *correlated equilibria*, which is a convex set that contains the set of Nash equilibria. These are "regret based" algorithms that revisit past decisions in an effort to evaluate what could have been a more fruitful course of action. See [24] for an extensive discussion.

An important assumption in [26] is that update mechanisms employ *static* functions of empirical frequencies. In this paper, we explore the possibility of *dynamic* functions of opponent actions in the spirit of dynamic compensation for feedback stabilization. It is well known that static output feedback need not be stabilizing, while dynamic output feedback generally can be stabilizing. We wish to explore this possibility in the context of repeated games.

The present approach is to view the problem as one of feedback stabilization. Contrary to standard feedback stabilization scenarios, one seeks to stabilize to an equilibrium point that is *unknown*, but must emerge through the non-cooperative interaction of repeated play. One paper that takes a similar feedback stabilization viewpoint is [14], in which an integral term is employed in the strategy update mechanism and a sufficient condition for convergence is derived.

In this paper, we focus on the use of *derivative* action. Derivative action, standard in classical control systems, is also a key component of biological motor control system models [34].

We will employ a strategy update mechanism that closely resembles traditional mechanisms but use both the empirical frequencies and their (approximate) derivatives. As such, the new approach differs conceptually from both aforementioned approaches of randomized search and no-regret methods.

We will establish convergence to Nash equilibrium in the ideal case of exact derivative measurements and near convergence in case of approximate derivative measurements. We will show how the use of approximate differentiators may or may not allow one to recover the ideal case. In addition to "best response" FP, we will also consider gradient-like "better response" strategy update mechanisms (e.g., [13]). We will illustrate all of these methods on the Shapley game. We outline the framework for multiplayer games, and illustrate convergence on the Jordan game.

Two papers that are closely related are [4], [10]. Reference [4] considers two dynamic processes. The first is that players use a strategy that is the best response to the previous action of the opponent. The second is a "relaxation" in which players use this best response only to adjust its current strategy, thereby introducing some inertia. This relaxation, which may be viewed as a sort of dynamic compensation, may have improved convergence properties. Reference [10] considers zero-sum games played in intervals. Players adjust their strategy based on an approximate forecast of the opponents strategy, which is reminiscent of the use of derivative action as a myopic predictor.

Other related papers with positive convergence results are [29], [33]. In [29], all players make a "calibrated" [16] forecast of their *joint* action and use this to derive a forecasted best response. This results in near convergence to the convex hull of Nash equilibria. In [33], all players use different time-scales to adjust their strategies. The authors show that such multiscale dynamics can enable convergence to a Nash equilibrium in certain cases, including the Shapley and Jordan games.

The remainder of this paper is organized as follows. Section 2 reviews standard fictitious play. Section 3 introduces the notion of "derivative action" fictitious play and analyzes the both the ideal case of exact derivative measurements and the non-ideal case of approximate differentiators. Section 4 introduces gradient play and analyzes exact and approximate implementations of derivative action. Section 5 discusses extensions to multi-player games. Finally, Section 6 presents concluding remarks.

Notation

— For $i \in \{1, 2, \ldots, n\}$, $-i$ denotes the complementary set $\{1, \ldots, i-1, i+1, \ldots, n\}$.
— Boldface $\mathbf{1}$ denotes the vector $\begin{pmatrix} 1 \\ \vdots \\ 1 \end{pmatrix} \in \mathcal{R}^n$.
— For $x \in \mathcal{R}^n$, $|x|$ denotes the usual 2-norm, i.e., $\sqrt{x^T x}$.
— For $x \in \mathcal{R}^n$, $\mathrm{diag}(x)$ denotes the diagonal $n \times n$ matrix with elements taken from x.
— $\Delta(n)$ denotes the simplex in $\mathcal{R}^n$, i.e.,
$$\left\{ s \in \mathcal{R}^n | s \geq 0 \text{ componentwise, and } \mathbf{1}^T s = 1 \right\}.$$
— $\mathrm{Int}(\Delta(n))$ denotes the set of interior points of a simplex, i.e., $s > 0$ componentwise.
— $\Pi_\Delta : \mathcal{R}^n \to \Delta(n)$ denotes the projection to the simplex,
$$\Pi_\Delta[x] = \arg \min_{s \in \Delta(n)} |x - s| .$$
— $\mathbf{v}_i \in \Delta(n)$ denotes the i^{th} vertex of the simplex $\Delta(n)$, i.e., the vector whose i^{th} term equals 1 and remaining terms equal 0.
— $\mathcal{H} : \mathrm{Int}(\Delta(n)) \to \mathcal{R}$ denotes the entropy function
$$\mathcal{H}(s) = -s^T \log(s).$$
— $\sigma : \mathcal{R}^n \to \mathrm{Int}(\Delta(n))$ denotes the "logit" or "soft-max" function
$$(\sigma(x))_i = \frac{e^{x_i}}{e^{x_1} + \cdots + e^{x_n}}.$$

This function is continuously differentiable. The Jacobian matrix of partial derivatives, denoted $\nabla\sigma(\cdot)$, is

$$\nabla\sigma(x) = \mathrm{diag}(\sigma(x)) - \sigma(x)\sigma^T(x).$$

II. FP Setup

A. Static Game

We consider a two-player game with players $\mathcal{P}_1$ and $\mathcal{P}_2$. Each player, $\mathcal{P}_i$, selects a strategy, $p_i \in \Delta(m_i)$, for given positive integers m_i, and receives a real-valued reward according to the utility function $\mathcal{U}_i(p_i, p_{-i})$. These utility functions take the form

$$\begin{aligned}\mathcal{U}_1(p_1, p_2) &= p_1^T M_1 p_2 + \tau\mathcal{H}(p_1)\\ \mathcal{U}_2(p_2, p_1) &= p_2^T M_2 p_1 + \tau\mathcal{H}(p_2),\end{aligned}$$

characterized by matrices M_i of appropriate dimension and $\tau \geq 0$.

The standard interpretation is that the p_i represent probabilistic strategies. Each player selects an integer action $a_i \in \{1, \ldots, m_i\}$ according to the probability distribution p_i. The reward to player $\mathcal{P}_i$ is

$$\mathbf{v}_{a_i}^T M_i \mathbf{v}_{a_{-i}} + \tau\mathcal{H}(p_i),$$

i.e., the reward to player $\mathcal{P}_i$ is the element of M_i in the a_i^{th} row and a_{-i}^{th} column, plus the weighted entropy of its strategy. For a given strategy pair, (p_1, p_2), the utilities represent the expected rewards

$$\mathcal{U}_i(p_i, p_{-i}) = E\left[\mathbf{v}_{a_i}^T M_i \mathbf{v}_{a_{-i}}\right] + \tau\mathcal{H}(p_i).$$

Define the *best response* mappings

$$\beta_i : \Delta(m_{-i}) \to \Delta(m_i)$$

by

$$\beta_i(p_{-i}) = \arg\max_{p_i \in \Delta(m_i)} \mathcal{U}_i(p_i, p_{-i}).$$

For $\tau > 0$, the best response turns out to be the logit or soft-max function (see Notation section)

$$\beta_i(p_{-i}) = \sigma(M_i p_{-i}/\tau).$$

For $\tau = 0$, the best response mapping can be set-valued.

A Nash equilibrium is a pair $(p_1^*, p_2^*) \in \Delta(m_1) \times \Delta(m_2)$ such that for all $p_i \in \Delta(m_i)$,

$$\mathcal{U}_i(p_i, p_{-i}^*) \leq \mathcal{U}_i(p_i^*, p_{-i}^*), \tag{1}$$

i.e., each player has no incentive to deviate from an equilibrium strategy provided that the other player maintains an equilibrium strategy. In terms of the best response mappings, a Nash equilibrium is a pair (p_1^*, p_2^*) such that

$$p_i^* = \beta_i(p_{-i}^*).$$

A Nash equilibrium is *completely mixed* if each component is strictly positive, i.e., $p_i^* \in \mathrm{Int}(\Delta(m_i))$. This distinction becomes relevant only in the case where $\tau = 0$.

B. Discrete-time FP

Now suppose that the game is repeated at every time $k \in \{0, 1, 2, \ldots\}$. In particular, we are interested in an "evolutionary" version of the game in which the strategies at time k, denoted by $p_i(k)$, are selected in response to the entire prior history of an opponent's actions.

Towards this end, let $a_i(k)$ denote the action of player $\mathcal{P}_i$ at time k, chosen according to the probability distribution $p_i(k)$, and let $\mathbf{v}_{a_i(k)} \in \Delta(m_i)$ denote the corresponding simplex vertex. The *empirical frequency*, $q_i(k)$, of player $\mathcal{P}_i$ is defined as the running average of the actions of player $\mathcal{P}_i$, which can be computed by the recursion

$$q_i(k+1) = q_i(k) + \frac{1}{k+1}(\mathbf{v}_{a_i(k)} - q_i(k)).$$

In discrete-time FP, the strategy of player $\mathcal{P}_i$ at time k is the optimal response to the running average of the opponent's actions, i.e.,

$$p_i(k) = \beta_i(q_{-i}(k)).$$

The case with $\tau = 0$ corresponds to classical FP. Setting τ positive rewards randomization, thereby forcing mixed strategies. As τ approaches zero, the best response mappings approximate selecting the maximal element since the probability of selecting a maximal element approaches one. The case with τ positive, often referred to as stochastic FP, can be viewed as a smoothed version of the matrix game [20], in which rewards are subject to random perturbations. Other interpretations, including connections to information theory, are discussed in [44].

C. Continuous-time FP

Now consider the continuous-time dynamics,

$$\begin{aligned}\dot{q}_1(t) &= \beta_1(q_2(t)) - q_1(t)\\ \dot{q}_2(t) &= \beta_2(q_1(t)) - q_2(t).\end{aligned} \tag{2}$$

We will call these equations *continuous-time FP*. These are the dynamics obtained by viewing discrete-time FP as stochastic approximation iterations and applying associated ordinary differential equation (ODE) analysis methods [5], [7], [32].

III. Dynamic FP

Standard discrete-time and continuous-time FP assume that the empirical frequencies, $q_i(\cdot)$, are available to all players, and the strategy of each player is the best response to the opponent's empirical frequency. This strategy is a *static* function of the empirical frequencies.

We wish to explore the possibility of *dynamic* processing of empirical frequencies. In standard control terminology, the analogous statement is that we wish to investigate the possible advantage of dynamic output feedback versus static output feedback. Clearly dynamic feedback is superior in a general setting, but our question is focused on the specific issue of FP and convergence of empirical frequencies.

In the entire discussion of dynamic FP, we will only consider $\tau > 0$.

A. Derivative Action

In continuous-time FP, the empirical frequencies are available to all players, and the strategy of each player is the best response to the opponent's empirical frequency, i.e.,

$$p_i(t) = \beta_i(q_{-i}(t)),$$

where $p_i(t)$ denotes the strategy of player $\mathcal{P}_i$ at time t.

Suppose now that in addition to empirical frequencies being available to all players, empirical frequency *derivatives*, $\dot{q}_i(t)$, are also available. Now consider

$$p_i(t) = \beta_i(q_{-i}(t) + \gamma \dot{q}_{-i}(t)),$$

i.e., each player's strategy is a best response to a combination of empirical frequencies *and* a weighted derivative of empirical frequencies.

This modification is very much in the spirit of standard PID controllers in engineered systems as well as motor control models [34] in biological systems. The classical control interpretation is that the derivative term serves as a short term prediction of the opponent's strategy, since

$$q_{-i}(t) + \gamma \dot{q}_{-i}(t) \approx q_{-i}(t + \gamma).$$

In this regard, the use of derivative action may be interpreted as using the best response to a forecasted opponent strategy.

This modification leads to the following (implicit) differential equation

$$\begin{aligned} \dot{q}_1 &= \beta_1(q_2 + \gamma \dot{q}_2) - q_1 \\ \dot{q}_2 &= \beta_2(q_1 + \gamma \dot{q}_1) - q_2, \end{aligned} \tag{3}$$

which we will refer to as "exact" *derivative action FP* (DAFP).

In actuality, the derivative is not directly measurable, but must be reconstructed from empirical frequency measurements. Towards this end, consider

$$\begin{aligned} \dot{q}_1 &= \beta_1(q_2 + \gamma\lambda(q_2 - r_2)) - q_1 \\ \dot{q}_2 &= \beta_2(q_1 + \gamma\lambda(q_1 - r_1)) - q_2 \\ \dot{r}_1 &= \lambda(q_1 - r_1) \\ \dot{r}_2 &= \lambda(q_2 - r_2), \end{aligned} \tag{4}$$

with $\lambda > 0$. An alternative expression is

$$\begin{aligned} \dot{q}_1 &= \beta_1(q_2 + \gamma \dot{r}_2) - q_1 \\ \dot{q}_2 &= \beta_2(q_1 + \gamma \dot{r}_1) - q_2, \end{aligned}$$

which we will refer to as "approximate" DAFP. The variables, r_i, are "filtered" versions of the empirical frequencies. The intention is that as λ increases, $\dot{r}_i$ closely tracks $\dot{q}_i$.

In the following sections, we will examine both exact and approximate DAFP. We will first focus on the case of unity derivative gain, i.e., $\gamma = 1$, which has a special interpretation. We will show that the positive results of exact DAFP need not be recovered through approximate DAFP. Accordingly, we will present a separate local analysis of approximate DAFP for general derivative gain values.

B. Exact DAFP with Unity Derivative Gain ($\gamma = 1$)

The particular case of derivative gain $\gamma = 1$ has the special interpretation of "system inversion", as illustrated in the block diagram of Figure 1. In words, the case of $\gamma = 1$ seeks to play a best response against the *current* strategy, as opposed to the empirical frequencies which reflect low-passed filtered strategies.

In case $\gamma = 1$, the equations of exact DAFP (3) become

$$\begin{aligned} \dot{q}_1 &= \beta_1(q_2 + \dot{q}_2) - q_1 \\ \dot{q}_2 &= \beta_2(q_1 + \dot{q}_1) - q_2. \end{aligned}$$

As previously noted, these form implicit differential equations, for which we will *assume* existence of solutions. Ultimately, exact DAFP will be replaced by the well posed approximate DAFP, so this assumption is not critical. Rather, exact DAFP will reveal an underlying structure that will enable the forthcoming convergence analysis.

Towards this end, introduce the variables

$$\begin{aligned} z_1 &= q_1 + \dot{q}_1 \\ z_2 &= q_2 + \dot{q}_2, \end{aligned}$$

and let

$$T : \mathcal{R}^{m_1} \times \mathcal{R}^{m_2} \to \Delta(m_1) \times \Delta(m_2)$$

be the mapping

$$\begin{pmatrix} z_1 \\ z_2 \end{pmatrix} \mapsto \begin{pmatrix} \beta_1(z_2) \\ \beta_2(z_1) \end{pmatrix}. \tag{5}$$

Let $z = \begin{pmatrix} z_1 \\ z_2 \end{pmatrix}$. Then we can restate exact DAFP dynamics (3) as

$$z = T(z),$$

i.e., exact DAFP must evolve over fixed points of T.

It turns out that these fixed points are Nash equilibria of the original game.

Proposition 3.1: The following are equivalent:

- $(z_1, z_2) \in \mathcal{R}^{m_1} \times \mathcal{R}^{m_2}$ *is a fixed point of* T *in (5).*
- $(z_1, z_2) \in \Delta(m_1) \times \Delta(m_2)$ *is a Nash equilibrium satisfying (1).*

Let

$$Q^* \subset \Delta(m_1) \times \Delta(m_2)$$

denote the set of Nash equilibria satisfying (1).

The following is an immediate consequences of Proposition 3.1.

Theorem 3.1: Any solution of exact DAFP dynamics (3) satisfies the differential inclusion

$$\begin{pmatrix} \dot{q}_1 \\ \dot{q}_2 \end{pmatrix} \in \begin{pmatrix} -q_1 \\ -q_2 \end{pmatrix} + Q^*.$$

In the case of a unique Nash equilibrium $Q^* = \{(q_1^*, q_2^*)\}$, *the unique solution to exact DAFP (3) is*

$$\begin{aligned} \dot{q}_1 &= -q_1 + q_1^* \\ \dot{q}_2 &= -q_2 + q_2^*, \end{aligned}$$

which converges (exponentially) to the unique Nash equilibrium.

In the multiple Nash equilibrium case, Theorem 3.1 does not, in itself, guarantee convergence of empirical frequencies. This is because only a subset of the entirety of solutions of the associated differential inclusion do converge to a Nash equilibrium (e.g., those solutions with continuous time-derivatives).

C. Noisy Derivative Measurements with Unity Derivative Gain ($\gamma = 1$)

Suppose we can make noisy measurements of the empirical frequency derivatives. Then exact DAFP can be written as

$$\begin{aligned} \dot{q}_1 &= \beta_1(q_2 + \dot{q}_2 + e_2) - q_1 \\ \dot{q}_2 &= \beta_2(q_1 + \dot{q}_1 + e_1) - q_2. \end{aligned} \tag{6}$$

The new variables, $e_i(t)$, denote the derivative measurement errors.

Let us assume that there exist unique solutions to (6) without specifying at this point how the e_i would be generated.

We will show that under certain conditions, the empirical frequencies converge to a neighborhood of the set of Nash equilibria, where the size of the neighborhood depends on the accuracy of the derivative approximation.

Introduce the following extension, $T_{\rm e}$, of the mapping T, defined in (5),

$$T_{\rm e} : \mathcal{R}^{m_1+m_2} \times \mathcal{R}^{m_1+m_2} \to \Delta(m_1) \times \Delta(m_2),$$

where

$$T_{\rm e}(z, e) = T(z + e).$$

Then we can write approximate derivative action FP (6) as

$$z = T_{\rm e}(z, e),$$

where, as before,

$$z_i = \dot{q}_i + q_i.$$

Lemma 3.1: Let $(q_1^*, q_2^*) \in Q^*$ be a Nash equilibrium. Assume that the matrix

$$\begin{pmatrix} -I & \frac{1}{\tau}\nabla\sigma(M_1 q_2^*/\tau)M_1 \\ \frac{1}{\tau}\nabla\sigma(M_2 q_1^*/\tau)M_2 & -I \end{pmatrix}$$

is nonsingular. Then there exists a $\delta > 0$ and unique continuously differentiable function, $\phi : \mathcal{R}^{m_1} \times \mathcal{R}_2^{m} \to \Delta(m_1) \times \Delta(m_2)$ defined on an δ–neighborhood of the origin such that

$$\phi(e) = T_{\rm e}(\phi(e), e).$$

Proof: At a Nash equilibrium, $q^* = \begin{pmatrix} q_1^* \\ q_2^* \end{pmatrix}$, the extended mapping satisfies

$$0 = T_{\rm e}(q^*, 0) - q^*.$$

Under the assumed nonsingularity, the function $T_{\rm e}(z, e) - z$ satisfies the conditions of the implicit function theorem [15]. ■

Theorem 3.2: Assume that Q^* is a finite set of Nash equilibria, each of which satisfies the nonsingularity assumption of Proposition 3.1. Suppose (q_1, q_2, e_1, e_2) satisfy noisy derivative measurement DAFP (6) with $\dot{q}_i(\cdot)$ continuous. For any $\varepsilon > 0$, there exists a $\delta > 0$ such that if $(e_1(t), e_2(t))$ eventually remain within a δ-neighborhood of the origin, then $(q_1(t), q_2(t))$ eventually remain within an ε-neighborhood of a single Nash equilibrium, i.e.,

$$\limsup_{t\geq 0} |e(t)| < \delta$$

implies

$$\limsup_{t\geq 0} |(q_1(t), q_2(t)) - (q_1^*, q_2^*)| < \varepsilon$$

for some $q^* \in Q^*$.

Proof: Enumerate the set of Nash equilibrium points,

$$Q^* = \left\{(q_1^*, q_2^*)^j : j = 1, 2, \ldots, N\right\}.$$

Let δ^j and $\phi^j(\cdot)$ denote the corresponding parameters and functions in Lemma 3.1. Pick

$$\delta < \min_j \delta^j$$

so that $|e| < \delta$ implies that for all j,

$$\left|\phi^j(e) - (q_1^*, q_2^*)^j\right| < \varepsilon.$$

Since the ϕ^j are continuous and the Nash equilibrium points are isolated, we can assume that the above ε-neighborhoods of equilibrium points do not overlap.

Now suppose that at some time $T > 0$,

$$\sup_{t\geq T} |e(t)| < \delta$$

Then necessarily for any $t \geq T$,

$$z(t) = T_{\rm e}(z(t), e(t)) \Longrightarrow z(t) = \phi^{j(t)}(e(t))$$

for some $j(t)$. Since the assumed continuity of $\dot{q}_i(\cdot)$ in turn implies the continuity of $z(\cdot)$, the associated $j(t)$ cannot change. Such a change would require a discontinuous evolution of $z(t)$ since the ε-neighborhoods of the Nash equilibria do not overlap.

Finally, from time T onward,

$$\begin{pmatrix} \dot{q}_1(t) \\ \dot{q}_2(t) \end{pmatrix} = -\begin{pmatrix} q_1(t) \\ q_2(t) \end{pmatrix} + \phi^{j(T)}(e(t)).$$

Standard arguments then show that $(q_1(t), q_2(t))$ eventually reach an ε-neighborhood of the associated Nash equilibrium $(q_1^*, q_2^*)^j$. ■

In case of a unique Nash equilibrium, the continuity assumption on the derivatives $\dot{q}_i(\cdot)$ may be dropped.

As we will see, the premise of Theorem 3.2 is may be prematurely optimistic. The reason is that the two-player interactions may prevent reconstruction of the derivative up to a small bounded error.

D. Approximate DAFP with Unity Derivative Gain ($\gamma = 1$)

Theorem 3.2 establishes a sort of continuity result for approximate derivative action FP. Namely, it is possible to converge to an arbitrary neighborhood of the set of Nash equilibrium points provided that we can construct sufficiently accurate approximations of empirical frequency derivatives.

Towards this end, we now consider approximate DAFP given by

$$\begin{aligned}\dot{q}_1 &= \beta_1(q_2 + \dot{r}_2) - q_1\\ \dot{q}_2 &= \beta_2(q_1 + \dot{r}_1) - q_2\\ \dot{r}_1 &= \lambda(q_1 - r_1)\\ \dot{r}_2 &= \lambda(q_2 - r_2)\end{aligned} \tag{7}$$

The construction of $\dot{r}_i$ depends on the empirical frequencies, which are measured quantities.

The motivation of approximate DAFP is that for large $\lambda > 0$, the quantity $\dot{r}_i$ serves as an estimate of $\dot{q}_i$. Indeed, it is easy to show that if

$$\sup_{t\geq 0} |\ddot{q}_i(t)| \leq \ddot{q}_{\max}$$

then

$$\limsup_{t\to\infty} |\dot{q}_i - \dot{r}_i| \leq \frac{1}{\lambda}\ddot{q}_{\max}$$

Unfortunately, such intuition may or may not hold. The problem is that the approximation error associated with reconstructing $\dot{q}_i$ is proportional to the magnitude of the *second* derivative $\ddot{q}_i$. These second derivatives, in turn, involve the derivatives $\dot{r}_i$, which of course involve λ. So as λ increases, the second derivative magnitudes $\ddot{q}_i$ can also increase, thereby cancelling the desired effect of superior tracking.

We will further investigate the obstacle of approximate differentiators by considering solutions to (7) with progressively larger values of λ. First, we state the following theorem from [3].

Theorem 3.3 ([3], Section 0.3, Theorem 4): Consider a sequence of absolutely continuous functions $x_k(\cdot)$ from a compact interval $[T_1, T_2]$ of $\mathcal{R}$ to $\mathcal{R}^n$ such that the sets $\{x_k(t)\}$ and $\{\dot{x}_k(t)\}$ are uniformly bounded for all $k > 0$ and $t \in [T_1, T_2]$. Then there exists a subsequence, again denoted by $x_k(\cdot)$, converging to an absolutely continuous function $x(\cdot)$ in the sense that

1) *$x_k(\cdot)$ converges uniformly to $x(\cdot)$ on $[T_1, T_2]$.*
2) *$\dot{x}_k(\cdot)$ converges weakly to $\dot{x}(\cdot)$ in $L^1([T_1, T_2], \mathcal{R}^n)$.*

Theorem 3.4: Let $(q_i^\lambda, r_i^\lambda)$ denote the λ-dependent solutions to (7). For any compact interval, $[T_1, T_2] \subset \mathcal{R}^+$, with $T_1 > 0$, there exist an unbounded increasing sequence $\{\lambda_k\}$ and absolutely continuous functions $\overline{q}_i$ with derivatives $\dot{\overline{q}}_i$ such that

1) *$q_i^{\lambda_k}$ and $r_i^{\lambda_k}$ both converge to $\overline{q}_i$ uniformly on $[T_1, T_2]$.*
2) *$\dot{q}_i^{\lambda_k}$ and $\dot{r}_i^{\lambda_k}$ both converge weakly to $\dot{\overline{q}}_i$ in $L^1([T_1, T_2], \mathcal{R}^{m_i})$.*

Proof: For any λ, the function $q_i^\lambda(t)$ is clearly uniformly bounded over $t \in \mathcal{R}^+$ since it evolves in the simplex. Since $\dot{q}_i^\lambda$ is formed by the difference of two simplex elements, it is similarly uniformly bounded. Standard Lyapunov analysis shows that for any λ

$$\left|q_i^\lambda(t) - r_i^\lambda(t)\right| \leq e^{-\lambda t}\left|q_i^\lambda(0) - r_i^\lambda(0)\right| + \frac{1}{\lambda}\sup_{\tau\geq 0}\left|\dot{q}_i^\lambda(\tau)\right|,$$

and so r_i^λ and $\dot{r}_i^\lambda$ are also uniformly bounded over $t \in \mathcal{R}^+$. Here we naturally assume that initial conditions are restricted to the simplex of appropriate dimension. As a result, for any increasing unbounded sequence $\{\lambda_k\}$, the sequence of functions $(q_i^\lambda, r_i^\lambda)$, over any $[T_1, T_2]$ with $T_1 > 0$, form a bounded equicontinuous family and satisfy the hypotheses of the Theorem 3.3. Therefore, there exists a subsequence, which we relabel λ_k, and absolutely continuous functions $\overline{q}_i$ and $\overline{r}_i$ with derivatives $\dot{\overline{q}}_i$ and $\dot{\overline{r}}_i$, respectively, such that

1) $q_i^{\lambda_k}$ converges to $\overline{q}_i$ uniformly on $[T_1, T_2]$.
2) $\dot{q}_i^{\lambda_k}$ converges weakly to $\dot{\overline{q}}_i$ in $L^1([T_1, T_2], \mathcal{R}^{m_i})$.
3) $r_i^{\lambda_k}$ converges to $\overline{r}_i$ uniformly on $[T_1, T_2]$.
4) $\dot{r}_i^{\lambda_k}$ converges weakly to $\dot{\overline{r}}_i$ in $L^1([T_1, T_2], \mathcal{R}^{m_i})$.

It follows that the sequence

$$\frac{1}{\lambda_k}\dot{r}_i^{\lambda_k} = q_i^{\lambda_k} - r_i^{\lambda_k}$$

is converging uniformly on $[T_1, T_2]$ to $\overline{q}_i - \overline{r}_i$. Since the λ_i are unbounded, it must follow that

$$\overline{q}_i - \overline{r}_i = 0.$$

This, in turn, implies that

$$\dot{\overline{q}}_i - \dot{\overline{r}}_i = 0,$$

which completes the proof. ∎

Theorem 3.5: In the context of Theorem 3.4, let $\overline{q}_i$ and $\dot{\overline{q}}_i$ be the respective limits of $q_i^{\lambda_k}$ and $\dot{q}_i^{\lambda_k}$ on the compact interval $[T_1, T_2]$. Define

$$b_i^\lambda(t) = \beta_i(q_{-i}^\lambda(t) + \dot{r}_{-i}^\lambda(t))$$

and

$$\overline{b}_i = \dot{\overline{q}}_i + \overline{q}_i.$$

Then the sequence $b_i^{\lambda_k}$ converges weakly to $\overline{b}_i$ in $L^1([T_1, T_2], \mathcal{R}^{m_i})$. Furthermore, if

$$\overline{b}_i(t) = \beta_i(\overline{q}_{-i}(t) + \dot{\overline{q}}_{-i}(t)), \tag{8}$$

then $(\overline{q}_1, \overline{q}_2)$ are solutions to exact DAFP dynamics (3) on $[T_1, T_2]$.

Proof: The weak convergence of the sequence $b_i^{\lambda_k}$ follows immediately from Theorem 3.4. Furthermore, on the interval $[T_1, T_2]$,

$$\dot{\overline{q}}_i(t) = \overline{b}_i(t) - \overline{q}_i(t)$$

Under the assumed equality condition (8), it follows that

$$\dot{\overline{q}}_i(t) = \beta_i(\overline{q}_{-i}(t) + \dot{\overline{q}}_{-i}(t)) - \overline{q}_i(t),$$

as desired. ∎

Theorem 3.5 establishes that using increasing values of λ can converge to a solution of the derivative action FP dynamics (3) under the equality assumption (8). This equality assumption is essentially a requirement of weak continuity of the function β_i viewed as an operator on $L^1([T_1, T_2], \mathcal{R}^{m_i})$. Even though β_i is uniformly continuous as a function over the simplex, this need not imply weak continuity as an operator. Indeed, asymmetries due to nonlinearities can destroy the desired weak continuity.

The convergence discussed in Theorems 3.4–3.5 refers to functional convergence as λ increases. They need not imply, for a fixed λ, convergence as time increases. Still, one may infer implications regarding such convergence in time. For example, if the weak continuity condition (8) holds, and if there is unique Nash equilibrium, then the limiting functional behavior is indeed exponential convergence to the Nash equilibrium.

E. Simulations: Derivative Action FP on the Shapley game

A counterexample of empirical frequency convergence in FP due to Shapley [40] is

$$M_1 = M_2 = \begin{pmatrix} 0 & 1 & 0 \\ 0 & 0 & 1 \\ 1 & 0 & 0 \end{pmatrix}.$$

Figure 2 shows the discrete-time and continuous-time evolution of the empirical frequencies of player $\mathcal{P}_1$ with the above matrices and $\tau = 0.01$. In discrete-time FP, the empirical frequencies exhibit an oscillatory behavior with an ever increasing period. In continuous-time, the oscillatory behavior is still present, but with a regular period.

Figure 3 shows the empirical frequency response of player $\mathcal{P}_1$ for the Shapley game with $\tau = 0.01$ under derivative action FP using approximate differentiators with $\lambda = 1, 10, 100$. The empirical frequencies approach to the (unique) Nash equilibrium $(1/3, 1/3, 1/3)$, and as λ increases, the oscillations associated with standard FP are progressively reduced. This behavior is mirrored in discrete-time as well. The plots are omitted here for the sake of brevity.

Two comments are in order regarding the simulations. First, the linearized dynamics near the Nash equilibrium are *not* exponentially stable in the continuous-time simulations. This underscores that the apparent "convergence" is to the limiting behavior of exact DAFP for increasing λ, but not convergence in time for increasing t. Second, a histogram of the *joint* actions of the two players in the discrete-time simulations reveals that the players are following a particular *correlated equilibrium* [23] of the Shapley example. The average payoff for each player is not consistent with the expected payoff from the associated Nash equilibrium. In this case, the average payoff is greater. The issue of consistency as related to learning in games is discussed further in [21].

Now consider a modified Shapley game

$$M_1 = \begin{pmatrix} 0 & 3 & 0 \\ 0 & 0 & 1 \\ 1 & 0 & 0 \end{pmatrix}, \quad M_2 = \begin{pmatrix} 0 & 1 & 0 \\ 0 & 0 & 1 \\ 1 & 0 & 0 \end{pmatrix}.$$

Note that this modification destroys a symmetry between players so that $(1/3, 1/3, 1/3)$ is no longer a Nash equilibrium. Rather, the new Nash equilibrium is (approximately)

$$q_1^* = \begin{pmatrix} 1/3 \\ 1/3 \\ 1/3 \end{pmatrix}, \quad q_2^* = \begin{pmatrix} 3/7 \\ 1/7 \\ 3/7 \end{pmatrix}.$$

Figure 4 shows the empirical frequency responses under derivative action FP with approximate differentiators with $\lambda = 100$. The constant dashed lines are the desired steady state values of the Nash equilibrium. Although the empirical frequencies apparently converge in time, they are converging to the wrong values

$$q_1(5) \approx \begin{pmatrix} 0.3988 \\ 0.3627 \\ 0.2385 \end{pmatrix}, \quad q_2(5) \approx \begin{pmatrix} 0.3650 \\ 0.1853 \\ 0.4497 \end{pmatrix}.$$

Here we see the effects of the lack of weak continuity in Theorem 3.5. Although the derivative is apparently converging weakly to the zero function, the empirical frequencies are not evolving towards a Nash equilibrium. Increasing λ did not improve this error in simulations.

It is possible to reduce, but not eliminate, this error by using a modified approximate differentiator. This is discussed in [39]. The following section shows that it is possible to eliminate this error using other than unity derivative gain.

F. Approximate DAFP with General Derivative Gain, $\gamma > 0$

We continue the analysis of approximate DAFP (4), but with arbitrary $\gamma > 0$. We will give a complete characterization of the values of γ that result in *local* asymptotic stability of a Nash equilibrium for large values of $\lambda > 0$. In the process, we will characterize when the introduction of derivative action in FP can enable the local asymptotic stability of a Nash equilibrium when standard FP is unstable. Interestingly, we will show that the case of unity gain, $\gamma = 1$, never leads to asymptotic stability. This was evident in the original Shapley game in that the apparent "convergence" to Nash equilibria was actually low amplitude oscillations.

Define N to be an orthonormal matrix whose columns span the null space of the row vector $\mathbf{1}^T \in \mathcal{R}^m$, i.e.,

$$\mathbf{1}^T N = 0 \quad \text{and} \quad N^T N = I. \tag{9}$$

For notational simplicity, we will not denote the dimension of N explicitly. Rather, it will be apparent from context.

A Nash equilibrium of (q_1^*, q_2^*) leads to an equilibrium point $(q_1^*, q_2^*, q_1^*, q_2^*)$ of approximate DAFP (4). Since q_i evolves in the unit simplex, we can write

$$q_i(t) = q_i^* + N\delta x_i(t)$$

for some uniquely specified $\delta x_i(t)$. Similar statements hold for r_i. Accordingly, we can write[1]

$$\begin{pmatrix} q_1(t) \\ q_2(t) \\ r_1(t) \\ r_2(t) \end{pmatrix} = \begin{pmatrix} q_1^* \\ q_2^* \\ q_1^* \\ q_2^* \end{pmatrix} + \begin{pmatrix} \mathcal{N} & \\ & \mathcal{N} \end{pmatrix} \delta x(t),$$

where

$$\mathcal{N} = \begin{pmatrix} N & 0 \\ 0 & N \end{pmatrix} \tag{10}$$

Equivalently, we can define

$$\delta x(t) = \begin{pmatrix} \mathcal{N}^T & \\ & \mathcal{N}^T \end{pmatrix} \left(\begin{pmatrix} q_1(t) \\ q_2(t) \\ r_1(t) \\ r_2(t) \end{pmatrix} - \begin{pmatrix} q_1^* \\ q_2^* \\ q_1^* \\ q_2^* \end{pmatrix} \right). \tag{11}$$

Linearizing (4) around $(q_1^*, q_2^*, q_1^*, q_2^*)$ results in

$$\frac{d}{dt}\delta x = \begin{pmatrix} -I & (1+\gamma\lambda)D_1 & 0 & -\gamma\lambda D_1 \\ (1+\gamma\lambda)D_2 & -I & -\gamma\lambda D_2 & 0 \\ \lambda I & 0 & -\lambda I & 0 \\ 0 & \lambda I & 0 & -\lambda I \end{pmatrix} \delta x, \tag{12}$$

with

$$D_i = \frac{1}{\tau} N^T \nabla\sigma(M_i q_{-i}^*/\tau) M_i N.$$

[1]We will repeatedly use δx to designate deviation from an equilibrium. In each case, the appropriate dimension should be reinterpreted accordingly.

Define

$$\mathcal{D} = \begin{pmatrix} 0 & D_1 \\ D_2 & 0 \end{pmatrix}. \qquad (13)$$

Then we can rewrite the linearization (12) as

$$\frac{d}{dt}\delta x = \begin{pmatrix} -I + (1+\gamma\lambda)\mathcal{D} & -\gamma\lambda\mathcal{D} \\ \lambda I & -\lambda I \end{pmatrix} \delta x.$$

The following theorem characterizes local asymptotic stability of approximate DAFP and establishes that derivative action FP can be locally convergent with a suitable derivative gain when standard FP is not convergent.

Theorem 3.6: Consider a two-player game under approximate DAFP (4) with a Nash equilibrium (q_1^*, q_2^*). *Assume that* $-I + \mathcal{D}$ *in (13) is non-singular. Let* $a_i + jb_i$ *denote the eigenvalues of* $-I + \mathcal{D}$. *The linearization (12) with* $\gamma > 0$ *is asymptotically stable for large* $\lambda > 0$ *if and only if*

$$\max_i a_i < \frac{1-\gamma}{\gamma}, \quad \textit{if } \max_i a_i < 0;$$

$$\max_i \frac{a_i}{a_i^2 + b_i^2} < \frac{\gamma}{1-\gamma} < \frac{1}{\max_i a_i}, \quad \textit{if } \max_i a_i \geq 0.$$

Proof: The proof follows arguments similar to the proof of the forthcoming Theorem 4.2 and is omitted. ∎

Since $-I + \mathcal{D}$ is the Jacobian matrix of the linearization of standard FP, Theorem 3.6 relates the potential stability of approximate DAFP to the eigenvalues of standard FP. In particular, Theorem 3.6 implies that the linearization of approximate DAFP is stable whenever the linearization of standard FP is asymptotically stable. Theorem 3.6 further implies that approximate DAFP may have a stable linearization in situations where standard FP does not.

If we apply Theorem 3.6 to the Shapley example, we obtain as a condition for local stability

$$0.0413 < \frac{\gamma}{1-\gamma} < 0.0638.$$

In particular, using $\gamma = 0.05$ leads to local asymptotic stability of derivative action FP. Simulations also result in convergent behavior. Using Theorem 3.6 with the modified Shapley example leads to convergence in simulations without the previously observed bias (not shown).

IV. Dynamic Gradient Play

In this section, we will consider an alternative form of continuous-time strategy evolution called gradient play (GP). GP may be viewed as a "better response" strategy, as opposed to a "best response" strategy. In FP, a player jumps to the best response to the empirical frequencies of the opponent. In GP, a player adjusts a current strategy in a gradient direction suggested by the empirical frequencies of the opponent.

We will first define standard GP and then introduce a dynamic version that uses derivative action. As before, we will analyze convergence for both exact and approximate implementations of the derivative term.

In the entire discussion of dynamic GP, we will consider $\tau = 0$, i.e., the non-smoothed game.

A. Standard GP

Recall that each player seeks to maximize its own utility in response to observations of an opponent's actions according to the utility function (with $\tau = 0$)

$$\mathcal{U}_i(p_i, p_{-i}) = p_i^T M_i p_{-i}$$

In this case, the utility function gradient is

$$\nabla_{p_i} \mathcal{U}_i(p_i, p_{-i}) = M_i p_{-i}.$$

In continuous-time GP, the strategy of each player is

$$p_i(t) = \Pi_\Delta[q_i(t) + M_i q_{-i}(t)],$$

i.e., a combination of a player's own empirical frequency and a (projected) gradient-step using the opponents empirical frequency.

The resulting empirical frequency dynamics are then

$$\begin{aligned} \dot{q}_1(t) &= \Pi_\Delta[q_1(t) + M_1 q_2(t)] - q_1(t) \\ \dot{q}_2(t) &= \Pi_\Delta[q_2(t) + M_2 q_1(t)] - q_2(t), \end{aligned} \qquad (14)$$

which we will call *continuous-time GP*. It is straightforward to show that the equilibrium points of continuous-time GP are precisely Nash equilibria.

As opposed to FP, gradient based evolution cannot converge to a *completely mixed* Nash equilibrium. To show this, we will use the following lemma.

Lemma 4.1: Suppose $v, w \in \mathcal{R}^n$ *satisfy,*

- $v \in \mathrm{Int}(\Delta(n))$ *and*
- $\Pi_\Delta[v + w] = v$.

Then

1) $NN^T w = 0$, *and*
2) *For sufficiently small* $y \in \mathcal{R}^n$,

$$\Pi_\Delta[v + w + y] = v + NN^T y,$$

where N *is the orthonormal matrix defined in (9).*

Proof: The proof of statement 1) uses the following property of convex projections [3, Section 0.6, Corollary 1]. For all $x \in \mathcal{R}^n$ and all $s \in \Delta(n)$,

$$(\Pi_\Delta(x) - x)^T (\Pi_\Delta(x) - s) \leq 0. \qquad (15)$$

Accordingly,

$$\begin{gathered} (\Pi_\Delta(v+w) - (v+w))^T (\Pi_\Delta(v+w) - s) \leq 0 \\ \Rightarrow \\ -w^T (v - s) \leq 0 \end{gathered}$$

for all $s \in \Delta(n)$. Since v is in the simplex interior, we can set $s = v + \rho NN^T w$ with $\rho > 0$ sufficiently small. This results in

$$w^T NN^T w \leq 0,$$

and so $NN^T w = 0$. As for statement 2), consider projecting $v + w + y$ to the affine set of vectors whose elements sum to unity. This set includes the simplex, and hence includes v. The resulting projection is

$$v + NN^T w + NN^T y = v + NN^T y,$$

where we used that $NN^Tw = 0$. For y sufficiently small, $v + NN^Ty$ lies in the simplex. ∎

Now suppose that $(q_1(t), q_2(t))$ are in the vicinity of a completely mixed Nash equilibrium (q_1^*, q_2^*). As before, we can set

$$\delta x(t) = \mathcal{N}^T \left(\begin{pmatrix} q_1(t) \\ q_2(t) \end{pmatrix} - \begin{pmatrix} q_1^* \\ q_2^* \end{pmatrix} \right),$$

where $\mathcal{N}$ is defined in (10). Lemma 4.1 implies that for δx sufficiently small, the resulting (linear!) dynamics are

$$\frac{d}{dt}\delta x = \begin{pmatrix} 0 & N^T M_1 N \\ N^T M_2 N & 0 \end{pmatrix} \delta x.$$

Since the dynamics matrix has zero trace, the corresponding equilibrium cannot be asymptotically stable.

B. Exact Derivative Action GP

We will consider a modification of gradient evolution in the spirit of the prior modification of FP. Introducing a derivative term in same manner as DAFP leads to the implicit differential equation

$$\begin{aligned} \dot{q}_1 &= \Pi_\Delta[q_1 + M_1(q_2 + \gamma\dot{q}_2)] - q_1 \\ \dot{q}_2 &= \Pi_\Delta[q_2 + M_2(q_1 + \gamma\dot{q}_1)] - q_2, \end{aligned} \tag{16}$$

which we will refer to as "exact" *derivative action GP* (DAGP).

We will show that in the idealized case of exact DAGP, there always exists a derivative gain, γ, such that a completely mixed Nash equilibrium is locally asymptotically stable. As in the case of DAFP, the introduction of approximate differentiators may or may not allow one to recover the ideal case (as analyzed in the forthcoming section).

Recall the matrix $\mathcal{N}$ defined in (10), and define

$$\mathcal{M} = \begin{pmatrix} 0 & M_1 \\ M_2 & 0 \end{pmatrix}. \tag{17}$$

It is straightforward to show that a completely mixed Nash equilibrium is *isolated* if and only if $\mathcal{N}^T\mathcal{M}\mathcal{N}$ is nonsingular.

Theorem 4.1: Assume that $\mathcal{N}^T\mathcal{M}\mathcal{N}$ is nonsingular. Let (q_1^*, q_2^*) be a (isolated) completely mixed Nash equilibrium. Then for sufficiently large $\gamma > 0$ and for initial conditions $(q_1(0), q_2(0))$ sufficiently close to (q_1^*, q_2^*), there exists a solution to exact DAGP (16) that exponentially converges to (q_1^*, q_2^*).

Proof: Choose $\gamma > 0$ sufficiently large so that

$$\begin{aligned} &(I - \gamma\mathcal{N}^T\mathcal{M}\mathcal{N})^{-1}\mathcal{N}^T\mathcal{M}\mathcal{N} \\ &\quad +((I - \gamma\mathcal{N}^T\mathcal{M}\mathcal{N})^{-1}\mathcal{N}^T\mathcal{M}\mathcal{N}) < 0. \end{aligned} \tag{18}$$

This is always possible given the assumed nonsingularity of $\mathcal{N}^T\mathcal{M}\mathcal{N}$.

Any solution to (16) can be written in terms of the new variables, δx, where

$$\begin{pmatrix} q_1 \\ q_2 \end{pmatrix} = \begin{pmatrix} q_1^* \\ q_2^* \end{pmatrix} + \mathcal{N}\delta x(t). \tag{19}$$

A straightforward calculation shows that for sufficiently small δx,

$$\begin{pmatrix} \dot{q}_1 \\ \dot{q}_2 \end{pmatrix} = \mathcal{N}\frac{d}{dt}\delta x = \mathcal{N}(I - \gamma\mathcal{N}^T\mathcal{M}\mathcal{N})^{-1}\mathcal{N}^T\mathcal{M}\mathcal{N}\delta x \tag{20}$$

is an algebraic solution for the derivatives in (16). This can be seen by substituting (19) and (20) into the right-hand-side of (16) and exploiting Lemma 4.1. Accordingly, since $\mathcal{N}^T\mathcal{N} = I$, we have that for sufficiently small δx, exact DAGP becomes

$$\frac{d}{dt}\delta x = (I - \gamma\mathcal{N}^T\mathcal{M}\mathcal{N})^{-1}\mathcal{N}^T\mathcal{M}\mathcal{N}\delta x. \tag{21}$$

Using the negative definiteness in (18), we have that

$$V(q(t)) = (q(t) - q^*)^T(q(t) - q^*) = \delta x(t)^T\delta x(t)$$

is a Lyapunov function for (21), and so $\delta x(t)$ remains sufficiently small, which in turn implies that (21) continues to hold true and that δx decays exponentially to zero. ∎

C. Approximate DAGP

We now consider "approximate" DAGP, given by

$$\begin{aligned} \dot{q}_1 &= \Pi_\Delta\left[q_1 + M_1(q_2 + \gamma\dot{r}_2)\right] - q_1 \\ \dot{q}_2 &= \Pi_\Delta\left[q_2 + M_2(q_1 + \gamma\dot{r}_1)\right] - q_2 \\ \dot{r}_1 &= \lambda(q_1 - r_1) \\ \dot{r}_2 &= \lambda(q_2 - r_2), \end{aligned} \tag{22}$$

As in approximate DAFP, the derivative terms in the right hand side are replaced by approximate differentiators.

In the following sections, we will show that the local asymptotic stability of exact DAGP may or may not be achieved under approximate DAGP. These results parallel those of approximate DAFP. Unlike DAFP, since we are dealing with a non-smoothed game, we require a separate analysis between completely mixed equilibrium and strict equilibrium.

1) Completely Mixed Nash Equilibria: The following theorem gives a complete characterization of when a completely mixed Nash equilibrium is locally asymptotically stable under approximate DAGP. As noted previously, a mixed equilibrium under standard GP is never asymptotically stable, clearly indicating that the introduction of derivative action can enable convergence. Unlike exact DAGP, the asymptotic stability is not always achievable.

A Nash equilibrium of (q_1^*, q_2^*) leads to an equilibrium point $(q_1^*, q_2^*, q_1^*, q_2^*)$ of approximate DAGP (22). As before, we can write

$$\begin{pmatrix} q_1(t) \\ q_2(t) \\ r_1(t) \\ r_2(t) \end{pmatrix} = \begin{pmatrix} q_1^* \\ q_2^* \\ q_1^* \\ q_2^* \end{pmatrix} + \begin{pmatrix} \mathcal{N} & \\ & \mathcal{N} \end{pmatrix} \delta x(t).$$

For a completely mixed equilibrium, it is straightforward to show that δx will (locally) evolve according to

$$\frac{d}{dt}\delta x = \begin{pmatrix} (1 + \gamma\lambda)\mathcal{N}^T\mathcal{M}\mathcal{N} & -\gamma\lambda\mathcal{N}^T\mathcal{M}\mathcal{N} \\ \lambda I & -\lambda I \end{pmatrix} \delta x \tag{23}$$

Theorem 4.2: Consider a two-player game under approximate DAGP with a completely mixed Nash equilibrium

(q_1^*, q_2^*)*, and assume that* $\mathcal{N}^T\mathcal{M}\mathcal{N}$ *is nonsingular. Let* $a_i + jb_i$ *be the eigenvalues of* $\mathcal{N}^T\mathcal{M}\mathcal{N}$*. The linear dynamics (23) are asymptotically stable for large* $\lambda > 0$ *if and only if*

$$\max_i \frac{a_i}{a_i^2 + b_i^2} < \gamma < \frac{1}{\max_i\{a_i\}}.$$

Proof: It will be convenient to examine the eigenvalues of a related matrix, J, defined as follows. We can scale the dynamics matrix in (23) by $\gamma > 0$ and perform the change of variables $\lambda \mapsto \frac{1}{\gamma}\lambda$ to produce the matrix

$$J = \begin{pmatrix} (1+\lambda)\gamma\mathcal{N}^T\mathcal{M}\mathcal{N} & -\gamma\lambda\mathcal{N}^T\mathcal{M}\mathcal{N} \\ \lambda I & -\lambda I \end{pmatrix}.$$

Since J is just a rescaled version of the dynamics matrix of (23), we have that J is a stability matrix for sufficiently large λ if and only if (23) is asymptotically stable for sufficiently large λ.

Let λ_J be any eigenvalue of J and $v_J = \begin{pmatrix} v_J^1 \\ v_J^2 \end{pmatrix}$ be the corresponding eigenvector. By definition, we have

$$\text{(i)} \qquad \lambda v_J^1 - \lambda v_J^2 = \lambda_J v_J^2 \;\Leftrightarrow\; v_J^2 = \frac{\lambda}{\lambda + \lambda_J} v_J^1, \text{ and}$$

$$\text{(ii)} \qquad (1+\lambda)\gamma\mathcal{N}^T\mathcal{M}\mathcal{N}v_J^1 - \lambda\gamma\mathcal{N}^T\mathcal{M}\mathcal{N}v_J^2 = \lambda_J v_J^1,$$

which is equivalent to

$$\text{(i)} \qquad \lambda v_J^1 - \lambda v_J^2 = \lambda_J v_J^2 \;\Leftrightarrow\; v_J^2 = \frac{\lambda}{\lambda + \lambda_J} v_J^1, \text{ and}$$

$$\text{(iii)} \qquad \left[(1+\lambda)\gamma - \frac{\lambda^2\gamma}{\lambda + \lambda_J}\right]\mathcal{N}^T\mathcal{M}\mathcal{N}v_J^1 = \lambda_J v_J^1,$$

where we used the assumption that $\mathcal{N}^T\mathcal{M}\mathcal{N}$ is nonsingular, in which case $\lambda + \lambda_J \neq 0$. Therefore, the eigenvalues of J and $\mathcal{N}^T\mathcal{M}\mathcal{N}$ are related through

$$\left[(1+\lambda)\gamma - \frac{\lambda^2\gamma}{\lambda + \lambda_J}\right]\mu = \lambda_J$$
$$\Leftrightarrow$$
$$\lambda_J^2 + \lambda_J[\lambda - \gamma(1+\lambda)\mu] - \gamma\lambda\mu = 0,$$

where μ is an eigenvalue of $\mathcal{N}^T\mathcal{M}\mathcal{N}$. Note that the stability of the above polynomial for $\mu = a + jb$ is equivalent to

(i) $(\lambda - \gamma(1+\lambda)a) > 0$,

(ii) $-\lambda\gamma a[\lambda - \gamma(1+\lambda)a]^2$
$-\lambda\gamma b\left([\lambda - \gamma(1+\lambda)a][-\gamma(1+\lambda)b] + \gamma\lambda b\right) > 0,$

by the (complex) Routh-Hurwitz criterion (e.g., [1], [41]). For large $\lambda > 0$, these stability conditions reduce to

$$1/\gamma > a, \text{ and } \gamma > a/(a^2 + b^2).$$

Note that matrix $\mathcal{N}^T\mathcal{M}\mathcal{N}$ has zero trace. Since $\mathcal{N}^T\mathcal{M}\mathcal{N}$ must be unstable, the resulting stability conditions for the matrix J for large λ become

$$1/\max_i\{a_i\} > \gamma > \max_i\{a_i/(a_i^2 + b_i^2)\}.$$

■

Both the Shapley and modified Shapley games satisfy the stability condition of Theorem 4.2 with $\gamma = 1$.

It is also possible to find examples where the mixed Nash equilibrium is unstable under the dynamics (23), for example a two-player/two-move identical interest game with $M_1 = M_2 = I$. In fact, it is straightforward to show for this specific case that all values of $\lambda > 0$ and $\gamma > 0$ result in instability of the mixed equilibrium.

2) Strict Nash Equilibria: We now show that approximate DAGP always results in locally stable behavior near *strict* Nash equilibria.

The pair (q_1^*, q_2^*) forms a *strict* Nash equilibrium (e.g., [23]) if for all $s \in \Delta(m)$, with $s \neq q_i^*$,

$$s^T M_i q_{-i}^* < (q_i^*)^T M_i q_{-i}^*,$$

i.e., the best response q_i^* to the strategy q_{-i}^* is strictly superior to other responses. Consequences of being a strict equilibrium are 1) both q_i^* lie on a vertex of the simplex $\Delta(m)$ (i.e., the q_i^* are pure strategies) and 2) there exists a $\rho > 0$ such that for all $s \in \Delta(m)$ and all $x \in \mathcal{R}^m$ with $|x| < \rho$,

$$s^T M_i(q_{-i}^* + x) < (q_i^*)^T M_i(q_{-i}^* + x). \qquad (24)$$

In other words, q_i^* is the best response to q_{-i}^* *and* vectors near q_{-i}^*.

Theorem 4.3: Consider a two-player game under approximate DAGP with a strict Nash equilibrium (q_1^*, q_2^*)*. The associated equilibrium* $(q_1^*, q_2^*, q_1^*, q_2^*)$ *of approximate DAGP (22) is locally asymptotically stable for any* $\gamma > 0$ *and* $\lambda > 0$*.*

Proof: Define the Lyapunov function

$$\mathcal{V}(q_1, q_2, r_1, r_2) = V_1(q_1) + V_2(q_2) + W_1(r_1, q_1) + W_2(r_2, q_2),$$

where

$$V_1(q_1) = \frac{1}{2}(q_1 - q_1^*)^T(q_1 - q_1^*)$$
$$V_2(q_2) = \frac{1}{2}(q_2 - q_2^*)^T(q_2 - q_2^*)$$
$$W_1(r_1, q_1) = \frac{\lambda}{2}(r_1 - q_1)^T(r_1 - q_1)$$
$$W_2(r_2, q_2) = \frac{\lambda}{2}(r_2 - q_2)^T(r_2 - q_2)$$

Then using the strict equilibrium property (24) and convex projection property (15), one can show that there exists a $\delta > 0$ such that

$$\mathcal{V}(q_1(t), q_2(t), r_1(t), r_2(t)) < \delta$$

implies

$$\begin{aligned} &\frac{d}{dt}\mathcal{V}(q_1(t), q_2(t), r_1(t), r_2(t)) \\ &\quad \le -\Big(|\dot{q}_1(t)|^2 + |\dot{q}_2(t)|^2 . \\ &\qquad + \lambda^2|r_1(t) - q_1(t)|^2 + \lambda^2|r_2(t) - q_2(t)|^2\Big) \\ &\qquad + \lambda|r_1(t) - q_1(t)|\,|\dot{q}_1(t)| + \lambda|r_2(t) - q_2(t)|\,|\dot{q}_2(t)| \\ &\quad \le -\frac{1}{2}\Big(|\dot{q}_1(t)|^2 + |\dot{q}_2(t)|^2 \\ &\qquad + \lambda^2|r_1(t) - q_1(t)|^2 + \lambda^2|r_2(t) - q_2(t)|^2\Big). \end{aligned}$$

The remainder of the proof follows standard Lyapunov arguments. ■

D. Simulations: Approximate DAGP on the Shapley game

Approximate DAGP (22) exhibited convergence on the standard Shapley example (not shown). Figures 5 shows the empirical frequencies of the two players in the modified Shapley example using $\gamma = 1$ and $\lambda = 100$. The empirical frequencies converge to the completely mixed Nash equilibrium as anticipated.

E. Local Stabilizability

If one allows even broader classes of strategic update mechanisms, it is possible to stabilize any mixed equilibrium. We can rewrite (23) as a feedback law with "control", u, and "measurement", y,

$$\frac{d}{dt}\delta x = \begin{pmatrix} \mathcal{N}^T\mathcal{M}\mathcal{N} & 0 \\ \lambda I & -\lambda I \end{pmatrix} \delta x + \begin{pmatrix} \mathcal{N}^T\mathcal{M}\mathcal{N} \\ 0 \end{pmatrix} u$$

$$y = \begin{pmatrix} \lambda I & -\lambda I \end{pmatrix} \delta x$$

$$u = \gamma y,$$

where $\mathcal{N}$ is defined in (25).

Proposition 4.1: Assume that $\mathcal{N}^T\mathcal{M}\mathcal{N}$ is nonsingular. Then

$$\left[\begin{pmatrix} \mathcal{N}^T\mathcal{M}\mathcal{N} & 0 \\ \lambda I & -\lambda I \end{pmatrix}, \begin{pmatrix} \mathcal{N}^T\mathcal{M}\mathcal{N} \\ 0 \end{pmatrix} \right]$$

form a controllable pair, and

$$\left[\begin{pmatrix} \lambda I & -\lambda I \end{pmatrix}, \begin{pmatrix} \mathcal{N}^T\mathcal{M}\mathcal{N} & 0 \\ \lambda I & -\lambda I \end{pmatrix} \right]$$

form an observable pair.

Proof: Standard rank tests prove the desired result. ∎

Proposition 4.1 implies that it is always possible to design a dynamic compensator that renders a completely mixed equilibrium stable. Note that such a compensator would *not* require knowledge of the Nash equilibrium since the measurement, y, can be expressed as in terms of $q - r$. This result is limited to a conceptual existence interpretation. One challenge is to compute the state space parameters of such a compensator without shared knowledge of utility matrices. Furthermore, the concept of "individual rationality" becomes less clear with general compensator dynamics. The present paper's approach finds sufficient conditions for local stability for the special structure of approximate derivative action, which is readily interpreted in the context of individual rationality.

V. Multiplayer Games

A. Multiplayer DAFP and DAGP

We now consider the case with n_{P} players, each with a utility function $\mathcal{U}_i(p_i, p_{-i})$. We will impose structural assumptions on the $\mathcal{U}_i$ as needed.

It turns out that all of the previous results hold in the multiplayer case with only notational changes.

First, consider multiplayer fictitious play. If we assume that each player has a differentiable best response function, $\beta_i(p_{-i})$, then we can write exact DAFP as

$$\begin{aligned} \dot{q}_1 &= \beta_1(q_{-1} + \gamma \dot{q}_{-1}) - q_1 \\ &\vdots \\ \dot{q}_{n_{\mathrm{P}}} &= \beta_{n_{\mathrm{P}}}(q_{-n_{\mathrm{P}}} + \gamma \dot{q}_{-n_{\mathrm{P}}}) - q_{n_{\mathrm{P}}}. \end{aligned}$$

The ensuing analysis of approximate derivative measurements, weak convergence, and approximate differentiator implementation remains the same. In particular, a suitably modified version of Theorem 3.6 regarding local asymptotic stability still holds.

In the case of gradient play, we will impose the following "pairwise" structure on the player utilities,

$$\mathcal{U}_i(p_i, p_{-i}) = p_i^T (\sum_{j \neq i} M_{ij} p_j),$$

characterized by matrices M_{ij}. Once again, all of the ensuing analysis holds. In particular, the local stability of complete mixed equilibria in Theorem 4.2 holds with a simple notational change. Namely, redefine the matrix $\mathcal{N}$ from (10) to

$$\mathcal{N} = \begin{pmatrix} N & & \\ & \ddots & \\ & & N \end{pmatrix}, \tag{25}$$

and redefine the matrix $\mathcal{M}$ as the block matrix whose ij^{th} block is M_{ij}, and whose ii^{th} block is 0.

We comment that in the multiplayer case, both DAFP and DAGP automatically respect any underlying "graph" structure in that a player only monitors the empirical frequencies of opponent players that enter into the utility function, e.g., [30]

B. Simulations: Derivative Action FP and GP on the Jordan game

We will illustrate the multiplayer case on a version of the Jordan anti-coordination game [28]. It is known that standard FP does not converge for this game. Reference [26] goes on to show that there is no algorithm that assures convergence to equilibrium in which player strategies are static functions of opponent empirical frequencies and players do not have access to opponent utilities. In this game, there are three players with two possible actions. The utilities reflect that player $\mathcal{P}_1$ wants to differ from player $\mathcal{P}_2$, player $\mathcal{P}_2$ wants to differ from player $\mathcal{P}_3$, and player $\mathcal{P}_3$ wants to differ from player $\mathcal{P}_1$. Following [26], an extension of the Jordan game can be written as

$$\begin{aligned} \mathcal{U}_1(p_1, p_2) &= p_1^T \left(\begin{smallmatrix} 0 & a^1 \\ 1 & 0 \end{smallmatrix} \right) p_2 + \tau \mathcal{H}(p_1) \\ \mathcal{U}_2(p_2, p_3) &= p_2^T \left(\begin{smallmatrix} 0 & a^2 \\ 1 & 0 \end{smallmatrix} \right) p_3 + \tau \mathcal{H}(p_2) \\ \mathcal{U}_3(p_3, p_1) &= p_3^T \left(\begin{smallmatrix} 0 & a^3 \\ 1 & 0 \end{smallmatrix} \right) p_1 + \tau \mathcal{H}(p_3), \end{aligned}$$

where the $a^i > 0$ are utility parameters. The case where $a^i = 1$ is the standard Jordan game. In case $\tau = 0$, the unique Nash equilibrium is

$$q_1^* = \begin{pmatrix} \frac{a^3}{a^3+1} \\ \frac{1}{a^3+1} \end{pmatrix}, \quad q_2^* = \begin{pmatrix} \frac{a^1}{a^1+1} \\ \frac{1}{a^1+1} \end{pmatrix}, \quad q_3^* = \begin{pmatrix} \frac{a^2}{a^2+1} \\ \frac{1}{a^2+1} \end{pmatrix}.$$

Note that game satisfies the pairwise utility structure of the previous section.

Figure 6 shows the oscillatory behavior of the three players' empirical frequencies under standard GP with $a^i = 1$. Figure 7 shows convergent behavior under approximate DAGP with standard approximate differentiators ($\lambda = 50$) and the same a^i parameters. Figure 8 shows convergent behavior for $a^1 = 2$, $a^2 = 1$, and $a^3 = 1/3$.

In all cases, multiplayer versions of Theorems 3.6 and 4.2 confirm the convergence of derivative action although standard methods are non-convergent.

VI. Concluding Remarks

We have introduced a notion of dynamic fictitious play and dynamic gradient play through the use of derivative action in a continuous-time form of repeated games. We have shown that in the ideal case of exact derivative measurements, derivative action can guarantee convergence to Nash equilibria. In the non-ideal case, we discussed how approximate differentiators in certain cases can recover the ideal case of exact derivative measurements.

These dynamics satisfy the "uncoupled" restriction of [26], but are able to bypass the uncoupled obstacle to convergence through the introduction of higher order dynamics, namely derivative action. Unlike randomized search approaches that also achieve convergence [27], the present approach is more akin to myopic local search.

One open question is how to better characterize the class of games for which derivative action, or more generally other types of dynamic compensation, can enable convergence. Since the present analysis is local, another concern is determining whether convergence is actually global for special classes of games.

In this paper, we did not formally establish any ties to discrete time. The dynamics under consideration are of the simpler sort to apply stochastic approximation results, i.e., continuous dynamics over compact sets, so issues associated with boundedness of iterations do not arise. Texts such as [32] or papers such as [5] provide tools to establish this connection. In particular, the results of [6], [7] combined with the local analysis of Theorems 3.6 and 4.2 establish that there is a positive probability of convergence to a mixed Nash equilibrium in discrete-time play given local stability of the continuous-time version. These issues are discussed in [2]. Reference [2] also discusses the "payoff based" case, where empirical frequencies of other players are not measured. Rather, each player only observes the private reward at each stage.

References

[1] S.D. Agashe. A new general routh-like algorithm to determine the number of rhp roots of a real or complex polynomial. *IEEE Transactions on Automatic Control*, **30**(4):406–409, 1985.

[2] G. Arslan and J.S. Shamma. Distributed convergence to Nash equilibria with local utility measurements. In *43rd IEEE Conference on Decision and Control*, 2004.

[3] J.P. Aubin and A. Cellina. *Differential Inclusions*. Springer-Verlag, New York, 1984.

[4] T. Basar. Relaxation techniques and asynchronous algorithms for online computation of non-cooperative equilibria. *Journal of Economic Dynamics and Control*, **11**:531–549, 1987.

[5] M. Benaim and M.W. Hirsch. A dynamical systems approach to stochastic approximation. *SIAM Journal on Control and Optimization*, **34**:437–472, 1996.

[6] M. Benaim and M.W. Hirsch. Mixed equilibria and dynamical systems arising from fictitious play in perturbed games. *Games and Economic Behavior*, **29**:36–72, 1999.

[7] M. Benaim, J. Hofbauer, and S. Sorin. Stochastic approximations and differential inclusions. preprint, http://ceco.polytechnique.fr/CAHIERS/pdf/2003-029.pdf, 2003.

[8] U. Berger. Fictitious play in $2 \times n$ games. Economics Working Paper Archive at WUSTL, mimeo, 2003.

[9] G.W. Brown. Iterative solutions of games by fictitious play. In T.C. Koopmans, editor, *Activity Analysis of Production and Allocation*, pages 374–376. Wiley, New York, 1951.

[10] J. Conlisk. Adaptation in games: Two solutions to the Crawford puzzle. *Journal of Economic Behavior and Organization*, **22**:25–50, 1993.

[11] V.P. Crawford. Learning behavior and mixed strategy Nash equilibria. *Journal of Economic Behavior and Organization*, **6**:69–78, 1985.

[12] G. Ellison and D. Fudenberg. Learning purified mixed equilibria. *Journal of Economic Theory*, **90**:83–115, 2000.

[13] S.D. Flam. Equilibrium, evolutionary stability and gradient dynamics. *International Game Theory Review*, **4**(4):357–370, 2002.

[14] S.D. Flam and J. Morgan. Newtonian mechanics and Nash play. Technical Report 5–2001, Dip. di Matematica e Applicazioni, Università di Napoli, 2001.

[15] W. Fleming. *Functions of Several Variables*. Springer-Verlag, 2nd edition, 1977.

[16] D.P. Foster and R.V. Vohra. Calibrated learning and correlated equilibrium. *Games and Economic Behavior*, **21**:40–55, 1997.

[17] D.P. Foster and H.P. Young. Learning, hypothesis testing, and Nash equilibrium. *Games and Economic Behavior*, **45**:73–96, 2003.

[18] D.P. Foster and H.P. Young. Regret testing: A simple payoff-based procedure for learning Nash equilibrium. preprint, http://www.econ.jhu.edu/people/young/RegretTesting.pdf, 2003.

[19] D.P. Foster and H.P. Young. On the nonconvergence of fictitious play in coordination games. *Games and Economic Behavior*, **25**, 79–96.

[20] D. Fudenberg and D. Kreps. Learning mixed equilibria. *Games and Economic Behavior*, **5**:320–367, 1993.

[21] D. Fudenberg and D.K. Levine. *The Theory of Learning in Games*. MIT Press, Cambridge, MA, 1998.

[22] D. Fudenberg and D.K. Levine. Conditional universal consistency. *Games and Economic Behavior*, **29**:104–130, 1999.

[23] D. Fudenberg and J. Tirole. *Game Theory*. MIT Press, Cambridge, MA, 1991.

[24] S. Hart. Adaptive heuristics. preprint, http://www.ma.huji.ac.il/˜hart/abs/heurist.html, 2004.

[25] S. Hart and A. Mas-Colell. A simple adaptive procedure leading to correlated equilibrium. *Econometrica*, **68**:1127–1150, 2000.

[26] S. Hart and A. Mas-Colell. Uncoupled dynamics do not lead to Nash equilibrium. *American Economic Review*, **93**(5):1830–1836, 2003.

[27] S. Hart and A. Mas-Colell. Stochastic uncoupled dynamics and Nash equilibrium. preprint, http://www.ma.huji.ac.il/˜hart/abs/uncoupl-st.html, 2004.

[28] J. Jordan. Three problems in game theory. *Games and Economic Behavior*, **5**:368–386, 1993.

[29] S.M. Kakade and D.P. Foster. Deterministic calibration and Nash equilibrium. preprint, http://www.cis.upenn.edu/˜skakade/papers/gt/calibration.pdf, 2004.

[30] M.J. Kearns, M.L. Littman, and S.P. Singh. Graphical models for game theory, 2001.

[31] V. Krishna and T. Sjöström. On the convergence of fictitious play. *Mathematics of Operations Research*, **23**(2):479–511, 1998.

[32] H.J. Kushner and G.G. Yin. *Stochastic Approximation Algorithms and Applications*. Springer-Verlag, 1997.

[33] D.S. Leslie and E.J. Collins. Convergent multiple-timescales reinforcement learning algorithms in normal form games. *The Annals of Applied Probability*, **4**(4):1231–1251, 2003.

[34] S.G. Massaquoi. *Modeling the Function of the Cerebellum in Scheduled Linear Servo Control of Simple Horizontal Planar Arm Movements*. PhD thesis, Massachusetts Institute of Technology, 1999.

[35] K. Miyasawa. On the convergence of learning processes in a 2×2 non-zero-sum two person game. Technical Report 33, Economic Research Program, Princeton University, 1961.

[36] D. Monderer and L.S. Shapley. Fictitious play property for games with identical interests. *Journal of Economic Theory*, **68**:258–265, 1996.

[37] J. Robinson. An iterative method of solving a game. *Ann. Math.*, **54**:296–301, 1951.

[38] A. Rubinstein. Comments on the interpretation of game theory. *Econometrica*, **59**(4):909–924, 1991.

[39] J.S. Shamma and G. Arslan. A feedback stabilization approach to fictitious play. In *Proceedings of the 42nd IEEE Conference on Decision and Control*, pages 4140–4145, 2003.

[40] L.S. Shapley. Some topics in two-person games. In L.S. Shapley M. Dresher and A.W. Tucker, editors, *Advances in Game Theory*, pages 1–29. University Press, Princeton, NJ, 1964.

[41] Y.Q. Shi, K.K. Yen, and C.M. Chen. The necessary conditions for a complex polynomial to be strictly Hurwitz and their applications in robust stability analysis. *IEEE Transaction on Automatic Control*, **38**(1):125–128, 1993.

[42] A. Vasin. On stability of mixed equilibria. *Nonlinear Analysis*, **38**:793–802, 1999.

[43] J.W. Weibull. *Evolutionary Game Theory*. MIT Press, Cambridge, MA, 1995.

[44] D. H. Wolpert. Information Theory – The Bridge Connecting Bounded Rational Game Theory and Statistical Physics. http://arxiv.org/PS_cache/cond-mat/pdf/0402/0402508.pdf, 2004.

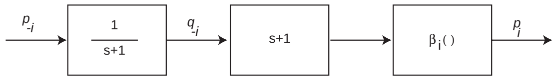

Fig. 1. Inversion schematic with $\gamma = 1$

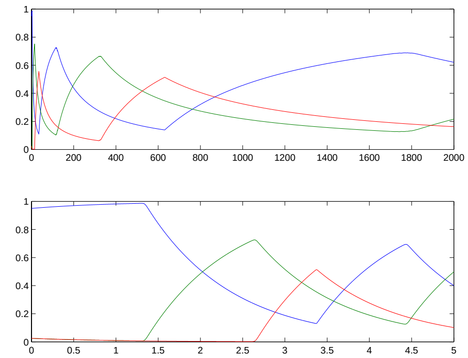

Fig. 2. Shapley game empirical frequencies: Discrete-time (top) & continuous-time (bottom)

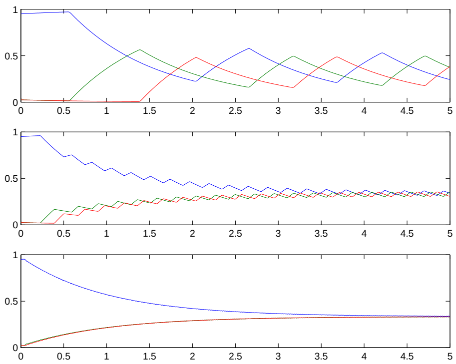

Fig. 3. Shapley game continuous time, $q_1^\lambda(t)$: Approximate differentiator $\lambda = 1$ (top), 10 (middle), 100 (bottom)

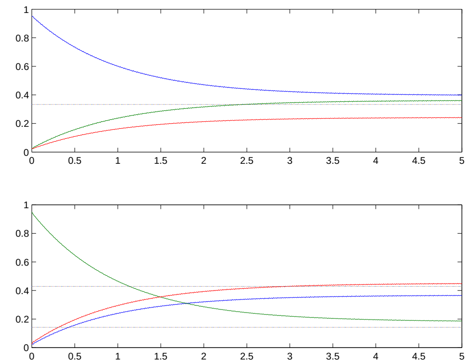

Fig. 4. Modified Shapley game, $q_1(t)$ (top) and $q_2(t)$ (bottom): Approximate differentiator $\lambda = 100$

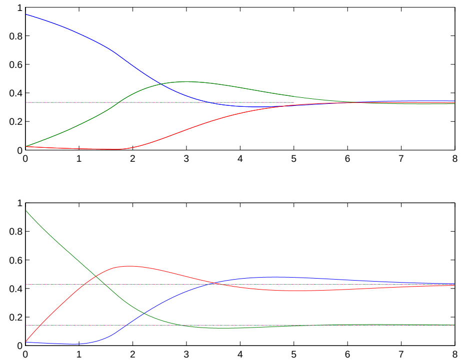

Fig. 5. Approximate DAGP on modified Shapley game

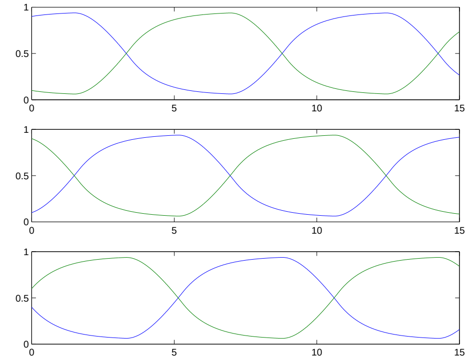

Fig. 6. Jordan game: Standard GP

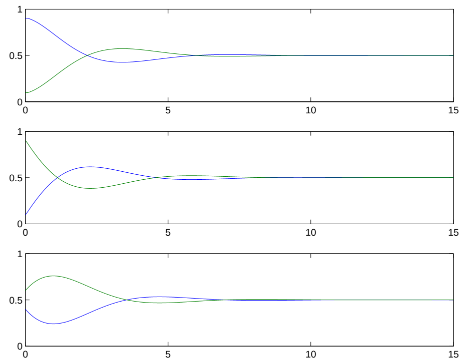

Fig. 7. Jordan game: Approximate DAGP

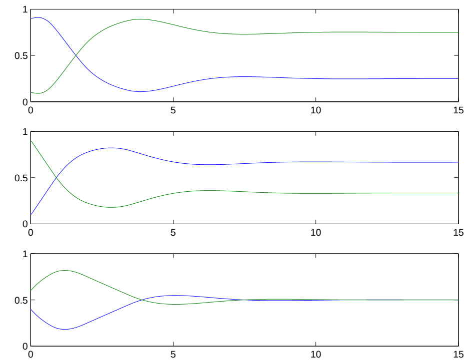

Fig. 8. Modified Jordan game: Approximate DAGP