Proceedings of the 42nd IEEE
Conference on Decision and Control
Maui, Hawaii USA, December 2003

ThP01-3

A Feedback Stabilization Approach to Fictitious Play†

Jeff S. Shamma
Department of Mechanical and
Aerospace Engineering
University of California Los Angeles
shamma@seas.ucla.edu

Gurdal Arslan
Department of Mechanical and
Aerospace Engineering
University of California Los Angeles
garslan@seas.ucla.edu

***Abstract*— We consider 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 show that the use of derivative action in processing opponent actions can lead to behavior converging to Nash equilibria. We analyze the use of approximate differentiators and reveal a potentially detrimental biasing effect. We go on to provide alternative mechanisms to diminish or eliminate this effect. We discuss two player games throughout and outline extensions to multiplayer games. We also provide convergent simulations throughout to standard counterexamples in the literature.**

I. OVERVIEW

We consider a repeated game in which players continually update strategies in response to observations of opponent actions. Our particular set-up is as follows. There are two players, each with a finite set of possible actions. Every time the game is played, each player select 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, individual utilities are *not* shared.

Now if *both* players presumed that the other player is using a constant strategy, their strategy update mechanisms become *intertwined*. This mechanism is called "fictitious play". 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 a strategies that are at a Nash equilibrium, i.e., neither player has an one-sided incentive to change strategy.

The procedure of fictitious play was introduced in 1951 [2], [11] as a mechanism to compute Nash equilibria. There is a substantial body of literature on the topic. Two overviews on parallel lines of research are [5] and [15]. Of particular concern is whether repeated play will indeed converge to a Nash equilibrium. It turns out that players' strategies may or may not converge. A convergence counterexample due to Shapley in 1964 has two players with three moves each [14]. A 1993 counterexample due to Jordan has three players with two moves each [9]. Regarding this lack of convergence, the paper [8] builds on Jordan's counterexample to show that a generalized version will not exhibit convergence for any strategy update mechanism (i.e., not just a best response mechanism), provided that players do not share their utility functions and update mechanisms are *static* functions of observed opponent actions.

There are methods that are guaranteed to converge to the set of *correlated* equilibria, which is larger than the set of Nash equilibria [3], [6], [7]. The recent paper [4] uses hypothesis testing to derive a method under which empirical frequencies spend "most" of the time in a neighborhood of Nash equilibria, but do not remain in that neighborhood.

In this paper, we explore the possibility of *dynamic* functions of opponent actions in the spirit of dynamic compensation for feedback stabilization. We will focus on the use of *derivative* action. We will employ a strategy update mechanism that resembles traditional mechanisms but use both the empirical frequencies and their derivatives. As opposed to revisiting past histories, the use of derivative action may be viewed as a myopic predictor of opponent actions.

For the sake of brevity, many citations and all proofs have been omitted from this presentation. A complete discussion may be found in [13], [12].

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 $\begin{pmatrix} 1 \\ \vdots \\ 1 \end{pmatrix} \in \mathcal{R}^n$.
- $\Delta(n)$ denotes the simplex in $\mathcal{R}^n$, i.e.,
$$\left\{ s \in \mathcal{R}^n | s \geq 0 \text{ componentwise, and } \mathbf{1}^T s = 1 \right\}$$
- $\Pi_\Delta : \mathcal{R}^n \to \Delta(n)$ 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.
- $\mathcal{H} : \text{Int}(\Delta(n)) \to \mathcal{R}$ denotes the entropy function
$$\mathcal{H}(s) = -s^T \log(s)$$
- $\sigma : \mathcal{R}^n \to \Delta(n)$ 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
$$\nabla\sigma(x) = \text{diag}(\sigma(x)) - \sigma(x)\sigma^T(x)$$
- For $\delta > 0$, $\text{dz}(\cdot, \delta) : \mathcal{R} \to \mathcal{R}$ denotes the deadzone function
$$\text{dz}(x, \delta) = \begin{cases} x - \delta, & \text{if } x \geq \delta \\ 0, & \text{if } -\delta < x < \delta \\ x + \delta, & \text{if } x \leq -\delta \end{cases}$$

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

II. Standard Fictitious Play for Two Player Games

A. Setup

We consider a two player game with players $\mathcal{P}_1$ and $\mathcal{P}_2$, each with positive integer dimensions m_1 and m_2, respectively. Each player, $\mathcal{P}_i$, selects a strategy, $p_i \in \Delta(m_i)$, and receives a real-valued reward according to the utility function $\mathcal{U}_i(p_i, p_{-i})$. These utility functions take the form

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

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

Define the *best response* mappings

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

by

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

The best response turns out to be the logit or soft-max function (see Notation section)

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

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

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

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 pair (p_1^*, p_2^*) such that

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

Now suppose that the game is repeated at every time $k \in \{0, 1, 2, \dots\}$. 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 standard discrete-time fictitious play (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.,

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

Now consider the continuous-time dynamics,

$$\dot{q}_1(t) = \beta_1(q_2(t)) - q_1(t) \tag{2a}$$
$$\dot{q}_2(t) = \beta_2(q_1(t)) - q_2(t) \tag{2b}$$

We call these equations *standard continuous-time FP*. These are the dynamics obtained by viewing standard discrete-time FP as stochastic approximation iterations and applying associated ordinary differential equation (ODE) analysis methods [10].

B. Convergent Cases

There are known convergent cases of standard fictitious play [5], in particular zero-sum games ($M_1 = -M_2^T$), identical interest games ($M_1 = M_2^T$), and two-player/two-move games. One can derive a unified framework which establishes convergence in all of the aforementioned special cases.

Define the function $V_1 : \Delta(m_1)\times\Delta(m_2) \to \mathcal{R}^+$ as

$$\begin{aligned} V_1(q_1,q_2) &= \max_{s\in\Delta(m_1)} \mathcal{U}_1(s,q_2) - \mathcal{U}_1(q_1,q_2) \\ &= (\beta_1(q_2) - q_1)^T M_1 q_2 + \tau(\mathcal{H}(\beta_1(q_2)) - \mathcal{H}(q_1)) \end{aligned}$$

Similarly define

$$V_2(q_2,q_1) = \max_{s\in\Delta(m_2)} \mathcal{U}_2(s,q_1) - \mathcal{U}_2(q_2,q_1)$$

Each V_i has the natural interpretation as the maximum possible reward improvement to player $\mathcal{P}_i$ by using the best response to q_{-i} rather than the specified q_i.

The functions V_1 and V_2 can be used to show that the continuous-time empirical frequencies converge to a Nash equilibrium by evaluating V_1 and V_2 along trajectories of standard continuous-time FP (2). The argument will either be to show that a weighted sum, $\alpha_1 V_1 + \alpha_2 V_2$, is a Lyapunov function, or show that the sum is integrable [13].

III. Dynamic Fictitious Play

A. Ideal Case: Derivative Measurements

In standard 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.,

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

Suppose now that in addition to empirical frequencies being available to all players, empirical frequency *derivatives*, $\dot{q}_i(t)$, are also available. This modification is very much in the spirit of standard PID controllers. Let $p_i(t)$ denote the strategy of player $\mathcal{P}_i$ at time t. Rather than use p_i as a best response to q_{-i}, consider

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

i.e., each player's strategy is a best response to a combination of empirical frequencies and the *derivative* of empirical frequencies. The classical control interpretation is that the derivative term serves as a short term prediction the opponent's strategy.

This modification leads to the following differential equation

$$\dot{q}_1 = \beta_1(q_2 + \dot{q}_2) - q_1 \tag{3a}$$
$$\dot{q}_2 = \beta_2(q_1 + \dot{q}_1) - q_2 \tag{3b}$$

which we refer to as *derivative action* FP.

Introduce the variables

$$z_1 = q_1 + \dot{q}_1$$
$$z_2 = q_2 + \dot{q}_2$$

and let

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

be the mapping

$$\begin{pmatrix} z_1 \\ z_2 \end{pmatrix} \mapsto \begin{pmatrix} \beta_1(z_2) \\ \beta_2(z_1) \end{pmatrix} \tag{4}$$

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

$$z = T(z)$$

i.e., derivative action FP must evolve over fixed points of T.

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

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

denote the set of Nash equilibria satisfying (1).

Theorem 3.1: Any solution of derivative action FP dynamics (3) satisfies the differential inclusion

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

Theorem 3.2: In the case of a unique Nash equilibrium $Q^* = \{(q_1^*, q_2^*)\}$, the unique solution to derivative action FP (3) is

$$\dot{q}_1 = -q_1 + q_1^*$$
$$\dot{q}_2 = -q_2 + q_2^*$$

which converges (exponentially) to the unique Nash equilibrium.

B. Approximate Derivative Measurements and Approximate Differentiators

Suppose we generate an approximation of the empirical frequency derivatives based on available data. Then we can write derivative action FP as

$$\dot{q}_1 = \beta_1(q_2 + \dot{q}_2 + e_2) - q_1 \quad (5a)$$
$$\dot{q}_2 = \beta_2(q_1 + \dot{q}_1 + e_1) - q_2 \quad (5b)$$

which we call approximate derivative action FP. The new variables, $e_i(t)$, denote the derivative approximation error.

One can show that under certain technical conditions [12] that for any $\varepsilon > 0$, there exists a $\delta > 0$ such that if $(e_1(t), e_2(t))$ eventually remain within an 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

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

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

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

Towards this end, consider

$$\dot{q}_1 = \beta_1(q_2 + \lambda(q_2 - r_2)) - q_1 \quad (6a)$$
$$\dot{q}_2 = \beta_2(q_1 + \lambda(q_1 - r_1)) - q_2 \quad (6b)$$
$$\dot{r}_1 = \lambda(q_1 - r_1) \quad (6c)$$
$$\dot{r}_2 = \lambda(q_2 - r_2) \quad (6d)$$

with $\lambda > 0$. An alternative expression for the empirical frequency evolution is

$$\dot{q}_1 = \beta_1(q_2 + \dot{r}_2) - q_1$$
$$\dot{q}_2 = \beta_2(q_1 + \dot{r}_1) - q_2$$

The variables, r_i, are "filtered" versions of the empirical frequencies. Intuitively, as λ increases, r_i closely tracks q_i, and so $\dot{r}_i$ is a good approximation for $\dot{q}_i$. Unfortunately, such intuition does not seem to hold. The problem is that the approximation error associated with $\dot{r}_i - \dot{q}_i$ 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 (6) with progressively larger values of λ. Towards this end, let $(q_i^\lambda, r_i^\lambda)$ denote the solutions to (6) for the specified value of λ.

Theorem 3.3: 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 $L^1([T_1, T_2], \mathcal{R}^{m_i})$.*

Theorem 3.4: In the context of Theorem 3.3, 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

$$\overline{b}_i(t) = \beta_i(\overline{q}_{-i}(t) + \dot{\overline{q}}_{-i}(t)) \quad (7)$$

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

Theorem 3.4 establishes that using increasing values of λ can converge to to a solution of the derivative action FP dynamics (3) under the equality assumption (7). 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 uniform continuity as an operator. Indeed, asymmetries due to nonlinearities can destroy the desired weak continuity.

C. Simulation Examples

Figure 1 shows the empirical frequency response of player $\mathcal{P}_1$ for the Shapley counterexample with $\tau = .01$ under *derivative action* FP using approximate differentiators with $\lambda = 1, 10, 100$. In standard fictitious play, empirical frequencies perpetually oscillate. Note that even with derivative action oscillations do occur, but with smaller magnitude. Indeed, a linearization analysis reveals that the dynamics are not locally exponentially stable.

Now consider a modified Shapley example

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

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) $q_1^* = (1/3, 1/3, 1/3)$ and $q_2^* = (3/7, 1/7, 3/7)$. Figure 2 shows the empirical frequency responses under derivative action FP with approximate differentiators with $\lambda = 1, 10, 100$. Although the empirical frequencies apparently converge, they are converging to $q_1(5) \approx$

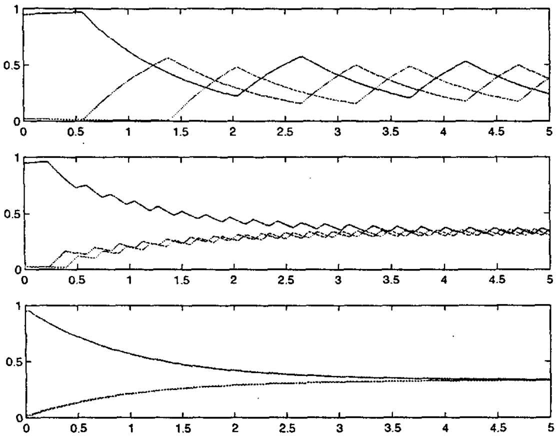

Fig. 1. Shapley counterexample continuous time: Approximate differentiator $\lambda = 1, 10, 100$

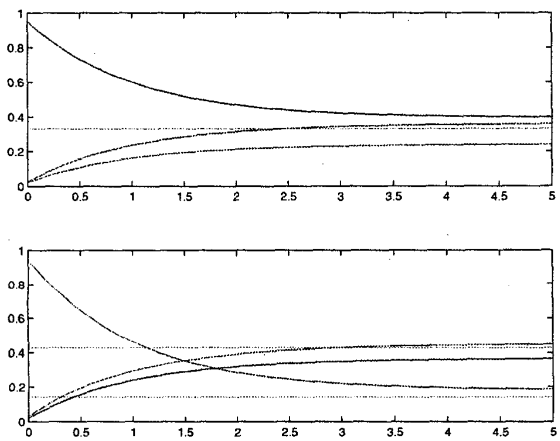

Fig. 2. Modified Shapley counterexample: Approximate differentiator $\lambda = 100$

$[0.3988,\ 0.3627,\ 0.2385]^T$, $q_2(5) \approx [0.3650,\ 0.1853,\ 0.4497]^T$. It is possible to reduce this error by using an alternative approximate differentiatior, in particular

$$\dot{r}_1 = \lambda \mathrm{dz}((q_1 - r_1), 2/\lambda)$$
$$\dot{r}_2 = \lambda \mathrm{dz}((q_2 - r_2), 2/\lambda)$$

The inclusion of a deadzone will shut off the derivative action whenever $q_i(t) - r_i(t)$ is smaller than $2/\lambda$. It is straightforward to show that Theorem 3.4 also holds using the above deadzone differentiator. Furthermore, the effect of the lack of weak convergence is significantly reduced. Figure 3 shows a continuous-time simulation using the deadzone differentiator with $\lambda = 200$. Both cases qualitatively resemble the standard differentiator simulation. However the empirical frequencies are converging to $q_1(5) \approx [0.3391,\ 0.3359,\ 0.3250]^T$ $q_2(5) \approx [0.4362,\ 0.1522,\ 0.4115]^T$.

IV. Dynamic Gradient Play

A. *Ideal Case: Derivative Measurements*

In this section, we will consider an alternative form of continuous time strategy evolution that is not directly related to fictitious play.

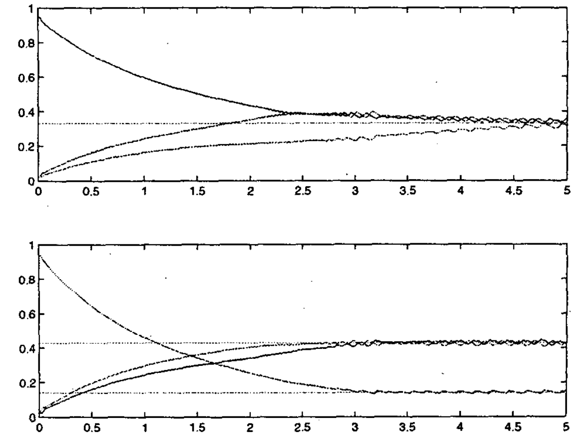

Fig. 3. Modified Shapley counterexample with deadzone differentiator and $\lambda = 200$

We will still explore the evolution of probability distributions, $q_i(t)$, but these should be interpreted as a strategy update mechanism in an effort to compute Nash equlibria. The "empirical frequency" interpretation from fictitious play is no longer immediate (although we will still use this term). The motivation is as follows. Theorem 3.4 suggests that the deficiency of approximate differentiators may be due to the nonlinearity of the $\dot{q}_i$ update mechanism which causes a lack of weak continuity. Updating strategies in a "more linear" fashion should alleviate this situation. One such possiblility is that of gradient evolution. We will review standard gradient play and propose a derivative action modification. As before, the inclusion of derivative action will allow certain desirable convergence properties.

Since each player seeks to maximize its own utility in response to observations of an opponent's actions, we can write as a gradient update mechanism

$$\dot{q}_i(t) = NN^T(M_i q_{-i}(t) - \tau \log(q_i(t)))$$

where N is an orthonormal matrix whose columns span the null space of the row vector $\mathbf{1}^T \in \mathcal{R}^m$.

As is the case with fictitious play, gradient based evolution need not converge. Indeed, the above dynamics do not converge for the Shapley counterexample.

We will consider a modification of gradient evolution in the spirit of the prior modification of fictitious play, namely, the use of derivative action.

We first investigate the special case of $\tau = 0$.

Introducing a derivative term in same manner as derivative action fictitious play leads to the implicit differential equation

$$\dot{q}_1 = \lim_{\varepsilon \downarrow 0} \frac{\Pi_\Delta\left[q_1 + \varepsilon(M_1(q_2 + \dot{q}_2)\right] - q_1}{\varepsilon} \tag{8a}$$

$$\dot{q}_2 = \lim_{\varepsilon \downarrow 0} \frac{\Pi_\Delta\left[q_2 + \varepsilon(M_2(q_1 + \dot{q}_1)\right] - q_2}{\varepsilon} \tag{8b}$$

which we refer to as *derivative action gradient play*. As in derivative action FP, the modification leads to an implicit differential equation. This reflects the intention that player $\mathcal{P}_1$'s strategy update mechanism involves processing q_2 *and* $\dot{q}_2$. Similarly, player $\mathcal{P}_2$'s strategy update mechanism involves processing q_1 and $\dot{q}_1$.

Define the matrices

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

$$E = \mathcal{N}\mathcal{N}^T \begin{pmatrix} 0 & M_1 \\ M_2 & 0 \end{pmatrix}$$

Assumption 4.1: The matrix $(I - E)$ is invertible. Furthermore,

$$\mathcal{N}^T(I-E)^{-1}E\mathcal{N} + (\mathcal{N}^T(I-E)^{-1}E\mathcal{N})^T < 0 \quad (9)$$

Proposition 4.1: Suppose there exists a solution to the derivative action gradient play (8). If for some $t \geq 0$,

$$q_i(t) \in \mathrm{Int}(\Delta(m))$$

then

$$\dot{q}(t) = (I-E)^{-1}Eq(t) \quad (10)$$

where $q(t) = \begin{pmatrix} q_1(t) \\ q_2(t) \end{pmatrix}$.

Theorem 4.1: Let $q^* = \begin{pmatrix} q_1^* \\ q_2^* \end{pmatrix}$ be a mixed Nash equilibrium such that

$$q_i^* \in Int(\Delta(m))$$

Then for initial conditions $q(0)$ sufficiently close to q^*, the solution to derivative action gradient play (8) satisfies the linear system (10) for all $t \geq 0$ and exponentially converges to q^*.

The main conclusion of Theorem 4.1 is that mixed Nash equilibria are locally attractive under derivative action fictitious play. A critical assumption was the negative definiteness of Assumption 4.1. It turns out that one can "generically" satisfy this assumption with suitable scaling of the utility matrices M_i.

Proposition 4.2: Let

$$M_1 = \gamma\tilde{M}_1, \quad M_2 = \gamma\tilde{M}_2$$

and suppose that for $i = 1, 2$,

$$N^T\tilde{M}_iN$$

is invertible. Define

$$\tilde{E} = \mathcal{N}\mathcal{N}^T \begin{pmatrix} 0 & \tilde{M}_1 \\ \tilde{M}_2 & 0 \end{pmatrix}$$

There exists a $\gamma^* > 0$ such that if

$$(I - \gamma\tilde{E})$$

is invertible for $\gamma > \gamma^*$, then the negative definiteness of Assumption 4.1 holds, i.e.,

$$\mathcal{N}^T(I-\gamma\tilde{E})^{-1}\gamma\tilde{E}\mathcal{N} + (\mathcal{N}^T(I-\gamma\tilde{E})^{-1}\gamma\tilde{E}\mathcal{N})^T < 0$$

We now consider derivative action gradient play in the case of nonzero τ. The natural extention of derivative action gradient play (8) is

$$\dot{q}_1 = NN^T(M_1(q_2 + \dot{q}_2) - \tau\log(q_1))$$
$$\dot{q}_2 = NN^T(M_2(q_1 + \dot{q}_1) - \tau\log(q_2))$$

where we presume that the $q_i(t)$ lie in simplex interiors. The problem with these dynamics is that the $\log(\cdot)$ terms may not perform their intended "barrier" function of maintaining the evolution in the simplex interior. The intuition is as follows. The $\dot{q}_1$ dynamics contain a $\log(\cdot)$ term, but also contain a $\dot{q}_2$ term, which in turn involves a $\log(\cdot)$ term. It is unclear whether one can guarantee *a priori* that these two terms will not somehow cancel each other.

We can avoid these issues by considering a slightly perturbed form of these dynamics, namely

$$\dot{q}_1 = NN^T(M_1q_2 + M_1(\dot{q}_2 + \tau NN^T\log(q_2)) - \tau\log(q_1)) \quad (11a)$$

$$\dot{q}_2 = NN^T(M_2q_1 + M_2(\dot{q}_1 + \tau NN^T\log(q_1)) - \tau\log(q_2)) \quad (11b)$$

which we will refer to as *perturbed derivative action gradient play*. Note the "removal" of the $\log(\cdot)$ terms in the processing of the opponent's derivative. This alternative form will be easier to analyze, but has the side effect of having equilibria that are not Nash equilibria. Fortunately, the size of the shift seems on the order of the shift in equilibria associated with smoothing a matrix game.

A direct interpretation of the alternative dynamics (11) is to introduce $\log(\cdot)$ terms directly into the linear system (10) associated with the $\tau = 0$ case, i.e.,

$$\dot{q} = (I-E)^{-1}Eq - \tau\mathcal{N}\mathcal{N}^T\log(q) \quad (12)$$

It is easy to see that solutions to the above (12) are also solutions to the perturbed derivative action gradient play (11).

Theorem 4.2: Consider perturbed derivative action gradient play (11) under Assumption 4.1. For initial conditions $q_i(0)$ in simplex interiors, there exists a solution for all $t \geq 0$ such that

$$\lim_{t\to 0} \dot{q}(t) = 0$$

B. Approximate Differentiator Implementation and Simulation Examples

As in the case of derivative action fictitious play, we can implement derivative action gradient play by employing approximate differentiators. Through a suitable modification, such equations can be written in a "fictitious play" form that is equivalent near equilibrium points.

We can write derivative action gradient play with $\tau = 0$ as

$$\dot{q}_1 = q_1 + NN^T(M_1(q_2 + \dot{r}_2)) - q_1$$
$$\dot{q}_2 = q_2 + NN^T(M_2(q_1 + \dot{r}_1)) - q_2$$
$$\dot{r}_1 = \lambda(q_1 - r_1)$$
$$\dot{r}_2 = \lambda(q_2 - r_2)$$

Now introduce simplex projections to write

$$\dot{q}_1 = \Pi_\Delta[q_1 + M_1(q_2 + \dot{r}_2)] - q_1 \quad (13a)$$
$$\dot{q}_2 = \Pi_\Delta[q_2 + M_2(q_1 + \dot{r}_1)] - q_2 \quad (13b)$$
$$\dot{r}_1 = \lambda(q_1 - r_1) \quad (13c)$$
$$\dot{r}_2 = \lambda(q_2 - r_2) \quad (13d)$$

With this representation, each player uses the strategy

$$p_i = \Pi_\Delta[q_i + M_{-i}(q_{-1} + \dot{r}_{-i})]$$

As before, simulations [12] show that empirical frequencies converge in the Shapley example. Furthermore, these converge in the modified Shapley example without recourse to deadzone differentiators.

V. MULTIPLAYER GAMES

Let us now consider the case with N 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.

If we assume that each player has a differentiable best response function

$$\beta_i(p_i, p_{-i})$$

then we can write derivative action as

$$\dot{q}_1 = \beta_1(q_{-1} + \dot{q}_{-1}) - q_1$$
$$\vdots$$
$$\dot{q}_N = \beta_N(q_{-N} + \dot{q}_{-N}) - q_N$$

The ensuing analysis of approximate derivative measurements, approximate differentiator implmentation, and weak convergence issues remains the same. Indeed, the specific structure of the utility functions other than differentiability of the best response map is not of concern.

In the dynamic gradient play scenario, we need to impose the following structure on the player utilities,

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

characterized by matrices M_{ij}. Let us assume for convenience that each player has the same dimension. As before, let N be an orthonormal matrix whose columns span the nullspace of $\mathbf{1}^T$. Redefine the matrix

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

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

$$E = \mathcal{N}\mathcal{N}^T\mathcal{M}$$

Then the underlying linear structure (analogous to (10)) is still

$$\dot{q} = (I - E)^{-1} E q$$

and the sufficient condition for negative definiteness (analogous to Proposition 4.2) is the invertibility of $\mathcal{N}^T\mathcal{M}\mathcal{N}$.

The full version [12] illustrates these methods on a smoothed version of the Jordan anti-coordination game [9]. It is known that standard FP does not converge for this game. Reference [8] 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 [8], an extension of the Jordan game can be written as

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

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

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

Simulations [12] show that both derivative action FP and gradient play converge to the Nash equilibrium. As before, a non-symmetric equilibrium can cause a bias in the convergence in derivative action FP, but deadzone differentiators resolve the problem.

VI. Concluding Remarks

We have introduced the use of derivative action in repeated games. We have shown that in the ideal case, derivative action can guarantee convergence to Nash equilibria in repeated games with best response dynamics. In the non-ideal case, we have shown that approximate differentiators can recover the ideal behavior, but may also exhibit detrimental biasing effects. We have illustrated how to mitigate these effects through alternative approximate differentiator mechanisms. We also considered derivative action in gradient dynamics and illustrated that the detrimental effect approximate differentiators is diminished.

Finally, we did not formally establish any ties to discrete time. The dynamics under consideration are of the simplest 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 [10] or papers such as [1] provide tools to establish this connection.

VII. References

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

[2] 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.

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

[4] D.P. Foster and H.P. Young. Learning, hypothesis testing, and Nash equilibrium. Technical report, 2002.

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

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

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

[8] S. Hart and A. Mas-Colell. Uncoupled dynamics cannot lead to Nash equilibrium. Technical Report DP–299, The Hebrew University of Jerusalem, Center for Rationality, 2002.

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

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

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

[12] J. S. Shamma and G. Arslan. Dynamic fictitious play and distributed convergence to nash equilibria, mimeo, 2003.

[13] J. S. Shamma and G. Arslan. On the convergence of continuous-time fictitious play, mimeo, 2003.

[14] 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.

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