On Some Extensions of Fictitious Play

Jerome Le Ny

May 20, 2006

Abstract

This report considers extensions of fictitious play, a well-known model of learning in games. We review stochastic fictitious play as introduced in [FK93] as an attempt to solve some issues associated with standard fictitious play. Since the original paper on stochastic fictitious play, the traditional way of studying the convergence of this algorithm is via the theory of stochastic approximation algorithms. This leads to the study of an ODE which captures the mean dynamics of the system. For this continuous-time system, we review some convergence results, as well as ideas from the control theory litterature aimed at stabilizing Shapley's game, for which it is known that standard fictitious play does not converge. We conclude with some potential future research directions.

Contents

1 Discrete-Time Fictitious Play and Smooth Fictitious Play

1.1 Review of Fictitious Play

In this section, we briefly review the fictitious play (FP) learning model and also introduce smooth fictitious play to try to correct some drawbacks of pure FP. This section also fixes the notation used in the report.

In this first part, when we introduce smooth fictitious play, we consider a game with I players, playing the strategic form game G at time $k = 1, 2, \dots$. Player i has a finite strategy space S_i with m_i possible actions. Later on, when we study convergence issues, we focus on the case where we have only two players (for multiplayer fictitious play, there are apparently modeling issues, see [FL99] p. 37).

Let $\Delta(n)$ denote the probability simplex in $\mathbb{R}^n$, i.e.,

$$\Delta(n) = \left\{ s \in \mathbb{R}^n | s \geq 0 \text{ and } \mathbf{1}^T s = 1 \right\},$$

where the inequality $\geq$ is meant componentwise, and $\mathbf{1}$ is always the all-one vector of appropriate dimension. At each state, player i selects a strategy $p_i \in \Delta(m_i)$ (i.e., in general a strategy can be mixed) and receives an *expected* stage payoff $g_i(p_i, p_{-i})$. Let M_i denote the payoff matrix for player i. Then when player i selects action $a_i \in S_i = \{1, \dots, m_i\}$ according to the probability vector p_i, he receives a reward

$$\mathbf{v}_{a_i}^T M_i \mathbf{v}_{a_{-i}},$$

where $\mathbf{v}_i$ denotes the vertex of the appropriate probability simplex whose ith coordinate equals 1 and the remaining coordinates equal 0. Thus the expected stage payoff is

$$g_i(p_i, p_{-i}) = p_i^T M_i p_{-i}.$$

In the following, by abuse of notation, we will also use the notation a_i instead of $\mathbf{v}_{a_i}$ for a deterministic strategy vector, especially in the argument of g_i. $[p_i]_{a_i}$ corresponds to the coordinate a_i of p_i, for $a_i \in S_i$. Finally, we let $\Delta = \Delta(m_1) \times \cdots \times \Delta(m_I)$.

In fictitious play, players assumes at each period that their opponent is using a *stationary* mixed strategy. They choose their actions to maximize that period's expected payoff given their prediction of the distribution of opponents' actions, which they form according to the empirical frequency denoted here by the vector $q_i(t) \in \Delta(m_i)$. $[q_i(t)]_{a_i}$ contains the running average frequency of player's i a_ith action; player $-i$ can form this vector as long as he is able to observe player's i actions, and *in particular the players do not need to know the utility functions of their opponents.* We have

$$q_i(k) = \frac{1}{k} \sum_{\tau=1}^{k} \mathbf{v}_{a_i(\tau)},$$

and note that we can evaluate q_i recursively as

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

where $a_i(k)$ is the action of player i at time k. We can start at a nonzero point (or fictitious past) $q_i(0)$. Then in FP, player i chooses at time k its action $a_i(k)$ according to

$$a_i(k) \in \arg \max_{s_i \in S_i} g_i(s_i, q_{-i}(k)).$$

Note that this means that at each stage, a player implements a pure strategy (except in some denegerate cases where the payoff function is such that several strategies realize the maximum, in which case the player could choose to randomize).

Recall from [FL99], chapter 2, the two main types of convergence considered for fictitious play.

1. Consider the sequence of play a(1), a(2),..., with $a(k) = (a_1(k), \dots, a_I(k))$. We say that the sequence $\{a(k)\}$ converges to $\bar{a}$ if there exists T such that $a(k) = \bar{a}$ for all $t \geq T$. It is easy to see that if $\{a(k)\}$ converges to $\bar{a}$, then $\bar{a}$ is a (pure) Nash Equilibrium. Moreover, for a FP path $\{a(k)\}$, if for some k, $a(k) = a^*$, where a^* is a strict NE of G, then $a(\tau) = a^*$ for all $\tau > k$.

2. Since the only pure-strategy profiles that FP can converge to are those that are NE, FP cannot converge to a pure-strategy profile in a game all of whose equilibria are mixed, like for example "matching pennies". There is an alternative notion of convergence for FP, *convergence of empirical distributions* (or in the time-average sense), which applies to mixed strategies as well. Again, we have

Proposition 1.1. *If the empirical distributions $q_i(k)$ over each player's choices converge, the strategy profile corresponding to the product of these distributions is a Nash Equilibrium.*

Here are *some* cases where convergence results for the empirical distributions under FP have been established: 2 player zero-sum games [Rob51], 2 × 2 games [Miy61], multiplayer games with identical player utilities [MS96], two player games in which one player has two moves [Ber05] ....

Denote the time average payoffs through time k as

$$U_i^k = \frac{1}{k+1} \sum_{\tau=0}^{k} g_i(a_i(\tau), a_{-i}(\tau)),$$

and the expected payoffs at time t as

$$\tilde{U}_i^k = \max_{a_i \in S_i} g_i(a_i, q_i(k)).$$

The following proposition is a slight modification of a lemma in the course notes, and can be found in [FL99], p. 42.

Proposition 1.2. *For any initial weights there is a sequence $\epsilon^k \to 0$ such that along any infinite horizon history $\tilde{U}_i^k \geq U_i^k + \epsilon^k$. That is, once there are enough data to outweigh the initial weights, players believe that their current period's expected payoff is at least as large as their average payoff to date.*

There are some issues with the ideas of FP. First, the empirical distributions need not converge. The first counterexample, due to Shapley [Sha64], is a two player 3 move game, which is a modified version of the "Rock-Scissors-Paper" game:

	L	M	R
T	$0,0$	$1,0$	$0,1$
M	$0,1$	$0,0$	$1,0$
D	$1,0$	$0,1$	$0,0$

In this game, there is a unique NE with expected payoffs 1/3 for both players. Therefore, by proposition 1.1, if FP converges, then $\tilde{U}_1^k + \tilde{U}_2^k \to 2/3$. Now it turns out that if the initial weights are $q_1(0) = (1,0,0)$ and $q_2(0) = (0,1,0)$ for example, the three diagonal profiles are never played. So by proposition 1.2, the players expect a sum of payoffs at least 1 for large k. This contradiction shows that we do not have convergence in the time average sense in this game.

The notion of convergence in empirical distribution has also some problems, as illustrated by Fudenberg and Kreps' example of persistent miscoordination [FK93], an example of which is reproduced in the course notes:

	A	B
A	$1,1$	$0,0$
B	$0,0$	$1,1$

In this game, the empirical joint distribution on pairs of actions does not equal the product of the two marginal distributions, so the empirical joint distribution corresponds to correlated as opposed to independent play. Specifically here, taking $q_1(0) = (1/2, 0)$ and $q_2(0) = (0, 1/2)$, the play follows the alternating sequence (A, B), (B, A), $(A, B), \ldots$, so the empirical frequencies converge to the Nash equilibrium $((1/2, 1/2), (1/2, 1/2))$, but the players never successfully coordinate. This behavior is not satisfactory since we assume that the players believe that they play against i.i.d. draws from the long-run empirical distribution of their opponents' play, and in particular one might wonder if players would ignore cycles in their opponents' play.

1.2 Stochastic Fictitious Play

The traditional process of fictitious play is deterministic, i.e., for generic payoffs and fictitious past, the players will use pure strategies in every period. We now introduce alternative models in the spirit of fictitious play, with two main motivations. Firstly, we would like a more satisfactory explanation for convergence to mixed-strategy equilibria: recall the miscoordination example above, where the players get a payoff considerably less than the amount they could have guaranteed themselves by randomizing $(1/2, 1/2)$ at each period. Secondly, we would like to avoid the discontinuity inherent in standard fictitious play, where a small change in the data can lead to an abrupt change in behavior.

1.2.1 Random Utility Model

Here is a possible model of learning to play mixed strategies ([Har73], [FK93], [FL99], p.105): *we perturb the original game* by changing the payoffs from $g_i(s)$ to $u_i(s) = g_i(s) + \eta_i(s_i)$, where η_i is a random variable, depending only on player i's action (that point is not clear, [BH99] seems no to require this assumption).

We assume that η_i has a distribution which is absolutely continuous with respect to the Lebesgue measure, in order to ensure uniqueness of the best response: this is one nice advantage of pertubing the game, we do not have to deal with correspondences any more. See [FK93], lemma 7.2., or [BH99], p.41. Then the *best response distribution* (or best response map) for player i is the deterministic map

$$\beta_i : S_{-i} \to S_i$$
$$[\beta_i(p_{-i})]_{a_i} = P(\eta_i \text{ s.t. } a_i = \arg\max_{s_i \in S_i} u_i(s_i, p_{-i}))$$

Definition 1.1. The Nash map $\nu : \Delta \to \Delta$ is defined by $\nu(p_1, \ldots, p_I) = (\beta_1(p_{-1}), \ldots, \beta_I(p_{-I}))$.

ν is Lipschitz, C^r of analytic provided the probability distribution functions of the random vectors (or matrices in [BH99]) η_i have the corresponding property. We require

Hypothesis 1.1. The Nash map ν is Lipschitz continuous.

In the learning model, this will mean that *if a player's assessment converges, his behavior will too*, i.e., we will avoid the miscoordination problem of standard FP. The function β_i is both continuous and close to the original best-response correspondence. For example, in the game of matching pennies, the best response of player 1 to the strategy of player 2 is a step function (play head with probability 1 if 2 plays head with any probability greater than 1/2); the best response distribution is a smoothed approximation of this step function, where generally even if the opposing player is playing a pure strategy, the smoothed best response will still be random. See [FL99], in particular about experiments in psychology, for justifications of introducing a smoothed best-response.

The notion of Nash equilibrium is defined as usual:

Definition 1.2. We say that the profile p is a *Nash distribution* if for all i, $p_i = \beta_i(p_{-i})$, i.e. $p = \nu(p)$.

Since we assumed ν to be continuous, the existence of a Nash distribution equilibrium follows from Brouwer's fixed point theorem.

1.2.2 Universal Consistency and Smooth Fictitious Play

Another explanation for smooth fictitious play besides the random utility model is that player may choose to randomize as a sort of protection from mistakes in their model of opponents' play. In particular, a desired feature of the learning rule is universal consistency, which requires, regardless of the opponents' play, that players almost surely get at least as much utility as they could have gotten had they known the frequency but not the order of observations in advance. In [FK93] p.118, it is argued that universal consistency can be accomplished by a smooth fictitious play procedure in which β_i is derived from maximizing a function of the form $g_i(p) + \tau v_i(p_i)$, $\tau > 0$, where g_i is the original utility function, and v_i belongs to the class of admissible functions satisfying the following properties:

1. v_i is a smooth, strictly concave differentiable function.
2. $\lim_{p_i \to \partial\Delta(m_i)} |v_i'(p_i)| = \infty$

As in the random utility model, these conditions imply *unicity* and *strict interiority* of the solution to the stage maximization problem so that every strategy is played with strictly positive probability regardless of the frequency of opponents' play.

An explicit example for the function v_i, that will be used in the following, is the entropy function:

$$v_i(p_i) = H(p_i) = \sum_{a_i} -[p_i]_{a_i} \log[p_i]_{a_i} = -p_i^T \log(p_i).$$

Then the utility functions become:

$$u_i(p_i, p_{-i}) = p_i^T M_i p_{-i} + \tau H(p_i).$$

In this case, we can explicitely solve for $\beta_i : \Delta(m_{-i}) \to \Delta(m_i)$, by

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

Define, for any dimension n, the "logit" or "soft-max" function

$$\begin{aligned} L &: \mathbb{R}^n \to \operatorname{Int}(\Delta(n)) \\ (L(x))_i &= \frac{e^{x_i}}{e^{x_i} + \ldots + e^{x_n}}. \end{aligned}$$

Then it turns out that we have

$$\beta_i(p_{-i}) = L(\frac{M_i p_{-i}}{\tau}),$$

or in other words,

$$[\beta_i(p_{-i})]_{a_i} = \frac{\exp(\frac{1}{\tau} g_i(a_i, p_{-i}))}{\sum_{s_i} \exp(\frac{1}{\tau} g_i(s_i, p_{-i}))}.$$

This is a special case of stochastic fictitious play referred to as *logistic fictitious play*, where each strategy is played in proportion to an exponential function of the utility it has historically yielded; it corresponds to the logit decision model that has been extensively used in empirical work. Note finally that as $\tau \to 0$, the probability that any strategy that is not a best response for g_i is played goes to 0. The case where $\tau = 0$ corresponds to the standard FP, and $\tau > 0$ forces mixed strategies.

To conclude our discussion on the two ways of perturbing the original game, we note the following interesting result

Theorem 1.3 ([HS02]). *Suppose that in the random utility model, the random variables η_i have strictly positive densities and are such that ν is continuously differentiable. Then there exists an admissible (in the sense above) deterministic perturbation V such that*

$$\nu(p) = arg \max_{y \in int\Delta} (g(y) + V(y)).$$

The proof is constructive and V appears as a Legendre transform.

2 Continuous-Time Fictitious Play

2.1 Continuous-Time Dynamics and Stochastic Approximation Algorithms

For both standard FP ($\tau = 0$ in the deterministic model) and smooth FP ($\tau > 0$), we have the recursion for the update of the empirical frequency

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

Then the strategy of player i at time t is given by the distribution

$$p_i(k) = [\nu(q(k))]_i = \beta_i(q_{-i}(k)). \tag{2}$$

The state of the game at time k is the vector $q(k) = (q_1(k), \ldots, q_I(k))$ listing the players' empirical frequencies at time k. When any model of stochastic FP is followed, the state sequence $q(t)$ is a nonstationary discrete-time Markov process, with values in the compact set Δ. Note also that we have

$$p_i(k) = E[\mathbf{v}_{a_i(k)} \,|\, q(k)]. \tag{3}$$

From (1) and (3), we have

$$E[q(k+1) - q(k) \,|\, q(k)] = \frac{1}{k+1}(\,\nu(q(k)) - q(k)\,). \tag{4}$$

The *game vector field* is defined on Δ as $F(x) = \nu(x) - x$. We can view Δ as a submanifold (with corner) in $\mathbb{R}^d$ and its tangent space at every point is identified with the linear subspace

$$T\Delta = \left\{ (y^1, \ldots, y^I) \in \mathbb{R}^{m_1} \times \ldots \times \mathbb{R}^{m_I} : \sum_{j=1}^{m_i} y^i_j = 0,\ i = 1, \ldots, I \right\}, \tag{5}$$

and then F is indeed a map from Δ to $T\Delta$. It measures the extent to which x is not a Nash distribution equilibrium. We also have

$$F(x) = (k+1)E[q(k+1) - q(k) \,|\, q(k) = x], \tag{6}$$

in other words, if the state at time k is x, then $F(x)$ is $k+1$ times the expected change in the state. The analysis of the convergence of stochastic FP traditionally relies on a close connection between limits of the sample paths $\{q(k)\}$ and the dynamics of the deterministic differential equation

$$\frac{dx}{dt} = F(x), \tag{7}$$

which represents the "mean" dynamics of the algorithm. From the form of the vector field (in particular $\nu(x) \in \text{int}\Delta$), it is easy to see that the trajectories of (7) starting in Δ remain in Δ. Note the obvious fact:

Proposition 2.1. *The zeroes of the game vector field F, i.e., the equilibrium points of (7), are the Nash distribution equilibria.*

We can rewrite the recursion for the state vector as

$$q(k+1) - q(k) = \frac{1}{k+1}[\,F(q(k)) + Z_{k+1}\,],$$

where the random variable $Z_{k+1} = (k+1)(q(k+1) - q(k)) - F(q(k))$ is a martingale difference, i.e., using (6),

$$E[Z_{k+1} \,|\, q(k)] = 0.$$

Such a recursion is a particular form of a *stochastic approximation algorithm* (see for example [KY03]). In the most classical form studied in the litterature, F is the negative of the gradient of a continuously differentiable function, in which case the algorithm corresponds to a noisy version of a gradient descent; here however, F can be quite arbitrary and this allows the algorithm to have a possibly complicated asymptotic behavior. See for example [Ben99, KY03].

Studying the relationship between the continuous time system (7) and the discrete time system (1) seems to be relatively hard, and in the rest of this paragraph, I merely try to give some additional motivation for considering the continuous-time system in the following.

Given a state sequence $\{q(k)\}_{k\in\mathbb{N}}$ resulting from infinitely repeated fictitious play, we say that a point q^* is a limit point of $\{q(k)\}_{k\in\mathbb{N}}$ if $\lim_{i\to\infty} q(k_i) = q^*$ for some sequence $k_i \to \infty$. The set of such limit points is the *state limit set* $L\{q(k)\}$ (Note that because the sequence $\{q(k)\}$ is a random variable, the state limit set is a set valued random variable).

The basic theorems in the theory of stochastic approximation algorithms ([KY03], chapters 5 and 6) guarantee only that $q(k)$ converges with probability one to an invariant set of the mean ODE (which is not useful here, because Δ itself is an invariant set). But sometimes the largest invariant set contains points to which convergence cannot occur. The idea of *chain recurrence*, introduced by Benaïm, can simplify the analysis in general, since it can be shown that the convergence must be to a subset of the invariant set that is chain recurrent. Intuitively, the chain recurrent states are those that can arise in the long run if the flow is subject to small shocks occurring at isolated moments in time. See for example [Ben99]. Here is an important theorem along these lines, cited here purely as a motivation for further study (compare with proposition 5.3 in [Ben99] to make the connection between attractor-free and chain recurrence).

Theorem 2.2 (The Limit Set Theorem [BH99])**.** *With probability* 1*, the state limit set* $L\{q(k)\}$ *has the following properties:*

1. $L\{q(k)\}$ *is an invariant set for the flow of the game vector field* F*.*
2. $L\{q(k)\}$ *is compact, connected and attractor-free.*

This theorem in particular allows Benaïm and Hirsch to prove rigorously convergence of stochastic fictitious play for 2 × 2 games. [HS02] gives rigorous convergence results on the global stability of Nash equilibria in stochastic FP for zero sum games, games with an interior evolutionary stable strategy, potential games, supermodular games (see the paper for the precise results).

As a last warning for the next section that looking only at the ω-limit sets of the continuous time system does not necessarily tell us much in terms of convergence of the discrete-time system with noise, consider the following example.

Example 2.1 ([HS02])**.** Consider a flow on a circle that moves clockwise everywhere except at a single rest point (Fig. 1). This rest point is the unique ω-limit point of the flow. Now suppose the flow represents the

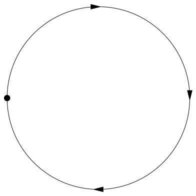

Figure 1: ω-limit sets and chain recurrence.

expected motion of some underlying stochastic process. If the stochastic process reaches the rest point, its expected motion is zero. Nevertheless, actual motion may occur with positive probability and in particular the process can jump past the rest point and begin another circuit. Therefore in the long run all regions of the circle are visited infinitely often. The long run behavior is captured by the notion of chain recurrence, as all points on the circle are chain recurrent under the flow.

2.2 Continuous-Time Fictitious Play: Convergence Results.

From now on we consider the two player case, and we look at the mean ODE only. We call these equations *continuous-time FP*:

$$\begin{cases} \dot{q}_1(t) &= \beta_1(q_2(t)) - q_1(t) \\ \dot{q}_2(t) &= \beta_2(q_1(t)) - q_2(t). \end{cases} \tag{8}$$

[SA04] proves the following theorem, in the case of logistic fictitious play.

Theorem 2.3. *Consider the following cases, for the particular case of logistic fictitious play:*

- $M_1 = -M_2^T$ *(unpertubed game is zero sum)*
- $M_1 = M_2^T$ *(identical interest unperturbed game)*
- $m_1 = m_2 = 2$ *(2×2 game, for generic payoffs, see the paper)*
- $m_1 = 2$ *($2 \times n$ game) and there is a finite number of Nash equilibria.*

Then the solutions of continuous-time FP (8) satisfy

$$\lim_{t\to\infty} (q_1(t) - \beta_1(q_2(t))) = 0$$
$$\lim_{t\to\infty} (q_2(t) - \beta_2(q_1(t))) = 0$$

Some comments are in order. First of all, the proofs are standard, using Lasalle's invariance principle for the zero sum case, Barbalat's lemma for the next two cases, and Poincaré-Bendixson's criterion for the last case ([Kha02]). Second, the first three cases involve a Lyapunov function which was introduced by Hofbauer et al. in [HS02, HH00]. At least for the zero sum case, the proof is an exact copy of the proof contained in [HH00], which actually proves the result in the general case of stochastic FP for any admissible cost function, not just logistic fictitious play. Although I did not try to do it precisely, it seems that the only preperties of the entropy function required in the proofs are those required for admissible cost functions, so it would be interesting to recast the theorem in the general framework of stochastic fictitious play. Last, [HH00] have a stronger result in the zero sum case than convergence to a Nash equilibrium: they show that the perturbed game has a *unique* Nash equilibrium, and this fact is actually used in [HS02] to prove the convergence of the discrete-time system. Unicity of the Nash equilibrium in a pertubed zero sum game makes sense intuitively: in the original game, the players are actually indifferent between the Nash equilibria which all realize the value of the game. Introducing an additional strictly concave cost function simply differentiates between potentially multiple equilibria.

3 Dynamic Fictitious Play

As we have discussed, convergence of fictitious play (in discrete-time or only for the continuous-time associated ODE) has been established for special cases, but there are counter-examples, such as Shapley's of Jordan's, which show that convergence is not guaranteed. These examples exhibit the same non-convergent behavior in stochastic fictitious play. Shamma et Arslan [SA05] cite litterature arguing that in most games, update mechanisms that are static functions of empirical frequencies will not converge to a mixed equilibrium (and recall that in the Shapley and Jordan examples, there is a unique Nash equilibrium, which is completely mixed). They introduce a modification to fictitious play, called "derivative action" fictitious play, which can lead in some cases to behaviors converging to Nash equilibria in previously nonconvergent situations. The idea is natural from a feedback control point of view, where it is known that static output feedback need not be stabilizing, while dynamic output feedback generally can be stabilizing.

Suppose that in addition to the empirical frequencies, all players also know the *empirical frequency derivatives* $\dot{q}_i(t)$. Now consider a mechanism where each player's strategy is a best response to a combination of empirical frequencies and a weighted derivative of empirical frequencies:

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

The interpretation is that the derivative term serves as a short term prediction of the opponent's strategy, since

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

and thus 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 equations:

$$\begin{aligned} \dot{q}_1 &= \beta_1(q_2 + \gamma \dot{q}_2) - q_1 \\ \dot{q}_2 &= \beta_1(q_1 + \gamma \dot{q}_1) - q_2 \ , \end{aligned} \tag{9}$$

which we will refer to as *exact derivative action FP* (DAFP). In the following, we will *assume* the existence of solutions to these equations, which is not guaranteed in general. In actuality, the derivative is not directly measurable, but must be reconstructed from empirical frequency measurements. Toward this end, we will consider the now well posed set of equations

$$\begin{aligned}
\dot{r}_1 &= \lambda(q_1 - r_1) \\
\dot{r}_2 &= \lambda(q_2 - r_2) \\
\dot{q}_1 &= \beta_1(q_2 + \gamma\lambda(q_2 - r_2)) - q_1 = \beta_1(q_2 + \gamma\dot{r}_2) - q_1 \\
\dot{q}_2 &= \beta_2(q_2 + \gamma\lambda(q_1 - r_1)) - q_2 = \beta_2(q_1 + \gamma\dot{r}_1) - q_2,
\end{aligned} \tag{10}$$

which will refer to as *approximate DAFP*. The intention is that, as λ increases, $\dot{r}_i$ closely tracks $\dot{q}_i$.

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

In this case, introducing the variables $z_i = q_i + \dot{q}_{-i}$, we can rewrite (9) as

$$z = \nu(z),$$

i.e., exact DAFP must evolve over fixed points of ν, which are the Nash equilibria of the game. Let $Q^* \subset \Delta(m_1) \times \Delta(m_2)$ be the set of Nash equilibria. Any solution of the exact DAFP dynamics 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 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 case of multiple Nash equilibria, this result does not in itself guarantee convergence of empirical frequencies.

3.2 Approximate DAFP with General Derivative Gain $\gamma > 0$

We consider now approximate DAFP with arbitrary $\gamma > 0$, and characterize the values of γ that result in *local* asymptotic stability of a Nash equilibrum for large values of $\lambda > 0$. Actually, a large part of the paper [SA05] deals with the case where $\gamma = 1$, which at this point is still puzzling to me, and therefore I will not discuss this case: indeed, as the next theorem taken from their paper shows, it turns out that the value $\gamma = 1$ never leads to asymptotic stability.

Let N be an orthonormal matrix whose columns span the null space of the row vector $\mathbf{1}^T \in \mathbb{R}^m$ (i.e., span the tangent space $T\Delta(m)$, see (5)):

$$\mathbf{1}^T N = 0 \quad \textit{and} \quad NN^T = I.$$

Then consider a Nash equilibrium (q_1^*, q_2^*), which leads to an equilibrium $(q_1^*, q_2^*, q_1^*, q_2^*)$ of approximate DAFP (10). We can write

$$\begin{pmatrix} q_1(t) \\ q_2(t) \\ r_1(t) \\ r_2(t) \end{pmatrix} = \begin{pmatrix} q_1^* \\ q_2^* \\ q_1^* \\ q_2^* \end{pmatrix} + \begin{pmatrix} N & 0 & 0 & 0 \\ 0 & N & 0 & 0 \\ 0 & 0 & N & 0 \\ 0 & 0 & 0 & N \end{pmatrix} \delta x(t),$$

or equivalently

$$\delta x(t) = \begin{pmatrix} N^T & 0 & 0 & 0 \\ 0 & N^T & 0 & 0 \\ 0 & 0 & N^T & 0 \\ 0 & 0 & 0 & N^T \end{pmatrix} \left(\begin{pmatrix} q_1(t) \\ q_2(t) \\ r_1(t) \\ r_2(t) \end{pmatrix} - \begin{pmatrix} q_1^* \\ q_2^* \\ q_1^* \\ q_2^* \end{pmatrix} \right).$$

Linearizing (10) around $(q_1^*, q_2^*, q_1^*, q_2^*)$, we obtain

$$\frac{d}{dt}\delta x = \begin{pmatrix} -I & (1+\gamma\lambda)D_1 & 0 & -\gamma\lambda D_1 \\ (1+\gamma\lambda)D_2 & -I & -\gamma\lambda D_2 & 0 \\ \lambda I & 0 & -\lambda I & 0 \\ 0 & \lambda I & 0 & -\lambda I \end{pmatrix} \delta x, \tag{11}$$

with

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

Define

$$\mathcal{D} = \begin{pmatrix} 0 & D_1 \\ D_2 & 0 \end{pmatrix}, \tag{12}$$

and rewrite the linearization above 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. The local asymptotic stability from linearization is also called the second method of Lyapunov in the control litterature [Kha02].

Theorem 3.1. *Consider a two-player game under approximate DAFP (10), in the particular case of the logit model, with a Nash equilibrium (q_1^*, q_2^*). Assume that $-I + \mathcal{D}$ in (12) is nonsingular, and let $a_i + jb_i$ denote the eigenvalues of $-I + \mathcal{D}$. The linearization (11) with $\gamma > 0$ is asymptotically stable for large $\lambda > 0$ if and only if*

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

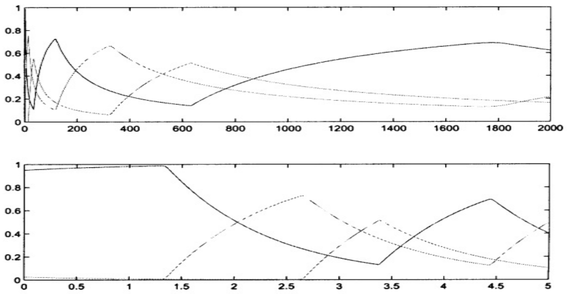

Figure 2: Oscillations in Shapley's game with stochastic FP.

Since $-I + \mathcal{D}$ is the Jacobian matrix of the linearization of standard stochastic FP, theorem 3.1 relates the potential stability of approximate DAFP to the eigenvalues of standard stochastic FP. In particular, the theorem implies that the linearization of approximate DAFP is asymptotically stable whenever the linearization of standard stochastic FP is asymptotically stable (case $\max_i a_i < 0$, we can take $\gamma = 1$). It also implies that approximate DAFP may have a stable linearization in situations where standard stochastic FP does not. Note also from the theorem that $\gamma = 1$ does not stabilize an unstable case of fictitious play.

Remark 3.1. As before, it would be interesting to go through the proofs in more details to know if we can generalize the result to stochastic FP with a general admissible cost function. Again at first glance, it seems that the proofs do not use the particular form of logistic fictitious play explicitely.

3.3 Simulations: DAFP for Shapley's Game

When we apply theorem 3.1 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 approximate DAFP.

Fig. 2 and 3 show oscillations of standard stochastic FP in Shapley's game and convergence with approximate DAFP. The convergence seems to be global from the simulations. Also the discrete time system associated to approximate DAFP apparently converges as well.

4 Conclusion

It seems at this point that we still do not have a clear picture of the convergence results for the various forms of fictitious play (which is only one possible model of learning proposed in the litterature). Among possible research directions, we can mention:

1. In the case of standard fictitious play (not stochastic FP), could we use a variant of the Lyapunov function introduced by Hofbauer and al. for stochastic continuous time FP in order to give unifying

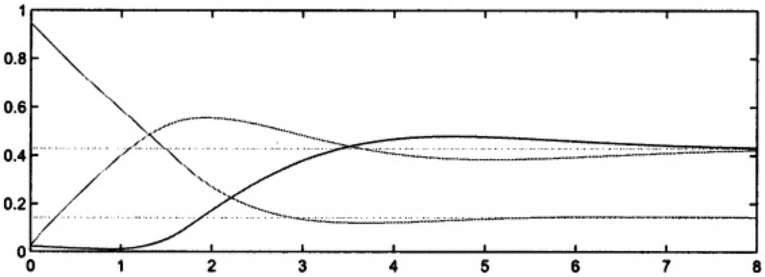

Figure 3: Convergence in a modified Shapley game with approximate DAFP.

proofs of convergence of the (deterministic) discrete-time system? Indeed, the current proofs do not seem to be particularly insightful (idea suggested by Prof. Ozdaglar).

2. Settle the question of the generalization of the proofs of Shamma et al. to the general version of stochastic FP.

3. [SA04] proves convergence of the continuous time system in the case of identical interests and 2 × n games: it seems that the rigorous results of convergence of discrete-time stochastic FP in these two cases have not yet appeared (convergence in zero-sum and 2 × 2 games for stochastic FP was proved rigorously earlier).

4. For dynamic FP, a lot remains to be done: extension to general stochastic FP, studying the global convergence of the continuous time system as opposed to the current local result, and stating a clear and rigorous result about the convergence of the discrete-time system, in particular: does it stabilize Shapley's game with probability one?

References

[Ben99] M. Benaim. Dynamics of stochastic approximation algorithms. In *Le Séminaire de Probabilités XXXIII*, volume 1709 of *Lecture Notes in Mathematics*, pages 1–68. Springer-Verlag, 1999.

[Ber05] U. Berger. Fictitious play in 2 × n games. *Journal of Economic Theory*, 120:139–154, 2005.

[BH99] M. Benaïm and M. Hirsch. Mixed equilibria and dynamical systems arising from ficticious play in perturbed games. *Games and Economic Behavior*, pages 36–72, 1999.

[FK93] D. Fudenberg and D. Kreps. Learning mixed equilibria. *Games and Economic Behavior*, 5:320–67, 1993.

[FL99] D. Fudenberg and D.K. Levine. *The Theory of Learning in Games.* MIT Press, Cambdrige, MA, 1999.

[Har73] J. Harsanyi. Games with randomly disturbed payoffs: A new rationale for mixed-strategy equilibrium points. *International Journal of Game Theory*, 2:1–23, 1973.

[HH00] J. Hofbauer and E. Hopkins. Learning in perturbed asymmetric games. Technical report, Universitä Wien and University of Edinburgh, 2000.

[HS02] J. Hofbauer and W.H. Sandholm. On the global convergence of stochastic fictitious play. *Econometrica*, 70(6):2265–2294, November 2002.

[Kha02] H.K. Khalil. *Nonlinear Systems*. Prentice Hall, 3rd edition, 2002.

[KY03] H.J. Kushner and G.G. Yin. *Stochastic Approximation Algorithms and Applications*. Springer-Verlag, New York, 2nd edition, 2003.

[Miy61] K. Miyasawa. On the convergence of learning processes in a 2 × 2 nonzero-sum two person game, 1961.

[MS96] D. Monderer and L.S. Shapley. Fictitious play property for games with identical interests. *Journal of Economic Theory*, 68:258–265, 1996.

[Rob51] J. Robinson. An iterative method of solving a game. *Annals of Mathematics*, 54:29–301, 1951.

[SA04] J.S. Shamma and G. Arslan. Unified convergence proofs of continuous-time fictitious play. *IEEE Transactions on Automatic Control*, pages 1137–1142, July 2004.

[SA05] J.S. Shamma and G. Arslan. Dynamic fictitious play, dynamic gradient play, and distributed convergence to Nash equilibria. *IEEE Transactions on Automatic Control*, 50(3), March 2005.

[Sha64] 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, Princeton, NJ, 1964. Princeton University Press.