A Review of Smooth Fictitious Play and Fictitious Play in Repeated Games

Justin Kang
Department of Electrical and Computer Engineering
University of Toronto
10 King's College Road, Toronto, Ontario M5S3G4, Canada
js.kang@mail.utoronto.ca

***Abstract*—In this review we compare two common methods for playing repeated games: Fictitious Play (FP) and Smooth (or Stochastic) Fictitious Play (sFP). We formulate both techniques theoretically and describe how FP can be considered as a limit of sFP. We then study their convergence in two-player zero-sum games by both surveying the literature on the theoretical guarantees of convergence, and investigating the convergence via simulation. In particular we use simulations to exhibit the difference between convergence in belief (for FP and sFP) and convergence in behaviour (for sFP but not FP). We discuss how this lack of convergence of FP is due to the play of pure strategies. We also discuss convergence of these techniques in more general bimatrix games, where there exist cases where they do not converge. We consider a simple 2×2 coordination game, where although FP and sFP both converge, sFP favours the pure strategy equilibria, and FP converges to the mixed equilibrium in belief, but fails to converge to the value of the game when the mixed NE is played. We also consider the example of Shapley's game and see that there are initial conditions where neither FP nor sFP converge in any sense. We investigate how the temperature parameter can be adjusted to prevent sFP from falling into a cycle in Shapley's game.**

I. Introduction

Fictitious Play (FP) was first introduced by by Brown [1], where he conjectured that such a scheme could be used to calculate the value of a zero-sum game. FP can be viewed as both a numerical technique for computing an equilibrium state of a game, or as a strategy for playing a game. In either case, the key advantage of FP is that it allows one to completely uncouple the dynamics of a game with only partial information about the behaviour of opponents. Only opponent actions are made public, while in comparison, a technique such as Best Response Play (BRP) requires full knowledge of opponent's mixed strategy. Due to this difference in accessible information, BRP is often referred to as a *full information* scheme, while FP is known as a *partial information* scheme.

Ever since Brown's initial work, there has been significant effort to prove that FP converges in a variety of games. There are different notions of convergence, for example, convergence in payoff (or convergence in value), convergence in belief, in which the long run empirical distribution of player action converges to a Nash equilibrium, or convergence in behaviour. As we will see, in some cases, one of these can imply the other, but it is not always the case. There are a variety of convergence results for fictitious play, including for two-player zero-sum games [2], 2×2 games [3], potential games [4], and games with an interior evolutionary stable state (ESS) [5].

However, FP is not above criticism, and its origins as a heuristic from the very early days of Game Theory means that there is room for refinement. This is precisely what happened when Fudenberg and Kreps [6] first introduced their notion of *smooth* or *stochastic* Fictitious Play (sFP). The core idea of sFP aims to address one of the key issues with FP: the fact that each agent always plays a pure strategy. Fudenberg and Kreps argue that the notion of convergence of an empirical average of pure strategies to a mixed strategy is not appropriate, since it does not capture the random nature in which mixed strategies are played. Instead, they return to the formulation of BRP, and consider a mixed strategy chosen from the perturbed best-response of an empirical average of opponent actions. Since their initial formulation, many convergence results have been established, primarily through the connection with stochastic approximation, and modelling the dynamics using ODEs.

In the following sections we formulate FP and sFP. Then we review results on convergence for various types of games, and present simulations of cases where the two schemes converge or fail to converge, allowing us to compare these results to theoretical expectations.

1) Notation: Throughout this paper, boldface upper and lower case symbols such as $\mathbf{A}, \mathbf{v}$ denote a matrix and a vector. a_{ij} denotes the element in i^{th} row and j^{th} column of $\mathbf{A}$. Let $\mathcal{U}_i(a)$ denote the utility function (i.e payoff) of player i (sometimes denoted P_i) given pure strategy $a \in \mathcal{M}_i$ is played. Let $\mathbf{x}_i \in \Delta_i$ denote a mixed strategy of P_i belonging to the simplex Δ_i, and let $\mathbf{x}_{-i} \in \Delta_{-i}$ denote the mixed strategy of the players P_j $j \neq i$, while $\mathbf{x} \in \Delta$ is the mixed strategy of all players.

II. Fictitious Play

We being this section by defining fictitious play.

Definition 1. *Fictitious Play is the scheme in which player P_i plays a pure strategy $\mathbf{e}_i^k$ in the k^{th} round based on the best pure response to the opponent empirical frequencies $\hat{\mathbf{x}}_{-i}^k$:*

$$\mathbf{e}_i^k \in \arg\max_{j \in \mathcal{M}_i} \mathcal{U}_i\left(\mathbf{e}_{ij}, \hat{\mathbf{x}}_{-i}^k\right), \quad k = 1, 2, \ldots, \tag{1}$$

and the full empirical frequency of all the players $\hat{\mathbf{x}}^k = (\hat{\mathbf{x}}_i^k, \hat{\mathbf{x}}_{-i}^k)$ *is updated based on the actions of the players in the* k^{th} *round* $\mathbf{e}_i^k$ *via:*

$$\hat{\mathbf{x}}_i^{k+1} = \frac{k}{k+1}\hat{\mathbf{x}}_i^k + \frac{1}{k+1}\mathbf{e}_i^k \ \ \forall i, \tag{2}$$

and $\hat{\mathbf{x}}_i^1$ *may be chosen arbitrarily (i.e. the first play is random).*

From the definition, we see that we can implement fictitious play with only a limited amount of information. Each player P_i requires:

1) Knowledge of $\mathcal{U}_i$, the player's own utility function.
2) The action of all the players each round $\mathbf{e}^k$.

A. Notions of Convergence

In each round, each player plays a *pure* strategy. Does this meant that it is only possible for FP to converge to a pure NE? In general, this is not true, and it is possible for FP to converge to a mixed strategy NE if we use the following condition for convergence.

We say FP converges to $\mathbf{x}^* = (\mathbf{x}_i^*, \mathbf{x}_{-i}^*) \in NE(\mathcal{G})$ of some game $\mathcal{G}$ if and only if

$$\lim_{k\to\infty} \hat{\mathbf{x}}^k = \mathbf{x}^*. \tag{3}$$

It should be noted. however, that the actual play in such a convergent FP may look very different for the play when the mixed NE is played, and there is no guarantee that the value in both cases will be the same.

Thus, it sometimes makes sense to consider convergence in *value*. That is, we have:

$$\lim_{k\to\infty} \frac{1}{k}\sum_k \mathcal{U}_i(\mathbf{e}_i^k) = \mathcal{U}_i(\mathbf{x}_i^*) \ \ \forall i, \ \ \mathbf{x}^* \in NE(\mathcal{G}). \tag{4}$$

We will see that in some cases, both of these forms of convergence are achieved in FP, while in others only one is, and in others neither hold.

Before we move on with our discussion, we should note that the FP of Definition 1 is somewhat different from the definition laid out by Brown in [1]. In Brown's original FP, players to not update their empirical frequencies simultaneously each turn, but instead in an alternating fashion. The convergence of this type of FP has been investigated in [7], however, the vast majority of literature on the topic uses our definition of FP laid out above.

III. Stochastic/Smooth Fictitious Play

To address some of the issues of FP, Fundenberg introduced the idea of sFP [6]. sFP is based around a perturbed BRP. A key idea in the definition of sFP is the use of a perturbed best response function, which is defined as follows:

$$\widetilde{BR}_i(\hat{\mathbf{x}}_{-i}) \triangleq \arg\max_{\mathbf{x}_i\in\Delta_i} \widetilde{\mathcal{U}}_i\left(\mathbf{x}_i, \hat{\mathbf{x}}_{-i}\right), \tag{5}$$

where:

$$\widetilde{\mathcal{U}}_i\left(\mathbf{x}_i, \hat{\mathbf{x}}_{-i}\right) = \mathcal{U}_i\left(\mathbf{x}_i, \hat{\mathbf{x}}_{-i}\right) + v_i(\mathbf{x}_i). \tag{6}$$

Where $v_i : int(\Delta_i) \mapsto \mathbb{R}$ is a deterministic perturbation penalty function. This function has the property that it is smooth and positive definite such that $\widetilde{\mathcal{U}}$ is concave, with a unique global maximum. The most common example of such a penalty function is the Gibbs Entropy:

$$v_i(\mathbf{x_i}) = \varepsilon \sum_j x_{ij} \log x_{ij}. \tag{7}$$

In this case, the best response function that results is:

$$\left[\widetilde{BR}_i(\hat{\mathbf{x}})\right]_j = \frac{\exp\left(\frac{1}{\varepsilon}\mathcal{U}\left(\mathbf{e}_{ij}, \hat{\mathbf{x}}_{-i}\right)\right)}{\sum_{j'}\exp\left(\frac{1}{\varepsilon}\mathcal{U}\left(\mathbf{e}_{ij'}, \hat{\mathbf{x}}_{-i}\right)\right)}, \tag{8}$$

which is the well known logit function. The main purpose of this perturbation is to turn the set valued map $BR(\hat{\mathbf{x}})$ into an argmax of a concave function, which is much more tractable. It should be noted that another common formulation of sFP is via a stochastic perturbation. For example, instead of a deterministic perturbation to the utility function, we define:

$$\widetilde{\mathcal{U}}_i\left(\mathbf{e}_{ij}, \hat{\mathbf{x}}_{-i}\right) = \mathcal{U}_i\left(\mathbf{e}_{ij}, \hat{\mathbf{x}}_{-i}\right) + \varepsilon_{ij}. \tag{9}$$

Then, our perturbed best response becomes:

$$\left[\widetilde{BR}_i(\hat{\mathbf{x}})\right]_j = \Pr\left(j = \arg\max_j \mathcal{U}_i(\mathbf{e}_{ij}, \hat{\mathbf{x}}_- i)\right) \tag{10}$$

Though this formulation seems quite different from that of (6), [8] showed that the perturbed best response of any stochastic perturbation ε_{ij}, can be written in the form (6). It should be noted, however, that the converse does not hold. In fact, in a discrete action scenario, if the number of actions $n > 4$, it can be shows that there exists no stochastic perturbation which is best-response-equivalent to a Gibbs Entropy deterministic perturbation (Proposition 2.2 of [8]).

Now that we have defined the perturbed best response function, we can define sFP.

Definition 2. *Smooth or Stochastic Fictitious Play is the scheme in which player* P_i *plays a mixed strategy* $\mathbf{x}_i^k$ *in the* k^{th} *round based on the best response to the opponent empirical frequencies* $\hat{\mathbf{x}}_{-i}^k$*, where:*

$$\mathbf{x}_i^k = \widetilde{BR}(\hat{\mathbf{x}}_{-i}^k) \quad k = 1, 2, \ldots, \tag{11}$$

and the full empirical frequency of all the players $\hat{\mathbf{x}}^k = (\hat{\mathbf{x}}_i^k, \hat{\mathbf{x}}_{-i}^k)$ *is updated based on the actions of the players in the* k^{th} *round* $\mathbf{e}^k$ *via:*

$$\hat{\mathbf{x}}_i^{k+1} = \frac{k}{k+1}\hat{\mathbf{x}}_i^k + \frac{1}{k+1}\mathbf{e}_i^k \ \ \forall i, \tag{12}$$

and $\hat{\mathbf{x}}_i^1$ *may be chosen arbitrarily (i.e. the first play is random).*

A. Notion of Convergence

Since in sFP, agents play mixed strategies, there is a natural way to define convergence to a pure or mixed strategy. However, in sFP agents do not play the original game, but rather a perturbed version $\tilde{\mathcal{G}}$. We say sFP converges to $\tilde{\mathbf{x}}^* = (\tilde{\mathbf{x}}_i^*, \tilde{\mathbf{x}}_{-i}^*) \in NE(\tilde{\mathcal{G}})$ if and only if:

$$\lim_{k\to\infty} \hat{\mathbf{x}}^k = \tilde{\mathbf{x}}^*. \tag{13}$$

A significant advantage of sFP is that, under mild conditions, if the above type of convergence is satisfied, sFP also converges to NE in behaviour.

B. Example

In this section, we will see how perturbing an agent's utility function $\mathcal{U}_i$ with a Gibbs Entropy term impacts the best response map. Consider a two player matrix game with payoffs:

$$\mathbf{U}_1 = \begin{bmatrix} -10 & 2 \\ 1 & -1 \end{bmatrix}, \mathbf{U}_2 = \begin{bmatrix} 5 & -2 \\ -1 & 1 \end{bmatrix}, \tag{14}$$

for players P_1 and P_2. For such a simple game, we can directly show that there exists only one NE:

$$(\mathbf{x}^*, \mathbf{y}^*) = \left(\frac{1}{9}\begin{bmatrix}2\\7\end{bmatrix}, \frac{1}{14}\begin{bmatrix}3\\11\end{bmatrix}\right) \approx \left(\begin{bmatrix}0.222\\0.778\end{bmatrix}, \frac{1}{14}\begin{bmatrix}0.214\\0.786\end{bmatrix}\right) \tag{15}$$

As can be surmised from Figure 1, the perturbed best response functions intersect at a different point from the unperturbed ones, and thus admit a slightly different equilibrium, which we can refer to as the perturbed equilibrium:

$$(\tilde{\mathbf{x}}^*, \tilde{\mathbf{y}}^*) \approx \left(\begin{bmatrix}0.1674\\0.8326\end{bmatrix}, \begin{bmatrix}0.2716\\0.7284\end{bmatrix}\right) \tag{16}$$

.

As the temperature factor $\epsilon \to 0$, the perturbed NE $(\tilde{\mathbf{x}}^*, \tilde{\mathbf{y}}^*) \to (\mathbf{x}^*, \mathbf{y}^*)$, and the sFP algorithm itself converges to FP. This happens because as $\varepsilon \to 0$, the perturbation term goes to zero. This is consistent with our earlier definition of the perturbed best response, because if we examine the limiting behaviour of (8) we have:

$$\left[\widetilde{BR}_i(\hat{\mathbf{x}})\right]_j = \frac{\exp\left(\frac{1}{\varepsilon}\mathcal{U}\left(\mathbf{e}_{ij}, \hat{\mathbf{x}}_{-i}\right)\right)}{\sum_{j'}\exp\left(\frac{1}{\varepsilon}\mathcal{U}\left(\mathbf{e}_{ij'}, \hat{\mathbf{x}}_{-i}\right)\right)} \to \mathbb{I}\left(j = \arg\max_j \mathcal{U}_i(\mathbf{e}_{ij}, \hat{\mathbf{x}}_{-}i)\right), \tag{17}$$

which is exactly the unperturbed best response.

This example, though simple, already shows why sFP is in some senses easier to analyze than FP. Since we are considering the perturbed game, best response maps become single-valued (i.e. functions). This is a significant advantage, because we can consider the dynamics of the mean differential equations of sFP to derive convergence results. Indeed, this is how most convergence results for sFP have been derived. In the case of FP. these differential equations become differential inclusions, and are much harder to analyze. Nevertheless, many of the classic results of FP have been re-derived via a stochastic approximation analysis of these differential inclusions.

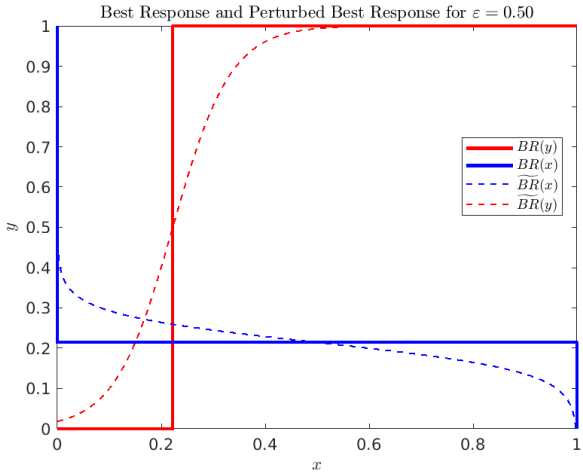

Fig. 1. Comparison of the many-valued best response map and the perturbed best response function. The perturbed curves intersect at a point which is distinct from the unperturbed case. As the temperature parameter $\varepsilon \to 0$, the perturbed curve approaches the unperturbed one.

IV. Repeated Two-Player Zero-Sum Games

Repeated two-player zero-sum games are an important class of games for which convergence results exist for both FP and sFP. Though the algorithms are related, as we will see, the proofs themselves are completely different, and separated by half a century.

A. Convergence of FP

Finite-action two-player zero-sum games can be played in a decoupled manner via FP. These types of games are among the simplest for analysis, and are a natural place to begin our study. In this section, we will examine FP and its convergence in two-player zero sum games. Specifically, we are interested in whether or not the game will converge to a NE solution. Our first study will begin by looking at the very first proof of convergence of FP, presented by Julia Robinson in 1950.

1) Robinson's Proof of Convergence: Robinson's proof of convergence relies on four key lemmas, and establishes a convergence in value. We will state her fundamental result.

Theorem 1. *Let $\hat{\mathbf{x}}^k$ and $\hat{\mathbf{y}}^k$ be the empirical frequency of actions for the two agents based on fictitious play. We have that:*

$$\lim_{n\to\infty} \min(\hat{\mathbf{x}}^k)^T\mathbf{U} = \lim_{k\to\infty} \max \mathbf{U}\hat{\mathbf{y}}^k = v, \tag{18}$$

where v is the value of the game defined to be:

$$\min_j \sum_i u_{ij}x_i^* = \max_i \sum_j u_{ij}y_j^*, \tag{19}$$

with some $(\mathbf{x}^*, \mathbf{y}^*) \in \Delta$.

It should be noted that this theorem establishes two important things. Firstly, the payoff in FP does converge in a two-player zero-sum game, and secondly, it converges to a min-max solution of the game in value.

Theorem 2. *If agents use FP in a two-player zero-sum game, the empirical frequency of actions $\hat{\mathbf{x}}^k$ converges to a NE, i.e*

$$\lim_{k\to\infty} \hat{\mathbf{x}}^k = \mathbf{x}^*. \tag{20}$$

Proof. By Theorem 1, we know that:

$$\lim_{k\to\infty} \max \mathbf{U}\hat{\mathbf{y}}^k = \lim_{k\to\infty} \min(\hat{\mathbf{x}}^k)^T\mathbf{U} \tag{21}$$

Additionally, we must have

$$\max \mathbf{U}\hat{\mathbf{y}}^k \geq (\hat{\mathbf{x}})^k\mathbf{U}\hat{\mathbf{y}}^k \tag{22}$$

which must also hold under the limit as $k \to \infty$. If we subtract the limit of (22) from (21), and note that $\lim_{k\to\infty} \min(\hat{\mathbf{x}}^k)^T\mathbf{U} \leq \lim_{k\to\infty}(\hat{\mathbf{x}}^k)^T\mathbf{U}\mathbf{y}' \ \ \forall \mathbf{y}' \in \Delta_y$:

$$\lim_{k\to\infty} (\hat{\mathbf{x}})^k\mathbf{U}\hat{\mathbf{y}}^k \leq \lim_{k\to\infty} (\hat{\mathbf{x}}^k)^T\mathbf{U}\mathbf{y}' \ \ \forall \mathbf{y}' \in \Delta_y. \tag{23}$$

Similarly, we can obtain:

$$\lim_{k\to\infty} (\hat{\mathbf{x}})^k\mathbf{U}\hat{\mathbf{y}}^k \geq \lim_{k\to\infty} (\mathbf{x}')^T\mathbf{U}\hat{\mathbf{y}}^k \ \ \forall \mathbf{x}' \in \Delta_x. \tag{24}$$

Thus we can conclude, that in two-player zero-sum games, the empirical frequency estimates converge to a NE. □

A new common form of analysis which is routed in stochastic approximation looks at the *mean dynamics* of FP which results in a differential inclusion:

$$\dot{\hat{\mathbf{x}}}_i \in BR(\hat{\mathbf{x}}_i) - \hat{\mathbf{x}}_i. \tag{25}$$

From this, we can also see that the rest points of FP must be NEs.

B. Convergence of sFP

The convergence of sFP in two-player zero sum games was established in [8]. The proof relies on several key theorems. We will summarize the results after a few preliminary definitions. We define individual user payoff vector and the noise vector as:

$$\boldsymbol{\pi}^{(m)} = [\pi_1^{(m)}, \ldots, \pi_n^m]^T = [\mathcal{U}_m(\mathbf{e}_{mn}, \hat{\mathbf{x}}_{-i}), \ldots, \mathcal{U}_m(\mathbf{e}_{m1}, \hat{\mathbf{x}}_{-i})]^T, \tag{26}$$

$$\boldsymbol{\varepsilon}^{(m)} = [\varepsilon_1^{(m)}, \ldots, \varepsilon_n^{(m)}]^T. \tag{27}$$

In the stochastic formulation of sFP, we define the *choice probability function*:

$$C_i^{(m)}(\boldsymbol{\pi}) = \Pr\left(\arg\max_j \pi_j^{(m)} + \varepsilon_j^{(m)} = i\right). \tag{28}$$

C_i can be interpreted as the probability of agent m choosing action i. in this case, $m = 1, 2$. Theorem 2.1 of [8] establishes that any stochastic perturbation in which $f_{\boldsymbol{\varepsilon}}(\boldsymbol{\epsilon})$ has positive density on $\mathbb{R}^n$ and results in a continuously differentiable C can be replaced with a deterministic perturbation, which allows the choice function to be written deterministically as:

$$C^{(m)}(\boldsymbol{\pi}^{(m)}) = \underset{\mathbf{x}\in int(\Delta^{(m)})}{\arg\max} \ (\mathbf{x}^T\boldsymbol{\pi}^{(m)} - V(\mathbf{x})). \tag{29}$$

Proposition 3.1 of [8] states that if for each agent, the distribution of the noise vector ε converges to a point mass (in distribution) as $k \to \infty$ then a rest point of the perterbed best-response dynamics converges to the NE of the game.

The final step examines the best-response dynamic:

$$\dot{\hat{\mathbf{x}}}_{i'} = \widetilde{BR}(\hat{\mathbf{x}}_{i'}) - \hat{\mathbf{x}}_{i'}. \tag{30}$$

We can use Theorem 2.1 to show that the best response dynamic can be written deteministically. Finally it can be shown that for two-player zero-sum games, a Lyaponov function exists for this dynamic, which establishes convergence. See [8] for details on the particular Lyaponov function.

C. Simulation

We simulated the repeated play of a two-player zero-sum game, where the utility of P_1 is given by:

$$\mathbf{U}_1 = \begin{bmatrix} -1 & 3 \\ 3 & -2 \end{bmatrix} = -\mathbf{U}_2, \tag{31}$$

which only has one NE:

$$(\mathbf{x}^*, \mathbf{y}^*) = \left(\frac{1}{9}\begin{bmatrix}0.2\\0.8\end{bmatrix}, \frac{1}{14}\begin{bmatrix}0.2\\0.8\end{bmatrix}\right). \tag{32}$$

Our main goal is to numerically investigate the convergence properties of FP and sFP. We use the standard FP algorithm, and a fixed Gibbs Entropy deterministic perturbation for sFP. Since we are using a fixed perturbation, we only expect convergence to an NE of the perturbed game. In Figure 2, we can see that as expected based on theory, both FP and sFP ($\varepsilon = 0.05$) converge to the NE of the game and the perturbed game respectively. Notably sFP converges more rapidly, and with significantly fewer sharp oscillations. Though sFP does converge more quickly, it should be noted that it converges to a perturbed equilibrium, which differs slightly from the unperturbed equilibrium.

In Figure 3, the importance of the temperature factor is further investigated. Specifically, we see that the first element of the perturbed NE $\tilde{\mathbf{x}}^*$ is increasing with the temperature parameter. As the temperature parameter goes to zero, we see that the behaviour of sFP begins to look very similar to that of FP, which is also in line with our expectations.

Though the empirical frequency of play has converged to that of the NE, the actual dynamics of the payoff under FP and sFP are quite different from the dynamics of playing the mixed strategy NE. Figure 4 shows that in all cases, the global average of all forms of play do converge to the value of the game: -2.2, as expected based on theory. If we look at the local properties of the payoff, however, we see that they vary significantly. In particular, the local average payoff of FP can oscillate wildly. This is because in FP, an agent

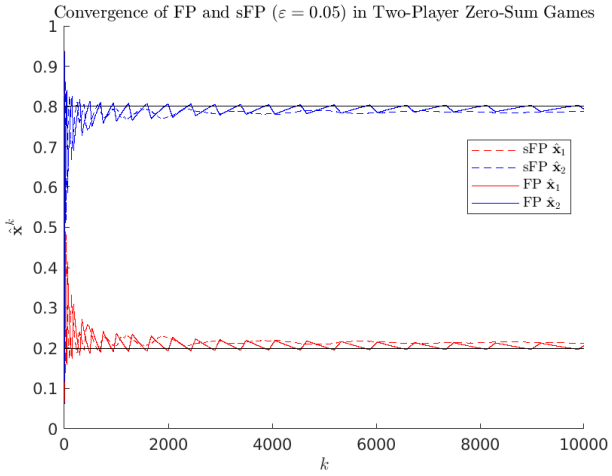

Fig. 2. Convergence of a 2×2 two-player zero-sum game. FP converges to the NE but has sharp oscillations. sFP converges more rapidly, and in a much smoother fashion, but converges to the perturbed NE which is close to the NE for the chosen temperature parameter.

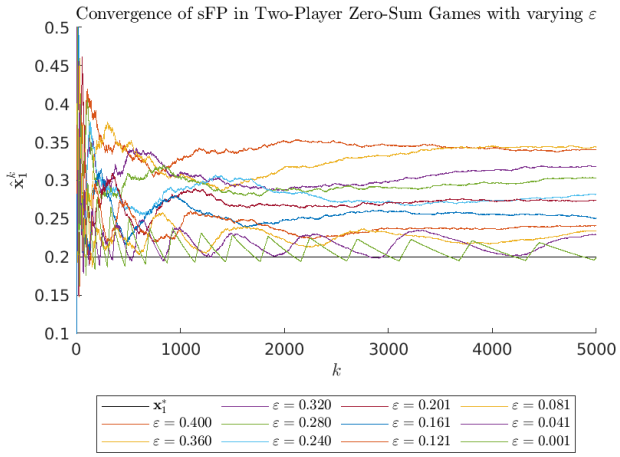

Fig. 3. Convergence of a two-player zero-sum under sFP to various perturbed equilibria. Note that as the temperature parameter goes to zero, the equilibrium approaches the NE of the unperturbed game, but the empirical frequency starts to oscillate more rapidly.

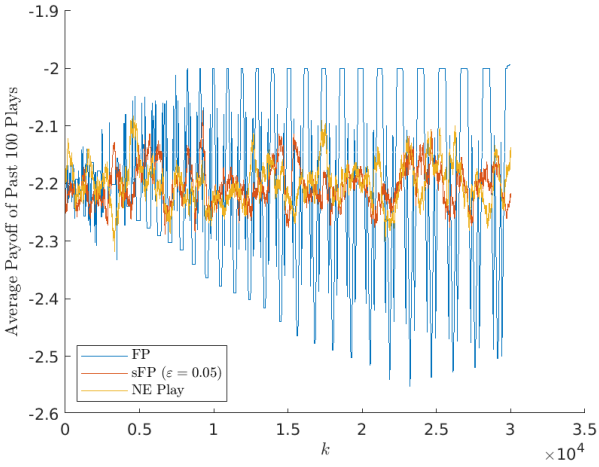

Fig. 4. The figure depicts the average payoff from the 100 closest games for the game (31) played via FP, sFP, and via the true mixed NE strategy. Note that average payoff in FP oscillates wildly, while the sFP more closely emulates the play of the mixed NE.

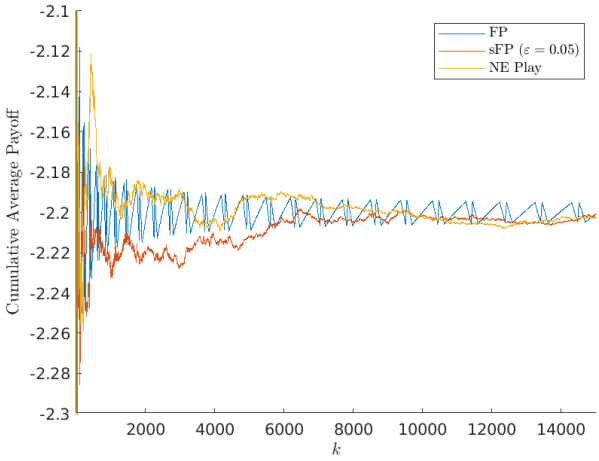

Fig. 5. The figure depicts the average cumulative payoff after k games for the game (31) played via FP, sFP, and via the true mixed NE strategy. Note that all three converge to the value of the mixed NE at -2.2.

can play the same pure strategy may times consecutively if the empirical average indicates that it should. Eventually, the empirical average adjusts, and the player changes its action. In sFP, even if the empirical average is biased slightly away from the NE, the agent still plays a mixed strategy which is close to that of the perturbed NE. This is why we find that the local average payoff of sFP much more closely resembles the behaviour of the NE play.

Figure 4 is an excellent depiction of one of the reasons why sFP was first formulated. One of the key motivations of [6], in which sFP was first proposed, is that the convergence of an empirical average to a mixed strategy, the predominant notion of convergence of FP, is flawed, because the agent makes deterministic plays in every round. Small changes in the empirical average can have sharp (discontinuous) changes in the chosen action. In sFP, small oscillations in the empirical average smoothly change the mixed strategy played by sFP, and since in sFP the agent plays a mixed strategy convergence to a mixed strategy can be naturally defined.

In this case, of the game we have considered in this section, though the payoff locally oscillates, the cumulative average payoff still converges to the value of the game, as depicted in Figure 5. As we will see, this is not something that can be taken for grated. In general bimatrix games there are examples where where FP can converge to an NE belief while the payoff is much worse than if the NE was played.

V. Repeated Bimatrix Games

Repeated bimatrix games are a more general class of games, and we can find examples of games where FP converges, and where it doesn't. In this section, we will review some of these results.

A. Convergence of FP in Repeated Bimatrix Games

Convergence of FP has been established for several variants of bimatrix games. Following Robinson's proof of convergence for two-player zero-sum games, Miyasawa [3] proved that in any 2×2 game, the empirical frequency converges to an NE. Monderer and Shapley [4] also proved the same for any bimatrix game of identical interests. Interestingly, it can be shows that Monderer's result and Robinson's result imply Miyasawa's result, since any nondegenerate 2×2 game can be shown to be "best response equivalent" to either a zero-sum game, or a game of identical interest.

It is important to note that in general, even if the empirical frequencies converge to the NE, the actual play may look quite different from play using an NE strategy. We will see an example in which the average payoff of FP is significantly lower than with NE play. Furthermore, we will see examples of 3×3 bimatrix games where FP fails to converge at all for certain initial conditions [9].

B. Convergence of sFP in Repeated Bimatrix Games

Convergence of sFP has been established for 2×2 games through [6], which first showed that it converges almost surely in a game with a unique mixed NE. This work was generalized by Kaniovski [10], and by Benaim [11], to any 2×2 game with countably many NEs. These works established convergence not only in the empirical frequency of actions, but also in behaviour. Essentially, what this means is that sFP converges to an NE such that one cannot distinguish between the play of the NE directly, and the sFP. Their proofs employed elements of stochastic approximation, but do not explicitly employ the connection between the connection between the mean differential equation of sFP and the convergence of sFP.

Hofbauer and Sandholm [8], by explicitly forming this connection established that sFP also converges in two-player zero-sum games, potential games, and several others.

C. FP vs. sFP in Coordination Game

Consider the following simple 2×2 coordination game, where both player have the identical payoff matrix:

$$\mathbf{U} = \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}. \tag{33}$$

Figure 6 shows 4 convergent paths of FP for distinct initial conditions. This game has three distinct NEs:

$$(\mathbf{x}_1^*, \mathbf{y}_1^*) = \left(\begin{bmatrix} 1 \\ 0 \end{bmatrix}, \begin{bmatrix} 1 \\ 0 \end{bmatrix} \right), \quad (\mathbf{x}_2^*, \mathbf{y}_2^*) = \left(\begin{bmatrix} 0 \\ 1 \end{bmatrix}, \begin{bmatrix} 0 \\ 1 \end{bmatrix} \right)$$
$$(\mathbf{x}_3^*, \mathbf{y}_3^*) = \left(\begin{bmatrix} 0.5 \\ 0.5 \end{bmatrix}, \begin{bmatrix} 0.5 \\ 0.5 \end{bmatrix} \right). \tag{34}$$

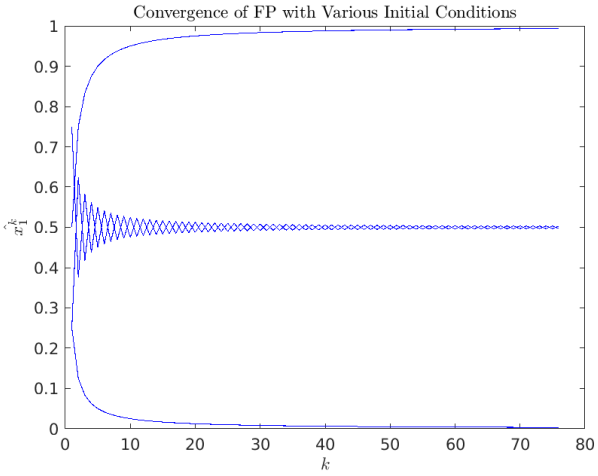

Fig. 6. Sample paths of convergence for the simple coordination game with various initial conditions.

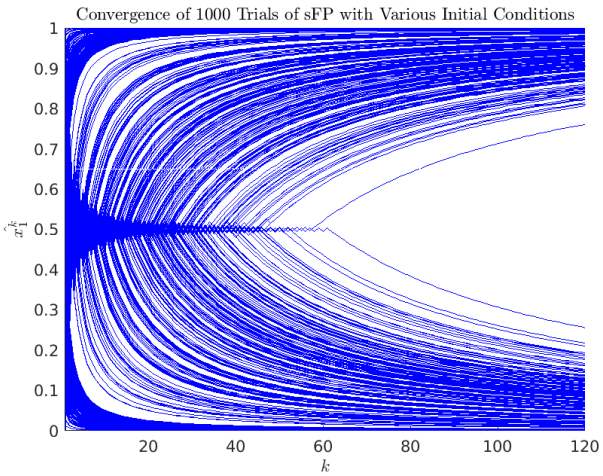

Fig. 7. Sample paths of convergence for the simple coordination game under sFP with various initial conditions. All the paths eventually diverge from the mixed equilibrium in favour of one of the two pure equlibria.

FP is capable of converging to all three, as we can see from the figure. However, for the case in which FP converge to the mixed NE, the payoff in each round is 0, as compared to the average payoff of the NE play which is 0.5. Thus in this coordination game, we can see that FP coordinates to achieve the worst possible payoff.

Our experiments found that in this game, 1000 trials of sFP $\varepsilon = 0.005$ did not converge to the mixed NE. We can see the trials of these runs in Figure 7. This tells us that it is likely that the mixed NE is unstable under sFP.

D. Nonconvergent 3×3 Games

In all the examples we have considered, sFP and FP have converged. 2×2 games have guarantees of convergence, but we only need to extend our consideration to 3×3 games before

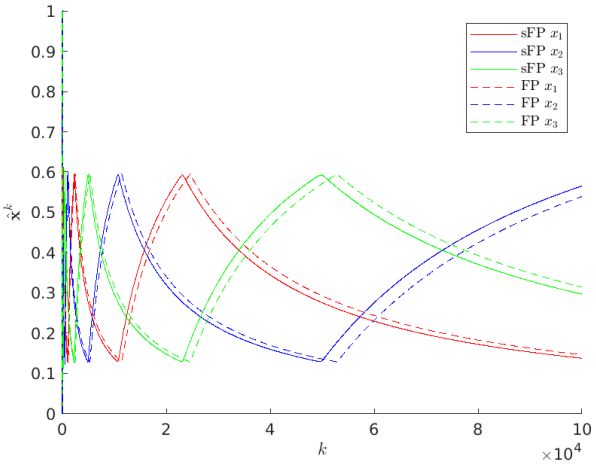

Fig. 8. sFP and FP failing to converge under certain initial conditions. The agents cycle over a subset of the states, each time playing the same action for a longer period of time before switching actions.

we begin find examples where FP and sFP do not converge. Consider the 3×3 game with the following payoffs:

$$\mathbf{U}_1 = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix}, \mathbf{U}_2 = \begin{bmatrix} 0 & 0 & 1 \\ 1 & 0 & 0 \\ 0 & & 0 \end{bmatrix}. \tag{35}$$

Shapley noted [9] that there is clearly an NE where both players chose actions with equal probability. However, under certain initial conditions, FP (and sFP) will fail to converge. One of these initial conditions is $\hat{\mathbf{x}}_1 = [001]^T, \hat{\mathbf{x}}_2 = [010]^T$. The trajectory of FP and sFP starting at these non-convergent conditions are plotted in Figure 8. There are however, still many initial conditions for which FP and sFP can be made to converge to the mixed NE.

1) Perturbation and Convergence in Shapley's Game: As we have just discussed, FP and sFP do not converge in Shapley's game (35) for the the previously specified initial condition. In this section, we attempt to make a contribution by investigating whether or not Shapley's game converges as the temperature parameter increases when we use a deterministic Gibbs Entropy perturbation for sFP. The figures in the appendix show the trajectory of the empirical average strategy on the simplex. From the figures, we can see that for small temperature parameters, sFP enters a limit cycle centred at the NE (see Figure 9). As we increase the temperature parameter, the limit cycle begins to orbit the NE more closely (see Figure 11). Eventually, as the temperature gets low enough, (see Figure 12), sFP slowly spirals inward approaching the NE. Finally, as the temperature parameter gets larger (see Figure 13), the solution rapidly approaches the NE.

These simulations show us that for large enough temperature parameters sFP can converge in (35), even for the troublesome initial condition.

VI. A Strongly Unstable Game

In the class of bimatrix we have only seen examples of games which *can* fail to converge for certain initial conditions. This is a fairly weak result on non-convergence, since it is still possible that FP and sFP may converge for many different initial conditions. Jordan [12] showed that there are repeated games for which all NEs are locally unstable in the strong sense. That is, for any $\epsilon > 0$ and for almost all initial empirical distributions that are within the euclidean ball of radius ϵ, FP does not converge to an NE. The game which he showed to exhibit this property is his three player matching pennies game. In this game each player has an identical binary action set $\{H, T\}$. P_1 tries to match their action to P_2's action, P_2 tried to match their action to P_3's action, and P_3 tries to match their action to the opposite of P_1's action. Later it was also shown [11] that sFP exhibits this same property.

VII. Summary and Concluding Remarks

In this review, we have seen that FP and sFP "converge" in a large class of games, including two player zero-sum games, 2×2 games, games of identical interest, and others. We have noted that convergence can sometimes be defined in a deceptive way. In particular, even when FP converges to a mixed NE in belief, the play of the game may look very different as compared to the play of the mixed NE. We saw how sFP can mitigate these issues by playing a mixed strategy in each round, and how it can converge to become behaviourally like NE play over time. We also saw how a simple cooperation game can exhibit many of the properties of both FP and sFP, and again show how FP can be behaviourally different when converging in belief to a mixed NE.

Finally we considered Shapley's game for which it is possible for both FP and sFP to fall into a limit cycle and fail to converge. We attempted to characterize the convergence of sFP in this case, and examine how the limit cycle can begin to break down with enough perturbation. Interestingly, we find that with enough perturbation, a previously non-convergent initial condition can exhibit rapid convergence. Further investigation will be need to characterize these dynamics.

References

[1] G. Brown, "Iterative solution of games by fictitious play," *Activity Analysis of Production and Allocation*.

[2] J. Robinson, "An Iterative Method of Solving a Game," *Ann. Math.*, vol. 54, no. 2, p. 296, sep 1951.

[3] K. Miyasawa, "On the convergence of the learning process in a 2 x 2 non-zero-sum two-person game," 1961.

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

[5] J. Hofbauer, "Imitation dynamics for games."

[6] D. Fudenberg and D. M. Kreps, "Learning Mixed Equilibria," *Games Econ. Behav.*, vol. 5, no. 3, pp. 320–367, jul 1993.

[7] U. Berger, "Brown's original fictitious play," *J. Econ. Theory*, vol. 135, no. 1, pp. 572–578, jul 2007.

[8] J. Hofbauer and W. H. Sandholm, "On the Global Convergence of Stochastic Fictitious Play," *Econometrica*, vol. 70, no. 6, pp. 2265–2294, nov 2002.

[9] L. S. Shapley, "Some Topics in Two Player Games," *Adv. Game Theory*, 1964.

[10] Y. M. Kaniovski and H. P. Young, "Learning dynamics in games with stochastic perturbations," *Games Econ. Behav.*, vol. 11, no. 2, pp. 330–363, 1995.

[11] M. Benaim and M. W. Hirsch, "Mixed Equilibria and Dynamical Systems Arising from Fictitious Play in Perturbed Games," *Games Econ. Behav.*, vol. 29, no. 1-2, pp. 36–72, oct 1999. [Online]. Available: https://linkinghub.elsevier.com/retrieve/pii/S0899825699907170

[12] J. Jordan, "Three Problems in Learning Mixed-Strategy Nash Equilibria," *Games Econ. Behav.*, vol. 5, no. 3, pp. 368–386, jul 1993. [Online]. Available: https://linkinghub.elsevier.com/retrieve/pii/S0899825683710225

APPENDIX

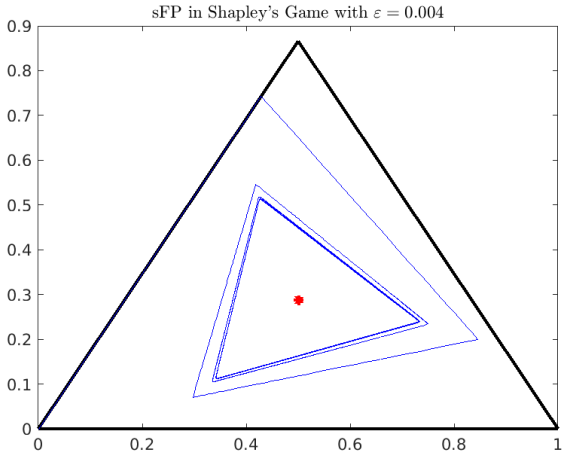

Fig. 9. (35) played via sFP with parameter 0.004.

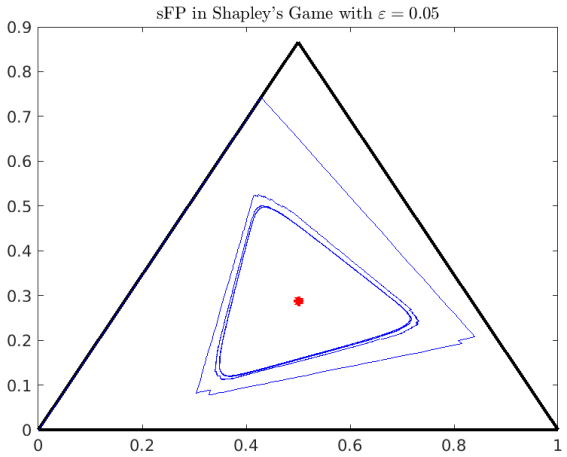

Fig. 10. (35) played via sFP with parameter 0.05.

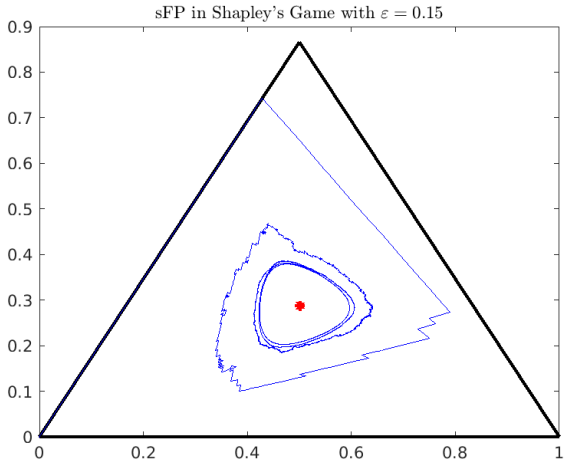

Fig. 11. (35) played via sFP with parameter 0.15.

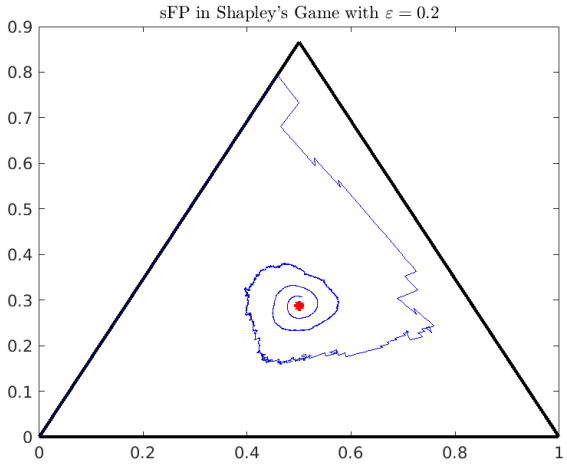

Fig. 12. (35) played via sFP with parameter 0.2.

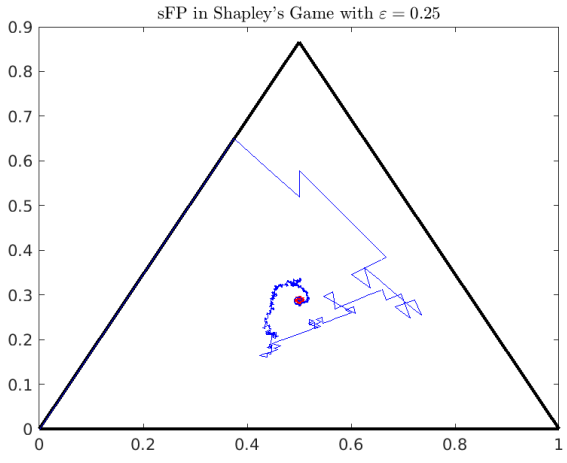

Fig. 13. (35) played via sFP with parameter 0.25.