Nash Equilibrium Solving for General Non-Zero-Sum Games from Non-Convergent Dynamics of Fictitious Play

Zhouming Wu[1,2], Yifen Mu[1], Xiaoguang Yang[3,4], Ge Chen[1,2]

1. Key Lab of Systems and Control, ISS, Academy of Mathematics and Systems Science, CAS
E-mail: wuzhouming@amss.ac.cn, mu@amss.ac.cn (corresponding author), chenge@amss.ac.cn

2. School of Mathematical Sciences, University of Chinese Academy of Sciences

3. Key Lab of Management, Decision and Information System, ISS, Academy of Mathematics and Systems Science, CAS
E-mail: xgyang@iss.ac.cn

4. School of Economies and Management, University of Chinese Academy of Sciences

Abstract: Solving Nash equilibrium (NE) is a fundamental problem in game theory. However, for general non-zero-sum games, computing NE remains challenging (in fact, it is PPAD-hard), and there are no known efficient algorithms so far. Apart from classical combinatorial methods, learning-based approaches offer an alternative for approximating NE. However, for general non-zero-sum games, the dynamics of learning algorithms turns out to be intricate, and often fail to converge. In this paper, we give a novel method to solve NE from the trajectory of fictitious play (FP), a classical learning algorithm in repeated games. By leveraging the equalization principle of NE, we demonstrate that hidden NE can be unveiled via the trajectory data. Specifically, we focus on two scenarios. For the 3×3 game with interior equilibrium when FP exhibits non-convergent behavior, such as cyclic or erratic trajectories, we can obtain an approximate NE by estimating the indifferent lines and solving their intersection points. For the games without an interior equilibrium, by identifying the intersection points between the indifferent lines and the edges of the players' strategy simplexes, and proposing a label-matching principle, we can still find the approximate NE. Hence via this method, we can solve all 3×3 games from finite trajectory date of FP no matter how complicated the dynamics is. Further experiments show that the method can be extended to more general normal form games and special kind of stochastic games. This paper reveals that there is more information hidden in the behavior of learning dynamics, and thus indicating a promising way to utilize trajectory data.

Key Words: NE Solving, Non-Zero-Sum Games, Fictitious Play, Repeated Games, Dynamical Systems, Equalization Principle

1 Introduction

As stated in [1], *"The Nash equilibrium, defined and shown universal by John F. Nash in 1950[2], is paramount in game theory."* Therefore solving Nash equilibrium (NE) is a fundamental problem, and there have been a lot of attempts from different perspectives in the history.

Early methods include Lemke-Howson algorithm [3], Sperner's lemma [4], support enumeration or vertex enumeration [5], and some topological methods [6, 7]. It turns out that these methods are useful for small-scale games, while the complexity would increase exponentially as games become larger and larger, since they could not avoid enumeration and searching in the whole space. In fact, later research on computational complexity reveals that such difficulty is essential and intractable. Solving NE is generally PPAD-hard [8] for normal form game [9–12] and stochastic game [13] .

On the other hand, almost from the same time with the birth of NE, researchers try to approximate it via learning in repeated games [14, 15]. Much progress has been made in this area.

One of the most classical learning algorithms is fictitious play(FP) [16]. It have been proved that FP can converge to NE in zero-sum games [17], 2×2 games [18], potential games [19], $2 \times n$ games [20] and other games with special properties, such as games solvable by iterated strict dominance [21], non-degenerate quasi-supermodular games with diminishing returns or of dimension $3\times n$ or 4×4 [22]. Here $2 \times n$ means one player has only two actions available while the other could have any number of choices. The other famous and popular learning framework is based on the idea of regret minimization [23–27]. It has been proved that these algorithms could converge to NE in zero-sum games and CE(correlated equilibrium) or CCE(coarse correlated equilibrium) in non-zero-sum games [28].

Recently with the rapid development of artificial intelligence in the past decade, learning in games has received much more attention, often appearing in MARL. For example, many variants of FP were proposed to adapt to more complex situations by combining with neural network [29] and Q-learning [30]. Regret-based algorithms are extended to CFR(counterfactual regret minimization) in order to deal with extensive form games [31], and successfully solve the Hold'em Poker game etc. [32, 33].

Unfortunately, learning algorithms still suffer from a lot of disadvantages. The convergence relies strongly on the special structure of games. For generic structures, the algorithm would only converge to some broader set, like the set of CE or CCE, making it hard to use in practice.

In fact, deeper research indicates that convergence to NE is even impossible in many scenarios [34]. As early as 1964, Shapley proposed the first counterexample of 3×3 game in which FP has cyclic behavior and never converges no matter what the initial point is [35]. Until recently, Shapley game is extended to a group of games by incorporating one parameter. As the parameter changes, bifurcation would emerge, and for some parameter, more complicated pattern and even

This work was supported by the National Key Research and Development Program of China under grant No.2022YFA1004600, the Strategic Priority Research Program of Chinese Academy of Sciences under Grant No. XDA27030201, the Natural Science Foundation of China under Grant T2293770.

erratic behavior appears [36, 37]. Similar results are also found in 4×4 game [39]. Latest research finds that Poincaré recurrence happens in some learning dynamics, which prevents the convergence to NE [40]. And more passively, [1] proves an impossiblility theorem that there exists such game that all learning algorithms cannot converge to NE.

As a result, now questions arise naturally: Can we acquire reasonable desicion suggestions when the learning algorithm yields non-convergent result? Is there any information hidden beneath the dynamics which can be utilized? We aim to answer these questions in this paper.

The key idea is to investigate the geometrical features of the trajectory of learning dynamics and combine it with the basic property of NE. We will focus on 3×3 games and FP dynamics in this paper, although the idea can be extended to more general cases.

To this end, we highlight the concepts of switching point and indifferent line. On the FP trajectory, players incessantly change their actions. When one player recognizes his opponent changing action from one to another, he could record his current mixed strategy as a switching point. By collecting sufficient data of switching points, we can approximate the indifferent lines where the opponent's two actions gain the same payoff. Then the classical equalization principle of NE indicates that the intersection point of these indifferent lines is just the NE of 3×3 games with interior equilibrium including Shapley game.

As for the 3×3 games without interior equilibrium, the indifferent lines could still help us to find NE with partial support. In fact, with the absence of interior equilibrium, the intersection points between indifferent lines and the edges of strategy simplex are all candidates of NE. For this case, we propose a labeling rule to these points and a matching principle to check, by which we can still solve these 3×3 games and avoid extra computation or enumeration.

This paper is organized as follows. Section 2 gives the problem formulation and some necessary preliminary knowledge about the non-convergent dynamics. Section 3 presents our method and proves its feasibility. Section 4 draws the conclusion and gives further discussion.

2 Problem Formulation and Preliminaries

2.1 Problem Formulation

Consider 2-player $n \times n$ normal form game. Player A and Player B have actions $i = 1, \cdots, n$ and $j = 1, \cdots, n$ respectively. When Player A chooses his i-th action and Player B chooses his j-th action, the payoff to each player is $u_A(i,j) = a_{ij}$ and $u_B(i,j) = b_{ij}$, forming corresponding payoff matrices $A = (a_{ij})$ and $B = (b_{ij})$.

A mixed strategy for Player A is denoted by $\mathbf{x} = (x_1, \cdots, x_n)^T, \sum_{i=1}^n x_i = 1, x_i \geq 0$, and Player B's mixed strategy is denoted by $\mathbf{y}$. Denote the sets of mixed strategies as two n-dimentional simplexes $\Delta_A = \Delta_B = \{\mathbf{x} \in \mathbb{R}^n | \sum_{i=1} x_i = 1, x_i \geq 0\}$, and let $\Delta := \Delta_A \times \Delta_B$ be their product. In particular, the i-action of Player $k = A, B$ can be represented by the vertex of simplex Δ_k, i.e. $e_i^k = (0, \cdots, 0, 1, 0, \cdots, 0)$, only the i-th component of e_i^k is 1.

When Player A uses a mixed strategy $\mathbf{x}$ and Player B uses $\mathbf{y}$, the expected payoffs are $u_A(\mathbf{x}, \mathbf{y}) = \mathbf{x}^T A\mathbf{y}$ and $u_B(\mathbf{x}, \mathbf{y}) = \mathbf{x}^T B\mathbf{y}$. Specifically, the expected payoff for Player A using action e_i^A against Player B's strategy $\mathbf{y}$ is $(A\mathbf{y})_i$, and these values constitute a n-dimensional payoff vector $A\mathbf{y}$. Analogously, $x^T B$ is Player B's payoff vector against $\mathbf{x}$.

Given the mixed strategy of the opponent, one can choose those strategies with the highest payoff, which can form a set called best response set. Formally,

Definition 2.1. *Given a mixed strategy profile $(\mathbf{x}, \mathbf{y})$, the best response of one player against the opponent is*

$$\mathrm{BR}_A(\mathbf{y}) := \{\mathbf{x} \in \Delta_A \mid \mathbf{x}^T A\mathbf{y} \geq (\mathbf{x}')^T A\mathbf{y},\ \forall \mathbf{x}' \neq \mathbf{x}\},$$
$$\mathrm{BR}_B(\mathbf{x}) := \{\mathbf{y} \in \Delta_A \mid \mathbf{x}^T B\mathbf{y} \geq \mathbf{x}^T A\mathbf{y}',\ \forall \mathbf{y}' \neq \mathbf{y}\}.$$

When the best response set is not a singleton, some literature would set tie-breaking rules in prior to determine one specific pure action, espicially in learning theory. For example, players would always choose the action with the lowest or highest index. In our setting we incorporate but not specify a tie-breaking rule, then both players can determine a certain action to choose when there are several alternatives.

Definition 2.2. *A mixed strategy profile $\sigma = (\bar{\mathbf{x}}, \bar{\mathbf{y}})$ is called a Nash equilibrium in normal form game (A, B), if*

$$\bar{\mathbf{x}}^T A\bar{\mathbf{y}} \geq (\mathbf{x}')^T A\bar{\mathbf{y}}, \quad \forall \mathbf{x}' \in \Delta_A, \mathbf{x}' \neq \bar{\mathbf{x}};$$
$$\bar{\mathbf{x}}^T B\bar{\mathbf{y}} \geq \bar{\mathbf{x}}^T B\mathbf{y}', \quad \forall \mathbf{y}' \in \Delta_B, \mathbf{y}' \neq \bar{\mathbf{y}}.$$

And $\sigma = (\tilde{\mathbf{x}}, \tilde{\mathbf{y}})$ is called a ε-Nash equilibrium, if

$$\tilde{\mathbf{x}}^T A\tilde{\mathbf{y}} \geq (\mathbf{x}')^T A\tilde{\mathbf{y}} - \varepsilon, \quad \forall \mathbf{x}' \in \Delta_A, \mathbf{x}' \neq \tilde{\mathbf{x}};$$
$$\tilde{\mathbf{x}}^T B\tilde{\mathbf{y}} \geq \tilde{\mathbf{x}}^T B\mathbf{y}' - \varepsilon, \quad \forall \mathbf{y}' \in \Delta_B, \mathbf{y}' \neq \tilde{\mathbf{y}}.$$

Generally speaking, finding NE is complicated[9]. To capture the property of NE, we need to define the support of a strategy.

Definition 2.3. *For some player's mixed strategy $\mathbf{x}$, its support is defined as the set of all pure actions which the player would choose with a non-zero probability, i.e.*

$$\mathrm{supp}(\mathbf{x}) = \{i \mid x_i > 0, \textit{ for } \mathbf{x} = (x_1, \cdots, x_n)\}.$$

And we say the support of a strategy profile $(\mathbf{x}, \mathbf{y})$ has size $l \times m$, if the support of Player A's mixed strategy $\mathbf{x}$ contains l actions, and the support of $\mathbf{y}$ contains m actions. Based on this concept, we have the following equalization principle.

Lemma 2.1. *(Equalization Principle[41], Theorem 7.1 [42]) Given a normal form game (A, B), the mixed strategy profile $(\mathbf{x}, \mathbf{y})$ is a Nash equilibrium iff for any player $k = A, B$, given the opponent's mixed strategy $\mathbf{y}$, his own strategy $\mathbf{x}$ satisfies*

1) $u_k(e_j^k, \mathbf{y})$ is the same for all $e_j^k \in \mathrm{supp}(\mathbf{x})$, and

2) $u_k(e_j^k, \mathbf{y}) \geq u_k(e_{j'}^k, \mathbf{y})$, where $e_j^k \in \mathrm{supp}(\mathbf{x})$ and $e_{j'}^k \notin \mathrm{supp}(\mathbf{x})$.

Now we can consider the repeated game in which both players adopt fictitious play (FP) as their learning dynamics. At time $t = 1, 2, \cdots$, Player A chooses action $a(t) = 1, \cdots, n$ and Player B chooses action $b(t) = 1, \cdots, n$. With slight abuse of notation, the empirical distribution $\mathbf{x}(T)$ of Player A is defined as

$$x_i(T) = \frac{\#\{a(t) \mid a(t) = i,\ t = 1, \cdots, T\}}{T}.$$

The empirical distribution $\mathbf{y}(T)$ of Player B can be defined similarly.

Based on the opponent's empirical distribution, each player will take the best response against it, i.e.

$$a(t+1) = \mathrm{BR}_A\left(\mathbf{y}(t)\right), \ b(t+1) = \mathrm{BR}_B\left(\mathbf{x}(t)\right). \quad (1)$$

Given the initial point $(\mathbf{x}(0), \mathbf{y}(0)) \in \Delta$, the whole dynamics can be represented by the following recursive equations

$$\begin{cases} \mathbf{x}(t+1) = \frac{1}{t+1}\Big(\mathrm{BR}_A\left(\mathbf{y}(t)\right) + t \cdot \mathbf{x}(t)\Big), \\ \mathbf{y}(t+1) = \frac{1}{t+1}\Big(\mathrm{BR}_B\left(\mathbf{x}(t)\right) + t \cdot \mathbf{y}(t)\Big). \end{cases} \quad (2)$$

In the dynamical system (2), $(\mathbf{x}(t), \mathbf{y}(t))$ is the state of system and evolves in the phrase space $\Delta = \Delta_A \times \Delta_B$. The sequence $\{(\mathbf{x}(t), \mathbf{y}(t)) \mid t = 1, 2, \cdots\}$ forms the trajectory of FP.

Remark 1. *Some literature also consider the continuous-time version of FP (CFP) in which time t belongs to $[0, \infty)$ [20, 36, 37], then the evolution of $(\mathbf{x}(t), \mathbf{y}(t))$ is driven by ordinary differential equations. Although CFP "often produces a clearer analytical picture which can subsequently serve as a scaffolding for the discrete-time analysis" [38], this paper concentrates on the original discrete-time FP, since our goal is to find NE of a game by analysing data on the trajectory of a learning algorithm, which can only be implemented by FP rather than CFP.*

Given the dynamical system (2) and initial point $(\mathbf{x}(0), \mathbf{y}(0))$, FP is said to *converge*, if there is some strategy profile $(\mathbf{x}^*, \mathbf{y}^*) \in \Delta$ such that the empirical distribution $\big(\mathbf{x}(t), \mathbf{y}(t)\big) \to (\mathbf{x}^*, \mathbf{y}^*)$ as $t \to \infty$. In the literature, it is a central question to distinguish the classes of games in which FP would converge, however Shapley's counterexample shatters this hope for general non-zero-sum games.

2.2 Phenomena in FP Dynamics: Preliminaries

As stated in Section 1, FP is very likely to not converge. Shapley has constructed a simple 3×3 game [35], whose payoff matrices are

$$A = \begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{pmatrix}, \quad B = \begin{pmatrix} 0 & 1 & 0 \\ 0 & 0 & 1 \\ 1 & 0 & 0 \end{pmatrix}. \quad (3)$$

Game (3) has an unique equilibrium $\big((\frac{1}{3}, \frac{1}{3}, \frac{1}{3}), (\frac{1}{3}, \frac{1}{3}, \frac{1}{3})\big)$. Shapley proved that for this game, under dynamics (2) $(\mathbf{x}(t), \mathbf{y}(t))$ does not converge, and the evolution is shown in Figure 1.

By Figure 1, we can see that only six action profiles would appear in the process. Meanwhile, $\mathbf{x}(t), \mathbf{y}(t)$ and the empirical distribution of action profiles follow a pattern similar to a limit cycle yet with quickly increasing consecutive time for each action or action profile.

Recently, the results of Shapley game are extended to a class of games by introducing a parameter $\beta \in (-1, 1)$ to the payoff matrices as below [36, 37].

$$A = \begin{pmatrix} 1 & 0 & \beta \\ \beta & 1 & 0 \\ 0 & \beta & 1 \end{pmatrix}, \quad B = \begin{pmatrix} -\beta & 1 & 0 \\ 0 & -\beta & 1 \\ 1 & 0 & -\beta \end{pmatrix}. \quad (4)$$

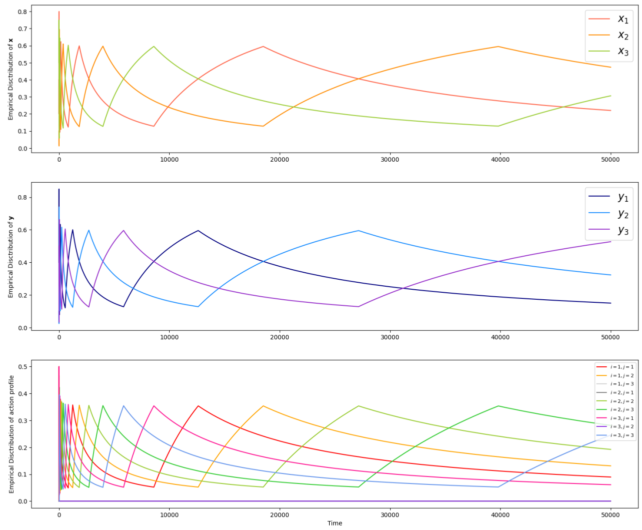

Fig. 1: **Evolution of FP in Shapley game**.

If $\beta = 0$, the game (4) is actually Shapley game. [36, 37] carefully analyze CFP dynamics for games in this class, and find that burification emerges as parameter β changes. For some particular β, even more complicated dynamics would appear. Among all the games, CFP dynamics will not converge to NE except only one specific game. In details,

- For $\beta \in (-1, 0]$, there exists a periodic closed orbit similar as in Shapley game. And from any initial point, the trajectory would tend to the orbit.
- For $\beta \in (0, \sigma)$, the periodic orbit still exists, while from different initial points, the trajectory would tend to either the orbit or NE.
- For $\beta = \sigma$, the orbit no longer exists and the trajectory converges to NE.
- For $\beta \in (\sigma, \tau)$, there also exists a periodic closed orbit, yet its orientation is converse to the orbit in Shapley Game. And the orbit now is of saddle-type, leading to erratic behavior of the system.
- For $\beta \in [\tau, 1)$, the anti-Shapley periodic orbit becomes attractive again.

Here $\sigma \approx 0.618, \tau \approx 0.915$, both are roots of some particular algebraic equations. Figure 2 illustrates CFP for Shapley game where $\beta = 0$.

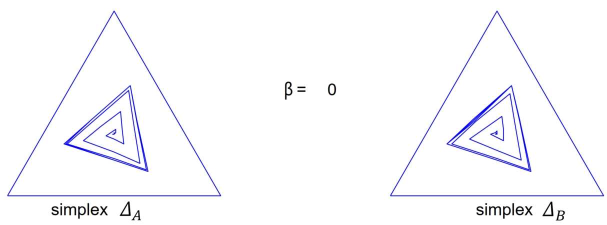

Fig. 2: **Trajectory of CFP in $\Delta_A \times \Delta_B$ of Shapley Game [36]**: The trajectory spirals away from NE lying in the center of each simplex towards a periodic orbit, and hence does not converge.

For more general games, the trajectory of FP would be much more complicated. [43] gives an example, whose payoff matrices are

$$A = \begin{pmatrix} -1.353259 & -1.268538 & 2.572738 \\ 0.162237 & -1.800824 & 1.584291 \\ -0.499026 & -1.544578 & 1.992332 \end{pmatrix}, \tag{5}$$
$$B = \begin{pmatrix} -1.839111 & -2.876997 & -3.366031 \\ -4.801713 & -3.854987 & -3.758662 \\ 6.740060 & 6.590451 & 6.898102 \end{pmatrix}.$$

For this game, the trajectory of CFP is showed in Figure 3. We can see that the state of system changes in a tiny range, while the orbit behaves like a butterfly. We will investigate this system further, and show that even erratic pattern would emerge in this system, see Figure 6.

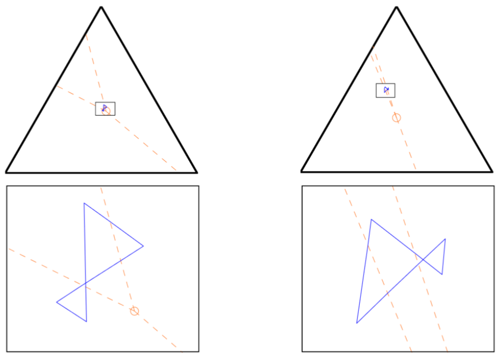

Fig. 3: **Trajectory of CFP in $\Delta_A \times \Delta_B$ of Game (5) [43]**: The two pictures below are partial enlargements of the two pictures above.

3 Methods and Results

As described above, the trajectory of FP dynamics in 3×3 game turns out to be much more complex. So a question arises now: if the FP dynamics i.e. system (2) does not converge or behaves in a complicated pattern such as a limit cycle, how can the players make decisions in the original stage game (A, B)? Is there any extra information in the trajectory?

In this section, we will show that the answer is *YES*. To this end, we need to investigate the trajectory of FP in more details. Besides, we need to incorporate the geometrical property of NE. Then we can suggest a new method and prove its effectiveness to calculate the approximate NE from the trajectory of FP for all 3×3 games.

We first give some necessary concepts for our method. According to system equation (2), the best response plays a central role in FP dynamics. We define the best response polygon as below.

Definition 3.1. *The best response polygon for Player A's i-th action is the set*

$$Z_i^A = \{\mathbf{y} \in \Delta_B \mid (A\mathbf{y})_i \geq (A\mathbf{y})_k, \forall k = 1, \cdots, n\}. \tag{6}$$

And the best response polygon for Player B's j-th action is the set

$$Z_j^B = \{\mathbf{x} \in \Delta_A \mid (\mathbf{x}^T B)_j \geq (\mathbf{x}^T B)_k, \forall k = 1, \cdots, n\}. \tag{7}$$

By Definition 3.1, Z_i^A depends totally on Player A's payoff matrix and includes all the strategies of Player B which make Player A's i-th action to be the best response against them. Since the set Z_i^A is defined by linear inequalities, it is a polygon in mathematics and characterized by the bounded intersection of a series of half spaces in $\mathbb{R}^n$.

We only need to consider the following non-trivial games.

Definition 3.2. *We say a game is non-trivial, if $\forall k = A, B$ and for any action i,*

$$\mu(Z_i^k) > 0.$$

Here μ is Borel measure in Euclidean space.

Otherwise, if for some k and i, $\mu(Z_i^k) = 0$, then either the i-th action of Player k is never the best response against any strategy of his opponent, or the best response polygon Z_i^k is at most a line segment. For the latter case, according to the continuity of payoff function $\mathbf{x}^T A\mathbf{y}$ or $\mathbf{x}^T B\mathbf{y}$, there must be another action i' which is also the best response to opponent's strategy in that line. As a result, the i-th action is weakly dominated. Hence this game is reduced to a lower-dimention game.

For some strategies of the opponent, among those inequalities in (6), if one equality holds, then there are more than one action being the best response. Such strategies form a line, called an indifferent line and defined formally below.

Definition 3.3. *The indifferent lines of Player A and B are defined as*

$$l_{jk}^A := \{\mathbf{y} \in \Delta_B \mid (A\mathbf{y})_j = (A\mathbf{y})_k > (A\mathbf{y})_i, \ \forall i \neq j, k\},$$
$$l_{jk}^B := \{\mathbf{x} \in \Delta_A \mid (\mathbf{x}^T B)_j = (\mathbf{x}^T B)_k > (\mathbf{x}^T B)_i, \ \forall i \neq j, k\}. \tag{8}$$

Figure 4 shows the best response polygons and indifferent lines for Shapley game.

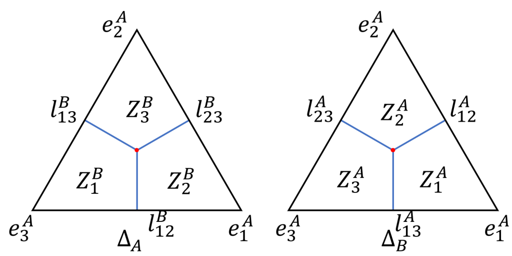

Fig. 4: **Best response polygons and indifferent lines of Shapley game**

Given these concepts, we can look closer to FP from a geometrical standpoint. For instance, if at time t the trajectory of FP $(\mathbf{x}(t), \mathbf{y}(t))$ falls in the interior of some best response polygon profile $Z_j^B \times Z_i^A$, the best response in (2) is (i, j). In this case, by the recursive equation, the trajectory will move closer with step size $\frac{1}{t+1}$ to the vertex profile (e_i^A, e_j^B) along the line connecting $(\mathbf{x}(t), \mathbf{y}(t))$ with (e_i^A, e_j^B)in the phrase space Δ. Such movement would continue until the trajectory crosses one indifferent line and enters a new best response polygon.

3.1 Games with Interior Equilibrium

First we consider those games in which the FP dynamics behave like in Shapley game. Then on the trajectory we can get at least two indifferent lines, which intersect in the interior of the simplex. Indeed, this intersection point is just an interior NE.

To see this, denote the interior equilibrium by $(\bar{\mathbf{x}}, \bar{\mathbf{y}})$, and $\mathrm{supp}(\bar{\mathbf{x}}, \bar{\mathbf{y}})$ has size 3×3. Accoding to equalization principle Lemma 2.1, $\bar{\mathbf{x}}$ makes all the actions of Player B gain the same payoff, and vice versa. Hence $\bar{\mathbf{x}}$ should lie in all the indifferent lines l^B_{jk}. Once the lines are available, we can calculate their intersection point and get $(\bar{\mathbf{x}}, \bar{\mathbf{y}})$.

So from the trajectory, how can we get the indifferent lines? In fact, the switching point on the FP trajectory indicates where the indifferent lines locate. Specifically, if one player, say Player A, changes his action from i to j at time t, then we can infer that Player B's frequency $\mathbf{y}(t)$ no longer makes i-th action to be A's best response. Correspondingly in Δ_B, the trajectory crosses the indifferent line l^A_{ij} between best response polygons Z^A_i and Z^A_j. Collect all the points $\mathbf{y}(t)$ satisfying $a(t-1) = i$ while $a(t) = j$, such points are called switching points. Since the indifferent lines is defined by linear equations in Definition 3.3, we could get the estimation $\hat{l}^{ij}_B$ by these switching points. Consequently, we get the first main result as Theorem 3.1.

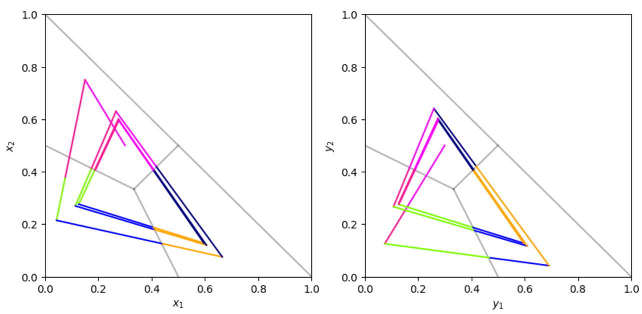

Fig. 5: **Trejectory of FP in Shapley Game**: Here we omit the last component x_3 of $\mathbf{x}$ by $x_3 = 1 - x_1 - x_2$. The gray lines represent indifferent lines while they are also the boundaries of best response polygons. Different colored lines represent different stage action profiles $(a(t), b(t))$ on the trajectory. The initial point is $\mathbf{x}(0) = \mathbf{y}(0) = (0.3, 0.5, 0.2)^T$.

Theorem 3.1. *For 3×3 game (A, B) with interior equilibrium $(\bar{\mathbf{x}}, \bar{\mathbf{y}}) \in \Delta$, if the trajectory of FP does not converge, then $(\hat{\mathbf{x}}, \hat{\mathbf{y}})$, the intersection point of $\{\hat{l}^{ij}_B\}$ and $\{\hat{l}^{ij}_A\}$ respectively, is an approximate NE.*

Proof. First, according the definition of interior NE, $(\bar{\mathbf{x}}, \bar{\mathbf{y}})$ has support size 3×3. And by the equalization principle (Lemma 2.1), $\bar{\mathbf{x}}$ makes Player B's three actions have the same payoff, hence $\bar{\mathbf{x}}$ must belong to all the indifferent lines based on (8), i.e. it is the intersection point of $\{l^B_{ij}\}$. Similarly $\bar{\mathbf{y}}$ is the intersection point of $\{l^A_{ij}\}$.

Next, since the trajectory of FP does not converge, on the trajectory, the three actions of each player must appear infinitely. Otherwise, if after certain times, some action would never appear on the trajectory for Player A or B, then the game would be reduced to a $2 \times n$ game. Then by the previous work [20], FP would converge. Contradiction.

Now from the trajectory of FP, we can get at least two estimations of indifferent lines $\{\hat{l}^B_{ij}\}$ and $\{\hat{l}^A_{ij}\}$. Hence we can calculate their intersection point $(\hat{\mathbf{x}}, \hat{\mathbf{y}})$. We prove that it is an approximate NE.

Apparently as the step size $\frac{1}{t+1}$ in (2) approaches to zero, we can make the estimation of indifferent lines by arbitrary accuracy, which further makes the intersection point $(\hat{\mathbf{x}}, \hat{\mathbf{y}})$ approach to $(\bar{\mathbf{x}}, \bar{\mathbf{y}})$. Denote $\hat{\mathbf{x}} = \bar{\mathbf{x}} + \varepsilon_1, \hat{\mathbf{y}} = \bar{\mathbf{y}} + \varepsilon_2$. We can make $\varepsilon = \max\{\|\varepsilon_1\|, \|\varepsilon_2\|\}$ sufficiently small.

Then for any $\mathbf{x}' \neq \hat{\mathbf{x}}, \mathbf{y}' \neq \hat{\mathbf{y}}$

$$\begin{aligned}
\hat{\mathbf{x}}^T A\hat{\mathbf{y}} =& \bar{\mathbf{x}}^T A\bar{\mathbf{y}} + \varepsilon_1^T A\bar{\mathbf{y}} + \bar{\mathbf{x}}^T A\varepsilon_2 + \varepsilon_1^T A\varepsilon_2 \\
\geq& \mathbf{x}'^T A\bar{\mathbf{y}} + \varepsilon_1^T A\bar{\mathbf{y}} + \bar{\mathbf{x}}^T A\varepsilon_2 + \varepsilon_1^T A\varepsilon_2 \\
=& (\mathbf{x}'^T A\bar{\mathbf{y}} + \mathbf{x}'^T A\varepsilon_2) - \mathbf{x}'^T A\varepsilon_2 \\
&+ \varepsilon_1^T A\bar{\mathbf{y}} + \bar{\mathbf{x}}^T A\varepsilon_2 + \varepsilon_1^T A\varepsilon_2 \\
\geq& \mathbf{x}'^T A\hat{\mathbf{y}} - 4\|A\| \cdot \varepsilon.
\end{aligned}$$

Similarly we have $\hat{\mathbf{x}}^T B\hat{\mathbf{y}} \geq \hat{\mathbf{x}}^T A\mathbf{y}' - 4\|B\| \cdot \varepsilon$. That proves the theorem. □

Remark 2. *The colored lines in Figure 5 and Figure 6 illustrate the trajectory of FP in Shapley game (3) and game (5). The gray lines are the actual indifferent lines, and their intersection is the interior NE. For the stages when players take different stage action profiles $(a(t), b(t))$, we distinguish them with different colors. As t increases, the switching points between two segments with different colors approach to the indifferent line. For both games, we could get at least two indifferent lines, and hence are able to calculate NE.*

Remark 3. *In some sense, Theorem 3.1 responds to the impossibility result in [1], which claims that "It is incapable of capturing the long-term behaviors of the players in all games". Our result indicates that sufficient but finite intermediate data of the learning dynamics is enough to solve NE, since the evolution of the dynamical system can provide adequate and extra information about the game structure, which can be further utilized.*

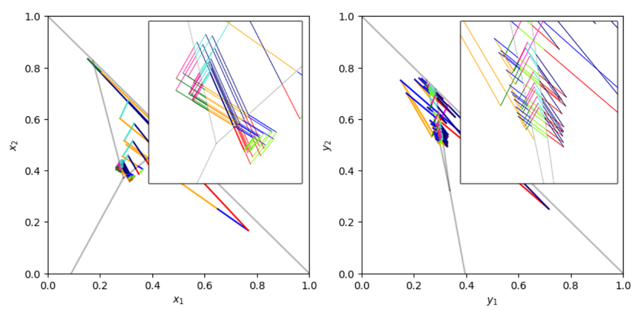

Fig. 6: **Trejectory of FP in Game (5)**: The initial point is $\mathbf{x}(0) = \mathbf{y}(0) = (0.3, 0.5, 0.2)^T$. We can see the trajectory crosses two indifferent lines, but the switching points still approach the indifferent lines. The upper right corner shows the enlarged local details.

3.2 Games without Interior Equilibrium

Section 3.1 gives a method to solve the NE for the games with interior equilibrium and thus Shapley-game-like FP dynamics. For the games without interior NE, two possible behaviors may happen for FP dynamics: either the intersection point of three indifferent lines does not belong to the phrase space Δ; or the intersection point does not exist at all. In this section, we will propose a label-matching priciple and show that one can still utilize the FP trajectory to solve NE in these cases. Besides, we need to examine the edges of each simplex in more details. For convenience, we still use $(\bar{\mathbf{x}}, \bar{\mathbf{y}})$ to denote the intersection point of indifferent lines, and suppose there is no $\bar{\mathbf{x}}$ in Δ_A, but $\bar{\mathbf{y}}$ may or may not belong to Δ_B.

If there exists a point in Δ_A which is the intersection of the edge $E_{ij}^A := \{\mathbf{x} \in \partial\Delta_A \mid x_i + x_j = 1\}$ and the indifferent line l_{kl}^B, we mark the point with a label $[a_i, a_j, b_k, b_l]$. The first two symbols $[a_i, a_j]$ indicate which edge it belongs to and the last two symbols $[b_k, b_l]$ represent which indifferent line goes through the point. As for the point in Δ_B, it can also be labelled with $[a_i, a_j, b_k, b_l]$ as the intersection of edge $E_{kl}^B \subseteq \partial\Delta_B$ and l_{ij}^A.

For example, we consider game (9) below. Figure 7 shows its best response polygons, indifferent lines and all the intersection points of the edges and them, as well as the labelling. Obviously, there are no intersection points of indifferent lines in both Δ_A and Δ_B, i.e. $\bar{\mathbf{x}} \notin \Delta_A, \bar{\mathbf{y}} \notin \Delta_B$. On the other hand, for each Δ_k, there are four intersection points of indifferent lines and edges. For example, in Δ_A, the edge E_{21}^A intersects with l_{13}^B, so we label their intersection point with $[a_1, a_2, b_1, b_3]$, which is colored red in Figure 7.

$$A = \begin{pmatrix} 0 & \frac{2}{3} & 1 \\ \frac{2}{3} & 0 & \frac{2}{3} \\ \frac{1}{3} & \frac{1}{3} & 0 \end{pmatrix}, B = \begin{pmatrix} 0 & 1 & -1 \\ 2 & 0 & \frac{5}{2} \\ \frac{1}{2} & -\frac{1}{2} & 0 \end{pmatrix} \tag{9}$$

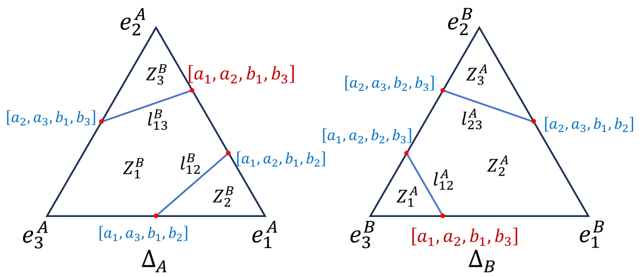

Fig. 7: **Best response polygons, indifferent lines and labelled points of Game (9)**

This labelling operation may encounter two special cases where a point may have more than one label. First, if the indifferent line, say l_{kl}^B, happens to go through the vertex e_k^A, then this vertex should be marked with two labels $[a_k, a_{k'}, b_k, b_l], k' = 1, 2, 3, k' \neq k$ since the vertex belongs to two edges. Second, if two indifferent lines intersect exactly at one point on an edge, this intersection point should also have two labels corresponding to the two indifferent lines. Figure 8 shows the first case and we discuss it in Remark 4 with more details.

Once the labels $[a_i, a_j, b_k, b_l]$ from Δ_A and $[a_i, a_j, b_k, b_l]$ from Δ_B are collected, it is enough to solve NE by the following label-matching principle.

Theorem 3.2. *For the 3×3 game without interior equilibrium, suppose we have collected all the labels of the intersection points $\mathbf{p} \in \Delta_A, \mathbf{q} \in \Delta_B$ of all the indifferent lines $\hat{l}_{ij}^k, k = A, B$ and all the edges of simplex Δ_k. Then for the points $\mathbf{p}$ and $\mathbf{q}$, if they have the same label, $(\mathbf{p}, \mathbf{q})$ is an approximate NE of the game.*

Proof. By the proof of Theorem 3.1, once the estimation of indifferent line, $\hat{l}_{ij}^k$, becomes more and more accurate, the intersection point would approach to the accurate one. On the other hand, if a strategy profile $(\mathbf{p}, \mathbf{q})$ is NE, then any point within its neighborhood is an approximate NE. Hence, we only need to prove that if the accurate intersection points $\mathbf{p}$ and $\mathbf{q}$ have the same label, they constitute a NE.

According to the definition of best response polygon Z_j^B, it is a connected, convex subset of Δ_A since it is defined by a group of linear inequalities in (7), and $\bigcup_{j=1}^3 Z_j^B = \Delta_A$. On the other hand, since there is no $\bar{\mathbf{x}}$ in Δ_A, i.e. the three polygons do not admit a common intersection point, there are only two indifferent lines partioning Δ_A into three polygons. So for the two lines, either their intersection point $\bar{\mathbf{x}}$ does not lie inside Δ_A, or they are parallel to each other.

First we consider the ideal case where every point has only one label. Without loss of generality, suppose two labelled points $\mathbf{p} \in \Delta_A$ and $\mathbf{q} \in \Delta_B$ have the same label $[a_1, a_2, b_2, b_3]$. The first two symbols $[a_1, a_2]$ means Player B's strategy $\mathbf{q}$ belongs to indifferent line l_{12}^A. In other words, against $\mathbf{q}$,

$$(A\mathbf{q})_1 = (A\mathbf{q})_2 > (A\mathbf{q})_3.$$

And since the label of $\mathbf{p}$ also has $[a_1, a_2]$, this ensures it lies on the edge E_{12}^A. Hence $\mathbf{p}$ satisfies the equivalization principle in Lemma 2.1. Similarly, $\mathbf{q}$ also satisfies the equivalization principle according to $[b_2, b_3]$. Hence $(\mathbf{p}, \mathbf{q})$ is a NE.

Now we consider the two special cases where some point has more than one label. Without loss of generality, suppose $(\mathbf{p}, \mathbf{q})$ has a matched label $[a_1, a_2, b_2, b_3]$, but $\mathbf{p}$ is the vertex e_1^A, hence $\mathbf{p}$ has another but unmatched label $[a_1, a_3, b_2, b_3]$. We now prove $(\mathbf{p}, \mathbf{q})$ is still a NE.

Since the matched label is $[a_1, a_2, \cdot, \cdot]$, the 1-st and 2-nd actions of Player A would have the same payoff against $\mathbf{q}$, i.e. $(A\mathbf{q})_1 = (A\mathbf{q})_2$. For any strategy $\mathbf{p}' = (p_1', p_2', p_3')^T \neq \mathbf{p}$, i.e. $p_1' < 1$, if $p_3' = 0$, we have

$$\begin{aligned}(\mathbf{p}')^T A\mathbf{q} &= \sum_{i=1}^{3} p_i' \cdot (A\mathbf{q})_i = p_1' \cdot (A\mathbf{q})_1 + p_2' \cdot (A\mathbf{q})_2 \\ &= (A\mathbf{q})_1 = \mathbf{p}^T A\mathbf{q}.\end{aligned}$$

If $p_3' > 0$, we have

$$\begin{aligned}\mathbf{p}^T A\mathbf{q} - (\mathbf{p}')^T A\mathbf{q} &= (1 - p_1' - p_2') \cdot (A\mathbf{q})_1 - p_3' \cdot (A\mathbf{q})_3 \\ &= p_3'[(A\mathbf{q})_1 - (A\mathbf{q})_3] > 0.\end{aligned}$$

Hence $\mathbf{p}$ satisfies the equalization principle in Lemma 2.1. Similarly, $\mathbf{q}$ also satisfies the principle. Thus $(\mathbf{p}, \mathbf{q})$ is a NE.

By similar arguments, we can prove for the other special case where two indifferent lines and one edge intersect exactly at the same point, if $\mathbf{p}$ and $\mathbf{q}$ have a matched label, they constitute a NE. That proves the theorem. □

Remark 4. *Consider a game with payoff matrices*

$$A = \begin{pmatrix} 0 & -\frac{4}{3} & \frac{1}{3} \\ \frac{1}{3} & 0 & -\frac{2}{3} \\ \frac{2}{3} & -\frac{1}{3} & 0 \end{pmatrix}, B = \begin{pmatrix} 0 & 1 & -2 \\ 2 & 0 & -3 \\ 0 & -1 & 0 \end{pmatrix}. \quad (10)$$

In Figure 8, we illustrate the special case where some indifferent line goes through one vertex of the simplex. For this game, e_3^A has two labels $[a_1, a_3, b_1, b_3]$ and $[a_2, a_3, b_1, b_3]$, implying that when Player A use his 3-rd action, Player B's 1-st and 3-rd are both best response against it, i.e. there are more than one best response for some pure action.

Such game is called degenerate in the literature [5, 20]. Many results are obtained under non-degenerate assumption [20, 22], while such degenerate game seems to be ignored. However, our analysis does not eliminate such degenerate game but incorporate it into an unified consideration.

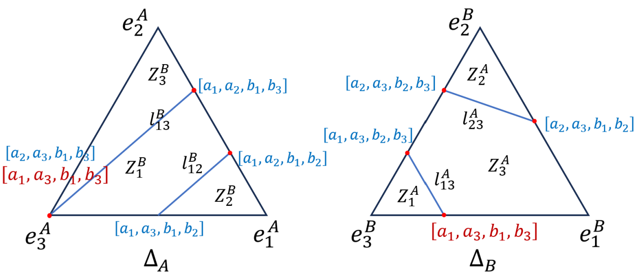

Fig. 8: **Best response polygons, indifferent lines and labelled points of Game (10).**

Remark 5. *Label-matching principle in Theorem 3.2 utilizes the geometrical feature of FP dynamics and labelling operation, and makes the computation and verification of NE easier than tranditional combinational methods, like support enumeration etc. [5]. In computation, now there is no need to enumerate all the possible forms of support of potential NE, such as $((x_1, x_2, 0), (y_1, y_2, 0))$, $((x_1, x_2, 0), (y_1, 0, y_3))$. In verification, given the intersection points $\mathbf{p}$ and $\mathbf{q}$ of the indifferent lines and the edges of the simplexes, labelling can help us quickly check whether $(\mathbf{p}, \mathbf{q})$ is a NE. Otherwise, we need to solve the optimal payoff of Player A against $\mathbf{q}$ and optimal payoff of Player B against $\mathbf{p}$. In other words, label-matching principle itself contains the information of optimization, hence we do not need extra computation.*

Remark 6. *We note that the method relies solely on the trajectory of the learning dynamics, eliminating the need for prior knowledge about the opponent's payoff matrix. This is the same with FP. In this sense, the process still constitutes an uncoupled dynamics [34].*

As a summary of Section 3, Theorem 3.1 and Theorem 3.2 together provide a geometrical method in solving all the 3×3 games. Our method is based on the trajectory of FP dynamics. Different from common interest which focuses on whether the latest iteration $(\mathbf{x}(t), \mathbf{y}(t))$ converges or not, we highlight the switching points where one player finds that the opponent's action changes from one to another. By collecting these switching points, we can establish the estimation of indifferent lines and hence calculate their intersection points. If the intersection point exists in Δ, by Theorem 3.1, it is the approximate NE according to equalization principle; if the intersection point does not exist in Δ, the labelling operation and label-matching principle in Theorem 3.2 ensure that we can find NE in the intersection points of the indifferent lines and the edges of the simplexes. The procedures are summarized in Algorithm 1.

Algorithm 1 Calculate NE in 3×3 Normal Form Game

Require: The $(\mathbf{x}(0), \mathbf{y}(0))$, iteration time T, threshold t_0
for $t = 1, 2, \cdots, T$ **do**
  $a(t) \leftarrow \mathrm{BR}(\mathbf{x}(t))$
  $b(t) \leftarrow \mathrm{BR}(\mathbf{y}(t))$
  if player $k \in \{A, B\}$ finds his oppoent changes action from i to j and $t \geq t_0$ **then**
    append the player k's frequnecy into set $\mathcal{D}_{ij}^k$
  end if
  $\mathbf{x}(t) \leftarrow \frac{1}{t}a(t) + \frac{t-1}{t}\mathbf{x}(t)$
  $\mathbf{y}(t) \leftarrow \frac{1}{t}b(t) + \frac{t-1}{t}\mathbf{y}(t)$
end for
for each data set $\mathcal{D}_{ij}^k$ **do**
  calculate the indifferent set $\hat{l}_{ij}^k = w^T z + b$, where $w, b \in \arg\max_{w,b} \sum_{z_j \in \mathcal{D}_i^k \to l} \mathrm{d}(z_j, w^T z + b)$
end for
calculate the intersection point $(\bar{\mathbf{x}}, \bar{\mathbf{y}})$ of all indifferent lines $\hat{l}_{ij}^k$ for each player
if $(\bar{\mathbf{x}}, \bar{\mathbf{y}})$ lies in the phrase space **then**
  return the approximate NE $(\bar{\mathbf{x}}, \bar{\mathbf{y}})$
else
  Label each intersection point with support 2×2, do label-matching, and return the matched point pair.
end if

4 Conclusions

In the framework of learning in games, NE is often regarded as a steady state of learning dynamics. This paper is motivated by the question how we can make decisions when the learning dynamics does not converge or even has erratic long-term behavior, which is often the case according to recent literature. By leveraging the equalization principle of NE and distinguishing the special opponent's-action-switching-point, this paper breaks the seemingly existing impasse and proposes a novel geometrical method to solve NE for the general non-zero-sum games.

To provide a complete description, we studied 3×3 non-zero-sum games. By dividing them into two classes and combining equalization principle with the points on FP trajectory, this paper provides a new perspective to deal with data generated from the learning algorithms, which may shed some light on general problems. Our further experiments show that for some 4×4 non-zero-sum games [44] and $2 \times 2 \times 2$ stochastic games with special structure [45], the method in this paper does still work, see our following complete paper.

On the other hand, the implementation of this idea in general large scale games still faces many challenges, due to exponential explosion of dimensions, invisibility to high-dimensional space by human mind and the lack of theoretical research on high-dimensional dynamical systems. We need to reconstruct the indifferent hyperplanes for high-dimensional games, and solve large-scale linear equations.

We leave them as future work.

References

[1] J. Milionis, C. Papadimitriou, G. Piliouras, and K. Spendlove, An impossibility theorem in game dynamics, *Proceedings of the National Academy of Sciences*, 120.41: e2305349120, 2023.

[2] J. Nash, Non-cooperative games. *Annals of mathematics*, 286-295, 1951.

[3] C. E. Lemke and J. T. Howson Jr., Equilibrium points of bimatrix games, *Journal of the Society for Industrial and Applied Mathematics*, 12:413423, 1964.

[4] H. Scarf, The Approximation of Fixed Points of a Continuous Mapping, *SIAM Journal on Applied Mathematics*, 15(5):13281343, 1967.

[5] N. Nisan, T. Roughgarden, E. Tardos, and V. V. Vazirani, Eds., *Algorithmic Game Theory*. Cambridge: Cambridge University Press, 2007.

[6] S. Smale, A convergent process of price adjustment and global newton methods, *Journal of Mathematical Economics*, 3(2):107-120, 1976.

[7] R. B. Kellogg, T. Y. Li and J. A. Yorke, A Method of Continuation for Calculating a Brouwer Fixed Point, *Fixed Points: Algorithms and Applications*, Academic Press: 133-147, 1977.

[8] C. H. Papadimitriou, On the Complexity of the Parity Argument and Other Inefficient Proofs of Existence, *Journal of Computer and System Sciences*, 48(3):498532, 1994.

[9] C. Daskalakis, P.W. Goldberg, and C.H. Papadimitriou, The complexity of computing a Nash equilibrium, *STOC 06: Proceedings of the 38th Annual ACM Symposium on Theory of Computing*, pages 7178, 2006.

[10] C. Daskalakis and C.H. Papadimitriou, Three-player games are hard, In *Electronic Colloquium in Computational Complexity*, TR05-139, 2005.

[11] X. Chen and X. Deng, 3-Nash is PPAD-Complete, *Electronic Colloquium on Computational Complexity (ECCC)*, 134, 2005.

[12] X. Chen and X. Deng, Settling the Complexity of Two-Player Nash Equilibrium, *the 47th Annual IEEE Symposium on Foundations of Computer Science (FOCS)*, 2006.

[13] X. Deng, N. Li, D. Mguni, J. Wang, Y. Yang, On the Complexity of Computing Markov Perfect Equilibrium in General-Sum Stochastic Games, *National Science Review*, 10(1), 2023.

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

[15] V. Krishna, T. Sjöström, Learning in Games: Fictitious Play Dynamics, *Cooperation: Game-Theoretic Approaches, NATO ASI Series*, S. Hart, A. Mas-Colell, Eds., vol 155:258-273, Springer, Berlin, Heidelberg, 1997.

[16] G. W. Brown, Iterative solution of games by fictitious play, *Activity analysis of production and allocation*, 13(1):374376, 1951.

[17] J. Robinson, An iterative method of solving a game, *Annals of Mathematics*, 54 (2):296301, 1951.

[18] K. Miyasawa, On the convergence of learning processes in a 2×2 non-zero-person game, *Technical Report Research Memorandum* No. 33, Econometric Research Program, Princeton University, 1961.

[19] D. Monderer, L. S. Shapley, Fictitious play property for games with identical interests, *Journal of Economic Theory*, 68(1):258265, 1996.

[20] U. Berger, Fictitious play in $2 \times$ n games, *Journal of Economic Theory*, 120(2):139154, 2005.

[21] J. Nachbar, Evolutionary selection dynamics in games: Convergence and limit properties, *International journal of game theory*, 19 : 59-89, 1990.

[22] U. Berger, Two more classes of games with the continuous-time fictitious play property. *Games and Economic Behavior*, 60(2): 247-261, 2007.

[23] J. Hannan, Approximation to Bayes risk in repeated play, *Contributions to the Theory of Games*, M. Dresher, A. W. Tucker, P. Wolfe, Eds., Vol. 3:97139, Princeton University Press, 1957.

[24] S. Hart, A. MasColell, A simple adaptive procedure leading to correlated equilibrium. *Econometrica*, 68(5): 1127-1150, 2000.

[25] Y. Freund, R. E. Schapire, Adaptive game playing using multiplicative weights, *Games and Economic Behavior*, 29(1):79103, 1999.

[26] N. Littlestone, M. K. Warmuth, The weighted majority algorithm, *Information and Computation*, 108(2):212261, 1994.

[27] N. Cesa-Bianchi and G. Lugosi, *Prediction, learning, and games*, Cambridge university press, 2006.

[28] T. Roughgarden, *Twenty lectures on algorithmic game theory*. Cambridge University Press, 2016.

[29] J. Heinrich, M. Lanctot, D. Silver, Fictitious self-play in extensive-form games, *International Conference on Machine Learning, PMLR*, 2015.

[30] M. O. Sayin, F. Parise and A. Ozdaglar, Fictitious play in zero-sum stochastic games, *SIAM Journal on Control and Optimization*, 60(4): 2095-2114, 2022.

[31] M. Zinkevich, M. Johanson, M. Bowling, and C. Piccione, Regret minimization in games with incomplete information, *Advances in neural information processing systems*, vol. 20, 2007.

[32] N. Brown, A. Lerer, S. Gross, and T. Sandholm, Deep counterfactual regret minimization, *International conference on machine learning, PMLR*, pp.793802, 2019.

[33] N. Brown and T. Sandholm, Superhuman ai for heads-up no-limit poker: Libratus beats top professionals, *Science*, 359(6374): 418424, 2018.

[34] S. Hart and A. Mas-Colell, Uncoupled Dynamics Do Not Lead to Nash Equilibrium, *The American Economic Review*, 93(5):1830-1836, 2003.

[35] L. Shapley, Some topics in two-person games, *Advances in Game Theory*, 52:1-29, 1964.

[36] C. Sparrow, S. van Strien, and C. Harris, Fictitious play in 3×3 games: The transition between periodic and chaotic behaviour, *Games and Economic Behavior*, 63(1):259291, 2008.

[37] S. Van Strien and C. Sparrow, Fictitious play in 3 Œ 3 games: Chaos and dithering behaviour, *Games and Economic Behavior*, 73(1):262286, 2011.

[38] P. Mertikopoulos, *Online optimization and learning in games: Theory and applications*, Habilitation à Diriger des Recherches (HDR), Université Grenoble-Alpes, 2019.

[39] H. Martin, Shapley Polygons in 4×4 Games, *Games*, 1(3): 189-220, 2010. Games 1(3), 2010, 2010,

[40] P. Mertikopoulos, C. Papadimitriou, G. Piliouras, Cycles in adversarial regularized learning, *Proceedings of the twenty-ninth annual ACM-SIAM symposium on discrete algorithms*: 2703-2717, 2018.

[41] A. R. Karlin, Y. Peres, *Game theory, alive*, American Mathematical Society, 2017.

[42] Y. Narahari, *Game theory and mechanism design*. World Scientific, 2014.

[43] G. Ostrovski and S. Van Strien, Payoff performance of fictitious play, *Journal of Dynamics and Games*, 1(4):621638, 2013.

[44] M. Hahn, Shapley Polygons in 4 x 4 Games, *Games*, 1: 189-220, 2010.

[45] G. Schoenmakers, J. Flesch and F.Thuijsman, Fictitious play in stochastic games, *Mathematical Methods of Operations Research*, 66:315-325, 2007.