

# Poincaré-Bendixson Limit Sets in Multi-Agent Learning

Anonymous Author(s)

Submission Id: 344

## ABSTRACT

A key challenge of evolutionary game theory and multi-agent learning is to characterize the limit behavior of game dynamics. Whereas convergence is often a property of learning algorithms in games satisfying a particular reward structure (e.g. zero-sum), it is well known, that for general payoffs even basic learning models, such as the replicator dynamics, are not guaranteed to converge. Worse yet, chaotic behavior is possible even in rather simple games, such as variants of Rock-Paper-Scissors games [52]. Although chaotic behavior in learning dynamics can be precluded by the celebrated Poincaré-Bendixson theorem, it is only applicable to low-dimensional settings. Are there other characteristics of a game, which can force regularity in the limit sets of learning?

We show that behaviors consistent with the Poincaré-Bendixson theorem (limit cycles, but no chaotic attractor) follows purely based on the topological structure of the interaction graph, even for high-dimensional settings with arbitrary number of players and arbitrary payoff matrices. We prove our result for a wide class of follow-the-regularized leader (FoReL) dynamics, which generalize replicator dynamics, for games where each player has two strategies at disposal, and for interaction graphs where payoffs of each agent are only affected by one other agent (i.e. interaction graphs of indegree one). Since chaos has been observed in a game with only two players and three strategies, this class of non-chaotic games is in a sense maximal. Moreover, we provide simple conditions, under which such behavior translates to efficiency guarantees, implying that FoReL learning achieves time average sum of payoffs at least as good as that of a Nash equilibrium; and connecting the topology of the dynamics to social welfare analysis.

## KEYWORDS

Replicator Dynamics, Follow-the-Regularized Leader, Polymatrix Games, Poincaré-Bendixson Theorem, Regret Minimization

### ACM Reference Format:

Anonymous Author(s). 2022. Poincaré-Bendixson Limit Sets in Multi-Agent Learning. In *Proc. of the 21st International Conference on Autonomous Agents and Multiagent Systems (AAMAS 2022)*, Auckland, New Zealand, May 9–13, 2022, IFAAMAS, 9 pages.

## 1 INTRODUCTION

Dynamical systems and evolutionary game theory have been instrumental in much of modern research on multi-agent learning [8, 12, 21, 47, 53, 54, 56]. In particular, characterizing convergence and limit sets of learning trajectories is key in understanding long-term behavior of multi-agent systems. However, even in simple games, such as Rock-Paper-Scissors [45, 52] models of evolution and learning are not guaranteed to converge; and even beyond

cycles, long-term behavior may lead to chaotic behavior, known to the dynamical systems community, e.g., from weather models [36]. Not only does chaos manifest itself even in simple games with two players, but moreover, a string of recent results seems to suggest that such chaotic, unpredictable behavior may indeed be the norm across a variety of simple low dimensional game dynamics [4–6, 13, 15, 18, 20, 43, 50, 57]. Importantly, these results are persistent even for the well known class of Follow-the-Regularized-leader (FoReL) dynamics [14, 39], despite the fact that FoReL dynamics include some of the most widely studied learning dynamics such as replicator dynamics [27, 55], the continuous time analogue of the Multiplicative Weights Update meta-algorithm [3], and are well known for their optimal regret properties. Finally, the emergence of chaotic behavior has been connected with increased social inefficiency showing that chaotic dynamics may lead to highly inefficient outcomes [16, 48]. Such profoundly negative results lead us to the following natural questions:

- Do there exist simple, robust conditions under which learning behaves well?
- Which type of games lie at the “edge of chaos”?
- Does dynamic simplicity translate to high efficiency, social welfare?

Traditionally, a lot of work has focused on showing that in specific classes of games (e.g. zero-sum or potential games), learning dynamics can lead to convergence and equilibration, see [12, 19, 51, 60] and references therein. There are few results that span over to general sum games and games of arbitrary payoff structures; however arguably such general approaches are essential in modern research on multi-agent learning. For instance, unstructured payoffs can naturally occur, when stochastic extensive form games are used to create empirical normal form games, by averaging payoffs from simulations for combinations of strategies [35, 40, 58]. Unstructured payoffs also arise in many real world applications – such as e.g. modelling impact of investing strategies of large funds on the stock market. While equilibration may not always be possible in such cases, one can still wish to ensure regularity of sorts in the learning outcomes of the multiagent system. In particular, the famous Poincaré-Bendixson theorem (Theorem 1) ensures that two-dimensional continuous learning and adaptation dynamics will never form truly chaotic outcomes. This comes however at a cost; even though no specific payoff structure is needed, the underlying learning dynamics need to be at most two dimensional.

**Our approach and results.** Rather than by making assumptions on the reward structure, or on the dimension, we explore a different type of constraint in games. We show, that the limit behavior of learning can be determined solely by the topological-combinatorial structure of the game, regardless of the number of players, or algebraic correlations between the payoffs (e.g. zero-sum). Firstly, we restrict ourselves to *binary* games [9, 38, 61] where players have two strategies to choose from. Secondly, we assume that every agent can be affected by the behavior of up to one other

agent. Finally, we add a technical restriction that the game is connected – it cannot be decomposed into two subgames completely independent of each other. Such games encompass, among others, all  $2 \times 2$  games [23], Jordan’s game [22, 25, 29], as well as easily identifiable subclasses of real-world systems, where the graph structure is evident – e.g. certain traffic networks [1, 34], supply chains [11] or problems of water allocation in deltas [2, 31]. Under these assumptions, in Section 3 we prove our main contribution in form of Theorems 3, 4 – that the limit behavior of FoReL learning of these games is always consistent with the Poincaré-Bendixson theorem.

Having excluded the presence of chaos, we further analyze quantitative properties of binary games, which admit cyclic interaction graphs. In Section 4 we show that under additional but structurally robust assumptions on the payoff matrices (i.e., assumptions that remain valid after small perturbations of the payoff matrices, so are suitable for e.g. empirical payoff matrices), one can derive positive results about the efficiency of the time-average behavior of the dynamics regardless of whether they are convergent or not. As it is typically the case in the Price of Anarchy (PoA) literature [33] we focus on the measure of *social welfare* – the sum of individual payoffs, but whereas the typical PoA literature tries to argue that regret-minimizing dynamics (such as FoReL) are at most a constant factor worse than the behavior of the worst case Nash equilibrium [48, 49], we instead show that FoReL dynamics are always at least as efficient as the worst case Nash equilibrium. Finally, in Section 5 we provide examples of games satisfying our assumptions, and their possible limit behaviors, as well as a counterexample – a simple, binary game, which breaks our assumptions, and induces chaotic learning dynamics.

**Related work.** First of all, we would like to point to several papers containing complementary results, in form of examples of simple FoReL systems with chaotic dynamics. In addition to the papers mentioned in the introduction, we would like to highlight a chaotic example of Sato et al. [52] in a 2-player, 3-action game as well as two complex/chaotic examples in 3-player binary games without structured interactions – by Plank [46], and recently by Peixe and Rodrigues [44]. Contrasting these to our assumptions (i.e. binary games, and previous neighbor interactions), we see that our results establish a maximal class of games for which such regularity results on limit sets are possible.

Research which considers non-convergence but focuses on non-chaoticity results is scarce. In the closest works to ours, [17, 41, 42], the authors leverage the Poincaré-Bendixson theorem to show that the limit behavior of bounded learning trajectories in certain learning systems can be either convergent or cyclic, but in particular no chaotic attractor is possible. However, they do so by assuming respectively low-dimensionality (three players limit) or non-generic structure on the set of allowable games, which allows for dimensionality reduction (i.e. a network of  $2 \times 2$  zero-sum / coordination games). In terms of connections between cyclic behavior and the efficiency of learning dynamics, [32] has shown that for a class of three agent, two strategy games a cyclic attractor can result in social welfare (sum of payoffs) that can be better than the Nash equilibrium payoff, however, that result is once again constrained to exact game theoretic model.

## 2 PRELIMINARIES

### 2.1 Normal form games

A *finite game in normal form* consists of a set of  $N$  players each with a finite set of *strategies*  $\mathcal{A}_i$ . The preferences of each player are represented by the payoff function  $u_i : \prod_i \mathcal{A}_i \rightarrow \mathbb{R}$ . To model behavior at scale, or probabilistic strategy choices, one assumes that players use *mixed strategies*, i.e. probability distributions  $(x_{i\alpha_i})_{\alpha_i \in \mathcal{A}_i} \in \Delta(\mathcal{A}_i) =: X_i$ . With slight abuse of notation, the expected payoff of the  $i$ -th player in the profile  $(x_{i\alpha_i})_{i, \alpha_i}$  is denoted by  $u_i$  again, and given by

$$u_i(x) = \sum_{\alpha_1 \in \mathcal{A}_1, \dots, \alpha_N \in \mathcal{A}_N} u_i(\alpha_1, \dots, \alpha_N) x_{1\alpha_1} \dots x_{N\alpha_N}. \quad (1)$$

A mixed strategy  $\hat{x}$  is a *Nash equilibrium* iff  $\forall i \forall x : x_j = \hat{x}_j, j \neq i$  we have  $u_i(x) \leq u_i(\hat{x})$ ; in other words no player can unilaterally increase their payoff by changing their strategy distribution. The *minimax value* for player  $i$  is given by  $\min_{x_{-i}} \max_{x_i} u_i(x)$ , where  $x_{-i} := (x_j)_{j \neq i}$ . It is the smallest possible value player  $i$  can be forced to attain by other players, without them knowing player  $i$  strategy. We call a game *binary* iff  $|\mathcal{A}_i| = 2$  for all  $i$ .

### 2.2 Graphical polymatrix games

To model the topology of interactions between players, we restrict our attention to a subset of normal form games, where the structure of interactions between players can be encoded by a graph of two-player normal form subgames, leading us to consider so-called *graphical polymatrix games* (GPGs) [28, 30, 59]. A simple directed graph is a pair  $(\mathcal{V}, \mathcal{E})$ , where  $\mathcal{V} = \{1, \dots, N\}$  is a finite set of *vertices* (representing the players), and  $\mathcal{E}$  is a set of ordered vertex pairs (*edges*), where the first element is called predecessor, and the second is called successor. Each edge  $(i, k)$  has an associated two-player normal form game, where only the successor  $k$  is assigned payoffs. These are represented by a matrix  $A^{i,k}$  with rows enumerating strategies of player  $k$ , and columns enumerating strategies of player  $i$ . For a given strategy profile  $s = \{s_i\}_i \in \prod_i \mathcal{A}_i$  the payoffs for player  $k$  in the full game are then determined as the sum

$$u_k(s) = \sum_{i:(i,k) \in \mathcal{E}} A^{i,k}(s_i, s_k). \quad (2)$$

The payoffs can be extended to mixed strategies in a standard multilinear fashion:

$$u_k(x) = \sum_{i:(i,k) \in \mathcal{E}} \sum_{x_{s_i}, x_{s_k}} A^{i,k}(s_i, s_k) x_{s_i} x_{s_k}. \quad (3)$$

A situation, where both the successor  $k$  and also the predecessor  $i$  obtain a reward can be modelled by including both edges  $(i, k)$  and  $(k, i)$  in the graph.

We say that a simple directed graph is weakly connected, if any two vertices can be connected by a set of edges, where the direction of the edges is not taken into account. This is a weaker condition than strong connectedness, where each pair of vertices needs to be connected by a *path*, i.e. a sequence of edges, together with associated vertices, where the successor in one edge needs to be the predecessor in the next one. The *indegree* of a vertex, is the amount of edges for which the vertex is the successor (in other words: the number of its predecessors). The *outdegree* is the amount of edges, for which the vertex is the predecessor, i.e. the number of

its successors. A *cycle* is a path, where the predecessor in the first edge is the successor in the last edge. For our exposition we shall identify cycles modulo shifts, i.e. if two paths consist of the same edges in shifted order, then they form the same cycle. In this paper we consider two types of weakly connected GPGs:

- (1) firstly, *cyclic* games, where the interaction between the agents forms a cycle, where each agent interacts only with the previous neighbor. We observe that in such a cyclic game the indegree and outdegree of each vertex is one. For simplicity, we will label the nodes of such  $N$ -player games by natural numbers  $i = 0, 1, \dots, N$ , and use the convention that  $i$ th node is the successor to  $i - 1$ st node, and that the 0th node is identified with the  $N$ th node.
- (2) Secondly, a more general class of graphical games, where each player's payoffs depend on up to one other player, i.e. the indegree of each vertex is at most one. For a vertex  $i \in \mathcal{V}$ , we will then denote the predecessor vertex by  $\hat{i}$ , if it exists. For cyclic games we have  $\hat{i} = i - 1$ .

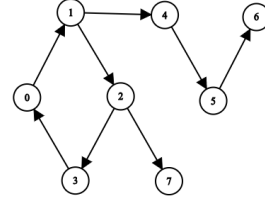
Below, we state and prove a simple lemma, which characterizes the one-predecessor assumption in terms of graph topology and clarifies the relation between cyclic and indegree-one graphs (c.f. Figure 1).

**LEMMA 1.** *Let  $(\mathcal{V}, \mathcal{E})$  be a weakly connected, simple, directed graph. If the indegree of each vertex is at most one, then, the graph can have up to one cycle. If the graph has no cycle, then it has to have at most one root vertex, i.e. a vertex of indegree zero, such that all other vertices are connected to it by a unique, directed path.*

**PROOF.** For the first part of the lemma, let us assume the contrary: that  $a_1, a_2$  are nodes of two distinct cycles within the same weakly connected component. The edges between  $a_1$  and  $a_2$  need to form a path (otherwise there would be a vertex with two predecessors). Assume the path leads from  $a_1$  to  $a_2$ , and let  $a_0$  be the first vertex which is both on the path and on  $a_2$  cycle. Then  $a_0$  has two predecessors, which leads to a contradiction.

For the second part of the lemma, we can argue as follows. If any vertex would have a sequence of predecessors which would not form a cycle, then by backtracking through the predecessors we could identify an infinite collection of distinct vertices. Therefore, there needs to be at least one root node for each vertex. The path from such root node to the given vertex needs to be unique, otherwise one could identify a vertex along the path with two predecessors. Finally, if there were two root nodes, from connectedness it follows that there must be a node with two predecessors on the edges between them.  $\square$

**REMARK 1.** *Under the assumptions of Lemma 1, if the graph has a cycle, then the cycle enjoys similar properties as a root node would; i.e. there are no paths from outside of the cycle to it (otherwise one vertex in the cycle would have two predecessors), and all vertices outside of the cycle have to be connected by a path from one of the vertices of the cycle (unique, up to the starting point within the cycle). Further on we will refer to such cycle as the root cycle.*



**Figure 1:** A weakly connected graph where each vertex is of indegree at most one.

### 2.3 Follow-the-regularized-leader equations

Denote by  $v_{i\alpha_i}(x) := u_i(\alpha_i; x_{-i})$  and  $v_i(x) = (v_{i\alpha_i}(x))_{\alpha_i \in \mathcal{A}_i}$ . To model the dynamics of learning we use a class of learning systems known as *follow-the-regularized-leader* systems (FoReL) [12, 53]. This class encompasses a variety of models ranging from gradient to replicator dynamics, and allows for natural description of agent learning as regularized maximization of individual payoffs.

FoReL dynamics for player  $i$  are defined by evolution of *utilities*  $y_i = \{y_{i\alpha_i}\}_{\alpha_i \in \mathcal{A}_i} \in \mathbb{R}^{|\mathcal{A}_i|}$  – that is real numbers representing a score each player assigns to each respective strategy – by the integral equation

$$\begin{aligned} y_i(t) &= y_i(0) + \int_0^t v_i(x(s)) ds, \\ x_i(t) &= Q_i(y_i(t)), \end{aligned} \quad (4)$$

where the *choice map*  $Q = (Q_1, \dots, Q_N)$ ,  $Q_i : \mathbb{R}^{|\mathcal{A}_i|} \rightarrow \mathcal{X}_i$ , which determines the evaluated strategy profile  $x(t)$  is given on each coordinate by:

$$Q_i(y_i) = \operatorname{argmax}_{x_i \in \mathcal{X}_i} \{ \langle y_i, x_i \rangle - h_i(x_i) \}. \quad (5)$$

In the above  $h_i : \mathcal{X}_i \rightarrow \mathbb{R} \cup \{-\infty, \infty\}$  is a convex regularizer function, representing a regularization/exploration term. The equation (4) represents how players adapt their mixed strategies to changing utility values. Observe, that without the regularization term, the map  $Q_i$  would simply put all weight on the strategy with the highest utility.

In binary games, each agent has only two strategies at his disposal, say  $\alpha_0, \alpha_1$ . The variable  $x_i$  denotes then the proportion of time player  $i$  plays strategy  $\alpha_0$ , and the proportion of  $\alpha_1$  is given by  $1 - x_i$ . Following [39], we introduce new variables  $z_i := y_{i\alpha_0} - y_{i\alpha_1} \in \mathbb{R}$ , representing the difference in utilities between playing strategy  $\alpha_0$  and  $\alpha_1$ . It is intuitively clear, and it was proved formally e.g. in [39] that  $Q_i(z_i + c, c)$  is constant in  $c$ , and therefore, without loss of generality, we can set  $c := 0$ , and restrict our considerations to a  $z$ -dependent choice map  $\hat{Q}_i(z_i) := Q_i(z_i, 0)$ . Provided that  $Q$  is sufficiently regular (e.g. continuous), the integral equation (4) can be converted to a system of differential equations

$$\dot{z} = V(z) \quad (6)$$

given coordinate-wise by

$$V_i(z) := v_{i\alpha_0}(\hat{Q}(z)) - v_{i\alpha_1}(\hat{Q}(z)), \quad (7)$$

for details again see [39].

REMARK 2. *An intuitively obvious, but technically important observation is that evolution of  $i$ th coordinates of the system (4), and, in turn (7) depends solely on the values of  $x_j$  or  $z_j$ , respectively, for nodes  $j$  that influence the payoffs of  $i$ . In particular, for GPGs we have  $\partial V_i / \partial z_j \neq 0$  implies that there is an edge from  $j$  to  $i$  in the game graph; and for GPGs with up to one predecessor, without loss of generality we can rewrite (6) as*

$$\dot{z}_i = V_i(z_i) = v_{i\alpha_0}(\hat{Q}_i(z_i)) - v_{i\alpha_1}(\hat{Q}_i(z_i)) \quad (8)$$

As previously hinted, for equation (7) to be well-posed, we need to pose certain regularity conditions on the regularizer. The following lemma determines desirable properties of monotonicity and smoothness of the choice map, when a player has exactly two strategies at disposal (so  $X_i = [0, 1]$ ).

LEMMA 2. *Assume that the regularizer  $h_i$  satisfies the following conditions:*

- (1)  $h_i \in C^2((0, 1)) \cap C^0([0, 1])$  (smoothness),
- (2)  $h'_i(x_i) \rightarrow -\infty$  as  $x_i \rightarrow 0$  and  $h'_i(x_i) \rightarrow \infty$  as  $x_i \rightarrow 1$  (steepness),
- (3)  $h''_i(x_i) > 0$  for  $x \in (0, 1)$  (strict convexity).

Then  $\hat{Q}_i \in C^1(\mathbb{R})$  and  $\hat{Q}'_i(z_i) > 0$ .

PROOF. For a given  $z_i$ ,  $\hat{Q}_i(z_i)$  is defined as the maximizer of  $\langle (z_i, 0), (x_i, 1 - x_i) \rangle - h_i(x_i)$  over  $x_i \in [0, 1]$ . We have

$$\langle (z_i, 0), (x_i, 1 - x_i) \rangle - h_i(x_i) = z_i x_i - h_i(x_i). \quad (9)$$

From steepness, continuity and strict convexity it follows that  $h_i(0) = h_i(1) = \infty$  so the maximum cannot be attained there. A necessary condition for maximum to be attained in  $(0, 1)$  is

$$z_i = h'_i(x_i). \quad (10)$$

From steepness and strict convexity it follows that equation (10) has a unique solution  $x_i =: \hat{Q}_i(z_i)$  for any  $z_i \in \mathbb{R}$ . From the inverse function theorem we have

$$\frac{\partial x_i}{\partial z_i} = \hat{Q}'_i(z_i) = 1/h''_i(x_i) > 0, \quad (11)$$

which also implies that  $\hat{Q}_i$  is  $C^1$ .  $\square$

Perhaps the best known example of a FoReL learning system are the replicator equations [55], where the regularizer is given by

$$h_i(x_i) := \sum_{\alpha_i} x_{i\alpha_i} \log x_{i\alpha_i}. \quad (12)$$

In particular, such regularizer satisfies the assumptions of Lemma 2, and yields the following equations for a binary GPG with up to one predecessor:

$$\begin{aligned} \dot{z}_i &= \sum_{j,k \in \{0,1\}} (-1)^{(j+k)} A^{\hat{i},i}(\alpha_j, \alpha_k) \frac{\exp(z_i)}{1 + \exp(z_i)} \\ &\quad - A^{\hat{i},i}(\alpha_1, \alpha_1) + A^{\hat{i},i}(\alpha_1, \alpha_0), \quad i = 1, \dots, N \end{aligned} \quad (13)$$

which translates to the following system in original  $(x)$  coordinates:

$$\begin{aligned} \dot{x}_i &= x_i(1 - x_i) \sum_{j,k \in \{0,1\}} (-1)^{(j+k)} A^{\hat{i},i}(\alpha_j, \alpha_k) x_j \\ &\quad - x_i(1 - x_i) \left( A^{\hat{i},i}(\alpha_1, \alpha_1) - A^{\hat{i},i}(\alpha_1, \alpha_0) \right) \quad i = 1, \dots, N. \end{aligned} \quad (14)$$

## 2.4 Limit sets, periodic orbits and chaos

A differential equation  $\dot{x} = F(x)$  given by a  $C^1$  vector field  $F : \Omega \rightarrow \mathbb{R}^n$  on a domain  $\Omega \subset \mathbb{R}^n$  admits a unique solution on a maximal open interval  $I = (I_l, I_r)$ ,  $I_l, I_r \in \mathbb{R} \cup \{\pm\infty\}$ , denoted by  $x(t) : I \rightarrow \mathbb{R}^n$ , for any initial condition  $x(0) = x_0 \in \Omega$ . Among possible solutions to such equation, we distinguish particular types of solutions defined by their qualitative properties: we say that a solution  $x(t)$  is an *equilibrium* iff  $x(t) = \text{const}$  for all  $t \in I$ . A solution is *periodic* iff  $x(t) = x(t + T)$  for some  $T > 0$  and all  $t \in I$ ; and it is a *connecting orbit* between equilibria  $x_1$  and  $x_2$  (allowing  $x_1 = x_2$ ), iff  $x(t) \rightarrow x_1$  as  $t \rightarrow \infty$  and  $x(t) \rightarrow x_2$  as  $t \rightarrow -\infty$ . A set  $\omega(x_0) \subset \Omega$  is a limit set for an initial condition  $x_0 \in \Omega$ , if  $\forall x \in \omega(x_0)$  there exists an unbounded, increasing sequence  $\{t_n\}_n \subset \mathbb{R}^+$ , such that  $x(t_n) \rightarrow x$ ,  $n \rightarrow \infty$ . Limit sets are *invariant* – they are formed by unions of solutions of the differential equation. They are also *compact* – bounded as subsets of  $\mathbb{R}^n$ , and closed under the limit operation for sequences from itself.

Fundamental research has been devoted to study the properties of solutions within limit sets, as they offer a qualitative description of long-term behavior of the system [24]. Since the discovery of chaotic attractors [36], it has become known that in the general setting, these solutions can have arbitrarily complicated shapes and exhibit seemingly random behavior, a clearly undesirable feature from the point of view of applications; and engineering systems with simple  $\omega$ -limit sets became of particular interest.

DEFINITION 1. *We say that a differential equation  $\dot{x} = F(x)$ ,  $x \in \Omega$  has the Poincaré-Bendixson property iff for all  $x \in \Omega$ , such that the solution  $x(t)$  is bounded, each limit set  $\omega(x)$  is either:*

- an equilibrium;
- a periodic solution;
- a union of equilibria and connecting orbits between these equilibria.

A well known result from the qualitative theory of differential equations shows that planar systems exhibit this trait.

THEOREM 1. *The Poincaré-Bendixson Theorem [7]. Let  $F = F(x)$ ,  $x \in \Omega \subset \mathbb{R}^2$  be a  $C^1$  vector field with finitely many zeroes. Then, the differential equation  $\dot{x} = F(x)$  has the Poincaré-Bendixson property.*

Already in  $\mathbb{R}^3$  there are known examples of systems having complicated, chaotic attractors [36]. However, dimensionality is not the only factor which could determine potential shapes of limit sets. In particular, for certain systems of arbitrary dimension, with structured “previous-neighbor” interactions between the variables, the limit sets can be as simple as in planar systems.

THEOREM 2. *Mallet-Paret & Smith [37]. Let  $(f_i(x_{i-1}, x_i))_{i=1}^n$ , be a  $C^1$  vector field on an open, convex set  $O \subset \mathbb{R}^n$ , and let  $x^0 := x^n$ . Assume that  $\frac{\partial f^i}{\partial x^{i-1}} \neq 0$  for all  $x \in O$ . Then, the system of differential equations*

$$\dot{x}_i = f_i(x_{i-1}, x_i), \quad i = 1, \dots, n, \quad (15)$$

*has the Poincaré-Bendixson property.*

The above theorem is key to proving our further results.

### 3 THE POINCARÉ-BENDIXSON THEOREM FOR GAMES

In this section we state and prove our main results on the topology of limit sets in Follow-the-regularized-Leader learning. We will first state and prove the Poincaré-Bendixson theorem for cyclic games:

**THEOREM 3.** *Let  $\dot{z} = V(z)$  be a system of differential equations given by the vector field (7) – the follow-the-regularized-leader learning dynamics – for a binary, cyclic game. For any smooth, steep, strictly convex collection of regularizers  $\{h_i\}_i$  such system possesses the Poincaré-Bendixson property.*

**PROOF.** Since  $u_i$  depends only on  $Q_i$  and  $Q_{i-1}$ , we have

$$\begin{aligned} V_i(\hat{Q}(z)) &= V_i(\hat{Q}_{i-1}(z_{i-1})) \\ &= v_{i\alpha_0}(Q_{i-1}(z_{i-1}, 0)) - v_{i\alpha_1}(Q_{i-1}(z_{i-1}, 0)). \end{aligned} \quad (16)$$

Our goal is to employ Theorem 2. Therefore, we would like to establish under which conditions

$$\frac{\partial V_i}{\partial z_{i-1}} \neq 0. \quad (17)$$

for all  $i$ . We have:

$$\frac{\partial V_i}{\partial z_{i-1}} = \frac{\partial v_{i\alpha_0}}{\partial x_{i-1}} \frac{\partial x_{i-1}}{\partial z_{i-1}} - \frac{\partial v_{i\alpha_1}}{\partial x_{i-1}} \frac{\partial x_{i-1}}{\partial z_{i-1}}. \quad (18)$$

Moreover, differentiation of mixed strategy payoffs yields

$$\begin{aligned} \frac{\partial v_{i\alpha_1}}{\partial x_{i-1}} - \frac{\partial v_{i\alpha_0}}{\partial x_{i-1}} &= A^{\hat{i},i}(\alpha_0, \alpha_0) - A^{\hat{i},i}(\alpha_1, \alpha_0) \\ &\quad + A^{\hat{i},i}(\alpha_1, \alpha_1) - A^{\hat{i},i}(\alpha_0, \alpha_1). \end{aligned} \quad (19)$$

From Lemma 2 we have  $\frac{\partial x_{i-1}}{\partial z_{i-1}} > 0$ , so the necessary condition to satisfy inequality (17) is:

$$\begin{aligned} &A^{\hat{i},i}(\alpha_0, \alpha_1) + A^{\hat{i},i}(\alpha_1, \alpha_0) \\ &\neq A^{\hat{i},i}(\alpha_0, \alpha_0) + A^{\hat{i},i}(\alpha_1, \alpha_1). \end{aligned} \quad (20)$$

Now let's consider the edge case, where  $A^{\hat{i},i}(\alpha_0, \alpha_1) + A^{\hat{i},i}(\alpha_1, \alpha_0) = A^{\hat{i},i}(\alpha_0, \alpha_0) + A^{\hat{i},i}(\alpha_1, \alpha_1)$  for some  $i$ . Then  $\partial v_{i\alpha_0} / \partial x_{i-1} = \partial v_{i\alpha_1} / \partial x_{i-1}$ . Consequently,  $\partial V_i / \partial z_{i-1} = 0$ , and hence  $i$ -th coordinate of all solutions has the form  $z_i(t) = a_i t + b$ , for some  $a_i, b_i$ . If  $a_i \neq 0$ , then all solutions diverge to infinity. If, however  $a_i = 0$ , then  $z_i(t) = \text{const}$ . Since  $V_{i+1}$  depends only on  $z_i$ , and  $z_{i+1} = a_{i+1} t + b_{i+1}$ ; the argument continues, until all coordinates of solutions are constant, or one coordinate diverges for all solutions.  $\square$

We are now ready to state and prove the theorem for GPGs with nodes of indegree at most one.

**THEOREM 4.** *Let  $\dot{z} = V(z)$  be a system of differential equations given by the follow-the-regularized leader dynamics of a binary, weakly connected, graphical polymatrix game, where each player has up to one predecessor. Then, for any smooth, steep, strictly convex collection of regularizers  $\{h_i\}_i$  (c.f. Lemma 2), such system possesses the Poincaré-Bendixson property.*

Firstly, we state the following lemma on inheritance of the Poincaré Bendixson property when extending dynamical systems.

**LEMMA 3.** *Consider the following  $y$ -augmented system of differential equations*

$$\begin{aligned} \dot{x} &= f(x), \\ \dot{y} &= g(x_i), \\ x &= \{x_1, \dots, x_n\} \in \mathbb{R}^n, y \in \mathbb{R}. \end{aligned} \quad (21)$$

for smooth  $f, g$ . If the original system

$$\dot{x} = f(x) \quad (22)$$

has the Poincaré-Bendixson property, then the augmented system (21) also has the Poincaré-Bendixson property.

**PROOF.** Let  $Z$  be an  $\omega$ -limit set corresponding to some solution  $(x(t), y(t))$  to the system (21). Consider  $X$  – an  $\omega$ -limit set to solution  $x(t)$  of (22).

From invariance of  $\omega$ -limit sets it follows set  $Z$  consists of a union of solutions of (21). For any solution  $\{x^*(t), y^*(t) : t \in \mathbb{R}\} \subset Z$ , we have  $\{x^*(t)\} \subset X$ . By the Poincaré-Bendixson property of the original system, we can distinguish three cases:

- (1)  $x^*(t)$  is an equilibrium of (22),
- (2)  $x^*(t)$  is a periodic orbit of (22),
- (3)  $x^*(t)$  is a connecting orbit of (22) – a part of a cycle of connecting orbits.

In the rest of the proof we will frequently use the integral form of solutions  $y(t)$  to (21), given by  $y(t) = y(0) + \int_0^t g(x_i(s)) ds$ .

Case (1): We will prove that  $(x^*(t), y^*(t))$  is stationary for (21). It is enough to show  $g(x_i^*) = 0$ . Assume otherwise. Then  $y^*(t) = y(0) + \int_0^t g(x_i^*) ds = y(0) + t g(x_i^*) \rightarrow \infty$  as  $t \rightarrow \infty$ . This contradicts the boundedness of an  $\omega$ -limit set.

Case (2) Let  $T$  be the period of  $x^*(t)$ . We will show that  $(x^*(t), y^*(t))$  is a periodic solution of (21) of the same period. We have

$$\begin{aligned} \frac{d}{dt}(y^*(t+T) - y^*(t)) &= \frac{d}{dt} \int_t^{T+t} g(x_i^*(s)) ds \\ &= g(x_i^*(T+t)) - g(x_i^*(t)) \\ &= 0, \end{aligned} \quad (23)$$

hence  $y^*(t+T) - y^*(t) = \text{const}$ . If this quantity would be non-zero, the diameter of the set  $\{y^*(t) : t \in \mathbb{R}\}$  would be infinite. However, the set  $Z$  is bounded, and therefore  $y^*(t+T) = y^*(t)$ .

Case (3): We will show that  $(x^*(t), y^*(t))$  is a connecting orbit between two equilibria for the full system (21). We shall only prove convergence with  $t \rightarrow \infty$ , the very same argument holds for  $t \rightarrow -\infty$  and  $\alpha$ -limit sets. The orbit  $(x^*(t), y^*(t))$  is bounded and therefore it has an accumulation point as  $t \rightarrow \infty$  given by  $(x^{**}, y^{**}) \in \omega(x^*(0), y^*(0))$ . The point  $x^{**}$  is an equilibrium for (22). We will show that  $(x^{**}, y^{**})$  is an equilibrium. It is enough to show that  $g(y^{**}) = 0$ . Assume otherwise. Then  $y^{**}(t) = y^{**} + t g(x_i^{**})$  which is unbounded. However, it is also a part of the  $\omega((x^*(0), y^*(0)))$ , since  $\omega$ -limit sets are invariant. Boundedness of  $\omega((x^*(0), y^*(0)))$  leads to a contradiction. The same process, repeated for all connecting orbits of (22), creates a cycle of connecting orbits for (21).  $\square$

Now, we can proceed to the proof of Theorem 4.

**PROOF.** By Lemma 1 and Remark 1 we know that the graph of the system has either a root vertex or a root cycle. We will first address the case of a root vertex. We will see that this case

is somewhat degenerate. Without loss of generality let us assume that it is labelled as the 1st vertex, and that the other vertices are numbered in order of increasing path distance from vertex 1 (i.e.  $j < i$  implies that the path from 1 to  $j$  is shorter than the path from 1 to  $i$ ) – this is possible by Lemma 1.

The payoffs of the root node are only affected by its own choice of strategy. Therefore, we can write  $\dot{z}_1 = u_1(\alpha_0) - u_1(\alpha_1)$ , and, consequently,  $z_1(t) = t(u_1(\alpha_0) - u_1(\alpha_1)) + z_1(0)$ . This system constitutes an autonomous ODE, which trivially has the Poincaré-Bendixson property (as it is either completely stationary, or is divergent). From then on, we can add nodes, starting from vertices connected to the root vertex, and then continuing in an inductive fashion. Then, either one of the nodes diverges, or they are all stationary, and trivially satisfy the Poincaré-Bendixson property. It should be noted that "divergence" in practice means that  $z_i(t)$ 's approach in the limit  $t \rightarrow \infty$  to either  $\infty$  or  $-\infty$ ; the former implies that the player  $i$  is placing almost all probability mass on strategy  $\alpha_0$ , and the latter – on  $\alpha_1$ .

The more interesting scenario arises for the root cycle, where periodic limit sets are possible. Enumerate these vertices by  $1, \dots, N_0$ , with  $N_0 \leq N$ , and assume that the vertices from  $N_0 + 1$  to  $N$  are arranged in the order of increasing path distance from vertices of the cycle (possible by Remark 1). Observe that the system

$$\begin{aligned} \dot{z}_i &= V_i(z_i), \\ i &= 1, \dots, N_0, \end{aligned} \quad (24)$$

is an autonomous system of differential equations (as there are no edges with successors in  $\{1, \dots, N_0\}$ , and predecessors outside of this set), and forms a binary, cyclic game in the sense of Theorem 3. As such, this subsystem possesses the Poincaré-Bendixson property. From then on, the proof continues similarly as for the root vertex. We add a vertex  $N_0 + 1$  which has an incoming edge from the root cycle, and, by Lemma 3 observe that the system

$$\begin{aligned} \dot{z}_i &= V_i(z_i), \\ i &= 1, \dots, N_0 + 1, \end{aligned} \quad (25)$$

again has the Poincaré-Bendixson property. The proof continues inductively w.r.to the vertices, until we conclude that the full system  $\dot{z} = V(z)$  has the Poincaré-Bendixson property.  $\square$

**REMARK 3.** *Our theorems apply only to fully mixed initial strategy profiles, as FoReL learning (4) is ill-defined for pure strategies. However, when one player has an initial pure strategy, the system can be suitably decomposed, and the Poincaré-Bendixson property still holds. More specifically, in a game where each agent has one predecessor, if agent  $i$  plays a pure strategy, then all the agents  $V_i := \{j : \text{there exists a path from } i \text{ to } j\}$  would eventually sequentially converge under all reasonable learning dynamics (including replicator) to their best response to strategy  $i$ . One can then apply Theorem 4 to the autonomous reduced system  $V \setminus V_i$ , where Equations (4) are again well defined.*

**REMARK 4.** *The assumption of connectedness is needed for the Poincaré-Bendixson property, as (by Lemma 1) it ensures that the graph of interactions has only one cycle. For games with multiple cycles, one can have yet another type of limit behavior. Consider a disjoint union of FoReL systems for two binary graphical games, both possessing the Poincaré-Bendixson property, such that the systems*

*have non-resonant periods of periodic orbits; e.g. one of the systems has a periodic solution of period 1 and the other system has a periodic orbit of period  $\sqrt{2}$ . Such orbits can be easily obtained from replicator dynamics for appropriately scaled mismatched pennies games, c.f. Section 5. Let  $(z_1(0), z_2(0))$  be a point belonging to the periodic solution of period 1, and  $(z_2(0), z_3(0))$  be a point belonging to the periodic solution of period  $\sqrt{2}$ . Then the solution of the full system starting from  $(z_1(0), z_2(0), z_3(0), z_4(0))$  forms a quasi-periodic motion, with the  $\omega$ -limit set of a toroidal shape, c.f. [10].*

## 4 FROM GEOMETRY TO EFFICIENCY: SOCIAL WELFARE ANALYSIS

The following result shows that for cyclic, binary games, under additional but structurally robust assumptions on the payoff matrices the time-average social welfare of our FoReL dynamics is at least as high, as the social welfare  $SW = \sum_i u_i$  of the worst Nash equilibrium. The proof crucially relies on the interplay of the optimal regret properties of FoReL dynamics combined with structural characterizations of the NE set of these games.

**THEOREM 5.** *In any binary, cyclic game with the property that for any agent  $i$ , the payoff entries are distinct and*

$$[A^{i-1,i}(\alpha_0, \alpha_0) - A^{i-1,i}(\alpha_1, \alpha_0)][A^{i-1,i}(\alpha_0, \alpha_1) - A^{i-1,i}(\alpha_1, \alpha_1)] < 0,$$

*the time-average of the social welfare of FoReL dynamics is at least that of the social welfare of the worst Nash equilibrium. Formally,*

$$\liminf \frac{1}{T} \int_0^T \sum_i u_i(x(t)) dt \geq \sum_i u_i(x_{NE}) \quad (26)$$

*where  $x_{NE}$  the worst case Nash equilibrium, i.e., a Nash equilibrium that minimizes the sum of utilities of all agents.*

In other words, the Nash equilibrium is the worst imaginable outcome for all players; and the dynamical, regret minimization approach yields superior payoffs.

**PROOF.** Lets consider the payoff matrix of each agent  $i$ . Recall, that by the cyclicity assumption, there is at most one agent  $k$  such that  $A^{k,i}$  is a non-zero matrix, i.e., the unique predecessor of  $i$ , that for simplicity of notation we call  $i - 1$ . By assumption, the four entries will be considered distinct. Next, we break down the analysis into two cases. As a first case, we consider the scenario where there exists at least one agent with a strictly dominant strategy. The FoReL dynamics of that agents will trivially converge to playing the strictly dominant strategy with probability one. Similarly, all agents reachable from agent  $i$  will similarly best respond to it. This is clearly the unique NE for the binary cyclic game, so in this case the limit behavior of FoReL dynamics exactly corresponds to the unique Nash behavior and the theorem follows immediately.

Next, let's consider the case where no agent has a strictly dominant strategy. In this case, we will construct a specific Nash equilibrium for the cyclic game (although it may have more than one). In this Nash equilibrium every agent  $i - 1$  plays the unique mixed strategy that makes its successor (agent  $i$ ) indifferent between its two strategies. Such a strategy exists for each agent, because otherwise there would exist an agent with a strictly dominant strategy. In fact by the assumption  $[A^{i-1,i}(\alpha_0, \alpha_0) - A^{i-1,i}(\alpha_1, \alpha_0)][A^{i-1,i}(\alpha_0, \alpha_1) -$

$A^{i-1,i}(\alpha_1, \alpha_1) < 0$  such a strategy would be the  $i-1$ st agent's min-max strategy if they participated in a zero-sum game with agent  $i$  defined by the payoff matrix of agent  $i$ . Indeed, this assumption, along with the fact that agent  $i$  does not have a dominant strategy, exactly encodes that the zero-sum game (defined by payoff matrix  $A^{i-1,i}$ ) has an interior Nash. Given its predecessors behavior, agent  $i$  will be receiving exactly its max-min payoff no matter which strategy they select, therefore this strategy profile where each agent  $i-1$  just plays the strategy that makes agent  $i$  indifferent between their two options is a Nash equilibrium, where each agent receives exactly their max-min payoffs. However, by [39] (Lemma C.1), continuous-time FoReL dynamics are no-regret with their time-average regret converging to zero at an optimal rate of  $O(1/T)$ , i.e. there exists an  $\Omega_i > 0$ , such that for all players  $i$  we have:

$$\max_{p_i \in \mathcal{X}_i} \frac{1}{T} \int_0^T (u_i(p_i; x_{-i}(t)) - u_i(x(t))) dt \leq \frac{\Omega_i}{T}. \quad (27)$$

However, the left hand side is greater or equal to

$$u_i(x_{NE}) - \frac{1}{T} \int_0^T u_i(x_i(t)), \quad (28)$$

since the mixed Nash equilibrium consists of max-min strategies. Therefore, the sum over  $i$  of the time-average performance is at least the sum of the max-min utilities minus a quickly vanishing term  $O(1/T)$  and the theorem follows.  $\square$

## 5 EXAMPLES

To illustrate our theoretical findings, we analyze the replicator dynamics (14) of two classes multidimensional binary, cyclic games, which exhibit non-convergence, and therefore non-trivial limit behavior. The goal of the examples is to show that all possible limit sets indicated in the Poincaré-Bendixson property, i.e. an equilibrium, a periodic solution and a cycle of connecting solutions are attainable for systems satisfying our assumptions. In addition, we plot the social welfare of simulated trajectories, relating them to the results of Theorem 5. Finally, we provide a counterexample; a three-dimensional replicator system which violates the assumptions of our theorems, and exhibits chaos. To determine the limit sets, we perform numerical integration of initial value problems for various starting conditions, via the *lsoda* differential equation integrator [26].

### 5.1 Matched-mismatched pennies game

Firstly, we analyze a four dimensional system of matched-mismatched pennies. Each player has a choice of two strategies,  $\alpha_0$  and  $\alpha_1$ . The payoffs for players 0, 2 are given by

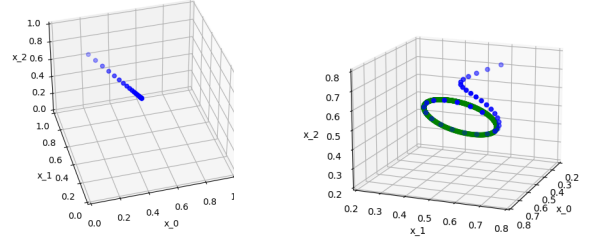
$$A^{3,0} = A^{1,2} = \begin{bmatrix} -1 & 1 \\ 1 & -1 \end{bmatrix} \quad (29)$$

and the payoffs of players 1, 3 are given by

$$A^{0,1} = A^{2,3} = \begin{bmatrix} 1 & -1 \\ -1 & 1 \end{bmatrix}. \quad (30)$$

Simply put, players 0 and 2 will try to mismatch the strategy with players 1 and 3, and players 1 and 3 will try to match them.

The system possesses three Nash equilibria, corresponding to the following strategy profiles:  $(0, 0, 1, 1)$ ,  $(1, 1, 0, 0)$ ,  $(0.5, 0.5, 0.5, 0.5)$ ,



**Figure 2: Limit sets in the matched-mismatched pennies system: an orbit converging to an equilibrium (left) and an orbit converging to a limit cycle (right).**

out of which the pure Nash equilibria are attracting, and the mixed Nash equilibrium has two center directions, one repelling direction, and one attracting direction. We will denote the mixed Nash by  $x^{MNE}$ . Due to symmetry of the system, the plane  $\{(t, s, t, s), t, s \in [0, 1]\}$ , is invariant, and consists purely of periodic orbits, and forms the center manifold to the mixed Nash equilibrium.

The numerical results are in line with Theorems 3, 4. The only limit sets observed through numerical simulations, are the mixed Nash equilibrium  $x^{MNE}$  itself (along a single-dimensional attracting set), and the limit cycles around it, which also appear to be of saddle nature and have a single attracting direction, see Figure 2. Most crucially, more complicated behavior like chaos or invariant tori does not emerge, despite the system being nontrivially embedded in four dimensions.

The mixed Nash yields the minimax payoff vector  $(0, 0, 0, 0)$  for each player, and the social welfare of 0. The payoff matrices satisfy the assumptions of Theorem 5, and the average payoffs along solutions are therefore at least non-negative. In fact, almost all (a set of full measure) initial conditions appear to converge to the pure equilibria at the boundary, their time-average payoffs exceeding the one of Nash equilibrium, and converge to the maximal welfare of 4, see Figure 3.

### 5.2 Asymmetric N-pennies game

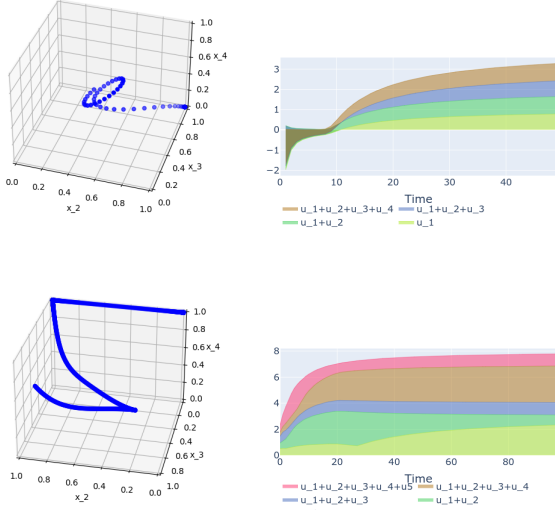
Our second system is a system of  $N$ -player asymmetric mismatched pennies, previously introduced in [32]. There are three players, and again each of them can choose between two strategies,  $\alpha_0$  and  $\alpha_1$ . The payoffs for players  $i$  w.r.t. the player  $i-1$  are given by the matrix

$$A^{i-1,i} = \begin{bmatrix} 0 & 1 \\ p & 0 \end{bmatrix}. \quad (31)$$

with  $p > 0$ .

For odd  $N$  there is no Nash equilibrium in pure strategies. In the replicator system, the pure strategy profiles are saddle-type stationary points of the ODE, linked by connecting orbits of mixed strategies. The system has a unique, mixed Nash equilibrium defined by  $x_i = \frac{1}{p+1}$ ,  $i \in \{1, \dots, N\}$ , where each player attains a payoff of  $\frac{p}{p+1}$ .

The system was thoroughly analyzed in [32], and the main result shown therein was that for  $N = 3$  and  $p > 7$  all mixed strategies



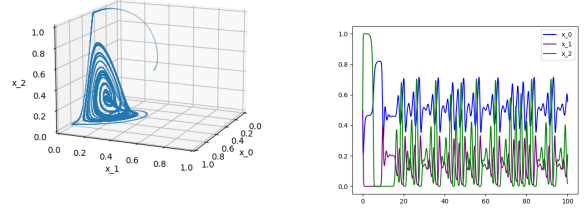
**Figure 3: Time-average payoffs and social welfare of a sample learning trajectory in the matched-mismatched pennies game (top), and in the asymmetric 5-pennies game with  $p = 3$  (bottom, projection onto first three variables).**

except for the diagonal converge to a sequence of orbits connecting boundary stationary points. Moreover, the social welfare attained close to the boundary exceeds the social welfare at the Nash equilibrium. We extend these results. From Theorem 3 we deduce that for all  $N$ , and for all  $p \neq -1$ , the only limit sets in the interior are equilibria, periodic orbits, and cycles of connecting orbits to equilibria. The payoff matrices satisfy the assumptions of Theorem 5, and in particular for all  $p > 0$ , the mixed equilibrium yields the minimax payoff for each player, and time averages of payoffs along other orbits have to exceed the minimax payoffs. We observe that for almost all initial conditions the dynamics is attracted to the boundary cycle of average payoff  $(p + 1) \frac{N-1}{2}$  (e.g. Figure 3), and indeed no chaotic emergent behavior is apparent.

### 5.3 A chaotic polymatrix replicator

Our last system serves as a counterexample; it shows that even in a binary 3-player game, but without structured interactions (i.e. no cyclicity, all possible connections in the game graph), learning trajectories of replicator dynamics can approach complex chaotic limit sets. The payoff matrices are given by

$$\begin{aligned}
 A^{1,1} &= \begin{bmatrix} \mu & 14 \\ 0 & 0 \end{bmatrix}, & A^{2,1} &= -A^{1,2} = \begin{bmatrix} -10 & 10 \\ 0 & 0 \end{bmatrix}, \\
 A^{3,1} &= A^{3,2} = A^{3,3} = -A^{2,2} = \begin{bmatrix} -2 & 2 \\ 0 & 0 \end{bmatrix}, & & \\
 A^{1,3} &= \begin{bmatrix} -25 & 29 \\ 0 & 0 \end{bmatrix}, & A^{2,3} &= \begin{bmatrix} 0 & -11 \\ 0 & 0 \end{bmatrix}.
 \end{aligned} \tag{32}$$



**Figure 4: A learning trajectory approaching a chaotic attractor in the polymatrix replicator [44] (left) and a plot of values of its coordinates (right). The game is characterized by unstructured interactions between payoffs, and therefore breaks the assumptions of Theorems 3, 4.**

After some transformations (for details, see [44]), we arrive at the following one-parameter system of differential equations

$$\begin{aligned}
 \dot{x}_0 &= x_0(1 - x_0)(12 - \mu + (\mu - 14)x_0 - 20x_1 - 4x_2), \\
 \dot{x}_1 &= x_1(1 - x_1)(-10 + 20x_0 + 4x_1 - 4x_2), \\
 \dot{x}_2 &= x_2(1 - x_2)(27 - 54x_0 + 11x_1 - 4x_2),
 \end{aligned} \tag{33}$$

where  $x_i$  represents the probability player  $i$  plays strategy  $\alpha_0$ , and  $1 - x_i$  is the probability it plays  $\alpha_1$ .

This system was recently introduced by Peixe and Rodrigues in [44], where it was formally shown, by a combination of theoretical and numerical results, that the system contains a persistent strange (chaotic) attractor for a range of parameter values  $\mu \in [1.4645, 9.5055]$ . We replicate their findings, by integrating a sample trajectory and observing its approach to the chaotic attractor for  $\mu = 2.8$ , see Figure 4. Due to lack of cyclicity, the game does not enjoy guarantees on payoff structure given by Theorem 5.

## 6 CONCLUSIONS

Numerous recent results in learning in games have established a clear separation between the idealized behavior of equilibration and the erratic, unpredictable and typically chaotic behavior of learning dynamics even in simple games and domains. Although at a first glance, this realization might seem as a set-back, when viewed from the right perspective it opens up a new way of understanding learning dynamics, by examining solution concepts from the topology of dynamical systems. Our results showcase the possibility of establishing links between the topological-combinatorial structure of multi-agent games (e.g. game graph, number of actions) to understand and constrain the topological complexity of game dynamics (Poincaré-Bendixson property), and finally link back to more traditional game theoretic analysis such as calculating the efficiency of the system via social welfare. These connections showcase promising advantages of this approach, and we hope to encourage more work along these lines.



## REFERENCES

- [1] Israel Alvarez and Alexander Poznyak. 2010. Game theory applied to urban traffic control problem. In *International Conference on Control, Automation and Systems*. 2164–2169.
- [2] Stefan Ambec and Yves Sprumont. 2002. Sharing a river. *Journal of Economic Theory* 107, 2 (2002), 453–462.
- [3] Sanjeev Arora, Elad Hazan, and Satyen Kale. 2012. The Multiplicative Weights Update Method: a Meta-Algorithm and Applications. *Theory of Computing* 8, 1 (2012), 121–164.
- [4] James P. Bailey and Georgios Piliouras. 2018. Multiplicative Weights Update in Zero-Sum Games. In *ACM Conference on Economics and Computation*. 321–338.
- [5] James P. Bailey and Georgios Piliouras. 2019. Multi-Agent Learning in Network Zero-Sum Games is a Hamiltonian System. In *18th International Conference on Autonomous Agents and Multiagent Systems*. 233–241.
- [6] Michel Benaïm, Josef Hofbauer, and Sylvain Sorin. 2012. Perturbations of set-valued dynamical systems, with applications to game theory. *Dynamic Games and Applications* 2, 2 (2012), 195–205.
- [7] Ivar Bendixson. 1901. Sur les courbes définies par des équations différentielles. *Acta Mathematica* 24, 1 (1901), 1–88.
- [8] Daan Bloembergen, Karl Tuyls, Daniel Hennes, and Michael Kaisers. 2015. Evolutionary dynamics of multi-agent learning: A survey. *Journal of Artificial Intelligence Research* 53 (2015), 659–697.
- [9] Matthias Blonski. 1999. Anonymous games with binary actions. *Games and Economic Behavior* 28, 2 (1999), 171–180.
- [10] Hendrik W Broer, George B Huitema, and Mikhail B Sevryuk. 2009. *Quasi-periodic motions in families of dynamical systems: order amidst chaos*. Springer.
- [11] Gerard P Cachon and Serguei Netessine. 2006. Game theory in supply chain analysis. *Models, methods, and applications for innovative decision making* (2006), 200–233.
- [12] Nicolo Cesa-Bianchi and Gábor Lugosi. 2006. *Prediction, learning, and games*. Cambridge university press.
- [13] Yun Kuen Cheung and Georgios Piliouras. 2020. Chaos, Extremism and Optimism: Volume Analysis of Learning in Games. In *Advances in Neural Information Processing Systems*, Vol. 33. 9039–9049.
- [14] Yun Kuen Cheung and Georgios Piliouras. 2019. Vortices Instead of Equilibria in MinMax Optimization: Chaos and Butterfly Effects of Online Learning in Zero-Sum Games. In *32nd Annual Conference on Learning Theory*, Vol. 99. 1–28.
- [15] Thiparat Chotitub, Fryderyk Faliński, Michał Misiurewicz, and Georgios Piliouras. 2020. Family of chaotic maps from game theory. *Dynamical Systems* (2020), 1–16.
- [16] Thiparat Chotitub, Fryderyk Faliński, Michał Misiurewicz, and Georgios Piliouras. 2020. The route to chaos in routing games: When is Price of Anarchy too optimistic? 33 (2020), 766–777.
- [17] Lampros Flokas, Emmanouil-Vasileios Vlatakis-Gkaragkounis, and Georgios Piliouras. 2019. Poincaré Recurrence, Cycles and Spurious Equilibria in Gradient-Descent-Ascent for Non-Convex Non-Concave Zero-Sum Games. *Advances in Neural Information Processing Systems*, 10450–10461.
- [18] Seth Frey and Robert L. Goldstone. 2013. Cyclic game dynamics driven by iterated reasoning. *PLoS ONE* 8, 2 (2013), e56416.
- [19] Drew Fudenberg, Fudenberg Drew, David K Levine, and David K Levine. 1998. *The theory of learning in games*. Vol. 2. MIT press.
- [20] Tobias Galla and J Doyne Farmer. 2013. Complex dynamics in learning complicated games. *Proceedings of the National Academy of Sciences* 110, 4 (2013), 1232–1236.
- [21] Nicola Gatti, Fabio Panozzo, and Marcello Restelli. 2013. Efficient evolutionary dynamics with extensive-form games. In *27th AAAI Conference on Artificial Intelligence*.
- [22] Andrea Gaunersdorfer and Josef Hofbauer. 1995. Fictitious play, Shapley polygons, and the replicator equation. *Games and Economic Behavior* 11, 2 (1995), 279–303.
- [23] David Goforth and David Robinson. 2004. *Topology of 2x2 Games*. Routledge.
- [24] Jack K. Hale. 2009. *Ordinary Differential Equations*. Dover Publications.
- [25] Sergiu Hart and Andreu Mas-Colell. 2003. Uncoupled dynamics do not lead to Nash equilibrium. *American Economic Review* 93, 5 (2003), 1830–1836.
- [26] AC Hindmarsh and LR Petzold. 2005. LSODA, ordinary differential equation solver for stiff or non-stiff system. *NEA* (2005).
- [27] Josef Hofbauer and Karl Sigmund. 1998. *Evolutionary Games and Population Dynamics*. Cambridge University Press.
- [28] Joseph T Howson Jr. 1972. Equilibria of polymatrix games. *Management Science* 18, 5-part-1 (1972), 312–318.
- [29] James S Jordan. 1993. Three problems in learning mixed-strategy Nash equilibria. *Games and Economic Behavior* 5, 3 (1993), 368–386.
- [30] Michael Kearns. 2007. Graphical games. *Algorithmic game theory* 3 (2007), 159–180.
- [31] Anna B Khmelnitskaya. 2010. Values for rooted-tree and sink-tree digraph games and sharing a river. *Theory and Decision* 69, 4 (2010), 657–669.
- [32] Robert D Kleinberg, Katrina Ligett, Georgios Piliouras, and Éva Tardos. 2011. Beyond the Nash Equilibrium Barrier. In *Symposium on Innovations in Computer Science*. 125–140.
- [33] Elias Koutsoupias and Christos Papadimitriou. 1999. Worst-case equilibria. In *Annual Symposium on Theoretical Aspects of Computer Science*. 404–413.
- [34] Lior Kuyler, Shimon Whiteson, Bram Bakker, and Nikos Vlassis. 2008. Multiagent Reinforcement Learning for Urban Traffic Control Using Coordination Graphs. In *Machine Learning and Knowledge Discovery in Databases*, Walter Daelemans, Bart Goethals, and Katharina Morik (Eds.). 656–671.
- [35] Marc Lanctot, Vinicius Zambaldi, Audrunas Gruslys, Angeliki Lazaridou, Karl Tuyls, Julien Pérolat, David Silver, and Thore Graepel. 2017. A unified game-theoretic approach to multiagent reinforcement learning. *Advances in Neural Information Processing Systems* (2017), 4193–4206.
- [36] Edward N Lorenz. 1963. Deterministic nonperiodic flow. *Journal of the atmospheric sciences* 20, 2 (1963), 130–141.
- [37] John Mallet-Paret and Hal L Smith. 1990. The Poincaré-Bendixson theorem for monotone cyclic feedback systems. *Journal of Dynamics and Differential Equations* 2, 4 (1990), 367–421.
- [38] Flavio M Menezes and Rohan Pitchford. 2006. Binary games with many players. *Economic Theory* 28, 1 (2006), 125–143.
- [39] Panayotis Mertikopoulos, Christos Papadimitriou, and Georgios Piliouras. 2018. Cycles in adversarial regularized learning. In *29th Annual ACM-SIAM Symposium on Discrete Algorithms*. 2703–2717.
- [40] Paul Muller, Shayegan Omidshafiei, Mark Rowland, Karl Tuyls, Pérolat Julien, Siqui Liu, Daniel Hennes, Luke Marris, Marc Lanctot, Edward Hughes, Zhe Wang, Guy Lever, Nicolas Heess, Thore Graepel, and Remi Munos. 2020. A Generalized Training Approach for Multiagent Learning. In *International Conference on Learning Representations*.
- [41] Sai Ganesh Nagarajan, David Balduzzi, and Georgios Piliouras. 2020. From Chaos to Order: Symmetry and Conservation Laws in Game Dynamics. In *37th International Conference on Machine Learning*, Vol. 119. 7186–7196.
- [42] Sai Ganesh Nagarajan, Sameh Mohamed, and Georgios Piliouras. 2018. Three body problems in evolutionary game dynamics: Convergence, periodicity and limit cycles. In *18th International Conference on Autonomous Agents and Multi-Agent Systems*. 685–693.
- [43] Gerasimos Palaiopoulos, Ioannis Panageas, and Georgios Piliouras. 2017. Multiplicative Weights Update with Constant Step-Size in Congestion Games: Convergence, Limit Cycles and Chaos. In *Advances in Neural Information Processing Systems*.
- [44] Telmo Peixe and Alexandre A Rodrigues. 2021. Persistent Strange attractors in 3D Polymatrix Replicators. *arXiv preprint arXiv:2103.11242* (2021).
- [45] Georgios Piliouras and Jeff S. Shamma. 2014. Optimization Despite Chaos: Convex Relaxations to Complex Limit Sets via Poincaré Recurrence. In *Proceedings of the 2014 Annual ACM-SIAM Symposium on Discrete Algorithms*. 861–873.
- [46] Manfred Plank. 1997. Some qualitative differences between the replicator dynamics of two player and n player games. *Nonlinear Analysis: Theory, Methods & Applications* 30, 3 (1997), 1411–1417.
- [47] Eduardo Rodrigues Gomes and Ryszard Kowalczyk. 2009. Dynamic analysis of multiagent Q-learning with  $\epsilon$ -greedy exploration. In *26th Annual International Conference on Machine Learning*. 369–376.
- [48] Tim Roughgarden. 2009. Intrinsic robustness of the price of anarchy. In *Proceedings of the forty-first annual ACM symposium on Theory of computing*. 513–522.
- [49] Tim Roughgarden. 2016. *Twenty lectures on algorithmic game theory*. Cambridge University Press.
- [50] James BT Sanders, J Doyne Farmer, and Tobias Galla. 2018. The prevalence of chaotic dynamics in games with many players. *Scientific reports* 8, 1 (2018), 1–13.
- [51] William H Sandholm. 2010. *Population games and evolutionary dynamics*. MIT Press.
- [52] Yuzuru Sato, Eizo Akiyama, and J Doyne Farmer. 2002. Chaos in learning a simple two-person game. *Proceedings of the National Academy of Sciences* 99, 7 (2002), 4748–4751.
- [53] Shai Shalev-Shwartz et al. 2011. Online learning and online convex optimization. *Foundations and trends in Machine Learning* 4, 2 (2011), 107–194.
- [54] Yoav Shoham and Kevin Leyton-Brown. 2008. *Multiagent systems: Algorithmic, game-theoretic, and logical foundations*. Cambridge University Press.
- [55] Peter D Taylor and Leo B Jonker. 1978. Evolutionary stable strategies and game dynamics. *Mathematical biosciences* 40, 1-2 (1978), 145–156.
- [56] Karl Tuyls and Ann Nowé. 2005. Evolutionary game theory and multi-agent reinforcement learning. *The Knowledge Engineering Review* 20, 1 (2005), 63–90.
- [57] Sebastian van Strien. 2011. Hamiltonian flows with random-walk behaviour originating from zero-sum games and fictitious play. *Nonlinearity* 24, 6 (2011), 1715.
- [58] Michael P Wellman. 2006. Methods for empirical game-theoretic analysis. In *Proceedings of the AAAI Conference on Artificial Intelligence*. 1552–1556.
- [59] Elena B Yanovskaya. 1968. Equilibrium points in polymatrix games. *Litovskii Matematicheskii Sbornik* 8 (1968), 381–384.
- [60] H Peyton Young. 2004. *Strategic learning and its limits*. Oxford University Press.
- [61] Sixie Yu, Kai Zhou, Jeffrey Brantingham, and Yevgeniy Vorobeychik. 2020. Computing Equilibria in Binary Networked Public Goods Games. In *Proceedings of the AAAI Conference on Artificial Intelligence*, Vol. 34. 2310–2317.