Non-chaotic limit sets in multi-agent learning

Czechowski, Aleksander; Piliouras, Georgios

DOI
10.1007/s10458-023-09612-x

Publication date
2023

Document Version
Final published version

Published in
Autonomous Agents and Multi-Agent Systems

Autonomous Agents and Multi-Agent Systems (2023) 37:29
https://doi.org/10.1007/s10458-023-09612-x

Non-chaotic limit sets in multi-agent learning

Aleksander Czechowski[1] · **Georgios Piliouras**[2]

Accepted: 25 June 2023

Abstract

Non-convergence is an inherent aspect of adaptive multi-agent systems, and even basic learning models, such as the replicator dynamics, are not guaranteed to equilibriate. Limit cycles, and even more complicated chaotic sets are in fact possible even in rather simple games, including variants of the Rock-Paper-Scissors game. A key challenge of multi-agent learning theory lies in characterization of these limit sets, based on qualitative features of the underlying game. Although chaotic behavior in learning dynamics can be precluded by the celebrated Poincaré–Bendixson theorem, it is only applicable directly to low-dimensional settings. In this work, we attempt to find other characteristics of a game that can force regularity in the limit sets of learning. We show that behavior consistent with the Poincaré–Bendixson theorem (limit cycles, but no chaotic attractor) follows purely from the topological structure of interactions, even for high-dimensional settings with an 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 binary games characterized interaction graphs where the payoffs of each player are only affected by one other player (i.e., interaction graphs of indegree one). Moreover, for cyclic games we provide further insight into the planar structure of limit sets, and in particular limit cycles. We propose simple conditions under which learning comes with efficiency guarantees, implying that FoReL learning achieves time-averaged sum of payoffs at least as good as that of a Nash equilibrium, thereby connecting the topology of the dynamics to social-welfare analysis.

Keywords Replicator dynamics · Follow-the-regularized leader · Polymatrix games · Poincaré–Bendixson theorem · Regret minimization

✉ Aleksander Czechowski
acze.uni@gmail.com

Georgios Piliouras
georgios.piliouras@gmail.com

[1] Delft University of Technology, Delft, The Netherlands

[2] Singapore University of Technology and Design, Singapore, Singapore

Published online: 13 July 2023

1 Introduction

Characterizing the convergence and limit behavior of learning is vital for understanding the long-term outcomes in multi-agent systems. Much of the research in this direction has been driven by methods of dynamical systems, and in particular evolutionary game theory [1–7]. Even in simple games, such as Rock-Paper-Scissors [8, 9], models of evolution and learning are not guaranteed to converge; beyond cycles, long-term behavior can lead to chaotic behavior, known to the dynamical systems community from, e.g., weather models [10]. Not only does chaos manifest itself even in simple games with two players, but moreover, a string of recent results suggests that such chaotic, unpredictable behavior may indeed be the norm across a variety of simple low-dimensional game dynamics [11–20]. Importantly, these results are persistent even for the well-known class of Follow-the-Regularized-leader (FoReL) dynamics [21, 22], despite the fact that FoReL dynamics include some of the most widely studied learning dynamics such as replicator dynamics [23, 24], which is the continuous-time analogue of the Multiplicative Weights Update meta-algorithm [25], well known for its optimal regret properties. Finally, the emergence of chaotic behavior has been connected with increased social inefficiency, which shows that chaotic dynamics can lead to highly inefficient outcomes [26, 27]. Such profoundly negative results raise the following questions:

- Do simple, robust conditions exist under which learning behaves well?
- Which types of games lie at the “edge of chaos”?
- Does dynamic simplicity translate to high-efficiency and 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 [5, 28–30] and references therein. Few results span over to general sum games and games of arbitrary payoff structures; however, such general approaches are arguably essential in modern research on multi-agent learning. Such payoffs can however occur naturally when stochastic extensive form games are used to create empirical normal form games, by averaging payoffs from simulations for combinations of strategies [31–33]. They can also appear in many real-world applications, such as, e.g., modeling the 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 a regularity of sorts in the learning outcomes of the multi-agent system. In particular, the famous Poincaré–Bendixson theorem (Theorem 1) ensures that two-dimensional continuous learning and adaptation dynamics never form truly chaotic outcomes. However, this comes at a cost: although no specific payoff structure is needed, the underlying learning dynamics must be at most two dimensional.

In this work, rather than by making assumptions on the reward structure or on the dimensionality, 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 [34–36], where players have two strategies. Secondly, we assume that every player can be affected by the behavior of up to one other player. Finally, we add a technical restriction that the game is connected, meaning that it cannot be decomposed into two subgames that are completely independent of each other. Such games encompass, among others, all 2x2 games [37], Jordan’s game [38–40], and easily identifiable subclasses of real-world systems where the graph structure is evident, such as certain traffic networks [41, 42], supply chains [43], or problems of water allocation in deltas [44, 45]. Under these assumptions, we prove in Sect. 3 our main contribution in the form of Theorems 3 and 4, which say 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 Sect. 4, we show that the projections of limit sets onto any pair of consecutive variables preserve their structure, and, in particular, that the projections of limit cycles are Jordan curves. Furthermore, in Sect. 5, under additional but structurally robust assumptions on the payoff matrices (i.e., assumptions that remain valid after small perturbations of the payoff matrices and so are suitable, for example, for empirical payoff matrices), we derive positive results about the efficiency of the time-averaged behavior of the dynamics regardless of whether they are convergent. As is typically the case in the price of anarchy (PoA) literature [46], we focus on the measure of *social welfare*, which is the sum of individual payoffs. Whereas the typical PoA literature argues that regret-minimizing dynamics (such as FoReL) are at most a constant factor worse than the behavior of the worst-case Nash equilibrium [27, 47], we instead show that FoReL dynamics are always at least as efficient as the worst-case Nash equilibrium. Finally, Sect. 6 provides examples of games satisfying our assumptions and their possible limit behavior, as well as two counterexamples of binary games which break assumptions on network topology and induce more complicated limit sets: invariant tori, and chaos.

This article is an extended and modified version of our earlier conference paper, *Poincaré–Bendixson Limit Sets in Multi-Agent Learning*, which appeared at AAMAS 2022 [48]. In this version we have added a new section with theoretical results on two-dimensional projections of limit sets in binary, cyclic games, and provided a new example on how additional connectivity in a network of pennies (i.e. a type of topological connection breaking our assumptions), can result in complicated limit sets in form of invariant tori. We have also added a discussion section, where we identify main limitations of our results, and the most promising directions of future research.

1.1 Related work

Firstly, we would like to highlight several papers containing examples of simple FoReL systems with chaotic dynamics. These show that the assumptions of binary actions and previous-neighbor interaction are essential, and similar results would not be possible for broader classes of games. Going beyond binary games, we have the chaotic example of Sato et al. [8] with two-players and three-actions. Furthermore [49, 50] provide further chaotic/complex attractor examples in n-player binary games without structured (e.g. previous neighbor) interactions. From this perspective, our results establish a maximal class of games for which such regularity results on limit sets are possible.

Research that considers non-convergence but focuses on non-chaoticity is scarce. In the closest works to ours, [51–53], 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, and in particular no chaotic attractor is possible. However, they do so by assuming low dimensionality (three-player limit) or a nongeneric structure on the set of allowable games, which allows for dimensionality reduction (i.e., a network of 2x2 zero-sum, or coordination games). In terms of connections between cyclic behavior and the efficiency of learning dynamics, [54] shows that, for a family of three players, two strategy games with a cyclic attractor can result in social welfare (sum of payoffs) that can be better than the Nash equilibrium payoff; however, the result is specific to this particular example.

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 $u_i : \prod_i \mathcal{A}_i \to \mathbb{R}$. To model the behavior at scale or probabilistic strategy choices, one assumes that players use *mixed strategies*, namely, probability distributions $(x_{i\alpha_i})_{\alpha_i \in \mathcal{A}_i} \in \Delta(\mathcal{A}_i) =: \mathcal{X}_i$. With a slight abuse of notation, the expected payoff of player i in the profile $(x_{i\alpha_i})_{i,\alpha_i}$ is denoted u_i and given by

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

A mixed strategy $\hat{x}$ is a *Nash equilibrium* iff $\forall\, i$ and $\forall x : x_j = \hat{x}_j,\, j \neq i$ 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 $x_{-i} := (x_j)_{j \neq i}$. This is the smallest possible value that player i can be forced to attain by other players, without them knowing the strategy of player i. 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) [55–57]. 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 the predecessor, and the second is called the 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 the strategies of player k, and columns enumerating the 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

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

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

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

A situation where both the successor k and the predecessor i obtain a reward can be modeled 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 considered. This is a weaker condition than strong connectedness, where each pair of vertices must be connected by a *path* (i.e., a sequence of edges together with associated vertices, where the successor in one edge is the predecessor in the next). The *indegree* of a vertex is the number of edges for which the vertex is the successor (i.e., the number of predecessors). The *outdegree* is the number of edges for which the vertex is the predecessor (i.e., the number of successors). A *cycle* is a path where the predecessor in the first edge is the successor in the last edge. For our exposition we 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. First, *cyclic* games, where the interaction between the players forms a cycle, where each player 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 label the nodes of such N-player games by natural numbers $i = 0, 1, \dots, n$ and use the convention that node i is the successor to node $i - 1$, and we can extend the indexing to all integers, in a way that indices congruent modulo n identify the same node.
2. Second, a more general class of graphical games, where each player's payoffs depend on at most one other player (i.e., the indegree of each vertex is at most one). For a vertex $i \in \mathcal{V}$, we denote the predecessor vertex by $\hat{\imath}$, if it exists. For cyclic games we have $\hat{\imath} = i - 1$.

Below, we state and prove a simple lemma that characterizes the one-predecessor assumption in terms of graph topology and clarifies the relation between cyclic and indegree-one graphs (cf. 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 at most one cycle. If the graph has no cycle, then it has 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, we 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 must 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 the cycle of a_2. Then a_0 has two predecessors, which leads to a contradiction.

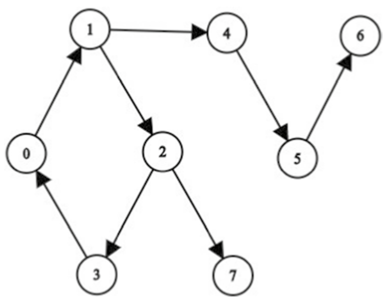

Fig. 1 A weakly connected graph where each vertex is at most of indegree one

For the second part of the lemma we argue as follows. If any vertex has a sequence of predecessors that does not form a cycle, and does not have a root node, then by backtracking through the predecessors we could identify an infinite collection of distinct vertices. Therefore, there must be at least one root node for each vertex. The path from such a root node to the given vertex must be unique, otherwise one could identify a vertex along the path with two predecessors. Finally, it is impossible to have two distinct root nodes, as connectedness imposes that there would have to exist a node with two predecessors between them. □

Remark 1 Under the assumptions of Lemma 1, if the graph has a cycle, then the cycle enjoys properties similar to those of a root node: no paths go from outside the cycle to the cycle (otherwise one vertex in the cycle would have two predecessors), and all vertices outside the cycle must be connected by a path from one of the vertices of the cycle (a unique path, up to the starting point within the cycle). Later, we shall refer to such cycle as the *root cycle*.

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) [5, 6]. This class encompasses a variety of models ranging from gradient descent, to replicator dynamics, and allows for natural description of 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

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

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

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

In the above $h_i : \mathcal{X}_i \to \mathbb{R} \cup \{-\infty, \infty\}$ 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 player has only two strategies at his disposal, say α_0, α_1. The variable x_i denotes then the proportion of time player i plays strategy α_0, and the proportion of α_1 is given by $1 - x_i$. Following [21], 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 α_0 and α_1. It is intuitively clear, and it was proved formally e.g. in [21] 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

$$\dot{z} = V(z), \tag{6}$$

given coordinate-wise by

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

for details again see [21].

Remark 2 An intuitively obvious, but technically important observation is that evolution of ith 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

$$\dot{z}_i = V_i(z_{\hat{i}}) = v_{i\alpha_0}(\hat{Q}_{\hat{i}}(z_{\hat{i}})) - v_{i\alpha_1}(\hat{Q}_{\hat{i}}(z_{\hat{i}})). \tag{8}$$

As previously hinted, for equation (7) to be well-posed, we need to enforce certain 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 $\mathcal{X}_i = [0, 1]$).

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. $h_i'(x_i) \to -\infty$ *as* $x_i \to 0$ *and* $h_i'(x_i) \to \infty$ *as* $x_i \to 1$ (*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

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

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

$$z_i = h_i'(x_i). \tag{10}$$

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

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

which also implies that the function $\hat{Q}_i$ is C^1 on its domain. □

Perhaps the best known example of a FoReL learning system are the *replicator equations* [23], where the regularizer is given by

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

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

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

which, via (10) and (12), translates to the following system in original (x) coordinates:

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

2.4 Limit sets, periodic orbits and chaos

A differential equation $\dot{x} = F(x)$ given by a C^1 vector field $F : \Omega \to \mathbb{R}^n$ 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 $x(t) : I \to \mathbb{R}^n$, 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 $x(t) \to x_1$ as $t \to \infty$ and $x(t) \to x_2$ as $t \to -\infty$. A set $\omega(x_0) \subset \Omega$ is an ω-limit set (sometimes also referred to as 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 $x(t_n) \to x$, $n \to \infty$. Limit sets are *invariant*: this means that they are formed by unions of solutions of the differential equation on maximal intervals (or, in other words, any solution intersecting one has to be contained in it). For bounded orbits $x(t)$, they are also *compact* – bounded as subsets of $\mathbb{R}^n$, and closed under the limit operation on 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 [58]. Since the discovery of chaotic attractors [10], 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 ω-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 $x_0 \in \Omega$, such that the solution $x(t)$ satisfying $x(0) = x^*$ is bounded, each limit set $\omega(x^*)$ such that $\omega(x^*) \subset \Omega$ 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* [59]. *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 [10]. 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 as simple as in planar systems. In what follows, we denote by $\Pi_i(x) = (x_{i-1}, x_i)$ a planar projection onto two consecutive variables.

Theorem 2 *Mallet-Paret & Smith* [60]. *Let* $x = (x_1, \dots, x_n)$, $(f_i(x_{i-1}, x_i))_{i=1}^n$, *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*

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

has the Poincaré–Bendixson property. In addition, for initial conditions $x^* \in O$, *such that the solution* $x(t)$ *satisfying* $x(0) = x^*$ *is bounded, the planar projections of the limit set*

$$\Pi_i : \omega(x^*) \to \mathbb{R}^2 \tag{16}$$

are one-to-one for all $i \in 1, \dots, n$.

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

$$\begin{aligned} V_i(\hat{Q}(z)) &= V_i(\hat{Q}_{i-1}(z_{i-1})) \\ &= v_{i\alpha_0}(Q_{i-1}(z_{i-1}, 0)) - v_{i\alpha_1}(Q_{i-1}(z_{i-1}, 0)). \end{aligned} \tag{17}$$

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

$$\frac{\partial V_i}{\partial z_{i-1}} \neq 0. \tag{18}$$

for all i. We have:

$$\frac{\partial V_i}{\partial z_{i-1}} = \frac{\partial v_{i\alpha_0}}{\partial x_{i-1}} \frac{\partial x_{i-1}}{\partial z_{i-1}} - \frac{\partial v_{i\alpha_1}}{\partial x_{i-1}} \frac{\partial x_{i-1}}{\partial z_{i-1}}. \tag{19}$$

Moreover, differentiation of mixed strategy payoffs yields

$$\begin{aligned} \frac{\partial v_{i\alpha_1}}{\partial x_{i-1}} - \frac{\partial v_{i\alpha_0}}{\partial x_{i-1}} =& A^{\hat{i},i}(\alpha_0, \alpha_0) - A^{\hat{i},i}(\alpha_1, \alpha_0) \\ &+ A^{\hat{i},i}(\alpha_1, \alpha_1) - A^{\hat{i},i}(\alpha_0, \alpha_1). \end{aligned} \tag{20}$$

From Lemma 2 we have $\frac{\partial x_{i-1}}{\partial z_{i-1}} > 0$, so the necessary condition to satisfy inequality (18) is:

$$A^{\hat{i},i}(\alpha_0, \alpha_1) + A^{\hat{i},i}(\alpha_1, \alpha_0) \neq A^{\hat{i},i}(\alpha_0, \alpha_0) + A^{\hat{i},i}(\alpha_1, \alpha_1). \tag{21}$$

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 $z_i(t) = a_i t + b_i$, for some a_i, b_i. If $a_i \neq 0$, then all solutions diverge to infinity. If, however $a_i = 0$, then $z_i(t) = const$. Since V_{i+1} depends only on z_i, we have $z_{i+1}(t) = a_{i+1} t + b_{i+1}$; the argument continues, until all coordinates of solutions are constant, or one coordinate diverges for all solutions. □

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 $\{h_i\}_i$, such system possesses the Poincaré–Bendixson property.*

First, we state the following lemma on inheritance of the Poincaré Bendixson property for augmented systems.

Lemma 3 *Consider the following y-augmented system of differential equations*

$$\begin{aligned} \dot{x} &= f(x), \\ \dot{y} &= g(x_i), \\ x &= \{x_1, \dots, x_n\} \in \mathbb{R}^n,\ y \in \mathbb{R}. \end{aligned} \tag{22}$$

for smooth f, g. If the original system

$$\dot{x} = f(x) \tag{23}$$

has the Poincaré–Bendixson property, then the augmented system (22) also has the Poincaré–Bendixson property.

Proof Let Z be an ω-limit set corresponding to some solution $(x(t), y(t))$ to the system (22). Consider X – an ω-limit set to solution $x(t)$ of (23).

From invariance of ω-limit sets it follows set Z consists of a union of solutions of (22). 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 (23),
2. $x^*(t)$ is a periodic orbit of (23),
3. $x^*(t)$ is a connecting orbit of (23) – 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 (22), given by $y(t) = y(0) + \int_0^t g(x_i(s))ds$.

Case (1): We prove that $(x^*(t), y^*(t))$ is stationary for (22). It is enough to show $g(x_i^*) = 0$. Assume otherwise. Then $\|y^*(t)\| = \|y(0) + \int_0^t g(x_i^*)ds\| = \|y(0) + tg(x_i^*)\| \to \infty$ as $t \to \pm\infty$. This contradicts the boundedness of an ω-limit set.

Case (2) Let T be the period of $x^*(t)$. We show that $(x^*(t), y^*(t))$ is a periodic solution of (22) of the same period. We have:

$$\begin{aligned} \frac{d}{dt}(y^*(t+T) - y^*(t)) &= \frac{d}{dt}\int_t^{T+t} g(x_i^*(s))ds \\ &= g(x_i^*(T+t)) - g(x_i^*(t)) \\ &= 0, \end{aligned} \tag{24}$$

hence $y^*(t+T) - y^*(t) = const$. 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 show that $(x^*(t), y^*(t))$ is a connecting orbit between two equilibria for the full system (22). We shall only prove convergence with $t \to \infty$, the very same argument holds for $t \to -\infty$ and α-limit sets. The orbit $(x^*(t), y^*(t))$ is bounded and therefore it has an accumulation point as $t \to \infty$ given by $(x^{**}, y^{**}) \in \omega(x^*(0), y^*(0))$. The point x^{**} is an equilibrium for (23). We will show that (x^{**}, y^{**}) is an equilibrium. It is enough to show that $g(y^{**}) = 0$. Assume otherwise. Then $y^{**}(t) = y^{**} + tg(x_i^{**})$ which is unbounded. However, it is also a part of $\omega((x^*(0), y^*(0)))$, since ω-limit sets are invariant. Boundedness of $\omega((x^*(0), y^*(0)))$ leads to a contradiction. The same process, repeated for all connecting orbits of (23), creates a cycle of connecting orbits for (22). □

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 $t \to \infty$ to either ∞ or $-\infty$; the former implies that the player i is placing almost all probability mass on strategy α_0, and the latter – on α_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 $N_0 \le N$, 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

$$\begin{aligned} \dot{z}_i &= V_i(z_{\hat{i}}), \\ i &= 1, \dots, N_0, \end{aligned} \tag{25}$$

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

$$\begin{aligned} \dot{z}_i &= V_i(z_{\hat{i}}), \\ i &= 1, \dots, N_0 + 1, \end{aligned} \tag{26}$$

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. □

Remark 3 Theorems 3, 4 apply to dynamics of fully mixed initial strategy profiles bounded away from pure strategies, as FoReL learning (4) is ill-defined for pure strategies. For some learning models such as as the replicator equations (14) the theorems can be applied to subsystems arising when certain players assume a pure strategy profile, as in these models pure strategy profiles define invariant learning spaces.

4 Projections of limit sets in cyclic games

For cyclic, binary games, we can derive even stronger results on the topological structure of bounded limit sets, and, in particular, limit cycles. By the second part of Theorem 2, we have a one-to-one correspondence between any such limit set, and its projection on any pair of consecutive coordinates. This is formalized via the theorem below.

Theorem 5 *Let $\dot{z} = V(z)$ be a system of differential equations given by the FoReL field (7), for a binary, cyclic game, with smooth, steep, and strictly convex regularizers $\{h_i\}_i$. Then, for bounded solutions $x(t)$ with $x(0) = x^*$, the planar projections $\Pi_i : \omega(x^*) \to \mathbb{R}^2$*

are one-to-one, for all $i \in 1, \dots, n$.

Proof We consider the following two cases. Firstly, by the proof of Theorem 3, if for all $i \in 1, \dots n$ we have

$$A^{\hat{i},i}(\alpha_0, \alpha_1) + A^{\hat{i},i}(\alpha_1, \alpha_0) \neq A^{\hat{i},i}(\alpha_0, \alpha_0) + A^{\hat{i},i}(\alpha_1, \alpha_1), \tag{27}$$

then the vector field (7) satisfies the assumptions of Theorem 2, and the assertion follows. Secondly, if for any i it holds that $A^{\hat{i},i}(\alpha_0, \alpha_1) + A^{\hat{i},i}(\alpha_1, \alpha_0) = A^{\hat{i},i}(\alpha_0, \alpha_0) + A^{\hat{i},i}(\alpha_1, \alpha_1)$, then $\partial V_i / \partial z_{i-1} = 0$, and the i-th coordinate of the solutions $z_i(t)$ is of the form $z_i(t) = a_i t + b_i$ for some real a_i, b_i. If $a_i \neq 0$, then all solutions diverge to infinity. If all a_i is zero for all i, then, by the same argument as in the proof $\partial V_{i+1} / \partial z_i = 0$, and the argument continues iteratively in the same fashion as in the proof of Theorem 3, until any of the coordinates is divergent, or all coordinates of all solutions are constant. In the former case, there are no limit sets, and in the latter, they are all singletons, so their projections are trivially continuous and injective. □

We recall that a Jordan curve is a continuous, injective image of a circle in $\mathbb{R}^2$. By the Jordan curve theorem, any Jordan curve divides $\mathbb{R}^2$ into two connected components, and forms their common boundary [61]. One of the consequences of theorem 5, is that projections of limit cycles onto consecutive variables carry over this simple, topological property.

Corollary 1 *Let $x^*(t)$ be a limit set formed by a non-periodic solution of the FoReL system (7) for a binary, cyclic game, with smooth, steep, and strictly convex regularizers. Then, the projections $\{\Pi_i(x(t)),\ t \in \mathbb{R}\}$ are Jordan curves.*

Proof Let $T > 0$ be the minimal period of x^*. Then $\mathbb{R}/\mathbb{Z} \in s \to \Pi_i x^*(sT)$ is a continuous, injective map of a circle in $\Pi_i \mathbb{R}^n = \mathbb{R}^2$. □

Remark 4 An analogous result, on projections on consecutive two variables, cannot be stated for general GPGs of indegree one. Consider a following three player game, where players 0 and 1 play matching-mismatching pennies, and in addition player 2 receives a uniform payoff only when player 1 plays one of his two strategies (regardless of what they do).

$$A^{0,1} = -A^{1,0} = \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}, \quad A^{1,2} = \begin{bmatrix} 1 & 0 \\ 1 & 0 \end{bmatrix}. \tag{28}$$

The learning dynamics for players 0 and 1 are formed by periodic solutions oscillating around the (0, 0) equilibrium, and the dynamics for player 2 are stationary, since there is no gain in changing strategies. The cross-product of any periodic solution with the stationary point is still a periodic solution, however, its projection onto the plane spanned by strategies of players 1 and 2 is a line segment.

5 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 (i.e., assumptions that remain valid after small perturbations of 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 set of Nash equilibria of these games.

Theorem 6 *In any binary, cyclic game with the property that for any player i, the payoff entries are distinct and*

$$[A^{i-1,i}(\alpha_0,\alpha_0)-A^{i-1,i}(\alpha_1,\alpha_0)][A^{i-1,i}(\alpha_0,\alpha_1)-A^{i-1,i}(\alpha_1,\alpha_1)]<0, \tag{29}$$

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

$$\liminf_{T\to\infty}\frac{1}{T}\int_0^T\sum_i u_i(x(t))dt \geq \sum_i u_i(x_{NE}), \tag{30}$$

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

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 player i. Recall, that by the cyclicity assumption, there is at most one player 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 player with a strictly dominant strategy. The FoReL dynamics of that player strategy profile will trivially converge to playing the strictly dominant strategy with probability one. Similarly, all players reachable from player 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 player has a strictly dominant strategy. In this case, we will construct a very specific Nash equilibrium for the cyclic game. In this Nash equilibrium every player $i-1$ plays the unique mixed strategy that makes its successor (player i) indifferent between its two strategies. Such a strategy follows from (29), and from the binary structure of the game. By (29), if players $i-1$, i participated in a zero-sum game defined by the payoff matrix of player i, then player $i-1$ would have no dominant pure strategy; however, since it would be a 2x2 game, they would have a min-max mixed strategy. This, along with the fact that player i does not have a dominant strategy, exactly encodes that this zero-sum game has an interior Nash. In such point, the player i will be receiving exactly its max-min payoff no matter which strategy they select, therefore the profile where each player $i-1$ just plays the strategy that makes player i indifferent between their two options is a Nash equilibrium for the full (cyclic) game, where each player receives exactly their max-min payoffs. However, by [21] (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:

$$\max_{p_i\in\mathcal{X}_i}\frac{1}{T}\int_0^T\left(u_i(p_i;x_{-i}(t))-u_i(x(t)\right)dt \leq \frac{\Omega_i}{T}. \tag{31}$$

However, the left hand side is greater or equal to

$$u_i(x_{NE}) - \frac{1}{T}\int_0^T u_i(x_i(t))), \tag{32}$$

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. □

We contrast the above result with the existing bounds on cost/social welfare of no-regret learners in smooth games e.g. in [47, 62] or [63], which give limits to how much at most it can deteriorate away from equilibrium. What we have demonstrated, is that in a wide class of learning games, their mixed Nash equilibrium is outperformed (in terms of social welfare) by the average welfare accumulated along the learning trajectories. A result similar to ours was proved in [54], for a specific game (asymmetric cyclic matching pennies).

6 Examples

To illustrate our theoretical results, we analyze the replicator dynamics (14) of two classes multidimensional binary cyclic games that 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 6. Finally, we provide two counterexamples, where additional connectivity in the payoff graph induces more complicated limit sets: invariant tori and chaos. To determine the limit sets, we numerically integrate the initial-value problems with various starting conditions via the *lsoda* differential equation integrator [64].

6.1 Matched–mismatched pennies game

First, we analyze a four-dimensional game of matched–mismatched pennies. Each player has a choice of two strategies, α_0 and α_1. The payoffs for players 0 and 2 are given by

$$A^{3,0} = A^{1,2} = \begin{bmatrix} -1 & 1 \\ 1 & -1 \end{bmatrix} \tag{33}$$

and the payoffs for players 1 and 3 are given by

$$A^{0,1} = A^{2,3} = \begin{bmatrix} 1 & -1 \\ -1 & 1 \end{bmatrix}. \tag{34}$$

Simply put, players 0 and 2 try to mismatch the strategy with players 1 and 3, and players 1 and 3 try to match them. The replicator equations are given by

$$\begin{aligned} \dot{x}_0 &= x_0(1-x_0)(2-4x_3), \\ \dot{x}_1 &= x_1(1-x_1)(4x_0-2), \\ \dot{x}_2 &= x_2(1-x_2)(2-4x_1), \\ \dot{x}_3 &= x_3(1-x_3)(4x_2-2). \end{aligned} \tag{35}$$

The system possesses three Nash equilibria, which correspond to the following strategy profiles: (0, 0, 1, 1), (1, 1, 0, 0), (0.5, 0.5, 0.5, 0.5), out of which the pure Nash equilibria are attracting, and the mixed Nash equilibrium has two center directions: one repelling and one attracting. We denote the mixed Nash equilibrium by x_{MNE}. Given the symmetry of the system, the plane $\{(t, s, t, s),\ t, s \in [0, 1]\}$ is invariant, consists purely of periodic orbits, and forms the center manifold to the mixed Nash equilibrium.

The numerical results are consistent with Theorems 3 and 4. The only limit sets observed by the 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 Fig. 2. Most crucially, more complicated behavior, such as chaos or invariant tori, does not emerge, despite the system being nontrivially embedded in four dimensions.

The mixed Nash equilibrium 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 6, 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, with their time-average payoffs exceeding that of the Nash equilibrium and converging to the maximal welfare of 4, see Fig. 3.

6.2 Asymmetric N-penny game

Our second system is a cyclic system of N-player asymmetric mismatched pennies, previously introduced in [54]. The payoffs for player i with respect to player $i-1$ (keeping the convention that $x_0 = x_N$) are given by the matrix

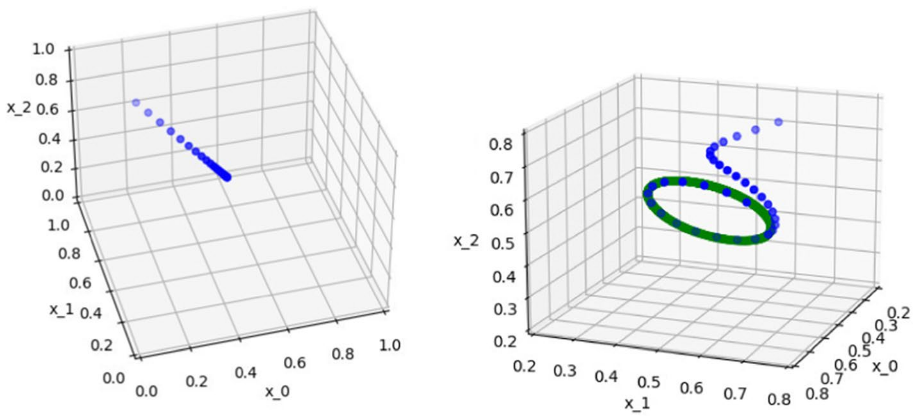

Fig. 2 Limit sets in the matched–mismatched pennies system: an orbit converging to an equilibrium (left) and an orbit converging to a limit cycle (right)

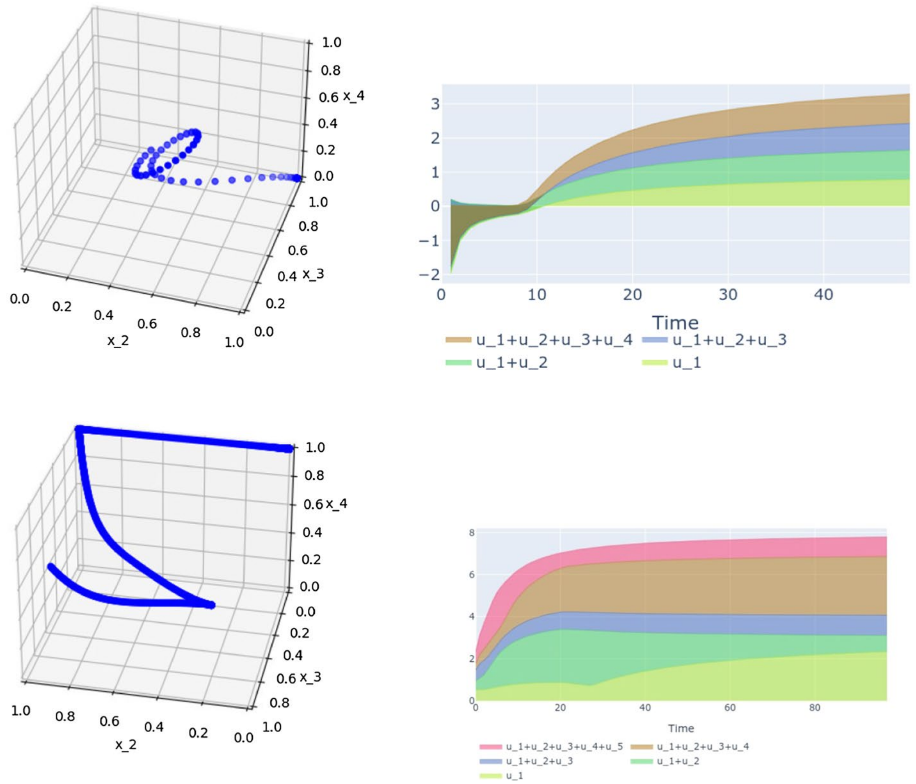

Fig. 3 Time-average payoffs and social welfare of a sample learning trajectory in the matched–mismatched pennies game (top), and in the asymmetric 5-penny game with $p = 3$ (bottom, projection onto first three variables)

$$A^{i-1,i} = \begin{bmatrix} 0 & 1 \\ p & 0 \end{bmatrix}. \tag{36}$$

with $p > 0$, yielding the following replicator equations

$$\dot{x}_i = x_i(1 - x_i)(1 - x_{i-1} - px_{i-1}). \tag{37}$$

For odd N, the game does not have a Nash equilibrium in pure strategies. The pure strategy profiles are saddle-type stationary points of the ordinary differential equation (37) 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 obtains payoff of $\frac{p}{p+1}$.

The system (37) was thoroughly analyzed in [54], and the main result given therein was that, for $N = 3$ and $p > 7$, all mixed strategies 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 6, 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 must exceed the minimax payoffs. For almost all initial conditions, the dynamics is attracted to the boundary cycle of average payoff $(p+1)\frac{N-1}{2}$ (see, e.g., Fig. 3), and indeed no chaotic emergent behavior appears.

6.3 Quasiperiodicity in a 6-player pennies game

We proceed to the first negative example. We consider a system, where six players play a custom combination of matched–mismatched pennies. The payoff matrices are given by

$$\begin{aligned} A^{0,1} = A^{1,0} = A^{2,3} &= \begin{bmatrix} -1 & 0 \\ 0 & -1 \end{bmatrix}, \\ A^{3,2} = A^{0,2} = A^{1,4} &= \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}, \\ A^{4,5} &= \begin{bmatrix} -\sqrt{2} & 0 \\ 0 & -\sqrt{2} \end{bmatrix}, \\ A^{5,4} &= \begin{bmatrix} \sqrt{2} & 0 \\ 0 & \sqrt{2} \end{bmatrix}. \end{aligned} \tag{38}$$

The system is built from three 2x2 sub-games. Players 0 and 1 play mismatching pennies, players 2 and 3 and players 4 and 5 both play matching–mismatching pennies. In addition there are matching-pennies type payoff contributions from player 0 to player 2, and from player 1 to player 4. Both nodes representing players 2 and 4 in the game graph have two predecessors each, which violates the assumptions of Theorem 4. The replicator system yields the following differential equations:

$$\begin{aligned} \dot{x}_0 &= x_0(1-x_0)(1-2x_1), \\ \dot{x}_1 &= x_1(1-x_1)(1-2x_0), \\ \dot{x}_2 &= x_2(1-x_2)(2x_3+2x_0-2), \\ \dot{x}_3 &= x_3(1-x_3)(1-2x_2), \\ \dot{x}_4 &= x_4(1-x_4)(2\sqrt{2}x_5+2x_1-\sqrt{2}-1), \\ \dot{x}_5 &= x_5(1-x_5)(\sqrt{2}-2\sqrt{2}x_4), \end{aligned} \tag{39}$$

where x_i represents the frequency of playing strategy α_0 (or simply, heads) for player i.

The dynamics of this game are best understood when interpreting them as a product of dynamics of the systems; the mismatching pennies dynamics between agents 0 and 1 are of saddle type, with mixed Nash equilibrium $(x_0^*, x_1^*) = (0.5, 0.5)$ at the center which

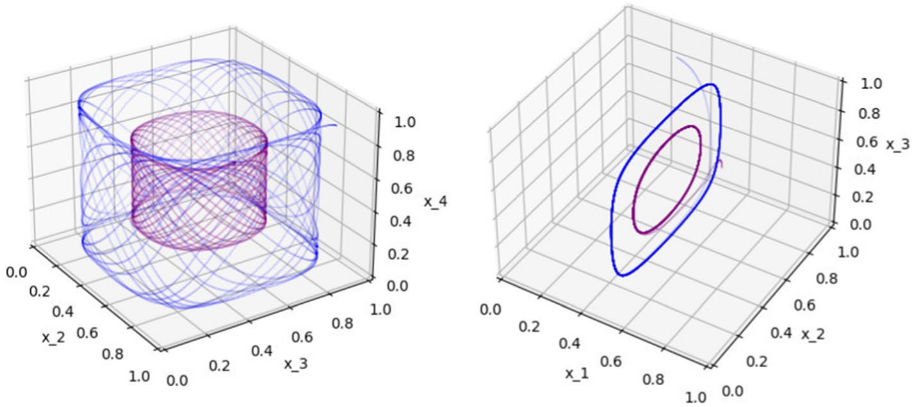

Fig. 4 Trajectories approaching two invariant tori in a six-player variant of matched–mismatched pennies

has an attracting direction along the diagonal; and two copies of matching–mismatching games, which contain continuum families of invariant cycles. The cross product of this sub-game mixed Nash equilibrium, and any chosen pair of invariant cycles from each subsystems forms a continuous family of invariant tori, as seen in Fig. 4, filled with dense orbits. Each of these tori inherits the attracting direction from the Nash equilibrium, thus forming a family of limit sets, which are neither equilibria, nor periodic. Due to lack of cyclicity, the game does not guarantee the payoff structure given by Theorem 6.

6.4 A chaotic polymatrix replicator

Our second negative example leads to an even more complex outcome; it shows that even in a binary three-player game, but without structured interactions (i.e., no cyclicity, all possible connections in the game graph), the learning trajectories of replicator dynamics can approach chaotic limit sets. The payoff matrices for this game are given by

$$
\begin{aligned}
A^{0,0} &= \begin{bmatrix} \mu & 14 \\ 0 & 0 \end{bmatrix}, \\
A^{1,0} = -A^{0,1} &= \begin{bmatrix} -10 & 10 \\ 0 & 0 \end{bmatrix}, \\
A^{2,0} = A^{2,1} = A^{2,2} = -A^{1,1} &= \begin{bmatrix} -2 & 2 \\ 0 & 0 \end{bmatrix}, \\
A^{0,2} &= \begin{bmatrix} -25 & 29 \\ 0 & 0 \end{bmatrix}, \\
A^{1,2} &= \begin{bmatrix} 0 & -11 \\ 0 & 0 \end{bmatrix}.
\end{aligned} \tag{40}
$$

After simple transformations, we arrive at the following one-parameter system of differential equations:

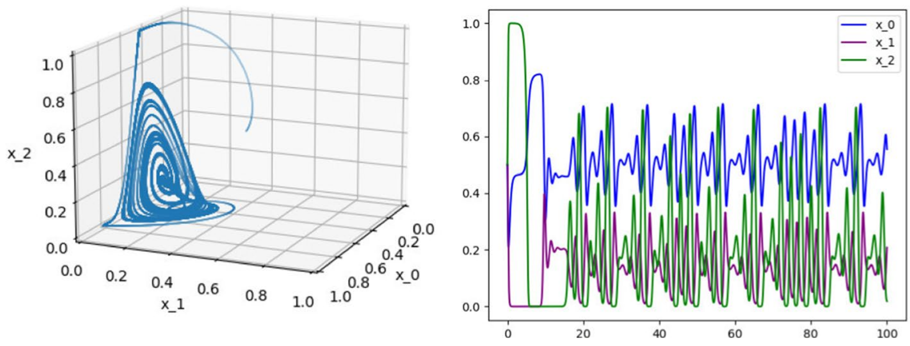

Fig. 5 A learning trajectory approaching a chaotic attractor in the polymatrix replicator [50] (left) and a plot of values of its coordinates (right). The game is characterized by unstructured interactions between payoffs and therefore breaks the assumptions of Theorems 3 and 4

$$
\begin{aligned}
\dot{x}_0 &= x_0(1-x_0)(12-\mu+(\mu-14)x_0-20x_1-4x_2),\\
\dot{x}_1 &= x_1(1-x_1)(-10+20x_0+4x_1-4x_2),\\
\dot{x}_2 &= x_2(1-x_2)(27-54x_0+11x_1-4x_2),
\end{aligned}
\tag{41}
$$

This system was recently introduced by Peixe and Rodrigues [50], who formally showed by combined theoretical and numerical approaches 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 Fig. 5. Similarly, as in the previous example, due to non-cyclicity there are no guarantees to be derived from Theorem 6 on the payoff structure.

7 Discussion

In this paper, we proved the existence of a strong connection between the structure of a game graph in GPGs, and the topological structure of the long-term FoReL learning outcomes, represented by the limit sets. For binary games of indegree one, this structure is indeed quite simple, as the Poincaré–Bendixson property chaotic limit sets cannot emerge. For cyclic games, the result is even stronger, as the limit sets can be unambiguously represented by their two-dimensional projections, and, for a large subclass of these games, carry social welfare guarantees similar to these of Nash equilibria, trajectories that need not converge.

Although these results can be applied to systems of arbitrarily high dimension, the strategy space of each individual agent is very limited – essentially one-dimensional, spanned by its choice of two pure strategies. This assumption limits the scope of applicability of provided theorems, and, without any additional dimensionality reduction, makes them unsuitable for high-dimensional deep learning scenarios [65]. On the other hand, our results extend to learning systems can include arbitrarily large amount of agents, which is a vast improvement, as the Poincaré–Bendixson theorem in its traditional, two-dimensional formulation (as given in Theorem 1) can only be directly applied to 2x2 games. This opens up possibilities of applying the theory to large, structured learning systems consisting of many primitive agents, such as the ones encountered in swarms [66], or mean field games [67].

8 Conclusions

Numerous recent results regarding 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. This realization might seem to be a setback, but when viewed from the correct perspective it unveils a new way of interpreting learning outcomes, by adopting solution concepts from the qualitative theory of dynamical systems. Our results showcase the possibility of establishing links between the topological-combinatorial structure of multi-agent games, such as game graph, or the number of actions, to understand and constrain the topological complexity of limit sets (Poincaré–Bendixson property) and finally link back to more traditional game theoretic analyses, such as calculating the efficiency of the system via social welfare. These connections showcase the promising advantages of this approach. and we hope that it will lead to more interesting results along these lines in the future.

Author contributions AC wrote sections 2, 3, 4, 6, 7. GP wrote sections 5, 8. Both authors contributed to Section 1. We would like to thank prof. Frans A. Oliehoek for his support, and helpful advice.

Funding AC was supported by the European Research Council (ERC) under the European Union's Horizon 2020 research and innovation programme (grant agreement No. 758824 – INFLUENCE). GP was supported in part by the National Research Foundation, Singapore and DSO National Laboratories under its AI Singapore Program (AISG Award No: AISG2-RP-2020-016), NRF 2018 Fellowship NRF-NRFF2018-07, NRF2019-NRF-ANR095 ALIAS grant, grant PIESGP-AI-2020-01, AME Programmatic Fund (Grant No. A20H6b0151) from the Agency for Science, Technology and Research (A*STAR) and Provost's Chair Professorship grant with number RGEPPV2101.

Data availibility Not applicable.

Declarations

Conflict of interest None.

Ethical approval Not applicable.

References

1. Bloembergen, D., Tuyls, K., Hennes, D., & Kaisers, M. (2015). Evolutionary dynamics of multi-agent learning: A survey. *Journal of Artificial Intelligence Research, 53*, 659–697.
2. Tuyls, K., & Nowé, A. (2005). Evolutionary game theory and multi-agent reinforcement learning. *The Knowledge Engineering Review, 20*(1), 63–90.
3. Rodrigues Gomes, E., & Kowalczyk, R. (2009). Dynamic analysis of multiagent q-learning with ε-greedy exploration. In *26th Annual International Conference on Machine Learning* (pp. 369–376).
4. Gatti, N., Panozzo, F., & Restelli, M. (2013). Efficient evolutionary dynamics with extensive-form games. In *27th AAAI Conference on Artificial Intelligence*.
5. Cesa-Bianchi, N., & Lugosi, G. (2006). *Prediction, learning, and games*. Cambridge University Press.
6. Shalev-Shwartz, S. (2011). Online learning and online convex optimization. *Foundations and Trends in Machine Learning, 4*(2), 107–194.
7. Shoham, Y., & Leyton-Brown, K. (2008). *Multiagent systems: Algorithmic, game-theoretic, and logical foundations*. Cambridge University Press.
8. Sato, Y., Akiyama, E., & Farmer, J. D. (2002). Chaos in learning a simple two-person game. *Proceedings of the National Academy of Sciences, 99*(7), 4748–4751.

9. Piliouras, G., & Shamma, J.S. (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* (pp. 861–873).
10. Lorenz, E. N. (1963). Deterministic nonperiodic flow. *Journal of the atmospheric sciences, 20*(2), 130–141.
11. van Strien, S. (2011). Hamiltonian flows with random-walk behaviour originating from zero-sum games and fictitious play. *Nonlinearity, 24*(6), 1715.
12. Palaiopanos, G., Panageas, I., & Piliouras, G. (2017) Multiplicative weights update with constant step-size in congestion games: Convergence, limit cycles and chaos. In *Advances in Neural Information Processing Systems* (pp. 5874–5884).
13. Benaïm, M., Hofbauer, J., & Sorin, S. (2012). Perturbations of set-valued dynamical systems, with applications to game theory. *Dynamic Games and Applications, 2*(2), 195–205.
14. Bailey, J.P., & Piliouras, G. (2018) Multiplicative weights update in zero-sum games. In *ACM Conference on Economics and Computation* (pp. 321–338).
15. Bailey, J.P., & Piliouras, G. (2019). Multi-agent learning in network zero-sum games is a hamiltonian system. In *18th International Conference on Autonomous Agents and Multiagent Systems* (pp. 233–241).
16. Cheung, Y., & Piliouras, G. (2020). Chaos, Extremism and Optimism: Volume Analysis of Learning in Games. In *Advances in Neural Information Processing Systems* (Vol. 33, pp. 9039–9049).
17. Sanders, J. B., Farmer, J. D., & Galla, T. (2018). The prevalence of chaotic dynamics in games with many players. *Scientific reports, 8*(1), 1–13.
18. Galla, T., & Farmer, J. D. (2013). Complex dynamics in learning complicated games. *Proceedings of the National Academy of Sciences, 110*(4), 1232–1236.
19. Frey, S., & Goldstone, R. L. (2013). Cyclic game dynamics driven by iterated reasoning. *PLOS ONE, 8*(2), 56416.
20. Chotibut, T., Falniowski, F., Misiurewicz, M., & Piliouras, G. (2021). Family of chaotic maps from game theory. *Dynamical Systems, 36*(1), 48–63.
21. Mertikopoulos, P., Papadimitriou, C. & Piliouras, G. (2018) Cycles in adversarial regularized learning. In *29th Annual ACM-SIAM Symposium on Discrete Algorithms* (pp. 2703–2717).
22. Cheung, Y.K., Piliouras, G. (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, pp. 1–28).
23. Taylor, P. D., & Jonker, L. B. (1978). Evolutionary stable strategies and game dynamics. *Mathematical biosciences, 40*(1–2), 145–156.
24. Hofbauer, J., & Sigmund, K. (1998). *Evolutionary games and population dynamics*. Cambridge University Press.
25. Arora, S., Hazan, E., & Kale, S. (2012). The multiplicative weights update method: a meta-algorithm and applications. *Theory of Computing, 8*(1), 121–164.
26. Chotibut, T., Falniowski, F., Misiurewicz, M., & Piliouras, G. (2020). The route to chaos in routing games: When is price of anarchy too optimistic? *Advances in Neural Information Processing Systems, 33*, 766–777.
27. Roughgarden, T. (2009). Intrinsic robustness of the price of anarchy. In *Proceedings of the Forty-first Annual ACM Symposium on Theory of Computing* (pp. 513–522).
28. Sandholm, W. H. (2010). *Population games and evolutionary dynamics*. MIT press.
29. Young, H. P. (2004). *Strategic learning and its limits*. Oxford University Press.
30. Fudenberg, D., Drew, F., Levine, D. K., & Levine, D. K. (1998). *The theory of learning in games*. MIT press.
31. Lanctot, M., Zambaldi, V., Gruslys, A., Lazaridou, A., Tuyls, K., Pérolat, J., Silver, D., Graepel, T. (2017) A unified game-theoretic approach to multiagent reinforcement learning. In *Advances in Neural Information Processing Systems* (pp. 4193–4206).
32. Muller, P., Omidshafiei, S., Rowland, M., Tuyls, K., Julien, P., Liu, S., Hennes, D., Marris, L., Lanctot, M., Hughes, E., Wang, Z., Lever, G., Heess, N., Graepel, T., & Munos, R. (2020). A generalized training approach for multiagent learning. In *International Conference on Learning Representations* (pp. 1–35).
33. Wellman, M.P. (2006). Methods for empirical game-theoretic analysis. In *Proceedings of the AAAI Conference on Artificial Intelligence* (pp. 1552–1556).
34. Menezes, F. M., & Pitchford, R. (2006). Binary games with many players. *Economic Theory, 28*(1), 125–143.
35. Blonski, M. (1999). Anonymous games with binary actions. *Games and Economic Behavior, 28*(2), 171–180.

36. Yu, S., Zhou, K., Brantingham, J., & Vorobeychik, Y. (2020). Computing equilibria in binary networked public goods games. In *Proceedings of the AAAI Conference on Artificial Intelligence* (Vol. 34, pp. 2310–2317).
37. Goforth, D., & Robinson, D. (2004). *Topology of 2x2 games*. Routledge.
38. Jordan, J. S. (1993). Three problems in learning mixed-strategy Nash equilibria. *Games and Economic Behavior, 5*(3), 368–386.
39. Gaunersdorfer, A., & Hofbauer, J. (1995). Fictitious play, shapley polygons, and the replicator equation. *Games and Economic Behavior, 11*(2), 279–303.
40. Hart, S., & Mas-Colell, A. (2003). Uncoupled dynamics do not lead to Nash equilibrium. *American Economic Review, 93*(5), 1830–1836.
41. Alvarez, I., & Poznyak, A. (2010). Game theory applied to urban traffic control problem. In *International Conference on Control, Automation and Systems* (pp. 2164–2169).
42. Kuyer, L., Whiteson, S., Bakker, B., & Vlassis, N. (2008). Multiagent reinforcement learning for urban traffic control using coordination graphs. In W. Daelemans, B. Goethals, & K. Morik (Eds.), *Machine Learning and Knowledge Discovery in Databases* (pp. 656–671). Springer.
43. Cachon, G.P. & Netessine, S. (2006). Game theory in supply chain analysis. Models, methods, and applications for innovative decision making (pp. 200–233).
44. Ambec, S., & Sprumont, Y. (2002). Sharing a river. *Journal of Economic Theory, 107*(2), 453–462.
45. Khmelnitskaya, A. B. (2010). Values for rooted-tree and sink-tree digraph games and sharing a river. *Theory and Decision, 69*(4), 657–669.
46. Koutsoupias, E., & Papadimitriou, C. (1999) Worst-case equilibria. In *Annual Symposium on Theoretical Aspects of Computer Science* (pp. 404–413).
47. Roughgarden, T. (2016). *Twenty lectures on algorithmic game theory*. Cambridge University Press.
48. Czechowski, A., & Piliouras, G. (2022). Poincaré-Bendixson limit sets in multi-agent learning. In *Proceedings of the 21st International Conference on Autonomous Agents and Multi-Agent Systems*.
49. Plank, M. (1997). Some qualitative differences between the replicator dynamics of two player and n player games. *Nonlinear Analysis: Theory, Methods & Applications, 30*(3), 1411–1417.
50. Peixe, T., & Rodrigues, A. (2022). Persistent strange attractors in 3d polymatrix replicators. *Physica D: Nonlinear Phenomena, 438*, 133346.
51. Nagarajan, S.G., Mohamed, S., & Piliouras, G. (2018). Three body problems in evolutionary game dynamics: Convergence, periodicity and limit cycles. In *18th International Conference on Autonomous Agents and Multi-Agent Systems* (pp. 685–693).
52. Nagarajan, S.G., Balduzzi, D., & Piliouras, G. (2020). From chaos to order: Symmetry and conservation laws in game dynamics. In *37th International Conference on Machine Learning* (Vol. 119, pp. 7186–7196).
53. Flokas, L., Vlatakis-Gkaragkounis, E.-V., & Piliouras, G. (2019) Poincaré recurrence, cycles and spurious equilibria in gradient-descent-ascent for non-convex non-concave zero-sum games. In *Proceedings of the 33rd International Conference on Neural Information Processing Systems* (pp. 10450–10461).
54. Kleinberg, R.D., Ligett, K., Piliouras, G., & Tardos, É. (2011). Beyond the Nash equilibrium barrier. In *Symposium on Innovations in Computer Science* (pp. 125–140).
55. Kearns, M. (2007). Graphical games. *Algorithmic game theory, 3*, 159–180.
56. Yanovskaya, E. B. (1968). Equilibrium points in polymatrix games. *Litovskii Matematicheskii Sbornik, 8*, 381–384.
57. Howson, J. T., Jr. (1972). Equilibria of polymatrix games. *Management Science, 18*(5–part–1), 312–318.
58. Hale, J. K. (2009). *Ordinary differential equations*. Courier Corporation.
59. Bendixson, I. (1901). Sur les courbes définies par des équations différentielles. *Acta Mathematica, 24*(1), 1–88.
60. Mallet-Paret, J., & Smith, H. L. (1990). The Poincaré-Bendixson theorem for monotone cyclic feedback systems. *Journal of Dynamics and Differential Equations, 2*(4), 367–421.
61. Jordan, C. (1887). *Cours D'analyse de l'École Polytechnique: T. Équations Différentielles Ordinaires. Équations Linéaires. Équations aux Dérivées partielles. Calcul des variations* (Vol. 3). Gauthier-Villars et fils.
62. Roughgarden, T. (2009). Intrinsic robustness of the price of anarchy. In *Proceedings of the Forty-first Annual ACM Symposium on Theory of Computing* (pp. 513–522).
63. Lykouris, T., Syrgkanis, V., Tardos, É. (2016). Learning and efficiency in games with dynamic population. In *Proceedings of the Twenty-seventh Annual ACM-SIAM Symposium on Discrete Algorithms* (pp. 120–129) SIAM.
64. Hindmarsh, A., & Petzold, L. (2005). Lsoda, ordinary differential equation solver for stiff or non-stiff system. NEA.

65. Wong, A., Bäck, T., Kononova, A.V., & Plaat, A. (2022). Deep multiagent reinforcement learning: Challenges and directions. Artificial Intelligence Review, 1–34.
66. Jadbabaie, A., Lin, J., & Morse, A. S. (2003). Coordination of groups of mobile autonomous agents using nearest neighbor rules. *IEEE Transactions on Automatic Control, 48*(6), 988–1001.
67. Lasry, J.-M., & Lions, P.-L. (2007). Mean field games. *Japanese Journal of Mathematics, 2*(1), 229–260.