Online Learning in Periodic Zero-Sum Games: von Neumann vs Poincaré

Work-in-progress paper †

Tanner Fiez
University of Washington
Seattle, Washington
fiezt@uw.edu

Ryann Sim
SUTD
Singapore, Singapore
ryann_sim@mymail.sutd.edu.sg

Stratis Skoulakis
SUTD
Singapore, Singapore
efstratios@sutd.edu.sg

Georgios Piliouras*
SUTD
Singapore, Singapore
georgios@sutd.edu.sg

Lillian Ratliff*
University of Washington
Seattle, Washington
ratliffl@uw.edu

ABSTRACT

Arguably, one of the most seminal results in game theory is von Neumann's minmax theorem that zero-sum games admit an essentially unique equilibrium solution. Ubiquitous, classic learning theoretic results show that online no-regret dynamics converge to equilibrium in zero-sum games in a time-average sense. Driven by AI applications, recent work has focused on the day-to-day behavior of such dynamics and show that it is typically cyclic, and formally Poincaré recurrent. We test the robustness of these results in the case of *dynamic zero-sum games* by focusing on the simplest such setting of periodically evolving games with a time-invariant equilibrium. This natural extension captures amongst other settings constant-sum games with a periodically evolving total sum. Interestingly, time-average convergence may fail even in the simplest such settings, in spite of the *equilibrium being fixed*, whereas recurrence provably generalizes well despite the complex, non-autonomous nature of these systems.

1 INTRODUCTION

The study of learning dynamics in zero-sum games is arguably as old a field as game theory itself, dating back to the seminal work of Brown and Robinson [7, 24], which followed shortly after the foundational maxmin theorem of von Neumann [29]. The dynamics of no-regret algorithms [25] are of particular interest in these settings as they are designed with an adversarial (even dynamically so) environment in mind. Arguably the most well known family of regret-minimizing dynamics is Follow-the-Regularized-Leader (FTRL, also known as Online Mirror Descent) dynamics. Well known results imply that such dynamics converge in time-average to the max-min equilibrium [10, 14].

Despite the classic nature of these questions, the actual day-to-day behavior of such dynamics in zero-sum games has received significantly less attention up until recent AI applications [15, 26] reinvigorated the interest for a deeper understanding of these settings. Since then a number of works have studied FTRL learning dynamics in zero-sum games (and variants thereof) with a particular focus on continuous-time dynamics [6, 18, 20–22, 28]. The unifying emergent picture is that such dynamics are "approximately cyclic" in a formal sense known as Poincaré recurrence. This condition is similar to the behavior exhibited by physical systems such as celestial mechanics (e.g. solar system). In fact, after appropriate change of coordinate systems such game dynamics are formally equivalent to Hamiltonian mechanics [4, 16]. Moreover, these results have acted as fundamental building blocks for understanding the limit behavior of their discrete-time variants [2, 3, 11, 12, 19].

Despite the plethora of emerging results, an important and well motivated aspect of this problem has received very little attention in comparison. *How do* FTRL *dynamics behave if the zero-sum game itself evolves over time?* Clearly, the answer depends critically on how the game is allowed to evolve over time.

Recently there have been some early forays in studying learning dynamics in dynamically evolving zero-sum games. Mai et al. [17], Skoulakis et al. [27] consider endogeneously evolving population zero-sum games, where a parametric game evolves itself adversarially towards the participating agents. A transformation that treats the game as an additional "hidden" agents allows for a reduction to a more standard static network zero-sum game for which both time-average convergence to equilibrium as well as Poincaré recurrence hold. Cardoso et al. [9] examine time-evolving zero-sum games and design new algorithms that provide a novel type of regret guarantee that they call Nash equilibrium regret. In a sense these algorithms are competitive against the NE of the long-term-averaged payoff matrix. Duvocelle et al. [13] study discrete-time FTRL dynamics in evolving games that are strictly/strongly monotone and provide sufficient conditions (e.g.when the evolving game stabilizes) under which the dynamics track/converge the evolving equilibrium. Unfortunately, zero-sum games do not satisfy these properties and hence these results do not apply to our setting.

Setting. We study learning dynamics in periodic zero-sum games on both finite and continuous strategy spaces under the assumption that there exists a common interior Nash equilibrium within the set of allowable games. From a standard game theoretic perspective this setting allows us to model competition between two service providers that wish to maximize their users, while the total market size evolves seasonally over time. This affects the resulting reward functions, even if the fundamentals of the market (e.g. the agent capabilities and thus the Nash equilibrium remain time-invariant).

† This paper is submitted under the work-in-progress category. We omit several proofs of theoretical results and leave them for a future published version of this work.
* Joint last authors.

Similarly, in AI applications of interest, data-driven empirical zero-sum games evolve over time as more data pours in. Moreover, training by cycling over uniformly random batches of a large dataset data can be modeled by playing a periodic set of games with similar properties (e.g., equilibrium).

Contributions. For the class of periodic zero-sum bilinear games with continuous and unconstrained strategy spaces, we examine the gradient descent-ascent (GDA) update rule in continuous time. In the class of periodic zero-sum matrix games (and network generalizations thereof) with finite strategy spaces we study the class of follow-the-regularized-leader (FTRL) learning dynamics. We show that GDA and FTRL learning dynamics are both Poincaré recurrent in our setting (Theorems 4 & 6). On the other hand, the time-average properties of GDA and FTRL dynamics are not well-behaved. Indeed, we show that even in the simplest games in this class, *the time-average* GDA *and* FTRL *strategies do not converge to the time-invariant Nash equilibrium* (Theorems 5 & 8). On the positive side, we are able to show that the time-average utilities of FTRL learning dynamics converge to the time-average game values in periodic zero-sum bimatrix games (Theorem 7).

Technical Novelty and Approach. Analyzing continuous-time GDA and FTRL dynamics in periodic zero-sum games presents a number of technical challenges which were not present in static fixed zero-sum games. In particular, each set of learning dynamics corresponds to a *non-autonomous* ordinary differential equation. This necessitates a number of novel ideas and techniques as the typical Poincaré recurrence theorem no longer applies. We begin by showing that a divergence-free vector field is sufficient to ensure volume preservation even in time-varying systems. Then, we prove that both continuous-time GDA and FTRL learning dynamics in periodic zero-sum games correspond to divergence-free vector fields so that volume is preserved. Following this, we are able to derive time-invariant functions that imply orbits in the systems remain bounded. Finally, to apply Poincaré's recurrence theorem, we construct a discrete-time subsequence that we are able to prove returns arbitrarily close to the initial condition infinitely often.

Organization. In Section 2, we formalize the classes of games that we study. We present characteristics of dynamical systems as they pertain to this work in Section 3. Section 4 and 5 contain our results analyzing GDA and FTRL learning dynamics in continuous and finite strategy periodic zero-sum bilinear and polymatrix games, respectively. We present illustrative numerical experiments in Section 6 and then conclude with a discussion in Section 7.

We remark that in this work-in-progress draft of the paper, we omit several proofs of theoretical results and leave them for a future version to be published. However, we do provide the key ideas that allow us to obtain the stated results in this paper.

2 GAME-THEORETIC PRELIMINARIES

This section formally defines the classes of games we study.

2.1 Continuous Strategy Periodic Zero-Sum Games

In the space of continuous strategy periodic zero-sum games, we focus our attention on periodic zero-sum bilinear games. We begin by formalizing the setup of a static zero-sum bilinear game and then define the periodic variant.

Zero-Sum Bilinear Games. Given a matrix $A \in \mathbb{R}^{n_1 \times n_2}$, a zero-sum bilinear game on continuous strategy spaces can be defined by the max-min problem

$$\max_{x_1 \in \mathbb{R}^{n_1}} \min_{x_2 \in \mathbb{R}^{n_2}} x_1^\top A x_2.$$

Formally, the action space of agents 1 and 2 are given by $\mathbb{R}^{n_1}$ and $\mathbb{R}^{n_2}$, respectively. The game is defined by the pair of payoff matrices $\{A, -A^\top\}$. Player 1 seeks to maximize the utility function $u_1(x_1, x_2) = x_1^\top A x_2$ while player 2 optimizes the utility $u_2(x_1, x_2) = -x_2^\top A^\top x_1$. The game is zero-sum since for any $x_1 \in \mathbb{R}^{n_1}$ and $x_2 \in \mathbb{R}^{n_2}$, the sum of utility over each player is zero.

Nash Equilibrium. For zero-sum bilinear games, a Nash equilibrium corresponds to a joint strategy (x_1^*, x_2^*) such that for each player i and $j \neq i$,

$$u_i(x_i^*, x_j^*) \geq u_i(x_i, x_j^*),\ \forall x_i \in \mathbb{R}^{n_i}.$$

Note that $(x_1^*, x_2^*) = (\mathbf{0}, \mathbf{0})$ is always a Nash equilibrium of a zero-sum bilinear game.

Periodic Zero-Sum Bilinear Games. We study the continuous time GDA learning dynamics in a class of games we refer to as periodic zero-sum bilinear games. The key distinction from a typical static zero-sum bilinear game is that the payoff matrix is no longer fixed in this class of games. Instead, the payoff matrix may change at each time instant as long as the sequence of payoffs is periodic with a time-invariant Nash equilibrium. The next definition formalizes the games we study on continuous strategy spaces.

Definition 1 (Periodic Zero-Sum Bilinear Game). *A periodic zero-sum bilinear game is defined by an infinite sequence of zero-sum bilinear games $\{A(t), -A(t)^\top\}_{t=0}^\infty$ in which the set of players and strategy spaces are fixed and the payoff matrix is such that $A(t) = A(t+T)$ for some finite period T and all $t \geq 0$.*

2.2 Finite Strategy Periodic Zero-Sum Games

For finite strategy periodic zero-sum games, we analyze periodic zero-sum polymatrix games. In what follows we define a zero-sum polymatrix game, which is a network generalization of a bimatrix game, and then detail the periodic variant considered in this paper.

Zero-Sum Polymatrix Games. An N-player polymatrix game is defined by an undirected graph $G = (V, E)$ where V corresponds to the set of agents and E corresponds to the set of edges between agents in which a bimatrix game is played between the endpoints [8]. Each agent $i \in V$ has a set of actions $\mathcal{A}_i = \{1, \ldots, n_i\}$ that can be selected at random from a distribution x_i called a mixed strategy. The set of mixed strategies of player $i \in V$ is the standard simplex in $\mathbb{R}^{n_i}$ and is denoted $\mathcal{X}_i = \Delta^{n_i - 1} = \{x_i \in \mathbb{R}^{n_i}_{\geq 0} : \sum_{\alpha \in \mathcal{A}_i} x_{i\alpha} = 1\}$ where $x_{i\alpha}$ denotes the probability mass on action $\alpha \in \mathcal{A}_i$. We denote the joint strategy space of the agents by $\mathcal{X} = \Pi_{i \in V} \mathcal{X}_i$.

The bimatrix game on edge (i, j) is described using a pair of matrices $A^{ij} \in \mathbb{R}^{n_i \times n_j}$ and $A^{ji} \in \mathbb{R}^{n_j \times n_i}$. An entry $A^{ij}_{\alpha\beta}$ for $(\alpha, \beta) \in \mathcal{A}_i \times \mathcal{A}_j$ represents the reward player i obtains for selecting action α given that player j chooses action β. The utility or payoff of agent $i \in V$ under the strategy profile $x \in \mathcal{X}$ is denoted by $u_i(x)$ and corresponds to the sum of payoffs from the bimatrix games the agent participates in. The payoff is equivalently expressed as $u_i(x_i, x_{-i})$ when distinguishing between the strategy of player i and all other players $-i$. More precisely,

$$u_i(x) = \sum\nolimits_{j:(i,j)\in E} x_i^\top A^{ij} x_j.$$

We further denote by $u_{i\alpha}(x) = \sum_{j:(i,j)\in E}(A^{ij}x_j)_\alpha$ the utility of player $i \in V$ under the strategy profile $x = (\alpha, x_{-i}) \in \mathcal{X}$ for $\alpha \in \mathcal{A}_i$. The game is called zero-sum if $\sum_{i\in V} u_i(x) = 0$ for all $x \in \mathcal{X}$. Observe that each individual bimatrix game on an edge is not necessarily zero-sum in a zero-sum polymatrix game.

Nash Equilibrium. A Nash equilibrium in a polymatrix game is defined as a mixed strategy profile $x^* \in \mathcal{X}$ such that for each player $i \in V$,

$$u_i(x_i^*, x_{-i}^*) \geq u_i(x_i, x_{-i}^*), \ \forall x_i \in \mathcal{X}_i.$$

We denote the support of $x_i^* \in \mathcal{X}_i$ by $\text{supp}(x_i^*) = \{\alpha \in \mathcal{A}_i : x_{i\alpha} > 0\}$. A Nash equilibrium is said to be an interior or fully mixed Nash equilibrium if $\text{supp}(x_i^*) = \mathcal{A}_i \ \forall i \in V$.

Periodic Zero-Sum Polymatrix Games. We analyze the continuous time FTRL learning dynamics in a class of games we refer to as periodic zero-sum polymatrix games. This class of games is such that each edge game in the zero-sum polymatrix game is evolving in a periodic fashion with time. We consider that this periodic evolution is such that there is a common interior Nash equilibrium that arises in each polymatrix game that arrives. The following definition formalizes the games we study on finite strategy spaces.

Definition 2 (Periodic Zero-Sum Polymatrix Game). *A periodic zero-sum polymatrix game is defined by an infinite sequence of zero-sum polymatrix games $\{G(t) = (V(t), E(t))\}_{t=0}^\infty$ in which the set of players, strategy spaces, and edges are fixed and each bimatrix game on an edge (i, j) is such that $A^{ij}(t) = A^{ij}(t+T)$ and $A^{ji}(t) = A^{ji}(t+T)$ for some finite period T and all $t \geq 0$. Moreover, we assume that there is some $x^* \in \mathcal{X}$ that is an interior Nash equilibrium of the polymatrix game $G(t)$ for all $t \geq 0$.*

3 TOPOLOGY OF DYNAMICAL SYSTEMS

We now cover concepts from dynamical systems theory that will help us analyze learning dynamics and prove Poincaré recurrence. It is worth noting that careful attention must be given to such preliminaries in this work since the ordinary differential equations we study are non-autonomous whereas typical recurrence analysis in the study of learning in games deals with autonomous ordinary differential equations.

Flows. Consider a differential equation $\dot{x} = f(t, x)$ on a topological space X. The existence and uniqueness theorem for ordinary differential equations guarantees that there exists a unique continuous function $\phi : \mathbb{R} \times X \to X$, which is termed the *flow*, that satisfies (*i*) $\phi(t, \cdot) : X \to X$, often denoted $\phi^t : X \to X$, is a homeomorphism for each $t \in \mathbb{R}$, (*ii*) $\phi(t+s, x) = \phi(t, \phi(s, x))$ for all $t, s \in \mathbb{R}$ and all $x \in X$, (*iii*) for each $x \in X$, $\frac{d}{dt}|_{t=0}\phi(t, x) = f(t, x)$,[1] and (*iv*) $\phi(t, x_0) = x(t)$ is the solution.

Periodic Systems and Poincaré Maps. A system $\dot{x} = f(t, x)$ is T-periodic if $f(t+T, x) = f(t, x)$ for all (x, t). Let $\phi^t : \mathbb{R}^n \to \mathbb{R}^n$ denote the mapping taking $x \in \mathbb{R}^n$ to the value at time t. For a T-periodic system, $\phi^{T+s} = \phi^s \circ \phi^T$ so that $\phi^{kT} = (\phi^T)^k$ for any integer k. The mapping $\phi^T : \mathbb{R}^n \to \mathbb{R}^n$ is called the *Poincaré map*.

If the differential equation is well-defined for all x and has a solution[2] for all $t \in [0, T]$, then for each initial condition (where we have suppressed the dependence on x_0), the Poincaré map ϕ^T defines a discrete time dynamical system. Given that we study T-periodic systems formed from classes of learning dynamics in periodic games, this concept is key to the analysis methods we pursue.

Conservation of Volume. The flow ϕ of a system of ordinary differential equations is called *volume preserving* if the volume of the image of any set $U \subseteq \mathbb{R}^d$ under ϕ^t is preserved. More precisely, for any set $U \subseteq \mathbb{R}^d$, $\text{vol}(\phi^t(U)) = \text{vol}(U)$. Whether or not a flow preserves volume can be determined by applying *Liouville's theorem*, which says the flow is volume preserving if and only if the divergence of f at any point $x \in \mathbb{R}^d$ equals zero: that is, $\text{div} f(t, x) = \text{tr}(Df(t, x)) = \sum_{i=1}^d \frac{\partial f(t,x)}{\partial x_i} = 0$.

The following result states that if the divergence of a T-periodic vector field $f(x, t)$ is divergence free so that the flow ϕ^t is volume preserving, then the Poincaré map ϕ^T is also volume preserving.

Theorem 1 (Arnol'd 1, Theorem 2, Chapter 3, Section 16.B). *If the T-periodic system $\dot{x} = f(t, x)$ is divergence-free, then ϕ^T preserves volume.*

As a result of this theorem, the Poincaré map plays an important role in our analysis of T-periodic systems as it enables us to infer that the discrete time dynamical system defined by ϕ^T is volume preserving given that the vector field of the continuous time system is divergence free. When the orbits of the system are also bounded, this allows us to construct a subsequence of the continuous time dynamical system that is autonomous and returns arbitrarily close to the initial condition infinitely often, a property known as Poincaré recurrence.

Poincaré Recurrence. If an autonomous dynamical system preserves volume and every orbit remains bounded, almost all trajectories return arbitrarily close to their initial position, and do so infinitely often [23]. Given a flow ϕ^t on a topological space X, a point $x \in X$ is *nonwandering* for ϕ^t if for each open neighborhood U containing x, there exists $T > 1$ such that $U \cap \phi^T(U) \neq \emptyset$. The set of all nonwandering points for ϕ^t, called the *nonwandering set*, is denoted $\Omega(\phi^t)$.

Theorem 2 (Poincaré Recurrence, Poincaré 23). *If a flow preserves volume and has only bounded orbits, then for each open set almost all orbits intersecting the set intersect it infinitely often: if ϕ^t is a volume preserving flow on a bounded set $Z \subset \mathbb{R}^d$, then $\Omega(\phi^t) = Z$.*

Given that the systems we study of learning dynamics in time-evolving games constitute non-autonomous dynamical systems, the usual statement of Poincaré recurrence is not sufficient for the problem. Thus, it will be useful for us to apply an alternative formulation of the Poincaré recurrence theorem that is applicable to discrete time maps. Our proof strategy will be to show that the following Poincaré recurrence theorem for discrete maps can be

[1] We take $t_0 = 0$ without loss of generality.

[2] The fundamental theorem of differential equations states that the solution of the differential equation will exist and be unique if, for every t, $f(\cdot, t)$ is Lipscthiz continuous and, for every x, $f(x, \cdot)$ is continuous at all but a measure zero set of points [1]. This condition is satisfied by the learning dynamics we analyze in this paper.

applied to the Poincaré map ϕ^T to guarantee the recurrent nature of the system orbits in the time-varying systems we analyze.

Theorem 3 (Poincaré Recurrence for Discrete-Time Maps, Barreira 5). *Let (X, Σ, μ) be a finite measure space and let $\phi: X \to X$ be a measure-preserving map. For any $E \in \Sigma$, the set of those points x of E for which there exists $N \in \mathbb{N}$ such that $\phi^n(x) \notin E$ for all $n > N$ has zero measure. In other words, almost every point of E returns to E. In fact,* almost every point returns infinitely often. *I.e., $\mu(\{x \in E : \exists\, N \text{ s.t. } \phi^n(x) \notin E \text{ for all } n > N\}) = 0$.*

4 GRADIENT DESCENT-ASCENT DYNAMICS

In this section, we study the continuous time GDA learning dynamics in periodic zero-sum bilinear games. The dynamics are such that each player seeks to maximize their utility by following the gradient with respect to their choice variable. This gives rise to the following system that we analyze in this section:

$$\dot{x}_1 = A(t)x_2(t)$$
$$\dot{x}_2 = -A^\top(t)x_1(t).$$

4.1 Poincaré Recurrence

We show in this section that the continuous time GDA learning dynamics are formally Poincaré recurrent. This serves as a warm-up for our study of FTRL learning dynamics and allows us to illustrate our analysis techniques in the simpler setting of two-player periodic zero-sum games. The following theorem formalizes our main result of this section.

Theorem 4. *The GDA learning dynamics are Poincaré recurrent in any periodic zero-sum bilinear game as given in Definition 1.*

Theorem 4 establishes that the recurrent nature of continuous time GDA dynamics in static zero-sum bilinear games is robust to the dynamic evolution of the payoff matrix in the periodic zero-sum bilinear games we analyze. Recalling the necessary techniques presented in Section 2 to show this type of result, we proceed by stating that volume is preserved and the orbits of the system remained bounded.

Toward stating the volume preservation property, we remark that it is straightforward to verify that the dynamics form a Hamiltonian system. Hamiltonian systems (time-varying or time-invariant) are volume preserving. Consequently, by explicitly deriving a Hamiltonian for the system, the volume preservation property of the dynamics is immediate. However, in our analysis, to mirror the proof techniques we employ in our study of FTRL dynamics, we show that the vector field is divergence free, which by Liouville's theorem is sufficient to conclude that the dynamics are volume preserving. We now state this result formally.

Lemma 1. *The GDA learning dynamics are volume preserving in any periodic zero-sum bilinear game as given in Definition 1.*

We now proceed by showing that the orbits of the system are bounded by demonstrating that a constant of motion exists for the dynamics. Indeed, in our analysis, we prove that the following function is time-invariant:

$$\Phi(t) = \frac{1}{2}\big(x_1^\top(t)x_1(t) + x_2^\top(t)x_2(t)\big).$$

This directly implies that the orbits of the system are bounded as is now stated.

Lemma 2. *The orbits generated by GDA learning dynamics are bounded in any periodic zero-sum bilinear game as given in Definition 1.*

Given the volume preservation and bounded orbit characteristics of the GDA learning dynamics in periodic zero-sum games, the proof of recurrence follows from applying Theorem 1 and Theorem 3 as we now outline. Observe that by definition of the periodic zero-sum bilinear game, the dynamics are T-periodic. Consider the discrete time dynamical system defined by the Poincaré map ϕ^T that arises. This system retains the volume preservation property of the continuous time system from Lemma 1 as a direct implication of Theorem 1 and the bounded orbits guarantee of the continuous time system from Lemma 2 since it holds at any set of times. Thus, we are able to apply Theorem 3 to the system defined by ϕ^T and conclude that the GDA dynamics are Poincaré recurrent.

4.2 Time-Average Convergence

The Poincaré recurrence analysis of the continuous time GDA learning dynamics in periodic zero-sum bilinear games indicates that the system has regularities which couple the evolving players and evolving game despite the failure to converge to a fixed point. A natural follow-up question to the cyclic transient behavior of the dynamics is whether the long-run converges to a game-theoretically meaningful outcome such as a Nash equilibrium.

We show that in periodic zero-sum bilinear games, the time-average of GDA learning dynamics may not converge to the time-invariant Nash equilibrium even for the simplest of periodic zero-sum bilinear games.

Theorem 5. *There exists periodic zero-sum bilinear games satisfying Definition 1 where the time-average strategies of the GDA dynamics fail to converge to the time-invariant equilibrium $(\mathbf{0}, \mathbf{0})$.*

To prove the above result, we consider a periodic zero-sum bilinear game with the action space of each player on $\mathbb{R}$ so that the evolving game simply rescales the vector field. The rescaling of the vector field is such that the dynamics return to the initial condition after each period of the evolving game. However, the time-average of the dynamics over a period of the game do not converge to the equilibrium since the dynamics spend more time on a portion of the orbit as a result of the rescaling.

5 FOLLOW-THE-REGULARIZED-LEADER DYNAMICS

In this section, we study FTRL learning dynamics in periodic zero-sum polymatrix games. Learning agents that follow FTRL learning dynamics in this class of games select a mixed strategy at each time that maximizes the difference between the cumulative payoff evaluated over the history of games along with opponent strategies and a regularization penalty. This adaptive strategy balances exploitation based on the past with exploration.

Formally, the continuous time FTRL learning dynamics for any player $i \in V$ in a periodic zero-sum polymatrix game with an initial payoff vector $y_i(0) \in \mathbb{R}^{n_i}$ are given by

$$\begin{aligned} y_i(t) &= y_i(0) + \int_0^t \sum_{j:(i,j)\in E} A^{ij}(\tau)x_j(\tau)d\tau \\ x_i(t) &= \arg\max_{x_i\in\mathcal{X}_i} \{\langle x_i, y_i(t)\rangle - h_i(x_i)\} \end{aligned} \tag{1}$$

where $h_i : \mathcal{X}_i \to \mathbb{R}$ is penalty term which encourages exploration away from the strategy which maximizes the cumulative payoffs in hindsight. We assume that the regularization function $h_i(\cdot)$ for each player $i \in V$ is continuous, strictly convex on $\mathcal{X}_i$, and smooth on the relative interior of every face of $\mathcal{X}_i$. These assumptions are sufficient to guarantee that the strategy update $x_i(t)$ is well-defined since a unique solution exists.

Prototypical instantiations of FTRL learning dynamics are the multiplicative weights update and the projected gradient dynamics. The multiplicative weights dynamics for a player $i \in V$ arise from the regularization function $h_i(x_i) = \sum_{\alpha\in\mathcal{A}_i} x_{i\alpha}\log x_{i\alpha}$ and correspond to the replicator dynamics from evolutionary game theory. The projected gradient dynamics for a player $i \in V$ derive from the Euclidean regularization $h_i(x_i) = \frac{1}{2}\|x_i\|_2^2$.

To simplify notation, the FTRL dynamics can equivalently be formulated as the following update

$$\begin{aligned} y_i(t) &= y(0) + \int_0^t v_i(x(\tau),\tau)d\tau \\ x_i(t) &= Q_i(y_i(t)). \end{aligned} \tag{2}$$

Observe that we denote by $v_i(x,\tau) = (u_{i\alpha}(x,\tau))_{\alpha\in\mathcal{A}_i}$ the vector of each pure strategy $\alpha \in \mathcal{A}_i$ utility for agent $i \in V$ under the joint profile $x = (\alpha, x_{-i}) \in \mathcal{X}$ at time $\tau \geq 0$. Moreover, $Q_i : \mathbb{R}^{n_i} \to \mathcal{X}_i$ is known as the choice map and defined as follows

$$Q_i(y_i(t)) = \arg\max_{x_i\in\mathcal{X}_i} \{\langle y_i(t), x_i\rangle - h_i(x_i)\}.$$

In this notation, the utility of the player $i \in V$ under the joint strategy $x = (x_i, x_{-i}) \in \mathcal{X}$ at time $t \geq 0$ is given by $u_i(x,\tau) = \langle v_i(x,\tau), x_i\rangle$. Observe that in our notation of utility we are now including the time index to make the dependence on the evolving game and payoffs explicit.

For any player $i \in V$ we denote by $h_i^* : \mathbb{R}^{n_i} \to \mathbb{R}$ the convex conjugate of the regularization function $h_i : \mathcal{X}_i \to \mathbb{R}$ which is given by the quantity

$$h_i^*(y_i(t)) = \max_{x_i\in\mathcal{X}_i} \{\langle x_i, y_i(t)\rangle - h_i(x_i)\}.$$

We now move on to studying the behavior of FTRL dynamics in periodic zero-sum polymatrix games.

5.1 Poincaré Recurrence

We prove in this section that the continuous time FTRL learning dynamics are Poincaré recurrent in periodic zero-sum polymatrix games. It is known that such a property holds in static zero-sum polymatrix games [18]. The following result demonstrates that this characteristic holds even in games that are evolving in a periodic fashion, providing a broad generalization.

Theorem 6. *The FTRL learning dynamics are Poincaré recurrent in any periodic zero-sum polymatrix game as given in Definition 2.*

In the remainder of this section we give an overview of the key technical approaches. The general approach is that we prove the Poincaré recurrence of a transformed system which then allows us to infer the equivalent property for the FTRL system.

For each player $i \in V$, consider a fixed strategy β_i, and for all $\alpha_i \in \mathcal{A}_i \setminus \beta_i$, define for all $t \geq 0$ the utility differences

$$z_{i\alpha_i}(t) = y_{i\alpha_i}(t) - y_{i\beta_i}(t).$$

The utility differences for each player $i \in V$ and strategy $\alpha_i \in \mathcal{A}_i \setminus \beta_i$ evolve following the differential equation

$$\dot{z}_{i\alpha_i} = v_{i\alpha_i}(x(t),t) - v_{i\beta_i}(x(t),t).$$

We proceed by show that this system is Poincaré recurrent. As an initial step, we show that the vector field $\dot{z}$ is divergence free and hence volume preserving.

Lemma 3. *The dynamics defined by the system $\dot{z}$ are volume preserving in any periodic zero-sum polymatrix game as given in Definition 2.*

Then, we prove that the following function is time-invariant:

$$\Phi(x^*, y(t)) = \sum_{i\in V} \big(h_i^*(y_i(t)) - \langle x_i^*, y_i(t)\rangle + h_i(x_i^*)\big).$$

The time-invariance of this function is sufficient to guarantee that the orbits generated by the $\dot{z}$ dynamics are bounded.

Lemma 4. *The orbits generated by the $\dot{z}$ dynamics are bounded in any periodic zero-sum polymatrix game as given in Definition 2.*

From this point, we follow similar arguments as for Theorem 4 to conclude the $\dot{z}$ dynamics are Poincaré recurrent. That is, we consider the discrete system defined by the Poincaré map corresponding to the $\dot{z}$ system and then argue that the volume preservation and bounded orbit properties carry over as a result of Theorem 1. Then, we conclude the Poincaré recurrence of the $\dot{z}$ system by applying Theorem 3. Finally, we show that Poincaré recurrence of the $\dot{z}$ system is sufficient to guarantee the Poincaré recurrence of the FTRL learning dynamics.

5.2 Time-Average Convergence

A number of well-known properties of zero-sum bimatrix games fail to generalize to zero-sum polymatrix games. Indeed, fundamental characteristics of zero-sum bimatrix games include that each agent has a unique utility value in any Nash equilibrium and that equilibrium strategies are exchangeable. However, Cai and Daskalakis [8] show that neither of these properties are guaranteed. Consequently, in general, seeking time-average convergence in the utility and strategy spaces to an equilibrium value and strategy is not well-defined even in static zero-sum polymatrix games.

For the reasons just outlined, we pursue a different notion of time-average convergence in periodic zero-sum polymatrix games. That is, we consider periodic zero-sum bimatrix games (2-player zero-sum polymatrix games) and show that the time-average utility of each agent converges to the time-average of game values over a period of the periodic game.

Theorem 7. *In periodic zero-sum bimatrix games satisfying Definition 2, if each player follows FTRL dynamics, then the time-average utility of each player converges to the time-average over a period of the values of the games.*

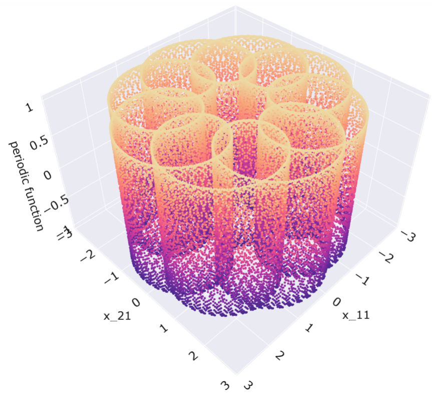


Figure 1: Periodically rescaled Matching Pennies showing the joint evolution of the players and game. (GDA)

This result paints a positive view of the time-average behavior of FTRL learning dynamics in periodic zero-sum games. However, the following result demonstrates that much like in the case of GDA in periodic zero-sum bilinear games, the time-average strategies are not guaranteed to converge to the invariant Nash equilibrium.

Theorem 8. *There exist periodic zero-sum bimatrix games satisfying Definition 1 in which the time-average strategies of* FTRL *dynamics fail to converge to the time-invariant Nash equilibrium.*

6 EXPERIMENTS

In this section, we present several experimental simulations that illustrate our theoretical results. First, for continuous time gradient descent-ascent dynamics we show that Poincarè recurrence holds in a periodic zero-sum bilinear game. We consider the ubiquitous Matching Pennies game with payoff matrix $\left(\begin{smallmatrix} 1 & -1 \\ -1 & 1 \end{smallmatrix}\right)$. We then use the following periodic rescaling with period 2π:

$$\alpha(t) = \begin{cases} \sin(t) & 0 \leq t \leq \frac{3\pi}{2} \\ \left(\frac{2}{\pi}\right)(t \bmod(2\pi) - 2\pi) & \frac{3\pi}{2} \leq t \leq 2\pi \end{cases}$$

When agents use GDA learning dynamics, we see from Figure 1 that the agents' trajectories when plotted alongside the value of the periodic rescaling are bounded.

A similar experimental result holds in the case of FTRL dynamics. We simulate replicator dynamics with the same periodic rescaling as in Figure 1, and the trajectories also remain bounded.

Lemmas 2 and 4 describe functions Φ which remain time-invariant. In the case of replicator dynamics, $\Phi(t)$ is the sum of Kullback-Liebler divergences measured between the strategy of each player and the mixed Nash $[1/2, 1/2]$. We simulated a 64-player polymatrix extension to the Matching Pennies game, where each agent plays against the opponent immediately adjacent to them, forming a 'toroid'-like chain of games. Furthermore, we randomly rescale each game with a different periodic function. Figure 2 depicts the claim presented in the lemmas: although each agent's specific divergence term $\mathrm{KL}(x_i^*||x_i(t))$ fluctuates, the sum $\sum_{i\in V}\mathrm{KL}(x_i^*||x_i(t))$ remains constant.

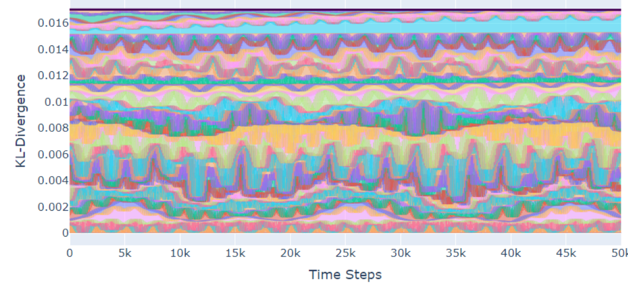
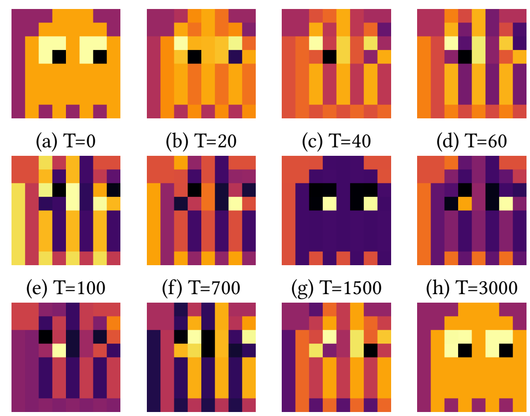


Figure 2: Each color represents a partial sum of $\mathrm{KL}(x_i^*||x_i(t))$ over the players in a 64 player periodically rescaled Matching Pennies game. The sum $\sum_{i\in V}\mathrm{KL}(x_i^*||x_i(t))$ is invariant.



(a) T=0 (b) T=20 (c) T=40 (d) T=60
(e) T=100 (f) T=700 (g) T=1500 (h) T=3000
(i) T=4000 (j) T=5000 (k) T=6000 (l) T=6226

Figure 3: Sequence of images showing Poincaré recurrence in an 8×8 zero-sum polymatrix game, where the changing color of each pixel on the grid represents the strategy of the player over time.

To generate Figure 3, we show the data from a simplified 64-player polymatrix game simulation, where the graph that represents player interactions is sparse. Here, the strategy of each player informs the RGB value of a corresponding pixel on a grid. If the system exhibits Poincaré recurrence, we should eventually see similar patterns emerge as the pixels change color over time (i.e., as their corresponding strategies evolve). As observed in Figure 3, the system returns near the initial image after 6226 iterations.

7 DISCUSSION

In this paper we study both GDA and FTRL learning dynamics in periodically varying zero-sum games. We prove that the recurrent nature of such dynamics carries over from static games to the classes of evolving games we study. Yet, in the settings we analyze, the time-average convergence behavior of static zero-sum games can fail to generalize. This work takes a step toward understanding the behavior of classical learning algorithms for games in the more realistic setting where the game itself is not fixed. An important question for future work is developing an understanding of dynamics in even more general dynamic games.

ACKNOWLEDGMENTS

Stratis Skoulakis gratefully acknowledges NRF 2018 Fellowship NRF-NRFF2018-07. Tanner Fiez acknowledges support from the DoD NDSEG Fellowship. Ryann Sim gratefully acknowledges support from the SUTD President's Graduate Fellowship (SUTD-PGF). Lillian Ratliff is supported by NSF CAREER Award number 1844729 and an Office of Naval Research Young Investigor Award. Georgios Piliouras gratefully acknowledges support from grant PIE-SGP-AI-2018-01, NRF2019-NRF-ANR095 ALIAS grant and NRF 2018 Fellowship NRF-NRFF2018-07.

REFERENCES

[1] Vladimir Igorevich Arnol'd. 2013. *Mathematical methods of classical mechanics.* Vol. 60. Springer Science & Business Media.

[2] James P. Bailey, Gauthier Gidel, and Georgios Piliouras. 2020. Finite Regret and Cycles with Fixed Step-Size via Alternating Gradient Descent-Ascent. In *COLT.*

[3] James P. Bailey and Georgios Piliouras. 2018. Multiplicative Weights Update in Zero-Sum Games. In *ACM Conference on Economics and Computation.*

[4] James P Bailey and Georgios Piliouras. 2019. Multi-Agent Learning in Network Zero-Sum Games is a Hamiltonian System. In *International Conference on Autonomous Agents and MultiAgent Systems.* 233–241.

[5] Luis Barreira. 2006. Poincaré recurrence: old and new. In *XIVth International Congress on Mathematical Physics.* World Scientific, 415–422.

[6] Victor Boone and Georgios Piliouras. 2019. From Darwin to Poincaré and von Neumann: Recurrence and Cycles in Evolutionary and Algorithmic Game Theory. In *International Conference on Web and Internet Economics.* 85–99.

[7] G.W. Brown. 1951. Iterative Solutions of Games by Fictitious Play. *In Activity Analysis of Production and Allocation, T.C. Koopmans (Ed.), New York: Wiley.* (1951).

[8] Yang Cai and Constantinos Daskalakis. 2011. On minmax theorems for multiplayer games. In *Symposium of Discrete Algorithms.* 217–234.

[9] Adrian Rivera Cardoso, Jacob Abernethy, He Wang, and Huan Xu. 2019. Competing against equilibria in Zero-Sum Games with evolving payoffs. *arXiv preprint arXiv:1907.07723* (2019).

[10] Nicolò Cesa-Bianchi and Gábor Lugoisi. 2006. *Prediction, Learning, and Games.* Cambridge University Press.

[11] Yun Kuen Cheung. 2018. Multiplicative weights updates with constant step-size in graphical constant-sum games. *Advances in Neural Information Processing Systems* 31 (2018), 3528–3538.

[12] 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 *COLT.*

[13] Benoit Duvocelle, Panayotis Mertikopoulos, Mathias Staudigl, and Dries Vermeulen. 2018. Learning in time-varying games. *arXiv preprint arXiv:1809.03066* (2018).

[14] Yoav Freund and Robert E Schapire. 1999. Adaptive game playing using multiplicative weights. *Games and Economic Behavior* 29, 1-2 (1999), 79–103.

[15] Ian Goodfellow, Jean Pouget-Abadie, Mehdi Mirza, Bing Xu, David Warde-Farley, Sherjil Ozair, Aaron Courville, and Yoshua Bengio. 2014. Generative adversarial nets. In *Advances in Neural Information Processing Systems.* 2672–2680.

[16] J. Hofbauer. 1996. Evolutionary dynamics for bimatrix games: A Hamiltonian system? *J. of Math. Biology* 34 (1996), 675–688.

[17] Tung Mai, Milena Mihail, Ioannis Panageas, Will Ratcliff, Vijay Vazirani, and Peter Yunker. 2018. Cycles in Zero-Sum Differential Games and Biological Diversity. In *ACM Conference on Economics and Computation.* 339–350.

[18] Panayotis Mertikopoulos, Christos Papadimitriou, and Georgios Piliouras. 2018. Cycles in adversarial regularized learning. In *Symposium of Discrete Algorithms.* 2703–2717.

[19] P. Mertikopoulos, H. Zenati, B. Lecouat, C.-S. Foo, V. Chandrasekhar, and G. Piliouras. 2018. Mirror descent in saddle-point problems: Going the extra (gradient) mile. *ArXiv e-prints* (July 2018). arXiv:1807.02629

[20] Sai Ganesh Nagarajan, David Balduzzi, and Georgios Piliouras. 2020. From chaos to order: Symmetry and conservation laws in game dynamics. In *International Conference on Machine Learning.* PMLR, 7186–7196.

[21] Julien Perolat, Remi Munos, Jean-Baptiste Lespiau, Shayegan Omidshafiei, Mark Rowland, Pedro Ortega, Neil Burch, Thomas Anthony, David Balduzzi, Bart De Vylder, Georgios Piliouras, Marc Lanctot, and Karl Tuyls. 2020. From Poincaré Recurrence to Convergence in Imperfect Information Games: Finding Equilibrium via Regularization. *arXiv preprint arXiv:2002.08456* (2020).

[22] Georgios Piliouras and Jeff S Shamma. 2014. Optimization despite chaos: Convex relaxations to complex limit sets via Poincaré recurrence. In *Symposium of Discrete Algorithms.* 861–873.

[23] Henri Poincaré. 1890. Sur le problème des trois corps et les équations de la dynamique. *Acta mathematica* 13, 1 (1890).

[24] J. Robinson. 1951. An Iterative Method of Solving a Game. *Annals of Mathematics* 54 (1951), 296–301.

[25] Shai Shalev-Shwartz et al. 2011. Online learning and online convex optimization. *Foundations and trends in Machine Learning* 4, 2 (2011), 107–194.

[26] David Silver, Aja Huang, Chris J Maddison, Arthur Guez, Laurent Sifre, George Van Den Driessche, Julian Schrittwieser, Ioannis Antonoglou, Veda Panneershelvam, Marc Lanctot, et al. 2016. Mastering the game of Go with deep neural networks and tree search. *nature* 529, 7587 (2016), 484–489.

[27] Stratis Skoulakis, Tanner Fiez, Ryann Sim, Georgios Piliouras, and Lillian Ratliff. 2021. Evolutionary Game Theory Squared: Evolving Agents in Endogenously Evolving Zero Sum Games. *AAAI Conference on Artificial Intelligence* (2021).

[28] Emmanouil-Vasileios Vlatakis-Gkaragkounis, Lampros Flokas, and Georgios Piliouras. 2019. Poincaré Recurrence, Cycles and Spurious Equilibria in Gradient-Descent-Ascent for Non-Convex Non-Concave Zero-Sum Games. In *Advances in Neural Information Processing Systems.* 10450–10461.

[29] John von Neumann. 1928. Zur Theorie der Gesellschaftsspiele. *Math. Ann.* 100 (1928), 295–300.