ONLINE LEARNING IN PERIODIC ZERO-SUM GAMES

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

Ryann Sim[*]
SUTD
Singapore
ryann_sim@mymail.sutd.edu.sg

Stratis Skoulakis[*]
SUTD
Singapore
efstratios@sutd.edu.sg

Georgios Piliouras[†]
SUTD
Singapore
georgios@sutd.edu.sg

Lillian Ratliff[†]
University of Washington
Seattle, Washington
ratliffl@uw.edu

ABSTRACT

A seminal result in game theory is von Neumann's minmax theorem, which states that zero-sum games admit an essentially unique equilibrium solution. Classical learning results build on this theorem to show that online no-regret dynamics converge to an equilibrium in a time-average sense in zero-sum games. In the past several years, a key research direction has focused on characterizing the day-to-day behavior of such dynamics. General results in this direction show that broad classes of online learning dynamics are cyclic, and formally Poincaré recurrent, in zero-sum games. We analyze the robustness of these online learning behaviors in the case of periodic zero-sum games with a time-invariant equilibrium. This model generalizes the usual repeated game formulation while also being a realistic and natural model of a repeated competition between players that depends on exogenous environmental variations such as time-of-day effects, week-to-week trends, and seasonality. Interestingly, time-average convergence may fail even in the simplest such settings, in spite of the equilibrium being fixed. In contrast, using novel analysis methods, we show that Poincaré recurrence provably generalizes despite the complex, non-autonomous nature of these dynamical systems.

1 Introduction

The study of learning dynamics in zero-sum games is arguably as old of a field as game theory itself, dating back to the seminal work of Brown and Robinson [7, 27], which followed shortly after the foundational minmax theorem of von Neumann [33]. The dynamics of online no-regret learning algorithms [10, 29] are of particular interest in zero-sum games as they are designed with an adversarial environment in mind. Moreover, well known results imply that such dynamics converge in a time-average sense to a minmax equilibrium in zero-sum games [10, 15].

Despite the classical nature of the study of online no-regret learning dynamics in zero-sum games, the actual transient behavior of such dynamics was historically not as understood. However, in the past several years this topic has gained attention with a number of works studying such dynamics in zero-sum games (and variants thereof) with a particular focus on continuous-time analysis [24, 25, 20, 6, 32, 23, 22]. The unifying emergent picture is that the dynamics are "approximately cyclic" in a formal sense known as Poincaré recurrence. Moreover, these results have acted as fundamental building blocks for understanding the limiting behavior of their discrete-time variants [2, 12, 21, 11, 3].

Despite the plethora of emerging results regarding online learning dynamics in zero-sum games, an important and well motivated aspect of this problem has only begun to receive attention.

How do online learning dynamics behave if the zero-sum game evolves over time?

Clearly, the answer to this question depends critically on how the game is allowed to evolve.

---

[*]Joint first authors
[†]Joint last authors

Problem Setting and Model. We study periodic zero-sum games with a time-invariant equilibrium, which is a class of games we formally define in Section 2. In a periodic zero-sum game, the payoffs that dictate the game are both T-periodic and zero-sum at all times. We consider both periodic zero-sum bilinear games on infinite, unconstrained strategy spaces and periodic zero-sum matrix games (along with network generalizations thereof) on finite strategy spaces. The goal of this work is to evaluate the robustness of the archetypal online learning behaviors in zero-sum games, Poincaré recurrence and time-average equilibrium convergence, to this natural model of game evolution.

Connections to Repeated Game Models. The time-evolving game model we study can be seen as a generalization of usual repeated game formulations. A time-invariant game is a trivial version of a periodic game, in which case we recover the repeated static game setting. For a general periodic zero-sum game with period T, each stage game now is chosen according to a fixed length T sequence of games, capturing interactions between the players with time-dependent payoffs.

Periodic zero-sum games can also fit into the frameworks of multi-agent contextual games [28] and dynamic games (see, e.g., [5]). In a multi-agent contextual game [28], the environment selects a context from a set before each round of play and this choice defines the game that is played. Periodic zero-sum games can be seen as a multi-agent contextual game where the environment draws contexts from the available set in a T-periodic fashion with each context defining a zero-sum game with a common equilibrium. In the class of dynamic games, there is a game state on which the payoffs may depend that evolves with dynamics. Periodic zero-sum games can be interpreted as a dynamic game where the state transitions do not depend on the strategies of the players, the state is T-periodic, and the payoffs are completely defined by the state. We remark that the focus of existing work on contextual games and dynamic games is distinct from the questions investigated in this paper.

The periodic zero-sum game model allows us to capture competitive settings where exogenous environmental variations manifest in an effectively periodic/epochal fashion. This naturally occurs in market competitions where time-of-day effects, week-to-week trends, and seasonality can dictate the game between players. To illustrate this point, consider a competition between service providers that wish to maximize their users, while the total market size evolves seasonally over time. This evolution affects the utility functions, even if the fundamentals of the market, and consequently the equilibrium, remain invariant.

Contributions and Approach. In this paper, for the classes of periodic zero-sum bilinear games and periodic zero-sum polymatrix games with time-invariant equilibrium, we investigate the day-to-day and time-average behaviors of continuous-time gradient descent-ascent (GDA) and follow-the-regularized-leader (FTRL) learning dynamics, respectively. This study highlights the careful attention that must be given to the dynamical systems in periodic zero-sum games which preclude standard proof techniques for Poincaré recurrence, while also revealing that intuition from existing results on static zero-sum games can be totally invalidated even by simple restricted examples in periodic zero-sum games.

Contribution 1: Poincaré Recurrence. A key technical challenge in this work is that the dynamical systems which emerge from learning dynamics in periodic zero-sum games correspond to *non-autonomous* ordinary differential equations, whereas learning dynamics in static zero-sum games correspond to *autonomous* ordinary differential equations. Consequently, the usual proof methods from static zero-sum games for showing Poincaré recurrence are insufficient on their own in periodic zero-sum games. We overcome this challenge by delicately piecing together properties of periodic systems to construct a discrete-time autonomous system that we are able to show is Poincaré recurrent. This approach allows to prove both the GDA and FTRL learning dynamics are Poincaré recurrent in the respective classes of periodic zero-sum games (Theorems 1 & 2). Finally, we show both periodicity and a time-invariant equilibrium are necessary for such results in evolving games (Proposition 1).

Contribution 2: Time-Average Strategy Equilibration Fails. Given that Poincaré recurrence provably generalizes from static zero-sum games to periodic zero-sum games, it may be expected that the time-average strategies in periodic zero-sum games converge to the time-invariant equilibrium as in static zero-sum games. Surprisingly, we show that counterexamples can be constructed to this intuition even in the simplest of periodic zero-sum games. In particular, we prove the negative result that the time-average GDA and FTRL strategies do not necessarily converge to the time-invariant equilibrium in the respective classes of zero-sum games (Propositions 2 & 3).

Contribution 3: Time-Average Equilibrium Utility Convergence. Despite the negative result on the time-average strategy convergence, in the special case of periodic zero-sum bimatrix games we are able to show a complimentary positive result on the time-average utility convergence. Specifically, we show that the time-average utilities of the FTRL learning dynamics converge to the average of the equilibrium utility values of all the zero-sum games included in a single period of our time-evolving games. (Theorem 3).

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 numerical experiments in Section 6 and finish with a discussion in Section 7. Proofs of are theoretical results are deferred to the appendix.

2 Game-Theoretic Preliminaries

2.1 Continuous Strategy Periodic Zero-Sum Games

For continuous strategy periodic zero-sum games, we study periodic zero-sum bilinear games. We begin by formalizing zero-sum bilinear games 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 game is defined by the pair of payoff matrices $\{A, -A^\top\}$ and the action space of agents 1 and 2 are given by $\mathbb{R}^{n_1}$ and $\mathbb{R}^{n_2}$, respectively. 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. 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 game remains zero-sum and the continuous-time sequence of payoffs is periodic. The next definition formalizes this class of games.

Definition 1 (Periodic Zero-Sum Bilinear Game). *A periodic zero-sum bilinear game is an infinite sequence of zero-sum bilinear games $\{A(t), -A(t)^\top\}_{t=0}^\infty$ in which the player set and strategy spaces are fixed and the payoff matrix is such that $A(t) = A(t + T)$ for a finite period T and all $t \geq 0$. Note that in such a game, $(0, 0)$ is always a time-invariant Nash equilibrium. Furthermore, we assume that the dependence of the payoff entries on time is smooth everywhere except for a finite set of points.*

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 is the player set and E is the edge set where a bimatrix game is played between the endpoints of each edge [8]. Each player $i \in V$ 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 mixed strategy set of player $i \in V$ is the simplex in $\mathbb{R}^{n_i}$ denoted by $\mathcal{X}_i = \Delta^{n_i - 1} = \{x_i \in \mathbb{R}_{\geq 0}^{n_i} : \sum_{\alpha \in \mathcal{A}_i} x_{i\alpha} = 1\}$ where $x_{i\alpha}$ denotes the probability of action $\alpha \in \mathcal{A}_i$. The joint strategy space is denoted by 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 $A^{ji} \in \mathbb{R}^{n_j \times n_i}$. The utility or payoff of agent $i \in V$ under the strategy profile $x \in \mathcal{X}$ is given by $u_i(x) = \sum_{j:(i,j) \in E} x_i^\top A^{ij} x_j$ and corresponds to the sum of payoffs from the bimatrix games the player participates in. 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 $x \in \mathcal{X}$. Each bimatrix edge game is not necessarily zero-sum in a zero-sum polymatrix game.

A *Nash equilibrium* in a polymatrix game is a mixed strategy profile $x^* \in \mathcal{X}$ such that for each player $i \in V$, $u_i(x_i^*, x_{-i}^*) \geq u_i(x_i, x_{-i}^*), \ \forall x_i \in \mathcal{X}_i$. A Nash equilibrium is said to be an interior if $\text{supp}(x_i^*) = \mathcal{A}_i \ \forall i \in V$ where $\text{supp}(x_i^*) = \{\alpha \in \mathcal{A}_i : x_{i\alpha} > 0\}$ is the support of $x_i^* \in \mathcal{X}_i$.

Periodic Zero-Sum Polymatrix Games. We analyze the continuous-time `FTRL` learning dynamics in a class of games we call periodic zero-sum polymatrix games. This class of games is such that the payoffs defined by the edge games evolve periodically. We consider that this periodic evolution is such that there is a common interior Nash equilibrium that arises in each zero-sum 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 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 $t \geq 0$. We assume there is a common interior Nash equilibrium $x^* \in \mathcal{X}$ of the polymatrix game $G(t)$ for all $t \geq 0$. Furthermore, we assume that the dependence of the payoff entries on time is smooth everywhere except for a finite set of points.*

3 Preliminaries on Dynamical Systems

We now cover concepts from dynamical systems theory that will help us analyze learning dynamics in periodic zero-sum games and prove Poincaré recurrence. Careful attention must be given to these preliminaries in this work since the dynamical systems we study are non-autonomous whereas typical recurrence analysis in the study of learning in games deals with autonomous dynamical systems.

3.1 Background on Dynamical Systems

We begin this section by providing dynamical systems background that is necessary both for defining Poincaré recurrence and sketching typical proof methods along with our approach.

Flows. Consider an ordinary differential equation $\dot{x} = f(t, x)$ on a topological space X. We can define the *flow* $\phi : \mathbb{R} \times X \to X$ of a dynamical system $\dot{x}$, for which the following holds: (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 $x \in X$, $\frac{d}{dt}|_{t=0}\phi(t, x) = f(t, x)$, and (iv) $\phi(t, x_0) = x(t)$ is the solution.

Existence and Uniqueness. We utilize Carathéodory's existence theorem to guarantee the existence of a flow for $\dot{x}$, even for some discontinuous functions f.

Theorem (Carathéodory's existence theorem [13, 16]). *Consider a differential equation $\dot{x} = f(t, x)$ on a rectangular domain $R = \{(t, y)| \ |t - t_0| \leq a, |x - x_0| \leq b\}$. If f satisfies the following conditions:*

1. *$f(t, x)$ is continuous in y for each fixed t,*

2. *$f(t, x)$ is measurable in t for each fixed y,*

3. *there is a Lebesgue-integrable function $m : [t_0 - a, t_0 + a] \to [0, \infty)$ such that $|f(t, x)| \leq m(t)$ for all $(t, x) \in R$,*

then the differential equation has a solution. Moreover, if f is also Lipschitz continuous, meaning $|f(t, x_1) - f(t, x_2)| \leq k(t)|x_1 - x_2|$ with some Lebesgue-integrable function $k : [t_0 - a, t_0 + a] \to [0, \infty)$, then there exists a unique solution of the differential equation.

In the settings we study, the above three conditions hold: Condition 1 holds because for every fixed t, the dynamics we study (specifically GDA and FTRL) are continuous functions of their state space. Condition 2 holds because the systems we study are finite and continuous almost-everywhere, and so by Lusin's theorem [18] are measurable for each fixed y. Finally, Condition 3 is always satisfied because the games we study always admit bounded orbits. Hence, it follows that a unique flow exists for all the dynamical systems studied in this paper.

Conservation of Volume. The flow ϕ of an 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, meaning that $\text{vol}(\phi^t(U)) = \text{vol}(U)$. *Liouville's theorem* states that a flow is volume preserving if the divergence of f at any point $x \in \mathbb{R}^d$ equals zero: that is, $\text{div} f(t, x) = \text{tr}(Df(t, x)) = \sum_{i=1}^{d} \frac{\partial f(t, x)}{\partial x_i} = 0$.

We now transition to give general Poincaré recurrence statements along with discussion of how the results are usually applied in game theory before we outline our proof methods.

3.2 Poincaré Recurrence in Autonomous Dynamical Systems

A number of works in the past several years show that online no-regret learning dynamics are Poincaré recurrent in repeated static zero-sum games (see, e.g., [24, 20, 6]). The proof methods for deriving such results crucially rely on the static nature of the game for the reason that the learning dynamics amount to an autonomous dynamical system. Informally, the standard Poincaré recurrence theorem states that 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 [26]; this property of a dynamical system is known as *Poincaré recurrence*. Thus, proving the Poincaré recurrence of dynamics in repeated static zero-sum games is tantamount to verifying the volume preservation and bounded orbit properties. We now formalize the Poincaré recurrence theorem.

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 (Poincaré Recurrence for Continuous-Time Systems [26]). *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 $\Omega(\phi^t) = Z$.*

In order to describe our proof methods in the following subsection for showing Poincaré recurrence in periodic zero-sum games, it will be useful for us to state an alternative formulation of the Poincaré recurrence theorem that is applicable to autonomous discrete-time systems.

Theorem (Poincaré Recurrence for Discrete-Time Maps [4]). *Let (X, Σ, μ) be a finite measure space and let $\phi \colon X \to X$ 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. *That is, $\mu\left(\{x \in E : \exists\, N \text{ s.t. } \phi^n(x) \notin E \text{ for all } n > N\}\right) = 0$.*

3.3   Poincaré Recurrence in Periodic Dynamical Systems

The proof methods described in the last section for showing the Poincaré recurrence of dynamics in static zero-sum games cannot directly be applied to time-evolving zero-sum games as a result of the non-autonomous nature of the systems. In fact, we can construct time-evolving zero-sum games without both periodic payoffs and a time-invariant equilibrium, where online learning dynamics are not Poincaré recurrent.[3] Despite this hurdle, we show that in the natural subclass of time-evolving games covered by periodic zero-sum games (periodic payoffs and a time-invariant equilibrium), we can develop proof methods to show the Poincaré recurrence of online learning dynamics. Given the previous claim regarding the possible non-recurrent nature of learning dynamics when there is not both periodic payoffs and a time-invariant equilibrium, this is perhaps the most general class of time-evolving zero-sum games with obtainable positive results in this direction.

We now give an overview of our approach, beginning by recalling properties of periodic systems.

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* or the *mapping at a period*.

If the differential equation is well-defined for all x and has a solution 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 autonomous dynamical system $x^+ = \phi^T(x)$. The learning dynamics we study in periodic zero-sum games form T-periodic dynamical systems. Thus, the discrete-time autonomous dynamical system $x^+ = \phi^T(x)$ formed by the Poincaré map is key to the analysis methods we pursue. In particular, our approach is to show that this system is Poincaré recurrent, which we then use to conclude that the original continuous-time non-autonomous system is Poincaré recurrent.

Given the previously presented Poincaré recurrence theorem for discrete-time maps, proving the Poincaré recurrence of the system $x^+ = \phi^T(x)$ requires verifying the volume preservation and bounded orbit properties, which then implies the measure preserving property. 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 and the resulting discrete-time dynamical system $x^+ = \phi^T(x)$ is also volume preserving.

Theorem (Volume preservation for T-Periodic Systems [1, 3.16.B, Thm 2]). *If the T-periodic system $\dot{x} = f(t, x)$ is divergence-free, then ϕ^T preserves volume.*

Similarly, if orbits of $\dot{x} = f(t, x)$ are bounded, then clearly the orbits of $x^+ = \phi^T(x)$ are bounded. Hence, to show the system $x^+ = \phi^T(x)$ is Poincaré recurrent, we prove $\dot{x} = f(t, x)$ has a divergence-free vector-field (equivalently, that the flow is volume preserving) and only bounded orbits. This will then be sufficient to conclude $\dot{x} = f(t, x)$ is Poincaré recurrent since the discrete-time system forms a subsequence of the continuous-time system.

---

[3]We formalize this statement in Proposition 1 of the following section.

4 Gradient Descent-Ascent in Periodic Zero-Sum Bilinear Games

This section focuses on 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 and are given by

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

4.1 Poincaré Recurrence

The focus of this section is on characterizing the transient behavior of the continuous-time GDA learning dynamics in periodic zero-sum games. Specifically, we show the following result.

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

Theorem 1 establishes that the recurrent nature of continuous-time GDA dynamics in static zero-sum bilinear games is robust to the dynamic evolution of the payoffs in periodic zero-sum bilinear games.

Prior to outlining the proof steps for Theorem 1, we elaborate on the claim from the previous section that without the periodicity property and a time-invariant equilibrium, such a result is unobtainable. In particular, we show the Poincaré recurrence of the GDA dynamics is not guaranteed without both properties by constructing counterexamples when only one of the properties holds.

Proposition 1. *There exists time-evolving zero-sum games such that there is a time-invariant equilibrium or the payoffs are periodic (but not both simultaneously) in which the* GDA *dynamics are not Poincaré recurrent.*

This proposition highlights the strength of our results regarding GDA, given that the assumptions needed to obtain them are more or less tight.

We now outline the key intermediate results we prove to obtain Theorem 1, following the techniques described in Section 3.3. For the GDA dynamics, we utilize the observation that the corresponding vector fields are divergence free to show that the learning 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 then proceed by showing that the GDA orbits are bounded by deriving a time-invariant function. This step relies on the fact that we have a time-invariant equilibrium.

Lemma 2. *The function* $\Phi(t) = \frac{1}{2}\left(x_1^\top(t)x_1(t) + x_2^\top(t)x_2(t)\right)$ *is time-invariant. Hence, the* GDA *orbits are bounded in any periodic zero-sum bilinear game as given in Definition 1.*

Given the volume preservation and bounded orbit characteristics of the continuous-time GDA learning dynamics in periodic zero-sum games, the proof of recurrence follows by applying the arguments described in Section 3.3.

4.2 Time-Average Convergence

The Poincaré recurrence 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.

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. To prove this, we consider a periodic zero-sum bilinear game with the action space of each player on $\mathbb{R}$ so that the evolving payoff simply rescales the vector field. We construct the payoff sequence so that the dynamics return back to the initial condition after a period of the game, while the time-average of the dynamics are not equal to the time-invariant equilibrium $(x_1^*, x_2^*) = (\mathbf{0}, \mathbf{0})$. Given the simplicity of this example, it effectively rules out hope to provide a meaningful time-average convergence guarantee in this class of games.

Proposition 2. *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* $(\mathbf{0}, \mathbf{0})$.

5 Follow-the-Regularized-Leader in Periodic Zero-Sum Polymatrix Games

We now analyze continuous-time FTRL learning dynamics in periodic zero-sum polymatrix games. Players 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 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

$$y_i(t) = y_i(0) + \int_0^t \sum_{j:(i,j) \in E} A^{ij}(\tau) x_j(\tau) d\tau$$
$$x_i(t) = \mathrm{argmax}_{x_i \in \mathcal{X}_i} \{\langle x_i, y_i(t) \rangle - h_i(x_i)\}$$

(1)

where $h_i : \mathcal{X}_i \to \mathbb{R}$ is a 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 ensure the update $x_i(t)$ is well-defined since a unique solution exists.

Common FTRL learning dynamics include 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. 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

$$y_i(t) = y(0) + \int_0^t v_i(x(\tau), \tau) d\tau$$
$$x_i(t) = Q_i(y_i(t)).$$

(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

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

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

5.1 Poincaré Recurrence

We now focus on characterizing the transient behavior of the continuous-time FTRL learning dynamics in periodic zero-sum polymatrix games. It is known that the continuous-time FTRL learning dynamics are Poincaré recurrent in static zero-sum polymatrix games [20]. The following result demonstrates that this characteristic holds even in games that are evolving in a periodic fashion with a time-invariant equilibrium, providing a broad generalization.

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

For the remainder of this subsection, we describe our proof methods. The general approach is that we prove the Poincaré recurrence of a transformed system using the techniques described in Section 3.3. This conclusion then allows us to infer the equivalent property for the original FTRL system.

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

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

Toward proving that this system is Poincaré recurrent, 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.*

We then construct a time-invariant function along the evolution of the system that is sufficient to guarantee that the orbits generated by the $\dot{z}$ dynamics are bounded. Recall that x^* denotes the common, time-invariant interior Nash equilibrium of the periodic zero-sum polymatrix game.

Lemma 4. *The function* $\Phi(x^*, y(t)) = \sum_{i \in V} \left(h_i^*(y_i(t)) - \langle x_i^*, y_i(t) \rangle + h_i(x_i^*) \right)$ *is time-invariant. Hence, 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 the arguments from Section 3.3 to conclude the $\dot{z}$ dynamics are Poincaré recurrent. Finally, we show that Poincaré recurrence of the $\dot{z}$ system is sufficient to guarantee the Poincaré recurrence of the FTRL learning dynamics.

The proofs of the results in this section can be found in Appendix D.

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 in zero-sum polymatrix games. Consequently, in general, time-average convergence to the set of equilibrium values in the utility and strategy spaces does not equate to the stronger notion of pointwise convergence.

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

Theorem 3. *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 game equilibrium utility values.*

Theorem 3 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 time-invariant Nash equilibrium.

Proposition 3. *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.*

We prove the negative result in this proposition by constructing a simple counterexample that corresponds to a time-varying rescaling of Matching Pennies.

The proofs of the results in this section can be found in Appendix E.

6 Experiments

In this section, we present several experimental simulations that illustrate our theoretical results. To begin, for continuous-time GDA dynamics we show that Poincaré 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 players follow the GDA learning dynamics, we see from Figure 1 that the trajectories when plotted alongside the value of the periodic rescaling are bounded. A similar experimental result holds in the case of FTRL dynamics. In the supplementary material, we simulate replicator dynamics with the same periodic rescaling as in Figure 1. The trajectories in the dual/payoff space also remain bounded due to the invariance of the Kullback-Leibler divergences (KL-divergence).

Lemmas 2 and 4 describe functions Φ which remain time-invariant. In the case of replicator dynamics, $\Phi(t)$ is the sum of Kullback-Leibler divergences measured between the strategy of each player and their mixed Nash strategy $[1/2, 1/2]$. 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.

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
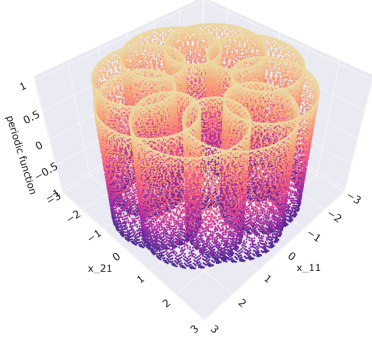
Figure 1: Bounded trajectories for a periodically rescaled Matching Pennies game updated using GDA.
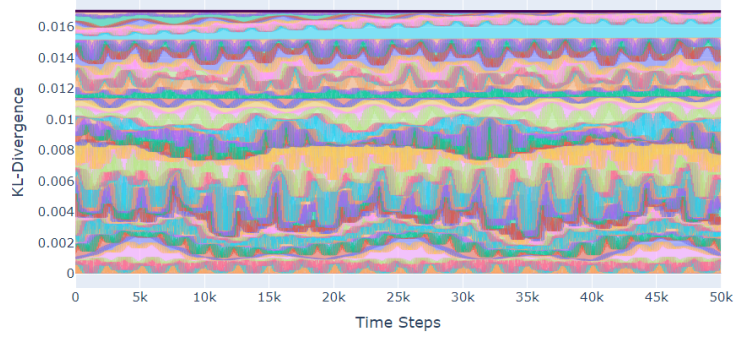

Figure 2: Weighted sum of KL-divergences for a 64-player periodically rescaled Matching Pennies game. Note that despite the complicated trajectories of each player, the weighted sum of their divergences remains constant.
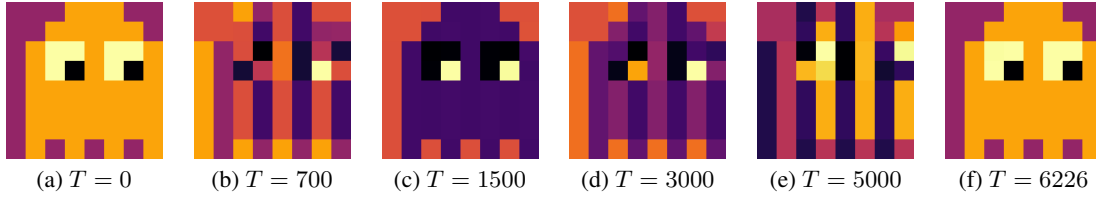

(a) $T = 0$ (b) $T = 700$ (c) $T = 1500$ (d) $T = 3000$ (e) $T = 5000$ (f) $T = 6226$

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 mixed strategy of the player over time. After time $T = 6226$, we see that an approximation of the original image is recovered, showing that the recurrence property holds.

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 mixed strategies evolve). As observed in Figure 3, the system returns near the initial image at time $T = 6226$. Further details about the experiments can be found in Appendix F.

7 Discussion

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 from 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.

We conclude by discussing related works on learning dynamics in evolving games. The existing literature considers a number of models that admit distinct results [19, 30, 9, 14]. In a class of time-evolving games where the evolution can be arbitrary [9], algorithms are designed that provide a novel type of regret guarantee called Nash equilibrium regret. In a sense these algorithms are competitive against the Nash equilibrium of the long-term-averaged payoff matrix. In an analysis of discrete-time FTRL dynamics in evolving games that are strictly/strongly monotone [14], sufficient conditions (e.g., when the evolving game stabilizes) under which the dynamics track/converge to the evolving equilibrium are derived. Unfortunately, zero-sum games do not satisfy these strong properties. Finally, in a model of endogenously evolving 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 has been developed [19, 30] under which both time-average convergence to equilibrium as well as Poincaré recurrence holds.

This growing literature indicates that time-varying games can exhibit distinct and often times more complex behavior than their classic static counterparts. As such, there is much potential for future work towards a better understanding of their learning dynamics.

Acknowledgements

References

[1] V. I. Arnol'd. *Mathematical methods of classical mechanics*, volume 60. Springer Science & Business Media, 2013.

[2] J. P. Bailey and G. Piliouras. Multiplicative weights update in zero-sum games. In *ACM Conference on Economics and Computation*, pages 321–338, 2018.

[3] J. P. Bailey, G. Gidel, and G. Piliouras. Finite regret and cycles with fixed step-size via alternating gradient descent-ascent. In *Conference on Learning Theory*, pages 391–407, 2020.

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

[5] T. Başar and G. J. Olsder. *Dynamic noncooperative game theory*. SIAM, 1998.

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

[7] G. Brown. Iterative solutions of games by fictitious play. *In Activity Analysis of Production and Allocation, T.C. Koopmans (Ed.), New York: Wiley.*, 1951.

[8] Y. Cai and C. Daskalakis. On minmax theorems for multiplayer games. In *Symposium of Discrete Algorithms*, pages 217–234, 2011.

[9] A. R. Cardoso, J. Abernethy, H. Wang, and H. Xu. Competing against nash equilibria in adversarially changing zero-sum games. In *International Conference on Machine Learning*, pages 921–930, 2019.

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

[11] Y. K. Cheung. Multiplicative weights updates with constant step-size in graphical constant-sum games. In *Advances in Neural Information Processing Systems*, pages 3528–3538, 2018.

[12] Y. K. Cheung and G. Piliouras. Vortices instead of equilibria in minmax optimization: Chaos and butterfly effects of online learning in zero-sum games. In *Conference on Learning Theory*, pages 807–834, 2019.

[13] E. A. Coddington and N. Levinson. *Theory of ordinary differential equations*. Tata McGraw-Hill Education, 1955.

[14] B. Duvocelle, P. Mertikopoulos, M. Staudigl, and D. Vermeulen. Learning in time-varying games. *arXiv preprint arXiv:1809.03066*, 2018.

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

[16] J. K. Hale. *Ordinary differential equations*. Robert E. Krieger Publishing Company, 1980.

[17] M. Heino, J. A. Metz, and V. Kaitala. The enigma of frequency-dependent selection. *Trends in Ecology & Evolution*, 13(9):367–370, 1998.

[18] N. Lusin. Sur les propriétés des fonctions mesurables. *CR Acad. Sci. Paris*, 154(25):1688–1690, 1912.

[19] T. Mai, M. Mihail, I. Panageas, W. Ratcliff, V. Vazirani, and P. Yunker. Cycles in Zero-Sum Differential Games and Biological Diversity. In *ACM Conference on Economics and Computation*, page 339–350, 2018.

[20] P. Mertikopoulos, C. Papadimitriou, and G. Piliouras. Cycles in adversarial regularized learning. In *Symposium of Discrete Algorithms*, pages 2703–2717, 2018.

[21] P. Mertikopoulos, B. Lecouat, H. Zenati, C.-S. Foo, V. Chandrasekhar, and G. Piliouras. Optimistic mirror descent in saddle-point problems: Going the extra (gradient) mile. In *International Conference on Learning Representations*, 2019.

[22] S. G. Nagarajan, D. Balduzzi, and G. Piliouras. From chaos to order: Symmetry and conservation laws in game dynamics. In *International Conference on Machine Learning*, pages 7186–7196, 2020.

[23] J. Perolat, R. Munos, J.-B. Lespiau, S. Omidshafiei, M. Rowland, P. Ortega, N. Burch, T. Anthony, D. Balduzzi, B. De Vylder, G. Piliouras, M. Lanctot, and K. Tuyls. From Poincaré Recurrence to Convergence in Imperfect Information Games: Finding Equilibrium via Regularization. *arXiv preprint arXiv:2002.08456*, 2020.

[24] G. Piliouras and J. S. Shamma. Optimization despite chaos: Convex relaxations to complex limit sets via Poincaré recurrence. In *Symposium of Discrete Algorithms*, pages 861–873, 2014.

[25] G. Piliouras, C. Nieto-Granda, H. I. Christensen, and J. S. Shamma. Persistent patterns: Multi-agent learning beyond equilibrium and utility. In *International Conference on Autonomous Agents and Multi-Agent Systems*, pages 181–188, 2014.

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

[27] J. Robinson. An iterative method of solving a game. *Annals of Mathematics*, 54:296–301, 1951.

[28] P. G. Sessa, I. Bogunovic, A. Krause, and M. Kamgarpour. Contextual games: Multi-agent learning with side information. In *Advances in Neural Information Processing Systems*, 2020.

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

[30] S. Skoulakis, T. Fiez, R. Sim, G. Piliouras, and L. Ratliff. Evolutionary game theory squared: Evolving agents in endogenously evolving zero sum games. In *AAAI Conference on Artificial Intelligence*, 2021.

[31] A. R. Tilman, J. B. Plotkin, and E. Akçay. Evolutionary games with environmental feedbacks. *Nature communications*, 11(1):1–11, 2020.

[32] E.-V. Vlatakis-Gkaragkounis, L. Flokas, and G. Piliouras. 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*, pages 10450–10461, 2019.

[33] J. von Neumann. Zur theorie der gesellschaftsspiele. *Mathematische Annalen*, 100:295–300, 1928.

[34] J. S. Weitz, C. Eksin, K. Paarporn, S. P. Brown, and W. C. Ratcliff. An oscillating tragedy of the commons in replicator dynamics with game-environment feedback. *National Academy of Sciences*, 113(47), 2016.

Appendices

Appendix Organization and Contents

The organization and contents of this appendix is as follows. Appendix A covers additional related work, which we provide a separate bibliography for at the end of the document. Following Appendix A, proofs for the theoretical results in the paper are presented in the order that they appeared. Specifically, Appendix B contains the proofs for the results presented in Section 4.1 on GDA dynamics and recurrence. This includes the proofs of Proposition 1, Lemma 1, Lemma 2, and Theorem 1. In Appendix C, we provide the proof of Proposition 2 from Section 4.2 on the time-average behavior of GDA. Appendix D contains the analysis for Section 5.1 on FTRL dynamics and recurrence, including the proofs of Lemma 3, Lemma 4, and Theorem 2. Appendix E covers the time-average behavior of FTRL dynamics as presented in Section 5.2. Specifically, the proofs of Theorem 3 and Proposition 3 are given in Appendix E. Finally, Appendix F contains additional experimental results.

A Additional Related Work

The discussions of related work presented in Section 1 and Section 7 focused on comparable theoretical results in static zero-sum games and studies on classes of certain evolving zero-sum games. We remark that there is also a rich literature studying evolutionary dynamics in action-dependent (endogenous) evolving games in problems strongly motivated by applications in science, economics, and sociology. Often this line of work falls under the hood of what is known as negative frequency dependent selection [17]. Negative frequency dependent selection is an evolutionary process in which the fitness of a strategy dissipates as it becomes more common. Indeed, a number of works (see, [34, 31] and the references therein) analyze evolutionary dynamics in specific formulations, typically in low-dimensional strategy spaces, and characterize the outcomes. In contrast, we study a broad class of learning dynamics in a general class of exogenously evolving games.

B GDA Recurrence Results: Proofs for Section 4.1

This appendix includes the proofs of Proposition 1, Lemma 1, Lemma 2, and Theorem 1.

B.1 Proof of Proposition 1

To prove this result, we construct time-evolving zero-sum games without both periodic payoffs and a time-invariant equilibrium in which the GDA dynamics are not Poincaré recurrent.

Example 1. Consider a time-evolving zero-sum game on scalar action spaces so that $x_1, x_2 \in \mathbb{R}$ with the time-evolving payoff matrix $A(t) = t^{-2}$. Without loss of generality, we can consider $t > 0$ so that the GDA dynamics are well-defined. This time-evolving zero-sum does not have periodic payoffs, but $(x_1^*, x_2^*) = (0, 0)$ is a time-invariant Nash equilibrium. We now show that the GDA dynamics are not Poincaré recurrent in this game. Before formally proving this, we remark that the intuition for why this statement holds is that since the payoff matrix goes to zero, the distance the dynamics can travel is bounded so it is impossible that the trajectory could return arbitrarily close to an initial condition infinitely often.

The GDA dynamics in this time-evolving zero-sum game are described by the system

$$\begin{bmatrix} \dot{x}_1 \\ \dot{x}_2 \end{bmatrix} = \begin{bmatrix} 0 & \frac{1}{t^2} \\ -\frac{1}{t^2} & 0 \end{bmatrix} \begin{bmatrix} x_1(t) \\ x_2(t) \end{bmatrix}.$$

The solution of a time-varying linear system of this form is given by

$$\begin{bmatrix} x_1(t) \\ x_2(t) \end{bmatrix} = \exp\left(\int_{t_0}^{t} \begin{bmatrix} 0 & \frac{1}{\tau^2} \\ -\frac{1}{\tau^2} & 0 \end{bmatrix} d\tau \right) \begin{bmatrix} x_1(t_0) \\ x_2(t_0) \end{bmatrix}.$$

We consider $t_0 > 1$ without loss of generality. To derive the explicit solution, we begin by computing the integral of the evolving payoff matrix and get that

$$\int_{t_0}^{t} \begin{bmatrix} 0 & \frac{1}{\tau^2} \\ -\frac{1}{\tau^2} & 0 \end{bmatrix} d\tau = \begin{bmatrix} 0 & \frac{1}{t_0} - \frac{1}{t} \\ \frac{1}{t} - \frac{1}{t_0} & 0 \end{bmatrix}.$$

Recalling the following identity

$$\exp\left(\theta\begin{bmatrix} 0 & -1 \\ 1 & 0 \end{bmatrix}\right) = \begin{bmatrix} \cos(\theta) & \sin(\theta) \\ -\sin(\theta) & \cos(\theta) \end{bmatrix},$$

we can determine that the matrix exponential is then given by

$$\exp\left(\begin{bmatrix} 0 & \frac{1}{t_0} - \frac{1}{t} \\ \frac{1}{t} - \frac{1}{t_0} & 0 \end{bmatrix}\right) = \begin{bmatrix} \cos(\frac{1}{t} - \frac{1}{t_0}) & -\sin(\frac{1}{t} - \frac{1}{t_0}) \\ \sin(\frac{1}{t} - \frac{1}{t_0}) & \cos(\frac{1}{t} - \frac{1}{t_0}) \end{bmatrix}.$$

Therefore, the solution of the system simplifies to be given by

$$\begin{bmatrix} x_1(t) \\ x_2(t) \end{bmatrix} = \begin{bmatrix} \cos(\frac{1}{t} - \frac{1}{t_0}) & -\sin(\frac{1}{t} - \frac{1}{t_0}) \\ \sin(\frac{1}{t} - \frac{1}{t_0}) & \cos(\frac{1}{t} - \frac{1}{t_0}) \end{bmatrix} \begin{bmatrix} x_1(t_0) \\ x_2(t_0) \end{bmatrix},$$

and equivalently,

$$\begin{bmatrix} x_1(t) \\ x_2(t) \end{bmatrix} = \begin{bmatrix} \cos(\frac{1}{t} - \frac{1}{t_0})x_1(t_0) - \sin(\frac{1}{t} - \frac{1}{t_0})x_2(t_0) \\ \sin(\frac{1}{t} - \frac{1}{t_0})x_1(t_0) + \cos(\frac{1}{t} - \frac{1}{t_0})x_2(t_0) \end{bmatrix}.$$

Given the explicit form of the solution, we now show that we can construct an initial condition that the strategies do not return back to infinitely often. This will immediately allow us to conclude the system is not Poincaré recurrent by definition. We remark that the following choice of initial condition is only for the simplicity of the proof and identical conclusions would hold for almost all initial conditions.

Let $x_1(t_0) = 1$ and $x_2(t_0) = 0$. Given this initial condition, the solution of the system simplifies to be given by

$$\begin{bmatrix} x_1(t) \\ x_2(t) \end{bmatrix} = \begin{bmatrix} \cos(\frac{1}{t} - \frac{1}{t_0}) \\ \sin(\frac{1}{t} - \frac{1}{t_0}) \end{bmatrix}.$$

Taking the limit of the solution as $t \to \infty$, we have that

$$\lim_{t\to\infty} x_1(t) = \lim_{t\to\infty} \cos\left(\frac{1}{t} - \frac{1}{t_0}\right) = \cos\left(\frac{1}{t_0}\right)$$

and

$$\lim_{t\to\infty} x_2(t) = \lim_{t\to\infty} \sin\left(\frac{1}{t} - \frac{1}{t_0}\right) = -\sin\left(\frac{1}{t_0}\right).$$

This shows that the dynamics converge to a fixed point $(\bar{x}_1, \bar{x}_2) = (\cos\left(\frac{1}{t_0}\right), -\sin\left(\frac{1}{t_0}\right))$. Since $(\bar{x}_1, \bar{x}_2) = (\cos\left(\frac{1}{t_0}\right), \sin\left(\frac{1}{t_0}\right)) \neq (1, 0) = (x_1(t_0), x_2(t_0))$ for $t_0 > 1$ unless $t_0 \to \infty$, this immediately implies that the GDA dynamics are not Poincaré recurrent in this time-evolving zero-sum game since they do not return infinitely often back to an arbitrarily small neighborhood around the initial condition.

Example 2. In the previous example, we showed that the GDA dynamics were not Poincaré recurrent in a time-evolving zero-sum game that had a time-invariant Nash equilibrium but not periodic payoffs. In this example, we provide theoretical evidence that the GDA dynamics are not Poincaré recurrent in a time-evolving zero-sum game with periodic payoffs but without a time-invariant Nash equilibrium.

Consider a time-evolving zero-sum game on scalar action spaces so that $x_1, x_2 \in \mathbb{R}$ with a periodic payoff matrix $A(t) = A(t + T)$ for any $t \geq 0$ and $T = 3$ that evolves over a period such that $A(t) = 1$ for $0 \leq t \leq 1$ and $A(t) = -1$ for $1 \leq t \leq 3$. We treat player 2 as a dummy player that just plays the fixed strategy of $x_2 = 1$ for all t. Thus, this time-evolving zero-sum game can be viewed as a trivial game that is equivalent to an optimization problem for player 1. Since player 1 is a utility maximizer, the Nash equilibrium of the game at each time simply corresponds to the strategy of player 1 that maximizes its utility. Thus, the Nash equilibrium when $A(t) = 1$ is $x_1^* = \infty$ and given $A(t) = -1$ it is $x_21^* = -\infty$. Therefore, this corresponds to a time-evolving zero-sum game that is periodic but there is not a time-invariant Nash equilibrium.

We now show that the GDA dynamics are not Poincaré recurrent in this game. The dynamics can be described by the system $\dot{x}_1 = A(t)$. Consequently, the solution is $x_1(t) = x_1(t_0) + t$ on the interval with $A(t) = 1$ and $x_1(t) = x_1(t_0) - 1$ on the interval with $A(t) = -1$. This means that after 1 period of the game, $x_1(t) = x_1(t_0) - 1$, which implies that $x_1(t) \to -\infty$ as $t \to \infty$. Thus the dynamics are not Poincaré recurrent in this time-evolving zero-sum game since they do not return infinitely often back to an arbitrarily small neighborhood around the initial condition.

Example 3. Additionally, we provide an experimental example to show the non-existence of Poincaré recurrence in the setting of FTRL dynamics. In particular, we simulate a time-evolving zero-sum game which has periodic payoffs but
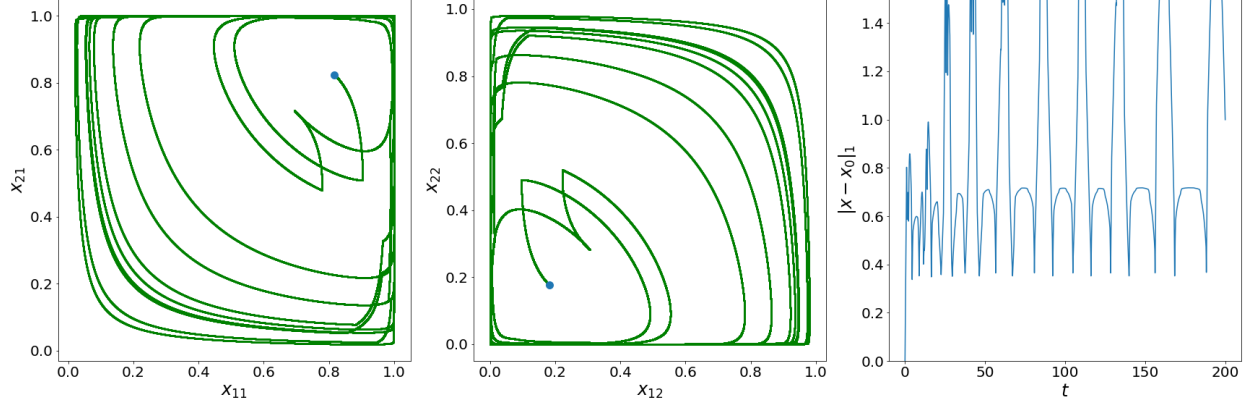
Figure 4: (Left, Center) Replicator trajectories for periodically evolving game without time-invariant equilibrium. (Right) L_1-norm plot showing that recurrence does not hold in this example.

does not have a time-invariant Nash equilibrium. Consider a time-evolving zero-sum game where the payoff matrix for the first quarter of a period is standard Matching Pennies $A = \left(\begin{smallmatrix} 1 & -1 \\ -1 & 1 \end{smallmatrix} \right)$. For the remaining three quarters of the period it is instead $A = \left(\begin{smallmatrix} 0.05 & -0.5 \\ -0.5 & 5 \end{smallmatrix} \right)$. Note that here the payoff matrix for the second player is just $-A$. The former game has a mixed Nash equilibrium where both players play $[0.5, 0.5]$ and the latter game has a mixed Nash equilibrium where both players play $[0.9091, 0.0909]$. We simulate this example with replicator dynamics, which is an instantiation of FTRL dynamics. First, we plot the trajectories of the player's strategies against each other. Moreover, we plot the L_1-norm between the joint trajectory of the players and the initial condition. The simulation results show that the trajectory does not return back arbitrarily close to the initial condition, thus the dynamics are not Poincaré recurrent in this periodic evolving game without a time-invariant equilibrium (Figure 4).

B.2 Proof of Lemma 1

We can show that the GDA dynamics are volume preserving by showing that the vector field is divergence free and then applying Liouville's theorem. Indeed,

$$\text{div}(\dot{x}) = \sum_{i=1}^{2} \sum_{j=1}^{n_i} \frac{\partial \dot{x}_{ij}}{\partial x_{ij}} = 0,$$

which follows from the fact that $\dot{x}_{ij}$ is independent of x_{ij} for each i, j. The divergence free property of the vector field then ensures that the flow ϕ^t of the differential equation is volume preserving by Liouville's theorem.

B.3 Proof of Lemma 2

To prove this statement, we claim the following function is time-invariant:

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

By taking the time-derivative of the $\Phi(t)$ we can verify the function is a constant of motion. Indeed, this holds based on the following analysis:

$$\begin{aligned} \frac{d\Phi}{dt} &= \frac{1}{2} \big(x_1^\top(t) \dot{x}_1 + \dot{x}_1^\top x_1(t) + x_2^\top(t) \dot{x}_2 + \dot{x}_2^\top x_2(t) \big) \\ &= x_1^\top(t) \dot{x}_1 + x_2^\top(t) \dot{x}_2 \\ &= x_1^\top(t) A(t) x_2(t) - x_2(t)^\top A^\top(t) x_1(t) \\ &= 0. \end{aligned}$$

Finally, observe that given a bounded initial condition, the time-invariance of $\Phi(t)$ directly implies that no strategy of any player can become unbounded so the flow ϕ^t of the differential equation has bounded orbits.

B.4 Proof of Theorem 1

Given the previous intermediate results, Theorem 1 follows from the arguments presented in Section 3.3. In particular, observe that by definition of the periodic zero-sum bilinear game, the GDA dynamics are T-periodic. Now, 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 since as presented in Section 3.3, if a T-periodic system is divergence-free then the discrete-time system defined by ϕ^T is also volume preserving [1, 3.16.B, Thm 2]. Similarly, the discrete-time system defined by the ϕ^T retains 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 the Poincaré recurrence theorem for discrete-time systems from Section 3.3 to the discrete-time system defined by ϕ^T to conclude the discrete-time system is Poincaré recurrent. This immediately implies that the GDA dynamics are Poincaré recurrent since the discrete-time system defined by ϕ^T forms a subsequence of the continuous-time system.

C GDA **Time-Average Result: Proof of Proposition 2**

Consider a periodic zero-sum bilinear game with $x_1, x_2 \in \mathbb{R}$ and a periodic payoff matrix $A(t)$ such that $A(t) = A(t+T)$ with $T = 3\pi$ for any $t \geq 0$. Moreover, let the payoff matrix evolve over a period as follows:

$$A(t) = \begin{cases} -1 & 0 \leq t \leq \pi \\ 1 & \pi \leq t \leq \frac{3\pi}{2} \\ -1 & \frac{3\pi}{2} \leq t \leq 3\pi. \end{cases}$$

The joint strategy $(x_1^*, x_2^*) = (0, 0)$ is the time-invariant Nash equilibrium. We now show that the time-average of the strategies produced by the GDA dynamics do not converge to the time-invariant Nash equilibrium.

The GDA dynamics in this periodic zero-sum bilinear game are given by

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

The solution to the differential equation that describes the GDA dynamics can be constructed in a piecewise manner. On each of the three intervals we have a linear system defined by

$$\begin{bmatrix} \dot{x}_1 \\ \dot{x}_2 \end{bmatrix} = \begin{bmatrix} 0 & A(t) \\ -A(t) & 0 \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \end{bmatrix}.$$

Now recall the following identity

$$\exp\left(\theta \begin{bmatrix} 0 & 1 \\ -1 & 0 \end{bmatrix}\right) = \begin{bmatrix} \cos(\theta) & \sin(\theta) \\ -\sin(\theta) & \cos(\theta) \end{bmatrix}.$$

Hence for initial condition $(x_1(0), x_2(0))$ and interval $[0, \pi)$ we know that $A(t) = -1$ for all t in the interval which implies that the solution on this interval is given by

$$\begin{bmatrix} x_1(t) \\ x_2(t) \end{bmatrix} = \begin{bmatrix} \cos(-t) & \sin(-t) \\ -\sin(-t) & \cos(-t) \end{bmatrix} \begin{bmatrix} x_1(0) \\ x_2(0) \end{bmatrix}.$$

On the interval $[\pi, 3\pi/2)$, $A(t) = 1$ so that

$$\begin{bmatrix} x_1(t) \\ x_2(t) \end{bmatrix} = \begin{bmatrix} \cos(t - \pi) & \sin(t - \pi) \\ -\sin(t - \pi) & \cos(t - \pi) \end{bmatrix} \begin{bmatrix} x_1(\pi) \\ x_2(\pi) \end{bmatrix}.$$

Finally, on $[3\pi/2, 3\pi)$, $A(t) = -1$ so that

$$\begin{bmatrix} x_1(t) \\ x_2(t) \end{bmatrix} = \begin{bmatrix} \cos(-(t - 3\pi/2)) & \sin(-(t - 3\pi/2)) \\ -\sin(-(t - 3\pi/2)) & \cos(-(t - 3\pi/2)) \end{bmatrix} \begin{bmatrix} x_1(3\pi/2) \\ x_2(3\pi/2) \end{bmatrix}.$$

Now, let us consider the initial condition $(x_1(0), x_2(0)) = (1, 0)$. Then,

$$(x_1(t), x_2(t)) = \begin{cases} (\cos(t), \sin(t)) = (\cos(-t), -\sin(-t)) & t \in [0, \pi) \\ (\cos(t), -\sin(t)) = (-\cos(t - \pi), \sin(t - \pi)) & t \in [\pi, 3\pi/2) \\ (-\cos(t), -\sin(t)) = (\sin(-(t - 3\pi/2)), \cos(-(t - 3\pi/2))) & t \in [3\pi/2, 3\pi) \end{cases}$$

15

Observe that from this solution we can determine that the GDA dynamics return to the initial condition at the end of a period. Thus, to assess convergence of the time-average it is sufficient to evaluate the time-average of the dynamics over a period of the evolving game. Integrating the solution over a period, we have that

$$\int_0^{3\pi} x_1(t)dt = \int_0^{3\pi/2} \cos(t)dt + \int_{3\pi/2}^{3\pi} -\cos(t)dt$$
$$= [\sin(3\pi/2) - \sin(0)] - [\sin(3\pi) - \sin(3\pi/2)]$$
$$= -2$$

and

$$\int_0^{3\pi} x_2(t)dt = \int_0^{\pi} \sin(t)dt + \int_{\pi}^{3\pi} -\sin(t)dt$$
$$= [-\cos(\pi) + \cos(0)] + [\cos(3\pi) - \cos(\pi)]$$
$$= 2$$

This implies that the time-average strategies of the players do not equal to zero, so the time-average of the GDA dynamics do not converge to the time-invariant Nash equilibrium. The fact that it is non-zero holds generally even changing the initial condition.

This completes the proof and shows that there exists periodic zero-sum bilinear games where the time-average GDA strategies do not converge to the time-invariant Nash equilibrium.

D FTRL **Poincaré Recurrence Results: Proofs for Section 5.1**

This appendix includes the proofs of Lemma 3, Lemma 4, and Theorem 2.

D.1 Proof of Lemma 3

Recall that this result states that the dynamics defined by the system $\dot{z}$ (given again in (3)) are volume preserving in any periodic zero-sum polymatrix game. Here, we also explain how the $\dot{z}$ dynamics were formulated. Then, we show that the divergence of this vector field is zero, from which we conclude the dynamics are volume preserving by Liouville's theorem. This proof follows closely arguments in [20].

For each player $i \in V$, given a fixed strategy $\beta \in \mathcal{A}_i$, for all $\alpha \in \mathcal{A}_i \setminus \beta$ the cumulative utility differences are defined by

$$z_{i\alpha}(t) = y_{i\alpha}(t) - y_{i\beta}(t).$$

This transformation from the cumulative utilities to the cumulative utility differences yields a linear map $\Pi_i : \mathbb{R}^{n_i} \to \mathbb{R}^{n_i - 1}$ from $y_i(t)$ to $z_i(t)$ for each player $i \in V$. Moreover, define by $\Pi = (\Pi_1, \ldots, \Pi_{|V|})$ the product map of the linear maps Π_i of each player $i \in V$. This map is surjective, but not injective.

Observe that the cumulative utility differences for each player $i \in V$ and all $\alpha \in \mathcal{A}_i \setminus \beta$ evolve following the differential equation

$$\dot{z}_{i\alpha}(t) = v_{i\alpha}(x(t), t) - v_{i\beta}(x(t), t). \tag{3}$$

The above differential equation is obtained directly by the form of y_i and the fundamental theorem of calculus. Moreover, recall that for any player $i \in V$ and pure strategy $\gamma \in \mathcal{A}_i$, the quantity $v_{i\gamma}(x(t), t)$ gives utility of player $i \in V$ at any time $t \geq 0$ for selecting the pure strategy $\gamma \in \mathcal{A}_i$.

To analyze the dynamics from the system in (3) we need it to be well-defined, which is not immediate since it depends on $x(t) = Q(y(t))$ and the mapping from $y(t)$ to $z(t)$ via Π is not invertible so that $y(t)$ cannot be expressed as a fuction of $z(t)$. However, despite this, the system is in fact well-defined. To see this, for each player $i \in V$, consider the reduced choice map $\hat{Q}_i : \mathbb{R}^{n_i - 1} \to \mathcal{X}_i$ defined as $\hat{Q}_i(z_i(t)) = Q_i(y_i(t))$ for some $y_i(t) \in \mathbb{R}^{n_i}$ such that $\Pi_i(y_i(t)) = z_i(t)$ which is guaranteed to exist since Π_i is surjective. Then, the fact that $\hat{Q}_i(z_i(t))$ is well-defined for each player $i \in V$ holds since by the construction $\Pi_i(y_i(t)) = \Pi_i(y_i'(t))$ if and only if $y_{i\alpha}'(t) = y_{i\alpha}(t) + c$ for $c \in \mathbb{R}$ and every $\alpha_i \in \mathcal{A}_i$ which immediately implies $Q_i(y_i'(t)) = Q_i(y_i(t))$ if and only if $\Pi_i(y_i(t)) = \Pi_i(y_i'(t))$. Finally, let $\hat{Q} = (\hat{Q}_1, \ldots, \hat{Q}_{|V|})$ be the combined reduced choice map and note that $Q(y(t)) = \hat{Q}(\Pi(y(t))) = \hat{Q}(z(t))$ by the construction. As a result, the dynamics from the system in (3) are equivalently given by the following system

$$\dot{z}_{i\alpha} = v_{i\alpha_i}(\hat{Q}_i(z_i(t)), t) - v_{i\beta_i}(\hat{Q}_i(z_i(t)), t).$$

This system is well-defined by the arguments above which ensures that the system in (3) is well-defined.

Now that we have shown the system is well-defined, we prove that is is volume preserving. To see this, observe that the vector field is divergence free. Indeed,

$$\text{div}(\dot{z}) = \sum_{i \in V} \sum_{\alpha \in \mathcal{A}_i} \frac{\partial \dot{z}_{i\alpha}}{\partial z_{i\alpha}} = \sum_{i \in V} \sum_{\alpha \in \mathcal{A}_i} \sum_{\gamma_i \in \mathcal{A}_i} \frac{\partial \dot{z}_{i\alpha}}{\partial x_{i\gamma}} \frac{\partial x_{i\gamma}}{\partial z_{i\alpha_i}} = 0.$$

Note that the equation above holds since for each player $i \in V$, the pure strategy utilities at any time $t \geq 0$ given by $v_i(x(t), t)$ where $v_{i\alpha}(x(t), t) = u_i((\alpha, x_{-i}(t)), t)$ do not depend on $x_i(t)$. Finally, the divergence free property of the vector field ensures that the flow ϕ^t of the differential equation is volume preserving by Liouville's theorem.

D.2   Proof of Lemma 4

Recall that Lemma 4 states that the orbits of the $\dot{z}$ dynamics are bounded. To prove this statement, we show that the function

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

is time-invariant where x^* denotes the time-invariant fully mixed Nash equilibrium and then argue that this is sufficient to ensure that orbits of the $\dot{z}$ dynamics are bounded.

To prove that the function $\Phi(x^*, y(t))$ is time-invariant, we show that the time-derivative of the function is equal to zero. The time-derivative of $\Phi(x^*, y(t))$ simplifies using the fact that $h_i(x_i^*)$ is time-independent to the following:

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

We begin by showing that the time-derivative of $\sum_{i \in V} h_i^*(y_i(t)) = 0$. This holds by the following computation that is explained below:

$$\frac{d}{dt} \sum_{i \in V} h_i^*(y_i(t)) = \sum_{i \in V} \langle \nabla h_i^*(y_i(t)), \dot{y}_i(t) \rangle \tag{4}$$

$$= \sum_{i \in V} \langle x_i(t), \dot{y}_i(t) \rangle \tag{5}$$

$$= \sum_{i \in V} \langle x_i(t), v_i(x(t), t) \rangle \tag{6}$$

$$= \sum_{i \in V} u_i(x(t), t) \tag{7}$$

$$= 0. \tag{8}$$

We obtain (4) by the chain rule, (5) by the maximizing argument of convex conjugates (see e.g., Shalev-Shwartz et al. 29, Chapter 2) that implies $x_i(t) = Q_i(y_i(t)) = \nabla h_i^*(y_i(t))$, (6) by the definition of $y_i(t)$ and the fundamental theorem of calculus, (7) by definition of the pure strategy utilities $v_i(x(t), t)$ and the utility $u_i(x(t), t)$, and (8) by the fact that the polymatrix game is zero-sum.

We now finish by showing that the time-derivative of $\sum_{i \in V} \langle x_i^*, y_i(t) \rangle = 0$. To begin, observe that the time-derivative can be described by

$$\frac{d}{dt} \sum_{i \in V} \langle x_i^*, y_i(t) \rangle = \sum_{i \in V} \langle x_i^*, \dot{y}_i(t) \rangle$$

$$= \sum_{i \in V} \sum_{j:(i,j) \in E} (x_i^*)^\top A^{ij}(t) x_j(t) \tag{9}$$

$$= \sum_{i \in V} \sum_{j:(i,j) \in E} (x_i^*)^\top A^{ij}(t) (x_j(t) - x_j^*). \tag{10}$$

Observe that (9) follows from the definition of $y_i(t)$ and the fundamental theorem of calculus and (10) comes about from subtracting $\sum_{i \in V} u_i(x^*, t)$ which is zero by the fact that the polymatrix game is zero-sum for any $t \geq 0$.

To continue, we remark that any zero-sum polymatrix game can be transformed to a payoff equivalent, pairwise constant-sum game [8]. This means that for each edge $(i, j) \in E$ there exists a matrix $B^{ij}(t)$ such that the following properties hold (see Lemma 3.1, 3.2, and 3.4, respectively Cai and Daskalakis 8):

Property 1. $A_{\alpha\beta}^{ij}(t) - A_{\alpha\gamma}^{ij}(t) = B_{\alpha\beta}^{ij}(t) - B_{\alpha\gamma}^{ij}(t)$ for any pure strategies $\alpha \in \mathcal{A}_i$ and $\beta, \gamma \in \mathcal{A}_j$.

Property 2. $B^{ij}(t) + (B^{ji}(t))^\top = c_{ij}(t) \cdot \mathbf{1}_{n_i \times n_j}$, where $c_{ij}(t)$ is a constant and $\mathbf{1}_{n_i \times n_j}$ is an $n_i \times n_j$ matrix of ones.

Property 3. In every joint pure strategy profile, every player $i \in V$ has the same utility in the game defined by the individual payoff matrices $\{A^{ij}(t)\}_{(i,j)\in E}$ as in the game defined by the individual payoff matrices $\{B^{ij}(t)\}_{(i,j)\in E}$.

Fixing a strategy $\gamma \in \mathcal{A}_j$, we can equivalently express any summand of (10) in the following manner that is justified below:

$$
\begin{aligned}
(x_i^*)^\top A^{ij}(x_j(t) - x_j^*) &= \sum_{\alpha \in \mathcal{A}_i} \sum_{\beta \in \mathcal{A}_j} x_{i\alpha}^* A_{\alpha\beta}^{ij}(x_{j\beta}(t) - x_{j\beta}^*) \\
&= \sum_{\alpha \in \mathcal{A}_i} \sum_{\beta \in \mathcal{A}_j} x_{i\alpha}^* \big(B_{\alpha\beta}^{ij}(t) - B_{\alpha\gamma}^{ij}(t) + A_{\alpha\gamma}^{ij}(t)\big)(x_{j\beta}(t) - x_{j\beta}^*) &\text{(11)} \\
&= (x_i^*)^\top B^{ij}(t)(x_j(t) - x_j^*) + \sum_{\alpha \in \mathcal{A}_i} x_{i\alpha}^* \big(A_{\alpha\gamma}^{ij}(t) - B_{\alpha\gamma}^{ij}(t)\big) \sum_{\beta \in \mathcal{A}_j} (x_{j\beta}(t) - x_{j\beta}^*). \\
&= (x_i^*)^\top B^{ij}(t)(x_j(t) - x_j^*). &\text{(12)}
\end{aligned}
$$

Observe that (11) results from applying Property 1 and (12) holds since both $x_j(t)$ and x_j^* are on the simplex so that $\sum_{\beta \in \mathcal{A}_j} x_{j\beta} = 1$ and $\sum_{\beta \in \mathcal{A}_j} x_{j\beta}^* = 1$ which implies $\sum_{\beta \in \mathcal{A}_j}(x_{j\beta} - x_{j\beta}^*) = 0$.

Thus, continuing from (10) and using that $(x_i^*)^\top A^{ij}(x_j(t) - x_j^*) = (x_i^*)^\top B^{ij}(t)(x_j(t) - x_j^*)$ from above and swapping the sum indexing and taking the transpose of the quadratic form $(x_i^*)^\top B^{ij}(t)(x_j(t) - x_j^*)$, we get that

$$
\begin{aligned}
\frac{d}{dt} \sum_{i \in V} \langle x_i^*, y_i(t) \rangle &= \sum_{i \in V} \sum_{j:(i,j)\in E} (x_i^*)^\top A^{ij}(t)(x_j(t) - x_j^*) \\
&= \sum_{i \in V} \sum_{j:(i,j)\in E} (x_i^*)^\top B^{ij}(t)(x_j(t) - x_j^*) &\text{(13)} \\
&= \sum_{j \in V} \sum_{i:(j,i)\in E} (x_j(t) - x_j^*)^\top (B^{ij}(t))^\top x_i^*.
\end{aligned}
$$

Moreover, we obtain the following expression that is justified below:

$$
\begin{aligned}
\frac{d}{dt} \sum_{i \in V} \langle x_i^*, y_i(t) \rangle &= \sum_{j \in V} \sum_{i:(j,i)\in E} (x_j(t) - x_j^*)^\top (B^{ij}(t))^\top x_i^* \\
&= \sum_{j \in V} \sum_{i:(j,i)\in E} (x_j(t) - x_j^*)^\top (c^{ji}(t)\mathbf{1}_{n_j \times n_i} - B^{ji}(t)) x_i^* &\text{(14)} \\
&= \sum_{j \in V} \sum_{i:(j,i)\in E} c^{ji}(t)(x_j(t) - x_j^*)^\top \mathbf{1}_{n_j \times n_i} x_i^* - \sum_{j \in V} \sum_{i:(j,i)\in E} (x_j(t) - x_j^*)^\top B^{ji}(t) x_i^* \\
&= -\sum_{j \in V} \sum_{i:(j,i)\in E} (x_j(t) - x_j^*)^\top B^{ji}(t) x_i^*. &\text{(15)}
\end{aligned}
$$

Note that (14) results from applying Property 2 and we obtain (15) using that $\sum_{j \in V} \sum_{i:(j,i)\in E} c^{ji}(t)(x_j(t) - x_j^*)^\top = 0$ since each summand is zero as can be seen by noting that $\sum_{\alpha \in \mathcal{A}_j} x_{j\alpha} = \sum_{\alpha \in \mathcal{A}_j} x_{j\alpha}^* = \sum_{\alpha \in \mathcal{A}_i} x_{i\alpha}^* = 1$ which gives

$$
c^{ji}(t)(x_j(t) - x_j^*)^\top \mathbf{1}_{n_j \times n_i} x_i^* = c^{ji}(x_j(t) - x_j^*)^\top \mathbf{1}_{n_j} = c^{ji}(t) - c^{ji}(t) = 0.
$$

We now analyze the summand in (15) for some $j \in V$. Fixing any pure strategy $\gamma_i \in \mathcal{A}_i$ for each $i \in V \setminus \{j\}$, obtain the following simplification that is explained below:

$$
\sum_{i:(j,i)\in E} (x_j(t) - x_j^*)^\top B^{ji}(t) x_i^* = \sum_{i:(j,i)\in E} \sum_{\alpha\in\mathcal{A}_j} \sum_{\beta\in\mathcal{A}_i} (x_{j\alpha}(t) - x_{j\alpha}^*) B^{ji}_{\alpha\beta}(t) x_{i\beta}^*
$$

$$
= \sum_{i:(j,i)\in E} \sum_{\alpha\in\mathcal{A}_j} \sum_{\beta\in\mathcal{A}_i} (x_{j\alpha}(t) - x_{j\alpha}^*)\big(A^{ji}_{\alpha\beta}(t) - A^{ji}_{\alpha\gamma_i}(t) + B^{ji}_{\alpha\gamma_i}(t)\big) x_{i\beta}^* \tag{16}
$$

$$
= \sum_{i:(j,i)\in E} (x_j(t) - x_j^*)^\top A^{ji}(t) x_i^* + \sum_{\alpha\in\mathcal{A}_j} (x_{j\alpha}(t) - x_{j\alpha}^*) \sum_{i:(j,i)\in E} (B^{ji}_{\alpha\gamma_i}(t) - A^{ji}_{\alpha\gamma_i}(t)) \sum_{\beta\in\mathcal{A}_i} x_{i\beta}^*
$$

$$
= \sum_{i:(j,i)\in E} (x_j(t) - x_j^*)^\top A^{ji}(t) x_i^* + \sum_{\alpha\in\mathcal{A}_j} (x_{j\alpha}(t) - x_{j\alpha}^*) \sum_{i:(j,i)\in E} (B^{ji}_{\alpha\gamma_i}(t) - A^{ji}_{\alpha\gamma_i}(t)) \tag{17}
$$

$$
= \sum_{i:(j,i)\in E} (x_j(t) - x_j^*)^\top A^{ji}(t) x_i^*. \tag{18}
$$

The equation in (16) follows from applying Property 1 and the equation in (17) holds since $\sum_{\beta\in\mathcal{A}_i} x_{i\beta}^* = 1$ as a result of the strategy spaces being on the simplex. Finally, to see how (18) is obtained, observe that for each $\alpha \in \mathcal{A}_j$ the terms $\sum_{i:(j,i)\in E} A^{ji}_{\alpha\gamma_i}(t)$ and $\sum_{i:(j,i)\in E} B^{ji}_{\alpha\gamma_i}(t)$ give the utility of player $j \in V$ in the games with payoffs $\{A^{ji}(t)\}_{(j,i)\in E}$ and $\{B^{ji}(t)\}_{(j,i)\in E}$ respectively under a joint pure strategy. Hence, by Property 3, the respective utilities are equal so that the difference is zero.

Finally, relating (18) back to (15), we conclude that the time-derivative is zero:

$$
\frac{d}{dt} \sum_{i\in V} \langle x_i^*, y_i(t) \rangle = -\sum_{j\in V} \sum_{i:(j,i)\in E} (x_j(t) - x_j^*)^\top B^{ji}(t) x_i^*
$$

$$
= -\sum_{j\in V} \sum_{i:(j,i)\in E} (x_j(t) - x_j^*)^\top A^{ji}(t) x_i^* = 0.
$$

The final equality holds since x^* is an interior Nash equilibrium, which implies $u_{j\alpha}(x^*, t) = u_j(x^*, t)$ for all strategies $\alpha \in \mathcal{A}_j$ and any linear combination thereof.

Hence,

$$
\frac{d\Phi(x^*, y(t))}{dt} = \frac{d}{dt} \sum_{i\in V} h_i^*(y_i(t)) + \frac{d}{dt} \sum_{i\in V} \langle x_i^*, y_i(t) \rangle = 0,
$$

which implies that $\Phi(x^*, y(t))$ is time-invariant.

Finally, by Lemma D.2 of Mertikopoulos et al. [20], the time-invariance of Φ is sufficient to ensure that the flow ϕ^t of the differential equation $\dot{z}$ has bounded orbits. This finishes the proof.

D.3 Proof of Theorem 2

Theorem 2 states that the FTRL dynamics are Poincaré recurrent in periodic zero-sum polymatrix games. The proof of this claim follows from Lemma 3, Lemma 4 and the methods described in Section 3.3. Indeed, to begin, observe that the $\dot{z}$ dynamics given in (3) are T-periodic. This follows immediately from the definition of a T-periodic system as described in Section 3.3 and the fact that the payoff matrices 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 3 since as presented in Section 3.3, if a T-periodic system is divergence-free then the discrete-time system defined by ϕ^T is also volume preserving [1, 3.16.B, Thm 2]. Furthermore, the bounded orbits guarantee of the continuous-time system from Lemma 4 imply the discrete-time system defined by ϕ^T has bounded orbits since it is a subsequence of the continuous-time system.

Thus, we are able to apply the Poincaré recurrence theorem to the system defined by ϕ^T to conclude that the discrete-time system is Poincaré recurrent. This implies that the $\dot{z}$ dynamics are Poincaré recurrent since the discrete-time system defined by ϕ^T forms a subsequence of the continuous-time system. Finally, the Poincaré recurrence of the $\dot{z}$ dynamics directly imply the Poincaré recurrence of the FTRL strategies. Indeed, since there is an increasing sequence of times t_n such that $z(t_n) \to z(0)$ by Poincaré recurrence, so using continuity there is also an increasing sequence of times t_n such that $x(t_n) = Q(y(t_n)) = \hat{Q}(z(t_n)) \to \hat{Q}(z(t_0)) = x(0)$ which means the FTRL dynamics are Poincaré recurrent.

E FTRL **Time-Average Convergence Results: Proofs for Section 5.2**

This appendix includes the proofs of Theorem 3 and Proposition 3.

E.1 Proof of Theorem 3

The outline of this proof is as follows. We begin by restating the relevant notation specialized to periodic zero-sum bimatrix games and then provide a more formal mathematical statement of the claim being proven. Following that we introduce a technical result regarding the bounded regret property of FTRL dynamics and state the implications that can be drawn from it. Finally, using the implications of bounded regret and properties of zero-sum bimatrix games we reach the conclusion.

Notation. Recall that we consider a periodic zero-sum bimatrix game for this result, which is a game that consists of a pair of players i and j and the bimatrix game between them at time $t \geq 0$ is described by the pair of payoffs $\{A^{ij}(t), A^{ji}(t)\} = \{A(t), -A^\top(t)\}$ and the sequence is periodic so that $\{A^{ij}(t+T), A^{ji}(t+T)\} = \{A^{ij}(t), A^{ji}(t)\}$ or equivalently $\{A(t+T), -A^\top(t+T)\} = \{A(t), -A^\top(t)\}$ for some finite period T and all time $t \geq 0$. Moreover, the bimatrix game at each time $t \geq 0$ is zero-sum which means that $u_i(x_i, x_j, t) + u_j(x_i, x_j, t) = 0$ for any strategy pair $x_i \in \mathcal{X}_i$ and $x_j \in \mathcal{X}_j$. Observe that we include the time-dependence t in the notation of player's utility to make it explicit the utility is time-dependent as a result of the time-varying payoff matrix. Finally, let the strategy pair $(x_i^*, x_j^*) \in \mathcal{X}_i \times \mathcal{X}_j$ denote the time-invariant Nash equilibrium.

Formal Statement of Result. Our goal is to prove that the time-average utility of each player converges to the time-average of the values of the games over a period. That is, we seek to show

$$\lim_{t \to \infty} \frac{1}{t} \int_0^t u_i(x(\tau), \tau)d\tau = \frac{1}{T} \int_0^T u_i(x_i^*, x_j^*, \tau)d\tau = \bar{V} \tag{19}$$

and

$$\lim_{t \to \infty} \frac{1}{t} \int_0^t u_j(x(\tau), \tau)d\tau = \frac{1}{T} \int_0^T u_j(x_j^*, x_i^*, \tau)d\tau = -\bar{V}, \tag{20}$$

where $\bar{V} := \frac{1}{T}\int_0^T V(\tau)d\tau$ denotes the time-average of the values of the games over a period and $V(\tau)$ denotes the value of the game at time τ for any $\tau \geq 0$.

Bounded Regret Property. The proof of the above statement requires an intermediate technical result. The following result of Mertikopoulos et al. [20] states that regardless of what other players do in a polymatrix game (not necessarily zero-sum), if a player follows FTRL learning dynamics then the regret of the player is bounded. It is important to remark that this result directly applies to periodic zero-sum polymatrix games. This follows from the fact that there is no assumptions on the behavior of other players, so the dynamics from the game can be viewed as arising from the behavior of the other players.

Proposition 4 (Theorem 3.1, Mertikopoulos et al. 20). *Let $h_{\max,i} = \max_{x_i \in \mathcal{X}_i} h_i(x_i)$ and $h_{\min,i} = \min_{x_i \in \mathcal{X}_i} h_i(x_i)$. If player $i \in V$ in a polymatrix game follows* FTRL *dynamics, then for every continuous trajectory of play $x_{-i}(t)$ of the opponents of player i the following regret bound holds:*

$$\max_{x_i \in \mathcal{X}_i} \frac{1}{t} \int_0^t \left[u_i(x_i, x_{-i}(\tau), \tau) - u_i(x(\tau), \tau) \right]d\tau \leq \frac{h_{\max,i} - h_{\min,i}}{t}.$$

Implications of Bounded Regret. Proposition 4 ensures that the following bounds hold for the regret of player i:

$$\frac{1}{t} \int_0^t \left[u_i(x_i^*, x_j(\tau), \tau) - u_i(x(\tau), \tau) \right]d\tau \leq \max_{x_i \in \mathcal{X}_i} \frac{1}{t} \int_0^t \left[u_i(x_i, x_j(\tau), \tau) - u_i(x(\tau), \tau) \right]d\tau$$

$$\leq \frac{h_{\max,i} - h_{\min,i}}{t}. \tag{21}$$

Similarly, Proposition 4 guarantees the following bounds hold for the regret of player j:

$$\frac{1}{t} \int_0^t \left[u_j(x_j^*, x_i(\tau), \tau) - u_j(x(\tau), \tau) \right]d\tau \leq \max_{x_j \in \mathcal{X}_j} \frac{1}{t} \int_0^t \left[u_j(x_j, x_i(\tau), \tau) - u_j(x(\tau), \tau) \right]d\tau$$

$$\leq \frac{h_{\max,j} - h_{\min,j}}{t}. \tag{22}$$

Observe that the lower bounds in (21) and (22) hold by replacing the maximizing argument over the strategy space of a player with a fixed strategy. In particular, the fixed strategy is taken to be the invariant Nash equilibrium strategy for player i or j.

Now, taking the limit as $t \to \infty$ of each side of (21) and using the zero-sum property of the bimatrix game at each time $\tau \geq 0$, we obtain the following:

$$\lim_{t \to \infty} \frac{1}{t} \int_0^t \left[u_i(x_i^*, x_j(\tau), \tau) - u_i(x(\tau), \tau) \right] d\tau = \lim_{t \to \infty} \frac{1}{t} \int_0^t \left[u_i(x_i^*, x_j(\tau), \tau) + u_j(x(\tau), \tau) \right] d\tau$$
$$\leq 0. \tag{23}$$

Similarly, taking the limit as $t \to \infty$ of each side of (22) and using the zero-sum property of the bimatrix game at each time $\tau \geq 0$, we get that:

$$\lim_{t \to \infty} \frac{1}{t} \int_0^t \left[u_j(x_j^*, x_i(\tau), \tau) - u_j(x(\tau), \tau) \right] d\tau = \lim_{t \to \infty} \frac{1}{t} \int_0^t \left[u_j(x_j^*, x_i(\tau), \tau) + u_i(x(\tau), \tau) \right] d\tau$$
$$\leq 0. \tag{24}$$

Time-Average Utility Convergence. We now proceed to show that

$$\lim_{t \to \infty} \frac{1}{t} \int_0^t u_i(x_i^*, x_j^*, \tau) d\tau \leq \lim_{t \to \infty} \frac{1}{t} \int_0^t u_i(x(\tau), \tau) d\tau \leq \lim_{t \to \infty} \frac{1}{t} \int_0^t u_i(x_i^*, x_j^*, \tau) d\tau. \tag{25}$$

The lower bound on the time-average utility of player i holds by the following analysis that is explained below:

$$\lim_{t \to \infty} \frac{1}{t} \int_0^t u_i(x(\tau), \tau) d\tau \geq \lim_{t \to \infty} \frac{1}{t} \int_0^t u_i(x_i^*, x_j(\tau), \tau) d\tau \tag{26}$$

$$\geq \lim_{t \to \infty} \frac{1}{t} \int_0^t \min_{x_j \in \mathcal{X}_j} u_i(x_i^*, x_j, \tau) d\tau \tag{27}$$

$$= \lim_{t \to \infty} \frac{1}{t} \int_0^t u_i(x_i^*, x_j^*, \tau) d\tau. \tag{28}$$

The inequality in (26) is a direct implication of (23). Moreover, the inequality in (27) is immediate by the fact that any fixed strategy of player j must give at least as much utility to player i as the strategy which minimizes the utility of player i. Finally, the last conclusion in (28) holds by the definition of a Nash equilibrium in a zero-sum bimatrix game.

The upper bound on the time-average utility of player i holds by the following similar analysis that is detailed below:

$$\lim_{t \to \infty} \frac{1}{t} \int_0^t u_i(x(\tau), \tau) d\tau \leq - \lim_{t \to \infty} \frac{1}{t} \int_0^t u_j(x_j^*, x_i(\tau), \tau) d\tau \tag{29}$$

$$= \lim_{t \to \infty} \frac{1}{t} \int_0^t u_i(x_i(\tau), x_j^*, \tau) d\tau \tag{30}$$

$$\leq \lim_{t \to \infty} \frac{1}{t} \int_0^t \max_{x_i \in \mathcal{X}_i} u_i(x_i, x_j^*, \tau) d\tau \tag{31}$$

$$= \lim_{t \to \infty} \frac{1}{t} \int_0^t u_i(x_i^*, x_j^*, \tau) d\tau. \tag{32}$$

The inequality in (29) follows directly from (24) and the equality in (30) is a result of the zero-sum property of the game at each time $\tau \geq 0$. Furthermore, the inequality in (31) holds by the fact that the the strategy of player i that maximizes the utility must give at least as much utility as any fixed strategy. The last conclusion in (32) again holds by the definition of a Nash equilibrium in a zero-sum bimatrix game.

The preceding arguments prove that the claimed inequalities in (25) hold. Observe that the time-average of the utility values of player i at the invariant Nash equilibrium converge to the time-average of the values of the games over a period as a result of the periodic nature of the game and the fact that the utility value at any Nash equilibrium in a zero-sum bimatrix game is unique. That is,

$$\lim_{t \to \infty} \frac{1}{t} \int_0^t u_i(x_i^*, x_j^*, \tau) d\tau = \frac{1}{T} \int_0^T u_i(x_i^*, x_j^*, \tau) d\tau = \bar{V}.$$

Thus, the squeeze theorem applied to (25) allows us to conclude the statement given in (19):

$$\lim_{t \to \infty} \frac{1}{t} \int_0^t u_i(x(\tau), \tau) d\tau = \frac{1}{T} \int_0^T u_i(x_i^*, x_j^*, \tau) d\tau = \bar{V}$$

Finally, by the zero-sum property of the bimatrix game at each time $\tau \geq 0$, the statement given in (20) immediately follows from the equation above. That is,

$$\lim_{t \to \infty} \frac{1}{t} \int_0^t u_j(x(\tau), \tau) d\tau = \frac{1}{T} \int_0^T u_i(x_j^*, x_i^*, \tau) d\tau = -\bar{V}.$$

This finishes the proof.

E.2 Proof of Proposition 3

We now provide the proof of Proposition 3 stating that there exists periodic zero-sum bimatrix games satisfying Definition 2 in which the time-average strategies of FTRL dynamics fail to converge to the time-invariant Nash equilibrium.

To prove this result we construct a specific periodic zero-sum bimatrix game that is the basis of the counterexample. In order to demonstrate that the FTRL strategies may not converge to the time-invariant Nash equilibrium, we consider the regularization function that leads to the replicator dynamics. Then, for replicator dynamics in the constructed game, we prove that the strategies are symmetric about the half period of the game and consequently return to the initial condition in a period of the game so that the time-average of the strategies in the limit corresponds to the time-average of the strategies over a half-period of the game. Finally, we use this property to show that the choice of the period of the game can ensure that the time-average strategies cannot converge to the time-invariant Nash equilibrium.

Counterexample Construction. A periodic zero-sum bimatrix game between players i and j is described by a periodic sequence of payoffs where the game at time $t \geq 0$ is described by the pair of payoffs $\{A^{ij}(t), A^{ji}(t)\} = \{A(t), -A^\top(t)\}$. To obtain a counterexample, we consider the periodic zero-sum bimatrix game defined by

$$A(t) = \gamma(t) A \quad \text{where} \quad A = \begin{bmatrix} 1 & -1 \\ -1 & 1 \end{bmatrix} \quad \text{and} \quad \gamma(t) = \sin\left(\frac{2\pi t}{T}\right).$$

This game corresponds to a periodic version of matching pennies and the period of the game is T. The joint strategy (x_i^*, x_j^*) where $x_i^* = (1/2, 1/2)$ and $x_j^* = (1/2, 1/2)$ is the unique time-invariant Nash equilibrium of the game.

Recall that in periodic zero-sum bimatrix games between players i and j, we denote the utility of each player at time $t \geq 0$ under the joint strategy $x(t)$ by $u_i(x(t), t)$ and $u_j(x(t), t)$ to emphasize the dependence on the time-dependent payoffs which are given by $\{\gamma(t)A, -\gamma(t)A^\top\}$ in this construction. Furthermore, in this problem construction we denote the utility of each player with payoffs $\{A, -A^\top\}$ under the joint strategy $x(t)$ by $u_i(x(t))$ and $u_j(x(t))$ where A is defined at the matching pennies payoff matrix defined above.

The regularization function $h_i(x_i) = \sum_{\alpha \in \mathcal{A}_i} x_{i\alpha} \log x_{i\alpha}$ in FTRL dynamics gives rise to the replicator dynamics commonly studied in evolutionary game theory. For the periodic zero-sum bimatrix game under consideration, the replicator dynamics for any strategy $\alpha \in \mathcal{A}_i$ of player i are given by

$$\dot{x}_{i\alpha}(t) = x_{i\alpha}(t)\big[u_{i\alpha}(x(t), t) - u_i(x(t), t)\big] = \gamma(t) x_{i\alpha}(t)\big[u_{i\alpha}(x(t)) - u_i(x(t))\big] := \gamma(t)\dot{x}'_{i\alpha}(t).$$

Similarly, the replicator dynamics for any strategy $\alpha \in \mathcal{A}_j$ of player j are given by

$$\dot{x}_{j\alpha}(t) = x_{j\alpha}(t)\big[u_{j\alpha}(x(t), t) - u_j(x(t), t)\big] = \gamma(t) x_{j\alpha}(t)\big[u_{j\alpha}(x(t)) - u_j(x(t))\big] := \gamma(t)\dot{x}'_{j\alpha}(t).$$

We now analyze the time-average of the dynamics of any strategy for each player in this game and show that they do not correspond to the time-invariant Nash equilibrium.

Time-Average Strategies. We begin by showing that for each player $k \in \{i, j\}$ and strategy of the player $\alpha \in \mathcal{A}_k$,

$$x_{k\alpha}\left(\tfrac{T}{2} + t\right) = x_{k\alpha}\left(\tfrac{T}{2} - t\right). \tag{33}$$

To see this, observe that for each player $k \in \{i, j\}$ and strategy of the player $\alpha \in \mathcal{A}_k$ and some initial condition t_0,

$$x_{k\alpha}(t) = x_{k\alpha}(t_0) + \int_{t_0}^t \dot{x}_{j\alpha}(\tau) d\tau = x_{k\alpha}(t) + \int_{t_0}^t \sin\left(\tfrac{2\pi}{T}\tau\right) \dot{x}'_{k\alpha}(\tau) d\tau. \tag{34}$$

To prove the claim in (33), we show that both $x_{k\alpha}\left(\frac{T}{2} + t\right)$ and $x_{k\alpha}\left(\frac{T}{2} - t\right)$ satisfy the same ordinary differential equation and initial condition. That is, we invoke the fundamental theorem of ordinary differential equations which says the solutions exist and are unique so that the claim holds.

Indeed, from (34) and the fundamental theorem of calculus

$$\frac{d}{dt} x_{k\alpha}\left(\tfrac{T}{2} + t\right) = \dot{x}'_{k\alpha}\left(\tfrac{T}{2} + t\right) \sin\left(\tfrac{2\pi}{T}\left(\tfrac{T}{2} + t\right)\right)$$
$$= \dot{x}'_{k\alpha}\left(\tfrac{T}{2} + t\right) \sin\left(\pi + \tfrac{2\pi}{T}t\right)$$
$$= -\dot{x}'_{k\alpha}\left(\tfrac{T}{2} + t\right) \sin\left(\tfrac{2\pi}{T}t\right).$$

Similarly,

$$\frac{d}{dt} x_{k\alpha}\left(\tfrac{T}{2} - t\right) = -\dot{x}'_{k\alpha}\left(\tfrac{T}{2} - t\right) \sin\left(\tfrac{2\pi}{T}\left(\tfrac{T}{2} - t\right)\right)$$
$$= -\dot{x}'_{k\alpha}\left(\tfrac{T}{2} - t\right) \sin\left(\pi - \tfrac{2\pi}{T}t\right)$$
$$= -\dot{x}'_{k\alpha}\left(\tfrac{T}{2} - t\right) \sin\left(\tfrac{2\pi}{T}t\right).$$

We conclude that the functions $x_{k\alpha}\left(\frac{T}{2} + t\right)$ and $x_{k\alpha}\left(\frac{T}{2} - t\right)$ satisfy the same ordinary differential equation. That is, the functional form of the ordinary differential equation is the same in both expressions. Furthermore, $x_{k\alpha}\left(\frac{T}{2} + 0\right) = x_{k\alpha}\left(\frac{T}{2} - 0\right) = x_{k\alpha}\left(\frac{T}{2}\right)$ so they satisfy the same initial condition. Hence, invoking the uniqueness property of the fundamental theorem of ordinary differential equations, the claim given in (33) holds.

The property in (33) implies for each player $k \in \{i, j\}$ and strategy of the player $\alpha \in \mathcal{A}_k$ that

$$\lim_{t \to \infty} \frac{1}{t} \int_0^t x_{k\alpha}(\tau)d\tau = \frac{1}{T} \int_0^T x_{k\alpha}(\tau)d\tau = \frac{2}{T} \int_0^{T/2} x_{k\alpha}(\tau)d\tau.$$

That is, the limiting time-average strategy is equal to the time-average strategy over half a period of the periodic game.

Now recall that the time-invariant Nash equilibrium strategy is given by the joint strategy (x_i^*, x_j^*) where $x_i^* = (1/2, 1/2)$ and $x_j^* = (1/2, 1/2)$. Thus, to finish the proof we need to show for some player $k \in \{i, j\}$ and strategy of $\alpha \in \mathcal{A}_k$ that

$$\frac{2}{T} \int_0^{T/2} x_{k\alpha}(\tau)d\tau \neq \frac{1}{2}.$$

To see that the claim above holds, recall from (34) that

$$x_{k\alpha}(t) = x_{k\alpha}(0) + \int_0^t \sin\left(\tfrac{2\pi}{T}\tau\right)\dot{x}'_{k\alpha}(\tau)d\tau.$$

Observe that for any $\tau \geq 0$,

$$-1 \leq \sin\left(\frac{2\pi}{T}\tau\right) \leq 1 \quad \text{and} \quad -2 \leq \dot{x}'_{k\alpha}(\tau) \leq 2.$$

The previous expressions combine to imply

$$x_{k\alpha}(0) - 2t \leq x_{k\alpha}(t) \leq x_{k\alpha}(0) + 2t$$

and consequently

$$x_{k\alpha}(0) - \frac{T}{2} \leq \frac{2}{T} \int_0^{T/2} x_{k\alpha}(\tau)d\tau \leq x_{k\alpha}(0) + \frac{T}{2}.$$

Finally, suppose that $x_{k\alpha}(0) > 1/2$. Then, when $T < 2(x_{k\alpha}(0) - 1/2)$, the time-average of the strategy

$$\frac{2}{T} \int_0^{T/2} x_{k\alpha}(\tau)d\tau > \frac{1}{2}.$$

Analogously, suppose that $x_{k\alpha}(0) < 1/2$. Then, when $T < 2(1/2 - x_{k\alpha}(0))$, the time-average of the strategy

$$\frac{2}{T} \int_0^{T/2} x_{k\alpha}(\tau)d\tau < \frac{1}{2}.$$

Thus, unless the players initialize at the time-invariant Nash equilibrium strategy, there is a choice of the period T of the periodic zero-sum matrix such that the time-average of the strategies do not converge to the time-invariant Nash equilibrium. This completes the proof.

F   Experiments

In this section, we present additional simulations and details that serve to strengthen the results in the main paper.

F.1   GDA Results

First, for continuous-time GDA dynamics we show that Poincaré recurrence holds in a periodic zero-sum bilinear game. We consider the ubiquitous Matching Pennies game with payoff matrix $A = \left(\begin{smallmatrix} 1 & -1 \\ -1 & 1 \end{smallmatrix} \right)$. We then use the following periodic rescaling with period 2π:

$$\alpha(t) = \begin{cases} \sin(t) & 0 \leq t \leq \frac{3\pi}{2} \\ \left(\frac{2}{\pi} \right)(t \bmod(2\pi) - 2\pi) & \frac{3\pi}{2} \leq t \leq 2\pi \end{cases} \tag{35}$$

Hence, the bilinear zero-sum game at time $t \geq 0$ is then described by the payoffs $\{\alpha(t)A(t), -\alpha(t)A(t)^\top\}$. 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.

Another key result in the space of GDA learning dynamics is that the time-average behavior fails to converge in general to the time invariant equilibrium $(0, 0)$. A simple counterexample can be constructed by considering a periodic zero-sum bilinear game with $x_1, x_2 \in \mathbb{R}$ and a periodic rescaling $\beta(t)$ such that $\beta(t) = \beta(t + T)$ with $T = 3\pi$ for any $t \geq 0$. Moreover, let the rescaling evolve over a period as follows:

$$\beta(t) = \begin{cases} -1 & 0 \leq t \leq \pi \\ 1 & \pi \leq t \leq \frac{3\pi}{2} \\ -1 & \frac{3\pi}{2} \leq t \leq 3\pi. \end{cases} \tag{36}$$

For the simulation, we consider the payoff matrices described by $\{\beta(t)A, -\beta(t)A^\top\}$ where $A = \left(\begin{smallmatrix} 1 & -1 \\ -1 & 1 \end{smallmatrix} \right)$. We show for this example that when both players use GDA, the time average strategy of each player remains bounded away from the Nash $[1/2, 1/2]$, as shown in Figure 5.
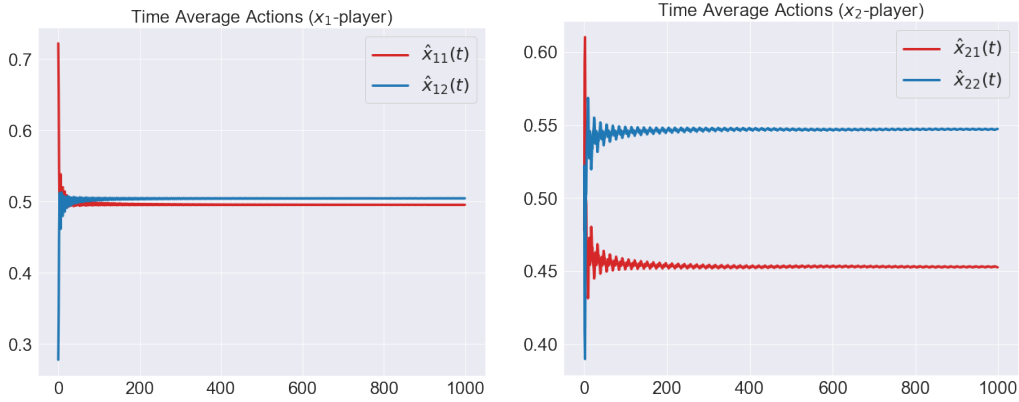

Figure 5: Time-average convergence away from the mixed Nash

F.2   FTRL Results

For the case of FTRL dynamics, we perform simulations on Matching Pennies updated with replicator dynamics. The reader is reminded of the definition of replicator dynamics as a continuous analogue of multiplicative weights update, as described in Section 5. In polymatrix games, replicator dynamics for each $i \in V$ uses regularization function $h_i(x_i) = \sum_{\alpha \in \mathcal{A}_i} x_{i\alpha} \log x_{i\alpha}$ in the FTRL dynamics. Like the GDA case, we also use the periodic rescaling described in Equation 35 and obtain recurrent dynamics, as seen in Figure 6.

Theorem 3 states that the time-average utility of each player converges to the time-average value of periodic zero-sum games when each player follows FTRL dynamics. However, Proposition 3 states that there exist periodic zero-sum bimatrix games where the time-average strategies of FTRL dynamics fail to converge to the time-invariant Nash equilibrium. Here, we show a simple example that exhibits both results. Consider a Matching Pennies game that is
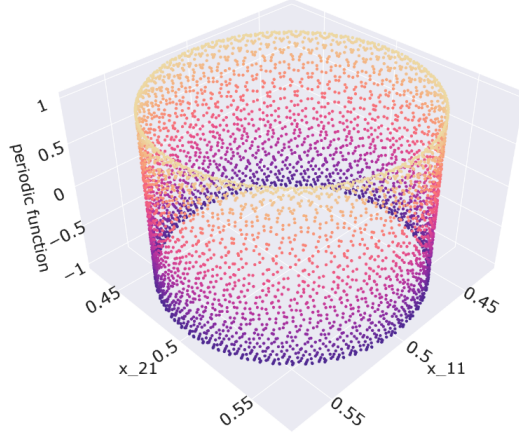
Figure 6: Periodically Rescaled Matching Pennies (Replicator)

rescaled with a $\sin$ function. Specifically, the periodic bimatrix game is given by

$$A(t) = \sin(t) \begin{bmatrix} 1 & -1 \\ -1 & 1 \end{bmatrix} \tag{37}$$

Even with this simple example, the time-average utilities for both players go to zero, while the time-average strategies do not converge to the $[1/2, 1/2]$ time-invariant Nash, as seen in Figure 7.
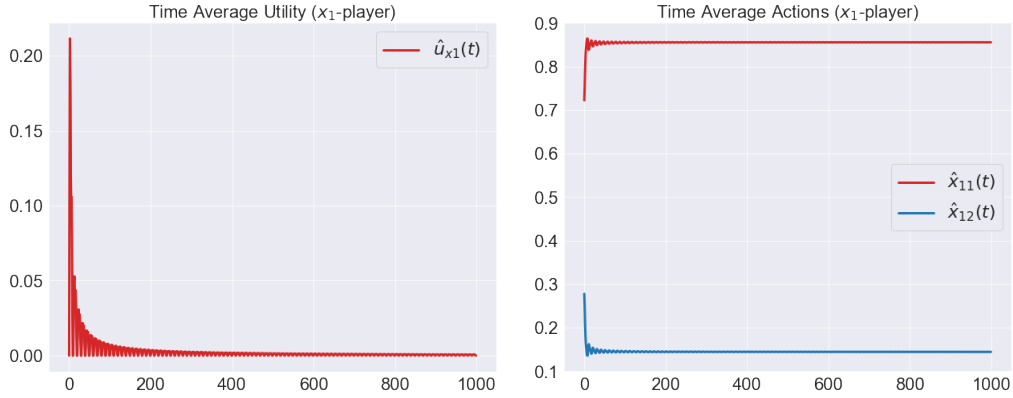

Figure 7: Time average results for MP rescaled with sin function. Notice that although the average utility goes to 0, the time average actions/strategy does not go to the time invariant Nash [1/2, 1/2].

F.3   Large Scale Simulation

We now perform simulations on larger-scale systems that present our findings in a more visually striking manner. Firstly, the time-invariant function presented in Lemma 3 and its proof can be demonstrated in any two-player periodic zero-sum polymatrix game. For replicator dynamics, the invariant function is the KL-divergence between each player's strategy and the unique mixed Nash. Using the same simulated data that was used to generate Figure 6, we show that the sum of divergences is indeed constant when both agents are using replicator dynamics. Figure 8 shows this phenomenon in the case where there are just two players playing a periodically rescaled Matching Pennies game. Specifically, the blue area represents the KL-divergence of the first player from the mixed Nash over time, and the green area represents the divergence of the second player.

To extend this formulation to the multiplayer setting, we implement a graphical polymatrix game where a number of agents are arranged in a line. Each agent then plays a bimatrix game against the agent directly adjacent to them, and the final agent also plays against the first agent. This results in a 'toroid'-like chain of games, where each agent plays
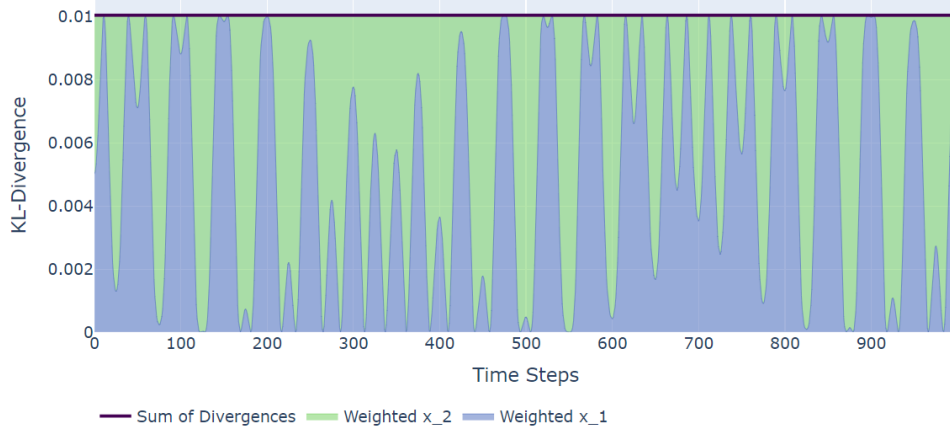
Figure 8: Time invariant function for two player periodically rescaled Matching Pennies game

against two other agents. In our simulation, each pair of agents plays the Matching Pennies game rescaled with $\sin$ against each other (Equation 37). With this system, we simulated a chain of games with 64 nodes (agents), resulting in much more complex dynamics than the two player case. Indeed, in Figure 10 we show zoomed-in plots of Figure 2. Here, similar to previous plots, each player's KL-divergence is represented by a different-colored area. The figure showcases that on a more granular level, the individual KL-divergences of each agent can become extremely erratic, and look nowhere near periodic. Nevertheless, we see from Figure 2 that the sum of KL-divergences remains constant.
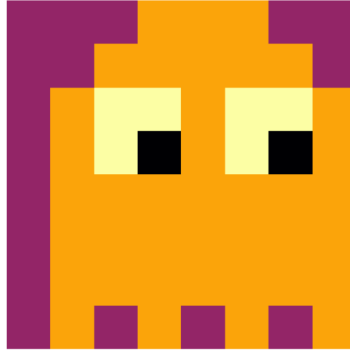

Figure 9: 8×8 grid of colors generated by sigmoid function

We also represent the trajectories of each agent by equating the strategy values of each player to RGB values in an 8x8 grid. In particular, the color of each pixel on the grid represents the probability of the respective player playing the first strategy, tuned with a sigmoid function. We then select initial conditions the correspond to RGB values such that they form the image shown in Figure 9. With the sigmoid function, any changes from the initial condition are reflected by changes in the color of the individual pixels. Thus, as agents play pairwise bimatrix games using replicator dynamics, the colors of the grid evolve. If recurrence holds, we expect to see the same image after some time. For the case of the Matching Pennies games rescaled with $\sin$, we obtain the various images found in Figure 3, which exhibit recurrence. We also provide an animation showing how the figure evolves over time in the supplementary Jupyter notebook.

F.4   Reproducibility Details

All experiments performed for this work were done using Python 3.7 and have been compiled into a Jupyter notebook for ease of viewing. Running the code requires only basic scientific computing packages such as NumPy and SciPy, as well as data visualization packages such as Matplotlib and Plotly. Most of the code in our submission has been edited such that it can be easily executed on a standard computer in a matter of minutes.
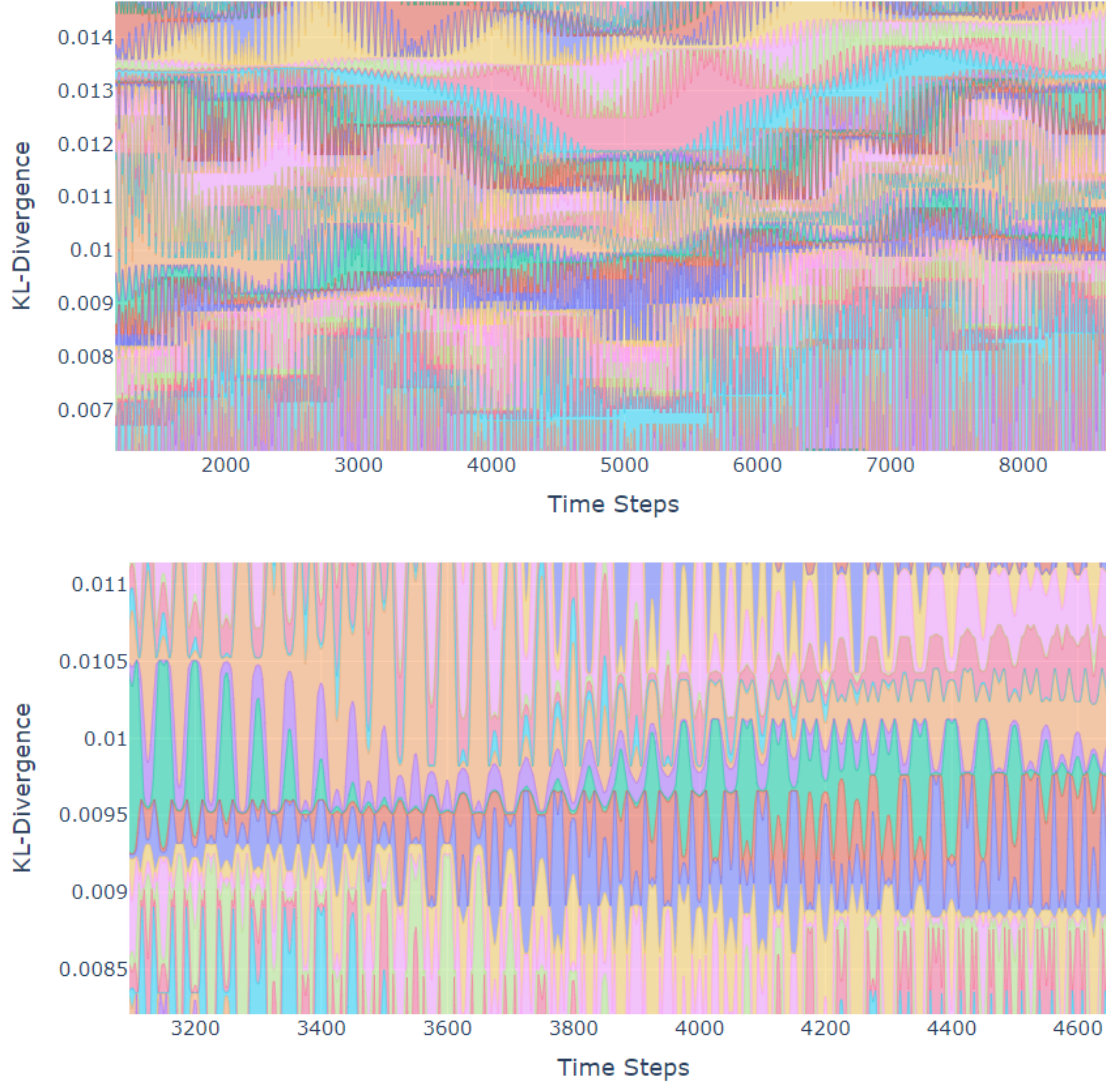
Figure 10: Zoomed-in time invariant functions for 64-player game.

Appendix References

[1] Alexander, J.A. & Mozer, M.C. (1995) Template-based algorithms for connectionist rule extraction. In G. Tesauro, D.S. Touretzky and T.K. Leen (eds.), *Advances in Neural Information Processing Systems 7*, pp. 609–616. Cambridge, MA: MIT Press.

[2] Bower, J.M. & Beeman, D. (1995) *The Book of GENESIS: Exploring Realistic Neural Models with the GEneral NEural SImulation System.* New York: TELOS/Springer–Verlag.

[3] Hasselmo, M.E., Schnell, E. & Barkai, E. (1995) Dynamics of learning and recall at excitatory recurrent synapses and cholinergic modulation in rat hippocampal region CA3. *Journal of Neuroscience* **15**(7):5249-5262.