CYCLES IN ADVERSARIAL REGULARIZED LEARNING

PANAYOTIS MERTIKOPOULOS*,
CHRISTOS PAPADIMITRIOU§, AND GEORGIOS PILIOURAS‡

Abstract. Regularized learning is a fundamental technique in online optimization, machine learning and many other fields of computer science. A natural question that arises in these settings is how regularized learning algorithms behave when faced against each other. We study a natural formulation of this problem by coupling regularized learning dynamics in zero-sum games. We show that the system's behavior is Poincaré recurrent, implying that almost every trajectory revisits any (arbitrarily small) neighborhood of its starting point infinitely often. This cycling behavior is robust to the agents' choice of regularization mechanism (each agent could be using a different regularizer), to positive-affine transformations of the agents' utilities, and it also persists in the case of networked competition, i.e., for zero-sum polymatrix games.

1. Introduction

Regularization is a fundamental and incisive method in optimization, its present *zeitgeist* and its entry into machine learning. Through the introduction of a new component in the objective, regularization techniques overcome ill-conditioning and overfitting, and they yield algorithms that achieve sparsity and parsimony without sacrificing efficiency [2, 5, 8].

In the context of online optimization, these features are exemplified in the family of learning algorithms known as *"Follow the Regularized Leader"* (FoReL) [41]. FoReL represents an important archetype of adaptive behavior for several reasons: it provides optimal min-max regret guarantees ($\mathcal{O}(t^{-1/2})$ in an adversarial setting), it offers significant flexibility with respect to the geometry of the problem at hand, and it captures numerous other dynamics as special cases (hedge, multiplicative weights, gradient descent, etc.) [2, 8, 15]. As such, given that these regret guarantees hold without any further assumptions about how payoffs/costs are determined at each stage, the dynamics of FoReL have been the object of intense scrutiny and study in algorithmic game theory.

The standard way of analyzing such no-regret dynamics in games involves a two-step approach. The first step exploits the fact that the empirical frequency of

* Univ. Grenoble Alpes, CNRS, Inria, LIG, F-38000 Grenoble, France.

§ UC Berkeley.

‡ Singapore University of Technology and Design.

E-mail addresses: panayotis.mertikopoulos@imag.fr, christos@berkeley.edu, georgios@sutd.edu.sg.

Key words and phrases. Regret; multiplicative weights update; zero-sum games; dueling algorithms.

Panayotis Mertikopoulos was partially supported by the French National Research Agency (ANR) project ORACLESS (ANR–GAGA–13–JS01–0004–01) and the Huawei Innovation Research Program ULTRON. Georgios Piliouras would like to acknowledge SUTD grant SRG ESD 2015 097 and MOE AcRF Tier 2 Grant 2016-T2-1-170.

play under a no-regret algorithm converges to the game's set of coarse correlated equilibria (CCE). The second involves proving some useful property of the game's CCE: For instance, leveraging (λ, μ)-robustness [33] implies that the social welfare at a CCE lies within a small constant of the optimum social welfare; as another example, the product of the marginal distributions of CCE in zero-sum games is Nash. In this way, the no-regret properties of FoReL can be turned into convergence guarantees for the players' empirical frequency of play (that is, in a time-averaged, correlated sense).

Recently, several papers have moved beyond this "black-box" framework and focused instead on obtaining stronger regret/convergence guarantees for systems of learning algorithms coupled together in games with a specific structure. Along these lines, Daskalakis et al. [9] and Rakhlin and Sridharan [31] developed classes of dynamics that enjoy a $\mathcal{O}(\log t/t)$ regret minimization rate in two-player zero-sum games. Syrgkanis et al. [43] further analyzed a recency biased variant of FoReL in more general multi-player games and showed that it is possible to achieve an $\mathcal{O}(t^{-3/4})$ regret minimization rate. The social welfare converges at a rate of $\mathcal{O}(t^{-1})$, a result which was extended to standard versions of FoReL dynamics in [11].

Whilst a regret-based analysis provides significant insights about these systems, it does not answer a fundamental behavioral question:

Does the system converge to a Nash equilibrium?
Does it even stabilize?

The dichotomy between a self-stabilizing, convergent system and a system with recurrent cycles is of obvious significance, but a regret-based analysis cannot distinguish between the two. Indeed, convergent, recurrent, and even chaotic [26] systems may exhibit equally strong regret minimization properties in general games, so the question remains: What does the long-run behavior of FoReL look like, really?

This question becomes particularly interesting and important under perfect competition (such as zero-sum games and variants thereof). Especially in practice, zero-sum games can capture optimization "duels" [18]: for example, two Internet search engines competing to maximize their market share can be modeled as players in a zero-sum game with a convex strategy space. In [18] it was shown that the time-average of a regret-minimizing class of dynamics converges to an approximate equilibrium of the game. Finally, zero-sum games have also been used quite recently as a model for deep learning optimization techniques in image generation and discrimination [14, 39].

In each of the above cases, min-max strategies are typically thought of as the axiomatically correct prediction. The fact that the time average of the marginals of a FoReL procedure converges to such states is considered as further evidence of the correctness of this prediction. However, the long-run behavior of the *actual* sequence of play (as opposed to its time-averages) seems to be trickier, and a number of natural questions arise:

- *Does optimization-driven learning converge under perfect competition?*
- *Does fast regret minimization necessarily imply (fast) equilibration in this case?*

Our results. We settle these questions with a resounding "no". Specifically, we show that the behavior of FoReL in zero-sum games with an interior equilibrium

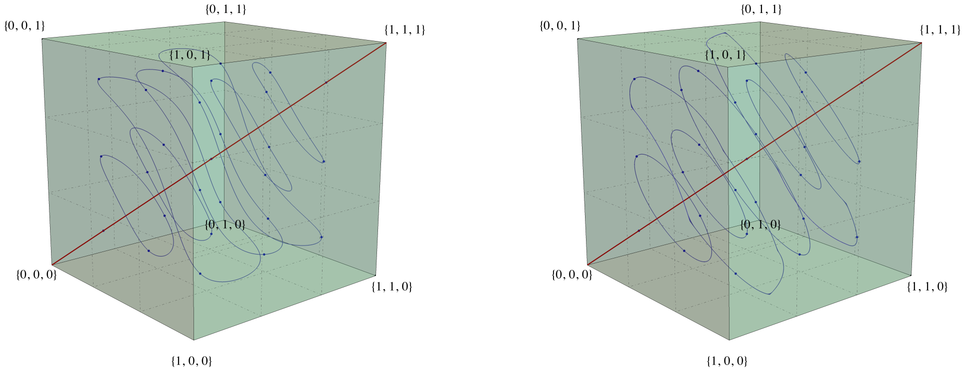

Figure 1. Evolution of the dynamics of FoReL in a 3-player zero-sum polymatrix game with entropic and Euclidean regularization (left and right respectively). The game considered is a graphical variant of Matching Pennies with three players. As can be seen, the trajectories of FoReL orbit the game's line of Nash equilibria (dark red). The kinks observed in the Euclidean case occur when the support of the trajectory of play changes; by contrast, the multiplicative weights dynamics (left) are interior, so they do not exhibit such kinks.

(e.g. Matching Pennies) is *Poincaré recurrent*, implying that almost every trajectory revisits any (arbitrarily small) neighborhood of its starting point infinitely often. Importantly, the observed cycling behavior is robust to the agents' choice of regularization mechanism (each agent could be using a different regularizer), and it applies to any positive affine transformation of zero-sum games (and hence all strictly competitive games [1]) even though these transformations lead to *different* trajectories of play. Finally, this cycling behavior also persists in the case of networked competition, i.e. for constant-sum polymatrix games [6, 7, 10].

Given that the no-regret guarantees of FoReL require a decreasing step-size (or learning rate),[1] we focus on a smooth version of FoReL described by a dynamical system in continuous time. The resulting FoReL dynamics enjoy a particularly strong $\mathcal{O}(t^{-1})$ regret minimization rate and they capture as a special case the replicator dynamics [38, 44, 45] and the projection dynamics [12, 24, 36], arguably the most widely studied game dynamics in biology, evolutionary game theory and transportation science [16, 35, 48]. In this way, our analysis unifies and generalizes many prior results on the cycling behavior of evolutionary dynamics [16, 28, 29, 37] and it provides a new interpretation of these results through the lens of optimization and machine learning.

From a technical point of view, our analysis touches on several issues. Our first insight is to focus not on the simplex of the players' mixed strategies, but on a *dual* space of payoff differences. The reason for this is that the vector of cumulative payoff differences between two strategies fully determines a player's mixed strategy under FoReL, and it is precisely these differences that ultimately drive the players' learning process. Under this transformation, FoReL exhibits a striking property,

[1] A standard trick is to decrease step-sizes by a constant factor after a window of "doubling" length [40].

incompressibility: the flow of the dynamics is volume-preserving, so a ball of initial conditions in this dual space can never collapse to a point.

That being said, the evolution of such a ball in the space of payoffs could be *transient*, implying in particular that the players' mixed strategies could converge (because the choice map that links payoff differences to strategies is nonlinear). To rule out such behaviors, we show that FoReL in zero-sum games with an interior Nash equilibrium has a further important property: it admits a *constant of motion*. Specifically, if $x^* = (x_i^*)_{i\in\mathcal{N}}$ is an interior equilibrium of the game and y_i is an arbitrary point in the payoff space of player i, this constant is given by the coupling function

$$G(y) = \sum_{i\in\mathcal{N}} [h_i^*(y_i) - \langle y_i, x_i^*\rangle],$$

where $h_i^*(y_i) = \max_{x_i}\{\langle y_i, x_i\rangle - h_i(x_i)\}$ is the convex conjugate of the regularizer h_i that generates the learning process of player i (for the details, see Sections 3 and 4). Coupled with the dynamics' incompressibility, this invariance can be used to show that FoReL is *recurrent*: after some finite time, almost every trajectory returns arbitrarily close to its initial state.

On the other hand, if the game does not admit an interior equilibrium, the coupling above is no longer a constant of motion. In this case, G decreases over time until the support of the players' mixed strategies matches that of a Nash equilibrium with maximal support: as this point in time is approached, G essentially becomes a constant. Thus, in general zero-sum games, FoReL wanders perpetually in the smallest face of the game's strategy space containing all of the game's equilibria; indeed, the only possibility that FoReL converges is if the game admits a unique Nash equilibrium in pure strategies – a fairly restrictive requirement.

2. Definitions from game theory

2.1. **Games in normal form.** We begin with some basic definitions from game theory. A *finite game in normal form* consists of a finite set of *players* $\mathcal{N} = \{1,\dots,N\}$, each with a finite set of *actions* (or *strategies*) $\mathcal{A}_i$. The preferences of player i for one action over another are determined by an associated *payoff function* $u_i\colon \mathcal{A} \equiv \prod_i \mathcal{A}_i \to \mathbb{R}$ which assigns a reward $u_i(\alpha_i;\alpha_{-i})$ to player $i \in \mathcal{N}$ under the *strategy profile* $(\alpha_i;\alpha_{-i})$ of all players' actions.[2] Putting all this together, a game in normal form will be written as a tuple $\Gamma \equiv \Gamma(\mathcal{N},\mathcal{A},u)$ with players, actions and payoffs defined as above.

Players can also use *mixed strategies*, i.e. mixed probability distributions $x_i = (x_{i\alpha_i})_{\alpha_i\in\mathcal{A}_i} \in \Delta(\mathcal{A}_i)$ over their action sets $\mathcal{A}_i$. The resulting probability vector x_i is called a *mixed strategy* and we write $\mathcal{X}_i = \Delta(\mathcal{A}_i)$ for the mixed strategy space of player i. Aggregating over players, we also write $\mathcal{X} = \prod_i \mathcal{X}_i$ for the game's *strategy space*, i.e. the space of all strategy profiles $x = (x_i)_{i\in\mathcal{N}}$.

In this context (and in a slight abuse of notation), the expected payoff of the i-th player in the profile $x = (x_1,\dots,x_N)$ is

$$u_i(x) = \sum_{\alpha_1\in\mathcal{A}_1}\cdots\sum_{\alpha_N\in\mathcal{A}_N} u_i(\alpha_1,\dots,\alpha_N)\, x_{1\alpha_1}\cdots x_{N\alpha_N}. \tag{2.1}$$

To keep track of the payoff of each pure strategy, we also write $v_{i\alpha_i}(x) = u_i(\alpha_i;x_{-i})$ for the payoff of strategy $\alpha_i \in \mathcal{A}_i$ under the profile $x \in \mathcal{X}$ and $v_i(x) = (v_{i\alpha_i}(x))_{\alpha_i\in\mathcal{A}_i}$

[2] In the above, we use the standard shorthand $(\beta_i;\alpha_{-i})$ for the profile $(\alpha_1,\dots,\beta_i,\dots,\alpha_N)$.

for the resulting *payoff vector* of player i. We then have

$$u_i(x) = \langle v_i(x), x_i \rangle = \sum_{\alpha_i \in \mathcal{A}_i} x_{i\alpha_i} v_{i\alpha_i}(x), \tag{2.2}$$

where $\langle v, x \rangle \equiv v^\top x$ denotes the ordinary pairing between v and x.

The most widely used solution concept in game theory is that of a *Nash equilibrium* (NE), defined here as a mixed strategy profile $x^* \in \mathcal{X}$ such that

$$u_i(x_i^*; x_{-i}^*) \geq u_i(x_i; x_{-i}^*) \tag{NE}$$

for every deviation $x_i \in \mathcal{X}_i$ of player i and all $i \in \mathcal{N}$. Writing $\operatorname{supp}(x_i^*) = \{\alpha_i \in \mathcal{A}_i : x_i^* > 0\}$ for the support of $x_i^* \in \mathcal{X}_i$, a Nash equilibrium $x^* \in \mathcal{X}$ is called *pure* if $\operatorname{supp}(x_i^*) = \{\alpha_i^*\}$ for some $\alpha_i^* \in \mathcal{A}_i$ and all $i \in \mathcal{N}$. At the other end of the spectrum, x^* is said to be *interior* (or *fully mixed*) if $\operatorname{supp}(x_i^*) = \mathcal{A}_i$ for all $i \in \mathcal{N}$. Finally, a *coarse correlated equilibrium* (CCE) is a distribution π over the set of action profiles $\mathcal{A} \equiv \prod_i \mathcal{A}_i$ such that, for every player $i \in \mathcal{N}$ and every action $\beta_i \in \mathcal{A}_i$, we have $\sum_{\alpha \in \mathcal{A}} v_i(\alpha)\pi(\alpha) \geq \sum_{\alpha_{-i} \in \mathcal{A}_{-i}} v_i(\beta_i, \alpha_{-i})\pi_i(\alpha_{-i})$, where $\pi_i(\alpha_{-i}) = \sum_{\alpha_i \in \alpha_i} \pi(\alpha_i, \alpha_{-i})$ is the marginal distribution of π with respect to i.

2.2. **Zero-sum games and zero-sum polymatrix games.** Perhaps the most widely studied class of finite games (and certainly the first to be considered) is that of 2-*player zero-sum games*, i.e. when $\mathcal{N} = \{1, 2\}$ and $u_1 = -u_2$. Letting $u \equiv u_1 = -u_2$, the *value* of a 2-player zero-sum game Γ is defined as

$$u_\Gamma = \max_{x_1 \in \mathcal{X}_1} \min_{x_2 \in \mathcal{X}_2} u(x_1, x_2) = \min_{x_2 \in \mathcal{X}_2} \max_{x_1 \in \mathcal{X}_1} u(x_1, x_2), \tag{2.3}$$

with equality following from von Neumann's celebrated min-max theorem [47]. As is well known, the solutions of this saddle-point problem form a closed, convex set consisting precisely of the game's Nash equilibria; moreover, the players' equilibrium payoffs are simply u_Γ and $-u_\Gamma$ respectively. As a result, Nash equilibrium is the standard game-theoretic prediction in such games.

An important question that arises here is whether the straightforward equilibrium structure of zero-sum games extends to the case of a *network* of competitors. Following [6, 7, 10], an N*-player pairwise zero-/constant-sum polymatrix game* consists of an (undirected) *interaction graph* $\mathcal{G} \equiv \mathcal{G}(\mathcal{N}, \mathcal{E})$ whose set of nodes $\mathcal{N}$ represents the competing players, with two nodes $i, j \in \mathcal{N}$ connected by an edge $e = (i, j)$ in $\mathcal{E}$ if and only if the corresponding players compete with each other in a two-player zero-/constant-sum game.

To formalize this, we assume that *a*) every player has a finite set of actions $\mathcal{A}_i$ (as before); and *b*) to each edge $e = \{i, j\} \in \mathcal{E}$ is associated a two-player game zero-/constant-sum Γ_e with player set $\mathcal{N}_e = \{i, j\}$, action sets $\mathcal{A}_i$ and $\mathcal{A}_j$, and payoff functions $u_{ij} = \gamma_{\{i,j\}} - u_{ji} \colon \mathcal{A}_i \times \mathcal{A}_j \to \mathbb{R}$ respectively.[3] The space of mixed strategies of player i is again $\mathcal{X}_i = \Delta(\mathcal{A}_i)$, but the player's payoff is now determined by aggregating over all games involving player i, i.e.

$$u_i(x) = \sum_{j \in \mathcal{N}_i} u_{ij}(x_i, x_j), \tag{2.4}$$

[3] In a zero-sum game, we have $\gamma_{\{i,j\}} = 0$ by default. Since the underlying interaction graph is assumed undirected, we also assume that the labeling of the players' payoff functions is symmetric. At the expense of concision, our analysis extends to directed graphs, but we stick with the undirected case for clarity.

where $\mathcal{N}_i = \{j \in \mathcal{N} : \{i,j\} \in \mathcal{E}\}$ denotes the set of "neighbors" of player i. In other words, the payoff to player i is simply the the sum of all payoffs in the zero-/constant-sum games that player i plays with their neighbors.

In what follows, we will also consider games which are payoff-equivalent to positive-affine transformations of pairwise constant-sum polymatrix games. Formally, we will allow for games Γ such that there exists a pairwise constant-sum polymatrix game Γ' and constants $a_i > 0$ and $b_i \in \mathbb{R}$ for each player i such that $u_i^{\Gamma}(\alpha) = a_i u_i^{\Gamma'}(\alpha) + b_i$ for each outcome $\alpha \in \mathcal{A}$.

3. No-regret learning via regularization

Throughout this paper, our focus will be on repeated decision making in low-information environments where the players don't know the rules of the game (perhaps not even that they are playing a game). In this case, even if the game admits a unique Nash equilibrium, it is not reasonable to assume that players are able to pre-compute their component of an equilibrium strategy – let alone assume that all players are fully rational, that there is common knowledge of rationality, etc.

With this in mind, we only make the bare-bones assumption that every player seeks to at least minimize their "regret", i.e. the average payoff difference between a player's mixed strategy at time $t \geq 0$ and the player's best possible strategy in hindsight. Formally, assuming that play evolves in continuous time, the *regret* of player i along the sequence of play $x(t)$ is defined as

$$\operatorname{Reg}_i(t) = \max_{p_i \in \mathcal{X}_i} \frac{1}{t} \int_0^t [u_i(p_i; x_{-i}(s)) - u_i(x(s))]\, ds, \tag{3.1}$$

and we say that player i has *no regret* under $x(t)$ if $\limsup_{t\to\infty} \operatorname{Reg}_i(t) \leq 0$.

The most widely used scheme to achieve this worst-case guarantee is known as *"Follow the Regularized Leader"* (FoReL), an exploitation-exploration class of policies that consists of playing a mixed strategy that maximizes the player's expected cumulative payoff (the exploitation part) minus a regularization term (exploration). In our continuous-time framework, this is described by the learning dynamics

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

where the so-called *choice map* $Q_i \colon \mathbb{R}^{\mathcal{A}_i} \to \mathcal{X}_i$ is defined as

$$Q_i(y_i) = \operatorname*{arg\,max}_{x_i \in \mathcal{X}_i} \{\langle y_i, x_i \rangle - h_i(x_i)\}. \tag{3.2}$$

In the above, the *regularizer function* $h_i \colon \mathcal{X}_i \to \mathbb{R}$ is a convex penalty term which smoothens the "hard" $\arg\max$ correspondence $y_i \mapsto \arg\max_{x_i \in \mathcal{X}_i} \langle y_i, x_i \rangle$ that maximizes the player's cumulative payoff over $[0,t]$. As a result, the "regularized leader" $Q_i(y_i) = \arg\max_{x_i \in \mathcal{X}_i} \{\langle y_i, x_i \rangle - h_i(x_i)\}$ is biased towards the *prox-center* $p_i = \arg\min_{x_i \in \mathcal{X}_i} h_i(x_i)$ of $\mathcal{X}_i$. For most common regularizers, the prox-center is interior (and usually coincides with the barycenter of $\mathcal{X}$), so the regularization in (3.2) encourages exploration by favoring mixed strategies with full support.

In Appendix A, we present in detail two of the prototypical examples of (FoReL): *i*) the *multiplicative weights* (MW) dynamics induced by the entropic regularizer function $h_i(x) = \sum_{\alpha_i \in \mathcal{A}_i} x_{i\alpha_i} \log x_{i\alpha_i}$ (which lead to the replicator dynamics of evolutionary game theory); and *ii*) the projection dynamics induced by the Euclidean regularizer $h_i(x) = \frac{1}{2}\|x_i\|^2$. For concreteness, we will assume in what follows that the regularizer of every player $i \in \mathcal{N}$ satisfies the following minimal requirements:

(1) h_i is continuous and strictly convex on $\mathcal{X}_i$.
(2) h_i is smooth on the relative interior of every face of $\mathcal{X}_i$ (including $\mathcal{X}_i$ itself).

Under these basic assumptions, the "regularized leader" $Q_i(y_i)$ is well-defined in the sense that (3.2) admits a unique solution. More importantly, we have the following no-regret guarantee:

Theorem 3.1. *A player following* (FoReL) *enjoys an $\mathcal{O}(1/t)$ regret bound, no matter what other players do. Specifically, if player $i \in \mathcal{N}$ follows* (FoReL)*, then, for every continuous trajectory of play $x_{-i}(t)$ of the opponents of player i, we have*

$$\operatorname{Reg}_i(t) \leq \frac{\Omega_i}{t}, \tag{3.3}$$

where $\Omega_i = \max h_i - \min h_i$ is a positive constant.

To streamline our discussion, we relegate the proof of Theorem 3.1 to Appendix C; we also refer to [20] for a similar regret bound for (FoReL) in the context of online convex optimization. Instead of discussing the proof, we close this section by noting that (3.3) represents a striking improvement over the $\Theta(t^{-1/2})$ worst-case bound for FoReL in discrete time [40]. In view of this, the continuous-time framework we consider here can be seen as particularly amenable to learning because it allows players seek to minimize their regret (and thus converge to coarse correlated equilibria) at the fastest possible rate.

4. Recurrence in adversarial regularized learning

In this section, our aim is to take a closer look at the ramifications of fast regret minimization under (FoReL) beyond convergence to the set of coarse correlated equilibria. Indeed, as is well known, this set is fairly large and may contain thoroughly non-rationalizable strategies: for instance, Viossat and Zapechelnyuk [46] recently showed that a coarse correlated equilibrium could assign positive selection probability *only* to strictly dominated strategies. Moreover, the time-averaging that is inherent in the definition of the players' regret leaves open the possibility of complex day-to-day behavior e.g. periodicity, recurrence, limit cycles or chaos [26, 27, 29, 37]. Motivated by this, we examine the long-run behavior of the (FoReL) in the popular setting of zero-sum games (with or without interior equilibria) and several extensions thereof.

A key notion in our analysis is that of (*Poincaré*) *recurrence*. Intuitively, a dynamical system is recurrent if, after a sufficiently long (but *finite*) time, almost every state returns arbitrarily close to the system's initial state.[4] More formally, given a dynamical system on $\mathcal{X}$ that is defined by means of a *semiflow* $\Phi\colon \mathcal{X} \times [0,\infty) \to \mathcal{X}$, we have:[5]

Definition 4.1. A point $x \in \mathcal{X}$ is said to be *recurrent* under Φ if, for every neighborhood U of x in $\mathcal{X}$, there exists an increasing sequence of times $t_n \uparrow \infty$ such that

[4] Here, "almost" means that the set of such states has full Lebesgue measure.

[5] Recall that a continuous map $\Phi\colon \mathcal{X}\times[0,\infty) \to \mathcal{X}$ is a *semiflow* if $\Phi(x,0) = x$ and $\Phi(x,t+s) = \Phi(\Phi(x,t),s)$ for all $x \in \mathcal{X}$ and all $s,t \geq 0$. Heuristically, $\Phi_t(x) \equiv \Phi(x,t)$ describes the trajectory of the dynamical system starting at x.

$\Phi(x, t_n) \in U$ for all n. Moreover, the flow Φ is called (*Poincaré*) *recurrent* if, for every measurable subset A of $\mathcal{X}$, the set of recurrent points in A has full measure.

An immediate consequence of Definition 4.1 is that, if a point is recurrent, there exists an increasing sequence of times $t_n \uparrow \infty$ such that $\Phi(x, t_n) \to x$. On that account, recurrence can be seen as the flip side of convergence: under the latter, (almost) every initial state of the dynamics eventually reaches some well-defined end-state; instead, under the former, the system's orbits fill the entire state space and return arbitarily close to their starting points infinitely often (so there is no possibility of convergence beyond trivial cases).

4.1. **Zero-sum games with an interior equilibrium.** Our first result is that (FoReL) is recurrent (and hence, non-convergent) in zero-sum games with an interior Nash equilibrium:

Theorem 4.2. *Let Γ be a 2-player zero-sum game that admits an interior Nash equilibrium. Then, almost every solution trajectory of* (FoReL) *is recurrent; specifically, for* (*Lebesgue*) *almost every initial condition $x(0) = Q(y(0)) \in \mathcal{X}$, there exists an increasing sequence of times $t_n \uparrow \infty$ such that $x(t_n) \to x(0)$.*

The proof of Theorem 4.2 is fairly complicated, so we outline the basic steps below:

(1) We first show that the dynamics of the score sequence $y(t)$ are *incompressible*, i.e. the volume of a set of initial conditions remains invariant as the dynamics evolve over time. By Poincaré's recurrence theorem (cf. Appendix B), if every solution orbit $y(t)$ of (FoReL) remains in a compact set for all $t \geq 0$, incompressibility implies recurrence.

(2) To counter the possibility of solutions escaping to infinity, we introduce a transformed system based on the differences between scores (as opposed to the scores themselves). To establish boundedness in these dynamics, we consider the "primal-dual" coupling

$$G(y) = \sum_{i \in \mathcal{N}} [h_i^*(y_i) - \langle y_i, x_i^* \rangle], \tag{4.1}$$

where x^* is an interior Nash equilibrium and $h_i^*(y_i) = \max_{x_i \in \mathcal{X}_i} \{\langle y_i, x_i \rangle - h_i(x_i)\}$ denotes the convex conjugate of h_i.[6] The key property of this coupling is that it remains invariant under (FoReL); however, its level sets are not bounded so, again, precompactness of solutions is not guaranteed.

(3) Nevertheless, under the score transformation described above, the level sets of G *are* compact. Since the transformed dynamics are invariant under said transformation, Poincaré's theorem finally implies recurrence.

Proof of Theorem 4.2. To make the above plan precise, fix some "benchmark" strategy $\hat{\alpha}_i \in \mathcal{A}_i$ for every player $i \in \mathcal{N}$ and, for all $\alpha_i \in \mathcal{A}_i \setminus \{\hat{\alpha}_i\}$, consider the corresponding score differences

$$z_{i\alpha_i} = y_{i\alpha_i} - y_{i,\hat{\alpha}_i}. \tag{4.2}$$

Obviously, $z_{i\hat{\alpha}_i} = y_{i\hat{\alpha}_i} - y_{i\hat{\alpha}_i}$ is identically zero so we can ignore it in the above definition. In so doing, we obtain a linear map $\Pi_i \colon \mathbb{R}^{\mathcal{A}_i} \to \mathbb{R}^{\mathcal{A}_i \setminus \{\hat{\alpha}_i\}}$ sending $y_i \mapsto z_i$; aggregating over all players, we also write Π for the product map $\Pi = (\Pi_1, \dots, \Pi_N)$

[6] This coupling is closely related to the so-called *Bregman divergence* – for the details, see [3, 19, 24, 40].

sending $y \mapsto z$. For posterity, note that this map is surjective but *not* injective,[7] so it does not allow us to recover the score vector y from the score difference vector z.

Now, under (FoReL), the score differences (4.2) evolve as

$$\dot{z}_{i\alpha_i} = v_{i\alpha_i}(x(t)) - v_{i\hat{\alpha}_i}(x(t)). \tag{4.3}$$

However, since the right-hand side (RHS) of (4.3) depends on $x = Q(y)$ and the mapping $y \mapsto z$ is not invertible (so y cannot be expressed as a function of z), the above does not a priori constitute an autonomous dynamical system (as required to apply Poincaré's recurrence theorem). Our first step below is to show that (4.3) does in fact constitute a well-defined dynamical system on z.

To do so, consider the reduced choice map $\hat{Q}_i \colon \mathbb{R}^{\mathcal{A}_i \setminus \{\hat{\alpha}_i\}} \to \mathcal{X}_i$ defined as

$$\hat{Q}_i(z_i) = Q_i(y_i) \tag{4.4}$$

for some $y_i \in \mathbb{R}^{\mathcal{A}_i}$ such that $\Pi_i(y_i) = z_i$. That such a y_i exists is a consequence of Π_i being surjective; furthemore, that $\hat{Q}_i(z_i)$ is well-defined is a consequence of the fact that Q_i is invariant on the fibers of Π_i. Indeed, by construction, we have $\Pi_i(y_i) = \Pi_i(y_i')$ if and only if $y'_{i\alpha_i} = y_{i\alpha_i} + c$ for some $c \in \mathbb{R}$ and all $\alpha_i \in \mathcal{A}_i$. Hence, by the definition of Q_i, we get

$$\begin{aligned} Q_i(y_i') &= \arg\max_{x_i \in \mathcal{X}_i} \bigl\{ \langle y_i, x_i \rangle + c \textstyle\sum_{\alpha_i \in \mathcal{A}_i} x_{i\alpha_i} - h_i(x_i) \bigr\} \\ &= \arg\max_{x_i \in \mathcal{X}_i} \{ \langle y_i, x_i \rangle - h_i(x_i) \} = Q_i(y_i), \end{aligned} \tag{4.5}$$

where we used the fact that $\sum_{\alpha_i \in \mathcal{A}_i} x_{i\alpha_i} = 1$. The above shows that $Q_i(y_i') = Q_i(y_i)$ if and only if $\Pi_i(y_i) = \Pi_i(y_i')$, so $\hat{Q}_i$ is well-defined.

Letting $\hat{Q} \equiv (\hat{Q}_1, \dots, \hat{Q}_N)$ denote the aggregation of the players' individual choice maps $\hat{Q}_i$, it follows immediately that $Q(y) = \hat{Q}(\Pi(y)) = \hat{Q}(z)$ by construction. Hence, the dynamics (4.3) may be written as

$$\dot{z} = V(z), \tag{4.6}$$

where

$$V_{i\alpha_i}(z) = v_{i\alpha_i}(\hat{Q}_i(z)) - v_{i\hat{\alpha}_i}(\hat{Q}_i(z)). \tag{4.7}$$

These dynamics obviously constitute an autonomous system, so our goal will be to use Liouville's formula and Poincaré's theorem in order to establish recurrence and then conclude that the induced trajectory of play $x(t)$ is recurrent by leveraging the properties of $\hat{Q}$.

As a first step towards applying Liouville's formula, we note that the dynamics (4.6) are *incompressible*. Indeed, we have

$$\frac{\partial V_{i\alpha_i}}{\partial z_{i\alpha_i}} = \sum_{\beta_i \in \mathcal{A}_i} \frac{\partial V_{i\alpha_i}}{\partial x_{i\beta_i}} \frac{\partial x_{i\beta_i}}{\partial z_{i\alpha_i}} = 0, \tag{4.8}$$

because v_i does not depend on x_i. We thus obtain $\operatorname{div}_z V(z) = 0$, i.e. the dynamics (4.6) are incompressible.

We now show that every solution orbit $z(t)$ of (4.6) is *precompact*, that is, $\sup_{t \geq 0} \lVert z(t) \rVert < \infty$. To that end, note that the coupling $G(y) = \sum_{i \in \mathcal{N}} [h_i^*(y_i) -$

[7] Specifically, $\Pi_i(y_i) = \Pi_i(y_i')$ if and only if $y'_{i\alpha_i} = y_{i\alpha_i} + c$ for some $c \in \mathbb{R}$ and all $\alpha_i \in \mathcal{A}_i$.

$\langle y_i, x_i^* \rangle]$ defined in (4.1) remains invariant under (FoReL) when Γ is a 2-player zero-sum game. Indeed, by Lemma C.1, we have

$$\begin{aligned} \frac{dG}{dt} &= \sum_{i \in \mathcal{N}} \langle v_i(x), x_i - x_i^* \rangle = \langle v_1(x), x_1 - x_1^* \rangle + \langle v_2(x), x_2 - x_2^* \rangle \\ &\qquad = u_1(x_1, x_2) - u_1(x_1^*, x_2) + u_2(x_1, x_2) - u_2(x_1, x_2^*) = 0, \end{aligned} \tag{4.9}$$

where we used the fact that $Q_i = \nabla h_i^*$ in the first line (cf. (C.2) above), and the assumption that x^* is an interior Nash equilibrium of a 2-player zero-sum game in the last one. We conclude that $G(y(t))$ remains constant under (FoReL), as claimed.

By Lemma D.2 in Appendix D, the invariance of $G(y(t))$ under (FoReL) implies that the score differences $z_{i\alpha_i}(t) = y_{i\alpha_i}(t) - y_{i\hat{\alpha}_i}(t)$ also remain bounded for all $t \geq 0$. Hence, by Liouville's formula and Poincaré's recurrence theorem, the dynamics (4.6) are *recurrent*, i.e. for (Lebesgue) almost every initial condition z_0 and every neighborhood U of z_0, there exists some τ_U such that $z(\tau_U) \in U$ (cf. Definition 4.1). Thus, taking a shrinking net of balls $\mathbb{B}_n(z_0) = \{z : \|z - z_0\| \leq 1/n\}$ and iterating the above, it follows that there exists an increasing sequence of times $t_n \uparrow \infty$ such that $z(t_n) \to z_0$. Therefore, to prove the corresponding claim for the induced trajectories of play $x(t) = Q(y(t)) = \hat{Q}(z(t))$ of (FoReL), fix an initial condition $x_0 \in \mathcal{X}^\circ$ and take some z_0 such that $x_0 = \hat{Q}(z_0)$. By taking t_n as above, we have $z(t_n) \to z_0$ so, by continuity, $x(t_n) = \hat{Q}(z_n) \to \hat{Q}(z_0) = x_0$. This shows that any solution orbit $x(t)$ of (FoReL) is recurrent and our proof is complete. □

Remark. We close this section by noting that the invariance of (4.1) under (FoReL) induces a foliation of $\mathcal{X}$, with each individual trajectory of (FoReL) living on a "leaf" of the foliation (a level set of G). Fig. 1 provides a schematic illustration of this foliation/cycling structure.

4.2. **Zero-sum games with no interior equilibria.** At first sight, Theorem 4.2 suggests that cycling is ubiquitous in zero-sum games; however, if the game does not admit an interior equilibrium, the behavior of (FoReL) turns out to be qualitatively different. To state our result for such games, it will be convenient to assume that the players' regularizer functions are *strongly convex*, i.e. each h_i can be bounded from below by a quadratic minorant:

$$h_i(tx_i + (1-t)x_i') \leq t h_i(x_i) + (1-t) h_i(x_i') - \tfrac{1}{2} K_i t(1-t) \|x_i - x_i'\|^2, \tag{4.10}$$

for all $x_i, x_i' \in \mathcal{X}_i$ and for all $t \in [0,1]$. Under this technical assumption, we have:

Theorem 4.3. *Let Γ be a 2-player zero-sum game that does not admit an interior Nash equilibrium. Then, for every initial condition of* (FoReL)*, the induced trajectory of play $x(t)$ converges to the boundary of $\mathcal{X}$. Specifically, if x^* is a Nash equilibrium of Γ with maximal support, $x(t)$ converges to the relative interior of the face of $\mathcal{X}$ spanned by* $\operatorname{supp}(x^*)$.

Theorem 4.3 is our most comprehensive result for the behavior of (FoReL) in zero-sum games, so several remarks are in order. First, we note that Theorem 4.3 complements Theorem 4.2 in a very natural way: specifically, if Γ admits an interior Nash equilibrium, Theorem 4.3 suggests that the solutions of (FoReL) will stay within the relative interior $\mathcal{X}^\circ$ of $\mathcal{X}$ (since an interior equilibrium is supported on all actions). Of course, Theorem 4.2 provides a stronger result because it states that, within $\mathcal{X}^\circ$, (FoReL) is recurrent. Hence, applying both results in tandem, we obtain the following heuristic for the behavior of (FoReL) in zero-sum games:

In the long run, (FoReL) *wanders in perpetuity*
in the smallest face of $\mathcal{X}$ containing the equilibrium set of Γ.

This leads to two extremes: On the one hand, if Γ admits an interior equilibrium, (FoReL) is recurrent and cycles in the level sets of the coupling function (4.1). At the other end of the spectrum, if Γ admits only a single, pure equilibrium, then (FoReL) converges to it (since it has to wander in a singleton set). In all other "in-between" cases, (FoReL) exhibits a hybrid behavior, converging to the face of $\mathcal{X}$ that is spanned by the maximal support equilibrium of Γ, and then cycling in that face in perpetuity.

The reason for this behavior is that the coupling (4.1) is no longer a constant of motion of (FoReL) if the game does not admit an interior equilibrium. As we show in Appendix C, the coupling (4.1) is strictly decreasing when the support of $x(t)$ is strictly greater than that of a Nash equilibrium x^* with maximal support. When the two match, the rate of change of (4.1) drops to zero, and we fall back to a "constrained" version of Theorem 4.2. We make this argument precise in Appendix C (where we present the proof of Theorem 4.3).

4.3. Zero-sum polymatrix games & positive affine payoff transformations.

We close this section by showing that the recurrence properties of (FoReL) are not unique to "vanilla" zero-sum games, but also occur when there is a *network of competitors* – i.e. in N-player zero-sum polymatrix games. In fact, the recurrence results carry over to any N-player game which is isomorphic to a constant-sum polymatrix game with an interior equilibrium up to a positive-affine payoff transformation (possibly different transformation for each agent). For example, this class of games contains all strictly competitive games [1]. Such transformations do not affect the equilibrium structure of the game, but can affect the geometry of the trajectories; nevertheless, the recurrent behavior persists as shown by the following result:

Theorem 4.4. *Let $\Gamma = (\Gamma_e)_{e\in\mathcal{E}}$ be a constant-sum polymatrix game (or a positive affine payoff transformation thereof). If Γ admits an interior Nash equilibrium, almost every solution trajectory of* (FoReL) *is recurrent; specifically, for (Lebesgue) almost every initial condition $x(0) = Q(y(0)) \in \mathcal{X}$, there exists an increasing sequence of times $t_n \uparrow \infty$ such that $x(t_n) \to x(0)$.*

We leave the case of zero-sum polymatrix games with no interior equilibria to future work.

5. Conclusions

Our results show that the behavior of regularized learning in adversarial environments is considerably more intricate than the strong no-regret properties of FoReL might at first suggest. Even though the empirical frequency of play under FoReL converges to the set of coarse correlated equilibria (possibly at an increased rate, depending on the game's structure), the actual trajectory of play under FoReL is recurrent and exhibits cycles in zero-sum games. We find this property particularly interesting as it suggests that "black box" guarantees are not the be-all/end-all of learning in games: the theory of dynamical systems is rife with complex phenomena and notions that arise naturally when examining the behavior of learning algorithms in finer detail.

Appendix A. Examples of FoReL dynamics

Example A.1 (Multiplicative weights and the replicator dynamics)*.* Perhaps the most widely known example of a regularized choice map is the so-called *logit choice map*

$$\Lambda_i(y) = \frac{(\exp(y_{i\alpha_i}))_{\alpha_i \in \mathcal{A}_i}}{\sum_{\beta_i \in \mathcal{A}_i} \exp(y_{i\beta_i})}. \tag{A.1}$$

This choice model was first studied in the context of discrete choice theory by McFadden [22] and it leads to the *multiplicative weights* (MW) dynamics:[8]

$$\begin{aligned} \dot{y}_i &= v_i(x), \\ x_i &= \Lambda_i(y_i). \end{aligned} \tag{MW}$$

As is well known, the logit map above is obtained by the model (3.2) by considering the entropic regularizer

$$h_i(x) = \sum_{\alpha_i \in \mathcal{A}_i} x_{i\alpha_i} \log x_{i\alpha_i}, \tag{A.2}$$

i.e. the (negative) *Gibbs–Shannon entropy function.* A simple differentiation of (MW) then shows that the players' mixed strategies evolve according to the dynamics

$$\dot{x}_{i\alpha_i} = x_{i\alpha_i} \left[v_{i\alpha_i}(x) - \sum_{\beta_i \in \mathcal{A}_i} x_{i\beta_i} v_{i\beta_i}(x) \right], \tag{RD}$$

This equation describes the *replicator dynamics* of [45], the most widely studied model for evolution under natural selection in population biology and evolutionary game theory. The basic relation between (MW) and (RD) was first noted in a single-agent environment by [34] and was explored further in game theory by [17, 23, 24, 42] and many others.

Example A.2 (Euclidean regularization and the projection dynamics)*.* Another widely used example of regularization is given by the *quadratic penalty*

$$h_i(x_i) = \frac{1}{2} \sum_{\alpha_i \in \mathcal{A}_i} x_{i\alpha_i}^2. \tag{A.3}$$

The induced choice map (3.2) is the (Euclidean) *projection map*

$$\Pi_i(y_i) = \arg\max\nolimits_{x_i \in \mathcal{X}_i} \left\{ \langle y_i, x_i \rangle - \tfrac{1}{2} \|x_i\|_2^2 \right\} = \arg\min\nolimits_{x_i \in \mathcal{X}_i} \|y_i - x_i\|_2^2, \tag{A.4}$$

leading to the *projected reinforcement learning* process

$$\begin{aligned} \dot{y}_i &= v_i(x), \\ x_i &= \Pi_i(y_i). \end{aligned} \tag{PL}$$

[8] The terminology "multiplicative weights" refers to the fact that (MW) is the continuous version of the discrete-time multiplicative weights update rule:

$$x_{i\alpha_i}(t+1) = \frac{x_{i\alpha_i}(t) e^{\eta_i v_{i\alpha_i}(x(t))}}{\sum_{\beta_i \in \mathcal{A}_i} x_{i\beta_i}(t) e^{\eta_i v_{i\beta_i}(x(t))}}, \tag{MWU}$$

where $\eta_i > 0$ is the scheme's "learning rate". For more details about (MWU), we refer the reader to [2].

The players' mixed strategies are then known to follow the *projection dynamics*

$$\dot{x}_{i\alpha_i} = \begin{cases} v_{i\alpha_i}(x) - |\mathrm{supp}(x_i)|^{-1} \sum_{\beta_i \in \mathrm{supp}(x_i)} v_{i\beta_i}(x) & \text{if } \alpha_i \in \mathrm{supp}(x_i), \\ 0 & \text{if } \alpha_i \notin \mathrm{supp}(x_i), \end{cases} \tag{PD}$$

over all intervals for which the support of $x(t)$ remains constant [24]. The dynamics (PD) were introduced in game theory by [12] as a geometric model of the evolution of play in population games; for a closely related approach, see also [21, 25] and references therein.

Appendix B. Liouville's formula and Poincaré recurrence

Below we present for completeness some basic results from the theory of dynamical systems.

Liouville's Formula. Liouville's formula can be applied to any system of autonomous differential equations with a continuously differentiable vector field ξ on an open domain of $\mathcal{S} \subset^k$. The divergence of ξ at $x \in \mathcal{S}$ is defined as the trace of the corresponding Jacobian at x, *i.e.*, $\mathrm{div}[\xi(x)] = \sum_{i=1}^{k} \frac{\partial \xi_i}{\partial x_i}(x)$. Since divergence is a continuous function we can compute its integral over measurable sets $A \subset \mathcal{S}$. Given any such set A, let $A(t) = \{\Phi(x_0, t) : x_0 \in A\}$ be the image of A under map Φ at time t. $A(t)$ is measurable and is volume is $\mathrm{vol}[A(t)] = \int_{A(t)} dx$. Liouville's formula states that the time derivative of the volume $A(t)$ exists and is equal to the integral of the divergence over $A(t)$:

$$\frac{d}{dt}[A(t)] = \int_{A(t)} \mathrm{div}[\xi(x)] dx.$$

A vector field is called divergence free if its divergence is zero everywhere. Liouville's formula trivially implies that volume is preserved in such flows.

Poincaré's recurrence theorem. The notion of recurrence that we will be using in this paper goes back to Poincaré and specifically to his study of the three-body problem. In 1890, in his celebrated work [30], he proved that whenever a dynamical system preserves volume almost all trajectories return arbitrarily close to their initial position, and they do so an infinite number of times. More precisely, Poincaré established the following:

Poincaré Recurrence: *[4, 30] If a flow preserves volume and has only bounded orbits then for each open set there exist orbits that intersect the set infinitely often.*

Appendix C. Technical proofs

The first result that we prove in this appendix is a key technical lemma concerning the evolution of the coupling function (4.1):

Lemma C.1. *Let $p_i \in \mathcal{X}_i$ and let $G_i(y_i) = h_i^*(y_i) - \langle y_i, p_i \rangle$ denote the coupling (4.1) for player $i \in \mathcal{N}$. If player $i \in \mathcal{N}$ follows (FoReL), we have*

$$\frac{d}{dt} G_i(y_i(t)) = \langle v_i(x(t)), x_i(t) - p_i \rangle, \tag{C.1}$$

for every trajectory of play $x_{-i}(t)$ of all players other than i.

Proof. We begin by recalling the "maximizing argument" identity

$$Q_i(y_i) = \nabla h_i^*(y_i) \tag{C.2}$$

which expresses the choice map Q_i as a function of the convex conjugate of h_i [40, p. 149]. With this at hand, a simple differentiation gives

$$\begin{aligned}\frac{d}{dt} G_i(y_i(t)) &= \frac{d}{dt} h_i^*(y_i(t)) - \langle \dot{y}_i(t), p_i \rangle \\ &= \langle \dot{y}_i(t), \nabla h_i^*(y_i(t)) - p_i \rangle \\ &= \langle v_i(x(t)), x_i(t) - p_i \rangle, \end{aligned} \tag{C.3}$$

where the last step follows from the fact that $x_i(t) = Q_i(y_i(t)) = \nabla h_i^*(y_i(t))$. □

Armed with this lemma, we proceed to prove the no-regret guarantees of (FoReL):

Proof of Theorem 3.1. Fix some base point $p_i \in \mathcal{X}_i$ and let $L_i(t) = G_i(y_i(t)) = h_i(y_i(t)) - \langle y_i(t), p_i \rangle$. Then, by Lemma C.1, we have

$$L_i'(t) = \langle v_i(x(t)), x_i(t) - p_i \rangle \tag{C.4}$$

and hence, after integrating and rearranging, we get

$$\int_0^t [u_i(p_i; x_{-i}(s)) - u_i(x(s))]\, ds = \int_0^t \langle v_i(x(s)), p_i - x_i(s) \rangle\, ds = L_i(0) - L_i(t), \tag{C.5}$$

where we used the fact that $u_i(p_i; x_{-i}) = \langle v_i(x), p_i \rangle$ – cf. Eq. (2.2) in Section 2. However, expanding the RHS of (C.5), we get

$$\begin{aligned} L_i(0) - L_i(t) &= h_i^*(y_i(0)) - \langle y_i(0), p_i \rangle - h_i^*(y_i(t)) + \langle y_i(t), p_i \rangle \\ &\leq h_i^*(y_i(0)) - \langle y_i(0), p_i \rangle + h_i(p_i) \\ &= h_i(p_i) - h_i(Q_i(y_i(0))) \\ &\leq \max h_i - \min h_i \equiv \Omega_i, \end{aligned} \tag{C.6}$$

where we used the defining property of convex conjugation in the second and third lines above – i.e. that $h_i^*(y_i) \geq \langle y_i, x_i \rangle - h_i(x_i)$ for all $x_i \in \mathcal{X}_i$, with equality if and only if $x_i = Q_i(y_i)$. Thus, maximizing (C.5) over $p_i \in \mathcal{X}_i$, we finally obtain

$$\operatorname{Reg}_i(t) = \max_{p_i \in \mathcal{X}_i} \frac{1}{t} \int_0^t [u_i(p_i; x_{-i}(s)) - u_i(x(s))]\, ds \leq \frac{\Omega_i}{t}, \tag{C.7}$$

as claimed. □

We now turn to two-player zero-sum games that do not admit interior equilibria. To describe such equilibria in more detail, we consider below the notion of *essential* and *non-essential* strategies:

Definition C.2. A strategy α_i of agent $i \in \{1, 2\}$ in a zero sum game is called *essential* if there exists a Nash equilibrium in which player i plays α_i with positive probability. A strategy that is not essential is called *non-essential.*

As it turns out, the Nash equilibria of a zero-sum game admit a very useful characterization in terms of essential and non-essential strategies:

Lemma C.3. *Let Γ be a 2-player zero-sum game that does not admit an interior Nash equilibrium. Then, there exists a mixed Nash equilibrium (x_1, x_2) such that a) each agent plays each of their essential strategies with positive probability; and* *b) for each agent deviating to a non-essential strategy results to a strictly worse performance than the value of the game.*

The key step in proving this characterization is Farkas' lemma; the version we employ here is due to Gale, Kuhn and Tucker [13]):

Lemma C.4 (Farkas' lemma)**.** *Let $\mathbf{P} \in \mathbb{R}^{m\times n}$ and $\mathbf{b} \in \mathbb{R}^m$. Then exactly one of the following two statements is true:*

- *There exists a $\mathbf{x} \in \mathbb{R}^m$ such that $\mathbf{P}^\top \mathbf{x} \geq 0$ and $\mathbf{b}^\top \mathbf{x} < 0$.*
- *There exists a $\mathbf{y} \in \mathbb{R}^n$ such that $\mathbf{P} \cdot \mathbf{y} = \mathbf{b}$ and $\mathbf{y} \geq 0$.*

With this lemma at hand, we have:

Proof of Lemma C.3. Assume without loss of generality that the value of the zero-sum game is zero. and that the first agent is a maximizing agent. Let A be the payoff matrix of the first agent and hence $A^T = A$ the payoff matrix of the second/minimizing agent. We will show first that for any non-essential strategy α_i of each agent there exists a Nash equilibrium strategy of his opponent such that the expected performance of α_i is strictly worse than the value of the game (i.e. zero).

It suffices to argue this for the first agent. Let α_i be one of his non-essential strategies then by definition there does not exist any Nash equilibrium strategy of that agent that chooses α_i with positive probability. This is equivalent to the negation of the following statement:

There exists a $\mathbf{x} \in \mathbb{R}^m$ such that $\mathbf{P}^\top \mathbf{x} \geq 0$ and $\mathbf{b}^\top \mathbf{x} < 0$
where

$$\mathbf{P}^\top = \begin{pmatrix} \mathbf{A}^\top \\ \mathbf{I}_{m\times m} \end{pmatrix} = \begin{pmatrix} a_{11} & a_{21} & \dots & a_{m1} \\ \vdots & \vdots & \dots & \vdots \\ a_{1n} & a_{2n} & \dots & a_{mn} \\ 1 & 0 & \dots & 0 \\ 0 & 1 & \dots & 0 \\ \vdots & \vdots & \dots & \vdots \\ 0 & 0 & \dots & 1 \end{pmatrix}, \tag{C.8}$$

and $\mathbf{b} = -\mathbf{e}_i = (0, \dots, 0, -1, 0, \dots, 0)^T$, the standard basis vector of dimension m that “chooses" the i-th strategy. By Farkas' lemma, there exists a $\mathbf{y} \in \mathbb{R}^{m+n}$ such that $\mathbf{Py} = \mathbf{b}$ and $\mathbf{y} \geq 0$. It is convenient to express $\mathbf{y} = (\mathbf{z}; \mathbf{w})$ where $\mathbf{z} \in \mathbb{R}^n$ and $\mathbf{w} \in \mathbb{R}^m$. Hence, for all $j \neq i \in \{1, 2, \dots, m\} : (\mathbf{Py})_j = (\mathbf{Az})_j + \mathbf{w}_j = 0$ and thus $(\mathbf{Az})_j \leq 0$. Finally, for $j = i : (\mathbf{Py})_i = (\mathbf{Az})_i + \mathbf{w}_i = -1$ and thus $(\mathbf{Az})_i < 0$. Hence $\mathbf{z}$ is a Nash equilibrium strategy for the second player such that when the first agent chooses the non-essential strategy α_i he receives payoff which is strictly worse than his value (zero).

To complete the proof, for each essential strategy of the first agent there exists one equilibrium strategy of his that chooses it with positive probability (by definition). Similarly, for each non-essential strategy of the second agent there exists one equilibrium strategy of the first agent such that makes the expected payoff of that non-essential strategy strictly worse than the value of the game. The barycenter of all the above equilibrium strategies is still an equilibrium strategy (by convexity) and has all the desired properties. □

With all this at hand, we are finally in a position to prove Theorem 4.3:

Proof of Theorem 4.3. We first show that the coupling $G(y) = \sum_{i\in\mathcal{N}}[h_i^*(y_i) - \langle y_i, x_i^*\rangle]$ defined in (4.1) given any fully mixed initial condition strictly increases under (FoReL) when Γ is a 2-player zero-sum game that does not have an equilibrium with full support.

Indeed, by (C.3) there exists a mixed Nash equilibrium (x_1^*, x_2^*) such that *i*) both players employ each of their essential strategies with positive probability over time; and *ii*) every player deviating to a non-essential strategy obtains a payoff lower than the value of the game. As a result, any player playing an interior (fully mixed) strategy against such an equilibrium strategy must receive less utility than their value. In more detail, we have

$$\begin{aligned}\frac{dG}{dt} &= \sum_{i\in\mathcal{N}}\langle v_i(x), x_i - x_i^*\rangle = \langle v_1(x), x_1 - x_1^*\rangle + \langle v_2(x), x_2 - x_2^*\rangle \\ &= u_1(x_1, x_2) - u_1(x_1^*, x_2) + u_2(x_1, x_2) - u_2(x_1, x_2^*) \\ &= -u_1(x_1^*, x_2) - u_2(x_1, x_2^*) \\ &< -u_1(x_1^*, x_2^*) - u_2(x_1^*, x_2^*) = 0, \end{aligned} \tag{C.9}$$

where we used the fact that $Q_i = \nabla h_i^*$ in the first line (cf. Appendix D), and the assumption that x^* is a Nash equilibrium of a 2-player zero-sum game such that any agent playing an interior (fully mixed) strategy against such an equilibrium strategy must receive less utility than their value (and hence the agent himself receives more utility than the value of the game). We thus conclude that $G(y(t))$ strictly increases under (FoReL), as claimed.

Let $x^* = (x_1^*, x_2^*)$ be the Nash equilibrium identified in (C.9) and let $L(t) = G(y(t)) = \sum_{i\in\mathcal{N}}[h_i^*(y_i(t)) - \langle y_i(t), x_i^*\rangle]$ denote the primal-dual coupling (4.1) between $y(t)$ and x^*. From (C.9), we have that starting from any fully mixed strategy profile $x(0) \in \prod_i \operatorname{int}(\mathcal{X}_i)$ and for all $t \geq 0$, $L'(t) = \langle \dot{y}(t), \nabla G(y(t))\rangle < 0$. However, G is bounded from below by $-\sum_i \max_{x_i\in\mathcal{X}_i} h_i(x_i)$, and since $G(y(t))$ is strictly decreasing, it must exhibit a finite limit.

We begin by noting that $x(t) = Q(y(t))$ is Lipschitz continuous in t. Indeed, v is Lipschitz continuous on $\mathcal{X}$ by linearity; furthermore, since the regularizer functions h_i are assumed K_i-strongly convex, it follows that Q_i is $(1/K_i)$-continuous by standard convex analysis arguments [32, Theorem 12.60]. In turn, this implies that the field of motion $V(y) \equiv v(Q(y))$ of (FoReL) is Lipschitz continuous, so the dynamics are well-posed and $y(t)$ is differentiable. Since $\dot{y} = v$ and, in addition, v is bounded on $\mathcal{X}$, we conclude that $\dot{y}$ is bounded so, in particular, $y(t)$ is Lipschitz continuous on $[0, \infty)$. We thus conclude that $x(t) = Q(y(t))$ is Lipschitz continuous as the composition of Lipschitz continuous functions.

We now further claim that $L'(t)$ is also Lipschitz continuous in t. Indeed, by (C.1), we have $L'(t) = \sum_{i\in\mathcal{N}}\langle v_i(x(t)), x_i(t) - x_i^*\rangle$; since v_i is Lipschitz continuous in x and $x(t)$ is Lipschitz continuous in t, our claim follows trivially. Hence, by Lemma D.3, we conclude that $\lim_{t\to\infty} L'(t) = \lim_{\to\infty}\sum_{i\in\mathcal{N}}\langle v_i(x(t), x_i(t) - x_i^*\rangle = 0$.

By (C.9), we know that $L'(t) < 0$ as long as $x(t)$ is interior. Hence, any ω-limit $\hat{x}$ of $x(t)$ cannot be interior (given that the embedded game does not have any interior Nash equilibria). Moreover, we can repeat this argument for any subspace such that the restriction of the game on that subspace (when ignoring the strategies that are played with probability zero) does not have a fully mixed NE. We thus

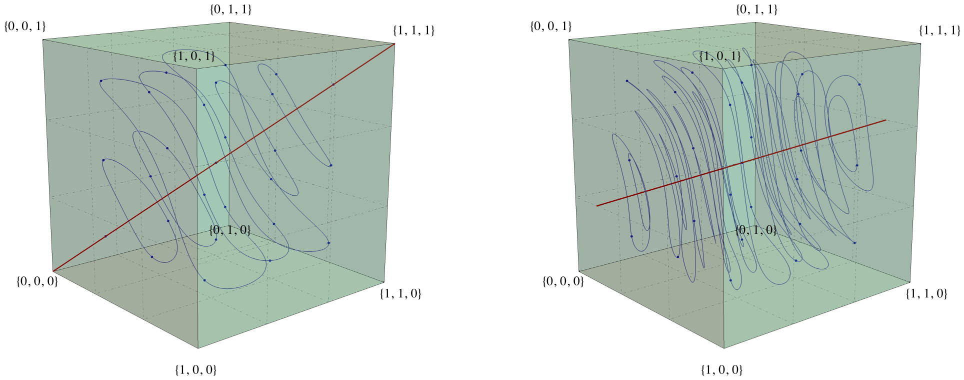

Figure 2. Evolution of the multiplicative weights dynamics (MW) in a 3-player zero-sum polymatrix game. In the left subfigure, each pair of players faces off in a game of standard (symmetric) Matching Pennies; in the right, the game on each pair is weighted by a different factor. In both cases, we plot the solution trajectories of (MW) for the same initial conditions. Even though the different weights change the trajectories of (MW) and the game's equilibrium set, the cycling behavior of the dynamics remains unaffected.

conclude that the support of $\hat{x}$ must be a subset of the support of x^*. Since Γ does not admit an interior equilibrium, x^* does not have full support, so every ω-limit of $x(t)$ lies on the boundary of $\mathcal{X}$, as claimed. □

We close this appendix with the proof of our result on constant-sum polymatrix games (and positive affine transformations thereof):

Proof of Theorem 4.4. Our proof follows closely that of Theorem 4.2; to streamline our presentation, we only highlight here the points that differ due to working with (an positive-affine transformations of) a network of constant-sum games (as opposed to a single 2-player zero-sum game).

The first such point is the incompressibility of the "reduced" dynamics (4.6). By definition, we have $u_i(x) = \sum_{j\in\mathcal{N}_i} u_{ij}(x_i, x_j)$, so we also have

$$v_{i\alpha_i}(x) = \sum_{j\in\mathcal{N}_i} u_{ij}(\alpha_i, x_j). \tag{C.10}$$

Since $u_{ij}(\alpha_i, x_j)$ does not depend on x_i, we readily obtain $\partial_{\alpha_i} v_{i\alpha_i}(x) = 0$ and incompressibility follows as before.

Let the network game in question be isomorphic to a network of constant-sum games after the following positive-affine transformation of utilities, $u_i(x) \leftarrow a_i u_i(x) + b_i$ where $a_i > 0$. The second point of interest is the use of the coupling $G(y) = \sum_{i\in\mathcal{N}} a_i[h_i^*(y_i) - \langle y_i, x_i^*\rangle]$ as a constant of motion for (FoReL). Indeed, adapting the derivation of (C.1), we now get

$$\begin{aligned}\frac{dG}{dt} &= \langle \dot{y}, \nabla G(y)\rangle = \sum_{i\in\mathcal{N}} \langle v_i(x), a_i(\nabla h_i^*(y_i) - x_i^*)\rangle = \sum_{i\in\mathcal{N}} \langle a_i v_i(x), x_i - x_i^*\rangle \\ &= \sum_{i\in\mathcal{N}} \sum_{j\in\mathcal{N}_i} \langle a_i v_{ij}(x), x_i - x_i^*\rangle\end{aligned}$$

$$= \sum_{\{i,j\}\in\mathcal{E}} [a_i u_{ij}(x_i,x_j) + b_i - a_i u_{ij}(x_i^*,x_j) - b_i + a_j u_{ji}(x_i,x_j) + b_j - a_j u_{ji}(x_i,x_j^*) - b_j]$$
$$= 0, \tag{C.11}$$

where the third line follows by regrouping the summands in the second line by edge, and the last line follows as in the case of (C.1). This implies that $G(y(t))$ remains constant along any solution of (FoReL), so the rest of the proof follows as in the case of Theorem 4.2. □

Appendix D. Auxiliary results

In this appendix, we provide two auxiliary results that are used in the proof of Theorem 4.2. The first one shows that if the score difference between two strategies grows large, the strategy with the lower score becomes extinct:

Lemma D.1. *Let $\mathcal{A}$ be a finite set and let h be a regularizer on $\mathcal{X} \equiv \Delta(\mathcal{A})$. If the sequence $y_n \in \mathbb{R}^{\mathcal{A}}$ is such that $y_{\beta,n} - y_{\alpha,n} \to \infty$ for some $\alpha, \beta \in \mathcal{A}$, then* $\lim_{n\to\infty} Q_\alpha(y_n) = 0$.

Proof. Set $x_n = Q(y_n)$ and, by descending to a subsequence if necessary, assume there exists some $\varepsilon > 0$ such that $x_{\alpha,n} \geq \varepsilon > 0$ for all n. Then, by the defining relation $Q(y) = \arg\max\{\langle y,x\rangle - h(x)\}$ of Q, we have:

$$\langle y_n, x_n\rangle - h(x_n) \geq \langle y_n, x'\rangle - h(x') \tag{D.1}$$

for all $x' \in \Delta$. Therefore, taking $x'_n = x_n + \varepsilon(e_\beta - e_\alpha)$, we readily obtain

$$\varepsilon(y_{\alpha,n} - y_{\beta,n}) \geq h(x_n) - h(x'_n) \geq \min h - \max h \tag{D.2}$$

which contradicts our original assumption that $y_{\alpha,n} - y_{\beta,n} \to -\infty$. With Δ compact, the above shows that $x^*_\alpha = 0$ for any limit point x^* of x_n, i.e. $Q_\alpha(y_n) \to 0$. □

A key step of the proof of Theorem 4.2 consists of showing that the level sets of the Fenchel coupling $G(p,y)$ become bounded under the coordinate reduction transformation $y \mapsto \Pi(y) = z$, so every solution orbit $z(t)$ of (4.6) also remains bounded. We encode this in the following lemma:

Lemma D.2. *Let $\mathcal{A}$ be a finite set, let h be a regularizer on $\mathcal{X} \equiv \Delta(\mathcal{A})$, and fix some interior $p \in \mathcal{X}$. If the sequence $y_n \in \mathbb{R}^{\mathcal{A}}$ is such that $\sup_n \lvert h^*(y_n) - \langle y_n, p\rangle\rvert < \infty$, the differences $y_{\beta,n} - y_{\alpha,n}$ also remain bounded for all $\alpha, \beta \in \mathcal{A}$.*

Proof. We argue by contradiction. Indeed, assume that the sequence $G_n \equiv h^*(y_n) - \langle y_n, p\rangle$ is bounded but $\limsup_{n\to\infty} \lvert y_{\alpha,n} - y_{\beta,n}\rvert = \infty$ for some $\alpha, \beta \in \mathcal{A}$. Letting $y_n^+ = \max_\alpha y_{\alpha,n}$ and $y_n^- = \min_{\alpha\in\mathcal{A}} y_{\alpha,n}$, this implies that $\limsup_{n\to\infty}(y_n^+ - y_n^-) = \infty$. Hence, by descending to a subsequence if necessary, there exist $\alpha^+, \alpha^- \in \mathcal{A}$ such that *a)* $y_n^\pm = y_{\alpha^\pm,n}$ for all n; and *b)* $y_{\alpha^+,n} - y_{\alpha^-,n} \to \infty$ as $n \to \infty$.

By construction, we have $y_{\alpha^-,n} = y_n^- \leq y_{\alpha,n} \leq y_n^+ = y_{\alpha^+,n}$ for all $\alpha \in \mathcal{A}$. Thus, by descending to a further subsequence if necessary, we may assume that the index set $\mathcal{A}$ can be partitioned into two nonempty sets $\mathcal{A}^+$ and $\mathcal{A}^-$ such that

(1) $y_n^+ - y_{\alpha,n}$ is bounded for all $\alpha \in \mathcal{A}^+$.
(2) $y_n^+ - y_{\alpha,n} \to \infty$ for all $\alpha \in \mathcal{A}^-$.

In more detail, consider the quantity

$$\delta_\alpha = \liminf_{n\to\infty}(y_n^+ - y_{\alpha,n}), \tag{D.3}$$

and construct the required partition $\{\mathcal{A}^+, \mathcal{A}^-\}$ according to the following procedure:

```
Set $\mathcal{A}^+ \leftarrow \{\alpha^+\}$, $\mathcal{A}^- = \mathcal{A} \setminus \mathcal{A}^+$
while $\delta_\alpha < \infty$ for some $\alpha \in \mathcal{A}^-$ do
    pick $\alpha^*$ such that $\delta_{\alpha^*} < \infty$;
    set $\mathcal{A}^+ \leftarrow \mathcal{A}^+ \cup \{\alpha^*\}$, $\mathcal{A}^- \leftarrow \mathcal{A}^- \setminus \{\alpha^*\}$;
    descend to a subsequence of $y_n$ that realizes $\delta_{\alpha^*}$;
    redefine $\delta_\alpha$ for all $\alpha \in \mathcal{A}$ based on chosen subsequence;
end while
return $\mathcal{A}^+, \mathcal{A}^-$
```

Thus, if we let $x_n = Q(y_n)$ we readily obtain:

$$\begin{aligned}\langle y_n, p - x_n\rangle &= \sum_{\alpha\in\mathcal{A}} y_{\alpha,n}(p_\alpha - x_{\alpha,n}) = \sum_{\alpha\in\mathcal{A}} (y_{\alpha,n} - y_n^+)(p_\alpha - x_{\alpha,n}) \\ &= \sum_{\alpha\in\mathcal{A}^+} (y_{\alpha,n} - y_n^+)(p_\alpha - x_{\alpha,n}) + \sum_{\alpha\in\mathcal{A}^-} (y_{\alpha,n} - y_n^+)(p_\alpha - x_{\alpha,n}), \end{aligned} \tag{D.4}$$

where we used the fact that $\sum_{\alpha\in\mathcal{A}} p_\alpha = \sum_{\alpha\in\mathcal{A}} x_{\alpha,n} = 1$ in the first line. The first sum above is bounded by assumption. As for the second one, the fact that $y_{\alpha^+,n} - y_{\alpha,n} = y_n^+ - y_{\alpha,n} \to \infty$ implies that $x_{\alpha,n} \to 0$ for all $\alpha \in \mathcal{A}^-$ (by Lemma D.1 above). We thus get $\liminf_n (p_\alpha - x_{\alpha,n}) > 0$ (recall that $p \in \Delta^\circ$), and hence, $\sum_{\alpha\in\mathcal{A}^-} (y_{\alpha,n} - y_n^+)(p_\alpha - x_{\alpha,n}) \to -\infty$.

From the above, we conclude that $\langle y_n, p - x_n\rangle \to -\infty$ as $n \to \infty$. However, by construction, we also have

$$G_n = h^*(y_n) - \langle y_n, x^*\rangle = \langle y_n, x_n\rangle - h(x_n) - \langle y_n, x^*\rangle = \langle y_n, p - x_n\rangle - h(x_n). \tag{D.5}$$

Since h is finite on x, it follows that $G_n \to -\infty$, contradicting our assumption that G_n is bounded. Retracing our steps, this implies that $\sup_n |y_{\alpha,n} - y_{\beta,n}| < \infty$, as claimed. □

The final result we state here is a technical result regarding the asymptotic behavior of the derivative of functions with a finite limit at infinity:

Lemma D.3. *Suppose that $L\colon [0,\infty) \to \mathbb{R}$ is differentiable with Lipschitz continuous derivative. If $\lim_{t\to\infty} L(t)$ exists and is finite, we have $\lim_{t\to\infty} L'(t) = 0$.*

Proof. Assume ad absurdum that $\lim_{t\to\infty} L'(t) \neq 0$. Then, without loss of generality, we may assume there exists some $\varepsilon > 0$ and an increasing sequence $t_n \uparrow \infty$ such that $L'(t_n) \geq \varepsilon$ for all $n \in \mathbb{N}$. Thus, if M denotes the Lipschitz constant of L' and $t \in [t_n, t_n + \varepsilon/(2M)]$, we readily obtain

$$|L'(t) - L'(t_n)| \leq M|t - t_n| \leq M \cdot \frac{\varepsilon}{2M} = \frac{\varepsilon}{2} \tag{D.6}$$

by the Lipschitz continuity of L'. Since $L'(t_n) \geq \varepsilon$ by assumption, we conclude that $L'(t) \geq \varepsilon/2$ for all $t \in [t_n, t_n + \varepsilon/(2M)]$. Hence, by integrating, we get $L(t_n + \varepsilon/(2M)) \geq L(t_n) + \varepsilon/(2M) \cdot (\varepsilon/2) = L(t_n) + \varepsilon^2/(4M)$ for all $n \in \mathbb{N}$. Taking $n \to \infty$ and recalling that $L_\infty \equiv \lim_{t\to\infty} L(t)$ exists and is finite, we get $L_\infty = L_\infty + \varepsilon^2/(4M) > L_\infty$, a contradiction. □

References

[1] I. Adler, C. Daskalakis, and C. H. Papadimitriou, *A note on strictly competitive games.*, in WINE, Springer, 2009, pp. 471–474.

[2] S. Arora, E. Hazan, and S. Kale, *The multiplicative weights update method: a meta-algorithm and applications.*, Theory of Computing, 8 (2012), pp. 121–164.

[3] H. Attouch, J. Bolte, P. Redont, and M. Teboulle, *Singular Riemannian barrier methods and gradient-projection dynamical systems for constrained optimization*, Optimization, 53 (2004), pp. 435–454.

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

[5] A. Ben-Tal and A. Nemirovski, *Lectures on modern convex optimization: analysis, algorithms, and engineering applications*, SIAM, 2001.

[6] Y. Cai, O. Candogan, C. Daskalakis, and C. Papadimitriou, *Zero-sum polymatrix games: A generalization of minmax*, Mathematics of Operations Research, 41 (2016), pp. 648–655.

[7] Y. Cai and C. Daskalakis, *On minmax theorems for multiplayer games*, in ACM-SIAM Symposium on Discrete Algorithms, SODA, 2011, pp. 217–234.

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

[9] C. Daskalakis, A. Deckelbaum, and A. Kim, *Near-optimal no-regret algorithms for zero-sum games*, in Proceedings of the Twenty-second Annual ACM-SIAM Symposium on Discrete Algorithms, SODA '11, Philadelphia, PA, USA, 2011, Society for Industrial and Applied Mathematics, pp. 235–254.

[10] C. Daskalakis and C. H. Papadimitriou, *On a network generalization of the minmax theorem*, in ICALP 2009: Proceedings of the 2009 International Colloquium on Automata, Languages, and Programming, 2009.

[11] D. J. Foster, T. Lykouris, K. Sridharan, and E. Tardos, *Learning in games: Robustness of fast convergence*, in Advances in Neural Information Processing Systems, 2016, pp. 4727–4735.

[12] D. Friedman, *Evolutionary games in economics*, Econometrica, 59 (1991), pp. 637–666.

[13] D. Gale, H. Kuhn, and A. W. Tucker, *(Linear Programming and the Theory of Games - Chapter XII) in Koopmans, Activity Analysis of Production and Allocation*, Wiley, 1951.

[14] I. Goodfellow, J. Pouget-Abadie, M. Mirza, B. Xu, D. Warde-Farley, S. Ozair, A. Courville, and Y. Bengio, *Generative adversarial nets*, in Advances in neural information processing systems, 2014, pp. 2672–2680.

[15] E. Hazan et al., *Introduction to online convex optimization*, Foundations and Trends® in Optimization, 2 (2016), pp. 157–325.

[16] J. Hofbauer and K. Sigmund, *Evolutionary Games and Population Dynamics*, Cambridge University Press, Cambridge, 1998.

[17] J. Hofbauer, S. Sorin, and Y. Viossat, *Time average replicator and best reply dynamics*, Mathematics of Operations Research, 34 (2009), pp. 263–269.

[18] N. Immorlica, A. T. Kalai, B. Lucier, A. Moitra, A. Postlewaite, and M. Tennenholtz, *Dueling algorithms*, in Proceedings of the forty-third annual ACM symposium on Theory of computing, ACM, 2011, pp. 215–224.

[19] K. C. Kiwiel, *Free-steering relaxation methods for problems with strictly convex costs and linear constraints*, Mathematics of Operations Research, 22 (1997), pp. 326–349.

[20] J. Kwon and P. Mertikopoulos, *A continuous-time approach to online optimization*, Journal of Dynamics and Games, 4 (2017), pp. 125–148.

[21] R. Lahkar and W. H. Sandholm, *The projection dynamic and the geometry of population games*, Games and Economic Behavior, 64 (2008), pp. 565–590.

[22] D. L. McFadden, *Conditional logit analysis of qualitative choice behavior*, in Frontiers in Econometrics, P. Zarembka, ed., Academic Press, New York, NY, 1974, pp. 105–142.

[23] P. Mertikopoulos and A. L. Moustakas, *The emergence of rational behavior in the presence of stochastic perturbations*, The Annals of Applied Probability, 20 (2010), pp. 1359–1388.

[24] P. Mertikopoulos and W. H. Sandholm, *Learning in games via reinforcement and regularization*, Mathematics of Operations Research, 41 (2016), pp. 1297–1324.

[25] A. Nagurney and D. Zhang, *Projected dynamical systems in the formulation, stability analysis, and computation of fixed demand traffic network equilibria*, Transportation Science, 31 (1997), pp. 147–158.

[26] G. Palaiopanos, I. Panageas, and G. Piliouras, *Multiplicative Weights Update with Constant Step-Size in Congestion Games: Convergence, Limit Cycles and Chaos*, ArXiv e-prints, (2017).

[27] C. Papadimitriou and G. Piliouras, *From nash equilibria to chain recurrent sets: Solution concepts and topology*, in ITCS, 2016.

[28] G. Piliouras, C. Nieto-Granda, H. I. Christensen, and J. S. Shamma, *Persistent patterns: Multi-agent learning beyond equilibrium and utility*, in AAMAS, 2014, pp. 181–188.

[29] G. Piliouras and J. S. Shamma, *Optimization despite chaos: Convex relaxations to complex limit sets via poincaré recurrence*, in Proceedings of the twenty-fifth annual ACM-SIAM symposium on Discrete algorithms, SIAM, 2014, pp. 861–873.

[30] H. Poincaré, *Sur le problème des trois corps et les équations de la dynamique*, Acta Math, 13 (1890), pp. 1–270.

[31] S. Rakhlin and K. Sridharan, *Optimization, learning, and games with predictable sequences*, in Advances in Neural Information Processing Systems, 2013, pp. 3066–3074.

[32] R. T. Rockafellar and R. J. B. Wets, *Variational Analysis*, vol. 317 of A Series of Comprehensive Studies in Mathematics, Springer-Verlag, Berlin, 1998.

[33] T. Roughgarden, *Intrinsic robustness of the price of anarchy*, in Proc. of STOC, 2009, pp. 513–522.

[34] A. Rustichini, *Optimal properties of stimulus-response learning models*, Games and Economic Behavior, 29 (1999), pp. 244–273.

[35] W. H. Sandholm, *Population Games and Evolutionary Dynamics*, MIT Press, Cambridge, MA, 2010.

[36] W. H. Sandholm, E. Dokumaci, and R. Lahkar, *The projection dynamic and the replicator dynamic*, Games and Economic Behavior, 64 (2008), pp. 666–683.

[37] Y. Sato, E. Akiyama, and J. D. Farmer, *Chaos in learning a simple two-person game*, Proceedings of the National Academy of Sciences, 99 (2002), pp. 4748–4751.

[38] P. Schuster and K. Sigmund, *Replicator dynamics*, Journal of Theoretical Biology, 100 (1983), pp. 533–538.

[39] D. Schuurmans and M. A. Zinkevich, *Deep learning games*, in Advances in Neural Information Processing Systems, 2016, pp. 1678–1686.

[40] S. Shalev-Shwartz, *Online learning and online convex optimization*, Foundations and Trends in Machine Learning, 4 (2011), pp. 107–194.

[41] S. Shalev-Shwartz and Y. Singer, *Convex repeated games and Fenchel duality*, in Advances in Neural Information Processing Systems 19, MIT Press, 2007, pp. 1265–1272.

[42] S. Sorin, *Exponential weight algorithm in continuous time*, Mathematical Programming, 116 (2009), pp. 513–528.

[43] V. Syrgkanis, A. Agarwal, H. Luo, and R. E. Schapire, *Fast convergence of regularized learning in games*, in Proceedings of the 28th International Conference on Neural Information Processing Systems, NIPS'15, Cambridge, MA, USA, 2015, MIT Press, pp. 2989–2997.

[44] P. D. Taylor, *Evolutionarily stable strategies with two types of player*, Journal of Applied Probability, 16 (1979), pp. 76–83.

[45] P. D. Taylor and L. B. Jonker, *Evolutionary stable strategies and game dynamics*, Mathematical Biosciences, 40 (1978), pp. 145–156.

[46] Y. Viossat and A. Zapechelnyuk, *No-regret dynamics and fictitious play*, Journal of Economic Theory, 148 (2013), pp. 825–842.

[47] J. von Neumann, *Zur Theorie der Gesellschaftsspiele*, Mathematische Annalen, 100 (1928), pp. 295–320. Translated by S. Bargmann as "On the Theory of Games of Strategy" in A. Tucker and R. D. Luce, editors, *Contributions to the Theory of Games IV*, volume 40 of *Annals of Mathematics Studies*, pages 13-42, 1957, Princeton University Press, Princeton.

[48] J. W. Weibull, *Evolutionary Game Theory*, MIT Press; Cambridge, MA: Cambridge University Press., 1995.