Evolutionary Game Theory Squared: Evolving Agents in Endogenously Evolving Zero-Sum Games

Extended Abstract [†]

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

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

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

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

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

ABSTRACT

The predominant paradigm in evolutionary game theory and more generally online learning in games is based on a clear distinction between a population of *dynamic agents* that interact given a *fixed, static game*. In this paper, we introduce and analyze a large class of competitive settings where both the agents and the games they play evolve strategically over time. We focus on zero-sum polymatrix games and replicator dynamics, the continuous-time analogue of multiplicative weights. In our setting, populations of agents compete against each other in a zero-sum environment that itself evolves adversarially to the current population mixture. We show that this system exhibits a number of regularities. First, the system exhibits *conservation laws* of an information-theoretic flavor. Secondly, the system is *Poincaré recurrent*, with effectively all possible initializations of agents and games lying on recurrent orbits that come arbitrarily close to their initial conditions infinitely often. Thirdly, the *time-average agent behavior and utility converge* to the Nash equilibrium values of the *time-average* game. Finally, we provide a polynomial time algorithm to efficiently predict this time-average behavior for any such co-evolving network game.

KEYWORDS

Evolutionary Game Theory, Multiagent Learning, Dynamical Systems

1 INTRODUCTION

The problem of analyzing evolutionary learning dynamics in games is of fundamental importance in several fields such as evolutionary game theory [8], online learning in games [2, 7], and multi-agent systems [9]. The dominant paradigm in each area is that of evolutionary agents adapting to each others behavior. In other words, the dynamism of the environment of each agent is driven by the other agents, whereas the game remains static. This separation between *evolving agents* and a *static game* is so standard that it typically goes unnoticed, however, this fundamental restriction does not allow us to capture many applications of interest. In artificial intelligence [3, 4, 6, 11, 15, 17] as well as biology, sociology, and economics [1, 5, 12–14, 16], the rules of interaction can themselves adapt to the collective history of the agent behavior.

In this paper, we provide a general framework for analyzing learning agents in time-evolving zero-sum games as well as rescaled network generalizations thereof. To begin, we develop a novel *reduction* that takes as input time-evolving games and reduces them to a game-theoretic graph that generalizes both graphical zero-sum games and evolutionary zero-sum games.

From an algorithmic learning perspective, we focus on the most studied evolutionary learning dynamic: replicator, the continuous-time analogue of the multiplicative weights update. In polymatrix games (defined by an undirected graph $G = (V, E)$), *replicator dynamics* [8] are given by

$$\dot{x}_{i\alpha} = x_{i\alpha}(u_{i\alpha}(x) - u_i(x)), \;\; \forall \alpha \in \mathcal{A}_i. \tag{1}$$

where $\mathcal{A}_i = \{1, \ldots, n_i\}$ is the set of actions for each player $i \in V$, $u_{i\alpha}(x)$ is the utility of player i under the strategy profile $x = (\alpha, x_{-i}) \in \mathcal{X}$ for $\alpha \in \mathcal{A}_i$ and $u_i(x)$ is the average utility player i obtains under strategy profile x. We prove the system is *Poincaré recurrent*, with effectively all initializations of agents and games lying on recurrent orbits that come arbitrarily close to their initial conditions infinitely often. As a crucial component of this result, we demonstrate the dynamics obey information-theoretic *conservation laws* that couple the behavior of all agents and games. Moreover, while the system never equilibrates, the conservation laws allow us to prove that the *time-average behavior and utility of the agents converge* to the time-average Nash of their evolving games with bounded regret. Finally, we provide a *polynomial time algorithm* that predicts these time-average quantities.

2 REDUCTION FROM TIME-VARYING GAMES TO POLYMATRIX GAMES

Numerous applications from artificial intelligence (AI) and machine learning (ML) to biology cast competition between populations (e.g., neural networks/algorithms or species/agents) and the environment (e.g., hyperparameters/network configurations or resources) as a time-evolving dynamical system. The basic abstraction takes the form of a population y of *species* which evolve dynamically in time as a function of itself and some *environment* parameters w whose evolution, in turn, depends on y.

[†] Published in AAAI 2021. The full paper can be found at [10].
[*] Joint last authors.

We generalize this concept to a class of time-evolving games defined by a set of populations $y = (y_1, \ldots, y_{n_y})$ and environments $w = (w_1, \ldots, w_{n_w})$, where $y_\ell \in \Delta^{n-1}$ for each $\ell \in \{1, \ldots, n_y\}$ and $w_k \in \Delta^{n-1}$ for each $k \in \{1, \ldots, n_w\}$. Here n is the number of actions for each player. Environments coevolve with only populations and not other environments, while any population coevolves only with environments and itself. Let $\mathcal{N}_k^w$ be the set of populations which coevolve with w_k and $\mathcal{N}_\ell^y$ be the set of environments which coevolve with y_ℓ. Moreover the payoff matrix has elements $A_{\alpha\beta}^{ij}$ for $(\alpha, \beta) \in \mathcal{A}_i \times \mathcal{A}_j$, which represents the reward player i obtains for selecting action α given that player j chooses action β. The time-evolving dynamics for each environment k and population ℓ are given component-wise by

$$\dot{w}_{k,i} = w_{k,i} \sum_{\ell \in \mathcal{N}_k^w} \sum_j w_{k,j} \left((A^{k,\ell} y_\ell)_i - (A^{k,\ell} y_\ell)_j \right), \tag{2}$$

$$\dot{y}_{\ell,i} = y_{\ell,i} \left((P_\ell(w) y_\ell)_i - y_\ell^\top P_\ell(w) y_\ell \right), \tag{3}$$

where $P_\ell(w) = P_\ell + \sum_{k \in \mathcal{N}_\ell^y} W^{\ell,k}$ where $P_\ell \in \mathbb{R}^{n \times n}$ represents the fixed payoff matrix of the game and $W^{\ell,k} \in \mathbb{R}^{n \times n}$ is defined such that the (i, j)–th entry is $(A^{\ell,k} w_k)_i - (A^{\ell,k} w_k)_j$.

Despite the complex nature of this dynamical system, we show that it is equivalent to replicator dynamics in a polymatrix game.

THEOREM 2.1. *Any time-evolving system defined by the dynamics in (2-3) is equivalent to replicator dynamics in a polymatrix game.*

The expressive power we gain from this reduction permits us to efficiently describe and characterize coevolutionary processes of higher complexity than past work since we can return to the familiar territory of analyzing dynamic agents in static games.

3 THEORETICAL ANALYSIS

We show that the replicator dynamics are Poincaré recurrent in N-player rescaled zero-sum polymatrix games with interior Nash equilibria. In particular, for almost all initial conditions $x(0) \in \mathcal{X}$, the replicator dynamics will return arbitrarily close to $x(0)$ an infinite number of times.

THEOREM 3.1. *The replicator dynamics given in* (1) *are Poincaré recurrent in any N-player rescaled zero-sum polymatrix game that has an interior Nash equilibrium.*

To prove Poincaré recurrence for (1), we first show recurrence for the flow corresponding to a system of ODEs which is diffeomorphic to the flow of the replicator system.

Given $x \in \mathcal{X}$, consider the transformed variable $z \in \mathbb{R}^{n_1 + \cdots + n_N - N}$ defined by

$$z_i = \left(\ln \frac{x_{i2}}{x_{i1}}, \ldots, \ln \frac{x_{in_i}}{x_{i1}} \right), \ \forall i \in V. \tag{4}$$

We then show Poincaré recurrence for this system by showing that the flow of $\dot{z} = F(z)$ is volume preserving (Lemma 3.1) and has bounded orbits (Lemma 3.2) from any interior initial condition.

LEMMA 3.1. *For any N-player rescaled zero-sum polymatrix game,* $\operatorname{tr}(DF(z)) = \sum_{i \in V} \sum_{\alpha \in \mathcal{A}_i} \frac{d}{dz_{i\alpha}} F_{i\alpha}(z) = 0$.

LEMMA 3.2. *Consider an N-player rescaled zero-sum polymatrix game such that for positive coefficients $\{\eta_i\}_{i \in V}$, $\sum_{i \in V} \eta_i u_i(x) = 0$ for $x \in \mathcal{X}$. If the game admits an interior Nash Equilibrium x^*, then $\Phi(t) = \sum_{i \in V} \sum_{\alpha \in \mathcal{A}_i} \eta_i x_{i\alpha}^* \ln x_{i\alpha}$ is time-invariant, meaning $\Phi(t) = \Phi(0)$ for $t \geq 0$. Hence, orbits from any interior initial condition $x(0)$ remain on the interior of the simplex.*

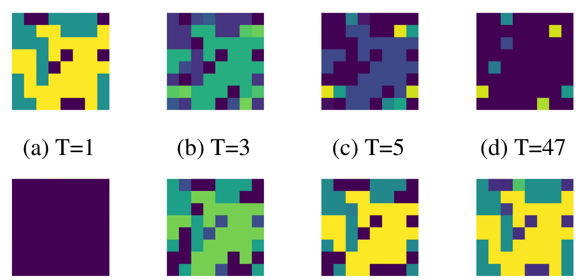
Finally, we also present results regarding time average behaviour and equilibrium computation for replicator dynamics in our setting.

THEOREM 3.2. *Consider an N-player rescaled zero-sum polymatrix game that admits a unique interior Nash equilibrium x^*. The trajectory $x(t)$ produced by replicator dynamics given in* (1) *is such that **i)** $\lim_{t \to \infty} \frac{1}{t} \int_0^t x(\tau) d\tau = x^*$ and **ii)** $\lim_{t \to \infty} \frac{1}{t} \int_0^t u_i(x(\tau)) d\tau = u_i(x^*)$.*

THEOREM 3.3. *Consider an N-player rescaled zero-sum polymatrix game such that for positive coefficients $\{\eta_i\}_{i \in V}$, $\sum_{i=1}^N \eta_i u_i(x) = 0$ for $x \in \mathcal{X}$. The optimal solution of the following linear program is a Nash equilibrium of the game:*

$$\min_{x \in \mathcal{X}} \{ \textstyle\sum_{i=1}^n \eta_i v_i | \, v_i \geq u_{i\alpha}(x), \ \forall \, i \in V, \ \forall \, \alpha \in \mathcal{A}_i \}$$

4 EXPERIMENTS



(a) T=1 (b) T=3 (c) T=5 (d) T=47

(i) T=68 (j) T=93 (k) T=117 (l) T=101701

Figure 1: Sequence of Pikachu images showing approximate recurrence in a 64-player zero-sum polymatrix game.

In the full version of the paper, we also present several experiments that verify some of our key results, and highlight other empirically observed properties outside the established theoretical results. For instance, to demonstrate the scalability of our model, we simulated a polymatrix zero sum Rock-Paper-Scissors game with 64 nodes. We use a relatively dense graph, where the initial condition of each player informs the RGB value of a corresponding pixel on an 8×8 grid. As can be seen in Figure 1, due to Poincaré recurrence, we see that the system returns close to the initial Pikachu image after over 100k time steps.

While the experiments are performed on highly controlled systems, the applications of our results are pertinent in various non-stationary environments. For instance, Google DeepMind trains populations of AI agents against each other and computes win probabilities in heads up competition. This results in a symmetric constant-sum game, which can be modeled with our framework. In addition, the progressive training of generative adversarial networks, where the Nash equilibrium changes over time, can also be captured by our model.

ACKNOWLEDGMENTS

Stratis Skoulakis gratefully acknowledges NRF 2018 Fellowship NRF-NRFF2018-07. Tanner Fiez acknowledges support from the DoD NDSEG Fellowship. Ryann Sim gratefully acknowledges support from the SUTD President's Graduate Fellowship (SUTD-PGF). Lillian Ratliff is supported by NSF CAREER Award number 1844729 and an Office of Naval Research Young Investigor Award. Georgios Piliouras gratefully acknowledges support from grant PIE-SGP-AI-2020-01, NRF2019-NRF-ANR095 ALIAS grant and NRF 2018 Fellowship NRF-NRFF2018-07.

REFERENCES

[1] Samuel Bowles, Jung-Kyoo Choi, and Astrid Hopfensitz. 2003. The co-evolution of individual behaviors and social institutions. *Journal of Theoretical Biology* 223, 2 (2003), 135–147.

[2] Nicolo Cesa-Bianchi and Gábor Lugosi. 2006. *Prediction, learning, and games*. Cambridge university press.

[3] Victor Costa, Nuno Lourenço, João Correia, and Penousal Machado. 2019. COEGAN: evaluating the coevolution effect in generative adversarial networks. In *Genetic and Evolutionary Computation Conference*. 374–382.

[4] Unai Garciarena, Roberto Santana, and Alexander Mendiburu. 2018. Evolved GANs for generating Pareto set approximations. In *Genetic and Evolutionary Computation Conference*. 434–441.

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

[6] Risto Miikkulainen, Jason Liang, Elliot Meyerson, Aditya Rawal, Daniel Fink, Olivier Francon, Bala Raju, Hormoz Shahrzad, Arshak Navruzyan, Nigel Duffy, et al. 2019. Evolving deep neural networks. In *Artificial Intelligence in the Age of Neural Networks and Brain Computing*. Elsevier, 293–312.

[7] N Nisan, T Roughgarden, E Tardos, and VV Vazirani. 2007. *Algorithmic Game Theory*. Cambridge university press.

[8] William H Sandholm. 2010. *Population games and evolutionary dynamics*. MIT press.

[9] Yoav Shoham and Kevin Leyton-Brown. 2008. *Multiagent systems: Algorithmic, game-theoretic, and logical foundations*. Cambridge University Press.

[10] Stratis Skoulakis, Tanner Fiez, Ryann Sim, Georgios Piliouras, and Lillian Ratliff. 2020. Evolutionary Game Theory Squared: Evolving Agents in Endogenously Evolving Zero-Sum Games. *arXiv e-prints* (2020), arXiv–2012.

[11] Kenneth O Stanley and Risto Miikkulainen. 2002. Evolving neural networks through augmenting topologies. *Evolutionary Computation* 10, 2 (2002), 99–127.

[12] Alexander J Stewart and Joshua B Plotkin. 2014. Collapse of cooperation in evolving games. *Proceedings of the National Academy of Sciences* 111, 49 (2014), 17558–17563.

[13] Andrew R Tilman, Joshua B Plotkin, and Erol Akçay. 2020. Evolutionary games with environmental feedbacks. *Nature communications* 11, 1 (2020), 1–11.

[14] Andrew R Tilman, James R Watson, and Simon Levin. 2017. Maintaining cooperation in social-ecological systems. *Journal of Theoretical Biology* 10, 2 (2017), 155–165.

[15] Chaoyue Wang, Chang Xu, Xin Yao, and Dacheng Tao. 2019. Evolutionary generative adversarial networks. *IEEE Transactions on Evolutionary Computation* 23, 6 (2019), 921–934.

[16] Joshua S. Weitz, Ceyhun Eksin, Keith Paarporn, Sam P. Brown, and William C. Ratcliff. 2016. An oscillating tragedy of the commons in replicator dynamics with game-environment feedback. *National Academy of Sciences* 113, 47 (2016).

[17] Yan Wu, Jeff Donahue, David Balduzzi, Karen Simonyan, and Timothy Lillicrap. 2019. LOGAN: Latent Optimisation for Generative Adversarial Networks. *arXiv preprint arXiv:1912.00953* (2019).