Learning and steering game dynamics towards desirable outcomes

Ilayda Canyakmaz[1] ILAYDA_CANYAKMAZ@SUTD.EDU.SG
Iosif Sakos[1] IOSIF_SAKOS@SUTD.EDU.SG
Wayne Lin[1] WAYNE_LIN@MYMAIL.SUTD.EDU.SG
Antonios Varvitsiotis[1] ANTONIOS@SUTD.EDU.SG
Georgios Piliouras[2] GPIL@GOOGLE.COM
[1]*Singapore University of Technology and Design,* [2]*Google DeepMind*

Editors: N. Ozay, L. Balzano, D. Panagou, A. Abate

Abstract

Game dynamics, which describe how agents' strategies evolve over time based on past interactions, can exhibit a variety of undesirable behaviours including convergence to suboptimal equilibria, cycling, and chaos. While central planners can employ incentives to mitigate such behaviors and steer game dynamics towards desirable outcomes, the effectiveness of such interventions critically relies on accurately predicting agents' responses to these incentives—a task made particularly challenging when the underlying dynamics are unknown and observations are limited. To address this challenge, this work introduces the Side Information Assisted Regression with Model Predictive Control (SIAR-MPC) framework. We extend the recently introduced SIAR method to incorporate the effect of control, enabling it to utilize side-information constraints inherent to game-theoretic applications to model agents' responses to incentives from scarce data. MPC then leverages this model to implement dynamic incentive adjustments. Our experiments demonstrate the effectiveness of SIAR-MPC in guiding systems towards socially optimal equilibria, stabilizing chaotic and cycling behaviors. Notably, it achieves these results in data-scarce settings of few learning samples, where well-known system identification methods paired with MPC show less effective results.

Keywords: game dynamics, system identification, model predictive control, sum of squares optimization, steering

1. Introduction

Game theory provides a mathematical framework for studying strategic interactions among self-interested decision-making agents, i.e., players. The Nash equilibrium (NE) is the central solution concept in game theory, describing a state where no player has an incentive to deviate (Nash, 1950). Over time, research has shifted from simply assuming that an NE exists and players will eventually play it, to understanding *how* equilibrium is reached (Smale, 1976; Papadimitriou and Piliouras, 2019). This shift has led to a focus on *learning* in games, exploring how strategies evolve over time based on past outcomes, adopting a dynamical systems perspective (Fudenberg and Levine, 1998; Sandholm, 2010). It has been shown that game dynamics do not necessarily converge to NE but instead can display a variety of undesirable behaviors, including cycling, chaos, Poincaré recurrence, or convergence to suboptimal equilibria (Akin and Losert, 1984; Sato et al., 2002; Hart and Mas-Colell, 2003; Mertikopoulos et al., 2018; Milionis et al., 2023). Motivated by these challenges, our primary objective in this work is to determine:

Can we steer game dynamics towards desirable outcomes?

To address this problem, we adopt the perspective of a central planner who seeks to influence player behaviour by designing incentives. Our goal is to achieve this with minimal effort, ensuring that the incentives are both cost-effective and efficient. More importantly, we operate in a setting with unknown game dynamics and limited observational data, reflecting real-world scenarios where information is often incomplete or uncertain. To tackle these challenges, we introduce a new computational framework called Side Information Assisted Regression with Model Predictive Control (SIAR-MPC), designed to steer game dynamics by integrating cutting-edge techniques for real-time system identification and control. In the system identification step, we predict agents' reactions to incentives, which is especially challenging for settings where observational data is limited, difficult to obtain, or costly. To address this problem, we extend the recently introduced SIAR method (Sakos et al., 2023), which was developed to identify agents' learning dynamics from a short burst of a system trajectory. To compensate for the absence of data, SIAR searches for polynomial regressors that approximate the dynamics, satisfying side-information constraints native to game theoretical applications. These constraints represent additional knowledge about agents' learning dynamics or assumptions about their behavior beyond the observed trajectory data. To adapt SIAR to our needs, we broaden its scope to incorporate the influence of control inputs, which represent the incentives designed by the central planner. This extension enables SIAR to model the controlled dynamics, resulting in SIAR with control (SIARc).

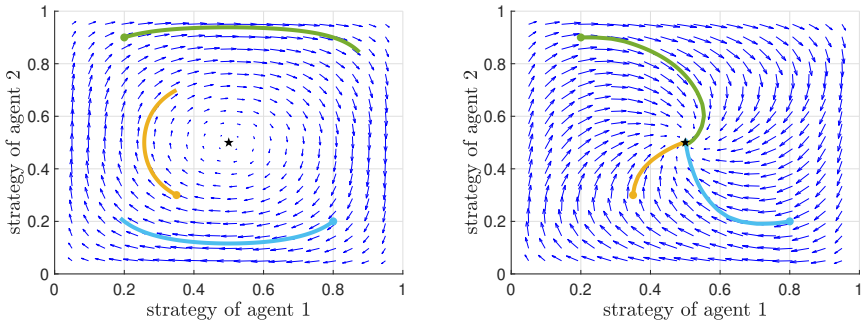

Figure 1: Replicator dynamics trajectories in the matching pennies game with and without control. Starting from three initial conditions, *(left)* without control the system cycles around the equilibrium; *(right)* with control, trajectories are guided towards the specified equilibrium (indicated by $\star$.)

Once agent responses to incentives are modeled with SIARc, we use MPC in the subsequent control step to develop a dynamic incentive scheme that steers the system towards desirable outcomes (see Fig. 1 for an example illustrating the impact of control). MPC is a control technique that leverages a mathematical model of the system to predict future behavior and calculate optimal control inputs that minimize a given objective function (Camacho and Bordons, 2007). A key advantage of MPC is its ability to handle input constraints, which is particularly relevant in our context since there are practical limits to the incentives that can be offered to agents. However, the effectiveness of MPC depends on having an accurate system model, highlighting the importance of the system identification step for its successful application. Our key contributions are:

- **Framework for steering game dynamics:** We introduce the SIAR-MPC framework for steering game dynamics towards desirable outcomes when the underlying agent behaviours are unknown and observational data are scarce.

- **Demonstration of performance across diverse game types:** We demonstrate the effectiveness of our framework across a diverse range of game types, from zero-sum games like Matching Pennies and Rock-Paper-Scissors to coordination games such as Stag Hunt. Our experimental results show that SIAR-MPC can successfully steer system dynamics towards socially optimal equilibria and stabilize chaotic and cycling learning dynamics.
- **Demonstration of performance in a data-scarce setting:** We show that SIAR-MPC consistently achieves convergence with low control costs in data-scarce settings characterized by limited learning samples, where other methods such as the unconstrained regression approach of Sparse Identification of Nonlinear Dynamics with control (SINDYc) as well as Physics-Informed Neural Network (PINN) coupled with MPC show less effective results.

2. Related Work

Control of Game Dynamics: Recently, incentive-based control has been applied extensively in multi-player environments (see, e.g., Riehl et al. (2018) and references therein). In addition, optimal control solutions have been given for specific evolutionary games and dynamics (Paarporn et al., 2018; Gong et al., 2022; Martins et al., 2023). However, in the aforementioned works the player behavior is known, and to the best of our knowledge a setup where the game dynamics are not a priori known to the controller has only been explored in the recent works of Zhang et al. (2024) and Huang et al. (2024). Zhang et al. (2024) study the problem of steering no-regret agents in normal- and extensive-form games, under both full and bandit feedback. However, in their setup the agents are adversarial and not bound to fixed dynamics, making their setup incompatible with system identification. On the other hand, Huang et al. (2024) study this problem in the richer environment of Markov games under the additional assumption that the players' dynamics belong to some known finite class. This assumption allows the controller to utilize simulators of the class of dynamics to optimize for controls that identify the exact update rule with high probability. In contrast, SIAR-MPC does not rely on this assumption and instead takes advantage of the approximation guarantees of polynomial regression to acquire an accurate representation of an unknown system.

System Identification: The field of system identification encompasses a variety of methods, with a significant emphasis on data-driven approaches such as deep learning (Brunton and Kutz, 2019; Cranmer et al., 2020), symbolic regression (Schmidt and Lipson, 2009; Udrescu and Tegmark, 2020; Udrescu et al., 2020), and statistical learning (Lu et al., 2019). A significant advancement in this field is the introduction of sparsity-promoting techniques, most notably the Sparse Identification of Nonlinear Dynamics (SINDy) method, which leverages the fact that the underlying dynamics can often be represented as a sparse combination of a set of candidate functions (Brunton et al., 2016; Kaiser et al., 2018). Despite their success, these data-driven methods rely on large amounts of data and often struggle to capture system dynamics accurately in data-scarce settings. Addressing this gap, Ahmadi and Khadir (2023) explored the use of sum-of-squares (SOS) optimization to identify dynamical systems from noisy observations of a few trajectories by incorporating contextual system information to compensate for data scarcity. This framework was used both in SINDy-SI (Machado and Jones, 2024), an expansion of SINDy that allows for polynomial nonnegativity constraints, and in the SIAR method (Sakos et al., 2023), which learns game dynamics with few training data by leveraging side-information constraints relevant to strategic agents. Physics Informed Neural Networks (PINNs) are a related method that integrate knowledge of physical laws into the training process to increase data efficiency (Chen et al., 2020; Stiasny et al., 2021; Canyakmaz et al., 2024).

3. Preliminaries

In this work, we model a multi-agent system as a time-evolving normal-form game of n players. Each player i is equipped with a finite set of strategies $\mathcal{A}_i$ of size m_i, and a time-varying reward function $u_i : \mathcal{A} \times \Omega \to \mathbb{R}$, where $\mathcal{A} \equiv \prod_{i=1}^n \mathcal{A}_i$ is the game's strategy space, of the form

$$u_i(a, \omega(t)) = u_i(a, 0) + \omega_{i,a}(t), \text{ for } t \in \mathbb{R}_+, a \in \mathcal{A}. \tag{1}$$

The value $\omega_{i,a}(t)$ denotes the control signal from the policy maker towards player i regarding the strategy profile $a := (a_1, \dots, a_n)$ at time t, where $a_i \in \mathcal{A}_i$. We refer to the ensemble $\omega_i(t) := \big(\omega_{i,a}(t)\big)_{a\in\mathcal{A}}$ as the control signal of i at time t, and to $\omega(t) := \big(\omega_1(t), \dots, \omega_n(t)\big)$ as the system's control signal at t. We restrict the control signals $\omega_i(t)$ in some semialgebraic sets Ω_i and refer to the product $\Omega \equiv \prod_{i=1}^n \Omega_i$ as the game's control space. The value $u_i(a) := u_i(a, 0)$ describes a time-independent base game, i.e., the reward of player i at strategy profile a in the absence of any control by the policy maker. The utilities $u_i(\cdot)$, $i=1, \dots, n$ will be considered common knowledge throughout the work. Altering the utilities is a natural method for influencing agent behavior as they encode all the information about the players' incentives in the game-theoretic model.

Moreover, each player i is allowed access to a set of mixed strategies $\mathcal{X}_i \equiv \Delta(\mathcal{A}_i)$, which is the $(m_i - 1)$-simplex that corresponds to the set of distributions over the pure strategies $\mathcal{A}_i$ of i, i.e., $x_i = (x_{i\alpha_i})_{\alpha_i\in\mathcal{A}_i}$. The players' reward function naturally extends to the space of mixed strategy profiles $\mathcal{X} \equiv \prod_{i=1}^n \mathcal{X}_i$ with $u_i\big(x, \omega\big) = \mathbb{E}[u_i(a, \omega)]$ for all $x \in \mathcal{X}$ and $\omega \in \Omega$, where the expectation is taken with respect to the distributions $x_1, \dots, x_n$ over the player's pure strategies. Finally, we assume that the evolution of the above game is dictated by some controlled learning dynamics of the form

$$\dot{x}(t) = f(x(t), \omega(t)) \quad \text{for } t \in \mathbb{R}_+, \quad x(0) \in \mathcal{X}, \tag{2}$$

where the update policies $f_i : \mathcal{X} \times \Omega_i \to \mathbb{R}^{m_i}$, $i = 1, \dots, n$ and the ensemble thereof, given by $f(x, \omega) := \big(f_1(x, \omega_1), \dots, f_n(x, \omega_n)\big)$, are considered unknown, and are going to be discovered in the identification step of the framework described below. Notice that the above assumption also implies that, at each time t, the control signal $\omega_i(t)$ of player i is observed by that player, while the strategy profile $x(t)$ is observed by all the players.

Throughout this work, given a strategy profile x, we adopt the common game-theoretic shorthand (x_i, x_{-i}) to distinguish between the strategy of player i and the strategies of the other players. Moreover, if x_i corresponds to a pure strategy a_i of i, we write (a_i, x_{-i}) to point to that fact.

4. The SIAR-MPC Framework

In this section, we describe the SIAR-MPC framework for the real-time identification and control of game dynamics. As outlined in the introduction, SIAR-MPC involves two steps. First, the system identification step aims to approximate the controlled dynamics in (2) using only a limited number of samples. Second, once the agents' reactions to payoffs are modeled, the control step employs MPC to steer the system towards a desirable outcome by optimizing specific objectives.

4.1. The System Identification Step

To model the controlled dynamics in (2) we extend the SIAR framework introduced in Sakos et al. (2023), which was in turn motivated by recent results in data-scarce system identification (Ahmadi and Khadir, 2023). SIAR relies on polynomial regression to approximate agents' learning dynamics

of the form $\dot{x}(t) = f\big(x(t)\big)$ based on a small number of observations $x(t_k), \dot{x}(t_k)$ (typically, $K = 5$ samples) taken along a short burst of a single system trajectory. To ensure the accuracy of the derived system model, SIAR searches for polynomial regressors that satisfy side-information constraints native to game-theoretic applications, which serve as a regularization mechanism.

For our control-oriented scenario, we extend the SIAR method to account for the influence of the control signal $\omega(t)$. We refer to this extended method as SIAR with control (SIARc). The aim of SIARc is to model the controlled dynamics in (2) for each agent i with a polynomial vector field $p_i(x, \omega)$. During the system identification phase, we assemble a dataset $x(t_k), \omega(t_k), \dot{x}(t_k)$, where $x(t_k)$ represents a snapshot of the system state, $\omega(t_k)$ is a randomly generated input reflecting various possible incentives given to players (cf. (1)), and $\dot{x}(t_k)$ is the velocity at time t_k, which is obtained either through direct measurement (if possible) or estimated from the state variables. The control input $\omega(t_k)$ is typically taken to be normally distributed with mean-zero and low variance: the noise is used to create variety between the data samples for better system identification, while the low variance maintains a low aggregated control cost during the system identification phase.

The process of training the SIARc model essentially involves solving an optimization problem to find a polynomial vector field that minimizes the mean square error relative to this dataset. However, straightforward regression often yields suboptimal models due to the limited available samples. To overcome this challenge, we search over regressors that satisfy additional side-information constraints, encapsulating essential game-theoretic application features and refining the search for applicable models. Formally, a generic SIARc instance is given by

$$\begin{aligned} \min_{p_1,\dots,p_n} \quad & \sum_{k=1}^{K}\sum_{i=1}^{n} \big\| p_i\big(x(t_k), \omega(t_k)\big) - \dot{x}_i(t_k) \big\|^2 \\ \text{s.t.} \quad & p_i \text{ are polynomial vector fields in } x \text{ and } \omega \\ & p_i \text{ satisfy side-information constraints.} \end{aligned} \tag{3}$$

In this work, we utilize two specific types of side-information constraints (though a broader array is available; see, e.g., Sakos et al. (2023)). The first side-information constraint ensures that the state space $\mathcal{X}$ of the system is robust forward invariant with respect to the controlled dynamics in (2). This implies that, for any initialization $x(0) \in \mathcal{X}$, we have that $x(t) \in \mathcal{X}$ for all subsequent times $t > 0$, and for any control signal $\omega(t) \in \Omega$. To search over regressors that satisfy robust forward invariance (RFI), we rely on a specific characterization of the property that dictates a set remains robustly forward invariant under system (2) only if $f(x, \omega)$ lies in the tangent cone of $\mathcal{X}$ at x for every control $\omega \in \Omega$. Using the characterization of the tangent cone at each simplex $\mathcal{X}_i$ (Nagumo, 1942; Blanchini, 1999), enforcing RFI is then reduced to verifying that, for all $x \in \mathcal{X}$ and $\omega \in \Omega$:

$$\sum_{a_i \in \mathcal{A}_i} p_{ia_i}(x, \omega) = 0 \quad \text{and} \quad p_{ia_i}(x, \omega) \geq 0, \quad \text{whenever } x_{ia_i} = 0. \tag{RFI}$$

The second side-information constraint is based on a fundamental assumption about agent behavior arising from their strategic nature. Agents, as strategic entities, are expected to act rationally, preferring actions that enhance their immediate benefits—a property known as positive correlation (PC) (Sandholm, 2010). Specifically, agents are inclined to choose actions that are likely to increase their expected utility, assuming other agents' behaviors remain unchanged, i.e., for all $x \in \mathcal{X}$ and $\omega \in \Omega$

$$\langle \nabla_{x_i} u_i(x, \omega),\, p_i(x, \omega) \rangle > 0, \text{ whenever } p(x, \omega) \neq 0. \tag{PC}$$

To enforce these side-information constraints computationally in our polynomial regression problem, we utilize SOS optimization (Parrilo, 2000; Prestel and Delzell, 2001; Lasserre, 2001; Parrilo, 2003; Lasserre, 2006). Both RFI and PC are represented as polynomial inequality or nonnegative constraints over the semialgebraic sets $\mathcal{X}$ and Ω (in PC's case, we need to relax the inequality (Sakos et al., 2023)). Using the SOS approach, instead of searching over polynomials $p(x)$ that are nonnegative over a semialgebraic set $\mathcal{S} \equiv \{x \mid g_j(x) \geq 0, h_\ell(x) = 0,\ j \in [m],\ \ell \in [r]\}$, we search over polynomials $p(x)$ that can be expressed as $p(x) = \sigma_0(x) + \sum_{j=1}^{m} \sigma_j(x) g_j(x) + \sum_{\ell=1}^{r} q_\ell(x) h_\ell(x)$ where q_ℓ are polynomials and σ_j are SOS polynomials. Such polynomials p are guaranteed to be nonnegative over $\mathcal{S}$, a condition that is also necessary under mild assumptions on the set $\mathcal{S}$ (Laurent, 2008, Theorem 3.20). Furthermore, for any given degree d, we can look for SOS certificates of degree d through semidefinite programming, creating a hierarchy of semidefinite problems. Our theoretical justification for this system identification step comes from showing that the theorems of Ahmadi and Khadir (2023) and Sakos et al. (2023), which guarantee the usefulness of polynomial dynamics in approximation, can be readily extended to our current setting with control inputs:

Theorem 1 (Informal) *Fix time horizon T, desired approximation accuracy $\epsilon > 0$, and desired side information accuracy $\delta > 0$. For any continuously differentiable dynamics f satisfying side information constraints, there exists polynomial dynamics p that δ-satisfies the same side-information constraints and is ϵ-close to f.*

Here, f and p are ϵ-close if for any initial point and sequence of controls $\boldsymbol{\omega} : [0, T] \to \Omega$, the distance between the trajectories and the distance between their velocities at any time $t \in [0, T]$ are both upper bounded by ϵ. δ-satisfiability of the side information constraints indicates approximately satisfying them to some tolerance δ and has an SOS certificate, meaning that we can perform regression over the space of polynomials that δ-satisfy the side-information constraints. A formal statement of this theorem, together with its proof, is presented in Appendix C of the supplementary material[1].

4.2. The Control Step

After estimating the controlled dynamics which describes how players' strategies (x) respond to the incentives (ω), our next goal is to steer the system towards desirable outcomes. To achieve this, we employ MPC, which formulates an optimization problem to identify the optimal sequence of control actions over a defined horizon, subject to constraints on control inputs. In our context, these control constraints represent practical limits on the incentives that can be offered to agents. MPC leverages a mathematical model of the system to predict future behavior over a specified prediction horizon $\mathcal{T}$. Each prediction is based on the current state measurement $x(t)$ and a sequence of future control signals $\omega_t := \{\omega_{0|t}, \ldots, \omega_{N-1|t}\} \subset \Omega$, which is calculated by solving a constrained optimization problem. Here, N is the number of control steps within the control horizon $\mathcal{T} := N \cdot \Delta t$ that determines the time period over which the control sequence is optimized. The optimal control sequence is obtained by solving the constrained optimization problem

$$\begin{aligned} \min_{\omega_t} \quad & \sum_{n=0}^{N} \|x_{n|t} - x^*\|^2 + \alpha \sum_{n=0}^{N-1} \|\omega_{n|t}\|^2 + \beta \sum_{n=1}^{N} \|\omega_{n|t} - \omega_{n-1|t}\|^2 \\ \text{s.t.} \quad & \omega_{n|t} \in \Omega,\ \ 0 \leq n \leq N \\ & x_{n+1|t} = x_{n|t} + \Delta t \cdot p(x_{n|t}, \omega_{n|t}),\ \ 1 \leq n \leq N \\ & x_{0|t} = x(t). \end{aligned} \tag{4}$$

1. This article contains supplementary material online at https://arxiv.org/abs/2404.01066

Here, $x_{n|t}$, $n = 1, \dots, N$ correspond to the forecasted trajectory, and x^* is a desirable target state. The first term of the objective function penalizes deviations of the predicted states $x_{n|t}$ from the target value x^*, the second term accounts for the control effort, and the final term penalizes large variations in control signal. The first control signal $\omega_{0|t}$ is then applied to the system and the optimization is repeated at time $t + \Delta t$ once the new state measurement $x(t + \Delta t)$ is obtained.

5. Experiments

In this section, we evaluate the SIAR-MPC framework across various types of normal-form games, each of independent interest, and compare its performance to pairing MPC with solutions obtained from SINDYc and PINN. The implementation of SINDYc follows the approach described in Kaiser et al. (2018), employing sequential least squares for solving the sparse regression problem as detailed in Brunton et al. (2016). For PINN, side-information constraints are integrated into the neural network training as terms in the loss function that penalize violations of the desired constraints (Raissi et al., 2019). Given the limited number of training samples—in most cases 5—we are restricted to a simple neural network architecture with two hidden layers of size 5. This limitation in the neural network's expressivity is counterbalanced by the simplicity of the ground-truth update policies f (which, in most cases, are polynomial). As activation functions we use the $\tanh$ function and the side-information constraints are enforced using $2,500$ collocation points. Further details on the construction of the loss function and the generation of collocation points can be found in Appendix B of the supplementary material.

5.1. Stag Hunt Game

The stag hunt game is a two-player, two-action coordination game that models a strategic interaction in which both players benefit by coordinating their actions towards a specific superior outcome (the hunt of a stag). However, if that outcome is not possible, each player prefers to take advantage of the lack of coordination and come out on top of their opponent by choosing the alternative (hunt a rabbit by themselves) rather than coordinating to an inferior outcome (hunt a rabbit together). While we perform experiments over various payoff matrices corresponding to stag hunt games in Section 5.3, for concreteness in this section we consider the example where the players' reward functions $u(\cdot) := u(\cdot, 0)$ in the absence of control (see (1)) are given by

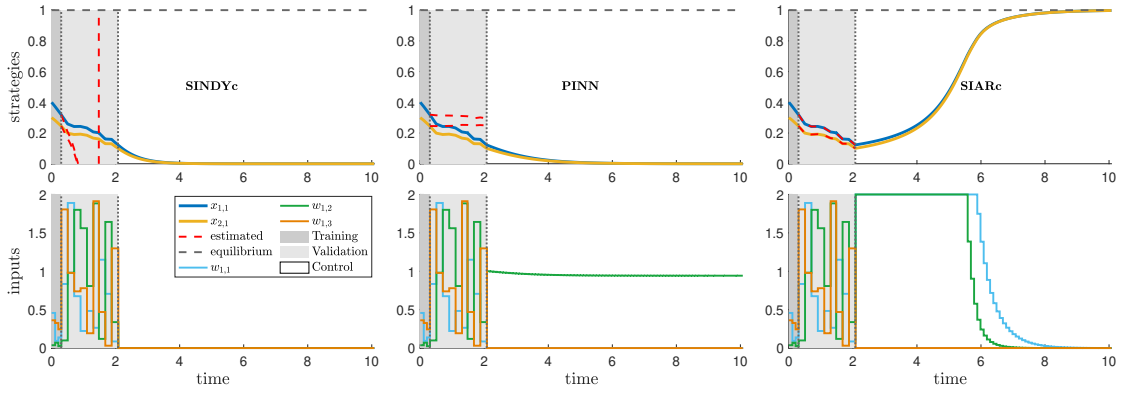

Figure 2: Performance comparison of SINDY-MPC *(left)*, PINN-MPC *(center)*, and SIAR-MPC *(right)* in steering the replicator dynamics for the stag hunt game.

$$u_1(a) = u_2(a) = \mathbf{A}_{a_1,a_2}, \text{ where } \mathbf{A} = \begin{pmatrix} 4 & 1 \\ 3 & 3 \end{pmatrix}. \tag{5}$$

For expositional purposes, we restrict the game's evolution to a subset of symmetric two-player two-action games given by the control signals $\omega_{1,a_1,a_2}(t) = \omega_{2,a_2,a_1}(t) \in [0,2]$ for all t. Furthermore, for tractability purposes we fix $\omega_{1,2,2}(t) := 0$. We assume that the agents' true behavior follow the replicator dynamics given, for player i and action a_i of i, by the polynomial update policies

$$f_{i,a_i}(x,\omega) = x_{i,a_i}(u_i(a_i, x_{-i}, \omega) - u_i(x,\omega)), \quad \text{for all} \quad x \in \mathcal{X} \quad \text{and} \quad \omega \in \Omega. \tag{6}$$

We search for f using the SIARc framework. Specifically, for each i and a_i, we are going to search for polynomial $p_{i,a_i} : \mathcal{X} \times \Omega \to \mathbb{R}$ such that $p_{i,a_i}\big(x(t),\omega(t)\big) \approx f_{i,a_i}\big(x(t),\omega(t)\big)$ for all $t \in \mathbb{R}_+$. As side-information constraints, we are going to impose the RFI and PC properties as given in the previous sections. Then, by substituting (5) to (PC) we have the following SIARc problem

$$\begin{aligned} \min_{p} \quad & \sum_{k=1}^{K} \big\| p\big(x(t_k),\omega(t_k)\big) - \dot{x}(t_k) \big\|^2 \\ \text{s.t.} \quad & p_{i,1}(x,\omega) + p_{i,2}(x,\omega) = 0, \ \forall i \\ & p_{i,1}\big((0,1), x_{-i}, \omega\big) \geq 0, \ \forall i \\ & p_{i,2}\big((1,0), x_{-i}, \omega\big) \geq 0, \ \forall i \\ & v_{i,1}(x,\omega) p_{i,1}(x,\omega) + v_{i,2}(x,\omega) p_{i,2} \geq 0, \ \forall i, \end{aligned} \tag{7}$$

where $x \in \mathcal{X}$, $\omega \in \Omega$, and $v_{i,a_i} : \mathcal{X} \times \Omega \to \mathbb{R}$ are given by

$$\begin{aligned} v_{1,1}(x,\omega) &= (4+\omega_{1,1})x_{2,1} + (1+\omega_{1,2})x_{2,2}, & v_{1,2}(x,\omega) &= (3+\omega_{2,1})x_{2,1} + 3x_{2,2}, \\ v_{2,1}(x,\omega) &= (4+\omega_{1,1})x_{1,1} + (1+\omega_{1,2})x_{1,2}, & v_{2,2}(x,\omega) &= (3+\omega_{2,1})x_{1,1} + 3x_{1,2}. \end{aligned} \tag{8}$$

Since the update policies f_{i,a_i} in (6) correspond to the replicator dynamics, the solution to the above optimization problem can be recovered by a 7-degree SOS relaxation (Sakos et al., 2023). Fig. 2 *(right)* shows the performance of the SIAR-MPC using this solution as a model for the MPC method. In the top panel of the figure, we have the trajectory $x(t)$ initialized at $x_0 = (0.4, 0.3)$ corresponding to the control signal $\omega(t)$ depicted in the bottom panel. The plot is divided into three sections corresponding to the system identification phase, an evaluation period, and a control phase. In the first section, we set the control signals to normally distributed noise with mean zero (bounded in Ω) and a sample of $K = 4$ datapoints from the resulting trajectory. At the end of this phase, we solve the optimization problem in (7) and acquire a model of the system's update policies. In the second section of the plot, we compare the ground-truth dynamics with the model's predicted trajectory (in dashed red lines); here, the control signal is chosen randomly. Finally, in the steering phase, we steer the ground truth using the output of the MPC as the control signal: the objective is to steer the system to the superior Nash equilibrium (NE) of the stag hunt game at $x^*_{1,1} = x^*_{2,1} = 1$, which is achieved by the SIAR-MPC framework at $t \approx 8$. In comparison, the SINDY-MPC and PINN-MPC solutions (Fig. 2 *(left)* and Fig. 2 *(center)*) fail to complete the steering objective.

5.2. Zero-Sum Games & Chaos

The stag hunt game provides an ideal setting for steering game dynamics to a NE due to the existence of a socially optimal NE with a positive-measure basin of attraction. In contrast, the class of zero-sum games lacks this property. Well-known game dynamics, such as replicator and log-barrier,

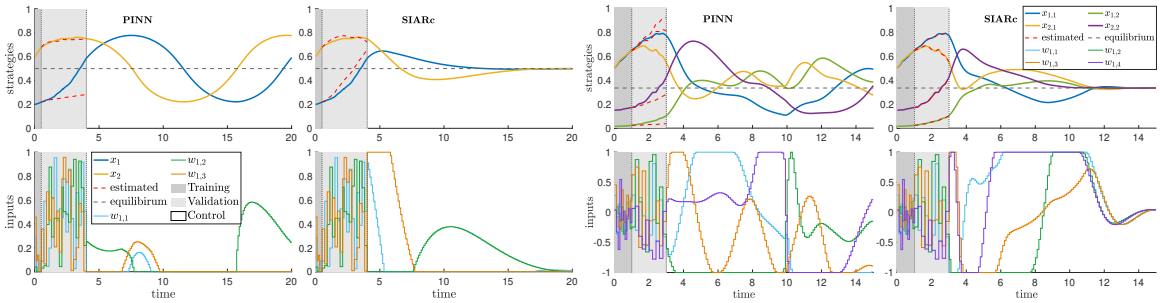

Figure 3: Performance comparison of PINN-MPC and SIAR-MPC steering the log-barrier dynamics for the matching pennies game *(left)* and the replicator dynamics for a 0.25-RPS game *(right)*.

are known to exhibit undesirable behaviors in zero-sum games, including cycling (see Fig. 1) and chaos. To address these challenges, we present two examples to demonstrate the performance of SIAR-MPC in steering log-barrier dynamics in the matching pennies game and chaotic replicator dynamics in an ϵ-perturbed rock-paper-scissors (ϵ-RPS) game.

Matching Pennies Game The matching pennies game is a two-player, two-action zero-sum game where one player benefits by the existence of coordination among the two, while the other player benefits by the lack thereof. Formally, a matching pennies game is encoded in the uncontrolled players' reward functions in (1) by

$$u_1(a) = -u_2(a) = \mathbf{A}_{a_1,a_2}, \text{ where } \mathbf{A} = \begin{pmatrix} 1 & -1 \\ -1 & 1 \end{pmatrix}. \tag{9}$$

For similar reasons as in the previous example, we are going to restrict the game's evolution to a subset of two-player two-action zero-sum games given by $\omega_{1,a}(t) = \omega_{2,a}(t) \in [0,1]$ for all t, and set $\omega_{1,2,2}(t) \coloneqq 0$. The log-barrier dynamics of the above time-varying game are, for player i and action a_i of i, given by the rational update policies

$$f_{i,a_i}(x,\omega) = x_{i,a_i}^2 \left(u_{i,a_i} - \frac{x_{i,1}^2 u_{i,1} + x_{i,2}^2 u_{i,2}}{x_{i,1}^2 + x_{i,2}^2} \right) \tag{10}$$

for all $t \in \mathbb{R}_+$ and $x \in \mathcal{X}$, where the shorthand $u_{i,x_j} \coloneqq u_i(x_j, x_{-i}, \omega)$ is used for compactness. Observe that, as is the case for the updated policies of the replicator dynamics in (6), f_i depends on $\omega(t)$ through u_i. In Fig. 3 *(left)*, we show that SIAR-MPC solution is able to steer a trajectory $x(t)$ of the above system initialized at $x(0) = (0.2, 0.6)$ to the unique mixed NE $x_{1,1}^* = x_{2,1}^* = 1/2$ of the matching pennies game with only $K = 6$ training samples.

ϵ-RPS Game In this example, we use the SIAR-MPC method to steer the replicator dynamics in an ϵ-RPS game, a two-player, three-action zero-sum game where the replicator dynamics exhibit chaotic behavior (Sato et al., 2002; Hu et al., 2019). In a nutshell, this means that any two initialization of the system—even the ones that are infinitesimally close to each other—may lead to completely different trajectories. In other words, the accurate estimation of the agents' update policies is futile due to the finite precision of any numerical method. An ϵ-RPS game is encoded by

$$u_1(a) = -u_2(a) = \mathbf{A}_{a_1,a_2}, \text{ where } \mathbf{A} = \begin{pmatrix} \epsilon & -1 & 1 \\ 1 & \epsilon & -1 \\ -1 & 1 & \epsilon \end{pmatrix}. \tag{11}$$

We consider time-evolving games in the subset of two-player three-action zero-sum games given by $\omega_{1,a}(t) = \omega_{2,a}(t) \in [-1, 1]$ for all t, and only four non-zero signals, namely, $\omega_{1,1,2}(t)$, $\omega_{1,1,3}(t)$, $\omega_{1,2,1}(t)$, and $\omega_{1,3,1}(t)$. This restriction on the controllers to be nonzero only on a subset of the outcomes is to demonstrate steerability even with incomplete controls. In Fig. 3 *(right)*, we show the successful steering of the chaotic replicator dynamics (6) in the 0.25-RPS game towards the unique mixed NE $x_1^* = x_2^* = (1/3, 1/3, 1/3)$ with only $K = 11$ learning samples.

5.3. Experiments with different initializations and payoff matrices

Finally, we conduct simulations across a large set of settings to gain a statistically significant understanding of each methods' performance. Table 1 presents the performance of each method across 100 initial conditions for the three examples discussed in Section 5. Table 2 shows the performance of each method under varying payoff matrices for stag hunt and zero-sum games. The results demonstrate that SIAR-MPC consistently achieves lower error values and control cost compared to other methods. All results here have been averaged across the state variables for compactness; more details on the experiments can be found in Appendix A of the supplementary material.

Method	**Stag Hunt**		**Matching Pennies**		**ϵ-RPS**	
	MSE (Ref.)	Cost	MSE (Ref.)	Cost	MSE (Ref.)	Cost
SIARc	3.48×10^{-2}	5.02×10^{1}	5.72×10^{-2}	2.26×10^{2}	9.66×10^{-3}	2.68×10^{1}
PINN	5.62×10^{-1}	1.09×10^{3}	7.70×10^{-2}	2.69×10^{2}	2.94×10^{-2}	1.03×10^{2}
SINDYc	6.67×10^{-1}	1.27×10^{3}	2.01×10^{-1}	3.00×10^{9}	5.77×10^{-2}	7.34×10^{34}

Table 1: Performance comparison of SIAR-MPC, PINN-MPC, and SINDY-MPC across three games described in Section 5.1-5.2 averaged over 100 initial conditions, evaluated on **MSE(Ref.)**: mean squared error between the estimated and reference trajectories and **Cost**: accumulated control cost.

Method	**Stag Hunt**		**2×2 Zero-sum Games**	
	MSE (Ref.)	Cost	MSE (Ref.)	Cost
SIARc	1.95×10^{-1}	3.71×10^{2}	1.88×10^{-2}	7.24×10^{1}
PINN	5.63×10^{-1}	1.08×10^{3}	3.99×10^{-2}	1.31×10^{2}
SINDYc	5.13×10^{-1}	1.70×10^{7}	4.88×10^{-2}	3.52×10^{37}

Table 2: Performance comparison of SIAR-MPC, PINN-MPC, and SINDY-MPC in stag hunt and 2×2 zero-sum games across 50 payoff matrices, based on the metrics described in Table 1.

6. Conclusion

In this work we introduced SIAR-MPC, a new computational framework for steering game dynamics towards desirable outcomes with limited data. SIAR-MPC extends SIAR for system identification of controlled game dynamics and integrates it with MPC for dynamic incentive adjustments. Our results demonstrated that SIAR-MPC effectively steers systems towards optimal equilibria, stabilizes chaotic and cycling dynamics. Future research can explore several potential directions. First, we intend to address a broader question: given the inherent limitations of game dynamics, where convergence to NE is not always guaranteed, what are the necessary and sufficient conditions for achieving global stability in controlled game dynamics? Second, we aim to extend our framework to encompass games beyond the normal form, thereby expanding its applicability to a wider range of strategic interactions. Finally, we plan to investigate the scalability of our approach by exploring the uncoupling assumption, which could enable its application to larger systems with numerous agents.

Acknowledgments

This research is supported by the MOE Tier 2 Grant (MOE-T2EP20223-0018), the National Research Foundation, Singapore, under its QEP2.0 programme (NRF2021-QEP2-02-P05), the CQT++ Core Research Funding Grant (SUTD) (RS-NRCQT-00002), the National Research Foundation Singapore and DSO National Laboratories under the AI Singapore Programme (Award Number: AISG2-RP-2020-016), and partially by Project MIS 5154714 of the National Recovery and Resilience Plan, Greece 2.0, funded by the European Union under the NextGenerationEU Program.

References

Amir Ali Ahmadi and Bachir El Khadir. Learning dynamical systems with side information. *SIAM Review*, 65(1):183–223, 2023.

Ethan Jon Akin and Viktor Losert. Evolutionary dynamics of zero-sum games. *Journal of Mathematical Biology*, 20(3):231–258, October 1984.

F. Blanchini. Set invariance in control. *Automatica*, 35(11):1747–1767, 1999.

Steven L. Brunton and J. Nathan Kutz. *Data-Driven Science and Engineering: Machine Learning, Dynamical Systems, and Control*. Cambridge University Press, 2019.

Steven L. Brunton, Joshua L. Proctor, and J. Nathan Kutz. Discovering governing equations from data by sparse identification of nonlinear dynamical systems. *Proceedings of the National Academy of Sciences*, 113(15):3932–3937, 2016.

E. F. Camacho and C. Bordons. *Model Predictive Control*. Advanced Textbooks in Control and Signal Processing. Springer London, 2 edition, 2007.

Ilayda Canyakmaz, Can Berk Saner, and Antonios Varvitsiotis. Physics-informed neural networks for privacy-preserving model sharing in power systems. In *2024 IEEE PES Innovative Smart Grid Technologies Europe (ISGT EUROPE)*, pages 1–5, 2024.

Zhao Chen, Yang Liu, and Hao Sun. Physics-informed learning of governing equations from scarce data. *Nature Communications*, 12, 2020.

Miles Cranmer, Alvaro Sanchez-Gonzalez, Peter Battaglia, Rui Xu, Kyle Cranmer, David Spergel, and Shirley Ho. Discovering symbolic models from deep learning with inductive biases. In *Proceedings of the 34th International Conference on Neural Information Processing Systems*, NIPS '20, Red Hook, NY, USA, 2020. Curran Associates Inc.

Drew Fudenberg and David K. Levine. *The theory of learning in games*. MIT Press, Cambridge, MA., 1998.

Lulu Gong, Weijia Yao, Jian Gao, and Ming Cao. Limit cycles analysis and control of evolutionary game dynamics with environmental feedback. *Automatica*, 145:110536, 2022.

Sergiu Hart and Andreu Mas-Colell. Uncoupled dynamics do not lead to Nash equilibrium. *The American Economic Review*, 93(5):1830–1836, December 2003.

Wenjun Hu, Gang Zhang, Haiyan Tian, and Zhiwei Wang. Chaotic dynamics in asymmetric rock-paper-scissors games. *IEEE Access*, 7:175614–175621, 2019.

Jiawei Huang, Vinzenz Thoma, Zebang Shen, Heinrich H Nax, and Niao He. Learning to steer markovian agents under model uncertainty. *arXiv preprint arXiv:2407.10207*, 2024.

Eurika Kaiser, J. Nathan Kutz, and Steven L. Brunton. Sparse identification of nonlinear dynamics for model predictive control in the low-data limit. *Proceedings of the Royal Society A: Mathematical, Physical and Engineering Sciences*, 474(2219):20180335, 2018.

Jean Bernard Lasserre. Global optimization with polynomials and the problem of moments. *SIAM Journal on Optimization*, 11(3):796–817, January 2001.

Jean Bernard Lasserre. A sum of squares approximation of nonnegative polynomials. *SIAM Journal on Optimization*, 16(3):751–765, January 2006.

Monique Laurent. Sums of squares, moment matrices and optimization over polynomials. In Mihai Putinar, Seth Sullivant, Douglas Norman Arnold, and Arnd Scheel, editors, *Emerging applications of algebraic geometry*, number 149 in The IMA volumes in mathematics and its applications, pages 157–270. Springer, New York City, New York, United States, 1 edition, September 2008.

Fei Lu, Ming Zhong, Sui Tang, and Mauro Maggioni. Nonparametric inference of interaction laws in systems of agents from trajectory data. *Proceedings of the National Academy of Sciences*, 116 (29):14424–14433, 2019.

Gabriel F. Machado and Morgan Jones. Sparse identification of nonlinear dynamics with side information (sindy-si). In *2024 American Control Conference (ACC)*, pages 2879–2884, 2024.

Nuno C. Martins, Jair Certório, and Richard J. La. Epidemic population games and evolutionary dynamics. *Automatica*, 153:111016, 2023.

Panayotis Mertikopoulos, Christos Papadimitriou, and Georgios Piliouras. Cycles in adversarial regularized learning. In *Proceedings of the 2018 Annual ACM-SIAM Symposium on Discrete Algorithms (SODA)*, pages 2703–2717, 2018.

Jason Milionis, Christos Papadimitriou, Georgios Piliouras, and Kelly Spendlove. An impossibility theorem in game dynamics. *Proceedings of the National Academy of Sciences*, 120(41): e2305349120, 2023.

Mitio Nagumo. Über die Lage der Integralkurven gewöhnlicher Differentialgleichungen. *Proceedings of the Physico-Mathematical Society of Japan*, 24:551–559, 1942.

John F. Nash. Equilibrium points in in n-person games. *Proceedings of the National Academy of Sciences*, 36(1):48–49, 1950.

Keith Paarporn, Ceyhun Eksin, Joshua S. Weitz, and Yorai Wardi. Optimal control policies for evolutionary dynamics with environmental feedback. In *2018 IEEE Conference on Decision and Control (CDC)*, pages 1905–1910, 2018.

Christos Papadimitriou and Georgios Piliouras. Game dynamics as the meaning of a game. *SIGecom Exch.*, 16(2):53–63, may 2019.

Pablo A. Parrilo. *Structured semidefinite programs and semialgebraic geometry methods in robustness and optimization*. Ph.D. thesis, California Institute of Technology, California, Pasadena, United States, May 2000.

Pablo A. Parrilo. Semidefinite programming relaxations for semialgebraic problems. *Mathematical Programming*, 96(2):293–320, May 2003.

Alexander Prestel and Charles Neal Delzell. *Positive polynomials: from Hilbert's 17th problem to real algebra*. Springer monographs in mathematics. Springer, Berlin, Germany, 1 edition, September 2001.

M. Raissi, P. Perdikaris, and G.E. Karniadakis. Physics-informed neural networks: A deep learning framework for solving forward and inverse problems involving nonlinear partial differential equations. *Journal of Computational Physics*, 378:686–707, 2019.

James Riehl, Pouria Ramazi, and Ming Cao. A survey on the analysis and control of evolutionary matrix games. *Annual Reviews in Control*, 45:87–106, 2018.

Iosif Sakos, Antonios Varvitsiotis, and Georgios Piliouras. Discovering how agents learn using few data, July 2023.

W.H. Sandholm. *Population Games and Evolutionary Dynamics*. Economic Learning and Social Evolution. MIT Press, 2010.

Yuzuru Sato, Eizo Akiyama, and J. Doyne Farmer. Chaos in learning a simple two-person game. *Proceedings of the National Academy of Sciences*, 99(7):4748–4751, April 2002.

Michael Schmidt and Hod Lipson. Distilling free-form natural laws from experimental data. *Science*, 324(5923):81–85, 2009.

Stephen Smale. Dynamics in general equilibrium theory. *The American Economic Review*, 66(2): 288–294, May 1976.

Jochen Stiasny, George S. Misyris, and Spyros Chatzivasileiadis. Physics-informed neural networks for non-linear system identification for power system dynamics. In *2021 IEEE Madrid PowerTech*, pages 1–6, 2021.

Silviu-Marian Udrescu and Max Tegmark. AI Feynman: A physics-inspired method for symbolic regression. *Science Advances*, 6(16):eaay2631, 2020.

Silviu-Marian Udrescu, Andrew Tan, Jiahai Feng, Orisvaldo Neto, Tailin Wu, and Max Tegmark. Ai feynman 2.0: Pareto-optimal symbolic regression exploiting graph modularity. In H. Larochelle, M. Ranzato, R. Hadsell, M.F. Balcan, and H. Lin, editors, *Advances in Neural Information Processing Systems*, volume 33, pages 4860–4871. Curran Associates, Inc., 2020.

Brian Hu Zhang, Gabriele Farina, Ioannis Anagnostides, Federico Cacciamani, Stephen Marcus McAleer, Andreas Alexander Haupt, Andrea Celli, Nicola Gatti, Vincent Conitzer, and Tuomas Sandholm. Steering no-regret learners to a desired equilibrium, 2024.