Inferring Short-Sightedness in Dynamic Noncooperative Games

Cade Armstrong*, Ryan Park*, Xinjie Liu, Kushagra Gupta, and David Fridovich-Keil

Abstract*—Dynamic game theory is an increasingly popular tool for modeling multi-agent, e.g. human-robot, interactions. Game-theoretic models presume that each agent wishes to minimize a private cost function that depends on others' actions. These games typically evolve over a fixed time horizon, which specifies the degree to which all agents care about the distant future. In practical settings, however, decision-makers may vary in their degree of short-sightedness. We conjecture that quantifying and estimating each agent's short-sightedness from online data will enable safer and more efficient interactions with other agents. To this end, we frame this inference problem as an *inverse* dynamic game. We consider a specific parametrization of each agent's objective function that smoothly interpolates myopic and farsighted planning. Games of this form are readily transformed into parametric mixed complementarity problems; we exploit the directional differentiability of solutions to these problems with respect to their hidden parameters in order to solve for agents' short-sightedness. We conduct several experiments simulating human behavior at a real-world crosswalk. The results of these experiments clearly demonstrate that by explicitly inferring agents' short-sightedness, we can recover more accurate game-theoretic models, which ultimately allow us to make better predictions of agents' behavior. Specifically, our results show up to a 30% more accurate prediction of myopic behavior compared to the baseline.

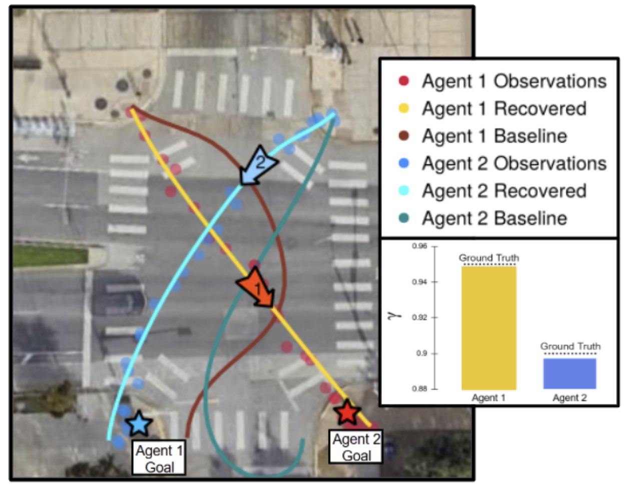


Fig. 1. Representative example showcasing how the proposed formulation improves our ability to model agents' behavior in noncooperative interactions. Here, two simulated pedestrians are crossing an intersection, and by correctly inferring their degree of short-sightedness (i.e., the γ parameter in the inset), our method recovers their trajectories than a baseline approach which does not account for agents' myopic decision-making.

I. Introduction

Robot planning problems often involve strategic interactions between both human and robotic decision makers. Dynamic game theory captures many of these problems' complexities, such as conflicting goals, coupled strategies, and long-term consequences. However, for an agent to use game theoretic planning algorithms, it must understand other agent's hidden objectives. When these objectives are not known *a priori*, we can cast inferring them as an inverse game [1]–[3]. In practice, agents can use algorithms for solving these problems to simultaneously infer others' objectives and plan trajectories which seamlessly interact with one another over time.

Existing inverse game solvers also often assume that decision makers care about each moment in time equally. In practice, both humans and robots can violate this assumption. For example, in an urban driving game modeling a busy intersection, a farsighted driver may slow down further in advance of the intersection than a short-sighted, or myopic, driver would. Robots assuming that the myopic driver is behaving in a farsighted manner may take prematurely evasive actions, resulting in unexpected and unnecessarily risky behavior. Thus, to interact safely and efficiently, autonomous robots must be able to infer the short-sightedness of other decision makers.

This paper aims to address the gap between the assumptions of existing game theoretic algorithms and decision makers' potential myopia. In particular, existing methods typically model each agent's objective as a sum of cost functions which can depend upon all agents' states and actions. In this work, we consider a time-discounted cost formulation, which smoothly interpolates between myopia and farsightedness for each agent individually, and develop a gradient-based algorithm for inverting the corresponding game and identifying agents' degree of short-sightedness. Our specific contributions are:

1) an explicitly myopic game formulation,
2) an efficient method to infer agents' individual degree of short-sightedness from data collected online,
3) and a series of simulation experiments which characterize the potential benefits of modeling agents' myopia.

In particular, we perform experiments simulating the interaction of two myopic human agents moving through a crosswalk with intersecting trajectories, as shown in Figure 1. Results demonstrate that our method recovers more accurate models of myopic agents than existing models, as well as being more robust to noise in partially observable settings. Specifically, our method produces up to a 30% improvement in trajectory reconstruction error in comparison to a baseline method, in partially observable environments.

This work was supported by a National Science Foundation CAREER award under Grant No. 2336840. All authors are with the Oden Institute for Computational Sciences and Engineering, University of Texas at Austin, 201 E 24th St, Austin, TX. {cadearmstrong, ryanjpark, xinjie-liu, kushagrag, dfk}@utexas.edu

*indicates equal contribution (Corresponding author: Cade Armstrong).

II. Related Works

A. Noncooperative Dynamic Games

An N-agent dynamic game models each agent's choice of strategy in terms of a optimization problem, whose objective and/or constraints can depend upon other agents' strategies. Dynamic games admit different solutions depending on the information available to each agent at every time step [4]. In this work, we consider the *Nash* equilibrium concept in the setting of *open-loop* strategies, i.e., we presume that at each time t agents only have information about the initial state of the game. Early works in this setting focused primarily on problems with convex, quadratic costs and linear game dynamics, without any additional constraints [4]–[7]. However, finding a Nash equilibrium in more general settings is often intractable [8]. Consequently, many recent works develop iterative trajectory optimization methods to find *approximate, local* open-loop Nash equilibria in games with nonlinear costs and complex constraint structures. Such approximations satisfy first-order optimality conditions of the underlying optimization problem for each agent, and second-order conditions for local optimality must generally be checked *a posteriori*. One class of such methods employs root-finding techniques for the underlying joint N-agent Karush-Kuhn-Tucker (KKT) system [9]–[11]. Another class employs iterative best-response algorithms to approximate Nash equilibria [12], [13].

B. Inverse Dynamic Games

Inverse dynamic games seek to identify unknown aspects of game models, e.g., agents' intentions and constraints, from observed interactions. Early approaches minimize the residual of the KKT system to identify possible game parameter values [14]–[16]. However, this type of approach requires *full* state-action demonstration to evaluate the residuals, posing a challenge for online estimation tasks, where not all agents' states and actions are available. Worse, the assumption that the demonstrations satisfy first-order optimality conditions means that such residual minimization approaches can perform poorly in the presence of observation noise [17]. To mitigate these issues, recent work formulates inverse game problems in terms of maximum likelihood estimation (MLE), and explicitly impose first-order optimality conditions as constraints [17]–[20]. In particular, our work builds directly upon this line of work, and specifically extends the solution approach of [18], which develops a differentiable game solver that enables inverse game problems to be solved via gradient descent on the unknown game parameters.

While MLE approaches to inverse games demonstrate strong performance in various scenarios, they are fundamentally limited to provide point estimates of unknown parameters. To this end, another line of work seeks to infer full Bayesian posterior *distributions* for the unknown game parameters, e.g., via particle filtering [21] and unscented Kalman filtering [22]. However, high computational costs hinder these approaches from scaling to larger systems. More recently, work in [23] proposes to embed a differentiable Nash solver into generative models to solve variational inference problems. However, this Bayesian line of inverse game approaches are only approximate in nature, and still suffer from high computational costs.

Finally, we note that inverse games have also been formulated in maximum-entropy settings [24], inferring opinion dynamics model [20], and in combination with neural network components [25].

III. Background

A. Dynamic Games

A dynamic game [4] is characterized by N agents, with the i^{th} agent's control input denoted $u_t^i \in \mathbb{R}^{m^i}\ \forall\, t$, and a state variable $x_t \in \mathbb{R}^n$ which follows given dynamics $x_{t+1} = f_t(x_t, u_t^1, u_t^2, \ldots, u_t^N)$ at discrete times $[T] := \{1, 2, ..., T\}$. Each agent has a cost function

$$J^i := \sum_{t=1}^{T} \Gamma^i(t;\gamma^i) C^i(x_t, u_t^i, u_t^{\neg i}; \theta^i), \tag{1}$$

which depends upon the state and its own actions, as well as others' actions $u_t^{\neg i}$, hidden parameters $\theta^i \in \mathbb{R}^k$, and *discount factor* γ^i. At each time t, agent i's cost is comprised of parametrized function $J^i(\cdot)$ and scaled by a parametrized *discounting function* $\Gamma^i(t;\gamma^i)$, which quantifies how important the cost at time t is to agent i.

In principle, the formulation of the game in (1) is general enough to handle arbitrary functions Γ^i. In this work, we assume that this discounting function is a decaying exponential, i.e. $\Gamma^i(t;\gamma^i) = (\gamma^i)^t$, where $\gamma^i \in (0, 1]$.

We shall refer to all agents' discount factors as a vector $\gamma = (\gamma^1, \ldots, \gamma^N) \in [0, 1]^N$ for convenience. Observe that the function $\Gamma^i(t;\gamma^i)$ rapidly approaches 0 as $t \to T$, unless $\gamma^i = 1$. Thus, the i^{th} agent's actions do not depend upon costs incurred past some effective time horizon $\tilde{T}^i \in [T]$.

For notational convenience, lowercase bold variables refer to aggregations over time and the omission of agent indices denotes aggregations of all agents. In addition, agent indices are always written as superscripts while time indices are always written as subscripts.

Thus, we refer to the sequence of game states as $\mathbf{x} = (x_1^\top, x_2^\top, \ldots, x_T^\top)^\top$, agent i's control sequence as $\mathbf{u}^i = (u_1^{i,\top}, u_2^{i,\top}, \ldots, u_T^{i,\top})^\top$, the sequence of all agents' actions as $\mathbf{u} = (\mathbf{u}^{1,\top}, \mathbf{u}^{2,\top}, \ldots, \mathbf{u}^{N,\top})^\top$, and all agents' hidden parameters as $\theta = (\theta^{1,\top}, \theta^{2,\top}, \ldots, \theta^{N,\top})^\top$. Then, we can write each cost function as $C_t^i(x_t, u_t^i, u_t^{\neg i}; \theta^i) = C_t^i(x_t, u_t; \theta^i)$ and thus agent i's overall cost is $J^i(\mathbf{x}, \mathbf{u}; \gamma^i, \theta^i)$ with a slight abuse of notation. In general, we may also assign each agent a set of inequality constraints $I^i(\mathbf{x}, \mathbf{u}; \theta^i) \geq 0$ and a set of equality constraints $E^i(\mathbf{x}, \mathbf{u}; \theta^i) = 0$. Finally, we can define a game as a tuple:

$$\begin{aligned}\mathcal{G}(\theta, \gamma) = (&\{J^i(\cdot;\gamma^i, \theta^i)\}_{i\in[N]}, \{I^i(\cdot;\theta^i)\}_{i\in[N]},\\ &\{E^i(\cdot;\theta^i)\}_{i\in[N]},\, x_1,\, T,\, N).\end{aligned} \tag{2}$$

Note that (2) does not include the dynamics $f_t(\cdot)$ explicitly; this is because the dynamics generate $(T-1)$ equality constraints of the form $x_{t+1} - f_t(x_t, u_t^1, u_t^2, \ldots, u_t^N) = 0$ for each of the N agents. For clarity, we assume that $E^i(\cdot)$ contains these terms.

B. Generalized Open-Loop Nash Equilibria

For a given set of parameters θ, a generalized open-loop Nash equilibrium (GOLNE) of the game $\mathcal{G}(\theta)$ in (2) is given by a point $(\mathbf{x}^*, \mathbf{u}^*)$ which jointly solves the following coupled optimization problems:

$$\forall i \in [N] \begin{cases} \min_{\mathbf{x},\mathbf{u}^i} \sum_{t=1}^{T} \Gamma^i(t;\gamma^i) C^i(x_t, u_t^i, u_t^{\neg i}; \theta^i) & (3a) \\ \text{s.t. } E^i(\mathbf{x}, \mathbf{u}; \theta^i) = 0 & (3b) \\ \quad I^i(\mathbf{x}, \mathbf{u}; \theta^i) \geq 0. \end{cases}$$

The strategies $\mathbf{u}^* = (\mathbf{u}^{1*}, \mathbf{u}^{2*}, ..., \mathbf{u}^{N*})$ have the property that $J^i(\mathbf{x}, \mathbf{u}^i, \mathbf{u}^{\neg i*}; \gamma^i, \theta^i) \geq J^i(\mathbf{x}^*, \mathbf{u}^*; \gamma^i, \theta^i)$ for all feasible $\mathbf{u}^i$ and $\mathbf{x}$, for all agents $i \in [N]$. Intuitively, no agent can reduce their cost by unilaterally deviating from a Nash equilibrium strategy, holding other agents' strategies fixed.

C. Mixed Complementarity Problem (MiCP)

Finding a generalized Nash equilibrium is computationally intractable [26]; therefore, it is common to relax the condition in Section III-B to hold only within an open neighborhood of the point $(\mathbf{x}^*, \mathbf{u}^*)$ [10], [11], [27]–[29]. Such points are called local Nash equilibria [30]. Practically, such points can be identified by solving agents' first order necessary conditions, which constitute a Mixed Complementarity Problem (MiCP) [31]. An MiCP is defined by decision variables $r \in \mathbb{R}^{\eta_r}, z \in \mathbb{R}^{\eta_z}$, as well as functions $c(r, z)$ and $h(r, z)$, such that

$$c(r, z) = 0 \tag{4a}$$

$$0 \leq z \perp h(r, z) \geq 0. \tag{4b}$$

MiCPs can often be solved efficiently via off-the-shelf solvers such as PATH [32].

IV. Problem Statement

We presume that an observer has obtained noisy sensor measurements of the game state over time, and denote these observations $\mathbf{y} = [y_1(x_1), y(x_2), y_3(x_3), \ldots, y_T(x_T)]$. We assume observations are drawn from Gaussian models

$$y_t(x_t) \sim \mathcal{N}\big(h_t(x_t), \Sigma_t\big) \tag{5}$$

with known covariances Σ_t, where $h_t(x_t)$ describes the expected output of the observation model at time t. Each y_t is independent from the others across time.

Given measurements $\mathbf{y}$, we seek to identify game parameters $\hat{\theta}, \hat{\gamma}$ that maximizes the likelihood to have observed $\mathbf{y}$ from a GOLNE of $\mathcal{G}(\hat{\theta}, \hat{\gamma})$. Equivalently, one can minimize covariance-weighted deviations from expected measurements, which we denote $\mathcal{P}$.

Problem 1. ***Inverse Game Problem:*** *Given a sequence of observations* $\mathbf{y}$*, find parameters* $\hat{\theta}$ *which solve*

$$\min_{\boldsymbol{x},\boldsymbol{u},\theta,\gamma} \overbrace{\sum_{t=1}^{T} (h_t(x_t) - y_t)^\top \Sigma_t^{-1} (h_t(x_t) - y_t)}^{\mathcal{P}(\boldsymbol{x}(\theta,\gamma))} \tag{6a}$$

$$\textit{s.t.} \quad (\boldsymbol{x}, \boldsymbol{u}) \textit{ is a GOLNE of } \mathcal{G}(\theta, \gamma). \tag{6b}$$

Intuitively, the problem can be separated into an outer *inverse* problem and an inner *forward* problem. The forward GOLNE problem (6b) is parametrized by the game parameters θ and discount factors γ, which are decision variables in the inverse problem (6a). This formulation is general enough to handle partial state observations, making it amenable to realistic observation models such as cameras or GPS, in principle. However, solving (6a) entails computational challenges. In particular, the KKT conditions of (6b) are certainly nonlinear in γ, making the overall problem non-convex. In addition, (6b) may contain inequality constraints, which imply that its KKT conditions involve complementarity conditions which make the overall problem non-smooth.

V. Solution Approach

In this section, we present a constrained gradient descent algorithm for identifying unknown parameters θ, γ in Problem 1. This technique will require us to take derivatives of solutions $(\mathbf{x}, \mathbf{u})$ to (6b) with respect to parameters (θ, γ). Therefore, we begin with a discussion of transforming a GOLNE into an MiCP, whose solutions are directionally differentiable with respect to problem parameters [31, Ch. 5].

A. Equilibrium Constraint as an MiCP

In this subsection, we show how to convert a GOLNE into a MiCP which encodes its first-order necessary conditions. Similar techniques have been used in [18], [33]–[35] to solve inverse optimal control and game problems.

We start with writing the agents' first-order necessary conditions. First, we introduce Lagrange multipliers λ^i and μ^i for the i^{th} agent's inequality and equality constraints, respectively, and write its Lagrangian as

$$\begin{aligned} \mathcal{L}^i(\mathbf{x}, \mathbf{u}, \lambda^i, \mu^i; \theta^i, \gamma^i) = C^i(\mathbf{x}, \mathbf{u}; \theta^i, \gamma^i) - \\ \lambda^{i,\top} I^i(\mathbf{x}, \mathbf{u}; \theta^i) - \mu^{i,\top} E^i(\mathbf{x}, \mathbf{u}; \theta^i). \end{aligned} \tag{7}$$

Then, when the gradients of the constraints are linearly independent at a candidate solution point (i.e., the linear independence constraint qualification is satisfied [31, Chapter 3.2]), the following Karush-Kuhn-Tucker (KKT) conditions must hold for each agent i:

$$\nabla_{\mathbf{x}} \mathcal{L}^i(\mathbf{x}, \mathbf{u}, \lambda^i, \mu^i; \theta^i, \gamma^i) = 0 \tag{8a}$$

$$\nabla_{u^i} \mathcal{L}^i(\mathbf{x}, \mathbf{u}, \lambda^i, \mu^i; \theta^i, \gamma^i) = 0 \tag{8b}$$

$$E^i(\mathbf{x}, \mathbf{u}; \theta^i) = 0 \tag{8c}$$

$$0 \leq \lambda^i \perp I^i(\mathbf{x}, \mathbf{u}; \theta^i) \geq 0. \tag{8d}$$

We can then structure agents' joint KKT conditions as a parametric MiCPs (4), in which the primal and dual variables are concatenated as follows:

$$r = [\mathbf{x}^\top, \mathbf{u}^\top, \mu^{1,\top}, \mu^{2,\top}, \ldots, \mu^{N,\top}]^\top \tag{9a}$$

$$z = [\lambda^{1,\top}, \lambda^{2,\top}, \ldots, \lambda^{N,\top}]^\top. \tag{9b}$$

For brevity, define $v = (r^\top, z^\top)^\top$. Then, with a slight abuse of notation, the parameterized MiCP for each agent can be written as described in (4), where functions $c(\cdot)$ and $h(\cdot)$ are defined as

$$c(v;\theta,\gamma) = \begin{bmatrix} \left(\nabla_{\mathbf{x}} L^i\right)_{i\in[N]} \\ \left(\nabla_{u^i} L^i\right)_{i\in[N]} \\ \left(E^i\right)_{i\in[N]} \end{bmatrix} \tag{10a}$$

$$h(v;\theta,\gamma) = \left[\left(I^i(v;\theta^i)\right)_{i\in[N]}\right] \tag{10b}$$

For brevity, we will define $F(v;\theta,\gamma) = [c(\cdot)^\top, h(\cdot)^\top]^\top$. For additional details on connections between GOLNEs and MiCPs in the context of open loop dynamic games, we direct the reader to [18], [31].

B. Optimizing Parameters with Gradient Descent

In this section, we present our algorithm for solving Problem 1. We first replace (6b) with the KKT conditions in (8), and transcribe them into an MiCP according to Section V-A. Then, we compute the total derivative of objective function $\mathcal{P}$ with respect to parameters (θ, γ), and update their values accordingly. These gradients can be computed via the chain rule:

$$\nabla_{(\theta,\gamma)}\mathcal{P}(\mathbf{x}(\theta,\gamma)) = (\nabla_{(\theta,\gamma)} v)^\top (\nabla_v \mathbf{x})^\top (\nabla_{\mathbf{x}} \mathcal{P}). \tag{11}$$

The only non-trivial term in (11) is $\nabla_{(\theta,\gamma)} v$. Next, we show how to take directional derivatives of (4) at a solution v^* where strictly complementarity holds.

First, we must consider the complementarity constraints on z and $h(\cdot)$. To this end, we construct an index set $\mathcal{I}$ which records all inactive inequality constraint dimensions in $h(v;\theta,\gamma)$. Indexing F at elements of this set yields a vector $[F]_{\mathcal{I}}$. From (4b), we know that these inactive inequalities $[F]_{\mathcal{I}}$ are strictly positive, and the Lagrange multipliers associated with these constraints are exactly 0. Then, presuming the continuity of F, small changes in (θ,γ) preserve the positivity of elements of $[F]_{\mathcal{I}}$ and force the corresponding Lagrange multipliers in $[v]_{\mathcal{I}}$ to remain 0. Thus, we find that $\nabla_{(\theta,\gamma)}[v]_{\mathcal{I}} = 0$.

Consider the remaining constraints from (8), which must be active due to strict complementarity; denote the indices corresponding to these constraint as $\mathcal{S}$. We can use the implicit function theorem and stationarity of F with respect to v to write

$$0 = \nabla_{(\theta,\gamma)}[F]_{\mathcal{S}} = \nabla_{(\theta,\gamma)}[F]_{\mathcal{S}} + (\nabla_v [F]_{\mathcal{S}})(\nabla_{(\theta,\gamma)} v) \tag{12a}$$

$$\implies \nabla_{(\theta,\gamma)} v = -(\nabla_v [F]_{\mathcal{S}})^{-1} \nabla_{(\theta,\gamma)}[F]_{\mathcal{S}}. \tag{12b}$$

Then, when $\nabla_v[F]_{\mathcal{S}}$ is invertible, we can find exact values of $\nabla_{(\theta,\gamma)} v$. For a more complete treatment, including a discussion of weak complementarity, readers are encouraged to consult [18], [31], [36].

Thus equipped, we iteratively update our estimate of (θ,γ) according to the gradient of (6a)—incorporating the aforementioned implicit derivatives as needed. Line 1 terminates after a maximum number of iterations (i.e., when $k > K$) or when parameters have converged (i.e., $\|\theta_{k+1} - \theta_k\|_2 \leq \epsilon$). In our experiments, $K = 500$ and $\epsilon = 10^{-4}$.

Finally, observe that the discounting function Γ only appears in each agent's Lagrangian, and thus we require that $\Gamma \in C^2$ to approximate directional derivatives of our MiCP. At each step, also note that we project the discount factor γ onto the set $[0,1]^N$ to ensure feasibility.

Algorithm 1: Myopia-aware inverse game

1 **Hyper-parameters:** Learning rate α
2 **Input:** initial θ, initial γ, observations $\mathbf{y}$
3 $\theta_0 \leftarrow \theta$
4 $\gamma_0 \leftarrow \gamma$
5 $k \leftarrow 0$
6 **while** *not converged* **do**
7 $\quad v_k \leftarrow \text{solveInnerMCP}(\theta_k, \gamma_k)$ ▷ (10)
8 $\quad \nabla_{(\theta_k,\gamma_k)}\mathcal{P} \leftarrow \text{calcGradient}(v_k, \theta_k, \gamma_k)$ ▷ (11), (12)
9 $\quad \theta_{k+1} \leftarrow \theta_k - \nabla_{\theta_k}\mathcal{P} \cdot \alpha$
10 $\quad \gamma_{k+1} \leftarrow \min(1, \max(0, \gamma_k - \nabla_{\gamma_k}\mathcal{P} \cdot \alpha))$
11 $\quad k \leftarrow k+1$
12 **return** $(\theta_k, \gamma_k, \mathbf{x}, \mathbf{u})$

VI. Experiments

In this section, we test the performance and robustness of the proposed approach in Section V and compare it to baseline inverse game approaches in simulated experiments. Results show that Line 1 performs better than an existing inverse game formulation across various performance metrics.

A. Example Game

Our experiment simulates $N = 2$ pedestrians navigating a crosswalk at an intersection on the UT Austin campus at W Dean Keaton St. and Whitis Ave. This intersection is unusual in that the entire intersection becomes safe for pedestrian crossing at once, encouraging students to cross diagonally and interact with one another. We simulate two short-sighted students who begin at the top right and left corners respectively, and their goals are to reach the bottom left and right corners respectively.

We model the students as wishing to minimize distance to their goal and control effort (they are tired), such that the i^{th} agent's cost can be written as:

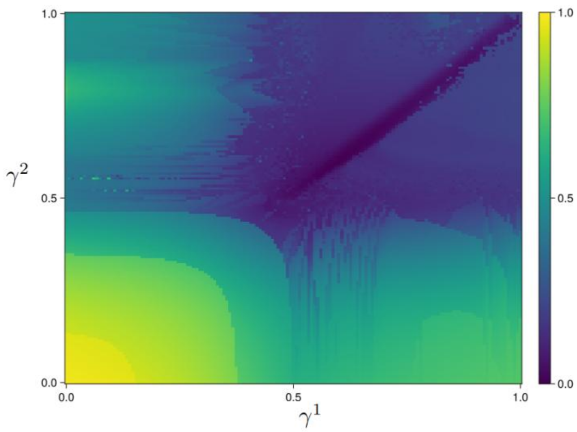
$$J^i = \sum_{t=0}^{T} (\gamma^i)^t \left[w^i_{\text{goal}} \|p^i_t - \theta^i_{\text{goal}}\|_2^2 + w^i_{\text{ctrl}} \|u^i_t\|_2^2\right]. \tag{13}$$

For brevity, we define $p^i_t \in \mathbb{R}^2$ to denote the part of x_t which corresponds to agent i's position at time t, w^i_{goal} and w^i_{ctrl} weight the goal and control cost terms, and the game is parameterized by the i^{th} agent's target position $\theta^i_{\text{goal}} \in \mathbb{R}^2$. The i^{th} agent has a collision avoidance constraint defined as

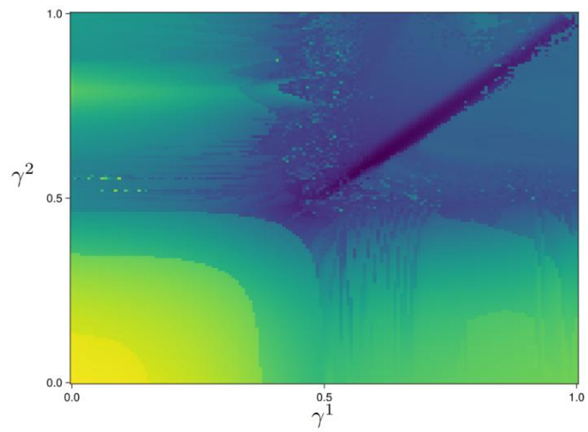
$$I^i(\cdot) = \left[\left(\|p^i_t - p^{\neg i}_t\|_2^2 - \delta \cdot d_{\min}\right)_{t\in[T]}\right] \geq 0, \tag{14}$$

where $d_{\min}$ denotes the minimum allowed distance between the agents and $\delta > 1$ is a safety margin.

The students' physical dynamics are modeled point masses, with bounds on the velocity and control inputs, assume a time discretization of $\Delta = 0.1\,\text{s}$, and play the game over $T = 25$



(a) Objective function, partially-observed case



(b) Objective function, fully-observed case

Fig. 2. Here we plot a heatmap of $\mathcal{P}$ from Problem 1 in a (a) fully observable and (b) partially observable setting. In both cases, both player's γ are varied while holding θ constant at the ground truth values and the resulting cost is scaled to be in between $[0, 1]$.

time steps. Explicitly, the game's state at time t is composed as $x_t = [x_t^{1,\top}, x_t^{2,\top}]^\top$, and for the i^{th} agent:

$$x_{t+1}^i = \begin{bmatrix} 1 & 0 & \Delta & 0 \\ 0 & 1 & 0 & \Delta \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix} \underbrace{\begin{bmatrix} p_{x,t}^i \\ p_{y,t}^i \\ v_{x,t}^i \\ v_{y,t}^i \end{bmatrix}}_{x_t^i} + \begin{bmatrix} \Delta^2/2 & 0 \\ 0 & \Delta^2/2 \\ \Delta & 0 \\ 0 & \Delta \end{bmatrix} \underbrace{\begin{bmatrix} a_{x,t}^i \\ a_{y,t}^i \end{bmatrix}}_{u_t^i}$$

In our testing, we found that Line 1 converged more reliably when these constraints are implemented as large penalties in each agent's cost function. Therefore, we model i^{th} agent's objective function in the inverse game as

$$J^i = \sum_{t=0}^{T} (\gamma^i)^t \left[w_{\text{goal}}^i \| p_t^i - \theta_{\text{goal}}^i \|_2^2 + w_{\text{ctrl}}^i \| u_t^i \|_2^2 \right] + w_{\text{coll}}^i \max(0, \delta \cdot d_{\text{min}} - \| p_t^i - p_t^{\neg i} \|_2^2), \quad (15)$$

where w_{coll}^i is a weighting parameter which governs the importance of collision avoidance for the i^{th} agent.

B. Experimental Setup

We aim to show that our method predicts trajectories more accurately than a baseline method, which solves the same inverse problem *but presumes agents have a discount factor of* 1.[1] Thus, the baseline estimator has two fewer hidden parameters to estimate than our method.

To ensure a fair comparison, we use the same MiCP backend [32] to satisfy the first-order necessary conditions of the GOLNE at line 7 of Line 1. Additionally, we set the default learning rate to $\alpha = 10^{-3}$ for both methods. As input, both methods receive the same observation sequence at every simulation, as well as the same initial θ; our method receives a uniformly random $\gamma \in (0, 1]$ as an initial estimate.

[1]For more details on our implementation, refer to our GitHub repository.

To compare the two methods, we calculate their respective error as

$$\text{error} = \| \mathbf{x} - \mathbf{x}_{\text{ground-truth}} \|_2^2, \quad (16)$$

enabling us to determine how well the methods predicted the agents' trajectories. We gather these results with a Monte Carlo study done over a range of observation noise covariances as per (5). Specifically, we sample observations from a Gaussian distribution with mean

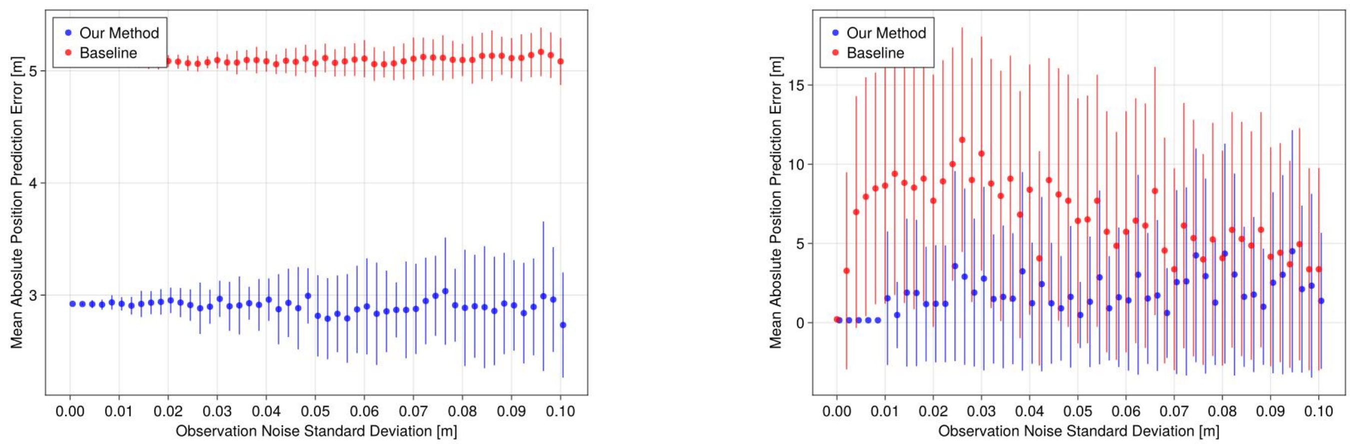
$$h(x_t) = \begin{bmatrix} p_{x,t}^1 \\ p_{y,t}^1 \\ p_{x,t}^2 \\ p_{y,t}^2 \end{bmatrix} \quad (17)$$

and covariance given by identity matrices scaled by 50 evenly spaced constants between 0 and $0.1\,\text{m}^2$. For each covariance matrix, we generated 50 noisy partial state observation sequences given to our method and the baseline method, resulting in 2500 observation sequences and parameter estimates. For fully observable trials we generate 50 observation sequences at the same noise levels, with $h(x_t) = x_t$.

C. Monte Carlo Results

Figure 1 shows a representative trial for our running example on the real-world crosswalk being simulated. It shows the two players initial positions, their respective ground-truth goals, their inferred goals for the baseline and our method, the observed trajectory being recovered, and their current and planned trajectories at a snapshot in time. Note that our method results in a more visually similar trajectory in comparison to the observations. In addition, according to the bar graph, our method accurately recovers the discount factor within 0.3%.

Next, in Figure 2 we visualize the objective function for Problem 1 as a heatmap, for a representative setting with θ fixed to the ground truth values, in both fully- and partially-observed settings. Here, the ground truth values of the agents' discount factors are $\gamma^1 = \gamma^2 = 0.6$. As shown, in both



(a) With only position measurements

(b) With full state measurements

Fig. 3. Line 1 reliably improves upon the baseline, indicating that modeling and estimating agents' short-sightedness leads to better trajectory predictions. These results hold in settings with both (a) partial state observations and (b) full state observations. Dots and bars denote means and standard deviations over 50 trials at each observation noise level.

settings, the inverse game's objective function has a global minimum at this point. However, we also observe that there is a narrow, diagonally-shaped low cost region: this implies that the equilibrium of the game is relatively insensitive to agents' discount factors as long as both agents have similar discount factors (and both are greater than ≈ 0.5). While it is intuitive, perhaps, that the relative performance of the agents should exhibit this kind of symmetry, it is not immediately clear that equilibria of the game should do so.

Nevertheless, this observation motivates us to introduce the following *regularization* scheme for the inverse game in Problem 1.

Problem 2. ***Regularized Inverse Game Problem:*** *Given a sequence of observations* $\mathbf{y}$*, find parameters* $\hat{\theta}$ *which solve*

$$\min_{\boldsymbol{x},\boldsymbol{u},\theta,\gamma} \quad \mathcal{P}\left(\boldsymbol{x}(\theta,\gamma)\right) + c_\gamma \|1-\gamma\|_2^2 \tag{18a}$$

$$\textit{s.t.} \qquad (\boldsymbol{x},\boldsymbol{u}) \textit{ is a GOLNE of } \mathcal{G}(\theta,\gamma). \tag{18b}$$

In Problem 2, we have introduced the term $\|1-\gamma\|_2^2$, which introduces a slight bias toward higher discount factors. Note, however, that this essentially encodes a prior belief that agents are more forward-looking; it is straightforward to instead penalize $\|\gamma\|_2^2$ and encode the belief that agents are more short-sighted. In the following Monte Carlo results, we report results for $c_\gamma = 10^{-3}$.

Figure 3 shows the mean trajectory estimation error (16) as a function of the observation noise covariance, for Line 1 and the $\gamma = 1$ baseline. Our method substantially outperforms the baseline, and exhibits about a 30% more accurate recovered trajectory in the partially observable case, and a decreasing improvement with more noise in the fully observable case. Interestingly, we can see that the mean error is only marginally sensitive to noise for both methods. Rather, the most noticeable effect of noise is an increase in the standard deviation of the errors. From these results, we conclude that our myopic formulation can better predict the behavior of short-sighted agents in comparison to a state-of-the-art inverse game baseline [18].

VII. Conclusions

This paper formulates a noncooperative game which models models agents as potentially short-sighted, by associating each agent to an unknown discount factor. Our approach rewrites the equilibrium conditions of the resulting game as an MiCP, and leverages the (directional) differentiability of equilibrium solutions with respect to unknown parameters. This construction allows us to solve an *inverse game* problem and identify hidden parameters—including agents' unknown discount factors—via a gradient descent procedure. We evaluate our method on a simulated crosswalk environment with noisy partial observations and demonstrated superior performance compared to an existing state-of-the-art method.

This work attempts to better describe myopic behavior through the use of discount factors. In this vein, we find two avenues for further research interesting. First, humans often make decisions assuming that they will get to see future events and react to them; in other words, humans often operate in feedback information structures. Future work may extend our formulation and method beyond the open-loop information structure used in this paper. Li *et al.* [19] provides a clear starting point for this direction. Second, humans can also vary how far into the future they care about depending on the present situation. For example, in a highway driving scenario, a person may not look very far into the future if traffic conditions are clear and no changes are expected. However, when merging on and off the highway or when traffic conditions worsen, a human driver may begin to look further in the future to avoid missing exits or being trapped in undesirable traffic situations. Thus, future work may explore discount factors that vary as a function of state as well as time.

References

[1] D. Hadfield-Menell, S. J. Russell, P. Abbeel, and A. Dragan, "Cooperative inverse reinforcement learning," *Advances in neural information processing systems*, vol. 29, 2016.

[2] V. Kuleshov and O. Schrijvers, "Inverse game theory: Learning utilities in succinct games," in *Web and Internet Economics: 11th International Conference, WINE 2015, Amsterdam, The Netherlands, December 9-12, 2015, Proceedings 11*, Springer, 2015, pp. 413–427.

[3] T. L. Molloy, J. I. Charaja, S. Hohmann, and T. Perez, *Inverse optimal control and inverse noncooperative dynamic game theory*. Springer, 2022.

[4] T. Başar and G. J. Olsder, *Dynamic Noncooperative Game Theory, 2nd Edition*. Society for Industrial and Applied Mathematics, 1998. DOI: 10.1137/1.9781611971132. eprint: https://epubs.siam.org/doi/pdf/10.1137/1.9781611971132. [Online]. Available: https://epubs.siam.org/doi/abs/10.1137/1.9781611971132.

[5] A. W. Starr and Y.-C. Ho, "Nonzero-sum differential games," *Journal of optimization theory and applications*, vol. 3, pp. 184–206, 1969.

[6] J. C. Engwerda, "On the open-loop nash equilibrium in lq-games," *Journal of Economic Dynamics and Control*, vol. 22, no. 5, pp. 729–762, 1998.

[7] J. C. Engwerda, "Computational aspects of the open-loop nash equilibrium in linear quadratic games," *Journal of Economic Dynamics and Control*, vol. 22, no. 8-9, pp. 1487–1506, 1998.

[8] C. Daskalakis, P. W. Goldberg, and C. H. Papadimitriou, "The complexity of computing a nash equilibrium," *Communications of the ACM*, vol. 52, no. 2, pp. 89–97, 2009.

[9] S. L. Cleac'h, M. Schwager, and Z. Manchester, "Algames: A fast solver for constrained dynamic games," *arXiv preprint arXiv:1910.09713*, 2019.

[10] B. Di and A. Lamperski, "First-order algorithms for constrained nonlinear dynamic games," *arXiv preprint arXiv:2001.01826*, 2020.

[11] E. L. Zhu and F. Borrelli, "A sequential quadratic programming approach to the solution of open-loop generalized nash equilibria," in *2023 IEEE International Conference on Robotics and Automation (ICRA)*, IEEE, 2023, pp. 3211–3217.

[12] R. Spica, E. Cristofalo, Z. Wang, E. Montijano, and M. Schwager, "A real-time game theoretic planner for autonomous two-player drone racing," *IEEE Transactions on Robotics*, vol. 36, no. 5, pp. 1389–1403, 2020.

[13] M. Wang, Z. Wang, J. Talbot, J. C. Gerdes, and M. Schwager, "Game theoretic planning for self-driving cars in competitive scenarios.," in *Robotics: Science and Systems*, 2019, pp. 1–9.

[14] S. Rothfuß, J. Inga, F. Köpf, M. Flad, and S. Hohmann, "Inverse optimal control for identification in non-cooperative differential games," *IFAC-PapersOnLine*, vol. 50, no. 1, pp. 14 909–14 915, 2017, ISSN: 2405-8963. DOI: https://doi.org/10.1016/j.ifacol.2017.08.2538.

[15] C. Awasthi and A. Lamperski, "Inverse differential games with mixed inequality constraints," in *Proc. of the IEEE American Control Conference (ACC)*, 2020.

[16] J. Inga, E. Bischoff, F. Köpf, and S. Hohmann, "Inverse dynamic games based on maximum entropy inverse reinforcement learning," *arXiv preprint arXiv:1911.07503*, 2019.

[17] L. Peters, V. Rubies-Royo, C. J. Tomlin, *et al.*, "Online and offline learning of player objectives from partial observations in dynamic games," in *Intl. Journal of Robotics Research (IJRR)*, 2023. [Online]. Available: https://journals.sagepub.com/doi/pdf/10.1177/02783649231182453.

[18] X. Liu, L. Peters, and J. Alonso-Mora, "Learning to play trajectory games against opponents with unknown objectives," *IEEE Robotics and Automation Letters*, vol. 8, no. 7, pp. 4139–4146, 2023. DOI: 10.1109/LRA.2023.3280809.

[19] J. Li, C.-Y. Chiu, L. Peters, S. Sojoudi, C. Tomlin, and D. Fridovich-Keil, "Cost inference for feedback dynamic games from noisy partial state observations and incomplete trajectories," in *Proceedings of the International Conference on Autonomous Agents and Multi-Agent Systems (AAMAS)*, 2023, pp. 1062–1070.

[20] H. Hu, J. DeCastro, D. Gopinath, G. Rosman, N. E. Leonard, and J. F. Fisac, "Think deep and fast: Learning neural nonlinear opinion dynamics from inverse dynamic games for split-second interactions," *arXiv preprint arXiv:2406.09810*, 2024.

[21] L. Peters, "Accommodating intention uncertainty in general-sum games for human-robot interaction," 2020.

[22] S. Le Cleac'h, M. Schwager, and Z. Manchester, "LUCIDGames: Online unscented inverse dynamic games for adaptive trajectory prediction and planning," *IEEE Robotics and Automation Letters (RA-L)*, vol. 6, no. 3, pp. 5485–5492, 2021.

[23] X. Liu, L. Peters, J. Alonso-Mora, U. Topcu, and D. Fridovich-Keil, *Auto-encoding bayesian inverse games*, 2024. arXiv: 2402.08902 [cs.RO]. [Online]. Available: https://arxiv.org/abs/2402.08902.

[24] N. Mehr, M. Wang, M. Bhatt, and M. Schwager, "Maximum-entropy multi-agent dynamic games: Forward and inverse solutions," *IEEE Transactions on Robotics*, vol. 39, no. 3, pp. 1801–1815, 2023. DOI: 10.1109/TRO.2022.3232300.

[25] C. Diehl, T. Klosek, M. Krueger, N. Murzyn, T. Osterburg, and T. Bertram, "Energy-based potential games for joint motion forecasting and control," in *Proc. of the Conf. on Robot Learning (CoRL)*, 2023.

[26] C. H. Papadimitriou, "The complexity of finding nash equilibria," *Algorithmic game theory*, vol. 2, p. 30, 2007.

[27] E. V. Mazumdar, M. I. Jordan, and S. S. Sastry, "On finding local nash equilibria (and only local nash equilibria) in zero-sum games," *arXiv preprint arXiv:1901.00838*, 2019.

[28] E. Mazumdar, L. J. Ratliff, and S. S. Sastry, "On gradient-based learning in continuous games," *SIAM Journal on Mathematics of Data Science*, vol. 2, no. 1, pp. 103–131, 2020.

[29] M. Heusel, H. Ramsauer, T. Unterthiner, B. Nessler, and S. Hochreiter, "Gans trained by a two time-scale update rule converge to a local nash equilibrium," *Advances in neural information processing systems*, vol. 30, 2017.

[30] L. J. Ratliff, S. A. Burden, and S. S. Sastry, "On the characterization of local nash equilibria in continuous games," *IEEE transactions on automatic control*, vol. 61, no. 8, pp. 2301–2307, 2016.

[31] F. Facchinei and J.-S. Pang, *Finite-Dimensional Variational Inequalities and Complementarity Problems*. Springer Science & Business Media, 2007.

[32] S. P. Dirkse and M. C. Ferris, "The path solver: A non-monotone stabilization scheme for mixed complementarity problems," *Optimization methods and software*, vol. 5, no. 2, pp. 123–156, 1995.

[33] K. Mombaur, A. Truong, and J.-P. Laumond, "From human to humanoid locomotion—an inverse optimal control approach," *Autonomous robots*, vol. 28, pp. 369–383, 2010.

[34] P. Englert, N. A. Vien, and M. Toussaint, "Inverse kkt: Learning cost functions of manipulation tasks from demonstrations," *The International Journal of Robotics Research*, vol. 36, no. 13-14, pp. 1474–1488, 2017.

[35] L. Peters, V. Rubies-Royo, C. J. Tomlin, *et al.*, "Online and offline learning of player objectives from partial observations in dynamic games," *The International Journal of Robotics Research*, vol. 42, no. 10, pp. 917–937, 2023.

[36] A. L. Dontchev and R. T. Rockafellar, *Implicit functions and solution mappings*. Springer, 2009, vol. 543.