Cost Inference for Feedback Dynamic Games from Noisy Partial State Observations and Incomplete Trajectories

Jingqi Li
University of California, Berkeley
Berkeley, United States
jingqili@berkeley.edu

Chih-Yuan Chiu
University of California, Berkeley
Berkeley, United States
chihyuan_chiu@berkeley.edu

Lasse Peters
Delft University of Technology
Delft, Netherlands
l.peters@tudelft.nl

Somayeh Sojoudi
University of California, Berkeley
Berkeley, United States
sojoudi@berkeley.edu

Claire Tomlin
University of California, Berkeley
Berkeley, United States
tomlin@eecs.berkeley.edu

David Fridovich-Keil
University of Texas, Austin
Austin, United States
dfk@utexas.edu

ABSTRACT

In multi-agent dynamic games, the Nash equilibrium state trajectory of each agent is determined by its cost function and the information pattern of the game. However, the cost and trajectory of each agent may be unavailable to the other agents. Prior work on using partial observations to infer the costs in dynamic games assumes an open-loop information pattern. In this work, we demonstrate that the feedback Nash equilibrium concept is more expressive and encodes more complex behavior. It is desirable to develop specific tools for inferring players' objectives in feedback games. Therefore, we consider the dynamic game cost inference problem under the feedback information pattern, using only partial state observations and incomplete trajectory data. To this end, we first propose an inverse feedback game loss function, whose minimizer yields a feedback Nash equilibrium state trajectory closest to the observation data. We characterize the landscape and differentiability of the loss function. Given the difficulty of obtaining the exact gradient, our main contribution is an efficient gradient approximator, which enables a novel inverse feedback game solver that minimizes the loss using first-order optimization. In thorough empirical evaluations, we demonstrate that our algorithm converges reliably and has better robustness and generalization performance than the open-loop baseline method when the observation data reflects a group of players acting in a feedback Nash game.

KEYWORDS

Inverse Games; Dynamic Game Theory; Nash Equilibrium

ACM Reference Format:
Jingqi Li, Chih-Yuan Chiu, Lasse Peters, Somayeh Sojoudi, Claire Tomlin, and David Fridovich-Keil. 2023. Cost Inference for Feedback Dynamic Games from Noisy Partial State Observations and Incomplete Trajectories. In *Proc. of the 22nd International Conference on Autonomous Agents and Multiagent Systems (AAMAS 2023), London, United Kingdom, May 29 – June 2, 2023*, IFAAMAS, 9 pages.

1 INTRODUCTION

The safety and efficiency of urban traffic relies heavily on the ability of each participant to predict the effects of their actions on others' decisions [23, 33]. For example, drivers on a highway may wish to halt an overtaking maneuver if they believe the other drivers are aggressive, and some drivers may decelerate their cars to avoid collision if they believe that another driver wishes to merge.

A powerful paradigm for modeling the interdependence of decisions in multi-agent settings is provided *general-sum dynamic games* [2, 13]. A Nash equilibrium solution of a game-theoretic model can be used to *simultaneously* predict the actions of all agents in the scene. This equilibrium solution is particularly expressive when the game possesses a feedback information structure. In this case, each equilibrium strategy explicitly accounts for the dynamically evolving information available to each player over time.

Despite the theoretical attractiveness of this modeling paradigm, in reality, autonomous agents often have only limited information available about the world around them. For example, in urban traffic an autonomous agent typically has incomplete knowledge of the objectives of other players. To address this challenge, recent works on *inverse dynamic game theory* [21, 28, 31] recover these objectives from past trajectory data. Moreover, in realistic applications, only noisy sensor measurements of agents' states are available. This partial observability further complicates the inverse game problem, and existing work [28] treats this case in the open-loop information structure.

In this work, we present a gradient-based solver for inverse dynamic games, under the state feedback information structure. Our solver can recover objectives from partial state observations of incomplete trajectories. Both of these effects are common in robotics due to noisy perception and occlusions. We show that our algorithm converges reliably in practice, and demonstrate the superior robustness and generalization performance as compared with a baseline method which learns cost functions under the open-loop assumption [28], when the observation data is from a group of players pursuing a feedback Nash equilibrium strategy.

Our contributions are threefold. Firstly, we characterize the solution set of the inverse feedback dynamic game problem. In particular, we show that the set of the global minima could be nonconvex and disconnected, and discuss regularization schemes to mitigate this problem. Secondly, we show the differentiability of the loss function in linear quadratic games and propose a computationally efficient procedure to approximate the gradient for nonlinear games. Finally, we propose an efficient first-order coordinate-descent solver

Proc. of the 22nd International Conference on Autonomous Agents and Multiagent Systems (AAMAS 2023), A. Ricci, W. Yeoh, N. Agmon, B. An (eds.), May 29 – June 2, 2023, London, United Kingdom.

for the inverse feedback game problem, using noisy partial observations of an incomplete expert state trajectory. Experimental results show that our method reliably converges for inverse feedback games with nonlinear dynamics and is able to learn nonconvex costs. Moreover, the converged cost function can accurately predict the feedback Nash equilibrium state trajectories even for unseen initial states.

2 RELATED WORK

2.1 Non-cooperative Dynamic Games

Non-cooperative dynamic game theory [2, 13] provides a formal framework for analyzing strategic interaction in a multi-agent setting [2, 6, 17]. In non-cooperative games, each player minimizes its own individual cost function; since players' costs may not be mutually aligned, the resulting equilibrium behavior is generally competitive. Among different equilibrium concepts, the Nash equilibrium has been extensively studied because of its representative power of capturing many non-cooperative behaviors arising in real-world multi-agent systems [9, 32].

Recent advances in the literature aim to develop efficient solutions to Nash equilibrium problems in dynamic games. Though the solutions for the open-loop and feedback Nash equilibrium in linear quadratic (LQ) games are well understood [2], for nonlinear games there is no closed-form solution in general. The work [30] characterizes the local Nash solution concept for open-loop Nash equilibrium. In the feedback setting, numerous approaches have been proposed under various special cases [14, 35]. A value iteration based approach for computing feedback Nash equilibria of nonlinear games without constraints is introduced in [10]. Recently, a set of KKT conditions for feedback Nash equilibria in constrained nonlinear games is derived in [16]. Computing a feedback Nash equilibrium is challenging due to the nested KKT conditions in different time steps.

Our work draws upon the ILQGames [8] framework, which at each iteration solves a linear-quadratic game that approximates the original game. The construction of the approximate game parallels the iterative linearization and quadraticization methods of iterative LQR [18], and the dynamic programming equations that characterize equilibrium strategies in linear quadratic dynamic games [2]. This approach differs from the ALGames [5] method, which computes an open-loop Nash equilibrium strategy.

2.2 *Inverse* Non-cooperative Dynamic Games

In contrast to the forward game problem of computing a strategy in dynamic games, an inverse game problem amounts to finding objectives for all agents such that the corresponding strategic (e.g., Nash equilibrium) interactions reproduce expert demonstrations. The inverse game problem is important because it paves the way for an agent to understand the preferences which explain other agents' behavior, which may facilitate more efficient multi-agent interaction and coordination.

The problem of inverse infinite-horizon LQ games is considered in [12], where the set of cost functions whose feedback Nash equilibrium strategies coincide with an expert strategy is derived. In [31, 36], the two-player inverse LQ game is solved by transforming the problem to an inverse optimal control under the assumption that the control input data of one player is known. Two methods based on the KKT conditions of an open-loop Nash equilibrium are proposed for open-loop general-sum differential games in [24]. Several necessary conditions for open-loop Nash equilibria are proposed in [22] and used for developing an inverse game solution for some classes of open-loop games.

Recently, an efficient bilevel optimization framework [28] based on the open-loop Nash equilibrium KKT conditions was proposed for solving inverse games with an open-loop Nash assumption. Another line of work on inferring costs in open-loop games [1, 7, 11] proposes to minimize the residual violation of the KKT conditions. This KKT residual framework assumes the knowledge of complete trajectory data and is a convex problem. Given the difficulty of evaluating KKT conditions for feedback Nash equilibria in nonlinear games [16], the extension of the KKT residual method to feedback nonlinear games may be subject to numerical difficulty.

A bilevel optimization approach for inverse feedback game problem is proposed in [25], with the assumption that both the expert state and control trajectories are observed without noise. In addition, an inverse game solver is proposed in [20] where they infer the players' cost functions with the assumption that the expert strategy follows a new concept called Maximum Entropy Nash Equilibrium. To the best of the authors' knowledge, there is no work on inferring cost functions of nonlinear dynamic games under feedback Nash equilibrium condition, from noisy partial state observation and incomplete trajectory data.

3 PRELIMINARIES

Consider an N-player, T-stage, deterministic, discrete-time dynamic game, with a state $x_t^i \in \mathbb{R}^{n_i}$ and control input $u_t^i \in \mathbb{R}^{m_i}$ for each player $i \in [N] := \{1, \cdots, N\}$, $t \in [T]$. Let the dimension of the joint state and control input be $n := \sum_{i=1}^N n_i$ and $m := \sum_{i=1}^N m_i$, respectively. We denote by $x_t := [x_t^1, \ldots, x_t^N] \in \mathbb{R}^n$ and $u_t := [u_t^1, \ldots, u_t^N] \in \mathbb{R}^m$ the joint state and joint control at time $t \in [T]$, respectively. The joint dynamics for the system is given by the differentiable dynamics map $f_t(\cdot, \cdot) : \mathbb{R}^n \times \mathbb{R}^m \to \mathbb{R}^n$:

$$x_{t+1} = f_t(x_t, u_t), \qquad \forall\, t = 1, \cdots, T. \tag{1}$$

We denote by $\mathbf{f} := \{f_t\}_{t=1}^T$ the set of dynamics across all the time instances within horizon T. We define $\mathbf{x} := \{x_t\}_{t=1}^T$ and $\mathbf{u} := \{u_t\}_{t=1}^T$ to be a state trajectory and control trajectory, respectively, if $x_{t+1} = f(x_t, u_t)$, for each $t \in [T]$. The objective of each agent i is to minimize its overall cost, given by the sum of its running costs $g_t^i : \mathbb{R}^n \times \mathbb{R}^m \to \mathbb{R}$ over the time horizon:

$$J^i(\mathbf{x}, \mathbf{u}) := \sum_{t=1}^T g_t^i(x_t, u_t) \tag{2}$$

Define $g_t := \{g_t^1, g_t^2, \cdots, g_t^N\}$, $t \in [T]$. We denote by $\mathbf{g} := \{g_t\}_{t=1}^T$ the set of cost functions for all the agents within horizon T.

To minimize (2), each player uses their observations of the environment to design a sequence of control inputs to deploy during the discrete time interval $[T]$. The information available to each player at each time characterizes the *information pattern* of the dynamic game, and plays a major role in shaping the optimal responses of each player [2]. Below, we explore two such information patterns—*feedback* and *open-loop*.

3.1 Nash Solutions in Feedback Strategies

Under the state feedback information pattern, each player observes the state x_t at each time t, and uses this information to design a *feedback strategy* $\gamma_t^i : \mathbb{R}^n \to \mathbb{R}^{m_i}$, given by: $u_t^i := \gamma_t^i(x_t)$, for each $i \in [N]$ and $t \in [T]$. Let $\gamma_t(x_t) := [\gamma_t^1(x_t), \gamma_t^2(x_t), \ldots, \gamma_t^N(x_t)] \in \mathbb{R}^m$.

Following the notation of [2], we denote by Γ_t^i the set of all state feedback strategies of player i, for each $i \in [N]$. Under this *feedback* information pattern, the Nash equilibrium of the dynamic game is as defined below.

Definition 1 (**Feedback Nash Equilibrium (FBNE)** [2, Ch. 6]). *The set of control strategies $\{\gamma_t^{1*}, \cdots, \gamma_t^{N*}\}_{t=1}^T$ is called a feedback Nash equilibrium if no player is incentivized to unilaterally alter its strategy. Formally:*

$$W_t^{i*}\Big(x_t, [\gamma_t^{1*}(x_t), \ldots, \gamma_t^{i*}(x_t), \ldots, \gamma_t^{N*}(x_t)]\Big) \tag{3}$$
$$\leq W_t^{i*}\Big(x_t, [\gamma_t^{1*}(x_t), \ldots, \gamma_t^{i}(x_t), \ldots, \gamma_t^{N*}(x_t)]\Big), \forall \gamma_t^i \in \Gamma_t^i, \forall t \in [T].$$

where $W_t^{i*}(\cdot,\cdot) : \mathbb{R}^n \times \mathbb{R}^m \to \mathbb{R}$, $t \in [T]$ is the optimal state-action function defined as follows,

$$\begin{aligned} W_T^{i*}(x_T, u_T) &:= g_T^i(x_T, u_T) \\ W_t^{i*}(x_t, u_t) &:= g_t^i(x_t, u_t) + V_{t+1}^{i*}(x_{t+1}), \forall t \in [T-1], \\ V_t^{i*}(x_t) &:= W_t^{i*}(x_t, [\gamma_t^{1^*}(x_t), \ldots, \gamma_t^{N^*}(x_t)]), \forall t \in [T]. \end{aligned} \tag{4}$$

We define $\mathbf{x}$ and $\mathbf{u}$ to be a FBNE state trajectory and a FBNE control trajectory, respectively, if $u_t^i = \gamma_t^{i*}(x_t)$, for each $i \in [N]$ and $t \in [T]$. We denote by $\xi(\mathbf{f}, \mathbf{g})$ the set of all FBNE state trajectories in the game defined by the dynamics $\mathbf{f}$ and cost functions $\mathbf{g}$.

Remark 1 (Strong Time Consistency). *The FBNE conditions of (3) implicitly enforce strong time-consistency [2, Def. 5.14] of the equilibrium strategies. That is, FBNE does not admit arbitrary feedback strategies, but imposes the additional condition that those strategies must also be in equilibrium for any subgame starting at a later stage from an arbitrary state.*

3.2 Nash Solutions in Open-loop Strategies

In contrast, under the open-loop information pattern, each player only observes the initial state x_1. In this case, the strategy for each player $i \in [N]$ is a map from x_1 to $\{u_1^i, u_2^i, \cdots, u_T^i\}$, which we denote by $\phi^i(\cdot) : \mathbb{R}^n \to \underbrace{\mathbb{R}^{m_i} \times \cdots \times \mathbb{R}^{m_i}}_{T}$. Let Φ^i be the set of all open-loop strategies of the player i, $i \in [N]$. The corresponding *open-loop Nash equilibrium* is defined as follows.

Definition 2 (**Open-Loop Nash Equilibrium (OLNE)** [2, Ch. 6]). *The tuple of control strategies $\{\phi_1^*, \cdots, \phi_N^*\}$ is called an open-loop Nash equilibrium if no player is incentivized to unilaterally alter its sequence of control inputs. Formally:*

$$J^i\Big(\mathbf{x}, [\phi^{1^*}(x_1), \cdots, \phi^{i^*}(x_1), \cdots, \phi^{N^*}(x_1)]\Big) \tag{5}$$
$$\leq J^i\Big(\mathbf{x}, [\phi^{1^*}(x_1), \cdots, \phi^{i}(x_1), \cdots, \phi^{N^*}(x_1)]\Big), \forall \phi^i \in \Phi^i, \forall x_1 \in \mathbb{R}^n.$$

Remark 2. *The OLNE definition does not imply the strong time-consistence as in the feedback counterpart [2].*

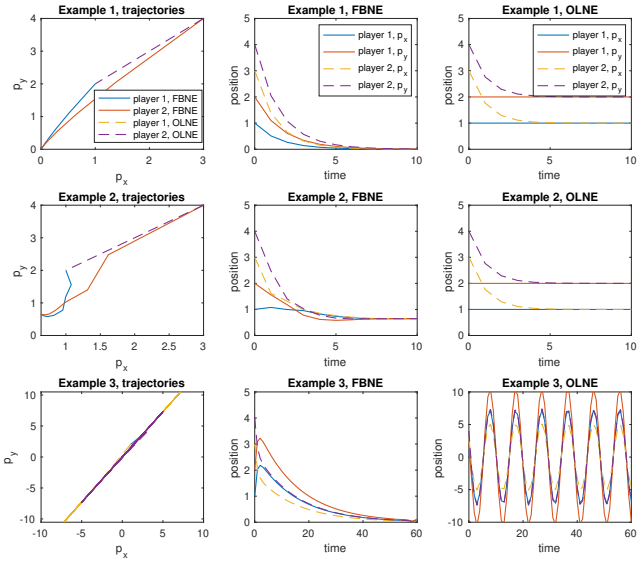

Fig. 1: Examples of cost functions that yield trajectories that are different under the OLNE and FBNE assumptions.

3.3 Feedback vs. Open-loop Nash Equilibria

In this subsection, we demonstrate the difference between open-loop and feedback Nash equilibria and show the necessity of developing specific solutions for cost inference problems with the feedback information pattern, instead of applying existing work with the open-loop assumption [29]. To this end, we introduce below several linear-quadratic (LQ) games where the open-loop Nash equilibrium (OLNE) and feedback Nash equilibrium (FBNE) state trajectories differ substantially. LQ games are a class of dynamic games with dynamics and player objectives of the form in (6) and (7), respectively,

$$x_{t+1} = A_t x_t + \sum_{i \in [N]} B_t^i u_t^i, \ \forall t \in [T], \tag{6}$$

$$g_t^i(x_t, u_t) = \frac{1}{2}(x_t^\top Q_t^i x_t + \sum_{j \in [N]} u_t^{j\top} R_t^{ij} u_t^j), \forall t \in [T], \forall i \in [N], \tag{7}$$

where matrices $\{A_t, B_t^i\}$, positive semidefinite matrix Q_t^i and positive definite matrix R_t^{ij} are defined with appropriate dimensions, for each $i, j \in [N]$ and $t \in [T]$.

Case Study: We consider a two-player LQ game with a state vector $x_t = [p_{x,t}^1, p_{y,t}^1, p_{x,t}^2, p_{y,t}^2]$, where $p_{x,t}^i$ and $p_{y,t}^i$ are the x- and y-coordinates of agent $i \in \{1, 2\}$, respectively. Let $u_t^i \in \mathbb{R}^2$ be the control input for the i-th agent, $i \in \{1, 2\}$. In this setting, we consider a class of games in which the first agent wants to drive the second agent to the origin, while the second agent wants to catch the first agent. The agents' joint dynamics and costs at time $t \in [T]$ are specified as follows:

$$\begin{aligned} x_{t+1} &= \begin{bmatrix} I_2 & 0 \\ 0 & I_2 \end{bmatrix} x_t + \begin{bmatrix} I_2 \\ 0 \end{bmatrix} u_t^1 + \begin{bmatrix} 0 \\ I_2 \end{bmatrix} u_t^2, \\ g_t^1(x_t, u_t) &= \|p_{x,t}^2\|_2^2 + \|p_{y,t}^2\|_2^2 + \|u_t^1\|_2^2, \\ g_t^2(x_t, u_t) &= \|p_{x,t}^2 - p_{x,t}^1\|_2^2 + \|p_{y,t}^2 - p_{y,t}^1\|_2^2 + \|u_t^2\|_2^2, \end{aligned} \tag{8}$$

where I_2 is the 2×2 identity matrix. We visualize the unique FBNE and OLNE state trajectories of this example in the first row in Fig. 1. If we modify the cost function of the first player such that it wants to lead the x- and y-position of the second player to be aligned with each other, i.e.,

$$\hat{g}_t^1(x_t, u_t) := \|p_{x,t}^2 - p_{y,t}^2\|_2^2 + \|u_t^1\|_2^2, \tag{9}$$

then, the unique FBNE and OLNE state trajectories are still different, as shown in the second row of Fig. 1. Moreover, observations of players may be noisy in practice. To illustrate this, we consider a task where the two agents want to catch each other, but the first player's observation of the second player's position is inaccurate. We modify the first player's cost in (8) as follows:

$$\hat{\hat{g}}_t^1(x_t, u_t) := \|p_{x,t}^1 - 2p_{x,t}^2\|_2^2 + \|p_{y,t}^1 - 2p_{y,t}^2\|_2^2 + \|u_t^1\|_2^2. \tag{10}$$

The third row of Fig. 1 reveals that the FBNE state trajectory is robust to inaccurate observations, but the unique OLNE state trajectory is not.

Thus, it is readily apparent that the OLNE and FBNE state strategies can be substantially different even for fixed cost functions. This difference in expressive power may be understood as a consequence of the strong time consistency property, which is enforced in the feedback information structure but not in the open-loop setting, per Remarks 1 and 2. A similar problem arises in the cost inference problem, where the existing OLNE cost inference algorithms may fail to infer the correct cost function in feedback games.

4 PROBLEM STATEMENT

Let $\mathbf{x}$ be an expert FBNE state trajectory under the nonlinear dynamics $\mathbf{f}$ but unknown cost functions $\{g_t^i\}_{t=1,i=1}^{T,N}$. Let $\mathcal{T} \subseteq [T]$ be the set of observed time indices of the trajectory $\mathbf{x}$. We denote by $\mathbf{y}_\mathcal{T} := \{y_t\}_{t\in\mathcal{T}}$ the observation data of $\mathbf{x}$, where $y_t \in \mathbb{R}^\ell$ is a partial observation of the state, composed of certain coordinates of x_t corrupted by noise. The task is to infer the cost function of each player such that those inferred costs jointly yield a FBNE state trajectory that is as close as possible to the observed trajectory. We parameterize the cost of the player i by a vector $\theta^i \in \mathbb{R}^{d_i}$, and let $\theta := [\theta^1, \theta^2, \ldots, \theta^N] \in \mathbb{R}^d$. Denote by $g_{t,\theta}^i(x_t, u_t) = \sum_{j=1}^{d_i} \theta_j^i b_{t,j}^i(x_t, u_t)$ player i's parameterized cost at time $t \in [T]$, for some basis functions $\{\{b_{t,j}^i\}_{j=1}^{d_i}\}_{t=1,i=1}^{T,N}$. Define $\mathbf{g}_\theta := \{g_{t,\theta}^i\}_{t=1,i=1}^{T,N}$. Formally, this problem is of the form:

$$\begin{aligned} \min_{\theta, x_1, \hat{\mathbf{x}}} \quad & -p(\mathbf{y}_\mathcal{T}|\hat{\mathbf{x}}) \\ \text{s.t.} \quad & \hat{\mathbf{x}} \in \xi(\mathbf{f}, \mathbf{g}_\theta, x_1), \end{aligned} \tag{11}$$

where $p(\cdot|\cdot)$ is the likelihood function corresponding to a known sensor model and $\xi(\mathbf{f}, \mathbf{g}_\theta, x_1)$ represents the set of state trajectories from the initial condition $x_1 \in \mathbb{R}^n$ following a FBNE strategy, under the cost set $\mathbf{g}_\theta$. Due to the noisy partial observation, x_1 is not assumed to be known and instead needs to be inferred as well in (11). Note that the above formulation can also be extended to the cases where multiple partially observed incomplete trajectories from different initial conditions are available.

Running example: We consider a highway platooning scenario where player 1 wants to guide player 2 to a particular lane of the road. The joint state vector is $x_t = [p_{x,t}^1, p_{y,t}^1, \beta_t^1, v_t^1, p_{x,t}^2, p_{y,t}^2, \beta_t^2, v_t^2]$.

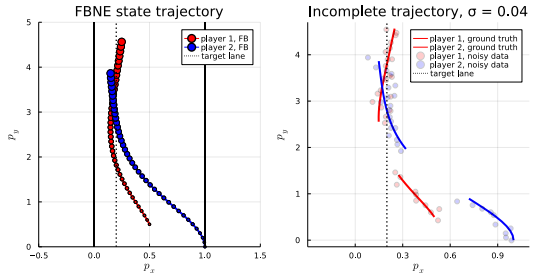

Fig. 2: Visualization of the running example.

The time horizon $T = 40$. The dynamics model for the player i is:

$$\begin{bmatrix} p_{x,t+1}^i \\ p_{y,t+1}^i \\ \beta_{t+1}^i \\ v_{t+1}^i \end{bmatrix} = \begin{bmatrix} p_{x,t}^i \\ p_{y,t}^i \\ \beta_t^i \\ v_t^i \end{bmatrix} + \Delta T \begin{bmatrix} v_t^i \cos(\beta_t^i) \\ v_t^i \sin(\beta_t^i) \\ \omega_t^i \\ a_t^i \end{bmatrix} \tag{12}$$

where ΔT is a time discretization constant and $u_t^i = [\omega_t^i, a_t^i] \in \mathbb{R}^2$ is the control input for player $i \in [N]$. Let p_x^* be the target lane that player 1 wants to guide player 2 to. We parameterize the cost function of the player i by $\theta^i \in \mathbb{R}^2$,

$$\begin{aligned} g_{t,\theta}^1(x_t, u_t) &= \theta_1^1 \|p_{x,t}^1\|_2^2 + \theta_2^1 \|p_{x,t}^2 - p_x^*\|_2^2 + \|u_t^1\|_2^2 \\ g_{t,\theta}^2(x_t, u_t) &= \theta_1^2 \|p_{x,t}^2 - p_{x,t}^1\|_2^2 + \theta_2^2 \|v_t^2 - 1\|_2^2 + \|u_t^2\|_2^2, \forall t \in [T]. \end{aligned} \tag{13}$$

The ground truth solution is $\theta^* = [0, 8, 4, 4]$. We assume that there is a period of occlusion happening from the time index $t = 11$ to $t = 19$, and the observed time index set is $\mathcal{T} = \{1, 2, \ldots, 10, 20, 21, \ldots, 40\}$. Also, it may be difficult for a human driver to measure other vehicles' velocity accurately, and therefore we assume that partial observation data $\mathbf{y}_\mathcal{T}$ excludes the velocity of both cars in the data set, and is further subject to Gaussian noise of standard deviation σ. The initial condition x_1 is not known and needs to be inferred. We visualize the ground truth solution in the first subplot of Fig. 2 and the noisy incomplete trajectory data in the second subplot of Fig. 2.

The many challenges of the above problem include: (a) partial observation; (b) noisy and incomplete expert trajectory data; and (c) the difficulty of evaluating and differentiating the objective in (11), due to the challenge of computing a FBNE strategy in nonlinear games [16]. In the following sections, we will characterize the complexity of this inverse feedback game problem and propose an efficient solution.

5 RESULTS: FROM CHARACTERIZATION TO COMPUTATION

In this section, we first characterize the complexity of the inverse feedback game problem (11). In particular, we will show the nonconvexity of the loss function and the existence of multiple isolated *global* minima. Based on this observation, we discuss regularization schemes that can mitigate this issue. Our main contribution is to characterize the differentiability of the inverse feedback game loss function in (11). Finally, we present a gradient approximation scheme that can be used in a first-order optimization formulation.

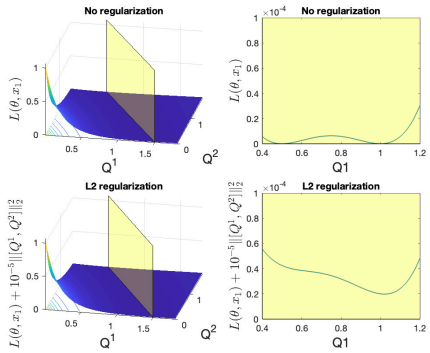

Fig. 3: Visualization of the loss function $L(\theta, x_1)$ of the LQ game specified in (16) and (17), and its L_2 regularization, with an initial condition $x_1 = 1$. We adopt Gaussian likelihood function. The yellow hyperplane is drawn according to $2Q^1+Q^2 = 3$. With L_2 regularization, the number of global minima is reduced.

5.1 Characterization of the Inverse Feedback Dynamic Game Problem

The inverse feedback dynamic game problem (11) is a constrained optimization problem, which is hard to solve due to the nonconvexity of the set $\xi(\mathbf{f}, \mathbf{g}_\theta, x_1)$. With a slight abuse of notation, we denote by $\hat{\mathbf{x}}(\mathbf{f}, \mathbf{g}_\theta, x_1) \in \xi(\mathbf{f}, \mathbf{g}_\theta, x_1)$ a FBNE state trajectory. To simplify the problem, we transform (11) to an unconstrained problem by substituting a forward game solution $\hat{\mathbf{x}}(\mathbf{f}, \mathbf{g}_\theta, x_1)$ into the likelihood function $p(\mathbf{y}_\mathcal{T}|\hat{\mathbf{x}})$, as follows:

$$\hat{L}(\theta, x_1) := -p(\mathbf{y}_\mathcal{T}|\hat{\mathbf{x}}(\mathbf{f}, \mathbf{g}_\theta, x_1)). \tag{14}$$

The minimizer of (14) is a local optimum to the original problem (11) and becomes global when $\xi(\mathbf{f}, \mathbf{g}_\theta, x_1)$ contains only a single element.

Before we dive into the nonlinear setting, let us first consider a simplified LQ case to highlight the main challenges associated with the optimization of this loss. In the LQ case, the *evaluation* of the loss (14) is straightforward if there exists a closed-form expression for $p(\mathbf{y}_\mathcal{T}|\hat{\mathbf{x}})$, e.g., under a Gaussian observation model. Even in that setting, however, it is important to realize that the problem remains nonconvex, as shown in Fig. 3. The following proposition makes this challenge explicit, and the proof can be found in the Appendix.

Proposition 1. *There exists an inverse LQ game problem* (11)*: (a) whose global minima are isolated, and (b) for which there exist multiple cost functions that exactly match expert data from any initial condition, when there is no observation noise.*

Remark 3. *Proposition 1 does not imply that any inverse LQ game problem will suffer from the multiple global minima issue. Instead, Proposition 1 suggests that simply normalizing the cost vector does not rule out the possibility of having multiple global solutions. That is, there exist two cost parameter vectors which are linearly independent, but generate the same FBNE state trajectories for any given initial state. This non-injective mapping from the cost parameter space to the FBNE state trajectory space is a fundamental problem in inverse feedback games, and is not particular to the formulation* (11)*. In practice, this multiple global minima issue could be mitigated by adding L_2 regularization, as visualized in Fig. 3.*

Though being nonconvex, the loss function $\hat{L}(\theta, x_1)$ is differentiable with respect to both θ and x_1 under the condition of Theorem 3.2 in [16], which follows from the implicit function theorem [15]. Inspired by the success of gradient-based methods in non-convex optimization with differentiable objective functions [3, 26, 34], one natural idea is to apply gradient descent to minimize $\hat{L}(\theta, x_1)$. In what follows, we discuss efficient ways to evaluate and differentiate $\hat{L}(\theta, x_1)$ in nonlinear games.

5.2 Efficient Computation for a FBNE State Trajectory in Nonlinear Games

It is easy to evaluate $\hat{L}(\theta, x_1)$ for LQ games, but when dynamics are nonlinear or objectives are non-quadratic, the problem becomes more challenging [16]. In forward games, this challenge can be addressed by using the ILQGames algorithm [8], which finds approximate local FBNE solutions in smooth non-LQ dynamic games. Given the effectiveness of this approximation scheme in those domains, we also adopt it as a submodule for evaluating the loss $\hat{L}(\theta, x_1)$. Akin to the ILQR method [18, 19], in each step of the ILQGames algorithm, the system dynamics $x_{t+1} = f(x_t, u_t)$ and the costs $\{g^i_t(x, u)\}^{T,N}_{t=1,i=1}$ are linearized and quadraticized, respectively, around a state trajectory $\mathbf{x}$ and a control trajectory $\mathbf{u}$. A FBNE strategy for each player of the derived LQ game is then used to update the state and control trajectories. This iteration continues until a convergence criterion is satisfied.

To be more specific, we approximate $\hat{L}(\theta, x_1)$ by a new loss function $\tilde{L}(\theta, x_1)$ defined as,

$$\hat{L}(\theta, x_1) \simeq \tilde{L}(\theta, x_1) := -p\big(\mathbf{y}_\mathcal{T}|\mathbf{x}(\tilde{\mathbf{f}}_\theta, \tilde{\mathbf{g}}_\theta, x_1)\big) \tag{15}$$

where $\mathbf{x}(\tilde{\mathbf{f}}_\theta, \tilde{\mathbf{g}}_\theta, x_1)$ represents a FBNE state trajectory from initial condition x_1, for the LQ game defined by the linearized dynamics $\tilde{\mathbf{f}}_\theta$, quadraticized cost set $\tilde{\mathbf{g}}_\theta := \{\tilde{\mathbf{g}}^i_{t,\theta}\}^{T,N}_{t=1,i=1}$ at the converged solution returned by ILQGames solver. Note that the linearized dynamics $\tilde{\mathbf{f}}_\theta$ depend upon θ via the state trajectory about which $\mathbf{f}$ is linearized; this trajectory is simulated under the feedback policy returned by ILQGames, where the policy depends upon costs $\mathbf{g}_\theta$.

5.3 Differentiating the Loss in the Inverse Feedback Game Problem

The challenge of computing a feedback Nash equilibrium strategy not only makes the evaluation of the loss function $\hat{L}(\theta, x_1)$ hard, but also renders differentiation difficult. In this work, we approximate the gradient of $\hat{L}(\theta, x_1)$ using a similar idea as the ILQGames algorithm in the previous section. In other words, we propose to use the LQ approximation of the nonlinear game specified by $\tilde{\mathbf{f}}_\theta$ and $\tilde{\mathbf{g}}_\theta$ to derive an approximation to the gradient of $\hat{L}(\theta, x_1)$. Note that $\tilde{g}^i_{t,\theta}(x, u) = \sum_{j=1}^{d_i} \theta^i_j \tilde{b}^i_{t,j,\theta}(x, u)$, where $\tilde{b}^i_{t,j,\theta}(x, u) : \mathbb{R}^n \times \mathbb{R}^m \to \mathbb{R}$ is the j-th quadraticized cost basis function. By the chain rule, we have

$$\frac{\partial \tilde{L}(\theta, x_1)}{\partial \theta^i_j} = -\nabla_{\mathbf{x}} p(\mathbf{y}_\mathcal{T}|\mathbf{x})\Big|_{\mathbf{x}(\tilde{\mathbf{f}}_\theta, \tilde{\mathbf{g}}_\theta, x_1)} \cdot \frac{\partial \mathbf{x}(\tilde{\mathbf{f}}_\theta, \tilde{\mathbf{g}}_\theta, x_1)}{\partial \theta^i_j},$$

$$\frac{\partial \mathbf{x}(\tilde{\mathbf{f}}_\theta, \tilde{\mathbf{g}}_\theta, x_1)}{\partial \theta^i_j} = \Big(\nabla_{\tilde{\mathbf{f}}_\theta} \mathbf{x}(\tilde{\mathbf{f}}_\theta, \tilde{\mathbf{g}}_\theta, x_1) \frac{\partial \tilde{\mathbf{f}}_\theta}{\partial \theta^i_j} + \nabla_{\tilde{\mathbf{g}}_\theta} \mathbf{x}(\tilde{\mathbf{f}}_\theta, \tilde{\mathbf{g}}_\theta, x_1) \frac{\partial \tilde{\mathbf{g}}_\theta}{\partial \theta^i_j}\Big).$$

The complexity of differentiating $\tilde{L}(\theta, x_1)$ comes from the fact that the linearized dynamics and the quadraticized costs are functions of θ implicitly, which makes the total derivative hard to compute. We propose to approximate the above gradient by treating the linearized $\tilde{\mathbf{f}}_\theta$ and each quadraticized cost basis function $\tilde{b}^i_{t,j,\theta}$ as constants with respect to θ, denoted by $\tilde{\mathbf{f}}$ and $\tilde{b}^i_{t,j}$, and only compute the partial derivative with respect to θ, rather than the total derivative:

$$\frac{\partial \tilde{L}(\theta, x_1)}{\partial \theta^i_j} \simeq -\nabla_{\mathbf{x}} p(\mathbf{y}_{\mathcal{T}}|\mathbf{x})\Big|_{\mathbf{x}(\tilde{\mathbf{f}}, \tilde{\mathbf{g}}_\theta, x_1)} \cdot \frac{\partial \mathbf{x}(\tilde{\mathbf{f}}, \{\sum_{j=1}^{d_i} \theta^i_j \tilde{b}^i_{t,j}\}^{T,N}_{t=1,i=1}, x_1)}{\partial \theta^i_j}.$$

This is based on the observation that at the convergence of the forward ILQGames solver, the linearized dynamics are a good approximation of the full nonlinear dynamics $\mathbf{f}$, so long as the cost parameter being perturbed remains sufficiently small. We adopt a similar approximation for the gradient $\nabla_{x_1}\tilde{L}(\theta, x_1)$ by fixing the linearized dynamics and quadraticized costs and obtaining the partial derivative with respect to x_1.

In summary, we approximate $\nabla\hat{L}(\theta, x_1)$ by $\nabla\tilde{L}(\theta, x_1)$. In practice, $\nabla\tilde{L}(\theta, x_1)$ can be efficiently computed by automatic differentiation [27, Ch. 8]. As exemplified in Fig. 4, the proposed gradient approximation is virtually always a descent direction and therefore aligns well with the true gradient of $\hat{L}(\theta, x_1)$.

5.4 An Inverse Feedback Game Solver

In this subsection, we present a solver for the inverse feedback game problem (11). In what follows, we first discuss how the three challenges mentioned in Section 4 are handled in our solver. We then introduce the proposed solver in Algorithm 1.

The first two challenges on noisy partial observation and incomplete trajectory data are handled by maintaining an estimate of the full initial condition and a noise-free state-input trajectory. As shown in Section 6, this procedure of joint reconstruction and filtering enables our solver to reliably recover player costs even in scenarios of substantial partial observability. The third difficulty of evaluating and differentiating the objective function in the inverse *feedback* game problem is mitigated by the efficient approximation outlined in Section 5.3. To jointly infer the initial condition, the cost and the state-input trajectory, we first adopt the coordinate gradient descent method, where gradient descent steps are first taken over the initial condition $\hat{x}_1$, and then taken over the cost parameter. We update the estimate of the noise-free full state-input trajectory by computing a FBNE state trajectory from the inferred initial condition and the cost.

We summarize our proposed solver in Algorithm 1. At the k-th iteration, we first compute an approximate FBNE state trajectory $\tilde{x}^{(k)}$ and the associated LQ approximation via the ILQGames algorithm of [8]. Using this LQ approximation, we estimate $\nabla_{x_1}\hat{L}(\theta, x_1^{(k)})$ using the procedure outlined in Section 5.3. We then update the initial condition $x_1^{(k)}$ by a step of gradient descent, where the stepsize is chosen by a suitable linesearch technique [27, Ch. 3] such that the loss $\hat{L}(\theta, x_1)$ is sufficiently decreased. Given the updated initial condition $x_1^{(k+1)}$, we find a new approximate FBNE state trajectory via the ILQGames algorithm again, which is then used to estimate $\nabla_\theta\hat{L}(\theta^{(k)}, x_1^{(k+1)})$ via the procedure in Section 5.3. With this gradient, we update $\theta^{(k)}$ by one step of gradient descent with linesearch. We repeat this procedure until, at convergence, we find a locally optimal solution $(\hat{\theta}, \hat{x}_1)$.

Algorithm 1: Inverse Iterative LQ (i^{2}LQ) Games

Data: Horizon $T > 0$, initial solution $\theta^{(0)} \in \mathbb{R}^d$, observed time index set $\mathcal{T} \subseteq [T]$, observation data $\mathbf{y}_{\mathcal{T}}$, max iteration number K , tolerance ϵ.
Result: Inferred cost parameter $\hat{\theta}$ and $\hat{x}_1$

for $k = 0, 1, \ldots, K$ **do**
- $(\tilde{\mathbf{x}}^{(k)}, \{\tilde{\gamma}^i_t\}^{T,N}_{t=1,i=1}, \tilde{\mathbf{f}}_{\theta^{(k)}}, \tilde{\mathbf{g}}_{\theta^{(k)}}) \leftarrow \text{ILQGames}(\mathbf{f}, \mathbf{g}_{\theta^{(k)}}, x_1^{(k)})$
- $\nabla_{x_1}\hat{L}(\theta^{(k)}, x_1^{(k)}) \leftarrow$ evaluated using $\tilde{\mathbf{f}}_{\theta^{(k)}}$ and $\tilde{\mathbf{g}}_{\theta^{(k)}}$ via Gradient Approximation in Section 5.3
- $x_1^{(k+1)} \leftarrow x_1^{(k)} - \eta\nabla_{x_1}\hat{L}(\theta^{(k)}, x_1^{(k)})$ with line search over η
- $(\check{x}^{(k)}, \{\check{\gamma}^i_t\}^{T,N}_{t=1,i=1}, \check{\mathbf{f}}_{\theta^{(k)}}, \check{\mathbf{g}}_{\theta^{(k)}}) \leftarrow \text{ILQGames}(\mathbf{f}, \mathbf{g}_{\theta^{(k)}}, x_1^{(k+1)})$
- $\nabla_\theta\hat{L}(\theta^{(k)}, x_1^{(k+1)}) \leftarrow$ evaluated using $\check{\mathbf{f}}_{\theta^{(k)}}$ and $\check{\mathbf{g}}_{\theta^{(k)}}$ via Gradient Approximation in Section 5.3
- $\theta^{(k+1)} \leftarrow \theta^{(k)} - \eta'\nabla_\theta\hat{L}(\theta^{(k)}, x_1^{(k+1)})$ with line search over η'
- **Return** $(\theta^{(k+1)}, x_1^{(k+1)})$ if $\|\theta^{(k)} - \theta^{(k-1)}\|_2 \le \epsilon$ or **Return** $(\theta^{(k')}, x_1^{(k')})$, where $k' \leftarrow \arg\min_k \tilde{L}(\theta^{(k)}, x_1^{(k)})$, if iteration number k reaches K.

end

6 EXPERIMENTS

In this section, we adopt the open-loop solution method of [28] as the baseline method and compare it to Algorithm 1. The experiment codes can be found in https://github.com/jamesjingqili/inverse-iLQGames.git. In particular, we evaluate Algorithm 1 in several Monte Carlo studies which aim to justify the following claims.

- The proposed gradient approximation often aligns with a descent direction in the loss function.
- Algorithm 1 is more robust than the open-loop baseline method [28] with respect to noise in, and incomplete observations of, the expert demonstration trajectory.
- The cost functions inferred by Algorithm 1 can be generalized to predict trajectories from unseen initial conditions.
- Algorithm 1 can infer nonconvex costs in nonlinear games.

6.1 Gradient Approximation Quality

We continue the 2-vehicle platooning example defined in (12) and (13). We measure the performance of Algorithm 1 in two settings, incomplete expert trajectory data with noisy partial state observation, and complete expert trajectory data with noisy full observation. In the first case, each player's partial observation only contains its x-position, y-position and heading angle. The time index set of the incomplete trajectory is $\mathcal{T} = [T] \setminus \{11, 12, \ldots, 19\}$. In the second case, the expert data includes the noisy observation of all the states of both players at all $t \in [T]$. The ground truth expert state trajectory follows a FBNE strategy from the initial condition $x_1 = [0, 0.5, \frac{\pi}{2}, 1, 1, 0, \frac{\pi}{2}, 1]$ and the target lane is $p_x^* = 0.0$. At each variance level $\sigma \in \{0.004, 0.008, \ldots, 0.04\}$, we generate 10 noisy observations of the ground truth expert trajectory, with

isotropic zero-mean Gaussian noise. For each noisy expert data set $\mathbf{y}_{\mathcal{T}}$, we minimize the negative log-likelihood objective in (11), i.e., $\sum_{t\in\mathcal{T}} \|y_t - h(x_t)\|_2^2$, where $h(\cdot) : \mathbb{R}^n \to \mathbb{R}^\ell$ maps a state x_t to its partial observation.

As shown in Fig. 4, the loss decreases monotonically on the average. This indicates that the gradient approximation proposed in Section 5.3 provides a reliable descent direction. The inverse feedback game problem becomes challenging when there is only partial state observation and incomplete trajectory data, and the quality of inferred costs may degrade when the observation noise is high.

6.2 Robustness, Generalization and the Ability to Infer Nonconvex Costs

We continue the previous 2-vehicle example and compare Algorithm 1 and the baseline in a Monte Carlo study, where we infer the costs under 10 different levels of Gaussian noise with increasing variance. In particular, we evaluate three metrics in Fig. 5: (a) the distance between the noisy expert data and the FBNE state trajectory which results from players' inferred costs; (b) the distance between the computed FBNE state trajectory (under the players' inferred costs) and the ground truth expert data. An example of such a comparison is shown in Fig. 6. Finally, we evaluate (c) the distance between the inferred FBNE state trajectories and the FBNE state trajectory under the ground truth costs for some randomly sampled initial conditions, which is also visualized in Fig. 7. Collectively, the results demonstrate that *Algorithm 1 has better robustness and generalization performance than the open-loop baseline when the expert data follows the FBNE assumption.*

To show that Algorithm 1 can infer nonconvex cost functions, we extend the previous 2-vehicle platooning example and assume that the 2-vehicle team encounters a third vehicle and the follower wants to stay close to the leader without colliding with the third vehicle. We model this scenario as a 3-vehicle game with a 12 dimensional state space and a horizon $T = 30$. The dynamics for each vehicle is the same as (12) and the costs are as follows,

$$\begin{aligned} g^1_{t,\theta}(x_t, u_t) =& \theta^1_1 \|p^1_{x,t}\|_2^2 + \theta^1_2 \|p^2_{x,t} - p^*_x\|_2^2 + \|v^1_t - 2\|_2^2 \\ &+ \|\beta^1_t - \frac{\pi}{2}\|_2^2 + \|u^1_t\|_2^2 \\ g^2_{t,\theta}(x_t, u_t) =& \theta^2_1 \|p^2_{x,t}\|_2^2 + \|\beta^2_t - \frac{\pi}{2}\|_2^2 + \theta^2_2 \|p^2_{x,t} - p^1_{x,t}\|_2^2 + \|v^2_t - 2\|_2^2 \\ &- \frac{1}{2}\log(\|p^2_{x,t} - p^3_{x,t}\|_2^2 + \|p^2_{y,t} - p^3_{y,t}\|_2^2) + \|u^2_t\|_2^2 \\ g^3_{t,\theta}(x_t, u_t) =& \theta^3_1 \|p^3_{x,t} - \frac{1}{2}\|_2^2 + \|u^3_t\|_2^2 \end{aligned}$$

where the ground truth $\theta^* \in \mathbb{R}^5$ is $[0, 4, 0, 4, 2]$. The ground truth expert state trajectory follows a FBNE strategy from the initial condition $x_1 = [0, 1, \frac{\pi}{2}, 2, 0.3, 0, \frac{\pi}{2}, 2, 0.5, 0.5, \frac{\pi}{2}, 2]$, where the last four elements encode the state of the third vehicle. The target lane in the expert data is $p^*_x = 0.2$.

Similar to the 2-vehicle experiment, we consider two settings, incomplete trajectory data with partial state observation and complete trajectory data with full state observation. The partial state observation includes all the states of each vehicle except for the velocity of all the vehicles, and the time indices set of the incomplete trajectory is $\mathcal{T} = [T] \setminus \{11, 12, \ldots, 19\}$. The nonconvex cost

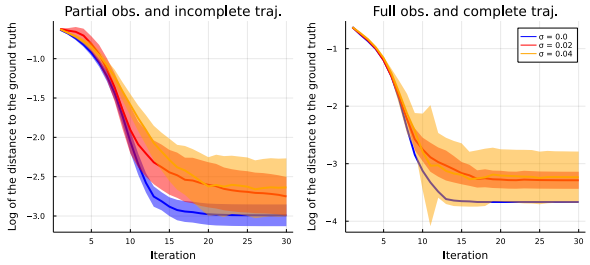

Fig. 4: Convergence of Algorithm 1 with the Gradient Approximation proposed in Section 5.3. The loss decreases monotonically on the average. The bold lines and shaded areas represent the mean values and their standard error, i.e., the variance divided by the square root of the sample size, respectively.

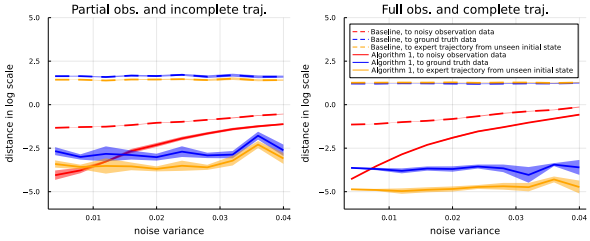

Fig. 5: 2-vehicle platooning scenario. The bold lines and shaded areas represent the mean values and their standard error, i.e., the variance divided by the square root of the sample size, respectively. As the noise variance growing, the converged loss value increases, as shown in the red curves. However, Algorithm 1 is still able to learn a more accurate cost and has less generalization error than the baseline, as shown in the blue and yellow curves, respectively.

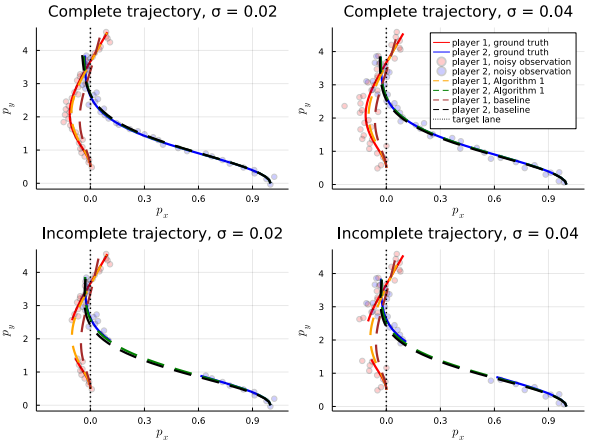

Fig. 6: Full and partial, noisy observation of the expert trajectories. Dashed lines represent predicted trajectories which result from inferred costs, and solid lines are ground truth. The trajectories predicted by Algorithm 1 are closer to the ground truth than the baseline.

of player 2 causes numerical problems in the baseline KKT OLNE solver [28]. Thus, we add an L_2 regularization $10^{-4}\|\theta\|_2^2$ to the loss $\hat{L}(\theta, x_1)$ and summarize the Monte Carlo study in Fig. 8, where we see Algorithm 1 is also able to learn better cost functions reflecting the true intentions of each vehicle in feedback games, even with only partial state observations and incomplete trajectory data.

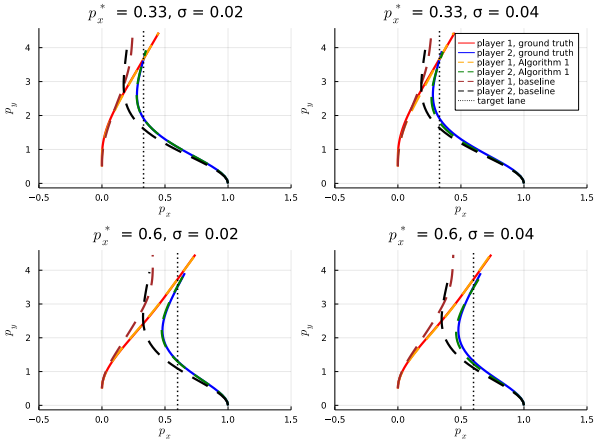

Fig. 7: Generalization performance comparison. p_x^* is the target lane position that player 1 wants to guide player 2 toward. All the costs are inferred from partial observations and incomplete trajectory data, with different noise variance specified in each of the subplot. The trajectories predicted by Algorithm 1 are closer to the ground truth than the baseline.

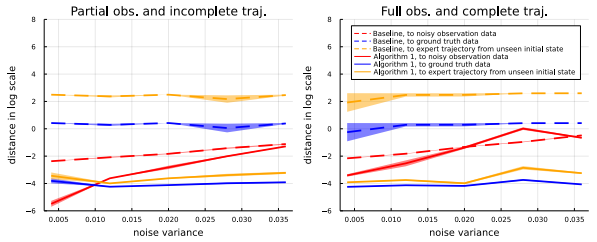

Fig. 8: 3-vehicle platooning scenario. The bold lines and shaded areas represent the mean values and their standard error, i.e., the variance divided by the square root of the sample size, respectively. As the noise variance growing, the converged loss value increases on the average, as shown in the red curves. However, Algorithm 1 is still able to learn a more accurate cost and has less generalization error than the baseline, as shown in the blue and yellow curves, respectively.

7 CONCLUSION

In this work, we propose an efficient cost inference algorithm for inverse feedback nonlinear games, with only partial state observation and incomplete trajectory data. Empirical results show that the proposed solver converges reliably for inverse games with non-convex costs and has superior generalization performance than a state-of-the-art open-loop baseline method when the expert demonstration reflects a group of agents acting in a dynamic feedback game. There are many future directions. We can investigate under what conditions the cost can be inferred exactly in feedback games. The active and online inference are also promising directions. In addition, we are eager to extend this work to settings of closed-loop interaction. In such an extension, rather than merely inferring the objectives of observed players, this information would be used to guide the decision-making of an autonomous agent in that scene.

APPENDIX

Proof of Proposition 1. Proposition 1 claims that there exists an inverse LQ game, which has isolated global minima and the induced FBNE state trajectories of those solutions match the expert demonstration. Here, we show such a counterexample, which supports the claim. Consider a 2-player horizon-3 LQ game with the linear dynamics

$$x_{t+1} = x_t + u_t^1 + u_t^2, \quad t \in \{1, 2, 3\}, \tag{16}$$

and the cost

$$\begin{aligned} g_t^1(x_t, u_t) &= \frac{1}{2}(Q^1\|x_t\|_2^2 + \|u_t^1\|_2^2), \quad t \in \{1, 2\}, \\ g_t^2(x_t, u_t) &= \frac{1}{2}(Q^2\|x_t\|_2^2 + 2\|u_t^2\|_2^2), \quad t \in \{1, 2\}, \\ g_3^1(x_3, u_3) &= \frac{1}{2}Q^1\|x_3\|_2^2, \; g_3^2(x_3, u_3) = \frac{1}{2}Q^2\|x_3\|_2^2. \end{aligned} \tag{17}$$

We assume that the ground truth solutions are $Q^1 = 1$, $Q^2 = 1$. We will show there is also one extra solution $\hat{Q}^1 = \frac{1}{2}$ and $\hat{Q}^2 = 2$, which yields the same FBNE state trajectory as the ground truth for any initial condition. We follow the same definition of the variable $\{Z_t^i\}_{t=1,i=1}^{3,2}$ as in [2]. By definition, we have $Z_t^i \geq Q^i > 0$, when $Q^1 \in \mathbb{R}_+$ and $Q^2 \in \mathbb{R}_+$. Following the notations in FBNE condition in Corollary 6.1 of [2], we consider the feedback matrices $\{P_t^i\}_{t=1,i=1}^{2,2}$,

$$\begin{bmatrix} P_t^1 \\ P_t^2 \end{bmatrix} = \underbrace{\begin{bmatrix} 1 + Z_{t+1}^1 & Z_{t+1}^1 \\ Z_{t+1}^2 & 2 + Z_{t+1}^2 \end{bmatrix}}_{G_t^i} \begin{bmatrix} Z_{t+1}^1 \\ Z_{t+1}^2 \end{bmatrix}, \quad \forall t \in \{1, 2\}, \tag{18}$$

where the matrix G_t^i is invertible because $\det(G_t^i) = 2 + Z_{t+1}^2 + 2Z_{t+1}^1 > 0$. The above analysis suggests that the FBNE state trajectory for all $Q^1 > 0$ and $Q^2 > 0$ are uniquely determined. We consider the time instant $t = 2$, and observe

$$\begin{bmatrix} P_2^1 \\ P_2^2 \end{bmatrix} = \begin{bmatrix} 1 + Q^1 & Q^1 \\ Q^2 & 2 + Q^2 \end{bmatrix}^{-1} \begin{bmatrix} Q^1 \\ Q^2 \end{bmatrix} = \frac{1}{2 + 2Q^1 + Q^2} \begin{bmatrix} 2Q^1 \\ Q^2 \end{bmatrix}. \tag{19}$$

We then have the closed-loop dynamics $x_3 = (1 - P_2^1 - P_2^2)x_2 = \frac{2}{2+2Q^1+Q^2}x_2$, which yields that for two pairs of positive variables (Q^1, Q^2) and $(\hat{Q}^1, \hat{Q}^2)$, a necessary condition for them to have the same FBNE trajectory is that $2Q^1 + Q^2 = 2\hat{Q}^1 + \hat{Q}^2$. We have $Z_2^1 = Q^1 + \frac{Q^1 + (2Q^1)^2}{(2+2Q^1+Q^2)^2}$, $Z_2^2 = Q^2 + \frac{Q^2 + 2(Q^2)^2}{(2+2Q^1+Q^2)^2}$. Similarly, for the time instant $t = 1$, we have $x_2 = (1 - P_1^1 - P_1^2)x_1 = \frac{2}{2+2Z_2^1+Z_2^2}x_1$. A necessary condition for $(\hat{Q}^1, \hat{Q}^2)$ to have the same FBNE state trajectory as (Q^1, Q^2) is that the following 2 equations are satisfied,

$$\begin{aligned} &2Q^1 + Q^2 = 2\hat{Q}^1 + \hat{Q}^2 \\ &2\Big(Q^1 + \frac{Q^1 + (2Q^1)^2}{(2 + 2Q^1 + Q^2)^2}\Big) + Q^2 + \frac{Q^2 + 2(Q^2)^2}{(2 + 2Q^1 + Q^2)^2} \\ &\quad = 2\Big(\hat{Q}^1 + \frac{\hat{Q}^1 + (2\hat{Q}^1)^2}{(2 + 2\hat{Q}^1 + \hat{Q}^2)^2}\Big) + \hat{Q}^2 + \frac{\hat{Q}^2 + 2(\hat{Q}^2)^2}{(2 + 2\hat{Q}^1 + \hat{Q}^2)^2}. \end{aligned} \tag{20}$$

We substitute $Q^1 = 1$, $Q^2 = 1$ and $\hat{Q}^2 = 3 - 2\hat{Q}^1$ into the second row of (20), which is reduced to a 2-degree polynomial of $\hat{Q}^2$. By the fundamental theorem of algebra [4], there exist at most 2 solutions for $\hat{Q}^2$. The two pairs of $(\hat{Q}^1, \hat{Q}^2)$ satisfying (20) are $(1, 1)$ and $(\frac{1}{2}, 2)$. The two global minima are isolated. Since the dimension of the state x_t is 1, for all initial states $x_1 \in \mathbb{R}$, the FBNE state trajectories under the costs specified by the two pairs cost parameters $(1, 1)$ and $(\frac{1}{2}, 2)$ coincide with each other. □

REFERENCES

[1] Chaitanya Awasthi and Andrew Lamperski. Inverse differential games with mixed inequality constraints. In *2020 American control conference (ACC)*, pages 2182–2187. IEEE, 2020.

[2] Tamer Başar and Geert Jan Olsder. *Dynamic Noncooperative Game Theory*. SIAM, 1998.

[3] Stephen Boyd, Stephen P Boyd, and Lieven Vandenberghe. *Convex optimization*. Cambridge university press, 2004.

[4] Augustin-Louis Cauchy. Cours d'analyse de l'ecole royale polytechnique, 1re partie. *Analyse algébrique. Debure freres, Paris*, 1821.

[5] Simon Le Cleac'h, Mac Schwager, and Zachary Manchester. Algames: A fast solver for constrained dynamic games. *arXiv preprint arXiv:1910.09713*, 2019.

[6] JB Cruz Jr. Survey of nash and stackelberg equilibrim strategies in dynamic games. In *Annals of Economic and Social Measurement, Volume 4, number 2*, pages 339–344. NBER, 1975.

[7] Peter Englert, Ngo Anh Vien, and Marc Toussaint. Inverse kkt: Learning cost functions of manipulation tasks from demonstrations. *The International Journal of Robotics Research*, 36(13-14):1474–1488, 2017.

[8] David Fridovich-Keil, Ellis Ratner, Lasse Peters, Anca D Dragan, and Claire J Tomlin. Efficient Iterative Linear-Quadratic Approximations for Nonlinear Multi-Player General-Sum Differential Games. *2020 IEEE international conference on robotics and automation (ICRA)*, pages 1475–1481, 2020.

[9] Volker Gabler, Tim Stahl, Gerold Huber, Ozgur Oguz, and Dirk Wollherr. A game-theoretic approach for adaptive action selection in close proximity human-robot-collaboration. In *2017 IEEE international conference on robotics and automation (ICRA)*, pages 2897–2903. IEEE, 2017.

[10] Jorge Herrera de la Cruz, Benjamin Ivorra, and Ángel M Ramos. An algorithm for solving a class of multiplayer feedback-nash differential games. *Mathematical Problems in Engineering*, 2019, 2019.

[11] Jairo Inga, Esther Bischoff, Florian Köpf, and Sören Hohmann. Inverse dynamic games based on maximum entropy inverse reinforcement learning. *arXiv preprint arXiv:1911.07503*, 2019.

[12] Jairo Inga, Esther Bischoff, Timothy L Molloy, Michael Flad, and Sören Hohmann. Solution Sets for Inverse Non-Cooperative Linear-Quadratic Differential Games. *IEEE Control Systems Letters*, 3(4):871–876, 2019.

[13] Rufus Isaacs. *Differential games: a mathematical theory with applications to warfare and pursuit, control and optimization*. Courier Corporation, 1999.

[14] Georgios Kossioris, Michael Plexousakis, Anastasios Xepapadeas, Aart de Zeeuw, and K-G Mäler. Feedback nash equilibria for non-linear differential games in pollution control. *Journal of Economic Dynamics and Control*, 32(4):1312–1331, 2008.

[15] Steven George Krantz and Harold R Parks. *The implicit function theorem: history, theory, and applications*. Springer Science & Business Media, 2002.

[16] Forrest Laine, David Fridovich-Keil, Chih-Yuan Chiu, and Claire Tomlin. The Computation of Approximate Generalized Feedback Nash Equilibria. *arXiv preprint arXiv:2101.02900*, 2021.

[17] Kang Woo Lee and Jeong-Hoon Hwang. Human–robot interaction as a cooperative game. *Trends in Intelligent Systems and Computer Engineering*, pages 91–103, 2008.

[18] Weiwei Li and Emanuel Todorov. Iterative Linear Quadratic Regulator Design for Nonlinear Biological Movement Systems. *ICINCO*, pages 222–229, 2004.

[19] David Mayne. A second-order gradient method for determining optimal trajectories of non-linear discrete-time systems. *International Journal of Control*, 3(1):85–95, 1966.

[20] Negar Mehr, Mingyu Wang, and Mac Schwager. Maximum-Entropy Multi-Agent Dynamic Games: Forward and Inverse Solutions. *ArXiv*, abs/2110.01027, 2021.

[21] Timothy L Molloy, Jairo Inga Charaja, Sören Hohmann, and Tristan Perez. Inverse optimal control and inverse noncooperative dynamic game theory, 2022.

[22] Timothy L. Molloy, Jason J. Ford, and Tristan Perez. Inverse noncooperative dynamic games. *IFAC-PapersOnLine*, 50(1):11788–11793, 2017. 20th IFAC World Congress.

[23] Timothy L Molloy, Grace S Garden, Tristan Perez, Ingo Schiffner, Debajyoti Karmaker, and Mandyam V Srinivasan. An inverse differential game approach to modelling bird mid-air collision avoidance behaviours. *IFAC-PapersOnLine*, 51(15):754–759, 2018.

[24] Timothy L Molloy, Jairo Inga, Michael Flad, Jason J Ford, Tristan Perez, and Sören Hohmann. Inverse open-loop noncooperative differential games and inverse optimal control. *IEEE Transactions on Automatic Control*, 65(2):897–904, 2019.

[25] Timothy L Molloy, Jairo Inga Charaja, Sören Hohmann, and Tristan Perez. Inverse noncooperative differential games. In *Inverse Optimal Control and Inverse Noncooperative Dynamic Game Theory*, pages 189–226. Springer, 2022.

[26] Yurii Nesterov. A method for unconstrained convex minimization problem with the rate of convergence o (1/kˆ 2). In *Doklady an ussr*, volume 269, pages 543–547, 1983.

[27] Jorge Nocedal and Stephen Wright. *Numerical Optimization*. Springer Science & Business Media, 2006.

[28] Lasse Peters, David Fridovich-Keil, Vicenç Rubies-Royo, Claire J Tomlin, and Cyrill Stachniss. Inferring Objectives in Continuous Dynamic Games from Noise-Corrupted Partial State Observations. *arXiv preprint arXiv:2106.03611*, 2021.

[29] Lasse Peters, David Fridovich-Keil, Claire J. Tomlin, and Zachary N. Sunberg. Inference-Based Strategy Alignment for General-Sum Differential Games. In *Proceedings of the 19th International Conference on Autonomous Agents and MultiAgent Systems*, AAMAS '20, page 1037–1045, Richland, SC, 2020. International Foundation for Autonomous Agents and Multiagent Systems.

[30] Lillian J Ratliff, Samuel A Burden, and S Shankar Sastry. On the characterization of local nash equilibria in continuous games. *IEEE transactions on automatic control*, 61(8):2301–2307, 2016.

[31] Simon Rothfuß, Jairo Inga, Florian Köpf, Michael Flad, and Sören Hohmann. Inverse optimal control for identification in non-cooperative differential games. *IFAC-PapersOnLine*, 50(1):14909–14915, 2017.

[32] Wilko Schwarting, Alyssa Pierson, Javier Alonso-Mora, Sertac Karaman, and Daniela Rus. Social behavior for autonomous vehicles. *Proceedings of the National Academy of Sciences*, 116(50):24972–24978, 2019.

[33] Wilko Schwarting, Alyssa Pierson, Sertac Karaman, and Daniela Rus. Stochastic dynamic games in belief space. *IEEE Transactions on Robotics*, 37(6):2157–2172, 2021.

[34] Ilya Sutskever, James Martens, George Dahl, and Geoffrey Hinton. On the importance of initialization and momentum in deep learning. In *International conference on machine learning*, pages 1139–1147. PMLR, 2013.

[35] Aneel Tanwani and Quanyan Zhu. Feedback Nash equilibrium for randomly switching differential–algebraic games. *IEEE Transactions on Automatic Control*, 65(8):3286–3301, 2019.

[36] Chengpu Yu, Yao Li, Shukai Li, and Jie Chen. Inverse linear quadratic dynamic games using partial state observations. *Automatica*, 145:110534, 2022.