

Stackelberg Game-Theoretic Trajectory Guidance for Multi-Robot Systems with Koopman Operator

Yuhan Zhao and Quanyan Zhu

Abstract— Guided trajectory planning involves a leader robot strategically directing a follower robot to collaboratively reach a designated destination. However, this task becomes notably challenging when the leader lacks complete knowledge of the follower’s decision-making model. There is a need for learning-based methods to effectively design the cooperative plan. To this end, we develop a Stackelberg game-theoretic approach based on the Koopman operator to address the challenge. We first formulate the guided trajectory planning problem through the lens of a dynamic Stackelberg game. We then leverage Koopman operator theory to acquire a learning-based linear system model that approximates the follower’s feedback dynamics. Based on this learned model, the leader devises a collision-free trajectory to guide the follower using receding horizon planning. We use simulations to elaborate on the effectiveness of our approach in generating learning models that accurately predict the follower’s multi-step behavior when compared to alternative learning techniques. Moreover, our approach successfully accomplishes the guidance task and notably reduces the leader’s planning time to nearly half when contrasted with the model-based baseline method¹.

I. INTRODUCTION

As a special kind of cooperation in multi-robot systems, guided cooperation allows multiple robots of heterogeneous capabilities to work together with leader-follower/mentor-apprentice types of interactions and jointly achieve the task objective. It has been studied in many fields including human-robot collaboration [1], [2], multi-agent path planning [3], [4], collective transportation [5]–[7]. In guided cooperation, a more resourceful robot can assist a less sophisticated robot in completing the task by leveraging the capabilities of the two agents. As an example, we consider navigation guidance between two unmanned ground vehicles (UGVs). One UGV has limited sensing and planning capabilities and has difficulty reaching the target destination independently. The other UGV, equipped with better sensing capabilities and planning resources, can find a collision-free trajectory to guide the first UGV to the destination.

Stackelberg games [8] provide a quantitative framework to capture such guided interactions. In the two-UGV example, the more powerful UGV acts as a leader and plans the future trajectory by anticipating the other UGV’s behavior; the less resourceful UGV refers to the leader UGV’s guidance and decides its next move. Their interactions can be formulated

Y. Zhao and Q. Zhu are with the Department of Electrical and Computer Engineering, New York University, Brooklyn, NY, 11201, USA. Email: {yhzha, qz494}@nyu.edu.

This work is partially supported by ECCS-1847056 from the National Science Foundation (NSF).

¹The simulation codes are available at <https://github.com/yuhan16/Stackelberg-Koopman-Learning>.

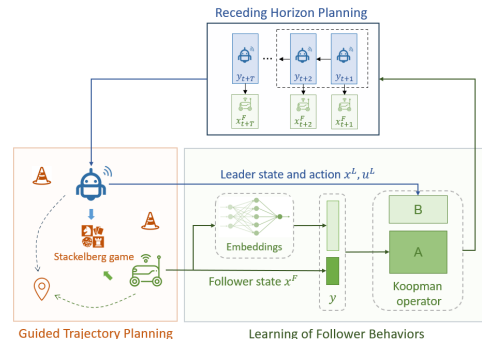


Fig. 1: Illustration of Stackelberg Koopman learning framework in trajectory guidance with an unknown follower robot. The guidance problem is modeled as a Stackelberg game. The leader leverages the Koopman operator to learn a linear model to predict the follower’s behavior. She performs receding horizon planning to find a collision-free trajectory and guide the follower to the target destination.

as a dynamic Stackelberg game. The resulting Stackelberg equilibrium (SE) solution provides a feasible agent-wise cooperative plan for navigation.

However, computing the SE requires the leader robot to know the follower robot’s decision-making model, including the objective function and the dynamics, which can be challenging to obtain in practice. In many cases, we can only observe the past interactive trajectories. Even if the follower’s model is known, model-based algorithms to compute the SE, such as mixed integer programming [9], have many computational challenges. There is a need for learning-based methods to find the SE and enable fast planning in guided cooperation tasks.

Koopman operator theory provides a data-driven method to learn a linear model of nonlinear systems from trajectories [10]–[13]. It also enables efficient model predictive control algorithms using the learned linear model [14]–[16]. Using the Koopman-operator approach, the leader can estimate the follower’s decision-making model using a linear system to anticipate the follower’s behavior. Based on the estimated model, the leader can compute an approximate SE solution to make cooperative plans.

In this work, we develop a Stackelberg game-theoretic approach based on the Koopman operator to address the challenging multi-robot guided cooperation problem, in which the follower’s planning parameters are unknown to the leader. We first formulate the guided cooperation as a dynamic Stackelberg game, where the follower makes *myopic* decisions to interact with the leader and decides his next move. To compute the SE, the leader leverages Koopman operator theory to learn an approximate linear model of the

follower's *feedback* dynamics with the best response. Based on the learned model, the leader computes the approximate SE using receding horizon planning to interact with the follower to complete the guidance task. We use simulations to demonstrate that our approach can successfully complete the guidance in an environment with obstacles. Compared with another two learning methods, our approach provides effective linear models to predict the follower's multi-step behavior. Besides, our approach reduces the leader's planning time by half compared with the model-based baseline method while achieving successful guidance in different scenarios.

II. RELATED WORK

Originating from the seminal work [17], [18], the Koopman operator was proposed to predict the evolution of autonomous dynamical systems using an infinite dimensional linear system. To facilitate computation, finite-dimensional approximation of the infinite linear system can be found by learning the Koopman invariant subspaces or eigenfunctions [19]–[21]. Recent studies in the Koopman operator have prospered its development in both theory and applications [22]–[24]. Various learning algorithms have been developed to estimate the Koopman operator more effectively. For example, Shi et al. in [25] have developed three deep learning algorithms to identify the Koopman embeddings for nonlinear control systems and learn the corresponding linear models for predictive control. Mamakoukas et al. in [26] have leveraged higher-order derivatives of general nonlinear dynamics to improve the learning accuracy of the Koopman-operator-based linear representation and the robustness of the prediction under measurement noise. In robotics, most work leverages the Koopman operator to learn an approximate linear model of the robot's dynamical system and performs model predictive control (MPC). For example, Koopman-based system identification methods for controlling soft robots with precision have been discussed in [27]–[29]. The work [30] by Folkestad et al. has applied Koopman operator theory to learn the nonlinear dynamical systems of multirotors and achieved fast landing using MPC. Similar approaches have been applied to wheeled mobile robots to achieve system identification and trajectory tracking [31]. However, few work has studied the application of Koopman operator in game theory and multi-robot systems. Our paper aims to bridge the gap and introduce the Koopman operator to multi-agent settings.

The Stackelberg game characterizes leader-follower cooperative interactions between agents and can be used to design and evaluate collaboration plans for guidance tasks [32]–[36]. For example, Zhao and Zhu in [33] have leveraged Stackelberg games to coordinate heterogeneous robots and jointly complete multi-object rearrangement tasks. Fisac et al. in [34] has studied efficient and safe trajectory planning for autonomous vehicles in the human driver-aware environment based on Stackelberg games. When there is incomplete information in the Stackelberg game, learning-based methods are helpful in finding the SE. Reinforcement learning has become a popular model-free approach to learning the

SE through interactions [37]–[39]. However, in Stackelberg game-based applications, it is pragmatic to assume that the leader knows her decision-making model but does not know the follower. It is sufficient for the leader to learn the follower's model to compute the SE and find collaboration plans. In this work, we propose an approach using Koopman operator theory to learn a follower's decision-making model and seek an approximate SE for guided trajectory planning.

III. STACKELBERG GAME FORMULATION

A. Trajectory Guidance as Dynamic Stackelberg Games

We consider a leader robot L (she) guiding a follower robot F (he) in the working space \mathcal{W} to reach the target destination while avoiding obstacles. Due to the limited sensing and planning capabilities, the follower can only make one-step (myopic) local planning to decide the next move. Therefore, he needs guidance from a more resourceful leader who can sense the entire environment and make multi-step predictions to find a collision-free trajectory.

Let $x_t^L \in \mathbb{R}^{n^L}$ and $x_t^F \in \mathbb{R}^{n^F}$ be the leader and follower's states at time t , including their positions. Let $u_t^L \in \mathcal{U}^L$ and $u_t^F \in \mathcal{U}^F$ be their controls and admissible control sets, respectively. We further denote $x_t := [x_t^L, x_t^F]$ as the joint state at time t and x^d as the common target state. During the interactive guidance, two robots first observe the joint state x_t . The leader determines an action u_t^L and announces it to the follower. Next, the follower myopically seeks a one-step-optimal action u_t^{F*} to move based on his objective function J^F and the leader's action u_t^L . Two robots then move to the new state x_{t+1} and start the next round of guidance. We assume M obstacles in \mathcal{W} and define the safety constraint $c_j^i(x^i) \geq 0$, $i \in \{L, F\}$, $j = 1, \dots, M$, to achieve obstacle avoidance. Using the follower's responses, the leader computes T -step actions $\mathbf{u}^L := \{u_t^L\}_{t=0}^{T-1}$ that generates a collision-free trajectory to guide the follower to the target destination. The trajectory guidance problem can be formulated as a dynamic Stackelberg game \mathcal{G} as follows:

$$\min_{\mathbf{u}^L \in \mathcal{U}^L} J^L(\mathbf{u}^L, \mathbf{u}^F) := \sum_{t=0}^{T-1} g_t^L(x_t, u_t^L, u_t^F) + g_T^L(x_T) \quad (1)$$

$$\text{s.t. } x_{t+1}^L = f^L(x_t, u_t^L), \quad t = 0, \dots, T-1, \quad (2)$$

$$c_j^L(x_t) \geq 0, \quad j = 1, \dots, M, \quad t = 1, \dots, T, \quad (3)$$

$$\min_{u_t^F \in \mathcal{U}^F} J^F(u_t^F; x_t, u_t^L) \quad t = 0, \dots, T-1, \quad (4a)$$

$$\text{s.t. } x_{t+1}^F = f^F(x_t^F, u_t^F), \quad (4b)$$

$$c_j^F(x_{t+1}^F) \geq 0, \quad j = 1, \dots, M, \quad (4c)$$

$$x_0^L, x_0^F \text{ given.}$$

Here, T is the leader's planning horizon. J^L and $\{g_t^L\}_{t=0}^T$ in (1) denote the leader's objective and stage costs. Eq. (2)–(3) capture the leader's dynamics and safety constraints. The follower's myopic planning problem is given by (4a)–(4c), where J^F in (4a) is his objective function that takes the current state x_t and the leader's action u_t^L as parameters. (4b) is the follower's dynamics and (4c) ensures one-step safety for the follower.

Remark. In this work, we consider regular obstacles that can be represented by norms. The safety constraint for obstacle $j = 1, \dots, M$ and robot $i \in \{L, F\}$ can be defined by

$$c_j^i(x^i) := \|\Lambda_j (p(x^i) - p_j^O)\|_l - r_j \geq 0.$$

Here, $p(x^i)$ is the position of robot i recovered from the state x^i ; p_j^O is the position of the j -th obstacle center; Λ_j and r_j determine the obstacle's shape and range; l denotes the l -norm. For example, we set $l = l_2$ for circular obstacles and $l = l_\infty$ for rectangle obstacles.

We use the open-loop SE $\langle \mathbf{u}_{SE}^L, \mathbf{u}_{SE}^F \rangle$ of the game \mathcal{G} as the agent-wise collaboration plan for the guidance task. The leader and follower take the actions in \mathbf{u}_{SE}^L and \mathbf{u}_{SE}^F , respectively, to generate a collision-free trajectory $\mathbf{x} := \{x_t\}_{t=1}^T$ starting from the initial state $x_0 := [x_0^L, x_0^F]$.

1) *Guidance in Trajectory Planning:* The guidance is an interactive process rather than unilateral instructions from the leader. The follower can make (myopic) decisions for the next move, and the leader's action serves as a reference to assist the follower in making effective local planning. Without guidance, the follower will not reach the target destination due to limited sensing and planning capabilities.

Here, the leader needs to predict the follower's behavior when computing the SE. The leader's prediction coincides with the follower's real behavior if the leader has complete information about the follower, including the follower's objective and dynamics. However, if the leader only has an estimated follower model, her prediction can deviate from the follower's real action. Thus, the follower always plans his action instead of following the leader's instructions.

B. Model-Based Approach for SE

The bilevel programming nature of the game \mathcal{G} challenges the computation of the SE. We leverage reformulation techniques to obtain tractable algorithms. Note that the follower is myopic and there are T inner-level problems (follower's problem) in \mathcal{G} . For the t -th inner-level problem, we penalize the safety constraints with log barrier functions and define

$$\tilde{J}_t^F(u_t^F; x_t, u_t^L) := J^F - \frac{1}{\mu} \sum_{j=1}^M \log(c_j^F(f^F(x_t^F, u_t^F))),$$

where $\mu > 0$ is the penalty parameter. Then, the t -th inner-level problem becomes

$$\min_{u_t^F \in \mathcal{U}^F} \tilde{J}_t^F(u_t^F; x_t, u_t^L).$$

We use the first-order condition $\nabla_{u_t^F} \tilde{J}_t^F = 0$ to approximate the optimality condition since \tilde{J}_t^F is nonconvex. Thus, the leader's problem becomes a single-agent Optimal Control Problem (OCP), denoted by \mathcal{P}_{loc} :

$$\begin{aligned} \min_{\mathbf{u}^L, \mathbf{u}^F} \quad & J^L(\mathbf{u}^L, \mathbf{u}^F) \\ \text{s.t.} \quad & x_{t+1}^L = f^L(x_t^L, u_t^L), \quad t = 0, \dots, T-1, \\ & c_j^L(x_t) \geq 0, \quad j = 1, \dots, M, \quad t = 1, \dots, T. \\ & \nabla_{u_t^F} \tilde{J}_t^F(u_t^F; x_t, u_t^L) = 0, \quad t = 0, \dots, T-1, \\ & x_0^L, x_0^F \text{ given,} \end{aligned} \quad (5)$$

which can be efficiently solved by trajectory optimization methods, such as direct collocation. We denote the solution of (5) as $\langle \tilde{\mathbf{u}}_{\text{loc}}^L, \tilde{\mathbf{u}}_{\text{loc}}^F \rangle$, which can be used to approximate the SE $\langle \mathbf{u}_{SE}^L, \mathbf{u}_{SE}^F \rangle$ of \mathcal{G} for generating cooperative plans.

C. Guidance with Unknown Follower

The model-based approach (5) provides an approximate solution of \mathcal{G} . However, the leader must know the follower's entire model (4a)-(4c). In many cases, the leader can only observe the follower's states and actions through interactions rather than the decision-making model. Therefore, we need learning-based methods to compute the SE of the game \mathcal{G} .

From the follower's problem (4a)-(4c), we define follower's feedback policy $\rho^F: \mathbb{R}^{n^F} \times \mathbb{R}^{n^L} \times \mathcal{U}^L \rightarrow \mathcal{U}^F$ by:

$$\rho^F(x_t^F; x_t^L, u_t^L) = \arg \min_{u_t^F \in \mathcal{U}^F} (4a)-(4c).$$

Note that the feedback policy ρ^F is a stationary function since the follower's problem remains the same for all t . The follower uses it to generate actions and update his dynamical system, yielding *feedback dynamics*:

$$x_{t+1}^F = f^F(x_t^F, \rho^F(x_t^F; x_t^L, u_t^L)) := \tilde{f}^F(x_t, u_t^L). \quad (6)$$

Once the leader knows the feedback policy ρ^F , her objective function can be written as $J^L(\mathbf{u}^L, \rho(\mathbf{x}, \mathbf{u}^L)) := \tilde{J}^L(\mathbf{u}^L)$. Her decision-making problem (i.e., the game \mathcal{G}) can be reduced to the following OCP, denoted by \mathcal{P}_{fb} :

$$\begin{aligned} \min_{\mathbf{u}^L \in \mathcal{U}^L} \quad & \tilde{J}^L(\mathbf{u}^L) \\ \text{s.t.} \quad & x_{t+1}^L = f^L(x_t^L, u_t^L), \quad t = 0, \dots, T-1, \\ & x_{t+1}^F = \tilde{f}^F(x_t, u_t^L), \quad t = 0, \dots, T-1, \\ & c_j^L(x_t) \geq 0, \quad j = 1, \dots, M, \quad t = 1, \dots, T. \\ & x_0^L, x_0^F \text{ given,} \end{aligned} \quad (7)$$

The optimal solution of (7), denoted by \mathbf{u}_{fb}^L , coincides with the SE of the leader's part \mathbf{u}_{SE}^L . In fact, the follower's response trajectory, denoted by \mathbf{u}_{fb}^F , coincides with \mathbf{u}_{SE}^F . In other words, the leader can solve (7) and use the optimal solution to guide the follower, which essentially captures the collaborative planning through the SE approach. Note that the leader in (7) does not predict the follower's action because the follower's optimal action generated by ρ^F is already incorporated in the feedback dynamics \tilde{f}^F . The Koopman operator provides a data-driven method to learn a linear model of nonlinear systems. In the next section, we present a Koopman-based approach to enable learning-based methods to approximate the SE.

IV. KOOPMAN OPERATOR FOR FEEDBACK DYNAMICS

A. Basics of Koopman Operator

We consider a discrete nonlinear dynamical system

$$x_{t+1} = f(x_t), \quad \text{where } f: \mathbb{R}^n \rightarrow \mathbb{R}^n.$$

The Koopman operator \mathcal{K} is an infinite-dimensional linear operator that acts on the embedding function $g: \mathbb{R}^n \rightarrow \mathbb{R}^q$ to measure the state evolution along the system trajectory:

$$\mathcal{K}g(x_t) = g \circ f(x_t) = g(x_{t+1}), \quad t = 0, 1, \dots$$

The Koopman operator has also been extended to dynamical systems with controls [15]. For a system $x_{t+1} = f(x_t, u_t)$ with $f : \mathbb{R}^n \times \mathbb{R}^m \rightarrow \mathbb{R}^n$, the Koopman operator acting on the extended embedding function $g : \mathbb{R}^n \times \mathbb{R}^m \rightarrow \mathbb{R}^q$ yields

$$\mathcal{K}g(x_t, u_t) = g \circ f(x_t, u_t) = g(x_{t+1}, u_{t+1}).$$

Using the embedding function g , the Koopman operator lifts the state and control spaces to a higher dimensional embedding space to obtain a new linear system that is easier to handle.

Note that Koopman operator \mathcal{K} is an infinite-dimensional operator. We aim to find a finite matrix representation $K \in \mathbb{R}^{q \times q}$ of \mathcal{K} such that $g(x_{t+1}, u_{t+1}) \approx Kg(x_t, u_t)$. To obtain K , we split the embedding function g into two parts $g = [g_x(x, u)^\top g_u(x, u)^\top]^\top$ and further assume that g_x is only related to the state, i.e., $g_x(x) := g_x(x, u)$. Then, we obtain

$$\begin{bmatrix} g_x(x_{t+1}) \\ g_u(x_{t+1}, u_{t+1}) \end{bmatrix} = \begin{bmatrix} K_{xx} & K_{xu} \\ K_{ux} & K_{uu} \end{bmatrix} \begin{bmatrix} g_x(x_t) \\ g_u(x_t, u_t) \end{bmatrix},$$

which gives a linear system in terms of embedding functions:

$$g_x(x_{t+1}) = K_{xx}g_x(x_t) + K_{xu}g_u(x_t, u_t). \quad (8)$$

Many parameterization methods on $g_u(x, u)$ are proposed to recover the control u_t . The linear parameterization is the most common, which sets $g_u(x, u) = u$. The advantage of linear parameterization is that it preserves the control u in the learned system, which makes it more convenient to perform MPC. In this work, we adopt linear parameterization.

Given a S -step state-action trajectory $\{\hat{x}_t, \hat{u}_t, \hat{x}_{t+1}\}_{t=0}^{S-1}$, we first compute the embedding state $\hat{g}_t := g_x(\hat{x}_t)$. The Koopman operator in (8) is computed by optimizing the S -step prediction loss:

$$K_{xx}^*, K_{xu}^* = \arg \min_{K_1, K_2} \sum_{t=0}^{S-1} \gamma^t \|\hat{g}_{t+1} - K_1 \hat{g}_t - K_2 \hat{u}_t\|_2^2,$$

where $\gamma \in (0, 1]$ is the weight decay factor. The use of S -step loss contributes to the accuracy of long-term prediction.

B. Learning Feedback Dynamics Using Koopman Operator

We leverage Koopman operator theory to learn the follower's feedback dynamics f^F in (6). Note that we need to learn both the Koopman operator K and the embedding function g_x . We parameterize g_x with some parameter θ (denoting h_θ as the parameterized function) and further encode the state itself to g_x , which gives $g_x(x) = [x^\top h_\theta(x)^\top]^\top$. Since f^F outputs the follower's state x^F , we denote $w^L := [x^L, u^L]$ as the augmented variable and let $y_t = g_x(x_t^F)$. x_t^F can be recovered by $x_t^F = Cy_t$ where $C = [I_{n^F} \ 0]$. Then, we obtain the following linear system

$$\begin{aligned} y_{t+1} &= K_{xx}y_t + K_{xu}w_t^L := Ay_t + Bw_t^L \\ &:= Ay_t + B_1x_t^L + B_2u_t^L, \end{aligned} \quad (9)$$

$$x_t^F = Cy_t.$$

Here, we split $B := [B_1 \ B_2]$ for clarity. We can leverage (9) to predict the follower's future state given the leader's state-action pair $\langle x_t^L, u_t^L \rangle$ and the initial x_0^F .

Using the interactive trajectory (length S) data set $\mathcal{D} = \{\{\hat{x}_{t,i}^L, \hat{x}_{t,i}^F, \hat{u}_{t,i}^L, \hat{u}_{t,i}^F, \hat{x}_{t+1,i}^L, \hat{x}_{t+1,i}^F\}_{i=1}^N\}_{t=0}^{S-1}$, the leader learns A, B and the embedding functions g_x by minimizing the loss

$$L(\theta, A, B) = \frac{1}{N} \sum_{i=1}^N \sum_{t=0}^{S-1} \gamma^t \|g_x(\hat{x}_{t,i}^F) - Ag_x(\hat{x}_{t,i}^L) - B\hat{u}_{t,i}^L\|_2^2,$$

which can be optimized by Stochastic Gradient Descent (SGD). Then, the leader uses the learned model in lieu of the follower's feedback dynamics in (7) and solves the following problem, denoted as \mathcal{P}_{kp} , to approximate the solution of (7):

$$\begin{aligned} \min_{\mathbf{u}^L \in \mathcal{U}^L} \quad & \tilde{J}^L(\mathbf{u}^L) \\ \text{s.t.} \quad & x_{t+1}^L = f^L(x_t, u_t^L), \quad t = 0, \dots, T-1, \\ & y_{t+1} = Ay_t + B_1x_t^L + B_2u_t^L, \quad t = 0, \dots, T-1, \\ & c_j^L(x_t^L) \geq 0, \quad j = 1, \dots, M, \quad t = 1, \dots, T, \\ & x_0^L, y_0(x_0^F) \text{ given.} \end{aligned} \quad (10)$$

We denote the solution of (10) as $\tilde{\mathbf{u}}_{\text{kp}}^L$, which approximates \mathbf{u}_{fb}^L and hence \mathbf{u}_{SE}^L . With the learned model, the leader uses $\tilde{\mathbf{u}}_{\text{kp}}^L$ to guide the follower. The follower's response under $\tilde{\mathbf{u}}_{\text{kp}}^L$, denoted by $\tilde{\mathbf{u}}_{\text{kp}}^F$, together with $\tilde{\mathbf{u}}_{\text{kp}}^L$, constitutes an approximation of the SE and a collaboration plan.

C. Scenario-based Receding Horizon Planning

We use receding horizon planning to perform the guidance. The leader periodically updates the guidance trajectory based on the learned model to overcome the approximation errors in Koopman learning. We note that guidance becomes less meaningful if the leader is too far away from the follower. Therefore, we design a scenario-based planning mechanism by introducing a guidance threshold. When the distance between the leader and the follower exceeds the threshold, the leader will return to the follower. Otherwise, the leader moves toward the destination to guide the follower. The follower's objective remains unchanged in the guidance. The scenario-based planning mechanism can be achieved by switching the leader's objective function, which is summarized in Alg. 1.

V. SIMULATIONS AND EVALUATIONS

A. Simulation Settings

We choose a $[0, 10] \times [0, 10]$ working space \mathcal{W} with 4 obstacles shown in Fig. 3. A discrete unicycle model is used as the leader and follower's dynamical systems:

$$x_{t+1}^i = x_t^i + \begin{bmatrix} v_t^i \cos(\theta_t^i) \\ v_t^i \sin(\theta_t^i) \\ \omega_t^i \end{bmatrix} dt, \quad i \in \{L, F\}.$$

The state $x^i = [p_x^i \ p_y^i \ \theta^i] \in \mathbb{R}^3$ denotes the (x, y) -positions and the heading angle of the robot i ; the control $u^i = [v^i \ \omega^i] \in [0, 2] \times [-2, 2]$ represents the linear and angular velocities. The leader's planning horizon is set as $T = 5$ and $dt = 0.2$. The target destination is $[9, 9]$.

Algorithm 1: Receding horizon planning for strategic trajectory guidance.

- 1 **Initialize:** Initial state x_0^L, x_0^F , destination x^d , scenario threshold λ ;
 - 2 Leader learns the Koopman operator by minimizing S -step loss ;
 - 3 **for** $t = 0, 1, 2, \dots$ **do**
 - 4 Leader computes embedding state y_t using x_t^F ;
 - 5 **if** $\text{dist}(x_t^L - x_t^F) > \lambda$ **then**
 - 6 Leader sets objective 1 to go to the follower ;
 - 7 **else**
 - 8 Leader sets objective 2 to head to destination ;
 - 9 $u^{L*} \leftarrow$ Leader solves OCP (10) ;
 - 10 Leader applies u_0^{L*} ;
 - 11 Follower observes x_t, u_0^{L*} and computes u_0^{F*} ;
 - 12 Leader and follower update dynamics x_{t+1}^L, x_{t+1}^F ;
 - 13 $x_t^L \leftarrow x_{t+1}^L, x_t^F \leftarrow x_{t+1}^F$;
 - 14 **if** Follower reaches destination **then**
 - 15 **break**;
-

The leader's stage cost is given by

$$g_t^L = \|x_t^L - x_t^F\|_{Q_1^L}^2 + \|x_t^L - x^d\|_{Q_2^L}^2 + \|u_t^L\|_{R^L}^2,$$

where the first term measures the guidance effectiveness and the second term drives the leader to the destination. The terminal cost of g_T^L is the same as the first two terms of g_t^L . We set $Q_1^L = \text{diag}(2, 2, 0)$, $R^L = \text{diag}(2, 1)$. To implement scenario-based guidance, we set the threshold $\lambda = 1$. When $\lambda < 1$, $Q_2^L = \text{diag}(1, 1, 0)$. Otherwise, $Q_2^L = \text{diag}(0.1, 0.1, 0)$. The purpose is to make the leader approach the follower because greater distance does not facilitate effective guidance. To simulate interactive data, we use the following *ground truth* cost function as the follower's objective:

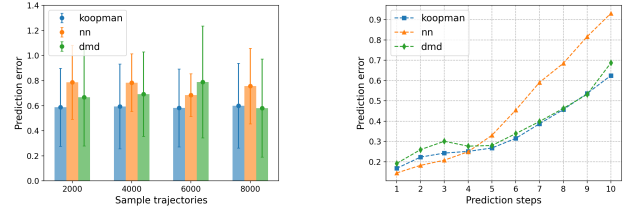
$$J_t^F(u_t^F; x_t^L, u_t^L) = \|x_{t+1}^F - x_{t+1}^L\|_{Q_1^F}^2 + \|x_{t+1}^F - x^d\|_{Q_2^F}^2 + Q_3^F \cdot \langle \vec{d}^L, \vec{d}^F \rangle + \|u_t^F\|_{R^F}^2.$$

Here, $\vec{d}^L = [\cos(\theta_{t+1}^L), \sin(\theta_{t+1}^L)] \in \mathbb{R}^2$ and $\vec{d}^F = [\cos(\theta_{t+1}^F), \sin(\theta_{t+1}^F)] \in \mathbb{R}^2$ are the leader's and the follower's heading directions, respectively. $x_{t+1}^L = f^L(x_t^L, u_t^L)$. The first two terms are similar to the ones in g_t^L . The third term aligns the follower's heading direction with the leader. We use $Q_1^F = \text{diag}(10, 10, 0)$, $Q_2^F = \text{diag}(0.1, 0.1, 0)$, $Q_3^F = -1$, $R^F = \text{diag}(2, 0.05)$.

The follower's best response (the inner-level problem) is solved by grid search to avoid local optimal solutions. We collect interactive trajectories with length $S = 30$ to generate the data set \mathcal{D} .

B. Koopman Learning Results

We use a neural network (NN) with three layers of 90 ReLU nonlinearities to parameterize the embedding function g_x in Sec. IV-B and matrices $\langle A, B \rangle$ for the Koopman operator. The training objective is set by a 30-step prediction



(a) Prediction error at the 10th step over 20 trajectories. (b) Step-wise prediction error of a single trajectory.

Fig. 2: Prediction errors with the learned model generated by three approaches. Our Koopman-based approach provides a smaller averaged prediction error and better long-term performance compared with the other two.

loss. We perform the training on different sizes of source data. For each size, we conduct the training for 10 times to evaluate the average learning performance. The model is trained over 10^3 epochs with SGD each time. All learning is performed on an RTX 6000 GPU, and Tab. I summarizes the training results using different numbers of data. We can observe that the training is stable as the number of samples increases, meaning that the data set size has less impact on learning the Koopman operator.

	$N = 2500$	$N = 5000$	$N = 7500$	$N = 10000$
Train loss	0.178(0.052)	0.173(0.045)	0.172(0.044)	0.169(0.042)
Test loss	0.188(0.053)	0.228(0.070)	0.182(0.045)	0.183(0.043)
Train time	18.3 min	35.4 min	52.8 min	70.8 min

TABLE I: Summary of Learning statistics of the Koopman operator. N is the size of the data set, and the training/testing data contains 80/20. The mean and variance of all losses are averaged over ten training episodes.

For comparison, we also implement another two methods to learn the follower's feedback dynamics \tilde{f}^F . The first one uses an NN of 64 ReLU nonlinearities to fit the one-step dynamics (labeled by nn-approach). The second one learns a linear model of \tilde{f}^F using Dynamic Mode Decomposition [40] (labeled by dmd-approach). They are training with $N = 10000$ trajectory samples.

We use the learned model by three approaches to predict the follower's response trajectory and evaluate the learning performance. The prediction is proceeded by (9). The prediction error is the difference between the predicted and ground truth states. We perform 10-step predictions over 20 ground truth interactive trajectories and plot the last step (the 10th step) prediction error in Fig. 2a. We observe that our approach provides the learned model that generates the smallest prediction error in the first three cases. In the last case with $N = 8000$, our approach only generates a slightly larger error compared with dmd-approach. However, the dmd-approach yields a larger variance than ours. For better visualization, we plot the step-wise prediction errors of three approaches using one ground truth trajectory in Fig. 2b. It is worth noting that the nn-approach yields a better prediction result in the first three steps. It is because the nn-approach is designed to fit the one-step dynamics directly and hence performs better for short-term prediction. However, it shows poor long-term prediction performance compared with our approach. This is due to the fact that our approach

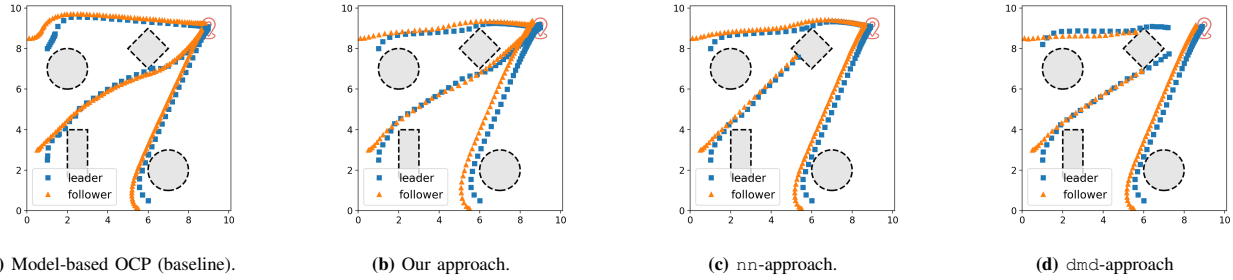


Fig. 3: Interactive guidance trajectories using receding horizon planning and the models learned by different approaches. The blue and the orange represent the leader and follower trajectories, respectively. The follower start from $[0, 8.5]$, $[0.5, 3]$, and $[5.5, 0]$ to reach the destination $[9, 9]$. The leader successfully guides the follower using our approach. *nn*-approach (and *dmd*-approach) fail to guide the follower in one (and two) case.

considers the dynamics evolution along trajectories. Also, due to the embedding function, our approach can capture more information in the dynamics compared with *dmd*-approach and hence has a smaller prediction error. Fig. 2b also suggests that the learned feedback dynamical model by our approach is more suitable for long-term planning.

C. Receding Horizon Planning Results

After obtaining the trained model, the leader runs Alg. 1 to achieve guidance. We simulate interactive guidance trajectories from different initial positions using three approaches. Besides, we also simulate baseline trajectories using the model-based OCP (5) to compare our approach, as shown in Fig. 3.

Our approach successfully guides the follower to the destination from three starting points. Compared with the baseline, the follower’s terminal position slightly deviates from the exact target $[9, 9]$ but is still within an acceptable range, which shows the effectiveness of our approach. This is mainly because of approximation accuracy in the learned model. However, the other two methods fail the guidance task in some cases. The *nn*-approach fails to overcome the diamond obstacle starting from the position $[0.5, 3]$ because the leader’s prediction on the follower has already diverged near the obstacle. The *dmd*-approach fails to guide the follower to circumvent the diamond obstacle because of the model accuracy. We observe that the leader has already navigated past the obstacle, and her prediction model indicates that the follower will also successfully avoid the obstacle by adhering to her trajectory. The absence of nonlinearity in the follower’s model learned through the *dmd*-approach leads to inadequacy in providing guidance near obstacles.

We take the guidance trajectory starting from $[5.5, 0]$ for detailed discussion since all approaches complete this guidance task. We plot the leader’s planning time along the trajectory in Fig. 4a. It is not surprising that the model-based OCP takes the longest time for planning. We note that our approach reduces planning time by about half compared with the modeled-based OCP baseline and still completes the guidance tasks, showing notable time efficiency in guidance planning. The *dmd*-approach achieves similar planning time as it uses linear models to predict follower dynamics. The *nn*-based model takes more time to plan as the learned model is nonlinear (an NN).

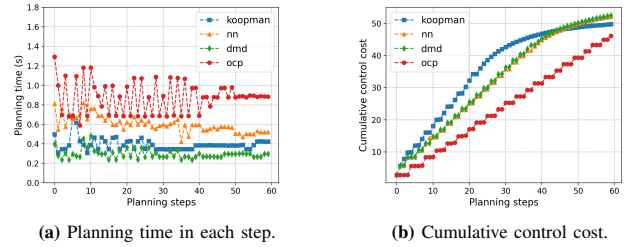


Fig. 4: Planning details for the interactive trajectory starting from $[5.5, 0]$.

We also plot the leader’s cumulative control cost in Fig. 4b to measure the guidance efficiency. The model-based OCP baseline yields the least control effort due to the model knowledge. Our approach provides a higher control cost at the beginning and soon becomes stable, meaning that the guidance is completed in a shorter time. In comparison, *dmd*-approach and *nn*-approach shows a higher guidance cost when approaching the destination. It indicates that our approach provides comparably efficient guidance regarding the total control cost.

VI. CONCLUSIONS

In this work, we have proposed a Stackelberg game-theoretic approach to address the challenge of guided trajectory planning when dealing with an unknown follower robot. Our approach not only formulates a game-theoretic framework for multi-robot cooperation in trajectory guidance but also develops an algorithm based on the Koopman operator. This algorithm effectively learns the follower’s decision-making model and achieves both rapid and control-efficient guidance. Simulations have corroborated our approach, demonstrating its superior performance in generating a more accurate linear model for predicting the follower’s long-term behavior compared to alternative learning methods. Benchmarked with the model-based baseline method, our approach greatly reduces the planning time while successfully accomplishing the guidance task.

There are remaining challenges to ensure safety when employing the Koopman operator for mission-critical applications. A potential avenue for future research involves an in-depth exploration of safety assurances, encompassing critical aspects such as the establishment of error bounds and the identification of valid operational regions.

REFERENCES

- [1] H. Bo, D. M. Mohan, M. Azhar, K. Sreekanth, and D. Campolo, "Human-robot collaboration for tooling path guidance," in *2016 6th IEEE International Conference on Biomedical Robotics and Biomechatronics (BioRob)*. IEEE, 2016, pp. 1340–1345.
- [2] J. Cacace, R. Caccavale, A. Finzi, and R. Grieco, "Combining human guidance and structured task execution during physical human-robot collaboration," *Journal of Intelligent Manufacturing*, pp. 1–15, 2022.
- [3] X. C. Ding, A. R. Rahmani, and M. Egerstedt, "Multi-uav convoy protection: An optimal approach to path planning and coordination," *IEEE transactions on Robotics*, vol. 26, no. 2, pp. 256–268, 2010.
- [4] M. Bibuli, M. Caccia, L. Lapiere, and G. Bruzzone, "Guidance of unmanned surface vehicles: Experiments in vehicle following," *IEEE Robotics & Automation Magazine*, vol. 19, no. 3, pp. 92–102, 2012.
- [5] T. Machado, T. Malheiro, S. Monteiro, W. Erlhagen, and E. Bicho, "Multi-constrained joint transportation tasks by teams of autonomous mobile robots using a dynamical systems approach," in *2016 IEEE international conference on robotics and automation (ICRA)*. IEEE, 2016, pp. 3111–3117.
- [6] X. Fu, D. Wang, J. Hu, J. Wei, and C.-B. Yan, "Leader-follower based two-agv cooperative transportation system in 5g environment," in *2022 IEEE 18th International Conference on Automation Science and Engineering (CASE)*. IEEE, 2022, pp. 67–72.
- [7] D. Koung, O. Kermorgant, I. Fantoni, and L. Belouaer, "Cooperative multi-robot object transportation system based on hierarchical quadratic programming," *IEEE Robotics and Automation Letters*, vol. 6, no. 4, pp. 6466–6472, 2021.
- [8] T. Başar and G. J. Olsder, *Dynamic noncooperative game theory*. SIAM, 1998.
- [9] P. Paruchuri, J. P. Pearce, J. Marecki, M. Tambe, F. Ordonez, and S. Kraus, "Playing games for security: An efficient exact algorithm for solving bayesian stackelberg games," in *Proceedings of the 7th international joint conference on Autonomous agents and multiagent systems-Volume 2*, 2008, pp. 895–902.
- [10] S. L. Brunton and J. N. Kutz, *Data-driven science and engineering: Machine learning, dynamical systems, and control*. Cambridge University Press, 2019.
- [11] P. Bevanda, S. Sosnowski, and S. Hirche, "Koopman operator dynamical models: Learning, analysis and control," *Annual Reviews in Control*, vol. 52, pp. 197–212, 2021.
- [12] S. L. Brunton, M. Budišić, E. Kaiser, and J. N. Kutz, "Modern koopman theory for dynamical systems," *SIAM Review*, vol. 64, no. 2, pp. 229–340, 2022.
- [13] I. Mezić, "Koopman operator, geometry, and learning of dynamical systems," *Not. Am. Math. Soc.*, vol. 68, no. 7, pp. 1087–1105, 2021.
- [14] M. Korda and I. Mezić, "Linear predictors for nonlinear dynamical systems: Koopman operator meets model predictive control," *Automatica*, vol. 93, pp. 149–160, 2018.
- [15] J. L. Proctor, S. L. Brunton, and J. N. Kutz, "Generalizing koopman theory to allow for inputs and control," *SIAM Journal on Applied Dynamical Systems*, vol. 17, no. 1, pp. 909–930, 2018.
- [16] S. L. Brunton, B. W. Brunton, J. L. Proctor, and J. N. Kutz, "Koopman invariant subspaces and finite linear representations of nonlinear dynamical systems for control," *PLoS one*, vol. 11, no. 2, p. e0150171, 2016.
- [17] B. O. Koopman, "Hamiltonian systems and transformation in hilbert space," *Proceedings of the National Academy of Sciences*, vol. 17, no. 5, pp. 315–318, 1931.
- [18] B. O. Koopman and J. v. Neumann, "Dynamical systems of continuous spectra," *Proceedings of the National Academy of Sciences*, vol. 18, no. 3, pp. 255–263, 1932.
- [19] N. Takeishi, Y. Kawahara, and T. Yairi, "Learning koopman invariant subspaces for dynamic mode decomposition," *Advances in neural information processing systems*, vol. 30, 2017.
- [20] B. Lusch, J. N. Kutz, and S. L. Brunton, "Deep learning for universal linear embeddings of nonlinear dynamics," *Nature communications*, vol. 9, no. 1, p. 4950, 2018.
- [21] E. Yeung, S. Kundu, and N. Hodas, "Learning deep neural network representations for koopman operators of nonlinear dynamical systems," in *2019 American Control Conference (ACC)*. IEEE, 2019, pp. 4832–4839.
- [22] A. Mauroy, Y. Susuki, and I. Mezić, *Koopman operator in systems and control*. Springer, 2020.
- [23] W. Manzoor, S. Rawashdeh, and A. Mohammadi, "Vehicular applications of koopman operator theory—a survey," *IEEE Access*, 2023.
- [24] E. Ling, L. Zheng, L. J. Ratliff, and S. Coogan, "Koopman operator applications in signalized traffic systems," *IEEE Transactions on Intelligent Transportation Systems*, vol. 23, no. 4, pp. 3214–3225, 2020.
- [25] H. Shi and M. Q.-H. Meng, "Deep koopman operator with control for nonlinear systems," *IEEE Robotics and Automation Letters*, vol. 7, no. 3, pp. 7700–7707, 2022.
- [26] G. Mamakoukas, M. L. Castano, X. Tan, and T. D. Murphey, "Derivative-based koopman operators for real-time control of robotic systems," *IEEE Transactions on Robotics*, vol. 37, no. 6, pp. 2173–2192, 2021.
- [27] D. Bruder, C. D. Remy, and R. Vasudevan, "Nonlinear system identification of soft robot dynamics using koopman operator theory," in *2019 International Conference on Robotics and Automation (ICRA)*. IEEE, 2019, pp. 6244–6250.
- [28] D. Bruder, X. Fu, R. B. Gillespie, C. D. Remy, and R. Vasudevan, "Data-driven control of soft robots using koopman operator theory," *IEEE Transactions on Robotics*, vol. 37, no. 3, pp. 948–961, 2020.
- [29] J. Wang, B. Xu, J. Lai, Y. Wang, C. Hu, H. Li, and A. Song, "An improved koopman-mpc framework for data-driven modeling and control of soft actuators," *IEEE Robotics and Automation Letters*, vol. 8, no. 2, pp. 616–623, 2022.
- [30] C. Folkestad, D. Pastor, and J. W. Burdick, "Episodic koopman learning of nonlinear robot dynamics with application to fast multirotor landing," in *2020 IEEE International Conference on Robotics and Automation (ICRA)*. IEEE, 2020, pp. 9216–9222.
- [31] C. Ren, H. Jiang, C. Li, W. Sun, and S. Ma, "Koopman-operator-based robust data-driven control for wheeled mobile robots," *IEEE/ASME Transactions on Mechatronics*, vol. 28, no. 1, pp. 461–472, 2022.
- [32] J. J. Koh, G. Ding, C. Heckman, L. Chen, and A. Roncone, "Cooperative control of mobile robots with stackelberg learning," in *2020 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*. IEEE, 2020, pp. 7985–7992.
- [33] Y. Zhao, B. Huang, J. Yu, and Q. Zhu, "Stackelberg strategic guidance for heterogeneous robots collaboration," in *2022 International Conference on Robotics and Automation (ICRA)*. IEEE, 2022, pp. 4922–4928.
- [34] J. F. Fisac, E. Bronstein, E. Stefansson, D. Sadigh, S. S. Sastry, and A. D. Dragan, "Hierarchical game-theoretic planning for autonomous vehicles," in *2019 International conference on robotics and automation (ICRA)*. IEEE, 2019, pp. 9590–9596.
- [35] D. Sadigh, S. Sastry, S. A. Seshia, and A. D. Dragan, "Planning for autonomous cars that leverage effects on human actions," in *Robotics: Science and systems*, vol. 2. Ann Arbor, MI, USA, 2016, pp. 1–9.
- [36] S. Nikolaidis, S. Nath, A. D. Procaccia, and S. Srinivasa, "Game-theoretic modeling of human adaptation in human-robot collaboration," in *Proceedings of the 2017 ACM/IEEE international conference on human-robot interaction*, 2017, pp. 323–331.
- [37] H. Zhang, W. Chen, Z. Huang, M. Li, Y. Yang, W. Zhang, and J. Wang, "Bi-level actor-critic for multi-agent coordination," in *Proceedings of the AAAI Conference on Artificial Intelligence*, vol. 34, no. 05, 2020, pp. 7325–7332.
- [38] L. Meng, J. Ruan, D. Xing, and B. Xu, "Learning in bi-level markov games," in *2022 International Joint Conference on Neural Networks (IJCNN)*. IEEE, 2022, pp. 1–8.
- [39] V. Könönen, "Asymmetric multiagent reinforcement learning," *Web Intelligence and Agent Systems: An international journal*, vol. 2, no. 2, pp. 105–121, 2004.
- [40] J. L. Proctor, S. L. Brunton, and J. N. Kutz, "Dynamic mode decomposition with control," *SIAM Journal on Applied Dynamical Systems*, vol. 15, no. 1, pp. 142–161, 2016.