An Inverse Optimal Control Approach for Learning and Reproducing Under Uncertainties

Sooyung Byeon, Dawei Sun, *Member, IEEE*, and Inseok Hwang, *Member, IEEE*

***Abstract*—This letter presents a novel inverse optimal control (IOC) approach that can account for uncertainties in measurements and system models. The proposed IOC approach aims to recover an objective function including a time-varying term, called variability, from a given demonstration. All uncertainties of the demonstration and the system model can be lumped into the variability such that the optimality condition violation is further reduced. The inferred objective function including the variability has two advantages over the objective function inferred by existing IOC approaches: first, the variability can enhance the capability of describing the given demonstration since it represents how the uncertainties of the system affect the objective function; and second, the proposed IOC approach can reproduce the trajectories such that we can predict the behavior of the system even with system modeling errors. We show that the variability exists and is unique under attainable assumptions. Illustrative numerical examples are presented to demonstrate the proposed method.**

***Index Terms*—Inverse optimal control, lumped uncertainty, variability, trajectory reproduction.**

I. Introduction

THIS letter presents a novel inverse optimal control (IOC) approach for *learning from an observed demonstration* that is perturbed by unknown uncertainties and *reproducing trajectories* in different situations. The IOC has been successfully applied to inferring an objective (i.e., cost or reward) function such that a demonstration (i.e., trajectory [1] or sequence [2]) is optimal for the inferred objective function. Since the objective function is a succinct representation of the demonstration, the IOC has been widely accepted in robotics [3], economics [2], human motor skills modeling [4], and human intent inference [5]. Two other imitation learning categories, learning trajectory [6] and learning control policy [7], are also useful, but they require a diversity of observed demonstrations and are limited to generalizing the learned information in unseen situations.

Manuscript received 15 September 2022; revised 4 November 2022; accepted 24 November 2022. Date of publication 6 December 2022; date of current version 13 December 2022. This work was supported by NSF under Grant CNS-1836952. Recommended by Senior Editor G. Cherubini. *(Corresponding author: Sooyung Byeon.)*

The authors are with the School of Aeronautics and Astronautics, Purdue University, West Lafayette, IN 47907 USA (e-mail: sbyeon@purdue.edu; sun289@purdue.edu; ihwang@purdue.edu).

Digital Object Identifier 10.1109/LCSYS.2022.3226882

The state-of-the-art IOC works have encompassed the data-driven approach to explain a demonstration in the optimal control framework. There are two categories of the recent IOC approaches based on the necessity for solving the (forward) optimal control problem in the learning process [1], [2]. The first category of the IOC approaches, called the *bilevel method*, iteratively updates the objective function and solves the optimal control problem for an updated objective function to fit the demonstration into the optimal solution [8]. However, this letter is computationally demanding when solving the optimal control problem iteratively. Motivated by this limitation, the second category of the IOC approaches, called the *minimum principle method*, utilizes the optimality conditions which must be satisfied by the demonstration. The Karush-Kuhn-Tucker (KKT) condition based method [9] enables us to parameterize the unknown objective function from the full trajectory of the optimal control problem. The Pontryagin minimum principle based method [1], [10] only requires an incomplete trajectory, and thus, it can solve the IOC problem with an unknown finite-time horizon or infinite-time horizon. An online method for control constrained IOC problem is also available [11].

However, there are two major limitations to the current IOC approaches in general. First, existing IOC formulations rarely identify uncertainties explicitly. For the linear quadratic regulator (LQR) case, it is possible to parameterize the objective function such that the inferred parameter is statistically consistent [12]. However, for more general cases, this approach might not be applicable. One noticeable approach for the nonlinear system dynamics is to infer a feasible set of the objective functions [13]. Finding all feasible objective functions may improve the ability to describe the system and demonstration, but the method [13] relies on a complex algorithm and it finds a conservative set. Second, existing IOC works have rarely reproduced trajectories of the system under uncertainties. *Reproducing trajectories* means to retrieve an encoded optimal control strategy learned by the IOC approach to predict the behavior of the system in a new situation [6]. The majority of the IOC studies have focused on inferring a true objective function from the given demonstration and knowledge of a nominal system. However, if the nominal system model includes an unknown modeling error, the true objective function and the nominal system model might fail to predict the system behavior. A model-free IOC approach [14] has been studied, but it accounts for the LQR case only.

We propose a novel IOC approach for *learning from demonstration* and *reproducing the system's trajectory* by inferring an objective function with a time-varying term, called *variability*. By introducing the variability, we can enhance the descriptiveness of the IOC approach to the given demonstration while avoiding using a complex set of parameters (e.g., over interpolation). The variability is designed to account for the combined effects of all uncertainties that induce optimality violation, motivated by the idea of lumped uncertainty in robust control [15]. Also, the inferred objective function and the variability can be utilized to predict the behavior of the system in different situations, e.g., different initial conditions. Our main contribution is to formulate and solve a novel IOC approach that enhances the capability for interpreting the demonstration and can describe the system behaviors in unseen situations. Furthermore, the range of the variability provided by the proposed approach can be used to reproduce the trajectory tubes [16] which assess the variation in the trajectories caused by uncertainties, and are useful for some applications related to safety verification.

The rest of this letter is organized as follows. Section II formulates the IOC problem. Section III presents the novel IOC approach using the variability. We demonstrate numerical examples in Section IV. Section V concludes the study.

II. Inverse Optimal Control Formulation

We consider the nonlinear difference equation for the discrete-time system model:

$$\mathbf{x}_{k+1} = f^*(\mathbf{x}_k, \mathbf{u}_k) \tag{1}$$

where $\mathbf{x}_k \in \mathbb{R}^n$ denotes the state, $\mathbf{u}_k \in U \subset \mathbb{R}^m$ denotes the control input within the feasible convex set U, and $k \in \mathbb{Z}$ is the time step. The true system model $f^* : \mathbb{R}^n \times U \to \mathbb{R}^n$ is assumed to be continuous and differentiable. An unknown objective function is defined as:

$$J(\mathbf{x}_{[0,T]}, \mathbf{u}_{[0,T]}, \theta^*) \triangleq g_T(\mathbf{x}_T) + \sum_{k=0}^{T-1} \theta^{*T} g(\mathbf{x}_k, \mathbf{u}_k) \tag{2}$$

where T denotes the time horizon, $\mathbf{x}_{[0,T]}$ and $\mathbf{u}_{[0,T]}$ denote the state trajectory and control trajectory for $0 \le k \le T$, respectively. Let $\mathbf{u}_T = 0$ without loss of generality. $\theta^* \in \Theta \subset \mathbb{R}^r$ denotes the true parameter of the objective function with a feasible parameter set Θ, and $g : \mathbb{R}^n \times U \to \mathbb{R}^r$ is the basis function. $g_T : \mathbb{R}^n \to \mathbb{R}$ denotes the final state objective function. The final (goal) state can be specified by $\mathbf{x}_T = \mathbf{x}_{goal} \in \mathbb{R}^n$ or be free. The objective function is a linear combination of the parameter and the basis function which has been widely accepted in the IOC research [1], [8], [11] since it makes the computation easier while maintaining the generality of the objective function. The optimal control problem is:

$$\begin{aligned} \inf_{\mathbf{u}_{[0,T]}} \quad & J(\mathbf{x}_{[0,T]}, \mathbf{u}_{[0,T]}, \theta^*) \\ \text{such that} \quad & \mathbf{x}_{k+1} = f^*(\mathbf{x}_k, \mathbf{u}_k), \quad \mathbf{u}_k \in U, \\ & \forall k \in [0, T], \quad \mathbf{x}_0 \in \mathbb{R}^n \end{aligned} \tag{3}$$

where $\mathbf{x}_0$ denotes the initial condition and is known. Let $\{\mathbf{x}^*_{[0,T]}, \mathbf{u}^*_{[0,T]}\}$ be the solution of the optimal control problem (3) and $\{\mathbf{x}_{[0,T]}, \mathbf{u}_{[0,T]}\}$ be the observed demonstration that is perturbed by measurement noise.

The IOC problem is defined as inferring an optimal parameter which minimizes the optimality violation from a potentially incomplete and noisy demonstration $\{\mathbf{x}_{[0,l]}, \mathbf{u}_{[0,l]}\}$ where $0 \le l \le T$ and T can be unknown to the IOC problem. Note that a special case with $l = T$ (complete trajectory with known time horizon) can be handled without loss of generality [2]. For the rest of this letter, we assume the true system model f^* is unknown for the IOC problem but only a nominal system model that includes the system modeling error

$$\mathbf{x}_{k+1} = f(\mathbf{x}_k, \mathbf{u}_k) \tag{4}$$

is known where $f : \mathbb{R}^n \times U \to \mathbb{R}^n$ is assumed to be continuous and differentiable. The Pontryagin minimum principle provides necessary conditions for the optimal control problem (3) [11]. The Hamiltonian H is given as:

$$H(\mathbf{x}_k, \mathbf{u}_k, \lambda_{k+1}, \theta) \triangleq \theta^T g(\mathbf{x}_k, \mathbf{u}_k) + \lambda_{k+1}^T f(\mathbf{x}_k, \mathbf{u}_k) \tag{5}$$

for all $k \ge 0$, $\theta \in \Theta$ denotes the nominal parameter of the objective function, and $\lambda_{k+1} \in \mathbb{R}^n$ denotes the costate. For simplicity, let us use abbreviations $H_k \triangleq H(\mathbf{x}_k, \mathbf{u}_k, \lambda_{k+1}, \theta)$, $g_k \triangleq g(\mathbf{x}_k, \mathbf{u}_k)$, and $f_k \triangleq f(\mathbf{x}_k, \mathbf{u}_k)$. The minimum principle gives the following necessary conditions for optimality.

$$\lambda_k = \nabla_{\mathbf{x}} H_k = \nabla_{\mathbf{x}} g_k \theta + \nabla_{\mathbf{x}} f_k \lambda_{k+1}, \quad \forall k \in [0, l-1] \tag{6}$$

$$0 = \nabla_{\mathbf{u}} H_k = \nabla_{\mathbf{u}} g_k \theta + \nabla_{\mathbf{u}} f_k \lambda_{k+1}, \quad \forall k \in K_l \tag{7}$$

where $\nabla_{\mathbf{x}} H_k \in \mathbb{R}^n$ and $\nabla_{\mathbf{u}} H_k \in \mathbb{R}^m$ denote the partial derivatives of the Hamiltonian for $\mathbf{x}$ and $\mathbf{u}$, respectively. $\nabla_{\mathbf{x}} g_k \in \mathbb{R}^{n \times r}$, $\nabla_{\mathbf{x}} f_k \in \mathbb{R}^{n \times n}$, $\nabla_{\mathbf{u}} g_k \in \mathbb{R}^{m \times r}$, and $\nabla_{\mathbf{u}} f_k \in \mathbb{R}^{m \times n}$ are similarly defined. K_l is defined as follows.

Definition 1 (Inactive Control Constraint Time Steps [11]): A set of time steps K_l is defined as

$$K_l \triangleq \{k : 0 \le k \le l, \mathbf{u}_k \in \mathrm{int} U\} \tag{8}$$

where $\mathrm{int} U$ denotes the interior of the control constraint U.

In general, the IOC problem is defined as estimating θ that minimizes the violation of the optimality condition (7) from the noisy demonstration $\{\mathbf{x}_{[0,l]}, \mathbf{u}_{[0,l]}\}$, nominal system model f with system modeling error, and basis function g [2]:

$$\begin{aligned} \inf_{\theta, \lambda_0} \quad & \sum_{k \in K_l} \|\nabla_{\mathbf{u}} H_k\|^2 \\ \text{such that} \quad & \lambda_k = \nabla_{\mathbf{x}} H_k, \quad \forall k \in [0, l-1] \end{aligned} \tag{9}$$

where $\|\cdot\|$ denotes the Euclidean norm.

We present an existing solution to the general IOC problem to utilize the solution for the proposed IOC approach and to make this letter self-contained. We assume the following to reformulate the IOC problem as a static optimization problem.

Assumption 1 (Jacobian Invertibility): $\nabla_{\mathbf{x}} f_k^* \in \mathbb{R}^{n \times n}$ and $\nabla_{\mathbf{x}} f_k \in \mathbb{R}^{n \times n}$ are invertible for all $k \ge 0$.

Note that Assumption 1 is widely accepted in the IOC studies [1], [2]. If Assumption 1 holds, the optimality conditions in (6) and (7) can be rewritten as

$$\mathbf{z}_{k+1} \triangleq \begin{bmatrix} \theta \\ \lambda_{k+1} \end{bmatrix} = \begin{bmatrix} I & 0 \\ -\nabla_{\mathbf{x}} f_k^{-1} \nabla_{\mathbf{x}} g_k & \nabla_{\mathbf{x}} f_k^{-1} \end{bmatrix} \begin{bmatrix} \theta \\ \lambda_k \end{bmatrix} \triangleq G_k \mathbf{z}_k \tag{10}$$

$$\begin{bmatrix}\nabla_{\mathbf{u}} g_k & \nabla_{\mathbf{u}} f_k\end{bmatrix}\begin{bmatrix}\theta \\ \lambda_{k+1}\end{bmatrix} \triangleq F_k \mathbf{z}_{k+1} = 0 \tag{11}$$

for all $k \in K_l$. The mixed constraint approach [2], [11] allows a slight violation in (7) while (6) holds exactly. To solve the mixed constraint IOC problem, another objective function:

$$J_l(\mathbf{z}_0) \triangleq \sum_{k\in K_l} \|F_k \mathcal{G}_k \mathbf{z}_0\|^2 = \mathbf{z}_0^T \mathcal{Q}_l \mathbf{z}_0 \tag{12}$$

where

$$\mathcal{G}_k \triangleq \prod_{i=0}^{k} G_i \in \mathbb{R}^{(n+r)\times(n+r)} \tag{13}$$

$$\mathcal{Q}_l \triangleq \sum_{k\in K_l} (F_k \mathcal{G}_k)^T (F_k \mathcal{G}_k) \in \mathbb{R}^{(n+r)\times(n+r)} \tag{14}$$

is defined and $\mathcal{Q}_l$ is a positive semidefinite matrix. It is known that $rank(\mathcal{Q}_l)$ is monotonically increasing for time step $k \geq 0$, and $\mathcal{Q}_l$ has full rank if the given demonstration is informative [2]. If $\mathcal{Q}_l$ has full rank, a unique IOC solution θ can be determined by solving the following quadratic programming:

$$\begin{aligned}\inf_{\mathbf{z}_0=[\theta^T\ \lambda_0^T]^T} \ & J_l(\mathbf{z}_0) \quad \text{subject to} \quad \mathcal{I}\mathbf{z}_0 = \theta \in \Theta \\ & \Theta = \{\theta \in \mathbb{R}^r : \|\theta\|^2 = 1, \theta^1 > 0\}\end{aligned} \tag{15}$$

where $\mathcal{I} \triangleq [I \quad 0] \in \mathbb{R}^{r\times(n+r)}$ and $\|\theta\|^2 = 1$ denotes the normalization of θ to eliminate the scalar ambiguity [14]. θ^1 denotes the first element of θ and we assume that the first basis function is relevant to the objective function in (2) and thus, θ^1 is nonzero. Note that (12)-(15) represent a solution method for (9) [2].

Although the solution to (15) provides a reasonable estimate θ, there are two limitations. First, the IOC approach only provides a (potentially biased) estimate of the parameter θ that could be too restrictive to describe the objective function under uncertainties. For instance, for an intent inference problem in [5], an agent's intent is mapped into a single parameter θ. However, if the demonstration is perturbed by uncertainties, the parameter could be inconsistent even if the intent is consistent. Instead of mapping a single parameter to an intent, it is more reasonable to associate each intent with a set of parameters to assess the effect of uncertainties. Second, the existing IOC approaches have rarely considered the trajectory reproduction problem with a time-varying objective parameter. From an application perspective, the IOC can be used to predict the future behavior of the system in a new situation based on the given demonstration [17], [18]. However, if the nominal system model f is perturbed by the modeling error, even if the true parameter θ^* is known, an accurate prediction is not feasible.

III. Variability for Lumping Uncertainties

A. Problem Reformulation

We generalize the objective function in (2) by introducing the variability $\Delta\theta_k \in \mathbb{R}^r$ to define a new IOC problem. The variability is designed not only for improving the descriptiveness of the objective function but also maintaining the structure of the objective function, i.e., a weighted combination of the basis function. This design prevents the use of overly complex parameters to describe an observed demonstration from the IOC perspective, and thus, it provides a good balance between the descriptiveness and complexity of the objective function. The variability can ameliorate the capability of reproducing trajectories by encoding the uncertainties that are reproducible under similar situations. Also, we can exploit an inferred range of variability to reproduce a set of trajectories, which can assess the variation of system behaviors under uncertainties. We define the time-varying parameter $\theta_k \in \mathbb{R}^r$:

$$\theta_k \triangleq \theta + \Delta\theta_k, \quad \forall k \in [0, T] \tag{16}$$

where θ is the solution to (15). Note that the proposed method relies on the existing IOC method in Section II to obtain θ as an initial point. Then, the objective function with θ_k is:

$$J_{\Delta\theta}(\mathbf{x}_{[0,T]}, \mathbf{u}_{[0,T]}, \theta_{[0,T]}) \triangleq g_T(\mathbf{x}_T) + \sum_{k=0}^{T-1} \theta_k^T g(\mathbf{x}_k, \mathbf{u}_k) \tag{17}$$

and the Hamiltonian in (5) is also redefined accordingly:

$$\bar{H}(\mathbf{x}_k, \mathbf{u}_k, \lambda_{k+1}, \theta_k) \triangleq \theta_k^T g(\mathbf{x}_k, \mathbf{u}_k) + \lambda_{k+1} f(\mathbf{x}_k, \mathbf{u}_k) \tag{18}$$

that can be simplified as $\bar{H}_k \triangleq \bar{H}(\mathbf{x}_k, \mathbf{u}_k, \lambda_{k+1}, \theta_k)$. Now, we propose a feasibility problem for finding the objective parameter $\theta_k = \theta + \Delta\theta_k, \forall k \in [0, l]$, where θ solves (15).

$$\begin{aligned}\inf_{\Delta\theta_{[0,l]}, \lambda_{[0,l+1]}} \ & C \\ \text{such that} \ & \lambda_k = \nabla_{\mathbf{x}} \bar{H}_k, \quad \forall k \in [0, l] \\ & 0 = \nabla_{\mathbf{u}} \bar{H}_k, \quad \forall k \in K_l\end{aligned} \tag{19}$$

for a constant $C \in \mathbb{R}$. A key difference from the existing IOC formulation in (9) is that the minimum principle violation can be further reduced even if the given demonstration is perturbed by uncertainties. If Assumption 1 holds, the costate can be propagated by:

$$\lambda_{k+1} = -\nabla_{\mathbf{x}} f_k^{-1} \nabla_{\mathbf{x}} g_k \theta + \nabla_{\mathbf{x}} f_k^{-1} \lambda_k, \quad \forall k \in [0, l] \tag{20}$$

where λ_0 is the solution to the IOC problem (15). Then, we can derive two constraints to satisfy the minimum principle exactly with the variability:

$$\lambda_k = \nabla_{\mathbf{x}} H_k = \nabla_{\mathbf{x}} \bar{H}_k - \nabla_{\mathbf{x}} g_k \Delta\theta_k, \quad \forall k \in [0, l] \tag{21}$$

$$\begin{aligned}0 = \nabla_{\mathbf{u}} \bar{H}_k &= \nabla_{\mathbf{u}} H_k + \nabla_{\mathbf{u}} g_k \Delta\theta_k \\ &= \nabla_{\mathbf{u}} g_k \theta + \nabla_{\mathbf{u}} f_k \lambda_{k+1} + \nabla_{\mathbf{u}} g_k \Delta\theta_k, \ \forall k \in K_l.\end{aligned} \tag{22}$$

Note that two different IOC problems (9) and (19) can have the same costate dynamics, i.e., $\lambda_k = \nabla_{\mathbf{x}} H_k = \nabla_{\mathbf{x}} \bar{H}_k$ by imposing a designed constraint $\nabla_{\mathbf{x}} g_k \Delta\theta_k = 0$. Equations (21) and (22) can be reformulated for all $k \in K_l$:

$$\begin{bmatrix}\nabla_{\mathbf{x}} g_k \\ \nabla_{\mathbf{u}} g_k\end{bmatrix} \Delta\theta_k \triangleq \nabla g_k \Delta\theta_k = \begin{bmatrix}0 \\ -\nabla_{\mathbf{u}} g_k \theta - \nabla_{\mathbf{u}} f_k \lambda_{k+1}\end{bmatrix} \triangleq \mathbf{v}_k \tag{23}$$

where $\nabla g_k \in \mathbb{R}^{(n+m)\times r}$ and $\mathbf{v}_k \in \mathbb{R}^{n+m}$ can be determined using the IOC solution $\{\theta, \lambda_0\}$ in (15), demonstration $\{\mathbf{x}_{[0,l]}, \mathbf{u}_{[0,l]}\}$, and the costate propagated by (20). We define the variability as the minimum weighted norm vector:

$$\Delta\theta_k = \arg\min_{\Delta\theta_k} \|\Delta\theta_k\|_W^2 \quad \text{subject to} \quad (23) \tag{24}$$

where $\|\Delta\theta_k\|_W^2 \triangleq \Delta\theta_k^T W \Delta\theta_k$. $W \in \mathbb{R}^{r\times r}$ denotes the user defined weight matrix and is positive definite.

B. Existence and Uniqueness of Variability

To guarantee the existence and uniqueness of the variability $\Delta\theta_k$, we assume and define the followings.

Assumption 2 (The Dimension of Basis Function): The dimension of basis function r is equal to or greater than the sum of state and control input dimensions $n+m$, i.e., $n+m \leq r$. Note that it is easy to satisfy Assumption 2 since the dimension of the basis function can be arbitrarily large at the cost of more time steps required to meet the rank condition $rank(\mathcal{Q}_l) = n+r$ [11]. For an irrelevant element of the basis function g^i, the corresponding parameter is zero, i.e., $\theta^{*i} = 0$.

Definition 2 (Full Rank Time Steps): Let Assumption 2 hold. A set of time steps R_l is defined as

$$R_l \triangleq \{k : 0 \leq k \leq l, rank(\nabla g_k) = n+m\} \tag{25}$$

and $k \in R_l$ implies that ∇g_k has full rank at time step k.

The full rank condition of ∇g_k is not restrictive if the demonstration is informative (i.e., the persistence of excitation [2]). Any arbitrary basis function element can be added to make ∇g_k have full rank as discussed in Assumption 2. Then, the existence and uniqueness of the variability can be shown.

Theorem 1: Let Assumption 2 hold and the variability $\Delta\theta_k$ follows (23) and (24). Then, for all $k \in K_l \cap R_l$, the variability $\Delta\theta_k$ exists and is unique as:

$$\Delta\theta_k = W^{-1}\nabla g_k^T(\nabla g_k W^{-1}\nabla g_k^T)^{-1}\mathbf{v}_k. \tag{26}$$

For all $k \notin K_l \cap R_l$, let $\Delta\theta_k = 0$ without loss of generality.

Proof: (Existence) Let $\nabla g_k^\dagger \in \mathbb{R}^{r\times(n+m)}$ be the Moore–Penrose pseudoinverse of the matrix ∇g_k. It is known that the linear system (23) has at least one solution $\Delta\theta_k$ if and only if $\nabla g_k \nabla g_k^\dagger \mathbf{v}_k = \mathbf{v}_k$ [2], [19]. Since Assumption 2 holds, for all $k \in R_l$, $\nabla g_k \nabla g_k^\dagger = I_r \in \mathbb{R}^{r\times r}$. Thus, at least one $\Delta\theta_k$ satisfies (23).

(Uniqueness) We have shown that there is at least one $\Delta\theta_k$ such that (23) is satisfied. Among all general solutions, the minimum weighted norm solution to (24) is uniquely given as (26) in [15] (Chap. 3). ∎

Remark 1: The derivation of (26) is based on the assumption that the variability $\Delta\theta_k$ is smaller than the parameter estimate θ, i.e., $\|\Delta\theta_k\|_W < \mu\|\theta\|_W$ where $\mu < 1$ is a small real positive constant. If $\|\Delta\theta_k\|_W$ is too large, it implies the basis function is inadequate or uncertainties are too severe. The variability reveals the degree of optimality violation of the given system and demonstration by its norm.

Remark 2: The variability depends on $\{\mathbf{x}_k, \mathbf{u}_k\}$ since the constraints (23) depend on the demonstration. Identifying an explicit function $\Delta\theta_k(\mathbf{x}_k, \mathbf{u}_k)$ from multiple demonstrations could be an interesting future work.

C. Reproducing Trajectory

To reproduce a new trajectory from the learned parameter θ_k, we present an optimal control problem. A standard optimal control problem in (3) can be solved with a trajectory optimization tool (e.g., CasADi [20]). However, the proposed time-varying objective parameter in (17) is usually not well fitted into the standard optimization problem, and thus we need to address the trajectory reproduction. Let $\bar{\mathbf{x}}_k \triangleq [\mathbf{x}_k^T \quad k]^T \in \mathbb{R}^{n+1}$ and $\bar{f} : \mathbb{R}^{n+1} \times U \to \mathbb{R}^{n+1}$ be the augmented state and the augmented system model, respectively:

$$\bar{\mathbf{x}}_{k+1} = \begin{bmatrix}\mathbf{x}_{k+1}\\ k+1\end{bmatrix} = \begin{bmatrix}f(\mathbf{x}_k, \mathbf{u}_k)\\ k+1\end{bmatrix} \triangleq \bar{f}(\bar{\mathbf{x}}_k, \mathbf{u}_k). \tag{27}$$

The time dependent variability θ_k can be formulated as the function of the augmented state, i.e., $\theta_k = \theta(\bar{\mathbf{x}}_k)$, where $\theta(\bar{\mathbf{x}}_k)$ is a continuous and differentiable function. The *piecewise cubic Hermite interpolating polynomial* can be used to preserve the shape and guarantee differentiability of θ_k [21]. The objective function (17) can be evaluated by

$$\bar{J}_{\Delta\theta}(\bar{\mathbf{x}}_{[0,T]}, \mathbf{u}_{[0,T]}, \theta_{[0,T]}) \triangleq g_T(\mathbf{x}_T) + \sum_{k=0}^{T-1}\theta(\bar{\mathbf{x}}_k)^T g(\mathbf{x}_k, \mathbf{u}_k). \tag{28}$$

Let $\tilde{\mathbf{x}}_0$ be the initial state of the trajectory reproduction. We aim to reproduce a whole sequence $\{\tilde{\bar{\mathbf{x}}}_{[0,T]}, \tilde{\mathbf{u}}_{[0,T]}\}$ by solving:

$$\begin{aligned}\inf_{\tilde{\mathbf{u}}_{[0,T]}} \quad & \bar{J}_{\Delta\theta}(\tilde{\bar{\mathbf{x}}}_{[0,T]}, \tilde{\mathbf{u}}_{[0,T]}, \theta_{[0,T]})\\ \text{such that} \quad & \tilde{\bar{\mathbf{x}}}_{k+1} = \bar{f}(\tilde{\bar{\mathbf{x}}}_k, \tilde{\mathbf{u}}_k), \quad \tilde{\mathbf{u}}_k \in U\\ & \forall k \in [0, T], \quad \tilde{\mathbf{x}}_0 \in \mathbb{R}^n\end{aligned} \tag{29}$$

where $\tilde{\bar{\mathbf{x}}}_k \triangleq [\tilde{\mathbf{x}}_k^T \quad k]^T \in \mathbb{R}^{n+1}$.

IV. Numerical Examples

To evaluate the proposed IOC approach, we present numerical examples with a two-link robot arm shown in Fig. 1. We employ the same physical dimension from [1], [21]: the length of each arm $l_1 = l_2 = 1\ m$; the length from each joint to the center of each arm $r_1 = r_2 = 0.5\ m$; the mass of each arm $m_1 = m_2 = 1\ kg$; and the moment of inertia of each arm $I_1 = I_2 = 0.5\ kg\cdot m^2$. The state, control input, and true system model are given as [22]:

$$\mathbf{x} = \begin{bmatrix}\alpha_1 & \alpha_2 & \dot{\alpha}_1 & \dot{\alpha}_2\end{bmatrix}^T, \quad \mathbf{u} = \begin{bmatrix}\tau_1 & \tau_2\end{bmatrix}^T \tag{30}$$

$$M_r(\boldsymbol{\alpha})\ddot{\boldsymbol{\alpha}} + C_r(\boldsymbol{\alpha}, \dot{\boldsymbol{\alpha}})\dot{\boldsymbol{\alpha}} + g_r(\boldsymbol{\alpha}) = \boldsymbol{\tau} \tag{31}$$

where $\boldsymbol{\alpha} = [\alpha_1 \quad \alpha_2]^T \in \mathbb{R}^2$ and $\dot{\boldsymbol{\alpha}} = [\dot{\alpha}_1 \quad \dot{\alpha}_2]^T \in \mathbb{R}^2$ denote the angle and angular velocity of each arm in *rad* and *rad*/*s*, respectively. $\boldsymbol{\tau} = [\tau_1 \quad \tau_2]^T \in \mathbb{R}^2$ denotes the torque input to each arm in $N\cdot m$. $M_r(\boldsymbol{\alpha}) \in \mathbb{R}^{2\times2}$ is the positive definite inertia matrix, $C_r(\boldsymbol{\alpha}, \dot{\boldsymbol{\alpha}}) \in \mathbb{R}^{2\times2}$ denotes the Coriolis matrix, and $g_r(\boldsymbol{\alpha}) \in \mathbb{R}^2$ denotes the gravity vector. We discretize the dynamics with time interval $\Delta t = 0.01\ s$ [1] and the total time step is $T = 200$. The control constraint is specified as $U = [-15\ \ 15]^2$. The true objective parameter $\theta^* \in \mathbb{R}^9$ and corresponding basis function are given as (adopted from [1]):

$$\theta^* = \begin{bmatrix}0.884 & 0.442 & 0.147 & 0 & 0 & 0 & 0 & 0 & 0\end{bmatrix}^T \tag{32}$$

$$g_k = \begin{bmatrix}\tau_1^2 & \tau_2^2 & \tau_1\tau_2 & \alpha_1^2 & \alpha_2^2 & \dot{\alpha}_1^2 & \dot{\alpha}_2^2 & \tau_1^3 & \tau_2^3\end{bmatrix}^T \tag{33}$$

with the zero final objective, i.e., $g_T(\mathbf{x}_T) = 0$. The goal state is set to $\mathbf{x}_{goal} = [40^\circ\ \ 50^\circ\ \ 0^\circ/s\ \ 0^\circ/s]^T$. Note that this example holds Assumption 2 since $n+m = 6 < r = 9$.

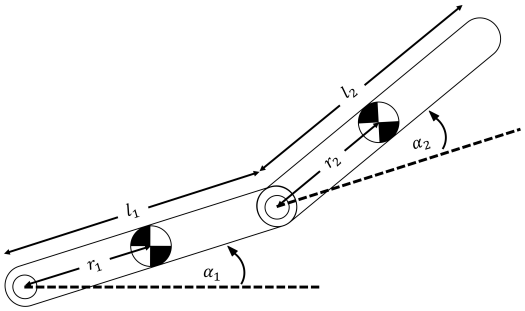

Fig. 1. Two-link robot arm with physical dimension.

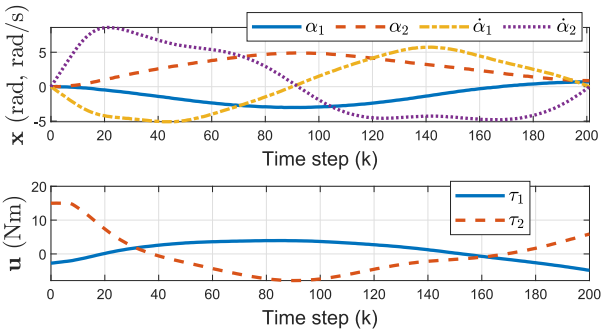

Fig. 2. Observed demonstration with Gaussian noise level $\sigma = 10^{-2}$.

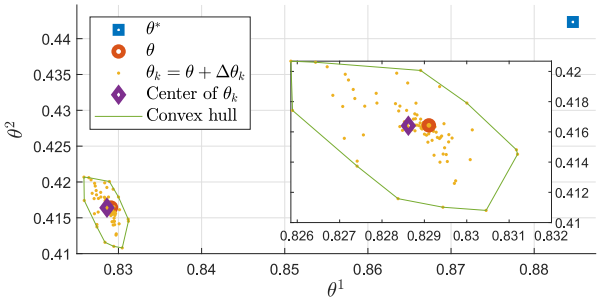

Fig. 3. First and second elements of true objective parameter $(\theta^{*1}, \theta^{*2})$, recovered objective parameter (θ^1, θ^2), and (θ_k^1, θ_k^2).

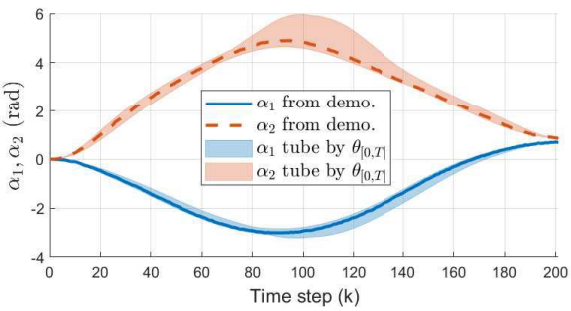

Fig. 4. Trajectory tubes reproduced by the convex hull of the inferred parameters $\boldsymbol{\theta_k}$. The tubes are obtained by reproducing trajectories using $\boldsymbol{\theta_k}$ for all $\boldsymbol{k}$ in convex hull indices and taking minimum and maximum values of all reproduced trajectories for each time step.

A. Learning From Demonstration

We test whether the proposed IOC approach can learn the appropriate objective parameter θ_k from a noisy demonstration and a nominal system f with system modeling errors.

1) Demonstration: Let $\{\mathbf{x}_{[0,T]}, \mathbf{u}_{[0,T]}\}$ be the observed demonstration with zero-mean Gaussian noise. The level of noise in standard deviation is $\sigma = 10^{-2}$ for the state and control input. The initial state is given as $\mathbf{x}_0 = [0\ 0\ 0\ 0]^T$. The demonstration is obtained by solving (3) with (31)-(33). CasADi [20] is used as a trajectory optimization tool and the source codes in [1] are used as a baseline of the numerical example. Fig. 2 shows the demonstration $\{\mathbf{x}_{[0,T]}, \mathbf{u}_{[0,T]}\}$.

2) Learning Objective Parameter: The proposed IOC approach is used to recover the objective parameter θ_k. A nominal system model f is given with 5% errors for $\{l_1, r_1, m_1, I_1\}$ and -5% errors for $\{l_2, r_2, m_2, I_2\}$, respectively. The weight matrix is given as $W = I_9$. θ is recovered by the conventional IOC in (15) and θ_k is recovered by the proposed IOC approach in (19). Fig. 3 shows that the recovered θ_k is represented as a set of distributed points, and a convex hull can show simplified information about the recovered objective parameter. The center of θ_k deviates from θ which means the proposed approach can obtain the center of the feasible parameter set that is not available from the conventional IOC approach. Interestingly, the convex hull is away from the true objective parameter θ^* with $\theta = [0.829\ 0.416\ 0.152\ 0.288\ -0.060\ -0.154\ 0.077\ 0.002\ 0.000]^T$ and θ includes nonzero values for irrelevant basis functions in the true model (32) and (33). The maximum $\|\Delta\theta_k\|_W$ is 0.018, which is smaller than $\|\theta\|_W = 1$, i.e., the volume of the convex hull in Fig. 3 is small, and thus, Remark 1 holds.

3) Reproducing Trajectory Tube From Convex Hull: We numerically examine that the obtained convex hull of θ_k can reproduce trajectory tubes containing the given demonstration. The trajectory tubes [16], which represent a range of the possible system behaviors under uncertainties using the convex hull of the inferred θ_k (in Fig. 3) and the fixed goal state $\mathbf{x}_{\text{goal}}$, are presented. In Fig. 4, the tubes are produced using the convex hull of the parameter θ_k, i.e., taking minimum and maximum of the reproduced states $\tilde{\mathbf{x}}_{[0,T]}$ using θ_k for all k in convex hull indices. The proposed IOC approach can provide the trajectory tube without knowledge of uncertainties. Note that the convex hull can efficiently account for multiple demonstrations from different initial states and different time horizons. Fig. 4 shows that the convex set of the θ_k is effective to predict the system behavior as a range of trajectories. The trajectory tube can be used to examine the safety of the system by inspecting an overlap between the trajectory tubes and dangerous environments (e.g., obstacles).

B. Reproducing in Different Initial Conditions

We reproduce trajectories with random initial states and multiple sets of uncertainties using the conventional IOC approach (15) and the proposed approach (19). For learning the objective parameters, a single demonstration is provided to two IOC approaches (i.e., as a training data set) with $\mathbf{x}_0 = [0\ 0\ 0\ 0]^T$. This demonstration is generated by solving (3) with the true system model f^* and true parameter θ^* and perturbed by multiple levels of Gaussian measurement noise. The true parameter θ^* and the learned parameters θ

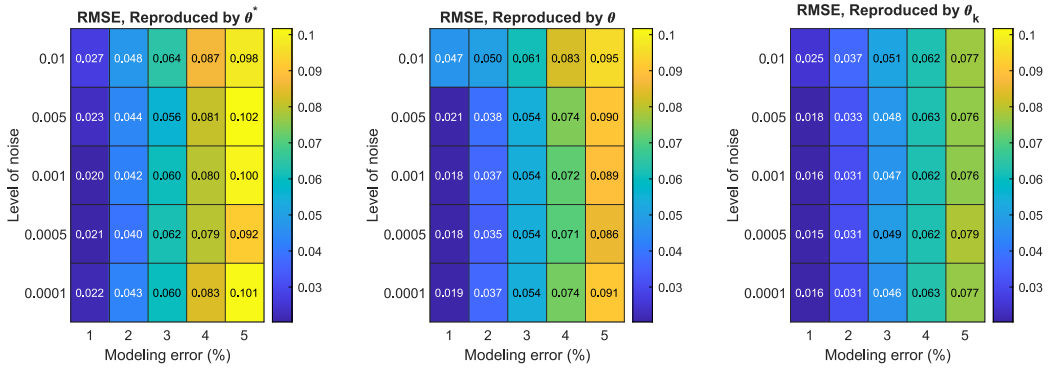

Fig. 5. Monte-Carlo simulation results with varying initial states, levels of noise, and multiple modeling errors. Each cell represents the prediction errors in RMSE between demonstrations $\mathbf{x}_k^{\text{test}}$ (which are not available to IOC approaches) and reproduced trajectories $\tilde{\mathbf{x}}_k$, $\forall \boldsymbol{k} \in [\mathbf{0}, \mathbf{200}]$. Evaluated RMSEs using 30 reproduced trajectories are averaged at each cell.

and θ_k are used for reproducing the system trajectories from multiple initial states, $\tilde{\mathbf{x}}_0 = [x\ y\ 0\ 0]^T$, where x and y are uniformly distributed within $[-5°, 5°]$. The reproduced trajectories $\tilde{\mathbf{x}}_{[0,T]}$ can be obtained by solving the optimal control problem with the nominal system model f with multiple levels of modeling error and three different parameters θ^*, θ, and θ_k, respectively. For comparison, the demonstrations $\mathbf{x}_{[0,T]}^{\text{test}}$ for each initial state $\tilde{\mathbf{x}}_0$ by solving (3) with f^* and θ^* are provided (i.e., testing data sets). A total of 30 reproduced trajectories with respect to each condition are simulated. Fig. 5 shows that the proposed IOC approach can reproduce the trajectories with smaller root mean square errors (RMSEs), compared to using the true objective function and the conventional IOC approach in all ranges of noises and system modeling errors. The evaluated RMSEs show that the objective function inferred by the proposed IOC approach improves the prediction accuracy by 22.36% and 16.16% (mean) compared to the two other approaches, respectively. Fig. 5 clearly shows that the proposed IOC approach can account for uncertainties in trajectory reproduction by encoding the variability.

V. Conclusion

We proposed a novel IOC approach to enhance the capability of the IOC by introducing a time-varying parameter, called the variability, for describing an objective function under uncertainties. The variability provides more information about the objective function and improves the quality of trajectory reproduction. The proposed approach is computationally efficient and guarantees the existence and uniqueness of the variability under reasonable assumptions.

References

[1] W. Jin, D. Kulić, S. Mou, and S. Hirche, "Inverse optimal control from incomplete trajectory observations," *Int. J. Robot. Res.*, vol. 40, nos. 6–7, pp. 848–865, 2021.

[2] T. L. Molloy, J. I. Charaja, S. Hohmann, and T. Perez, *Inverse Optimal Control and Inverse Noncooperative Dynamic Game Theory*. Cham, Switzerland: Springer, 2022.

[3] M. Menner, P. Worsnop, and M. N. Zeilinger, "Constrained inverse optimal control with application to a human manipulation task," *IEEE Trans. Control Syst. Technol.*, vol. 29, no. 2, pp. 826–834, Mar. 2021.

[4] S. Levine and V. Koltun, "Continuous inverse optimal control with locally optimal examples," in *Proc. 29th Int. Coference Int. Conf. Mach. Learn.*, 2012, pp. 475–482.

[5] N. Yokoyama, "Inference of aircraft intent via inverse optimal control including second-order optimality condition," *J. Guid., Control, Dyn.*, vol. 41, no. 2, pp. 349–359, 2018.

[6] S. Calinon, F. D'Halluin, E. L. Sauser, D. G. Caldwell, and A. G. Billard, "Learning and reproduction of gestures by imitation," *IEEE Robot. Autom. Mag.*, vol. 17, no. 2, pp. 44–54, Jun. 2010.

[7] N. Jaquier, D. Ginsbourger, and S. Calinon, "Learning from demonstration with model-based Gaussian process," in *Proc. Conf. Robot Learn.*, vol. 100, 2020, pp. 247–257.

[8] P. Abbeel and A. Y. Ng, "Apprenticeship learning via inverse reinforcement learning," in *Proc. 21st Int. Conf. Mach. Learn.*, 2004, p. 1.

[9] P. Englert, N. A. Vien, and M. Toussaint, "Inverse KKT: Learning cost functions of manipulation tasks from demonstrations," *Int. J. Robot. Res.*, vol. 36, nos. 13–14, pp. 1474–1488, 2017.

[10] W. Jin, D. Kulić, J. F.-S. Lin, S. Mou, and S. Hirche, "Inverse optimal control for multiphase cost functions," *IEEE Trans. Robot.*, vol. 35, no. 6, pp. 1387–1398, Dec. 2019.

[11] T. L. Molloy, J. J. Ford, and T. Perez, "Online inverse optimal control for control-constrained discrete-time systems on finite and infinite horizons," *Automatica*, vol. 120, Oct. 2020, Art. no. 109109.

[12] H. Zhang, J. Umenberger, and X. Hu, "Inverse optimal control for discrete-time finite-horizon linear quadratic regulators," *Automatica*, vol. 110, Dec. 2019, Art. no. 108593.

[13] A. M. Panchea and N. Ramdani, "Inverse parametric optimization in a set-membership error-in-variables framework," *IEEE Trans. Autom. Control*, vol. 62, no. 12, pp. 6536–6543, Dec. 2017.

[14] S. G. Clarke, S. Byeon, and I. Hwang, "A low complexity approach to model-free stochastic inverse linear quadratic control," *IEEE Access*, vol. 10, pp. 9298–9308, 2022.

[15] P. Paraskevopoulos, *Modern Control Engineering*, 1st ed. New York, NY, USA: CRC Press, 2002.

[16] A. B. Kurzhanski and P. Varaiya, *Dynamics and Control of Trajectory Tubes*. Cham, Switzerland: Birkhäuser, 2014.

[17] I. Maroger, O. Stasse, and B. Watier, "Inverse optimal control to model human trajectories during locomotion," *Comput. Methods Biomech. Biomed. Eng.*, vol. 25, no. 5, pp. 499–511, 2022.

[18] D. Clever, R. M. Schemschat, M. L. Felis, and K. Mombaur, "Inverse optimal control based identification of optimality criteria in whole-body human walking on level ground," in *Proc. 6th IEEE Int. Conf. Biomed. Robot. Biomechatronics (BioRob)*, 2016, pp. 1192–1199.

[19] A. Ben-Israel and T. N. E. Greville, *Generalized Inverses: Theory and Applications*. New York, NY, USA: Springer, 2003.

[20] J. A. E. Andersson, J. Gillis, G. Horn, J. B. Rawlings, and M. Diehl, "CasADi: A software framework for nonlinear optimization and optimal control," *Math. Program. Comput.*, vol. 11, no. 3, pp. 1–36, 2019.

[21] S. Ishida, T. Harada, P. Carreno-Medrano, D. Kulić, and G. Venture, "Human motion imitation using optimal control with time-varying weights," in *Proc. IEEE/RSJ Int. Conf. Intell. Robots Syst. (IROS)*, 2021, pp. 608–615.

[22] M. W. Spong and M. Vidyasagar, *Robot Dynamics and Control*. New York, NY, USA: Wiley, 2008.