Proceedings of Machine Learning Research vol 242:375–386, 2024

Inverse Optimal Control as an Errors-in-Variables Problem

Rahel Rickenbach RRAHEL@ETHZ.CH
Anna Scampicchio ASCAMPICC@ETHZ.CH
Melanie N. Zeilinger MZEILINGER@ETHZ.CH
Institute for Dynamic Systems and Control, Leonhardstrasse 21, 8051 Zurich

Abstract

Inverse optimal control (IOC) is about estimating an unknown objective of interest given its optimal control sequence. However, truly optimal demonstrations are often difficult to obtain, e.g., due to human errors or inaccurate measurements. This paper presents an IOC framework for objective estimation from multiple sub-optimal demonstrations in constrained environments. It builds upon the Karush-Kuhn-Tucker optimality conditions, and addresses the Errors-In-Variables problem that emerges from the use of sub-optimal data. The approach presented is applied to various systems in simulation, and consistency guarantees are provided for linear systems with zero mean additive noise, polytopic constraints, and objectives with quadratic features.[1]

Keywords: Inverse optimal control, Errors-in-Variables, Total Least Squares

1. Introduction

Applications in robotics and control often involve complex and demanding tasks in constrained environments. While strategies such as Model Predictive Control (Kouvaritakis and Cannon, 2016) have been successfully deployed to address these challenges, their performance depends heavily on the design of the objective function, which is often nontrivial. In fact, the translation of a complex goal description into a suitable objective is often unintuitive, and its tuning can be delicate. Thus, inverse optimal control (IOC) methods (Lin et al., 2021; Ab Azar et al., 2020) provide a promising tool to tackle this issue: starting from a partially specified objective (e.g., described by a set of basis functions), they aim to estimate the missing parameters from optimal control sequences (also known as *demonstrations*).

However, available control sequences may be suboptimal in practice, as they are, e.g., provided by humans, or affected by noise. The methods in (Englert et al., 2017) and (Menner et al., 2019), which rely on the Karush-Kuhn-Tucker (KKT) conditions (Kuhn and Tucker, 2014), address these issues by allowing for a slight suboptimality in the demonstrations using a least-squares approximation of the stationarity condition. However, the estimate obtained is not consistent: it does not deal with the errors entering the regressors due to the sub-optimality of the given controls – in other words, the Errors-In-Variables (EIV) nature of the problem (Griliches, 1974) is not addressed. A possibility to address this issue consists in jointly estimating demonstrations and optimal objective parameters, as done by (Hatz et al., 2012). The same idea is used in (Menner and Zeilinger, 2020), where a probabilistic model taking Gaussian measurement noises into account is considered. However, the majority of IOC approaches rely on deterministic models and do not consider stochastic noise in the dynamics. Another exception is the approach in (Nakano, 2023), where a stochastic demonstration

1. The datasets generated and/or analysed during the current study are available in the eth research collection repository, https://doi.org/10.3929/ethz-b-000673653.

model is considered, but knowledge of the optimal input sequence is still assumed. An intrinsic probabilistic view of the problem is given by the formulation in Inverse Reinforcement Learning (IRL), where data are modeled as Markov Decision Processes (Feinberg and Shwartz, 2012) and thus intrinsically deal with suboptimal demonstrations. Solutions are typically obtained with entropy maximization (Ziebart et al., 2008; Finn et al., 2016). However, in contrast to IOC, the IRL formulation does not allow for a natural inclusion of constraints.

Contribution This work presents an IOC framework capable of dealing with suboptimal and noisy demonstrations. It starts from a formulation based on the KKT conditions and rephrases it in terms of an EIV problem (Griliches, 1974): to the best of the authors' knowledge, this is novel in the IOC and IRL literature. We present two approaches considering different assumptions to solve it: when the distribution of noises is known, we start from a Bayesian interpretation of the problem (Dellaportas and Stephens, 1995) and rely on Markov Chain Monte Carlo (MCMC) (Gilks et al., 1995), while when no such an information is given, we leverage total least squares (Golub and Van Loan, 1980). Our approach relates to that in (Menner and Zeilinger, 2020), because it jointly estimates demonstration and optimal objective parameters working with probabilistic models; however, differently from that work, we provide principled ways of estimating hyper-parameters, and study the case in which disturbances are non-Gaussian. Our approaches not only allow for an improved estimation performance with respect to the state-of-the-art, as shown in the numerical tests, but also provide consistency guarantees in the case of linear systems with zero-mean additive noise, polytopic constraints and quadratic objectives.

Notation We denote with $\mathbb{0}_{n,m}$ and $\mathbb{1}_{n,m}$ a matrix of dimension $n \times m$ filled with zeros and ones, respectively, and with $\mathbb{I}_n$ the $n-$dimensional identity matrix. Given a vector $a \in \mathbb{R}^b$ and a function $h(a) : \mathbb{R}^b \to \mathbb{R}^c$, then $\nabla_a h(a)$ returns the Jacobian of $h(a)$ with respect to a. A Gaussian random vector $\mathfrak{a}$ with mean $\mu_\mathfrak{a}$ and covariance $\Sigma_\mathfrak{a}$ is given as $\mathcal{N}(\mathfrak{a}; \mu_\mathfrak{a}, \Sigma_\mathfrak{a})$, while an Inverse Wishart $\mathfrak{A}$ random matrix with scale matrix $W_\mathfrak{A}$ and $m_\mathfrak{A}$ degrees of freedom will be denoted as $IW(\mathfrak{A}; W_\mathfrak{A}, m_\mathfrak{A})$.

2. Problem Statement

We introduce the IOC problem in Section 2.1, and present in Section 2.2 the solution strategy from (Englert et al., 2017) building upon the KKT conditions. This is followed by the definition of the sub-optimal demonstrations considered and of the least-squares approximation of the inverse KKT approach in Sections 2.3 and 2.4, respectively.

2.1. Forward and Inverse Optimal Control Problem

We consider a known, deterministic, discrete-time system $x_{k+1} = f(x_k, u_k)$, where $x_k \in \mathbb{R}^n$ indicates the state and $u_k \in \mathbb{R}^m$ the input vector at time step k, generating state and input trajectories of length $N + 1$ and N, respectively. For notational simplicity and a compact representation of the IOC problem, we stack the input variables over their respective horizon length in $U = (u_0^\top, \dots, u_{N-1}^\top)^\top \in \mathbb{R}^{Nm}$. As for the state x_k, we represent it through the operator

$$F_k(U, x_0) = \begin{cases} x_0 & \text{if } k = 0, \\ f(F_{k-1}(U, x_0), u_{k-1}) & \text{if } k \geq 1, \end{cases} \tag{1}$$

and collect the sequence of all $F_k(U, x_0)$ in $\mathcal{F}_{U,x_0} = \{F_0(U, x_0), \dots, F_N(U, x_0)\}$. The optimal control problem considered is referred to as the *forward problem* and, with horizon length N and I known inequality constraints $\{g_i(\cdot, \cdot)\}_{i=1}^I$, reads as

$$\min_U \sum_{k=0}^{N-1} \theta^\top \phi(F_k(U, x_0), u_k) \quad \text{subject to} \begin{cases} g_i(F_k(U, x_0), u_k) \leq 0 \\ x_0 = x(0) \\ k = 0, \dots, N, \ i = 1, \dots, I. \end{cases} \tag{2}$$

The objective is modeled as a linear combination of given features, collected in the vector $\phi(\cdot) \in \mathbb{R}^q$, with an unknown coefficient vector $\theta \in \mathbb{R}^q$. We further assume that f, g_i and ϕ are continuously differentiable for all $i = 1, \dots, I$. Ultimately, the *inverse optimal control* problem follows as inferring the unknown parameter vector θ from the optimal input sequence, denoted with $U^* = (u_0^{*,\top}, \dots, u_{N-1}^{*,\top})$.

2.2. The Inverse KKT Approach

Considering optimal demonstrations, an estimate of the unknown parameter vector θ can be obtained based on the KKT conditions, which provide necessary conditions for the solution of an optimization problem that fulfills suitable constraint qualifications (Kuhn and Tucker, 2014). To present these KKT conditions, we introduce the Lagrangian multipliers $\lambda_{i,k} \in \mathbb{R}$, for all $k = 0, \dots, N$ and $i = 1, \dots, I$, and combine them for all i and a specific time step k in the vectors $\lambda_k = (\lambda_{0,k}, \dots, \lambda_{I,k})^\top \in \mathbb{R}^I$, as well as further summarizing them in $\lambda = (\lambda_0^\top, \dots, \lambda_N^\top)^\top \in \mathbb{R}^{IN}$. Additionally, we express the inequality constraints via $G(x_k, u_k) = (g_0(x_k, u_k), \dots, g_I(x_k, u_k))^\top \in \mathbb{R}^I$, such that the Lagrangian of problem (2) follows as

$$\mathcal{L}(\theta, \lambda, \mathcal{F}_{U,x_0}, U) = \textstyle\sum_{k=0}^{N-1} (\theta^\top \phi(F_k(U, x_0), u_k) + \lambda_k^\top G(F_k(U, x_0), u_k)).$$

Then, for all $k = 0, \dots, N$ and $i = 1, \dots, I$, the KKT conditions in accordance to the optimal demonstration U^* and the initial condition x_0^* are given as

$$\nabla_U \mathcal{L}(\theta, \lambda, \mathcal{F}_{U,x_0}, U)|_{x_0 = x_0^*, U = U^*} = \mathbb{0}_{mN,1} \tag{3a}$$

$$\lambda_{i,k} g_i(F_k(U^*, x_0^*), u_k^*) = 0 \tag{3b}$$

$$g_i(F_k(U^*, x_0^*), u_k^*) \leq 0, \lambda_{i,k} \geq 0. \tag{3c}$$

Solving (3) for θ and λ returns the coefficient vector θ^* of the corresponding forward problem, as well as its Lagrangian multipliers λ^*.

2.3. Sub-Optimal Demonstrations

The considered sub-optimal demonstrations follow a unified structure given in Assumption 1.

Assumption 1 *The available sequences, indicated with $U_d = (u_{d,0}^\top, \dots, u_{d,N-1}^\top)^\top$ for $d = 1, \dots, D$, are noisy realizations of the optimal sequence U^*:*

$$U_d = U^* + n_d, \tag{4}$$

where $n_d = (n_{d,0}, \dots, n_{d,N-1})^\top \in \mathbb{R}^{Nm}$ is a vector collecting the realization of the additive noise.

To address various causes of sub-optimal data, such as, e.g., a non-expert demonstrator or the presence of measurement noise, we consider the following two scenarios.

Problem 1 Demonstrations are i.i.d., normally distributed around U^*, i.e., $n_{d,k} \sim \mathcal{N}(\mathbb{0}_{m,1}, \Sigma_u)$ for all $d = 1, ..., D$ and $k = 0, ..., N-1$. The covariance matrix Σ_u thereby reflects the difficulty of obtaining an optimal input and is potentially unknown.

Problem 2 Additive noises $n_{d,k}$, for all $d = 1, ..., D$ and $k = 0, ..., N-1$, follow an unknown distribution.

2.4. A Least Squares Approximation of Inverse KKT

Working with sub-optimal demonstrations not only requires a nontrivial constraint handling, but can additionally cause violations of the stationarity condition in equation (3a), yielding

$$\nabla_U \mathcal{L}(\theta^*, \lambda^*, \mathcal{F}_{U,x_0}, U)|_{x_0=x_{d,0}, U=U_d} \neq \mathbb{0}_{mN,1}.$$

To deal with this issue, approaches such as those in (Englert et al., 2017; Menner et al., 2019) propose to reformulate (3a) in terms of a least-squares optimization problem, which we indicate with $(\theta^\top, \lambda_{[1:D]}^\top)^\top = KKT(U_{[1:D]})$. Its objective then reads as

$$\min_{\theta, \lambda_{[1:D]}} \sum_{d=0}^{D} \|\nabla_U \mathcal{L}(\theta, \lambda^d, \mathcal{F}_{U,x_0}, U)|_{x_0=x_{d,0}, U=U_d}\|^2 \tag{5}$$

and conditions (3b) and (3c) are considered in the constraints. However, as detailed in Subsection 3.1, the resulting problem has an EIV nature, and a plain least-squares estimate is not consistent – i.e., it is biased and does not converge to its optimal value as $D \to +\infty$.

3. IOC as an Errors-in-Variables Problem

In this section, we introduce the EIV-regression problem emerging from the inverse KKT problem with sub-optimal demonstrations, and present an IOC framework that takes inspiration from existing EIV solution strategies. The first approach builds upon an MCMC sampler, addresses Problem 1, and is presented in Section 3.2. The second approach, presented in Section 3.3, addresses Problem 2 and follows the idea of total least squares. Finally, practical aspects of the proposed framework are discussed in Section 3.4

3.1. Stationarity Condition as an EIV-Regression

Denoting with $J_\theta = \nabla_U(\sum_{k=0}^{N-1} \phi(F_k(U, x_0), u_k))$ and with $J_\lambda = \nabla_U(\sum_{k=0}^{N-1} G(F_k(U, x_0), u_k)))$, the stationarity condition in (3a) can be interpreted as the following linear regression problem

$$\underbrace{\begin{pmatrix} J_\theta & J_\lambda \end{pmatrix} |_{x_0=x_0^*, U=U^*}}_{J(U)} \underbrace{\begin{pmatrix} \theta \\ \lambda \end{pmatrix}}_{\beta} = \underbrace{\mathbb{0}_{mN,1}}_{Y}. \tag{6}$$

There, β acts as the regression parameter vector, $J(U)$ as a stack of feature vectors, and U as the independent variable vector. The dependent variables, collected in Y, are equal to zero when $U = U^*$. However, when operating with U_d as in (4), noise enters through the independent variable in the regressors, yielding an EIV problem. As a direct consequence, a naive least-squares estimate results biased and not consistent (Söderström, 2019). To address this issue, we present two optimization-based approaches in which the objective reflects the quality of a potential solution vector β in accordance with (6) and the available data, and consider conditions (3b) and (3c) in the constraints.

3.2. Approach 1: Maximum A Posteriori Estimate

This first approach addresses Problem 1 and relies on the information about the distribution of the additive noise in (4). The goal of jointly estimating U^* and β is formulated as a Maximum-a-Posteriori problem as follows.
We treat the independent variable U as a random vector with prior distribution $U \sim \mathcal{N}(U_0, \Sigma_{U_0})$. It should capture the optimal U^* and yield a likelihood model $U_d|U \sim \mathcal{N}(U, \Sigma_U)$ describing the suboptimal demonstrations according to (4). Moreover, the likelihood associated with (6) is $Y_d|U, \beta \sim \mathcal{N}(J(U)\beta, \Sigma_Y)$, where the covariance Σ_Y describes the discrepancy from the vector $\mathbb{0}_{mN,1}$ and is deterministically chosen. To complete the Bayesian description of our problem, we consider the following priors: $\beta \sim \mathcal{N}(\beta_0, \Sigma_\beta)$ and $\Sigma_U \sim IW(W_U, m_U)$. The resulting Bayesian network is depicted in Figure 1. Noting that U_d and Y_d are conditionally independent given U, the posterior for the unknown β, U and Σ_U reads as follows:

$$\begin{aligned} &p(\beta, \Sigma_U, U|\Sigma_Y, Y_{[1:D]}, U_{[1:D]}) \\ &\propto \textstyle\prod_{d=1}^{D} \mathcal{N}(Y_d; J(U)\beta, \Sigma_Y) \cdot \mathcal{N}(U_d; U, \Sigma_U) \cdot \mathcal{N}(U; U_0, \Sigma_{U_0}) \cdot \mathcal{N}(\beta; \beta_0, \Sigma_\beta) \cdot IW(\Sigma_U; W_U, m_U). \end{aligned} \tag{7}$$

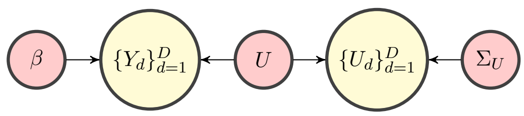

Figure 1: Visualization of the Bayesian network considered.

The rationale of our proposed approach consists in finding U and β maximizing such a posterior, while taking into consideration the constraints (3b) and (3c). This Maximum-A-Posteriori approach can be written as follows:

$$\min_{\beta, U} \; P(U, \beta|\Sigma_U) \qquad \text{subject to} \begin{cases} \lambda_{i,k} g_i(F_k(U, x_0), u_k) = 0 \\ g_i(F_k(U, x_0), u_k) \leq 0, \quad \lambda_{i,k} \geq 0 \\ k = 0, \ldots, N, i = 1, \ldots, I, \end{cases} \tag{8}$$

where the cost $P(U, \beta|\Sigma_U)$ is obtained by taking the negative logarithm of (7) and neglecting the distribution of Σ_U. Before solving (8), we perform a full Bayesian treatment of the problem to provide (A) an estimate for Σ_U and (B) good candidate initial values for U and β. This is done with an MCMC strategy based on a single-component Metropolis-Hastings sampling scheme (Gilks et al., 1995). We sketch the procedure in the following paragraph, and defer to the Supplementary material for the detailed derivation.

MCMC strategy The (single-component) Metropolis-Hastings algorithm allows for the generation of a Markov Chain for which the invariant distribution is a target distribution of interest. In our case, we aim at simulating the posterior (7). This is done by iteratively sampling from the so-called *full conditionals*, which in this case are $p(\beta|Y_{[1:D]}, \Sigma_Y)$, $p(U|U_{[1:D]}, Y_{[1:D]}, \beta, \Sigma_U)$ and $p(\Sigma_U|U_{[1:D]}, U)$. We first focus on the case in which the dynamics in (1) are linear, the objective features in $\phi(\cdot)$ are quadratic, and constraints in $\{g_i(\cdot,\cdot)\}_{i=1}^{I}$ are polytopic. In this scenario, $J(U)$ can be re-written as an affine transformation of U, i.e., $J(U)\beta = (MU+E)\beta$. The full conditionals, inspecting the likelihood-prior products, read as follows:

$$p(\beta|Y_{[1:D]}, U) \propto \prod_{d=1}^{D} \mathcal{N}\Big(Y_d; (MU + E)\beta, \Sigma_Y\Big) \cdot \mathcal{N}(\beta; \beta_0, \Sigma_\beta) \tag{9}$$

$$p(U|\beta, Y_{[1:D]}, U_{[1:D]}, \Sigma_U) \propto \prod_{d=1}^{D} \mathcal{N}(Y_d; (MU+E)\beta, \Sigma_Y) \cdot \mathcal{N}(U_d; U, \Sigma_U) \cdot \mathcal{N}(U; U_0, \Sigma_{U_0}) \quad (10)$$

$$p(\Sigma_U|U_{[1:D]}, U) \propto \prod_{d=1}^{D} \mathcal{N}(U_d; U, \Sigma_U) \cdot IW(\Sigma_U; W_U, m_U). \quad (11)$$

By leveraging standard properties of conjugate priors (Gelman et al., 2004), the distributions above have a closed-form expression as Gaussian ((9), (10)) and Inverse Wishart (11). Since these are easy to sample, the single-component Metropolis-Hastings scheme becomes a Gibbs sampler. Because all the involved distributions are well-defined, the stochastic simulation scheme is ergodic; at convergence, we perform a Monte Carlo integration of the full conditionals to obtain the sample mean for Σ_U, β and U. The first is used to set the cost in (8), while the other two are used as intialization of the solver. With this construction, the following Theorem holds. Its proof can be found in the Appendix.

Theorem 1 *Consider linear system dynamics, polytopic constraints, and quadratic features in ϕ. Then, if the adopted solver for problem (8) converges to the global minimum, its solution returns the maximum a posteriori estimate of the posterior probability function in (7), fulfilling conditions (3b) and (3c). Furthermore, the estimate is consistent.*

In the more general case of nonlinear dynamics, (10) is not Gaussian anymore, so a Metropolis-Hastings step has to be adopted. As a further consequence, the cost loses bi-convexity, and convergence is ensured only to a local optimum.

Finally, note that the sampling-based solution can easily be extended to non-Gaussian (but known) distributions by applying a general Metropolis-Hastings step on the required full-conditionals.

3.3. Approach 2: Total Least Square Estimate

This section addresses Problem 2, where a likelihood function cannot be constructed because no information on the noise distribution is available. In this approach, we employ total least squares to estimate the regression parameter vector β and the residuals $r_{[1:D]}$, which describe the noise realizations. A first formulation reads as follows:

$$\min_{r_{[1:D]},\beta} \sum_{d=1}^{D} r_d^\top \Sigma_U^{-1} r_d \quad \text{subject to} \begin{cases} J(U_d - r_d)\beta = \mathbb{0}_{mN,1} \\ \lambda_{i,k} g_i(F_k(U_d - r_d, x_0), u_{d,k} - r_{d,k}) = 0 \\ g_i(F_k(U - n_d, x_0), u_{d,k} - r_{d,k}) \leq 0, \lambda_{i,k} \geq 0 \\ k = 0, \dots, N, i = 1, \dots, I, d = 1, \dots, D. \end{cases} \quad (12)$$

To reduce the number of necessary optimization variables, we leverage the assumption that a unique optimal demonstration U^* exists, and extend the optimization problem in (12) with the equality constraint $U = U_d - r_d$ for all $d = 1, ..., D$. This allows for the transformation of (12) into an equivalent problem, optimizing for a single demonstration U instead of all residuals by replacing each usage of $U_d - r_d$ with U. Thus, (12) becomes

$$\min_{U,\beta} \sum_{d=1}^{D} (U - U_d)^\top \Sigma_U^{-1} (U - U_d) \quad \text{subject to} \begin{cases} J(U)\beta = \mathbb{0}_{mN,1} \\ \lambda_{i,k} g_i(F_k(U, x_0), u_k) = 0 \\ g_i(F_k(U, x_0), u_k) \leq 0, \lambda_{i,k} \geq 0 \\ k = 0, \dots, N, i = 1, \dots, I. \end{cases} \quad (13)$$

However, such a problem depends on Σ_U, which is initially unknown. To obviate this issue, we propose an alternating scheme which (A) starts from an initial guess for U and β, (B) estimates the unknown covariance as $\Sigma_U = \frac{1}{D}\sum_{d=1}^{D}(U - U_d)^\top (U - U_d)$, (C) solves (13), and (D) iterates the steps (B-C). The performance of such an approach is summarized in the following Theorem, which is proved in the Appendix.

Theorem 2 *Consider linear system dynamics, polytopic constraints, and quadratic features in ϕ. Furthermore, assume that n_k has mean zero. Then, if the solver for problem* (13) *converges to the global minimum, the proposed procedure returns a consistent estimate for β.*

3.4. Practical Aspects

In this section, we discuss two main considerations arising in the developed optimization schemes: namely, the initialization and the adopted solvers.

3.4.1. Initial Guess

Especially for non-convex optimization problems, or problems with only a few available demonstrations, a suitable initial guess and solver initialization are critical to improve the final estimate or reduce computation time. If no prior knowledge is available, we suggest setting U_0 equal to the sample mean, and β_0 as its corresponding estimate $KKT(U_0)$. The matrices Σ_Y and W_U can be either set to a covariance approximation with respect to U_0, risking exaggerated certainty of the estimated parameters (Kass and Steffey, 1989), or to a diagonal matrix, reflecting the independence of the investigated noise.

3.4.2. Solver

For linear demonstrating system dynamics, polytopic constraints and quadratic features in $\phi(\cdot)$, the optimization problems in (8) and (13) are bi-convex and fulfill all requirements for global convergence of the Global Optimization Algorithm (Floudas, 2000). In any other case, they denote a general Mathematical Problem with Equilibrium Constraints (MPEC). While such problems can be hard for most common solvers (e.g., IPOPT (Wächter and Biegler, 2006) or SQP-based ones (Boggs and Tolle, 1995)), the strongest convergence guarantees are provided for combinatorial methods, such as pivoting (Fang et al., 2012) or active set methods (Giallombardo and Ralph, 2008; Leyffer and Munson, 2007). The iterative application of nonlinear programming methods to a relaxed version of the MPECs is actively investigated (Scholtes, 2001; Kadrani et al., 2009; Kanzow and Schwartz, 2013), leading to the often-used IPOPT-C solver (Raghunathan and Biegler, 2005). In this work, the MPEC is addressed by combining the relaxation approach by (Scholtes, 2001) with IPOPT. The relaxation constant and its decrease factor entering the formulation are individually adjusted for each experiment, which is a key and delicate task to ensure that the solver converges to an acceptable solution.

4. Experimental Results

The proposed framework is tested on three different systems in simulation: a spring-damper system (Section 4.1), the kinematic bicycle model (Section 4.2), and a two-compartment Bayesian glucose model (Section 4.3). The continuous system dynamics are discretized using the backward Euler method with sampling time $T_s = 0.1$ for the first two, and $T_s = 1$ for the third system. While Approach 1 is employed on the first two systems, Approach 2 is tested on systems 2 and 3. Throughout the experiments, we consider different values for the input noise covariance, which are calculated with respect to the mean input values of the optimal demonstration over the considered horizon; we denote such a mean value with $u_{m,D}$. The relaxation constant and decrease factor of our MPEC solver (see Section 3.4.2) are adjusted individually and chosen between 6 and 25, as well as 0.75 and 0.9 respectively.[2] All estimates obtained with our framework are further denoted with U_{EIV} and θ_{EIV}, and their quality is evaluated by means of the root mean square error (RMSE) with respect to the true values U^* and β^*. In each set-up we repeat the experiments 10 times, and compare the performance against the β obtained with the inverse KKT least square relaxation proposed in (Menner et al., 2019), and the mean of all demonstrations $U_m = \frac{1}{D}\sum_{d=1}^{D} U_d$.

4.1. Spring-Damper System

The system dynamics read as

$$\dot{x}_1 = x_2, \quad \dot{x}_2 = \frac{1}{m}(-cx_1 - dx_2 + u),$$

with $m = 1.0$ kg, $c = 0.2$ kg $\cdot$ s^{-2} and $d = 0.1$ kg $\cdot$ s^{-1}. The optimal control forward problem has a horizon length of $N = 10$ and includes the inequality constraint $u \leq 0.7$. The unknown parameter is set to $\theta = (10, 5, 7)^\top$, the feature vector to $\phi = ((x_1 - 3)^2, (x_2 - 0)^2, u^2)^\top$, and the initial value is chosen as $x_0 = (1, 0.1)^\top$. Mean and standard deviations of the RMSE are presented in Table 4.1. The average runtime of one experiment employing Approach 1 is measured as 61.3s.

Method		Σ_U		
		5% $u_{m,D}$	10% $u_{m,D}$	20% $u_{m,D}$
KKT	θ_{KKT}	2.78 ± 0.207	1.59 ± 0.319	2.08 ± 0.734
	U_{m}	0.005 ± 0.01	0.02 ± 0.009	0.02 ± 0.005
EIV	$\boldsymbol{\theta_{\mathrm{EIV}}}$	$\mathbf{0.16 \pm 0.105}$	$\mathbf{0.28 \pm 0.138}$	$\mathbf{0.55 \pm 0.455}$
	$\mathbf{U_{EIV}}$	$\mathbf{0.002 \pm 0.001}$	$\mathbf{0.004 \pm 0.002}$	$\mathbf{0.01 \pm 0.006}$

Table 1: Mean and standard deviations of the RMSE for parameter and trajectory estimates on the spring-damper system.

4.2. Kinematic Bicycle Model

Representing a more realistic use case, Approach 1 and 2 have been tested with the kinematic bicycle model (Rajamani, 2011), indicated with θ_{EIV_1},U_{EIV_1} and θ_{EIV_2},U_{EIV_2} respectively. The demonstrating system dynamics read as:

$$\dot{x}_1 = u_1 \cos(x_3), \quad \dot{x}_2 = u_1 \sin(x_3), \quad \dot{x}_3 = u_1 \tan(x_4)/L, \quad \dot{x}_4 = u_2.$$

We set $L = 0.115$ m. The considered forward problem has a horizon length of $N = 10$ and includes the inequality constraint $u_1 \leq 2.2$. The unknown parameter is set to $\theta = (10, 10, 3, 3, 8, 5)$ and the input noise acts on the steering input u_2. The feature vector is set to $\phi = ((x_1 - 3)^2, (x_2 - 3)^2, (x_3 - 0)^2, (x_4 - 0)^2, u_1^2, u_2^2)^\top$ and the initial value is chosen as $x_0 = (0, 0, 0, 0)^\top$. Mean and

2. Note that our experiments required careful tuning of the MPEC relaxation parameters to ensure convergence to a suitable stationary point. This could likely be overcome and solve time significantly reduced by using a different and more complex MPEC solver.

standard deviations of the RMSE are presented in Table 2. The average runtime of one experiment employing Approach 1 and 2 is given as 110.1s and 52.2s respectively. We can observe that in this experiment, Approach 2 yields the best performance. This is likely due to the Metropolis-Hastings step needed in Approach 1 to deal with the nonlinearity of the problem: the results might be slightly inaccurate due to the finite-sample nature of the approach and/or to the dependence on the choice of the proposal distribution.

Method		Σ_U		
		5% $u_{m,D}$	10% $u_{m,D}$	20% $u_{m,D}$
KKT	θ_{KKT}	3.20 ± 0.001	3.20 ± 0.002	3.18 ± 0.007
	U_{m}	0.0006 ± 0.0001	0.001 ± 0.0002	0.002 ± 0.0005
EIV_1	θ_{EIV_1}	0.29 ± 0.096	0.48 ± 0.161	0.98 ± 0.114
	U_{EIV_1}	0.0007 ± 0.0001	0.001 ± 0.0002	0.002 ± 0.0004
EIV_2	$\boldsymbol{\theta}_{\mathbf{EIV_2}}$	$\mathbf{0.10 \pm 0.076}$	$\mathbf{0.21 \pm 0.171}$	$\mathbf{0.23 \pm 0.145}$
	$\mathbf{U}_{\mathbf{EIV_1}}$	$\mathbf{0.0004 \pm 0.0001}$	$\mathbf{0.0007 \pm 0.0002}$	$\mathbf{0.001 \pm 0.0004}$

Table 2: Mean and standard deviations of the RMSE for parameter and trajectory estimates on the kinematic bicycle model.

4.3. Two-Compartment Bayesian Glucose Model

As biomedical applications are considered an important potential use case of inverse optimal control, we finally test the approach proposed in Section 3.3 on the two-compartment Bayesian glucose model from (Callegari et al., 2003). The demonstrating system dynamics are

$$\dot{x}_1 = (-p_1 - k_{21} - x_3)x_1 + k_{12}x_2 + p_1 Q_{1b}, \quad \dot{x}_2 = k_{21}x_1 - k_{12}x_2, \quad \dot{x}_3 = -p_2 x_3 + p_3(u - I_b).$$

For a full explanation on these equations, we defer to (Callegari et al., 2003), and the parameter values are set according to patient 1 in Table 1 reported therein. The forward problem has a horizon length of $N = 20$ and the insulin input is constrained as $u \geq 0$. The unknown parameter is set to $\theta = (1, 0.1, 10)$, the feature vector to $\phi = ((x_1 - Q_{1b})^2, (x_2 - \frac{k_{21}}{k_{12}}Q_{1b})^2, u^2)^\top$ and the initial value is chosen as $x_0 = (Q_{1b} + 330, \frac{k_{21}}{k_{12}}Q_{1b})^\top$. Mean and standard deviations of the RMSE are presented in Table 4.3. The average runtime of one experiment employing Approach 2 is measured as 99.8s.

Method		Σ_U		
		5% $u_{m,D}$	10% $u_{m,D}$	20% $u_{m,D}$
KKT	θ_{KKT}	0.12 ± 0.007	0.13 ± 0.015	0.24 ± 0.046
	U_{m}	0.70 ± 0.091	1.40 ± 0.123	3.25 ± 0.182
EIV	$\boldsymbol{\theta}_{\mathbf{EIV}}$	$\mathbf{0.09 \pm 0.042}$	$\mathbf{0.10 \pm 0.033}$	$\mathbf{0.12 \pm 0.015}$
	$\mathbf{U}_{\mathbf{EIV}}$	$\mathbf{0.10 \pm 0.061}$	$\mathbf{0.20 \pm 0.113}$	$\mathbf{0.41 \pm 0.320}$

Table 3: Mean and standard deviations of the RMSE for parameter and trajectory estimates for the two-compartment glucose model of (Callegari et al., 2003).

5. CONCLUSIONS

We presented two IOC approaches for the estimation of an unknown parameter vector in partially known objectives from sub-optimal demonstrations. Both strategies build upon the KKT conditions and allow for the consideration of inequality constraints in the optimization problem of interest. The key idea in the proposed methods consists in addressing the EIV nature of the problem to obtain unbiased estimates. We consider two scenarios: in the first, we assume that the input noise entering the dynamics is distributed according to a Gaussian (but it could be extended to any known distribution), while in the second such information is unavailable. We tackle the first situation with a Bayesian strategy, leveraging MCMC to initialize the actual constrained optimization problem, while for the second we employ a formulation based on total least squares. Differently from other approaches in the literature, both of our proposed strategies learn the noise input covariance from data; additionally, the first approach can also include prior information and allow for uncertainty quantification. Theoretical consistency guarantees are provided for linear systems with zero-mean additive noise, polytopic constraints, and quadratic objectives. Results in simulation show that (i) the estimated input sequence is closer to the optimal one with respect to a naive sample mean of the demonstrating sequences, and (ii) the proposed approaches outperform a previously presented method relying on a least-squares relaxation of the classical KKT inversion approach, especially in scenarios with higher input noise levels.

Appendix

In the following, we present the proofs of Theorems 1 and 2 stated in Sections 3.2 and 3.3, respectively.

Proof of Theorem 1

Given the irreducibility and aperiodicity of the chosen MCMC sampling algorithms, obtaining a maximum a posteriori estimate follows directly from the ergodic theorem, as well as global convergence of the bi-convex problem in (8). To prove the consistency of this estimate, we reformulate the objective in (8) with respect to an arbitrary positive definite covariance and divide it by the amount of data D. This reads as

$$-\Bigg(\sum_{d=1}^{D}\frac{1}{D}\Big((J(U)\beta)^\top\Sigma_Y^{-1}(J(U)\beta)+Y_d^\top\Sigma_Y^{-1}Y_d-2(J(U)\beta)^\top\Sigma_Y^{-1}Y_d+U^\top\Sigma_U^{-1}U - 2U^\top\Sigma_U^{-1}U_d+U_d^\top\Sigma_U^{-1}U_d\Big)+\frac{1}{D}((U-U_0)^\top\Sigma_{U_0}^{-1}(U-U_0)+(\beta-\beta_0)^\top\Sigma_{\beta_0}^{-1}(\beta-\beta_0))\Bigg).$$

By the Law of Large Numbers, acknowledging that the mean of Y is equal to 0, and replacing U_d with $U^*+N_{U,d}$, then for $D\to\infty$ we obtain

$$\begin{aligned}&(J(U)\beta)^\top\Sigma_Y^{-1}(J(U)\beta)+U^\top\Sigma_U^{-1}U-2U^\top\Sigma_U^{-1}U^*+U^{*\top}\Sigma_U^{-1}U^*+C\\ &=(J(U)\beta)^\top\Sigma_Y^{-1}(J(U)\beta)+(U-U^*)^\top\Sigma_U^{-1}(U-U^*)+C,\end{aligned}\tag{14}$$

where C is a constant term that depends on β and U. The proof is concluded by noting that (14) is minimized by taking U^* and β^*. ■

Proof of Theorem 2

The proof of Theorem 2 directly follows along the line of the consistency proof in Theorem 1, showing that the cost function of optimization problem (13) is minimized by U^*. Consequently, the estimate of β follows as β^*. ■

The supplementary material for this paper, including a synopsis of MCMC and the full derivation of the approach presented in Section 3.2, can be found at http://arxiv.org/abs/2312.03532

References

Nematollah Ab Azar, Aref Shahmansoorian, and Mohsen Davoudi. From inverse optimal control to inverse reinforcement learning: A historical review. *Annual Reviews in Control*, 2020.

Paul T. Boggs and Jon W. Tolle. Sequential quadratic programming. *Acta numerica*, 4:1–51, 1995.

Tizano Callegari, Andrea Caumo, and Claudio Cobelli. Bayesian two-compartment and classic single-compartment minimal models: comparison on insulin modified IVGTT and effect of experiment reduction. *IEEE transactions on biomedical engineering*, pages 1301–1309, 2003.

Petros Dellaportas and David A. Stephens. Bayesian analysis of errors-in-variables regression models. *Biometrics*, pages 1085–1095, 1995.

Peter Englert, Ngo Anh Vien, and Marc Toussaint. Inverse KKT: Learning cost functions of manipulation tasks from demonstrations. *The International Journal of Robotics Research*, 2017.

Haw-ren Fang, Sven Leyffer, and Todd Munson. A pivoting algorithm for linear programming with linear complementarity constraints. *Optimization Methods and Software*, 27(1):89–114, 2012.

Eugene A. Feinberg and Adam Shwartz. *Handbook of Markov decision processes: methods and applications*, volume 40. Springer Science & Business Media, 2012.

Chelsea Finn, Sergey Levine, and Pieter Abbeel. Guided cost learning: Deep inverse optimal control via policy optimization. In *International conference on machine learning*, pages 49–58. PMLR, 2016.

Christodoulos A. Floudas. *Deterministic Global Optimization*. Springer, 2000.

Andrew Gelman, John B. Carlin, Hal S. Stern, and Donald B. Rubin. *Bayesian Data Analysis*. Chapman and Hall/CRC, 2nd ed. edition, 2004.

Giovanni Giallombardo and Daniel Ralph. Multiplier convergence in trust-region methods with application to convergence of decomposition methods for mpecs. *Mathematical Programming*, 112:335–369, 2008.

Walter Gilks, Sylvia Richardson, and Daniel Spiegelhalter, editors. *Markov Chain Monte Carlo in Practice*. Chapman and Hall/CRC, New York, December 1995.

Gene H. Golub and Charles F. Van Loan. An analysis of the total least squares problem. *SIAM journal on numerical analysis*, 17(6):883–893, 1980.

Zvi Griliches. Errors in variables and other unobservables. *Econometrica: Journal of the Econometric Society*, pages 971–998, 1974.

Kathrin Hatz, Johannes P. Schloder, and Hans Georg Bock. Estimating parameters in optimal control problems. *SIAM Journal on Scientific Computing*, 34, 2012.

Abdeslam Kadrani, Jean-Pierre Dussault, and Abdelhamid Benchakroun. A new regularization scheme for mathematical programs with complementarity constraints. *SIAM Journal on Optimization*, 20(1):78–103, 2009.

Christian Kanzow and Alexandra Schwartz. A new regularization method for mathematical programs with complementarity constraints with strong convergence properties. *SIAM Journal on Optimization*, 23(2):770–798, 2013.

Robert E. Kass and Duane Steffey. Approximate bayesian inference in conditionally independent hierarchical models (parametric empirical bayes models). *Journal of the American Statistical Association*, pages 717–726, 1989.

Basil Kouvaritakis and Mark Cannon. Model predictive control. *Switzerland: Springer International Publishing*, 38, 2016.

Harold W. Kuhn and Albert W. Tucker. Nonlinear programming. In *Traces and emergence of nonlinear programming*. Springer, 2014.

Sven Leyffer and Todd S. Munson. A globally convergent filter method for MPECs. *Preprint ANL/MCS-P1457-0907, Argonne National Laboratory, Mathematics and Computer Science Division*, 2007.

Jonathan Feng-Shun Lin, Pamela Carreno-Medrano, Mahsa Parsapour, Maram Sakr, and Dana Kulić. Objective learning from human demonstrations. *Annual Reviews in Control*, 2021.

Marcel Menner and Melanie N. Zeilinger. Maximum likelihood methods for inverse learning of optimal controllers. *IFAC-PapersOnLine*, 2020.

Marcel Menner, Peter Worsnop, and Melanie N. Zeilinger. Constrained inverse optimal control with application to a human manipulation task. *IEEE Transactions on Control Systems Technology*, 2019.

Yumiharu Nakano. Inverse stochastic optimal controls. *Automatica*, 149, 2023.

Arvind U Raghunathan and Lorenz T Biegler. An interior point method for mathematical programs with complementarity constraints (mpccs). *SIAM Journal on Optimization*, pages 720–750, 2005.

Rajesh Rajamani. *Vehicle dynamics and control*. Springer Science & Business Media, 2011.

Stefan Scholtes. Convergence properties of a regularization scheme for mathematical programs with complementarity constraints. *SIAM Journal on Optimization*, pages 918–936, 2001.

Torsten Söderström. A user perspective on errors-in-variables methods in system identification. *Control Engineering Practice*, 89:56–69, August 2019.

Andreas Wächter and Lorenz T. Biegler. On the implementation of an interior-point filter line-search algorithm for large-scale nonlinear programming. *Mathematical programming*, 2006.

Brian D. Ziebart, Andrew L. Maas, J. Andrew Bagnell, Anind K. Dey, et al. Maximum entropy inverse reinforcement learning. In *Proceedings of the Twenty-Third AAAI Conference on Artificial Intelligence*, volume 8, 2008.