Inverse linear-quadratic discrete-time finite-horizon optimal control for indistinguishable homogeneous agents: A convex optimization approach

(article starts on next page)

Inverse linear-quadratic discrete-time finite-horizon optimal control for indistinguishable homogeneous agents: A convex optimization approach[☆]

Han Zhang [a,b,c], Axel Ringh [d,*,1]

[a] *Department of Automation, School of Electronic Information and Electrical Engineering, Shanghai Jiao Tong University, Shanghai, China*
[b] *Key Laboratory of System Control and Information Processing, Ministry of Education of China, Shanghai 200240, China*
[c] *Shanghai Engineering Research Center of Intelligent Control and Management, Shanghai 200240, China*
[d] *Department of Mathematical Sciences, Chalmers University of Technology and University of Gothenburg, 41296 Gothenburg, Sweden*

ARTICLE INFO

ABSTRACT

The inverse linear-quadratic optimal control problem is a system identification problem whose aim is to recover the quadratic cost function and hence the closed-loop system matrices based on observations of optimal trajectories. In this paper, the discrete-time, finite-horizon case is considered, where the agents are also assumed to be homogeneous and indistinguishable. The latter means that the agents all have the same dynamics and objective functions and the observations are in terms of "snap shots" of all agents at different time instants, but what is not known is "which agent moved where" for consecutive observations. This absence of linked optimal trajectories makes the problem challenging. We first show that this problem is globally identifiable. Then, for the case of noiseless observations, we show that the true cost matrix, and hence the closed-loop system matrices, can be recovered as the unique global optimal solution to a convex optimization problem. Next, for the case of noisy observations, we formulate an estimator as the unique global optimal solution to a modified convex optimization problem. Moreover, the statistical consistency of this estimator is shown. Finally, the performance of the proposed method is demonstrated by a number of numerical examples.

1. Introduction

First proposed in Kalman (1964), Inverse Optimal Control (IOC) is an inverse problem whose aim is to reconstruct the cost function and predict the closed-loop system's behaviour, based on knowledge of the underlying system dynamics and observations of the system. The problem can be categorized as a system identification problem where the system is assumed to be governed by an optimal control model with known structure, and hence it is a so-called gray-box system identification problem (Ljung, 1999, p. 13). IOC problems are of great interest, not least due to the fact that many processes in nature have been observed to be optimal with respect to some criteria (Alexander, 1996). "Reverse-engineering" the agents' objective function not only enables us to understand how their decisions are made, but also provides a way to predict and imitate their behaviours. For example, the IOC problem has found applications in human motion analysis (Berret, Chiovetto, Nori, & Pozzo, 2011; Westermann, Lin, & Kulić, 2020), robot locomotion (Mombaur, Truong, & Laumond, 2010), robot manipulation (Jin, Kulić, Lin, Mou, & Hirche, 2019; Menner, Worsnop, & Zeilinger, 2021), systems biology (Tsiantis, Balsa-Canto, & Banga, 2018) and machine learning (Finn, Levine, & Abbeel, 2016; Köpf, Inga, Rothfuß, Flad, & Hohmann, 2017).

Nevertheless, for some scenarios with multiple agents, such as human crowds, bacteria, bird flocks, and schools of fish, the observation data is naturally collected by observing all the agents simultaneously in discrete time, for instance via video cameras. Due to the similarity between the agents and the fact that the population is often huge, it is often too expensive or simply impossible to track each individual in the group. Under such conditions, the agents can be assumed to be homogeneous and "indistinguishable". This "indistinguishablility" means that we

* Corresponding author.
  E-mail addresses: zhanghan_tc@sjtu.edu.cn (H. Zhang), axelri@chalmers.se (A. Ringh).
[1] The majority of the work was done when the author was with the Department of Electronic and Computer Engineering, The Hong Kong University of Science and Technology, Clear Water Bay, Kowloon, Hong Kong, China.

cannot tell "which agent goes where" for consecutive observations. "Reverse-engineering" the objective function for such agents is challenging due to the fact that:

(1) The control input measurement is usually not available. Moreover, the "indistinguishable" characteristics means that the agents' optimal trajectories are not available. Therefore, the existing IOC methods, such as Aswani, Shen, and Siddiq (2018), Jin et al. (2019), Molloy, Ford, and Perez (2018), Yu, Li, Fang, and Chen (2021), Zhang, Li and Hu (2019) and Zhang, Umenberger and Hu (2019), that minimizes the difference between the observed and expected optimal trajectories, or minimizes the violation of optimality conditions, cannot be applied.

(2) The structural identifiability under such "indistinguishable" observations is not guaranteed. In particular, given observed data for optimal "indistinguishable" agents, there is a priori no guarantee that there are not different objective functions or closed-loop systems that corresponds to it.

(3) The observations are usually contaminated by noise due to limitations of experiments and measuring methods. The identification methods should be statistically consistent and robust to observation noise in order to provide an accurate estimate of the objective function or the closed-loop system.

On the other hand, the Linear-Quadratic (LQ) optimal control formulation is one of the most commonly used optimal control methods in practice. Due to its simplicity, it is often used to approximate complex control problems and optimal behaviours (Fridovich-Keil, Ratner, Peters, Dragan, & Tomlin, 2020; Ji et al., 2019; Toumi, Malhamé, & Le Ny, 2020). It is therefore not surprising that the inverse LQ optimal control problem (with distinguishable observations) has been studied in different settings and in various degrees of generality; see, e.g., Anderson and Moore (2007, Sec. 5.6) and Boyd, El Ghaoui, Feron, and Balakrishnan (1994, Sec. 10.5) for the continuous-time infinite-horizon case, Li, Yao, and Hu (2020) and Li, Zhang, Yao, and Hu (2018) for the continuous-time finite-horizon case, Priess, Conway, Choi, Popovich, and Radcliffe (2014) for the discrete-time infinite-horizon case, and Keshavarz, Wang, and Boyd (2011), Yu et al. (2021), Zhang, Li et al. (2019) and Zhang, Umenberger et al. (2019) for the discrete-time finite-horizon case, respectively. Moreover, although the inverse LQ optimal control problem can be seen as a special case of the IOC problem for general nonlinear systems (Hatz, Schloder, & Bock, 2012; Keshavarz et al., 2011; Molloy et al., 2018; Molloy, Ford, & Perez, 2020; Pauwels, Henrion, & Lasserre, 2016; Rouot & Lasserre, 2017), the special structure of the LQ formulation allows for precise answers to structural identifiability, and in the finite time horizon case for the identification of a linear time-varying system from a limited amount of data. Further, to the best of our knowledge, the IOC problem for "indistinguishable" homogeneous agents has not been considered in existing literature.

In this work, we consider the linear-quadratic discrete-time finite-horizon IOC problem with indistinguishable observations. In particular, we assume that each agent is governed by the same discrete-time finite-horizon Linear Quadratic Regulator (LQR). Moreover, in this initial work on "indistinguishable" homogeneous agents, we also assume that there is no interaction between the agents. This is a simplifying assumption which we intend to relax in the future. In the current work, we focus on studying how to tackle this "indistinguishability" issue in the observations. In particular, this means that different trajectories of agents (which all follow the same dynamics) are observed, but such trajectories are "indistinguishable" in the sense that the matching

between each state observation and the corresponding agent is not available. The goal is to develop an inverse LQ optimal control algorithm for an external observer that can be used to identify the homogeneous agents' common objective function using "indistinguishable" observations. More precisely, the main contributions are:

(i) We show that the corresponding model structure is strictly globally identifiable.

(ii) In the case of exact measurements, we formulate a convex optimization problem and prove that the unique globally optimal solution is the quadratic cost term of interest.

(iii) In the case of noisy observations, we formulate an estimator of the sought quadratic cost term as the unique globally optimal solution to a modified convex optimization problem. Moreover, we also show that this estimator is (asymptotically) statistically consistent. The latter means that the estimate converges to the true parameter value as the number of agents tends to infinity.

The article is organized as follows: in Section 2, we formulate the problem, specify the model structure, and show that the model structure is strictly globally identifiable. Section 3 investigates the inverse problem in the absence of measurement noise, and we formulate a convex optimization problem whose unique global optimal solution coincides with the parameters of interest. In Section 4, we adapt the results of the previous section to the case when the observations have measurement noise. In particular, it is shown that the unique global optimal solution of the modified stochastic optimization problem is a statistically consistent estimator. Next, discussions regarding some general numerical difficulties with the discrete-time finite-horizon LQ IOC problem are included in Section 5, and in Section 6, we present the numerical results. Finally, the article is concluded in Section 7.

Notation: The following notation is utilized throughout the article: $\mathbb{S}^n$ denotes the set of $n \times n$ symmetric matrices, and $\mathbb{S}^n_+$ denotes the set of symmetric $n \times n$ matrices that are positive semidefinite. For symmetric matrices, $G_1 \succeq G_2$ denotes the Loewner partial order of G_1 and G_2, namely, $G_1 - G_2 \in \mathbb{S}^n_+$. Moreover, $G_1 \succ G_2$ denotes that $G_1 - G_2$ is strictly positive definite; see, e.g., Horn and Johnson (2013, Sec. 7.7). $\| \cdot \|_F$ denotes the Frobenius norm, and $\mathbf{1}$ denotes an all-one vector with appropriate length. Furthermore, we use ***italic bold font*** to denote stochastic elements. Finally, we denote $\prod_{k=1}^m A_k = A_m A_{m-1} \ldots A_1$.

2. Problem formulation, model structure, and identifiability

Suppose that each observed agent i ($i = 1, \ldots, M$) is governed by the following discrete-time finite-horizon LQR:

$$\underset{x^i_{1:N}, \, u^i_{1:N-1}}{\text{minimize}} \quad x^{iT}_N \bar{S} x^i_N + \sum_{t=1}^{N-1} \left(x^{iT}_t \bar{Q} x^i_t + u^{iT}_t \bar{R} u^i_t \right) \tag{1a}$$

$$\text{subject to} \quad x^i_{t+1} = A x^i_t + B u^i_t, \ t = 1, \ldots, N-1, \tag{1b}$$

$$x^i_1 = \bar{x}^i, \tag{1c}$$

where $A, \bar{Q}, \bar{S} \in \mathbb{R}^{n \times n}$, $B \in \mathbb{R}^{n \times m}$, and $\bar{R} \in \mathbb{R}^{m \times m}$. The agents are assumed to be homogeneous, namely, they have the same dynamics as well as the same objective functions. This means that the difference between the agents is only their initial values. For the sake of simplicity, as stated in the Introduction, in this work we assume there is no interactions between the agents. We intend to relax this assumption in future work. We will also assume that $\bar{S} = \bar{Q}$ and $\bar{R} = I$ throughout this paper. It is further assumed that A is invertible, that B has full column rank, and that (A, B) is controllable. The rationale for making the assumption that A is invertible lies in the fact that discrete-time systems are

often obtained by sampling of continuous-time systems. More precisely, if $\hat{A}$ is the system matrix of the continuous-time system and $\Delta t > 0$ is the sample period, then the system matrix $A = \exp(\hat{A}\Delta t)$ of the sampled discrete-time systems is always invertible (Zhang, Umenberger et al., 2019).

The optimal control input sequence $\bar{u}^i_{1:N-1}$ of (1) is given by $\bar{u}^i_t = \bar{K}_t x^i_t$, $t = 1, \ldots, N-1$, where

$$\bar{K}_t = -(B^T \bar{P}_{t+1} B + I)^{-1} B^T \bar{P}_{t+1} A,$$

and $\bar{P}_{2:N}$ is the solution to the following discrete-time Riccati Equation (DRE):

$$\bar{P}_t = A^T \bar{P}_{t+1} A + \bar{Q} - A^T \bar{P}_{t+1} B(B^T \bar{P}_{t+1} B + I)^{-1} B^T \bar{P}_{t+1} A,$$
$$t = 2, \ldots, N-1$$
$$\bar{P}_N = \bar{Q}. \tag{2}$$

For the problem of LQ IOC for indistinguishable agents, which is considered in this paper, the goal is to recover the matrix $\bar{Q}$ in the objective function given the knowledge of (possibly noisy) state observations. However, note that since the agents are indistinguishable, this means that the observations at different time instants are *unpaired*. More precisely, let $X_t := [x^1_t, \ldots, x^M_t]$ be the state at time t of M agents following the dynamics in (1), then the observations are modelled as

$$\left\{ Y^\phi_t \right\}^N_{t=1} = \left\{ \left[x^{\phi_t(1)}_t, \ldots, x^{\phi_t(M)}_t \right] \right\}^N_{t=1},$$

where $\phi_t : \{1, \ldots, M\} \mapsto \{1, \ldots, M\}$, $t = 1, \ldots, N$, are unknown permutations (the superscript ϕ in Y^ϕ_t is used to emphasizes the fact that the observations are shuffled). Nevertheless, $\phi_1(\cdot)$ can, without loss of generality, be assumed to be the identity mapping since the ordering of the initial states does not matter. Therefore we will henceforth restrict our attention to the set $\{\phi_t\}^N_{t=2}$.

Problem 1 (*IOC for Indistinguishable LQ Homogeneous Agents*). Given the unpaired state observations $\{Y^\phi_t\}^N_{t=1}$ of M homogeneous agents that are all governed by (1), recover the corresponding parameter $\bar{Q}$ in the objective function.

Before we continue, note that in the formulation in Problem 1 we assume that we have data from the entire planning horizon N of the forward problem (1). This means that (implicitly) we assume that the planning horizon N is known. Next, we would like to discuss the identifiability of such problem. According to the definition of identifiability in Ljung (1999, Def. 4.5, 4.6), *the identifiability is a property of the model structure $\mathcal{M}$ itself and have nothing to do with any concrete data*, where the model structure $\mathcal{M}$ is a parameterized collection of models that describes the relations between the input and the output signal of the system (Ljung & Chen, 2013). Hence, as a pre-step, we need to first define the underlying model structure $\mathcal{M}$ of Problem 1.

In Problem 1, we see the initial values of the trajectories as the "input signal", and $\{Y^\phi_t\}^N_{t=2}$ as the "output signal". Note that any permutation $\phi_t(\cdot)$ can be uniquely represented by a permutation matrix $\mathscr{P}_t \in \mathcal{P}$, where

$$\mathcal{P} := \{\mathscr{P} \in \{0, 1\}^{M \times M} \mid \mathscr{P}\mathbf{1} = \mathbf{1}, \ \mathscr{P}^T\mathbf{1} = \mathbf{1}\},$$

i.e., the set of 0-1-matrices with exactly one element that is one in each row and in each column. Post-multiplying any matrix with such a permutation matrix results in a matrix with permuted columns. Therefore, Y^ϕ_t can be represented as

$$Y^\phi_t = X_t \bar{\mathscr{P}}_t, \quad t = 2, \ldots, N,$$

where $\bar{\mathscr{P}}_t$ is the true (unknown) permutation, and hence for the system output it holds that

$$Y^\phi_t = \left(\prod_{k=1}^{t-1} A_{cl}(k; \bar{Q}) X_1 \right) \bar{\mathscr{P}}_t. \tag{3}$$

where $A_{cl}(k; \bar{Q})$ is the closed-loop system matrix at time instant k that is generated by LQR (1) using the cost matrix $\bar{Q}$. By using the property of the vectorization operator (Horn & Johnson, 1994, Lem. 4.3.1), vectorizing the above equation we have that

$$\text{vec}(Y^\phi_t) = \left[\bar{\mathscr{P}}^T_t \otimes \prod_{k=1}^{t-1} A_{cl}(k; \bar{Q}) \right] \text{vec}(X_1).$$

This is a valid input–output relation for any $Q \succeq 0$ and $\{\mathscr{P}_t\}^N_{t=2} \subset \mathcal{P}$. Thus, we have the following definition of the model structure $\mathcal{M}$ for Problem 1:

$$\mathcal{M}(Q, \{\mathscr{P}_t\}^N_{t=2}) := \{\mathcal{M}_t(Q, \mathscr{P}_t)\}^N_{t=2}$$
$$= \{\mathscr{P}^T_t \otimes \prod_{k=1}^{t-1} A_{cl}(k; Q)\}^N_{t=2}.$$

Next, we adopt the definition of identifiability in Ljung (1999, Def. 4.6, 4.7). More precisely in our case:

Definition 2.1 (*Identifiability*). $\mathcal{M}$ is globally identifiable at $(Q', \{\mathscr{P}'_t\}^N_{t=2}) \in \mathbb{S}^n_+ \times \mathscr{P}^{N-1}_t$ if it holds that

$$\mathcal{M}(Q, \{\mathscr{P}_t\}^N_{t=2}) = \mathcal{M}(Q', \{\mathscr{P}'_t\}^N_{t=2}), \ Q \in \mathbb{S}^n_+, \mathscr{P}_t \in \mathcal{P}$$
$$\implies Q = Q', \mathscr{P}_t = \mathscr{P}'_t, \quad \forall t = 2, \ldots, N.$$

$\mathcal{M}$ is strictly globally identifiable if it is globally identifiable at all $Q' \in \mathbb{S}^n_+, \mathscr{P}'_t \in \mathcal{P}$.

Proposition 2.1. *If $N \geq n + 1$, then the model structure $\mathcal{M}(Q, \{\mathscr{P}_t\}^N_{t=2})$ is strictly globally identifiable.*

Proof. Let $(Q', \{\mathscr{P}'_t\}^N_{t=2}) \in \mathbb{S}^n_+ \times \mathcal{P}^{N-1}$, and assume that $\mathcal{M}(Q, \{\mathscr{P}_t\}^N_{t=2}) = \mathcal{M}(Q', \{\mathscr{P}'_t\}^N_{t=2})$ for some $(Q, \{\mathscr{P}_t\}^N_{t=2}) \in \mathbb{S}^n_+ \times \mathcal{P}^{N-1}$. Therefore, $\mathcal{M}_t(Q, \mathscr{P}_t) = \mathcal{M}_t(Q', \mathscr{P}'_t)$ for $t = 2, \ldots, N$. On the other hand, recall that any permutation matrix $\mathscr{P}_t$ is a 0–1 matrix with exactly one element that is one in each row and column. This means that for all t, $\mathcal{M}_t(Q, \mathscr{P}_t)$ is composed of all-zero and $\prod_{k=1}^{t-1} A_{cl}(k; Q)$ sub-matrix blocks. Together with the fact that $\mathcal{M}_t(Q, \mathscr{P}_t) = \mathcal{M}_t(Q', \mathscr{P}'_t)$ for $t = 2, \ldots, N$, it implies that $\mathscr{P}_t = \mathscr{P}'_t$ and $\prod_{k=1}^{t-1} A_{cl}(k; Q) = \prod_{k=1}^{t-1} A_{cl}(k; Q')$ holds for $t = 2, \ldots, N$. Since $A_{cl}(t; Q)$ is invertible for all $t = 1, \ldots, N-1$ (Zhang, Umenberger et al., 2019), by induction $A_{cl}(t; Q) = A_{cl}(t; Q')$ holds for all $t = 1, \ldots, N-1$. Now, if $N \geq n+1$, this implies that $Q = Q'$ (Zhang, Umenberger et al., 2019, Theorem. 2.1). This gives the global identifiability of $\mathcal{M}(Q', \{\mathscr{P}'_t\})$. Finally, since $(Q', \{\mathscr{P}'_t\}^N_{t=2})$ was arbitrarily chosen from $\mathbb{S}^n_+ \times \mathcal{P}^{N-1}$, $\mathcal{M}(Q', \{\mathscr{P}'_t\}^N_{t=2})$ is strictly globally identifiable. $\square$

As a final note in this section, we observe that by identifying $\bar{Q}$ we implicitly also identify $\{\bar{\mathscr{P}}_t\}^N_{t=2}$. More precisely, if $\bar{Q}$ is identified, then by solving the forward problem (1), the estimates of each agent's trajectory can be easily obtained; pairing these estimates with the data gives the permutations. Moreover, as will be shown next, it turns out that $\bar{Q}$ can be identified without explicitly identifying the permutations.

3. IOC for indistinguishable agents in the noiseless case

After justifying the identifiability of the problem, we now investigate the IOC problem for indistinguishable homogeneous agents in the noiseless case, namely, it is assumed that $y^i_t = x^{\phi_t(i)}_t$, $\forall t$ and i. More precisely, we construct the IOC algorithm for indistinguishable observations as a convex optimization problem. However, before presenting the optimization problem, let us first sketch the main intuition behind it.

To this end, we first note that the optimal solution to the forward problem (1) is characterized by the DRE (2). However,

the latter is a nonlinear equation in $\bar{P}_t$, and in order to tackle to problem we first relax it to a linear matrix inequality (LMI). To do so, note that since $\bar{Q} \in \mathbb{S}_+^n$ we know that $\{\bar{P}_t\}_{t=1}^N \subset \mathbb{S}_+^n$, and hence it follows that $B^T\bar{P}_t B + I \succ 0, \forall t = 1, \ldots, N$. Moreover, the expression $A^T\bar{P}_{t+1}A + \bar{Q} - A^T\bar{P}_{t+1}B(B^T\bar{P}_{t+1}B+I)^{-1}B^T\bar{P}_{t+1}A - \bar{P}_t$ is actually the Schur complement of

$$\bar{F}_t := \begin{bmatrix} B^T\bar{P}_{t+1}B + I & B^T\bar{P}_{t+1}A \\ A^T\bar{P}_{t+1}B & A^T\bar{P}_{t+1}A + \bar{Q} - \bar{P}_t \end{bmatrix},$$

namely, $\bar{F}_t \backslash (B^T\bar{P}_{t+1}B + I)$, which is well-defined since $B^T\bar{P}_{t+1}B + I$ has full rank. By properties of the Schur complement (see, e.g., Horn & Johnson, 2013, p. 495), if we let $\bar{F}_t \succeq 0$ for $t = 1, \ldots, N-1$, this is equivalent to relaxing the DRE into the matrix inequality

$$A^T\bar{P}_{t+1}A - \bar{P}_t + \bar{Q} - A^T\bar{P}_{t+1}B(B^T\bar{P}_{t+1}B+I)^{-1}B^T\bar{P}_{t+1}A \succeq 0. \quad (4)$$

On the other hand, introducing $\bar{K}_t := -(B^T\bar{P}_{t+1}B + I)^{-1}B^T\bar{P}_{t+1}A$, the above relaxation can be further written as

$$(A + B\bar{K}_t)^T\bar{P}_{t+1}(A + B\bar{K}_t) - \bar{P}_t + \bar{Q} + \bar{K}_t^T\bar{K}_t \succeq 0.$$

Now, pre- and post-multiply on both sides of the above inequality with the state vector x_t^i, we get that

$$x_{t+1}^{iT}\bar{P}_{t+1}x_{t+1}^i - x_t^{iT}\bar{P}_t x_t^i + x_t^{iT}\bar{Q}x_t^i + \|\bar{u}_t^i\|^2 \geq 0,$$

for $t = 1, \ldots, N-1$, since $x_{t+1}^i = (A + B\bar{K}_t)x_t$ and $\bar{u}_t^i = \bar{K}_t x_t^i$. Summing the above inequality from $t = 1$ to $N - 1$, and using that the trace operator is invariant under cyclic permutation of the elements, we get that

$$-\operatorname{tr}(\bar{P}_1 x_1^i x_1^{iT}) + \operatorname{tr}(\bar{P}_N x_N^i x_N^{iT}) + \sum_{t=1}^{N-1} \operatorname{tr}(\bar{Q}x_t^i x_t^{iT}) + \|\bar{u}_t^i\|^2 \geq 0.$$

Summing this inequality over i, and using the property that

$$Y_t^\phi Y_t^{\phi T} = \sum_{i=1}^M x_t^{\phi_t(i)} x_t^{\phi_t(i)T} = \sum_{i=1}^M x_t^i x_t^{iT} = X_t X_t^T, \quad (5)$$

we get that

$$-\operatorname{tr}(\bar{P}_1 Y_1^\phi Y_1^{\phi T}) + \operatorname{tr}(\bar{P}_N Y_N^\phi Y_N^{\phi T})$$
$$+ \sum_{t=1}^{N-1}\left(\operatorname{tr}(\bar{Q}Y_t Y_t^{\phi T}) + \sum_{i=1}^M \|\bar{u}_t^i\|^2\right) \geq 0.$$

The above inequality is only derived for the "true" parameters $\bar{Q}$ and $\{\bar{P}_t\}_{t=1}^N$ (in which case it is in fact an equality since (4) is an equality in this case), but as we shall see shortly (Lemma 3.1) the inequality is in fact true for all parameters $Q \in \mathbb{S}_+^n$ and $\{P_t\}_{t=1}^N \subset \mathbb{S}_+^n$ such that

$$F_t := \begin{bmatrix} B^T P_{t+1}B + I & B^T P_{t+1}A \\ A^T P_{t+1}B & A^T P_{t+1}A + Q - P_t \end{bmatrix} \succeq 0,$$
$$t = 1, \ldots, N - 1, \quad (6a)$$
$$P_N = Q. \quad (6b)$$

Therefore, let us define the domain

$$\mathscr{D} := \{Q \in \mathbb{S}_+^n, \{P_t \in \mathbb{S}_+^n\}_{t=1}^N : (6) \text{ holds}\}, \quad (7)$$

and the function $H : \mathscr{D} \mapsto \mathbb{R}$ as

$$H(Q, \{P_t\}) := -\operatorname{tr}(P_1 Y_1^\phi Y_1^{\phi T})$$
$$+ \operatorname{tr}(P_N Y_N^\phi Y_N^{\phi T}) + \sum_{t=1}^{N-1} \operatorname{tr}(Q Y_t^\phi Y_t^{\phi T}). \quad (8)$$

For this function, we have the following properties:

Lemma 3.1. *For any* $(Q, \{P_t\}_{t=1}^N) \in \mathscr{D}$, *it holds that*

$$H(Q, \{P_t\}) + \sum_{i=1}^M \sum_{t=1}^{N-1} \|\bar{u}_t^i\|^2 \geq 0,$$

where $\{\bar{u}_t^i\}_{t=1}^{N-1}$ *is the optimal control input sequence in* (1) *that corresponds to the "true" matrix* $\bar{Q}$ *and the trajectory* $\{x_t^i\}_{t=1}^N$. *Moreover, let* $\{\bar{P}_t\}_{t=1}^N$ *be the solution to the DRE* (2) *that corresponds to* $\bar{Q}$. *Then* $H(\bar{Q}, \{\bar{P}_t\}) + \sum_{i=1}^M \sum_{t=1}^{N-1} \|\bar{u}_t^i\|^2 = 0$.

Proof. Since $\{P_t\}_{t=1}^N$, Q is feasible, it is clear that $P_t \in \mathbb{S}_+^n$, $\forall t = 1, \ldots, N$ and hence $B^T P_{t+1}B + I \succ 0, \forall t = 1, \ldots, N-1$. Moreover, since (6a) holds, by positive semidefiniteness of the Schur complement (Horn & Johnson, 2013, p. 495) it holds that

$$A^T P_{t+1}A - P_t + Q - A^T P_{t+1}B(B^T P_{t+1}B + I)^{-1}B^T P_{t+1}A \succeq 0. \quad (9)$$

Introducing $K_t := -(B^T P_{t+1}B + I)^{-1}B^T P_{t+1}A$, we can re-write the above matrix inequalities as

$$(A + BK_t)^T P_{t+1}(A + BK_t) - P_t + Q + K_t^T K_t \succeq 0, \quad t = 1, \ldots, N-1.$$

Rearranging the terms, we have

$$A^T P_{t+1}A - P_t + Q \succeq -K_t^T B^T P_{t+1}A$$
$$- A^T P_{t+1}BK_t - K_t^T(B^T P_{t+1}B + I)K_t. \quad (10)$$

On the other hand, for the i'th trajectory, consider the following term

$$J_t^i = x_{t+1}^{iT} P_{t+1}x_{t+1}^i - x_t^{iT}P_t x_t^i + x_t^{iT}Q x_t^i + \|\bar{u}_t^i\|^2, \quad (11)$$

where $\bar{u}_t^i$ is the control input that corresponds to $\bar{Q}$ and the trajectory $\{x_t^i\}_{t=1}^N$ in the forward problem (1). Using the fact that $x_{t+1}^i = (A + B\bar{K}_t)x_t^i$ and $\bar{u}_t^i = \bar{K}_t x_t^i$, it holds for (11) that

$$J_t^i = x_t^{iT}\underbrace{\left[(A + B\bar{K}_t)^T P_{t+1}(A + B\bar{K}_t) - P_t + Q + \bar{K}_t^T\bar{K}_t\right]}_{=: \mathscr{H}_t} x_t^i.$$

Expanding the terms of $\mathscr{H}_t$, we have

$$\mathscr{H}_t = A^T P_{t+1}A - P_t + Q$$
$$+ \bar{K}_t^T B^T P_{t+1}A + A^T P_{t+1}B\bar{K}_t + \bar{K}_t^T(B^T P_{t+1}B + I)\bar{K}_t,$$

and using the matrix inequality (10) we get that

$$\mathscr{H}_t \succeq -K_t^T B^T P_{t+1}A - A^T P_{t+1}BK_t - K_t^T(B^T P_{t+1}B + I)K_t$$
$$+ \bar{K}_t^T B^T P_{t+1}A + A^T P_{t+1}B\bar{K}_t + \bar{K}^T(B^T P_{t+1}B + I)\bar{K}_t. \quad (12)$$

Recalling that $K_t = -(B^T P_{t+1}B + I)^{-1}B^T P_{t+1}A$ and using the fact that

$$\bar{K}_t^T B^T P_{t+1}A = \bar{K}_t^T(B^T P_{t+1}B + I)(B^T P_{t+1}B + I)^{-1}B^T P_{t+1}A$$
$$= -\bar{K}_t^T(B^T P_{t+1}B + I)K_t,$$

and that

$$K_t^T B^T P_{t+1}A = K_t^T(B^T P_{t+1}B + I)(B^T P_{t+1}B + I)^{-1}B^T P_{t+1}A$$
$$= -K_t^T(B^T P_{t+1}B + I)K_t,$$

the matrix inequality (12) can be further rewritten as

$$\mathscr{H}_t \succeq 2K_t^T(B^T P_{t+1}B + I)K_t - K_t^T(B^T P_{t+1}B + I)K_t$$
$$- \bar{K}_t^T(B^T P_{t+1}B + I)K_t - K_t^T(B^T P_{t+1}B + I)\bar{K}_t$$
$$+ \bar{K}_t^T(B^T P_{t+1}B + I)\bar{K}_t$$
$$= (K_t - \bar{K}_t)^T(B^T P_{t+1}B + I)(K_t - \bar{K}_t).$$

Hence it holds that

$$J_t^i = x_t^{iT}\mathscr{H}_t x_t^i \geq x_t^{iT}(K_t - \bar{K}_t)^T(B^T P_{t+1}B + I)(K_t - \bar{K}_t)x_t^i. \quad (13)$$

On the other hand, recall that $Y_t^\phi = \left[x_t^{\phi_t(1)}, \ldots, x_t^{\phi_t(M)} \right]$. Therefore, it follows that

$$
\mathrm{tr}(P_t Y_t^\phi Y_t^{\phi T}) = \mathrm{tr}\left(P_t \left[x_t^{\phi_t(1)}, \ldots, x_t^{\phi_t(M)} \right] \begin{bmatrix} x_t^{\phi_t(1)T} \\ \vdots \\ x_t^{\phi_t(M)T} \end{bmatrix} \right)
$$

$$
= \mathrm{tr}\left(P_t \left(\sum_{i=1}^M x_t^{\phi_t(i)} x_t^{\phi_t(i)T} \right) \right) = \sum_{i=1}^M x_t^{\phi_t(i)T} P_t x_t^{\phi_t(i)}
$$

$$
= \sum_{i=1}^M x_t^{iT} P_t x_t^i, \quad t = 1, \ldots, N. \tag{14}
$$

Similarly, it also holds that $\mathrm{tr}(QY_t^\phi Y_t^{\phi T}) = \sum_{i=1}^M x_t^{iT} Q x_t^i$, $t = 2, \ldots, N$. Together with (11) and (13), we therefore have that

$$
H(Q, \{P_t\}) + \sum_{i=1}^M \sum_{t=1}^{N-1} \|\bar{u}_t^i\|^2 = \sum_{i=1}^M \sum_{t=1}^{N-1} \|\bar{u}_t^i\|^2
$$

$$
+ \sum_{t=1}^{N-1} \left\{ \mathrm{tr}(P_{t+1} Y_{t+1}^\phi Y_{t+1}^{\phi T}) - \mathrm{tr}(P_t Y_t^\phi Y_t^{\phi T}) + \mathrm{tr}(Q Y_t^\phi Y_t^{\phi T}) \right\}
$$

$$
= \sum_{t=1}^{N-1} \left\{ \sum_{i=1}^M \underbrace{x_{t+1}^{iT} P_{t+1} x_t^i - x_t^{iT} P_t x_t^i + x_t^{iT} Q x_t^i + \|\bar{u}_t^i\|^2}_{= J_t^i} \right\}
$$

$$
\geq \sum_{t=1}^{N-1} \left\{ \sum_{i=1}^M \left[x_t^{iT} (K_t - \bar{K}_t)^T (B^T P_{t+1} B + I)(K_t - \bar{K}_t) x_t^i \right] \right\}
$$

$$
\geq 0, \tag{15}
$$

which proves the first part of the statement.

Finally, by the arguments that lead up to (8) we know that the lower bound is reached by $(\bar{Q}, \{\bar{P}_t\}_{t=1}^N)$. □

In particular, note that $\{\bar{u}_t^i\}_{t=1}^{N-1}$ in Lemma 3.1 has nothing to do with the variable $(Q, \{P_t\})$. Therefore, the lemma effectively shows that the function H is bounded from below on the domain $\mathscr{D}$ and that the lower bound is attained by the "true" parameters $(\bar{Q}, \{\bar{P}_t\}_{t=1}^N)$. Therefore, we construct the following optimization problem to reconstruct Q:

$$
\underset{Q, \{P_t\}_{t=1}^N}{\text{minimize}} \quad H(Q, \{P_t\}) \tag{16a}
$$

$$
\text{subject to} \quad (Q, \{P_t\}_{t=1}^N) \in \mathscr{D}. \tag{16b}
$$

This is a semidefinite programming problem, and hence a convex optimization problem, that can be solved using standard numerical solvers, e.g., Löfberg (2004). We know that $\bar{Q}$ and $\{\bar{P}_t\}_{t=1}^N$ is an optimal solution to the problem. Next, we show that it is unique. For this, we need the following result:

Lemma 3.2 (*Persistent Excitation*). *If $Y_1^\phi Y_1^{\phi T} \succ 0$, then it holds that $Y_t^\phi Y_t^{\phi T} \succ 0$, $\forall t = 2, \ldots, N$.*

Proof. By using the property $x_{t+1}^i = A_{cl}(t; \bar{Q}) x_t^i$, as well as (5), it follows that

$$
Y_{t+1}^\phi Y_{t+1}^{\phi T} = \sum_{i=1}^M x_{t+1}^i x_{t+1}^{iT} = A_{cl}(t; \bar{Q}) \left(\sum_{i=1}^M x_t^i x_t^{iT} \right) A_{cl}^T(t; \bar{Q}).
$$

Since for all $Q \in \mathbb{S}_+^n$, $A_{cl}(t; Q)$ is invertible for all $t = 1, \ldots, N-1$ (Zhang, Umenberger et al., 2019), and since positive definiteness is invariant under congruence (Horn & Johnson, 2013, Obs. 7.1.8), by induction the statement follows. □

Remark 3.1. The "persistent excitation" condition $Y_1^\phi Y_1^{\phi T} \succ 0$ is equivalent to that there are n linearly independent initial values x_1^i amongst all M initial values.

Now we are ready to present the main theorem for the noiseless case.

Theorem 3.1. *Assume that $N \geq n + 1$, $Y_1^\phi Y_1^{\phi T} \succ 0$, and let $(Q^*, \{P_t^*\}_{t=1}^N)$ be an optimal solution to (16). Then $Q^* = \bar{Q}$, where $\bar{Q}$ is the "true" parameter that is used to generate $\{Y_t^\phi\}_{t=2}^N$ with unknown permutations.*

Proof. By Lemma 3.1, we know that $\left(\bar{Q}, \{\bar{P}_t\}_{t=1}^N \right)$ is an optimal solution to (16). What remains to show is the uniqueness.

Since $(Q^*, \{P_t^*\}_{t=1}^N)$ is an optimal solution, it must be feasible, thus (15) also holds for $(Q^*, \{P_t^*\}_{t=1}^N)$. Moreover, by the assumption that $Y_1^\phi Y_1^{\phi T} \succ 0$ and Lemma 3.2, it holds that $Y_t^\phi Y_t^{\phi T} \succ 0$, $\forall t = 2, \ldots, N$, and hence there exists a strictly positive definite matrix $(Y_t^\phi Y_t^{\phi T})^{\frac{1}{2}}$ such that $Y_t^\phi Y_t^{\phi T} = (Y_t^\phi Y_t^{\phi T})^{\frac{1}{2}} (Y_t^\phi Y_t^{\phi T})^{\frac{1}{2}}$ (Horn & Johnson, 2013, p. 440). Letting $K_t^* := -(B^T P_{t+1}^* B + I)^{-1} B^T P_{t+1}^* A$, by further term manipulation in (15), we have

$$
H(Q^*, \{P_t^*\}) + \sum_{i=1}^M \sum_{t=1}^{N-1} \|\bar{u}_t^i\|^2
$$

$$
\geq \sum_{t=1}^{N-1} \left\{ \sum_{i=1}^M \left[x_t^{iT} (K_t^* - \bar{K}_t)^T (B^T P_{t+1}^* B + I)(K_t^* - \bar{K}_t) x_t^i \right] \right\}
$$

$$
= \sum_{t=1}^{N-1} \left\{ \mathrm{tr} \left[(K_t^* - \bar{K}_t)^T (B^T P_{t+1}^* B + I)(K_t^* - \bar{K}_t) Y_t^\phi Y_t^{\phi T} \right] \right\}
$$

$$
= \sum_{t=1}^{N-1} \left\{ \left\| (B^T P_{t+1}^* B + I)^{\frac{1}{2}} (K_t^* - \bar{K}_t)(Y_t^\phi Y_t^{\phi T})^{\frac{1}{2}} \right\|_F^2 \right\}
$$

$$
\geq 0.
$$

As stated in Lemma 3.1, the lower bound zero in the above inequality is reached by $\left(\bar{Q}, \{\bar{P}_t\}_{t=1}^N \right)$, and it also holds that $H(Q^*, \{P_t^*\}_{t=1}^N) + \sum_{i=1}^M \sum_{t=1}^{N-1} \|\bar{u}_t^i\|^2 = 0$ since $(Q^*, \{P_t^*\}_{t=1}^N)$ is also an optimal solution. Therefore, since all terms in the above sum are nonnegative, it must hold that

$$
\left\| (B^T P_{t+1}^* B + I)^{\frac{1}{2}} (K_t^* - \bar{K}_t)(Y_t^\phi Y_t^{\phi T})^{\frac{1}{2}} \right\|_F = 0,
$$

$$
\iff (B^T P_{t+1}^* B + I)^{\frac{1}{2}} (K_t^* - \bar{K}_t)(Y_t^\phi Y_t^{\phi T})^{\frac{1}{2}} = 0,
$$

$$
\iff K_t^* - \bar{K}_t = 0, \quad t = 1, \ldots, N-1
$$

since $(B^T P_{t+1}^* B + I)^{\frac{1}{2}}$ and $(Y_t^\phi Y_t^{\phi T})^{\frac{1}{2}}$ are invertible.

The above argument shows that $K_t^* = \bar{K}_t$ for all $t = 1, \ldots, N-1$. Nevertheless, note that $\{K_t^*\}_{t=1}^{N-1}$ is not the control gain that is generated by Q^* using DRE (2). Instead, it is defined by $K_t^* = -(B^T P_{t+1}^* B + I)^{-1} B^T P_{t+1}^* A$, for $t = 1, \ldots, N-1$, where $(Q^*, \{P_t^*\}_{t=1}^{N-1})$ is an optimizer of (16). Hence the result from Zhang, Umenberger et al. (2019, Theorem 2.1) cannot be directly applied to show that $Q^* = \bar{Q}$. Nevertheless, to show the latter, first note that we can always write $P_t^* = \bar{P}_t + \Delta P_t$, $t = 1, \ldots, N$, and $Q^* = \bar{Q} + \Delta Q$, for some symmetric matrices ΔQ and ΔP_t, for $t = 1, \ldots, N$. Since both $(Q^*, \{P_t^*\}_{t=1}^N)$ and $(\bar{Q}, \{\bar{P}_t\}_{t=1}^N)$ are optimal solutions to (16), they must have the same optimal value, and hence

$$
H(Q^*, \{P_t^*\}_{t=1}^N) - H(\bar{Q}, \{\bar{P}_t\}_{t=1}^N) = 0,
$$

which implies that

$$
-\mathrm{tr}(\Delta P_1 Y_1^\phi Y_1^{\phi T}) + \mathrm{tr}(\Delta P_N Y_N^\phi Y_N^{\phi T}) + \sum_{t=1}^{N-1} \mathrm{tr}(\Delta Q Y_t^\phi Y_t^{\phi T}) = 0, \tag{17}
$$

where we have used the linearity of the trace operator. On the other hand, it holds that

$$K_t^* = -(B^T P_{t+1}^* B + I)^{-1} B^T P_{t+1}^* A$$
$$\iff (B^T P_{t+1}^* B + I)K_t^* = -B^T P_{t+1}^* A$$
$$\iff B^T P_{t+1}^* (A + BK_t^*) = -K_t^*.$$

Since $K_t^* = \bar{K}_t, \forall t = 1, \ldots, N-1$, $A + BK^* = A + B\bar{K}_t = A_{cl}(t; \bar{Q})$, and by the fact that $A_{cl}(t; \bar{Q})$ is invertible for all $t = 1, \ldots, N - 1$ (Zhang, Umenberger et al., 2019), it follows that

$$B^T P_{t+1}^* = -K_t^* A_{cl}(t; \bar{Q})^{-1} = -\bar{K}_t A_{cl}(t; \bar{Q})^{-1} = B^T \bar{P}_{t+1}$$
$$\implies B^T \Delta P_{t+1} = 0, \forall t = 1, \ldots, N-1. \tag{18}$$

Next, since $(Q^*, \{P_t^*\}_{t=1}^N)$ is feasible, (9) also holds for $(Q^*, \{P_t^*\}_{t=1}^N)$. This means that

$$A^T(\bar{P}_{t+1} + \Delta P_{t+1})A - (\bar{P}_t + \Delta P_t) + (\bar{Q} + \Delta Q)$$
$$- A^T(\bar{P}_{t+1} + \Delta P_{t+1})B[B^T(\bar{P}_{t+1} + \Delta P_{t+1})B + I]^{-1}$$
$$\times B^T(\bar{P}_{t+1} + \Delta P_{t+1})A \succeq 0, \quad t = 1, \ldots, N-1$$
$$\bar{P}_N + \Delta P_N = \bar{Q} + \Delta Q.$$

On the other hand, since $(\bar{Q}, \{\bar{P}_t\}_{t=1}^N)$ satisfies (2), by also using (18) we get that

$$A^T \Delta P_{t+1} A - \Delta P_t + \Delta Q \succeq 0, \quad t = 1, \ldots, N-1, \tag{19}$$
$$\Delta P_N = \Delta Q. \tag{20}$$

Using (18), we can further manipulate the terms of (19) and get

$$\underbrace{(A + B\bar{K}_t)^T}_{A_{cl}(t;\bar{Q})} \Delta P_{t+1} \underbrace{(A + B\bar{K}_t)}_{A_{cl}(t;\bar{Q})} - \Delta P_t + \Delta Q \succeq 0, \quad \forall t = 1, \ldots, N-1$$
$$\implies X_t^T A_{cl}(t; \bar{Q})^T \Delta P_{t+1} A_{cl}(t; \bar{Q}) X_t - X_t^T \Delta P_t X_t + X_t^T \Delta Q X_t \succeq 0,$$
$$\implies X_{t+1}^T \Delta P_{t+1} X_{t+1} - X_t^T \Delta P_t X_t + X_t^T \Delta Q X_t \succeq 0,$$
$$\implies \text{tr}(\Delta P_{t+1} X_{t+1} X_{t+1}^T) - \text{tr}(\Delta P_t X_t X_t^T) + \text{tr}(\Delta Q X_t X_t^T) \geq 0,$$
$$\implies \text{tr}(\Delta P_{t+1} Y_{t+1}^\phi Y_{t+1}^{\phi T}) - \text{tr}(\Delta P_t Y_t^\phi Y_t^{\phi T}) + \text{tr}(\Delta Q Y_t^\phi Y_t^{\phi T}) \geq 0,$$
$$t = 1, \ldots, N-1, \tag{21}$$

where $X_t = [x_t^1, \ldots, x_t^M]$, and where we also used that $\text{tr}(\Delta P_t Y_t^\phi Y_t^{\phi T}) = \text{tr}(\Delta P_t X_t X_t^T)$ and that $\text{tr}(\Delta P_{t+1} Y_{t+1}^\phi Y_{t+1}^{\phi T}) = \text{tr}(\Delta P_{t+1} X_{t+1} X_{t+1}^T)$.

Summing (21) from $t = 1$ to $N - 2$, we have

$$\text{tr}(\Delta P_{N-1} Y_{N-1}^\phi Y_{N-1}^{\phi T}) - \text{tr}(\Delta P_1 Y_1^\phi Y_1^{\phi T})$$
$$+ \sum_{t=1}^{N-2} \text{tr}(\Delta Q Y_t^\phi Y_t^{\phi T}) \geq 0. \tag{22}$$

In view of (17) and (22), we have that

$$0 = -\text{tr}(\Delta P_1 Y_1^\phi Y_1^{\phi T}) + \text{tr}(\Delta P_N Y_N^\phi Y_N^{\phi T}) + \sum_{t=1}^{N-1} \text{tr}(\Delta Q Y_t^\phi Y_t^{\phi T})$$
$$= -\text{tr}(\Delta P_1 Y_1^\phi Y_1^{\phi T}) + \text{tr}(\Delta P_{N-1} Y_{N-1}^\phi Y_{N-1}^{\phi T}) + \sum_{t=1}^{N-2} \text{tr}(\Delta Q Y_t^\phi Y_t^{\phi T})$$
$$- \text{tr}(\Delta P_{N-1} Y_{N-1}^\phi Y_{N-1}^{\phi T}) + \text{tr}(\Delta P_N Y_N^\phi Y_N^{\phi T}) + \text{tr}(\Delta Q Y_{N-1}^\phi Y_{N-1}^{\phi T})$$
$$\geq -\text{tr}(\Delta P_{N-1} Y_{N-1}^\phi Y_{N-1}^{\phi T}) + \text{tr}(\Delta P_N Y_N^\phi Y_N^{\phi T}) + \text{tr}(\Delta Q Y_{N-1}^\phi Y_{N-1}^{\phi T})$$
$$= -\text{tr}(\Delta P_{N-1} X_{N-1} X_{N-1}^T) + \text{tr}(\Delta P_N X_N X_N^T) + \text{tr}(\Delta Q X_{N-1} X_{N-1}^T)$$

Since $X_N = A_{cl}(N - 1; \bar{Q})X_{N-1}$ and by also using (18), from the equation above it follows that

$$0 \geq \text{tr}\Big[\Big(A_{cl}(N - 1; \bar{Q})^T \Delta P_N A_{cl}(N - 1; \bar{Q}) - \Delta P_{N-1} + \Delta Q\Big)$$
$$\times X_{N-1} X_{N-1}^T\Big]$$
$$= \text{tr}\Big[(A^T \Delta P_N A - \Delta P_{N-1} + \Delta Q)X_{N-1} X_{N-1}^T\Big].$$

Since $X_{N-1}X_{N-1}^T$ is positive definite (by Lemma 3.2), it follows that $A^T \Delta P_N A - \Delta P_{N-1} + \Delta Q \preceq 0$. Together with (19), we can therefore conclude that $A^T \Delta P_N A - \Delta P_{N-1} + \Delta Q = 0$. Using (18), we therefore have that $A_{cl}(N - 1; \bar{Q})^T \Delta P_N A_{cl}(N - 1; \bar{Q}) - \Delta P_{N-1} + \Delta Q = 0$, which implies that $-\text{tr}(\Delta P_{N-1} X_{N-1} X_{N-1}^T) + \text{tr}(\Delta P_N X_N X_N^T) + \text{tr}(\Delta Q X_{N-1} X_{N-1}^T) = 0$. Therefore, summing (21) from $t = 1$ to $N - 3$ and reiterating the above analysis, we can conclude that $A^T \Delta P_{N-1} A - \Delta P_{N-2} + \Delta Q = 0$. Doing this recursively, we have that

$$A^T \Delta P_{t+1} A - \Delta P_t + \Delta Q = 0, \quad t = 1, \ldots, N-1. \tag{23}$$

Equipped with (23), we can use the same argument as in the proof of Zhang, Umenberger et al. (2019, Theorem 2.1) to conclude that $\Delta Q = 0$. Thus $Q^* = \bar{Q}$, i.e., the optimal solution of (16) is unique and hence the theorem follows. □

Before proceeding, we make a few remarks about the formulation in (16). As will be seen later, these remarks naturally hold also in the case of noisy observations.

Remark 3.2. The optimization problem (16) can also be applied to the IOC problem for "distinguishable" agents, i.e., when the trajectories of the agents are available. Nevertheless, as one can see from (16), in this formulation of the problem there is no fundamental difference between IOC problems for "distinguishable" and "indistinguishable" agents.

Remark 3.3. The size of the variables and the number of variables and constraints in the optimization problem (16) only depends on the size of the state space, n, and the length of the time horizon, N, respectively. In particular, the overall size of the problem (16) is independent of the number of agents, M. The only quantity that scales with M is the computations of the outer products $Y_t^\phi Y_t^{\phi T}, t = 1, \ldots, N$. The latter scales linearly in M and can be done before solving the optimization problem. Therefore, the problem can be efficiently solved for scenarios with a large number of agents.

4. IOC for indistinguishable agents in the noisy case

Next, we extend the results to the case of noisy observations of the unpaired states. In particular, we show that a modified version of (16) gives a statistically consistent estimator. To this end, let $(\Omega, \mathcal{F}, \mathbb{P})$ be a probability space which carries random vectors $\boldsymbol{x}_1^i \in \mathbb{R}^n$, $\{\boldsymbol{v}_t^i \in \mathbb{R}^n\}_{t=1}^N$, for $i = 1, \ldots, M$. Based on that, the following assumptions are made for the rest of the paper.

Assumption 1 (*I.I.D. Random Variables*). The random vectors $\boldsymbol{x}_1^i \in \mathbb{R}^n$, $\{\boldsymbol{v}_t^i \in \mathbb{R}^n\}_{t=1}^N$, for $i = 1, \ldots, M$, are all independent. Moreover, random vectors $\boldsymbol{x}_1^i \in \mathbb{R}^n$, for $i = 1, \ldots, M$, are identically distributed, and the random vectors $\boldsymbol{v}_t^i \in \mathbb{R}^n$, for $t = 1, \ldots, N$ and $i = 1, \ldots, M$, are identically distributed. Finally, $\mathbb{E}[\boldsymbol{v}_t^i] = 0$, and $\text{cov}(\boldsymbol{v}_t^i, \boldsymbol{v}_t^i) := \Sigma$ is a priori known, where $\|\Sigma\|_F < \infty$.

Assumption 2 (*Stochastic Persistent Excitation*). It holds that $\text{cov}(\boldsymbol{x}_1^i, \boldsymbol{x}_1^i) \succ 0$, where $\mathbb{E}[\|\boldsymbol{x}_1^i\|^2] < \infty$.

Assumption 3 (*Bounded Parameter*). The "true" $\bar{Q}$ that governs the agents lives in a compact set $\bar{\mathbb{S}}_+^n(\varphi) := \{Q \in \mathbb{S}_+^n : \|Q\|_F^2 \leq \varphi\}$, for some $0 < \varphi < \infty$.

Our goal is to seek the "true" $\bar{Q}$ in $\bar{\mathbb{S}}_+^n(\varphi)$. To this end, we define a domain $\mathscr{D}(\varphi)$ similar to $\mathscr{D}$ in (7), namely,

$$\mathscr{D}(\varphi) := \{Q \in \bar{\mathbb{S}}_+^n(\varphi), \{P_t \in \mathbb{S}_+^n\}_{t=1}^N : (6) \text{ holds}\}. \tag{24}$$

Note that in practice, we can set φ arbitrarily large if there is no prior knowledge on the norm bound of possible Q.

Equipped with the stochastic problem set-up, let the initial value x_1^i be a realization of $\boldsymbol{x}_1^i$. Then the optimal state trajectory and optimal control input of the "forward" optimal control (1) can be seen as mappings from Ω to $\mathbb{R}^n$ and $\mathbb{R}^m$, respectively. This means that the states and control signals are in fact stochastic variables, which are parameterized by Q and implicitly determined by

$$\{\bar{\boldsymbol{x}}_t^i(\omega)\}_{t=2}^N, \{\bar{\boldsymbol{u}}_t^i(\omega)\}_{t=1}^{N-1} = \tag{25}$$
$$\begin{cases} \arg\min & J\left(\{\boldsymbol{x}_t^i(\omega)\}_{t=2}^N, \{\boldsymbol{u}_t^i(\omega)\}_{t=1}^{N-1}; \bar{Q}, \boldsymbol{x}_1^i(\omega)\right) \\ \text{subject to} & (1b), (1c), \end{cases}$$

for all $\omega \in \Omega$, and where J is the cost function in (1a). From now on, we omit the "bar" and simply write $\{\boldsymbol{x}_t^i\}$ to denote the corresponding optimal state for the sake of brevity. Next, we assume that the noisy observations of a state is given by $\tilde{\boldsymbol{x}}_t^i = \boldsymbol{x}_t^i + v_t^i$, and that the observed data thus take the form $\boldsymbol{Y}_t^\phi = [\boldsymbol{y}_t^1, \ldots, \boldsymbol{y}_t^M] = [\tilde{\boldsymbol{x}}_t^{\phi t(1)}, \ldots, \tilde{\boldsymbol{x}}_t^{\phi t(M)}]$. Written in a form similar to (3), the measured output can be expressed as

$$\boldsymbol{Y}_t^\phi = (\boldsymbol{X}_t + \boldsymbol{V}_t)\,\bar{\mathscr{P}}_t = \left(\left(\prod_{k=1}^{t-1} A_{cl}(k; \bar{Q})\boldsymbol{X}_1\right) + \boldsymbol{V}_t\right)\bar{\mathscr{P}}_t,$$

where $\boldsymbol{X}_t = [\boldsymbol{x}_t^1, \ldots, \boldsymbol{x}_t^M]$ and $\boldsymbol{V}_t = [v_t^1, \ldots, v_t^M]$. This together with Assumption 1 implies that the columns of $\boldsymbol{Y}_t^\phi$ are I.I.D. A calculation similar to (14) then shows that

$$\mathbb{E}\left[\text{tr}\left(Q\boldsymbol{Y}_t^\phi \boldsymbol{Y}_t^{\phi T}\right)\right] = \mathbb{E}\left[\text{tr}\left(Q\boldsymbol{X}_t\boldsymbol{X}_t^T\right)\right] + M\,\text{tr}\left(Q\,\Sigma\right)$$

where Σ is the covariance of the noise. In particular, the last equality holds since $\boldsymbol{x}_t^i = \prod_{k=1}^{t-1} A_{cl}(k; \bar{Q})\boldsymbol{x}_1^i$, and hence $\boldsymbol{x}_t^i$ is a stochastic variable that is independent of v_t^i, for $t = 1, \ldots, N$ and $i = 1, \ldots, M$. Similar expressions hold for the other terms in (8), and based on this we construct the problem

$$\underset{Q, \{P_t\}_{t=1}^N}{\text{minimize}} \quad \mathbb{E}\left[H_S^{\boldsymbol{Y}}(Q, \{P_t\})\right] \tag{26a}$$

$$\text{subject to} \quad (Q, \{P_t\}_{t=1}^N) \in \mathscr{D}(\varphi), \tag{26b}$$

where

$$H_S^{\boldsymbol{Y}}(Q, \{P_t\}) := \frac{1}{M}\Big[-\text{tr}(P_1 \boldsymbol{Y}_1^\phi \boldsymbol{Y}_1^{\phi T})$$
$$+ \text{tr}(P_N \boldsymbol{Y}_N^\phi \boldsymbol{Y}_N^{\phi T}) + \sum_{t=1}^{N-1} \text{tr}(Q \boldsymbol{Y}_t^\phi \boldsymbol{Y}_t^{\phi T})$$
$$+ M\big(\text{tr}(P_1 \Sigma) - \text{tr}(P_N \Sigma) - (N-1)\text{tr}(Q \Sigma)\big)\Big]. \tag{26c}$$

In particular, note that by a direct calculation it follows that the cost function (26a) can be written as

$$\mathbb{E}\left[H_S^{\boldsymbol{Y}}(Q, \{P_t\})\right] = \frac{1}{M}\mathbb{E}\Big[-\text{tr}(P_1 \boldsymbol{X}_1 \boldsymbol{X}_1^T)$$
$$+ \text{tr}(P_N \boldsymbol{X}_N \boldsymbol{X}_N^T) + \sum_{t=1}^{N-1} \text{tr}(Q \boldsymbol{X}_t \boldsymbol{X}_t^T)\Big]. \tag{27}$$

With this, we can now prove the following result akin Theorem 3.1.

Proposition 4.1. *Let $\bar{Q}$ be the true parameter in (25), and let $\{\bar{P}_t\}_{t=1}^N$ be the corresponding solution to the DRE (2). Under Assumptions 1, 2, and 3, if $N \geq n+1$, then $\left(\bar{Q}, \{\bar{P}_t\}_{t=1}^N\right)$ is the unique optimal solution to (26).*

Proof. Since (27) holds, by adapting the arguments in the proof of Lemma 3.1 it follows that

$$\mathbb{E}[H_S^{\boldsymbol{Y}}(Q, \{P_t\})] + \frac{1}{M}\sum_{i=1}^M \sum_{t=1}^{N-1} \mathbb{E}\left[\|\bar{\boldsymbol{u}}_t^i\|^2\right]$$

$$\geq \frac{1}{M}\sum_{i=1}^M \sum_{t=1}^{N-1} \mathbb{E}\left[\boldsymbol{x}_t^{iT}(K_t - \bar{K}_t)^T(B^T P_{t+1} B + I)(K_t - \bar{K}_t)\boldsymbol{x}_t^i\right]$$

$$= \frac{1}{M}\sum_{t=1}^{N-1} \text{tr}\left[(K_t - \bar{K}_t)^T(B^T P_{t+1} B + I)(K_t - \bar{K}_t)\mathbb{E}\left[\boldsymbol{X}_t \boldsymbol{X}_t^T\right]\right]$$

$$\geq 0, \tag{28}$$

where $\bar{\boldsymbol{u}}_t^i$ is the stochastic optimal control input signal of the agent i at time instant t. Next, by using (28) and following along the lines of the proof of Lemma 3.1, it can be seen that $(\bar{Q}, \{\bar{P}_t\})$ is an optimal solution to (26), and that

$$\mathbb{E}\left[H_S^{\boldsymbol{Y}}(\bar{Q}, \{\bar{P}_t\})\right] + \frac{1}{M}\sum_{i=1}^M \sum_{t=1}^{N-1} \mathbb{E}\left[\|\bar{\boldsymbol{u}}_t^i\|^2\right] = 0. \tag{29}$$

Therefore, what remains is to show the uniqueness of the optimal solution. To this end, by Assumption 2 the covariance matrix $\text{cov}(\boldsymbol{x}_1^i, \boldsymbol{x}_1^i)$ is strictly positive definite, and therefore the second-order moment of the initial value $\mathbb{E}[\boldsymbol{x}_1^i \boldsymbol{x}_1^{iT}] = \text{cov}(\boldsymbol{x}_1^i, \boldsymbol{x}_1^i) + \mathbb{E}[\boldsymbol{x}_1^i]\mathbb{E}[\boldsymbol{x}_1^i]^T \succ 0$. By the fact that $A_{cl}(t; Q)$ is invertible for all $Q \in \mathbb{S}_+^n$, it holds that

$$\mathbb{E}[\boldsymbol{x}_t^i \boldsymbol{x}_t^{iT}] = \left[\prod_{k=1}^{t-1} A_{cl}(k; Q)\right]\mathbb{E}[\boldsymbol{x}_1^i \boldsymbol{x}_1^{iT}]\left[\prod_{k=t-1}^{1} A_{cl}^T(k; Q)\right] \succ 0,$$

cf. Lemma 3.2, and hence $\mathbb{E}[\boldsymbol{X}_t \boldsymbol{X}_t^T] \succ 0$. Now suppose that there exists some other $(Q^*, \{P_t^*\})$ that is also optimal to (26). By (28), (29), we have

$$0 = \mathbb{E}[H_S^{\boldsymbol{Y}}(Q^*, \{P_t^*\})] + \frac{1}{M}\sum_{i=1}^M \sum_{t=1}^{N-1} \mathbb{E}\left[\|\bar{\boldsymbol{u}}_t^i\|^2\right]$$

$$\geq \frac{1}{M}\sum_{t=1}^{N-1} \text{tr}\left[(K_t^* - \bar{K}_t)^T(B^T P_{t+1} B + I)(K_t^* - \bar{K}_t)\mathbb{E}\left[\boldsymbol{X}_t \boldsymbol{X}_t^T\right]\right]$$

$$= \frac{1}{M}\sum_{t=1}^{N-1} \left\|(B^T P_{t+1} B + I)^{\frac{1}{2}}(K_t^* - \bar{K}_t)\mathbb{E}\left[\boldsymbol{X}_t \boldsymbol{X}_t^T\right]^{\frac{1}{2}}\right\|_F^2$$

$$\geq 0,$$

which together with the fact that $\mathbb{E}[\boldsymbol{X}_t \boldsymbol{X}_t^T] \succ 0$ implies that

$$(B^T P_{t+1} B + I)^{\frac{1}{2}}(K_t^* - \bar{K}_t)\mathbb{E}\left[\boldsymbol{X}_t \boldsymbol{X}_t^T\right]^{\frac{1}{2}} = 0,$$

for $t = 1, \ldots, N-1$. The latter in turn implies that

$$K_t^* - \bar{K}_t = 0, \forall t = 1, \ldots, N-1,$$

where K_t^* denote the same expression as the one in the proof of Theorem 3.1. To this end, following the same analysis as in the proof of Theorem 3.1, we can conclude that $Q^* = \bar{Q}$ and hence the statement holds. $\square$

Since the distribution of the initial values $\boldsymbol{x}_1^i$ and the additive noise v_t^i are not known, we cannot express the expected value of the objective function (26c) explicitly and hence it is not possible

to solve (26) directly. Therefore, we derive an empirical estimate of the expectation based on the observations. To this end, first recall that due to Assumption 1, $\boldsymbol{y}_t^i$ are I.I.D. for $i = 1, \ldots, M$. This means that we can rewrite the cost (26a) as

$$\mathbb{E}\left[H_S^{\boldsymbol{Y}}(Q, \{P_t\})\right] = \mathbb{E}\left[- \operatorname{tr}(P_1 \boldsymbol{y}_1 \boldsymbol{y}_1^T) + \operatorname{tr}(P_N \boldsymbol{y}_N \boldsymbol{y}_N^T) + \sum_{t=1}^{N-1} \operatorname{tr}(Q \boldsymbol{y}_t \boldsymbol{y}_t^T)\right]$$
$$+ \operatorname{tr}(P_1 \Sigma) - \operatorname{tr}(P_N \Sigma) - (N-1)\operatorname{tr}(Q \Sigma), \quad (30)$$

where, for $t = 1, \ldots, N$, $\boldsymbol{y}_t$ is a random variable with the same distribution as $\boldsymbol{y}_t^i$ for $i = 1, \ldots, M$. This means that an empirical estimate of the expectation of (26c) can be obtained as

$$\mathbb{E}\left[\left(-\boldsymbol{y}_1^T P_1 \boldsymbol{y}_1 + \boldsymbol{y}_N^T P_N \boldsymbol{y}_N + \sum_{t=1}^{N-1} \boldsymbol{y}_t^T Q \boldsymbol{y}_t\right)\right]$$
$$\approx \frac{1}{M} \sum_{i=1}^{M} \left(-\boldsymbol{y}_1^{iT} P_1 \boldsymbol{y}_1^i + \boldsymbol{y}_N^{iT} P_N \boldsymbol{y}_N^i + \sum_{t=1}^{N-1} \boldsymbol{y}_t^{iT} Q \boldsymbol{y}_t^i\right).$$

Based on this, we formulate the estimation problem

$$\underset{Q, \{P_t\}_{t=1}^{N}}{\text{minimize}} \quad H_E^{\boldsymbol{Y}}(Q, \{P_t\}) \quad (31a)$$

$$\text{subject to} \quad (Q, \{P_t\}_{t=1}^{N}) \in \mathscr{D}(\varphi), \quad (31b)$$

where

$$H_E^{\boldsymbol{Y}}(Q, \{P_t\}) := \frac{1}{M}\left[- \operatorname{tr}(P_1 \boldsymbol{Y}_1^\phi \boldsymbol{Y}_1^{\phi T})\right.$$
$$+ \operatorname{tr}(P_N \boldsymbol{Y}_N^\phi \boldsymbol{Y}_N^{\phi T}) + \sum_{t=1}^{N-1} \operatorname{tr}(Q \boldsymbol{Y}_t^\phi \boldsymbol{Y}_t^{\phi T})\right]$$
$$+ \left(\operatorname{tr}(P_1 \Sigma) - \operatorname{tr}(P_N \Sigma) - (N-1)\operatorname{tr}(Q \Sigma)\right). \quad (31c)$$

The problem (31) defines the estimator, and for a given realization $\{Y_t^\phi\}_{t=1}^{N}$ of the stochastic variables $\{\boldsymbol{Y}_t^\phi\}_{t=1}^{N}$, it can be solved in order to obtain an estimate. In particular, we use the notation $H_E^{\boldsymbol{Y}}(Q, \{P_t\})|_{\boldsymbol{Y}=Y}$ to denote the cost function (31c) evaluated at a particular realization. We now want to show that this estimator is in fact (asymptotically) statistically consistent. However, note that since we approximate the expected value in the objective function by the empirical average, the objective function changes and hence the "bounded-from-below" argument (28) does not necessarily hold for $H_E^{\boldsymbol{Y}}(Q, \{P_t\})|_{\boldsymbol{Y}=Y}$ on the domain $\mathscr{D}(\varphi)$. This issue needs to be addressed in order to make (31) well-posed. This is an important first step towards showing that the estimator is statistically consistent.

Lemma 4.1. *The domain $\mathscr{D}(\varphi)$, defined in (24), is compact, and $H_E^{\boldsymbol{Y}}(Q, \{P_t\})|_{\boldsymbol{Y}=Y}$ is bounded on $\mathscr{D}(\varphi)$.*

Proof. Consider the domain $\mathscr{D}(\varphi)$ and recall that, by the property of Schur complement, (9) holds on the feasible domain (6a) and (6b). Since the Frobenius norm is monotone with respect to the Loewner partial order, it holds that

$$\|P_t\|_F \leq \|A^T P_{t+1} A + Q - A^T P_{t+1} B(B^T P_{t+1} B + I)^{-1} B^T P_{t+1} A\|_F$$

By the Cauchy–Schwarz and the triangular inequality, we have that

$$\|P_t\|_F \leq \|A\|_F^2 \cdot \|P_{t+1}\|_F + \|Q\|_F$$
$$+ \|A\|_F^2 \cdot \|B\|_F^2 \cdot \|P_{t+1}\|_F^2 \cdot \|(B^T P_{t+1} B + I)^{-1}\|_F. \quad (32)$$

Next, since $P_t \in \mathbb{S}_+^n$, it holds that $B^T P_t B + I \succeq I$, and hence that $(B^T P_t B + I)^{-1} \preceq I$ (Horn & Johnson, 2013, Cor. 7.7.4). By monotonicity of the Frobenius norm with respect to the Loewner partial order, it therefore follows that $\|(B^T P_t B + I)^{-1}\|_F \leq \|I\|_F =$

$\sqrt{n}$. Now, since $\|Q\|_F^2 \leq \varphi$ and $P_N = Q$, using this together with (32), it follows that $\|P_{N-1}\|_F$ is bounded. Recursively applying this backwards for the time indices t, it follows that $\|P_t\|_F$ is bounded for all $t = 1, \ldots, N$. This implies that the domain $\mathscr{D}(\varphi)$ is compact. Finally, since $H_E^{\boldsymbol{Y}}(Q, \{P_t\})|_{\boldsymbol{Y}=Y}$ is continuous, it is bounded on $\mathscr{D}$. $\square$

Remark 4.1. Note that Assumption 3 is critical in the proof of Lemma 4.1, since we can thus optimize over $\mathscr{D}(\varphi)$ instead of $\mathscr{D}$. In fact, $H_E^{\boldsymbol{Y}}(Q, \{P_t\})|_{\boldsymbol{Y}=Y}$ might not be bounded from below if we only impose $Q \in \mathbb{S}_+^n$. To see this, assume that there exists a realization $\{Y_t^\phi\}$ of $\{\boldsymbol{Y}_t^\phi\}$ such that $Y_1^\phi Y_1^{\phi T} - M\Sigma = 0$ and $Y_t^\phi Y_t^{\phi T} - M\Sigma \prec 0$, $t = 2, \ldots, N$. Let $(Q, \{P_t\})$ satisfies DRE (2), and note that then $(\alpha Q, \{\alpha P_t\})$ also satisfies DRE (2) for any positive α. Therefore, $(\alpha Q, \{\alpha P_t\}) \in \mathscr{D}$ for all $\alpha > 0$, and thus $\mathscr{D}$ is not a bounded set. Moreover, as $\alpha \to \infty$ it holds that

$$H_E^{\boldsymbol{Y}}(\alpha Q, \{\alpha P_t\})|_{\boldsymbol{Y}=Y} = \frac{1}{M}\left[- \operatorname{tr}\left(\alpha P_1(Y_1^\phi Y_1^{\phi T} - M\Sigma)\right)\right.$$
$$+ \operatorname{tr}\left(\alpha P_N(Y_N^\phi Y_N^{\phi T} - M\Sigma)\right) + \sum_{t=1}^{N-1} \operatorname{tr}\left(\alpha Q(Y_t^\phi Y_t^{\phi T} - M\Sigma)\right)\right]$$
$$\to -\infty,$$

hence $H_E^{\boldsymbol{Y}}(\alpha Q, \{\alpha P_t\})|_{\boldsymbol{Y}=Y}$ is not bounded from below on $\mathscr{D}$.

Next, we show that the "Uniform Law of Large Numbers" holds for $H_E^{\boldsymbol{Y}}(Q, \{P_t\})$.

Lemma 4.2 (*Uniform Law of Large Numbers*). *Under Assumptions 1, 2 and 3, it holds that*

$$\sup_{(Q, \{P_t\}_{t=1}^{N}) \in \mathscr{D}(\varphi)} \left|H_E^{\boldsymbol{Y}}(Q, \{P_t\}) - \mathbb{E}\left[H_S^{\boldsymbol{Y}}(Q, \{P_t\})\right]\right| \xrightarrow{a.s.} 0,$$

as $M \to \infty$.

Proof. It is clear that $H_S^{\boldsymbol{Y}}(Q, \{P_t\})$ is continuous with respect to $\{Y_t^\phi\}_{t=1}^{N}$ and therefore it is a measurable function of $\{Y_t^\phi\}_{t=1}^{N}$ for each Q and $\{P_t\}$. On the other hand, Assumption 2 implies that $\mathbb{E}[\|\boldsymbol{x}_t^i\|^2] < \infty$ for $i = 1, \ldots, M$ (Zhang, Umenberger et al., 2019, cf. the proof of Theorem 4.1). By Assumption 1, we have $\mathbb{E}[\|\boldsymbol{v}_t^i\|^2] < \infty$. Since $\boldsymbol{x}_t^i$ is independent of $\boldsymbol{v}_t^i$, and since $\mathbb{E}[\boldsymbol{v}_t^i] = 0$, it follows that

$$\mathbb{E}[\|\boldsymbol{y}_t^i\|^2] = \mathbb{E}[(\boldsymbol{x}_t^i + \boldsymbol{v}_t^i)^T(\boldsymbol{x}_t^i + \boldsymbol{v}_t^i)] = \mathbb{E}[\|\boldsymbol{x}_t^i\|^2] + \mathbb{E}[\|\boldsymbol{v}_t^i\|^2] < \infty.$$

In addition, from Lemma 4.1, we know that there exists constants $\{\bar{\varphi}_t\}_{t=1}^{N}$ such that for all $(Q, \{P_t\}_{t=1}^{N}) \in \mathscr{D}(\varphi)$ we have that $\|P_t\|_F \leq \bar{\varphi}_t$. Using the form of (26a) given in (30), by the Cauchy–Schwarz and the triangular inequality it therefore holds that

$$H_S^{\boldsymbol{Y}}(Q, \{P_t\}) = - \operatorname{tr}(P_1 \boldsymbol{y}_1 \boldsymbol{y}_1^T) + \operatorname{tr}(P_N \boldsymbol{y}_N \boldsymbol{y}_N^T) + \sum_{t=1}^{N-1} \operatorname{tr}(Q \boldsymbol{y}_t \boldsymbol{y}_t^T)$$
$$+ \operatorname{tr}(P_1 \Sigma) - \operatorname{tr}(P_N \Sigma) - (N-1)\operatorname{tr}(Q \Sigma)$$
$$\leq \left|\boldsymbol{y}_1^T P_1 \boldsymbol{y}_1\right| + \left|\boldsymbol{y}_N^T P_N \boldsymbol{y}_N\right| + \sum_{t=1}^{N-1} \left|\boldsymbol{y}_t^T Q \boldsymbol{y}_t\right|$$
$$+ |\operatorname{tr}(P_1 \Sigma)| + |\operatorname{tr}(P_N \Sigma)| + (N-1)|\operatorname{tr}(Q \Sigma)|$$
$$\leq \|\boldsymbol{y}_1\|^2 \|P_1\|_F + \|\boldsymbol{y}_N\|^2 \|P_N\|_F + \sum_{t=1}^{N-1} \|\boldsymbol{y}_t\|^2 \|Q\|_F$$
$$+ \|P_1\|_F \|\Sigma\|_F + \|P_N\|_F \|\Sigma\|_F + (N-1)\|Q\|_F \|\Sigma\|_F$$
$$\leq \bar{\varphi}_1(\|\boldsymbol{y}_1\|^2 + \|\Sigma\|_F) + \bar{\varphi}_N(\|\boldsymbol{y}_N\|^2 + \|\Sigma\|_F)$$
$$+ \varphi \sum_{t=1}^{N-1}(\|\boldsymbol{y}_t\|^2 + \|\Sigma\|_F) := d(\{\boldsymbol{y}_t\}),$$

and it is clear that $\mathbb{E}[d(\{\boldsymbol{y}_t\})] < \infty$ since $\mathbb{E}[\|\boldsymbol{y}_t^i\|^2] < \infty$. Therefore, by Jennrich (1969, Thm. 2) the result follows. $\quad\square$

We are now ready to prove the main result of this section.

Theorem 4.1 (*Statistical Consistency*). *Suppose that* $(Q_M^*, \{P_{t,M}^*\}_{t=1}^N)$ *is an optimal solution to* (31) *when observing M agents. Then* $Q_M^* \xrightarrow{p} \bar{Q}$ *as* $M \to \infty$, *where* $\bar{Q}$ *is the true parameter used in the objective function of "forward" problem* (25).

Proof. The theorem is proved by showing that all the conditions in van der Vaart (1998, Thm. 5.7) are satisfied. To this end, the first condition follows from Lemma 4.2, since convergence a.s. implies convergence in probability (Kallenberg, 1997, Lem 3.2). Next, the second condition holds since by Proposition 4.1 the optimal solution to (26) is unique, and since by Lemma 4.1 the set $\mathscr{D}(\varphi)$ is compact; see van der Vaart (1998, p. 46). Therefore, all conditions in van der Vaart (1998, Thm. 5.7) are satisfied, and the statement hence follows. $\quad\square$

5. On numerical ill-conditioning

Proposition 2.1 shows that the model is globally identifiable, and Theorem 3.1 shows that the optimization problem (16) has a unique optimal solution at the "true" $\bar{Q}$. Hence, in theory the latter can be recovered by solving the optimization problem. Nevertheless, recovering this optimal solution turns out to be numerically difficult, in particular for certain problem instances − this will be demonstrated with examples in Section 6. Here, we argue that this has to do with an intrinsic numerical ill-conditioning of the inverse problem for these problem instances.

To this end, recall that Pontryagin's Maximum Principle (PMP) gives a necessary and sufficient condition for optimality in the forward problem (1). Namely, $\{x_t\}_{t=1}^N$ and $\{u_t\}_{t=1}^{N-1}$ are the optimal trajectory and control signal of (1), respectively, if and only if there exists adjoint states $\{\lambda_t\}_{t=2}^N$ such that

$$\lambda_t = A^T \lambda_{t+1} + Q x_t, \ t = 2, \ldots, N-1, \tag{33a}$$

$$\lambda_N = Q x_N, \tag{33b}$$

$$u_t = -B^T \lambda_{t+1}, \ t = 1, \ldots, N-1. \tag{33c}$$

Based on (33a)–(33b), we can write a linear system of equations for the adjoint variables of the ith agent, namely

$$\begin{bmatrix} I & -A^T & & \\ & I & \ddots & \\ & & \ddots & -A^T \\ & & & I \end{bmatrix} \underbrace{\begin{bmatrix} \lambda_2^i \\ \vdots \\ \lambda_N^i \end{bmatrix}}_{=: \lambda^i} = (I \otimes Q) \underbrace{\begin{bmatrix} x_2^i \\ \vdots \\ x_N^i \end{bmatrix}}_{=: x_{2:N}^i}.$$

Solving this for the adjoint variables gives

$$\lambda^i = \begin{bmatrix} I & A^T & (A^T)^2 & \cdots & (A^T)^{N-2} \\ & I & A^T & \cdots & (A^T)^{N-3} \\ & & \ddots & \ddots & \vdots \\ & & & I & A^T \\ & & & & I \end{bmatrix} (I \otimes Q) x_{2:N}^i,$$

and substituting the latter into (33c) we obtain

$$- \underbrace{\begin{bmatrix} u_1^i \\ \vdots \\ u_{N-1}^i \end{bmatrix}}_{=: u^i} = (I_{N-1} \otimes B^T) \lambda^i$$

$$= \begin{bmatrix} B^T & B^T A^T & B^T (A^T)^2 & \cdots & B^T (A^T)^{N-2} \\ & B^T & B^T A^T & \cdots & B^T (A^T)^{N-3} \\ & & \ddots & \ddots & \vdots \\ & & & B^T & B^T A^T \\ & & & & B^T \end{bmatrix} (I_{N-1} \otimes Q) x_{2:N}^i, \tag{34}$$

$$\underbrace{}_{= [\mathcal{S}_2\Gamma,\ldots,\mathcal{S}_N\Gamma] = \mathcal{S}(I_{N-1}\otimes\Gamma)}$$

where $\mathcal{S} := [\mathcal{S}_2, \ldots, \mathcal{S}_N]$, $\mathcal{S}_N := I_{m(N-1)}$, $\mathcal{S}_{N-k} \in \mathbb{R}^{m(N-1)\times m(N-1)}$ is a block-matrix with identity matrices of size $n \times n$ on the kth upper block-diagonal, and

$$\Gamma := \begin{bmatrix} (A)^{N-2}B & (A)^{N-3}B & \cdots & B \end{bmatrix}^T.$$

Using (34) and the fact that

$$x_t^i = \begin{bmatrix} A^{t-2}B & \cdots & AB & B \end{bmatrix} \begin{bmatrix} u_1^{iT}, \ldots, u_{t-1}^{iT} \end{bmatrix}^T + A^{t-1} x_1^i,$$

we have

$$x_{2:N}^i = \underbrace{\begin{bmatrix} B & 0 & \cdots & 0 & 0 \\ AB & B & \cdots & 0 & 0 \\ \vdots & \vdots & & \vdots & \vdots \\ A^{N-1}B & A^{N-2}B & \cdots & AB & B \end{bmatrix}}_{=(I_{N-1}\otimes\Gamma^T)\mathcal{S}^T} u^i + \underbrace{\begin{bmatrix} A \\ A^2 \\ \vdots \\ A^{N-1} \end{bmatrix}}_{=: \tilde{A}} x_1^i,$$

$$= -\underbrace{(I_{N-1} \otimes \Gamma^T)\mathcal{S}^T \mathcal{S}(I_{N-1} \otimes \Gamma)(I_{N-1} \otimes Q)}_{=: \mathscr{F}(Q)} x_{2:N}^i + \tilde{A} x_1^i$$

$$\implies (I_{(N-1)n} + \mathscr{F}(Q)) x_{2:N}^i = \tilde{A} x_1^i$$

Since (1) has a unique solution, $I_{(N-1)n} + \mathscr{F}(Q)$ is intrinsically invertible. Therefore, $(I_{(N-1)n} + \mathscr{F}(Q))^{-1}\tilde{A}$ is another representation of the same model structure as $\mathcal{M}(Q, \{\mathcal{P}_t\}_{t=2}^N)$, if the former is also combined with the permutation matrices $\{\mathscr{P}_t\}_{t=2}^N$; for the sake of brevity we omit the details.

From above, it can be seen that the model structure is identifiable at $\bar{Q}$ if and only if

$$\Upsilon := \left\{ \operatorname{Im}(\Delta Q) \mid \Delta Q \in \mathbb{S}^n, \bar{Q} + \Delta Q \in \mathbb{S}_+^n \right\} \cap \operatorname{Ker}(\Gamma) = \{0\}. \tag{35}$$

Assuming that the system (A, B) is controllable implies that the controllability matrix

$$\Gamma_n = \begin{bmatrix} (A)^{n-1}B & \cdots & AB & B \end{bmatrix}^T$$

has full column-rank, and hence that $\operatorname{Ker}(\Gamma) = \{0\}$. The latter, in turn, means that (35) is fulfilled. Nevertheless, in practise, if Γ is ill-conditioned, the kernel of Γ can be "expanded" from a numerical perspective. In this case we can have $\Gamma \Delta Q \approx 0$ for some ΔQ that is not close to zero, and the set Υ might numerically not be the singleton $\{0\}$. Thus, it is possible that

$$(I_{(N-1)n} + \mathscr{F}(\bar{Q} + \Delta Q))^{-1}\tilde{A} \approx (I_{(N-1)n} + \mathscr{F}(\bar{Q}))^{-1}\tilde{A}$$

for some feasible ΔQ that is not close to zero. Therefore, when the controllability matrix is ill-conditioned, we might have very similar models that corresponds to very different Q's. In general, we expect that it will be numerically challenging to recover the "true" $\bar{Q}$ in these settings, regardless of which method is used. However, although it might be numerically difficult to recover $\bar{Q}$ in these circumstances, the proof of Lemma 3.1 ensures that the control gains corresponding to $\bar{Q} + \Delta Q$ numerically coincides with the "true" control gain, which is sufficient for predicting the agents' behaviours. Finally, a similar argument holds in the case of noisy observations.

6. Numerical experiments and discussions

In this section, we present a number of numerical experiments, performed on a number of different discrete-time systems, to illustrate the properties of the proposed algorithm. In particular, the discrete-time systems are all generated by sampling continuous-time systems $\dot{x} = \hat{A}x + \hat{B}u$ via $A = e^{\hat{A}\Delta t}$ and $B = \int_0^{\Delta t} e^{\hat{A}t} dt \hat{B}$, where the sampling period $\Delta t = 0.05$.

All numerical examples are run on a MacBook Pro with Apple M1 eight-core CPU and 16 GB of RAM. The solutions are obtained by implementing the optimization problems in Matlab using YALMIP (Löfberg, 2004) and solving them using MOSEK (MOSEK ApS, 2019).

6.1. Noiseless case

In this experiment, the dimension of the system is set to $n = 3$, and $m = 1$. More specifically, we randomly generate system matrices $\hat{A} \in \mathbb{R}^{3\times3}$ and $\hat{B} \in \mathbb{R}^{3\times1}$ with entries drawn from a normal distribution with mean value zero and standard deviation one, i.e., with entries drawn from the distribution $\mathcal{N}(0, 1)$. These are then sampled to generate discrete-time systems, as described above. Moreover, the "true" $\bar{Q}$ is randomly generated as $\bar{Q} = GG^T$, where $G \in \mathbb{R}^{3\times3}$ with entries drawn from $\mathcal{N}(0, 1)$. We let $\bar{Q} \in \bar{\mathbb{S}}_+^3(5)$; any randomly generated $\bar{Q}$ would be discarded if it does not belong to $\bar{\mathbb{S}}_+^3(5)$ and another random $\bar{Q}$ would be generated. In this way, 500 random triplets $(A, B, \bar{Q})$ are generated. For each such random triplet, we set the time horizon to $N = 20$ and generate $M = 15$ random starting points x_1^i. The latter are drawn from a uniform distribution on $[-10, 10] \times [-10, 10] \times [-10, 10]$. The forward problem (1) is solved for each starting point, and the noiseless data is then used to solve the optimization problem in (16), except that the cost function is scaled with 10^{-4} in order to give a better numerical scaling for the problem. While this does not change any analytic properties, the obtained optimal solutions were observed to have a smaller relative error in general. The results are presented in Figs. 1 and 2.

The lower plot in Fig. 1 illustrates the absolute value of the scaled objective function value at the theoretical optimal solution $(\bar{Q}, \{\bar{P}_t\})$, together with the absolute value of the difference between the objective function value at $(\bar{Q}, \{\bar{P}_t\})$ and at the solution $(Q_{est}, \{P_{t,est}\})$ obtained with the solver. As can be seen, the difference is in general several orders of magnitude smaller than the optimal value of the cost function, despite the fact that the obtained Q_{est} is sometimes relatively far from $\bar{Q}$ (cf. upper plot in Fig. 1). This indicates that the cost function is "flat" in a region around the optimal solution, which makes the "true" $\bar{Q}$ hard to recover numerically with high accuracy. We believe that this is highly related to the fact that the controllability matrix is very ill-conditioned (the condition numbers of which varies from 2.899×10^2 to 7.439×10^5). Mitigation of this numerical difficulties is left for future work.

Nevertheless, as can be seen from Fig. 2, the corresponding control gain and closed loop system matrix are well-recovered, which serves the purpose of "predicting the agent's behaviour". In fact, the latter is in general recovered with better accuracy than the former. This indicates that for certain systems, a larger mismatch in Q can still give small mismatches in the control gains and the closed-loop system matrices. Hence it is harder to identify Q numerically in these cases. Since the control gain and the closed-loop system are time-varying, the smallest and largest relative error over all time points are shown in Fig. 2.

The fact that the closed-loop system matrix in general seems to be better recovered than $\bar{Q}$ seems to indicate that the "flatness"
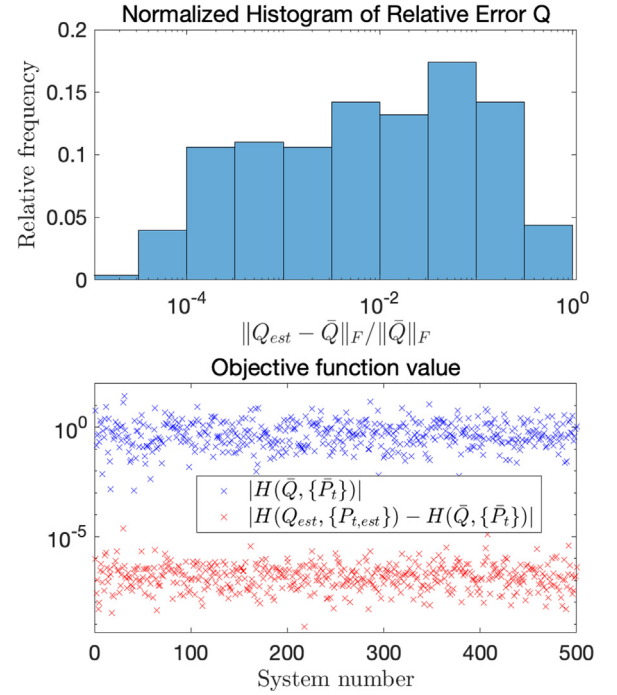
Fig. 1. The upper plots shows a (normalized) histogram of the relative error in the estimate Q_{est} obtained with noiseless data, as described in Section 6.1. The lower plots shows the absolute value of the (scaled) globally optimal objective function value, as well as absolute value of the difference between the globally optimal objective function value and the objective function value in the obtained estimate.

of the cost function for certain problem instances is (at least partly) related to the discussion in Section 5. Namely, that for certain problem instances, substantially different Q's can give rise to very similar closed-loop system matrices.

6.2. Noisy case

Next, we illustrate the statistical consistency of the method. This is done on a dynamical system which does not have a numerically ill-conditioned controllability Gramian. More specifically, the continuous-time dynamics is given by

$$\hat{A} = \begin{bmatrix} 0 & 0 \\ 0 & 0 \end{bmatrix}, \qquad \hat{B} = \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix},$$

which is the kinematic dynamics of a point mass that moves on a two-dimensional plane. In fact, the condition number of the controllability Gramian for this dynamical system is 1. Moreover, with the corresponding discretized system matrices (A, B), for any $\bar{Q} \succ 0$ the forward problem (1) describes an agent that moves towards the origin. Hence, a group of homogeneous agents that are all governed by (1) is a model for the simplified setting of "non-interacting" agents moving towards a common goal at the origin. As mentioned in the introduction, in the future we intend to extend the work to interacting agents.
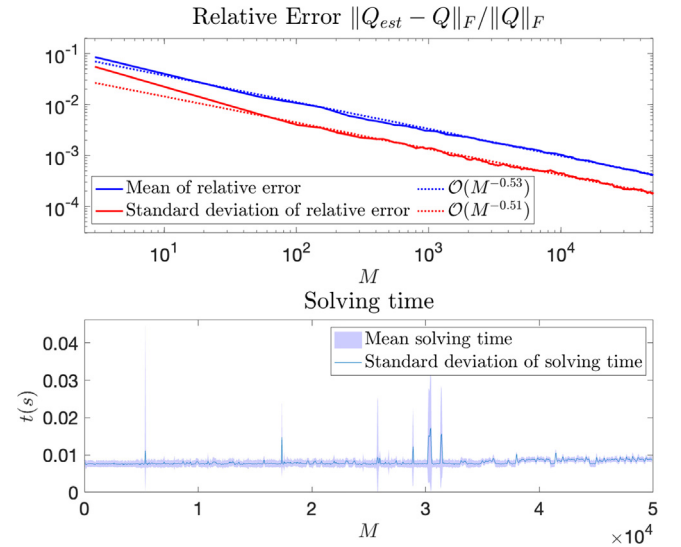
The "true" $\bar{Q}$ is generated as described in Section 6.1, and we set the time horizon to $N = 20$. We then generate 49953 random starting points x_1^i, drawn from a uniform distribution supported on $[-10, 10] \times [-10, 10]$. For each agent, the forward problem (1) is solved, and noise is added on the obtained optimal states (including the initial state). The additive noise is drawn from a multi-variate zero-mean normal distribution with covariances matrix drawn from a Wishart distribution of degree 2, i.e., with the same number of degrees of freedom as the dimension of the
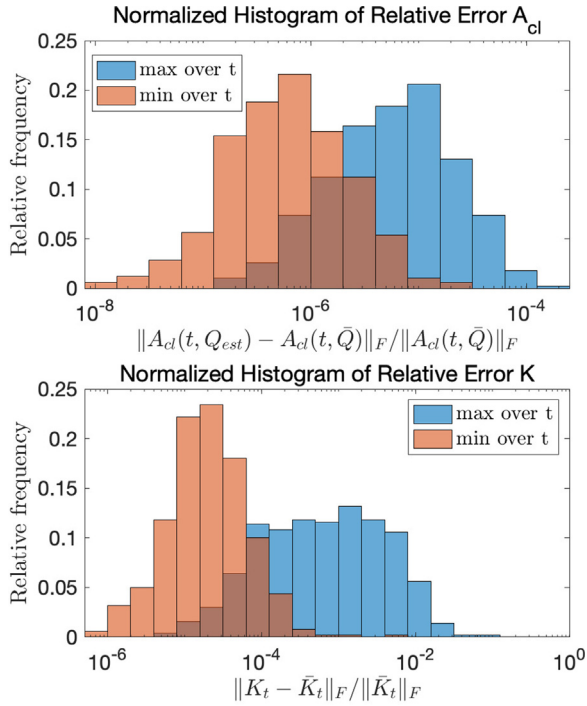
Fig. 2. Histograms (normalized) of the relative error in the estimates of the closed-loop system matrix $A_{cl}(t, Q_{est})$ and the gain K_t, obtained from noiseless data as described in Section 6.1. Since the quantities are time-varying, the largest and smallest relative errors are shown.

state space. The Wishart distribution has a random covariance generated as $0.02GG^T$, where each element in $G \in \mathbb{R}^{2 \times 2}$ was drawn from $\mathcal{N}(0, 1)$. The trajectories are then divided into groups of size $M = 3 + 50(k - 1)$, for $k = 1, \ldots, 1000$, where each larger group contains all the trajectories of a smaller group. For the fixed triplet $(A, B, \bar{Q})$, the above process is repeated for 100 times so that, in total, we get 100 noisy data sets with a varying number of agents in each. Moreover, the signal-to-noise ratio (SNR) in the data sets varies between 29.2479 and 29.3767 dB.[2] For each data set and each trajectory number M, which varies from 3 to 49953, the problem (31) is solved, using the corresponding cost function $H_S^{(\mathbf{Y},M)}(Q, \{P_t\})|_{\mathbf{Y}=Y}$. That means that for each fixed M, we get 100 estimates of $\bar{Q}$, and from these 100 estimates we calculate the mean and the standard deviation of the relative error $\|Q_{est} - \bar{Q}\|_F / \|\bar{Q}\|_F$. The result, as a function of M, is shown in Fig. 3.

From the upper plot in Fig. 3 we see that both the mean and the standard deviation of the relative error of the estimates decreases with increasing M, in line with the statistical consistency of the estimate as proved in Theorem 4.1. Moreover, in the log–log plot of the mean and the standard deviation of the estimates v.s. M, we can see that the relation is approximately linear. Fitting a log-linear model to the data, i.e., fitting an affine function to the logarithmic data, we get that `Mean of relative error` $\approx \mathcal{O}(M^{-0.53})$ and `Standard deviation of relative error` $\approx \mathcal{O}(M^{-0.51})$. The corresponding lines are also shown in Fig. 3. The orders are close to -0.5, and hence we suspect that the convergence rate is $\mathcal{O}(M^{-0.5})$ and that $\sqrt{M}(Q_M - \bar{Q})$ is asymptotically normal, just like most M-estimators such as maximum

---

[2] The SNR in a data set is computed as the mean of the SNR for all trajectories in that data set, where the SNR of a trajectory is computed as the sum of the squared norm of all states divided by the sum of the squared norm of all noise realizations.

Fig. 3. The upper plot shows the mean and standard deviation of the relative error of Q_{est} as a function of the number of agents. The estimates are obtained using noisy data, as described in Section 6.2. Moreover, the lower plot shows the time it took (in seconds) to solve the corresponding optimization problem.

log-likelihood (van der Vaart, 1998, p. 51). Further analysis of this is left for future work.

Finally, from the lower plot in Fig. 3 we see that, as expected from Remark 3.3, the time to solve the corresponding optimization problem does note scale with the number of agents M.[3]

7. Conclusions

In this work we have considered the linear-quadratic inverse optimal control problem in discrete time and with finite time horizon, but where the observed homogeneous agents are indistinguishable. In the case of exact measurements of the states, we show that the true parameter $\bar{Q}$ can be recovered as the unique globally optimal solution to a semidefinite programming problem. Moreover, the size of this convex optimization problem is independent of the number of agents observed, and the formulation is thus suitable also for scenarios with a large number of agents. Furthermore, in the case of noisy state observations the optimization problem is modified, and a statistically consistent estimator is obtained as the unique globally optimal solution to another semidefinite programming problem. However, from numerical simulations it seems that for certain problem instances the cost functions in both semidefinite programmes are relatively flat around the globally optimal solutions, and hence accurate estimates of the parameters are difficult to recover. Nevertheless, estimates of the time-varying control gains and closed-loop system matrices are obtained with higher accuracy. An analysis of numerical ill-conditioning of the problem seem to suggest that this could be linked to the fact that for certain problem instances, different Q's can give rise to similar close-loop system behaviour. An important open question is therefore if and how the numerical conditioning of the estimators can be improved.

References

Alexander, R. McNeill (1996). *Optima for animals*. Princeton, NJ: Princeton University Press.

---

[3] The solving times are the times returned by YALMIP.

Anderson, Brian D. O., & Moore, John B. (2007). *Optimal control: linear quadratic methods*. Mineola, NY: Dover publications.

Aswani, Anil, Shen, Zuo-Jun, & Siddiq, Auyon (2018). Inverse optimization with noisy data. *Operations Research*, 66(3), 870–892.

Berret, Bastien, Chiovetto, Enrico, Nori, Francesco, & Pozzo, Thierry (2011). Evidence for composite cost functions in arm movement planning: an inverse optimal control approach. *PLoS Computational Biology*, 7(10), Article e1002183.

Boyd, Stephen, El Ghaoui, Laurent, Feron, Eric, & Balakrishnan, Venkataramanan (1994). *Linear matrix inequalities in system and control theory*. Philadelphia, PA: SIAM.

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

Fridovich-Keil, David, Ratner, Ellis, Peters, Lasse, Dragan, Anca D, & Tomlin, Claire J (2020). Efficient iterative linear-quadratic approximations for nonlinear multi-player general-sum differential games. In *2020 IEEE international conference on robotics and automation (ICRA)* (pp. 1475–1481). IEEE.

Hatz, Kathrin, Schloder, Johannes P., & Bock, Hans Georg (2012). Estimating parameters in optimal control problems. *SIAM Journal on Scientific Computing*, 34(3), A1707–A1728.

Horn, Roger A., & Johnson, Charles R. (1994). *Topics in matrix analysis*. New York, NY: Cambridge University Press.

Horn, Roger A., & Johnson, Charles R. (2013). *Matrix analysis*. New York, NY: Cambridge University Press.

Jennrich, Robert I. (1969). Asymptotic properties of non-linear least squares estimators. *The Annals of Mathematical Statistics*, 40(2), 633–643.

Ji, Xuewu, Yang, Kaiming, Na, Xiaoxiang, Lv, Chen, Liu, Yulong, & Liu, Yahui (2019). Feedback game-based shared control scheme design for emergency collision avoidance: a fuzzy-linear quadratic regulator approach. *Journal of Dynamic Systems, Measurement, and Control*, 141(8).

Jin, Wanxin, Kulić, Dana, Lin, Jonathan Feng-Shun, Mou, Shaoshuai, & Hirche, Sandra (2019). Inverse optimal control for multiphase cost functions. *IEEE Transactions on Robotics*, 35(6), 1387–1398.

Kallenberg, Olav (1997). *Foundations of modern probability*. Springer.

Kalman, Rudolf E. (1964). When is a linear control system optimal? *Journal of Basic Engineering*, 86(1), 51–60.

Keshavarz, Arezou, Wang, Yang, & Boyd, Stephen (2011). Imputing a convex objective function. In *2011 IEEE international symposium on intelligent control* (pp. 613–619). IEEE.

Köpf, Florian, Inga, Jairo, Rothfuß, Simon, Flad, Michael, & Hohmann, Sören (2017). Inverse reinforcement learning for identification in linear-quadratic dynamic games. *IFAC-PapersOnLine*, 50(1), 14902–14908.

Li, Yibei, Yao, Yu, & Hu, Xiaoming (2020). Continuous-time inverse quadratic optimal control problem. *Automatica*, 117, Article 108977.

Li, Yibei, Zhang, Han, Yao, Yu, & Hu, Xiaoming (2018). A convex optimization approach to inverse optimal control. In *2018 37th Chinese control conference (CCC)* (pp. 257–262). IEEE.

Ljung, Lennart (1999). *System identification (2nd Ed.): theory for the user*. USA: Prentice Hall PTR.

Ljung, Lennart, & Chen, Tianshi (2013). Convexity issues in system identification. In *2013 10th IEEE international conference on control and automation (ICCA)* (pp. 1–9).

Löfberg, Johan (2004). YALMIP: A toolbox for modeling and optimization in MATLAB. In *2004 IEEE international conference on robotics and automation (IEEE cat. no. 04CH37508)* (pp. 284–289). IEEE.

Menner, Marcel, Worsnop, Peter, & Zeilinger, Melanie N. (2021). Constrained inverse optimal control with application to a human manipulation task. *IEEE Transactions on Control Systems Technology*, 29(2), 826–834.

Molloy, Timothy L., Ford, Jason J., & Perez, Tristan (2018). Finite-horizon inverse optimal control for discrete-time nonlinear systems. *Automatica*, 87, 442–446.

Molloy, Timothy L., Ford, Jason J., & Perez, Tristan (2020). Online inverse optimal control for control-constrained discrete-time systems on finite and infinite horizons. *Automatica*, 120, Article 109109.

Mombaur, Katja, Truong, Anh, & Laumond, Jean-Paul (2010). From human to humanoid locomotion – an inverse optimal control approach. *Autonomous Robots*, 28(3), 369–383.

MOSEK ApS (2019). The MOSEK optimization toolbox for MATLAB manual. Version 9.0.

Pauwels, Edouard, Henrion, Didier, & Lasserre, Jean-Bernard (2016). Linear conic optimization for inverse optimal control. *SIAM Journal on Control and Optimization*, 54(3), 1798–1825.

Priess, M. Cody, Conway, Richard, Choi, Jongeun, Popovich, John M., & Radcliffe, Clark (2014). Solutions to the inverse LQR problem with application to biological systems analysis. *IEEE Transactions on Control Systems Technology*, 23(2), 770–777.

Rouot, Jérémy, & Lasserre, Jean-Bernard (2017). On inverse optimal control via polynomial optimization. In *2017 IEEE 56th annual conference on decision and control (CDC)* (pp. 721–726). IEEE.

Toumi, Noureddine, Malhamé, Roland, & Le Ny, Jerome (2020). A tractable mean field game model for the analysis of crowd evacuation dynamics. In *2020 59th IEEE conference on decision and control (CDC)* (pp. 1020–1025). IEEE.

Tsiantis, Nikolaos, Balsa-Canto, Eva, & Banga, Julio R. (2018). Optimality and identification of dynamic models in systems biology: an inverse optimal control framework. *Bioinformatics*, 34(14), 2433–2440.

van der Vaart, Adrianus W. (1998). *Asymptotic statistics*. Cambridge, United Kingdom: Cambridge University Press.

Westermann, Kevin, Lin, Jonathan Feng-Shun, & Kulić, Dana (2020). Inverse optimal control with time-varying objectives: application to human jumping movement analysis. *Scientific Reports*, 10(1), 1–15.

Yu, Chengpu, Li, Yao, Fang, Hao, & Chen, Jie (2021). System identification approach for inverse optimal control of finite-horizon linear quadratic regulators. *Automatica*, 129, Article 109636.

Zhang, Han, Li, Yibei, & Hu, Xiaoming (2019). Inverse optimal control for finite-horizon discrete-time linear quadratic regulator under noisy output. In *2019 IEEE 58th conference on decision and control (CDC)* (pp. 6663–6668). IEEE.

Zhang, Han, Umenberger, Jack, & Hu, Xiaoming (2019). Inverse optimal control for discrete-time finite-horizon linear quadratic regulators. *Automatica*, 110, Article 108593.
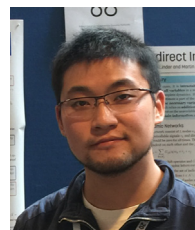
Han Zhang received his Ph.D. from Department of Mathematics, KTH Royal Institute of Technology, Sweden in 2019. He obtained both this B.S. and M.S. degree from Deptartment of Automation, Shanghai Jiao Tong University in 2011 and 2014 respectively. After spending two years in industry, he is now an assistant professor in Department of Automation, Shanghai Jiao Tong University. His main research interests are inverse optimal control, game theory and their application in robotics.

Axel Ringh received a M.Sc. degree in Engineering Physics in 2014, and a Ph.D. degree in Applied and Computational Mathematics in 2019, both from KTH Royal Institute of Technology, Stockholm. From 2019 to 2021 he was a postdoctoral researcher with the Department of Electronic and Computer Engineering, The Hong Kong University of Science and Technology, Hong Kong, and since 2021 he is assistant professor at the Department of Mathematical Sciences, Chalmers University of Technology and University of Gothenburg, Gothenburg. His current research is in the intersection of applied mathematics and areas such as control theory, signal processing, inverse problems, and machine learning.