Discrete-time inverse linear quadratic optimal control over finite time-horizon under noisy output measurements

Han Zhang^{1,2,3} · Yibei Li⁴ · Xiaoming Hu⁴

Received: 31 May 2021 / Revised: 9 July 2021 / Accepted: 12 July 2021 / Published online: 15 November 2021
© The Author(s) 2021

Abstract

In this paper, the problem of inverse quadratic optimal control over finite time-horizon for discrete-time linear systems is considered. Our goal is to recover the corresponding quadratic objective function using noisy observations. First, the identifiability of the model structure for the inverse optimal control problem is analyzed under relative degree assumption and we show the model structure is strictly globally identifiable. Next, we study the inverse optimal control problem whose initial state distribution and the observation noise distribution are unknown, yet the exact observations on the initial states are available. We formulate the problem as a risk minimization problem and approximate the problem using empirical average. It is further shown that the solution to the approximated problem is statistically consistent under the assumption of relative degrees. We then study the case where the exact observations on the initial states are not available, yet the observation noises are known to be white Gaussian distributed and the distribution of the initial state is also Gaussian (with unknown mean and covariance). EM-algorithm is used to estimate the parameters in the objective function. The effectiveness of our results are demonstrated by numerical examples.

Keywords Inverse optimal control · Linear quadratic regulator · Statistical consistency · EM-algorithm

1 Introduction

Since first proposed by [1], there has been a numerous applications of inverse optimal control (IOC) [2–4]. For example, IOC has been widely developed as a powerful tool to help us understand the optimality criteria underlying biological motions, which could then used for control synthesis of

bionic robots. The goal of “forward” optimal control problem is to find the optimal control input and the corresponding state trajectory given the cost function, system dynamics and the initial value of the states. In contrast, for a given system dynamics, the IOC problem focuses on recovering the undetermined parameters in the cost function that corresponds to the observed system states or output measurements (possibly noisy).

In this paper, the problem of IOC for discrete-time linear quadratic regulators (LQRs) over finite-time horizon is considered. As an extension of our early work [5], the identifiability of the corresponding model structure is also fully studied. Specifically speaking, our goal is to reconstruct the unknown parameters in the quadratic objective function using the given discrete-time linear system dynamics and its noisy measurement of output.

The inverse linear quadratic optimal control problem has been widely studied during the past decades, particularly in the continuous infinite time-horizon case such as [6–8]. In most work, it is assumed that the optimal feedback gain K is measured exactly, which is a constant matrix in the infinite time-horizon. A classical method [9] is to formulate the problem into an optimization problem with linear matrix

✉ Xiaoming Hu
hu@kth.se

Han Zhang
zhanghan_tc@sjtu.edu.cn

Yibei Li
yibei@kth.se

¹ Department of Automation, Shanghai Jiao Tong University, Shanghai 200240, China

² Key Laboratory of System Control and Information Processing, Ministry of Education of China, Shanghai 200240, China

³ Shanghai Engineering Research Center of Intelligent Control and Management, Shanghai 200240, China

⁴ Department of Mathematics, KTH Royal Institute of Technology, 100 44 Stockholm, Sweden

inequalities (LMI) constraints and solve for the optimal Q and R when the feedback gain K is known. Nevertheless, in many scenarios, the feedback control gain K cannot be known exactly and we could only measure the optimal state sequence or the corresponding system output. In [10], given the noisy state observations, the discrete infinite time-horizon case is studied in which the optimal feedback gain K is time-invariant. Their approach is to first identify the feedback matrix K from noisy observations by a maximum-likelihood estimator. Then matrices Q and R could be solved in a similar way as the method proposed in [9].

However, note that in the finite-time horizon case, the feedback gain K_t is time-varying, since “identifying time-varying system” is in general not an easy task, such “two steps” approach, i.e., “first identify the feedback gain using classical system identification methods, then search for the corresponding parameters in the quadratic objective function” cannot be directly applied. Furthermore, one critical property which is ignored for the “first identification step” is that the K_t ’s are actually generated by an LQR. Hence, it is difficult to guarantee the performance of such method in the finite-time horizon case and the optimality behind the observations could have been utilized to further improve the performance.

Since linear quadratic regulators can be seen as a special case of nonlinear optimal control problems, hence more generally speaking, the problem of IOC for quadratic regulators also lies in the scope of IOC problems for nonlinear systems. Usually, the considered objective function is composed by a weighted sum of base functions, where various parameter identification methods or learning techniques are applied. For example, in [11, 12], the optimality conditions for the forward optimal control problem is analyzed and the undetermined parameters in the cost function are estimated by minimizing the violation of these conditions. Nevertheless, [13] showed that such methods are not statistically consistent and sensitive to observation noise. Instead, a statistically consistent formulation is proposed in [13], where the resulting optimization problem is difficult to solve. The discrete finite time case is also considered in [14, 15] based on the Pontryagin’s Maximum Principle (PMP) for the optimal control problem. An optimization problem is then formulated whose constraints are two of the three conditions of PMP, and the residual of the third PMP condition is minimized. It is assumed that the optimal control input signal is known exactly therein while in our case, we only assume that the initial values and noisy system output is known. Moreover, though the authors claim that the problem of identifiability is considered, the “identifiability” considered therein does not coincide with the traditional definition of identifiability. In addition, the statistical consistency of the estimation is not

claimed therein, hence the performance under noisy observations can not be guaranteed. In [16], the authors consider the discrete-time IOC for nonlinear systems when some segments of the trajectories and input observations are missing. However, their analysis for the identifiability does not accord with the other definition of identifiability and statistical consistency cannot be guaranteed.

The aforementioned problems of estimating the parameters in a nonlinear optimal control problem could indeed include inverse LQ optimal control as a special case. However, taking advantage of the special structure of LQR, we can further prove the identifiability for the model structure in Sect. 3 and the statistical consistency for the estimation in Sect. 4 as well. Furthermore, under Gaussian assumptions, we use EM-algorithm to obtain a maximum likelihood estimation of the objective function as shown in Sect. 5 due to its special structure. The performance of the algorithms are shown by numerical examples in Sect. 6.

Notations We denote n dimensional positive semi-definite matrices cone as $\mathbb{S}_+^n$ while $\mathbb{S}_{++}^n$ is the set of strictly positive semi-definite matrices. $\|\cdot\|$ denotes the l_2 norm of a vector. We denotes the Frobenius norm of a matrix as $\|\cdot\|_F$. $\text{tr}(\cdot)$ denotes the trace of a matrix and $\text{tr}(H_1^T H_1) = \|H_1\|_F^2$. $\otimes$ denotes Kronecker product and $[H]_i$ denotes the i th row of matrix H .

2 Problem formulation

The “forward” optimal LQ problem reads

$$\min_{x_{2:N}, u_{1:N-1}} J = x_N^T S x_N + \sum_{t=1}^{N-1} (u_t^T R u_t + x_t^T Q x_t) \quad (1)$$

$$\text{s.t. } x_{t+1} = A x_t + B u_t, \quad x_1 = \bar{x}, \quad (2)$$

where $S, Q \in \mathbb{S}_+^n$, $R \in \mathbb{S}_{++}^m$, $x_t \in \mathbb{R}^n$ and $u_t \in \mathbb{R}^m$.

Suppose the noisy output $y_t = C x_t + v_t$ is available, where $v_t \in \mathbb{R}^l$ denotes the noise term, $y_t \in \mathbb{R}^l$ and the matrices (A, B, C) are known. For the sake of simplicity, we only consider the case which $R = I$ and $S = 0$ in this paper.

The IOC problem considered in this paper aims to find Q in the objective function using:

1. the exact samples of initial value $x_1 = \bar{x}$ and the noisy output $y_{2:N}$, or
2. the noisy output $y_{1:N}$ under Gaussian assumption.

We further assume (A, B) is controllable and B has full column rank. Moreover, since a discrete-time system is usually

sampled from a continuous linear system $\dot{x} = \hat{A}x + \hat{B}u$, this means that $A = e^{\hat{A}\Delta t}$, $B = \int_0^{\Delta t} e^{\hat{A}\tau} \hat{B} d\tau$ and by the fact of matrix exponential being invertible, it is reasonable to further suppose A being invertible.

3 Model identifiability

Before moving on, we would like to investigate the fundamental question: given the observation data, can we uniquely reconstruct the parameter Q in the objective function? This question actually involves two aspects: identifiability of the model structure as well as persistent excitation. Identifiability is a property of the model structure which determines if the model structure can generate the same output using two different parameters while persistent excitation of the input signal ensures that we can actually recover the undetermined parameters in the identifiable model by the measured corresponding output. Hence, to do further analysis, we need to define the model structure for the IOC problem first.

Model structure is nothing but a parameterized candidate model that describes the input–output relationship [17]. In the IOC problem we just formulated, we regard the initial value $\bar{x}$ as the input signal and the system output y_t as the output signal. Therefore, the model structure $\mathcal{M}_{2:N-1}(Q)$ considered in this paper is defined as the following:

Definition 1 (Model structure) For $\mathcal{M}_t : \mathbb{R}^n \mapsto \mathbb{R}^l, t = 2, \dots, N-1$, the model structure $\mathcal{M}_{2:N-1}(\cdot)$ takes the form:

$$\mathcal{M}_t(Q) = C \prod_{k=1}^{t-1} A_{cl}(k; Q), \quad (3)$$

where $A_{cl}(k; Q)$ is the closed-loop system matrix generated by the “forward problem” (1) and (2).

We further make the following assumption for the system:

Assumption 1 The discrete-time linear system defined by (A, B, C) is a square system, i.e., $l = m$ and has relative degree $(r_1, \dots, r_m)$. Namely,

$$c_i A^j B = 0, \quad j = 0 : r_i - 2, \quad c_i A^{r_i-1} B \neq 0,$$

$$\mathcal{L} = \begin{bmatrix} c_1 A^{r_1-1} B \\ \vdots \\ c_m A^{r_m-1} B \end{bmatrix} \text{ is nonsingular,}$$

where c_i denotes the i th row of C .

We first look into more details of the optimal control sequences.

Theorem 1 Suppose that $N \geq n+1$ and (A, B) is controllable. If $Q \neq Q', Q, Q' \in \mathbb{S}_+^n$, then $K_t(Q) = K_t(Q')$ cannot happen consecutively for more than $n-1$ steps.

Proof Recall that the optimal solution to (2) is derived by the discrete-time Riccati Equation (DRE)

$$\begin{aligned} P_t &= A^T P_{t+1} A + Q - A^T P_{t+1} B (B^T P_{t+1} B \\ &\quad + I)^{-1} B^T P_{t+1} A, \quad t = 1, \dots, N-1, \\ P_N &= S = 0, \end{aligned}$$

whose solution sequence is denoted by $DRE(Q, S)$. We then prove the theorem by contradiction. Let $P_t = DRE(Q, 0)$, $P'_t = DRE(Q', 0)$, $\Delta Q = Q - Q'$ and $\Delta P_t = P_t - P'_t$ for $t = 2 : N$. Assume there exists $t_1 \in [1, N-1]$ such that $K_t(Q) = K_t(Q')$ hold for $t \in [t_1 - n + 1, t_1] \subset [1, N-1]$. Then it holds that

$$\begin{cases} B^T \Delta P_{t+1} = 0, \\ A^T \Delta P_{t+1} A - \Delta P_t + \Delta Q = 0, \end{cases} \quad (4)$$

where $t = t_1 - n + 1 : t_1$.

Using (4) and the fact that A is invertible, $B^T \Delta P_{t+1} = 0$ could be computed recursively for $t = t_1 - n + 1 : t_1$ and stacked into the following compact form:

$$\underbrace{\begin{bmatrix} B^T \\ B^T A^T \\ \vdots \\ B^T (A^T)^{(n-1)} \end{bmatrix}}_{\Gamma} \Delta P_{t_1+1} = - \underbrace{\begin{bmatrix} 0 \\ B^T \Delta Q A^{-1} \\ \vdots \\ B^T \left(\sum_{k=0}^{n-2} (A^T)^k \Delta Q A^k \right) (A^{-1})^{(n-1)} \end{bmatrix}}_{\Omega}.$$

Since (A, B) is controllable, Γ has full column rank. Hence, ΔP_{t_1+1} is the unique solution of the linear equation $\Gamma \mathcal{P} = \Omega$.

On the other hand, consider the following algebraic equation:

$$\begin{cases} B^T \Delta P = 0, \\ A^T \Delta P A - \Delta P + \Delta Q = 0. \end{cases} \quad (5)$$

We can apply the same techniques to (5) and get the same equation $\Gamma \mathcal{P} = \Omega$ while ΔP is a solution to it. Since $\Gamma \mathcal{P} = \Omega$ has a unique solution, hence it follows that $\Delta P_{t_1+1} = \Delta P$.

Moreover, recall that it holds for any $t_1 - n + 1 \leq t \leq t_1$ that $A^T \Delta P_{t+1} A - \Delta Q = \Delta P_t$ and $A^T \Delta P A + \Delta Q = \Delta P$, thus we have $\Delta P_t = \Delta P, \forall t = t_1 - n + 2 : t_1 + 1$.

Using (5) and $P'_t = DRE(Q', 0)$, we can then show that $\bar{P}_t \triangleq P'_t + \Delta P$ is in fact the solution to $DRE(Q, \Delta P)$ for all $t \in [1, N - 1]$, i.e.,

$$\begin{aligned} \bar{P}_t &= P'_t + \Delta P \\ &= A^T P'_{t+1} A + Q' - A^T P'_{t+1} B (B^T P'_{t+1} B + I)^{-1} B^T P'_{t+1} A \\ &\quad + A^T \Delta P A + \Delta Q \\ &= A^T (P'_{t+1} + \Delta P) A + Q - A^T (P'_{t+1} \\ &\quad + \Delta P) B (B^T (P'_{t+1} + \Delta P) B + I)^{-1} B^T (P'_{t+1} + \Delta P) A, \end{aligned}$$

where $\bar{P}_N = P'_N + \Delta P = \Delta P$.

Recall that on the time interval $t = t_1 - n + 2 : t_1 + 1$, P_t and $\bar{P}_t$ satisfy the same DRE with weighting matrix Q and boundary $P_{t_1+1} = \bar{P}_{t_1+1} = P'_{t_1+1} + \Delta P$. Then based on the backwards recursion of DRE, it is obvious that $P_t = \bar{P}_t$ for all $1 \leq t \leq t_1 + 1$. It means that P_t and $\bar{P}_t$ are generated by the same forward Riccati equation with initial condition P_1 . We further show uniqueness of solution for the forward DRE for $t \in [1, N - 1]$ by considering its adjoint system and using the fact that A is nonsingular

$$\begin{cases} X_{t+1} = AX_t - BB^T Y_{t+1}, & X_1 = I, \\ Y_{t+1} = -(A^T)^{-1} Q X_t + (A^T)^{-1} Y_t, & Y_1 = P_1, \end{cases} \quad (6)$$

whose solution can be uniquely determined for all $t \geq 1$. Recall that $P_t = DRE(Q, 0)$ and let $\{X_t\}$ be the regular matrix solution of its closed-loop system $X_{t+1} = (A + BK_t)X_t$ with $X_1 = I$. Since $A + BK_t$ is always nonsingular [18], the nonsingular X_t together with $Y_t = P_t X_t$ compose the unique solution to (6). Therefore, for the given P_1 , there exists a unique solution $Y_t X_t^{-1}$ that satisfies the forward DRE for $t = 1 : N$, which contradicts with the fact $\bar{P}_N = \Delta P \neq P_N$. Hence the claim follows. $\square$

Now, we are ready to present the theorem for the identifiability of the model structure $\mathcal{M}_{2:N-1}(Q)$.

Theorem 2 Under Assumption 1, suppose that (A, B) is controllable, $N \geq 2n$, A is invertible, B has full column rank, then the model structure (3) is strictly globally identifiable.

Proof We prove the statement by contradiction. Suppose two different weighting matrices $Q_1 \neq Q_2$ could generate the same model structure, namely, $\mathcal{M}_{2:N-1}(Q_1) = \mathcal{M}_{2:N-1}(Q_2)$. For the model structure $\mathcal{M}_{t,i}(Q)$ that corresponds to the i th system output at time instant t . By Theorem 1, assume $t' \leq n - 1$ is the first time instant such that $K_{t'}(Q_1) \neq K_{t'}(Q_2)$. Therefore, we have

$$\begin{aligned} \mathcal{M}_{t,i}(Q) &= c_i \prod_{j=t'}^{t-1} (A + BK_j(Q)) \prod_{j=1}^{t'-1} A_{cl}(j; Q) \\ &= c_i A^{t-t'} \prod_{j=1}^{t'-1} A_{cl}(j; Q), \quad t = 1 : r_i + t' - 1, \quad i = 1 : m, \\ \mathcal{M}_{r_i+t',i}(Q) &= c_i [A^{r_i} + A^{r_i-1} BK_{t'}(Q)] \prod_{j=1}^{t'-1} A_{cl}(j; Q), \quad i = 1 : m. \end{aligned}$$

Hence, if it holds for the model structures $\mathcal{M}_t(Q_1) = \mathcal{M}_t(Q_2)$ at all $t = 1 : N - 1$, then it must hold that

$$\begin{aligned} \mathcal{M}_{r_i+t',i}(Q_1) &= \mathcal{M}_{r_i+t',i}(Q_2), \quad i = 1 : m \\ &\Leftrightarrow c_i [A^{r_i} + A^{r_i-1} BK_{t'}(Q)] \prod_{j=1}^{t'-1} A_{cl}(j; Q_1) \\ &= c_i [A^{r_i} + A^{r_i-1} BK_{t'}(Q)] \prod_{j=1}^{t'-1} A_{cl}(j; Q_2), \quad i = 1 : m. \end{aligned}$$

Using the fact that $A_{cl}(t; Q)$ is invertible for all $t = 1 : N - 1$ [5], we can cancel the product of closed-loop system matrices in the above equation and get

$$\begin{aligned} c_i A^{r_i-1} BK_{t'}(Q) &= c_i A^{r_i-1} BK_{t'}(Q), \quad i = 1 : m \\ &\Leftrightarrow \mathcal{L}K_{t'}(Q_1) = \mathcal{L}K_{t'}(Q_2). \end{aligned}$$

Since $\mathcal{L}$ is invertible by definition, it holds that $K_{t'}(Q_1) = K_{t'}(Q_2)$ and we have a contradiction. Hence the model structure is globally identifiable at Q_1 . Since Q_1 is arbitrarily chosen, this means that the model structure is strictly globally identifiable. $\square$

Remark 1 In the above theorem, it is easy to see that it is actually sufficient for the time-horizon length to have $N \geq n + \max_i(r_i)$ to make the model structure strictly globally identifiable. However, we state the theorem as it is just to make the theorem easier to comprehend for the reader. On the other hand, for the case $l > m$ (the output dimension is greater than the input dimension), we just need to assume that there is a subset of outputs that makes the system have relative degree $(r_1, \dots, r_m)$.

Here, we would like to remark that in our problem formulation observability of the system is not assumed. This is due to the fact that lack of observability can happen in various situations such as sensor unavailability, data shared with privacy, uncertain environment or tasks with a limited number of descriptive features, and it is not necessary to impose reconstructability of the state from noisy observations. On the other hand, for the identifiability of the IOC problem, it is required that the influence of different cost functions should be reflected sufficiently in the observation sequences. More specifically, from the proof of Theorem 2, we can see that the existence of relative degree plays a key role in guaranteeing

the distinguishability of the system output under different parameter specifications, namely, $\bar{y}_t(\bar{Q}, \bar{x}) = \bar{y}_t(Q, \bar{x})$ with $t = 2, \dots, N$ if and only if $Q = \bar{Q}$. It means that under such assumption, the LQR problems would generate exactly the same output sequences in the noiseless case only when the same weighting matrix Q is used. However, such property can not always be guaranteed for systems without a relative degree, which is illustrated by the following example.

Example 1 Consider the following controllable square system

$$A = \begin{bmatrix} 3 & 0 & 0 \\ 1 & 3 & -1 \\ 2 & 1 & 1 \end{bmatrix}, B = \begin{bmatrix} 1 & 0 \\ 0 & 0 \\ 1 & 5 \end{bmatrix}, C = \begin{bmatrix} c_1 \\ c_2 \end{bmatrix} = \begin{bmatrix} 0 & 1 & 0 \\ -1 & -1 & 1 \end{bmatrix},$$

which, however, does not have a relative degree.

For the quadratic cost function, we take $N = 5$. We could show that the following weighting matrices would generate the same output sequences $\bar{y}_t$,

$$Q_1 = \begin{bmatrix} 3 & 1 & -2 \\ 1 & 2 & -1 \\ -2 & -1 & 2 \end{bmatrix}, Q_2 = \begin{bmatrix} 4 & 1 & -2 \\ 1 & 2 & -1 \\ -2 & -1 & 2 \end{bmatrix}.$$

First, we check the first output. Since $c_1 B = 0$, $c_1 A B \neq 0$, and $\bar{y}_{t+1,i}(Q, \bar{x}) = c_i \Pi_{j=1}^t (A + B K_j) \bar{x}$, straightforward computation gives

$$\begin{aligned} \bar{y}_{2,1}(Q, \bar{x}) &= c_1 A \bar{x}, \\ \bar{y}_{3,1}(Q, \bar{x}) &= c_1 A^2 \bar{x} + c_1 A B K_1 \bar{x}, \\ \bar{y}_{4,1}(Q, \bar{x}) &= c_1 A^3 \bar{x} + c_1 A^2 B K_1 \bar{x} + c_1 A B K_2 B K_1 \bar{x}, \\ \bar{y}_{5,1}(Q, \bar{x}) &= c_1 A^4 \bar{x} + c_1 A^3 B K_1 \bar{x} + c_1 A^2 B K_2 B K_1 \bar{x} \\ &\quad + c_1 A B K_3 B K_2 B K_1 \bar{x}. \end{aligned}$$

By computing K_t from DRE, one could easily show that $\bar{y}_{t,1}(\bar{Q}, \bar{x}) = \bar{y}_{t,1}(Q, \bar{x})$ for any $\bar{x}$ and $t = 1, \dots, N$ since all the coefficient matrices of $\bar{x}$ in the right hand side of the above equations are equal for $K_t(Q)$ and $K_t(\bar{Q})$. The same claim also holds for $\bar{y}_{t,2}$ and the statement follows. $\square$

4 IOC using exact initial values and noisy output

We first consider the IOC problem in which exact samples of the initial value $x_1 = \bar{x}$ and noisy output $y_{2:N}$ are available. Suppose that $\bar{x} \in \mathbb{R}^n$, $\{v_t \in \mathbb{R}^n\}_{t=2}^N$ are independent random vectors with unknown distributions. We make further the following assumptions:

Assumption 2 It holds that $N \geq 2n$, $l \geq m$, and there exists a subset of outputs that makes the system have relative degree $(r_1, \dots, r_m)$.

Assumption 3 (Persistent excitation) $\forall \eta \in \mathbb{R}^n$, $\exists r(\eta) \in \mathbb{R} \setminus \{0\}$ such that $\mathbb{P}(\bar{x} \in \mathcal{B}_\varepsilon(r\eta)) > 0, \forall \varepsilon > 0$, where $\mathcal{B}_\varepsilon(r\eta)$ is the open ε -ball centered at $r\eta$.

Assumption 4 The “real” $\bar{Q}$ belongs to a compact set $\bar{\mathbb{S}}_+^n(\varphi) = \{Q \in \mathbb{S}_+^n \mid \|Q\|_F^2 \leq \varphi\}$.

Assumption 5 $\mathbb{E}(\|\bar{x}\|^2) < +\infty, \mathbb{E}(v_t) = 0$ and $\mathbb{E}(\|v_t\|^2) < +\infty$.

With the assumptions above, the forward LQR problem (1) and (2) can actually be expressed as

$$\min_{u_{1:N-1}(\omega), x_{2:N}(\omega)} \{J(u_{1:N-1}(\omega), x_{2:N}(\omega); Q; \bar{x}(\omega)) \mid (2), \omega \in \Omega\}, \quad (7)$$

where Ω is the sample space. Hence the optimal trajectory $\{x_t^*\}$ as well as the observed system output $\{y_t^* = Cx_t^* + v_t\}$ are random vectors implicitly determined by the random variable $\bar{x}$ and the parameter Q . Based on (7) we can define a risk function as

$$\mathcal{R}(Q) = \mathbb{E}_\xi[f(Q; \xi)], \quad (8)$$

where $\xi = [\bar{x}^T, Y^T]^T$, $Y = [y_2^T, \dots, y_N^T]^T$ and $f: \bar{\mathbb{S}}_+^n(\varphi) \times \mathbb{R}^{n+(N-1)l} \mapsto \mathbb{R}$,

$$f(Q; \xi) = \sum_{t=2}^N \|y_t - Cx_t^*(Q; \bar{x})\|^2, \quad (9)$$

and $x_{2:N}^*(Q; \bar{x})$ is the optimal trajectory to (7).

We minimize the risk function to obtain an estimate of the Q matrix, namely,

$$\min_{Q \in \bar{\mathbb{S}}_+^n(\varphi)} \mathcal{R}(Q). \quad (10)$$

The main problem now is that whether the risk minimization problem has a unique solution. For a given closed-loop system matrix $A + BK_t$ of a finite-horizon LQR problem, the uniqueness of the corresponding quadratic cost function has been studied in discrete systems [18] and continuous systems [19], respectively. However, additional difficulties would rise when we could only measure the system output. The following lemma shows the uniqueness of the solution to (10) under the assumption of relative degrees.

Lemma 1 Under Assumptions 2–5, the risk minimization problem (10) has a unique optimizer $\bar{Q}$, where $\bar{Q}$ is the true value used in the forward problem (7).

Proof Since $y_t = Cx_t^*(\bar{Q}, \bar{x}) + v_t$, $\mathbb{E}(v_t) = 0$, $t = 2 : N$, (8) can be simplified as $\mathcal{R}(Q) = L(Q) + \sum_{t=2}^N \mathbb{E}(\|v_t\|^2)$, where $L(Q) = \mathbb{E}(\sum_{t=2}^N \|\bar{y}_t(\bar{Q}, \bar{x}) - \bar{y}_t(Q, \bar{x})\|^2)$, $\bar{y}_t(\bar{Q}, \bar{x}) = Cx_t^*(\bar{Q}, \bar{x})$ and $\bar{y}_t(Q, \bar{x}) = Cx_t^*(Q, \bar{x})$. It is clear that $Q = \bar{Q}$ minimizes the risk function $\mathcal{R}(Q)$. What remains to show is the uniqueness.

Recall that by our assumption, there exists a subset of outputs that makes the system have relative degree $(r_1, \dots, r_m)$. On the other hand, it holds that $\bar{y}_t(Q, \bar{x}) = \mathcal{M}_t(Q)\bar{x}$. Note that

$$\begin{aligned} L(Q) &= \sum_{t=2}^N \mathbb{E}[\|\bar{y}_t(\bar{Q}, \bar{x}) - \bar{y}_t(Q, \bar{x})\|^2] \\ &= \sum_{t=2}^N \mathbb{E}[\|(\mathcal{M}_t(\bar{Q}) - \mathcal{M}_t(Q))\bar{x}\|^2]. \end{aligned}$$

By Theorem 2, we know that given $N \geq 2n$, if $Q \neq \bar{Q}$, then there exists some time instants t_i s such that $\mathcal{M}_{t_i}(Q) \neq \mathcal{M}_{t_i}(\bar{Q})$. On the other hand, by Assumption 3, $\exists r(\eta) \neq 0$, such that $\mathbb{P}(\bar{x} \in \mathcal{B}_\varepsilon(r\eta)) > 0, \forall \varepsilon$. Since $r \neq 0$, it holds that $\mathcal{M}_{t_i}(\bar{Q})(r\eta) \neq \mathcal{M}_{t_i}(Q)(r\eta)$. Furthermore, since $\mathcal{M}_t(Q)\eta$ is continuous with respect to η , $\forall Q \in \mathbb{S}_+^n$, this implies $\exists \varepsilon_1$, such that $\mathcal{M}_{t_i}(Q)\bar{x} \neq \mathcal{M}_{t_i}(\bar{Q})\bar{x}, \forall \bar{x} \in \mathcal{B}_{\varepsilon_1}(r\eta)$ and $\mathbb{P}(\bar{x} \in \mathcal{B}_{\varepsilon_1}(r\eta)) > 0$. Thus,

$$L(Q) \geq \int_{\mathcal{B}_{\varepsilon_1}(r\eta)} \|(\mathcal{M}_{t_i}(Q) - \mathcal{M}_{t_i}(\bar{Q}))\bar{x}(\omega)\|^2 \mathbb{P} d\omega > 0.$$

Hence, $\bar{Q}$ is the unique minimizer to (8) and the statement follows. $\square$

The risk minimization problem is expected to be solved numerically. To tackle the issue that the distributions of $\bar{x}$ and $v_{2:N}$ are unknown, we approximate (8) using the empirical mean

$$\mathcal{R}_M(Q) = \frac{1}{M} \sum_{i=1}^M f(Q; \xi^{(i)}), \quad (11)$$

where $\xi^{(i)}$ are i.i.d.

We consider first the optimal control problem (7). By Pontryagin's Maximum Principle (PMP), if $u_{1:N-1}^*$ and $x_{1:N}^*$ are the optimal control and corresponding trajectory, then there exist adjoint variables $\lambda_{2:N}^*$ such that

$$\begin{cases} x_{t+1}^* = Ax_t^* + Bu_t^*, & t = 1 : N-1, \\ \lambda_t^* = A^T \lambda_{t+1}^* + Qx_t^*, & t = 2 : N-1, \\ \lambda_N^* = 0, \\ u_t = -B^T \lambda_{t+1}^*, & t = 1 : N-1. \end{cases} \quad (12)$$

Since what we consider is a forward LQR problem, the PMP conditions are also sufficient. Hence, we can express $x_{2:N}^*$ in (9) using (12). Thus, the approximated risk minimization problem reads

$$\begin{aligned} \min_{Q \in \mathbb{S}_+^n(\varphi), x_{2:N}^{(i)}, \lambda_{2:N}^{(i)}} \mathcal{R}_M^x(Q) &= \frac{1}{M} \sum_{i=1}^M f(Q; \xi^{(i)}) \\ \text{s.t. } x_{t+1}^{(i)} &= Ax_t^{(i)} - BB^T \lambda_{t+1}^{(i)}, & t = 2 : N-1, \\ \lambda_t^{(i)} &= A^T \lambda_{t+1}^{(i)} + Qx_t^{(i)}, & t = 2 : N-1, \\ x_2^{(i)} &= A\bar{x}^{(i)} - BB^T \lambda_2^{(i)}, \\ \lambda_N^{(i)} &= 0, & i = 1 : M. \end{aligned} \quad (13)$$

The superscript “star” is omitted for simplicity. Note that the optimizer $(Q_M^*(\omega), x_{2:N}^{(i)*}(\omega), \lambda_{2:N}^{(i)*}(\omega))$ is stochastic and is optimal in the sense that it optimizes (11) for every $\omega \in \Omega$.

Based on Lemma 1, we could then show the statistical consistency of the approximated risk minimization problem (13).

Theorem 3 Suppose $Q_M^* \in \mathbb{S}_+^n(\varphi)$, $\{x_{2:N}^{(i)*}\}$ and $\{\lambda_{2:N}^{(i)*}\}$ solve (13), then $Q_M^* \xrightarrow{p} \bar{Q}$, where $\bar{Q}$ is the true value used in the “forward” problem (7).

Proof Denote $z_t = (x_t^T, \lambda_t^T)^T, t = 2 : N$, then the first two equations in (12) can be written as the following implicit dynamics

$$\underbrace{\begin{bmatrix} I & BB^T \\ 0 & A^T \end{bmatrix}}_E z_{t+1} = \underbrace{\begin{bmatrix} A & 0 \\ -Q & I \end{bmatrix}}_F z_t, \quad t = 2 : N-1.$$

Therefore, we can rewrite (12) in the following compact matrix form as

$$\underbrace{\begin{bmatrix} \tilde{E} & \tilde{F} \\ -F & E \\ & \ddots \\ & -F & E \end{bmatrix}}_{\mathcal{F}(Q)} \underbrace{\begin{bmatrix} z_2 \\ \vdots \\ z_N \end{bmatrix}}_Z = \underbrace{\begin{bmatrix} A\bar{x}^{(i)} \\ 0 \\ \vdots \\ 0 \end{bmatrix}}_{b(\bar{x})}, \quad (14)$$

where

$$\tilde{E} = \begin{bmatrix} I & BB^T \\ 0 & 0 \end{bmatrix}, \quad \tilde{F} = \begin{bmatrix} 0 & 0 \\ 0 & I \end{bmatrix}.$$

Note that for an arbitrary $Q \in \mathbb{S}_+^n$, since (14) is a sufficient and necessary condition for the optimality of the corresponding forward LQR problem that in turn has a unique solution, $\mathcal{F}(Q)$ must be invertible for all $Q \in \mathbb{S}_+^n$. Thus, it follows that $Z = \mathcal{F}(Q)^{-1}b(\bar{x}) = \mathcal{F}(Q)^{-1}\tilde{A}\bar{x}$, where $\tilde{A} = [A^T, 0, \dots, 0]^T$. Hence $f(Q; \xi)$ can be rewritten as

$$f(Q; \xi) = \|Y - GZ\|^2 = \|Y - G\mathcal{F}(Q)^{-1}\tilde{A}\bar{x}\|^2,$$

where $G = I_{N-1} \otimes [C, 0_{l \times n}]$ and $Y = [y_2^T, \dots, y_N^T]^T$.

It can be shown in a similar way to [5] that the uniform law of large numbers [20] applies, namely,

$$\sup_{Q \in \mathbb{S}_+^n(\varphi)} \left\| \frac{1}{M} \sum_{i=1}^M f(Q, \xi^{(i)}) - \mathbb{E}_{\xi} [f(Q; \xi)] \right\| \xrightarrow{p} 0. \quad (15)$$

In addition to (15), by Lemma 1 we have also shown that $\bar{Q}$ is the unique optimizer to (10), then $Q_M^* \xrightarrow{p} \bar{Q}$ follows directly from Theorem 5.7 in [21]. $\square$

5 IOC under Gaussian assumption

In this section, we consider the case in which exact samples on the initial values are not available and we can only observe the noisy measurement $y_1 = Cx_1 + v_1$. Recall that in Sect. 3, we mentioned that we see the initial value as the “input signal” of the model structure while the system output observation as the “output signal”. This means that in this scenario, the problem is actually an *error-in-variable system identification problem*, which is a hard problem in general. To tackle this problem, we have the following assumption in the remainder of this section.

Assumption 6 $\bar{x} = x_1 \sim \mathcal{N}(m_1, \Sigma_1)$, $v_t \sim \mathcal{N}(0, \Sigma_v)$ and $v_{1:N}$ are white. Σ_v is known, but m_1 and Σ_1 are not a priori known.

Recall that in the proof of Theorem 3, we are able to write $Y = [y_2^T, \dots, y_N^T]^T = G\mathcal{F}(Q)^{-1}\tilde{A}\bar{x} + [v_2^T, \dots, v_N^T]^T$. Namely, it holds that

$$y_t = G_t\mathcal{F}(Q)^{-1}\tilde{A}\bar{x} + v_t, \quad t = 2 : N,$$

where $G_t = [G]_{(t-1)+1:l}$. We can regard $y_{1:N}$ as the output of the following system:

$$\begin{aligned} \eta_{t+1} &= \eta_t, \quad t = 1 : N-1, \quad \eta_1 = \bar{x}, \\ y_t &= \underbrace{G_t\mathcal{F}(Q)^{-1}\tilde{A}}_{\tilde{C}_t(Q)} \eta_t + v_t, \quad t = 1 : N, \end{aligned} \quad (16)$$

where $\tilde{C}_1(Q) = C$ and $\tilde{C}_t(Q) = G_t\mathcal{F}(Q)^{-1}\tilde{A}$, $t = 2 : N$. This means that the IOC problem can be viewed as a system identification problem for linear Gaussian model under Assumption 6, which can be solved using maximum log-likelihood method. If M sets of output sequences are available, it means we have M i.i.d samples $x_1^{(1:M)}$ of the initial value. Since $v_t^{(1:M)}$ is also i.i.d., $y_t^{(i)} = Cx_t^{(i)} + v_t^{(i)}$ and $x_t^{(i)}$ depends only on the initial value and Q , $y_t^{(i)}$ is independent of $y_j^{(j)}$, $\forall i \neq j$.

EM-algorithm will be used to get a maximum log-likelihood estimate for Q . Based on (16), η_1 is chosen as the latent variable and we use θ to parametrize Q , m_1 and Σ_1 . Now we are ready to compute the auxiliary function $\mathcal{Q}(\theta, \theta_k)$ for EM-algorithm.

Proposition 1 Under Assumption 6, the auxiliary function $\mathcal{Q}(\theta, \theta_k)$ that corresponds to the model (16) is given by

$$\mathcal{Q}(\theta, \theta_k) \propto \mathcal{Q}_1(\theta, \theta_k) + \mathcal{Q}_2(\theta, \theta_k),$$

where

$$\begin{aligned} \mathcal{Q}_1(\theta, \theta_k) &= - \sum_{i=1}^M \left\{ \ln \det \Sigma_1 + \text{tr}[\Sigma_1^{-1} P_{1|N}^{(i)}] \right. \\ &\quad \left. + \sum_{t=2}^N \text{tr}[\Sigma_1^{-1} (\hat{\eta}_{1|N}^{(i)} - m_1)(\hat{\eta}_{1|N}^{(i)} - m_1)^T] \right\}, \end{aligned} \quad (17)$$

$$\begin{aligned} \mathcal{Q}_2(\theta, \theta_k) &= - \sum_{i=1}^M \left\{ \sum_{t=2}^N \text{tr}[\Sigma_v^{-1} \tilde{C}_t(Q) P_{1|N}^{(i)} \tilde{C}_t(Q)^T] \right. \\ &\quad \left. + \sum_{t=2}^N \text{tr}[\Sigma_v^{-1} (y_t^{(i)} - \tilde{C}_t(Q) \hat{\eta}_{1|N}^{(i)})(y_t^{(i)} - \tilde{C}_t(Q) \hat{\eta}_{1|N}^{(i)})^T] \right\}, \end{aligned} \quad (18)$$

$$\text{and } \hat{\eta}_{1|N}^{(i)} := \mathbb{E}_{\theta_k}[\eta_1^{(i)} | y_{1:N}^{(i)}], P_{1|N}^{(i)} = \text{cov}_{\theta_k}[\eta_1^{(i)} | y_{1:N}^{(i)}].$$

Proof By Bayes' rule, it follows that

$$p_{\theta}(\eta_1^{(1:M)}, y_{1:N}^{(1:M)}) = \prod_{i=1}^M p_{\theta}(\eta_1^{(i)}) \prod_{t=1}^N p_{\theta}(y_t^{(i)} | \eta_1^{(i)}).$$

Hence by the definition of the auxiliary function $\mathcal{Q}(\theta, \theta_k)$, we have

$$\begin{aligned} \mathcal{Q}(\theta, \theta_k) &= \int \ln p_{\theta}(\eta_1^{(1:M)}, y_{1:N}^{(1:M)}) p_{\theta_k}(\eta_1^{(1:M)} | y_{1:N}^{(1:M)}) d\eta_1^{(1:M)} \\ &= \int \left(\sum_{i=1}^M \ln p_{\theta}(\eta_1^{(i)}, y_{1:N}^{(i)}) \right) \prod_{i=1}^M p_{\theta_k}(\eta_1^{(i)} | y_{1:N}^{(i)}) d\eta_1^{(1:M)} \\ &= \sum_{i=1}^M \int \ln p_{\theta}(\eta_1^{(i)}, y_{1:N}^{(i)}) p_{\theta_k}(\eta_1^{(i)} | y_{1:N}^{(i)}) d\eta_1^{(i)} \\ &= \sum_{i=1}^M \left\{ \int \ln p_{\theta}(\eta_1^{(i)}) p_{\theta_k}(\eta_1^{(i)} | y_{1:N}^{(i)}) d\eta_1^{(i)} \right. \\ &\quad \left. + \sum_{t=2}^N \int \ln p_{\theta}(y_t^{(i)} | \eta_1^{(i)}) p_{\theta_k}(\eta_1^{(i)} | y_{1:N}^{(i)}) d\eta_1^{(i)} \right\}. \end{aligned}$$

Note that $y_1^{(i)} = C\eta_1^{(i)} + v_1^{(i)}$, hence $p_{\theta}(y_1^{(i)} | \eta_1^{(i)}) \sim \mathcal{N}(C\eta_1^{(i)}, \Sigma_v)$ and it has nothing to do with θ . By using the cyclic permutation property of matrix trace and the fact that $\mathbb{E}_{\theta_k}[\eta_1^{(i)} \eta_1^{(i)T} | y_{1:N}^{(i)}] = \hat{\eta}_{1|N}^{(i)} \hat{\eta}_{1|N}^{(i)T} + P_{1|N}^{(i)}$, we have

$$\begin{aligned}
\mathcal{Q}(\theta, \theta_k) &\propto -\sum_{i=1}^M \left\{ \ln \det \Sigma_1 + \int \|\eta_1^{(i)}\|_{\Sigma_1^{-1}}^2 p_{\theta_k}(\eta_1^{(i)} | y_{1:N}^{(i)}) d\eta_1^{(i)} \right. \\
&\quad \left. + \sum_{i=2}^N \int \|y_i^{(i)} - \tilde{c}_i(Q) \eta_1^{(i)}\|_{\Sigma_v^{-1}}^2 p_{\theta_k}(\eta_1^{(i)} | y_{1:N}^{(i)}) d\eta_1^{(i)} \right\} \\
&= \mathcal{Q}_1(\theta, \theta_k) + \mathcal{Q}_2(\theta, \theta_k),
\end{aligned}$$

where the constant and irrelevant terms are ignored. $\square$

We can obtain the mean and covariance for the initial value $\hat{\eta}_1^{(i)}$ and $P_{1|N}^{(i)}$ by fixed point smoothing [22]. Each iteration of EM-algorithm involves maximizing $\mathcal{Q}(\theta, \theta_k)$ with respect to θ . Since m_1 and Σ_1 only appear in $\mathcal{Q}_1(\theta, \theta_k)$ while Q appears only in $\mathcal{Q}_2(\theta, \theta_k)$, they can be optimized separately. The first-order optimality condition for maximizing $\mathcal{Q}_1(\theta, \theta_k)$ with respect to m_1 and Σ_1 reads:

$$\begin{aligned}
\partial_{m_1} \mathcal{Q}_1(\theta^*, \theta_k) &= -\partial_{m_1} \left(\sum_{i=1}^M (m_1^* - \hat{\eta}_{1|N}^{(i)})^T \Sigma_1^{*-1} (m_1^* - \hat{\eta}_{1|N}^{(i)}) \right) \\
&= -2 \sum_{i=1}^M \Sigma_1^{*-1} (m_1^* - \hat{\eta}_{1|N}^{(i)}) = 0,
\end{aligned} \quad (19)$$

$$\begin{aligned}
\partial_{\Sigma_1} \mathcal{Q}_1(\theta^*, \theta_k) &= -\sum_{i=1}^M \left\{ \Sigma_1^{*-1} - \Sigma_1^{*-1} P_{1|N}^{(i)} \Sigma_1^{*-1} \right. \\
&\quad \left. - \Sigma_1^{*-1} (m_1^* - \hat{\eta}_{1|N}^{(i)}) (m_1^* - \hat{\eta}_{1|N}^{(i)})^T \Sigma_1^{*-1} \right\} = 0.
\end{aligned} \quad (20)$$

Solving (19) and (20), we get

$$m_1^* = \frac{1}{M} \sum_{i=1}^M \eta_{1|N}^{(i)}, \quad (21)$$

$$\begin{aligned}
\Sigma_1^{-1} \sum_{i=1}^M \left[\Sigma_1 - (m_1^* - \hat{\eta}_{1|N}^{(i)}) (m_1^* - \hat{\eta}_{1|N}^{(i)})^T - P_{1|N}^{(i)} \right] \Sigma_1^{-1} &= 0 \\
\Leftrightarrow \sum_{i=1}^M \left[\Sigma_1 - (m_1^* - \hat{\eta}_{1|N}^{(i)}) (m_1^* - \hat{\eta}_{1|N}^{(i)})^T - P_{1|N}^{(i)} \right] &= 0 \\
\Leftrightarrow \Sigma_1^* &= \frac{1}{M} \sum_{i=1}^M \left[(m_1^* - \hat{\eta}_{1|N}^{(i)}) (m_1^* - \hat{\eta}_{1|N}^{(i)})^T + P_{1|N}^{(i)} \right].
\end{aligned} \quad (22)$$

Since (21) and (22) give the unique maximizer of $\mathcal{Q}_1(\theta, \theta_k)$, it is also the global maximizer of $\mathcal{Q}_1(\theta, \theta_k)$. In contrast it is not possible to obtain the analytic solution for the optimizer of $\mathcal{Q}_2(\theta, \theta_k)$, thus we have to solve Q^* numerically. The problem of maximizing $\mathcal{Q}_2(\theta, \theta_k)$ can be rewritten as

$$\begin{aligned}
\min_{Q \in \mathbb{S}_+^n(\varphi)} \sum_{i=1}^M \left\{ \|Y^{(i)} - G \mathcal{F}(Q)^{-1} \tilde{A} \hat{\eta}_{1|N}^{(i)}\|_{\Sigma_v^{-1}}^2 \right. \\
\left. + \sum_{i=1}^N \text{tr} \left[\Sigma_v^{-1} G_i \mathcal{F}(Q)^{-1} \tilde{A} P_{1|N}^{(i)} \tilde{A}^T \mathcal{F}(Q)^{-T} G_i^T \right] \right\},
\end{aligned} \quad (23)$$

where $Y^{(i)} = [y_2^{(i)T}, \dots, y_N^{(i)T}]^T$, $\Sigma_v = I_{N-1} \otimes \Sigma_v$. Using the function $\hat{f}_\varepsilon$ proposed by [23] and apply the same “trick” mentioned in the end of Sect. 4, we can use standard non-linear optimization solvers to solve (23).

6 Numerical examples

In our numerical simulation, a group of discrete-time linear systems sampled from continuous linear systems $\dot{x} = \hat{A}x + \hat{B}u$ with the sampling period $\Delta t = 0.1$ are generated, where

$$\hat{A} = \begin{bmatrix} 0 & 1 \\ a_1 & a_2 \end{bmatrix}, \quad B = \begin{bmatrix} 0 \\ 1 \end{bmatrix}, \quad C = [c_1 \ c_2];$$

and a_1, a_2, c_1, c_2 are randomly generated from $\mathcal{N}(0, 3)$, respectively. However, we need to screen for satisfaction of the relative degree condition. The “real” $\bar{Q}$ is generated by letting $\bar{Q} = \hat{Q} \hat{Q}^T$, where each element of $\hat{Q}$ is sampled from the uniform distribution on $[-1, 1]$ and the time-horizon length N is set to 40. The feasible compact set for Q is set as $\mathbb{S}_+^n(5)$ and those randomly generated $\bar{Q}$ that does not belong to $\mathbb{S}_+^n(5)$ is discarded and re-generated until we have 200 groups of randomly generated “forward” problems. Each element of the initial conditions $\bar{x}^{(1:M)}$ is sampled from a uniform distribution supported on $[-5, 5]$. For each forward problem, 200 trajectories are generated, i.e., $M = 200$. We add 20dB of white Gaussian noises to $Cx_{2:N}^{(1:M)}$ to obtain $y_{2:N}^{(1:M)}$. MATLAB function `fmincon` is used to solve the risk-minimizing problem. We use $Q = I$ as the initial iteration values when solving the optimization problem in MATLAB for all IOC estimations. Denote the estimation of $\bar{Q}$ as Q_{est} .

From Fig. 1, we see that the relative error $\|Q_{\text{est}} - \bar{Q}\|_F / \|\bar{Q}\|_F$ largely decreases as M increases. This empirically justifies Theorem 3 of the statistical consistency.

Next, the result of IOC estimation under Gaussian assumptions would be illustrated. Here, 100 groups of forward problems, in which 50 trajectories are generated for each group, i.e., $M = 50$. The initial values are randomly generated by sampling from $\mathcal{N}(m_1 = 0, \Sigma_1 = 10I)$. White Gaussian noises are added to $Cx_{1:N}^{(1:M)}$ to get $y_{1:N}^{(1:M)}$ and $v_i \sim \mathcal{N}(0, \Sigma_v = 0.01I)$. We use $m_1 = I$, $\Sigma_1 = 20I$ and $Q = I$ as the initial guesses for EM-algorithm for all groups. The

Fig. 1 The relative errors of the estimation by minimizing $\mathcal{R}_M(Q)$. **a** $M = 20$. **b** $M = 200$

Fig. 2 The relative errors of the estimation by EM-algorithm under Gaussian assumption, $M = 50$, $\Sigma_v = 0.01I$

stopping criterion for iteration is $\|Q_{k+1} - Q_k\|_F \leq 10^{-4}$. The results are shown as Fig. 2.

As illustrated in Fig. 2, the relative error $\|Q_{\text{est}} - \bar{Q}\|_F / \|\bar{Q}\|_F$ is largely small, but we do get a few estimations with “huge” relative errors. This is due to the fact that EM-algorithms only guarantee that the likelihood does not descend but guarantee neither the finding of the global maximizer for the log-likelihood function nor the convergence of Q_k to some local maxima. Mitigation of such issues will be our future work.

7 Conclusions

In this investigation, IOC for finite-horizon discrete-time LQRs is studied. We first investigate the identifiability of the model structure and conditions for identifiability is given. Next, the problem of IOC is considered under the two scenarios. We first consider the case where the distributions of the initial state and the observation noise are completely unknown, yet the initial states and some noisy system output are available. The problem is formulated as a risk minimization problem. By utilizing the property of identifiability, we are able to prove the statistical consistency for such the

estimation method. We next consider the case where the initial states are not available, but the distribution as well as that of the observation noises is Gaussian. Maximum-likelihood estimation is produced by EM-algorithms. The performance of our methods is illustrated by numerical examples.

Open Access This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>.

References

1. Kalman, R. E. (1964). When is a linear control system optimal? *Journal of Basic Engineering*, 86(1), 51–60.
2. Mombaur, K., Truong, A., & Laumond, J.-P. (2010). From human to humanoid locomotion—an inverse optimal control approach. *Autonomous Robots*, 28(3), 369–383.
3. Finn, C., Levine, S., & Abbeel, P. (2016). Guided cost learning: Deep inverse optimal control via policy optimization. In *International conference on machine learning*, pp. 49–58. New York, NY, USA.
4. Berret, B., & Jean, F. (2016). Why don't we move slower? The value of time in the neural control of action. *Journal of Neuroscience*, 36(4), 1056–1070.
5. Zhang, H., Li, Y., & Hu, X. (2019). Inverse optimal control for finite-horizon discrete-time linear quadratic regulator under noisy output. In *IEEE 58th conference on decision and control (CDC)*, pp. 6663–6668. Nice, France.
6. Anderson, B. D. O., & Moore, J. B. (2007). *Optimal control: Linear quadratic methods*. Courier Corporation.
7. Jameson, A., & Kreindler, E. (1973). Inverse problem of linear optimal control. *SIAM Journal on Control*, 11(1), 1–19.
8. Fujii, T. (1987). A new approach to the LQ design from the viewpoint of the inverse regulator problem. *IEEE Transactions on Automatic Control*, 32(11), 995–1004.
9. Boyd, S., El Ghaoui, L., Feron, E., & Balakrishnan, V. (1994). *Linear matrix inequalities in system and control theory* (vol. 15). SIAM.
10. Priess, M. C., Conway, R., Choi, J., Popovich, J. M., & Radcliffe, C. (2015). Solutions to the inverse LQR problem with application to biological systems analysis. *IEEE Transactions on Control Systems Technology*, 23(2), 770–777.
11. Keshavarz, A., Wang, Y., & Boyd, S. (2011). Imputing a convex objective function. In *IEEE international symposium on intelligent control (ISIC)*, pp. 613–619. Denver, CO, USA.
12. Bertsimas, D., Gupta, V., & Paschalidis, I. C. (2015). Data-driven estimation in equilibrium using inverse optimization. *Mathematical Programming*, 153(2), 595–633.
13. Aswani, A., Shen, Z.-J., & Siddiq, A. (2018). Inverse optimization with noisy data. *Operations Research*, 66(3), 870–892.
14. Molloy, T. L., Tsai, D., Ford, J. J., & Perez, T. (2016). Discrete-time inverse optimal control with partial-state information: A

soft-optimality approach with constrained state estimation. In *IEEE 55th conference on decision and control (CDC)*, pp. 1926–1932. Las Vegas, NV, USA.

15. Molloy, T. L., Ford, J. J., & Perez, T. (2018). Finite-horizon inverse optimal control for discrete-time nonlinear systems. *Automatica*, 87, 442–446.
16. Jin, W., Kulić, D., Mou, S., & Hirche, S. (2018). Inverse optimal control with incomplete observations. [arXiv: 1803.07696](https://arxiv.org/abs/1803.07696) (arXiv Preprint).
17. Ljung, L., & Chen, T. (2013). Convexity issues in system identification. In *The 10th IEEE international conference on control and automation (ICCA)*, pp. 1–9. Hangzhou, China.
18. Zhang, H., Umenberger, J., & Hu, X. (2019). Inverse optimal control for discrete-time finite-horizon linear quadratic regulators. *Automatica*, 110, 108593.
19. Li, Y., Yao, Yu., & Hu, X. (2020). Continuous-time inverse quadratic optimal control problem. *Automatica*, 117, 108977.
20. Jennrich, R. I. (1969). Asymptotic properties of non-linear least squares estimators. *The Annals of Mathematical Statistics*, 40(2), 633–643.
21. der Vaart, V., & Aad, W. (2000). *Asymptotic statistics* (vol. 3). Cambridge University Press.
22. Söderström, T. (2012). *Discrete-time stochastic systems: Estimation and control*. Springer.
23. Nesterov, Y. (2007). Smoothing technique and its applications in semidefinite optimization. *Mathematical Programming*, 110(2), 245–259.

Han Zhang received his Ph.D. from Department of Mathematics, KTH Royal Institute of Technology, Sweden in 2019. He obtained both his B.Sc. and M.Sc. degree from Department 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 control of multi-agent systems, distributed optimization and

optimal control.

Yibei Li received her B.E. degree in Automation from Harbin Institute of Technology, China, in 2015. She received her licentiate degree in Applied and Computational Mathematics from KTH Royal Institute of Technology, Sweden, in 2019. She currently is a Ph.D. candidate at the Division of Optimization and Systems Theory, Department of Mathematics, KTH. Her research interests include inverse optimization, nonlinear control and multi-agent systems. She received the Guan Zhao-Zhi Best

Paper Award at Chinese Control Conference (CCC) in 2020.

Xiaoming Hu received the B.Sc. degree from the University of Science and Technology of China in 1983, and the M.Sc. and Ph.D. degrees from the Arizona State University in 1986 and in 1989, respectively. He served as a research assistant at the Institute of Automation, the Chinese Academy of Sciences, from 1983 to 1984. From 1989 to 1990 he was a Gustafsson Postdoctoral Fellow at KTH Royal Institute of Technology, Stockholm, where he is currently a professor of Optimization and

Systems Theory. His main research interests are in nonlinear dynamical systems, nonlinear observation and filtering, and multi-agent systems.