Inverse Optimal Control for Finite-Horizon Discrete-time Linear Quadratic Regulator Under Noisy Output

Han Zhang\textsuperscript{1}, Xiaoming Hu\textsuperscript{1}

Abstract— In this paper, the problem of inverse optimal control for finite-horizon discrete-time Linear Quadratic Regulators (LQRs) is considered. The goal of the inverse optimal control problem is to recover the corresponding objective function by the noisy observations. We consider the problem of inverse optimal control in two scenarios: 1) the distributions of the initial state and the observation noise are unknown, yet the exact observations on the initial states and the noisy observations on system output are available; 2) the exact observations on the initial states are not available, yet the observation noises are known white Gaussian and the distribution of the initial state is also Gaussian (with unknown mean and covariance). For the first scenario, we formulate the problem as a risk minimization problem and show that its solution is statistically consistent. For the second scenario, we fit the problem into the framework of maximum-likelihood and Expectation Maximization (EM) algorithm is used to solve this problem. The performance for the estimations are shown by numerical examples.

I. INTRODUCTION

The problem of inverse optimal control is first proposed by [11] and has found a multitude of applications in the field of robotics, economics and bionics [15], [6], [3]. It is well-known that the goal of a “forward” optimal control problem is to find the optimal control input as well as the optimal trajectory when the cost function, system dynamics and initial conditions are given. On the other hand, the inverse optimal control problem aims to find the parameters in the corresponding cost function, provided that the observations on the system outputs (which are contaminated by noises) are available and the system dynamics are known.

In this paper, the problem of inverse optimal control for discrete-time Linear Quadratic Regulators (LQRs) over finite-time horizon is considered. Namely, we aim to find the parameters in the quadratic objective function given the discrete-time linear system dynamics and observations of the noisy system output.

Inverse optimal control for LQR, particularly in the continuous infinite time-horizon case, has been studied by a number of authors [1], [8], [7]. They assume the optimal feedback gain $K$ is known exactly and focus on recovering the objective function. It was shown in [5] that the search for matrices $Q$ and $R$ can be formulated with Linear Matrix Inequalities (LMI) when the feedback gain $K$ is known. Nevertheless, in many scenarios, the feedback control gain $K$ can not be known exactly. In addition, we are interested in case of finite-time horizon, wherein the feedback gain $K_t$ is time-varying. [17] considers the discrete infinite time-horizon case with noisy observations, in which the optimal feedback gain $K$ is time-invariant. Their approach is to identify the feedback matrix $K$ first and solve for $Q$ and $R$ similar to the method proposed in [5]. Such an approach is not applicable in the finite-time horizon case, since the feedback gain $K_t$ is time-varying. Furthermore, the idea of “identify the feedback gain $K_t$, then compute the corresponding $Q$” suffers from the huge number of parameters in the identification stage, i.e., the number of $K_t$’s is proportional to the length of the time-horizon and the length of the time-horizon is usually large. In addition, such identification does not use the knowledge that the $K_t$’s are actually generated by an LQR hence the performance of such method can not be guaranteed in the finite-time horizon case.

Spoken from a broader perspective, the problem of inverse optimal control has received considerable attention in the community. In particular, people have been focused on systems with nonlinear dynamics and objective functions composed by weighted-sum of base functions in recent years. In [12] and [4], the authors analyze the optimality conditions for the optimal control problem and a method based on minimizing the violation of these conditions is proposed. Nevertheless, as pointed out by [2], the approaches proposed in [12] and [4] are not statistically consistent and sensitive to observation noise. [2] present a statistically consistent formulation, but results in a difficult optimization problem. [14], [13] also consider the discrete finite time-horizon case. They consider the Pontryagin’s Maximum Principle (PMP) for the optimal control problem and pose an optimization problem whose constraints are two of the three conditions of PMP; they then minimize the residual of the third PMP condition. In addition, they assume the optimal control input is known exactly while in our case, the estimation of the parameter relies on exact samples of initial values or noisy system output (only). The question of identifiability, i.e. uniqueness of the solution, for this approach is addressed, however, the statistical consistency of the estimation is not claimed. In a very recent work [10], the authors consider the discrete-time inverse optimal control problem for nonlinear systems when some segments of the trajectories and input observations are missing.

Without any doubt, the problem of inverse optimal control for LQR can be viewed as a special case of aforementioned problems for nonlinear systems. However, special structure of LQR is utilized and we are able to show the statistical consistency for the estimation. Furthermore, under Gaussian assumptions, we are able to use EM-algorithm to solve the
problem due to its special structure. The performance of the algorithms are shown by numerical examples.

Notations: $S^n_p$ denotes the cone of $n$ dimensional positive semi-definite matrices, $S^n_{p+}$ denotes the set of strictly positive semi-definite matrices, $\| \cdot \|_F$ denotes the Frobenius norm of a matrix, $\| \cdot \|_2$ denotes the $l_2$ norm of a vector, $tr(\cdot)$ denotes the trace of a matrix and $tr(G_1^T G_1) = \|G_1\|^2_F$. We denote $\otimes$ as the Kronecker product. The $i$'th row of matrix $G$ is denoted as $[G]_i$.

II. PROBLEM FORMULATION

It is well-known that the “forward” optimal LQ problem reads

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

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

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

Suppose we can observe the noisy output $y_t = C x_t + v_t$, where $v_t \in \mathbb{R}^l$ denotes the noise term, $v_t \in \mathbb{R}^l$ and the matrices ($A,B,C$) are known. For simplicity, in this paper, we consider the case of $R = I$ and $S = 0$. The inverse optimal control problem 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.

In addition, it is assumed that ($A,B$) is controllable and $B$ has full column rank. Moreover, we assume that $A$ is invertible. To see that the assumption is reasonable, consider a discrete-time system sampled from a continuous linear system $\dot{x} = \dot{A} x + B u$, where the sample period $\Delta t$ is small. Hence for the discretized linear system, we have $A = e^{A \Delta t}$, $B = \int_0^{\Delta t} e^{A \tau} B d\tau$. It is clear that $A \in e^{A \Delta t}$ is invertible.

III. MAIN RESULT

A. Inverse Optimal Control Using Exact Initial Values and Noisy Output

Inspired by [2], we first formulate the inverse optimal control problem when exact samples of the initial value $x_1 = \bar{x}$ and noisy output $y_{2:N}$ are available. Suppose the probability space $\{\Omega, \mathcal{F}, \mathbb{P}\}$ carries independent random vectors $\bar{x} \in \mathbb{R}^n, \{v_t \in \mathbb{R}^m\}_{t=2}^n$ distributed according to some unknown distributions. The following assumptions are made in the remainder of this section:

Assumption 1. $\mathbb{E}(\|\bar{x}\|^2) < +\infty, \mathbb{E}(v_t) = 0$ and $\mathbb{E}(\|v_t\|^2) < +\infty$.

Assumption 2. $\forall \eta \in \mathbb{R}^n, \exists (\eta) \in \mathbb{R} \setminus \{0\}$ such that $\mathbb{P}(\bar{x} \in \mathcal{B}_\varepsilon(\eta)) > 0, \forall \varepsilon > 0$, where $\mathcal{B}_\varepsilon(\eta)$ is the open $\varepsilon$-ball centered at $\eta$.

Assumption 3. The discrete-time linear system defined by ($A,B,C$) is a square-system, i.e., $l = m$ and has relative degree [19] ($r_1, \cdots, r_m$).

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

With the assumptions above, we see that when solving the “forward” optimal LQ problem, the initial value $x_1$ used is actually a realization of the random vector $\bar{x}$, i.e., $x_1 = \bar{x}(\omega)$; hence the “forward” LQR problem can actually be seen as

$$\min_{u_{1:N-1}(\omega),x_{2:N}(\omega)} \{ J(u_{1:N-1}(\omega), x_{2:N}(\omega); Q; \bar{x}(\omega)) \} \bigg|_{\omega}$$

given $\omega \in \Omega$, (3)

Note that the optimal control input and trajectory $\{u^*_t\}, \{x^*_t\}$ are now random vectors implicitly determined by the random variable $\bar{x}$ and the parameter $Q$. With the formulation of the “forward problem" [3], we can define the risk function

$$\mathcal{R}(Q) = \mathbb{E}_\bar{x} [J(Q; \xi)]$$

where $\xi = [x^T, Y^T]^T, Y = [y^T_1, \cdots, y^T_N]^T$ and $f : S^m_+ (\varphi) \times \mathbb{R}^{n \times (N+1)} \mapsto \mathbb{R}$

$$f(Q; \xi) = \sum_{i=2}^{N} \| y_i - C x^*_i(Q; \bar{x}) \|^2, \quad (5)$$

and $x^*_{2:N}(Q; \bar{x})$ is the optimal trajectory to (3).

In order to solve the inverse optimal control problem, it is rational to minimize the risk function, namely,

$$\min_{Q \in S^m_+ (\varphi)} \mathcal{R}(Q). \quad (6)$$

Nevertheless, since the distributions of $\bar{x}$ and $v_{2:N}$ are unknown, it is not possible to solve (6) directly. However, in principle, can be approximated by

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

where $\xi^{(i)}$ are i.i.d random samples. The statistical consistency for this approximation will be shown later.

On the other hand, by Pontryagin’s Maximum Principle (PMP), if $u^*_{1:N-1}$ and $x^*_{1:N}$ are the optimal control and corresponding trajectory, then there exists adjoint variables $\lambda^*_{2:N}$ such that

$$x^*_{t+1} = A x^*_t + B u^*_t, \quad t = 1 : N - 1$$

$$\lambda^*_t = A^T \lambda^*_{t+1} + Q x^*_t, \quad t = 2 : N - 1, \quad \lambda^*_N = 0$$

$$u_t = - B^T \lambda^*_{t+1}, \quad t = 1 : N - 1. \quad (8)$$

Although PMP in general provides necessary conditions for optimality, note that it becomes also sufficient conditions for optimality since the solution to the “forward” LQR problem is unique. Hence we can express $x^*_{2:N}$ in [3] using [5]. Thus
where $G = I_{N-1} \otimes [C, 0_{t \times n}]$ and $Y = \begin{bmatrix} y_g^T, \cdots, y_N^T \end{bmatrix}^T$.

It is clear that $f(Q; \xi)$ is continuous with respect to $\xi$, hence it is a measurable function of $\xi$ at each $Q$. Further, $\mathcal{F}(Q)$ is continuous and hence $\mathcal{F}(Q)^{-1}$ is continuous. Then $f(Q; \xi)$ is also continuous with respect to $Q$.

On the other hand, since $\mathcal{F}(Q)^{-1}$ is continuous and $Q$ lives in a compact set, then $\mathcal{F}(Q)^{-1} F$ is bounded, i.e., $\|\mathcal{F}(Q)^{-1} F\| < \bar{\phi}$ for some finite positive $\bar{\phi}$. It follows that $\mathbb{E}(\|Z^*\|^2) = \mathbb{E}(\|\mathcal{F}(Q)^{-1} A \bar{x}\|^2) \leq \|\mathcal{F}(Q)^{-1}\|^2 \|\bar{A}\|^2 \mathbb{E}(\|\bar{x}\|^2) < +\infty$, where $Z^*$ corresponds to the “true” $Q$.

Recall that $y_t = C x_t^* + v_t$ and this implies that $Y = G Z^* + \zeta$, where $\zeta = \begin{bmatrix} \nu_2^T, \cdots, \nu_N^T \end{bmatrix}^T$.

By Assumption [1] $\mathbb{E}(\|v_t\|^2) < +\infty$, which implies $\mathbb{E}(\|\zeta\|^2) < +\infty$. Therefore, $\mathbb{E}(\|Y\|^2) = \mathbb{E}(\|G Z^* + \zeta\|^2) \leq 2 \{\mathbb{E}(\|G Z^*\|^2) + \mathbb{E}(\|\zeta\|^2)\} \leq 2 \{\|G\|^2 \mathbb{E}(\|Z^*\|^2) + \mathbb{E}(\|\zeta\|^2)\} < +\infty$.

Hence it holds that

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

$$\leq 2 \left(\|Y\|^2 + \|G \mathcal{F}(Q)^{-1} \bar{A} \bar{x}\|^2\right)$$

$$\leq 2 \left(\|Y\|^2 + \|G\|^2 \mathbb{E}(\|\mathcal{F}(Q)^{-1}\|^2 \|\bar{A}\|^2 \|\bar{x}\|^2)\right)$$

$$\leq 2 \left(\|Y\|^2 + \|G\|^2 \mathbb{E}(\|\mathcal{F}(Q)^{-1}\|^2 \|\bar{A}\|^2 \|\bar{x}\|^2)\right) =: d(\xi),$$

and it is clear that $\mathbb{E}(d(\xi)) < +\infty$ since $\mathbb{E}(\|Y\|^2) < +\infty$ and $\mathbb{E}(\|\zeta\|^2) < +\infty$. By the analysis above, we conclude that the uniform law of large numbers [9] applies, namely,

$$\sup_{Q \in \mathcal{S}_{+}^{N}(\varphi)} \|\sum_{i=1}^{M} f(Q, \xi^{(i)}) - \mathbb{E}[f(Q; \xi)]\|_{\mathbb{F}} \rightarrow 0.$$  

(11)

Besides (11), if we are able to show $Q$ is the unique optimizer to (6), then $Q^*_M \xrightarrow{P} Q$ follows directly from Theorem 5.7 in [20].

Note that by assumption, $\bar{x}$, $\{v_t\}$ are independent, hence $x_t^*(Q, \bar{x}) + v_t$, $\mathbb{E}(v_t) = 0$, $t = 2 : N$ (4) can be simplified as

$$\mathcal{F}(Q) = L(Q) + \sum_{t=2}^{N} \mathbb{E}(\|v_t\|^2),$$

where $L(Q) = \sum_{t=2}^{N} \|y_t(Q, x) - C x_t(Q, x)\|^2$, $y_t(Q, \bar{x}) = C x_t^*(Q, \bar{x})$, and $\bar{y}_{t+1}(Q, \bar{x}) = C x_t^*(Q, \bar{x})$. It is clear that $Q$ minimizes the risk function $\mathcal{F}(Q)$. What remains to show is the uniqueness.

By Assumption [3] the system has relative degree $(r_1, \cdots, r_m)$ and it holds by definition that

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

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

is nonsingular,

where $c_i$ denotes the $i$'th row of $C$. Therefore it follows that $\bar{y}_{t+1}(Q, \bar{x}) = c_i \Pi_{j=1}^{m} (A + BK_j) \bar{x}$

$$= c_i A^r \bar{x}, 
 t = 1 : r_i - 1, 
 i = 1 : m, 
 \bar{y}_{t+1}(Q, \bar{x}) = c_i A^r \bar{x} + c_i A^{r_i - 1} B K_i(Q) \bar{x}.$$


By Lemma III.1 we know that if $Q \neq \tilde{Q}$, then $K_1(Q) \neq K_1(\tilde{Q})$. Hence there exists $\eta \in \mathbb{R}^n$, such that $K_1(Q)\eta \neq K_1(\tilde{Q})\eta$. On the other hand, by Assumption 2 $\exists r(\eta) \neq 0$, such that $\mathbb{P}(\bar{x} \in B_r(\eta)) > 0, \forall \varepsilon$. Since $r \neq 0$ and $\mathcal{L}$ is nonsingular, it follows that $\mathcal{L}K_1(Q)\eta \neq \mathcal{L}K_1(\tilde{Q})\eta$. Furthermore, since $\mathcal{L}K_1(Q)\eta$ is continuous with respect to $\eta$, $\forall Q \in \mathbb{S}^n$, this implies $\exists \eta \in \mathbb{S}^n$, such that $\mathcal{L}K_1(Q)\bar{x} \neq \mathcal{L}K_1(\tilde{Q})\bar{x}$, $\forall \bar{x} \in B_{\varepsilon}(\eta)$ and $\mathbb{P}(\bar{x} \in B_{\varepsilon}(\eta)) > 0$. Thus $\mathcal{L}K_1(Q)\eta \geq \mathcal{L}K_1(\tilde{Q})\eta \|X(\epsilon)\|_2^2P(\omega) > 0$. Hence $\tilde{Q}$ is the unique minimizer to (4) and the statement follows.

We have shown that the risk minimization problem (9) is statistically consistent. As stated in the proof, the $\tilde{Q}$ can be rewritten in the compact form of minimization of $\sum_{i=1}^{tr} \epsilon^2 \sum_{i=1}^{N} |Y(i) - G_i \tilde{A} \tilde{x}|^2$. To solve the problem, we introduce the following convex function $f_{\epsilon}: \mathbb{R}^n \mapsto \mathbb{R}$, where $f_{\epsilon}(Q) = \epsilon \ln tr(e^{Q/\epsilon}) = \epsilon \ln \sum_{i=1}^{N} e^{\sigma_i(Q)/\epsilon}$ [16], where $\sigma_i(Q)$ is the $i$th largest eigenvalue of $Q$. It holds that $\sigma_1(Q) \leq f_{\epsilon}(Q) \leq \sigma_1(Q) + \epsilon \ln n$. Hence when $\epsilon$ is small, the function $f_{\epsilon}(Q)$ approximates the largest eigenvalue of $Q$ well. On the other hand, the gradient of $f_{\epsilon}(Q)$ is $\nabla_Q f_{\epsilon}(Q) = [\sum_{i=1}^{N} e^{\sigma_i(Q)/\epsilon}]^{-1}[\sum_{i=1}^{N} e^{\sigma_i(Q)/\epsilon} \nu_i \nu_i^T]$, where $(\sigma_i(Q), \nu_i)$ are eigen-pairs of $Q$ with $\|\nu_i\| = 1, \forall i$. Note that for $\epsilon$ small enough, the gradient only numerically depends on the eigenvectors that correspond to the largest eigenvalues [16], which makes the gradient easy to compute. With the setup above, we approximate the semi-positive definite constraint $Q \in \mathbb{S}^n_+$ with $f_{\epsilon}(Q) \leq 0$ and we can solve the optimization problems with standard nonlinear optimization solvers.

B. Inverse Optimal Control Using Noisy Output Under Gaussian Assumption

In Section III-A it is assumed that the distributions of the noise and the initial value are unknown, yet we can get exact samples of the initial value. But in some scenarios, the exact sample on initial value is not available and we can only observe the noisy measurement $y_1 = Cx_1 + v_1$. In this scenario, we have the following assumption.

Assumption 5. $\bar{x} = x_1 \sim \mathcal{N}(m_1, \Sigma_1)$, $v_1 \sim \mathcal{N}(0, \Sigma_v)$ and $v_{1:N}$ are white. $\Sigma_v$ is known, but $m_1$ and $\Sigma_1$ is not a prior known.

Recall the proof of Theorem III.1 we can write $Y = [y_2^T, \cdots, y_N^T]^T = G \mathcal{F}(Q)^{-1} \tilde{A} \tilde{x} + [v_2^T, \cdots, v_N^T]^T$. More precisely, it holds that

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

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

$$y_{t+1} = \tilde{y}_t, \quad t = 1 : N - 1, \quad \tilde{y}_t = \tilde{x}$$

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

where $\tilde{C}_1(Q) = C$ and $\tilde{C}_2(Q) = G_t \mathcal{F}(Q)^{-1} \tilde{A}, \quad t = 2, N$. Now, under Assumption 5 the inverse optimal control problem can be viewed as a system identification problem for Linear Gaussian Model, which can be solved using maximum log-likelihood method. If there are $M$ output sequences available, then we have $M$ i.i.d samples $x_1^{(1:M)}$ of the initial value. Since $v_t^{(i:j)}$ is also i.i.d., $y_t^{(i)}$ and $y_t^{(j)}$ are independent for all $i \neq j$. We would like to use EM-algorithm to solve the maximum log-likelihood problem.

Note that in (12), all of the stochasticity comes from the initial value $\eta_1 = \bar{x}$ as well as the output measurements $y_{1:N}$, hence we choose $\eta_1$ to be the latent variable and let $\theta$ parametrize $Q, m_1$ and $\Sigma_1$. We now compute the auxiliary function $Q(\theta, \theta_k)$ of the EM-algorithm.

Proposition III.1. Under Assumption 5 [10] the auxiliary function $Q(Q, Q_k)$ that corresponds to the model (12) is given by

$$Q(\theta, \theta_k) \propto Q_1(\theta, \theta_k) + Q_2(\theta, \theta_k),$$

where

$$Q_1(\theta, \theta_k) = \sum_{i=1}^{M} \left\{ \ln \det \Sigma_i + tr \left[ \Sigma_i^{-1} P_i^{(i)} \right] \right\} + \sum_{t=2}^{N} \left[ \sum_{i=1}^{M} \left( \hat{\eta}_i^{(i)} - \Sigma_i^{-1} \Sigma_i^{-1} \right)^T \right]$$

$$Q_2(\theta, \theta_k) = \sum_{i=1}^{M} \sum_{t=2}^{N} \left[ \sum_{i=1}^{M} \left( y_t^{(i)} - \hat{C}_i(Q) \hat{\eta}_i^{(i)} \right)^T \right],$$

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

Proof. By Bayes’ rule, it follows that

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

Hence by the definition of the auxiliary function $Q(Q, Q_k)$,
we have

\[ Q(\theta, \theta_k) = \int \ln p_\theta(\tilde{y}_i^{(1:M)}, y_{1:N}^{(1:M)}) p_{\theta_k}(\tilde{y}_i^{(1:M)} | y_{1:N}^{(1:M)}) d\tilde{y}_i^{(1:M)} \]

\[ = \int \left( \sum_{i=1}^{M} \ln p_\theta(\tilde{y}_i^{(1)}, y_{1:N}^{(1)}) \prod_{i=1}^{M} p_{\theta_k}(\tilde{y}_i^{(1)} | y_{1:N}^{(1)}) d\tilde{y}_i^{(1)} \right) \]

\[ = \sum_{i=1}^{M} \int \ln p_\theta(\tilde{y}_i^{(1)}, y_{1:N}^{(1)}) p_{\theta_k}(\tilde{y}_i^{(1)} | y_{1:N}^{(1)}) d\tilde{y}_i^{(1)} \]

\[ = \sum_{i=1}^{M} \left\{ \int \ln p_\theta(\tilde{y}_i^{(1)}) p_{\theta_k}(\tilde{y}_i^{(1)} | y_{1:N}^{(1)}) d\tilde{y}_i^{(1)} \right\} + \sum_{i=1}^{M} \int \ln p_\theta(y_i^{(1)}, \tilde{y}_i^{(1)}) p_{\theta_k}(\tilde{y}_i^{(1)} | y_{1:N}^{(1)}) d\tilde{y}_i^{(1)} \}

Note that \( y_i^{(i)} = C \eta_i^{(i)} + v_i^{(i)} \), hence \( p_\theta(y_i^{(i)} | \eta_i^{(i)}) \sim N(C \eta_i^{(i)}, \Sigma_v) \) and it is irrelevant with respect to \( \theta \).

By using the property of matrix trace and the fact of \( \Sigma_{\theta_k} \eta_i^{(i)} y_{1:N}^{(1)} = \eta_i^{(i)} y_{1:N}^{(1)} \) + \( P_{1:N} \), also by ignoring the constant and irrelevant terms, it follows that

\[ Q(\theta, \theta_k) \approx -\sum_{i=1}^{M} \left\{ \ln \det \Sigma_1 + \| \eta_i^{(i)} - m_1 \|^2 \Sigma_1^{-1} \right\} \]

\[ + \sum_{i=2}^{N} \int \| y_i^{(i)} - \tilde{C}_1(Q) \eta_i^{(i)} \|^2 \Sigma_1^{-1} \right\} d\eta_i^{(i)} \}

\[ = Q_1(\theta, \theta_k) + Q_2(\theta, \theta_k). \]

Note that \( \eta_i^{(i)} \) and \( P_{1:N} \) can be obtained by fixed point smoothing [18]. We now check (13) and (14) in detail. In each iteration of EM-algorithm, we need to maximize \( Q(\theta, \theta_k) \) with respect to \( \theta \). Note that \( m_1 \) and \( \Sigma_1 \) only show up in \( Q_1(\theta, \theta_k) \); \( Q \) shows up only in \( Q_2(\theta, \theta_k) \), therefore we can optimize them separately. Consider the derivative of \( Q_1(\theta, \theta_k) \) with respect to \( m_1 \) and \( \Sigma_1 \), the first order optimality condition for maximizing \( Q_1(\theta, \theta_k) \) reads:

\[ \frac{\partial m_1 Q_1(\theta^*, \theta_k)}{\partial m_1} = -\frac{\partial m_1}{\partial m_1} \left( \sum_{i=1}^{M} (m_i^{(i)} - \eta_{1:N}^{(i)})^T \Sigma_1^{-1} (m_i^{(i)} - \eta_{1:N}^{(i)}) \right) \]

\[ = -2 \sum_{i=1}^{M} \Sigma_1^{-1} (m_i^{(i)} - \eta_{1:N}^{(i)} \right) = 0, \]

\[ \frac{\partial \Sigma_1 Q_1(\theta^*, \theta_k)}{\partial \Sigma_1} = -\sum_{i=1}^{M} \left\{ \Sigma_1^{-1} - (m_i^{(i)} - \eta_{1:N}^{(i)})^T \Sigma_1^{-1} \right\} \]

\[ = -\Sigma_1^{-1} (m_i^{(i)} - \eta_{1:N}^{(i)})^T \Sigma_1^{-1} = 0. \]

Solving (16) and (17), we get

\[ m_i^{(i)} = \frac{1}{M} \sum_{i=1}^{M} \eta_{1:N}^{(i)}. \]

\[ \sum_{i=1}^{M} \left[ \Sigma_1 - (m_i^{(i)} - \eta_{1:N}^{(i)}) (m_i^{(i)} - \eta_{1:N}^{(i)})^T - P_{1:N} \right] \Sigma_1^{-1} = 0, \]

\[ \Leftrightarrow \sum_{i=1}^{M} \left[ \Sigma_1 - (m_i^{(i)} - \eta_{1:N}^{(i)}) (m_i^{(i)} - \eta_{1:N}^{(i)})^T - P_{1:N} \right] = 0, \]

\[ \Leftrightarrow \Sigma_1 = \frac{1}{M} \sum_{i=1}^{M} \left[ (m_i^{(i)} - \eta_{1:N}^{(i)}) (m_i^{(i)} - \eta_{1:N}^{(i)})^T + P_{1:N} \right]. \]
Next, we illustrate the performance of the estimation by using EM-algorithm under Gaussian assumptions. The groups of discrete-time linear systems \((A, B, C)\) and the “real” \(\tilde{Q}\) are generated in the same way as mentioned above. 100 groups for discrete-time linear systems are generated. For each system \((A, B, C)\), we generate 30 trajectories, i.e., \(M = 30\). The initial values are sampled from \(N(m_1 = 0, \Sigma_1 = 10I)\). White Gaussian noises are added to \(C x_1^{1:N} \) to get \(y_1^{1:N}\) and \(v_t \sim N(0, \Sigma_v = 0.01I)\). We use \(m_1 = I, \Sigma_1 = 20I\) and \(Q = I\) as the initial values for the iterations in EM-algorithm for all groups of linear systems. The iterations are stopped when \(\|Q_{k+1} - Q_k\|_F \leq 10^{-4}\). The results are shown as Fig. 2.

![Fig. 1. The relative errors of the estimation by minimizing \(D_M(Q)\).](image1)

As it can be seen in Fig. 2, the relative error \(\|Q_{est} - \tilde{Q}\|_F / \|Q\|_F\) is roughly small, while only a few of them is “huge”. This is because EM-algorithm does not guarantee to find the global maximizer for the log-likelihood function and \(Q_k\) converges to some local maxima.

V. CONCLUSION

In this paper, the problem of inverse optimal control for finite-horizon discrete-time LQRs is considered under the following two scenarios: 1) the distributions of the initial state and the observation noise are unknown, yet the exact observations on the initial states and the noisy observations on system output are available; 2) the exact observations on the initial states are not available, yet the observation noises are known white Gaussian and the distribution of the initial state is also Gaussian (with unknown mean and covariance). For the first scenario, the problem is formulated as a risk minimization problem and the statistical consistency for the estimation is proven. For the second scenario, we fit the problem into the framework of maximum-likelihood and EM-algorithm is used to solve this problem. We show the performance for the estimations empirically by numerical examples.

REFERENCES