On the Sample Complexity of the Linear Quadratic Gaussian Regulator

Abed AlRahman Al Makdah and Fabio Pasqualetti

***Abstract*— In this paper we provide direct data-driven expressions for the Linear Quadratic Regulator (LQR), the Kalman filter, and the Linear Quadratic Gaussian (LQG) controller using a finite dataset of noisy input, state, and output trajectories. We show that our data-driven expressions are consistent, since they converge as the number of experimental trajectories increases, we characterize their convergence rate, and we quantify their error as a function of the system and data properties. These results complement the body of literature on data-driven control and finite-sample analysis, and they provide new ways to solve canonical control and estimation problems that do not assume, nor require the estimation of, a model of the system and noise and do not rely on solving implicit equations.**

I. Introduction

Consider the discrete-time, linear, time-invariant system

$$\begin{aligned} x(t+1) &= Ax(t) + Bu(t) + w(t), \\ y(t) &= Cx(t) + v(t), \qquad t \geq 0, \end{aligned} \tag{1}$$

where $x(t) \in \mathbb{R}^n$ denotes the state, $u(t) \in \mathbb{R}^m$ the control input, $y(t) \in \mathbb{R}^p$ the measured output, $w(t)$ and $v(t)$ the process and measurement noise at time t. The LQG control problem asks for the input that minimizes the cost function

$$\lim_{T\to\infty} \mathbb{E}\left[\frac{1}{T}\Big(\sum_{t=0}^{T-1} x(t)^\mathsf{T} Q_x x(t) + u(t)^\mathsf{T} R_u u(t)\Big)\right], \tag{2}$$

where $Q_x \succeq 0$, $R_u \succ 0$ are weight matrices and T is the control horizon. With the standard assumptions that[1]

(A1) the process and measurement noise sequences and the initial state are independent at all times and satisfy $w(t) \sim \mathcal{N}(0, Q_w)$, $v(t) \sim \mathcal{N}(0, R_v)$, and $x(0) \sim \mathcal{N}(0, \Sigma_0)$, with $Q_w \succeq 0$, $R_v \succ 0$, and $\Sigma_0 \succ 0$;

(A2) the pairs (A, B) and $(A, Q_w^{\frac{1}{2}})$ are controllable, and the pairs (A, C) and $(A, Q_x^{\frac{1}{2}})$ are observable;

the input that solves the LQG problem can obtained by concatenating the Kalman filter for (1) with the (static) controller that solves the LQR problem for (1) with weight matrices Q_x and R_u [1]. That is,

$$u^*(t) = K_{\text{LQR}} x_{\text{KF}}(t), \tag{3}$$

where $x_{\text{KF}}(t)$ is the Kalman estimate of the state $x(t)$. The classic, model-based computation of the LQR gain and Kalman filter in (3) requires the complete knowledge of the system (1), including the noise statistics. Motivated by the recent successes of data-driven and machine-learning methods, we seek here a solution to the LQG problem that relies only on a (finite) dataset of experimental data, without the need to estimate the system dynamics and noise statistics.

Related work. Data-driven methods for system analysis and control have flourished in the last years and are revolutionizing the field [2]. The methods developed in this paper fall in the category of direct data-driven methods [3], where controls are obtained directly from data bypassing the classic system identification step [4]. In line with earlier work and differently from optimization-based approaches [5], [6], we pursue here closed-form data-driven expressions, which are typically computationally advantageous [7], are transparent, and can reveal novel insights into the problems [8].

This paper focuses on data-driven LQG control, while most of the literature on data-driven control has focused on the LQR problem with noiseless data [9]–[11]. Recent work [12] has studied the design of data-driven controllers from noisy data [13], [14], the design of data-driven Kalman filters [15], imitation-based LQG control design [16], and some versions of the output-weighted LQG control problem [17], [18]. Compared to [18], in particular, this paper does not assume perfect knowledge of the Markov parameters or any part of the system dynamics and noise, and it does not estimate them to solve the state-weighted LQG problem. To the best of our knowledge, this paper contains the first direct, closed-form data-driven solution to the state-weighted LQG problem, with finite-sample performance guarantees.

The recent literature on the analysis of the sample complexity of estimation and control problems is also relevant to this work. In particular, [19], [20] follow an indirect approach, where sample complexity bounds are derived for the identification of the system dynamics and such errors are propagated towards the design of LQR and LQG controllers. Differently from this paper, this analysis is valid only for stable systems and output-weighted LQG costs. Bounds on the performance of the learned LQG controller are also derived in [21] assuming a sufficiently small error in the system identification step [22]–[25], and in [26] where the optimal LQR is learned in a model-free setting using gradient methods. Although this paper makes use of similar technical tools, the approach pursued here is direct and does not rely on the identification of the system matrices, nor on optimization algorithms to design or tune robust controllers. Further, this paper considers the canonical LQG setting, rather than the noisy LQR problem or the output-weighted LQG problem with noisy controls, and it provides closed-form expressions for the optimal controllers rather than their performance.

This paper is based upon work supported in part by awards ONR-N00014-19-1-2264, AFOSR-FA9550-20-1-0140, and AFOSR-FA9550-19-1-0235. A. A. Al Makdah and F. Pasqualetti are with the Department of Electrical and Computer Engineering and Mechanical Engineering at the University of California, Riverside, {aalm005,fabiopas}@ucr.edu.

[1]These assumptions also hold throughout this paper.

Contributions of the paper. The main contributions of this paper is the characterization of direct data-driven formulas for the LQR gain, Kalman filter, and LQG gain using a dataset of trajectories of the input, state, and output of the system (1). Importantly, since the experimental data is noisy and the system dynamics and noise statistics are unknown, we show that our formulas are consistent, as they converge to the true expressions when the amount of experimental data increases. Additionally, we characterize the convergence rate of our expressions, as well as their error when the data is of finite size. Finally, we provide illustrative examples and remarks to highlight how the properties of the system and of the experimental data affect the accuracy of our formulas.

Organization of the paper. The remainder of the paper is organized as follows. Section II formalizes our problem setting and contains some preliminary results. Section III contains our main results and examples, and Section IV concludes the paper. Finally, all proofs are in the Appendix.

Notation. A Gaussian random variable x with mean μ and covariance Σ is denoted as $x \sim \mathcal{N}(\mu, \Sigma)$. The $n \times n$ identity matrix is denoted by I_n, and the $n \times m$ zero matrix is denoted by $0_{n\times m}$. The expectation operator is denoted by $\mathbb{E}[\cdot]$. The trace of a square matrix A is denoted by $\mathrm{tr}\,[A]$. A positive definite (semidefinite) matrix A is denoted as $A \succ 0$ ($A \succeq 0$). The Kronecker product is denoted by $\otimes$, and the vectorization operator is denoted by $\mathrm{vec}(\cdot)$. The left (right) pseudo inverse of a tall (fat) matrix A is denoted by $A^\dagger$. A block-diagonal matrix with block matrices A and B is denoted by $\mathrm{blkdiag}(A, B)$. The smallest (largest) singular value of a matrix A is denoted by $\sigma_{\min}(A)$ $\big(\sigma_{\max}(A)\big)$.

II. Problem formulation and preliminary results

In this work we aim to compute the LQG inputs in a data-driven setting where datasets from offline experiments are available but the system matrices and noise statistics are unknown. In particular, we have access to the following data:

$$U = \begin{bmatrix} u^1 & \cdots & u^N \end{bmatrix},\ X = \begin{bmatrix} x^1 & \cdots & x^N \end{bmatrix},\ Y = \begin{bmatrix} y^1 & \cdots & y^N \end{bmatrix}, \tag{4}$$

where x^i and y^i are the i-th state and output trajectories of (1) generated by the input u^i. That is, for $i \in \{1, \dots, N\}$,

$$u^i = \begin{bmatrix} u^i(0) \\ \vdots \\ u^i(T-1) \end{bmatrix}, x^i = \begin{bmatrix} x^i(0) \\ \vdots \\ x^i(T) \end{bmatrix}, y^i = \begin{bmatrix} y^i(0) \\ \vdots \\ y^i(T) \end{bmatrix},$$

where T is the horizon of the control experiments. We make the following assumption on the experimental inputs.

Assumption 2.1: ***(Experimental inputs)*** The inputs in (4) are independent and identically distributed, that is, $u^i(t) \sim \mathcal{N}(0, \Sigma_u)$, with $\Sigma_u \succ 0$, for all $i \in \{1, \dots, N\}$ and times. □

In our analysis we will make use of an equivalent characterization of the LQG inputs derived in [27, Theorem 2.1], which shows that these inputs can also be computed as

$$u^*(t+n) = K_{\text{LQG}} \begin{bmatrix} u^*(t) \\ \vdots \\ u^*(t+n-1) \\ y^*(t+1) \\ \vdots \\ y^*(t+n) \end{bmatrix}, \tag{5}$$

where the static gain K_{LQG} depends on the system and noise matrices, and y^* is the output of (1) with input u^*.

Remark 1: ***(State vs output measurements)*** We assume here that the state of the system (1) can be directly measured. This assumption is easily satisfied in certain lab experiments, where additional sensors (e.g., a motion capture system for robotic applications) can be deployed during the design stage to measure the system state and collect training data. Further, state measurements are necessary to solve the state-weighted LQG problem, since the state weight matrix Q_x uses specific coordinates that cannot be inferred from output measurements only [28], but they can be substituted with input and output measurements for different versions of the LQG problem. See also [27] for a reformulation of the LQG problem that uses only input and output measurements. □

III. Data-driven formulas for LQG control

In this section we derive our main results, that is, direct data-driven formulas for the LQR controller, the Kalman filter, and the LQG controller using the data (4). Additionally, we show that these formulas are consistent, i.e., they converge to the true model-based expressions as the data grows, and we finally quantify their error when the data is finite.

We start by introducing some additional notation. Let

$$X_t = \begin{bmatrix} x^1(t)^\mathsf{T} & \cdots & x^N(t)^\mathsf{T} \end{bmatrix}^\mathsf{T}, \tag{6}$$

and, given input and state trajectories $u_\text{v} \in \mathbb{R}^{mT}$ and $x_\text{v} \in \mathbb{R}^{nT}$, let $u_\text{m} \in \mathbb{R}^{m\times T}$ and $x_\text{m} \in \mathbb{R}^{n\times T}$ be the matrices obtained by reorganizing the inputs and states in the vectors u_v and x_v in chronological order. The next result characterizes the LQR gain from data.

Theorem 3.1: ***(Data-driven LQR gain)*** Let $x_0 \in \mathbb{R}^n$ and

$$\begin{bmatrix} u_\text{v} \\ x_\text{v} \end{bmatrix} = \begin{bmatrix} H \\ M \end{bmatrix} P^{-1/2} \left(\begin{bmatrix} I_n & 0_{n\times mT} \end{bmatrix} P^{-1/2} \right)^\dagger x_0, \tag{7}$$

where

$$\begin{aligned} &H = \begin{bmatrix} 0_{mT\times n} & I_{mT} \end{bmatrix}, M = X \begin{bmatrix} X_0 \\ U \end{bmatrix}^\dagger, \text{ and} \\ &P = M^\mathsf{T} (I_{T+1} \otimes Q_x) M + \mathrm{blkdiag}\left(0_{n\times n}, I_T \otimes R_u\right). \end{aligned} \tag{8}$$

Let $x_\text{v}^* \in \mathbb{R}^{nT}$ be the trajectory of (1) with initial state x_0, control input $u^*(t) = K_{\text{LQR}} x(t)$, and $w(t) = 0$ at all times. Then, the data-driven estimate $K_{\text{LQR}}^\text{D} = u_\text{m} x_\text{m}^\dagger$ of K_{LQR}

$$\|K_{\text{LQR}} - K_{\text{LQR}}^\text{D}\|_2 \le \frac{1}{\sigma_{\min}(x_\text{m}^*)\,(1 - \kappa(x_\text{m}^*))} \left(\frac{c_1}{\sqrt{N}} + c_2 \rho^T \right),$$

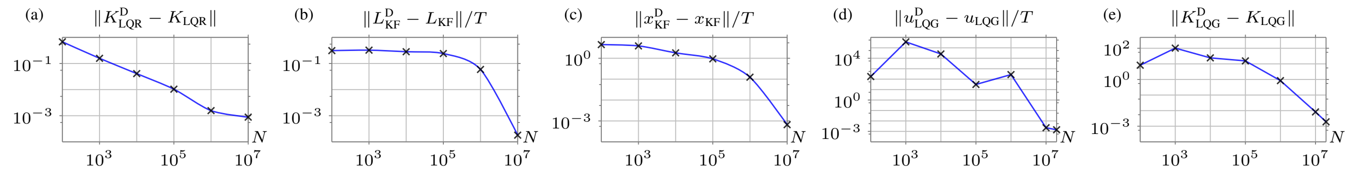

Fig. 1. This figure shows the error between the data-driven and the model-based gains as a function of the size of the data in (4) for the setting described in Example 1, 2, and 3. Panel (a) shows the error between the model-based LQR gain and the data-based LQR gain obtained from Theorem 3.1 as a function of N for the setting described in Example 1. Panel (b) shows the error between the model-based Kalman filter and the data-based Kalman filter obtained from Theorem 3.2, and panel (c) shows the error between the corresponding state estimates as a function of N for the setting in Example 2. Panel (d) shows the error between the model-based LQG input generated by (3) and the data-based LQG input generated by (11) as a function of N. Panel (e) shows the error between the model-based LQG gain in (5) and the data-based LQG gain obtained from Theorem 3.3 as a function of N for the setting in Example 3. We observe that all the quantities in the plots decrease as the number of trajectories, N, increases, which agrees with our theoretical results.

for sufficiently large N and probability at least $1-6\delta$, where

$$\kappa(x^*_{\rm m}) = \frac{\sigma_{\max}(x_{\rm m}-x^*_{\rm m})}{\sigma_{\min}(x^*_{\rm m})},$$

where the constants c_1 and c_2 are independent of N and are defined in (29), $\rho<1$, and $\delta\in[0,1/6]$. □

A proof of Theorem 3.1 is postponed to Appendix B. Some comments are in order. First, Theorem 3.1 provides a direct, data-driven way to estimate the LQR gain from noisy data, namely, $K^{\rm D}_{\rm LQR}$, and characterizes the error between the true and the estimated gains. Such error vanishes as the number (N) and length (T) of the experimental trajectories grow.[2] Further, the term $\kappa(x^*_{\rm m})$ also vanishes as the number of experimental trajectories increases (see Theorem B.3). Second, the vectors $u_{\rm v}$ and $x_{\rm v}$ contain an estimate of the optimal input and state trajectories that minimize the LQR cost with matrices Q_x and R_u for the system (1) with initial state x_0 and without process noise. Notably, these trajectories are estimated using the noisy dataset (4). Thus, this result extends the analysis in [7]. Third, Theorem 3.1 is valid when N is sufficiently large. In particular, N needs to be at least large enough to satisfy $\kappa(x^*_{\rm m})<1$ (see Appendix B for other conditions on N). Also, the result holds with probability $1-6\delta$, and the specific choice of δ affects the magnitude of the constant c_1. Fourth and finally, although formulas with similar convergence rates for the estimation of the LQR exist [19], [21], Theorem 3.1 provides an alternative, direct, closed-form expression of the gain, as opposed to indirect and optimization-based approaches. This will allow us to estimate the LQG controller.

Example 1: ***(Estimating the LQR gain from noisy data)*** Consider System (1) with

$$A=\begin{bmatrix}0.7 & 1.2\\ 0 & 0.4\end{bmatrix},\quad B=\begin{bmatrix}0\\1\end{bmatrix},\quad C=\begin{bmatrix}1 & 0\end{bmatrix},$$

$Q_x=5I_2$, $Q_w=2I_2$, $R_u=R_v=\Sigma_u=1$, and $\Sigma_0=I_2$. We collect open-loop trajectories as in (4) generated by inputs satisfying Assumption 2.1 with horizon $T{=}50$. The model-based LQR gain is $K_{\rm LQR}=[0.241 \quad 0.788]$. We use Theorem 3.1 to compute the data-driven LQR gain, $K^{\rm D}_{\rm LQR}$ for different values of N. Fig. 1(a) shows the error $\|K^{\rm D}_{\rm LQR}{-}K_{\rm LQR}\|$ as a function of the number of trajectories. □

[2]The constant c_1, as well as other constants defined later in the paper, depend also on the horizon T. While a detailed characterization of the effects of this dependency requires a dedicated analysis, notice that our expressions remain consistent if N grows sufficiently faster than T. The formulas in the paper quantify the error for finite choices of these two parameters.

We now focus on estimating the Kalman filter from noisy data with unknown system dynamics and noise statistics.

Theorem 3.2: ***(Data-driven Kalman filter)*** Let U_t and Y_t be the submatrices of U and Y in (4) obtained by selecting only the inputs and outputs up to time t. Let

$$L^{\rm D}_t = X_t\begin{bmatrix}U_{t-1}\\ Y_t\end{bmatrix}^{\dagger}, \tag{9}$$

where X_t is as in (6). Then, for every $t\in[0,T]$,

$$\left\|x_{\rm KF}(t)-L^{\rm D}_t\begin{bmatrix}u_0^{t-1}\\ y_0^t\end{bmatrix}\right\|_2 \le \frac{c_3}{\sqrt{N}}\left\|\begin{bmatrix}u_0^{t-1}\\ y_0^t\end{bmatrix}\right\|_2, \tag{10}$$

with probability at least $1-2\delta$, where u_0^t and y_0^t are the vectors of inputs and outputs of (1), respectively, from time 0 up to time t, c_3 is a constant independent of N as defined in (35), and $\delta\in[0,1/2]$. □

A proof of Theorem 3.2 is postponed to Appendix C. Theorem 3.2 provides a way to construct an approximate Kalman filter using a finite set of experimental data, without knowing the system dynamics and the statistics of the noise. As can be seen from (10), the error vanishes with rate $1/\sqrt{N}$ as the number of experimental data grows, showing the consistency of the data-driven Kalman filter expressions (9).

Example 2: ***(Estimating the Kalman filter from noisy data)*** Following the setting introduced in Example 1, we use Theorem 3.2 to obtain the data-driven Kalman filter, $L^{\rm D}_{\rm KF}$, and the corresponding data-driven state estimate, $x^{\rm D}_{\rm KF}$. Fig. 1(b) and 1(c) show the errors $\|L^{\rm D}_{\rm KF}{-}L_{\rm KF}\|$ and $\|x^{\rm D}_{\rm KF}{-}x_{\rm KF}\|$. □

Theorems 3.1 and 3.2 allow us to compute the LQG inputs from time 0 up to time T. In particular, recalling the structure of the LQG inputs due to the separation principle [1],

$$u_{\rm dLQG}(t) = K^{\rm D}_{\rm LQR}\, L^{\rm D}_t \underbrace{\begin{bmatrix}u_{\rm dLQG}(0)\\ \vdots\\ u_{\rm dLQG}(t-1)\\ y_{\rm dLQG}(0)\\ \vdots\\ y_{\rm dLQG}(t)\end{bmatrix}}_{x^{\rm D}_{\rm KF}(t)}, \tag{11}$$

where x_{KF}^{D} is the state estimate obtained using our data-driven scheme. Fig. 1(d) shows how these data-driven inputs compare to the model-based LQG inputs as a function of the amount of data. As expected, the performance gap between the data-driven and the model-based schemes shrinks as the amount of data increases. We next provide an estimate of the LQG gain (5), which allows us to compute LQG inputs beyond the horizon T of the experimental trajectories. We start by collecting $M \geq n + nm + np$ closed-loop input-output trajectories of system (1) driven by the LQG inputs generated from (11). In particular,

$$U_{\text{dLQG}} = \begin{bmatrix} u_{\text{dLQG}}^1 & \cdots & u_{\text{dLQG}}^M \end{bmatrix}, \; Y_{\text{dLQG}} = \begin{bmatrix} y_{\text{dLQG}}^1 & \cdots & y_{\text{dLQG}}^M \end{bmatrix}, \tag{12}$$

where y_{dLQG}^i is the i-th output trajectory of (1) generated by the LQG input u_{dLQG}^i in (11). That is, for $i \in \{1, \dots, M\}$,

$$u_{\text{dLQG}}^i = \begin{bmatrix} u_{\text{dLQG}}^i(0) \\ \vdots \\ u_{\text{dLQG}}^i(T-1) \end{bmatrix}, \quad y_{\text{dLQG}}^i = \begin{bmatrix} y_{\text{dLQG}}^i(0) \\ \vdots \\ y_{\text{dLQG}}^i(T) \end{bmatrix}.$$

Theorem 3.3: ***(Data-driven LQG gain)*** Let U_{dLQG}^n and Y_{dLQG}^n be the submatrices of U_{dLQG} and Y_{dLQG} in (12) obtained by selecting only the inputs from time $T-n$ up to time $T-1$ and the outputs from time $T-n+1$ up to time T, respectively. Define the data-driven LQG gain as

$$K_{\text{LQG}}^{\text{D}} = K_{\text{LQR}}^{\text{D}} L_t^{\text{D}} \underbrace{\begin{bmatrix} U_{\text{dLQG}} \\ Y_{\text{dLQG}} \end{bmatrix} \begin{bmatrix} U_{\text{dLQG}}^n \\ Y_{\text{dLQG}}^n \end{bmatrix}^{\dagger}}_{c_4},$$

Then, the data-driven estimate of the LQG gain satisfies

$$\|K_{\text{LQG}} - K_{\text{LQG}}^{\text{D}}\|_2 \leq \|c_4\|_2 \left(\frac{c_5 + c_6 \rho^T}{\sqrt{N}} + c_7 \rho^T \right), \tag{13}$$

for sufficiently large T and N and probability at least $1 - 8\delta$, where the constants c_5, c_6, and c_7 are independent of N and are defined in (38), $\rho < 1$, and $\delta \in [0, 1/8]$. □

We postpone the proof of Theorem 3.3 to Appendix D. Theorem (3.3) provides a direct data-driven expression of the LQG gain that converges with polynomial rate as the experimental data increases. To the best of our knowledge, this result is the first of its kind, and it provides a new way to compute the LQG controller using offline experimental data and a finite number of online experiments, without knowing or identifying the system and noise matrices.

Example 3: ***(Estimating the LQG gain from noisy data)*** Following the setting introduced in Example 1 and Example 2, we use Theorem 3.3 to obtain the data-driven LQG gain, $K_{\text{LQG}}^{\text{D}}$. Fig. 1(e) shows the error $\|K_{\text{LQG}}^{\text{D}} - K_{\text{LQG}}\|$ as a function of the number of trajectories, N. □

IV. Conclusion

In this paper we derive direct data-driven expressions for the LQR gain, Kalman filter, and LQG controller using a dataset of input, state, output trajectories. We show the convergence of these expressions as the size of the dataset increases, we characterize their convergence rate, and we quantify the error incurred when using a dataset of finite size. Our expressions are direct, as they do not use a model of the system nor require the estimation of a model, and provide new insights into the solution of canonical control and estimation problems. Directions of future research include the direct data-driven solution to $\mathcal{H}_2$ and $\mathcal{H}_\infty$ problems, as well as the extension of the results to accomodate for incomplete, heterogeneous and, possibly, corrupted datasets.

Appendix

A. Technical lemmas

Lemma A.1: ***(Product of Gaussian matrices [19, Lemma 1])*** Let $A = [a_1, \cdots, a_N]$ and $B = [b_1, \cdots, b_N]$, where $a_i \in \mathbb{R}^n$ and $b_i \in \mathbb{R}^m$ are independent random vectors with $a_i \sim \mathcal{N}(0, \Sigma_a)$ and $b_i \sim \mathcal{N}(0, \Sigma_b)$ for $i = 1, \cdots, N$. Let $\delta \in [0, 1]$ and $N \geq 2(n + m) \log(1/\delta)$. Then, with probability at least $1 - \delta$

$$\|AB^{\mathsf{T}}\|_2 \leq 4\|\Sigma_a\|_2^{1/2} \|\Sigma_b\|_2^{1/2} \sqrt{N(n+m)\log(9/\delta)}.$$

□

Lemma A.2: ***(Singular values of a Gaussian matrix)*** Let $\delta \in [0, 1]$, and let $A \in \mathbb{R}^{n \times N}$ be a random matrix with independent entries distributed as $\mathcal{N}(0, 1)$. Then, for $N \geq 8n + 16 \log(1/\delta)$, each of the following inequalities hold with probability probability at least $1 - \delta$

$$\sigma_{\min}(A) \geq \sqrt{N}/2, \qquad \sigma_{\max}(A) \leq 3\sqrt{N}/2,$$

where $\sigma_{\min}$ ($\sigma_{\max}$) is the smallest (largest) singular value. □

Proof: For notational convenience, we use $\sigma_{\min}$, $\sigma_{\max}$, and δ' to denote $\sigma_{\min}(A)$, $\sigma_{\max}(A)$, and $2\log(1/\delta)$, respectively. From [29, Corollary 5.35], we have each of the following inequalities holds with probability at least $1 - \delta$

$$\sigma_{\min} \geq \sqrt{N} - \sqrt{n} - \sqrt{\delta'}, \quad \sigma_{\max} \leq \sqrt{N} + \sqrt{n} + \sqrt{\delta'}. \tag{14}$$

Assume that $N \geq 8n + 8\delta'$. Then,

$$\sqrt{N}/2 \geq \sqrt{n} + \sqrt{\delta'}, \tag{15}$$

where we have used the inequality $2(a^2 + b^2) \geq (a + b)^2$. The proof follows by substituting (15) into (14). ■

B. Proof of Theorem 3.1

Let $u_{\text{v}}^* \in \mathbb{R}^{mT}$ and $x_{\text{v}}^* \in \mathbb{R}^{nT}$ be the optimal LQR trajectories of (1) from the initial state x_0. Then, $K_{\text{LQR}} = u_{\text{m}}^* x_{\text{m}}^{*\dagger}$ asymptotically as the control horizon T grows. Further, from [7], [30], the trajectories u_{v}^* and x_{v}^* can be obtained using (7) when the state data is not corrupted by the process noise. Let X_{clean} be such data, that is, the state trajectories of (1) with inputs U and noise $w(t) = 0$ at all times. Notice that in our setting X is different from X_{clean} since the process noise is nonzero when the data is collected. Because of this deviation in the data, the vectors u_{v} and x_{v} in (7) are a perturbed version of the optimal trajectories u_{v}^* and x_{v}^*. Accordingly, $K_{\text{LQR}}^{\text{D}} = u_{\text{m}} x_{\text{m}}^{\dagger}$ is a perturbed version of K_{LQR}. To quantify the deviation between $K_{\text{LQR}}^{\text{D}}$ and K_{LQR}, we quantify (i) the deviation in the data induced by the process noise, (ii) the sensitivity of the map (7) that generates LQR trajectories, and (iii) how the induced errors propagate to compute $K_{\text{LQR}}^{\text{D}}$.

(i) Data deviation induced by the process noise. Note that

$$X = \underbrace{\left[\begin{array}{c:c} O & F_u \end{array}\right]}_{F} \underbrace{\left[\begin{array}{c} X_0 \\ \hdashline \overline{U} \end{array}\right]}_{\overline{U}} + F_w W, \tag{16}$$

where $W \in \mathbb{R}^{nT \times N}$ is a matrix that contains the corresponding N process noise realizations of horizon $T-1$, and

$$O = \begin{bmatrix} I_n \\ A \\ \vdots \\ A^T \end{bmatrix}, \; F_u = \begin{bmatrix} 0 & \cdots & 0 \\ B & \cdots & 0 \\ \vdots & \ddots & \vdots \\ A^{T-1}B & \cdots & B \end{bmatrix}, \; F_w = \begin{bmatrix} 0 & \cdots & 0 \\ I_n & \cdots & 0 \\ \vdots & \ddots & \vdots \\ A^{T-1} & \cdots & I_n \end{bmatrix}.$$

Note that $X_{\text{clean}} = F\overline{U}$. Let the data matrices in (4) and (6) be partitioned as

$$U = \begin{bmatrix} U_{\text{d}} & U_{\text{n}} \end{bmatrix}, X = \begin{bmatrix} X_{\text{d}} & X_{\text{n}} \end{bmatrix}, X_0 = \begin{bmatrix} X_{0,\text{d}} & X_{0,\text{n}} \end{bmatrix}, \tag{17}$$

where U_{d}, X_{d}, and $X_{0,\text{d}}$ contain the first $N_{\text{d}} \geq mT + n$ columns of U, X, and X_0, respectively, and let $\overline{U} = [\overline{U}_{\text{d}}, \overline{U}_{\text{n}}]$ be partitioned similarly. For notational convenience, we define $Q_T = (I_{T+1} \otimes Q_x)$ and $R_T = \text{blkdiag}(0_{n\times n}, I_T \otimes R_u)$. Noting that $\overline{U}_{\text{d}}\overline{U}_{\text{d}}^{\dagger} = I_{n+mT}$, we rewrite u_{v} in (7) as

$$u_{\text{v}} = HP^{-1/2}\left(\begin{bmatrix} I_n & 0_{n\times mT} \end{bmatrix} P^{-1/2}\right)^{\dagger} x_0, \tag{18}$$

with

$$P = \left(\widetilde{X}_{\text{c}}\overline{U}_{\text{d}}^{\dagger}\right)^{\mathsf{T}} Q_T \left(\widetilde{X}_{\text{c}}\overline{U}_{\text{d}}^{\dagger}\right) + R_T, \text{ and } \widetilde{X}_{\text{c}} = X\overline{U}^{\dagger}\overline{U}_{\text{d}}. \tag{19}$$

Further, let

$$X_{\text{c}} = X_{\text{clean}}\overline{U}^{\dagger}\overline{U}_{\text{d}} \quad \text{and} \quad \Delta_X = \widetilde{X}_{\text{c}} - X_{\text{c}}. \tag{20}$$

Notice that if the process noise, W, is zero, then $\Delta_X = 0$ and $\widetilde{X}_{\text{c}} = X_{\text{c}}$ and, from (7), $u_{\text{v}} = u_{\text{v}}^*$ and $x_{\text{v}} = x_{\text{v}}^*$. Thus, we use Δ_X as a proxy for the deviation between X and X_{clean}, which is induced by the process noise, $F_w W$. The next Lemma provides a non-asymptotic upper bound to $\|\Delta_X\|_2$.

Lemma B.1: ***(Non-asymptotic bound on*** $\|\Delta_X\|_2$***)*** Let Δ_X be as in (20), and let $\delta \in [0, 1/3]$. Assume that $N > \max\{N_1, N_{\text{d}}\}$, with $N_1 = 2((n+m)T+n)\log(1/\delta)$ and $N_{\text{d}} \geq 8(mT+n) + 16\log(1/\delta)$. Then, with probability at least $1-3\delta$,

$$\|\Delta_X\|_2 \leq d_1\sqrt{\frac{N_{\text{d}}((n+m)T+n)\log(9/\delta)}{N}}, \tag{21}$$

where $d_1 = 24\|F_w\|_2\|\overline{Q}_w\|_2^{1/2}$ and $\overline{Q}_w = I_T \otimes Q_w$. □

Proof: Let $\overline{U} = \Sigma_{\overline{u}}^{1/2}Z$ and $\overline{U}_{\text{d}} = \Sigma_{\overline{u}}^{1/2}Z_{\text{d}}$, where $\Sigma_{\overline{u}} = \text{blkdiag}(\Sigma_0, I_T \otimes \Sigma_u)$, $Z \in \mathbb{R}^{n+mT\times N}$ is a random matrix whose columns are independent copies of $\mathcal{N} \sim (0, I_{n+mT})$, and Z_{d} contains the first N_{d} columns of Z. From (19), (20),

$$\begin{aligned}\|\Delta_X\|_2 &= \|F_w W\overline{U}^{\mathsf{T}}(\overline{U}\overline{U}^{\mathsf{T}})^{-1}\overline{U}_{\text{d}}\|_2 = \|F_w WZ^{\mathsf{T}}(ZZ^{\mathsf{T}})^{-1}Z_{\text{d}}\|_2 \\ &\leq \|F_w\|_2\|WZ^{\mathsf{T}}\|_2\|(ZZ^{\mathsf{T}})^{-1}\|_2\|Z_{\text{d}}\|_2.\end{aligned}$$

The proof follows by using Lemma A.1 to bound $\|WZ^{\mathsf{T}}\|_2$, Lemma A.2 to bound $\|(ZZ^{\mathsf{T}})^{-1}\|_2$ and $\|Z_{\text{d}}\|_2$, and using the union bound to compute the probability. ■

(ii) Sensitivity of map (7) w.r.t. Δ_X. We focus our analysis on the map $f : \mathbb{R}^{n(T+1)N_{\text{d}}} \times \mathbb{R}^{(n+mT)N_{\text{d}}} \to \mathbb{R}^{n+mT}$ that generates u_{v} as in (18). Then, $u_{\text{v}}^* = f(\text{vec}(X_{\text{c}}), \text{vec}(\overline{U}_{\text{d}}))$. Since f is Fréchet-differentiable with respect to $\text{vec}(X_{\text{c}})$ [30], [31], we can write its first-order Taylor-series expansion as

$$\begin{aligned} f(\text{vec}(\widetilde{X}_{\text{c}}), \text{vec}(\overline{U}_{\text{d}})) &= f(\text{vec}(X_{\text{c}}), \text{vec}(\overline{U}_{\text{d}})) \\ &\quad + \nabla f_X\left(\text{vec}(X_{\text{c}}), \text{vec}(\overline{U}_{\text{d}})\right)\text{vec}(\Delta_X),\end{aligned} \tag{22}$$

where ∇f_X is the Jacobian matrix of $f(\text{vec}(X_{\text{c}}), \text{vec}(\overline{U}_{\text{d}}))$ with respect to $\text{vec}(X_{\text{c}})$. We quantify the sensitivity of the map (18) to the change in X_{c} by ∇f_X (large values of ∇f_X implies higher sensitivity). Next, we derive an upper bound on $\|\nabla f_X\|_2$, and upper bounds on $\|u_{\text{v}} - u_{\text{v}}^*\|_2$ and $\|x_{\text{v}} - x_{\text{v}}^*\|_2$ using the first-order approximation in (22).

Lemma B.2: ***(Non-asymptotic bound on*** $\|\nabla f_X\|_2$***)*** Let $\overline{U}_{\text{d}}$, X_{c}, and $\nabla f_X\left(\text{vec}(X_{\text{c}}), \text{vec}(\overline{U}_{\text{d}})\right)$ be as in (20) and (22). Also, let $\delta \in [0,1]$ and assume that $N_{\text{d}} \geq 8(n+mT) + 16\log(1/\delta)$. Then, with probability at least $1-\delta$,

$$\left\|\nabla f_X\left(\text{vec}(X_{\text{c}}), \text{vec}(\overline{U}_{\text{d}})\right)\right\|_2 \leq 4d_2\sqrt{\frac{n(T+1)}{N_{\text{d}}}}, \tag{23}$$

where $d_2 > 0$ is independent of N_{d}. □

Proof: The proof can be adapted from the proof of [30, Lemma IV.4], then using Lemma A.2 and $\|\cdot\|_2 \leq \|\cdot\|_{\text{F}}$. ■

Theorem B.3: ***(Non-asymptotic bound on the deviation of the LQR trajectories)*** Let u_{v} and x_{v} be as in (7) and u_{v}^* and x_{v}^* be the optimal LQR trajectories of length T of (1) from the initial state x_0. Let $\delta \in [0, 1/6]$ and assume that $N \geq \max\{N_1, N_2, N_3\}$, with $N_1 = 2((n+m)T+n)\log(1/\delta)$, $N_2 = 8(mT+n) + 16\log(1/\delta)$, and $N_3 = ((n+m)T+n)\log(9/\delta)$. Then, with probability at least $1-4\delta$,

$$\|u_{\text{v}} - u_{\text{v}}^*\|_2 \leq d_3\sqrt{\frac{((n+m)T+n)\log(9/\delta)}{N}}. \tag{24}$$

Further, with probability at least $1-6\delta$,

$$\|x_{\text{v}} - x_{\text{v}}^*\|_2 \leq d_4\sqrt{\frac{((n+m)T+n)\log(9/\delta)}{N}}, \tag{25}$$

with

$$\begin{aligned} d_3 &= 4d_1d_2\sqrt{qn(T+1)}, \\ d_4 &= \|F\|_2 d_3 + 16\|F_w\|_2\|\Sigma_{\overline{u}}^{-1/2}\|_2\|\overline{Q}_w\|_2^{1/2}\left(\|\overline{u}_{\text{v}}^*\|_2 + d_3\right),\end{aligned}$$

where d_1, F, and F_w are as in (21) and (16), respectively, $d_2 > 0$ is independent of N, $q = \text{Rank}(\Delta_X) \leq n(T+1)$, $\overline{u}_{\text{v}}^* = [x_0^{\mathsf{T}}, u_{\text{v}}^{*\mathsf{T}}]^{\mathsf{T}}$, and $\overline{Q}_w$ and $\Sigma_{\overline{u}}$ are as in Lemma B.1. □

Proof: Inequality (24) follows from (22) by using Lemma B.1, Lemma B.2, and $\|\text{vec}(\Delta_X)\|_2 = \|\Delta_X\|_{\text{F}} \leq \sqrt{q}\|\Delta_X\|_2$, with $q = \text{Rank}(\Delta_X)$. Next, we derive (25). For notational convenience, we use Δ_u and Δ_x to denote $u_{\text{v}} - u_{\text{v}}^*$ and

$x_{\rm v}-x_{\rm v}^*$, respectively. From (7), we can write

$$\begin{aligned}\|\Delta_x\|_2 &= \left\|\widetilde{X}_{\rm c}\overline{U}_{\rm d}^\dagger\begin{bmatrix}x_0\\u_{\rm v}\end{bmatrix} - X_{\rm c}\overline{U}_{\rm d}^\dagger\begin{bmatrix}x_0\\u_{\rm v}^*\end{bmatrix}\right\|_2 \\ &= \left\|X_{\rm c}\overline{U}_{\rm d}^\dagger\begin{bmatrix}0\\\Delta_u\end{bmatrix} + \Delta_X\overline{U}_{\rm d}^\dagger\begin{bmatrix}x_0\\u_{\rm v}^*\end{bmatrix} + \Delta_X\overline{U}_{\rm d}^\dagger\begin{bmatrix}0\\\Delta_u\end{bmatrix}\right\|_2 \\ &\le \|X_{\rm c}\overline{U}_{\rm d}^\dagger\|_2\|\Delta_u\|_2 + \|\Delta_X\overline{U}_{\rm d}^\dagger\|_2\|\bar{u}_{\rm v}^*\|_2 + \|\Delta_X\overline{U}_{\rm d}^\dagger\|_2\|\Delta_u\|_2.\end{aligned} \tag{26}$$

Note that $\overline{U}_{\rm d}\overline{U}_{\rm d}^\dagger = I_{n+mT}$. Then we have

$$\begin{aligned}\|X_{\rm c}\overline{U}_{\rm d}^\dagger\|_2 &= \|X_{\rm clean}\overline{U}^\dagger\|_2 = \|F\|_2, \\ \|\Delta_X\overline{U}_{\rm d}^\dagger\|_2 &= \|(X-X_{\rm clean})\overline{U}^\dagger\|_2 = \|F_wW\overline{U}^\dagger\|_2 \\ &\le \|F_w\|_2\|W\overline{U}^\mathsf{T}\|_2\|(\overline{U}\,\overline{U}^\mathsf{T})^{-1}\|_2,\end{aligned}$$

Inequality (25) follows from (26) by using (24), Lemma A.1, and Lemma A.2 to bound $\|\Delta_u\|_2$, $\|W\overline{U}^\mathsf{T}\|_2$, and $\|(\overline{U}\,\overline{U}^\mathsf{T})^{-1}\|_2$, respectively, and noting that for $N\ge N_3$ we have $\frac{\delta'}{N}\le\sqrt{\frac{\delta'}{N}}$, with $\delta'=((n+m)T+n)\log(9/\delta)$. The probabilities follow from the union bound. ■

(iii) Error between $K_{\rm LQR}$ and K_{LQR}^D. We are now ready to conclude the proof of Theorem 3.1. Notice that

$$u_{\rm m}^* = K_{\rm LQR}x_{\rm m}^*, \text{ and } u_{\rm m}^*+\delta_u = K_{\rm LQR}^{\rm D}(x_{\rm m}^*+\delta_x), \tag{27}$$

where $\delta_u = u_{\rm m}-u_{\rm m}^*$ and $\delta_x = x_{\rm m}-x_{\rm m}^*$. Note that $u_{\rm m}$ and $x_{\rm m}$ are the matrices obtained by reorganizing the inputs and states in the vectors $u_{\rm v}$ and $x_{\rm v}$ in chronological order. For notational convenience, we use K and $K^{\rm D}$ to denote $K_{\rm LQR}$ and $K_{\rm LQR}^{\rm D}$. Let $\Delta_K = K-K^{\rm D}$. In what follows, subscript i denotes the i-th row, with $i\in\{1,\cdots,m\}$. Using [32, Theorem 5.1] and assuming that $x_{\rm m}^*$ is of full row rank,[3]

$$\|\Delta_{K,i}\|_2 \le d_5\left(\epsilon\|K_i\|_2\|x_{\rm m}^*\|_2 + \|\delta_{u,i}\|_2 + \epsilon\alpha\|r\|_2\right), \tag{28}$$

where

$$d_5 = \frac{\alpha}{1-\alpha\epsilon\|x_{\rm m}^*\|_2}, \quad \epsilon = \frac{\|\delta_x\|_2}{\|x_{\rm m}^*\|_2}, \quad r = u_{{\rm m},i}^* - K_ix_{\rm m}^*,$$

and $\alpha = \|x_{\rm m}^*\|_2\|x_{\rm m}^{*\dagger}\|_2$ is the spectral condition number of $x_{\rm m}^*$. From [7, Theorem 3.2], we have $\|r\|_2\le d_6\rho^T$, where $d_6>0$ and $\rho<1$, which are independent of N. Since $\|x_{\rm m}^*\|_2 = \sigma_{\max}(x_{\rm m}^*)$, $\|(x_{\rm m}^*)^\dagger\|_2 = 1/\sigma_{\min}(x_{\rm m}^*)$. Then, we can write d_5 as

$$d_5 = \frac{1}{\sigma_{\min}(x_{\rm m}^*)\,(1-\kappa(x_{\rm m}^*))}, \quad \text{with} \quad \kappa(x_{\rm m}^*) = \frac{\sigma_{\max}(\delta_x)}{\sigma_{\min}(x_{\rm m}^*)},$$

For sufficiently large N such that $\sigma_{\max}(\Delta X)<\sigma_{\min}(X^*)$, we have $\kappa(x_{\rm m}^*)<1$ and $\epsilon\alpha<1$. Then, we can write (28) as

$$\begin{aligned}\|\Delta_{K,i}\|_2 &\le d_5\Big(\|K_i\|_2\|\delta_x\|_2 + \|\delta_{u,i}\|_2 + d_6\rho^T\Big), \\ &\overset{\text{(a)}}{\le} d_5\Big(\|K\|_2\|\Delta_x\|_2 + \|\Delta_u\|_2 + d_6\rho^T\Big),\end{aligned}$$

where in step (a), we have used $\|\delta_x\|_2 \le \|\delta_x\|_{\rm F} = \|{\rm vec}(\delta_x)\|_2 = \|\Delta_x\|_2$, and $\|\delta_{u,i}\|_2 = \|\delta_{u,i}\|_{\rm F} \le \|\delta_u\|_{\rm F} = \|{\rm vec}(\delta_u)\|_2 = \|\Delta_u\|_2$, where Δ_x and Δ_u are as in Theorem B.3. Noting that $\|\Delta K\|_{\rm F} = \sqrt{{\rm tr}\left[\Delta K(\Delta K)^\mathsf{T}\right]} = \sqrt{\sum_{i=1}^m\|\Delta K_i\|_2^2}$ and using the bounds in Theorem B.3, we have with probability at least $1-6\delta$

$$\|\Delta_K\|_2 \le \frac{1}{\sigma_{\min}(x_{\rm m}^*)\,(1-\kappa(x_{\rm m}^*))}\left(\frac{c_1}{\sqrt{N}}+c_2\rho^T\right),$$

where,

$$\begin{aligned}c_1 &= (d_3+\|K_{\rm LQR}\|_2d_4)\sqrt{m\left((n+m)T+n\right)\log(9/\delta)}, \\ c_2 &= d_6\sqrt{m},\end{aligned} \tag{29}$$

and d_3 and d_4 are as in Theorem B.3. Finally, the probability follows from the union bound. This concludes the proof.

C. Proof of Theorem 3.2

The Kalman filter computes the estimate $x_{\rm KF}(t)$ given $\{u(0),\dots,u(t-1),y(0),\dots,y(t)\}$ that minimizes the cost

$$\sum_{t=0}^T \mathbb{E}\left[(x(t)-x_{\rm KF}(t))^T(x(t)-x_{\rm KF}(t))\right], \tag{30}$$

which is then used to generate LQG inputs. Equivalently, $x_{\rm KF}(t)$ can be obtained with the following linear estimator,

$$\begin{aligned}x_{\rm KF}(t) &= \underbrace{\begin{bmatrix}L_{t,0}^u & \cdots & L_{t,t-1}^u\end{bmatrix}}_{L_t^u}\underbrace{\begin{bmatrix}u(0)\\\vdots\\u(t-1)\end{bmatrix}}_{u_0^{t-1}} + \underbrace{\begin{bmatrix}L_{t,0}^y & \cdots & L_{t,t}^y\end{bmatrix}}_{L_t^y}\underbrace{\begin{bmatrix}y(0)\\\vdots\\y(t)\end{bmatrix}}_{y_0^t}, \\ &= \underbrace{\begin{bmatrix}L_t^u & L_t^y\end{bmatrix}}_{L_t^{\rm KF}}\underbrace{\begin{bmatrix}u_0^{t-1}\\y_0^t\end{bmatrix}}_{z_t},\end{aligned}$$

where $L_t^{\rm KF}\in\mathbb{R}^{n\times mt+p(t+1)}$, with $L_t^u\in\mathbb{R}^{n\times mt}$ and $L_t^y\in\mathbb{R}^{n\times p(t+1)}$, is the estimator gain that minimizes (30). Let $e(t) = x(t)-x_{\rm KF}(t)$ and $\Sigma_{e,t}\in\mathbb{R}^{n\times n}\succeq 0$ denote the estimation error and the estimation error covariance matrix, respectively. For an optimal linear estimator, $L_t^{\rm KF}$, we have $e(t)\sim\mathcal{N}(0,\Sigma_{e,t})$, and we can write the state $x(t)$ as

$$x(t) = L_t^{\rm KF}z_t + e(t).$$

Let

$$x_t = [x^1(t),\dots,x^N(0)], \quad e_t = [e^1(t),\dots,e^N(0)], \tag{31}$$

where $x^i(t)$ and $e^i(t)$ denote the state and the state estimation error incurred by $L_t^{\rm KF}$ at time t for the i-th trajectory of the data (4), respectively. Further, let $Z_t = [U_{t-1}^\mathsf{T}, Y_t^\mathsf{T}]^\mathsf{T}$, where U_t and Y_t are the submatrices of U and Y in (4) obtained by selecting the inputs and outputs up to some t. Then,

$$x_t = L_t^{\rm KF}Z_t + e_t. \tag{32}$$

To estimate the optimal filter L_t from the data (4), we consider the following least squares problem

$$L_t^{\rm D} = \arg\min_{L_t}\|x_t - L_tZ_t\|_{\rm F}^2. \tag{33}$$

Problem (33) admits a unique solution since Z_t is full-row rank, which is given by (9). Next, we bound $\|L_t^{\rm D}-L_t^{\rm KF}\|_2$.

Theorem C.1: ***(Non-asymptotic bound on $\|L_t^D - L_t^{KF}\|_2$)*** Let $L_t^{\rm KF}$ and $L_t^{\rm D}$ be as in (32) and (9), respectively, and

[3] This condition is typically satisfied for generic choices of the initial state.

let $\delta \in [0,1/2]$. Assume that $N \geq \max\{N_1, N_2\}$, with $N_1 = 2((n+m)T+n)\log(1/\delta)$ and $N_2 = 8(mT+n) + 16\log(1/\delta)$. Then, with probability at least $1-2\delta$,

$$\|L_t^{\mathsf{D}} - L_t^{\mathsf{KF}}\|_2 \leq d_7 \sqrt{\frac{((m+p)t+n+p)\log(9/\delta)}{N}}, \quad (34)$$

with

$$d_7 \triangleq 16\|\Sigma_Z^{-1/2}\|_2 \|\Sigma_{e,t}\|_2^{1/2},$$

where $\Sigma_Z \in \mathbb{R}^{(m+p)t+p\times(m+p)t+p} \succ 0$ comprises the noise statistics, Σ_0, and Σ_u in Assumption 2.1, and $\Sigma_{e,t}$ is the optimal estimation error covariance matrix at time t. □

Proof: Let $Z = \Sigma_Z^{1/2} G$ and where $\Sigma_Z \succ 0$ is as in the theorem statement, and $G \in \mathbb{R}^{(m+p)t+p\times N}$ is a random matrix whose columns are independent random vectors distributed as $\mathcal{N} \sim (0, I_{(m+p)t+p})$. From (9) and (32),

$$\begin{aligned}\|L_t^{\mathsf{D}} - L_t^{\mathsf{KF}}\|_2 &= \|e_t Z^\dagger\|_2 = \|e_t G^\mathsf{T}(GG^\mathsf{T})^{-1}\Sigma_Z^{-1/2}\|_2 \\ &\leq \|\Sigma_Z^{-1/2}\|_2 \|e_t G^\mathsf{T}\|_2 \|(GG^\mathsf{T})^{-1}\|_2.\end{aligned}$$

The proof follows by using Lemma A.1 to bound $\|e_t G^\mathsf{T}\|_2$, and Lemma A.2 to bound $\|(GG^\mathsf{T})^{-1}\|_2$. Finally, The probability follows from the union bound. ■

To conclude the proof of Theorem 3.2, we have

$$\begin{aligned}\left\| x_{\mathrm{KF}}(t) - L_t^{\mathsf{D}} \begin{bmatrix} u_0^{t-1} \\ y_0^t \end{bmatrix} \right\|_2 &= \left\| L_t^{\mathsf{KF}} \begin{bmatrix} u_0^{t-1} \\ y_0^t \end{bmatrix} - L_t^{\mathsf{D}} \begin{bmatrix} u_0^{t-1} \\ y_0^t \end{bmatrix} \right\|_2 \\ &\leq \|L_t^{\mathsf{KF}} - L_t^{\mathsf{D}}\|_2 \left\| \begin{bmatrix} u_0^{t-1} \\ y_0^t \end{bmatrix} \right\|_2,\end{aligned}$$

where u_0^t and y_0^t are the vectors of inputs and outputs of (1), respectively, from time 0 up to time t. Using Theorem C.1,

$$\left\| x_{\mathrm{KF}}(t) - L_t^{\mathsf{D}} \begin{bmatrix} u_0^{t-1} \\ y_0^t \end{bmatrix} \right\|_2 \leq \frac{c_3}{\sqrt{N}} \left\| \begin{bmatrix} u_0^{t-1} \\ y_0^t \end{bmatrix} \right\|_2, \quad (35)$$

where $c_3 = d_7\sqrt{((m+p)t+n+p)\log(9/\delta)}$, and d_7 is as in Theorem C.1. The above inequality holds with probability at least $1-2\delta$, which follows from Theorem C.1 for $\delta \in [0,1/2]$. This concludes the proof of Theorem 3.2.

D. Proof of Theorem 3.3

Consider the closed-loop trajectories in (12), and let U_{dLQG}^n and Y_{dLQG}^n be the submatrices of U_{dLQG} and Y_{dLQG} in (12) obtained by selecting only the inputs from time $T-n$ up to time $T-1$ and the outputs from time $T-n+1$ up to time T, respectively. We can write the data-based and the model-based LQG inputs at time T for the trajectories in (12) as

$$\underbrace{\begin{bmatrix} u_{\mathrm{dLQG}}^1(T) & \cdots & u_{\mathrm{dLQG}}^M(T) \end{bmatrix}}_{U_{\mathrm{dLQG}}(T)} = K_{\mathrm{LQR}}^{\mathsf{D}} L_T^{\mathsf{D}} \begin{bmatrix} U_{\mathrm{dLQG}} \\ Y_{\mathrm{dLQG}} \end{bmatrix},$$

$$\underbrace{\begin{bmatrix} u_{\mathrm{LQG}}^1(T) & \cdots & u_{\mathrm{LQG}}^M(T) \end{bmatrix}}_{U_{\mathrm{LQG}}(T)} = K_{\mathrm{LQR}} L_T^{\mathsf{KF}} \underbrace{\begin{bmatrix} U_{\mathrm{dLQG}} \\ Y_{\mathrm{dLQG}} \end{bmatrix}}_{Z}.$$

For notational convenience, let $\Delta K_{\mathrm{LQR}} = K_{\mathrm{LQR}}^{\mathsf{D}} - K_{\mathrm{LQR}}$, $\Delta L = L_T^{\mathsf{D}} - L_T^{\mathsf{KF}}$, and $\Delta U = U_{\mathrm{dLQG}}(T) - U_{\mathrm{LQG}}(T)$. Then,

$$\begin{aligned}\Delta U &= K_{\mathrm{LQR}}^{\mathsf{D}} L_T^{\mathsf{D}} Z - K_{\mathrm{LQR}} L_T^{\mathsf{KF}} Z \\ &= K_{\mathrm{LQR}} \Delta L Z + \Delta K_{\mathrm{LQR}} L_T^{\mathsf{KF}} Z + \Delta K_{\mathrm{LQR}} \Delta L Z.\end{aligned} \quad (36)$$

For sufficiently large T, we use (5) to write

$$U_{\mathrm{dLQG}}(T) = K_{\mathrm{LQG}}^{\mathsf{D}} \begin{bmatrix} U_{\mathrm{dLQG}}^n \\ Y_{\mathrm{dLQG}}^n \end{bmatrix}, U_{\mathrm{LQG}}(T) = K_{\mathrm{LQG}} \underbrace{\begin{bmatrix} U_{\mathrm{dLQG}}^n \\ Y_{\mathrm{dLQG}}^n \end{bmatrix}}_{Z_n}.$$

Then, $K_{\mathrm{LQG}}^{\mathsf{D}} = U_{\mathrm{dLQG}}(T) Z_n^\dagger$ and $K_{\mathrm{LQG}} = U_{\mathrm{LQG}}(T) Z_n^\dagger$. For notational convenience, let $\Delta K_{\mathrm{LQG}} = K_{\mathrm{LQG}}^{\mathsf{D}} - K_{\mathrm{LQG}}$, and let $\|\cdot\|$ denote $\|\cdot\|_2$. Then, using (36), we can write

$$\begin{aligned}\|\Delta K_{\mathrm{LQG}}\|_2 &= \|(U_{\mathrm{dLQG}}(T) - U_{\mathrm{LQG}}(T)) Z_n^\dagger\| = \|\Delta U Z_n^\dagger\| \\ &\leq \|K_{\mathrm{LQR}}\| \|\Delta L\| \|Z Z_n^\dagger\| + \|\Delta K_{\mathrm{LQR}}\| \|L_T^{\mathsf{KF}}\| \|Z Z_n^\dagger\| \\ &\quad + \|\Delta K_{\mathrm{LQR}}\| \|\Delta L\| \|Z Z_n^\dagger\|.\end{aligned} \quad (37)$$

Let $\delta \in [0,1/8]$ and assume that $N \geq \max\{N_1, N_2, N_3\}$, where N_1, N_2, and N_3 are as in Theorem B.3. Then, inequality (13) follows by using Theorem 3.1 and Theorem C.1 to bound $\|\Delta K_{\mathrm{LQR}}\|$ and $\|\Delta L\|$ in (37), respectively, with probability at least $1-8\delta$ and with

$$\begin{aligned} c_5 &= \frac{c_1\|L_t^{\mathsf{KF}}\| + c_1 c_3}{\sigma_{\min}(x_{\mathrm{m}}^*)\,(1-\kappa(x_{\mathrm{m}}^*))} + c_3\|K_{\mathrm{LQR}}\|, \\ c_6 &= \frac{c_2 c_3}{\sigma_{\min}(x_{\mathrm{m}}^*)(1-\kappa(x_{\mathrm{m}}^*))},\ c_7 = \frac{c_2\|L_t^{\mathsf{KF}}\|}{\sigma_{\min}(x_{\mathrm{m}}^*)(1-\kappa(x_{\mathrm{m}}^*))},\end{aligned} \quad (38)$$

where c_1, c_2, x_{m}^*, and $\kappa(x_{\mathrm{m}}^*)$ are as in Theorem 3.1, and c_3 is as in Theorem 3.2. Finally, the probability follows using the union bound. This concludes the proof of Theorem 3.3.

REFERENCES

[1] K. Zhou, J. C. Doyle, and K. Glover. *Robust and Optimal Control*. Prentice Hall, 1996.

[2] B. Recht. A tour of reinforcement learning: The view from continuous control. *Annual Review of Control, Robotics, and Autonomous Systems*, 2:253–279, 2019.

[3] G. Baggio, D. S. Bassett, and F. Pasqualetti. Data-driven control of complex networks. *Nature Communications*, 12(1429), 2021.

[4] M. Gevers. Identification for control: From the early achievements to the revival of experiment design. *European Journal of Control*, 11:1–18, 2005.

[5] B. Hu, K. Zhang, N. Li, M. Mesbahi, M. Fazel, and T. Başar. Towards a theoretical foundation of policy optimization for learning control policies. *Annual Review of Control, Robotics, and Autonomous Systems*, 6(1):123–158, 2023.

[6] F. Dörfler, P. Tesi, and C. De Persis. On the role of regularization in direct data-driven LQR control. In *IEEE Conf. on Decision and Control*, pages 1091–1098, Cancún, Mexico, December 2022. IEEE.

[7] F. Celi, G. Baggio, and F. Pasqualetti. Closed-form estimates of the LQR gain from finite data. In *IEEE Conf. on Decision and Control*, pages 4016–4021, Cancún, Mexico, December 2022.

[8] F. Celi and F. Pasqualetti. Data-driven meets geometric control: Zero dynamics, subspace stabilization, and malicious attacks. *IEEE Control Systems Letters*, 6:2569–2574, 2022.

[9] I. Markovsky and P. Rapisarda. On the linear quadratic data-driven control. In *European Control Conference*, pages 5313–5318. IEEE, 2007.

[10] G. R. Gonçalves da Silva, A. S. Bazanella, C. Lorenzini, and L. Campestrini. Data-driven LQR control design. *IEEE Control Systems Letters*, 3(1):180–185, 2019.

[11] M. Rotulo, C. De Persis, and P. Tesi. Data-driven linear quadratic regulation via semidefinite programming. *IFAC-PapersOnLine*, 53(2):3995–4000, 2020.

[12] C. Y. Chang and A. Bernstein. Robust data-driven control for systems with noisy data. *arXiv preprint arXiv:2207.09587*, 2022.

[13] C. De Persis and P. Tesi. Low-complexity learning of linear quadratic regulators from noisy data. *Automatica*, 128:109548, 2021.

[14] B. Gravell, P. M. Esfahani, and T. Summers. Learning optimal controllers for linear systems with multiplicative noise via policy gradient. *IEEE Transactions on Automatic Control*, 66(11):5283–5298, 2020.

[15] X. Zhang, B. Hu, and T. Başar. Learning the Kalman filter with fine-grained sample complexity. *arXiv preprint arXiv:2301.12624*, 2023.

[16] T. Guo, A. A. Al Makdah, V. Krishnan, and F. Pasqualetti. Imitation and transfer learning for LQG control. *IEEE Control Systems Letters*, 7:2149–2154, 2023.

[17] W. Favoreel, B. D. Moor, P. V. Overschee, and M. Gevers. Model-free subspace-based LQG-design. In *American Control Conference*, volume 5, pages 3372–3376, San Diego, CA, Jun. 1999.

[18] R. E. Skelton and G. Shi. The data-based LQG control problem. In *IEEE Conf. on Decision and Control*, volume 2, pages 1447–1452, Lake Buena Vista, FL, December 1994.

[19] S. Dean, H. Mania, N. Matni, B. Recht, and S. Tu. On the sample complexity of the linear quadratic regulator. *Foundations of Computational Mathematics*, 20(4):633–679, 2020.

[20] Y. Zheng, L. Furieri, M. Kamgarpour, and N. Li. Sample complexity of linear quadratic gaussian (LQG) control for output feedback systems. In *Learning for Dynamics and Control*, volume 144 of *Proceedings of Machine Learning Research*, pages 559–570, Virtual, Jun. 2021.

[21] H. Mania, S. Tu, and B. Recht. Certainty equivalence is efficient for linear quadratic control. In *Advances in Neural Information Processing Systems*, volume 32, pages 10154–10164, Vancouver, Canada, dec 2019. Curran Associates, Inc.

[22] S. Oymak and N. Ozay. Revisiting Ho–Kalman-based system identification: Robustness and finite-sample analysis. *IEEE Transactions on Automatic Control*, 67(4):1914–1928, 2022.

[23] A. Tsiamis and G. J. Pappas. Finite sample analysis of stochastic system identification. In *IEEE Conf. on Decision and Control*, pages 3648–3654, Nice, France, dec 2019.

[24] Y. Zheng and N. Li. Non-asymptotic identification of linear dynamical systems using multiple trajectories. *IEEE Control Systems Letters*, 5(5):1693–1698, 2021.

[25] S. Tu, R. Frostig, and M. Soltanolkotabi. Learning from many trajectories. *arXiv preprint arXiv:2203.17193*, 2023.

[26] H. Mohammadi, A. Zare, M. Soltanolkotabi, and M. R. Jovanović. Convergence and sample complexity of gradient methods for the model-free linear quadratic regulator problem. *IEEE Transactions on Automatic Control*, pages 1–1, 2021.

[27] A. A. Al Makdah, V. Krishnan, V. Katewa, and F. Pasqualetti. Behavioral feedback for optimal LQG control. In *IEEE Conf. on Decision and Control*, pages 4660–4666, Cancún, Mexico, December 2022.

[28] A. Tsiamis, I. Ziemann, N. Matni, and G. J. Pappas. Statistical learning theory for control: A finite sample perspective. *arXiv preprint arXiv:2209.05423*, 2022.

[29] R. Vershynin. Introduction to the non-asymptotic analysis of random matrices. *arXiv preprint arXiv:1011.3027*, 2010.

[30] F. Celi, G. Baggio, and F. Pasqualetti. Distributed data-driven control of network systems. *IEEE Open Journal of Control Systems*, pages 93–107, 2023.

[31] T. Kollo and D. von Rosen. *Advanced Multivariate Statistics with Matrices*. Mathematics and Its Applications. Springer, Berlin, 2005.

[32] P. Å. Wedin. Perturbation theory for pseudo-inverses. *BIT Numerical Mathematics*, 13:217–232, 1973.