Behavioral Feedback for Optimal LQG Control

Abed AlRahman Al Makdah, Vishaal Krishnan, Vaibhav Katewa, and Fabio Pasqualetti

Abstract— In this work, we revisit the Linear Quadratic Gaussian (LQG) optimal control problem from a behavioral perspective. Motivated by the suitability of behavioral models for data-driven control, we begin with a reformulation of the LQG problem in the space of input-output behaviors and obtain a complete characterization of the optimal solutions. In particular, we show that the optimal LQG controller can be expressed as a static behavioral-feedback gain, thereby eliminating the need for dynamic state estimation characteristic of state space methods. The static form of the optimal LQG gain also makes it amenable to its computation by gradient descent, which we investigate via numerical experiments. Furthermore, we highlight the advantage of this approach in the data-driven control setting of learning the optimal LQG controller from expert demonstrations.

I. INTRODUCTION

Data-driven control has received increasing interest during the past few years. Specifically, this interest has been surging towards optimal control problems [1]–[4]. The Linear Quadratic Gaussian (LQG) is one of the most fundamental optimal control problems, which deals with partially-observed linear dynamical systems in the presence of additive white gaussian noise [5]. When the system is known, the LQG problem enjoys an elegant closed-form solution obtained via the separation principle [5, Theorem 14.7]. In the context of data-driven control, however, the LQG problem is less studied in the literature due to some major challenges: (i) the states of the system cannot be directly measured for learning purposes, (ii) the optimal policy is expressed in the dynamic controller form where it is not unique [5], and (iii) the set of stabilizing controllers can be disconnected [6]. On the other hand, the Linear Quadratic Regulator (LQR) optimal control problem has received more attention in the context of data-driven control [1], [3], [4]. Some of the reasons that make the LQR problem attractive is that the optimal policy can be expressed as a static feedback gain and it is unique [5, Theorem 14.2]. Moreover, the set of stabilizing feedback gains for the LQR problem is connected and the LQR cost function is gradient dominant [7], [8]. These properties are useful for providing convergence guarantees for gradient-based methods for solving the LQR problem [9], [10] as well as for model-free policy optimization methods

[11]. However, LQR controllers require measuring the full states, and are used in deterministic settings, which limits the use of LQR controllers in practical control applications. In this paper, we make the LQG problem more accessible for data-driven methods. In particular, we show that the optimal LQG controller can be expressed as a static feedback gain by reformulation of the model-based LQG problem in the space of input-output behaviors. Then, we highlight the advantages of having a static LQG gain in the context of data-driven control and gradient-based algorithms.

Related work. The LQG control problem has been studied extensively in the literature [5], [12], where fundamental properties have been characterized, such as the existence of optimal solution, how to obtain it using separation principle [5], and its lack of stability margin guarantees in closed-loop [13]. However, in the context of data-driven control, the LQR problem has received more attention than the LQG problem. The landscape properties for the LQR problem with state-feedback control has been studied in [7], [8], which has paved the way for subsequent works investigating convergence properties of gradient methods for solving the LQR problem [9]–[11]. Recent studies have revisited the LQG problem in the context of data-driven methods (e.g. [14]–[16]). In [6], the authors characterize the optimization landscape for the LQG problem, showing that the set of stabilizing dynamic controllers can be disconnected. In the context of data-driven control, the behavioral approach has garnered much attention in recent years [17]–[20], as it circumvents the need for state space representation. Owing it this fact, it belongs in the same category as the difference operator representation and ARMAX models [21, Sec. 2.3 and Sec. 7.4], and shares several connections with these classes of models. We refer the reader to [22] for a comprehensive overview of the behavioral approach.

Despite recent interest in the behavioral approach, a fundamental understanding of the LQG problem from a behavioral perspective is still lacking, and our work addresses this gap. Different from the literature, our work seeks to characterize the optimal behavioral feedback controllers for the LQG problem in model-based setting, and to demonstrate their suitability for data-driven control and gradient-based methods for controller design. More specifically, we show that the optimal LQG controller can be expressed as a static behavioral-feedback gain, which underlies its advantages for developing data-driven methods to learn LQG controllers.

Contributions. This paper features three main contributions. First, we introduce equivalent representations for stochastic discrete-time, linear, time-invariant systems and the LQG optimal control problem in the behavioral space

(Lemma 5.2 and Lemma 5.4, respectively). Second, we show that, in the behavioral space, the optimal LQG controller can be expressed as a static behavioral-feedback gain, which can be solved for directly from the LQG problem represented in the behavioral space (Theorem 2.1). Third, we highlight the advantages of having a static feedback LQG gain over a dynamic LQG controller in the context of data-driven control and gradient-based algorithms (section III).

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. The expectation operator is denoted by $\mathbb{E}[\cdot]$. The spectral radius and the trace of a square matrix A are denoted by $\rho(A)$ and $\mathrm{tr}[A]$, respectively. A positive definite (semidefinite) matrix A is denoted as $A \succ 0$ ($A \succeq 0$). The Kronecker product is denoted by $\otimes$, and vectorization operator is denoted by $\mathrm{vec}(\cdot)$. The left (right) pseudo inverse of a tall (fat) matrix A is denoted by $A^\dagger$.

II. PROBLEM SETUP AND MAIN RESULTS

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, $w(t)$ the process noise, and $v(t)$ the measurement noise at time t. We assume that $w(t) \sim \mathcal{N}(0, Q_w)$, with $Q_w \succeq 0$, $v(t) \sim \mathcal{N}(0, R_v)$, with $R_v \succ 0$, and $x(0) \sim \mathcal{N}(0, \Sigma_0)$, with $\Sigma_0 \succeq 0$, are independent of each other at all times. For the system (1), the Linear Quadratic Gaussian (LQG) problem asks to find a control input that minimizes the cost

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

where $Q_x \succeq 0$ and $R_u \succ 0$ are weighing matrices of appropriate dimension. We assume that (A, B) and $(A, Q_w^{1/2})$ are controllable, and (A, C) and $(A, Q_x^{1/2})$ are observable. As a classic result [5], the optimal control input that solves the LQG problem can be generated by a dynamic controller of the form

$$
\begin{aligned}
x_c(t+1) &= Ex_c(t) + Fy(t), \quad t \geq 0, \\
u(t) &= Gx_c(t) + Hy(t),
\end{aligned} \tag{3}
$$

where $x_c(t)$ denotes the state at time t, and $E \in \mathbb{R}^{n \times n}$, $F \in \mathbb{R}^{n \times p}$, $G \in \mathbb{R}^{m \times n}$, and $H \in \mathbb{R}^{m \times p}$ are the dynamic, input, output and feedthrough matrices of the compensator, respectively. The optimal LQG controller can be conveniently obtained using the separation principle by concatenating the Kalman filter for (1) with the (static) controller that solves the Linear Quadratic Regulator problem for (1) with weight matrices Q_x and R_u. Specifically, after some manipulation, the optimal input that solves the LQG problem reads as (3), we refer the reader to Appendix A for the details.

In what follows, we will make use of an equivalent representation of the system (1). To this aim, let

$$
z(t) \triangleq [U(t-1)^\mathsf{T}, Y(t)^\mathsf{T}, W(t-1)^\mathsf{T}, V(t)^\mathsf{T}]^\mathsf{T}, \tag{4}
$$

where

$$
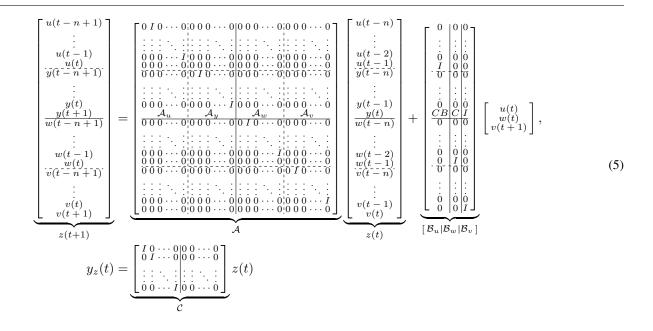\begin{aligned}
U(t-1) &\triangleq \left[u(t-n)^\mathsf{T}, \cdots, u(t-1)^\mathsf{T} \right]^\mathsf{T}, \\
Y(t) &\triangleq \left[y(t-n)^\mathsf{T}, \cdots, y(t)^\mathsf{T} \right]^\mathsf{T}, \\
W(t-1) &\triangleq \left[w(t-n)^\mathsf{T}, \cdots, w(t-1)^\mathsf{T} \right]^\mathsf{T}, \\
V(t) &\triangleq \left[v(t-n)^\mathsf{T}, \cdots, v(t)^\mathsf{T} \right]^\mathsf{T}.
\end{aligned}
$$

We can write an equivalent representation of (1) in the behavioral space z as (5) (see Appendix B for the derivation). In fact, given a sequence of control inputs and noise values, the state z contains the system output y over time, and can

$$ \tag{5} $$

$$
y_z(t) = \underbrace{\begin{bmatrix} I & 0 & \cdots & 0 & 0 & 0 & \cdots & 0 \\ 0 & I & \cdots & 0 & 0 & 0 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots & \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & I & 0 & 0 & \cdots & 0 \end{bmatrix}}_{\mathcal{C}} z(t)
$$

be used to reconstruct the exact value of the system state x. This also implies that a controller for the system (1) can equivalently be designed using the dynamics (5). In fact, we show that any *dynamic* controller for (1) can be equivalently represented as a *static* controller for (5), see Appendix C. Next, we reformulate the LQG problem (2) for the behavioral dynamics (5) and characterize its optimal solution. The LQG problem (2) can be equivalently written in the behavioral space as:

$$\mathcal{J}_z \triangleq \lim_{T \to \infty} \mathbb{E}\left[\frac{1}{T}\Big(\sum_{t=n}^{T-1} z(t)^{\mathsf{T}} Q_z z(t) + u(t)^{\mathsf{T}} R_u u(t)\Big)\right], \tag{6}$$

subject to (5), where Q_z is presented in Appendix D along with the derivation of (6), and R_u is as in (2). The solution to the LQG problem in the behavioral space is given by a static controller, which we characterize next.

*Theorem 2.1: (**Behavioral solution to the LQG problem**)* Let u^* be the minimizer of (6) subject to (5). Then,

$$u^*(t) = \underbrace{-\left(R_u + \mathcal{B}_u^{\mathsf{T}} M \mathcal{B}_u\right)^{-1} \mathcal{B}_u^{\mathsf{T}} M \mathcal{A} P \mathcal{C}^{\mathsf{T}} \left(\mathcal{C} P \mathcal{C}^{\mathsf{T}}\right)^{\dagger}}_{\mathcal{K}} y_z(t) \tag{7}$$

where $M \succeq 0$ and $P \succeq 0$ are the unique solutions of the following coupled Riccati equations:

$$M = \mathcal{A}^{\mathsf{T}} M \mathcal{A} - \mathcal{A}^{\mathsf{T}} M \mathcal{B}_u S_M \mathcal{B}_u^{\mathsf{T}} M \mathcal{A} + Q_z$$
$$+ \left(I - P\mathcal{C}^{\mathsf{T}} S_P \mathcal{C}\right)^{\mathsf{T}} \mathcal{A}^{\mathsf{T}} M \mathcal{B}_u S_M \mathcal{B}_u^{\mathsf{T}} M \mathcal{A} \left(I - P\mathcal{C}^{\mathsf{T}} S_P \mathcal{C}\right)$$
$$P = \mathcal{A} P \mathcal{A}^{\mathsf{T}} - \mathcal{A} P \mathcal{C}^{\mathsf{T}} S_P \mathcal{C} P \mathcal{A}^{\mathsf{T}} + \mathcal{B}_w Q_w \mathcal{B}_w^{\mathsf{T}} + \mathcal{B}_v R_v \mathcal{B}_v^{\mathsf{T}}$$
$$+ \left(I - M \mathcal{B}_u S_M \mathcal{B}_u^{\mathsf{T}}\right)^{\mathsf{T}} \mathcal{A} P \mathcal{C}^{\mathsf{T}} S_P \mathcal{C} P \mathcal{A}^{\mathsf{T}} \left(I - M \mathcal{B}_u S_M \mathcal{B}_u^{\mathsf{T}}\right)$$

with $S_M \triangleq (R_u + \mathcal{B}_u^{\mathsf{T}} M \mathcal{B}_u)^{-1}$ and $S_P \triangleq (\mathcal{C} P \mathcal{C}^{\mathsf{T}})^{\dagger}$. $\square$

The proof of Theorem 2.1 is postponed to Appendix E. The gain $\mathcal{K}$ is not unique since $\mathcal{C} P \mathcal{C}^{\mathsf{T}}$ is generally not invertible. In some cases, such as with SISO systems, the gain $\mathcal{K}$ becomes unique, which gives solving for the optimal LQG controller in the behavioral space an advantage over solving for it in the state space. The issue of non-uniqueness of $\mathcal{K}$ stems from the fact that y_z has components that are dependent on each other, which makes the left kernel of $\mathcal{C} P \mathcal{C}^{\mathsf{T}}$ non-empty. We can avoid this issue by carefully choosing the time window of U and Y that form the behavioral space in (4), but we leave this aspect for future work. Note that, solving the coupled Riccati equations that characterize the LQG gain in Theorem 2.1 can be challenging. One method to solve for the LQG gain is to solve for the LQR and the Kalman gains, then use (16) and Lemma 5.3.

*Example 1: (**LQG controller in the behavioral space**)* Consider the system (1) with $A = 1.1$, $B = 1$, $C = 1$, $Q_w = 0.5$, and $R_v = 0.8$. Also, consider the optimal control problem (2) with $Q_x = R_u = 1$. The Kalman and the LQR gains are $K_{\text{kf}} = 0.5474$ and $K_{\text{lqr}} = 0.7034$, respectively, which can be written as (3) using (16) with $E = 0.1716$, $F = 0.0973$, $G = -0.7034$, and $H = -0.3991$. Using (4), we define the behavioral space as $z(t) \triangleq$
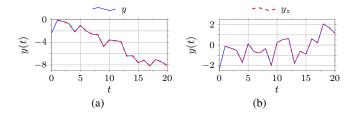

Fig. 1. This figure shows the free response and the LQG feedback response of (1) and (5) for the setting defined in Example 1. In both panels, the solid blue line and the dashed red line represent the output of (1) and the output of (5) that corresponds to $y(t)$, respectively. Panel (a) shows the free response of (1) and (5), we observe that the response of both systems are equal, which agrees with Lemma 5.2. Panel (b) shows the feedback response of (1) and (5) to the LQG controller (16) and the behavioral LQG controller in Theorem 2.1, respectively. We observe that both systems have equal responses, which agrees with Lemma 5.4 and Theorem 2.1. Notice that the response of (5) starts at time $t = n = 1$ since we have to wait $n = 1$ time steps in order to get the equivalent initial condition for (5).

$[u(t-1), y(t-1), y(t), w(t-1), v(t-1), v(t)]^{\mathsf{T}}$ for $t \geq 1$. Using Lemma 5.2, we write the equivalent dynamics of (1) in the behavioral space as (5) with $\mathcal{A}_u = 0.4977$, $\mathcal{A}_y = \begin{bmatrix} 0.5475 & 0.6023 \end{bmatrix}$, $\mathcal{A}_w = 0.4977$, and $\mathcal{A}_y = \begin{bmatrix} -0.5475 & -0.6023 \end{bmatrix}$. Using Theorem 2.1, the LQG gain is $\mathcal{K} = [0.1716, 0, -0.3991]$. Fig. 1(a) shows the free response of (1) and (5) with equal initial conditions. Fig. 1(b) shows the response of (1) and (5) to the LQG controllers (16) and (7), respectively. $\square$

III. Implications of behavioral representation in numerical methods

In this section, we highlight some implications of our behavioral representation and results. In particular, we provide an analysis of learning the LQG controller from finite expert demonstrations, and an analysis of solving for the behavioral LQG gain via a gradient descent method. First, we present the following Lemma regarding the sparsity of the LQG gain in (7), which we use in our subsequent analysis.

*Lemma 3.1: (**Sparsity of the optimal LQG gain**)* Consider the LQG gain written in the behavioral space as

$$u(t) = \begin{bmatrix} \mathcal{K}_1 & \mathcal{K}_2 & \mathcal{K}_3 \end{bmatrix} \begin{bmatrix} U(t-1) \\ y(t-n) \\ \overline{Y}(t) \end{bmatrix}, \tag{8}$$

where $\overline{Y}(t) \triangleq \begin{bmatrix} y(t-n+1)^{\mathsf{T}}, \cdots, y(t)^{\mathsf{T}} \end{bmatrix}^{\mathsf{T}}$. Then $\mathcal{K}_2 = 0.\square$ A proof of Lemma 3.1 is presented in Appendix F.

A. Learning LQG controller from expert demonstrations

Consider the system (1), assume that the system is stabilized by an expert that uses optimal LQG controller. We also assume that the system and the noise statistics are unknown. Our objective is to learn the optimal LQG controller from finite expert demonstrations, which are composed of input and output data. In the behavioral representation, this boils down to learning the gain $\mathcal{K}$ of the subspace $u(t) = \mathcal{K} y_z(t)$ for $t \geq n$. Using Lemma 3.1, we only need to learn $\mathcal{K}_1$ and

$\mathcal{K}_3$, which are obtained as $[\mathcal{K}_1 \; \mathcal{K}_3] = U_N Y_N^\dagger + \mathcal{K}_{\text{null}}$, where

$$U_N \triangleq \begin{bmatrix} u(t) \cdots u(t+k-1) \end{bmatrix},$$

$$Y_N \triangleq \begin{bmatrix} u(t-n) & \cdots & u(t-n+k-1) \\ \vdots & \ddots & \vdots \\ u(t-1) & \cdots & u(t-2+k) \\ y(t-n+1) & \cdots & y(t-n+k) \\ \vdots & \ddots & \vdots \\ y(t) & \cdots & y(t-1+k) \end{bmatrix}, \quad (9)$$

for $t \geq n$, where k is the number of columns, and $\mathcal{K}_{\text{null}}$ is any matrix with appropriate dimension whose rows belong to the left null space of Y_N. Note that $\mathcal{K}_{\text{null}}$ will disappear when multiplied by the feedback $y_z(t)$, i.e., $\mathcal{K}_{\text{null}} y_z(t) = 0$. Therefore, without loss of generality, we set $\mathcal{K}_{\text{null}} = 0$.

Lemma 3.2: (Sufficient number of expert data to compute the optimal LQG gain) Consider input and output expert samples $U = [u(t), \cdots, u(t+N-1)]$ and $Y = [y(t), \cdots, y(t+N-1)]$ generated by LQG controller to stabilize system (1), such that U is full-rank. Then, $N = n + nm + np$ expert samples are sufficient to compute the LQG gain $\mathcal{K}$. $\qquad\square$

A proof of Lemma 3.2 is presented in Appendix G. We note that the rank condition on the input matrix U in the statement of Lemma 3.2 is a reasonable assumption owing to the fact that system (1) is driven by i.i.d. process noise w and that the measurement noise v is also i.i.d. Furthermore, note that we can learn the dynamic controller matrices E, F, G, and H in (3) (up to a similarity transformation) using subspace identification methods for deterministic systems (see [23]) with U and Y treated as the output and input signals to (3), respectively. Using [23, Theorem 2], we need at least $N = 2(n+1)(m+p+1) - 1$ expert samples to learn (3), which is more than the sufficient number of expert samples to learn $\mathcal{K}$ (Lemma 3.2).

Example 2: (Learning LQG controller from expert data) Consider the system in Example 1 where the system dynamics and the noise statistics are assumed to be unknown. The system is driven by an expert that uses an LQG policy. According to Lemma 3.2, we collect $N = n + nm + np = 3$ expert input-output samples to form the data matrices

$$U_N = \begin{bmatrix} -0.2269 & -0.1231 \end{bmatrix}, \quad Y_N = \begin{bmatrix} 1.7878 & -0.2269 \\ 1.3371 & 0.211 \end{bmatrix}.$$

Using the data, we obtain $[\mathcal{K}_1 \; \mathcal{K}_3] = [0.1716 \; -0.3991]$ with $\mathcal{K}_{\text{null}} = 0$, which matches the LQG gain in Example 1. $\quad\square$

B. Gradient descent in the behavioral space

In this section, we use gradient descent to solve for $\mathcal{K}$:

$$\mathcal{K}^{(i+1)} = \mathcal{K}^{(i)} - \alpha^{(i)} \nabla \mathcal{J}_z(\mathcal{K}^{(i)}) \quad \text{for } i = 0, 1, 2, \cdots \quad (10)$$

where the index i refers to the iteration number, $\alpha^{(i)}$ is the step size at iteration i, and $\nabla \mathcal{J}_z(\mathcal{K}^{(i)})$ is computed using (26). We initialize the gradient descent method with a stabilizing gain $\mathcal{K}^{(0)}$. We determine the step size $\alpha^{(i)}$ by the Armijo rule [24, Chapter 1.3]: initialize $\alpha^{(0)} = 1$, repeat $\alpha^{(i)} = \beta \alpha^{(i)}$ until

$$\mathcal{J}_z(\mathcal{K}^{(i+1)}) \leq \mathcal{J}_z(\mathcal{K}^{(i)}) - \sigma \alpha^{(i)} \left\| \nabla \mathcal{J}_z(\mathcal{K}^{(i)}) \right\|_F^2$$
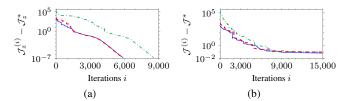

Fig. 2. This figure shows the convergence performance (measured by the suboptimality gap) of the gradient descent applied to the system in Example 3. The solid blue line, dashed red line and the dash-dotted green line represent different initial conditions, respectively. Panels (a) and (b) show the convergence performance of the gradient descent over $\mathcal{K}$ and the gradient descent over the controller matrices E, F, G and H, respectively.

is satisfied, with $\beta, \sigma \in (0, 1)$.

Example 3: (Gradient descent) We consider the example in [13] discretized with sampling time $T_s = 0.4$,

$$A = \begin{bmatrix} 1.4918 & 0.5967 \\ 0 & 1.4918 \end{bmatrix}, \quad B = \begin{bmatrix} 0.1049 \\ 0.4918 \end{bmatrix}, \quad C = \begin{bmatrix} 1 & 0 \end{bmatrix},$$

$$Q_w = \begin{bmatrix} 4.6477 & 3.7575 \\ 3.7575 & 3.0639 \end{bmatrix}, \quad Q_x = \begin{bmatrix} 3.0639 & 3.7575 \\ 3.7575 & 4.6477 \end{bmatrix},$$

$R_v = 2.5$ and $R_u = 0.5966$. The LQG gain from Theorem 2.1 is $\mathcal{K} = [-0.0366, -0.103, 0, 5.8461, -4.7434]$. Using Lemma 3.1, we only need to do the search over $\mathcal{K}_1$ and $\mathcal{K}_3$ since $\mathcal{K}_2 = 0$. We use gradient descent in (10) to solve for the LQG gain. We choose a stabilizing initial gain $\mathcal{K}^{(0)}$ that randomly place the closed-loop eigenvalues within $[0.45, 0.92]$. We use the Armijo rule to compute the step size with $\alpha^{(0)} = 1$, $\beta = 0.8$, and $\sigma = 0.7$. We set the stopping criteria to be when the gradient vanishes or when the maximum number of iterations is reached (in this example we set it to 15000 iterations). For numerical comparison, we use gradient descent to solve for the optimal LQG dynamic controller in the form of (3) as in [6], where we optimize the LQG cost (2) and apply the gradient search over the control parameters E, F, G, and H.[1] Fig. 2 shows the convergence performance of the gradient descent for different initial conditions. We observe that the gradient descent over $\mathcal{K}$ in Fig. 2(a) converges to $\mathcal{K}^* = [-0.0366, -0.1030, 0, 5.8460, -4.7434]$ before reaching the maximum number iterations for different initial conditions. Starting from initial conditions equivalent to the ones in Fig. 2(a), the gradient descent over the controller matrices E, F, G and H in Fig. 2(b) did not converge within 15000 iterations. $\quad\square$

IV. CONCLUSION AND FUTURE WORK

In this work, we revisited the LQG optimal control problem from a behavioral perspective. We introduced equivalent representations for the class of stochastic discrete-time, linear, time-invariant systems and the LQG optimal control problem in the space of input-output behaviors. In particular,

---

[1]In [6], $H = 0$ since it is assumed that the control input $u(t)$ at time t depends on the history $\{u(0), \cdots, u(t-1), y(0), \cdots, y(t-1)\}$. In this paper, $u(t)$ depends also on $y(t)$, therefore H is nonzero (see Appendix A). We computed the gradient of $\mathcal{J}$ w.r.t. the controller matrices E, F, G and H as in [6] adapted to the case where H is nonzero. We have not included the derivations in this paper due to space constraint.

we showed that the optimal LQG controller can be expressed as a static behavioral-feedback gain, which can be solved for directly from the LQG problem in the behavioral space. Finally, we highlighted the advantages of having a static LQG gain over a dynamic LQG controller in the context of data-driven control and gradient-based algorithms, which arise from the fact that the behavioral approach circumvents the need for a state space representation and the fact that the optimal behavioral-feedback is a static gain. There still remain several unexplored questions, including the investigation of the optimization landscape of the LQG problem in the behavioral space, which will pave the way for an improved understanding of the convergence properties of data-driven and gradient algorithms, as well as, for investigating the uniqueness of the optimal LQG gain.

REFERENCES

[1] G. Baggio, V. Katewa, and F. Pasqualetti. Data-driven minimum-energy controls for linear systems. *IEEE Control Systems Letters*, 3(3):589–594, 2019.

[2] J. Coulson, J. Lygeros, and F. Dörfler. Data-enabled predictive control: In the shallows of the DeePC. In *European Control Conference*, pages 307–312, Naples, Italy, 2019.

[3] S. Tu and B. Recht. The gap between model-based and model-free methods on the linear quadratic regulator: An asymptotic viewpoint. In *Conference on Learning Theory*, volume 99 of *Proceedings of Machine Learning Research*, pages 3036–3083, Phoenix, AZ, Jun. 2019. PMLR.

[4] F. Celi, G. Baggio, and F. Pasqualetti. Distributed learning of optimal controls for linear systems. In *IEEE Conf. on Decision and Control*, pages 5764–5769, Austin, TX, December 2021.

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

[6] Y. Zheng, Y. Tang, and N. Li. Analysis of the optimization landscape of linear quadratic gaussian (lqg) control. *arXiv preprint arXiv:2102.04393*, 2021.

[7] M. Fazel, R. Ge, S. Kakade, and M. Mesbahi. Global convergence of policy gradient methods for the linear quadratic regulator. In *International Conference on Machine Learning*, pages 1467–1476, Stockholm, Sweden, 2018. PMLR.

[8] H. Mohammadi, A. Zare, M. Soltanolkotabi, and M. R. Jovanović. Global exponential convergence of gradient methods over the nonconvex landscape of the linear quadratic regulator. In *IEEE Conf. on Decision and Control*, pages 7474–7479, Nice, France, Dec. 2019.

[9] J. Bu, A. Mesbahi, M. Fazel, and M. Mesbahi. Lqr through the lens of first order methods: Discrete-time case. *arXiv preprint arXiv:1907.08921*, 2019.

[10] I. Fatkhullin and B. Polyak. Optimizing static linear feedback: Gradient method. *SIAM Journal on Control and Optimization*, 59(5):3887–3911, 2021.

[11] 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.

[12] D. P. Bertsekas. *Dynamic Programming and Optimal Control, Vol. 1*. Athena Scientific, 2 edition, 2001.

[13] J. C. Doyle. Guaranteed margins for LQG regulators. *IEEE Transactions on automatic Control*, 23(4):756–757, 1978.

[14] S. Lale, K. Azizzadenesheli, B. Hassibi, and A. Anandkumar. Logarithmic regret bound in partially observable linear dynamical systems. In *Advances in Neural Information Processing Systems*, volume 33, pages 20876–20888, Virtual, Dec. 2020. Curran Associates, Inc.

[15] L. Furieri, Y. Zheng, and M. Kamgarpour. Learning the globally optimal distributed lq regulator. In *Learning for Dynamics and Control*, volume 120 of *Proceedings of Machine Learning Research*, pages 287–297, Virtual, Jun. 2020. PMLR.

[16] 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*

[17] of Machine Learning Research, pages 559–570, Virtual, Jun. 2021. PMLR.

[18] J. C. Willems, P. Rapisarda, I. Markovsky, and B. L. M. De Moor. A note on persistency of excitation. *Systems & Control Letters*, 54(4):325–329, 2005.

[18] C. De Persis and P. Tesi. Formulas for data-driven control: Stabilization, optimality and robustness. *IEEE Transactions on Automatic Control*, 65(3):909–924, 2020.

[19] L. Furieri, B. Guo, A. Martin, and G. Ferrari-Trecate. A behavioral input-output parametrization of control policies with suboptimality guarantees. *arXiv preprint arXiv:2102.13338*, 2021.

[20] V. Krishnan and F. Pasqualetti. On direct vs indirect data-driven predictive control. In *IEEE Conf. on Decision and Control*, pages 736–741, Austin, TX, December 2021.

[21] G. C. Goodwin and K. S. Sin. *Adaptive filtering prediction and control*. Courier Corporation, 2014.

[22] I. Markovsky and F. Dörfler. Behavioral systems theory in data-driven analysis, signal processing, and control. *Annual Reviews in Control*, 52:42–64, 2021.

[23] P. V. Overschee and B. D. Moor. *Subspace identification for linear systems: Theory-Implementation-Applications*. Kluwer Academic Publishers, 1996.

[24] D. P. Bertsekas. *Nonlinear Programming*. Athena Scientific, 1995.

[25] A. A. Al Makdah, V. Katewa, and F. Pasqualetti. Accuracy prevents robustness in perception-based control. In *American Control Conference*, Denver, CO, USA, July 2020.

APPENDIX

A. Optimal LQG controller

The optimal LQG controller that solves (2) is written as

$$\hat{x}(t+1) = (I_n - K_{\text{kf}}C)(A - BK_{\text{LQR}})\hat{x}(t) + K_{\text{kf}}y(t+1),$$
$$u(t) = -K_{\text{LQR}}\hat{x}(t), \tag{11}$$

where K_{kf} and K_{LQR} are the Kalman and LQR gains, respectively. To write the controller (11) in the form of (3), we need the following lemma.

*Lemma 5.1: (**Equivalent compensator forms**)* Consider the compensator (3) and a compensator of the form:

$$\xi_c(t+1) = \overline{E}\xi_c(t) + \overline{F}y(t+1),$$
$$u(t) = \overline{G}\xi_c(t), \tag{12}$$

with $\xi_c \in \mathbb{R}^n$ denoting the state, and $\overline{E} \in \mathbb{R}^{n\times n}$, $\overline{F} \in \mathbb{R}^{n\times p}$, and $\overline{G} \in \mathbb{R}^{m\times n}$ denoting the dynamic, input, and output matrices of the compensator, respectively. Let $x_c(0) = \xi_c(0)$ and $y(0) = 0$, then, the compensators (3) and (12) output the same $u(t)$ given the same input $y(t)$ if:

$$E = \overline{E}, \quad F = \overline{E}\,\overline{F}, \quad G = \overline{G}, \quad H = \overline{G}\,\overline{F}. \tag{13}$$

Proof: Using (3) with $y(0) = 0$, we can write

$$u(t) = GE^t x_c(0) + \begin{bmatrix} GE^{t-2}F & \cdots & GF & H \end{bmatrix} y, \tag{14}$$

where $y = [y(1)^{\mathsf{T}}, \cdots, y(t)^{\mathsf{T}}]^{\mathsf{T}}$. Using (12), we can write

$$u(t) = \overline{G}\,\overline{E}^t \xi_c(0) + \begin{bmatrix} \overline{G}\,\overline{E}^{t-1}\overline{F} & \cdots & \overline{G}\,\overline{F} \end{bmatrix} y. \tag{15}$$

Under the same y, (14) is equal to (15) for $E = \overline{E}$, $F = \overline{E}\,\overline{F}$, $G = \overline{G}$, and $H = \overline{G}\,\overline{F}$. ∎

Using Lemma 5.1 and (11), we can write the LQG controller in the form of (3) with

$$
\begin{aligned}
E &= (I_n - K_{\text{kf}}C)(A - BK_{\text{LQR}}), \\
F &= (I_n - K_{\text{kf}}C)(A - BK_{\text{LQR}})K_{\text{kf}}, \\
G &= -K_{\text{LQR}}, \\
H &= -K_{\text{LQR}}K_{\text{kf}}.
\end{aligned} \tag{16}
$$

B. System representation in the behavioral space

The following Lemma provides an equivalent representation of (1) in the behavioral space, z, which is written in (5).

*Lemma 5.2: (**Equivalent dynamics**)* Let z be as in (4). Then,

$$z(t+1) = \mathcal{A}z(t) + \mathcal{B}_u u(t) + \mathcal{B}_w w(t) + \mathcal{B}_v v(t+1),$$

where $\mathcal{A}$, $\mathcal{B}_u$, $\mathcal{B}_w$, and $\mathcal{B}_v$ are as in (5), and

$$\mathcal{A}_u \triangleq \mathcal{F}_2 - CA^{n+1}\mathcal{O}^\dagger \mathcal{F}_1, \qquad \mathcal{A}_y \triangleq CA^{n+1}\mathcal{O}^\dagger,$$
$$\mathcal{A}_w \triangleq \mathcal{F}_4 - CA^{n+1}\mathcal{O}^\dagger \mathcal{F}_3, \qquad \mathcal{A}_v \triangleq -CA^{n+1}\mathcal{O}^\dagger,$$

$$\mathcal{O} \triangleq \begin{bmatrix} C \\ CA \\ \vdots \\ CA^n \end{bmatrix}, \quad \mathcal{F}_1 \triangleq \begin{bmatrix} 0 & \cdots & 0 \\ CB & \cdots & 0 \\ \vdots & \ddots & \vdots \\ CA^{n-1}B & \cdots & CB \end{bmatrix},$$
$$\mathcal{F}_2 \triangleq \begin{bmatrix} CA^n B \cdots CAB \end{bmatrix},$$

and the matrices $\mathcal{F}_3$ and $\mathcal{F}_4$ are obtained by replacing B with I in $\mathcal{F}_1$ and $\mathcal{F}_2$, respectively.

Proof: We can write the evolution of $y(t+1)$ as

$$y(t+1) = CA^{n+1}x(t-n) + \mathcal{F}_2 U(t-1) + \mathcal{F}_4 W(t-1) \\ + CBu(t) + Cw(t) + v(t+1), \quad (17)$$

where, $\mathcal{F}_2$ and $\mathcal{F}_4$ are as in Lemma 5.2, and $U(t-1)$ and $W(t-1)$ are as in (4). Also, we can write $Y(t)$ in (4) in terms of $U(t-1)$, $W(t-1)$, and $V(t)$:

$$Y(t) = \mathcal{O}x(t-n) + \mathcal{F}_1 U(t-1) + \mathcal{F}_3 W(t-1) + V(t), \quad (18)$$

where $\mathcal{O}$, $\mathcal{F}_1$, and $\mathcal{F}_3$ are same as in Lemma 5.2 and $V(t)$ is as in (4). Then using (18), we substitute $x(t-n)$ into (17) to get

$$y(t+1) = \begin{bmatrix} \mathcal{A}_u & \mathcal{A}_y & \mathcal{A}_w & \mathcal{A}_v \end{bmatrix} \underbrace{\begin{bmatrix} U(t-1) \\ Y(t) \\ W(t-1) \\ V(t) \end{bmatrix}}_{z(t)} + CBu(t) \\ + Cw(t) + v(t+1),$$

where $\mathcal{A}_u$, $\mathcal{A}_y$, $\mathcal{A}_w$, and $\mathcal{A}_v$ are as in Lemma 5.2. ∎

C. From dynamic to static controller

*Lemma 5.3: (**From dynamic to static controllers**)* Let the control input u be the output of the dynamic controller (3). Then, equivalently,

$$u(t) = \begin{bmatrix} GE^n\mathcal{T}_1^\dagger & \mathcal{T}_2 - GE^n\mathcal{T}_1^\dagger \mathcal{M} \end{bmatrix} y_z(t), \quad (19)$$

where y_z is as in (5), and

$$\mathcal{T}_1 \triangleq \begin{bmatrix} G \\ GE \\ \vdots \\ GE^{n-1} \end{bmatrix}, \quad \mathcal{T}_2 \triangleq \begin{bmatrix} GE^{n-1}F \cdots GFH \end{bmatrix},$$

$$\mathcal{M} \triangleq \begin{bmatrix} H & 0 & 0 & \cdots & 0 & 0 \\ GF & H & 0 & \cdots & 0 & 0 \\ GEF & GF & H & \cdots & 0 & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots & \vdots \\ GE^{n-2}F & GE^{n-3}F & \cdots & \cdots & H & 0 \end{bmatrix}.$$

Proof: Using (3), we can write

$$u(t) = GE^n x_c(t-n) + \mathcal{T}_2 Y(t), \quad (20)$$

where $\mathcal{T}_2$ and $Y(t)$ are as in Lemma 5.3 and (4), respectively. Further, we can write $U(t-1)$ in (4) as

$$U(t-1) = \mathcal{T}_1 x_c(t-n) + \mathcal{M}Y(t), \quad (21)$$

where $\mathcal{T}_1$ and $\mathcal{M}$ are as in Lemma 5.3. Using (21) we substitute $x_c(t-n)$ into (20) to get

$$u(t) = \begin{bmatrix} GE^n\mathcal{T}_1^\dagger & \mathcal{T}_2 - GE^n\mathcal{T}_1^\dagger \mathcal{M} \end{bmatrix} \underbrace{\begin{bmatrix} U(t-1) \\ Y(t) \end{bmatrix}}_{y_z}.$$
∎

D. LQG problem in the behavioral space

*Lemma 5.4: (**LQG problem in the behavioral space**)* The input u^* is the minimizer of (2) subject to (1) if and only if it is the minimizer of

$$\mathcal{J}_z \triangleq \lim_{T\to\infty} \mathbb{E}\left[\frac{1}{T}\left(\sum_{t=0}^{T-1} z(t)^\mathsf{T} Q_z z(t) + u(t)^\mathsf{T} R_u u(t) \right) \right] \quad (22)$$

subject to (5), where $Q_z = \mathcal{H}^\mathsf{T} Q_x \mathcal{H}$ and

$$\mathcal{H} \triangleq \begin{bmatrix} \mathcal{G}_1 - A^n\mathcal{O}^\dagger \mathcal{F}_1 & A^n\mathcal{O}^\dagger & \mathcal{G}_2 - A^n\mathcal{O}^\dagger \mathcal{F}_3 & -A^n\mathcal{O}^\dagger \end{bmatrix},$$
$$\mathcal{G}_1 \triangleq \begin{bmatrix} A^{n-1}B & \cdots & B \end{bmatrix}, \quad \mathcal{G}_2 \triangleq \begin{bmatrix} A^{n-1} & \cdots & I_n \end{bmatrix}.$$

Proof: We begin by proving that the costs in (2) and (22) are equivalent. We can express $x(t)$ for $t \geq n$ as

$$x(t) = A^n x(t-n) + \mathcal{G}_1 U(t-1) + \mathcal{G}_2 W(t-1), \quad (23)$$

where $\mathcal{G}_1$ and $\mathcal{G}_2$ are as in Lemma 5.4, and $U(t-1)$ and $W(t-1)$ are as in (4). Using (18), we can substitute $x(t-n)$ in terms of $U(t-1)$, $Y(t)$, $W(t-1)$, and $V(t)$ into (23) to get $x(t) = \mathcal{H}z(t)$, where $\mathcal{H}$ is as in Lemma 5.4. Substituting $x(t) = \mathcal{H}z(t)$ into the cost (2) yields the cost (22). Further, Lemma 5.2 shows that the systems (1) and (5) are equivalent. Therefore, the minimizer of (2) subject to (1) is the minimizer of (22) subject to (5). ∎

E. Proof of Theorem 2.1

For the proof of Theorem 2.1, we need the following technical results from the literature.

*Lemma 5.5: (**Steady-state cost**)* For a controller $u(t) = \mathcal{K}y_z(t)$ with stabilizing gain $\mathcal{K}$, the cost (6) at steady-state is written as

$$\mathcal{J}_z(\mathcal{K}) = \mathrm{tr}\left[Q_\mathcal{K} P \right], \quad (24)$$

where $Q_\mathcal{K} \triangleq Q_z + \mathcal{C}^\mathsf{T}\mathcal{K}^\mathsf{T} R_u \mathcal{K}\mathcal{C}$, and $P \succeq 0$ is the unique solution of the following Lyapunov equation

$$P = \mathcal{A}_c P \mathcal{A}_c^\mathsf{T} + \mathcal{B}_w Q_w \mathcal{B}_w^\mathsf{T} + \mathcal{B}_v R_v \mathcal{B}_v^\mathsf{T}. \quad (25)$$

with $\mathcal{A}_c \triangleq \mathcal{A} + \mathcal{B}_u\mathcal{K}\mathcal{C}$.

Proof: Since $u(t) = \mathcal{K}y_z(t)$ is stabilizing, the closed-loop matrix $\mathcal{A}_c = \mathcal{A} + \mathcal{B}_u\mathcal{K}\mathcal{C}$ is stable. We can write

$$\mathbb{E}\left[z(t)z(t)^\mathsf{T}\right] = \mathcal{A}_c\mathbb{E}\left[z(t-1)z(t-1)^\mathsf{T}\right]\mathcal{A}_c^\mathsf{T} + \mathcal{B}_wQ_w\mathcal{B}_w^\mathsf{T}$$
$$+ \mathcal{B}_vR_v\mathcal{B}_v^\mathsf{T},$$

where we have used the fact that $z(t-1)$, $w(t-1)$, and $v(t)$ are uncorrelated, and $\mathbb{E}\left[w(t-1)w(t-1)^\mathsf{T}\right] = Q_w$ and $\mathbb{E}\left[v(t)v(t)^\mathsf{T}\right] = R_v$. Since $\mathcal{A}_c$ is stable, $\lim_{t\to\infty}\mathbb{E}\left[z(t)z(t)^\mathsf{T}\right]$ converges to a finite value, and at steady state we have $P \triangleq \lim_{t\to\infty}\mathbb{E}\left[z(t)z(t)^\mathsf{T}\right] = \lim_{t\to\infty}\mathbb{E}\left[z(t-1)z(t-1)^\mathsf{T}\right]$, where P satisfies (25). The cost (6) is written as

$$\mathcal{J}_z \triangleq \lim_{T\to\infty}\mathbb{E}\left[\frac{1}{T}\left(\sum_{t=0}^{T-1} z(t)^\mathsf{T}\left(Q_z + \mathcal{C}^\mathsf{T}\mathcal{K}^\mathsf{T}R_u\mathcal{K}\mathcal{C}\right)z(t)\right)\right]$$
$$= \lim_{t\to\infty}\mathbb{E}\left[\mathrm{tr}\left[z(t)^\mathsf{T}Q_\mathcal{K}z(t)\right]\right] = \mathrm{tr}\left[Q_\mathcal{K}\lim_{t\to\infty}\mathbb{E}\left[z(t)z(t)^\mathsf{T}\right]\right]$$
$$= \mathrm{tr}\left[Q_\mathcal{K}P\right],$$

where $Q_\mathcal{K} \triangleq Q_z + \mathcal{C}^\mathsf{T}\mathcal{K}^\mathsf{T}R_u\mathcal{K}\mathcal{C}$. The proof is complete. ∎

Lemma 5.6: **(Property of the solution to Lyapunov equation, [25])** Let A, B, Q be matrices of appropriate dimensions with $\rho(A) < 1$. Let Y satisfy $Y = AYA^\mathsf{T} + Q$. Then, $\mathrm{tr}\left[BY\right] = \mathrm{tr}\left[Q^\mathsf{T}M\right]$, where M satisfies $M = A^\mathsf{T}MA + B^\mathsf{T}$. □

Proof of Theorem 2.1: Using Lemma 5.5, we can write the cost (6) at steady-state as (24). Next, we compute the derivative of $\mathcal{J}_z(\mathcal{K})$ with respect to the variable $\mathcal{K}$. Taking the differential of (25) with respect to the variable $\mathcal{K}$, we get

$$dP = \mathcal{A}_cdP\mathcal{A}_c^\mathsf{T} + d\mathcal{A}_cP\mathcal{A}_c^\mathsf{T} + \mathcal{A}_cPd\mathcal{A}_c^\mathsf{T} \triangleq \mathcal{A}_cdP\mathcal{A}_c^\mathsf{T} + X$$
$$\implies \mathrm{tr}\left[Q_\mathcal{K}dP\right] \overset{(a)}{=} \mathrm{tr}\left[XM\right] \overset{(b)}{=} 2\mathrm{tr}\left[\mathcal{C}P\mathcal{A}_c^\mathsf{T}M\mathcal{B}_ud\mathcal{K}\right],$$

where $M \succeq 0$ satisfies $M = \mathcal{A}_c^\mathsf{T}M\mathcal{A}_c + Q_\mathcal{K}$, (a) follows from Lemma 5.6, and (b) follows from $\mathrm{tr}\left[d\mathcal{A}_cP\mathcal{A}_c^\mathsf{T}M\right] = \mathrm{tr}\left[(d\mathcal{A}_cP\mathcal{A}_c^\mathsf{T}M)^\mathsf{T}\right]$ and using the trace cyclic property. Taking the differential of $Q_\mathcal{K}$, we get

$$dQ_\mathcal{K} = \mathcal{C}^\mathsf{T}d\mathcal{K}^\mathsf{T}R_u\mathcal{K}\mathcal{C} + \mathcal{C}^\mathsf{T}\mathcal{K}^\mathsf{T}R_ud\mathcal{K}\mathcal{C}$$
$$\implies \mathrm{tr}\left[dQ_\mathcal{K}P\right] \overset{(c)}{=} 2\mathrm{tr}\left[\mathcal{C}P\mathcal{C}^\mathsf{T}\mathcal{K}^\mathsf{T}R_ud\mathcal{K}\right],$$

where (c) follows similarly as (b). For notational convenience, we denote $\mathcal{J}_z(\mathcal{K})$ by $\mathcal{J}_z$. Taking the differential of $\mathcal{J}_z$ in (24), we get,

$$d\mathcal{J}_z = d\mathrm{tr}\left[Q_\mathcal{K}P\right] = \mathrm{tr}\left[dQ_\mathcal{K}P\right] + \mathrm{tr}\left[Q_\mathcal{K}dP\right]$$
$$= 2\mathrm{tr}\left[\left(\mathcal{C}P\mathcal{C}^\mathsf{T}\mathcal{K}^\mathsf{T}R_u + \mathcal{C}P\mathcal{A}_c^\mathsf{T}M\mathcal{B}_u\right)d\mathcal{K}\right]$$
$$\implies \frac{d\mathcal{J}_z}{d\mathcal{K}} = 2\left(R_u\mathcal{K}\mathcal{C}P\mathcal{C}^\mathsf{T} + \mathcal{B}_u^\mathsf{T}M\mathcal{A}_cP\mathcal{C}^\mathsf{T}\right) \quad (26)$$
$$= 2\left(R_u + \mathcal{B}_u^\mathsf{T}M\mathcal{B}_u\right)\mathcal{K}\mathcal{C}P\mathcal{C}^\mathsf{T} + 2\mathcal{B}_u^\mathsf{T}M\mathcal{A}P\mathcal{C}^\mathsf{T}$$

The stationary optimality condition implies $\frac{d\mathcal{J}_z}{d\mathcal{K}} = 0$, we get

$$\mathcal{K} = -\left(R_u + \mathcal{B}_u^\mathsf{T}M\mathcal{B}_u\right)^{-1}\mathcal{B}^\mathsf{T}M\mathcal{A}P\mathcal{C}^\mathsf{T}\left(\mathcal{C}P\mathcal{C}^\mathsf{T}\right)^\dagger + \mathcal{K}_{\text{null}},$$
$$(27)$$

where we have used the right pseudo inverse of $\mathcal{C}P\mathcal{C}^\mathsf{T}$ since it is rank deficient, and $\mathcal{K}_{\text{null}}$ is any matrix with appropriate

dimension whose rows belong to the left null space of $\mathcal{C}P\mathcal{C}^\mathsf{T}$. Next we derive the Riccati equations of M and P. Let $S_M \triangleq (R_u + \mathcal{B}_u^\mathsf{T}M\mathcal{B}_u)^{-1}$ and $S_P \triangleq (\mathcal{C}P\mathcal{C}^\mathsf{T})^\dagger$. Substituting the expression of $\mathcal{K}$ in (27) into (25), we get

$$P = \mathcal{A}P\mathcal{A}^\mathsf{T} - \mathcal{A}P\mathcal{C}^\mathsf{T}S_p\mathcal{C}P\mathcal{A}^\mathsf{T}M\mathcal{B}_uS_M\mathcal{B}_u^\mathsf{T}$$
$$- \mathcal{B}_uS_M\mathcal{B}_u^\mathsf{T}M\mathcal{A}P\mathcal{C}^\mathsf{T}S_P\mathcal{C}P\mathcal{A}^\mathsf{T} + \mathcal{B}_wQ_w\mathcal{B}_w^\mathsf{T} + \mathcal{B}_vR_v\mathcal{B}_v^\mathsf{T}$$
$$+ \mathcal{B}_uS_M\mathcal{B}_u^\mathsf{T}M\mathcal{A}P\mathcal{C}^\mathsf{T}\underbrace{S_P\left(\mathcal{C}P\mathcal{C}^\mathsf{T}\right)S_P}_{\overset{(d)}{=}S_p}\mathcal{C}P\mathcal{A}^\mathsf{T}M\mathcal{B}_uS_M\mathcal{B}_u^\mathsf{T}$$

$$\overset{(e)}{=}\mathcal{A}P\mathcal{A}^\mathsf{T} - \mathcal{A}P\mathcal{C}^\mathsf{T}S_p\mathcal{C}P\mathcal{A}^\mathsf{T}M\mathcal{B}_uS_M\mathcal{B}_u^\mathsf{T}$$
$$- \mathcal{B}_uS_M\mathcal{B}_u^\mathsf{T}M\mathcal{A}P\mathcal{C}^\mathsf{T}S_P\mathcal{C}P\mathcal{A}^\mathsf{T} + \mathcal{B}_wQ_w\mathcal{B}_w^\mathsf{T} + \mathcal{B}_vR_v\mathcal{B}_v^\mathsf{T}$$
$$+ \mathcal{B}_uS_M\mathcal{B}_u^\mathsf{T}M\mathcal{A}P\mathcal{C}^\mathsf{T}S_P\mathcal{C}P\mathcal{A}^\mathsf{T}M\mathcal{B}_uS_M\mathcal{B}_u^\mathsf{T}$$
$$+ \mathcal{A}P\mathcal{C}^\mathsf{T}S_p\mathcal{C}P\mathcal{A}^\mathsf{T} - \mathcal{A}P\mathcal{C}^\mathsf{T}S_p\mathcal{C}P\mathcal{A}^\mathsf{T}$$
$$= \mathcal{A}P\mathcal{A}^\mathsf{T} - \mathcal{A}P\mathcal{C}^\mathsf{T}S_P\mathcal{C}P\mathcal{A}^\mathsf{T} + \mathcal{B}_wQ_w\mathcal{B}_w^\mathsf{T} + \mathcal{B}_vR_v\mathcal{B}_v^\mathsf{T}$$
$$+ \left(I - M\mathcal{B}_uS_M\mathcal{B}_u^\mathsf{T}\right)^\mathsf{T}\mathcal{A}P\mathcal{C}^\mathsf{T}S_P\mathcal{C}P\mathcal{A}^\mathsf{T}\left(I - M\mathcal{B}_uS_M\mathcal{B}_u^\mathsf{T}\right),$$

where (d) follows from the Moore-Penrose conditions, and in (e) we have added and subtracted the term $\mathcal{A}P\mathcal{C}^\mathsf{T}S_p\mathcal{C}P\mathcal{A}^\mathsf{T}$. The Riccati equation of M is derived in similar manner. ∎

F. Proof of Lemma 3.1

$\mathcal{K}_2$ in Lemma 3.1 corresponds to the first block of $\mathcal{T}_2 - GE^n\mathcal{T}_1^\dagger\mathcal{M}$ in (19). We start by expanding $GE^n\mathcal{T}_1^\dagger\mathcal{M}$. Since $\mathcal{T}_1^\dagger$ is full column rank, we have

$$\mathcal{T}_1^\dagger = \left(\mathcal{T}_1^\mathsf{T}\mathcal{T}_1\right)^{-1}\mathcal{T}_1^\mathsf{T}$$
$$= \underbrace{\left(G^\mathsf{T}G + \cdots + (E^{n-1})^\mathsf{T}G^\mathsf{T}GE^{n-1}\right)^{-1}}_{\triangleq S}\mathcal{T}_1^\mathsf{T},$$

then we have

$$GE^n\mathcal{T}_1^\dagger\mathcal{M} =$$
$$GE^nS\left[\ G^\mathsf{T}H + \cdots + (E^{n-1})^\mathsf{T}G^\mathsf{T}GE^{n-2}F\ \vdots\ \mathrm{X}\ \right],$$

where X denotes any matrix. Then, we take the first block of $GE^n\mathcal{T}_1^\dagger\mathcal{M}$ and the first block of $\mathcal{T}_2$ to write $\mathcal{K}_2$ as

$$\mathcal{K}_2 = GE^{n-1}F$$
$$- GE^nS\left(G^\mathsf{T}H + \cdots + (E^{n-1})^\mathsf{T}G^\mathsf{T}GE^{n-2}F\right)$$
$$\overset{(a)}{=}GE^{n-1}F$$
$$- GE^n\underbrace{S\left(\overline{G}^\mathsf{T}\overline{G} + \cdots + (\overline{E}^{n-1})^\mathsf{T}\overline{G}^\mathsf{T}\overline{G}\overline{E}^{n-1}\right)}_{\overset{(b)}{=}I}\overline{F}$$
$$\overset{(c)}{=}GE^{n-1}F - GE^{n-1}F = 0,$$

where in steps (a), (b) and (c) we have used Lemma 5.1. ∎

G. Proof of Lemma 3.2

Since the rank of Y_N in (9) is $\mathrm{Rank}(Y_N) \leq nm + np$, $k = nm + np$ columns are enough for $\mathrm{Rank}(Y_N)$ to stop increasing. To construct Y_N with $k = nm + np$ columns, $nm + np + n$ samples are required. Therefore, $N = nm + np + n$ expert samples are sufficient to learn the LQG gain $\mathcal{K}$. This completes the proof. ∎