

Imitation and Transfer Learning for LQG Control

Taosha Guo¹, Abed AlRahman Al Makdah², Vishaal Krishnan³, and Fabio Pasqualetti⁴, *Member, IEEE*

Abstract—In this letter we study an imitation and transfer learning setting for Linear Quadratic Gaussian (LQG) control, where (i) the system dynamics, noise statistics and cost function are unknown and expert data is provided (that is, sequences of optimal inputs and outputs) to learn the LQG controller, and (ii) multiple control tasks are performed for the same system but with different LQG costs. We show that the LQG controller can be learned from a set of expert trajectories of length $n(l+2)-1$, with n and l the dimension of the system state and output, respectively. Further, the controller can be decomposed as the product of an estimation matrix, which depends only on the system dynamics, and a control matrix, which depends on the LQG cost. This data-based separation principle allows us to transfer the estimation matrix across different LQG tasks, and to reduce the length of the expert trajectories needed to learn the LQG controller to $2n + m - 1$ with m the dimension of the inputs (for single-input systems with $l = 2$, this yields approximately a 50% reduction of the required expert data).

Index Terms—Data driven control, transfer learning, optimal control, linear systems.

I. INTRODUCTION

IMITATION and transfer learning are popular techniques to learn optimal policies while reducing the amount of labeled data. In imitation learning, an agent is given access to samples of expert (optimal) behavior and seeks to learn a policy that mimics this behavior. In transfer learning, a model trained on one task is used as the starting point for a model on a second related task. The key idea is that certain features learned by the model on the first task can be used as a general-purpose set of features for the second task, allowing the model to learn the second task efficiently. While these techniques have proven useful in multiple learning scenarios, including image classification and natural language processing, their use and utility in control settings have mostly escaped scrutiny.

Manuscript received 15 March 2023; revised 20 May 2023; accepted 7 June 2023. Date of publication 12 June 2023; date of current version 27 June 2023. This work was supported in part by the Office of Naval Research (ONR) under Award N00014-19-1-2264, and in part by the Air Force Office of Scientific Research (AFOSR) under Award FA9550-19-1-0235 and Award FA9550-20-1-0140. Recommended by Senior Editor J. Daafouz. (*Corresponding author: Taosha Guo.*)

Taosha Guo and Fabio Pasqualetti are with the Department of Mechanical Engineering, University of California at Riverside, Riverside, CA 92507 USA (e-mail: tguo@engr.ucr.edu; fabiopas@engr.ucr.edu).

Abed AlRahman Al Makdah is with the Department of Electrical and Computer Engineering, University of California at Riverside, Riverside, CA 92507 USA (e-mail: aalmakdah@engr.ucr.edu).

Vishaal Krishnan is with the School of Engineering and Applied Sciences, Harvard University, Cambridge, MA 02138 USA (e-mail: vkrishnan@seas.harvard.edu).

Digital Object Identifier 10.1109/LCSYS.2023.3285167

In this letter we investigate the use of imitation and transfer learning for Linear Quadratic Gaussian (LQG) control, which seeks a control policy for a stochastic linear system that minimizes the expected value of a quadratic function of the state and input [1]. We consider multiple control tasks, where the system dynamics are fixed but the quadratic cost function varies.¹ We assume that the system dynamics, noise statistics, and cost functions are unknown, and that datasets are available containing optimal input and output trajectories for the different cost functions (source tasks). The questions that we answer include whether it is possible to learn the LQG controllers from expert data, the required size of the dataset, and whether the source datasets can be leveraged to learn the controller for a target task. We show that our data-based controller enjoys a separation property similar to the well-known separation principle [1], and that the lower-dimensional data-based estimation module can be transferred upon changes of the cost function to reduce the amount of expert data required for control design.

Related Work: A number of approaches to direct and indirect data-driven control have recently been proposed. Most approaches focus on learning optimal policies from open-loop data for a fixed task and cost, e.g., see [2], [3], [4]. Differently from these works, this letter considers an imitation and transfer learning framework, where control policies are constructed by imitating expert demonstrations and transferring information across multiple, similar control tasks. Multi-task scenarios have received less attention, with [5], [6] and [7] being recent exceptions for system identification and control design, respectively. In [7], in particular, the notion of a common lower-dimensional representation among the tasks is used to reduce the amount of data required for control design across tasks. Similarly to [7], this letter also exploits a lower-dimensional representation for efficient transfer learning. However, differently from [7] and leveraging [8], this letter focuses on the LQG control problem and provides a precise, quantitative characterization of the lower-dimensional representation for multi-task LQG design from expert demonstrations, as well as tight bounds on the required data. This letter also differs from [9], [10], which study the sample complexity of learning LQG controllers in state-space form from open-loop data.

Contribution of this letter: The main contributions of this letter are as follows. First, we formalize an imitation and transfer learning setting for LQG control. We show that the LQG controller can be learned using an optimal input-output trajectory of length $n(l+2) - 1$, where n denotes the dimension

¹An example of our setting is the control of autonomous vehicles with cost functions that capture different levels of fuel consumption and travel times.

of the system and l the number of outputs. Further, we show that the proposed LQG controller is unique for the case of single-input systems, while it admits multiple representations for multi-input systems. Second, we prove the existence of a data-based separation principle since the data-based LQG controller can be written as the product of two matrices: the estimation matrix, which depends only on the system dynamics, and the controller matrix, which depends on the system dynamics and the quadratic cost function. Further, for the case of single-input systems, we show how the estimation matrix can be learned uniquely using the expert datasets (we also discuss and validate a procedure for the multi-input case). Third, we show how the data-based separation principle can be used for transfer learning because the estimation matrix remains invariant upon changes of the LQG cost function. By doing so, we show that an expert dataset of length $2n+m-1$ is sufficient to learn the LQG controller, thus confirming the benefits of transfer learning also for control design (for instance, for single input systems with $l = 2$, our transfer learning technique reduces the amount of expert data by about 50%). As minor results, we show that the estimation matrix is of full row rank, thus suggesting the minimality of the internal representation, and that the system dimension can be learned using a single expert input-output trajectory of finite length.

II. PROBLEM FORMULATION AND PRELIMINARY RESULTS

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

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

where $x(t) \in \mathbb{R}^n$ denotes the state, $u(t) \in \mathbb{R}^m$ the control input, $y(t) \in \mathbb{R}^l$ the measured output, $w(t)$ the process noise, and $v(t)$ the measurement noise. We assume that the process and measurement noise sequences are independent at all times and satisfy $w(t) \sim \mathcal{N}(0, W)$ and $v(t) \sim \mathcal{N}(0, V)$, with $W \succeq 0$ and $V \succ 0$. Further, we assume that (A, B) and $(A, W^{\frac{1}{2}})$ are controllable, and that (A, C) is observable.

For the system (1), the Linear Quadratic Gaussian (LQG) control problem asks for an input that minimizes the cost

$$\lim_{T \rightarrow \infty} \mathbb{E} \left[\frac{1}{T} \left(\sum_{t=0}^{T-1} x(t)^\top Q x(t) + u(t)^\top R u(t) \right) \right], \quad (2)$$

where $Q \succeq 0$, $R \succ 0$ are weight matrices and T is the control horizon. We assume that $(A, Q^{\frac{1}{2}})$ is observable.

As a classic result [1], the optimal input that solves the LQG problem can be generated by a dynamic controller:

$$\begin{aligned} \hat{x}(t+1) &= E\hat{x}(t) + Fu(t) + Gy(t+1), \\ u(t) &= H\hat{x}(t), \end{aligned} \quad (3)$$

where the controller matrices $E \in \mathbb{R}^{n \times n}$, $F \in \mathbb{R}^{n \times m}$, $G \in \mathbb{R}^{n \times l}$ and $H \in \mathbb{R}^{m \times n}$ can be obtained by combining the Kalman filter for (1) with the static controller that solves the Linear Quadratic Regulator (LQR) problem for (1) with weight matrices Q and R (separation principle). In this case, $\hat{x}(t) \in \mathbb{R}^n$ denotes the estimate of $x(t)$ generated by the Kalman filter and the controller matrices that satisfy

$$\begin{aligned} E &= (I - L_f C)A, & F &= (I - L_f C)B, \\ G &= L_f, & H &= K_{LQR}, \end{aligned} \quad (4)$$

where K_{LQR} and L_f are the LQR and Kalman gains, respectively. Although different choices are possible, we assume that the controller (3) uses the matrices (4) for simplicity and to further highlight the connections between our results and the classic separation-based solution to the LQG control problem. Additionally, we make the following technical assumption.²

Assumption 1 (Observability and Controllability of the Controller): Let $K_{LQR,i}$ be the i -th row of the LQR gain K_{LQR} . Then, the pair $(E, K_{LQR,i})$ is observable for every $i \in \{1, \dots, m\}$, and the pair (E, L_f) is controllable.

The optimal inputs generated by the dynamic controller (3) can also be obtained using a static gain and a finite window of past inputs and outputs [8]. In particular, the optimal LQG inputs u^* satisfy the following relation:

$$u^*(t+n) = K_{LQG} \begin{bmatrix} U_n(t) \\ Y_n(t+1) \end{bmatrix}, \quad (5)$$

where

$$K_{LQG} = H \begin{bmatrix} F_u + E^n F_x^\dagger (I - M_u) & F_y - E^n F_x^\dagger M_y \end{bmatrix}, \quad (6)$$

and $U_n(t)$, $Y_n(t+1)$ are constructed as follows from u^* and its corresponding output y^* from the system (1),

$$U_n(t) = \begin{bmatrix} u^*(t) \\ \vdots \\ u^*(t+n-1) \end{bmatrix}, \quad Y_n(t+1) = \begin{bmatrix} y^*(t+1) \\ \vdots \\ y^*(t+n) \end{bmatrix}, \quad (7)$$

and

$$\begin{aligned} M_u &= \begin{bmatrix} 0 & & & & & \\ HF & & & & & \\ \vdots & \ddots & & & & \\ HE^{n-2}F & \dots & HF & 0 & & \end{bmatrix}, & F_x &= \begin{bmatrix} H \\ HE \\ \vdots \\ HE^{n-1} \end{bmatrix}, \\ F_u &= [E^{n-1}F \quad \dots \quad F], \end{aligned}$$

with M_y and F_y constructed in the same way as M_u and F_u by replacing F by G . The expression (5) is convenient for learning purposes and it will be at the basis of our approach. The next technical result will be useful for our derivations (a proof can be found in the Appendix).

Lemma 1 (Properties of Input-Output Matrices): Let³

$$H_{r,c} = \begin{bmatrix} U_r(t) & \dots & U_r(t+c-1) \\ Y_r(t+1) & \dots & Y_r(t+c) \end{bmatrix}, \quad (8)$$

with $r \in \mathbb{N}_{\geq 0}$, $t \in \mathbb{N}_{\geq 0}$, and $U_r(t)$, $Y_r(t+1)$ as in (7). Then

$$\text{Rank}(H_{r,c}) = \min\{mr + lr, c, n + lr\}.$$

The static LQG controller K_{LQG} can be computed using the static relation (5) and a sufficiently long, yet finite, optimal input-output trajectory. In fact, using optimal input and output sequences, the LQG gain (5) can be written as

$$K_{LQG} = \underbrace{\begin{bmatrix} u^*(t+n) & \dots & u^*(t+n+c-1) \end{bmatrix}}_{\tilde{U}_c} H_{n,c}^\dagger, \quad (9)$$

with $c \geq n(l+1)$. Lemma 1 implies that the data matrix $H_{n,c}$ in (9) is of full row rank for single-input systems. In this case, the gain K_{LQG} is unique and can be reconstructed exactly from

²This assumption is satisfied for generic choices of system parameters.

³This result holds also when the input and output sequences are taken from (3) but are not generated by the optimal LQG compensator, that is, when the compensator is defined with arbitrary matrices E , F , G , and H .

Theorem 2 (Learning L_{est} When $m = 1$): Let D_1, \dots, D_N be the expert trajectories of length $T \geq n(l+2) - 1$. Then,

$$\text{Ker}(L_{\text{est}}) = \bigcap_{i=1}^N \text{Ker}(\bar{U}_{c_s}^i H_{n,c_s}^{i\dagger}), \quad (11)$$

where $\bar{U}_{c_s}^i$ and H_{n,c_s}^i are constructed as in (8) using the expert dataset D_i with $c_s = T - n + 1$.

Proof: Let K_{LQG}^i be the LQG controller of the i -th task. From Theorem 1 we have that $K_{\text{LQG}}^i = K^i L_{\text{est}}$. Then,

$$\begin{aligned} \text{Ker}(K_{\text{LQG}}^i) &= \text{Ker}(L_{\text{est}}) + L_{\text{est}}^\dagger (\text{Im}(L_{\text{est}}) \cap \text{Ker}(K^i)) \\ &= \text{Ker}(\bar{U}_{c_s}^i H_{n,c_s}^{i\dagger}), \end{aligned}$$

where the last equality is due to Lemma 1 since $m = 1$, H_{n,c_s}^i is of full row rank and $K_{\text{LQG}}^i = \bar{U}_{c_s}^i H_{n,c_s}^{i\dagger}$. The claimed statement now follows from Assumption 2. ■

From Theorem 2, the kernel of the estimation matrix L_{est} can be learned from a finite number of LQG datasets, with each dataset comprising optimal input and output trajectories of finite length. Hence, the estimation matrix L_{est} can also be learned up to multiplication by an invertible matrix using a basis of the orthogonal complement to $\text{Ker}(L_{\text{est}})$. That is,

$$L_{\text{est}} = P \cdot \underbrace{\text{Basis} \left(\left(\bigcap_{i=1}^N \text{Ker}(\bar{U}_{c_s}^i H_{n,c_s}^{i\dagger}) \right)^\perp \right)}_{\hat{L}_{\text{est}}}, \quad (12)$$

for some invertible matrix P . Then, using Theorem 1 and for any choice of the weight matrices Q and R , the LQG controller for (1) can always be written as the product $KP^{-1}\hat{L}_{\text{est}}$, where only the matrix $\hat{K} = KP^{-1}$ depends on the weight matrices Q and R and, from Theorem 2, the estimation matrix \hat{L}_{est} can be learned given a sufficiently large and diverse dataset of expert trajectories. These observations imply that the controller for the target LQG task can be computed by simply learning the control matrix \hat{K}_{target} as a solution to the linear system

$$\bar{U}_{c_t}^{\text{target}} = \hat{K}_{\text{target}} \hat{L}_{\text{est}} H_{n,c_t}^{\text{target}}, \quad (12)$$

where $\bar{U}_{c_t}^{\text{target}}$ and $H_{n,c_t}^{\text{target}}$ are constructed as in (8) from the target dataset D_{target} , with $c_t = \bar{T} - n + 1$. The next result quantifies the length of the expert trajectory D_{target} . We make the following assumption on the target dataset

Assumption 3 (Persistence of Excitation): For every value of c_t , the target dataset satisfies

$$\text{Ker}(L_{\text{est}}) \cap \text{Im}(H_{n,c_t}^{\text{target}}) = \{0\}.$$

We remark that Assumption 3 is generically satisfied since the entries of $H_{n,c_t}^{\text{target}}$ are driven by the system noise.

Theorem 3 (Length of Expert Trajectory to Learn the Target LQG Controller): Let \bar{T} be the length of the expert trajectory in D_{target} . The LQG controller for the target task can be learned whenever $\bar{T} \geq 2n + m - 1$.

Proof: We first show that $\text{Rank}(F_y - (a \otimes I_n)\tilde{M}_y) = n$, which implies that L_{est} is of full row rank. With standard

manipulation, the matrix $F_y - (a \otimes I_n)\tilde{M}_y$ can be rewritten as

$$\underbrace{\begin{bmatrix} E^{n-1} & \dots & I \end{bmatrix}}_J \underbrace{\begin{bmatrix} 1 & & & & \\ -a_{n-1} & 1 & & & \\ \vdots & \ddots & \ddots & \ddots & \\ -a_1 & \dots & -a_{n-1} & 1 \end{bmatrix}}_S \underbrace{\begin{bmatrix} G & & & \\ & \ddots & & \\ & & G & \end{bmatrix}}_O.$$

Notice that S is invertible and that $\text{Rank}(JO) = \text{Rank}(F_y) = n$ due to Assumption 1. Then, $\text{Rank}(F_y - (a \otimes I_n)\tilde{M}_y) = \text{Rank}(JSO) = \text{Rank}(JO) = n$. Due to Assumption 3 and Lemma 1, the matrix $H_{n,c_t}^{\text{target}}$ has full column rank $n + m$ ($n + m \leq n(l+1)$) when $c_t = n + m$ (equivalently, $\bar{T} = 2n + m - 1$) and $\text{Ker}(L_{\text{est}}) \cap \text{Im}(H_{n,c_t}^{\text{target}}) = \{0\}$. Thus, $L_{\text{est}} H_{n,c_t}^{\text{target}}$ is invertible, and finally $K_{\text{target}} = \bar{U}_{\text{target}} (L_{\text{est}} H_{n,c_t}^{\text{target}})^{-1}$. ■

Using (9), we notice that the LQG controller can be learned uniquely from a single trajectory of length $T \geq n(l+2) - 1$ for the single-input case, since the data matrix in (9) becomes of full row rank. This bound reflects the complexity of learning the LQG controller in an imitation learning framework. On the other hand, leveraging the separation principle in Theorem 1, the matrix \hat{L}_{est} can be learned from $N \geq n + 1$ expert datasets and used to learn the LQG controller of any target task. By doing so, Theorem 3 states that the expert trajectory of the target task needs only to be of length $2n$ for the single-input case. This reduced bound reflects the benefits of the imitation and transfer learning setting, where data from earlier tasks is used to solve future LQG tasks. For instance, when $m = 1$ and $l = 2$, the imitation and transfer approach requires about 50% less expert data compared to the imitation approach alone.

Remark 1 (Learning the Dimension of the System From Data): The reconstruction of the LQG gain in (9) and of the estimation matrix in Theorem 2 requires the knowledge of the dimension of the system to properly construct the required matrices. If unknown, the dimension of the system can be learned by solving the following minimization problem:

$$n = \min\{r \in \mathbb{N} : \text{Rank}(H_{r,r}) = \text{Rank}(H_{r+1,r+1})\}.$$

This follows from Lemma 1, since the rank of $H_{r,r}$ equals $n(l+1)$ when $r = n$ and the value of l can be easily inferred from the expert data (l equals the dimension of y^*). We note that this remark is also valid for multi-input systems.

Remark 2 (Learning L_{est} When $m > 1$): Lemma 1 implies that the data matrix in (9) is not full row rank when $m > 1$ and that the LQG gain in (5) is not unique. Although the decomposition in Theorem 1 still holds, the computation of L_{est} from data is more involved than the procedure presented in Theorem 2. Here we discuss two different ways for this computation but, in the interest of space and clarity, we leave a detailed treatment for future research. First, let K_{LQG}^i be the LQG controller of the i -th source task. Then,

$$K_{\text{LQG}}^i = \bar{U}_{c_s}^i H_{n,c_s}^{i\dagger} + X_i Z_i = K^i L_{\text{est}} \quad (13)$$

for some matrix X_i , where Z_i is a basis of the left null space of H_{n,c_s}^i . By stacking these expressions together we obtain

$$\begin{bmatrix} K^1 \\ \vdots \\ K^N \end{bmatrix} L_{\text{est}} = \begin{bmatrix} \bar{U}_{c_s}^1 H_{n,c_s}^{1\dagger} + X_1 Z_1 \\ \vdots \\ \bar{U}_{c_s}^N H_{n,c_s}^{N\dagger} + X_N Z_N \end{bmatrix}. \quad (14)$$

Since L_{est} has $n+m$ rows, where n is obtained from Remark 1, the row space of every gain K_{LQG}^i must belong to the same $(n+m)$ -dimensional subspace. Then, the matrix on the right hand side of (14) must have a left null space of dimension at least $mN - (n+m)$ for an appropriate choice of the matrices X_1, \dots, X_N . This condition can be used to find the matrices X_1, \dots, X_N that satisfy (13) for a sufficiently large number N . Finally, similar to Theorem 2,

$$\text{Ker}(L_{\text{est}}) = \bigcap_{i=1}^N \text{Ker}(U_{n,c_s}^i H_{n,c_s}^{i\dagger} + X_i Z_i). \quad (15)$$

Second, using the notation in Theorem 2 and (12) and the fact that $\text{vec}(U_{c_s}^i) = (H_{n,c_s}^{i\dagger} \otimes K^i) \text{vec}(L_{\text{est}})$, L_{est} can also be computed by solving the following bi-linear problem:

$$\min_{L, K^1, \dots, K^N} \sum_{i=1}^N \|\text{vec}(\tilde{U}_{c_s}^i) - (H_{n,c_s}^{i\dagger} \otimes K^i) \text{vec}(L)\|. \quad (16)$$

The convergence properties of the two approaches above deserve a full discussion that is beyond the scope of this letter; in the next section we provide some numerical evidence.

IV. ILLUSTRATIVE EXAMPLE

We use the following model of a batch reactor system that is open-loop unstable:

$$A = \begin{bmatrix} 1.178 & 0.001 & 0.511 & -0.403 \\ -0.051 & 0.661 & -0.011 & 0.061 \\ 0.076 & 0.335 & 0.560 & 0.382 \\ 0 & 0.335 & 0.089 & 0.849 \end{bmatrix}, B = \begin{bmatrix} 0.004 \\ 0.467 \\ 0.213 \\ 0.213 \end{bmatrix}, \\ C = [-0.44 \quad -0.51 \quad 0.09 \quad 0.44], \quad (17)$$

with process and measurement noise covariance $W = 1.5I_4$ and $V = 0.6$. The weight matrices of the target task are

$$Q_{\text{target}} = \begin{bmatrix} 6 & 1 & 1 & -3 \\ 1 & 1 & 0 & -1 \\ 1 & 0 & 3 & 0 \\ -3 & -1 & 0 & 2 \end{bmatrix}, \text{ and } R_{\text{target}} = 1.$$

We compare the model-based approach in Theorem 1 with the data-driven approach in Theorem 2. Using (6) we obtain

$$K_{\text{LQG}}^{\text{target}} = [-0.01 \quad 0.16 \quad -0.54 \quad 1.02 \quad 2.6 \quad -13.34 \quad 21.25 \quad -10.60].$$

For our data-based approach, we have collected expert trajectories D_1, \dots, D_N of length $T = 11$ from $N = 5$ source tasks, with weighting matrices $Q_i = iI_4$ and $R_i = I_2$ respectively. Using Theorem 2, we compute the estimation matrix \hat{L}_{est} as

$$\begin{bmatrix} -0.01 & 0.09 & -0.30 & 0.45 & 0.08 & -0.42 & 0.65 & -0.30 \\ 0.02 & -0.18 & 0.54 & -0.61 & 0.07 & -0.30 & 0.42 & -0.19 \\ 0.05 & -0.34 & 0.57 & 0.56 & 0.12 & -0.28 & -0.10 & 0.38 \\ -0.04 & 0.23 & -0.32 & -0.29 & 0.32 & -0.60 & -0.17 & 0.52 \\ -0.05 & 0.17 & 0.03 & -0.04 & -0.43 & 0.26 & 0.54 & 0.65 \end{bmatrix}$$

It can be verified that $K_{\text{LQG}}^{\text{target}T} \in \text{Im}(\hat{L}_{\text{est}}^T)$, that is, there exists a matrix \hat{K}_{target} such that $K_{\text{LQG}}^{\text{target}} = \hat{K}_{\text{target}} \hat{L}_{\text{est}}$. This verifies that the estimation matrix \hat{L}_{est} generates an internal representation from which the LQG inputs can be computed.

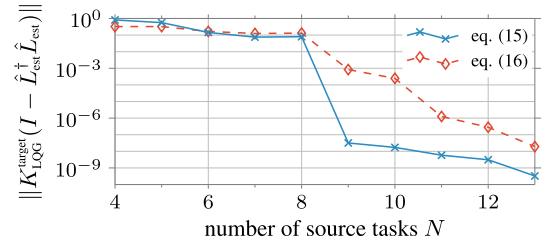


Fig. 1. This figure shows the error $\|K_{\text{LQG}}^{\text{target}}(I - \hat{L}_{\text{est}}^{\dagger} \hat{L}_{\text{est}})\|$ as a function of the number of source tasks. The error converges for (15) and (16) as the number of the source tasks increases, which implies that both approaches in Remark (2) reconstruct exactly the estimation matrix.

Consider now the same system (17) with two inputs, where the new input matrix and its corresponding cost matrix are:

$$B = \begin{bmatrix} 0.004 & -0.087 \\ 0.467 & 0.001 \\ 0.213 & -0.235 \\ 0.213 & -0.016 \end{bmatrix}, \text{ and } R_{\text{target}} = \begin{bmatrix} 1 & 0 \\ 0 & 4 \end{bmatrix}.$$

We follow the same steps as in the single input case to compute $K_{\text{LQG}}^{\text{target}}$ using the model-based approach in (6), and then following the procedures in Remark (2), we compute \hat{L}_{est} using (15) and (16) respectively. In Fig. 1 we plot the error $\|K_{\text{LQG}}^{\text{target}}(I - \hat{L}_{\text{est}}^{\dagger} \hat{L}_{\text{est}})\|$ for both approaches as the number of source tasks increases. The convergence of the error implies that \hat{L}_{est} obtained using the methods in Remark (2) becomes the correct estimation matrix for the target LQG controller.

V. CONCLUSION

In this letter we study an imitation and transfer learning setting for LQG control, where expert input-output trajectories are used to learn a data-based LQG controller. We show how the LQG controller can be computed from data, quantify the length of the expert trajectories needed to learn the controller, and show how the controller can be decomposed as the product of an estimation matrix, which depends only on the system dynamics, and a controller matrix, which depends also on the LQG cost. This separation principle allows us to reuse the estimation matrix across different LQG tasks, thus reducing the length of the required expert trajectories. Aspects of this letter requiring additional investigation include a detailed treatment of the multi-input case, the study of transfer methods when the system dynamics also change, the extension to more general optimal control problems, and a proof of the minimality of the proposed internal representation.

APPENDIX

A. Proof of Lemma 1

Proof: From (3) and (7) we have

$$U_r(t) = \underbrace{\begin{bmatrix} H \\ H\bar{E} \\ \vdots \\ H\bar{E}^{r-1} \end{bmatrix}}_{\bar{F}_x} \hat{x}(t) + \underbrace{\begin{bmatrix} 0 & 0 & \dots & 0 \\ HG & 0 & \dots & 0 \\ \vdots & \ddots & \ddots & \vdots \\ H\bar{E}^{r-2}G & \dots & HG & 0 \end{bmatrix}}_{\bar{F}_y} Y_r(t+1),$$

where $\bar{E} = E + FH$. Then, we obtain

$$H_{r,c} = \underbrace{\begin{bmatrix} \bar{F}_x & \bar{F}_y \\ 0 & I \end{bmatrix}}_M \underbrace{\begin{bmatrix} \hat{x}(t) & \cdots & \hat{x}(t+c-1) \\ Y_r(t+1) & \cdots & Y_r(t+c) \end{bmatrix}}_N.$$

Further, $\text{Rank}(H_{r,c}) \leq \min\{\text{Rank}(M), \text{Rank}(N)\}$, and $\text{Rank}(H_{r,c}) = \text{Rank}(M)$ whenever N is of full row rank [11]. Notice that $\text{Rank}(M) \leq \min\{mr + lr, n + lr\}$, and $\text{Rank}(M) = n + lr$ if $mr \geq n$ and the pair (\bar{E}, H) is observable. To conclude, [12, Corollary 2] implies that $\text{Rank}(N) = \min\{n + lr, c\}$. ■

B. Proof of Theorem 1

We start with an alternative expression for K_{LQG} .

Lemma 2 (Alternative Expression for K_{LQG}): Let $K_{\text{LQG},i}$ and $K_{\text{LQR},i}$ be the i -th row of K_{LQG} and K_{LQR} , respectively, and define the matrices P_i such that

$$P_i \underbrace{\begin{bmatrix} K_{\text{LQR}} \\ K_{\text{LQR}}E \\ \vdots \\ K_{\text{LQR}}E^{n-1} \end{bmatrix}}_{F_x} = \underbrace{\begin{bmatrix} K_{\text{LQR},i} \\ K_{\text{LQR},i}E \\ \vdots \\ K_{\text{LQR},i}E^{n-1} \end{bmatrix}}_{F_x^i}.$$

for all $i \in \{1, \dots, m\}$. Then,

$$K_{\text{LQG},i} = K_{\text{LQR},i} \begin{bmatrix} F_u + E^n F_x^{i\dagger} (P_i - M_u^i) & F_y - E^n F_x^{i\dagger} M_y^i \end{bmatrix}, \quad (18)$$

where $F_x^i = P_i F_x$, $M_u^i = P_i M_u$, and $M_y^i = P_i M_y$.

Proof: Using the compensator dynamics (3) we obtain

$$U_n(t) = F_x \hat{x}(t) + M_u U_n(t) + M_y Y_n(t+1),$$

and

$$\hat{x}(t+n) = E^n \hat{x}(t) + F_u U_n(t) + F_y Y_n(t+1). \quad (19)$$

Due to Assumption 1, F_x^i is invertible so that

$$\hat{x}(t) = F_x^{i\dagger} \left((P_i - M_u^i) U_n(t) - M_y^i Y_n(t+1) \right),$$

for any $i \in \{1, \dots, m\}$ and, consequently,

$$\begin{aligned} \hat{x}(t+n) &= E^n F_x^{i\dagger} \left((P_i - M_u^i) U_n(t) - M_y^i Y_n(t+1) \right) \\ &\quad + F_u U_n(t) + F_y Y_n(t+1). \end{aligned} \quad (20)$$

Notice that the gain $K_{\text{LQR},i}$ must satisfy, at all times,

$$K_{\text{LQR},i} \hat{x}(t+n) = K_{\text{LQG},i} \begin{bmatrix} U_n(t) \\ Y_n(t+1) \end{bmatrix}.$$

Substituting (20) into $\hat{x}(t+n)$ in (19) yields the result. ■

We are now ready to prove Theorem 1.

Proof of Theorem 1: Notice that (18) can be rewritten as

$$K_{\text{LQG},i} = \begin{bmatrix} K_{\text{LQR},i} & K_{\text{LQR},i} \end{bmatrix} \begin{bmatrix} F_u & F_y \\ E^n F_x^{i\dagger} (I - M_u^i) & -E^n F_x^{i\dagger} M_y^i \end{bmatrix}.$$

Further, using the Cayley-Hamilton Theorem, we have

$$\begin{aligned} K_{\text{LQR},i} E^n &= K_{\text{LQR},i} (a_0 I_n + a_1 E + \cdots + a_{n-1} E^{n-1}) \\ &= \underbrace{\begin{bmatrix} a_0 & a_1 & \cdots & a_{n-1} \end{bmatrix}}_a F_x^i, \end{aligned}$$

where a_0, \dots, a_{n-1} are the negative coefficients of the characteristic polynomial of E . Then, since F_x^i is invertible (Assumption 1), we have $K_{\text{LQR},i} E^n F_x^{i\dagger} (P_i - M_u^i) = a(P_i - M_u^i)$ and $K_{\text{LQR},i} E^n F_x^{i\dagger} M_y^i = a M_y^i$, and (18) becomes

$$K_{\text{LQG},i} = \begin{bmatrix} K_{\text{LQR},i} & 1 \end{bmatrix} \begin{bmatrix} F_u & F_y \\ a(P_i - M_u^i) & -a M_y^i \end{bmatrix}. \quad (21)$$

Notice that

$$M_u^i = \underbrace{\begin{bmatrix} K_{\text{LQR},i} & & \\ & \ddots & \\ & & K_{\text{LQR},i} \end{bmatrix}}_{K_{\text{diag}}} \tilde{M}_u, \quad M_y^i = \begin{bmatrix} K_{\text{LQR},i} & & \\ & \ddots & \\ & & K_{\text{LQR},i} \end{bmatrix} \tilde{M}_y,$$

where \tilde{M}_u and \tilde{M}_y are defined in (10), and

$$a K_{\text{diag}} = \begin{bmatrix} a_0 K_{\text{LQR},i} & \cdots & a_{n-1} K_{\text{LQR},i} \end{bmatrix} = K_{\text{LQR},i} (a \otimes I_n),$$

Thus, (21) becomes

$$K_{\text{LQG},i} = \begin{bmatrix} K_{\text{LQR},i} & 1 \end{bmatrix} \begin{bmatrix} F_u - (a \otimes I_n) \tilde{M}_u & F_y - (a \otimes I_n) \tilde{M}_y \\ a P_i & 0 \end{bmatrix}.$$

By using $[(aP_1)^\top \cdots (aP_m)^\top]^\top = a \otimes I_m$, we obtain

$$\begin{bmatrix} K_{\text{LQG},1} \\ \vdots \\ K_{\text{LQG},m} \end{bmatrix} = \begin{bmatrix} K_{\text{LQR}} & I_m \end{bmatrix} \begin{bmatrix} F_u - (a \otimes I_n) \tilde{M}_u & F_y - (a \otimes I_n) \tilde{M}_y \\ a \otimes I_m & 0 \end{bmatrix}.$$

This concludes the proof of Theorem 1. ■

REFERENCES

- [1] K. Zhou, J. C. Doyle, and K. Glover, *Robust and Optimal Control*. Upper Saddle River, NJ, USA: Prentice-Hall, 1996.
- [2] B. Recht, "A tour of reinforcement learning: The view from continuous control," *Annu. Rev. Control Robot. Auton. Syst.*, vol. 2, pp. 253–279, May 2019.
- [3] K. Zhang, B. Hu, and T. Başar, "Policy optimization for \mathcal{H}_2 linear control with \mathcal{H}_∞ robustness guarantee: Implicit regularization and global convergence," in *Proc. Int. Conf. Mach. Learn.*, vol. 120, Jun. 2020, pp. 179–190.
- [4] I. Markovsky and F. Dörfler, "Behavioral systems theory in data-driven analysis, signal processing, and control," *Annu. Rev. Control*, vol. 52, pp. 42–64, Dec. 2021.
- [5] L. Xin, L. Ye, G. Chiu, and S. Sundaram, "Identifying the dynamics of a system by leveraging data from similar systems," in *Proc. Amer. Control Conf.*, Atlanta, GA, USA, Jun. 2022, pp. 818–824.
- [6] Y. Chen, A. M. Ospina, F. Pasqualetti, and E. Dall'Anese, "Multi-task system identification of similar linear time-invariant dynamical systems," 2023, *arXiv:2301.01430*.
- [7] T. T. Zhang et al., "Multi-task imitation learning for linear dynamical systems," 2022, *arXiv:2212.00186*.
- [8] A. A. Al Makkah, V. Krishnan, V. Katewa, and F. Pasqualetti, "Behavioral feedback for optimal LQG control," in *Proc. IEEE Conf. Decis. Control*, Cancún, Mexico, Dec. 2022, pp. 4660–4666.
- [9] S. Lale, K. Azizzadenesheli, B. Hassibi, and A. Anandkumar, "Logarithmic regret bound in partially observable linear dynamical systems," in *Proc. Int. Conf. Adv. Neural Inf. Process. Syst.*, vol. 33, Dec. 2020, pp. 20876–20888.
- [10] Y. Zheng, L. Furieri, M. Kamgarpour, and N. Li, "Sample complexity of linear quadratic Gaussian (LQG) control for output feedback systems," in *Proc. Int. Conf. Mach. Learn.*, vol. 144, Jun. 2021, pp. 559–570.
- [11] R. A. Horn and C. R. Johnson, *Matrix Analysis*. Cambridge, U.K.: Cambridge Univ. Press, 1985.
- [12] J. C. Willems, P. Rapisarda, I. Markovsky, and B. L. M. De Moor, "A note on persistency of excitation," *Syst. Control Lett.*, vol. 54, no. 4, pp. 325–329, 2005.