Proceedings of Machine Learning Research vol 144:1–12, 2021 3rd Annual Conference on Learning for Dynamics and Control

Sample Complexity of Linear Quadratic Gaussian (LQG) Control for Output Feedback Systems

Yang Zheng * ZHENGY@G.HARVARD.EDU †
School of Engineering and Applied Sciences, Harvard University, USA

Luca Furieri * LUCA.FURIERI@EPFL.CH ‡
Automatic Control Laboratory, ETH Zurich, Switzerland,
Laboratoire d'Automatique, EPFL, Switzerland

Maryam Kamgarpour MARYAMK@ECE.UBC.CA ‡
Electrical and Computer Engineering, University of British Columbia, Canada

Na Li NALI@SEAS.HARVARD.EDU †
School of Engineering and Applied Sciences, Harvard University, USA

Editors: A. Jadbabaie, J. Lygeros, G. J. Pappas, P. A. Parrilo, B. Recht, C. J. Tomlin, M. N. Zeilinger

Abstract

This paper studies a class of partially observed Linear Quadratic Gaussian (LQG) problems with *unknown* dynamics. We establish an end-to-end sample complexity bound on learning a robust LQG controller for open-loop stable plants. This is achieved using a robust synthesis procedure, where we first estimate a model from a single input-output trajectory of finite length, identify an H-infinity bound on the estimation error, and then design a robust controller using the estimated model and its quantified uncertainty. Our synthesis procedure leverages a recent control tool called Input-Output Parameterization (IOP) that enables robust controller design using convex optimization. For open-loop stable systems, we prove that the LQG performance degrades *linearly* with respect to the model estimation error using the proposed synthesis procedure. Despite the hidden states in the LQG problem, the achieved scaling matches previous results on learning Linear Quadratic Regulator (LQR) controllers with full state observations.

1. Introduction

There has been a surging interest in applying machine learning techniques to the control of dynamical systems with continuous action spaces (see e.g., Duan et al. (2016); Recht (2019)). An increasing body of recent studies have started to address theoretical and practical aspects of deploying learning-based control policies in dynamical systems (Recht, 2019). An extended review of related work is given in Appendix A of Zheng* and Furieri* et al. (2020)[1].

For data-driven reinforcement learning (RL) control, the existing algorithmic frameworks can be broadly divided into two categories: (a) model-based RL, in which an agent first fits a model for the system dynamics from observed data and then uses this model to design a policy using either the certainty equivalence principle (Åström and Wittenmark, 2013; Tu and Recht, 2019; Mania et al.,

* Yang Zheng and Luca Furieri contributed equally to this work.
† Yang Zheng and Na Li are supported by NSF career, AFOSR YIP, and ONR YIP.
‡ Luca Furieri and Maryam Kamgarpour are gratefully supported by ERC Starting Grant CONENE.
1. The report Zheng* and Furieri* et al. (2020) also contains technical proofs and complementary discussion.

© 2021 Y. Zheng, L. Furieri, M. Kamgarpour & N. Li.

2019) or classical robust control tools (Zhou et al., 1996; Dean et al., 2019; Tu et al., 2017); and (b) model-free RL, in which the agent attempts to learn an optimal policy directly from the data without explicitly building a model for the system (Fazel et al., 2018; Malik et al., 2018; Furieri et al., 2020). Another interesting line of work formulates model-free control by using past trajectories to predict future trajectories based on so called *fundamental lemma* (Coulson et al., 2019; Berberich et al., 2019; De Persis and Tesi, 2019). For both model-based and model-free methods, it is critical to establish formal guarantees on their sample efficiency, stability and robustness. Recently, the Linear Quadratic Regulator (LQR), one of the most well-studied optimal control problems, has been adopted as a benchmark to understand how machine learning interacts with continuous control (Dean et al., 2019; Tu and Recht, 2019; Mania et al., 2019; Fazel et al., 2018; Recht, 2019; Dean et al., 2018; Malik et al., 2018). It was shown that the simple certainty equivalent model-based method requires asymptotically less samples than model-free policy gradient methods for LQR (Tu and Recht, 2019). Besides, the certainty equivalent control (Mania et al., 2019) (scaling as $\mathcal{O}(N^{-1})$, where N is the number of samples) is more sample-efficient than robust model-based methods (scaling as $\mathcal{O}(N^{-1/2})$) that account for uncertainty explicitly (Dean et al., 2019).

In this paper, we take a step further towards a theoretical understanding of model-based learning methods for Linear Quadratic Gaussian (LQG) control. As one of the most fundamental control problems, LQG deals with partially observed linear dynamical systems driven by additive white Gaussian noises (Zhou et al., 1996). As a significant challenge compared to LQR, the internal system states cannot be directly measured for learning and control purposes. When the system model is known, LQG admits an elegant closed-form solution, combining a Kalman filter together with an LQR feedback gain (Bertsekas, 2011; Zhou et al., 1996). For unknown dynamics, however, much fewer results are available for the achievable closed-loop performance. One natural solution is the aforementioned *certainty equivalence principle*: collect some data of the system evolution, fit a model, and then solve the original LQG problem by treating the fitted model as the truth (Åström and Wittenmark, 2013). It has been recently proved in Mania et al. (2019) that this certainty equivalent principle enjoys a good statistical rate for sub-optimality gap that scales as the *square* of the model estimation error. However, this procedure does not come with a robust stability guarantee, and it might fail to stabilize the system when data is not sufficiently large. Sample-complexity of Kalman filters has also been recently characterized in Tsiamis et al. (2020).

Leveraging recent advances in control synthesis (Furieri et al., 2019) and non-asymptotic system identification (Oymak and Ozay, 2019; Tu et al., 2017; Sarkar et al., 2019; Zheng and Li, 2020), we establish an end-to-end sample-complexity result of learning LQG controllers that robustly stabilize the true system with a high probability. In particular, our contribution is on developing a novel tractable robust control synthesis procedure, whose sub-optimality can be tightly bounded as a function of the model uncertainty. By incorporating a non-asymptotic $\mathcal{H}_\infty$ bound on the system estimation error, we establish an end-to-end sample complexity bound for learning robust LQG controllers. Dean et al. (2019) performed similar analysis for learning LQR controllers with full state measurements. Instead, our method includes noisy output measurements without reconstruting an internal state-space representation for the system. Despite the challenge of hidden states, for open-loop stable systems, our method achieves the same scaling for the sub-optimality gap as Dean et al. (2019), that is, $\mathcal{O}(\epsilon)$, where ϵ is the model uncertainty level. Specifically, the highlights of our work include:

- Our design methodology is suitable for general multiple-input multiple-output (MIMO) LTI systems that are open-loop stable. Based on a recent control tool, called Input-Ouput Parameterization (IOP) (Furieri et al., 2019), we derive a new convex parameterization of robustly stabilizing controllers. Any feasible solution from our procedure corresponds to a controller that is robust against model uncertainty. Our framework directly aims for a class of general LQG problems, going beyond the recent results (Dean et al., 2019; Boczar et al., 2018) that are built on the system-level parameterization (SLP) (Wang et al., 2019).

- We quantify the performance degradation of the robust LQG controller, scaling *linearly* with the model error, which is consistent with Dean et al. (2019); Boczar et al. (2018). Our analysis requires a few involved bounding arguments in the IOP framework (Furieri et al., 2019) due to the absence of direct state measurements. We note that this linear scaling is inferior to the simple certainty equivalence controller (Mania et al., 2019), for which the performance degradation scales as the *square* of parameter errors for both LQR and LQG, but without guarantees on the robust stablility against model errors. This brings an interesting trade-off between optimality and robustness, which is also observed in the LQR case (Dean et al., 2019).

The rest of this paper is organized as follows. We introduce Linear Quadratic Gaussian (LQG) control for unknown systems and overview our contributions in Section 2. In Section 3, we first leverage the IOP framework to develop a robust controller synthesis procedure taking into account estimation errors explicitly, and then derive our main sub-optimality result. This enables our end-to-end sample complexity analysis discussed in Section 4. We conclude this paper in Section 5. Proofs are postponed to the appendices of Zheng* and Furieri* et al. (2020)

Notation. We use lower and upper case letters (*e.g.* x and A) to denote vectors and matrices, and lower and upper case boldface letters (*e.g.* $\mathbf{x}$ and $\mathbf{G}$) are used to denote signals and transfer matrices, respectively. Given a stable transfer matrix $\mathbf{G} \in \mathcal{RH}_\infty$, where $\mathcal{RH}_\infty$ denotes the subspace of stable transfer matrices, we denote its $\mathcal{H}_\infty$ norm by $\|\mathbf{G}\|_\infty := \sup_\omega \sigma_{\max}(\mathbf{G}(e^{j\omega}))$.

2. Problem Statement and Our contributions

2.1. LQG formulation

We consider the following *partially observed* output feedback system

$$\begin{aligned} x_{t+1} &= A_\star x_t + B_\star u_t, \\ y_t &= C_\star x_t + v_t, \\ u_t &= \pi(y_t, \ldots, y_0) + w_t\,, \end{aligned} \tag{1}$$

where $x_t \in \mathbb{R}^n$ is the state of the system, $u_t \in \mathbb{R}^m$ is the control input and $\pi(\cdot)$ is an output-feedback control policy, $y_t \in \mathbb{R}^p$ is the observed output, and $w_t \in \mathbb{R}^m, v_t \in \mathbb{R}^p$ are Gaussian noise with zero-mean and covariance $\sigma_w^2 I, \sigma_v^2 I$. The setup in (1) is convenient from an external input-output perspective, where the noise w_t affects the input u_t directly.[2] This setup was also considered in Tu et al. (2017); Boczar et al. (2018); Zheng et al. (2019). When $C = I, v_t = 0$, *i.e.*, the state x_t is directly measured, the system is called *fully observed*. Throughout this paper, we make the following assumption.

Assumption 1 *$(A_\star, B_\star)$ is stabilizable and $(C_\star, A_\star)$ detectable.*

2. Letting $\hat{w}_t = B_\star w_t$, (1) is an instance of the classical LQG formulation where the process noise $\hat{w}_t$ has variance $\sigma_w^2 B_\star B_\star^{\mathsf{T}}$. The setting (1) enables a concise closed-loop representation in (6), facilitating the suboptimality analysis in the IOP framework. We leave the general case with unstructured covariance for future work.

The classical Linear Quadratic Gaussian (LQG) control problem is defined as

$$\begin{aligned} \min_{u_0,u_1,\dots} \quad & \lim_{T\to\infty} \mathbb{E}\left[\frac{1}{T}\sum_{t=0}^{T}\left(y_t^{\mathsf{T}} Q y_t + u_t^{\mathsf{T}} R u_t\right)\right] \\ \text{subject to} \quad & (1), \end{aligned} \tag{2}$$

where Q and R are positive definite. Without loss of generality and for notational simplicity, we assume that $Q = I_p$, $R = I_m$, $\sigma_w = \sigma_v = 1$. When the dynamics (1) are known, this problem has a well-known closed-form solution by solving two algebraic Riccati equations (Zhou et al., 1996). The optimal solution is $u_t = K\hat{x}_t$ with a fixed $p \times n$ matrix K and $\hat{x}_t$ is the state estimation from the observation $y_0, \dots, y_t$ using the Kalman filter. Compactly, the optimal controller to (2) can be written in the form of transfer function $\mathbf{u}(z) = \mathbf{K}(z)\mathbf{y}(z)$, with $\mathbf{K}(z) = C_k(zI - A_k)^{-1}B_k + D_k$, where $z \in \mathbb{C}$, $\mathbf{u}(z)$ and $\mathbf{y}(z)$ are the $z-$transform of the input u_t and output y_t, and the transfer function $K(z)$ has a state-space realization expressed as

$$\begin{aligned} \xi_{t+1} &= A_k \xi_t + B_k y_t, \\ u_t &= C_k \xi_t + D_k y_t, \end{aligned} \tag{3}$$

where $\xi \in \mathbb{R}^q$ is the controller internal state, and A_k, B_k, C_k, D_k depends on the system matrices A, B, C and solutions to algebraic Riccati equations. We refer interested readers to Zhou et al. (1996); Bertsekas (2011) for more details.

Throughout the paper, we make another assumption, which is required in both plant estimation algorithms (Oymak and Ozay, 2019) and the robust controller synthesis phase.

Assumption 2 *The plant dynamics are open-loop stable,* i.e.*,* $\rho(A_\star) < 1$*, where* $\rho(\cdot)$ *denotes the spectral radius.*

2.2. LQG for unknown dynamics

In the case where the system dynamics $A_\star, B_\star, C_\star$ are unknown, one natural idea is to conduct experiments to estimate $\hat{A}, \hat{B}, \hat{C}$ (Ljung, 2010) and to design the corresponding controller based on the estimated dynamics, which is known as *certainty equivalence control*. When the estimation is accurate enough, the certainty equivalent controller leads to good closed-loop performance (Mania et al., 2019). However, the certainty equivalent controller does not take into account estimation errors, which might lead to instability in practice. It is desirable to explicitly incorporate the estimation errors

$$\|\hat{A} - A_\star\|, \quad \|\hat{B} - B_\star\|, \quad \|\hat{C} - C_\star\|, \tag{4}$$

into the controller synthesis, and this requires novel tools from robust control. Unlike the fully observed LQR case (Dean et al., 2019), in the partially observed case of (1), it remains unclear how to directly incorporate state-space model errors (4) into robust controller synthesis. Besides, the state-space realization of a partially observed system is not unique, and different realizations from the estimation procedure might have an impact on the controller synthesis.

Instead of the state-space form (1), the system dynamics can be described uniquely in the frequency domain in terms of the transfer function as

$$\mathbf{G}_\star(z) = C_\star(zI - A_\star)^{-1}B_\star\,,$$

where $z \in \mathbb{C}$. Based on an estimated model $\hat{\mathbf{G}}$ and an upper bound ϵ on its estimation error $\|\mathbf{\Delta}\|_\infty := \|\mathbf{G}_\star - \hat{\mathbf{G}}\|_\infty$, we consider a robust variant of the LQG problem that seeks to minimize the worst-case LQG performance of the closed-loop system

$$\begin{aligned} \min_{\mathbf{K}} \sup_{\|\mathbf{\Delta}\|_\infty < \epsilon} \quad & \lim_{T\to\infty} \mathbb{E}\left[\frac{1}{T}\sum_{t=0}^{T}\left(y_t^\mathsf{T} Q y_t + u_t^\mathsf{T} R u_t\right)\right], \\ \text{subject to} \quad & \mathbf{y} = (\hat{\mathbf{G}} + \mathbf{\Delta})\mathbf{u} + \mathbf{v}\,, \quad \mathbf{u} = \mathbf{K}\mathbf{y} + \mathbf{w}, \end{aligned} \tag{5}$$

where $\mathbf{K}$ is a proper transfer function. When $\mathbf{\Delta} = 0$, (5) recovers the standard LQG formulation (2). Additionally, note that the cost of (5) is finite if and only if $\mathbf{K}$ stabilizes all plants $\mathbf{G}$ such that $\mathbf{G} = \hat{\mathbf{G}} + \mathbf{\Delta}$ and $\|\mathbf{\Delta}\|_\infty < \epsilon$.

2.3. Our contributions

Although classical approaches exist to compute controllers that stabilize all plants $\hat{\mathbf{G}}$ with $\|\mathbf{\Delta}\|_\infty \leq \epsilon$, (Zhou et al., 1996), these methods typically do not quantify the closed-loop LQG performance degradation in terms of the uncertainty size ϵ. In this paper, we exploit the recent IOP framework (Furieri et al., 2019) to develop a tractable inner approximation of (5) (see Theorem 3.2). In our main Theorem 3.3, if ϵ is small enough, we bound the suboptimality performance gap as

$$\frac{\hat{J} - J_\star}{J_\star} \leq M\epsilon\,,$$

where $J_\star$ is the globally optimal LQG cost to (2), and $\hat{J}$ is the LQG cost when applying the robust controller from our procedure to the true plant $\mathbf{G}_\star$, $\epsilon > \|\mathbf{G}_\star - \hat{\mathbf{G}}\|_\infty$ is an upper bound of estimation error, and M is a constant that depends explicitly on true dynamics $\mathbf{G}_\star$, its estimation $\hat{\mathbf{G}}$, and the true LQG controller $\mathbf{K}_\star$; see (17) for a precise expression. Adapting recent non-asymptotic estimation results from input-output trajectories (Oymak and Ozay, 2019; Sarkar et al., 2019; Tu et al., 2017; Zheng and Li, 2020), we derive an end-to-end sample complexity of learning LQG controllers as

$$\frac{\hat{J} - J_\star}{J_\star} \leq \mathcal{O}\left(\frac{T}{\sqrt{N}} + \rho(A_\star)^T\right),$$

with high probability provided N is sufficiently large, where T is the length of finite impulse response (FIR) model estimation, and N is the number of samples in input-output trajectories (see Corollary 4.3). When the true plant $\mathbf{G}_\star$ is an FIR, the sample complexity scales as $\mathcal{O}\left(N^{-1/2}\right)$.

If $\mathbf{G}_\star$ is unstable, the residual $\mathbf{\Delta} = \mathbf{G}_\star - \hat{\mathbf{G}}$ might be unstable as well, and thus $\|\mathbf{\Delta}\|_\infty = \infty$. Instead, $\mathbf{\Delta}$ is always a stable residual when $\mathbf{G}_\star$ is stable, and thus $\|\mathbf{\Delta}\|_\infty$ is finite. Also, it is hard to utilize a single trajectory for identifying unstable systems. Finally, unstable residues in the equality constraints of IOP pose a challenge for controller implementation; we refer to Section 6 of Zheng et al. (2019) for details. We therefore leave the case of unstable systems for future work.

3. Robust controller synthesis

We first derive a tractable convex approximation for (5) using the recent input-output parameterization (IOP) framework (Furieri et al., 2019). This allows us to compute a robust controller using convex optimization. We then provide sub-optimality guarantees in terms of the uncertainty size ϵ. The overall principle is parallel to that of Dean et al. (2019) for the LQR case.

3.1. An equivalent IOP reformulation of (5)

Similar to the Youla parameterization (Youla et al., 1976) and the SLP (Wang et al., 2019), the IOP framework (Furieri et al., 2019) focuses on the *system responses* of a closed-loop system. In particular, given an arbitrary control policy $\mathbf{u} = \mathbf{K}\mathbf{y}$, straightforward calculations show that the closed-loop responses from the noises $\mathbf{v}, \mathbf{w}$ to the output $\mathbf{y}$ and control action $\mathbf{u}$ are

$$\begin{bmatrix}\mathbf{y}\\ \mathbf{u}\end{bmatrix} = \begin{bmatrix}(I-\mathbf{G}_\star\mathbf{K})^{-1} & (I-\mathbf{G}_\star\mathbf{K})^{-1}\mathbf{G}_\star\\ \mathbf{K}(I-\mathbf{G}_\star\mathbf{K})^{-1} & (I-\mathbf{K}\mathbf{G}_\star)^{-1}\end{bmatrix}\begin{bmatrix}\mathbf{v}\\ \mathbf{w}\end{bmatrix}. \tag{6}$$

To ease the notation, we can define the closed-loop responses as

$$\begin{bmatrix}\mathbf{Y} & \mathbf{W}\\ \mathbf{U} & \mathbf{Z}\end{bmatrix} := \begin{bmatrix}(I-\mathbf{G}_\star\mathbf{K})^{-1} & (I-\mathbf{G}_\star\mathbf{K})^{-1}\mathbf{G}_\star\\ \mathbf{K}(I-\mathbf{G}_\star\mathbf{K})^{-1} & (I-\mathbf{K}\mathbf{G}_\star)^{-1}\end{bmatrix}. \tag{7}$$

Given any stabilizing $\mathbf{K}$, we can write $\mathbf{K} = \mathbf{U}\mathbf{Y}^{-1}$ and the square root of the LQG cost in (2) as

$$J(\mathbf{G}_\star, \mathbf{K}) = \left\| \begin{bmatrix}\mathbf{Y} & \mathbf{W}\\ \mathbf{U} & \mathbf{Z}\end{bmatrix} \right\|_{\mathcal{H}_2}, \tag{8}$$

with the closed-loop responses $(\mathbf{Y}, \mathbf{U}, \mathbf{W}, \mathbf{Z})$ defined in (7); see Furieri et al. (2019); Zheng et al. (2020) for more details on IOP. We present a full equivalence derivation of (8) in (Zheng* and Furieri* et al., 2020, Appendix G).

Our first result is a reformulation of the robust LQG problem (5) in the IOP framework.

Theorem 3.1 *The robust LQG problem (5) is equivalent to*

$$\min_{\hat{\mathbf{Y}},\hat{\mathbf{W}},\hat{\mathbf{U}},\hat{\mathbf{Z}}} \max_{\|\mathbf{\Delta}\|_\infty<\epsilon} \quad J(\mathbf{G}_\star, \mathbf{K}) = \left\| \begin{bmatrix}\hat{\mathbf{Y}}(I-\mathbf{\Delta}\hat{\mathbf{U}})^{-1} & \hat{\mathbf{Y}}(I-\mathbf{\Delta}\hat{\mathbf{U}})^{-1}(\hat{\mathbf{G}}+\mathbf{\Delta})\\ \hat{\mathbf{U}}(I-\mathbf{\Delta}\hat{\mathbf{U}})^{-1} & (I-\hat{\mathbf{U}}\mathbf{\Delta})^{-1}\hat{\mathbf{Z}}\end{bmatrix} \right\|_{\mathcal{H}_2} \tag{9}$$

$$\text{subject to} \quad \begin{bmatrix}I & -\hat{\mathbf{G}}\end{bmatrix}\begin{bmatrix}\hat{\mathbf{Y}} & \hat{\mathbf{W}}\\ \hat{\mathbf{U}} & \hat{\mathbf{Z}}\end{bmatrix} = \begin{bmatrix}I & 0\end{bmatrix}, \tag{10}$$

$$\begin{bmatrix}\hat{\mathbf{Y}} & \hat{\mathbf{W}}\\ \hat{\mathbf{U}} & \hat{\mathbf{Z}}\end{bmatrix}\begin{bmatrix}-\hat{\mathbf{G}}\\ I\end{bmatrix} = \begin{bmatrix}0\\ I\end{bmatrix}, \tag{11}$$

$$\hat{\mathbf{Y}}, \hat{\mathbf{W}}, \hat{\mathbf{U}}, \hat{\mathbf{Z}} \in \mathcal{RH}_\infty, \ \|\hat{\mathbf{U}}\|_\infty \leq \frac{1}{\epsilon},$$

where the optimal robust controller is recovered from the optimal $\hat{\mathbf{U}}$ *and* $\hat{\mathbf{Y}}$ *as* $\mathbf{K} = \hat{\mathbf{U}}\hat{\mathbf{Y}}^{-1}$.

The proof relies on a novel robust variant of IOP for parameterizing robustly stabilizing controller in a convex way. We provide the detailed proof and a review of the IOP framework in (Zheng* and Furieri* et al., 2020, Appendix C). Here, it is worth noting that the feasible set in (9) is convex in the decision variables $(\hat{\mathbf{Y}}, \hat{\mathbf{W}}, \hat{\mathbf{U}}, \hat{\mathbf{Z}})$, which represent four closed-loop maps on the estimated plant $\hat{\mathbf{G}}$. Using the small-gain theorem (Zhou et al., 1996), the convex requirement $\|\hat{\mathbf{U}}\|_\infty \leq \frac{1}{\epsilon}$ ensures that any controller $\mathbf{K} = \hat{\mathbf{U}}\hat{\mathbf{Y}}^{-1}$, with $\hat{\mathbf{Y}}, \hat{\mathbf{U}}$ feasible for (9), stabilizes the real plant $\mathbf{G}_\star$ for all $\mathbf{\Delta}$ such that $\|\mathbf{\Delta}\|_\infty < \epsilon$.

Due to the uncertainty $\mathbf{\Delta}$, the cost in (9) is nonconvex in the decision variables. We therefore proceed with deriving an upper-bound on the functional $J(\mathbf{G}_\star, \mathbf{K})$, which will be exploited to derive a quasi-convex approximation of the robust LQG problem (5).

3.2. Upper bound on the non-convex cost in (9)

It is easy to derive (see Appendix B of Zheng* and Furieri* et al. (2020)) that

$$\begin{aligned} J(\mathbf{G}_\star, \mathbf{K})^2 = &\|\hat{\mathbf{Y}}(I - \mathbf{\Delta}\hat{\mathbf{U}})^{-1}\|_{\mathcal{H}_2}^2 + \|\hat{\mathbf{U}}(I - \mathbf{\Delta}\hat{\mathbf{U}})^{-1}\|_{\mathcal{H}_2}^2 \\ &+ \|(I - \hat{\mathbf{U}}\mathbf{\Delta})^{-1}\hat{\mathbf{Z}}\|_{\mathcal{H}_2}^2 + \|\hat{\mathbf{Y}}(I - \mathbf{\Delta}\hat{\mathbf{U}})^{-1}(\hat{\mathbf{G}} + \mathbf{\Delta})\|_{\mathcal{H}_2}^2. \end{aligned} \tag{12}$$

Similarly to Dean et al. (2019) for the LQR case, it is relatively easy to bound the first three terms on the right hand side of (12) using small-gain arguments. However, dealing with outputs makes it challenging to bound the last term. The corresponding result is summarized in the following proposition (see Appendix D.1 of Zheng* and Furieri* et al. (2020) for proof).

Proposition 3.1 *If $\|\hat{\mathbf{U}}\|_\infty < \frac{1}{\epsilon}$, $\|\mathbf{\Delta}\|_\infty < \epsilon$ and $\hat{\mathbf{G}} \in \mathcal{RH}_\infty$, then, we have*

$$\|\hat{\mathbf{Y}}(I - \mathbf{\Delta}\hat{\mathbf{U}})^{-1}(\hat{\mathbf{G}} + \mathbf{\Delta})\|_{\mathcal{H}_2} \leq \frac{\|\hat{\mathbf{W}}\|_{\mathcal{H}_2} + \epsilon\|\hat{\mathbf{Y}}\|_{\mathcal{H}_2}(2 + \|\hat{\mathbf{U}}\|_\infty\|\hat{\mathbf{G}}\|_\infty)}{1 - \epsilon\|\hat{\mathbf{U}}\|_\infty}, \tag{13}$$

where $\hat{\mathbf{W}} = \hat{\mathbf{Y}}\hat{\mathbf{G}}$.

We are now ready to present an upper bound on the LQG cost. The proof is based on Proposition 3.1 and basic inequalities; see (Zheng* and Furieri* et al., 2020, Appendix D.2).

Proposition 3.2 *If $\|\hat{\mathbf{U}}\|_\infty \leq \frac{1}{\epsilon}$, $\|\mathbf{\Delta}\|_\infty < \epsilon$ and $\hat{\mathbf{G}} \in \mathcal{RH}_\infty$, the robust LQG cost in (9) is upper bounded by*

$$J(\mathbf{G}_\star, \mathbf{K}) \leq \frac{1}{1 - \epsilon\|\hat{\mathbf{U}}\|_\infty} \left\| \begin{bmatrix} \sqrt{1 + h(\epsilon, \|\hat{\mathbf{U}}\|_\infty)}\hat{\mathbf{Y}} & \hat{\mathbf{W}} \\ \hat{\mathbf{U}} & \hat{\mathbf{Z}} \end{bmatrix} \right\|_{\mathcal{H}_2}, \tag{14}$$

where $\hat{\mathbf{Y}}, \hat{\mathbf{W}}, \hat{\mathbf{U}}, \hat{\mathbf{Z}}$ satisfy the constraints in (9), and the factor $h(\epsilon, \|\hat{\mathbf{U}}\|_\infty)$ is defined as

$$h(\epsilon, \|\hat{\mathbf{U}}\|_\infty) := \epsilon\|\hat{\mathbf{G}}\|_\infty(2 + \|\hat{\mathbf{U}}\|_\infty\|\hat{\mathbf{G}}\|_\infty) + \epsilon^2(2 + \|\hat{\mathbf{U}}\|_\infty\|\hat{\mathbf{G}}\|_\infty)^2. \tag{15}$$

3.3. Quasi-convex formulation

Building on the LQG cost upper bound (14), we derive our first main result on a tractable approximation of (9). The proof is reported in (Zheng* and Furieri* et al., 2020, Appendix D.3).

Theorem 3.2 *Given $\hat{\mathbf{G}} \in \mathcal{RH}_\infty$, a model estimation error ϵ, and any constant $\alpha > 0$, the robust LQG problem (9) is upper bounded by the following problem*

$$\begin{aligned} \min_{\gamma \in [0, 1/\epsilon)} \frac{1}{1 - \epsilon\gamma} \quad \min_{\hat{\mathbf{Y}}, \hat{\mathbf{W}}, \hat{\mathbf{U}}, \hat{\mathbf{Z}}} & \left\| \begin{bmatrix} \sqrt{1 + h(\epsilon, \alpha)}\hat{\mathbf{Y}} & \hat{\mathbf{W}} \\ \hat{\mathbf{U}} & \hat{\mathbf{Z}} \end{bmatrix} \right\|_{\mathcal{H}_2} \\ \text{subject to} \quad & (10) - (11),\ \hat{\mathbf{Y}}, \hat{\mathbf{W}}, \hat{\mathbf{Z}} \in \mathcal{RH}_\infty,\ \|\hat{\mathbf{U}}\|_\infty \leq \min(\gamma, \alpha), \end{aligned} \tag{16}$$

where $h(\epsilon, \alpha) = \epsilon\|\hat{\mathbf{G}}\|_\infty(2 + \alpha\|\hat{\mathbf{G}}\|_\infty) + \epsilon^2(2 + \alpha\|\hat{\mathbf{G}}\|_\infty)^2$.

The hyper-parameter α in (16) plays two important roles: 1) *robust stability*: the resulting controller has a guarantee of robust stability against model estimation error up to $\frac{1}{\alpha}$, thus suggesting $\alpha < \frac{1}{\epsilon}$, as we will clarify it later; 2) *quasi-convexity*: the inner optimization problem (16) is convex when fixing γ, and the outer optimization is quasi-convex with respect to γ, which can effectively be solved using the golden section search.

Remark 3.1 (Numerical computation)

1. *The inner optimization in (16) is convex but infinite dimensional. A practical numerical approach is to apply a finite impulse response (FIR) truncation on the decision variables $\hat{\mathbf{Y}}, \hat{\mathbf{U}}, \hat{\mathbf{W}}, \hat{\mathbf{Z}}$, which leads to a finite dimensional convex semidefinite program (SDP) for each fixed value of γ; see Appendix B of Zheng* and Furieri* et al. (2020). The degradation in performance decays exponentially with the FIR horizon (Dean et al., 2019).*

2. *Since $\hat{\mathbf{G}} \in \mathcal{RH}_\infty$ is stable, the IOP framework is numerically robust (Zheng et al., 2019),* i.e.*, the resulting controller $\mathbf{K} = \hat{\mathbf{U}}\hat{\mathbf{Y}}^{-1}$ is stabilizing even when numerical solvers induce small computational residues in (16).*

3.4. Sub-optimality guarantee

Our second main result offers a sub-optimality guarantee on the performance of the robust controller synthesized using the robust IOP framework (16) in terms of the estimation error ϵ. The proof is reported in (Zheng* and Furieri* et al., 2020, Appendix E).

Theorem 3.3 *Let $\mathbf{K}_\star$ be the optimal LQG controller in (2), and the corresponding closed-loop responses be $\mathbf{Y}_\star, \mathbf{U}_\star, \mathbf{W}_\star, \mathbf{Z}_\star$. Let $\hat{\mathbf{G}}$ be the plant estimation with error $\|\mathbf{\Delta}\|_\infty < \epsilon$, where $\mathbf{\Delta} = \mathbf{G}_\star - \hat{\mathbf{G}}$. Suppose that $\epsilon < \frac{1}{5\|\mathbf{U}_\star\|_\infty}$, and choose the constant hyper-parameter $\alpha \in \left[\frac{\sqrt{2}\|\mathbf{U}_\star\|_\infty}{1-\epsilon\|\mathbf{U}_\star\|_\infty}, \frac{1}{\epsilon}\right)$. We denote the optimal solution to (16) as $\gamma_\star, \hat{\mathbf{Y}}_\star, \hat{\mathbf{U}}_\star, \hat{\mathbf{W}}_\star, \hat{\mathbf{Z}}_\star$. Then, the controller $\mathbf{K} = \hat{\mathbf{U}}_\star\hat{\mathbf{Y}}_\star^{-1}$ stabilizes the true plant $\mathbf{G}_\star$ and the relative LQG error is upper bounded by*

$$\frac{J(\mathbf{G}_\star, \mathbf{K})^2 - J(\mathbf{G}_\star, \mathbf{K}_\star)^2}{J(\mathbf{G}_\star, \mathbf{K}_\star)^2} \leq 20\epsilon\|\mathbf{U}_\star\|_\infty + h(\epsilon, \alpha) + g(\epsilon, \|\mathbf{U}_\star\|_\infty), \tag{17}$$

where $h(\cdot,\cdot)$ is defined in (15) and

$$g(\epsilon, \|\mathbf{U}_\star\|_\infty) = \epsilon\|\mathbf{G}_\star\|_\infty(2 + \|\mathbf{U}_\star\|_\infty\|\mathbf{G}_\star\|_\infty) + \epsilon^2(2 + \|\mathbf{U}_\star\|_\infty\|\mathbf{G}_\star\|_\infty)^2. \tag{18}$$

Theorem 3.3 shows that the relative error in the LQG cost grows as $\mathcal{O}(\epsilon)$ as long as ϵ is sufficiently small, and in particular $\epsilon < \frac{1}{5\|\mathbf{U}_\star\|_\infty}$. Previous results in Dean et al. (2019) proved a similar convergence rate of $\mathcal{O}(\epsilon)$ for LQR using a robust synthesis procedure based on the SLP (Wang et al., 2019). Our robust synthesis procedure using the IOP framework extends Dean et al. (2019) to a class of LQG problems. Note that our bound is valid for open-loop stable plants, while the method Dean et al. (2019) works for all systems at the cost of requiring direct state observations. Similar to Dean et al. (2019) and related work, the hyper-parameters ϵ and α have to be tuned in practice without knowing $\mathbf{U}_\star$.

Remark 3.2 (Optimality versus Robustness) *Note that recent results in Mania et al. (2019) show that the certainty equivalent controller achieves a better sub-optimality scaling of $\mathcal{O}(\epsilon^2)$ for both fully observed LQR and partially observed LQG settings, at the cost of a much stricter requirement on admissible uncertainty ϵ. Quoting from Mania et al. (2019),* "the price of obtaining a faster rate for LQR is that the certainty equivalent controller becomes less robust to model uncertainty." *Our result in Theorem 3.3 shows that this trade-off may hold true for the LQG problem as well.*

4. Sample complexity

Based on our main results in the previous section, here we discuss how to estimate the plant $\mathbf{G}_\star$, provide a non-asymptotic $\mathcal{H}_\infty$ bound on the estimation error, and finally establish an end-to-end sample complexity of learning LQG controllers. By Assumption 2, *i.e.*, $\mathbf{G}_\star \in \mathcal{RH}_\infty$, we can write

$$\mathbf{G}_\star(z) = \sum_{i=0}^{\infty} \frac{1}{z^i} G_{\star,i} = \sum_{i=0}^{T-1} \frac{1}{z^i} G_{\star,i} + \sum_{i=T}^{\infty} \frac{1}{z^i} G_{\star,i}, \tag{19}$$

where $G_{\star,i} \in \mathbb{R}^{p\times m}$ denotes the i-th spectral component. Given the state-space representation (1), we have $G_{\star,0} = 0\,, G_{\star,i} = C_\star A_\star^{i-1} B_\star,\ \forall i \geq 1\,.$ As $\rho(A_\star) < 1$, the spectral component $G_{\star,i}$ decays exponentially. Thus, we can use a finite impulse response (FIR) truncation of order T for $\mathbf{G}_\star$:

$$G_\star = \begin{bmatrix} 0 & C_\star B_\star & \cdots & C_\star A_\star^{T-2} B_\star \end{bmatrix} \in \mathbb{R}^{p\times Tm}.$$

Many recent algorithms have been proposed to estimate $G_\star$, e.g., Oymak and Ozay (2019); Tu et al. (2017); Sarkar et al. (2019); Zheng and Li (2020), and these algorithms differ from the estimation setup; see Zheng and Li (2020) for a comparison. All of them can be used to establish an end-to-end sample complexity. Here, we discuss a recent ordinary least-squares (OLS) algorithm (Oymak and Ozay, 2019). This OLS estimator is based on collecting a trajectory $\{y_t, u_t\}_{t=0}^{T+N-1}$, where u_t is Gaussian with variance $\sigma_u^2 I$ for every t. The OLS algorithm details are provided in Zheng* and Furieri* et al. (2020, Appendix F.1). From the OLS solution $\hat{G}$, we form the estimated plant

$$\hat{\mathbf{G}} := \sum_{k=0}^{T-1} \frac{1}{z^k} \hat{G}_k. \tag{20}$$

To bound the $\mathcal{H}_\infty$ norm of the estimation error $\mathbf{\Delta} := \mathbf{G}_\star - \hat{\mathbf{G}}$, we define $\Phi(A_\star) = \sup_{\tau\geq 0} \frac{\|A_\star^\tau\|}{\rho(A_\star)^\tau}$. We assume $\Phi(A_\star)$ is finite (Oymak and Ozay, 2019). We start from (Oymak and Ozay, 2019, Theorem 3.2) to derive two new corollaries. The proofs rely on connecting the spectral radius of $\hat{G} - G_\star$ with the $\mathcal{H}_\infty$ norm of $\hat{\mathbf{G}} - \mathbf{G}_\star$; see Zheng* and Furieri* et al. (2020, Appendix F.2) for details.

Corollary 4.1 *Under the OLS estimation setup in Theorem 3.2 of Oymak and Ozay (2019), with high probability, the FIR estimation $\hat{\mathbf{G}}$ in (20) satisfies*

$$\|\mathbf{G}_\star - \hat{\mathbf{G}}\|_\infty \leq \frac{R_w + R_v + R_e}{\sigma_u}\sqrt{\frac{T}{N}} + \Phi(A_\star)\,\|C_\star\|\,\|B_\star\|\,\frac{\rho(A_\star)^T}{1-\rho(A_\star)}. \tag{21}$$

where N is the length of one input-output trajectory, and $R_w, R_v, R_e \in \mathbb{R}$ are problem-dependent[3].

Corollary 4.2 *Fix an $\epsilon > 0$. Let the length of FIR truncation satisfy*

$$T > \frac{1}{\log\left(\rho(A_\star)\right)} \log \frac{\epsilon(1-\rho(A_\star))}{2\Phi(A_\star)\|C_\star\|\|B_\star\|}. \tag{22}$$

3. See Oymak and Ozay (2019) for precise formula and probabilistic guarantees. Note that the dynamics in (1) are slightly different from Oymak and Ozay (2019), with an extra matrix $B_\star$ in front of w_t. By replacing the matrix F in Oymak and Ozay (2019) with $G_\star$, all the analysis and bounds stay the same.

Under the OLS estimation setup in Theorem 3.2 of Oymak and Ozay (2019), and further letting

$$N > \max\left\{\frac{4T}{\sigma_u^2\epsilon^2}(R_w + R_v + R_e)^2, cTm\log^2(2Tm)\log^2(2Nm)\right\}, \tag{23}$$

with high probability, the FIR estimation $\hat{\mathbf{G}}$ *in* (20) *satisfies* $\|\mathbf{G}_\star - \hat{\mathbf{G}}\|_\infty < \epsilon$.

The lower bound on the FIR length T in (22) guarantees that the FIR truncation error is less than $\epsilon/2$, while the lower bound on N in (23) makes sure that the FIR approximation part is less than $\epsilon/2$, thus leading to the desired bound with high probability. We note that the terms R_w, R_v, R_e depend on the system dimensions and on the FIR length as $\mathcal{O}(\sqrt{T}(p+m+n))$. Corollary 4.2 states that the number of time steps to achieve identification error ϵ in $\mathcal{H}_\infty$ norm scales as $\mathcal{O}(T^2/\epsilon^2)$.

We finally give an end-to-end guarantee by combining Corollary 4.2 with Theorem 3.3:

Corollary 4.3 *Let* $\mathbf{K}_\star$ *be the optimal LQG controller to* (2)*, and the corresponding closed-loop responses be* $\mathbf{Y}_\star, \mathbf{U}_\star, \mathbf{W}_\star, \mathbf{Z}_\star$*. Choose an estimation error* $0 < \epsilon < \frac{1}{5\|\mathbf{U}_\star\|_\infty}$*, and the hyper-parameter* $\alpha \in \left[\frac{\sqrt{2}\|\mathbf{U}_\star\|_\infty}{1-\epsilon\|\mathbf{U}_\star\|_\infty}, \frac{1}{\epsilon}\right)$*. Estimate* $\hat{\mathbf{G}}$ (20) *with a trajectory of length* N *in* (23) *and an FIR truncation length* T *in* (22)*. Then, with high probability, the robust controller* $\mathbf{K}$ *from* (16) *yields a relative error in the LQG cost satisfying* (17)*.*

Since the bound on the trajectory length N in (23) scales as $\tilde{\mathcal{O}}(\epsilon^{-2})$ (where the logarithmic factor comes from the FIR length T), the suboptimality gap on the LQG cost roughly scales as $\tilde{\mathcal{O}}\left(\frac{1}{\sqrt{N}}\right)$ when the FIR length T is chosen large-enough accordingly. In particular, when the true plant is FIR of order $\overline{T}$ and $T \geq \overline{T}$, via combining Corollary 4.1 with Theorem 3.3, we see that with high probability, the suboptimality gap behaves as

$$\frac{J(\mathbf{G}_\star, \mathbf{K})^2 - J(\mathbf{G}_\star, \mathbf{K}_\star)^2}{J(\mathbf{G}_\star, \mathbf{K}_\star)^2} \sim \mathcal{O}\left(\frac{1}{\sqrt{N}}\right).$$

Despite the additional difficulty of hidden states, our sample complexity result is on the same level as that obtained in Dean et al. (2019) where a robust SLP procedure is used to design a robust LQR controller with full observations.

5. Conclusion and future work

We have developed a robust controller synthesis procedure for partially observed LQG problems, by combining non-asymptotic identification methods with IOP for robust control. Our procedure is consistent with the idea of Coarse-ID control in Dean et al. (2019), and extends the results in Dean et al. (2019) from fully observed state feedback systems to partially observed output-feedback systems that are open-loop stable.

One interesting future direction is to extend these results to open-loop unstable systems. We note that non-asymptotic identification for partially observed open-loop unstable system is challenging in itself; see Zheng and Li (2020) for a recent discussion. It is also non-trivial to derive a robust synthesis procedure with guaranteed performance, and using a pre-stabilizing controller might be useful (Simchowitz et al., 2020; Zheng et al., 2019). Finally, extending our results to an online adaptive setting and performing regret analysis are exciting future directions as well.

References

Karl J Åström and Björn Wittenmark. *Adaptive control*. Courier Corporation, 2013.

Julian Berberich, Anne Romer, Carsten W Scherer, and Frank Allgöwer. Robust data-driven state-feedback design. *arXiv preprint arXiv:1909.04314*, 2019.

Dimitri P Bertsekas. Dynamic programming and optimal control 3rd edition, volume II. *Belmont, MA: Athena Scientific*, 2011.

Ross Boczar, Nikolai Matni, and Benjamin Recht. Finite-data performance guarantees for the output-feedback control of an unknown system. In *2018 IEEE Conference on Decision and Control (CDC)*, pages 2994–2999. IEEE, 2018.

Jeremy Coulson, John Lygeros, and Florian Dörfler. Data-enabled predictive control: In the shallows of the deepc. In *2019 18th European Control Conference (ECC)*, pages 307–312. IEEE, 2019.

Claudio De Persis and Pietro Tesi. Formulas for data-driven control: Stabilization, optimality and robustness. *IEEE Transactions on Automatic Control*, 2019.

Sarah Dean, Horia Mania, Nikolai Matni, Benjamin Recht, and Stephen Tu. Regret bounds for robust adaptive control of the Linear Quadratic Regulator. In *Advances in Neural Information Processing Systems*, pages 4188–4197, 2018.

Sarah Dean, Horia Mania, Nikolai Matni, Benjamin Recht, and Stephen Tu. On the sample complexity of the Linear Quadratic Regulator. *Foundations of Computational Mathematics*, pages 1–47, 2019.

Yan Duan, Xi Chen, Rein Houthooft, John Schulman, and Pieter Abbeel. Benchmarking deep reinforcement learning for continuous control. In *International Conference on Machine Learning*, pages 1329–1338, 2016.

Maryam Fazel, Rong Ge, Sham M Kakade, and Mehran Mesbahi. Global convergence of policy gradient methods for the Linear Quadratic Regulator. *arXiv preprint arXiv:1801.05039*, 2018.

Luca Furieri, Yang Zheng, Antonis Papachristodoulou, and Maryam Kamgarpour. An input–output parametrization of stabilizing controllers: Amidst youla and system level synthesis. *IEEE Control Systems Letters*, 3(4):1014–1019, 2019.

Luca Furieri, Yang Zheng, and Maryam Kamgarpour. Learning the globally optimal distributed LQ regulator. In *Learning for Dynamics and Control*, pages 287–297, 2020.

Lennart Ljung. Perspectives on system identification. *Annual Reviews in Control*, 34(1):1–12, 2010.

Dhruv Malik, Ashwin Pananjady, Kush Bhatia, Koulik Khamaru, Peter L Bartlett, and Martin J Wainwright. Derivative-free methods for policy optimization: Guarantees for linear quadratic systems. *arXiv preprint arXiv:1812.08305*, 2018.

Horia Mania, Stephen Tu, and Benjamin Recht. Certainty equivalent control of LQR is efficient. *arXiv preprint arXiv:1902.07826*, 2019.

Samet Oymak and Necmiye Ozay. Non-asymptotic identification of LTI systems from a single trajectory. In *2019 American Control Conference (ACC)*, pages 5655–5661. IEEE, 2019.

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

Tuhin Sarkar, Alexander Rakhlin, and Munther A Dahleh. Finite-time system identification for partially observed LTI systems of unknown order. *arXiv preprint arXiv:1902.01848*, 2019.

Max Simchowitz, Karan Singh, and Elad Hazan. Improper learning for non-stochastic control. *arXiv preprint arXiv:2001.09254*, 2020.

Anastasios Tsiamis, Nikolai Matni, and George Pappas. Sample complexity of kalman filtering for unknown systems. In *Learning for Dynamics and Control*, pages 435–444. PMLR, 2020.

Stephen Tu and Benjamin Recht. The gap between model-based and model-free methods on the Linear Quadratic Regulator: An asymptotic viewpoint. In *Conference on Learning Theory*, pages 3036–3083, 2019.

Stephen Tu, Ross Boczar, Andrew Packard, and Benjamin Recht. Non-asymptotic analysis of robust control from coarse-grained identification. *arXiv preprint arXiv:1707.04791*, 2017.

Yuh-Shyang Wang, Nikolai Matni, and John C Doyle. A system level approach to controller synthesis. *IEEE Trans. on Aut. Contr.*, 2019.

Dante Youla, Hamid Jabr, and Jr Bongiorno. Modern Wiener-Hopf design of optimal controllers–Part II: The multivariable case. *IEEE Trans. on Aut. Contr.*, 21(3):319–338, 1976.

Yang Zheng and Na Li. Non-asymptotic identification of linear dynamical systems using multiple trajectories. *arXiv preprint arXiv:2009.00739*, 2020.

Yang Zheng, Luca Furieri, Maryam Kamgarpour, and Na Li. System-level, input-output and new parameterizations of stabilizing controllers, and their numerical computation. *arXiv preprint arXiv:1909.12346*, 2019.

Yang Zheng, Luca Furieri, Antonis Papachristodoulou, Na Li, and Maryam Kamgarpour. On the equivalence of Youla, System-level and Input-output parameterizations. *IEEE Transactions on Automatic Control*, pages 1–8, 2020.

Zheng* and Furieri*, Maryam Kamgarpour, and Na Li. Sample complexity of linear quadratic gaussian (LQG) control for output feedback systems. *arXiv preprint arXiv:2011.09929, [PDF]*, 2020.

Kemin Zhou, John Comstock Doyle, and Keith Glover. *Robust and optimal control*, volume 40. Prentice hall New Jersey, 1996.