



On the Sample Complexity of the Linear Quadratic Regulator

Sarah Dean¹ · Horia Mania¹ · Nikolai Matni² · Benjamin Recht¹ · Stephen Tu¹

Received: 8 February 2018 / Revised: 3 December 2018 / Accepted: 22 February 2019 /
Published online: 5 August 2019

© SFOCM 2019

Abstract

This paper addresses the optimal control problem known as the linear quadratic regulator in the case when the dynamics are unknown. We propose a multistage procedure, called *Coarse-ID control*, that estimates a model from a few experimental trials, estimates the error in that model with respect to the truth, and then designs a controller using both the model and uncertainty estimate. Our technique uses contemporary tools from random matrix theory to bound the error in the estimation procedure. We also employ a recently developed approach to control synthesis called *System Level Synthesis* that enables robust control design by solving a quasi-convex optimization problem. We provide end-to-end bounds on the relative error in control cost that are optimal in the number of parameters and that highlight salient properties of the system to be controlled such as closed-loop sensitivity and optimal control magnitude. We show experimentally that the Coarse-ID approach enables efficient computation of a stabilizing controller in regimes where simple control schemes that do not take the model uncertainty into account fail to stabilize the true system.

Keywords Optimal control · Robust control · System identification · Statistical learning theory · Reinforcement learning · System level synthesis

Mathematics Subject Classification 49N05 · 93D09 · 93D21 · 93E12 · 93E35

Communicated by Michael Overton.

This work was generously supported in part by NSF award CCF-1359814, ONR awards N00014-14-1-0024 and N00014-17-1-2191, the DARPA Fundamental Limits of Learning Program, a Sloan Research Fellowship, and a Google Faculty Award. SD is additionally supported by an NSF Graduate Research Fellowship under Grant No. DGE 1752814. NM is additionally supported by grants from the AFOSR and NSF, and by gifts from Huawei and Google.

Extended author information available on the last page of the article

1 Introduction

Having surpassed human performance in video games [41] and Go [51], there has been a renewed interest in applying machine learning techniques to planning and control. In particular, there has been a considerable amount of effort in developing new techniques for *continuous control* where an autonomous system interacts with a physical environment [16,34]. A tremendous opportunity lies in deploying these data-driven systems in more demanding interactive tasks including self-driving vehicles, distributed sensor networks, and agile robotics. As the role of machine learning expands to more ambitious tasks, however, it is critical these new technologies be safe and reliable. Failure of such systems could have severe social and economic consequences including the potential loss of human life. How can we guarantee that our new data-driven automated systems are robust?

Unfortunately, there are no clean baselines delineating the possible control performance achievable given a fixed amount of data collected from a system. Such baselines would enable comparisons of different techniques and would allow engineers to trade off between data collection and action in scenarios with high uncertainty. Typically, a key difficulty in establishing baselines is in proving *lower bounds* that state the minimum amount of knowledge needed to achieve a particular performance, regardless of method. However, in the context of controls, even *upper bounds* describing the worst-case performance of competing methods are exceptionally rare. Without such estimates, we are left to compare algorithms on a case-by-case basis, and we may have trouble diagnosing whether poor performance is due to algorithm choice or some other errors such as a software bug or a mechanical flaw.

In this paper, we attempt to build a foundation for a theoretical understanding of how machine learning interfaces with control by analyzing one of the most well-studied problems in classical optimal control, the *Linear Quadratic Regulator* (LQR). Here we assume that the system to be controlled obeys *linear* dynamics, and we wish to minimize some *quadratic* function of the system state and control action. This problem has been studied for decades in control: it has a simple, closed-form solution on the infinite time horizon and an efficient, dynamic programming solution on finite-time horizons. When the dynamics are unknown, however, there are far fewer results about achievable performance.

Our contribution is to analyze the LQR problem when the dynamics of the system are unknown, and we can measure the system's response to varied inputs. A naïve solution to this problem would be to collect some data of how the system behaves over time, fit a model to this data, and then solve the original LQR problem assuming this model is accurate. Unfortunately, while this procedure might perform well given sufficient data, it is difficult to determine how many experiments are necessary in practice. Furthermore, it is easy to construct examples where the procedure fails to find a stabilizing controller.

As an alternative, we propose a method that couples our uncertainty in estimation with the control design. Our main approach uses the following framework of *Coarse-ID control* to solve the problem of LQR with unknown dynamics:

1. Use supervised learning to learn a coarse model of the dynamical system to be controlled. We refer to the system estimate as the *nominal system*.

2. Using either prior knowledge or statistical tools like the bootstrap, build probabilistic guarantees about the distance between the nominal system and the true, unknown dynamics.
3. Solve a robust optimization problem over controllers that optimizes performance of the nominal system while penalizing signals with respect to the estimated uncertainty, ensuring stable and robust execution.

We will show that for a sufficient number of observations of the system, this approach is guaranteed to return a control policy with small relative cost. In particular, it guarantees asymptotic stability of the closed-loop system. In the case of LQR, step 1 of Coarse-ID control simply requires solving a linear least squares problem, step 2 uses a finite sample theoretical guarantee or a standard bootstrap technique, and step 3 requires solving a small semidefinite program. Analyzing this approach, on the other hand, requires contemporary techniques in nonasymptotic statistics and a novel parameterization of control problems that renders nonconvex problems convex [38,59].

We demonstrate the utility of our method on a simple simulation. In the presented example, we show that simply using the nominal system to design a control policy frequently results in unstable closed-loop behavior, even when there is an abundance of data from the true system. However, the Coarse-ID approach finds a stabilizing controller with few system observations.

1.1 Problem Statement and Our Contributions

The standard optimal control problem aims to find a control sequence that minimizes an expected cost. We assume a dynamical system with *state* $x_t \in \mathbb{R}^n$ can be acted on by a *control* $u_t \in \mathbb{R}^p$ and obeys the stochastic dynamics

$$x_{t+1} = f_t(x_t, u_t, w_t) \tag{1}$$

where w_t is a random process with w_t independent of $w_{t'}$ for all $t \neq t'$. Optimal control then seeks to minimize

$$\begin{aligned} &\text{minimize } \mathbb{E} \left[\frac{1}{T} \sum_{t=1}^T c_t(x_t, u_t) \right] \\ &\text{subject to } x_{t+1} = f_t(x_t, u_t, w_t) \end{aligned} \tag{2}$$

Here, c_t denotes the state-control cost at every time step, and the input u_t is allowed to depend on the current state x_t and all previous states and actions. In this generality, problem (2) encapsulates many of the problems considered in the reinforcement learning literature.

The simplest optimal control problem with continuous state is the linear quadratic regulator (LQR), in which costs are a fixed quadratic function of state and control and the dynamics are linear and time-invariant:

$$\begin{aligned} &\text{minimize } \mathbb{E} \left[\frac{1}{T} \sum_{t=1}^T x_t^* Q x_t + u_{t-1}^* R u_{t-1} \right] \\ &\text{subject to } x_{t+1} = A x_t + B u_t + w_t \end{aligned} \tag{3}$$

Here Q (resp. R) is a $n \times n$ (resp. $p \times p$) positive definite matrix, A and B are called the *state-transition matrices*, and $w_t \in \mathbb{R}^n$ is Gaussian noise with zero mean and covariance Σ_w . Throughout, M^* denotes the Hermitian transpose of the matrix M .

In what follows, we will be concerned with the *infinite time horizon* variant of the LQR problem where we let the time horizon T go to infinity and minimize the average cost. When the dynamics are known, this problem has a celebrated closed-form solution based on the solution of matrix Riccati equations [64]. Indeed, the optimal solution sets $u_t = Kx_t$ for a fixed $p \times n$ matrix K , and the corresponding optimal cost will serve as our gold-standard baseline to which we will compare the achieved cost of all algorithms.

In the case when the state-transition matrices are unknown, fewer results have been established about what cost is achievable. We will assume that we can conduct experiments of the following form: given some initial state x_0 , we can evolve the dynamics for T time steps using any control sequence $\{u_0, \dots, u_{T-1}\}$, measuring the resulting output $\{x_1, \dots, x_T\}$. If we run N such independent experiments, what infinite time horizon control cost is achievable using only the data collected? For simplicity of bookkeeping, in our analysis we further assume that we can prepare the system in initial state $x_0 = 0$.

In what follows, we will examine the performance of the Coarse-ID control framework in this scenario. We will estimate the errors accrued by least squares estimates (\hat{A}, \hat{B}) of the system dynamics. This estimation error is not easily handled by standard techniques because the design matrix is highly correlated with the model to be estimated. Regardless, for theoretical tractability, we can build a least squares estimate using only the final sample (x_T, x_{T-1}, u_{T-1}) of each of the N experiments. Indeed, in Sect. 2 we prove the following

Proposition 1 *Define the matrices*

$$G_T = [A^{T-2}B \ A^{T-3}B \ \dots \ B] \quad \text{and} \quad F_T = [A^{T-2} \ A^{T-3} \ \dots \ I_n]. \quad (4)$$

Assume we collect data from the linear, time-invariant system initialized at $x_0 = 0$, using inputs $u_t \stackrel{i.i.d.}{\sim} \mathcal{N}(0, \sigma_u^2 I_p)$ for $t = 0, \dots, T - 1$, with $T \geq 2$. Suppose that the process noise is $w_t \stackrel{i.i.d.}{\sim} \mathcal{N}(0, \sigma_w^2 I_n)$ and that

$$N \geq 8(n + p) + 16 \log(4/\delta).$$

Then, with probability at least $1 - \delta$, the least squares estimator using only the final sample of each trajectory satisfies both the inequality

$$\|\hat{A} - A\|_2 \leq \frac{16\sigma_w}{\sqrt{\lambda_{\min}(\sigma_u^2 G_T G_T^* + \sigma_w^2 F_T F_T^*)}} \sqrt{\frac{(2n + p) \log(36/\delta)}{N}}, \quad (5)$$

and the inequality

$$\|\hat{B} - B\|_2 \leq \frac{16\sigma_w}{\sigma_u} \sqrt{\frac{(2n + p) \log(36/\delta)}{N}}. \quad (6)$$

The details of the estimation procedure are described in Sect. 2. Note that this estimation result seems to yield an optimal dependence in terms of the number of parameters: (A, B) together have $n(n + p)$ parameters to learn and each measurement consists of n values. Moreover, this proposition further illustrates that not all linear systems are equally easy to estimate. The matrices $G_T G_T^*$ and $F_T F_T^*$ are finite-time *controllability Gramians* for the control and noise inputs, respectively. These are standard objects in control: each eigenvalue/vector pair of such a Gramian characterizes how much input energy is required to move the system in that particular direction of the state space. Therefore, $\lambda_{\min}(\sigma_u^2 G_T G_T^* + \sigma_w^2 F_T F_T^*)$ quantifies the least controllable, and hence most difficult to excite and estimate, mode of the system. This property is captured nicely in our bound, which indicates that for systems for which all modes are easily excitable (i.e., all modes of the system amplify the applied inputs and disturbances), the identification task becomes easier.

While we cannot compute the operator norm error bounds (5) and (6) without knowing the true system matrices (A, B) , we present a data-dependent bound in Proposition 3. Moreover, as we show in Sect. 2.3, a simple bootstrap procedure can efficiently upper bound the errors $\epsilon_A := \|A - \hat{A}\|_2$ and $\epsilon_B := \|B - \hat{B}\|_2$ from simulation.

With our estimates (\hat{A}, \hat{B}) and error bounds (ϵ_A, ϵ_B) in hand, we can turn to the problem of synthesizing a controller. We can assert with high probability that $A = \hat{A} + \Delta_A$ and $B = \hat{B} + \Delta_B$, for $\|\Delta_A\|_2 \leq \epsilon_A$ and $\|\Delta_B\|_2 \leq \epsilon_B$, where the size of the error terms is determined by the number of samples N collected. In light of this, it is natural to pose the following robust variant of the standard LQR optimal control problem (3), which computes a robustly stabilizing controller that seeks to minimize the worst-case performance of the system given the (high probability) norm bounds on the perturbations Δ_A and Δ_B :

$$\begin{aligned} & \text{minimize} && \sup_{\substack{\|\Delta_A\|_2 \leq \epsilon_A \\ \|\Delta_B\|_2 \leq \epsilon_B}} \lim_{T \rightarrow \infty} \frac{1}{T} \sum_{t=1}^T \mathbb{E} [x_t^* Q x_t + u_{t-1}^* R u_{t-1}] \\ & \text{subject to} && x_{t+1} = (\hat{A} + \Delta_A)x_t + (\hat{B} + \Delta_B)u_t + w_t \end{aligned} \tag{7}$$

Although classic methods exist for computing such controllers [21,45,53,60], they typically require solving nonconvex optimization problems, and it is not readily obvious how to extract interpretable measures of controller performance as a function of the perturbation sizes ϵ_A and ϵ_B . To that end, we leverage the recently developed System Level Synthesis (SLS) framework [59] to create an alternative robust synthesis procedure. Described in detail in Sect. 3, SLS lifts the system description into a higher-dimensional space that enables efficient search for controllers. At the cost of some conservatism, we are able to guarantee robust stability of the resulting closed-loop system for all admissible perturbations and bound the performance gap between the resulting controller and the optimal LQR controller. This is summarized in the following proposition.

Proposition 2 *Let (\hat{A}, \hat{B}) be estimated via the independent data collection scheme used in Proposition 1 and \hat{K} synthesized using robust SLS. Let \hat{J} denote the infinite time horizon LQR cost accrued by using the controller \hat{K} and J_* denote the optimal LQR cost achieved when (A, B) are known. Then the relative error in the LQR cost is bounded as*

$$\frac{\widehat{J} - J_\star}{J_\star} \leq \mathcal{O} \left(\mathcal{C}_{\text{LQR}} \sqrt{\frac{(n+p) \log(1/\delta)}{N}} \right) \quad (8)$$

with probability $1 - \delta$ provided N is sufficiently large.

The complexity term \mathcal{C}_{LQR} depends on the rollout length T , the true dynamics, the matrices (Q, R) which define the LQR cost, and the variances σ_u^2 and σ_w^2 of the control and noise inputs, respectively. The $1 - \delta$ probability comes from the probability of estimation error from Proposition 1. The particular form of \mathcal{C}_{LQR} and concrete requirements on N are both provided in Sect. 4.

Though the optimization problem formulated by SLS is infinite dimensional, in Sect. 5 we provide two finite-dimensional upper bounds on the optimization that inherit the stability guarantees of the SLS formulation. Moreover, we show via numerical experiments in Sect. 6 that the controllers synthesized by our optimization do indeed provide stabilizing controllers with small relative error. We further show that settings exist wherein a naïve synthesis procedure that ignores the uncertainty in the state-space parameter estimates produces a controller that performs poorly (or has unstable closed-loop behavior) relative to the controller synthesized using the SLS procedure.

1.2 Related Work

We first describe related work in the *estimation of unknown dynamical systems* and then turn to connections in the literature on *robust control with uncertain models*. We will end this review with a discussion of a few works from the reinforcement learning community that have attempted to address the LQR problem and related variants.

Estimation of Unknown Dynamical Systems

Estimation of unknown systems, especially linear dynamical systems, has a long history in the system identification subfield of control theory. While the text of Ljung [36] covers the classical asymptotic results, our interest is primarily in nonasymptotic results. Early results [10,57] on nonasymptotic rates for parameter identification featured conservative bounds which are exponential in the system degree and other relevant quantities. More recently, Bento et al. [46] show that when the A matrix is stable and induced by a sparse graph, then one can recover the support of A from a single trajectory using ℓ_1 -penalized least squares. Furthermore, Hardt et al. [26] provide the first polynomial-time guarantee for identifying stable linear systems with outputs. Their guarantees, however, are in terms of predictive output performance of the model and require an assumption on the true system that is more stringent than stability. It is not clear how their statistical risk guarantee can be used in a downstream robust synthesis procedure.

Next, we turn our attention to system identification of linear systems in the frequency domain. A comprehensive text on these methods (which differ from the aforementioned state-space methods) is the work by Chen and Gu [11]. For stable systems, Helmicki et al. [29] propose to identify a finite impulse response (FIR) approximation by directly estimating the first r impulse response coefficients. This method is analyzed in a nonadversarial probabilistic setting by [23,54], who prove that a poly-

nomial number of samples are sufficient to recover a FIR filter which approximates the true system in both ℓ_p -norm and \mathcal{H}_∞ -norm. However, transfer function methods do not easily allow for optimal control with state variables, since they only model the input/output behavior of the system.

In parallel with the system identification community, identification of autoregressive time-series models is a widely studied topic in the statistics literature (see, e.g., Box et al. [7] for the classical results). Goldenshluger and Zeevi [24] show that the coefficients of a stationary auto-regressive model can be estimated from a single trajectory of length polynomial in $1/(1 - \rho)$ via least squares, where ρ denotes the stability radius of the process. They also prove that their rate is minimax optimal. More recently, several authors [33,39,42] have studied generalization bounds for non-i.i.d. data, extending the standard learning theory guarantees for independent data. At the crux of these arguments lie various mixing assumptions [63], which limits the analysis to only hold for stable dynamical systems. Results in this line of research suggest that systems with smaller mixing time (i.e., systems that are more stable) are easier to identify (i.e., take less samples). Our result in Proposition 1, however, suggests instead that identification benefits from more easily excitable systems. While our analysis holds when we have access to full state observations, empirical testing suggests that Proposition 1 reflects reality more accurately than arguments based on mixing. In follow-up work, we have begun to reconcile this issue for stable linear systems [52].

Robust Controller Design

For end-to-end guarantees, parameter estimation is only half the picture. Our procedure provides us with a family of system models described by a nominal estimate and a set of unknown but bounded model errors. It is therefore necessary to ensure that the computed controller has stability and performance guarantees for any such admissible realization. The problem of robustly stabilizing such a family of systems is one with a rich history in the controls community. When modeling errors to the nominal system are allowed to be arbitrary norm-bounded linear time-invariant (LTI) operators in feedback with the nominal plant, traditional small-gain theorems and robust synthesis techniques can be applied to exactly solve the problem [13,64]. However, when the errors are known to have more structure, there are more sophisticated techniques based on structured singular values and corresponding μ -synthesis techniques [15,19,44,62] or integral quadratic constraints (IQCs) [40]. While theoretically appealing and much less conservative than traditional small-gain approaches, the resulting synthesis methods are both computationally intractable (although effective heuristics do exist) and difficult to interpret analytically. In particular, we know of no results in the literature that bound the degradation in performance of controlling an uncertain system in terms of the size of the perturbations affecting it.

To circumvent this issue, we leverage a novel parameterization of robustly stabilizing controllers based on the SLS framework for controller synthesis [59]. We describe this framework in more detail in Sect. 3. Originally developed to allow for scaling optimal and robust controller synthesis techniques to large-scale systems, the SLS framework can be viewed as a generalization of the celebrated Youla parameterization [61]. We show that SLS allows us to account for model uncertainty in a transparent and analytically tractable way.

PAC Learning and Reinforcement Learning

In terms of end-to-end guarantees for LQR, our work is most comparable to that of Fiechter [22], who shows that the *discounted* LQR problem is PAC-learnable. Fiechter analyzes an identify-then-control scheme similar to the one we propose, but there are several key differences. First, our probabilistic bounds on identification are much sharper, by leveraging modern tools from high-dimensional statistics. Second, Fiechter implicitly assumes that the true closed-loop system with the estimated controller is not only stable but also contractive. While this very strong assumption is nearly impossible to verify in practice, contractive closed-loop assumptions are actually pervasive throughout the literature, as we describe below. To the best of our knowledge, our work is the first to properly lift this technical restriction. Third, and most importantly, Fiechter proposes to directly solve the discounted LQR problem with the identified model and does not take into account any uncertainty in the controller synthesis step. This is problematic for two reasons. First, it is easy to construct an instance of a discounted LQR problem where the *optimal* solution does not stabilize the true system (see, e.g., [47]). Therefore, even in the limit of infinite data, there is no guarantee that the closed-loop system will be stable. Second, even if the optimal solution does stabilize the underlying system, failing to take uncertainty into account can lead to situations where the synthesized controller does not. We will demonstrate this behavior in our experiments.

We are also particularly interested in the LQR problem as a baseline for more complicated problems in reinforcement learning (RL). LQR should be a relatively easy problem in RL because one can learn the dynamics from anywhere in the state space, vastly simplifying the problem of exploration. Hence, it is important to establish how well a pure exploration followed by exploitation strategy can fare on this simple baseline.

There are indeed some related efforts in RL and online learning. Abbasi-Yadkori and Szepesvari [1] propose to use the optimism in the face of uncertainty (OFU) principle for the LQR problem, by maintaining confidence ellipsoids on the true parameter, and using the controller which, in feedback, minimizes the cost objective the most among all systems in the confidence ellipsoid. Ignoring the computational intractability of this approach, their analysis reveals an exponential dependence in the system order in their regret bound and also makes the very strong assumption that the optimal closed-loop systems are contractive for every A, B in the confidence ellipsoid. The regret bound is improved by Ibrahimi et al. [30] to depend linearly on the state dimension under additional sparsity constraints on the dynamics.

In response to the computational intractability of the OFU principle, researchers in RL and online learning proposed the use of Thompson sampling [49] for exploration. Abeille and Lazaric [2] show that the regret of a Thompson sampling approach for LQR scales as $\tilde{O}(T^{2/3})$ and improve the result to $\tilde{O}(\sqrt{T})$ in [3], where $\tilde{O}(\cdot)$ hides poly-logarithmic factors. However, their results are only valid for the scalar $n = d = 1$ setting. Ouyang et al. [43] show that in a *Bayesian* setting, the expected regret can be bounded by $\tilde{O}(\sqrt{T})$. While this matches the bound of [1], the Bayesian regret is with respect to a particular Gaussian prior distribution over the true model, which differs from the frequentist setting considered in [1–3]. Furthermore, these works also make

the same restrictive assumption that the optimal closed-loop systems are uniformly contractive over some known set.

Jiang et al. [31] propose a general exploration algorithm for contextual decision processes (CDPs) and show that CDPs with low *Bellman rank* are PAC-learnable; in the LQR setting, they show the Bellman rank is bounded by n^2 . While this result is appealing from an information-theoretic standpoint, the proposed algorithm is computationally intractable for continuous problems. Hazan et al. [27,28] study the problem of prediction in a linear dynamical system via a novel spectral filtering algorithm. Their main result shows that one can compete in a regret setting in terms of prediction error. As mentioned previously, converting prediction error bounds into concrete bounds on sub-optimality of control performance is an open question. Fazel et al. [20] show that randomized search algorithms similar to policy gradient can learn the optimal controller with a polynomial number of samples in the noiseless case; an explicit characterization of the dependence of the sample complexity on the parameters of the true system is not given.

2 System Identification Through Least Squares

To estimate a coarse model of the unknown system dynamics, we turn to the simple and classical method of linear least squares. By running experiments in which the system starts at $x_0 = 0$ and the dynamics evolve with a given input, we can record the resulting state observations. The set of inputs and outputs from each such experiment will be called a *rollout*. For system estimation, we excite the system with Gaussian noise for N rollouts, each of length T . The resulting dataset is $\{(x_t^{(\ell)}, u_t^{(\ell)}) : 1 \leq \ell \leq N, 0 \leq t \leq T\}$, where t indexes the time in one rollout and ℓ indexes independent rollouts. Therefore, we can estimate the system dynamics by

$$(\widehat{A}, \widehat{B}) \in \arg \min_{(A, B)} \sum_{\ell=1}^N \sum_{t=0}^{T-1} \frac{1}{2} \|Ax_t^{(\ell)} + Bu_t^{(\ell)} - x_{t+1}^{(\ell)}\|_2^2. \tag{9}$$

For the Coarse-ID control setting, a good estimate of error is just as important as the estimate of the dynamics. Statistical theory and tools allow us to quantify the error of the least squares estimator. First, we present a theoretical analysis of the error in a simplified setting. Then, we describe a computational bootstrap procedure for error estimation from data alone.

2.1 Least Squares Estimation as a Random Matrix Problem

We begin by explicitly writing the form of the least squares estimator. First, fixing notation to simplify the presentation, let $\Theta := [A \ B]^* \in \mathbb{R}^{(n+p) \times n}$ and let $z_t := \begin{bmatrix} x_t \\ u_t \end{bmatrix} \in \mathbb{R}^{n+p}$. Then, the system dynamics can be rewritten, for all $t \geq 0$,

$$x_{t+1}^* = z_t^* \Theta + w_t^* .$$

Then in a single rollout, we will collect

$$X := \begin{bmatrix} x_1^* \\ x_2^* \\ \vdots \\ x_T^* \end{bmatrix}, \quad Z := \begin{bmatrix} z_0^* \\ z_1^* \\ \vdots \\ z_{T-1}^* \end{bmatrix}, \quad W := \begin{bmatrix} w_0^* \\ w_1^* \\ \vdots \\ w_{T-1}^* \end{bmatrix}. \tag{10}$$

The system dynamics give the identity $X = Z\Theta + W$. Resetting state of the system to $x_0 = 0$ each time, we can perform N rollouts and collect N datasets like (10). Having the ability to reset the system to a state independent of past observations will be important for the analysis in the following section, and it is also practically important for potentially unstable systems. Denote the data for each rollout as $(X^{(\ell)}, Z^{(\ell)}, W^{(\ell)})$. With slight abuse of notation, let X_N be composed of vertically stacked $X^{(\ell)}$, and similarly for Z_N and W_N . Then, we have

$$X_N = Z_N\Theta + W_N.$$

The full data least squares estimator for Θ is (assuming for now invertibility of $Z_N^*Z_N$),

$$\widehat{\Theta} = (Z_N^*Z_N)^{-1}Z_N^*X_N = \Theta + (Z_N^*Z_N)^{-1}Z_N^*W_N. \tag{11}$$

Then the estimation error is given by

$$E := \widehat{\Theta} - \Theta = (Z_N^*Z_N)^{-1}Z_N^*W_N. \tag{12}$$

The magnitude of this error is the quantity of interest in determining confidence sets around estimates $(\widehat{A}, \widehat{B})$. However, since W_N and Z_N are not independent, this estimator is difficult to analyze using standard methods. While this type of analysis is an open problem of interest, in this paper we turn instead to a simplified estimator.

2.2 Theoretical Bounds on Least Squares Error

In this section, we work out the statistical rate for the least squares estimator which uses just the last sample of each trajectory $(x_T^{(\ell)}, x_{T-1}^{(\ell)}, u_{T-1}^{(\ell)})$. This estimation procedure is made precise in Algorithm 1. Our analysis ideas are analogous to those used to prove statistical rates for standard linear regression, and they leverage recent tools in nonasymptotic analysis of random matrices. The result is presented above in Proposition 1.

In the context of Proposition 1, a single data point from each T -step rollout is used. We emphasize that this strategy results in independent data, which can be seen by defining the estimator matrix directly. The previous estimator (11) is amended as follows: the matrices defined in (10) instead include only the final time step of each trial, $X_N = [x_T^{(1)} \ x_T^{(2)} \ \dots \ x_T^{(N)}]^*$, and similar modifications are made to Z_N and W_N . Estimator (11) uses these modified matrices, which now contain independent rows.

Algorithm 1 Estimation of linear dynamics with independent data

```

1: for  $\ell$  from 1 to  $N$  do
2:    $x_0^{(\ell)} = 0$ 
3:   for  $t$  from 0 to  $T - 1$  do
4:      $x_{t+1}^{(\ell)} = Ax_t^{(\ell)} + Bu_t^{(\ell)} + w_t^{(\ell)}$  with  $w_t^{(\ell)} \stackrel{\text{i.i.d.}}{\sim} \mathcal{N}(0, \sigma_w^2 I_n)$  and  $u_t^{(\ell)} \stackrel{\text{i.i.d.}}{\sim} \mathcal{N}(0, \sigma_u^2 I_p)$ .
5:   end for
6: end for
7:  $(\hat{A}, \hat{B}) \in \arg \min_{(A,B)} \sum_{\ell=1}^N \frac{1}{2} \|Ax_{T-1}^{(\ell)} + Bu_{T-1}^{(\ell)} - x_T^{(\ell)}\|_2^2$ 
    
```

To see this, recall the definition of G_T and F_T from (4),

$$G_T = [A^{T-2}B \ A^{T-3}B \ \dots \ B], \quad F_T = [A^{T-2} \ A^{T-3} \ \dots \ I_n],$$

defined for $T \geq 2$. We can unroll the system dynamics and see that

$$x_{T-1} = G_T \begin{bmatrix} u_0 \\ u_1 \\ \vdots \\ u_{T-2} \end{bmatrix} + F_T \begin{bmatrix} w_0 \\ w_1 \\ \vdots \\ w_{T-2} \end{bmatrix}. \tag{13}$$

Using Gaussian excitation, $u_t \sim \mathcal{N}(0, \sigma_u^2 I_p)$ gives

$$\begin{bmatrix} x_{T-1} \\ u_{T-1} \end{bmatrix} \sim \mathcal{N}\left(0, \begin{bmatrix} \sigma_u^2 G_T G_T^* + \sigma_w^2 F_T F_T^* & 0 \\ 0 & \sigma_u^2 I_p \end{bmatrix}\right). \tag{14}$$

Since $F_T F_T^* > 0$, as long as both σ_u, σ_w are positive, this is a nondegenerate distribution.

Therefore, bounding the estimation error can be achieved via proving a result on the error in random design linear regression with vector-valued observations. First, we present a lemma which bounds the spectral norm of the product of two independent Gaussian matrices.

Lemma 1 Fix a $\delta \in (0, 1)$ and $N \geq 2(n + m) \log(1/\delta)$. Let $f_k \in \mathbb{R}^m, g_k \in \mathbb{R}^n$ be independent random vectors $f_k \sim \mathcal{N}(0, \Sigma_f)$ and $g_k \sim \mathcal{N}(0, \Sigma_g)$ for $1 \leq k \leq N$. With probability at least $1 - \delta$,

$$\left\| \sum_{k=1}^N f_k g_k^* \right\|_2 \leq 4 \|\Sigma_f\|_2^{1/2} \|\Sigma_g\|_2^{1/2} \sqrt{N(m + n) \log(9/\delta)}.$$

We believe this bound to be standard and include a proof in the appendix for completeness. Lemma 1 shows that if X is $n_1 \times N$ with i.i.d. $\mathcal{N}(0, 1)$ entries and Y is $N \times n_2$ with i.i.d. $\mathcal{N}(0, 1)$ entries, and X and Y are independent, then with probability at least $1 - \delta$ we have

$$\|XY\|_2 \leq 4\sqrt{N(n_1 + n_2) \log(9/\delta)}.$$

Next, we state a standard nonasymptotic bound on the minimum singular value of a standard Wishart matrix (see, e.g., Corollary 5.35 of [56]).

Lemma 2 *Let $X \in \mathbb{R}^{N \times n}$ have i.i.d. $\mathcal{N}(0, 1)$ entries. With probability at least $1 - \delta$,*

$$\sqrt{\lambda_{\min}(X^*X)} \geq \sqrt{N} - \sqrt{n} - \sqrt{2 \log(1/\delta)}.$$

We combine the previous lemmas into a statement on the error of random design regression.

Lemma 3 *Let $z_1, \dots, z_N \in \mathbb{R}^n$ be i.i.d. from $\mathcal{N}(0, \Sigma)$ with Σ invertible. Let $Z^* := [z_1 \dots z_N]$. Let $W \in \mathbb{R}^{N \times p}$ with each entry i.i.d. $\mathcal{N}(0, \sigma_w^2)$ and independent of Z . Let $E := (Z^*Z)^\dagger Z^*W$, and suppose that*

$$N \geq 8n + 16 \log(2/\delta). \tag{15}$$

For any fixed matrix Q , we have with probability at least $1 - \delta$,

$$\|QE\|_2 \leq 16\sigma_w \|Q\Sigma^{-1/2}\|_2 \sqrt{\frac{(n+p) \log(18/\delta)}{N}}.$$

Proof First, observe that Z is equal in distribution to $X\Sigma^{1/2}$, where $X \in \mathbb{R}^{N \times n}$ has i.i.d. $\mathcal{N}(0, 1)$ entries. By Lemma 2, with probability at least $1 - \delta/2$,

$$\sqrt{\lambda_{\min}(X^*X)} \geq \sqrt{N} - \sqrt{n} - \sqrt{2 \log(2/\delta)} \geq \sqrt{N}/2.$$

The last inequality uses (15) combined with the inequality $(a + b)^2 \leq 2(a^2 + b^2)$. Furthermore, by Lemma 1 and (15), with probability at least $1 - \delta/2$,

$$\|X^*W\|_2 \leq 4\sigma_w \sqrt{N(n+p) \log(18/\delta)}.$$

Let \mathcal{E} denote the event which is the intersection of the two previous events. By a union bound, $\mathbb{P}(\mathcal{E}) \geq 1 - \delta$. We continue the rest of the proof assuming the event \mathcal{E} holds. Since X^*X is invertible,

$$QE = Q(Z^*Z)^\dagger Z^*W = Q(\Sigma^{1/2} X^* X \Sigma^{1/2})^\dagger \Sigma^{1/2} X^*W = Q\Sigma^{-1/2} (X^*X)^{-1} X^*W.$$

Taking operator norms on both sides,

$$\|QE\|_2 \leq \|Q\Sigma^{-1/2}\|_2 \|(X^*X)^{-1}\|_2 \|X^*W\|_2 = \|Q\Sigma^{-1/2}\|_2 \frac{\|X^*W\|_2}{\lambda_{\min}(X^*X)}.$$

Combining the inequalities above,

$$\frac{\|X^*W\|_2}{\lambda_{\min}(X^*X)} \leq 16\sigma_w \sqrt{\frac{(n+p) \log(18/\delta)}{N}}.$$

The result now follows. □

Using this result on random design linear regression, we are now ready to analyze the estimation errors of the identification in Algorithm 1 and provide a proof of Proposition 1.

Proof Consider the least squares estimation error (12) with modified single-sample-per-rollout matrices. Recall that rows of the design matrix Z_N are distributed as independent normals, as in (14). Then applying Lemma 3 with $Q_A = [I_n \ 0]$ so that $Q_A E$ extracts only the estimate for A , we conclude that with probability at least $1 - \delta/2$,

$$\|\widehat{A} - A\|_2 \leq \frac{16\sigma_w}{\sqrt{\lambda_{\min}(\sigma_u^2 G_T G_T^* + \sigma_w^2 F_T F_T^*)}} \sqrt{\frac{(2n + p) \log(36/\delta)}{N}}, \tag{16}$$

as long as $N \geq 8(n + p) + 16 \log(4/\delta)$. Now applying Lemma 3 under the same condition on N with $Q_B = [0 \ I_p]$, we have with probability at least $1 - \delta/2$,

$$\|\widehat{B} - B\|_2 \leq \frac{16\sigma_w}{\sigma_u} \sqrt{\frac{(2n + p) \log(36/\delta)}{N}}. \tag{17}$$

The result follows by application of the union bound. □

There are several interesting points to make about the guarantees offered by Proposition 1. First, as mentioned in the introduction, there are $n(n + p)$ parameters to learn and our bound states that we need $O(n + p)$ measurements, each measurement providing n values. Hence, this appears to be an optimal dependence with respect to the parameters n and p . Second, note that intuitively, if the system amplifies the control and noise inputs in all directions of the state space, as captured by the minimum eigenvalues of the control and disturbance Gramians $G_T G_T^*$ or $F_T F_T^*$, respectively, then the system has a larger “signal-to-noise” ratio, and the system matrix A is easier to estimate. On the other hand, this measure of the excitability of the system has no impact on learning B . Unlike in Fiechter’s work [22], we do not need to assume that $G_T G_T^*$ is invertible. As long as the process noise is not degenerate, it will excite all modes of the system.

Finally, we note that Proposition 1 offers a data-independent guarantee for the estimation of the parameters (A, B) . We can also provide data-dependent guarantees, which will be less conservative in practice. The next result shows how we can use the observed states and inputs to obtain more refined confidence sets than the ones offered by Proposition 1. The proof is deferred to Appendix B.

Proposition 3 *Assume we have N independent samples $(y^{(\ell)}, x^{(\ell)}, u^{(\ell)})$ such that*

$$y^{(\ell)} = Ax^{(\ell)} + Bu^{(\ell)} + w^{(\ell)},$$

where $w^{(\ell)}$ are i.i.d. $\mathcal{N}(0, \sigma_w^2 I_n)$ and are independent from $x^{(\ell)}$ and $u^{(\ell)}$. Also, let us assume that $N \geq n + p$. Then, with probability $1 - \delta$, we have

$$\begin{bmatrix} (\widehat{A} - A)^\top \\ (\widehat{B} - B)^\top \end{bmatrix} [(\widehat{A} - A) (\widehat{B} - B)] \leq C(n, p, \delta) \left(\sum_{\ell=1}^N \begin{bmatrix} x^{(\ell)} \\ u^{(\ell)} \end{bmatrix} [(x^{(\ell)})^\top (u^{(\ell)})^\top] \right)^{-1},$$

where $C(n, p, \delta) = \sigma_w^2(\sqrt{n+p} + \sqrt{n} + \sqrt{2 \log(1/\delta)})^2$. If the matrix on the right-hand side has zero as an eigenvalue, we define the inverse of that eigenvalue to be infinity.

Proposition 3 is a general result that does not require the inputs $u^{(\ell)}$ to be normally distributed and it allows the states $x^{(\ell)}$ to be arbitrary as long as all the samples $(y^{(\ell)}, x^{(\ell)}, u^{(\ell)})$ are independent and the process noise $w^{(\ell)}$ is normally distributed. Nonetheless, both Propositions 1 and 3 require estimating (A, B) from independent samples. In practice, one would collect rollouts from the system, which consist of many dependent measurements. In that case, using all the data is preferable. Since the guarantees offered in this section do not apply in that case, in the next section we study a different procedure for estimating the size of the estimation error.

2.3 Estimating Model Uncertainty with the Bootstrap

In the previous sections, we offered theoretical guarantees on the performance of the least squares estimation of A and B from independent samples. However, there are two important limitations to using such guarantees in practice to offer upper bounds on $\epsilon_A = \|A - \widehat{A}\|_2$ and $\epsilon_B = \|B - \widehat{B}\|_2$. First, using only one sample per system rollout is empirically less efficient than using all available data for estimation. Second, even optimal statistical analyses often do not recover constant factors that match practice. For purposes of robust control, it is important to obtain upper bounds on ϵ_A and ϵ_B that are not too conservative. Thus, we aim to find $\widehat{\epsilon}_A$ and $\widehat{\epsilon}_B$ such that $\epsilon_A \leq \widehat{\epsilon}_A$ and $\epsilon_B \leq \widehat{\epsilon}_B$ with high probability.

We propose a vanilla bootstrap method for estimating $\widehat{\epsilon}_A$ and $\widehat{\epsilon}_B$. Bootstrap methods have had a profound impact in both theoretical and applied statistics since their introduction [18]. These methods are used to estimate statistical quantities (e.g., confidence intervals) by sampling synthetic data from an empirical distribution determined by the available data. For the problem at hand, we propose the procedure described in Algorithm 2.¹

For $\widehat{\epsilon}_A$ and $\widehat{\epsilon}_B$ estimated by Algorithm 2, we intuitively have

$$\mathbb{P}(\|A - \widehat{A}\|_2 \leq \widehat{\epsilon}_A) \approx 1 - \delta \quad \text{and} \quad \mathbb{P}(\|B - \widehat{B}\|_2 \leq \widehat{\epsilon}_B) \approx 1 - \delta.$$

There are many known guarantees for the bootstrap, particularly for the parametric version we use. We do not discuss these results here; for more details, see texts by Van Der Vaart and Wellner [55], Shao and Tu [50], and Hall [25]. Instead, in Appendix F we show empirically the performance of the bootstrap for our estimation problem. For mission critical systems, where empirical validation is insufficient, the statistical

¹ We assume that σ_u and σ_w are known. Otherwise, they can be estimated from data.

Algorithm 2 Bootstrap estimation of ϵ_A and ϵ_B

1: **Input:** confidence parameter δ , number of trials M , data $\{(x_t^{(i)}, u_t^{(i)})\}_{1 \leq i \leq N, 1 \leq t \leq T}$, and $(\widehat{A}, \widehat{B})$ a minimizer of $\sum_{\ell=1}^N \sum_{t=0}^{T-1} \frac{1}{2} \|Ax_t^{(\ell)} + Bu_t^{(\ell)} - x_{t+1}^{(\ell)}\|_2^2$.
 2: **for** M trials **do**
 3: **for** ℓ from 1 to N **do**
 4: $\widehat{x}_0^{(\ell)} = x_0^{(\ell)}$
 5: **for** t from 0 to $T - 1$ **do**
 6: $\widehat{x}_{t+1}^{(\ell)} = \widehat{A}\widehat{x}_t^{(\ell)} + \widehat{B}\widehat{u}_t^{(\ell)} + \widehat{w}_t^{(\ell)}$ with $\widehat{w}_t^{(\ell)} \stackrel{\text{i.i.d.}}{\sim} \mathcal{N}(0, \sigma_w^2 I_n)$ and $\widehat{u}_t^{(\ell)} \stackrel{\text{i.i.d.}}{\sim} \mathcal{N}(0, \sigma_u^2 I_p)$.
 7: **end for**
 8: **end for**
 9: $(\widetilde{A}, \widetilde{B}) \in \arg \min_{(A, B)} \sum_{\ell=1}^N \sum_{t=0}^{T-1} \frac{1}{2} \|A\widehat{x}_t^{(\ell)} + B\widehat{u}_t^{(\ell)} - \widehat{x}_{t+1}^{(\ell)}\|_2^2$.
 10: record $\widetilde{\epsilon}_A = \|\widetilde{A} - \widehat{A}\|_2$ and $\widetilde{\epsilon}_B = \|\widetilde{B} - \widehat{B}\|_2$.
 11: **end for**
 12: **Output:** $\widehat{\epsilon}_A$ and $\widehat{\epsilon}_B$, the $100(1 - \delta)$ th percentiles of the $\widetilde{\epsilon}_A$'s and the $\widetilde{\epsilon}_B$'s.

error bounds presented in Sect. 2.2 give guarantees on the size of ϵ_A, ϵ_B . In general, data-dependent error guarantees will be less conservative. In follow-up work, we offer guarantees similar to the ones presented in Sect. 2.2 for estimation of linear dynamics from dependent data [52].

3 Robust Synthesis

With estimates of the system $(\widehat{A}, \widehat{B})$ and operator norm error bounds (ϵ_A, ϵ_B) in hand, we now turn to control design. In this section, we introduce some useful tools from *System Level Synthesis* (SLS), a recently developed approach to control design that relies on a particular parameterization of signals in a control system [38,59]. We review the main SLS framework, highlighting the key constructions that we will use to solve the robust LQR problem. As we show in this and the following section, using the SLS framework, as opposed to traditional techniques from robust control, allows us to (a) compute robust controllers using semidefinite programming and (b) provide sub-optimality guarantees in terms of the size of the uncertainties on our system estimates.

3.1 Useful Results from System Level Synthesis

The SLS framework focuses on the *system responses* of a closed-loop system. As a motivating example, consider linear dynamics under a fixed static state-feedback control policy K , i.e., let $u_k = Kx_k$. Then, the closed-loop map from the disturbance process $\{w_0, w_1, \dots\}$ to the state x_k and control input u_k at time k is given by

$$\begin{aligned} x_k &= \sum_{t=1}^k (A + BK)^{k-t} w_{t-1}, \\ u_k &= \sum_{t=1}^k K(A + BK)^{k-t} w_{t-1}. \end{aligned} \tag{18}$$

Letting $\Phi_x(k) := (A + BK)^{k-1}$ and $\Phi_u(k) := K(A + BK)^{k-1}$, we can rewrite Eq. (18) as

$$\begin{bmatrix} x_k \\ u_k \end{bmatrix} = \sum_{t=1}^k \begin{bmatrix} \Phi_x(k-t+1) \\ \Phi_u(k-t+1) \end{bmatrix} w_{t-1}, \tag{19}$$

where $\{\Phi_x(k), \Phi_u(k)\}$ are called the *closed-loop system response elements* induced by the static controller K .

Note that even when the control is a linear function of the state and its past history (i.e., a linear dynamic controller), expression (19) is valid. Though we conventionally think of the control policy as a function mapping states to input, whenever such a mapping is linear, both the control input and the state can be written as linear functions of the disturbance signal w_t . With such an identification, the dynamics require that the $\{\Phi_x(k), \Phi_u(k)\}$ must obey the constraints

$$\Phi_x(k+1) = A\Phi_x(k) + B\Phi_u(k), \quad \Phi_x(1) = I, \quad \forall k \geq 1, \tag{20}$$

As we describe in more detail below in Theorem 1, these constraints are in fact both necessary and sufficient. Working with closed-loop system responses allows us to cast optimal control problems as optimization problems over elements $\{\Phi_x(k), \Phi_u(k)\}$, constrained to satisfy the affine Eq. (20). Comparing Eqs. (18) and (19), we see that the former is nonconvex in the controller K , whereas the latter is affine in the elements $\{\Phi_x(k), \Phi_u(k)\}$.

As we work with infinite horizon problems, it is notationally more convenient to work with *transfer function* representations of the above objects, which can be obtained by taking a z -transform of their time-domain representations. The frequency-domain variable z can be informally thought of as the time-shift operator, i.e., $z\{x_k, x_{k+1}, \dots\} = \{x_{k+1}, x_{k+2}, \dots\}$, allowing for a compact representation of LTI dynamics. We use boldface letters to denote such transfer functions signals in the frequency domain, e.g., $\Phi_x(z) = \sum_{k=1}^{\infty} \Phi_x(k)z^{-k}$. Then, constraints (20) can be rewritten as

$$[zI - A - B] \begin{bmatrix} \Phi_x \\ \Phi_u \end{bmatrix} = I,$$

and the corresponding (not necessarily static) control law $\mathbf{u} = \mathbf{K}\mathbf{x}$ is given by $\mathbf{K} = \Phi_u \Phi_x^{-1}$. The relevant frequency-domain connections for LQR are illustrated in Appendix C.

We formalize our discussion by introducing notation that is common in the controls literature. For a thorough introduction to the functional analysis commonly used in control theory, see Chapters 2 and 3 of [64]. Let \mathbb{T} (resp. \mathbb{D}) denote the unit circle (resp. open unit disk) in the complex plane. The restriction of the Hardy spaces $\mathcal{H}_\infty(\mathbb{T})$ and $\mathcal{H}_2(\mathbb{T})$ to matrix-valued real-rational functions that are analytic on the complement of \mathbb{D} will be referred to as \mathcal{RH}_∞ and \mathcal{RH}_2 , respectively. In controls parlance, this corresponds to (discrete-time) stable matrix-valued transfer functions. For these two function spaces, the \mathcal{H}_∞ and \mathcal{H}_2 norms simplify to

$$\|\mathbf{G}\|_{\mathcal{H}_\infty} = \sup_{z \in \mathbb{T}} \|G(z)\|_2, \quad \|\mathbf{G}\|_{\mathcal{H}_2} = \sqrt{\frac{1}{2\pi} \int_{\mathbb{T}} \|G(z)\|_F^2 dz}. \tag{21}$$

Finally, the notation $\frac{1}{z}\mathcal{RH}_\infty$ refers to the set of transfer functions \mathbf{G} such that $z\mathbf{G} \in \mathcal{RH}_\infty$. Equivalently, $\mathbf{G} \in \frac{1}{z}\mathcal{RH}_\infty$ if $\mathbf{G} \in \mathcal{RH}_\infty$ and \mathbf{G} is strictly proper.

The most important transfer function for the LQR problem is the map from the state sequence to the control actions: the control policy. Consider an arbitrary transfer function \mathbf{K} denoting the map from state to control action, $\mathbf{u} = \mathbf{K}\mathbf{x}$. Then the closed-loop transfer matrices from the process noise \mathbf{w} to the state \mathbf{x} and control action \mathbf{u} satisfy

$$\begin{bmatrix} \mathbf{x} \\ \mathbf{u} \end{bmatrix} = \begin{bmatrix} (zI - A - B\mathbf{K})^{-1} \\ \mathbf{K}(zI - A - B\mathbf{K})^{-1} \end{bmatrix} \mathbf{w}. \tag{22}$$

We then have the following theorem parameterizing the set of stable closed-loop transfer matrices, as described in Eq. (22), that are achievable by a given stabilizing controller \mathbf{K} .

Theorem 1 (State-Feedback Parameterization [59]) *The following are true:*

- The affine subspace defined by

$$[zI - A - B] \begin{bmatrix} \Phi_x \\ \Phi_u \end{bmatrix} = I, \quad \Phi_x, \Phi_u \in \frac{1}{z}\mathcal{RH}_\infty \tag{23}$$

parameterizes all system responses (22) from \mathbf{w} to (\mathbf{x}, \mathbf{u}) , achievable by an internally stabilizing state-feedback controller \mathbf{K} .

- For any transfer matrices $\{\Phi_x, \Phi_u\}$ satisfying (23), the controller $\mathbf{K} = \Phi_u \Phi_x^{-1}$ is internally stabilizing and achieves the desired system response (22).

Note that in particular, $\{\Phi_x, \Phi_u\} = \{(zI - A - B\mathbf{K})^{-1}, \mathbf{K}(zI - A - B\mathbf{K})^{-1}\}$ as in (22) are elements of the affine space defined by (23) whenever \mathbf{K} is a causal stabilizing controller.

We will also make extensive use of a robust variant of Theorem 1.

Theorem 2 (Robust Stability [38]) *Let Φ_x and Φ_u be two transfer matrices in $\frac{1}{z}\mathcal{RH}_\infty$ such that*

$$[zI - A - B] \begin{bmatrix} \Phi_x \\ \Phi_u \end{bmatrix} = I + \Delta. \tag{24}$$

Then the controller $\mathbf{K} = \Phi_u \Phi_x^{-1}$ stabilizes the system described by (A, B) if and only if $(I + \Delta)^{-1} \in \mathcal{RH}_\infty$. Furthermore, the resulting system response is given by

$$\begin{bmatrix} \mathbf{x} \\ \mathbf{u} \end{bmatrix} = \begin{bmatrix} \Phi_x \\ \Phi_u \end{bmatrix} (I + \Delta)^{-1} \mathbf{w}. \tag{25}$$

Corollary 1 *Under the assumptions of Theorem 2, if $\|\Delta\| < 1$ for any induced norm $\|\cdot\|$, then the controller $\mathbf{K} = \Phi_u \Phi_x^{-1}$ stabilizes the system described by (A, B) .*

Proof Follows immediately from the small-gain theorem, see for example Sect. 9.2 in [64]. □

3.2 Robust LQR Synthesis

We return to the problem setting where estimates $(\widehat{A}, \widehat{B})$ of a true system (A, B) satisfy

$$\|\Delta_A\|_2 \leq \epsilon_A, \quad \|\Delta_B\|_2 \leq \epsilon_B$$

where $\Delta_A := \widehat{A} - A$ and $\Delta_B := \widehat{B} - B$ and where we wish to minimize the LQR cost for the worst instantiation of the parametric uncertainty.

Before proceeding, we must formulate the LQR problem in terms of the system responses $\{\Phi_x(k), \Phi_u(k)\}$. It follows from Theorem 1 and the standard equivalence between infinite horizon LQR and \mathcal{H}_2 optimal control that, for a disturbance process distributed as $w_t \stackrel{i.i.d.}{\sim} \mathcal{N}(0, \sigma_w^2 I)$, the standard LQR problem (3) can be equivalently written as

$$\underset{\Phi_x, \Phi_u}{\text{minimize}} \sigma_w^2 \left\| \begin{bmatrix} Q^{\frac{1}{2}} & 0 \\ 0 & R^{\frac{1}{2}} \end{bmatrix} \begin{bmatrix} \Phi_x \\ \Phi_u \end{bmatrix} \right\|_{\mathcal{H}_2}^2 \quad \text{s.t. Eq. (23)}. \tag{26}$$

We provide a full derivation of this equivalence in Appendix C. Going forward, we drop the σ_w^2 multiplier in the objective function as it affects neither the optimal controller nor the sub-optimality guarantees that we compute in Sect. 4.

We begin with a simple sufficient condition under which any controller \mathbf{K} that stabilizes $(\widehat{A}, \widehat{B})$ also stabilizes the true system (A, B) . To state the lemma, we introduce one additional piece of notation. For a matrix M , we let \mathfrak{R}_M denote the resolvent

$$\mathfrak{R}_M := (zI - M)^{-1}. \tag{27}$$

We now can state our robustness lemma.

Lemma 4 *Let the controller \mathbf{K} stabilize $(\widehat{A}, \widehat{B})$ and (Φ_x, Φ_u) be its corresponding system response (22) on system $(\widehat{A}, \widehat{B})$. Then if \mathbf{K} stabilizes (A, B) , it achieves the following LQR cost*

$$J(A, B, \mathbf{K}) := \left\| \begin{bmatrix} Q^{\frac{1}{2}} & 0 \\ 0 & R^{\frac{1}{2}} \end{bmatrix} \begin{bmatrix} \Phi_x \\ \Phi_u \end{bmatrix} \left(I + \begin{bmatrix} \Delta_A & \Delta_B \end{bmatrix} \begin{bmatrix} \Phi_x \\ \Phi_u \end{bmatrix} \right)^{-1} \right\|_{\mathcal{H}_2}. \tag{28}$$

Furthermore, letting

$$\widehat{\Delta} := \begin{bmatrix} \Delta_A & \Delta_B \end{bmatrix} \begin{bmatrix} \Phi_x \\ \Phi_u \end{bmatrix} = (\Delta_A + \Delta_B \mathbf{K}) \mathfrak{R}_{\widehat{A} + \widehat{B} \mathbf{K}}. \tag{29}$$

a sufficient condition for \mathbf{K} to stabilize (A, B) is that $\|\widehat{\Delta}\|_{\mathcal{H}_\infty} < 1$.

Proof Follows immediately from Theorems 1, 2 and Corollary 1 by noting that for system responses (Φ_x, Φ_u) satisfying

$$[zI - \widehat{A} - \widehat{B}] \begin{bmatrix} \Phi_x \\ \Phi_u \end{bmatrix} = I,$$

it holds that

$$[zI - A - B] \begin{bmatrix} \Phi_x \\ \Phi_u \end{bmatrix} = I + \widehat{\Delta}$$

for $\widehat{\Delta}$ as defined in Eq. (29). □

We can therefore recast the robust LQR problem (7) in the following equivalent form

$$\begin{aligned} &\underset{\Phi_x, \Phi_u}{\text{minimize}} \quad \sup_{\substack{\|\Delta_A\|_2 \leq \epsilon_A \\ \|\Delta_B\|_2 \leq \epsilon_B}} J(A, B, \mathbf{K}) \\ &\text{s.t.} \quad [zI - \widehat{A} - \widehat{B}] \begin{bmatrix} \Phi_x \\ \Phi_u \end{bmatrix} = I, \quad \Phi_x, \Phi_u \in \frac{1}{z} \mathcal{RH}_\infty. \end{aligned} \tag{30}$$

The resulting robust control problem is one subject to real-parametric uncertainty, a class of problems known to be computationally intractable [8]. Although effective computational heuristics (e.g., DK iteration [64]) exist, the performance of the resulting controller on the true system is difficult to characterize analytically in terms of the size of the perturbations.

To circumvent this issue, we take a slightly conservative approach and find an upper bound to the cost $J(A, B, \mathbf{K})$ that is independent of the uncertainties Δ_A and Δ_B . First, note that if $\|\widehat{\Delta}\|_{\mathcal{H}_\infty} < 1$, we can write

$$J(A, B, \mathbf{K}) \leq \|(I + \widehat{\Delta})^{-1}\|_{\mathcal{H}_\infty} J(\widehat{A}, \widehat{B}, \mathbf{K}) \leq \frac{1}{1 - \|\widehat{\Delta}\|_{\mathcal{H}_\infty}} J(\widehat{A}, \widehat{B}, \mathbf{K}). \tag{31}$$

Because $J(\widehat{A}, \widehat{B}, \mathbf{K})$ captures the performance of the controller \mathbf{K} on the nominal system $(\widehat{A}, \widehat{B})$, it is not subject to any uncertainty. It therefore remains to compute a tractable bound for $\|\widehat{\Delta}\|_{\mathcal{H}_\infty}$, which we do using the following fact.

Proposition 4 For any $\alpha \in (0, 1)$ and $\widehat{\Delta}$ as defined in (29)

$$\|\widehat{\Delta}\|_{\mathcal{H}_\infty} \leq \left\| \begin{bmatrix} \frac{\epsilon_A}{\sqrt{\alpha}} \Phi_x \\ \frac{\epsilon_B}{\sqrt{1-\alpha}} \Phi_u \end{bmatrix} \right\|_{\mathcal{H}_\infty} =: H_\alpha(\Phi_x, \Phi_u). \tag{32}$$

Proof Note that for any block matrix of the form $[M_1 \ M_2]$, we have

$$\|[M_1 \ M_2]\|_2 \leq \left(\|M_1\|_2^2 + \|M_2\|_2^2 \right)^{1/2}. \tag{33}$$

To verify this assertion, note that

$$\begin{aligned} \|[M_1 \ M_2]\|_2^2 &= \lambda_{\max}(M_1 M_1^* + M_2 M_2^*) \\ &\leq \lambda_{\max}(M_1 M_1^*) + \lambda_{\max}(M_2 M_2^*) = \|M_1\|_2^2 + \|M_2\|_2^2. \end{aligned}$$

With (33) in hand, we have

$$\begin{aligned} \left\| \begin{bmatrix} \Delta_A & \Delta_B \end{bmatrix} \begin{bmatrix} \Phi_x \\ \Phi_u \end{bmatrix} \right\|_{\mathcal{H}_\infty} &= \left\| \begin{bmatrix} \frac{\sqrt{\alpha}}{\epsilon_A} \Delta_A & \frac{\sqrt{1-\alpha}}{\epsilon_B} \Delta_B \end{bmatrix} \begin{bmatrix} \frac{\epsilon_A}{\sqrt{\alpha}} \Phi_x \\ \frac{\epsilon_B}{\sqrt{1-\alpha}} \Phi_u \end{bmatrix} \right\|_{\mathcal{H}_\infty} \\ &\leq \left\| \begin{bmatrix} \frac{\sqrt{\alpha}}{\epsilon_A} \Delta_A & \frac{\sqrt{1-\alpha}}{\epsilon_B} \Delta_B \end{bmatrix} \right\|_2 \left\| \begin{bmatrix} \frac{\epsilon_A}{\sqrt{\alpha}} \Phi_x \\ \frac{\epsilon_B}{\sqrt{1-\alpha}} \Phi_u \end{bmatrix} \right\|_{\mathcal{H}_\infty} \\ &\leq \left\| \begin{bmatrix} \frac{\epsilon_A}{\sqrt{\alpha}} \Phi_x \\ \frac{\epsilon_B}{\sqrt{1-\alpha}} \Phi_u \end{bmatrix} \right\|_{\mathcal{H}_\infty}, \end{aligned}$$

completing the proof. □

The following corollary is then immediate.

Corollary 2 *Let the controller \mathbf{K} and resulting system response (Φ_x, Φ_u) be as defined in Lemma 4. Then if $H_\alpha(\Phi_x, \Phi_u) < 1$, the controller $\mathbf{K} = \Phi_u \Phi_x^{-1}$ stabilizes the true system (A, B) .*

Applying Proposition 4 in conjunction with bound (31), we arrive at the following upper bound to the cost function of the robust LQR problem (7), which is independent of the perturbations (Δ_A, Δ_B) :

$$\sup_{\substack{\|\Delta_A\|_2 \leq \epsilon_A \\ \|\Delta_B\|_2 \leq \epsilon_B}} J(A, B, \mathbf{K}) \leq \left\| \begin{bmatrix} Q^{\frac{1}{2}} & 0 \\ 0 & R^{\frac{1}{2}} \end{bmatrix} \begin{bmatrix} \Phi_x \\ \Phi_u \end{bmatrix} \right\|_{\mathcal{H}_2} \frac{1}{1 - H_\alpha(\Phi_x, \Phi_u)} = \frac{J(\widehat{A}, \widehat{B}, \mathbf{K})}{1 - H_\alpha(\Phi_x, \Phi_u)}. \tag{34}$$

The upper bound is only valid when $H_\alpha(\Phi_x, \Phi_u) < 1$, which guarantees the stability of the closed-loop system as in Corollary 2. We remark that Corollary 2 and the bound in (34) are of interest independent of the synthesis procedure for \mathbf{K} . In particular, they can be applied to the optimal LQR controller \widehat{K} computed using the nominal system $(\widehat{A}, \widehat{B})$.

As the next lemma shows, the right-hand side of Eq. (34) can be efficiently optimized by an appropriate decomposition. The proof of the lemma is immediate.

Lemma 5 *For functions $f : \mathcal{X} \rightarrow \mathbb{R}$ and $g : \mathcal{X} \rightarrow \mathbb{R}$ and constraint set $C \subseteq \mathcal{X}$, consider*

$$\min_{x \in C} \frac{f(x)}{1 - g(x)}.$$

Assuming that $f(x) \geq 0$ and $0 \leq g(x) < 1$ for all $x \in C$, this optimization problem can be reformulated as an outer single-variable problem and an inner-constrained

optimization problem (the objective value of an optimization over the empty set is defined to be infinity):

$$\min_{x \in C} \frac{f(x)}{1 - g(x)} = \min_{\gamma \in [0,1]} \frac{1}{1-\gamma} \min_{x \in C} \{f(x) \mid g(x) \leq \gamma\}$$

Then combining Lemma 5 with the upper bound in (34) results in the following optimization problem:

$$\begin{aligned} &\text{minimize}_{\gamma \in [0,1]} \frac{1}{1-\gamma} \min_{\Phi_x, \Phi_u} \left\| \begin{bmatrix} Q^{\frac{1}{2}} & 0 \\ 0 & R^{\frac{1}{2}} \end{bmatrix} \begin{bmatrix} \Phi_x \\ \Phi_u \end{bmatrix} \right\|_{\mathcal{H}_2} \\ &\text{subject to} \quad [zI - \widehat{A} - \widehat{B}] \begin{bmatrix} \Phi_x \\ \Phi_u \end{bmatrix} = I, \quad \left\| \begin{bmatrix} \frac{\epsilon_A}{\sqrt{\alpha}} \Phi_x \\ \frac{\epsilon_B}{\sqrt{1-\alpha}} \Phi_u \end{bmatrix} \right\|_{\mathcal{H}_\infty} \leq \gamma \quad (35) \\ &\quad \Phi_x, \Phi_u \in \frac{1}{z} \mathcal{RH}_\infty. \end{aligned}$$

We note that this optimization objective is jointly quasi-convex in (γ, Φ_x, Φ_u) . Hence, as a function of γ alone the objective is quasi-convex and furthermore is smooth in the feasible domain. Therefore, the outer optimization with respect to γ can effectively be solved with methods like golden section search. We remark that the inner optimization is a convex problem, though an infinite-dimensional one. We show in Sect. 5 that a simple finite impulse response truncation yields a finite-dimensional problem with similar guarantees of robustness and performance.

We further remark that because $\gamma \in [0, 1)$, any feasible solution (Φ_x, Φ_u) to optimization problem (35) generates a controller $\mathbf{K} = \Phi_u \Phi_x^{-1}$ satisfying the conditions of Corollary 2 and hence stabilizes the true system (A, B) . Therefore, even if the solution is approximated, as long as it is feasible, it will be stabilizing. As we show in the next section, for sufficiently small estimation error bounds ϵ_A and ϵ_B , we can further bound the sub-optimality of the performance achieved by our robustly stabilizing controller relative to that achieved by the optimal LQR controller K_\star .

4 Sub-optimality Guarantees

We now return to analyzing the Coarse-ID control problem. We upper bound the performance of the controller synthesized using optimization (35) in terms of the size of the perturbations (Δ_A, Δ_B) and a measure of complexity of the LQR problem defined by A, B, Q , and R . The following result is one of our main contributions.

Theorem 3 *Let J_\star denote the minimal LQR cost achievable by any controller for the dynamical system with transition matrices (A, B) , and let K_\star denote the optimal controller. Let $(\widehat{A}, \widehat{B})$ be estimates of the transition matrices such that $\|\Delta_A\|_2 \leq \epsilon_A$, $\|\Delta_B\|_2 \leq \epsilon_B$. Then, if \mathbf{K} is synthesized via (35) with $\alpha = 1/2$, the relative error in the LQR cost is*

$$\frac{J(A, B, \mathbf{K}) - J_\star}{J_\star} \leq 5(\epsilon_A + \epsilon_B \|K_\star\|_2) \|\mathfrak{R}_{A+BK_\star}\|_{\mathcal{H}_\infty}, \tag{36}$$

as long as $(\epsilon_A + \epsilon_B \|K_\star\|_2) \|\mathfrak{R}_{A+BK_\star}\|_{\mathcal{H}_\infty} \leq 1/5$.

This result offers a guarantee on the performance of the SLS synthesized controller regardless of the estimation procedure used to estimate the transition matrices. Together with our result (Proposition 1) on system identification from independent data, Theorem 3 yields a sample complexity upper bound on the performance of the robust SLS controller \mathbf{K} when (A, B) are not known. We make this guarantee precise in Corollary 3. The rest of the section is dedicated to proving Theorem 3.

Recall that K_\star is the optimal LQR static state-feedback matrix for the true dynamics (A, B) , and let $\mathbf{\Delta} := -[\Delta_A + \Delta_B K_\star] \mathfrak{R}_{A+BK_\star}$. We begin with a technical result.

Lemma 6 Define $\zeta := (\epsilon_A + \epsilon_B \|K_\star\|_2) \|\mathfrak{R}_{A+BK_\star}\|_{\mathcal{H}_\infty}$, and suppose that $\zeta < (1 + \sqrt{2})^{-1}$. Then $(\gamma_0, \tilde{\Phi}_x, \tilde{\Phi}_u)$ is a feasible solution of (35) with $\alpha = 1/2$, where

$$\gamma_0 = \frac{\sqrt{2}\zeta}{1 - \zeta}, \quad \tilde{\Phi}_x = \mathfrak{R}_{A+BK_\star}(I + \mathbf{\Delta})^{-1}, \quad \tilde{\Phi}_u = K_\star \mathfrak{R}_{A+BK_\star}(I + \mathbf{\Delta})^{-1}. \tag{37}$$

Proof By construction $\tilde{\Phi}_x, \tilde{\Phi}_u \in \frac{1}{2} \mathcal{RH}_\infty$. Therefore, we are left to check three conditions:

$$\gamma_0 < 1, \quad [zI - \hat{A} - \hat{B}] \begin{bmatrix} \tilde{\Phi}_x \\ \tilde{\Phi}_u \end{bmatrix} = I, \text{ and } \left\| \begin{bmatrix} \frac{\epsilon_A}{\sqrt{\alpha}} \tilde{\Phi}_x \\ \frac{\epsilon_B}{\sqrt{1-\alpha}} \tilde{\Phi}_u \end{bmatrix} \right\|_{\mathcal{H}_\infty} \leq \frac{\sqrt{2}\zeta}{1 - \zeta}. \tag{38}$$

The first two conditions follow by simple algebraic computations. Before we check the last condition, note that $\|\mathbf{\Delta}\|_{\mathcal{H}_\infty} \leq (\epsilon_A + \epsilon_B \|K_\star\|_2) \|\mathfrak{R}_{A+BK_\star}\|_{\mathcal{H}_\infty} = \zeta < 1$. Now observe that,

$$\begin{aligned} \left\| \begin{bmatrix} \frac{\epsilon_A}{\sqrt{\alpha}} \tilde{\Phi}_x \\ \frac{\epsilon_B}{\sqrt{1-\alpha}} \tilde{\Phi}_u \end{bmatrix} \right\|_{\mathcal{H}_\infty} &= \sqrt{2} \left\| \begin{bmatrix} \epsilon_A \mathfrak{R}_{A+BK_\star} \\ \epsilon_B K_\star \mathfrak{R}_{A+BK_\star} \end{bmatrix} (I + \mathbf{\Delta})^{-1} \right\|_{\mathcal{H}_\infty} \\ &\leq \sqrt{2} \|(I + \mathbf{\Delta})^{-1}\|_{\mathcal{H}_\infty} \left\| \begin{bmatrix} \epsilon_A \mathfrak{R}_{A+BK_\star} \\ \epsilon_B K_\star \mathfrak{R}_{A+BK_\star} \end{bmatrix} \right\|_{\mathcal{H}_\infty} \\ &\leq \frac{\sqrt{2}}{1 - \|\mathbf{\Delta}\|_{\mathcal{H}_\infty}} \left\| \begin{bmatrix} \epsilon_A I \\ \epsilon_B K_\star \end{bmatrix} \mathfrak{R}_{A+BK_\star} \right\|_{\mathcal{H}_\infty} \\ &\leq \frac{\sqrt{2}(\epsilon_A + \epsilon_B \|K_\star\|_2) \|\mathfrak{R}_{A+BK_\star}\|_{\mathcal{H}_\infty}}{1 - \|\mathbf{\Delta}\|_{\mathcal{H}_\infty}} \leq \frac{\sqrt{2}\zeta}{1 - \zeta}. \end{aligned}$$

□

Proof of Theorem 3 Let $(\gamma_\star, \Phi_x^\star, \Phi_u^\star)$ be an optimal solution to problem (35) and let $\mathbf{K} = \Phi_u^\star(\Phi_x^\star)^{-1}$. We can then write

$$J(A, B, \mathbf{K}) \leq \frac{1}{1 - \|\hat{\mathbf{\Delta}}\|_{\mathcal{H}_\infty}} J(\hat{A}, \hat{B}, \mathbf{K}) \leq \frac{1}{1 - \gamma_\star} J(\hat{A}, \hat{B}, \mathbf{K}),$$

where the first inequality follows from bound (31), and the second follows from the fact that $\|\hat{\Delta}\|_{\mathcal{H}_\infty} \leq \gamma_\star$ due to Proposition 4 and the constraint in optimization problem (35).

From Lemma 6, we know that $(\gamma_0, \tilde{\Phi}_x, \tilde{\Phi}_u)$ defined in Eq. (37) is also a feasible solution. Therefore, because $K_\star = \tilde{\Phi}_u \tilde{\Phi}_x^{-1}$, we have by optimality,

$$\begin{aligned} \frac{1}{1 - \gamma_\star} J(\hat{A}, \hat{B}, \mathbf{K}) &\leq \frac{1}{1 - \gamma_0} J(\hat{A}, \hat{B}, K_\star) \leq \frac{J(A, B, K_\star)}{(1 - \gamma_0)(1 - \|\Delta\|_{\mathcal{H}_\infty})} \\ &= \frac{J_\star}{(1 - \gamma_0)(1 - \|\Delta\|_{\mathcal{H}_\infty})}, \end{aligned}$$

where the second inequality follows by the argument used to derive (31) with the true and estimated transition matrices switched. Recall that $\|\Delta\|_{\mathcal{H}_\infty} \leq \zeta$ and that $\gamma_0 = \sqrt{2}\zeta/(1 + \zeta)$. Therefore,

$$\frac{J(A, B, \mathbf{K}) - J_\star}{J_\star} \leq \frac{1}{1 - (1 + \sqrt{2})\zeta} - 1 = \frac{(1 + \sqrt{2})\zeta}{1 - (1 + \sqrt{2})\zeta} \leq 5\zeta,$$

where the last inequality follows because $\zeta < 1/5 < 1/(2 + 2\sqrt{2})$. The conclusion follows. □

With this sub-optimality result in hand, we are now ready to give an end-to-end performance guarantee for our procedure when the independent data estimation scheme is used.

Corollary 3 *Let $\lambda_G = \lambda_{\min}(\sigma_u^2 G_T G_T^* + \sigma_w^2 F_T F_T^*)$, where F_T, G_T are defined in (4). Suppose the independent data estimation procedure described in Algorithm 1 is used to produce estimates (\hat{A}, \hat{B}) and \mathbf{K} is synthesized via (35) with $\alpha = 1/2$. Then, there are universal constants C_0 and C_1 such that the relative error in the LQR cost satisfies*

$$\frac{J(A, B, \mathbf{K}) - J_\star}{J_\star} \leq C_0 \sigma_w \|\mathfrak{R}_{A+BK_\star}\|_{\mathcal{H}_\infty} \left(\frac{1}{\sqrt{\lambda_G}} + \frac{\|K_\star\|_2}{\sigma_u} \right) \sqrt{\frac{(n + p) \log(1/\delta)}{N}} \tag{39}$$

with probability $1 - \delta$ if $N \geq C_1(n + p)\sigma_w^2 \|\mathfrak{R}_{A+BK_\star}\|_{\mathcal{H}_\infty}^2 (1/\lambda_G + \|K_\star\|_2^2/\sigma_u^2) \log(1/\delta)$.

Proof Recall from Proposition 1 that for the independent data estimation scheme, we have

$$\epsilon_A \leq \frac{16\sigma_w}{\sqrt{\lambda_G}} \sqrt{\frac{(n + 2p) \log(32/\delta)}{N}}, \text{ and } \epsilon_B \leq \frac{16\sigma_w}{\sigma_u} \sqrt{\frac{(n + 2p) \log(32/\delta)}{N}}, \tag{40}$$

with probability $1 - \delta$, as long as $N \geq 8(n + p) + 16 \log(4/\delta)$.

To apply Theorem 3, we need $(\epsilon_A + \epsilon_B \|K_\star\|_2) \|\mathfrak{R}_{A+BK_\star}\|_{\mathcal{H}_\infty} < 1/5$, which will hold as long as $N \geq \mathcal{O}\left\{ (n + p)\sigma_w^2 \|\mathfrak{R}_{A+BK_\star}\|_{\mathcal{H}_\infty}^2 (1/\lambda_G + \|K_\star\|_2^2/\sigma_u^2) \log(1/\delta) \right\}$. A direct plug in of (40) in (36) yields the conclusion. □

This result fully specifies the complexity term \mathcal{C}_{LQR} promised in the introduction:

$$\mathcal{C}_{\text{LQR}} := C_0 \sigma_w \left(\frac{1}{\sqrt{\lambda_G}} + \frac{\|K_\star\|_2}{\sigma_u} \right) \|\mathfrak{R}_{A+BK_\star}\|_{\mathcal{H}_\infty}.$$

Note that \mathcal{C}_{LQR} decreases as the minimum eigenvalue of the sum of the input and noise controllability Gramians increases. This minimum eigenvalue tends to be larger for systems that amplify inputs in all directions of the state space. \mathcal{C}_{LQR} increases as function of the operator norm of the gain matrix K_\star and the \mathcal{H}_∞ norm of the transfer function from disturbance to state of the closed-loop system. These two terms tend to be larger for systems that are “harder to control.” The dependence on Q and R is implicit in this definition since the optimal control matrix K_\star is defined in terms of these two matrices. Note that when R is large in comparison with Q , the norm of the controller K_\star tends to be smaller because large inputs are more costly. However, such a change in the size of the controller could cause an increase in the \mathcal{H}_∞ norm of the closed-loop system. Thus, our upper bound suggests an odd balance. Stable and highly damped systems are easy to control but hard to estimate, whereas unstable systems are easy to estimate but hard to control. Our theorem suggests that achieving a small relative LQR cost requires for the system to be somewhere in the middle of these two extremes.

Finally, we remark that the above analysis holds more generally when we apply additional constraints to the controller in the synthesis problem (35). In this case, the sub-optimality bounds presented in Theorem 3 and Corollary 3 are true with respect to the minimal cost achievable by the constrained controller with access to the true dynamics. In particular, the bounds hold unchanged if the search is restricted to static controllers, i.e., $u_t = Kx_t$. This is true because the optimal controller is static and therefore feasible for the constrained synthesis problem.

5 Computation

As posed, the main optimization problem (35) is a semi-infinite program, and we are not aware of a way to solve this problem efficiently. In this section, we describe two alternative formulations that provide upper bounds to the optimal value and that can be solved in polynomial time.

5.1 Finite Impulse Response Approximation

An elementary approach to reducing the aforementioned semi-infinite program to a finite-dimensional one is to only optimize over the first L elements of the transfer functions Φ_x and Φ_u , effectively taking a finite impulse response (FIR) approximation. Since these are both stable maps, we expect the effects of such an approximation to be negligible as long as the optimization horizon L is chosen to be sufficiently large—in what follows, we show that this is indeed the case.

By restricting our optimization to FIR approximations of Φ_x and Φ_u , we can cast the \mathcal{H}_2 cost as a second-order cone constraint. The only difficulty arises in posing the \mathcal{H}_∞ constraint as a semidefinite program. Though there are several ways to cast \mathcal{H}_∞

constraints as linear matrix inequalities, we use the formulation in Theorem 5.8 of Dumitrescu’s text to take advantage of the FIR structure in our problem [17]. We note that using Dumitrescu’s formulation, the resulting problem is affine in α when γ is fixed, and hence, we can solve for the optimal value of α . Then the resulting system response elements can be cast as a dynamic feedback controller using Theorem 2 of [4].

5.1.1 Sub-optimality Guarantees

In this subsection, we show that optimizing over FIR approximations incurs only a small degradation in performance relative to the solution to the infinite horizon problem. In particular, this degradation in performance decays exponentially in the FIR horizon L , where the rate of decay is specified by the decay rate of the spectral elements of the optimal closed-loop system response $\mathfrak{R}_{A+BK_\star}$.

Before proceeding, we introduce additional concepts and notation needed to formalize guarantees in the FIR setting. A linear-time-invariant transfer function is stable if and only if it is exponentially stable, i.e., $\Phi = \sum_{t=0}^\infty z^{-t} \Phi(t) \in \mathcal{RH}_\infty$ if and only if there exists positive values C and $\rho \in [0, 1)$ such that for every spectral element $\Phi(t)$, $t \geq 0$, it holds that

$$\|\Phi(t)\|_2 \leq C\rho^t. \tag{41}$$

In what follows, we pick C_\star and ρ_\star to be any such constants satisfying $\|\mathfrak{R}_{A+BK_\star}(t)\|_2 \leq C_\star\rho_\star^t$ for all $t \geq 0$.

We introduce a version of the optimization problem (30) with a finite number of decision variables:

$$\begin{aligned} &\text{minimize}_{\gamma \in (0,1)} \frac{1}{1-\gamma} \min_{\Phi_x, \Phi_u, V} \left\| \begin{bmatrix} Q^{\frac{1}{2}} & 0 \\ 0 & R^{\frac{1}{2}} \end{bmatrix} \begin{bmatrix} \Phi_x \\ \Phi_u \end{bmatrix} \right\|_{\mathcal{H}_2} \\ &\text{subject to} \quad [zI - \widehat{A} - \widehat{B}] \begin{bmatrix} \Phi_x \\ \Phi_u \end{bmatrix} = I + \frac{1}{z^L} V, \\ &\quad \left\| \begin{bmatrix} \frac{\epsilon_A}{\sqrt{\alpha}} \Phi_x \\ \frac{\epsilon_B}{\sqrt{1-\alpha}} \Phi_u \end{bmatrix} \right\|_{\mathcal{H}_\infty} + \|V\|_2 \leq \gamma \\ &\quad \Phi_x = \sum_{t=1}^L \frac{1}{z^t} \Phi_x(t), \quad \Phi_u = \sum_{t=1}^L \frac{1}{z^t} \Phi_u(t). \end{aligned} \tag{42}$$

In this optimization problem, we search over finite response transfer functions Φ_x and Φ_u . Given a feasible solution Φ_x, Φ_u of problem (42), we can implement the controller $\mathbf{K}_L = \Phi_u \Phi_x^{-1}$ with an equivalent state-space representation (A_K, B_K, C_K, D_K) using the response elements $\{\Phi_x(k)\}_{k=1}^L$ and $\{\Phi_u(k)\}_{k=1}^L$ via Theorem 2 of [4].

The slack term V accounts for the error introduced by truncating the infinite response transfer functions of problem (30). Intuitively, if the truncated tail is sufficiently small, then the effects of this approximation should be negligible on performance. The next result formalizes this intuition.

Theorem 4 Set $\alpha = 1/2$ in (42) and choose two real values $C_\star > 0$ and $\rho_\star \in [0, 1)$ such that $\|\mathfrak{R}_{(A+BK_\star)}(t)\|_2 \leq C_\star \rho_\star^t$ for all $t \geq 0$. Then, if \mathbf{K}_L is synthesized via (42), the relative error in the LQR cost is

$$\frac{J(A, B, \mathbf{K}_L) - J_\star}{J_\star} \leq 10(\epsilon_A + \epsilon_B \|K_\star\|_2) \|\mathfrak{R}_{A+BK_\star}\|_{\mathcal{H}_\infty},$$

as long as

$$\epsilon_A + \epsilon_B \|K_\star\|_2 \leq \frac{1 - \rho_\star}{10C_\star} \quad \text{and} \quad L \geq \frac{4 \log \left(\frac{C_\star}{(\epsilon_A + \epsilon_B \|K_\star\|_2) \|\mathfrak{R}_{A+BK_\star}\|_{\mathcal{H}_\infty}} \right)}{1 - \rho_\star}.$$

The proof of this result, deferred to Appendix D, is conceptually the same as that of the infinite horizon setting. The main difference is that care must be taken to ensure that the approximation horizon L is sufficiently large so as to ensure stability and performance of the resulting controller. From the theorem statement, we see that for such an appropriately chosen FIR approximation horizon L , our performance bound is the same, up to universal constants, to that achieved by the solution to the infinite horizon problem. Furthermore, the approximation horizon L only needs to grow logarithmically with respect to one over the estimation rate in order to preserve the same statistical rate as the controller produced by the infinite horizon problem. Finally, an end-to-end sample-complexity result analogous to that stated in Corollary 3 can be easily obtained by simply substituting in the sample-complexity bounds on ϵ_A and ϵ_B specified in Proposition 1.

5.2 Static Controller and a Common Lyapunov Approximation

As we have reiterated above, when the dynamics are known, the optimal LQR control law takes the form $u_t = Kx_t$ for properly chosen static gain matrix K . We can reparameterize the optimization problem (35) to restrict our attention to such static control policies:

$$\begin{aligned} & \text{minimize}_{\gamma \in [0, 1)} \frac{1}{1 - \gamma} \min_{\Phi_x, \Phi_u, K} \left\| \begin{bmatrix} Q^{\frac{1}{2}} & 0 \\ 0 & R^{\frac{1}{2}} \end{bmatrix} \begin{bmatrix} \Phi_x \\ \Phi_u \end{bmatrix} \right\|_{\mathcal{H}_2} \\ & \text{subject to} \quad [zI - \widehat{A} - \widehat{B}] \begin{bmatrix} \Phi_x \\ \Phi_u \end{bmatrix} = I, \quad \left\| \begin{bmatrix} \frac{\epsilon_A}{\sqrt{\alpha}} \Phi_x \\ \frac{\epsilon_B}{\sqrt{1-\alpha}} \Phi_u \end{bmatrix} \right\|_{\mathcal{H}_\infty} \leq \gamma \quad (43) \\ & \quad \Phi_x, \Phi_u \in \frac{1}{z} \mathcal{RH}_\infty, \quad K = \Phi_u \Phi_x^{-1}. \end{aligned}$$

Under this reparameterization, the problem is no longer convex. Here we present a simple application of the *common Lyapunov relaxation* that allows us to find a controller K using semidefinite programming.

Note that the equality constraints imply:

$$I = [zI - \widehat{A} - \widehat{B}] \begin{bmatrix} \Phi_x \\ \Phi_u \end{bmatrix} = [zI - \widehat{A} - \widehat{B}] \begin{bmatrix} I \\ K \end{bmatrix} \Phi_x = (zI - \widehat{A} - \widehat{B}K) \Phi_x,$$

revealing that we must have

$$\Phi_x = (zI - \widehat{A} - \widehat{B}K)^{-1} \text{ and } \Phi_u = K(zI - \widehat{A} - \widehat{B}K)^{-1}.$$

With these identifications, (43) can be reformulated as

$$\begin{aligned} &\text{minimize}_{\gamma \in (0,1)} \frac{1}{1-\gamma} \min_K \left\| \begin{bmatrix} Q^{\frac{1}{2}} & 0 \\ 0 & R^{\frac{1}{2}}K \end{bmatrix} (zI - \widehat{A} - \widehat{B}K)^{-1} \right\|_{\mathcal{H}_2} \\ &\text{subject to} \quad \left\| \begin{bmatrix} \frac{\epsilon_A}{\sqrt{\alpha}} \\ \frac{\epsilon_B}{\sqrt{1-\alpha}}K \end{bmatrix} (zI - \widehat{A} - \widehat{B}K)^{-1} \right\|_{\mathcal{H}_\infty} \leq \gamma \end{aligned} \tag{44}$$

Using standard techniques from the robust control literature, we can upper bound this problem via the semidefinite program

$$\begin{aligned} &\text{minimize}_{X,Z,W,\alpha,\gamma} \frac{1}{(1-\gamma)^2} \{ \text{Trace}(QW_{11}) + \text{Trace}(RW_{22}) \} \\ &\text{subject to} \quad \begin{bmatrix} X & X & Z^* \\ X & W_{11} & W_{12} \\ Z & W_{21} & W_{22} \end{bmatrix} \geq 0 \\ &\quad \begin{bmatrix} X - I & (\widehat{A} + \widehat{B}K)X & 0 & 0 \\ X(\widehat{A} + \widehat{B}K)^* & X & \epsilon_A X & \epsilon_B Z^* \\ 0 & \epsilon_A X & \alpha\gamma^2 I & 0 \\ 0 & \epsilon_B Z & 0 & (1-\alpha)\gamma^2 I \end{bmatrix} \geq 0. \end{aligned} \tag{45}$$

Note that this optimization problem is affine in α when γ is fixed. Hence, in practice we can find the optimal value of α as well. A static controller can then be extracted from this optimization problem by setting $K = ZX^{-1}$. A full derivation of this relaxation can be found in Appendix E. Note that this compact SDP is simpler to solve than the truncated FIR approximation. As demonstrated experimentally in the following section, the cost of this simplification is that the common Lyapunov approach provides a controller with slightly higher LQR cost.

6 Numerical Experiments

We illustrate our results on estimation, controller synthesis, and LQR performance with numerical experiments of the end-to-end Coarse-ID control scheme. The least squares estimation procedure (9) is carried out on a simulated system in Python, and the bootstrapped error estimates are computed in parallel using PyWren [32].

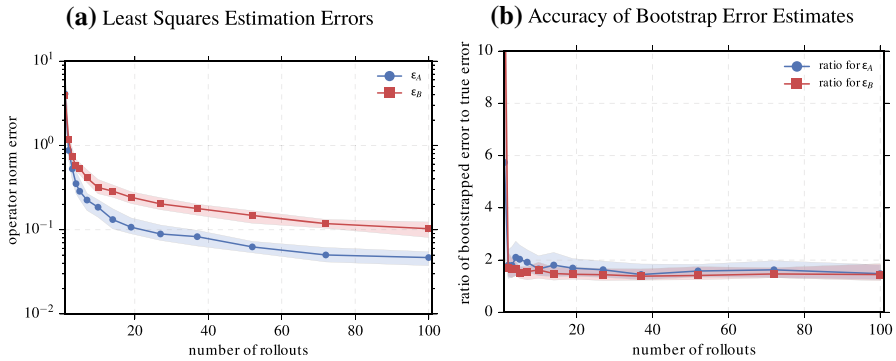


Fig. 1 The resulting errors from 100 repeated least squares identification experiments with rollout length $T = 6$ is plotted against the number of rollouts. In **(a)**, the median of the least squares estimation errors decreases with N . In **(b)**, the ratio of the bootstrap estimates to the true estimates hovers at 2. Shaded regions display quartiles

All of the synthesis and performance experiments are run in MATLAB. We make use of the YALMIP package for prototyping convex optimization [37] and use the MOSEK solver under an academic license [5]. In particular, when using the FIR approximation described in Sect. 5.1, we find it effective to make use of YALMIP’s `dualize` function, which considerably reduces the computation time.

6.1 Estimation of Example System

We focus experiments on a particular example system. Consider the LQR problem instance specified by

$$A = \begin{bmatrix} 1.01 & 0.01 & 0 \\ 0.01 & 1.01 & 0.01 \\ 0 & 0.01 & 1.01 \end{bmatrix}, \quad B = I, \quad Q = 10^{-3}I, \quad R = I. \quad (46)$$

The dynamics correspond to a marginally unstable graph Laplacian system where adjacent nodes are weakly connected, each node receives direct input, and input size is penalized relatively more than state. Dynamics described by graph Laplacians arise naturally in consensus and distributed averaging problems. For this system, we perform the full data identification procedure in (9), using inputs with variance $\sigma_u^2 = 1$ and noise with variance $\sigma_w^2 = 1$. The errors are estimated via the bootstrap (Algorithm 2) using $M = 2000$ trials and confidence parameter $\delta = 0.05$.

The behavior of the least squares estimates and the bootstrap error estimates is illustrated in Fig. 1. The rollout length is fixed to $T = 6$, and the number of rollouts used in the estimation is varied. As expected, increasing the number of rollouts corresponds to decreasing errors. For large enough N , the bootstrapped error estimates are of the same order of magnitude as the true errors. In Appendix G, we show plots for the setting in which the number of rollouts is fixed to $N = 6$, while the rollout length is varied.

6.2 Controller Synthesis on Estimated System

Using the estimates of the system in (46), we synthesize controllers using two robust control schemes: the convex problem in 42 with filters of length $L = 32$ and V set to 0, and the common Lyapunov (CL) relaxation of the static synthesis problem (43). Once the FIR responses $\{\Phi_x(k)\}_{k=1}^F$ and $\{\Phi_u(k)\}_{k=1}^F$ are found, we need a way to implement the system responses as a controller. We represent the dynamic controller $\mathbf{K} = \Phi_u \Phi_x^{-1}$ by finding an equivalent state-space realization (A_K, B_K, C_K, D_K) via Theorem 2 of [4]. In what follows, we compare the performance of these controllers with the nominal LQR controller (the solution to (3) with \hat{A} and \hat{B} as model parameters), and explore the trade-off between robustness, complexity, and performance.

The relative performance of the nominal controller is compared with robustly synthesized controllers in Fig. 2. For both robust synthesis procedures, two controllers are compared: one using the true errors on A and B , and the other using the bootstrap estimates of the errors. The robust static controller generated via the common Lyapunov approximation performs slightly worse than the more complex FIR controller, but it still achieves reasonable control performance. Moreover, the conservative bootstrap estimates also result in worse control performance, but the degradation of performance is again modest.

Furthermore, the experiments show that the nominal controller often outperforms the robust controllers *when it is stabilizing*. On the other hand, the nominal controller is not guaranteed to stabilize the true system, and as shown in Fig. 2, it only does so in roughly 80 of the 100 instances after $N = 60$ rollouts. It is also important to note a distinction between stabilization for nominal and robust controllers. When the nominal controller is not stabilizing, there is no indication to the user (though sufficient conditions for stability can be checked using our result in Corollary 4 or structured singular value methods [48]). On the other hand, the robust synthesis procedure will return as infeasible, alerting the user by default that the uncertainties are too high. We observe similar results when we fix the number of trials but vary the rollout length. These figures are provided in Appendix G.

Figure 3 explores the trade-off between performance and complexity for the computational approximations, both for FIR truncation and for the common Lyapunov relaxation. We examine the trade-off in terms of both the bound on the LQR cost (given by the value of the objective) and the actual achieved value. It is interesting that for smaller numbers of rollouts (and therefore larger uncertainties), the benefit of using more complex FIR models is negligible, in terms of both the actual costs and the upper bound. This trend makes sense: as uncertainties decrease to zero, the best robust controller should approach the nominal controller, which is associated with infinite impulse response (IIR) transfer functions. Furthermore, for the experiments presented here, FIR length of $L = 32$ seems to be sufficient to characterize the performance of the robust synthesis procedure in (35). Additionally, we note that static controllers are able to achieve costs of a similar magnitude.

The SLS framework guarantees a stabilizing controller for the true system, provided that the computational approximations are feasible for *any* value of γ between 0 and 1, as long as the system errors (ϵ_A, ϵ_B) are upper bounds on the true errors. Figure 4

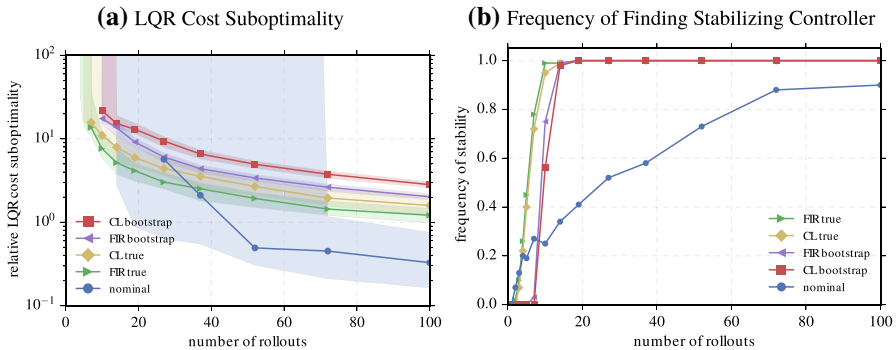


Fig. 2 Performance of controllers synthesized on the results of the 100 identification experiments is plotted against the number of rollouts. Controllers are synthesis nominally, using FIR truncation, and using the common Lyapunov (CL) relaxation. In (a), the median sub-optimality of nominal and robustly synthesized controllers is compared, with shaded regions displaying quartiles, which go off to infinity in the case that a stabilizing controller was not found. In (b), the frequency that the synthesis methods found stabilizing controllers

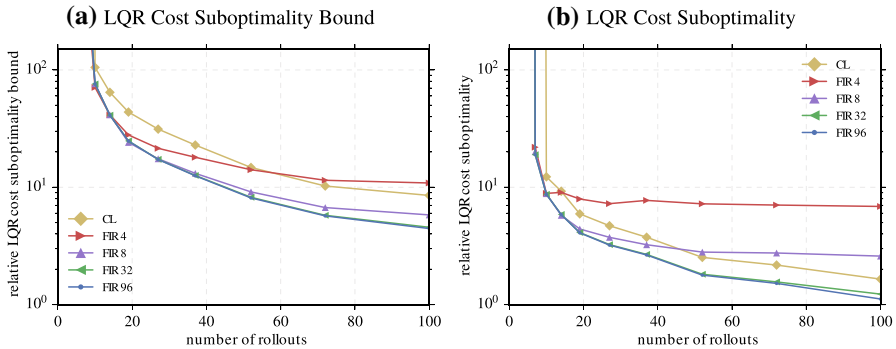


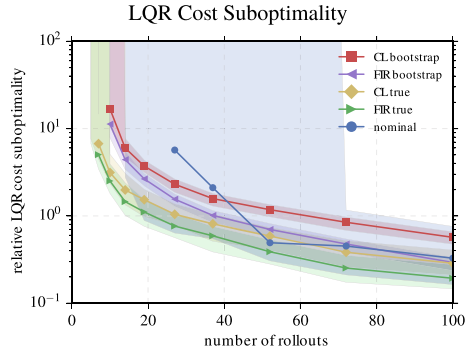
Fig. 3 Performance of controllers synthesized with varying FIR filter lengths on the results of 10 of the identification experiments using true errors. The median sub-optimality of robustly synthesized controllers does not appear to change for FIR lengths greater than 32, and the common Lyapunov (CL) synthesis tracks the performance in both upper bound and actual cost

displays the controller performance for robust synthesis when γ is set to 0.999. Simply ensuring a stable model and neglecting to optimize the nominal cost yields controllers that perform nearly an order of magnitude better than those where we search for the optimal value of γ . This observation aligns with common practice in robust control: constraints ensuring stability are only active when the cost tries to drive the system up against a safety limit. We cannot provide end-to-end sample-complexity guarantees for this method and leave such bounds as an enticing challenge for future work.

7 Conclusions and Future Work

Coarse-ID control provides a straightforward approach to merging nonasymptotic methods from system identification with contemporary Systems Level Synthesis

Fig. 4 Performance of controllers synthesized on the results of 100 identification experiments is plotted against the number of rollouts. The plot compares the median sub-optimality of nominal controllers with fixed- γ robustly synthesized controllers ($\gamma = 0.999$)



approaches to robust control. Indeed, many of the principles of Coarse-ID control were well established in the 1990s [11,12,29], but fusing together an end-to-end result required contemporary analysis of random matrices and a new perspective on controller synthesis. These results can be extended in a variety of directions, and we close this paper with a discussion of some of the shortcomings of our approach and of several possible applications of the Coarse-ID framework to other control settings.

Other Performance Metrics Though we focused exclusively on LQR in this paper, we note that all of our results on robust synthesis and end-to-end performance analysis extend to other metrics popular in control. Indeed, *any* norm on the system responses $\{\Phi_x(k), \Phi_u(k)\}$ can be solved robustly using our approach; Lemma 4 holds for any norm. In turn, we can mimic the derivation in Sect. 3 to yield a constrained optimization problem with respect to the nominal dynamics and a norm on the uncertainty $\hat{\Delta}$. This means that our sub-optimality bound in Corollary 3 holds true when we replace $\mathcal{H}_2(\mathbb{T})$ with $\mathcal{H}_\infty(\mathbb{T})$. Furthermore, similar results can be derived for other norms, so long as care is placed on the associated sub-multiplicative properties of the norms in question. For example, in follow-up work we analyze robustness under the \mathcal{L}_1 norm in the context of constraints on the states and control signals [14].

Improving the End-to-End Analysis There are several places where our analysis could be substantially improved. The most obvious is that in our estimator for the state-transition matrices, our algorithm only uses the final time step of each rollout. This strategy is data inefficient, and empirically, accuracy only improves when including all of the data. Analyzing the full least squares estimator is nontrivial because the design matrix strongly depends on data to be estimated. This poses a challenging problem in random matrix theory that has applications in a variety of control and reinforcement learning settings. In follow-up work, we have begun to address this issue for stable linear systems [52].

In the context of SLS, we use a very coarse characterization of the plant uncertainty to bound the quantity in Lemma 4 and to yield a tractable optimization problem. Indeed, the only property we use about the error between our nominal system and the true system is that the maps

$$x \mapsto (A - \hat{A})x \text{ and } u \mapsto (B - \hat{B})u$$

are contractions. Nowhere do we use the fact that these are linear operators, or even the fact that they are the same operator from time step to time step. Indeed, there are stronger bounds that could be engineered using the theory of integral quadratic constraints [40] that would take into account these additional properties. Such tighter bounds could yield considerably less conservative control schemes in both theory and practice.

Additionally, it would be of interest to understand the loss in performance incurred by the common Lyapunov relaxation we use in our experiments. Empirically, we see that the approximation leads to good performance, suggesting that it does not introduce much conservatism into the synthesis task. Further, our numerical experiments suggest that optimizing a nominal cost subject to robust stability constraints, as opposed to directly optimizing the SLS upper bound, leads to better empirical performance. Future work will seek to understand whether this is a phenomenological observation specific to the systems used in our experiments, or if there is a deeper principle at play that leads to tighter sub-optimality guarantees.

Lower Bounds Finding lower bounds for control problems when the model is unknown is an open question. Even for LQR, it is not at all clear how well the system (A, B) needs to be known in order to attain good control performance. While we produce reasonable worst-case upper bounds for this problem, we know of no lower bounds. Such bounds would offer a reasonable benchmark for how well one could ever expect to do with no priors on the linear system dynamics.

Integrating Coarse-ID Control in Other Control Paradigms The end-to-end Coarse-ID control framework should be applicable in a variety of settings. For example, in model predictive control (MPC), controller synthesis problems are approximately solved on finite-time horizons, one step is taken, and then this process is repeated [6]. MPC is an effective solution which substitutes fast optimization solvers for clever, complex control design. We believe it will be straightforward to extend the Coarse-ID paradigm to MPC, using a similar perturbation argument as in Sect. 3. The main challenges in MPC lie in how to guarantee that safety constraints are maintained throughout execution without too much conservatism in control costs.

Another interesting investigation lies in the area of adaptive control, where we could investigate how to incorporate new data into coarse models to further refine constraint sets and costs. Indeed, some work has already been done in this space. We propose to investigate how to operationalize and extend the notion of *optimistic exploration* proposed in the context of continuous control by Abbasi-Yadkori and Szepesvari [1]. The idea behind optimistic improvement is to select the model that would give the best optimization cost if the current model was true. In this way, we fail fast, either receiving a good cost or learning quickly that our model is incorrect. It would be worth investigating whether the Coarse-ID framework can make it simple to update a least squares estimate for the system parameters and then provide an efficient mechanism for choosing the next optimistic control.

Finally, Coarse-ID control could be relevant to nonlinear control applications. In nonlinear control, iterative LQR schemes are remarkably effective [35]. Hence, it would be interesting to understand how parametric model errors can be estimated and mitigated in a control loop that employs iterative LQR or similar dynamic programming methods.

Sample Complexities of Reinforcement Learning for Continuous Control Finally, we imagine that the analysis in this paper may be useful for understanding popular reinforcement learning algorithms that are also being tested for continuous control. Reinforcement learning directly attacks a control cost in question without resorting to any specific identification scheme. While this suffers from the drawback that generally speaking, no parameter convergence can be guaranteed, it is ideally suited to ignoring modes that do not affect control performance. For instance, it might not be important to get a good estimate of very stable modes or of lightly damped modes that do not substantially affect the performance.

There are two parallel problems here. First, it would be of interest to determine system identification algorithms that are tuned to particular control tasks. In the Coarse-ID control approach, the estimation and control are completely decoupled. However, it may be beneficial to inform the identification algorithm about the desired cost, resulting in improved sample complexity.

From a different perspective, policy gradient and Q-learning methods applied to LQR could yield important insights about the pros and cons of such methods. There are classic papers [9] on Q-learning for LQR, but these use asymptotic analysis. Recently, the first such analysis for policy gradient has appeared, though the precise scaling with respect to system parameters is not yet understood [20]. Providing clean nonasymptotic bounds here could help provide a rapprochement between machine learning and adaptive control, with optimization negotiating the truce.

Acknowledgements We thank Ross Boczar, Qingqing Huang, Laurent Lessard, Michael Littman, Manfred Morari, Andrew Packard, Anders Rantzer, Daniel Russo, and Ludwig Schmidt for many helpful comments and suggestions. We also thank the anonymous referees for making several suggestions that have significantly improved the paper and its presentation.

A Proof of Lemma 1

First, recall Bernstein’s lemma. Let X_1, \dots, X_p be zero-mean independent r.v.s satisfying the Orlicz norm bound $\|X_i\|_{\psi_1} \leq K$. Then as long as $p \geq 2 \log(1/\delta)$, with probability at least $1 - \delta$,

$$\sum_{i=1}^p X_i \leq K \sqrt{2n \log(1/\delta)} .$$

Next, let Q be an $m \times n$ matrix. Let $u_1, \dots, u_{M_\varepsilon}$ be a ε -net for the m -dimensional ℓ_2 ball, and similarly let $v_1, \dots, v_{N_\varepsilon}$ be a ε covering for the n -dimensional ℓ_2 ball. For each $\|u\|_2 = 1$ and $\|v\|_2 = 1$, let u_i, v_j denote the elements in the respective nets such that $\|u - u_i\|_2 \leq \varepsilon$ and $\|v - v_j\|_2 \leq \varepsilon$. Then,

$$\begin{aligned} u^* Q v &= (u - u_i + u_i)^* Q v = (u - u_i)^* Q v + u_i^* Q (v - v_j + v_j) \\ &= (u - u_i)^* Q v + u_i^* Q (v - v_j) + u_i^* Q v_j . \end{aligned}$$

Hence,

$$u^* Qv \leq 2\varepsilon \|Q\|_2 + u_i^* Qv_j \leq 2\varepsilon \|Q\|_2 + \max_{1 \leq i \leq M_\varepsilon, 1 \leq j \leq N_\varepsilon} u_i^* Qv_j .$$

Since u, v are arbitrary on the sphere,

$$\|Q\|_2 \leq \frac{1}{1 - 2\varepsilon} \max_{1 \leq i \leq M_\varepsilon, 1 \leq j \leq N_\varepsilon} u_i^* Qv_j .$$

Now we study the problem at hand. Choose $\varepsilon = 1/4$. By a standard volume comparison argument, we have that $M_\varepsilon \leq 9^m$ and $N_\varepsilon \leq 9^n$, and that

$$\left\| \sum_{k=1}^N f_k g_k^* \right\|_2 \leq 2 \max_{1 \leq i \leq M_\varepsilon, 1 \leq j \leq N_\varepsilon} \sum_{k=1}^N (u_i^* f_k)(g_k^* v_j) .$$

Note that $u_i^* f_k \sim N(0, u_i^* \Sigma_f u_i)$ and $g_k^* v_j \sim N(0, v_j^* \Sigma_g v_j)$. By independence of f_k and g_k , $(u_i^* f_k)(g_k^* v_j)$ is a zero mean sub-exponential random variable, and therefore, $\|(u_i^* f_k)(g_k^* v_j)\|_{\psi_1} \leq \sqrt{2} \|\Sigma_f\|_2^{1/2} \|\Sigma_g\|_2^{1/2}$. Hence, for each pair u_i, v_j we have with probability at least $1 - \delta/9^{m+n}$,

$$\sum_{k=1}^N (u_i^* f_k)(g_k^* v_j) \leq 2 \|\Sigma_f\|_2^{1/2} \|\Sigma_g\|_2^{1/2} \sqrt{N(m+n) \log(9/\delta)} .$$

Taking a union bound over all pairs in the ε -net yields the claim.

B Proof of Proposition 3

For this proof, we need a lemma similar to Lemma 1. The following is a standard result in high-dimensional statistics [58], and we state it here without proof.

Lemma 7 *Let $W \in \mathbb{R}^{N \times n}$ be a matrix with each entry i.i.d. $\mathcal{N}(0, \sigma_w^2)$. Then, with probability $1 - \delta$, we have*

$$\|W\|_2 \leq \sigma_w (\sqrt{N} + \sqrt{n} + \sqrt{2 \log(1/\delta)}) .$$

As before we use Z to denote the $N \times (n + p)$ matrix with rows equal to $z_\ell^\top = [(x^{(\ell)})^\top \ (u^{(\ell)})^\top]$. Also, we denote by W the $N \times n$ matrix with columns equal to $w^{(\ell)}$. Therefore, the error matrix for the ordinary least squares estimator satisfies

$$E = \begin{bmatrix} (\widehat{A} - A)^\top \\ (\widehat{B} - B)^\top \end{bmatrix} = (Z^\top Z)^{-1} Z^\top W ,$$

when the matrix Z has rank $n + p$. Under the assumption that $N \geq n + p$ we consider the singular value decomposition $Z = U \Lambda V^\top$, where $V, \Lambda \in \mathbb{R}^{(n+p) \times (n+p)}$ and $U \in \mathbb{R}^{N \times (n+p)}$. Therefore, when Λ is invertible,

$$E = V(\Lambda^\top \Lambda)^{-1} \Lambda^\top U^\top W = V \Lambda^{-1} U^\top W .$$

This implies that

$$EE^T = V\Lambda^{-1}U^TWW^T U\Lambda^{-1}V^T \leq \|U^T W\|_2^2 V\Lambda^{-2}V^T = \|U^T W\|_2^2 (Z^T Z)^{-1}.$$

Since the columns of U are orthonormal, it follows that the entries of $U^T W$ are i.i.d. $\mathcal{N}(0, \sigma_w^2)$. Hence, the conclusion follows by Lemma 7.

C Derivation of the LQR Cost as an \mathcal{H}_2 Norm

In this section, we consider the transfer function description of the infinite horizon LQR optimal control problem. In particular, we show how it can be recast as an equivalent \mathcal{H}_2 optimal control problem in terms of the system response variables defined in Theorem 1.

Recall that stable and achievable system responses (Φ_x, Φ_u) , as characterized in Eq. (23), describe the closed-loop map from disturbance signal \mathbf{w} to the state and control action (\mathbf{x}, \mathbf{u}) achieved by the controller $\mathbf{K} = \Phi_u \Phi_x^{-1}$, i.e.,

$$\begin{bmatrix} \mathbf{x} \\ \mathbf{u} \end{bmatrix} = \begin{bmatrix} \Phi_x \\ \Phi_u \end{bmatrix} \mathbf{w}.$$

Letting $\Phi_x = \sum_{t=1}^{\infty} \Phi_x(t)z^{-t}$ and $\Phi_u = \sum_{t=1}^{\infty} \Phi_u(t)z^{-t}$, we can then equivalently write for any $t \geq 1$

$$\begin{bmatrix} x_t \\ u_t \end{bmatrix} = \sum_{k=1}^t \begin{bmatrix} \Phi_x(k) \\ \Phi_u(k) \end{bmatrix} w_{t-k}. \tag{47}$$

For a disturbance process distributed as $w_t \stackrel{\text{i.i.d.}}{\sim} \mathcal{N}(0, \sigma_w^2 I_n)$, it follows from Eq. (47) that

$$\begin{aligned} \mathbb{E} [x_t^* Q x_t] &= \sigma_w^2 \sum_{k=1}^t \text{Tr}(\Phi_x(k)^* Q \Phi_x(k)), \\ \mathbb{E} [u_t^* R u_t] &= \sigma_w^2 \sum_{k=1}^t \text{Tr}(\Phi_u(k)^* R \Phi_u(k)). \end{aligned}$$

We can then write

$$\begin{aligned} \lim_{T \rightarrow \infty} \frac{1}{T} \sum_{t=1}^T \mathbb{E} [x_t^* Q x_t + u_t^* R u_t] &= \sigma_w^2 \left[\sum_{t=1}^{\infty} \text{Tr}(\Phi_x(t)^* Q \Phi_x(t)) + \text{Tr}(\Phi_u(t)^* R \Phi_u(t)) \right] \\ &= \sigma_w^2 \sum_{t=1}^{\infty} \left\| \begin{bmatrix} Q^{\frac{1}{2}} & 0 \\ 0 & R^{\frac{1}{2}} \end{bmatrix} \begin{bmatrix} \Phi_x(t) \\ \Phi_u(t) \end{bmatrix} \right\|_F^2 \\ &= \frac{\sigma_w^2}{2\pi} \int_{\mathbb{T}} \left\| \begin{bmatrix} Q^{\frac{1}{2}} & 0 \\ 0 & R^{\frac{1}{2}} \end{bmatrix} \begin{bmatrix} \Phi_x \\ \Phi_u \end{bmatrix} \right\|_F^2 dz \end{aligned}$$

$$= \sigma_w^2 \left\| \begin{bmatrix} Q^{\frac{1}{2}} & 0 \\ 0 & R^{\frac{1}{2}} \end{bmatrix} \begin{bmatrix} \Phi_x \\ \Phi_u \end{bmatrix} \right\|_{\mathcal{H}_2}^2,$$

where the second to last equality is due to Parseval’s theorem.

D Proof of Theorem 4

To understand the effect of restricting the optimization to FIR transfer functions, we need to understand the decay of the transfer functions $\mathfrak{R}_{\widehat{A} + \widehat{B}K_\star}$ and $K_\star \mathfrak{R}_{\widehat{A} + \widehat{B}K_\star}$. To this end, we consider $C_\star > 0$ and $\rho_\star \in (0, 1)$ such that $\|(A + BK_\star)^t\|_2 \leq C_\star \rho_\star^t$ for all $t \geq 0$. Such C_\star and ρ_\star exist because K_\star stabilizes the system (A, B) . The next lemma quantifies how well K_\star stabilizes the system $(\widehat{A}, \widehat{B})$ when the estimation error is small.

Lemma 8 *Suppose $\epsilon_A + \epsilon_B \|K_\star\|_2 \leq \frac{1 - \rho_\star}{2C_\star}$. Then,*

$$\|(\widehat{A} + \widehat{B}K_\star)^t\|_2 \leq C_\star \left(\frac{1 + \rho_\star}{2} \right)^t, \text{ for all } t \geq 0.$$

Proof The claim is obvious when $t = 0$. Fix an integer $t \geq 1$ and denote $M = A + BK_\star$. Then, if $\Delta = \Delta_A + \Delta_B K_\star$, we have $\widehat{A} + \widehat{B}K_\star = M + \Delta$.

Consider the expansion of $(M + \Delta)^t$ into 2^t terms. Label all these terms as $T_{i,j}$ for $i = 0, \dots, t$ and $j = 1, \dots, \binom{t}{i}$ where i denotes the degree of Δ in the term. Since Δ has degree i in $T_{i,j}$, the term $T_{i,j}$ has the form $M^{\alpha_1} \Delta M^{\alpha_2} \Delta \dots \Delta M^{\alpha_{i+1}}$, where each α_k is a nonnegative integer and $\sum_k \alpha_k = t - i$. Then, using the fact that $\|M^k\|_2 \leq C_\star \rho_\star^k$ for all $k \geq 0$, we have $\|T_{i,j}\|_2 \leq C^{i+1} \rho^{t-i} \|\Delta\|_2^i$. Hence by triangle inequality:

$$\begin{aligned} \|(M + \Delta)^t\|_2 &\leq \sum_{i=0}^t \sum_j \|T_{i,j}\|_2 \\ &\leq \sum_{i=0}^t \binom{t}{i} C_\star^{i+1} \rho_\star^{t-i} \|\Delta\|_2^i \\ &= C_\star \sum_{i=0}^t \binom{t}{i} (C_\star \|\Delta\|_2)^i \rho_\star^{t-i} \\ &= C_\star (C_\star \|\Delta\|_2 + \rho_\star)^t \\ &\leq C_\star \left(\frac{1 + \rho_\star}{2} \right)^t, \end{aligned}$$

where the last inequality uses the fact $\|\Delta\|_2 \leq \epsilon_A + \epsilon_B \|K_\star\|_2 \leq \frac{1 - \rho_\star}{2C_\star}$. □

For the remainder of this discussion, we use the following notation to denote the restriction of a system response to its first L time steps:

$$\Phi_x(1 : L) = \sum_{t=1}^L \frac{1}{z^t} \Phi_x(t), \quad \Phi_u(1 : L) = \sum_{t=1}^L \frac{1}{z^t} \Phi_u(t). \tag{48}$$

To prove Theorem 4, we must relate the optimal controller K_\star with the optimal solution of the optimization problem (42). In the next lemma, we use K_\star to construct a feasible solution for problem (42). As before, we denote $\zeta = (\epsilon_A + \epsilon_B \|K_\star\|_2) \|\mathfrak{R}_{\widehat{A} + \widehat{B}K_\star}\|_{\mathcal{H}_\infty}$.

Lemma 9 *Set $\alpha = 1/2$ in problem (42), and assume that $\epsilon_A + \epsilon_B \|K_\star\|_2 \leq \frac{1-\rho_\star}{2C_\star}$, $\zeta < 1/5$, and*

$$L \geq \frac{4 \log\left(\frac{C_\star}{\zeta}\right)}{1 - \rho_\star}. \tag{49}$$

Then, optimization problem (42) is feasible, and the following is one such feasible solution:

$$\begin{aligned} \tilde{\Phi}_x &= \mathfrak{R}_{\widehat{A} + \widehat{B}K_\star}(1 : L), \quad \tilde{\Phi}_u = K_\star \mathfrak{R}_{\widehat{A} + \widehat{B}K_\star}(1 : L), \\ \tilde{V} &= -\mathfrak{R}_{\widehat{A} + \widehat{B}K_\star}(L + 1), \quad \tilde{\gamma} = \frac{4\zeta}{1 - \zeta}. \end{aligned} \tag{50}$$

Proof From Lemma 8 and the assumption on ζ , we have that $\|(\widehat{A} + \widehat{B}K_\star)^t\|_2 \leq C_\star \left(\frac{1+\rho_\star}{2}\right)^t$ for all $t \geq 0$. In particular, since $\mathfrak{R}_{\widehat{A} + \widehat{B}K_\star}(L + 1) = (\widehat{A} + \widehat{B}K_\star)^L$, we have $\|\tilde{V}\| = \|(\widehat{A} + \widehat{B}K_\star)^L\| \leq C_\star \left(\frac{1+\rho_\star}{2}\right)^L \leq \zeta$. The last inequality is true because we assumed L is sufficiently large.

Once again, since $\mathfrak{R}_{\widehat{A} + \widehat{B}K_\star}(L + 1) = (\widehat{A} + \widehat{B}K_\star)^L$, it can be easily seen that our choice of $\tilde{\Phi}_x$, $\tilde{\Phi}_u$, and \tilde{V} satisfy the linear constraint of problem (42). It remains to prove that

$$\sqrt{2} \left\| \begin{bmatrix} \epsilon_A \tilde{\Phi}_x \\ \epsilon_B \tilde{\Phi}_u \end{bmatrix} \right\|_{\mathcal{H}_\infty} + \|\tilde{V}\|_2 \leq \tilde{\gamma} < 1.$$

The second inequality holds because of our assumption on ζ . We already know that $\|\tilde{V}\|_2 \leq \zeta$. Now, we bound:

$$\begin{aligned} \left\| \begin{bmatrix} \epsilon_A \tilde{\Phi}_x \\ \epsilon_B \tilde{\Phi}_u \end{bmatrix} \right\|_{\mathcal{H}_\infty} &\leq (\epsilon_A + \epsilon_B \|K_\star\|_2) \|\mathfrak{R}_{\widehat{A} + \widehat{B}K_\star}(1 : L)\|_{\mathcal{H}_\infty} \\ &\leq (\epsilon_A + \epsilon_B \|K_\star\|_2) (\|\mathfrak{R}_{\widehat{A} + \widehat{B}K_\star}\|_{\mathcal{H}_\infty} + \|\mathfrak{R}_{\widehat{A} + \widehat{B}K_\star}(L + 1 : \infty)\|_{\mathcal{H}_\infty}). \end{aligned}$$

These inequalities follow from the definition of $(\tilde{\Phi}_x, \tilde{\Phi}_u)$ and the triangle inequality.

We recall that $\mathfrak{R}_{\widehat{A} + \widehat{B}K_\star} = \mathfrak{R}_{A+BK_\star} (I + \mathbf{\Delta})^{-1}$, where $\mathbf{\Delta} = -(\Delta_A + \Delta_B K_\star) \mathfrak{R}_{A+BK_\star}$. Then, since $\|\mathbf{\Delta}\|_{\mathcal{H}_\infty} \leq \zeta$ (due to Proposition 4), we have $\|\mathfrak{R}_{\widehat{A} + \widehat{B}K_\star}\|_{\mathcal{H}_\infty} \leq \frac{1}{1-\zeta} \|\mathfrak{R}_{A+BK_\star}\|_{\mathcal{H}_\infty}$.

We can upper bound

$$\begin{aligned} \|\mathfrak{R}_{\hat{A}+\hat{B}K_\star}(L+1:\infty)\|_{\mathcal{H}_\infty} &\leq \sum_{t=L+1}^\infty \|\mathfrak{R}_{\hat{A}+\hat{B}K_\star}(t)\|_2 \\ &\leq C_\star \left(\frac{1+\rho_\star}{2}\right)^L \sum_{t=0}^\infty \left(\frac{1+\rho_\star}{2}\right)^t = \frac{2C_\star}{1-\rho_\star} \left(\frac{1+\rho_\star}{2}\right)^L. \end{aligned}$$

Then, since we assumed that ϵ_A and ϵ_B are sufficiently small and that L is sufficiently large, we obtain

$$(\epsilon_A + \epsilon_B \|K_\star\|_2) \|\mathfrak{R}_{\hat{A}+\hat{B}K_\star}(L+1:\infty)\|_{\mathcal{H}_\infty} \leq \zeta.$$

Therefore,

$$\left\| \begin{bmatrix} \epsilon_A \tilde{\Phi}_x \\ \epsilon_B \tilde{\Phi}_u \end{bmatrix} \right\|_{\mathcal{H}_\infty} \leq \frac{\zeta}{1-\zeta} + \zeta \leq \frac{2\zeta}{1-\zeta}.$$

The conclusion follows. □

Proof of Theorem 4 As all of the assumptions of Lemma 9 are satisfied, optimization problem (42) is feasible. We denote $(\Phi_x^\star, \Phi_u^\star, V_\star, \gamma_\star)$ the optimal solution of problem (42). We denote

$$\hat{\Delta} := \Delta_A \Phi_x^\star + \Delta_B \Phi_u^\star + \frac{1}{zL} V_\star.$$

Then, we have

$$[zI - A - B] \begin{bmatrix} \Phi_x^\star \\ \Phi_u^\star \end{bmatrix} = I + \hat{\Delta}.$$

Applying the triangle inequality, and leveraging Proposition 4, we can verify that

$$\|\hat{\Delta}\|_{\mathcal{H}_\infty} \leq \sqrt{2} \left\| \begin{bmatrix} \epsilon_A \Phi_x^\star \\ \epsilon_B \Phi_u^\star \end{bmatrix} \right\|_{\mathcal{H}_\infty} + \|V_\star\|_2 \leq \gamma_\star < 1,$$

where the last two inequalities are true because the optimal solution is a feasible point of the optimization problem (42).

We now apply Lemma 4 to characterize the response achieved by the FIR approximate controller K_L on the true system (A, B) :

$$\begin{aligned} J(A, B, K_L) &= \left\| \begin{bmatrix} Q^{\frac{1}{2}} & 0 \\ 0 & R^{\frac{1}{2}} \end{bmatrix} \begin{bmatrix} \Phi_x^\star \\ \Phi_u^\star \end{bmatrix} (I + \hat{\Delta})^{-1} \right\|_{\mathcal{H}_2} \\ &\leq \frac{1}{1-\gamma_\star} \left\| \begin{bmatrix} Q^{\frac{1}{2}} & 0 \\ 0 & R^{\frac{1}{2}} \end{bmatrix} \begin{bmatrix} \Phi_x^\star \\ \Phi_u^\star \end{bmatrix} \right\|_{\mathcal{H}_2}. \end{aligned}$$

Denote by $(\tilde{\Phi}_x, \tilde{\Phi}_u, \tilde{V}, \tilde{\gamma})$ the feasible solution constructed in Lemma 9, and let $J_L(\hat{A}, \hat{B}, K_*)$ denote the truncation of the LQR cost achieved by controller K_* on system (\hat{A}, \hat{B}) to its first L time steps.

Then,

$$\begin{aligned} \frac{1}{1 - \gamma_*} \left\| \begin{bmatrix} Q^{\frac{1}{2}} & 0 \\ 0 & R^{\frac{1}{2}} \end{bmatrix} \begin{bmatrix} \Phi_x^* \\ \Phi_u^* \end{bmatrix} \right\|_{\mathcal{H}_2} &\leq \frac{1}{1 - \tilde{\gamma}} \left\| \begin{bmatrix} Q^{\frac{1}{2}} & 0 \\ 0 & R^{\frac{1}{2}} \end{bmatrix} \begin{bmatrix} \tilde{\Phi}_x \\ \tilde{\Phi}_u \end{bmatrix} \right\|_{\mathcal{H}_2} \\ &= \frac{1}{1 - \tilde{\gamma}} J_L(\hat{A}, \hat{B}, K_*) \\ &\leq \frac{1}{1 - \tilde{\gamma}} J(\hat{A}, \hat{B}, K_*) \\ &\leq \frac{1}{1 - \tilde{\gamma}} \frac{1}{1 - \|\Delta\|_{\mathcal{H}_\infty}} J_*, \end{aligned}$$

where $\Delta = -(\Delta_A + \Delta_B K_*)\mathfrak{R}_{A+BK_*}$. The first inequality follows from the optimality of $(\Phi_x^*, \Phi_u^*, V_*, \gamma_*)$, the equality and second inequality from the fact that $(\tilde{\Phi}_x, \tilde{\Phi}_u)$ are truncations of the response of K_* on (\hat{A}, \hat{B}) to the first L time steps, and the final inequality by following similar arguments to the proof of Theorem 3, and in applying Theorem 2.

Noting that

$$\|\Delta\|_{\mathcal{H}_\infty} = \|(\Delta_A + \Delta_B K_*)\mathfrak{R}_{A+BK_*}\|_{\mathcal{H}_\infty} \leq \zeta < 1,$$

we then have that

$$J(A, B, \mathbf{K}_L) \leq \frac{1}{1 - \tilde{\gamma}} \frac{1}{1 - \zeta} J_*,$$

Recalling that $\tilde{\gamma} = \frac{4\zeta}{1 - \zeta}$, we obtain

$$\frac{J(A, B, \mathbf{K}_L) - J_*}{J_*} \leq \frac{1 - \zeta}{1 - 5\zeta} \frac{1}{1 - \zeta} - 1 = \frac{5\zeta}{(1 - 5\zeta)} \leq 10\zeta,$$

where the last equality is true when $\zeta \leq 1/10$. The conclusion follows. □

E A Common Lyapunov Relaxation for Proportional Control

We unpack each of the norms in (44) as linear matrix inequalities. First, by the KYP Lemma, the \mathcal{H}_∞ constraint is satisfied if and only if there exists a matrix P_∞ satisfying

$$\begin{aligned} &\begin{bmatrix} (\hat{A} + \hat{B}K)^* P_\infty (\hat{A} + \hat{B}K) - P_\infty & (\hat{A} + \hat{B}K)^* P_\infty \\ P_\infty (\hat{A} + \hat{B}K) & P_\infty \end{bmatrix} \\ &+ \begin{bmatrix} \gamma^{-2} \begin{bmatrix} \frac{\epsilon_A}{\sqrt{\alpha}} \\ \frac{\epsilon_B}{\sqrt{1-\alpha}} K \end{bmatrix}^* \begin{bmatrix} \frac{\epsilon_A}{\sqrt{\alpha}} \\ \frac{\epsilon_B}{\sqrt{1-\alpha}} K \end{bmatrix} & 0 \\ 0 & -I \end{bmatrix} \leq 0. \end{aligned}$$

Applying the Schur complement Lemma, we can reformulate this as the equivalent matrix inequality

$$\begin{bmatrix} -P_\infty^{-1} & 0 & 0 & (\widehat{A} + \widehat{B}K) & I \\ 0 & -\gamma^2 I & 0 & \frac{\epsilon_A}{\sqrt{\alpha}} I & 0 \\ 0 & 0 & -\gamma^2 I & \frac{\epsilon_B}{\sqrt{1-\alpha}} K & 0 \\ (\widehat{A} + \widehat{B}K)^* & \frac{\epsilon_A}{\sqrt{\alpha}} I & \frac{\epsilon_B}{\sqrt{1-\alpha}} K^* & -P_\infty & 0 \\ I & 0 & 0 & 0 & -I \end{bmatrix} \leq 0.$$

Then, conjugating by the matrix $\text{diag}(I, I, P_\infty^{-1}, I)$ and setting $X_\infty = P_\infty^{-1}$, we are left with

$$\begin{bmatrix} -X_\infty & 0 & 0 & (\widehat{A} + \widehat{B}K)X_\infty & I \\ 0 & -\gamma^2 I & 0 & \frac{\epsilon_A}{\sqrt{\alpha}} X_\infty & 0 \\ 0 & 0 & -\gamma^2 I & \frac{\epsilon_B}{\sqrt{1-\alpha}} K X_\infty & 0 \\ X_\infty(\widehat{A} + \widehat{B}K)^* & \frac{\epsilon_A}{\sqrt{\alpha}} X_\infty & \frac{\epsilon_B}{\sqrt{1-\alpha}} X_\infty K^* & -X_\infty & 0 \\ I & 0 & 0 & 0 & -I \end{bmatrix} \leq 0.$$

Finally, applying the Schur complement lemma again gives the more compact inequality

$$\begin{bmatrix} -X_\infty + I & 0 & 0 & (\widehat{A} + \widehat{B}K)X_\infty \\ 0 & -\gamma^2 I & 0 & \frac{\epsilon_A}{\sqrt{\alpha}} X_\infty \\ 0 & 0 & -\gamma^2 I & \frac{\epsilon_B}{\sqrt{1-\alpha}} K X_\infty \\ X_\infty(\widehat{A} + \widehat{B}K)^* & \frac{\epsilon_A}{\sqrt{\alpha}} X_\infty & \frac{\epsilon_B}{\sqrt{1-\alpha}} X_\infty K^* & -X_\infty \end{bmatrix} \leq 0.$$

For convenience, we permute the rows of this inequality and conjugate by $\text{diag}(I, I, \sqrt{\alpha}I, \sqrt{1-\alpha}I)$ and use the equivalent form

$$\begin{bmatrix} -X_\infty + I & (\widehat{A} + \widehat{B}K)X_\infty & 0 & 0 \\ X_\infty(\widehat{A} + \widehat{B}K)^* & -X_\infty & \epsilon_A X_\infty & \epsilon_B X_\infty K^* \\ 0 & \epsilon_A X_\infty & -\alpha\gamma^2 I & 0 \\ 0 & \epsilon_B K X_\infty & 0 & -(1-\alpha)\gamma^2 I \end{bmatrix} \leq 0.$$

For the \mathcal{H}_2 norm, we have that under proportional control K , the average cost is given by $\text{Trace}((Q + K^*RK)X_2)$ where X_2 is the steady-state covariance. That is, X_2 satisfies the Lyapunov equation

$$X_2 = (\widehat{A} + \widehat{B}K)X_2(\widehat{A} + \widehat{B}K)^* + I.$$

But note that we can relax this expression to a matrix inequality

$$X_2 \succeq (\widehat{A} + \widehat{B}K)X_2(\widehat{A} + \widehat{B}K)^* + I, \tag{51}$$

and $\text{Trace}((Q + K^*RK)X_2)$ will remain an upper bound on the squared \mathcal{H}_2 norm. Rewriting this matrix inequality with Schur complements and combining with our derivation for the \mathcal{H}_∞ norm, we can reformulate (44) as a nonconvex semidefinite program

$$\begin{aligned} & \underset{X_2, X_\infty, K, \gamma}{\text{minimize}} \frac{1}{(1-\gamma)^2} \text{Trace}((Q + K^*RK)X_2) \\ & \text{subject to} \begin{bmatrix} X_2 - I & (\widehat{A} + \widehat{B}K)X_2 \\ X_2(\widehat{A} + \widehat{B}K)^* & X_2 \end{bmatrix} \succeq 0 \\ & \begin{bmatrix} X_\infty - I & (\widehat{A} + \widehat{B}K)X_\infty & 0 & 0 \\ X_\infty(\widehat{A} + \widehat{B}K)^* & X_\infty & \epsilon_A X_\infty & \epsilon_B X_\infty K^* \\ 0 & \epsilon_A X_\infty & \alpha\gamma^2 I & 0 \\ 0 & \epsilon_B K X_\infty & 0 & (1-\alpha)\gamma^2 I \end{bmatrix} \succeq 0. \end{aligned} \tag{52}$$

The common Lyapunov relaxation simply imposes that $X_2 = X_\infty$. Under this identification, we note that the first LMI becomes redundant and we are left with the SDP

$$\begin{aligned} & \underset{X, K, \gamma}{\text{minimize}} \frac{1}{(1-\gamma)^2} \text{Trace}((Q + K^*RK)X) \\ & \text{subject to} \begin{bmatrix} X - I & (\widehat{A} + \widehat{B}K)X & 0 & 0 \\ X(\widehat{A} + \widehat{B}K)^* & X & \epsilon_A X & \epsilon_B X K^* \\ 0 & \epsilon_A X & \alpha\gamma^2 I & 0 \\ 0 & \epsilon_B K X & 0 & (1-\alpha)\gamma^2 I \end{bmatrix} \succeq 0. \end{aligned}$$

Now though this appears to be nonconvex, we can perform the standard variable substitution $Z = KX$ and rewrite the cost to yield (45).

F Numerical Bootstrap Validation

We evaluate the efficacy of the bootstrap procedure introduced in Algorithm 2. Recall that even though we provide theoretical bounds in Proposition 1, for practical purposes and for handling dependent data, we want bounds that are the least conservative possible.

For given state dimension n , input dimension p , and scalar ρ , we generate upper triangular matrices $A \in \mathbb{R}^{n \times n}$ with all diagonal entries equal to ρ and the upper triangular entries i.i.d. samples from $\mathcal{N}(0, 1)$, clipped at magnitude 1. By construction, matrices will have spectral radius ρ . The entries of $B \in \mathbb{R}^{n \times p}$ were sampled i.i.d. from $\mathcal{N}(0, 1)$, clipped at magnitude 1. The variance terms σ_u^2 and σ_w^2 were fixed to be 1.

Recall from Sect. 2.3 that M represents the number of trials used for the bootstrap estimation, and $\widehat{\epsilon}_A, \widehat{\epsilon}_B$ are the bootstrap estimates for ϵ_A, ϵ_B . To check the validity of the bootstrap procedure, we empirically estimate the fraction of time A and B lie in the balls $B_{\widehat{\epsilon}_A}(\widehat{\epsilon}_A)$ and $B_{\widehat{\epsilon}_B}(\widehat{\epsilon}_B)$, where $B_X(r) = \{X' : \|X' - X\|_2 \leq r\}$.

Our findings are summarized in Figs. 5 and 6. Although not plotted, the theoretical bounds found in Sect. 2 would be orders of magnitude larger than the true ϵ_A and ϵ_B , while the bootstrap bounds offer a good approximation.

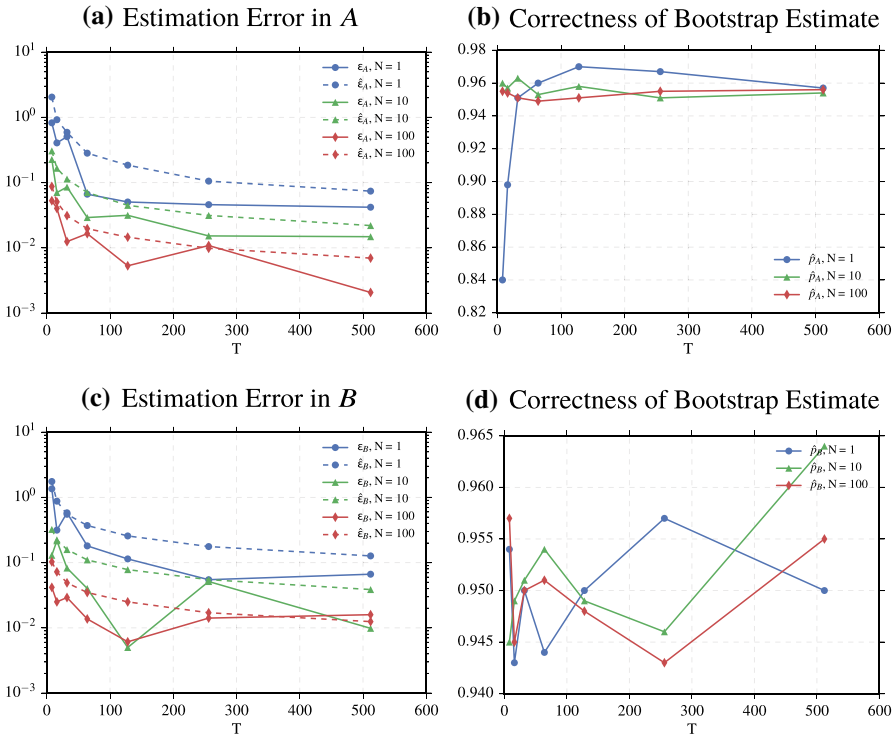


Fig. 5 In these simulations: $n = 3$, $p = 1$, $\rho = 0.9$, and $M = 2000$. In (a), the spectral distances to A (shown in the solid lines) are compared with the bootstrap estimates (shown in the dashed lines). In (b), the probability A lies in $B_{\hat{A}}(\hat{\epsilon}_A)$ estimated from 2000 trials. In (c), the spectral distances to B_* are compared with the bootstrap estimates. In (d), the probability B lies in $B_{\hat{B}}(\hat{\epsilon}_B)$ estimated from 2000 trials

G Experiments with Varying Rollout Lengths

Here we include results of experiments in which we fix the number of trials ($N = 6$) and vary the rollout length. Figure 7 displays the estimation errors. The estimation errors on A decrease more quickly than in the fixed rollout length case, consistent with the idea that longer rollouts of easily excitable systems allow for better identification due to higher signal-to-noise ratio. Figure 8 shows that stabilizing performance of the nominal is somewhat better than in the fixed rollout length case (Fig. 2). This fact is likely related to the smaller errors on the estimation of A (Fig. 7).

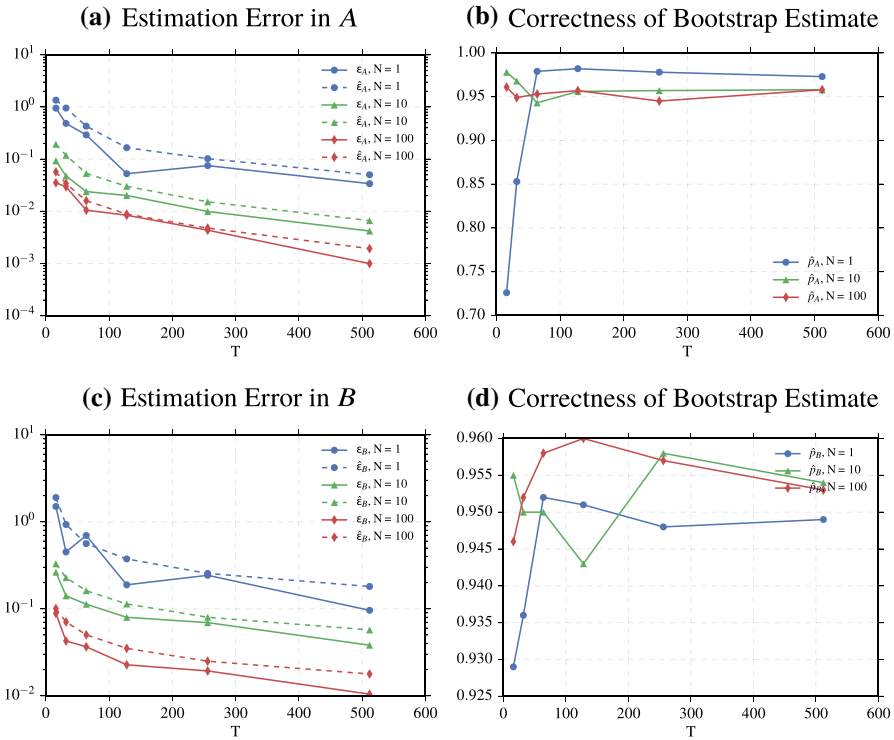


Fig. 6 In these simulations: $n = 6$, $p = 2$, $\rho = 1.01$, and $M = 2000$. In (a), the spectral distances to A are compared with the bootstrap estimates. In (b), the probability A lies in $B_{\hat{A}}(\hat{\epsilon}_A)$ estimated from 2000 trials. In (c), the spectral distances to B are compared with the bootstrap estimates. In (d), the probability B lies in $B_{\hat{B}}(\hat{\epsilon}_B)$ estimated from 2000 trials

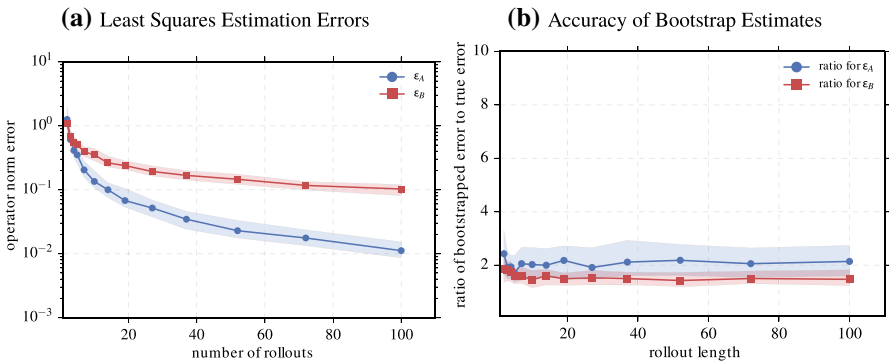


Fig. 7 The resulting errors from 100 identification experiments with a total of $N = 6$ rollouts are plotted against the length rollouts. In (a), the median of the least squares estimation errors decreases with T . In (b), the ratio of the bootstrap estimates to the true estimates display quartiles

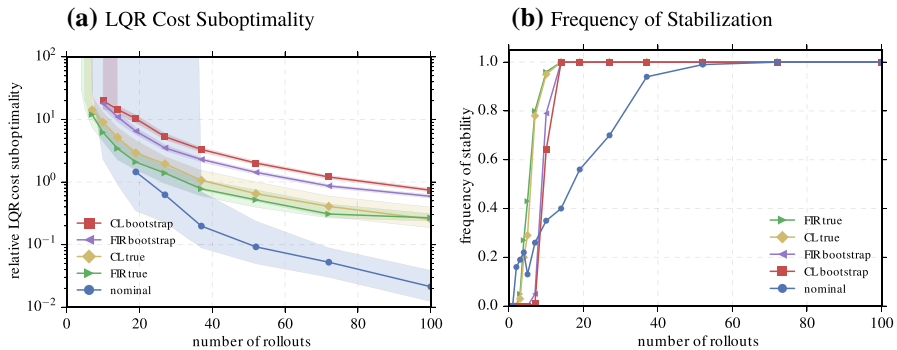


Fig. 8 Performance of controllers synthesized on the results of the 100 identification experiments is plotted against the length of rollouts. In **(a)**, the median sub-optimality of nominal and robustly synthesized controllers are compared, with shaded regions displaying quartiles, which go off to infinity when stabilizing controllers are not frequently found. In **(b)**, the frequency synthesis methods found stabilizing controllers

References

1. Y. Abbasi-Yadkori and C. Szepesvári. Regret Bounds for the Adaptive Control of Linear Quadratic Systems. In *Proceedings of the 24th Annual Conference on Learning Theory*, pages 1–26, 2011.
2. M. Abeille and A. Lazaric. Thompson sampling for linear-quadratic control problems. In *AISTATS 2017-20th International Conference on Artificial Intelligence and Statistics*, 2017.
3. M. Abeille and A. Lazaric. Improved regret bounds for thompson sampling in linear quadratic control problems. In *International Conference on Machine Learning*, pages 1–9, 2018.
4. J. Anderson and N. Matni. Structured state space realizations for sls distributed controllers. In *2017 55th Annual Allerton Conference on Communication, Control, and Computing (Allerton)*, pages 982–987. IEEE, 2017.
5. M. ApS. *The MOSEK optimization toolbox for MATLAB manual. Version 8.1 (Revision 25)*, 2015. URL <http://docs.mosek.com/8.1/toolbox/index.html>.
6. F. Borrelli, A. Bemporad, and M. Morari. *Predictive control for linear and hybrid systems*. Cambridge University Press, New York, NY, USA, 2017.
7. G. E. Box, G. M. Jenkins, G. C. Reinsel, and G. M. Ljung. *Time series analysis: forecasting and control*. John Wiley & Sons, 2015.
8. R. P. Braatz, P. M. Young, J. C. Doyle, and M. Morari. Computational complexity of μ calculation. *IEEE Transactions on Automatic Control*, 39 (5): 1000–1002, 1994.
9. S. J. Bradtke, B. E. Ydstie, and A. G. Barto. Adaptive linear quadratic control using policy iteration. In *Proceedings of 1994 American Control Conference-ACC'94*, volume 3, pages 3475–3479. IEEE, 1994.
10. M. C. Campi and E. Weyer. Finite sample properties of system identification methods. *IEEE Transactions on Automatic Control*, 47 (8): 1329–1334, 2002.
11. J. Chen and G. Gu. *Control-Oriented System Identification: An \mathcal{H}_∞ Approach*. Wiley, 2000.
12. J. Chen and C. N. Nett. The Caratheodory-Fejer problem and H_∞ identification: a time domain approach. In *Proceedings of 32nd IEEE Conference on Decision and Control*, pages 68–73. IEEE, 1993.
13. M. A. Dahleh and I. J. Diaz-Bobillo. *Control of uncertain systems: a linear programming approach*. Prentice-Hall, Inc., Upper Saddle River, NJ, USA, 1994.
14. S. Dean, S. Tu, N. Matni, and B. Recht. Safely Learning to Control the Constrained Linear Quadratic Regulator. *arXiv:1809.10121*, 2018.
15. J. Doyle. Analysis of feedback systems with structured uncertainties. *IEE Proceedings D - Control Theory and Applications*, 129(6), 1982. ISSN 0143-7054. <https://doi.org/10.1049/ip-d.1982.0053>.
16. Y. Duan, X. Chen, R. Houthoofd, J. Schulman, and P. Abbeel. Benchmarking deep reinforcement learning for continuous control. In *International Conference on Machine Learning*, pages 1329–1338, 2016.

17. B. Dumitrescu. *Positive trigonometric polynomials and signal processing applications*, volume 103. Springer Science & Business Media, 2007.
18. B. Efron. *Bootstrap Methods: Another Look at the Jackknife*, pages 569–593. Springer New York, New York, NY, 1992. https://doi.org/10.1007/978-1-4612-4380-9_41.
19. M. K. H. Fan, A. L. Tits, and J. C. Doyle. Robustness in the presence of mixed parametric uncertainty and unmodeled dynamics. *IEEE Transactions on Automatic Control*, 36(1), 1991. ISSN 0018-9286. <https://doi.org/10.1109/9.62265>.
20. M. Fazel, R. Ge, S. Kakade, and M. Mesbahi. Global convergence of policy gradient methods for the linear quadratic regulator. In *Proceedings of the 35th International Conference on Machine Learning*, volume 80, pages 1467–1476. PMLR, 10–15 Jul 2018.
21. E. Feron. Analysis of robust \mathcal{H}_2 performance using multiplier theory. *SIAM Journal on Control and Optimization*, 35 (1): 160–177, 1997. 10.1137/S0363012994266504.
22. C.-N. Fiechter. PAC adaptive control of linear systems. In *Proceedings of the Tenth Annual Conference on Computational Learning Theory, COLT '97*, pages 72–80, New York, NY, USA, 1997. ACM. <https://doi.org/10.1145/267460.267481>.
23. A. Goldenshluger. Nonparametric estimation of transfer functions: rates of convergence and adaptation. *IEEE Transactions on Information Theory*, 44 (2): 644–658, 1998. <https://doi.org/10.1109/18.661510>.
24. A. Goldenshluger and A. Zeevi. Nonasymptotic bounds for autoregressive time series modeling. *Ann. Statist.*, 29 (2): 417–444, 04 2001. 10.1214/aos/1009210547.
25. P. Hall. *The Bootstrap and Edgeworth Expansion*. Springer Science & Business Media, 2013.
26. M. Hardt, T. Ma, and B. Recht. Gradient descent learns linear dynamical systems. *The Journal of Machine Learning Research*, 19 (1): 1025–1068, 2018.
27. E. Hazan, K. Singh, and C. Zhang. Learning linear dynamical systems via spectral filtering. In *Advances in Neural Information Processing Systems*, pages 6702–6712, 2017.
28. E. Hazan, H. Lee, K. Singh, C. Zhang, and Y. Zhang. Spectral filtering for general linear dynamical systems. In *Advances in Neural Information Processing Systems*, pages 4634–4643, 2018.
29. A. J. Helmicki, C. A. Jacobson, and C. N. Nett. Control oriented system identification: a worst-case/deterministic approach in \mathcal{H}_∞ . *IEEE Transactions on Automatic Control*, 36 (10): 1163–1176, 1991. 10.1109/9.90229.
30. M. Ibrahim, A. Javanmard, and B. V. Roy. Efficient reinforcement learning for high dimensional linear quadratic systems. In F. Pereira, C. J. C. Burges, L. Bottou, and K. Q. Weinberger, editors, *Advances in Neural Information Processing Systems 25*, pages 2636–2644, 2012.
31. N. Jiang, A. Krishnamurthy, A. Agarwal, J. Langford, and R. E. Schapire. Contextual decision processes with low Bellman rank are PAC-learnable. In *Proceedings of the 34th International Conference on Machine Learning*, volume 70, pages 1704–1713. PMLR, 06–11 Aug 2017.
32. E. Jonas, Q. Pu, S. Venkataraman, I. Stoica, and B. Recht. Occupy the cloud: Distributed computing for the 99%. In *Proceedings of the 2017 Symposium on Cloud Computing*, pages 445–451. ACM, 2017. 10.1145/3127479.3128601.
33. V. Kuznetsov and M. Mohri. Generalization bounds for non-stationary mixing processes. *Machine Learning*, 106 (1): 93–117, 2017. <https://doi.org/10.1007/s10994-016-5588-2>.
34. S. Levine, C. Finn, T. Darrell, and P. Abbeel. End-to-end training of deep visuomotor policies. *Journal of Machine Learning Research*, 17 (1): 1334–1373, Jan. 2016.
35. W. Li and E. Todorov. Iterative Linear Quadratic Regulator Design for Nonlinear Biological Movement Systems. In *International Conference on Informatics in Control, Automation and Robotics*, 2004.
36. L. Ljung. *System Identification: Theory for the User*. Prentice-Hall, Inc., Upper Saddle River, NJ, USA, 1999.
37. J. Löfberg. YALMIP : A toolbox for modeling and optimization in MATLAB. In *IEEE International Symposium on Computer Aided Control System Design*, 2004.
38. N. Matni, Y. Wang, and J. Anderson. Scalable system level synthesis for virtually localizable systems. In *2017 IEEE 56th Annual Conference on Decision and Control (CDC)*, pages 3473–3480, Dec 2017. <https://doi.org/10.1109/CDC.2017.8264168>.
39. D. J. McDonald, C. R. Shalizi, and M. Schervish. Nonparametric risk bounds for time-series forecasting. *Journal of Machine Learning Research*, 18 (1): 1044–1083, Jan. 2017.
40. A. Megretski and A. Rantzer. System analysis via integral quadratic constraints. *IEEE Transactions on Automatic Control*, 42(6): 819–830, June 1997. <https://doi.org/10.1109/9.587335>.
41. V. Mnih, K. Kavukcuoglu, D. Silver, A. A. Rusu, J. Veness, M. G. Bellemare, A. Graves, M. Riedmiller, A. K. Fidjeland, G. Ostrovski, S. Petersen, C. Beattie, A. Sadik, I. Antonoglou, H. King, D. Kumaran,

- D. Wierstra, S. Legg, and D. Hassabis. Human-level control through deep reinforcement learning. *Nature*, 518: 529–533, 02 2015.
42. M. Mohri and A. Rostamizadeh. Stability bounds for stationary ϕ -mixing and β -mixing processes. *Journal of Machine Learning Research*, 11: 789–814, 2010.
 43. Y. Ouyang, M. Gagrani, and R. Jain. Control of unknown linear systems with thompson sampling. In *2017 55th Annual Allerton Conference on Communication, Control, and Computing (Allerton)*, pages 1198–1205, Oct 2017. <https://doi.org/10.1109/ALLERTON.2017.8262873>.
 44. A. Packard and J. Doyle. The complex structured singular value. *Automatica*, 29 (1): 71 – 109, 1993. [https://doi.org/10.1016/0005-1098\(93\)90175-S](https://doi.org/10.1016/0005-1098(93)90175-S).
 45. F. Paganini. Necessary and sufficient conditions for robust \mathcal{H}_2 performance. In *Proceedings of 1995 34th IEEE Conference on Decision and Control*, volume 2, pages 1970–1975 vol.2, Dec 1995. <https://doi.org/10.1109/CDC.1995.480635>.
 46. J. Pereira, M. Ibrahimi, and A. Montanari. Learning networks of stochastic differential equations. In *Advances in Neural Information Processing Systems*, pages 172–180, 2010.
 47. R. Postoyan, L. Buşoniu, D. Nešić, and J. Daafouz. Stability analysis of discrete-time infinite-horizon optimal control with discounted cost. *IEEE Transactions on Automatic Control*, 62 (6): 2736–2749, June 2017. <https://doi.org/10.1109/TAC.2016.2616644>.
 48. L. Qiu, B. Bernhardsson, A. Rantzer, E. Davison, P. Young, and J. Doyle. A formula for computation of the real stability radius. *Automatica*, 31 (6): 879 – 890, 1995. [https://doi.org/10.1016/0005-1098\(95\)00024-Q](https://doi.org/10.1016/0005-1098(95)00024-Q).
 49. D. J. Russo, B. Van Roy, A. Kazerouni, I. Osband, and Z. Wen. A tutorial on thompson sampling. *Foundations and Trends on Machine Learning*, 11(1): 1–96, 2018. <https://doi.org/10.1561/22000000070>.
 50. J. Shao and D. Tu. *The Jackknife and Bootstrap*. Springer Science & Business Media, 2012.
 51. D. Silver, A. Huang, C. J. Maddison, A. Guez, L. Sifre, G. van den Driessche, J. Schrittwieser, I. Antonoglou, V. Panneershelvam, M. Lanctot, S. Dieleman, D. Grewe, J. Nham, N. Kalchbrenner, I. Sutskever, T. Lillicrap, M. Leach, K. Kavukcuoglu, T. Graepel, and D. Hassabis. Mastering the game of go with deep neural networks and tree search. *Nature*, 529: 484–489, 01 2016.
 52. M. Simchowitz, H. Mania, S. Tu, M. I. Jordan, and B. Recht. Learning without mixing: Towards a sharp analysis of linear system identification. In *Proceedings of the 31st Conference On Learning Theory*, volume 75, pages 439–473. PMLR, 06–09 Jul 2018.
 53. M. Sznajder, T. Amishima, P. Parrilo, and J. Tierno. A convex approach to robust \mathcal{H}_2 performance analysis. *Automatica*, 38 (6): 957 – 966, 2002. [https://doi.org/10.1016/S0005-1098\(01\)00299-0](https://doi.org/10.1016/S0005-1098(01)00299-0).
 54. S. Tu, R. Boczar, A. Packard, and B. Recht. Non-Asymptotic Analysis of Robust Control from Coarse-Grained Identification. *arXiv:1707.04791*, 2017.
 55. A. W. Van Der Vaart and J. A. Wellner. *Weak Convergence and Empirical Processes*. Springer Science & Business Media, 1996.
 56. R. Vershynin. Introduction to the non-asymptotic analysis of random matrices. *arXiv:1011.3027*, 2010.
 57. M. Vidyasagar and R. L. Karandikar. A learning theory approach to system identification and stochastic adaptive control. *Journal of Process Control*, 18 (3): 421 – 430, 2008. <https://doi.org/10.1016/j.jprocont.2007.10.009>.
 58. M. J. Wainwright. *High-Dimensional Statistics: A Non-Asymptotic Viewpoint*. Cambridge Series in Statistical and Probabilistic Mathematics. Cambridge University Press, 2019. <https://doi.org/10.1017/9781108627771>.
 59. Y. Wang, N. Matni, and J. C. Doyle. A system level approach to controller synthesis. *IEEE Transactions on Automatic Control*, pages 1–1, 2019. <https://doi.org/10.1109/TAC.2018.2890753>.
 60. F. Wu and A. Packard. Optimal lqg performance of linear uncertain systems using state-feedback. In *Proceedings of 1995 American Control Conference - ACC'95*, volume 6, pages 4435–4439, June 1995. <https://doi.org/10.1109/ACC.1995.532775>.
 61. D. Youla, H. Jabr, and J. Bongiorno. Modern wiener-hopf design of optimal controllers—part ii: The multivariable case. *IEEE Transactions on Automatic Control*, 21 (3): 319–338, 1976. <https://doi.org/10.1109/TAC.1976.1101223>.
 62. P. M. Young, M. P. Newlin, and J. C. Doyle. μ analysis with real parametric uncertainty. In *Proceedings of the 30th IEEE Conference on Decision and Control*, volume 2, pages 1251–1256, Dec 1991. <https://doi.org/10.1109/CDC.1991.261579>.
 63. B. Yu. Rates of convergence for empirical processes of stationary mixing sequences. *The Annals of Probability*, 22 (1): 94–116, 1994.

64. K. Zhou, J. C. Doyle, and K. Glover. *Robust and Optimal Control*. Prentice-Hall, Inc., Upper Saddle River, NJ, USA, 1996.

Publisher's Note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Affiliations

Sarah Dean¹ · Horia Mania¹ · Nikolai Matni² · Benjamin Recht¹ · Stephen Tu¹

✉ Benjamin Recht
brecht@berkeley.edu

Sarah Dean
dean_sarah@berkeley.edu

Horia Mania
hmania@berkeley.edu

Nikolai Matni
nmatni@berkeley.edu

Stephen Tu
stephent@berkeley.edu

¹ Department of Electrical Engineering and Computer Sciences, University of California, Berkeley, CA 94720, USA

² Department of Computing and Mathematical Sciences, California Institute of Technology, Pasadena, CA 91125, USA