Safely Learning to Control the Constrained Linear Quadratic Regulator

Sarah Dean, Stephen Tu, Nikolai Matni, and Benjamin Recht*

September 25, 2018

Abstract

We study the constrained linear quadratic regulator with unknown dynamics, addressing the tension between safety and exploration in data-driven control techniques. We present a framework which allows for system identification through persistent excitation, while maintaining safety by guaranteeing the satisfaction of state and input constraints. This framework involves a novel method for synthesizing robust constraint-satisfying feedback controllers, leveraging newly developed tools from *system level synthesis*. We connect statistical results with cost sub-optimality bounds to give non-asymptotic guarantees on both estimation and controller performance.

1 Introduction

While data-driven design has considerable potential in contemporary control systems where precise modeling of the dynamics is intractable (e.g., systems with complex contact forces), one of the biggest hurdles to overcome for practical deployment is maintaining safe execution during the learning process.

Motivated by this issue, we study the data-driven design of a controller for the *constrained* Linear Quadratic Regulator (LQR) problem. In constrained LQR, we design a controller for a (potentially unknown) linear dynamical system that minimizes a given quadratic cost, subject to the additional requirement that both the state and input stay within a specified safe region. This is a problem that has received much attention within the model predictive control (MPC) community.

For the LQR problem with no constraints, a natural method of exploration for learning the dynamics is to excite the system by injecting white noise. When safety is not an issue, this method is effective and recently Dean et al. [1] provide an end-to-end sample complexity

*S. Dean, S. Tu, N. Matni, and B. Recht are with the Department of Electrical Engineering and Computer Sciences, University of California, Berkeley, CA, 94709 USA (email: dean_sarah@berkeley.edu, stephent@berkeley.edu, nmatni@berkeley.edu, brecht@berkeley.edu)

on this "identify-then-control" scheme. However, this method fails to consider safety or constraint satisfaction.

We directly address the tension between exploration for learning and safety, which are fundamentally at odds. We do this by synthesizing a controller which simultaneously excites and regulates the system; we propose to learn by additively injecting bounded noise to the control inputs computed by a safe controller. By leveraging the recently developed system level synthesis (SLS) framework for control design [2], we give a computationally tractable algorithm which returns a controller that (a) guarantees the closed loop system remains within the specified constraint set and (b) ensures that enough noise can be injected into the system to obtain a statistical guarantee on learning. To the best of our knowledge, our algorithm is the first to simultaneously achieve both objectives. Furthermore, the controller synthesis is solved by a convex optimization problem whose feasibility is a certificate of safety, and no considerations of robust invariant sets are required.

Our second contribution is to provide a sub-optimality bound on control performance for constrained LQR. Using the same SLS framework, we quantify the excess cost incurred by playing a controller designed on the uncertain dynamics obtained from learning, in terms of both the size of the uncertainty sets and a type of constraint robustness margin of the optimal constrained controller for the true system. This allows us to provide the first end-to-end sample complexity guarantee for the control of constrained systems.

1.1 Related Work

Estimation and control of the unconstrained LQR problem has been studied in the non-asymptotic setting [3, 1]. However, the identification schemes rely on pure excitation and system restarts, which is unsuitable in the constrained setting. The online learning literature simultaneously considers learning and control, where strategies are based on *optimism in the face of uncertainty* (OFU) [4] or *Thompson Sampling* [5]. These approaches guarantee system estimation only up to optimal closed-loop equivalence, and do not consider safety. Building on a statistical result by Simchowitz et al. [6] which allows for non-asymptotic guarantees on parameter estimation from a single trajectory of a linear system, Dean et al. [7] provide a robust online method that guarantees parameter estimation and stability throughout.

The design of controllers that guarantee robust constraint satisfaction has long been considered in the context of model predictive control [8], including methods that model uncertainty in the dynamics directly [9], or model it as a bounded state disturbance for computational efficiency [10, 11]. Strategies for incorporating estimation of the dynamics include experiment-design inspired costs [12], decoupling learning from constraint satisfaction [13], and set-membership methods rather than parameter estimation [14]. Due to the receding horizon nature of model predictive controllers, this literature relies on set invariance theory for infinite horizon guarantees [15]. Our framework considers the infinite horizon problem directly, and therefore we do not require computation of invariant sets.

Finally, the machine-learning community has begun to consider safety in reinforcement learning, where much work positions itself as being for general dynamical systems in lieu of providing statistical guarantees [16, 17, 18, 19]. Some works assume the existence of an initial safe controller for learning [20], and robust MPC methods have been proposed to modify potentially unsafe learning inputs [21]. Our framework gives an alternative procedure for designing such a controller using coarse system estimates. Most similar to this work is that of Lu et al. [22], who propose a method to allow excitation on top of a safe controller, but consider only finite-time safety and require non-convex optimization to obtain formal guarantees.

2 Problem Setting and Preliminaries

We fix an underlying linear dynamical system $x_{k+1} = A_\star x_k + B_\star u_k + w_k$, with full state observation, initial condition $x_0 \in \mathbb{R}^n$, sequence of inputs $\{u_k\} \subseteq \mathbb{R}^d$, and disturbance process $\{w_k\} \subseteq \mathbb{R}^n$. The dynamics matrices $(A_\star, B_\star)$ are unknown. For estimates of the system $(\widehat{A}, \widehat{B})$, define $\|A_\star - \widehat{A}\|_p := \|\Delta_A\|_p = \varepsilon_{A,p}$ for $p = 2, \infty$, and similarly for Δ_B and $\varepsilon_{B,p}$.

We assume some prior knowledge is given in the form of initial estimates $(\widehat{A}_0, \widehat{B}_0)$ and uncertainty measures $(\varepsilon^0_{A,p}, \varepsilon^0_{B,p})$. We note that the initial estimates may be coarse grained, and the goal of the learning procedure will be to refine this uncertainty prior to optimal control design.

2.1 System Level Synthesis

Many approaches to optimal control for systems with constraints involve receding horizon control, where an *open loop* finite-time trajectory is computed at each timestep; indeed, parameterizing optimal control problems by a state feedback controller generally leads to nonconvex optimization. Instead, we can parametrize the problem in terms of convolution with the closed-loop system response,

$$x_k = \sum_{t=1}^{k+1} \Phi_x(t) w_{k-t},\ u_k = \sum_{t=1}^{k+1} \Phi_u(t) w_{k-t} \tag{2.1}$$

where we defined $w_{-1} = x_0$ the fixed initial condition. The *system sevel synthesis* (SLS) framework shows that for any elements $\{\Phi_x(t), \Phi_u(t)\}$ constrained to obey, for all $k \geq 1$,

$$\Phi_x(k+1) = A_\star \Phi_x(k) + B_\star \Phi_u(k)\ ,\ \ \Phi_x(1) = I\ ,$$

there exists a controller that achieves the desired system responses (2.1). The state-feedback parameterization result in Theorem 1 of Wang et al. [2] formalizes this observation, and therefore any optimal control problem over linear systems can be cast as an optimization problem over system response elements. We use boldface letters to denote transfer functions, e.g. $\mathbf{\Phi}_x(z) = \sum_{k=1}^\infty \Phi_x(k) z^{-k}$ and signals, $\mathbf{x} = \sum_{k=0}^\infty x_k z^{-k}$. The affine constraints can be rewritten as

$$\begin{bmatrix} zI - A_\star & -B_\star \end{bmatrix} \begin{bmatrix} \mathbf{\Phi}_x \\ \mathbf{\Phi}_u \end{bmatrix} = I\ ,$$

and the corresponding control law $\mathbf{u} = \mathbf{K}\mathbf{x}$ is given by $\mathbf{K} = \mathbf{\Phi}_u \mathbf{\Phi}_x^{-1}$.

2.2 Notation

In this paper, we restrict our attention to the function space $\mathcal{RH}_\infty$, consisting of (discrete-time) stable matrix-valued transfer functions. We use $\frac{1}{z}\mathcal{RH}_\infty$ to denote the set of transfer functions $\mathbf{G}$ such that $z\mathbf{G} \in \mathcal{RH}_\infty$. We further use the notation $\mathcal{RH}_\infty(C,\rho) := \{\mathbf{M} = \sum_{k=0}^{\infty} M(k)z^{-k} \mid \|M(k)\|_2 \leq C\rho^k,\ k = 0,1,2,...\}$ for transfer functions that satisfy a certain decay rate in the spectral norm of their impulse response elements.

When working with transfer functions and signals, denote the coefficient of the term of degree k as $\mathbf{G}[k] = G(k)$ and $\mathbf{x}[k] = x_k$. We will also denote $G[k:1]$ as the block row vector of system response elements of $\mathbf{G}$

$$G[k:1] = \begin{bmatrix} G(k) & \dots & G(1) \end{bmatrix} .$$

As is standard, we let $\|x\|_p$ denote the ℓ_p-norm of a vector x. For a matrix M, we let $\|M\|_p$ denote its $\ell_p \to \ell_p$ operator norm. We will consider the $\mathcal{H}_2$, $\mathcal{H}_\infty$, and $\mathcal{L}_1$ norms, which are infinite horizon analogs of the Frobenius, spectral, and $\ell_\infty \to \ell_\infty$ operator norms of a matrix, respectively: $\|\mathbf{M}\|_{\mathcal{H}_2} = \sqrt{\sum_{k=0}^{\infty}\|M(k)\|_F^2}$, $\|\mathbf{M}\|_{\mathcal{H}_\infty} = \sup_{\|\mathbf{w}\|_2=1} \|\mathbf{M}\mathbf{w}\|_2$, and $\|\mathbf{M}\|_{\mathcal{L}_1} = \sup_{\|\mathbf{w}\|_\infty=1} \|\mathbf{M}\mathbf{w}\|_\infty$.

Finally, for two numbers a, b, we let $a \lesssim b$ (resp. $a \gtrsim b$) denote that there exists an absolute constant $C > 0$ such that $a \leq Cb$ (resp. $a \geq Cb$).

2.3 Optimal Control Problem

We now describe the constrained optimal control problem (OCP) that we would want to solve given perfect knowledge of $(A_\star, B_\star)$. This formulation acts as our baseline:

$$\begin{aligned} \min_{\mathbf{u}\in\mathcal{K}} \ & J_\sigma(A_\star, B_\star, \mathbf{K}) \\ \text{s.t. } & x_0 \text{ fixed}; \quad F_x x_k \leq b_x, \quad F_u u_k \leq b_u \\ & \qquad \forall k\ , \forall \{w_k : \|w_k\|_\infty \leq \sigma_w\} . \end{aligned} \tag{2.2}$$

This problem is to be interpreted as follows. First, let the set $\mathcal{K}$ enumerate all inputs that result from linear dynamic stabilizing feedback controllers for $(A_\star, B_\star)$ of the form $\mathbf{u} = \mathbf{K}\mathbf{x}$. This is made possible by the system level synthesis framework described above. The cost

$$J_\sigma(A_\star, B_\star, \mathbf{K}) = \lim_{T\to\infty} \frac{1}{T} \sum_{k=0}^{T-1} \mathbb{E}_w[x_k^\top Q x_k + u_k^\top R u_k]$$

where the system $(A_\star, B_\star)$ is in feedback with $\mathbf{K}$, $\{w_k\}$ is any distribution that satisfies $\mathbb{E}[w_k] = 0$ and $\mathbb{E}[w_k w_k^\top] = \sigma^2 I$ and is independent across time, i.e., $w_k \perp w_\ell$ for $\ell \neq k$. On the other hand, the constraints read that for every possible realization $\{w_k\}$ satisfying $\sup_{k\geq 0}\|w_k\|_\infty \leq \sigma_w$, the trajectory $\{x_k\}$ and the inputs $\{u_k\}$ coming from the system dynamics $(A_\star, B_\star)$ in feedback with the law $\mathbf{K}$ are contained within the state and input constraint polytopes $F_x x \leq b_x$ and $F_u u \leq b_u$, respectively.

We note that the OCP given in (2.2) is a convex, but infinite-dimensional problem. It is an idealized baseline to compare our actual solutions to; our sub-optimality guarantees will be with respect to the optimal cost achieved by this idealized problem. This is a desirable baseline, since it optimizes for average case performance J_σ but ensures safety for the worst-case behavior, consistent with MPC literature [10, 23]. We remark that an alternative to (2.2) is to replace the worst case constraint behavior with probabilistic chance constraints [24]. We do not work with chance constraints because they are generally difficult to directly enforce on an infinite horizon; arguments around recursive feasibility using robust invariant sets are common in the MPC literature to deal with this issue.

3 Constraint-Satisfying Control

We begin by formulating a method for robustly operating a system while maintaining safety. First, a system level synthesis approach to the constrained LQR problem is described and then modified to be robust to uncertainties in system dynamics. Finally, we discuss a reduction to tractable a finite-dimensional optimization problem.

3.1 A System Level Approach

Using the SLS formulation, we define an optimization problem that solves the OCP (2.2).

Proposition 3.1. *The following convex optimization problem solves OCP* (2.2)*.*

$$\begin{aligned} \min_{\mathbf{\Phi}_x, \mathbf{\Phi}_u} \quad & \left\| \begin{bmatrix} Q^{1/2} & \\ & R^{1/2} \end{bmatrix} \begin{bmatrix} \mathbf{\Phi}_x \\ \mathbf{\Phi}_u \end{bmatrix} \right\|_{\mathcal{H}_2} \\ \text{s.t.} \quad & \begin{bmatrix} zI - A_\star & -B_\star \end{bmatrix} \begin{bmatrix} \mathbf{\Phi}_x \\ \mathbf{\Phi}_u \end{bmatrix} = I, \\ & G_x(\mathbf{\Phi}_x; k) \leq b_x, \quad G_u(\mathbf{\Phi}_u; k) \leq b_u \quad \forall\ k \geq 0\,. \end{aligned} \tag{3.1}$$

where

$$\begin{aligned} G_x(\mathbf{\Phi}_x; k)_j &= F_{x,j}^\top \Phi_x(k+1) x_0 + \sigma_w \| F_{x,j}^\top \Phi_x[k:1] \|_1\,, \\ G_u(\mathbf{\Phi}_u; k)_j &= F_{u,j}^\top \Phi_u(k+1) x_0 + \sigma_w \| F_{u,j}^\top \Phi_u[k:1] \|_1\,, \end{aligned}$$

with j indexing the rows of F_x and F_u.

With $\mathbf{K} = \mathbf{\Phi}_u(\mathbf{\Phi}_x)^{-1}$, we define the LQR cost on the true system (omitting the constant multiple σ) as

$$J(A_\star, B_\star, \mathbf{K}) = \left\| \begin{bmatrix} Q^{1/2} & \\ & R^{1/2} \end{bmatrix} \begin{bmatrix} \mathbf{\Phi}_x \\ \mathbf{\Phi}_u \end{bmatrix} \right\|_{\mathcal{H}_2} .$$

We remark that the feasibility of the convex synthesis problem in (3.1) for an initial condition x_0 implies that x_0 is a member of a robust control invariant set.

Proof. By the state-feedback parameterization result in Theorem 1 of [2], the SLS parametrization encompasses all internally stabilizing state-feedback controllers acting on the true system $(A_\star, B_\star)$. Thus, it is necessary only to show that the optimization problem in (3.1) is consistent with that of (2.2) under the system level parametrization. The equivalence between the LQR cost and the $\mathcal{H}_2$ system norm is standard and omitted for brevity – see the Appendix of [1] for this reformulation in terms of system responses.

Therefore, it remains to consider the inequality constraints. Because the constraints must be satisfied robustly, it is equivalent to consider

$$\max_{\{w_k\}} F_x \sum_{t=1}^{k+1} \Phi_x(t) w_{k-t} \leq b_x$$
$$\iff F_x \Phi_x(k+1) x_0 + \max_{\{w_k\}} F_x \sum_{t=1}^{k} \Phi_x(t) w_{k-t} \leq b_x \,.$$

Then considering elements in the second term for $j = 1, \dots, n_c$,

$$\begin{aligned}\max_{\{w_k\}} F_{x,j}^\top \sum_{t=1}^{k} \Phi_x(t) w_{k-t} &= \sum_{t=1}^{k} \max_{\|w\|_\infty \leq \sigma_w} F_{x,j}^\top \Phi_x(t) w \\ &= \sum_{t=1}^{k} \sigma_w \|F_{x,j}^\top \Phi_x(t)\|_1 = \sigma_w \|F_{x,j}^\top \Phi_x[k:1]\|_1 \,.\end{aligned}$$

Thus the inequality constraint on the function $G_x(\Phi_x; k)$ is an equivalent condition. A similar computation holds for the input constraint. $\square$

3.2 Robust Control

Further motivation for reformulating the optimal control problem in terms of system responses is the ability to transparently consider uncertainties in the dynamics. Recall that we consider controller synthesis under model errors, where only nominal estimates of the system are known. Then the model mismatch impacts the closed-loop system in a transparent way:

Proposition 3.2. *Define* $\widehat{\mathbf{\Delta}} = \begin{bmatrix} \Delta_A & \Delta_B \end{bmatrix} \begin{bmatrix} \widehat{\mathbf{\Phi}}_x \\ \widehat{\mathbf{\Phi}}_u \end{bmatrix}$. *If* $\widehat{\mathbf{\Phi}}_x$, $\widehat{\mathbf{\Phi}}_u$ *satisfy* $\begin{bmatrix} zI - \widehat{A} & -\widehat{B} \end{bmatrix} \begin{bmatrix} \widehat{\mathbf{\Phi}}_x \\ \widehat{\mathbf{\Phi}}_u \end{bmatrix} = I$ *and* $\|\widehat{\mathbf{\Delta}}\|_{\mathcal{H}_\infty} < 1$, *then on the true system, the controller* $\widehat{\mathbf{K}} = \widehat{\mathbf{\Phi}}_u (\widehat{\mathbf{\Phi}}_x)^{-1}$ *achieves the system response* $\begin{bmatrix} \widehat{\mathbf{\Phi}}_x \\ \widehat{\mathbf{\Phi}}_x \end{bmatrix} (I + \widehat{\mathbf{\Delta}})^{-1}$ *and cost bounded by*

$$J(A_\star, B_\star, \widehat{\mathbf{K}}) \leq \frac{1}{1 - \|\widehat{\mathbf{\Delta}}\|_{\mathcal{H}_\infty}} J(\widehat{A}, \widehat{B}, \widehat{\mathbf{K}}) \,.$$

Proof. We make use of the robust stability result in Theorem 2 of Matni et al. [25] and note that $\|\widehat{\mathbf{\Delta}}\| < 1$ is a sufficient condition for the existence of the inverse $(I + \widehat{\mathbf{\Delta}})^{-1}$ for any induced norm $\|.\|$ by the small gain theorem. Then by the sub-multiplicativity of the $\mathcal{H}_2$ and $\mathcal{H}_\infty$ norms,

$$J(A_\star, B_\star, \widehat{\mathbf{K}}) \leq \left\| \begin{bmatrix} Q^{1/2}\widehat{\mathbf{\Phi}}_x \\ R^{1/2}\widehat{\mathbf{\Phi}}_u \end{bmatrix} \right\|_{\mathcal{H}_2} \|(I + \widehat{\mathbf{\Delta}})^{-1}\|_{\mathcal{H}_\infty} .$$

□

Motivated by this result, consider the following robust optimization problem:

$$\begin{aligned} \min_{\mathbf{\Phi}_x, \mathbf{\Phi}_u} \ & \frac{1}{1-\gamma} J(\widehat{A}, \widehat{B}, \mathbf{K}) \\ \text{s.t. } & \begin{bmatrix} zI - \widehat{A} & -\widehat{B} \end{bmatrix} \begin{bmatrix} \mathbf{\Phi}_x \\ \mathbf{\Phi}_u \end{bmatrix} = I, \\ & \sqrt{2} \left\| \begin{bmatrix} \varepsilon_{A,2}\mathbf{\Phi}_x \\ \varepsilon_{B,2}\mathbf{\Phi}_u \end{bmatrix} \right\|_{\mathcal{H}_\infty} \leq \gamma, \quad \left\| \begin{bmatrix} \varepsilon_{A,\infty}\mathbf{\Phi}_x \\ \varepsilon_{B,\infty}\mathbf{\Phi}_u \end{bmatrix} \right\|_{\mathcal{L}_1} \leq \tau, \\ & G_x^\tau(\Phi_x; k) \leq b_x, \quad G_u^\tau(\Phi_u; k) \leq b_u \ \ \forall\ k\ . \end{aligned} \tag{3.2}$$

where γ, τ are fixed parameters and

$$\begin{aligned} G_x^\tau(\mathbf{\Phi}_x; k)_j &= G_x(\mathbf{\Phi}_x; k)_j + \tfrac{\tau\sigma_w c_0}{1-\tau} \|F_{x,j}^\top \Phi_x[k+1:1]\|_1 \ , \\ G_u^\tau(\mathbf{\Phi}_u; k)_j &= G_u(\mathbf{\Phi}_u; k)_j + \tfrac{\tau\sigma_w c_0}{1-\tau} \|F_{x,j}^\top \Phi_u[k+1:1]\|_1 \ , \end{aligned}$$

where we define $c_0 = \max(1, \frac{1}{\sigma_w}\|x_0\|_\infty)$.

Theorem 3.3. *Any controller designed from a feasible solution to the robust control problem (3.2) for any $0 \leq \gamma, \tau < 1$ will stabilize the true system. Furthermore, the state and input constraints will be satisfied.*

Proof. First, note that

$$\begin{aligned} \|\widehat{\mathbf{\Delta}}\|_{\mathcal{H}_\infty} &\leq \sqrt{2} \left\| \begin{bmatrix} \varepsilon_{A,2}\mathbf{\Phi}_x \\ \varepsilon_{B,2}\mathbf{\Phi}_u \end{bmatrix} \right\|_{\mathcal{H}_\infty} < 1 \ , \\ \|\widehat{\mathbf{\Delta}}\|_{\mathcal{L}_1} &\leq \left\| \begin{bmatrix} \varepsilon_{A,\infty}\mathbf{\Phi}_x \\ \varepsilon_{B,\infty}\mathbf{\Phi}_u \end{bmatrix} \right\|_{\mathcal{L}_1} < 1\ . \end{aligned} \tag{3.3}$$

Then by Proposition 3.2, the true system trajectory will be given by

$$\mathbf{x} = \mathbf{\Phi}_x(I + \widehat{\mathbf{\Delta}})^{-1}\mathbf{w}, \quad \mathbf{u} = \mathbf{\Phi}_u(I + \widehat{\mathbf{\Delta}})^{-1}\mathbf{w}\ .$$

Therefore, the state constraints are satisfied as long as

$$\begin{aligned} b_x &\geq \max_{\mathbf{w}} F_x(\mathbf{\Phi}_x(I + \widehat{\mathbf{\Delta}})^{-1}\mathbf{w})[k] \\ &= \max_{\mathbf{w}} F_x(\mathbf{\Phi}_x\mathbf{w})[k] - F_x(\mathbf{\Phi}_x\widehat{\mathbf{\Delta}}(I + \widehat{\mathbf{\Delta}})^{-1}\mathbf{w})[k]\ . \end{aligned}$$

The first term reduces to $G_x(\Phi_x; k)$ as in the non-robust case. Because information about $\widehat{\mathbf{\Delta}}$ is not known, we resort to a sufficient condition to bound the second term, letting $\tilde{\mathbf{w}} = \widehat{\mathbf{\Delta}}(I + \widehat{\mathbf{\Delta}})^{-1}\mathbf{w}$,

$$\begin{aligned}
&|F_{x,j}^\top(\mathbf{\Phi}_x \tilde{\mathbf{w}})[k]| \\
&\quad \le \|F_{x,j}^\top \Phi_x[k+1:1]\|_1 \|\tilde{\mathbf{w}}\|_\infty \\
&\quad \le \|F_{x,j}^\top \Phi_x[k+1:1]\|_1 \frac{\|\widehat{\mathbf{\Delta}}\|_{\mathcal{L}_1}}{1 - \|\widehat{\mathbf{\Delta}}\|_{\mathcal{L}_1}} \times \\
&\qquad\qquad \max(\sigma_w, \|x_0\|_\infty) \\
&\quad \le \|F_{x,j}^\top \Phi_x[k+1:1]\|_1 \frac{\tau}{1-\tau} \sigma_w c_0 \,.
\end{aligned}$$

Consequently, a sufficient condition for satisfying state constraints is to have for all $j = 1, \dots, n_c$,

$$G_x(\mathbf{\Phi}_x; k)_j + \frac{\tau}{1-\tau}\sigma_w c_0 \|F_{x,j}^\top \Phi_x[k+1:1]\|_1 \le b_{x,j} \,.$$

Therefore, the constraints on $\mathbf{\Phi}_x$ imply that the state constraints are satisfied. Similar logic shows that the constraints on $\mathbf{\Phi}_u$ imply that the input constraints are satisfied. $\square$

3.3 Finite Dimensional Reduction

To make controller synthesis tractable, we can solve a finite approximation to optimization problem (3.2) wherein we only optimize over the first L impulse response elements of $\mathbf{\Phi}_x$ and $\mathbf{\Phi}_u$, treating them as finite impulse response (FIR) filers. We show that in this setting, the optimization variables and constraints admit finite-dimensional representations. We first reformulate the constraints. Starting with the affine constraint, we have for $k = 1, ..., L-1$

$$\begin{gathered}
\Phi_x(1) = I, \ \Phi_x(k+1) = \widehat{A}\Phi_x(k) + \widehat{B}\Phi_u(k) \\
V = \widehat{A}\Phi_x(L) + \widehat{B}\Phi_u(L) \,,
\end{gathered} \tag{3.4}$$

where we will also optimize over V, a term which captures the tail of the system responses that we ignore in the synthesis.

Next, considering the system norm constraints, the $\mathcal{H}_\infty$ norm can be reduced to a compact SDP over Φ_x, Φ_u, γ as in Theorem 5.8 of Dumitrescu [26], described explicitly for this setting in Appendix G.3 of Dean et al. [7]. For the $\mathcal{L}_1$ norm, the constraint becomes an $\ell_\infty \to \ell_\infty$ operator norm bound,

$$\left\| \begin{bmatrix} \varepsilon_{A,\infty}\mathbf{\Phi}_x \\ \varepsilon_{B,\infty}\mathbf{\Phi}_u \end{bmatrix} \right\|_{\mathcal{L}_1} = \left\| \begin{bmatrix} \varepsilon_{A,\infty}\Phi_x[L:1] \\ \varepsilon_{B,\infty}\Phi_u[L:1] \end{bmatrix} \right\|_\infty + \|V\|_\infty \le \tau \,, \tag{3.5}$$

where the tail variable V enters transparently. For $k = 1, \dots, L-1$, the inequality constraints on $G_x^\tau(\Phi_x; k)$ and $G_u^\tau(\Phi_u; k)$ remain. For any $k \ge L$, the expression reduces to, for $j =$

$1, \dots, n_c$

$$\begin{aligned}\sigma_w(1+\frac{\tau c_0}{1-\tau})\|F_{x,j}^\top \Phi_x[L:1]\|_1 \le b_{x,j}\,,\\ \sigma_w(1+\frac{\tau c_0}{1-\tau})\|F_{u,j}^\top \Phi_u[L:1]\|_1 \le b_{u,j}\,.\end{aligned} \tag{3.6}$$

Therefore, the synthesis problem becomes

$$\begin{aligned}\min_{\Phi_x,\Phi_u,V} \quad & \tfrac{1}{1-\gamma}\textstyle\sum_0^L \mathbf{Tr}(\Phi_x(t)^\top Q\Phi_x(t)+\Phi_u(t)^\top R\Phi_u(t))\\ \text{s.t.} \quad & (3.4),\ \mathrm{SDP}(\Phi_x,\Phi_u,\gamma-\|V\|_2),\ (3.5),\ (3.6)\,,\\ & 1\le k\le L-1: G_x^\tau(\Phi_x;k)\le b_x,\ G_u^\tau(\Phi_u;k)\le b_u.\end{aligned} \tag{3.7}$$

This is a finite dimensional SDP. The controller given by $\mathbf{K}=\mathbf{\Phi}_u\mathbf{\Phi}_x^{-1}$ can be written in an equivalent state-space realization (A_K, B_K, C_K, D_K) via Theorem 2 of Anderson et al. [27].

4 Suboptimality Guarantees

How much is control performance degraded by uncertainties about the dynamics? In this section, we derive a sub-optimality bound which answers this question for the constrained LQR problem. First, consider the addition of an outer minimization over γ and τ:[1]

$$\min_{\gamma,\tau}\ (3.2)\,. \tag{4.1}$$

Denote the solution to the true optimal control problem as $(\mathbf{\Phi}_x^\star, \mathbf{\Phi}_u^\star)$, then define $\mathbf{K}_\star = \mathbf{\Phi}_u^\star\mathbf{\Phi}_x^{\star\,-1}$ and $J_\star = J(A_\star, B_\star, \mathbf{K}_\star)$. Additionally, define constants related to the optimal system norm and the dynamics uncertainties:

$$\zeta_\infty = (\varepsilon_{A,\infty}+\varepsilon_{B,\infty}\|\mathbf{K}_\star\|_{\mathcal{L}_1})\|\mathbf{\Phi}_x^\star\|_{\mathcal{L}_1}\,,$$

and $\zeta_2 = (\varepsilon_{A,2}+\varepsilon_{B,2}\|\mathbf{K}_\star\|_{\mathcal{H}_\infty})\|\mathbf{\Phi}_x^\star\|_{\mathcal{H}_\infty}$.

Theorem 4.1. *Define the constraint robustness margins of the optimal constrained controller as*

$$\mathrm{margin}_x = \inf_{\substack{k\ge 0\\ 0\le j\le n_c}} \frac{b_{x,j}-G_x(\mathbf{\Phi}_x^\star;k)_j}{\sigma_w c_0\|F_{x,j}^\top\Phi_x^\star[k+1:1]\|_1}$$

and similarly for margin_u*. Then, as long as* $\zeta_2\le\frac{1}{4\sqrt{2}}$ *and* $\zeta_\infty\le\min\left(\frac{\mathrm{margin}_x}{10},\frac{\mathrm{margin}_u}{10},\frac{1}{4}\right)$*, we have that the cost achieved by* $\widehat{\mathbf{K}}=\widehat{\mathbf{\Phi}}_u\widehat{\mathbf{\Phi}}_x^{-1}$ *synthesized from the minimizers of* (4.1) *satisfies*

$$\tfrac{J(A_\star,B_\star,\widehat{\mathbf{K}})-J_\star}{J_\star}\le 2\sqrt{2}(\varepsilon_{A,2}+\varepsilon_{B,2}\|\mathbf{K}_\star\|_{\mathcal{H}_\infty})\|\mathbf{\Phi}_x^\star\|_{\mathcal{H}_\infty}\,.$$

[1] The objective (3.2) is unimodal in γ,τ individually, and therefore this outer minimization can be achieved by searching over the box $[0,1)\times[0,1)$. For less computational complexity, the minimization need only be over a single outer variable: $\max(\gamma,\tau)$. In this case, the sub-optimality bound will retain the same flavor, but the norm distinctions between cost and constraints will be less clear.

While this result is stated in terms of quantities related to the unknown true system, we note that a similar data-dependent expression could be derived that depends only on the estimated system. We further remark that the condition on constraint robustness margins may be restrictive; for systems operating close to their constraints, our theorem requires near-perfect knowledge before guaranteeing sub-optimality.

Proof. Using Proposition 3.2 along with the norm bounds (3.3) and the constraints in optimization problem (4.1),

$$J(A_\star, B_\star, \widehat{\mathbf{K}}) \leq \frac{1}{1-\widehat{\gamma}} J(\widehat{A}, \widehat{B}, \widehat{\mathbf{K}})\,.$$

Next, we will use the following lemma

Lemma 4.2. *Under the conditions of Theorem 4.1, we have that the following is a feasible solution to* (4.1)

$$\tilde{\mathbf{\Phi}}_x = \mathbf{\Phi}_x^\star (I - \mathbf{\Delta})^{-1},\ \ \tilde{\mathbf{\Phi}}_u = \mathbf{K}_\star \mathbf{\Phi}_x^\star (I - \mathbf{\Delta})^{-1},$$
$$\tilde{\gamma} = \frac{\sqrt{2}\zeta_2}{1-\sqrt{2}\zeta_2}, \quad \tilde{\tau} = \frac{\zeta_\infty}{1-\zeta_\infty}\,.$$

where we define $\mathbf{\Delta} = -\begin{bmatrix}\Delta_A & \Delta_B\end{bmatrix}\begin{bmatrix}\mathbf{\Phi}_x^\star \\ \mathbf{\Phi}_u^\star\end{bmatrix}$.

The proof of Lemma 4.2 follows by checking that the proposed solution satisfies all the constraints and is presented in Appendix 9. Applying Lemma 4.2,

$$\frac{1}{1-\widehat{\gamma}} J(\widehat{A}, \widehat{B}, \widehat{\mathbf{K}}) \leq \frac{1}{1-\tilde{\gamma}} J(\widehat{A}, \widehat{B}, \mathbf{K}_\star)\,.$$

This is true because $(\widehat{\mathbf{K}}, \widehat{\gamma})$ is the optimal solution to (4.1), so objective function with feasible $(\tilde{\mathbf{\Phi}}_u \tilde{\mathbf{\Phi}}_x^{-1} = \mathbf{K}_\star, \tilde{\gamma})$ is an upper bound. Then we have

$$\begin{aligned} J(A_\star, B_\star, \widehat{\mathbf{K}}) &\leq \frac{1}{1-\tilde{\gamma}} J(\widehat{A}, \widehat{B}, \mathbf{K}_\star) \\ &\leq \frac{1}{1-\tilde{\gamma}} \frac{1}{1-\|\mathbf{\Delta}\|_2} J(A_\star, B_\star, \mathbf{K}_\star) \\ &\leq (1 + 4\sqrt{2}\zeta_2) J(A_\star, B_\star, \mathbf{K}_\star)\,. \end{aligned}$$

The second inequality follows from an application of Proposition 3.2 with the roles of the nominal and true systems switched. The final follows from bounding $\|\mathbf{\Delta}\|_2$ by $\sqrt{2}\zeta$ and noticing that $\frac{x}{1-x} \leq 2x$ for $0 \leq x \leq \frac{1}{2}$, where we set $x = 2\sqrt{2}\zeta_2$. □

Here, we briefly remark that a similar sub-optimality bound can be derived for the finite problem in (3.7). In short, controllers synthesized from the optimization problem $\min_{\gamma,\tau}$ (3.7) will satisfy a sub-optimality bound of the form in Theorem 4.1 with an additional factor due to the FIR truncation. The formal statement and proof of this result are deferred to Appendix 9, but we highlight here that the cost penalty incurred due to FIR approximation decays exponentially in the horizon L over which the approximation is taken.

5 Learning with Control

Finally, we connect the previous results on robust control with system estimation. To show a priori guarantees on statistical learning we adopt control actions that both keep the system safe and provide excitation,

$$\mathbf{u} = \mathbf{K}\mathbf{x} + \eta \tag{5.1}$$

where each $\eta = (\eta_0, \eta_1, \dots)$ is stochastic and ℓ_∞-bounded, i.e. $\|\eta_k\|_\infty \leq \sigma_\eta$. Given a trajectory sequence $\{(x_k, u_k)\}_{k=0}^T$, we propose to learn the dynamics $(A_\star, B_\star)$ via least-squares regression on a trajectory of length T:

$$(\widehat{A}, \widehat{B}) \in \arg\min_{(A,B)} \sum_{t=0}^{T-1} \frac{1}{2}\|Ax_t + Bu_t - x_{t+1}\|_2^2 \,. \tag{5.2}$$

We will prove a statistical rate on the least-squares estimate $(\widehat{A}, \widehat{B})$ in terms of the system response and the trajectory length.

The bulk of the proof for the statistical rate comes from a general theorem regarding linear-response time series data from Simchowitz et al. [6]. Recently, this proof was adopted by Dean et al. [7] to show a rate of estimation in the setting given by (5.1) when both η and the disturbance w are Gaussian distributed. We modify the reduction given by Dean et al. to the case when the excitation and disturbance are no longer Gaussian, but instead zero-mean and bounded. We assume that w_t and η_t are both zero-mean sequences with independent coordinates and finite fourth moments. In particular, we assume $\mathbb{E}_{w_t}[w_t(i)^2] = \sigma_w^2$, $\mathbb{E}_{w_t}[w_t(i)^4] \lesssim \sigma_w^4$, $\mathbb{E}_{\eta_t}[\eta_t(i)^2] = \sigma_\eta^2$, $\mathbb{E}_{\eta_t}[\eta_t(i)^4] \lesssim \sigma_\eta^4$. These assumptions are quickly verified for common distributions such as uniform on a compact interval or over a discrete set of points. The main estimation result is the following.

Theorem 5.1. *Fix a failure probability $\delta \in (0, 1)$. Suppose the stochastic disturbance $\{w_t\}$ and the input disturbance $\{\eta_t\}$ satisfy the assumptions above. Assume for simplicity that $\sigma_\eta \leq \sigma_w$, and that the stabilizing controller $\mathbf{K}$ achieves a SLS response $\mathbf{\Phi}_x \in \frac{1}{z}\mathcal{RH}_\infty(C_x, \rho)$, $\mathbf{\Phi}_u \in \frac{1}{z}\mathcal{RH}_\infty(C_u, \rho)$. Let $C_K^2 := nC_x^2 + dC_u^2$. Then as long as the trajectory length T satisfies the condition:*

$$T \gtrsim T_0 := (n+d)\log\Bigg(\frac{dC_u^2}{\delta} + \frac{\sigma_w^2}{\sigma_\eta^2}\frac{\rho^2 C_u^2 C_K^2}{\delta(1-\rho^2)}\left(1 + \|B_\star\|_2^2 + \frac{\|x_0\|_2^2}{\sigma_w^2 T}\right)\Bigg)\,, \tag{5.3}$$

we have the following bound on the least-squares estimation errors that holds with probability at least $1-\delta$,

$$\max\{\|\Delta_A\|_2, \|\Delta_B\|_2\} \lesssim \frac{\sigma_w C_u}{\sigma_\eta}\sqrt{\frac{n+d}{T}} \times \sqrt{\log\left(\frac{dC_u}{\delta} + \frac{\sigma_w}{\sigma_\eta}\frac{\rho C_u C_K}{\delta(1-\rho^2)}\left(1 + \|B_\star\|_2 + \frac{\|x_0\|_2}{\sigma_w\sqrt{T}}\right)\right)}\,.$$

The proof of this result is presented in Appendix 8. We remark on the interpretation of statistical learning bounds. A priori guarantees, like the one presented here, depend on quantities related to the underlying true system. As such, they are not directly useful when the system is unknown,[2] but rather they indicate qualities of systems that make them easier or harder to estimate.

Corollary 5.2. *If the robust control synthesis problem* (3.2) *is feasible for any* $0 < \gamma, \tau < 1$, *initial system estimates* (A_0, B_0), *initial dynamics uncertainties* $(\varepsilon_A^0, \varepsilon_B^0)$, σ_w *replaced with*[3] $\sigma_\eta \|B_\star\|_\infty + \sigma_w$, *and* $b_{u,j}$ *replaced with* $b_{u,j} - \sigma_\eta \|F_{u,j}\|_1$, *then the resulting control law* $\mathbf{u} = \mathbf{K}\mathbf{x} + \eta$ *with stochastic* $\|\eta\|_\infty \leq \sigma_\eta$ *stabilizes the true system, satisfies state and input constraints, and allows for learning at the rate given in Theorem 5.1.*

Proof. (Sketch) The proposed control law is equivalent to the system controlled by a deterministic control law plus an enlarged process noise distribution $\tilde{w}_k = B_\star \eta_k + w_k$. Therefore, the stability and constraint satisfaction follow from Theorem 3.3. Since the control law is of the form (5.1), the results of Theorem 5.1 hold. □

Finally, we connect the sub-optimality result to the statistical learning bound for an end-to-end sample complexity bound on the constrained LQR problem.

Corollary 5.3. *Assume initial feasibility of the learning problem. For simplicity, assume* $\mathrm{margin}_x \leq \min(\mathrm{margin}_u, \frac{5}{2})$. *Then for*

$$T \gtrsim T_0 \frac{\sigma_w^2 C_u^2}{\sigma_\eta^2} \max\left\{ \frac{(n+d)}{\mathrm{margin}_x^2} (1 + \|\mathbf{K}_\star\|_{\mathcal{L}_1})^2 \|\mathbf{\Phi}_x^\star\|_{\mathcal{L}_1}^2, \right. \\ \left. (1 + \|\mathbf{K}_\star\|_{\mathcal{H}_\infty})^2 \|\mathbf{\Phi}_x^\star\|_{\mathcal{H}_\infty}^2 \right\}, \tag{5.4}$$

the cost achieved by $\widehat{\mathbf{K}} = \widehat{\mathbf{\Phi}}_u (\widehat{\mathbf{\Phi}}_x)^{-1}$ *synthesized from* (3.2) *on the least-squares estimates* $\widehat{A}, \widehat{B}$ *satisfies with probability at least* $1 - \delta$,

$$\frac{J(A_\star, B_\star, \widehat{\mathbf{K}}) - J_\star}{J_\star} \lesssim \\ \frac{\sigma_w C_u}{\sigma_\eta} \sqrt{\frac{n+d}{T}} (1 + \|\mathbf{K}_\star\|_{\mathcal{H}_\infty}) \|\mathbf{\Phi}_x^\star\|_{\mathcal{H}_\infty} \mathcal{O}(\sqrt{\log(d/\delta)}) \,.$$

Proof. (Sketch) This result follows by combining the statistical guarantee in Theorem 5.1 with the sub-optimality bound in Theorem 4.1. Note that we use the naïve bound $\varepsilon_{A,\infty} \leq \sqrt{n}\varepsilon_{A,2}$ and similarly $\varepsilon_{B,\infty} \leq \sqrt{d}\varepsilon_{B,2}$; this results in an extra factor of $(n+d)$ appearing in (5.4). □

[2] Statistical bounds in terms of data-dependent quantities can also be worked out; however, modern methods like bootstrapping generally provide tighter statistical guarantees [28].

[3] Note that since the quantity $\|B_\star\|_\infty$ would not generally be known, it can be bounded by $\|B_0\|_\infty + \varepsilon_{B,\infty}^0$.

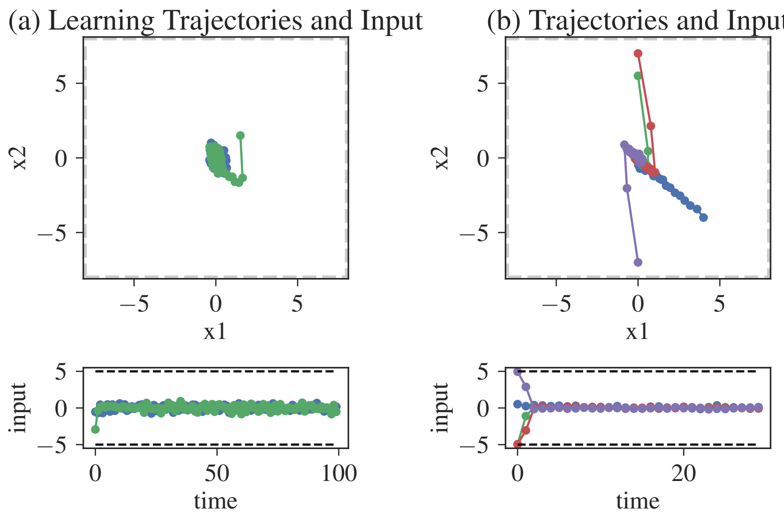

Figure 1: Safe learning trajectories synthesized with coarse initial estimates (a), then robust execution with reduced model errors (b).

Notice that this result depends both on the true system and the initial system estimates by way of the learning controller, which affects T_0 and constants in the $\mathcal{O}(\sqrt{\log(d/\delta)})$ term. The system constraints enter through their effect on margin_x, and while they may impact the waiting time, they do not influence the ultimate cost sub-optimality.

6 Numerical Experiments

We demonstrate the utility of this framework on the double integrator example. In this case, the true dynamics are given by

$$x_{k+1} = \begin{bmatrix} 1 & 0.1 \\ 0 & 1 \end{bmatrix} x_k + \begin{bmatrix} 0 \\ 1 \end{bmatrix} u_k + w_k$$

with the constraints as states bounded between -8 and 8, and inputs bounded in between -4 and 4. We have $\sigma_w = 0.1$. Our initial estimate comes from a randomly generated initial perturbation of the true system with $\varepsilon_{A,\infty} = \varepsilon_{B,\infty} = 0.1$. Safe controllers are generated with finite truncation length $L = 15$, and for larger initial conditions, the system is warm-started with a finite-time robust controller with horizon 20 to reduce the initial condition.

Figure 1 displays safe trajectories and input sequences for several example initial conditions. In 1a, the plotted trajectories are used for learning: the controller both regulates and excites the system ($\sigma_\eta = 0.5$), and is robust to initial uncertainties. Figure 1b demonstrates an ability to operate closer to the margin when there is less uncertainty: in this case, there is no added excitation ($\sigma_\eta = 0$) and the system estimates are better specified ($\varepsilon_\infty = 0.001$), so larger initial conditions are feasible.

Figure 2a displays the decreasing estimation errors over time, demonstrating learning. Shaded areas represent quartiles over 400 trials. Figure 2b displays the trade-off between

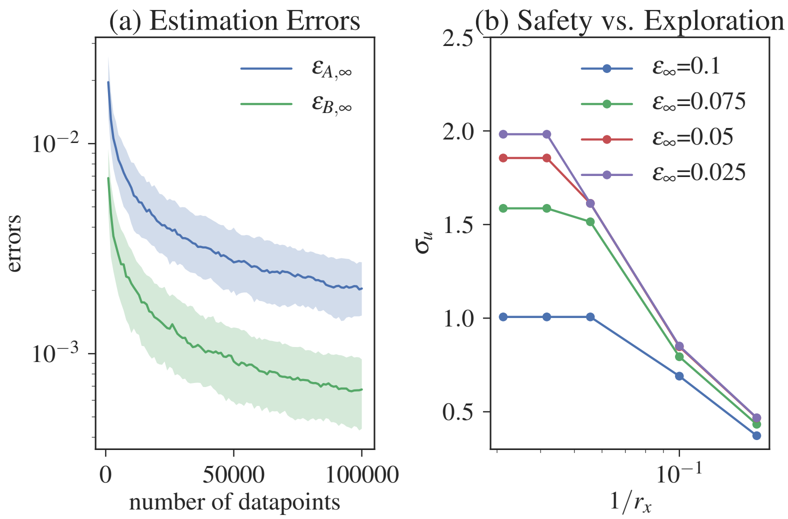

Figure 2: Over time, estimation errors decrease (a). As safety requirements increase, the maximum feasible excitation decreases (b).

safety and exploration by showing the largest value of σ_η for which the robust synthesis is feasible, given a size for the state constraint set r_x. Here, we leave $x_0 = 0$, and examine a variety of errors in the dynamics estimates. As the uncertainties in the dynamics decrease, higher levels of both safety and exploration are achievable.

7 Discussion

In this paper, we propose a method for learning unknown linear systems while ensuring that they satisfy state and input constraints. By synthesizing a controller that both excites and regulates the system, we address the trade-off between safety and exploration directly. We further derive an end-to-end finite sample bound on the performance of LQR controllers synthesized from collected data.

There are several directions for possible extensions of this work. To mitigate the conservativeness of the robust controller, tighter bounds on the uncertainty in the system response $\widehat{\mathbf{\Delta}}$ could be derived for structured settings, where more than just the norm of the error is known. To connect this work to experiment design literature, the objective in the synthesis problem (3.2) could be replaced with an exploration inspired cost function for the learning stage.

Alternatively, the constrained LQR problem could be cast in the setting of online learning, where one seeks to minimize cost at all times, including during learning. This would require an analysis of recursive feasibility, to understand the transition that occurs when controllers are updated based on refined system estimates. It would also likely require a direct quantification of performance loss when the robustness margin conditions are not satisfied. Finally, we remark that the exploration vs. safety trade-off is compelling for nonlinear systems.

ACKNOWLEDGMENT

We thank Francesco Borrelli and the members of the MPC Lab at UC Berkeley for their helpful comments and feedback. SD is supported by an NSF Graduate Research Fellowship under Grant No. DGE 1752814. ST is supported by a Google PhD fellowship. BR is generously supported in part by ONR awards N00014-17-1-2191, N00014-17-1-2401, and N00014-18-1-2833, the DARPA Assured Autonomy (FA8750-18-C-0101) and Lagrange (W911NF-16-1-0552) programs, and an Amazon AWS AI Research Award.

References

[1] S. Dean, H. Mania, N. Matni, B. Recht, and S. Tu, "On the sample complexity of the linear quadratic regulator," *arXiv preprint arXiv:1710.01688*, 2017.

[2] Y.-S. Wang, N. Matni, and J. C. Doyle, "A system level approach to controller synthesis," *arXiv preprint arXiv:1610.04815*, 2016.

[3] C.-N. Fiechter, "Pac adaptive control of linear systems," in *Conference on Learning Theory*, 1997.

[4] Y. Abbasi-Yadkori and C. Szepesvári, "Regret Bounds for the Adaptive Control of Linear Quadratic Systems," in *Conference on Learning Theory*, 2011.

[5] M. Abeille and A. Lazaric, "Thompson Sampling for Linear-Quadratic Control Problems," in *AISTATS*, 2017.

[6] M. Simchowitz, H. Mania, S. Tu, M. I. Jordan, and B. Recht, "Learning without mixing: Towards a sharp analysis of linear system identification," in *Conference on Learning Theory*, 2018.

[7] S. Dean, H. Mania, N. Matni, B. Recht, and S. Tu, "Regret bounds for robust adaptive control of the linear quadratic regulator," in *Neural Information Processing Systems*, 2018.

[8] A. Bemporad and M. Morari, "Robust model predictive control: A survey," in *Robustness in identification and control*. Springer, 1999, pp. 207–226.

[9] M. V. Kothare, V. Balakrishnan, and M. Morari, "Robust constrained model predictive control using linear matrix inequalities," *Automatica*, vol. 32, no. 10, pp. 1361–1379, 1996.

[10] D. Q. Mayne, M. M. Seron, and S. Raković, "Robust model predictive control of constrained linear systems with bounded disturbances," *Automatica*, vol. 41, no. 2, pp. 219–224, 2005.

[11] P. J. Goulart, E. C. Kerrigan, and J. M. Maciejowski, "Optimization over state feedback policies for robust control with constraints," *Automatica*, vol. 42, no. 4, pp. 523–533, 2006.

[12] C. A. Larsson, C. R. Rojas, and H. Hjalmarsson, "Mpc oriented experiment design," *IFAC Proceedings Volumes*, vol. 44, no. 1, pp. 9966–9971, 2011.

[13] A. Aswani, H. Gonzalez, S. S. Sastry, and C. Tomlin, "Provably safe and robust learning-based model predictive control," *Automatica*, vol. 49, no. 5, pp. 1216–1226, 2013.

[14] M. Tanaskovic, L. Fagiano, R. Smith, and M. Morari, "Adaptive receding horizon control for constrained mimo systems," *Automatica*, vol. 50, no. 12, pp. 3019–3029, 2014.

[15] F. Blanchini, "Set invariance in control," *Automatica*, vol. 35, no. 11, pp. 1747–1767, 1999.

[16] F. Berkenkamp and A. P. Schoellig, "Safe and Robust Learning Control with Gaussian Processes," in *European Control Conference (ECC)*, 2015.

[17] F. Berkenkamp, M. Turchetta, A. Schoellig, and A. Krause, "Safe Model-based Reinforcement Learning with Stability Guarantees," in *Neural Information Processing Systems*, 2017.

[18] G. Dalal, K. Dvijotham, M. Vecerik, T. Hester, C. Paduraru, and Y. Tassa, "Safe Exploration in Continuous Action Spaces," *arXiv preprint arXiv:1801.08757*, 2018.

[19] Y. Chow, O. Nachum, E. Duenez-Guzman, and M. Ghavamzadeh, "A Lyapunov-based Approach to Safe Reinforcement Learning," *arXiv preprint arXiv:1805.07708*, 2018.

[20] T. Koller, F. Berkenkamp, M. Turchetta, and A. Krause, "Learning-based model predictive control for safe exploration and reinforcement learning," *arXiv preprint arXiv:1803.08287*, 2018.

[21] K. P. Wabersich and M. N. Zeilinger, "Linear model predictive safety certification for learning-based control," *arXiv preprint arXiv:1803.08552*, 2018.

[22] T. Lu, M. Zinkevich, C. Boutilier, B. Roy, and D. Schuurmans, "Safe Exploration for Identifying Linear Systems via Robust Optimization," *arXiv preprint arXiv:1711.11165*, 2017.

[23] F. Oldewurtel, C. N. Jones, and M. Morari, "A tractable approximation of chance constrained stochastic mpc based on affine disturbance feedback," in *47th IEEE Annual Conference on Decision and Control*, 2008.

[24] M. Farina, L. Giulioni, and R. Scattolini, "Stochastic linear model predictive control with chance constraints–a review," *Journal of Process Control*, vol. 44, pp. 53–67, 2016.

[25] N. Matni, Y.-S. Wang, and J. Anderson, "Scalable system level synthesis for virtually localizable systems," in *IEEE 56th Annual Conference on Decision and Control*, 2017.

[26] B. Dumitrescu, *Positive trigonometric polynomials and signal processing applications.* Springer, 2007, vol. 103.

[27] J. Anderson and N. Matni, "Structured State Space Realizations for SLS Distributed Controllers," in *Allerton*, 2017.

[28] B. Efron, "Bootstrap methods: another look at the jackknife," in *Breakthroughs in statistics.* Springer, 1992, pp. 569–593.

[29] R. Ibragimov and S. Sharakhmetov, "The Exact Constant in the Rosenthal Inequality for Random Variables with Mean Zero," *Theory of Probability & Its Applications*, vol. 46, no. 1, pp. 127–132, 2002.

8 Learning Results

First, we prove a simple small-ball result for random variables with finite fourth moments.

Proposition 8.1. *Let $X \in \mathbb{R}$ be a zero-mean random variable with finite fourth moment, which satisfies the conditions*

$$\mathbb{E}[X^4] \le C(\mathbb{E}[X^2])^2 \,.$$

Let $a \in \mathbb{R}$ be a fixed scalar and $\theta \in (0,1)$. We have that

$$\mathbb{P}\{|a+X| \ge \sqrt{\theta \mathbb{E}[X^2]}\} \ge (1-\theta)^2 / \max\{4, 3C\} \,.$$

Proof. First, we note that we can assume $a \ge 0$ without loss of generality, since we can perform a change of variables $X \leftarrow -X$. We have that $\mathbb{E}[(a+X)^4] = a^4 + 6a^2\mathbb{E}[X^2] + 4a\mathbb{E}[X^3] + \mathbb{E}[X^4]$. Therefore, by the Paley-Zygmund inequality and Young's inequality, we have that

$$\begin{aligned}
\mathbb{P}\{|a+X| \ge \sqrt{\theta \mathbb{E}[X^2]}\} &\ge \mathbb{P}\{|a+X| \ge \sqrt{\theta \mathbb{E}[(a+X)^2]}\} \\
&= \mathbb{P}\{(a+X)^2 \ge \theta \mathbb{E}[(a+X)^2]\} \\
&\ge (1-\theta)^2 \frac{(\mathbb{E}[(a+X)^2])^2}{\mathbb{E}[(a+X)^4]} \\
&= (1-\theta)^2 \frac{a^4 + 2a^2\mathbb{E}[X^2] + (\mathbb{E}[X^2])^2}{a^4 + 8a^2\mathbb{E}[X^2] + 3\mathbb{E}[X^4]} \,.
\end{aligned}$$

Now define the function $f(a)$ for $a \geq 0$ as

$$f(a) = \frac{a^4 + 2a^2\mu + \mu^2}{a^4 + 8a^2\mu + 3\beta}\,.$$

Clearly $f(0) = \mu^2/(3\beta)$ and $f(\infty) = 1$. Since by Jensen's inequality we know that $\mu^2 \leq \beta$, this means $f(0) < f(\infty)$. On the other hand,

$$\frac{d}{da}f(a) = \frac{4a\,(a^2+\mu)\,(\mu\,(3a^2-4\mu)+3\beta)}{(a^4+8a^2\mu+3\beta)^2}\,.$$

Assume that $\mu \neq 0$ (otherwise the claim is trivially true). The only critical points of the function f on non-negative reals is at $a = 0$ and $a = \sqrt{\frac{4\mu^2-3\beta}{3\mu}}$ if $4\mu^2 \geq 3\beta$. If $4\mu^2 < 3\beta$, then $a = 0$ is the only critical point of f, so $f(a) \geq f(0) = \mu^2/(3\beta) \geq 1/(3C)$. On the other hand, if $4\mu^2 \geq 3\beta$ holds, Some algebra yields that

$$f(\sqrt{(4\mu^2-3\beta)/(3\mu)}) = \frac{7\mu^2-3\beta}{16\mu^2-3\beta} \geq \inf_{\gamma\in[3,4]} \frac{7-\gamma}{16-\gamma} = 1/4\,.$$

The inequality above holds since $3\beta \in [3\mu^2, 4\mu^2]$. Hence,

$$f(a) \geq \min\{1/4, 1/(3C)\} = 1/\max\{4, 3C\}\,.$$

□

This gives us the main new technical ingredient needed to prove Theorem 5.1.

Proposition 8.2. *Let $\mu \in \mathbb{R}^d$ and $M \in \mathbb{R}^{d\times d}$ be a full rank matrix. Let $w \in \mathbb{R}^d$ be a random vector such that its fourth moment is finite and each coordinate w_i is independent and zero-mean, i.e. $\mathbb{E}[w_i] = 0$. Then, for any fixed $v \in \mathbb{R}^d$,*

$$\mathbb{P}_w\left(|\langle v, \mu + Mw\rangle| \geq \sqrt{\lambda_{\min}(M\Sigma M^\top)/2}\right) \geq \frac{1}{C\cdot C_w}\,,$$

where $\Sigma = \mathbb{E}_w[ww^\top]$, C is an absolute constant, and $C_w = \max_{1\leq i\leq d}\mathbb{E}[w_i^4]/(\mathbb{E}[w_i^2])^2$. In particular, we can take $C = 192$.

Proof. Fix a $q \in \mathbb{R}^d$. By Rosenthal's inequality, we have that $\mathbb{E}[|\langle q, w\rangle|^4] \leq C\cdot C_w\mathbb{E}[|\langle q, w\rangle|^2]^2$ for an absolute constant–we can take $C = 4$ (see e.g. [29]). We now apply Proposition 8.1 with $a = \langle v, \mu\rangle$, $X = \langle M^\top v, w\rangle$, $C = 4C_w$, and $\theta = 1/2$. The claim now follows. □

With Proposition 8.2 in place, the proof of Theorem 5.1 is identical to that of Proposition C.1 in Dean et al. [7], except in the final step of establishing the block martingale condition: instead of using the small-ball calculation for Gaussian distributions we replace it with the estimate provided by Proposition 8.2.

9 Extended Sub-optimality Results and Proofs

This section contains proofs necessary for the sub-optimality results. It further develops a sub-optimality bound for the case of FIR truncation. We begin by considering the frequency response elements of composed transfer functions. First, define, for $\mathbf{D} = \sum_{k=1}^{\infty} D(k)z^{-k}$,

$$\mathrm{Toep}_k(D) = \begin{bmatrix} D(1) & & \\ \vdots & \ddots & \\ D(k) & \dots & D(1) \end{bmatrix} .$$

Then we have the following lemma.

Lemma 9.1. *For $\mathbf{M} = \sum_{k=1}^{\infty} M(k)z^{-k}$ and $\mathbf{D} = \sum_{k=1}^{\infty} D(k)z^{-k}$ the frequency response elements of $\tilde{\mathbf{M}} = \mathbf{MD}$ are given by*

$$\tilde{M}(k) = M[k:1]D(1:k) .$$

where we use the notation $D(1:k)$ for the vertical concatenation of $D(1)$ through $D(k)$. Further, we have

$$\tilde{M}[k:1] = M[k:1]\mathrm{Toep}_k(D) .$$

Proof. For the first statement, notice that we can write

$$\mathbf{MD} = \sum_{k=1}^{\infty}\sum_{t=1}^{k} M(t)D(k-t)z^{-k} = \sum_{k=1}^{\infty} M[k:1]D(1:k)z^{-k} .$$

Then for the second, we have

$$\begin{aligned}\tilde{M}[k:1] &= \begin{bmatrix} M[k:1]D(1:k) & \dots & M[1]D(1) \end{bmatrix} \\ &= M[k:1]\underbrace{\begin{bmatrix} D(1) & & \\ \vdots & \ddots & \\ D(k) & \dots & D(1) \end{bmatrix}}_{\mathrm{Toep}_k(D)} .\end{aligned}$$

□

9.1 Proof of Lemma 4.2

Proof of Lemma 4.2. The proposed feasible solution is

$$\tilde{\mathbf{\Phi}}_x = \mathbf{\Phi}_x^\star(I-\mathbf{\Delta})^{-1},\ \tilde{\mathbf{\Phi}}_u = \mathbf{K}_\star\mathbf{\Phi}_x^\star(I-\mathbf{\Delta})^{-1},\ \ \tilde{\gamma} = \frac{\sqrt{2}\zeta_2}{1-\sqrt{2}\zeta_2},\ \ \tilde{\tau} = \frac{\zeta_\infty}{1-\zeta_\infty} .$$

By construction, $\tilde{\mathbf{\Phi}}_x$ and $\tilde{\mathbf{\Phi}}_u$ satisfy the equality constraints. Considering the $\mathcal{H}_\infty$ norm constraint,

$$\begin{aligned}\sqrt{2}\left\|\begin{bmatrix}\varepsilon_{A,2}\tilde{\mathbf{\Phi}}_x\\ \varepsilon_{B,2}\tilde{\mathbf{\Phi}}_u\end{bmatrix}\right\|_{\mathcal{H}_\infty} &= \sqrt{2}\left\|\begin{bmatrix}\varepsilon_{A,2}I\\ \varepsilon_{B,2}\mathbf{K}_\star\end{bmatrix}\mathbf{\Phi}_x^\star(I-\mathbf{\Delta})^{-1}\right\|_{\mathcal{H}_\infty}\\ &\leq \sqrt{2}(\varepsilon_{A,2}+\varepsilon_{B,2}\|\mathbf{K}_\star\|_{\mathcal{H}_\infty})\|\mathbf{\Phi}_x^\star\|_{\mathcal{H}_\infty}\frac{1}{1-\|\mathbf{\Delta}\|_{\mathcal{H}_\infty}} \leq \sqrt{2}\zeta_2\frac{1}{1-\sqrt{2}\zeta_2} = \tilde{\gamma}\,.\end{aligned}$$

The decomposition is valid since $\sqrt{2}\zeta_2 < 1$ by assumption. Then consider the $\mathcal{L}_1$ norm constraint,

$$\left\|\begin{bmatrix}\varepsilon_{A,\infty}\tilde{\mathbf{\Phi}}_x\\ \varepsilon_{B,\infty}\tilde{\mathbf{\Phi}}_u\end{bmatrix}\right\|_{\mathcal{L}_1} \leq (\varepsilon_{A,\infty}+\varepsilon_{B,\infty}\|\mathbf{K}_\star\|_{\mathcal{L}_1})\|\mathbf{\Phi}_x^\star\|_{\mathcal{L}_1}\frac{1}{1-\|\mathbf{\Delta}\|_{\mathcal{L}_1}} \leq \zeta_\infty\frac{1}{1-\zeta_\infty} = \tilde{\tau}$$

where the decomposition of the inverse is valid by assumption on the size of ζ_∞.

Then it remains to show that the robust state and input constraints are satisfied. Recall that they are given by

$$G_x^\tau(\tilde{\mathbf{\Phi}}_x;k)_j = F_{x,j}^\top\tilde{\Phi}_x(k+1)x_0 + \sigma_w\|F_{x,j}^\top\tilde{\Phi}_x[k:1]\|_1 + \frac{\tau\sigma_w c_0}{1-\tau}\|F_{x,j}^\top\tilde{\Phi}_x[k+1:1]\|_1\,.$$

Note that $\tilde{\mathbf{\Phi}}_x = \mathbf{\Phi}_x^\star + \mathbf{\Phi}_x^\star\mathbf{\Delta}(I-\mathbf{\Delta})^{-1}$. Define the frequency response elements of $\mathbf{\Delta}(I-\mathbf{\Delta})^{-1}$ by $D(k)$. Then, using Lemma 9.1 we have

$$\tilde{\Phi}(k) = \Phi_x^\star(k) + \Phi_x^\star[k:1]D(1:k)\ ,\quad \tilde{\Phi}[k:1] = \Phi_x^\star[k:1] + \Phi_x^\star[k:1]\mathrm{Toep}_k(D)\,.$$

Lemma 9.2. *We have that for any vector v, and D representing the frequency response of $\mathbf{\Delta}(1+\mathbf{\Delta})^{-1}$,*

$$\|v^\top\mathrm{Toep}_k(D)\|_1 \leq \frac{\zeta_\infty}{1-\zeta_\infty}\|v\|_1,$$

as long as $\zeta_\infty < 1$.

Proof.

$$\begin{aligned}\|v^\top\mathrm{Toep}_k(D)\|_1 &= \|\mathrm{Toep}_k(D)^\top v\|_1 \leq \|\mathrm{Toep}_k(D)^\top\|_1\|v\|_1 = \|\mathrm{Toep}_k(D)\|_\infty\|v\|_1\\ &\leq \|\mathbf{\Delta}(1+\mathbf{\Delta})^{-1}\|_{\mathcal{L}_1}\|v\|_1 \leq \frac{\zeta_\infty}{1-\zeta_\infty}\|v\|_1\,.\end{aligned}$$

Above, we make use of the fact that the ℓ_1 and ℓ_∞ norms are duals, and therefore $\|A\|_\infty = \|A^\top\|_1$. The second inequality holds because $\mathrm{Toep}_k(D)$ is a truncation of the semi-infinite Toeplitz matrix associated with the operator $\mathbf{\Delta}(1+\mathbf{\Delta})^{-1}$. The final decomposition is valid because $\|\mathbf{\Delta}\|_{\mathcal{L}_1} \leq \zeta_\infty < 1$. $\square$

Now we are ready to consider the state constraint indexed by j and k,

$$\begin{aligned} G_x^{\tilde\tau}(\tilde{\mathbf{\Phi}}_x; k)_j &= F_{x,j}^\top(\Phi_x^\star(k+1) + \Phi_x^\star[k+1:1]D(1:k+1))x_0 \\ &+ \sigma_w \|F_{x,j}^\top \Phi_x^\star[k:1](I + \mathrm{Toep}_k(D))\|_1 \\ &+ \frac{\tilde\tau}{1-\tilde\tau}\sigma_w c_0 \|F_{x,j}^\top \Phi_x^\star[k+1:1](I + \mathrm{Toep}_{k+1}(D))\|_1 \end{aligned}$$

Considering each term individually,

$$\begin{aligned} &F_{x,j}^\top(\Phi_x^\star(k+1) + \Phi_x^\star[k+1:1]D(1:k+1))x_0 \\ &\qquad = F_{x,j}^\top \Phi_x^\star(k+1)x_0 + F_{x,j}^\top \Phi_x^\star[k+1:1]D(1:k+1)x_0 \, , \end{aligned}$$

Then defining E_1 to contain an identity in the first block and zeros elsewhere,

$$\begin{aligned} F_{x,j}^\top \Phi_x^\star[k+1:1]D(1:k+1)x_0 &\le \|F_{x,j}^\top \Phi_x^\star[k+1:1]\mathrm{Toep}_{k+1}(D)E_1\|_1 \|x_0\|_\infty \\ &\le \|F_{x,j}^\top \Phi_x^\star[k+1:1]\|_1 \frac{\zeta_\infty}{1-\zeta_\infty}\|x_0\|_\infty \\ &\le \|F_{x,j}^\top \Phi_x^\star[k+1:1]\|_1 2\zeta_\infty \|x_0\|_\infty \, , \end{aligned}$$

The first inequality is Hölder's inequality, the second by Lemma 9.2, and the last for $\zeta_\infty \le \frac{1}{2}$. Next, the second term:

$$\begin{aligned} \sigma_w \|F_{x,j}^\top \Phi_x^\star[k:1](I + \mathrm{Toep}_k(D))\|_1 &\le \sigma_w \left(1 + \frac{\zeta_\infty}{1-\zeta_\infty}\right) \|F_{x,j}^\top \Phi_x^\star[k:1]\|_1 \\ &\le \sigma_w(1 + 2\zeta_\infty)\|F_{x,j}^\top \Phi_x^\star[k:1]\|_1 \, , \end{aligned}$$

where the first inequality holds by Lemma 9.2 and the final for $\zeta_\infty \le \frac{1}{2}$. Finally, the last term,

$$\frac{\tilde\tau}{1-\tilde\tau}\sigma_w c_0 \|F_{x,j}^\top \Phi_x^\star[k+1:1](I + \mathrm{Toep}_{k+1}(D))\|_1 \le \frac{\tilde\tau}{1-\tilde\tau}\sigma_w c_0 \frac{1}{1-\zeta_\infty}\|F_{x,j}^\top \Phi_x^\star[k+1:1]\|_1$$

where we again apply Lemma 9.2. Focusing on the constant factors, and plugging in the definition of $\tilde\tau$

$$\begin{aligned} \frac{\tilde\tau}{1-\tilde\tau}\frac{1}{1-\zeta_\infty} &= \frac{\zeta_\infty}{(1-\zeta_\infty)(1-\frac{\zeta_\infty}{1-\zeta_\infty})}\frac{1}{1-\zeta_\infty} = \frac{\zeta_\infty}{1-2\zeta_\infty}\frac{1}{1-\zeta_\infty} \\ &\le 2\zeta_\infty(1+4\zeta_\infty) \le 6\zeta_\infty \, . \end{aligned}$$

where the first inequality follows for $2\zeta_\infty \le \frac{1}{2}$ and the second for $2\zeta_\infty \le 1$.

The resulting sum is

$$\begin{aligned} G_x^{\tilde\tau}(\tilde{\mathbf{\Phi}}_x; k)_j &\le G_x(\mathbf{\Phi}_x^\star; k)_j + 2\zeta_\infty \sigma_w \|F_{x,j}^\top \Phi_x^\star[k:1]\|_1 \\ &+ 2\zeta_\infty \|F_{x,j}^\top \Phi_x^\star[k+1:1]\|_1(\|x_0\|_\infty + 3\sigma_w c_0) \\ &\le G_x(\mathbf{\Phi}_x^\star; k)_j + 10\zeta_\infty \sigma_w c_0 \|F_{x,j}^\top \Phi_x^\star[k+1:1]\|_1 \le b_{x,j} \end{aligned}$$

where the final inequality follows from our assumption about the size of ζ_∞ satisfying the margin condition. A similar computation with the input constraints shows the same result. Therefore, the proposed solution is feasible. □

9.2 Finite Dimensional Sub-optimality

We can recover sub-optimality guarantees in the case that we optimize over only a finite set of system response variables. Define the truncated responses $\mathbf{\Phi}_x^L = \sum_{k=1}^L \Phi_x(t) z^{-k}$ and similarly for $\mathbf{\Phi}_u^L$, and notice that

$$\begin{bmatrix} zI - A & -B \end{bmatrix} \begin{bmatrix} \mathbf{\Phi}_x \\ \mathbf{\Phi}_u \end{bmatrix} = I \iff \begin{bmatrix} zI - A & -B \end{bmatrix} \begin{bmatrix} \mathbf{\Phi}_x^L \\ \mathbf{\Phi}_u^L \end{bmatrix} = I + \frac{1}{z^L} \Phi_x(L+1) \, . \tag{9.1}$$

This reformulation allows for the optimization over the FIR filters $\mathbf{\Phi}_x^L$ and $\mathbf{\Phi}_u^L$, plus an additional variable that represents the tail of the true response. Applying this observation to the robust synthesis problem (3.2) yields the following finite dimensional problem:

$$\begin{aligned} \min_{\mathbf{\Phi}_x^L, \mathbf{\Phi}_u^L, V} \ & \frac{1}{1-\gamma} J(\widehat{A}, \widehat{B}, \mathbf{K}) \\ \text{s.t.} \ & \begin{bmatrix} zI - \widehat{A} & -\widehat{B} \end{bmatrix} \begin{bmatrix} \mathbf{\Phi}_x^L \\ \mathbf{\Phi}_u^L \end{bmatrix} = I + \frac{1}{z^L} V, \\ & \sqrt{2} \left\| \begin{bmatrix} \varepsilon_{A,2} \mathbf{\Phi}_x^L \\ \varepsilon_{B,2} \mathbf{\Phi}_u^L \end{bmatrix} \right\|_{\mathcal{H}_\infty} + \|V\|_2 \le \gamma, \quad \left\| \begin{bmatrix} \varepsilon_{A,\infty} \mathbf{\Phi}_x^L \\ \varepsilon_{B,\infty} \mathbf{\Phi}_u^L \end{bmatrix} \right\|_{\mathcal{L}_1} + \|V\|_\infty \le \tau, \\ & G_x^\tau(\Phi_x^L; k) \le b_x, \quad G_u^\tau(\Phi_u^L; k) \le b_u \quad \forall \ k \, . \end{aligned} \tag{9.2}$$

We now show that so long as the horizon L is suitably large, all of the properties that hold for the solution of infinite horizon problem (3.2) carry over to the solution of the finite dimensional approximation (9.2).

Proposition 9.3. *Any controller designed from a feasible solution to the finite robust control problem (9.2) for any $0 \le \gamma, \tau < 1$ stabilizes the true system and ensures that state and input constraints will be satisfied.*

Proof. Consider the affine constraint on $\mathbf{\Phi}_x^L$ and $\mathbf{\Phi}_u^L$. We can equivalently write, using the observation in (9.1),

$$\begin{bmatrix} zI - \widehat{A} & -\widehat{B} \end{bmatrix} \begin{bmatrix} \mathbf{\Phi}_x^L \\ \mathbf{\Phi}_u^L \end{bmatrix} = I + \frac{1}{z^L} V \implies \begin{bmatrix} zI - A_\star & -B_\star \end{bmatrix} \begin{bmatrix} \mathbf{\Phi}_x^L \\ \mathbf{\Phi}_u^L \end{bmatrix} = I + \underbrace{\frac{1}{z^L} V + \widehat{\mathbf{\Delta}}}_{\widehat{\mathbf{\Delta}}_L}$$

where we define $\widehat{\mathbf{\Delta}} = \Delta_A \mathbf{\Phi}_x^L + \Delta_B \mathbf{\Phi}_u^L$. The noting that

$$\begin{aligned} \|\widehat{\mathbf{\Delta}}_L\|_{\mathcal{H}_\infty} &\le \sqrt{2} \left\| \begin{bmatrix} \varepsilon_{A,2} \mathbf{\Phi}_x \\ \varepsilon_{B,2} \mathbf{\Phi}_u \end{bmatrix} \right\|_{\mathcal{H}_\infty} + \|V\|_2 \ , \\ \|\widehat{\mathbf{\Delta}}_L\|_{\mathcal{L}_1} &\le \left\| \begin{bmatrix} \varepsilon_{A,\infty} \mathbf{\Phi}_x \\ \varepsilon_{B,\infty} \mathbf{\Phi}_u \end{bmatrix} \right\|_{\mathcal{L}_1} + \|V\|_\infty \ , \end{aligned}$$

by Proposition 3.2, the system is stabilized and the true system trajectory is given by

$$\mathbf{x} = \mathbf{\Phi}_x (I + \widehat{\mathbf{\Delta}}_L)^{-1} \mathbf{w}, \quad \mathbf{u} = \mathbf{\Phi}_u (I + \widehat{\mathbf{\Delta}}_L)^{-1} \mathbf{w} \, .$$

The rest of the proof follows exactly as in the proof of Theorem 3.3

□

To bound the sub-optimality of this robust synthesis, we define constants related to the decay rate of the system responses. Let $C_\star$ and $\rho_\star$ be any constants such that $\|\Phi_x^\star(t)\|_p \le C_\star \rho_\star^t$ for $p = 2, \infty$. Note that these constants must exist since the optimal closed loop will be exponentially stable. Let C_Δ and ρ_Δ similarly be any constants that bound the norm of the frequency response elements of $(I + \mathbf{\Delta})^{-1}$; these constants can be bounded by multiples of $C_\star$ and $\rho_\star$, as worked out in Appendix G of Dean et al. [7]. Then, set $C = \frac{C_\star C_\Delta}{\log(\rho / \max(\rho_\star, \rho_\Delta))}$ and $\rho = \frac{\max(\rho_\star, \rho_\Delta)+1}{2}$.

Additionally, we will have a similar condition on the robustness margin. This time, the minimization is over a finite set of indices. Define

$$\mathrm{margin}_x = \min_{\substack{0 \le j \le n_c \\ 0 \le k \le L}} \frac{b_{x,j} - G_x(\mathbf{\Phi}_x^\star; k)_j}{\sigma_w c_0 \|F_{x,j}^\top \Phi_x^\star[k+1:1]\|_1} ,$$

similarly for margin_u, and let $\mathrm{margin} = \min(\mathrm{margin}_x, \mathrm{margin}_u)$.

Theorem 9.4. *Fix a $C_{FIR} > 1$, and suppose that the truncation length is such that*

$$L \ge \frac{\log((1 - C_{FIR}^{-1})^{-1} C)}{\log(1/\rho)} - 1 .$$

As long as $\zeta_\infty \le \min\left(\frac{\mathrm{margin}}{2C_{FIR}(1+C_{FIR})} - (C_{FIR} - 1), \frac{1}{2(1+C_{FIR})}\right)$ and $\zeta_2 \le \frac{1}{2\sqrt{2}(1+C_{FIR})}$, we have that the cost achieved by $\widehat{\mathbf{K}}(L) = \widehat{\mathbf{\Phi}}_u \widehat{\mathbf{\Phi}}_x^{-1}$ synthesized from the minimizers of $\min_{\gamma,\tau}$ (9.2) satisfies

$$\frac{J(A_\star, B_\star, \widehat{\mathbf{K}}(L)) - J_\star}{J_\star} \le 2\sqrt{2} C_{FIR}(1 + C_{FIR})(\varepsilon_{A,2} + \varepsilon_{B,2}\|\mathbf{K}_\star\|_{\mathcal{H}_\infty})\|\mathbf{\Phi}_x^\star\|_{\mathcal{H}_\infty} + C_{FIR} - 1 .$$

The proof of Theorem 9.4 relies on the construction of a feasible solution to the optimization problem.

Lemma 9.5. *Under the assumptions of Theorem 9.4, the following is a feasible solution to (9.2):*

$$\tilde{\mathbf{\Phi}}_x = \mathbf{\Phi}_x^\star (I - \mathbf{\Delta})^{-1}[1:L], \quad \tilde{\mathbf{\Phi}}_u = \mathbf{K}_\star \mathbf{\Phi}_x^\star (I - \mathbf{\Delta})^{-1}[1:L], \quad \tilde{V} = \mathbf{\Phi}_x^\star (I - \mathbf{\Delta})^{-1}[L+1],$$

$$\tilde{\gamma} = \frac{\sqrt{2}\zeta_2}{1 - \sqrt{2}\zeta_2} + C\rho^{L+1}, \quad \tilde{\tau} = \frac{\zeta_\infty}{1 - \zeta_\infty} + C\rho^{L+1}.$$

Proof. First, we consider the affine constraint,

$$\begin{bmatrix} zI - \widehat{A} & -\widehat{B} \end{bmatrix} \begin{bmatrix} \tilde{\mathbf{\Phi}}_x \\ \tilde{\mathbf{\Phi}}_u \end{bmatrix} = I + \frac{1}{z^L} V .$$

Applying the observation in (9.1),

$$\begin{bmatrix} zI-\widehat{A} & -\widehat{B} \end{bmatrix} \begin{bmatrix} \tilde{\mathbf{\Phi}}_x \\ \tilde{\mathbf{\Phi}}_u \end{bmatrix} = I + \frac{1}{z^L}V \iff \begin{bmatrix} zI-\widehat{A} & -\widehat{B} \end{bmatrix} \begin{bmatrix} \mathbf{\Phi}_x^\star \\ \mathbf{\Phi}_u^\star \end{bmatrix} (I-\mathbf{\Delta})^{-1} = I$$
$$\iff \left(\begin{bmatrix} zI-A_\star & -B_\star \end{bmatrix} + \begin{bmatrix} \Delta_A & \Delta_B \end{bmatrix} \right) \begin{bmatrix} \mathbf{\Phi}_x^\star \\ \mathbf{\Phi}_u^\star \end{bmatrix} (I-\mathbf{\Delta})^{-1} = I$$

The right hand side of this equation is true by the definition of $\mathbf{\Delta}$ and the affine constraint on $\mathbf{\Phi}_x^\star$ and $\mathbf{\Phi}_u^\star$, and thus the affine constraint on $\tilde{\mathbf{\Phi}}_x$ and $\tilde{\mathbf{\Phi}}_u$ is satisfied.

Next, we bound the norm of $\tilde{V}$. Using Lemma 9.1 and the definition of constants, we have that

$$\begin{aligned} \|\mathbf{\Phi}_x^\star(I-\mathbf{\Delta})^{-1}[k]\|_p &\leq \sum_{t=1}^{k} \|\Phi_x^\star(t)\|_p \|D(k-t)\|_p \\ &\leq \sum_{t=1}^{k} C_\star \rho_\star^t C_\Delta \rho_\Delta^{k-t} \leq C_\star C_\Delta k \max(\rho_\star, \rho_\Delta)^k \,, \end{aligned}$$

where $D(t)$ are the frequency response elements of $(I+\mathbf{\Delta})^{-1}$. Then noting that for any $x \in (0,1)$, $\alpha > 0$, and $k \geq 1$,

$$\log(kx^k) = \log(k) + k\log(x) \leq \frac{1}{\alpha}k + \log(\alpha) + k\log(x),$$

which is true because $\log\left(\frac{k}{\alpha}\right) \leq \frac{k}{\alpha}$, we have for $p = 2, \infty$

$$\|V\|_p \leq C_\star C_\Delta \alpha \left(e^{1/\alpha}\max(\rho_\star,\rho_\Delta)\right)^{L+1} \leq C\rho^{L+1} \,.$$

The final inequality follows from choosing $\alpha = \log\left(\frac{\max(\rho_\star,\rho_\Delta)+1}{2\max(\rho_\star,\rho_\Delta)}\right)^{-1}$.

Next, we consider the $\mathcal{H}_\infty$ norm constraint

$$\begin{aligned} \sqrt{2}\left\| \begin{bmatrix} \varepsilon_{A,2}\tilde{\mathbf{\Phi}}_x \\ \varepsilon_{B,2}\tilde{\mathbf{\Phi}}_u \end{bmatrix} \right\|_{\mathcal{H}_\infty} &\leq \sqrt{2}(\varepsilon_{A,2} + \varepsilon_{B,2}\|\mathbf{K}_\star\|_{\mathcal{H}_\infty}) \|\mathbf{\Phi}_x^\star(I-\mathbf{\Delta})^{-1}\|_{\mathcal{H}_\infty} \\ &\leq \sqrt{2}(\varepsilon_{A,2} + \varepsilon_{B,2}\|\mathbf{K}_\star\|_{\mathcal{H}_\infty}) \|\mathbf{\Phi}_x^\star\|_{\mathcal{H}_\infty} \frac{1}{1-\|\mathbf{\Delta}\|_{\mathcal{H}_\infty}} \\ &\leq \sqrt{2}\zeta_2 \frac{1}{1-\sqrt{2}\zeta_2} = \tilde{\gamma} - C\rho^{L+1} \leq \tilde{\gamma} - \|V\|_2 \,. \end{aligned}$$

The first inequality follows because the proposed $\tilde{\mathbf{\Phi}}_x$ is a truncation of $\mathbf{\Phi}_x^\star(I-\mathbf{\Delta})^{-1}$ and similarly for $\tilde{\mathbf{\Phi}}_u$. The second follow because $\|\mathbf{\Delta}\|_{\mathcal{H}_\infty} \leq \zeta_2 < 1$. Then similarly considering the $\mathcal{L}_1$ norm constraint,

$$\left\| \begin{bmatrix} \varepsilon_{A,\infty}\tilde{\mathbf{\Phi}}_x \\ \varepsilon_{B,\infty}\tilde{\mathbf{\Phi}}_u \end{bmatrix} \right\|_{\mathcal{L}_1} \begin{aligned} &\leq (\varepsilon_{A,\infty} + \varepsilon_{B,\infty}\|\mathbf{K}_\star\|_{\mathcal{L}_1})\|\mathbf{\Phi}_x^\star(I-\mathbf{\Delta})^{-1}\|_{\mathcal{L}_1} \\ &\leq (\varepsilon_{A,\infty} + \varepsilon_{B,\infty}\|\mathbf{K}_\star\|_{\mathcal{L}_1})\|\mathbf{\Phi}_x^\star\|_{\mathcal{L}_1}\frac{1}{1-\|\mathbf{\Delta}\|_{\mathcal{L}_1}} \\ &\leq \zeta_\infty \frac{1}{1-\zeta_\infty} = \tilde{\tau} - C\rho^{L+1} \leq \tilde{\tau} - \|V\|_\infty . \end{aligned}$$

Then it remains only to show that the robust state and input constraints are satisfied. Notice that $\tilde{\mathbf{\Phi}}_x$ is a truncation of $\mathbf{\Phi}_x^\star(1+\mathbf{\Delta})^{-1}$. For $k > L$, the frequency response elements are zero, and therefore the constraints are trivially satisfied. For for $0 \leq k \leq L$, we have, by Lemma 9.1, using block-Toeplitz notation, that

$$\tilde{\Phi}(k) = \Phi_x^\star(k) + \Phi_x^\star[k:1]D(1:k)$$

where we now take $D(k)$ represent the frequency response elements of $\mathbf{\Delta}(1+\mathbf{\Delta})^{-1}$. Then the proof of constraint satisfaction follows as in Section 9.1, up to the computations with the definition of $\tilde{\tau}$

$$\begin{aligned} \frac{\tilde{\tau}}{1-\tilde{\tau}}\frac{1}{1-\zeta_\infty} &= \frac{\frac{\zeta_\infty}{(1-\zeta_\infty)} + C\rho^{L+1}}{(1-\frac{\zeta_\infty}{1-\zeta_\infty} - C\rho^{L+1})}\frac{1}{1-\zeta_\infty} \\ &= \frac{\frac{\zeta_\infty}{1-\zeta_\infty} + C\rho^{L+1}}{1-C\rho^{L+1} - (2-C\rho^{L+1})\zeta_\infty} \leq \frac{2\zeta_\infty + C\rho^{L+1}}{1-C\rho^{L+1} - (2-C\rho^{L+1})\zeta_\infty} \\ &= C_{FIR}\frac{2\zeta_\infty + C\rho^{L+1}}{1-(1+C_{FIR})\zeta_\infty} = \frac{2C_{FIR}\zeta_\infty}{1-(1+C_{FIR})\zeta_\infty} + \frac{C_{FIR}-1}{1-(1+C_{FIR})\zeta_\infty} \\ &\leq 4C_{FIR}\zeta_\infty + (C_{FIR}-1)(1+2(1+C_{FIR})\zeta_\infty) \end{aligned}$$

where the first inequality holds for $\zeta_\infty \leq \frac{1}{2}$ and the second for $\zeta_\infty \leq \frac{1}{2(1+C_{FIR})}$. We use the fact that our assumption on L gives $C_{FIR} \geq \frac{1}{1-C\rho^{L+1}}$. Then

$$4C_{FIR}\zeta_\infty + (C_{FIR}-1)(1+2(1+C_{FIR})\zeta_\infty) \leq \underbrace{2(C_{FIR}+C_{FIR}^2-1)}_{c_1}\zeta_\infty + \underbrace{(C_{FIR}-1)}_{c_2} .$$

The resulting sum is

$$\begin{aligned} G_x^{\tilde{\tau}}(\tilde{\mathbf{\Phi}}_x;k)_j &\leq G_x(\mathbf{\Phi}_x^\star;k)_j + 2\zeta_\infty\sigma_w\|F_{x,j}^\top\Phi_x^\star[k:1]\|_1 \\ &\quad + \zeta_\infty\|F_{x,j}^\top\Phi_x^\star[k+1:1]\|_1(2\|x_0\|_\infty + c_1\sigma_w c_0) + \|F_{x,j}^\top\Phi_x^\star[k+1:1]\|_1 c_2\sigma_w c_0 \\ &\leq G_x(\mathbf{\Phi}_x^\star;k)_j + (2+c_1)\zeta_\infty\sigma_w c_0\|F_{x,j}^\top\Phi_x^\star[k+1:1]\|_1 \\ &\quad + \|F_{x,j}^\top\Phi_x^\star[k+1:1]\|_1 c_2\sigma_w c_0 \\ &\leq b_{x,j} \end{aligned}$$

where the final inequality follows from our assumption about the size of ζ_∞,

$$2C_{FIR}(1+C_{FIR})\zeta_\infty \le \frac{b_{x,j} - G_x(\mathbf{\Phi}_x^\star; k)_j}{\sigma_w c_0 \|F_{x,j}^\top \Phi_x^\star[k+1:1]\|_1} - C_{FIR} + 1\,.$$

A similar computation with the input constraints shows the same result. Therefore, the proposed solution is feasible. $\square$

We are now ready to prove the main sub-optimality result.

Proof of Theorem 9.4. Recall that we denote the minimizers of $\min_{\gamma,\tau}$ (9.2) as $(\widehat{\mathbf{\Phi}}_x, \widehat{\mathbf{\Phi}}_u, \widehat{V}, \widehat{\gamma}, \widehat{\tau})$. Let

$$\widehat{\mathbf{\Delta}}_L := \widehat{\mathbf{\Delta}} + \frac{1}{z^L}\widehat{V}.$$

Then we have that that

$$\|\widehat{\mathbf{\Delta}}_L\|_{\mathcal{H}_\infty} \le \sqrt{2}\left\|\begin{bmatrix}\varepsilon_{A,2}\widehat{\mathbf{\Phi}}_x \\ \varepsilon_{B,2}\widehat{\mathbf{\Phi}}_u\end{bmatrix}\right\|_{\mathcal{H}_\infty} + \|\widehat{V}\|_2 \le \widehat{\gamma},$$

By the constraints of the optimization problem in (9.2).

We now apply observation 9.1 and Proposition 3.2 with $\widehat{\mathbf{\Delta}}_L$ to characterize the response achieved by the FIR approximate controller $\widehat{\mathbf{K}}(L)$ on the true system $(A_\star, B_\star)$, giving the following:

$$\begin{aligned} J(A_\star, B_\star, \widehat{\mathbf{K}}(L)) &= \left\|\begin{bmatrix} Q^{\frac{1}{2}} & \\ & R^{\frac{1}{2}}\end{bmatrix}\begin{bmatrix}\widehat{\mathbf{\Phi}}_x \\ \widehat{\mathbf{\Phi}}_u\end{bmatrix}(I+\widehat{\mathbf{\Delta}}_L)^{-1}\right\|_{\mathcal{H}_2} \\ &\le \frac{1}{1-\widehat{\gamma}}\left\|\begin{bmatrix} Q^{\frac{1}{2}} & \\ & R^{\frac{1}{2}}\end{bmatrix}\begin{bmatrix}\widehat{\mathbf{\Phi}}_x \\ \widehat{\mathbf{\Phi}}_u\end{bmatrix}\right\|_{\mathcal{H}_2}. \end{aligned}$$

The inequality follows from $\|\widehat{\mathbf{\Delta}}\|_{\mathcal{H}_\infty} \le \widehat{\gamma} < 1$.

Denote by $(\tilde{\mathbf{\Phi}}_x, \tilde{\mathbf{\Phi}}_u, \tilde{V}, \tilde{\gamma}, \tilde{\tau})$ the feasible solution constructed in Lemma 9.5. Then,

$$\begin{aligned} \frac{1}{1-\widehat{\gamma}}\left\|\begin{bmatrix} Q^{\frac{1}{2}} & \\ & R^{\frac{1}{2}}\end{bmatrix}\begin{bmatrix}\widehat{\mathbf{\Phi}}_x \\ \widehat{\mathbf{\Phi}}_u\end{bmatrix}\right\|_{\mathcal{H}_2} &\le \frac{1}{1-\tilde{\gamma}}\left\|\begin{bmatrix} Q^{\frac{1}{2}} & \\ & R^{\frac{1}{2}}\end{bmatrix}\begin{bmatrix}\tilde{\mathbf{\Phi}}_x \\ \tilde{\mathbf{\Phi}}_x\end{bmatrix}\right\|_{\mathcal{H}_2} = \frac{1}{1-\tilde{\gamma}} J_L(\widehat{A}, \widehat{B}, \mathbf{K}_\star) \\ &\le \frac{1}{1-\tilde{\gamma}} J(\widehat{A}, \widehat{B}, \mathbf{K}_\star) \le \frac{1}{1-\tilde{\gamma}}\frac{1}{1-\|\mathbf{\Delta}\|_2} J(A_\star, B_\star, \mathbf{K}_\star), \end{aligned}$$

where the first inequality follows from the optimality of $(\widehat{\mathbf{\Phi}}_x, \widehat{\mathbf{\Phi}}_u, \widehat{V}, \widehat{\gamma})$, the equality and second inequality from the fact that $(\tilde{\mathbf{\Phi}}_x, \tilde{\mathbf{\Phi}}_u)$ are truncations of the response of $\mathbf{K}_\star$ on $(\widehat{A}, \widehat{B})$ to the first L time steps. The second to last inequality follows from an application of

Proposition 3.2 with the roles of the nominal and true systems switched. We therefore have, bounding $\|\mathbf{\Delta}\|_2$ by $\sqrt{2}\zeta$,

$$
\begin{aligned}
J(A_\star, B_\star, \widehat{\mathbf{K}}(L)) &\le \frac{1}{1-\tilde{\gamma}} \frac{1}{1-\sqrt{2}\zeta} J_\star = \frac{1}{1 - C\rho^{L+1} - (2 - C\rho^{L+1})\sqrt{2}\zeta} J_\star \\
&= C_{FIR} \left(1 + \frac{\sqrt{2}(1 + C_{FIR})\zeta_2}{1 - \sqrt{2}(1 + C_{FIR})\zeta_2} \right) J_* \\
&\le C_{FIR} \left(1 + 2\sqrt{2}(1 + C_{FIR})\zeta_2 \right) J_*
\end{aligned}
$$

which follows for $\zeta_2 \le \frac{1}{2\sqrt{2}(1+C_{FIR})}$. Then finally,

$$
\frac{J(A_\star, B_\star, \widehat{\mathbf{K}}(L)) - J_\star}{J_\star} \le C_{FIR} - 1 + 2\sqrt{2}C_{FIR}(1 + C_{FIR})\zeta_2 \,.
$$

$\square$