Certainty Equivalence is Efficient for Linear Quadratic Control

Horia Mania
University of California, Berkeley
hmania@berkeley.edu

Stephen Tu
University of California, Berkeley
stephentu@berkeley.edu

Benjamin Recht
University of California, Berkeley
brecht@berkeley.edu

Abstract

We study the performance of the *certainty equivalent controller* on Linear Quadratic
(LQ) control problems with unknown transition dynamics. We show that for both
the fully and partially observed settings, the sub-optimality gap between the cost
incurred by playing the certainty equivalent controller on the true system and the
cost incurred by using the optimal LQ controller enjoys a fast statistical rate, scaling
as the *square* of the parameter error. To the best of our knowledge, our result is the
first sub-optimality guarantee in the partially observed Linear Quadratic Gaussian
(LQG) setting. Furthermore, in the fully observed Linear Quadratic Regulator
(LQR), our result improves upon recent work by Dean et al. [11], who present an
algorithm achieving a sub-optimality gap linear in the parameter error. A key part
of our analysis relies on perturbation bounds for discrete Riccati equations. We
provide two new perturbation bounds, one that expands on an existing result from
Konstantinov et al. [25], and another based on a new elementary proof strategy.

1 Introduction

One of the most straightforward methods for controlling a dynamical system with unknown transitions
is based on the *certainty equivalence principle*: a model of the system is fit by observing its time
evolution, and a control policy is then designed by treating the fitted model as the truth [6]. Despite
the simplicity of this method, it is challenging to guarantee its efficiency because small modeling
errors may propagate to large, undesirable behaviors on long time horizons. As a result, most work
on controlling systems with unknown dynamics has explicitly incorporated robustness against model
uncertainty [11, 12, 23, 30, 41, 42].

In this work, we show that for the standard baseline of controlling an unknown linear dynamical
system with a quadratic objective function known as Linear Quadratic (LQ) control, certainty
equivalent control synthesis achieves *better* cost than prior methods that account for model
uncertainty. Our results hold for both the fully observed Linear Quadratic Regulator (LQR) and the partially
observed Linear Quadratic Gaussian (LQG) setting. For offline control, where one collects some
data and then designs a fixed control policy to be run on an infinite time horizon, we show that the
gap between the performance of the certainty equivalent controller and the optimal control policy
scales *quadratically* with the error in the model parameters for both LQR and LQG. To the best of
our knowledge, we provide the first sub-optimality guarantee for LQG. Moreover, in the LQR setting
our work improves upon the recent result of Dean et al. [11], who present an algorithm that achieves
a sub-optimality gap linear in the parameter error. In the case of online LQR control, where one

adaptively improves the control policy as new data comes in, our offline result implies that a simple, polynomial time algorithm using ε-greedy exploration suffices for nearly optimal $\widetilde{\mathcal{O}}(\sqrt{T})$ regret.

2 Main Results for the Linear Quadratic Regulator

An instance of the linear quadratic regulator (LQR) is defined by four matrices: two matrices $A_\star \in \mathbb{R}^{n \times n}$ and $B_\star \in \mathbb{R}^{n \times d}$ that define the linear dynamics and two positive semidefinite matrices $Q \in \mathbb{R}^{n \times n}$ and $R \in \mathbb{R}^{d \times d}$ that define the cost function. Given these matrices, the goal of LQR is to solve the optimization problem

$$\min_{\mathbf{u}_0, \mathbf{u}_1, \dots} \lim_{T \to \infty} \mathbb{E}\left[\frac{1}{T} \sum_{t=0}^{T} \mathbf{x}_t^\top Q \mathbf{x}_t + \mathbf{u}_t^\top R \mathbf{u}_t\right] \text{ s.t. } \mathbf{x}_{t+1} = A_\star \mathbf{x}_t + B_\star \mathbf{u}_t + \mathbf{w}_t, \tag{1}$$

where $\mathbf{x}_t$, $\mathbf{u}_t$ and $\mathbf{w}_t$ denote the state, input (or action), and noise at time t, respectively. The expectation is over the initial state $\mathbf{x}_0 \sim \mathcal{N}(0, I_n)$ and the i.i.d. noise $\mathbf{w}_t \sim \mathcal{N}(0, \sigma_w^2 I_n)$. When the problem parameters $(A_\star, B_\star, Q, R)$ are known the optimal policy is given by linear feedback, $\mathbf{u}_t = K_\star \mathbf{x}_t$, where $K_\star = -(R + B_\star^\top P_\star B_\star)^{-1} B_\star^\top P_\star A_\star$ where $P_\star$ is the (positive definite) solution to the discrete Riccati equation

$$P_\star = A_\star^\top P_\star A_\star - A_\star^\top P_\star B_\star (R + B_\star^\top P_\star B_\star)^{-1} B_\star^\top P_\star A_\star + Q \tag{2}$$

and can be computed efficiently [4, see e.g.]. Problem (1) considers an average cost over an infinite horizon. The optimal controller for the finite horizon variant is also static and linear, but time-varying. The LQR solution in this case can be computed efficiently via dynamic programming.

In this work we are interested in the control of a linear dynamical system with unknown transition parameters $(A_\star, B_\star)$ based on estimates $(\widehat{A}, \widehat{B})$. The cost matrices Q and R are assumed known. We analyze the *certainty equivalence approach*: use the estimates $(\widehat{A}, \widehat{B})$ to solve the optimization problem (1) while disregarding the modeling error, and use the resulting controller on the true system $(A_\star, B_\star)$. We interchangeably refer to the resulting policy as the *certainty equivalent controller* or, following Dean et al. [11], the *nominal controller*. We denote by $\widehat{P}$ the solution to the Riccati equation (2) associated with the parameters $(\widehat{A}, \widehat{B})$ and let $\widehat{K}$ be the corresponding controller. We denote by $J(A, B, K)$ the cost (1) obtained by using the actions $\mathbf{u}_t = K\mathbf{x}_t$ on the system (A, B), and we use $\widehat{J}$ and $J_\star$ to denote $J(A_\star, B_\star, \widehat{K})$ and $J(A_\star, B_\star, K_\star)$, respectively.

Let $\varepsilon \geq 0$ such that $\|A_\star - \widehat{A}\| \leq \varepsilon$ and $\|B_\star - \widehat{B}\| \leq \varepsilon$. (Here and throughout this work we use $\|\cdot\|$ to denote the Euclidean norm for vectors as well as the spectral (operator) norm for matrices.) Dean et al. [11] introduced a robust controller that achieves $\widehat{J} - J_\star \leq C_1(A_\star, B_\star, Q, R)\varepsilon$ for some complexity term $C_1(A_\star, B_\star, Q, R)$ that depends on the problem parameters. We show that the nominal controller $\mathbf{u}_t = \widehat{K}\mathbf{x}_t$ achieves $\widehat{J} - J_\star \leq C_2(A_\star, B_\star, Q, R)\varepsilon^2$. Both results require ε to be sufficiently small (as a function of the problem parameters) and it is important to note that ε must be much smaller for the nominal controller to be guaranteed to stabilize the system than for the robust controller proposed by Dean et al. [11]. However, our result shows that once the estimation error ε is small enough, the nominal controller performs better: the sub-optimality gap scales as $\mathcal{O}(\varepsilon^2)$ versus $\mathcal{O}(\varepsilon)$. Both the more stringent requirement on ε and better performance of nominal control compared to robust control, when the estimation error is sufficiently small, were observed empirically by Dean et al. [11].

Before we can formally state our result we need to introduce a few more concepts and assumptions. It is common to assume that the cost matrices Q and R are positive definite. Under an additional observability assumption, this condition can be relaxed to Q being positive semidefinite.

Assumption 1. *The cost matrices Q and R are positive definite. Since scaling both Q and R does not change the optimal controller $K_\star$, we can assume without loss of generality that $\underline{\sigma}(R) \geq 1$, where $\underline{\sigma}(\cdot)$ denotes the minimum singular value.*

A square matrix M is *stable* if its spectral radius $\rho(M)$ is (strictly) smaller than one. Recall that the spectral radius is defined as $\rho(M) = \max\{|\lambda| : \lambda \text{ is an eigenvalue of } M\}$. A linear dynamical system (A, B) in feedback with K is fully described by the *closed loop matrix* $A + BK$. More precisely, in this case $\mathbf{x}_{t+1} = (A + BK)\mathbf{x}_t + \mathbf{w}_t$. For a static linear controller $\mathbf{u}_t = K\mathbf{x}_t$ to achieve finite LQR cost it is necessary and sufficient that the closed loop matrix is stable.

In order to quantify the growth or decay of powers of a square matrix M, we define

$$\tau(M, \rho) := \sup \left\{ \|M^k\| \rho^{-k} : k \geq 0 \right\}. \tag{3}$$

In other words, $\tau(M, \rho)$ is the smallest value such that $\|M^k\| \leq \tau(M, \rho)\rho^k$ for all $k \geq 0$. We note that $\tau(M, \rho)$ might be infinite, depending on the value of ρ, and it is always greater or equal than one. If ρ is larger than $\rho(M)$, we are guaranteed to have a finite $\tau(M, \rho)$ (this is a consequence of Gelfand's formula). In particular, if M is a stable matrix, we can choose $\rho < 1$ such that $\tau(M, \rho)$ is finite. Also, we note that $\tau(M, \rho)$ is a decreasing function of ρ; if $\rho \geq \|M\|$, we have $\tau(M, \rho) = 1$. At a high level, the quantity $\tau(M, \rho)$ measures the degree of transient response of the linear system $\mathbf{x}_{t+1} = M\mathbf{x}_t + \mathbf{w}_t$. In particular, when M is stable, $\tau(M, \rho)$ can be upper bounded by the $\mathcal{H}_\infty$-norm of the system defined by M, which is the ℓ_2 to ℓ_2 operator norm of the system and a fundamental quantity in robust control [see 40, for more details].

Throughout this work we use the quantities $\Gamma_\star := 1 + \max\{\|A_\star\|, \|B_\star\|, \|P_\star\|, \|K_\star\|\}$ and $L_\star := A_\star + B_\star K_\star$. We use $\Gamma_\star$ as a uniform upper bound on the spectral norms of the relevant matrices for the sake of algebraic simplicity. We are ready to state our meta theorem. The proofs for all the results can be found in the full version of the paper [28].

Theorem 1. *Suppose $d \leq n$. Let $\gamma > 0$ such that $\rho(L_\star) \leq \gamma < 1$. Also, let $\varepsilon > 0$ such that $\|\widehat{A} - A_\star\| \leq \varepsilon$ and $\|\widehat{B} - B_\star\| \leq \varepsilon$ and assume $\|\widehat{P} - P_\star\| \leq f(\varepsilon)$ for some function f such that $f(\varepsilon) \geq \varepsilon$. Then, under Assumption 1 the certainty equivalent controller $\mathbf{u}_t = \widehat{K}\mathbf{x}_t$ achieves*

$$\widehat{J} - J_\star \leq 200\, \sigma_w^2\, d\, \Gamma_\star^9\, \frac{\tau(L_\star, \gamma)^2}{1 - \gamma^2} f(\varepsilon)^2, \tag{4}$$

as long as $f(\varepsilon)$ is small enough so that the right hand side is smaller than σ_w^2.

In Section 4 we present two upper bounds $f(\varepsilon)$ on $\|\widehat{P} - P_\star\|$: one based on a proof technique proposed by Konstantinov et al. [25] and one based on our direct approach. Both of these upper bounds satisfy $f(\varepsilon) = \mathcal{O}(\varepsilon)$ for ε sufficiently small. For simplicity, in this section we only specialize our meta-theorem (Theorem 1) using the perturbation result from our direct approach.

To state a specialization of Theorem 1 we need a few more concepts. A linear system (A, B) is called *controllable* when the *controllability matrix* $\begin{bmatrix} B & AB & A^2B & \ldots & A^{n-1}B \end{bmatrix}$ has full row rank. Controllability is a fundamental concept in control theory; it states that there exists a sequence of inputs to the system (A, B) that moves it from any starting state to any final state in at most n steps. In this work we quantify how controllable a linear system is. We denote, for any integer $\ell \geq 1$, the matrix $\mathcal{C}_\ell := \begin{bmatrix} B & AB & \ldots & A^{\ell-1}B \end{bmatrix}$ and call the system (ℓ, ν)-*controllable* if the n-th singular value of $\mathcal{C}_\ell$ is greater or equal than ν, i.e. $\underline{\sigma}(\mathcal{C}_\ell) = \sqrt{\lambda_{\min}\left(\mathcal{C}_\ell \mathcal{C}_\ell^\top\right)} \geq \nu$. Intuitively, the larger ν is, the less control effort is needed to move the system between two different states.

Assumption 2. *We assume the unknown system $(A_\star, B_\star)$ is (ℓ, ν)-controllable, with $\nu > 0$.*

Assumption 2 was used in a different context by Cohen et al. [9]. For any controllable system and any $\ell \geq n$ there exists $\nu > 0$ such that the system is (ℓ, ν)-controllable. Therefore, (ℓ, ν)-controllability is really not much stronger of an assumption than controllability. As ℓ grows minimum singular value $\underline{\sigma}(\mathcal{C}_\ell)$ also grows and therefore a larger ν can be chosen so that the system is still (ℓ, ν) controllable.

Note that controllability is not necessary for LQR to have a well-defined solution: the weaker requirement is that of *stabilizability*, in which there exists a feedback matrix K so that $A_\star + B_\star K$ is stable. The result of Dean et al. [11] only requires stabilizability. While our upper bound on $\|\widehat{P} - P_\star\|$ requires controllability, the result of Konstantinov et al. [25] only requires stabilizability. However, our upper bound on $\|\widehat{P} - P_\star\|$ is sharper for some classes of systems (see Section 4). A direct plug in of our perturbation result, presented in Section 4, into Theorem 1 yields the following guarantee.

Theorem 2. *Suppose that $d \leq n$. Let ρ and γ be two real values such that $\rho(A_\star) \leq \rho$ and $\rho(L_\star) \leq \gamma < 1$. Also, let $\varepsilon > 0$ such that $\|\widehat{A} - A_\star\| \leq \varepsilon$ and $\|\widehat{B} - B_\star\| \leq \varepsilon$ and define $\beta = \max\{1, \varepsilon\tau(A_\star, \rho) + \rho\}$. Under Assumptions 1 and 2, the certainty equivalent controller $\mathbf{u}_t = \widehat{K}\mathbf{x}_t$ satisfies the suboptimality gap*

$$\widehat{J} - J_\star \leq \mathcal{O}(1)\, \sigma_w^2\, d\, \ell^5\, \Gamma_\star^{15}\, \tau(A_\star, \rho)^6 \beta^{4(\ell-1)} \frac{\tau(L_\star, \gamma)^2}{1 - \gamma^2} \frac{\max\{\|Q\|^2, \|R\|^2\}}{\min\left\{\underline{\sigma}(Q)^2, \underline{\sigma}(R)^2\right\}} \left(1 + \frac{1}{\nu}\right)^2 \varepsilon^2, \tag{5}$$

as long as the right hand side is smaller than σ_w^2. Here, $\mathcal{O}(1)$ denotes a universal constant.

The exact form of Equation 5, such as the polynomial dependence on $\ell, \Gamma_\star$, etc, can be improved at the expense of conciseness of the expression. In our proof we optimized for the latter. The factor $\max\{\|Q\|^2, \|R\|^2\} / \min\{\underline{\sigma}(Q)^2, \underline{\sigma}(R)^2\}$ is the squared condition number of the cost function, a natural quantity in the context of the optimization problem (1), which can be seen as an infinite dimensional quadratic program with a linear constraint. The term $\frac{\tau(L_\star, \gamma)^2}{1-\gamma^2}$ quantifies the rate at which the optimal controller drives the state towards zero. Generally speaking, the less stable the optimal closed loop system is, the larger this term becomes.

An interesting trade-off arises between the factor $\ell^5 \beta^{4(\ell-1)}$ (which arises from upper bounding perturbations of powers of $A_\star$ on a time interval of length ℓ) and the factor ν (the lower bound on $\underline{\sigma}(\mathcal{C}_\ell)$), which is increasing in ℓ. Hence, the parameter ℓ should be seen as a free-parameter that can be tuned to minimize the right hand side of (5). Now, we specialize Theorem 2 to a few cases.

Case: $A_\star$ is contractive, i.e. $\|A_\star\| < 1$. In this case, we can choose $\rho = \|A_\star\|$ and ε small enough so that $\varepsilon \leq 1 - \|A_\star\|$. Then, (5) simplifies to:

$$\widehat{J} - J_\star \leq \mathcal{O}(1)\, d\, \sigma_w^2\, \ell^5\, \Gamma_\star^{15} \frac{\tau(L_\star, \gamma)^2}{1-\gamma^2} \frac{\max\{\|Q\|^2, \|R\|^2\}}{\min\{\underline{\sigma}(Q)^2, \underline{\sigma}(R)^2\}} \left(1 + \frac{1}{\nu}\right)^2 \varepsilon^2 \,.$$

Case: $B_\star$ has rank n. In this case, we can choose $\ell = 1$. Then, (5) simplifies to:

$$\widehat{J} - J_\star \leq \mathcal{O}(1)\, d\, \sigma_w^2\, \Gamma_\star^{15} \tau(A_\star, \rho)^6 \frac{\tau(L_\star, \gamma)^2}{1-\gamma^2} \frac{\max\{\|Q\|^2, \|R\|^2\}}{\min\{\underline{\sigma}(Q)^2, \underline{\sigma}(R)^2\}} \left(1 + \frac{1}{\nu}\right)^2 \varepsilon^2 \,.$$

2.1 Comparison to Theorem 4.1 of Dean et al. [11].

Dean et al. [11] show that when their robust synthesis procedure is run with estimates $(\widehat{A}, \widehat{B})$ satisfying $\max\{\|\widehat{A} - A_\star\|, \|\widehat{B} - B_\star\|\} \leq \varepsilon \leq [5(1 + \|K_\star\|)\Psi_\star]^{-1}$, the resulting controller satisfies:

$$\widehat{J} - J_\star \leq 10(1 + \|K_\star\|)\Psi_\star J_\star \varepsilon + \mathcal{O}(\varepsilon^2) \,. \tag{6}$$

Here, the quantity $\Psi_\star := \sup_{z \in \mathbb{T}} \|(zI_n - L_\star)^{-1}\|$ is the $\mathcal{H}_\infty$-norm of the optimal closed loop system $L_\star$. In order to compare Equation 6 to Equation 5, we upper bound the quantity $\Psi_\star$ in terms of $\tau(L_\star, \gamma)$ and γ. In particular, by a infinite series expansion of the inverse $(zI_n - L_\star)^{-1}$ we can show $\Psi_\star \leq \frac{\tau(L_\star, \gamma)}{1-\gamma}$. Also, we have $J_\star = \sigma_w^2 \operatorname{tr}(P_\star) \leq \sigma_w^2 n \Gamma_\star$. Therefore, Equation 6 gives us that:

$$\widehat{J} - J_\star \leq \mathcal{O}(1) n \sigma_w^2 \Gamma_\star^2 \frac{\tau(L_\star, \gamma)}{1-\gamma} \varepsilon + \mathcal{O}(\varepsilon^2) \,.$$

We see that the dependence on the parameters $\Gamma_\star$ and $\tau(L_\star, \gamma)$ is significantly milder compared to Equation 5. Furthermore, this upper bound is valid for larger ε than the upper bound given in Theorem 2. Comparing these upper bound suggests that there is a price to pay for obtaining a fast rate, and that in regimes of moderate uncertainty (moderate size of ε), being robust to model uncertainty is important. This observation is supported by the empirical results of Dean et al. [11].

A similar trade-off between slow and fast rates arises in the setting of first-order convex stochastic optimization. The convergence rate $\mathcal{O}(1/\sqrt{T})$ of the stochastic gradient descent method can be improved to $\mathcal{O}(1/T)$ under a strong convexity assumption. However, the performance of stochastic gradient descent, which can achieve a $\mathcal{O}(1/T)$ rate, is sensitive to poorly estimated problem parameters [29]. Similarly, in the case of LQR, the nominal controller achieves a fast rate, but it is much more sensitive to estimation error than the robust controller of Dean et al. [11].

End-to-end guarantees. Theorem 2 can be combined with finite sample learning guarantees (e.g. [11, 15, 33, 34]) to obtain an end-to-end guarantee similar to Proposition 1.2 of Dean et al. [11]. In general, estimating the transition parameters from N samples yields an estimation error that scales as $\mathcal{O}(1/\sqrt{N})$. Therefore, Theorem 2 implies that $\widehat{J} - J_\star \leq \mathcal{O}(1/N)$ instead of the $\widehat{J} - J_\star \leq \mathcal{O}(1/\sqrt{N})$ rate from Proposition 1.2 of Dean et al. [11]. This is similar to the case of linear regression, where $\mathcal{O}(1/\sqrt{N})$ estimation error for the parameters translates to a $\mathcal{O}(1/N)$ *fast rate* for prediction error. Furthermore, Simchowitz et al. [34] and Sarkar and Rakhlin [33] showed

that faster estimation rates are possible for some linear dynamical systems. Theorem 2 translates such rates into control suboptimality guarantees in a transparent way.

Our result explains the behavior observed in Figure 4 of Dean et al. [11]. The authors propose two procedures for synthesizing robust controllers for LQR with unknown transitions: one which guarantees robustness of the performance gap $\widehat{J} - J_\star$, and one which only guarantees the stability of the closed loop system. Dean et al. [11] observed that the latter performs better in the small estimation error regime, which happens because the robustness constraint of the synthesis procedure becomes inactive when the estimation error is small enough. Then, the second robust synthesis procedure effectively outputs the certainty equivalent controller, which we now know to achieve a fast rate.

2.2 Nearly optimal $\widetilde{\mathcal{O}}(\sqrt{T})$ regret in the adaptive setting

The regret formulation of adaptive LQR was first proposed by Abbasi-Yadkori and Szepesvári [1]. The task is to design an adaptive algorithm $\{\mathbf{u}_t\}_{t \geq 0}$ to minimize regret, as defined by $\mathsf{Regret}(T) := \sum_{t=1}^{T} \mathbf{x}_t^\top Q \mathbf{x}_t + \mathbf{u}_t^\top R \mathbf{u}_t - T J_\star$. Abbasi-Yadkori and Szepesvári [1] study the performance of optimism in the face of uncertainty (OFU) and show that it has $\widetilde{\mathcal{O}}(\sqrt{T})$ regret, which is nearly optimal for this problem formulation. However, the OFU algorithm requires repeated solutions to a non-convex optimization problem for which no known efficient algorithm exists.

To deal with the computational issues of OFU, Dean et al. [12] propose to analyze the behavior of ε-greedy exploration using the suboptimality gap results of Dean et al. [11]. In the context of continuous control, ε-greedy exploration refers to the application of the control law $\mathbf{u}_t = \pi(\mathbf{x}_t, \mathbf{x}_{t-1}, ..., \mathbf{x}_0) + \eta_t$ with $\eta_t \sim \mathcal{N}(0, \sigma_{\eta,t}^2 I_d)$, where π is the policy, updated in epochs, and $\sigma_{\eta,t}^2$ is the variance of the exploration noise. Dean et al. [12] set the variance of the exploration noise as $\sigma_{\eta,t}^2 \sim t^{-1/3}$, and show that their method achieves $\widetilde{\mathcal{O}}(T^{2/3})$ regret. They use epochs of size 2^i and decompose the regret roughly as $\mathsf{Regret}(T) = \mathcal{O}\left(T(\widehat{J} - J_\star) + T\sigma_{\eta,T}^2\right)$. Since the estimation error of the model parameters scales as $\mathcal{O}((\sigma_{\eta,T}\sqrt{T})^{-1})$, and since the suboptimality gap $\widehat{J} - J_\star$ of the robust controller is linear in the estimation error, we have $\mathsf{Regret}(T) = \mathcal{O}\left(\frac{\sqrt{T}}{\sigma_{\eta,T}} + T\sigma_{\eta,T}^2\right)$. Then, setting $\sigma_{\eta,t}^2 \sim t^{-1/3}$ balances these two terms and yields $\widetilde{\mathcal{O}}(T^{2/3})$ regret. However, Theorem 2, which states that the gap $\widehat{J} - J_\star$ for the nominal controller depends quadratically on the estimation rate, implies that online certainty equivalent control achieves $\mathsf{Regret}(T) = \mathcal{O}\left(\frac{1}{\sigma_{\eta,T}^2} + T\sigma_{\eta,T}^2\right)$. Here, the optimal variance of the exploration noise scales as $\sigma_{\eta,t}^2 \sim t^{-1/2}$, yielding $\widetilde{\mathcal{O}}(\sqrt{T})$ regret. We note that the observation that certainty equivalence coupled with ε-greedy exploration achieves $\widetilde{\mathcal{O}}(\sqrt{T})$ regret was first made by Faradonbeh et al. [16].

Corollary 1. *(Informal) ε-greedy exploration with exploration schedule $\sigma_{\eta,t}^2 \sim t^{-1/2}$ combined with certainty equivalent control yields an adaptive LQR algorithm with regret bounded as $\widetilde{\mathcal{O}}(\sqrt{T})$.*

3 Main Results for the Linear Quadratic Gaussian Problem

Now we consider partially observable systems. In this case the system dynamics have the form:

$$\mathbf{x}_{t+1} = A_\star \mathbf{x}_t + B_\star \mathbf{u}_t + \mathbf{w}_t , \quad \mathbf{w}_t \sim \mathcal{N}(0, \sigma_w^2 I) , \tag{7a}$$

$$\mathbf{y}_t = C_\star \mathbf{x}_t + \mathbf{v}_t , \quad \mathbf{v}_t \sim \mathcal{N}(0, \sigma_v^2 I) . \tag{7b}$$

In (7), only the output process $\mathbf{y}_t$ is observed. The LQG problem is defined as[1]:

$$\min_{\mathbf{u}_0, \mathbf{u}_1, \ldots} \lim_{T \to \infty} \mathbb{E}\left[\frac{1}{T}\sum_{t=0}^{T} \mathbf{y}_t^\top Q \mathbf{y}_t + \mathbf{u}_t^\top R \mathbf{u}_t\right] \quad \text{s.t. } (7a), (7b) . \tag{8}$$

---

[1]Note that many texts define the LQG cost in terms of $\mathbf{x}_t^\top Q \mathbf{x}_t$ instead of $\mathbf{y}_t^\top Q \mathbf{y}_t$. We choose the latter because we do not want the cost to be tied to a particular (unknown) state representation.

Here, the input $\mathbf{u}_t$ is allowed to depend on the history[2] $\mathcal{H}_t := (\mathbf{u}_0, ..., \mathbf{u}_{t-1}, \mathbf{y}_0, ..., \mathbf{y}_{t-1})$. The optimal solution to (8) is to set $\mathbf{u}_t = K_\star \widehat{\mathbf{x}}_t$, with $K_\star$ the optimal LQR solution to $(A_\star, B_\star, C_\star^\mathsf{T} Q C_\star, R)$ and $\widehat{\mathbf{x}}_t := \mathbb{E}[\mathbf{x}_t | \mathcal{H}_t]$. The MSE estimate $\widehat{\mathbf{x}}_t$ can be solved efficiently via Kalman filtering:

$$\widehat{\mathbf{x}}_{t+1} = A_\star \widehat{\mathbf{x}}_t + B_\star \mathbf{u}_t + L_\star (\mathbf{y}_t - C_\star \widehat{\mathbf{x}}_t), \tag{9a}$$

$$L_\star = -A_\star \Sigma_\star C_\star^\mathsf{T} (C_\star \Sigma_\star C_\star^\mathsf{T} + V)^{-1}, \tag{9b}$$

$$\Sigma_\star = A_\star \Sigma_\star A_\star^\mathsf{T} + \sigma_w^2 I - A_\star \Sigma_\star C_\star^\mathsf{T} (C_\star \Sigma_\star C_\star^\mathsf{T} + \sigma_v^2 I)^{-1} C_\star \Sigma_\star A_\star^\mathsf{T}. \tag{9c}$$

There is an inherent ambiguity in the dynamics (7a)-(7b) which makes LQG more delicate than LQR. In particular, for any invertible T, the LQG problem (8) with parameters $(A_\star, B_\star, C_\star, Q, R)$ is equivalent to the LQG problem with parameters $(T A_\star T^{-1}, T B_\star, C_\star T^{-1}, Q, R)$ and appropriately rescaled noise processes. To deal with this ambiguity, we assume that we have estimates $(\widehat{A}, \widehat{B}, \widehat{C}, \widehat{L})$ such that there exists an unitary T such that:

$$\max\{\|\widehat{A} - T A_\star T^{-1}\|, \|\widehat{B} - T B_\star\|, \|\widehat{C} - C_\star T^{-1}\|, \|\widehat{L} - T L_\star\|\} \le \varepsilon. \tag{10}$$

Recent work [32, 35, 38] has shown how to obtain this style of estimates with guarantees from input/output data. As in Section 2, we assume that the cost matrices (Q, R) are known. Then, we study the performance of the certainty equivalence controller defined by:

$$\widehat{\mathbf{x}}_{t+1} = \widehat{A} \widehat{\mathbf{x}}_t + \widehat{B} \mathbf{u}_t + \widehat{L}(\mathbf{y}_t - \widehat{C} \widehat{\mathbf{x}}_t), \quad \mathbf{u}_t = \widehat{K} \widehat{\mathbf{x}}_t, \quad \widehat{K} = \mathsf{LQR}(\widehat{A}, \widehat{B}, \widehat{C}^\mathsf{T} Q \widehat{C}, R). \tag{11}$$

Similarly to Theorem 1 for LQR, we state a meta theorem for LQG. Unlike Theorem 1, however, we need a stronger type of Riccati perturbation guarantee which also allows for perturbation of the Q matrix. Specifically, we suppose there exists γ_0 such that for any $\gamma \le \gamma_0$ and $(\widehat{A}, \widehat{B}, \widehat{Q})$ with $\max\{\|\widehat{A} - A\|, \|\widehat{B} - B\|, \|\widehat{Q} - Q\|\} \le \gamma$ the solutions P and $\widehat{P}$ of the Riccati equations with parameters (A, B, Q, R) and $(\widehat{A}, \widehat{B}, \widehat{Q}, R)$ satisfy

$$\|P - \widehat{P}\| \le f(\gamma), \tag{12}$$

for an increasing function f with $f(\gamma) \ge \gamma$. The constant γ_0 and function f are allowed to depend on the parameters (A, B, Q, R). In Section 4, we present a perturbation bound (Proposition 1) that satisfies these properties. Similarly to Section 2, we define $\Gamma_\star := 1 + \max\{\|A_\star\|, \|B_\star\|, \|C_\star\|, \|K_\star\|, \|L_\star\|, \|P_\star\|\}$. The following theorem is our main result for LQG.

Theorem 3. *Suppose that $(A_\star, B_\star)$ is stabilizable, $(C_\star, A_\star)$ is observable, and that Assumption 1 holds. Let ε be an upper bound on $\|\widehat{A} - T A_\star T^{-1}\|$, $\|\widehat{B} - T B_\star\|$, $\|\widehat{C} - C_\star T^{-1}\|$, and $\|\widehat{L} - T L_\star\|$ for some unitary transformation T. Suppose that assumption (12) holds with parameters $(T A_\star T^{-1}, T B_\star, T^{-\mathsf{T}} C_\star^\mathsf{T} Q C_\star T^{-1}, R)$ and that ε is sufficiently small so that $3\|C_\star\|_+ \|Q\|_+ \varepsilon \le \gamma_0$ and $\overline{\varepsilon} \le 1$, where $\overline{\varepsilon} := \frac{7 \Gamma_\star^3}{\underline{\sigma}(R)} f(3\|C_\star\|_+^2 \|Q\|_+ \varepsilon)$. Let $\widehat{K}$ be defined as in (11), and define $N_\star$ as*

$$N_\star := \begin{bmatrix} A_\star + B_\star K_\star & B_\star K_\star \\ 0 & A_\star - L_\star C_\star \end{bmatrix}, \tag{13}$$

where the pair $(K_\star, L_\star)$ is optimal for the LQG problem defined by $(A_\star, B_\star, C_\star, Q, R)$. Let $\gamma > 0$ be such that $\rho(N_\star) < \gamma < 1$. Then as long as $\overline{\varepsilon} \le \frac{1-\gamma}{20 \Gamma_\star \tau(N_\star, \gamma)}$, the interconnection of (11) with (7) using $(\widehat{A}, \widehat{B}, \widehat{C}, \widehat{K}, \widehat{L})$ is stable. Furthermore, the cost $J(\widehat{A}, \widehat{B}, \widehat{C}, \widehat{K}, \widehat{L})$ satisfies:

$$J(\widehat{A}, \widehat{B}, \widehat{C}, \widehat{K}, \widehat{L}) - J_\star \le \mathcal{O}(1) \max\{\sigma_w^2, \sigma_v^2\}(\mathrm{tr}(C_\star^\mathsf{T} Q C_\star) + \mathrm{tr}(R)) \frac{\tau^6(N_\star, \gamma)}{(1 - \gamma^2)^3} \Gamma_\star^6 \overline{\varepsilon}^2.$$

The proof of Theorem 3 appears in Appendix F. We note that such a γ exists since $\rho(N_\star) < 1$; by the stability and observability assumptions in Theorem 3, we have that both $A_\star + B_\star K_\star$ and $A_\star - L_\star C_\star$ are stable (c.f. Appendix E of Kailath et al. [24]). Theorem 3 is a meta-theorem showing how perturbation bounds on the solution of Riccati equations translates into suboptimality bounds on the performance of certainty equivalent control. Combining Theorem 3 with Proposition 1, we have the following explicit result, an analogue of Theorem 2 for LQG. To simplify notation we denote by $\mathsf{dare}(A, B, Q, R)$ the solution to the discrete algebraic Riccati equation defined by the parameters A, B, Q, and R.

---

[2] The one step delay in $\mathbf{y}_t$ is a standard assumption in controls which slightly simplifies the Kalman filtering expressions. Our results generalize to the setting where the history also contains the current observation $\mathbf{y}_t$.

Theorem 4. *Suppose that $(A_\star, B_\star)$ is stabilizable, $(C_\star, A_\star)$ is observable, and that Assumption 1 holds. Let ε be an upper bound on $\|\widehat{A} - TA_\star T^{-1}\|$, $\|\widehat{B} - TB_\star\|$, $\|\widehat{C} - C_\star T^{-1}\|$, and $\|\widehat{L} - TL_\star\|$ for some unitary transformation T. Let $P_\star = \mathsf{dare}(A_\star, B_\star, C_\star^\mathsf{T} Q C_\star, R)$ and suppose that $\underline{\sigma}(P_\star) \geq 1$. Let $N_\star$ be as in (13) and fix γ such that $\rho(N_\star) < \gamma < 1$. As long as ε satisfies $\varepsilon \leq \frac{(1-\gamma^2)^2}{\tau^4(N_\star, \gamma)} \frac{1}{\Gamma_\star^{11} \|Q\|}$, we have the following sub-optimality bound:*

$$J(\widehat{A}, \widehat{B}, \widehat{C}, \widehat{K}, \widehat{L}) - J_\star \leq \mathcal{O}(1) \max\{\sigma_w^2, \sigma_v^2\}(\mathrm{tr}(C_\star^\mathsf{T} Q C_\star) + \mathrm{tr}(R)) \frac{\|Q\|^2}{\underline{\sigma}(R)^2} \Gamma_\star^{26} \frac{\tau^{10}(N_\star, \gamma)}{(1-\gamma^2)^5} \varepsilon^2 .$$

Several remarks are in order. First, the assumption that $\underline{\sigma}(P_\star) \geq 1$ is without loss of generality, since we can always rescale Q and R without affecting the control solution. Next, we compare our results here to a classic result from Doyle [13], which states that there are no gain margins for LQG. We remark that the notion of a gain margin is a robustness property that holds uniformly over a class of perturbations of varying degree. Our results do not hold uniformly; we use quantities such as $\tau(N_\star, \gamma)$ and $\Gamma_\star$ to quantify how much mismatch a given LQG instance can tolerate.

4   Riccati Perturbation Theory

As discussed in Sections 2 and 3, a key piece of our analysis is bounding the solutions to discrete Riccati equations as we perturb the problem parameters. Specifically, we are interested in quantities b, L such that $\|\widehat{P} - P_\star\| \leq L\varepsilon$ if $\varepsilon < b$, where ε represents a bound on the perturbation. We note that it is not possible to find universal values b, L. Consider the systems $(A_\star, B_\star) = (1, \varepsilon)$ and $(\widehat{A}, \widehat{B}) = (1, 0)$; the latter system is not stabilizable and hence $\widehat{P}$ does not even exist. Therefore, b and L must depend on the system parameters.

While there is a long line of work analyzing perturbations of Riccati equations, we are not aware of any result that offers explicit and easily interpretable b and L for a fixed $(A_\star, B_\star, Q, R)$; see Konstantinov et al. [26] for an overview of this literature. In this section, we present two new results for Riccati perturbation which offer interpretable bounds. The first one expands upon the operator-theoretic proof of Konstantinov et al. [25]; its proof can be found in Appendix B.1. In this result we assume the cost matrix Q can also be perturbed, which is needed for our LQG guarantee. In order to be consistent we denote the true cost by function by $Q_\star$ and the estimated one by $\widehat{Q}$.

Proposition 1. *Let $\gamma \geq \rho(L_\star)$ and also let ε such that $\|\widehat{A} - A_\star\|$, $\|\widehat{B} - B_\star\|$, and $\|\widehat{Q} - Q_\star\|$ are at most ε. Let $\|\cdot\|_+ = \|\cdot\| + 1$. We assume that $R \succ 0$, $(A_\star, B_\star)$ is stabilizable, $(Q^{1/2}, A_\star)$ observable, and $\underline{\sigma}(P_\star) \geq 1$.*

$$\|\widehat{P} - P_\star\| \leq \mathcal{O}(1)\, \varepsilon \, \frac{\tau(L_\star, \gamma)^2}{1 - \gamma^2} \|A_\star\|_+^2 \|P_\star\|_+^2 \|B_\star\|_+ \|R^{-1}\|_+,$$

as long as

$$\varepsilon \leq \mathcal{O}(1) \frac{(1-\gamma^2)^2}{\tau(L_\star, \gamma)^4} \|A_\star\|_+^{-2} \|P_\star\|_+^{-2} \|B_\star\|_+^{-3} \|R^{-1}\|_+^{-2} \min\left\{ \|L_\star\|_+^{-2}, \|P_\star\|_+^{-1} \right\}.$$

We note that the assumption $\underline{\sigma}(P_\star) \geq 1$ can be made without loss of generality when the other assumptions are satisfied. Since $R \succ 0$ and $(Q^{1/2}, A)$ observable, the value function matrix $P_\star$ is guaranteed to be positive definite. Then, by rescaling Q and R we can ensure that $\underline{\sigma}(P_\star) \geq 1$.

We now present our direct approach, which uses Assumption 2 to give a bound which is sharper for some systems $(A_\star, B_\star)$ then the one provided by Proposition 1. Recall that any controllable system is always (ℓ, ν)-controllable for some ℓ and ν.

Proposition 2. *Let $\rho \geq \rho(A_\star)$ and also let $\varepsilon \geq 0$ such that $\|\widehat{A} - A_\star\| \leq \varepsilon$ and $\|\widehat{B} - B_\star\| \leq \varepsilon$. Let $\beta := \max\{1, \varepsilon\tau(A_\star, \rho) + \rho\}$. Under Assumptions 1 and 2 we have*

$$\|\widehat{P} - P_\star\| \leq 32\, \varepsilon\, \ell^{\frac{5}{2}} \tau(A_\star, \rho)^3 \beta^{2(\ell-1)} \left(1 + \frac{1}{\nu}\right)(1 + \|B_\star\|)^2 \|P_\star\| \frac{\max\{\|Q\|, \|R\|\}}{\min\{\underline{\sigma}(R), \underline{\sigma}(Q)\}},$$

as long as ε is small enough so that the right hand side is smaller or equal than one.

The proof of this result is deferred to Appendix B.2. We note that Proposition 2 can also be extended to handle perturbations in the cost matrix Q. Proposition 2 requires an (ℓ, ν)-controllable system $(A_\star, B_\star)$, whereas Proposition 1 only requires a stabilizable system, which is a milder assumption. However, Proposition 2 can offer a sharper guarantee. For example, consider the linear system with two dimensional states ($n = 2$) given by $A_\star = 1.01 \cdot I_2$ and $B_\star = \begin{bmatrix} 1 & 0 \\ 0 & \beta \end{bmatrix}$. Both Q and R are chosen to be the identity matrix I_2. This system $(A_\star, B_\star)$ is readily checked to be $(1, \beta)$-controllable. It is also straightforward to verify that as β tends to zero, Proposition 1 gives a bound of $\|\widehat{P} - P_\star\| = \mathcal{O}(\varepsilon/\beta^4)$, whereas Proposition 2 gives a sharper bound of $\|\widehat{P} - P_\star\| = \mathcal{O}(\varepsilon/\beta^3)$.

5 Related Work

For the offline LQR batch setting, Fiechter [18] proved that the sub-optimality gap $\widehat{J} - J_\star$ scales as $\mathcal{O}(\varepsilon)$ for certainty equivalent control. A crucial assumption of his analysis is that the nominal controller stabilizes the true unknown system. We give bounds on when this assumption is valid. Recently, Dean et al. [11] proposed a robust controller synthesis procedure which takes model uncertainty into account and whose suboptimality gap scales as $\mathcal{O}(\varepsilon)$. Tu and Recht [39] show that the gap $\widehat{J} - J_\star$ of certainty equivalent control scales asymptotically as $\mathcal{O}(\varepsilon^2)$; we provide a non-asymptotic analogue of this result. Fazel et al. [17] and Malik et al. [27] analyze a model-free approach to policy optimization for LQR, in which the controller is directly optimized from sampled rollouts. Malik et al. [27] showed that, after collecting N rollouts, a derivative free method achieves a discounted cost gap that scales as $\mathcal{O}(1/\sqrt{N})$ or $\mathcal{O}(1/N)$, depending on the oracle model used.

In the online LQR adaptive setting it is well understood that using the certainty equivalence principle without adequate exploration can result in a lack of parameter convergence [see e.g. 5]. Abbasi-Yadkori and Szepesvári [1] showed that optimism in the face of uncertainty (OFU), when applied to online LQR, yields $\widetilde{\mathcal{O}}(\sqrt{T})$ regret. Faradonbeh et al. [14] removed some un-necessary assumptions of the previous analysis. Ibrahimi et al. [22] showed that when the underlying system is sparse, the dimension dependent constants in the regret bound can be improved. The main issue with OFU for LQR is that there are no known computationally tractable ways of implementing it. In order to deal with this, both Dean et al. [12] and Abbasi-Yadkori et al. [2] propose polynomial time algorithms for adaptive LQR based on ε-greedy exploration which achieve $\widetilde{\mathcal{O}}(T^{2/3})$ regret. Only recently progress has been made on offering $\widetilde{\mathcal{O}}(\sqrt{T})$ regret guarantees for computationally tractable algorithms. Abeille and Lazaric [3] show that Thompson sampling achieves $\widetilde{\mathcal{O}}(\sqrt{T})$ (frequentist) regret for the case when the state and inputs are both scalars. In a Bayesian setting Ouyang et al. [31] showed that Thompson sampling achieves $\widetilde{\mathcal{O}}(\sqrt{T})$ *expected* regret. Faradonbeh et al. [16] argue that certainty equivalence control with an epsilon-greedy-like scheme achieves $\widetilde{\mathcal{O}}(\sqrt{T})$ regret, though their work does not provide any explicit dependencies on instance parameters. Finally, Cohen et al. [10] also give an efficient algorithm based on semidefinite programming that achieves $\widetilde{\mathcal{O}}(\sqrt{T})$ regret. Their main result requires the initial parameter error to scale as $\mathcal{O}(1/T^{1/4})$. While they propose a $\mathcal{O}(\sqrt{T})$ length warmup period to get around this, our analysis of ε-greedy control does not require $o_T(1)$ accuracy of the initial parameters. Moreover, there are specialized algorithms for solving Riccati equations that are more efficient than general semidefinite programming solvers.

The literature for LQG is less complete, with most of the focus on the estimation side. Hardt et al. [19] show that gradient descent can be used to learn a model with good predictive performance, under strong technical assumptions on the A matrix. A line of work [20, 21] has focused on using spectral filtering techniques to learn a predictive model with low regret. Beyond predictive performance, several works [32, 35, 38] show how to learn the system dynamics up to a similarity transform from input/output data. Finally, we remark that Boczar et al. [8] give sub-optimality guarantees for output-feedback of a single-input-single-output (SISO) linear system with no process noise.

A key part of our analysis involves bounding the perturbation of solutions to the discrete algebraic Riccati equation. While there is a rich line of work studying perturbations of Riccati equations [25, 26, 36, 37], the results in the literature are either asymptotic in nature or difficult to use and interpret. We clarify the operator-theoretic result of Konstantinov et al. [25] and provide an explicit upper bound on the perturbation based on their proof strategy. Also, we take a new direct approach

and use an extended notion of controllability to give a constructive and simpler result. While the result of Konstantinov et al. [25] applies more generally to systems that are stabilizable, we give examples of linear systems for which our new perturbation result is tighter.

Finally, while we focus on a continuous control problem, we note that the performance of certainty equivalence had been studied in the context of tabular MDPs, e.g. Azar et al. [7] derived matching upper and lower bounds on the performance of value-iteration and policy-iteration that use estimated transition probabilities.

6 Conclusion

Though a naïve Taylor expansion suggests that the fast rates we derive here must be achievable, precisely computing such rates has been open since the 80s. All of the pieces we used here have existed in the literature for some time, and perhaps it has just required a bit of time to align contemporary rate-analyses in learning theory with earlier operator theoretic work in optimal control. There remain many possible extensions to this work. The robust control approach of Dean et al. [11] applies to many different objective functions besides quadratic costs, such as $\mathcal{H}_\infty$ and $\mathcal{L}_1$ control. It would be interesting to know whether fast rates for control are possible for other objective functions. Finally, determining the optimal minimax rate for both LQR and LQG would allow us to understand the tradeoffs between nominal and robust control at a more fine grained level.

References

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

[2] Y. Abbasi-Yadkori, N. Lazić, and C. Szepesvári. Model-Free Linear Quadratic Control via Reduction to Expert Prediction. In *AISTATS*, 2019.

[3] M. Abeille and A. Lazaric. Improved Regret Bounds for Thompson Sampling in Linear Quadratic Control Problems. In *International Conference on Machine Learning*, 2018.

[4] B. D. O. Anderson and J. B. Moore. *Optimal Control: Linear Quadratic Methods*. 2007.

[5] K. J. Åström and B. Wittenmark. On Self Tuning Regulators. *Automatica*, 9:185–199, 1973.

[6] K. J. Åström and B. Wittenmark. *Adaptive Control*. 2013.

[7] M. G. Azar, R. Munos, and H. J. Kappen. Minimax pac bounds on the sample complexity of reinforcement learning with a generative model. *Machine learning*, 91(3):325–349, 2013.

[8] R. Boczar, N. Matni, and B. Recht. Finite-Data Performance Guarantees for the Output-Feedback Control of an Unknown System. In *57th IEEE Conference on Decision and Control*, 2018.

[9] A. Cohen, A. Hassidim, T. Koren, N. Lazic, Y. Mansour, and K. Talwar. Online Linear Quadratic Control. In *International Conference on Machine Learning*, 2018.

[10] A. Cohen, T. Koren, and Y. Mansour. Learning Linear-Quadratic Regulators Efficiently with only $\sqrt{T}$ Regret. *arXiv:1902.06223*, 2019.

[11] S. Dean, H. Mania, N. Matni, B. Recht, and S. Tu. On the Sample Complexity of the Linear Quadratic Regulator. *arXiv:1710.01688*, 2017.

[12] 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.

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

[14] M. K. S. Faradonbeh, A. Tewari, and G. Michailidis. Optimism-Based Adaptive Regulation of Linear-Quadratic Systems. *arXiv:1711.07230*, 2017.

[15] M. K. S. Faradonbeh, A. Tewari, and G. Michailidis. Finite Time Identification in Unstable Linear Systems. *Automatica*, 96:342–353, 2018.

[16] M. K. S. Faradonbeh, A. Tewari, and G. Michailidis. Input Perturbations for Adaptive Regulation and Learning. *arXiv:1811.04258*, 2018.

[17] M. Fazel, R. Ge, S. M. Kakade, and M. Mesbahi. Global Convergence of Policy Gradient Methods for the Linear Quadratic Regulator. In *International Conference on Machine Learning*, 2018.

[18] C.-N. Fiechter. PAC Adaptive Control of Linear Systems. In *Conference on Learning Theory*, 1997.

[19] M. Hardt, T. Ma, and B. Recht. Gradient Descent Learns Linear Dynamical Systems. *Journal of Machine Learning Research*, 19(29):1–44, 2018.

[20] E. Hazan, K. Singh, and C. Zhang. Learning Linear Dynamical Systems via Spectral Filtering. In *Neural Information Processing Systems*, 2017.

[21] E. Hazan, H. Lee, K. Singh, C. Zhang, and Y. Zhang. Spectral Filtering for General Linear Dynamical Systems. In *Neural Information Processing Systems*, 2018.

[22] M. Ibrahimi, A. Javanmard, and B. V. Roy. Efficient Reinforcement Learning for High Dimensional Linear Quadratic Systems. In *Neural Information Processing Systems*, 2012.

[23] G. N. Iyengar. Robust Dynamic Programming. *Mathematics of Operations Research*, 30(2): 257–280, 2005.

[24] T. Kailath, A. H. Sayed, and B. Hassibi. *Linear Estimation*. 2000.

[25] M. M. Konstantinov, P. H. Petkov, and N. D. Christov. Perturbation analysis of the discrete riccati equation. *Kybernetika*, 29(1):18–29, 1993.

[26] M. M. Konstantinov, D.-W. Gu, V. Mehrmann, and P. H. Petkov. *Perturbation theory for matrix equations*, volume 9. Gulf Professional Publishing, 2003.

[27] D. Malik, A. Pananjady, K. Bhatia, K. Khamaru, P. L. Bartlett, and M. J. Wainwright. Derivative-Free Methods for Policy Optimization: Guarantees for Linear Quadratic Systems. In *AISTATS*, 2019.

[28] H. Mania, S. Tu, and B. Recht. Certainty Equivalence is Efficient for Linear Quadratic Control. *arXiv:1902.07826*, 2019.

[29] A. Nemirovski, A. Juditsky, G. Lan, and A. Shapiro. Robust Stochastic Approximation Approach to Stochastic Programming. *SIAM Journal on Optimization*, 19(4):1574–1609, 2009.

[30] A. Nilim and L. El Ghaoui. Robust Control of Markov Decision Processes with Uncertain Transition Matrices. *Operations Research*, 53(5):780–798, 2005.

[31] Y. Ouyang, M. Gagrani, and R. Jain. Control of Unknown Linear Systems with Thompson Sampling. In *Allerton*, 2017.

[32] S. Oymak and N. Ozay. Non-asymptotic Identification of LTI Systems from a Single Trajectory. *arXiv:1806.05722*, 2018.

[33] T. Sarkar and A. Rakhlin. Near optimal finite time identification of arbitrary linear dynamical systems. In *International Conference on Machine Learning*, 2019.

[34] 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.

[35] M. Simchowitz, R. Boczar, and B. Recht. Learning Linear Dynamical Systems with Semi-Parametric Least Squares. In *Conference on Learning Theory*, 2019.

[36] J.-g. Sun. Perturbation theory for algebraic riccati equations. *SIAM Journal on Matrix Analysis and Applications*, 19(1):39–65, 1998.

[37] J.-g. Sun. Sensitivity analysis of the discrete-time algebraic riccati equation. *Linear algebra and its applications*, 275:595–615, 1998.

[38] A. Tsiamis and G. J. Pappas. Finite Sample Analysis of Stochastic System Identification. *arXiv:1903.09122*, 2019.

[39] S. Tu and B. Recht. The Gap Between Model-Based and Model-Free Methods on the Linear Quadratic Regulator: An Asymptotic Viewpoint. In *Conference on Learning Theory*, 2019.

[40] S. Tu, R. Boczar, A. Packard, and B. Recht. Non-Asymptotic Analysis of Robust Control from Coarse-Grained Identification. *arXiv:1707.04791*, 2017.

[41] H. Xu and S. Mannor. Distributionally Robust Markov Decision Processes. *Mathematics of Operations Research*, 37(2):288–300, 2012.

[42] K. Zhou, J. C. Doyle, and K. Glover. *Robust and Optimal Control*. 1995.

A   Proof of Theorem 1

In this section we prove our meta theorem; we show how an upper bound $\|\widehat{P} - P_\star\| \leq f(\varepsilon)$ can be used to quantify the mismatch between the performance of the the nominal controller and the optimal controller. First, we upper bound $\|\widehat{K} - K_\star\|$ and offer a condition on this mismatch size so that $A_\star + B_\star\widehat{K}$ is a stable matrix. The next two optimization results are helpful in proving $\|\widehat{K} - K_\star\|$ is small.

Lemma 1. *Let f_1, f_2 be two μ-strongly convex twice differentiable functions. Let $\mathbf{x}_1 = \arg\min_{\mathbf{x}} f_1(\mathbf{x})$ and $\mathbf{x}_2 = \arg\min_{\mathbf{x}} f_2(\mathbf{x})$. Suppose $\|\nabla f_1(\mathbf{x}_2)\| \leq \varepsilon$, then $\|\mathbf{x}_1 - \mathbf{x}_2\| \leq \frac{\varepsilon}{\mu}$.*

Proof. Taylor expanding ∇f_1, we have:
$$\nabla f_1(\mathbf{x}_2) = \nabla f_1(\mathbf{x}_1) + \nabla^2 f_1(\tilde{\mathbf{x}})(\mathbf{x}_2 - \mathbf{x}_1) = \nabla^2 f_1(\tilde{\mathbf{x}})(\mathbf{x}_2 - \mathbf{x}_1) \,.$$
for $\tilde{\mathbf{x}} = t\mathbf{x}_1 + (1 - t)\mathbf{x}_2$ with some $t \in [0, 1]$. Therefore:
$$\mu\|\mathbf{x}_1 - \mathbf{x}_2\| \leq \|\nabla^2 f_1(\tilde{\mathbf{x}})(\mathbf{x}_2 - \mathbf{x}_1)\| = \|\nabla f_1(\mathbf{x}_2)\| \leq \varepsilon \,.$$
$\square$

Lemma 2. *Define $f_i(\mathbf{u}; \mathbf{x}) = \frac{1}{2}\mathbf{u}^\mathsf{T} R\mathbf{u} + \frac{1}{2}(A_i\mathbf{x} + B_i\mathbf{u})^\mathsf{T} P_i(A_i\mathbf{x} + B_i\mathbf{u})$ for $i = 1, 2$, with R, P_1, and P_2 positive definite matrices. Let K_i be the unique matrix such that $\mathbf{u}_i := \arg\min_{\mathbf{u}} f_i(\mathbf{u}; \mathbf{x}) = K_i\mathbf{x}$ for any vector $\mathbf{x}$. Also, denote $\Gamma := 1 + \max\{\|A_1\|, \|B_1\|, \|P_1\|, \|K_1\|\}$. Suppose there exists ε such that $0 \leq \varepsilon < 1$ and $\|A_1 - A_2\| \leq \varepsilon$, $\|B_1 - B_2\| \leq \varepsilon$, and $\|P_1 - P_2\| \leq \varepsilon$. Then, we have*
$$\|K_1 - K_2\| \leq \frac{7\varepsilon\Gamma^3}{\underline{\sigma}(R)} \,.$$

Proof. We first compute the gradient $\nabla f_i(\mathbf{u}; \mathbf{x})$ with respect to $\mathbf{u}$:
$$\nabla f_i(\mathbf{u}; \mathbf{x}) = (B_i^\mathsf{T} P_i B_i + R)\mathbf{u} + B_i^\mathsf{T} P_i A_i\mathbf{x} \,.$$
Now, we observe that:
$$\|B_1^\mathsf{T} P_1 B_1 - B_2^\mathsf{T} P_2 B_2\| \leq 7\Gamma^2\varepsilon \text{ and } \|B_1^\mathsf{T} P_1 A_1 - B_2^\mathsf{T} P_2 A_2\| = 7\Gamma^2\varepsilon.$$
Hence, for any vector $\mathbf{x}$ with $\|\mathbf{x}\| \leq 1$, we have
$$\|\nabla f_1(\mathbf{u}; \mathbf{x}) - \nabla f_2(\mathbf{u}; \mathbf{x})\| \leq 7\Gamma^2\varepsilon(\|\mathbf{u}\| + 1).$$
We can bound $\|\mathbf{u}_1\| \leq \|K_1\|\|\mathbf{x}\| \leq \|K_1\|$. Then, from Lemma 1 we obtain
$$\underline{\sigma}(R)\|(K_1 - K_2)\mathbf{x}\| = \underline{\sigma}(R)\|\mathbf{u}_1 - \mathbf{u}_2\| \leq 7\Gamma^3\varepsilon.$$
$\square$

Recall that $\Gamma_\star := 1 + \max\{\|A_\star\|, \|B_\star\|, \|P_\star\|, \|K_\star\|\}$. Now, we upper bound $\|\widehat{K} - K_\star\|$.

Proposition 3. *Let $\varepsilon > 0$ such that $\|\widehat{A} - A_\star\| \leq \varepsilon$ and $\|\widehat{B} - B_\star\| \leq \varepsilon$. Also, let $\|\widehat{P} - P_\star\| \leq f(\varepsilon)$ for some function f such that $f(\varepsilon) \geq \varepsilon$. Then, under Assumption 1 we have*
$$\|\widehat{K} - K_\star\| \leq 7\Gamma_\star^3 f(\varepsilon). \tag{14}$$
Let γ be a real number such that $\rho(L_\star) < \gamma < 1$. Then, if $f(\varepsilon)$ is small enough so that the right hand side of (14) is smaller than $\frac{1 - \gamma}{2\tau(L_\star, \gamma)}$, we have
$$\tau\left(A_\star + B_\star K, \frac{1 + \gamma}{2}\right) \leq \tau(L_\star, \gamma).$$

Proof. By our assumptions $\|\widehat{A} - A_\star\|$, $\|\widehat{B} - B_\star\|$, and $\|\widehat{P} - P_\star\|$ are smaller than $f(\varepsilon)$, and $\underline{\sigma}(R) \geq 1$. Then, Lemma 2 ensures that
$$\|\widehat{K} - K_\star\| \leq 7\Gamma_\star^3 f(\varepsilon).$$

Finally, when ε is small enough so that the right hand side of (14) is smaller or equal than $\frac{1 - \gamma}{2\tau(A_\star + B_\star K, \gamma)}$, we can apply Lemma 5, presented in Section B, to guarantee that $\|(A_\star + B_\star\widehat{K})^k\| \leq \tau(A_\star + B_\star K_\star, \gamma)\left(\frac{1 + \gamma}{2}\right)^k$ for all $k \geq 0$. $\square$

In order to finish the proof of Theorem 1 we need to quantify the suboptimality gap $\widehat{J} - J_\star$ in terms of the controller mismatch $\widehat{K} - K_\star$. For a stable matrix L and a symmetric matrix M, we let $\mathsf{dlyap}(L, M)$ denote the solution X to the Lyapunov equation $L^\top X L - X + M = 0$. The following lemma offers a useful second order expansion of the average LQR cost.

Lemma 3 (Lemma 12 of Fazel et al. [17]). *Let K be an arbitrary static linear controller that stabilizes $(A_\star, B_\star)$. Denote $\Sigma(K) := \mathsf{dlyap}((A_\star + B_\star K)^\top, \sigma_w^2 I_n)$ the covariance matrix of the stationary distribution of the closed loop system $A_\star + B_\star K$. We have that:*

$$J(A_\star, B_\star, K) - J_\star = \mathrm{tr}(\Sigma(K)(K - K_\star)^\mathsf{T}(R + B_\star^\mathsf{T} P_\star B_\star)(K - K_\star)) . \tag{15}$$

Now, we have the necessary ingredients to complete the proof of Theorem 1. Equation 15 implies:

$$J(A_\star, B_\star, K) - J_\star \leq \|\Sigma(K)\| \|R + B_\star^\mathsf{T} P_\star B_\star\| \|K - K_\star\|_F^2 .$$

Proposition 3 states that $\widehat{K}$ stabilizes the system $(A_\star, B_\star)$ when the estimation error is small enough. More precisely, under the assumptions of Theorem 1, we have $\tau\left(A_\star + B_\star \widehat{K}, \frac{1+\gamma}{2}\right) \leq \tau(L_\star, \gamma)$. When $\widehat{L} = A_\star + B_\star \widehat{K}$ is a stable matrix we know that $\Sigma(K) = \sigma^2 \sum_{t \geq 0} (L^\top)^t L^t$. Then, by the triangle inequality we can bound

$$\|\Sigma(K)\| \leq \frac{\sigma_w^2 \tau(L_\star, \gamma)^2}{1 - \left(\frac{\gamma+1}{2}\right)^2} \leq \frac{4\sigma_w^2 \tau(L_\star, \gamma)^2}{1 - \gamma^2} .$$

Recalling that $\Gamma_\star := 1 + \max\{\|A_\star\|, \|B_\star\|, \|P_\star\|, \|K_\star\|\}$, we have $\|R + B_\star^\mathsf{T} P_\star B_\star\| \leq \Gamma^3$. Then,

$$J(K) - J_\star \leq 4\sigma_w^2 \Gamma^3 \frac{\tau(L_\star, \gamma)^2}{1 - \gamma^2} \|K - K_\star\|_F^2$$

$$\leq 4\sigma_w^2 \min\{n, d\} \Gamma^3 \frac{\tau(L_\star, \gamma)^2}{1 - \gamma^2} \|K - K_\star\|^2$$

$$\leq 200 \sigma_w^2 d \Gamma^9 \frac{\tau(L_\star, \gamma)^2}{1 - \gamma^2} f(\varepsilon)^2,$$

where we used Proposition 3 and the assumption on $f(\varepsilon)$.

B   Perturbation of the discrete algebraic Riccati equation

B.1   Proof of Proposition 1

Given parameters (A, B, Q) (R is assumed fixed throughout; Q is assumed positive semidefinite throughout) we denote by $F(X, A, B, Q)$ the matrix expression

$$F(X, A, B, Q) = X - A^\top X A + A^\top X B (R + B^\top X B)^{-1} B^\top X A - Q$$

$$= X - A^\top X \left(I + B R^{-1} B^\top X\right)^{-1} A - Q . \tag{16}$$

Then, solving the Riccati equation associated with (A, B, Q) corresponds to finding the unique positive definite matrix X such that $F(X, A, B, Q) = 0$. We denote by $P_\star$ the solution of the Riccati equation corresponding to the true system parameters $(A_\star, B_\star)$ and we denote by $\widehat{P}$ the solution associated with $(\widehat{A}, \widehat{B}, \widehat{Q})$. Our goal is to upper bound $\|\widehat{P} - P_\star\|$ in terms of ε, where $\varepsilon > 0$ such that $\|\widehat{A} - A_\star\| \leq \varepsilon$, $\|\widehat{B} - B_\star\| \leq \varepsilon$, and $\|\widehat{Q} - Q_\star\| \leq \varepsilon$.

We denote $\Delta_P = \widehat{P} - P_\star$. The proof strategy goes as follows. Given the identities $F(P_\star, A_\star, B_\star, Q_\star) = 0$ and $F(\widehat{P}, \widehat{A}, \widehat{B}, \widehat{Q}) = 0$ we construct an operator Φ such that Δ_P is its unique fixed point. Then, we show that the fixed point of Φ must have small norm when ε is sufficiently small.

We denote $S_\star = B_\star R^{-1} B_\star^\top$ and $\widehat{S} = \widehat{B} R^{-1} \widehat{B}^\top$. Also, recall that $L_\star = A_\star + B_\star K_\star$. For any matrix X such that $I + S_\star(P_\star + X)$ is invertible we have

$$F(P_\star + X, A_\star, B_\star, Q_\star) = X - L_\star^\top X L_\star + L_\star^\top X \left[I + S_\star(P_\star + X)\right]^{-1} S_\star X L_\star . \tag{17}$$

To check this identity one needs to add $F(P_\star, A_\star, B_\star, Q_\star)$, which is equal to zero, to the right hand side of (17) and use the identity $(I + B_\star R^{-1} B_\star^\top P_\star)^{-1} A_\star = A_\star + B_\star K_\star$. This last identity can be checked by recalling $K_\star = -(R + B_\star^\top P_\star B_\star)^{-1} B_\star^\top P_\star A_\star$ and using the matrix inversion formulat.

To write (17) more compactly we define the following two matrix operators

$$\mathcal{T}(X) = X - L_\star^\top X L_\star \text{ and } \mathcal{H}(X) = L_\star^\top X \left(I + S_\star(P_\star + X)\right)^{-1} S_\star X L_\star.$$

Then, Equation 17 becomes $F(P_\star + X, A_\star, B_\star, Q_\star) = \mathcal{T}(X) + \mathcal{H}(X)$. Since Equation 17 is satisfied by any matrix X with $I + S_\star(P_\star + X)$ invertible, the matrix equation

$$F(P_\star + X, A_\star, B_\star, Q_\star) - F(P_\star + X, \widehat{A}, \widehat{B}, \widehat{Q}) = \mathcal{T}(X) + \mathcal{H}(X) \tag{18}$$

has a unique symmetric solution X such that $P_\star + X \succeq 0$. That solution is $X = \Delta_P$ because any solution of (18) must satisfy $F(P_\star + X, \widehat{A}, \widehat{B}, \widehat{Q}) = 0$.

The linear map $\mathcal{T} \colon X \mapsto X - L_\star^\top X L_\star$ has eigenvalues equal to $1 - \lambda_i \lambda_j$, where λ_i and λ_j are eigenvalues of the closed loop matrix $L_\star$. Since $L_\star$ is a stable matrix, the linear map $\mathcal{T}$ must be invertible. Now, we define the operator

$$\Phi(X) = \mathcal{T}^{-1} \left(F(P_\star + X, A_\star, B_\star, Q_\star) - F(P_\star + X, \widehat{A}, \widehat{B}, \widehat{Q}) - \mathcal{H}(X) \right).$$

Then, solving for X in Equation 18 is equivalent to finding X satisfying $P_\star + X \succeq 0$ such that $X = \Phi(X)$. Hence, Φ has a unique symmetric fixed point X such that $P_\star + X \succeq 0$ and that is $X = \Delta_P$. Now, we consider the set

$$\mathcal{S}_\nu := \left\{ X \colon \|X\| \le \nu, X = X^\top, P_\star + X \succeq 0 \right\}$$

and we show that for an appropriately chosen ν the operator Φ maps $\mathcal{S}_\nu$ into itself and is also a contraction over the set $\mathcal{S}_\nu$. If we show these two properties, Φ is guaranteed to have a fixed point in the set $\mathcal{S}_\nu$. However, since Δ_P is the only possible fixed point of Φ in a set $\mathcal{S}_\nu$ we find $\|\Delta_P\| \le \nu$.

We denote $\Delta_A = \widehat{A} - A_\star$, $\Delta_B = \widehat{B} - B_\star$, $\Delta_Q = \widehat{Q} - Q_\star$, and $\Delta_S = \widehat{S} - S_\star$. By assumption we have $\|\Delta_A\| \le \varepsilon$, $\|\Delta_B\| \le \varepsilon$, $\|\Delta_Q\| \le \varepsilon$. Then, $\|\Delta_S\| \le 3\|B_\star\|\|R^{-1}\|\varepsilon$ because $\varepsilon \le \|B_\star\|$.

Lemma 4. *Suppose the matrices X, X_1, X_2 belong to $\mathcal{S}_\nu$, with $\nu \le \min\{1, \|S_\star\|^{-1}\}$. Furthermore, we assume that $\|\Delta_A\| \le \varepsilon$, $\|\Delta_B\| \le \varepsilon$, and $\|\Delta_Q\| \le \varepsilon$ with $\varepsilon \le \min\{1, \|B_\star\|\}$. Finally, let $\underline{\sigma}(P_\star) \ge 1$. Then*

$$\|\Phi(X)\| \le 3\frac{\tau(L_\star, \gamma)^2}{1 - \gamma^2} \left[\|L_\star\|^2 \|S_\star\| \nu^2 + \varepsilon \|A_\star\|_+^2 \|P_\star\|_+^2 \|B_\star\|_+ \|R^{-1}\|_+ \right],$$

$$\|\Phi(X_1) - \Phi(X_2)\| \le 32\frac{\tau(L_\star, \gamma)^2}{1 - \gamma^2} \left[\|L_\star\|^2 \|S_\star\| \nu + \varepsilon \|A_\star\|_+^2 \|P_\star\|_+^3 \|B_\star\|_+^3 \|R^{-1}\|_+^2 \right] \|X_1 - X_2\|.$$

The proof of this lemma is defered to Appendix E. Now, we choose

$$\nu = 6\,\varepsilon\,\frac{\tau(L_\star, \gamma)^2}{1 - \gamma^2} \|A_\star\|_+^2 \|P_\star\|_+^2 \|B_\star\|_+ \|R^{-1}\|_+. \tag{19}$$

Since ε is assumed to be small enough, we know

$$\nu \le \min \left\{ \frac{1 - \gamma^2}{128\tau(L_\star, \gamma)^2 \|L_\star\|^2 \|S_\star\|}, \|S_\star\|^{-1}, \frac{1}{2} \right\}.$$

Then, the operator Φ satisfies $\|\Phi(X_1) - \Phi(X_2)\| \le \frac{1}{2}\|X_1 - X_2\|$ for all X_1 and X_2 in $\mathcal{S}_\nu$. Moreover, we have $\|\Phi(X)\| \le \nu$ for all $X \in \mathcal{S}_\nu$. Since $\nu \le \underline{\sigma}(P_\star)$, we know that $P_\star + \Phi(X) \succeq 0$

Therefore, Φ maps $\mathcal{S}_\nu$ into itself and is a contraction over $\mathcal{S}_\nu$. Hence, Φ has a fixed point in $\mathcal{S}_\nu$ since $\mathcal{S}_\nu$ is a closed set. However, we already argued that the unique fixed point of Φ is Δ_P. Therefore, $\Delta_P \in \mathcal{S}_\nu$ and $\|\Delta_P\| \le \nu$. Proposition 1 is now proven.

B.2 Proof of Proposition 2

Since both noisy and noiseless LQR have the same associated Riccati equation and the same optimal controller, we can focus on the noiseless case in this section. Namely, noiseless LQR takes the form

$$\min_{\mathbf{u}} \sum_{t=0}^{\infty} \mathbf{x}_t^\top Q \mathbf{x}_t + \mathbf{u}_t^\top R \mathbf{u}_t \text{ , where } \mathbf{x}_{t+1} = A_\star \mathbf{x}_t + B_\star \mathbf{u}_t,$$

for a given initial state $\mathbf{x}_0$. Then, we know that the cost achieved by the optimal controller when the system is initialized at $\mathbf{x}_0$ is equal to $\mathbf{x}_0^\top P_\star \mathbf{x}_0$.

We denote by $J(A, B, \mathbf{x}_0, \{\mathbf{u}_t\}_{t \geq 0})$ the cost achieved on a linear system (A, B) initialized at $\mathbf{x}_0$ by the input sequence $\{\mathbf{u}_t\}_{t \geq 0}$. When the input sequence is given by a time invariant linear gain matrix K we slightly abuse notation and denote the cost by $J(A, B, \mathbf{x}_0, K)$. In this case, $J(A, B, \mathbf{x}_0, K) = \mathbf{x}_0^\top P \mathbf{x}_0$, where P is the solution to the associated Riccati equation.

Now, let $\mathbf{x}_0$ be an arbitrary unit state vector in $\mathbb{R}^n$. Then,

$$\mathbf{x}_0^\top \widehat{P} \mathbf{x}_0 - \mathbf{x}_0^\top P_\star \mathbf{x}_0 = J(\widehat{A}, \widehat{B}, \mathbf{x}_0, \widehat{K}) - J(A_\star, B_\star, \mathbf{x}_0, K_\star)$$
$$\leq J(\widehat{A}, \widehat{B}, \mathbf{x}_0, \{\widehat{\mathbf{u}}_t\}_{t \geq 0}) - J(A_\star, B_\star, \mathbf{x}_0, K_\star)$$

for any sequence of inputs $\{\widehat{\mathbf{u}}_t\}_{t \geq 0}$. We denote by $\widehat{\mathbf{x}}_t$ the states produced by $\widehat{\mathbf{u}}_t$ on the system $(\widehat{A}, \widehat{B})$ and by $\mathbf{x}_t$ and $\mathbf{u}_t$ the states and actions obtained on the system $(A_\star, B_\star)$ when the optimal controller $\mathbf{u}_t = K_\star \mathbf{x}_t$ is used. To prove Proposition 2 we choose a sequence of actions $\{\widehat{\mathbf{u}}_t\}_{t \geq 0}$ such that $J(\widehat{A}, \widehat{B}, \mathbf{x}_0, \{\widehat{\mathbf{u}}_t\}_{t \geq 0}) \approx J(A_\star, B_\star, \mathbf{x}_0, K_\star)$.

For any sequence of inputs $\{\widehat{\mathbf{u}}_t\}_{t \geq 0}$ such that the series defining the cost $J(\widehat{A}, \widehat{B}, \mathbf{x}_0, \{\widehat{\mathbf{u}}_t\}_{t \geq 0})$ is absolutely convergent, we can write

$$J(\widehat{A}, \widehat{B}, \mathbf{x}_0, \widehat{K}) - J(A_\star, B_\star, \mathbf{x}_0, K_\star) = \sum_{j=0}^{\infty} \sum_{i=0}^{\ell-1} \left[\widehat{\mathbf{x}}_{\ell j+i}^\top Q \widehat{\mathbf{x}}_{\ell j+i} - \mathbf{x}_{\ell j+i}^\top Q \mathbf{x}_{\ell j+i} \right] \qquad (20)$$
$$+ \sum_{j=0}^{\infty} \sum_{i=0}^{\ell-1} \left[\widehat{\mathbf{u}}_{\ell j+i}^\top R \widehat{\mathbf{u}}_{\ell j+i} - \mathbf{u}_{\ell j+i}^\top R \mathbf{u}_{\ell j+i} \right].$$

Then, the key idea is to choose a sequence of inputs $\{\widehat{\mathbf{u}}_t\}_{t \geq 0}$ such that the system $(\widehat{A}, \widehat{B})$ tracks the system $(A_\star, B_\star, K_\star)$, i.e., $\widehat{\mathbf{x}}_{\ell j} = \mathbf{x}_{\ell j}$ for any $j \geq 0$ ($\widehat{\mathbf{x}}_0 = \mathbf{x}_0$ because both systems are initialized at the same state). This can be done because $(\widehat{A}, \widehat{B})$ is $(\ell, \tau/2)$-controllable when $(A_\star, B_\star)$ is (ℓ, τ)-controllable and the estimation error is sufficiently small, as shown in Lemma 6. First, we present a result that quantifies the effect of matrix perturbations on powers of matrices.

Lemma 5. *Let M be an arbitrary matrix in $\mathbb{R}^{n \times n}$ and let $\rho \geq \rho(M)$. Then, for all $k \geq 1$ and real matrices Δ of appropriate dimensions we have*

$$\|(M + \Delta)^k\| \leq \tau(M, \rho)(\tau(M, \rho)\|\Delta\| + \rho)^k,$$
$$\|(M + \Delta)^k - M^k\| \leq k\, \tau(M, \rho)^2 (\tau(M, \rho)\|\Delta\| + \rho)^{k-1}\|\Delta\|.$$

Recall that $\tau(M, \rho)$ is defined in Equation 3.

The proof is deferred to Appendix C. Lemma 5 quantifies the effect of a perturbation Δ, applied to a matrix M on the spectral radius of $M + \Delta$. We are interested in quantifying the sizes of these perturbations for all $k = 1, 2, \ldots, \ell$. Depending on $\|\Delta\|$, M, and ρ the sum $\tau(M, \rho)\|\Delta\| + \rho$ can either be greater than one or smaller than one. For notational simplicity, in the rest of the proof we denote $\beta = \max\{1, \varepsilon\tau(A_\star, \rho) + \rho\}$. Then, we have $\|(A_\star + \Delta)^k\| \leq \tau(A_\star, \rho)\beta^{\ell-1}$ and $\|(A_\star + \Delta)^k - A_\star^k\| \leq \ell\tau(A_\star, \rho)^2\beta^{\ell-1}\varepsilon$ for all $k \leq \ell - 1$ and all real matrices Δ with $\|\Delta\| \leq \varepsilon$.

We denote $C_\ell = \begin{bmatrix} B_\star & A_\star B_\star & \ldots & A_\star^{\ell-1} B_\star \end{bmatrix}$ and $\widehat{C}_\ell = \begin{bmatrix} \widehat{B} & \widehat{A}\widehat{B} & \ldots & \widehat{A}^{\ell-1}\widehat{B} \end{bmatrix}$. Before presenting the next result we recall that for any block matrix M with blocks $M_{i,j}$ we have $\|M\|^2 \leq \sum_{i,j} \|M_{i,j}\|^2$. The next lemma gives us control over the smallest positive singular value of the controllability matrix $\widehat{C}_\ell$ in terms of the corresponding value for C_ℓ.

15

Lemma 6. *Suppose the linear $(A_\star, B_\star)$ is (ℓ, ν)-controllable and let ρ be a real number such that $\rho \geq \rho(A_\star)$. Then, if $\|\widehat{A} - A_\star\| \leq \varepsilon$ and $\|\widehat{B} - B_\star\| \leq \varepsilon$, we have*

$$\underline{\sigma}(\widehat{C}_\ell) \geq \tau - 3\varepsilon\ell^{\frac{3}{2}}\tau(A_\star, \rho)^2 \max\{1, \tau(A_\star, \rho)\|\Delta\| + \rho\}^{\ell-1}(\|B_\star\| + 1).$$

The proof is deferred to Appendix D. Lemma 6 tells us that by the assumption made in Proposition 2 on ε, we have $\underline{\sigma}(\widehat{C}_\ell) \geq \frac{\tau_\ell}{2}$. Hence, we know that for any $\mathbf{x}_0 \in \mathbb{R}^n$ and $\mathbf{u}_0, \mathbf{u}_1, \ldots, \mathbf{u}_{\ell-1} \in \mathbb{R}^d$, there exist $\widehat{\mathbf{u}}_0, \widehat{\mathbf{u}}_1, \ldots, \widehat{\mathbf{u}}_{\ell-1} \in \mathbb{R}^d$ such that

$$A_\star^\ell \mathbf{x}_0 + \sum_{i=0}^{\ell-1} A_\star^i B_\star \mathbf{u}_{\ell-1-i} = \widehat{A}^\ell \mathbf{x}_0 + \sum_{i=0}^{\ell-1} \widehat{A}^i \widehat{B} \widehat{\mathbf{u}}_{\ell-1-i} \tag{21}$$

because the system $(\widehat{A}, \widehat{B})$ is controllable. This equation implies that $\widehat{\mathbf{x}}_\ell = \mathbf{x}_\ell$.

We denote the concatenation of $\mathbf{u}_i$, for i from 0 to $\ell - 1$ by $\mathbf{u}^{(\ell)}$. We define $\widehat{\mathbf{u}}^{(\ell)}$ analogously. Therefore, Equation (21) can be rewritten as

$$\left(A_\star^\ell - \widehat{A}^\ell\right)\mathbf{x}_0 + \left(\mathcal{C}_\ell - \widehat{\mathcal{C}}_\ell\right)\mathbf{u}^{(\ell)} = \widehat{\mathcal{C}}_\ell(\widehat{\mathbf{u}}^{(\ell)} - \mathbf{u}^{(\ell)}). \tag{22}$$

Recall that $\beta = \max\{1, \tau(A_\star, \rho)\|\Delta\| + \rho\}$. Combining Lemma 5 and the upper bound on operator norms of block matrices we find $\|\widehat{\mathcal{C}}_\ell - \mathcal{C}_\ell\| \leq \varepsilon\ell^{\frac{3}{2}}\tau(A_\star, \rho)^2\beta^{\ell-1}(\|B_\star\| + 1)$.

We are free to choose $\widehat{\mathbf{u}}^{(\ell)}$ anyway we wish as long as Equation (22) is true. Therefore, we can choose $\widehat{\mathbf{u}}^{(\ell)}$ such that $\widehat{\mathbf{u}}^{(\ell)} - \mathbf{u}^{(\ell)}$ is perpendicular to the nullspace of $\widehat{\mathcal{C}}_\ell$. Then,

$$\frac{\tau_\ell}{2}\|\widehat{\mathbf{u}}^{(\ell)} - \mathbf{u}^{(\ell)}\| \leq \|\widehat{\mathcal{C}}_\ell(\widehat{\mathbf{u}}^{(\ell)} - \mathbf{u}^{(\ell)})\| \leq \varepsilon\ell\tau(A_\star, \rho)^2\beta^{\ell-1}\|\mathbf{x}_0\| + \|\widehat{\mathcal{C}}_\ell - \mathcal{C}_\ell\|\|\mathbf{u}^{(\ell)}\|$$

$$\leq \varepsilon\ell\tau(A_\star, \rho)^2\beta^{\ell-1}\|\mathbf{x}_0\| + \varepsilon\ell^{\frac{3}{2}}\tau(A_\star, \rho)^2\beta^{\ell-1}(\|B_\star\| + 1)\|\mathbf{u}^{(\ell)}\|.$$

Hence,

$$\|\widehat{\mathbf{u}}^{(\ell)} - \mathbf{u}^{(\ell)}\| \leq \frac{2\varepsilon\ell^{\frac{3}{2}}}{\tau_\ell}\tau(A_\star, \rho)^2\beta^{\ell-1}(\|B_\star\| + 1)\left(\|\mathbf{x}_0\| + \|\mathbf{u}^{(\ell)}\|\right)$$

$$=: \eta\left(\|\mathbf{x}_0\| + \|\mathbf{u}^{(\ell)}\|\right). \tag{23}$$

Let us consider the block Toeplitz matrix

$$\mathcal{T}_\ell = \begin{bmatrix} 0 & 0 & 0 & \ldots & 0 \\ B_\star & 0 & 0 & \cdots & 0 \\ A_\star B_\star & B_\star & 0 & \cdots & 0 \\ \vdots & \cdots & \cdots & \cdots & \cdots \\ A_\star^{\ell-2}B_\star & A_\star^{\ell-3}B_\star & \cdots & B_\star & 0 \end{bmatrix}.$$

From Lemma 5 and the upper bound on operator norms of block matrices we have $\|\mathcal{T}_\ell - \widehat{\mathcal{T}}_\ell\| \leq \varepsilon\ell^2\tau(A_\star, \rho)^2\beta^{\ell-2}(\|B_\star\| + 1)$. Let $\mathbf{x}^{(\ell)}$ be the concatenation of the vectors $\mathbf{x}_0, \mathbf{x}_1, \ldots, \mathbf{x}_{\ell-1}$. Then,

$$\mathbf{x}^{(\ell)} = \mathcal{T}_\ell \mathbf{u}^{(\ell)} + \begin{bmatrix} I_n \\ A_\star \\ \vdots \\ A^{\ell-1} \end{bmatrix}\mathbf{x}_0.$$

Hence,

$$\|\mathbf{x}^{(\ell)} - \widehat{\mathbf{x}}^{(\ell)}\| \leq \|\mathcal{T}_\ell\mathbf{u}^{(\ell)} - \widehat{\mathcal{T}}_\ell\widehat{\mathbf{u}}^{(\ell)}\| + \varepsilon\ell^{\frac{3}{2}}\tau(A_\star, \rho)^2\beta^{\ell-2}\|\mathbf{x}_0\|$$

$$\leq \|\mathcal{T}_\ell\mathbf{u}^{(\ell)} - \widehat{\mathcal{T}}_\ell\mathbf{u}^{(\ell)}\| + \|\widehat{\mathcal{T}}_\ell\mathbf{u}^{(\ell)} - \widehat{\mathcal{T}}_\ell\widehat{\mathbf{u}}^{(\ell)}\| + \varepsilon\ell^{\frac{3}{2}}\tau(A_\star, \rho)^2\beta^{\ell-2}\|\mathbf{x}_0\|$$

$$\leq \|\mathcal{T}_\ell - \widehat{\mathcal{T}}_\ell\|\|\mathbf{u}^{(\ell)}\| + \|\widehat{\mathcal{T}}_\ell\|\|\mathbf{u}^{(\ell)} - \widehat{\mathbf{u}}^{(\ell)}\| + \varepsilon\ell^{\frac{3}{2}}\tau(A_\star, \rho)^2\beta^{\ell-2}\|\mathbf{x}_0\|$$

$$\leq \varepsilon\ell^2\tau(A_\star, \rho)^2\beta^{\ell-2}(\|B_\star\| + 1)\|\mathbf{u}^{(\ell)}\| + \varepsilon\ell^{\frac{3}{2}}\tau(A_\star, \rho)^2\beta^{\ell-2}\|\mathbf{x}_0\|$$

$$\quad + \ell\tau(A_\star, \rho)\beta^{\ell-2}(\|B_\star\| + 1)\|\mathbf{u}^{(\ell)} - \widehat{\mathbf{u}}^{(\ell)}\|$$

$$\leq 2\varepsilon\ell^{\frac{5}{2}}\tau(A_\star, \rho)^3\beta^{2(\ell-1)}(1 + \tau^{-1})(\|B_\star\| + 1)^2\left[\|\mathbf{x}_0\| + \|\mathbf{u}^{(\ell)}\|\right]$$

$$=: \mu\left[\|\mathbf{u}^{(\ell)}\| + \|\mathbf{x}_0\|\right]. \tag{24}$$

In Equations (23) and (24) we proved that the inputs and states of the system $(\widehat{A}, \widehat{B})$ are close to the inputs and states of the system $(A_\star, B_\star)$ from time 0 to ℓ. Since the inputs to the system $(\widehat{A}, \widehat{B})$ satisfy Equation (22), we know that $\widehat{\mathbf{x}}_{\ell j} = \mathbf{x}_{\ell j}$ for all j. We can repeat the same argument as above, with $\mathbf{x}_{\ell j}$ taking the place of $\mathbf{x}_0$, to show that the inputs and states of the two systems are close to each other from time ℓj to $\ell(j+1)$. Let us denote by $\mathbf{x}_j^{(\ell)}$ the concatenation of the vectors $\mathbf{x}_{\ell j}, \mathbf{x}_{\ell j+1}, \ldots,$ $\mathbf{x}_{\ell j+\ell-1}$ and let $\mathbf{u}_j^{(\ell)}$ be defined analogously. Then,

$$\|\widehat{\mathbf{u}}_j^{(\ell)} - \mathbf{u}_j^{(\ell)}\| \le \eta \left[\|\mathbf{u}_j^{(\ell)}\| + \|\mathbf{x}_{\ell j}\| \right], \text{ and } \|\widehat{\mathbf{x}}_j^{(\ell)} - \mathbf{x}_j^{(\ell)}\| \le \mu \left[\|\mathbf{u}_j^{(\ell)}\| + \|\mathbf{x}_{\ell j}\| \right]. \quad (25)$$

Now, we note that

$$
\begin{aligned}
\mathbf{x}_0^\top \widehat{P} \mathbf{x}_0 - \mathbf{x}_0^\top P_\star \mathbf{x}_0 \le & \sum_{j=0}^{\infty} \sum_{i=0}^{\ell-1} \left[\widehat{\mathbf{x}}_{\ell j+i}^\top Q \widehat{\mathbf{x}}_{\ell j+i} - \mathbf{x}_{\ell j+i}^\top Q \mathbf{x}_{\ell j+i} \right] \\
& + \sum_{j=0}^{\infty} \sum_{i=0}^{\ell-1} \left[\widehat{\mathbf{u}}_{\ell j+i}^\top R \widehat{\mathbf{u}}_{\ell j+i} - \mathbf{u}_{\ell j+i}^\top R \mathbf{u}_{\ell j+i} \right] \\
\le & \sum_{j=0}^{\infty} 2\|Q\| \|\mathbf{x}_j^{(\ell)}\| \|\mathbf{x}_j^{(\ell)} - \widehat{\mathbf{x}}_j^{(\ell)}\| + \|Q\| \|\mathbf{x}_j^{(\ell)} - \widehat{\mathbf{x}}_j^{(\ell)}\|^2 \\
& + \sum_{j=0}^{\infty} 2\|R\| \|\mathbf{u}_j^{(\ell)}\| \|\mathbf{u}_j^{(\ell)} - \widehat{\mathbf{u}}_j^{(\ell)}\| + \|R\| \|\mathbf{u}_j^{(\ell)} - \widehat{\mathbf{u}}_j^{(\ell)}\|^2 .
\end{aligned}
$$

Now, we use the upper bounds from (25). We always have $\eta \le \mu$. Since Proposition 2 assumes ε is small enough, we also have $\mu \le 1$. Using these upper bounds, we find

$$
\begin{aligned}
\widehat{J} - J_\star \le & \, \mu \sum_{j=0}^{\infty} 2\|Q\| \|\mathbf{x}_j^{(\ell)}\| \left[\|\mathbf{u}_j^{(\ell)}\| + \|\mathbf{x}_{\ell j}\| \right] + \|Q\| \left[\|\mathbf{u}_j^{(\ell)}\| + \|\mathbf{x}_{\ell j}\| \right]^2 \\
& + \mu \sum_{j=0}^{\infty} 2\|R\| \|\mathbf{u}_j^{(\ell)}\| \left[\|\mathbf{u}_j^{(\ell)}\| + \|\mathbf{x}_{\ell j}\| \right] + \|R\| \left[\|\mathbf{u}_j^{(\ell)}\| + \|\mathbf{x}_{\ell j}\| \right]^2 .
\end{aligned}
$$

Then, we get $\widehat{J} - J_\star \le 8\mu \max\{\|Q\|, \|R\|\} \sum_{j=0}^{\infty} \|\mathbf{x}_j^{(\ell)}\|^2 + \|\mathbf{u}_j^{(\ell)}\|^2$ after using the inequalities $(a+b)^2 \le 2(a^2 + b^2)$ and $2ab \le a^2 + b^2$. Now, As long as $\|\mathbf{x}_0\| \le 1$ we have

$$\min\{\underline{\sigma}(Q), \underline{\sigma}(R)\} \sum_{j=0}^{\infty} \|\mathbf{x}_j^{(\ell)}\|^2 + \|\mathbf{u}_j^{(\ell)}\|^2 \le \sum_{t=0}^{\infty} \mathbf{x}_t^\top Q \mathbf{x}_t + \mathbf{u}_t^\top R \mathbf{u}_t = \mathbf{x}_0^\top P_\star \mathbf{x}_0 \le \|P_\star\|. \quad (26)$$

Since the initial state is an arbitrary unit norm vector, our upper bound on $\mathbf{x}_0^\top (\widehat{P} - P_\star) \mathbf{x}_0$ becomes

$$\lambda_{\max}\left(\widehat{P} - P_\star \right) \le 16\varepsilon \ell^{\frac{5}{2}} \tau(A_\star, \rho)^3 \beta^{2(\ell-1)} (1 + \nu^{-1})(1 + \|B_\star\|)^2 \|P_\star\| \frac{\max\{\|Q\|, \|R\|\}}{\min\{\underline{\sigma}(Q), \underline{\sigma}(R)\}}. \quad (27)$$

Now, we can reverse the roles of $(\widehat{A}, \widehat{B})$ and $(A_\star, B_\star)$ and repeat the same argument and obtain an upper bound on $\lambda_{\max}\left(P_\star - \widehat{P} \right)$ analogous to Equation (27), but which has $\|P_\star\|$ replaced by $\|\widehat{P}\|$ on the right hand side. However, (27) implies that $\|\widehat{P}\| \le \|P_\star\| + 1 \le 2\|P_\star\|$ because we assumed that ε is small enough such that the right hand side of (27) is less than one, and because $P_\star \succeq I_n$. The conclusion follows.

We note that the proof can be easily extended to the case when the cost matrix Q is also being perturbed. Moreover, the only essential step in the argument where we used $Q \succ 0$ is (26). The goal of (26) is to upper bound $\sum_{j=0}^{\infty} \|\mathbf{x}_j^{(\ell)}\|^2 + \|\mathbf{u}_j^{(\ell)}\|^2$. This quantity can be upper bounded even when Q is not positive definite, but the system $(Q^{1/2}, A)$ is observable; which is a necessary requirement for LQR on the parameters (A, B, Q, R) to stabilize the system (A, B).

C Proof of Lemma 5

This proof is a simple modification of Lemma D.1 in [11]. We replicate the argument here for completeness.

Fix an integer $k \geq 1$. Consider the expansion of $(M + \Delta)^k$ into 2^k terms. Label all these terms as $T_{i,j}$ for $i = 0, ..., k$ and $j = 1, ..., \binom{k}{i}$ where i denotes the degree of Δ in the term (hence there are $\binom{k}{i}$ terms with a degree of i for Δ). Using the fact that $\|M^k\| \leq \tau(M, \rho)\rho^k$ for all $k \geq 0$, we can bound $\|T_{i,j}\| \leq \tau(M, \rho)^{i+1}\rho^{k-i}\|\Delta\|^i$. Hence by triangle inequality:

$$
\begin{aligned}
\|(M + \Delta)^k\| &\leq \sum_{i=0}^{k} \sum_{j} \|T_{i,j}\| \\
&\leq \sum_{i=0}^{k} \binom{k}{i} \tau(M, \rho)^{i+1}\rho^{k-i}\|\Delta\|^i \\
&= \tau(M, \rho) \sum_{i=0}^{k} \binom{k}{i} (\tau(M, \rho)\|\Delta\|)^i \rho^{k-i} \\
&= \tau(M, \rho)(\tau(M, \rho)\|\Delta\| + \rho)^k.
\end{aligned}
$$

To prove the first part of the lemma we follow the same argument. We find

$$
\begin{aligned}
\|(M + \Delta)^k - M^k\| &\leq \sum_{i=1}^{k} \sum_{j} \|T_{i,j}\| \\
&\leq \sum_{i=1}^{k} \binom{k}{i} \tau(M, \rho)^{i+1}\rho^{k-i}\|\Delta\|^i \\
&= \tau(M, \rho) \sum_{i=1}^{k} \binom{k}{i} (\tau(M, \rho)\|\Delta\|)^i \rho^{k-i} \\
&= \tau(M, \rho) \left[(\tau(M, \rho)\|\Delta\| + \rho)^k - \rho^k \right] \\
&\leq k C_M^2 (\tau(M, \rho)\|\Delta\| + \rho)^{k-1}\|\Delta\|,
\end{aligned}
$$

where the last inequality follows from the mean value theorem applied to the function $z \mapsto z^k$.

D Proof of Lemma 6

We can write

$$
\underline{\sigma}\left(\begin{bmatrix} \widehat{B} & \widehat{A}\widehat{B} & \ldots & \widehat{A}^{\ell-1}\widehat{B} \end{bmatrix} \right) = \min_{v \in \mathcal{S}^{n-1}} \left\| v^\top \begin{bmatrix} \widehat{B} & \widehat{A}\widehat{B} & \ldots & \widehat{A}^{\ell-1}\widehat{B} \end{bmatrix} \right\|.
$$

Fix an arbitrary unit vector v in $\mathbb{R}^d$. Then,

$$
\begin{aligned}
&\left\| v^\top \begin{bmatrix} B_\star & A_\star B_\star & \ldots & A_\star^{\ell-1}B_\star \end{bmatrix} - v^\top \begin{bmatrix} \widehat{B} & \widehat{A}\widehat{B} & \ldots & \widehat{A}^{\ell-1}\widehat{B} \end{bmatrix} \right\| \\
&\leq \left\| v^\top \begin{bmatrix} B_\star & A_\star B_\star & \ldots & A_\star^{\ell-1}B_\star \end{bmatrix} - v^\top \begin{bmatrix} B_\star & \widehat{A}B_\star & \ldots & \widehat{A}^{\ell-1}B_\star \end{bmatrix} \right\| \\
&\quad + \left\| v^\top \begin{bmatrix} B_\star & \widehat{A}B_\star & \ldots & \widehat{A}^{\ell-1}B_\star \end{bmatrix} - v^\top \begin{bmatrix} \widehat{B} & \widehat{A}\widehat{B} & \ldots & \widehat{A}^{\ell-1}\widehat{B} \end{bmatrix} \right\| \\
&\leq \varepsilon \ell^{\frac{3}{2}} \tau(A_\star, \rho)^2 \beta^{\ell-1}\|B_\star\| + \varepsilon\sqrt{\ell}\tau(A_\star, \rho, \beta)^{\ell-1} \\
&\leq \varepsilon \ell^{\frac{3}{2}} \tau(A_\star, \rho)^2 \beta^{\ell-1} (\|B_\star\| + 1).
\end{aligned}
$$

We used $\ell \geq 1$, $\tau(A_\star, \rho) \geq 1$, Lemma 5, and the upper bound $\|M\|^2 \leq \sum_{i,j} \|M_{i,j}\|^2$ on the operator norm of a block matrix. The conclusion follows by the triangle inequality.

18

E Proof of Lemma 4

We wish to upper bound $\|\Phi(X)\|$ and $\|\Phi(X_1) - \Phi(X_2)\|$ for X, X_1, and X_2 in $\mathcal{S}_\nu$. First, we upper bound the operator norm of the linear operator $\mathcal{T}^{-1}$, the inverse of $\mathcal{T}\colon X \mapsto X - L_\star^\top X L_\star$. Since $L_\star$ is a stable matrix, the linear map $\mathcal{T}$ must be invertible. Moreover, when $L_\star$ is stable and $X - L_\star^\top X L_\star = M$ for some matrix M, we know that $X = \sum_{k=0}^\infty (L_\star^k)^\top M L_\star^k$. Therefore, by the triangle inequality, the operator norm of $\mathcal{T}^{-1}$ can be upper bounded by $\|\mathcal{T}^{-1}\| \le \frac{\tau(L_\star, \rho)^2}{1 - \rho^2}$. Before we proceed with the rest of the proof we present a technical lemma which will be used several times.

Lemma 7. *Let M and N be two positive semidefinite matrices of the same dimension. Then $\|N(I + MN)^{-1}\| \le \|N\|$.*

Proof. We assume that M and N are invertible. If they are not, we can work with the matrices $M + \nu I$ and $N + \nu I$ and take the limit of ν going to zero. Then, we have $N(I + MN)^{-1} = NN^{-1}(N^{-1} + M)^{-1} = (N^{-1} + M)^{-1} \preceq N$. $\qquad\square$

Next, recall that $\mathcal{H}(X) = L_\star^\top X \left(I + S_\star(P_\star + X)\right)^{-1} S_\star X L_\star$. Then, Lemma 7 yields

$$\|\mathcal{H}(X)\| \le \|L_\star\|^2 \|S_\star\| \|X\|^2.$$

We turn our attention to the difference $F(P_\star + X, A_\star, B_\star) - F(P_\star + X, \widehat{A}, \widehat{B})$. We use the notation P_X as a shorthand for $P_\star + X$. Then, by Equation 16 we find

$$F(P_X, \widehat{A}, \widehat{B}, \widehat{Q}) - F(P_X, A_\star, B_\star, Q_\star) = A_\star^\top P_X (I + S_\star P_X)^{-1} A_\star - \widehat{A}^\top P_X \left(I + \widehat{S} P_X\right)^{-1} \widehat{A} - \Delta_Q$$

$$= A_\star^\top P_X (I + S_\star P_X)^{-1} \Delta_S P_X (I + \widehat{S} P_X)^{-1} A_\star - A_\star^\top P_X (I + \widehat{S} P_X)^{-1} \Delta_A$$

$$- \Delta_A^\top P_X (I + \widehat{S} P_X)^{-1} A_\star - \Delta_A^\top P_X (I + \widehat{S} P_X)^{-1} \Delta_A - \Delta_Q. \tag{28}$$

Then,

$$\|F(P_\star + X, \widehat{A}, \widehat{B}, \widehat{Q}) - F(P_\star + X, A_\star, B_\star, Q_\star)\|$$
$$\le \|A_\star\|^2 \|P_X\|^2 \|\Delta_S\| + 2\|A_\star\| \|P_X\| \varepsilon + \|P_X\| \varepsilon^2 + \varepsilon,$$

where we used Lemma 7. Since $X \in \mathcal{S}_\nu$, we know $\|X\| \le \nu$ and hence $\|P_X\| \le \|P_\star\| + \nu$. We assumed that $\nu \le 1/2$ and so $\|P_X\| \le \|P_\star\| + 1$. Now, we know that $\|\Delta_S\| \le 2\|B_\star\| \|R^{-1}\| \varepsilon + \|R^{-1}\| \varepsilon^2$ and since we assumed $\varepsilon \le \|B_\star\|$, we have $\|\Delta_S\| \le 3\|B_\star\| \|R^{-1}\| \varepsilon$. Therefore,

$$\|\Phi(X)\| \le \frac{\tau(L_\star, \rho)^2}{1 - \rho^2} \left[\|L_\star\|^2 \|S_\star\| \nu^2 + 3\|A_\star\|_+^2 \|P_\star\|_+^2 \|B_\star\|_+ \|R^{-1}\|_+ \varepsilon \right].$$

We use Lemma 7, the assumption $\nu \le \|S_\star\|^{-1}$, and the definition of $\mathcal{H}$ to upper bound

$$\|\mathcal{H}(X_1) - \mathcal{H}(X_2)\| \le \|L_\star\|^2 \left[\|S_\star\|^2 \nu^2 + 2\|S_\star\| \nu \right] \|X_1 - X_2\| \le 3\|L_\star\|^2 \|S_\star\| \nu \|X_1 - X_2\|.$$

Let us denote $\mathcal{G}(X) = F(P_\star + X, \widehat{A}, \widehat{B}, \widehat{Q}) - F(P_\star + X, A_\star, B_\star, Q_\star)$. In order to upper bound $\|\mathcal{G}(X_1) - \mathcal{G}(X_2)\|$ we first upper bound the norm of $(I + S_\star P_X)^{-1}$ and $(I + \widehat{S} P_X)^{-1}$. Since $\|X\| \le \nu \le 1/2$ and since $P_\star \succeq I$, by Lemma 7 we get

$$\|(I + S_\star P_X)^{-1}\| = \|P_X^{-1} P_X (I + S_\star P_X)^{-1}\| \le \|P_X^{-1}\| \|P_X (I + S_\star P_X)^{-1}\| \le 2\|P_X\|.$$

Therefore, after some algebraic manipulations, we obtain

$$\|\mathcal{G}(X_1) - \mathcal{G}(X_2)\| \le 32\varepsilon \|A_\star\|_+^2 \|P_\star\|_+^3 \|B_\star\|_+^3 \|R^{-1}\|_+^2 \|X_1 - X_2\|.$$

F Proofs for LQG

F.1 Proof of Theorem 3

The proof strategy works as follows. Note that unlike the case for LQR, we were unable to reuse or derive an exact second order perturbation bound analogous to Lemma 3 for LQG. Hence we resort to an argument based on a simple Taylor expansion.

We first assume that $\widehat{K}$ satisfies $\|\widehat{K} - K_\star T^{-1}\| \leq \varepsilon$ and that $\varepsilon \leq 1$. Let $\Theta = (A, B, C, W, V, Q, R)$ be the parameters that specific an instance of LQG in (8). Here, we will be slightly more general than in (8) and allow the process noise $\mathbf{w}_t$ and observation noise $\mathbf{v}_t$ to have non-isotropic covariances W and V, respectively. Let $J(\widehat{A}, \widehat{B}, \widehat{C}, \widehat{K}, \widehat{L}; \Theta)$ denote the following cost:

$$J(\widehat{A}, \widehat{B}, \widehat{C}, \widehat{K}, \widehat{L}; \Theta) = \lim_{T \to \infty} \mathbb{E}\left[\frac{1}{T} \sum_{t=1}^T \mathbf{y}_t^\mathsf{T} Q \mathbf{y}_t + \mathbf{u}_t^\mathsf{T} R \mathbf{u}_t\right] \text{ s.t.}$$

$$\mathbf{x}_{t+1} = A\mathbf{x}_t + B\mathbf{u}_t + \mathbf{w}_t, \quad \mathbf{w}_t \sim \mathcal{N}(0, W),$$
$$\mathbf{y}_t = C\mathbf{x}_t + \mathbf{v}_t, \quad \mathbf{v}_t \sim \mathcal{N}(0, V),$$
$$\mathbf{u}_t = \widehat{K}\widehat{\mathbf{x}}_t,$$
$$\widehat{\mathbf{x}}_{t+1} = \widehat{A}\widehat{\mathbf{x}}_t + \widehat{B}\mathbf{u}_t + \widehat{L}(\mathbf{y}_t - \widehat{C}\widehat{\mathbf{x}}_t).$$

As observed earlier, we have that $J(\widehat{A}, \widehat{B}, \widehat{C}, \widehat{K}, \widehat{L}; \Theta) = J(\widehat{A}, \widehat{B}, \widehat{C}, \widehat{K}, \widehat{L}, \tilde{\Theta})$, with $\tilde{\Theta} = (TAT^{-1}, TB, CT^{-1}, TWT^\mathsf{T}, V, Q, R)$ for any invertible T. Hence, we can assume that $T = I$ moving forward.

We consider the closely related cost $J_s(\widehat{A}, \widehat{B}, \widehat{C}, \widehat{K}, \widehat{L}; \Theta)$ defined as:

$$J_s(\widehat{A}, \widehat{B}, \widehat{C}, \widehat{K}, \widehat{L}; \Theta) = \lim_{T \to \infty} \mathbb{E}\left[\frac{1}{T} \sum_{t=1}^T \mathbf{x}_t^\mathsf{T} Q \mathbf{x}_t + \mathbf{u}_t^\mathsf{T} R \mathbf{u}_t\right] \text{ s.t.}$$

$$\mathbf{x}_{t+1} = A\mathbf{x}_t + B\mathbf{u}_t + \mathbf{w}_t, \quad \mathbf{w}_t \sim \mathcal{N}(0, W),$$
$$\mathbf{y}_t = C\mathbf{x}_t + \mathbf{v}_t, \quad \mathbf{v}_t \sim \mathcal{N}(0, V),$$
$$\mathbf{u}_t = \widehat{K}\widehat{\mathbf{x}}_t,$$
$$\widehat{\mathbf{x}}_{t+1} = \widehat{A}\widehat{\mathbf{x}}_t + \widehat{B}\mathbf{u}_t + \widehat{L}(\mathbf{y}_t - \widehat{C}\widehat{\mathbf{x}}_t).$$

Observe that:

$$J(\widehat{A}, \widehat{B}, \widehat{C}, \widehat{K}, \widehat{L}; \Theta) = J_s(\widehat{A}, \widehat{B}, \widehat{C}, \widehat{K}, \widehat{L}; (A, B, C, W, V, Q_c, R)) + \mathrm{tr}(QV), \quad Q_c := C^\mathsf{T} Q C.$$

Let $J_{s,\star}$ denote the optimal value for J_s. This decomposition shows that:

$$J(\widehat{A}, \widehat{B}, \widehat{C}, \widehat{K}, \widehat{L}; \Theta) - J_\star = J_s(\widehat{A}, \widehat{B}, \widehat{C}, \widehat{K}, \widehat{L}; (A, B, C, W, V, Q_c, R)) - J_{s,\star}.$$

Hence we will work with J_s from here on. In the sequel, we will drop the dependence of J on Θ. We write $\widehat{\Xi} = (\widehat{A}, \widehat{B}, \widehat{C}, \widehat{K}, \widehat{L})$ and $\Xi_\star = (A, B, C, K_\star, L_\star)$. By Taylor's theorem,

$$J_s(\widehat{\Xi}) - J_{s,\star} = [DJ_s(\Xi_\star)](\Delta) + \frac{1}{2}[D^2 J_s(\tilde{\Xi})](\Delta, \Delta)$$

$$= \frac{1}{2}[D^2 J_s(\tilde{\Xi})](\Delta, \Delta).$$

where the first equality holds by the optimality of $\Xi_\star$ and $\tilde{\Xi}$ is an element along the ray between $\widehat{\Xi}$ and $\Xi_\star$.

We start by defining $\mathbf{e}_t = \widehat{\mathbf{x}}_t - \mathbf{x}_t$. Observe that:

$$\begin{bmatrix} \mathbf{x}_t \\ \widehat{\mathbf{x}}_t \end{bmatrix} = S \begin{bmatrix} \mathbf{x}_t \\ \mathbf{e}_t \end{bmatrix}, \quad S = \begin{bmatrix} I & 0 \\ I & I \end{bmatrix}. \tag{29}$$

We write:

$$\begin{bmatrix} \mathbf{x}_{t+1} \\ \widehat{\mathbf{x}}_{t+1} \end{bmatrix} = \begin{bmatrix} A\mathbf{x}_t + B\widehat{K}\widehat{\mathbf{x}}_t + \mathbf{w}_t \\ \widehat{A}\widehat{\mathbf{x}}_t + \widehat{B}\widehat{K}\widehat{\mathbf{x}}_t + \widehat{L}(C\mathbf{x}_t + \mathbf{v}_t - \widehat{C}\widehat{\mathbf{x}}_t) \end{bmatrix}$$

$$= \begin{bmatrix} A & B\widehat{K} \\ \widehat{L}C & (\widehat{A} + \widehat{B}\widehat{K}) - \widehat{L}\widehat{C} \end{bmatrix} \begin{bmatrix} \mathbf{x}_t \\ \widehat{\mathbf{x}}_t \end{bmatrix} + \begin{bmatrix} I & 0 \\ 0 & \widehat{L} \end{bmatrix} \begin{bmatrix} \mathbf{w}_t \\ \mathbf{v}_t \end{bmatrix}$$

$$=: M(\widehat{A}, \widehat{B}, \widehat{C}, \widehat{K}, \widehat{L}) \begin{bmatrix} \mathbf{x}_t \\ \widehat{\mathbf{x}}_t \end{bmatrix} + \begin{bmatrix} I & 0 \\ 0 & \widehat{L} \end{bmatrix} \begin{bmatrix} \mathbf{w}_t \\ \mathbf{v}_t \end{bmatrix}. \tag{30}$$

20

We abbreviate $\widehat{M} := M(\widehat{A}, \widehat{B}, \widehat{C}, \widehat{K}, \widehat{L})$. Applying a similarity transform to $\widehat{M}$:

$$S^{-1}\widehat{M}S = \begin{bmatrix} I & 0 \\ -I & I \end{bmatrix} \begin{bmatrix} A & B\widehat{K} \\ \widehat{L}C & (\widehat{A} + \widehat{B}\widehat{K}) - \widehat{L}\widehat{C} \end{bmatrix} \begin{bmatrix} I & 0 \\ I & I \end{bmatrix}$$

$$= \begin{bmatrix} A + B\widehat{K} & B\widehat{K} \\ (\widehat{A} - A) + (\widehat{B} - B)\widehat{K} + \widehat{L}(C - \widehat{C}) & (\widehat{A} - \widehat{L}\widehat{C}) + (\widehat{B} - B)\widehat{K} \end{bmatrix}$$

$$=: \widehat{N}\,.$$

Therefore $\widehat{M}$ is stable iff $\widehat{N}$ is stable.

By (30), if $\widehat{M}$ is stable, then the stationary distribution of $(\mathbf{x}_t, \widehat{\mathbf{x}}_t)$ is given by $\mathcal{N}(0, \widehat{\Sigma})$ with $\widehat{\Sigma} = $ dlyap $\left(\widehat{M}^{\mathsf{T}}, \begin{bmatrix} W & 0 \\ 0 & \widehat{L}V\widehat{L}^{\mathsf{T}} \end{bmatrix} \right)$. Therefore, we have that:

$$J_s(\widehat{A}, \widehat{B}, \widehat{C}, \widehat{K}, \widehat{L}) = \text{tr}\left(\begin{bmatrix} Q_c & 0 \\ 0 & \widehat{K}^{\mathsf{T}}R\widehat{K} \end{bmatrix} \widehat{\Sigma} \right)\,.$$

Now for any $(\tilde{A}, \tilde{B}, \tilde{C}, \tilde{K}, \tilde{L})$, observe we can write $\tilde{M} = M(\tilde{A}, \tilde{B}, \tilde{C}, \tilde{K}, \tilde{L})$ in terms of $M_\star = M(A, B, C, K_\star, L_\star)$ as:

$$\tilde{M} = \tilde{M} - M_\star + M_\star$$

$$= \begin{bmatrix} 0 & B(\tilde{K} - K_\star) \\ (\tilde{L} - L_\star)C & (\tilde{A} - A) + (\tilde{B} - B)\tilde{K} + B(\tilde{K} - K_\star) - ((\tilde{L} - L_\star)\tilde{C} + L_\star(\tilde{C} - C)) \end{bmatrix} + M_\star$$

$$=: \Delta + M_\star\,.$$

Using the assumption that $\varepsilon \leq 1$, we can bound $\|\Delta\|$ by:

$$\|\Delta\| \leq \|B\|\varepsilon_K + \|C\|\varepsilon_L + \varepsilon_A + (\|K_\star\| + 1)\varepsilon_B + \|B\|\varepsilon_K + (\|C\| + 1)\varepsilon_L + \|L_\star\|\varepsilon_C$$

$$\leq \varepsilon_A + (\|K_\star\| + 1)\varepsilon_B + \|L_\star\|\varepsilon_C + 2(\|C\| + 1)\varepsilon_L + 2\|B\|\varepsilon_K$$

$$\leq 2\Gamma_\star(\varepsilon_A + \varepsilon_B + \varepsilon_C + \varepsilon_K + \varepsilon_L)$$

$$\leq 10\Gamma_\star\varepsilon\,.$$

Let $\rho \geq \rho(M_\star)$. If $10\Gamma_\star \cdot \varepsilon \leq \frac{1-\rho}{2\tau(M_\star, \rho)}$, then we have $\rho(\tilde{M}) \leq (1 + \rho)/2$ and $\tau(\tilde{M}, (1 + \rho)/2) \leq \tau(M_\star, \rho)$. This holds for any $(\tilde{A}, \tilde{B}, \tilde{C}, \tilde{K}, \tilde{L})$ along the ray of $(\widehat{A}, \widehat{B}, \widehat{C}, \widehat{K}, \widehat{L})$ and (A, B, C, K, L). Therefore for any point along this ray we have that $\tilde{M}$ is stable and that for any fixed H, $\|\text{dlyap}(\tilde{M}^{\mathsf{T}}, H)\| \leq \frac{\|H\|\tau(M_\star, \rho)}{1 - \text{Avg}(\rho, 1)^2}$.

Next, we consider $\tilde{\Xi} = (\tilde{A}, \tilde{B}, \tilde{C}, \tilde{K}, \tilde{L})$ and differentiate J_s twice w.r.t. $\tilde{\Xi}$, assuming that $\tilde{\Xi}$ corresponds to a point along the ray $\widehat{\Xi} = (\widehat{A}, \widehat{B}, \widehat{C}, \widehat{K}, \widehat{L})$ and $\Xi_\star = (A, B, C, K_\star, L_\star)$. By the chain rule:

$$[D^2 J_s(\tilde{\Xi})](\Delta, \Delta)$$

$$= \text{tr}\left(\begin{bmatrix} Q_c & 0 \\ 0 & \tilde{K}^{\mathsf{T}}R\tilde{K} \end{bmatrix} [D^2\Sigma(\tilde{\Xi})](\Delta, \Delta) \right) + \text{tr}\left(\begin{bmatrix} 0 & 0 \\ 0 & \Delta_K^{\mathsf{T}}R\tilde{K} + \tilde{K}^{\mathsf{T}}R\Delta_K \end{bmatrix} [D\Sigma(\tilde{\Xi})](\Delta) \right)$$

$$+ 2\,\text{tr}\left(\begin{bmatrix} 0 & 0 \\ 0 & \Delta_K^{\mathsf{T}}R\Delta_K \end{bmatrix} \Sigma(\tilde{\Xi}) \right)$$

$$\leq (\text{tr}(Q_c) + \text{tr}(R)(\|K_\star\| + 1)^2)\|[D^2\Sigma(\tilde{\Xi})](\Delta, \Delta)\|$$

$$+ 2\,\text{tr}(R)(\|K_\star\| + 1)\varepsilon_K\|[D\Sigma(\tilde{\Xi})](\Delta)\| + 2\,\text{tr}(R)\varepsilon_K^2\|\Sigma(\tilde{\Xi})\|$$

$$\leq (\text{tr}(Q_c) + \text{tr}(R)\Gamma_\star^2)\left(\|[D^2\Sigma(\tilde{\Xi})](\Delta, \Delta)\| + 2\varepsilon_K\|[D\Sigma(\tilde{\Xi})](\Delta)\| + 2\varepsilon_K^2\|\Sigma(\tilde{\Xi})\| \right)\,.$$

Given the composition $h(x) = g(f(x))$, we have by the chain rule:

$$[Dh(x)](\Delta) = [Dg(f(x))]([Df(x)](\Delta))\,, \tag{31}$$

$$[D^2h(x)](\Delta, \Delta) = [D^2g(f(x))]([Df(x)](\Delta), [Df(x)](\Delta)) + [Dg(f(x))]([D^2f(x)](\Delta, \Delta))\,. \tag{32}$$

We apply this with $g(M,Q) = \mathsf{dlyap}(M,Q)$ and

$$f(\tilde{\Xi}) = \left(\begin{bmatrix} A & B\tilde{K} \\ \tilde{L}C & (\tilde{A}+\tilde{B}\tilde{K})-\tilde{L}\tilde{C} \end{bmatrix}^{\mathsf{T}}, \begin{bmatrix} W & 0 \\ 0 & \tilde{L}V\tilde{L}^{\mathsf{T}} \end{bmatrix} \right) .$$

Therefore the composition $h = g \circ f$ is $h(\Xi) = \Sigma(\Xi)$. Differentiating f,

$$[Df(\tilde{\Xi})](\Delta) = \left(\begin{bmatrix} 0 & B\Delta_K \\ \Delta_L C & \Delta_A + \tilde{B}\Delta_K + \Delta_B\tilde{K} - (\tilde{L}\Delta_C + \Delta_L\tilde{C}) \end{bmatrix}^{\mathsf{T}}, \begin{bmatrix} 0 & 0 \\ 0 & \tilde{L}V\Delta_L^{\mathsf{T}} + \Delta_L V\tilde{L}^{\mathsf{T}} \end{bmatrix} \right) ,$$

$$[D^2 f(\tilde{\Xi})](\Delta,\Delta) = \left(\begin{bmatrix} 0 & 0 \\ 0 & 2\Delta_B\Delta_K - 2\Delta_L\Delta_C \end{bmatrix}^{\mathsf{T}}, \begin{bmatrix} 0 & 0 \\ 0 & 2\Delta_L V\Delta_L^{\mathsf{T}} \end{bmatrix} \right) .$$

Letting $P = g(M,Q)$, we have:

$$[Dg(M,Q)](\Delta) = g(M, \Delta_Q + M^{\mathsf{T}} P\Delta_M + \Delta_M^{\mathsf{T}} PM) ,$$

$$[D^2 g(M,Q)](\Delta,\Delta) = 2g(M, \Delta_M^{\mathsf{T}} P\Delta_M + M^{\mathsf{T}} g(M,\Delta_Q)\Delta_M + \Delta_M^{\mathsf{T}} g(M,\Delta_Q)M)$$
$$+ 2g(M, \Delta_M^{\mathsf{T}} g(M, \Delta_M^{\mathsf{T}} PM + M^{\mathsf{T}} P\Delta_M)M + M^{\mathsf{T}} g(M, \Delta_M^{\mathsf{T}} PM + M^{\mathsf{T}} P\Delta_M)\Delta_M) . \tag{33}$$

The derivation for (33) is postponed for now.

Let $\mathcal{S}_M(Q)$ be the linear operator $Q \mapsto g(M,Q)$. Observe that $\|\mathcal{S}_M\| \leq \frac{\tau^2(M,\rho)}{1-\rho^2}$ and $\|M^{\mathsf{T}} g(M,Q)\| \leq \frac{\tau^2(M,\rho)}{1-\rho^2}\|Q\|$. With this, we can bound,

$$\|[Dg(M,Q)](\Delta)\| \leq \frac{\tau^2(M,\rho)}{1-\rho^2}\|\Delta_Q\| + 2\frac{\tau^4(M,\rho)}{(1-\rho^2)^2}\|Q\|\|\Delta_M\| ,$$

$$\|[D^2 g(M,Q)](\Delta,\Delta)\| \leq 10\frac{\tau^6(M,\rho)}{(1-\rho^2)^3}\|Q\|\|\Delta_M\|^2 + 4\frac{\tau^4(M,\rho)}{(1-\rho^2)^2}\|\Delta_Q\|\|\Delta_M\| .$$

Therefore,

$$\|[Dg(f(\tilde{\Xi}))]([Df(\tilde{\Xi})](\Delta))\|$$

$$\leq 2\frac{\tau^2(\tilde{M},\tilde{\rho})}{1-\tilde{\rho}^2}\|\tilde{L}\|\|V\|\varepsilon_L + 20\frac{\tau^4(\tilde{M},\tilde{\rho})}{(1-\tilde{\rho}^2)^2}\max\{\|W\|,\|\tilde{L}V\tilde{L}^{\mathsf{T}}\|\}\Gamma_\star\varepsilon$$

$$\leq 2\frac{\tau^2(M_\star,\rho)}{1-\mathsf{Avg}(\rho,1)^2}\Gamma_\star\|V\|\varepsilon_L + 20\frac{\tau^4(M_\star,\rho)}{(1-\mathsf{Avg}(\rho,1)^2)^2}\max\{\|W\|,\Gamma_\star^2\|V\|\}\Gamma_\star\varepsilon$$

$$\leq 22\frac{\tau^4(M_\star,\rho)}{(1-\mathsf{Avg}(\rho,1)^2)^2}\max\{\|W\|,\|V\|\}\Gamma_\star^3\varepsilon .$$

and,

$$\|[Dg(f(\tilde{\Xi}))]([D^2 f(\tilde{\Xi})](\Delta,\Delta))\|$$

$$\leq 2\frac{\tau^2(\tilde{M},\tilde{\rho})}{1-\tilde{\rho}^2}\|V\|\varepsilon_L^2 + 4\frac{\tau^4(\tilde{M},\tilde{\rho})}{(1-\tilde{\rho}^2)^2}\max\{\|W\|,\|\tilde{L}V\tilde{L}^{\mathsf{T}}\|\}(\varepsilon_B\varepsilon_K + \varepsilon_L\varepsilon_C)$$

$$\leq 2\frac{\tau^2(M_\star,\rho)}{1-\mathsf{Avg}(\rho,1)^2}\|V\|\varepsilon_L^2 + 8\frac{\tau^4(M_\star,\rho)}{(1-\mathsf{Avg}(\rho,1)^2)^2}\max\{\|W\|,\Gamma_\star^2\|V\|\}\varepsilon^2$$

$$\leq 10\frac{\tau^4(M_\star,\rho)}{(1-\mathsf{Avg}(\rho,1)^2)^2}\max\{\|W\|,\|V\|\}\Gamma_\star^2\varepsilon^2 .$$

and also,

$$\|[D^2 g(f(\tilde{\Xi}))]([Df(\tilde{\Xi})](\Delta),[Df(\tilde{\Xi})](\Delta))\|$$

$$\leq 10\frac{\tau^6(\tilde{M},\tilde{\rho})}{(1-\tilde{\rho}^2)^3}\max\{\|W\|,\|\tilde{L}V\tilde{L}^{\mathsf{T}}\|\}(10\Gamma_\star\varepsilon)^2 + 4\frac{\tau^4(\tilde{M},\tilde{\rho})}{(1-\tilde{\rho}^2)^2}(2\|\tilde{L}\|\|V\|\varepsilon_L)(10\Gamma_\star\varepsilon)$$

$$\leq 1000\frac{\tau^6(M_\star,\rho)}{(1-\mathsf{Avg}(\rho,1)^2)^3}\max\{\|W\|,\|V\|\}\Gamma_\star^4\varepsilon^2 + 80\frac{\tau^4(M_\star,\rho)}{(1-\mathsf{Avg}(\rho,1)^2)^2}\|V\|\Gamma_\star\varepsilon^2$$

$$\leq 1080\frac{\tau^6(M_\star,\rho)}{(1-\mathsf{Avg}(\rho,1)^2)^3}\max\{\|W\|,\|V\|\}\Gamma_\star^4\varepsilon^2 .$$

Therefore,

$$\|\Sigma(\tilde{\Xi})\| \leq \frac{\tau^2(M_\star, \rho)}{1 - \mathsf{Avg}(\rho, 1)^2} \max\{\|W\|, \|V\|\}\Gamma_\star^2 \,,$$

$$\|[D\Sigma(\tilde{\Xi})](\Delta)\| \leq 22\frac{\tau^4(M_\star, \rho)}{(1 - \mathsf{Avg}(\rho, 1)^2)^2} \max\{\|W\|, \|V\|\}\Gamma_\star^3\varepsilon \,,$$

$$\|[D^2\Sigma(\tilde{\Xi})](\Delta, \Delta)\| \leq 1090\frac{\tau^6(M_\star, \rho)}{(1 - \mathsf{Avg}(\rho, 1)^2)^3} \max\{\|W\|, \|V\|\}\Gamma_\star^4\varepsilon^2 \,.$$

Combining these bounds with the calculations for J_s,

$$
\begin{aligned}
&[D^2 J_s(\tilde{\Xi})](\Delta, \Delta) \\
&\leq (\mathrm{tr}(Q_c) + \mathrm{tr}(R)\Gamma_\star^2)\left(\|[D^2\Sigma(\tilde{\Xi})](\Delta, \Delta)\| + 2\varepsilon_K\|[D\Sigma(\tilde{\Xi})](\Delta)\| + 2\varepsilon_K^2\|\Sigma(\tilde{\Xi})\|\right) \\
&\leq 1136(\mathrm{tr}(Q_c) + \mathrm{tr}(R)\Gamma_\star^2)\frac{\tau^6(M_\star, \rho)}{(1 - \mathsf{Avg}(\rho, 1)^2)^3} \max\{\|W\|, \|V\|\}\Gamma_\star^4\varepsilon^2 \\
&\leq 1136(\mathrm{tr}(Q_c) + \mathrm{tr}(R))\frac{\tau^6(M_\star, \rho)}{(1 - \mathsf{Avg}(\rho, 1)^2)^3} \max\{\|W\|, \|V\|\}\Gamma_\star^6\varepsilon^2 \\
&\leq 1136 \cdot 64(\mathrm{tr}(Q_c) + \mathrm{tr}(R))\frac{\tau^6(M_\star, \rho)}{(1 - \rho^2)^3} \max\{\|W\|, \|V\|\}\Gamma_\star^6\varepsilon^2 \,.
\end{aligned}
$$

To finish the first part of proof, we show the derivation for (33). Observe that for E small:

$$
\begin{aligned}
(A + E)^{-1} &= (A(I + A^{-1}E))^{-1} = (I + A^{-1}E)^{-1}A^{-1} = \sum_{k=0}^{\infty}(-A^{-1}E)^k A^{-1} \\
&= A^{-1} - A^{-1}EA^{-1} + A^{-1}EA^{-1}EA^{-1} + O(\|E\|^3) \,.
\end{aligned}
$$

Therefore, letting $S = I - M^\mathsf{T} \otimes M^\mathsf{T}$,

$$
\begin{aligned}
&(I - (M + \Delta_M)^\mathsf{T} \otimes (M + \Delta_M)^\mathsf{T})^{-1}\mathrm{vec}(Q + \Delta_Q) \\
&= S^{-1}\mathrm{vec}(Q + \Delta_Q) \\
&\quad + S^{-1}(\Delta_M^\mathsf{T} \otimes M^\mathsf{T} + M^\mathsf{T} \otimes \Delta_M^\mathsf{T} + \Delta_M^\mathsf{T} \otimes \Delta_M^\mathsf{T})S^{-1}\mathrm{vec}(Q + \Delta_Q) \\
&\quad + S^{-1}(\Delta_M^\mathsf{T} \otimes M^\mathsf{T} + M^\mathsf{T} \otimes \Delta_M^\mathsf{T})S^{-1}(\Delta_M^\mathsf{T} \otimes M^\mathsf{T} + M^\mathsf{T} \otimes \Delta_M^\mathsf{T})S^{-1}\mathrm{vec}(Q + \Delta_Q) + O(\|\Delta\|^3) \\
&= S^{-1}\mathrm{vec}(Q) + S^{-1}\mathrm{vec}(\Delta_Q) \\
&\quad + S^{-1}(\Delta_M^\mathsf{T} \otimes M^\mathsf{T} + M^\mathsf{T} \otimes \Delta_M^\mathsf{T})S^{-1}\mathrm{vec}(Q) + S^{-1}(\Delta_M^\mathsf{T} \otimes \Delta_M^\mathsf{T})S^{-1}\mathrm{vec}(Q) \\
&\quad + S^{-1}(\Delta_M^\mathsf{T} \otimes M^\mathsf{T} + M^\mathsf{T} \otimes \Delta_M^\mathsf{T})S^{-1}\mathrm{vec}(\Delta_Q) \\
&\quad + S^{-1}(\Delta_M^\mathsf{T} \otimes M^\mathsf{T} + M^\mathsf{T} \otimes \Delta_M^\mathsf{T})S^{-1}(\Delta_M^\mathsf{T} \otimes M^\mathsf{T} + M^\mathsf{T} \otimes \Delta_M^\mathsf{T})S^{-1}\mathrm{vec}(Q) + O(\|\Delta\|^3) \,.
\end{aligned}
$$

Now to finish the proof, we need to obtain a $\bar{\varepsilon}$ such that

$$\max\{\|\widehat{A} - TA_\star T^{-1}\|, \|\widehat{B} - TB_\star\|, \|\widehat{C} - C_\star T^{-1}\|, \|\widehat{L} - TL_\star\|, \|\widehat{K} - K_\star T^{-1}\|\} \leq \bar{\varepsilon} \,.$$

We do this by observing that:

$$K_\star T^{-1} = \mathsf{LQR}(TA_\star T^{-1}, TB_\star, T^{-\mathsf{T}}C_\star^\mathsf{T}QC_\star T^{-1}, R) \,.$$

Hence with $\widehat{K} = \mathsf{LQR}(\widehat{A}, \widehat{B}, \widehat{C}^\mathsf{T}Q\widehat{C}, R)$, we can use the Riccati perturbation between $\overline{P}_\star = \mathsf{dare}(TA_\star T^{-1}, TB_\star, T^{-\mathsf{T}}C_\star^\mathsf{T}QC_\star T^{-1}, R)$ and $\widehat{P} = \mathsf{dare}(\widehat{A}, \widehat{B}, \widehat{C}^\mathsf{T}Q\widehat{C}, R)$. We can bound, assuming that that $\varepsilon_C \leq \|C_\star\|$:

$$\|T^{-\mathsf{T}}C_\star^\mathsf{T}QC_\star T^{-1} - \widehat{C}^\mathsf{T}Q\widehat{C}\| \leq (2\|C_\star\| + \varepsilon_C)\|Q\|\varepsilon_C \leq 3\|C_\star\|\|Q\|\varepsilon_C \,.$$

Therefore by our hypothesis and the assumption that f is monotonically increasing,

$$\|\widehat{P} - \overline{P}_\star\| \leq f(\max\{\varepsilon_A, \varepsilon_B, 3\|C_\star\|\|Q\|\varepsilon_C\}) \leq f(3\|C_\star\|_+\|Q\|_+ \max\{\varepsilon_A, \varepsilon_B, \varepsilon_C\}) \,.$$

as long as $3\|C_\star\|_+ \|Q\|_+ \max\{\varepsilon_A, \varepsilon_B, \varepsilon_C\} \le \gamma_0$. By Lemma 2, this allows us to bound:

$$\|\widehat{K} - K_\star T^{-1}\| \le \frac{7\Gamma_\star^3}{\underline{\sigma}(R)} \max\{\varepsilon_A, \varepsilon_B, f(3\|C_\star\|_+ \|Q\|_+ \max\{\varepsilon_A, \varepsilon_B, \varepsilon_C\})\}$$

$$\le \frac{7\Gamma_\star^3}{\underline{\sigma}(R)} f(3\|C_\star\|_+ \|Q\|_+ \max\{\varepsilon_A, \varepsilon_B, \varepsilon_C\}) \,.$$

We can set $\overline{\varepsilon} = \frac{7\Gamma_\star^3}{\underline{\sigma}(R)} f(3\|C_\star\|_+ \|Q\|_+ \max\{\varepsilon_A, \varepsilon_B, \varepsilon_C\})$ because we assume that $f(\gamma) \ge \gamma$ and $\underline{\sigma}(R) \ge 1$.

We now complete the proof by observing that for $M_\star = \begin{bmatrix} A_\star & B_\star K_\star \\ L_\star C_\star & A_\star + B_\star K_\star - L_\star C_\star \end{bmatrix}$ and $N_\star = \begin{bmatrix} A_\star + B_\star K_\star & B_\star K_\star \\ 0 & A_\star - L_\star C_\star \end{bmatrix}$, using S defined in (29), we have that $N_\star = S^{-1} M_\star S$. This identity has several consequences. First, we have $\rho(N_\star) = \rho(M_\star)$. Second, for any integer $k \ge 1$ we have $S^{-1} M_\star^k S = N_\star^k$, and therefore

$$\|N_\star^k\| \le \operatorname{cond}(S) \|M_\star^k\| = \frac{3 + \sqrt{5}}{2} \|M_\star^k\| \,.$$

A nearly identical argument shows that $\|M_\star^k\| \le \frac{3+\sqrt{5}}{2} \|N_\star^k\|$. Therefore, $\tau(M_\star, \rho) \asymp \tau(N_\star, \rho)$ for any ρ.