Towards a Dimension-Free Understanding of Adaptive Linear Control

Juan C. Perdomo
jcperdomo@berkeley.edu
University of California, Berkeley

Max Simchowitz
msimchow@berkeley.edu
University of California, Berkeley

Alekh Agarwal
alekha@microsoft.com
Microsoft Research

Peter Bartlett
peter@berkeley.edu
University of California, Berkeley

March 18, 2021

Abstract

We study the problem of adaptive control of the linear quadratic regulator for systems in very high, or even infinite dimension. We demonstrate that while sublinear regret requires finite dimensional inputs, the ambient state dimension of the system need not be bounded in order to perform online control. We provide the first regret bounds for LQR which hold for infinite dimensional systems, replacing dependence on ambient dimension with more natural notions of problem complexity. Our guarantees arise from a novel perturbation bound for certainty equivalence which scales with the *prediction error* in estimating the system parameters, without requiring consistent parameter recovery in more stringent measures like the operator norm. When specialized to finite dimensional settings, our bounds recover near optimal dimension and time horizon dependence.

1 Introduction

Reinforcement learning (RL) has matured considerably in recent years, setting its sights on increasingly ambitious tasks in ever more complex environments. With this increased complexity, it is neither possible nor desirable to learn models of the environment that are uniformly accurate across all possible states. In particular, to scale learning methods to complex sensorimotor state observations, it is critical to focus model estimation on the parts of the state space that are most relevant to the cost and which can be influenced by the available control actions. This paper investigates the possibility of meeting this challenge in high-dimensional control tasks. We study the problem of learning the optimal *Linear Quadratic Regulator* (LQR) where the states live in a potentially infinite dimensional *Reproducing Kernel Hilbert Space* (RKHS), a linear control problem in which the dynamics, optimal value function, and optimal control policy are infinite dimensional, and therefore cannot be efficiently estimated to uniform precision. We focus on the regret setting, known as *online* LQR, in which a learner faces an unknown linear dynamical system and must adaptively tune a control policy to compete with the optimal policy.

Recent work has studied the statistical complexity of finite-dimensional LQRs at length (both in online and batch settings) [Dean et al., 2019, 2018, Mania et al., 2019, Faradonbeh et al., 2018, Simchowitz and Foster, 2020, Fazel et al., 2018, Tu and Recht, 2019]. However, all known results scale explicitly with the ambient dimension of the state space, which is infinite in our setting. In this work, we develop more fine-grained complexity measures to understand the hardness of linear control. While our measures are always crudely bounded by the state dimension, they behave more like an intrinsic dimension that adapts to the problem structure. Hence, they can be well-defined in many infinite dimensional RKHS settings as well, where they depend most strongly on the decay of the spectrum of the noise covariance. Such complexity measures based on intrinsic dimension are well-understood in supervised learning, and have been extended recently to the bandit and discrete action RL settings. Extending these ideas to continuous control is more challenging as several aspects of existing theory critically leverage the estimation of system parameters in operator norm, which necessitates an explicit dimension dependence. This motivates the following question:

> *Is it possible to obtain dimension-free sample complexity and regret guarantees in continuous control? What is the right measure of complexity?*

We answer the first question in the affirmative and show that the *spectral properties of the noise covariance*, such as its trace and eigenvalue decay, provide a much sharper characterization of the problem complexity than the state dimension in many parameter regimes.

1.1 Problem Setting & Background

We consider the problem of adaptive control of the linear quadratic regulator, or *online LQR*. To enable dimension-free results, we assume that the states $\mathbf{x}_t$ lie in a Hilbert space $\mathcal{H}_{\mathbf{x}}$. While states $\mathbf{x}_t$ are potentially infinite-dimensional, as shown in Theorem 1.1, it is necessary to assume the inputs $d_u \in \mathbb{R}^{d_u}$ are finite dimensional in order to guarantee sublinear regret. Given an arbitrary linear operator $X : \mathcal{H}_1 \to \mathcal{H}_2$ between Hilbert spaces, we let $\|X\|_{\mathrm{op}}$, $\|X\|_{\mathsf{HS}}$, $\|X\|_{\mathsf{tr}}$ denote its operator, Hilbert-Schmidt (HS), and trace norms. These norms may be infinite in general, and we say X is bounded, Hilbert Schmidt, or trace class if the corresponding norm is finite. We let X^{H} (resp. $\mathbf{x}^{\mathsf{H}}$) denote adjoints of operators (resp. vectors) and $\mathbf{x} \otimes \mathbf{x}$ denote outer products. The dynamics evolve according to:

$$\mathbf{x}_{t+1} = A_\star \mathbf{x}_t + B_\star \mathbf{u}_t + \mathbf{w}_t, \quad \mathbf{w}_t \overset{\text{i.i.d}}{\sim} \mathcal{N}(0, \Sigma_{\mathbf{w}}), \tag{1.1}$$

where $A_\star : \mathcal{H}_{\mathbf{x}} \to \mathcal{H}_{\mathbf{x}}$ and $B_\star : \mathbb{R}^{d_u} \to \mathcal{H}_{\mathbf{x}}$ are bounded linear operators, and $\Sigma_{\mathbf{w}}$ is *trace class* (i.e. $\mathrm{tr}\left[\Sigma_{\mathbf{w}}\right] = \mathbb{E}\|\mathbf{w}_t\|^2 < \infty$), self-adjoint, and PSD.[1] In LQR, the goal is to select a policy that minimizes cumulative quadratic losses $\langle \mathbf{x}_t, Q\mathbf{x}_t\rangle + \langle \mathbf{u}_t, R\mathbf{u}_t\rangle$, where Q, R are bounded, positive-definite operators. Given a bounded linear operator $K : \mathcal{H}_{\mathbf{x}} \to \mathbb{R}^{d_u}$, the infinite-horizon cost of the static feedback law $\mathbf{u}_t = K\mathbf{x}_t$ is

$$\mathcal{J}(K) := \lim_{T\to\infty} \frac{1}{T}\mathbb{E}\left[\textstyle\sum_{t=1}^T \langle \mathbf{x}_t, Q\mathbf{x}_t\rangle + \langle \mathbf{u}_t, R\mathbf{u}_t\rangle\right], \quad \text{subject to } \mathbf{u}_t = K\mathbf{x}_t. \tag{1.2}$$

We assume that $(A_\star, B_\star)$ is *stabilizable*, meaning there exists a controller K such that $\mathcal{J}(K)$ is finite, which is true if and only if the spectral radius $\rho(A_\star + B_\star K) := \limsup_{i\to\infty} \|(A_\star + B_\star K)^i\|_{\mathrm{op}}^{1/i} < 1$. We define the optimal control policy $K_\star := \inf_{K:\mathcal{H}_{\mathbf{x}}\to\mathbb{R}^{d_u}} \mathcal{J}(K)$. Under general conditions, $K_\star$ is unique, does not depend on the noise covariance $\Sigma_{\mathbf{w}}$, and its induced feedback law attains the optimal infinite horizon cost over *all* control policies. In the online LQR protocol, the system matrices $(A_\star, B_\star)$ are unknown, and the learner's goal is to adaptively learn to control the system so as attain low regret. We define the regret of a learning algorithm $\mathcal{A}$ (which chooses actions $\mathbf{u}_t$ based on the history of previous states and actions) as

$$\mathrm{Regret}_T(\mathcal{A}) := \left(\textstyle\sum_{t=1}^T \langle \mathbf{x}_t, Q\mathbf{x}_t\rangle + \langle \mathbf{u}_t, R\mathbf{u}_t\rangle\right) - T\mathcal{J}_\star, \quad \mathcal{J}_\star := \mathcal{J}(K_\star). \tag{1.3}$$

While we are not aware of other work which studies online control of LQR for infinite dimensional systems, this model has a long and rich history within the control theory community dating back at least to the 1970s (see for example Curtain and Zwart [2012], Bensoussan et al. [2007], and the references therein).

Dimension-Free Problem Parameter and Asymptotic Notation A central object in the analysis of LQR is the solution to a discrete algebraic Ricatti equation (DARE) $P_\star$ (see Eq. (2.1)), which represents the value function for the optimal controller (see Appendix A.4 for further details). The bounds in our setting

[1]The choice of Gaussian noise $\mathbf{w}_t$ is made for simplicity, our analysis can be easily extended to work for any stochastic, sub-Gaussian distribution.

are parameterized by the operator norms of $P_\star$, the system matrices $(A_\star, B_\star)$, and the noise covariance $\Sigma_{\mathbf{w}}$. These terms are considered dimension-free in prior literature (e.g. Mania et al. [2019], Simchowitz and Foster [2020]), and indeed do not scale with the dimension when the state dimension is finite. We define the quantity $M_\star$ as a uniform bound on these dimension free system parameters.

$$M_\star := \max\{\|A_\star\|_{\mathrm{op}}^2, \|B_\star\|_{\mathrm{op}}^2, \|P_\star\|_{\mathrm{op}}, \|\Sigma_{\mathbf{w}}\|_{\mathrm{op}}, 1\} \tag{1.4}$$

We use $a \lesssim b$ to denote that $a \le c \cdot b$, where c is a universal constant independent of any problem parameters. We let $\log_+(x) := \max\{\log(x), 1\}$ For a time horizon T, we use $\tilde{\mathcal{O}}(f(T))$ to denote a term that, for T sufficiently large, is bounded by $f(T)$ times logarithmic factors in relevant problem parameters. We define a weaker asymptotic notation $\mathcal{O}_\star(f(T))$ to denote a term bounded by $f(T)^{1+o(1)}$, times logarithmic factors, where $o(1) \to 0$ as $T \to \infty$.

1.2 The Challenges of Dimension-Free Linear Control

Though there is now a mature theory of dimension-free learning rates in prediction and online decision making [Bartlett and Mendelson, 2002, Zhang, 2005, Srinivas et al., 2010, Rakhlin and Sridharan, 2014], dimension-free rates in reinforcement learning have remained more elusive. This is because a learned model of transition dynamics that is accurate under one policy may be highly inaccurate on states visited under another policy.

Addressing policy mismatch in learning requires some handle on the complexity of the class of state distributions that a learner can encounter under available policies. Numerous strategies have been proposed, via both combinatorial quantities like Eluder Dimension [Russo and Van Roy, 2013] and linear-algebraic notions such as Bellman Rank [Jiang et al., 2017]. Targeting dimension-free rates more specifically, recent work has studied MDPs with linear transitions, where the parameters lie in an RKHS [Yang et al., 2020b, Yang and Wang, 2020, Agarwal et al., 2020]. These developments assume that the dynamics can be factorized as the inner product of two vectors that have bounded RKHS norm. The dynamical matrices that arise in LQR, however, are considerably richer objects: *bounded operators* on the RKHS, rather than mere elements of it, and this leads to fundamental differences between the settings. In a similar vein, Kakade et al. [2020] consider a nonlinear dynamical model with kernelized dynamics, finite-dimensional state, and well-conditioned Gaussian noise. Their setting is in general incomparable to ours, yet in the finite-dimensional LQR setting in which we overlap, our techniques yield more refined bounds. See the discussion following Theorem 3.2 for further comparison.

It is not obvious that dimension-free learning in LQR is even possible. In fact, we show that the worst-case regret necessarily scales with the *ambient input dimension*:

Theorem 1.1. (informal) Fix any integer $r_{\mathbf{x}} \ge 1$, and input dimension d_u. Consider identity costs $Q = I_{\mathcal{H}_{\mathbf{x}}}$, $R = I_{d_u}$, and noise covariance $\Sigma_{\mathbf{w}}$ with trace $\mathrm{tr}[\Sigma_{\mathbf{w}}] \le r_{\mathbf{x}}$. Then, there exists a family of stabilizable instances (A, B) with Hilbert-Schmidt norm bounded by 2 such that any algorithm must suffer $\Omega(T)$ regret for $T \le r_{\mathbf{x}} d_u^2$, and $\Omega(\sqrt{T r_{\mathbf{x}} d_u^2})$ regret thereafter.

The formal statement of the lower bound, its proof, and further discussion are given in Appendix H. Notably, the above theorem stands in stark contrast to analogous results for linear MDPs and those presented in Kakade et al. [2020], which apply to infinite dimensional inputs.

Addressing high-dimensional states in LQR has posed a challenge in both theory and practice [Sagaut, 2006, Liu and Vandenberghe, 2010]. All relevant prior work has incurred a dependence on the ambient state dimension. Model-based methods, which estimate system parameters and propose a policy based on those estimates, have required consistent recovery of those parameters (say, in operator or Frobenius norm), which is far stronger than a prediction error guarantee [Dean et al., 2018, 2019, Mania et al., 2019, Cohen et al., 2019]. Model-free methods, which eschew learning the system parameters in favor of directly optimizing the policy or value function, have been observed to suffer an even worse dependence on ambient dimension [Tu and Recht, 2019]. To summarize, dimension-free statistical learning encounters major obstacles when translated to linear control.

1.3 Summary of Results

Certainty Equivalence Our results are based on *certainty equivalence*, [Theil, 1957, Simon, 1956], first analyzed for the online LQR setting by Mania et al. [2019]. Given a stabilizable (A, B), we let $K_\infty(A, B)$ denote the optimal controller K, which minimizes the cost functional in Eq. (1.2) with $(A_\star, B_\star)$ set to (A, B). It is a well-known fact that $K_\infty(A, B)$ has a closed form expression in terms of the system parameters (A, B) and the DARE (see formal preliminaries in Section 2). Given estimates $(\widehat{A}, \widehat{B})$ of $(A_\star, B_\star)$, the *certainty equivalence controller* is $\widehat{K} = K_\infty(\widehat{A}, \widehat{B})$; that is, the optimal control policy as if the true system were $(\widehat{A}, \widehat{B})$. To be well-posed, this controller requires $(\widehat{A}, \widehat{B})$ to be stabilizable, which occurs when $(\widehat{A}, \widehat{B})$ is sufficiently close to $(A_\star, B_\star)$ in operator norm (see, e.g., Proposition C.3).

A Regret Bound We analyze a simple explore-then-commit style algorithm, OnlineCE, based on certainty equivalence, similar to that of Mania et al. [2019]. For now, we assume access to initial "warm-start" estimates (A_0, B_0) whose distance from $(A_\star, B_\star)$ in operator norm is a small constant. We show how to get rid of this assumption later. The algorithm proceeds by first synthesizing an exploratory controller $K_0 = K_\infty(A_0, B_0,)$, and then collecting T_{exp} steps of samples with inputs $\mathbf{u}_t = K_0\mathbf{x}_t + \mathbf{v}_t$, where $\mathbf{v}_t$ is i.i.d. Gaussian noise injected for exploration. In the second phase, the algorithms constructs refined estimates $(\widehat{A}, \widehat{B})$ by performing ridge regression on the collected data, synthesizes the certainty equivalence controller $\widehat{K} = K_\infty(\widehat{A}, \widehat{B})$, and selects inputs $\mathbf{u}_t = \widehat{K}\mathbf{x}_t$ for the remainder of the protocol.

In Theorem 3.1, we demonstrate that this relatively simple algorithm enjoys regret that scales polynomially with the eigendecay of the noise covariance $\Sigma_{\mathbf{w}}$, Hilbert-Schmidt norm of $A_\star$, input dimension d_u, and the operator norms of relevant system operators. When specialized to common rates of eigendecay, we attain the following regret bounds.

Theorem 3.2 (informal). Let σ_j be the eigenvalues of $\Sigma_{\mathbf{w}}$. If the initial estimates (A_0, B_0) are sufficiently close to $(A_\star, B_\star)$, then OnlineCE suffers regret at most:

- (polynomial decay) $\mathcal{O}_\star(\sqrt{\mathcal{C}_P d_{\max}^2 T^{1+1/\alpha}})$, if $\sigma_j = j^{-\alpha}$ for $\alpha > 1$,
- (exponential decay) $\mathcal{O}_\star(\sqrt{\mathcal{C}_P d_{\max}^2 d_u T})$, if $\sigma_j = \exp(-\alpha j)$ for $\alpha > 0$.
- (finite dimension) $\widetilde{\mathcal{O}}(\sqrt{\mathcal{C}_P (d_u + d_x)^3 T})$ if $\mathcal{H}_{\mathbf{x}} = \mathbb{R}^{d_x}$.

In the above expressions, $\mathcal{C}_P$ is a polynomial in $M_\star$ and $d_{\max} := \max\{\mathrm{tr}\left[\Sigma_{\mathbf{w}}\right], d_u, W_{\mathrm{tr}}\}$ where W_{tr} is slightly larger than $\mathrm{tr}\left[\Sigma_{\mathbf{w}}\right] + \|B_\star\|_{\mathsf{HS}}^2$

Interestingly, this result shows that achieving dimension-free rates for online linear control does not require new algorithmic ideas, but rather a refined analysis of classical ones, like certainty equivalence. In particular, when specialized to finite dimension, our regret bound has the same dimension dependence as the one of Mania et al. [2019]. Simchowitz and Foster [2020] show that this dependence is sharp in the regime where state and input spaces have the same dimension. More generally, however, the spectrum of $\Sigma_{\mathbf{w}}$ is a significantly sharper complexity measure as indicated before and evidenced in the following example.

An Illustrative Example Let $(\mathbf{e}_i)_{i=1}^\infty$ be an orthonormal basis for $\mathcal{H}_{\mathbf{x}}$, $(\mathbf{f}_i)_{i=1}^{d_u}$ an orthonormal basis for $\mathbb{R}^{d_u}$, and consider the following problem instance:

$$A_\star = \frac{1}{2}\sum_{i=1}^{d_u} \mathbf{e}_i \otimes \mathbf{e}_i + \sum_{i>d_u} \frac{1}{i^2} \cdot \mathbf{e}_i \otimes \mathbf{e}_i, \quad B_\star = \sum_{i=1}^{d_u} \mathbf{e}_i \otimes \mathbf{f}_i, \quad \Sigma_{\mathbf{w}} = \sum_{i=1}^{\infty} \frac{1}{i^2} \cdot \mathbf{e}_i \otimes \mathbf{e}_i,$$

where $Q = I_{\mathcal{H}_{\mathbf{x}}}$ and $R = I_{d_u}$. In this example, $A_\star$ is infinite-dimensional, yet only has a finite dimensional controllable subspace. Therefore, in order to learn the optimal policy, it is not necessary for the learner to estimate the whole system, but rather the parts of it that are relevant for control as determined by noise and the controllability properties of $A_\star$ and $B_\star$. While guarantees from previous work are vacuous in this setting since the ambient system dimension is infinite, the example corresponds to $\alpha = 2$ in case 1 of Theorem 3.2, with $d_{\max} = O(d_u)$ and $M_\star \leq 1$, yielding an $\mathcal{O}_\star(d_u T^{3/4})$ regret bound based on our theory.

Suboptimality Bounds from Prediction Error During the exploration phase, the OnlineCE algorithm selects inputs according to $\mathbf{u}_t = K_0\mathbf{x}_t + \mathbf{v}_t$. Let $\Sigma_{\mathbf{x},0}$ denote the stationary covariance over states induced by this policy (see Eq. (2.3) for details). The ridge regression step in OnlineCE recovers $B_\star$ in Hilbert-Schmidt norm (since $B_\star$ is finite rank), but recovers $A_\star$ only in the *covariance-weighted* Hilbert-Schmidt norm $\|(A_\star - \widehat{A})\Sigma_{\mathbf{x},0}^{1/2}\|_{\mathsf{HS}}$, corresponding to the prediction error. The key technical innovation in this paper that underlies all of our results is a perturbation bound on the suboptimality of the certainty equivalence controller $\widehat{K}$ in terms of this prediction error.

Theorem 2.1 (informal). Let K_0 be any state-feedback controller that stabilizes $(A_\star, B_\star)$, and let $\Sigma_{\mathbf{x},0}$ denote the induced state covariance with $\sigma_{\mathbf{u}}^2 = 1$. Then, if $(\widehat{A}, \widehat{B})$ are within a small but constant operator norm error of $(A_\star, B_\star)$,

$$\mathcal{J}(\widehat{K}) - \mathcal{J}_\star \leq \mathcal{C}_J \cdot (\varepsilon_{\mathrm{cov}})^{2-o(1)}, \text{ where } \varepsilon_{\mathrm{cov}} := \max\left\{\|(\widehat{A} - A_\star)\Sigma_{\mathbf{x},0}^{1/2}\|_{\mathsf{HS}}, \|\widehat{B} - B_\star\|_{\mathsf{HS}}\right\}.$$

Here, $\mathcal{C}_J$ is a polynomial in $M_\star$ and $o(1)$ denotes a term that tends to 0 as $\varepsilon_{\mathrm{cov}} \to 0$.

Importantly, a $\varepsilon_{\mathrm{cov}}^2$ scaling of the perturbation is known to be optimal [Mania et al., 2019], and in finite dimensions, the $o(1)$ term in the exponent of our bound can be discarded.

Alignment Condition As stated, the regret guarantee guarantee for OnlineCE requires access to a warm-start estimates (A_0, B_0). We use this to perform a projection step ensuring that the certainty equivalent controller $\widehat{K}$ is stabilizing. While this condition is stronger than those that have previously appeared in the literature, which typically assume that the learner initially has access to an *arbitrary* stabilizing controller K_{init}, we show that, under a certain alignment condition, it is possible to achieve this warm start condition:

Proposition 3.1 (informal). Under a certain alignment condition, which holds if all eigenvalues of $\Sigma_{\mathbf{w}}$ are strictly positive (though decaying to zero, see Assumption 3), after collecting $\mathcal{O}(1)$ many samples, ridge regression returns estimates (A_0, B_0) of $(A_\star, B_\star)$ satisfying the requisite closeness condition for Theorem 3.1.

One can stitch together an initial estimation phase described by the theorem above with the analysis of OnlineCE to provide an algorithm that dispenses with the access to warm-start estimates, yet this requires the above alignment condition. Crucially, we use this condition for coarse recovery up to a constant tolerance. Hence, the initial estimation phase adds only a constant burn-in to the regret. Importantly, the alignment condition *does not* afford us consistent parameter recovery.

1.4 Related Work

In the interest of brevity, we provide an abridged discussion of related work here, and defer an extended discussion to Appendix A.2. The learning community has seen a surge in interest in linear control and in system identification [Vidyasagar and Karandikar, 2006, Hardt et al., 2018, Pereira et al., 2010, Oymak and Ozay, 2019, Simchowitz et al., 2018, Sarkar and Rakhlin, 2019, Dean et al., 2019]. We consider the online LQR setting first proposed by Abbasi-Yadkori and Szepesvári [2011], and subsequently studied by Faradonbeh et al. [2018], Cohen et al. [2019], Dean et al. [2018], Mania et al. [2019], Abeille and Lazaric [2020], Simchowitz et al. [2020]. Our work is based on the analysis of certainty equivalence for online control, first studied by Mania et al. [2019] and refined by Simchowitz and Foster [2020]. Concurrent work has also studied *model-free* approaches for control [Fazel et al., 2018, Krauth et al., 2019, Abbasi-Yadkori et al., 2019, Tu and Recht, 2019]. As noted, all analyses incur dependence on ambient system dimension.

Sample complexity guarantees depending on instrinsic measures of complexity (rather than ambient dimension) are well-known in supervised learning [Bartlett and Mendelson, 2002, Zhang, 2005] and bandit problems [Srinivas et al., 2010]. More recently, these results have been extended to the reinforcement learning literature as well, for a class of problems defined as linear MDPs [Jin et al., 2020, Agarwal et al., 2020, Yang et al., 2020a]. Further discussion comparing the linear MDP regime to LQR is deferred to Appendix A.2.

2 From Prediction Error Bounds to Controller Suboptimality

In this section, we establish the main perturbation bounds regarding the suboptimality of a certainty-equivalent controller $\widehat{K} = K_\infty(\widehat{A}, \widehat{B})$ given estimates $(\widehat{A}, \widehat{B})$ that satisfy a prediction error bound under a particular exploratory distribution. In doing so, we highlight a key change of measure lemma that allows us to evaluate the behavior of the system under *any* state-feedback law given only a prediction error bound under a single exploratory policy. We begin by stating some further preliminaries.

Formal Preliminaries As in finite-dimensional settings, the optimal controller $K_\infty(A, B)$ for LQR in infinite dimension can be computed in terms of the PSD operator $P_\infty(A, B) : \mathcal{H}_{\mathbf{x}} \to \mathcal{H}_{\mathbf{x}}$ which solves the Discrete Algebraic Riccati Equation (DARE),[2]

$$P_\infty(A, B) \quad \text{solves} \quad P = A^{\mathsf{H}} P A - A^{\mathsf{H}} P B (R + B^{\mathsf{H}} P B)^{-1} B^{\mathsf{H}} P A + Q. \tag{2.1}$$

$$K_\infty(A, B) \quad := -(R + B^{\mathsf{H}} P B)^{-1} B^{\mathsf{H}} P A, \quad \text{where } P = P_\infty(A, B) \tag{2.2}$$

We define $P_\star := P_\infty(A_\star, B_\star)$ and recall $K_\star := K_\infty(A_\star, B_\star)$.

The DARE is intimately related to the discrete Lyapunov operator, dlyap. Given a bounded linear operator $A : \mathcal{H}_{\mathbf{x}} \to \mathcal{H}_{\mathbf{x}}$ that is stable (i.e. $\rho(A) < 1$), and a symmetric bounded operator $\Lambda : \mathcal{H}_{\mathbf{x}} \to \mathcal{H}_{\mathbf{x}}$, $\mathsf{dlyap}\,(A,\ \Lambda)$ denotes the solution to the equation $X = A^{\mathsf{H}} X A + \Lambda$. A classic result in Lyapunov theory states that the solution X is unique, and is given by $\mathsf{dlyap}\,(A,\ \Lambda) = \sum_{j=0}^{\infty} (A^{\mathsf{H}})^j \Lambda A^j$. For any controller K such that K is stabilizing for (A, B) we define $P_\infty(K; A, B) := \mathsf{dlyap}\left(A + BK,\ Q + K^{\mathsf{H}} R K\right)$, which can be viewed as the value function induced by the controller K (see Appendix A.4 for details). Two consequences of this interpretation are that $P_\infty(K; A, B) = P_\infty(A, B)$ for $K = K_\infty(A, B)$, and $P_\infty(A, B) \preceq P_\infty(K'; A, B)$ for any other stabilizing controller K'.

We adopt the following notation to refer to the steady-state covariance operator for the true system $(A_\star, B_\star)$, where $\mathbf{u}_t$ is chosen by combining a state feedback policy K with isotropic Gaussian noise $\mathbf{v}_t$:

$$\Sigma_\star(K, \sigma_{\mathbf{u}}^2) := \lim_{t \to \infty} \mathbb{E}[\mathbf{x}_t \otimes \mathbf{x}_t], \text{ s.t. } \mathbf{x}_{t+1} = A_\star \mathbf{x}_t + B_\star \mathbf{u}_t + \mathbf{w}_t,$$
$$\text{where } \mathbf{u}_t = K\mathbf{x}_t + \mathbf{v}_t, \quad \mathbf{w}_t \overset{\text{i.i.d}}{\sim} \mathcal{N}(0, \Sigma_{\mathbf{w}}), \mathbf{v}_t \overset{\text{i.i.d}}{\sim} \mathcal{N}(0, \sigma_{\mathbf{u}}^2 I). \tag{2.3}$$

We let $\Sigma_\star(K) = \Sigma_\star(K, 0)$. A short calculation reveals that,

$$\Sigma_\star(K, \sigma_{\mathbf{u}}^2) = \mathsf{dlyap}\left((A_\star + B_\star K)^{\mathsf{H}},\ \Sigma_{\mathbf{w}} + \sigma_{\mathbf{u}}^2 B_\star B_\star^{\mathsf{H}}\right).$$

Lastly, for the remainder of the presentation, we make the following assumption on the costs:

Assumption 1. The cost operators Q, R satisfy $Q, R \succ I$.

Since scaling both Q and R by a constant does not change the form of the optimal controller, this assumption is without loss of generality if the operators are already positive definite.

2.1 Performance Difference and Change of Measure

The suboptimality of a controller K for an instance $(A_\star, B_\star)$ admits the following closed form, often referred to as the *performance-difference lemma* [Fazel et al., 2018]:

$$\begin{aligned}\mathcal{J}(K) - \mathcal{J}(K_\star) &= \operatorname{tr}\left[(R + B_\star^{\mathsf{H}} P_\star B_\star)^\top \cdot (K - K_\star)\Sigma_\star(K)(K - K_\star)^{\mathsf{H}}\right] \\ &\leq \left\|R + B_\star^{\mathsf{H}} P_\star B_\star\right\|_{\mathrm{op}} \|(K - K_\star)\Sigma_\star(K)^{\frac{1}{2}}\|_{\mathsf{HS}}^2. \end{aligned} \tag{2.4}$$

Hence, the correct geometry in which K should approximate $K_\star$ is in the HS norm, weighted by its steady-state covariances $\Sigma_\star(K)$. Weighting by the $\Sigma_\star(K)^{\frac{1}{2}}$ is crucial for dimension-free bounds since recovery of $K_\star$ in the unweighted HS norm would incur dependence on ambient dimension.

[2] See for example Zabczyk [1974, 1975], Lee et al. [1972], Curtain and Zwart [2012].

One could achieve low error in the $\Sigma_\star(K)^{\frac{1}{2}}$-weighted norm if one already had access to samples with covariance $\Sigma_\star(K)$, but this logic becomes circular. Instead, we ensure that $\|(K - K_\star)\Sigma_{\mathbf{x},0}^{1/2}\|_{\mathsf{HS}}^2$ is small, where $\Sigma_{\mathbf{x},0} = \Sigma_\star(K_0, \sigma_{\mathbf{u}}^2)$ is the state covariance under an arbitrary stabilizing controller K_0 and some additional Gaussian excitation. Our first technical contribution shows that, up to constant factors, the exploratory covariance $\Sigma_{\mathbf{x},0}$ dominates the target $\Sigma_\star(K)$:

Lemma 2.1. *Let K be any stabilizing controller for $(A_\star, B_\star)$ and let $\Sigma_\star(K)$ be its induced state covariance. Then, for any stabilizing controller K_0 and any variance $\sigma_{\mathbf{u}}^2 \geq 1$, we have that $\Sigma_{\mathbf{x},0} = \Sigma_\star(K_0, \sigma_{\mathbf{u}}^2)$ satisfies*

$$\Sigma_\star(K) \preceq \mathcal{C}_{K,\sigma_{\mathbf{u}}^2} \cdot \Sigma_{\mathbf{x},0},$$

where $\mathcal{C}_{K,\sigma_{\mathbf{u}}^2} = \max\left\{2, \frac{128}{\sigma_{\mathbf{u}}^2} \|\Sigma_{\mathbf{w}}\|_{\mathrm{op}} \|K - K_0\|_{\mathrm{op}}^2 \|P_K\|_{\mathrm{op}}^3 \log\left(3\|P_K\|_{\mathrm{op}}\right)^2\right\}$ *for* $P_K = P_\infty(K; A_\star, B_\star)$.

Hence, it suffices to replace the performance difference bound in Eq. (2.4) with an estimate in the norm induced by the exploratory covariance. More specifically, we have that,

$$\|(K - K_\star)\Sigma_\star(K)^{\frac{1}{2}}\|_{\mathsf{HS}}^2 \;\leq\; \mathcal{C}_{K,\sigma_{\mathbf{u}}^2}\|(K - K_\star)\Sigma_{\mathbf{x},0}^{1/2}\|_{\mathsf{HS}}^2,$$

therefore ensuring good performance under the exploratory distribution is enough to ensure good performance under *any* other induced distribution.

In finite dimensions with full-rank noise $\lambda_{\min}(\Sigma_{\mathbf{w}}) > 0$, the above comparison follows quite directly. The key challenge in our setting is ruling out the possibility that one controller K "pushes" large eigenvalues of $\Sigma_{\mathbf{w}}$ into one region of the state space, in which the other controller K_0 induces small excitation. The key insight is that the closed loop systems $A_\star + B_\star K$ and $A_\star + B_\star K_0$ differ only along the column space of $B_\star$, and these directions are excited in the $\Sigma_{\mathbf{x},0}$ covariance due to the injection of Gaussian noise. Importantly, the bound holds *without* further controllability assumptions, which may fail in infinite dimensions.

2.2 A Dimension-Free Perturbation Bound

Building on the above insight, we show that it suffices to have a prediction error bound (i.e estimate $A_\star$ in the $\Sigma_{\mathbf{x},0}$-induced HS norm) in order to synthesize a close to optimal controller. Since d_u is finite, we can recover $B_\star$ in the unweighted HS norm. Specifically, fix a stabilizing controller K_0, and set $\Sigma_{\mathbf{x},0} := \Sigma_\star(K_0, \sigma_{\mathbf{u}}^2)$. Now, define the error terms,

$$\varepsilon_{\mathrm{cov}} = \max\{\|(\widehat{A} - A_\star)\Sigma_{\mathbf{x},0}^{1/2}\|_{\mathsf{HS}}, \|\widehat{B} - B_\star\|_{\mathsf{HS}}\}, \quad \varepsilon_{\mathsf{op}} := \max\{\|\widehat{A} - A_\star\|, \|\widehat{B} - B_\star\|\}. \tag{2.5}$$

Here, $\varepsilon_{\mathrm{cov}}$ corresponds to the relevant HS norms (weighted for $A_\star$, unweighted for $B_\star$). In addition to $\varepsilon_{\mathrm{cov}}$, we also consider a uniform (unweighted) operator norm bound error $\varepsilon_{\mathsf{op}}$. This error needs to be smaller than some problem dependent constant to ensure that $(\widehat{A}, \widehat{B})$ is stabilizable, and that $\widehat{K} = K_\infty(\widehat{A}, \widehat{B})$ stabilizes $(A_\star, B_\star)$. To this end, our perturbation bound imposes the following condition:

Condition 2.1. The error $\varepsilon_{\mathsf{op}}$ defined in Eq. (2.5) satisfies $\varepsilon_{\mathsf{op}} \leq \mathcal{C}_{\mathsf{stable}} := 1/229\,\|P_\star\|_{\mathrm{op}}^3$.

We now state the main perturbation bound. We assume $1 \leq \sigma_{\mathbf{u}}^2 \lesssim 1$, as in our online algorithm.

Theorem 2.1. *Let $\sigma_{\mathbf{u}}^2 \geq 1$, $K_0 = K_\infty(A_0, B_0)$, and $\Sigma_{\mathbf{x},0} := \Sigma_\star(K_0, \sigma_{\mathbf{u}}^2)$. If (A_0, B_0), $(\widehat{A}, \widehat{B})$ both satisfy Condition 2.1, then, for $\widehat{K} := K_\infty(\widehat{A}, \widehat{B})$,*

$$\mathcal{J}(\widehat{K}) - \mathcal{J}(K_\star) \lesssim \sigma_{\mathbf{u}}^4 M_\star^{36} \cdot \mathcal{L}\exp(\tfrac{1}{50}\sqrt{\mathcal{L}}) \cdot \varepsilon_{\mathrm{cov}}^2, \qquad \textit{where } \mathcal{L} := \log\left(e + \frac{2e\|\widehat{A} - A_\star\|_{\mathrm{op}}^2 \mathrm{tr}[\Sigma_{\mathbf{x},0}]}{\varepsilon_{\mathrm{cov}}^2}\right),$$

and $M_\star$ is defined as in Eq. (1.4). Moreover, in finite dimensions with $\Sigma_{\mathbf{x},0} \succ 0$, $\mathcal{L}$ can be replaced by $\log(1 + \mathrm{cond}(\Sigma_{\mathbf{x},0}))$, *where* $\mathrm{cond}(\cdot)$ *denotes condition number.*

The main novelty of the above perturbation bound, relative to previous analysis of certainty equivalent control is that, assuming that $\varepsilon_{\mathsf{op}}$ is smaller than a constant[3], the suboptimality gap $\mathcal{J}(\widehat{K}) - \mathcal{J}(K_\star)$ depends only on the weighted HS norm or prediction error $\varepsilon_{\mathrm{cov}}^2$. In particular, we observe that $\mathcal{L}\exp(\frac{1}{50}\sqrt{\mathcal{L}})$ grows more slowly as $\varepsilon_{\mathrm{cov}} \to 0$ than any power $(\varepsilon_{\mathrm{cov}})^\alpha$; hence, for any $\alpha > 0$, there is some c_α such that $\mathcal{J}(\widehat{K}) - \mathcal{J}(K_\star) \leq c_\alpha M_\star^{36} \varepsilon_{\mathrm{cov}}^{2-\alpha}$.

Furthermore, in the finite dimensional setting with full rank noise covariance, the condition number of $\Sigma_{\mathbf{x},0}$ is bounded; thus, $\mathcal{L} = \mathcal{O}(1)$ in this regime, and the scaling is exactly $\varepsilon_{\mathrm{cov}}^2$, which is known to be optimal [Mania et al., 2019, Simchowitz et al., 2020]. Moreover, Theorem 2.1 depends only on the operator norm of natural control theoretic quantities, and hides no dimension like terms. Hence, up to $M_\star$ dependence, it matches the best-possible perturbations bounds attainable in the finite dimensional setting. Lastly, little effort was made in sharpening the dependence on $M_\star$, which we believe can be refined considerably.

Proof Sketch of Theorem 2.1. The proof uses arguments in Section 2.1 to reduce to bounding $\|(K - K_\star)\Sigma_{\mathbf{x},0}^{1/2}\|_{\mathsf{HS}}$. A direct computation (Proposition C.2) shows that this term can be bounded in terms of $\varepsilon_{\mathrm{cov}}$, and the weighted error $\|\Sigma_{\mathbf{x},0}^{1/2}(\widehat{P} - P_\star)\Sigma_{\mathbf{x},0}^{1/2}\|_{\mathsf{HS}}$, where $\widehat{P} := P_\infty(\widehat{A}, \widehat{B})$ is the certainty equivalent value function. In Proposition C.4, we bound the latter by considering a linear interpolating curve $(A(t), B(t))$, $t \in [0, 1]$ between $(A_\star, B_\star)$ and $(\widehat{A}, \widehat{B})$, and the value function $P(t) = P_\infty(A(t), B(t))$ along that curve. $P(t)$ can be shown to be continuously differentiable for all $t \in [0, 1]$ under Condition 2.1, and hence it suffices to bound the supremum of $\|\Sigma_{\mathbf{x},0}^{1/2} P'(t) \Sigma_{\mathbf{x},0}^{1/2}\|_{\mathsf{HS}}$ over $t \in [0, 1]$. Bounding this term requires the majority of our technical effort, and relies on both the change-of-measure bounds similar to Lemma 2.1, and a careful application of the self-bounding ODE method introduced by Simchowitz and Foster [2020]. The full proof of the end-to-end bound, and all its constituent results, is outlined in Appendix C. $\square$

3 Algorithms & Regret Bounds

Having concluded our discussion of certainty equivalence, in this section we now leverage our earlier results to prove that a simple explore-then-commit style algorithm, $\mathsf{OnlineCE}$, achieves sublinear regret for LQR in infinite dimension.

In order to more clearly communicate the salient features of our analysis, we divide this section into two parts. First, we prove a regret bound assuming the learner has access to initial system estimates (A_0, B_0) such that these operators lie within some problem dependent constant of the true system operators $(A_\star, B_\star)$. The following subsection shows how under an appropriate alignment condition, one can incorporate an initial phase to the algorithm that achieves these warm start estimates (A_0, B_0) while only adding a constant term (in T) to the overall regret.

3.1 Regret with Warm-Start

We now describe the $\mathsf{OnlineCE}$ algorithm assuming access to *warm-start* estimates (A_0, B_0) of the true system parameters $(A_\star, B_\star)$. In particular, we assume

Condition 3.1 (Warm Start)**.** The pair (A_0, B_0) satisfy $\max\{\|A_0 - A_\star\|_{\mathrm{op}}, \|B_0 - B_\star\|_{\mathrm{op}}\} \leq 1/2 \cdot \mathcal{C}_{\mathsf{stable}}$.

Under a minor technical extension, $\mathcal{C}_{\mathsf{stable}} = 1/(229\,\|P_\star\|_{\mathrm{op}}^3)$ can be replaced by a data dependent quantity $\lesssim \|P_\infty(A_0, B_0)\|_{\mathrm{op}}^{-3}$, which can be can be verified using only a confidence interval around $(A_\star, B_\star)$. For simplicity, we present the main algorithm without this modification, which we defer to Appendix F.1. Given these estimates, our algorithm, $\mathsf{OnlineCE}$, consists of the following explore-then-commit strategy:

1. Synthesize $K_0 := K_\infty(A_0, B_0)$ and choose inputs $\mathbf{u}_t = K_0\mathbf{x}_t + \mathbf{v}_t$ where $\mathbf{v}_t \sim \mathcal{N}(0, I)$ $(\sigma_{\mathbf{u}}^2 = 1)$ for T_{exp} many iterations, collecting observations $\{(\mathbf{x}_t, \mathbf{v}_t)\}_{t=1}^{T_{\mathsf{exp}}}$.

[3] This condition can be relaxed for (A_0, B_0); it suffices that K_0 is an arbitrary, stabilizing controller for $(A_\star, B_\star)$.

2. Compute system estimates $(\widehat{A}, \widehat{B})$ via ridge regression on data $\{(\mathbf{x}_t, \mathbf{v}_t)\}_{t=1}^{T_{\mathsf{exp}}}$ followed by a projection onto a safe set around warm start estimates (A_0, B_0) (see Appendix F for full pseudocode).

3. Synthesize $\widehat{K} = K_\infty(\widehat{A}, \widehat{B})$ and choose inputs $\mathbf{u}_t = \widehat{K}\mathbf{x}_t$ for the remaining time steps.

Throughout, we let $\sigma_j(\Lambda)$ denote the j-th largest eigenvalue of a PSD operator Λ. Furthermore, we define $W_{\mathrm{tr}} := \|B_\star\|_{\mathsf{HS}}^2 + \sum_{j=1}^{\infty} \sigma_j(\Sigma_{\mathbf{w}}) \log(j)$, which captures the magnitude of noise in the system under an exploratory policy. Note that for W_{tr} to be finite, $\Sigma_{\mathbf{w}}$ needs to be slightly stronger than trace class. We recall our earlier asymptotic notation: $\widetilde{\mathcal{O}}(f(T))$ suppresses logarithms, and $\mathcal{O}_\star(f(T)) := \widetilde{\mathcal{O}}(f(T)^{1+o(1)})$ where $o(1) \to 0$ as T goes to infinity.

We now state our main regret bound for $\mathsf{OnlineCE}$. For simplicity, we assume that the initial state is drawn from the steady state covariance, $\mathbf{x}_1 \sim \mathcal{N}(0, \Sigma_{\mathbf{x},0})$ (see Appendix G.4 for further discussion).

Theorem 3.1. *Let $(\sigma_j)_{j=1}^\infty = (\sigma_j(\Sigma_{\mathbf{x},0}))_{j=1}^\infty$ be the eigenvalues of $\Sigma_{\mathbf{x},0}$ and define,*

$$d_\lambda := |\{\sigma_j : \sigma_j \ \geq \ \lambda\}|, \quad \mathcal{C}_{\mathsf{tail},\lambda} := \tfrac{1}{\lambda}\textstyle\sum_{j>d_\lambda} \sigma_j.$$

If the learner has access to warm start estimates satisfying Condition 3.1, with probability $1-\delta$, $\mathsf{OnlineCE}$ satisfies

$$\mathrm{Regret}_T(\mathsf{OnlineCE}) \ \leq \ \mathcal{O}_\star\left(\sqrt{M_\star^{42} d_{\max}^2 (d_\lambda + \mathcal{C}_{\mathsf{tail},\lambda}) T}\right)$$

where $M_\star$ is as in Eq. (1.4) and $d_{\max} := \max\{\mathrm{tr}\left[\Sigma_{\mathbf{w}}\right], d_u, W_{\mathrm{tr}}\}$.

We state this first theorem in terms of the eigenvalues of $\Sigma_{\mathbf{x},0}$, but the bounds can also be stated in terms of $\Sigma_{\mathbf{w}}$ under particular eigenvalue decay assumptions. We carry out this translation in Appendix D.2 through novel eigenvalue comparison inequalities for Lyapunov operators which may be of independent interest. In particular, the following result formalizes our earlier statement about the spectrum of $\Sigma_{\mathbf{w}}$ encoding the right problem complexity as we remarked in Section 1.3.

Theorem 3.2. *Let $(\sigma_j(\Sigma_{\mathbf{w}}))_{j=1}^\infty$ be the (descending) eigenvalues of $\Sigma_{\mathbf{w}}$. In the setting of Theorem 3.1, the* $\mathsf{OnlineCE}$ *algorithm suffers regret at most,*

- *(polynomial decay) $\mathcal{O}_\star\left(\sqrt{M_\star^{46} d_{\max}^2 T^{1+1/\alpha}}\right)$, if $\sigma_j(\Sigma_{\mathbf{w}}) = j^{-\alpha}$ for $\alpha > 1$,*
- *(exponential decay) $\mathcal{O}_\star\left(\sqrt{M_\star^{45} d_{\max}^2 d_u T}\right)$, if $\sigma_j(\Sigma_{\mathbf{w}}) = \exp(-\alpha j)$ for $\alpha > 0$.*
- *(finite dimension) $\widetilde{\mathcal{O}}\left(\sqrt{M_\star^{42}(d_u + d_x)^3 T}\right)$, if $\mathcal{H}_{\mathbf{x}} = \mathbb{R}^{d_x}$ and $\Sigma_{\mathbf{w}} = I$.*

For finite dimensional systems, the extra terms depending on $\mathcal{L}$ become $\mathcal{O}(1)$, and we achieve the optimal $\sqrt{d_{\max}^3 T}$ regret, which is optimal in the regime where $d_x \asymp d_u$ [Simchowitz and Foster, 2020]. The results for polynomial and exponential decay illustrate how the ambient state dimension d_x is only a very coarse measure of complexity for linear control. Our analysis of certainty equivalence shows that the magnitude of the system noise, $\mathrm{tr}\left[\Sigma_{\mathbf{w}}\right]$, is a more accurate measure of problem hardness. While this measure is $\Omega(d_x)$ in the case of a full rank noise covariance, it can be considerably smaller as per our example in the introduction.

Comparison to Kakade et al. [2020] The authors consider a setting where the dynamics are kernelized into a feature map, and the dynamical parameters are linear operators in the kernel space. Their model requires a finite dimensional state with well-conditioned Gaussian noise. Furthermore, their bounds scale with the ambient state dimension (via the optimal control cost $J^\star$), even if their associated intrinsic kernel dimension is $\mathcal{O}(1)$. It is unclear how to extend their techniques to high-dimensional noise with spectral decay, due to the subtleties of their change-of-measure argument (Lemma 3.9).

Moreover, the techniques in this paper (Lemma 2.1 and Lemma D.15) can also be used to attain refined bounds on their algorithm-dependent intrinsic dimension quantity, $\gamma_T(\lambda)$, which improves on worst-case analyses based on kernel structure.

3.2 Finding warm start estimates

We now move on to discussing how to achieve initial warm start estimates that satisfy Condition 3.1. To do so, we require two additional assumptions. The first is standard within the online control literature (see for example Dean et al. [2018], Mania et al. [2019], Cohen et al. [2019]):

Assumption 2 (Initial Controller)**.** The learner has initial access to a controller K_{init} that stabilizes $(A_\star, B_\star)$.

The second is an alignment condition specific to our setting. Before stating it, we define $A_{\mathrm{cl}_\star} := (A_\star + B_\star K_{\mathsf{init}})$ and let $U(\Lambda_r + \Lambda_{/r})V$ be its SVD where $\Lambda_r = \mathrm{diag}(s_1, \dots, s_r)$ is a diagonal operator containing the first r singular values and $\Lambda_{/r}$ is another diagonal operator whose first r entries are 0 and the rest contain the tail singular values from s_{r+1} on. Furthermore, we define $\Sigma_{\mathbf{x},\mathsf{init}} := \Sigma_\star(K_{\mathsf{init}}, \sigma_{\mathbf{u}}^2)$

Assumption 3 (Alignment)**.** There exists $r < \infty$ and $\rho > 0$ such that,

$$V^{\mathsf{H}} \Lambda_r^2 V \preceq \rho \Sigma_{\mathbf{x},\mathsf{init}} \text{ and } s_{r+1} < \frac{1}{16} \mathcal{C}_{\mathsf{stable}}.$$

Since $\Sigma_{\mathbf{x},\mathsf{init}} \succeq \Sigma_{\mathbf{w}}$, Assumption 3 holds for some $r < \infty$ as long as $\Sigma_{\mathbf{w}}$ is positive definite in the sense that all its eigenvalues are strictly larger than, though decaying to, 0. Using this alignment condition, we show that the $\mathsf{WarmStart}$ algorithm returns estimates (A_0, B_0) satisfying Condition 3.1. Given an initial state $\mathbf{x}_1 = 0$, $\mathsf{WarmStart}$ chooses actions according to $\mathbf{u}_t = K_{\mathsf{init}} \mathbf{x}_t + \mathbf{v}_t$ for T_{init} many iterations, where $T_{\mathsf{init}} = \mathcal{O}(1)$ is a constant independent of the horizon T. Having collected a constant number of samples, the algorithm returns ridge regression estimates. Ssee Appendix F for formal description of $\mathsf{WarmStart}$.

Proposition 3.1. (informal) If Assumptions 2 and 3 hold, there exists a constant T_{init}, independent of the time horizon T, such that after collecting T_{init} samples $(\mathbf{x}_t, \mathbf{v}_t)$ under the exploration policy, $\mathbf{u}_t = K_{\mathsf{init}} \mathbf{x}_t + \mathbf{v}_t$, for $\mathbf{v}_t \sim \mathcal{N}(0, \sigma_{\mathbf{u}}^2 I)$, with probability $1 - \delta$, ridge regression returns estimates A_0, B_0 satisfying Condition 3.1.

Using this estimation result, we can prove the following corollary:

Corollary 3.1. *If Assumptions 2 and 3 hold, then with probability* $1 - \delta$*, the regret incurred by* $\mathsf{WarmStart}$ *satisfies*

$$\mathrm{Regret}_T(\mathsf{WarmStart}) \lesssim \log(1/\delta) \left(\sigma_{\mathbf{u}}^2 \mathrm{tr}\left[R\right] + \left\| Q + K_{\mathsf{init}}^{\mathsf{H}} R K_{\mathsf{init}} \right\|_{\mathrm{op}} \mathrm{tr}\left[\Sigma_\star(K_{\mathsf{init}}, \sigma_{\mathbf{u}}^2) \right] \right) T_{\mathsf{init}}.$$

In Appendix G.4, we describe how running $\mathsf{WarmStart}$ followed by $\mathsf{OnlineCE}$ satisfies an end-to-end regret guarantee whose asymptotics exactly match those of the $\mathsf{OnlineCE}$ algorithm.

4 Conclusion

In summary, this paper presents the first dimension-free regret guarantees for online LQR. We show that with a warm start, a simple approach based on certainty equivalence achieves sublinear regret for any $\Sigma_{\mathbf{w}}$ whose eigendecay is ever so slightly faster than trace-class and transition operator $A_\star$ that has finite Hilbert-Schmidt norm. While our bounds are nearly optimal when specialized to finite dimension, they provide a step towards a much sharper understanding of problem complexity in broader settings.

We believe that there are a number of promising directions for future work in this area. For example, it would be interesting to understand whether the alignment condition we introduce is necessary for certainty equivalence to succeed, or whether having a small prediction error bound (i.e $\|(\widehat{A} - A_\star)\Sigma_{\mathbf{x},0}^{1/2}\|_{\mathsf{HS}} \leq \varepsilon$) is in fact sufficient to guarantee that the certainty equivalent controller stabilizes the true system. Perhaps other algorithmic ideas, such as system level synthesis [Wang et al., 2019], could be used to circumvent operator-norm closeness. Furthermore, instead of depending on the ambient state dimension, our bounds instead depend on spectral decay of the noise covariance $\Sigma_{\mathbf{w}}$. It is an open question whether this measure of complexity can be made even sharper in some settings.

Lastly, given that our lower bounds rule out the possibility of infinite dimensional inputs, it is worth exploring what additional structural assumptions are necessary in order to incorporate richer control structures. More broadly, it would be exciting to understand whether dimension-free rates are possible in more general continuous control settings, or to extend the techniques from this paper to settings with partial observability like LQG.

Acknowledgements

We gratefully acknowledge the support of Microsoft through the BAIR Open Research Commons and of the NSF through grant DMS-2023505. JCP is supported by an NSF Graduate Research Fellowship. MS is generously supported by an Open Philanthropy Fellowship grant. We thank Yasin Abbasi-Yadkori for pointing us to his self-normalized inequality for vectors in Hilbert space, Lemma E.10.

References

Yasin Abbasi-Yadkori. *Online learning for linearly parametrized control problems.* PhD thesis, University of Alberta, 2012.

Yasin Abbasi-Yadkori and Csaba Szepesvári. Regret bounds for the adaptive control of linear quadratic systems. In *Proceedings of the 24th Annual Conference on Learning Theory*, pages 1–26, 2011.

Yasin Abbasi-Yadkori, Peter Bartlett, and Varun Kanade. Tracking adversarial targets. In *International Conference on Machine Learning*, pages 369–377, 2014.

Yasin Abbasi-Yadkori, Nevena Lazic, and Csaba Szepesvári. Model-free linear quadratic control via reduction to expert prediction. In *The 22nd International Conference on Artificial Intelligence and Statistics*, pages 3108–3117. PMLR, 2019.

Marc Abeille and Alessandro Lazaric. Thompson sampling for linear-quadratic control problems. In Aarti Singh and Jerry Zhu, editors, *Proceedings of the 20th International Conference on Artificial Intelligence and Statistics*, volume 54 of *Proceedings of Machine Learning Research*, pages 1246–1254. PMLR, 20–22 Apr 2017.

Marc Abeille and Alessandro Lazaric. Improved regret bounds for thompson sampling in linear quadratic control problems. *Proceedings of Machine Learning Research*, 80, 2018.

Marc Abeille and Alessandro Lazaric. Efficient optimistic exploration in linear-quadratic regulators via lagrangian relaxation. In *International Conference on Machine Learning*, pages 23–31. PMLR, 2020.

Alekh Agarwal, Mikael Henaff, Sham M. Kakade, and Wen Sun. Pc-pg: Policy cover directed exploration for provable policy gradient learning. In *Advances in Neural Information Processing Systems*, 2020.

Naman Agarwal, Brian Bullins, Elad Hazan, Sham Kakade, and Karan Singh. Online control with adversarial disturbances. In *Proceedings of the 36th International Conference on Machine Learning*, volume 97 of *Proceedings of Machine Learning Research*, pages 111–119. PMLR, 2019.

Peter L Bartlett and Shahar Mendelson. Rademacher and gaussian complexities: Risk bounds and structural results. *Journal of Machine Learning Research*, 3(Nov):463–482, 2002.

Alain Bensoussan, Giuseppe Da Prato, Michel C Delfour, and Sanjoy K Mitter. *Representation and control of infinite dimensional systems.* Springer Science & Business Media, 2007.

Dimitri P Bertsekas. *Dynamic programming and optimal control*, volume 1.

Asaf Cassel, Alon Cohen, and Tomer Koren. Logarithmic regret for learning linear quadratic regulators efficiently. In *International Conference on Machine Learning*, pages 1328–1337. PMLR, 2020.

Jie Chen and Guoxiang Gu. *Control-oriented system identification: an H-infinity approach*, volume 19. Wiley-Interscience, 2000.

Xiaohui Chen and Yun Yang. Hanson–wright inequality in hilbert spaces with application to k-means clustering for non-euclidean data. *Bernoulli*, 27(1):586–614, 02 2021.

Alon Cohen, Avinatan Hasidim, Tomer Koren, Nevena Lazic, Yishay Mansour, and Kunal Talwar. Online linear quadratic control. In *International Conference on Machine Learning*, pages 1029–1038. PMLR, 2018.

Alon Cohen, Tomer Koren, and Yishay Mansour. Learning linear-quadratic regulators efficiently with only sqrt(t) regret. In *International Conference on Machine Learning*, pages 1300–1309. PMLR, 2019.

Ruth F Curtain and Hans Zwart. *An introduction to infinite-dimensional linear systems theory*, volume 21. Springer Science & Business Media, 2012.

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

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

Silvestru Sever Dragomir. Some trace inequalities for operators in hilbert spaces. *arXiv preprint arXiv:1409.6295*, 2014.

Mohamad Kazem Shirani Faradonbeh, Ambuj Tewari, and George Michailidis. On optimality of adaptive linear-quadratic regulators. *arXiv preprint arXiv:1806.10749*, 2018.

Amir-massoud Farahmand, Andre Barreto, and Daniel Nikovski. Value-aware loss function for model-based reinforcement learning. In *Artificial Intelligence and Statistics*, pages 1486–1494, 2017.

Maryam Fazel, Rong Ge, Sham Kakade, and Mehran Mesbahi. Global convergence of policy gradient methods for the linear quadratic regulator. In *International Conference on Machine Learning*, pages 1467–1476. PMLR, 2018.

Alexander Goldenshluger. Nonparametric estimation of transfer functions: rates of convergence and adaptation. *IEEE Transactions on Information Theory*, 44(2):644–658, 1998.

Moritz Hardt, Tengyu Ma, and Benjamin Recht. Gradient descent learns linear dynamical systems. *The Journal of Machine Learning Research*, 19(1):1025–1068, 2018.

Elad Hazan, Sham Kakade, and Karan Singh. The nonstochastic control problem. In *Algorithmic Learning Theory*, pages 408–421. PMLR, 2020.

Arthur J Helmicki, Clas A Jacobson, and Carl N Nett. Control oriented system identification: a worst-case/deterministic approach in h/sub infinity. *IEEE Transactions on Automatic control*, 36(10):1163–1176, 1991.

G. Hewer. An iterative technique for the computation of the steady state gains for the discrete optimal regulator. *IEEE Transactions on Automatic Control*, 16(4):382–384, 1971. doi: 10.1109/TAC.1971.1099755.

Petros A Ioannou and Jing Sun. *Robust adaptive control*. Courier Corporation, 2012.

Nan Jiang, Akshay Krishnamurthy, Alekh Agarwal, John Langford, and Robert E Schapire. Contextual decision processes with low bellman rank are pac-learnable. In *International Conference on Machine Learning*, pages 1704–1713. PMLR, 2017.

Chi Jin, Zhuoran Yang, Zhaoran Wang, and Michael I Jordan. Provably efficient reinforcement learning with linear function approximation. In *Conference on Learning Theory*, pages 2137–2143. PMLR, 2020.

Sham Kakade, Akshay Krishnamurthy, Kendall Lowrey, Motoya Ohnishi, and Wen Sun. Information theoretic regret bounds for online nonlinear control. In *Advances in Neural Information Processing Systems*, volume 33, pages 15312–15325, 2020.

Karl Krauth, Stephen Tu, and Benjamin Recht. Finite-time analysis of approximate policy iteration for the linear quadratic regulator. In *Advances in Neural Information Processing Systems*, pages 8514–8524, 2019.

Miroslav Krstic, Petar V Kokotovic, and Ioannis Kanellakopoulos. *Nonlinear and adaptive control design*. John Wiley & Sons, Inc., 1995.

Kwang Yun Lee, Shui-nee Chow, and Robert O Barr. On the control of discrete-time distributed parameter systems. *SIAM Journal on Control*, 10(2):361–376, 1972.

Zhang Liu and Lieven Vandenberghe. Interior-point method for nuclear norm approximation with application to system identification. *SIAM Journal on Matrix Analysis and Applications*, 31(3):1235–1256, 2010.

Lennart Ljung. System identification. *Wiley encyclopedia of electrical and electronics engineering*, pages 1–19, 1999.

Horia Mania, Stephen Tu, and Benjamin Recht. Certainty equivalence is efficient for linear quadratic control. In *Advances in Neural Information Processing Systems*, pages 10154–10164, 2019.

Horia Mania, Michael I Jordan, and Benjamin Recht. Active learning for nonlinear system identification with guarantees. *arXiv preprint arXiv:2006.10277*, 2020.

Yi Ouyang, Mukul Gagrani, and Rahul Jain. Control of unknown linear systems with thompson sampling. In *2017 55th Annual Allerton Conference on Communication, Control, and Computing (Allerton)*, pages 1198–1205. IEEE, 2017.

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

José Pereira, Morteza Ibrahimi, and Andrea Montanari. Learning networks of stochastic differential equations. In *Advances in Neural Information Processing Systems*, pages 172–180, 2010.

Alexander Rakhlin and Karthik Sridharan. Online non-parametric regression. In *Conference on Learning Theory*, pages 1232–1264. PMLR, 2014.

Daniel Russo and Benjamin Van Roy. Eluder dimension and the sample complexity of optimistic exploration. In *Advances in Neural Information Processing Systems*, volume 26, 2013.

Pierre Sagaut. *Large eddy simulation for incompressible flows: an introduction.* Springer Science & Business Media, 2006.

Tuhin Sarkar and Alexander Rakhlin. Near optimal finite time identification of arbitrary linear dynamical systems. In *International Conference on Machine Learning*, pages 5610–5618. PMLR, 2019.

Max Simchowitz and Dylan Foster. Naive exploration is optimal for online lqr. In *International Conference on Machine Learning*, pages 8937–8948. PMLR, 2020.

Max Simchowitz, Horia Mania, Stephen Tu, Michael I Jordan, and Benjamin Recht. Learning without mixing: Towards a sharp analysis of linear system identification. In *Conference On Learning Theory*, pages 439–473. PMLR, 2018.

Max Simchowitz, Ross Boczar, and Benjamin Recht. Learning linear dynamical systems with semi-parametric least squares. In *Conference on Learning Theory*, pages 2714–2802. PMLR, 2019.

Max Simchowitz, Karan Singh, and Elad Hazan. Improper learning for non-stochastic control. In *Conference on Learning Theory*, pages 3320–3436. PMLR, 2020.

Herbert A Simon. Dynamic programming under uncertainty with a quadratic criterion function. *Econometrica, Journal of the Econometric Society*, pages 74–81, 1956.

Niranjan Srinivas, Andreas Krause, Sham Kakade, and Matthias Seeger. Gaussian process optimization in the bandit setting: No regret and experimental design. In *Proceedings of the 27th International Conference on International Conference on Machine Learning*, ICML'10, page 1015–1022, 2010.

Wen Sun, Nan Jiang, Akshay Krishnamurthy, Alekh Agarwal, and John Langford. Model-based rl in contextual decision processes: Pac bounds and exponential improvements over model-free approaches. In *Conference on Learning Theory*, pages 2898–2933. PMLR, 2019.

Henri Theil. A note on certainty equivalence in dynamic planning. *Econometrica: Journal of the Econometric Society*, pages 346–349, 1957.

Anastasios Tsiamis and George J Pappas. Finite sample analysis of stochastic system identification. In *2019 IEEE 58th Conference on Decision and Control (CDC)*, pages 3648–3654. IEEE, 2019.

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

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

Mathukumalli Vidyasagar and Rajeeva L Karandikar. A learning theory approach to system identification and stochastic adaptive control. In *Probabilistic and randomized methods for design under uncertainty*, pages 265–302. Springer, 2006.

Yuh-Shyang Wang, Nikolai Matni, and John C Doyle. A system-level approach to controller synthesis. *IEEE Transactions on Automatic Control*, 64(10):4079–4093, 2019.

Emmet F Whittlesey. Analytic functions in banach spaces. *Proceedings of the American Mathematical Society*, 16(5):1077–1083, 1965.

Lin Yang and Mengdi Wang. Reinforcement learning in feature space: Matrix bandit, kernels, and regret bound. In *International Conference on Machine Learning*, pages 10746–10756. PMLR, 2020.

Zhuoran Yang, Chi Jin, Zhaoran Wang, Mengdi Wang, and Michael Jordan. Provably efficient reinforcement learning with kernel and neural function approximations. *Advances in Neural Information Processing Systems*, 33, 2020a.

Zhuoran Yang, Chi Jin, Zhaoran Wang, Mengdi Wang, and Michael I Jordan. Bridging exploration and general function approximation in reinforcement learning: Provably efficient kernel and neural value iterations. *arXiv preprint arXiv:2011.04622*, 2020b.

Jerzy Zabczyk. Remarks on the control of discrete-time distributed parameter systems. *SIAM Journal on Control*, 12(4):721–735, 1974.

Jerzy Zabczyk. On optimal stochastic control of discrete-time systems in hilbert space. *SIAM Journal on Control*, 13(6):1217–1234, 1975.

Tong Zhang. Learning bounds for kernel regression using effective data dimensionality. *Neural Computation*, 17(9):2077–2098, 2005.

Table of Contents: Appendix

A Extended Preliminaries, Organization, and Related Work

A.1 Organization

The appendix is organized as follows:

- Appendix A includes extended related work, a review of notation, and further preliminaries on linear quadratic control.
- Part I contains all technical control theoretic contributions. Appendix B describes the various change of covariance theorems used to swap between controllers. Appendix C establishes the perturbation bounds for certainty equivalence. Appendix D states and proves various technical lemmas used throughout.
- Part II addresses estimation of system parameters from a single trajectory.
- In Part III, Appendix F contains a formal statement of the algorithms, Appendix G proves an upper bound on the regret of OnlineCE, and Appendix H describes the lower bound requiring finite dimensions.

A.2 Extended Related Work

The problem of learning the parameters of a linear system is historically referred to as *system identification*, and has been studied at length for systems of finite dimension. Classical asymptotic results are detailed in Ljung [1999]; early non-asymptotic results [Vidyasagar and Karandikar, 2006, Hardt et al., 2018, Pereira et al., 2010] suffered from opaque and possible exponential dependencies on system parameters. Dean et al. [2019] presented the first finite-sample guarantees for control synthesis from statistical data, and estimation techniques were later refined in subsequent works [Simchowitz et al., 2018, Sarkar and Rakhlin, 2019], and extended to systems with partial observation [Oymak and Ozay, 2019, Simchowitz et al., 2019, Tsiamis and Pappas, 2019]. Parallel work has studied estimation in the frequency domain [Tu et al., 2017, Helmicki et al., 1991, Goldenshluger, 1998, Chen and Gu, 2000].

Building on system-identification, *adaptive control* considers the problem of refining estimates of system parameters during a control task, so as to converge to a near-optimal policy. Classical results are detailed in [Krstic et al., 1995, Ioannou and Sun, 2012]. The online LQR setting considered in this work is a special case. The study of online LQR was initiated by Abbasi-Yadkori and Szepesvári [2011], who gave a computationally intractable algorithm based on optimism in the face of uncertainty (OFU). Their algorithm obtained $\sqrt{T}$ regret, albeit with a potentially exponential dependence on dimension. Dean et al. [2018] gave an efficient algorithm based on System Level Synthesis which obtained a $T^{2/3}$ regret bound with polynomial dependence on the relevant problem parameters. Mania et al. [2019], Faradonbeh et al. [2018], and Cohen et al. [2019] simultaneously presented efficient algorithms enjoying $\sqrt{T}$ regret (as well as polynomial dependence in other problem parameters). Cassel et al. [2020] and Simchowitz and Foster [2020] demonstrated that the $\sqrt{T}$ rate was indeed optimal. The latter provided matching upper and lower bounds of $\tilde{\Theta}(\sqrt{d_x d_u^2 T})$ in terms of time horizon and problem dimensions. Other approaches have studied Thompson sampling [Abeille and Lazaric, 2017, Ouyang et al., 2017, Abeille and Lazaric, 2018], though regret guarantees which depend transparently on both time horizon and dimension remain elusive. Finally, Abeille and Lazaric [2020] provided an efficient implementation of the OFU algorithm introduced by Abbasi-Yadkori and Szepesvári [2011], which attained $\sqrt{T}$ regret, and sacrifices suboptimal dependence in problem dimension for improved dependence in other problem parameters. Similar regret guarantees were subsequently attained by Kakade et al. [2020] in a non-linear control setting similar to the one studied in Mania et al. [2020]. There has also been work on a number of related online control settings [Abbasi-Yadkori et al., 2014, Cohen et al., 2018], notably the nonstochastic control setting proposed by Agarwal et al. [2019] and expanded upon in Hazan et al. [2020] and Simchowitz et al. [2020].

The majority of the above approaches to online and adaptive control are *model-based*: that is, they learn a representation of the model of the dynamics, and update their control policy accordingly. This present work builds on past study of *certainty equivalence* [Mania et al., 2019, Simchowitz et al., 2020]; other approaches include robust control synthesis via SLS [Dean et al., 2018, 2019], OFU [Abbasi-Yadkori and Szepesvári, 2011, Abeille and Lazaric, 2020], SDP relaxations [Cohen et al., 2018], and other convex methods [Agarwal et al., 2019, Simchowitz et al., 2020]. Concurrent work has also studied *model-free* approaches in the batch [Fazel et al., 2018, Krauth et al., 2019] and online [Abbasi-Yadkori et al., 2019] settings, though recent work suggest these approaches suffer from high variance, worse dimension dependence, and overall inferior sample complexity [Tu and Recht, 2019].

Outside the controls literature, sample complexity and regret guarantees that do not explicitly depend on the ambient dimension, but on more intrinsic measures for learning in RKHSs are well-known in the supervised learning setting (see e.g. [Bartlett and Mendelson, 2002, Zhang, 2005]) as well as in bandit problems [Srinivas et al., 2010]. More recently, these results have been extended to the reinforcement learning literature as well, for a class of problems defined as linear MDPs [Jin et al., 2020, Agarwal et al., 2020, Yang et al., 2020a]. While linear MDPs also make linearity assumptions on the system dynamics, the precise assumption is quite different from those present in LQR. In a linear MDP, the conditional distribution of the next state, given the current state and action is assumed to be linear under a known featurization of the state, action pair. In contrast, LQRs only require the conditional mean to be linear and do not guarantee certain nice properties of a linear MDP, such as the conditional expectation of any function of the next state being linear in the features of the current state, action pair. Consequently, the results from linear MDPs are incomparable to our work.

Somewhat related to our development here, the observation of using an accuracy measure for model estimation that is informed by the value function parameterization has been recently leveraged in the reinforcement learning literature [Farahmand et al., 2017, Sun et al., 2019], though the analysis techniques are quite different and the algorithms are not applicable to the continuous control setting.

A.3 Notation Review

Setting. Recall the state $\mathbf{x}_t \in \mathcal{H}_{\mathbf{x}}$, input $\mathbf{u}_t \in \mathbb{R}^{d_x}$, noise $\mathbf{w}_t \in \mathcal{H}_{\mathbf{x}}$, true dynamics operators $(A_\star, B_\star)$, noise covariance $\Sigma_{\mathbf{w}}$, and cost function $\mathcal{J}(K)$, defined for any state-feedback controller $K : \mathcal{H}_{\mathbf{x}} \to \mathbb{R}^{d_u}$ which is stabilizing for $(A_\star, B_\star)$. The optimal cost of the LQR problem is $\mathcal{J}_\star$, which is achieved the controller $K_\star$.

Given operators (A, B), $P_\infty(A, B)$ and $K_\infty(A, B)$ denote the value function and optimal controller (solving Eq. (2.1) and Eq. (2.2)). The dlyap operator, and its related quantities, are described in the extended preliminaries below (Appendix A.4). Given a controller K which stabilizes (A, B), we set

$$P_\infty(K; A, B) := \mathsf{dlyap}\left(A + BK,\ Q + K^{\mathsf{H}} RK\right), \tag{A.1}$$

where Q and R the costs operators. As detailed in Assumption 1, we assumed throughout the entirety of our analysis that $\sigma_{\min}(Q) > 1$ and $\sigma_{\min}(R) > 1$.

Exploration. We use K_0 to denote exploratory controllers, and $\Sigma_{\mathbf{x},0}$ to denote the induced exploratory covariance with inputs $\mathbf{u}_t = K_0 \mathbf{x}_t + \mathbf{v}_t$, where $\mathbf{v}_t \sim \mathcal{N}(0, \sigma_{\mathbf{u}}^2 I)$ captures additional Gaussian excitation. We let $\Sigma_0 := \sigma_{\mathbf{u}}^2 B_\star B_\star^{\mathsf{H}} + \Sigma_{\mathbf{w}} = \mathbb{E}[(B_\star \mathbf{v}_t + \mathbf{w}_t) \otimes (B_\star \mathbf{v}_t + \mathbf{w}_t)]$, so that $\Sigma_{\mathbf{x},0} = \mathsf{dlyap}\left((A_\star + B_\star K_0)^{\mathsf{H}},\ \Sigma_0\right)$. We recall the dimension-free parameter from Eq. (1.4),

$$M_\star := \max\{\|A_\star\|_{\mathrm{op}}^2, \|B_\star\|_{\mathrm{op}}^2, \|P_\star\|_{\mathrm{op}}^2, \|\Sigma_{\mathbf{w}}\|_{\mathrm{op}}, 1\}. \tag{A.2}$$

Certainty Equivalence We let $\widehat{A}, \widehat{B}$ denote estimates of the true system operators $A_\star, B_\star$. Likewise, we use $\widehat{P} := P_\infty(\widehat{A}, \widehat{B})$ and $\widehat{K} := K_\infty(\widehat{A}, \widehat{B})$ to describe the value function (solution to the the DARE over $\widehat{A}, \widehat{B}$) and optimal controller for the estimated system. Similarly, we define $P_\star := P_\infty(A_\star, B_\star)$.

Linear Algebra. We let $\mathcal{H}_{\mathbf{x}}$ denote a Hilbert space containing the states. Inputs $\mathbf{u}_t$ lie in $\mathbb{R}^{d_u}$ where $d_u < \infty$. We use upper case X for linear operators, and lower case bound $\mathbf{x}$ for vectors. X^{H} and $\mathbf{x}^{\mathsf{H}}$ denote adjoints. We let $\langle \cdot, \cdot \rangle$ and $\| \cdot \|$ denote norms and inner products in the relevant Hilbert space. $\| \cdot \|_{\mathrm{op}}$, $\| \cdot \|_{\mathsf{HS}}$, and $\|\cdot\|_{\mathsf{tr}}$ denote operator, Hilbert-Schmidt (abbreviated as HS), and trace norms, respectively. Occasionally, we use $[XY]_{\mathrm{algn}} := \max_{W:\|W\|_{\mathrm{op}}} \mathrm{tr}\,[XWY]$.

Lastly, we adopt the shorthand $\log_+(x) := \max\{\log(x), 1\}$ and use $a \lesssim b$ to denote that $a \ \le\ C \cdot b$ where C is a universal constant.

A.4 Extended Preliminaries

The discrete Lyapunov operator and higher order operators Recall our earlier definition of the Lyapunov operator,

Definition A.1 (Lyapunov Operator)**.** Let $A : \mathcal{H}_{\mathbf{x}} \to \mathcal{H}_{\mathbf{x}}$ be a bounded stable linear operator and let $\Lambda : \mathcal{H}_{\mathbf{x}} \to \mathcal{H}_{\mathbf{x}}$ be self-adjoint, $\mathsf{dlyap}\,(A,\ \Lambda)$ is a symmetric bounded operator that solves the matrix equation,

$$X = A^{\mathsf{H}} X A + \Lambda. \tag{A.3}$$

Furthermore, it has a closed-form expression given by:

$$\mathsf{dlyap}\,(A,\ \Lambda) = \sum_{j=0}^{\infty} (A^{\mathsf{H}})^j \Lambda A^j. \tag{A.4}$$

Throughout our analysis, we will make repeated use of the higher order Lyapunov operator.

Definition A.2 (Higher Order Lyapunov Operator)**.**

$$\mathsf{dlyap}_{(m)}\,(A,\ \Sigma) = \sum_{j=0}^{\infty} (A^{\mathsf{H}})^j \Lambda A^j (j+1)^m \tag{A.5}$$

As is shown in Lemma D.10, we have $\mathsf{dlyap}_{(1)}\left(A^{\mathsf{H}},\ \Sigma\right) = \mathsf{dlyap}\,(A,\ \mathsf{dlyap}\,(A,\ \Sigma))$.

The Lyapunov operator satisfies a number of important properties which feature prominently in our technical discussion. We state the following lemma describing some of the most important properties and point the reader to Appendix D.2 for further results on Lyapunov theory.

Lemma A.1 (Lemma B.5 in Simchowitz and Foster [2020])**.** *The following relationships hold for* dlyap*:*

1. *If A_{cl} is stable and $Y \preceq Z$, then* $\mathsf{dlyap}\,(A_{\mathrm{cl}},\ Y) \preceq \mathsf{dlyap}\,(A_{\mathrm{cl}},\ Z)$.
2. *If $Q \succeq I$ and $A + BK$ is stable, then*
$$\pm\mathsf{dlyap}\,(A + BK,\ Y) \preceq \mathsf{dlyap}\,(A + BK,\ I) \cdot \|Y\|_{\mathrm{op}} \preceq P_\infty(K; A, B) \cdot \|Y\|_{\mathrm{op}}$$
3. *If $Q \succ I$, then $P_\infty(A, B) \succ I$.*
4. *If A_{cl} is stable, then* $\|\mathsf{dlyap}\,(A_{\mathrm{cl}},\ I)\|_{\mathrm{op}} = \left\|\mathsf{dlyap}\left(A_{\mathrm{cl}}^{\mathsf{H}},\ I\right)\right\|_{\mathrm{op}}$.

Stationary state covariances For controllers K such that $A_\star + B_\star K$ is stable, we define the covariance operator:

$$\Sigma_\star(K, \sigma_{\mathbf{u}}^2) := \lim_{t\to\infty} \mathbb{E}[\mathbf{x}_t \otimes \mathbf{x}_t],\ \text{s.t. } \mathbf{x}_{t+1} = A_\star \mathbf{x}_t + B_\star \mathbf{u}_t + \mathbf{w}_t,$$
$$\text{where } \mathbf{u}_t = K\mathbf{x}_t + \mathbf{v}_t, \quad \mathbf{w}_t \overset{\text{i.i.d}}{\sim} \mathcal{N}(0, \Sigma_{\mathbf{w}}), \mathbf{v}_t \overset{\text{i.i.d}}{\sim} \mathcal{N}(0, \sigma_{\mathbf{u}}^2 I).$$

We specialize $\Sigma_\star(K) = \Sigma_\star(K, 0)$. A short calculation shows that:

$$\Sigma_\star(K, \sigma_{\mathbf{u}}^2) = \mathsf{dlyap}\left((A_\star + B_\star K)^{\mathsf{H}},\ \Sigma_{\mathbf{w}} + \sigma_{\mathbf{u}}^2 B_\star B_\star^{\mathsf{H}}\right). \tag{A.6}$$

Interpretation of P-matrices as value functions The P-matrices $P_\infty(A,B)$ and $P_\infty(K;A,B)$ can be interpreted in terms of value functions. Specifically, consider an LQR problem with cost operators Q,R and initial state $\mathbf{x}_0$, but *no noise*:

$$\mathbf{x}_{t+1} = A_\star \mathbf{x}_t + B_\star \mathbf{u}_t, \quad t \geq 0. \tag{A.7}$$

We can then verify that the operator $P_\infty(K;A,B)$ defined in Eq. (A.1) satisfies

$$\langle \mathbf{x}, P_\infty(K;A,B)\mathbf{x}\rangle = \sum_{t\geq 0} \langle \mathbf{x}, Q\mathbf{x}\rangle + \langle K\mathbf{u}, RK\mathbf{u}_t\rangle \text{ subject to Eq. (A.7)}, \quad \text{with } \mathbf{x}_0 = \mathbf{x}.$$

The controller $K_\infty(A,B)$ can be show to be the minimizer of the above cost over all policies, regardless of starting state (see e.g. Bertsekas). The following calculation provides another perspective of P capturing the costs of the LQR problem. From Eq. (1.2),

$$\begin{aligned}
\mathcal{J}(K) &= \lim_{T\to\infty} \frac{1}{T}\mathbb{E}\left[\textstyle\sum_{t=1}^{T} \langle \mathbf{x}_t, Q\mathbf{x}_t\rangle + \langle \mathbf{u}_t, R\mathbf{u}_t\rangle\right], \quad \text{subject to } \mathbf{u}_t = K\mathbf{x}_t \\
&= \lim_{T\to\infty} \operatorname{tr}\left[(Q + K^{\mathsf{H}}RK)\mathbb{E}[\mathbf{x}_t \otimes \mathbf{x}_t]\right] \\
&= \operatorname{tr}\left[(Q + K^{\mathsf{H}}RK)\Sigma_\star(K)\right] \\
&= \operatorname{tr}\left[(Q + K^{\mathsf{H}}RK)\mathsf{dlyap}\left((A_\star + B_\star K)^{\mathsf{H}},\ \Sigma_{\mathbf{w}}\right)\right] \\
&= \operatorname{tr}\left[\mathsf{dlyap}\left(A_\star + B_\star K,\ Q + K^{\mathsf{H}}RK\right)\Sigma_{\mathbf{w}}\right] \\
&= \operatorname{tr}\left[P_\infty(K;A_\star,B_\star)\Sigma_{\mathbf{w}}\right].
\end{aligned}$$

The cost of a controller K is therefore captured by the trace inner product of $P_\infty(K;A_\star,B_\star)$ and $\Sigma_{\mathbf{w}}$.

Part I
Perturbation Bounds and Technical Lemmas

B The Change of Covariance Theorems

In this section, we state and prove the various change-of-covariance theorems required in the paper, including Lemma 2.1, and its generalization Theorem B.1. We begin by stating the more general result and then illustrate how Lemma 2.1 follows from this statement. We conclude by proving another comparison inequality between the covariance operators induced by different stabilizing controllers.

To state the general theorem, recall the *higher order* Lyapunov operator defined in Definition A.2:

$$\mathsf{dlyap}_{(m)}(A,\ \Sigma) = \sum_{j=0}^{\infty} (A^{\mathsf{H}})^j \Lambda A^j (j+1)^m. \tag{B.1}$$

The result is as follows:

Theorem B.1. *Let $K_1 : \mathcal{H}_{\mathbf{x}} \to \mathbb{R}^{d_u}$ be such that $A + BK_1$ is stable and let $\Lambda \in \mathbb{S}_+^{\mathcal{H}_{\mathbf{x}}}$ be a trace class, positive semi-definite operator. Then, for any K_2 such that $A + BK_2$ is also stable, we have that*

$$\mathsf{dlyap}\left((A + BK_1)^{\mathsf{H}},\ \Lambda\right) \preceq \mathsf{dlyap}\left((A + BK_2)^{\mathsf{H}},\ \overline{\Lambda}\right)$$

where $\overline{\Lambda}$ is a bounded operator defined as

$$\overline{\Lambda} := 2\Lambda + 4B(K_1 - K_2)\ \mathsf{dlyap}_{(2)}\left((A + BK_1)^{\mathsf{H}},\ \Lambda\right)\ (K_1 - K_2)^{\mathsf{H}} B^{\mathsf{H}}$$

Proof of Theorem B.1. We argue by constructing a hypothetical dynamical system and analyzing its behavior in two ways. In particular, define

$$A_{\mathrm{cl},1} := A + BK_1 \text{ and } A_{\mathrm{cl},2} := A + BK_2$$

and consider the system,

$$\mathbf{x}_{t+1} = A_{\mathrm{cl},1}\mathbf{x}_t + \mathbf{w}_t, \text{ where } \mathbf{x}_0 = 0 \text{ and } \forall\, t \geq 0,\ \mathbf{w}_t \overset{\text{i.i.d}}{\sim} \mathcal{N}(0, \Lambda). \tag{B.2}$$

Then, for $\Sigma_1 := \mathsf{dlyap}\left(A_{\mathrm{cl},1}^{\mathsf{H}},\ \Lambda\right)$ we have that

$$\Sigma_1 = \lim_{T\to\infty} \Sigma_{1;T}, \text{where } \Sigma_{1;T} := \mathbb{E}[\mathbf{x}_T \otimes \mathbf{x}_T],$$

and where the above limit exists due to monotonicity of $\Sigma_{1;T}$. To prove our claim, let us express the evolution of Eq. (B.2) as

$$\mathbf{x}_{t+1} = A_{\mathrm{cl},2}\mathbf{x}_t + (A_{\mathrm{cl},1} - A_{\mathrm{cl},2})\mathbf{x}_t + \mathbf{w}_t = A_{\mathrm{cl},2}\mathbf{x}_t + B\underbrace{(K_1 - K_2)\mathbf{x}_t}_{:=\mathbf{u}_t} + \mathbf{w}_t.$$

Define the deterministic operator $G_T := \left[I \mid A_{\mathrm{cl},2} \mid \cdots \mid A_{\mathrm{cl},2}^{T-1}\right] : \mathcal{H}_{\mathbf{x}}^T \to \mathcal{H}_{\mathbf{x}}$, and the random vector

$$\mathbf{z}_{[T]} := \begin{bmatrix} B\mathbf{u}_{T-1} + \mathbf{w}_{T-1} \\ \cdots \\ B\mathbf{u}_1 + \mathbf{w}_1 \\ B\mathbf{u}_0 + \mathbf{w}_0 \end{bmatrix} \in \mathcal{H}_{\mathbf{x}}^T$$

Now, we can rewrite $\mathbf{x}_T$ as $\mathbf{x}_T = G_T\, \mathbf{z}_{[T]}$, so that

$$\Sigma_{1;T} := \mathbb{E}[\mathbf{x}_T \otimes \mathbf{x}_T] = G_T \mathbb{E}[\mathbf{z}_{[T]} \otimes \mathbf{z}_{[T]}] G_T^{\mathsf{H}} \tag{B.3}$$

Given a general linear operator $X : \mathcal{U} \to \mathcal{V}$, define $\mathrm{diag}_T(X) : \mathcal{U}^T \to \mathcal{V}^T$ as the block diagonal operator with T copies of X on its diagonal. Observe that if it holds that, for some psd $\overline{\Lambda} : \mathcal{H}_{\mathbf{x}} \to \mathcal{H}_{\mathbf{x}}$, $\mathbb{E}[\mathbf{z}_{[T]} \otimes \mathbf{z}_{[T]}] \preceq \mathrm{diag}_T(\overline{\Lambda})$, then by Eq. (B.3),

$$\begin{aligned}
\Sigma_1 &= \lim_{T\to\infty} G_T \mathbb{E}[\mathbf{z}_{[T]} \otimes \mathbf{z}_{[T]}] G_T^{\mathsf{H}} \\
&\preceq \lim_{T\to\infty} G_T \mathrm{diag}_T(\overline{\Lambda}) G_T \\
&= \lim_{T\to\infty} \sum_{t=0}^{T-1} A_{\mathrm{cl},2}^t \overline{\Lambda} \left(A_{\mathrm{cl},2}^{\mathsf{H}}\right)^t = \mathsf{dlyap}\left((A + BK_2)^{\mathsf{H}},\ \overline{\Lambda}\right).
\end{aligned}$$

Hence, it remains to bound $\mathbb{E}[\mathbf{z}_{[T]} \otimes \mathbf{z}_{[T]}]$. Let us introduce the shorthand $\mathbf{u}_{[T]} := (\mathbf{u}_{T-1}, \mathbf{u}_{T-2}, \ldots, \mathbf{u}_0)$, and similarly for $\mathbf{w}_{[T]}$ and $\mathbf{x}_{[T]}$. Then, using the definition of $\mathbf{z}_{[T]}$ and the fact that $\mathbf{u}_t = (K_1 - K_2)\mathbf{x}_t$,

$$\begin{aligned}
\mathbb{E}[\mathbf{z}_{[T]} \otimes \mathbf{z}_{[T]}] &= \mathbb{E}[(\mathrm{diag}_T(B)\mathbf{u}_{[T]} + \mathbf{w}_{[T]})^{\otimes 2}] \\
&= \mathbb{E}[(\mathrm{diag}_T(B(K_1 - K_2))\mathbf{x}_{[T]} + \mathbf{w}_{[T]})^{\otimes 2}] \\
&\preceq 2\mathrm{diag}_T(B(K_1 - K_2)) \cdot \mathbb{E}[\mathbf{x}_{[T]} \otimes \mathbf{x}_{[T]}] \cdot \mathrm{diag}_T(B(K_1 - K_2))^{\mathsf{H}} + 2\mathbb{E}[\mathbf{w}_{[T]} \otimes \mathbf{w}_{[T]}] \\
&= 2\mathrm{diag}_T(B(K_1 - K_2)) \cdot \mathbb{E}[\mathbf{x}_{[T]} \otimes \mathbf{x}_{[T]}] \cdot \mathrm{diag}_T(B(K_1 - K_2))^{\mathsf{H}} + 2\mathrm{diag}_T(\Lambda).
\end{aligned} \tag{B.4}$$

The inequality on the third line follows from the fact that,

$$(a + b) \otimes (a + b) = 2\left(a \otimes a + b \otimes b\right) - (a - b) \otimes (a - b).$$

In the last line, we have used $\mathbb{E}[\mathbf{w}_{[T]} \otimes \mathbf{w}_{[T]}] = \mathrm{diag}_T(\Lambda)$ since $\mathbf{w}_{[T]} = (\mathbf{w}_{T-1}, \dots, \mathbf{w}_0)$, and $\mathbf{w}_t \overset{\text{i.i.d}}{\sim} \mathcal{N}(0, \Lambda)$ for $t \geq 0$. To bound $\mathbb{E}[\mathbf{x}_{[T]} \otimes \mathbf{x}_{[T]}]$, we introduce the block Toeplitz operator:

$$\mathsf{Toep}_T : \mathcal{H}_{\mathbf{x}}^T \to \mathcal{H}_{\mathbf{x}}^T, \text{ with blocks } \mathsf{Toep}[i,j] = A_{\mathrm{cl},1}^{j-i-1}\mathbb{I}_{j \geq i+1}.$$

Then, we have the identity $\mathbf{x}_{[T]} = \mathsf{Toep}_T \mathbf{w}_{[T]}$, which implies that

$$\mathbb{E}\mathbf{x}_{[T]} \otimes \mathbf{x}_{[T]} = \mathsf{Toep}_T \mathbb{E}\left[\mathbf{w}_{[T]} \otimes \mathbf{w}_{[T]}\right] \mathsf{Toep}_T^{\mathsf{H}} = \mathsf{Toep}_T \cdot \mathrm{diag}_T(\Lambda) \cdot \mathsf{Toep}_T^{\mathsf{H}}, \tag{B.5}$$

where $\mathbf{w}_t \overset{\text{i.i.d}}{\sim} \mathcal{N}(0, \Lambda)$ for $t \geq 0$. To conclude, let us decompose Toep_T into single-band operators via $Y_n : \mathcal{H}_{\mathbf{x}}^T \to \mathcal{H}_{\mathbf{x}}^T$ for $n \in \{1, \dots, T-1\}$, via:

$$\mathsf{Toep}_T = \sum_{n=1}^{T-1} Y_n, \text{ where } Y_n : \mathcal{H}_{\mathbf{x}}^T \to \mathcal{H}_{\mathbf{x}}^T \text{ has blocks } Y_n[i,j] = \mathbb{I}_{j=i+n} \cdot A_{\mathrm{cl},1}^{n-1}. \tag{B.6}$$

Hence, continuing from Eq. (B.5), and applying the substitution in Eq. (B.6), by Lemma D.1 we have that

$$\begin{aligned}
\mathbb{E}[\mathbf{x}_{[T]} \otimes \mathbf{x}_{[T]}] &\preceq \mathsf{Toep}_T \cdot \mathrm{diag}_T(\Lambda) \cdot \mathsf{Toep}_T^{\mathsf{H}} && \text{(Eq. (B.5))} \\
&= \left(\sum_{n=1}^{T-1} Y_n\right) \cdot \mathrm{diag}_T(\Lambda) \cdot \left(\sum_{n=1}^{T-1} Y_n\right)^{\mathsf{H}} && \text{(Eq. (B.6))} \\
&\preceq 2\sum_{n=1}^{T-1} n^2 \cdot Y_n \mathrm{diag}_T(\Lambda) Y_n^{\mathsf{H}} && \text{(Lemma D.1)}
\end{aligned}$$

A simple computation reveals that $X_n := Y_n \mathrm{diag}_T(\Lambda) Y_n^{\mathsf{H}}$ is block diagonal with i-th block given by,

$$X_n[i,i] = \mathbb{I}_{n \leq i} A_{\mathrm{cl},1}^{n-1} \Lambda \left(A_{\mathrm{cl},1}^{\mathsf{H}}\right)^{n-1}.$$

Thus, $\left(\sum_{n=1}^T n^2 \cdot Y_n \mathrm{diag}_T(\Lambda) Y_n^{\mathsf{H}}\right)$ is block diagonal with i-th block given by

$$\begin{aligned}
\sum_{n=1}^{i} n^2 \cdot A_{\mathrm{cl},1}^{n-1} \Lambda A_{\mathrm{cl},1}^{(n-1)\mathsf{H}} &\preceq \sum_{n \geq 0} (n+1)^2 \cdot A_{\mathrm{cl},1}^{n} \Lambda A_{\mathrm{cl},1}^{n\mathsf{H}} \\
&= \sum_{n \geq 0} (n+1)^2 \cdot (A + BK_1)^n \Lambda \left((A+BK_1)^n\right)^{\mathsf{H}} = \mathsf{dlyap}_{(2)}\left(A_{\mathrm{cl},1},\ \Lambda\right)
\end{aligned}$$

Hence, $\mathbb{E}\mathbf{x}_{[T]} \otimes \mathbf{x}_{[T]} \preceq 2\mathrm{diag}_T\left(\mathsf{dlyap}_{(2)}\left(A_{\mathrm{cl},1},\ \Lambda\right)\right)$. Combining with Eq. (B.4) gives

$$\begin{aligned}
\mathbb{E}[\mathbf{z}_{[T]} \otimes \mathbf{z}_{[T]}] &\preceq 2\mathrm{diag}_T(B(K_1 - K_2)) \cdot \mathbb{E}[\mathbf{x}_{[T]} \otimes \mathbf{x}_{[T]}] \cdot \mathrm{diag}_T(B(K_1 - K_2))^{\mathsf{H}} + 2\mathrm{diag}_T(\Lambda), \\
&\preceq 4\mathrm{diag}_T(B(K_1 - K_2)) \cdot \mathrm{diag}_T\left(\mathsf{dlyap}_{(2)}\left(A_{\mathrm{cl},1},\ \Lambda\right)\right) \cdot \mathrm{diag}_T(B(K_1 - K_2))^{\mathsf{H}} + 2\mathrm{diag}_T(\Lambda) \\
&= 4\mathrm{diag}_T\left(B(K_1 - K_2)\mathsf{dlyap}_{(2)}\left(A_{\mathrm{cl},1},\ \Lambda\right)(K_1 - K_2)^{\mathsf{H}} B^{\mathsf{H}}\right) + 2\mathrm{diag}_T(\Lambda) \\
&= \mathrm{diag}_T\left(4B(K_1 - K_2)\mathsf{dlyap}_{(2)}\left(A_{\mathrm{cl},1},\ \Lambda\right)(K_1 - K_2)^{\mathsf{H}} B^{\mathsf{H}} + 2\Lambda\right).
\end{aligned}$$

This concludes the proof.

□

B.1 Proof of Lemma 2.1

Proof. By Theorem B.1,

$$\mathsf{dlyap}\left((A_\star + B_\star K)^{\mathsf{H}},\ \Sigma_{\mathbf{w}}\right) \preceq \mathsf{dlyap}\left((A_\star + B_\star K_0)^{\mathsf{H}},\ Z\right) \tag{B.7}$$

where $Z = 2\Sigma_{\mathbf{w}} + 4B_\star B_\star^{\mathsf{H}} \|K - K_0\|_{\mathrm{op}}^2 \left\|\mathsf{dlyap}_{(2)}\left((A_\star + B_\star K)^{\mathsf{H}},\ \Sigma_{\mathbf{w}}\right)\right\|_{\mathrm{op}}$.

The remainder of the argument consists of simply bounding Z in terms of a constant times $B_\star B_\star^{\mathsf{H}} + \Sigma_{\mathbf{w}}$. By Lemma D.11, we have that for $P_K = P_\infty(K; A_\star, B_\star)$,

$$\left\|\mathsf{dlyap}_{(2)}\left((A_\star + B_\star K)^{\mathsf{H}},\ \Sigma_{\mathbf{w}}\right)\right\|_{\mathrm{op}} \ \leq\ n^2 \|P_K\|_{\mathrm{op}} \|\Sigma_{\mathbf{w}}\|_{\mathrm{op}} + (n^2 + 2n + 2) \|\Sigma_{\mathbf{w}}\|_{\mathrm{op}} \|P_K\|_{\mathrm{op}}^4 \exp\left(-\|P_K\|_{\mathrm{op}}^{-1} n\right).$$

Since $\|P_K\|_{\mathrm{op}} > 1$ (Lemma A.1), if we set $n = 4\|P_K\|_{\mathrm{op}} \log\left(3\|P_K\|_{\mathrm{op}}\right) \ \geq\ \left\lceil \|P_K\|_{\mathrm{op}} \log\left(5\|P_K\|_{\mathrm{op}}^3\right)\right\rceil$. This implies that,

$$n^2 \|P_K\|_{\mathrm{op}} \|\Sigma_{\mathbf{w}}\|_{\mathrm{op}} \ \geq\ (n^2 + 2n + 2) \|\Sigma_{\mathbf{w}}\|_{\mathrm{op}} \|P_K\|_{\mathrm{op}}^4 \exp\left(-\|P_K\|_{\mathrm{op}}^{-1} n\right)$$

and hence,

$$\left\|\mathsf{dlyap}_{(2)}\left((A_\star + B_\star K)^{\mathsf{H}},\ \Sigma_{\mathbf{w}}\right)\right\|_{\mathrm{op}} \ \leq\ 32 \|\Sigma_{\mathbf{w}}\|_{\mathrm{op}} \|P_K\|_{\mathrm{op}}^3 \log\left(3\|P_K\|_{\mathrm{op}}\right)^2.$$

Therefore,

$$\begin{aligned} Z &\preceq 2\Sigma_{\mathbf{w}} + \frac{128}{\sigma_{\mathbf{u}}^2} \|\Sigma_{\mathbf{w}}\|_{\mathrm{op}} \|K - K_0\|_{\mathrm{op}}^2 \|P_K\|_{\mathrm{op}}^3 \log\left(3\|P_K\|_{\mathrm{op}}\right)^2 B_\star B_\star^{\mathsf{H}} \sigma_{\mathbf{u}}^2 \\ &\preceq \max\left\{2, \frac{128}{\sigma_{\mathbf{u}}^2} \|\Sigma_{\mathbf{w}}\|_{\mathrm{op}} \|K - K_0\|_{\mathrm{op}}^2 \|P_K\|_{\mathrm{op}}^3 \log\left(3\|P_K\|_{\mathrm{op}}\right)^2\right\} (\Sigma_{\mathbf{w}} + \sigma_{\mathbf{u}}^2 B_\star B_\star^{\mathsf{H}}). \end{aligned}$$

Going back to (B.7),

$$\mathsf{dlyap}\left((A_\star + B_\star K)^{\mathsf{H}},\ \Sigma_{\mathbf{w}}\right) \preceq \mathcal{C}_{K,\sigma_{\mathbf{u}}^2} \cdot \Sigma_{\mathbf{x},0}$$

for $\mathcal{C}_{K,\sigma_{\mathbf{u}}^2} = \max\left\{2, \frac{128}{\sigma_{\mathbf{u}}^2} \|\Sigma_{\mathbf{w}}\|_{\mathrm{op}} \|K - K_0\|_{\mathrm{op}}^2 \|P_K\|_{\mathrm{op}}^3 \log\left(3\|P_K\|_{\mathrm{op}}\right)^2\right\}$. □

B.2 A Change of Controller Lemma

Lemma B.1. *Let K_1 be a stabilizing controller for the instance $A_\star, B_\star$. Then, for any stabilizing K_2,*

$$\mathsf{dlyap}\left((A_\star + B_\star K_1)^{\mathsf{H}},\ \Sigma_{\mathbf{x},0}\right) \preceq \mathcal{C}_K \cdot \mathsf{dlyap}\left((A_\star + B_\star K_2)^{\mathsf{H}},\ \Sigma_{\mathbf{x},0}\right)$$

for

$$\mathcal{C}_K := 2\left(1 + \frac{64\|K_2 - K_1\|_{\mathrm{op}}^2}{\sigma_{\mathbf{u}}^2} \|\Sigma_{\mathbf{x},0}\|_{\mathrm{op}} \|P_1\|_{\mathrm{op}}^3 \log(2\|P_1\|_{\mathrm{op}})^2\right)$$

where $P_1 := P_\infty(K_1; A_\star, B_\star)$

Proof. We apply Theorem B.1 to get that,

$$\mathsf{dlyap}\left((A_\star + B_\star K_1)^{\mathsf{H}},\ \Sigma_{\mathbf{x},0}\right) \preceq \mathsf{dlyap}\left((A_\star + B_\star K_2)^{\mathsf{H}},\ \overline{\Lambda}\right)$$

for

$$\overline{\Lambda} = 2\Sigma_{\mathbf{x},0} + 4B_\star(K_1 - K_2)\mathsf{dlyap}_{(2)}\left((A_\star + B_\star K_1)^{\mathsf{H}},\ \Sigma_{\mathbf{x},0}\right)(K_1 - K_2)^{\mathsf{H}} B_\star^{\mathsf{H}}.$$

Since $\Sigma_{\mathbf{x},0} = \mathsf{dlyap}\left((A_\star + B_\star K_0)^{\mathsf{H}},\ \sigma_{\mathbf{u}}^2 B_\star B_\star^{\mathsf{H}} + \Sigma_{\mathbf{w}}\right) \succeq \sigma_{\mathbf{u}}^2 B_\star B_\star^{\mathsf{H}}$, then letting $\Delta_{\mathsf{op}} = \|K_2 - K_1\|_{\mathrm{op}}$ we have that

$$\overline{\Lambda} \preceq 2\left(1 + \frac{2\Delta_{\mathsf{op}}^2}{\sigma_{\mathbf{u}}^2}\left\|\mathsf{dlyap}_{(2)}\left((A_\star + B_\star K_1)^{\mathsf{H}},\ \Sigma_{\mathbf{x},0}\right)\right\|_{\mathrm{op}}\right)\Sigma_{\mathbf{x},0}. \tag{B.8}$$

Using Lemma D.11, for any $n \geq 0$, we can upper bound $\left\|\mathsf{dlyap}_{(2)}\left((A_\star + B_\star K_1)^{\mathsf{H}},\ \Sigma_{\mathbf{x},0}\right)\right\|_{\mathrm{op}}$ by

$$n^2\left\|\mathsf{dlyap}\left((A_\star + B_\star K_1)^{\mathsf{H}},\ \Sigma_{\mathbf{x},0}\right)\right\|_{\mathrm{op}} + (n^2 + 2n + 2)\|\Sigma_{\mathbf{x},0}\|_{\mathrm{op}}\|P_1\|_{\mathrm{op}}^4 \exp\left(-n\|P_1\|_{\mathrm{op}}^{-1}\right). \tag{B.9}$$

And, by properties of dlyap (Lemma A.1),

$$\begin{aligned}
\left\|\mathsf{dlyap}\left((A_\star + B_\star K_1)^{\mathsf{H}},\ \Sigma_{\mathbf{x},0}\right)\right\|_{\mathrm{op}} &\leq \left\|\mathsf{dlyap}\left((A_\star + B_\star K_1)^{\mathsf{H}},\ I\right)\right\|_{\mathrm{op}}\|\Sigma_{\mathbf{x},0}\|_{\mathrm{op}} \\
&= \|\mathsf{dlyap}\left((A_\star + B_\star K_1),\ I\right)\|_{\mathrm{op}}\|\Sigma_{\mathbf{x},0}\|_{\mathrm{op}} \\
&\leq \|P_1\|_{\mathrm{op}}\|\Sigma_{\mathbf{x},0}\|_{\mathrm{op}}.
\end{aligned}$$

We can therefore upper bound Eq. (B.9) by:

$$n^2\|\Sigma_{\mathbf{x},0}\|_{\mathrm{op}}\|P_1\|_{\mathrm{op}} + (n^2 + 2n + 2)\|\Sigma_{\mathbf{x},0}\|_{\mathrm{op}}\|P_1\|_{\mathrm{op}}^4 \exp\left(-n\|P_1\|_{\mathrm{op}}^{-1}\right).$$

Setting $n = n_0 = \left\lceil \|P_1\|_{\mathrm{op}} \log\left(5\|P_1\|_{\mathrm{op}}^3\right)\right\rceil \leq 4\|P_1\|_{\mathrm{op}}\log(2\|P_1\|_{\mathrm{op}})$ (since $\|P_1\|_{\mathrm{op}} > 1$), we get that

$$(n_0^2 + 2n_0 + 2)\|\Sigma_{\mathbf{x},0}\|_{\mathrm{op}}\|P_1\|_{\mathrm{op}}^4 \exp\left(-n_0\|P_1\|_{\mathrm{op}}^{-1}\right) \leq n_0^2\|\Sigma_{\mathbf{x},0}\|_{\mathrm{op}}\|P_1\|_{\mathrm{op}}.$$

Therefore,

$$\left\|\mathsf{dlyap}_{(2)}\left((A_\star + B_\star K_1)^{\mathsf{H}},\ \Sigma_{\mathbf{x},0}\right)\right\|_{\mathrm{op}} \leq 32\|\Sigma_{\mathbf{x},0}\|_{\mathrm{op}}\|P_1\|_{\mathrm{op}}^3 \log(2\|P_1\|_{\mathrm{op}})^2$$

Plugging this upper bound into Eq. (B.8) finishes the proof. □

C Proof of Certainty Equivalence Perturbation Bounds

In this section, we provide the full proof of our end-to-end perturbation bound, Theorem 2.1, and the constituent results that comprise the argument. These supporting results are outlined in Appendix C.1 and Theorem 2.1 is proven in Appendix C.2. The constituent results given in Appendix C.1 are then proven in the subsequent sections. As indicated in the main body of the paper, we assume throughout our presentation that $Q \succ I$ and $R \succ I$.

C.1 Main Constituent Results

C.1.1 Performance Difference

We begin by stating our own variant of the now-ubiquitous performance difference lemma for LQR, which we use to bound the suboptimality of a controller K in terms of its Hilbert-Schmidt difference from $K_\star$, weighted by the initial exploration covariance $\Sigma_{\mathbf{x},0}$. This result follows by combining the standard LQR performance difference lemma with our change of measure result, Lemma 2.1.

Lemma C.1. *Let K be a stabilizing controller for $A_\star, B_\star$, then*

$$\mathcal{J}(K) - \mathcal{J}(K_\star) \ \le \ \mathcal{C}_{K,\sigma_{\mathbf{u}}^2} \left\| R + B_\star^{\mathsf{H}} P_\star B_\star \right\|_{\mathrm{op}} \left\| (K - K_\star)\Sigma_{\mathbf{x},0}^{1/2} \right\|_{\mathsf{HS}}^2$$

for $\mathcal{C}_{K,\sigma_{\mathbf{u}}^2} = \max\left\{2, \frac{128}{\sigma_{\mathbf{u}}^2} \|\Sigma_{\mathbf{w}}\|_{\mathrm{op}} \|K - K_0\|_{\mathrm{op}}^2 \|P_K\|_{\mathrm{op}}^3 \log\left(3\|P_K\|_{\mathrm{op}}\right)^2\right\}$ *for* $P_K = P_\infty(K; A_\star, B_\star)$ *defined as in Lemma 2.1.*

Proof. The proof follows by applying Lemma 2.1 on the standard performance difference lemma for LQR. By Lemma 12 in Fazel et al. [2018] (as presented in Lemma 4 from Mania et al. [2019]),

$$J_\star(K) - J_\star(K_\star) = \mathrm{tr}\left[\Sigma_\star(K)(K - K_\star)^{\mathsf{H}}(R + B_\star^{\mathsf{H}} P_\star B_\star)(K - K_\star)\right].$$

Now, by Lemma 2.1, $\Sigma_\star(K) \preceq \mathcal{C}_{K,\sigma_{\mathbf{u}}^2} \cdot \Sigma_{\mathbf{x},0}$., and therefore

$$\begin{aligned} \mathrm{tr}\left[\Sigma_\star(K)(K - K_\star)^{\mathsf{H}}(R + B_\star^{\mathsf{H}} P_\star B_\star)(K - K_\star)\right] \quad &\le \ \mathcal{C}_{K,\sigma_{\mathbf{u}}^2} \mathrm{tr}\left[\Sigma_{\mathbf{x},0}(K - K_\star)^{\mathsf{H}}(R + B_\star^{\mathsf{H}} P_\star B_\star)(K - K_\star)\right] \\ &\le \ \mathcal{C}_{K,\sigma_{\mathbf{u}}^2} \left\| R + B_\star^{\mathsf{H}} P_\star B_\star \right\|_{\mathrm{op}} \left\| (K - K_\star)\Sigma_{\mathbf{x},0}^{1/2} \right\|_{\mathsf{HS}}^2. \end{aligned}$$

□

C.1.2 Intermediate K Perturbation

Next, we give a bound controlling the error $\|(K_\star - \widehat{K})\Sigma_{\mathbf{x},0}^{1/2}\|_{\mathsf{HS}}$ in terms of the maximum of the errors,

$$\max\left\{\|\widehat{B} - B_\star\|_{\mathsf{HS}}, \|(\widehat{A} - A_\star)\Sigma_{\mathbf{x},0}^{1/2}\|_{\mathsf{HS}}, \|\Sigma_{\mathbf{x},0}^{1/2}(\widehat{P} - P_\star)\Sigma_{\mathbf{x},0}^{1/2}\|_{\mathsf{HS}}\right\}.$$

The following proposition is proven in Appendix C.3.

Proposition C.2. *Recall our earlier definitions, $\widehat{P} = P_\infty(\widehat{A}, \widehat{B})$, $\widehat{K} = K_\infty(\widehat{A}, \widehat{B})$. Assume that the following inequality holds for some $\sigma_{\mathbf{u}}^2 \ \ge \ 1$, and $\varepsilon \ \le \ 1$,*

$$\max\left\{\|\widehat{B} - B_\star\|_{\mathsf{HS}}, \|(\widehat{A} - A_\star)\Sigma_{\mathbf{x},0}^{1/2}\|_{\mathsf{HS}}, \|\Sigma_{\mathbf{x},0}^{1/2}(\widehat{P} - P_\star)\Sigma_{\mathbf{x},0}^{1/2}\|_{\mathsf{HS}}\right\} \ \le \ \varepsilon.$$

Furthermore, suppose $(\widehat{A}, \widehat{B})$ is also stabilizable, let $P_0 = P_\infty(K_0; A_\star, B_\star)$, and set

$$M_K = \max\left\{\|A_\star\|_{\mathrm{op}}^2, \|B_\star\|_{\mathrm{op}}^2, \|P_\star\|_{\mathrm{op}}, \|\widehat{P}\|_{\mathrm{op}}, \|P_0\|_{\mathrm{op}}, \|\Sigma_{\mathbf{w}}\|_{\mathrm{op}}, 1\right\},$$

be a uniform bound on the operator norm on relevant operators. Then,

$$\left\|(K_\star - \widehat{K})\Sigma_{\mathbf{x},0}^{1/2}\right\|_{\mathsf{HS}} \ \le \ 9\sigma_{\mathbf{u}}\varepsilon M_K^4.$$

The two challenges in applying Proposition C.2 are (a) verifying that the nominal system $(\widehat{A}, \widehat{B})$ is stabilizable, so that the uniform bound M is finite, and (b) bounding the weighted error $\Sigma_{\mathbf{x},0}^{1/2}(\widehat{P} - P_\star)\Sigma_{\mathbf{x},0}^{1/2}$. We address these two parts of the argument in what follows.

C.1.3 Operator-Norm P-Perturbation

Proposition C.3. *Fix two instances (A_1, B_1), (A_2, B_2), and define $\varepsilon_{\mathrm{op}} = \max\{\|B_1 - B_2\|_{\mathrm{op}}, \|A_1 - A_2\|_{\mathrm{op}}\}$. Suppose $P_1 = P_\infty(A_1, B_1)$, $K_1 := K_\infty(A_1, B_1)$, and fix a tolerance parameter $\eta \in (0, 1]$. Then,*

- *If $\varepsilon_{\mathrm{op}} \le \eta/(16\|P_1\|_{\mathrm{op}}^3)$, then $P_2 = P_\infty(A_2, B_2)$ and $P_\infty(K_1; A_2, B_2)$ are bounded operators, and*

$$P_2 \preceq P_\infty(K_1; A_2, B_2) \preceq P_1 + \eta\|P_1\|_{\mathrm{op}} I \preceq (1 + \|P_1\|)\eta P_1 \tag{C.1}$$

- *If the stronger condition $\varepsilon_{\mathrm{op}} \le \eta/(16(1 + \eta)^4\|P_1\|_{\mathrm{op}}^3)$ holds, then,$\|P_2 - P_1\|_{\mathrm{op}} \le \eta\|P_1\|_{\mathrm{op}}$.*

The proof is deferred to Appendix C.4. One important consequence of the above bound is that, by setting $\eta = 1/11$, Proposition C.3 shows that the closeness condition, Condition 2.1, implies that $\|\widehat{P}\|_{\mathrm{op}} \le 1.2\|P_\star\|_{\mathrm{op}}$ (see Lemma C.8). This is an essential ingredient in the subsequent covariance-weighted bound.

C.1.4 Covariance-Weighted Perturbation of $P_\star$

To state the covariance-weighted perturbation of $P_\star$, we recall the uniform closeness condition: Condition 2.1.

$$\varepsilon_{\mathsf{op}} := \max\left\{\left\|\widehat{A} - A_\star\right\|_{\mathrm{op}}, \left\|\widehat{B} - B_\star\right\|_{\mathrm{op}}\right\} \;\le\; \frac{1}{229\,\|P_\star\|_{\mathrm{op}}^3} \tag{C.2}$$

For the following proposition, we assume access to a stabilizing controller K_0, and define $P_0 := \|P_\infty(K_0; A_\star, B_\star)\|_{\mathrm{op}}$. We set $M_{P_\star} := \|P_\star\|_{\mathrm{op}}$ and $M_{P_0} = \|P_0\|_{\mathrm{op}}$

Proposition C.4. *Suppose $\widehat{A}, \widehat{B}$ satisfy Condition 2.1 and recall our earlier definition $\widehat{P} = P_\infty(\widehat{A}, \widehat{B})$. Then,*

$$\left\|\Sigma_{\mathbf{x},0}^{1/2}(\widehat{P} - P_\star)\Sigma_{\mathbf{x},0}^{1/2}\right\|_{\mathsf{HS}} \lesssim \mathcal{C}_P \cdot \varepsilon_P \cdot \sqrt{\log_+(\kappa_P)} \cdot \phi(\kappa_P)^{\alpha_{\mathrm{op}}},$$

where $\phi(u) := e^{\sqrt{\log(u)}}$ and $\varepsilon_P, \mathcal{C}_P, \kappa_P, \alpha_{\mathrm{op}}$ are defined as:

$$\begin{aligned}
\varepsilon_P^2 &:= 2\|(A_\star - \widehat{A})\Sigma_{\mathbf{x},0}^{1/2}\|_{\mathsf{HS}}^2 + 2.4 M_{P_\star}\|B_\star - \widehat{B}\|_{\mathsf{HS}}^2 \|\Sigma_{\mathbf{x},0}\|_{\mathrm{op}} \\
\mathcal{C}_P &:= M_{P_0}^4 M_{P_\star}^{3/2}\sqrt{(1 + \tfrac{\|\Sigma_{\mathbf{x},0}\|_{\mathrm{op}}}{\sigma_{\mathbf{u}}^2})\,\|\Sigma_{\mathbf{x},0}\|_{\mathrm{op}}} \\
\kappa_P &:= 1 + 2\frac{\|\widehat{A} - A_\star\|_{\mathrm{op}}^2 \mathrm{tr}[\Sigma_{\mathbf{x},0}]}{\varepsilon_P^2} \\
\alpha_{\mathrm{op}} &:= 2M_{P_\star}^{5/2}\varepsilon_{\mathrm{op}} \le 1/100.
\end{aligned}$$

Furthermore, the above bound holds for any $\varepsilon_P' \ge \varepsilon_P$. For the case of finite-dimensional systems where $\Sigma_{\mathbf{x},0} \succ 0$, κ_P can be replaced by $1 + \mathrm{cond}(\Sigma_{\mathbf{x},0})$, where $\mathrm{cond}(\cdot)$ denotes the condition number.

Proposition C.4 is the most involved and technically innovative in our analysis. Its proof is presented in Appendix C.5.

C.2 Proof of the End-to-End Perturbation Bound: Theorem 2.1

We recall:

$$M_\star := \max\{\|A_\star\|_{\mathrm{op}}^2, \|B_\star\|_{\mathrm{op}}^2, \|P_\star\|_{\mathrm{op}}, \|\Sigma_{\mathbf{w}}\|_{\mathrm{op}}, 1\}. \tag{C.3}$$

Proof. The proof follows by combining the performance difference lemma (Lemma C.1), our controller perturbation bound (Proposition C.2), and the riccati perturbation bound (Proposition C.4).

Applying performance difference Starting from the performance difference lemma, we have that

$$J(\widehat{K}) - J(K_\star) \;\le\; \mathcal{C}_{\widehat{K},\sigma_{\mathbf{u}}^2}\left\|R + B_\star^{\mathsf{H}} P_\star B_\star\right\|_{\mathrm{op}} \left\|(\widehat{K} - K_\star)\Sigma_{\mathbf{x},0}^{1/2}\right\|_{\mathsf{HS}}^2$$

for $\mathcal{C}_{\widehat{K},\sigma_{\mathbf{u}}^2} = \max\left\{2, \frac{128}{\sigma_{\mathbf{u}}^2}\|\Sigma_{\mathbf{w}}\|_{\mathrm{op}}\left\|\widehat{K} - K_0\right\|_{\mathrm{op}}^2 \left\|P_{\widehat{K}}\right\|_{\mathrm{op}}^3 \log\left(2\left\|P_{\widehat{K}}\right\|_{\mathrm{op}}\right)^2\right\}$ where $P_{\widehat{K}} = P_\infty(\widehat{K}; A_\star, B_\star)$.

We now simplify these terms. Since $K_0 = K_\infty(A_0, B_0)$ and $\widehat{K} = K_\infty(\widehat{A}, \widehat{B})$ where both pairs of systems (A_0, B_0), $(\widehat{A}, \widehat{B})$ satisfy the uniform closeness conditions, we can apply Lemma C.8 to conclude that

$$P_\infty(\widehat{K}; A_\star, B_\star) \preceq 1.2 P_\star, \quad \left\|P_\infty(\widehat{A}, \widehat{B})\right\|_{\mathrm{op}} \;\le\; 1.1\|P_\star\|_{\mathrm{op}}, \quad \text{and } \|P_\infty(A_0, B_0)\|_{\mathrm{op}} \;\le\; 1.1\|P_\star\|_{\mathrm{op}}.$$

Furthermore, by Lemma D.7, $\|\widehat{K}\|_{\mathrm{op}}^2 \;\le\; \|P_\infty(\widehat{A}, \widehat{B})\|_{\mathrm{op}}$ and similarly, $\|K_0\|_{\mathrm{op}}^2 \;\le\; \|P_\infty(A_0, B_0)\|_{\mathrm{op}}$. Therefore,

$$\|\widehat{K} - K_0\|_{\mathrm{op}}^2 \;\le\; 2(\|\widehat{K}\|_{\mathrm{op}}^2 + \|K_0\|_{\mathrm{op}}^2) \;\le\; 4.4\|P_\star\|_{\mathrm{op}}.$$

Recalling that $M_\star = \|P_\star\|_{\mathrm{op}} \geq 1$ and $\sigma_{\mathbf{u}}^2 \geq 1$,

$$\frac{128}{\sigma_{\mathbf{u}}^2} \|\Sigma_{\mathbf{w}}\|_{\mathrm{op}} \left\|\widehat{K} - K_0\right\|_{\mathrm{op}}^2 \|P_{\widehat{K}}\|_{\mathrm{op}}^3 \log\left(2\|P_{\widehat{K}}\|_{\mathrm{op}}\right)^2 \lesssim M_\star^5 \log(M_\star)^2.$$

Therefore, under the uniform closeness assumptions, we have that,

$$J(\widehat{K}) - J(K_\star) \lesssim M_\star^7 \log(M_\star)^2 \left\|(\widehat{K} - K_\star)\Sigma_{\mathbf{x},0}^{1/2}\right\|_{\mathsf{HS}}^2, \tag{C.4}$$

where we have used the calculation $\left\|R + B_\star^{\mathsf{H}} P_\star B_\star\right\|_{\mathrm{op}} \leq \|R\|_{\mathrm{op}} + \left\|B_\star^{\mathsf{H}} P_\star B_\star\right\|_{\mathrm{op}} \lesssim M_\star^2$.

Controller perturbation Now, we include our controller perturbation bound (Proposition C.2) which states that for M_K defined as,

$$M_K = \max\left\{\|A_\star\|_{\mathrm{op}}^2, \|B_\star\|_{\mathrm{op}}^2, \|P_\star\|_{\mathrm{op}}, \|\widehat{P}\|_{\mathrm{op}}, \|P_0\|_{\mathrm{op}}, \|\Sigma_{\mathbf{w}}\|_{\mathrm{op}}, 1\right\},$$

we have that

$$\left\|(K_\star - \widehat{K})\Sigma_{\mathbf{x},0}^{1/2}\right\|_{\mathsf{HS}} \leq 9\sigma_{\mathbf{u}}\varepsilon M_K^4.$$

for

$$\varepsilon = \max\left\{\left\|\widehat{B} - B_\star\right\|_{\mathsf{HS}}^2, \left\|(\widehat{A} - A_\star)\Sigma_{\mathbf{x},0}^{1/2}\right\|_{\mathsf{HS}}^2, \left\|\Sigma_{\mathbf{x},0}^{1/2}(\widehat{P} - P_\star)\Sigma_{\mathbf{x},0}^{1/2}\right\|_{\mathsf{HS}}^2\right\}.$$

Under the uniform operator norm closeness, we again have $\|\widehat{P}\|_{\mathrm{op}} \lesssim \|P_\star\|_{\mathrm{op}} \lesssim M_\star$, and similarly for P_0. Therefore, M_K as defined in Proposition C.2 satisfies $M_K \lesssim M_\star$. Using the controller perturbation, we then conclude that,

$$\left\|(\widehat{K} - K_\star)\Sigma_{\mathbf{x},0}^{1/2}\right\|_{\mathsf{HS}}^2 \lesssim \sigma_{\mathbf{u}} M_\star^9 \log(M_\star)^2 \varepsilon. \tag{C.5}$$

Riccati perturbation The last step in the proof is to apply our Riccati perturbation bound from Proposition C.4. The bound states that,

$$\left\|\Sigma_{\mathbf{x},0}^{1/2}(\widehat{P} - P_\star)\Sigma_{\mathbf{x},0}^{1/2}\right\|_{\mathsf{HS}}^2 \leq \mathcal{C}_P^2 \cdot \varepsilon_P^2 \cdot \log_+(\kappa_P) \cdot \phi(\kappa_P)^{1/50}$$

where $\phi(u) := e^{\sqrt{\log(u)}}$, and the remaining quantities are defined as,

$$\begin{aligned}
\varepsilon_P^2 &:= 2\|(A_\star - \widehat{A})\Sigma_{\mathbf{x},0}^{1/2}\|_{\mathsf{HS}}^2 + 2.4\|P_\star\|_{\mathrm{op}}\left\|B_\star - \widehat{B}\right\|_{\mathsf{HS}}^2 \|\Sigma_{\mathbf{x},0}\|_{\mathrm{op}} \\
\mathcal{C}_P &\lesssim \|P_\infty(A_0, B_0)\|_{\mathrm{op}}^4 \|P_\star\|_{\mathrm{op}}^{3/2} \sqrt{\left(1 + \tfrac{\|\Sigma_{\mathbf{x},0}\|_{\mathrm{op}}}{\sigma_{\mathbf{u}}^2}\right)\|\Sigma_{\mathbf{x},0}\|_{\mathrm{op}}} \\
\kappa_P &:= 1 + 2\tfrac{\|\widehat{A} - A_\star\|_{\mathrm{op}}^2 \mathrm{tr}[\Sigma_{\mathbf{x},0}]}{\varepsilon_P^2}
\end{aligned}$$

We note that by Condition 2.1, $\|P_\infty(A_0, B_0)\|_{\mathrm{op}} \lesssim \|P_\star\|_{\mathrm{op}}$. Furthermore, by Lemma A.1 and the fact $\sigma_{\mathbf{u}}^2 \geq 1$,

$$\begin{aligned}
\|\Sigma_{\mathbf{x},0}\|_{\mathrm{op}} &= \left\|\mathsf{dlyap}\left((A_\star + B_\star K_0)^{\mathsf{H}},\ B_\star B_\star^{\mathsf{H}}\sigma_{\mathbf{u}}^2 + \Sigma_{\mathbf{w}}\right)\right\|_{\mathrm{op}} \\
&\leq \left\|\mathsf{dlyap}\left((A_\star + B_\star K_0)^{\mathsf{H}},\ I\right)\right\|_{\mathrm{op}} \left\|B_\star B_\star^{\mathsf{H}}\sigma_{\mathbf{u}}^2 + \Sigma_{\mathbf{w}}\right\|_{\mathrm{op}} \\
&= \|\mathsf{dlyap}(A_\star + B_\star K_0,\ I)\|_{\mathrm{op}} \left\|B_\star B_\star^{\mathsf{H}}\sigma_{\mathbf{u}}^2 + \Sigma_{\mathbf{w}}\right\|_{\mathrm{op}} \\
&\leq M_\star(\sigma_{\mathbf{u}}^2 M_\star),
\end{aligned}$$

where we have used the fact that $I \preceq Q$ and hence

$$\|\mathsf{dlyap}\,(A_\star + B_\star K_0,\ I)\|_{\mathrm{op}} \ \le\ \left\|\mathsf{dlyap}\left(A_\star + B_\star K_0,\ Q + K_\star^{\mathsf{H}} R K_\star\right)\right\|_{\mathrm{op}} = \|P_\star\|_{\mathrm{op}}.$$

Using this calculation, we can then bound the relevant quantities as:

$$\begin{aligned}
\mathcal{C}_P^2 &\lesssim M_\star^{11}(1 + \|\Sigma_{\mathbf{x},0}\|_{\mathrm{op}}/\sigma_{\mathbf{u}}^2)\,\|\Sigma_{\mathbf{x},0}\|_{\mathrm{op}} \lesssim \sigma_{\mathbf{u}}^2 M_\star^{15}\\
\varepsilon_P^2 &\lesssim M_\star^3 \varepsilon^2\\
\log_+(\kappa_P) \ &\le\ \log\left(e + \frac{2e\|\widehat{A} - A_\star\|_{\mathrm{op}}^2 \mathrm{tr}\,[\Sigma_{\mathbf{x},0}]}{\varepsilon^2}\right) := \mathcal{L},
\end{aligned}$$

where we also note that Proposition C.4 allows us to replace $\mathcal{L}$ by $\log(1 + \mathrm{cond}(\Sigma_{\mathbf{x},0}))$ in finite dimension with $\Sigma_{\mathbf{x},0} \succ 0$. Using these simplifications, we get that,

$$\left\|\Sigma_{\mathbf{x},0}^{1/2}(\widehat{P} - P_\star)\Sigma_{\mathbf{x},0}^{1/2}\right\|_{\mathsf{HS}}^2 \lesssim \sigma_{\mathbf{u}}^2 M_\star^{18}\sqrt{\mathcal{L}}\exp(\frac{1}{50}\sqrt{\mathcal{L}})\varepsilon^2. \tag{C.6}$$

Wrapping up Combining this last equation with the bounds from Eq. (C.4) and Eq. (C.5), the total power of $M_\star$ is $M_\star^{(7+9+18)}\log(M_\star)^2 \ \le\ M_\star^{36}$, yielding:

$$J(\widehat{K}) - J(K_\star) \lesssim \sigma_{\mathbf{u}}^4 M_\star^{36}\mathcal{L}^{1/2}\exp(\frac{1}{50}\sqrt{\mathcal{L}})\varepsilon^2.$$

□

C.3 Proof of Intermediate K Perturbation: Proposition C.2

Recall the definition

$$M_K = \max\left\{\|A_\star\|_{\mathrm{op}}^2, \|B_\star\|_{\mathrm{op}}^2, \|P_\star\|_{\mathrm{op}}, \|\widehat{P}\|_{\mathrm{op}}, \|P_0\|_{\mathrm{op}}, \|\Sigma_{\mathbf{w}}\|_{\mathrm{op}}, 1,\right\},$$

where $P_0 = P_\infty(K_0; A_\star, B_\star)$. To simplify notation, we use $M = M_K$. Furthermore, we assume $\sigma_{\mathbf{u}}^2 \ \ge\ 1$, as is chosen in our algorithm later on. Our proof appeals to the following lemma from Mania et al. [2019]:

Lemma C.5 (Mania et al. [2019]). *Let f_1, f_2 be γ-strongly convex functions. Let $\mathbf{x}_i = \arg\min f_i(\mathbf{x})$ for $i = 1, 2$. If $\|\nabla f_1(\mathbf{x}_2)\| \ \le\ \varepsilon$ then, $\|\mathbf{x}_1 - \mathbf{x}_2\| \ \le\ \varepsilon/\gamma$*

Proof of Proposition C.2. The proof is inspired by that of Lemma 2 in Mania et al. [2019]. Consider the functions $f^\star$ and $\widehat{f}$ defined as,

$$\begin{aligned}
f^\star(X) &:= \frac{1}{2}\left\|R^{1/2}X\right\|_{\mathsf{HS}}^2 + \frac{1}{2}\left\|P_\star^{1/2}B_\star X\right\|_{\mathsf{HS}}^2 + \left\langle B_\star^{\mathsf{H}} P_\star A_\star \Sigma_{\mathbf{x},0}^{1/2},\ X\right\rangle_{\mathsf{HS}}\\
\widehat{f}(X) &:= \frac{1}{2}\left\|R^{1/2}X\right\|_{\mathsf{HS}}^2 + \frac{1}{2}\left\|\widehat{P}^{1/2}\widehat{B}X\right\|_{\mathsf{HS}}^2 + \left\langle \widehat{B}^{\mathsf{H}}\widehat{P}\widehat{A}\Sigma_{\mathbf{x},0}^{1/2},\ X\right\rangle_{\mathsf{HS}},
\end{aligned}$$

where $\langle A,\ B\rangle_{\mathsf{HS}} = \mathrm{tr}\left[A^{\mathsf{H}}B\right]$ denotes the Hilbert-Schmidt inner product. Both functions are strongly convex with strong convexity parameter lower bounded by $\sigma_{\min}(R) \ \ge\ 1$. We observe that,

$$\begin{aligned}
\nabla f^\star(X) &= (B_\star^{\mathsf{H}} P_\star B_\star + R)X + B_\star^{\mathsf{H}} P_\star A_\star \Sigma_{\mathbf{x},0}^{1/2}.\\
\nabla \widehat{f}(X) &= (\widehat{B}^{\mathsf{H}}\widehat{P}\widehat{B} + R)X + \widehat{B}^{\mathsf{H}}\widehat{P}\widehat{A}\Sigma_{\mathbf{x},0}^{1/2}.
\end{aligned}$$

and hence,

$$\begin{aligned}
X_\star &= \arg\min_X f^\star(X) = -(B_\star^{\mathsf{H}} P_\star B_\star + R)^{-1} B_\star^{\mathsf{H}} P_\star A_\star \Sigma_{\mathbf{x},0}^{1/2} = K_\star \Sigma_{\mathbf{x},0}^{1/2}\\
\widehat{X} &= \arg\min_X \widehat{f}(X) = -(\widehat{B}^{\mathsf{H}}\widehat{P}\widehat{B} + R)^{-1}\widehat{B}^{\mathsf{H}}\widehat{P}\widehat{A}\Sigma_{\mathbf{x},0}^{1/2} = \widehat{K}\Sigma_{\mathbf{x},0}^{1/2}.
\end{aligned}$$

Next, we show that the norm of the difference between both gradients is small. We will repeatedly use the fact that $\|AB\|_{\mathsf{HS}} \le \|A\|_{\mathrm{op}} \|B\|_{\mathsf{HS}}$ (and similarly $\|AB\|_{\mathsf{HS}} \le \|B\|_{\mathrm{op}} \|A\|_{\mathsf{HS}}$) throughout the remainder of the proof.

$$\left\|\nabla f^\star(X) - \nabla \widehat{f}(X)\right\|_{\mathsf{HS}} \le \left\|B_\star^{\mathsf{H}} P_\star B_\star - \widehat{B}^{\mathsf{H}} \widehat{P} \widehat{B}\right\|_{\mathsf{HS}} \|X\|_{\mathrm{op}} + \left\|(B_\star^{\mathsf{H}} P_\star A_\star - \widehat{B}^{\mathsf{H}} \widehat{P} \widehat{A})\Sigma_{\mathbf{x},0}^{1/2}\right\|_{\mathsf{HS}}. \tag{C.7}$$

Bounding the first term in the above decomposition,

$$\begin{aligned}
\left\|\widehat{B}^{\mathsf{H}} \widehat{P} \widehat{B} - B_\star^{\mathsf{H}} P_\star B_\star\right\|_{\mathsf{HS}} &\overset{(i)}{\le} \left\|\widehat{B}^{\mathsf{H}} \widehat{P} \widehat{B} - B_\star^{\mathsf{H}} \widehat{P} B_\star\right\|_{\mathsf{HS}} + \left\|B_\star^{\mathsf{H}} (\widehat{P} - P_\star) B_\star\right\|_{\mathsf{HS}} \\
&\le \left\|\widehat{B}^{\mathsf{H}} \widehat{P} \widehat{B} \pm \widehat{B}^{\mathsf{H}} \widehat{P} B_\star - B_\star^{\mathsf{H}} \widehat{P} B_\star\right\|_{\mathsf{HS}} + \varepsilon \\
&\le \left\|\widehat{B}^{\mathsf{H}} \widehat{P} (\widehat{B} - B_\star)\right\|_{\mathsf{HS}} + \left\|(\widehat{B} - B_\star) \widehat{P} B_\star\right\|_{\mathsf{HS}} + \varepsilon \\
&\le 4M^{3/2}\varepsilon,
\end{aligned}$$

where in the last line, we used $\|\widehat{B}\|_{\mathrm{op}} \le \varepsilon + \|B_\star\|_{\mathrm{op}} \le 2\sqrt{M}$, since $\varepsilon \le 1$ and $M \ge \max\{1, \|B_\star\|_{\mathrm{op}}^2\}$. In inequality (i), we used the following calculation:

$$\begin{aligned}
\left\|B_\star^{\mathsf{H}} (\widehat{P} - P_\star) B_\star\right\|_{\mathsf{HS}}^2 &= \operatorname{tr}\left[B_\star^{\mathsf{H}} (\widehat{P} - P_\star) B_\star B_\star^{\mathsf{H}} (\widehat{P} - P_\star) B_\star\right] \\
&\le \operatorname{tr}\left[B_\star^{\mathsf{H}} (\widehat{P} - P_\star) \Sigma_{\mathbf{x},0} (\widehat{P} - P_\star) B_\star\right] && (B_\star B_\star^{\mathsf{H}} \preceq \Sigma_{\mathbf{x},0}) \\
&= \operatorname{tr}\left[\Sigma_{\mathbf{x},0}^{1/2} (\widehat{P} - P_\star) B_\star B_\star^{\mathsf{H}} (\widehat{P} - P_\star) \Sigma_{\mathbf{x},0}^{1/2}\right] \\
&\le \operatorname{tr}\left[\Sigma_{\mathbf{x},0}^{1/2} (\widehat{P} - P_\star) \Sigma_{\mathbf{x},0} (\widehat{P} - P_\star) \Sigma_{\mathbf{x},0}^{1/2}\right].
\end{aligned}$$

The last line is equal to $\left\|\Sigma_{\mathbf{x},0}^{1/2} (\widehat{P} - P_\star) \Sigma_{\mathbf{x},0}^{1/2}\right\|_{\mathsf{HS}}^2$ which is less than ε^2 by assumption. Next, we bound the second term in Eq. (C.7),

$$\left\|(\widehat{B}^{\mathsf{H}} \widehat{P} \widehat{A} - B_\star^{\mathsf{H}} P_\star A_\star)\Sigma_{\mathbf{x},0}^{1/2}\right\|_{\mathsf{HS}} \le \underbrace{\left\|(\widehat{B}^{\mathsf{H}} \widehat{P} \widehat{A} - B_\star^{\mathsf{H}} \widehat{P} A_\star)\Sigma_{\mathbf{x},0}^{1/2}\right\|_{\mathsf{HS}}}_{T_1} + \underbrace{\left\|B_\star^{\mathsf{H}} (\widehat{P} - P_\star) A_\star \Sigma_{\mathbf{x},0}^{1/2}\right\|_{\mathsf{HS}}}_{T_2}.$$

We bound T_1 as follows:

$$\begin{aligned}
\left\|(\widehat{B}^{\mathsf{H}} \widehat{P} \widehat{A} \pm \widehat{B}^{\mathsf{H}} \widehat{P} A_\star - B_\star^{\mathsf{H}} \widehat{P} A_\star)\Sigma_{\mathbf{x},0}^{1/2}\right\|_{\mathsf{HS}} &\le \left\|\widehat{B}^{\mathsf{H}} \widehat{P} (\widehat{A} - A_\star)\Sigma_{\mathbf{x},0}^{1/2}\right\|_{\mathsf{HS}} + \left\|(\widehat{B} - B_\star)^{\mathsf{H}} \widehat{P} A_\star \Sigma_{\mathbf{x},0}^{1/2}\right\|_{\mathsf{HS}} \\
&\le \varepsilon \left\|\widehat{B} \widehat{P}\right\|_{\mathrm{op}} + \varepsilon \left\|\widehat{P} A_\star \Sigma_{\mathbf{x},0}^{1/2}\right\|_{\mathrm{op}} \\
&\le \left(2M^{3/2} + M^{3/2} \left\|\Sigma_{\mathbf{x},0}^{1/2}\right\|_{\mathrm{op}}\right)\varepsilon,
\end{aligned}$$

where again we used $\|\widehat{B}\| \le \sqrt{2M}$. Before bounding T_2, we observe that:

$$\begin{aligned}
A_\star \Sigma_{\mathbf{x},0} A_\star^{\mathsf{H}} &\preceq 2(A_\star + B_\star K_0)\Sigma_{\mathbf{x},0}(A_\star + B_\star K_0)^{\mathsf{H}} + 2(B_\star K_0)\Sigma_{\mathbf{x},0}(B_\star K_0)^{\mathsf{H}} \\
&\preceq 2\Sigma_{\mathbf{x},0} + 2\|\Sigma_{\mathbf{x},0}\|_{\mathrm{op}} \|K_0\|_{\mathrm{op}}^2 B_\star B_\star^{\mathsf{H}} \\
&\preceq 2(1 + \|\Sigma_{\mathbf{x},0}\|_{\mathrm{op}} \|K_0\|_{\mathrm{op}}^2)\Sigma_{\mathbf{x},0}.
\end{aligned} \tag{C.8}$$

To go from the first to the second line, we have used the fact that since

$$\Sigma_{\mathbf{x},0} = \mathsf{dlyap}\left((A_\star + B_\star K_0)^{\mathsf{H}},\ B_\star B_\star^{\mathsf{H}} \sigma_{\mathbf{u}}^2 + \Sigma_{\mathbf{w}}\right),$$

by definition of dlyap,

$$(A_\star + B_\star K_0)\Sigma_{\mathbf{x},0}(A_\star + B_\star K_0)^{\mathsf{H}} = \Sigma_{\mathbf{x},0} - B_\star B_\star^{\mathsf{H}}\sigma_{\mathbf{u}}^2 - \Sigma_{\mathbf{w}} \preceq \Sigma_{\mathbf{x},0}.$$

Applying (C.8), we can now bound T_2 using a similar calculation as before,

$$\begin{aligned}
\left\| B_\star^{\mathsf{H}}(\widehat{P} - P_\star)A_\star\Sigma_{\mathbf{x},0}^{1/2}\right\|_{\mathsf{HS}}^2 &= \mathrm{tr}\left[B_\star^{\mathsf{H}}(\widehat{P} - P_\star)A_\star\Sigma_{\mathbf{x},0}A_\star^{\mathsf{H}}(\widehat{P} - P_\star)B_\star\right] \\
&\leq\ 2(1 + \|K_0\|_{\mathrm{op}}^2 \|\Sigma_{\mathbf{x},0}\|_{\mathrm{op}})\mathrm{tr}\left[B_\star^{\mathsf{H}}(\widehat{P} - P_\star)\Sigma_{\mathbf{x},0}(\widehat{P} - P_\star)B_\star\right] \\
&= 2(1 + \|K_0\|_{\mathrm{op}}^2 \|\Sigma_{\mathbf{x},0}\|_{\mathrm{op}})\left\|\Sigma_{\mathbf{x},0}^{1/2}(\widehat{P} - P_\star)\Sigma_{\mathbf{x},0}^{1/2}\right\|_{\mathsf{HS}}^2 \\
&\leq\ 2(1 + \|K_0\|_{\mathrm{op}}^2 \|\Sigma_{\mathbf{x},0}\|_{\mathrm{op}})\varepsilon^2.
\end{aligned}$$

Returning to Eq. (C.7),

$$\begin{aligned}
\|\nabla f^\star(X) - \nabla_{\mathbf{u}}\widehat{f}(X)\| \ &\leq\ 5M^{3/2}\varepsilon\|X\|_{\mathrm{op}} \\
&+ \left(2M^{3/2} + M^{3/2}\left\|\Sigma_{\mathbf{x},0}^{1/2}\right\|_{\mathrm{op}} + \sqrt{2(1 + \|K_0\|_{\mathrm{op}}^2 \|\Sigma_{\mathbf{x},0}\|_{\mathrm{op}})}\right)\varepsilon,
\end{aligned}$$

and for $X = X_\star = K_\star\Sigma_{\mathbf{x},0}^{1/2}$, we have that,

$$\begin{aligned}
\|\nabla\widehat{f}(X_\star)\| \ &\leq\ 4M^{3/2}\varepsilon\left\|K_\star\Sigma_{\mathbf{x},0}^{1/2}\right\|_{\mathrm{op}} \\
&+ \left(2M^{3/2} + M^{3/2}\left\|\Sigma_{\mathbf{x},0}^{1/2}\right\|_{\mathrm{op}} + \sqrt{2(1 + \|K_0\|_{\mathrm{op}}^2 \|\Sigma_{\mathbf{x},0}\|_{\mathrm{op}})}\right)\varepsilon.
\end{aligned}$$

We now simplify the above. By Lemma D.7, we can take $\|K_\star\|_{\mathrm{op}} \leq \|P_\star\|_{\mathrm{op}}^{1/2} \leq M^{1/2}$, and $\|K_0\|_{\mathrm{op}} \leq \|P_{K_0}\|_{\mathrm{op}}^{1/2} \leq M^{1/2}$, where $P_{K_0} = P_\infty(K_0; A_\star, B_\star)$. Moreover, Lemma D.9 yields $\|\Sigma_{\mathbf{x},0}\|_{\mathrm{op}} \leq \|P_{K_0}\|_{\mathrm{op}}^2 (\|\Sigma_{\mathbf{w}}\|_{\mathrm{op}} + \sigma_{\mathbf{u}}^2\|B_\star\|_{\mathrm{op}}^2) \leq \sigma_{\mathbf{u}}^2 M^3$ (since $\sigma_{\mathbf{u}}^2 \geq 1$). Hence,

$$\|\nabla\widehat{f}(X_\star)\| \leq 4M^{7/2}\varepsilon + (2M^{3/2} + \sigma_{\mathbf{u}}M^{5/2} + \sqrt{2(1 + \sigma_{\mathbf{u}}^2 M^4)})\varepsilon \leq 9\sigma_{\mathbf{u}}M^4\varepsilon.$$

Lastly, by Lemma C.5,

$$\left\|X_\star - \widehat{X}\right\|_{\mathsf{HS}} = \left\|(K_\star - \widehat{K})\Sigma_{\mathbf{x},0}^{1/2}\right\|_{\mathsf{HS}}\ \leq\ 9\sigma_{\mathbf{u}}M^4\varepsilon\frac{1}{\sigma_{\min}(R)}.$$

The precise statement follows by applying our assumption that $\sigma_{\min}(R) > 1$. □

C.4 Proof of Operator Norm P-Perturbation: Proposition C.3

We begin the proof with the following lemma, which ensures that the Lyapunov function of a stable matrix A_1 is also a Lyapunov function for a sufficiently nearby matrix A_2:

Lemma C.6. *Let A_1, A_2 be two matrices with A_1 stable. Set $P_1 = \mathsf{dlyap}(A_1, \Sigma)$, where $\Sigma \succeq I$. Fix an $\alpha \in (0, 1)$, and suppose that $\|A_1 - A_2\|_{\mathrm{op}}^2 \leq \frac{\alpha^2}{16\|P_1\|_{\mathrm{op}}^3}$. Then, $A_2^{\mathsf{H}}P_1A_2 \preceq P_1(1 - \frac{1-\alpha}{\|P_1\|_{\mathrm{op}}})$, and iterating,*

$$(A_2^{\mathsf{H}})^j P_1 A_2^j \preceq P_1(1 - \frac{1-\alpha}{\|P_1\|_{\mathrm{op}}})^j, \quad \forall j \geq 0.$$

Proof of Lemma C.6. From Lemma D.9, $A_1^{\mathsf{H}} P_1 A_1 \preceq P_1(1 - \|P_1\|_{\mathrm{op}}^{-1})$. Set $\Delta = A_2 - A_1$. For any $\tau > 0$, invoking Lemma D.3,

$$\begin{aligned}
A_2^{\mathsf{H}} P_1 A_2 &\preceq (A_1 + \Delta)^{\mathsf{H}} P_1 (A_1 + \Delta) \\
&= A_1^{\mathsf{H}} P_1 A_1 + \Delta^{\mathsf{H}} P_1 \Delta + \Delta^{\mathsf{H}} P_1 A_1 + A_1^{\mathsf{H}} P_1 \Delta \\
&\preceq (1+\tau) A_1^{\mathsf{H}} P_1 A_1 + \left(1 + \frac{1}{\tau}\right) \Delta^{\mathsf{H}} P_1 \Delta \\
&\preceq (1+\tau) P_1 (1 - \|P_1\|_{\mathrm{op}}^{-1}) + \left(1 + \frac{1}{\tau}\right) \|\Delta\|_{\mathrm{op}}^2 \|P_1\|_{\mathrm{op}} \\
&\preceq P_1 \left\{ (1 - \|P_1\|_{\mathrm{op}}^{-1})(1+\tau) + \left(1 + \frac{1}{\tau}\right) \|\Delta\|_{\mathrm{op}}^2 \|P_1\|_{\mathrm{op}} \right\},
\end{aligned}$$

where the last line follows from the fact that $P_1 \succeq I$ since $Q \succeq I$. Using our assumption that $\|\Delta\|_{\mathrm{op}}^2 \le \frac{\alpha^2}{16\|P_1\|_{\mathrm{op}}^3}$ and optimizing over τ, a short calculation shows that the expression between brackets above is bounded by $(1 - (1-\alpha)/\|P_1\|_{\mathrm{op}})$. ☐

The next proposition provides a perturbation bound for the dlyap operator:

Proposition C.7. *Let A_1 be a linear operator, $\Sigma \succeq I$, and $P_1 = \mathsf{dlyap}(A_1, \Sigma)$. Define $\|\cdot\|_\circ$ to be any spectral norm* (i.e. $\circ \in \{\mathrm{op}, \mathsf{HS}, \mathsf{tr}\}$). *Then, for any $\alpha \in [0,1)$, linear operator A_2 with $\|A_1 - A_2\|_{\mathrm{op}}^2 \le \frac{\alpha^2}{16\|P_1\|_{\mathrm{op}}^3}$, and any symmetric Σ_0, $\mathsf{dlyap}(A_2, \Sigma_0)$ is a bounded operator, and*

$$\|\mathsf{dlyap}(A_2, \Sigma_0) - \mathsf{dlyap}(A_1, \Sigma_0)\|_\circ \le 2\mathcal{C}_\circ \|A_1 - A_2\|_{\mathrm{op}} \|P_1\|_{\mathrm{op}}^{7/2} (1-\alpha)^{-2},$$

where $\mathcal{C}_\circ = \|P_1^{-1/2} \Sigma_0 P_1^{-1/2}\|_\circ$.

Proof of Proposition C.7. Set $\Delta = A_1 - A_2$, and define the terms $E_n := P_1^{1/2}(A_1^n - A_2^n)$, $\gamma := 1 - \frac{1-\alpha}{\|P_1\|_{\mathrm{op}}}$. Since A_1, A_2 satisfy the necessary closeness conditions, by Lemmas C.6 and D.9, we have that,

$$\max\{\|P_1^{1/2} A_1^n\|_{\mathrm{op}}, \|P_2^{1/2} A_1^n\|_{\mathrm{op}}\} \le \sqrt{\|P_1\|_{\mathrm{op}} \gamma^n}. \tag{C.9}$$

Next, by closed form expression for dlyap,

$$\begin{aligned}
\mathsf{dlyap}(A_2, \Sigma_0) - \mathsf{dlyap}(A_1, \Sigma_0) &= \sum_{n \ge 0} (A_2^n)^{\mathsf{H}} \Sigma_0 A_2^n - (A_1^n)^{\mathsf{H}} \Sigma_0 A_1^n \\
&= \sum_{n \ge 1} (A_2^n - A_1^n)^{\mathsf{H}} \Sigma_0 A_2^n + A_1^n \Sigma_0 (A_2^n - A_1^n) \\
&= -\sum_{n \ge 1} E_n^{\mathsf{H}} P_1^{-1/2} \Sigma_0 A_2^n + (A_1^n)^{\mathsf{H}} \Sigma_0 P_1^{-1/2} E_n,
\end{aligned}$$

and thus,

$$\begin{aligned}
\|\mathsf{dlyap}(A_2, \Sigma_0) - \mathsf{dlyap}(A_1, \Sigma_0)\|_\circ &\le \sum_{n \ge 1} \|E_n\|_{\mathrm{op}} \left(\|P_1^{-1/2} \Sigma_0 A_2^n\|_\circ + \|(A_1^n)^{\mathsf{H}} \Sigma_0 P_1^{-1/2}\|_\circ \right) \\
&\le \sum_{n \ge 1} \|E_n\|_{\mathrm{op}} \|P_1^{-1/2} \Sigma_0 P_1^{-1/2}\|_\circ \left(\|P_1^{1/2} A_2^n\|_{\mathrm{op}} + \|P_1^{1/2} A_1^n\|_{\mathrm{op}} \right) \\
&\le \underbrace{\|P_1^{-1/2} \Sigma_0 P_1^{-1/2}\|_\circ}_{:= \mathcal{C}_\circ} \sum_{n \ge 1} \|E_n\|_{\mathrm{op}} \left(\|P_1^{1/2} A_2^n\| + \|P_1^{1/2} A_1^n\| \right) \\
&\le 2\mathcal{C}_\circ \sum_{n \ge 1} \|E_n\|_{\mathrm{op}} \sqrt{\|P_1\|_{\mathrm{op}} \gamma^n}. && \text{(Eq. (C.9))}
\end{aligned}$$

Next, we bound $\|E_n\|_{\mathrm{op}}$. Using the identity, $A_1^n - A_2^n = \sum_{i=0}^n A_1^i \Delta A_2^{n-i-1}$,

$$\begin{aligned}\|E_n\|_{\mathrm{op}} &\le \sum_{i=0}^n \|P_1^{1/2}A_1^i\|_{\mathrm{op}}\|\Delta\|_{\mathrm{op}}\|A_2^{n-i-1}\|_{\mathrm{op}} \\ &\le \sum_{i=0}^n \|P_1^{1/2}A_1^i\|_{\mathrm{op}}\|\Delta\|_{\mathrm{op}}\|P_1^{1/2}A_2^{n-i-1}\|_{\mathrm{op}}, && (P_1 \succeq I) \\ &\le \|\Delta\|_{\mathrm{op}}\|P_1\|_{\mathrm{op}}\sum_{i=0}^n \sqrt{\gamma^{n-1}}. && \text{(Eq. (C.9))}\end{aligned}$$

Hence, combining the above,

$$\begin{aligned}\|\mathsf{dlyap}\,(A_2,\ \Sigma_0) - \mathsf{dlyap}\,(A_1,\ \Sigma_0)\,\|_\circ &\le 2\mathcal{C}_\circ\,\|\Delta\|_{\mathrm{op}}\,\|P_1\|_{\mathrm{op}}^{3/2}\sum_{n\ge1} n\gamma^{n-1/2} \\ &\le 2\mathcal{C}_\circ\,\|\Delta\|_{\mathrm{op}}\,\|P_1\|_{\mathrm{op}}^{3/2}\sum_{n\ge1} n\gamma^{n-1} && (\gamma \le 1) \\ &= 2\mathcal{C}_\circ\,\|\Delta\|_{\mathrm{op}}\,\|P_1\|_{\mathrm{op}}^{3/2}\,(1-\gamma)^{-2}.\end{aligned}$$

Substituting in $\gamma = (1 - \frac{1-\alpha}{\|P_1\|_{\mathrm{op}}})$, the above becomes $2\mathcal{C}_\circ\,\|\Delta\|_{\mathrm{op}}\,\|P_1\|_{\mathrm{op}}^{7/2}\,(1-\alpha)^{-2}$, concluding the proof. □

We may now conclude the proof of Proposition C.3

Proof of Proposition C.3. We prove the each part of the proposition individually.

Part 1 Define $\Sigma_1 := K_1^\top R K_1 + Q$. Then,

$$\begin{aligned}P_2 &\preceq P_\infty(K_1; A_2, B_2) = \mathsf{dlyap}\,(A_2 + B_2K_1,\ \Sigma_1) \\ &\preceq \mathsf{dlyap}\,(A_1 + B_1K_1,\ \Sigma_1) + I\cdot\|\mathsf{dlyap}\,(A_1 + B_1K_1,\ \Sigma_1) - \mathsf{dlyap}\,(A_2 + B_2K_1,\ \Sigma_1)\,\|_{\mathrm{op}} \\ &= P_1 + I\cdot\|\mathsf{dlyap}\,(A_1 + B_1K_1,\ \Sigma_1) - \mathsf{dlyap}\,(A_2 + B_2K_1,\ \Sigma_1)\,\|_{\mathrm{op}} \qquad \text{(C.10)}\end{aligned}$$

If $\varepsilon_0^2 := \|A_1+B_1K_1-(A_2+B_2K_1)\|_{\mathrm{op}}^2 \le \frac{1}{16\|P_1\|_{\mathrm{op}}^3}$, then, invoking Proposition C.7 with $\Sigma_0 \leftarrow \Sigma_1$, $\|\cdot\|_\circ \leftarrow \|\cdot\|_{\mathrm{op}}$, $\alpha \leftarrow 1/2$, and noting how $\mathcal{C}_\circ \le 1$, we can conclude that,

$$\|\mathsf{dlyap}\,(A_1 + B_1K_1,\ \Sigma_1) - \mathsf{dlyap}\,(A_2 + B_2K_1,\ \Sigma_1)\,\|_{\mathrm{op}} \le 8\varepsilon_0\,\|P_1\|_{\mathrm{op}}^{7/2}.$$

Observe that if $\varepsilon_0 \le \eta/(8\,\|P_1\|_{\mathrm{op}}^{5/2})$ for some $\eta \in (0,1]$, then (since $\|P_1\|_{\mathrm{op}} \ge 1$ by Lemma D.7) it holds that $\varepsilon_0^2 \le \frac{1}{64\|P_1\|_{\mathrm{op}}^3}$. Plugging into the inequality above, we get our desired result,

$$P_2 \preceq P_1 + \eta\,\|P_1\|_{\mathrm{op}}\cdot I.$$

Therefore, to finish the proof of Part 1, we only need to verify that, under the assumptions of the proposition, $\varepsilon_0 \le \eta/(8\,\|P_1\|_{\mathrm{op}}^{5/2})$. By the definition of ε_0 and Lemma D.7,

$$\varepsilon_0 \le \|A_1 - A_2\|_{\mathrm{op}} + \|K_1\|_{\mathrm{op}}\,\|B_1 - B_2\|_{\mathrm{op}} \le (1 + \|K_1\|_{\mathrm{op}})\varepsilon_{\mathrm{op}} \le 2\,\|P_1\|_{\mathrm{op}}^{1/2}\,\varepsilon_{\mathrm{op}}.$$

Since, $\varepsilon_{\mathrm{op}} \le \eta/(16\,\|P_1\|_{\mathrm{op}}^3)$, this calculation above proves that $\varepsilon_0 \le \eta/(8\,\|P_1\|_{\mathrm{op}}^{5/2})$.

Part 2 Fix a parameter $\eta_0 \in (0,1)$ to be chosen, and let η be as in the theorem statement. Switching the roles of the indices $i=1$ and $i=2$ in the first part of the proposition, we find that if $\varepsilon_{\mathrm{op}} \le \eta_0/(16\,\|P_2\|_{\mathrm{op}}^3)$, we can establish the following PSD inequalities:

$$P_1 \preceq P_\infty(K_2; A_1, B_1) \preceq P_2 + \eta_0 \|P_2\|_{\mathrm{op}} I. \tag{C.11}$$

If in addition $\varepsilon_{\mathrm{op}} \le \eta/(16\,\|P_2\|_{\mathrm{op}}^3)$, then we have $\|P_2\|_{\mathrm{op}} \le (1+\eta)\|P_1\|_{\mathrm{op}}$. Therefore, if $\varepsilon_{\mathrm{op}} \le \eta_0/(16(1+\eta)^3\,\|P_1\|_{\mathrm{op}}^3)$ (replacing P_2 in the denominator by P_1), then $P_1 \preceq P_2 + \eta_0(1+\eta)\|P_1\|_{\mathrm{op}} I$.

Selecting $\eta_0 \leftarrow \eta/(1+\eta)$, we have that if $\varepsilon_{\mathrm{op}} \le \eta/(16(1+\eta)^4\,\|P_2\|_{\mathrm{op}}^3)$, $P_1 \preceq P_2 + \eta\|P_1\|_{\mathrm{op}} I$. Hence, combining with Part 1, we have that $\|P_1 - P_2\|_{\mathrm{op}} \le \eta\|P_1\|_{\mathrm{op}}$. □

C.5 Proof of Weighted P-Perturbation: Proposition C.4

Our argument is based on the self-bounding ODE method introduced by Simchowitz and Foster [2020], where the perturbation bound is derived from considering an interpolating curve $A(t), B(t)$ between the ground-truth instances $(A_\star, B_\star)$ and estimated instance $(\widehat{A}, \widehat{B})$.

Definition C.1 (Interpolating Curves)**.** Given estimates $(\widehat{A}, \widehat{B})$ of $(A_\star, B_\star)$, for $t \in [0,1]$ we define,

$$A(t) := A_\star + t(\widehat{A} - A_\star), \quad B(t) := B_\star + t(\widehat{B} - B_\star).$$

In addition, we define the optimal controller $K(t)$, closed-loop matrix $A_{\mathrm{cl}}(t)$, and value function $P(t)$ as

$$K(t) := K_\infty(A(t), B(t)), \quad A_{\mathrm{cl}}(t) := A(t) + B(t)K(t), \quad P(t) := P_\infty(A(t), B(t)).$$

Finally, define the following error term in the closed loop matrix

$$\Delta_{A_{\mathrm{cl}}}(t) := \widehat{A} - A_\star + (\widehat{B} - B_\star)K(t).$$

At the core of the technique is verifying that the instances along the curve $(A(t), B(t))$ remain stabilizable, so that the operators $K(t), P(t)$ are well defined. The following guarantee, established in Appendix C.5.3, ensures that this is the case.

Lemma C.8. *Assume that $(\widehat{A}, \widehat{B})$ satisfy Condition 2.1. Then:*

1. *The instances $(A(t), B(t))$ are stabilizable along $t \in [0,1]$, and $\sup_{t\in[0,1]} \|P(t)\|_{\mathrm{op}} \le 1.1\,\|P_\star\|_{\mathrm{op}}$*
2. *$\sup_{u,t\in[0,1]} \|P_\infty(K(t); A(u\cdot t), B(u\cdot t))\|_{\mathrm{op}} \le 1.2\,\|P_\star\|_{\mathrm{op}}$*

Since instances $(A(t), B(t))$ are stabilizable along $t \in [0,1]$, Lemma D.18 ensures that $P'(t)$ is *Frechet differentiable* in the operator norm, and its derivative is

$$\frac{\mathrm{d}}{\mathrm{d}t} P(t) = \mathsf{dlyap}\left(A_{\mathrm{cl}}(t),\ E(t)\right), \tag{C.12}$$

$$\text{where } E(t) := A_{\mathrm{cl}}(t)^{\mathsf{H}} P(t) \Delta_{A_{\mathrm{cl}}}(t) + \Delta_{A_{\mathrm{cl}}}(t)^{\mathsf{H}} P(t) A_{\mathrm{cl}}(t). \tag{C.13}$$

The formal definition of the Frechet derivative is deferred to Definition D.2; in this section, we shall only use it as a condition to call relevant lemmas. The heart of proposition is based off the following observation which follows from part (c) of Lemma D.20:

$$\left\|\Sigma_{\mathbf{x},0}^{1/2}(P_\star - \widehat{P})\Sigma_{\mathbf{x},0}^{1/2}\right\|_{\mathsf{HS}} = \left\|\Sigma_{\mathbf{x},0}^{1/2}(P(0) - P(1))\Sigma_{\mathbf{x},0}^{1/2}\right\|_{\mathsf{HS}} \le \sup_{t\in[0,1]} \left\|\Sigma_{\mathbf{x},0}^{1/2} P'(t) \Sigma_{\mathbf{x},0}^{1/2}\right\|_{\mathsf{HS}}.$$

To prove the main result, it remains to show the following bound on the derivative:

$$\forall t \in [0,1], \quad \left\|\Sigma_{\mathbf{x},0}^{1/2} P'(t) \Sigma_{\mathbf{x},0}^{1/2}\right\|_{\mathsf{HS}} \le \mathcal{C}_P \cdot \varepsilon_P \cdot \sqrt{\log_+(\kappa_P)} \cdot \phi(\kappa_P)^{\alpha_{\mathrm{op}}}. \tag{C.14}$$

We verify that Eq. (C.14) holds in Appendix C.5.1.

A key step along the way is the following "change of system" lemma, which allows us to bound the norm of the error in a covariance induced by one closed-loop system A_1+B_1K by that induced by another, A_2+B_2K. This proposition is, in some sense, the mirror image of the change-of-covariance theorems in Appendix B. In those results, we keep the system fixed but change the controller. Here, we keep the controller fixed, but change the system.

Proposition C.9. *Let (A_1, B_1) and (A_2, B_2) be two systems, K a controller, and define for $u \in [0,1]$:*

$$\bar{A}(u) := A_1 + u(A_2 - A_1) \quad \bar{B}(u) := B_1 + u(B_2 - B_2), \quad \bar{A}_{\mathrm{cl}}(u) := \bar{A}(u) + \bar{B}(u)K.$$

Assume that K stabilizes all instances $(\bar{A}(u), \bar{B}(u))$, so that

$$\bar{M} := \max_{u\in[0,1]} \left\|\bar{P}(u)\right\|_{\mathrm{op}} < \infty, \qquad \text{where } \bar{P}(u) := \mathsf{dlyap}\left(\bar{A}_{\mathrm{cl}}(u),\ Q + K^{\mathsf{H}}RK\right).$$

Finally, define $\bar{\Delta}_{A_{\mathrm{cl}}} := \frac{d}{du}\bar{A}_{\mathrm{cl}}(u) = (A_2 - A_1) + (B_2 - B_1)K$, let $\Sigma \succeq 0$ be a trace class operator, and define the error terms:

$$\varepsilon^2_{\mathrm{sys},i} := \mathrm{tr}\left[\bar{\Delta}^{\mathsf{H}}_{A_{\mathrm{cl}}}\bar{\Delta}_{A_{\mathrm{cl}}}\mathsf{dlyap}\left((A_i + B_iK)^{\mathsf{H}},\ \Sigma\right)\right], \quad i \in \{1, 2\}.$$

Then, setting the operator $[X, Y]_{\mathrm{algn}} := \max\{\|XWY\|_{\mathsf{tr}} : \|W\|_{\mathrm{op}} = 1\}$ the following bound holds,

$$\varepsilon^2_{\mathrm{sys},2} \le \psi_{\mathrm{diff}}(\varepsilon_{\mathrm{sys},1}) \cdot \varepsilon^2_{\mathrm{sys},1}$$
$$\text{where } \psi_{\mathrm{diff}}(\varepsilon) := 2\exp\left(\frac{3}{2}\bar{M}\left\|\bar{\Delta}_{A_{\mathrm{cl}}}\right\|_{\mathrm{op}}\sqrt{\log_+\left(\frac{e\cdot\left[\bar{\Delta}^{\mathsf{H}}_{A_{\mathrm{cl}}}\bar{\Delta}_{A_{\mathrm{cl}}}, \Sigma\right]_{\mathrm{algn}}\bar{M}^3}{\varepsilon^2}\right)}\right).$$

Moreover, these relationships hold for any $\varepsilon \ge \varepsilon_{\mathrm{sys},1}$.

The proof relies on a careful application of the self-bounding ODE method. The full proof is given in Appendix C.5.2.

C.5.1 Establishing Eq. (C.14)

Proof. Our proof proceeds in multiple steps. First, we majorize the weighted derivative $\Sigma^{1/2}_{\mathbf{x},0}P'(t)\Sigma^{1/2}_{\mathbf{x},0}$ in terms of two PSD operators $Y_1(t), Y_2(t)$; a careful analysis shows it suffices to bound the operator norm of $Y_1(t)$ and the trace of $Y_2(t)$. The term $Y_1(t)$ is straightforward to control; the term $Y_2(t)$ requires appeal to the change-of-system error bound in Proposition C.9. The proof concludes with an application of one of our change-of-covariance bounds (Lemma B.1), and some further simplifications.

Majorization by $Y_1(t), Y_2(t)$ Define for all $t \in [0,1]$ and $\alpha > 0$ the quantities:

$$E_1(t) := A_{\mathrm{cl}}(t)^{\mathsf{H}}P(t)A_{\mathrm{cl}}(t), \quad E_2(t) := \Delta_{A_{\mathrm{cl}}}(t)^{\mathsf{H}}P(t)\Delta_{A_{\mathrm{cl}}}(t), \quad E_{[\alpha]}(t) := \frac{\alpha}{2}E_1(t) + \frac{1}{2\alpha}E_2(t).$$

By the AM-GM inequality in Lemma D.3, $E(t) \preceq E_{[\alpha]}(t)$ for all $t \in [0,1]$ and $\alpha > 0$ (recall that $E(t)$ is defined in Eq. (C.13)). Since the solution of the Lyapunov equation preserves PSD order (Lemma A.1),

$$\underbrace{-\Sigma^{1/2}_{\mathbf{x},0}\mathsf{dlyap}\left(A_{\mathrm{cl}}(t),\ E_{[\alpha]}(t)\right)\Sigma^{1/2}_{\mathbf{x},0}}_{:=-Y_{[\alpha]}(t)} \preceq \Sigma^{1/2}_{\mathbf{x},0}P'(t)\Sigma^{1/2}_{\mathbf{x},0} \preceq \underbrace{\Sigma^{1/2}_{\mathbf{x},0}\mathsf{dlyap}\left(A_{\mathrm{cl}}(t),\ E_{[\alpha]}(t)\right)\Sigma^{1/2}_{\mathbf{x},0}}_{:=Y_{[\alpha]}(t)}.$$

Now, observe that since $\mathsf{dlyap}(\cdot,\ \cdot)$ is linear in its second argument, we can express

$$Y_{[\alpha]}(t) = \frac{\alpha}{2}Y_1(t) + \frac{1}{2\alpha}Y_2(t), \quad \text{where}$$
$$Y_1(t) := \Sigma_{\mathbf{x},0}^{1/2}\mathsf{dlyap}\left(A_{\mathrm{cl}}(t),\ A_{\mathrm{cl}}(t)^{\mathsf{H}}P(t)A_{\mathrm{cl}}(t)\right)\Sigma_{\mathbf{x},0}^{1/2}$$
$$Y_2(t) := \Sigma_{\mathbf{x},0}^{1/2}\mathsf{dlyap}\left(A_{\mathrm{cl}}(t),\ \Delta_{A_{\mathrm{cl}}}(t)^{\mathsf{H}}P(t)\Delta_{A_{\mathrm{cl}}}(t)\right)\Sigma_{\mathbf{x},0}^{1/2}.$$

From Lemma D.5, we show that the PSD domination of $\Sigma_{\mathbf{x},0}^{1/2}P'(t)\Sigma_{\mathbf{x},0}^{1/2}$ by the weighted linear combinations of the form $Y_{[\alpha]}(t)$ imply that

$$\left\|\Sigma_{\mathbf{x},0}^{1/2}P'(t)\Sigma_{\mathbf{x},0}^{1/2}\right\|_{\mathsf{HS}} \ \leq\ \sqrt{\|Y_1(t)\|_{\mathrm{op}}\cdot \mathrm{tr}[Y_2(t)]}. \tag{C.15}$$

We now proceed to bound the terms $\|Y_1(t)\|_{\mathrm{op}}$ and $\mathrm{tr}[Y_2(t)]$ individually. We let $M = 1.2\,\|P_\star\|_{\mathrm{op}}$, which upper bounds $\sup_{t\in[0,1]}\|P(t)\|_{\mathrm{op}}$ by Lemma C.8.

Bounding $\|Y_1(t)\|_{\mathrm{op}}$ By Lemma D.9 we have that $(A_{\mathrm{cl}}(t)^{\mathsf{H}})^jP(t)A_{\mathrm{cl}}(t)^j \preceq P(t)(1-\|P(t)\|_{\mathrm{op}}^{-1})^j$. Therefore, using our uniform upper bound on $\|P(t)\|$ from Lemma C.8, the following inequality holds for T_1:

$$\begin{aligned}
\|Y_1(t)\|_{\mathrm{op}} &= \left\|\Sigma_{\mathbf{x},0}^{1/2}\mathsf{dlyap}\left(A_{\mathrm{cl}}(t),\ A_{\mathrm{cl}}(t)^{\mathsf{H}}P(t)A_{\mathrm{cl}}(t)\right)\Sigma_{\mathbf{x},0}^{1/2}\right\|_{\mathrm{op}}\\
&= \left\|\Sigma_{\mathbf{x},0}\sum_{j=0}^{\infty}\left(A_{\mathrm{cl}}(t)^{\mathsf{H}}\right)^{j+1}P(t)A_{\mathrm{cl}}(t)^{j+1}\right\|_{\mathrm{op}}\\
&\leq\ \left\|\Sigma_{\mathbf{x},0}^{1/2}P(t)\Sigma_{\mathbf{x},0}^{1/2}\right\|_{\mathrm{op}}\sum_{j=0}^{\infty}(1-\|P(t)\|_{\mathrm{op}}^{-1})^{j+1}\\
&\leq\ \|\Sigma_{\mathbf{x},0}\|_{\mathrm{op}}\|P(t)\|_{\mathrm{op}}^2\\
&\leq\ \|\Sigma_{\mathbf{x},0}\|_{\mathrm{op}}M^2.
\end{aligned}\tag{C.16}$$

Bounding $\mathrm{tr}[Y_2(t)]$ via change of system Commuting traces and rewriting $\sum_{j=0}^{\infty}A_{\mathrm{cl}}(t)^j\Sigma_{\mathbf{x},0}\left(A_{\mathrm{cl}}(t)^{\mathsf{H}}\right)^j$ as $\mathsf{dlyap}\left(A_{\mathrm{cl}}(t)^{\mathsf{H}},\ \Sigma_{\mathbf{x},0}\right)$, we have

$$\begin{aligned}
\mathrm{tr}\left[Y_2(t)\right] &= \mathrm{tr}\left[\Sigma_{\mathbf{x},0}^{1/2}\mathsf{dlyap}\left(A_{\mathrm{cl}}(t),\ \Delta_{A_{\mathrm{cl}}}(t)^{\mathsf{H}}P(t)\Delta_{A_{\mathrm{cl}}}(t)\right)\Sigma_{\mathbf{x},0}^{1/2}\right]\\
&= \mathrm{tr}\left[\mathsf{dlyap}\left(A_{\mathrm{cl}}(t)^{\mathsf{H}},\ \Sigma_{\mathbf{x},0}\right)\Delta_{A_{\mathrm{cl}}}(t)^{\mathsf{H}}P(t)\Delta_{A_{\mathrm{cl}}}(t)\right]\\
&\leq\ M\cdot\mathrm{tr}\left[\Delta_{A_{\mathrm{cl}}}(t)\mathsf{dlyap}\left(A_{\mathrm{cl}}(t)^{\mathsf{H}},\ \Sigma_{\mathbf{x},0}\right)\Delta_{A_{\mathrm{cl}}}(t)^{\mathsf{H}}\right].
\end{aligned}\tag{C.17}$$

In the last line, we used that for any $X \succeq 0$, and operators X, A, $\mathrm{tr}\left[XA^{\mathsf{H}}YA\right] = \mathrm{tr}\left[AXA^{\mathsf{H}}Y\right] \leq \|Y\|_{\mathrm{op}}\,\mathrm{tr}\left[AXA^{\mathsf{H}}\right]$.

Note however that our estimation guarantees hold under the true system $A_\star, B_\star$, whereas the above bound holds under a closed loop system involving the matrices $A(t), B(t)$. To this end, we invoke the change of system guarantee, Proposition C.9. We instantiate Proposition C.9 with the following substitutions:

- Take $K \leftarrow K(t)$, $(A_1, B_1) \leftarrow (A(t), B(t))$, and $(A_2, A_2) \leftarrow (A_\star, B_\star)$.
- By Lemma C.8, $\bar{M}$ can be upper bounded by $M := 1.2\|P_\star\|_{\mathrm{op}} := 1.2M_\star$.
- $\bar{\Delta}_{A_{\mathrm{cl}}} = A_\star - A(t) + (B_\star - B(t))K(t)) = -t\Delta_{A_{\mathrm{cl}}}(t)$. Hence, bounding $\sup_{t\in[0,1]}\|K(t)\|_{\mathrm{op}} \leq \sqrt{M}$:

$$\begin{aligned}
\|\bar{\Delta}_{A_{\mathrm{cl}}}\|_{\mathrm{op}} &\leq \|\Delta_{A_{\mathrm{cl}}}(t)\|_{\mathrm{op}} \leq \|A_\star - A(t)\|_{\mathrm{op}} + \|B_\star - B(t)\|_{\mathrm{op}}\|K(t)\|_{\mathrm{op}}\\
&\leq \varepsilon_{\mathsf{op}}(1+\|K(t)\|_{\mathrm{op}}) \leq \varepsilon_{\mathsf{op}}(1+\sqrt{M}) \leq 2\sqrt{M}\varepsilon_{\mathsf{op}} \leq 2.4\sqrt{M_{P_\star}}.
\end{aligned}\tag{C.18}$$

- Define the aligned error,

$$\varepsilon_{\mathrm{algn}}^2 := \max_{t\in[0,1]} \left[\Delta_{A_{\mathrm{cl}}}(t)^{\mathsf{H}}\Delta_{A_{\mathrm{cl}}}(t), \Sigma_{\mathbf{x},0}\right]_{\mathrm{algn}}. \tag{C.19}$$

- With these substitutions, we can take

$$\psi(\varepsilon) := 2\exp\left(3\cdot 1.2^{3/2}\cdot M_{P_\star}^{3/2}\varepsilon_{\mathsf{op}}\sqrt{\log_+\left(1.2^3 e M_{P_\star}^3\cdot\frac{\varepsilon_{\mathrm{algn}}^2}{\varepsilon^2}\right)}\right). \tag{C.20}$$

Applying these substitutions, Proposition C.9 entails

$$\mathrm{tr}\left[Y_2(t)\right] \le M\mathrm{tr}\left[\Delta_{A_{\mathrm{cl}}}(t)^{\mathsf{H}}\Delta_{A_{\mathrm{cl}}}(t)\mathsf{dlyap}\left((A(t)+B(t)K(t))^{\mathsf{H}},\ \Sigma_{\mathbf{x},0}\right)\right] \le M\psi(\varepsilon)\varepsilon, \quad \forall\varepsilon\ge T_3 \tag{C.21}$$

$$\text{where } T_3 := \mathrm{tr}\left[\Delta_{A_{\mathrm{cl}}}(t)^{\mathsf{H}}\Delta_{A_{\mathrm{cl}}}(t)\mathsf{dlyap}\left((A_\star+B_\star K(t))^{\mathsf{H}},\ \Sigma_{\mathbf{x},0}\right)\right]. \tag{C.22}$$

Bounding T_3 via change of controller Focusing on the remaining trace term T_3, by Lemma B.1 we can switch the controller $K(t)$ inside the Lyapunov operator to K_0 and pay a multiplicative constant,

$$\begin{aligned} T_3 &= \mathrm{tr}\left[\Delta_{A_{\mathrm{cl}}}(t)^{\mathsf{H}}\Delta_{A_{\mathrm{cl}}}(t)\mathsf{dlyap}\left((A_\star+B_\star K(t))^{\mathsf{H}},\ \Sigma_{\mathbf{x},0}\right)\right] \\ &\le\ \mathcal{C}_K\cdot\mathrm{tr}\left[\Delta_{A_{\mathrm{cl}}}(t)^{\mathsf{H}}\Delta_{A_{\mathrm{cl}}}(t)\mathsf{dlyap}\left((A_\star+B_\star K_0)^{\mathsf{H}},\ \Sigma_{\mathbf{x},0}\right)\right], \end{aligned} \tag{C.23}$$

where, for $P_0 = P_\infty(K_0; A_\star, B_\star)$ and $M_{P_0} := \|P_0\|_{\mathrm{op}}$, we define

$$\mathcal{C}_K := 2\left(1+\frac{64\max_{t\in[0,1]}\|K_0-K(t)\|_{\mathrm{op}}^2}{\sigma_{\mathbf{u}}^2}\|\Sigma_{\mathbf{x},0}\|_{\mathrm{op}} M_{P_0}^3\log(2M_{P_0})^2\right).$$

Next, we recall that $\Sigma_{\mathbf{x},0} = \mathsf{dlyap}\left((A_\star+B_\star K_0)^{\mathsf{H}},\ \Sigma_0\right)$ where $\Sigma_0 = B_\star B_\star\sigma_{\mathbf{u}}^2+\Sigma_{\mathbf{w}}$ is equal to the steady-state covariance induced by the initial controller K_0. By Lemma D.10, we can rewrite the solution to Lyapunov equation as,

$$\mathsf{dlyap}\left((A_\star+B_\star K_0)^{\mathsf{H}},\ \Sigma_{\mathbf{x},0}\right) = \mathsf{dlyap}_{(1)}\left((A_\star+B_\star K_0)^{\mathsf{H}},\ \Sigma_0\right).$$

By applying Lemma D.12, and recalling that $M_{P_0} = \|P_0\|_{\mathrm{op}}$,$[X,Y]_{\mathrm{algn}} := \max\{\|XWY\|_{\mathsf{tr}} : \|W\|_{\mathrm{op}}=1\}$, we can upper bound $\mathrm{tr}\left[\Delta_{A_{\mathrm{cl}}}(t)^{\mathsf{H}}\Delta_{A_{\mathrm{cl}}}(t)\mathsf{dlyap}_{(1)}\left((A_\star+B_\star K_0)^{\mathsf{H}},\ \Sigma_0\right)\right]$ as follows,

$$\begin{aligned} &\le n\cdot\mathrm{tr}\left[\Delta_{A_{\mathrm{cl}}}(t)^{\mathsf{H}}\Delta_{A_{\mathrm{cl}}}(t)\mathsf{dlyap}\left((A_\star+B_\star K_0)^{\mathsf{H}},\ \Sigma_0\right)\right]+(n+1)\left[\Delta_{A_{\mathrm{cl}}}(t)^{\mathsf{H}}\Delta_{A_{\mathrm{cl}}}(t),\Sigma_0\right]_{\mathrm{algn}}\|P_0\|_{\mathrm{op}}^3\exp(-\|P_0\|_{\mathrm{op}}^{-1}n) \\ &= n\cdot\mathrm{tr}\left[\Delta_{A_{\mathrm{cl}}}(t)^{\mathsf{H}}\Delta_{A_{\mathrm{cl}}}(t)\Sigma_{\mathbf{x},0}\right]+(n+1)\left[\Delta_{A_{\mathrm{cl}}}(t)^{\mathsf{H}}\Delta_{A_{\mathrm{cl}}}(t),\Sigma_0\right]_{\mathrm{algn}} M_{P_0}^3\exp(-M_{P_0}^{-1}n) \\ &\le n\cdot\mathrm{tr}\left[\Delta_{A_{\mathrm{cl}}}(t)^{\mathsf{H}}\Delta_{A_{\mathrm{cl}}}(t)\Sigma_{\mathbf{x},0}\right]+(n+1)\varepsilon_{\mathrm{algn}}^2 M_{P_0}^3\exp(-M_{P_0}^{-1}n). \end{aligned}$$

In the last inequality, we used that by Lemma D.6, $\left[\Delta_{A_{\mathrm{cl}}}(t)^{\mathsf{H}}\Delta_{A_{\mathrm{cl}}}(t),\Sigma_{\mathbf{x},0}\right]_{\mathrm{algn}} \le \varepsilon_{\mathrm{algn}}^2$ where $\varepsilon_{\mathrm{algn}}^2$ is defined in Eq. (C.19). Next, by AM-GM and the fact that $\mathrm{tr}\left[XY\right] \le \mathrm{tr}\left[X'Y\right]$ for $0 \preceq X \preceq X'$ and $Y \succeq 0$,

$$\begin{aligned} \mathrm{tr}\left[\Delta_{A_{\mathrm{cl}}}(t)^{\mathsf{H}}\Delta_{A_{\mathrm{cl}}}(t)\Sigma_{\mathbf{x},0}\right] &= \mathrm{tr}\left[\left(\widehat{A}-A_\star+K(t)(\widehat{B}-B_\star)\right)^{\mathsf{H}}\left(\widehat{A}-A_\star+K(t)(\widehat{B}-B_\star)\right)\Sigma_{\mathbf{x},0}\right] \\ &\le 2\mathrm{tr}\left[\left((\widehat{A}-A_\star)^{\mathsf{H}}(\widehat{A}-A_\star)+K(t)^{\mathsf{H}}(\widehat{B}-B_\star)^{\mathsf{H}}(\widehat{B}-B_\star)K(t)\right)\Sigma_{\mathbf{x},0}\right] \\ &= 2\|(\widehat{A}-A_\star)\Sigma_{\mathbf{x},0}^{1/2}\|_{\mathsf{HS}}^2+2\|(\widehat{B}-B_\star)K(t)\Sigma_{\mathbf{x},0}^{1/2}\|_{\mathsf{HS}}^2 \\ &\le 2\|(\widehat{A}-A_\star)\Sigma_{\mathbf{x},0}^{1/2}\|_{\mathsf{HS}}^2+2\max_{t\in[0,1]}\|K(t)\|_{\mathrm{op}}^2\|(\widehat{B}-B_\star)\Sigma_{\mathbf{x},0}^{1/2}\|_{\mathsf{HS}}^2 \\ &\le 2\|(\widehat{A}-A_\star)\Sigma_{\mathbf{x},0}^{1/2}\|_{\mathsf{HS}}^2+2M\|(\widehat{B}-B_\star)\Sigma_{\mathbf{x},0}^{1/2}\|_{\mathsf{HS}}^2 \\ &\le 2\|(\widehat{A}-A_\star)\Sigma_{\mathbf{x},0}^{1/2}\|_{\mathsf{HS}}^2+2.4M_{P_\star}\|\widehat{B}-B_\star\|_{\mathsf{HS}}^2\|\Sigma_{\mathbf{x},0}\|_{\mathrm{op}} := \varepsilon_P^2, \end{aligned} \tag{C.24}$$

where above, we used the fact that $\|K(t)\|_{\mathrm{op}}^2 \le \|P(t)\|_{\mathrm{op}}$ by Lemma D.7, which by Appendix C.5.3 is at most $M := 1.2 M_{P_\star}$ for all $t \in [0,1]$. Hence,

$$\mathrm{tr}\left[\Delta_{A_{\mathrm{cl}}}(t)^{\mathsf{H}}\Delta_{A_{\mathrm{cl}}}(t)\mathsf{dlyap}_{(1)}\left((A_\star + B_\star K_0)^{\mathsf{H}},\ \Sigma_0\right)\right] \le n\varepsilon_P^2 + (n+1)\varepsilon_{\mathrm{algn}}^2 M_{P_0}^3 \exp(-M_{P_0}^{-1} n).$$

To optimize over n, select $n = \lceil M_{P_0} \log \frac{M_{P_0}^3 \varepsilon_{\mathrm{algn}}^2}{\varepsilon_P^2} \rceil$. Then, going back to Eq. (C.23) and putting things together, we get that:

$$\begin{aligned} T_3 &\le \mathcal{C}_K \mathrm{tr}\left[\Delta_{A_{\mathrm{cl}}}(t)^{\mathsf{H}}\Delta_{A_{\mathrm{cl}}}(t)\mathsf{dlyap}_{(1)}\left((A_\star + B_\star K_0)^{\mathsf{H}},\ \Sigma_0\right)\right] \\ &\le \mathcal{C}_K(2n+1)\varepsilon_P^2 \\ &\le 3\mathcal{C}_K M_{P_0}\varepsilon_P^2 \log_+\left(\frac{eM_{P_0}^3\varepsilon_{\mathrm{algn}}^2}{\varepsilon_P^2}\right) := \bar{T}_3. \end{aligned} \tag{C.25}$$

Concluding the proof Combining Eqs. (C.21), (C.22) and (C.25), and $M = 1.2M_{P_\star}$ gives

$$\max_{t\in[0,1]} \mathrm{tr}[Y_2(t)] \le M\psi(\bar{T}_3)\bar{T}_3 = 3.6 \cdot \mathcal{C}_K M_{P_\star} M_{P_0} \varepsilon_P^2 \log_+\left(\frac{eM_{P_0}^3\varepsilon_{\mathrm{algn}}^2}{\varepsilon_P^2}\right)\psi(\bar{T}_3). \tag{C.26}$$

Now, recall the definition from Eq. (C.20) that

$$\psi(\varepsilon) := 2\exp\left(3 \cdot 1.2^{3/2} \cdot M_{P_\star}^{3/2}\varepsilon_{\mathsf{op}}\sqrt{\log_+\left(1.2^3 e M_{P_\star}^3 \cdot \frac{\varepsilon_{\mathrm{algn}}^2}{\varepsilon^2}\right)}\right).$$

Since this quantity is decreasing in ε, and since $\bar{T}_3 \ge 3M_{P_0}\mathcal{C}_K\varepsilon_P^2 \ge 6M_{P_0}\varepsilon_P^2$ (recall $\mathcal{C}_K \ge 2$),

$$\begin{aligned} \psi(\bar{T}_3) &\le \psi(3M_{P_0}\mathcal{C}_K\varepsilon_P^2) \\ &\le 2\exp\left(3 \cdot 1.2^{3/2} \cdot M_{P_\star}^{3/2}\varepsilon_{\mathsf{op}}\sqrt{\log_+\left(1.2^3 e M_{P_\star}^3 \cdot \frac{\varepsilon_{\mathrm{algn}}^2}{6M_{P_0}\varepsilon_P^2}\right)}\right) \\ &\le 2\exp\left(4M_{P_\star}^{3/2}\varepsilon_{\mathrm{op}}\sqrt{\log_+\left(\frac{M_{P_\star}^2\varepsilon_{\mathrm{algn}}^2}{\varepsilon_P^2}\right)}\right), \end{aligned}$$

where in the last inequality, we use the simplifications $3 \cdot 1.2^{3/2} \le 4$, $1.2^3 \cdot e/6 \le 1$, and $M_{P_\star} \le M_{P_0}$, since M_{P_0} is the norm of a suboptimal value function. Hence, continuing from Eq. (C.26),

$$\max_{t\in[0,1]} \mathrm{tr}[Y_2(t)] \lesssim \mathcal{C}_K M M_{P_0}\varepsilon_P^2 \log_+\left(\frac{eM_{P_0}^3\varepsilon_{\mathrm{algn}}^2}{\varepsilon_P^2}\right)\exp\left(4M_{P_\star}^{3/2}\varepsilon_{\mathrm{op}}\sqrt{\log_+\left(\frac{M_{P_\star}^2\varepsilon_{\mathrm{algn}}^2}{\varepsilon_P^2}\right)}\right). \tag{C.27}$$

Recall from Eq. (C.24) that $\varepsilon_P^2 := 2\|(\widehat{A} - A_\star)\Sigma_{\mathbf{x},0}^{1/2}\|_{\mathsf{HS}}^2 + 2.4M_{P_\star}\|\widehat{B} - B_\star\|_{\mathsf{HS}}^2\|\Sigma_{\mathbf{x},0}\|_{\mathrm{op}}$. A similar computation to that used to derive the bound in Eq. (C.24) (this time, invoking the domination property in Lemma D.6) lets us bound

$$\begin{aligned} \varepsilon_{\mathrm{algn}}^2 &:= \max_{t\in[0,1]}\left[\Delta_{A_{\mathrm{cl}}}(t)^{\mathsf{H}}\Delta_{A_{\mathrm{cl}}}(t), \Sigma_{\mathbf{x},0}\right]_{\mathrm{algn}} \\ &\le 2\left[(\widehat{A} - A_\star)^{\mathsf{H}}(\widehat{A} - A_\star), \Sigma_{\mathbf{x},0}\right]_{\mathrm{algn}} + 2.4M_{P_\star}\left[(\widehat{B} - B_\star)^{\mathsf{H}}(\widehat{B} - B_\star), \Sigma_{\mathbf{x},0}\right]_{\mathrm{algn}} \\ &\le 2\|\widehat{A} - A_\star\|_{\mathrm{op}}^2 \mathrm{tr}\left[\Sigma_{\mathbf{x},0}\right] + 2.4M_{P_\star}\|\widehat{B} - B_\star\|_{\mathsf{HS}}^2\|\Sigma_{\mathbf{x},0}\|_{\mathrm{op}}. \end{aligned}$$

Using the elementary inequality $\frac{a+b}{c+b} \le 1 + \frac{a}{c+b}$ for $a, b, c \ge 0$, we obtain

$$\frac{\varepsilon_{\text{algn}}^2}{\varepsilon_P^2} \le 1 + \frac{2\|\widehat{A} - A_\star\|_{\text{op}}^2 \text{tr}\left[\Sigma_{\mathbf{x},0}\right]}{\varepsilon_P^2} := \kappa_P.$$

The above bound on $\text{tr}[Y_2(t)]$ then simplifies to:

$$\max_{t\in[0,1]} \text{tr}[Y_2(t)] \lesssim \mathcal{C}_K M M_{P_0} \varepsilon_P^2 \log\left(eM_{P_0}^3 \kappa_P\right) \exp\left(4M_{P_\star}^{3/2} \varepsilon_{\text{op}} \sqrt{\log_+\left(M_{P_\star}^2 \kappa_P\right)}\right). \tag{C.28}$$

Combining with **??** and Eq. (C.16) with $M = 1.2M_{P_\star}$ yields

$$\begin{aligned}\left\|\Sigma_{\mathbf{x},0}^{1/2} P'(t) \Sigma_{\mathbf{x},0}^{1/2}\right\|_{\mathsf{HS}} &\le \sqrt{\|Y_1(t)\|_{\text{op}} \cdot \text{tr}[Y_2(t)]} \\ &\lesssim \sqrt{M_{P_0} M_{P_\star}^3 \mathcal{C}_K \|\Sigma_{\mathbf{x},0}\|_{\text{op}} \varepsilon_P^2 \log\left(eM_{P_0}^3 \kappa_P\right)} \exp\left(2M_{P_\star}^{3/2} \varepsilon_{\text{op}} \sqrt{\log_+\left(M_{P_\star}^2 \kappa_P\right)}\right).\end{aligned}$$

Simplifying the bound Recall that $M = 1.2\|P_\star\|_{\text{op}}$, and $M_{P_\star} := \|P_\star\|_{\text{op}}$. Hence, $M_{P_0} \ge M_{P_\star}$, since $P_0 \succeq P_\star$, as $P_\star$ is the optimal value function for the pair $(A_\star, B_\star)$. Hence, from Lemma D.7. $M_{P_0} \ge M_{P_\star} \ge 1$, $\|K_0\|_{\text{op}}^2 \le \|P_0\|_{\text{op}} := M_{P_0}$, and $\|K(t)\|_{\text{op}}^2 \le 1.1 \|P_\star\|_{\text{op}} = 1.1M_{P_\star}$. We note also that

$$\|K_0 - K(t)\|_{\text{op}}^2 \le 2\|K_0\|_{\text{op}}^2 + 2\|K(t)\|_{\text{op}}^2 \le 2(M_{P_0}^2 + M^2) = 2(M_{P_0}^2 + 1.2^2 M_{P_\star}^2),$$

which is less than or equal to $9M_{P_0}^2$. Noting $M_{P_0} \ge 1$, we can bound

$$\begin{aligned}\mathcal{C}_K &\lesssim \left(1 + \frac{M_{P_0}^2}{\sigma_{\mathbf{u}}^2} \|\Sigma_{\mathbf{x},0}\|_{\text{op}} M_{P_0}^3 \log(2M_{P_0})^2\right) \\ &\lesssim M_{P_0}^5 \log(2M_{P_0})^2 \left(1 + \frac{\|\Sigma_{\mathbf{x},0}\|_{\text{op}}}{\sigma_{\mathbf{u}}^2}\right).\end{aligned}$$

Therefore, we get that $\left\|\Sigma_{\mathbf{x},0}^{1/2} P'(t) \Sigma_{\mathbf{x},0}^{1/2}\right\|_{\mathsf{HS}}$ is upper bounded by:

$$\lesssim \sqrt{M_{P_0}^6 M_{P_\star}^3 \left(1 + \tfrac{\|\Sigma_{\mathbf{x},0}\|_{\text{op}}}{\sigma_{\mathbf{u}}^2}\right) \|\Sigma_{\mathbf{x},0}\|_{\text{op}} \log\left(eM_{P_0}^3 \kappa_P\right)} \cdot \log(2M_{P_0}) \cdot \exp\left(2M_\star^{3/2} \varepsilon_{\text{op}} \sqrt{\log_+\left(M_{P_\star}^2 \kappa_P\right)}\right) \cdot \varepsilon_P.$$

To conclude, we bound

$$\begin{aligned}\log(2M_{P_0}) \cdot \sqrt{\log\left(eM_{P_0}^3 \kappa_P\right)} &\le \log(2M_{P_0}) \left(\sqrt{\log\left(e\kappa_P\right)} + \sqrt{\log(M_{P_0}^3)}\right) \\ &\lesssim M_{P_0} \sqrt{\log_+(\kappa_P)}.\end{aligned}$$

Therefore, the bound further simplifies to:

$$\begin{aligned}&\lesssim M_{P_0}^4 M_{P_\star}^{3/2} \sqrt{\left(1 + \tfrac{\|\Sigma_{\mathbf{x},0}\|_{\text{op}}}{\sigma_{\mathbf{u}}^2}\right) \|\Sigma_{\mathbf{x},0}\|_{\text{op}} \log_+\left(\kappa_P\right)} \exp\left(2M_\star^{5/2} \varepsilon_{\text{op}} \sqrt{\log_+(\kappa_P)}\right) \cdot \varepsilon_P \\ &\lesssim M_{P_0}^4 M_{P_\star}^{3/2} \sqrt{\left(1 + \tfrac{\|\Sigma_{\mathbf{x},0}\|_{\text{op}}}{\sigma_{\mathbf{u}}^2}\right) \|\Sigma_{\mathbf{x},0}\|_{\text{op}} \log_+\left(\kappa_P\right)} \cdot \phi(\kappa_P)^{2M_{P_\star}^{5/2} \varepsilon_{\text{op}}} \cdot \varepsilon_P,\end{aligned}$$

where $\phi(z) = \exp(\sqrt{\log z})$.

Further simplifications in finite dimensions To conclude, we remark on how κ_P can be replaced by $1 + \mathrm{cond}(\Sigma_{\mathbf{x},0})$, where $\mathrm{cond}(\Sigma_{\mathbf{x},0})$ denotes the condition number, in finite dimensions. To see this, we note that the term $\left\|\widehat{A} - A_\star\right\|_{\mathrm{op}}^2 \mathrm{tr}\left[\Sigma_{\mathbf{x},0}\right]$ in κ_P arose from an upper bound on $\left[(\widehat{A} - A_\star)^{\mathsf{H}}(\widehat{A} - A_\star), \Sigma_{\mathbf{x},0}\right]_{\mathrm{algn}}$. By Lemma D.6, one can similarly bound $\left[(\widehat{A} - A_\star)^{\mathsf{H}}(\widehat{A} - A_\star), \Sigma_{\mathbf{x},0}\right]_{\mathrm{algn}} \le \|\Sigma_{\mathbf{x},0}\|_{\mathrm{op}} \left\|\widehat{A} - A_\star\right\|_{\mathsf{HS}}^2$. In finite dimensions with invertible $\Sigma_{\mathbf{x},0}$, we can therefore compute:

$$\left\|\widehat{A} - A_\star\right\|_{\mathsf{HS}}^2 \le \lambda_{\min}(\Sigma_{\mathbf{x},0})^{-1} \left\|(\widehat{A} - A_\star)\Sigma_{\mathbf{x},0}^{1/2}\right\|_{\mathsf{HS}}^2 \le \frac{\varepsilon_P^2}{2\lambda_{\min}(\Sigma_{\mathbf{x},0})}.$$

Thus we can replace:

$$\log(1 + \frac{2\left[(\widehat{A} - A_\star)^{\mathsf{H}}(\widehat{A} - A_\star), \Sigma_{\mathbf{x},0}\right]_{\mathrm{algn}}}{\varepsilon_P^2}) \le \log(1 + \frac{2\|\Sigma_{\mathbf{x},0}\|_{\mathrm{op}}\,\varepsilon_P^2}{2\varepsilon_P^2\lambda_{\min}(\Sigma_{\mathbf{x},0})}) = \log(1 + \mathrm{cond}(\Sigma_{\mathbf{x},0})).$$

□

C.5.2 Proof of Proposition C.9

Before beginning the proof, let us recall the setup. We let (A_1, B_1) and (A_2, B_2) be two systems, and K a controller, and define for $u \in [0, 1]$

$$\bar{A}(u) := A_1 + u(A_2 - A_1) \quad \bar{B}(u) := B_1 + u(B_2 - B_1), \quad \bar{A}_{\mathrm{cl}}(u) := A(u) + B(u)K.$$

Assume that K stabilizes all instances $(\bar{A}(u), \bar{B}(u))$, so that

$$\bar{M} := \max_{u\in[0,1]} \left\|\bar{P}(u)\right\|_{\mathrm{op}} < \infty, \qquad \text{where } \bar{P}(u) := \mathsf{dlyap}\left(\bar{A}_{\mathrm{cl}}(u),\ Q + K^{\mathsf{H}}RK\right).$$

Finally, define $\bar{\Delta}_{A_{\mathrm{cl}}} := \frac{d}{du}\bar{A}_{\mathrm{cl}}(u) = (A_2 - A_1) + (B_2 - B_1)K$.

Proof. The proof is based on analyzing the behavior of the curve,

$$z(u) : [0, 1] \to \mathbb{R} := \mathrm{tr}\left[\bar{\Delta}_{A_{\mathrm{cl}}}^{\mathsf{H}}\bar{\Delta}_{A_{\mathrm{cl}}}\mathsf{dlyap}\left(\bar{A}_{\mathrm{cl}}(u)^{\mathsf{H}},\ \Sigma\right)\right].$$

Since $\bar{M} < \infty$, $\bar{A}_{\mathrm{cl}}(u)$ is stable for all $u \in [0, 1]$, and hence the Lyapunov operator is well defined implying that $z(u)$ is finite.

In order to simplify our presentation, we let $Z(u) = \mathsf{dlyap}\left(\bar{A}_{\mathrm{cl}}(u)^{\mathsf{H}},\ \Sigma\right)$. Since $\bar{A}_{\mathrm{cl}}(u)$ is a linear curve supported on stable operators, Lemma D.19 ensures that $Z(u)$ is continuously Frechet-differentiable, with Frechet derivative equal to

$$Z'(u) = \mathsf{dlyap}\left(\bar{A}_{\mathrm{cl}}(u)^{\mathsf{H}},\ \bar{\Delta}_{A_{\mathrm{cl}}}Z(u)\bar{A}_{\mathrm{cl}}(u)^{\mathsf{H}} + \bar{A}_{\mathrm{cl}}(u)Z(u)\bar{\Delta}_{A_{\mathrm{cl}}}^{\mathsf{H}}\right). \tag{C.29}$$

From part (a) of Lemma D.20, $z(u)$ is continuously differentiable (as a real-valued curve), and its derivative is $z'(u) = \mathrm{tr}[\bar{\Delta}_{A_{\mathrm{cl}}}^{\mathsf{H}}\bar{\Delta}_{A_{\mathrm{cl}}}, Z'(u)]$. By applying the PSD AM-GM inequality twice (Lemma D.3), we have that for any $\alpha > 0$,

$$Z'(u) = \bar{\Delta}_{A_{\mathrm{cl}}}Z(u)\bar{A}_{\mathrm{cl}}(u)^{\mathsf{H}} + \bar{A}_{\mathrm{cl}}(u)Z(u)\bar{\Delta}_{A_{\mathrm{cl}}}^{\mathsf{H}} \preceq \frac{1}{2}\alpha \cdot \bar{\Delta}_{A_{\mathrm{cl}}}Z(u)\bar{\Delta}_{A_{\mathrm{cl}}}^{\mathsf{H}} + \frac{1}{2}\alpha^{-1}\bar{A}_{\mathrm{cl}}(u)Z(u)\bar{A}_{\mathrm{cl}}(u)^{\mathsf{H}}. \tag{C.30}$$

Combining Eq. (C.30) and Eq. (C.29), and optimizing over α, we can upper bound the derivative of $z(u)$,

$$z'(u) = \mathrm{tr}\left[\bar{\Delta}_{A_{\mathrm{cl}}}^{\mathsf{H}}\bar{\Delta}_{A_{\mathrm{cl}}}\mathsf{dlyap}\left(\bar{A}_{\mathrm{cl}}(u)^{\mathsf{H}},\ \bar{\Delta}_{A_{\mathrm{cl}}}Z(u)\bar{A}_{\mathrm{cl}}(u)^{\mathsf{H}} + \bar{A}_{\mathrm{cl}}(u)Z(u)\bar{\Delta}_{A_{\mathrm{cl}}}^{\mathsf{H}}\right)\right]$$

$$\leq \left(\underbrace{\mathrm{tr}\left[\bar{\Delta}_{A_{\mathrm{cl}}}^{\mathsf{H}}\bar{\Delta}_{A_{\mathrm{cl}}}\mathsf{dlyap}\left(\bar{A}_{\mathrm{cl}}(u)^{\mathsf{H}},\ \bar{\Delta}_{A_{\mathrm{cl}}}Z(u)\bar{\Delta}_{A_{\mathrm{cl}}}^{\mathsf{H}}\right)\right]}_{:=R_1}\right)^{1/2} \tag{C.31}$$

$$\times\left(\underbrace{\mathrm{tr}\left[\bar{\Delta}_{A_{\mathrm{cl}}}^{\mathsf{H}}\bar{\Delta}_{A_{\mathrm{cl}}}\mathsf{dlyap}\left(\bar{A}_{\mathrm{cl}}(u)^{\mathsf{H}},\ \bar{A}_{\mathrm{cl}}(u)Z(u)\bar{A}_{\mathrm{cl}}(u)^{\mathsf{H}}\right)\right]}_{:=R_2}\right)^{1/2}. \tag{C.32}$$

We now bound each of R_1 and R_2 individually.

Bounding R_1 Beginning with R_1,

$$\begin{aligned}
&\mathrm{tr}\left[\bar{\Delta}_{A_{\mathrm{cl}}}^{\mathsf{H}}\bar{\Delta}_{A_{\mathrm{cl}}}\mathsf{dlyap}\left(\bar{A}_{\mathrm{cl}}(u)^{\mathsf{H}},\ \bar{\Delta}_{A_{\mathrm{cl}}}Z(u)\bar{\Delta}_{A_{\mathrm{cl}}}^{\mathsf{H}}\right)\right]\\
&\quad\leq\ \left\|\bar{\Delta}_{A_{\mathrm{cl}}}\right\|_{\mathrm{op}}^{2}\mathrm{tr}\left[\mathsf{dlyap}\left(\bar{A}_{\mathrm{cl}}(u)^{\mathsf{H}},\ \bar{\Delta}_{A_{\mathrm{cl}}}Z(u)\bar{\Delta}_{A_{\mathrm{cl}}}^{\mathsf{H}}\right)\right]\\
&\quad\leq\ \left\|\bar{\Delta}_{A_{\mathrm{cl}}}\right\|_{\mathrm{op}}^{2}\cdot\left\|\mathsf{dlyap}\left(\bar{A}_{\mathrm{cl}}(u),\ I\right)\right\|_{\mathrm{op}}\mathrm{tr}\left[\bar{\Delta}_{A_{\mathrm{cl}}}Z(u)\bar{\Delta}_{A_{\mathrm{cl}}}^{\mathsf{H}}\right]\\
&\quad\leq\ \left\|\bar{\Delta}_{A_{\mathrm{cl}}}\right\|_{\mathrm{op}}^{2}\left\|\bar{P}(u)\right\|_{\mathrm{op}}\mathrm{tr}\left[\bar{\Delta}_{A_{\mathrm{cl}}}Z(u)\bar{\Delta}_{A_{\mathrm{cl}}}^{\mathsf{H}}\right].\\
&\quad= \left\|\bar{\Delta}_{A_{\mathrm{cl}}}\right\|_{\mathrm{op}}^{2}\left\|\bar{P}(u)\right\|_{\mathrm{op}}\cdot z(u)\\
&\quad= \left\|\bar{\Delta}_{A_{\mathrm{cl}}}\right\|_{\mathrm{op}}^{2}\bar{M}\cdot z(u)
\end{aligned}$$

In the second line, we have used Lemma D.8. Furthermore, the second to last line is justified by the following observation. Since $I \preceq Q$ and $I \preceq R$,

$$\mathsf{dlyap}\left(\bar{A}_{\mathrm{cl}}(u),\ I\right) \preceq \mathsf{dlyap}\left(\bar{A}_{\mathrm{cl}}(u),\ Q + K^{\mathsf{H}}RK\right) = \bar{P}(u).$$

Bounding R_2 We first notice that,

$$\begin{aligned}
\mathsf{dlyap}\left(\bar{A}_{\mathrm{cl}}(u)^{\mathsf{H}},\ \bar{A}_{\mathrm{cl}}(u)Z(u)\bar{A}_{\mathrm{cl}}(u)^{\mathsf{H}}\right) &= \sum_{j=0}^{\infty}\bar{A}_{\mathrm{cl}}(u)^{j+1}\mathsf{dlyap}\left(\bar{A}_{\mathrm{cl}}(u)^{\mathsf{H}},\ \Sigma\right)(A_{\mathrm{cl}}^{\mathsf{H}})^{j+1}\\
&= \bar{A}_{\mathrm{cl}}(u)\cdot\mathsf{dlyap}\left(\bar{A}_{\mathrm{cl}}(u)^{\mathsf{H}},\ \mathsf{dlyap}\left(\bar{A}_{\mathrm{cl}}(u)^{\mathsf{H}},\ \Sigma\right)\right)\cdot\bar{A}_{\mathrm{cl}}(u)^{\mathsf{H}}\\
&= \mathsf{dlyap}\left(\bar{A}_{\mathrm{cl}}(u)^{\mathsf{H}},\ \mathsf{dlyap}\left(\bar{A}_{\mathrm{cl}}(u)^{\mathsf{H}},\ \Sigma\right)\right) - \mathsf{dlyap}\left(\bar{A}_{\mathrm{cl}}(u)^{\mathsf{H}},\ \Sigma\right)\\
&= \mathsf{dlyap}_{(1)}\left(\bar{A}_{\mathrm{cl}}(u)^{\mathsf{H}},\ \Sigma\right) - \mathsf{dlyap}\left(\bar{A}_{\mathrm{cl}}(u)^{\mathsf{H}},\ \Sigma\right)\\
&\preceq \mathsf{dlyap}_{(1)}\left(\bar{A}_{\mathrm{cl}}(u)^{\mathsf{H}},\ \Sigma\right).
\end{aligned}$$

The third line the calculation above follows from definition of the solution to the Lyapunov equation, i.e Eq. (A.4). The second to last line is justified by Lemma D.10. Setting $X := \bar{\Delta}_{A_{\mathrm{cl}}}^{\mathsf{H}}\bar{\Delta}_{A_{\mathrm{cl}}} \succeq 0$,

$$R_2 = \mathrm{tr}\left[X\mathsf{dlyap}\left(\bar{A}_{\mathrm{cl}}(u)^{\mathsf{H}},\ \bar{A}_{\mathrm{cl}}(u)Z(u)\bar{A}_{\mathrm{cl}}(u)^{\mathsf{H}}\right)\right] \leq \mathrm{tr}\left[X\mathsf{dlyap}_{(1)}\left(\bar{A}_{\mathrm{cl}}(u)^{\mathsf{H}},\ \Sigma\right)\right].$$

Recall $[X,Y]_{\mathrm{algn}} := \max\{\|XWY\|_{\mathsf{tr}} : \|W\|_{\mathrm{op}} = 1\}$. By Lemma D.12, the following bound holds for any $n \ge 0$:

$$\begin{aligned}
R_2 \le \mathrm{tr}\left[X\mathsf{dlyap}_{(1)}\left(\bar{A}_{\mathrm{cl}}(u)^{\mathsf{H}},\ \Sigma\right)\right] &\le n \cdot \mathrm{tr}\left[X\mathsf{dlyap}\left(\bar{A}_{\mathrm{cl}}(u)^{\mathsf{H}},\ \Sigma\right)\right] \\
&\quad + (n+1)\,[X,\Sigma]_{\mathrm{algn}} \left\|\bar{P}(u)\right\|_{\mathrm{op}}^{3} \exp(-\left\|\bar{P}(u)\right\|_{\mathrm{op}}^{-1} n) \\
&= n \cdot z(u) \\
&\quad + (n+1)\,[X,\Sigma]_{\mathrm{algn}} \left\|\bar{P}(u)\right\|_{\mathrm{op}}^{3} \exp(-\left\|\bar{P}(u)\right\|_{\mathrm{op}}^{-1} n) \\
&\le\ n \cdot z(u) + (n+1)\,[X,\Sigma]_{\mathrm{algn}}\,\bar{M}^3 \exp(-\bar{M}^{-1} n).
\end{aligned}$$

Concluding the proof Combining our bounds for R_1 and R_2, then for any $n \ge 0$, we have

$$\begin{aligned}
z'(u) \ &\le\ \sqrt{\left(\left\|\bar{\Delta}_{A_{\mathrm{cl}}}\right\|_{\mathrm{op}}^{2} \bar{M} \cdot z(u)\right)\left(n \cdot z(u) + (n+1)\,[X,\Sigma]_{\mathrm{algn}}\,\bar{M}^3 \exp(-\bar{M}^{-1}n)\right)} \\
&\overset{(i)}{\le}\ \sqrt{n\bar{M}}\left\|\bar{\Delta}_{A_{\mathrm{cl}}}\right\|_{\mathrm{op}} z(u) + \sqrt{n\bar{M}}\left\|\bar{\Delta}_{A_{\mathrm{cl}}}\right\|_{\mathrm{op}} \left(z(u)^{1/2} \cdot \sqrt{\frac{n+1}{n}\,[X,\Sigma]_{\mathrm{algn}}\,\bar{M}^3 \exp(-\bar{M}^{-1}n)}\right) \\
&\overset{(ii)}{\le}\ \sqrt{n\bar{M}}\left\|\bar{\Delta}_{A_{\mathrm{cl}}}\right\|_{\mathrm{op}} z(u) + \sqrt{n\bar{M}}\left\|\bar{\Delta}_{A_{\mathrm{cl}}}\right\|_{\mathrm{op}} \left(\frac{z(u)}{2} + \frac{n+1}{2n}\,[X,\Sigma]_{\mathrm{algn}}\,\bar{M}^3 \exp(-\bar{M}^{-1}n)\right) \\
&\le \underbrace{\frac{3}{2}\sqrt{n\bar{M}}\left\|\bar{\Delta}_{A_{\mathrm{cl}}}\right\|_{\mathrm{op}}}_{:=a} \cdot z(u) + \underbrace{\sqrt{n\bar{M}}\left\|\bar{\Delta}_{A_{\mathrm{cl}}}\right\|_{\mathrm{op}} \cdot [X,\Sigma]_{\mathrm{algn}}\,\bar{M}^3 \exp(-\bar{M}^{-1}n)}_{:=b},
\end{aligned}$$

where (i) uses concavity of the square-root, and (ii) applies AM-GM. In other words, the scalar function $z(u)$ exhibits the self-bounding property $z'(u) \le az(u) + b$ for a, b defined above. Hence, for any slack parameter $\eta > 0$,[4] a standard ODE comparison inequality (see, e.g. Simchowitz and Foster [2020], Lemma D.1) implies that $z(u) \le \tilde{z}(u)$, where $\tilde{z}(u)$ solves the analogous ODE with equality:

$$\tilde{z}(u) = a\tilde{z}'(u) + b + \eta, \quad \tilde{z}(0) = z(0) + \eta. \tag{C.33}$$

The differential equation above has solution:

$$\tilde{z}(u) \ \le\ \left(z(0) + \eta + \frac{b+\eta}{a}\right)\exp(a \cdot u) - \frac{b+\eta}{a}.$$

Taking $\eta \to 0$ and bounding $z(u) \le \tilde{z}(u)$ yields

$$z(1) \ \le\ z(0)\exp(a) + \frac{b}{a}\exp(a).$$

Substituting in the relevant quantities, the following bound holds for any $n \ge 1$:

$$z(1) \le \exp\left(\frac{3}{2}\sqrt{n\bar{M}}\left\|\bar{\Delta}_{A_{\mathrm{cl}}}\right\|_{\mathrm{op}}\right)\left(z(0) + [X,\Sigma]_{\mathrm{algn}}\,\bar{M}^3 \exp(-\bar{M}^{-1}n)\right),$$

Taking $n = \lceil \bar{M}\log\left(\frac{[X,\Sigma]_{\mathrm{algn}}\bar{M}^3}{z(0)}\right)\rceil \le \bar{M}\log_+\left(\frac{e[X,\Sigma]_{\mathrm{algn}}\bar{M}^3}{z(0)}\right)$ lets us bound

$$[X,\Sigma]_{\mathrm{algn}}\,\bar{M}^3 \exp(-\bar{M}^{-1}n) \le z(0),$$

[4] This is useful to ensure strict domination of one ODE by another

and hence

$$z(1) \le 2\exp\left(\frac{3}{2}\bar{M}\left\|\bar{\Delta}_{A_{\mathrm{cl}}}\right\|_{\mathrm{op}}\sqrt{\log_{+}\left(\frac{e\cdot[X,\Sigma]_{\mathrm{algn}}\bar{M}^3}{z(0)}\right)}\right)z(0).$$

Recalling $X = \bar{\Delta}_{A_{\mathrm{cl}}}^{\mathsf{H}}\bar{\Delta}_{A_{\mathrm{cl}}}$, and recognizing $z(1)$ as $\varepsilon_{\mathrm{sys},2}$ and $z(0)$ as $\varepsilon_{\mathrm{sys},1}$, the bound follows. Note also that this bound on the derivative is necessarily non-decreasing in the $\varepsilon_{\mathrm{sys},2}$ argument. □

C.5.3 Proof of Lemma C.8

Proof. We apply the first bullet point of Proposition C.3 to establish both parts of the lemma.

Part 1: Take $(A_1, B_1) = (A_\star, B_\star)$, and $(A_2, B_2) = (A(t), B(t))$. Since (A_2, B_2) lies on the segment joining $(A_\star, B_\star)$ and $(\widehat{A}, \widehat{B})$, we have

$$\max\{\|A_1 - A_2\|_{\mathrm{op}}, \|B_1 - B_2\|_{\mathrm{op}}\} \le \max\{\|A_\star - \widehat{A}\|_{\mathrm{op}}, \|B_\star - \widehat{B}\|_{\mathrm{op}}\} \le \varepsilon_{\mathrm{op}}$$

Hence, if $\varepsilon_{\mathrm{op}} \le \frac{\eta}{16\|P_1\|_{\mathrm{op}}^3}$ for $\eta = 1/11$, Proposition C.3 implies $\|P(t)\|_{\mathrm{op}} \le (1+\eta)\|P_\star\|_{\mathrm{op}} \le 1.1\|P_\star\|_{\mathrm{op}}$.

Part 2 For part 2, take $(A_1, B_1) = (A(t), B(t))$ and $(A_2, B_2) = (A(u\cdot t), B(u\cdot t))$. Again, it holds that $\max\{\|A_1 - A_2\|_{\mathrm{op}}, \|B_1 - B_2\|_{\mathrm{op}}\} \le \varepsilon_{\mathrm{op}}$. Then, if $\varepsilon_{\mathrm{op}} \le \eta/(16\|P_1\|_{\mathrm{op}}^3)$, where $P_1 = P_\infty(A_1, B_1)$, Proposition C.3 implies that

$$\|P_\infty(K(t); A(u\cdot t), B(u\cdot t))\|_{\mathrm{op}} \le (1+\eta)\|P_1\|_{\mathrm{op}}.$$

From part 1 of the present lemma, $\|P_1\|_{\mathrm{op}} \le (1+\eta)\|P_\star\|_{\mathrm{op}}$. Hence, if $\varepsilon_{\mathrm{op}} \le \frac{\eta}{16(1+\eta)^3\|P_\star\|_{\mathrm{op}}^3}$, we have $\|P_\infty(K(t); A(u\cdot t), B(u\cdot t))\|_{\mathrm{op}} \le (1+\eta)^2\|P_\star\|_{\mathrm{op}}$. Computing $(1+\eta)^2 \le 1.2$, and noting that $16(1+\eta)^3/\eta \le 229$ concludes the proof.

□

D Technical Lemmas

This section states and prove the main technical tools used throughout the paper. Appendix D.1 contains linear algebraic tools, Appendix D.2 gives tools for controlling terms involving Lyapunov operators, Appendix D.3 states and proves a comparison theorem between the eigendecay of a PSD operator Λ and its image $\mathsf{dlyap}\,(A,\ \Lambda)$. Finally, Appendix D.4 addresses the relevant differentiability considerations that arise in infinite dimensional spaces.

D.1 Linear Algebra

Lemma D.1. *Let $Z \succeq 0$ and $Y_1, \dots, Y_T$ be bounded linear operators on a Hilbert space $\mathcal{H}_{\mathbf{x}}$. Then,*

$$\left(\sum_{t=1}^{T} Y_t\right) Z \left(\sum_{t=1}^{T} Y_t\right)^{\mathsf{H}} \preceq 2\sum_{t=1}^{T} t^2\, Y_t Z Y_t^{\mathsf{H}}.$$

Proof. Let $\mathbf{x}$ be any vector in $\mathcal{H}_{\mathbf{x}}$, then

$$\begin{aligned}\left\langle \mathbf{x},\ \left(\textstyle\sum_{t=1}^{T} Y_t\right) Z \left(\textstyle\sum_{t=1}^{T} Y_t\right)^{\mathsf{H}} \mathbf{x}\right\rangle_{\mathcal{H}_{\mathbf{x}}} &= \left\langle Z^{1/2}\left(\textstyle\sum_{t=1}^{T} Y_t\right)^{\mathsf{H}}\mathbf{x},\ Z^{1/2}\left(\textstyle\sum_{t=1}^{T} Y_t\right)^{\mathsf{H}}\mathbf{x}\right\rangle_{\mathcal{H}_{\mathbf{x}}} \\ &= \left\|\textstyle\sum_{t=1}^{T} Z^{1/2}\left(\textstyle\sum_{t=1}^{T} Y_t\right)^{\mathsf{H}}\mathbf{x}\right\|_{\mathcal{H}_{\mathbf{x}}}^2.\end{aligned}$$

Therefore it suffices to show that, for any $\mathbf{x}_1, \dots, \mathbf{x}_T \in \mathcal{H}_{\mathbf{x}}$, $\left\|\sum_{t=1}^{T} \mathbf{x}_t\right\|_{\mathcal{H}_{\mathbf{x}}}^2 \le 2\sum_{t=1}^{T} t^2 \cdot \|\mathbf{x}_t\|_{\mathcal{H}_{\mathbf{x}}}^2$. We argue by Cauchy Schwartz,

$$\left\|\sum_{t=1}^{T} \mathbf{x}_t\right\|_{\mathcal{H}_{\mathbf{x}}}^2 = \left\|\sum_{t=1}^{T} \frac{1}{t}\cdot t\cdot \mathbf{x}_t\right\|_{\mathcal{H}_{\mathbf{x}}}^2 \le \left(\sum_{t=1}^{T}\frac{1}{t^2}\right)\cdot \sum_{t=1}^{T}\|t\cdot \mathbf{x}_t\|_{\mathcal{H}_{\mathbf{x}}}^2.$$

Since $\sum_{t=1}^{T} \frac{1}{t^2} < \frac{\pi^2}{6} \le 2$, the bound follows. ☐

Lemma D.2. *Let M be a positive, semi-definite linear operator, then*

$$\log\det(I + M)\ \le\ \operatorname{tr}[M]$$

Proof. If M has eigenvalues $\{\sigma_i\}_{i=1}^{\infty}$, then $I + M$ has eigenvalues $\{1+\sigma_i\}_{i=1}^{\infty}$. Since the determinant of a linear operator is equal to the product of its eigenvalues, we have that,

$$\log\det(I+M) = \sum_i \log(1+\sigma_i)\ \le\ \sum_i \sigma_i = \operatorname{tr}[M],$$

where we have used the numerical inequality $\log(1+x)\ \le\ x$ for all $x\ \ge\ 0$. ☐

Lemma D.3. *Let $X, Y, P : \mathcal{H}_{\mathbf{x}} \to \mathcal{H}_{\mathbf{x}}$ be linear operators and let P be positive semi-definite, then for any $\alpha > 0$ we have that*

$$XPY^{\mathsf{H}} \preceq \frac{\alpha}{2} XPX^{\mathsf{H}} + \frac{1}{2\alpha} YPY^{\mathsf{H}}.$$

Proof. Letting $\mathbf{v} \in \mathcal{H}_{\mathbf{x}}$, the proof follows by direct application of Cauchy-Schwarz and the AM-GM inequality:

$$\begin{aligned}\left\langle \mathbf{v},\ XPY^{\mathsf{H}}\mathbf{v}\right\rangle_{\mathcal{H}_{\mathbf{x}}} &= \left\langle \sqrt{\alpha}P^{\frac{1}{2}}X^{\mathsf{H}}\mathbf{v},\ \frac{1}{\sqrt{\alpha}}P^{\frac{1}{2}}Y^{\mathsf{H}}\mathbf{v}\right\rangle_{\mathcal{H}_{\mathbf{x}}} \\ &\le \left\|\sqrt{\alpha}P^{\frac{1}{2}}X^{\mathsf{H}}\mathbf{v}\right\|_{\mathcal{H}_{\mathbf{x}}}\left\|\frac{1}{\sqrt{\alpha}}P^{\frac{1}{2}}Y^{\mathsf{H}}\mathbf{v}\right\|_{\mathcal{H}_{\mathbf{x}}} \\ &\le \frac{\alpha}{2}\left\|P^{\frac{1}{2}}X^{\mathsf{H}}\mathbf{v}\right\|^2_{\mathcal{H}_{\mathbf{x}}} + \frac{1}{2\alpha^{-1}}\left\|P^{\frac{1}{2}}Y^{\mathsf{H}}\mathbf{v}\right\|^2_{\mathcal{H}_{\mathbf{x}}} \\ &= \frac{\alpha}{2}\left\langle \mathbf{v},\ XPX^{\mathsf{H}}\mathbf{v}\right\rangle_{\mathcal{H}_{\mathbf{x}}} + \frac{1}{2\alpha}\left\langle \mathbf{v},\ YPY^{\mathsf{H}}\mathbf{v}\right\rangle_{\mathcal{H}_{\mathbf{x}}}.\end{aligned}$$

□

Lemma D.4. *Let $X : \mathcal{H}_{\mathbf{x}} \to \mathcal{H}_{\mathbf{x}}$ be a self-adjoint operator and let $Y \in \mathbb{S}_+^{\mathcal{H}_{\mathbf{x}}}$ be a trace class, positive semi-definite operator. If, $-Y \preceq X \preceq Y$ then $\|X\|_{\mathsf{tr}} \le \operatorname{tr}[Y]$.*

Proof. Let $X = \sum_{j=1}^{\infty} \lambda_j \mathbf{q}_j \mathbf{q}_j^{\mathsf{H}}$ denote the spectral decomposition of X. Then,

$$\|X\|_{\mathsf{tr}} = \sum_{j\ge1}|\lambda_j| = \sum_{j\ge1}|\mathbf{q}_j^{\mathsf{H}}X\mathbf{q}_j|,$$

since $-Y \preceq X \preceq Y$, $|\mathbf{q}_j^{\mathsf{H}}X\mathbf{q}_j| \le \mathbf{q}_j^{\mathsf{H}}Y\mathbf{q}_j$ for each j. Thus, $\|X\|_{\mathsf{tr}} \le \sum_{j=1}^{\infty}\mathbf{q}_j^{\mathsf{H}}Y\mathbf{q}_j = \operatorname{tr}[Y]$, since the elements $\mathbf{q}_j$ form an orthonormal basis. □

Lemma D.5. *Let X, Y_1, Y_2 be symmetric operators with $Y_1, Y_2 \succeq 0$. For all $\alpha > 0$, define $Y_{[\alpha]} := \frac{\alpha}{2}Y_1 + \frac{1}{2\alpha}Y_2$, and suppose that $-Y_{[\alpha]} \preceq X \preceq Y_{[\alpha]}$ for any $\alpha > 0$. Then,*

$$\|X\|_{\mathsf{HS}} \le \sqrt{\|Y_1\|_{\mathrm{op}}\operatorname{tr}[Y_2]}.$$

Proof. Let $X = \sum_{j=1}^{\infty} \lambda_j \mathbf{q}_j \mathbf{q}_j^{\mathsf{H}}$ denote the spectral decomposition of X. Then,

$$\|X\|^2_{\mathsf{HS}} = \sum_{j\ge1}\lambda_j^2 = \sum_{j\ge1}(\mathbf{q}_j^{\mathsf{H}}X\mathbf{q}_j)^2.$$

For any fixed j, we have $|\mathbf{q}_j^{\mathsf{H}}X\mathbf{q}_j| \le \inf_{\alpha>0}\mathbf{q}_j^{\top}Y_{[\alpha]}\mathbf{q}$, since for all α, $Y_{[\alpha]} \preceq X \preceq Y_{[\alpha]}$. Moreover, we compute

$$\begin{aligned}|\mathbf{q}_j^{\mathsf{H}}X\mathbf{q}_j| \le \inf_{\alpha>0}\mathbf{q}_j^{\mathsf{H}}Y_{[\alpha]}\mathbf{q}_j, &= \inf_{\alpha>0}\frac{\alpha}{2}\mathbf{q}_j^{\mathsf{H}}Y_1\mathbf{q}_j + \frac{1}{2\alpha}\mathbf{q}_j^{\mathsf{H}}Y_2\mathbf{q}_j \\ &= \sqrt{\mathbf{q}_j^{\mathsf{H}}Y_1\mathbf{q}_j\cdot\mathbf{q}_j^{\mathsf{H}}Y_2\mathbf{q}_j} \le \sqrt{\|Y_1\|_{\mathrm{op}}}\sqrt{\mathbf{q}_j^{\mathsf{H}}Y_2\mathbf{q}_j},\end{aligned}$$

where we note that $\min_{\alpha>0}\frac{a}{2\alpha} + \frac{b}{2\alpha} = \sqrt{ab}$ for nonegative $a, b \ge 0$. Combining the above two displays,

$$\|X\|^2_{\mathsf{HS}} \le \sum_{j=1}^{\infty}\left(\sqrt{\|Y_1\|_{\mathrm{op}}}\sqrt{\mathbf{q}_j^{\mathsf{H}}Y_2\mathbf{q}_j},\right)^2 = \|Y_1\|_{\mathrm{op}}\sum_{j=1}^{\infty}\mathbf{q}_j^{\mathsf{H}}Y_2\mathbf{q}_j.$$

Since $\mathbf{q}_j$ are an orthonormal basis, $\sum_{j=1}^{\infty}\mathbf{q}_j^{\mathsf{H}}Y_2\mathbf{q}_j = \operatorname{tr}[Y_2]$. The bound follows. □

Lemma D.6. *Define the operator $[X, Y]_{\mathrm{algn}} := \max\{\|XWY\|_{\mathsf{tr}} : \|W\|_{\mathrm{op}} = 1\}$. Then for any $X' \succeq X \succeq 0$, $[X, Y]_{\mathrm{algn}} \le [X', Y]_{\mathrm{algn}}$. Similarly, for $Y' \succeq Y \succeq 0$, then $[X, Y']_{\mathrm{algn}} \le [X, Y]_{\mathrm{algn}}$.*

Proof. Let us prove the first point; the second is analogous. Fix $\varepsilon > 0$, and let W be such that $\|XWY\|_{\mathsf{tr}} \geq [X, Y]_{\mathrm{algn}} - \varepsilon$. Let $WY = U\Sigma V^{\mathsf{H}}$ denote the singular value decomposition of WY. Then, $\|XWY\|_{\mathsf{tr}} = \left\|XU\Sigma V^{\mathsf{H}}\right\|_{\mathsf{tr}} = \left\|XU\Sigma U^{\mathsf{H}}\right\|_{\mathsf{tr}}$, since VU^{H} is orthonormal, and thus conjugating by it does not alter the trace norm. Since $X' \succeq X \succeq 0$ and $U\Sigma U^{\mathsf{H}} \succeq 0$,

$$\begin{aligned}\left\|XU\Sigma U^{\mathsf{H}}\right\|_{\mathsf{tr}} &= \mathrm{tr}\left[XU\Sigma U^{\mathsf{H}}\right] \leq \mathrm{tr}\left[X'U\Sigma U^{\mathsf{H}}\right] \\ &= \left\|X'U\Sigma U^{\mathsf{H}}\right\|_{\mathsf{tr}} = \left\|X'U\Sigma V^{\mathsf{H}}\right\|_{\mathsf{tr}} = \|X'WY\|_{\mathsf{tr}} \leq [X', Y]_{\mathrm{algn}}.\end{aligned}$$

The bound follows. ∎

D.2 Lyapunov Theory

Lemma D.7. *Let (A_1, B_1) be a stabilizable instance with stabilizing controller K_1, and let $P_1 = P_\infty(K_1; A_1, B_1)$ be the associated value function. Then, if $R \succeq I$ and $Q \succeq I$, it holds that $P_1 \succeq I$ and $\|K_1\|_{\mathrm{op}} \leq \|P_1\|_{\mathrm{op}}^{1/2}$. In particular, $P_\infty(A_1, B_1) \succeq I$, and $\|K_\infty(A_1, B_1)\|_{\mathrm{op}} \leq \|P_\infty(A_1, B_1)\|_{\mathrm{op}}{}^{1/2}$.*

Proof. We have the identity:

$$P_1 = P_\infty(K_1; A_1, B_1) = \mathsf{dlyap}\left(A_1 + B_1K_1,\ Q + K_1^\top RK_1\right) \succeq Q + K_1RK_1$$

Thus, $P_1 \succeq Q \succeq I$, and since $R \succeq I$, $K_1^\top K_1 \preceq K_1^\top RK_1 \preceq P_1$, so that $\|K_1\|_{\mathrm{op}} \leq \|P_1\|_{\mathrm{op}}^{1/2}$. ∎

Lemma D.8. *Let Y be a trace class operator, and suppose $\mathsf{dlyap}\left(X^{\mathsf{H}},\ I\right)$ is bounded. Then,*

$$\|\mathsf{dlyap}\,(X,\ Y)\|_{\mathsf{tr}} \quad\leq\quad \left\|\mathsf{dlyap}\left(X^{\mathsf{H}},\ I\right)\right\|_{\mathrm{op}} \|Y\|_{\mathsf{tr}}.$$

In particular, if $Y \succeq 0$ is PSD, then

$$\mathrm{tr}[\mathsf{dlyap}\,(X,\ Y)] \quad\leq\quad \left\|\mathsf{dlyap}\left(X^{\mathsf{H}},\ I\right)\right\|_{\mathrm{op}} \mathrm{tr}[Y].$$

Proof. We write out Y in its spectral decomposition, $Y = \sum_{i=0}^{\infty} v_i \otimes v_i \cdot \lambda_i$, and use the form of the Lyapunov solution to get that

$$\begin{aligned}\mathsf{dlyap}\,(X,\ Y) = \sum_{j=0}^{\infty}(X^{\mathsf{H}})^jYX^j &= \sum_{i=0}^{\infty}\sum_{j=0}^{\infty}(X^{\mathsf{H}})^j(v_i \otimes v_i)X^j \cdot \lambda_i \\ &\preceq \sum_{i=0}^{\infty}\sum_{j=0}^{\infty}(X^{\mathsf{H}})^j(v_i \otimes v_i)X^j \cdot |\lambda_i| := Z.\end{aligned}$$

Similarly,

$$\mathsf{dlyap}\,(X,\ Y) \succeq -\sum_{i=0}^{\infty}\sum_{j=0}^{\infty}(X^{\mathsf{H}})^j(v_i \otimes v_i)X^j \cdot |\lambda_i| = -Z.$$

Therefore, from Lemma D.4

$$\begin{aligned}\|\mathsf{dlyap}\,(X,\ Y)\|_{\mathsf{tr}} \leq \mathrm{tr}[Z] &= \sum_{i=0}^{\infty}\|\mathsf{dlyap}\,(X,\ v_i \otimes v_i)\|_{\mathsf{tr}} \cdot |\lambda_i| \\ &= \sum_{i=0}^{\infty}|\left\langle v_i,\ \mathsf{dlyap}\left(X^{\mathsf{H}},\ I\right)v_i\right\rangle_{\mathcal{H}_{\mathsf{x}}}| \cdot |\lambda_i| \\ &\leq \left\|\mathsf{dlyap}\left(X^{\mathsf{H}},\ I\right)\right\|_{\mathrm{op}} \sum_{i=0}^{\infty}|\lambda_i|.\end{aligned}$$

Noting that $\sum_{i=0}^{\infty} \lambda_i = \|Y\|_{\mathsf{tr}}$ finishes the first part of the proof. For the second part, notice that if $Y \succeq 0$, then the nuclear and trace norms norms coincide.

□

Lemma D.9. *Let A be stable and $P = \mathsf{dlyap}\,(A,\ \Sigma)$ be the corresponding value function for some $\Sigma \succeq I$, then for all integers $j\ \geq\ 0$,*

$$(A_{\mathrm{cl}}^{\mathsf{H}})^j P A^j \preceq P(1 - \|P\|_{\mathrm{op}}^{-1})^j.$$

In particular, for all $j \geq 0$,

$$\|A^j\|_{\mathrm{op}}^2 \leq \|P\|_{\mathrm{op}}(1 - \|P\|_{\mathrm{op}}^{-1})^j.$$

Hence, for any over $\Sigma' \succeq 0$, $\|\mathsf{dlyap}\,(A,\ \Sigma)\,\|_{\mathrm{op}} \leq \|\Sigma\|_{\mathrm{op}}\|P\|_{\mathrm{op}}^2$.

Proof. Since P satisfies the Lyapunov equation, $A^{\mathsf{H}}PA - P + \Sigma = 0$, we can write for any $x \in \mathcal{H}_{\mathbf{x}}$:

$$\begin{aligned}
\left\langle x,\ A^{\mathsf{H}}PAx\right\rangle_{\mathcal{H}_{\mathbf{x}}} &= \langle x,\ Px\rangle_{\mathcal{H}_{\mathbf{x}}} - \langle x,\ \Sigma x\rangle_{\mathcal{H}_{\mathbf{x}}} \\
&= \langle x,\ Px\rangle_{\mathcal{H}_{\mathbf{x}}}\left(1 - \frac{\langle x,\ \Sigma x\rangle_{\mathcal{H}_{\mathbf{x}}}}{\langle x,\ Px\rangle_{\mathcal{H}_{\mathbf{x}}}}\right) \\
&\leq\ \langle x,\ Px\rangle_{\mathcal{H}_{\mathbf{x}}}(1 - \frac{1}{\|P\|_{\mathrm{op}}}).
\end{aligned}$$

In the last line, we have use the assumption that $\Sigma \succeq I$. Hence, $A^{\mathsf{H}}PA \preceq P(1 - \|P\|_{\mathrm{op}}^{-1})$ where $\|P\|_{\mathrm{op}} > 1$. Repeating this argument, we can in fact show that,

$$(A^{\mathsf{H}})^j P A^j \preceq P(1 - \|P\|_{\mathrm{op}}^{-1})^j.$$

For the second guarantee, since $\Sigma \succeq I$, we have $P \succeq \Sigma \succeq I$. Hence,

$$\left\|A^j\right\|_{\mathrm{op}}^2 = \left\|(A^j)^{\mathsf{H}}A^j\right\|_{\mathrm{op}}\ \leq\ \left\|(A^j)^{\mathsf{H}}PA^j\right\|_{\mathrm{op}}\ \leq\ \|P\|_{\mathrm{op}}(1 - \|P\|_{\mathrm{op}}^{-1})^j.$$

□

For the next lemma, we recall the higher order Lyapunov operator from Definition A.2,

$$\mathsf{dlyap}_{(m)}\,(A,\ \Lambda) := \sum_{j=0}^{\infty}(A^{\mathsf{H}})^j \Lambda A^j (j+1)^m.$$

Lemma D.10. *Let A be a stable linear operator and Σ be self-adjoint, the the following identity holds:*

$$\sum_{j=0}^{\infty} A^j \mathsf{dlyap}\left(A^{\mathsf{H}},\ \Sigma\right)(A^{\mathsf{H}})^j = \sum_{j=0}^{\infty} A^j \Sigma (A^{\mathsf{H}})^j \cdot (j+1).$$

Equivalently,

$$\mathsf{dlyap}\left(A^{\mathsf{H}},\ \mathsf{dlyap}\left(A^{\mathsf{H}},\ \Sigma\right)\right) = \mathsf{dlyap}_{(1)}\left(A^{\mathsf{H}},\ \Sigma\right).$$

Proof. To simplify the proof, let $\Gamma = \mathsf{dlyap}\left(A^{\mathsf{H}},\ \Sigma\right)$. Since $\Gamma = \mathsf{dlyap}\left(A^{\mathsf{H}},\ \Sigma\right)$ is the solution to the Lyapunov equation $A\Gamma A^{\mathsf{H}} + \Sigma - \Gamma = 0$, we have that

$$A\Gamma A^{\mathsf{H}}\Sigma = \Gamma - \Sigma.$$

By repeating this argument, we can in fact show that:

$$(A)^j\Gamma(A^{\mathsf{H}})^j = \Gamma - \sum_{i=0}^{j-1} A^i\Gamma(A^{\mathsf{H}})^i.$$

Using the fact that $\Gamma = \sum_{j=0}^{\infty} A^j\Sigma(A^{\mathsf{H}})^j$, it follows that $(A)^j\Gamma(A^{\mathsf{H}})^j = \sum_{i=j}^{\infty} A^i\Sigma(A^{\mathsf{H}})^i$. Therefore, we can rewrite $\mathsf{dlyap}\left(A^{\mathsf{H}},\ \mathsf{dlyap}\left(A^{\mathsf{H}},\ \Sigma\right)\right)$ as follows,

$$\begin{aligned}\sum_{j=0}^{\infty} A^j\Gamma(A^{\mathsf{H}})^j &= \sum_{j=0}^{\infty}\sum_{i=j}^{\infty} A^i\Sigma(A^{\mathsf{H}})^i \\ &= \sum_{j=0}^{\infty} A^j\Sigma(A^{\mathsf{H}})^j \cdot (j+1).\end{aligned}$$

which is exactly $\mathsf{dlyap}_{(1)}\left(A^{\mathsf{H}},\ \Sigma\right)$. □

Lemma D.11. *Let A be stable and $P = \mathsf{dlyap}\left(A,\ Q + K^{\mathsf{H}}RK\right)$ be the corresponding value function, then for all integers $n \geq 0$,*

For $m = 1$:

$$\mathsf{dlyap}_{(1)}\left(A^{\mathsf{H}},\ \Sigma\right) \preceq n \cdot \mathsf{dlyap}\left(A^{\mathsf{H}},\ \Sigma\right) + (n+1)\left\|\Sigma\right\|_{\mathrm{op}}\left\|P\right\|_{\mathrm{op}}^{3}\exp(-\left\|P\right\|_{\mathrm{op}}^{-1} n)\cdot I$$

For $m = 2$:

$$\mathsf{dlyap}_{(2)}\left(A^{\mathsf{H}},\ \Sigma\right) \preceq n^2 \cdot \mathsf{dlyap}\left(A^{\mathsf{H}},\ \Sigma\right) + (n^2+2n+2)\left\|\Sigma\right\|_{\mathrm{op}}\left\|P\right\|_{\mathrm{op}}^{4}\exp(-\left\|P\right\|_{\mathrm{op}}^{-1} n)\cdot I$$

Proof. We begin by expanding out the definition of $\mathsf{dlyap}_{(m)}\left(A(t)^{\mathsf{H}},\ \Sigma\right)$,

$$\begin{aligned}\mathsf{dlyap}_{(m)}\left(A^{\mathsf{H}},\ \Sigma\right) &= \sum_{j=0}^{\infty} A^j\Sigma(A^{\mathsf{H}})^j\cdot(j+1)^m \\ &\preceq n^m\sum_{j=0}^{n-1} A^j\Sigma(A^{\mathsf{H}})^j + \sum_{j=n}^{\infty} A^j\Sigma(A^{\mathsf{H}})^j\cdot(j+1)^m \\ &\preceq n^m\mathsf{dlyap}\left(A^{\mathsf{H}},\ \Sigma\right) + \sum_{j=n}^{\infty} A^j\Sigma(A^{\mathsf{H}})^j\cdot(j+1)^m, \end{aligned} \tag{D.1}$$

where we have let the sum go to infinity in the first term. Focusing on the second term,

$$\begin{aligned}\sum_{j=n}^{\infty} A^j\Sigma(A^{\mathsf{H}})^j\cdot(j+1)^m &\preceq I\cdot\left\|\Sigma\right\|_{\mathrm{op}}\sum_{j=n}^{\infty}\left\|A^j(A^{\mathsf{H}})^j\right\|_{\mathrm{op}}\cdot(j+1)^m \\ &= I\cdot\left\|\Sigma\right\|_{\mathrm{op}}\sum_{j=n}^{\infty}\left\|(A^{\mathsf{H}})^jA^j\right\|_{\mathrm{op}}\cdot(j+1)^m && (\left\|NN^{\mathsf{H}}\right\|_{\mathrm{op}} = \left\|N^{\mathsf{H}}N\right\|_{\mathrm{op}}) \\ &\preceq I\cdot\left\|\Sigma\right\|_{\mathrm{op}}\sum_{j=n}^{\infty}\left\|(A^{\mathsf{H}})^jPA^j\right\|_{\mathrm{op}}\cdot(j+1)^m && (P \succeq I) \\ &\preceq \left\|\Sigma\right\|_{\mathrm{op}}\left\|P\right\|_{\mathrm{op}}\sum_{j=n}^{\infty}(j+1)\cdot(1-\left\|P\right\|_{\mathrm{op}}^{-1})^j\cdot I. && (\text{Lemma D.9})\end{aligned}$$

The result then follows by applying the following two identities regarding geometric series, which hold for $c \in (0,1)$,

$$\sum_{j=n}^{\infty}(1-c)^j \cdot (j+1) = \frac{(1-c)^n(cn+1)}{c^2}$$
$$\sum_{j=n}^{\infty}(1-c)^j \cdot (j+1)^2 = \frac{(1-c)^n(c^2n^2+2cn-c)}{c^3},$$

and using the fact that $(1-c)^t \leq \exp(-ct)$. ☐

Lemma D.12. *Let A be stable and $P = \mathsf{dlyap}\left(A,\ Q + K^{\mathsf{H}}RK\right)$ be the corresponding value function, then for all integers $n \geq 0$, and all PSD, bounded operators X, and $\Sigma \succeq 0$,*

$$\operatorname{tr}\left[X\mathsf{dlyap}_{(1)}\left(A^{\mathsf{H}},\ \Sigma\right)\right] \leq n \cdot \operatorname{tr}\left[X\mathsf{dlyap}\left(A^{\mathsf{H}},\ \Sigma\right)\right] + (n+1)\left[X,\Sigma\right]_{\mathrm{algn}} \|P\|_{\mathrm{op}}^3 \exp(-\|P\|_{\mathrm{op}}^{-1} n),$$

where $[X,Y]_{\mathrm{algn}} := \max\{\|XWY\|_{\mathsf{tr}} : \|W\|_{\mathrm{op}} = 1\}$.

Proof. From Eq. (D.1) and the fact that $\operatorname{tr}[WY] \leq \operatorname{tr}[WZ]$ for $W \succeq 0$ and $Y \preceq Z$,

$$\operatorname{tr}\left[X\mathsf{dlyap}_{(1)}\left(A^{\mathsf{H}},\ \Sigma\right)\right] \leq n\operatorname{tr}\left[X\mathsf{dlyap}\left(A^{\mathsf{H}},\ \Sigma\right)\right] + \operatorname{tr}\left[X\sum_{j=n}^{\infty}A^j\Sigma(A^{\mathsf{H}})^j \cdot (j+1)\right].$$

We can then bound,

$$\operatorname{tr}\left[XA^j\Sigma(A^{\mathsf{H}})^j\right] \leq \|A^j\|_{\mathrm{op}}^2 \left\|X \cdot \frac{A^j}{\|A^j\|_{\mathrm{op}}} \cdot \Sigma\right\|_{\mathsf{tr}} \leq \|A^j\|_{\mathrm{op}}^2 \left[X,\Sigma\right]_{\mathrm{algn}} = \|A^j(A^j)^{\mathsf{H}}\|_{\mathrm{op}} \left[X,\Sigma\right]_{\mathrm{algn}}, \tag{D.2}$$

where we note the above bound holds even if $A^j = 0$. The bound now follows from the computation given in the proof of Lemma D.11. ☐

D.3 Spectrum Comparison under Lyapunov Operator

Lemma D.13 (Iterated Weyl's Eigenvalue Inequality)**.** *Let $X_1, \ldots, X_n$ be a sequence of PSD operators, and let $\lambda_j(\cdot)$ denote the j-eigenvalue. Then, $\lambda_j(\sum_{i=1}^n X_i) \leq \sum_{i=1}^n \lambda_{\lceil j/n \rceil}(X_i)$.*

Proof. Consider $n = 2$. Then, for any indices k_1, k_2 such that $k_1 + k_2 - 1 \leq j$, $\lambda_i(X_1 + X_2) \leq \lambda_{k_1}(X_1) + \lambda_{k_2}(X_2)$. Hence, for general n,

$$\lambda_i\left(\sum_{i=1}^n X_i\right) \leq \lambda_{k_{1:n-1}}\left(\sum_{i=1}^{n-1} X_i\right) + \lambda_{k_n}(X_n),$$

where $k_n + k_{1:n-1} \leq j+1$. Iterating,

$$\lambda_{k_{1:n-1}}\left(\sum_{i=1}^{n-1} X_i\right) \leq \lambda_{k_{n-1}}(X_{n-1}) + \lambda_{k_{1:n-2}}\left(\sum_{i=1}^{n-2} X_i\right),$$

where $k_{n-1} + k_{1:n-2} \leq k_{1:n-1} + 1$. Continuing, we have that for any $k_1, \ldots, k_n$ with $\sum_{i=1}^n k_i \leq j + (n-1)$,

$$\lambda_j\left(\sum_{i=1}^n X_i\right) \leq \sum_{i=1}^n \lambda_{k_i}(X_i).$$

Taking $k_i = \lceil j/n \rceil$, we have $\sum_{i=1}^n k_i = n\lceil j/n \rceil$. Now, if j/n is integral, then $n\lceil j/n \rceil = j$. Otherwise, $\lceil j/n \rceil = \lceil (j-1)/n \rceil \leq 1 + \frac{j-1}{n}$, so $n\lceil j/n \rceil \leq n + j - 1$. In either case, $\sum_{i=1}^n k_i \leq j + (n-1)$, as needed. ☐

Lemma D.14. *Let X be a bounded operator on a Hilbert space $\mathcal{H}$, and $\Lambda \succeq 0$ be a PSD operator on $\mathcal{H}$. Then $\sigma_j(X\Lambda X^{\mathsf{H}}) \le \|X\|_{\mathrm{op}}^2 \sigma_j(\Lambda)$.*

Proof. Since $\sigma_j(X\Lambda X^{\mathsf{H}}) = \sigma_j(X\Lambda^{1/2})^2$, it suffices to show that for any two bounded operators A, B, $\sigma_j(AB) \le \|A\|\sigma_j(B)$. This follows using the variation representation of singular values:

$$\begin{aligned}\sigma_j(AB) &= \min_{\text{subspaces } \mathcal{V}\subseteq\mathcal{H} \text{ of dimension } j} \left(\max_{\mathbf{q}\in\mathcal{V}:\|\mathbf{q}\|_{\mathcal{H}}=1} \|AB\mathbf{q}\|_{\mathcal{H}} \right) \\ &\le \|A\|_{\mathrm{op}} \cdot \min_{\text{subspaces } \mathcal{V}\subseteq\mathcal{H} \text{ of dimension } j} \left(\max_{\mathbf{q}\in\mathcal{V}:\|\mathbf{q}\|_{\mathcal{H}}=1} \|B\mathbf{q}\|_{\mathcal{H}} \right) \\ &= \|A\|_{\mathrm{op}}\sigma_j(B).\end{aligned}$$

□

Lemma D.15. *Let A be a stable operator, $\Lambda \succeq 0$, and let $\Sigma = \mathsf{dlyap}\left(A^{\mathsf{H}},\ \Lambda\right)$, and $P = \mathsf{dlyap}\left(A,\ I\right)$. Then, for all indices $j, n \ge 1$, we can bound*

$$\sigma_j(\Sigma) \le \|P\|_{\mathrm{op}}^2 \left(\sigma_{\lceil \frac{j}{n+1} \rceil}(\Lambda) + (1 - \|P\|_{\mathrm{op}}^{-1})^n \|\Lambda\|_{\mathrm{op}} \right)$$

$$\sum_{j \ge k} \sigma_j(\Sigma) \le (n+1)\|P\|_{\mathrm{op}}^2 \left(\sum_{j \ge \lceil \frac{k}{n+1} \rceil} \sigma_j(\Lambda) + (1 - \|P\|_{\mathrm{op}}^{-1})^n \mathrm{tr}\left[\Lambda\right] \right).$$

Moreover, from monotonicity of dlyap *(Lemma A.1), the above also holds for $P = \mathsf{dlyap}\left(A,\ Z\right)$ for any $Z \succeq I$. In particular, if $n \ge \|P\|_{\mathrm{op}} \log \frac{2\|\Lambda\|_{\mathrm{op}}}{\lambda}$, then,*

$$d_\lambda(\Sigma) \le (n+1) d_{\lambda/(2\|P\|_{\mathrm{op}}^2)}(\Lambda).$$

Proof. Fix an integer n, and define the matrices

$$X_i := A^i \Lambda (A^{\mathsf{H}})^i, \quad Y_n := \sum_{i>n} X_i.$$

By Lemma D.13,

$$\sigma_j(\Sigma) = \sigma_j(\sum_{i=1}^n X_i + Y_n) \le \sum_{i=1}^n \sigma_{\lceil \frac{j}{n+1} \rceil}(X_i) + \sigma_{\lceil \frac{j}{n+1} \rceil}(Y_n). \tag{D.3}$$

Now, for any index k, we can bound

$$\sigma_k(X_i) = \sigma_k(A^i \Lambda (A^{\mathsf{H}})^i) = \sigma_k(A^i \Lambda^{1/2}) \le \|A^i\|_{\mathrm{op}}^2 \sigma_k(\Lambda).$$

Hence,

$$\begin{aligned}\sigma_j(\Sigma) &\le \left(\sum_{i=1}^n \|A^i\|_{\mathrm{op}}^2 \right) \sigma_{\lceil \frac{j}{n+1} \rceil}(\Lambda) + \sigma_{\lceil \frac{j}{n+1} \rceil}(Y_n) \\ &\le \|P\|_{\mathrm{op}}^2 \sigma_{\lceil \frac{j}{n+1} \rceil}(\Lambda) + \sigma_{\lceil \frac{j}{n+1} \rceil}(Y_n),\end{aligned} \tag{D.4}$$

where in the last line we use Lemma D.9 to bound $\|A^i\|_{\mathrm{op}}^2$ by a geometric series.

Now, let us bound the operator norm and trace of Y_n. Again, by Lemma D.9, we have

$$\|Y_n\|_{\mathrm{op}} = \left\| \sum_{i>n} A^i \Lambda (A^{\mathsf{H}})^i \right\|_{\mathrm{op}} \le \sum_{i>n} \|A^i\|_{\mathrm{op}}^2 \|\Lambda\|_{\mathrm{op}} \le \|P\|_{\mathrm{op}}^2 (1 - \|P\|_{\mathrm{op}}^{-1})^n \|\Lambda\|_{\mathrm{op}} \tag{D.5}$$

and similarly

$$\|Y_n\|_{\mathsf{tr}} \le \|P\|_{\mathrm{op}}^2 (1 - \|P\|_{\mathrm{op}}^{-1})^n \mathrm{tr}\,[\Lambda]\,. \tag{D.6}$$

Combining these bounds, we have

$$\begin{aligned}
\sigma_j(\Sigma) &\le \|P\|_{\mathrm{op}}^2 \sigma_{\lceil \frac{j}{n+1} \rceil}(\Lambda) + \sigma_{\lceil \frac{j}{n+1} \rceil}(Y_n), && \text{(by Eq. (D.4))}\\
&\le \|P\|_{\mathrm{op}}^2 \sigma_{\lceil \frac{j}{n+1} \rceil}(\Lambda) + \|Y_n\|_{\mathrm{op}}\,, && \text{(by Eq. (D.5))}\\
&\le \|P\|_{\mathrm{op}}^2 \left(\sigma_{\lceil \frac{j}{n+1} \rceil}(\Lambda) + (1 - \|P\|_{\mathrm{op}}^{-1})^n \|\Lambda\|_{\mathrm{op}}\right).
\end{aligned}$$

and,

$$\begin{aligned}
\sum_{j \ge k} \sigma_j(\Sigma) &\le \|P\|_{\mathrm{op}}^2 \sum_{j \ge k} \sigma_{\lceil \frac{j}{n+1} \rceil}(\Lambda) + \sum_{j \ge k} \sigma_{\lceil \frac{j}{n+1} \rceil}(Y_n) && \text{(by Eq. (D.4))}\\
&\overset{(i)}{\le} (n+1)\|P\|_{\mathrm{op}}^2 \sum_{j \ge \lceil \frac{k}{n+1} \rceil} \sigma_j(\Lambda) + (n+1) \sum_{j \ge \lceil \frac{k}{n+1} \rceil} \sigma_j(Y_n) && \text{(D.7)}\\
&\le (n+1)\|P\|_{\mathrm{op}}^2 \sum_{j \ge \lceil \frac{k}{n+1} \rceil} \sigma_j(\Lambda) + (n+1)\mathrm{tr}\,[Y_n]\\
&\le (n+1)\|P\|_{\mathrm{op}}^2 \left(\sum_{j \ge \lceil \frac{k}{n+1} \rceil} \sigma_j(\Lambda) + (1 - \|P\|_{\mathrm{op}}^{-1})^n \mathrm{tr}\,[\Lambda] \right), && \text{(by Eq. (D.6))}
\end{aligned}$$

concluding the proof. Here, in inequality (i), we used that summing over the indices $\lceil \frac{j}{n+1} \rceil j$ for $j \ge k$ sums over only indices $j' \ge \lceil \frac{k}{n+1} \rceil$, and includes each such index at most n times.

To see why the last statement holds, for $a := (n+1) d_{\lambda/(2\|P\|_{\mathrm{op}})}$, by the first statement,

$$\sigma_a(\Sigma) \;\le\; \|P\|_{\mathrm{op}}^2 \left(\sigma_{d_{\lambda/(2\|P\|_{\mathrm{op}})}}(\Lambda) + \exp(-n \|P\|_{\mathrm{op}}^{-1}) \|\Lambda\|_{\mathrm{op}}\right).$$

By definition of d_λ, $\|P\|_{\mathrm{op}}^2 \sigma_{d_{\lambda/(2\|P\|_{\mathrm{op}})}}(\Lambda) \;\le\; \lambda/2$ and for $n \;\ge\; \|P\|_{\mathrm{op}} \log(2\mathrm{tr}\,[\Lambda]\,.\lambda)$ the second term is also smaller than $\lambda/2$. Hence $d_\lambda(\Sigma)$ is at most $(n+1) d_{\lambda/(2\|P\|_{\mathrm{op}}^2)}(\Lambda)$.

□

D.4 Frechet Differentiability

Definition D.1. Let $\mathscr{B}_{\mathrm{op}}$ denote the Banach space consisting of all operator-norm bounded operators on $\mathcal{H}_{\mathbf{x}}$, and let $\mathscr{S}_{\mathrm{op}}$ denote the subspace of self-adjoint bounded operators on $\mathcal{H}_{\mathbf{x}}$.

Lemma D.16. *Let $X \in \mathscr{B}_{\mathrm{op}}$ have spectral radius $\rho(X) < 1$, and for $Y \in \mathscr{S}_{\mathrm{op}}$, define $\mathcal{T}_X[Y] = Y - X^{\mathsf{H}} Y X$. Then, $\mathcal{T}_X[\cdot]$ is a bounded linear operator on $\mathscr{S}_{\mathrm{op}}$, and has an inverse which is also bounded. Hence, the solution to $\mathcal{T}_X[Y] = Q$ is bounded for any $Q \in \mathscr{S}_{\mathrm{op}}$, and given by* $\mathsf{dlyap}\,(X,\; Q)$.

Definition D.2 (Frechet Derivative)**.** Let $\mathscr{B}_1, \mathscr{B}_2$ be a banach space with norms $\|\cdot\|_{\mathscr{B}_1}$ and $\|\cdot\|_{\mathscr{B}_2}$. Given a subset $\mathcal{U} \subset \mathscr{B}_1$, we say $f : \mathcal{U} \to \mathscr{B}_2$ is *Frechet Differentiable* at $x \in \mathscr{B}_1$ if there exists linear mapping $\mathrm{D}f(x) : \mathscr{B}_1 \to \mathscr{B}_2$ such that

$$\lim_{\|h\|_{\mathscr{B}_1} \to 0} \frac{\|f(x+h) - \mathrm{D}f(x)[h])\|_{\mathscr{B}_2}}{\|h\|} = 0.$$

We say that f is *Frechet continuously differentriable* at x if f is Frechet differentiable in an open neighborhood about x, and $x \mapsto \mathrm{D}f(x)[h]$ is a continuous in the dual norm $\|g\|_{\mathscr{B}_1^*} := \sup_{\|h\|_{\mathscr{B}}=1} g(h)$.

An important case is where $f(t) : \mathbb{R} \to \mathscr{B}$ is a curve. In this case, the Frechet Derivative exists if there exists an $f'(t)$ such that

$$\lim_{\varepsilon \to 0} \frac{\|f(t+\varepsilon) - f'(t)\|_{\mathscr{B}}}{\varepsilon} = 0.$$

Lemma D.17 (Implicit Function Theorem [Whittlesey, 1965])**.** *Let $\mathscr{B}_1, \mathscr{B}_2$ be a two Banach spaces, and let $f : \mathscr{B}_1 \times \mathscr{B}_2 \to \mathscr{B}_2$ be a continuously Frechet differentiable function at $(x_0, y_0) \in \mathscr{B}_1 \times \mathscr{B}_2$. Suppose that the mapping $h_2 \mapsto \mathrm{D}f(x_0, y_0)[0, h_2]$ is a Banach space isopmorphism on $\mathscr{B}_2$ (i.e, it admits a a bounded linear inverse). Then, there exists a neighborhood $\mathcal{U}$ around x_0 and $\mathcal{V}$ around y_0 and a Frechet-differentiable function $g : \mathcal{U} \to \mathcal{V}$ on which $g(x)$ is the unique solution to $f(x, g(x)) = 0$.*

Lemma D.18. *Let (A_1, B_1) and (A_2, B_2) be two instance, and define the curve $A(t) = A_1 + t(B_2 - B_1)$ and $B(t) = B_1 + t(B_2 - B_1)$. If $(A(t), B(t))$ is stabilizable for all $t \in [0, 1]$,*

- *$P(t) := P_\infty(A(t), B(t))$ is Frechet differentiable on $[0, 1]$ in the operator norm $\|\cdot\|_{\mathrm{op}}$.*
- *The Frechet derivative of $P(t)$ is given by*

$$P'(t) = \mathsf{dlyap}\left(A_{\mathrm{cl}}(t),\ E(t)\right) \text{ where } E(t) := A_{\mathrm{cl}}(t)^{\mathsf{H}} P(t) \Delta_{A_{\mathrm{cl}}}(t) + \Delta_{A_{\mathrm{cl}}}(t)^{\mathsf{H}} P(t) A_{\mathrm{cl}}(t), \tag{D.8}$$

where $\Delta_{A_{\mathrm{cl}}}(t) := (A_2 - A_1) + (B_2 - B_1) K_\infty\left(A(t), B(t)\right)$.

Proof. The result is the infinite-dimensional analogue of Lemma 3.1 in Simchowitz and Foster [2020]. Define the function $\mathcal{F}(t, P) := \mathbb{R} \times \mathscr{S}_{\mathrm{op}}$ as

$$\mathcal{F}(t, P) := A(t)^{\mathsf{H}} P A(t) - P + Q - A(t)^{\mathsf{H}} P B(t) (R + B(t)^{\mathsf{H}} P B(t))^{-1} B(t)^{\mathsf{H}} P A(t)$$

Then, $P(t)$ solves $\mathcal{F}(t, P) = 0$. One can verify that $\mathcal{F}(t, P)$ is a Frechet continuously differentiable function[5] The computation of Lemma 3.1 in Simchowitz and Foster [2020] (carried out now in infinite dimensions) shows that,

$$\left\{ \frac{\mathrm{d}}{\mathrm{d}t} \mathcal{F}(t, P) \cdot (\delta t) + \frac{\mathrm{d}}{\mathrm{d}P} \mathcal{F}(t, P) \cdot (\delta P) \right\} |_{P = P(t)} = \mathcal{T}_{A(t) + B(t) K(t)}[\delta P] + E(t) \cdot \delta t,$$

where $K(t) = K_\infty(A(t), B(t))$, and where $E(t)$ is defined in Eq. (D.8), and $\mathcal{T}_X[Y] = Y - X^{\mathsf{H}} Y X$. Since $(A(t), B(t))$ is stabilizable, $A(t) + B(t) K(t)$ is stable and thus $\mathcal{T}_{A(t)+B(t)K(t)}[\cdot]$ is a bounded linear operator on $\mathscr{S}_{\mathrm{op}}$ with bounded inverse. It follows from the Lemma D.17 that $P(t)$ admits a Frechet Derivative $P'(t)$, and its derivative must $\mathcal{T}_{A(t)+B(t)K(t)}[\delta P] + E(t) = 0$. Hence, by Lemma D.16, $P'(t)$ is given by Eq. (D.8). □

Lemma D.19. *Let $A(t) = A_1 + (A_2 - A_1)t$ be a curve on $[0, 1]$ with $A(t) \in \mathscr{B}_{\mathrm{op}}$ stabilizable. Set $\Delta = A_2 - A_1$.Then, $t \mapsto \mathsf{dlyap}\left(A(t),\ \Sigma\right)$ is Frechet differentiable in the operator norm, and*

$$\frac{\mathrm{d}}{\mathrm{d}t} \mathsf{dlyap}\left(A(t),\ \Sigma\right) = \mathsf{dlyap}\left(A(t),\ N\right) \tag{D.9}$$

where $N := \Delta^{\mathsf{H}} \mathsf{dlyap}\left(A(t),\ \Sigma\right) A(t) + A(t)^{\mathsf{H}} \mathsf{dlyap}\left(A(t),\ \Sigma\right) \Delta$.

Proof. Set $P(t) := \mathsf{dlyap}\left(A(t),\ \Sigma\right)$. Recall the operator $\mathcal{T}_X[Y] = Y - X^{\mathsf{H}} Y X$. Then, $P(t)$ solves $\mathcal{F}(t, P) = 0$, where $\mathcal{F}(t, P) = \mathcal{T}_{(A(t)}[P] - \Sigma$. It is straightforward to check that $\mathcal{F}(t, P)$ is Frechet continuously differentiable. A direct computation reveals

$$\left\{ \frac{\mathrm{d}}{\mathrm{d}t} \mathcal{F}(t, P) \cdot (\delta t) + \frac{\mathrm{d}}{\mathrm{d}P} \mathcal{F}(t, P) \cdot (\delta P) \right\} |_{P = P(t)} = \mathcal{T}_{A(t)}[\delta P] + N(t) \cdot \delta t,$$

[5] Indeed, the first term in a polynomial in bounded operators, and the second term is a product of surge polynomials, and the matrix inverse $(R + B(t)^{\mathsf{H}} P B(t))^{-1}$. Since $R + B(t)^{\mathsf{H}} P B(t) \succeq R \succ 0$ is finite dimensional, one can express perturbations of $(R + B(t)^{\mathsf{H}} P B(t))^{-1}$ as a convergence series, which can be used to verify continuous differentiability.

where $N(t)$ is defined in the lemma. Moreover, since $A(t)$ is stabilizable, $\mathcal{T}_{A(t)}$ is a bounded linear operator with bounded inverse (Lemma D.16). Hence, the implicit function theorem (Lemma D.17) shows that $P(t)$ admits a Frechet Derivative $P'(t)$, and its derivative must satisfy $\mathcal{T}_{A(t)+B(t)K(t)}[\delta P]+N(t) = 0$. Consequently, by Lemma D.16, $P'(t)$ is given by Eq. (D.8). □

Lemma D.20 (Traces of Derivatives on $\mathscr{S}_{\mathrm{op}}$)**.** *Let $P(t) : [0,1] \to \mathscr{S}_{\mathrm{op}}$ be a continuously Frechet-differentiable curve. Moreover, let Γ be a fixed trace-class operator on $\mathcal{H}_{\mathbf{x}}$. Then,*

(a) The function $f(t) = \mathrm{tr}\left[\Gamma P(t)\right]$ is a continously differentiable function, with derivative $f'(t) = \mathrm{tr}\left[\Gamma P'(t)\right]$.

(b) If Γ is also PSD,

$$\left\|\Gamma^{1/2}(P(1) - P(0))\Gamma^{1/2}\right\|_{\mathsf{tr}} \le \max_{t\in[0,1]} \left\|\Gamma^{1/2}P'(t)\Gamma^{1/2}\right\|_{\mathsf{tr}}.$$

(c) Similarly, if Γ is PSD,

$$\left\|\Gamma^{1/2}(P(1) - P(0))\Gamma^{1/2}\right\|_{\mathsf{HS}} \le \max_{t\in[0,1]} \left\|\Gamma^{1/2}P'(t)\Gamma^{1/2}\right\|_{\mathsf{HS}}.$$

Proof. For any trace-class Γ, the mapping $P \mapsto \mathrm{tr}\left[\Gamma P\right]$ is a bounded linear functional on the space $\mathscr{S}_{\mathrm{op}}$. Hence, the map commutes with Frechet derivatives, and thus part (a) follows.

To prove part (b), write $\left\|\Gamma^{1/2}(P(1) - P(0))\Gamma^{1/2}\right\|_{\mathsf{tr}} = \max_{X:\|X\|=1} \mathrm{tr}(X\Gamma^{1/2}(P(1)-P(0)\Gamma^{1/2}) = \mathrm{tr}(\tilde{\Gamma}(P(1)-P(0)))$, where $\tilde{\Gamma} = \Gamma^{1/2}X\Gamma^{1/2}$. One can check that $\tilde{\Gamma}$ is also trace class (see, e.g. Dragomir [2014]) and hence $f(t) = \mathrm{tr}\left[\tilde{\Gamma}P(t)\right]$ is a continuously differentiable function by the first part of the lemma, with derivative $f'(t) = \mathrm{tr}\left[\tilde{\Gamma}P'(t)\right]$. Thus, from the mean-value theorem, $\mathrm{tr}(\tilde{\Gamma}(P(1) - P(0))) = f'(t) = \mathrm{tr}\left[\tilde{\Gamma}P'(t)\right]$ for some $t \in [0,1]$. Thus,

$$\begin{aligned}\left\|\Gamma^{1/2}(P(1) - P(0))\Gamma^{1/2}\right\|_{\mathsf{tr}} &= \max_{X:\|X\|_{\mathrm{op}}=1} \mathrm{tr}\left[\Gamma^{1/2}X\Gamma^{1/2}(P(1) - P(0))\right] \\ &\le \max_{X:\|X\|_{\mathrm{op}}=1} \max_{t\in[0,1]} \mathrm{tr}\left[\Gamma^{1/2}X\Gamma^{1/2}P'(t)\right].\end{aligned}$$

Swapping the maxima and rearranging the trace, we see that the resulting term is just

$$\max_{X:\|X\|_{\mathrm{op}}=1} \max_{t\in[0,1]} \mathrm{tr}\left[X\Gamma^{1/2}P'(t)\Gamma^{1/2}\right] = \max_{t\in[0,1]} \left\|\Gamma^{1/2}P'(t)\Gamma^{1/2}\right\|_{\mathsf{tr}}.$$

The proof of part (c) is nearly identical, except that the constraint on the variational parameter X is strengthed to $\|X\|_{\mathsf{HS}} \le 1$. □

Part II

Estimation Rates

E Estimating System Operators

In this section, we prove estimation rates for the system operators $A_\star, B_\star$ estimated via ridge regression on data collected under a single trajectory. In particular, throughout this section, as per our definition of the dynamics in Eq. (1.1), we assume that data $\{(\mathbf{x}_t, \mathbf{u}_t,)\}_{t=1}^{T+1}$ is generated according to,

$$\mathbf{x}_{t+1} = A_{\mathrm{cl},\star}\mathbf{x}_t + B_\star \mathbf{u}_t + \mathbf{w}_t,$$

where $\mathbf{u}_t = K_0 \mathbf{x}_t + \mathbf{v}_t$. Here, K_0 is a stabilizing controller, $\mathbf{v}_t \overset{\text{i.i.d}}{\sim} \mathcal{N}(0, \sigma_{\mathbf{u}}^2 I)$, and $\mathbf{w}_T \overset{\text{i.i.d}}{\sim} \mathcal{N}(0, \Sigma_{\mathbf{w}})$. As before, we let $\Sigma_{\mathbf{x},0} := \Sigma_\star(K_0, \sigma_{\mathbf{u}}^2) = \lim_{t\to} \mathbb{E}[\mathbf{x}_t \otimes \mathbf{x}_t] = \mathsf{dlyap}\left((A_\star + B_\star K_0)^{\mathsf{H}},\ B_\star B_\star^{\mathsf{H}} \sigma_{\mathbf{u}}^2 + \Sigma_{\mathbf{w}}\right)$ be the stationary state covariance when inputs are chosen according to the exploratory policy above. We define $A_{\mathrm{cl},\star} := A_\star + B_\star K_0$ and estimate $A_{\mathrm{cl},\star}, B_\star$ via two separate regressions:

$$\widehat{A}_{\mathrm{cl}} := \arg\min_{A_{\mathrm{cl}}} \frac{1}{T}\sum_{t=1}^{T} \frac{1}{2}\|\mathbf{x}_{t+1} - A_{\mathrm{cl}}\mathbf{x}_t\|^2 + \frac{\lambda}{2}\|A_{\mathrm{cl}}\|_{\mathsf{HS}}^2 \tag{E.1}$$

$$\widehat{B} := \arg\min_{B} \frac{1}{T}\sum_{t=1}^{T} \frac{1}{2}\|\mathbf{x}_{t+1} - B\mathbf{v}_t\|^2. \tag{E.2}$$

The following lemma provides closed form expressions for the estimation error.

Lemma E.1. *Let $\widehat{A}_{\mathrm{cl}}$ and $\widehat{B}$ be defined as in Eq. (E.1) and Eq. (E.2), then*

$$A_{\mathrm{cl},\star} - \widehat{A}_{\mathrm{cl}} = \lambda A_{\mathrm{cl},\star}\left(\frac{1}{T}\sum_{t=1}^{T}\mathbf{x}_t \otimes \mathbf{x}_t + \lambda I\right)^{-1} - \left(\frac{1}{T}\sum_{t=1}^{T}(B_\star \mathbf{v}_t + \mathbf{w}_t)\otimes \mathbf{x}_t\right)\left(\frac{1}{T}\sum_{t=1}^{T}\mathbf{x}_t \otimes \mathbf{x}_t + \lambda I\right)^{-1}$$

$$B_\star - \widehat{B} = -\left(\sum_{t=1}^{T}(A_{\mathrm{cl},\star}\mathbf{x}_t + \mathbf{w}_t)\otimes \mathbf{v}_t\right)\left(\sum_{t=1}^{T}\mathbf{v}_t \otimes \mathbf{v}_t\right)^{-1}.$$

Proof. For the first statement, by taking the first order optimality conditions for the optimization problem in (E.1), we have that,

$$\widehat{A}_{\mathrm{cl}} = \left(A_{\mathrm{cl},\star}\left(\frac{1}{T}\sum_{t=1}^{T}\mathbf{x}_t \otimes \mathbf{x}_t\right) + \frac{1}{T}\sum_{t=1}^{T}(B_\star \mathbf{v}_t + \mathbf{w}_t)\otimes \mathbf{x}_t\right)\left(\frac{1}{T}\sum_{t=1}^{T}\mathbf{x}_t \otimes \mathbf{x}_t + \lambda I\right)^{-1}.$$

Subtracting out the following quantity from both sides,

$$A_{\mathrm{cl},\star}\left(\frac{1}{T}\sum_{t=1}^{T}\mathbf{x}_t \otimes \mathbf{x}_t + \lambda I\right)\left(\frac{1}{T}\sum_{t=1}^{T}\mathbf{x}_t \otimes \mathbf{x}_t + \lambda I\right)^{-1}$$

and multiplying by -1 we get the first identity. The second follows directly from computing the optimality conditions for (E.2). $\square$

E.1 Estimating $B_\star$

We start by presenting the estimation guarantees for $B_\star$ since they are significantly easier to prove than those for $A_\star$. This is because the covariates $\mathbf{v}_t$ are independent and finite dimensional.

Proposition E.2 (B estimation)**.** *Let $T \gtrsim \max\left\{d_u \log(d_u), \, d_u \log\left(\frac{1}{\delta}\right)\right\}$ then with probability $1-\delta$,*

$$\left\|B_\star - \widehat{B}\right\|_{\mathsf{HS}}^2 \lesssim \frac{d_u\left(\mathrm{tr}\left[\Sigma_{\mathbf{x},0}\right] + \|\Sigma_{\mathbf{x},0}\|_{\mathsf{HS}} \log\left(\frac{d_u}{\delta}\right)\right)}{\sigma_{\mathbf{u}}^2 T}.$$

Proof. Using Lemma E.1 we know that the error in estimating $B_\star$ has the following form, which we can upper bound as follows,

$$\left\|\sum_{t=1}^T \left(A_{\mathrm{cl}_\star}\mathbf{x}_t + \mathbf{w}_t\right) \otimes \mathbf{v}_t \left(\sum_{t=1}^T \mathbf{v}_t \otimes \mathbf{v}_t\right)^{-1}\right\|_{\mathsf{HS}}^2 \leq \frac{1}{\lambda_{\min}\left(\sum_{t=1}^T \mathbf{v}_t \otimes \mathbf{v}_t\right)^2} \left\|\sum_{t=1}^T \left(A_{\mathrm{cl}_\star}\mathbf{x}_t + \mathbf{w}_t\right) \otimes \mathbf{v}_t\right\|_{\mathsf{HS}}^2.$$

Upper bounding Hilbert-Schmidt norm Focusing on the Hilbert-Schmidt norm term, we let $\mathbf{e}_1 \dots \mathbf{e}_{d_u}$ be the standard basis for $\mathbb{R}^{d_u}$. Then, we can expand the Hilbert Schmidt norm as,

$$\left\|\sum_{t=1}^T \left(A_{\mathrm{cl}_\star}\mathbf{x}_t + \mathbf{w}_t\right) \otimes \mathbf{v}_t\right\|_{\mathsf{HS}}^2 = \sum_{i=1}^{d_u} \underbrace{\left\| \sum_{t=1}^T \left(A_{\mathrm{cl}_\star}\mathbf{x}_t + \mathbf{w}_t\right) \langle \mathbf{v}_t, \mathbf{e}_i \rangle \right\|^2}_{:=E_i}. \tag{E.3}$$

Now, we deal with each E_i individually. To do so, we notice that $\sum_{t=1}^T \left(A_{\mathrm{cl}_\star}\mathbf{x}_t + \mathbf{w}_t\right) \langle \mathbf{v}_t, \mathbf{e}_i \rangle$ is a mean zero gaussian vector in $\mathcal{H}_{\mathbf{x}}$ with covariance operator equal to,

$$\begin{aligned}\mathbb{E}\left[\sum_{j,k=1}^T \left(A_{\mathrm{cl}_\star}\mathbf{x}_j + \mathbf{w}_j\right) \otimes \left(A_{\mathrm{cl}_\star}\mathbf{x}_k + \mathbf{w}_k\right)^{\mathsf{H}} \langle \mathbf{v}_j, \mathbf{e}_i \rangle \langle \mathbf{v}_k, \mathbf{e}_i \rangle\right] &= \sum_{t=1}^T \left(A_{\mathrm{cl}_\star}\mathbb{E}[\mathbf{x}_t \otimes \mathbf{x}_t] A_{\mathrm{cl}_\star}^{\mathsf{H}} + \Sigma_{\mathbf{w}}\right)\sigma_{\mathbf{u}}^2 \\ &\preceq \sigma_{\mathbf{u}}^2 T \cdot \Sigma_{\mathbf{x},0}.\end{aligned}$$

The first equality follows from the fact that for $j \neq k$, $\langle \mathbf{v}_j, \mathbf{e}_i \rangle$ and $\langle \mathbf{v}_k, \mathbf{e}_i \rangle$ are independent and both mean zero. Therefore, only the diagonal terms remain and by definition of $\mathbf{v}_t$, $\mathbb{E}\langle \mathbf{v}_t, \mathbf{e}_i \rangle^2 = \sigma_{\mathbf{u}}^2$ for all i and t. The upper bound in the second line is a consequence of the fact that, by definition of the dynamics and dlyap, $\mathbb{E}\left[\mathbf{x}_t \otimes \mathbf{x}_t\right] \preceq \Sigma_{\mathbf{x},0}$. Moreover, since $\Sigma_{\mathbf{x},0}$ is the solution to a Lyapunov equation,

$$A_{\mathrm{cl}_\star}\Sigma_{\mathbf{x},0}A_{\mathrm{cl}_\star}^{\mathsf{H}} = \Sigma_{\mathbf{x},0} - \Sigma_{\mathbf{w}} - \sigma_{\mathbf{u}}^2 B_\star B_\star^{\mathsf{H}},$$

the $\Sigma_{\mathbf{w}}$ cancel out. Now, by applying the Hanson-Wright inequality (Lemma E.8), with probability $1-\delta_i$, E_i is upper bounded by,

$$2T\sigma_{\mathbf{u}}^2 \cdot \mathrm{tr}\left[\Sigma_{\mathbf{x},0}\right] + 5T\sigma_{\mathbf{u}}^2 \cdot \log(1/\delta_i) \left\|\Sigma_{\mathbf{x},0}\right\|_{\mathsf{HS}}.$$

Letting $\delta_i = \frac{1}{2d_u}\delta$, we get that with probability, $1-\delta/2$, the expression in (E.3) is less than or equal to

$$5d_u T\sigma_{\mathbf{u}}^2 \cdot \left(\mathrm{tr}\left[\Sigma_{\mathbf{x},0}\right] + \log\left(\frac{2d_u}{\delta}\right)\left\|\Sigma_{\mathbf{x},0}\right\|_{\mathsf{HS}}\right).$$

Lower bounding minimum eigenvalue To finish off the proof, we lower bound the minimum eigenvalue of $\sum_{t=1}^T \mathbf{v}_t \otimes \mathbf{v}_t$. Recall that $\mathbf{v}_t \overset{\text{i.i.d}}{\sim} \mathcal{N}(0, \sigma_{\mathbf{u}}^2 I)$ for all t. Hence, by Lemma E.9 (and our lower bound on T), with probability $1-\delta/2$,

$$\lambda_{\min}\left(\sum_{i=1}^n \mathbf{v}_i \otimes \mathbf{v}_i\right)^2 \geq (9/1600)^2 T^2 \sigma_{\mathbf{u}}^4.$$

Here, we have used that M, as defined in Lemma E.9, is upper bounded by $\sigma_{\mathbf{u}}^2 d_u$ and that $\sigma_{\min}(\sigma_{\mathbf{u}}^2 I) = \sigma_{\mathbf{u}}^2$.

Wrapping up Putting everything together, with probability $1-\delta$, for T larger than the stated threshold,

$$\left\|B_\star - \widehat{B}\right\|_{\mathsf{HS}}^2 \lesssim \frac{d_u \left(\operatorname{tr}\left[\Sigma_{\mathbf{x},0}\right] + \log\left(\frac{d_u}{\delta}\right) \|\Sigma_{\mathbf{x},0}\|_{\mathsf{HS}}\right)}{\sigma_{\mathbf{u}}^2 T}.$$

□

E.2 Estimating $A_\star$

For this subsection, we define the following the following quantities to simplify our notation,

$$\begin{aligned}
\widehat{\Sigma}_{\mathbf{x},0} &:= \frac{1}{T}\sum_{t=1}^{T} \mathbf{x}_t \otimes \mathbf{x}_t \\
V_\lambda &:= \{\mathbf{q} \in \mathcal{H}_{\mathbf{x}} : \langle \mathbf{q}, \Sigma_{\mathbf{x},0}\mathbf{q}\rangle \ \geq \ \lambda\} \\
t_0 &:= c_0 \|P_0\|_{\mathrm{op}} \log_+\left(\lambda^{-1}\|P_0\|_{\mathrm{op}}^2 \|\Sigma_0\|_{\mathrm{op}}\right),
\end{aligned}$$

where $P_0 := P_\infty(K_0; A_\star, B_\star)$ and c_0 is a universal constant.

Let S_λ denote the orthogonal projection operator onto the subspace V_λ. Similarly, let $\overline{S}_\lambda$ denote the projection operator onto the orthogonal complement of V_λ. Recall our definition of $\Sigma_0 := B_\star B_\star^{\mathsf{H}} \sigma_{\mathbf{u}}^2 + \Sigma_{\mathbf{w}}$. Lastly, we will make extensive reference to d_λ and $\mathcal{C}_{\mathsf{tail},\lambda}$ as defined in Theorem 3.1. In particular, if we let $(\sigma_j)_{j=1}^\infty = (\sigma_j(\Sigma_{\mathbf{x},0}))_{j=1}^\infty$ be the eigenvalues of $\Sigma_{\mathbf{x},0}$ then,

$$d_\lambda := |\{\sigma_j : \sigma_j \ \geq \ \lambda\}|, \quad \mathcal{C}_{\mathsf{tail},\lambda} := \tfrac{1}{\lambda}\textstyle\sum_{j>d_\lambda} \sigma_j.$$

Lastly, recall $A_\star := A_\star + B_\star K_0$.

Proposition E.3 (A estimation)**.** *Let T be such that*

$$T \gtrsim \max\left\{d_\lambda \log_+\left(\operatorname{tr}\left[\Sigma_{\mathbf{x},0}\right]\lambda^{-1}\right),\ d_\lambda \log_+(1/\delta)\right\} + \|P_0\|_{\mathrm{op}} \log_+\left(\lambda^{-1}\|P_0\|_{\mathrm{op}}^2\|\Sigma_0\|_{\mathrm{op}}\right).$$

Then, with probability $1-\delta$,

$$\left\|(\widehat{A}_{\mathrm{cl}} - A_{\mathrm{cl}_\star})\Sigma_{\mathbf{x},0}^{1/2}\right\|_{\mathsf{HS}}^2 \lesssim \lambda \|A_{\mathrm{cl}_\star}\|_{\mathsf{HS}}^2 + \frac{W_{\mathrm{tr}}}{T}(d_\lambda + \mathcal{C}_{\mathsf{tail},\lambda}) \log_+\left(\frac{\operatorname{tr}\left[\Sigma_{\mathbf{x},0}\right]}{\delta\lambda}\right).$$

Furthermore, the above guarantee holds for any $\overline{A_{\mathrm{cl}}}$ equal to

$$\overline{A_{\mathrm{cl}}} := \operatorname*{arg\,min}_{A_{\mathrm{cl}} \in \mathcal{B}} \left\langle (A_{\mathrm{cl}} - \widehat{A}_{\mathrm{cl}}), (\widehat{\Sigma}_{\mathbf{x},0} + \lambda I)(A_{\mathrm{cl}} - \widehat{A}_{\mathrm{cl}})\right\rangle,$$

where $\mathcal{B}$ is a compact and convex set containing $A_{\mathrm{cl}_\star}$.

Proof. Using the identity from Lemma E.1,

$$\begin{aligned}
(A_{\mathrm{cl}_\star} - \widehat{A}_{\mathrm{cl}})\left(\frac{1}{T}\sum_{t=1}^{T}\mathbf{x}_t \otimes \mathbf{x}_t + \lambda I\right)^{1/2} &= \lambda A_{\mathrm{cl}_\star}\left(\frac{1}{T}\sum_{t=1}^{T}\mathbf{x}_t \otimes \mathbf{x}_t + \lambda I\right)^{-1/2} \\
&- \left(\frac{1}{T}\sum_{t=1}^{T}(B_\star \mathbf{v}_t + \mathbf{w}_t)\otimes \mathbf{x}_t\right)\left(\frac{1}{T}\sum_{t=1}^{T}\mathbf{x}_t \otimes \mathbf{x}_t + \lambda I\right)^{-1/2}.
\end{aligned}$$

Taking the Hilbert-Schmidt norm of both sides, we get that

$$\left\|(A_{\mathrm{cl}_\star} - \widehat{A}_{\mathrm{cl}})\left(\frac{1}{T}\sum_{t=1}^{T}\mathbf{x}_t \otimes \mathbf{x}_t + \lambda I\right)^{1/2}\right\|_{\mathsf{HS}}^2$$

is less than or equal to,

$$2\underbrace{\left\|\lambda A_{\mathrm{cl}_\star}\left(\frac{1}{T}\sum_{t=1}^T \mathbf{x}_t\otimes\mathbf{x}_t+\lambda I\right)^{-1/2}\right\|_{\mathsf{HS}}^2}_{N_1}+2\underbrace{\left\|\left(\frac{1}{T}\sum_{t=1}^T(B_\star\mathbf{v}_t+\mathbf{w}_t)\otimes\mathbf{x}_t\right)\left(\frac{1}{T}\sum_{t=1}^T\mathbf{x}_t\otimes\mathbf{x}_t+\lambda I\right)^{-1/2}\right\|_{\mathsf{HS}}^2}_{N_2}.$$

Since projections onto convex sets are non-expansive, incorporating a projection step as in the definition of $\overline{A_{\mathrm{cl}}}$, doesn't change the above guarantee since

$$\left\|(A_{\mathrm{cl}_\star}-\overline{A_{\mathrm{cl}}})\left(\widehat{\Sigma}_{\mathbf{x},0}+\lambda I\right)^{1/2}\right\|_{\mathsf{HS}}^2 \;\leq\; \left\|(A_{\mathrm{cl}_\star}-\widehat{A}_{\mathrm{cl}})\left(\widehat{\Sigma}_{\mathbf{x},0}+\lambda I\right)^{1/2}\right\|_{\mathsf{HS}}^2.$$

Bounding the bias Next, we bound each of these terms separately. For the first, we have that,

$$N_1 \;\leq\; \lambda\left\|A_{\mathrm{cl}_\star}\right\|_{\mathsf{HS}}^2.$$

Bounding the noise For the second term, we apply Lemma E.4, to get that with probability $1-\delta/2$,

$$\begin{aligned}N_2 \;&\leq\; \frac{1}{T}\left\|\left(\sum_{t=1}^T(B_\star\mathbf{v}_t+\mathbf{w}_t)\otimes\mathbf{x}_t\right)\left(\sum_{t=1}^T\mathbf{x}_t\otimes\mathbf{x}_t+\lambda I\right)^{-1/2}\right\|_{\mathsf{HS}}^2\\ &\lesssim \frac{1}{T}W_{\mathrm{tr}}\cdot(d_\lambda+\mathcal{C}_{\mathsf{tail},\lambda})\log_+\left(\frac{\mathrm{tr}\left[\Sigma_{\mathbf{x},0}\right]}{\delta\lambda}\right).\end{aligned}$$

Therefore, with probability $1-\delta/2$,

$$\left\|(\widehat{A}_{\mathrm{cl}}-A_{\mathrm{cl}_\star})(\widehat{\Sigma}_{\mathbf{x},0}+\lambda I)^{1/2}\right\|_{\mathsf{HS}}^2\lesssim\lambda\left\|A_{\mathrm{cl}_\star}\right\|_{\mathsf{HS}}^2+\frac{1}{T}W_{\mathrm{tr}}\cdot(d_\lambda+\mathcal{C}_{\mathsf{tail},\lambda})\log_+\left(\frac{\mathrm{tr}\left[\Sigma_{\mathbf{x},0}\right]}{\delta\lambda}\right). \tag{E.4}$$

Now, since we have chosen T to be large enough, we apply Proposition E.6 to get that with probability $1-\delta/2$

$$\Sigma_{\mathbf{x},0}\preceq c\cdot(\widehat{\Sigma}_{\mathbf{x},0}+\lambda I)$$

where c is a universal constant. This finishes the proof since,

$$\begin{aligned}\left\|(\widehat{A}_{\mathrm{cl}}-A_{\mathrm{cl}_\star})\Sigma_{\mathbf{x},0}^{1/2}\right\|_{\mathsf{HS}}^2&=\mathrm{tr}\left[(\widehat{A}_{\mathrm{cl}}-A_{\mathrm{cl}_\star})\Sigma_{\mathbf{x},0}(\widehat{A}_{\mathrm{cl}}-A_{\mathrm{cl}_\star})\right]\\ &\leq\; c\cdot\mathrm{tr}\left[(\widehat{A}_{\mathrm{cl}}-A_{\mathrm{cl}_\star})(\widehat{\Sigma}_{\mathbf{x},0}+\lambda I)(\widehat{A}_{\mathrm{cl}}-A_{\mathrm{cl}_\star})\right]\\ &=c\cdot\left\|(\widehat{A}_{\mathrm{cl}}-A_{\mathrm{cl}_\star})(\widehat{\Sigma}_{\mathbf{x},0}+\lambda I)^{1/2}\right\|_{\mathsf{HS}}^2.\end{aligned}$$

The final result then follows by applying a union bound and combining this last inequality with (E.4). □

Lemma E.4. *The following inequality holds with probability* $1-\delta$,

$$\left\|\sum_{t=1}^T(B_\star\mathbf{v}_t+\mathbf{w}_t)\otimes\mathbf{x}_t\left(\sum_{t=1}^T\mathbf{x}_t\otimes\mathbf{x}_t+\lambda T\cdot I\right)^{-1/2}\right\|_{\mathsf{HS}}^2\lesssim W_{\mathrm{tr}}\cdot\log_+\left(\frac{\mathrm{tr}\left[\Sigma_{\mathbf{x},0}\right]}{\delta\lambda}\right)(d_\lambda+\mathcal{C}_{\mathsf{tail},\lambda})$$

Proof. Define $\mathbf{z}_t := B_\star \mathbf{v}_t + \mathbf{w}_t$. Note that $\mathbf{z}_t \overset{\text{i.i.d}}{\sim} \mathcal{N}(0, \Sigma_0)$, where Σ_0 is again defined as,

$$\Sigma_0 = B_\star B_\star^{\mathsf{H}} \sigma_{\mathbf{u}}^2 + \Sigma_{\mathbf{w}}.$$

Now, let $\sum_{j=1}^\infty \mathbf{q}_j \otimes \mathbf{q}_j \cdot \sigma_j$ be the eigendecomposition of this linear operator Σ_0 where the $\mathbf{q}_j$ form an orthonormal basis of $\mathcal{H}_{\mathbf{x}}$. By definition of the Hilbert-Schmidt norm,

$$\left\| \sum_{t=1}^T \mathbf{z}_t \otimes \mathbf{x}_t \left(\sum_{t=1}^T \mathbf{x}_t \otimes \mathbf{x}_t + \lambda T \cdot I \right)^{-1/2} \right\|_{\mathsf{HS}}^2 = \sum_{j=1}^\infty \underbrace{\| \left(\sum_{t=1}^T \mathbf{x}_t \otimes \mathbf{x}_t + \lambda T \cdot I \right)^{-1/2} \sum_{t=1}^T \mathbf{x}_t \langle \mathbf{z}_t, \mathbf{q}_j \rangle \|^2}_{:= E_j}. \tag{E.5}$$

Due to our choice of basis, we have that $\langle \mathbf{z}_t, \mathbf{q}_j \rangle$ is a zero-mean sub-Gaussian random variable with sub-Gaussian parameter σ_j. In particular, since the $\mathbf{z}_t$ are drawn i.i.d from $\mathcal{N}(0, \Sigma_0)$, we have that for all $\gamma \in \mathbb{R}$,

$$\mathbb{E} \exp(\gamma \langle \mathbf{q}_j, \mathbf{z}_t \rangle) = \exp\left(\frac{\gamma^2}{2} \langle \Sigma_0 \mathbf{q}_j, \mathbf{q}_j \rangle \right) = \exp\left(\frac{\gamma^2}{2} \sigma_j \right).$$

Self-normalized inequality Using this decompostion, we can bound each E_j via a self-normalized martingale bound for vectors in Hilbert space. In particular by Lemma E.10, with probability $1 - \delta_j$,

$$\begin{aligned} E_j \;&\leq\; 2\sigma_j \log \left(\frac{\det\left(I + \frac{1}{\lambda T} \sum_{t=1}^T \mathbf{x}_t \otimes \mathbf{x}_t \right)^{1/2}}{\delta_j} \right) \\ &= \sigma_j \log\det \left(I + \frac{1}{\lambda T} \sum_{t=1}^T \mathbf{x}_t \otimes \mathbf{x}_t \right) + 2\sigma_j \log(1/\delta_j). \end{aligned}$$

Setting $\delta_j = \frac{3}{\pi^2} j^{-2} \cdot \delta$, with probability $1 - \delta_j$, E_j has the following upper bound,

$$E_j \;\leq\; \sigma_j \log\det \left(I + \frac{1}{\lambda T} \sum_{t=1}^T \mathbf{x}_t \otimes \mathbf{x}_t \right) + 2\sigma_j \log\left(\frac{\pi^2}{3\delta} \right) + 4\sigma_j \log(j).$$

Using a union bound and summing up over all j, we get that with probability $1 - \sum_{j=1}^\infty \delta_j = 1 - \delta/2$,

$$\begin{aligned} \left\| \sum_{t=1}^T \mathbf{z}_t \otimes \mathbf{x}_t \left(\sum_{t=1}^T \mathbf{x}_t \otimes \mathbf{x}_t + \lambda T \cdot I \right)^{-1/2} \right\|_{\mathsf{HS}}^2 \;&\leq\; \left(\log\det \left(I + \frac{1}{\lambda T} \sum_{t=1}^T \mathbf{x}_t \otimes \mathbf{x}_t \right) + 2 \log\left(\frac{\pi^2}{3\delta} \right) \right) \sum_{j=1}^\infty \sigma_j \\ &\quad + 4 \sum_{j=1}^\infty \sigma_j \log(j). \end{aligned}$$

Bounding log determinant & simplifying Lastly, we can apply Lemma E.5, to conclude that with probability $1 - \delta/2$,

$$\log\det \left(I + \frac{1}{\lambda T} \sum_{t=1}^T \mathbf{x}_t \otimes \mathbf{x}_t \right) \lesssim d_\lambda \log\left(1 + \operatorname{tr}\left[\Sigma_{\mathbf{x},0} \right] \frac{2}{\delta \lambda} \right) + \lambda^{-1} \log_+(1/\delta) \cdot \operatorname{tr}\left[\overline{S}_\lambda \Sigma_{\mathbf{x},0} \overline{S}_\lambda \right].$$

Now using the fact that $\sum_{j=1}^\infty \sigma_j = \operatorname{tr}\left[B_\star B_\star^{\mathsf{H}} \sigma_{\mathbf{u}}^2 + \Sigma_{\mathbf{w}} \right] = \operatorname{tr}\left[\Sigma_0 \right]$, the target quantity is less than or equal to a universal constant times:

$$\operatorname{tr}\left[\Sigma_0 \right] \left(d_\lambda \log\left(1 + \operatorname{tr}\left[\Sigma_{\mathbf{x},0} \right] \frac{2}{\delta \lambda} \right) + \lambda^{-1} \log_+(1/\delta) \cdot \operatorname{tr}\left[\overline{S}_\lambda \Sigma_{\mathbf{x},0} \overline{S}_\lambda \right] + \log\left(\frac{\pi^2}{3\delta} \right) \right) + W_{\mathrm{tr}}.$$

Furthermore, by definition of W_{tr}, $\mathrm{tr}\left[\Sigma_0\right] \;\le\; W_{\mathrm{tr}}$. A short calculation shows this above satisfies,

$$\lesssim W_{\mathrm{tr}} \cdot \log_+\left(\frac{\mathrm{tr}\left[\Sigma_{\mathbf{x},0}\right]}{\delta\lambda}\right)\left(d_\lambda + \mathcal{C}_{\mathsf{tail},\lambda}\right).$$

□

Lemma E.5. *With probability* $1-\delta$*,*

$$\log\left(\det\left(I + \frac{1}{\lambda T}\sum_{t=1}^{T}\mathbf{x}_t \otimes \mathbf{x}_t\right)\right) \;\le\; d_\lambda \log\left(1 + \mathrm{tr}\left[\Sigma_{\mathbf{x},0}\right]\frac{2}{\delta\lambda}\right) + 7\lambda^{-1}\log_+(2/\delta)\cdot \mathrm{tr}\left[\overline{S}_\lambda \Sigma_{\mathbf{x},0}\overline{S}_\lambda\right].$$

Proof. Let $\sum_{i=1}^{\infty}\mathbf{q}_i \otimes \mathbf{q}_i \cdot \sigma_i$ be the eigendecomposition of $\frac{1}{\lambda T}\sum_{t=1}^{T}\mathbf{x}_t \otimes \mathbf{x}_t$ and recall the following definitions.

$$V_\lambda = \{\mathbf{z} \in \mathcal{H}_{\mathbf{x}} : \langle \mathbf{z}, \Sigma_{\mathbf{x},0}\mathbf{z}\rangle \;\ge\; \lambda\}$$
$$d_\lambda = \mathsf{dim}(V_\lambda).$$

Partitioning the spectrum We recall that for any self-adjoint, PSD linear operator M, $\det(M)$ is equal to the product $\Pi_{i=1}^{\infty}\sigma_i$ of the eigenvalues $\{\sigma_i\}_{i=1}^{\infty}$ of M. Therefore, $\log\det(M)$ is equal to $\sum \log(\sigma_i)$ and $\log\det(I+M) = \sum \log(1+\sigma_i)$. Using this identity, we now bound our target quantity as,

$$\begin{aligned}
\log\left(\det\left(I + \frac{1}{\lambda T}\sum_{t=1}^{T}\mathbf{x}_t \otimes \mathbf{x}_t\right)\right) &= \sum_{i=1}^{\infty}\log(1+\sigma_i) \\
&= \sum_{i:\mathbf{q}_i \in V_\lambda}\log(1+\sigma_i) + \sum_{i:\mathbf{q}_i \notin V_\lambda}\log(1+\sigma_i) \\
&\le\; d_\lambda \log\left(1 + \sum_{i=1}^{\infty}\sigma_i\right) + \sum_{i:\mathbf{q}_i \notin V_\lambda}\sigma_i \\
&= d_\lambda \log\left(1 + \mathrm{tr}\left[\frac{1}{\lambda T}\sum_{t=1}^{T}\mathbf{x}_t \otimes \mathbf{x}_t\right]\right) + \mathrm{tr}\left[\overline{S}_\lambda \frac{1}{\lambda T}\sum_{t=1}^{T}\mathbf{x}_t \otimes \mathbf{x}_t \overline{S}_\lambda\right].
\end{aligned}$$

To go from the second to the third line, we used the inequality $\log(1+x) \;\le\; x$ for all $x \;\ge\; 0$ and the fact that $\dim(V_\lambda) = d_\lambda$. In the last line, we have used definition of $\overline{S}_\lambda$ as a projection operator. In further detail,

$$\sum_{i:\mathbf{q}_i \in V_\lambda}\log(1+\sigma_i) \;\le\; d_\lambda \log(1+\sigma_1) \;\le\; d_\lambda \log\left(1 + \sum_{i=1}^{\infty}\sigma_i\right).$$

Bounding the top Now, by Markov's inequality, with probability $1-\delta/2$,

$$\mathrm{tr}\left[\frac{1}{\lambda T}\sum_{t=1}^{T}\mathbf{x}_t \otimes \mathbf{x}_t\right] \;\le\; \frac{2\sum_{t=1}^{T}\mathbb{E}[\mathbf{x}_t \otimes \mathbf{x}_t]}{\lambda\delta \cdot T} \;\le\; \mathrm{tr}\left[\Sigma_{\mathbf{x},0}\right]\frac{2}{\delta\lambda}.$$

Bounding the tail Define $\tilde{\mathbf{x}}$ as,

$$\tilde{\mathbf{x}} = \begin{bmatrix}\overline{S}_\lambda \mathbf{x}_1 \\ \overline{S}_\lambda \mathbf{x}_2 \\ \cdots\end{bmatrix}.$$

Then,

$$\mathrm{tr}\left[\overline{S}_\lambda \frac{1}{\lambda T}\sum_{t=1}^{T}\mathbf{x}_t \otimes \mathbf{x}_t \overline{S}_\lambda\right] = \frac{1}{\lambda T}\sum_{t=1}^{T}\|\overline{S}_\lambda \mathbf{x}_t\|^2 = \frac{1}{\lambda T}\|\tilde{\mathbf{x}}\|^2,$$

where the last norm is taken in the relevant Hilbert space. Now, we notice that $\tilde{\mathbf{x}}$ is a zero mean Gaussian. Furthermore, the trace norm of its covariance operator can be upper bounded as follows,

$$\mathbb{E}\mathrm{tr}\left[\tilde{\mathbf{x}} \otimes \tilde{\mathbf{x}}\right] = \sum_{t=1}^{T} \mathbb{E}\left[\mathrm{tr}\left[\overline{S}_\lambda \mathbf{x}_t \otimes \mathbf{x}_t \overline{S}_\lambda\right]\right] \preceq T \cdot \mathrm{tr}\left[\overline{S}_\lambda \Sigma_{\mathbf{x},0} \overline{S}_\lambda\right].$$

Therefore, we can apply the Hanson-Wright inequality (Lemma E.8) to conclude that with probability $1-\delta/2$,

$$\begin{aligned}\mathrm{tr}\left[\overline{S}_\lambda \frac{1}{\lambda T}\sum_{t=1}^{T} \mathbf{x}_t \otimes \mathbf{x}_t \overline{S}_\lambda\right] &\le 2\lambda^{-1}\mathrm{tr}\left[\overline{S}_\lambda \Sigma_{\mathbf{x},0}\overline{S}_\lambda\right] + 5\lambda^{-1}\log(2/\delta)\mathrm{tr}\left[\overline{S}_\lambda \Sigma_{\mathbf{x},0}\overline{S}_\lambda\right] \\ &\le 7\lambda^{-1}\log_+(2/\delta)\mathrm{tr}\left[\overline{S}_\lambda \Sigma_{\mathbf{x},0}\overline{S}_\lambda\right].\end{aligned}$$

Finishing the proof In conclusion, by combining the previous parts, we get that with probability $1-\delta$,

$$\log\left(\det\left(I + \frac{1}{\lambda T}\sum_{t=1}^{T}\mathbf{x}_t \otimes \mathbf{x}_t\right)\right) \le d_\lambda \log\left(1 + \mathrm{tr}\left[\Sigma_{\mathbf{x},0}\right]\frac{2}{\delta\lambda}\right) + 7\lambda^{-1}\log_+(2/\delta)\cdot \mathrm{tr}\left[\overline{S}_\lambda \Sigma_{\mathbf{x},0}\overline{S}_\lambda\right].$$

□

Proposition E.6. *For t_0 and d_λ defined as in introduction to Appendix E.2, if*

$$T - t_0 \gtrsim \max\left\{d_\lambda \log_+\left(\frac{\mathrm{tr}\left[\Sigma_{\mathbf{x},0}\right]}{\lambda}\right),\ d_\lambda \log_+(1/\delta)\right\}$$

then with probability $1-\delta$,

$$\Sigma_{\mathbf{x},0} \preceq c \cdot \left(\widehat{\Sigma}_{\mathbf{x},0} + \lambda I\right)$$

where c is a universal constant.

Proof. Recall our definition of V_λ as the subspace of $\mathcal{H}_{\mathbf{x}}$ corresponding to the directions where the state covariance $\Sigma_{\mathbf{x},0}$ has large eigenvalues,

$$V_\lambda = \{\mathbf{q} : \langle \mathbf{q}, \Sigma_{\mathbf{x},0}\mathbf{q}\rangle \ \ge\ \lambda\}.$$

Furthermore, we review the definitions of S_λ, the orthogonal projection operator onto the subspace V_λ, and $\overline{S}_\lambda$, the orthogonal projection operator on the complement of V_λ. Since $\Sigma_{\mathbf{x},0}$ is a trace class operator, V_λ is a finite dimensional subspace. Therefore, $d_\lambda = \mathsf{dim}(V_\lambda) < \infty$.

For any vector $\mathbf{z} \in \mathcal{H}_{\mathbf{x}}$, we have that,

$$\begin{aligned}\langle \mathbf{z}, \Sigma_{\mathbf{x},0}\mathbf{z}\rangle &\le 2\langle S_\lambda \mathbf{z}, \Sigma_{\mathbf{x},0} S_\lambda \mathbf{z}\rangle + 2\left\langle \overline{S}_\lambda \mathbf{z}, \Sigma_{\mathbf{x},0}\overline{S}_\lambda \mathbf{z}\right\rangle \\ &\le 2\langle S_\lambda \mathbf{z}, \Sigma_{\mathbf{x},0} S_\lambda \mathbf{z}\rangle + 2\lambda.\end{aligned}$$

In order to complete the proof, it suffices to show that,

$$S_\lambda \Sigma_{\mathbf{x},0} S_\lambda \preceq c \cdot S_\lambda \widehat{\Sigma}_{\mathbf{x},0} S_\lambda.$$

In short, we have reduced the proof to showing a PSD upper bound for finite dimensional operators. By Lemma E.7, for $t \ \ge\ t_0$, we have that $S_\lambda \mathbf{x}_t \sim \mathcal{N}(0, \Sigma_t)$ for $\Sigma_t \succeq \frac{\lambda}{2} I$. Therefore, for $T - t_0$ greater than the lower bound in the statement of the proposition, we can apply Lemma E.9, to conclude that with probability $1-\delta$,

$$S_\lambda \Sigma_{\mathbf{x},0}^{1/2} S_\lambda \preceq c \cdot S_\lambda \widehat{\Sigma}_{\mathbf{x},0} S_\lambda.$$

This concludes the proof. □

Lemma E.7. *For $t \geq t_0$, $S_\lambda \mathbb{E}\left[\mathbf{x}_t \otimes \mathbf{x}_t\right] S_\lambda \succeq \frac{\lambda}{2} I$, where $I \in \mathbb{R}^{d_\lambda \times d_\lambda}$.*

Proof. Let $\mathbf{v} \in V_\lambda$ be a unit vector. Then, by definition of V_λ, $\langle \mathbf{v}, \Sigma_{\mathbf{x},0} \mathbf{v} \rangle \geq \lambda$. Furthermore, by properties of the dynamical system,

$$\mathbb{E}\left[\mathbf{x}_t \otimes \mathbf{x}_t\right] = \sum_{j=0}^{t-2} A_{\mathrm{cl}_\star}^j \left(B_\star B_\star \sigma_{\mathbf{u}}^2 + \Sigma_{\mathbf{w}}\right) \left(A_{\mathrm{cl}_\star}^j\right)^{\mathsf{H}}.$$

Before moving on, we recall the form of the steady-state covariance operator $\Sigma_{\mathbf{x},0}$,

$$\Sigma_{\mathbf{x},0} = \mathsf{dlyap}\left(A_{\mathrm{cl}_\star}^{\mathsf{H}},\ B_\star B_\star \sigma_{\mathbf{u}}^2 + \Sigma_{\mathbf{w}}\right) = \sum_{j=0}^{\infty} A_{\mathrm{cl}_\star}^j \left(B_\star B_\star \sigma_{\mathbf{u}}^2 + \Sigma_{\mathbf{w}}\right) \left(A_{\mathrm{cl}_\star}^j\right)^{\mathsf{H}}.$$

By the previous two equations, we have that:

$$\langle \mathbf{v}, \Sigma_{\mathbf{x},0} \mathbf{v} \rangle - \langle \mathbf{v}, \mathbb{E} \mathbf{x}_t \otimes \mathbf{x}_t \mathbf{v} \rangle = \left\langle \mathbf{v}, \sum_{j=t-1}^{\infty} A_{\mathrm{cl}_\star}^j \left(B_\star B_\star \sigma_{\mathbf{u}}^2 + \Sigma_{\mathbf{w}}\right) \left(A_{\mathrm{cl}_\star}^j\right)^{\mathsf{H}} \mathbf{v} \right\rangle.$$

Therefore for any $\mathbf{v} \in V_\lambda$,

$$\begin{aligned}
\langle \mathbf{v}, \mathbb{E} \mathbf{x}_t \otimes \mathbf{x}_t \mathbf{v} \rangle &= \langle \mathbf{v}, \Sigma_{\mathbf{x},0} \mathbf{v} \rangle - \left\langle \mathbf{v}, \sum_{j=t-1}^{\infty} A_{\mathrm{cl}_\star}^j \left(B_\star B_\star \sigma_{\mathbf{u}}^2 + \Sigma_{\mathbf{w}}\right) \left(A_{\mathrm{cl}_\star}^j\right)^{\mathsf{H}} \mathbf{v} \right\rangle \\
&\geq \lambda - \left\langle \mathbf{v}, \sum_{j=t-1}^{\infty} A_{\mathrm{cl}_\star}^j \left(B_\star B_\star \sigma_{\mathbf{u}}^2 + \Sigma_{\mathbf{w}}\right) \left(A_{\mathrm{cl}_\star}^j\right)^{\mathsf{H}} \mathbf{v} \right\rangle.
\end{aligned}$$

To finish the proof, we show that for any $\mathbf{v} \in V_\lambda$, the second term is no smaller than $-\lambda/2$. To do so, we proceed as in Lemma D.11. Recall that $B_\star B_\star \sigma_{\mathbf{u}}^2 + \Sigma_{\mathbf{w}} = \Sigma_0$,

$$\begin{aligned}
\left\langle \mathbf{v}, \sum_{j=t-1}^{\infty} A_{\mathrm{cl}_\star}^j \left(B_\star B_\star \sigma_{\mathbf{u}}^2 + \Sigma_{\mathbf{w}}\right) \left(A_{\mathrm{cl}_\star}^j\right)^{\mathsf{H}} \mathbf{v} \right\rangle &\leq \left\|B_\star B_\star \sigma_{\mathbf{u}}^2 + \Sigma_{\mathbf{w}}\right\|_{\mathrm{op}} \sum_{j=t-1}^{\infty} \left\|A_{\mathrm{cl}_\star}^j (A_{\mathrm{cl}_\star}^{\mathsf{H}})^j\right\|_{\mathrm{op}} \\
&= \|\Sigma_0\|_{\mathrm{op}} \sum_{j=t-1}^{\infty} \left\|(A_{\mathrm{cl}_\star}^{\mathsf{H}})^j A_{\mathrm{cl}_\star}^j\right\|_{\mathrm{op}} \\
&\leq \|\Sigma_0\|_{\mathrm{op}} \sum_{j=t-1}^{\infty} \left\|(A_{\mathrm{cl}_\star}^{\mathsf{H}})^j P_0 A_{\mathrm{cl}_\star}^j\right\|_{\mathrm{op}} \\
&\leq \|\Sigma_0\|_{\mathrm{op}} \|P_0\|_{\mathrm{op}} \sum_{j=t-1}^{\infty} (1 - \|P_0\|_{\mathrm{op}}^{-1})^j \\
&= \|\Sigma_0\|_{\mathrm{op}} \|P_0\|_{\mathrm{op}}^2 (1 - \|P_0\|_{\mathrm{op}}^{-1})^{t-1} \\
&\leq \|\Sigma_0\|_{\mathrm{op}} \|P_0\|_{\mathrm{op}}^2 \exp\left(-\frac{(t-1)}{\|P_0\|_{\mathrm{op}}}\right).
\end{aligned}$$

The fourth inequality follows from applying Lemma D.9. For $t \geq 1 + \lceil \|P_0\|_{\mathrm{op}} \log\left(\frac{2}{\lambda} \|P_0\|_{\mathrm{op}}^2 \|\Sigma_0\|_{\mathrm{op}}\right) \rceil$, the expression above is smaller than $\frac{\lambda}{2}$. Lastly, we note that since $\|P_0\|_{\mathrm{op}} > 1$,

$$t_0 := 3 \|P_0\|_{\mathrm{op}} \log_+\left(\frac{2}{\lambda} \|P_0\|_{\mathrm{op}}^2 \|\Sigma_0\|_{\mathrm{op}}\right) \geq 1 + \left\lceil \|P_0\|_{\mathrm{op}} \log\left(\frac{2}{\lambda} \|P_0\|_{\mathrm{op}}^2 \|\Sigma_0\|_{\mathrm{op}}\right) \right\rceil.$$

□

E.3 Formal Statement and Proof of Proposition 3.1

Proposition 3.1. Assume that Assumption 3 holds, and define $\Delta_r := \frac{1}{16}\mathcal{C}^2_{\mathsf{stable}} - s^2_{r+1}$, $\lambda_{\mathsf{safe}} := c\frac{\Delta_r}{\rho}$, where ρ, r are defined as in Assumption 3 and c is a universal constant. Then, for $T \geq T_{\mathsf{init}}$, and (A_0, B_0) computed as in the WarmStart algorithm, with probability $1-\delta$

$$\max\left\{\|A_0 - A_\star\|_{\mathrm{op}}, \|B_0 - B_\star\|_{\mathrm{op}}\right\} \leq \frac{1}{2}\mathcal{C}_{\mathsf{stable}}.$$

Here, $P_{\mathsf{init}} = P_\infty(K_{\mathsf{init}}; A_\star, B_\star)$, $\Sigma_{\mathbf{x},\mathsf{init}} := \Sigma_\star(K_{\mathsf{init}}, \sigma^2_{\mathbf{u}})$, and T_{init} is equal to a universal constant times the maximum of the following three quantities,

a) $d_{\lambda_{\mathsf{safe}}} \log_+\left(\frac{\mathrm{tr}[\Sigma_{\mathbf{x},\mathsf{init}}]}{\delta\lambda_{\mathsf{safe}}}\right) + \|P_{\mathsf{init}}\|_{\mathrm{op}} \log_+\left(\frac{1}{\lambda_{\mathsf{safe}}}\|P_{\mathsf{init}}\|^2_{\mathrm{op}}\left\|B_\star B_\star^{\mathsf{H}}\sigma^2_{\mathbf{u}} + \Sigma_{\mathbf{w}}\right\|_{\mathrm{op}}\right)$

b) $\frac{W_{\mathrm{tr}}}{\lambda_{\mathsf{safe}}\Delta_r}\log_+\left(\frac{\mathrm{tr}[\Sigma_{\mathbf{x},\mathsf{init}}]}{\delta\lambda_{\mathsf{safe}}}\right)(d_{\lambda_{\mathsf{safe}}} + \mathcal{C}_{\mathsf{tail},\lambda_{\mathsf{safe}}})$

c) $\max\{1, \|K_{\mathsf{init}}\|^2_{\mathrm{op}}\} \cdot \frac{d_u\left(\mathrm{tr}[\Sigma_{\mathbf{x},\mathsf{init}}] + \|\Sigma_{\mathbf{x},\mathsf{init}}\|_{\mathsf{HS}}\log\left(\frac{d_u}{\delta}\right)\right)}{\sigma^2_{\mathbf{u}}\mathcal{C}^2_{\mathsf{stable}}}$.

Proof. The overall proof strategy is to show there exists a threshold such that the error in both $A_\star$ and $B_\star$ is small. We deal with operator error separately. Assume $\widehat{A}_{\mathrm{cl}}, A_0$, and B_0 are computed as in WarmStart (see Appendix F). Furthermore, redefine $A_{\mathrm{cl}_\star} := A_\star + B_\star K_{\mathsf{init}}$.

$B_\star$ recovery This part is simple since we have consistent parameter recovery for $B_\star$. Recalling Proposition E.2, the following statement holds with probability $1-\delta/2$,

$$\|B_\star - B_0\|^2_{\mathsf{HS}} \lesssim \frac{d_u\left(\mathrm{tr}\left[\Sigma_{\mathbf{x},\mathsf{init}}\right] + \|\Sigma_{\mathbf{x},\mathsf{init}}\|_{\mathsf{HS}}\log\left(\frac{d_u}{\delta}\right)\right)}{\sigma^2_{\mathbf{u}}T}.$$

Since $T_{\mathsf{init}} \gtrsim \max\{1, \|K_{\mathsf{init}}\|^2_{\mathrm{op}}\}\frac{d_u\left(\mathrm{tr}[\Sigma_{\mathbf{x},\mathsf{init}}] + \|\Sigma_{\mathbf{x},\mathsf{init}}\|_{\mathsf{HS}}\log\left(\frac{d_u}{\delta}\right)\right)}{\sigma^2_{\mathbf{u}}\mathcal{C}^2_{\mathsf{stable}}}$ we have that with probability $1-\delta/2$,

$$\|B_\star - B_0\|_{\mathsf{HS}} \leq \frac{1}{4\max\{1, \|K_{\mathsf{init}}\|_{\mathrm{op}}\}}\mathcal{C}_{\mathsf{stable}}.$$

$A_\star$ recovery In order to guarantee a close estimate of $A_\star$, we use the error decomposition from Lemma E.1 and show that this quantity is small in operator norm if the alignment condition holds. For the sake of making notation concise, we let,$\widehat{\Sigma}_{\mathbf{x},\mathsf{init}} := \frac{1}{T}\sum_{t=1}^T \mathbf{x}_t \otimes \mathbf{x}_t$ and $A_{\mathrm{cl}_\star} = A_\star + B_\star K_{\mathsf{init}}$. From our earlier error decomposition,

$$\begin{aligned}\left\|\widehat{A}_{\mathrm{cl}} - A_{\mathrm{cl}_\star}\right\|^2_{\mathrm{op}} &= \left\|\lambda A_{\mathrm{cl}_\star}\left(\widehat{\Sigma}_{\mathbf{x},\mathsf{init}} + \lambda I\right)^{-1} - \left(\frac{1}{T}\sum_{t=1}^T (B_\star \mathbf{v}_t + \mathbf{w}_t)\otimes \mathbf{x}_t\right)\left(\widehat{\Sigma}_{\mathbf{x},\mathsf{init}} + \lambda I\right)^{-1}\right\|^2_{\mathrm{op}} \\ &\leq 2\underbrace{\left\|\lambda A_{\mathrm{cl}_\star}\left(\widehat{\Sigma}_{\mathbf{x},\mathsf{init}} + \lambda I\right)^{-1}\right\|^2_{\mathrm{op}}}_{:=N_1} + 2\underbrace{\left\|\left(\frac{1}{T}\sum_{t=1}^T (B_\star \mathbf{v}_t + \mathbf{w}_t)\otimes \mathbf{x}_t\right)\left(\widehat{\Sigma}_{\mathbf{x},\mathsf{init}} + \lambda I\right)^{-1}\right\|^2_{\mathsf{HS}}}_{:=N_2}.\end{aligned}$$

Bounding the noise The bound on N_2 follows a simple application of Lemma E.4.

$$\begin{aligned}\left\|\left(\frac{1}{T}\sum_{t=1}^T (B_\star \mathbf{v}_t + \mathbf{w}_t)\otimes \mathbf{x}_t\right)\left(\widehat{\Sigma}_{\mathbf{x},\mathsf{init}} + \lambda I\right)^{-1}\right\|^2_{\mathsf{HS}} &\leq \frac{1}{\lambda T^2}\left\|\left(\sum_{t=1}^T (B_\star \mathbf{v}_t + \mathbf{w}_t)\otimes \mathbf{x}_t\right)\left(\widehat{\Sigma}_{\mathbf{x},\mathsf{init}} + \lambda I\right)^{-1/2}\right\|^2_{\mathsf{HS}} \\ &\leq \frac{1}{\lambda T}\left\|\left(\sum_{t=1}^T (B_\star \mathbf{v}_t + \mathbf{w}_t)\otimes \mathbf{x}_t\right)\left(T\widehat{\Sigma}_{\mathbf{x},\mathsf{init}} + \lambda T I\right)^{-1/2}\right\|^2_{\mathsf{HS}}.\end{aligned}$$

Applying Lemma E.4, we get that with probability $1-\delta/4$,

$$N_2 \lesssim \frac{W_{\text{tr}}}{\lambda T} \cdot \log_+ \left(\frac{\text{tr}\left[\Sigma_{\mathbf{x},\mathsf{init}}\right]}{\delta\lambda}\right)(d_\lambda + \mathcal{C}_{\mathsf{tail},\lambda}).$$

Bounding the bias The bound on N_1 follows from the alignment condition. In order to apply it, we first perform the following simplification of the bias term.

$$\begin{aligned}\left\|\lambda A_{\text{cl}_\star}\left(\widehat{\Sigma}_{\mathbf{x},\mathsf{init}}+\lambda I\right)^{-1}\right\|_{\text{op}}^2 &\leq \lambda\left\|A_{\text{cl}_\star}\left(\widehat{\Sigma}_{\mathbf{x},\mathsf{init}}+\lambda I\right)^{-1/2}\right\|_{\text{op}}^2 \\ &= \lambda\left\|\left(\widehat{\Sigma}_{\mathbf{x},\mathsf{init}}+\lambda I\right)^{-1/2} A_{\text{cl}_\star}^{\mathsf{H}} A_{\text{cl}_\star}\left(\widehat{\Sigma}_{\mathbf{x},\mathsf{init}}+\lambda I\right)^{-1/2}\right\|_{\text{op}}. \end{aligned} \tag{E.6}$$

As in the alignment condition, we let $U(\Lambda_r + \Lambda_{/r})V^{\mathsf{H}}$ be the SVD of $A_{\text{cl}_\star}$. Therefore, $A_{\text{cl}_\star}^{\mathsf{H}} A_{\text{cl}_\star}$ is equal to $V(\Lambda_r^2 + \Lambda_{/r}^2)V^{\mathsf{H}}$. Applying the triangle inequality, we can then bound (E.6) by,

$$\lambda\left\|\left(\widehat{\Sigma}_{\mathbf{x},\mathsf{init}}+\lambda I\right)^{-1/2} V\Lambda_r^2 V^{\mathsf{H}}\left(\widehat{\Sigma}_{\mathbf{x},\mathsf{init}}+\lambda I\right)^{-1/2}\right\|_{\text{op}} + \lambda\left\|\left(\widehat{\Sigma}_{\mathbf{x},\mathsf{init}}+\lambda I\right)^{-1/2} V\Lambda_{/r}^2 V^{\mathsf{H}}\left(\widehat{\Sigma}_{\mathbf{x},\mathsf{init}}+\lambda I\right)^{-1/2}\right\|_{\text{op}}. \tag{E.7}$$

We can bound the second term above by $\|\Lambda_{/r}^2\|_{\text{op}} = s_{r+1}^2$. Next, we have chosen T_{init} large enough so that $\widehat{\Sigma}_{\mathbf{x},\mathsf{init}} + \lambda I$ is a PSD upper bound on $\Sigma_{\mathbf{x},\mathsf{init}}$ as per Proposition E.6. Together with the alignment condition, we have that with probability $1-\delta/4$,

$$V\Lambda_r^2 V^{\mathsf{H}} \preceq \rho\Sigma_{\mathbf{x},\mathsf{init}} \preceq c\cdot\rho(\widehat{\Sigma}_{\mathbf{x},\mathsf{init}}+\lambda I),$$

for some universal constant c. This implies that,

$$\left(\widehat{\Sigma}_{\mathbf{x},\mathsf{init}}+\lambda I\right)^{-1/2} V\Lambda_r^2 V^{\mathsf{H}}\left(\widehat{\Sigma}_{\mathbf{x},\mathsf{init}}+\lambda I\right)^{-1/2} \preceq c\rho I,$$

and hence the first term in Eq. (E.7) is smaller than $c\cdot\lambda\rho$. Putting everything together, we get that with probability $1-\delta/4$, N_1 is less than or equal to $c\rho\lambda + s_{r+1}^2$. Therefore, with probability $1-\delta/2$,

$$\left\|\widehat{A}_{\text{cl}} - A_{\text{cl}_\star}\right\|_{\text{op}}^2 \leq c_0\frac{W_{\text{tr}}}{\lambda T}\cdot\log_+\left(\frac{\text{tr}\left[\Sigma_{\mathbf{x},\mathsf{init}}\right]}{\delta\lambda}\right)(d_\lambda + \mathcal{C}_{\mathsf{tail},\lambda}) + c_1\rho\lambda + s_{r+1}^2. \tag{E.8}$$

Now, we let $\Delta_r := \frac{1}{16}\mathcal{C}_{\mathsf{stable}}^2 - s_{r+1}^2$ which is strictly great than 0 by Assumption 3. Setting $\lambda = \lambda_{\mathsf{safe}} := c'\frac{\Delta_r}{\rho}$ for some universal constant c', we have that $c_1\rho\lambda \leq \frac{1}{2}\Delta_r$. Furthermore, for T such that,

$$T \gtrsim \frac{W_{\text{tr}}}{\lambda_{\mathsf{safe}}\Delta_r}\log_+\left(\frac{\text{tr}\left[\Sigma_{\mathbf{x},0}\right]}{\delta\lambda_{\mathsf{safe}}}\right)(d_{\lambda_{\mathsf{safe}}} + \mathcal{C}_{\mathsf{tail},\lambda_{\mathsf{safe}}}),$$

the first term in Eq. (E.8) is less than or equal to $\Delta_r/2$ and hence $\|\widehat{A}_{\text{cl}} - A_{\text{cl}_\star}\|_{\text{op}}^2 \leq \frac{1}{16}\mathcal{C}_{\mathsf{stable}}^2$. Computing $A_0 = \widehat{A}_{\text{cl}} - B_0 K_{\mathsf{init}}$, we get that:

$$\begin{aligned}\|A_0 - A_\star\|_{\text{op}} &= \left\|(\widehat{A}_{\text{cl}} - B_0K_{\mathsf{init}}) \pm A_{\text{cl}_\star} - A_\star\right\|_{\text{op}} \\ &\leq \left\|\widehat{A}_{\text{cl}} - A_{\text{cl}_\star}\right\|_{\text{op}} + \|(B_\star - B_0)K_{\mathsf{init}}\|_{\text{op}} \\ &\leq \frac{1}{2}\mathcal{C}_{\mathsf{stable}}.\end{aligned}$$

This concludes the proof. □

E.4 Estimation Lemmas

Lemma E.8 (Theorem 2.6 in Chen and Yang [2021]). *Let* $\mathbf{v}_i \in \mathcal{H}_\mathbf{x}$ *be independent random vectors in a Hilbert space* $\mathcal{H}_\mathbf{x}$ *such that* $\mathbf{v}_i \sim \mathcal{N}(0, \Sigma_i)$ *and* $\Sigma_i \preceq \Sigma$ *for all* i*. Then,*

$$\mathbb{P}\left[\sum_{i=1}^n \|\mathbf{v}_i\|^2 \;\geq\; 2n\mathrm{tr}\left[\Sigma\right] + 5t \left\|\Sigma\right\|_{\mathsf{HS}}\right] \;\leq\; \exp\left(-t\right)$$

Proof. The lemma is a restatement of Theorem 2.6 in Chen and Yang [2021]. Since the $\mathbf{v}_i$ are Gaussians, the inequality immediately following Equation 4.2 in the proof of Theorem 2.6 in Chen and Yang [2021] can be restated as,

$$\mathbb{P}\left[\sum_{i=1}^n \|\mathbf{v}_i\|^2 \;\geq\; n\mathrm{tr}\left[\Sigma\right] + t\right] \;\leq\; \exp\left(-\lambda t + 2n\lambda^2 \left\|\Sigma\right\|_{\mathsf{HS}}^2\right) \text{ for all } 0 \;\leq\; \lambda < \frac{1}{4\left\|\Sigma\right\|_{\mathrm{op}}}.$$

Since $\left\|\Sigma\right\|_{\mathrm{op}} \;\leq\; \left\|\Sigma\right\|_{\mathsf{HS}}$, if we set $\lambda = (5\left\|\Sigma\right\|_{\mathsf{HS}})^{-1}$, we get that

$$\mathbb{P}\left[\sum_{i=1}^n \|\mathbf{v}_i\|^2 \;\geq\; n\mathrm{tr}\left[\Sigma\right] + t\right] \;\leq\; \exp\left(-\frac{t}{5\left\|\Sigma\right\|_{\mathsf{HS}}} + \frac{2}{25}n\right).$$

Lastly, setting $-t' = -\frac{t}{5\|\Sigma\|_{\mathsf{HS}}} + \frac{2}{25}n$,

$$\mathbb{P}\left[\sum_{i=1}^n \|\mathbf{v}_i\|^2 \;\geq\; n\mathrm{tr}\left[\Sigma\right] + 5t' \left\|\Sigma\right\|_{\mathsf{HS}} + \frac{2}{5}n\left\|\Sigma\right\|_{\mathsf{HS}}\right] \;\leq\; \exp\left(-t'\right).$$

The proof follows since $\left\|\Sigma\right\|_{\mathsf{HS}} \;\leq\; \mathrm{tr}\left[\Sigma\right]$. □

Lemma E.9 (Lemma E.4 in Simchowitz and Foster [2020]). *Let* $\mathcal{F}_t$ *be a filtration such that* $\mathbf{z}_t \mid \mathcal{F}_{t-1} \sim \mathcal{N}(0, \Sigma_t)$ *where* $\Sigma_t \in \mathbb{R}^{d\times d}$ *is* $\mathcal{F}_{t-1}$*-measurable, and* $\Sigma_t \succeq \Sigma$*. Furthermore, assume that*

$$\mathbb{E}\,\mathrm{tr}\left[\frac{1}{T}\sum_{t=1}^T \mathbf{z}_t \otimes \mathbf{z}_t\right] \;\leq\; M.$$

Then for,

$$T \;\geq\; 223\max\left\{2d\log(\frac{100}{3}) + d\log\left(\frac{M}{\lambda_{\min}(\Sigma)}\right),\; (d+1)\log\left(\frac{2}{\delta}\right)\right\},$$

with probability $1-\delta$*,*

$$\frac{1}{T}\sum_{t=1}^T \mathbf{z}_t \otimes \mathbf{z}_t \succeq \frac{9}{1600}\Sigma.$$

Lemma E.10 (Corollary 3.6 in Abbasi-Yadkori [2012]). *Let* $(\mathcal{F}_k, k \;\geq\; 1)$ *be a filtration and let* $(\mathbf{m}_k, k \;\geq\; 1)$ *be an* $\mathcal{H}_\mathbf{x}$*-valued stochastic process adapted to* $\mathcal{F}_k$*, and* $(\eta_k, k \;\geq\; 2)$ *be a real valued martingale difference process adapted to* $\mathcal{F}_k$*. Furthermore, assume that* η_k *is conditionally sub-Gaussian in the sense that there exists a* $\sigma > 0$ *such that,* $\mathbb{E}\left[\exp(\eta_k\gamma)\right] \;\leq\; \exp\left(\gamma^2\sigma/2\right)$ *for all* $\gamma \in \mathbb{R}$*. Consider the martingale and operator-valued processes,*

$$S_t := \sum_{k=1}^t \eta_{k+1}\cdot\mathbf{m}_k, \quad V_t := \sum_{k=1}^{t-1}\mathbf{m}_k \otimes \mathbf{m}_k, \quad \overline{V}_t := \lambda I + V_t \text{ for } t \;\geq\; 0.$$

Then, for any $\delta \in (0,1)$ *with probability* $1-\delta$*,*

$$\forall t \;\geq\; 2, \quad \|\overline{V}_t^{-1/2}S_t\| \;\leq\; 2\sigma\log\left(\frac{\det(I + \lambda^{-1}V_t)}{\delta}\right).$$

Part III
Regret Bounds

F Algorithm Descriptions: OnlineCE and WarmStart

Below, we let $\mathcal{B} := \{A_{\mathrm{cl}} : \|A_{\mathrm{cl}} - (A_0 + B_0 K_0)\|_{\mathrm{op}} \le .5\mathcal{C}_{\mathsf{stable}}\}$ denote an operator norm ball around the warm start estimate $A_0 + B_0 K_0$ and let $\widehat{\Sigma}_{\mathbf{x},0} = T^{-1}\sum_{t=1}^{T} \mathbf{x}_t \otimes \mathbf{x}_t$ be the empirical state covariance. The precise description of T_{exp} and λ may be found in the proof of Theorem 3.1.

OnlineCE

Input: Warm start estimates (A_0, B_0), confidence δ, horizon length T

1. Synthesize controller $K_0 = K_\infty(A_0, B_0)$
2. Collect data under exploration policy

   For $t = 1, 2, \dots, T_{\mathsf{exp}}$:

   - Observe state $\mathbf{x}_t$
   - Choose input $\mathbf{u}_t = K_0 \mathbf{x}_t + \mathbf{v}_t$ where $\mathbf{v}_t \sim \mathcal{N}(0, I)$

3. Estimate $B_\star$

$$\widehat{B} = \arg\min_{B} \sum_{t=1}^{T_{\mathsf{exp}}} \frac{1}{2T_{\mathsf{exp}}} \|\mathbf{x}_{t+1} - B\mathbf{v}_t\|_{\mathcal{H}_{\mathbf{x}}}^2$$

4. Estimate $A_\star$

   (a) Compute initial estimate via ridge regression

$$\widetilde{A_{\mathrm{cl}}} := \arg\min_{A_{\mathrm{cl}}} \tfrac{1}{T_{\mathsf{exp}}} \textstyle\sum_{t=1}^{T_{\mathsf{exp}}} \tfrac{1}{2} \|\mathbf{x}_{t+1} - A_{\mathrm{cl}}\mathbf{x}_t\|_{\mathcal{H}_{\mathbf{x}}}^2 + \tfrac{\lambda}{2} \|A_{\mathrm{cl}}\|_{\mathsf{HS}}^2$$

   (b) Project to safe set

$$\widehat{A}_{\mathrm{cl}} := \arg\min_{A_{\mathrm{cl}} \in \mathcal{B}} \left\langle (A_{\mathrm{cl}} - \widetilde{A_{\mathrm{cl}}}), (\widehat{\Sigma}_{\mathbf{x},0} + \lambda I)(A_{\mathrm{cl}} - \widetilde{A_{\mathrm{cl}}}) \right\rangle$$

   (c) Refine estimate

$$\widehat{A} := \widehat{A}_{\mathrm{cl}} - \widehat{B} K_0$$

5. Synthesize certainty equivalence controller $\widehat{K} = K_\infty(\widehat{A}, \widehat{B})$
6. Choose inputs according to $\widehat{K}$ for remainder of horizon.

   For $t = T_{\mathsf{exp}} + 1, \dots, T$:

   - Observe state $\mathbf{x}_t$
   - Choose input $\mathbf{u}_t = \widehat{K}\mathbf{x}_t$

WarmStart

Input: Initial controller K_{init} which stabilizes $(A_\star, B_\star)$.

1. Collect data under exploration policy

   For $t = 1, 2, \ldots, T_{\mathsf{init}}$:

   - Observe state $\mathbf{x}_t$
   - Choose input $\mathbf{u}_t = K_0\mathbf{x}_t + \mathbf{v}_t$ where $\mathbf{v}_t \sim \mathcal{N}(0, I)$

2. Estimate $B_\star$

$$B_0 = \arg\min_B \sum_{t=1}^{T_{\mathsf{exp}}} \frac{1}{2T_{\mathsf{exp}}} \|\mathbf{x}_{t+1} - B\mathbf{v}_t\|_{\mathcal{H}_{\mathbf{x}}}^2$$

3. Estimate $A_\star$

   (a) Compute initial estimate via ridge regression

$$\widehat{A}_{\mathrm{cl}} := \arg\min_{A_{\mathrm{cl}}} \tfrac{1}{T_{\mathsf{exp}}} \textstyle\sum_{t=1}^{T_{\mathsf{exp}}} \tfrac{1}{2} \|\mathbf{x}_{t+1} - A_{\mathrm{cl}}\mathbf{x}_t\|_{\mathcal{H}_{\mathbf{x}}}^2 + \tfrac{\lambda_{\mathsf{safe}}}{2} \|A_{\mathrm{cl}}\|_{\mathsf{HS}}^2$$

   (b) Refine estimate: $A_0 = \widehat{A}_{\mathrm{cl}} - B_0 K_{\mathsf{init}}$.

F.1 Incorporating Data Dependent Conditions for WarmStart

Let $P_0 := P_\infty(A_0, B_0)$ denote the optimal value function of the initial estimate (A_0, B_0). From our perturbation bounds on the solution to the DARE (Proposition C.3), we have that if

$$\varepsilon_{\mathrm{op},0} := \max\{\|A_0 - A_\star\|_{\mathrm{op}}, \|B_0 - B_\star\|_{\mathrm{op}}\} \le \eta/(16(1+\eta)^4 \|P_\infty(A_0, B_0)\|_{\mathrm{op}}^3,$$

for some parameter $\eta \in (0,1)$, then $\|P_\star - P_0\|_{\mathrm{op}} \le \eta \|P_0\|_{\mathrm{op}}$. By choosing η sufficiently small, we can see that that if $\varepsilon_{\mathrm{op},0} \le \frac{1}{c_1\|P_0\|_{\mathrm{op}}^3}$, then the warm start condition Condition 2.1 holds. Hence, by modifying constants, we can replace $\mathcal{C}_{\mathsf{stable}}$ in the warm-start condition with a data-dependent condition $\varepsilon_{\mathrm{op},0} \le \frac{1}{c_1\|P_0\|_{\mathrm{op}}^3}$ (for constant c_1), which depends only on the value function of the initial estimate. One can can show that this condition is guaranteed to be met as soon as $\varepsilon_{\mathrm{op},0} \le \frac{1}{c_2\|P_\star\|_{\mathrm{op}}^3}$, which means that the data-dependent warm start condition does not significantly alter what is required from the alignment condition.

F.2 Implementation via Representer Theorems

In the case where the states $\mathbf{x}_t$ are infinite dimensional feature mappings equal to $\phi(\mathbf{y}_t)$, where ϕ is a kernel and $\mathbf{y}_t$ is a finite dimensional observation, we can employ standard representer theorem arguments for kernel ridge regression in order to efficiently implement the algorithms above. More specifically, the estimates $\widehat{A}$ and $\widehat{B}$ can be represented via outer products of the data points $(\mathbf{x}_t, \mathbf{v}_t)$.

Furthermore, since inputs are finite dimensional, we can compute inverses and solve the Riccati equation by iterating the finite horizon version until convergence (see discussion in Fazel et al. [2018], Appendix A). That is, for $P_1 = Q$, Hewer [1971] shows that the following fixed point iteration is contractive and that $P_\infty(A, B)$ is equal to the limit of,

$$P_{t+1} = Q + A^{\mathsf{H}} P_t A - A^{\mathsf{H}} P_t B(R + B^{\mathsf{H}} P_t B)^{-1} B^{\mathsf{H}} P_t A.$$

Having solved the Riccati equation, we can then compute controller by taking products of linear operators, for which we have tractable representations.

G Regret Bounds

For the sake of the analysis in this section, we define the quantity $\text{Regret}_T(\mathcal{A}, \mathbf{z})$ as the regret incurred by the algorithm $\mathcal{A}$ over T time steps starting from (a possibly random) initial state $\mathbf{z}$.

Lemma G.1. *Let $\mathcal{A}$ be an algorithm that chooses actions according to $\mathbf{u}_t = K\mathbf{x}_t + \mathbf{v}_t$ for $\mathbf{v}_t \sim \mathcal{N}(0, \sigma_{\mathbf{u}}^2 I)$ and let $\mathbf{z} \sim \mathcal{N}(0, \Sigma_{\mathbf{z}})$ with $\|\Sigma_{\mathbf{z}}\|_{\text{op}} \le B_{\mathbf{z}}$. Then, with probability $1-\delta$, $\text{Regret}_T(\mathcal{A}, \mathbf{z})$ is less than or equal to,*

$$7\log_+(2/\delta)\left(T\left(\sigma_{\mathbf{u}}^2 \text{tr}\left[R\right] + \text{tr}\left[Q_0 \mathsf{dlyap}\left(A_{\text{cl}}^{\mathsf{H}},\ \Sigma_{\mathbf{w}} + B_\star B_\star^{\mathsf{H}} \sigma_{\mathbf{u}}^2\right)\right]\right) + B_{\mathbf{z}} \text{tr}\left[P_\infty(K; A_\star, B_\star)\right]\right) - TJ_\star,$$

where $Q_0 := Q + K^{\mathsf{H}}RK$ and $A_{\text{cl}} := A_\star + B_\star K$.

Proof. By definition, the regret of the algorithm is equal to:

$$\sum_{t=1}^{T} \left\langle \mathbf{x}_t, (Q + K^{\mathsf{H}}RK)\mathbf{x}_t \right\rangle + \langle \mathbf{v}_t, R\mathbf{v}_t\rangle - TJ_\star. \tag{G.1}$$

The lemma follows from first showing that the relevant random variables concentrate around their expectations and then upper bounding these expectations.

Bounding exploration cost Using the Hanson-Wright inequality (Lemma E.8), we argue that with probability $1-\delta/2$, since $\mathbf{v}_t \overset{\text{i.i.d}}{\sim} \mathcal{N}(0, \sigma_{\mathbf{u}}^2 I)$,

$$\sum_{t=1}^{T} \langle \mathbf{v}_t, R\mathbf{v}_t\rangle^2 = \|R^{1/2}\mathbf{v}_t\|^2 \le 7\sigma_{\mathbf{u}}^2 T \log_+(2/\delta) \text{tr}\left[R\right].$$

More precisely, we have applied Hanson-Wright to the series of random variables $\tilde{\mathbf{v}}_t = R^{1/2}\mathbf{v}_t$ and used the calculation, $\mathbb{E}\text{tr}\left[R\mathbf{v}_t \otimes \mathbf{v}_t\right] = \text{tr}\left[R\right]\sigma_{\mathbf{u}}^2$.

Bounding state cost Now let $Q_0 := Q + K^{\mathsf{H}}RK$, we have that,

$$\sum_{t=1}^{T} \left\langle \mathbf{x}_t, (Q + K^{\mathsf{H}}RK)\mathbf{x}_t \right\rangle = \sum_{t=1}^{T} \|Q_0^{1/2}\mathbf{x}_t\|^2 = \|\tilde{\mathbf{x}}\|^2,$$

where $\tilde{\mathbf{x}}$ is defined as, $\mathbf{z}_0 := \left[Q_0^{1/2}\mathbf{x}_1 \ \ldots Q_0^{1/2}\mathbf{x}_T\right]^{\mathsf{H}}$. Since all the $\mathbf{x}_t$ are Gaussian, $\tilde{\mathbf{x}_t}$ is also Gaussian, albeit in a different Hilbert space. Applying Hanson-Wright, we get that with probability $1-\delta/2$,

$$\|\tilde{\mathbf{x}}\|^2 \le 7\log_+(2/\delta)\mathbb{E}\text{tr}\left[\tilde{\mathbf{x}} \otimes \tilde{\mathbf{x}}\right].$$

Bounding expectation Letting $A_{\text{cl}} := A_\star + B_\star K$ we have that by definition of the dynamical system, for $j \ge 0$,

$$\mathbf{x}_{1+j} = A_{\text{cl}}^j \mathbf{x}_1 + \sum_{k=1}^{j} A_{\text{cl}}^{k-1}(B_\star \mathbf{v}_{1+j-k} + \mathbf{w}_{1+j-k}).$$

Therefore,

$$\mathbb{E}\left[\mathbf{x}_{1+j} \otimes \mathbf{x}_{1+j}\right] \preceq A_{\text{cl}}^j \mathbb{E}\left[\mathbf{x}_1 \otimes \mathbf{x}_1\right]\left(A_{\text{cl}}^{\mathsf{H}}\right)^j + \mathsf{dlyap}\left(A_{\text{cl}}^{\mathsf{H}},\ \Sigma_{\mathbf{w}} + B_\star B_\star^{\mathsf{H}} \sigma_{\mathbf{u}}^2\right).$$

Using this, we can upper bound $\mathbb{E}\mathrm{tr}\left[\tilde{\mathbf{x}} \otimes \tilde{\mathbf{x}}\right]$ as follows,

$$\begin{aligned}\mathbb{E}\mathrm{tr}\left[\tilde{\mathbf{x}} \otimes \tilde{\mathbf{x}}\right] &= \sum_{t=1}^{T} \mathbb{E}\mathrm{tr}\left[Q_0 \mathbf{x}_t \otimes \mathbf{x}_t\right] \\ &\leq \ \mathrm{tr}\left[Q_0 \sum_{j=0}^{\infty} A_{\mathrm{cl}}^j \mathbb{E}\left[\mathbf{x}_1 \otimes \mathbf{x}_1\right] \left(A_{\mathrm{cl}}^{\mathsf{H}}\right)^j\right] + T\mathrm{tr}\left[Q_0 \mathsf{dlyap}\left(A_{\mathrm{cl}}^{\mathsf{H}},\ \Sigma_{\mathbf{w}} + B_\star B_\star^{\mathsf{H}} \sigma_{\mathbf{u}}^2\right)\right] \qquad \text{(G.2)} \\ &\leq \ B_{\mathbf{z}} \mathrm{tr}\left[P_\infty(K; A_\star, B_\star)\right] + T\mathrm{tr}\left[Q_0 \mathsf{dlyap}\left((A_\star + B_\star K)^{\mathsf{H}},\ \Sigma_{\mathbf{w}} + B_\star B_\star^{\mathsf{H}} \sigma_{\mathbf{u}}^2\right)\right]. \qquad \text{(G.3)}\end{aligned}$$

The final inequality is justified by the following series of manipulations,

$$\begin{aligned}\mathrm{tr}\left[Q_0 \sum_{j=0}^{\infty} A_{\mathrm{cl}}^j \mathbb{E}\left[\mathbf{x}_1 \otimes \mathbf{x}_1\right] \left(A_{\mathrm{cl}}^{\mathsf{H}}\right)^j\right] \ &\leq \ \left\|\mathbb{E}\left[\mathbf{x}_1 \otimes \mathbf{x}_1\right]\right\|_{\mathrm{op}} \mathrm{tr}\left[\sum_{j=0}^{\infty} \left(A_{\mathrm{cl}}^{\mathsf{H}}\right)^j (Q + K^{\mathsf{H}} P K) A_{\mathrm{cl}}^j\right] \\ &= \left\|\mathbb{E}\left[\mathbf{x}_1 \otimes \mathbf{x}_1\right]\right\|_{\mathrm{op}} \mathrm{tr}\left[P_\infty(K; A_\star, B_\star)\right].\end{aligned}$$

Wrapping up Combining our results so far, we have that with probability $1-\delta$ the regret as expressed in Eq. (G.1) is less than or equal to,

$$7\log_+(2/\delta)\left(T\left(\sigma_{\mathbf{u}}^2 \mathrm{tr}\left[R\right] + \mathrm{tr}\left[Q_0 \mathsf{dlyap}\left(A_{\mathrm{cl}}^{\mathsf{H}},\ \Sigma_{\mathbf{w}} + B_\star B_\star^{\mathsf{H}} \sigma_{\mathbf{u}}^2\right)\right]\right) + B_{\mathbf{z}} \mathrm{tr}\left[P_\infty(K; A_\star, B_\star)\right]\right) - TJ_\star.$$

□

G.1 Proof of Corollary 3.1

Proof. The proof of this proposition consists of a simple application of Lemma G.1 and then bounding the relevant terms to show that the regret is $\mathcal{O}(T_{\mathsf{init}})$. Recall that WarmStart chooses inputs according to $\mathbf{u}_t = K_{\mathsf{init}} \mathbf{x}_t + \mathbf{v}_t$. Furthermore, the initial state is exactly 0 so $B_{\mathbf{z}} = 0$ in the statement of the lemma. Therefore, with probability $1-\delta$, for $A_{\mathrm{cl}} = A_\star + B_\star K_{\mathsf{init}}$, the regret is smaller than:

$$7T_{\mathsf{init}} \log_+(2/\delta)\left(\sigma_{\mathbf{u}}^2 \mathrm{tr}\left[R\right] + \mathrm{tr}\left[(Q + K_{\mathsf{init}}^{\mathsf{H}} R K_{\mathsf{init}}) \mathsf{dlyap}\left(A_{\mathrm{cl}}^{\mathsf{H}},\ \Sigma_{\mathbf{w}} + B_\star B_\star^{\mathsf{H}} \sigma_{\mathbf{u}}^2\right)\right]\right) - T_{\mathsf{init}} J_\star.$$

We note that $\mathrm{tr}\left[(Q + K_{\mathsf{init}}^{\mathsf{H}} R K_{\mathsf{init}}) \mathsf{dlyap}\left(A_{\mathrm{cl}}^{\mathsf{H}},\ \Sigma_{\mathbf{w}} + B_\star B_\star^{\mathsf{H}} \sigma_{\mathbf{u}}^2\right)\right] \leq \left\|Q + K_{\mathsf{init}}^{\mathsf{H}} R K_{\mathsf{init}}\right\|_{\mathrm{op}} \mathrm{tr}\left[\Sigma_\star(K_{\mathsf{init}}, \sigma_{\mathbf{u}}^2)\right]$. Since $J_\star \geq 0$, this expression above is then upper bounded by,

$$7\log_+(2/\delta)\left(\sigma_{\mathbf{u}}^2 \mathrm{tr}\left[R\right] + \left\|Q + K_{\mathsf{init}}^{\mathsf{H}} R K_{\mathsf{init}}\right\|_{\mathrm{op}} \mathrm{tr}\left[\Sigma_\star(K_{\mathsf{init}}, \sigma_{\mathbf{u}}^2)\right]\right) T_{\mathsf{init}}.$$

□

G.2 Proof of Theorem 3.1

Proof. By definition of the algorithm in Appendix F, we can split up the regret into two separate phases: an initial explore phase and then a commit phase.

$$\mathrm{Regret}_{T_{\mathsf{exp}}}(\mathsf{explore}, \mathbf{x}_1) + \mathrm{Regret}_{T - T_{\mathsf{exp}}}\left(\mathsf{commit}, \mathbf{x}_{T_{\mathsf{exp}}+1}\right).$$

The explore phase corresponds to the regret incurred during the first part of the algorithm wherein inputs are chosen according to $\mathbf{u}_t = K_0\mathbf{x}_t + \mathbf{v}_t$ for T_{exp} many iterations. The commit algorithm then chooses inputs $\mathbf{u}_t = \widehat{K}\mathbf{x}_t$ for the remaining $T - T_{\mathsf{exp}}$ time steps.

Exploration regret We now upper bound the regret in each phase individually. For the exploration phase, we can apply Lemma G.1, to get that with probability $1-\delta/4$, for $A_{\mathrm{cl}} = A_\star + B_\star K_0$ and $Q_0 = Q + K_0^{\mathsf{H}} R K_0$, the regret experienced in this phase bounded by,

$$7\log_+(8/\delta) T_{\mathsf{exp}}\left(\sigma_{\mathbf{u}}^2 \mathrm{tr}\left[R\right] + \mathrm{tr}\left[Q_0 \mathsf{dlyap}\left(A_{\mathrm{cl}}^{\mathsf{H}},\ \Sigma_{\mathbf{w}} + B_\star B_\star^{\mathsf{H}}\sigma_{\mathbf{u}}^2\right)\right]\right).$$

Since $\mathbf{x}_1 \sim \mathcal{N}(0, \Sigma_{\mathbf{x},0})$ is drawn from the same exploration distribution, the remaining constant term from Lemma G.1 vanishes since there is no need to consider the change in distribution. As in the proof of the WarmStart regret bound, we observe that,

$$\mathrm{tr}\left[(Q + K_0^{\mathsf{H}} R K_0)\mathsf{dlyap}\left(A_{\mathrm{cl}}^{\mathsf{H}},\ \Sigma_{\mathbf{w}} + B_\star B_\star^{\mathsf{H}}\sigma_{\mathbf{u}}^2\right)\right] \ \leq\ \left\|Q + K_0^{\mathsf{H}} R K_0\right\|_{\mathrm{op}} \mathrm{tr}\left[\Sigma_{\mathbf{x},0}\right].$$

Given that the initial estimates satisfy Condition 2.1, by Lemma D.7 and Lemma C.8 we have that

$$\|K_0\|_{\mathrm{op}}^2 \ \leq\ \|P_\infty(A_0, B_0)\|_{\mathrm{op}} \lesssim \|P_\star\|_{\mathrm{op}}.$$

Hence,

$$\left\|Q + K_0^{\mathsf{H}} R K_0\right\|_{\mathrm{op}} \ \leq\ \|Q\|_{\mathrm{op}} + \|K_0\|_{\mathrm{op}}^2 \|R\|_{\mathrm{op}} \lesssim M_\star^2.$$

Therefore, with probability $1-\delta/4$,

$$\mathrm{Regret}_{T_{\mathsf{exp}}}(\mathsf{explore}, \mathbf{x}_1) \lesssim \log_+(1/\delta) T_{\mathsf{exp}}\left(\sigma_{\mathbf{u}}^2 \mathrm{tr}\left[R\right] + M_\star^2 \mathrm{tr}\left[\Sigma_{\mathbf{x},0}\right]\right). \tag{G.4}$$

Bounding regret during commit phase Moving on to bounding regret during the second phase, our first observation is that due to the projection step, the system estimates $\widehat{A}, \widehat{B}$ lie inside an operator norm ball of radius $.5\mathcal{C}_{\mathsf{stable}}$ around the warm start estimates (A_0, B_0). Since the warm start estimates are themselves $.5\mathcal{C}_{\mathsf{stable}}$ close to $(A_\star, B_\star)$, we conclude that $(\widehat{A}, \widehat{B})$ satisfy Condition 2.1 and by Lemma C.8, $\widehat{K} = K_\infty(\widehat{A}, \widehat{B})$ is guaranteed to be stabilizing for the true system $(A_\star, B_\star)$. Furthermore, $\|P_0\|_{\mathrm{op}} \lesssim M_\star$ and $\|\widehat{P}\|_{\mathrm{op}} \lesssim M_\star$.

Next, we use a similar regret analysis as before and apply Lemma G.1, to conclude that with probability $1-\delta/4$, $\mathrm{Regret}_{T-T_{\mathsf{exp}}}\left(\mathsf{commit}, \mathbf{x}_{T_{\mathsf{exp}}+1}\right)$ is upper bounded by,

$$7T\log_+(8/\delta)\left(\mathrm{tr}\left[Q_0 \mathsf{dlyap}\left(A_{\mathrm{cl}}^{\mathsf{H}},\ \Sigma_{\mathbf{w}}\right)\right] - \mathcal{J}_\star\right) + 7\log_+(8/\delta)\left\|\Sigma_{\mathbf{x},0}\right\|_{\mathrm{op}} \mathrm{tr}\left[P_\infty(K_0; A_\star, B_\star)\right], \tag{G.5}$$

where $A_{\mathrm{cl}} = A_\star + B_\star\widehat{K}$ and $Q_0 = Q + \widehat{K}^{\mathsf{H}} R \widehat{K}$. Next, since

$$\mathrm{tr}\left[Q_0 \mathsf{dlyap}\left(A_{\mathrm{cl}}^{\mathsf{H}},\ \Sigma_{\mathbf{w}}\right)\right] = \mathrm{tr}\left[(Q + \widehat{K}^{\mathsf{H}} R \widehat{K})\Sigma_\star(\widehat{K})\right] = \mathcal{J}(\widehat{K}),$$

we can rewrite Eq. (G.5) as,

$$7T\log_+(8/\delta)(J(\widehat{K}) - J(K_\star)) + 7\log_+(8/\delta)\left\|\Sigma_{\mathbf{x},0}\right\|_{\mathrm{op}} \mathrm{tr}\left[P_\infty(K_0; A_\star, B_\star)\right].$$

Using Theorem 2.1,

$$\mathcal{J}(\widehat{K}) - \mathcal{J}(K_\star) \lesssim M_\star^{36} \cdot \mathcal{L}\exp(\tfrac{1}{50}\sqrt{\mathcal{L}}) \cdot \varepsilon^2, \qquad \text{where } \mathcal{L} := \log\left(e + \frac{2e\|\widehat{A} - A_\star\|_{\mathrm{op}}^2 \mathrm{tr}[\Sigma_{\mathbf{x},0}]}{\varepsilon^2}\right),$$

where $\varepsilon = \max\left\{\left\|(\widehat{A} - A_\star)\Sigma_{\mathbf{x},0}^{1/2}\right\|_{\mathsf{HS}}, \left\|\widehat{B} - B_\star\right\|_{\mathsf{HS}}\right\}$ and $\sigma_{\mathbf{u}}^2$ has been set to 1 as in the description of the OnlineCE in Appendix F. To finish the proof, we now plug in our estimation rates from Part II to upper bound ε and optimize over T_{exp}.

Plugging in estimation rates Using Proposition E.2, for $T_{\mathsf{exp}} \gtrsim d_u \log\left(\frac{d_u}{\delta}\right)$, with probability $1-\delta/4$,

$$\left\|B_\star - \widehat{B}\right\|_{\mathsf{HS}}^2 \lesssim \frac{d_u \mathrm{tr}\left[\Sigma_{\mathbf{x},0}\right] \log\left(\frac{d_u}{\delta}\right)}{\sigma_{\mathbf{u}}^2 T_{\mathsf{exp}}}.$$

Incorporating the analogous proposition for $A_\star$, Proposition E.3, for

$$T_{\mathsf{exp}} \gtrsim d_\lambda \log_+\left(\frac{\mathrm{tr}\left[\Sigma_{\mathbf{x},0}\right]}{\delta\lambda}\right) + \|P_\star\|_{\mathrm{op}} \log_+\left(\frac{\|P_\star\|_{\mathrm{op}}^2}{\lambda}\left\|B_\star B_\star^{\mathsf{H}}\sigma_{\mathbf{u}}^2 + \Sigma_{\mathbf{w}}\right\|_{\mathrm{op}}\right),$$

with probability $1-\delta/4$

$$\left\|(\widehat{A}_{\mathrm{cl}} - A_{\mathrm{cl}_\star})\Sigma_{\mathbf{x},0}^{1/2}\right\|_{\mathsf{HS}}^2 \lesssim \lambda \|A_{\mathrm{cl}_\star}\|_{\mathsf{HS}}^2 + \cdot \frac{W_{\mathrm{tr}}}{T_{\mathsf{exp}}}(d_\lambda + \mathcal{C}_{\mathsf{tail},\lambda}) \log_+\left(\frac{\mathrm{tr}\left[\Sigma_{\mathbf{x},0}\right]}{\delta\lambda}\right).$$

Setting $\lambda = c\frac{W_{\mathrm{tr}}}{T_{\mathsf{exp}}\|A_{\mathrm{cl}_\star}\|_{\mathsf{HS}}^2}$ for some universal constant c, the second term dominates the first and we get that,

$$\left\|(\widehat{A}_{\mathrm{cl}} - A_{\mathrm{cl}_\star})\Sigma_{\mathbf{x},0}^{1/2}\right\|_{\mathsf{HS}}^2 \lesssim \frac{W_{\mathrm{tr}}}{T_{\mathsf{exp}}}(d_\lambda + \mathcal{C}_{\mathsf{tail},\lambda}) \log\left(\frac{\mathrm{tr}\left[\Sigma_{\mathbf{x},0}\right] T_{\mathsf{exp}} \|A_{\mathrm{cl}_\star}\|_{\mathsf{HS}}^2}{W_{\mathrm{tr}}\delta}\right).$$

From our definition of ε, we see that it is upper bounded by, the sum of the errors in $A_\star$ and $B_\star$ and hence,

$$\varepsilon^2 \lesssim \frac{d_u \mathrm{tr}\left[\Sigma_{\mathbf{x},0}\right] + W_{\mathrm{tr}}(d_\lambda + \mathcal{C}_{\mathsf{tail},\lambda})}{T_{\mathsf{exp}}} \log\left(\frac{d_u \mathrm{tr}\left[\Sigma_{\mathbf{x},0}\right] \|A_{\mathrm{cl}_\star}\|_{\mathsf{HS}}^2 T}{W_{\mathrm{tr}}\delta^2}\right),$$

where above we also upper bounded $T_{\mathsf{exp}} \le T$.

Wrapping up All that remains is to optimize over T_{exp} to balance the regret between both phases. In particular, if we choose,

$$T_{\mathsf{exp}} = \sqrt{\frac{T \cdot M_\star^{36}\left(d_u \mathrm{tr}\left[\Sigma_{\mathbf{x},0}\right] + W_{\mathrm{tr}}(d_\lambda + \mathcal{C}_{\mathsf{tail},\lambda})\right)}{\sigma_{\mathbf{u}}^2 \mathrm{tr}\left[R\right] + M_\star^2 \mathrm{tr}\left[\Sigma_{\mathbf{x},0}\right]}},$$

we get that with probability $1-\delta$, up to constants and log factors, the total regret is bounded by:

$$\sqrt{(\mathrm{tr}\left[R\right] + M_\star^2 \mathrm{tr}\left[\Sigma_{\mathbf{x},0}\right]) M_\star^{36}\left(d_u \mathrm{tr}\left[\Sigma_{\mathbf{x},0}\right] + W_{\mathrm{tr}}(d_\lambda + \mathcal{C}_{\mathsf{tail},\lambda})\right) T} \cdot \varphi(T)$$

where $\varphi(T) = \exp\left(\sqrt{\log\left(1 + \sqrt{T}\mathrm{tr}\left[\Sigma_{\mathbf{x},0}\right]\right)}\right)$. Simplifying the bound a bit further, we know that $\mathrm{tr}\left[R\right] \le \|R\|_{\mathrm{op}} d_u$. And, by Lemma D.8,

$$\begin{aligned}
\mathrm{tr}\left[\Sigma_{\mathbf{x},0}\right] &= \mathrm{tr}\left[\mathsf{dlyap}\left((A_\star + B_\star K_0)^{\mathsf{H}},\ B_\star B_\star^{\mathsf{H}}\sigma_{\mathbf{u}}^2 + \Sigma_{\mathbf{w}}\right)\right] \\
&\le \left\|\mathsf{dlyap}\left((A_\star + B_\star K_0)^{\mathsf{H}},\ Q + K_0^{\mathsf{H}} R K_0\right)\right\|_{\mathrm{op}} \mathrm{tr}\left[B_\star B_\star^{\mathsf{H}}\sigma_{\mathbf{u}}^2 + \Sigma_{\mathbf{w}}\right] \\
&\le M_\star(M_\star + \mathrm{tr}\left[\Sigma_{\mathbf{w}}\right]) \\
&\lesssim M_\star^2 \mathrm{tr}\left[\Sigma_{\mathbf{w}}\right].
\end{aligned}$$

Hence, if we let $d_{\max} := \max\{\mathrm{tr}\left[\Sigma_{\mathbf{w}}\right], W_{\mathrm{tr}}, d_u\}$ we get that:

$$\mathrm{tr}\left[R\right] + M_\star^2 \mathrm{tr}\left[\Sigma_{\mathbf{x},0}\right] \lesssim M_\star^4 d_{\max} \text{ and } d_u \mathrm{tr}\left[\Sigma_{\mathbf{x},0}\right] + W_{\mathrm{tr}}(d_\lambda + \mathcal{C}_{\mathsf{tail},\lambda}) \lesssim M_\star^2 d_{\max}^2.$$

Therefore, we can upper bound the total regret of $\mathsf{OnlineCE}$ is with high probability $\mathcal{O}_\star\left(\sqrt{M_\star^{42} d_{\max}^2 (d_\lambda + \mathcal{C}_{\mathsf{tail},\lambda}) T}\right)$. □

G.3 Proof of Theorem 3.2

Proof. The proof of the theorem follows by bounding d_λ and $\mathcal{C}_{\mathsf{tail},\lambda}$ based on the decay rates of $\Sigma_{\mathbf{w}}$. We overload notation and define $d_\lambda(\Lambda)$ for $\Lambda \succeq 0$ to be the number of eigenvalues of Λ that are larger than λ. Likewise, we define $\mathcal{C}_{\mathsf{tail},\lambda}(\Lambda)$ to be the sum of the eigenvalues of Λ that are smaller than λ, divided by λ. The proof follows by applying two inequalities which follow from Lemma D.15. In particular, since $B_\star$ is finite rank and by the warm start property, the following are true for $n \gtrsim n_0 := \|P_\star\|_{\mathrm{op}} \log\left(\|P_\star\|_{\mathrm{op}}^2 W_{\mathrm{tr}}/\lambda\right)$,

$$
\begin{aligned}
d_\lambda(\Sigma_{\mathbf{x},0}) &\leq n_0 d_{\lambda/(2\|P_\star\|_{\mathrm{op}}^2)}(\Sigma_{\mathbf{w}}) + d_u \|\Sigma_{\mathbf{x},0}\|_{\mathrm{op}} \\
\mathcal{C}_{\mathsf{tail},\lambda}(\Sigma_{\mathbf{x},0}) &\leq \frac{1}{\lambda} n_0 \|P_\star\|_{\mathrm{op}}^2 \left(\sum_{j\,\geq\,\lceil a\rceil}^{\infty} \sigma_j(\Sigma_{\mathbf{w}}) + \lambda\right) + d_u \|\Sigma_{\mathbf{x},0}\|_{\mathrm{op}},
\end{aligned}
$$

where $a = d_\lambda(\Sigma_{\mathbf{w}})/n_0$. We analyze each case separately.

Polynomial decay If $\sigma_j(\Sigma_{\mathbf{w}}) = j^{-\alpha}$, then a short calculation shows that $d_\lambda(\Sigma_{\mathbf{w}}) = \lfloor \lambda^{-1/\alpha} \rfloor$. Therefore,

$$
d_\lambda(\Sigma_{\mathbf{x},0}) \leq n_0 \left(\frac{\lambda}{\|P_\star\|_{\mathrm{op}}^2}\right)^{-1/\alpha} + d_u \|\Sigma_{\mathbf{x},0}\|_{\mathrm{op}}.
$$

Since $\lambda = c \cdot \frac{W_{\mathrm{tr}}}{T\|A_{\mathrm{cl},\star}\|_{\mathsf{HS}}^2}$ for some constant c, $d_\lambda(\Sigma_{\mathbf{x},0})$ scales no faster than $n_0 T^{1/\alpha} + d_u \|\Sigma_{\mathbf{x},0}\|_{\mathrm{op}}$. For $\mathcal{C}_{\mathsf{tail},\lambda}$, we have that

$$
\sum_{j\,\geq\,\lceil \frac{\lambda^{-1/\alpha}}{n_0}\rceil} j^{-\alpha} \leq \int_{\frac{\lambda^{-1/\alpha}}{n_0}} j^{-\alpha} = \frac{1}{\alpha-1}\lambda^{(\alpha-1)/\alpha} n_0^{\alpha-1}.
$$

Therefore, after dividing by λ, we get that $\mathcal{C}_{\mathsf{tail},\lambda}$ scales as $n_0 \|P_\star\|_{\mathrm{op}}^2 T^{1/\alpha} \|A_{\mathrm{cl},\star}\|_{\mathsf{HS}}^2 + d_u \|\Sigma_{\mathbf{x},0}\|_{\mathrm{op}}$. Since $\|\Sigma_{\mathbf{x},0}\|_{\mathrm{op}} \lesssim M_\star^3$, we get that $d_\lambda + \mathcal{C}_{\mathsf{tail},\lambda}$ are $\widetilde{\mathcal{O}}(M_\star^4 T^{1/\alpha})$.

Exponential decay Moving on to the case where the eigenvalues of $\Sigma_{\mathbf{w}}$ decay exponentially fast, a similar calculation to the previous one shows that $d_\lambda(\Sigma_{\mathbf{w}}) = \lfloor \alpha^{-1}\log(1/\lambda)\rfloor$. Therefore,

$$
d_\lambda(\Sigma_{\mathbf{x},0}) \lesssim \frac{n_0}{\alpha}\log\left(\frac{\|P_\star\|_{\mathrm{op}}^2}{\lambda}\right) + d_u \|\Sigma_{\mathbf{x},0}\|_{\mathrm{op}}.
$$

Hence $d_\lambda(\Sigma_{\mathbf{x},0})$ scales as $\widetilde{\mathcal{O}}(n_0 + d_u M_\star^3)$. For the tail term, we observe that $\sum_{j=1}^{\infty}\exp(-\alpha j) = 1/(\exp(\alpha)-1)$. This term therefore scales no faster than $n_0\|P_\star\|_{\mathrm{op}}^2 + d_u\|\Sigma_{\mathbf{x},0}\|_{\mathrm{op}}$. This shows that $d_\lambda + \mathcal{C}_{\mathsf{tail},\lambda}$ are $\widetilde{\mathcal{O}}(M_\star^3 d_u)$.

Finite dimension In finite dimension with full rank noise, it is clear that for large enough T, $d_\lambda(\Sigma_{\mathbf{x},0})$ is bounded by d_x and that $\mathcal{C}_{\mathsf{tail},\lambda}$ is equal to 0. Furthermore, using now standard analysis such as the ones present in Simchowitz et al. [2018], it follows that $\|\widehat{A} - A_\star\|_{\mathsf{HS}}$ goes to 0 at the same rate as $\|(\widehat{A} - A_\star)\Sigma_{\mathbf{x},0}^{1/2}\|_{\mathsf{HS}}$. Therefore, the terms depending on $\mathcal{L}$ in Theorem 3.1 become $\mathcal{O}(1)$.

□

G.4 Combining WarmStart and OnlineCE

In the analysis of OnlineCE (Theorem 3.1), we assumed that $\mathbf{x}_1 \sim \mathcal{N}(0, \Sigma_{\mathbf{x},0})$ was distributed according to the steady state distribution of induced by the exploration policy. This assumption can be relaxed, since any initial distribution over $\mathbf{x}_1$ will converge exponentially quickly to the steady steady in Wasserstein distance due to a mixing argument (variants of this argument are ubiquitous in the analysis of online LQR, and for brevity we omit them. The curious reader can see appendices of Dean et al. [2018], Abeille and Lazaric [2020] for examples. The mixing time will be a polynomial in $\|P_\infty(K_0; A_\star, B_\star)\|_{\mathrm{op}}$, which we show is $\lesssim \|P_\star\|_{\mathrm{op}}$.

Hence to stitch the two regret bounds together, we simply run the initial phase to garner estimates (A_0, B_0), begin to play controller $K_0 = K_\infty(A_0, B_0,)$, allow a constant-length burnin for the state to converge to the distribution of $\Sigma_{\mathbf{x},0}$, and then execute OnlineCE. Again, for the sake of brevity, we omit the details.

As a final remark, when stitching both algorithms together, one could in principle omit synthesizing the controller K_0 as outlined in the first step of the OnlineCE algorithm and run the entire exploration phase just using K_{init}. Doing so would only increase the constants $M_\star$ since they would now depend on $\|P_\infty(K_{\mathsf{init}}; A_\star, B_\star)\|_{\mathrm{op}}$. The asymptotics of the algorithm would remain unchanged. However, the projection step onto a safe set around (A_0, B_0) is crucial for our analysis in order to ensure that the certainty equivalent controller is stabilizing for the true system.

H Lower Bound

In this section, we state and prove lower bounds demonstrating the necessity of finite input dimension; these results follow from applications of the lower bound due to Simchowitz and Foster [2020]. Our first bound is as follows:

Theorem 1.1. Let $c, c' > 0$ denote universal constants. Fix any trace bound $\gamma \geq 1$ and input dimension $d_u \in \mathbb{N}$ with $d_u \geq \sqrt{\log(1+\gamma)}$. Consider the set $\mathcal{U}$ of instances with state dimension $d_x = \lfloor \gamma \rfloor$ defined by

$$\mathcal{U} := \{(A, B) := \|A - \frac{1}{2}I\|_{\mathsf{HS}} \leq \frac{1}{4}, \quad \|B\|_{\mathsf{HS}} \leq \frac{1}{4}\}.$$

Then, the LQR regret with cost matrices $Q = I_{d_x}$, $R = I_{d_u}$, and noise $\|\Sigma_{\mathbf{w}}\|_{\mathrm{op}} = 1$, $\mathrm{tr}[\Sigma_{\mathbf{w}}] \leq \gamma$ satisfies

$$\min_{\mathsf{alg}} \max_{(A,B)\in\mathcal{U}} \mathbb{E}_{A,B}[\mathrm{Regret}_T(\mathsf{alg})] \geq c \cdot \begin{cases} T & T \in [c'\gamma \log(1+\gamma), \gamma d_u^2] \\ \sqrt{\gamma d_u^2 \cdot T} & T \geq \gamma d_u^2 \end{cases}.$$

We prove the bound in the following subsection. The bound considers instances that lie in a finite dimensional Hilbert space of dimension $d_x = \lfloor \gamma \rfloor$. In particular, all the instances in the packing are operator norm bounded, Hilbert-Schmidt, and in fact finite rank. The difficulty introduced by high-dimensional inputs is one of a needle in the haystack: to find the optimal control policy, the learner needs to learn to align their controller with the true $B_\star$ matrix, and doing so incurs dependence on ambient dimension. In essence, this is because the learner is free to pick any direction she chooses, so the complexity of the problem behaves more like, say, a linear bandit problem in dimension d_u than a statistical learning problem which admits a more refined notion of intrinsic dimension.

In Appendix H.2, we state and sketch the proof of a lower bound that holds *even if* all the instances are controllable, demonstrating that little can be done to remove the finite dimensionality requirement of inputs.

H.1 Proof of Theorem 1.1

Throughout, we let $c_i, i > 1$ denote universal constants. Recall that $\gamma \geq 1$ is the trace bound, and $d_x = \lfloor \gamma \rfloor$ is the state dimension. We select $\Sigma_{\mathbf{w}} = I_{d_x}$, which has trace $\mathrm{tr}\,[\Sigma_{\mathbf{w}}] = d_x \leq \gamma$.

Our lower bound follows from specializing the lower bound due to Simchowitz and Foster [2020]. To begin, we state a variant of their main lower bound.

Proposition H.1 (Variant of Theorem 1 in Simchowitz and Foster [2020])**.** *Let $c_1, c_2, p > 0$ denote universal constants. Consider a finite dimensional LQR system $(A_\star, B_\star)$, with finite input dimension d_u, state dimension d_x, cost matrices $R, Q \succeq I$, optimal controller $K_\star$, value function $P_\star$, and noise $\Sigma_{\mathbf{w}} = I_{d_x}$. Suppose $\nu := \sigma_{\min}(A_\star + B_\star K_\star)/\|R + B_\star^\top P_\star B_\star\|_{\mathrm{op}} > 0$. Then, defining the convex set,*

$$\mathcal{B} := \{(A, B) : \|A - A_\star\|_{\mathsf{HS}}^2 + \|B - B_\star\|_{\mathsf{HS}}^2\} \leq \frac{1}{16}, \tag{H.1}$$

it holds that,

$$\min_{\mathsf{alg}} \max_{(A,B)\in\mathcal{B}} \mathbb{E}_{A,B}[\mathrm{Regret}_T(\mathsf{alg})] \geq c_2\sqrt{d_x d_u^2 T} \cdot \frac{\min\{1,\nu^2\}}{\|P_\star\|_{\mathrm{op}}^2},$$

provided that $T \geq c_1 \|P_\star\|_{\mathrm{op}}^p \min\{d_u^2 d_x, \frac{d_x \max\{1,\nu^4\}\max\{1,\|B_\star\|_{\mathrm{op}}^4\}}{d_u^2}, d_x \log(1+d_x\|P_\star\|_{\mathrm{op}})\}$.

Proof. Proposition H.1 is obtained by specializing the proof of the lower bound from Theorem 1 in Simchowitz and Foster [2020] to $m = d_u$, and noting that the instances in the construction of the lower bound lie in an ball which satisfies the conditions of Lemma 4.1 their work (note that, by enlarging the constant term in the unspecified polynomial in that lemma, we can ensure that the constant is sufficiently small). □

First, for simplicity, we specialize Proposition H.1 with a concrete instance. Take $A_\star = \frac{1}{2}I$, $R = Q = I$ and let $B_\star = \mathbf{0}$.

Lemma H.2. *Let* $P_\star$ *denote the value function for the instance* $(A_\star, B_\star)$, $K_\star$ *the optimal controller, and* $A_{\mathrm{cl}_\star} := A_\star + B_\star K_\star$ *the optimal closed loop systems. Then,* $P_\star = \frac{4}{3}I$, $K_\star = 0$, *and* $A_{\mathrm{cl}_\star} = A_\star = \frac{1}{2}I$.

Proof. Since $B_\star = \mathbf{0}$, $A_{\mathrm{cl}_\star} = A_\star$. Moreover, zero $B_\star$ means that the control inputs do not affect the system. Hence, optimal performance is optimized by selecting no control input, so as to minimize input cost; that is, $K_\star = 0$. Hence, the value function is $\mathsf{dlyap}\,(A_{\mathrm{cl}_\star},\, Q) = \sum_{j\geq 0}(\frac{1}{2}I)^j I(\frac{1}{2})^j = I \cdot \sum_{j\geq 0} 4^{-j} = \frac{1}{1-1/4}I = \frac{4}{3}I$. □

Invoking Lemma H.4, and using $d_u^2 \geq \log(1+\gamma) \geq \log(1+d_x)$ to simplify terms, we find that for universal (i.e. dimension independent) constants c_3, c_4, the instances centered in $\mathcal{B} := \{(A,B) : \|A - A_\star\|_{\mathsf{HS}}^2 + \|B\|_{\mathsf{HS}}^2\} \leq \frac{1}{4}\}$ satisfy

$$\min_{\mathsf{alg}} \max_{(A,B)\in\mathcal{B}} \mathbb{E}_{A,B}[\mathrm{Regret}_T(\mathsf{alg})] \geq c_3\sqrt{d_x d_u^2 \cdot T}, \quad \forall T \geq c_4 d_u^2 d_x. \tag{H.2}$$

To conclude, we address the case where $T \leq c_4 d_u^2 d_x$. For $T \geq 4c_4 d_x \log(1+d_x)$, consider a smaller input dimension $d := \sqrt{\lfloor T/c_4 d_x \rfloor}$. Let $\tilde{\mathcal{B}}_d := \{(A,\tilde{B}) \in \mathbb{R}^{d_x^2} \times \mathbb{R}^{d_x d} : \|A - A_\star\|_{\mathsf{HS}}^2 + \|\tilde{B}\|_{\mathsf{HS}}^2\} \leq \frac{1}{4}\}$ denote the analogous set of local instances to the above construction with input dimension d. From the choice of d and condition on T, we have $d^2 \geq \log(1+d_x)$ and $T \geq c_4 d^2 d_u$, so the above lower bound established in dimension d_u entails that, for another universal constant c_5,

$$\min_{\tilde{\mathsf{alg}}} \max_{(A,\tilde{B})\in\tilde{\mathcal{B}}_d} \mathbb{E}_{A,\tilde{B}}[\mathrm{Regret}_T(\tilde{\mathsf{alg}})] \geq c_3\sqrt{d_x d^2 \cdot T} \geq c_5 T,$$

where the last inequality follows from the choice of $d := \sqrt{\lfloor T/c_4 d_x \rfloor}$. Stated otherwise, it holds that for all $T \in [4c_4 d_x \log(1+d_x), c_4 d_x d_u^2]$,

$$\exists\, d \in [1, d_u] \text{ such that } \min_{\tilde{\mathsf{alg}}} \max_{(A,\tilde{B})\in\tilde{\mathcal{B}}_d} \mathbb{E}_{A,\tilde{B}}[\mathrm{Regret}_T(\tilde{\mathsf{alg}})] \geq c_5 T.$$

We now embed the above instances of input dimension d into input dimension d_u via the following lemma.

Lemma H.3. *Fix a state dimension* d_x, *and an input dimension* $d \leq d_u$. *Given a matrix* $\tilde{B} \in \mathbb{R}^{d_x \times d}$, *let* $\iota(\tilde{B})$ *denote its canonical embedding into* $\mathbb{R}^{d_x \times d_u}$ *by padding the remaining* $d_u - d$ *columns with zeros. Overloading notation, given a subset of* $\tilde{\mathcal{U}} \subset \mathbb{R}^{d_x^2} \times \mathbb{R}^{d_x d_u}$, *define its embedding*

$$\iota(\tilde{\mathcal{U}}) := \{(A,B) : \exists (A,\tilde{B}) \in \tilde{\mathcal{U}} \text{ with } B = \iota(\tilde{B})\}.$$

Then,

$$\inf_{\mathsf{alg}} \max_{(A,B)\in\mathcal{U}} \mathbb{E}_{A,B,}[\mathrm{Regret}_T(\mathsf{alg})] \geq \inf_{\tilde{\mathsf{alg}}} \max_{(A,\tilde{B})\in\tilde{\mathcal{U}}} \mathbb{E}_{A,\tilde{B}}[\mathrm{Regret}_T(\tilde{\mathsf{alg}})],$$

where on both sides, the noise covariance is $\Sigma_{\mathbf{w}} = I$ *and the state cost* $Q = I_{d_x}$. *On the left hand side, the input cost is* I_{d_u}, *and on the right,* $R = I_d$.

Thus, letting $\bar{\mathcal{B}} := \bigcup_{d\in[1,d_u]} \iota(\tilde{\mathcal{B}}_d)$ denote the union of all the embeddings of the sets $\bar{\mathcal{B}}$, it holds that for all $T \in [4c_4 d_x \log(1+d_x), c_4 d_x d_u^2]$,

$$\min_{\mathsf{alg}} \max_{((A,B)\in\bar{\mathcal{B}}} \mathbb{E}_{A,B}[\mathrm{Regret}_T(\mathsf{alg})] \geq c_5 T.$$

Moreover, since $\bar{\mathcal{B}} \supset \mathcal{B}$, incorporating Eq. (H.2) we conclude that,

$$\min_{\mathsf{alg}} \max_{((A,B)\in\bar{\mathcal{B}}} \mathbb{E}_{A,B}[\mathrm{Regret}_T(\mathsf{alg})] \geq \begin{cases} c_5 T & T \in [4c_4 d_x \log(1+d_x), c_4 d_x d_u^2] \\ c_3\sqrt{d_x d_u^2 \cdot T} & T \geq c_4 d_x d_u^2 \end{cases}.$$

Replacing d_x with $\lfloor\gamma\rfloor$, absorbing constants, and simplifying, we find that for universal constants c_6, c_7:

$$\min_{\mathsf{alg}} \max_{((A,B)\in\bar{\mathcal{B}}} \mathbb{E}_{A,B}[\mathrm{Regret}_T(\mathsf{alg})] \geq c_6 \begin{cases} T & T \in [c_7\gamma \log(1+\gamma), \gamma d_u^2] \\ \sqrt{\gamma d_u^2 \cdot T} & T \geq \gamma d_u^2 \end{cases}.$$

Finally, we observe that

$$\bar{\mathcal{B}} \subset \mathcal{U} := \{(A,B) := \|A - \frac{1}{2}I\|_{\mathsf{HS}} \leq \frac{1}{4}, \quad \|B\|_{\mathsf{HS}} \leq \frac{1}{4}\}.$$

Proof of Lemma H.3. We begin with the following claim.

Claim H.1. Let alg be an algorithm which interacts with instances $(A,B) \in \mathcal{U}$. Then, there is an algorithm $\tilde{\mathsf{alg}}$ which interacts with instances $(A,\tilde{B}) \in \tilde{\mathcal{U}}$ for which,

$$\forall A, \tilde{B}, B = \iota(\tilde{B}), \quad \mathbb{E}_{A,\tilde{B},\tilde{\mathsf{alg}}}[\sum_{t=1}^{T} \|\mathbf{x}_t\|^2 + \|\mathbf{u}_t\|^2] \leq \mathbb{E}_{A,B,\mathsf{alg}}[\sum_{t=1}^{T} \|\mathbf{x}_t\|^2 + \|\mathbf{u}_t\|^2],$$

with equality if alg always plays inputs $\mathbf{u}_t$ whose last $d_u - d$ coordinates are 0.

Proof. Given an input $\mathbf{u}_t \in \mathbb{R}^{d_u}$, write $\mathbf{u}_t = (\tilde{\mathbf{u}}_t, \mathring{\mathbf{u}}_t)$ as the decomposition of $\mathbf{u}_t$ into its first d, and last $d_u - d$ coordinates. Observe that, for instances $(A,B) \in \mathcal{U}$, the last $d_u - d$ coordinates $\mathring{\mathbf{u}}_t$ *do not* affect the dynamics. Hence, the iterates $(\mathbf{x}_t, \tilde{\mathbf{u}}_t)$ produced by alg on (A,B) in $\mathcal{U}$ coincide with the iterates obtained by the algorithm $\tilde{\mathsf{alg}}$ which, given an instances $(A,\tilde{B}) \in \tilde{\mathcal{U}}$ proceeds as follows:

- $\tilde{\mathsf{alg}}$ maintains "internal" inputs $\bar{\mathbf{u}}_t$ corresponding to the inputs that would have been selected by the original algorithm alg with input dimension d_u.
- For each t, $\tilde{\mathsf{alg}}$ feeds alg the past iterates $\mathbf{x}_{1:t}, \bar{\mathbf{u}}_{1:t-1}$, and receive internal input $\bar{\mathbf{u}}_t$.
- Then, $\tilde{\mathsf{alg}}$ plays the input $\mathbf{u}_t \in \mathbb{R}^d$ obtained by projecting $\bar{\mathbf{u}}_t$ onto its first d coordinates.

Given an instance $(A,\tilde{B}) \in \tilde{U}$, and its embedding $(A,B) = (A,\iota(\tilde{B}))$, the iterates $(\mathbf{x}_t, \tilde{\mathbf{u}}_t)$ produced by alg on (A,B) have the same distribution as the iterates $(\mathbf{x}_t, \mathbf{u}_t) \in \mathbb{R}^{d_x} \times \mathbb{R}^d$ produced by $\tilde{\mathsf{alg}}$. Hence, for all $A, \tilde{B}, B = \iota(\tilde{B})$, it holds that

$$\begin{aligned}\mathbb{E}_{A,\tilde{B},\tilde{\mathsf{alg}}}[\sum_{t=1}^{T} \|\mathbf{x}_t\|^2 + \|\mathbf{u}_t\|^2] &= \mathbb{E}_{A,B,\mathsf{alg}}[\sum_{t=1}^{T} \|\mathbf{x}_t\|^2 + \|\tilde{\mathbf{u}}_t\|^2] \\ &\leq \mathbb{E}_{A,B,\mathsf{alg}}[\sum_{t=1}^{T} \|\mathbf{x}_t\|^2 + \|\mathbf{u}_t\|^2],\end{aligned}$$

with equality if remaining $d_u - d$ coordinates of the inputs prescribed by alg are identically 0 for all t. □

Arguing along the lines of Claim H.1, we can also see that the optimal infinite horizon control policy for $(A, B) \in \mathcal{U}$ also only selects inputs supported on the first d coordinates, and thus

$$J^\star_{A,B} = J^\star_{A,\tilde{B}}, \quad \forall (A, \tilde{B}) \in \mathcal{U}, B = \iota(\tilde{B}).$$

Consequently,

$$\begin{aligned}
\max_{(A,B)\in\mathcal{U}} \mathbb{E}_{A,B}[\mathrm{Regret}_T(\mathsf{alg})] &= \max_{(A,B)\in\mathcal{U}} \mathbb{E}_{A,B,\mathsf{alg}}[\sum_{t=1}^T \|\mathbf{x}_t\|^2 + \|\mathbf{u}_t\|^2] - TJ^\star_{A,B} \\
&\ge \max_{(A,B)\in\mathcal{U}} \mathbb{E}_{A,\tilde{B},\tilde{\mathsf{alg}}}[\sum_{t=1}^T \|\mathbf{x}_t\|^2 + \|\mathbf{u}_t\|^2] - TJ^\star_{A,\tilde{B}}, \quad \text{where } B = \iota(\tilde{B}) \\
&\ge \max_{(A,B)\in\tilde{\mathcal{U}}} \mathbb{E}_{A,\tilde{B}}[\mathrm{Regret}_T(\tilde{\mathsf{alg}})],
\end{aligned}$$

as needed. □

H.2 Lower Bound that Maintains Controllability

In this section, we state a lower bound that maintains controllability, in order to demonstrate how controllability does not ameliorate the requirement that the input dimension d_u be bounded. To capture this scenario, set

$$d_x \le d_u, \quad A_\star = \frac{1}{2}I, \quad B_\star = \begin{bmatrix} I_{d_x} & 0_{d_x} \end{bmatrix}. \tag{H.3}$$

Theorem H.1. *Let c, c' be universal constants. Let $c, c' > 0$ denote universal constants. Fix any trace bound $\gamma \ge 1$ and input dimension $d_u \in \mathbb{N}$ with $d_u \ge \gamma$. Consider the set $\mathcal{U}$ of instances with state dimension $d_x = \lfloor \gamma \rfloor$ defined by*

$$\mathcal{U} := \left\{ (A, B) := \|A - \frac{1}{2}I\|_{\mathsf{HS}} \le \frac{1}{4}, \quad \|B - B_\star\|_{\mathsf{HS}} \le \frac{1}{4} \right\}, \quad \text{where } B_\star \text{ is in Eq. (H.3)}.$$

Then, the LQR regret with cost matrices $Q = I_{d_x}$, $R = I_{d_u}$, $\|\Sigma_{\mathbf{w}}\|_{\mathrm{op}} = 1$, and $\mathrm{tr}[\Sigma_{\mathbf{w}}] \le \gamma$, satisfies,

$$\min_{\mathsf{alg}} \max_{(A,B)\in\mathcal{U}} \mathbb{E}_{A,B}[\mathrm{Regret}_T(\mathsf{alg})] \ge \sqrt{\gamma d_u^2 \cdot T},$$

for all $T \ge c'\gamma d_u^2$. In particular, if $T \propto \gamma d_u^2$, the minimax regret on $\mathcal{U}$ is linear in T.

Note that, for all instances $(A, B) \in \mathcal{U}$, not only is A stable ($\|A\|_{\mathrm{op}} \le 3/4$), but the column space has rank d_x, and smallest singular value at least $3/4$ (since the first d_x columns of $B_\star$ are the identity, and all matrices are in $\mathcal{U}$ are a bounded perturbation thereof). Hence, the systems $(A, B) \in \mathcal{U}$ are all one-step controllable. Nevertheless, the regret still scales with d_u^2.

The proof of Theorem H.1 is nearly the same as that of Theorem 1.1; the main difference is verifying the bounds on $P_\star$ and $\sigma_{\min}$ required to instantiate Proposition H.1.

Lemma H.4. *Regardless of the choice of dimension, we have that $\|P_\star\|_{\mathrm{op}} \le 4/3$, and $\sigma_{\min}(A_\star + B_\star K_\star) \ge 1/5$.*

Proof of Lemma H.4. For simplicity, we drop the stars in the subscript. Let us characterize the optimal solution. Define $\mathcal{F}(P) := A^\top PA - (A^\top PB)(R + B^\top PB)^{-1}(B^\top PA) + Q - P$. Let J denote the projection onto the subspace spanned by the columns of B, and $J_\perp = I - J$. We guess a solution of the form

$$P = p_1 J + p_2 J_\perp.$$

Next, we show that such a P solves $\mathcal{F}(P) = 0$ when p_1, p_2 are appropriately chosen. With this matrix, $R + B^\top PB = R + p_1 I_{d_u} = (1+p)I_{d_u}$, so $B(R + B^\top PB)^{-1}B^\top = (1+p_1)J$, where J is the projection onto the subspace spanned by the columns of B. Let $J^\perp = I - J$. Note that in the overactuated case, J is the identity. Hence,

$$\begin{aligned}\mathcal{F}(P) &= \frac{P}{4} - \frac{p_1^2}{4(1+p_1)}J - P + I \\ &= (1 + \frac{p_1}{4} - p_1 - \frac{p_1^2}{4(1+p_1)})J + (1 + \frac{p_2}{4} - p_2)J_\perp \\ &= (1 - \frac{3p_1}{4} - \frac{p_1^2}{4(1+p_1)})J + (1 + \frac{-3p_2}{4})J_\perp.\end{aligned}$$

To solve this equation, set $p_2 = \frac{4}{3}$, and set

$$\begin{aligned}&1 - \frac{3p_1}{4} - \frac{p_1^2}{4(1+p_1)} = 0 \\ &-4(1+p_1) + 3p_1(1+p_1) + p_1^2 = 0 \\ &4p_1^2 - p_1 - 4 = 0.\end{aligned}$$

Taking the positive solution of the quadratic equation, $p_1 = \frac{1+\sqrt{65}}{8} \leq 4/3$.

Now, the optimal control policy is $K = -(R + B^\top PB)^{-1}B^\top PA = \frac{1}{2} \cdot \frac{p_1}{1+p_1}B^\top$ (using the form of B, $R = I$, and P), yielding $BK = -\frac{1}{2} \cdot (\frac{p_1}{1+p_2})J$. Hence, $A + BK = \frac{1}{2}\left(J_\perp + \frac{1}{1+p_2}J\right)$, and thus, $\sigma_{\min}(A+BK) \geq \frac{1}{2(1+p_2)} \geq 1/5$. $\square$