Naive Exploration is Optimal for Online LQR

Max Simchowitz [1] **Dylan J. Foster** [2]

Abstract

We consider the problem of online adaptive control of the linear quadratic regulator, where the true system parameters are unknown. We prove new upper and lower bounds demonstrating that the optimal regret scales as $\widetilde{\Theta}(\sqrt{d_{\mathbf{u}}^2 d_{\mathbf{x}} T})$, where T is the number of time steps, $d_{\mathbf{u}}$ is the dimension of the input space, and $d_{\mathbf{x}}$ is the dimension of the system state. Notably, our lower bounds rule out the possibility of a $\mathrm{poly}(\log T)$-regret algorithm, which had been conjectured due to the apparent strong convexity of the problem. Our upper bound is attained by a simple variant of *certainty equivalent control*, where the learner selects control inputs according to the optimal controller for their estimate of the system while injecting exploratory random noise (Mania et al., 2019).

Central to our upper and lower bounds is a new approach for controlling perturbations of Riccati equations called the *self-bounding ODE method*, which we use to derive suboptimality bounds for the certainty equivalent controller synthesized from estimated system dynamics. This in turn enables regret upper bounds which hold for *any stabilizable instance* and scale with natural control-theoretic quantities.

1. Introduction

Reinforcement learning has recently achieved great success in application domains including Atari (Mnih et al., 2015), Go (Silver et al., 2016), and robotics (Lillicrap et al., 2015). All of these breakthroughs leverage data-driven methods for continuous control in large state spaces. Their success, along with challenges in deploying RL in the real world, has led to renewed interest on developing continuous control algorithms with improved reliability and sample efficiency. In particular, on the theoretical side, there has been a push to develop a non-asymptotic theory of data-driven continuous control, with an emphasis on understanding key algorithmic principles and fundamental limits.

In the non-asymptotic theory of reinforcement learning, much attention has been focused on the so-called "tabular" setting where states and actions are discrete, and the optimal rates for this setting are by now relatively well-understood (Jaksch et al., 2010; Dann & Brunskill, 2015; Azar et al., 2017). Theoretical results for continuous control setting have been more elusive, with progress spread across various models (Kakade et al., 2003; Munos & Szepesvári, 2008; Jiang et al., 2017; Jin et al., 2020), but the linear-quadratic regulator (LQR) problem has recently emerged as a candidate for a standard benchmark for continuous control and RL. For tabular reinforcement learning problems, it is widely understood that careful exploration is essential for sample efficiency. Recently, however, it was shown that for the online variant of the LQR problem, relatively simple exploration strategies suffice to obtain the best-known performance guarantees (Mania et al., 2019). In this paper, we address a curious question raised by these results: Is sophisticated exploration helpful for LQR, or is linear control in fact substantially easier than the general reinforcement learning setting? More broadly, we aim to shed light on the question:

To what extent to do sophisticated exploration strategies improve learning in online linear-quadratic control?

Is ε-Greedy Optimal for Online LQR? In the LQR problem, the system state $\mathbf{x}_t$ evolves according to

$$\mathbf{x}_{t+1} = A\mathbf{x}_t + B\mathbf{u}_t + \mathbf{w}_t, \quad \text{where} \quad \mathbf{x}_1 = 0, \qquad (1.1)$$

and where $\mathbf{u}_t \in \mathbb{R}^{d_{\mathbf{u}}}$ is the learner's control input, $\mathbf{w}_t \in \mathbb{R}^{d_{\mathbf{x}}}$ is a noise process drawn as $\mathbf{w}_t \overset{\text{i.i.d.}}{\sim} \mathcal{N}(0, I)$, and $A \in \mathbb{R}^{d_{\mathbf{x}} \times d_{\mathbf{x}}}$, $B \in \mathbb{R}^{d_{\mathbf{x}} \times d_{\mathbf{u}}}$ are unknown system matrices.

Initially the learner has no knowledge of the system dynamics, and their goal is to repeatedly select control inputs and observing states over T rounds so as to minimize their total cost $\sum_{t=1}^{T} c(\mathbf{x}_t, \mathbf{u}_t)$, where $c(x, u) = x^\top R_{\mathbf{x}} x + u^\top R_{\mathbf{u}} u$ is a known quadratic function. In the online variant of the LQR problem, we measure performance via *regret* to the

[1]UC Berkeley [2]Massachusetts Institute of Technology. Correspondence to: Max Simchowitz <msimchow@berkeley.edu>.

Proceedings of the 37^{th} International Conference on Machine Learning, Vienna, Austria, PMLR 119, 2020.

optimal linear controller:

$$\mathrm{Regret}_{A,B,T}[\pi] = \left[\sum_{t=1}^{T} c(\mathbf{x}_t, \mathbf{u}_t)\right] - T \min_K \mathcal{J}_{A,B}[K], \tag{1.2}$$

where K is a linear state feedback policy and—letting $\mathbb{E}_{A,B,K}[\cdot]$ denote expectation under this policy—where

$$\mathcal{J}_{A,B}[K] := \lim_{T\to\infty} \frac{1}{T}\mathbb{E}_{A,B,K}\left[\sum_{t=1}^{T} c(\mathbf{x}_t, \mathbf{u}_t)\right],$$

is the average infinite-horizon cost of K, which is finite as long as K is *stabilizing* in the sense that $\rho(A+BK) < 1$, where $\rho(\cdot)$ denotes the spectral radius.[1] We further define $\mathcal{J}^\star_{A,B} := \min_K \mathcal{J}_{A,B}[K]$.

This setting has enjoyed substantial development beginning with the work of (Abbasi-Yadkori & Szepesvári, 2011), and following a line of successive improvements (Dean et al., 2018; Faradonbeh et al., 2018a; Cohen et al., 2019; Mania et al., 2019), the best known algorithms for online LQR have regret scaling as $\sqrt{T}$.

We investigate a question that has emerged from this research: The role of exploration in linear control. The first approach in this line of work, (Abbasi-Yadkori & Szepesvári, 2011), proposed a sophisticated though computationally inefficient strategy based on *optimism in the face of uncertainty*, upon which (Cohen et al., 2019) improved to ensure optimal $\sqrt{T}$-regret and polynomial runtime. Another approach which enjoys $\sqrt{T}$-regret, due to (Mania et al., 2019), employs a variant of the classical ε-greedy exploration strategy (Sutton & Barto, 2018) known in control literature as *certainty equivalence*: At each timestep, the learner computes the greedy policy for the current estimate of the system dynamics, then follows this policy, adding exploration noise proportional to ε. While appealing in its simplicity, ε-greedy has severe drawbacks for general reinforcement learning problems: For tabular RL, it leads to exponential blowup in the time horizon (Kearns et al., 2000), and for multi-armed bandits, bandit linear optimization, and contextual bandits, it leads to suboptimal dependence on the time horizon T (Langford & Zhang, 2007).

This begs the question: Can we improve beyond $\sqrt{T}$ regret for online LQR using more sophisticated exploration strategies? Or is exploration in LQR simply much easier than in general reinforcement learning settings? One natural hope would be to achieve logarithmic (i.e. $\mathrm{poly}(\log T)$) regret. After all, online LQR has strongly convex loss functions, and this is a sufficient condition for logarithmic regret in many simpler online learning and optimization problems (Vovk, 2001; Hazan et al., 2007; Rakhlin & Sridharan, 2014), as well as LQR with known dynamics but potentially changing costs (Agarwal et al., 2019b). More subtly, the $\sqrt{T}$ online LQR regret bound of (Mania et al., 2019) requires that the pair $(A_\star, B_\star)$ be *controllable*;[2] it was not known if naive exploration attains this rate for arbitrary *stabilizable* problem instances, or if it necessarily leverages controllability to ensure its efficiency.

[1] For potentially asymmetric matrix $A \in \mathbb{R}^{d\times d}$, $\rho(A) := \max\{|\lambda| \mid \lambda \text{ is an eigenvalue for } A\}$.

1.1. Contributions

We prove new upper and lower bounds which characterize the minimax optimal regret for online LQR as $\widetilde{\Theta}(\sqrt{d_{\mathbf{u}}^2 d_{\mathbf{x}} T})$. Beyond dependence on the horizon T, dimensions $d_{\mathbf{x}}, d_{\mathbf{u}}$, and logarithmic factors, our bounds depend only on *operator* norms of transparent, control theoretic quantities, which do not hide additional dimension dependence. Our main lower bound is Theorem 1, which implies that no algorithm can improve upon $\sqrt{T}$ regret for online LQR, and so simple ε-greedy exploration is indeed *rate-optimal*.

Theorem 1 (informal). For every sufficiently non-degenerate problem instance and every (potentially randomized) algorithm, there exists a nearby problem instance on which the algorithm must suffer regret at least $\widetilde{\Omega}(\sqrt{d_{\mathbf{u}}^2 d_{\mathbf{x}} T})$.

Perhaps more surprisingly, our main upper bound shows that a simple variant of certainty equivalence is also *dimension-optimal*, in that it asymptotically matches the $\sqrt{d_{\mathbf{u}}^2 d_{\mathbf{x}} T}$ lower bound of Theorem 1.

Theorem 2 (informal). Certainty equivalent control with continual ε-greedy exploration (Algorithm 1) has regret at most $\widetilde{O}\left(\sqrt{d_{\mathbf{u}}^2 d_{\mathbf{x}} T} + d_{\mathbf{x}}^2\right)$ for every stabilizable online LQR instance.

Our upper bound *does not* require controllability, and is the first bound for *any* algorithm to attain the optimal dimension dependence. In comparison, result of (Mania et al., 2019) guarantees $\sqrt{(d_{\mathbf{x}} + d_{\mathbf{u}})^3 T}$ regret *and* imposes strong additional assumptions. In the many control settings where $d_{\mathbf{u}} \ll d_{\mathbf{x}}$, our bound constitutes a significant improvement. Other approaches *not* based on certainty equivalence suffer considerably larger dimension dependence (Cohen et al., 2019). Together, Theorem 1 and Theorem 2 characterize the asymptotic minimax regret for online LQR, showing that there is little room for improvement over naive exploration.

Our results leverage a new perturbation bound for controllers synthesized via certainty equivalence. Unlike prior bounds due to (Mania et al., 2019), our guarantee depends

[2] $(A_\star, B_\star)$ are said to be controllable if and only the *controllability Gramian* $\mathcal{C}_n\mathcal{C}_n^\top := \sum_{i=0}^n A_\star^i B_\star B_\star^\top (A_\star^i)^\top$ is strictly positive definite for some $n \geq 0$. For any n for which $\mathcal{C}_n \succ 0$, the upper bounds of (Mania et al., 2019) scale polynomially in $n, 1/\lambda_{\min}(\mathcal{C}_n\mathcal{C}_n^\top)$. Controllability implies stabilizability, but the converse is not true.

only on natural control-theoretic quantities, and crucially does not require controllability of the system.

Theorem 3 (informal). Fix an instance (A, B). Let $(\widehat{A}, \widehat{B})$, and let $\widehat{K}$ denote the optimal infinite horizon controller from instance $(\widehat{A}, \widehat{B})$. Then if $(\widehat{A}, \widehat{B})$ are sufficiently close to (A, B), we have

$$\mathcal{J}_{A,B}[\widehat{K}] - \mathcal{J}^{\star}_{A,B} \leq 142\|P\|_{\mathrm{op}}^{8} \cdot (\|\widehat{A} - A\|_{\mathrm{F}}^{2} + \|\widehat{B} - B\|_{\mathrm{F}}^{2}),$$

where P is the solution to the DARE for the system (A, B).

For simplicity, the bound above assumes the various normalization conditions on the noise and cost matrices, described in Section 1.4. With these conditions, our perturbation bound only requires that the operator norm distance between $(\widehat{A}, \widehat{B})$ and (A, B) be at most $1/\mathrm{poly}(\|P\|_{\mathrm{op}})$. Hence, we establish perturbation bounds for which both the scaling of the deviation and the region in which the bound applies can be quantified in terms of a single quantity: the norm of DARE solution P. We prove this bound through a new technique we term the *Self-Bounding ODE method*, described below. Beyond removing the requirement of controllability, we believe this method is simpler and more transparent than past approaches.

1.2. Our Approach

Both our lower and upper bounds are facilitated by the *self-bounding ODE method*, a new technique for establishing perturbation bounds for the Riccati equations that characterize the optimal value function and controller for LQR. The method sharpens existening perturbation bounds, weakens controllability and stability assumptions required by previous work (Dean et al., 2018; Faradonbeh et al., 2018a; Cohen et al., 2019; Mania et al., 2019), and yields an upper bound whose leading terms depend only on the horizon T, dimension parameters $d_{\mathbf{x}}, d_{\mathbf{u}}$, and the control-theoretic parameters sketched in the prequel.

In more detail, if (A, B) is stabilizable and $R_{\mathbf{x}}, R_{\mathbf{u}} \succ 0$, there exists a unique PSD solution $P_{\infty}(A, B)$ for the *discrete algebraic Riccati equation* (DARE),

$$P = A^{\top}PA + R_{\mathbf{x}} - A^{\top}PB(R_{\mathbf{u}} + B^{\top}PB)^{-1}B^{\top}PA \quad (1.3)$$

The unique optimal infinite-horizon controller is given by

$$K_{\infty}(A, B) = -(R_{\mathbf{u}} + B^{\top}P_{\infty}(A, B)B)^{-1}B^{\top}P_{\infty}(A, B)A,$$

and the matrix $P_{\infty}(A, B)$ induces a positive definite quadratic form which can be interpreted as a value function for the LQR problem.

Both our upper and lower bounds make use of novel perturbation bounds to control the change in P_{∞} and K_{∞} when we move from a nominal instance (A, B) to a nearby instance $(\widehat{A}, \widehat{B})$. For our upper bound, these are used to show that a good estimator for the nominal instance leads to a good controller, while for our lower bounds, they show that the converse is true. The self-bounding ODE method allows us to prove perturbation guarantees that depend only on the norm of the value function $\|P_{\infty}(A, B)\|_{\mathrm{op}}$ for the nominal instance, which is a weaker assumption that subsumes previous conditions. The key observation underpinning the method is that the norm of the directional derivative of $\frac{d}{dt}P_{\infty}(A(t), B(t))\big|_{t=u}$ at a point $t = u$ along a line $(A(t), B(t))$ is bounded in terms of the magnitude of $\|P_{\infty}(A(u), B(u))\|$; we call this the *self-bounding* property. From this relation, we show that bounding the norm of the derivatives reduces to solving a scalar ordinary differential equation, whose derivative saturates the scalar analogue of this self-bounding property. Notably, this technique does not require that the system be controllable, and in particular does not yield guarantees which depend on the smallest singular value of the controllability matrix as in (Mania et al., 2019). Moreover, given estimates $(\widehat{A}, \widehat{B})$ and an upper-bound on their deviation from the true system $(A_{\star}, B_{\star})$, our bound allows the learner to check whether the certainty-equivalent controller synthesized from $\widehat{A}, \widehat{B}$ stabilizes the true system and satisfies the preconditions for our perturbation bounds.

On the lower bound side, we begin with a nominal instance (A_0, B_0) and consider a packing of alternative instances within a small neighborhood. Specifically, if K_0 is the optimal controller for (A_0, B_0), we consider perturbations of the form $(A_{\Delta}, B_{\Delta}) = (A_0 - \Delta K_0, B_0 + \Delta)$ for $\Delta \in \mathbb{R}^{d_{\mathbf{u}} d_{\mathbf{x}}}$. The self-bounding ODE method facilitates a perturbation analysis which implies that the optimal controller K_{Δ} on each alternative (A_{Δ}, B_{Δ}) deviates from K_0 by $\|K_0 - K_{\Delta}\|_{\mathrm{F}} \geq \Omega(\|\Delta\|_{\mathrm{F}})$ for non-degenerate instances. Using this reasoning, we show that any low-regret algorithm can approximately recover the perturbation Δ.

On the other hand, if the learner selects inputs $\mathbf{u}_t = K_0\mathbf{x}_t$ according to the optimal control policy for the nominal instance, all alternatives are *indistinguishable* from the nominal instance. Indeed, the structure of our perturbations ensures that $A_{\Delta} + B_{\Delta}K_0 = A_0 + B_0K_0$ for all choices of Δ. Thus, since low regret implies identification of the perturbation, any low regret learner must substantially deviate from the nominal controller K_0. Equivalently, this can be understood as a consequence of the fact that playing $\mathbf{u}_t = K_0\mathbf{x}_t$ yields a degenerate covariance matrix for the random variable $(\mathbf{x}_t, \mathbf{u}_t)$, and thus some deviation from K_0 is required to ensure this covariance is full rank. The regret scales proportionally to the deviation from K_0, which scales proportionally to the minimum eigenvalues of the aforementioned covariance matrix, but the estimation error rate scales as $1/T$ (the typical "fast rate") times the *inverse* of these eigenvalues. Balancing the tradeoffs leads to the

"slow" $\sqrt{T}$ lower bound. Crucially, our argument exploits a fundamental tension between control and indentification in linear systems, first described by Polderman (1986), and summarized in Polderman (1989).

Our upper bound refines the certainty equivalent control strategy proposed in (Mania et al., 2019) by re-estimating the system parameters on a doubling epoch schedule to advantage of the endogenous excitation supplied by the $\mathbf{w}_t$-sequence. A careful analysis of the least squares estimator shows that the error in a $d_{\mathbf{x}}d_{\mathbf{u}}$-dimensional subspace decays as $\mathcal{O}\left(1/\sqrt{t}\right)$, and in the remaining $d_{\mathbf{x}}^2$ dimensions decays at a *fast rate* of $\mathcal{O}\left(1/t\right)$.

Related Work Non-asymptotic guarantees for learning linear dynamical systems have been the subject of intense recent interest (Dean et al.; Hazan et al., 2017; Tu & Recht, 2018; Hazan et al., 2018; Simchowitz et al., 2018; Sarkar & Rakhlin, 2019; Simchowitz et al., 2019; Mania et al., 2019; Sarkar et al., 2019). The online LQR setting we study was introduced by (Abbasi-Yadkori & Szepesvári, 2011), which considers the problem of controlling an unknown linear system under stationary stochastic noise.[3] They showed that an algorithm based on the optimism in the face of uncertainty (OFU) principle enjoys $\sqrt{T}$, but their algorithm is computationally inefficient and their regret bound depends exponentially on dimension. The problem was revisited by (Dean et al., 2018), who showed that an explicit explore-exploit scheme based on ε-greedy exploration and certainty equivalence achieves $T^{2/3}$ regret efficiently, and left the question of obtaining $\sqrt{T}$ regret efficiently as an open problem. This issue was subsequently addressed by (Faradonbeh et al., 2018a) and (Mania et al., 2019), who showed that certainty equivalence obtains $\sqrt{T}$ regret, and (Cohen et al., 2019), who achieve $\sqrt{T}$ regret using a semidefinite programming relaxation for the OFU scheme. The regret bounds in (Faradonbeh et al., 2018a) do not specify dimension dependence, and (for $d_{\mathbf{x}} \geq d_{\mathbf{u}}$), the dimension scaling of (Cohen et al., 2019) can be as large as $\sqrt{d_{\mathbf{x}}^{16}T}$;[4] (Mania et al., 2019) incurs an almost-optimal dimension dependence of $\sqrt{d_{\mathbf{x}}^3 T}$ (suboptimal when $d_{\mathbf{u}} \ll d_{\mathbf{x}}$), but at the expense of imposing a strong controllability assumption.

The question of whether regret for online LQR could be improved further (for example, to $\log T$) remained open, and was left as a conjecture by (Faradonbeh et al., 2018b). Our lower bounds resolve this conjecture by showing that $\sqrt{T}$-regret is optimal. Moreover, by refining the upper bounds of (Mania et al., 2019), our results show that the asymptotically optimal regret is $\widetilde{\Theta}(\sqrt{d_{\mathbf{u}}^2 d_{\mathbf{x}} T})$, and that this achieved by certainty equivalence. Beyond attaining the optimal dimension dependence, our upper bounds also enjoy refined dependence on problem parameters, and do not require a-priori knowledge of these parameters.

Logarithmic regret bounds are ubiquitous in online learning and optimization problems with strongly convex loss functions (Vovk, 2001; Hazan et al., 2007; Rakhlin & Sridharan, 2014). (Agarwal et al., 2019b) demonstrate that for the problem of controlling an *known* linear dynamic system with adversarially chosen, strongly convex costs, logarithmic regret is also attainable. Our $\sqrt{T}$ lower bound shows that the situation for the online LQR with an *unknown* system parallels that of bandit convex optimization, where (Shamir, 2013) showed that $\sqrt{T}$ is optimal even for strongly convex quadratics. That is, in spite of strong convexity of the losses, issues of partial observability prevent fast rates in both settings.

Our lower bound carefully exploits the online LQR problem structure to show that $\sqrt{T}$ is optimal. To obtain optimal dimension dependence for the lower bound, we build on well-known lower bound technique for adaptive sensing based on Assouad's lemma (Arias-Castro et al., 2012) (see also (Assouad, 1983; Yu, 1997)).

Finally, a parallel line of research provides Bayesian and frequentist regret bounds for online LQR based on Thompson sampling (Ouyang et al., 2017; Abeille & Lazaric, 2017), with (Abeille & Lazaric, 2018) demonstrating $\sqrt{T}$-regret for the scalar setting. Unfortunately, Thompson sampling is not computationally efficient for the LQR.

1.3. Organization

Section 1.4 introduces basic notation and definitions. Section 2 introduces our main results: In Section 2.1 and Section 2.2 we state our main lower and upper bounds respectively and give an overview of the proof techniques, and in Section 2.3 we instantiate and compare these bounds for the simple special case of strongly stable systems. In Section 3 we introduce the self-bounding ODE method and show how it is used to prove key perturbation bounds used in our main results. All additional proofs and proof details are given in the appendix, whose organization is described at length in Appendix A. Future directions and open problems are discussed in Section 4.

1.4. Preliminaries

Assumptions We restrict our attention *stabilizable* systems (A, B) for which there exists a stabilizing controller K such that $\rho(A + BK) < 1$. Note that this does not require that the system be controllable. We further assume that $R_{\mathbf{u}} = I$ and $R_{\mathbf{x}} \succeq I$. The first can be enforced by a change of basis in input space, and the second can be enforced by

[3] A more recent line of work studies a more general *non-stochastic* noise regime (see (Agarwal et al., 2019a) et seq.), which we do not consider in this work.

[4] The regret bound of (Cohen et al., 2019) scales as $d_{\mathbf{x}}^3\sqrt{T} \cdot (\mathcal{J}^\star_{A_\star,B_\star})^5$; typically, $\mathcal{J}^\star_{A_\star,B_\star}$ scales linearly in $d_{\mathbf{x}}$

rescaling the state space, increasing the regret by at most a multiplicative factor of $\min\{1, 1/\sigma_{\min}(R_{\mathbf{x}})\}$. We also assume that the process noise $\mathbf{w}_t$ has identity covariance. We note that non-identity noise can be adressed via a change of variables, and in Appendix I.8 we sketch extensions of our results to (a) independent, sub-Gaussian noise with bounded below covariance, and (b) more general martingale noise, where we remark on how to achieve optimal rates in the regime $d_{\mathbf{x}} \lesssim d_{\mathbf{u}}^2$.

Algorithm Protocol and Regret Formally, the learner's (potentially randomized) decision policy is modeled as a sequence of mappings $\pi = (\pi_t)_{t=1}^T$, where each function π_t maps the history $(\mathbf{x}_1, \dots, \mathbf{x}_t, \mathbf{u}_1, \dots, \mathbf{u}_{t-1})$ and an internal random seed ξ to an output control signal $\mathbf{u}_t$. For a linear system evolving according to Eq. (1.1) and policy π, we let $\mathbb{P}_{A,B,\pi}$ and $\mathbb{E}_{A,B,\pi}[\cdot]$ denote the probability and expectation with respect to the dynamics (1.1) and randomization of π. For such a policy, we use the notation $\mathrm{Regret}_{A,B,T}[\pi]$ as in Eq. (1.2) for regret, which is a random variable with law $\mathbb{P}_{A,B,\pi}[\cdot]$. We prove high-probability upper bounds on $\mathrm{Regret}_{A,B,T}[\pi]$, and prove lower bounds on the expected regret $\mathbb{E}\mathrm{Regret}_{A,B,T}[\pi] := \mathbb{E}_{A,B,\pi}[\mathrm{Regret}_{A,B,T}[\pi]]$.[5]

Additional Notation For vectors $x \in \mathbb{R}^d$, $\|x\|$ denotes the ℓ_2 norm. For matrices $X \in \mathbb{R}^{d_1 \times d_2}$, $\|X\|_{\mathrm{op}}$ denotes the spectral norm, and $\|X\|_{\mathrm{F}}$ the Frobenius norm. When $d_1 \le d_2$, $\sigma_1(X), \dots, \sigma_{d_1}(X)$ denote the singular values of X, arranged in decreasing order. We say $f \lesssim g$ to denote that $f(x) \le Cg(x)$ for a universal constant C, and $f \lessapprox g$ to denote informal inequality. We write $f \eqsim g$ if $g \lesssim f \lesssim g$.

For "starred" systems $(A_\star, B_\star)$, we adopt the shorthand $P_\star := P_\infty(A_\star, B_\star)$, $K_\star := K_\infty(A_\star, B_\star)$ for the optimal controller, $\mathcal{J}_\star := \mathcal{J}^\star_{A_\star,B_\star} := \mathcal{J}_{A_\star,B_\star}[K_\star]$ for optimal cost, and $A_{\mathrm{cl},\star} := A_\star + B_\star K_\star$ for the optimal closed loop system. We define $\Psi_\star := \max\{1, \|A_\star\|_{\mathrm{op}}, \|B_\star\|_{\mathrm{op}}\}$ and $\Psi_{B_\star} := \max\{1, \|B_\star\|_{\mathrm{op}}\}$. For systems (A_0, B_0), we let $\mathcal{B}_{\mathrm{op}}(\epsilon; A_0, B_0) = \{(A, B) \mid \|A - A_0\|_{\mathrm{op}} \vee \|B - B_0\|_{\mathrm{op}} \le \epsilon\}$ denote the set of nearby systems in operator norm.

2. Main Results

We now state our main upper and lower bounds for online LQR and give a high-level overview of the proof techniques behind both results. At the end of the section, we instantiate and compare the two bounds for the simple special case of strongly stable systems.

[5] One might consider as a stronger benchmark described the expected loss of the optimal policy for *fixed horizon* T. A fortiori, our lower bounds apply for this benchmark as well: In view of the proof of Lemma F.3 in Appendix G.2, this benchmark differs from $T\,\mathcal{J}^\star_{A,B}$ by a constant factor which depends on (A, B) but *does not grow with* T.

Both our upper and lower bounds start with the following question: Suppose that the learner is selecting near optimal control inputs $\mathbf{u}_t \approx K_\star \mathbf{x}_t$, where $K_\star = K_\infty(A_\star, B_\star)$ is the optimal controller for the system $(A_\star, B_\star)$. What information can she glean about the system?

2.1. Lower Bound

We provide a *local minimax* lower bound, which captures the difficulty of ensuring low regret on both a *nominal instance* $(A_\star, B_\star)$ and on the hardest nearby alternative. For a distance parameter $\epsilon > 0$, we define the local minimax complexity at scale ϵ as

$$\mathcal{R}_{A_\star,B_\star,T}(\epsilon) := \min_{\pi} \max_{A,B} \Big\{ \mathbb{E}\mathrm{Regret}_{A,B,T}[\pi] : \|A - A_\star\|_{\mathrm{F}}^2 \vee \|B - B_\star\|_{\mathrm{F}}^2 \le \epsilon \Big\}.$$

Local minimax complexity captures the idea certain instances $(A_\star, B_\star)$ are more difficult than others, and allows us to provide lower bounds that scale only with control-theoretic parameters of the nominal instance. Of course, the local minimax lower bound immediately implies a lower bound on the global minimax complexity as well.[6]

Intuition Behind the Lower Bound. We show that if the learner plays near-optimally on every instance in the neighborhood of $(A_\star, B_\star)$, then there is a $d_{\mathbf{x}} d_{\mathbf{u}}$-dimensional subspace of system parameters that the learner must explore by deviating from $K_\star$ when the underlying instance is $(A_\star, B_\star)$. Even though the system parameters can be estimated at a fast rate, such deviations preclude logarithmic regret.

In more detail, if the learner plays near-optimally, she is not be able to distinguish between whether the instance she is interacting with is $(A_\star, B_\star)$, or another system of the form

$$(A, B) = (A_\star - K_\star \Delta, B_\star + \Delta), \tag{2.1}$$

for some perturbation $\Delta \in \mathbb{R}^{d_{\mathbf{x}} \times d_{\mathbf{u}}}$. This is because all the obsevations $(\mathbf{x}_t, \mathbf{u}_t)$ generated by the optimal controller lie in the subspace $\{(x, u) : u - K_\star x = 0\}$, and likewise all observations generated by any near-optimal controller approximately lie in this subspace. Since the learner cannot distinguish between $(A_\star, B_\star)$ and (A, B), she will also play $\mathbf{u}_t \approx K_\star \mathbf{x}_t$ on (A, B). This leads to poor regret when the instance is (A, B), since the optimal controller in this case has $\mathbf{u}_t = K_\infty(A, B)\mathbf{x}_t$. This is made concrete by the next lemma, which shows to a first-order approximation that if Δ is large, the distance between $K_\star$ and $K_\infty(A, B)$ must also be large.

[6] Some care must be taken in defining the global complexity, or it may well be infinite. One sufficient definition, which captures prior work, is to consider minimax regret over all instances subject to a global bound on $\|P_\star\|$, $\|B_\star\|$, and so on.

Lemma 2.1 (Derivative Computation (Abeille & Lazaric (2018), Proposition 2)). *Let $(A_\star, B_\star)$ be stabilizable, and recall $A_{\mathrm{cl},\star} := A_\star + B_\star K_\star$. Then,*

$$\frac{d}{dt} K_\infty(A_\star - t\Delta K_\star, B_\star + t\Delta)\big|_{t=0} = -(R_{\mathbf{u}} + B_\star^\top P_\star B_\star)^{-1} \cdot \Delta^\top P_\star A_{\mathrm{cl},\star}.$$

In particular, when the closed loop system $A_{\mathrm{cl},\star}$ is (approximately) well-conditioned, the optimal controllers for $(A_\star, B_\star)$ and for (A, B) are $\Omega(\|\Delta\|_{\mathrm{F}})$-apart, and so the learner cannot satisfy both $\mathbf{u}_t \approx K_\star \mathbf{x}_t$ and $\mathbf{u}_t \approx K_\infty(A,B)\mathbf{x}_t$ simultaneously. More precisely, for the learner to ensure $\sum_t \|\mathbf{x}_t - K_\infty(A,B)\mathbf{u}_t\|_{\mathrm{F}}^2 \lessapprox d_{\mathbf{x}} d_{\mathbf{u}} \epsilon^2$ on every instance, she must deviate from optimal by at least $\sum_{t=1}^T \|\mathbf{x}_t - K_\star \mathbf{u}_t\|_{\mathrm{F}}^2 \gtrapprox d_{\mathbf{u}} T/\epsilon^2$ on the optimal instance; the $d_{\mathbf{u}}$ factor here comes from the necessity of exploring all control-input directions. Balancing these terms leads to the final $\Omega(\sqrt{T d_{\mathbf{u}} d_{\mathbf{x}}^2})$ lower bound (proven in Appendix F).

Theorem 1. *Let $c_1, p > 0$ denote universal constants. For $m \in [d_{\mathbf{x}}]$, define $\nu_m := \sigma_m(A_{\mathrm{cl},\star})/\|R_{\mathbf{u}} + B_\star^\top P_\star B_\star\|_{\mathrm{op}}$. Then if $\nu_m > 0$, we have*

$$\mathcal{R}_{A_\star, B_\star, T}(\epsilon_T) \gtrsim \sqrt{d_{\mathbf{u}}^2 m T} \cdot \frac{1 \wedge \nu_m^2}{\|P_\star\|_{\mathrm{op}}^2},$$

where $\epsilon_T = \sqrt{d_{\mathbf{u}}^2 m / T}$, provided that T is at least $c_1(\|P_\star\|_{\mathrm{op}}^p (d_{\mathbf{u}} m \vee \frac{d_{\mathbf{x}}^2 \Psi_{B_\star}^4 (1 \vee \nu_m^{-4})}{m d_{\mathbf{u}}^2} \vee d_{\mathbf{x}} \log(1 + d_{\mathbf{x}} \|P_\star\|_{\mathrm{op}}))$.

Let us briefly discuss some key features of Theorem 1.

- The only system-dependent parameters appearing in the lower bound are the operator norm bounds $\Psi_{B_\star}$ and $\|P_\star\|_{\mathrm{op}}$, which only depend on the nominal instance. The latter parameter is finite whenever the system is stabilizable, and does not explicitly depend on the spectral radius or strong stability parameters.
- The lower bound takes $\epsilon_T \propto T^{-1/2}$, so the alternative instances under consideration converge to the nominal instance $(A_\star, B_\star)$ as $T \to \infty$.
- The theorem can be optimized for each instance by tuning the dimension parameter $m \in [d_{\mathbf{x}}]$: The leading $\sqrt{d_{\mathbf{u}}^2 m T}$ term is increasing in m, while the parameter ν_m scales with $\sigma_m(A_{\mathrm{cl},\star})$ and thus is decreasing in m. The simplest case is when $\sigma_m(A_{\mathrm{cl},\star})$ is bounded away from 0 for $m \gtrsim d_{\mathbf{x}}$; here we obtain the optimal $\sqrt{d_{\mathbf{u}}^2 d_{\mathbf{x}} T}$ lower bound. In particular, if $d_{\mathbf{u}} \le d_{\mathbf{x}}/2$, we can choose $m = \frac{1}{2} d_{\mathbf{x}}$ to get $\sigma_m(A_{\mathrm{cl},\star}) \ge \sigma_{\min}(A_\star)$.

2.2. Upper Bound

While playing near-optimally prevents the learner from ruling out perturbations of the form Eq. (2.1), she can rule perturbations in orthogonal directions. Indeed, if $\mathbf{u}_t \approx K_\star \mathbf{x}_t$, then $\mathbf{x}_{t+1} \approx (A_\star + B_\star K_\star)\mathbf{x}_t + \mathbf{w}_t$. As a result, the persistent noise process $\mathbf{w}_t$ allows the learner recover the closed loop dynamics matrix $A_{\mathrm{cl},\star} = A_\star + B_\star K_\star$ to Frobenius error $d_{\mathbf{x}}\epsilon$ after just $T \gtrapprox 1/\epsilon^2$ steps, regardless of whether she incorporates additional exploration (Simchowitz et al., 2018). Hence, for perturbations perpendicular to those in Eq. (2.1), the problem closely resembles a setting where $\log T$ is achievable.

Our main algorithm, Algorithm 1, is detailed in Appendix H. It is an ε-greedy scheme that takes advantage of this principle. The full pseudocode and analysis are deferred to Appendix H, but we sketch the intuition here. The algorithm takes as input a stabilizing controller K_0 and proceeds in epochs k of length $\tau_k = 2^k$. After an initial burn-in period ending with epoch k_{safe}, the algorithm can ensure the reliability of its synthesized controllers, and uses a (projected) least-squares estimate $(\widehat{A}_k, \widehat{B}_k)$ of $(A_\star, B_\star)$ to synthesize a controller $\widehat{K}_k = K_\infty(\widehat{A}_k, \widehat{B}_k)$ known as the *certainty equivalent* controller. The learner then selects inputs by adding white Gaussian noise with variance σ_k^2: $\mathbf{u}_t = \widehat{K}_t \mathbf{x}_t + \mathcal{N}(0, \sigma_k^2 I)$. We show that this scheme exploits the rapid estimation along directions orthogonal to those in Eq. (2.1), leading to optimal dimension dependence.

To begin, we show (Theorem 3) that the cost of the certainty-equivalent controller is bounded by the estimation error for $\widehat{A}_k$ and $\widehat{B}_k$, i.e.

$$\mathcal{J}_{A_\star, B_\star}[\widehat{K}_k] - \mathcal{J}_\star \lesssim \mathrm{poly}(\|P_\star\|_{\mathrm{op}}) \cdot (\|\widehat{A}_k - A_\star\|_{\mathrm{F}}^2 + \|\widehat{B}_k - B_\star\|_{\mathrm{F}}^2),$$

once $(\widehat{A}_k, \widehat{B}_k)$ are sufficiently accurate, as guaranteed by the burn-in period. Through a regret decomposition based on the Hanson-Wright inequality (Lemma I.1), we next show that the bulk of the algorithm's regret scales as the sum of the suboptimality in the controller for a given epoch, plus the cost of the exploratory noise: $\sum_{k=k_{\mathrm{safe}}}^{\log_2 T} \tau_k \left(\mathcal{J}_{A_\star, B_\star}[\widehat{K}_k] - \mathcal{J}_\star\right) + d_{\mathbf{u}} \tau_k \sigma_k^2 \lessapprox \sum_{k=k_{\mathrm{safe}}}^{\log_2 T} \tau_k \left(\|\widehat{A}_k - A_\star\|_{\mathrm{F}}^2 + \|\widehat{B}_k - B_\star\|_{\mathrm{F}}^2\right) + d_{\mathbf{u}} \tau_k \sigma_k^2$. In the above, we also incur a term of approximately $\sum_{k=k_{\mathrm{safe}}}^{\log_2 T} \sqrt{(d_{\mathbf{x}} + d_{\mathbf{u}})\tau_k} \lesssim \sqrt{T(d_{\mathbf{x}} + d_{\mathbf{u}})}$, which is lower order than the overall regret of $\sqrt{T d_{\mathbf{x}} d_{\mathbf{u}}^2}$. This term arises from the random fluctuations of the costs around their expectation, and crucially, the Hanson-Wright inequality allows us to pay of the *square root* of the dimension.[7]

[7] The use of the Hanson-Wright crucially leverages independence of the noise process; for general sub-Gaussian martingale noise, an argument based on martingale concentration would mean that the fluctuations contribute $(d_{\mathbf{x}} + d_{\mathbf{u}})\sqrt{T}$ to the regret up to logarithmic factors, yielding an overall regret of $\sqrt{\max\{d_{\mathbf{x}}, d_{\mathbf{u}}^2\} d_{\mathbf{x}} T}$. This is suboptimal regret for $d_{\mathbf{x}} \gg d_{\mathbf{u}}^2$, but still an improvement over the $\sqrt{(d_{\mathbf{x}} + d_{\mathbf{u}})^3 T}$-bound of (Mania et al., 2019). It is un-

Paralleling the lower bound, the analysis crucially relies on the exploratory noise to bound the error in the $d_{\mathbf{x}}d_{\mathbf{u}}$-dimensional subspace corresponding to Eq. (2.1), as the error in this subspace grows as $\frac{d_{\mathbf{x}}d_{\mathbf{u}}}{\sigma_k^2\tau_k}$. However, for the directions parallel to those in Eq. (2.1), the estimation error is at most $d_{\mathbf{x}}^2/\tau_k$, and so the total regret is bounded as $\mathrm{Regret}_{A_\star,B_\star,T}[\mathsf{Alg}] \lessapprox \sum_{k=k_{\mathrm{safe}}}^{\log_2 T} \tau_k \left(\frac{d_{\mathbf{x}}^2}{\tau_k} + \frac{d_{\mathbf{x}}d_{\mathbf{u}}}{\tau_k\sigma_k^2}\right) + d_{\mathbf{u}}\tau_k\sigma_k^2 \approx d_{\mathbf{x}}^2\log T + \sum_{k=1}^{\log_2 T}\frac{d_{\mathbf{x}}d_{\mathbf{u}}}{\sigma_k^2} + d_{\mathbf{u}}\tau_k\sigma_k^2$. Trading off $\sigma_k^2 = \sqrt{d_{\mathbf{x}}/\tau_k}$ gives regret $d_{\mathbf{x}}^2\log T + \sum_{k=1}^{\log_2 T}\sqrt{d_{\mathbf{x}}d_{\mathbf{u}}^2\tau_k} \approx d_{\mathbf{x}}^2\log T + \sqrt{d_{\mathbf{x}}d_{\mathbf{u}}^2 T}$. We emphasize that to ensure that the $d_{\mathbf{x}}^2$ term in this bound scales only with $\log T$ due to rapid exploration perpendicular to Eq. (2.1), and it is crucial that the algorithm uses doubling epochs to take advantage of this. We now state the full guarantee.

Theorem 2. *When Algorithm 1 is invoked with stabilizing controller K_0 and confidence parameter $\delta \in (0, 1/T)$, it guarantees that with probability at least $1-\delta$, $\mathrm{Regret}_T[\mathsf{Alg}; A_\star, B_\star]$ is bounded as*

$$\lesssim \sqrt{d_{\mathbf{u}}^2 d_{\mathbf{x}} T \cdot \Psi_{B_\star}^2 \|P_\star\|_{\mathrm{op}}^{11} \log\tfrac{\|P_\star\|_{\mathrm{op}}}{\delta}} + d^2\cdot\mathcal{P}_0\Psi_{B_\star}^6\|P_\star\|_{\mathrm{op}}^{11}(1+\|K_0\|_{\mathrm{op}}^2)\log\frac{d\Psi_{B_\star}\mathcal{P}_0}{\delta}\log^2\frac{1}{\delta},$$

where $\mathcal{P}_0 := \mathcal{J}_{A_\star,B_\star}[K_0]/d_{\mathbf{x}}$ is the normalized cost of K_0, and $d = d_{\mathbf{x}} + d_{\mathbf{u}}$.

Ignoring dependence on problem parameters, the upper bound of Theorem 2 scales asymptotically as $\sqrt{d_{\mathbf{u}}^2 d_{\mathbf{x}} T}$, matching our lower bound. Like the lower bound, the theorem depends on the instance $(A_\star, B_\star)$ only through the operator norm bounds $\Psi_{B_\star}$ and $\|B_\star\|_{\mathrm{op}}$. Similar to previous work (Dean et al., 2018; Mania et al., 2019), the regret bound has additional dependence on the stabilizing controller K_0 through $\|K_0\|_{\mathrm{op}}$ and $\mathcal{P}_0$, but these parameters only affect the lower-order terms.

2.3. Consequences for Strongly Stable Systems

To emphasize the dependence on dimension and time horizon in our results, we now present simplified findings for a special class of *strongly stable* systems.

Definition 2.1 (Strongly Stable System (Cohen et al., 2018))**.** We say that $A_\star$ is (γ,κ)-strongly stable if there exists a transform T such that $\|T\|_{\mathrm{op}}\cdot\|T^{-1}\|_{\mathrm{op}} \le \kappa$ and $\|TA_\star T^{-1}\|_{\mathrm{op}} \le 1-\gamma$. When $A_\star$ is (γ,κ)-strongly stable, we define $\gamma_{\mathrm{sta}} := \gamma/\kappa^2$.

For the simplified results in this section we make the following assumption.

clear if one can do better in this setting without improved concentration bounds for quadratic forms of martingale vectors, because it is unclear how an algorithm can ameliorate these random fluctuations.

Assumption 1. *The nominal instance $(A_\star, B_\star)$ is such that $A_\star$ is (γ,κ)-strongly stable and $\|B_\star\|_{\mathrm{op}} \le 1$. Furthermore, $R_{\mathbf{x}} = R_{\mathbf{u}} = I$.*

For strongly stable systems under Assumption 1, our main lower bound (Theorem 1) takes the following particularly simple form.

Corollary 1 (Lower Bound for Strongly Stable Systems)**.** *Suppose that Assumption 1 holds, and that $d_{\mathbf{u}} \le \frac{1}{2}d_{\mathbf{x}}$ and $\sigma_{\min}(A_\star) > 0$.[8] Then for any $T \ge (d_{\mathbf{x}}d_{\mathbf{u}} + d_{\mathbf{x}}\log d_{\mathbf{x}})\mathrm{poly}(1/\gamma_{\mathrm{sta}}, 1/\sigma_{\min}(A_\star))$, we have*

$$\mathcal{R}_{A_\star,B_\star,T}(\varepsilon_T) \gtrsim \sqrt{d_{\mathbf{u}}^2 d_{\mathbf{x}} T}\cdot\sigma_{\min}(A_\star)^2\gamma_{\mathrm{sta}}^4,$$

where $\varepsilon_T := \sqrt{d_{\mathbf{u}}^2 d_{\mathbf{x}}/T}$.

The upper bound from Theorem 2 takes on a similarly simple form, and is seen to be nearly matching.

Corollary 2 (Upper Bound for Strongly Stable Systems)**.** *Suppose that Assumption 1 holds. Then Algorithm 1 with stabilizing controller $K_0 = 0$ and confidence parameter $\delta \in (0, 1/T)$, ensures that probability at least $1-\delta$, $\mathrm{Regret}_T[\mathsf{Alg}; A_\star, B_\star]$ is bounded as*

$$\lesssim \sqrt{d_{\mathbf{u}}^2 d_{\mathbf{x}} T\cdot\gamma_{\mathrm{sta}}^{-11}\log\frac{1}{\delta\gamma_{\mathrm{sta}}}} + (d_{\mathbf{x}}+d_{\mathbf{u}})^2\gamma_{\mathrm{sta}}^{-12}\log\frac{d}{\delta\gamma_{\mathrm{sta}}}\log^2\frac{1}{\delta}.$$

We observe that the leading $\sqrt{d_{\mathbf{u}}^2 d_{\mathbf{x}} T}$ terms in the upper and lower bounds differ only by factors polynomial in γ_{sta}, as well as a $\sigma_{\min}(A_\star)$ factor incurred by the lower bound. The lower order term $(d_{\mathbf{x}}+d_{\mathbf{u}})^2$ in the upper bound appears unavoidable, but we leave a complementary lower bound for future work. Both corollaries hold because strong stability immediately implies a bound on $\|P_\star\|_{\mathrm{op}}$.

Proof of Corollary 1 and Corollary 2. First, observe that under Assumption 1, $\Psi_{B_\star} \le 1$. Next, note that if $d_{\mathbf{u}} < d_{\mathbf{x}}/2$, then for $m = \lceil d_{\mathbf{x}}/2\rceil$, $\sigma_m(A_{\mathrm{cl},\star}) = \sigma_m(A_\star + B_\star K_\star) \ge \sigma_{m+d_{\mathbf{u}}}(A_\star + B_\star K_\star) \ge \sigma_{\min}(A_\star)$. This gives $\nu_m \ge \sigma_{\min}(A_\star)/(1+\|P_\star\|_{\mathrm{op}})$. Finally, Lemma B.7 (stated and proven in Appendix B.3.1) gives $\|P_\star\|_{\mathrm{op}} \le \gamma_{\mathrm{sta}}^{-1}$. Plugging these three observations into Theorem 1 and Theorem 2 concludes the proof. □

3. Perturbation Bounds via the Self-Bounding ODE Method

Both Theorem 1 and Theorem 2 scale only with the natural system parameter $\|P_\star\|_{\mathrm{op}}$, and avoid explicit dependence

[8] The assumption $d_{\mathbf{u}} \le \frac{1}{2}d_{\mathbf{x}}$ can be replaced with $d_{\mathbf{u}} \le \alpha d_{\mathbf{x}}$ for any $\alpha < 1$, and can be removed entirely for special instances. See Corollary 7 in Appendix G.7 for more details.

on the spectral radius or strong stability parameters found in prior work. This is achieved using the *self-bounding ODE* method, a new technique for deriving bounds on perturbations to the DARE solution $P_\infty(A, B)$ and corresponding controller $K_\infty(A, B)$ as the matrices A and B are varied. This method gives a general recipe for establishing perturbation bounds for solutions to implicit equations. It depends only on the norms of the system matrices and DARE solution $P_\infty(A, B)$, and it applies to all stabilizable systems, even those that are not controllable.

In this section we give an overview of the self-bounding ODE method and use it to prove a simplified version of the main perturbation bound used in our main upper and lower bounds. To state the perturbation bound, we first define the following problem-dependent constants.

$$C_{\mathrm{safe}}(A, B) = 54\|P_\infty(A, B)\|_{\mathrm{op}}^5, \quad \text{and}$$
$$C_{\mathrm{est}}(A, B) = 142\|P_\infty(A, B)\|_{\mathrm{op}}^8. \tag{3.1}$$

The parameter $C_{\mathrm{safe}}(A, B)$ determines the radius of admissible perturbations, while the parameter $C_{\mathrm{est}}(A, B)$ determines the quality of controllers synthesized from the resulting perturbation. The main perturbation bound is as follows.

Theorem 3. *Let $(A_\star, B_\star)$ be a stabilizable system. Given an alternate pair of matrices $(\widehat{A}, \widehat{B})$, for each $\circ \in \{\mathrm{op}, F\}$ define $\epsilon_\circ := \max\{\|\widehat{A} - A_\star\|_\circ, \|\widehat{B} - B_\star\|_\circ\}$. Then if $\epsilon_{\mathrm{op}} \le 1/C_{\mathrm{safe}}(A_\star, B_\star)$,*

1. $\|P_\infty(\widehat{A}, \widehat{B})\|_{\mathrm{op}} \lesssim \|P_\star\|_{\mathrm{op}}$ *and* $\|K_\star - K_\infty(\widehat{A}, \widehat{B})\|_{\mathrm{op}} \lesssim \frac{1}{\|P_\star\|_{\mathrm{op}}^{3/2}}$.
2. $\mathcal{J}_{A_\star, B_\star}[K_\infty(\widehat{A}, \widehat{B})] - \mathcal{J}^\star_{A_\star, B_\star} \le C_{\mathrm{est}}(A_\star, B_\star)\epsilon_{\mathrm{F}}^2$.

This theorem is a simplification of a stronger version, Theorem 5, stated and proven in Appendix B.1. Additional perturbation bounds are detailed in Appendix B.1; notably, Theorem 11 shows that the condition $\epsilon_{\mathrm{op}} \le 1/C_{\mathrm{safe}}(A_\star, B_\star)$ can be replaced by a condition that can be certificated from an approximate estimate of the system. In the remainder of this section, we sketch how to use the self-bounding ODE method to prove the following slightly more general version of the first part of Theorem 3.

Proposition 4. *Let $(A_\star, B_\star)$ be a stabilizable system and let $(\widehat{A}, \widehat{B})$ be an alternate pair of matrices. Then, if $u := 8\|P_\star\|_{\mathrm{op}}^2\epsilon_{\mathrm{op}} < 1$, the pair $(\widehat{A}, \widehat{B})$ is stabilizable and the following bounds hold:*

1. $\|P_\infty(\widehat{A}, \widehat{B})\|_{\mathrm{op}} \le (1-u)^{-1/2}\|P_\star\|_{\mathrm{op}}$.
2. *For each $\circ \in \{\mathrm{op}, F\}$, $\|K_\infty(\widehat{A}, \widehat{B}) - K_\star\|_\circ \le 7(1-u)^{-7/4}\|P_\star\|_{\mathrm{op}}^{7/2}\epsilon_\circ$.*

To begin proving the proposition, set $\Delta_A := \widehat{A} - A_\star$ and $\Delta_B := \widehat{B} - B_\star$. We consider a linear curve between the two instances, parameterized by $t \in [0, 1]$:

$$(A(t), B(t)) = (A_\star + t\Delta_A, B_\star + t\Delta_B)\,. \tag{3.2}$$

At each point t for which $(A(t), B(t))$ is stabilizable, the DARE has a unique solution, which allows us to define associated optimal cost matrices, controllers, and closed-loop dynamics matrices:

$$P(t) := P_\infty(A(t), B(t)), \quad K(t) := K_\infty(A(t), B(t))$$
$$\text{and} \quad A_{\mathrm{cl}}(t) := A(t) + B(t)K(t). \tag{3.3}$$

Our strategy will be to show that $P(t)$ and $K(t)$ are in fact smooth curves, and then obtain uniform bounds on $\|P'(t)\|_\circ$ and $\|K'(t)\|_\circ$ over the interval $[0, 1]$, yielding perturbation bounds via the mean value theorem. To start, we express the derivatives of the DARE in terms of Lyapunov equations.

Definition 3.1 (Discrete Lyapunov Equation)**.** Let $X, Y \in \mathbb{R}^{d_{\mathtt{x}} \times d_{\mathtt{x}}}$ with $Y = Y^\top$ and $\rho(X) < 1$. We let $\mathcal{T}_X[P] := X^\top P X - X$, and let $\mathsf{dlyap}(X, Y)$ denote the unique PSD solution $\mathcal{T}_X[P] = Y$. We let $\mathsf{dlyap}[X] := \mathsf{dlyap}(X, I)$.

The following lemma (proven in Appendix C.2) serves as the basis for our computations, and also establishes the requisite smoothness required to take derivatives.

Lemma 3.1 (Derivative and Smoothness of the DARE)**.** *Let $(A(t), B(t))$ be an analytic curve, and define $\Delta_{A_{\mathrm{cl}}}(t) := A'(t) + B'(t)K_\infty(A(t), B(t))$. Then for any t such that $(A(t), B(t))$ is stabilizable, the functions $P(u)$ and $K(u)$ are analytic in a neighborhood around t, and we have $P'(u) = \mathsf{dlyap}(A_{\mathrm{cl}}(u), Q_1(u))$, where $Q_1(u) := A_{\mathrm{cl}}(u)^\top P(u)\Delta_{A_{\mathrm{cl}}}(u) + \Delta_{A_{\mathrm{cl}}}(u)^\top P(u)A_{\mathrm{cl}}(u)$.*

Lemma 3.1 expresses $P'(t)$ as the solution to an ordinary differential equation. While the lemma guarantees local existence of the derivatives, it is not clear that the entire curve $(A(t), B(t)), t \in [0, 1]$ is stabilizable. However, since ODEs are locally guaranteed to have solutions, we should only expect trouble when the corresponding ODE becomes ill-defined, i.e. if $P'(t)$ escapes to infinity. We circumvent this issue by observing that $P'(t)$ satisfies the following self-bounding property.

Lemma 3.2 (Bound on First Derivatives)**.** *Let $(A(t), B(t))$ be an analytic curve. Then, for all t at which $(A(t), B(t))$ is stabilizable, we have $\|P'(t)\|_\circ \le 4\|P(t)\|_{\mathrm{op}}^3\epsilon_\circ$, and $\|K'(t)\|_\circ \le 7\|P(t)\|_{\mathrm{op}}^{7/2}\epsilon_\circ$.*

The bound on $P'(t)$ above follows readily from the expression for $P'(t)$ derived in Lemma 3.1, and the bound on $K'(t)$ uses that K is an explicit, analytic function of P; see Appendix C.2 for a full proof. Intuitively, the self-bounding property states that if P does not escape to infinity, then

$P'(t)$ cannot escape either. Since the rate of growth for $P(t)$ is in turn bounded by $P'(t)$, this suggests that there is an interval for t on which P and P' self-regulate one another, ensuring a well-behaved solution.

3.1. Norm Bounds for Self-Bounding ODEs

Informally, the self-bounding ODE method argues that if a vector-valued ODE $y(t)$ satisfies a self-bounding property of the form $\|y'(t)\| \le g(\|y(t)\|)$ wherever it is defined, then the ODE can be compared to a scalar ODE $z'(t) \approx g(z(t))$ with initial condition $z(0) \approx \|y(0)\|$. Specifically, it admits a solution $y(t)$ which is well-defined on an interval roughly as large as that of $z(t)$. We develop the method in a general setting where $y(t)$ (when defined) is the zero of a sufficiently regular function.

Definition 3.2 (Valid Implicit Function)**.** A function $F(\cdot,\cdot) : \mathbb{R}^m \times \mathbb{R}^d \to \mathbb{R}^d$ is a called a *valid implicit function* with domain $\mathcal{U} \subseteq \mathbb{R}^d$ if F is continuously differentiable, and if for any continuously differentiable curve $x(t)$ and any $t \in [0,1]$, either (a) $F(x(t), y) = 0$ has no solution $y \in \mathcal{U}$, or (b) it has a unique solution $y(t) \in \mathcal{U}$, and there exists an open interval around t and a $\mathcal{C}^1$ curve $y(u)$ defined on this interval for which $F(x(u), y(u)) = 0$.

This setting captures as a special case the characterization of $P(t)$ from Lemma 3.1. As a consequence of the lemma, we may take $F = \mathcal{F}_{\mathsf{DARE}}$, where, identifying $\mathbb{S}^{d_{\mathbf{x}}}$ as a $\binom{d_{\mathbf{x}}+1}{2}$-dimensional euclidean space, $\mathcal{F}_{\mathsf{DARE}} : (\mathbb{R}^{d_{\mathbf{x}}^2} \times \mathbb{R}^{d_{\mathbf{x}} d_{\mathbf{u}}}) \times \mathbb{S}^{d_{\mathbf{x}}} \to \mathbb{S}^{d_{\mathbf{x}}}$ is the function whose zero-solution defines the DARE: $\mathcal{F}_{\mathsf{DARE}}((A,B),P) := A^\top PA - P - A^\top PB(R_{\mathbf{u}} + B^\top PB)^{-1}B^\top PA + R_{\mathbf{x}}$. Then $\mathcal{F}_{\mathsf{DARE}}$ is a valid implicit function with unique solutions in the set of positive-definite matrices $\mathcal{U} := \mathbb{S}^{d_{\mathbf{x}}}_{++}$. To proceed, we introduce our self-bounding condition.

Definition 3.3 (Self-bounding)**.** Let $g : \mathbb{R} \to \mathbb{R}_{\ge 0}$ be non-negative and non-decreasing, let F be a valid implicit function with domain $\mathcal{U}$, and let $\|\cdot\|$ be a norm. For a continuously differentiable curve $x(t)$ defined on $[0,1]$, we say that F is $(g, \|\cdot\|)$-*self bounded* on $x(t)$ if $F(x(0), y) = 0$ has a solution $y \in \mathcal{U}$ and $\|y'(t)\| \le g(\|y\|)$ for all $t \in [0,1]$ for which $F(x(t), y)\ y \in \mathcal{U}$. We call the tuple $(F, \mathcal{U}, g, \|\cdot\|, x(\cdot))$ a *self-bounding* tuple.

Lemma 3.2 shows that $\mathcal{F}_{\mathsf{DARE}}$ is $(g, \|\cdot\|_{\mathrm{op}})$-self bounding on the curve the $(A(t), B(t))$ with $g(z) = cz^3$ for $c \propto \epsilon_{\mathrm{op}}$. For functions $g(z)$ with this form we have the following general bound on $\|y(t)\|$.

Corollary 3. *Let $(F, \mathcal{U}, g, \|\cdot\|, x(\cdot))$ be a self-bounding tuple, where $g(z) = cz^p$ for $c > 0$ and $p > 1$. Then, if $\alpha := c(p-1)\|y(0)\|^{p-1} < 1$, there exists a unique continuously differentiable function $y(t) \in \mathcal{U}$ defined on $[0,1]$ which satisfies $F(x(t), y(t)) = 0$, and this solution satisfies $\forall t \in [0,1]$, $\|y(t)\| \le (1-\alpha)^{-1/(p-1)}\|y(0)\|$, and $\|y'(t)\| \le c(1-\alpha)^{-p/(p-1)}\|y(0)\|^p$.*

Corollary 3 is a consequence of a similar result for general functions g (Theorem 13, in Appendix D). The condition on the parameter α directly arises from the requirement that the scalar ODE $w'(u) = cw(u)^3$ has a solution on $[0,1]$.

Finishing the Proof of Proposition 4 Finally, we use Corollary 3 to conclude the proof of Proposition 4.

Proof of Proposition 4. Lemma 3.2 states that for any $t \in [0,1]$ for which $(A(t), B(t))$ is stabilizable (i.e., $\mathcal{F}_{\mathsf{DARE}}([A(t), B(t)], \cdot)$ has a solution), we have the bound

$$\|P'(t)\|_{\mathrm{op}} \le 4\|P(t)\|_{\mathrm{op}}^3 \epsilon_{\mathrm{op}}.$$

Applying Corollary 3 with $p = 2$ and $c = 4\epsilon_{\mathrm{op}}$, we see that if $\alpha := 8\epsilon_{\mathrm{op}}\|P_\star\|_{\mathrm{op}}^2 < 1$, then $P(t)$ is continuously differentiable on the interval $[0,1]$ and $\forall t \in [0,1]$, $\|P(t)\|_{\mathrm{op}} \le \|P_\star\|_{\mathrm{op}}/\sqrt{1-\alpha}$. By Lemma 3.2, $K(t)$ is well defined as well, and satisfies $\max_{t\in[0,1]} \|K'(t)\|_\circ \le 7\epsilon_\circ \max_{t\in[0,1]} \|P(t)\|_{\mathrm{op}}^{7/2} \le (1-\alpha)^{-7/4}\|P_\star\|_{\mathrm{op}}$. The desired bound on $\|K_\infty(A_\star, B_\star) - K_\infty(\widehat{A}, \widehat{B})\|_\circ$ follows from the mean value theorem. □

4. Concluding Remarks

We have established that the asymptotically optimal regret for the online LQR problem is $\widetilde{\Theta}(\sqrt{d_{\mathbf{u}}^2 d_{\mathbf{x}} T})$, and that this rate is attained by ε-greedy exploration. We are hopeful that the our new analysis techniques, especially our perturbation bounds, will find broader use within the non-asymptotic theory of control and beyond. Going forward our work raises a number of interesting conceptual questions. Are there broader classes of "easy" reinforcement learning problems beyond LQR for which naive exploration attains optimal sample complexity, or is LQR a fluke? Conversely, is there a more demanding (eg, robust) version of the LQR problem for which more sophisticated exploration techniques such as robust synthesis (Dean et al., 2018) or optimism in the face of uncertainty (Abbasi-Yadkori & Szepesvári, 2011; Cohen et al., 2019) are required to attain optimal regret? On the purely technical side, recall that while our upper and lower bound match in terms of dependence on $d_{\mathbf{u}}$, $d_{\mathbf{x}}$, and T, they differ in their polynomial dependence on $\|P_\star\|_{\mathrm{op}}$. Does closing this gap require new algorithmic techniques, or will a better analysis suffice?

Acknowledgements Max Simchowitz is generously supported by an Open Philanthropy graduate student fellowship. Dylan Foster acknowledges the support of NSF TRIPODS award #1740751.

References

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

Abbasi-Yadkori, Y., Pál, D., and Szepesvári, C. Improved algorithms for linear stochastic bandits. In *Advances in Neural Information Processing Systems*, pp. 2312–2320, 2011.

Abeille, M. and Lazaric, A. Thompson sampling for linear-quadratic control problems. In *Artificial Intelligence and Statistics*, pp. 1246–1254, 2017.

Abeille, M. and Lazaric, A. Improved regret bounds for thompson sampling in linear quadratic control problems. In *International Conference on Machine Learning*, pp. 1–9, 2018.

Adamczak, R. A note on the Hanson-Wright inequality for random vectors with dependencies. *Electronic Communications in Probability*, 20, 2015.

Agarwal, N., Bullins, B., Hazan, E., Kakade, S., and Singh, K. Online control with adversarial disturbances. In *International Conference on Machine Learning*, pp. 111–119, 2019a.

Agarwal, N., Hazan, E., and Singh, K. Logarithmic regret for online control. In *Advances in Neural Information Processing Systems 32*, pp. 10175–10184. 2019b.

Arias-Castro, E., Candes, E. J., and Davenport, M. A. On the fundamental limits of adaptive sensing. *IEEE Transactions on Information Theory*, 59(1):472–481, 2012.

Assouad, P. Deux remarques sur l'estimation. *Comptes rendus des séances de l'Académie des sciences. Série 1, Mathématique*, 296(23):1021–1024, 1983.

Azar, M. G., Osband, I., and Munos, R. Minimax regret bounds for reinforcement learning. In *Proceedings of the 34th International Conference on Machine Learning-Volume 70*, pp. 263–272. JMLR. org, 2017.

Bertsekas, D. P. *Dynamic Programming and Optimal Control, Vol. I*. Athena Scientific, 2005.

Bof, N., Carli, R., and Schenato, L. Lyapunov theory for discrete time systems. *arXiv preprint arXiv:1809.05289*, 2018.

Boyd, S. Lecture 13: Linear quadratic Lyapunov theory. *EE363 Course Notes, Stanford University*, 2008.

Cohen, A., Hasidim, A., Koren, T., Lazic, N., Mansour, Y., and Talwar, K. Online linear quadratic control. In *International Conference on Machine Learning*, pp. 1028–1037, 2018.

Cohen, A., Koren, T., and Mansour, Y. Learning linear-quadratic regulators efficiently with only $\sqrt{T}$ regret. In *International Conference on Machine Learning*, pp. 1300–1309, 2019.

Dann, C. and Brunskill, E. Sample complexity of episodic fixed-horizon reinforcement learning. In *Advances in Neural Information Processing Systems*, pp. 2818–2826, 2015.

Dean, S., Mania, H., Matni, N., Recht, B., and Tu, S. On the sample complexity of the linear quadratic regulator. *Foundations of Computational Mathematics*, pp. 1–47.

Dean, S., Mania, H., Matni, N., Recht, B., and Tu, S. Regret bounds for robust adaptive control of the linear quadratic regulator. In *Advances in Neural Information Processing Systems*, pp. 4188–4197, 2018.

Faradonbeh, M. K. S., Tewari, A., and Michailidis, G. Input perturbations for adaptive regulation and learning. *arXiv preprint arXiv:1811.04258*, 2018a.

Faradonbeh, M. K. S., Tewari, A., and Michailidis, G. On optimality of adaptive linear-quadratic regulators. *arXiv preprint arXiv:1806.10749*, 2018b.

Fazel, M., Ge, R., Kakade, S., and Mesbahi, M. Global convergence of policy gradient methods for the linear quadratic regulator. In *International Conference on Machine Learning*, pp. 1466–1475, 2018.

Hazan, E., Agarwal, A., and Kale, S. Logarithmic regret algorithms for online convex optimization. *Machine Learning*, 69(2):169–192, 2007.

Hazan, E., Singh, K., and Zhang, C. Learning linear dynamical systems via spectral filtering. In *Advances in Neural Information Processing Systems*, pp. 6702–6712, 2017.

Hazan, E., Lee, H., Singh, K., Zhang, C., and Zhang, Y. Spectral filtering for general linear dynamical systems. In *Advances in Neural Information Processing Systems*, pp. 4634–4643, 2018.

Hsu, D., Kakade, S., Zhang, T., et al. A tail inequality for quadratic forms of subgaussian random vectors. *Electronic Communications in Probability*, 17, 2012.

Jaksch, T., Ortner, R., and Auer, P. Near-optimal regret bounds for reinforcement learning. *Journal of Machine Learning Research*, 11(Apr):1563–1600, 2010.

Jiang, N., Krishnamurthy, A., Agarwal, A., Langford, J., and Schapire, R. E. Contextual decision processes with low Bellman rank are PAC-learnable. In *Proceedings of the 34th International Conference on Machine Learning-Volume 70*, pp. 1704–1713. JMLR. org, 2017.

Jin, C., Yang, Z., Wang, Z., and Jordan, M. I. Provably efficient reinforcement learning with linear function approximation. *Conference on Learning Theory (COLT)*, 2020.

Kakade, S., Kearns, M. J., and Langford, J. Exploration in metric state spaces. In *Proceedings of the 20th International Conference on Machine Learning (ICML-03)*, pp. 306–312, 2003.

Kearns, M. J., Mansour, Y., and Ng, A. Y. Approximate planning in large POMDPs via reusable trajectories. In *Advances in Neural Information Processing Systems*, pp. 1001–1007, 2000.

Langford, J. and Zhang, T. The epoch-greedy algorithm for contextual multi-armed bandits. In *Proceedings of the 20th International Conference on Neural Information Processing Systems*, pp. 817–824. Citeseer, 2007.

Lillicrap, T. P., Hunt, J. J., Pritzel, A., Heess, N., Erez, T., Tassa, Y., Silver, D., and Wierstra, D. Continuous control with deep reinforcement learning. *arXiv preprint arXiv:1509.02971*, 2015.

Lincoln, B. and Rantzer, A. Relaxing dynamic programming. *IEEE Transactions on Automatic Control*, 51(8): 1249–1260, 2006.

Mania, H., Tu, S., and Recht, B. Certainty equivalence is efficient for linear quadratic control. In *Advances in Neural Information Processing Systems*, pp. 10154–10164, 2019.

Mnih, V., Kavukcuoglu, K., Silver, D., Rusu, A. A., Veness, J., Bellemare, M. G., Graves, A., Riedmiller, M., Fidjeland, A. K., Ostrovski, G., et al. Human-level control through deep reinforcement learning. *Nature*, 518(7540): 529, 2015.

Munos, R. and Szepesvári, C. Finite-time bounds for fitted value iteration. *Journal of Machine Learning Research*, 9 (May):815–857, 2008.

Ouyang, Y., Gagrani, M., and Jain, R. Control of unknown linear systems with Thompson sampling. In *2017 55th Annual Allerton Conference on Communication, Control, and Computing (Allerton)*, pp. 1198–1205. IEEE, 2017.

Polderman, J. W. On the necessity of identifying the true parameter in adaptive LQ control. *Systems & control letters*, 8(2):87–91, 1986.

Polderman, J. W. Adaptive LQ control: Conflict between identification and control. *Linear algebra and its applications*, 122:219–244, 1989.

Rakhlin, A. and Sridharan, K. Online nonparametric regression. In *Conference on Learning Theory*, 2014.

Ran, A. and Vreugdenhil, R. Existence and comparison theorems for algebraic riccati equations for continuous-and discrete-time systems. *Linear Algebra and its applications*, 99:63–83, 1988.

Rudelson, M. and Vershynin, R. Hanson-Wright inequality and sub-gaussian concentration. *Electronic Communications in Probability*, 18, 2013.

Sarkar, T. and Rakhlin, A. Near optimal finite time identification of arbitrary linear dynamical systems. In *International Conference on Machine Learning*, pp. 5610–5618, 2019.

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

Shamir, O. On the complexity of bandit and derivative-free stochastic convex optimization. In *Conference on Learning Theory*, pp. 3–24, 2013.

Silver, D., Huang, A., Maddison, C. J., Guez, A., Sifre, L., Van Den Driessche, G., Schrittwieser, J., Antonoglou, I., Panneershelvam, V., Lanctot, M., et al. Mastering the game of Go with deep neural networks and tree search. *nature*, 529(7587):484, 2016.

Simchowitz, M., Mania, H., Tu, S., Jordan, M. I., and Recht, B. Learning without mixing: Towards a sharp analysis of linear system identification. In *Conference On Learning Theory*, pp. 439–473, 2018.

Simchowitz, M., Boczar, R., and Recht, B. Learning linear dynamical systems with semi-parametric least squares. In *Conference on Learning Theory*, pp. 2714–2802, 2019.

Sutton, R. S. and Barto, A. G. *Reinforcement learning: An introduction*. 2018.

Tilli, P. Singular values and eigenvalues of non-Hermitian block Toeplitz matrices. *Linear Algebra and its Applications*, 272(1-3):59–89, 1998.

Tu, S. and Recht, B. Least-squares temporal difference learning for the linear quadratic regulator. In *International Conference on Machine Learning*, pp. 5005–5014, 2018.

Vovk, V. Competitive on-line statistics. *International Statistical Review*, 69(2):213–248, 2001.

Yu, B. Assouad, Fano, and Le Cam. In *Festschrift for Lucien Le Cam*, pp. 423–435. Springer, 1997.

A. Organization and Notation

A.1. Notation

Notation	Definition
T	problem horizon
$d_{\mathbf{x}}, d_{\mathbf{u}}$	state/input dimension
$\mathbf{x}_t, \mathbf{u}_t$	state/input at time t
$\mathbf{w}_t$	noise at time t
$R_{\mathbf{x}}, R_{\mathbf{u}}$	control costs
$\mathrm{Regret}_{A,B,T}[\pi]$	Regret of a policy (as a random variable)
$\mathbb{E}\mathrm{Regret}_{A,B,T}[\pi]$	Expected Regret of a policy
$\mathcal{R}_{A_\star,B_\star,T}(\epsilon)$	$\min_\pi \max_{A,B} \left\{\mathbb{E}\mathrm{Regret}_{A,B,T}[\pi] : \|A-A_\star\|_{\mathrm{F}}^2 \vee \|B-B_\star\|_{\mathrm{F}}^2 \le \epsilon\right\}.$
$P_\infty(A,B)$	Solution to the DARE
$K_\infty(A,B)$	Optimal Controller for DARE
$\mathcal{J}_{A,B}[K]$	Infinite horizon control cost of K on instance (A,B)
$\|A\|_{\mathcal{H}_\infty}$	$\max_{z\in\mathbb{C}:\|z\|=1} \|(zI-A)^{-1}\|_{\mathrm{op}}$
$\mathcal{B}_{\mathrm{op}}(\epsilon; A_0, B_0)$	$\{(A,B) \mid \|A-A_0\|_{\mathrm{op}} \vee \|B-B_0\|_{\mathrm{op}} \le \epsilon\}$
$\mathsf{dlyap}(X,Y)$	Solves $\mathcal{T}_X[P]=Y$, where $\mathcal{T}_X[P] := X^\top P X - X$. Requires $\rho(X)<1$, $Y=Y^\top$. Given by $\sum_{i\ge 0}(X^i)^\top Y X^i$.
System parameters	
$(A_\star, B_\star)$	**Upper bound:** Ground truth for upper bound. **Lower bound:** Nominal instance for local minimax complexity.
$P_\star$	$P_\infty(A_\star,B_\star)$
$K_\star$	$K_\infty(A_\star,B_\star)$
$A_{\mathrm{cl},\star}$	$A_\star + B_\star K_\star$
$\mathcal{J}^\star$	$\mathcal{J}^\star_{A_\star,B_\star} := \min_K \mathcal{J}_{A_\star,B_\star}[K] = \mathcal{J}_{A_\star,B_\star}[K_\star]$
$\Psi_\star$	$\max\{1, \|A_\star\|_{\mathrm{op}}, \|B_\star\|_{\mathrm{op}}\}$
$\Psi_{B_\star}$	$\max\{1, \|B_\star\|_{\mathrm{op}}\}$

A.2. Organization of the Appendices

The appendix is divided into three parts. Part I establishes the main technical tools used throughout the upper and lower bounds. Appendix B describes and proves our main perturbation bounds, deferring additional proof details to Appendix C. Appendix D proves guarantees for the Self-Bounding ODE method, summarized in Corollary 3, as well as a slightly more general statement for generic self-bounding relations, Theorem 13. This part of the appendix concludes with Appendix E.1, which describes a set of tools for analyzing ordinary least squares estimation, which we use in the proofs of both our upper and lower bounds.

Part II provides the proof of our lower bound, Theorem 1. Appendix F presents a complete proof in terms of numerous constituent lemmas, and Appendix G proves these supporting lemmas. Part III mirror the structure of Part II, with Appendix H presenting formal pseudocode for our algortithm and a proof of the upper bound, and Appendix I verifying the relevant constituent lemmas from Appendix H.

Part I

Technical Tools

B. Main Perturbation Bounds

Preliminaries Throughout, we shall use extensively the dlyap operator, which we recall here.

Definition 3.1 (Discrete Lyapunov Equation)**.** Let $X, Y \in \mathbb{R}^{d_{\mathbf{x}} \times d_{\mathbf{x}}}$ with $Y = Y^\top$ and $\rho(X) < 1$. We let $\mathcal{T}_X[P] := X^\top P X - X$, and let $\mathsf{dlyap}(X, Y)$ denote the unique PSD solution $\mathcal{T}_X[P] = Y$. We let $\mathsf{dlyap}[X] := \mathsf{dlyap}(X, I)$.

We shall need to describe the "P"-matrix analogue of the functional $\mathcal{J}$.

Definition B.1. Suppose that $(A_\star + B_\star K)$ is stable. We define $P_\infty(K; A_\star, B_\star) := \mathsf{dlyap}(A_\star + B_\star K, R_{\mathbf{x}} + K^\top R_{\mathbf{u}} K$.

It is a standard fact (see e.g. Lemma B.6) that $\mathcal{J}_{A_\star, B_\star}[K] = \mathrm{tr}(P_\infty(K; A_\star, B_\star))$ whenever $A_\star + B_\star K$ is stable. We also recall the definition of the $\mathcal{H}_\infty$-norm.

Definition B.2 ($\mathcal{H}_\infty$ norm)**.** For any stable $\tilde{A} \in \mathbb{R}^{d_{\mathbf{x}}^2}$ (e.g. $A + BK_\infty(A, B)$), we define $\|\tilde{A}\|_{\mathcal{H}_\infty} := \sup_{z \in \mathbb{C}: |z|=1} \|(zI - \tilde{A})^{-1}\|_{\mathrm{op}}$.

Organization of Appendix B The remainder of this appendix is organized as follows. Appendix B.1 states our main perturbation upper bounds, and provides proofs in terms of various supporting propositions. Appendix B.2 walks the reader through the relevant computations of various derivatives. Appendix B.3 states numerous technical tools which we use in the proofs of our main perturbation bounds, and finally Appendix B.4 proves the supporting propositions leveraged in Appendix B.1. Many supporting proofs are deferred to Appendix C.

B.1. Main Results

B.1.1. Main Perturbation Upper Bound

Recall $C_{\mathrm{safe}}(A_\star, B_\star) = 54\|P_\star\|_{\mathrm{op}}^5$, and $C_{\mathrm{est}}(A_\star, B_\star) = 142\|P_\star\|_{\mathrm{op}}^8$. We state a strengthening of our main perturbation bound from the main text (Theorem 3) here.

Theorem 5. *Let $(A_\star, B_\star)$ be a stabilizable system. Given an alternate pair of matrices $(\widehat{A}, \widehat{B})$, for each $\circ \in \{\mathrm{op}, F\}$ define $\epsilon_\circ := \max\{\|\widehat{A} - A_\star\|_\circ, \|\widehat{B} - B_\star\|_\circ\}$. Then if $\epsilon_{\mathrm{op}} \le 1/C_{\mathrm{safe}}(A_\star, B_\star)$,*

1. $\|P_\infty(\widehat{A}, \widehat{B})\|_{\mathrm{op}} \le 1.0835\|P_\star\|_{\mathrm{op}}$ *and* $\|B_\star(K_\star - K_\infty(\widehat{A}, \widehat{B}))\|_2 < \frac{1}{5\|P_\star\|_{\mathrm{op}}^{3/2}}$.
2. $\mathcal{J}_{A_\star, B_\star}[K_\infty(\widehat{A}, \widehat{B})] - \mathcal{J}^\star_{A_\star, B_\star} \le C_{\mathrm{est}}(A_\star, B_\star)\epsilon_{\mathrm{F}}^2$.
3. $\|P_\infty(K_\infty(\widehat{A}, \widehat{B}); A_\star, B_\star) - P_\star\|_{\mathrm{op}} \le C_{\mathrm{est}}(A_\star, B_\star)\epsilon_{\mathrm{op}}^2$.
4. *Moreover,* $P_\infty(K_\infty(\widehat{A}, \widehat{B}); A_\star, B_\star) \preceq (21/20)P_\star$.

Proof. Throughout, we use $P_\star \succeq I$ (see Lemma F.2). This theorem requires two consituent results. First, we have a perturbation bound for P_∞ and K_∞, which refines Proposition 4, and is proven in Section B.4.1.

Proposition 6. *Let $(A_\star, B_\star)$ be a stabilizable system, and define the* DARE *solution $P_\star := P_\infty(A_\star, B_\star)$ and controller $K_\star = K_\infty(A_\star, B_\star)$. Given an alternate pair of matrices $(\widehat{A}, \widehat{B})$, define for norms $\circ \in \{\mathrm{op}, \mathrm{F}\}$ the error $\epsilon_\circ := \max\{\|A_\star - \widehat{A}\|_\circ, \|B_\star - \widehat{B}\|_\circ\}$. Then, if $\alpha := 8\|P_\star\|_{\mathrm{op}}^2\epsilon_{\mathrm{op}} < 1$, the pair $(\widehat{A}, \widehat{B})$ is stabilizable, and*

$$\begin{aligned}
&\|P_\infty(\widehat{A}, \widehat{B})\|_{\mathrm{op}} \le (1-\alpha)^{-1/2}\|P_\star\|_{\mathrm{op}}, \\
&\|R_{\mathbf{u}}^{1/2}(K_\infty(\widehat{A}, \widehat{B}) - K_\star)\|_\circ \le 7(1-\alpha)^{-7/4}\|P_\star\|_{\mathrm{op}}^{7/2}\,\epsilon_\circ, \\
&\|B_\star(K_\infty(\widehat{A}, \widehat{B}) - K_\star)\|_\circ \le 8(1-\alpha)^{-7/4}\|P_\star\|_{\mathrm{op}}^{7/2}\,\epsilon_\circ.
\end{aligned}$$

In addition, if $\epsilon_{\mathrm{op}} \le 32\|P_\star\|_{\mathrm{op}}^3$, *then*

$$\|P_\star^{1/2} B(K_\infty(\widehat{A},\widehat{B}) - K_\star)\|_\circ \le 9(1-\alpha)^{-7/4}\|P_\star\|_{\mathrm{op}}^{7/2}\epsilon_\circ.$$

Next, we have a perturbation bound for the $\mathcal{J}$-functional as the controller K-is varied. The proof is deferred to Section B.4.2.

Proposition 7. *Fix any controller* K *satisfying* $\|B_\star(K-K_\star)\|_2 \le 1/5\|P_\star\|_{\mathrm{op}}^{3/2}$. *Then,*

$$\mathcal{J}_{A_\star,B_\star}[K] - \mathcal{J}_{A_\star,B_\star} \le \|P_\star\|_{\mathrm{op}} \max\{\|K-K_\star\|_{\mathrm{F}}^2, \|P_\star^{1/2}B_\star(K-K_\star)\|_{\mathrm{F}}^2\},$$
$$\|P_\infty(K;A_\star,B_\star) - P_\infty(A_\star,B_\star)\|_{\mathrm{op}} \le \|P_\star\|_{\mathrm{op}} \max\{\|K-K_\star\|_{\mathrm{op}}^2, \|P_\star^{1/2}B_\star(K-K_\star)\|_{\mathrm{op}}^2\}.$$

Now, observe that $\epsilon_{\mathrm{op}} \le 1/54\|P_\star\|_{\mathrm{op}}^5 < 1/8\|P_\star\|_{\mathrm{op}}^2$ and $\alpha = 8\|P_\star\|_{\mathrm{op}}^2\epsilon_{\mathrm{op}}$, Proposition 6 gives that

$$\|P_\infty(\widehat{A},\widehat{B})\|_{\mathrm{op}} \le \|P_\star\|_{\mathrm{op}}/\sqrt{1-8/54} \le 1.0835\|P_\star\|_{\mathrm{op}},$$

and that

$$\begin{aligned}
5\|P_\star\|_{\mathrm{op}}^{3/2}\cdot\|B_\star(K_\infty(\widehat{A},\widehat{B})-K_\star)\|_{\mathrm{op}} &\le 8(1-\alpha)^{-7/4}\|P_\star\|_{\mathrm{op}}^{7/2}\,\epsilon_{\mathrm{op}} \\
&\le 40(1-\alpha)^{-7/4}\|P_\star\|_{\mathrm{op}}^{5}\,\epsilon_{\mathrm{op}} \\
&\le 40/54\cdot(1-8/54)^{-7/4} < 1.
\end{aligned}$$

Hence, for such ϵ_{op}, we find from Proposition 7 followed by Proposition C.3.1 that

$$\begin{aligned}
\mathcal{J}_{A_\star,B_\star}[K] - \mathcal{J}_{A_\star,B_\star} &\le \|P_\star\|_{\mathrm{op}}\max\{\|R_{\mathbf{u}}^{1/2}(K-K_\star)\|_{\mathrm{F}}^2, \|P_\star^{1/2}B_\star(K-K_\star)\|_{\mathrm{F}}^2\} \\
&\le 81\|P_\star\|_{\mathrm{op}}^8(1-\alpha)^{-7/2}\epsilon_{\mathrm{F}}^2 \\
&\le 142\|P_\star\|_{\mathrm{op}}^8\epsilon_{\mathrm{F}}^2,
\end{aligned}$$

and similarly, using $\|P_\star\|_{\mathrm{op}} \ge 1$,

$$\|P_\infty(K;A_\star,B_\star) - P_\infty(A_\star,B_\star)\|_{\mathrm{op}} \le 142\|P_\star\|_{\mathrm{op}}^8\epsilon_{\mathrm{op}}^2 \le \frac{1}{20},$$

yielding $P_\infty(K;A_\star,B_\star) \preceq (1+\frac{1}{20})P_\star$ as $P_\star \succeq I$.

□

B.1.2. Perturbation of $\mathcal{H}_\infty$ Norm and Lyapunov Functions

Next, we establish perturbation bounds on the $\mathcal{H}_\infty$ norm of the closed loop system, and show that all perturbed closed loop systems share a common Lyapunov function.

Theorem 8. *Let* $A_\star, B_\star$ *be stabilizable, and let* $(\widehat{A},\widehat{B})$ *satisfy the conditions of Theorem 5, with* $R_{\mathbf{x}} \succeq I$, *and* $R_{\mathbf{u}} = I$. *Define* $A_{\mathrm{cl},\star} := A_\star + B_\star K_\star$, *and given* $(\widehat{A},\widehat{B}) \in \mathcal{B}_{\mathrm{op}}(\epsilon_\star, A_\star, B_\star)$, *define and* $A_{\mathrm{cl},\widehat{\star}} := A_\star + B_\star K_\infty(\widehat{A},\widehat{B})$. *Then,*

1. $I \preceq \mathsf{dlyap}[A_{\mathrm{cl},\star}] \preceq P_\star$.
2. $\|A_{\mathrm{cl},\widehat{\star}}\|_{\mathcal{H}_\infty} \le 2\|A_{\mathrm{cl},0}\|_{\mathcal{H}_\infty} \le 4\|\mathsf{dlyap}[A_{\mathrm{cl},\star}]\|_{\mathrm{op}}^{3/2} \le 4\|P_\star\|_{\mathrm{op}}^{3/2}$.
3. $A_{\mathrm{cl},\widehat{\star}}^\top\cdot\mathsf{dlyap}[A_{\mathrm{cl},\star}]\cdot A_{\mathrm{cl},\widehat{\star}} \preceq (1-\frac{1}{2}\|\mathsf{dlyap}[A_{\mathrm{cl},\star}\|_{\mathrm{op}}^{-1})\,\mathsf{dlyap}[A_{\mathrm{cl},\star}] \preceq (1-\frac{1}{2}\|P_\star\|_{\mathrm{op}}^{-1})\,\mathsf{dlyap}[A_{\mathrm{cl},\star}]$.

Proof of Part 1. We can directly verify $\mathsf{dlyap}[A_{\mathrm{cl},\star}] \succeq I$ from the definition, and $\mathsf{dlyap}[A_{\mathrm{cl},\star}] \preceq P_\star$ by Lemma B.5.

Proof of Part 2. We use a general-purpose perturbation bound for the $\mathcal{H}_\infty$ norm, proved in B.4.3.

Proposition 9 ($\mathcal{H}_\infty$ Bounds)**.** *Fix* $u \in (0,1)$, *and matrixes* $A_{\mathrm{safe}}, A_1 \in \mathbb{R}^{d_{\mathbf{x}}^2}$ *with* A_{safe} *stable. Then if* $\|A_1 - A_{\mathrm{safe}}\| \le \frac{\alpha}{\|A_{\mathrm{safe}}\|_{\mathcal{H}_\infty}}$, $\|A_1\|_{\mathcal{H}_\infty} \le \frac{1}{1-\alpha}\|A_{\mathrm{safe}}\|_{\mathcal{H}_\infty}$.

From Part 1 of Theorem 5,

$$\|A_{\mathrm{cl},\star} - A_{\mathrm{cl},\widehat{\star}}\|_{\mathrm{op}} \le \|B_\star(K_\infty(A_\star, B_\star) - K_\infty(\widehat{A}, \widehat{B}))\|_{\mathrm{op}} < \frac{1}{5\|P_\star\|_{\mathrm{op}}^{3/2}}. \tag{B.1}$$

By Lemma B.11 followed by Lemma B.5, we have that

$$\|A_{\mathrm{cl},\star}\|_{\mathcal{H}_\infty} \le 2\|\mathsf{dlyap}[A_{\mathrm{cl},\star}]\|_{\mathrm{op}}^{3/2} \le 2\|P_\star\|_{\mathrm{op}}^{3/2}.$$

Therefore, since $\|P_\star\|_{\mathrm{op}} \ge 1$, we have

$$\|A_{\mathrm{cl},\star} - A_{\mathrm{cl},\widehat{\star}}\|_{\mathrm{op}} < \frac{1}{(5/2)\|A_{\mathrm{cl},\star}\|_{\mathcal{H}_\infty}} \le \frac{1}{2\|A_{\mathrm{cl},\star}\|_{\mathcal{H}_\infty}}.$$

Propostion 9 then implies that $\|A_{\mathrm{cl},\widehat{\star}}\|_{\mathcal{H}_\infty} \le 2\|A_{\mathrm{cl},\star}\|_{\mathcal{H}_\infty}$. Moreover, by Lemma B.11, we can upper bound this in term by $4\|\mathsf{dlyap}[A_{\mathrm{cl},\star}]\|_{\mathrm{op}}^{3/2} \le 4\|P_\star\|_{\mathrm{op}}^{3/2}$.

Proof of Part 3. Here, we use a perturbation bound which we prove from first principles, without the self-bounding ODE method (proved in Appendix B.4.4).

Proposition 10. *Suppose that A is a stable matrix, and suppose that $\widehat{A}$ satisfies*

$$\|\widehat{A} - A\|_{\mathrm{op}} \le \frac{1}{4}\min\left\{\frac{1}{\|\mathsf{dlyap}[A]\|_{\mathrm{op}}\|A\|_{\mathrm{op}}}, \|\mathsf{dlyap}[A]\|_{\mathrm{op}}^{-1/2}\right\},$$

Then, $\widehat{A}^\top \mathsf{dlyap}[A]\widehat{A} \preceq (1 - \frac{1}{2}\|\mathsf{dlyap}[A]\|_{\mathrm{op}}^{-1}) \cdot \mathsf{dlyap}[A]$.

By Lemma B.8, we have $\|A_{\mathrm{cl},\star}\|_{\mathrm{op}} \le \|P_\star\|_{\mathrm{op}}^{1/2}$. Since $\|\mathsf{dlyap}[A_{\mathrm{cl},\star}]\|_{\mathrm{op}} \le \|P_\star\|_{\mathrm{op}}$, combining with Eq. B.1 gives

$$\|A_{\mathrm{cl},\star} - A_{\mathrm{cl},\widehat{\star}}\|_{\mathrm{op}} < \frac{1}{5\|P_\star\|_{\mathrm{op}}^{3/2}} < \frac{1}{4\|\mathsf{dlyap}[A_{\mathrm{cl},\star}]\|_{\mathrm{op}}\|A_{\mathrm{cl},\star}\|_{\mathrm{op}}}.$$

Similarly, we have $\|A_{\mathrm{cl},\star} - A_{\mathrm{cl},\widehat{\star}}\|_{\mathrm{op}} < \frac{1}{\|\mathsf{dlyap}[A_{\mathrm{cl},\star}]\|_{\mathrm{op}}^{1/2}}$, which means that, in particular, $A_{\mathrm{cl},\star}, A_{\mathrm{cl},\widehat{\star}}$ satisfy the conditions for $A, \widehat{A}$ in Proposition 10. This means that $A_{\mathrm{cl},\widehat{\star}}^\top \mathsf{dlyap}[A_{\mathrm{cl},\star}]A_{\mathrm{cl},\widehat{\star}} \preceq (1 - \frac{1}{2}\|\mathsf{dlyap}[A_{\mathrm{cl},\star}]\|_{\mathrm{op}}^{-1}) \cdot \mathsf{dlyap}[A_{\mathrm{cl},\star}](1 - \frac{1}{2}\|P_\star\|_{\mathrm{op}}^{-1})$. The last inequality follows from Part 1. □

B.1.3. Continuity of the Safe Set

We show that the size of the so-called "safe" set is continuous in nearby instances. This allows us to use an instance (A_0, B_0) to guage whether the perturbed system $(\widehat{A}, \widehat{B})$ is sufficiently close to $(A_\star, B_\star)$ to ensure correctness of the perturbation bounds.

Theorem 11. *Let (A_0, B_0) be a stabilizable system. Then, for any pair of systems $(\widehat{A}, \widehat{B}), (A_\star, B_\star) \in \mathcal{B}_{\mathrm{op}}(\frac{1}{3C_{\mathrm{safe}}(A_0,B_0)}, A_0, B_0)$ is stabilizable, and satisfies $\max\{\|A_\star - \widehat{A}\|_{\mathrm{op}}, \|\widehat{B} - B_\star\|_{\mathrm{op}}\} \le 1/C_{\mathrm{safe}}(A_\star, B_\star)$. Moreover, $\|P_\infty(A_\star, B_\star)\|_{\mathrm{op}} \le 1.0835\|P_\infty(A_0, B_0)\|_{\mathrm{op}}$.*

Proof. Let $\epsilon_0 := \max\{\|A_0 - \widehat{A}\|_{\mathrm{op}}, \|B_0 - B_\star\|_{\mathrm{op}}\} \le 1/C_{\mathrm{safe}}(A_0, B_0)$. Applying Theorem 5 Part 1 with $(\widehat{A}, \widehat{B}) \leftarrow (A_\star, B_\star)$ and $(A_\star, B_\star) \leftarrow (A_0, B_0)$, we have $\|P_\infty(A_\star, B_\star)\|_{\mathrm{op}} \le 1.0835\|P_\infty(A_0, B_0)\|_{\mathrm{op}}$. Hence, $C_{\mathrm{safe}}(A_\star, B_\star) \le 1.5C_{\mathrm{safe}}(A_0, B_0)$. Hence $(\widehat{A}, \widehat{B}), (A_\star, B_\star) \in \mathcal{B}_{\mathrm{op}}(\frac{2}{3C_{\mathrm{safe}}(A_0,B_0)}, A_\star, B_\star) \subseteq \mathcal{B}_{\mathrm{op}}(\frac{(2)\cdot(1.5)}{3C_{\mathrm{safe}}(A_\star,B_\star)}, A_\star, B_\star)$, which means by triangle inequality that $\max\{\|\widehat{A} - A_\star\|_{\mathrm{op}}, \|\widehat{B} - B_\star\|_{\mathrm{op}}\} \le 1/C_{\mathrm{safe}}(A_\star, B_\star)$. □

B.1.4. Quality of First-Order Taylor Approximation

We bound the error of the first-order taylor expression in the following theorem.

Theorem 12. *There exists universal constants $c, p > 0$ such that the following holds. Let $(A_\star, B_\star)$ be stabilizable, and let $\epsilon_\circ := \max\{\|\widehat{A} - A_\star\|_\circ, \|\widehat{B} - B_\star\|_\circ\}$, and suppose that $\epsilon_{\mathrm{op}} \le 1/C_{\mathrm{safe}}(A_\star, B_\star)$ and $R_{\mathbf{x}} \succeq I$, $R_{\mathbf{u}} = I$. Let*

$$K' := \frac{d}{dt}K_\infty(A_\star + t(A_\star - \widehat{A}), B_\star + t(B_\star - \widehat{B})).$$

Then, $\|K_\infty(\widehat{A}, \widehat{B}) - (K_\star + K')\|_\circ \le c\|P_\star\|_{\mathrm{op}}^p \epsilon_{\mathrm{op}}^2 \epsilon_\circ^2$.

Proof. Consider the curve $K(t) = K_\infty(A(t), B(t))$ for $A(t) = (1-t)A_\star + t\widehat{A}$ and $B(t) = (1-t)B_\star + t\widehat{B}$. By Theorem 5, the curve $A(t), B(t)$ for $t \in [0,1]$ consists of all stabilizable matrices with $\|P_\infty(A(t), B(t))\|_{\mathrm{op}} \lesssim \|P_\star\|_{\mathrm{op}}$. By Lemma 3.1, the curve $K(t)$ is analytic on $[0,1]$. Moreover, from Lemma B.3 below, we have $\|K''(t)\|_\circ \le c_0\|P(t)\|_{\mathrm{op}}^p \epsilon_{\mathrm{op}}^2 \epsilon_\circ^2 \le c\|P_\star\|_{\mathrm{op}}^p \epsilon_{\mathrm{op}}^2 \epsilon_\circ^2$ for universal constants c_0, p. The bound now follows by Taylor's theorem. □

B.2. Key Derivative Computations

In the following computations, let $\Delta_A = \widehat{A} - A_\star$ and $\Delta_A := \widehat{B} - B_\star$. We recall $\epsilon_\circ := \max\{\|\Delta_A\|_\circ, \|\Delta_B\|_\circ\}$.

We consider derivatives allow curves $(A(t), B(t)) = (A_\star + t\Delta_A, B_\star + t\Delta_B)$, and associated functions $P(t) := P_\infty(A(t), B(t))$ and $K_\infty(A(t), B(t))$ defined at stabilizes $A(t), B(t)$. All proofs are given in Section C.2.

We begin by recalling the derivative computation from the main text, which also establishes local smoothness of $K(t)$ and $P(t)$.

Lemma 3.1 (Derivative and Smoothness of the DARE)**.** *Let $(A(t), B(t))$ be an analytic curve, and define $\Delta_{A_{\mathrm{cl}}}(t) := A'(t) + B'(t)K_\infty(A(t), B(t))$. Then for any t such that $(A(t), B(t))$ is stabilizable, the functions $P(u)$ and $K(u)$ are analytic in a neighborhood around t, and we have $P'(u) = \mathsf{dlyap}(A_{\mathrm{cl}}(u), Q_1(u))$, where $Q_1(u) := A_{\mathrm{cl}}(u)^\top P(u)\Delta_{A_{\mathrm{cl}}}(u) + \Delta_{A_{\mathrm{cl}}}(u)^\top P(u)A_{\mathrm{cl}}(u)$.*

Note that the above lemma allows for general analytic curves $(A(t), B(t))$. For our purposes, we restrict to linear curves given a above. For K', we have the following computation

Lemma B.1 (Computation of K')**.** *The first derivative of the optimal controller can be expressed as*

$$K' = -(R_{\mathbf{u}} + B^\top P B)^{-1}\left(\Delta_B^\top P A_{\mathrm{cl}} + B^\top P(\Delta_{A_{\mathrm{cl}}}) + B^\top P' A_{\mathrm{cl}}\right). \tag{B.2}$$

Of importance to our lower bound is the setting where the perturbations are of the form $(\Delta_A, \Delta_B) := (\Delta K_\star, \Delta)$. In this case, the expression for the derivative of K simplifies considerably. we recall the following from the main text

Lemma 2.1 (Derivative Computation (Abeille & Lazaric (2018), Proposition 2))**.** *Let $(A_\star, B_\star)$ be stabilizable, and recall $A_{\mathrm{cl},\star} := A_\star + B_\star K_\star$. Then,*

$$\frac{d}{dt}K_\infty(A_\star - t\Delta K_\star, B_\star + t\Delta)\big|_{t=0} = -(R_{\mathbf{u}} + B_\star^\top P_\star B_\star)^{-1} \cdot \Delta^\top P_\star A_{\mathrm{cl},\star}.$$

Proof. Observe that for the perturbation in question, $\Delta_{A_{\mathrm{cl}}}(0) = \Delta K_\star - \Delta K(0) = \Delta K_\star - \Delta K_\star = 0$. By Lemma 3.1 and the fact that $\mathsf{dlyap}(X, 0) = 0$, we have that $P'(0) = 0$. Thus, the term $B^\top P(\Delta_{A_{\mathrm{cl}}}) + B^\top P' A_{\mathrm{cl}}$ in Eq. (C.2) is 0 at $t = 0$. The result follows. □

B.2.1. Bounds on the Derivatives

Here, we state bounds on the various derivatives. Recall $\epsilon_\circ := \max\{\|\Delta_A\|_\circ, \|\Delta_B\|_\circ\}$. These bounds are established in Sections C.3.1 and C.3.2, respectively.

Lemma 3.2 (Bound on First Derivatives)**.** *Let $(A(t), B(t))$ be an analytic curve. Then, for all t at which $(A(t), B(t))$ is stabilizable, we have $\|P'(t)\|_\circ \le 4\|P(t)\|_{\mathrm{op}}^3 \epsilon_\circ$,and $\|K'(t)\|_\circ \le 7\|P(t)\|_{\mathrm{op}}^{7/2} \epsilon_\circ$.*

In fact, it will be more useful to prove the following related bound.

Lemma B.2. $\|R_{\mathbf{u}}^{1/2} K'\|_\circ \vee \|P^{1/2} B K'\|_\circ \vee \|BK'\|_\circ \le 7\|P\|_{\mathrm{op}}^{7/2}\epsilon_\circ$.

For our lower bounds, we shall also use a second-order derivative bound

Lemma B.3 (Bound on K'')**.** *If $\epsilon_\circ = \max\{\|A_\star - \widehat{A}\|_\circ, \|B_\star - \widehat{B}\|_\circ\}$ and $K(t) = K_\infty(A(t), B(t))$ for $A(t) = (1-t)A_\star + t\widehat{A}$ and $B(t) = (1-t)B_\star + t\widehat{B}$, that at any t at which $(A(t), B(t))$ is stabilizable,*

$$\|K''(t)\|_\circ \le \mathrm{poly}(\|P(t)\|_{\mathrm{op}})\epsilon_{\mathrm{op}}\epsilon_\circ.$$

B.3. Main Control Theory Tools

B.3.1. Properties of the dlyap Operator

We begin by describing relevant facts about dlyap operator. The first is standard (see e.g. (Bof et al., 2018; Boyd, 2008)), and gives a closed-form expression for the function.

Lemma B.4. *Let $Y = Y^\top$ and $\rho(X) < 1$. Then $\mathcal{T}_X(Y) := X^\top Y X - Y$ is an invertible map from $\mathbb{S}^{d_{\mathbf{x}}} \to \mathbb{S}^{d_{\mathbf{x}}}$, and*

$$\mathsf{dlyap}(X, Y) = \mathcal{T}_X^{-1}(Y) = \sum_{k=0}^{\infty} (X^\top)^k Y X^k. \tag{B.3}$$

Next, we show that dlyap is order-preserving in the following sense.

Lemma B.5 (Elementary dlyap bounds)**.** *The following bounds hold*

1. *If $Y \preceq Z$ and A_{safe} is stable, then $\mathsf{dlyap}(A_{\mathrm{safe}}, X) \preceq \mathsf{dlyap}(A_{\mathrm{safe}}, Y)$.*
2. *$Y \succeq 0$ and A_{safe} is stable, $\mathsf{dlyap}(A_{\mathrm{safe}}, Y) \succeq Y$.*
3. *Suppose $R_{\mathbf{x}} \succeq I$, and let $A + BK$ is stable. Then,*
$$\pm \mathsf{dlyap}(A + BK, Y) \preceq \mathsf{dlyap}(A + BK, I)\|Y\|_{\mathrm{op}} \preceq \|Y\|_{\mathrm{op}} \cdot P_\infty[K; A, B].$$
4. *When $R_{\mathbf{x}} \succeq I$, $\mathsf{dlyap}[A + BK] \preceq P_\infty[K; A, B]$, and $I \preceq \mathsf{dlyap}[A + BK_\infty(A, B)] \preceq P_\infty(A, B)$.*
5. *If A_{safe} is stable, $\|\mathsf{dlyap}[A_{\mathrm{safe}}]\|_{\mathrm{op}} = \|\mathsf{dlyap}[A_{\mathrm{safe}}^\top]\|_{\mathrm{op}}$.*

Next, we give a standard identity which relates the cost functions J to the dlyap operator.

Lemma B.6 (PSD bounds on P)**.** *Let $(A_\star, B_\star)$ be a stabilizable system, and let $A_\star + B_\star K$ be stable. Set $K_\star = K_\infty(A_\star, B_\star)$. Then,*

$$P_\infty[K; A_\star, B_\star] \succeq P_\infty(A_\star, B_\star) = P_\infty(K_\star; A_\star, B_\star).$$

Moreover, we have $\mathcal{J}^\star_{A_\star, B_\star}[K] = \mathrm{tr}(P_\infty[K; A, B])$, and in particular, $\mathcal{J}^\star_{A_\star, B_\star} = \mathcal{J}_{A_\star, B_\star}[K_\star] = \mathrm{tr}(P_\infty(A_\star, B_\star))$. As a conseuqnece, if $R_{\mathbf{x}} \succeq I$, then $\mathcal{J}^\star_{A_\star, B_\star}[K] \geq \mathcal{J}^\star_{A_\star, B_\star} \geq d_{\mathbf{x}}$ by Lemma B.5 part 4.

The following is a consequence of the above lemmas, and is useful for deriving interpretable corollaries of our main results.

Lemma B.7. *Suppose that $R_{\mathbf{x}} = I$. If $A_\star$ is (γ, κ)-strongly stable, then $\|P_\star\|_{\mathrm{op}} \leq \gamma_{\mathrm{sta}}^{-1}$ and $\frac{1}{d_{\mathbf{x}}} \mathcal{J}_{A_\star, B_\star}[0] \leq \gamma_{\mathrm{sta}}^{-1}$. More generally, if $(A_\star + B_\star K)$ is (γ, κ)-strongly stable, then $\|P_\star\|_{\mathrm{op}} \leq \gamma_{\mathrm{sta}}^{-1}(1 + \|K\|_{\mathrm{op}}^2)$.*

Proof of Lemma B.7. By considering the controller $K = 0$, Lemma B.6 implies $P_\star \preceq \mathsf{dlyap}[A_\star, R_{\mathbf{x}}]$ and $\mathcal{J}_{A_\star, B_\star}[0] = \mathrm{tr}(\mathsf{dlyap}[A_\star, R_{\mathbf{x}}]) \leq \frac{\|\mathsf{dlyap}[A_\star, R_{\mathbf{x}}]\|_{\mathrm{op}}}{d_{\mathbf{x}}}$. If $R_{\mathbf{x}} = I$, and we can bound

$$\|\mathsf{dlyap}[A_\star, I]\|_{\mathrm{op}} \leq \sum_{t \geq 0} \|A^t\|_{\mathrm{op}}^2.$$

If there exists a transform T with $\sigma_{\max}(T)/\sigma_{\min}(T) \leq \kappa$ such that $\|T A_\star T^{-1}\|_{\mathrm{op}} \leq 1 - \gamma$, then $\|A^t\|_{\mathrm{op}} \leq \kappa(1 - \gamma)$. Hence, $\|P_\star\|_{\mathrm{op}} \leq \|\mathsf{dlyap}[A_\star, I]\|_{\mathrm{op}} \leq \kappa^2 \sum_{t \geq 0} \gamma^2 \leq \frac{\kappa^2}{1-(1-\gamma)^2} \leq \frac{\kappa^2}{1-(1-\gamma)} = \kappa^2 \gamma^{-1}$. More generally, we have that $P_\star \preceq \mathsf{dlyap}[A_\star + B_\star K, R_{\mathbf{x}} + K^\top R_{\mathbf{u}} K]$ for $R_{\mathbf{x}}, R_{\mathbf{u}} = I$, $R_{\mathbf{x}} + K^\top R_{\mathbf{u}} K \preceq (1 + \|K\|_{\mathrm{op}}^2) I$, and the bound follows by invoking Lemma B.5. □

B.3.2. Helpful Norm Bounds

Lemma B.8 (Helpful norm bounds)**.** *Let $(A_\star, B_\star)$ be given, with $P_\star = P_\infty(A_\star, B_\star)$, $K_\star = K_\infty(A_\star, B_\star)$, and $A_{\mathrm{cl},\star} = A_\star + B_\star K_\star$. If $R_{\mathbf{x}} \succeq I$, $R_{\mathbf{u}} = I$, then the following bounds hold:*

1. *$P_\star \succeq I$, so that $\|P_\star^{-1}\|_2 \leq 1$, and $\|P_\star\|_{\mathrm{op}} \geq 1$.*
2. *$\|K_\star\|_{\mathrm{op}}^2 \leq \|P_\star\|_{\mathrm{op}}$ and $\|A_{\mathrm{cl},\star}\|_{\mathrm{op}}^2 \leq \|P_\star\|_{\mathrm{op}}$.*
3. *More generally, if $(A_\star + B_\star K)$ is stable, $K^\top K \preceq P_\infty(K; A_\star, B_\star) = \mathsf{dlyap}(A_\star + B_\star K, R_{\mathbf{x}} + K^\top R_{\mathbf{u}} K)$.*

B.3.3. BOUNDS ON $P_\infty(K; A_\star, B_\star)$ AND $\mathcal{J}_{A_\star,B_\star}[K]$

We now state a variant of a result due to (Fazel et al., 2018), which bounds the effect of perturbations on $P_\infty(K; A_\star, B_\star) - P_\infty(A_\star, B_\star)$.

Lemma B.9 (Generalization of Lemma 12 of (Fazel et al., 2018), see also Eq 3.2 in (Ran & Vreugdenhil, 1988))**.** *Let K be an arbitrary static controller which stabilizes $A_\star, B_\star$. Then,*

$$P_\infty(K; A_\star, B_\star) - P_\infty(A_\star, B_\star) = \mathsf{dlyap}(A_\star + B_\star K, (K - K_\star)^\top (R_{\mathbf{u}} + B_\star^\top P_\star B_\star)(K - K_\star)).$$

As a consequence of Lemma B.9 and B.5, we have the following corollary.

Corollary 4. *Let K be any arbitrary static controller which stabilities $A_\star, B_\star$, and suppose $R_{\mathbf{u}} = I$. Define the* adjoint[9] *as $\Sigma^{\mathrm{adj}}_{A_\star,B_\star}[K] := \mathsf{dlyap}(A_\star + B_\star K, I)$ covariance matrix. Then,*

$$\mathcal{J}_{A_\star,B_\star}[K] - \mathcal{J}_{A_\star,B_\star} \le \|\Sigma^{\mathrm{adj}}_{A_\star,B_\star}[K]\|_{\mathrm{op}} \max\{\|R_{\mathbf{u}}^{1/2}(K - K_\star)\|_{\mathrm{F}}^2, \|P_\star^{1/2} B (K - K_\star)\|_{\mathrm{F}}^2\},$$
$$\|P_\infty(K; A_\star, B_\star) - P_\infty(A_\star, B_\star)\|_{\mathrm{op}} \le \|\Sigma^{\mathrm{adj}}_{A_\star,B_\star}[K]\|_{\mathrm{op}} \max\{\|R_{\mathbf{u}}^{1/2}(K - K_\star)\|_{\mathrm{op}}^2, \|P_\star^{1/2} B (K - K_\star)\|_{\mathrm{op}}^2\}.$$

B.3.4. LINEAR LYAPUNOV THEORY

We now state a classical result in Lyapunov theory (see, e.g. (Boyd, 2008)). Recall the notation $\mathsf{dlyap}[A] := \mathsf{dlyap}(A, I)$.

Lemma B.10. *For any $x \in \mathbb{R}^d_{\mathbf{x}}$ and stable A, we have $\mathsf{dlyap}[A] \succeq I$ and*

$$A^\top \mathsf{dlyap}[A] A \preceq (1 - \|\mathsf{dlyap}[A]\|_{\mathrm{op}}^{-1}) \cdot \mathsf{dlyap}[A].$$

Lemma B.11. *For any stable A, $\|A\|_{\mathcal{H}_\infty} \le 2\|\mathsf{dlyap}[A]\|_{\mathrm{op}}^{3/2}$. More generally, suppose that $P \succeq I$ is a matrix satisfying $(Ax)^\top \mathsf{dlyap}[A](Ax) \le (1 - \rho) x^\top P x$. Then,*

$$\|A\|_{\mathcal{H}_\infty} \le \sum_{t \ge 0} \|A^t\|_2 \le 2 \frac{\sqrt{\|P\|_{\mathrm{op}}}}{\rho}.$$

B.4. Proofs for Supporting Perturbation Upper Bounds

B.4.1. PROOF OF PROPOSITION 6

The proof is analogous to that of Proposition 4, except we also apply the derivative bound on $R_{\mathbf{u}}^{1/2} K'(t)$ $BK'(t)$ from Lemma B.2. That bound also gives

$$\|R_{\mathbf{u}}^{1/2} K'(t)\|_\circ \le 7\|P(t)\|_{\mathrm{op}}^{7/2} \epsilon_\circ$$
$$\|B_\star K'(t)\|_\circ = \epsilon_{\mathrm{op}} \|K'(t)\|_\circ + \|B(t) K'(t)\|_\circ \le (1 + \epsilon_{\mathrm{op}}) 7 \|P(t)\|_{\mathrm{op}}^{7/2} \epsilon_\circ \le 8 \|P(t)\|_{\mathrm{op}}^{7/2} \epsilon_\circ,$$

so that the desired bound follow by the mean value theorem.

Moreover, we have

$$\|P_\star^{1/2} B_\star K'(t)\|_\circ \le \|P_\star^{1/2} P(t)^{-1/2}\|_{\mathrm{op}} \|P B_\star K'(t)\|_\circ \le \|P_\star^{1/2} P(t)^{-1/2}\|_{\mathrm{op}} 8 \|P(t)\|_{\mathrm{op}}^{7/2} \epsilon_\circ.,$$

which translates to a bound of

$$\|P_\star^{1/2} B (K_\infty(\widehat{A}, \widehat{B}) - K_\star)\|_\circ \le \max_{t \in [0,1]} \|P_\star^{1/2} P(t)^{-1/2}\|_{\mathrm{op}} \le 7 \|P(t)\|_{\mathrm{op}}^{7/2} \epsilon_\circ.$$

Finally, by the mean value theorem, we can bound for $\epsilon_{\mathrm{op}} \le 1/32 \|P_\star\|_{\mathrm{op}}^3$ and $\alpha = 8\epsilon_{\mathrm{op}} \|P_\star\|_{\mathrm{op}}^2 \le 1/4$,

$$\|P(t) - P_\star\|_{\mathrm{op}} \le 4 \max_{t \in [0,1]} \|P(t)\|_{\mathrm{op}}^3 \epsilon_{\mathrm{op}} \le 4 \|P_\star\|_{\mathrm{op}}^3 (1 - \alpha)^{-3/2} \le \frac{1}{8} (4/3)^{3/2}.$$

[9] Note that the canonical state covariance matrix $\Sigma_{A_\star,B_\star}[K]$ is given by $\mathsf{dlyap}((A_\star + B_\star K)^\top, I)$. By Lemma B.6, we have that $\|\Sigma_{A_\star,B_\star}[K]\|_{\mathrm{op}} = \|\Sigma^{\mathrm{adj}}_{A_\star,B_\star}[K]\|_{\mathrm{op}}$

Since $P(t) \succeq I$, then, this implies that for $t \in [0,1]$, $P(t) \succeq (1 - \frac{1}{8}(4/3)^{3/2})P_\star$, yielding $\|P_\star^{1/2} P(t)^{-1/2}\|_{\mathrm{op}} \leq \sqrt{(1 - \frac{1}{8}(4/3)^{3/2})} \leq 9/8$. Hence, for this ϵ_{op}, we have

$$\|P_\star^{1/2} B_\star (K_\infty(\widehat{A}, \widehat{B}) - K_\star)\|_\circ \leq 9\|P(t)\|_{\mathrm{op}}^{7/2} \epsilon_\circ.$$

□

B.4.2. Proof of Proposition 7

Since the final bound we derive does not depend on the control basis, we may assume without loss of generality that $R_{\mathbf{u}} = I$. Recall the steady state covariance matrix $\Sigma^{\mathrm{adj}}_{A_\star, B_\star}[K_\star] := \mathsf{dlyap}(A_\star + B_\star K, I)$. We shall prove the following lemma.

Lemma B.12. *Suppose that* $\|B_\star(K - K_\star)\|_2 \leq 1/5 \|\|\Sigma^{\mathrm{adj}}_{A_\star, B_\star}[K_\star]\|\|_{\mathrm{op}}^{3/2}$, *then* $\|\Sigma^{\mathrm{adj}}_{A_\star, B_\star}[K]\| \leq 2\|\Sigma^{\mathrm{adj}}_{A_\star, B_\star}[K_\star]\|$.

Note that a similar result was given by Lemma 16 (Fazel et al., 2018); we give our proof using the self-bounding ODE method to demonstrate the generality of its scope, and to avoid dependence on system matrices. Noting that $\|\Sigma^{\mathrm{adj}}_{A_\star, B_\star}[K_\star]\|_{\mathrm{op}} \leq \|P_\star\|_{\mathrm{op}}$ as verified above, it is enough that $\|B_\star(K - K_\star)\|_2 \leq 1/5\|P_\star\|_{\mathrm{op}}$ to ensure that $\|\Sigma^{\mathrm{adj}}_{A_\star, B_\star}[K]\| \leq 2P_\star$. When this holds, we have by Corollary 4, we have (assuming $R_{\mathbf{u}} = I$)

$$\begin{aligned}
\mathcal{J}_{A_\star, B_\star}[K] - \mathcal{J}_{A_\star, B_\star} &\leq \|\Sigma^{\mathrm{adj}}_{A_\star, B_\star}[K]\|_{\mathrm{op}} \max\{\|R_{\mathbf{u}}^{1/2}(K - K_\star)\|_{\mathrm{F}}^2, \|P_\star^{1/2} B_\star (K - K_\star)\|_{\mathrm{F}}^2\} \\
&\leq \|P_\star\|_{\mathrm{op}} \max\{\|R_{\mathbf{u}}^{1/2}(K - K_\star)\|_{\mathrm{F}}^2, \|P_\star^{1/2} B_\star (K - K_\star)\|_{\mathrm{F}}^2\}, \\
\|P_\infty(K; A_\star, B_\star) - P_\star\|_{\mathrm{op}} &\leq \|P_\star\|_{\mathrm{op}} \max\{\|R_{\mathbf{u}}(K - K_\star)\|_{\mathrm{op}}^2, \|P_\star^{1/2} B_\star (K - K_\star)\|_{\mathrm{op}}^2\},
\end{aligned}$$

as needed.

Proof of Lemma B.12. We shall now use the self-bounding machinery developed above to bound $\Sigma^{\mathrm{adj}}_{A_\star, B_\star}[K]$. Introduce the straight curve $\widetilde{K}(t) := K_\star + t\Delta_K$, where $\Delta_K = K - K_\star$, and where the $\widetilde{(\cdot)}$ is to avoid confusion with the curve $K(t) = K_\infty(A(t), B(t))$. Let $\Sigma(t) = \mathsf{dlyap}(A_\star + B_\star \widetilde{K}(t), I)$, so that $\Sigma(0) = \Sigma^{\mathrm{adj}}_{A_\star, B_\star}[K_\star]$ and $\Sigma(1) = \Sigma^{\mathrm{adj}}_{A_\star, B_\star}[K]$.

By the definition of dlyap, we have that at all t for which $K(t)$ stabilizes $A_\star, B_\star$,

$$\Sigma(t) = (A_\star + B_\star \widetilde{K}(t))^\top \Sigma(t) (A_\star + B_\star \widetilde{K}(t)) + I.$$

We shall now prove that $\Sigma(t)$ satisfies a self-bounding relation analogous to Proposition 4.

Claim B.13. *For all* $t \in [0,1]$ *for which* $\Sigma(t)$ *is defined,* $\|\Sigma'(t)\|_{\mathrm{op}} \leq 2\|\Sigma(t)\|^{5/2} \|B_\star \Delta_K\|_{\mathrm{op}}$.

Proof. Taking a derivative with respect to Σ, we have

$$\Sigma'(t) = (A_\star + B_\star \widetilde{K}(t))^\top \Sigma'(t) (A_\star + B_\star \widetilde{K}(t)) + Q_\Sigma(t),$$

where $Q_\Sigma(t) = (B_\star \Delta_K)^\top \Sigma(t)(A_\star + B_\star \widetilde{K}(t)) + (A_\star + B_\star \widetilde{K}(t))^\top \Sigma(t) B_\star \Delta_K$. Thus, we can render

$$\Sigma'(t) = \mathsf{dlyap}(A_\star + B_\star \widetilde{K}(t), Q_\Sigma(t)).$$

By an argument analogus to Lemma B.5, we have $\pm\Sigma'(t) \preceq \|Q_\Sigma(t)\| \mathsf{dlyap}(A_\star + B_\star \widetilde{K}(t), I) = \|Q_\Sigma(t)\| \Sigma(t)$, yielding the self-bounding relation

$$\|\Sigma'(t)\|_{\mathrm{op}} \leq \|Q_\Sigma(t)\|_{\mathrm{op}} \|\Sigma(t)\|_{\mathrm{op}}.$$

Moreover, we can bound for $t \in [0,1]$

$$\begin{aligned}
\|Q_\Sigma(t)\|_{\mathrm{op}} &\leq 2\|\Sigma(t)\|_{\mathrm{op}} \|B_\star \Delta_K\|_{\mathrm{op}} \|A_\star + B_\star \widetilde{K}(t)\|_{\mathrm{op}} \\
&\leq 2\|\Sigma(t)\|_{\mathrm{op}}^{3/2} \|B_\star \Delta_K\|_{\mathrm{op}},
\end{aligned}$$

where we use thaat $\|A_\star + B_\star \widetilde{K}(t)\|_{\mathrm{op}}^2 \leq \|\mathsf{dlyap}(A_\star + B_\star \widetilde{K}(t), I)\|_{\mathrm{op}} = \|\Sigma(t)\|_{\mathrm{op}}$. □

We check explicitly that $\Sigma(t)$ corresponds to the solution of a valid implicit function with domain $\mathcal{U} := \{\Sigma^{\mathrm{adj}} : \Sigma^{\mathrm{adj}} > 0\}$ (using the more general second condition that ensures that $t \mapsto \Sigma(t)$ is a continuously differentiable funciton, which follows from the form of dlyap). Applying Corollary 3 with $p = 5/2$ and $c = 2\|B_\star \Delta_K\|_{\mathrm{op}}$, this yields that if $\alpha = (p-1)c\|\Sigma(0)\|_{\mathrm{op}}^{3/2} = 3\|\Sigma(0)\|_{\mathrm{op}}^{3/2}\|B_\star\Delta_K\|_{\mathrm{op}} < 1$, then, $\|\Sigma(1)\|_{\mathrm{op}} \le (1-u)^{-2/3}\|\Sigma(0)\|_{\mathrm{op}}$. In particular, if $\|B_\star\Delta_K\|_{\mathrm{op}} \le 1/5\|\Sigma(0)\|_{\mathrm{op}}^{3/2}$, then we can show $\|\Sigma(1)\|_{\mathrm{op}} \le 2\|\Sigma(0)\|_{\mathrm{op}}$. □

B.4.3. Proof of Proposition 9

Introduce the curve $A(t) = A_{\mathrm{safe}} + t\Delta_A$, where $\Delta_A = A_1 - A_{\mathrm{safe}}$, define $Y_z(t) := (zI - A(t))^{-1}$. Then, $\|A(t)\|_{\mathcal{H}_\infty} = \sup_{z\in\mathbb{T}} \|Y_z(t)\|_2$. Let us now use the self-bounding method to bound $\|Y_z(t)\|$. We can observe that

$$Y_z'(t) = (zI - A(t))^{-1}\Delta_A(zI - A(t))^{-1},$$

so that $\|Y_z'(t)\|_2 \le \|Y_z(t)\|_2^2\|\Delta_A\|$.. Since $Y_z(t)$ corresponds to the zeros of the valid implicit function $F)z(A, Y) = Y \cdot (zI - A(t)) - I$, Theorem 13 implies that, if $\|\Delta_A\| \le \frac{u}{\|A_{\mathrm{safe}}\|_{\mathcal{H}_\infty}} = \min_{z\in\mathbb{T}} \frac{u}{\|Y_z(0)\|_2}$, then we have $\|Y_z(1)\| \le \frac{1}{1-\alpha}\|Y_z(0)\|$ for all $z \in \mathbb{T}$. Hence,

$$\|A_1\|_{\mathcal{H}_\infty} = \max_{z\in\mathbb{T}} \|Y_z(1)\| \le \max_{z\in\mathbb{T}} \|Y_z(0)\| = \frac{1}{1-\alpha}\|A_{\mathrm{safe}}\|_{\mathcal{H}_\infty},$$

as needed. □

B.4.4. Proof of Proposition 10

Observe that we have

$$\begin{aligned}
&(\widehat{A}x)^\top \mathsf{dlyap}[A](\widehat{A}x \\
&\le (Ax)^\top \mathsf{dlyap}[A](Ax) + x^\top(\widehat{A} - A)^\top \mathsf{dlyap}[A]Ax + x^\top(\widehat{A} - A)^\top \mathsf{dlyap}[A](\widehat{A} - A)x \\
&\le (1 - \|\mathsf{dlyap}[A]\|_{\mathrm{op}}^{-1}) \cdot x^\top \mathsf{dlyap}[A]x + \|x\|_2^2 \left(\|\widehat{A} - A\|_{\mathrm{op}}\|A\|_{\mathrm{op}} + \|\widehat{A} - A\|_{\mathrm{op}}^2\right) \|\mathsf{dlyap}[A]\|_{\mathrm{op}}) \\
&\le (1 - \|\mathsf{dlyap}[A]\|_{\mathrm{op}}^{-1} + \left(\|\widehat{A} - A\|_{\mathrm{op}}\|A\|_{\mathrm{op}} + \|\widehat{A} - A\|_{\mathrm{op}}^2\right) \|\mathsf{dlyap}[A]\|_{\mathrm{op}})) \cdot x^\top \mathsf{dlyap}[A]x,
\end{aligned}$$

where we used that $\mathsf{dlyap}[A] \succeq I$. In particular, if

$$\|\widehat{A} - A\|_{\mathrm{op}} \le \frac{1}{4}\min\left\{\frac{1}{\|A\|_{\mathrm{op}}\|\mathsf{dlyap}[A]\|_{\mathrm{op}}}, \|\mathsf{dlyap}[A]\|_{\mathrm{op}}^{-1/2}\right\},$$

then, the above is at most, $(1 - \frac{1}{2}\|\mathsf{dlyap}[A]\|_{\mathrm{op}}^{-1}) \cdot x^\top \mathsf{dlyap}[A]x$. □

C. Supporting Proofs for Appendix B

C.1. Proofs for Main Technical Tools (Section B.3)

We begin with the following lemma, which follows from a standard computation.

C.1.1. Proof of Lemma B.5

Proof. Let $\rho(A_0) < 1$, and so from (B.3) we have that for any Z with $Y \preceq Z$ that

$$\mathsf{dlyap}(A_0, Y) = \sum_{k=0}^{\infty} (A_0^k)^\top Y A_0^k \preceq \sum_{k=0}^{\infty} (A_0^k)^\top Z (A_0^k).$$

Second, if $Y \succeq 0$, $\sum_{k=0}^{\infty} (A_0^k)^\top Y (A_0^k) \mathsf{dlyap}(A_0, Y) \succeq Y$.

The third statement is a direct consequence of the first. Moreover, since $I \preceq R_{\mathbf{x}} \preceq R_{\mathbf{x}} + K^\top R_{\mathbf{u}} K$, taking $Z = \|Y\|(R_{\mathbf{x}} + K^\top R_{\mathbf{u}} K)$ yields the fourth inequality.

For the last statement, let $\|x\|_2 = 1$. Then, we have

$$\begin{aligned} x^\top \mathsf{dlyap}[A_0]x = \sum_{k=0}^{\infty} x^\top (A_0^\top)^k (A_0^k)x &= \sum_{k=0}^{\infty} \mathrm{tr}(A_0^k x x^\top (A_0^k)^\top) \\ &= \mathsf{dlyap}(A_0, xx^\top) \\ &\preceq \|xx^\top\|_{\mathrm{op}} \|\mathsf{dlyap}(A+BK, I)\|_{\mathrm{op}} I = \|\mathsf{dlyap}(A+BK, I)\|_{\mathrm{op}} I. \end{aligned}$$

□

C.1.2. Proof of Lemma B.6

We begin with the following lemma, whose proof is a straightforward computation.

Lemma C.1. *Let $A_\star, B_\star$ be stabilizable. For a controler K such that $A_\star + B_\star K$ is stable, we define the* value function

$$V^K(x) := \sum_{t=0}^{\infty} c(x_t^{K,x}, K x_t^{K,x}), \quad \textit{where } x_0^{K,x} = x, \quad \textit{and} \quad x_t^{K,x} = (A_\star + B_\star K) x_{t-1}^{K,x}.$$

We then have $x_t^{K,x} = (A_\star + B_\star K)^t x$, $\sum_{t=0}^{\infty} (x_t^{K,x})^\top Y x_t^{K,x} = \mathsf{dlyap}(A_\star + B_\star K, Y)$, and in particular,

$$V^K(x) = x^\top \mathsf{dlyap}(A_\star + B_\star K, R_{\mathbf{x}} + K^\top R_{\mathbf{u}} K)x = x^\top P_\infty(K) x,$$

We now prove Lemma B.6.

Proof. Introduce the shorthand $P_\star = P_\infty(A_\star, B_\star)$, $P_\infty(K) = P_\infty(K; A_\star, B_\star)$. and in particular, $V^{K_\infty}(x) = x^\top P_\infty(x)$. It is well known that $x^\top P_\infty x = V^{K_\infty}(x)$ and that $V^{K_\infty}(x) = \inf_K V^K(x) \le V^K(x)$ (Bertsekas, 2005). Hence, $P_\infty(K) \succeq P_\infty$. Finally, observe that by using that $A + BK$ is stable, we have

$$\begin{aligned} \mathcal{J}_{A,B}[L] &= \lim_{t\to\infty} \frac{1}{t} \sum_{i=1}^{t} \mathbb{E}_{A,B,K}[\mathbf{x}_i^\top R_{\mathbf{x}} \mathbf{x}_i + \mathbf{u}_i^\top R_{\mathbf{u}} \mathbf{u}_i] \\ &= \lim_{t\to\infty} \mathbb{E}_{A,B,K}[\mathbf{x}_t^\top R_{\mathbf{x}} \mathbf{x}_t + \mathbf{u}_t^\top R_{\mathbf{u}} \mathbf{u}_t] \\ &= \mathrm{tr}\left(\sum_{s=0}^{\infty} \left((A+BK)^\top\right)^s (R_{\mathbf{x}} + K^\top R_{\mathbf{u}} K)(A+BK)^s \right), \\ &= \mathrm{tr}(P_\infty(K)). \end{aligned}$$

The identity for P_∞ is the special case where $K = K_\infty$. □

C.1.3. Proof of Lemma B.8

Proof. We address each bound in succession.

1. $\sigma_{\min}(P_\star) \ge 1$ by Lemma B.5.
2. We have that

$$P_\infty = \mathsf{dlyap}(A_\star + B_\star K_\star, R_{\mathbf{x}} + K_\star^\top R_{\mathbf{u}} K_\star) \succeq R_{\mathbf{x}} + K_\star^\top R_{\mathbf{u}} K_\star \succeq K_\star^\top K_\star,$$

since $R_{\mathbf{u}} \succeq I$ and $R_{\mathbf{x}} \succeq I$. Moreover, we have that

$$\begin{aligned} P_\infty = \mathsf{dlyap}(A_\star + B_\star K_\star R_{\mathbf{x}} + K_\star^\top R_{\mathbf{u}} K_\star) &= \sum_{t=0}^{t} ((A_\star + B_\star K_\star)^\top)^t (R_{\mathbf{x}} + K_\star^\top R_{\mathbf{u}} K_\star)(A_\star + B_\star K_\star)^t \\ &\succeq \sum_{t=0}^{t} ((A_\star + B_\star K_\star)^\top)^t (A_\star + B_\star K_\star)^t \\ &\succeq (A_\star + B_\star K_\star)^\top (A_\star + B_\star K_\star). \end{aligned}$$

□

C.1.4. PROOF OF LEMMA B.9

Proof. The first inequality is precisely Lemmas 12 in (Fazel et al., 2018). In light of Lemma C.1, it suffices to show that

$$V^K(x) - V^{K_\star}(x) = x^\top \mathsf{dlyap}(A_\star + BK_\star, (K-K_\star)^\top(R_{\mathbf{u}} + B_\star^\top P_\star B_\star)(K-K_\star))$$

Lemma 10 in (Fazel et al., 2018) implies (noting $E_{K_\star} = 0$ for E_K defined therein) that

$$\begin{aligned} V^K(x) - V^{K_\star}(x) &= \sum_{t=0}^{\top} (x_t^{K,x})^\top (K-K_\star)^\top (R_{\mathbf{u}} + B_\star^\top P_\star B_\star)(K-K_\star) x_t^{K,x} \\ &= \mathsf{dlyap}(A_\star + B_\star K, (K-K_\star)^\top (R_{\mathbf{u}} + B_\star^\top P_\star B_\star)(K-K_\star)), \end{aligned}$$

where the second inequality uses Lemmas 12 in (Fazel et al., 2018).

□

C.1.5. PROOF OF LEMMA B.11

Let us prove the more general claim.

$$\begin{aligned} \|A\|_{\mathcal{H}_\infty} \le \sum_{t=0}^{\infty} \|A^i\|_{\mathrm{op}} &= \sum_{t=0}^{\infty} \sqrt{\|(A^i)^\top (A^i)\|_{\mathrm{op}}} \\ &\le \sum_{t=0}^{\infty} \sqrt{\frac{1}{\sigma_{\min}(\mathsf{dlyap}[A])} \|(A^i)^\top P (A^i)\|_{\mathrm{op}}} \\ &\le \sum_{t=0}^{\infty} \frac{1}{\sigma_{\min}(P)} \sqrt{(1-\rho)^i \|P\|_{\mathrm{op}}} \\ &\le \|P\|_{\mathrm{op}}^{1/2} \sum_{t=0}^{\infty} \sqrt{1-\rho}^{\,i} \quad \text{since } P \succeq I \\ &= \|P\|_{\mathrm{op}}^{1/2} \frac{1}{1-\sqrt{1-\rho}} \\ &\le \|P\|_{\mathrm{op}}^{1/2} \frac{1+\sqrt{1-\rho}}{1-(1-\rho} \\ &\le 2\|P\|_{\mathrm{op}}^{1/2}/\rho. \end{aligned}$$

C.2. Derivative Computations

C.2.1. PROOF OF LEMMA 3.1

Recall the function

$$\mathcal{F}_{\mathsf{DARE}}([A,B],P) = A^\top PA - P - A^\top PB(R_{\mathbf{u}} + B^\top PB)^{-1} B^\top PA + R_{\mathbf{x}}.$$

Let us compute the differentiable of this map. To keep notation, let us suppress the dependence of the A, B arguments on t. We have that

$$\begin{aligned} \mathsf{D}\mathcal{F}_{\mathsf{DARE}}[\mathrm{d}P, \mathrm{d}t]\big|_{A(t),B(t),P} &= \mathsf{D}(A^\top PA) - \mathsf{D}P + \mathsf{D}(A^\top PB) \cdot (R_{\mathbf{u}} + B^\top PB)^{-1} B^\top PA \\ &\quad - (A^\top PB)(R_{\mathbf{u}} + B^\top PB)^{-1} \cdot (B^\top PA)\mathsf{D} \\ &\quad - (A^\top PB) \cdot \mathsf{D}((R_{\mathbf{u}} + B^\top PB)^{-1}) \cdot B^\top PA \\ &= \mathsf{D}(A^\top PA) + \mathsf{D}(A^\top PB) \cdot K + K^\top \cdot \mathsf{D}(B^\top PA) \\ &\quad - (A^\top PB) \cdot \mathsf{D}((R_{\mathbf{u}} + B^\top PB)^{-1}) \cdot B^\top PA, \end{aligned}$$

where for compactness, we substituted in the formula

$$K = K(t, P) = -(R_{\mathbf{u}} + B(t)^\top P B(t))^{-1} B(t)^\top P A(t). \tag{C.1}$$

Recall that for a symmetric matrix, we have $((X^{-1})' = -X^{-1}X'X^{-1})$. Thus, substituting in the definition of K, we can write the last term in the expression above as

$$\begin{aligned}
&- (A^\top P B) \cdot \mathsf{D}((R_{\mathbf{u}} + B^\top P B))' B^\top P A \\
&= (A^\top P B)(R_{\mathbf{u}} + B^\top P B)^{-1}(R_{\mathbf{u}} + B^\top P B)\mathsf{D}(R_{\mathbf{u}} + B^\top P B)^{-1} B^\top P A \\
&= K^\top \mathsf{D}(R_{\mathbf{u}} + B^\top P B) K.
\end{aligned}$$

Hence, gathering terms, we have

$$\mathsf{D}\mathcal{F}_{\mathsf{DARE}}[\mathrm{d}P, \mathrm{d}t]\big|_{A(t),B(t),P} = \mathsf{D}(A^\top P A) - \mathsf{D}(P) + \mathsf{D}(A^\top P B)K + K^\top \cdot \mathsf{D}(B^\top P A) + K^\top \mathsf{D}(R_{\mathbf{u}} + B^\top P B)K.$$

Let us now adopt shorthand $(\cdot)' := \frac{d}{dt}(\cdot)$. Expanding the derivatives using the product rule, we then have

$$\begin{aligned}
\mathsf{D}\mathcal{F}_{\mathsf{DARE}}[\mathrm{d}P, \mathrm{d}t]\big|_{A(t),B(t),P} = {} & A^\top \mathsf{D}P A - \mathsf{D}(P) + A^\top \mathsf{D}B K + K^\top B^\top \mathsf{D}A + K^\top B^\top \cdot \mathsf{D}P \cdot BK \\
& + A'^\top P A + A^\top P A' + A'^\top B K + (BK)^\top P A' \\
& + A^\top P (B'K) + (B'K)^\top P A + (B'K)^\top P B K + (BK)^\top P (B'K).
\end{aligned}$$

Grouping terms, this is equal to

$$\begin{aligned}
\mathsf{D}\mathcal{F}_{\mathsf{DARE}}[\mathrm{d}P, \mathrm{d}t]\big|_{A(t),B(t),P} &= (A + BK)^\top \mathrm{d}P (A + BK) - \mathrm{d}P\big|_{A(t),B(t),P} \\
&\quad + A'^\top P (A + BK) + (A + BK)^\top P A'\big|_{A(t),B(t),P} \\
&\quad + (B'K)'^\top P (A + BK) + (A + BK)^\top P (B'K)\big|_{A(t),B(t),P} \\
&= (A + BK)^\top \cdot \mathrm{d}P \cdot (A + BK) - \mathrm{d}P\big|_{A(t),B(t),P} \\
&\quad + \underbrace{(A'(t) + B'(t)K)^\top P (A(t) + B(t)K) + (A(t) + B(t)K) P (A'(t) + B'(t)K)}_{:= Q_1(t,P),\ \text{and } K = (t,P) \text{ as in Eq. (C.1)}} \\
&= \mathcal{T}_{A(t)+B(t)K}[\mathrm{d}P] + Q(t, P)\mathrm{d}t.
\end{aligned}$$

In particular, if $\mathcal{F}_{\mathsf{DARE}}([A(t), B(t)], P) = 0$, then for $K(t, p)$ as in Eq. (C.1), the matrix $A(t) + B(t)K(t, P)$ is stable. Hence, $\mathcal{T}_{A(t)+B(t)K(t,P)}[\cdot]$ is invertible on $\mathbb{S}^d$. Moreover, since the second term has no-explicit depending on $\mathrm{d}P$, we find that $(\mathrm{d}P, \mathrm{d}t) \mapsto \mathsf{D}\mathcal{F}_{\mathsf{DARE}}[\mathrm{d}P, \mathrm{d}t]\big|_{A(t),B(t),P}$ is full-rank, with zero solution

$$\mathrm{d}P = \mathcal{T}^{-1}_{A(t)+B(t)K(t,P)}[Q_1(t, P)\mathrm{d}t] = \mathsf{dlyap}(A(t) + B(t)K(t, P), Q_1(t, P)).$$

By the implicit function theorem, this implies that there if $\mathcal{F}_{\mathsf{DARE}}([A(t), B(t)], P) = 0$, then there exists a neighborhood around t on which the function $u \mapsto P(u)$ is analytic (recall $\mathcal{F}_{\mathsf{DARE}}$ is analytic), and $\mathcal{F}_{\mathsf{DARE}}([A(u), B(u)], P(u)) = 0$ on this neighborhood.By the above display then, we have $P'(u) = \mathsf{dlyap}(A(u)+B(u)K(u), Q_1(u))$, where $Q_1(u) \leftarrow Q_1(u, P(u))$ and $K(t) \leftarrow K(u, P(u))$ are specializations of the above to the curve $u \mapsto P(u)$.

□

C.2.2. Computation of K' (Lemma B.1)

Throughout, we suppress dependence on t, and the computations are understood to hold only at those t for which $(A(t), B(t))$ is stabilizable.

Proof. Note that we can take derivatives freely by Lemma 3.1. Invoking the product rule and the identity $((X^{-1})' = -X^{-1}X'X^{-1})$,

$$\begin{aligned}
K' &= (R_{\mathbf{u}} + B^\top P B)^{-1} \cdot (R_{\mathbf{u}} + B^\top P B)' \cdot (R_{\mathbf{u}} + B^\top P B)^{-1} B^\top P A - (R_{\mathbf{u}} + B^\top P B)^{-1} \cdot (B^\top P A)' \\
&= -(R_{\mathbf{u}} + B^\top P B)^{-1}(R_{\mathbf{u}} + B^\top P B)' \cdot K - (R_{\mathbf{u}} + B^\top P B)^{-1}(B^\top P A)' \\
&= -(R_{\mathbf{u}} + B^\top P B)^{-1} \left((R_{\mathbf{u}} + B^\top P B)' K + (B^\top P A)' \right).
\end{aligned}$$

We simplify the expression inside the parentheses as

$$\begin{aligned}(R_{\mathbf{u}} + B^\top PB)'K + (B^\top PA)' &= B'^\top P(A+BK) + B^\top P(A' + B'K) + B^\top P'(A+BK) \\ &= B'^\top PA_{\mathrm{cl}} + B^\top P(\Delta_{A_{\mathrm{cl}}}) + B^\top P' A_{\mathrm{cl}}.\end{aligned}$$

Since $B' = \Delta_B$, this yields the result. □

C.2.3. Computation of P''

Again, suppress dependence on t. We compute P'', which Lemma 3.1 ensures exists whenever $(A(t), B(t))$ is stabilizable.

Lemma C.2 (Computation of P'')**.** *The second derivative of the optimal cost matrix has the form*

$$P'' = \mathsf{dlyap}(A_{\mathrm{cl}}, Q_2),$$

where $Q_2 := A_{\mathrm{cl}}'^\top P' A_{\mathrm{cl}} + A_{\mathrm{cl}}^\top P' A_{\mathrm{cl}}' + Q_1'$ is a symmetric matrix defined in terms of

$$Q_1' := A_{\mathrm{cl}}'^\top P(\Delta_{A_{\mathrm{cl}}}) + A_{\mathrm{cl}}^\top P' \Delta_{A_{\mathrm{cl}}} + A_{\mathrm{cl}}^\top P(B'K') + (B'K')^\top P A_{\mathrm{cl}} + \Delta_{A_{\mathrm{cl}}}^\top P' A_{\mathrm{cl}} + \Delta_{A_{\mathrm{cl}}}^\top P A_{\mathrm{cl}}'\,.$$

Proof. Applying the product rule to the expression for P' from Lemma 3.1, we have

$$\begin{aligned}P'' &= A_{\mathrm{cl}}^\top P'' A_{\mathrm{cl}} + A_{\mathrm{cl}}'^\top P' A_{\mathrm{cl}} + A_{\mathrm{cl}}^\top P' A_{\mathrm{cl}}' \\ &\quad + A_{\mathrm{cl}}'^\top P \Delta_{A_{\mathrm{cl}}} + A_{\mathrm{cl}}^\top P' \Delta_{A_{\mathrm{cl}}} + A_{\mathrm{cl}}^\top P(\Delta_{A_{\mathrm{cl}}})' + (\Delta_{A_{\mathrm{cl}}})'^\top P A_{\mathrm{cl}} + \Delta_{A_{\mathrm{cl}}}^\top P' A_{\mathrm{cl}} + \Delta_{A_{\mathrm{cl}}}^\top P A_{\mathrm{cl}}' \\ &= \mathsf{dlyap}(A_{\mathrm{cl}}, Q_2),\end{aligned}$$

where $Q_2 := A_{\mathrm{cl}}'^\top P' A_{\mathrm{cl}} + A_{\mathrm{cl}}^\top P' A_{\mathrm{cl}}' + A_{\mathrm{cl}}'^\top P \Delta_{A_{\mathrm{cl}}} + A_{\mathrm{cl}}^\top P' \Delta_{A_{\mathrm{cl}}} + A_{\mathrm{cl}}^\top P(\Delta_{A_{\mathrm{cl}}})' + (\Delta_{A_{\mathrm{cl}}})'^\top P A_{\mathrm{cl}} + \Delta_{A_{\mathrm{cl}}}^\top P' A_{\mathrm{cl}} + \Delta_{A_{\mathrm{cl}}}^\top P A_{\mathrm{cl}}'$.
We conclude by observing that $(\Delta_{A_{\mathrm{cl}}})' = (A'' + B''K + B'K') = B'K'$, since A and B are linear in t. □

C.3. Norm Bounds for Derivatives

C.3.1. Norm bounds for First Derivatives

In this section, we work through obtaining concrete bounds on the derivatives of $P(t), K(t)$ using the expressions derived in the previous section. As above, we assume that $R_{\mathbf{u}} \succeq I$ and $R_{\mathbf{x}} \succeq I$. We state some more bounds that will be of use to use.

Lemma C.3 (Norm-Bounds for Derivative Quantities)**.** *Let $(A_\star, B_\star)$ be given, with $P_\star = P_\infty(A_\star, B_\star)$, $K_\star = K_\infty(A_\star, B_\star)$, and $A_{\mathrm{cl},\star} = A_\star + B_\star K_\star$. If $R_{\mathbf{u}}, R_{\mathbf{x}} \succeq I$, then the following bounds hold:*

1. *Let $R_0 := R_{\mathbf{u}} + B_\star^\top P_\star B_\star$. Then for any $X, Y \in \{B_\star, P_\star^{1/2} B_\star, R_{\mathbf{u}}^{1/2}, I\}$, $\|X R_0^{-1} Y^\top\|_{\mathrm{op}} \le 1$.*
2. *For $\circ \in \{\mathrm{op}, \mathrm{F}\}$, we have $\|\Delta_{A_{\mathrm{cl}}}\|_\circ \le 2\|P\|_{\mathrm{op}}^{1/2} \epsilon_\circ$.*

Proof. First, we have that $\|X R_0^{-1} Y^\top\|_{\mathrm{op}} \le \|X R_0^{-1/2}\|_{\mathrm{op}} \|Y R_0^{-1/2}\|_{\mathrm{op}} \le \sqrt{\|X R_0^{-1} X^\top\|_{\mathrm{op}} \|Y R_0^{-1} Y^\top\|_{\mathrm{op}}}$. Since $R_{\mathbf{u}}, P \succeq I$, we can verify that $XX^\top, YY^\top \preceq R_0$, which means that $\|X R_0^{-1} X^\top\|_{\mathrm{op}}, \|Y R_0^{-1} Y^\top\|_{\mathrm{op}} \le 1$.

Second, for $\|\cdot\|_\circ$ denoting either the operator or Frobenius norm, we bound $\|\Delta_{A_{\mathrm{cl}}}\|_\circ = \|\Delta_A + \Delta_B K\|_\circ \le \|\Delta_A\|_\circ + \|\Delta_B\|_\circ \|K\|_{\mathrm{op}} = \epsilon_\circ(1 + \|K\|_{\mathrm{op}}) \le 2\sqrt{\|P\|_{\mathrm{op}}}\epsilon_\circ$. □

C.3.2. Proof of Lemma 3.2 and Lemma B.2

Recall that $P' = \mathsf{dlyap}(A_{\mathrm{cl}}, Q_1)$, where $Q_1 := A_{\mathrm{cl}}^\top P(\Delta_{A_{\mathrm{cl}}}) + (\Delta_{A_{\mathrm{cl}}})^\top P A_{\mathrm{cl}}$. Hence, using Lemma B.5 with $R_{\mathbf{x}} \succeq I$, followed by Lemmas B.8 and C.3, we can bound

$$\begin{aligned}\|P'\|_\circ &= \|\mathsf{dlyap}(A_{\mathrm{cl}}, Q_1)\|_\circ \\ &\le \|P\|_{\mathrm{op}} \|Q_1\|_\circ \le 2\|P\|_{\mathrm{op}}^2 \|A_{\mathrm{cl}}\|_{\mathrm{op}} \|\Delta_{A_{\mathrm{cl}}}\|_\circ \\ &\le 2\|P\|_{\mathrm{op}}^2 \cdot \|P\|_{\mathrm{op}}^{1/2} \cdot 2\|P\|_{\mathrm{op}}^{1/2} \epsilon_\circ = 4\|P\|_{\mathrm{op}}^3.\end{aligned}$$

Next, recall from Lemma B.1 that we have the identity

$$K' = -R_0^{-1}\left(\Delta_B^\top P A_{\mathrm{cl}} + B^\top P(\Delta_{A_{\mathrm{cl}}}) + B^\top P' A_{\mathrm{cl}}\right),$$

where $R_0 := R_{\mathbf{u}} + B^\top P B$. Next bound each of the three terms that arise. Again using $\|R_0^{-1}\|_{\mathrm{op}} \le 1$ and $\|A_{\mathrm{cl}}\|_{\mathrm{op}} \le \|P\|_{\mathrm{op}}^{1/2}$ (Lemma B.8), we have

$$\|R_0^{-1}\Delta_B P A_{\mathrm{cl}}\|_\circ \le \|P\|_{\mathrm{op}}^{3/2}\epsilon_\circ.$$

Next, since $\|R_0^{-1}B^\top P^{1/2}\|_{\mathrm{op}} \le 1$ (Lemma C.3), we have

$$\begin{aligned}\|(R_{\mathbf{u}} + B^\top P B)^{-1}\left(B^\top P(\Delta_{A_{\mathrm{cl}}}) + B^\top P' A_{\mathrm{cl}}\right)\|_\circ &\le \|P\|_{\mathrm{op}}^{1/2}\|\Delta_{A_{\mathrm{cl}}}\|_{\mathrm{op}} + \|P^{-1/2}\|_{\mathrm{op}}\|P'\|_{\mathrm{op}}\|A_{\mathrm{cl}}\|_{\mathrm{op}} \\ &\le 2\|P\|_{\mathrm{op}}\epsilon_\circ + \|P^{-1/2}\|_{\mathrm{op}}\|P'\|_{\mathrm{op}}\|P\|_{\mathrm{op}}^{1/2} \\ &\le 2\|P\|_{\mathrm{op}}\epsilon_\circ + 4\|P\|_{\mathrm{op}}^{7/2}\epsilon_\circ.\end{aligned}$$

where the second to last line uses Lemma B.8, and the last line uses $\|P^{-1/2}\|_{\mathrm{op}} \le 1$, as well as $\|P'\|_{\mathrm{op}} \le 4\|P\|_{\mathrm{op}}^3$. Putting the bounds together, we have $\|K'\|_\circ \le 7\|P\|_{\mathrm{op}}^{7/2}\epsilon_\circ$.

□.

We also restate and prove an an analogous bound that pre-conditions $K'(t)$ by appropriate matrices.

Lemma B.2. $\|R_{\mathbf{u}}^{1/2}K'\|_\circ \vee \|P^{1/2}BK'\|_\circ \vee \|BK'\|_\circ \le 7\|P\|_{\mathrm{op}}^{7/2}\epsilon_\circ$.

Proof. The bound is analogous to the bound on K' from Lemma 3.2, but now uses right multiplication of R_0^{-1} which adresses left-multiplication by $B, P^{1/2}B, R_{\mathbf{u}}^{1/2}$. □

C.3.3. Norm Bounds for Second Derivatives

Next, we turn to bounding P'' and K''. We shall need some intermediate lemmas. Let us bound the intermediate term A'_{cl}

Lemma C.4. *It holds that* $\max\{\|\Delta_{A_{\mathrm{cl}}}\|_\circ, \|A'_{\mathrm{cl}}\|_\circ\} \le 9\|P\|_2^{7/2}\epsilon_\circ$, *and* $\|\Delta'_{A_{\mathrm{cl}}}\|_\circ \le \epsilon_\circ\epsilon_{\mathrm{op}}\|P\|_{\mathrm{op}}^{7/2}$.

Proof. $A'_{\mathrm{cl}} = \Delta_{A_{\mathrm{cl}}} + BK'$. From Lemma C.3, $\|\Delta_{A_{\mathrm{cl}}}\|_\circ \le 2\sqrt{\|P\|_{\mathrm{op}}}\epsilon_\circ$. Moreover, from Lemma B.2, $\|BK'\|_\circ \le 7\|P\|_{\mathrm{op}}^{7/2}\epsilon_\circ$. Thus, $\|A'_{\mathrm{cl}}\|_\circ \le 9\|P\|_{\mathrm{op}}^{7/2}\epsilon_\circ$. The second bound uses $\Delta'_{A_{\mathrm{cl}}} = \Delta_B K'$, and the same bound on $\|K'\|_\circ$. □

Next, we bound the norm of P''.

Lemma C.5. *We have the bound* $\|P''\|_\circ \le \mathrm{poly}(\|P_\star\|_{\mathrm{op}})\epsilon_{\mathrm{op}}\epsilon_\circ$.

Proof. Recall that $P'' = \mathsf{dlyap}(A_{\mathrm{cl}}, Q_2)$, where

$$\begin{aligned}Q_2 = {} & A_{\mathrm{cl}}'^\top P' A_{\mathrm{cl}} + A_{\mathrm{cl}}^\top P' A'_{\mathrm{cl}} \\ & + A_{\mathrm{cl}}'^\top P(\Delta_{A_{\mathrm{cl}}}) + A_{\mathrm{cl}}^\top P'\Delta_{A_{\mathrm{cl}}} + A_{\mathrm{cl}}^\top P(B'K') + (B'K')^\top P A_{\mathrm{cl}} + \Delta_{A_{\mathrm{cl}}}^\top P' A_{\mathrm{cl}} + \Delta_{A_{\mathrm{cl}}}^\top P A'_{\mathrm{cl}}.\end{aligned}$$

Hence, $\|P''\|_{\mathrm{op}} \le \|P\|_{\mathrm{op}}\|Q_2\|_{\mathrm{op}}$. We upper bound the norm of Q_2 by

$$\|Q_2\|_\circ \le 2\left(\|A'_{\mathrm{cl}}\|_\circ\|P'\|_{\mathrm{op}}\|A_{\mathrm{cl}}\|_{\mathrm{op}} + \|A'_{\mathrm{cl}}\|_\circ\|\Delta_{A_{\mathrm{cl}}}\|_{\mathrm{op}}\|P\|_{\mathrm{op}} + \|B'\|_\circ\|K'\|_{\mathrm{op}}\|P A_{\mathrm{cl}}\|_{\mathrm{op}} + \|A_{\mathrm{cl}}\|_{\mathrm{op}}\|P'\|_{\mathrm{op}}\|\Delta_{A_{\mathrm{cl}}}\|_\circ\right).$$

Using Lemma B.8 and Lemma 3.2, one can show that

$$\|P''\|_\circ \le \mathrm{poly}(\|P_\star\|_{\mathrm{op}})\epsilon_{\mathrm{op}}\epsilon_\circ.$$

□

Proof of Lemma B.3 . From Lemma B.1, we have that

$$K' = -(R_{\mathbf{u}} + B^\top PB)^{-1}\left(\Delta_B^\top PA_{\mathrm{cl}} + B^\top P(\Delta_{A_{\mathrm{cl}}}) + B^\top P' A_{\mathrm{cl}}\right). \tag{C.2}$$

Denote $Q_3 := \Delta_B^\top PA_{\mathrm{cl}} + B^\top P(\Delta_{A_{\mathrm{cl}}}) + B^\top P' A_{\mathrm{cl}}$, and $R_0 := R_{\mathbf{u}} + B^\top PB$. Then, we have

$$\begin{aligned} K'' &= R_0^{-1}Q_3'(t) + R_0^{-1}(R_{\mathbf{u}} + B^\top PB)' R_0^{-1} Q_3(t) \\ &= R_0^{-1}Q_3'(t) + R_0^{-1}(R_{\mathbf{u}} + B^\top PB)' K'. \end{aligned}$$

Lets first handle the term $R_0^{-1}Q_3'(t)$. From Lemma C.3, we have that $\|R_0^{-1}\|_{\mathrm{op}} \le 1, \|R_0^{-1}B\|_{\mathrm{op}}$. Thus,

$$\begin{aligned} \|R_0^{-1}Q_3'(t)\|_\circ &\le \|R_0^{-1}\|_{\mathrm{op}}\|\Delta_B^\top PA_{\mathrm{cl}})'\|_\circ + \|R_0^{-1}B\|_{\mathrm{op}}\|(P'A_{\mathrm{cl}})'\|_\circ \\ &\le \|(\Delta_B^\top PA_{\mathrm{cl}})'\|_\circ + \|(P'A_{\mathrm{cl}})'\|_\circ \\ &\le \|\Delta_B\|_\circ\|P\|_{\mathrm{op}}A_{\mathrm{cl}}'\|P'\|_{\mathrm{op}}\|A_{\mathrm{cl}}\|_{\mathrm{op}} + \|P''\|_\circ\|A_{\mathrm{cl}}\|_{\mathrm{op}} + \|P'\|_\circ\|A_{\mathrm{cl}}'\|_{\mathrm{op}} \\ &\le \mathrm{poly}(\|P_\star\|_{\mathrm{op}})\epsilon_{\mathrm{op}}\epsilon_\circ, \end{aligned}$$

where we invoke the derivative computations above. Similarly, we can show that

$$\begin{aligned} \|R_0^{-1}(R_{\mathbf{u}} + B^\top PB)'K'\|_\circ &\le \|R_0^{-1}\Delta_B PBK'\|_\circ + \|R_0^{-1}B^\top P'BK'\|_\circ + \|R_0^{-1}BP\Delta_B K'\|_\circ \\ &\le (\|R_0^{-1}\|_{\mathrm{op}}\|P\|_{\mathrm{op}}\epsilon_{\mathrm{op}} + \|R_0^{-1}B^\top\|_{\mathrm{op}}\|P'\|_{\mathrm{op}})\|BK'\|_\circ + \|R_0^{-1}B\|_{\mathrm{op}}\|P\|_{\mathrm{op}}\epsilon_\circ\|K'\|_{\mathrm{op}} \\ &\le (\|P\|_{\mathrm{op}}\epsilon_{\mathrm{op}} + \|P'\|_{\mathrm{op}})\|BK'\|_\circ + \|P\|_{\mathrm{op}}\epsilon_\circ\|K'\|_{\mathrm{op}} \le \mathrm{poly}(\|P\|_{\mathrm{op}})\epsilon_\circ\epsilon_{\mathrm{op}}. \end{aligned}$$

□

D. Self-Bounding ODE Method

We begin by stating Theorem 13, which provides a generic guarantee for self-bounding ODES (Definition 3.3).

Theorem 13. *Let $(F, \mathcal{U}, g, \|\cdot\|, x(\cdot))$ be a self-bounding tuple. Suppose that for some $\eta > 0$, $h(\cdot)$ satisfies $h(z) \ge g(z) + \eta$ for all $z \ge \|y(0)\|$, and that the scalar ODE*

$$w(0) = \|y(0)\| + \eta, \quad w'(t) = h(w(t)$$

has a continuously differentiable solution on $[0, 1]$. Then, there exists a unique continuously differentiable function $y(t) \in \mathcal{U}$ defined on $[0, 1]$ which satisfies $F(x(t), y(t)) = 0$, and this solution satisfies $\|y(t)\| \le w(t) \le w(1)$, $\|y'(t)\| \le g(w(t)) \le g(w(1))$ for all $t \in [0, 1]$.

We shall prove the above theorem, and then derive Corollary 3 as a consequence. We begin the proof of this theorem with a simple scalar comparison inequality.

Lemma D.1 (Scalar Comparison Inequalities for Curves)**.** *Suppose that $x(t), w(t)$ are continuously differentiable curves defined on $[0, u)$. Suppose further that, for a function $f(\cdot, \cdot)$, $x'(t) = f(x(t), t)$, and that $w'(t) = g(x(t))$. In addition, suppose*

1. *$w(0) > x(0)$*
2. *$g(\cdot) \ge 0$*
3. *For $t \in [0, u)$ such that $x(t) \ge w(0)$, $g(x(t)) > f(x(t), t)$.*

Then, $x(t) < w(t)$ for $t \in [0, u)$.

Proof. Define $\delta(t) = w(t) - x(t)$. Since $\delta(0) > 0$, there exists an $s > 0$ such that $\delta(t)$ for $t \in [0, s)$. Choose the maximal such $s := \sup\{t : \delta(t') \ge 0, \forall t' < t\}$, and suppose for the sake of contradiction that $s < u$. Then, by continuity, $\delta(s) = 0$, and therefore $\delta'(s) = g(w(s)) - f(x(s), s) = g(x(s)) - f(x(s), s)$, since $x(s) = w(u)$ for $\delta(s) = 0$.

Next, note that since $g(\cdot) \ge 0$, $w(t)$ is non-decreasing on $[0, u)$, and thus $w(s) \ge w(0)$ for all $s \in [0, u)$. Since $x(s) = w(s)$ at s, we have $x(s) \ge w(0)$ as well. Thus, $\delta'(s) = g(x(s)) - f(x(s), s) > 0$, by the assumption of the lemma. Hence, for an $\epsilon > 0$ sufficiently small, $\delta(s - \epsilon) < \delta(s) = 0$. This contradicts the fact that of $\delta(t') = 0$ for all $t' < s$. □

Next, we extend the above scalar comparison inequality to a comparison inequality between scalar ODEs, and vector ODEs.

Lemma D.2 (Norm Comparison for Vector ODE)**.** *Let $\|\cdot\|$ denote an arbitrary norm. Suppose that $v(t) \in \mathbb{R}^d$ is a continuously differentiable curve defined on $[0, u)$ such that $\|v'(t)\| \le g(\|v(t)\|)$ for a non-decreasing function g. Fix $\eta > 0$, and let $h(z)$ denote a function such that $h(z) \ge \max\{0, g(z) + \eta\}$ for all $z \ge \|v(0)\|$. Then, if the ODE*

$$w(0) = \|v(0)\| + \eta, \quad w'(t) = h(w(t))$$

has a continuously differentiable solution defined on $[0, u)$, then $\|v(t)\| \le w(t)$ for all $t \in [0, u)$

Proof of Lemma D.2. The main challenge is that $\|\cdot\|$ may be non-smooth. We circumvent this with a Gaussian approximation. Let $c_Z := \mathbb{E}_{Z\sim\mathcal{N}(0,I)}[\|Z\|]$, and for every $\eta > 0$, and define $\Psi_\eta(v) := \mathbb{E}_{Z\sim\mathcal{N}(0,I)}[\|v + \frac{\eta}{2c_Z}Z\|]$. Defining $c_Z := \mathbb{E}_{Z\sim\mathcal{N}(0,1)}[\|Z\|]$. Moreover, we can see that $0 \le \|v\| \le \Psi_\eta(v) \le \|v\| + \eta/2 < \|v\| + \eta$ by Jensen's inequality and the triangle inequality. Consider the curve $x(t)$

$$x(t) = \Psi_\eta(v(t)), t \in [0, u),$$

Note then that the curve satisfies

$$x(0) = \Psi_\eta(v(0)), \quad \text{and} \quad x'(t)) = f(t, x(t)) = \frac{\mathrm{d}}{\mathrm{d}t}(\Psi_\eta(v(t))),$$

where $f(t, x(t))$ does not depend implicitly on $x(t)$, but only on t through the function $t \mapsto v(t)$.

Now, let g be a monotone function satisfying $\|v'(t)\| \le g(\|v(t)\|)$, and let h be the assumed function satifying $h(z) \ge g(z) + \eta$ for all $z \ge \|v(0)\| + \eta$. We define the associated ODE

$$w(0) = \|v(0)\| + \eta, \quad w'(t) = h(w(t)),$$

which we assume is also defined on $[0, u)$. We would like to show that $w(t) > x(t)$ for $t \in [0, u)$. To this end, we would like to verify the conditions of Lemma D.1. First, we have $w(0) = \|v(0)\| + \eta > \Psi_\eta(v(0)) = x(0)$, by above application of the triangle inequality.

For the second condition, we have

$$\begin{aligned}
f(t, x(t)) := \frac{\mathrm{d}}{\mathrm{d}t}(\Psi_\eta(v(t))) &= \frac{\mathrm{d}}{\mathrm{d}t}\mathbb{E}_{Z\sim\mathcal{N}(0,I)}\left[\left\|v + \frac{\eta}{2c_Z}Z\right\|\right] \\
&\le \mathbb{E}_{Z\sim\mathcal{N}(0,1)}\left[\left\|v'(t) + \frac{\eta}{2c_Z}Z\right\|\right] \\
&= \Psi_\eta(v'(t)) < \|v'(t)\| + \eta \\
&\le g(\|v(t)\|) + \eta \quad \text{(since } g \text{ satisfies } \|v'(t)\| \le g(\|v(t)\|)\text{)} \\
&\le g(\Psi_\eta(v(t)) + \eta \quad \text{(since } g \text{ is monotone)} \\
&= g(x(t)) + \eta.
\end{aligned}$$

Now, if $h(z) \ge g(z) + \eta$ for any $z \ge w(0)$, then, we see that, for any $t \in [0, u)$ such that $x(t) \ge w(0)$, we have $f(t, x(t)) \le h(x(t))$. Lemma D.1 therefore implies that $x(t) \le w(t)$ for $t \in [0, u)$. But $x(t) = \Psi_\eta(v(t)) \ge \|v(t)\|$. □

Let us now prove the general guarantee for self-bounding functions.

Proof of Theorem 13. Observe that by the valid-function assumption and the assumption that $F(x(0), y(0))$ has a solution, there exists some interval $[0, u)$ on which a solution $y(t)$ to $F(x(t), y(t)) = 0$ exists. Let u denote the maximal value of $u \le 2$ for which this holds.

First, let us bound $\|y(t)\|$ for $t \in \mathcal{I} := [0, u) \cap [0, 1]$. By assumption, there is a function $h(z) \ge g(z) + \eta$, where $g(z)$ is non-negative and non-decreasing, such that the scalar ODE $w'(t) = h(w(t))$ has a solution on $[0, 1]$ with $w(0) = \|y(0)\| + \eta$. By Lemma D.2, we then that $\|y(t)\| \le w(t)$ on $\mathcal{I}$. Moreover, since $w'(t) \ge 0$ since h is non-negative, we have $\|y(t)\| \le w(t) \le w(1)$ on $\mathcal{I}$.

We conclude by showing that $\mathcal{I} = [0, 1]$. Suppose for the sake of contradiction that $\mathcal{I} \neq [0, 1]$. Then $u \in (0, 1]$. Moreover, by Definition 3.2, $F(x(u), \cdot) = 0$ has no solution, since otherwise, $y(t)$ would be defined on $[0, u + \epsilon)$ for some $\epsilon > 0$, contradicting the maximality of u. Therefore, to contradict our hypothesis $\mathcal{I} \neq [0, 1]$, it suffices to show that $F(x(u), \cdot) = 0$ has a solution. To this end, define

$$\widetilde{y}(s) := \int_0^s y'(t)dt,$$

which is well defined and continuous for $s \in [0, u)$, since $y'(s)$ is continuously differentiable on this interval. Moreover, $\|y'(t)\| \le g(y(t)) \le g(w(t)) \le g(w(1))$ on $[0, u)$ since $y(t) \le w(t) \le w(1)$. Therefore, $y'(t)$ is uniformly bound on $[0, u)$, so that $\widetilde{y}(u) = \lim_{s\to u} \widetilde{y}(s)$ is well-defined at u, and in fact continuous on $[0, u]$.

Since $\widetilde{y}(s)$ is continuous on $[0, u]$, and since $F(\cdot, \cdot)$ and $x(s)$ are continuous, $\lim_{s\to u} F(x(s), \widetilde{y}(s)) = F(x(u), \widetilde{y}(u))$. But by the fundamental theorem of Calculus, we see that $\widetilde{y}(s) = y(s)$ for $s \in [0, u)$, so that $F(x(u), \widetilde{y}(s)) = F(A(s), B(s), y(s)) = 0$ for $s \in [0, u)$. Thus, $\lim_{s\to u} F(A(s), B(s), \widetilde{y}(s)) = 0$, and hence $F(x(u), \widetilde{y}(u)) = 0$. This shows that $F(x(u), \cdot) = 0$ has a solution, as needed. □

We now prove the corollary for the specific function form $g(z) = cz^p$.

Proof of Corollary 3. Fix $\eta > 0$ to be selected later. By assumption, we have

$$\|y'(t)\| \le g(\|y(t)\|), \quad g(z) = cz^p.$$

Moreover, for an $\eta > 0$ to be selected, and for $z \ge \|y(0)\|$, we have

$$g(z) + \eta \le \underbrace{(1 + \frac{\eta}{c\|y(0)\|^p})}_{:=c_\eta} z^p := h(z).$$

Now, consider the ODE

$$w'_\eta(t) = h(w_\eta(t)), \quad w_\eta(0) = \|y(0)\|_{\mathrm{op}} + \eta.$$

Let us show that, for η sufficiently small, this ODE exists on $[0, 1]$. Indeed, the solution to this the ODE is

$$\frac{1}{(p-1)w_\eta^{p-1}(0)} - \frac{1}{(p-1)w_\eta^{p-1}(t)} = c_\eta t.$$

So that a continuously differentiable solution $w_\eta(t)$ exists for $t \in [0, 1]$ as long as

$$c_\eta < \frac{1}{(p-1)w_\eta^{p-1}(0)} = \frac{1}{(p-1)(\|y(0)\| + \eta)^{p-1}}, \tag{D.1}$$

and the solution is given by

$$w_\eta(t) = \left(\frac{1}{(\|y(0)\| + \eta)^{p-1}} - (p-1)c_\eta t\right)^{-1/(p-1)}.$$

In particular, if $c < \frac{1}{(p-1)(\|y(0)\|)^{p-1}}$, then since $\lim_{\eta\to 0} c_\eta = c$, there exists an $\eta_0 > 0$ sufficiently small so that the condition in (D.1) the above display holds for all $\eta \in (0, \eta_0)$. Therefore, by Theorem 13,

$$\max_{t\in[0,1]} \|y(t)\| \le \inf_{\eta\in(0,\eta_0)} w_\eta(t) = \left(\frac{1}{\|y(0)\|^{p-1}} - ct\right)^{-(p-1)} \le \left(\frac{1}{\|y(0)\|^{p-1}} - c(p-1)\right)^{-1/(p-1)}.$$

In particular, when $\alpha = c(p-1)\|y(0)\|^{p-1} < 1$, then

$$\max_{t\in[0,1]} \|y(t)\| \le (1-\alpha)^{-1/(p-1)}\|y(0)\|.$$

Hence, for all $t \in [0, 1]$, we have that $\|y(t)\| \le c(1-\alpha)^{-p/(p-1)}\|y(0)\|$. □

E. Concentration and Estimation Bounds

E.1. Ordinary Least Squares Tools

In what follows, we develop a general toolkit for analyzing the performance of ordinary least squares. First, let $\{\mathbf{z}_t\}_{t\ge 1} \in (\mathbb{R}^d)^{\mathbb{N}}$ and $\{\mathbf{y}_t\}_{t\ge 1} \in (\mathbb{R}^m)^{\mathbb{N}}$ denote sequences of random vectors adapted to a filtration $\{\mathcal{F}_t\}_{t\ge 0}$. We define the empirical covariance matrix $\mathbf{\Lambda}_T := \sum_{t=1}^T \mathbf{z}_t\mathbf{z}_t^\top$. We begin with a standard self-normalized tail bound (cf. (Abbasi-Yadkori et al., 2011)).

Lemma E.1 (Self-Normalized Tail Bound)**.** *Suppose that $\{\boldsymbol{e}_t\}_{t\ge 1} \in \mathbb{R}^{\mathbb{N}}$ is a scalar $\mathcal{F}_t$-adapted sequence such that $\boldsymbol{e}_t \mid \mathcal{F}_{t-1}$ is σ^2 sub-Gaussian. Fix a matrix $V_0 \succeq 0$. Then with probability $1-\delta$,*

$$\left\|\sum_{t=1}^T \mathbf{x}_t\boldsymbol{e}_t\right\|^2_{(V_0+\mathbf{\Lambda}_T)^{-1}} \le 2\sigma^2 \log\left\{\frac{1}{\delta}\det(V_0^{-1/2}(V_0+\mathbf{\Lambda}_T)V_0^{-1/2})\right\}.$$

As a corollary, we have the following Frobenius norm bound for regression.

Lemma E.2 (Frobenius Norm Least Squares, Coarse Bound)**.** *Suppose that the sequence $\boldsymbol{e}_t := \mathbf{y}_t - \Theta_\star\mathbf{z}_t \in \mathbb{R}^m$ is σ^2-sub-Gaussian conditioned on $\mathcal{F}_{t-1}$, and define the least squares estimator $\widehat{\Theta}_T := \left(\sum_{t=1}^T \mathbf{y}_t\mathbf{z}_t\right)(\sum_{t=1}^T \mathbf{z}_t\mathbf{z}_t^\top)^\dagger$. Then,*

$$\mathbb{P}\left[\left\{\|\widehat{\Theta}_T - \Theta_\star\|_{\mathrm{F}}^2 \ge 3m\lambda_{\min}(\mathbf{\Lambda}_T)^{-1}\log\left\{\frac{m\det(3\Lambda_0^{-1/2}(\mathbf{\Lambda}_T)\Lambda_0^{-1/2}}{\delta}\right\}\right\} \cap \{\mathbf{\Lambda}_T \succeq \Lambda_0\}\right] \le \delta,$$

and

$$\mathbb{P}\left[\left\{\|\widehat{\Theta}_T - \Theta_\star\|_{\mathrm{op}}^2 \ge 6\lambda_{\min}(\mathbf{\Lambda}_T)^{-1}(d\log 5 + \log\left\{\frac{\det(3\Lambda_0^{-1/2}(\mathbf{\Lambda}_T)\Lambda_0^{-1/2}}{\delta}\right\}\right\} \cap \{\mathbf{\Lambda}_T \succeq \Lambda_0\}\right] \le \delta. \tag{E.1}$$

Unfortunately, this tail bound will lead to a dimension dependence of $\Omega(d)$, which may be suboptimal if $\mathbf{\Lambda}_T$ has eigenvalues of varying magnitude. Instead, we opt for a related bound that pays for Rayleigh quotients between Λ_0 and $\mathbf{\Lambda}_T$.

Lemma E.3 (Frobenius Norm Least Squares, Refined Bound)**.** *Fix a matrix $\Lambda_0 \succ 0$, and let $v_1, \dots, v_d$ denote its eigenbasis ordered by decreasing eigenvalue. Define the Raleigh quotients $\boldsymbol{\kappa}_j := \frac{v_j^\top \mathbf{\Lambda}_T v_j}{v_j^\top \Lambda_0 v_j}$. Then, in the setting of Lemma E.2, the least squares estimator $\widehat{\Theta}_T$ admits the following bound on its Frobenius error:*

$$\mathbb{P}\left[\left\{\|\widehat{\Theta}_T - \Theta_\star\|_{\mathrm{F}}^2 \ge 3m\sigma^2\sum_{j=1}^d \lambda_j(\Lambda_0^{-1})\boldsymbol{\kappa}_j\log\frac{3\boldsymbol{\kappa}_j}{\delta}\right\} \cap \{\mathbf{\Lambda}_T \succeq \Lambda_0\}\right] \le \delta.$$

Lemma E.4 (Covariance Lower Bound)**.** *Suppose that $\mathbf{z}_t \mid \mathcal{F}_{t-1} \sim \mathcal{N}(\overline{\mathbf{z}}_t, \Sigma_t)$, where $\overline{\mathbf{z}}$ and $\Sigma_t \in \mathbb{R}^d$ are $\mathcal{F}_{t-1}$-measurable and $\Sigma_t \succeq \Sigma \succ 0$. Let $\mathcal{E}$ be any event for which $\overline{\mathbf{\Lambda}}_T := \mathbb{E}[\mathbf{\Lambda}_T\mathbb{I}(\mathcal{E})]$ satisfies $\mathrm{tr}(\overline{\mathbf{\Lambda}}_T) \le TJ$ for some $J \ge 0$. Then, for*

$$T \ge \frac{2000}{9}\left(2d\log\tfrac{100}{3} + d\log\frac{J}{\lambda_{\min}(\Sigma)}\right),$$

it holds that, for $\Lambda_0 := \frac{9T}{1600}\Sigma$

$$\mathbb{P}\left[\left\{\mathbf{\Lambda}_T \not\succeq \frac{9T}{1600}\Sigma\right\} \cap \mathcal{E}\right] \le 2\exp\left(-\tfrac{9}{2000(d+1)}T\right).$$

E.2. Basic Concentration Bounds

Here we state some useful concentration bounds for Gaussian distributions.

Lemma E.5 (Proposition 1.1 in (Hsu et al., 2012))**.** *Let $\mathbf{g} \sim \mathcal{N}(0, I_d)$ be an isotropic Gaussian vector, and let A be a symmetric matrix. Then,*

$$\mathbb{P}\left[\left|\mathbf{g}^\top A\mathbf{g} - \mathrm{tr}(A)\right| > 2t^{1/2}\|A\|_{\mathrm{F}} + 2t\|A\|_{\mathrm{op}}\right] \le 2e^{-t}.$$

By replacing the Frobenius and operator norms in the above inequality with the Hilbert-Schmidt norm, we obtain the following corollary.

Corollary 5. *Let $A \succeq 0$, and let $\mathbf{g} \sim \mathcal{N}(0, I_d)$. Then, with probability $1-\delta$ for any $\delta < 1/e$,*

$$\mathbf{g}^\top A \mathbf{g} \lesssim \mathrm{tr}(A) \log \frac{1}{\delta} = \mathbb{E}[\mathbf{g}^\top A \mathbf{g}_t] \log \frac{1}{\delta}.$$

E.3. Proofs from Appendix E.1

E.3.1. Proof of Lemma E.2

We assume without loss of generality that $\sigma^2 = 1$. Let $\boldsymbol{e}_t = \mathbf{y}_t - \Theta_\star \mathbf{z}_t$. Let $\mathbf{X} \in \mathbb{R}^{Td}$ denote the matrix whose rows are $\boldsymbol{e}_t$, and $\boldsymbol{E}^{(i)} \in \mathbb{R}^T$ denote the vector $(\boldsymbol{e}_{1,i}, \dots, \boldsymbol{e}_{T,i})$, where $\boldsymbol{e}_{t,i}$ is the i-th coordinate of $\boldsymbol{e}_t$. Let $\lambda_{\min} := \lambda_{\min}(\Lambda_0)$. Then,

$$\begin{aligned}
\|\widehat{\Theta}_T - \Theta_\star\|_{\mathrm{F}}^2 &= \sum_{i=1}^m \left\| \boldsymbol{\Lambda}_T^{-1} \mathbf{X}^\top \boldsymbol{E}^{(i)} \right\|_2^2 \\
&\le \lambda_{\min}(\boldsymbol{\Lambda}_T)^{-1} \sum_{i=1}^m \left\| \boldsymbol{\Lambda}_T^{-1/2} \mathbf{X}^\top \boldsymbol{E}^{(i)} \right\|_2^2 \\
&= \sum_{i=1}^m \lambda_{\min}(\boldsymbol{\Lambda}_T)^{-1} \left\| \mathbf{X}^\top \boldsymbol{E}^{(i)} \right\|_{\boldsymbol{\Lambda}_T^{-1}}^2 \\
&\le \frac{3}{2} \sum_{i=1}^m \lambda_{\min}(\boldsymbol{\Lambda}_T)^{-1} \left\| \mathbf{X}^\top \boldsymbol{E}^{(i)} \right\|_{(\boldsymbol{\Lambda}_T + \frac{1}{2}\Lambda_0)^{-1}}^2,
\end{aligned}$$

where the last line holds for $\Lambda_0 \preceq \boldsymbol{\Lambda}_T$. Invoking Lemma E.1, we have that with probability at least $1-\delta$, it holds for any fixed $i \in [m]$ that

$$\left\| \mathbf{X}^\top \boldsymbol{E}^{(i)} \right\|_{(\boldsymbol{\Lambda}_T + \Lambda_0)^{-1}}^2 \le 2 \log \left\{ \frac{1}{\delta} \det((\frac{\Lambda_0}{2})^{-1/2} (\frac{\Lambda_0}{2} + \boldsymbol{\Lambda}_T)(\frac{\Lambda_0}{2})^{-1/2}) \right\}.$$

Since $\Lambda_0 \preceq \boldsymbol{\Lambda}$, we have $\frac{\Lambda_0}{2} + \boldsymbol{\Lambda}_T \le \frac{3}{2}\boldsymbol{\Lambda}$, when the above can be bounded by

$$\left\| \mathbf{X}^\top \boldsymbol{E}^{(i)} \right\|_{(\boldsymbol{\Lambda}_T + \Lambda_0)^{-1}}^2 \le 2 \log \left\{ \frac{1}{\delta} \det(3(\Lambda_0)^{-1/2} \boldsymbol{\Lambda}_T \Lambda_0^{-1/2}) \right\}.$$

Union bounding over $i \in [m]$ and summing the bound concludes.

E.3.2. Proof of Lemma E.3

We assume without loss of generality that $\sigma^2 = 1$. Let $\boldsymbol{e}_t = \mathbf{y}_t - \Theta_\star \mathbf{z}_t$. Let $\mathbf{X} \in \mathbb{R}^{Td}$ denote the matrix whose rows are $\boldsymbol{e}_t$, and $\boldsymbol{E}^{(i)} \in \mathbb{R}^T$ denote the vector $(\boldsymbol{e}_{1,i}, \dots, \boldsymbol{e}_{T,i})$, where $\boldsymbol{e}_{t,i}$ is the i-th coordinate of $\boldsymbol{e}_t$. Let $\lambda_{\min} := \lambda_{\min}(\Lambda_0)$. Moreover, let $v_1, \dots, v_d$ denote an eigenbasis for the matrix Λ_0, which we note is non-random. When $\boldsymbol{\Lambda}_T = \mathbf{X}^\top \mathbf{X} \succeq \Lambda_0$, we can then render

$$\begin{aligned}
\|\widehat{\Theta}_T - \Theta_\star\|_{\mathrm{F}}^2 &= \sum_{i=1}^m \left\| (\mathbf{X}^\top \mathbf{X})^{-1} \mathbf{X}^\top \boldsymbol{E}^{(i)} \right\|_2^2 \\
&\ge \sum_{i=1}^m \left\| \Lambda_0^{-1} \mathbf{X}^\top \boldsymbol{E}^{(i)} \right\|_2^2 \\
&= \sum_{i=1}^m \sum_{j=1}^d \left(v_j^\top \Lambda_0^{-1} \mathbf{X}^\top \boldsymbol{E}^{(i)} \right)^2 \\
&= \sum_{i=1}^m \sum_{j=1}^d \lambda_j(\Lambda_0)^{-2} \left(v_j^\top \mathbf{X}^\top \boldsymbol{E}^{(i)} \right)^2.
\end{aligned}$$

Define the vector $\mathbf{X}_j = \mathbf{X}v_j$. Then, $\left(v_j^\top \mathbf{X}^\top \boldsymbol{E}^{(i)}\right)^2$ can be bounded as as

$$\begin{aligned}\left(v_j^\top \mathbf{X}^\top \boldsymbol{E}^{(i)}\right)^2 &= \|\mathbf{X}_j\|_2^2 (\boldsymbol{E}^{(i)\top} \mathbf{X}_j \underbrace{(\mathbf{X}_j^\top \mathbf{X}_j)^{-1}}_{=\|\mathbf{X}_j\|_2^{-2}} \mathbf{X}_j^\top \boldsymbol{E}^{(i)} \\ &\le \frac{3}{2}\|\mathbf{X}_j\|_2^2 (\mathbf{X}_j^\top \boldsymbol{E}^{(i)})^\top (\frac{1}{2}\lambda_j(\Lambda_0) + \mathbf{X}_j^\top \mathbf{X}_j)^{-1} \mathbf{X}^\top \boldsymbol{E}^{(i)},\end{aligned}$$

where we use the fact that $\Lambda_0 \preceq \boldsymbol{\Lambda}_T = \mathbf{X}^\top \mathbf{X}$. Hence, by the self normalized tail inequality Lemma E.1, it holds with probability $1-\delta$ that

$$\left(v_j^\top \mathbf{X}^\top \boldsymbol{E}^{(i)}\right)^2 \le 3\|\mathbf{X}_j\|_2^2 \log \frac{\frac{1}{2}\lambda_j(\Lambda_0) + \|\mathbf{X}_j\|_2^2}{\frac{1}{2}\lambda_j(\Lambda_0)\delta} \le 3\|\mathbf{X}_j\|_2^2 \log \frac{3\|\mathbf{X}_j\|_2^2}{\lambda_j(\Lambda_0)\delta}.$$

Hence, recalling that $\boldsymbol{\kappa}_j := \frac{v_j^\top \boldsymbol{\Lambda}_T v_j}{v_j^\top \Lambda_0 v_j}$, we conclude that, with probability $1-\delta$,

$$\begin{aligned}\|\widehat{\Theta}_T - \Theta_\star\|_{\mathrm{F}}^2 &\le \sum_{i=1}^m \sum_{j=1}^d 3\lambda_j(\Lambda_0)\|\mathbf{X}_j\|_2^2 \log \frac{3\|\mathbf{X}_j\|_2^2}{\lambda_j(\Lambda_0)\delta} \\ &= 3m \sum_{j=1}^d \lambda_j(\Lambda_0)^{-1} \boldsymbol{\kappa}_j \log \frac{3\kappa_j}{\delta}.\end{aligned}$$

□

E.3.3. Proof of Lemma E.4

By the the Paley-Zygmund inequality (specifically, the variant in Simchowitz et al. (2018, Equation 3.12)), one can easily show that the sequence $(\mathbf{z}_t)$ satisfies the $(1, \Sigma, \frac{3}{10})$-block martingale small ball property (Simchowitz et al., 2018, Definition 2.1). Then, for any matrix $\Lambda_+ \succeq 0$, Simchowitz et al. (2018, Section D.2) (correcting the section for a lost normalization factor of T) shows that

$$\begin{aligned}\mathbb{P}\left[\left\{\boldsymbol{\Lambda}_T \not\succeq \frac{T}{16}(\frac{3}{10})^2\Sigma\right\} \cap \{\boldsymbol{\Lambda}_T \preceq \Lambda_+\}\right] &\le \exp\left(-\frac{1}{10}T(\frac{3}{10})^2 + 2d\log(\frac{100}{3}) + \log\det \Lambda_+(T\Sigma)^{-1}\right) \\ &\le \exp\left(-\frac{9T}{1000} + 2d\log(\frac{100}{3}) + d\log\frac{\|\Lambda_+\|_{\mathrm{op}}}{T\lambda_{\min}(\Sigma)})\right). \qquad \text{(E.2)}\end{aligned}$$

Now, notice that if we select $\Lambda_+ = \frac{\mathrm{tr}(\overline{\boldsymbol{\Lambda}}_T)}{\delta} I$, the bound $\|\Lambda\|_{\mathrm{op}} \le \mathrm{tr}(\Lambda)$ for $\Lambda \succeq 0$ and an application of Markov's inequality show that, $\mathbb{P}[\{\boldsymbol{\Lambda}_T \not\preceq \Lambda_+\} \cap \mathcal{E}] \le \delta$. Hence, we have

$$\begin{aligned}\mathbb{P}\left[\boldsymbol{\Lambda}_T \not\succeq \frac{T}{16}(\frac{3}{10})^2\right] &\le \inf_{\delta>0} \exp\left(-\frac{9T}{1000} + 2d\log(\frac{100}{3}) + d\log\frac{\mathrm{tr}(\overline{\boldsymbol{\Lambda}}_T)}{T\lambda_{\min}(\Sigma)\delta})\right) + \delta \\ &\le \inf_{\delta>0} \delta^{-d} \exp\left(-\frac{9T}{1000} + 2d\log(\frac{100}{3}) + d\log\frac{\mathrm{tr}(\overline{\boldsymbol{\Lambda}}_T)}{T\lambda_{\min}(\Sigma)\delta})\right) + \delta.\end{aligned}$$

Note that balancing $a\delta^{-d} = \delta$ selects $\delta = a^{1/d+1}$, giving that the above is at most

$$2\exp\left(-\frac{1}{d+1}\left(\frac{9T}{1000} - 2d\log(\frac{100}{3}) - d\log\frac{\mathrm{tr}(\overline{\boldsymbol{\Lambda}}_T)}{T\lambda_{\min}(\Sigma)})\right)\right).$$

We conclude by bounding $\mathrm{tr}(\overline{\boldsymbol{\Lambda}}_T) \le JT$ by assumption and applying some elementary algebra.

Part II

Lower Bound

F. Proof of Lower Bound (Theorem 1)

We now prove the main lower bound, Theorem 1. The proof follows the plan outlined in Section 2: We construct a packing of alternative instances, show that low regret on a given instance implies low estimation error, and then deduce from an information-theoretic argument that this implies high regret an alternative instance. All omitted proofs for intermediate lemmas are given in Appendix G. Recall throughout that we assume $\sigma_w^2 = 1$.

F.1. Alternative Instances and Packing Construction

We construct a packing of alternate instances (A_e, B_e) which take the form $(A_\star + K_\star\Delta_e, B_\star + \Delta_e)$, for appropriately chosen perturbations Δ_e described shortly. As discussed in Section 2.1, this packing is chosen because the learner *cannot* distinguish between alternatives if she commits to playing the optimal policy $\mathbf{u}_t = K_\star \mathbf{x}_t$, and must therefore deviate from this policy in order to distinguish between alternatives. We further recall Lemma 2.1, which describes how the optimal controllers from these instances varying with the perturbation Δ.

Lemma 2.1 (Derivative Computation (Abeille & Lazaric (2018), Proposition 2))**.** *Let $(A_\star, B_\star)$ be stabilizable, and recall $A_{\mathrm{cl},\star} := A_\star + B_\star K_\star$. Then,*

$$\frac{d}{dt} K_\infty(A_\star - t\Delta K_\star, B_\star + t\Delta)\big|_{t=0} \\ = -(R_{\mathbf{u}} + B_\star^\top P_\star B_\star)^{-1} \cdot \Delta^\top P_\star A_{\mathrm{cl},\star}.$$

In particular, if A_{cl} is non-degenerate, then to first order, the Frobenius distance between between the optimal controllers for $A_\star, B_\star$ and the alternatives (A_e, B_e) is $\Omega(\|\Delta\|_{\mathrm{F}})$.

To obtain the correct dimension dependence, it is essential that the packing is sufficiently large; a single alternative instance will not suffice. Our goal is to make the packing as large as possible while ensuring that if one can recover the optimal controller for a given instance, they can also recover the perturbation Δ.

Let $n = d_{\mathbf{u}}$, and let $m \le d_{\mathbf{x}}$ be the free parameter from the theorem statement. We construct a collection of instances indexed by sign vectors $e \in \{\text{-}1, 1\}^{[n]\times[m]}$. Let $w_1, \dots, w_n$ denote an eigenbasis basis of $(R_{\mathbf{u}} + B_\star^\top P_\star B_\star)^{-1}$, and $v_1, \dots, v_m$ denote the first m right-singular vectors of $A_{\mathrm{cl},\star}P_\star$. Then for each $e \in \{\text{-}1, 1\}^{[n]\times[m]}$, the corresponding instances is

$$(A_e, B_e) := (A_\star - \Delta_e K_\star, B_\star + \Delta_e), \quad \text{where } \Delta_e = \epsilon_{\mathrm{pack}} \sum_{i=1}^{n}\sum_{j=1}^{m} e_{i,j} w_i v_j^\top. \tag{F.1}$$

It will be convenient to adopt the shorthand $K_e := K_\infty(A_e, B_e)$, $P_e = P_\infty(A_e, B_e)$ and $\mathcal{J}_e = \mathcal{J}^\star_{A_e,B_e}$, and $\Psi_e = \max\{1, \|A_e\|_{\mathrm{op}}, \|B_e\|_{\mathrm{op}}\}$. The following lemma—proven in Appendix G.1—gathers a number of bounds on the error between (A_e, B_e) and $(A_\star, B_\star)$ and their corresponding system parameters. Perhaps most importantly, the lemma shows that to first order, K_e can be approximated using the derivative expression in Lemma 2.1.

Lemma F.1. *There exist universal polynomial functions $\mathfrak{p}_1, \mathfrak{p}_2$ such that, for any $\epsilon_{\mathrm{pack}} \in (0, 1)$, if $\epsilon_{\mathrm{pack}}^2 \le \mathfrak{p}_1(\|P_\star\|_{\mathrm{op}})^{-1}/nm$, the following bounds hold:*

1. **Parameter errror:** $\max\{\|A_e - A_\star\|_{\mathrm{F}}, \|B_e - B_\star\|_{\mathrm{F}}\} \le \sqrt{\|P_\star\|_{\mathrm{op}}}\sqrt{mn}\epsilon_{\mathrm{pack}}$.
2. **Boundedness of value functions:** $\Psi_e \le 2^{1/5}\Psi_\star$ *and* $\|P_e - P_\star\|_{\mathrm{op}} \le 2^{1/5}\|P_\star\|_{\mathrm{op}}$.
3. **Controller error:** $\|K_e - K_\star\|_{\mathrm{F}}^2 \le 2\|P_\star\|_{\mathrm{op}}^3 mn\epsilon_{\mathrm{pack}}^2$.
4. **First-order error:** $\|K_\star + \frac{d}{dt}K_\infty(A_\star - t\Delta K_\star, B_\star + t\Delta_e)\big|_{t=0} - K_e\|_{\mathrm{F}}^2 \le \mathfrak{p}_2(\|P_\star\|_{\mathrm{op}})^2(mn)^2\epsilon_{\mathrm{pack}}^4$.

Notably, item 4 ensures that the first order approximation in Lemma 2.1 is accurate for ϵ_{pack} sufficiently small.

Going forward, we we choose the polynomials in the above lemma $\mathfrak{p}_1, \mathfrak{p}_2$ to satisfy $\mathfrak{p}_1(x), \mathfrak{p}_2(x) \geq x$ (without loss of generality). We use that $\|P_\star\|_{\mathrm{op}} \geq 1$ repeatedly throughout the proof.

Lemma F.2 (Lower bound on $\|P_\star\|_{\mathrm{op}}$)**.** *If $R_{\mathbf{x}} \succeq I$, then $P_\star \succeq I$, and in particular $\|P_\star\|_{\mathrm{op}} \geq 1$.*

Proof. This is Part 4 of a more general statement, Lemma B.5, given in Appendix B. □

Henceforth, we take ϵ_{pack} sufficiently small so as to satisfy the conditions of Lemma F.1.

Assumption 2 (Small ϵ_{pack})**.** $\epsilon_{\mathrm{pack}}^2 \leq \frac{1}{mn}(\mathfrak{p}_1(\|P_\star\|_{\mathrm{op}})^{-1} \wedge \frac{1}{20}\mathfrak{p}_2(\|P_\star\|_{\mathrm{op}})^{-1})$.

F.2. Low Regret Implies Estimation for Controller

We now show that if one can achieve low regret on every instance, then one can estimate the infinite-horizon optimal controller K_e. Suppressing dependence on T, we introduce the shorthand $\mathbb{E}\mathrm{Regret}_e[\pi] := \mathbb{E}\mathrm{Regret}_{A_e,B_e,T}[\pi]$. Going forward, we restrict ourselves to algorithms whose regret is sufficiently small on every packing instance; the trivial case where this is not satisfied is handled at the end of the proof.

Assumption 3 (Uniform Correctness)**.** *For all instances (A_e, B_e), the algorithm π ensures that* $\mathbb{E}\mathrm{Regret}_e[\pi] \leq \frac{T}{6d_{\mathbf{x}}\|P_\star\|_{\mathrm{op}}\Psi_\star^2} - \gamma_{\mathrm{err}}$*, where* $\gamma_{\mathrm{err}} := 6\|P_\star\|_{\mathrm{op}}^3\Psi_\star^2$.

We now define an intermediate term which captures which captures the extent to which the control inputs under instance e deviate from those prescribed by the optimal infinite horizon controller K_e on the first $T/2$ rounds:

$$\text{K-Err}_e[\pi] := \mathbb{E}_{A_e,B_e,\pi}\left[\sum_{t=1}^{T/2}\|\mathbf{u}_t - K_e\mathbf{x}_t\|^2\right].$$

The following lemma, proven in Appendix G.2, shows that regret is lower bounded by $\text{K-Err}_e[\pi]$, and hence any algorithm with low regret under this instance must play controls close to $K_e\mathbf{x}_t$.

Lemma F.3. *There is a universal constant $c_{\mathrm{err}} > 0$ such that if Assumptions 2 and 3 hold and $T \geq c_{\mathrm{err}}\|P_\star\|_{\mathrm{op}}^2\Psi_\star^4$, then*

$$\mathbb{E}\mathrm{Regret}_e[\pi] \geq \frac{1}{2}\text{K-Err}_e[\pi] - \gamma_{\mathrm{err}}.$$

In light of Lemma F.3, the remainder of the proof will focus on lower bounding the deviation K-Err_e. As a first step, the next lemma—proven in Appendix G.3—shows that the optimal controller can be estimated well through least squares whenever K-Err_e is small. More concretely, we consider a least squares estimator which fits a controller using the first half of the algorithm's trajectory. The estimator returns

$$\widehat{K}_{\mathrm{LS}} := \operatorname*{arg\,min}_K \sum_{t=1}^{T/2}\|\mathbf{u}_t - K\mathbf{x}_t\|^2, \tag{F.2}$$

when $\sum_{t=1}^{T/2}\mathbf{x}_t\mathbf{x}_t^\top \succeq c_{\mathrm{min}}T \cdot I$, and returns $\widehat{K}_{\mathrm{LS}} = 0$ otherwise.

Lemma F.4. *If $T \geq c_0 d_{\mathbf{x}}\log(1 + d_{\mathbf{x}}\|P_\star\|_{\mathrm{op}})$ and Assumptions 2 and 3 hold, and if c_{min} is chosen to be an appropriate numerical constant, then the least squares estimator Equation (F.2) guarantees*

$$\text{K-Err}_e[\pi] \geq c_{\mathrm{LS}}T \cdot \mathbb{E}_{A_e,B_e,\pi}\left[\|\widehat{K}_{\mathrm{LS}} - K_e\|_{\mathrm{F}}^2\right] - 1,$$

where c_0 and c_{LS} are universal constants.

Henceforth we take T large enough such that Lemma F.3 and Lemma F.4 apply.

Assumption 4. *We have that $T \geq c_0 d_{\mathbf{x}}\log(1 + d_{\mathbf{x}}\|P_\star\|_{\mathrm{op}}) \vee c_{\mathrm{err}}\|P_\star\|_{\mathrm{op}}^2\Psi_\star^4$.*

F.3. Information-Theoretic Lower Bound for Estimation

We have established that low regret under the instance (A_e, B_e) requires a small deviation from K_e in the sense that $\mathrm{K\text{-}Err}_e[\pi]$ is small, and have shown in turn that any algorithm with low regret yields an estimator for the optimal controller K_e (Lemma F.4). We now provide necessary condition for estimating the optimal controller, which will lead to the final tradeoff between regret on the nominal instance and the alternative instance. This condition is stated in terms of a quantity related to $\mathrm{K\text{-}Err}_e$:

$$\mathrm{K}_\star\text{-}\mathrm{Err}_e[\pi] := \mathbb{E}_{A_e,B_e,\pi}\left[\sum_{t=1}^{T/2} \|\mathbf{u}_t - K_\star \mathbf{x}_t\|^2\right].$$

Both $\mathrm{K}_\star\text{-}\mathrm{Err}_e[\pi]$ and $\mathrm{K\text{-}Err}_e[\pi]$ concern the behavior of the algorithm under instance (A_e, B_e), but former measures deviation from $K_\star$ ("exploration error") while the latter measures deviation from the optimal controller K_e. Our proof essentially argues the following. Let $(\boldsymbol{e}, \boldsymbol{e}')$ be a pair of random indices on the hypercube, where $\boldsymbol{e}$ is uniform on $\{-1, 1\}^{nm}$, and $\boldsymbol{e}'$ is obtained by flipping a single, uniformly selected entry of $\boldsymbol{e}$. Moroever, let $\mathbb{P}_{\boldsymbol{e}}, \mathbb{P}_{\boldsymbol{e}'}$ denote the respective laws for our algorithm under these two instances. We show that—because our instances take the form $(A_\star - \Delta K_\star, B + \Delta)$—$\mathrm{K}_\star\text{-}\mathrm{Err}_e[\pi]$ captures the KL divergence between these two instances:

$$\mathbb{E}_{\boldsymbol{e}}\mathrm{K}_\star\text{-}\mathrm{Err}_{\boldsymbol{e}}[\pi] \approx \mathbb{E}_{\boldsymbol{e},\boldsymbol{e}'}\mathrm{KL}(\mathbb{P}_{\boldsymbol{e}}, \mathbb{P}_{\boldsymbol{e}'}),$$

where the expectations are taken with respect to the distribution over $(\boldsymbol{e}, \boldsymbol{e}')$. In other words, the average error $\mathbb{E}_{\boldsymbol{e}}\mathrm{K}_\star\text{-}\mathrm{Err}_{\boldsymbol{e}}[\pi]$ corresponds to the average one-flip KL-divergence between instances. This captures the fact that the instances can only be distinguished by playing controls which deviate from $\mathbf{u}_t = K_\star \mathbf{x}_t$.

As a consequence, using a technique based on Assouad's lemma (Assouad, 1983) due to (Arias-Castro et al., 2012), we prove an information-theoretic lower bound that shows that any algorithm that can recover the index vector e in Hamming distance on every instance must have $\mathrm{K}_\star\text{-}\mathrm{Err}_e[\pi]$ is large on some instances.

As described above, the following lemma concerns the case where the alternative instance index e is drawn uniformly from the hypercube. Let $\mathbb{E}_{\boldsymbol{e}}$ denote expectation $\boldsymbol{e} \overset{\mathrm{unif}}{\sim} \{\text{-}1, 1\}^{[n]\times[m]}$, and let $d_{\mathrm{ham}}(e, e')$ denote the Hamming distance.

Lemma F.5. *Let $\widehat{e}$ be any estimator depending only on $(\mathbf{x}_1, \dots, \mathbf{x}_{T/2})$ and $(\mathbf{u}_1, \dots, \mathbf{u}_{T/2})$. Then*

$$\textit{either} \quad \mathbb{E}_{\boldsymbol{e}}\mathrm{K}_\star\text{-}\mathrm{Err}_{\boldsymbol{e}}[\pi] \geq \frac{n}{4\epsilon_{\mathrm{pack}}^2}, \quad \textit{or} \quad \mathbb{E}_{\boldsymbol{e}}\mathbb{E}_{A_{\boldsymbol{e}},B_{\boldsymbol{e}},\mathsf{Alg}}\left[d_{\mathrm{ham}}(\boldsymbol{e}, \widehat{e})\right] \geq \frac{nm}{4}.$$

The above lemma is proven in Appendix G.4. To apply this result to the least squares estimator $\widehat{K}_{\mathrm{LS}}$, we prove the following lemma (Appendix G.5), which shows that any estimator $\widehat{K}$ with low Frobenius error relative to K_e can be used to recover e in Hamming distance.

Lemma F.6. *Let $\widehat{e}_{i,j}(\widehat{K}) := \mathrm{sign}\,(w_i^\top(\widehat{K} - K_\star)v_j)$, and define $\nu_k := \|R_{\mathbf{u}} + B_\star^\top P_\star B_\star\|_{\mathrm{op}}/\sigma_k(A_{\mathrm{cl},\star})$. Then under Assumption 2,*

$$d_{\mathrm{ham}}(\widehat{e}_{i,j}(\widehat{K}), e_{i,j}) \leq \frac{2\|\widehat{K} - K_e\|_{\mathrm{F}}}{\nu_m^2\epsilon_{\mathrm{pack}}^2} + \frac{1}{20}nm.$$

Combining Lemmas F.4, F.5, and F.6, we arrive at a dichotomy: either the average exploration error $\mathrm{K}_\star\text{-}\mathrm{Err}_e[\pi]$ is large, or the regret proxy $\mathrm{K\text{-}Err}_e[\pi]$ is large.

Corollary 6. *Let $\boldsymbol{e} \overset{\mathrm{unif}}{\sim} \{\text{-}1, 1\}^{[n]\times[m]}$. Then if Assumptions 2, 3,and 4 hold,*

$$\textit{either} \quad \underbrace{\mathbb{E}_{\boldsymbol{e}}\mathrm{K}_\star\text{-}\mathrm{Err}_{\boldsymbol{e}}[\pi] \geq \frac{n}{4\epsilon_{\mathrm{pack}}^2}}_{\text{(sufficient exploration)}}, \quad \textit{or} \quad \underbrace{\mathbb{E}_{\boldsymbol{e}}\mathrm{K\text{-}Err}_{\boldsymbol{e}}[\pi] \geq \frac{c_{\mathrm{LS}}}{10}Tnm\nu_m^2\epsilon_{\mathrm{pack}}^2 - \gamma_{\mathrm{ls}}}_{\text{(large deviation from optimal)}}. \tag{F.3}$$

Proof. Let $\widehat{e} = \widehat{e}(\widehat{K}_{\mathrm{LS}})$, where $\widehat{e}$ is the estimator from Lemma F.6, and $\widehat{K}_{\mathrm{LS}}$ is as defined in Lemma F.4. Since this estimator only depends on $\mathbf{x}_1, \dots, \mathbf{x}_{T/2}$ and $\mathbf{u}_1, \dots, \mathbf{u}_{T/2}$, we see that if the first condition in Equation (F.3) (sufficient exploration)

fails, then by Lemma F.5, we have $\mathbb{E}_{\boldsymbol{e}}\mathbb{E}_{A_{\boldsymbol{e}},B_{\boldsymbol{e}},\mathsf{Alg}}\left[d_{\mathrm{ham}}(\widehat{e},\boldsymbol{e})\right] \geq \frac{nm}{4} = \frac{nm}{5} + \frac{nm}{20}$. Thus, by Lemma F.6, we have $\frac{2\mathbb{E}_{\boldsymbol{e}}\mathbb{E}_{A_{\boldsymbol{e}},B_{\boldsymbol{e}},\mathsf{Alg}}\|\widehat{K}-K_{\boldsymbol{e}}\|_{\mathrm{F}}}{\nu_m^2\epsilon_{\mathrm{pack}}^2} \geq \frac{nm}{5}$, yielding $\mathbb{E}_{\boldsymbol{e}}\mathbb{E}_{A_{\boldsymbol{e}},B_{\boldsymbol{e}},\mathsf{Alg}}\|\widehat{K}-K_{\boldsymbol{e}}\|_{\mathrm{F}}^2 \geq \frac{1}{10}nm\nu_m^2\epsilon_{\mathrm{pack}}^2$. The bound now follows from Lemma F.4. □

F.4. Completing the Proof

To conclude the proof, we show (Appendix G.6) that $\mathbb{E}_{\boldsymbol{e}}\mathrm{K\text{-}Err}_{\boldsymbol{e}} \approx \mathbb{E}_{\boldsymbol{e}}\mathrm{K}_\star\text{-}\mathrm{Err}_{\boldsymbol{e}}$, so that the final bound follows by setting $\epsilon_{\mathrm{pack}}^2 \approx \sqrt{1/mT}$.

Lemma F.7. *Under Assumptions 2 and 3, we have* $\mathbb{E}_{\boldsymbol{e}}\mathrm{K}_\star\text{-}\mathrm{Err}_{\boldsymbol{e}}[\pi] \leq 2\mathbb{E}_{\boldsymbol{e}}\mathrm{K\text{-}Err}_{\boldsymbol{e}}[\pi] + 4nmT\|P_\star\|_{\mathrm{op}}^4\epsilon_{\mathrm{pack}}^2$.

Combining Lemma F.7 with Corollary 6, we have

$$\max_e \mathrm{K\text{-}Err}_e[\pi] \geq \mathbb{E}_{\boldsymbol{e}}\mathrm{K\text{-}Err}_{\boldsymbol{e}}[\pi] \geq \left(\frac{n}{8\epsilon_{\mathrm{pack}}^2} - 2nmT\|P_\star\|_{\mathrm{op}}^4\epsilon_{\mathrm{pack}}^2\right) \wedge \frac{c_{\mathrm{LS}}}{10}Tnm\nu_m^2\epsilon_{\mathrm{pack}}^2.$$

Setting $\epsilon_{\mathrm{pack}}^2 = \frac{1}{32\|P_\star\|_{\mathrm{op}}^2\sqrt{mT}}$ and substituting in $n = d_{\mathbf{u}}$, we find that as long as T is large enough such that Assumptions 2-4 hold,

$$\begin{aligned}\max_e \mathrm{K\text{-}Err}_e[\pi] &\gtrsim d_{\mathbf{u}}\sqrt{mT}/\|P_\star\|_{\mathrm{op}}^2 \wedge \sqrt{d_{\mathbf{u}}^2mT}\nu_m^2/\|P_\star\|_{\mathrm{op}}^2 - 1\\ &\gtrsim (1\wedge\nu_m^2)\sqrt{md_{\mathbf{u}}^2T}/\|P_\star\|_{\mathrm{op}}^2) - 1.\end{aligned}$$

Thus, by Lemma F.4, we have that for a sufficiently small numerical constant C_{lb} (which we choose to have value at most 1 without loss of generality),

$$\max_e \mathbb{E}\mathrm{Regret}_e[\pi] \geq 2C_{\mathrm{lb}}\frac{(1\wedge\nu_m^2)\sqrt{d_{\mathbf{u}}^2mT}}{\|P_\star\|_{\mathrm{op}}^2} - \frac{1}{2} - \gamma_{\mathrm{err}} \geq 2C_{\mathrm{lb}}\frac{(1\wedge\nu_m^2)\sqrt{d_{\mathbf{u}}^2mT}}{\|P_\star\|_{\mathrm{op}}^2} - 7d_{\mathbf{x}}\|P_\star\|_{\mathrm{op}}^3\Psi_\star^2.$$

It follows that once

$$T \geq c_1\left(\|P_\star\|_{\mathrm{op}}^p(nm \vee \frac{d_{\mathbf{x}}^2\Psi_\star^4(1\vee\nu_m^{-4})}{mn^2} \vee d_{\mathbf{x}}\log(1+d_{\mathbf{x}}\|P_\star\|_{\mathrm{op}})\right), \tag{F.4}$$

where c_1 and p sufficiently numerical constants, Assumptions 2 and 4 are indeed satisfied, so we have

$$\max_e \mathbb{E}\mathrm{Regret}_e[\pi] \geq C_{\mathrm{lb}}\frac{(1\wedge\nu_m^2)\sqrt{d_{\mathbf{u}}^2mT}}{\|P_\star\|_{\mathrm{op}}^2} := \mathcal{R}.$$

We now justify Assumption 3. Suppose the assumption fails, i.e. for some instance e the algorithm has $\mathbb{E}\mathrm{Regret}_e[\pi] \geq \frac{T}{6\Psi_\star^2d_{\mathbf{x}}} - \gamma_{\mathrm{err}}$. Then since $C_{\mathrm{lb}} \leq 1$ and $\|P_\star\|_{\mathrm{op}} \geq 1$, we see that if $\sqrt{T} \geq 12\Psi_\star^2d_{\mathbf{x}}/\sqrt{mn^2}$, then $\mathbb{E}\mathrm{Regret}_e[\pi] \geq 2\mathcal{R} - \gamma_{\mathrm{err}} \geq \mathcal{R}$. By taking c_1 sufficiently large, we see that whenever Equation (F.4) holds, we have $\mathbb{E}\mathrm{Regret}_e[\pi] \geq 2\mathcal{R} - \gamma_{\mathrm{err}} \geq \mathcal{R}$ as desired.

To conclude, we verify that the construction is consistent with the scale parameter ϵ_T from the theorem statement:

$$\|A_e - A_\star\|_{\mathrm{F}}^2 \vee \|B_e - B_\star\|_{\mathrm{F}}^2 \overset{(i)}{\leq} nm\epsilon_{\mathrm{pack}}^2\|P_\star\|_{\mathrm{op}} \overset{(ii)}{\leq} n\sqrt{m/T} \leq \epsilon_T,$$

where (i) follows by Lemma F.1, and (ii) follows by plugging in our choice for ϵ_{pack}. □

G. Additional Proof Details for Lower Bound (Appendix F)

G.1. Proof of Lemma F.1

Observe that

$$\begin{aligned}\max\{\|A_e - A_\star\|_{\mathrm{op}}, \|B_e - B_\star\|_{\mathrm{op}}\} &\leq \max\{\|A_e - A_\star\|_{\mathrm{F}}, \|B_e - B_\star\|_{\mathrm{F}}\}\\ &\leq \max\{\|K_\star\|_{\mathrm{op}}^{1/2}, 1\}\|\Delta\|_{\mathrm{F}}\\ &\leq \sqrt{mn}\epsilon_{\mathrm{pack}}\max\{\|K_\star\|_{\mathrm{op}}^{1/2}, 1\} \leq \sqrt{\|P_\star\|_{\mathrm{op}}\sqrt{mn}}\epsilon_{\mathrm{pack}},\end{aligned}$$

where the last inequality is by Lemma B.8. This prove the first point of the lemma. Next, if $\epsilon_{\mathrm{pack}}^2 \le \frac{1}{\|P_\star\|_{\mathrm{op}}} C_{\mathrm{safe}}^2(A_\star, B_\star)/nm$, then,

$$\max\{\|A_e - A_\star\|_{\mathrm{op}}, \|B_e - B_\star\|_{\mathrm{op}}\} \le \frac{1}{C_{\mathrm{safe}}(A_\star, B_\star)} \le (1 - 2^{1/5}),$$

which implies $\Psi_e \le 2^{1/5} \max\{1, \|A_\star\|_{\mathrm{op}}, \|B_\star\|_{\mathrm{op}}\}$. Moreover Theorem 5 yields

$$\|P_e - P_\star\|_{\mathrm{op}} \le 1.085\|P_\star\|_{\mathrm{op}} \le 2^{1/5}\|P_\star\|_{\mathrm{op}}.$$

For the next point, Theorem 12 bounds the error of the Taylor approximation, and implies that for some polynomial $\mathfrak{p}$,

$$\begin{aligned}
&\| - (R_{\mathbf{u}} + B_\star^\top P_\star B_\star)^{-1} \cdot \Delta^\top P_\star A_{\mathrm{cl},\star} + K_\star - K_e\|_{\mathrm{F}}^2 \\
&\le \mathfrak{p}(\|P_\star\|_{\mathrm{op}}) \max\{\|A_e - A_\star\|_{\mathrm{op}}, \|B_e - B_\star\|_{\mathrm{op}}\}^2 \max\{\|A_e - A_\star\|_{\mathrm{F}}, \|B_e - B_\star\|_{\mathrm{F}}\}^2 \\
&\le \mathfrak{p}(\|P_\star\|_{\mathrm{op}}) \max\{\|A_e - A_\star\|_{\mathrm{F}}, \|B_e - B_\star\|_{\mathrm{F}}\}^4 \\
&\le (nm)^2 \underbrace{\|P_\star\|_{\mathrm{op}} \mathfrak{p}(\|P_\star\|_{\mathrm{op}})}_{:=\mathfrak{p}_2(\mathfrak{p}(\|P_\star\|_{\mathrm{op}}))} \epsilon_{\mathrm{pack}}^4.
\end{aligned}$$

Finally, point 4 follows by bounding

$$\begin{aligned}
\|K_\star - K_e\|_{\mathrm{F}} &\le \| - (R_{\mathbf{u}} + B_\star^\top P_\star B_\star)^{-1} \cdot \Delta^\top P_\star A_{\mathrm{cl},\star} + K_\star - K_e\|_{\mathrm{F}} + \|(R_{\mathbf{u}} + B_\star^\top P_\star B_\star)^{-1} \cdot \Delta^\top P_\star A_{\mathrm{cl},\star}\|_{\mathrm{F}} \\
&\le mn\epsilon_{\mathrm{pack}}^2 \mathfrak{p}_2(\|P_\star\|_{\mathrm{op}}) + \|(R_{\mathbf{u}} + B_\star^\top P_\star B_\star)^{-1} \cdot \Delta^\top P_\star A_{\mathrm{cl},\star}\|_{\mathrm{F}} \\
&\le mn\epsilon_{\mathrm{pack}}^2 \mathfrak{p}_2(\|P_\star\|_{\mathrm{op}}) + \|\Delta\|_{\mathrm{F}} \|P_\star\|_{\mathrm{op}} \|A_{\mathrm{cl},\star}\|_{\mathrm{op}} \\
&\le mn\epsilon_{\mathrm{pack}}^2 \mathfrak{p}_2(\|P_\star\|_{\mathrm{op}}) + \|\Delta\|_{\mathrm{F}} \|P_\star\|_{\mathrm{op}}^{3/2} && \text{(Lemma B.8)} \\
&\le mn\epsilon_{\mathrm{pack}}^2 \mathfrak{p}_2(\|P_\star\|_{\mathrm{op}}) + \epsilon_{\mathrm{pack}} \sqrt{mn} \|P_\star\|_{\mathrm{op}}^{3/2}.
\end{aligned}$$

By taking $\epsilon_{\mathrm{pack}}^2 \le 1/mn\mathrm{poly}(\|P_\star\|_{\mathrm{op}})$, the expression above can be made to be at most $2mn\epsilon_{\mathrm{pack}}^2 \|P_\star\|_{\mathrm{op}}^3$.

G.2. Proof of Lemma F.3

Our strategy is to relate $\mathrm{Regret}_T[\pi; A_e, B_e]$ and $\text{K-Err}_e[\pi]$ to the benchmark inducted by following the true optimal policy $\pi_\star = \pi_\star(A, B)$ which minimizes $\mathbb{E}_{A_e, B_e, \pi} \sum_{t=1}^T c(\mathbf{x}_t, \mathbf{u}_t)]$ over all possible policies π.

To begin, consider an arbitrary stabilizable system (A, B). Let $K_\infty := K_\infty(A, B)$ and $P_\infty = P_\infty(A, B)$. For T fixed and a control policy π, let

$$\text{K-Err}[\pi] := \mathbb{E}_{A,B,\pi} \left[\sum_{t=1}^{T/2} \|\mathbf{u}_t - K_\infty(A, B)\mathbf{x}_t\|_2^2 \right].$$

We define the Q-functions and value functions associated with the LQR problem as follows.

$$\mathbf{Q}_{t;T}(x, u) := \mathbb{E}_{A,B,\pi_\star} \left[\sum_{s=t}^{T} c(\mathbf{x}_s, \mathbf{u}_s) \mid \mathbf{x}_t = x, \mathbf{u}_t = u \right] \qquad \mathbf{V}_{t;T}(x) := \inf_u \mathbf{Q}_{t;T}(x, u),$$

where $\mathbb{E}_{A,B,\pi_\star(A,B)}[\cdot \mid \mathbf{x}_t = x, \mathbf{u}_t = u]$ denotes that the state at time t is $\mathbf{x}_t = x$, inputs is $\mathbf{u}_t = u$, and all future inputs are according to the policy $\pi_\star(A, B)$. Note then that $\pi_\star$ always perscribes the action $\mathbf{u}_t := \arg\min \mathbf{Q}_{t;T}(\mathbf{x}_t, u)$ at time t. We can now characterize the form of the $\mathbf{Q}_{t;T}$ and $\pi_\star$ using the following lemma.

Lemma G.1 (Optimal Finite-Horizon Controllers (Bertsekas, 2005))**.** *Define the elements*

$$\begin{aligned}
P_{t+1} &:= R_{\mathbf{x}} + A^\top P_t A - A^\top P_t B \Sigma_t^{-1} B^\top P_t A, \\
\Sigma_{t+1} &:= R_{\mathbf{u}} + B^\top P_t B, \\
K_{t+1} &:= -\Sigma_{t+1}^{-1} B^\top P_t A,
\end{aligned}$$

with the convention that $P_0 = R_{\mathbf{x}}$. Then, $\mathbf{V}_{t;T}(x) = x^\top P_{T-t} x$, and $\mathbf{Q}_{t;T}(x, u) - \mathbf{V}_{t;T}(x) = \|u - K_{T-t}\, x\|_{\Sigma_{T-t}}^2$, and $(\pi_\star)_{t;T}(\mathbf{x}_t) = K_{T-t}\, \mathbf{x}_t$.

For completeness, we prove the lemma in Section G.2.1. Having defined the true optimal policy, we that the regret is lower bounded as follows.

Lemma G.2. *Fix a system A, B, and suppose that $\mathrm{Regret}_T[\pi; A, B] \le T\,\mathcal{J}^\star_{A,B}$. Then,*

$$\mathrm{Regret}_T[\pi; A, B] \ge \frac{1}{2}\text{K-Err}[\pi] - \mathcal{J}^\star_{A,B}\left(2T\left(\max_{t\ge T/2}\eta_t\right) + \sum_{t\ge 0}\eta_t\right),$$

where we define the errors $\eta_t := \|\Sigma_t^\top (K_\infty - K_t) R_{\mathbf{x}}^{-1/2}\|_2^2$.

Proof of Lemma G.2. We compare both the cost under π and the cost under a comparator to $\mathbf{V}_{1;T}(0)$, the value of the optimal policy starting at $\mathbf{x}_1 = 0$.

$$\begin{aligned}\mathrm{Regret}_T[\pi; A, B] &= \mathbb{E}_{A,B,\pi}\left[\sum_{t=1}^T c_t(\mathbf{x}_t, \mathbf{u}_t)\right] - \mathbf{V}_{1;T}(0) - (T\mathcal{J}^\star_{A,B}K_\infty - \mathbf{V}_{1;T}(0)) \\ &\ge \mathbb{E}_{A,B,\pi}\left[\sum_{t=1}^T c_t(\mathbf{x}_t, \mathbf{u}_t)\right] - \mathbf{V}_{1;T}(0) - (T\mathbb{E}_{A,B,K_\infty}\left[\sum_{t=1}^T c_t(\mathbf{x}_t, \mathbf{u}_t)\right] - \mathbf{V}_{1;T}(0)),\end{aligned}$$

where we use the fact that the infinite horizon regret induced by $K_\star$ on a finite time horizon T is upper bounded by T-times the infinite horizon cost (this can be verified by direct computation).

Next, we use the performance difference lemma, which states that for any policy π',

$$\begin{aligned}\mathbb{E}_{A,B,\pi'}\left[\sum_{t=1}^T c_t(\mathbf{x}_t, \mathbf{u}_t)\right] - \mathbf{V}_{0;T}(0) &= \sum_{t=1}^T \mathbb{E}_{A,B,\pi'}[\mathbf{Q}_{t;T}(\mathbf{x}_t, \mathbf{u}_t) - \mathbf{V}_{t;T}(\mathbf{x}_t)] \\ &= \sum_{t=1}^T \mathbb{E}_{A,B,\pi'}\left[\|\mathbf{u}_t - K_{T-t}\mathbf{x}_t\|^2_{\Sigma_{T-t}}\right]. \qquad \text{(Lemma G.1)}\end{aligned}$$

Therefore,

$$\mathrm{Regret}_T[\pi; A, B] = \underbrace{\sum_{t=1}^T \mathbb{E}_{A,B,\pi}\left[\|\mathbf{u}_t - K_{T-t}\mathbf{x}_t\|^2_{\Sigma_{T-t}}\right]}_{\text{(policy suboptimality)}} - \underbrace{\sum_{t=1}^T \mathbb{E}_{A,B,K_\infty}\left[\|\mathbf{u}_t - K_{T-t}\mathbf{x}_t\|^2_{\Sigma_{T-t}}\right]}_{\text{(comparator suboptimality)}}.$$

Comparator Suboptimality. We begin with two claims.

Claim G.3. $\|(K_\infty - K_{T-t})\mathbf{x}_t\|^2_{\Sigma_{T-t}} \le \eta_{T-t} \cdot \mathbf{x}_t^\top R_{\mathbf{x}}\mathbf{x}$.

Proof. We have that

$$\begin{aligned}\|(K_\infty - K_{T-t})\,\mathbf{x}_t\|^2_{\Sigma_{T-t}} &= \left\|\Sigma_{T-t}^{1/2}(K_\infty - K_{T-t})R_{\mathbf{x}}^{-1/2}R_{\mathbf{x}}^{1/2}\mathbf{x}_t\right\|^2 \\ &\le \left\|\Sigma_{T-t}^{1/2}(K_\infty - K_{T-t})R_{\mathbf{x}}^{-1/2}\right\|_2^2 \left\|R_{\mathbf{x}}^{1/2}\mathbf{x}_t\right\|_2^2 := \eta_{T-t}\cdot\mathbf{x}_t^\top R_{\mathbf{x}}\mathbf{x}_t.\end{aligned}$$

□

Claim G.4. $\mathbb{E}_{A,B,K_\infty}[\sum_{t=1}^T \mathbf{x}_t^\top R_{\mathbf{x}}\mathbf{x}_t] \le T\mathcal{J}^\star_{A,B}$.

Proof. We have

$$\begin{aligned}\mathbb{E}_{A,B,K_\infty}[\sum_{t=1}^{T}\mathbf{x}_t^\top R_{\mathbf{x}}\mathbf{x}_t] &\le \mathbb{E}_{A,B,K_\infty}[\sum_{t=1}^{T}\mathbf{x}_t^\top (R_{\mathbf{x}}+K_\infty^\top R_{\mathbf{u}}K_\infty)\mathbf{x}_t]\\ &= \mathrm{tr}(\sum_{t=1}^{T}\sum_{s=0}^{t}((A+BK_\infty)^s)^\top (R_{\mathbf{x}}+K_\infty^\top R_{\mathbf{u}}K_\infty)((A+BK_\infty)^s))\\ &\le \mathrm{tr}(\sum_{t=1}^{T}\sum_{s=0}^{\infty}((A+BK_\infty)^s)^\top (R_{\mathbf{x}}+K_\infty^\top R_{\mathbf{u}}K_\infty)((A+BK_\infty)^s))\\ &= T\mathrm{tr}(P_\infty(A,B)) = T\mathcal{J}^\star_{A,B},\end{aligned}$$

where the last equalities are by Lemma B.6. □

Invoking these two claims, we have

$$\sum_{t=1}^{T}\mathbb{E}_{A,B,K_\infty}\left[\|\mathbf{u}_t - K_{T-t;T}\mathbf{x}_t\|^2_{\Sigma_{t;T}}\right] \le \sum_{t=1}^{T}\mathbb{E}_{A,B,K_\infty}\eta_t\left[\mathbf{x}_t^\top R_{\mathbf{x}}\mathbf{x}_t\right] \le \mathcal{J}^\star_{A,B}\sum_{t=1}^{T}\eta_{T-t} \le \mathcal{J}^\star_{A,B}\sum_{t=0}^{\infty}\eta_t.$$

Policy Suboptimality. We first make the following claim.

Claim G.5. *Let $(\mathcal{X}, \langle\cdot,\cdot\rangle_{\mathcal{X}})$ denote an inner product space with induced norm $\|\cdot\|_{\mathcal{X}}$. Then for any $x, y \in \mathcal{X}$, $\|x+y\|^2_{\mathcal{X}} \ge \frac{1}{2}\|x\|^2_{\mathcal{X}} - \|y\|^2_{\mathcal{X}}$.*

Proof. We have $\|x+y\|^2_{\mathcal{X}} = \|x\|^2_{\mathcal{X}} + \|y\|^2_{\mathcal{X}} + 2\langle x,y\rangle_{\mathcal{X}}$. Note that, for any $\alpha > 0$, we have $|2\langle x,y\rangle_{\mathcal{X}}| = |2\langle \alpha^{1/2}x, \alpha^{-1/2}y\rangle_{\mathcal{X}}| \le \alpha\|x\|^2_{\mathcal{X}} + \alpha^{-1}\|y\|^2_{\mathcal{X}}$. Setting $\alpha = \frac{1}{2}$, we have $|2\langle x,y\rangle_{\mathcal{X}}| \le \frac{1}{2}\|x\|^2_{\mathcal{X}} + 2\|y\|^2_{\mathcal{X}}$. Hence $\|x+y\|^2_{\mathcal{X}} = \|x\|^2_{\mathcal{X}} + \|y\|^2_{\mathcal{X}} + 2\langle x,y\rangle_{\mathcal{X}} \ge \|x\|^2_{\mathcal{X}} + \|y\|^2_{\mathcal{X}} - (\frac{1}{2}\|x\|^2_{\mathcal{X}} + 2\|y\|^2_{\mathcal{X}}) = \frac{1}{2}\|x\|^2_{\mathcal{X}} - \|y\|^2_{\mathcal{X}}$. □

We can now lower bound

$$\begin{aligned}&\mathbb{E}_{A,B,\pi}\left[\sum_{t=1}^{T}\|\mathbf{u}_t - K_{T-t}\mathbf{x}_t\|^2_{\Sigma_{T-t}}\right]\\ &\ge \mathbb{E}_{A,B,\pi}\left[\sum_{t=1}^{T/2}\|\mathbf{u}_t - K_{T-t}\mathbf{x}_t\|^2_{\Sigma_{T-t}}\right]\\ &\ge \mathbb{E}_{A,B,\pi}\left[\sum_{t=1}^{T/2}\frac{1}{2}\|\mathbf{u}_t - K_\infty\mathbf{x}_t\|^2_{\Sigma_{T-t}} - \|(K_{T-t}-K_\infty)\mathbf{x}_t\|^2_{\Sigma_{T-t}}\right] && \text{(Claim G.5)}\\ &\ge \mathbb{E}_{A,B,\pi}\left[\sum_{t=1}^{T/2}\frac{1}{2}\|\mathbf{u}_t - K_\infty\mathbf{x}_t\|^2_2 - \|(K_{T-t}-K_\infty)\mathbf{x}_t\|^2_{\Sigma_{T-t}}\right]. && (\Sigma_{T-t} \succeq R_{\mathbf{u}} \succeq I)\end{aligned}$$

The expression above is equal to

$$\begin{aligned}&= \frac{1}{2}\text{K-Err}[\pi] - \sum_{t=1}^{T/2}\mathbb{E}_{A,B,\pi}\left[\|(K_{T-t}-K_\infty)\mathbf{x}_t\|^2_{\Sigma_{T-t}}\right]\\ &\ge \frac{1}{2}\text{K-Err}[\pi] - \sum_{t=1}^{T/2}\eta_t\mathbb{E}_{A,B,\pi}\left[\mathbf{x}_t^\top R_{\mathbf{x}}\mathbf{x}_t\right] && \text{(Claim G.3)}\\ &\ge \frac{1}{2}\text{K-Err}[\pi] - 2\max_{t=1}^{T/2}\eta_{T-t}\mathcal{J}^\star_{A,B} && \text{(Claim G.6)}\\ &\ge \frac{1}{2}\text{K-Err}[\pi] - 2\mathcal{J}^\star_{A,B}\max_{t\ge T/2}\eta_t, && \text{(Claim G.6)}\end{aligned}$$

where the last inequality uses the following claim.

Claim G.6. *If* $\mathrm{Regret}_{A,B,T}[\pi] \le T\mathcal{J}^\star_{A,B}$ *(in particular, under Assumption 3), then for any* $\tau \le T$ *and* $Q \preceq R_{\mathbf{x}}$*, we have* $\sum_{t=1}^{\tau} \mathbb{E}_{A,B,\pi}\left[\mathbf{x}_t^\top Q \mathbf{x}_t\right] \le 2T\mathcal{J}^\star_{A,B}$.

The claim is stated for an arbitrary matrix Q so that it can be specialized where necessary.

Proof. We have $\sum_{t=1}^{\tau} \mathbb{E}_{A,B,\pi}\left[\mathbf{x}_t^\top Q \mathbf{x}_t\right] \le \sum_{t=1}^{T} \mathbb{E}_{A,B,\pi}\left[\mathbf{x}_t^\top R_{\mathbf{x}} \mathbf{x}_t\right] \le \sum_{t=1}^{T} \mathbb{E}_{A,B,\pi}\left[\mathbf{x}_t^\top R_{\mathbf{x}} \mathbf{x}_t + \mathbf{u}_t^\top R_{\mathbf{u}} \mathbf{u}\right] = \mathrm{Regret}_{A,B,T}[\pi] + T\mathcal{J}^\star_{A,B} \le 2T\mathcal{J}^\star_{A,B}$. □

Combining the comparator suboptimality and policy suboptimality bounds completes the proof of Lemma G.2. □

The next lemma shows that the error sequence η_t has geometric decrease.

Lemma G.7 (Bound on η_t)**.** *Let* (A, B) *be statibilzable. Then, for* η_t *defined above, we have*

$$\eta_t \le \left(1 + \tfrac{1}{\nu}\right)^{-t}, \quad \textit{where} \quad \nu = 2\|P_\infty(A,B)\|_{\mathrm{op}} \Psi(A,B)^2.$$

Proof of Lemma G.7. Since $R_{\mathbf{x}} \succeq I$,

$$\eta_t \le \frac{1}{\lambda_{\min}(R_{\mathbf{x}})} \left\| \Sigma_{t;T}^{1/2} (K_\infty - K_{T-t}) \right\|_{\mathrm{op}}^2 \le \left\| \Sigma_{t;T}^{1/2} (K_\infty - K_{T-t}) \right\|_{\mathrm{op}}^2.$$

Next, observe that from Lemma G.1 we have

$$\left\| \Sigma_{T-t}^{1/2} (K_\infty - K_{T-t}) \right\|_{\mathrm{op}}^2 = \sup_{\|x\| \le 1} \|(K_\infty - K_{T-t})x\|_{\Sigma_{T-t}}^2 = \sup_{\|x\| \le 1} [\mathbf{Q}_{t;T}(x, K_\infty x) - \mathbf{V}_{T;t}(x)].$$

Since $\mathbf{Q}_{t;T}(x, K_\infty x)$ is a finite horizon Q-function for a stationary process with non-negative rewards, we have $\mathbf{Q}_{t;T}(x,u) \le \mathbf{Q}_\infty(x,u)$. Therefore, the above is

$$\begin{aligned}
&\le \sup_{\|x\| \le 1} [\mathbf{Q}_\infty(x, K_\infty x) - \mathbf{V}_{T;t}(x)] \\
&= \sup_{\|x\| \le 1} [\mathbf{V}_\infty(x) - \mathbf{V}_{T-t}(x)] \\
&= \sup_{\|x\| \le 1} \left[x^\top P_\infty x - x^\top P_{t;T} x\right] \\
&= \|P_\infty - P_{T-t}\|_{\mathrm{op}},
\end{aligned}$$

where we use that P_∞ is the value function for the infinite horizon process (Lincoln & Rantzer (2006, Proposition 1)). By reparametrizing, we have verified that

$$\eta_t \le \|P_\infty - P_t\|_{\mathrm{op}},$$

To conclude, we apply Lemma G.8, which implies that $\|P_\infty - P_{t;T}\|_2 \le \left(1 + \frac{1}{\nu}\right)^{-(T-t+1)}$, where ν is as in the lemma statement:

Lemma G.8 ((Dean et al., 2018), Lemma E.6)**.** *Consider the Riccati recursion*

$$P_{t+1} = R_{\mathbf{x}} + A^\top P_t A - A^\top B P_t (R_{\mathbf{u}} + B^\top P_t B)^{-1} B^\top P_t A,$$

where $R_{\mathbf{x}}$ *and* $R_{\mathbf{u}}$ *are positive definite and* $P_0 = 0$*. When* P_∞ *is the unique solution of the DARE, we have*

$$\|P_t - P_\infty\|_{\mathrm{op}} \le \|P_\infty\|_{\mathrm{op}} \left(1 + \frac{1}{\nu}\right)^{-t}, \tag{G.1}$$

where $\nu = 2\|P_\infty\|_{\mathrm{op}} \cdot \left(\frac{\|A\|_{\mathrm{op}}^2}{\lambda_{\min}(R_{\mathbf{x}})} \vee \frac{\|B\|_{\mathrm{op}}^2}{\lambda_{\min}(R_{\mathbf{u}})} \right)$.[10]

[10] The bound stated in (Dean et al., 2018) is sightly incorrect in that it is missing a factor of $\|P_\infty\|_{\mathrm{op}}$. The reader can verify the correctness of our statement by examining Lincoln & Rantzer (2006, Proposition 1).

We can take $\nu \le 2\|P_\infty\|_{\mathrm{op}} \max\{\|A\|_{\mathrm{op}}^2, \|B\|_{\mathrm{op}}^2\} \le 2\Psi(A,B)\|P_\infty\|_{\mathrm{op}}^2$, as $R_{\mathbf{x}}, R_{\mathbf{u}} \succeq I$. □

We can now conclude the proof of Lemma F.3.

Proof of Lemma F.3. From Lemma G.2, we have the lower bound

$$\mathrm{Regret}_T[\pi; A_e, B_e] \ge \frac{1}{2}\text{K-Err}_e[\pi] - \mathcal{J}_e\left(2T\left(\max_{t \ge T/2}\eta_t\right) + \sum_{t\ge 0}\eta_t\right),$$

Recall $\nu := 2\|P_e\|_{\mathrm{op}}\Psi_e^2$ from Lemma G.7, and that $\eta_t \le \|P_e\|_{\mathrm{op}}(1+\nu^{-1})^t \le \exp(-t/\nu)$. Therefore

$$\sum_{t\ge 0}\eta_t \le 2\|P_e\|_{\mathrm{op}}^2\Psi_e^2,$$

Hence, if $T \ge 2\nu\log(2T)$, we have that

$$2T\left(\max_{t\ge T/2}\eta_t\right) \le \|P_e\|_{\mathrm{op}} \le \|P_e\|_{\mathrm{op}}^2,$$

where we use $P_\infty(\cdot,\cdot) \succeq I$ (Lemma B.5). Hence, for such T,

$$\begin{aligned}
\mathrm{Regret}_T[\pi; A, B] &\ge \frac{1}{2}\text{K-Err}_e[\pi] - 3\|P_e\|_{\mathrm{op}}^2\Psi_e^2 \\
&\ge \frac{1}{2}\text{K-Err}_e[\pi] - J_e 3\|P_e\|_{\mathrm{op}}^2\Psi_e^2 \\
&\ge \frac{1}{2}\text{K-Err}_e[\pi] - d_{\mathbf{x}}\underbrace{3\|P_e\|_{\mathrm{op}}^3\Psi_e^2}_{=\gamma_e}.
\end{aligned}$$

Since $\nu \ge 1$, the condition $T \ge 2\nu\log(2T)$ holds as long as $T \ge c'\nu^2 = c'\|P_e\|_{\mathrm{op}}^2\Psi_e^4 c'$. Reparametrizing in terms of $P_\star, \Psi_\star$ in view of Lemma F.1 concludes the proof. □

G.2.1. Proof of Lemma G.1

We first recall a standard expression for the value function Bertsekas (2005, Section 4.1):

$$\mathbf{V}_t(x) = \|x\|_{P_{T-t}}^2 + \sum_{s=t+1}^{T}\mathrm{tr}(P_{T-s}).$$

To obtain the expression for the $\mathbf{Q}_t$, we have

$$\begin{aligned}
\mathbf{Q}_t(x,u) &= c(x,u) + \mathbb{E}_{\mathbf{w}_t}[\mathbf{V}_{t+1}(Ax+Bu+\mathbf{w}_t)] \\
&= c(x,u) + (Ax+Bu)^\top P_{T-(t+1)}(Ax+Bu) + \mathbb{E}[\mathbf{w}_t^\top P_{T-(t+1)}\mathbf{w}_t] + \sum_{s=t+2}^{T}\mathrm{tr}(P_{T-s}) \\
&= c(x,u) + (Ax+Bu)^\top P_{T-(t+1)}(Ax+Bu) + \sum_{s=t+1}^{T}\mathrm{tr}(P_{T-s}).
\end{aligned}$$

Note that $\mathbf{V}_t(x) := \min_u \mathbf{Q}_t(x,u)$, and $\mathbf{Q}_t(x,u)$ is a quadratic function. We can compute

$$\begin{aligned}
\arg\min_u \mathbf{Q}_t(x,u) &= \arg\min_u c(x,u) + (Ax+Bu)^\top P_{T-(t+1)}(Ax+Bu) + \beta_t \\
&= \arg\min_u x^\top R_{\mathbf{x}}x + u^\top R_{\mathbf{u}}u + (Ax+Bu)^\top P_{T-(t+1)}(Ax+Bu) + \beta_t \\
&= \arg\min_u u^\top(R_{\mathbf{u}} + B^\top P_{t+1;T}B)u + 2u^\top B^\top P_{T-(t+1)}Ax \\
&= -(R_{\mathbf{u}} + B^\top P_{T-(t+1)}B)^{-1}B^\top P_{t+1;T}Ax \\
&= K_{t;T}x.
\end{aligned}$$

Moreover, since $\mathbf{Q}_t(x,u)$ is quadratic in u with quadratic form $\Sigma_{T-t} = (R_{\mathbf{u}} + B^\top P_{T-(t+1)} B)$ and the gradient $\nabla_u \mathbf{Q}_t(x,u)$ vanishes at the minimizer $u = K_{t;T}x$, we have

$$\mathbf{Q}_t(x,u) - \mathbf{V}_t(x) = \mathbf{Q}_t(x,u) - \mathbf{Q}_t(x, K_{t;T}x) = \|u - K_{t;T}x\|^2_{\Sigma_{T-t}}.$$

□

G.3. Proof of Lemma F.4

Again, let us begin proving the lemma for an arbitrary stabilizable (A,B), and then specialize to the packing instances (A_e, B_e). For a fixed policy π, and let all probabilities and expectations be under the law $\mathbb{P}_{\pi;A,B}$. Our strategy follows from (Arias-Castro et al., 2012). Let $K_\infty = K_\infty(A,B)$ and let $\boldsymbol{\delta}_t := \mathbf{u}_t - K_\infty \mathbf{x}_t$, and note that $\text{K-Err}[\pi] = \mathbb{E}_{A,B,\pi}[\sum_{t=1}^{\frac{T}{2}} \|\boldsymbol{\delta}_t\|_2^2]$. Define the covariance matrix

$$\boldsymbol{\Lambda}_{\frac{T}{2}} := \sum_{t=1}^{\frac{T}{2}} \mathbf{x}_t \mathbf{x}_t^\top.$$

For some constant $c > 0$ to be chosen at the end of the proof, consider a 'thresholded" least squares estimator defined as follows:

$$\begin{aligned}
\widehat{K}_{\mathrm{LS}} &:= \mathbb{I}\left\{\boldsymbol{\Lambda}_{\frac{T}{2}} \succeq c\frac{T}{2} I\right\} \cdot \left(\sum_{t=1}^{\frac{T}{2}} \mathbf{u}_t \mathbf{x}_t^\top\right) \boldsymbol{\Lambda}_{\frac{T}{2}}^{-1} \\
&= \mathbb{I}\left\{\boldsymbol{\Lambda}_{\frac{T}{2}} \succeq c\frac{T}{2} I\right\} \left(\sum_{t=1}^{\frac{T}{2}} \boldsymbol{\delta}_t \mathbf{x}_t^\top\right) \boldsymbol{\Lambda}_{\frac{T}{2}}^{-1} + \mathbb{I}\left\{\boldsymbol{\Lambda}_{\frac{T}{2}} \succeq c\frac{T}{2} I\right\} \left(\sum_{t=1}^{\frac{T}{2}} K_\infty \mathbf{x}_t \mathbf{x}_t^\top\right) \boldsymbol{\Lambda}_{\frac{T}{2}}^{-1} \\
&= \mathbb{I}\left\{\boldsymbol{\Lambda}_{\frac{T}{2}} \succeq c\frac{T}{2} I\right\} K_\infty + \mathbb{I}\left\{\boldsymbol{\Lambda}_{\frac{T}{2}} \succeq c\frac{T}{2} I\right\} \left(\sum_{t=1}^{\frac{T}{2}} \boldsymbol{\delta}_t \mathbf{x}_t^\top\right) \boldsymbol{\Lambda}_{\frac{T}{2}}^{-1}.
\end{aligned}$$

Hence, introducing the matrices $\mathbf{X} := \left[\mathbf{x}_1 \mid \mathbf{x}_2 \mid \dots \mathbf{x}_{\frac{T}{2}}\right]$, and $\boldsymbol{\Delta} := \left[\boldsymbol{\delta}_1 \mid \boldsymbol{\delta}_2 \mid \dots \boldsymbol{\delta}_{\frac{T}{2}}\right]$,

$$\begin{aligned}
\mathbb{E}\left[\|\widehat{K}_{\mathrm{LS}} - K_\infty\|_{\mathrm{F}}^2\right] &= \|K_\infty\|_{\mathrm{F}}^2 \mathbb{P}\left[\boldsymbol{\Lambda}_{\frac{T}{2}} \not\succeq c\frac{T}{2} I\right] + \mathbb{E}\left[\mathbb{I}\left\{\boldsymbol{\Lambda}_{\frac{T}{2}} \succeq c\frac{T}{2} I\right\} \left\| \left(\sum_{t=1}^{\frac{T}{2}} \cdot \boldsymbol{\delta}_t \mathbf{x}_t^\top\right) \boldsymbol{\Lambda}_{\frac{T}{2}}^{-1} \right\|_F^2\right] \\
&= \|K_\infty\|_{\mathrm{F}}^2 \mathbb{P}\left[\boldsymbol{\Lambda}_{\frac{T}{2}} \not\succeq c\frac{T}{2} I\right] + \mathbb{E}\left[\mathbb{I}\left\{\mathbf{X}\mathbf{X}^\top \succeq c\frac{T}{2} I\right\} \left\|(\boldsymbol{\Delta}\mathbf{X}^\top)(\mathbf{X}\mathbf{X}^\top)^{-1}\right\|_F^2\right] \\
&= \|K_\infty\|_{\mathrm{F}}^2 \mathbb{P}\left[\boldsymbol{\Lambda}_{\frac{T}{2}} \not\succeq c\frac{T}{2} I\right] + \mathbb{E}\left[\mathbb{I}\left\{\mathbf{X}\mathbf{X}^\top \succeq c\frac{T}{2} I\right\} \left\|\boldsymbol{\Delta}\mathbf{X}^\dagger\right\|_F^2\right] \\
&\le \|K_\infty\|_{\mathrm{F}}^2 \mathbb{P}\left[\boldsymbol{\Lambda}_{\frac{T}{2}} \not\succeq c\frac{T}{2} I\right] + \mathbb{E}\left[\mathbb{I}\left\{\mathbf{X}\mathbf{X}^\top \succeq c\frac{T}{2} I\right\} \|\boldsymbol{\Delta}\|_F^2 \left\|\mathbf{X}^\dagger\right\|_2^2\right] \\
&\le \|K_\infty\|_{\mathrm{F}}^2 \mathbb{P}\left[\boldsymbol{\Lambda}_{\frac{T}{2}} \not\succeq c\frac{T}{2} I\right] + \frac{1}{c\frac{T}{2}} \mathbb{E}\left[\|\boldsymbol{\Delta}\|_F^2\right] \\
&= \|K_\infty\|_{\mathrm{F}}^2 \mathbb{P}\left[\boldsymbol{\Lambda}_{\frac{T}{2}} \not\succeq c\frac{T}{2} I\right] + \frac{2}{cT} \text{K-Err}[\pi] \\
&\le \mathcal{J}_\infty(A,B) \mathbb{P}\left[\boldsymbol{\Lambda}_{\frac{T}{2}} \not\succeq c\frac{T}{2} I\right] + \frac{2}{cT} \text{K-Err}[\pi].
\end{aligned}$$

where the second-to-last line follows since $\|\boldsymbol{\Delta}\|_F^2 = \sum_{t=1}^{\frac{T}{2}} \|\boldsymbol{\delta}_t\|_2^2$, and the last line uses by Lemma B.8 which bounds $K_\infty^\top K_\infty \preceq P_\infty(A,B)$, so that $\|K_\infty\|_{\mathrm{F}}^2 \le \mathrm{tr}(P_\infty(A,B)) = \mathcal{J}^\star_{A,B}$.

In order to conclude the proof, we need to select show that, for some constant c sufficiently small and $\frac{T}{2}$ sufficiently large, $\mathbb{P}\left[\mathbf{\Lambda}_{\frac{T}{2}} \not\succeq c\frac{T}{2}I\right]$ is neglible. Let us now apply Lemma E.4. Let $\mathcal{F}_t$ denote the filtration generated by $(\mathbf{x}_s, \mathbf{u}_s)_{s\le t}$ and $\mathbf{u}_{t+1}$. Observe that $\mathbf{x}_t \mid \mathcal{F}_{t-1} \sim \mathcal{N}(\overline{\mathbf{x}}_t, I)$, where $\overline{\mathbf{x}}_t$ is $\mathcal{F}_{t-1}$-measurable.

Let us now specialize to an instance $(A, B) = (A_e, B_e)$. We can then bound

$$\mathbb{E}[\mathrm{tr}(\sum_{t=1}^{T/2} \mathbf{x}_t\mathbf{x}_t^\top)] \le \mathbb{E}\left[\sum_{t=1}^{T} \mathbf{x}_t^\top R_{\mathbf{x}}\mathbf{x}_t + \mathbf{u}_t^\top R_{\mathbf{u}}\mathbf{u}\right] \le 2\mathcal{J}^\star_{A,B}T,$$

by the Assumption 3 and Claim G.6. Hence, $\mathrm{tr}(\mathbb{E}[\mathbf{\Lambda}_{T/2}]) \le \frac{T}{2}\cdot(4\mathcal{J}^\star_{A_e,B_e})$. Therefore, if

$$T \ge \frac{2000}{9}\left(2d_{\mathbf{x}}\log\tfrac{100}{3} + d_{\mathbf{x}}\log 4\mathcal{J}^\star_{A_e,B_e}\right),$$

we have that

$$\mathbb{P}\left[\mathbf{\Lambda}_{T/2} \not\succeq \frac{9(T/2)}{1600}\right] \le 2\exp\left(-\tfrac{9}{2000(d_{\mathbf{x}}+1)}T\right).$$

In particular, there exists a universal constants c, c_{LS} such that (recalling$\mathcal{J}^\star_{A_e,B_e} = \mathrm{tr}(P_\infty(A_e,B_e))$

$$T \ge cd_{\mathbf{x}}\log(1 + d_{\mathbf{x}}\|P_\infty(A_e,B_e)\|_{\mathrm{op}}) \ge cd_{\mathbf{x}}\log(1+\mathcal{J}^\star_{A_e,B_e}).$$

then for a universal constant c_{LS}, we have

$$\mathbb{E}\left[\|\widehat{K}_{\mathrm{LS}} - K_\infty\|_F^2\right] \le \gamma_{\mathrm{ls}} + \frac{1}{Tc_{\mathrm{LS}}}\text{K-Err}[\pi].$$

Moreover, for (A_e, B_e), we can upper bound $\|P_e\|_{\mathrm{op}} \lesssim \|P_\star\|_{\mathrm{op}}$ (and amend c accordingly) using Lemma F.1, concluding the proof.

G.4. Proof of Lemma F.5

Let $\tau = T/2$. Recall that our packing consists of systems (A_e, B_e) indexed by sign-vectors $e \in \{\text{-}1,1\}^{[n]\times[m]}$:

$$(A_e, B_e) := (A_\star - \Delta_e K_\star, B_\star + \Delta_e), \qquad \text{where } \Delta_e = \epsilon\sum_{i=1}^{n}\sum_{j=1}^{m} e_{i,j}w_i v_j^\top.$$

To keep notation compact, let $q := (i,j)$ denote a stand-in for the double indices (i,j), with $q_1 = i$ and $q_2 = j$. Given an indexing vector $e \in \{\text{-}1,1\}^{[n]\times[m]}$, e^c denote the vector consisting of coordinates of e other than (q_1, q_2). For $a \in \{\text{-}1,1\}$, we set

$$\Delta_{a,q,e_q^c} := \epsilon\left(ae_{q_1,q_2} + \sum_{q'\neq q} e_{q_1',q_2'}w_{q_1'}v_{q_2'}^\top\right).$$

and define A_{a,q,e_q^c}, B_{a,q,e_q^c} analogously, let $\mathbb{P}_{a,q,e_q^c}$ denote the law of the first $\tau = T/2$ rounds under $\mathbb{P}_{A_{a,q,e_q^c},B_{a,q,e_q^c},\mathsf{Alg}}[\cdot]$.

We now consider an indexing vector $\boldsymbol{e}$ drawn uniformly from $\{\text{-}1,1\}^{[n]\times[m]}$. We will then let $\mathbb{P}_{a,q}$ denote the law $\mathbb{P}_{\boldsymbol{e}_q^c}[\mathbb{P}_{a,q,\boldsymbol{e}_q^c}]$, maginalizing over the entries $\boldsymbol{e}_q^c$. Our proof now follows from the argument in (Arias-Castro et al., 2012). We note then that, for any q and any $\widehat{e}$ that depends only on the first $\tau = T/2$ time steps, we can bound

$$\begin{aligned}\mathbb{E}_{\boldsymbol{e}}\mathbb{E}_{A_{\boldsymbol{e}},B_{\boldsymbol{e}}}\left[|\boldsymbol{e}_q - \widehat{e}_q|\right] &= \mathbb{E}_{\boldsymbol{e}_q\overset{\mathrm{unif}}{\sim}\{\text{-}1,1\}}\mathbb{E}_{\boldsymbol{e}_q^c}\mathbb{E}_{A_{\boldsymbol{e}_q,q,\boldsymbol{e}_q^c},B_{\boldsymbol{e}_q,q,\boldsymbol{e}_q^c}}\left[|\boldsymbol{e}_q - \widehat{e}_q|\right] \\ &= \mathbb{E}_{\boldsymbol{e}_q\overset{\mathrm{unif}}{\sim}\{\text{-}1,1\}}\mathbb{E}_{\boldsymbol{e}_q,q}\left[|\boldsymbol{e}_q - \widehat{e}_q|\right] \ge \frac{1}{2}\left(1 - \mathrm{TV}\left(\mathbb{P}_{\text{-}1,q}, \mathbb{P}_{1,q}\right)\right).\end{aligned}$$

Hence, by Cauchy Schwarz,

$$\mathbb{E}\left[\sum_q |\boldsymbol{e}_q - \widehat{e}_q|\right] \geq \sum_q \frac{1}{2}(1 - \mathrm{TV}(\mathbb{P}_{0,q}, \mathbb{P}_{1,q})) \\ \geq \frac{nm}{2}\sum_q \left(1 - \sqrt{\frac{1}{nm}\sum_q \mathrm{TV}(\mathbb{P}_{0,q}, \mathbb{P}_{1,q})^2}\right).$$

Moreover, by Jensen's ineqality followed by a symmetrized Pinsker's ienqualty,

$$\mathrm{TV}(\mathbb{P}_{0,q}, \mathbb{P}_{1,q})^2 \leq \mathbb{E}_{\boldsymbol{e}_q^c}\left[\mathrm{TV}(\mathbb{P}_{-1,q,\boldsymbol{e}_q^c}, \mathbb{P}_{1,q,\boldsymbol{e}_q^c})^2\right] \\ \leq \frac{1}{2}\mathbb{E}_{\boldsymbol{e}_q^c}\left[\frac{\mathrm{KL}(\mathbb{P}_{-1,q,\boldsymbol{e}_q^c}, \mathbb{P}_{1,q,\boldsymbol{e}_q^c})}{2} + \frac{\mathrm{KL}(\mathbb{P}_{1,q,\boldsymbol{e}_q^c}, \mathbb{P}_{\text{-}1,q,\boldsymbol{e}_q^c})}{2}\right].$$

We now require the following lemma to compute the relevant KL-divergences, which we prove below.

Lemma G.9. *Let $\Delta^{(0)}, \Delta^{(1)} \in \mathbb{R}^{d_{\mathbf{x}} \times d_{\mathbf{u}}}$, $\tau \in \mathbb{N}$, and let $A_\star, B_\star$ be the nominal systems defined above. For $i \in \{0, 1\}$, let $\mathbb{P}_i$ denote the law of the first τ iterates under $\mathbb{P}_{A_\star - \Delta^{(i)}K_\star, B_\star + \Delta^{(i)}, \mathsf{Alg}}[\cdot]$. Then,*

$$\mathrm{KL}(\mathbb{P}_0, \mathbb{P}_1) = \frac{1}{2}\mathrm{tr}\left(\left(\Delta^{(0)} - \Delta^{(1)}\right)\Lambda_\tau(\Delta^{(0)})\left(\Delta^{(0)} - \Delta^{(1)}\right)^\top\right).$$

where we have defined the matrix

$$\Lambda_\tau(\Delta) := \mathbb{E}_{A_\star - \Delta K_\star, B_\star + \Delta, \mathsf{Alg}}\left[\sum_{t=1}^{\tau}(\mathbf{u}_t - K_\star \mathbf{x}_t)(\mathbf{u}_t - K_\star \mathbf{x}_t)^\top\right].$$

We can now compute

$$\mathrm{KL}(\mathbb{P}_{\text{-}1,q,\boldsymbol{e}_q^c}, \mathbb{P}_{1,q,\boldsymbol{e}_q^c}) = \mathrm{tr}((\Delta_{1,q,e_q^c} - \Delta_{\text{-}1,q,e_q^c})^\top \Lambda_\tau(\Delta_{-1,q,e_q^c})(\Delta_{1,q,e_q^c} - \Delta_{\text{-}1,q,e_q^c})) \\ = 2\epsilon^2 \mathrm{tr}(u_{q_1} w_{q_2}^\top \Lambda_\tau(\Delta_{\text{-}1,q,e_q^c}) w_{q_2} u_{q_1}^\top) \\ = 2\epsilon^2 w_{q_2}^\top \Lambda_\tau(\Delta_{\text{-}1,q,e_q^c}) w_{q_2}.$$

Hence, we have

$$\mathrm{TV}(\mathbb{P}_{0,q}, \mathbb{P}_{1,q}))^2 \leq \frac{1}{2}\cdot 2\epsilon^2 w_{q_2}^\top \left(\mathbb{E}_{\boldsymbol{e}_q^c}\left[\frac{\Lambda_T(\Delta_{0,q,e_q^c}) + \Lambda_T(\Delta_{1,q,e_q^c})}{2}\right]\right) w_{q_2} \\ = \epsilon^2 w_{q_2}^\top \left(\mathbb{E}_{\boldsymbol{e}}\left[\Lambda_\tau(\Delta_{\boldsymbol{e}})\right]\right) w_{q_2}.$$

Hence, since $\{w_j\}$ for an orthonormal basis,

$$\sum_q \mathrm{TV}(\mathbb{P}_{0,q}, \mathbb{P}_{1,q}))^2 = \sum_q \epsilon w_{q_2}^\top \left(\mathbb{E}_{\boldsymbol{e}}\left[\Lambda_\tau(\Delta_{\boldsymbol{e}})\right]\right) w_{q_2} \\ = \epsilon \sum_{i=1}^m \sum_{j=1}^n w_j^\top \left(\mathbb{E}_{\boldsymbol{e}}\left[\Lambda_\tau(\Delta_{\boldsymbol{e}})\right]\right) w_j \\ \leq m\epsilon^2 \mathrm{tr}(\mathbb{E}_{\boldsymbol{e}}[\Lambda_\tau(\Delta_{\boldsymbol{e}})]).$$

We simplify further as

$$\begin{aligned}\operatorname{tr}(\mathbb{E}_{\boldsymbol{e}}[\Lambda_T(\Delta_{\boldsymbol{e}}]) &= \mathbb{E}_{\boldsymbol{e}}[\operatorname{tr}(\Lambda_T(\Delta_{\boldsymbol{e}})] \\ &= \mathbb{E}_{\boldsymbol{e}}\left[\operatorname{tr}\left(\mathbb{E}_{A_\star-\Delta_{\boldsymbol{e}}K_\star,B_\star+\Delta_{\boldsymbol{e}},\mathsf{Alg}}\left[\sum_{t=1}^{\tau}(\mathbf{u}_t-K_\star\mathbf{x}_t)(\mathbf{u}_t-K_\star\mathbf{x}_t)^\top\right]\right)\right] \\ &= \mathbb{E}_{\boldsymbol{e}}\left[\mathbb{E}_{A_{\boldsymbol{e}},B_{\boldsymbol{e}},\mathsf{Alg}}\left[\sum_{t=1}^{\tau}\operatorname{tr}\left((\mathbf{u}_t-K_\star\mathbf{x}_t)(\mathbf{u}_t-K_\star\mathbf{x}_t)^\top\right)\right]\right] \\ &= \mathbb{E}_{\boldsymbol{e}}\left[\mathbb{E}_{A_{\boldsymbol{e}},B_{\boldsymbol{e}},\mathsf{Alg}}\left[\sum_{t=1}^{\tau}\|\mathbf{u}_t-K_\star\mathbf{x}_t\|^2\right]\right].\end{aligned}$$

Therefore, we conclude

$$\mathbb{E}\left[\sum_q|\boldsymbol{e}_q-\widehat{\boldsymbol{e}}_q|\right]\geq\frac{nm}{2}\sum_q\left(1-\sqrt{\frac{\epsilon^2}{n}\mathbb{E}_{\boldsymbol{e}}\left[\mathbb{E}_{A_{\boldsymbol{e}},B_{\boldsymbol{e}},\mathsf{Alg}}\left[\sum_{t=1}^{\tau}\|\mathbf{u}_t-K_\star\mathbf{x}_t\|^2\right]\right]}\right).$$

This concludes the proof of the proposition. □

G.4.1. Proof of Lemma G.9

By convexity of KL and Jensen's inequality, one can see that the KL under a randomized algorithm $\mathsf{Alg}_{\mathrm{rand}}$ is upper bounded by the largest KL divergence attained by one of the deterministic algorithms corresponding to a realization of its random seeds. Hence, we may assume without loss of generality that Alg is deterministic.

By first conditioning the performance of Alg on its random seed, then integrating the KL combu Note that by we may assume that Alg is deterministic. Let $\mathcal{F}_{t-1}$ denote the filtration generated by $(\mathbf{x}_{1:t-1},\mathbf{u}_{1:t-1})$.

$$\mathrm{KL}(\mathbb{P}_0,\mathbb{P}_1)=\sum_{t=1}^{\tau}\mathbb{E}_{A(\Delta^{(0)}),B(\Delta^{(0)}),\mathsf{Alg}}[\mathrm{KL}(\mathbb{P}_0(\mathbf{x}_t,\mathbf{u}_t\mid\mathcal{F}_{t-1}),\mathbb{P}_{\Delta_2,T}(\mathbf{x}_t,\mathbf{u}_t\mid\mathcal{F}_{t-1})],$$

where $\mathbb{P}_0(\mathbf{x}_t,\mathbf{u}_t\mid\mathcal{F}_{t-1})$ denotes the conditional probability law. Note that $\mathbf{u}_t$ is deterministic given $\mathcal{F}_{t-1}$. Moreover, $\mathbf{x}_t\mid\mathcal{F}_{t-1}$ has the distribution of $\mathcal{N}((A-\Delta^{(i)}K_\star)\mathbf{x}_t+(B+\Delta^{(i)})\mathbf{u}_t,I)$ under $\mathbb{P}_i(\cdot\mid\mathcal{F}_{t-1})$. Hence, using the standard formula for Gaussian KL,

$$\begin{aligned}&\mathrm{KL}(\mathbb{P}_i(\mathbf{x}_t,\mathbf{u}_t\mid\mathcal{F}_{t-1}),\mathbb{P}_i(\mathbf{x}_t,\mathbf{u}_t\mid\mathcal{F}_{t-1})) \\ &=\frac{1}{2}\|(A-\Delta^{(0)}K_\star)\mathbf{x}_t+(B+\Delta^{(0)})-(A-\Delta^{(1)}K_\star)\mathbf{x}_t+(B+\Delta^{(1)}))\|_2^2 \\ &=\frac{1}{2}\|(\Delta^{(0)}-\Delta^0)(\mathbf{u}_t-K_\star\mathbf{x}_t)\|_2^2 \\ &=\frac{1}{2}\mathrm{tr}((\Delta^{(0)}-\Delta^1)^\top(\mathbf{u}_t-K_\star\mathbf{x}_t)(\mathbf{u}_t-K_\star\mathbf{x}_t)^\top(\Delta^{(0)}-\Delta^1)).\end{aligned}$$

The lemma now follows from summing from $t=1,\dots,\tau$ and taking expectations.

G.5. Proof of Lemma F.6

We have $\mathbb{I}(e_{i,j}\neq\widehat{e}_{i,j}(\widehat{K}))=\mathbb{I}(e_{i,j}\widehat{e}_{i,j}(\widehat{K})\neq1)=\mathbb{I}(e_{i,j}w_i^\top(\widehat{K}-K_\star)v_j\leq0)$. Define the Taylor approximation error matrix $\Delta_{2,e}:=K_\star-(R_{\mathbf{u}}+B_\star^\top P_\star B_\star)^{-1}(\Delta_e A_{\mathrm{cl},\star}P_\star)-K_e$. We then have

$$\begin{aligned}e_{i,j}w_i^\top(\widehat{K}-K_\star)v_j &\geq e_{i,j}w_i^\top(K_e-K_\star)v_j-|w_i^\top(\widehat{K}-K_\star)v_j| \\ &\geq e_{i,j}w_i^\top(R_{\mathbf{u}}+B_\star^\top P_\star B_\star)^{-1}(\Delta_e A_{\mathrm{cl},\star}P_\star)v_j-|w_i^\top\Delta_{2,e}v_j|-|w_i^\top(\widehat{K}-K_\star)v_j| \\ &= e_{i,j}\underbrace{\frac{\sigma_j(A_{\mathrm{cl},\star}P_\star)}{\sigma_i(R_{\mathbf{u}}+B_\star^\top P_\star B_\star)}}_{\leq\nu_m}w_i^\top\Delta_e v_j-\left(|w_i^\top\Delta_{2,e}v_j|+|w_i^\top(\widehat{K}-K_\star)v_j|\right),\end{aligned}$$

where we use the definition of w_i and v_j, as less as $\sigma_j(A_{\mathrm{cl},\star}P_\star) \geq \sigma_j(A_{\mathrm{cl},\star})$ since $P_\star \succeq I$. Since $\{w_{i'}\}$ and $\{v_{j'}\}$ form an orthornomal basis, we have $w_i^\top \Delta_e v_j = w_i^\top \sum_{i'=1}^n \sum_{j'=1}^m (\epsilon_{\mathrm{pack}} e_{i',j'} w_{i'} v_{j'}^\top) v_j = \epsilon_{\mathrm{pack}} e_{i,j}$. Hence,

$$e_{i,j} w_i^\top (\widehat{K} - K_\star) v_j \geq \nu_m \epsilon_{\mathrm{pack}} - \left(|w_i^\top \Delta_{2,e} v_j| + |w_i^\top (\widehat{K} - K_\star) v_j| \right).$$

It follows that for any $u \in (0,1)$,

$$\begin{aligned}
\mathbb{I}\left(e_{i,j} w_i^\top (\widehat{K} - K_\star) v_j \leq 0 \right) &\leq \mathbb{I}\left(|w_i^\top (\widehat{K} - K_\star) v_j| \geq \sqrt{u} \nu_m \epsilon_{\mathrm{pack}} \right) + \mathbb{I}\left(|w_i^\top \Delta_{2,e} v_j| \geq (1-\sqrt{u}) \nu_m \epsilon_{\mathrm{pack}} \right) \\
&\leq \frac{|w_i^\top (\widehat{K} - K_\star) v_j|^2}{u \nu_m^2 \epsilon_{\mathrm{pack}}^2} + \frac{|w_i^\top \Delta_{2,e} v_j|}{(1-\sqrt{u})^2 \nu_m^2 \epsilon_{\mathrm{pack}}^2}.
\end{aligned}$$

Since w_i, v_j form an orthonormal basis, we have

$$\begin{aligned}
d_{\mathrm{ham}}(e_{i,j}, \widehat{e}_{i,j}(\widehat{K})) &= \sum_{i=1}^n \sum_{j=1}^m \mathbb{I}\left(e_{i,j} w_i^\top (\widehat{K} - K_\star) v_j \leq 0 \right) \\
&\leq \frac{\|\widehat{K} - K\|_{\mathrm{F}}}{u \nu_m^2 \epsilon_{\mathrm{pack}}^2} + \frac{\|\Delta_{2,e}\|_{\mathrm{F}}^2}{(1-\sqrt{u})^2 \nu_m^2 \epsilon_{\mathrm{pack}}^2}.
\end{aligned}$$

Finally, since $\|\Delta_{2,e}\|_{\mathrm{F}}^2 \leq (nm)^2 \epsilon_{\mathrm{pack}}^4 \mathfrak{p}_2(\|P_\star\|_{\mathrm{op}})^2$ by Lemma F.1, we have that for $u = 1/\sqrt{2}$ and for $\epsilon_{\mathrm{pack}}^2 \leq \frac{1}{20nm} \mathfrak{p}_2(\|P_\star\|_{\mathrm{op}}) \leq \frac{1}{nm}(1-1/\sqrt{2})\sqrt{20}/\mathfrak{p}_2(\|P_\star\|_{\mathrm{op}})$ that the above is at most

$$d_{\mathrm{ham}}(e_{i,j}, \widehat{e}_{i,j}(\widehat{K})) \leq \frac{2\|\widehat{K} - K\|_{\mathrm{F}}}{\nu_m^2 \epsilon_{\mathrm{pack}}^2} - \frac{nm}{20}.$$

□

G.6. Proof of Lemma F.7

Introduce the shorthand $\text{K-Err}_e := \text{K-Err}_{T/2}[\pi; A_e, B_e]$. We then have

$$\begin{aligned}
\mathbb{E}_{\boldsymbol{e}} \text{K}_\star\text{-Err}_{\boldsymbol{e}}[\pi] &= \mathbb{E}_{\boldsymbol{e}} \mathbb{E}_{A_{\boldsymbol{e}}, B_{\boldsymbol{e}}, \pi} \left[\sum_{t=1}^{T/2} \|\mathbf{x}_t - K_\star \mathbf{u}_t\|^2 \right] \\
&\leq 2\mathbb{E}_{\boldsymbol{e}} \left[\mathbb{E}_{A_{\boldsymbol{e}}, B_{\boldsymbol{e}}, \pi} \left[\sum_{t=1}^{T/2} \|\mathbf{x}_t - K_{\boldsymbol{e},\infty} \mathbf{u}_t\|^2 + \|(K_{\boldsymbol{e},\infty} - K_\star)\mathbf{x}_t\|^2 \right] \right] \\
&= 2\mathbb{E}_{\boldsymbol{e}} \text{K-Err}_{\boldsymbol{e}}[\pi] + 2\mathbb{E}_{\boldsymbol{e}} \mathrm{tr}\left((K_{\boldsymbol{e},\infty} - K_\star)^\top \mathbb{E}_{A_{\boldsymbol{e}}, B_{\boldsymbol{e}}, \pi} \left[\sum_{t=1}^{T/2} \mathbf{x}_t \mathbf{x}_t^\top \right] (K_{\boldsymbol{e},\infty} - K_\star) \right) \\
&\leq 2\mathbb{E}_{\boldsymbol{e}} \text{K-Err}_{\boldsymbol{e}}[\pi] + 2\left(\max_e \|K_e - K_\star\|_{\mathrm{F}}^2 \right) \cdot \mathbb{E}_{\boldsymbol{e}} \left\| \mathbb{E}_{A_{\boldsymbol{e}}, B_{\boldsymbol{e}}, \pi} \left[\sum_{t=1}^{T/2} \mathbf{x}_t \mathbf{x}_t^\top \right] \right\|_{\mathrm{op}} \\
&\leq 2\mathbb{E}_{\boldsymbol{e}} \text{K-Err}_{\boldsymbol{e}}[\pi] + 4nm\|P_\star\|_{\mathrm{op}}^3 \epsilon_{\mathrm{pack}}^2 \cdot \mathbb{E}_{\boldsymbol{e}} \left\| \mathbb{E}_{A_{\boldsymbol{e}}, B_{\boldsymbol{e}}, \pi} \left[\sum_{t=1}^{T/2} \mathbf{x}_t \mathbf{x}_t^\top \right] \right\|_{\mathrm{op}},
\end{aligned} \tag{G.2}$$

where the last inequality uses Lemma F.1.

Lemma G.10. *Suppose ϵ is sufficiently small. Given matrices A_e, B_e and optimal controller K_e,*

$$\begin{aligned}
\left\| \mathbb{E}_{A_e, B_e, \pi} \left[\sum_{t=1}^{T/2} \mathbf{x}_t \mathbf{x}_t^\top \right] \right\|_{\mathrm{op}} &\leq (3/2) T \|P_e\|_{\mathrm{op}} + 2J_e \|B_e\|_{\mathrm{op}}^2 . \text{K-Err}_e[\pi] \\
&\leq 2T\|P_\star\|_{\mathrm{op}} + 3\mathcal{J}_\star \Psi_\star^2 . \text{K-Err}_e[\pi],
\end{aligned}$$

where the last inequality uses Lemma F.1.

In particular, note that by Assumption 3 and Lemma F.3, we have the bound

$$\mathbb{E}_{\boldsymbol{e}}[\text{K-Err}_{\boldsymbol{e}}[\pi]] \le 2\mathbb{E}_{\boldsymbol{e}}\mathbb{E}\text{Regret}_{\boldsymbol{e}}[\pi] + \gamma_{\text{err}} \le 2\gamma_{\text{err}}T \le \frac{T}{3d_{\mathbf{x}}\Psi_\star^3}.$$

Then, noting $\mathcal{J}_\star \le d_{\mathbf{x}}\|P_\star\|_{\text{op}}$, we can bound $\mathbb{E}_{\boldsymbol{e}}\left\|\mathbb{E}_{A_{\boldsymbol{e}},B_{\boldsymbol{e}},\pi}\left[\sum_{t=1}^{T/2}\mathbf{x}_t\mathbf{x}_t^\top\right]\right\|_{\text{op}} \le 3T\|P_\star\|_{\text{op}}$. Combining with Eq. (G.2), we have

$$\mathbb{E}_{\boldsymbol{e}}\text{K}_\star\text{-Err}_{\boldsymbol{e}}[\pi] \le 2\mathbb{E}_{\boldsymbol{e}}\text{K-Err}_{\boldsymbol{e}}[\pi] + 4nmT\|P_\star\|_{\text{op}}^4\epsilon_{\text{pack}}^2.$$

□

G.6.1. Proof of Lemma G.10

Let $\mathbf{x}_t$ denote the sequence induced by playing the algorithm π. Recalling the notation $\boldsymbol{\delta}_t = \mathbf{u}_t - B_eK_e\mathbf{x}_t$, we then have

$$\mathbf{x}_t = A_e\mathbf{x}_{t-1} + \mathbf{u}_t + \mathbf{w}_t = (A_e + B_eK_e)\mathbf{x}_{t-1} + B_e\boldsymbol{\delta}_t + \mathbf{w}_t. \tag{G.3}$$

We further define the comparison sequence

$$\overline{\mathbf{x}}_t := (A_e + B_eK_e)\overline{\mathbf{x}}_{t-1} + \mathbf{w}_t \tag{G.4}$$

in which we play the optimal infinite-horizon inputs for (A_e, B_e). As shorthand, let $\mathbb{E}_e[\cdot] := \mathbb{E}_{A_e,B_e,\pi}[\cdot]$, and recall that $\text{K-Err}_e := \text{K-Err}_{T/2}[\pi; A_e, B_e]$. We can bound the desired operator norm of the algorithms

$$\left\|\mathbb{E}_e\left[\sum_{t=1}^{T/2}\mathbf{x}_t\mathbf{x}_t^\top\right]\right\|_{\text{op}} \le \left\|\mathbb{E}_e\left[\sum_{t=1}^{T/2}\overline{\mathbf{x}}_t\overline{\mathbf{x}}_t^\top\right]\right\|_{\text{op}} + \left\|\mathbb{E}_e\left[\sum_{t=1}^{T/2}\overline{\mathbf{x}}_t\overline{\mathbf{x}}_t^\top - \mathbf{x}_t\mathbf{x}_t^\top\right]\right\|_{\text{op}}.$$

It therefore suffices to establish the bounds

$$\left\|\mathbb{E}_e\left[\sum_{t=1}^{T/2}\overline{\mathbf{x}}_t\overline{\mathbf{x}}_t^\top\right]\right\|_{\text{op}} \le T\|P_e\|_{\text{op}} \tag{G.5}$$

$$\left\|\mathbb{E}_e\left[\sum_{t=1}^{T/2}\overline{\mathbf{x}}_t\overline{\mathbf{x}}_t^\top - \mathbf{x}_t\mathbf{x}_t^\top\right]\right\|_{\text{op}} \le \frac{1}{2}T\|P_e\|_{\text{op}} + 2J_e\|B_e\|_{\text{op}}^2\text{K-Err}_e. \tag{G.6}$$

Let us first prove Equation (G.5). We can compute

$$\begin{aligned}\left\|\mathbb{E}_e\left[\sum_{t=1}^{T/2}\overline{\mathbf{x}}_t\overline{\mathbf{x}}_t^\top\right]\right\| &\le \left\|\sum_{t=1}^{T/2}\sum_{s=0}^{-1}(A_e + B_eK_e)^s\left((A_e + B_eK_e)^s\right)^\top\right\|_{\text{op}} \\ &\le \frac{T}{2}\|\mathsf{dlyap}((A_e + B_eK_e)^\top, I)\|_{\text{op}} = \frac{T}{2}\|\mathsf{dlyap}((A_e + B_eK_e), I)\|_{\text{op}} \le \frac{T}{2}\|P_e\|_{\text{op}},\end{aligned}$$

where the last two steps are by Lemma B.5.

Next, we prove Equation (G.6). By Jensen's inequality, the triangle inequality, and Cauchy-Schwarz, we can bound

$$\begin{aligned}\left\|\mathbb{E}_e\left[\sum_{t=1}^{T/2}\overline{\mathbf{x}}_t\overline{\mathbf{x}}_t^\top - \mathbf{x}_t\mathbf{x}_t^\top\right]\right\|_{\text{op}} &\le \mathbb{E}_e\left[\sum_{t=1}^{T}\|\overline{\mathbf{x}}_t\overline{\mathbf{x}}_t^\top - \mathbf{x}_t\mathbf{x}_t^\top\|_{\text{op}}\right] \\ &\le \mathbb{E}_e\left[\sum_{t=1}^{T/2}2\|\overline{\mathbf{x}}_t - \mathbf{x}_t\|\|\overline{\mathbf{x}}_t\| + \|\overline{\mathbf{x}}_t - \mathbf{x}_t\|^2\right] \\ &\le 2\sqrt{\mathbb{E}_e\left[\sum_{t=1}^{T/2}\|\overline{\mathbf{x}}_t\|^2\right]}\sqrt{\mathbb{E}_e\left[\sum_{t=1}^{T/2}\|\overline{\mathbf{x}}_t - \mathbf{x}_t\|^2\right]} + \mathbb{E}_e\left[\sum_{t=1}^{T/2}\|\overline{\mathbf{x}}_t - \mathbf{x}_t\|^2\right].\end{aligned}$$

From Equations (G.3) and (G.4), we have that

$$\begin{aligned}\overline{\mathbf{x}}_t - \mathbf{x}_t &= (A_e + B_e K_e)\overline{\mathbf{x}}_{t-1} + \mathbf{w}_t - ((A_e + B_e K_e)\mathbf{x}_{t-1} + B_e\boldsymbol{\delta}_t + \mathbf{w}_t) \\ &= (A_e + B_e K_e)(\overline{\mathbf{x}}_{t-1} - \mathbf{x}_{t-1}) - B_e\boldsymbol{\delta}_t \\ &= -\sum_{s=1}^{t}(A_e + B_e K_e)^{t-s} B_e \boldsymbol{\delta}_s.\end{aligned}$$

Therefore, we have that

$$\begin{aligned}\sum_{t=1}^{T/2}\|\overline{\mathbf{x}}_t - \mathbf{x}_t\|_2^2 &\le \sum_{t=1}^{T/2}\sum_{s=1}^{t}\|A_e + B_e K_e)^{t-s} B_e\boldsymbol{\delta}_s\|_2^2 \\ &\le \sum_{t=1}^{T/2}\boldsymbol{\delta}_t^\top\left(B_e^\top\sum_{s=0}^{\infty}(A_e + B_e K_e)^{s\top}(A_e + B_e K_e)^s\right)B_e\boldsymbol{\delta}_t^\top \\ &= \sum_{t=1}^{T/2}\boldsymbol{\delta}_t^\top(B_e^\top \mathsf{dlyap}(A_e + B_e K_e, I)B_e\boldsymbol{\delta}_t \\ &\le \|B_e\|_{\mathrm{op}}^2\|P_e\|_{\mathrm{op}}\sum_{t=1}^{T/2}\|\boldsymbol{\delta}_t\|_2^2,\end{aligned}$$

where we use Lemma B.5 in the last inequality. Taking expectations, we have

$$\sum_{t=1}^{T/2}\|\overline{\mathbf{x}}_t - \mathbf{x}_t\|_2^2 \le \|B_e\|_{\mathrm{op}}^2\|P_e\|_{\mathrm{op}}\text{K-Err}_e.$$

This yields

$$\begin{aligned}\left\|\mathbb{E}_e\left[\sum_{t=1}^{T/2}\overline{\mathbf{x}}_t\overline{\mathbf{x}}_t^\top - \mathbf{x}_t\mathbf{x}_t^\top\right]\right\|_{\mathrm{op}} &\le 2\sqrt{\mathbb{E}_e\left[\sum_{t=1}^{T/2}\|\overline{\mathbf{x}}_t\|^2\right]\|B_e\|_{\mathrm{op}}^2\|P_e\|_{\mathrm{op}}\text{K-Err}_e} + \|B_e\|_{\mathrm{op}}^2\|P_e\|_{\mathrm{op}}\text{K-Err}_e \\ &\le 2\sqrt{T/2\cdot J_e\|B_e\|_{\mathrm{op}}^2\|P_e\|_{\mathrm{op}}\text{K-Err}_e} + \|B_e\|_{\mathrm{op}}^2\|P_e\|_{\mathrm{op}}\text{K-Err}_e \\ &= \sqrt{2T\cdot J_e\|B_e\|_{\mathrm{op}}^2\|P_e\|_{\mathrm{op}}\text{K-Err}_e} + \|B_e\|_{\mathrm{op}}^2\|P_e\|_{\mathrm{op}}\text{K-Err}_e,\end{aligned}$$

where use the bound that $\sum_{t=1}^{T/2}\mathbb{E}[\|\overline{\mathbf{x}}_t\|^2] \le (T/2)J_e$ using similar arguments to Lemma G.6. The above can be bounded by

$$\begin{aligned}&\le \frac{1}{2}T\|P_e\|_{\mathrm{op}} + J_e\|B_e\|_{\mathrm{op}}^2\text{K-Err}_e + +\|B_e\|_{\mathrm{op}}^2\|P_e\|_{\mathrm{op}}\text{K-Err}_e \\ &\le \frac{1}{2}T\|P_e\|_{\mathrm{op}} + 2J_e\|B_e\|_{\mathrm{op}}^2\text{K-Err}_e,\end{aligned}$$

since $J_e = \mathrm{tr}(P_e)$.

G.7. Additional Corollaries of Theorem 1

For scaled identity systems, we can remove the requirement that $d_{\mathbf{u}} \le (1 - \Omega(1))d_{\mathbf{u}}$.

Corollary 7 (Scaled Idenity System)**.** *Suppose that $A_\star = (1-\gamma)I$ for $\gamma \in (0,1)$, that $B_\star = U^\top$ where U has orthonormal columns, and $R_{\mathbf{x}}, R_{\mathbf{u}} = I$. Then, for $T \ge c_1\gamma^{-p}\left(d_{\mathbf{u}}d_{\mathbf{x}} \vee \frac{d_{\mathbf{x}}(1-\gamma)^{-4}}{d_{\mathbf{u}}^2}\right) \vee c_1 d_{\mathbf{x}}\log(1 + d_{\mathbf{x}}\gamma^{-1})$,*

$$\mathcal{R}_{A_\star,B_\star,T}\left(\sqrt{d_{\mathbf{u}}^2 d_{\mathbf{x}}/T}\right) \gtrsim \gamma^{-4}(1-\gamma)^2\sqrt{d_{\mathbf{u}}^2 d_{\mathbf{x}}T}.$$

Proof of Corollary 7. By the same arguments as in Corollary 1, we have $\Psi_\star \le 1$ and $\|P_\star\|_{\mathrm{op}} \le \gamma^{-1}$. To conclude, let us lower bound $\sigma_{\min}(A_{\mathrm{cl},\star}) \gtrsim 1-\gamma$, which yields $\nu_{d_{\mathbf{x}}} \gtrsim \frac{1-\gamma}{\gamma}$. Reparameterize $a = (1-\gamma)$. Then for $A_\star = aI$ and $B_\star = U^\top$. Then, we can see that the DARE decouples into scalar along the columns of U and their orthogonal complement. That is, if p, k is the solution to

$$(1-a^2)p = -p^2a^2(1+p)^{-1} + 1, \quad k = -(1+p)^{-1}pa, \tag{G.7}$$

then $A_{\mathrm{cl},\star} = (A_\star - kUU^\top) = (a-k)UU^\top + a(I - UU^\top)$, so that

$$\sigma_{\min}(A_{\mathrm{cl},\star}) \ge \min\{a, a-k\} = \min\{a, \frac{a}{1+p}\} = \frac{a}{1+p}.$$

To conclude, we solve (G.7) and show that p is bounded above by a universal consant. For scalar (a, b), the solution to the DARE is

$$(1-a^2)p + p^2(1-a^2) = -p^2a^2 + (1+p) \quad \text{and thus} \quad -a^2p + p^2 - 1 = 0.$$

The solution p is then given by

$$p = \frac{a^2 \pm \sqrt{a^4+4}}{2} \le \frac{1+\sqrt{5}}{2},$$

as needed. □

Part III

Upper Bound

H. Algorithm and Proof of Upper Bound (Theorem 2)

We now formally describe our main algorithm, Algorithm 1, and prove that it attains the upper bound in Theorem 2. The algorithm is a variant of certainty equivalent control with continual ε-greedy exploration. In line with previous work (Dean et al.; 2018; Cohen et al., 2019; Mania et al., 2019), the algorithm takes as input a controller K_0 that is guaranteed to stabilize the system but otherwise may be arbitrarily suboptimal relative to $K_\star$. The algorithm proceeds in epochs of doubling length. At the beginning of epoch k, the algorithm uses an ordinary least squares subroutine (Algorithm 2) to form an estimate $(\widehat{A}_k, \widehat{B}_k)$ for the system dynamics using data collected in the previous epoch. The algorithm then checks whether the estimate is sufficiently close to $(A_\star, B_\star)$ for the perturbation bounds developed in Theorem 3 take effect; such closeness guarantees that the optimal controller for $(\widehat{A}_k, \widehat{B}_k)$ stabilizes the system and has low regret. If the test fails, the algorithm falls back on the stabilizing controller K_0 for the remainder of the epoch, adding exploratory noise with constant scale. Otherwise, if the test succeeds, the algorithm forms the certainty equivalent controller $\widehat{K}_k := K_\infty(\widehat{A}_k, \widehat{B}_k)$ and plays this for the remainder of the epoch, adding exploratory noise whose scale is carefully chosen to balance exploration and exploitation.

Preliminaries Before beginning the proof, let us first give some additional definitions and notation. We adopt the shorthand $d := d_{\mathbf{x}} + d_{\mathbf{u}}$ and define $k_{\mathrm{fin}} = \lceil \log_2 T \rceil$. For every controllers K for which $(A+BK)$ is stable, we define $P_\infty(K; A, B) := \mathsf{dlyap}(A+BK, R_{\mathbf{x}} + K^\top R_{\mathbf{u}} K$. It is a standard fact (see e.g. Lemma B.6) that such controllers have $\mathcal{J}_{A,B}[K] = \mathrm{tr}(P_\infty(K; A, B))$.

We make will make heavy use of the following system parameters for the controllers used within Algorithm 1:

$$\begin{aligned}
&P_k := P_\infty(\widehat{K}_k; A_\star, B_\star), \quad \mathcal{P}_0 := \frac{\mathcal{J}_0}{d_{\mathbf{x}}} \le \|P_\infty(K_0; A_\star, B_\star)\|_{\mathrm{op}}, \\
&\mathcal{J}_k := \mathcal{J}_{A_\star, B_\star}[\widehat{K}_k], \qquad \mathcal{J}_0 := J_{A_\star, B_\star}[K_0], \\
&A_{\mathrm{cl},k} := A_\star + B_\star \widehat{K}_k, \qquad A_{\mathrm{cl},0} := A_\star + B_\star K_0.
\end{aligned}$$

Algorithm 1: Certainty Equivalent Control with Continual Exploration

Input: Stabilizing controller K_0, confidence parameter δ.
Initialize: $\mathsf{safe} \leftarrow \mathsf{False}$.
Play $\mathbf{u}_1 \sim \mathcal{N}(0, I)$.
for $k = 2, 3, \dots$ **do**
  Let $\tau_k \leftarrow 2^k$.
  /* OLS estimator and covariance matrix using samples $\tau_{k-1}, \dots, \tau_k - 1$. See Algorithm 2. */
  Set $(\widehat{A}_k, \widehat{B}_k, \mathbf{\Lambda}_k) \leftarrow \mathsf{OLS}(k)$.
  if $\mathsf{safe} = \mathsf{False}$ **then**
    $\mathsf{Conf}_k \leftarrow 6\lambda_{\min}(\mathbf{\Lambda}_k)^{-1}\big(d \log 5 + \log\big(4k^2 \det(3(\mathbf{\Lambda}_k)/\delta)\big)$ (infinite if $\mathbf{\Lambda}_k \not\succ 0$).
    if $\mathbf{\Lambda}_k \succeq I$ *and* $1/\mathsf{Conf}_k \geq 9C_{\mathrm{safe}}(\widehat{A}_k, \widehat{B}_k)^2$ **then**
      $\mathsf{safe} \leftarrow \mathsf{True}$, $k_{\mathrm{safe}} \leftarrow k$.
      $\mathcal{B}_{\mathrm{safe}}, \sigma_{\mathrm{in}}^2 \leftarrow \mathsf{SafeRoundInit}(\widehat{A}_k, \widehat{B}_k, \mathsf{Conf}_k, \delta)$. // Confidence ball (Algorithm 3).
    else for $t = \tau_k, \dots, 2\tau_k - 1$, play $\mathbf{u}_t = K_0 \mathbf{x}_t + \mathbf{g}_t$, where $\mathbf{g}_t \sim \mathcal{N}(0, I)$.
  else
    Let $(\widetilde{A}_k, \widetilde{B}_k)$ denote the euclidean projection of $(\widehat{A}_k, \widehat{B}_k)$ onto $\mathcal{B}_{\mathrm{safe}}$.
    $\widehat{K}_k \leftarrow K_\infty(\widetilde{A}_k, \widetilde{B}_k)$.
    for $t = \tau_k, \dots, 2\tau_k - 1$ **do**
      Play $\mathbf{u}_t = \widehat{K}_k \mathbf{x}_t + \sigma_k \mathbf{g}_t$, where $\mathbf{g}_t \sim \mathcal{N}(0, I)$, and $\sigma_k^2 := \min\{1, \sigma_{\mathrm{in}}^2 \tau_k^{-1/2}\}$.

H.1. Proof

We begin the proof by showing that the initial estimation phase (in which the algorithm uses the stabilizing controller K_0) ensures that various regularity conditions hold for the epochs $k \geq k_{\mathrm{safe}}$ (in which the algorithm uses the certainty-equivalent controller). One such regularity condition bounds the $\mathcal{H}_\infty$-norm, which describes the *worst-case* response of a system to perturbations. We recall the following definition from Appendix B:

Definition H.1 ($\mathcal{H}_\infty$ norm)**.** For any stable $\tilde{A} \in \mathbb{R}^{d_{\mathbf{x}}^2}$ (e.g. $A + BK_\infty(A, B)$), we define $\|\tilde{A}\|_{\mathcal{H}_\infty} := \sup_{z \in \mathbb{C}: |z|=1} \|(zI - \tilde{A})^{-1}\|_{\mathrm{op}}$.

The following result is proved in Appendix I.1.

Lemma H.1 (Correctness of Perturbations)**.** *On the event*

$$\mathcal{E}_{\mathrm{safe}} := \left\{ \left\| \left[\widehat{A}_{k_{\mathrm{safe}}} - A_\star \mid \widehat{B}_{k_{\mathrm{safe}}} - B_\star \right] \right\|_{\mathrm{op}}^2 \leq \mathsf{Conf}_{k_{\mathrm{safe}}} \right\},$$

the following bounds hold for all $k \geq k_{\mathrm{safe}}$:

1. $\mathcal{J}_k - \mathcal{J}_\star \leq C_{\mathrm{est}}(A_\star, B_\star) \left(\|\widehat{A} - A_\star\|_{\mathrm{F}}^2 + \|\widehat{B} - B_\star\|_{\mathrm{F}}^2 \right) \lesssim \|P_\star\|_{\mathrm{op}}^8 \left(\|\widehat{A} - A_\star\|_{\mathrm{F}}^2 + \|\widehat{B} - B_\star\|_{\mathrm{F}}^2 \right)$.
2. $\mathcal{J}_k \lesssim \mathcal{J}_\star$, *and* $\|P_k\|_{\mathrm{op}} \lesssim \|P_\star\|_{\mathrm{op}}$.
3. $\|\widehat{K}_k\|_{\mathrm{op}}^2 \leq \frac{21}{20} \|P_\star\|_{\mathrm{op}}$.
4. $\|A_{\mathrm{cl},k}\|_{\mathcal{H}_\infty} \lesssim \|A_{\mathrm{cl},\star}\|_{\mathcal{H}_\infty} \lesssim \|P_\star\|_{\mathrm{op}}^{3/2}$.
5. $A_{\mathrm{cl},k}^\top \mathsf{dlyap}[A_{\mathrm{cl},\star}] A_{\mathrm{cl},k} \preceq (1 - \frac{1}{2}\|\mathsf{dlyap}[A_{\mathrm{cl},\star}]\|_{\mathrm{op}}^{-1})$, *where* $I \preceq \mathsf{dlyap}[A_{\mathrm{cl},\star}] \preceq P_\star$, *where we recall the shorthand* $\mathsf{dlyap}[A_{\mathrm{cl},\star}] = \mathsf{dlyap}(A_{\mathrm{cl},\star}, I)$.
6. $\sigma_{\mathrm{in}}^2 \eqsim \sqrt{d_{\mathbf{x}}} \|P_\star\|_{\mathrm{op}}^{9/2} \Psi_{B_\star} \sqrt{\log \frac{\|P_\star\|_{\mathrm{op}}}{\delta}}$.

We will verify at the end of the proof that $\mathcal{E}_{\mathrm{safe}}$ indeed holds with high probability. We remark that Part 5 of the above lemma plays a role similar to that of "sequential strong stability" in Cohen et al. (2019, Definition 2). By using $\mathsf{dlyap}[A_{\mathrm{cl},\star}]$ as a common Lyapunov function, we remove the complications involved in applying sequential strong stability.

Algorithm 2: $\mathsf{OLS}(k)$

Input: Examples $\mathbf{x}_{\tau_k-1}, \dots, \mathbf{x}_{\tau_k}, \mathbf{u}_{\tau_{k-1}}, \dots, \mathbf{u}_{\tau_k-1}$.

Return $(\widehat{A}_k, \widehat{B}_k, \mathbf{\Lambda}_k)$, where

$$\begin{bmatrix}\widehat{A}_k & \widehat{B}_k\end{bmatrix} \leftarrow \left(\sum_{t=\tau_{k-1}}^{\tau_k-1} \mathbf{x}_{t+1}\begin{bmatrix}\mathbf{x}_t \\ \mathbf{u}_t\end{bmatrix}^\top\right)\mathbf{\Lambda}_k^\dagger, \quad \text{and} \quad \mathbf{\Lambda}_k \leftarrow \sum_{t=\tau_{k-1}}^{\tau_k-1}(\mathbf{x}_t, \mathbf{u}_t)(\mathbf{x}_t, \mathbf{u}_t)^\top.$$

Building on this result, we provide (Appendix I.2) a decomposition of the algorithm's regret which holds conditioned on $\mathcal{E}_{\mathrm{safe}}$.

Lemma H.2 (Regret Decomposition on Safe Rounds)**.** *There is an event $\mathcal{E}_{\mathrm{reg}}$ which holds with probability at least $1-\frac{\delta}{8}$ such that, on $\mathcal{E}_{\mathrm{reg}} \cap \mathcal{E}_{\mathrm{safe}}$, following bound holds*

$$\begin{aligned}\sum_{t=\tau_{k_{\mathrm{safe}}}}^{T}(\mathbf{x}_t^\top R_{\mathbf{x}}\mathbf{x}_t + \mathbf{u}_t^\top R_{\mathbf{u}}\mathbf{u}_t - \mathcal{J}_\star) \lesssim \sum_{k=k_{\mathrm{safe}}}^{k_{\mathrm{fin}}}\tau_k(J_k - \mathcal{J}_\star) + \log T \max_{k\le\log_T}\|\mathbf{x}_{\tau_k}\|_2^2 \\ + \sqrt{T}\left(d_{\mathbf{u}}\sigma_{\mathrm{in}}^2\Psi_{B_\star}^2\|P_\star\|_{\mathrm{op}}) + \sqrt{d\log(1/\delta)}\|P_\star\|_{\mathrm{op}}^4\right) \\ + \log^2\frac{1}{\delta}(1+\sqrt{d}\sigma_{\mathrm{in}}^2\Psi_{B_\star}^2)\|P_\star\|_{\mathrm{op}}^4.\end{aligned} \tag{H.1}$$

Let us unpack the terms that arise in Equation (H.1). The term $\sum_{k=k_{\mathrm{safe}}}^{k_{\mathrm{fin}}}\tau_k(\mathcal{J}_k - \mathcal{J}_\star)$ captures the suboptimality of the controlers $\widehat{K}_k$ selected at each epoch. We bound this term by using that, in light of Lemma H.2, we have $\mathcal{J}_k - \mathcal{J}_\star \propto \|\widehat{A}_k - A_\star\|_{\mathrm{F}}^2 + \|\widehat{B}_k - B_\star\|_{\mathrm{F}}^2$. The next term, $\log T \cdot \max_{k\le\log_T}\|\mathbf{x}_{\tau_k}\|^2$, is of lower order, and roughly captures the penalty for switching controllers at each epoch. The term proportional to $\sqrt{T}$ captures both the penalty for injecting exploratory noise into the system (which incurs a dependence on $d_{\mathbf{u}}$), as well as random fluctuations in the cost coming from the underlying noise process. Finally, the term on the last line of the display is also of lower order ($\mathrm{poly}(\log T)$). To proceed, we show that the norms $\|\mathbf{x}_{\tau_k}\|$ appearing in the second term are well-behaved.

Lemma H.3. *There is an event $\mathcal{E}_{\mathrm{bound}}$ which holds with probability at least $1-\frac{\delta}{8}$ such that, conditioned on $\mathcal{E}_{\mathrm{safe}} \cap \mathcal{E}_{\mathrm{bound}}$,*

$$\|\mathbf{x}_{\tau_k}\| \lesssim \sqrt{\Psi_{B_\star}\mathcal{J}_0\log(1/\delta)}\|P_\star\|_{\mathrm{op}}^{3/2}, \quad \forall k \ge k_{\mathrm{safe}}.$$

This bound is quite crude, but is sufficient for our purposes. We give a concise proof (Appendix I.3) using that in light of Lemma H.1, $\mathsf{dlyap}[A_{\mathrm{cl},\star}]$ acts as a Lyapunov function for all the systems $A_{\mathrm{cl},k}$ conditioned on $\mathcal{E}_{\mathrm{safe}}$.

To bound the error terms $\mathcal{J}_k - \mathcal{J}_\star$ appearing in Equation (H.1) we prove (Appendix I.4) the following bound, which ensures the correctness of the estimators $(\widehat{A}_k, \widehat{B}_k)$ once $k \ge k_{\mathrm{safe}}$.

Lemma H.4. *Define $\tau_{\mathrm{ls}} := d\left(\|P_\star\|_{\mathrm{op}}^3\mathcal{P}_0 + \|P_\star\|_{\mathrm{op}}^{11}\Psi_{B_\star}^6\right)\log\frac{d\|P_\star\|_{\mathrm{op}}}{\delta}$. There is an event $\mathcal{E}_{\mathrm{ls}}$, which holds with probability at least $1-\delta/8$, such that conditioned on $\mathcal{E}_{\mathrm{ls}} \cap \mathcal{E}_{\mathrm{safe}} \cap \mathcal{E}_{\mathrm{bound}}$,*

$$\|\widehat{A}_k - A_\star\|_{\mathrm{F}}^2 + \|\widehat{B}_k - B_\star\|_{\mathrm{F}}^2 \lesssim \frac{d_{\mathbf{u}}d_{\mathbf{x}}}{\sigma_{\mathrm{in}}^2\sqrt{\tau_k}}\|P_\star\|_{\mathrm{op}}^2\log\frac{\|P_\star\|_{\mathrm{op}}}{\delta} + \frac{d_{\mathbf{x}}^2}{\tau_k}\|P_\star\|_{\mathrm{op}}^2\log^2\frac{1}{\delta} \quad \forall k: \ \ \tau_k \ge c\tau_{\mathrm{ls}},$$

where $c > 0$ is a universal constant.

We now put all of these pieces together to prove the final regret bound. Henceforth, we condition on the event $\mathcal{E}_{\mathrm{safe}} \cap \mathcal{E}_{\mathrm{bound}} \cap \mathcal{E}_{\mathrm{reg}} \cap \mathcal{E}_{\mathrm{ls}}$. To begin, consider the sum of errors $\mathcal{J}_k - \mathcal{J}_\star$ in Equation (H.1). We apply Lemma I.2 followed by the

Algorithm 3: SafeRoundInit$(\widehat{A}, \widehat{B}, \mathsf{Conf}, \delta)$

Input: Stabilizable pair $(\widehat{A}, \widehat{B}, \mathsf{Conf}, \delta)$.

Return $\mathcal{B}_{\mathrm{safe}} := \mathcal{B}_{\mathrm{op}}(\mathsf{Conf}; \widehat{A}, \widehat{B})$ and $\sigma_{\mathrm{in}}^2 := \sqrt{d_{\mathbf{x}}}\|P_\infty(\widehat{A}, \widehat{B})\|_{\mathrm{op}}^{9/2} \max\{1, \|\widehat{B}\|_{\mathrm{op}}\}\sqrt{\log \frac{\|P_\infty(\widehat{A},\widehat{B})\|_{\mathrm{op}}}{\delta}}$.

bound on $\mathcal{J}_k \lesssim \mathcal{J}_\star$ from Lemma H.1, which yields

$$\begin{aligned}
&\sum_{k=k_{\mathrm{safe}}}^{k_{\mathrm{fin}}} \tau_k(\mathcal{J}_k - \mathcal{J}_\star) + \log T \max_{k \le \log_T} \|\mathbf{x}_{\tau_k}\|^2 \\
&\le \sum_{k > \tau_{\mathrm{ls}}} \tau_k(\mathcal{J}_k - \mathcal{J}_\star) + \mathcal{J}_\star \sum_{k : \tau_k \le c\tau_{\mathrm{ls}}} \tau_k + \sqrt{\Psi_{B_\star}\mathcal{J}_0 \log(1/\delta)}\|P_\star\|_{\mathrm{op}}^{3/2} \log T \\
&\lesssim \left\{ \sum_{k > \tau_{\mathrm{ls}}} \tau_k(\mathcal{J}_k - \mathcal{J}_\star) \right\} + \tau_{\mathrm{ls}}\mathcal{J}_\star + \sqrt{\Psi_{B_\star}\mathcal{J}_0 \log(1/\delta)}\|P_\star\|_{\mathrm{op}}^{3/2} \log T \\
&\lesssim \left\{ \sum_{k > \tau_{\mathrm{ls}}} \tau_k(\mathcal{J}_k - \mathcal{J}_\star) \right\} + d_{\mathbf{x}}\|P_\star\|_{\mathrm{op}}\tau_{\mathrm{ls}} \log \frac{1}{\delta},
\end{aligned}$$

where the last line uses that $\delta \le 1/T$ to combine the lower-order terms in the line preceding it. Next, using the bound $\mathcal{J}_k - \mathcal{J}_\star \lesssim \|P_\star\|_{\mathrm{op}}^8 \left(\|A_\star - \widehat{A}\|_{\mathrm{F}}^2 + \|B_\star - \widehat{B}\|_{\mathrm{F}}^2 \right)$ from Lemma H.1 followed by the error bound in Lemma H.4, we have

$$\begin{aligned}
\sum_{k > \tau_{\mathrm{ls}}} \tau_k(\mathcal{J}_k - \mathcal{J}_\star) &\le \|P_\star\|_{\mathrm{op}}^8 \sum_{k > \tau_{\mathrm{ls}}} \frac{d_{\mathbf{u}} d_{\mathbf{x}}}{\sigma_{\mathrm{in}}^2 \sqrt{\tau_k}} \|P_\star\|_{\mathrm{op}}^2 \log \frac{\|P_\star\|_{\mathrm{op}}}{\delta} + \frac{d_{\mathbf{x}}^2}{\tau_k}\|P_\star\|_{\mathrm{op}}^2 \log^2 \frac{1}{\delta} \\
&\lesssim \frac{d_{\mathbf{u}} d_{\mathbf{x}} \sqrt{T}}{\sigma_{\mathrm{in}}^2}\|P_\star\|_{\mathrm{op}}^{10} \log \frac{\|P_\star\|_{\mathrm{op}}}{\delta} + \underbrace{d_{\mathbf{x}}^2 \|P_\star\|_{\mathrm{op}}^{10} \log^3 \frac{1}{\delta}}_{\lesssim d_{\mathbf{x}}\|P_\star\|_{\mathrm{op}}\tau_{\mathrm{ls}} \log \frac{1}{\delta}},
\end{aligned}$$

where again we use $\log T \lesssim \log(1/\delta)$. Combining the computations so far shows that

$$\sum_{k=k_{\mathrm{safe}}}^{k_{\mathrm{fin}}} \tau_k(\mathcal{J}_k - \mathcal{J}_\star) + \log T \max_{k \le \log_T} \|\mathbf{x}_{\tau_k}\|^2 \le \frac{d_{\mathbf{u}} d_{\mathbf{x}} \sqrt{T}}{\sigma_{\mathrm{in}}^2}\|P_\star\|_{\mathrm{op}}^{10} \log \frac{\|P_\star\|_{\mathrm{op}}}{\delta} + d_{\mathbf{x}}\|P_\star\|_{\mathrm{op}}\tau_{\mathrm{ls}} \log \frac{1}{\delta}.$$

Hence, on $\mathcal{E}_{\mathrm{safe}} \cap \mathcal{E}_{\mathrm{bound}} \cap \mathcal{E}_{\mathrm{reg}} \cap \mathcal{E}_{\mathrm{ls}}$m the regret in the episodes $k \ge k_{\mathrm{safe}}$ decomposes into a component scaling with $\sqrt{T}$ and a component scaling with $\log T$:

$$\begin{aligned}
&\sum_{t=\tau_{k_{\mathrm{safe}}}}^{T} (\mathbf{x}_t^\top R_{\mathbf{x}} \mathbf{x}_t + \mathbf{u}_t^\top R_{\mathbf{u}} \mathbf{u}_t - \mathcal{J}_\star) \\
&\lesssim \underbrace{\sqrt{T}\left(d_{\mathbf{u}}\sigma_{\mathrm{in}}^2 \Psi_{B_\star}^2 \|P_\star\|_{\mathrm{op}} + \sqrt{d\log(1/\delta)}\|P_\star\|_{\mathrm{op}}^4 + \frac{d_{\mathbf{u}} d_{\mathbf{x}}}{\sigma_{\mathrm{in}}^2}\|P_\star\|_{\mathrm{op}}^{10} \log \frac{\|P_\star\|_{\mathrm{op}}}{\delta} \right)}_{\sqrt{T}\text{-component)}} \\
&\quad + \underbrace{(1 + \sqrt{d}\sigma_{\mathrm{in}}^2 \Psi_{B_\star}^2)\|P_\star\|_{\mathrm{op}}^4 \log^2 \frac{1}{\delta} + d_{\mathbf{x}}\tau_{\mathrm{ls}} \log \frac{1}{\delta}}_{((\mathrm{poly}(\log T)\text{-component)}}.
\end{aligned}$$

Using that $\sigma_{\mathrm{in}}^2 \eqsim \sqrt{d_{\mathbf{x}}}\|P_\star\|_{\mathrm{op}}^{9/2}\Psi_{B_\star}\sqrt{\log \frac{\|P_\star\|_{\mathrm{op}}}{\delta}}$ (Lemma H.1) and recalling that $d = d_{\mathbf{x}} + d_{\mathbf{u}}$, we upper bound these terms as

$$\begin{aligned}
(\sqrt{T}\text{-component}) &\lesssim \sqrt{T d_{\mathbf{u}}^2 d_{\mathbf{x}} \Psi_{B_\star}^2 \|P_\star\|_{\mathrm{op}}^{11} \log \frac{\|P_\star\|_{\mathrm{op}}}{\delta}}, \\
(\mathrm{poly}(\log T)\text{-component}) &\lesssim d_{\mathbf{x}}\tau_{\mathrm{ls}} \log \frac{1}{\delta}.
\end{aligned}$$

We conclude that conditioned on $\mathcal{E}_{\mathrm{safe}} \cap \mathcal{E}_{\mathrm{bound}} \cap \mathcal{E}_{\mathrm{reg}} \cap \mathcal{E}_{\mathrm{ls}}$,

$$\sum_{t=\tau_{k_{\mathrm{safe}}}}^{T} (\mathbf{x}_t^\top R_{\mathbf{x}} \mathbf{x}_t + \mathbf{u}_t^\top R_{\mathbf{u}} \mathbf{u}_t - \mathcal{J}_\star) \lesssim \sqrt{T d_{\mathbf{u}}^2 d_{\mathbf{x}} \Psi_{B_\star}^2 \|P_\star\|_{\mathrm{op}}^{11} \log \frac{\|P_\star\|_{\mathrm{op}}}{\delta}} + d_{\mathbf{x}} \tau_{\mathrm{ls}} \log \frac{1}{\delta}. \tag{H.2}$$

To finish the proof, we (a) verify that $\mathcal{E}_{\mathrm{safe}}$ indeed holds with high probability, and (b) bound the regret contribution of the initial rounds (proof given in Appendix I.5).

Lemma H.5. *The event $\mathcal{E}_{\mathrm{safe}}$ holds with probability $1 - \frac{\delta}{2}$, and the following event $\mathcal{E}_{\mathrm{reg,init}}$ holds with probability $1 - \frac{\delta}{8}$:*

$$\sum_{t=1}^{\tau_{k_{\mathrm{safe}}}-1} \mathbf{x}_{t,0}^\top R_{\mathbf{x}} \mathbf{x}_{t,0} + \mathbf{u}_{t,0}^\top R_{\mathbf{u}} \mathbf{u}_{t,0} \lesssim \mathcal{P}_0 d^2 \Psi_{B_\star}^2 \|P_\star\|_{\mathrm{op}}^{10} (1 + \|K_0\|_{\mathrm{op}}^2) \log \frac{d \Psi_{B_\star}^2 \mathcal{P}_0}{\delta} \log \tfrac{1}{\delta}.$$

Thus, $\mathcal{E}_{\mathrm{safe}} \cap \mathcal{E}_{\mathrm{bound}} \cap \mathcal{E}_{\mathrm{reg}} \cap \mathcal{E}_{\mathrm{ls}} \mathcal{E}_{\mathrm{reg,init}}$ holds with total probability at least $1 - \delta$, and conditioned on this event Lemma H.5 and Equation (H.2) imply

$$\begin{aligned}
\mathrm{Regret}_T[\mathsf{Alg}; A_\star, B_\star] &= \sum_{t=1}^{T} (\mathbf{x}_t^\top R_{\mathbf{x}} \mathbf{x}_t + \mathbf{u}_t^\top R_{\mathbf{u}} \mathbf{u}_t - \mathcal{J}_\star) \\
&\lesssim \sqrt{T d_{\mathbf{u}}^2 d_{\mathbf{x}} \Psi_{B_\star}^2 \|P_\star\|_{\mathrm{op}}^{11} \log \frac{\|P_\star\|_{\mathrm{op}}}{\delta}} \\
&\quad + \mathcal{P}_0 d^2 \Psi_{B_\star}^2 \|P_\star\|_{\mathrm{op}}^{10} (1 + \|K_0\|_{\mathrm{op}}^2) \log \frac{d \Psi_{B_\star}^2 \mathcal{P}_0}{\delta} \log \tfrac{1}{\delta} + d_{\mathbf{x}} \tau_{\mathrm{ls}} \log \frac{1}{\delta}.
\end{aligned}$$

Recalling that $\tau_{\mathrm{ls}} := d \left(\|P_\star\|_{\mathrm{op}}^3 \mathcal{P}_0 + \|P_\star\|_{\mathrm{op}}^{11} \Psi_{B_\star}^6 \right) \log \frac{d \|P_\star\|_{\mathrm{op}}}{\delta}$, that $\mathcal{P}_0, \|P_\star\|_{\mathrm{op}} \Psi_{B_\star} \geq 1$, and that $d\|P_\star\|_{\mathrm{op}} \leq d\mathcal{J}_\star \leq d\mathcal{J}_0 = d^2 \mathcal{P}_0$, we move to a simplified upper bound:

$$\begin{aligned}
\mathrm{Regret}_T[\mathsf{Alg}; A_\star, B_\star] &\lesssim \sqrt{T d_{\mathbf{u}}^2 d_{\mathbf{x}} \Psi_{B_\star}^2 \|P_\star\|_{\mathrm{op}}^{11} \log \frac{\|P_\star\|_{\mathrm{op}}}{\delta}} \\
&\quad + d^2 \mathcal{P}_0 \Psi_{B_\star}^6 \|P_\star\|_{\mathrm{op}}^{11} (1 + \|K_0\|_{\mathrm{op}}^2) \log \frac{d^2 \Psi_{B_\star}^2 \mathcal{P}_0}{\delta} \log^2 \frac{1}{\delta}.
\end{aligned}$$

Since the square of $d\Psi_{B_\star}$ inside the logarithm contributes only a constant factor, we may remove it in the final bound. This concludes the proof.

□

I. Additional Proof Details for Upper Bound (Appendix H)

I.1. Proof of Lemma H.1 (Correctness of Perturbations)

On the event $\mathcal{E}_{\mathrm{safe}}$ of Lemma H.5, the condition defining k_{safe} yields

$$\left\| \left[\widehat{A}_{k_{\mathrm{safe}}} - A_\star \mid \widehat{B}_{k_{\mathrm{safe}}} - B_\star \right] \right\|_{\mathrm{op}}^2 \leq \mathsf{Conf}_{k_{\mathrm{safe}}} \leq 1/3 C_{\mathrm{safe}}(\widehat{A}_{k_{\mathrm{safe}}}, \widehat{B}_{k_{\mathrm{safe}}}).$$

By the continuity of C_{safe} given by Theorem 11, we then have that, for any $(\widehat{A}, \widehat{B}) \in B_{\mathrm{safe}}$,

$$\left\| \left[\widehat{A} - A_\star \mid \widehat{B} - B_\star \right] \right\|_{\mathrm{op}}^2 \leq C_{\mathrm{safe}}(A_\star, B_\star).$$

In particular, the projection step ensures that the above holds for any $(\widehat{A}_k, \widehat{B}_k)$. Let us now go point by point. Theorem 5 then implies that

1. $P_k \preceq \frac{21}{20} P_\star$, and thus $J_k \lesssim \mathcal{J}_\star$.

2. $\mathcal{J}_k - \mathcal{J}_\star = \mathcal{J}_{A_\star,B_\star}[K_\infty(\widehat{A}_k,\widehat{B}_k)] - \mathcal{J}^\star_{A_\star,B_\star} \le C_{\mathrm{est}}(A_\star,B_\star)\epsilon_{\mathrm{F}}^2$.

3. By Lemma B.8,

$$\|\underbrace{K_\infty(\widehat{A}_k,\widehat{B}_k)}_{:=\widehat{K}_k}\|_{\mathrm{op}}^2 \le \|\mathsf{dlyap}(A_\star + B_\star\widehat{K}_k, R_{\mathbf{x}} + \widehat{K}_k^\top R_{\mathbf{u}}\widehat{K}_k)\|_{\mathrm{op}} = \|P_k\|_{\mathrm{op}} \le \frac{21}{20}\|P_\star\|_{\mathrm{op}}.$$

The next two points of the lemma follow from Theorem 8.

For the last point, recall that

$$\sigma_{\mathrm{in}}^2 = \sqrt{d_{\mathbf{x}}}\|P_\infty(\widehat{A}_{k_{\mathrm{safe}}},\widehat{B}_{k_{\mathrm{safe}}})\|_{\mathrm{op}}^{9/2}\max\{1,\|\widehat{B}_{k_{\mathrm{safe}}}\|_{\mathrm{op}}\}\sqrt{\log\frac{\|P_\infty(\widehat{A}_{k_{\mathrm{safe}}},\widehat{B}_{k_{\mathrm{safe}}})\|_{\mathrm{op}}}{\delta}}.$$

Since $\mathsf{Conf}_{k_{\mathrm{safe}}} \lesssim 1$, we have $\max\{1,\|\widehat{B}_{k_{\mathrm{safe}}}\|_{\mathrm{op}}\} \eqsim \Psi_{B_\star}$. Let us show $\|P_\star\|_{\mathrm{op}} \eqsim \|P_\infty(\widehat{A}_{k_{\mathrm{safe}}},\widehat{B}_{k_{\mathrm{safe}}})\|_{\mathrm{op}}$. By Lemma B.6, $P_\infty(\widehat{A}_{k_{\mathrm{safe}}},\widehat{B}_{k_{\mathrm{safe}}}) \preceq P_{k_{\mathrm{safe}}}$, which is $\precsim P_\star$ by point 1 of this lemma. On the other hand, $\|P_\star\|_{\mathrm{op}} \lesssim \|P_\infty(\widehat{A}_{k_{\mathrm{safe}}},\widehat{B}_{k_{\mathrm{safe}}})\|_{\mathrm{op}}$ by Theorem 11.

I.2. Proof of Main Regret Decomposition (Lemma H.2)

We establish Lemma H.2 by establishing a more general regret decomposition for arbitrary feedback controllers K, noise-input variances σ_u, and control costs R_1, R_u. This will allow us to reuse the same computations for similar calculations in the initial estimation phase (Lemma H.5), and for covariance matrix upper bounds as well.

Definition I.1 (Control Evolution Distribution)**.** We define the law $\mathcal{D}(K,\sigma_u,x_1)$ to denote the law of the following dynamical system evolutation: $\mathbf{x}_1 = x_1$, and for $t \ge 2$, the system evolves according to the following distribution:

$$\mathbf{x}_t = A_\star\mathbf{x}_{t-1} + \mathbf{w}_t, \quad \mathbf{u}_t = K\mathbf{x}_t + \sigma_u\mathbf{g}_t, \tag{I.1}$$

where $\mathbf{w}_t \sim \mathcal{N}(0, I_{d_{\mathbf{x}}})$ and $\mathbf{g}_t \sim \mathcal{N}(0, I_{d_{\mathbf{u}}})$.

We begin with the following characterization, proven in Appendix I.6, of the quadratic forms that will arise in our regret bounds. Note that we use arbitrary cost matrices $R_1, R_2 \succeq 0$.

Lemma I.1. *Let K be a stabilizing controller, and let $(\mathbf{x}_t,\mathbf{u}_t)_{t\ge1}$ denote the linear dynamical system described by the evolution of the law $\mathcal{D}(K,\sigma_u,x_1)$. For cost matrices $R_1, R_2 \succeq 0$, define the random variable*

$$\mathsf{Cost}(R_1,R_2;x_1,t,\sigma_u) := \sum_{s=1}^t \mathbf{x}_t^\top R_1\mathbf{x}_t + \mathbf{u}_t^\top R_2\mathbf{u}_t = \overline{\mathbf{g}}^\top\Lambda_{\overline{\mathbf{g}}}\overline{\mathbf{g}} + x_1^\top\Lambda_{\mathbf{x}_1}x_1 + 2\overline{\mathbf{g}}^\top\Lambda_{\mathrm{cross}}x_1.$$

Further, define $R_K = R_1 + K^\top R_2 K$, $A_K = A_\star + B_\star K$, $P_K = \mathsf{dlyap}(A_K, R_K)$, and $J_K := \mathrm{tr}(P_K)$.

1. *In expectation, we have*

$$\mathbb{E}[\mathsf{Cost}(R_1,R_2;x_1,t,\sigma_u)] \le tJ_K + 2\sigma_u^2 t d_{\mathbf{u}}\left(\|R_2\|_{\mathrm{op}} + \|B_\star\|_{\mathrm{op}}^2\|P_K\|_{\mathrm{op}}\right) + x_1^\top P_K x_1$$

2. *Set $d_{\mathrm{eff}} := \min\{d_{\mathbf{u}}, \mathrm{rank}(R_1) + \mathrm{rank}(R_2)\}$. With a probability $1-\delta$, we have*

$$\begin{aligned}\mathsf{Cost}(R_1,R_2;x_1,t,\sigma_u) \le{}& tJ_K + 2\sigma_u^2 d_{\mathrm{eff}} t\left(\|R_2\|_{\mathrm{op}} + \|B_\star\|_{\mathrm{op}}^2\|P_K\|_{\mathrm{op}}\right)\\ &+ \mathcal{O}\left(\sqrt{dt\log\tfrac{1}{\delta}} + \log\tfrac{1}{\delta}\right)\left((1+\sigma_u^2\|B_\star\|_{\mathrm{op}}^2)\|R_K\|_{\mathrm{op}}\|A_K\|_{\mathcal{H}_\infty}^2 + \sigma_u^2\|R_2\|_{\mathrm{op}}^2\right)\\ &+ 2x_1^\top P_K x_1.\end{aligned}$$

3. *More crudely, we can also bound, with probability $1-\delta$,*

$$\mathsf{Cost}(R_1,R_2;x_1,t,\sigma_u) \lesssim t\log\frac{1}{\delta}\left(J_K + 2\sigma_u^2 d_{\mathrm{eff}}\left(\|R_2\|_{\mathrm{op}} + \|B_\star\|_{\mathrm{op}}^2\|P_K\|_{\mathrm{op}}\right)\right) + 2x_1^\top P_K x_1. \tag{I.2}$$

Let us now apply the above lemma to our present setting. For $k \geq k_{\mathrm{safe}}$, define the terms

$$\mathsf{Cost}_{\mathrm{noise},k} := d_{\mathbf{u}} \left(\|R_{\mathbf{u}}\|_{\mathrm{op}} + \|B_\star\|_{\mathrm{op}}^2 \|P_k\|_{\mathrm{op}} \right)$$
$$\mathsf{Cost}_{\mathrm{conc},k} := \left((1 + \sigma_k^2 \|B_\star\|_{\mathrm{op}}^2) \|R_{\mathbf{x}} + \widehat{K}_k^\top R_{\mathbf{u}} \widehat{K}_k\|_{\mathrm{op}} \|A_{\mathrm{cl},k}\|_{\mathcal{H}_\infty}^2 \right) + \sigma_k^2 \|R_{\mathbf{u}}\|_{\mathrm{op}}.$$

By Lemma I.1 and the fact $\mathcal{J}_\star \leq J_k$,

$$\sum_{t=\tau_{k_{\mathrm{safe}}}}^{T} (\mathbf{x}_t^\top R_{\mathbf{x}} \mathbf{x}_t + \mathbf{u}_t^\top R_{\mathbf{u}} \mathbf{u}_t - \mathcal{J}_\star) \lesssim \sum_{k=k_{\mathrm{safe}}}^{k_{\mathrm{fin}}} \tau_k (J_k - \mathcal{J}_\star) + \tau_k \sigma_k^2 \mathsf{Cost}_{\mathrm{noise},k}$$
$$+ \sum_{k=k_{\mathrm{safe}}}^{k_{\mathrm{fin}}} (\sqrt{\tau_k d \log(1/\delta)} + \log(1/\delta) \mathsf{Cost}_{\mathrm{conc},k} + \sum_{k=k_{\mathrm{safe}}}^{k_{\mathrm{fin}}} \mathbf{x}_{\tau_k}^\top P_k \mathbf{x}_{\tau_k}.$$

Let us first bound the $\mathsf{Cost}_{\mathrm{noise},k}$-terms. Since $1 \leq \|P_k\|_{\mathrm{op}} \lesssim \|P_\star\|_{\mathrm{op}}$ on event $\mathcal{E}_{\mathrm{safe}}$ (Lemma H.1) and $\|R_{\mathbf{u}}\|_{\mathrm{op}} = 1$, we have

$$\mathsf{Cost}_{\mathrm{noise},k} \leq d_{\mathbf{u}} (\|R_{\mathbf{u}}\|_{\mathrm{op}} + \|B_\star\|_{\mathrm{op}}^2 \|P_k\|_{\mathrm{op}}) \lesssim d_{\mathbf{u}} \Psi_{B_\star}^2 \|P_\star\|_{\mathrm{op}}.$$

Since $\sigma_k^2 \leq \sigma_{\mathrm{in}}^2 \tau_k^{-1/2}$ and $\|P_k\|_{\mathrm{op}} \lesssim \|P_\star\|_{\mathrm{op}}$, we then obtain

$$\sum_{k=k_{\mathrm{safe}}}^{k_{\mathrm{fin}}} \tau_k \sigma_k^2 \mathsf{Cost}_{\mathrm{noise},k} \lesssim \sqrt{T} d_{\mathbf{u}} \sigma_{\mathrm{in}}^2 \Psi_{B_\star}^2 \|P_\star\|_{\mathrm{op}}.$$

Next, let us bound

$$\mathsf{Cost}_{\mathrm{conc},k} := \left((1 + \sigma_k^2 \|B_\star\|_{\mathrm{op}}^2) \|R_{\mathbf{x}} + \widehat{K}_k^\top R_{\mathbf{u}} \widehat{K}_k\|_{\mathrm{op}} \|A_{\mathrm{cl},k}\|_{\mathcal{H}_\infty}^2 \right) + \sigma_k^2 \|R_{\mathbf{u}}\|_{\mathrm{op}}.$$

Observe that $R_{\mathbf{x}} + \widehat{K}_k^\top R_{\mathbf{u}} \widehat{K}_k \preceq \mathsf{dlyap}[A_{\mathrm{cl},k}, R_{\mathbf{x}} + \widehat{K}_k^\top R_{\mathbf{u}} \widehat{K}_k] = P_k$. On the good event $\mathcal{E}_{\mathrm{safe}}$, we have $\|P_k\|_{\mathrm{op}} \lesssim \|P_\star\|_{\mathrm{op}}$, $\|A_{\mathrm{cl},k}\|_{\mathcal{H}_\infty} \lesssim \|A_{\mathrm{cl},\star}\|_{\mathcal{H}_\infty} \leq \|P_\star\|_{\mathrm{op}}^{3/2}$ (Lemma H.1), and by definition. $\|B_\star\|_{\mathrm{op}}^2 \leq \Psi_{B_\star}^2$. Thus, the above is at most (again taking $R_{\mathbf{u}} = I$)

$$\mathsf{Cost}_{\mathrm{conc},k} \lesssim \|P_\star\|_{\mathrm{op}}^4 + \sigma_k^2 \left(\|R_{\mathbf{u}}\|_{\mathrm{op}} + \Psi_{B_\star}^2 \|P_\star\|_{\mathrm{op}}^4 \right) \leq \|P_\star\|_{\mathrm{op}}^4 \left(1 + \Psi_{B_\star}^2 \sigma_k^2 \right).$$

Therefore,

$$\sum_{k=k_{\mathrm{safe}}}^{k_{\mathrm{fin}}} (\sqrt{\tau_k d \log(1/\delta)} + \log(1/\delta)) \mathsf{Cost}_{\mathrm{conc},k} \lesssim \sqrt{T d \log(1/\delta)} \|P_\star\|_{\mathrm{op}}^4 + \log(T) \log(1/\delta) \|P_\star\|_{\mathrm{op}}^4$$
$$+ \sigma_{\mathrm{in}}^2 \log(T) \log(1/\delta) \sqrt{d} \Psi_{B_\star}^2 \|P_\star\|_{\mathrm{op}}^4.$$
$$\leq \sqrt{T d \log(1/\delta)} \|P_\star\|_{\mathrm{op}}^4 + \log^2 \frac{1}{\delta} (1 + \sqrt{d} \sigma_{\mathrm{in}}^2 \Psi_{B_\star}^2) \|P_\star\|_{\mathrm{op}}^4.$$

where we use $\log(T) \leq \log(1/\delta)$. Finally, we have the bound

$$\sum_{k=k_{\mathrm{safe}}}^{k_{\mathrm{fin}}} \mathbf{x}_{\tau_k}^\top P_k \mathbf{x}_{\tau_k} \lesssim \log T \max_{k \leq \log T} \mathbf{x}_{\tau_k}^\top P_k \mathbf{x}_{\tau_k}$$
$$\leq \log T \max_{k \leq \log T} \|\mathbf{x}_{\tau_k}\|_2^2 \|P_k\|_{\mathrm{op}}$$
$$\lesssim \log T \max_{k \leq \log T} \|\mathbf{x}_{\tau_k}\|_2^2 \|P_\star\|_{\mathrm{op}}.$$

Hence, putting things together, we have

$$\sum_{t=\tau_{k_{\mathrm{safe}}}}^{T} (\mathbf{x}_t^\top R_{\mathbf{x}} \mathbf{x}_t + \mathbf{u}_t^\top R_{\mathbf{u}} \mathbf{u}_t - \mathcal{J}_\star) \lesssim \sum_{k=k_{\mathrm{safe}}}^{k_{\mathrm{fin}}} \tau_k (J_k - \mathcal{J}_\star) + \log T \max_{k \leq \log_T} \|\mathbf{x}_{\tau_k}\|_2^2$$
$$+ \sqrt{T} \left(d_{\mathbf{u}} \sigma_{\mathrm{in}}^2 \Psi_{B_\star}^2 \|P_\star\|_{\mathrm{op}}) + \sqrt{d \log(1/\delta)} \|P_\star\|_{\mathrm{op}}^4 \right)$$
$$+ \log^2 \frac{1}{\delta} (1 + \sqrt{d} \sigma_{\mathrm{in}}^2 \Psi_{B_\star}^2) \|P_\star\|_{\mathrm{op}}^4.$$

Reparameterizing $\delta \leftarrow \frac{\delta}{6T}$ and taking a union bound preserves the above inequality up to constants (since $\log T \le \log \frac{1}{\delta}$), and reduces the failure probability across all episodes to $\delta/6$.

I.3. Bounding the States: Lemma H.3

Lemma I.2. *Let $\mathbf{x}_t$ denote the t-th iterate form the law $\mathcal{D}(K, x_1, \sigma_u)$. Then, with probability at least $1-\delta$,*

$$\|\mathbf{x}_t - A_K^{t-1} x_1\| \le \mathcal{O}\left(\sqrt{J_K(1+\sigma_u^2\|B_\star\|_2^2)\log\frac{1}{\delta}}\right).$$

Let $\alpha_0 = \sqrt{\mathcal{J}_0 \Psi_{B_\star}^2 \log\frac{1}{\delta}}$, We conclude by arguing an upper bound on $\tau_{k_{\mathrm{safe}}}$. We rely on the following guaranteeand $\alpha_1 = \sqrt{\mathcal{J}_{\max} \Psi_{B_\star}^2 \log\frac{1}{\delta}}$. For $k > k_{\mathrm{safe}}$, define the vector $\boldsymbol{e}_k := \mathbf{x}_{\tau_k} - A_{\mathrm{cl},k-1}^{\tau_{k-1}} \mathbf{x}_{\tau_{k-1}}$. Since $\delta < 1/T$, $\sigma_k^2 \le 1$, and $\mathcal{J}_k \lesssim \mathcal{J}_\star$, a union bound and reparametrization of δ implies that, $\boldsymbol{e}_k \lesssim \alpha_1$ and $\|\mathbf{x}_{\tau_{k_{\mathrm{safe}}}}\|_2 \le \alpha_0$ with probability $1-\delta/8$. Now, we can write

$$\begin{aligned}
\mathbf{x}_{\tau_k} &= \boldsymbol{e}_k + A_{\mathrm{cl},k-1}^{\tau_{k-1}} \mathbf{x}_{\tau_{k-1}} \\
&= \boldsymbol{e}_k + A_{\mathrm{cl},k-1}^{\tau_{k-1}} \left(\boldsymbol{e}_{k-1} + A_{\mathrm{cl},k-2}^{\tau_{k-2}} \mathbf{x}_{\tau_{k-2}}\right) \\
&= \sum_{j=k_{\mathrm{safe}}+1}^{k} \left(\prod_{i=j}^{k-1} A_{\mathrm{cl},i}^{\tau_i}\right) \boldsymbol{e}_j + \left(\prod_{i=k_{\mathrm{safe}}}^{k-1} A_{\mathrm{cl},i}^{\tau_i}\right) \mathbf{x}_{\tau_{k_{\mathrm{safe}}}}.
\end{aligned}$$

Since $\mathsf{dlyap}[A_{\mathrm{cl},\star}] \succeq I$, we have

$$\left\|\mathsf{dlyap}[A_{\mathrm{cl},\star}]^{1/2} \prod_{i=j}^{k-1} A_{\mathrm{cl},i}\right\|_{\mathrm{op}} \le \sqrt{\left(\prod_{i=j}^{k-1} A_{\mathrm{cl},i}^{\tau_i}\right)^\top \mathsf{dlyap}[A_{\mathrm{cl},\star}] \left(\prod_{i=j}^{k-1} A_{\mathrm{cl},i}^{\tau_i}\right)}.$$

Moreover, by Lemma H.1, we have that for all $i \ge k_{\mathrm{safe}}$, $A_{\mathrm{cl},i}\mathsf{dlyap}[A_{\mathrm{cl},\star}]A_{\mathrm{cl},i} \preceq (1 - \frac{1}{2\|\mathsf{dlyap}[A_{\mathrm{cl},\star}]\|_{\mathrm{op}}})\mathsf{dlyap}[A_{\mathrm{cl},\star}]$. This yields that

$$\begin{aligned}
\left\|\mathsf{dlyap}[A_{\mathrm{cl},\star}]^{1/2} \prod_{i=j}^{k-1} A_{\mathrm{cl},i}\right\|_{\mathrm{op}} &\le \sqrt{\left(\prod_{i=j}^{k-1}\left(1 - \frac{1}{2\|\mathsf{dlyap}[A_{\mathrm{cl},\star}]\|_{\mathrm{op}}}\right)^{\tau_i}\right)^2 \|\mathsf{dlyap}[A_{\mathrm{cl},\star}]\|_{\mathrm{op}}} \\
&= \left(1 - \frac{1}{2\|\mathsf{dlyap}[A_{\mathrm{cl},\star}]\|_{\mathrm{op}}}\right)^{\sum_{i=j}^{k-1}\tau_i} \sqrt{\|\mathsf{dlyap}[A_{\mathrm{cl},\star}]\|_{\mathrm{op}}} \\
&= \left(1 - \frac{1}{2\|\mathsf{dlyap}[A_{\mathrm{cl},\star}]\|_{\mathrm{op}}}\right)^{\tau_{k-1}} \sqrt{\|\mathsf{dlyap}[A_{\mathrm{cl},\star}]\|_{\mathrm{op}}}.
\end{aligned}$$

Hence, we have that, with probability $1 - \mathcal{O}(\delta)$,

$$\begin{aligned}
&\|\mathsf{dlyap}[A_{\mathrm{cl},\star}]^{1/2}\mathbf{x}_{\tau_k}\|_2 \lesssim \\
&\quad \alpha_1\sqrt{\|\mathsf{dlyap}[A_{\mathrm{cl},\star}]\|_{\mathrm{op}}}\left(1 + k\left(1 - \frac{1}{2\|\mathsf{dlyap}[A_{\mathrm{cl},\star}]\|_{\mathrm{op}}}\right)^{\tau_{k-1}}\right) + \alpha_0\sqrt{\|\mathsf{dlyap}[A_{\mathrm{cl},\star}]\|_{\mathrm{op}}}.
\end{aligned}$$

Since $\max_k k(1-\rho)^k \lesssim \frac{1}{\rho}$ and since $I \preceq \mathsf{dlyap}[A_{\mathrm{cl},\star}] \preceq P_\star$ (see Lemma H.1), this implies the crude bound

$$\begin{aligned}
\|\mathbf{x}_{\tau_k}\|_2 &\le \lesssim \sqrt{\|\mathsf{dlyap}[A_{\mathrm{cl},\star}]\|_{\mathrm{op}}}(\alpha_0 + \alpha_1\|\|\mathsf{dlyap}[A_{\mathrm{cl},\star}]\|_{\mathrm{op}}\|_{\mathrm{op}}) \le \|P_\star\|_{\mathrm{op}}^{3/2}(\alpha_0+\alpha_1) \\
&\lesssim \sqrt{\Psi_{B_\star}(\mathcal{J}_0 + \mathcal{J}_\star)\log(1/\delta)}\|P_\star\|_{\mathrm{op}}^{3/2} \\
&\lesssim \sqrt{\Psi_{B_\star}\mathcal{J}_0\log(1/\delta)}\|P_\star\|_{\mathrm{op}}^{3/2}.
\end{aligned}$$

□

I.4. Proof of Estimation Bound (Lemma H.4)

Definition I.2 (Round-wise projections)**.** Given $v \in \mathbb{R}^d$, let $v = (v^x, v^u)$ denote its decomposition along the x and u directions. For a given round $k \geq k_{\mathrm{safe}}$, let $\mathcal{V}_k := \{v \in \mathbb{R}^d : v^x + \widehat{K}_k v^u = 0\}$, and let $\mathcal{V}_k^\perp$ denotes it orthogonal complement. Finally, let P_k denote the orthogonal projection onto $\mathcal{V}_k$, and let $\mathsf{P}_k^\perp := (I - \mathsf{P}_k)$ denote the projection on $\mathcal{V}_k^\perp$.

The first step in our bound will be to lower bound the relevant, centered covariances.

Lemma I.3 (Round-wise covariance lower bound)**.** *Let $k \geq k_{\mathrm{safe}} + 1$, at let $t \in \{\tau_k, \dots, \tau_{k+1} - 1\}$. Then, on $\mathcal{E}_{\mathrm{safe}}$. If σ_k^2 satisfies $\sigma_k^2 \leq \frac{1}{6.2\|P_\star\|_{\mathrm{op}}}$, we have that*

$$\mathbb{E}\left[(\mathbf{z}_t - \mathbb{E}[\mathbf{z}_t \mid \mathcal{F}_{t-1}])(\mathbf{z}_t - \mathbb{E}[\mathbf{z}_t \mid \mathcal{F}_{t-1}])^\top\right] \succeq \Gamma_k := \frac{\sigma_k^2}{6.2\|P_\star\|_{\mathrm{op}}}.\mathsf{P}_k + \frac{1}{2}\mathsf{P}_k^\perp.$$

See Section I.4.1 for the proof. We now convert the above bound into a Löwner lower bound, then conclude by giving an upper bound on $\tau_{k_{\mathrm{safe}}}$. We rely on the following guarantee $\mathbf{\Lambda}_k$. To state the bound, we introduce some additional notation.

Definition I.3. We say that $f(x) \gtrsim_\star g(x)$ if $f \geq Cg$ for a sufficiently large contstant C.

Further, let $v_{k,1}, \dots, v_{k,d}$ denote an eigenbasis of Γ_k. Let us prove the following.

Lemma I.4. . *The following bounds hold simultaneously with probability $1 - \delta/2$, if $\mathcal{E}_{\mathrm{bound}} \cap \mathcal{E}_{\mathrm{safe}}$ holds:*

1. *$i \in \{d_{\mathbf{x}} + 1, \dots, d\}$, we have $v_{k,i}^\top \mathbf{\Lambda}_k v_{k,i} \lesssim \tau_k \sigma_k^2$ if $\tau_k \gtrsim_\star \sqrt{\log(d/\delta)}$.*
2. *Suppose that $\tau_k \geq \|P_\star\|_{\mathrm{op}}^3 \mathcal{J}_0 \vee \Psi_{B_\star}^4 \sigma_{\mathrm{in}}^4$. Then,*
$$v_{k,i}^\top \mathbf{\Lambda}_k v_{k,i} \lesssim \tau_k \|P_\star\|_{\mathrm{op}} \log(d/\delta).$$
3. *If $\tau_k \gtrsim_\star \tau_{ls} := d\left(\|P_\star\|_{\mathrm{op}}^3 \mathcal{P}_0 + \|P_\star\|_{\mathrm{op}}^{11} \Psi_{B_\star}^6\right) \log \frac{d\|P_\star\|_{\mathrm{op}}}{\delta}$, then the above two conditionds hold, σ_k^2 sastisfies the conditions of Lemma I.3, and $\mathbf{\Lambda}_k \succeq c\tau_k \Gamma_k$ for some universal constant $c > 0$.*

The proof is defered to Appendix I.4.2. From lemma E.3 with covariate dimension d and output dimension $d_{\mathbf{x}}$, we have that with probability $1 - \delta/2$ on the events of Lemma E.4 that

$$\|\widehat{A} - A_\star\|_{\mathrm{F}}^2 + \|\widehat{B} - B_\star\|_{\mathrm{F}}^2 \lesssim d_{\mathbf{x}} \sum_{j=1}^{d} \lambda_j(\tau_k \Gamma_k)^{-1} \boldsymbol{\kappa}_j \log \frac{3\boldsymbol{\kappa}_j}{\delta}.$$

where $\boldsymbol{\kappa}_i \lesssim \frac{v_{k,i}^\top \mathbf{\Lambda}_k v_{k,i}}{\tau_k \lambda_i(\Gamma_k)}$. Let us decompose the above sum intro the sum over the first k indices, and the second. For $i \in [d_{\mathbf{x}}]$, we have $\lambda_i(\Gamma_k) \geq \frac{1}{2}$, and on the events of Lemma E.4, we can bound $v_{k,i}^\top \mathbf{\Lambda}_k v_{k,i} \lesssim \tau_k \|P_\star\|_{\mathrm{op}} \log(1/\delta)$, yielding $\boldsymbol{\kappa}_i \lesssim \|P_\star\|_{\mathrm{op}} \log(1/\delta)$.

$$\begin{aligned}
\sum_{j=1}^{d_{\mathbf{x}}} \lambda_j(\tau_k \Gamma_k)^{-1} \boldsymbol{\kappa}_j \log \frac{3\boldsymbol{\kappa}_j}{\delta} &\lesssim \frac{d_{\mathbf{x}}^2 \|P_\star\|_{\mathrm{op}} \log(1/\delta)}{\tau_k} \log\left(\frac{\|P_\star\|_{\mathrm{op}} \log(1/\delta)}{\delta}\right) \\
&\lesssim \frac{d_{\mathbf{x}}^2 \|P_\star\|_{\mathrm{op}} \log \frac{1}{\delta}}{\tau_k} \log \frac{\|P_\star\|_{\mathrm{op}}}{\delta} \lesssim \frac{d_{\mathbf{x}}^2 \|P_\star\|_{\mathrm{op}}^2 \log^2 \frac{1}{\delta}}{\tau_k}.
\end{aligned}$$

For $i > d_{\mathbf{x}}$, $\lambda_i(\Gamma_k) \gtrsim \frac{\sigma_k^2}{\|P_\star\|_{\mathrm{op}}}$, on the events of Lemma E.4, we can bound $v_{k,i}^\top \mathbf{\Lambda}_k v_{k,i} \lesssim 1$, yielding $\boldsymbol{\kappa}_i \lesssim \|P_\star\|_{\mathrm{op}}$, and thus

$$\begin{aligned}
\sum_{j=d_{\mathbf{x}}+1}^{d} d_{\mathbf{x}} \lambda_j(\tau_k \Gamma_k)^{-1} \boldsymbol{\kappa}_j \log \frac{3\boldsymbol{\kappa}_j}{\delta} &\lesssim d_{\mathbf{u}} d_{\mathbf{x}} \|P_\star\|_{\mathrm{op}}^2 \log \|P_\star\|_{\mathrm{op}} \frac{1}{\delta \sigma_k^2 \tau_k} \\
&\lesssim d_{\mathbf{u}} d_{\mathbf{x}} \|P_\star\|_{\mathrm{op}}^2 \log \frac{\|P_\star\|_{\mathrm{op}}}{\delta} \frac{1}{\sigma_{\mathrm{in}}^2 \sqrt{\tau_k}}.
\end{aligned}$$

Combining the two summations gives the desired bound:

$$\|\widehat{A} - A_\star\|_{\mathrm{F}}^2 + \|\widehat{B} - B_\star\|_{\mathrm{F}}^2 \lesssim \frac{d_{\mathbf{u}} d_{\mathbf{x}}}{\sigma_{\mathrm{in}}^2 \sqrt{\tau_k}} \|P_\star\|_{\mathrm{op}}^2 \log \frac{\|P_\star\|_{\mathrm{op}}}{\delta} + \frac{d_{\mathbf{x}}^2}{\tau_k} \|P_\star\|_{\mathrm{op}}^2 \log^2 \frac{1}{\delta}.$$

Finally, reparametrizing $\delta \leftarrow \delta/8$ gives the desired probability. □

I.4.1. Proof of Lemma I.3

Define

$$\Sigma_k = \begin{bmatrix} I_{d_{\mathbf{x}}} \\ \widehat{K}_k^\top \end{bmatrix} \begin{bmatrix} I_{d_{\mathbf{x}}} & \widehat{K}_k \end{bmatrix} + \begin{bmatrix} 0 & 0 \\ 0 & \sigma_k^2 I \end{bmatrix} = \begin{bmatrix} I_{d_{\mathbf{x}}} & \widehat{K}_k \\ \widehat{K}^\top & \widehat{K}_k^\top \widehat{K}_k + \sigma_k^2 I \end{bmatrix}.$$

We see that for $k \geq \boldsymbol{k}_0 + 1$ and $t \geq \tau_k$, we have that $\mathbf{z}_t \mid \mathcal{F}_{t-1} \sim \mathcal{N}(\overline{\mathbf{z}}_t, \Sigma_k)$, where $\overline{\mathbf{z}}_t$ and Σ_t are $\mathcal{F}_{t-1}$-measurable. Our goal will now be to lower bound $\Sigma_k \succsim \gamma_1 \mathsf{P}_k + \gamma_2 \mathsf{P}_k^\perp$. To this end, let $v \in \mathcal{S}^{d-1}$, and write $v = \mathsf{P}_k v + \mathsf{P}_k^\perp v := v_\parallel + v_\perp$. Observe then that

$$\left|v_\perp^\top \Sigma_k v_\parallel\right| \leq \left|v_\perp^\top \begin{bmatrix} I_{d_{\mathbf{x}}} \\ \widehat{K}_k^\top \end{bmatrix} \begin{bmatrix} I_{d_{\mathbf{x}}} & \widehat{K}_k \end{bmatrix} v_\parallel\right| + \left|v_\perp^\top \begin{bmatrix} 0 & 0 \\ 0 & \sigma_k^2 I \end{bmatrix} v_\parallel\right| = \left|v_\perp^\top \begin{bmatrix} 0 & 0 \\ 0 & \sigma_k^2 I \end{bmatrix} v_\parallel\right| \leq \sigma_k^2 \|v_\perp\| \|v_\parallel\|,$$

where we use that $\begin{bmatrix} I_{d_{\mathbf{x}}} & \widehat{K}_k \end{bmatrix} v_\parallel = 0$. On the other hand, since $v_\perp \in \mathsf{null}\left(\begin{bmatrix} I_{d_{\mathbf{x}}} & \widehat{K}_k \end{bmatrix}\right)^\perp$, we have

$$\begin{aligned} v_\perp^\top \Sigma_k v_\perp &\geq v_\perp^\top \begin{bmatrix} I_{d_{\mathbf{x}}} \\ \widehat{K}_k^\top \end{bmatrix} \begin{bmatrix} I_{d_{\mathbf{x}}} & \widehat{K}_k \end{bmatrix} v_\perp \\ &= \| \begin{bmatrix} I_{d_{\mathbf{x}}} & \widehat{K}_k \end{bmatrix} v_\perp \|^2 \\ &\geq \|v_\perp\|^2 \sigma_{d_{\mathbf{x}}}\left(\begin{bmatrix} I_{d_{\mathbf{x}}} & \widehat{K}_k \end{bmatrix}\right)^2 = \|v_\perp\|^2 \lambda_{\min}(I_{d_{\mathbf{x}}} + \widehat{K}_k^\top \widehat{K}_k) \geq \|v_\perp\|^2. \end{aligned}$$

We can therefore bound, for any $\alpha > 0$,

$$\begin{aligned} v^\top \Sigma_k v &= v_\perp^\top \Sigma_k v_\perp + 2 v_\perp^\top \Sigma_k v_\parallel + v_\parallel^\top \Sigma_k v_\parallel \\ &\geq \|v_\perp\|^2 - 2\sigma_k^2 \|v_\perp\| \|v_\parallel\| + \lambda_{\min}(\Sigma_k) \|v_\perp\|^2 \\ &\geq \|v_\perp\|^2 - \sigma_k^2(\alpha \|v_\parallel\|^2 + \frac{1}{\alpha}\|v_\parallel\|^2) + \lambda_{\min}(\Sigma_k)\|v_\perp\|^2. \end{aligned}$$

Taking $\alpha = \lambda_{\min}(\Sigma_k)/2\sigma_k^2$, we have

$$v^\top \Sigma_k v \geq \|v_\perp\|^2 \underbrace{(1 - \sigma_k^2 \cdot \frac{2\sigma_k^2}{\lambda_{\min}(\Sigma_k)})}_{:=\gamma_1} + \underbrace{\frac{1}{2}\lambda_{\min}(\Sigma_k)}_{:=\gamma_2} \|v_\perp\|^2.$$

Hence, we have show that, for γ_1, γ_2 defined in the above display, $\Sigma_k \succeq \gamma_1 \mathsf{P}_k^\perp + \gamma_2 \mathsf{P}_k$. Let us now lower bound each of these quantities. From Dean et al. (2018, Lemma F.6), since $\|\widehat{K}_k\|^2 \lesssim \|P_\star\|_{\mathrm{op}}$ (Lemma H.1), and $\|P_\star\|_{\mathrm{op}} \geq 1 \geq \sigma_k^2$,

$$\begin{aligned} \lambda_{\min}(\Sigma_k) &\geq \sigma_k^2 \min\left\{\frac{1}{2}, \frac{1}{2\|\widehat{K}_k\|_{\mathrm{op}}^2 + \sigma_k^2}\right\} \geq \sigma_k^2 \min\left\{\frac{1}{2}, \frac{1}{2.1\|P_\star\|_{\mathrm{op}} + \sigma_k^2}\right\} \\ &\geq \sigma_k^2 \min\left\{\frac{1}{2}, \frac{1}{3.1\|P_\star\|_{\mathrm{op}}}\right\} = \frac{\sigma_k^2}{6.2\|P_\star\|_{\mathrm{op}}}. \end{aligned}$$

Hence, for $\sigma_k^2 \leq \frac{1}{6.2\|P_\star\|_{\mathrm{op}}}$, we have $\gamma_1 \geq \frac{1}{2}$, and $\gamma_2 \geq \frac{\sigma_k^2}{3.1\|P_\star\|_{\mathrm{op}}}$.

I.4.2. Proof of Lemma I.8

Proof. All union bounds will be absorbed into δ factors, as $\delta \leq 1/T$ and $T \geq d$. We decompose $v_{k,i} = v_{k,i}^x + v_{k,i}^u$ along its x and u coordinate. It suffices to show that each bound holds individually with probability $1 - \delta$ for a fixed i, and k, since the union bound over k can be absorbed into the δ factor (as $\delta \leq 1/T$), and dimension addressed by reparametrizing $\delta \leftarrow \delta/d$.

Point 1: For $i \in \{d_{\mathbf{x}}+1, \dots, k\}$, $v_{k,i}$ lies in the vector space $\mathcal{V}_k$. Therefore

$$\begin{aligned} v_{k,i}^\top \mathbf{\Lambda}_k v_{k,i} &= \sum_{t=\tau_k}^{2\tau_k-1} v_{k,i}^\top \begin{bmatrix} \mathbf{x}_t \\ \mathbf{u}_t \end{bmatrix} \begin{bmatrix} \mathbf{x}_t \\ \mathbf{u}_t \end{bmatrix}^\top v_{k,i} \\ &= \sum_{t=\tau_k}^{2\tau_k-1} v_{k,i}^\top \begin{bmatrix} \mathbf{x}_t \\ \widehat{K}_k \mathbf{x}_t + \sigma_k \mathbf{g}_t \end{bmatrix} \begin{bmatrix} \mathbf{x}_t \\ \widehat{K}_k \mathbf{x}_t + \sigma_k \mathbf{g}_t \end{bmatrix}^\top v_{k,i} \\ &= \sum_{t=\tau_k}^{2\tau_k-1} v_i^\top \begin{bmatrix} \\ \sigma_k \mathbf{g}_t \end{bmatrix} \begin{bmatrix} \mathbf{x}_t \\ \mathbf{x}_t + \sigma_k \mathbf{g}_t \end{bmatrix}^\top v_{k,i} \\ &= \sigma_k^2 \sum_{t=\tau_k}^{2\tau_k-1} \langle v_{k,i}^u, \mathbf{g}_t \rangle^2 \sim \|v_{k,i}^u\|_2^2 \sigma_k^2 \cdot \chi^2(\tau_k). \end{aligned}$$

By standard χ^2-concentration, the above is $\lesssim \tau_k \|v_{k,i}^u\|_2^2 \sigma_k^2 \le \tau_k \sigma_k^2$ for $\tau_k \ge \sqrt{\log(1/\delta)}$.

Point 2: For arbitrary i, set $R_1 := v_{k,i}^x (v_{k,i}^x)^\top$ and $R_2 := v_{k,i}^u (v_{k,u}^u)^\top$.

$$v_{k,i}^\top \mathbf{\Lambda}_k v_{k,i} = \sum_{t=\tau_k}^{2\tau_k-1} v_{k,i}^\top \begin{bmatrix} \mathbf{x}_t \\ \mathbf{u}_t \end{bmatrix} \begin{bmatrix} \mathbf{x}_t \\ \mathbf{u}_t \end{bmatrix}^\top v_{k,i} \le 2 \sum_{t=\tau_k}^{2\tau_k-1} \mathbf{x}_t^\top R_1 \mathbf{x}_1 + \mathbf{u}_t^\top R_2 \mathbf{u}_t.$$

Thus, Lemma I.1 ensures that, with probability $1-\delta$, we have that for the matrix $P := \mathsf{dlyap}(A_\star + B_\star \widehat{K}_k, R_1 + \widehat{K}_k^\top R_2 \widehat{K}_k)$,

$$v_{k,i}^\top \mathbf{\Lambda}_k v_{k,i} \lesssim \tau_k \log \frac{1}{\delta} \left(\mathrm{tr}(P) + 2\sigma_k^2 d_{\mathrm{eff}} \left(\|R_2\|_{\mathrm{op}} + \|B_\star\|_{\mathrm{op}}^2 \|P\|_{\mathrm{op}} \right) \right) + \|P\|_{\mathrm{op}} \|\mathbf{x}_{\tau_k}\|^2,$$

where $d_{\mathrm{eff}} \le \mathrm{rank}(R_1) + \mathrm{rank}(R_2) = 2$. Since $R_1 \preceq I \preceq R_{\mathbf{x}}$ and $R_2 \preceq I = R_{\mathbf{u}}$, we have $R_1 + \widehat{K}_k^\top R_2 \widehat{K}_k \preceq R_{\mathbf{x}} + \widehat{K}_k^\top R_{\mathbf{u}} \widehat{K}_k$, and thus (by Lemma B.5), $P = \mathsf{dlyap}(A_\star + B_\star \widehat{K}_k, R_1 + \widehat{K}_k^\top R_2 \widehat{K}_k) \preceq \mathsf{dlyap}(A_\star + B_\star \widehat{K}_k, R_{\mathbf{x}} + \widehat{K}_k^\top R_{\mathbf{u}} \widehat{K}_k) = P_k$. Moerover Lemma H.1, we get $\|P_k\|_{\mathrm{op}} \lesssim \|P_\star\|_{\mathrm{op}}$. Moreover, since P can be shown to have rank at most 2, $\mathrm{tr}(P) \lesssim \|P_\star\|_{\mathrm{op}}$. Finally, $\|x_{\tau_k}\|_2 \le \sqrt{\mathcal{J}_0 \log(1/\delta)} \|P_\star\|_{\mathrm{op}}^{3/2}$ from Lemma I.3,

$$\begin{aligned} v_{k,i}^\top \mathbf{\Lambda}_k v_{k,i} &\lesssim \tau_k \log \frac{1}{\delta} \left(\|P_\star\|_{\mathrm{op}} + \sigma_k^2 \left(1 + \|B_\star\|_{\mathrm{op}}^2 \|P_\star\|_{\mathrm{op}} \right) \right) + \mathcal{J}_0 \log(1/\delta) \|P_\star\|_{\mathrm{op}}^4 \\ &\lesssim \tau_k \log \frac{1}{\delta} \left(\|P_\star\|_{\mathrm{op}} + \sigma_k^2 \|P_\star\|_{\mathrm{op}} \Psi_{B_\star}^2 \right) + \mathcal{J}_0 \log(1/\delta) \|P_\star\|_{\mathrm{op}}^4. \end{aligned}$$

In particular, if $\tau_k \ge \mathcal{J}_0 \|P_\star\|_{\mathrm{op}}^3$, and $\sigma_k \le 1/\Psi_{B_\star}^2$ (for which it suffices $\tau_k \le \sigma_{\mathrm{in}}^4 \Psi_{B_\star}^4$), we have

$$v_{k,i}^\top \mathbf{\Lambda}_k v_{k,i} \lesssim \tau_k \|P_\star\|_{\mathrm{op}} \log(1/\delta).$$

Point 3: Suppose now that $\tau_k \ge \|P_\star\|_{\mathrm{op}}^3 \mathcal{J}_0 \log(1/\delta) \vee \Psi_{B_\star}^4 \sigma_{\mathrm{in}}^4$. Then, by using the expectation bound statement of Lemma I.1, and summing over inidices i, we have

$$\mathbb{E}[\mathrm{tr}(\mathbf{\Lambda}_k) \cap \mathcal{E}_{\mathrm{bound}} \cap \mathcal{E}_{\mathrm{safe}}] \lesssim d\tau_k \|P_\star\|_{\mathrm{op}}.$$

Hence, by Lemma E.4, if

$$\tau_k \gtrsim_\star d \log \left\{ \frac{d\|P_\star\|_{\mathrm{op}}}{\sigma_{\min}(\Gamma_k)} \right\}.$$

then with probability $1 - e^{-\tau_k/d}$ on $\mathcal{E}_{\mathrm{bound}} \cap \mathcal{E}_{\mathrm{safe}}$, we have that $\mathbf{\Lambda}_k \gtrsim \tau_k \Gamma_k$. Note that since $\tau_k \ge \|P_\star\|_{\mathrm{op}}^3 \mathcal{J}_0 \log(1/\delta) \ge d \log(1/\delta)$ (since $\mathcal{J}_0 \ge d$ and $\|P_\star\|_{\mathrm{op}} \ge 1$), we have also $1 - e^{-\tau_k/d} \ge 1 - \delta$.

Now, if in addition $\tau_k \gtrsim_\star \sigma_{\mathrm{in}}^4 \|P_\star\|_{\mathrm{op}}^2$, Lemma I.4.1 entails that $\Gamma_k \succsim \frac{\sigma_k^2}{\|P_\star\|_{\mathrm{op}}} = \frac{\sigma_{\mathrm{in}}^2}{\sqrt{\tau_k}\|P_\star\|_{\mathrm{op}}}$ (note that for such k, $\sigma_{\mathrm{in}}^2/\sqrt{\tau_k} \le 1$). With a few simplifications, we see then that if

$$\tau_k \gtrsim_\star d \log \tau_k + d \log \left\{ \frac{d\|P_\star\|_{\mathrm{op}}}{1 \wedge \sigma_{\mathrm{in}}^2} \right\} \vee \|P_\star\|_{\mathrm{op}}^3 \mathcal{J}_0 \log(1/\delta) \vee \sigma_{\mathrm{in}}^4 (\max\{\Psi_{B_\star}^4, \|P_\star\|_{\mathrm{op}}^2\}),$$

then with probability $1 - \mathcal{O}(\delta)$, $\mathbf{\Lambda}_k \succsim \tau_k \Gamma_k$. Since $\tau_k \gtrsim_\star d \log \tau_k$ for $\tau_k \gtrsim d \log d$, we need simply $\tau_k \gtrsim_\star d \log \left\{ \frac{d\|P_\star\|_{\mathrm{op}}}{1 \wedge \sigma_{\mathrm{in}}^2} \right\} \vee \|P_\star\|_{\mathrm{op}}^3 \mathcal{J}_0 \log(1/\delta) \vee \sigma_{\mathrm{in}}^4 (\max\{\Psi_{B_\star}^4, \|P_\star\|_{\mathrm{op}}^2\})$ to ensure $\mathbf{\Lambda}_k \succsim \tau_k \Gamma_k$ with probability $1 - \mathcal{O}(\delta)$. Shrinking δ by a constant reduces the failure probability to $1 - \delta$. Lastly, using $\sigma_{\mathrm{in}}^2 \ge 1$ by definition, and $\sigma_{\mathrm{in}}^2 \lesssim \sqrt{d_{\mathbf{x}}} \|P_\star\|_{\mathrm{op}}^{9/2} \Psi_{B_\star} \sqrt{\log \frac{\|P_\star\|_{\mathrm{op}}}{\delta}}$ by Lemma H.1, we can bound

$$\begin{aligned}
&d \log \left\{ \frac{d\|P_\star\|_{\mathrm{op}}}{1 \wedge \sigma_{\mathrm{in}}^2} \right\} \vee \|P_\star\|_{\mathrm{op}}^3 \mathcal{J}_0 \log(1/\delta) \vee \sigma_{\mathrm{in}}^4 (\Psi_{B_\star}^4 \vee \|P_\star\|_{\mathrm{op}}^2) \\
&\lesssim d \log d \|P_\star\|_{\mathrm{op}} + \|P_\star\|_{\mathrm{op}}^3 \mathcal{J}_0 \log(1/\delta) + d_{\mathbf{x}} \|P_\star\|_{\mathrm{op}}^9 \Psi_{B_\star}^2 \log \frac{\|P_\star\|_{\mathrm{op}}}{\delta} (\Psi_{B_\star}^4 \vee \|P_\star\|_{\mathrm{op}}^2) \\
&\lesssim d \left(\|P_\star\|_{\mathrm{op}}^3 \mathcal{P}_0 + \|P_\star\|_{\mathrm{op}}^{11} \Psi_{B_\star}^6 \right) \log \frac{d\|P_\star\|_{\mathrm{op}}}{\delta} := \tau_{\mathrm{ls}}.
\end{aligned}$$

where in the last line we use $\Psi_{B_\star}, \|P_\star\|_{\mathrm{op}} \ge 1$, $d_{\mathbf{x}} \le d$, and $\mathcal{P}_0 = \mathcal{J}_0 / d_{\mathbf{x}}$. □

I.5. Proof of Lemma H.5 ($k < k_{\mathrm{safe}}$)

We analyze the rounds $k < k_{\mathrm{safe}}$, which correspond to the rounds before the least-squares procedure produces a sufficiently close approximation to $(A_\star, B_\star)$ that we can safely implement certainty equivalent control.

In order to avoid directly conditioning on events $\{k_{\mathrm{safe}} \le (\dots)\}$, let us define the sequence $\mathbf{z}_{t,0} := (\mathbf{x}_{t,0}, \mathbf{u}_{t,0})$ on the same probability space as $(\mathbf{x}_t, \mathbf{u}_t)$ to denote the system driven by the same noise $\mathbf{w}_t$, and with the same random perturbations $\mathbf{g}_t$, but where the evolution is with respect to the dynamics

$$\mathbf{x}_{t,0} = A_\star + B_\star \mathbf{u}_{t,0} \quad \mathbf{u}_{t,0} = K_0 \mathbf{x}_{t,0} + \mathbf{g}_t,$$

that is, the dynamics defined by the distribution $\mathcal{D}(K_0, \sigma_u^2 = 1, x_1 = 0)$. Observe that, for any $t < \tau_{k_{\mathrm{safe}}}$, it holds that $\mathbf{x}_{t,0} = \mathbf{x}_t$ and $\mathbf{u}_{t,0} = \mathbf{u}_t$, so it will suffice to reason about this sequence.

Proof that $\mathcal{E}_{\mathrm{safe}}$ holds As above, to reason rigorously about probabilities, we introduce $\widehat{A}_{k,0}$, $\widehat{B}_{k,0}$ as the OLS estimators on the $\mathbf{z}_{k,0} := (\mathbf{x}_{k,0}, \mathbf{u}_{k,0})$ sequence, and define the covariance matrix

$$\mathbf{\Lambda}_{k,0} := \sum_{t=\tau_k}^{2\tau_k - 1} \mathbf{z}_{k,0} \mathbf{z}_{k,0}^\top.$$

We also define the induced confidence term:

$$\mathsf{Conf}_{k,0} = 6 \lambda_{\min}(\mathbf{\Lambda}_{k,0})^{-1} \Big(d \log 5 + \log \Big\{ \tfrac{4k^2 \det(3(\mathbf{\Lambda}_{k,0})}{\delta} \Big\} \Big).$$

Lemma I.5. *The following event holds with probability $1 - \delta$:*

$$\mathcal{E}_{\mathrm{conf}} := \left\{ \forall k \le k_{\mathrm{safe}} \textit{ with } \mathbf{\Lambda}_k \succeq I, \quad \left\| \left[\widehat{A}_k - A_\star \mid \widehat{B}_k - B_\star \right] \right\|_2^2 \le \mathsf{Conf}_k \right\}.$$

Proof. Applying (E.1) in Lemma E.2 with $\Lambda_0 = I$, we see that for any fixed k for which $\mathbf{\Lambda}_{k,0} \succeq I$, $\mathsf{Conf}_{k,0}$ is a valid $\delta/4k^2$-confidence interval; that is $\|[A_\star - \widehat{A}_{k,0} \mid B_\star - \widehat{B}_{k,0}]\|_{\mathrm{op}} \le \mathsf{Conf}_{k,0}$. By a union bound, the confidence intervals are valid with probability $1 - \delta/2$, simultaneously. Since the the sequence $\mathbf{x}_{t,0}$ coincides with $\mathbf{x}_t$ for $t \le \tau_{k_{\mathrm{safe}}}$, and $\mathbf{u}_{t,0}$ with $\mathbf{u}_t$ for $t \le \tau_{k_{\mathrm{safe}}} - 1$, we see that $\mathsf{Conf}_{k,0} = \mathsf{Conf}_k$ for all $k \le k_{\mathrm{safe}}$. □

Proof of Regret Bound We begin with the following regret bound.

Lemma I.6. *For $\delta < 1/T$, the following hold with probability $1-\delta$,*

$$\sum_{t=1}^{\tau_{k_{\mathrm{safe}}}-1} \mathbf{x}_{t,0}^\top R_{\mathbf{x}} \mathbf{x}_{t,0} + \mathbf{u}_{t,0}^\top R_{\mathbf{u}} \mathbf{u}_{t,0} \lesssim d\tau_{k_{\mathrm{safe}}} \Psi_{B_\star}^2 \mathcal{P}_0 \log(\frac{1}{\delta}).$$

Proof. It suffices to show that the $(\mathbf{x}_{t,0}, \mathbf{u}_{t,0})$ sequences satisfies the following bound:

$$\sum_{t=1}^{\tau_{k_0}-1} \mathbf{x}_{t,0}^\top R_{\mathbf{x}} \mathbf{x}_{t,0} + \mathbf{u}_{t,0}^\top R_{\mathbf{u}} \mathbf{u}_{t,0} \lesssim \tau_{k_0} \left(\mathcal{J}_0(1+\|B_\star\|_{\mathrm{op}}^2) + \mathrm{tr}(R_{\mathbf{u}})\right) \log(\frac{1}{\delta}),$$

where the inequality suffices since $\mathcal{P}_0 \geq 1$ (indeed, $\mathcal{J}_0 \geq \mathcal{J}_\star \geq d$ by Lemma B.6), and thus $\mathcal{J}_0(1+\|B_\star\|_{\mathrm{op}}^2) + \mathrm{tr}(R_{\mathbf{u}}) = d_{\mathbf{x}}\mathcal{P}_0(1+\|B_\star\|_{\mathrm{op}}^2) + d_{\mathbf{u}} \leq d\mathcal{P}_0\Psi_\star$.

For the second, we have from Lemma I.9 and the fact that $\mathbf{x}_1 = 0$ that there is a Gaussian quadratic form $\overline{\mathbf{g}}^\top \Lambda_{\overline{\mathbf{g}}} \overline{\mathbf{g}}$ which is equal to $\sum_{t=1}^{\tau_{k_0}-1} \mathbf{x}_{t,0}^\top R_{\mathbf{x}} \mathbf{x}_{t,0}$, and where $\mathrm{tr}(\Lambda_{\overline{\mathbf{g}}}) \leq \tau_{k_0}\left(\mathcal{J}_0(1+\|B_\star\|_{\mathrm{op}}^2) + \mathrm{tr}(R_{\mathbf{u}})\right)$. The second bound now follows from the crude statement of Hanson Wright in Corollary 5. The last statement follows by a union bound, noting that we need to bound over $k_{\max} = \log_2 T \leq T \leq 1/\delta$, rounds, and absorbing constants. □

We conclude by arguing an upper bound on $\tau_{k_{\mathrm{safe}}}$. We rely on the following guarantee.

Lemma I.7. *Suppose $\mathcal{E}_{\mathrm{safe}}$ holds. Then for all $k < k_{\mathrm{safe}}$ for which $\mathbf{\Lambda}_k \succeq I$, we must have that $\mathsf{Conf}_k \gtrsim \epsilon_{\mathrm{safe}}$, where $\epsilon_{\mathrm{safe}} = \|P_\star\|_{\mathrm{op}}^{-10}$.*

Proof. For all $k < k_{\mathrm{safe}}$ for which $\mathbf{\Lambda}_k \succeq I$, we must have that $\mathsf{Conf}_k > 1/C_{\mathrm{safe}}(\widehat{A}_k, \widehat{B}_k)\}$. If $\mathsf{Conf}_k \leq c/C_{\mathrm{safe}}(A_\star, B_\star)^2$ for a sufficiently small c, then the same perturbation argument as in Theorem 11 entails that we have $\mathsf{Conf}_k \leq 1/9C_{\mathrm{safe}}(\widehat{A}_k, \widehat{B}_k)^2$, yielding a contradiction. Finally, we subsitute in $C_{\mathrm{safe}}(A_\star, B_\star)^2 \lesssim \|P_\star\|_{\mathrm{op}}^{10}$ by Equation (3.1). □

Recall that we say $f \gtrsim_\star f$ if "$f \geq Cg$" for a sufficiently large constant C (Definition I.3). In light of the above lemma, Part 2 will follow as soon as we can show that, for any $\epsilon \in (0,1)$,,

$$\text{if } \tau_k \gtrsim_\star \frac{d(1+\|K_0\|_{\mathrm{op}}^2)}{\epsilon} \log \frac{\Psi_{B_\star}^2 \mathcal{J}_0}{\delta}, \quad \text{then } \mathsf{Conf}_{k,0} \leq \epsilon, \text{ and } \mathbf{\Lambda}_{k,0} \succeq I \text{ w.p. } 1 - \mathcal{O}(\delta). \tag{I.3}$$

We begin with a lower bound the matrices $\mathbf{\Lambda}_{k,0}$:

Lemma I.8. *for a sufficiently large constant C. Finally, set $\tau_{\min} = d\log(1+\Psi_{B_\star}\mathcal{J}_0)$. Then, for any k such that $\tau_k \gtrsim_\star \tau_{\min} \vee d\log(\frac{1}{\delta})$, , it holds that*

$$\mathbb{E}[\mathrm{tr}(\mathbf{\Lambda}_{k,0})] \lesssim \Psi_{B_\star}^2 \mathcal{J}_0 \tau_k, \quad \mathbb{P}\left[\lambda_{\min}(\mathbf{\Lambda}_{k,0}) \gtrsim_\star \frac{\tau_k}{1+\|K_0\|_2^2}\right] \leq \delta.$$

The bound above is proven in Section I.5.1. We can now verify Eq. (I.3), concluding the proof of Part 2.

Proof of Eq. (I.3). Suppose that k is such that $\tau_k \gtrsim_\star \tau_{\min} \vee d\log(\frac{1}{\delta})$. Then, by the above lemma, and using $\det(cX) = c^d \det(X)$ for $X \in \mathbb{R}^{d\times d}$, we have, with probability $1 - \mathcal{O}(\delta)$,

$$\begin{aligned}
\mathsf{Conf}_{k,0} &\lesssim \frac{1+\|K_0\|_2^2}{\tau_k}(d + \log\frac{k^2}{\delta} + \log\det((\mathbf{\Lambda}_{k,0})) \\
&\leq \frac{1+\|K_0\|_2^2}{\tau_k}(d + \log\frac{k^2}{\delta} + d\log\mathrm{tr}((\mathbf{\Lambda}_{k,0})),
\end{aligned}$$

where we use that $X \succeq 0$, we have $\log\det(X) = \sum_{i=1}^d \log\lambda_i(X) \le d\log\mathrm{tr}(X)$. By Markov's inequality, we have with probability $1-\delta$ that $\mathrm{tr}((\mathbf{\Lambda}_{k,0}) \le \mathbb{E}[\mathrm{tr}((\mathbf{\Lambda}_{k,0})]/\delta \lesssim \Psi_{B_\star}^2 \mathcal{J}_0 \tau_k \le \Psi_{B_\star}^2 \mathcal{J}_0/\delta$, since $\tau_k \le T \le 1/\delta$. Hence, with some elementary operators, we can bound

$$\mathsf{Conf}_{k,0} \lesssim \frac{d}{\tau_k}\log\frac{\Psi_{B_\star}^2\mathcal{J}_0}{\delta}.$$

Hence, for $\tau_k \gtrsim_\star \frac{d}{\epsilon}\log\frac{\Psi_{B_\star}^2\mathcal{J}_0}{\delta}$, we have with probability $1-\delta$ that we have $\mathsf{Conf}_k \le \epsilon$. □

I.5.1. Proof of Lemma I.8

1. We first need to argue a lower bound on matrices Σ_t such that that $\mathbf{z}_{t,0} \mid \mathcal{F}_{t-1} \sim \mathcal{N}(\overline{\mathbf{z}}_{t,0}, \Sigma_{t,0})$, where $\overline{\mathbf{z}}_{t,0}, \Sigma_{t,0}$ are $\mathcal{F}_{t-1}$ measurable. It is straightforward to show that

$$\Sigma_{t,0} = \begin{bmatrix} I & K_0 \\ K_0^\top & K_0^\top K_0 + I \end{bmatrix},$$

which by Dean et al. (2018, Lemma F.6), has least singular value bounded below as

$$\lambda_{\min}(\Sigma_{t,0}) \ge \min\left\{\frac{1}{2}, \frac{1}{1+2\|K_0\|^2}\right\} \ge \frac{1}{2+2\|K_0\|^2}.$$

2. Next, we need an upper bound on

$$\begin{aligned}
\mathbb{E}[\mathrm{tr}(\mathbf{\Lambda}_{k,0})] &= \mathbb{E}[\sum_{t=\tau_{k-1}}^{\tau_k-1} \|\mathbf{x}_t\|^2 + \|\mathbf{z}_t\|^2] \\
&\le \mathbb{E}[\sum_{t=1}^{\tau_k-1} \|\mathbf{x}_t\|^2 + \|\mathbf{z}_t\|^2] \\
&\le \tau_k(1+\|B_\star\|^2)\mathrm{tr}(\mathsf{dlyap}(A_{K_0}, I + K_0^\top K_0)) + \mathrm{tr}\tau_k(I) \\
&\le 2\tau_k(1+\|B_\star\|^2)\mathrm{tr}(\mathsf{dlyap}(A_{K_0}, I + K_0^\top K_0)) \\
&\le 2\tau_k(1+\|B_\star\|^2)J_{K_0} = 4(\tau_k - \tau_{k-1})(1+\|B_\star\|^2)J_{K_0} \\
&\le 4(\tau_k - \tau_{k-1})(1+\Psi_{B_\star}^2)J_{K_0} \lesssim \tau_k\Psi_{B_\star}^2 J_{K_0},
\end{aligned}$$

where we use that $I \preceq \mathsf{dlyap}(A_{K_0}, I + K_0^\top K_0)) \preceq \mathsf{dlyap}(A_{K_0}, R_{\mathbf{x}} + K_0^\top R_{\mathbf{u}} K_0)) = J_{K_0}$ for $R_{\mathbf{u}}, R_{\mathbf{x}} \ge I$. This proves the trace upper bound.

3. Using the second to last inequality in the above display, we see that for

$$\tau_k - \tau_{k-1} = \frac{1}{2}\tau_k \ge \underbrace{\frac{2000}{9}\left(2d\log\tfrac{100}{3} + d\log(8(1+1\|K_0\|)^2(1+\Psi_{B_\star}^2)J_{K_0})\right)}_{:=\underline{\tau}},$$

Lemma E.4 implies (taking $\mathcal{E} = \Omega$ to be the probability space and $T = \tau_k/2$) that, if $\tau_k \gtrsim_\star \tau_{\min}$, we have

$$\mathbb{P}\left[\mathbf{\Lambda}_{k_0} \not\succeq \frac{9\tau_k}{3200}\Sigma_0\right] \le 2\exp\left(-\tfrac{9}{4000(d+1)}\tau_k\right).$$

Routine manipulations of give dlyap, $1 + \|K_0\|^2 \le \mathsf{dlyap}(A_{K_0}, I + \|K_0\|^2) \le \mathsf{dlyap}(A_{K_0}, R_{\mathbf{x}} + K_0^\top R_{\mathbf{u}} K_0) = J_{K_0}$ for $R_{\mathbf{u}}, R_{\mathbf{x}} \succeq I$. Hence, with a bit of algebra, we can bound

$$\underline{\tau} \lesssim \tau_{\min} := d\log(1 + \Psi_{B_\star} J_{K_0}).$$

Using the lower bound on Σ_0 concludes the proof.

□

I.6. Proof of Lemma I.1

In order to prove Lemma I.1, we first show that we can represent the Cost functional as a quadratic form in Gaussian variables.

Lemma I.9. *Let $(\mathbf{x}_1, \mathbf{x}_2, \sigma_u)$ denote the linear dynamical system described by the evolution of $\mathcal{D}(K, x_1)$. Then for any $t \geq 1$, there exists a standard Gaussian? vector $\overline{\mathbf{g}} \in \mathbb{R}^{\mathcal{O}(td)}$ such that for any cost matrices $R_1, R_2 \succeq 0$, we have*

$$\mathsf{Cost}(R_1, R_2; x_1, \sigma_u, t) = \overline{\mathbf{g}}^\top \Lambda_{\overline{\mathbf{g}}} \overline{\mathbf{g}} + x_1^\top \Lambda_{\mathbf{x}_1} x_1 + 2\overline{\mathbf{g}}^\top \Lambda_{\mathrm{cross}} x_1,$$

where, letting $R_K = R_1 + K^\top R_2 K$, $A_K = A_\star + B_\star K$, $P_K = \mathsf{dlyap}(A_K, R_K)$ $J_K := \mathrm{tr}(P_K)$, and $d_{\mathrm{eff}} := \min\{d_{\mathbf{u}}, \dim(R_1) + \dim(R_2)\}$,

$$\begin{aligned}
&\mathrm{tr}(\Lambda_{\overline{\mathbf{g}}}) \leq tJ_k + 2\sigma_u^2 t d_{\mathrm{eff}} \left(\|R_2\|_{\mathrm{op}} + \|B_\star\|_{\mathrm{op}}^2 \|P_K\|_{\mathrm{op}}\right), \\
&\|\Lambda_{\overline{\mathbf{g}}}\|_{\mathrm{op}} \lesssim (1 + \sigma_u^2 \|B_\star\|_{\mathrm{op}}^2)\|R_K\|_{\mathrm{op}} \|A_K\|_{\mathcal{H}_\infty}^2 + \sigma_u^2 \|R_2\|_{\mathrm{op}}^2, \\
&\Lambda_{\mathbf{x}_1} \preceq P_K, \\
&\|\Lambda_{\mathrm{cross}} x_1\|_2 \leq \sqrt{\|\Lambda_{\overline{\mathbf{g}}}\|_{\mathrm{op}} \cdot x_1^\top P_K x}.
\end{aligned}$$

Let us continue to prove Lemma I.1. The expectation result follow since $\mathbb{E}[\mathsf{Cost}(R_1, R_2; x_1, \sigma_u, t)] = \mathrm{tr}(\Lambda_{\overline{\mathbf{g}}}) + x_1^\top \Lambda_{\mathbf{x}_1} x_1$ for a Gaussian quadratic form.

For the high probability result, observe that by Gaussian concentration and Lemma I.9, we have with probability $1 - \delta$

$$2\overline{\mathbf{g}}^\top \Lambda_{\mathrm{cross}} x_1, \lesssim \sqrt{\log(1/\delta)\|\Lambda_{\mathrm{cross}} x_1\|_2} \lesssim \sqrt{\log(1/\delta)\|\Lambda_{\overline{\mathbf{g}}}\|_{\mathrm{op}} \cdot x_1^\top P_K x_1}.$$

Hence, by AM-GM, $2\overline{\mathbf{g}}^\top \Lambda_{\mathrm{cross}} x_1 \leq \mathcal{O}\left(\log(1/\delta)\|\Lambda_{\overline{\mathbf{g}}}\|_{\mathrm{op}}\right) + x_1^\top P_K x_1$. On the other hand, by Hanson-Wright

$$\begin{aligned}
\overline{\mathbf{g}}^\top \Lambda_{\overline{\mathbf{g}}} \overline{\mathbf{g}} &\leq \mathrm{tr}(\mathrm{tr}(\Lambda_{\overline{\mathbf{g}}})) + \mathcal{O}\left(\|\Lambda_{\overline{\mathbf{g}}}\|_{\mathrm{F}} \sqrt{\log(1/\delta)} + \|\Lambda_{\overline{\mathbf{g}}}\|_{\mathrm{op}} \log(1/\delta)\right) \\
&\leq \mathrm{tr}(\Lambda_{\overline{\mathbf{g}}}) + \mathcal{O}\left(\sqrt{td\log(1/\delta)} + \|\Lambda_{\overline{\mathbf{g}}}\|_{\mathrm{op}} \log(1/\delta)\right),
\end{aligned} \tag{I.4}$$

where we use the dimension of $\Lambda_{\overline{\mathbf{g}}}$ in the last line. Combining with the previous result, and adding in $x_1^\top \Lambda_{\mathbf{x}_1} x_1 \leq x_1^\top P_K x_1$, we have that with probability $1 - \delta$,

$$\mathsf{Cost}(R_1, R_2; x_1, \sigma_u, t) \leq \mathrm{tr}(\Lambda_{\overline{\mathbf{g}}}) + \mathcal{O}\left((\sqrt{td\log(1/\delta)} + \log(1/\delta))\|\Lambda_{\overline{\mathbf{g}}}\|_{\mathrm{op}} + x_1^\top P_K x_1\right).$$

The first high-probability statement follows by substituing in $\mathrm{tr}(\Lambda_{\overline{\mathbf{g}}})$ and $\|\Lambda_{\overline{\mathbf{g}}}\|_{\mathrm{op}}$. Then second statement follows from returning to Eq. I.4 and using $\|X\|_{\mathrm{op}}, \|X\|_{\mathrm{F}} \leq \mathrm{tr}(X)$ for $X \succeq 0$. □

We shall now prove Lemma I.9, but first, we establish some useful preliminaries.

I.6.1. Linear Algebra Preliminaries

Definition I.4 (Toeplitz Operator)**.** For $\ell \in \mathbb{N}$, and $j, \ell \geq i$, define the matrices

$$\mathsf{Toep}_{i,j,\ell}(A) := \begin{bmatrix} A^i \mathbb{I}_{i \geq 0} & A^{i+1} \mathbb{I}_{i \geq -1} & \dots & A^{i+\ell} \mathbb{I}_{i \geq -\ell} \\ A^{i-1} \mathbb{I}_{i \geq 1} & A^i \mathbb{I}_{i \geq 0} & \dots & A^{i+\ell-1} \mathbb{I}_{i \geq 1-\ell} \\ \dots & \dots & \dots & \dots \\ A^{i-j} \mathbb{I}_{i \geq j} & \dots & \dots & A^{i+\ell-j} \mathbb{I}_{i+\ell-j \geq 0} \end{bmatrix}, \quad \mathsf{ToepCol}_{i,j}(A) := \begin{bmatrix} A^{j-1} \\ A^{j-2} \\ \dots \\ \mathbb{I}_{i \geq 1} A^{i-1} \end{bmatrix}.$$

We shall use the following lemma.

Lemma I.10. *For any $i \leq j, \ell$, we have $\|\mathsf{ToepCol}_{i,j}\|_{\mathrm{op}} \leq \|\mathsf{Toep}_{i,j,\ell}(A)\|_{\mathrm{op}} \leq \|A\|_{\mathcal{H}_\infty}$, and, for $Y \in \mathbb{R}^{d_{\mathbf{x}}^2}$, and $\mathrm{diag}_{j-i}(Y)$ denoting a $j - i$-block block matrix with blocks Y on the diagonal, we have the bound*

$$\mathrm{tr}(\mathsf{ToepCol}_{i,j}(A)^\top \mathrm{diag}_{j-i}(Y) \mathsf{ToepCol}_{i,j}(A)) \preceq (j - i) \cdot \mathrm{tr}(\mathsf{dlyap}(Y, A))$$

Proof. The first bound is a consequence of the fact that $\mathsf{Toep}_{i,j,\ell}(A)$ is a submatrix of the infinite-dimensional linear operator mapping inputs sequences in $\ell_2(\mathbb{R}^{d_{\mathbf{x}}})$ to outputs $\ell_2(\mathbb{R}^{d_{\mathbf{x}}})$; thus, the operator norm of $\mathsf{Toep}_{i,j,\ell}(A)$ is bounded by the operator norm of this infinite dimensional linear operator, which is equal to $\|A\|_{\mathcal{H}_\infty}$(see e.g. Tilli (1998, Corollary 4.2)). The second bound follows from direct computation, as

$$\mathrm{tr}(\mathsf{ToepCol}_{i,j}(A)^\top \mathrm{diag}_{j-i}(Y)\mathsf{ToepCol}_{i,j}(A)) \le \mathrm{tr}(\sum_{s=0}^{\infty} A^\top Y A) = \mathrm{tr}(\mathsf{dlyap}(A, Y)).$$

□

I.6.2. Proof of Lemma I.9

Lemma I.11 (Form of the Covariates)**.** *Introduce the vector* $\mathbf{x}_{[t]} = (\mathbf{x}_t, \dots, \mathbf{x}_1)$ *and* $\mathbf{u}_{[t]} := (\mathbf{u}_t, \dots, \mathbf{u}_1)$*, set* $\overline{\mathbf{w}}_{[t-1]} = (\mathbf{w}_{t-1}, \dots, \mathbf{w}_1)$ *and* $\mathbf{g}_{[t]} = (\mathbf{g}_t, \dots, \mathbf{g}_1)$*. Then, we can write*

$$\begin{bmatrix} \mathbf{x}_{[t]} \\ \mathbf{u}_{[t]} \end{bmatrix} = M_{K,t} \begin{bmatrix} \mathbf{w}_{[t-1]} \\ \mathbf{g}_{[t]} \end{bmatrix} + \underbrace{\begin{bmatrix} I_t \\ \mathrm{diag}_t(K) \end{bmatrix} \mathsf{ToepCol}_{1,t}(0)}_{:=M_{0,t}} \mathbf{x}_1,$$

where we have defined the matrix

$$M_{K,t} = \begin{bmatrix} \mathsf{Toep}_{0,t,t-1}(A_K) & \sigma_u \mathsf{Toep}_{-1,t,t}(A_K)\mathrm{diag}_t(B_\star) \\ K\mathsf{Toep}_{0,t,t-1}(A_K) & \sigma_u \mathrm{diag}_t(I) + \sigma_u K \mathsf{Toep}_{-1,t,t}(A_K)\mathrm{diag}_t(B_\star) \end{bmatrix}.$$

Further, let

$$\left[\frac{A}{B}\right]_{\mathrm{diag}} := \begin{bmatrix} A & 0 \\ 0 & B \end{bmatrix}.$$

In light of the the above lemma, we have for $\overline{\mathbf{g}} := \begin{bmatrix} \mathbf{w}_{[t-1]} \\ \mathbf{g}_{[t]} \end{bmatrix}$, we have that

$$\begin{aligned}
&\sum_{s=1}^{t} \mathbf{x}_s^\top R_1 \mathbf{x}_s + \mathbf{u}_s^\top R_2 \mathbf{u}_s \\
&= \begin{bmatrix} \mathbf{x}_{[t]} \\ \mathbf{u}_{[t]} \end{bmatrix}^\top \left[\frac{\mathrm{diag}(R_1)}{\mathrm{diag}(R_2)}\right]_{\mathrm{diag}} \begin{bmatrix} \mathbf{x}_{[t]} \\ \mathbf{u}_{[t]} \end{bmatrix} \\
&= (M_{K,t}\overline{\mathbf{g}}_t + M_{0,t}\mathbf{x}_1)^\top \left[\frac{\mathrm{diag}(R_1)}{\mathrm{diag}(R_2)}\right]_{\mathrm{diag}} (M_{K,t}\overline{\mathbf{g}}_t + M_{0,t}\mathbf{x}_1) \\
&= \overline{\mathbf{g}}_t^\top \underbrace{M_{K,t}^\top \left[\frac{\mathrm{diag}(R_1)}{\mathrm{diag}(R_2)}\right]_{\mathrm{diag}} M_{K,t}}_{:=\Lambda_{\overline{\mathbf{g}}}} \overline{\mathbf{g}}_t + 2\mathbf{x}_1^\top \underbrace{M_{0,t}^\top \left[\frac{\mathrm{diag}(R_1)}{\mathrm{diag}(R_2)}\right]_{\mathrm{diag}} M_{K,t}}_{:=\Lambda_{\mathrm{cross}}} \overline{\mathbf{g}}_t \\
&\quad + \mathbf{x}_1^\top \underbrace{M_{0,t}^\top \left[\frac{\mathrm{diag}(R_1)}{\mathrm{diag}(R_2)}\right]_{\mathrm{diag}} M_{0,t}}_{:=\Lambda_{\mathbf{x}_1}} \mathbf{x}_1.
\end{aligned}$$

We can evaluate each term separately.

Bounding $\mathrm{tr}(\Lambda_{\overline{\mathbf{g}}})$**.** Let us recall

$$M_{K,t} = \begin{bmatrix} \mathsf{Toep}_{0,t,t-1}(A_K) & \sigma_u \mathsf{Toep}_{-1,t,t}(A_K)\mathrm{diag}_t(B_\star) \\ K\mathsf{Toep}_{0,t,t-1}(A_K) & \sigma_u \mathrm{diag}_t(I) + \sigma_u K \mathsf{Toep}_{-1,t,t}(A_K)\mathrm{diag}_t(B_\star) \end{bmatrix}.$$

Recall $R_K := R_1 + K^\top R_2 K$. We find that the diagonal terms of $\Lambda_{\overline{\mathbf{g}}}$ coincinde with the diagonals of the matrix $\Lambda_{\overline{\mathbf{g}},\mathrm{diag}}$ defined as

$$\left[\frac{\mathsf{Toep}_{0,t,t-1}(A_K)^\top \mathrm{diag}_{t-1}(R_K)\mathsf{Toep}_{0,t,t-1}(A_K)}{\sigma_u^2\mathrm{diag}_t(B_\star)^\top \mathsf{Toep}_{-1,t,t}(A_K)^\top \mathrm{diag}_t(R_K)\mathsf{Toep}_{-1,t,t}(A_K)\mathrm{diag}_t(B_\star) + \sigma_u^2\mathrm{diag}_t(R_2) + (\text{cross term})}\right]_{\mathrm{diag}}$$
$$\preceq \left[\frac{\mathsf{Toep}_{0,t,t-1}(A_K)^\top \mathrm{diag}_{t-1}(R_K)\mathsf{Toep}_{0,t,t-1}(A_K)}{2\sigma_u^2\mathrm{diag}_t(B_\star)^\top \mathsf{Toep}_{-1,t,t}(A_K)^\top \mathrm{diag}_t(R_K)\mathsf{Toep}_{-1,t,t}(A_K)\mathrm{diag}_t(B_\star) + 2\sigma_u^2\mathrm{diag}_t(R_2)}\right]_{\mathrm{diag}},$$

where (cross term) denotes the cross term between the term $\sigma_u^2\mathrm{diag}_t(B_\star)^\top \mathsf{Toep}_{-1,t,t}(A_K)^\top \mathrm{diag}_t(R_K)\mathsf{Toep}_{-t,t,t}(A_K)\mathrm{diag}_t(B_\star)+\sigma_u^2\mathrm{diag}_t(R_2)$, which we bound in the second inequality by Young's inequality.

By Lemma I.10, we have

$$\mathrm{tr}(\mathsf{Toep}_{0,t,t-1}(A_K)^\top \mathrm{diag}_{t-1}(R_K)\mathsf{Toep}_{0,t,t-1}(A_K)) \le t\cdot \mathrm{tr}(\mathsf{dlyap}(A_K,R_K)) = J_K.$$

Similarly, since $\mathsf{dlyap}(A_K,R_K) = P_K$, and thus $\mathrm{rank}(P_K)\le \mathrm{rank}(R_K)\le \mathrm{rank}(R_1)+\mathrm{rank}(R_2)$,

$$\begin{aligned}
&\mathrm{tr}(\mathrm{diag}_t(B_\star)^\top \mathsf{Toep}_{-1,t,t}(A_K))^\top \mathrm{diag}_t(R_K)\mathsf{Toep}_{-1,t,t}(A_K))\mathrm{diag}_t(B_\star))\\
&\le t\cdot \mathrm{tr}(B_\star^\top \mathsf{dlyap}(A_K,R_K)B_\star)\\
&= t\cdot \mathrm{tr}(B_\star^\top P_K B_\star)\\
&\le t\|B_\star\|_{\mathrm{op}}^2\|P_\star\|_{\mathrm{op}}\min\{\mathrm{rank}(B_\star),\mathrm{rank}(P_K)\} \le t d_{\mathrm{eff}}\|B_\star\|_{\mathrm{op}}^2\|P_\star\|_{\mathrm{op}}.
\end{aligned}$$

Finally, we can bound $\mathrm{tr}(2\sigma_u^2\mathrm{diag}_t(R_2)) \le 2t\sigma_u^2\mathrm{rank}(R_2)\|R_2\|_{\mathrm{op}} \le 2d_{\mathrm{eff}}t\sigma_u^2\|R_2\|_{\mathrm{op}}$, yielding

$$\mathrm{tr}(\Lambda_{\overline{\mathbf{g}}}) = \mathrm{tr}(\Lambda_{\overline{\mathbf{g}},\mathrm{diag}}) \le tJ_k + 2\sigma_u^2 t d_{\mathrm{eff}}\left(\|R_2\|_{\mathrm{op}} + \|B_\star\|_{\mathrm{op}}^2\|P_K\|_{\mathrm{op}}\right).$$

Bounding $\|\Lambda_{\overline{\mathbf{g}}}\|_{\mathrm{op}}$. Observe that, for any PSD matrix $M = \begin{bmatrix} A & X \\ X^\top & B\end{bmatrix}$, we have that

$$M \preceq 2\left[\frac{A}{B}\right]_{\mathrm{diag}}.$$

Since $\Lambda_{\overline{\mathbf{g}}} \succeq 0$ (it is a non-negative form), in particular, we hae $\Lambda_{\overline{\mathbf{g}}} \preceq 2\Lambda_{\overline{\mathbf{g}},\mathrm{diag}}$. Thus

$$\begin{aligned}
\|\Lambda_{\overline{\mathbf{g}}}\|_{\mathrm{op}} &\lesssim \|\Lambda_{\overline{\mathbf{g}},\mathrm{diag}}\|_{\mathrm{op}}\\
&\lesssim \sigma_u^2\left(\|R_2\|_{\mathrm{op}}^2 + \|R_K\|_{\mathrm{op}}\|B_\star\|_{\mathrm{op}}^2\|\|\mathsf{Toep}_{-1,t,t}(A_K)\|_{\mathrm{op}}^2\right) + \|R_K\|_{\mathrm{op}}\|\mathsf{Toep}_{0,t,t-1}(A_K)\|_{\mathrm{op}}^2.
\end{aligned}$$

Since we can bound $\|\mathsf{Toep}_{-1,t,t}(A_K)\|_{\mathrm{op}}^2 \le \|A_K\|_{\mathcal{H}_\infty}$ by Lemma I.10, we obtain

$$\|\Lambda_{\overline{\mathbf{g}}}\|_{\mathrm{op}} \lesssim \|R_K\|\|A_K\|_{\mathcal{H}_\infty}^2 + \sigma_u^2\left(\|R_2\|_{\mathrm{op}} + \|R_K\|_{\mathrm{op}}\|B_\star\|_2^2\|A_K\|_{\mathcal{H}_\infty}^2\right),$$

where we use that $\sigma_u \le 1$.

Bounding $\Lambda_{\mathbf{x}_1}$. Let us recall that

$$M_{0,t} := \begin{bmatrix} I_t \\ \mathrm{diag}_t(K)\end{bmatrix}\mathsf{ToepCol}_{1,t}(A_K).$$

Thus,

$$\begin{aligned}
\Lambda_{\mathbf{x}_1} = M_{0,t}^\top\left[\frac{\mathrm{diag}(R_1)}{\mathrm{diag}(R_2)}\right]_{\mathrm{diag}}M_{0,t} &= \mathsf{ToepCol}_{1,t}(A_K)^\top \mathrm{diag}_t(R_1 + K^\top R_2 K)\mathsf{ToepCol}_{1,t}(A_K)\\
&= \mathsf{ToepCol}_{1,t}(A_K)^\top \mathrm{diag}_t(R_K)\mathsf{ToepCol}_{1,t}(A_K)\\
&\preceq \mathsf{dlyap}(A_K,R_K) = P_K.
\end{aligned}$$

Bounding Λ_{cross}. We can directly verify that there exists a matrix A with $AA^\top = \Lambda_{\overline{\mathbf{g}}}$ and a matrix B with $BB^\top = \Lambda_{\mathbf{x}_1}$ such that $\Lambda_{\mathrm{cross}} = 2AB^\top$. Hence,

$$\|\Lambda_{\mathrm{cross}} x_1\|_{\mathrm{op}} \le \sqrt{\|\Lambda_{\overline{\mathbf{g}}}\|_{\mathrm{op}} \cdot x_1^\top \Lambda_{\mathbf{x}_1} x_1} \le \sqrt{\|\Lambda_{\overline{\mathbf{g}}}\|_{\mathrm{op}} \cdot x_1^\top P_K x_1}.$$

I.7. Proof of Lemma I.2

Set $\overline{\mathbf{g}} = \begin{bmatrix} \mathbf{w}_{[t-1]} \\ \mathbf{g}_{[t-1]} \end{bmatrix}$. Then we have

$$\mathbf{x}_t - A_K^{t-1}\mathbf{x}_1 = \mathsf{ToepCol}_{1,t-1}(A_K)\mathbf{w}_{[t-1]} + \sigma_u \mathsf{ToepCol}_{1,t-1}(A_K)\mathrm{diag}(B_\star)\mathbf{g}_{[t]}\mathbf{g}_{[t]}.$$

We now observe that

$$\begin{aligned}
\mathbb{E}[\|\mathbf{x}_t - A_K^{t-1}\mathbf{x}_1\|_2^2 \mid \mathbf{x}_1] &= \mathrm{tr}(\mathsf{ToepCol}_{1,t-1}(A_K)\mathsf{ToepCol}_{1,t-1}(A_K)^\top) \\
&\quad + \sigma_u^2 \mathrm{tr}(\mathrm{diag}_{t-1}(B_\star^\top)\mathsf{ToepCol}_{1,t-1}(A_K)\mathsf{ToepCol}_{1,t-1}(A_K)^\top \mathrm{diag}_{t-1}(B_\star)) \\
&\le (1+\sigma_u^2\|B_\star\|_2^2)\mathrm{tr}(\mathsf{ToepCol}_{1,t-1}(A_K)\mathsf{ToepCol}_{1,t-1}(A_K)^\top) \\
&\le (1+\sigma_u^2\|B_\star\|_2^2)\|\mathsf{ToepCol}_{1,t-1}(A_K)\|_{\mathrm{F}}^2 \\
&\le (1+\sigma_u^2\|B_\star\|_2^2)\mathrm{tr}(\mathsf{dlyap}(A_K, I)) \\
&\le (1+\sigma_u^2\|B_\star\|_2^2)J_K,
\end{aligned}$$

where the last inequality uses Lemma B.5. Since $\mathbf{x}_t - A_K^{t-1}\mathbf{x}_1$ is a Gaussian quadratic form, the simplified Hanson Wright inequality (Corollary 5) gives

$$\|\mathbf{x}_t - A_K^{t-1}\mathbf{x}_1\|_2^2 \lesssim (1+\sigma_u^2\|B_\star\|_2^2)J_K \log\frac{1}{\delta}.$$

□

I.8. Extension to General Noise Models

Our upper bounds hold for general noise distributions with the following properties:

1. The noise satisfies a Hanson-Wright style inequality, so that an analogue of Lemma I.1 holds. Recall that Lemma I.1 establishes that the true costs concentrate around their expectations.
2. The noise process is a σ_+-sub-Gaussian martingale difference sequence, in the sense that $\mathbb{E}[\mathbf{w}_t \mid \mathbf{w}_1, \ldots, \mathbf{w}_{t-1}] = 0$ and for any $v \in \mathbb{R}^{d_{\mathbf{x}}}$, $\mathbb{E}[\exp(\langle v, \mathbf{w}_t\rangle \mid \mathbf{w}_1, \ldots, \mathbf{w}_{t-1}] \le \exp(\frac{1}{2}\|v\|^2\sigma_+^2)$. This is necessary for the self-normalized tail bound (Lemma E.1 of Abbasi-Yadkori et al. (2011)).
3. The noise satisfies the block-martingale small ball condition from Simchowitz et al. (2018), which ensures the covariates are well-conditioned during the estimation phase (in particular, that an analogue of Lemma E.4 holds)

In more detail, suppose that the noise is σ_+-sub-Gaussian, and that $\mathbb{E}[\mathbf{w}_t\mathbf{w}_t^\top \mid \mathbf{w}_1, \ldots, \mathbf{w}_{t-1}] \succeq \Sigma_- \succ 0$. Then by applying the Paley-Zygmund inequality (analogously to Eq. 3.12 in (Simchowitz et al., 2018)), one can show that the $(1, \frac{1}{2}\Sigma_-, p)-$block-martingale small-ball property holds with

$$\begin{aligned}
p &= \frac{1}{4} \cdot \min_{v \ne 0} \frac{\mathbb{E}[\langle \mathbf{w}_t, z\rangle^2 \mid \mathbf{w}_{1:t-1}]^2}{\mathbb{E}[\langle \mathbf{w}_t, z\rangle^4 \mid \mathbf{w}_{1:t-1}]} \\
&\gtrsim \frac{\lambda_{\min}(\Sigma_-)^2}{\sigma_+^4},
\end{aligned}$$

where in the last inequality, we upper bound $\mathbb{E}[\langle \mathbf{w}_t, z\rangle^4]$ using the standard moment bound for sub-Gaussian variables. Hence, a sub-Gaussian upper bound and covariance lower bound are enough to guarantee point 3 above holds.

Point 1 is more delicate, because Hanson-Wright inequalities are known under only restrictive assumptions: namely, for vectors which have independent sub-Gaussian coordinates (Rudelson & Vershynin, 2013), or for those satisfying a Lipschitz-concentration property (Adamczak, 2015). For the first condition to be satisfied, we need to assume that there exists a matrix $\Sigma_+ \succ 0$ such that the vectors $\tilde{\mathbf{w}}_t := \Sigma_+^{-1/2}\mathbf{w}_t$ are (a) jointly independent, and (b) have jointly independent, sub-Gaussian coordinates. For the second condition to hold, we must assume that the concatenated vectors $(\tilde{\mathbf{w}}_1, \dots, \tilde{\mathbf{w}}_t)$ satisfy the Lipschitz-concentration property (Adamczak, 2015, Definition 2.1). If either condition holds, then we can obtain the same regret as in our main theorem by modifying Lemma I.9 to use a quadratic form for the sequence $(\tilde{\mathbf{w}}_1, \dots, \tilde{\mathbf{w}}_t)$, and then applying one of the Hanson-Wright variants above to attain Lemma I.1.

In general, it is not known if sub-Gaussian martingale noise satisfies a Hanson-Wright inequality. In this case, we can demonstrate the concentration of costs around their expectation via a combination of the Azuma-Hoeffding/Azuma-Bernstein inequality with truncation and mixing arguments. This type of argument bounds the fluctuations of the costs around their mean as roughly $(d_{\mathbf{x}} + d_{\mathbf{u}})\sqrt{T}$, which is worse than the square root scaling $\sqrt{d_{\mathbf{x}} + d_{\mathbf{u}}} \cdot \sqrt{T}$ enjoyed by the Hanson-Wright inequality. Up to logarithmic factors, this would yield regret of $(d_{\mathbf{x}} + d_{\mathbf{u}})\sqrt{T} + \sqrt{d_{\mathbf{x}} d_{\mathbf{u}}^2 T} = \sqrt{d_{\mathbf{x}} \max\{d_{\mathbf{x}}, d_{\mathbf{u}}^2\} T}$, which is sub-optimal for $d_{\mathbf{x}} \gg d_{\mathbf{u}}^2$. It is not clear if *any* algorithm can do better in this regime (without a sharper inequality for the concentration of costs around their means), since it is not clear how to ameliorate these random fluctuations. Nevertheless, the final regret bound of $\sqrt{d_{\mathbf{x}} \max\{d_{\mathbf{x}}, d_{\mathbf{u}}^2\} T}$ still improves upon the dimension dependence in the upper bound of $\sqrt{(d_{\mathbf{x}} + d_{\mathbf{u}})^3 T}$ attained by (Mania et al., 2019).