Journal of Machine Learning Research 22 (2021) 1-112 Submitted 11/20; Revised 10/21; Published 10/21

Exact Asymptotics for Linear Quadratic Adaptive Control

Feicheng Wang FEICHENG_WANG@FAS.HARVARD.EDU
Department of Statistics
Harvard University
Cambridge, MA 02138-2901, USA

Lucas Janson LJANSON@FAS.HARVARD.EDU
Department of Statistics
Harvard University
Cambridge, MA 02138-2901, USA

Editor: Jon McAuliffe

Abstract

Recent progress in reinforcement learning has led to remarkable performance in a range of applications, but its deployment in high-stakes settings remains quite rare. One reason is a limited understanding of the behavior of reinforcement algorithms, both in terms of their regret and their ability to learn the underlying system dynamics—existing work is focused almost exclusively on characterizing *rates*, with little attention paid to the constants multiplying those rates that can be critically important in practice. To start to address this challenge, we study perhaps the simplest non-bandit reinforcement learning problem: linear quadratic adaptive control (LQAC) . By carefully combining recent finite-sample performance bounds for the LQAC problem with a particular (less-recent) martingale central limit theorem, we are able to derive asymptotically-*exact* expressions for the regret, estimation error, and prediction error of a rate-optimal stepwise-updating LQAC algorithm. In simulations on both stable and unstable systems, we find that our asymptotic theory also describes the algorithm's finite-sample behavior remarkably well.

Keywords: reinforcement learning, adaptive control, linear dynamical system, system identification, safety, uncertainty quantification, exact asymptotics

1. Introduction

Many dynamic systems such as robots, power grids, or living cells can be described at any given time t by a system state x_t that depends on both its previous state x_{t-1} and some internal or external control u_{t-1} that is applied to direct the system to achieve its desired function. Both adaptive control and reinforcement learning address the problem of choosing the controls u_t when the system dynamics, i.e., the relationship between x_{t+1} and (x_t, u_t), are *unknown.* But the behavior of the algorithms developed in these fields has been characterized only coarsely, even in the simplest systems, preventing their deployment in high-stakes applications that require precise guarantees on safety and performance.

In this paper we will consider a canonical model for such systems, the discrete-time linear dynamical system:

$$x_{t+1} = Ax_t + Bu_t + \varepsilon_t, \tag{1}$$

where $x_t \in \mathbb{R}^n$ represents the state of the system at time t and starts at some initial state x_0, $u_t \in \mathbb{R}^d$ represents the action or control applied at time t, $\varepsilon_t \overset{\text{i.i.d.}}{\sim} \mathcal{N}(0, \sigma^2 I_n)$ is the system noise, and $A \in \mathbb{R}^{n\times n}$ and $B \in \mathbb{R}^{n\times d}$ are matrices determining the system's linear dynamics; the fact that they do not depend on t makes this a *time-homogeneous* dynamical model. We assume the noise ε_t follows a Gaussian distribution, but our proof could be immediately generalized to allow for sub-Gaussian noise distributions. The states x_t and controls u_t are assumed to have been transformed so that x_t closer to zero represents the system better-performing its function, and u_t closer to zero represents lower control cost/effort. The goal is to find an algorithm U that, at each time t, outputs a control $u_t = U(H_t)$ that is computed using the entire thus-far-observed history of the system $H_t = \{x_t, u_{t-1}, x_{t-1}, \dots, u_1, x_1, u_0\}$ to maximize the system's function while minimizing control effort.

We formalize this tradeoff by augmenting the linear dynamics (1) with the popular *quadratic* cost function, so that at every time t, the system incurs the cost $x_t^\top Q x_t + u_t^\top R u_t$, for some known positive-definite matrices $Q \in \mathbb{R}^{n\times n}$ and $R \in \mathbb{R}^{d\times d}$. In order to abstract away finite-sample issues arising from different time horizons T, we will focus on the infinite-horizon problem, which seeks to minimize the expected average limiting cost:

$$\mathcal{J}(U) = \lim_{T\to\infty} \mathbb{E}\mathcal{J}(U,T), \qquad \mathcal{J}(U,T) = \frac{1}{T}\sum_{t=1}^{T}\left(x_t^\top Q x_t + u_t^\top R u_t\right). \tag{2}$$

When the system dynamics A and B are known, the cost-minimizing algorithm is known and called the linear-quadratic regulator (LQR): $U^*(H_t) = Kx_t$, where $K \in \mathbb{R}^{d\times n}$ is the efficiently-computable solution to a system of equations that only depend on A, B, Q, and R; we will review the exact expressions for K in Section 1.3. Like the Gaussian linear model in regression and supervised learning, the aforementioned linear-quadratic problem is foundational to control theory because it is conceptually simple yet it provides a remarkably good description for some real-world systems (e.g., biological systems (Priess et al., 2014), aircraft flight control (Choi and Seo, 1999), or power supply (Shabaani and Jalili-Kharaajoo, 2003)), and insights from its study often translate to innovations and improved understanding in far-more-complex models.

In this paper we consider the case when the system dynamics A and B are unknown, which we call *linear-quadratic adaptive control* (LQAC), to distinguish it from the LQR setting when A and B are assumed known. Intuitively, one might hope that after enough time observing a system controlled by almost any algorithm, one should be able to estimate A and B (and hence K) fairly well and thus be able to apply an algorithm quite close to U^*. Indeed the key challenge in LQAC, as in any reinforcement learning problem, is to trade off *exploration* (actions that help estimate A and B) with *exploitation* (actions that minimize cost). We will quantify the cost of an LQAC algorithm by its *average regret*:[1]

$$\mathcal{R}(U,T) = \mathcal{J}(U,T) - \mathcal{J}(U^*,T).$$

A flurry of recent work has proposed new algorithms for LQAC and studied their regret and estimation error; we review this literature in Section 1.2. These studies have produced

[1] Not to be confused with the more-common *cumulative regret*, given by $T\mathcal{R}(U,T)$. Since one is simply T times the other, it makes no mathematical difference which one is considered, but we prefer a regret formulation that does not diverge to infinity.

finite-sample bounds (in terms of the problem parameters) on various performance metrics which capture the rates at which those metrics depend on various values, especially time T. These recent breakthroughs have advanced the field significantly, but two significant hurdles remain to using their insights to enable reliable, safe, high-performance reinforcement learning.

- Many of the benefits of a theoretical characterization of the performance of an algorithm (e.g., its regret or estimation error) involve quantifying *differences*, such as the difference in performance between two algorithms applied to the same system or the performance difference between applying the same algorithm to two different systems. But a difference between two rigorous, but loose, bounds that have the same rate can be misleading, since the difference in the looseness of the bounds can overwhelm the difference in the true performance.

- When an expression characterizing an algorithm's performance depends explicitly on the system dynamics (in our case, A and B), it cannot actually be evaluated in practice because the system dynamics are by assumption unknown. Thus in order to enable certain critical aspects of reinforcement learning such as safety, non-stationarity detection, and generalization to new systems, there is a pressing need to characterize algorithmic behavior in terms only of *observable* quantities.

1.1 Our Contribution

This paper presents asymptotically-exact expressions for a number of quantities of interest for a simple LQAC algorithm that achieves the optimal rate of regret. That is, we prove that the performance of the algorithm converges *exactly* to the expressions we present. We have two types of results: asymptotically-exact expressions in terms of non-random system *parameters*, and asymptotically-exact expressions in terms of only *observable* random variables.

Theory for a rate-optimal algorithm with stepwise-update estimates. The LQAC algorithm we consider in all of the theory in this paper is very simple and intuitive, using a least-squares estimate of the system dynamics at each time point to estimate the optimal controller K and adding a vanishing exploration noise to that certainty-equivalent control which can be tuned to achieve the optimal rate of regret. All our theory is for a single system trajectory (no independent restarts), and in contrast to existing literature on LQAC we allow our algorithm to update its estimate of the dynamics at *every* time step, although we show our theoretical results can easily be extended to the more common setting of logarithmic updating as well.

Asymptotically-exact expressions characterizing LQAC performance metrics. For a number of different performance metrics of interest for the LQAC problem, we provide asymptotically-exact expressions (a) purely in terms of the non-random, unknown system parameters, and (b) purely in terms of the random, observable system history. In particular, we provide both types (a) and (b) of asymptotically-exact expressions for

(i) the *regret* at any current or future time point,

(ii) the distribution of the *estimation error* of the least-squares estimate of system dynamics A and B, and

(iii) the distribution of the *prediction error* of the least-squares estimate of a future state.

We further use (ii) to derive the estimation error of the least-squares estimate of the optimal controller K, and to identify a function of the dynamics, $A + BK$, that can be estimated at a much faster rate than just A or B (although to reiterate, our expressions characterize not just the rates but the exact constants multiplying those rates as well). Our observable expressions for (ii) and (iii) immediately give us asymptotically-exact online confidence regions for the system dynamics (and optimal controller K) and prediction regions for a future state, respectively. These prediction regions also allow us to perform online testing of the modeling assumptions of, e.g., linearity and stationarity, see Section 3.2.3.

Numerical validation of our theory We apply our algorithm to both a stable and an unstable simulated system to compare our asymptotic expressions to the performance metrics they characterize, and we find quite good agreement, even at very early time steps.

1.2 Related Work

Our study of the asymptotics of the LQAC problem has connections with many works across control theory, machine learning, and statistics, and we defer a more thorough exposition of related work to Section 5, while here only focusing on the most relevant literature.

The LQAC algorithm we consider in this paper falls into the class of algorithms which has been referred to as *certainty equivalent* controllers in the literature. The key idea is to estimate the system dynamics and then apply a control that would be optimal if the estimate were correct.

Following this strategy blindly is known to be inconsistent (Becker et al., 1985; Lai and Robbins, 1982), but consistency can still be achieved with the addition of a persistently exciting input (Åström and Wittenmark, 1973). Another type of fix is to add a vanishing noise term to the input, which was shown by Dean et al. (2018) to achieve $\tilde{\mathcal{O}}(T^{-1/3})$ average regret (their control is estimated by a robust optimization instead of directly from system dynamics estimates) and later by Faradonbeh et al. (2018a,b); Mania et al. (2019) to achieve $\tilde{\mathcal{O}}(T^{-1/2})$ average regret. The recent work of Simchowitz and Foster (2020) refined the existing regret bounds and showed $\tilde{\mathcal{O}}(T^{-1/2})$ to be the *optimal* rate of average regret. To our knowledge, all LQAC algorithms that have been proved to achieve the optimal rate of regret update their estimate of the system dynamics *logarithmically* often,[2] and their bounds on regret and estimation error hold in finite samples but have conservative constants multiplying the rate.

There is work on *system identification* and in particular on *optimal experimental design* that relates to our characterization of the estimation error of the learned system dynamics. These works focus mainly on minimizing estimation error with little or no consideration for the regret, and hence only consider algorithms with average regret bounded away from zero as this allows the optimal rate of estimation error of $\mathcal{O}(T^{-1/2})$. For such algorithms (which essentially correspond to our Algorithm 1 with $\beta = 1$), these works do provide

[2] The only exception is Abeille and Lazaric (2018), whose Thompson sampling algorithm updates its estimates at every step, but their proof only holds for scalar systems ($n = 1$).

asymptotically-exact expressions for the estimation error (Ljung, 1997; Bombois et al., 2006; Gerencsér et al., 2009; Hjalmarsson, 2009; Wahlberg et al., 2010; Huang et al., 2012; Stojanovic and Filipovic, 2014; Stojanovic et al., 2016; Gerencsér et al., 2017). More recent work provides finite-sample bounds on the estimation error of such algorithms, but with conservative constants multiplying the rate (Abbasi-Yadkori et al., 2011; Simchowitz et al., 2018; Sarkar et al., 2019; Dean et al., 2019; Oymak and Ozay, 2019; Sarkar et al., 2019; Khosravi and Smith, 2020; Sattar and Oymak, 2020; Foster et al., 2020; Zheng and Li, 2020; Sun et al., 2020). At the intersection of these approaches, Tu and Recht (2019) showed that a certainty equivalent estimate has an asymptotic sample complexity advantage over a model-free algorithm based on policy gradient.

The main distinction between our paper and all these related works is that we consider a stepwise-updating, regret-rate-optimal LQAC algorithm and provide characterizations of the regret, estimation error, and prediction error that are asymptotically-*exact*. To achieve these results, our proofs combine recent finite-sample bounds (Dean et al., 2018; Mania et al., 2019) with martingale central limit theorems developed in the statistics literature (Lai and Wei, 1982; Anderson and Kunitomo, 1992).

1.3 Preliminaries

We make the following mild assumption on A and B, without which *no* algorithm could even achieve finite average regret.

Assumption 1 (Stability)**.** *Assume the system is* stabilizable*, i.e., there exists K_0 such that the spectral radius (maximum absolute eigenvalue) of $A + BK_0$ is strictly less than 1. We also assume that Q, R are both positive definite.*

Under Assumption 1, there is a unique optimal controller (Arnold and Laub, 1984) that can be computed from A and B, given by the linear feedback controller $u_t = Kx_t$, where

$$K = -(R + B^\top PB)^{-1}B^\top PA. \tag{3}$$

Here P is the unique positive definite solution to the discrete algebraic Riccati equation (DARE):

$$P = A^\top PA - A^\top PB(R + B^\top PB)^{-1}B^\top PA + Q \tag{4}$$

2. Algorithm

The algorithm whose performance we characterize in Section 3 is given in Algorithm 1. At the end of each step in line 5, we apply a plug-in version of the LQR controller, $\hat{K}_t x_t$, plus added exploration noise that vanishes asymptotically with variance $\tau^2 t^{-(1-\beta)} \log^\alpha(t)$. Larger β corresponds to more exploration noise, and we will see that $\beta = 1/2$ gives the optimal rate of regret and is the only β value for which a nonzero α is needed in our theory.[3] $\hat{K}_t$ is taken as the solution to the DARE (Eqs. 3 and 4) with inputs $\hat{A}_{t-1}, \hat{B}_{t-1}$

[3] $\beta = 1$ and $\alpha = 0$ would make the added exploration noise non-vanishing and give the optimal rate of system identification estimation error; see Appendix A.2 for the extension of our results to the case of $\beta = 1$.

computed in line 3. Line 4 then checks whether the state or controller is too large, and if so, $\hat{K}_t$ is set to K_0, which by assumption stabilizes the system. The cutoffs for 'too large' are determined by inputs C_x and C_K, with the latter assumed to be greater than $\|K\|$. We will prove (Proposition 17) the cutoffs are only breached, and hence K_0 applied, finitely often with probability 1, and none of K_0, C_x, or C_K appear in any of our expressions characterizing the asymptotic performance of Algorithm 1.

To prevent the regret from exploding in early time steps (see Fig. I.5), our experiments suggest it is important to keep the check on large states x_t but not the check on large $\hat{K}_t$. We note that $\hat{K}_t$ is computed from $\hat{A}_{t-1}$ and $\hat{B}_{t-1}$ as opposed to $\hat{A}_t$ and $\hat{B}_t$—we expect this to have little impact on the performance but it is needed for the proof of the key Lemma 34. In particular, in the lemma we need the $\hat{K}_{t+1}$ to only be dependent on the history stricly before x_t.

Since the algorithm asymptotically always just applies a noisy plug-in version of the LQR controller, it is simple, intuitive, and computationally efficient.[4] All our theory and experimental results are exactly based on Algorithm 1 without any modification, and in particular, we always analyze a single trajectory (no independent restarts) and our estimates of A and B are updated *stepwise*, i.e., at every time step. This last point is a significant departure from existing literature which focuses on *logarithmic* updating. We show in Figs. 1c and I.1c that updating stepwise reduces regret compared to updating logarithmically often, but in fact our theory also applies to a logarithmically-updated version of Algorithm 1, as made precise in the following remark.

Remark 2 (Logarithmically-updated estimates)**.** *All our theoretical results in Section 3 also hold when $\hat{A}_t$ and $\hat{B}_t$ are only updated $\Theta(\log(t))$ times per t steps. More precisely, assume $\{t_i\}_{i=1}^{\infty}$ are the times at which $\hat{K}_t$ is updated. As long as there exists a constant C such that $\limsup_{i\to\infty} \frac{t_{i+1}}{t_i} \le C$, all results in Section 3 still hold.*

3. Theoretical Results

Almost all of our asymptotic results are based on the following new result which shows that the *Gram matrix* $\sum_{i=0}^{t-1} \begin{bmatrix} x_i \\ u_i \end{bmatrix} \begin{bmatrix} x_i \\ u_i \end{bmatrix}^\top \in \mathbb{R}^{(n+d)\times(n+d)}$ is asymptotically equal in a certain sense to the deterministic matrix $D_t D_t^\top$, where

$$D_t := t^{\beta/2} \log^{\alpha/2}(t) \left[\begin{array}{cc} I_n & 0 \\ K & I_d \end{array} \right] \left[\begin{array}{cc} C_t^{1/2} & 0 \\ 0 & \sqrt{\frac{\tau^2}{\beta}} I_d \end{array} \right], \tag{7}$$

and

$$C_t = t^{1-\beta} \log^{-\alpha}(t) \sum_{p=0}^{\infty} (A+BK)^p ((A+BK)^p)^\top \sigma^2 + \frac{\tau^2}{\beta} \sum_{q=0}^{\infty} (A+BK)^q BB^\top ((A+BK)^q)^\top.$$

[4] The least squares estimator can be computed efficiently in a recursive manner (Engel et al., 2004).

Algorithm 1 Stepwise Noisy Certainty Equivalent Control

```
Require: Initial state $x_0$, stabilizing control matrix $K_0$, scalars $C_x > 0$, $C_K > \|K\|$, $\tau^2 > 0$,
    $\beta \in [1/2, 1)$, and $\alpha > 3/2$ when $\beta = 1/2$.
Let $u_0 = K_0 x_0 + \tau w_0$ and $u_1 = K_0 x_1 + \tau w_1$, with $w_0, w_1 \overset{iid}{\sim} \mathcal{N}(0, I_d)$.
for $t = 2, 3, \dots$ do
    Compute
        $(\hat{A}_{t-1}, \hat{B}_{t-1}) \in \underset{(A', B')}{\arg\min} \sum_{k=0}^{t-2} \left\| x_{k+1} - A' x_k - B' u_k \right\|_2^2$    (5)
    and if stabilizable, plug them into the DARE (Eqs. 3 and 4) to compute $\hat{K}_t$, otherwise
    set $\hat{K}_t = K_0$.
    If $\|x_t\| > C_x \log(t)$ or $\|\hat{K}_t\| > C_K$, reset $\hat{K}_t = K_0$.
    Let
        $u_t = \hat{K}_t x_t + \eta_t, \qquad \eta_t = \tau \sqrt{t^{-(1-\beta)} \log^{\alpha}(t)}\, w_t, \qquad w_t \overset{iid}{\sim} \mathcal{N}(0, I_d)$    (6)
end for
```

Theorem 3. *Algorithm 1 applied to a system described by Eq. 1 under Assumption 1 satisfies*

$$D_t^{-1} \sum_{i=0}^{t-1} \begin{bmatrix} x_i \\ u_i \end{bmatrix} \begin{bmatrix} x_i \\ u_i \end{bmatrix}^\top (D_t^\top)^{-1} \xrightarrow{P} I_{n+d}. \tag{8}$$

The proof of Theorem 3 can be found at Appendix B. The main idea was to first prove Eq. 8 under the simplifying approximation that $\hat{K}_t = K$, and then to derive novel uniform rate bounds on the estimation error $\hat{K}_t - K$ by extending existing bounds (Mania et al., 2019; Dean et al., 2018) to the setting of stepwise update. We require the $\log^{\alpha}(t)$ term because in that case the exploration error terms will dominate the $\hat{K}_t - K$ estimation error terms, and we are only able to tightly control the former with our analysis. Theorem 3 is the key ingredient that will allow us to asymptotically exactly characterize many of the important properties of Algorithm 1.

3.1 Parametric Expressions

We have three different types of asymptotically-exact expressions characterizing the system performance in terms of only the non-random problem parameters (i.e., the algorithm, system, and cost function parameters): the regret (Section 3.1.1), the distribution of the estimation error $[\hat{A}_t - A, \hat{B}_t - B]$ (Section 3.1.2), and the distribution of the prediction error $(\hat{A}_t x_t + \hat{B}_t u_t) - (A x_t + B u_t)$ (Section 3.1.3).

3.1.1 Asymptotically exact expression for the regret (parametric)

Our first result in fact does not follow from Theorem 3 but requires instead a careful decomposition of the regret paired with novel rate bounds.

Theorem 4. *The average regret of the controller U defined by Algorithm 1 applied through time horizon T to a system described by Eq. 1 under Assumption 1 satisfies, as $T \to \infty$,*

$$\frac{\mathcal{R}(U,T)}{\tau^2\beta^{-1}\,\mathbf{Tr}(B^\top PB+R)T^{\beta-1}\log^{\alpha}(T)} \xrightarrow{P} 1, \tag{9}$$

with $\beta = 1/2$ therefore achieving the optimal rate (Simchowitz and Foster, 2020) of $\mathcal{R}(U,T) = \tilde{\mathcal{O}}_p(T^{-1/2})$.

The proof can be found at Appendix C.

It is counterintuitive that the regret monotonically decreases as the exploration noise level τ decreases, as we would expect insufficient exploration to harm the regret at some point. The main intuition behind the theorem is that the regret is dominated by the rate of the exploration noise term (regardless of the constant $\tau > 0$ multiplying it). Hence, asymptotically only the exploration cost stands out, and that cost indeed decreases when τ decreases.

To our knowledge, this is the first time an LQAC algorithm's regret has been characterized asymptotically *exactly*, i.e., Eq. 9 not only captures the rate but also the constant multiplying that rate. With an exact expression for the asymptotic regret, a user can understand exactly how the regret of Algorithm 1 depends on the system parameters, and would be able to compare this expression directly with exact expressions for other algorithms (if they existed).

3.1.2 Asymptotic distribution of the estimation error (parametric)

Theorem 3 provides the key ingredient in a martingale central limit theorem (CLT) for the estimators $\hat{A}_t, \hat{B}_t$ (Anderson and Kunitomo, 1992), which gives the exact asymptotic distribution of the estimation error in terms of only the system parameters.

Theorem 5. *Algorithm 1 applied to a system described by Eq. 1 under Assumption 1 satisfies, as $t \to \infty$,*

$$\mathrm{vec}\left(\left[\hat{A}_t - A, \hat{B}_t - B\right] D_t\right) \xrightarrow{D} \mathcal{N}(0, \sigma^2 I_{n(n+d)}). \tag{10}$$

The proof of Theorem 5 can be found at Appendix D. Again, to our knowledge, this is the first time an LQAC algorithm's estimation error has been characterized asymptotically *exactly* and, similarly, such a result can help a user understand exactly how the distribution of the estimation error of Algorithm 1 depends on the system parameters.

Remark 6 (A convergence rate disparity)**.** *Plugging the definition of D_t Eq. 7 into Eq. 10 gives different convergence rates for two different parts of $[\hat{A}_t - A, \hat{B}_t - B]$. In particular, as $t \to \infty$,*

$$\begin{aligned}\mathrm{vec}\left(\left[t^{\beta/2}\log^{\alpha/2}(t)C_t^{1/2}(\hat{A}_t - A + (\hat{B}_t - B)K), \quad \sqrt{\tfrac{\tau^2}{\beta}}t^{\beta/2}\log^{\alpha/2}(t)(\hat{B}_t - B)\right]\right)& \\ \xrightarrow{D} \mathcal{N}(0, \sigma^2 I_{n(n+d)}).&\end{aligned} \tag{11}$$

Thus $\hat{A}_t - A + (\hat{B}_t - B)K$ converges at the rate of $\left(t^{\beta/2}\log^{\alpha/2}(t)C_t^{1/2}\right)^{-1} = \mathcal{O}_p(t^{-1/2})$ for any β, while $\hat{B}_t - B$ converges at the slower β-dependent rate of $\mathcal{O}_p(t^{-\beta/2}\log^{-\alpha/2}(t))$. The *faster convergence rate of $\hat{A}_t - A + (\hat{B}_t - B)K$ implies strong dependency between $\hat{A}_t - A$ and $\hat{B}_t - B$: $\hat{A}_t - A \approx -(\hat{B}_t - B)K$.*

Remark 7 (Regret-estimation trade-off)**.** *Because of the asymptotic linear relationship $\hat{A}_t - A \approx -(\hat{B}_t - B)K$, the estimation error $[\hat{A}_t - A, \hat{B}_t - B]$ can be characterized by the asymptotic variance of* $\mathrm{vec}[\hat{B}_t - B]$*: $\frac{\beta\sigma^2}{\tau^2} t^{-\beta} \log^{-\alpha}(t) I_{nd}$. Combining this with Theorem 4 gives the following asymptotic identity that precisely characterizes a fundamental regret-estimation trade-off for Algorithm 1 with any β: as $t \to \infty$,*

$$t\mathcal{R}(U,t) \cdot \mathrm{Cov}(\mathrm{vec}(\hat{B}_t - B)) \xrightarrow{P} \mathbf{Tr}(B^\top P B + R)\sigma^2 I_{nd}.$$

Because K is a function of $[A, B]$ (and asymptotically, $\hat{K}_t$ is the same function of $[\hat{A}_{t-1}, \hat{B}_{t-1}]$), by the Delta method, we can use its matrix of derivatives $\frac{dK}{d[A,B]} := \frac{d\,\mathrm{vec}(K)}{d\,\mathrm{vec}([A,B])} \in \mathbb{R}^{nd \times n(n+d)}$ to translate the asymptotic distribution of $[\hat{A}_t - A, \hat{B}_t - B]$ from Theorem 5 to the asymptotic distribution of $\hat{K}_t - K$.

Corollary 8. *Assume $A + BK$ is full rank. Then Algorithm 1 applied to a system described by Eq. 1 under Assumption 1 satisfies, as $t \to \infty$,*

$$\sqrt{\frac{\tau^2}{\sigma^2\beta}} t^{\beta/2} \log^{\alpha/2}(t) \left(\left(\frac{dK}{d[A,B]} \right) \left(\begin{bmatrix} -K^\top \\ I_d \end{bmatrix} \otimes I_n \right) \right)^{-1} \mathrm{vec}\left(\hat{K}_t - K \right) \xrightarrow{D} \mathcal{N}(0, I_{nd}). \quad (12)$$

The proof of Corollary 8 can be found at Appendix F.1. Eq. 12 quantifies the distance from the current control matrix $\hat{K}_t$ to the optimal control matrix K, and shows implicitly but asymptotically exactly how the distribution of that distance depends on the system dynamics.

3.1.3 Asymptotic distribution of the prediction error (parametric)

If we consider the entire history $\{x_i, u_i\}_{i=0}^t$ to be the input of the prediction rule whose goal is to predict the next state x_{t+1}, then the optimal (in terms of mean squared error) prediction is given by $\mathbb{E}[x_{t+1} \mid \{x_i, u_i\}_{i=0}^t] = Ax_t + Bu_t$, and a natural choice at time t would be to use the least-squares prediction rule given by $\hat{A}_t x_t + \hat{B}_t u_t$. By combining Theorem 5's asymptotic distribution for $[\hat{A}_t - A, \hat{B}_t - B]$ with a careful handling of the asymptotic dependence between (x_t, u_t) and $[\hat{A}_t - A, \hat{B}_t - B]$, we can derive the asymptotic distribution of the error $\hat{A}_t x_t + \hat{B}_t u_t - (Ax_t + Bu_t)$ of the least-squares prediction rule.

Theorem 9. *Algorithm 1 applied to a system described by Eq. 1 under Assumption 1 satisfies, as $t \to \infty$,*

$$\left(x_t^\top \left(\sum_{p=0}^{\infty} (A + BK)^p \left((A + BK)^p \right)^\top \right)^{-1} x_t + \beta\sigma^2 \left\| w_t \right\|^2 \right)^{-1/2} \\ t^{1/2} \left((\hat{A}_t - A)x_t + (\hat{B}_t - B)u_t \right) \xrightarrow{D} \mathcal{N}(0, I_n). \quad (13)$$

The proof of Theorem 9 can be found at Appendix E. This expression is parametric in the sense that the first parenthetical only depends on the system parameters and the random variables x_t and w_t that are used by the algorithm in the time step immediately before the prediction is made. Note that the convergence rate of $\tilde{\mathcal{O}}_p(t^{-1/2})$ does not depend on β, as foreshadowed by Remark 6, but the constant in the convergence does depend on β. Thus, Eq. 13 shows that the optimal asymptotic prediction error is attained at $\beta = 1/2$ (x_t's asymptotic distribution does not depend on β, so asymptotically the only β dependence is in the term $\beta\sigma\|w_t\|^2$), a conclusion we could not have reached had we only considered the rate. Theorem 9 can easily be extended to characterize the full prediction error of $x_{t+1} - (\hat{A}_t x_t + \hat{B}_t u_t)$ by simply adding σ^2 to the first parenthetical.

3.2 Observable Expressions

The previous subsection provides three asymptotically-exact expressions (regret, estimation error, and prediction error) in terms of only the system parameters; in this subsection, we provide three analogous asymptotically exact expressions in terms of only observable random variables.

3.2.1 Asymptotically exact expression for the regret (observable)

Define $\hat{P}_t$ as the plug-in estimator using Eq. 4:

$$\hat{P}_t = \hat{A}_t^\top \hat{P}_t \hat{A}_t - \hat{A}_t^\top \hat{P}_t \hat{B}_t (R + \hat{B}_t^\top \hat{P}_t \hat{B}_t)^{-1} \hat{B}_t^\top \hat{P}_t \hat{A}_t + Q.$$

Then by consistency of $\hat{A}_t$ and $\hat{B}_t$ (see Theorem 5), and therefore also $\hat{P}_t$, the plug-in version of Eq. 9 is an immediate corollary of Theorem 4.

Corollary 10. *The average regret of the controller U defined by Algorithm 1 applied through time horizon T to a system described by Eq. 1 under Assumption 1 satisfies, as $t \to \infty$ and $T \to \infty$,*

$$\frac{\mathcal{R}(U,T)}{\tau^2 \beta^{-1}\, \mathbf{Tr}(\hat{B}_t^\top \hat{P}_t \hat{B}_t + R) T^{\beta-1} \log^\alpha(T)} \xrightarrow{P} 1. \tag{14}$$

The proof of Corollary 10 can be found at Appendix F.2. Notice when $t \le T$, Corollary 10 tells us that we can consistently estimate the regret at a future time point. Furthermore, the Delta method applied to Theorem 5 gives the asymptotic distribution of the denominator in Eq. 14.

3.2.2 Asymptotic distribution of the estimation error (observable)

Combining the asymptotic equivalence of Gram matrix and $D_t D_t^\top$ from Theorem 3, the asymptotic distribution of the estimation error from Theorem 5, and Slutsky's theorem immediately produces the following very useful corollary.

Corollary 11. *Algorithm 1 applied to a system described by Eq. 1 under Assumption 1 satisfies*

$$\mathbf{Tr}\left(\left[\hat{A}_t - A, \hat{B}_t - B\right] \sum_{i=0}^{t-1} \begin{bmatrix} x_i \\ u_i \end{bmatrix} \begin{bmatrix} x_i \\ u_i \end{bmatrix}^\top \left[\hat{A}_t - A, \hat{B}_t - B\right]^\top \right) \xrightarrow{D} \sigma^2 \chi^2_{n(n+d)}.$$

The proof of Corollary 11 can be found at Appendix F.3. The reason it is useful is it allows us to construct an asymptotically exact ellipsoidal confidence region for the system dynamics A and B. In particular, the following confidence region has asymptotic coverage exactly $1-\alpha$ and is entirely and efficiently computable from data observable through time t:

$$\left\{A, B \,:\, \sigma^{-2}\,\mathbf{Tr}\left(\left[\hat{A}_t - A, \hat{B}_t - B\right] \sum_{i=0}^{t-1} \begin{bmatrix} x_i \\ u_i \end{bmatrix} \begin{bmatrix} x_i \\ u_i \end{bmatrix}^\top \left[\hat{A}_t - A, \hat{B}_t - B\right]^\top\right) \leq \chi^2_{n(n+d),1-\alpha}\right\}, \tag{15}$$

where $\chi^2_{n(n+d),1-\alpha}$ is the $1-\alpha$ quantile of a $\chi^2_{n(n+d)}$ random variable. To our knowledge, this is the first asymptotically exact confidence region for the system dynamics in the LQAC problem. Note the confidence region in Eq. 15 is identical to the confidence region one would compute if the data points $\{x_i, u_i\}_{i=0}^{t-1}$ were i.i.d., but the theory that led us to this result is far more challenging than in the i.i.d. setting.

Analogously to Corollary 8, we can also use the Delta method to derive a confidence region for K.

Corollary 12. *Assume $A+BK$ is full rank. Then Algorithm 1 applied to a system described by Eq. 1 under Assumption 1 satisfies*

$$\mathrm{vec}(\hat{K}_t - K)^\top \left(\left(\frac{dK}{d[A,B]}\right)_t \left(\sum_{i=0}^{t-1} \begin{bmatrix} x_i \\ u_i \end{bmatrix} \begin{bmatrix} x_i \\ u_i \end{bmatrix}^\top \otimes I_n\right)^{-1} \left(\frac{dK}{d[A,B]}\right)_t^\top\right)^{-1} \mathrm{vec}(\hat{K}_t - K) \xrightarrow{D} \sigma^2 \chi^2_{nd},$$

where $\left(\frac{dK}{d[A,B]}\right)_t \in \mathbb{R}^{nd \times n(n+d)}$ is defined as $\frac{dK}{d[A,B]}$ evaluated at $\hat{A}_{t-1}, \hat{B}_{t-1}$.

The proof of Corollary 12 can be found at Appendix F.4. Corollary 12 gives the following asymptotically exact ellipsoidal $1-\alpha$ confidence region for K:

$$\left\{K \,:\, \sigma^{-2}\mathrm{vec}(\hat{K}_t - K)^\top \left(\left(\frac{dK}{d[A,B]}\right)_t \left(\sum_{i=0}^{t-1} \begin{bmatrix} x_i \\ u_i \end{bmatrix} \begin{bmatrix} x_i \\ u_i \end{bmatrix}^\top \otimes I_n\right)^{-1} \left(\frac{dK}{d[A,B]}\right)_t^\top\right)^{-1} \cdot \right.$$
$$\left. \mathrm{vec}(\hat{K}_t - K) \leq \chi^2_{nd,1-\alpha}\right\}.$$

3.2.3 Asymptotic distribution of the prediction error (observable)

We can obtain an observable expression for the asymptotic distribution of the prediction error as a direct corollary of Theorem 3 and 9.

Corollary 13. *Algorithm 1 applied to a system described by Eq. 1 under Assumption 1 satisfies:*

$$\left(\sigma^2 \begin{bmatrix} x_t \\ u_t \end{bmatrix}^\top \left(\sum_{i=0}^{t-1} \begin{bmatrix} x_i \\ u_i \end{bmatrix} \begin{bmatrix} x_i \\ u_i \end{bmatrix}^\top\right)^{-1} \begin{bmatrix} x_t \\ u_t \end{bmatrix}\right)^{-1/2} \left((\hat{A}_t - A)x_t + (\hat{B}_t - B)u_t\right) \xrightarrow{D} \mathcal{N}(0, I_n).$$

The proof can be found in Appendix F.5, and is a special case of a more general result that allows the users to choose their own desired input by replacing $u_t = \hat{K}_t x_t + \eta_t$ with $u_t = \hat{K}_t x_t + \xi_t$ for any ξ_t constant or independent of the data. Again, Corollary 13 can easily be extended to characterize the full prediction error of $x_{t+1} - (\hat{A}_t x_t + \hat{B}_t u_t)$ by simply adding σ^2 to the first parenthetical, leading to the following prediction region:

$$\left\{ x_{t+1} \,:\, \sigma^{-2} \left(1 + \begin{bmatrix} x_t \\ u_t \end{bmatrix}^\top \left(\sum_{i=0}^{t-1} \begin{bmatrix} x_i \\ u_i \end{bmatrix} \begin{bmatrix} x_i \\ u_i \end{bmatrix}^\top \right)^{-1} \begin{bmatrix} x_t \\ u_t \end{bmatrix} \right)^{-1} \|(\hat{A}_t x_t + \hat{B}_t u_t) - x_{t+1}\|^2 \leq \chi^2_{n,1-\alpha} \right\}. \tag{16}$$

Having at each time t a computable region with a high probability of containing the next state x_{t+1} is a crucial ingredient in ensuring the *safety* of a learning system, as it both provides a warning about where the system will be next and gives the system the opportunity to change or cancel the control u_t if the prediction region intersects an unsafe part of the state space.

As an additional application of the prediction region Eq. 16, since x_{t+1} is observed at the next time step, we can use the agreement between our prediction region and the true x_{t+1} to test certain assumptions about our system. For instance, the hypothesis test which rejects if x_{t+1} does not fall within the prediction region constructed at time t constitutes a asymptotically valid level-α test of our stationary linear dynamics encoded in Eq. 1. For instance, if we are confident about the linearity of our system but worried that it may be non-stationary, we could use this test to detect whether the dynamics have changed within the first $t+1$ time steps, and more generally, such tests could be strung together to constitute a change detection algorithm (Grünwald et al., 2019; Wang and You, 2020).

Note that the naive prediction region

$$\left\{ x_{t+1} \,:\, \sigma^{-2} \|(\hat{A}_t x_t + \hat{B}_t u_t) - x_{t+1}\|^2 \leq \chi^2_{n,1-\alpha} \right\}. \tag{17}$$

also has asymptotically exact coverage even though it ignores the estimation error in $[\hat{A}_t, \hat{B}_t]$. However, our experiments show that our prediction region from Eq. 16 achieves much better finite-sample coverage by accounting for the estimation error of $[\hat{A}_t, \hat{B}_t]$; see Fig. 1e.

4. Experiments

We verify our algorithm's performance in one stable and one unstable dynamical system. We focus on comparing the finite sample performance of our algorithm to our theoretical predictions, and defer comparison between our algorithm and other existing algorithms for future work (see Dean et al. (2018) for a comparison between an algorithm similar to our algorithm except it updates $\hat{K}_t$ logarithmically often and other algorithms which we will review in Appendix 5). In the main text, we will only display the figures with $\beta = 1/2$ and $\alpha = 2$ in the stable system; the remaining figures and details of the experimental setup can be found in Appendix I. [5]

[5] Source code for reproducing our results can be found at https://github.com/Feicheng-Wang/LQAC_code.

4.1 A Representative Simulation

Fig. 1 summarizes the results of our experiment with $\beta = 1/2$ and $\alpha = 2$ in a stable system (for the analogous figure in an unstable system see Fig. I.1). The main takeaways are:

- Fig. 1a shows that Algorithm 1's stepwise update leads to lower regret than update logarithmically often, although the difference is small compared with the variability of the regret. The difference is qualitatively similar but quantitatively larger in the unstable system, and the difference can be quite large for poor choices of K_0, but pretty robust for choices of C_K; see Fig. I.1a, I.2 and I.6.

- Fig. 1b verifies that the ratio of the true observed regret with either of our regret expressions in Theorem 4 and Corollary 10 is converging to 1. Note that the large confidence band is due to the huge variance in the regret itself. The analogous plots for $\beta \neq 1/2$ and the unstable system can be found at Fig. I.1b and I.3; larger β speeds up the convergence speed.

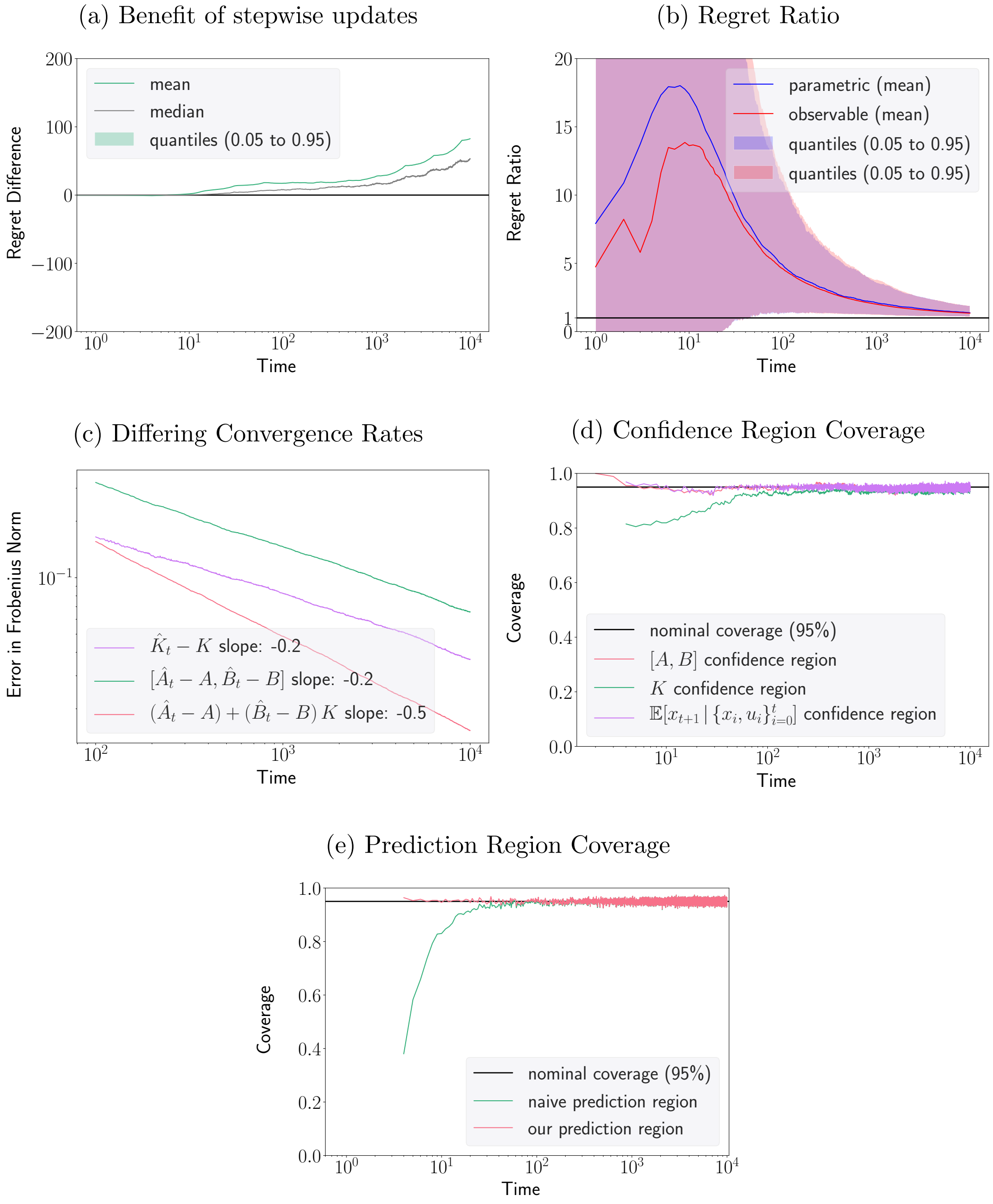
- Fig. 1c verifies the convergence rate disparity in Remark 6 that $\hat{A}_t - A$, $\hat{B}_t - B$, and $\hat{K}_t - K$ have a slow convergence rate $\tilde{\mathcal{O}}(t^{-\beta/2})$, while $\hat{A}_t - A + (\hat{B}_t - B)K$ has a fast convergence rate $\tilde{\mathcal{O}}(t^{-1/2})$; see Fig. I.1c.

- Fig. 1d shows that, the finite sample coverage of our confidence regions and prediction region closely matches our asymptotic theory in Corollary 11, Corollary 12, and Corollary 13. Also Fig. 1e shows that our prediction region Eq. 16 have better finite sample coverage than the naive region Eq. 17. In this simulation, the observable expressions have slightly better coverage. Similar results hold for other choices of β and the unstable systems (Fig. I.1d and I.5).

5. Detailed Review of Related Work

The LQAC problem lies at the intersection of adaptive control and reinforcement learning and has drawn considerable attention in the past decade. This line of work differs from much of the work in reinforcement learning that is based on games or other virtual simulators that can be rerun infinitely many times (Vinyals et al. (2017), Silver et al. (2017)) because it is run in one-shot. However, many real-world applications cannot be easily restarted over and over again, and repeating experiments can be prohibitively expensive. Aside from the CE approach taken in this paper and reviewed in Section 1.2, we classify LQAC algorithms into two broad categories:

- **Optimism in the Face of Uncertainty**: This method uses non-convex optimization to repeatedly select a near optimal control (in the regret sense) from a confidence set, achieves the optimal rate of regret (Abbasi-Yadkori and Szepesvári, 2011; Ibrahimi et al., 2012; Faradonbeh et al., 2017). Later Cohen et al. (2019) extended this work by replacing non-convex optimization with semi-definite programming and still achieves the optimal regret.


(a) Benefit of stepwise updates

(b) Regret Ratio

(c) Differing Convergence Rates

(d) Confidence Region Coverage

(e) Prediction Region Coverage

Figure 1: *(See next page for caption)*

Figure 1: Summary of 1000 independent experiments applying Algorithm 1 with $\beta = 1/2$, $\alpha = 2$, $C_x = 1$, and $C_K = 5$ on the stable system described in Appendix I.1.1. (a) Difference between the regret of Algorithm 1 using stepwise and logarithmic updates. (b) The ratio of the empirical regret and our parametric or observable expressions for the regret. (c) The average Frobenius norm of various estimation errors considered in this paper, with slopes fitted on a log-log scale so that the estimation error is $\tilde{\mathcal{O}}(t^{\text{slope}})$. The effect of α was removed from the slopes of $\hat{K}_t - K$ and $[\hat{A}_t - A, \hat{B}_t - B]$ by dividing the error by $\log^{\alpha/2}(t)$. (d) Coverage of our 95% confidence regions for $[A, B]$, K, and $\mathbb{E}[x_{t+1} \mid \{x_i, u_i\}_{i=0}^t] = Ax_t + Bu_t$. (e) Coverage of our 95% prediction region for $x_{t+1} \mid \{x_i, u_i\}_{i=0}^t$, along with coverage of the naive prediction region given in Eq. 17.

- **Thompson Sampling**: Starting with a prior distribution for the system parameters, one can use Bayes' rule to update a posterior distribution online and can use samples from that posterior to choose controls that balance exploration and exploitation. The pioneering work (Abeille and Lazaric, 2017) applying this idea to LQAC demonstrated a suboptimal $\tilde{\mathcal{O}}(T^{-1/3})$ average regret, which is later improved to the optimal rate $\tilde{\mathcal{O}}(T^{-1/2})$ by Ouyang et al. (2017); Faradonbeh et al. (2018b). Abeille and Lazaric (2018) is the only work which we know of that achieves the optimal rate with stepwise updates, although their proofs only apply in scalar systems (i.e., $n = 1$).

Logarithmic Regret We pause here to clarify that any result achieving logarithmic regret is in a different setting from ours (in our setting, a lower bound of $\tilde{\mathcal{O}}(T^{-1/2})$ was proven in Simchowitz and Foster (2020)). For example, when the system parameters A and B are known or partially known, a logarithmic rate of regret is achievable due to the extra information in A and B which allows faster estimation of K (Foster and Simchowitz, 2020; Cassel et al., 2020). Or, when the states are only partially observed, although the controller receives less information, the optimal controller also has less information, which turns out to allow a logarithmic rate of regret (Lale et al., 2020; Tsiamis and Pappas, 2020). As a final example, when the cost is not an explicit function of the controls u_t, a logarithmic rate of regret is achievable using a controller called a self-tuning regulator, which is similar to our certainty equivalent controller except that it targets a different optimal controller U^* (because the cost function is different) and applies constant size probing steps logarithmically often (Lai and Wei, 1986; Lai, 1986; Guo and Chen, 1991; Guo, 1995).

Sequential Analysis and Time Series Establishing asymptotic normality is common in sequential analysis (Lai, 2001) and time series or state space model analysis (Kohn and Ansley, 1986; Pedroni, 2004), but the focus in these fields is on stationary and Markovian time series (although we assume our system is stabilizable, the data generated by applying our adaptive controller to that system is non-Markovian and non-stationary as the controller depends on the whole history) and on simpler forms of dependence than we consider.

6. Discussion

This paper's main contributions are asymptotically exact expressions for the regret and the distributions of the estimation and prediction errors of a stepwise updating noisy certainty equivalent control algorithm in terms of either the system parameters or observable random variables.

These results improve the field's understanding of the LQAC problem and open up a number of new research directions:

1. **Theoretical improvements**. Our simulations support our suspicion that all of our results except for Theorem 4 and Corollary 10 hold under more general version of Algorithm 1 that allows $\beta = 1/2$ and $\alpha = 0$, the summation in Line 3 to go up to $t-1$, and the removal of Line 4. We expect such extensions to require significantly stronger theoretical machinery, and we hope that future work will prove these extensions and analogues to Theorem 4 and Corollary 10 which account for an expected additional term of order $\mathcal{O}(T^{-1/2})$.

2. **Safe reinforcement learning**. Existing work in safe reinforcement learning relies heavily on prediction regions derived from Bayesian inference (Berkenkamp et al., 2017; Koller et al., 2018). Our Corollary 13 provides a tight frequentist asymptotic prediction region that, unlike Bayesian inference, does not assume a prior on the system parameters, providing a potential starting point for new safe reinforcement learning algorithms.

3. **Non-asm:InitialStableCondition reinforcement learning**. As mentioned in the last paragraph of Section 3, our prediction region can be used for change point detection in non-stationary systems. Many existing work designed for reinforcement learning algorithms in the non-stationary environment relies on some form of change point detection, although they focus on discrete state and action spaces (Da Silva et al., 2006; Auer et al., 2009; Padakandla et al., 2019). Thus, our work may be useful for designing new reinforcement learning algorithms in non-stationary settings with continuous state and action spaces.

Acknowledgments

We are grateful to Na Li, Haoyi Yang, and Yue Li for helpful discussions regarding this project.

Appendix A. Preliminaries

A.1 Notation

Let us first review the definition of $\mathcal{O}(\cdot)$, and generalize the notation to contain relative constants θ, as well as introducing a new notation representing constant functions that we know exactly the order as well as the coefficient in front of the largest order term.

Definition 14. *Let f and g both be real valued function, and suppose $g(x)$ is strictly positive for any x large enough. Then*

1. *$f(x) = \mathcal{O}(g(x))$ if and only if $\exists x_0$, $|f(x)| \leq Mg(x)$ for any $x \geq x_0$.*

2. *$f(x) = \tilde{\mathcal{O}}(g(x))$ if and only if $\exists x_0$ and $\exists k \in \mathbb{Z}$, $|f(x)| \leq Mg(x)\log^k(g(x))$ for any $x \geq x_0$.*

3. *$f(x) = \mathcal{O}(g(x))$ is a fixed function with regard to x such that $\exists C > 0$, and $\lim_{x\to\infty}|f(x)/g(x)| = C$*

4. *$f(x) = \mathcal{O}(\theta; g(x))$ is a fixed function with regard to x such that $\exists C(\theta) > 0$, and $\lim_{x\to\infty}|f(x)/g(x)| = C(\theta)$*

5. *For a set of random variables X_n and a corresponding set of constants a_n, the notation*
$$X_n = o_p(a_n).$$
means that the set of values X_n/a_n converges to zero in probability as n approaches an appropriate limit. Equivalently, $X_n = o_p(a_n)$ can be written as $X_n/a_n = o_p(1)$, where $X_n = o_p(1)$ is defined as
$$X_n \xrightarrow{P} 0.$$

6. *For a set of random variables X_n and Y_n, where Y_n is almost surely non-zero, the notation*
$$X_n = o(Y_n) \ a.s.$$
means that
$$X_n/Y_n \xrightarrow{a.s.} 0.$$

7. *The notation*
$$X_n = \mathcal{O}_p(a_n).$$
means that the set of values X_n/a_n is stochastically bounded. That is, for any $\epsilon > 0$, there exists a finite $M > 0$ and a finite $N > 0$ such that,
$$\mathbb{P}(|X_n/a_n| > M) < \epsilon, \forall n > N.$$

8. $$X_n = \mathcal{O}(a_n) \ a.s.$$
if for almost every $\omega \in \Omega$, there exists a number $C(\omega)$ such that $|X_n(\omega)| \leq C(\omega)a_n$. In other words, $X_n = \mathcal{O}(a_n)$ a.s. if there exists a random variable C such that $|X_n| \leq Ca_n$ a.s. Equivalently,
$$X_n = \mathcal{O}(a_n) \ a.s. \iff \limsup_{n\to\infty} \frac{|X_n|}{a_n} < \infty \ a.s.$$

9. *The notation*
$$X_n = \tilde{\mathcal{O}}_p(a_n).$$
means that the set of values X_n/a_n is stochastically bounded up to a constant order of $log(a_n)$. That is, for any $\epsilon > 0$, there exists a finite $M > 0$, a finite $k \in \mathbb{Z}$, and a finite $N > 0$ such that,
$$\mathbb{P}(|X_n/\log^k(a_n)a_n| > M) < \epsilon, \forall n > N.$$

All these definitions can be generalized to vectors or matrices with entry-wise definition. *Without extra specification, all norms $\|\cdot\|$ (for both vectors and matrices) are meant to be L_2 norm $\|\cdot\|_2$, i.e., operator-2 norm for the matrix.*

Some relationships between these notations are worth keeping in mind: (see Eq.(7) and Eq.(8) in Janson (2011))

$$X_n = o(a_n) \text{ a.s.} \Longrightarrow X_n = o_p(a_n). \tag{18}$$

$$X_n = O(a_n) \text{ a.s.} \Longrightarrow X_n = O_p(a_n). \tag{19}$$

To carefully track down the constant chosen manually, when we state order bounds like $\mathcal{O}(\theta; g(x))$, θ should not contain variables such as δ which are set fixed when we prove high probability bounds but could be varying later, but could contain global constants such as A, B, K, P, Q, R, dimension d, n and C_x, C_u, τ, β that are fixed throughout the whole algorithm.

In order to differentiate $\mathcal{O}(\cdot)$ from fixed constants, we denote $\mathcal{O}(\theta)$ as constant terms which could be potentially varying and only related with θ. That means for the same $\mathcal{O}(\theta)$ symbol in two different places, they can be different constants. One special symbol is $\mathcal{O}(1)$ which represents constant that does not rely on any parameters.

A.2 Extending results to $\beta = 1$

Although the main text only considered vanishing exploration noise (i.e., $\beta < 1$), for completeness (and because it is straightforward to do so) we will also consider the case of $\beta = 1$ and $\alpha \leq 0$ for all of our results.

A.3 Proof dependency tree

In order to make the proof more readable and easier to understand, we put the proof outlines first and summarize most useful middle steps by lemmas. These lemmas' proofs often involve more technical details and is deferred to later parts in the appendix. While this may help readers have better understanding in the high level ideas behind the long proof, we realize that it may also cause loops in the proof structure. Thus, we provide a tree (Fig. A.1) which describes the exact proof dependency structure to make sure that there is no circular argument. In Fig. A.1, all conclusions lies in a perfect tree graph except for the loop marked in red between Lemma 18 and Proposition 17. This is not a contradiction because the proof of Proposition 17 only relies on a subset of conclusions in Lemma 18:

Eqs. 24 and 25, which do not require Proposition 17 to hold. Some of the proofs relies on Eq. 81, which is not included in the graph but still self-consistent (does not rely on other results in the paper).

Appendix B. The proof of Theorem 3

Theorem. *Algorithm 1 applied to a system described by Eq. 1 under Assumption 1 satisfies*

$$D_t^{-1}\sum_{i=0}^{t-1}\begin{bmatrix}x_i\\u_i\end{bmatrix}\begin{bmatrix}x_i\\u_i\end{bmatrix}^\top(D_t^\top)^{-1}\xrightarrow{P} I_{n+d}. \tag{20}$$

B.1 Proof Outline

Proof Let us first examine the Gram matrix $\sum_{i=0}^{t-1}\begin{bmatrix}x_i\\u_i\end{bmatrix}\begin{bmatrix}x_i\\u_i\end{bmatrix}^\top$. Denote

$$M_t := \sum_{i=1}^{t-1} x_i x_i^\top / t^\beta \log^\alpha(t), \tag{21}$$

and

$$\begin{aligned}\Delta_t :=& \sum_{i=1}^{t-1} u_i x_i^\top / t^\beta \log^\alpha(t) - KM_t\\ =& \sum_{i=1}^{t-1}((\hat{K}_t - K)x_i + \eta_i)x_i^\top / t^\beta \log^\alpha(t).\end{aligned} \tag{22}$$

We will show that

$$\sum_{i=0}^{t-1} u_i u_i^\top / t^\beta \log^\alpha(t) = KM_tK^\top + \Delta_t K^\top + K\Delta_t^\top + \frac{\tau^2}{\beta}I_d + o_p(1),$$

and thus we can write our Gram matrix as

$$\begin{aligned}\sum_{i=1}^{t-1}\begin{bmatrix}x_i\\u_i\end{bmatrix}\begin{bmatrix}x_i\\u_i\end{bmatrix}^\top / t^\beta \log^\alpha(t) &= \begin{bmatrix}\sum_{i=0}^{t-1} x_i x_i^\top & \sum_{i=1}^{t-1} x_i u_i^\top\\ \sum_{i=0}^{t-1} u_i x_i^\top & \sum_{i=1}^{t-1} u_i u_i^\top\end{bmatrix} / t^\beta \log^\alpha(t)\\ &= \begin{bmatrix} M_t & M_tK^\top + \Delta_t^\top\\ KM_t + \Delta_t & KM_tK^\top + \Delta_tK^\top + K\Delta_t^\top + \frac{\tau^2}{\beta}I_d\end{bmatrix} + o_p(1)\\ &= \begin{bmatrix} I_n & 0\\ K & I_d\end{bmatrix}\begin{bmatrix} M_t & \Delta_t^\top\\ \Delta_t & \frac{\tau^2}{\beta}I_d\end{bmatrix}\begin{bmatrix} I_n & K^\top\\ 0 & I_d\end{bmatrix} + o_p(1).\end{aligned}$$

Therefore, in order to satisfy

$$D_t^{-1}\sum_{i=0}^{t-1}\begin{bmatrix}x_i\\u_i\end{bmatrix}\begin{bmatrix}x_i\\u_i\end{bmatrix}^\top(D_t^\top)^{-1}\xrightarrow{P} I_{n+d},$$

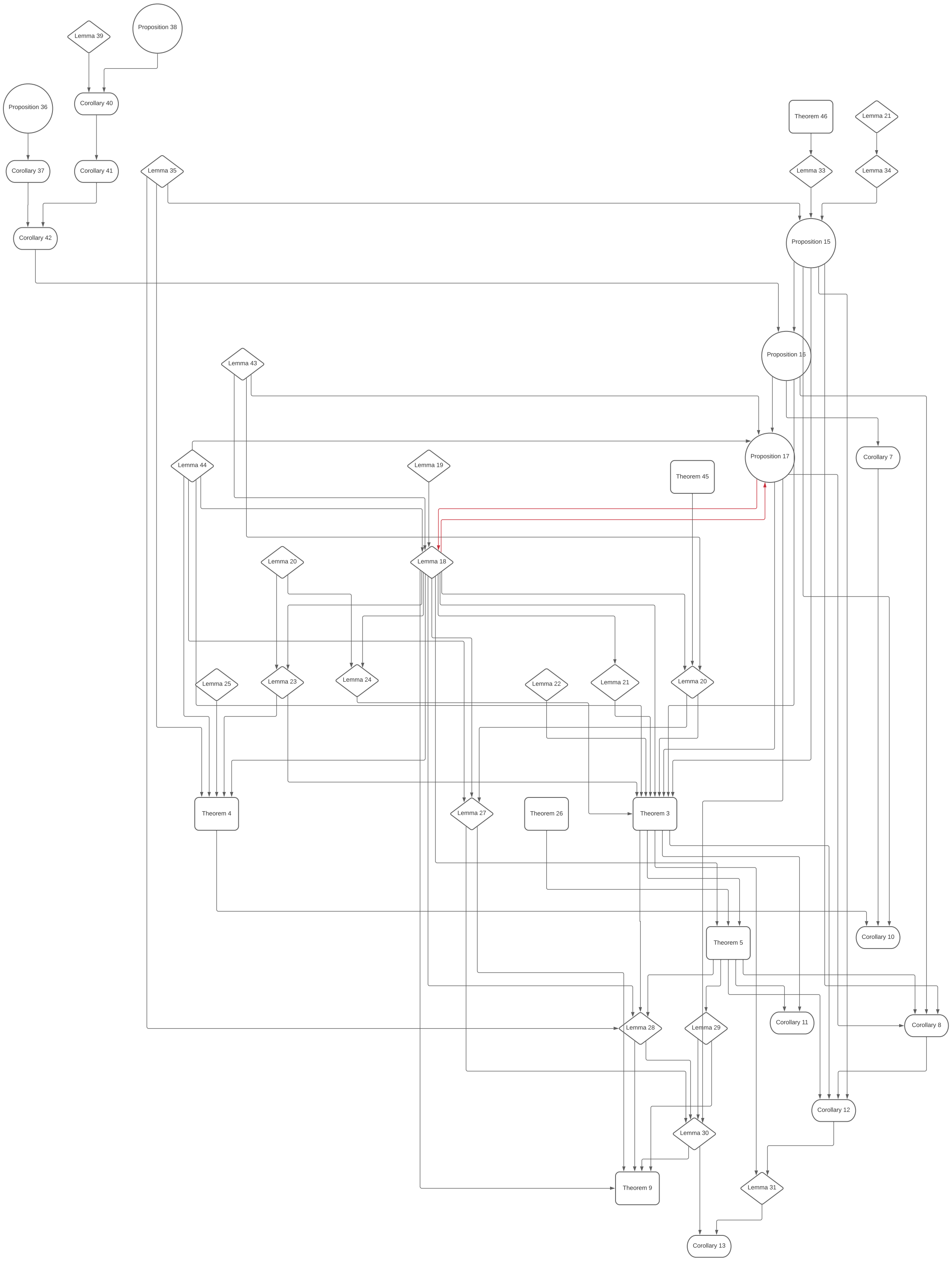


Figure A.1: Proof dependency tree

we can pick $D_t^{-1} := \begin{bmatrix} C_t^{-1/2} & 0 \\ 0 & \sqrt{\frac{\beta}{\tau^2}} I_d \end{bmatrix} \begin{bmatrix} I_n & 0 \\ -K & I_d \end{bmatrix} / t^{\beta/2} \log^{\alpha/2}(t)$. C_t is a deterministic matrix which satisfies $C_t^{-1/2} M_t^{1/2} \xrightarrow{P} I_n$ and $C_t^{-1/2} \Delta_t = o_p(1)$ (we will give C_t's exact expression in Eq. 29). With this choice of D_t^{-1}, we have

$$
\begin{aligned}
& D_t^{-1} \sum_{i=0}^{t-1} \begin{bmatrix} x_i \\ u_i \end{bmatrix} \begin{bmatrix} x_i \\ u_i \end{bmatrix}^\top (D_t^\top)^{-1} \\
=& \begin{bmatrix} C_t^{-1/2} & 0 \\ 0 & \sqrt{\frac{\beta}{\tau^2}} I_d \end{bmatrix} \begin{bmatrix} I_n & 0 \\ -K & I_d \end{bmatrix} \left(\sum_{i=0}^{t-1} \begin{bmatrix} x_i \\ u_i \end{bmatrix} \begin{bmatrix} x_i \\ u_i \end{bmatrix}^\top / t^\beta \log^\alpha(t) \right) \begin{bmatrix} I_n & -K^\top \\ 0 & I_d \end{bmatrix} \begin{bmatrix} C_t^{-1/2} & 0 \\ 0 & \sqrt{\frac{\beta}{\tau^2}} I_d \end{bmatrix} \\
=& \begin{bmatrix} C_t^{-1/2} & 0 \\ 0 & \sqrt{\frac{\beta}{\tau^2}} I_d \end{bmatrix} \begin{bmatrix} I_n & 0 \\ -K & I_d \end{bmatrix} \left(\begin{bmatrix} I_n & 0 \\ K & I_d \end{bmatrix} \begin{bmatrix} M_t & \Delta_t^\top \\ \Delta_t & \frac{\tau^2}{\beta} I_d \end{bmatrix} \begin{bmatrix} I_n & K^\top \\ 0 & I_d \end{bmatrix} + o_p(1) \right) \\
& \cdot \begin{bmatrix} I_n & -K^\top \\ 0 & I_d \end{bmatrix} \begin{bmatrix} C_t^{-1/2} & 0 \\ 0 & \sqrt{\frac{\beta}{\tau^2}} I_d \end{bmatrix} \\
=& \begin{bmatrix} C_t^{-1/2} & 0 \\ 0 & \sqrt{\frac{\beta}{\tau^2}} I_d \end{bmatrix} \begin{bmatrix} I_n & 0 \\ -K & I_d \end{bmatrix} \begin{bmatrix} I_n & 0 \\ K & I_d \end{bmatrix} \begin{bmatrix} M_t & \Delta_t^\top \\ \Delta_t & \frac{\tau^2}{\beta} I_d \end{bmatrix} \begin{bmatrix} I_n & K^\top \\ 0 & I_d \end{bmatrix} \\
& \cdot \begin{bmatrix} I_n & -K^\top \\ 0 & I_d \end{bmatrix} \begin{bmatrix} C_t^{-1/2} & 0 \\ 0 & \sqrt{\frac{\beta}{\tau^2}} I_d \end{bmatrix} + o_p(1) \quad \text{(we can move } o_p(1) \text{ outside because } C_t^{-1/2} \to 0) \\
=& \begin{bmatrix} C_t^{-1/2} & 0 \\ 0 & \sqrt{\frac{\beta}{\tau^2}} I_d \end{bmatrix} \begin{bmatrix} M_t & \Delta_t^\top \\ \Delta_t & \frac{\tau^2}{\beta} I_d \end{bmatrix} \begin{bmatrix} C_t^{-1/2} & 0 \\ 0 & \sqrt{\frac{\beta}{\tau^2}} I_d \end{bmatrix} + o_p(1) \\
=& \begin{bmatrix} C_t^{-1/2} M_t C_t^{-1/2} & \sqrt{\frac{\beta}{\tau^2}} C_t^{-1/2} \Delta_t^\top \\ \sqrt{\frac{\beta}{\tau^2}} \Delta_t C_t^{-1/2} & I_d \end{bmatrix} + o_p(1) \\
=& I_{n+d} + o_p(1).
\end{aligned}
$$

Components needing further explanation In the final step of the above derivation there are still several points that remains unclear, namely

- $\sum_{i=0}^{t-1} u_i u_i^\top / t^\beta \log^\alpha(t) = K M_t K^\top + \Delta_t K^\top + K \Delta_t^\top + \frac{\tau^2}{\beta} I_d + o_p(1)$,
- $C_t^{-1/2} M_t^{1/2} \xrightarrow{P} I_n$, and
- $C_t^{-1/2} \Delta_t = o_p(1)$.

As we will see, the order of Δ_t is decided by the convergence rate of $\hat{K}_t - K$. Because of that, the first step in our proof is to identify the convergence rate of $\hat{K}_t - K$. Then we will prove the three remaining points in The proof of Eq. 20. To summarize, our proof can be mainly separated into two big steps:

1. Identify the convergence rate of $\hat{K}_t - K$. (see Appendix B.2)

2. Prove Eq. 20 holds:

$$D_t^{-1}\sum_{i=0}^{t-1}\begin{bmatrix}x_i\\u_i\end{bmatrix}\begin{bmatrix}x_i\\u_i\end{bmatrix}^\top (D_t^\top)^{-1} \xrightarrow{P} I_{n+d}.$$

- Summarize uniform high probability bound for some random variables, which will serve as basic tools for later proof. (see Appendix B.3.1)
- Prove $C_t^{-1/2}M_t^{1/2} \xrightarrow{P} I_n$. (see Appendix B.3.2)
- Prove $C_t^{-1/2}\Delta_t = o_p(1)$. (see Appendix B.3.3)
- Prove $\sum_{i=0}^{t-1} u_i u_i^\top / t^\beta = KM_tK^\top + \Delta_t K^\top + K\Delta_t^\top + \frac{\tau^2}{\beta} I_d + o_p(1)$. (see Appendix B.3.4)

Now we will examine these steps in order. ■

B.2 Convergence rate of $\hat{K}_t - K$

As said in the previous part, the main purpose of this section is to derive the convergence rate of $\hat{K}_t - K$, which is one crucial step in our proof. Denote the stabilizing controller computed by Line 3 Algorithm 1 as $\tilde{K}_{t+1}$, i.e.,

$$\tilde{K}_{t+1} = \begin{cases}\text{Solve DARE Eqs. 3 and 4 with } A = \hat{A}_t, B = \hat{B}_t, & \text{for } (\hat{A}_t, \hat{B}_t) \text{ stabilizable}\\ K_0, & \text{for } (\hat{A}_t, \hat{B}_t) \text{ not stabilizable}\end{cases}.$$

By Line 4 Algorithm 1, $\hat{K}_{t+1}$ can be written as:

$$\hat{K}_{t+1} = \begin{cases} K_0, & \text{when } \|x_t\| > C_x \log(t) \text{ or } \|\hat{K}_t\| > C_K \\ \tilde{K}_{t+1}, & \text{otherwise}\end{cases}.$$

In particular, the proof can be separated into three parts:

1. Derive the convergence rate of $\hat{A}_t$ and $\hat{B}_t$.
2. Show that $\tilde{K}_{t+1}$ enjoy the same convergence rate as $\hat{A}_t$ and $\hat{B}_t$.
3. Show that $\hat{K}_{t+1}$ is only different from $\tilde{K}_{t+1}$ finitely often, and as a result, $\hat{K}_{t+1}$ also enjoy the same convergence rate as $\hat{A}_t$ and $\hat{B}_t$.

Correspondingly we have the following three propositions:

Proposition 15 (Similar to Proposition C.1 in Dean et al. (2018)). *Let $x_0 \in \mathbb{R}^n$ be any initial state. Assume Assumption 1 is satisfied. When applying Algorithm* 1,

$$\max\left\{\|\hat{A}_t - A\|, \|\hat{B}_t - B\|\right\} = \mathcal{O}(t^{-\frac{\beta}{2}} \log^{\frac{-\alpha+1}{2}}(t)) \ a.s.$$

The proof of Proposition 15 can be found in Appendix G.1.

Proposition 16. *Let $x_0 \in \mathbb{R}^n$ be any initial state. Assume Assumption 1 is satisfied. When applying Algorithm* 1,

$$\max\left\{\|\hat{A}_t - A\|, \|\hat{B}_t - B\|, \|\tilde{K}_{t+1} - K\|\right\} = \mathcal{O}(t^{-\frac{\beta}{2}} \log^{\frac{-\alpha+1}{2}}(t)) \ a.s.$$

The proof of Proposition 16 can be found in Appendix G.2.

Proposition 17. *Let $x_0 \in \mathbb{R}^n$ be any initial state. Assume Assumption 1 is satisfied. When applying Algorithm 1,*

$$\max\left\{\|\hat{A}_t - A\|, \|\hat{B}_t - B\|, \|\hat{K}_{t+1} - K\|\right\} = \mathcal{O}(t^{-\frac{\beta}{2}} \log^{\frac{-\alpha+1}{2}}(t)) \ a.s. \tag{23}$$

The proof of Proposition 17 can be found in Appendix G.3.

Propositions 15, 16, and 17 all hold additionally for a version of Algorithm 1 that only updates logarithmically often; see Appendix G. The takeaway from this section is the uniform bound for $\|\hat{K}_{t+1} - K\|$ Eq. 23, which is the only property of $\hat{K}_t$ we need for the rest of the proof.

B.3 Proving Eq. 20

B.3.1 Uniform Bounds

In this section we will show several basic uniform bounds that will be used frequently in the later The proof of Theorem 3.

Lemma 18.

- $$\|\varepsilon_t\|, \|\eta_t\| = \mathcal{O}(\log^{1/2}(t)) \ a.s. \tag{24}$$

- $$\|B\eta_t + \varepsilon_t\| = \mathcal{O}(\log^{1/2}(t)) \ a.s. \tag{25}$$

Assume Eq. 23, then:

- $$\|\delta_t\| = \|\hat{K}_t - K\| = \mathcal{O}(t^{-\frac{\beta}{2}} \log^{\frac{-\alpha+1}{2}}(t)) \ a.s. \tag{26}$$

- *For $t > q$,*
$$\|(L + B\delta_{t-1}) \cdots (L + B\delta_q)\| = \mathcal{O}(\rho_L^{t-q}) \ a.s. \tag{27}$$

- $$\|x_t\|, \|u_t\| = \mathcal{O}(\log^{1/2}(t)) \ a.s. \tag{28}$$

where $\delta_t := \hat{K}_t - K$, $L := A + BK$, and $\rho_L := \frac{2+\rho(L)}{3}$. ***Additionally, when** $t = 0, 1$ **all these terms are bounded by** $\mathcal{O}(1)$ **a.s.***

The proof can be found in Appendix H.1.1. Following Definition 14 Item 8, Lemma 18 presents uniform upper bounds for $t \geq 0$. We will see that all states x_t and actions u_t can be expressed in recursive summations, which can be bounded easily if we have uniform upper bound for each of their components.

Let us briefly explain why these orders makes sense.

- The first two inequalities come from the tail bound for standard Gaussian random variables, whose maximum scales as $\log^{1/2}(t)$.
- The third inequality Eq. 26 directly follows from Eq. 23.
- The fourth inequality Eq. 27 holds with exponential decay because the L has spectual radius < 1 and by Eq. 26, δ_t is shrinking to 0.
- The fifth inequality Eq. 28 holds because the system is stabilizable and the effect of previous states and actions are exponentially decaying, leaving the main factor in the norm to come from the recent system noises. By the first two inequalities $\|x_t\|$ is uniformly bounded by $\log^{1/2}(t)$ scale.

B.3.2 Showing $C_t^{-1/2}M_t^{1/2} \xrightarrow{P} I_n$

We wish to show that $M_t = \sum_{i=0}^{t-1} x_i x_i^\top / t^\beta \log^\alpha(t) = C_t(1 + o_p(1))$, where

$$C_t = \log^{-\alpha}(t) t^{1-\beta} \sum_{p=0}^{\infty} L^p (L^p)^\top \sigma^2 + \frac{\tau^2}{\beta} \sum_{q=0}^{\infty} L^q B B^\top (L^q)^\top \tag{29}$$

Recall the system definition Eq. 1:

$$x_{t+1} = Ax_t + Bu_t + \varepsilon_t.$$

and the input Eq. 6

$$u_t = \hat{K}_t x_t + \eta_t.$$

Recursively applying these two equations produces the following formula for x_t in terms of x_0, $\{\varepsilon_p\}_{p=0}^{t-1}$, and $\{\eta_p\}_{p=0}^{t-1}$.

Lemma 19. *For any $t \geq 1$,*

$$x_t = \sum_{p=0}^{t-1} (A + B\hat{K}_{t-1}) \cdots (A + B\hat{K}_{p+1})(B\eta_p + \varepsilon_p) + (A + B\hat{K}_{t-1}) \cdots (A + BK_0)x_0, \tag{30}$$

and

$$u_t = \sum_{p=0}^{t-1} \hat{K}_t (A + B\hat{K}_{t-1}) \cdots (A + B\hat{K}_{p+1})(B\eta_p + \varepsilon_p) + \hat{K}_t (A + B\hat{K}_{t-1}) \cdots (A + BK_0)x_0 + \eta_t.$$

Here when $p = t - 1$, we define the product $(A + B\hat{K}_{t-1}) \cdots (A + B\hat{K}_{p+1}) := I_n$.

The proof can be found in Appendix H.1.2. As a result, we can rewrite $\sum_{i=0}^{t-1} x_i x_i^\top$ into a summation in terms of $\{\varepsilon_i, \eta_i\}_{i=0}^{t-1}$. First consider the terms without x_0.

$$\sum_{i=1}^{t-1}\sum_{p=0}^{i-1}\sum_{q=0}^{i-1} \Big[(A+B\hat{K}_{i-1})\cdots(A+B\hat{K}_{p+1})(B\eta_p+\varepsilon_p)\Big]\Big[(A+B\hat{K}_{i-1})\cdots(A+B\hat{K}_{q+1})(B\eta_q+\varepsilon_q)\Big]^\top.$$

This whole expression can be separated into four components with the following bounds:

Lemma 20. *Assume Eq. 23, then:*

1.

$$\sum_{i=1}^{t-1}\sum_{p=0}^{i-1}\sum_{q=0}^{i-1}\left[(A+BK)^{i-p-1}\right](B\eta_p+\varepsilon_p)(B\eta_q+\varepsilon_q)^\top\left[(A+BK)^{i-q-1}\right]^\top = t^\beta \log^\alpha(t)(C_t+o_p(1)).$$

2.

$$\sum_{i=1}^{t-1}\sum_{p=0}^{i-1}\sum_{q=0}^{i-1}\Big[(A+B\hat{K}_{i-1})\cdots(A+B\hat{K}_{p+1})-(A+BK)^{i-p-1}\Big] \cdot(B\eta_p+\varepsilon_p)(B\eta_q+\varepsilon_q)^\top\left[(A+BK)^{i-q-1}\right]^\top = \mathcal{O}(t^{1-\beta/2}\log^{\frac{-\alpha+3}{2}}(t))\ \textit{a.s.}$$

3.

$$\sum_{i=1}^{t-1}\sum_{p=0}^{i-1}\sum_{q=0}^{i-1}\left[(A+BK)^{i-p-1}\right](B\eta_p+\varepsilon_p)(B\eta_q+\varepsilon_q)^\top\Big[(A+B\hat{K}_{i-1})\cdots(A+B\hat{K}_{q+1})-(A+BK)^{i-q-1}\Big]^\top = \mathcal{O}(t^{1-\beta/2}\log^{\frac{-\alpha+3}{2}}(t))\ \textit{a.s.}$$

4.

$$\sum_{i=1}^{t-1}\sum_{p=0}^{i-1}\sum_{q=0}^{i-1}\Big[(A+B\hat{K}_{i-1})\cdots(A+B\hat{K}_{p+1})-(A+BK)^{i-p-1}\Big](B\eta_p+\varepsilon_p)(B\eta_q+\varepsilon_q)^\top \cdot\Big[(A+B\hat{K}_{i-1})\cdots(A+B\hat{K}_{q+1})-(A+BK)^{i-q-1}\Big]^\top = \mathcal{O}(t^{1-\beta/2}\log^{\frac{-\alpha+3}{2}}(t))\ \textit{a.s.}$$

The proof can be found in Appendix H.1.3.

It remains to consider the remaining terms with x_0, which is relatively straight-forward, since the effect of the initial state is exponentially decaying when $t \to \infty$.

Lemma 21. *Assume Eq. 23, then*

1. $\sum_{i=0}^{t-1}\Big[(A+B\hat{K}_{i-1})\cdots(A+BK_0)x_0\Big]\Big[\sum_{q=0}^{i-1}(A+B\hat{K}_{i-1})\cdots(A+B\hat{K}_{q+1})(B\eta_q+\varepsilon_q)\Big]^T = \tilde{\mathcal{O}}(1)$ *a.s.*

2. $\sum_{i=0}^{t-1}\left[(A+B\hat{K}_{i-1})\cdots(A+BK_0)x_0\right]\left[(A+B\hat{K}_{i-1})\cdots(A+BK_0)x_0\right]^T = \mathcal{O}(1)$ *a.s.*

The proof can be found in Appendix H.1.4. As mentioned in Eq. 19, $\mathcal{O}$ a.s. notation is stronger than $\mathcal{O}_p$ notation. Summing up all the results in Lemma 20 and Lemma 21 we can finally conclude that

$$\sum_{i=0}^{t-1} x_i x_i^\top = t^\beta \log^\alpha(t)(C_t + o_p(1)) + \mathcal{O}_p(t^{1-\beta/2}\log^{\frac{-\alpha+3}{2}}(t)).$$

Thus

$$M_t = \sum_{i=0}^{t-1} x_i x_i^\top / t^\beta \log^\alpha(t) = C_t + o_p(1) + \mathcal{O}_p(t^{1-3\beta/2}\log^{\frac{-3\alpha+3}{2}}(t)), \tag{31}$$

where C_t is defined in Eq. 29 This is already very close to our objective $C_t^{-1/2}M_t^{1/2} \xrightarrow{P} I_n$, but we still need to show that C_t is an invertible matrix. C_t is already a positive semi-definite (PSD) matrix because it is a weighted summation of PSD matrices $L^p(L^p)^\top$ and $L^qBB^\top(L^q)^\top$. The only thing we need to ensure is that C_t is a full rank matrix. And that is indeed true because the $p=0$ term is the identity matrix, and adding more PSD matrices $L^p(L^p)^\top$ and $L^qBB^\top(L^q)^\top$ will not change its positive definite nature. Following Eq. 29, we have (because $\beta < 1$ or $\beta = 1$ and $\alpha \le 0$)

$$C_t = \log^{-\alpha}(t)t^{1-\beta}\sum_{p=0}^{\infty} L^p\left(\sigma^2 I_n + 1_{\{\beta=1,\alpha=0\}}\tau^2 BB^\top\right)(L^p)^\top(I_n + o(1)). \tag{32}$$

Thus

$$C_t^{-1} = t^{\beta-1}\log^\alpha(t)\left(\sum_{p=0}^{\infty} L^p\left(\sigma^2 I_n + 1_{\{\beta=1,\alpha=0\}}\tau^2 BB^\top\right)(L^p)^\top\right)^{-1}(I_n+o(1)) = \mathcal{O}(t^{\beta-1}\log^\alpha(t)). \tag{33}$$

Noticing that

$$\mathcal{O}(t^{\beta-1}\log^\alpha(t))\mathcal{O}_p(t^{1-3\beta/2}\log^{\frac{-3\alpha+3}{2}}(t)) = \mathcal{O}_p(t^{-\beta/2}\log^{\frac{-\alpha+3}{2}}(t)) = o_p(1),$$

we have from Eq. 31

$$C_t^{-1}M_t \xrightarrow{P} I_n. \tag{34}$$

With the help of the following lemma we conclude that $C_t^{-1/2}M_t^{1/2} \xrightarrow{P} I_n$.

Lemma 22. *Assume we have two matrix sequences $\{A_t\}_{t=1}^\infty$ and $\{B_t\}_{t=1}^\infty$, where A_t and B_t are $p \times p$ positive definite matrices, then*

$$A_t^2 B_t^2 \xrightarrow{P} I_p.$$

iff

$$A_t B_t \xrightarrow{P} I_p.$$

The proof can be found in Appendix H.1.5 (Thanks for the help from Haoyi Yang and Yue Li in proving this lemma).

B.3.3 Proving $C_t^{-1/2}\Delta_t = o_p(1)$

Recall the definition of Δ_t from Eq. 22:

$$\Delta_t := \left(\sum_{i=0}^{t-1}(\hat{K}_i - K)x_i x_i^\top + \sum_{i=0}^{t-1}\eta_i x_i^\top\right)/t^\beta \log^\alpha(t).$$

The order of Δ_t depends on the order of its two components:

Lemma 23. *Assume Eq. 23, then*

1. $\sum_{i=0}^{t-1}(\hat{K}_i - K)x_i x_i^\top = \mathcal{O}(t^{1-\beta/2}\log^{\frac{-\alpha+3}{2}}(t))$ *a.s.*
2. $\sum_{i=0}^{t-1}\eta_i x_i^\top = o\left(t^{\beta/2}\log^{\frac{\alpha+3}{2}}(t)\right)$ *a.s.*

The proof can be found in Appendix H.1.6. The first term has larger order than the second term when $1/2 \leq \beta < 1$ or $\beta = 1$ and $\alpha \leq 0$. As a result, we have

$$\Delta_t = \mathcal{O}(t^{1-3\beta/2}\log^{\frac{-3\alpha+3}{2}}(t)) \text{ a.s.} \quad (\text{when } \beta \in [1/2, 1)) \tag{35}$$

Observe from Eq. 33:

$$C_t^{-1} = \mathcal{O}(t^{\beta-1}\log^\alpha(t)).$$

Then when $\beta > 1/2$ or $\beta = 1/2, \alpha > 3/2$

$$\begin{aligned} C_t^{-1/2}\Delta_t =& \mathcal{O}(t^{-1/2+\beta/2}\log^{\alpha/2}(t)t^{1-3\beta/2}\log^{\frac{-3\alpha+3}{2}}(t)) \\ =& \mathcal{O}(t^{1/2-\beta}\log^{\frac{-2\alpha+3}{2}}(t)) \\ =& o(1) \text{ a.s.} \end{aligned}$$

B.3.4 Proving $\sum_{i=0}^{t-1} u_i u_i^\top / t^\beta \log^\alpha(t) = KM_tK^\top + \Delta_t K^\top + K\Delta_t^\top + \frac{\tau^2}{\beta}I_d + o_p(1)$

Finally we need to check

$$\sum_{i=0}^{t-1} u_i u_i^\top = \sum_{i=0}^{t-1}((K+\delta_i)x_i + \eta_i)((K+\delta_i)x_i + \eta_i)^\top,$$

where $\delta_i = \hat{K}_i - K$. There are six different kinds of terms in the above equation, namely $\sum_{i=0}^{t-1} Kx_i x_i^T K^\top$, $\sum_{i=0}^{t-1} Kx_i x_i^\top \delta_i^\top$ and $\sum_{i=0}^{t-1} \delta_i x_i x_i^T K^\top$, $\sum_{i=0}^{t-1} Kx_i\eta_i^\top$ and $\sum_{i=0}^{t-1}\eta_i x_i^T K^\top$, $\sum_{i=0}^{t-1}\delta_i x_i x_i^\top \delta_i^\top$, $\sum_{i=0}^{t-1}\delta_i x_i \eta_i^\top$ and $\sum_{i=0}^{t-1}\eta_i x_i^\top \delta_i^\top$, and $\sum_{i=0}^{t-1}\eta_i\eta_i^\top$. The first three terms can be written as

$$\sum_{i=0}^{t-1} Kx_i x_i^T K^\top / t^\beta \log^\alpha(t) = KM_tK^\top,$$

and

$$\left(\sum_{i=0}^{t-1} Kx_i x_i^\top \delta_i^\top + \sum_{i=0}^{t-1}\delta_i x_i x_i^T K^\top + \sum_{i=0}^{t-1} Kx_i\eta_i^\top + \sum_{i=0}^{t-1}\eta_i x_i^T K^\top\right)/t^\beta \log^\alpha(t) = K\Delta_t^T + \Delta_t K^T.$$

The remaining terms can be summarized by

Lemma 24. *Assume Eq. 23, then*

1. $\sum_{i=0}^{t-1} \delta_i x_i x_i^\top \delta_i^\top = \mathcal{O}(t^{1-\beta} \log^{-\alpha+2}(t))$ *a.s.*
2. $\sum_{i=0}^{t-1} \delta_i x_i \eta_i^\top = (\sum_{i=0}^{t-1} \eta_i x_i^\top \delta_i^\top)^\top = o\left(\log^2(t)\right)$ *a.s.*
3. $\sum_{i=0}^{t-1} \eta_i \eta_i^\top = t^\beta \frac{\tau^2}{\beta} \log^\alpha(t)(I_d + o_p(1))$

The proof of Lemma 24 can be found in Appendix H.1.7. Combining all parts in Lemma 24 we have when $\beta > 1/2$ or $\beta = 1/2, \alpha > 1$, the third item dominates the other two. To sum up, we have

$$\sum_{i=0}^{t-1} u_i u_i^\top / t^\beta \log^\alpha(t) = K M_t K^\top + \Delta_t K^\top + K \Delta_t^\top + \frac{\tau^2}{\beta} I_d + o_p(1). \tag{36}$$

Summary Now we have completed all missing proof pieces in the proof of Eq. 20, which finishes The proof of Theorem 3.

Appendix C. The proof of Theorem 4

Theorem. *The average regret of the controller U defined by Algorithm 1 applied through time horizon T to a system described by Eq. 1 under Assumption 1 satisfies, as $T \to \infty$,*

$$\frac{\mathcal{R}(U, T)}{\tau^2 \beta^{-1} \mathbf{Tr}(B^\top P B + R) T^{\beta-1} \log^\alpha(T)} \xrightarrow{P} 1,$$

with $\beta = 1/2$ therefore achieving the optimal rate (Simchowitz and Foster, 2020) of $\mathcal{R}(U, T) = \tilde{\mathcal{O}}_p(T^{-1/2})$.

C.1 Proof Outline

Proof We are interested in the cost

$$\sum_{t=1}^{T} x_t^\top Q x_t + u_t^\top R u_t \quad \text{with } u_t = \hat{K}_t x_t + \eta_t.$$

Recall the Eq. 30 from Lemma 19 that

$$x_t = \sum_{p=0}^{t-1} (A + B\hat{K}_{t-1}) \cdots (A + B\hat{K}_{p+1})(B\eta_p + \varepsilon_p) + (A + B\hat{K}_{t-1}) \cdots (A + BK_0) x_0.$$

Notice that the state x_t has the same expression as if the system had noise $\tilde{\varepsilon}_t = B\eta_t + \varepsilon_t$ and controller $\tilde{u}_t = \hat{K}_t x_t$. We wish to switch to the new system because there are some existing tools with controls in the form of $\tilde{u}_t = \hat{K}_t x_t$.

We will first show in Appendix C.2 that the difference between the original cost and transformed cost is

$$\sum_{t=1}^{T} u_t^\top R u_t - \tilde{u}_t^\top R \tilde{u}_t = \frac{\tau^2}{\beta} T^\beta \log^\alpha(T) \, \mathbf{Tr}(R)(1 + o_p(1)),$$

and then prove in Appendix C.3 the new system cost is

$$\sum_{t=1}^{T} x_t^\top Q x_t + \tilde{u}_t^\top R \tilde{u}_t = T\sigma^2 \mathbf{Tr}(P) + \frac{\tau^2}{\beta} T^\beta \log^\alpha(T) \mathbf{Tr}(B^\top P B)(1 + o_p(1)).$$

Combining the above two equations, we conclude that

$$\begin{aligned}\mathcal{J}(U,T) &= \frac{1}{T}\left[\sum_{t=1}^{T} x_t^\top Q x_t + u_t^\top R u_t\right] \\ &= \sigma^2 \mathbf{Tr}(P) + \tau^2 \beta^{-1} \mathbf{Tr}(B^\top P B + R) T^{\beta-1} \log^\alpha(T)(1 + o_p(1)).\end{aligned}$$

Based on similar analysis we prove in Appendix C.4 that

$$\mathcal{J}(U^*,T) = \sigma^2 \mathbf{Tr}(P) + \mathcal{O}_p(T^{-1/2}\log(T)).$$

Recall that we choose $\beta \in [1/2, 1]$, and $\alpha > 3/2$ when $\beta = 1/2$, which means $T^{\beta-1}\log^\alpha(T)$ is of larger order than $T^{-1/2}\log(T)$. Finally we finish the proof with

$$\begin{aligned}\mathcal{R}(U,T) &= \mathcal{J}(U,T) - \mathcal{J}(U^*,T) \\ &= \tau^2\beta^{-1} \mathbf{Tr}(B^\top P B + R) T^{\beta-1} \log^\alpha(T)(1 + o_p(1)).\end{aligned}$$

■

C.2 Cost difference induced by transformation

The difference is expressed as

$$\begin{aligned}\sum_{t=1}^{T} u_t^\top R u_t - \tilde{u}_t^\top R \tilde{u}_t =& \sum_{t=1}^{T} (\hat{K}_t x_t + \eta_t)^\top R (\hat{K}_t x_t + \eta_t) - \sum_{t=1}^{T} (\hat{K}_t x_t)^\top R (\hat{K}_t x_t) \\ =& 2\sum_{t=1}^{T} (\hat{K}_t x_t)^\top R \eta_t + \sum_{t=1}^{T} \eta_t^\top R \eta_t.\end{aligned}$$

We show in Eq. 83 that

$$\sum_{t=1}^{T} (\hat{K}_t x_t)^\top R \eta_t = o\left(T^{\beta/2} \log^{\frac{\alpha+3}{2}}(T)\right) \text{ a.s.},$$

which is a direct corollary of Lemma 23.

Next we consider the order of $\sum_{t=1}^{T} \eta_t^\top R \eta_t$. Since $\eta_t \sim \mathcal{N}(0, \tau^2 t^{-1+\beta} \log^\alpha(t) I_d)$,

$$\begin{aligned}\mathbb{E}\sum_{t=1}^{T} \eta_t^\top R \eta_t &= \sum_{t=1}^{T} \mathbf{Tr}(\mathbb{E}\eta_t\eta_t^\top R) \\ &= \sum_{t=1}^{T} \tau^2 t^{-1+\beta} \log^\alpha(t) \mathbf{Tr}(R) \\ &\qquad \text{(see the proof in Eq. 81)} \\ &= \tau^2 \frac{T^\beta}{\beta} \log^\alpha(T) \mathbf{Tr}(R)(1 + o(1)).\end{aligned}$$

While the variance of $\sum_{t=1}^{T} \eta_t^\top R\eta_t$ is $\mathcal{O}(\sum_{t=1}^{T} t^{-2+2\beta}\log^{2\alpha}(t)) = \mathcal{O}(T^{-1+2\beta}\log^{2\alpha}(T))$, which means the standard error $\mathcal{O}(T^{-1/2+\beta}\log^{\alpha}(T))$ is of lower order than the expectation. Thus

$$\sum_{t=1}^{T} \eta_t^\top R\eta_t = \tau^2\frac{T^\beta}{\beta}\log^\alpha(T)\,\mathbf{Tr}(R)(1+o_p(1)).$$

As a conclusion, the error caused by this transformation is of order $\tilde{\mathcal{O}}_p(T^\beta)$, and the dominating term is $\sum_{t=1}^{T}\eta_t^\top R\eta_t$.

$$\sum_{t=1}^{T} u_t^\top Ru_t - \tilde{u}_t^\top R\tilde{u}_t = \tau^2\frac{T^\beta}{\beta}\log^\alpha(T)\,\mathbf{Tr}(R)(1+o_p(1)). \tag{37}$$

C.3 Cost of transformed system

Next we proceed as if our system was x_t with system noise $\tilde{\varepsilon}_t = B\eta_t + \varepsilon_t$ and controller $\tilde{u}_t = \hat{K}_t x_t$. The key idea of the following proof is from Appendix C of Fazel et al. (2018).

We are interested in the cost

$$\sum_{t=1}^{T} x_t^\top Qx_t + \tilde{u}_t^\top R\tilde{u}_t \quad \text{with } \tilde{u}_t = \hat{K}_t x_t,$$

which can be written as

$$\begin{aligned}
\sum_{t=1}^{T} x_t^\top Qx_t + \tilde{u}_t^\top R\tilde{u}_t =& \sum_{t=1}^{T} x_t^\top Qx_t + (\hat{K}_t x_t)^\top R\hat{K}_t x_t\\
=& \sum_{t=1}^{T} x_t^\top (Q + \hat{K}_t^\top R\hat{K}_t)x_t\\
=& \sum_{t=1}^{T} \Big[x_t^\top (Q + \hat{K}_t^\top R\hat{K}_t)x_t + x_{t+1}^\top Px_{t+1} - x_t^\top Px_t\Big] + x_1^\top Px_1 - x_{T+1}^\top Px_{T+1}\\
=& \sum_{t=1}^{T} \Big[x_t^\top (Q + \hat{K}_t^\top R\hat{K}_t)x_t + ((A+B\hat{K}_t)x_t + \tilde{\varepsilon}_t)^\top P((A+B\hat{K}_t)x_t + \tilde{\varepsilon}_t) - x_t^\top Px_t\Big]\\
&+ \tilde{\mathcal{O}}_p(1) \quad \text{(by Lemma 18)}\\
=& \sum_{t=1}^{T} \Big[x_t^\top (Q + \hat{K}_t^\top R\hat{K}_t)x_t + x_t^\top (A+B\hat{K}_t)^\top P(A+B\hat{K}_t)x_t - x_t^\top Px_t\\
&+ 2\tilde{\varepsilon}_t^\top P(A+B\hat{K}_t)x_t + \tilde{\varepsilon}_t^\top P\tilde{\varepsilon}_t\Big] + \tilde{\mathcal{O}}_p(1).
\end{aligned} \tag{38}$$

We constructed the specific form of the first term on purpose. The following lemma translates the first term into a quadratic term with respect to $\hat{K}_t - K$.

Lemma 25. *For any $\hat{K}$ with suitable dimension,*

$$\begin{aligned}
&x^\top (Q + \hat{K}^\top R\hat{K})x + x^\top (A+B\hat{K})^\top P(A+B\hat{K})x - x^\top Px\\
&= x^\top (\hat{K}-K)^\top (R + B^\top PB)(\hat{K}-K)x.
\end{aligned}$$

The proof can be found in Appendix H.2.1. As a result

$$\begin{aligned}\sum_{t=1}^{T} x_t^\top Q x_t + \tilde{u}_t^\top R \tilde{u}_t =& \sum_{t=1}^{T} x_t^\top (\hat{K}_t - K)^\top (R + B^\top P B)(\hat{K}_t - K) x_t \\ &+ 2\tilde{\varepsilon}_t^\top P(A + B\hat{K}_t)x_t + \tilde{\varepsilon}_t^\top P \tilde{\varepsilon}_t + \tilde{\mathcal{O}}_p(1).\end{aligned}$$

Now we have three terms, and we will examine them in order.

1. The first term we consider is $\sum_{t=1}^{T} x_t^\top (\hat{K}_t - K)^\top (R + B^\top P B)(\hat{K}_t - K) x_t$. Recall from Lemma 18 that

$$\|x_t\|, \|u_t\| = \mathcal{O}(\log^{1/2}(t)) \text{ a.s.}$$

and

$$\|\hat{K}_t - K\| = \mathcal{O}(t^{-\frac{\beta}{2}} \log^{\frac{-\alpha+1}{2}}(t)) \text{ a.s.}$$

As a result

$$\begin{aligned}&\sum_{t=1}^{T} x_t^\top (\hat{K}_t - K)^\top (R + B^\top P B)(\hat{K}_t - K) x_t \\ \leq & \sum_{t=1}^{T} \|x_t\|^2 \|\hat{K}_t - K\|^2 \|R + B^\top P B\| \\ = & \sum_{t=1}^{T} \mathcal{O}(\log(t)) \mathcal{O}(t^{-\beta} \log^{-\alpha+1}(t)) \text{ a.s.} \\ = & \mathcal{O}(T^{1-\beta} \log^{-\alpha+2}(T)) \text{ a.s.} \qquad \text{(by Eq. 81)}\end{aligned}$$

2. The second term we consider is $\sum_{t=1}^{T} \tilde{\varepsilon}_t^\top P(A + B\hat{K}_t)x_t$. Similar as before, we notice that $\tilde{\varepsilon}_t = \varepsilon_t + B\eta_t \perp\!\!\!\perp (A + B\hat{K}_t)x_t$. Then

$$\mathbb{E} \sum_{t=1}^{T} \tilde{\varepsilon}_t^\top P(A + B\hat{K}_t)x_t = 0.$$

Next consider

$$\begin{aligned}
&\mathbb{E}(\sum_{t=1}^{T}\tilde{\varepsilon}_t^\top P(A+B\hat{K}_t)x_t)^2\\
=&\sum_{t=1}^{T}\mathbb{E}(\tilde{\varepsilon}_t^\top P(A+B\hat{K}_t)x_t)^2\\
\le&\sum_{t=1}^{T}\mathbb{E}\|\tilde{\varepsilon}_t\|^2\|P\|^2\|(A+B\hat{K}_t)\|^2\|x_t\|^2\\
&\quad(\|\hat{K}_t\|\le C_K \text{ based on Algorithm 1 design})\\
\le&\sum_{t=1}^{T}\|P\|^2(\|A\|+\|B\|C_K)^2\mathbb{E}\|\tilde{\varepsilon}_t\|^2\mathbb{E}\|x_t\|^2\\
=&\mathcal{O}(1)\mathbb{E}\sum_{t=1}^{T}\|x_t\|^2\\
&\quad(\text{because of Lemma 35 } \mathbb{E}\sum_{t=1}^{T}\|x_t\|^2=\mathcal{O}(T\log^2(T)))\\
=&\mathcal{O}(T\log^2(T)).
\end{aligned}$$

Thus

$$\sum_{t=1}^{T}\tilde{\varepsilon}_t^\top P(A+B\hat{K}_t)x_t=\mathcal{O}_p(T^{1/2}\log(T)). \tag{39}$$

3. The third term we consider is $\sum_{t=1}^{T}\tilde{\varepsilon}_t^\top P\tilde{\varepsilon}_t$. The expectation is

$$\begin{aligned}
&\mathbb{E}\sum_{t=1}^{T}\tilde{\varepsilon}_t^\top P\tilde{\varepsilon}_t\\
=&\sum_{t=1}^{T}\mathbf{Tr}(P\mathbb{E}\tilde{\varepsilon}_t\tilde{\varepsilon}_t^\top)\\
=&\sum_{t=1}^{T}\mathbf{Tr}(P(\sigma^2 I_n+\tau^2 t^{\beta-1}\log^\alpha(t)BB^\top))\\
=&T\sigma^2\,\mathbf{Tr}(P)+\frac{\tau^2}{\beta}T^\beta\log^\alpha(T)\,\mathbf{Tr}(B^\top PB)(1+o(1))\quad(\text{By Eq. 81}).
\end{aligned}$$

On the other hand, the variance is the sum of variances for each single summand with total order $\mathcal{O}(T)$. As a result, when $\beta>1/2$ or $\beta=1/2,\alpha>0$

$$\sum_{t=1}^{T}\tilde{\varepsilon}_t^\top P\tilde{\varepsilon}_t=T\sigma^2\,\mathbf{Tr}(P)+\frac{\tau^2}{\beta}T^\beta\log^\alpha(T)\,\mathbf{Tr}(B^\top PB)(1+o_p(1)). \tag{40}$$

Summing up all three parts we have: when $\beta > 1/2$, or $\beta = 1/2, \alpha > 1$,

$$\sum_{t=1}^{T} x_t^\top Q x_t + \tilde{u}_t^\top R \tilde{u}_t = T\sigma^2 \mathbf{Tr}(P) + \frac{\tau^2}{\beta} T^\beta \log^\alpha(T) \mathbf{Tr}(B^\top P B)(1 + o_p(1)). \tag{41}$$

Taking the transformation part into consideration (Eq. 37):

$$\sum_{t=1}^{T} u_t^\top R u_t - \tilde{u}_t^\top R \tilde{u}_t = \tau^2 \frac{T^\beta}{\beta} \log^\alpha(T) \mathbf{Tr}(R)(1 + o_p(1)).$$

Finally we have when $\beta > 1/2$, or $\beta = 1/2, \alpha > 1$

$$\begin{aligned}
\mathcal{J}(U,T) &= \frac{1}{T} \left[\sum_{t=1}^{T} x_t^\top Q x_t + u_t^\top R u_t \right] \\
&= \sigma^2 \mathbf{Tr}(P) + \tau^2 \beta^{-1} \mathbf{Tr}(B^\top P B + R) T^{\beta-1} \log^\alpha(T)(1 + o_p(1)).
\end{aligned}$$

Finally we only need to prove that the optimal average cost can be expressed as:

$$\mathcal{J}(U^*,T) = \sigma^2 \mathbf{Tr}(P) + \mathcal{O}_p(T^{-1/2} \log(T)).$$

C.4 Optimal average cost

Denote the states and actions following policy $U^*(H_t) = Kx_t$ as x_t' and u_t'. Following Eq. 38 we know that

$$\begin{aligned}
&\sum_{t=1}^{T} (x_t')^\top Q x_t' + (u_t')^\top R u_t' \\
=& \sum_{t=1}^{T} \Big[(x_t')^\top (Q + K^\top R K) x_t' + (x_t')^\top (A + BK)^\top P (A + BK) x_t' - (x_t')^\top P x_t' \\
&+ 2\varepsilon_t^\top P (A + B\hat{K}_t) x_t + \varepsilon_t^\top P \varepsilon_t \Big] + \tilde{\mathcal{O}}_p(1).
\end{aligned}$$

Following Lemma 25, since our $\hat{K}$ is exactly K:

$$(x_t')^\top (Q + K^\top R K) x_t' + (x_t')^\top (A + BK)^\top P (A + BK) x_t' - (x_t')^\top P x_t' = 0$$

The remaining terms can be considered in exactly same way as Eq. 39 and Eq. 40, which turn out to be:

$$\sum_{t=1}^{T} \tilde{\varepsilon}_t^\top P (A + B\hat{K}_t) x_t = \mathcal{O}_p(T^{1/2} \log(T)),$$

and

$$\sum_{t=1}^{T} \varepsilon_t^\top P \varepsilon_t = T\sigma^2 \mathbf{Tr}(P) + \mathcal{O}_p(T^{1/2}).$$

Finally we arrive at the conclusion that

$$\begin{aligned}\mathcal{J}(U^*, T) &= \frac{1}{T}\left(\sum_{t=1}^{T}(x_t')^\top Q x_t' + (u_t')^\top R u_t'\right)\\ &= \frac{1}{T}\left(\mathcal{O}_p(T^{1/2}\log(T)) + T\sigma^2\,\mathbf{Tr}(P) + \mathcal{O}_p(T^{1/2})\right)\\ &= \sigma^2\,\mathbf{Tr}(P) + \mathcal{O}_p(T^{-1/2}\log(T)).\end{aligned}$$

Appendix D. The proof of Theorem 5

Theorem. *Algorithm 1 applied to a system described by Eq. 1 under Assumption 1 satisfies, as $t \to \infty$,*

$$\mathrm{vec}\left[\left[\hat{A}_t - A, \hat{B}_t - B\right] D_t\right] \xrightarrow{D} \mathcal{N}(0, \sigma^2 I_{n(n+d)}).$$

Proof One can find the definition of D_t in Eq. 7. The proof heavily relies on the following theorems from Anderson and Kunitomo (1992). For better understanding, we directly state those theorems with the same notation as our paper.

Theorem 26 (Theorems 1 and 3 in Anderson and Kunitomo (1992))**.** *Let $\{x_i, u_i, \varepsilon_i\}$, $i = 0, 1\cdots$, be a sequence of random vectors described by Eq. 1 under Assumption 1, and let $\{\mathcal{F}_i\}$ be an increasing sequence of σ-fields such that $\{x_i, u_i\}$ is $\mathcal{F}_{i-1}$ measureable and ε_i is $\mathcal{F}_i$ measurable. Let the matrix D_t be a deterministic matrix such that*

$$D_t^{-1}\sum_{i=0}^{t-1}\begin{bmatrix}x_i\\u_i\end{bmatrix}\begin{bmatrix}x_i\\u_i\end{bmatrix}^\top (D_t^\top)^{-1} \xrightarrow{P} C, \tag{42}$$

where C is a constant matrix, and

$$\max_{1\le i\le t}\begin{bmatrix}x_i\\u_i\end{bmatrix}^\top (D_t D_t^\top)^{-1}\begin{bmatrix}x_i\\u_i\end{bmatrix} \xrightarrow{P} 0. \tag{43}$$

Suppose further that $\mathbb{E}(\varepsilon_i|\mathcal{F}_{i-1}) = 0$ a.s., $\mathbb{E}(\varepsilon_i\varepsilon_i^\top|\mathcal{F}_{i-1}) = \Sigma_i$ a.s.,

$$\sum_{i=0}^{t-1}\left[\Sigma_i \otimes D_t^{-1}\begin{bmatrix}x_i\\u_i\end{bmatrix}\begin{bmatrix}x_i\\u_i\end{bmatrix}^\top (D_t^\top)^{-1}\right] \xrightarrow{P} \Sigma\otimes C, \tag{44}$$

where Σ is a constant positive semi-definite matrix and

$$\sup_{i\ge 1}\mathbb{E}\left[\varepsilon_i^\top\varepsilon_i \mathbf{1}_{\varepsilon_i^\top\varepsilon_i > a}|\mathcal{F}_{i-1}\right] \xrightarrow{P} 0, \tag{45}$$

as $a \to \infty$. Then

$$\mathrm{vec}\left[\left[\hat{A}_t - A, \hat{B}_t - B\right] D_t\right] \xrightarrow{D} \mathcal{N}(0, C^{-1}\otimes\Sigma). \tag{46}$$

As we have seen in Algorithm 1 the controller $\hat{K}_t$ is fully determined by $\{x_i, u_i\}_{i=0}^{t-1}$. Pick

$$\mathcal{F}_{t-1} = \sigma(\{x_i, u_i, \eta_i\}_{i=0}^{t}, \{\varepsilon_i\}_{i=0}^{t-1}).$$

Now we verified the design vector $\begin{bmatrix} x_t \\ u_t \end{bmatrix}$ at stage t is $\mathcal{F}_{t-1}$ measurable. Since $\varepsilon_t \overset{\text{i.i.d.}}{\sim} \mathcal{N}(0, \sigma^2 I_d)$, we know that $\varepsilon_t \perp\!\!\!\perp \mathcal{F}_{t-1}$, and $\{\varepsilon_t\}$ is a martingale difference sequence with respect to an increasing sequence of σ-fields $\{\mathcal{F}_t\}$. Eq. 44 holds by the fact that all variances $\Sigma_i = \sigma^2 I_d$ and Eq. 42. For Eq. 45, notice that we can remove the sup since every term has the same value, so the conclusion follows from a standard property of Gaussian distributions.

Actually, Eq. 42 is already shown in Theorem 3. Eq. 43 requires less effort to prove as we defined D_t by

$$D_t := t^{\beta/2} \log^{\alpha/2}(t) \left[\begin{array}{cc} I_n & 0 \\ K & I_d \end{array} \right] \left[\begin{array}{cc} C_t^{1/2} & 0 \\ 0 & \sqrt{\frac{\tau^2}{\beta}} I_d \end{array} \right]. \tag{47}$$

As a result, Eq. 43 is not surprising since z_t should be only of constant order.

D.1 The proof of Eq. 43

Since

$$D_t D_t^\top = t^\beta \log^\alpha(t) \left[\begin{array}{cc} I_n & 0 \\ K & I_d \end{array} \right] \left[\begin{array}{cc} C_t & 0 \\ 0 & \frac{\tau^2}{\beta} I_d \end{array} \right] \left[\begin{array}{cc} I_n & K^\top \\ 0 & I_d \end{array} \right],$$

we have

$$\begin{aligned} (D_t D_t^\top)^{-1} =& t^{-\beta} \log^{-\alpha}(t) \left[\begin{array}{cc} I_n & -K^\top \\ 0 & I_d \end{array} \right] \left[\begin{array}{cc} C_t^{-1} & 0 \\ 0 & \frac{\tau^2}{\beta} I_d \end{array} \right] \left[\begin{array}{cc} I_n & 0 \\ -K & I_d \end{array} \right] \\ =& \mathcal{O}(t^{-\beta} \log^{-\alpha}(t)) \quad \text{(by Eq. 33)}. \end{aligned} \tag{48}$$

Recall that Eq. 43 is

$$\max_{1 \le i \le t} \begin{bmatrix} x_i \\ u_i \end{bmatrix}^\top (D_t D_t^\top)^{-1} \begin{bmatrix} x_i \\ u_i \end{bmatrix} \xrightarrow{P} 0.$$

It suffices to show

$$t^{-\beta/2} \log^{-\alpha/2}(t) \max_{1 \le i \le t} \|x_i\| \xrightarrow{P} 0 \ \text{ and } \ t^{-\beta/2} \log^{-\alpha/2}(t) \max_{1 \le i \le t} \|u_i\| \xrightarrow{P} 0.$$

Actually we already shown in Lemma 18 that

$$\|x_t\|, \|u_t\| = \mathcal{O}(\log^{1/2}(t)) \text{ a.s.}$$

This is a uniform bound over t, thus a direct corollary is

$$\max_{1 \le i \le t} \|x_i\|, \max_{1 \le i \le t} \|u_i\| = \mathcal{O}(\log^{1/2}(t)) \text{ a.s.}$$

That immediately implies

$$t^{-\beta/2} \max_{1 \le i \le t} \|x_i\| \xrightarrow{a.s.} 0 \ \text{ and } \ t^{-\beta/2} \max_{1 \le i \le t} \|u_i\| \xrightarrow{a.s.} 0.$$

■

Appendix E. The proof of Theorem 9

Here we state and prove a generalization of Theorem 9 that allows for the case when $\beta = 1$ and $\alpha \le 0$.

Theorem. *Algorithm 1 applied to a system described by Eq. 1 under Assumption 1 satisfies, as $t \to \infty$,*

$$\left(x_t^\top \left(\sum_{p=0}^{\infty} (A+BK)^p \left(I_n + 1_{\{\beta=1,\alpha=0\}} \frac{\tau^2}{\sigma^2} BB^\top \right) ((A+BK)^p)^\top \right)^{-1} x_t + \beta\sigma^2 \|w_t\|^2 \right)^{-1/2} \cdot t^{1/2} \left((\hat{A}_t - A)x_t + (\hat{B}_t - B)u_t \right) \xrightarrow{D} \mathcal{N}(0, I_n). \tag{49}$$

Proof We can generalize the input noise η_t to ξ_t which is any random vector independent of the data before t: $\{\varepsilon_i, \eta_i\}_{i=0}^{t-1}$. Hereafter, $u_t = \hat{K}_t x_t + \xi_t$ (but u_i for $i < t$ is still $\hat{K}_i x_i + \eta_i$).

The proof will proceed by showing that $(\hat{A}_t, \hat{B}_t)$ acts as if it were independent of (x_t, u_t), and then effectively conditioning on (x_t, u_t) and using $(\hat{A}_t, \hat{B}_t)$'s asymptotic distribution from Theorem 5.

Define $\rho_L := \frac{2+\rho(L)}{3}$ as in Lemma 18. Define replacements of x_t and u_t which are independent of $\hat{A}_{t-\left\lfloor -\frac{\log(t)}{\log(\rho_L)} \right\rfloor}$ and $\hat{B}_{t-\left\lfloor -\frac{\log(t)}{\log(\rho_L)} \right\rfloor}$:

$$\tilde{x}_t := \sum_{p=t-\left\lfloor -\frac{\log(t)}{\log(\rho_L)} \right\rfloor}^{t-1} (A+BK)^{t-p-1} (B\eta_p + \varepsilon_p), \tag{50}$$

and

$$\tilde{u}_t := K\tilde{x}_t + \xi_t = K \sum_{p=t-\left\lfloor -\frac{\log(t)}{\log(\rho_L)} \right\rfloor}^{t-1} (A+BK)^{t-p-1} (B\eta_p + \varepsilon_p) + \xi_t. \tag{51}$$

We can show that the difference between $\tilde{x}_t, \tilde{u}_t$ and x_t, u_t is very small:

Lemma 27.

$$x_t = \tilde{x}_t + O(t^{-\frac{\beta}{2}} \log^{\frac{-\alpha+2}{2}}(t)) \ \textit{a.s.}$$

$$u_t = \tilde{u}_t + O(t^{-\frac{\beta}{2}} \log^{\frac{-\alpha+2}{2}}(t)) \ \textit{a.s.}$$

The proof can be found in Appendix H.3.1. At the same time, the difference between $\hat{A}_{t-\left\lfloor -\frac{\log(t)}{\log(\rho_L)} \right\rfloor}$, $\hat{B}_{t-\left\lfloor -\frac{\log(t)}{\log(\rho_L)} \right\rfloor}$ and $\hat{A}_t, \hat{B}_t$ is also small:

Lemma 28.

$$\hat{A}_t = \hat{A}_{t-\left\lfloor -\frac{\log(t)}{\log(\rho_L)} \right\rfloor} + \mathcal{O}_p(t^{-\beta} \log^{-\alpha+3/2}(t)).$$

$$\hat{B}_t = \hat{B}_{t-\left\lfloor -\frac{\log(t)}{\log(\rho_L)} \right\rfloor} + \mathcal{O}_p(t^{-\beta} \log^{-\alpha+3/2}(t)).$$

The proof can be found in Appendix H.3.2. These substitutions are very close to our original concern, and they have the good independence property:

$$\left(\hat{A}_{t-\left\lfloor -\frac{\log(t)}{\log(\rho_L)}\right\rfloor} - A, \hat{B}_{t-\left\lfloor -\frac{\log(t)}{\log(\rho_L)}\right\rfloor} - B\right) \perp\!\!\!\perp (\tilde{x}_t, \tilde{u}_t).$$

This is because $\hat{A}_{t-\left\lfloor -\frac{\log(t)}{\log(\rho_L)}\right\rfloor}$ and $\hat{B}_{t-\left\lfloor -\frac{\log(t)}{\log(\rho_L)}\right\rfloor}$ are only functions of the system up to time $t - \left\lfloor -\frac{\log(t)}{\log(\rho_L)}\right\rfloor - 1$, while $\tilde{x}_t$ and $\tilde{u}_t$ are independent with event before time $t - \left\lfloor -\frac{\log(t)}{\log(\rho_L)}\right\rfloor$ by definitions in Eq. 50 and 51. Our initial target is to identify the distribution of $(\hat{A}_t - A)x_t + (\hat{B}_t - B)u_t$. We will start from its substitution

$$\begin{aligned}
&(\hat{A}_{t-\left\lfloor -\frac{\log(t)}{\log(\rho_L)}\right\rfloor} - A)\tilde{x}_t + (\hat{B}_{t-\left\lfloor -\frac{\log(t)}{\log(\rho_L)}\right\rfloor} - B)\tilde{u}_t \\
=&((\hat{A}_{t-\left\lfloor -\frac{\log(t)}{\log(\rho_L)}\right\rfloor} - A) + K(\hat{B}_{t-\left\lfloor -\frac{\log(t)}{\log(\rho_L)}\right\rfloor} - B))\tilde{x}_t + (\hat{B}_{t-\left\lfloor -\frac{\log(t)}{\log(\rho_L)}\right\rfloor} - B)\xi_t.
\end{aligned}$$

Because of this independence after substitution, the first term is independent with the second term, and their asymptotic distribution can be described by Eq. 11.

Lemma 29. *For any ξ_t independent of the data before t: $\{\varepsilon_i, \eta_i\}_{i=0}^{t-1}$:*

$$\left(\tilde{x}_t^\top \left(\sum_{p=0}^{\infty} L^p \left(I_n + 1_{\{\beta=1,\alpha=0\}}\frac{\tau^2}{\sigma^2}BB^\top\right)(L^p)^\top\right)^{-1} \tilde{x}_t + \frac{\beta\sigma^2}{\tau^2}t^{1-\beta}\log^{-\alpha}(t)\left\|\xi_t\right\|^2\right)^{-1/2}$$
$$\cdot\, t^{1/2}\left[(\hat{A}_{t-\left\lfloor -\frac{\log(t)}{\log(\rho_L)}\right\rfloor} - A)\tilde{x}_t + (\hat{B}_{t-\left\lfloor -\frac{\log(t)}{\log(\rho_L)}\right\rfloor} - B)(K\tilde{x}_t + \xi_t)\right] \xrightarrow{D} \mathcal{N}(0, I_n).$$

The proof of Lemma 29 can be found in Appendix H.3.3. With the help of Lemma 27 and Lemma 28 , we can change all the replacements back to the original form:

Lemma 30. *For any ξ_t independent of the data before t: $\{\varepsilon_i, \eta_i\}_{i=0}^{t-1}$,*

$$\left(x_t^\top \left(\sum_{p=0}^{\infty} L^p \left(I_n + 1_{\{\beta=1,\alpha=0\}}\frac{\tau^2}{\sigma^2}BB^\top\right)(L^p)^\top\right)^{-1} x_t + \frac{\beta\sigma^2}{\tau^2}t^{1-\beta}\log^{-\alpha}(t)\left\|\xi_t\right\|^2\right)^{-1/2}$$
$$\cdot\, t^{1/2}\left[(\hat{A}_t - A)x_t + (\hat{B}_t - B)(\hat{K}_t x_t + \xi_t)\right] \xrightarrow{D} \mathcal{N}(0, I_n).$$

The proof of Lemma 30 can be found in Appendix H.3.4. Since η_t is independent with$\{\varepsilon_i, \eta_i\}_{i=0}^{t-1}$, which satisfies the condition of ξ_t, we can restate the result with η_t replaced by ξ_t:

$$\left(x_t^\top \left(\sum_{p=0}^{\infty} L^p \left(I_n + 1_{\{\beta=1,\alpha=0\}}\frac{\tau^2}{\sigma^2}BB^\top\right)(L^p)^\top\right)^{-1} x_t + \frac{\beta\sigma^2}{\tau^2}t^{1-\beta}\log^{-\alpha}(t)\left\|\eta_t\right\|^2\right)^{-1/2}$$
$$\cdot\, t^{1/2}\left[(\hat{A}_t - A)x_t + (\hat{B}_t - B)(\hat{K}_t x_t + \eta_t)\right] \xrightarrow{D} \mathcal{N}(0, I_n).$$

Finally, we have the desired conclusion using $\eta_t = \tau\sqrt{t^{\beta-1}\log^{\alpha}(t)}\, w_t$:

$$\left(x_t^\top \left(\sum_{p=0}^{\infty} L^p \left(I_n + 1_{\{\beta=1,\alpha=0\}}\frac{\tau^2}{\sigma^2}BB^\top\right)(L^p)^\top\right)^{-1} x_t + \beta\sigma^2 \left\|w_t\right\|^2\right)^{-1/2}$$

$$\cdot\, t^{1/2}\left((\hat{A}_t - A)x_t + (\hat{B}_t - B)u_t\right) \xrightarrow{D} \mathcal{N}(0, I_n).$$

■

Appendix F. The proof of Corollaries

F.1 The proof of Corollary 8

Corollary. *Assume $A+BK$ is full rank. Algorithm 1 applied to a system described by Eq. 1 under Assumption 1 satisfies*

$$\sqrt{\frac{\tau^2}{\sigma^2\beta}} t^{\beta/2}\log^{\alpha/2}(t)\left(\left(\frac{dK}{d[A,B]}\right)\left(\begin{bmatrix}-K^\top \\ I_d\end{bmatrix}\otimes I_n\right)\right)^{-1}\operatorname{vec}\left(\hat{K}_t - K\right)\xrightarrow{D}\mathcal{N}(0, I_{nd}).$$

Proof Before we prove this result, we should first examine that the matrix $\left(\frac{dK}{d[A,B]}\right)\left(\begin{bmatrix}-K^\top \\ I_d\end{bmatrix}\otimes I_n\right)$ is indeed invertible. Since $\left(\begin{bmatrix}-K^\top \\ I_d\end{bmatrix}\otimes I_n\right)$ has an identity matrix component I_{dn}, it is sufficient to show that $\frac{dK}{d[A,B]}$ is full rank.

F.1.1 $\frac{dK}{d[A,B]}$ is full rank

We can ignore the effect of K_0 and consider $\hat{K}_t$ to be the same as certainty equivalent controller $\tilde{K}_t$ which is directly calculated by plugging $\hat{A}_{t-1}, \hat{B}_{t-1}$ into DARE Eqs. 3 and 4. This is because $\hat{K}_t = K_0$ only happens finitely often and thus does not affect asymptotic properties; see Appendix G.3.

Before we start, we need to define how we solve $\frac{dK}{d[A,B]} \in \mathbb{R}^{nd\times n(n+d)}$ and then prove that $\frac{dK}{d[A,B]}$ is indeed a full rank matrix. Lemmas 3.1 and B.1 from Simchowitz and Foster (2020) gives the relationship between the derivatives of K, P, A, B:

$$dK = -(R + B^\top PB)^{-1}(dB^\top P(A+BK) + B^\top P(dA + dBK) + B^\top dP(A+BK)), \quad (52)$$

where dP can be solved from

$$(A+BK)^\top dP(A+BK) - dP + (dA+dBK)^\top P(A+BK) + (A+BK)^\top P(dA+dBK) = 0. \quad (53)$$

Now we can solve $\frac{dK}{d[A,B]}$ by Eq. 52 and Eq. 53. Denote the kernel space of the derivative matrix $\frac{dK}{d[A,B]}$ as $\mathcal{S}$. It suffices to show that $\mathcal{S}$'s dimension is $n(n+d) - nd = n^2$, which implies $\frac{dK}{d[A,B]}$ is full rank with rank nd. The equivalent definition of kernel space $\mathcal{S}$ is the small perturbation vec$[dA, dB]$ such that K does not change ($dK = 0$):

$$dK = \frac{dK}{d[A,B]}\text{vec}(dA, dB) = 0.$$

Any vector in kernel space $\mathcal{S}$ can be considered as vec$[dA, dB]$ which satisfies $dK = 0$ in Eq. 52, and that means:

$$dB^\top P(A+BK) + B^\top P(dA + dBK) + B^\top dP(A+BK) = 0. \tag{54}$$

On the other hand, Eq. 53 describes a linear recursive relationship between dP and $dA + dBK$, so that we can solve dP with the infinite summation:

$$\begin{aligned}
dP =&(A+BK)^\top dP(A+BK) + (dA+dBK)^\top P(A+BK) + (A+BK)^\top P(dA+dBK)\\
=&((A+BK)^\top)^2 dP(A+BK)^2\\
&+ (A+BK)^\top \Big((dA+dBK)^\top P(A+BK) + (A+BK)^\top P(dA+dBK)\Big)(A+BK)\\
&+ (dA+dBK)^\top P(A+BK) + (A+BK)^\top P(dA+dBK)\\
&\text{(recursively plugging in the first equation)}\\
=&\sum_{i=0}^{\infty}((A+BK)^\top)^i \Big((dA+dBK)^\top P(A+BK) + (A+BK)^\top P(dA+dBK)\Big)(A+BK)^i.
\end{aligned}$$

Also recall that $A + BK$ is assumed to be full rank matrix, and we can show that P is also full rank; see Appendix G.2. Thus we can explicitly solve dB from Eq. 54 as a linear equation with regard to $dA + dBK$:

$$dB^\top = -(P(A+BK))^{-1}(B^\top P(dA+dBK) + B^\top dP(A+BK)).$$

This tells us the kernel space $\mathcal{S}$ is the image of a function of its linear subspace $dA+dBK \in \mathbb{R}^{n^2}$, which means $dim(\mathcal{S}) \le n^2$. Notice by kernel space definition its dimension should be at least $dim(\mathcal{S}) \ge n(n+d) - nd = n^2$, where the equality is achieved when $\frac{dK}{d[A,B]}$ has full rank nd. Combining these two equations we have $dim(\mathcal{S}) = n^2$. Finally we arrived at the desired conclusion that dimension of $\frac{dK}{d[A,B]} \in \mathbb{R}^{nd\times n(n+d)}$'s kernel space $\mathcal{S}$ is exactly n^2, which means $\frac{dK}{d[A,B]}$ is full rank.

Next we describe the rest of the proof:

F.1.2 Proof by the Delta method

By Taylor expansion and the consistency of $[\hat{A}_t, \hat{B}_t]$ (see Proposition 15), we have

$$\text{vec}\left(\hat{K}_t - K\right) = \left(\frac{dK}{d[A,B]}\right)\text{vec}\left[\hat{A}_t - A, \hat{B}_t - B\right](1 + o_p(1)).$$

From Remark 6 we know

$$\hat{A}_t - A = (\hat{B}_t - B)(-K)(1 + o_p(1)).$$

Then

$$\text{vec}\left(\hat{K}_t - K\right) = \left(\frac{dK}{d[A,B]}\right)\text{vec}\left((\hat{B}_t - B)\begin{bmatrix}-K, & I_d\end{bmatrix}\right)(1+o_p(1)).$$

which can be written as

$$\text{vec}\left(\hat{K}_t - K\right) = \left(\frac{dK}{d[A,B]}\right)\left(\begin{bmatrix}-K^\top \\ I_d\end{bmatrix}\otimes I_n\right)\text{vec}\left(\hat{B}_t - B\right)(1+o_p(1)).$$

By Eq. 11,

$$\sqrt{\frac{\tau^2}{\sigma^2\beta}}t^{\beta/2}\log^{\alpha/2}(t)\text{vec}\left(\hat{B}_t - B\right) \xrightarrow{D} \mathcal{N}(0, I_{nd}).$$

Combining the above two equations, finally we have

$$\sqrt{\frac{\tau^2}{\sigma^2\beta}}t^{\beta/2}\log^{\alpha/2}(t)\text{vec}\left(\hat{K}_t - K\right) \xrightarrow{D} \left(\frac{dK}{d[A,B]}\right)\left(\begin{bmatrix}-K^\top \\ I_d\end{bmatrix}\otimes I_n\right)\mathcal{N}(0, I_{nd}).$$

From the fact that $\frac{dK}{d[A,B]}$ is full rank and that $\left(\begin{bmatrix}-K^\top \\ I_d\end{bmatrix}\otimes I_n\right)$ has an identity matrix component I_{dn}, we can take matrix inverse and get

$$\sqrt{\frac{\tau^2}{\sigma^2\beta}}t^{\beta/2}\log^{\alpha/2}(t)\left(\left(\frac{dK}{d[A,B]}\right)\left(\begin{bmatrix}-K^\top \\ I_d\end{bmatrix}\otimes I_n\right)\right)^{-1}\text{vec}\left(\hat{K}_t - K\right) \xrightarrow{D} \mathcal{N}(0, I_{nd}).$$

■

F.2 The proof of Corollary 10

Corollary. *The average regret of the controller U defined by Algorithm 1 applied through time horizon T to a system described by Eq. 1 under Assumption 1 satisfies, as $t \to \infty$ and $T \to \infty$,*

$$\frac{\mathcal{R}(U,T)}{\tau^2\beta^{-1}\,\mathbf{Tr}(\hat{B}_t^\top \hat{P}_t \hat{B}_t + R)T^{\beta-1}\log^\alpha(T)} \xrightarrow{P} 1. \tag{55}$$

Proof This is a direct corollary from Theorem 4, which states

$$\frac{\mathcal{R}(U,T)}{\tau^2\beta^{-1}\,\mathbf{Tr}(B^\top PB + R)T^{\beta-1}\log^\alpha(T)} \xrightarrow{P} 1,$$

and from Proposition 15 and Corollary 40 which implies the consistency of $\hat{B}_t$ and $\hat{P}_t$. By Slutsky's theorem we can replace the parameters B and P in Eq. 55 with $\hat{B}_t$ and $\hat{P}_t$. ■

F.3 The proof of Corollary 11

Corollary. *Algorithm 1 applied to a system described by Eq. 1 under Assumption 1 satisfies*

$$\mathbf{Tr}\left(\left[\hat{A}_t - A, \hat{B}_t - B\right] \sum_{i=0}^{t-1} \begin{bmatrix} x_i \\ u_i \end{bmatrix} \begin{bmatrix} x_i \\ u_i \end{bmatrix}^\top \left[\hat{A}_t - A, \hat{B}_t - B\right]^\top\right) \xrightarrow{D} \sigma^2 \chi^2_{n(n+d)}.$$

Proof For notational simplicity denote $\hat{\Theta}_t := \left[\hat{A}_t, \hat{B}_t\right]$ and $\Theta := \left[A, B\right]$. By Theorem 5 we know

$$\operatorname{vec}\left((\hat{\Theta}_t - \Theta)D_t\right) \xrightarrow{D} \mathcal{N}(0, \sigma^2 I_{n(n+d)}). \tag{56}$$

Potentially we can derive an ellipsoid "confidence region" with the above formula by

$$\mathbf{Tr}\left((\hat{\Theta}_t - \Theta)\left(D_t D_t^\top\right)(\hat{\Theta}_t - \Theta)^\top\right) \xrightarrow{D} \sigma^2 \chi^2_{n(n+d)}. \tag{57}$$

However, since a true confidence region should not require any knowledge on oracle parameters, we need to replace $D_t D_t^\top$ with some observable expression, which turns out to be:

$$\mathbf{Tr}\left((\hat{\Theta}_t - \Theta)\left(\sum_{i=0}^{t-1} \begin{bmatrix} x_i \\ u_i \end{bmatrix} \begin{bmatrix} x_i \\ u_i \end{bmatrix}^\top\right)(\hat{\Theta}_t - \Theta)^\top\right) \xrightarrow{D} \sigma^2 \chi^2_{n(n+d)}.$$

Next we will explain why it is valid to replace $D_t D_t^\top$ by $\sum_{i=0}^{t-1} \begin{bmatrix} x_i \\ u_i \end{bmatrix} \begin{bmatrix} x_i \\ u_i \end{bmatrix}^\top$. We know from Eq. 57 that

$$\mathbf{Tr}\left((\hat{\Theta}_t - \Theta)\left(D_t I_{n+d} D_t^\top\right)(\hat{\Theta}_t - \Theta)^\top\right) \xrightarrow{D} \sigma^2 \chi^2_{n(n+d)},$$

and we can replace I_{n+d} by $D_t^{-1} \sum_{i=0}^{t-1} \begin{bmatrix} x_i \\ u_i \end{bmatrix} \begin{bmatrix} x_i \\ u_i \end{bmatrix}^\top (D_t^\top)^{-1} + o_p(1)$ thanks to Theorem 3. As a result,

$$\mathbf{Tr}\left((\hat{\Theta}_t - \Theta)D_t\left(D_t^{-1} \sum_{i=0}^{t-1} \begin{bmatrix} x_i \\ u_i \end{bmatrix} \begin{bmatrix} x_i \\ u_i \end{bmatrix}^\top (D_t^\top)^{-1} + o_p(1)\right) D_t^\top (\hat{\Theta}_t - \Theta)^\top\right) \xrightarrow{D} \sigma^2 \chi^2_{n(n+d)}.$$

By Eq. 56, vec $\left((\hat{\Theta}_t - \Theta)D_t\right)$ is of constant order, and thus the $o_p(1)$ can be ignored. Finally, we have

$$\mathbf{Tr}\left((\hat{\Theta}_t - \Theta)\left(\sum_{i=0}^{t-1} \begin{bmatrix} x_i \\ u_i \end{bmatrix} \begin{bmatrix} x_i \\ u_i \end{bmatrix}^\top\right)(\hat{\Theta}_t - \Theta)^\top\right) \xrightarrow{D} \sigma^2 \chi^2_{n(n+d)}.$$

■

F.4 The proof of Corollary 12

Corollary. *Algorithm 1 applied to a system described by Eq. 1 under Assumption 1 satisfies*

$$\mathrm{vec}(\hat{K}_t - K)^\top \left(\left(\frac{dK}{d[A,B]} \right)_t \left(\sum_{i=0}^{t-1} \begin{bmatrix} x_i \\ u_i \end{bmatrix} \begin{bmatrix} x_i \\ u_i \end{bmatrix}^\top \otimes I_n \right)^{-1} \left(\frac{dK}{d[A,B]} \right)_t^\top \right)^{-1} \mathrm{vec}(\hat{K}_t - K) \xrightarrow{D} \sigma^2 \chi^2_{nd}, \tag{58}$$

where $\left(\frac{dK}{d[A,B]}\right)_t \in \mathbb{R}^{nd \times n(n+d)}$ *is defined as* $\frac{dK}{d[A,B]}$ *evaluated at* $\hat{A}_{t-1}, \hat{B}_{t-1}$.

Proof Again, let us denote $\hat{\Theta}_t := \big[\hat{A}_t, \hat{B}_t\big]$ and $\Theta := \big[A, B\big]$. Starting from Theorem 5

$$\mathrm{vec}\left((\hat{\Theta}_t - \Theta) D_t \right) \xrightarrow{D} \mathcal{N}(0, \sigma^2 I_{n(n+d)}),$$

we need to transfer D_t to its observable version in terms of the Gram matrix. More specifically, we need to find another matrix E_t which is observable and satisfies:

- $D_t^{-1} E_t \xrightarrow{P} I_{n+d}$ because we want to use Slutsky's theorem.
- $E_t E_t^\top = \sum_{i=0}^{t-1} \begin{bmatrix} x_i \\ u_i \end{bmatrix} \begin{bmatrix} x_i \\ u_i \end{bmatrix}^\top$ because $D_t^{-1} \sum_{i=0}^{t-1} \begin{bmatrix} x_i \\ u_i \end{bmatrix} \begin{bmatrix} x_i \\ u_i \end{bmatrix}^\top (D_t^\top)^{-1} \xrightarrow{P} I_{n+d}$.

For now let us assume we have already found such matrix E_t, and thus we can replace D_t with E_t:

$$\mathrm{vec}\left((\hat{\Theta}_t - \Theta) E_t \right) \xrightarrow{D} \mathcal{N}(0, \sigma^2 I_{n(n+d)}).$$

That is:

$$(E_t^\top \otimes I_n) \mathrm{vec}\left(\hat{\Theta}_t - \Theta \right) \xrightarrow{D} \mathcal{N}(0, \sigma^2 I_{n(n+d)}).$$

Further denote $F_t := E_t^\top \otimes I_n$, and then

$$F_t \mathrm{vec}\left(\hat{\Theta}_t - \Theta \right) \xrightarrow{D} \mathcal{N}(0, \sigma^2 I_{n(n+d)}). \tag{59}$$

By Taylor expansion and the consistency of $\hat{\Theta}_t$ (see Proposition 15), we have

$$\mathrm{vec}\left(\hat{K}_t - K \right) = \left(\frac{dK}{d\Theta} \right)_t \mathrm{vec}\left(\hat{\Theta}_t - \Theta \right) (1 + o_p(1)).$$

Since we will prove $D_t^{-1} E_t \xrightarrow{P} I_{n+d}$ in Appendix F.4.1, E_t is asymptotically invertible, which means we can take inverse of $F_t = E_t^\top \otimes I_n$ in asymptotic equations:

$$\mathrm{vec}\left(\hat{K}_t - K \right) = \left(\frac{dK}{d\Theta} \right)_t (F_t)^{-1} F_t \mathrm{vec}\left(\hat{\Theta}_t - \Theta \right) (1 + o_p(1)).$$

We have already shown in Appendix F.1 that $\frac{dK}{d\Theta}$ is full rank, in the same way we can prove that $\left(\frac{dK}{d\Theta}\right)_t$ is almost surely full rank (the only difference is that we replaced A, B with $\hat{A}_{t-1}, \hat{B}_{t-1}$). Recall the QR decomposition, we can re-express $\left(\frac{dK}{d\Theta}\right)_t (F_t)^{-1}$ as $\left(\frac{dK}{d\Theta}\right)_t (F_t)^{-1} =$

Q_tU_t, where $Q_t \in \mathbb{R}^{nd\times nd}$ is an invertible matrix, and $U_t \in \mathbb{R}^{nd\times n(n+d)}$ satisfies $U_tU_t^\top = I_{nd}$. This implies that

$$\text{vec}\left(\hat{K}_t - K\right) = Q_tU_tF_t\text{vec}\left(\hat{\Theta}_t - \Theta\right)(1+o(1)) \text{ a.s.}$$

From this and Eq. 59 we know

$$Q_t^{-1}\text{vec}\left(\hat{K}_t - K\right) = U_tF_t\text{vec}\left(\hat{\Theta}_t - \Theta\right)(1+o(1)) \xrightarrow{D} \mathcal{N}(0, \sigma^2 I_{nd}).$$

That is,

$$\text{vec}\left(\hat{K}_t - K\right)^\top (Q_t^\top)^{-1}Q_t^{-1}\text{vec}\left(\hat{K}_t - K\right) \xrightarrow{D} \sigma^2\chi^2_{nd}.$$

$$\text{vec}\left(\hat{K}_t - K\right)^\top (Q_tU_tU_t^\top Q_t^\top)^{-1}\text{vec}\left(\hat{K}_t - K\right) \xrightarrow{D} \sigma^2\chi^2_{nd}.$$

Recall that $\left(\frac{dK}{d\Theta}\right)_t (F_t)^{-1} = Q_tU_t$, and thus

$$\text{vec}\left(\hat{K}_t - K\right)^\top \left(\left(\frac{dK}{d\Theta}\right)_t (F_t^\top F_t)^{-1}\left(\frac{dK}{d\Theta}\right)_t^\top\right)^{-1} \text{vec}\left(\hat{K}_t - K\right) \xrightarrow{D} \sigma^2\chi^2_{nd}.$$

By definition

$$\begin{aligned} F_t^\top F_t =& (E_t^\top \otimes I_n)^\top (E_t^\top \otimes I_n) \\ =& (E_t \otimes I_n)(E_t^\top \otimes I_n) \\ =& E_tE_t^\top \otimes I_n \\ =& \sum_{i=0}^{t-1} \begin{bmatrix} x_i \\ u_i \end{bmatrix} \begin{bmatrix} x_i \\ u_i \end{bmatrix}^\top \otimes I_n. \end{aligned}$$

Finally we can say

$$\text{vec}\left[\hat{K}_t - K\right]^\top \left(\left(\frac{dK}{d\Theta}\right)_t \left(\sum_{i=0}^{t-1} \begin{bmatrix} x_i \\ u_i \end{bmatrix} \begin{bmatrix} x_i \\ u_i \end{bmatrix}^\top \otimes I_n\right)^{-1} \left(\frac{dK}{d\Theta}\right)_t^\top\right)^{-1} \text{vec}\left[\hat{K}_t - K\right] \xrightarrow{D} \sigma^2\chi^2_{nd}.$$

The only remaining task is to find a valid E_t which satisfies $D_t^{-1}E_t \xrightarrow{P} I_{n+d}$ and $E_tE_t^\top = \sum_{i=0}^{t-1} \begin{bmatrix} x_i \\ u_i \end{bmatrix} \begin{bmatrix} x_i \\ u_i \end{bmatrix}^\top$. Although we already have Theorem 3, $E_t = \left(\sum_{i=0}^{t-1} \begin{bmatrix} x_i \\ u_i \end{bmatrix} \begin{bmatrix} x_i \\ u_i \end{bmatrix}^\top\right)^{1/2}$ is still not necessarily a valid choice, because we can only show $D_t^{-1}\left(\sum_{i=0}^{t-1} \begin{bmatrix} x_i \\ u_i \end{bmatrix} \begin{bmatrix} x_i \\ u_i \end{bmatrix}^\top\right)^{1/2}$ is asymptotically an orthogonal matrix, but not identity matrix.

F.4.1 Finding a valid E_t

Recall Eq. 36 that

$$\sum_{i=0}^{t-1} u_i u_i^\top / t^\beta \log^\alpha(t) = K M_t K^\top + \Delta_t K^\top + K \Delta_t^\top + \frac{\tau^2}{\beta} I_d + o_p(1).$$

Now denote

$$\Delta_u := \sum_{i=0}^{t-1} u_i u_i^\top / t^\beta \log^\alpha(t) - \Big(K M_t K^\top + \Delta_t K^\top + K \Delta_t^\top \Big) = \frac{\tau^2}{\beta} I_d + o_p(1), \tag{60}$$

which is asymptotically proportional to the identity matrix, and is also symmetric. Recall that D_t is defined as

$$D_t := t^{\beta/2} \log^{\alpha/2}(t) \begin{bmatrix} I_n & 0 \\ K & I_d \end{bmatrix} \begin{bmatrix} C_t^{1/2} & 0 \\ 0 & \sqrt{\frac{\tau^2}{\beta}} I_d \end{bmatrix}.$$

We will verify that the following construction of E_t is a valid choice:

$$E_t := t^{\beta/2} \log^{\alpha/2}(t) \begin{bmatrix} I_n & 0 \\ K & I_d \end{bmatrix} \begin{bmatrix} (M_t - \Delta_t^\top \Delta_u^{-1} \Delta_t)^{1/2} & \Delta_t^\top \Delta_u^{-1/2} \\ 0 & \Delta_u^{1/2} \end{bmatrix}.$$

We shall examine the two conditions $D_t^{-1} E_t \xrightarrow{P} I_{n+d}$ and $E_t E_t^\top = \sum_{i=0}^{t-1} \begin{bmatrix} x_i \\ u_i \end{bmatrix} \begin{bmatrix} x_i \\ u_i \end{bmatrix}^\top$ in order.

Proving $D_t^{-1} E_t \xrightarrow{P} I_{n+d}$ It suffices to show:

$$\begin{bmatrix} C_t^{-1/2} & 0 \\ 0 & \sqrt{\frac{\beta}{\tau^2}} I_d \end{bmatrix} \begin{bmatrix} (M_t - \Delta_t^\top \Delta_u^{-1} \Delta_t)^{1/2} & \Delta_t^\top \Delta_u^{-1/2} \\ 0 & \Delta_u^{1/2} \end{bmatrix} \xrightarrow{P} I_{n+d}.$$

Eqs. 33, 34, and Eqs. 35 and 60 states that

- $C_t^{-1} = \mathcal{O}(t^{\beta-1} \log^\alpha(t))$
- $M_t = C_t(1 + o_p(1))$
- $\Delta_t = \mathcal{O}_p(t^{1-3\beta/2} \log^{\frac{-3\alpha+3}{2}}(t))$
- $\Delta_u = \frac{\tau^2}{\beta} I_d + o_p(1)$

With these facts, $C_t^{-1/2} \Delta_t^\top \Delta_u^{-1/2} = \mathcal{O}_p(t^{1/2-\beta} \log^{\frac{-2\alpha+3}{2}}(t)) \xrightarrow{P} 0$ and $\sqrt{\frac{\tau^2}{\beta}} I_d \Delta_u^{1/2} \xrightarrow{P} I_d$ are immediate. It only remains to show that $C_t^{-1/2}(M_t - \Delta_t^\top \Delta_u^{-1} \Delta_t)^{1/2} \xrightarrow{P} I_n$. Notice

$$C_t^{-1/2}(M_t - \Delta_t^\top \Delta_u^{-1} \Delta_t)^{1/2} = (C_t^{-1} M_t - C_t^{-1} \Delta_t^\top \Delta_u^{-1} \Delta_t)^{1/2},$$

and Eq. 34 shows that $C_t^{-1}M_t \xrightarrow{P} I_n$. It only remains to show

$$C_t^{-1}\Delta_t^\top \Delta_u^{-1}\Delta_t \xrightarrow{P} 0,$$

which is true because when $\beta > 1/2$ or $\beta = 1/2$ and $\alpha > 3/2$:

$$\begin{aligned}
&C_t^{-1}\Delta_t^\top \Delta_u^{-1}\Delta_t \\
=&\mathcal{O}(t^{\beta-1}\log^{\alpha}(t))\mathcal{O}_p(t^{1-3\beta/2}\log^{\frac{-3\alpha+3}{2}}(t))\mathcal{O}_p(1)\mathcal{O}_p(t^{1-3\beta/2}\log^{\frac{-3\alpha+3}{2}}(t)) \\
=&\mathcal{O}_p(t^{-2\beta+1}\log^{-2\alpha+3}(t)) \\
=&o_p(1).
\end{aligned}$$

Proving $E_tE_t^\top = \sum_{i=0}^{t-1}\begin{bmatrix} x_i \\ u_i \end{bmatrix}\begin{bmatrix} x_i \\ u_i \end{bmatrix}^\top$.

$$\begin{aligned}
E_tE_t^\top =&t^\beta \log^\alpha(t)\left[\begin{array}{cc} I_n & 0 \\ K & I_d \end{array}\right]\left[\begin{array}{cc} (M_t - \Delta_t^\top \Delta_u^{-1}\Delta_t)^{1/2} & \Delta_t^\top \Delta_u^{-1/2} \\ 0 & \Delta_u^{1/2} \end{array}\right] \\
&\cdot \left[\begin{array}{cc} (M_t - \Delta_t^\top \Delta_u^{-1}\Delta_t)^{1/2} & 0 \\ \Delta_u^{-1/2}\Delta_t & \Delta_u^{1/2} \end{array}\right]\left[\begin{array}{cc} I_n & K^\top \\ 0 & I_d \end{array}\right] \\
=&t^\beta \log^\alpha(t)\left[\begin{array}{cc} I_n & 0 \\ K & I_d \end{array}\right]\left[\begin{array}{cc} M_t & \Delta_t^\top \\ \Delta_t & \Delta_u \end{array}\right]\left[\begin{array}{cc} I_n & K^\top \\ 0 & I_d \end{array}\right] \\
=&t^\beta \log^\alpha(t)\left[\begin{array}{cc} M_t & M_tK^\top + \Delta_t^\top \\ KM_t + \Delta_t & KM_tK^\top + \Delta_tK^\top + K\Delta_t^\top + \Delta_u \end{array}\right] \\
=&\sum_{i=0}^{t-1}\begin{bmatrix} x_i \\ u_i \end{bmatrix}\begin{bmatrix} x_i \\ u_i \end{bmatrix}^\top.
\end{aligned}$$

Last step is by definitions Eq. 21, 22, and 60. We will re-use the following equation later:

$$\sum_{i=0}^{t-1}\begin{bmatrix} x_i \\ u_i \end{bmatrix}\begin{bmatrix} x_i \\ u_i \end{bmatrix}^\top = t^\beta \log^\alpha(t)\left[\begin{array}{cc} I_n & 0 \\ K & I_d \end{array}\right]\left[\begin{array}{cc} M_t & \Delta_t^\top \\ \Delta_t & \Delta_u \end{array}\right]\left[\begin{array}{cc} I_n & K^\top \\ 0 & I_d \end{array}\right]. \tag{61}$$

■

F.5 The proof of Corollary 13

Corollary. *Algorithm 1 applied to a system described by Eq. 1 under Assumption 1 satisfies:*

$$\left(\sigma^2\begin{bmatrix} x_t \\ u_t \end{bmatrix}^\top\left(\sum_{i=0}^{t-1}\begin{bmatrix} x_i \\ u_i \end{bmatrix}\begin{bmatrix} x_i \\ u_i \end{bmatrix}^\top\right)^{-1}\begin{bmatrix} x_t \\ u_t \end{bmatrix}\right)^{-1/2}\left((\hat{A}_t - A)x_t + (\hat{B}_t - B)u_t\right) \xrightarrow{D} \mathcal{N}(0, I_n).$$

where $u_t = \hat{K}_tx_t + \xi_t$ *for any* ξ_t *independent of the data before* t*:* $\{\varepsilon_i, \eta_i\}_{i=0}^{t-1}$.

Proof

This one final lemma connects Lemma 30 to our desired conclusion by changing the parametric expression to the observable one:

Lemma 31. *For any ξ_t independent of the data before t: $\{\varepsilon_i, \eta_i\}_{i=0}^{t-1}$,*

$$\left(x_t^\top \left(\sum_{p=0}^{\infty} L^p \left(I_n + 1_{\{\beta=1,\alpha=0\}} \frac{\tau^2}{\sigma^2} BB^\top \right) (L^p)^\top \right)^{-1} x_t + \frac{\beta\sigma^2}{\tau^2} t^{1-\beta} \log^{-\alpha}(t) \|\xi_t\|^2 \right)^{-1/2}$$

$$\cdot\, t^{1/2} \left(\sigma^2 \begin{bmatrix} x_t \\ u_t \end{bmatrix}^\top \left(\sum_{i=0}^{t-1} \begin{bmatrix} x_i \\ u_i \end{bmatrix} \begin{bmatrix} x_i \\ u_i \end{bmatrix}^\top \right)^{-1} \begin{bmatrix} x_t \\ u_t \end{bmatrix} \right)^{1/2} \xrightarrow{P} 1.$$

The proof of Lemma 31 can be found in Appendix H.3.5. Finally, we can say

$$\left(\sigma^2 \begin{bmatrix} x_t \\ u_t \end{bmatrix}^\top \left(\sum_{i=0}^{t-1} \begin{bmatrix} x_i \\ u_i \end{bmatrix} \begin{bmatrix} x_i \\ u_i \end{bmatrix}^\top \right)^{-1} \begin{bmatrix} x_t \\ u_t \end{bmatrix} \right)^{-1/2} \left((\hat{A}_t - A)x_t + (\hat{B}_t - B)u_t \right) \xrightarrow{D} \mathcal{N}(0, I_n).$$

■

Appendix G. The proof of Propositions

G.1 The proof of Proposition 15

Proposition (Similar to Proposition C.1 in Dean et al. (2018))**.** *Let $x_0 \in \mathbb{R}^n$ be any initial state. Assume Assumption 1 is satisfied. When applying Algorithm 1,*

$$\max\left\{ \|\hat{A}_t - A\|, \|\hat{B}_t - B\| \right\} = \mathcal{O}(t^{-\frac{\beta}{2}} \log^{\frac{-\alpha+1}{2}}(t)) \ a.s.$$

G.1.1 Proof Outline

Proof We shall see that all the properties we derived in this section only require the safety condition Algorithm 1 Line 4 without any other requirement on the controller $\hat{K}_t$, and thus also apply to Algorithm 1 with logarithmic updates; see Remark 2.

According to Algorithm 1 Line 4, we keep our controller $\hat{K}_t$ bounded $\|\hat{K}_t\| \le C_K$, which means the next state can not be too far from the previous state. At the same time, whenever the state is too large ($\|x_t\| > C_x \log(t)$), it is tuned down by safe controller K_0. Overall speaking, the state x_t is always controlled with at most $\log(t)$ growth. We will see in Lemma 33 that when state growth is controlled, we have a decent bound on $\hat{A}_t, \hat{B}_t$.

In other words, as long as we still run Algorithm 1 Line 4 at every time step, which is enough to "control" the system by itself, any $\hat{A}_t, \hat{B}_t$ generated with Line 3 satisfies

$$\max\left\{ \|\hat{A}_t - A\|, \|\hat{B}_t - B\| \right\} = \mathcal{O}(t^{-\frac{\beta}{2}} \log^{\frac{-\alpha+1}{2}}(t)) \text{ a.s.}$$

regardless of the estimation result before time t.

Lemma 33 follows from a result by Simchowitz et al. (2018) on the estimation of linear response time-series. We present that result in the context of our problem. Let $\Theta := [A, B]$, and define $z_t := \begin{bmatrix} x_t \\ u_t \end{bmatrix}$. Then, the OLS estimator Eq. 5 is

$$(\hat{A}_T, \hat{B}_T) = \hat{\Theta}_T \in \arg\min_{\Theta} \sum_{t=0}^{T-1} \frac{1}{2} \left\| x_{t+1} - \Theta z_t \right\|_2^2 . \tag{62}$$

We know that the accuracy of the OLS estimator is related to the covariance structure of the predictors, which are $\{z_t\}_{t=0}^{T}$ in our context. To capture such covariance structure, we need the following definiton:

Definition 32 (BMSB condition)**.** *The $\{\mathcal{F}_t\}_{t\geq 0}$-adapted process $\{z_t\}_{t=0}^{T}$ is said to satisfy the (k, ν, ξ)-*block martingale small-ball *(BMSB) condition if for any $0 \leq j \leq T-k$ and $v \in \mathcal{S}^{n+d-1} := \{x \in \mathbb{R}^{n+d} : \|x\| = 1\}$, one has that*

$$\frac{1}{k} \sum_{i=1}^{k} \mathbb{P}\left(|\langle v, z_{j+i} \rangle| \geq \nu | \mathcal{F}_j\right) \geq \xi \ \ a.s.$$

This condition is used for characterizing the size of the minimum eigenvalue of the matrix $\sum_{t=0}^{T-1} z_t z_t^\top$. A larger ν guarantees a larger lower bound of the minimum eigenvalue. In the context of our problem the result by Simchowitz et al. (2018) translates as follows.

Lemma 33 (A slightly different version of Theorem C.2 in Dean et al. (2018))**.** *For $\delta \in (0, \frac{(n+d)\xi^2}{2}]$, for every T, k, ν, and ξ such that $\{z_t\}_{t=0}^{T}$ satisfies the (k, ν, ξ)-BMSB and*

$$T/k \geq \frac{10(n+d)}{\xi^2} \log\left(\frac{100(n+d) \sum_{t=0}^{T-1} \mathbf{Tr}(\mathbb{E} z_t z_t^\top)}{T \nu^2 \xi^2 \delta^{1+\frac{1}{n+d}}} \right). \tag{63}$$

the estimate $\hat{\Theta}_T$ defined in Eq. 62 satisfies the following statistical rate

$$\mathbb{P}\left(\left\| \hat{\Theta}_T - \Theta \right\| > \frac{90\sigma}{\xi\nu} \sqrt{\frac{n+d}{T}\left(1 + \log\left(\frac{10(n+d)\sum_{t=0}^{T-1} \mathbf{Tr}(\mathbb{E} z_t z_t^\top)}{T \delta^{1+\frac{1}{n+d}} \nu^2 \xi} \right)\right)} \right) \leq 3\delta. \tag{64}$$

The proof of Lemma 33 can be found in Appendix H.4.1.

We will show that $\sum_{t=1}^{T} \mathbf{Tr}(\mathbb{E} z_t z_t^\top)$ grows linearly with T (ignoring logarithmic terms), which means in Eq. 63 the LHS grows faster than the RHS, and is thus always satisfied if T is large enough. Lemma 33 is saying that for any T larger than some constant, we can control the L_2 norm of the system parameter estimate $\hat{\Theta}_T$, which implies we can control the L_2 norm of both $\hat{A}_T$ and $\hat{B}_T$.

Still there is one more gap from our Proposition 15, which requires *uniform* control on $\hat{A}_T$ and $\hat{B}_T$. Fortunately, we have the blessing that this high-probability bound is in the log scale w.r.t δ. Because of that, we can choose a series of decaying $\delta_T = 1/T^2$ for each different estimate $\hat{\Theta}_T$, so that $\sum_{T=C}^{\infty} 1/T^2 \leq 1/C$ and we can achieve a uniform high probability bound on $\hat{A}_T$ and $\hat{B}_T$ for all $T > C$, which directly leads to the desired conclusion once we plug in appropriate values for k, ν, and ξ:

$$\max\left\{\|\hat{A}_t - A\|, \|\hat{B}_t - B\|\right\} = \mathcal{O}(t^{-\frac{\beta}{2}} \log^{\frac{-\alpha+1}{2}}(t)) \text{ a.s.}$$

To sum up, there are three main steps in our proof of Proposition 15:

- Verify $\{z_t\}_{t=0}^T$ satisfies the (k, ν, ξ)-BMSB condition in our setting.
- Replace $\mathbf{Tr}(\mathbb{E}z_t z_t^\top)$ in Lemma 33 by an explicit upper bound in terms of T.
- Prove a uniform high probability bound for $\hat{A}_T$ and $\hat{B}_T$ by choosing with $\delta_T = 1/T^2$ with Lemma 33.

■

G.1.2 Verifying $\{z_t\}_{t=0}^T$ satisfies the (k, ν, ξ)-BMSB condition

In order to apply Lemma 33, we need to find k, ν, and ξ such that $\{z_t\}_{t=0}^T$ satisfies the (k, ν, ξ)-BMSB condition.

Lemma 34 (Similar to Lemma C.3 in Dean et al. (2018))**.** *If we assume Assumption 1, then apply Algorithm 1, the process $\{z_t\}_{t\geq 0}^T$ satisfies the (k, ν, ξ)-BMSB condition for*

$$(k, \nu, \xi) = \left(1, \sqrt{\sigma_{\eta,T}^2 \min\left(\frac{1}{2}, \frac{\sigma^2}{2\sigma^2 C_K^2 + \tau^2}\right)}, \frac{3}{10}\right),$$

where $\sigma_{\eta,T}^2 = \tau^2 T^{\beta-1} \log^\alpha(T)$.

See Appendix H.4.2 for the proof of Lemma 34.

G.1.3 Upper bound of $\mathbf{Tr}(\mathbb{E}z_t z_t^\top)$ in terms of T

The benefit of a non-random upper bound of $\mathbf{Tr}(\mathbb{E}z_t z_t^\top)$ w.r.t T is two-fold.

- We can know exactly how large our T should be for Eq. 63 to hold.
- Furthermore, we can also substitute the upper bound in to Eq. 64.

Lemma 35 shows that we have an upper bound of $\mathbf{Tr}(\mathbb{E}z_t z_t^\top)$ that is $\tilde{\mathcal{O}}(T)$.

Lemma 35 (Similar to Lemma C.4 in Dean et al. (2018))**.** *If we assume Assumption 1, then apply Algorithm 1, the process $\{z_t\}_{t\geq 0}^T$ satisfies*

$$\sum_{t=0}^{T-1} \mathbf{Tr}\left(\mathbb{E}z_t z_t^\top\right) = \mathcal{O}(T \log^2(T)). \tag{65}$$

See Appendix H.4.3 for the proof of Lemma 35.

G.1.4 Uniform upper bound for $\max\left\{\|\hat{A}_t - A\|, \|\hat{B}_t - B\|\right\}$

With Lemma 34 and Lemma 35 in hand, we can translate Lemma 33 into our problem setting. Fixing $\delta \in (0, \frac{(n+d)\xi^2}{2}]$, we already proved by Lemma 34 that the process $z_t = \begin{bmatrix} x_t \\ u_t \end{bmatrix}$ satisfies the

$$(k, \nu, \xi) = \left(1, \sqrt{\sigma_{\eta,T}^2 \min\left(\frac{1}{2}, \frac{\sigma^2}{2\sigma^2 C_K^2 + \tau^2}\right)}, \frac{3}{10}\right) \text{ BMSB condition.} \tag{66}$$

If we choose $\delta = \frac{1}{3T^2}$ and T such that Eq. 63 holds with (k, ν, ξ) in Eq. 66, we can apply Lemma 33. By Eq. 65, we only need T to satisfy

$$\begin{aligned}
T/k \geq & \frac{10(n+d)}{\xi^2} \log\left(\frac{100(n+d)\tilde{\mathcal{O}}(T)}{T\nu^2\xi^2\delta^{1+\frac{1}{n+d}}}\right) \\
= & \mathcal{O}(1) \log\left(\frac{\tilde{\mathcal{O}}(T)}{T\sigma_{\eta,T}^2 \frac{\sigma^2}{2\sigma^2 C_K^2 + \tau^2} T^{-2(1+\frac{1}{n+d})}}\right) \qquad (\xi = \frac{3}{10} \text{ is fixed constant}) \\
= & \tilde{\mathcal{O}}(1) \qquad (\text{Recall that } \sigma_{\eta,T}^2 = T^{\beta-1}\log^{\alpha}(T)).
\end{aligned}$$

Since T is growing faster than $\tilde{\mathcal{O}}(1)$, the above condition is essentially saying that our T should be larger than some constant $\mathcal{O}(1)$. Suppose that is the case, then following Lemma 33 and Lemma 35, the estimate $\hat{\Theta}_T$ defined in Eq. 62 satisfies the following statistical rate

$$\begin{aligned}
& \mathbb{P}\left(\left\|\hat{\Theta}_T - \Theta\right\| > \frac{90\sigma}{\xi\nu}\sqrt{\frac{n+d}{T}\left(1 + \log\left(\frac{10(n+d)\tilde{\mathcal{O}}(T)}{T\delta^{1+\frac{1}{n+d}}\nu^2\xi}\right)\right)}\right) \\
\leq & \mathbb{P}\left(\left\|\hat{\Theta}_T - \Theta\right\| > \frac{90\sigma}{\xi\nu}\sqrt{\frac{n+d}{T}\left(1 + \log\left(\frac{10(n+d)\sum_{t=1}^T \mathbf{Tr}(\mathbb{E}z_t z_t^\top)}{T\delta^{1+\frac{1}{n+d}}\nu^2\xi}\right)\right)}\right) \\
\leq & 3\delta.
\end{aligned}$$

Notice that $\hat{\Theta}_T = [\hat{A}_T, \hat{B}_T]$, and we know that $\max\left\{\begin{smallmatrix}\|\hat{A}_T - A\|, \\ \|\hat{B}_T - B\|\end{smallmatrix}\right\} \leq \left\|\hat{\Theta}_T - \Theta\right\|_2$. That is to say

$$\mathbb{P}\left(\max\left\{\begin{smallmatrix}\|\hat{A}_T - A\|, \\ \|\hat{B}_T - B\|\end{smallmatrix}\right\} > \frac{90\sigma}{\xi\nu}\sqrt{\frac{n+d}{T}\left(1 + \log\left(\frac{10(n+d)\tilde{\mathcal{O}}(1)}{T\delta^{1+\frac{1}{n+d}}\nu^2\xi}\right)\right)}\right) \leq 3\delta.$$

Next we substitute $k = 1$, $\xi = \frac{3}{10}$, $\nu = \sqrt{\sigma_{\eta,T}^2 \min\left(\frac{1}{2}, \frac{\sigma^2}{2\sigma^2 C_K^2 + \tau^2}\right)}$, and $\delta = \frac{1}{3T^2}$ into the previous equation

$$\mathbb{P}\left(\max\left\{\substack{\|\hat{A}_T-A\|,\\ \|\hat{B}_T-B\|}\right\} > \frac{\mathcal{O}(1)}{\sqrt{\sigma^2_{\eta,T}}}\sqrt{\frac{n+d}{T}\left(1+\log\left(\frac{\tilde{\mathcal{O}}(T)}{T(3T^{-2})^{1+\frac{1}{n+d}}\sigma^2_{\eta,T}}\right)\right)}\right) \leq \frac{1}{T^2}.$$

By merging all constant parameters in to the $\mathcal{O}$ style expression, and noticing that $\sigma^2_{\eta,T} = \tau^2 T^{\beta-1}\log^{\alpha}(T)$, where $\beta \in [1/2, 1)$, we have for any $T > \mathcal{O}(1)$:

$$\mathbb{P}\left(\max\left\{\substack{\|\hat{A}_T-A\|,\\ \|\hat{B}_T-B\|}\right\} > \mathcal{O}(T^{\frac{1-\beta}{2}}\log^{-\alpha/2}(T))\sqrt{\frac{n+d}{T}\mathcal{O}(\log(T))}\right) \leq \frac{1}{T^2},$$

which implies

$$\mathbb{P}\left(\max\left\{\substack{\|\hat{A}_T-A\|,\\ \|\hat{B}_T-B\|}\right\} > \mathcal{O}(T^{-\frac{\beta}{2}}\log^{\frac{-\alpha+1}{2}}(T))\right) \leq \frac{1}{T^2}.$$

Notice that

$$\sum_{T=C+1}^{\infty}\frac{1}{T^2} \leq \sum_{T=C+1}^{\infty}\frac{1}{T(T-1)} \leq \sum_{T=C+1}^{\infty}\frac{1}{T-1}-\frac{1}{T} = \frac{1}{C}.$$

Therefore we can derive a uniform confidence bound on the estimation error of parameters $\hat{A}_t$ and $\hat{B}_t$: For any integer $C > \mathcal{O}(1)$:

$$\begin{aligned}&\mathbb{P}\left(\exists t > C,\ s.t.\ \max\left\{\substack{\|\hat{A}_t-A\|,\\ \|\hat{B}_t-B\|}\right\} > \mathcal{O}(T^{-\frac{\beta}{2}}\log^{\frac{-\alpha+1}{2}}(T))\right)\\ \leq &\sum_{t=C+1}^{\infty}\mathbb{P}\left(\max\left\{\substack{\|\hat{A}_t-A\|,\\ \|\hat{B}_t-B\|}\right\} > \mathcal{O}(T^{-\frac{\beta}{2}}\log^{\frac{-\alpha+1}{2}}(T))\right)\\ \leq &\frac{1}{C}.\end{aligned}$$

Notice that this is a uniform upper bound for all $t > C$. Recall Definition 14 Item 8, where we define $X_n = \mathcal{O}(a_n)$ a.s. as: for almost every $\omega \in \Omega$, there exists a number $C(\omega)$ such that $|X_n(\omega)| \leq C(\omega)a_n$, where Ω denotes the sample space of $\{X_n\}_n$. The previous equation is telling us the union of such event ω happens with at least probability $1 - 1/C$, and by taking $C \to \infty$ that is exactly the definition of $\mathcal{O}(a_n)$ a.s., and thus:

$$\max\left\{\|\hat{A}_t - A\|, \|\hat{B}_t - B\|\right\} = \mathcal{O}(t^{-\frac{\beta}{2}}\log^{\frac{-\alpha+1}{2}}(t)) \text{ a.s.}$$

The same bound holds for logarithmic updates. The reason is that for time t, the closest estimation update will always be within t/c time steps of t, which does not change the order:

$$\mathcal{O}((t/c)^{-\frac{\beta}{2}}\log^{\frac{-\alpha+1}{2}}(t/c)) = \mathcal{O}(t^{-\frac{\beta}{2}}\log^{\frac{-\alpha+1}{2}}(t)).$$

G.2 The proof of Proposition 16

Proposition. *Let $x_0 \in \mathbb{R}^n$ be any initial state. Assume Assumption 1 is satisfied. When applying Algorithm 1,*

$$\max\left\{\|\hat{A}_t - A\|, \|\hat{B}_t - B\|, \|\tilde{K}_{t+1} - K\|\right\} = \mathcal{O}(t^{-\frac{\beta}{2}}\log^{\frac{-\alpha+1}{2}}(t)) \ a.s.$$

G.2.1 Proof Outline

Proof When the problem parameters (A, B, Q, R) are known the optimal policy is given by linear feedback, $u_t = Kx_t$, where $K = -(R + B^\top PB)^{-1}B^\top PA$ and P is the (positive definite) solution to the discrete Riccati equation

$$P = A^\top PA - A^\top PB(R + B^\top PB)^{-1}B^\top PA + Q. \tag{67}$$

In the following context any time we mention $\hat{P}_t$ and $\tilde{K}_{t+1}$, we are refering to the corresponding certainty equivalent responses.

$$\hat{P}_t = \hat{A}_t^\top \hat{P}_t\hat{A}_t - \hat{A}_t^\top \hat{P}_t\hat{B}_t(R + \hat{B}_t^\top \hat{P}_t\hat{B}_t)^{-1}\hat{B}_t^\top \hat{P}_t\hat{A}_t + Q.$$

$$\tilde{K}_{t+1} = -(R + \hat{B}_t^\top \hat{P}_t\hat{B}_t)^{-1}\hat{B}_t^\top \hat{P}_t\hat{A}_t.$$

Since we already controlled the estimation error of $\hat{A}_t - A$ and $\hat{B}_t - B$, one natural thing to ask is that, if we have control over $\hat{A}_t - A$ and $\hat{B}_t - B$, do we have control over $\tilde{K}_{t+1} - K$? This can be achieved by two steps:

1. Show that we can control $\tilde{K}_{t+1}$ once $\hat{A}_t$, $\hat{B}_t$, and $\hat{P}_t$ are controlled.
2. Show that we can control $\hat{P}_t$ once $\hat{A}_t$ and $\hat{B}_t$ are controlled.

■

G.2.2 Show that we can control $\tilde{K}_{t+1}$ once $\hat{A}_t$, $\hat{B}_t$, and $\hat{P}_t$ are controlled

This is already stated by Proposition 1 in Mania et al. (2019). Denote the quantity

$$\Gamma_1 := 1 + \max\{\|A\|, \|B\|, \|P\|, \|K\|\}.$$

Proposition 36 (Proposition 1 in Mania et al. (2019)). *Let $\epsilon > 0$ such that $\|\hat{A} - A\| \le \epsilon$ and $\|\hat{B} - B\| \le \epsilon$. Also, let $\|\hat{P} - P\| \le \epsilon_P$ such that $\epsilon_P \ge \epsilon$. Assume $\underline{\sigma}(R) \ge 1$ we have*

$$\|\hat{K} - K\| \le 7\Gamma_1^3\epsilon_P.$$

The $\underline{\sigma}(R)$ represents the minimum eigenvalue of R. we can discard the constraint of $\underline{\sigma}(R) \ge 1$ by the following observation. If we replace our Q and R by $Q/\underline{\sigma}(R)$ and $R/\underline{\sigma}(R)$, then the corresponding solution P for Eq. 67 will be $P/\underline{\sigma}(R)$. Notice that changing Q and R by the same proportion does not change the LQR problem. With that being said, our LS estimator $\hat{A}_t$, $\hat{B}_t$, and the nominal controller $\tilde{K}_t$ will remain the same. By this transformation the minimum eigenvalue condition is satisfied, and we only need to control

$$\|\hat{P} - P\|/\underline{\sigma}(R) \le \epsilon_P$$

such that $\epsilon_P \ge \epsilon$, and we will have $\|\hat{K} - K\| \le 7\Gamma_2^3\epsilon_P$, where

$$\Gamma_2 := 1 + \max\{\|A\|, \|B\|, \|P\|/\underline{\sigma}(R), \|K\|\}.$$

Here we can replace this denominator $\underline{\sigma}(R)$ by any constant smaller than $\underline{\sigma}(R)$, and the whole story would still work. Since later we will also require $\underline{\sigma}(P) \ge 1$, we can choose the shared denominator to be $\min\{\underline{\sigma}(R), \underline{\sigma}(P)\}$. To sum up we have the following corollary of Proposition 36.

Corollary 37. *Let $\epsilon > 0$ such that $\|\hat{A} - A\| \le \epsilon$ and $\|\hat{B} - B\| \le \epsilon$. Also, let $\|\hat{P} - P\| \le \min\{\underline{\sigma}(R), \underline{\sigma}(P)\}\epsilon_P$ such that $\epsilon_P \ge \epsilon$. Then we have*

$$\|\hat{K} - K\| \le 7\Gamma_3^3\,\epsilon_P.$$

where $\Gamma_3 := 1 + \max\{\|A\|, \|B\|, \|P\|/\min\{\underline{\sigma}(R), \underline{\sigma}(P)\}, \|K\|\}$.

Now we only need to prove that $\|\hat{P} - P\| = \mathcal{O}(\epsilon)$ given $\|\hat{A} - A\| \le \epsilon$ and $\|\hat{B} - B\| \le \epsilon$.

G.2.3 Show that we can control $\hat{P}_t$ once $\hat{A}_t$ and $\hat{B}_t$ are controlled

Consider a general square matrix M. In order to quantify the decay rate of $\|M^k\|$, we define

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

In other words, $\tau(M, \rho)$ is the smallest value such that $\|M^k\| \le \tau(M, \rho)\rho^k$ for all $k \ge 0$. We note that $\tau(M, \rho)$ might be infinite, depending on the value of ρ, and it is always greater than or equal to one. If ρ is larger than $\rho(M)$, we are guaranteed to have a finite $\tau(M, \rho)$ (this is a consequence of Gelfand's formula). In particular, if M is a stable matrix, we can choose $\rho < 1$ such that $\tau(M, \rho)$ is finite. Also, we note that $\tau(M, \rho)$ is a decreasing function of ρ; if $\rho \ge \|M\|$, we have $\tau(M, \rho) = 1$.

Recall that $L := A + BK$. The following proposition that upper bounds $\|\hat{P} - P\|$ holds in a more general LQG setting where the matrix Q is unknown:

Proposition 38 (Proposition 2 in Mania et al. (2019))**.** *Let $\gamma \ge \rho(L)$ and also let ϵ be such that $\|\hat{A} - A\|$, $\|\hat{B} - B\|$, and $\|\hat{Q} - Q\|$ are at most ϵ. Let $\|\cdot\|_+ = \|\cdot\| + 1$. We assume that $R \succ 0$, (A, B) is stabilizable, $(Q^{1/2}, A)$ observable, and $\underline{\sigma}(P) \ge 1$.*

$$\|\hat{P} - P\| \le \mathcal{O}(1)\,\epsilon\,\frac{\tau(L,\gamma)^2}{1-\gamma^2}\|A\|_+^2\|P\|_+^2\|B\|_+\|R^{-1}\|_+,$$

as long as

$$\epsilon \le \mathcal{O}(1)\frac{(1-\gamma^2)^2}{\tau(L,\gamma)^4}\|A\|_+^{-2}\|P\|_+^{-2}\|B\|_+^{-3}\|R^{-1}\|_+^{-2}\min\left\{\|L\|_+^{-2}, \|P\|_+^{-1}\right\}.$$

Here $\mathcal{O}(1)$ are pure constants without dependence of any other parameters. We already assumed in Assumption 1 that (A, B) stabilizable, but we have not defined 'observable' yet. An equivalent statement of observable can be found here.

Lemma 39 (Lemma 2.1 in (Payne and Silverman, 1973))**.** *The pair (C, A) is observable if and only if $Ax = \lambda x$, $Cx = 0$ imply $x = 0$*

Since we already assumed Q is positive definite, $Qx = 0$ imply $x = 0$, and thus $(Q^{1/2}, A)$ is observable. In the LQAC setting we know Q exactly, so we can remove the estimation bound condition on Q.

Now we can restate Proposition 38 in the LQAC setting:

Corollary 40. *Let ϵ such that $\|\hat{A}_t - A\|$, and $\|\hat{B}_t - B\|$ are at most ϵ. Let $\|\cdot\|_+ = \|\cdot\| + 1$. We assume that $R \succ 0$, (A, B) is stabilizable, and $\underline{\sigma}(P) \geq 1$.*

$$\|\hat{P}_t - P\| \leq \mathcal{O}(1)\, \epsilon \frac{\tau(L, \rho(L))^2}{1 - \rho(L)^2} \|A\|_+^2 \|P\|_+^2 \|B\|_+ \|R^{-1}\|_+ = \mathcal{O}(\epsilon).$$

as long as

$$\epsilon \leq \mathcal{O}(1) \frac{(1 - \rho(L)^2)^2}{\tau(L, \rho(L))^4} \|A\|_+^{-2} \|P\|_+^{-2} \|B\|_+^{-3} \|R^{-1}\|_+^{-2} \min\left\{\|L\|_+^{-2}, \|P\|_+^{-1}\right\} = \mathcal{O}(1).$$

Here, the upper bound condition on ϵ is to ensure that $\hat{A}_t, \hat{B}_t$ is stabilizable, so that $\hat{P}_t$ is well defined. Furthermore, following the paragraph after Proposition 2 in Mania et al. (2019), the assumption $\underline{\sigma}(P) \geq 1$ can be made without loss of generality when the other assumptions are satisfied. The reason is that, when $R \succ 0$ and $(Q^{1/2}, A)$ observable, the value function matrix P is guaranteed to be positive definite. Similar to how we got Corollary 37, by replacing Q, R and P with $Q/\min\{\underline{\sigma}(R), \underline{\sigma}(P)\}$, $R/\min\{\underline{\sigma}(R), \underline{\sigma}(P)\}$ and $P/\min\{\underline{\sigma}(R), \underline{\sigma}(P)\}$, we can remove the constraint $\underline{\sigma}(P) \geq 1$.

Corollary 41. *Suppose $\|\hat{A}_t - A\| \leq \epsilon$ and $\|\hat{B}_t - B\| \leq \epsilon$. Let $\|\cdot\|_+ = \|\cdot\| + 1$. We assume that $R \succ 0$ and (A, B) is stabilizable.*

$$\begin{aligned}\|\hat{P}_t - P\| \leq& \min\{\underline{\sigma}(R), \underline{\sigma}(P)\} \mathcal{O}(1)\, \epsilon \frac{\tau(L, \rho(L))^2}{1 - \rho(L)^2} \|A\|_+^2 \\ & \left\|\frac{P}{\min\{\underline{\sigma}(R), \underline{\sigma}(P)\}}\right\|_+^2 \|B\|_+ \left\|\left(\frac{R}{\min\{\underline{\sigma}(R), \underline{\sigma}(P)\}}\right)^{-1}\right\|_+ \\ =& \mathcal{O}(\epsilon).\end{aligned}$$

as long as

$$\begin{aligned}\epsilon \leq& \mathcal{O}(1) \frac{(1 - \rho(L)^2)^2}{\tau(L, \rho(L))^4} \|A\|_+^{-2} \left\|\frac{P}{\min\{\underline{\sigma}(R), \underline{\sigma}(P)\}}\right\|_+^{-2} \\ & \|B\|_+^{-3} \left\|\left(\frac{R}{\min\{\underline{\sigma}(R), \underline{\sigma}(P)\}}\right)^{-1}\right\|_+^{-2} \min\left\{\|L\|_+^{-2}, \left\|\frac{P}{\min\{\underline{\sigma}(R), \underline{\sigma}(P)\}}\right\|_+^{-1}\right\} \\ =& \mathcal{O}(1).\end{aligned}$$

G.2.4 Combining the two results together

With Corollary 37 and Corollary 41 the following corollary is straightforward.

Corollary 42. *Let $\epsilon > 0$ such that $\epsilon \leq \mathcal{O}(1)$, $\|\hat{A}_t - A\| \leq \epsilon$ and $\|\hat{B}_t - B\| \leq \epsilon$. Then, we have*

$$\|\tilde{K}_{t+1} - K\| \leq 7\Gamma^3\, \epsilon_P = \mathcal{O}(\epsilon).$$

Here $\Gamma := 1 + \max\{\|A\|, \|B\|, \|P\|/\min\{\underline{\sigma}(R), \underline{\sigma}(P)\}, \|K\|\}$.

Proof With Corollary 41 we can find ϵ_P such that $\|\hat{P}_t - P\| = \mathcal{O}(\epsilon)$. Thus, the condition of Corollary 37 is satisfied. ■

G.2.5 Concluding the proof of Proposition 16

Proof With Proposition 15 and Corollary 42, it is straightforward to give a new corollary with uniform control on all $\|\hat{A}_t - A\|$, $\|\hat{B}_t - B\|$, and $\|\tilde{K}_{t+1} - K\|$. Recall that we already proved the high probability bound in Proposition 15 that

$$\max\left\{\|\hat{A}_t - A\|, \|\hat{B}_t - B\|\right\} = \mathcal{O}(t^{-\frac{\beta}{2}} \log^{\frac{-\alpha+1}{2}}(t)) \text{ a.s.}$$

Basically, to satisfy the constraint in Corollary 42, we only need our bound (named ϵ) in Proposition 15 to satisfy

$$\epsilon = \mathcal{O}(t^{-\frac{\beta}{2}} \log^{\frac{-\alpha+1}{2}}(t)) \leq \mathcal{O}(1) \text{ a.s.}$$

which is always true when t is large enough. (This also ensures $\hat{A}_t, \hat{B}_t$ to be stabilizable so that K_0 is only used finitely many times.) That means,

$$\|\tilde{K}_{t+1} - K\| = \mathcal{O}(\epsilon) = \mathcal{O}(t^{-\frac{\beta}{2}} \log^{\frac{-\alpha+1}{2}}(t)) \text{ a.s.}$$

Finally, we can say

$$\max\left\{\|\hat{A}_t - A\|, \|\hat{B}_t - B\|, \|\tilde{K}_{t+1} - K\|\right\} = \mathcal{O}(t^{-\frac{\beta}{2}} \log^{\frac{-\alpha+1}{2}}(t)) \text{ a.s.}$$

■

G.3 The proof of Proposition 17

Proposition. *Let $x_0 \in \mathbb{R}^n$ be any initial state. Assume Assumption 1 is satisfied. When applying Algorithm 1*

$$\max\left\{\|\hat{A}_t - A\|, \|\hat{B}_t - B\|, \|\hat{K}_{t+1} - K\|\right\} = \mathcal{O}(t^{-\frac{\beta}{2}} \log^{\frac{-\alpha+1}{2}}(t)) \textit{ a.s.}$$

Proof Going thorough the whole Algorithm 1, there are two conditions that might cause the difference between $\hat{K}_t$ and $\tilde{K}_t$:

1. $\|\tilde{K}_t\| > C_K$, and
2. $\|x_t\| > C_{x,t} = C_x \log(t)$.

Our objective is to show that, with probability 1, $\hat{K}_t \neq \tilde{K}_t$ will happen only finitely often.

The first case $\|\tilde{K}_t\| > C_K$ The first case is when $\|\tilde{K}_t\| > C_K$, this will not happen infinitely often. The first case $\|\tilde{K}_t\| > C_K$ can only happen when

$$\|\tilde{K}_t - K\| \geq \|\tilde{K}_t\| - \|K\| > C_K - \|K\|. \tag{68}$$

By Proposition 16, we know that $\|\tilde{K}_t - K\|$ is exponentially decaying:

$$\max\left\{\|\tilde{K}_t - K\|\right\} = \mathcal{O}(t^{-\frac{\beta}{2}} \log^{\frac{-\alpha+1}{2}}(t)) \text{ a.s.}$$

As a result, Eq. 68 will hold only finitely many times, a.s.

The second case $\|x_t\| > C_{x,t} = C_x \log(t)$ To examine how often this would happen, we need to dig into more details of the decomposition of $\|x_t\|$. Recall the previously derived formula from Lemma 19:

$$x_t = \sum_{p=0}^{t-1} (A + B\hat{K}_{t-1}) \cdots (A + B\hat{K}_{p+1})(B\eta_p + \varepsilon_p) + (A + B\hat{K}_{t-1}) \cdots (A + B\hat{K}_0)x_0.$$

We hope to get an upper bound for $\|x_t\|$. Apparently the main difficulty here is to bound the norm of $(A + B\hat{K}_{t-1}) \cdots (A + B\hat{K}_{p+1})$. The following lemma serves as a key.

Lemma 43. *Suppose we have a constant square matrix M with spectral radius $\rho(M) < 1$, and a sequence of uniformly bounded random variables $\{\delta_t\}_{t=0}^{\infty}$, satisfying $\|\delta_t\| \xrightarrow{a.s.} 0$. Denote the constant $\rho_M := \frac{2+\rho(M)}{3} < 1$. Then we have, for any $t, q \in \mathbb{N}$, $t > q$:*

$$\|(M + \delta_{t-1}) \cdots (M + \delta_q)\| = \mathcal{O}(\rho_M^{t-q}) \text{ a.s.}$$

And as a direct corollary

$$\|M^{t-q}\| = \mathcal{O}(\rho_M^{t-q}).$$

The proof can be found in Appendix H.5.1.

Notice that by our Algorithm 1, $\|\hat{K}_t\| \leq C_K$ always holds, thus there exists a uniform upper bound on $\|B\delta_t\| := \|B(\hat{K}_t - K)\| \leq \|B\|(C_K + \|K\|)$. Now we can separate the whole $\|(A + B\hat{K}_{t-1}) \cdots (A + B\hat{K}_{p+1})\|$ into two parts. If we denote $\rho_0 := \max(\frac{2+\rho(A+BK_0)}{3}, \frac{2+\rho(A+BK)}{3})$, then with Lemma 43, we can simultaneously bound both parts.

1. The first part contains the $A + B\hat{K}_k$ where $\hat{K}_k = K_0$, this part of product is denoted as I_1. In this part, $A + B\hat{K}_k = A + BK_0$. Suppose this part has p_1 same items, by Lemma 43 we know $I_1 \leq \mathcal{O}(\rho_0^{p_1})$ a.s.

2. The second part contains the $(A + B\hat{K}_k)$ where $\hat{K}_k = \tilde{K}_k$ to be our true certainty equivalent controller, this part of the product is denoted as I_2. If we denote $\delta_k := (\hat{K}_k - K)$, then, in this part, $(A+B\hat{K}_k) = (A+BK+B\delta_k)$. Remember our conclusion in Proposition 16 that $\|\tilde{K}_k - K\| \xrightarrow{a.s.} 0$, thus $\|\delta_k\| \xrightarrow{a.s.} 0$, assuming this part has p_2 items, then since $\|B\delta_k\| \leq \|B\|(C_K + \|K\|)$, by Lemma 43

$$I_2 \leq \mathcal{O}(\rho_0^{p_2}) \text{ a.s.}$$

We know $p_1 + p_2 = t - p - 1$. Combining these two parts we have

$$\|(A + B\hat{K}_{t-1}) \cdots (A + B\hat{K}_{p+1})\| \leq \mathcal{O}(\rho_0^{t-p}) \text{ a.s.}$$

Finally we have the bound on x_t:

$$\|x_t\| = \left\| \sum_{p=0}^{t-1} (A + B\hat{K}_{t-1}) \cdots (A + B\hat{K}_{p+1})(B\eta_p + \varepsilon_p) + (A + B\hat{K}_{t-1}) \cdots (A + BK_0)x_0 \right\|$$

$$
\begin{aligned}
&\leq \left(\sum_{p=0}^{t-1} \left\| (A + B\hat{K}_{t-1}) \cdots (A + B\hat{K}_{p+1}) \right\| \|B\eta_p + \varepsilon_p\| + \left\| (A + B\hat{K}_{t-1}) \cdots (A + BK_0) \right\| \|x_0\| \right) \\
&\leq \sum_{p=0}^{t-1} \mathcal{O}(\rho_0^{t-p}) \|B\eta_p + \varepsilon_p\| + \mathcal{O}(\rho_0^{t}) \|x_0\| \text{ a.s.}
\end{aligned}
$$

Then

$$
\|x_t\| = \mathcal{O}\left(\sum_{p=0}^{t-1} \rho_0^{t-p} \|B\eta_p + \varepsilon_p\| + \rho_0^t \|x_0\| \right) \text{ a.s.}
$$

By Gaussian tail bounds (see Lemma 18), we know that

$$
\|B\eta_t + \varepsilon_t\| = \mathcal{O}(\log^{1/2}(t)) \text{ a.s.}
$$

Then

$$
\|x_t\| = \mathcal{O}\left(\sum_{p=0}^{t-1} \rho_0^{t-p} \log^{1/2}(t) \right) + o(1) \text{ a.s.}
$$

Because ρ_0^{t-p} is geometric sequence,

$$
\|x_t\| \leq \mathcal{O}(\log^{1/2}(t)) \text{ a.s.}
$$

Thus for almost any $\omega \in \Omega$, $\|x_t\| > C_{x,t} = C_x \log(t)$ will happen only finitely many times.

Finally, because two conditions $\|\tilde{K}_t\| > C_K$ and $\|x_t\| > C_{x,t} = C_x \log(t)$ will happen only finitely many times, $\hat{K}_t$ and $\tilde{K}_t$ eventually are the same. Following Proposition 16,

$$
\max\left\{ \|\hat{A}_t - A\|, \|\hat{B}_t - B\|, \|\hat{K}_{t+1} - K\| \right\} = \mathcal{O}(t^{-\frac{\beta}{2}} \log^{\frac{-\alpha+1}{2}}(t)) \text{ a.s.}
$$

■

Appendix H. The proof of lemmas

H.1 Lemmas in Appendix B

H.1.1 The proof of Lemma 18

Lemma.

-
$$
\|\varepsilon_t\|, \|\eta_t\| = \mathcal{O}(\log^{1/2}(t)) \textit{ a.s.} \tag{69}
$$

-
$$
\|B\eta_t + \varepsilon_t\| = \mathcal{O}(\log^{1/2}(t)) \textit{ a.s.} \tag{70}
$$

Assume Eq. 23, then:

- $$\|\delta_t\| = \|\hat{K}_t - K\| = \mathcal{O}(t^{-\frac{\beta}{2}} \log^{\frac{-\alpha+1}{2}}(t)) \text{ a.s.} \tag{71}$$

- $$\|(L + B\delta_{t-1}) \cdots (L + B\delta_q)\| = \mathcal{O}(\rho_L^{t-q}) \text{ a.s.} \tag{72}$$

- $$\|x_t\|, \|u_t\| = \mathcal{O}(\log^{1/2}(t)) \text{ a.s.} \tag{73}$$

where $\delta_t := \hat{K}_t - K$, $L := A + BK$, and $\rho_L := \frac{2+\rho(L)}{3}$. ***Additionally, when*** $t = 0, 1$ ***all these terms are bounded by*** $\mathcal{O}(1)$ ***a.s.***

Proof Outline:

The proof of Eq. 69 and Eq. 70 The following lemma give the proof that Eq. 69 and Eq. 70 holds with probability at least $1 - \delta$, which can be shown by the tail bound for i.i.d Gaussian random variables.

Lemma 44. *For the noise $\eta_t \overset{i.i.d.}{\sim} \mathcal{N}(0, \tau^2 t^{1-\beta} \log^{\alpha}(t))$ and $\varepsilon_t \overset{i.i.d.}{\sim} \mathcal{N}(0, \sigma^2)$, we have that for any $\delta \in (0, 1)$, with probability $1 - \delta$, the following two equations holds for any $t \geq 1$:*

$$\|\varepsilon_t\|, \|\eta_t\|, \|B\eta_t + \varepsilon_t\| \leq \mathcal{O}(1) \log^{1/2}(t^2/\delta).$$

We will prove Lemma 44 shortly. By Definition 14 Item 8, this implies

$$\|\varepsilon_t\|, \|\eta_t\| \leq \mathcal{O}(\log^{1/2}(t)) \text{ a.s.}$$

and

$$\|B\eta_t + \varepsilon_t\| \leq \mathcal{O}(\log^{1/2}(t)) \text{ a.s.}$$

The proof of Eq. 71 and Eq. 72 Eq. 71 directly follows from Eq. 23. Eq. 72 follows from Lemma 43 given that we have $\delta_t \xrightarrow{a.s.} 0$ from Proposition 17:

$$\|(L + B\delta_{t-1}) \cdots (L + B\delta_q)\| \leq \mathcal{O}(\rho_L^{t-q}) \text{ a.s.}$$

The proof of Eq. 73 Finally we need to prove Eq. 73 that

$$\|x_t\|, \|u_t\| = \mathcal{O}(\log^{1/2}(t)) \text{ a.s.}$$

With the fact from Lemma 19 that

$$\begin{aligned} x_t =& \sum_{p=0}^{t-1} (A + B\hat{K}_{t-1}) \cdots (A + B\hat{K}_{p+1})(B\eta_p + \varepsilon_p) + (A + B\hat{K}_{t-1}) \cdots (A + BK_0)x_0 \\ =& \sum_{p=0}^{t-1} (L + B\delta_{t-1}) \cdots (L + B\delta_{p+1})(B\eta_p + \varepsilon_p) + (L + B\delta_{t-1}) \cdots (L + B\delta_0)x_0, \end{aligned}$$

combined with the conclusion of Eq. 72 and Eq. 70, we derive a norm bound on x_t:

$$\begin{aligned}
\|x_t\| \leq &\sum_{p=0}^{t-1} \|(L+B\delta_{t-1})\cdots(L+B\delta_{p+1})\|\|B\eta_p+\varepsilon_p\| + \|(L+B\delta_{t-1})\cdots(L+B\delta_0)\|\|x_0\| \text{ a.s.}\\
=&\sum_{p=0}^{t-1} \mathcal{O}(\rho_L^{t-p})\|B\eta_p+\varepsilon_p\| + \mathcal{O}(\rho_L^t)\|x_0\| \text{ a.s.}\\
=&\sum_{p=0}^{t-1} \mathcal{O}(\rho_L^{t-p})\mathcal{O}(\log^{1/2}(p)) + o(1) \text{ a.s.}\\
\leq&\sum_{p=0}^{t-1} \mathcal{O}(\rho_L^{t-p})\mathcal{O}(\log^{1/2}(t)) + o(1) \text{ a.s.}\\
=&\mathcal{O}(\log^{1/2}(t)) \text{ a.s.}
\end{aligned}$$

Recall that we have already shown Eq. 69:

$$\|\eta_t\| = \mathcal{O}(\log^{1/2}(t)) \text{ a.s.}$$

That means

$$\begin{aligned}
&\|u_i\|\\
=&\|(A+B\hat{K}_i)x_i+\eta_i\|\\
\leq&(\|A\|+\|B\|C_K)\|x_i\|+\|\eta_i\|\\
=&\mathcal{O}(\log^{1/2}(t)) \text{ a.s.}
\end{aligned}$$

■

The proof of Lemma 44 **Proof** For any Gaussian variable $X \sim \mathcal{N}(0, \sigma^2)$,

$$\mathbb{P}(X > t\sigma) \leq e^{-t^2/2},$$

and

$$\mathbb{P}(X^2 > t^2\sigma^2) = 2\mathbb{P}(X > t\sigma) \leq 2e^{-t^2/2}.$$

For any multivariate normal vector sequence $X_t \sim \mathcal{N}(0, \sigma^2 I_n)$,

$$\mathbb{P}(\|X_t\|^2 > nt\sigma^2) = \mathbb{P}\left(\sum_{i=1}^n X_{t,i}^2 > nt\sigma^2\right) \leq \sum_{i=1}^n \mathbb{P}(X_i^2 > t\sigma^2) \leq 2ne^{-t/2}.$$

That means for any constant $c > 0$,

$$\mathbb{P}(\|X_t\|^2 > n2\log(ct^2/\delta)\sigma^2) \leq 2ne^{-2\log(ct^2/\delta)/2} = \frac{2n\delta}{ct^2}.$$

We can sum up all choices of t to get a uniform bound. A well known equation states that $\sum_{t=1}^{\infty} 1/t^2 = \frac{\pi^2}{6}$. Then

$$\mathbb{P}(\exists t \geq 1 : \|X_t\|^2 > 2n\sigma^2 \log(ct^2/\delta)) \leq \sum_{t=1}^{\infty} \frac{2n\delta}{ct^2}.$$

We can choose $c = \frac{1}{2n} \sum_{t=1}^{\infty} 1/t^2 = \frac{\pi^2}{6 \cdot 2n}$, so that

$$\mathbb{P}(\exists t \geq 1 : \|X_t\|^2 > 2n\sigma^2 \log(ct^2/\delta)) \leq \delta.$$

That is to say, with probability at least $1 - \delta$, we have for any $t \geq 1$,

$$\|X_t\| \leq \mathcal{O}(1) \log^{1/2}(ct^2/\delta) = \mathcal{O}(1)(\log(t^2/\delta) + \log(c))^{1/2}.$$

Since $\log(c)$ can be dominated by $\log(t^2/\delta)$, the above equation can simply be written as

$$\|X_t\| \leq \mathcal{O}(1) \log^{1/2}(t^2/\delta).$$

This bound holds for ε_t which has constant variance and is also true for η_t which has shrinking variance. Thus, with probability at least $1 - \delta$:

$$\|\varepsilon_t\|, \|\eta_t\| \leq \mathcal{O}(1) \log^{1/2}(t^2/\delta).$$

Consider the fact that $\|B\eta_t + \varepsilon_t\| \leq \|B\|\|\eta_t\| + \|\varepsilon_t\|$, which means $\|B\eta_t + \varepsilon_t\|$ can still be bounded by:

$$\|B\eta_t + \varepsilon_t\| \leq \mathcal{O}(1) \log^{1/2}(t^2/\delta).$$

■

H.1.2 The proof of Lemma 19

Lemma.

$$x_t = \sum_{p=0}^{t-1} (A + B\hat{K}_{t-1}) \cdots (A + B\hat{K}_{p+1})(B\eta_p + \varepsilon_p) + (A + B\hat{K}_{t-1}) \cdots (A + BK_0)x_0.$$

$$u_t = \sum_{p=0}^{t-1} \hat{K}_t(A + B\hat{K}_{t-1}) \cdots (A + B\hat{K}_{p+1})(B\eta_p + \varepsilon_p) + \hat{K}_t(A + B\hat{K}_{t-1}) \cdots (A + BK_0)x_0 + \eta_t.$$

Here when $p = t - 1$, $(A + B\hat{K}_{t-1}) \cdots (A + B\hat{K}_{p+1}) := I_n$.

Proof Consider the following relationship:

$$u_t = \hat{K}_t x_t + \eta_t.$$

$$x_t = Ax_{t-1} + Bu_{t-1} + \varepsilon_{t-1}$$

$$
\begin{aligned}
&= Ax_{t-1} + B(\hat{K}_{t-1}x_{t-1} + \eta_{t-1}) + \varepsilon_{t-1} \\
&= (A + B\hat{K}_{t-1})x_{t-1} + B\eta_{t-1} + \varepsilon_{t-1}.
\end{aligned}
$$

Iteratively do this calculation to the end:

$$
x_t = \sum_{p=0}^{t-1} (A + B\hat{K}_{t-1}) \cdots (A + B\hat{K}_{p+1})(B\eta_p + \varepsilon_p) + (A + B\hat{K}_{t-1}) \cdots (A + BK_0)x_0.
$$

$$
u_t = \sum_{p=0}^{t-1} \hat{K}_t(A + B\hat{K}_{t-1}) \cdots (A + B\hat{K}_{p+1})(B\eta_p + \varepsilon_p) + \hat{K}_t(A + B\hat{K}_{t-1}) \cdots (A + BK_0)x_0 + \eta_t.
$$

■

H.1.3 The proof of Lemma 20

Lemma. *Assume Eq. 23, then*

1.

$$
\begin{aligned}
&\sum_{i=1}^{t-1} \sum_{p=0}^{i-1} \sum_{q=0}^{i-1} \left[(A + BK)^{i-p-1}\right] (B\eta_p + \varepsilon_p)(B\eta_q + \varepsilon_q)^\top \left[(A + BK)^{i-q-1}\right]^\top \\
&\quad = t \sum_{p=0}^{\infty} L^p (L^p)^\top \sigma^2 + t^\beta \frac{\tau^2}{\beta} \log^\alpha(t)(1 + o_p(1)) \sum_{q=0}^{\infty} L^q BB^\top [L^q]^\top \\
&\quad = t^\beta \log^\alpha(t)(C_t + o_p(1)).
\end{aligned}
$$

2.

$$
\begin{aligned}
&\sum_{i=1}^{t-1} \sum_{p=0}^{i-1} \sum_{q=0}^{i-1} \left[(A + B\hat{K}_{i-1}) \cdots (A + B\hat{K}_{p+1}) - (A + BK)^{i-p-1}\right] \\
&\quad \cdot (B\eta_p + \varepsilon_p)(B\eta_q + \varepsilon_q)^\top \left[(A + BK)^{i-q-1}\right]^\top = \mathcal{O}_p(t^{1-\beta/2} \log^{\frac{-\alpha+3}{2}}(t)).
\end{aligned}
$$

3.

$$
\begin{aligned}
&\sum_{i=1}^{t-1} \sum_{p=0}^{i-1} \sum_{q=0}^{i-1} \left[(A + BK)^{i-p-1}\right] (B\eta_p + \varepsilon_p)(B\eta_q + \varepsilon_q)^\top \\
&\quad \cdot \left[(A + B\hat{K}_{t-1}) \cdots (A + B\hat{K}_{q+1}) - (A + BK)^{i-q-1}\right]^\top = \mathcal{O}_p(t^{1-\beta/2} \log^{\frac{-\alpha+3}{2}}(t)).
\end{aligned}
$$

4.

$$
\begin{aligned}
&\sum_{i=1}^{t-1} \sum_{p=0}^{i-1} \sum_{q=0}^{i-1} \left[(A + B\hat{K}_{t-1}) \cdots (A + B\hat{K}_{p+1}) - (A + BK)^{i-p-1}\right] \\
&\quad \cdot (B\eta_p + \varepsilon_p)(B\eta_q + \varepsilon_q)^\top \left[(A + B\hat{K}_{t-1}) \cdots (A + B\hat{K}_{q+1}) - (A + BK)^{i-q-1}\right]^\top \\
&\quad = \mathcal{O}_p(t^{1-\beta/2} \log^{\frac{-\alpha+3}{2}}(t)).
\end{aligned}
$$

Proof The first step is to show the order of 2nd, 3rd and 4th part because they follow by the same method, especially the second part is just a transpose of the third part. Then we can focus on analyzing the first part, which is replacing all controllers $\hat{K}_t$ by optimal controller K.

Second Part With Lemma 18 in hand, now we are in good shape to start our proof with the second part showing

$$\sum_{i=1}^{t-1}\sum_{p=0}^{i-1}\sum_{q=0}^{i-1}\left[(A+B\hat{K}_{i-1})\cdots(A+B\hat{K}_{p+1})-(A+BK)^{i-p-1}\right]$$
$$(B\eta_p+\varepsilon_p)(B\eta_q+\varepsilon_q)^\top\left[(A+BK)^{i-q-1}\right]^\top=\mathcal{O}_p(t^{1-\beta/2}\log^{\frac{-\alpha+3}{2}}(t)).$$

Since we have already shown the uniform bound of $(B\eta_p+\varepsilon_p)(B\eta_q+\varepsilon_q)^\top$ in Lemma 18, and that $\left[(A+BK)^{i-q-1}\right]^\top$ has an exponential decay rate, the main difficulty in bounding the second part is to give a tight bound on $\left[(A+B\hat{K}_{i-1})\cdots(A+B\hat{K}_{p+1})-(A+BK)^{i-p-1}\right]$.

Recall the conclusion of Lemma 18:

$$\|(L+B\delta_{i-1})\cdots(L+B\delta_{p+1})\|=\mathcal{O}(\rho_L^{i-p})\text{ a.s.}\tag{74}$$

Thus

$$\begin{aligned}
&\|(A+B\hat{K}_{i-1})\cdots(A+B\hat{K}_{p+1})-(A+BK)^{i-p-1}\|\\
=&\|(L+B\delta_{i-1})\cdots(L+B\delta_{p+1})-L^{i-p-1}\|\\
\le&\|B\delta_{i-1}(L+B\delta_{i-2})\cdots(L+B\delta_{p+1})\|+\|LB\delta_{i-2}(L+B\delta_{i-3})\cdots(L+B\delta_{p+1})\|+\cdots\|L^{i-p-2}B\delta_{p+1}\|\\
&\text{(For example, }(L+B\delta_3)(L+B\delta_2)(L+B\delta_1)-L^3=\delta_3(L+B\delta_2)(L+B\delta_1)+L\delta_2(L+B\delta_1)+L^2B\delta_1)\\
\le&\|B\delta_{i-1}\|\|(L+B\delta_{i-2})\cdots(L+B\delta_{p+1})\|+\|B\delta_{i-2}\|\|L(L+B\delta_{i-3})\cdots(L+B\delta_{p+1})\|+\cdots\|B\delta_{p+1}\|\|L^{i-p-2}\|\\
\le&\mathcal{O}(\rho_L^{i-p})(\|\delta_{i-1}\|+\cdots+\|\delta_{p+1}\|)\text{ a.s.}\qquad\text{(using Eq. 74)}
\end{aligned}\tag{75}$$

Now the L2 norm of the second term can be bounded as

$$
\begin{aligned}
&\left\| \sum_{i=1}^{t-1}\sum_{p=0}^{i-1}\sum_{q=0}^{i-1} \left[(A+B\hat{K}_{i-1})\cdots(A+B\hat{K}_{p+1}) - (A+BK)^{i-p-1}\right] \right.\\
&\qquad \left. \cdot\, (B\eta_p+\varepsilon_p)(B\eta_q+\varepsilon_q)^\top \left[(A+BK)^{i-q-1}\right]^\top \right\|\\
&\qquad \le \sum_{i=1}^{t-1}\sum_{p=0}^{i-1}\sum_{q=0}^{i-1} \left\| \left[(A+B\hat{K}_{i-1})\cdots(A+B\hat{K}_{p+1}) - (A+BK)^{i-p-1}\right] \right\|\\
&\qquad\qquad \cdot \left\| (B\eta_p+\varepsilon_p)(B\eta_q+\varepsilon_q)^\top \right\| \left\| \left[(A+BK)^{i-q-1}\right]^\top \right\|\\
&\qquad \le \sum_{i=1}^{t-1}\sum_{p=0}^{i-1}\sum_{q=0}^{i-1} \mathcal{O}(\rho_L^{i-p})(\|\delta_{i-1}\|+\cdots+\|\delta_{p+1}\|)\mathcal{O}(\rho_L^{i-q})\|(B\eta_p+\varepsilon_p)(B\eta_q+\varepsilon_q)^\top\| \text{ a.s.}\\
&\qquad \le \sum_{i=1}^{t-1}\sum_{p=0}^{i-1}\sum_{q=0}^{i-1} \mathcal{O}(\rho_L^{i-p})(\|\delta_{i-1}\|+\cdots+\|\delta_{p+1}\|)\mathcal{O}(\rho_L^{i-q})(\|B\eta_p+\varepsilon_p\|^2+\|B\eta_q+\varepsilon_q\|^2) \text{ a.s.}
\end{aligned}
\tag{76}
$$

At first glance it seems like there is no way this would generate the desired bound, because the $\|\delta_{i-1}\|+\cdots+\|\delta_{p+1}\|$ term could diverge when i is large. However, thanks to the exponentially decaying term $\mathcal{O}(\rho_L^{i-p})$, we can avoid this by changing the order of summation:

$$
\begin{aligned}
\sum_{p=0}^{i-1}\mathcal{O}(\rho_L^{i-p})(\|\delta_{i-1}\|+\cdots+\|\delta_{p+1}\|) =& \sum_{p=0}^{i-1}\mathcal{O}(\rho_L^{i-p})\sum_{j=p+1}^{i-1}\|\delta_j\|\\
=& \sum_{p=0}^{i-1}\sum_{j=p+1}^{i-1}\mathcal{O}(\rho_L^{i-p})\|\delta_j\|\\
=& \sum_{j=1}^{i-1}\sum_{p=0}^{j-1}\mathcal{O}(\rho_L^{i-p})\|\delta_j\| \qquad \text{(exchange the order of summation)}\\
=& \sum_{j=1}^{i-1}\|\delta_j\|\sum_{p=0}^{j-1}\mathcal{O}(\rho_L^{i-p})\\
=& \sum_{j=1}^{i-1}\|\delta_j\|\mathcal{O}(\rho_L^{i-j})
\end{aligned}
\tag{77}
$$

The final form is almost the same as the beginning, except that the summation of δ_i disappears. Restart from Eq. 76, and remember to use Eq. 77 (**Additionally, when $p = 0, 1$, $\mathcal{O}(\log(p))$ is meant to be $\mathcal{O}(1)$ a.s.**):

$$
\left\| \sum_{i=1}^{t-1}\sum_{p=0}^{i-1}\sum_{q=0}^{i-1} \left[(A+B\hat{K}_{i-1})\cdots(A+B\hat{K}_{p+1}) - (A+BK)^{i-p-1}\right] \right.
$$

$$
\begin{aligned}
&\cdot (B\eta_p+\varepsilon_p)(B\eta_q+\varepsilon_q)^\top \left[(A+BK)^{i-q-1}\right]^\top \Bigg\| \\
&\le \sum_{i=1}^{t-1}\sum_{p=0}^{i-1}\sum_{q=0}^{i-1}\mathcal{O}(\rho_L^{i-p})(\|\delta_{i-1}\|+\cdots+\|\delta_{p+1}\|)\mathcal{O}(\rho_L^{i-q})(\|B\eta_p+\varepsilon_p\|^2+\|B\eta_q+\varepsilon_q\|^2) \text{ a.s.} \quad \text{(by Lemma 18)} \\
&\le \sum_{i=1}^{t-1}\sum_{p=0}^{i-1}\sum_{q=0}^{i-1}\mathcal{O}(\rho_L^{i-p})(\|\delta_{i-1}\|+\cdots+\|\delta_{p+1}\|)\mathcal{O}(\rho_L^{i-q})(\mathcal{O}(\log(p))+\mathcal{O}(\log(q))) \text{ a.s.} \\
&\le \sum_{i=1}^{t-1}\sum_{p=0}^{i-1}\sum_{q=0}^{i-1}\mathcal{O}(\rho_L^{i-p})(\|\delta_{i-1}\|+\cdots+\|\delta_{p+1}\|)\mathcal{O}(\rho_L^{i-q})\mathcal{O}(\log(t)) \text{ a.s.} \\
&= \mathcal{O}(\log(t))\sum_{i=1}^{t-1}\sum_{p=0}^{i-1}\mathcal{O}(\rho_L^{i-p})(\|\delta_{i-1}\|+\cdots+\|\delta_{p+1}\|) \text{ a.s.} \\
&= \mathcal{O}(\log(t))\sum_{i=1}^{t-1}\sum_{j=1}^{i-1}\|\delta_j\|\mathcal{O}(\rho_L^{i-j}) \text{ a.s.} \\
&= \mathcal{O}(\log(t))\sum_{j=1}^{t-1}\|\delta_j\|\sum_{i=j+1}^{t-1}\mathcal{O}(\rho_L^{i-j}) \text{ a.s.} \qquad \text{(by Eq. 77)} \\
&= \mathcal{O}(\log(t))\sum_{j=1}^{t-1}\|\delta_j\| \text{ a.s.} \\
&= \mathcal{O}(\log(t))\left(\sum_{j=1}^{t-1}\mathcal{O}(j^{-\frac{\beta}{2}}\log^{\frac{-\alpha+1}{2}}(j))\right) \text{ a.s.} \qquad \text{(by Eq. 81)} \\
&= \mathcal{O}(\log(t))\mathcal{O}(t^{1-\frac{\beta}{2}}\log^{\frac{-\alpha+1}{2}}(t)) \text{ a.s.} \\
&= \mathcal{O}(t^{1-\beta/2}\log^{\frac{-\alpha+3}{2}}(t)) \text{ a.s.} \qquad (78)
\end{aligned}
$$

We know that for any matrix A, $\|A\| \le \|A\|_F \le \sqrt{r}\|A\|$, where r is the rank of matrix A. Thus Eq. 78 implies an upper bound on the Frobenius norm, and the Frobenius norm implies entry-wise upper bound:

$$
\begin{aligned}
&\sum_{i=1}^{t-1}\sum_{p=0}^{i-1}\sum_{q=0}^{i-1}\left[(A+B\hat{K}_{i-1})\cdots(A+B\hat{K}_{p+1})-(A+BK)^{i-p-1}\right] \\
&\qquad \cdot (B\eta_p+\varepsilon_p)(B\eta_q+\varepsilon_q)^\top \left[(A+BK)^{i-q-1}\right]^\top = \mathcal{O}(t^{1-\beta/2}\log^{\frac{-\alpha+3}{2}}(t)) \text{ a.s.}
\end{aligned}
$$

Third Part This part is the transpose of the second part, thus shares the same result with the second part.

Fourth Part We wish to show that

$$
\Bigg\| \sum_{i=1}^{t-1}\sum_{p=0}^{i-1}\sum_{q=0}^{i-1}\left[(A+B\hat{K}_{i-1})\cdots(A+B\hat{K}_{p+1})-(A+BK)^{i-p-1}\right]
$$

$$\cdot (B\eta_p + \varepsilon_p)(B\eta_q + \varepsilon_q)^\top \left[(A + B\hat{K}_{i-1}) \cdots (A + B\hat{K}_{q+1}) - (A + BK)^{i-q-1}\right]^\top \Bigg\|$$

$$= \mathcal{O}_p(t^{1-\beta/2} \log^{\frac{-\alpha+3}{2}}(t)).$$

By Lemma 43 we have

$$\|(A + B\hat{K}_{i-1}) \cdots (A + B\hat{K}_{q+1})\| = \mathcal{O}(\rho_L^{i-q}) \text{ a.s.},$$

and

$$\|(A + BK)^{i-q-1}\| = \mathcal{O}(\rho_L^{i-q}) \text{ a.s.}$$

Thus,

$$(A + B\hat{K}_{i-1}) \cdots (A + B\hat{K}_{q+1}) - (A + BK)^{i-q-1} = \mathcal{O}(\rho_L^{i-q}) \text{ a.s.}$$

Combining this with Eq. 75,

$$\|\sum_{i=1}^{t-1}\sum_{p=0}^{i-1}\sum_{q=0}^{i-1} \left[(A + B\hat{K}_{i-1}) \cdots (A + B\hat{K}_{p+1}) - (A + BK)^{i-p-1}\right]$$

$$\cdot (B\eta_p + \varepsilon_p)(B\eta_q + \varepsilon_q)^\top \left[(A + B\hat{K}_{i-1}) \cdots (A + B\hat{K}_{q+1}) - (A + BK)^{i-q-1}\right]^\top \|$$

$$\leq \sum_{i=1}^{t-1}\sum_{p=0}^{i-1}\sum_{q=0}^{i-1} \mathcal{O}(\rho_L^{i-p})(\|\delta_{i-1}\| + \cdots + \|\delta_{p+1}\|)\|(B\eta_p + \varepsilon_p)(B\eta_q + \varepsilon_q)^\top\|\mathcal{O}(\rho_L^{i-q}) \text{ a.s.}$$

$$\leq \sum_{i=1}^{t-1}\sum_{p=0}^{i-1}\sum_{q=0}^{i-1} \mathcal{O}(\rho_L^{i-p})(\|\delta_{i-1}\| + \cdots + \|\delta_{p+1}\|)\mathcal{O}(\rho_L^{i-q})(\|B\eta_p + \varepsilon_p\|^2 + \|B\eta_q + \varepsilon_q\|^2) \text{ a.s.},$$

which is exactly the same as the final line of Eq. 76. Then following the same proof procedure as in the second part we can get the same order as in the second part: $\mathcal{O}(t^{1-\beta/2} \log^{\frac{-\alpha+3}{2}}(t))$ a.s.

Summarize second, third, and fourth parts To sum up, all three parts are bounded by the same order $\mathcal{O}(t^{1-\beta/2} \log^{\frac{-\alpha+3}{2}}(t))$ a.s.

First Part It remains to show

$$\sum_{i=1}^{t-1}\sum_{p=0}^{i-1}\sum_{q=0}^{i-1} \left[(A + BK)^{i-p-1}\right] (B\eta_p + \varepsilon_p)(B\eta_q + \varepsilon_q)^\top \left[(A + BK)^{i-q-1}\right]^\top$$

$$= t\sum_{p=0}^{\infty} L^p(L^p)^\top\sigma^2 + t^\beta \frac{\tau^2}{\beta} \log^\alpha(t) \sum_{q=0}^{\infty} L^q BB^\top [L^q]^\top (I_n + o_p(1)).$$

Recall $L = A + BK$. We divide the left hand side into two separate parts:

- The part where $p \neq q$. We will show this part is dominated by the $p = q$ part and is only of order $\mathcal{O}_p(t^{1/2})$.

$$G_t \coloneqq \sum_{i=1}^{t-1}\sum_{p\neq q}^{i-1} L^{i-p-1}(B\eta_p + \varepsilon_p)(B\eta_q + \varepsilon_q)^\top [L^{i-q-1}]^\top = \mathcal{O}_p(t^{1/2}).$$

- The part where $p = q$. We will show that

$$\begin{aligned}&\sum_{i=1}^{t-1}\sum_{p=q=0}^{i-1} L^{i-p-1}(B\eta_p+\varepsilon_p)(B\eta_q+\varepsilon_q)^\top[L^{i-q-1}]^\top\\&\quad= t\sum_{p=0}^{\infty} L^p(L^p)^\top\sigma^2 + t^\beta\frac{\tau^2}{\beta}\log^\alpha(t)\sum_{q=0}^{\infty} L^qBB^\top[L^q]^\top(I_n+o_p(1)).\end{aligned}$$

Let us first consider the part where $p \neq q$. We will show the order of G_t by considering its expectation and variance. Since G_t is a summation of cross terms and $\mathbb{E}(B\eta_p+\varepsilon_p) = 0$, $\mathbb{E}(G_t) = 0$. Now it remains to consider the variance

$$\begin{aligned}\mathbb{E}(\|G_t\|_F^2) =&\mathbb{E}(\mathbf{Tr}(G_t^2))\\ =&\mathbb{E}\Bigg(\mathbf{Tr}\Bigg(\sum_{p\neq q}^{t-1}\sum_{i=p\vee q+1}^{t-1}\sum_{j=p\vee q+1}^{t-1} L^{i-p-1}(B\eta_p+\varepsilon_p)(B\eta_q+\varepsilon_q)^\top[L^{i-q-1}]^\top\\&\cdot L^{j-p-1}(B\eta_p+\varepsilon_p)(B\eta_q+\varepsilon_q)^\top[L^{j-q-1}]^\top\Bigg)\Bigg)\\ &+\mathbb{E}\Bigg(\mathbf{Tr}\Bigg(\sum_{p\neq q}^{t-1}\sum_{i=p\vee q+1}^{t-1}\sum_{j=p\vee q+1}^{t-1} L^{i-p-1}(B\eta_p+\varepsilon_p)(B\eta_q+\varepsilon_q)^\top[L^{i-q-1}]^\top\\&\cdot L^{j-q-1}(B\eta_q+\varepsilon_q)(B\eta_p+\varepsilon_p)^\top[L^{j-p-1}]^\top\Bigg)\Bigg)\\&\text{(terms with odd power go away in expectation)}\ .\end{aligned}$$

It is sufficient to consider the first term in the previous expression, and the other term can be analyzed in exactly the same way. Notice the following relationship on any square matrix A with dimension n

$$\mathbf{Tr}^2(A) \le n\|A\|_F^2 \le n\cdot n\|A\|^2.$$

That is

$$\mathbf{Tr}(A) \le n\|A\|.$$

Then

$$\begin{aligned}&\mathbb{E}\Bigg(\mathbf{Tr}\Bigg(\sum_{p\neq q}^{t-1}\sum_{i=p\vee q+1}^{t-1}\sum_{j=p\vee q+1}^{t-1} L^{i-p-1}(B\eta_p+\varepsilon_p)(B\eta_q+\varepsilon_q)^\top[L^{i-q-1}]^\top\\&\qquad\cdot L^{j-p-1}(B\eta_p+\varepsilon_p)(B\eta_q+\varepsilon_q)^\top[L^{j-q-1}]^\top\Bigg)\Bigg)\\&\quad\le n\mathbb{E}\Bigg\|\sum_{p\neq q}^{t-1}\sum_{i=p\vee q+1}^{t-1}\sum_{j=p\vee q+1}^{t-1} L^{i-p-1}(B\eta_p+\varepsilon_p)(B\eta_q+\varepsilon_q)^\top[L^{i-q-1}]^\top\\&\qquad\quad\cdot L^{j-p-1}(B\eta_p+\varepsilon_p)(B\eta_q+\varepsilon_q)^\top[L^{j-q-1}]^\top\Bigg\|\end{aligned}$$

$$
\begin{aligned}
&\le \mathbb{E}\sum_{p\neq q}^{t-1}\sum_{i=p\vee q+1}^{t-1}\sum_{j=p\vee q+1}^{t-1}\mathcal{O}(\rho_L^{i-p}\rho_L^{i-q})\|B\eta_p+\varepsilon_p\|_2^2\|B\eta_q+\varepsilon_q\|_2^2 O(\rho_L^{j-p}\rho_L^{j-q}) \qquad \text{(by Lemma 43)}\\
&=\mathcal{O}\left(\sum_{p\neq q}^{t-1}\sum_{i=p\vee q+1}^{t-1}\sum_{j=p\vee q+1}^{t-1}\rho_L^{2i-p-q}\rho_L^{2j-p-q}\right)\\
&=\mathcal{O}\left(\sum_{p>q}^{t-1}\rho_L^{2(p-q)}\right) \quad \text{(WLOG consider the part where } p>q)\\
&=\mathcal{O}\left(\sum_{q=0}^{t-1}1\right)\\
&=\mathcal{O}(t).
\end{aligned}
$$

Thus the entry-wise standard error of G_t is of order $\mathcal{O}(t^{1/2})$. Combining this with the fact that $\mathbb{E}G_t=0$, we have

$$
G_t \coloneqq \sum_{i=1}^{t-1}\sum_{p\neq q}^{i-1}L^{i-p-1}(B\eta_p+\varepsilon_p)(B\eta_q+\varepsilon_q)^\top[L^{i-q-1}]^\top=\mathcal{O}_p(t^{1/2}). \tag{79}
$$

and it remains to consider

$$
R\coloneqq\sum_{i=1}^{t-1}\sum_{p=0}^{i-1}L^{i-p-1}(B\eta_p+\varepsilon_p)(B\eta_p+\varepsilon_p)^\top[L^{i-p-1}]^\top.
$$

Consider the expectation of R: $\mathbb{E}(R)=R_0$, where

$$
R_0\coloneqq\sum_{i=0}^{t-1}\sum_{p=0}^{i-1}L^{i-p-1}(p^{\beta-1}\log^\alpha(p)BB^\top\tau^2+I_n\sigma^2)[L^{i-p-1}]^\top.
$$

Let us first show $R-R_0=\mathcal{O}_p(t^{1/2})$, and after that we only need to consider R_0, which is the dominating term. We know that $B\eta_p+\varepsilon_p$ has a finite fourth moment, so the sum of the variances of each element of $R-R_0$ can be written as

$$
\begin{aligned}
\mathbb{E}\|R-R_0\|_F^2 =&\mathbb{E}(\mathbf{Tr}((R-R_0)^2))\\
\le&\mathbb{E}(\mathbf{Tr}(\sum_{p=0}^{t-1}\sum_{i=p+1}^{t-1}\sum_{j=p+1}^{t-1}L^{i-p-1}[(B\eta_p+\varepsilon_p)(B\eta_p+\varepsilon_p)^\top-(p^{\beta-1}\log^\alpha(p)BB^\top\tau^2+I_m\sigma^2)]\\
&\cdot[L^{i-p-1}]^\top L^{j-p-1}[(B\eta_p+\varepsilon_p)(B\eta_p+\varepsilon_p)^\top-(p^{\beta-1}\log^\alpha(p)BB^\top\tau^2+I_m\sigma^2)][L^{j-p-1}]^\top))\\
\le&n\mathbb{E}\|\sum_{p=0}^{t-1}\sum_{i=p+1}^{t-1}\sum_{j=p+1}^{t-1}L^{i-p-1}[(B\eta_p+\varepsilon_p)(B\eta_p+\varepsilon_p)^\top-(p^{\beta-1}\log^\alpha(p)BB^\top\tau^2+I_m\sigma^2)]\\
&\cdot[L^{i-p-1}]^\top L^{j-p-1}[(B\eta_p+\varepsilon_p)(B\eta_p+\varepsilon_p)^\top-(p^{\beta-1}\log^\alpha(p)BB^\top\tau^2+I_m\sigma^2)][L^{j-p-1}]^\top\|\\
\le&\mathcal{O}(\mathbb{E}\sum_{p=0}^{t-1}\sum_{i=p+1}^{t-1}\sum_{j=p+1}^{t-1}\rho_L^{2i-2p}\|(B\eta_p+\varepsilon_p)(B\eta_p+\varepsilon_p)^\top-(p^{\beta-1}\log^\alpha(p)BB^\top\tau^2+I_m\sigma^2)\|^2\rho_L^{2j-2p})
\end{aligned}
$$

$$
\begin{aligned}
=&\mathcal{O}(\sum_{p=0}^{t-1}\mathbb{E}\|(B\eta_p+\varepsilon_p)(B\eta_p+\varepsilon_p)^\top-(p^{\beta-1}\log^\alpha(p)BB^\top\tau^2+I_m\sigma^2)\|^2)\\
=&\mathcal{O}(t).
\end{aligned}
$$

Thus $R-R_0=\mathcal{O}_p(t^{1/2})$. Now we only need to focus on:

$$
R_0=\sum_{i=1}^{t-1}\sum_{p=0}^{i-1}L^{i-p-1}(p^{\beta-1}\log^\alpha(p)BB^\top\tau^2+I_m\sigma^2)[L^{i-p-1}]^\top.
$$

Again, when $p=0,1$, $p^{\beta-1}\log^\alpha(p)$ should be considered as 1. Let us start from the identity matrix part $\sum_{i=1}^{t-1}\sum_{p=0}^{i-1}L^{i-p-1}I_m\sigma^2[L^{i-p-1}]^\top$.

$$
\begin{aligned}
\sum_{i=1}^{t-1}\sum_{p=0}^{i-1}L^{i-p-1}[L^{i-p-1}]^\top=&\sum_{i=1}^{t-1}\sum_{q=0}^{i-1}L^q[L^q]^\top\\
=&\sum_{i=1}^{t-1}(\sum_{p=0}^{\infty}L^p(L^p)^\top-\sum_{q=i}^{\infty}L^q[L^q]^\top)\\
=&t\sum_{p=0}^{\infty}L^p(L^p)^\top-\sum_{i=1}^{t-1}\sum_{q=i}^{\infty}L^q[L^q]^\top.
\end{aligned}
$$

Notice

$$
\begin{aligned}
\|\sum_{i=1}^{t-1}\sum_{q=i}^{\infty}L^q[L^q]^\top\|\leq&\sum_{i=1}^{t-1}\sum_{q=i}^{\infty}\mathcal{O}(\rho_L^{2q})\\
=&\sum_{i=1}^{t-1}\mathcal{O}(\rho_L^{2i})\\
=&\mathcal{O}(1).
\end{aligned}
$$

Thus

$$
\sum_{i=1}^{t-1}\sum_{p=0}^{i-1}L^{i-p-1}[L^{i-p-1}]^\top=t\sum_{p=0}^{\infty}L^p(L^p)^\top+\mathcal{O}(1).
$$

On the other hand (**when $p = 0, 1$, $p^{\beta-1}\log^{\alpha}(p)$ is meant to be** 1),

$$\begin{aligned}
&\sum_{i=1}^{t-1}\sum_{p=0}^{i-1} L^{i-p-1}p^{\beta-1}\log^{\alpha}(p)BB^{\top}[L^{i-p-1}]^{\top} \\
&\quad= \sum_{p=0}^{t-2}\sum_{i=p+1}^{t-1} L^{i-p-1}p^{\beta-1}\log^{\alpha}(p)BB^{\top}[L^{i-p-1}]^{\top} \\
&\quad= \sum_{p=0}^{t-2} p^{\beta-1}\log^{\alpha}(p)\sum_{q=0}^{t-p-2} L^{q}BB^{\top}[L^{q}]^{\top} \\
&\quad= \sum_{p=0}^{t-2} p^{\beta-1}\log^{\alpha}(p)\left(\sum_{q=0}^{\infty} L^{q}BB^{\top}[L^{q}]^{\top} - \sum_{q=t-p-1}^{\infty} L^{q}BB^{\top}[L^{q}]^{\top}\right) \\
&\quad= \sum_{p=0}^{t-2} p^{\beta-1}\log^{\alpha}(p)\left(\sum_{q=0}^{\infty} L^{q}BB^{\top}[L^{q}]^{\top} - \mathcal{O}(\rho_L^{2(t-p-1)})\right) \\
&\quad= \sum_{p=0}^{t-2} p^{\beta-1}\log^{\alpha}(p)\sum_{q=0}^{\infty} L^{q}BB^{\top}[L^{q}]^{\top} + \sum_{p=0}^{t-2} p^{\beta-1}\log^{\alpha}(p)\mathcal{O}\left(\rho_L^{2(t-p-1)}\right) \\
&\quad\leq \sum_{p=0}^{t-2} p^{\beta-1}\log^{\alpha}(p)\sum_{q=0}^{\infty} L^{q}BB^{\top}[L^{q}]^{\top} + \sum_{p=0}^{t-2} \mathcal{O}(1)\mathcal{O}\left(\rho_L^{2(t-p-1)}\right) \\
&\quad= \sum_{p=0}^{t-2} p^{\beta-1}\log^{\alpha}(p)\sum_{q=0}^{\infty} L^{q}BB^{\top}[L^{q}]^{\top} + \mathcal{O}(1).
\end{aligned} \tag{80}$$

Now it remains to calculate $\sum_{p=0}^{t-2} p^{\beta-1}\log^{\alpha}(p)$. Let us consider a more general case $\sum_{p=0}^{t} p^{\gamma}\log^{\alpha}(p)$ where $\gamma > -1$ and α is any real number. It is clear that this summation goes to infinity when $t \to \infty$. Recall the Stolz–Cesàro theorem:

Theorem 45 (Stolz–Cesàro)**.** *Let $\{a_t\}_{t\geq 1}$ and $\{b_t\}_{t\geq 1}$ be two sequences of real numbers. Assume that $\{b_t\}_{t\geq 1}$ is a strictly monotone and divergent sequence and the following limit exists:*

$$\lim_{t\to\infty}\frac{a_{t+1}-a_t}{b_{t+1}-b_t} = l$$

Then, the limit

$$\lim_{t\to\infty}\frac{a_t}{b_t} = l$$

In Theorem 45, we choose a_t and b_t to be $\sum_{p=0}^{t} p^{\gamma}\log^{\alpha}(p)$ and $t^{\gamma+1}\log^{\alpha}(t)$, respectively.

$$\begin{aligned}
\lim_{t\to\infty}\frac{a_t-a_{t-1}}{b_t-b_{t-1}} &= \lim_{t\to\infty}\frac{t^{\gamma}\log^{\alpha}(t)}{t^{\gamma+1}\log^{\alpha}(t)-(t-1)^{\gamma+1}\log^{\alpha}(t-1)} \\
&= \lim_{t\to\infty}\frac{1}{t-(\frac{t-1}{t})^{\gamma}(t-1)\left(\frac{\log(t-1)}{\log(t)}\right)^{\alpha}}
\end{aligned}$$

$$\begin{aligned}
&= \lim_{t\to\infty} \frac{1/t}{1-(1-\frac{1}{t})^{\gamma+1}\left(1+\frac{\log(t-1)-\log(t)}{\log(t)}\right)^{\alpha}} \\
&= \lim_{t\to\infty} \frac{1/t}{1-(1-\frac{\gamma+1}{t}+o(\frac{1}{t}))\left(1+\frac{-\frac{1}{t}+o(\frac{1}{t})}{\log(t)}\right)^{\alpha}} \\
&= \lim_{t\to\infty} \frac{1/t}{1-(1-\frac{\gamma+1}{t}+o(\frac{1}{t}))\left(1+o(\frac{1}{t})\right)^{\alpha}} \\
&= \lim_{t\to\infty} \frac{1/t}{1-(1-\frac{\gamma+1}{t}+o(\frac{1}{t}))e^{\alpha\log\left(1+o(\frac{1}{t})\right)}} \\
&= \lim_{t\to\infty} \frac{1/t}{1-(1-\frac{\gamma+1}{t}+o(\frac{1}{t}))e^{\alpha o(\frac{1}{t})}} \\
&= \lim_{t\to\infty} \frac{1/t}{1-(1-\frac{\gamma+1}{t}+o(\frac{1}{t}))\left(1+o(\frac{\alpha}{t})\right)} \\
&= \lim_{t\to\infty} \frac{1/t}{\frac{\gamma+1}{t}+o(\frac{1}{t})} \\
&= \frac{1}{\gamma+1}.
\end{aligned}$$

By Theorem 45, we know

$$\lim_{t\to\infty}\frac{a_t}{b_t} = \lim_{t\to\infty}\frac{\sum_{p=0}^{t} p^{\gamma}\log^{\alpha}(p)}{t^{\gamma+1}\log^{\alpha}(t)} = \frac{1}{\gamma+1}$$

That is to say, for any $\gamma > -1$:

$$\sum_{p=0}^{t} p^{\gamma}\log^{\alpha}(p) = \frac{1}{\gamma+1}t^{\gamma+1}\log^{\alpha}(t)(1+o(1)). \tag{81}$$

Following Eqs. 80 and 81,

$$\begin{aligned}
&\sum_{i=1}^{t-1}\sum_{p=0}^{i-1} L^{i-p-1}p^{\beta-1}\log^{\alpha}(p)BB^{\top}[L^{i-p-1}]^{\top} \\
=&\sum_{p=0}^{t} p^{\beta-1}\log^{\alpha}(p)\sum_{q=0}^{\infty} L^{q}BB^{\top}[L^{q}]^{\top} + \mathcal{O}(1) \\
=&\frac{t^{\beta}}{\beta}\log^{\alpha}(t)(1+o(1))\sum_{q=0}^{\infty} L^{q}BB^{\top}[L^{q}]^{\top} + \mathcal{O}(1) \\
=&\frac{t^{\beta}}{\beta}\log^{\alpha}(t)\sum_{q=0}^{\infty} L^{q}BB^{\top}[L^{q}]^{\top}(I_n+o(1)).
\end{aligned}$$

To sum up,

$$R_0 = t\sum_{p=0}^{\infty} L^{p}(L^{p})^{\top}\sigma^{2} + \frac{t^{\beta}}{\beta}\log^{\alpha}(t)\sum_{q=0}^{\infty} L^{q}BB^{\top}[L^{q}]^{\top}(I_n+o(1)).$$

Recall that $R - R_0 = \mathcal{O}_p(t^{1/2})$, so

$$
\begin{aligned}
R &= \sum_{i=1}^{t-1}\sum_{p=0}^{i-1} L^{i-p-1}(B\eta_p + \varepsilon_p)(B\eta_p + \varepsilon_p)^\top [L^{i-p-1}]^\top \\
&= t\sum_{p=0}^{\infty} L^p (L^p)^\top \sigma^2 + t^\beta \frac{\tau^2}{\beta}\log^\alpha(t)\sum_{q=0}^{\infty} L^q BB^\top [L^q]^\top (I_n + o_p(1)).
\end{aligned}
$$

Recall Eq. 79:

$$
\sum_{i=1}^{t-1}\sum_{p\neq q}^{i-1} L^{i-p-1}(B\eta_p + \varepsilon_p)(B\eta_q + \varepsilon_q)^\top [L^{i-q-1}]^\top = \mathcal{O}_p(t^{1/2}).
$$

Finally we proved the order of the first part:

$$
\begin{aligned}
&\sum_{i=1}^{t-1}\sum_{p=0}^{i-1}\sum_{q=0}^{i-1} \left[(A+BK)^{i-p-1}\right](B\eta_p + \varepsilon_p)(B\eta_q + \varepsilon_q)^\top \left[(A+BK)^{i-q-1}\right]^\top \\
&= \sum_{i=1}^{t-1}\sum_{p=0}^{i-1}\sum_{q=0}^{i-1} L^{i-p-1}(B\eta_p + \varepsilon_p)(B\eta_q + \varepsilon_q)^\top [L^{i-q-1}]^\top \\
&= t\sum_{p=0}^{\infty} L^p (L^p)^\top \sigma^2 + t^\beta \frac{\tau^2}{\beta}\log^\alpha(t)\sum_{q=0}^{\infty} L^q BB^\top [L^q]^\top (I_n + o_p(1)) \\
&= t^\beta \log^\alpha(t)(C_t + o_p(1)) \qquad \text{(by } C_t \text{ definition Eq. 29).}
\end{aligned}
$$

■

H.1.4 The proof of Lemma 21

Lemma. *Assume Eq. 23, then*

1. $\sum_{i=0}^{t-1}\left[(A+B\hat{K}_{i-1})\cdots(A+BK_0)x_0\right]\left[\sum_{q=0}^{i-1}(A+B\hat{K}_{i-1})\cdots(A+B\hat{K}_{q+1})(B\eta_q+\varepsilon_q)\right]^T = \tilde{\mathcal{O}}(1)$ *a.s.*
2. $\sum_{i=0}^{t-1}\left[(A+B\hat{K}_{i-1})\cdots(A+BK_0)x_0\right]\left[(A+B\hat{K}_{i-1})\cdots(A+BK_0)x_0\right]^T = \mathcal{O}(1)$ *a.s.*

Proof This can be proved using a similar technique as in Appendix H.1.3. **Recall that when $q = 0, 1$, $\log^\alpha(q)$ is taken to be 1.**

$$
\begin{aligned}
&\left\|\sum_{i=1}^{t-1}\sum_{q=0}^{i-1}(A+B\hat{K}_{i-1})\cdots(A+BK_0)x_0(B\eta_q+\varepsilon_q)^\top\left[(A+B\hat{K}_{i-1})\cdots(A+B\hat{K}_{q+1})\right]^\top\right\| \\
&\qquad \le \sum_{i=1}^{t-1}\sum_{q=0}^{i-1}\|(L+B\delta_{t-1})\cdots(L+B\delta_0)\|\|x_0\|\|B\eta_q+\varepsilon_q\|\|(L+B\delta_{t-1})\cdots(L+B\delta_{q+1})\|^\top
\end{aligned}
$$

$$
\begin{aligned}
&\leq \sum_{i=1}^{t-1}\sum_{q=0}^{i-1} \mathcal{O}(\rho_L^i)\|x_0\|\|B\eta_q + \varepsilon_q\|\mathcal{O}(\rho_L^{i-q}) \text{ a.s.} \qquad \text{(by Lemma 18)}\\
&\leq \sum_{i=1}^{t-1}\sum_{q=0}^{i-1} \mathcal{O}(\rho_L^{2i-q})\mathcal{O}(1)\mathcal{O}(\log^{1/2}(q)) \text{ a.s.} \quad \text{(by Lemma 18)}\\
&\leq \sum_{i=1}^{t-1}\sum_{q=0}^{i-1} \mathcal{O}(\rho_L^{2i-q})\tilde{\mathcal{O}}(1) \text{ a.s.}\\
&= \sum_{i=1}^{t-1} \mathcal{O}(\rho_L^{i})\tilde{\mathcal{O}}(1) \text{ a.s.}\\
&\leq \tilde{\mathcal{O}}(1) \text{ a.s.}
\end{aligned}
$$

Also,

$$
\begin{aligned}
&\left\|\sum_{i=1}^{t-1}\left[(A+B\hat{K}_{i-1})\cdots(A+BK_0)x_0\right]\left[(A+B\hat{K}_{i-1})\cdots(A+BK_0)x_0\right]^T\right\|\\
&\qquad\leq \sum_{i=1}^{t-1}\mathcal{O}(\rho_L^i)\|x_0\|^2\mathcal{O}(\rho_L^i) \text{ a.s.} \qquad \text{(by Lemma 18)}\\
&\qquad\leq \sum_{i=1}^{t-1}\mathcal{O}(\rho_L^{2i}) \text{ a.s.}\\
&\qquad\leq \mathcal{O}(1) \text{ a.s.}
\end{aligned}
$$

■

H.1.5 The proof of Lemma 22

Lemma. *Assume we have two matrix sequences $\{A_t\}_{t=1}^{\infty}$ and $\{B_t\}_{t=1}^{\infty}$, where A_t and B_t are $p \times p$ positive definite matrices, and*

$$A_t^2 B_t^2 \xrightarrow{P} I_p.$$

Then

$$A_t B_t \xrightarrow{P} I_p.$$

Proof The basic idea is to utilize the equivalence of entry-wise convergence and F-norm convergence and the fact that the F-norm is invariant under orthogonal transformation. We know that positive definite matrices can be diagonalized by orthogonal transformation, and these diagonal matrices are easier to deal with. Starting from our only equation

$$A_t^2 B_t^2 \xrightarrow{P} I_p.$$

Entry-wise convergence implies F-norm convergence:

$$\|A_t^2 B_t^2 - I_p\|_F \xrightarrow{P} 0.$$

By the positive definiteness of A_t and B_t, we can assume they have the diagnolization $A_t = U_{At}\Lambda_{At}U_{At}^\top$ and $B_t = U_{Bt}\Lambda_{Bt}U_{Bt}^\top$, where Λ_{At} and Λ_{Bt} are diagonal matrices with diagonal values $\lambda_{Ai,t}$ and $\lambda_{Bi,t}$ $(i = 1, 2, \cdots, p)$, and U_{At} and U_{Bt} are orthogonal matrices. With this transformation, we have

$$\|U_{At}\Lambda_{At}^2U_{At}^\top U_{Bt}\Lambda_{Bt}^2U_{Bt}^\top - I_p\|_F \xrightarrow{P} 0.$$

Since orthogonal transformation does not affect F-norm, on RHS inside the F-norm, we can multiply $U_{At}^\top$ on the left and U_{Bt} and on the right and get

$$\|\Lambda_{At}^2U_{At}^\top U_{Bt}\Lambda_{Bt}^2 - U_{At}^\top U_{Bt}\|_F \xrightarrow{P} 0.$$

Because F-norm convergence to zero is equivalent to entry-wise convergence to zero,

$$\Lambda_{At}^2U_{At}^\top U_{Bt}\Lambda_{Bt}^2 - U_{At}^\top U_{Bt} \xrightarrow{P} 0.$$

Denote $T_t := U_{At}^\top U_{Bt}$, then

$$\Lambda_{At}^2T_t\Lambda_{Bt}^2 - T_t \xrightarrow{P} 0.$$

If we consider the ijth element of the above equation:

$$\lambda_{Ai,t}^2T_{ij}\lambda_{Bj,t}^2 - T_{ij} \xrightarrow{P} 0,$$

which is

$$(\lambda_{Ai,t}\lambda_{Bj,t} - 1)(\lambda_{Ai,t}\lambda_{Bj,t} + 1)T_{ij} \xrightarrow{P} 0.$$

Since by positive definiteness we have $\lambda_{Ai,t}, \lambda_{Bj,t} > 0$, the above equation implies

$$(\lambda_{Ai,t}\lambda_{Bj,t} - 1)T_{ij} \xrightarrow{P} 0.$$

This holds for every i, j pair. If we write out this equation back to matrix form, we would get

$$\Lambda_{At}T_t\Lambda_{Bt} - T_t \xrightarrow{P} 0.$$

By the same trick this is equivalent to the F-norm form

$$\|\Lambda_{At}T_t\Lambda_{Bt} - T_t\|_F \xrightarrow{P} 0,$$

$$\|\Lambda_{At}U_{At}^\top U_{Bt}\Lambda_{Bt} - U_{At}^\top U_{Bt}\|_F \xrightarrow{P} 0.$$

On RHS inside the F-norm, we can multiply U_{At} on the left and $U_{Bt}^\top$ and on the right and get

$$\|U_{At}\Lambda_{At}U_{At}^\top U_{Bt}\Lambda_{Bt}U_{Bt}^T - I_p\|_F \xrightarrow{P} 0.$$

Plug in our definition $A_t = U_{At}\Lambda_{At}U_{At}^\top$ and $B_t = U_{Bt}\Lambda_{Bt}U_{Bt}^\top$:

$$\|A_tB_t - I_p\|_F \xrightarrow{P} 0.$$

And this implies

$$A_tB_t \xrightarrow{P} I_p.$$

■

H.1.6 The proof of Lemma 23

Lemma. *Assume Eq. 23, then*

1. $\sum_{i=0}^{t-1}(\hat{K}_i - K)x_i x_i^\top = \mathcal{O}(t^{1-\beta/2}\log^{\frac{-\alpha+3}{2}}(t))$ *a.s.*
2. $\sum_{i=0}^{t-1}\eta_i x_i^\top = o\left(t^{\beta/2}\log^{\frac{\alpha+3}{2}}(t)\right)$ *a.s.*

Proof

First part $\sum_{i=0}^{t-1}(\hat{K}_i - K)x_i x_i^\top$ By Lemma 18 we have a uniform bound for $\delta_i = \hat{K}_i - K$ and x_i. We can derive the result in the first part by directly plugging in the bound for $\|\delta_i\|$ and $\|x_i\|$.

By Lemma 18

$$\|x_i\| \leq \mathcal{O}(\log^{1/2}(t)) \text{ a.s.}$$

Thus

$$\begin{aligned}
\left\|\sum_{i=0}^{t-1}(\hat{K}_i - K)x_i x_i^\top\right\| =& \sum_{i=0}^{t-1}\|\delta_i\|\|x_i x_i^\top\| \\
\leq& \mathcal{O}(\log(t))\sum_{i=0}^{t-1}\|\delta_i\| \text{ a.s.} \qquad \text{(by Lemma 18)} \\
\leq& \mathcal{O}(\log(t))\sum_{i=0}^{t-1}\mathcal{O}(i^{-\beta/2}\log^{\frac{-\alpha+1}{2}}(i)) \text{ a.s.} \qquad \text{(by Lemma 18)} \\
\leq& \mathcal{O}(\log(t)t^{1-\beta/2}\log^{\frac{-\alpha+1}{2}}(t)) \text{ a.s.} \qquad \text{(by Eq. 81)} \\
\leq& \mathcal{O}(t^{1-\beta/2}\log^{\frac{-\alpha+3}{2}}(t)) \text{ a.s.}
\end{aligned}$$

which means (by bounding entry-wise terms by the operator norm)

$$\sum_{i=0}^{t-1}(\hat{K}_i - K)x_i x_i^\top = \mathcal{O}(t^{1-\beta/2}\log^{\frac{-\alpha+3}{2}}(t)) \text{ a.s.}$$

Second Part $\sum_{i=0}^{t-1}\eta_i x_i^\top$ Following Lemma 2 (iii) from Lai and Wei (1982):

Lemma 46. *Let $\{\epsilon_n\}$ be a martingale difference sequence with respect to an increasing sequence of σ-fields $\{\mathcal{F}_n\}$ such that $\sup_n \mathbb{E}(\varepsilon_n^2|\mathcal{F}_{n-1}) < \infty$ a.s. Let v_n be an $\mathcal{F}_{n-1}$-measurable random variable for every n. Then*

$$\sum_{i=1}^{n} v_i\epsilon_i < \infty \textit{ a.s. on } \{\sum_{i=1}^{\infty} v_i^2 < \infty\}.$$

And for any $\eta > 1/2$

$$\sum_{i=1}^{n} v_i\epsilon_i = o\left((\sum_{i=1}^{n} v_i^2)^{1/2}\log^{\eta}(\sum_{i=1}^{n} v_i^2)\right) \textit{ a.s. on } \{\sum_{i=1}^{\infty} v_i^2 = \infty\}.$$

As a result, with probability 1

$$\begin{aligned}\sum_{i=1}^{n} v_i\epsilon_i =&o\left((\sum_{i=1}^{n} v_i^2)^{1/2}\log(\sum_{i=1}^{n} v_i^2)\right)1_{\sum_{i=1}^{\infty} v_i^2=\infty} + \mathcal{O}(1)1_{\sum_{i=1}^{\infty} v_i^2<\infty} \text{ a.s.}\\ =&o\left((\sum_{i=1}^{n} v_i^2)^{1/2}\log(\sum_{i=1}^{n} v_i^2)\right) + \mathcal{O}(1) \text{ a.s.}\end{aligned} \tag{82}$$

We can apply Lemma 46 to our context by noticing

$$\sum_{i=0}^{t-1}\eta_i x_i^\top = \sum_{i=0}^{t-1}\eta_i i^{\frac{1-\beta}{2}}\log^{-\alpha/2}(i)(i^{\frac{\beta-1}{2}}\log^{\alpha/2}(i)x_i^\top).$$

Here we normalized all η_i to have a fixed normal distribution $\eta_i i^{\frac{1-\beta}{2}}\log^{-\alpha/2}(i)\sim\mathcal{N}(0,\tau^2 I_d)$. Apply Eq. 82 entry-wise, where v_i corresponds to a fixed entry of $i^{\frac{\beta-1}{2}}\log^{\alpha/2}(i)x_i^\top$ and ϵ_i corresponds to a fixed entry of $\eta_i i^{\frac{1-\beta}{2}}\log^{-\alpha/2}(i)$. v_i is bounded by $i^{\frac{\beta-1}{2}}\log^{\alpha/2}(i)\|x_i\|$. Thus

$$\sum_{i=0}^{t-1}\eta_i x_i^\top = o\left(V_t^{1/2}\log(V_t)\right) + \mathcal{O}(1) \text{ a.s.},$$

where $V_t := \sum_{i=0}^{t-1}(i^{\frac{\beta-1}{2}}\log^{\alpha/2}(i)\|x_i\|)^2$. Applying the bounds in Lemma 18 (**recall that when** $i=0,1$**,** $i^{\beta-1}\log^{\alpha}(i)$ **is taken to be** 1):

$$\begin{aligned}V_t =&\sum_{i=0}^{t-1}(i^{\frac{\beta-1}{2}}\log^{\alpha/2}(i)\|x_i\|)^2\\ =&\sum_{i=0}^{t-1} i^{-1+\beta}\log^{\alpha}(i)\mathcal{O}(\log(t)) \text{ a.s.}\quad \text{(by Lemma 18)}\\ =&\mathcal{O}(t^\beta\log^\alpha(t))\mathcal{O}(\log(t)) \text{ a.s.}\quad \text{(by Eq. 81)}\\ =&\mathcal{O}(t^\beta\log^{\alpha+1}(t)) \text{ a.s.}\end{aligned}$$

Thus,

$$\begin{aligned}\sum_{i=0}^{t-1}\eta_i x_i^\top =&o\left(V_t^{1/2}\log(V_t)\right) + \mathcal{O}(1)\\ =&o\left(\mathcal{O}(t^\beta\log^{\alpha+1}(t))^{1/2}\log(\mathcal{O}(t^\beta\log^{\alpha+1}(t)))\right) + \mathcal{O}(1)\\ =&o\left(\mathcal{O}(t^{\beta/2}\log^{\frac{\alpha+1}{2}}(t)\log(t))\right) + \mathcal{O}(1)\\ =&o\left(t^{\beta/2}\log^{\frac{\alpha+3}{2}}(t)\right) \text{ a.s.}\end{aligned}$$

■

In exactly the same way, we can show that

$$\sum_{i=1}^{t}(\hat{K}_i x_i)^\top R\eta_i = o\left(t^{\beta/2}\log^{\frac{\alpha+3}{2}}(t)\right) \text{ a.s.} \tag{83}$$

We first standardize η_i

$$\sum_{i=1}^{t}(\hat{K}_i x_i)^\top R\eta_i = \sum_{i=0}^{t-1}(i^{\frac{\beta-1}{2}}\log^{\alpha/2}(i)(\hat{K}_i x_i)^\top R)\eta_i i^{\frac{1-\beta}{2}}\log^{-\alpha/2}(i),$$

and then v_i is bounded by

$$\begin{aligned}
&i^{\frac{\beta-1}{2}}\log^{\alpha/2}(i)\left\|(\hat{K}_i x_i)^\top R\right\| \\
&\quad\le i^{\frac{\beta-1}{2}}\log^{\alpha/2}(i)\left\|\hat{K}_i\right\|\|R\|\|x_i\| \\
&\quad\le i^{\frac{\beta-1}{2}}\log^{\alpha/2}(i)C_K\|R\|\|x_i\| \quad \text{(by Algorithm 1's design)},
\end{aligned}$$

which is different from v_i in $\sum_{i=0}^{t-1}\eta_i x_i^\top$ by a constant factor $C_K\|R\|$. The rest of the proof is all the same.

H.1.7 The proof of Lemma 24

Lemma. *Assume Eq. 23, then*

1. $\sum_{i=0}^{t-1}\delta_i x_i x_i^\top\delta_i^\top = \mathcal{O}(t^{1-\beta}\log^{-\alpha+2}(t))$ *a.s.*
2. $\sum_{i=0}^{t-1}\delta_i x_i\eta_i^\top = (\sum_{i=0}^{t-1}\eta_i x_i^\top\delta_i^\top)^\top = o\left(\log^2(t)\right)$ *a.s.*
3. $\sum_{i=0}^{t-1}\eta_i\eta_i^\top = t^\beta\frac{\tau^2}{\beta}\log^\alpha(t)(I_d + o_p(1))$

Proof

First part $\sum_{i=0}^{t-1}\delta_i x_i x_i^\top\delta_i^\top$ Recall the conclusion from Lemma 18: $\|x_t\| = \mathcal{O}(\log^{1/2}(t))$ a.s. and $\|\delta_t\| = \mathcal{O}(t^{-\frac{\beta}{2}}\log^{\frac{-\alpha+1}{2}}(t))$ a.s.

$$\begin{aligned}
\left\|\sum_{i=1}^{t-1}\delta_i x_i x_i^\top\delta_i^\top\right\| \le &\sum_{i=1}^{t-1}\|\delta_i\|^2\|x_i\|^2 \\
\le&\mathcal{O}(\log(t))\sum_{i=1}^{t-1}\mathcal{O}(i^{-\beta}\log^{-\alpha+1}(i)) \text{ a.s.} \qquad \text{(by Lemma 18)} \\
=&\mathcal{O}(t^{1-\beta}\log^{-\alpha+2}(t)) \text{ a.s.} \qquad \text{(by Eq. 81)}
\end{aligned}$$

This implies (by bounding the entries by the operator norm, and including the $i = 0$ term as $\mathcal{O}(1)$):

$$\sum_{i=0}^{t-1}\delta_i x_i x_i^\top\delta_i^\top = \mathcal{O}(t^{1-\beta}\log^{-\alpha+2}(t)) \text{ a.s.}$$

Second part $\sum_{i=0}^{t-1} \eta_i x_i^\top \delta_i^\top$ The representative of the third term is $\sum_{i=0}^{t-1} \eta_i x_i^\top \delta_i^\top$. The proof idea is similar to that in Lemma 23 when we prove the bound for $\sum_{i=0}^{t-1} \eta_i x_i^\top$. Here we have an extra shrinking term δ_i which makes things easier.

Again, we can apply Lemma 46 to our context by noticing

$$\sum_{i=0}^{t-1} \eta_i x_i^\top \delta_i^\top = \sum_{i=0}^{t-1} \eta_i i^{\frac{1-\beta}{2}} \log^{-\alpha/2}(i)(i^{\frac{\beta-1}{2}} \log^{\alpha/2}(i) x_i^\top \delta_i^\top).$$

Here we normalized all η_i to have a fixed normal distribution. Apply Lemma 46 entry-wise, where v_i corresponds to a fixed entry of $i^{\frac{\beta-1}{2}} \log^{\alpha/2}(i) x_i^\top \delta_i^\top$ and ϵ_i corresponds to a fixed entry of the normalized η_i. Our v_i is bounded by $i^{\frac{\beta-1}{2}} \log^{\alpha/2}(i)\|x_i\|\|\delta_i\|$. Thus,

$$\sum_{i=0}^{t-1} \eta_i x_i^\top \delta_i^\top = o\left(V_t^{1/2} \log(V_t)\right) + \mathcal{O}(1).$$

where $V_t := \sum_{i=0}^{t-1} (i^{\frac{\beta-1}{2}} \log^{\alpha/2}(i)\|x_i\|\|\delta_i\|)^2$. Apply the high probability bound in Lemma 18 and we have

$$\begin{aligned}
V_t =& \sum_{i=1}^{t-1} (i^{\frac{\beta-1}{2}} \log^{\alpha/2}(i)\|x_i\|\|\delta_i\|)^2 \\
=& \sum_{i=1}^{t-1} i^{-1+\beta} \log^{\alpha}(i) \mathcal{O}(\log(t)) \mathcal{O}(t^{-\beta} \log^{-\alpha+1}(t)) \text{ a.s.} \qquad \text{(by Lemma 18)} \\
=& \mathcal{O}(t^{\beta} \log^{\alpha}(t)) \mathcal{O}(\log(t)) \mathcal{O}(t^{-\beta} \log^{-\alpha+1}(t)) \text{ a.s.} \qquad \text{(by Eq. 81)} \\
=& \mathcal{O}(\log^2(t)) \text{ a.s.}
\end{aligned}$$

That is to say, $V_t = \mathcal{O}(\log^2(t))$ a.s. (adding the $i = 0$ term as $\mathcal{O}(1)$). Thus,

$$\begin{aligned}
\sum_{i=0}^{t-1} \eta_i x_i^\top \delta_i^\top =& o\left(V_t^{1/2} \log(V_t)\right) + \mathcal{O}(1) \text{ a.s.} \\
=& o\left(\mathcal{O}(\log^2(t))^{1/2} \log(\mathcal{O}(\log^2(t)))\right) + \mathcal{O}(1) \text{ a.s.} \\
=& o\left(o(\log^2(t))\right) + \mathcal{O}(1) \text{ a.s.} \\
=& o\left(\log^2(t)\right) \text{ a.s.}
\end{aligned}$$

Third part $\sum_{i=0}^{t-1} \eta_i \eta_i^\top$ By Eq. 81:

$$\mathbb{E}(\sum_{i=0}^{t-1} \eta_i \eta_i^\top) = \sum_{i=0}^{t-1} \tau^2 i^{\beta-1} \log^{\alpha}(i) I_d = t^{\beta} \frac{\tau^2}{\beta} \log^{\alpha}(t)(I_d + o(1)).$$

With a little abuse of notation we use $\mathrm{Var}(\cdot)$ as entry-wise variance of a matrix. Again, $i = 0, 1$ terms are meant to be $\mathcal{O}(1)$.

$$\begin{aligned}\mathrm{Var}(\sum_{i=0}^{t-1} \eta_i \eta_i^\top) =& \sum_{i=0}^{t-1} \mathrm{Var}(\eta_i \eta_i^\top) \\ =&\mathcal{O}\left(\sum_{i=0}^{t-1} i^{2(\beta-1)} \log^{2\alpha}(i)\right) \\ \leq&\mathcal{O}\left(\sum_{i=0}^{t-1} i^{2(\beta-1)} \log^{2\max\{0,\alpha\}}(i)\right) \\ \leq&\mathcal{O}\left(\sum_{i=0}^{t-1} i^{2(\beta-1)} \log^{2\max\{0,\alpha\}}(t)\right) \\ =&\tilde{\mathcal{O}}\left(\sum_{i=0}^{t-1} i^{2(\beta-1)}\right) \\ =&\tilde{\mathcal{O}}(t^{2\beta-1}).\end{aligned}$$

When $\beta > 1/2$ the last equation follows by Eq. 81 and when $\beta = 1/2$ it is summation of harmonic series which is $\tilde{\mathcal{O}}(1)$. Thus the standard error is only of order $\tilde{\mathcal{O}}(t^{\beta-1/2})$, which is smaller than $\mathbb{E}(\sum_{i=0}^{t-1} \eta_i \eta_i^\top)$. That is to say,

$$\sum_{i=0}^{t-1} \eta_i \eta_i^\top = t^\beta \frac{\tau^2}{\beta} \log^\alpha(t)(I_d + o_p(1)).$$

■

H.2 Lemmas in Appendix C

H.2.1 The proof of Lemma 25

Lemma. *For any $\hat{K}$ with suitable dimension,*

$$\begin{aligned}&x^\top(Q + \hat{K}^\top R\hat{K})x + x^\top(A + B\hat{K})^\top P(A + B\hat{K})x - x^\top Px \\ &= x^\top(\hat{K} - K)^\top(R + B^\top PB)(\hat{K} - K)x.\end{aligned}$$

Recall P is the middle step described by the DARE. It should satisfy Eq. 3

$$K = -(R + B^\top PB)^{-1}B^\top PA.$$

As a result,

$$(R + B^\top PB)K + B^\top PA = 0. \tag{84}$$

Also it is well known that (Jamieson et al., 2018):

$$Q + K^\top RK + (A + BK)^\top P(A + BK) = P. \tag{85}$$

Let $\hat{K}$ be another controller, then we have the following useful equation stated by Lemma 25.

$$\begin{aligned}
&x^\top(Q+\hat{K}^\top R\hat{K})x + x^\top(A+B\hat{K})^\top P(A+B\hat{K})x - x^\top Px\\
&\quad = x^\top(Q+(\hat{K}-K+K)^\top R(\hat{K}-K+K))x\\
&\qquad + x^\top(A+B(\hat{K}-K)+BK)^\top P(A+B(\hat{K}-K)+BK)x\\
&\qquad - x^\top Px\\
&\quad = x^\top(Q+K^\top RK+(A+BK)^\top P(A+BK))x\\
&\qquad + 2x^\top(\hat{K}-K)^\top(RK+B^\top P(A+BK))x\\
&\qquad + x^\top(\hat{K}-K)^\top(R+B^\top PB)(\hat{K}-K)x\\
&\qquad - x^\top Px\\
&\quad = x^\top(Q+K^\top RK+(A+BK)^\top P(A+BK))x - x^\top Px\\
&\qquad + 2x^\top(\hat{K}-K)^\top((R+B^\top PB)K+B^\top PA)x\\
&\qquad + x^\top(\hat{K}-K)^\top(R+B^\top PB)(\hat{K}-K)x\\
&\quad = x^\top(\hat{K}-K)^\top(R+B^\top PB)(\hat{K}-K)x \qquad \text{(by Eqs. 84 and 85).}
\end{aligned}$$

H.3 Lemmas in Appendix E

H.3.1 The proof of Lemma 27

Lemma.

$$x_t = \tilde{x}_t + O(t^{-\frac{\beta}{2}}\log^{\frac{-\alpha+2}{2}}(t)) \; a.s.$$
$$u_t = \tilde{u}_t + O(t^{-\frac{\beta}{2}}\log^{\frac{-\alpha+2}{2}}(t)) \; a.s.$$

where

$$\tilde{x}_t := \sum_{p=t-\left\lfloor -\frac{\log(t)}{\log(\rho_L)}\right\rfloor}^{t-1} (A+BK)^{t-p-1}(B\eta_p+\varepsilon_p), \tag{86}$$

and

$$\tilde{u}_t := K\tilde{x}_t + \xi_t = K \sum_{p=t-\left\lfloor -\frac{\log(t)}{\log(\rho_L)}\right\rfloor}^{t-1} (A+BK)^{t-p-1}(B\eta_p+\varepsilon_p) + \xi_t.$$

Proof Recall Lemma 19 states that

$$x_t = \sum_{p=0}^{t-1}(A+B\hat{K}_{t-1})\cdots(A+B\hat{K}_{p+1})(B\eta_p+\varepsilon_p) + (A+B\hat{K}_{t-1})\cdots(A+BK_0)x_0.$$

Similarly, we can rewrite x_t as if starting from time $t-\left\lfloor -\frac{\log(t)}{\log(\rho_L)}\right\rfloor$:

$$\begin{aligned}x_t = &\sum_{p=t-\left\lfloor -\frac{\log(t)}{\log(\rho_L)}\right\rfloor}^{t-1} (A+B\hat{K}_{t-1})\cdots(A+B\hat{K}_{p+1})(B\eta_p+\varepsilon_p)+\\
&(A+B\hat{K}_{t-1})\cdots(A+B\hat{K}_{t-\left\lfloor -\frac{\log(t)}{\log(\rho_L)}\right\rfloor})x_{t-\left\lfloor -\frac{\log(t)}{\log(\rho_L)}\right\rfloor}.\end{aligned} \tag{87}$$

By Lemma 18, we know

$$\begin{aligned}(A+B\hat{K}_{t-1})\cdots(A+B\hat{K}_{t-\left\lfloor -\frac{\log(t)}{\log(\rho_L)}\right\rfloor}) \leq &\mathcal{O}(\rho_L^{-\log(t)/\log(\rho_L)}) \text{ a.s.}\\ =&\mathcal{O}(e^{-\log(t)}) \text{ a.s.}\\ =&\mathcal{O}(t^{-1}) \text{ a.s.}\end{aligned}$$

and

$$\|x_t\|, \|u_t\| \leq \mathcal{O}(\log^{1/2}(t)) \text{ a.s.}$$

Thus

$$(A+B\hat{K}_{t-1})\cdots(A+B\hat{K}_{t-\left\lfloor -\frac{\log(t)}{\log(\rho_L)}\right\rfloor})x_{t-\left\lfloor -\frac{\log(t)}{\log(\rho_L)}\right\rfloor} = \mathcal{O}(t^{-1}\log^{1/2}(t)) \text{ a.s.}$$

Next, comparing Eq. 86 with Eq. 87, we still need to bound the difference between $(A+B\hat{K}_{t-1})\cdots(A+B\hat{K}_{p+1})$ and $(A+BK)^{t-p-1}$. Again by Lemma 18,

$$\begin{aligned}&\left\|\sum_{p=t-\left\lfloor -\frac{\log(t)}{\log(\rho_L)}\right\rfloor}^{t-1}\left[(A+B\hat{K}_{t-1})\cdots(A+B\hat{K}_{p+1})-(A+BK)^{t-p-1}\right](B\eta_p+\varepsilon_p)\right\|\\
&\leq \sum_{p=t-\left\lfloor -\frac{\log(t)}{\log(\rho_L)}\right\rfloor}^{t-1}\mathcal{O}(\rho_L^{t-p})(\|\delta_{t-1}\|+\cdots+\|\delta_{p+1}\|)\mathcal{O}(\log^{1/2}(t)) \text{ a.s.} \qquad \text{(by Eqs. 75 and 24)}\\
&= \sum_{p=t-\left\lfloor -\frac{\log(t)}{\log(\rho_L)}\right\rfloor}^{t-1}\|\delta_{p+1}\|\mathcal{O}(\rho_L^{t-p})\mathcal{O}(\log^{1/2}(t)) \text{ a.s.} \qquad \text{(by Eq. 77)}\\
&\leq \mathcal{O}((t/2)^{-\frac{\beta}{2}}\log^{\frac{-\alpha+1}{2}}(t/2))\sum_{p=t-\left\lfloor -\frac{\log(t)}{\log(\rho_L)}\right\rfloor}^{t-1}\mathcal{O}(\rho_L^{t-p})\mathcal{O}(\log^{1/2}(t)) \text{ a.s.}\\
&\qquad \text{(by Eq. 26 and that asymptotically } p > t/2\text{)}\\
&= \mathcal{O}(t^{-\frac{\beta}{2}}\log^{\frac{-\alpha+1}{2}}(t))\mathcal{O}(\log^{1/2}(t)) \text{ a.s.}\\
&= \mathcal{O}(t^{-\frac{\beta}{2}}\log^{\frac{-\alpha+2}{2}}(t)) \text{ a.s.}\end{aligned}$$

This is larger than $\mathcal{O}(t^{-1}\log^{1/2}(t))$. To summarize,

$$x_t = \tilde{x}_t + \mathcal{O}(t^{-\frac{\beta}{2}}\log^{\frac{-\alpha+2}{2}}(t)) \text{ a.s.}$$

Since $u_t - \tilde{u}_t = K(x_t - \tilde{x}_t)$,

$$u_t = \tilde{u}_t + O(t^{-\frac{\beta}{2}}\log^{\frac{-\alpha+2}{2}}(t)) \text{ a.s.}$$

$\blacksquare$

H.3.2 The proof of Lemma 28

Lemma.

$$\hat{A}_t = \hat{A}_{t-\left\lfloor -\frac{\log(t)}{\log(\rho_L)} \right\rfloor} + \mathcal{O}_p(t^{-\beta} \log^{-\alpha+3/2}(t)).$$

$$\hat{B}_t = \hat{B}_{t-\left\lfloor -\frac{\log(t)}{\log(\rho_L)} \right\rfloor} + \mathcal{O}_p(t^{-\beta} \log^{-\alpha+3/2}(t)).$$

Proof We can bound the distance of neighboring estimators by the following recursive LS formula. Denote $\hat{\Theta}_t := [\widehat{A}_t, \widehat{B}_t]$, $z_i := \begin{bmatrix} x_i \\ u_i \end{bmatrix}$, $H_t := (\sum_{i=0}^{t-1} z_i z_i^\top)^{-1}$. Then the LS estimator Eq. 5 is

$$\hat{\Theta}_t = \sum_{i=0}^{t-1} z_{i+1} z_i^\top (\sum_{i=0}^{t-1} z_i z_i^\top)^{-1} = \sum_{i=0}^{t-1} z_{i+1} z_i^\top H_t.$$

For simplicity, denote $a_t := \left\lfloor -\frac{\log(t)}{\log(\rho_L)} \right\rfloor$, then our objective is to bound the difference $\hat{\Theta}_t - \hat{\Theta}_{t-a_t}$.

$$\hat{\Theta}_{t-a_t} = \sum_{i=0}^{t-a_t-1} z_{i+1} z_i^\top H_{t-a_t}.$$

As a result,

$$\hat{\Theta}_t = (\hat{\Theta}_{t-a_t} H_{t-a_t}^{-1} + \sum_{i=t-a_t}^{t-1} z_{i+1} z_i^\top) H_t.$$

And

$$\begin{aligned} \hat{\Theta}_t - \hat{\Theta}_{t-a_t} &= \left(\hat{\Theta}_{t-a_t} (H_{t-a_t}^{-1} - H_t^{-1}) + \sum_{i=t-a_t}^{t-1} z_{i+1} z_i^\top \right) H_t \\ &= \left(-\hat{\Theta}_{t-a_t} \left(\sum_{i=t-a_t}^{t-1} z_i z_i^\top \right) + \sum_{i=t-a_t}^{t-1} z_{i+1} z_i^\top \right) H_t \\ &= \left(-\hat{\Theta}_{t-a_t} \left(\sum_{i=t-a_t}^{t-1} z_i z_i^\top \right) + \sum_{i=t-a_t}^{t-1} (\Theta z_i + \varepsilon_i) z_i^\top \right) H_t \\ &= (\Theta - \hat{\Theta}_{t-a_t}) \left(\sum_{i=t-a_t}^{t-1} z_i z_i^\top \right) H_t + \sum_{i=t-a_t}^{t-1} \varepsilon_i z_i^\top H_t. \end{aligned} \tag{88}$$

Following Eqs. 8 and 48,

$$H_t = \mathcal{O}_p(t^{-\beta} \log^{-\alpha}(t)). \tag{89}$$

Next will bound the first and second term separately.

First term $(\Theta - \hat{\Theta}_{t-a_t})(\sum_{i=t-a_t}^{t-1} z_i z_i^\top)H_t$ By Lemma 18,

$$z_t = \mathcal{O}(\log^{1/2}(t)) \text{ a.s.}$$

Recall that from Eqs. 10 and 48, $\Theta - \hat{\Theta}_{t-a_t} = \mathcal{O}_p(t^{-\beta/2}\log^{-\alpha/2}(t))$. As a result,

$$\begin{aligned}(\Theta - \hat{\Theta}_{t-a_t})(\sum_{i=t-a_t}^{t-1} z_i z_i^\top)H_t =& \mathcal{O}_p\left(t^{-\beta/2}\log^{-\alpha/2}(t)\right)\mathcal{O}_p(a_t\log(t)t^{-\beta}\log^{-\alpha}(t)) \\ =& \mathcal{O}_p(t^{-3\beta/2}\log^{-3\alpha/2+2}(t)).\end{aligned} \tag{90}$$

We will see that this order is smaller than the second term, so that the second term is dominating.

Second term $\sum_{i=t-a_t}^{t-1} \varepsilon_i z_i^\top H_t$ Consider the variance of the jk-th element of $\sum_{i=t-a_t}^{t-1} \varepsilon_i z_i^\top$, which is applicable to any choice of j and k. Fix j, k. Define $\mathcal{F}_{t-1}$ as the filtration which contains every variable except for $\varepsilon_{t-1,j}$. We know that $\varepsilon_{t-1,j} \perp\!\!\!\perp \mathcal{F}_{t-1}$ and $\varepsilon_{t-1,j} \sim \mathcal{N}(0, \sigma^2)$.

$$\begin{aligned}&\operatorname{Var}\left(\sum_{i=t-a_t}^{t-1} \varepsilon_{ij}(z_i)_k\right) \\ &= \operatorname{Var}\left(\mathbb{E}\left(\sum_{i=t-a_t}^{t-1} \varepsilon_{ij}(z_i)_k \middle| \mathcal{F}_{t-1}\right)\right) + \mathbb{E}\left(\operatorname{Var}\left(\sum_{i=t-a_t}^{t-1} \varepsilon_{ij}(z_i)_k \middle| \mathcal{F}_{t-1}\right)\right) \\ &= \operatorname{Var}\left(\sum_{i=t-a_t}^{t-2} \varepsilon_{ij}(z_i)_k\right) + \mathbb{E}\left((z_{t-1})_k^2\sigma^2\right) \\ &= \sigma^2 \sum_{i=t-a_t}^{t-1} \mathbb{E}\left((z_i)_k^2\right) \qquad \text{(by recursively conditioning on } \mathcal{F}_{t-2}, \cdots, \mathcal{F}_{t-a_t}) \\ &\leq \sigma^2 \sum_{i=t-a_t}^{t-1} \mathbb{E}\|z_i\|^2 \\ &\leq \sigma^2 a_t \mathcal{O}(\log^2(t)) \qquad \text{(by Eq. 104)} \\ &\leq \sigma^2 \mathcal{O}(\log^3(t)) \qquad \left(\text{by } a_t := \left\lfloor -\frac{\log(t)}{\log(\rho_L)} \right\rfloor\right)\end{aligned}$$

Since $\mathbb{E}\left(\sum_{i=t-a_t}^{t-1} \varepsilon_{ij}(z_i^\top)_k\right) = 0$, we have $\sum_{i=t-a_t}^{t-1} \varepsilon_{ij}(z_i^\top)_k = \mathcal{O}_p(\log^{3/2}(t))$, which implies

$$\sum_{i=t-a_t}^{t-1} \varepsilon_i z_i^\top = \mathcal{O}_p(\log^{3/2}(t)).$$

By Eq. 89,

$$\sum_{i=t-a_t}^{t-1} \varepsilon_i z_i^\top H_t = \mathcal{O}_p(\log^{3/2}(t))\mathcal{O}_p(t^{-\beta}\log^{-\alpha}(t)) = \mathcal{O}_p(t^{-\beta}\log^{-\alpha+3/2}(t)). \tag{91}$$

This is larger than the first term. Combining Eqs. 88, 90, and 91 we have

$$\hat{\Theta}_t - \hat{\Theta}_{t-a_t} = \mathcal{O}_p(t^{-\beta}\log^{-\alpha+3/2}(t)).$$

■

H.3.3 The proof of Lemma 29

Lemma. *For any ξ_t independent of the data before t: $\{\varepsilon_i, \eta_i\}_{i=0}^{t-1}$:*

$$\left(\tilde{x}_t^\top \left(\sum_{p=0}^{\infty} L^p \left(I_n + 1_{\{\beta=1,\alpha=0\}}\frac{\tau^2}{\sigma^2}BB^\top\right)(L^p)^\top\right)^{-1} \tilde{x}_t + \frac{\beta\sigma^2}{\tau^2}t^{1-\beta}\log^{-\alpha}(t)\left\|\xi_t\right\|^2\right)^{-1/2}$$
$$\cdot\, t^{1/2}\left((\hat{A}_{t-\left\lfloor -\frac{\log(t)}{\log(\rho_L)}\right\rfloor} - A)\tilde{x}_t + (\hat{B}_{t-\left\lfloor -\frac{\log(t)}{\log(\rho_L)}\right\rfloor} - B)(K\tilde{x}_t + \xi_t)\right) \xrightarrow{D} \mathcal{N}(0, I_n).$$

Proof We will start from finding the conditional distribution of

$$(\hat{A}_{t-\left\lfloor -\frac{\log(t)}{\log(\rho_L)}\right\rfloor} - A)\tilde{x}_t + (\hat{B}_{t-\left\lfloor -\frac{\log(t)}{\log(\rho_L)}\right\rfloor} - B)\tilde{u}_t\Big|\tilde{x}_t = x, \tilde{u}_t = Kx + \xi.$$

where x and ξ are constants. This should be easy because $\hat{A}_{t-\left\lfloor -\frac{\log(t)}{\log(\rho_L)}\right\rfloor} - A, \hat{B}_{t-\left\lfloor -\frac{\log(t)}{\log(\rho_L)}\right\rfloor} - B \perp\!\!\!\perp \tilde{x}_t, \tilde{u}_t$, which means we can directly apply the asymptotic normality result from Theorem 5. Recall Eq. 11 that

$$t^{\beta/2}\log^{\alpha/2}(t)\mathrm{vec}\left(\begin{bmatrix}\hat{A}_t - A + (\hat{B}_t - B)K, & \hat{B}_t - B\end{bmatrix}\begin{bmatrix} C_t^{1/2} & 0 \\ 0 & \sqrt{\frac{\tau^2}{\beta}}I_d\end{bmatrix}\right) \xrightarrow{D} \mathcal{N}(0, \sigma^2 I_{n(n+d)}),$$

where $C_t = t^{1-\beta}\log^{-\alpha}(t)\sum_{p=0}^{\infty} L^p\left(\sigma^2 I_n + 1_{\{\beta=1,\alpha=0\}}\tau^2 BB^\top\right)(L^p)^\top(I_n + o_p(1))$ (by Eq. 32). Here there are two different convergence speeds and we need to consider them separately. More precisely,

$$\mathrm{vec}\left(\begin{bmatrix} (\hat{A}_t - A + (\hat{B}_t - B)K)t^{\beta/2}\log^{\alpha/2}(t)C_t^{1/2}\sigma^{-1} & (\hat{B}_t - B)t^{\beta/2}\log^{\alpha/2}(t)\sqrt{\frac{\tau^2}{\sigma^2\beta}}I_d\end{bmatrix}\right)$$
$$\xrightarrow{D} \mathcal{N}(0, I_{n+d}\otimes I_n).$$

That is to say, for any constant vector x and ξ_t independent of data before t, we have

$$\mathrm{vec}\Bigg(\begin{bmatrix} (\hat{A}_t - A + (\hat{B}_t - B)K)t^{\beta/2}\log^{\alpha/2}(t)C_t^{1/2}\sigma^{-1} & (\hat{B}_t - B)t^{\beta/2}\log^{\alpha/2}(t)\sqrt{\frac{\tau^2}{\sigma^2\beta}}I_d\end{bmatrix}$$
$$\cdot\begin{bmatrix} t^{-\beta/2}\log^{-\alpha/2}(t)C_t^{-1/2}\sigma x \\ t^{(1-\beta)/2}\log^{-\alpha/2}(t)\sqrt{\frac{\sigma^2\beta}{\tau^2}}\xi_t\end{bmatrix}\Bigg/\left\|\begin{bmatrix} t^{-\beta/2}\log^{-\alpha/2}(t)C_t^{-1/2}\sigma x \\ t^{(1-\beta)/2}\log^{-\alpha/2}(t)\sqrt{\frac{\sigma^2\beta}{\tau^2}}\xi_t\end{bmatrix}\right\|\Bigg) \xrightarrow{D} \mathcal{N}(0, I_n).$$

The above equation holds because we are multiplying independent unit vector to the left hand side, so the result is still a normal distribution. Simplifying the equation:

$$\left(x^\top\left(\sum_{p=0}^{\infty}L^p\left(I_n+1_{\{\beta=1,\alpha=0\}}\frac{\tau^2}{\sigma^2}BB^\top\right)(L^p)^\top\right)^{-1}x+\frac{\beta\sigma^2}{\tau^2}t^{1-\beta}\log^{-\alpha}(t)\|\xi_t\|^2\right)^{-1/2}$$
$$\cdot t^{1/2}\left[(\hat{A}_t-A)x+(\hat{B}_t-B)(Kx+\xi_t)\right]\xrightarrow{D}\mathcal{N}(0,I_n).$$

We can replace t with $t-\left\lfloor-\frac{\log(t)}{\log(\rho_L)}\right\rfloor$:

$$\left(x^\top\left(\sum_{p=0}^{\infty}L^p\left(I_n+1_{\{\beta=1,\alpha=0\}}\frac{\tau^2}{\sigma^2}BB^\top\right)(L^p)^\top\right)^{-1}x+\frac{\beta\sigma^2}{\tau^2}\left(t-\left\lfloor-\frac{\log(t)}{\log(\rho_L)}\right\rfloor\right)^{1-\beta}\log^{-\alpha}\left(t-\left\lfloor-\frac{\log(t)}{\log(\rho_L)}\right\rfloor\right)$$
$$\cdot\left(t-\left\lfloor-\frac{\log(t)}{\log(\rho_L)}\right\rfloor\right)^{1/2}\left[(\hat{A}_{t-\left\lfloor-\frac{\log(t)}{\log(\rho_L)}\right\rfloor}-A)x+(\hat{B}_{t-\left\lfloor-\frac{\log(t)}{\log(\rho_L)}\right\rfloor}-B)(Kx+\xi_t)\right]\xrightarrow{D}\mathcal{N}(0,I_n).$$

Because $\left(t-\left\lfloor-\frac{\log(t)}{\log(\rho_L)}\right\rfloor\right)^{1/2}t^{-1/2}\to 1$, we can drop the first three instances of $\left\lfloor-\frac{\log(t)}{\log(\rho_L)}\right\rfloor$:

$$\left(x^\top\left(\sum_{p=0}^{\infty}L^p\left(I_n+1_{\{\beta=1,\alpha=0\}}\frac{\tau^2}{\sigma^2}BB^\top\right)(L^p)^\top\right)^{-1}x+\frac{\beta\sigma^2}{\tau^2}t^{1-\beta}\log^{-\alpha}(t)\left\|\xi_t\right\|^2\right)^{-1/2}$$
$$\cdot t^{1/2}\left[(\hat{A}_{t-\left\lfloor-\frac{\log(t)}{\log(\rho_L)}\right\rfloor}-A)x+(\hat{B}_{t-\left\lfloor-\frac{\log(t)}{\log(\rho_L)}\right\rfloor}-B)(Kx+\xi_t)\right]\xrightarrow{D}\mathcal{N}(0,I_n).$$

Here we actually used the fact that for $c_t,a_t,b_t>0$, when $a_t/b_t\to 1$, then $(c_t+a_t)/(c_t+b_t)\to 1$. This is because

$$\left|\frac{c_t+a_t}{c_t+b_t}-\frac{a_t}{b_t}\right|=\left|\frac{(b_t-a_t)c_t}{(c_t+b_t)b_t}\right|\le\left|\frac{b_t-a_t}{b_t}\right|\to 0.$$

In our specific context c_t is the constant $x^\top\left(\sum_{p=0}^{\infty}L^p\left(I_n+1_{\{\beta=1,\alpha=0\}}\frac{\tau^2}{\sigma^2}BB^\top\right)(L^p)^\top\right)^{-1}x$.

Since $\tilde{x}_t\perp\!\!\!\perp\hat{A}_{t-\left\lfloor-\frac{\log(t)}{\log(\rho_L)}\right\rfloor}-A,\hat{B}_{t-\left\lfloor-\frac{\log(t)}{\log(\rho_L)}\right\rfloor}-B$, we can replace x with $\tilde{x}_t$ by conditioning on $\tilde{x}_t=x$, replace all x with $\tilde{x}_t$, and finally remove the conditioning since they all converge in distribution to standard normal and $\tilde{x}_t$ asymptotically have same distribution.

$$\left(\tilde{x}_t^\top\left(\sum_{p=0}^{\infty}L^p\left(I_n+1_{\{\beta=1,\alpha=0\}}\frac{\tau^2}{\sigma^2}BB^\top\right)(L^p)^\top\right)^{-1}\tilde{x}_t+\frac{\beta\sigma^2}{\tau^2}t^{1-\beta}\log^{-\alpha}(t)\left\|\xi_t\right\|^2\right)^{-1/2}$$
$$\cdot t^{1/2}\left((\hat{A}_{t-\left\lfloor-\frac{\log(t)}{\log(\rho_L)}\right\rfloor}-A)\tilde{x}_t+(\hat{B}_{t-\left\lfloor-\frac{\log(t)}{\log(\rho_L)}\right\rfloor}-B)(K\tilde{x}_t+\xi_t)\right)\xrightarrow{D}\mathcal{N}(0,I_n).\tag{92}$$

■

H.3.4 The proof of Lemma 30

Lemma. *For any ξ_t independent of the data before t: $\{\varepsilon_i, \eta_i\}_{i=0}^{t-1}$,*

$$\left(x_t^\top \left(\sum_{p=0}^{\infty} L^p \left(I_n + 1_{\{\beta=1,\alpha=0\}}\frac{\tau^2}{\sigma^2}BB^\top\right)(L^p)^\top\right)^{-1} x_t + \frac{\beta\sigma^2}{\tau^2}t^{1-\beta}\log^{-\alpha}(t)\,\|\xi_t\|^2\right)^{-1/2}$$
$$\cdot\, t^{1/2}\left[(\hat{A}_t - A)x_t + (\hat{B}_t - B)(\hat{K}_t x_t + \xi_t)\right] \xrightarrow{D} \mathcal{N}(0, I_n).$$

Proof Since we already proved Lemma 29, the only thing we need to do is to replace $\tilde{x}_t$ with x_t, K with $\hat{K}_t$, and $\hat{A}_{t-\left\lfloor -\frac{\log(t)}{\log(\rho_L)}\right\rfloor}$, $\hat{B}_{t-\left\lfloor -\frac{\log(t)}{\log(\rho_L)}\right\rfloor}$ with $\hat{A}_t$, $\hat{B}_t$.

Replacing $\tilde{x}_t$ with x_t First, we can replace

$$\left(\tilde{x}_t^\top \left(\sum_{p=0}^{\infty} L^p \left(I_n + 1_{\{\beta=1,\alpha=0\}}\frac{\tau^2}{\sigma^2}BB^\top\right)(L^p)^\top\right)^{-1} \tilde{x}_t + \frac{\beta\sigma^2}{\tau^2}t^{1-\beta}\log^{-\alpha}(t)\,\|\xi_t\|^2\right)$$

with $\left(x_t^\top \left(\sum_{p=0}^{\infty} L^p \left(I_n + 1_{\{\beta=1,\alpha=0\}}\frac{\tau^2}{\sigma^2}BB^\top\right)(L^p)^\top\right)^{-1} x_t + \frac{\beta\sigma^2}{\tau^2}t^{1-\beta}\log^{-\alpha}(t)\,\|\xi_t\|^2\right)$ in Eq. 92 because $\tilde{x}_t = x_t + o_p(1)$ by Lemma 27.

$$\left(x_t^\top \left(\sum_{p=0}^{\infty} L^p \left(I_n + 1_{\{\beta=1,\alpha=0\}}\frac{\tau^2}{\sigma^2}BB^\top\right)(L^p)^\top\right)^{-1} x_t + \frac{\beta\sigma^2}{\tau^2}t^{1-\beta}\log^{-\alpha}(t)\,\|\xi_t\|^2\right)^{-1/2}$$
$$\cdot\, t^{1/2}\left((\hat{A}_{t-\left\lfloor -\frac{\log(t)}{\log(\rho_L)}\right\rfloor} - A)\tilde{x}_t + (\hat{B}_{t-\left\lfloor -\frac{\log(t)}{\log(\rho_L)}\right\rfloor} - B)(K\tilde{x}_t + \xi_t)\right) \xrightarrow{D} \mathcal{N}(0, I_n).$$

Since $x_t^\top \left(\sum_{p=0}^{\infty} L^p \left(I_n + 1_{\{\beta=1,\alpha=0\}}\frac{\tau^2}{\sigma^2}BB^\top\right)(L^p)^\top\right)^{-1} x_t$ is bounded away from 0 with high probability (x_t has the component ε_{t-1}),

$$\left(x_t^\top \left(\sum_{p=0}^{\infty} L^p \left(I_n + 1_{\{\beta=1,\alpha=0\}}\frac{\tau^2}{\sigma^2}BB^\top\right)(L^p)^\top\right)^{-1} x_t + \frac{\beta\sigma^2}{\tau^2}t^{1-\beta}\log^{-\alpha}(t)\,\|\xi_t\|^2\right)^{-1/2} = \mathcal{O}_p(1).$$

By Lemma 27, $\tilde{x}_t = x_t + O_p(t^{-\frac{\beta}{2}}\log^{\frac{-\alpha+2}{2}}(t))$. Recall Proposition 17 states that $\|\hat{A}_{t-\left\lfloor -\frac{\log(t)}{\log(\rho_L)}\right\rfloor} - A\|, \|\hat{B}_{t-\left\lfloor -\frac{\log(t)}{\log(\rho_L)}\right\rfloor} - B\|, \|\hat{K}_t - K\| = \mathcal{O}_p(t^{-\beta/2}\log^{\frac{-\alpha+1}{2}}(t))$. Thus, the error induced by replacing the remaining $\tilde{x}_t$ with x_t in Eq. 92 is

$$\mathcal{O}_p(1)t^{1/2}\mathcal{O}_p(t^{-\frac{\beta}{2}}\log^{\frac{-\alpha+2}{2}}(t))\mathcal{O}_p(t^{-\frac{\beta}{2}}\log^{\frac{-\alpha+1}{2}}(t)) = \mathcal{O}_p(t^{1/2-\beta}\log^{-\alpha+3/2}(t)).$$

Under our condition $\beta > 1/2$ or $\beta = 1/2, \alpha > 3/2$, this error is of order $o_p(1)$, which is negligible. Now we can replace all $\tilde{x}_t$ with x_t:

$$\left(x_t^\top \left(\sum_{p=0}^{\infty} L^p \left(I_n + 1_{\{\beta=1,\alpha=0\}}\frac{\tau^2}{\sigma^2}BB^\top\right)(L^p)^\top\right)^{-1} x_t + \frac{\beta\sigma^2}{\tau^2}t^{1-\beta}\log^{-\alpha}(t)\,\|\xi_t\|^2\right)^{-1/2}$$

$$\cdot\, t^{1/2}\left[(\hat{A}_{t-\left\lfloor -\frac{\log(t)}{\log(\rho_L)}\right\rfloor} - A)x_t + (\hat{B}_{t-\left\lfloor -\frac{\log(t)}{\log(\rho_L)}\right\rfloor} - B)(Kx_t + \xi_t)\right] \xrightarrow{D} \mathcal{N}(0, I_n).$$

Replacing K by $\hat{K}_t$ Since $\|\hat{B}_{t-\left\lfloor -\frac{\log(t)}{\log(\rho_L)}\right\rfloor} - B\|, \|\hat{K}_t - K\| = \mathcal{O}_p(t^{-\beta/2}\log^{\frac{-\alpha+1}{2}}(t))$ (see Proposition 17), and $x_t = \mathcal{O}_p(\log^{1/2}(t))$, the final difference is still of order $\mathcal{O}_p(t^{1/2-\beta}\log^{-\alpha+3/2}(t)) = o_p(1)$. Thus

$$\left(x_t^\top \left(\sum_{p=0}^{\infty} L^p \left(I_n + 1_{\{\beta=1,\alpha=0\}}\frac{\tau^2}{\sigma^2}BB^\top\right)(L^p)^\top\right)^{-1} x_t + \frac{\beta\sigma^2}{\tau^2}t^{1-\beta}\log^{-\alpha}(t)\,\|\xi_t\|^2\right)^{-1/2}$$
$$\cdot\, t^{1/2}\left[(\hat{A}_{t-\left\lfloor -\frac{\log(t)}{\log(\rho_L)}\right\rfloor} - A)x_t + (\hat{B}_{t-\left\lfloor -\frac{\log(t)}{\log(\rho_L)}\right\rfloor} - B)(\hat{K}_t x_t + \xi_t)\right] \xrightarrow{D} \mathcal{N}(0, I_n).$$

Replacing $\hat{A}_{t-\left\lfloor -\frac{\log(t)}{\log(\rho_L)}\right\rfloor}$, $\hat{B}_{t-\left\lfloor -\frac{\log(t)}{\log(\rho_L)}\right\rfloor}$ with $\hat{A}_t$, $\hat{B}_t$ By Lemma 28,

$$\hat{A}_t - \hat{A}_{t-\left\lfloor -\frac{\log(t)}{\log(\rho_L)}\right\rfloor}, \hat{B}_t - \hat{B}_{t-\left\lfloor -\frac{\log(t)}{\log(\rho_L)}\right\rfloor} = \mathcal{O}_p(t^{-\beta}\log^{-\alpha+3/2}(t)).$$

Notice the x_t and ξ_t are multiplied by

$$\left(x_t^\top \left(\sum_{p=0}^{\infty} L^p \left(I_n + 1_{\{\beta=1,\alpha=0\}}\frac{\tau^2}{\sigma^2}BB^\top\right)(L^p)^\top\right)^{-1} x_t + \frac{\beta\sigma^2}{\tau^2}t^{1-\beta}\log^{-\alpha}(t)\,\|\xi_t\|^2\right)^{-1/2},$$

thus their order is only $\mathcal{O}_p(1)$. The difference induced by replacing $\hat{A}_{t-\left\lfloor -\frac{\log(t)}{\log(\rho_L)}\right\rfloor}, \hat{B}_{t-\left\lfloor -\frac{\log(t)}{\log(\rho_L)}\right\rfloor}$ with $\hat{A}_t, \hat{B}_t$ is of order $\mathcal{O}_p(t^{1/2-\beta}\log^{-\alpha+3/2}(t))$. When $\beta > 1/2$ or $\beta = 1/2, \alpha > 3/2$, this error is of order $o_p(1)$. Finally, after replacement we have

$$\left(x_t^\top \left(\sum_{p=0}^{\infty} L^p \left(I_n + 1_{\{\beta=1,\alpha=0\}}\frac{\tau^2}{\sigma^2}BB^\top\right)(L^p)^\top\right)^{-1} x_t + \frac{\beta\sigma^2}{\tau^2}t^{1-\beta}\log^{-\alpha}(t)\,\|\xi_t\|^2\right)^{-1/2}$$
$$\cdot\, t^{1/2}\left[(\hat{A}_t - A)x_t + (\hat{B}_t - B)(\hat{K}_t x_t + \xi_t)\right] \xrightarrow{D} \mathcal{N}(0, I_n).$$

■

H.3.5 The proof of Lemma 31

Lemma. *For any ξ_t independent of the data before t: $\{\varepsilon_i, \eta_i\}_{i=0}^{t-1}$,*

$$\left(x_t^\top \left(\sum_{p=0}^{\infty} L^p \left(I_n + 1_{\{\beta=1,\alpha=0\}}\frac{\tau^2}{\sigma^2}BB^\top\right)(L^p)^\top\right)^{-1} x_t + \frac{\beta\sigma^2}{\tau^2}t^{1-\beta}\log^{-\alpha}(t)\,\|\xi_t\|^2\right)^{-1/2}$$

$$\cdot t^{1/2}\left(\sigma^2 \begin{bmatrix} x_t \\ u_t \end{bmatrix}^\top \left(\sum_{i=0}^{t-1} \begin{bmatrix} x_i \\ u_i \end{bmatrix} \begin{bmatrix} x_i \\ u_i \end{bmatrix}^\top\right)^{-1} \begin{bmatrix} x_t \\ u_t \end{bmatrix}\right)^{1/2} \xrightarrow{P} 1.$$

Proof By $u_t = \hat{K}_t x_t + \xi_t$, it suffices to show

$$\begin{aligned}&\left(x_t^\top \left(\sum_{p=0}^{\infty} L^p \left(I_n + 1_{\{\beta=1,\alpha=0\}} \frac{\tau^2}{\sigma^2} BB^\top\right)(L^p)^\top\right)^{-1} x_t + \frac{\beta\sigma^2}{\tau^2} t^{1-\beta} \log^{-\alpha}(t) \left\|\xi_t\right\|^2\right)^{-1} \\ &\quad \cdot t^{1/2}\left(\sigma^2 \begin{bmatrix} x_t \\ \hat{K}_t x_t + \xi_t \end{bmatrix}^\top \left(\sum_{i=0}^{t-1} \begin{bmatrix} x_i \\ u_i \end{bmatrix} \begin{bmatrix} x_i \\ u_i \end{bmatrix}^\top\right)^{-1} \begin{bmatrix} x_t \\ \hat{K}_t x_t + \xi_t \end{bmatrix}\right) \xrightarrow{P} 1.\end{aligned}$$

By Eq. 61:

$$\sum_{i=0}^{t-1} \begin{bmatrix} x_i \\ u_i \end{bmatrix} \begin{bmatrix} x_i \\ u_i \end{bmatrix}^\top / t^\beta \log^\alpha(t) = \left[\begin{array}{cc} I_n & 0 \\ K & I_d \end{array}\right] \left[\begin{array}{cc} M_t & \Delta_t^\top \\ \Delta_t & \Delta_u \end{array}\right] \left[\begin{array}{cc} I_n & K^\top \\ 0 & I_d \end{array}\right].$$

Thus

$$\begin{aligned}&\begin{bmatrix} x_t \\ \hat{K}_t x_t + \xi_t \end{bmatrix}^\top \left(\sum_{i=0}^{t-1} \begin{bmatrix} x_i \\ u_i \end{bmatrix} \begin{bmatrix} x_i \\ u_i \end{bmatrix}^\top\right)^{-1} \begin{bmatrix} x_t \\ \hat{K}_t x_t + \xi_t \end{bmatrix} t^\beta \log^\alpha(t) \\ &\quad = \begin{bmatrix} x_t \\ \hat{K}_t x_t + \xi_t \end{bmatrix}^\top \left[\begin{array}{cc} I_n & K^\top \\ 0 & I_d \end{array}\right]^{-1} \left[\begin{array}{cc} M_t & \Delta_t^\top \\ \Delta_t & \Delta_u \end{array}\right]^{-1} \left[\begin{array}{cc} I_n & 0 \\ K & I_d \end{array}\right]^{-1} \begin{bmatrix} x_t \\ \hat{K}_t x_t + \xi_t \end{bmatrix} \\ &\quad = \begin{bmatrix} x_t \\ \hat{K}_t x_t + \xi_t \end{bmatrix}^\top \left[\begin{array}{cc} I_n & -K^\top \\ 0 & I_d \end{array}\right] \left[\begin{array}{cc} (M_t - \Delta_t^\top \Delta_u^{-1} \Delta_t)^{-1} & -(M_t - \Delta_t^\top \Delta_u^{-1} \Delta_t)^{-1} \Delta_t^\top \Delta_u^{-1} \\ -((M_t - \Delta_t^\top \Delta_u^{-1} \Delta_t)^{-1} \Delta_t^\top \Delta_u^{-1})^\top & (\Delta_u - \Delta_t M_t^{-1} \Delta_t^\top)^{-1} \end{array}\right] \\ &\qquad \cdot \left[\begin{array}{cc} I_n & 0 \\ -K & I_d \end{array}\right] \begin{bmatrix} x_t \\ \hat{K}_t x_t + \xi_t \end{bmatrix} \qquad \text{(by block matrix inversion)} \\ &\quad = \begin{bmatrix} x_t \\ \hat{K}_t x_t + \xi_t - K x_t \end{bmatrix}^\top \left[\begin{array}{cc} (M_t - \Delta_t^\top \Delta_u^{-1} \Delta_t)^{-1} & -(M_t - \Delta_t^\top \Delta_u^{-1} \Delta_t)^{-1} \Delta_t^\top \Delta_u^{-1} \\ -((M_t - \Delta_t^\top \Delta_u^{-1} \Delta_t)^{-1} \Delta_t^\top \Delta_u^{-1})^\top & (\Delta_u - \Delta_t M_t^{-1} \Delta_t^\top)^{-1} \end{array}\right] \\ &\qquad \cdot \begin{bmatrix} x_t \\ \hat{K}_t x_t + \xi_t - K x_t \end{bmatrix} \\ &\quad = x_t^\top (M_t - \Delta_t^\top \Delta_u^{-1} \Delta_t)^{-1} x_t - 2 x_t^\top (M_t - \Delta_t^\top \Delta_u^{-1} \Delta_t)^{-1} \Delta_t^\top \Delta_u^{-1} (\hat{K}_t x_t + \xi_t - K x_t) \\ &\qquad + (\hat{K}_t x_t + \xi_t - K x_t)^\top (\Delta_u - \Delta_t M_t^{-1} \Delta_t^\top)^{-1} (\hat{K}_t x_t + \xi_t - K x_t).\end{aligned}$$

By Eq. 31, Eq. 35, Eq. 60:

$$\begin{aligned}
M_t =& \log^{-\alpha}(t)t^{1-\beta}\left(\sum_{p=0}^{\infty} L^p\left(\sigma^2 I_n + 1_{\{\beta=1,\alpha=0\}}\tau^2 BB^\top\right)(L^p)^\top\right)(I_n+o(1))\\
M_t^{-1} =& \log^{\alpha}(t)t^{-1+\beta}\left(\sum_{p=0}^{\infty} L^p\left(\sigma^2 I_n + 1_{\{\beta=1,\alpha=0\}}\tau^2 BB^\top\right)(L^p)^\top\right)^{-1}(I_n+o(1))\\
\Delta_t =& \mathcal{O}_p(t^{1-3\beta/2}\log^{\frac{-3\alpha+3}{2}}(t))\\
\Delta_u =& \frac{\tau^2}{\beta}(I_d+o_p(1)).
\end{aligned} \tag{93}$$

As a result, when $\beta > 1/2$ or $\beta = 1/2, \alpha > 3/2$

$$\begin{aligned}
\Delta_t^\top \Delta_u^{-1}\Delta_t =& \mathcal{O}_p(t^{2-3\beta}\log^{-3\alpha+3}(t)) = o_p(t^{1-\beta}\log^{-\alpha}(t))\\
(M_t - \Delta_t^\top \Delta_u^{-1}\Delta_t)^{-1} =& M_t^{-1}(I_n - o_p(1))^{-1} = M_t^{-1}(I_n+o_p(1))\\
\Delta_t M_t^{-1}\Delta_t^\top =& \mathcal{O}_p(t^{1-3\beta/2}\log^{\frac{-3\alpha+3}{2}}(t))\mathcal{O}_p(t^{\beta-1}\log^{\alpha}(t))\mathcal{O}_p(t^{1-3\beta/2}\log^{\frac{-3\alpha+3}{2}}(t))\\
=& \mathcal{O}_p(t^{1-2\beta}\log^{-2\alpha+3}(t)) = o_p(1)\\
(\Delta_u - \Delta_t M_t^{-1}\Delta_t^\top)^{-1} =& \Delta_u^{-1}(I_d+o_p(1)).
\end{aligned}$$

Notice by Lemma 18, $\hat{K}_t - K = \mathcal{O}_p(t^{-\frac{\beta}{2}}\log^{\frac{-\alpha+1}{2}}(t))$. Then

$$\begin{aligned}
&\begin{bmatrix} x_t \\ \hat{K}_t x_t + \xi_t \end{bmatrix}^\top \left(\sum_{i=0}^{t-1}\begin{bmatrix} x_i \\ u_i\end{bmatrix}\begin{bmatrix} x_i \\ u_i\end{bmatrix}^\top\right)^{-1}\begin{bmatrix} x_t \\ \hat{K}_t x_t + \xi_t \end{bmatrix} t^\beta \log^\alpha(t)\\
&= x_t^\top (M_t - \Delta_t^\top \Delta_u^{-1}\Delta_t)^{-1}x_t + 2x_t^\top (M_t - \Delta_t^\top \Delta_u^{-1}\Delta_t)^{-1}\Delta_t^\top \Delta_u^{-1}(\hat{K}_t x_t + \xi_t - Kx_t)\\
&\quad + (\hat{K}_t x_t + \xi_t - Kx_t)^\top (\Delta_u - \Delta_t M_t^{-1}\Delta_t^\top)^{-1}(\hat{K}_t x_t + \xi_t - Kx_t)\\
&= x_t^\top M_t^{-1}(I_n+o_p(1))x_t + 2x_t^\top M_t^{-1}(I_n+o_p(1))\mathcal{O}_p(t^{1-3\beta/2}\log^{\frac{-3\alpha+3}{2}}(t))\Delta_u^{-1}(\mathcal{O}_p(t^{-\frac{\beta}{2}}\log^{\frac{-\alpha+1}{2}}(t))x_t + \xi_t)\\
&\quad + (\mathcal{O}_p(t^{-\frac{\beta}{2}}\log^{\frac{-\alpha+1}{2}}(t))x_t + \xi_t)^\top \Delta_u^{-1}(I_d+o_p(1))(\mathcal{O}_p(t^{-\frac{\beta}{2}}\log^{\frac{-\alpha+1}{2}}(t))x_t + \xi_t).
\end{aligned}$$

Quadratic terms of x_t Let us first consider all those quadratic terms of x_t:

- $x_t^\top M_t^{-1}(I_n+o_p(1))x_t$.
- $$\begin{aligned}
&2x_t^\top M_t^{-1}(I_n+o_p(1))\mathcal{O}_p(t^{1-3\beta/2}\log^{\frac{-3\alpha+3}{2}}(t))\Delta_u^{-1}\mathcal{O}_p(t^{-\frac{\beta}{2}}\log^{\frac{-\alpha+1}{2}}(t))x_t\\
&\quad = 2x_t^\top M_t^{-1}(I_n+o_p(1))\mathcal{O}_p(t^{1-2\beta}\log^{\frac{-4\alpha+4}{2}}(t))x_t\\
&\quad = x_t^\top M_t^{-1}o_p(1)x_t.
\end{aligned}$$
- $$x_t^\top \mathcal{O}_p\left(t^{-\frac{\beta}{2}}\log^{\frac{-\alpha+1}{2}}(t)\right)\Delta_u^{-1}(I_d+o_p(1))\mathcal{O}_p\left(t^{-\frac{\beta}{2}}\log^{\frac{-\alpha+1}{2}}(t)\right)x_t$$

$$
\begin{aligned}
&= x_t^\top \mathcal{O}_p\left(t^{-\beta}\log^{-\alpha+1}(t)\right)x_t\\
&= x_t^\top M_t^{-1} t^{1-\beta}\log^{-\alpha}(t)\mathcal{O}_p\left(t^{-\beta}\log^{-\alpha+1}(t)\right)x_t \qquad \text{(by Eq. 93)}\\
&= x_t^\top M_t^{-1}\mathcal{O}_p\left(t^{1-2\beta}\log^{-2\alpha+1}(t)\right)x_t\\
&= x_t^\top M_t^{-1} o_p(1) x_t.
\end{aligned}
$$

Thus the later two items are dominated by the first term, and the quadratic terms of x_t can be summarized by $x_t^\top M_t^{-1}(I_n + o_p(1))x_t = x_t^\top M_t^{-1}x_t(1 + o_p(1))$.

Quadratic terms of ξ_t That is already in a simple single item form, so we just keep it as $\xi_t^\top \Delta_u^{-1}(I_d + o_p(1))\xi_t = \xi_t^\top \Delta_u^{-1}\xi_t(1 + o_p(1))$.

Cross terms between x_t and ξ_t Finally consider the cross terms of x_t and ξ_t:

$$
\begin{aligned}
&2x_t^\top M_t^{-1}(I_n + o_p(1))\mathcal{O}_p(t^{1-3\beta/2}\log^{\frac{-3\alpha+3}{2}}(t))\Delta_u^{-1}\xi_t + 2\mathcal{O}_p(t^{-\frac{\beta}{2}}\log^{\frac{-\alpha+1}{2}}(t))x_t^\top \Delta_u^{-1}(I_d + o_p(1))\xi_t\\
&\quad= 2x_t^\top \mathcal{O}_p(t^{-\frac{\beta}{2}}\log^{\frac{-\alpha+3}{2}}(t))\xi_t + 2x_t^\top \mathcal{O}_p(t^{-\frac{\beta}{2}}\log^{\frac{-\alpha+1}{2}}(t))\xi_t \qquad \text{(by Eq. 93)}\\
&\quad= x_t^\top \mathcal{O}_p(t^{-\frac{\beta}{2}}\log^{\frac{-\alpha+3}{2}}(t))\xi_t\\
&\quad= x_t^\top o_p(t^{\frac{\beta-1}{2}}\log^{\frac{\alpha}{2}}(t))\xi_t \qquad \text{(because } \beta > 1/2 \text{ or } \beta = 1/2 \text{ and } \alpha > 3/2)\\
&\quad= x_t^\top M_t^{-1/2} o_p(1)\Delta_u^{-1/2}\xi_t \qquad \text{(by Eq. 93)}\\
&\quad\le o_p(1)\|x_t^\top M_t^{-1/2}\|\|\Delta_u^{-1/2}\xi_t\|\\
&\quad\le o_p(1)\left(x_t^\top M_t^{-1}x_t + \xi_t^\top \Delta_u^{-1}\xi_t\right),
\end{aligned}
$$

which is dominated by the quadratic part. To sum up, we have

$$
\begin{aligned}
&\begin{bmatrix} x_t \\ \hat{K}_t x_t + \xi_t \end{bmatrix}^\top \begin{bmatrix} \sum_{i=0}^{t-1} x_i x_i^\top & \sum_{i=1}^{t-1} x_i u_i^\top \\ \sum_{i=0}^{t-1} u_i x_i^\top & \sum_{i=1}^{t-1} u_i u_i^\top \end{bmatrix}^{-1} \begin{bmatrix} x_t \\ \hat{K}_t x_t + \xi_t \end{bmatrix} t^\beta \log^\alpha(t)\\
&\quad= (x_t^\top M_t^{-1}x_t + \xi_t^\top \Delta_u^{-1}\xi_t)(1 + o_p(1))\\
&\quad= \left(x_t^\top \log^\alpha(t)t^{-1+\beta}\left(\sum_{p=0}^\infty L^p\left(\sigma^2 I_n + 1_{\{\beta=1,\alpha=0\}}\tau^2 BB^\top\right)(L^p)^\top\right)^{-1} x_t + \xi_t^\top \frac{\beta}{\tau^2}\xi_t\right)(1 + o_p(1)).
\end{aligned}
$$

In other words

$$
\begin{aligned}
&t\sigma^2 \begin{bmatrix} x_t \\ \hat{K}_t x_t + \xi_t \end{bmatrix}^\top \begin{bmatrix} \sum_{i=0}^{t-1} x_i x_i^\top & \sum_{i=1}^{t-1} x_i u_i^\top \\ \sum_{i=0}^{t-1} u_i x_i^\top & \sum_{i=1}^{t-1} u_i u_i^\top \end{bmatrix}^{-1} \begin{bmatrix} x_t \\ \hat{K}_t x_t + \xi_t \end{bmatrix}\\
&\quad= \left(x_t^\top \left(\sum_{p=0}^\infty L^p\left(I_n + 1_{\{\beta=1,\alpha=0\}}\frac{\tau^2}{\sigma^2}BB^\top\right)(L^p)^\top\right)^{-1} x_t + \frac{\beta\sigma^2}{\tau^2}t^{1-\beta}\log^{-\alpha}(t)\|\xi_t\|^2\right)(1 + o_p(1)).
\end{aligned}
$$

■

H.4 Lemmas in Appendix G.1

H.4.1 The proof of Lemma 33

Lemma (A slightly different version of Theorem C.2 in Dean et al. (2018))**.** *Fixing $\delta \in (0, \frac{(n+d)\xi^2}{2}]$, for every T, k, ν, and ξ such that $\{z_t\}_{t=0}^T$ satisfies the (k,ν,ξ)-BMSB and*

$$T/k \geq \frac{10(n+d)}{\xi^2}\log\left(\frac{100(n+d)\sum_{t=1}^T \mathbf{Tr}(\mathbb{E}z_t z_t^\top)}{T\nu^2\xi^2\delta^{1+\frac{1}{n+d}}}\right).$$

the estimate $\hat{\Theta}_T$ defined in Eq. 62 satisfies the following statistical rate

$$\mathbb{P}\left[\left\|\hat{\Theta}_T - \Theta\right\|_2 > \frac{90\sigma}{\xi\nu}\sqrt{\frac{n+d}{T}\left(1+\log\left(\frac{10(n+d)\sum_{t=1}^T \mathbf{Tr}(\mathbb{E}z_t z_t^\top)}{T\delta^{1+\frac{1}{n+d}}\nu^2\xi}\right)\right)}\right] \leq 3\delta.$$

First let us review the main theorem in (Simchowitz et al., 2018). Lemma 33 is actually a corollary of that. To capture the excitation behavior observed in the case of linear systems we introduce a general martingale small-ball condition which quantifies the growth of the covariates X_t for vectors (notice that this is different from Definition 32).

Definition 47 (BMSB condition 2)**.** *Given an $\{\mathcal{F}_t\}_{t\geq 1}$-adapted random process $\{X_t\}_{t\geq 1}$ taking values in $\mathbb{R}^d$, we say that it satisfies the (k, Γ_{sb}, ξ)-matrix block martingale small-ball (BMSB) condition for $\Gamma_{sb} \succ 0$ if, for any $w \in \mathcal{S}^{d-1}$ and $j \geq 0$, $\frac{1}{k}\sum_{i=1}^k \mathbb{P}(|\langle w, X_{j+i}\rangle| \geq \sqrt{w^\top \Gamma_{sb} w}|\mathcal{F}_j) \geq \xi$ a.s.*

Theorem 48 (Theorem 2.4 in Simchowitz et al. (2018))**.** *Fix $\delta \in (0,1)$, $T \in \mathbb{N}$ and $0 \prec \Gamma_{sb} \preceq \bar{\Gamma}$. Then if $\{z_t, x_{t+1}\}_{t\geq 0} \in (\mathbb{R}^{d+n} \times \mathbb{R}^n)^T$ is a random sequence such that (a) $x_{t+1} = \Theta z_t + \varepsilon_t$, where $\varepsilon_t|\mathcal{F}_t$ is σ^2-sub-Gaussian and mean zero, (b) $z_0, \ldots, z_{T-1}$ satisfies the (k, Γ_{sb}, ξ)-small ball condition, and (c) such that $\mathbb{P}[\sum_{t=0}^{T-1} z_t z_t^\top \npreceq T\bar{\Gamma}] \leq \delta$. Then if*

$$T \geq \frac{10k}{\xi^2}\left(\log\left(\frac{1}{\delta}\right) + 2(d+n)\log(10/\xi) + \log\det(\bar{\Gamma}\Gamma_{sb}^{-1})\right),$$

we have $\hat{\Theta}_T$ defined in Eq. 62 satisfies the following statistical rate

$$\mathbb{P}\left[\left\|\hat{\Theta}_T - \Theta\right\| > \frac{90\sigma}{\xi}\sqrt{\frac{n + (n+d)\log\frac{10}{\xi} + \log\det\bar{\Gamma}\Gamma_{sb}^{-1} + \log\left(\frac{1}{\delta}\right)}{T\sigma_{\min}(\Gamma_{sb})}}\right] \leq 3\delta.$$

Now the main task is to translate this theorem to Lemma 33. First we need to derive the (a), (b), (c) three conditions from the assumptions in Lemma 33. Let us check the conditions one by one.

Condition (a) Theorem 48 states the model should be in the form of $x_t = \Theta z_t + \varepsilon_t$, where $\varepsilon_t|\mathcal{F}_t$ is σ^2-sub-Gaussian and mean zero. It is obvious that the system noise satisfy the sub-Gaussian and mean zero condition.

Condition (b) $z_1, \dots, z_T$ satisfies the (k, Γ_{sb}, ξ)-small ball condition.

Based on Definition 47, if we pick $\Gamma_{sb} = \nu^2 I_{n+d}$, then the condition becomes

$$\frac{1}{k}\sum_{i=1}^{k} \mathbb{P}(|\langle w, z_{j+i}\rangle| \geq \sqrt{w^\top \Gamma_{sb} w} = \nu | \mathcal{F}_j) \geq \xi \text{ a.s.} \tag{94}$$

Since we already assume $\{z_t\}_{t=0}^{T}$ satisfies the (k, ν, ξ)-BMSB (see Definition 32) in Lemma 33, Eq. 94 holds by definition.

Condition (c) We need to show that $\mathbb{P}[\sum_{t=0}^{T-1} z_t z_t^\top \npreceq T\bar{\Gamma}] \leq \delta$ for some choice $\bar{\Gamma}$. Let us take

$$\bar{\Gamma} = \frac{(n+d)\mathbb{E}\{\sum_{t=0}^{T-1} z_t z_t^\top\}}{T\delta} \succ 0. \tag{95}$$

First we need to show that $\bar{\Gamma} = \frac{(n+d)\mathbb{E}\{\sum_{t=0}^{T-1} z_t z_t^\top\}}{T\delta} \succeq \Gamma_{sb}$, and we can prove this from Eq. 94:

$$\text{For any } 0 \leq j \leq T-k, \quad \frac{1}{k}\sum_{i=1}^{k} \mathbb{P}(|\langle w, z_{j+i}\rangle| \geq \nu | \mathcal{F}_j) \geq \xi.$$

From a high level perspective, this equation allows us to have a lower bound on the minimum eigenvalue of $\mathbb{E}\{\sum_{t=0}^{T-1} z_t z_t^\top\}$, and then we can choose a δ small enough so that $\bar{\Gamma} \succeq \Gamma_{sb} = \nu^2 I_{n+d}$. By Markov inequality, for any $0 \leq j \leq T-k$,

$$\frac{\frac{1}{k}\sum_{i=1}^{k} \mathbb{E}|\langle w, z_{j+i}\rangle|}{\nu} \geq \xi.$$

This is equivalent to

$$\left(\frac{1}{k}\sum_{i=1}^{k} \mathbb{E}|\langle w, z_{j+i}\rangle|\right)^2 \geq \xi^2\nu^2.$$

By Cauchy–Schwarz inequality:

$$\frac{1}{k}\sum_{i=1}^{k} \mathbb{E}|\langle w, z_{j+i}\rangle|^2 \geq \frac{1}{k}\sum_{i=1}^{k} \mathbb{E}^2|\langle w, z_{j+i}\rangle| \geq \left(\frac{1}{k}\sum_{i=1}^{k} \mathbb{E}|\langle w, z_{j+i}\rangle|\right)^2 \geq \xi^2\nu^2.$$

Thus $\frac{1}{k}\sum_{i=1}^{k} \mathbb{E}|\langle w, z_{jk+i}\rangle|^2 \geq \xi^2\nu^2$. By summing up this inequality with $j = 0, 1, \cdots, \lfloor\frac{T-1}{k}\rfloor - 1$, we have

$$\frac{1}{\lfloor\frac{T-1}{k}\rfloor}\sum_{j=0}^{\lfloor\frac{T-1}{k}\rfloor-1} \frac{1}{k}\left(\sum_{i=1}^{k} \mathbb{E}|\langle w, z_{jk+i}\rangle|^2\right) \geq \xi^2\nu^2.$$

We can clean up the summation by merging $\sum_j$ and $\sum_i$ into one summation:

$$\frac{1}{k\lfloor\frac{T-1}{k}\rfloor}\sum_{t=1}^{k\lfloor\frac{T-1}{k}\rfloor} \mathbb{E}|\langle w, z_t\rangle|^2 \geq \xi^2\nu^2.$$

Recall that w is any vector in $\mathcal{S}^{d-1}$, so the above equation can be translated into

$$\begin{aligned}
\xi^2\nu^2 \leq & \min_{w\in\mathcal{S}^{d-1}} \frac{1}{k\lfloor\frac{T-1}{k}\rfloor}\sum_{t=1}^{k\lfloor\frac{T-1}{k}\rfloor} \mathbb{E}|\langle w, z_t\rangle|^2 \\
&= \mathbb{E}\left(\frac{1}{k\lfloor\frac{T-1}{k}\rfloor}\sum_{t=1}^{k\lfloor\frac{T-1}{k}\rfloor} z_t z_t^T\right) w \\
&= \min_{w\in\mathcal{S}^{d-1}} w^T \mathbb{E}\left(\frac{1}{k\lfloor\frac{T-1}{k}\rfloor}\sum_{t=1}^{k\lfloor\frac{T-1}{k}\rfloor} z_t z_t^T\right) w \\
&\leq \sigma_{\min}\left(\mathbb{E}\left(\sum_{t=0}^{T-1} z_t z_t^\top\right)/(k\lfloor\frac{T-1}{k}\rfloor)\right).
\end{aligned}$$

This means

$$\begin{aligned}
\lambda_{\min}\left(\bar{\Gamma}\right) =& \lambda_{\min}\left(\frac{(n+d)\mathbb{E}\left(\sum_{t=0}^{T-1} z_t z_t^\top\right)}{T\delta}\right) \\
=& \lambda_{\min}\left(\frac{(n+d)\mathbb{E}\left(\sum_{t=0}^{T-1} z_t z_t^\top/(k\lfloor\frac{T-1}{k}\rfloor)\right)}{T\delta}(k\lfloor\frac{T-1}{k}\rfloor)\right) \\
\geq& \frac{(n+d)\xi^2\nu^2}{T\delta}k\lfloor\frac{T-1}{k}\rfloor \\
\geq& \frac{(n+d)\xi^2\nu^2}{T\delta}\frac{T}{2} \qquad \text{(achieved when T is even and } k = T/2) \\
=& \frac{(n+d)\xi^2\nu^2}{2\delta}.
\end{aligned}$$

We wish to have $\frac{(n+d)\xi^2\nu^2}{2\delta} \geq \nu^2$ so that $\lambda_{\min}\left(\bar{\Gamma}\right) \geq \nu^2$ and $\bar{\Gamma} \succeq \Gamma_{sb} = \nu^2 I_{n+d}$. One sufficient condition is

$$\delta \leq \frac{(n+d)\xi^2}{2}.$$

Next we need to show $\mathbb{P}[\sum_{t=0}^{T-1} z_t z_t^\top \npreceq T\bar{\Gamma}] \leq \delta$. For simplicity denote $Z_T = \sum_{t=0}^{T-1} z_t z_t^\top$, which is a positive semi-definite matrix.

$$\begin{aligned}
\mathbb{P}[\sum_{t=0}^{T-1} z_t z_t^\top \npreceq T\bar{\Gamma}] =& \mathbb{P}[Z_T \npreceq \frac{\mathbb{E}\{Z_T\}(n+d)}{\delta}] \qquad \text{(by Eq. 95)} \\
=& \mathbb{P}[\mathbb{E}^{-1/2}\left(Z_T\right) Z_T \mathbb{E}^{-1/2}\left(Z_T\right) \npreceq \frac{I_{n+d}(n+d)}{\delta}] \\
=& \mathbb{P}[\lambda_{\max}\{\mathbb{E}^{-1/2}\left(Z_T\right) Z_T \mathbb{E}^{-1/2}\left(Z_T\right)\} \geq \frac{(n+d)}{\delta}] \\
\leq& \mathbb{P}[\mathbf{Tr}\{\mathbb{E}^{-1/2}\left(Z_T\right) Z_T \mathbb{E}^{-1/2}\left(Z_T\right)\} \geq \frac{(n+d)}{\delta}]
\end{aligned}$$

$$\begin{aligned}
&\leq \mathbb{E}[\mathbf{Tr}\{\mathbb{E}^{-1/2}\left(Z_T\right) Z_T \mathbb{E}^{-1/2}\left(Z_T\right)\}]\delta/(n+d) \qquad \text{(by Markov inequality)}\\
&= \mathbf{Tr}[\mathbb{E}\{\mathbb{E}^{-1/2}\left(Z_T\right) Z_T \mathbb{E}^{-1/2}\left(Z_T\right)\}]\delta/(n+d)\\
&= \mathbf{Tr}[I_{n+d}]\delta/(n+d)\\
&= \delta.
\end{aligned}$$

Result Now that we verified all conditions of Theorem 48, we can now translate the conclusion of Theorem 48 into our setting. Theorem 48 requires

$$T \geq \frac{10k}{\xi^2}\left(\log\left(\frac{1}{\delta}\right) + 2(d+n)\log(10/\xi) + \log\det(\bar{\Gamma}\Gamma_{sb}^{-1})\right).$$

First by our choice of Γ_{sb} and $\bar{\Gamma}$ we have

$$\begin{aligned}
\log\det(\bar{\Gamma}\Gamma_{sb}^{-1}) &= \log\det\left(\frac{(n+d)\mathbb{E}\{\sum_{t=0}^{T-1} z_t z_t^\top\}}{T\delta}\nu^{-2}\right)\\
&= \log\left(\left(\frac{(n+d)}{T\delta\nu^2}\right)^{n+d}\det\left(\mathbb{E}\{\sum_{t=0}^{T-1} z_t z_t^\top\}\right)\right)\\
&\leq \log\left(\left(\frac{(n+d)}{T\delta\nu^2}\right)^{n+d}\left(\sum_{t=0}^{T-1}\mathbf{Tr}(\mathbb{E}z_t z_t^\top)\right)^{n+d}\right) \qquad (96)\\
&= (n+d)\log\left(\frac{(n+d)}{T\delta\nu^2}\sum_{t=0}^{T-1}\mathbf{Tr}(\mathbb{E}z_t z_t^\top)\right).
\end{aligned}$$

With this in hand, we know that

$$\begin{aligned}
&\frac{10k}{\xi^2}\left(\log\left(\frac{1}{\delta}\right) + 2(d+n)\log(10/\xi) + \log\det(\bar{\Gamma}\Gamma_{sb}^{-1})\right)\\
&\quad\leq \frac{10k}{\xi^2}\left(\log\left(\frac{1}{\delta}\right) + 2(d+n)\log(10/\xi) + (n+d)\log\left(\frac{(n+d)}{T\delta\nu^2}\sum_{t=0}^{T-1}\mathbf{Tr}(\mathbb{E}z_t z_t^\top)\right)\right)\\
&\quad= \frac{10(n+d)k}{\xi^2}\left(\log\left(\delta^{-\frac{1}{n+d}}\right) + \log(100/\xi^2) + \log\left(\frac{(n+d)}{T\delta\nu^2}\sum_{t=0}^{T-1}\mathbf{Tr}(\mathbb{E}z_t z_t^\top)\right)\right)\\
&\quad= \frac{10(n+d)k}{\xi^2}\log\left(\frac{100(n+d)\sum_{t=0}^{T-1}\mathbf{Tr}(\mathbb{E}z_t z_t^\top)}{T\nu^2\xi^2\delta^{1+\frac{1}{n+d}}}\right).
\end{aligned}$$

Thus one sufficient condition for the requirement in Theorem 48 is

$$T/k \geq \frac{10(n+d)}{\xi^2}\log\left(\frac{100(n+d)\sum_{t=0}^{T-1}\mathbf{Tr}(\mathbb{E}z_t z_t^\top)}{T\nu^2\xi^2\delta^{1+\frac{1}{n+d}}}\right).$$

Finally we need to translate the conclusion of Theorem 48:

$$\mathbb{P}\left[\left\|\hat{\Theta}_T - \Theta\right\| > \frac{90\sigma}{\xi}\sqrt{\frac{n + (n+d)\log\frac{10}{\xi} + \log\det\bar{\Gamma}\Gamma_{sb}^{-1} + \log\left(\frac{1}{\delta}\right)}{T\sigma_{\min}(\Gamma_{sb})}}\right] \leq 3\delta.$$

By Eq. 96 and $\Gamma_{sb} = \nu^2 I_{n+d}$ we have

$$\mathbb{P}\left[\left\|\hat{\Theta}_T - \Theta\right\| > \frac{90\sigma}{\xi}\sqrt{\frac{n + (n+d)\log\frac{10}{\xi} + (n+d)\log\left(\frac{(n+d)}{T\delta\nu^2}\sum_{t=0}^{T-1}\mathbf{Tr}(\mathbb{E}z_t z_t^\top)\right) + \log\left(\frac{1}{\delta}\right)}{T\nu^2}}\right] \le 3\delta.$$

Notice that

$$\begin{aligned}
&n + (n+d)\log\frac{10}{\xi} + (n+d)\log\left(\frac{(n+d)}{T\delta\nu^2}\sum_{t=0}^{T-1}\mathbf{Tr}(\mathbb{E}z_t z_t^\top)\right) + \log\left(\frac{1}{\delta}\right)\\
&\le (n+d)\left(1 + \log\frac{10}{\xi} + \log\left(\frac{(n+d)}{T\delta\nu^2}\sum_{t=0}^{T-1}\mathbf{Tr}(\mathbb{E}z_t z_t^\top)\right) + \log\delta^{-\frac{1}{n+d}}\right)\\
&= (n+d)\left(1 + \log\left(\frac{10(n+d)\sum_{t=0}^{T-1}\mathbf{Tr}(\mathbb{E}z_t z_t^\top)}{T\delta^{1+\frac{1}{n+d}}\nu^2\xi}\right)\right).
\end{aligned}$$

Combining this with the previous inequality we have

$$\mathbb{P}\left[\left\|\hat{\Theta}_T - \Theta\right\| > \frac{90\sigma}{\xi\nu}\sqrt{\frac{n+d}{T}\left(1 + \log\left(\frac{10(n+d)\sum_{t=0}^{T-1}\mathbf{Tr}(\mathbb{E}z_t z_t^\top)}{T\delta^{1+\frac{1}{n+d}}\nu^2\xi}\right)\right)}\right] \le 3\delta.$$

H.4.2 The proof of Lemma 34

Lemma (Similar to Lemma C.3 in Dean et al. (2018))**.** *If we assume Assumption 1, then apply Algorithm 1, the process $\{z_t\}_{t\ge 0}^T$ satisfies the (k, ν, ξ)-BMSB condition for*

$$(k, \nu, \xi) = \left(1, \sqrt{\sigma_{\eta,T}^2 \min\left(\frac{1}{2}, \frac{\sigma^2}{2\sigma^2 C_K^2 + \tau^2}\right)}, \frac{3}{10}\right),$$

where $\sigma_{\eta,T}^2 = \tau^2 T^{\beta-1}\log^{\alpha}(T)$.

Proof

By Definition 32 the statement means, for any $v \in \mathcal{S}^{n+d}$ and $0 \le t \le T-1$:

$$\mathbb{P}\left(|\langle v, z_{t+1}\rangle| \ge \sqrt{\sigma_{\eta,T}^2 \min\left(\frac{1}{2}, \frac{\sigma^2}{2\sigma^2 C_K^2 + \tau^2}\right)}\middle|\mathcal{F}_t\right) \ge 3/10.$$

Recall that

$$x_{t+1} = Ax_t + Bu_t + \varepsilon_t.$$

$$u_{t+1} = \hat{K}_{t+1}x_{t+1} + \eta_{t+1} = \hat{K}_{t+1}(Ax_t + Bu_t + \varepsilon_t) + \eta_{t+1}.$$

Denote the filtration $\mathcal{F}_t = \sigma(x_0, \eta_0, \varepsilon_0 \ldots, \eta_{t-1}, \varepsilon_{t-1}, \eta_t) = \sigma(x_0, u_0, x_1, \cdots, x_t, u_t)$. It is clear that the process $\{z_t\}_{t\ge 0}$ is $\{\mathcal{F}_t\}_{t\ge 0}$-adapted.

Recall that $\hat{K}_{t+1}$ is decided by $\hat{A}_t, \hat{B}_t$ in Algorithm 1, where our estimator $\hat{A}_t, \hat{B}_t$ is designed to be only dependent on $x_0, u_0, x_1, \cdots, u_{t-1}, x_t$, which means

$$\hat{K}_{t+1} \in \mathcal{F}_t = \sigma(x_0, u_0, x_1, \cdots, x_t, u_t).$$

For all $t \geq 1$, denote

$$\xi_{t+1} := \hat{K}_{t+1}(Ax_t + Bu_t) \in \mathcal{F}_t.$$

Now we are ready to prove Lemma 34. We have

$$\begin{bmatrix} x_{t+1} \\ u_{t+1} \end{bmatrix} = \begin{bmatrix} Ax_t + Bu_t \\ \xi_{t+1} \end{bmatrix} + \begin{bmatrix} I_n & 0 \\ \hat{K}_{t+1} & I_d \end{bmatrix} \begin{bmatrix} \varepsilon_t \\ \eta_{t+1} \end{bmatrix}.$$

Given $\mathcal{F}_t$, $\begin{bmatrix} x_{t+1} \\ u_{t+1} \end{bmatrix}$ only has randomness in $\begin{bmatrix} I_n & 0 \\ \hat{K}_{t+1} & I_d \end{bmatrix} \begin{bmatrix} \varepsilon_t \\ \eta_{t+1} \end{bmatrix}$, where $\begin{bmatrix} I_n & 0 \\ \hat{K}_{t+1} & I_d \end{bmatrix}$ is fixed given $\mathcal{F}_t$, and $\begin{bmatrix} \varepsilon_t \\ \eta_{t+1} \end{bmatrix}$ follows $\mathcal{N}\left(0, \begin{bmatrix} \sigma^2 I_n & 0 \\ 0 & \sigma^2_{\eta,t+1} I_d \end{bmatrix}\right)$. That implies

$$\begin{bmatrix} x_{t+1} \\ u_{t+1} \end{bmatrix} \Bigg| \mathcal{F}_t \sim \mathcal{N}\left(\begin{bmatrix} Ax_t + Bu_t \\ \xi_{t+1} \end{bmatrix}, \begin{bmatrix} \sigma^2 I_n & \sigma^2 \hat{K}_{t+1}^\top \\ \sigma^2 \hat{K}_{t+1} & \sigma^2 \hat{K}_{t+1} \hat{K}_{t+1}^\top + \sigma^2_{\eta,t+1} I_d \end{bmatrix} \right).$$

Denote $\mu_{z,t+1}$ and $\Sigma_{z,t+1}$ as the mean and covariance of this multivariate normal distribution. Recall that we denoted $z_{t+1} = \begin{bmatrix} x_{t+1} \\ u_{t+1} \end{bmatrix}$. Let $v \in \mathcal{S}^{n+d}$ and then $\langle v, z_{t+1} \rangle \Big| \mathcal{F}_t \sim \mathcal{N}(\langle v, \mu_{z,t+1} \rangle, v^\top \Sigma_{z,t+1} v)$. Therefore,

$$\begin{aligned} \mathbb{P}\left(|\langle v, z_{t+1} \rangle| \geq \sqrt{\sigma_{\min}(\Sigma_{z,t+1})} \Bigg| \mathcal{F}_t \right) &\geq \mathbb{P}\left(|\langle v, z_{t+1} \rangle| \geq \sqrt{v^\top \Sigma_{z,t+1} v} \Bigg| \mathcal{F}_t \right) \\ &\geq \mathbb{P}\left(|\langle v, z_{t+1} - \mu_{z,t+1} \rangle| \geq \sqrt{v^\top \Sigma_{z,t+1} v} \Bigg| \mathcal{F}_t \right) \\ &\geq 3/10. \end{aligned} \tag{97}$$

Here we used the fact that for any $\mu, \sigma^2 \in \mathbb{R}$ and $\omega \sim \mathcal{N}(0, \sigma^2)$, we have:

$$\mathbb{P}(|\mu + \omega| \geq \sigma) \geq \mathbb{P}(|\omega| \geq \sigma) \geq 3/10.$$

Recall in Algorithm 1, we force all our controllers $\hat{K}_t$ to have norm $\|\hat{K}_t\| \leq C_K$, where C_K is a constant. Then, by a simple argument based on a Schur complement (Lemma 49):

$$\begin{aligned} \sigma_{\min}(\Sigma_{z,t+1}) \geq & \sigma^2_{\eta,t} \min\left(\frac{1}{2}, \frac{\sigma^2}{2\left\| \hat{K}_{t+1} \sigma^2 \hat{K}_{t+1}^\top \right\|_2 + \sigma^2_{\eta,t}} \right) \\ \geq & \sigma^2_{\eta,t} \min\left(\frac{1}{2}, \frac{\sigma^2}{2\sigma^2 C_K^2 + \sigma^2_{\eta,t}} \right) \\ \geq & \sigma^2_{\eta,T} \min\left(\frac{1}{2}, \frac{\sigma^2}{2\sigma^2 C_K^2 + \tau^2} \right). \end{aligned}$$

The desired conclusion directly follows:

$$\begin{aligned}
&\mathbb{P}\left(|\langle v, z_{t+1}\rangle| \geq \sqrt{\sigma_{\eta,T}^2 \min\left(\frac{1}{2}, \frac{\sigma^2}{2\sigma^2 C_K^2 + \tau^2}\right)}\middle| \mathcal{F}_t\right) \\
&\geq \mathbb{P}\left(|\langle v, z_{t+1}\rangle| \geq \sqrt{\sigma_{\min}(\Sigma_{z,t+1})}\middle| \mathcal{F}_t\right) \\
&\geq 3/10 \qquad \text{by Eq. 97.}
\end{aligned}$$

■

Schur complement

Lemma 49 (Lemma F.1 in Mania et al. (2019)). *Let Σ be a $n \times n$ positive-definite matrix and let K be a real $d \times n$ matrix. Then, for any $\sigma_u \in \mathbb{R}$ we have that*

$$\sigma_{\min}\left(\begin{bmatrix} \Sigma & \Sigma K^\top \\ K\Sigma & K\Sigma K^\top + \sigma_u^2 I \end{bmatrix}\right) \geq \sigma_u^2 \min\left(\frac{1}{2}, \frac{\sigma_{\min}(\Sigma)}{2\left\|K\Sigma K^\top\right\|_2 + \sigma_u^2}\right).$$

H.4.3 The proof of Lemma 35

Lemma (Similar to Lemma C.4 in Dean et al. (2018)). *If we assume Assumption 1, then apply Algorithm 1, the process $\{z_t\}_{t\geq 0}^T$ satisfies*

$$\sum_{t=0}^{T-1} \mathbf{Tr}\left(\mathbb{E} z_t z_t^\top\right) = \mathcal{O}(T \log^2(T)).$$

Proof

Now, note that

$$\mathbf{Tr}\left(\mathbb{E} z_t z_t^\top\right) = \mathbb{E}\left(\mathbf{Tr}\, z_t z_t^\top\right) = \mathbb{E}\|z_t\|^2 = \mathbb{E}\left(\|x_t\|^2 + \|u_t\|^2\right).$$

Since $\|u_t\| = \|\hat{K}_t x_t + \eta_t\| \leq \|\hat{K}_t\|\|x_t\| + \|\eta_t\| \leq C_K\|x_t\| + \|\eta_t\|$, we will show that if we can bound $\|x_t\|$, then we can also get a bound for $\|u_t\|$ in the same order. Next we will focus on deriving the bound for $\|x_t\|$.

Define $C_{x,t} := C_x \log(t)$. Since $\rho(A + BK_0) < 1$, there exists some integer m that $\|(A + BK_0)^m\| < (\frac{\rho(A+BK_0)+1}{2})^m$. Let us denote $\rho := \frac{\rho(A+BK_0)+1}{2} < 1$ just for this Lemma 35.

For each $t > m + 1$, one of the following two statement must be true:

- $\|x_{t-i}\| > C_{x,t-i}, (i = 2, \cdots, m + 1)$.
- $\exists i \in \{2, \cdots, m + 1\}$, which satisfies $\|x_{t-i}\| \leq C_{x,t-i}$.

We can derive an upper bound for $\|x_t\|$ in both cases, and thus have an upper bound for every $\|x_t\|$ by adding up those two bounds in two different cases.

1. If $\|x_{t-i}\| > C_{x,t-i}, (i = 2, \cdots, m+1)$, recall that if $\|x_t\| > C_{x,t}$, then we assert our controller in the next step to be probing noise: $u_{t+1} = K_0 x_{t+1} + \eta_{t+1}$. By assumption we already had $\|x_k\| > C_{x,k}$, for $k = t-m-1, t-m, \cdots, t-2$. That means we have a consecutive m steps of probing noise with $u_k = K_0 x_k + \eta_k$, for $k = t-m, t-(m-1), \cdots, t-1$. Now we have

$$x_{k+1} = (A + BK_0)x_k + B\eta_k + \varepsilon_k, \quad \text{for} \quad k = t-m, t-(m-1), \cdots, t-1.$$

That is

$$x_t = (A + BK_0)^m x_{t-m} + \sum_{k=0}^{m-1} (A + BK_0)^k (B\eta_{t-1-k} + \varepsilon_{t-1-k}).$$

which implies

$$\|x_t\| \leq \|(A + BK_0)^m\| \|x_{t-m}\| + \sum_{k=0}^{m-1} \|(A + BK_0)^k\| \|(B\eta_{t-1-k} + \varepsilon_{t-1-k})\|. \tag{98}$$

2. If $\exists i \in \{2, \cdots, m+1\}$, which satisfies $\|x_{t-i}\| \leq C_{x,t-i}, (i = 2, \cdots, m+1)$, then consider the following relationship

$$\begin{aligned} x_t &= Ax_{t-1} + Bu_{t-1} + \varepsilon_{t-1} \\ &= (A + B\hat{K}_{t-1})x_{t-1} + B\eta_{t-1} + \varepsilon_{t-1}. \end{aligned}$$

Therefore by our algorithm design that $\|\hat{K}_t\| \leq C_K$ for any t

$$\begin{aligned} \|x_t\| &\leq \|A + B\hat{K}_{t-1}\| \|x_{t-1}\| + \|B\eta_{t-1} + \varepsilon_{t-1}\| \\ &\leq (\|A\| + \|B\| \|\hat{K}_{t-1}\|) \|x_{t-1}\| + \|B\eta_{t-1} + \varepsilon_{t-1}\| \\ &\leq (\|A\| + \|B\| C_K) \|x_{t-1}\| + \|B\eta_{t-1} + \varepsilon_{t-1}\| \\ &\leq (\|A\| + \|B\| C_K)^i \|x_{t-i}\| + \sum_{k=0}^{i-1} (\|A\| + \|B\| C_K)^k \|B\eta_{t-1-k} + \varepsilon_{t-1-k}\| \\ &\leq \max\{1, (\|A\| + \|B\| C_K)^m\} C_{x,t} + \sum_{k=0}^{m-1} (\|A\| + \|B\| C_K)^k \|B\eta_{t-1-k} + \varepsilon_{t-1-k}\|. \end{aligned} \tag{99}$$

By adding up Eqs. 98 and 99, we have a bound that is applicable to both cases. Notice our previous assumption that $\|(A + BK_0)^m\| \leq \rho^m$, where $\rho < 1$, further take $\|(A + BK_0)^k\|$, and $(\|A\| + \|B\| C_K)^k$ to be all bounded by a constant $M \geq 1$ for $k = 0, 1, \cdots, m$, which is of order $M = \mathcal{O}(1)$ (because $m = \mathcal{O}(1)$). By Eqs. 98 and 99

$$\begin{aligned}
\|x_t\| &\leq \|(A+BK_0)^m\|\|x_{t-m}\| + \sum_{k=0}^{m-1} \|(A+BK_0)^k\|\|B\eta_{t-1-k} + \varepsilon_{t-1-k}\| \\
&\qquad + \max\{1, (\|A\| + \|B\|C_K)^m\} C_{x,t-i} + \sum_{k=0}^{m-1} (\|A\| + \|B\|C_K)^k \|B\eta_{t-k-1} + \varepsilon_{t-k-1}\| \\
&\leq \rho^m \|x_{t-m}\| + M\left(C_{x,t-i} + 2\sum_{k=0}^{m-1} \|B\eta_{t-k-1} + \varepsilon_{t-k-1}\|\right) \\
&\leq \rho^m \|x_{t-m}\| + M\left(C_{x,t} + 2\sum_{k=0}^{m-1} \|B\eta_{t-k-1} + \varepsilon_{t-k-1}\|\right).
\end{aligned} \tag{100}$$

Eq. 100 is very promising because it has a shrinking weight on $\|x_{t-m}\|$. Let us use a simplified notation for the remainder:

$$J_t := M\left(C_{x,t} + 2\sum_{k=0}^{m-1} \|B\eta_{t-k-1} + \varepsilon_{t-k-1}\|\right).$$

In $\mathbb{E}[J_t^2]$ there are three types of components:

- $M = \mathcal{O}(1)$
- $C_{x,t} = C_x \log(t)$
- $\mathbb{E}(\sum_{k=0}^{m-1} \|B\eta_{t-k-1} + \varepsilon_{t-k-1}\|)^2 = \mathcal{O}(1)$.

Since

$$\mathbb{E}[J_t^2] \leq M^2 \cdot 2(C_{x,t}^2 + 4\mathbb{E}(\sum_{k=0}^{m-1} \|B\eta_{t-k-1} + \varepsilon_{t-k-1}\|)^2) = \mathcal{O}(\log^2(t)), \tag{101}$$

we can control $\mathbb{E}\|x_t\|^2$ by

$$\begin{aligned}
\mathbb{E}\|x_t\|^2 &\leq \mathbb{E}\left(\rho^m \|x_{t-m}\| + J_t\right)^2 \\
&= \rho^{2m}\mathbb{E}\|x_{t-m}\|^2 + \mathbb{E}J_t^2 + 2\rho^m \mathbb{E}\|x_{t-m}\||J_t| \\
&\leq \rho^{2m}\mathbb{E}\|x_{t-m}\|^2 + \mathbb{E}J_t^2 + \frac{1-\rho^{2m}}{2}\mathbb{E}\|x_{t-m}\|^2 + \frac{2\rho^{2m}}{1-\rho^{2m}}\mathbb{E}J_t^2 \\
&= \frac{1+\rho^{2m}}{2}\mathbb{E}\|x_{t-m}\|^2 + \frac{1+\rho^{2m}}{1-\rho^{2m}}\mathbb{E}J_t^2 \\
&\qquad \text{(because } 2ab \leq a^2 + b^2 \text{ with } a^2 = \frac{1-\rho^{2m}}{2}\|x_{t-m}\|^2 \text{ and } b^2 = \frac{2\rho^{2m}}{1-\rho^{2m}}J_t^2\text{).}
\end{aligned} \tag{102}$$

By Eqs. 101 and 102,

$$
\begin{aligned}
\mathbb{E}\|x_t\|^2 &\le \frac{1+\rho^{2m}}{2}\mathbb{E}\|x_{t-m}\|^2 + \mathcal{O}(\log^2(t)) \\
&\le (\frac{1+\rho^{2m}}{2})^2\mathbb{E}\|x_{t-2m}\|^2 + \frac{1+\rho^{2m}}{2}\mathcal{O}(\log^2(t)) + \mathcal{O}(\log^2(t)) \\
&\le (\frac{1+\rho^{2m}}{2})^{\lfloor \frac{t}{m} \rfloor}\mathbb{E}\|x_{t-m\lfloor \frac{t}{m} \rfloor}\|^2 + \sum_{i=0}^{\lfloor \frac{t}{m} \rfloor -1} (\frac{1+\rho^{2m}}{2})^i\mathcal{O}(\log^2(t)) \qquad (103)\\
&\le \mathbb{E}\|x_{t-m\lfloor \frac{t}{m} \rfloor}\|^2 + \mathcal{O}(\log^2(t)) \\
&\qquad \text{(Recall that } \rho < 1 \text{, and thus } \frac{1+\rho^{2m}}{2} < 1).
\end{aligned}
$$

Now it only remains to show that $\mathbb{E}\|x_{t-m\lfloor \frac{t}{m} \rfloor}\|^2$ is bounded by some constant. Notice that

$$
\begin{aligned}
\mathbb{E}\|x_t\|^2 \le& \mathbb{E}\left((\|A\| + \|B\|\|\hat{K}_t\|)\|x_{t-1}\| + \|B\|\|\eta_t\| + \|\varepsilon_t\|\right)^2 \\
\le& 3\left((\|A\| + \|B\|C_K)^2\mathbb{E}\|x_{t-1}\|^2 + \|B\|^2\mathbb{E}\|\eta_t\|^2 + \|\varepsilon_t\|^2\right) \\
\le& 3\left((\|A\| + \|B\|C_K)^2\mathbb{E}\|x_{t-1}\|^2 + \|B\|^2\tau^2 + \sigma^2\right).
\end{aligned}
$$

By iteratively applying this inequality down to $\mathbb{E}\|x_0\|^2$, we know that for $t \le m$:

$$\mathbb{E}\|x_t\|^2 = \mathcal{O}(1).$$

Thus following from Eq. 103 we have

$$\mathbb{E}\|x_t\|^2 = \mathcal{O}(\log^2(t)).$$

Since we already controlled the expectation of $\|x_t\|^2$, it is straightforward to control the expectation of $\|u_t\|^2$:

$$u_t = \hat{K}_t x_t + \eta_t.$$

$$
\begin{aligned}
\mathbb{E}\|u_t\|^2 &\le \mathbb{E}\|\hat{K}_t x_t + \eta_t\|^2 \\
&\le 2\mathbb{E}(\|\hat{K}_t\|^2\|x_t\|^2 + \|\eta_t\|^2) \\
&\le 2\mathbb{E}(C_K^2\|x_t\|^2 + \|\eta_t\|^2) \\
&\le \mathcal{O}(\log^2(t)).
\end{aligned}
$$

Thus,

$$\mathbb{E}\|z_t\|^2 = \mathbb{E}\|x_t\|^2 + \mathbb{E}\|u_t\|^2 \le \mathcal{O}(\log^2(t)) \qquad (104)$$

Then we have

$$\mathbb{E}\sum_{t=0}^{T-1}\|x_t\|^2, \mathbb{E}\sum_{t=0}^{T-1}\|u_t\|^2, \mathbb{E}\sum_{t=0}^{T-1}\|z_t\|^2 \le \mathcal{O}(T\log^2(T)).$$

■

H.5 Lemma in Appendix G.3

H.5.1 The proof of Lemma 43

Lemma. *Suppose we have a constant square matrix M with spectral radius $\rho(M) < 1$, and a sequence of uniformly bounded random variables $\{\delta_t\}_{t=0}^{\infty}$, satisfying $\|\delta_t\| \xrightarrow{a.s.} 0$. Denote the constant $\rho_M := \frac{2+\rho(M)}{3} < 1$. Then we have, for any $t, q \in \mathbb{N}$, $t > q$:*

$$\|(M+\delta_{t-1})\cdots(M+\delta_q)\| = \mathcal{O}(\rho_M^{t-q}) \ a.s.$$

And as a direct corollary

$$\|M^{t-q}\| = \mathcal{O}(\rho_M^{t-q}).$$

Proof

Our assumption of stability only says $\rho(M) < 1$, but our analysis prefers similar exponential decay with regard to spectral norm. First, we need a conversion between spectral radius and spectral norm. Define

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

For simplicity, let us denote

$$\tau(M) := \tau\left(M, \frac{1+\rho(M)}{2}\right).$$

and with Gelfand's Formula

$$\rho(M) = \lim_{k\to\infty}\left\|M^k\right\|^{\frac{1}{k}}.$$

Thus $\tau(M)$ is finite because $\frac{1+\rho(M)}{2} > \rho(M)$. Since $\{\delta_t\}_{t=0}^{\infty}$ is uniformly bounded, we can assume an upper bound U_δ for $\|M+\delta_i\|$. Let us now consider the spectral norm of $(M+\delta_{t-1})\cdots(M+\delta_q)$.

$$\begin{aligned}
\|(M+\delta_{t-1})\cdots(M+\delta_q)\| \leq & \sum_{m=0}^{t-q}\|M^{t-q-m}\| \sum_{q\leq k_1<\cdots<k_m\leq t-1}\prod_{j=1}^{m}\|\delta_{k_j}\| \\
\leq & \sum_{m=0}^{t-q}\tau(M)\left(\frac{1+\rho(M)}{2}\right)^{t-q-m} \sum_{q\leq k_1<\cdots<k_m\leq t-1}\prod_{j=1}^{m}\|\delta_{k_j}\| \\
= & \tau(M)\sum_{m=0}^{t-q}\left(\frac{1+\rho(M)}{2}\right)^{t-q-m} \sum_{q\leq k_1<\cdots<k_m\leq t-1}\prod_{j=1}^{m}\|\delta_{k_j}\| \\
= & \tau(M)\left(\frac{1+\rho(M)}{2}+\|\delta_{t-1}\|\right)\cdots\left(\frac{1+\rho(M)}{2}+\|\delta_{q+1}\|\right).
\end{aligned}$$

Since $\|\delta_t\| \to 0$ a.s., for every ω in the sample space Ω, such that there exists some $T_1(\omega)$, whenever $t > T_1(\omega)$, $\frac{1+\rho(M)}{2} + \|\delta_t\| < \frac{2+\rho(M)}{3} < 1$, then

$$\|(M+\delta_{t-1})\cdots(M+\delta_q)\| \leq \tau(M)\left(\frac{1+\rho(M)}{2}+\|\delta_{t-1}\|\right)\cdots\left(\frac{1+\rho(M)}{2}+\|\delta_{q+1}\|\right)$$

$$\leq \tau(M)\rho_M^{t-q-T_1(\omega)}(\frac{1+\rho(M)}{2}+U_\delta)^{T_1(\omega)}.$$

Following Definition 14 Item 8:

$$\|(M+\delta_{t-1})\cdots(M+\delta_q)\| = \mathcal{O}(\rho_M^{t-q}) \text{ a.s.}$$

■

Appendix I. Experiment Details

I.1 Experiment Setting

I.1.1 Experiment Setting on Stable System

We set $A = \begin{bmatrix} 0.8 & 0.1 \\ 0 & 0.8 \end{bmatrix}$ and $B = \begin{bmatrix} 0 \\ 1 \end{bmatrix}$, with system noise $\sigma = 1$, injected noise baseline $\tau = 1$, $Q = I_2$, $R = 1$ and initial state $x_0 = [0,0]^\top$. As for the algorithmic hyper-parameters, we set the warning threshold for states x_t at $C_x = 1$ (so that $C_{x,t} = \log(t)$), the known stable controller $K_0 = [0,0]$, and the upper bound of the L2-norm for our controller $\hat{K}_t$ at $C_K = 5$. Note that this is conservative by about a factor of 10, since the true optimal controller in this system is $K \approx \begin{bmatrix} -0.10, -0.48 \end{bmatrix}$. Recall that the choice of these hyper-parameters does not actually affect our theoretical coverage (as long as $C_K > \|K\|$) or regret guarantees, but in practice their values prevent the system from incurring very large regret in the first few time steps. Even for this, they are only needed because we do not assume we are given an initial controller that is very close to K; in contrast, for instance, Dean et al. (2018) started from a controller fitted with 100 samples of white noise actions. All stable system results are based on 1,000 independent runs of Algorithm 1 for $T = 10,000$ time steps.

I.1.2 Experiment Setting on Unstable System

The unstable system we simulate is highly unstable, and is largely the same as that in Appendix H of Dean et al. (2018). We set $A = \begin{bmatrix} 2 & 0 & 0 \\ 4 & 2 & 0 \\ 0 & 4 & 2 \end{bmatrix}$ and $B = I_3$, with system noise $\sigma = 1$, injected noise baseline $\tau = 1$, $Q = 10I_3$, $R = I_3$ and initial state $x_0 = [0,0,0]^\top$. As for the hyper-parameters, we set the warning threshold for states x_t at $C_x = 1$ (so that $C_{x,t} = \log(t)$), and we examined two different choices for the known stabilizing controller: $K_0 = -\begin{bmatrix} 1.5 & 0 & 0 \\ 0 & 1.5 & 0 \\ 0 & 0 & 1.5 \end{bmatrix}$ and $K_0 = -\begin{bmatrix} 1.5 & 0 & 0 \\ 3.5 & 1.5 & 0 \\ 0 & 3.5 & 1.5 \end{bmatrix}$. The former choice incurs quite a bit higher regret than the latter, and hence we refer to the former as the 'bad' stabilizing controller and to the latter as the 'good' stabilizing controller. We set the upper bound of the L2-norm for our controller $\hat{K}_t$ at the level of $C_K = 1000$. Our choice of K_0 is different from the starting point in Dean et al. (2018), where they started from a $T = 250$ burn in period estimate, and did not report the regret in the first 250 steps. All unstable system results are based on 1,000 independent runs of Algorithm 1 for $T = 5,000$ time steps.

I.2 Experiment on Unstable System

In contrast to the stable system simulation summarized in Section 4.1, in this section we simulate the severely unstable system described in Appendix I.1.2. In this setting, the specification of K_0 is critical due to the costs incurred at the early time steps, an unavoidable consequence of starting from limited information in a system that can rapidly spiral (nearly) out of control.

I.2.1 Summary of results on unstable system

We begin with the analogue of Fig. 1 for the unstable system, given in Fig. I.1. The main takeaways are the same as the discussion in Appendix 4.1.

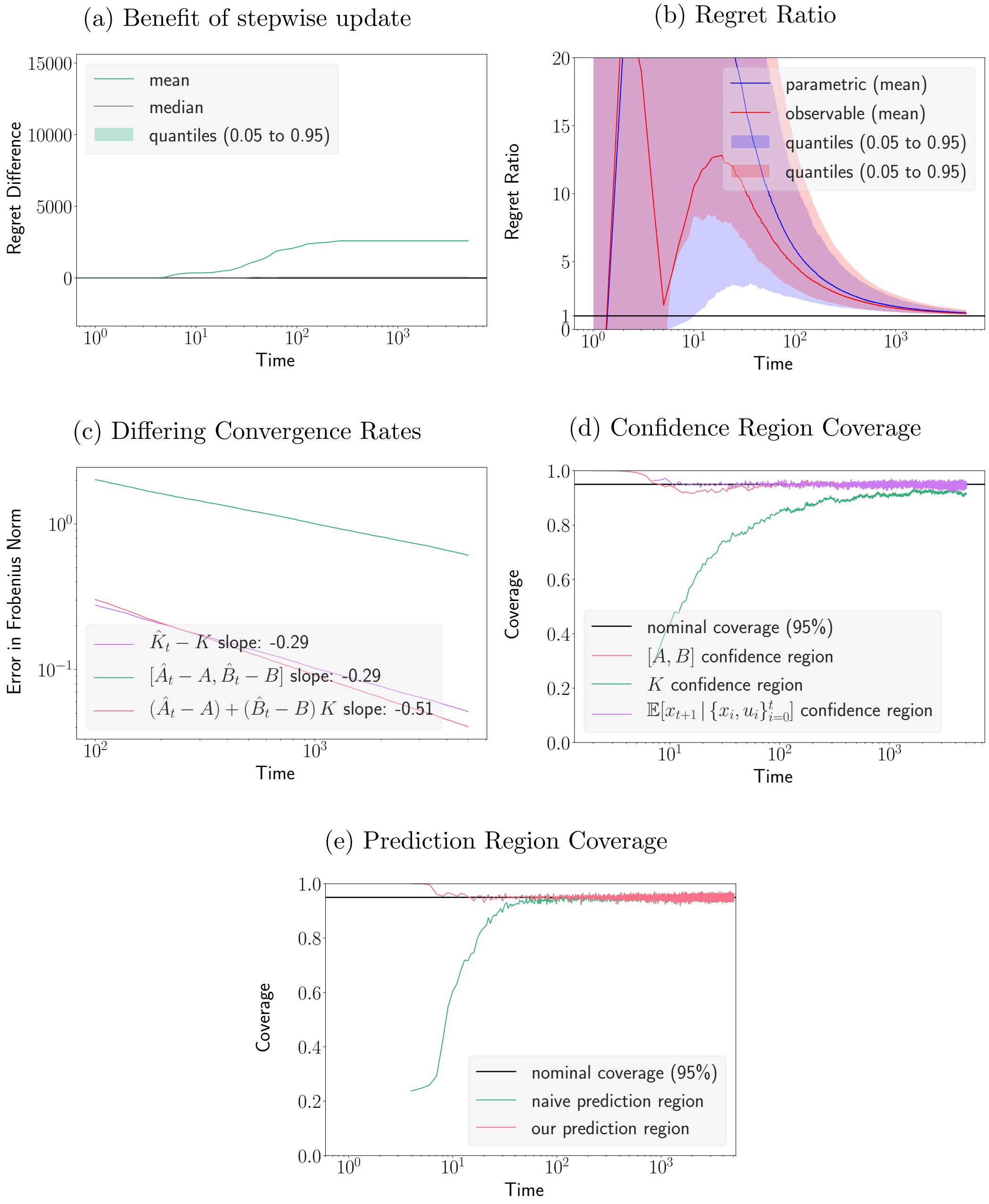
I.2.2 Large Regret From Early Time Steps

For the 'bad' choice of stabilizing controller $K_0 = -\begin{bmatrix} 1.5 & 0 & 0 \\ 0 & 1.5 & 0 \\ 0 & 0 & 1.5 \end{bmatrix}$, we plot the log regret in subplot (a) of Fig. I.2. We observe a rapidly increasing regret in the first roughly 200 time steps, which dominates all the regret in the remaining steps. We offer a brief explanation why the cost in the early time steps is very large despite assuming knowledge of a stabilizing yet sub-optimal controller K_0. Notice $A + BK_0 = \begin{bmatrix} 0.5 & 0 & 0 \\ 4 & 0.5 & 0 \\ 0 & 4 & 0.5 \end{bmatrix}$. Thus $(A + BK_0)^2 = \begin{bmatrix} 0.25 & 0 & 0 \\ 4 & 0.25 & 0 \\ 16 & 4 & 0.25 \end{bmatrix}$, $(A + BK_0)^3 = \begin{bmatrix} 2^{-3} & 0 & 0 \\ 3 & 2^{-3} & 0 \\ 24 & 3 & 2^{-3} \end{bmatrix}$, $(A + BK_0)^4 = \begin{bmatrix} 2^{-4} & 0 & 0 \\ 2 & 2^{-4} & 0 \\ 24 & 2 & 2^{-4} \end{bmatrix}$, $(A+BK_0)^5 = \begin{bmatrix} 2^{-5} & 0 & 0 \\ 1.25 & 2^{-5} & 0 \\ 20 & 1.25 & 2^{-5} \end{bmatrix}$, $(A+BK_0)^6 = \begin{bmatrix} 2^{-6} & 0 & 0 \\ 0.75 & 2^{-6} & 0 \\ 15 & 0.75 & 2^{-6} \end{bmatrix}$. So although we have a controlled system with maximum eigenvalue 0.5, the power of $(A + BK_0)^k$ can still be very large in the bottom left corner for $k = 2, 3, 4, 5, 6$. Because of this, the randomness in the states is enlarged and propagated to several future steps. It turns out that, at the first 200 steps we used this high cost safety policy K_0 a lot as we do not have a good estimate of optimal controller K, and that is the real reason for this high burn-in period cost. As we will see later, if we change the stabilizing controller K_0 to be closer to the optimal K, the regret will be much smaller.

I.2.3 Comparison with Thompson Sampling

For comparison, we implement a straightforward version of Thompson sampling as follows. Denote $\Theta := [A, B]$. We use a prior of

$$\mathrm{vec}[\Theta_{prior}] \sim \mathcal{N}(\mathrm{vec}[\Theta], I_{n(n+d)}).$$



(a) Benefit of stepwise update

(b) Regret Ratio

(c) Differing Convergence Rates

(d) Confidence Region Coverage

(e) Prediction Region Coverage

Figure I.1: Summary of 1000 independent experiments applying Algorithm 1 with $\beta = 0.5$, $\alpha = 2$, $C_x = 1$, $C_K = 5$, and $K_0 = -\begin{bmatrix}1.5 & 0 & 0\\ 3.5 & 1.5 & 0\\ 0 & 3.5 & 1.5\end{bmatrix}$ to the unstable system described in Appendix I.1.2. (a) Difference between the regret of Algorithm 1 using stepwise and logarithmic updates. (b) The ratio of the empirical regret and our parametric or observable expressions for the regret. (c) The average Frobenius norm of various estimation errors considered in this paper, with slopes fitted on a log-log scale so that the estimation error is $\tilde{\mathcal{O}}(t^{\text{slope}})$. The effect of α was removed from the slopes of $\hat{K}_t - K$ and $[\hat{A}_t - A, \hat{B}_t - B]$ by dividing the error by $\log^{\alpha/2}(t)$. (d) Coverage of our 95% confidence regions for $[A, B]$, K, and $\mathbb{E}[x_{t+1} \,|\, \{x_i, u_i\}_{i=0}^t] = Ax_t + Bu_t$. (e) Coverage of our 95% prediction region for $x_{t+1} \,|\, \{x_i, u_i\}_{i=0}^t$, along with coverage of the naive prediction region given in Eq. 17.

Using the Bayesian updating equations and denoting the least-squares estimate of Θ by $\hat{\Theta}_t = [\hat{A}_t, \hat{B}_t]$, the posterior at time t is given by

$$\operatorname{vec}[\Theta_t^{TS}] \sim \mathcal{N}\Bigg(\operatorname{vec}\left[\left(\Theta + \hat{\Theta}_t \sum_{i=0}^{t-1} \begin{bmatrix}x_i\\ u_i\end{bmatrix}\begin{bmatrix}x_i\\ u_i\end{bmatrix}^\top\right)\left(I_{n+d} + \sum_{i=0}^{t-1}\begin{bmatrix}x_i\\ u_i\end{bmatrix}\begin{bmatrix}x_i\\ u_i\end{bmatrix}^\top\right)^{-1}\right], \\ \left(I_{n+d} + \sum_{i=0}^{t-1}\begin{bmatrix}x_i\\ u_i\end{bmatrix}\begin{bmatrix}x_i\\ u_i\end{bmatrix}^\top\right)^{-1} \otimes I_n\Bigg).$$

At each step, we draw a sample Θ_t^{TS} from this posterior and use it as the input to the DARE for calculating $\hat{K}_t$. Since a system is stabilizable if $\operatorname{rank}([A - \lambda I, B]) = n$ for any eigenvalue λ of A (Hautus, 1970), the Gaussian posterior puts probability 1 on stabilizable $\Theta = [A, B]$ and hence defines a unique solution to the DARE with probability 1 as well.

We report the Thompson sampling regret in subplot (b) of Fig. I.2, and see that it also suffers from rapidly increasing regret at early time points.f

I.2.4 Improved Regret When Using ‘Good’ K_0

When we switch from the ‘bad’ stabilizing controller to the ‘good’ one specified in Appendix I.1.2 as $K_0 = -\begin{bmatrix}1.5 & 0 & 0\\ 3.5 & 1.5 & 0\\ 0 & 3.5 & 1.5\end{bmatrix}$, we get that $A + BK_0 = \begin{bmatrix}0.5 & 0 & 0\\ 0.5 & 0.5 & 0\\ 0 & 0.5 & 0.5\end{bmatrix}$, which is a much better starting point than the previous $\begin{bmatrix}0.5 & 0 & 0\\ 4 & 0.5 & 0\\ 0 & 4 & 0.5\end{bmatrix}$, and the regret in this setting is indeed much better (see subplot (c) of Fig. I.2) and resembles that of the stable system described in Appendix I.1.1.

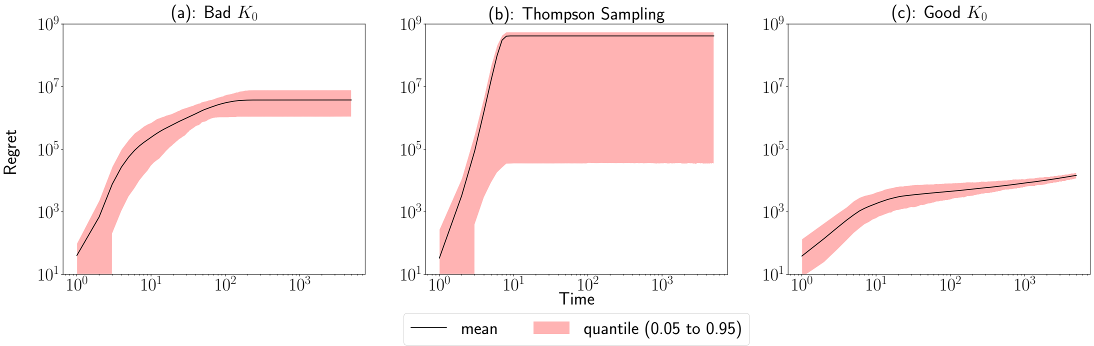


Figure I.2: Regret on the log scale based on 1000 independent experiments on the unstable system for $\beta = 0.5$ and $\alpha = 0$. (a): Bad safety controller $K_0 = -\begin{bmatrix} 1.5 & 0 & 0 \\ 0 & 1.5 & 0 \\ 0 & 0 & 1.5 \end{bmatrix}$; (b): Thompson Sampling; (c): Good safety controller $K_0 = -\begin{bmatrix} 1.5 & 0 & 0 \\ 3.5 & 1.5 & 0 \\ 0 & 3.5 & 1.5 \end{bmatrix}$.

I.3 Choices of β other than 0.5

Our simulations consider choices of β beyond 0.5 and even beyond those covered by our theory. In particular, we consider $\beta = 0.1, 0.3, 0.5, 0.7, 0.9$ and observe promising evidence that some of our asymptotic coverage results may generalize to the setting of $\beta < 1/2$.

I.3.1 Regret

According to Theorem 4 the dominating term for regret should be $T^{\beta} \log^{\alpha}(T)\, \mathbf{Tr}((B^{\top}PB + R)\frac{\tau^2}{\beta})$ for any $\beta \in [1/2, 1)$ and $\max\{\beta, \alpha - 1\} > 1/2$, and that indeed matches with our experimental results (see Fig. I.3). The asymptotic regret expression from Theorem 4 is represented as the black solid curve, which converges to the empirical regret for $\beta > 0.5$, but not $\beta < 0.5$.

I.3.2 Confidence region coverage

Fig. I.4 shows that the finite sample coverage of our confidence regions and prediction region closely matches the asymptotic theory from Corollary 11, Corollary 12 and Corollary 13 for any choice among $\beta = 0.1, 0.3, 0.5, 0.7, 0.9$, with the exception of confidence regions for K, which seem to only work for the $\beta \geq 0.5$ covered by our theory.

I.4 Algorithm design

We now investigate how the details of Algorithm 1 (the stabilizing controller K_0 and the thresholds on x_t and $\|\hat{K}_t\|$) impact the regret.

The threshold $C_{x,t}$ controls extreme tail behavior Although we only trigger the threshold $C_{x,t}$ rarely, without it we can see some extreme behavior with low probability. In particular, when this threshold constraint is removed, we occasionally observe very large

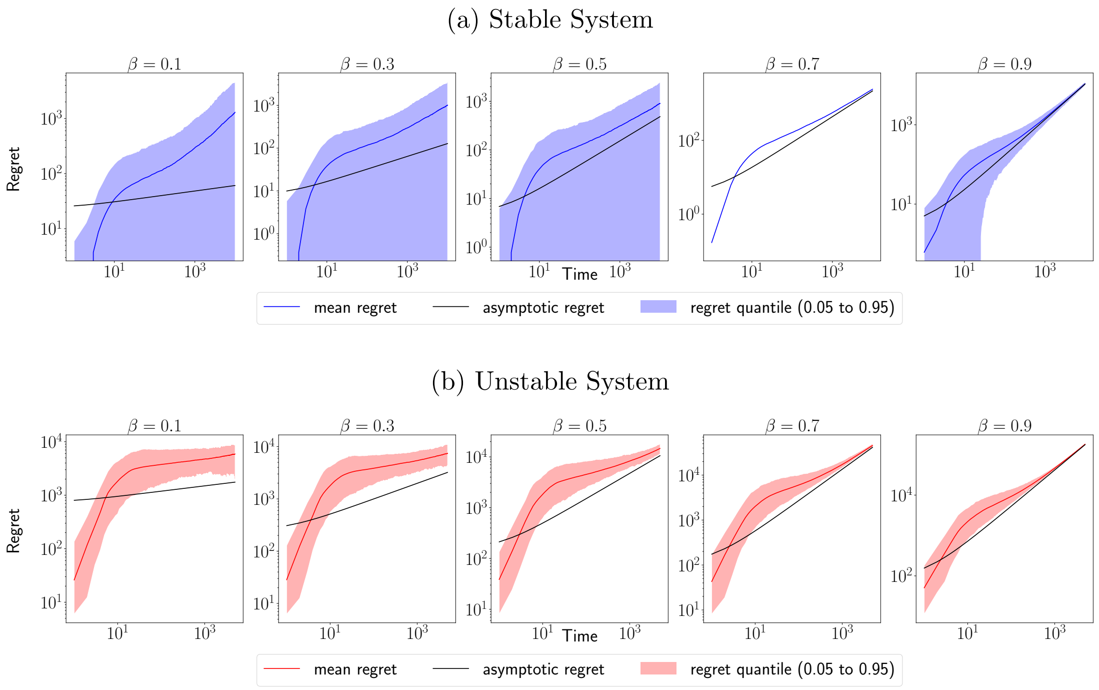
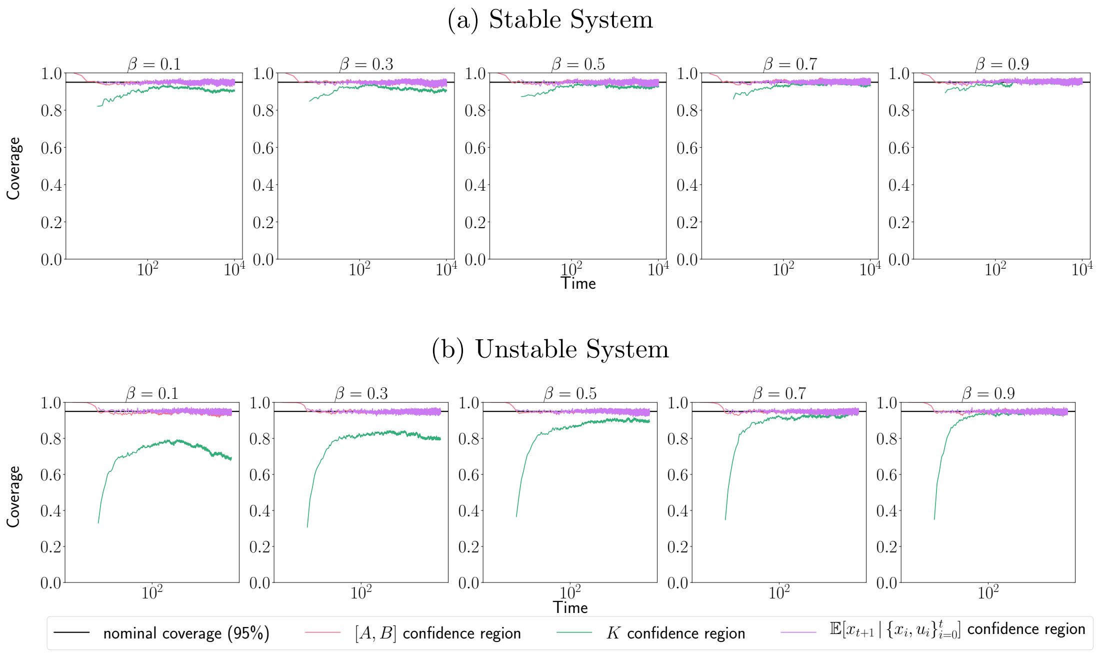


Figure I.3: Regret on the log scale based on 1000 independent experiments for $\beta = 0.1, 0.3, 0.5, 0.7, 0.9$ and $\alpha = 0$. (a): stable system; (b): unstable system.

regret in early time steps due to the poor estimate $\hat{K}_t$, which causes instability of the system (see Fig. I.5 and compare it to the purple line and shaded region in Fig. I.6). The mean value is even higher than the 0.95 quantile curve because of several extremely large regrets induced by the unstable closed-loop system. And compared to when $C_{x,t}$ is used in Fig. I.6, the 0.95 quantile when $C_{x,t}$ is not used is considerably higher, although its median is quite similar to the mean when $C_{x,t}$ is used.

Stepwise updating improves regret over logarithmic updating As our theory provides guarantees for Algorithm 1 with both stepwise and logarithmic updating, we run experiments to compare the regret of these two choices. Figs. 1c and I.1c show the difference in regret between Algorithm 1 and the same algorithm but that only updates its estimates of the system parameters logarithmically often, i.e., at times $t = 1, 2, 4, 8, \ldots$ On average, we see a steady logarithmic increase in regret from switching from stepwise updates to logarithmic frequency.

A stabilizing controller K_0 closer to K improves performance Although K_0 is a stabilizing controller by assumption, bad choices of K_0 can still make $(A + BK_0)^k$ large for some finite k (see Appendix I.2.2 for a concrete example). Thus, unsurprisingly, choosing K_0 to be as near as possible to the optimal controller K produces smaller regret, as evidenced by Fig. I.2.


(a) Stable System

(b) Unstable System

Figure I.4: Coverage on the log scale based on 1000 independent experiments for $\beta = 0.1, 0.3, 0.5, 0.7, 0.9$ and $\alpha = 0$. (a): stable system; (b): unstable system.

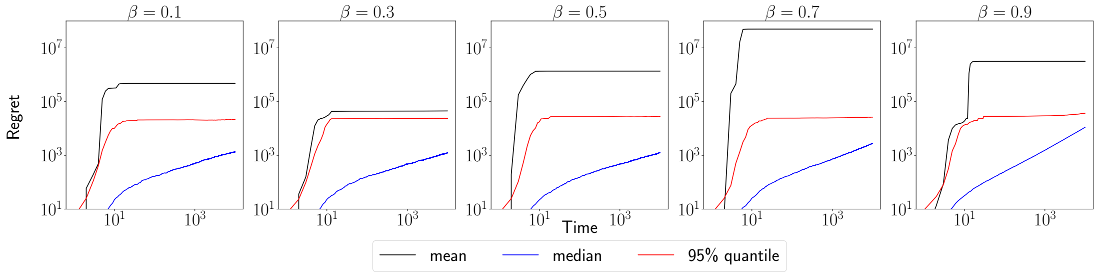


Figure I.5: Regret on the log scale with no $C_{x,t}$ threshold on $\|x_t\|$ based on 1000 independent experiments on stable system for $\beta = 0.1, 0.3, 0.5, 0.7, 0.9$ with $C_K = 5$ and $\alpha = 0$.

Regret is robust to conservative choices of C_K To check the sensitivity of the choice of $C_K = 5$ in the stable system, we also tried a looser bound $C_K = 1000$. We found that the norm of $\hat{K}_t$ never surpassed the $C_K = 1000$ bound. This larger C_K made little difference for settings covered by our theory ($\beta \geq 0.5$), and surprisingly seems to actually improve the regret for smaller β (see Fig. I.6).

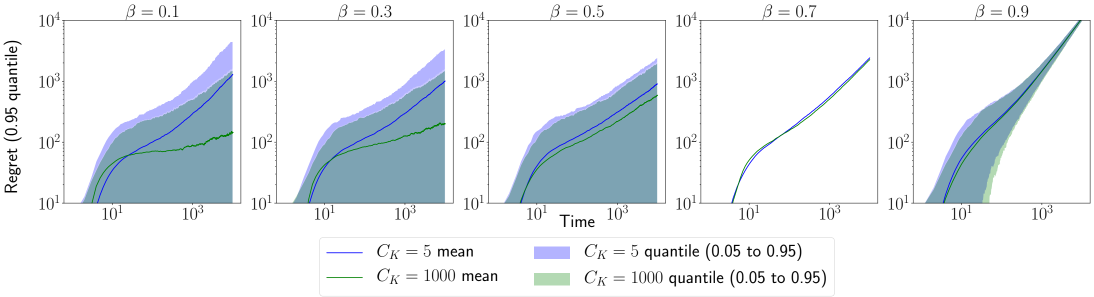


Figure I.6: Regret on the log scale based on 1000 independent experiments on stable system for $\beta = 0.1, 0.3, 0.5, 0.7, 0.9$ with $\alpha = 0$ comparing $C_K = 5$ and $C_K = 1000$.

References

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

Yasin Abbasi-Yadkori, Dávid Pál, and Csaba Szepesvári. Online least squares estimation with self-normalized processes: An application to bandit problems. *arXiv preprint arXiv:1102.2670*, 2011.

Marc Abeille and Alessandro Lazaric. Thompson sampling for linear-quadratic control problems. In *AISTATS 2017-20th International Conference on Artificial Intelligence and Statistics*, 2017.

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

Theodore W Anderson and Naoto Kunitomo. Asymptotic distributions of regression and autoregression coefficients with martingale difference disturbances. *Journal of Multivariate Analysis*, 40(2):221–243, 1992.

William F Arnold and Alan J Laub. Generalized eigenproblem algorithms and software for algebraic riccati equations. *Proceedings of the IEEE*, 72(12):1746–1754, 1984.

Karl Johan Åström and Björn Wittenmark. On self tuning regulators. *Automatica*, 9(2): 185–199, 1973.

Peter Auer, Thomas Jaksch, and Ronald Ortner. Near-optimal regret bounds for reinforcement learning. In *Advances in neural information processing systems*, pages 89–96, 2009.

Arthur Becker, P Kumar, and Ching-Zong Wei. Adaptive control with the stochastic approximation algorithm: Geometry and convergence. *IEEE Transactions on Automatic Control*, 30(4):330–338, 1985.

Felix Berkenkamp, Matteo Turchetta, Angela Schoellig, and Andreas Krause. Safe model-based reinforcement learning with stability guarantees. In *Advances in neural information processing systems*, pages 908–918, 2017.

Xavier Bombois, Gérard Scorletti, Michel Gevers, Paul MJ Van den Hof, and Roland Hildebrand. Least costly identification experiment for control. *Automatica*, 42(10):1651–1662, 2006.

Asaf Cassel, Alon Cohen, and Tomer Koren. Logarithmic regret for learning linear quadratic regulators efficiently. *arXiv preprint arXiv:2002.08095*, 2020.

Jae Weon Choi and Young Bong Seo. Lqr design with eigenstructure assignment capability [and application to aircraft flight control]. *IEEE Transactions on Aerospace and Electronic Systems*, 35(2):700–708, 1999.

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

Bruno C Da Silva, Eduardo W Basso, Ana LC Bazzan, and Paulo M Engel. Dealing with non-stationary environments using context detection. In *Proceedings of the 23rd international conference on Machine learning*, pages 217–224, 2006.

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

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

Yaakov Engel, Shie Mannor, and Ron Meir. The kernel recursive least-squares algorithm. *IEEE Transactions on signal processing*, 52(8):2275–2285, 2004.

Mohamad Kazem Shirani Faradonbeh, Ambuj Tewari, and George Michailidis. Finite time analysis of optimal adaptive policies for linear-quadratic systems. *arXiv preprint arXiv:1711.07230*, 2017.

Mohamad Kazem Shirani Faradonbeh, Ambuj Tewari, and George Michailidis. Input perturbations for adaptive regulation and learning,". *arXiv preprint arXiv:1811.04258*, 2018a.

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

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

Dylan J Foster and Max Simchowitz. Logarithmic regret for adversarial online control. *arXiv preprint arXiv:2003.00189*, 2020.

Dylan J Foster, Alexander Rakhlin, and Tuhin Sarkar. Learning nonlinear dynamical systems from a single trajectory. *arXiv preprint arXiv:2004.14681*, 2020.

László Gerencsér, Håkan Hjalmarsson, and Jonas Mårtensson. Identification of arx systems with non-stationary inputs—asymptotic analysis with application to adaptive input design. *Automatica*, 45(3):623–633, 2009.

L. Gerencsér, H. Hjalmarsson, and L. Huang. Adaptive input design for lti systems. *IEEE Transactions on Automatic Control*, 62(5):2390–2405, 2017. doi: 10.1109/TAC.2016.2612946.

Peter Grünwald, Rianne de Heide, and Wouter Koolen. Safe testing. *arXiv preprint arXiv:1906.07801*, 2019.

Lei Guo. Convergence and logarithm laws of self-tuning regulators. *Automatica*, 31(3): 435–450, 1995.

Lei Guo and Han-Fu Chen. The astrom-wittenmark self-tuning regulator revisited and els-based adaptive trackers. *IEEE Transactions on Automatic Control*, 36(7):802–812, 1991.

MLJ Hautus. Stabilization controllability and observability of linear autonomous systems. In *Indagationes mathematicae (proceedings)*, volume 73, pages 448–455. North-Holland, 1970.

Håkan Hjalmarsson. System identification of complex and structured systems. In *2009 European Control Conference (ECC)*, pages 3424–3452. IEEE, 2009.

L. Huang, H. Hjalmarsson, and L. Gerencsér. Adaptive experiment design for armax systems? In *2012 IEEE 51st IEEE Conference on Decision and Control (CDC)*, pages 907–912, 2012. doi: 10.1109/CDC.2012.6425920.

Morteza Ibrahimi, Adel Javanmard, and Benjamin V Roy. Efficient reinforcement learning for high dimensional linear quadratic systems. In *Advances in Neural Information Processing Systems*, pages 2636–2644, 2012.

K. Jamieson, Scribes Atinuke Ademola-Idowu, and Y. Shi. Lecture 20 : Linear dynamics and lqg 3 2 linear system optimal control 2 . 1 linear quadratic regulator (lqr) : Discrete-time finite horizon. 2018.

Svante Janson. Probability asymptotics: notes on notation. *arXiv preprint arXiv:1108.3924*, 2011.

Mohammad Khosravi and Roy S Smith. Nonlinear system identification with prior knowledge of the region of attraction. *arXiv preprint arXiv:2003.12330*, 2020.

Robert Kohn and Craig F Ansley. Prediction mean squared error for state space models with estimated parameters. *Biometrika*, 73(2):467–473, 1986.

Torsten Koller, Felix Berkenkamp, Matteo Turchetta, and Andreas Krause. Learning-based model predictive control for safe exploration. In *2018 IEEE Conference on Decision and Control (CDC)*, pages 6059–6066. IEEE, 2018.

TL Lai. Asymptotically efficient adaptive control in stochastic regression models. *Advances in Applied Mathematics*, 7(1):23–45, 1986.

TL Lai and Herbert Robbins. Iterated least squares in multiperiod control. *Advances in Applied Mathematics*, 3(1):50–73, 1982.

Tze Lai and Ching-Zong Wei. Extended least squares and their applications to adaptive control and prediction in linear systems. *IEEE Transactions on Automatic Control*, 31 (10):898–906, 1986.

Tze Leung Lai. Sequential analysis: some classical problems and new challenges. *Statistica Sinica*, pages 303–351, 2001.

Tze Leung Lai and Ching Zong Wei. Least squares estimates in stochastic regression models with applications to identification and control of dynamic systems. *The Annals of Statistics*, 10(1):154–166, 1982.

Sahin Lale, Kamyar Azizzadenesheli, Babak Hassibi, and Anima Anandkumar. Logarithmic regret bound in partially observable linear dynamical systems. *arXiv preprint arXiv:2003.11227*, 2020.

Lennart Ljung. *System Identification: Theory for the User*. Pearson, 2nd edition, 1997.

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

Yi Ouyang, Mukul Gagrani, and Rahul Jain. Learning-based control of unknown linear systems with thompson sampling. *arXiv preprint arXiv:1709.04047*, 2017.

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

Sindhu Padakandla, Shalabh Bhatnagar, et al. Reinforcement learning in non-stationary environments. *arXiv preprint arXiv:1905.03970*, 2019.

H Payne and L Silverman. On the discrete time algebraic riccati equation. *IEEE Transactions on Automatic Control*, 18(3):226–234, 1973.

Peter Pedroni. Panel cointegration: asymptotic and finite sample properties of pooled time series tests with an application to the ppp hypothesis. *Econometric theory*, 20(3):597–625, 2004.

M Cody Priess, Richard Conway, Jongeun Choi, John M Popovich, and Clark Radcliffe. Solutions to the inverse lqr problem with application to biological systems analysis. *IEEE Transactions on control systems technology*, 23(2):770–777, 2014.

Tuhin Sarkar, Alexander Rakhlin, and Munther A Dahleh. Finite-time system identification for partially observed lti systems of unknown order. *arXiv preprint arXiv:1902.01848*, 2019.

Yahya Sattar and Samet Oymak. Non-asymptotic and accurate learning of nonlinear dynamical systems. *arXiv preprint arXiv:2002.08538*, 2020.

Karam Shabaani and Mahdi Jalili-Kharaajoo. Application of adaptive lqr with repetitive control for ups systems. In *Proceedings of 2003 IEEE Conference on Control Applications, 2003. CCA 2003.*, volume 2, pages 1124–1129. IEEE, 2003.

David Silver, Julian Schrittwieser, Karen Simonyan, Ioannis Antonoglou, Aja Huang, Arthur Guez, Thomas Hubert, Lucas Baker, Matthew Lai, Adrian Bolton, et al. Mastering the game of go without human knowledge. *Nature*, 550(7676):354–359, 2017.

Max Simchowitz and Dylan J Foster. Naive exploration is optimal for online lqr. *arXiv preprint arXiv:2001.09576*, 2020.

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

Vladimir Stojanovic and Vojislav Filipovic. Adaptive input design for identification of output error model with constrained output. *Circuits, Systems, and Signal Processing*, 33(1):97–113, 2014.

Vladimir Stojanovic, Novak Nedic, Dragan Prsic, and Ljubisa Dubonjic. Optimal experiment design for identification of arx models with constrained output in non-gaussian noise. *Applied Mathematical Modelling*, 40(13-14):6676–6689, 2016.

Yue Sun, Samet Oymak, and Maryam Fazel. Finite sample system identification: Optimal rates and the role of regularization. In *Learning for Dynamics and Control*, pages 16–25. PMLR, 2020.

Anastasios Tsiamis and George Pappas. Online learning of the kalman filter with logarithmic regret. *arXiv preprint arXiv:2002.05141*, 2020.

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

Oriol Vinyals, Timo Ewalds, Sergey Bartunov, Petko Georgiev, Alexander Sasha Vezhnevets, Michelle Yeo, Alireza Makhzani, Heinrich Küttler, John Agapiou, Julian Schrittwieser, et al. Starcraft ii: A new challenge for reinforcement learning. *arXiv preprint arXiv:1708.04782*, 2017.

Bo Wahlberg, Håkan Hjalmarsson, and Mariette Annergren. On optimal input design in system identification for control. In *49th IEEE Conference on Decision and Control (CDC)*, pages 5548–5553. IEEE, 2010.

Hongyi Wang and Dianlong You. Online streaming feature selection via multi-conditional independence and mutual information entropy. *International Journal of Computational Intelligence Systems*, 2020.

Yang Zheng and Na Li. Non-asymptotic identification of linear dynamical systems using multiple trajectories. *arXiv preprint arXiv:2009.00739*, 2020.