# Identification and Adaptive Control of Markov Jump Systems: Sample Complexity and Regret Bounds

Yahya Sattar*† Zhe Du*‡ Davoud Ataee Tarzanagh‡ Necmiye Ozay‡
Laura Balzano‡ Samet Oymak†

**Abstract**

Learning how to effectively control unknown dynamical systems is crucial for intelligent autonomous systems. This task becomes a significant challenge when the underlying dynamics are changing with time. Motivated by this challenge, this paper considers the problem of controlling an unknown Markov jump linear system (MJS) to optimize a quadratic objective. By taking a model-based perspective, we consider identification-based adaptive control for MJSs. We first provide a system identification algorithm for MJS to learn the dynamics in each mode as well as the Markov transition matrix, underlying the evolution of the mode switches, from a single trajectory of the system states, inputs, and modes. Through mixing-time arguments, sample complexity of this algorithm is shown to be $\mathcal{O}(1/\sqrt{T})$. We then propose an adaptive control scheme that performs system identification together with certainty equivalent control to adapt the controllers in an episodic fashion. Combining our sample complexity results with recent perturbation results for certainty equivalent control, we prove that when the episode lengths are appropriately chosen, the proposed adaptive control scheme achieves $\mathcal{O}(\sqrt{T})$ regret, which can be improved to $\mathcal{O}(\text{polylog}(T))$ with partial knowledge of the system. Our proof strategy introduces innovations to handle Markovian jumps and a weaker notion of stability common in MJSs. Our analysis provides insights into system theoretic quantities that affect learning accuracy and control performance. Numerical simulations are presented to further reinforce these insights.

## 1 Introduction

A canonical problem at the intersection of machine learning and control is that of adaptive control of an unknown dynamical system. An intelligent autonomous system is likely to encounter such a task; from an observation of the inputs and outputs, it needs to both learn and effectively control the dynamics. A commonly used control paradigm is the Linear Quadratic Regulator (LQR), which is theoretically well understood when system dynamics are linear and known. LQR also provides an interesting benchmark, when system dynamics are unknown, for reinforcement learning (RL) with continuous state and action spaces and for adaptive control [2, 4, 9, 16, 36, 47].

A generalization of linear dynamical systems called Markov jump linear systems (MJSs) models dynamics that switch between multiple linear systems, called modes, according to an underlying finite Markov chain. MJS allows for modeling a richer set of problems where the underlying dynamics can abruptly change over time. One can, similarly, generalize the LQR paradigm to MJS by using mode-dependent cost matrices, which allow different control goals under different modes. While the MJS-LQR problem is also well understood when one has perfect knowledge of the system dynamics [12, 14], in practice, it is not always possible to have a perfect knowledge of the system dynamics and the Markov transition matrix. For instance, a Mars rover optimally exploring an unknown heterogeneous terrain, optimal solar power generation on a cloudy day, or

*Equal contribution.
†Dept. of Electrical and Computer Engineering, Univ. of California, Riverside. Email: {ysatt001, soymak}@ucr.edu.
‡Dept. of Electrical Engineering and Computer Science, Univ. of Michigan Ann Arbor. Email: {zhedu, tarzanaq, necmiye, girasole}@umich.edu.

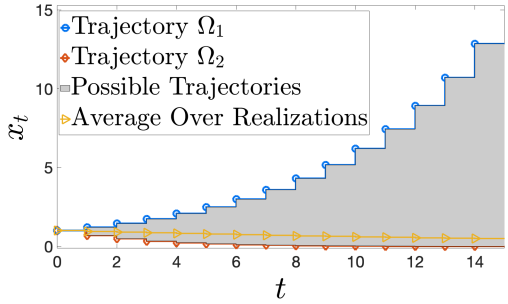



Figure 1: State trajectories for a two-modes MJS $\begin{cases} x_{t+1} = 1.2x_t \\ x_{t+1} = 0.7x_t \end{cases}$ with Markov matrix $\begin{bmatrix} 0.6 & 0.4 \\ 0.3 & 0.7 \end{bmatrix}$ and $\mathbf{x}_0 = 1$. Red and blue curves: mode switching sequences $\Omega_1 = \{1, 1, \dots\}$ and $\Omega_2 = \{2, 2, \dots\}$. Yellow curve: average over all realizations. Gray area: region for all possible trajectories.

controlling investments in financial markets may be modeled as MJS-LQR problems with unknown system dynamics [6, 8, 46, 60, 64]. Earlier works have aimed at analyzing the asymptotic properties (i.e., stability) of adaptive controllers for unknown MJSs both in continuous-time [7] and discrete-time [66] settings, however, despite the practical importance of MJSs, non-asymptotic sample complexity results and regret analysis for MJSs are lacking. The high-level challenge here is the hybrid nature of the problem that requires consideration of both the system dynamics and the underlying Markov transition matrix. A related challenge is that, typically, the stability of MJS is understood only in the *mean-square sense*. This is in stark contrast to the deterministic stability (e.g., as in LQR), where the system is guaranteed to converge towards an equilibrium point in the absence of noise. In contrast, the convergence of MJS trajectories towards an equilibrium depends heavily on how the switching between modes occurs. Figure 1 shows an example (adapted from [14]) of an MJS that is stable in the mean-square sense despite having an unstable mode. Clearly, under an unfavorable mode switching sequence, the system trajectory can still blow up. High-probability light tail bounds are therefore not applicable without very strong assumptions on the joint spectral radius of different modes (cf. [56]). Perhaps more surprisingly, there are examples of MJS with all modes individually stable, however due to switching, the system exhibits an unstable behavior on average, and the MJS is not mean-square stable (see Example 3.17 of [14]). Therefore, finding controllers to individually stabilize the mode dynamics does not guarantee that overall system will be stable when mode switches over time. This more relaxed notion of *mean-square stability* presents major challenges in learning, controlling, and statistical analysis.

**Contributions:** In this paper, we provide the first comprehensive system identification and regret guarantees for learning and controlling Markov jump linear systems using a single trajectory while assuming only mean-square stability (see Def. 3.1). Importantly, our guarantees are optimal in the trajectory length $T$. Specifically, our contributions are as follows[1]:

- **System identification:** For an MJS with $s$ modes, the system dynamics involve a Markov transition matrix $\mathsf{T} \in \mathsf{R}^{s \times s}$ and $s$ state-input matrix pairs $(\mathsf{A}_i, \mathsf{B}_i)_{i=1}^s$. We provide an algorithm (Alg. 1) to estimate these dynamics with an error rate of $\mathcal{O}((n+p)\log(T)\sqrt{s/T})$, where $n$ and $p$ are the state and input dimensions respectively, and the $\mathcal{O}(1/\sqrt{T})$ dependence on the trajectory length $T$ is optimal.
- **$\mathcal{O}(\sqrt{T})$-regret bound:** We employ our system identification guarantees for the MJS-LQR. When the system dynamics are unknown, we show that the certainty-equivalent adaptive MJS-LQR Algorithm (Alg. 2) achieves a regret bound of $\mathcal{O}(\sqrt{T})$. Remarkably, this coincides with the optimal regret bound for the standard LQR problem obtained via certainty equivalence [47].
- **$\mathcal{O}(\mathsf{polylog}(T))$-regret with partial knowledge:** We also consider the practically relevant setting where the state matrices are unknown but the input matrices are known. We show that the regret bound can be significantly improved to $\mathcal{O}(\mathrm{polylog}(T))$. This bound also coincides with the polylogarithmic regret bound for the standard LQR with the knowledge of the input matrix $\mathsf{B}$ [10].

[1] orders of magnitude here are up to polylogarithmic factors

**Technical tools:** Besides these key contributions to MJS control, our proof strategy introduces multiple innovations. To address Markovian mode transitions, we introduce a mixing-time argument to jointly track the approximate-dependence across the states and the modes. This in turn helps ensure each mode has sufficient samples and these samples are sufficiently informative. Secondly, as clarified further below, due to mean-square stability and mode transitions, it becomes non-trivial to determine whether the states have a light-tailed distribution (e.g., sub-gaussian or sub-exponential). To circumvent this, we develop intricate system identification arguments that allow for heavy-tailed states. Such arguments can potentially benefit other RL problems with heavy-tailed data.

## 2 Related Work

Our work is related to several topics in model-based reinforcement learning, system identification, and adaptive control. A comparison with the related works, in the LQR setting, is provided in Table 1.

• **System Identification:** Learning dynamical models has a long history in the control community, with major theoretical results being related to asymptotic properties under strong assumptions on persistence of excitation [45]. The problem becomes harder for hybrid and switched systems where the initial focus was on computational complexity as opposed to sample complexity of learning [40, 53]. There are some recent results on asymptotic consistency [30] in the switched system setting, a special case of MJS where the modes change in an independently and identically distributed manner. Identification of MJS has also attracted attention from different communities in the case when mode sequence is hidden [25, 63].

• **Sample Complexity of System Identification:** There is a recent surge of interest towards understanding the sample complexity of learning linear dynamical systems from a single trajectory under mild assumptions [52], using statistical tools like martingales [55, 59, 61] or mixing-time arguments [35, 50]. Recently, [34] provides precise rates for the finite-time identification of LTI systems using a single trajectory. The literature gets scarcer for switched systems. In [38], a novel approach based on Lyapunov equation is proposed for systems with stochastic switches, but no theory is built. [56] is one of the early works – and it seems to be the only work not assuming persistence of excitation – to provide finite sample analysis for learning systems with stochastic switches, yet with additional strong assumptions like independent switches and small joint spectral radius. The proof techniques developed within our work aim to obviate such assumptions. Our paper tackles the open problem of learning MJS from finite samples, obtained from a single trajectory, with theoretical guarantees under mild assumptions.

• **Learning-based Control and Regret Analysis:** As a direct application of single-trajectory system identification results, one can provide more sophisticated adaptive control guarantees from regret perspective [1, 2, 16, 23, 29, 47]. Specifically, [58] achieves $\mathcal{O}(\sqrt{T})$ regret lower bound for adaptive LQR control with unknown system dynamics, while with partial knowledge of the system [10] or persistence of excitation assumptions [37], one can achieve logarithmic regret [10, 37], as no additional excitation noise is needed to guarantee learnability of the system. However, in the MJS setting, due to the lack of well established identification analysis, prior works provide guarantees [7, 66] from the stability aspect. The case of input design without system state dynamics is considered in [5], which can be thought of as a generalization of linear bandits to have a Markovian structure in the reward function without any continuous dynamic structure. However, only a regret lower bound is provided in [5]. Finally, we refer the reader to the survey papers [28, 48, 54] for a broad overview of the recent developments on non-asymptotic system identification, adaptive control and reinforcement learning from the perspective of optimization and control.

• **Model-free Approaches:** Somehow orthogonal to the above developments, but still highly relevant, are approaches that sidestep system identification and try to learn an optimal controller (policy) directly (among many others, see e.g., [24, 49, 68, 69]). These works analyze the optimization landscape of LQR and related optimal control problems and provide polynomial-time algorithms that lead to a globally convergent search in the space of controllers. Importantly, these optimization algorithms do not require the knowledge of the system parameters as long as relevant quantities like gradients can be approximated from simulated system trajectories. More recently, this line of work is extended to MJSs in [33], significantly expanding their utility. However, these works require multiple trajectories to estimate the gradients as opposed to a controller that

adapts at run-time, therefore, they provide a complementary perspective to the single trajectory adaptation and regret analysis in our work.

A preliminary version of this work has been submitted to the American Control Conference 2022 [19], where we provide preliminary guarantees for the data-driven control of MJS. In contrast to this paper, Algorithm 1 in [19] performs a fairly-sophisticated double-subsampling to estimate the unknown MJS dynamics $(A_i, B_i)_{i=1}^s$ and $T$ with guarantees. On the other hand, Algorithm 1 in this paper uses all of the bounded samples from an MJS trajectory to estimate the unknown MJS dynamics. In this paper, we provide new and substantially improved results for the adaptive control of MJS with unknown state matrices but the input matrices are known. We also analyze the impact of "mean-square stability" (see Def. 3.1) on our regret bounds by replacing it with "uniform stability" (see Sec. 5.2). Hence, this paper also provides regret bounds for the adaptive control of MJS under uniform stability assumption. Furthermore, it also provides the necessary technical framework and the associated proofs. Lastly, compared to [19], the exposition of this paper is significantly improved by adding detailed discussions on the design of initial stabilizing controllers for the adaptive control of MJS and analyzing the sub-optimality gap for the offline control of MJS.

Table 1: Comparison with prior works in the LQR setting.

| Model | Reference | Regret | Computational Complexity | Cost | Stabilizability/ Controllability |
|---|---|---|---|---|---|
| LTI | [2] | $\sqrt{T}$ | Exponential | Strongly Convex | Controllable |
| | [32] | $\sqrt{T}$ | Exponential | Convex | Controllable |
| | [3] (one dim. systems) | $\sqrt{T}$ | Polynomial | Strongly Convex | Stabilizable |
| | [15] | $T^{2/3}$ | Polynomial | Convex | Stabilizable |
| | [47] | $\sqrt{T}$ | Polynomial | Strongly Convex | Controllable |
| | [13] | $\sqrt{T}$ | Polynomial | Strongly Convex | Strongly Stabilizable |
| | [22, 58] | $\sqrt{T}$ | Polynomial | Strongly Convex | Stabilizable |
| | [10] (known $A$ or $B$) | polylog$(T)$ | Polynomial | Strongly Convex | Strongly Stabilizable |
| MJS | Ours | $s^2\sqrt{T}$ | Polynomial | Strongly Convex | MSS |
| | Ours (known $B_{1:s}$) | $s^2$polylog$(T)$ | Polynomial | Strongly Convex | MSS |

# 3 Preliminaries and Problem Setup

**Notations:** We use boldface uppercase (lowercase) letters to denote matrices (vectors). For a matrix $V$, $\rho(V)$ denotes its spectral radius. We use $\|\cdot\|$ to denote the Euclidean norm of vectors as well as the spectral norm of matrices. Similarly, we use $\|\cdot\|_1$ to denote the $_1$-norm of a matrix/vector. The Kronecker product of two matrices $M$ and $N$ is denoted as $M \otimes N$. $V_{1:s}$ denotes a set of $s$ matrices $\{V_i\}_{i=1}^s$ of same dimensions. We define $[s] := \{1, 2, \ldots, s\}$ and $\|V_{1:s}\| := \max_{i\in[s]} \|V_i\|$. The $i$-th row or column of a matrix $M$ is denoted by $[M]_{i,:}$ or $[M]_{:,i}$ respectively. Orders of magnitude notation $\tilde{\mathcal{O}}(\cdot)$ hides $\log(\frac{1}{\delta})$ or $\log^2(\frac{1}{\delta})$ terms.

## 3.1 Markov Jump Linear Systems

In this paper we consider the identification and adaptive control of MJSs which are governed by the following state equation,

$$x_{t+1} = A_{\omega(t)} x_t + B_{\omega(t)} u_t + w_t \quad \text{s.t.} \qquad \omega(t) \sim \text{Markov Chain}(T), \tag{3.1}$$

where $x_t \in \mathbb{R}^n$, $u_t \in \mathbb{R}^p$ and $w_t \in \mathbb{R}^n$ are the state, input, and process noise of the MJS at time $t$ with $\{w_t\}_{t=0}^{\infty} \overset{\text{i.i.d.}}{\sim} \mathcal{N}(0, \sigma_w^2 I_n)$. There are $s$ modes in total, and the dynamics of mode $i$ is given by the state matrix $A_i$ and input matrix $B_i$. The active mode at time $t$ is indexed by $\omega(t) \in [s]$. Throughout, we assume the state $x_t$ and the mode $\omega(t)$ can be observed at time $t \geq 0$. The mode switching sequence $\{\omega(t)\}_{t=0}^{\infty}$ follows a Markov chain with transition matrix $T \in \mathbb{R}_+^{s\times s}$ such that for all $t \geq 0$, the $ij$-th element of $T$ denotes the

conditional probability $[\mathsf{T}]_{ij} := \mathsf{P}\big(\omega(t+1) = j \mid \omega(t) = i\big),\ \forall\ i, j \in [s]$. Throughout, we assume the initial state $\mathsf{x}_0$, the mode switching sequence $\{\omega(t)\}_{t=0}^{\infty}$, and the noise $\{\mathsf{w}_t\}_{t=0}^{\infty}$ are mutually independent. We use MJS($\mathsf{A}_{1:s}, \mathsf{B}_{1:s}, \mathsf{T}$) to refer to an MJS with state equation (3.1), parameterized by ($\mathsf{A}_{1:s}, \mathsf{B}_{1:s}, \mathsf{T}$).

For mode-dependent state-feedback controller $\mathsf{K}_{1:s}$ that yields the input $\mathsf{u}_t = \mathsf{K}_{\omega(t)}\mathsf{x}_t$, we use $\mathsf{L}_i := \mathsf{A}_i + \mathsf{B}_i\mathsf{K}_i$ to denote the closed-loop state matrix for mode $i$. We use $\mathsf{x}_{t+1} = \mathsf{L}_{\omega(t)}\mathsf{x}_t$ to denote the noise-free autonomous MJS, either open-loop ($\mathsf{L}_i = \mathsf{A}_i$) or closed-loop ($\mathsf{L}_i = \mathsf{A}_i + \mathsf{B}_i\mathsf{K}_i$). Due to the randomness in $\{\omega(t)\}_{t=0}^{\infty}$, it is common to consider the stability of MJS in the mean-square sense which is defined as follows.

**Definition 3.1 (Mean-square stability [14])** *We say MJS in (3.1) with $\mathsf{u}_t = 0$ is mean-square stable (MSS) if there exists $\mathsf{x}_\infty, \Sigma_\infty$ such that for any initial state $\mathsf{x}_0$ and mode $\omega(0)$, as $t \to \infty$, we have*

$$\|\mathsf{E}[\mathsf{x}_t] - \mathsf{x}_\infty\| \to 0, \quad \|\mathsf{E}[\mathsf{x}_t\mathsf{x}_t^\top] - \Sigma_\infty\| \to 0, \tag{3.2}$$

*where the expectation is over the Markovian mode switching sequence $\{\omega(t)\}_{t=0}^{\infty}$, the noise $\{\mathsf{w}_t\}_{t=0}^{\infty}$ and the initial state $\mathsf{x}_0$. In the noise-free case ($\mathsf{w}_t = 0$), we have $\mathsf{x}_\infty = 0$, $\Sigma_\infty = 0$. We say MJS in (3.1) with $\mathsf{w}_t = 0$ is (mean-square) stabilizable if there exists mode-dependent controller $\mathsf{K}_{1:s}$ such that the closed-loop MJS $\mathsf{x}_{t+1} = (\mathsf{A}_{\omega(t)} + \mathsf{B}_{\omega(t)}\mathsf{K}_{\omega(t)})\mathsf{x}_t$ is MSS. We call such $\mathsf{K}_{1:s}$ a stabilizing controller.*

The MSS of a noise-free autonomous MJS is related to the spectral radius of an augmented state matrix $\tilde{\mathsf{L}} \in \mathsf{R}^{sn^2 \times sn^2}$ with $ij$-th $n^2 \times n^2$ block given by $[\tilde{\mathsf{L}}]_{ij} := [\mathsf{T}]_{ji}\mathsf{L}_j \otimes \mathsf{L}_j$. As discussed in [14, Theorem 3.9], $\tilde{\mathsf{L}}$ can be viewed as the mapping from $\mathsf{E}[\mathsf{x}_t\mathsf{x}_t^\top]$ to $\mathsf{E}[\mathsf{x}_{t+1}\mathsf{x}_{t+1}^\top]$, thus a noise-free autonomous MJS is MSS if and only if $\rho(\tilde{\mathsf{L}}) < 1$. The analysis of this work highly depends on certain "mixing" of the MJS – the distributions of both state $\mathsf{x}_t$ and mode $\omega(t)$ can converge close enough to some stationary distributions within finite time, which is guaranteed by the following assumption.

**Assumption 1** *The MJS in (3.1) has ergodic Markov chain and is stabilizable.*

Ergodicity guarantees that the distribution of $\omega(t)$ converges to a unique strictly positive stationary distribution [27, Theorem 4.3.5]. Throughout, we let $\pi_\infty$ denote the stationary distribution of $\mathsf{T}$ and $\pi_{\min} := \min_i \pi_\infty(i)$. We further define the mixing time [43] of $\mathsf{T}$ as $t_{\mathsf{MC}} := \inf\big\{t \in \mathsf{N} : \max_{i \in [s]} \|([\mathsf{T}^t]_{i,:})^\top - \pi_\infty\|_1 \le 0.5\big\}$, to quantify its convergence rate. In our analysis, ergodicity and $t_{\mathsf{MC}}$ ensures that the MJS trajectory could have enough "visits" to every mode $i \in [s]$ thus providing us enough data to learn $[\mathsf{T}]_{i,:}$, $\mathsf{A}_i$ and $\mathsf{B}_i$. On the other hand, stability (or stabilizability) characterized by the spectral radius of $\tilde{\mathsf{L}}$ guarantees the convergence/mixing of $\mathsf{x}_t$, which allows us to obtain weakly dependent sub-trajectories from a single trajectory of MJS, upon which the sample complexity of learning the matrices $\mathsf{A}_{1:s}$ and $\mathsf{B}_{1:s}$ can be established.

## 3.2 Problem Formulation

In this work we consider two major problems under the MJS setting: System identification and adaptive control, with identification being the core part of adaptive control.

**(A) System Identification.** This problem seeks to estimate unknown system dynamics from data, i.e. from input-output trajectory(ies), when one has the flexibility to design the inputs so that the collected data has nice statistical properties. In the MJS setting, one needs to estimate both the state/input matrices $\mathsf{A}_{1:s}, \mathsf{B}_{1:s}$ for every mode $i \in [s]$ as well as the Markov transition matrix $\mathsf{T}$. In this work, we seek to estimate the MJS dynamics from a single trajectory of states, inputs and mode observations $\{\mathsf{x}_t, \mathsf{u}_t, \omega(t)\}_{t=0}^{T}$ and provide finite sample guarantees. As mentioned earlier, MJS presents unique statistical analysis challenges due to Markovian jumps and a weaker notion of stability. Section 4 presents our system identification guarantees overcoming these challenges. These guarantees are further integrated into model-based control for MJS-LQR in Section 5.

**(B) Online Linear Quadratic Regulator.** In this paper, we consider the following finite-horizon Markov jump system linear quadratic regulator (MJS-LQR) problem:

$$\begin{aligned} \inf_{\mathbf{u}_{0:T}} \quad & J(\mathbf{u}_{0:T}) := \sum_{t=0}^{T} \mathsf{E}\big[\mathsf{x}_t^\top \mathsf{Q}_{\omega(t)}\mathsf{x}_t + \mathsf{u}_t^\top \mathsf{R}_{\omega(t)}\mathsf{u}_t\big], \\ \text{s.t.} \quad & \mathsf{x}_t, \omega(t) \sim \text{MJS}(\mathsf{A}_{1:s}, \mathsf{B}_{1:s}, \mathsf{T}). \end{aligned} \tag{3.3}$$

**Algorithm 1:** MJS-SYSID

**Input:** A mean square stabilizing controller $\mathsf{K}_{1:s}$; process and exploration noise variances $\sigma_{\mathbf{w}}^2$ and $\sigma_{\mathbf{z}}^2$; MJS trajectory $\{\mathsf{x}_t, \mathsf{z}_t, \omega(t)\}_{t=0}^{T}$ generated using input $\mathsf{u}_t = \mathsf{K}_{\omega(t)}\mathsf{x}_t + \mathsf{z}_t$ with $\mathsf{z}_t \overset{\text{i.i.d.}}{\sim} \mathcal{N}(0, \sigma_{\mathbf{z}}^2 \mathsf{I}_p)$; and data clipping thresholds $c_{\mathbf{x}}$, $c_{\mathbf{z}}$.

**Estimate** $\mathsf{A}_{1:s}, \mathsf{B}_{1:s}$: **for all** modes $i \in [s]$ **do**

$S_i = \{t \mid \omega(t) = i, \quad \|\mathsf{x}_t\| \le c_{\mathbf{x}}\sigma_{\mathbf{w}}\sqrt{\log(T)}, \quad \|\mathsf{z}_t\| \le c_{\mathbf{z}}\sigma_{\mathbf{z}}\}$ // sub-sample data

$\hat{\Theta}_{1,i}, \hat{\Theta}_{2,i} = \arg\min_{\Theta_1, \Theta_2} \sum_{k \in S_i} \|\mathsf{x}_{k+1} - \Theta_1 \mathsf{x}_k/\sigma_{\mathbf{w}} - \Theta_2 \mathsf{z}_k/\sigma_{\mathbf{z}}\|^2$ // regress with data $S_i$

$\hat{\mathsf{B}}_i = \hat{\Theta}_{2,i}/\sigma_{\mathbf{z}}$, $\hat{\mathsf{A}}_i = (\hat{\Theta}_{1,i} - \hat{\mathsf{B}}_i \mathsf{K}_i)/\sigma_{\mathbf{w}}$

**Estimate** $\mathsf{T}$: $[\hat{\mathsf{T}}]_{ji} = \frac{\sum_{t=1}^{T} \mathbf{1}_{\{\omega(t)=i, \omega(t-1)=j\}}}{\sum_{t=1}^{T} \mathbf{1}_{\{\omega(t-1)=j\}}}$ //empirical frequency of transitions

**Output:** $\hat{\mathsf{A}}_{1:s}, \hat{\mathsf{B}}_{1:s}, \hat{\mathsf{T}}$

Here, the goal is to design control inputs to minimize the expected quadratic cost function composed of positive semi-definite cost matrices $\mathsf{Q}_{1:s}$ and $\mathsf{R}_{1:s}$ under the MJS dynamics (3.1). The quadratic cost incurred by the state $\mathsf{x}_t$ represents the deviation from target values, e.g. desired velocity, position, angle, etc., whereas, the quadratic term in $\mathsf{u}_t$ represents the control effort, e.g. energy consumption. The flexibility of having mode-dependent cost matrices allows one to design different control requirements or trade-offs under different circumstances. For the MJS-LQR problem (3.3), we assume the following.

**Assumption 2** *For all* $i \in [s]$, $\mathsf{R}_i > 0, \mathsf{Q}_i > 0$.

Assumptions 1 and 2 together guarantee the solvability of MJS-LQR when the dynamics are known [14, Corollary A.21]. In the remaining of the paper, we use MJS-LQR($\mathsf{A}_{1:s}, \mathsf{B}_{1:s}, \mathsf{T}, \mathsf{Q}_{1:s}, \mathsf{R}_{1:s}$) to denote MJS-LQR problem (3.3) composed of MJS($\mathsf{A}_{1:s}, \mathsf{B}_{1:s}, \mathsf{T}$) and cost matrices $\mathsf{Q}_{1:s}, \mathsf{R}_{1:s}$.

Recall our assumption that the states $\mathsf{x}_t$ and the modes $\omega(t)$ can be observed at time $t \ge 0$. With these observations, instead of a fixed and open-loop input sequence, one can design closed-loop policies that generate real-time control inputs based on the current observations, e.g. mode-dependent state-feedback controllers. When the dynamics $\mathsf{A}_{1:s}, \mathsf{B}_{1:s}, \mathsf{T}$ of the MJS are known, one can solve for the optimal controllers recursively via coupled discrete-time algebraic Riccati equations [14]. In this work, we assume the dynamics are unknown, and only the design parameters $\mathsf{Q}_{1:s}$ and $\mathsf{R}_{1:s}$ are known. Control schemes in this scenario are typically referred to as adaptive control, which usually involves procedures of learning, either the dynamics or directly the controllers. Adaptive control suffers additional costs as (i) the lack of the exact knowledge of the system and (ii) the exploration-exploitation trade-off – the necessity to sacrifice short-term input optimality to boost learning, so that overall long-term optimality can be improved.

Because of this, to evaluate the performance of an adaptive scheme, one is interested in the notion of regret – how much more cost it will incur if one could have applied the optimal controllers? In our setting, we compare the resulting cost against the optimal cost $T \cdot J^\star$ where $J^\star$ is the optimal infinite-horizon average cost

$$J^\star := \limsup_{T \to \infty} \frac{1}{T} \inf_{\mathbf{u}_{0:T}} J(\mathbf{u}_{0:T}), \tag{3.4}$$

i.e., if one applies the optimal controller for infinitely long, how much cost one would get on average for each single time step. Compared to the regret analysis of standard adaptive LQR problem [15], in MJS-LQR setting, the cost analysis requires additional consideration of Markov chain mixing, which is addressed in this paper.

# 4 System Identification for MJS

Our MJS identification procedure is given in Algorithm 1. We assume one has access to an initial stabilizing controller $\mathsf{K}_{1:s}$, which is a standard assumption in data-driven control [3, 13, 15, 32, 58] for LTI systems. For

MJSs, a thorough discussion on the validity of this assumption is provided in Section 6.1. Note that, if the open-loop MJS is already MSS, then one can simply set $\mathsf{K}_{1:s} = 0$ and carry out MJS identification. Given an MJS trajectory $\{\mathsf{x}_t, \mathsf{z}_t, \omega(t)\}_{t=0}^{T}$ generated using the input $\mathsf{u}_t = \mathsf{K}_{\omega(t)}\mathsf{x}_t + \mathsf{z}_t$ (where $\mathsf{z}_t \sim \mathcal{N}(0, \sigma_{\mathsf{z}}^2 \mathsf{I})$ is the exploration noise), we sub-sample it, for each mode $i \in [s]$, to obtain $s$ sub-trajectories with bounded states $\mathsf{x}_t$ and excitations $\mathsf{z}_t$. This sub-sampling is required because of the mean-square stability, which can at most guarantee that the states are bounded in expectation. As a result of sub-sampling only bounded states/excitations, we obtain samples with manageable distributional properties. After appropriate scaling, we regress over these samples to obtain the estimates $\hat{\mathsf{A}}_i, \hat{\mathsf{B}}_i$ for each $i \in [s]$. Lastly, using the empirical frequency of observed modes, we obtain the estimate $\hat{\mathsf{T}}$.

The following theorem gives our main results on learning the dynamics of an unknown MJS from finite samples obtained from a single trajectory. One can refer to Theorems B.1 and B.17 in Appendix B for the detailed theorem statements and proofs.

**Theorem 4.1 (Identification of MJS)** *Suppose we run Algorithm 1 with $c_{\mathsf{x}} = \mathcal{O}(\sqrt{n})$ and $c_{\mathsf{z}} = \mathcal{O}(\sqrt{p})$. Let $\rho = \rho(\tilde{\mathsf{L}})$, where $\tilde{\mathsf{L}}$ is the augmented state matrix of the closed-loop MJS defined in Sec. 3.1. Suppose the trajectory length obeys $T \geq \tilde{\mathcal{O}}\big(\frac{\bar{s} t_{\mathsf{MC}} \log^2(T)}{\pi_{\min}(1-\rho)}(n+p)\big)$. Then, under Assumption 1, with probability at least $1-\delta$, for all $i \in [s]$, we have*

$$\max\{\|\hat{\mathsf{A}}_i - \mathsf{A}_i\|, \|\hat{\mathsf{B}}_i - \mathsf{B}_i\|\} \leq \tilde{\mathcal{O}}\left( \frac{(\sigma_{\mathsf{z}} + \sigma_{\mathsf{w}})}{\sigma_{\mathsf{z}}} \frac{(n+p)\log(T)}{\pi_{\min}(1-\rho)} \sqrt{\frac{s}{T}} \right),$$
$$\text{and} \quad \|\hat{\mathsf{T}} - \mathsf{T}\|_\infty \leq \tilde{\mathcal{O}}\left( \frac{1}{\pi_{\min}} \sqrt{\frac{\log(T)}{T}} \right). \tag{4.1}$$

**Corollary 4.2** *Consider the same setting of Theorem 4.1. Additionally, when $\mathsf{B}_{1:s}$ are known, setting $\sigma_{\mathsf{z}} = 0$ and solving only for the state matrices leads to the stronger upper bound $\|\hat{\mathsf{A}}_i - \mathsf{A}_i\| \leq \tilde{\mathcal{O}}(\frac{(n+p)\log(T)}{\pi_{\min}(1-\rho)}\sqrt{\frac{s}{T}})$ for all $i \in [s]$.*

**Proof sketch** [Theorem 4.1] Our proof strategy addresses the key challenges introduced by MJS and mean-square stability. We only emphasize the core technical challenges here. The idea is to think of the set $S_i$ as a union of $L$ subsets $S_i^{(\tau)}$ defined as follows:

$$S_i^{(\tau)} := \{\tau + kL \;\big|\; \omega(\tau + kL) = i,\ \|\mathsf{x}_{\tau+kL}\| \leq c\sigma_{\mathsf{w}}\beta_+\sqrt{n},\ \|\mathsf{z}_{\tau+kL}\| \leq c\sigma_{\mathsf{z}}\sqrt{p}\}, \tag{4.2}$$

where $0 \leq \tau \leq L-1$ is a fixed offset and $k = 1, 2, \ldots, \lfloor \frac{T-L}{L} \rfloor$. The spacing of samples by $L \geq 1$ in each subset $S_i^{(\tau)}$ aims to reduce the statistical dependence across the samples belonging to that subset, to obtain weakly-dependent sub-trajectories. This weak dependence is due to the Markovian mode switching sequence $\{\omega(t)\}_{t\geq 0}$ – unique to the MJS setting – and the system's memory (contributions from the past states). Thus $L$ is primarily a function of the mixing-time ($t_{\mathsf{MC}}$) of the Markov chain and the spectral radius ($\rho(\tilde{\mathsf{L}})$) of the MJS. At a high-level, by choosing sufficiently large $L$ (e.g., $\mathcal{O}(T)$), we can upper/lower bound the empirical covariance matrix of the concatenated state vector $\mathsf{h}_{\tau_k} := [\frac{1}{\sigma_{\mathsf{w}}}\mathsf{x}_{\tau_k}^\top \ \frac{1}{\sigma_{\mathsf{z}}}\mathsf{z}_{\tau_k}^\top]^\top$ for all $\tau_k \in S_i^{(\tau)}$.

Unlike related works on system identification and regret analysis [15, 36, 37, 52, 59], mean-square stability does not lead to strong high-probability bounds, as one can only bound $\|\mathsf{x}_t\|$ or $\mathsf{x}_t\mathsf{x}_t^\top$ in expectation. Therefore, in Algorithm 1, we sample only *bounded* state-excitation pairs $(\mathsf{x}_t, \mathsf{z}_t)$ on each mode $i \in [s]$. This boundedness enables us to control the covariance matrix of $\mathsf{h}_{\tau_k}$, despite MSS and potentially heavy-tailed states, via non-asymptotic tool-sets (e.g., Thm 5.44 of [65]). However, heavy-tailed empirical covariance lower bounds require independence, and our sub-sampled data are only "approximately independent" (coupled over modes and history). To make matters worse, the fact that we sub-sample only bounded states introduces further dependencies. To resolve this, we introduce a novel strategy to construct (for the purpose of analysis) an independent subset of *processed states* from this larger dependent set. The independence is ensured by conditioning on the mode-sequence and truncating the contribution of earlier states. We then use perturbation-based techniques (see e.g., [57]) to deal with actual (non-truncated) states. The final ingredient

is showing that, for each mode $i \in [s]$, with high probability, this carefully-crafted subset contains enough samples to ensure a well-conditioned covariance (with excitation provided by $z_t$ and $w_t$). With this in place, after stitching together the estimation error from $L$ sub-trajectories $\{(x_{\tau_k}, z_{\tau_k}, \omega(\tau_k))\}_{\tau_k \in S_i^T}$ for $0 \le \tau \le L-1$, least-squares will accurately estimate $A_i$ and $B_i$ from the data $\{(x_t, z_t, \omega(t))\}_{t \in S_i}$ with statistical error rate of $\tilde{O}(1/\sqrt{T})$. ∎

Note that, our system identification result achieves near-optimal ($\mathcal{O}(1/\sqrt{T})$) dependence on the trajectory length $T$. However, the effective sample complexity of our system identification algorithm is $\mathcal{O}(s(n+p)^2 \log^2(T)/\pi_{\min}^2)$, that is, the sample complexity grows quadratically in the state dimension $n$, which can potentially be improved to linear via a more refined analysis of the state-covariance (see e.g., [16, 59] for standard LTI systems). It also grows with the inverse of the minimum mode frequency as $\pi_{\min}^{-2}$. Note that, $\pi_{\min}$ dictates the trajectory fraction of the least-frequent mode, thus, in the result $\pi_{\min}^{-1}$ multiplier is unavoidable. In Corollary 4.2, we show that, when $B_{1:s}$ is assumed to be known, $A_{1:s}$ can be estimated regardless of the exploration strength $\sigma_z$. This is because the excitation for the state matrix arises from noise $w_t$. As we will see in Section 5, the distinct $\sigma_z$ dependencies in Theorem 4.1 and Corollary 4.2 will lead to different regret bounds for MJS-LQR (albeit both bounds will be optimal up to polylog$(T)$).

# 5 Adaptive Control for MJS-LQR

Our adaptive MJS-LQR control scheme is given in Algorithm 2. It is performed on an epoch-by-epoch basis; a fixed controller is used for each epoch, and from epoch to epoch, the controller is updated using a newly collected MJS trajectory. Note that a new epoch is just a continuation of previous epochs instead of restarting the MJS. Similar to the discussion in Section 4, we assume, at the beginning of epoch 0, that one has access to a stabilizing controller $K_{1:s}^{(0)}$. During epoch $i$, the controller $K_{1:s}^{(i)}$ is used together with additive exploration noise $z_t^{(i)} \overset{\text{i.i.d.}}{\sim} \mathcal{N}(0, \sigma_{z,i}^2 I_p)$ to boost learning. At the end of epoch $i$, the trajectory during that epoch is used to obtain a new MJS dynamics estimate $A_{1:s}^{(i)}, B_{1:s}^{(i)}, T^{(i)}$ using Algorithm 1. Then, we set the controller $K_{1:s}^{(i+1)}$ for epoch $i+1$ to be the optimal controller for the infinite-horizon MJS-LQR$(A_{1:s}^{(i)}, B_{1:s}^{(i)}, T^{(i)}, Q_{1:s}, R_{1:s})$, which can be computed as follows:

For a generic infinite-horizon MJS-LQR$(A_{1:s}, B_{1:s}, T, Q_{1:s}, R_{1:s})$, its optimal controller is given by $K_{1:s}$ such that for all $j \in [s]$,

$$K_j := -\left(R_j + B_j^\top \phi_j(P_{1:s}) B_j\right)^{-1} B_j^\top \phi_j(P_{1:s}) A_j, \tag{5.1}$$

where $\phi_j(P_{1:s}) := \sum_{k=1}^s [T]_{jk} P_k$ and $P_{1:s}$ is the solution to the following coupled discrete-time algebraic Riccati equations (cDARE):

$$P_j = A_j^\top \phi_j(P_{1:s}) A_j + Q_j - A_j^\top \phi_j(P_{1:s}) B_j \left(R_j + B_j^\top \phi_j(P_{1:s}) B_j\right)^{-1} B_j^\top \phi_j(P_{1:s}) A_j, \tag{5.2}$$

for all $j \in [s]$. In practice, cDARE can be solved efficiently via value iteration or LMIs [14]. Note that cDARE may not be solvable for arbitrary parameters, but our theory guarantees that when epoch lengths are appropriately chosen, cDARE parameterized by $A_{1:s}^{(i)}, B_{1:s}^{(i)}, T^{(i)}, Q_{1:s}, R_{1:s}$ is solvable for every epoch $i$. This control design based on the estimated dynamics is also referred to as certainty equivalent control.

To achieve theoretically guaranteed performance, i.e., sub-linear regret, the key is to have a subtle scheduling of epoch lengths $T_i$ and exploration noise variance $\sigma_{z,i}^2$. We choose $T_i$ to increase exponentially with rate $\gamma > 1$, and set $\sigma_{z,i}^2 = \sigma_w^2/\sqrt{T_i}$, which collectively guarantee $\tilde{\mathcal{O}}(\sqrt{T})$ regret when combined with the system identification result from Theorem 4.1. Intuitively, this scheduling can be interpreted as follows: (i) the increase of epoch lengths guarantees we have more accurate MJS estimates thus more optimal controllers; (ii) as the controller becomes more optimal we can gradually decrease the exploration noise and deploy (exploit) the controller for a longer time. Note that the scheduling rate $\gamma$ has a similar role to the discount factor in reinforcement learning: smaller $\gamma$ aims to reduce short-term cost while larger $\gamma$ aims to reduce long-term cost.

**Algorithm 2:** Adaptive MJS-LQR

**Input:** Initial epoch length $T_0$; initial stabilizing controller $\mathsf{K}_{1:s}^{(0)}$; epoch incremental ratio $\gamma > 1$; and data clipping thresholds $c_{\mathbf{x}}$, $c_{\mathbf{z}}$

**for** $i = 0, 1, 2, \ldots$ **do**

  Set epoch length $T_i = \lfloor T_0 \gamma^i \rfloor$.

  Set exploration noise variance $\sigma_{\mathbf{z},i}^2 = \frac{\sigma_{\mathbf{w}}^2}{\sqrt{T_i}}$.

  Evolve the MJS for $T_i$ steps with $\mathbf{u}_t^{(i)} = \mathsf{K}_{\omega(t)^{(i)}}^{(i)} \mathbf{x}_t^{(i)} + \mathbf{z}_t^{(i)}$ with $\mathbf{z}_t^{(i)} \overset{\text{i.i.d.}}{\sim} \mathcal{N}(0, \sigma_{\mathbf{z},i}^2 I_p)$ and record the trajectory $\{\mathbf{x}_t^{(i)}, \mathbf{z}_t^{(i)}, \omega^{(i)}(t)\}_{t=0}^{T_i}$.

  $\mathsf{A}_{1:s}^{(i)}, \mathsf{B}_{1:s}^{(i)}, \mathsf{T}^{(i)} = \text{MJS-SYSID}(\mathsf{K}_{1:s}^{(i)}, \sigma_{\mathbf{w}}^2, \sigma_{\mathbf{z},i}^2, \{\mathbf{x}_t^{(i)}, \mathbf{z}_t^{(i)}, \omega^{(i)}(t)\}_{t=0}^{T_i}, c_{\mathbf{x}}, c_{\mathbf{z}})$.

  Set the controller $\mathsf{K}_{1:s}^{(i+1)}$ for the next epoch to be the optimal controller for the infinite-horizon MJS-LQR$(\mathsf{A}_{1:s}^{(i)}, \mathsf{B}_{1:s}^{(i)}, \mathsf{T}^{(i)}, \mathsf{Q}_{1:s}, \mathsf{R}_{1:s})$.

**end**

## 5.1 Regret Analysis

We define filtration $\mathcal{F}_{-1}, \mathcal{F}_0, \mathcal{F}_1, \ldots$ such that $\mathcal{F}_{-1} := \sigma(\mathbf{x}_0, \omega(0))$ is the sigma-algebra generated by the initial state and initial mode, and $\mathcal{F}_i := \sigma(\mathbf{x}_0, \omega(0), \{\{\omega^{(j)}(t)\}_{t=1}^{T_j}\}_{j=0}^{i}, \mathbf{w}_0, \{\{\mathbf{w}_t^{(j)}\}_{t=1}^{T_j}\}_{j=0}^{i}, \mathbf{z}_0, \{\{\mathbf{z}_t^{(j)}\}_{t=1}^{T_j}\}_{j=0}^{i})$ is the sigma-algebra generated by the randomness up to epoch $i$. Note that the initial state $\mathbf{x}_0^{(i)}$ of epoch $i$ is also the final state $\mathbf{x}_{T_{i-1}}^{(i-1)}$ of epoch $i-1$, therefore, $\mathbf{x}_0^{(i)}$ is $\mathcal{F}_{i-1}$-measurable, and so is $\omega^{(i)}(0)$. Suppose time step $t$ belongs to epoch $i$, then we define the following conditional expected cost at time $t$ as:

$$c_t = \mathbb{E}[\mathbf{x}_t^\top Q_{\omega(t)} \mathbf{x}_t + \mathbf{u}_t^\top R_{\omega(t)} \mathbf{u}_t \mid \mathcal{F}_{i-1}], \tag{5.3}$$

and cumulative cost as $J_T = \sum_{t=1}^{T} c_t$. We define the total regret and epoch-$i$ regret as

$$\text{Regret}(T) = J_T - T J^\star, \quad \text{Regret}_i = \Big(\sum_{t=1}^{T_i} c_{T_0 + \cdots + T_{i-1} + t}\Big) - T_i J^\star. \tag{5.4}$$

Then, we have $\text{Regret}(T) = \mathcal{O}(\sum_{i=1}^{\mathcal{O}(\log_\gamma(T/T_0))} \text{Regret}_i)$, where regret of epoch 0 is ignored as it does not scale with time $T$. Let $\mathsf{K}_{1:s}^\star$ denote the optimal controller for the infinite-horizon MJS-LQR$(\mathsf{A}_{1:s}, \mathsf{B}_{1:s}, \mathsf{T}, \mathsf{Q}_{1:s}, \mathsf{R}_{1:s})$ problem. $\tilde{\mathsf{L}}^{(0)}$ and $\tilde{\mathsf{L}}^\star$ denote the closed-loop augmented state matrices under the initial controller $\mathsf{K}_{1:s}^{(0)}$ and $\mathsf{K}_{1:s}^\star$ respectively, and we let $\bar{\rho} := \max\{\rho(\tilde{\mathsf{L}}^{(0)}), \rho(\tilde{\mathsf{L}}^\star)\}$. With these definitions, we have the following sub-linear regret guarantee. Please refer to Theorem C.11 in the appendix for the complete version and proof.

**Theorem 5.1 (Sub-linear regret)** *Assume that the initial state $\mathbf{x}_0 = 0$, and Assumptions 1 and 2 hold. In Algorithm 2, suppose hyper parameters $c_{\mathbf{x}} = \mathcal{O}(\sqrt{n})$, $c_{\mathbf{z}} = \mathcal{O}(\sqrt{p})$, and $T_0 \geq \tilde{\mathcal{O}}\big(\frac{\sqrt{\bar{s} t_{\text{MC}} \log^2(T_0)}}{\pi_{\min}(1-\bar{\rho})}(n+p)\big)$. Then, with probability at least $1-\delta$, Algorithm 2 achieves*

$$\text{Regret}(T) \leq \tilde{\mathcal{O}}\left(\frac{s^2 p(n^2+p^2)\sigma_{\mathbf{w}}^2}{\pi_{\min}^2} \log^2(T)\sqrt{T}\right) + \mathcal{O}\left(\frac{\sqrt{ns}\log^3(T)}{\delta}\right). \tag{5.5}$$

**Proof sketch** [Theorem 5.1] For simplicity, we only show the dominant $\tilde{\mathcal{O}}(\cdot)$ term here and leave the complete proof to appendix. Define the estimation error after epoch $i$ as $\varepsilon_{\mathbf{A},\mathbf{B}}^{(i)} := \max_{j \in [s]} \max\{\|\mathsf{A}_j^{(i)} - \mathsf{A}_j\|, \|\mathsf{B}_j^{(i)} - \mathsf{B}_j\|\}$ and $\varepsilon_{\mathbf{T}}^{(i)} := \|\mathsf{T}^{(i)} - \mathsf{T}\|_\infty$. Analyzing the finite-horizon cost and combining the infinite-horizon perturbation results in [18], we can bound epoch-$i$ regret as $\text{Regret}_i \leq \mathcal{O}\left(T_i \sigma_{\mathbf{z},i}^2 + T_i \sigma_{\mathbf{w}}^2 \left(\varepsilon_{\mathbf{A},\mathbf{B}}^{(i-1)} + \varepsilon_{\mathbf{T}}^{(i-1)}\right)^2\right)$. Plugging in $\sigma_{\mathbf{z},i}^2 = \frac{\sigma_{\mathbf{w}}^2}{\sqrt{T_i}}$ and the upper bounds on the estimation errors $\varepsilon_{\mathbf{A},\mathbf{B}}^{(i)} \leq \tilde{\mathcal{O}}\left(\frac{\sigma_{\mathbf{z},i}+\sigma_{\mathbf{w}}}{\sigma_{\mathbf{z},i}\pi_{\min}} \frac{\sqrt{\bar{s}(n+p)\log(T_i)}}{\sqrt[4]{T_i}}\right)$ and $\varepsilon_{\mathbf{T}}^{(i)} \leq \tilde{\mathcal{O}}\left(\sqrt{\frac{\log(T_i)}{T_i}}\right)$

from Theorem 4.1, we have $\text{Regret}_i \le \tilde{\mathcal{O}}\left(\frac{s^2p(n^2+p^2)\sigma_w^2}{\pi_{\min}^2}\gamma\sqrt{T_i}\log^2(T_i)\right)$. Finally, since $T_i = \mathcal{O}(T_0\gamma^i)$ from Alg. 2, we have $\text{Regret}(T) = \sum_{i=1}^{O(\log_\gamma(\frac{T}{T_0}))}\text{Regret}_i \le \tilde{\mathcal{O}}\left(\frac{s^2p(n^2+p^2)\sigma_w^2}{\pi_{\min}^2}\sqrt{T}\log(\frac{T}{T_0})\left(\frac{\sqrt{\gamma}}{\sqrt{\gamma}-1}\right)^3\left(\gamma\log(\frac{T}{T_0})-\sqrt{\gamma}\log(\gamma)\right)\right) = \tilde{\mathcal{O}}\left(\frac{s^2p(n^2+p^2)\sigma_w^2}{\pi_{\min}^2}\text{polylog}(T)\sqrt{T}\right)$. ■

One can see the interplay between $T$ and $\gamma$ from the term $\left(\frac{\sqrt{\gamma}}{\sqrt{\gamma}-1}\right)^3\left(\gamma\log(\frac{T}{T_0})-\sqrt{\gamma}\log(\gamma)\right)$ in the proof sketch. Specifically, when horizon $T$ is smaller, a smaller $\gamma$ minimizes the upper bound, and vice versa. This further provides a mathematical justification for $\gamma$ being similar to the discount factor in reinforcement learning in early discussions.

## 5.2 Two Special Cases

### 5.2.1 Tighter probability bound under uniform stability

Note that the regret upper bound (5.5) in Theorem 5.1 has the second term depending on the failure probability $\delta$ through $\frac{1}{\delta}$. Though this term has a much milder dependency on the time horizon $T$, when setting $\delta$ to be small, it can still easily outweigh the other $\tilde{\mathcal{O}}(\cdot)$ term in (5.5), which only has $\log(\frac{1}{\delta})$ dependency, and can result in overly pessimistic regret bounds. The main cause of this $\frac{1}{\delta}$ term is that in the regret analysis, one needs to factor in the cumulative impact of initial state of every epoch, i.e. $\sum_i\|x_0^{(i)}\|^2$. Since MSS guarantees the stability and state convergence only in the mean-square sense, we can, at best, only bound $\mathbb{E}[\|x_0^{(i)}\|^2]$ and then use the Markov inequality: with probability at least $1-\delta$, $\|x_0^{(i)}\|^2 \le \mathbb{E}[\|x_0^{(i)}\|^2]/\delta$. Furthermore, in Appendix C.4, we construct an MJS example that is MSS but no dependencies better than $\frac{1}{\delta}$ can be established. Fortunately, there exists an easy workaround to get rid of this $\frac{1}{\delta}$ dependency if the MJS is uniformly stable [42, 44], which enforces stability under arbitrary switching sequences thus is stronger than MSS. It allows us to bound $x_0^{(i)}$ using tail inequalities much tighter than the Markov inequality and obtain $\|x_0^{(i)}\|^2 \le \mathcal{O}(\log(\frac{1}{\delta}))$. In the end, in the regret bound, $\frac{1}{\delta}$ can be improved to $\log(\frac{1}{\delta})$.

One type of uniform stability assumption that can help us in this case is regarding the closed-loop MJS under the optimal controllers. We let $K_{1:s}^\star$ denote the optimal controller for the infinite-horizon MJS-LQR$(A_{1:s}, B_{1:s}, T, Q_{1:s}, R_{1:s})$ and define closed-loop state matrices $L_i^\star = A_i + B_iK_i^\star$ for all $i \in [s]$. We let $\theta^\star$ denote the joint spectral radius of $L_{1:s}^\star$, i.e. $\theta^\star := \lim_{l\to\infty}\max_{\omega_{1:l}\in[s]^l}\|L_{\omega_1}^\star\cdots L_{\omega_l}^\star\|^{\frac{1}{l}}$, and we say $L_{1:s}^\star$ is uniformly stable if and only if $\theta^\star < 1$. The resulting regret bound is outlined in the following theorem, with its complete version and proof provided in Theorem C.12 of Appendix C.4.1.

**Theorem 5.2 (Regret Under Uniform Stability)** *Assume that the initial state $x_0 = 0$, Assumptions 1 and 2 hold, and $L_{1:s}^\star$ is* uniformly stable. *If hyper-parameters $T_0$, $c_x$, and $c_z$ are chosen as su ciently large, with probability at least $1-\delta$, Algorithm 2 achieves*

$$\text{Regret}(T) \le \tilde{\mathcal{O}}\left(\frac{s^2p(n^2+p^2)\sigma_w^2}{\pi_{\min}^2}\log^2(T)\sqrt{T}\right). \tag{5.6}$$

### 5.2.2 Partial knowledge of dynamics

From Corollary 4.2, we know that when input matrices $B_{1:s}$ are known, no further exploration noise is needed to identify the state matrices $A_{1:s}$ or Markov matrix $T$. This can also be applied to the adaptive MJS-LQR setting, and the resulting regret bound can improve (from $\mathcal{O}(\log^2(T)\sqrt{T})$ to $\mathcal{O}(\log^3(T))$) since exploration noise incurs additional costs. The result is given by the following corollary, and we omit the proof due to its similarity to the proofs of Theorems 5.1 and 5.2.

**Corollary 5.3 (Poly-logarithmic regret)** *When $B_{1:s}$ are known, it su ces to set the exploration noise to be $\sigma_{z,i} = 0$ for all $i$ in Algorithm 2. Then, the regret bound in Theorem 5.1 becomes* $\text{Regret}(T) \le$

$\tilde{\mathcal{O}}\left(\frac{s^2p(n^2+p^2)\sigma_w^2}{\pi_{\min}^2}\log^3(T)\right)+\mathcal{O}\left(\frac{\sqrt{ns}\log^3(T)}{\delta}\right)$. *Additionally, the regret bound in Theorem 5.2 becomes* $\mathrm{Regret}(T) \leq \tilde{\mathcal{O}}\left(\frac{s^2p(n^2+p^2)\sigma_w^2}{\pi_{\min}^2}\log^3(T)\right)$.

# 6 Discussion

In this section, we discuss how one may obtain the initial stabilizing controller for MJS as required in the input to Algorithms 1 and 2 and the application of our results to offline data-driven control.

## 6.1 Initial Stabilizing Controllers

Having access to an initial stabilizing controller has become a very common assumption in system identification (see for instance [41] and references therein) and adaptive control [3, 13, 15, 32, 58] for LTI systems. On the other hand, for work where no initial stabilizing controller is required, there is usually a separate warm-up phase at the beginning, where coarse dynamics is learned, upon which a stabilizing controller is computed. Recent non-asymptotic system identification results [21, 55] on potentially unstable LTI systems can be used to obtain coarse dynamics without stabilizing controller. One can use random linear feedback to construct a confidence set of the dynamics such that any point in this set can produce a stabilizing controller by solving Riccati equations [20]. In the model-free setting, [39] provides asymptotic results and relies on persistent excitation assumption. [11] designs subtle scaled one-hot vector input and collects the trajectory to estimate the dynamics, then a stabilizing controller can be solved via semi-definite programming. For MJS or general switched systems, to the best of our knowledge, there is no work on stabilizing unknown dynamics using single trajectory with guarantees. One challenge is, as we discussed in Section 1, the individual mode stability and overall stability does not imply each other due to mode switching. However, as outlined below, we can approach this problem leveraging what is recently done for the LTI case in the aforementioned literature (modulo some additional assumptions).

Similar to the LTI case, suppose we could obtain some coarse dynamics estimate $\hat{A}_{1:s}, \hat{B}_{1:s}, \hat{T}$, then we can solve for the optimal controller $\hat{K}_{1:s}$ of the infinite-horizon MJS-LQR($\hat{A}_{1:s}, \hat{B}_{1:s}, \hat{T}, Q_{1:s}, R_{1:s}$) via Riccati equations. To investigate when $\hat{K}_{1:s}$ can stabilize the MJS, the key is to obtain sample complexity guarantees for this coarse dynamics, i.e. dependence of estimation error $\|\hat{A}_i - A_i\|$, $\|\hat{B}_i - B\|$, and $\|\hat{T} - T\|$ on sample size. Fortunately [18] provides the required estimation accuracy under which $\hat{K}_{1:s}$ is guaranteed to be stabilizing. Thus, combining [18] with the estimation error bounds (in terms of sample size), the required accuracy can be translated to the required number of samples. Note that learning $T$ is the same as learning a Markov chain, thus using the mode transition pair frequencies in an arbitrary single MJS trajectory, we can obtain an estimate $\hat{T}$ as in Algorithm 1, and its sample complexity is given in Lemma B.1 in Appendix B. The more challenging part is the identification scheme and corresponding sample complexity for $\hat{A}_{1:s}$ and $\hat{B}_{1:s}$. Here, we outline two potential schemes.

- Suppose we could generate $N$ i.i.d. MJS rollout trajectories, each with length $T$ (small $T$, e.g. $T = 1$, is preferred to avoid potential unstable behavior and for the ease of the implementation). We can obtain least squares estimates $\hat{A}_{1:s}, \hat{B}_{1:s}$ using only $\{x_T, x_{T-1}, u_{T-1}, \omega(T-1)\}$ from each trajectory, which is similar to the scheme in [16] for LTI systems. Since only i.i.d. data is used in the computation, one can easily obtain the sample complexity in terms of $N$.

- If each mode in the MJS can run in isolation (i.e. for any $i \in [s]$, $\omega(t) = i$ for all $t$) so that it acts as an LTI system, we could use recent advances [21, 55] on single-trajectory open-loop LTI system identification to obtain coarse estimates together with sample complexity for $\hat{A}_i$ and $\hat{B}_i$ for every mode $i$.

We also note that while finding an initial stabilizing controller is theoretically very interesting and challenging, most results we know of are limited to simulated or numerical examples (see for instance [41] and references therein). This is because, from a practical standpoint, an initial stabilizing controller is almost

required in model-based approaches since running experiments with open-loop unstable plants can be very dangerous as the state could explode quickly.

### 6.2 Offline Data-Driven Control

In many scenarios, we may not be able to perform learning and control in real time due to limited onboard computing resources or measurement sensors. In this case, the dynamics is usually learned in a one-shot way at the beginning, and the resulting controller will be deployed forever without any further update. The controller suboptimality in this non-adaptive setting does not improve over time, thus the regret will increase linearly over time rather than sublinearly as in our work. The natural performance metric in this case is the time-averaged regret, which can also be viewed as the slope of the cumulative regret with respect to time. The system identification scheme and corresponding sample complexity developed in this paper can also help address this problem.

Suppose we obtain MJS estimate $\hat{\mathsf{A}}_{1:s}, \hat{\mathsf{B}}_{1:s}, \hat{\mathsf{T}}$ from a length-$T_0$ rollout trajectory using Algorithm 1 and solve for the controller $\hat{\mathsf{K}}_{1:s}$ that is optimal for the infinite-horizon MJS-LQR$(\hat{\mathsf{A}}_{1:s}, \hat{\mathsf{B}}_{1:s}, \hat{\mathsf{T}}, \mathsf{Q}_{1:s}, \mathsf{R}_{1:s})$ via Riccati equations. Let $\hat{J} := \limsup_{T\to\infty} \frac{1}{T} J(\hat{\mathsf{K}}_{\omega(t)} \mathsf{x}_t)$ denote the infinite-horizon average cost incurred when we deploy $\hat{\mathsf{K}}_{1:s}$ indefinitely. Combining our identification sample complexity result in Theorem 4.1 with the infinite-horizon MJS-LQR perturbation result in [18], we can easily obtain an upper bound on the suboptimality, $\hat{J} - J^\star \le \tilde{\mathcal{O}}\left(\log^2(T_0)/T_0\right)$, which provides the required rollout trajectory length $T_0$ if certain suboptimality is desired.

## 7 Numerical Experiments

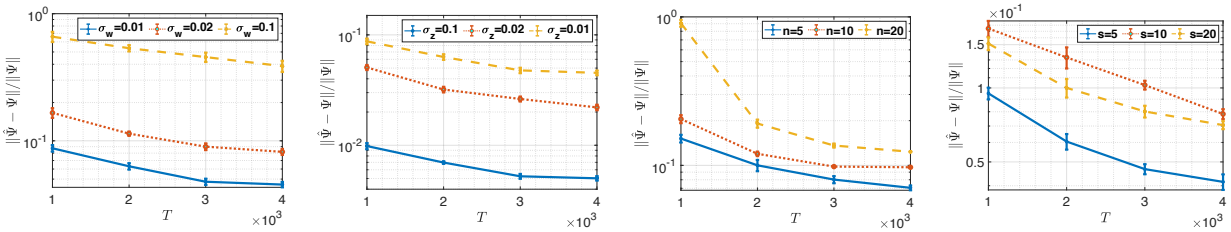
We provide experiments to investigate the efficiency and verify the theory of the proposed algorithms on synthetic datasets. Throughout, we show results from a synthetic experiment where entries of the true system matrices $(A_{1:s}, B_{1:s})$ were generated randomly from a standard normal distribution. We further scale each $\mathsf{A}_i$ to have $\|\mathsf{A}_i\| \le 0.5$. Since this guarantees the MJS itself is MSS, as we discussed in Sec 4, we set controller $\mathsf{K}_{1:s} = 0$ in system identification Algorithm 1 and initial stabilizing controller $\mathsf{K}_{1:s}^{(0)} = 0$ in adaptive MJS-LQR Algorithm 2. For the cost matrices $(Q_{1:s}, R_{1:s})$, we set $Q_j = \underline{Q}_j \underline{Q}_j^\top$, and $R_j = \underline{R}_j \underline{R}_j^\top$ where $\underline{Q}_j \in \mathsf{R}^{n\times n}$ and $\underline{R}_j \in \mathsf{R}^{p\times p}$ were generated from a standard normal distribution. The Markov matrix $T \in \mathsf{R}_+^{s\times s}$ was sampled from a Dirichlet distribution $\mathtt{Dir}((s-1)\cdot I_s + 1)$, where $I_s$ denotes the identity matrix. We assume that we had equal probability of starting in any initial mode.

Since for system identification, our main contribution is estimating $\mathsf{A}_{1:s}$ and $\mathsf{B}_{1:s}$ of the MJS, we omit the plots for estimating $\mathsf{T}$. Let $\hat{\Psi}_j = [\hat{\mathsf{A}}_j, \hat{\mathsf{B}}_j]$ and $\Psi_j = [\mathsf{A}_j, \mathsf{B}_j]$. We use $\|\hat{\Psi} - \Psi\|/\|\Psi\| := \max_{j\in[s]} \|\hat{\Psi}_j - \Psi_j\|/\|\Psi_j\|$ to investigate the convergence behaviour of MJS-SYSID Algorithm 1. The clipping constants in this algorithm, i.e., $C_{\mathsf{sub}}$, $c_{\mathbf{x}}$, and $c_{\mathbf{z}}$ are chosen based on their lower bounds provided in Theorem 5.1. In all the aforementioned algorithms, the depicted results are averaged over 10 independent replications.

### 7.1 Performance of MJS-SYSID

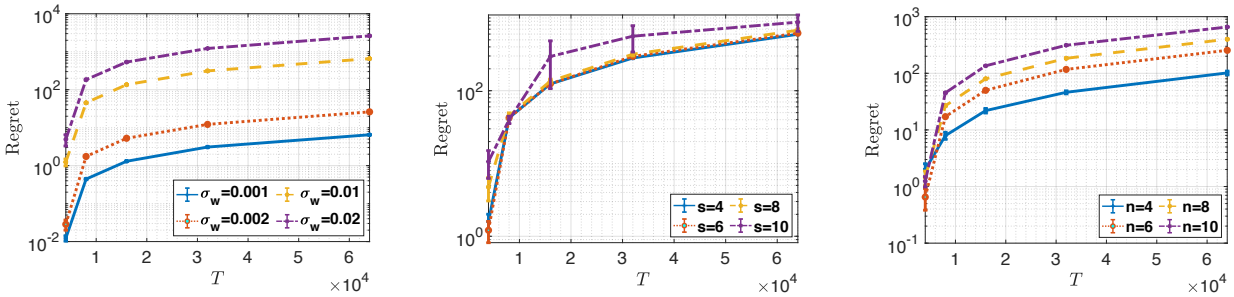
In this section, we investigate the performance of our MJS-SYSID method, i.e., Algorithm 1. We first empirically evaluate the effect of the noise variances $\sigma_{\mathbf{w}}$ and $\sigma_{\mathbf{z}}$. In particular, we study how the system errors vary with (i) $\sigma_{\mathbf{w}} = 0.01, \sigma_{\mathbf{z}} \in \{0.01, 0.02, 0.1\}$ and (ii) $\sigma_{\mathbf{z}} = 0.01, \sigma_{\mathbf{w}} \in \{0.01, 0.02, 0.1\}$. The number of states, inputs, and modes are set to $n = 5$, $p = 3$, and $s = 5$, respectively. Fig. 2 (a) and (b) demonstrate how the relative estimation error $\|\hat{\Psi} - \Psi\|/\|\Psi\|$ changes as $T$ increases. Each curve on the plot represents a fixed $\sigma_{\mathbf{w}}$ and $\sigma_{\mathbf{z}}$. These empirical results are all consistent with the theoretical bound of MJS-SYSID given in (4.1). In particular, the estimation errors degrade with increasing $\sigma_{\mathbf{w}}$ and decreasing $\sigma_{\mathbf{z}}$, respectively.

Now, we fix $\sigma_{\mathbf{w}}$=$\sigma_{\mathbf{z}}$=0.01 and investigate the performance of the MJS-SYSID with varying number of states, inputs, and modes. Fig. 2 (c) and (d) show how the estimation error $\|\hat{\Psi} - \Psi\|/\|\Psi\|$ changes with (left) $s = 5$, $n \in \{5, 10, 20\}$, $p = n - 2$ and (right) $n = 5$, $p = n - 2$, $s \in \{5, 10, 20\}$. As we can see, the MJS-SYSID has better performance with small $n$, $p$ and $s$ which is consistent with (4.1).




(a) process noise $\sigma_{\mathbf{w}}$ (b) exploration noise $\sigma_{\mathbf{z}}$ (c) state dimension $n$ (d) number of modes $s$

Figure 2: Performance profiles of MJS-SYSID with varying: (a) process noise $\sigma_{\mathbf{w}}$, (b) exploration noise $\sigma_{\mathbf{z}}$, (c) state dimension $n$, and (d) number of modes $s$.




(a) process noise $\sigma_{\mathbf{w}}$ (b) number of modes $s$ (c) state dimension $n$

Figure 3: Performance profiles of Adaptive MJS-LQR with varying: (a) process noise $\sigma_{\mathbf{w}}$, (b) number of modes $s$, (c) state dimension $n$.

### 7.2 Performance of Adaptive MJS-LQR

In our next series of experiments, we explore the sensitivity of the regret bounds to the system parameters. In these experiments, we set the initial epoch length $T_0 = 2000$ and incremental ratio $\gamma = 2$. We select five epochs to run Algorithm 2. As an intermediate step for computing controller $\mathsf{K}_{1:s}^{(i+1)}$ in Algorithm 2, the coupled Riccati equations (5.2) are solved via value iteration, and the iteration stops when the parameter variation between two iterations falls below $10^{-6}$, or iteration number reaches $10^4$.

Fig. 3 demonstrates how regret bounds vary with (a) $\sigma_{\mathbf{w}} \in \{0.001, 0.002, 0.01, 0.02\}$, $n = 10$, $p = s = 5$; (b) $\sigma_{\mathbf{w}} = 0.01$, $n = 10$, $p = 5$, $s \in \{4, 6, 8, 10\}$, and (c) $\sigma_{\mathbf{w}} = 0.01$, $s = 10$, $p = 5$, $n \in \{4, 6, 8, 10\}$. We see that the regret degrades as $\sigma_{\mathbf{w}}$, $n$, and $s$ increase. We also see that when $\sigma_{\mathbf{w}}$ is large ($T$ is small), the regret becomes worse quickly as $n$ and $s$ grow larger. These results are consistent with the theoretical bounds in Theorem 5.1.

## 8 Conclusions and Discussion

Markov jump systems are fundamental to a rich class of control problems where the underlying dynamics are changing with time. Despite its importance, statistical understanding (system identification and regret bounds) of MJS have been lacking due to the technicalities such as Markovian transitions and weaker notion of mean-square stability. At a high-level, this work overcomes (much of) these challenges to provide finite sample system identification and model-based adaptive control guarantees for MJS. Notably, resulting estimation error and regret bounds are optimal in the trajectory length and coincide with the standard LQR up to polylogarithmic factors. As a future work, it would be interesting and of practical importance to investigate the case when mode is not observed, which makes both system identification and adaptive quadratic control problems non-trivial.

We want to mention possible negative societal impacts. While our work is theoretical and has many

potential positive impacts in reinforcement learning, robotics, and autonomous systems, there are also potential negative applications in the military (e.g. with drone control) and for malicious actors (e.g. computer network hackers), among others. Additionally, all our work was built on stochastic noise assumptions, whereas in reality intelligent autonomous systems may instead encounter adversarial behavior. There is potential here for future work to extend our approach to non-stochastic noise or even non-Markovian / non-random switching among states.

## Acknowledgements

Y. Sattar and S. Oymak were supported in part by NSF grant CNS-1932254 and S. Oymak was supported in part by NSF CAREER award CCF-2046816 and ARO MURI grant W911NF-21-1-0312. Z. Du and N. Ozay were supported in part by ONR under grant N00014-18-1-2501 and N. Ozay was supported in part by NSF under grant CNS-1931982 and ONR under grant N00014-21-1-2431. D. Ataee Tarzanagh, Z. Du, and L. Balzano were supported in part by NSF CAREER award CCF-1845076 and AFOSR YIP award FA9550-19-1-0026.

## References

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

[2] Yasin Abbasi-Yadkori and Csaba Szepesvári. Regret bounds for the adaptive control of linear quadratic systems. In *Proc. of COLT*, pages 1–26. JMLR Workshop and Conference Proceedings, 2011.

[3] 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. PMLR, 2018.

[4] Marc Abeille and Alessandro Lazaric. Efficient optimistic exploration in linear-quadratic regulators via lagrangian relaxation. In *ICML*, pages 23–31. PMLR, 2020.

[5] Sevi Baltaoglu, Lang Tong, and Qing Zhao. Online learning and optimization of markov jump affine models. *arXiv preprint arXiv:1605.02213*, 2016.

[6] Lars Blackmore, Stanislav Funiak, and Brian C Williams. Combining stochastic and greedy search in hybrid estimation. In *AAAI*, pages 282–287, 2005.

[7] Peter E Caines and Ji-Feng Zhang. On the adaptive control of jump parameter systems via nonlinear filtering. *SIAM J. Control Optim.*, 33(6):1758–1777, 1995.

[8] DO Cajueiro. *Stochastic optimal control of jumping Markov parameter processes with applications to finance*. PhD thesis, PhD thesis, 2002, Instituto Tecnológico de Aeronáutica-ITA, Brazil, 2002.

[9] Marco C Campi and PR Kumar. Adaptive linear quadratic gaussian control: the cost-biased approach revisited. *SIAM J. Control Optim.*, 36(6):1890–1907, 1998.

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

[11] Xinyi Chen and Elad Hazan. Black-box control for linear dynamical systems. In *Conference on Learning Theory*, pages 1114–1143. PMLR, 2021.

[12] Howard J Chizeck, Alan S Willsky, and D Castanon. Discrete-time markovian-jump linear quadratic optimal control. *International Journal of Control*, 43(1):213–231, 1986.

[13] Alon Cohen, Tomer Koren, and Yishay Mansour. Learning linear-quadratic regulators efficiently with only $\sqrt{T}$ regret. In *International Conference on Machine Learning*, pages 1300–1309. PMLR, 2019.

[14] Oswaldo Luiz Valle Costa, Marcelo Dutra Fragoso, and Ricardo Paulino Marques. *Discrete-time Markov jump linear systems*. Springer, 2006.

[15] 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.

[16] Sarah Dean, Horia Mania, Nikolai Matni, Benjamin Recht, and Stephen Tu. On the sample complexity of the linear quadratic regulator. *FOCM*, pages 1–47, 2019.

[17] Zhe Du, Necmiye Ozay, and Laura Balzano. Mode clustering for markov jump systems. *arXiv preprint arXiv:1910.02193*, 2019.

[18] Zhe Du, Yahya Sattar, Davoud Ataee Tarzanagh, Laura Balzano, Samet Oymak, and Necmiye Ozay. Certainty equivalent quadratic control for markov jump systems. *arXiv preprint arXiv:2105.12358*, 2021.

[19] Zhe Du, Yahya Sattar, Davoud Ataee Tarzanagh, Laura Balzano, Necmiye Ozay, and Samet Oymak. Data-driven control of markov jump systems: Sample complexity and regret bounds. https://www.dropbox.com/s/s0xp09lxzgtsv4f/AdaptiveMJS_LQR_ACC.pdf, 2021.

[20] Mohamad Kazem Shirani Faradonbeh, Ambuj Tewari, and George Michailidis. Finite-time adaptive stabilization of linear systems. *IEEE Transactions on Automatic Control*, 64(8):3498–3505, 2018.

[21] Mohamad Kazem Shirani Faradonbeh, Ambuj Tewari, and George Michailidis. Finite time identification in unstable linear systems. *Automatica*, 96:342–353, 2018.

[22] Mohamad Kazem Shirani Faradonbeh, Ambuj Tewari, and George Michailidis. On adaptive linear–quadratic regulators. *Automatica*, 117:108982, 2020.

[23] Mohamad Kazem Shirani Faradonbeh, Ambuj Tewari, and George Michailidis. Optimism-based adaptive regulation of linear-quadratic systems. *IEEE Transactions on Automatic Control*, 2020.

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

[25] Emily B Fox, Erik B Sudderth, Michael I Jordan, and Alan S Willsky. Bayesian nonparametric methods for learning markov switching processes. *IEEE Signal Processing Magazine*, 27(6):43–54, 2010.

[26] David A Freedman. On tail probabilities for martingales. *the Annals of Probability*, pages 100–118, 1975.

[27] Robert G Gallager. *Stochastic processes: theory for applications*. Cambridge University Press, 2013.

[28] Joseph E Gaudio, Travis E Gibson, Anuradha M Annaswamy, Michael A Bolender, and Eugene Lavretsky. Connections between adaptive control and optimization in machine learning. In *2019 IEEE 58th Conference on Decision and Control (CDC)*, pages 4563–4568. IEEE, 2019.

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

[30] Pedro Hespanhol and Anil Aswani. Statistical consistency of set-membership estimator for linear systems. *IEEE Control Systems Letters*, 4(3):668–673, 2020.

[31] Daniel Hsu, Sham Kakade, Tong Zhang, et al. A tail inequality for quadratic forms of subgaussian random vectors. *Electronic Communications in Probability*, 17, 2012.

[32] Morteza Ibrahimi, Adel Javanmard, and Benjamin Van Roy. Efficient reinforcement learning for high dimensional linear quadratic systems. In *NeurIPS*, pages 2645–2653, 2012.

[33] Joao Paulo Jansch-Porto, Bin Hu, and Geir Dullerud. Policy learning of mdps with mixed continuous/discrete variables: A case study on model-free control of markovian jump systems. In *Learning for Dynamics and Control*, pages 947–957. PMLR, 2020.

[34] Yassir Jedra and Alexandre Proutiere. Finite-time identification of stable linear systems optimality of the least-squares estimator. In *2020 59th IEEE Conference on Decision and Control (CDC)*, pages 996–1001. IEEE, 2020.

[35] Vitaly Kuznetsov and Mehryar Mohri. Generalization bounds for non-stationary mixing processes. *Machine Learning*, 106(1):93–117, 2017.

[36] Sahin Lale, Kamyar Azizzadenesheli, Babak Hassibi, and Anima Anandkumar. Explore more and improve regret in linear quadratic regulators. *arXiv preprint arXiv:2007.12291*, 2020.

[37] Sahin Lale, Kamyar Azizzadenesheli, Babak Hassibi, and Anima Anandkumar. Logarithmic regret bound in partially observable linear dynamical systems. In *Advances in Neural Information Processing Systems*, 2020.

[38] Sahin Lale, Oguzhan Teke, Babak Hassibi, and Anima Anandkumar. Stability and identification of random asynchronous linear time-invariant systems. In *Learning for Dynamics and Control*, pages 651–663. PMLR, 2021.

[39] Andrew Lamperski. Computing stabilizing linear controllers via policy iteration. In *2020 59th IEEE Conference on Decision and Control (CDC)*, pages 1902–1907. IEEE, 2020.

[40] F Lauer and G Bloch. Hybrid system identification: Theory and algorithms for learning switching models, vol. 478. *Cham, Switzerland: Springer*, 2018.

[41] Bruce Lee and Andrew Lamperski. Non-asymptotic closed-loop system identification using autoregressive processes and hankel model reduction. In *2020 59th IEEE Conference on Decision and Control (CDC)*, pages 3419–3424. IEEE, 2020.

[42] Ji-Woong Lee and Geir E. Dullerud. Uniform stabilization of discrete-time switched and markovian jump linear systems. *Automatica*, 42(2):205–218, 2006.

[43] David A Levin and Yuval Peres. *Markov chains and mixing times*, volume 107. American Mathematical Soc., 2017.

[44] Daniel Liberzon. *Switching in systems and control*. Springer Science & Business Media, 2003.

[45] Lennart Ljung. System identification. *Wiley Encyclopedia of Electrical and Electronics Engineering*, pages 1–19, 1999.

[46] KA Loparo and F Abdel-Malek. A probabilistic approach to dynamic power system security. *IEEE transactions on circuits and systems*, 37(6):787–798, 1990.

[47] Horia Mania, Stephen Tu, and Benjamin Recht. Certainty equivalence is efficient for linear quadratic control. In *NeurIPS*, 2019.

[48] Nikolai Matni, Alexandre Proutiere, Anders Rantzer, and Stephen Tu. From self-tuning regulators to reinforcement learning and back again. In *2019 IEEE 58th Conference on Decision and Control (CDC)*, pages 3724–3740. IEEE, 2019.

[49] Hesameddin Mohammadi, Mahdi Soltanolkotabi, and Mihailo R Jovanović. On the linear convergence of random search for discrete-time lqr. *IEEE Control Systems Letters*, 5(3):989–994, 2020.

[50] Mehryar Mohri and Afshin Rostamizadeh. Stability bounds for non-iid processes. In *Advances in Neural Information Processing Systems*, pages 1025–1032, 2008.

[51] Samet Oymak. Stochastic gradient descent learns state equations with nonlinear activations. In *Conference on Learning Theory*, pages 2551–2579, 2019.

[52] Samet Oymak and Necmiye Ozay. Non-asymptotic identification of lti systems from a single trajectory. *American Control Conference*, 2019.

[53] Necmiye Ozay, Mario Sznaier, Constantino M Lagoa, and Octavia I Camps. A sparsification approach to set membership identification of switched affine systems. *IEEE Transactions on Automatic Control*, 57(3):634–648, 2011.

[54] Benjamin Recht. A tour of reinforcement learning: The view from continuous control. *Annual Review of Control, Robotics, and Autonomous Systems*, 2:253–279, 2019.

[55] Tuhin Sarkar and Alexander Rakhlin. Near optimal finite time identification of arbitrary linear dynamical systems. In *ICML*, pages 5610–5618. PMLR, 2019.

[56] Tuhin Sarkar, Alexander Rakhlin, and Munther Dahleh. Nonparametric system identification of stochastic switched linear systems. In *2019 IEEE 58th Conference on Decision and Control (CDC)*, pages 3623–3628. IEEE, 2019.

[57] Yahya Sattar and Samet Oymak. Non-asymptotic and accurate learning of nonlinear dynamical systems. *arXiv preprint arXiv:2002.08538*, 2020.

[58] Max Simchowitz and Dylan Foster. Naive exploration is optimal for online lqr. In *ICML*, pages 8937–8948. PMLR, 2020.

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

[60] Lars EO Svensson, Noah Williams, et al. Optimal monetary policy under uncertainty: a markov jump-linear-quadratic approach. *Federal Reserve Bank of St. Louis Review*, 90(4):275–293, 2008.

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

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

[63] Jitendra Tugnait. Adaptive estimation and identification for discrete systems with markov jump parameters. *IEEE Transactions on Automatic control*, 27(5):1054–1065, 1982.

[64] Valery Ugrinovskii* and Hemanshu R Pota. Decentralized control of power systems via robust control of uncertain markov jump parameter systems. *International Journal of Control*, 78(9):662–677, 2005.

[65] Roman Vershynin. *Introduction to the non-asymptotic analysis of random matrices*, page 210–268. Cambridge University Press, 2012.

[66] F Xue and L Guo. Necessary and sufficient conditions for adaptive stablizability of jump linear systems. *Communications in Information and Systems*, 1(2):205–224, 2001.

[67] Anru Zhang and Mengdi Wang. Spectral state compression of markov processes. *IEEE transactions on information theory*, 66(5):3202–3231, 2019.

[68] Kaiqing Zhang, Bin Hu, and Tamer Basar. Policy optimization for $\mathcal{H}_2$ linear control with $\mathcal{H}_\infty$ robustness guarantee: Implicit regularization and global convergence. In *Learning for Dynamics and Control*, pages 179–190. PMLR, 2020.

[69] Yang Zheng, Yujie Tang, and Na Li. Analysis of the optimization landscape of linear quadratic gaussian (lqg) control. *arXiv preprint arXiv:2102.04393*, 2021.

# A Preliminaries

In additional to the notations defined in Section 1, we define a few more here to be used throughout the appendix. For a matrix $V$, $\underline{\sigma}(V)$, $\|V\|_1$, and $\|V\|_F$ denote its smallest singular value, $_1$ norm and Frobenius norm, respectively. We use $\mathsf{vec}(V)$ to denote the vectorization of a matrix $V$ and define $\|V\|_+ := \|V\| + 1$. We define $\underline{\sigma}(V_{1:s}) := \min_{i\in[s]} \underline{\sigma}(V_i)$ and $\|V_{1:s}\|_+ := \max_{i\in[s]} \|V_i\|_+$. We use $I_n$ to denote the identity matrix of dimension $n$. $1_n$ denotes the all 1 vector of dimension $n$ and $1_{\{\cdot\}}$ denotes the indicator function. Lastly, we use $\lesssim$ and $\gtrsim$ for inequalities that hold up to a constant factor.

To begin, we define the following quantity which will be used throughout to quantify the decay of a square matrix $M$.

**Definition A.1** *For a square matrix $M$ with $\rho(M) \le 1$, we have*

$$\tau(M) := \sup_{k\in\mathbb{N}} \{\|M^k\|/\rho(M)^k\}. \tag{A.1}$$

Note that $\tau(M)$ is finite by Gelfand's formula, and it is easy to see that $\tau(M) \ge 1$. This quantity measures the transient response of a non-switching system with state matrix $M$ and can be upper bounded by its $\mathcal{H}_\infty$ norm [62]. In this work, we will mainly use this quantity to evaluate the augmented state matrix for an MJS defined in Section 3.1.

For a Markov chain with transition matrix $T$, we let $_0 \in \mathbb{R}^s$ denote the initial state distribution and $_t$ denote the transient state distribution, i.e. $\mathbb{P}(\omega(t) = i) = _t(i)$. Then, it is easy to see $_t = _0 T^t$. Note that $_t$ is essentially a convex combination of rows of matrix $T^t$, then by triangle inequality, we have $\| _t - _\infty\|_1 \le \max_{i\in[s]} \|([T^t]_{i,\cdot})^\top - _\infty\|_1$. Thus, for an ergodic Markov matrix $T$, we define the following to quantify the convergence of $\| _t - _\infty\|_1$.

**Definition A.2** *For an ergodic Markov matrix $T \in \mathbb{R}^{s\times s}$, let $\tau_{\mathsf{MC}} > 0$ and $\rho_{\mathsf{MC}} \in [0,1)$ be two constants [43, Theorem 4.9] such that*

$$\max_{i\in[s]} \|([T^t]_{i,\cdot})^\top - _\infty\|_1 \le \tau_{\mathsf{MC}} \rho_{\mathsf{MC}}^t. \tag{A.2}$$

*Furthermore, we define*

$$t_{\mathsf{MC}}( ) := \min\left\{ t \in \mathbb{N} : \max_{i\in[s]} \frac{1}{2} \|([T^t]_{i,\cdot})^\top - _\infty\|_1 \le \right\}. \tag{A.3}$$

*When parameter is omitted, it denotes $t_{\mathsf{MC}} := t_{\mathsf{MC}}(\frac{1}{4})$, i.e. the mixing time defined in Section 3.1.*

Note that $\tau(M)$ and $\tau_{\mathsf{MC}}$ have similar roles except $\tau(M)$ is usually used to study state matrices while $\tau_{\mathsf{MC}}$ is for Markov matrices. For $M$, we have $\|M^k\| \le \tau(M)\rho(M)^k$, and for a Markov matrix $\|T^t - 1_s {}_\infty\|_1 \le \tau_{\mathsf{MC}}\rho_{\mathsf{MC}}^t$.

In this section, we define a few notations to ease the exposition in the appendix. Note that, for notations under parameterized form, i.e., notations which are functions of $(\delta, \rho, \tau)$ etc., one can choose these parameters freely to get different deterministic quantities.

Table 2 introduces notations and constants related to the choice of tuning parameter $c_x$, $c_z$, and the shortest trajectory (initial epoch) length such that theoretical performance guarantees can be achieved. Recall that $K_{1:s}^{(0)}$ is the stabilizing controller for epoch 0 in Algorithm 2. We let $L_i^{(0)} := A_i + B_i K_i^{(0)}$, for all $i \in [s]$, denote the closed-loop state matrix, and $\tilde{L}^{(0)} \in \mathbb{R}^{sn^2\times sn^2}$ denotes the augmented closed-loop state matrix

Table 2: Notations — Tuning Parameters and Trajectory Length

| | |
|---|---|
| $\bar{\sigma}_{\mathbf{z}}$ (*depending on context*) | $\sigma_{\mathbf{z}}$ or $\sigma_{\mathbf{z},0}$ or $\sqrt{\|\Sigma_{\mathbf{z}}\|}$ |
| $\bar{\sigma}_{\mathbf{w}}$ (*depending on context*) | $\sigma_{\mathbf{w}}$ or $\sqrt{\|\Sigma_{\mathbf{w}}\|}$ |
| $C_{\mathbf{z}}$ | $\bar{\sigma}_{\mathbf{z}}/\bar{\sigma}_{\mathbf{w}}$ |
| $\bar{\sigma}^2$ | $\|B_{1:s}\|^2\bar{\sigma}_{\mathbf{z}}^2+\bar{\sigma}_{\mathbf{w}}^2$ |
| $\underline{c}_{\mathbf{x}}(\rho,\tau)$ | $3\sqrt{\frac{18n\sqrt{s\tau}\bar{\sigma}^2}{\pi_{\min}\bar{\sigma}_{\mathbf{w}}^2(1-\rho)}}$ |
| $\underline{c}_{\mathbf{z}}$ | $\max\left\{(\sqrt{3}+\sqrt{6})\sqrt{p},\sqrt{3\log(\frac{6}{\pi_{\min}})}\right\}$ |
| $\bar{\tau}$ | $\max\{\tau(\tilde{L}^{(0)}),\tau(\tilde{L}^{\star})\}$ |
| $\bar{\rho}$ | $\max\{\rho(\tilde{L}^{(0)}),\frac{1+\rho^\star}{2}\}$ |
| $C_{MC}$ | $t_{MC}\cdot\max\{3,3-3\log(\pi_{\max}\log(s))\}$ |
| $\underline{T}_{MC,1}(\mathcal{C},\delta)$ | $\left(68\mathcal{C}\pi_{\max}\pi_{\min}^{-2}\log(\frac{s}{\delta})\right)^2$ |
| $\underline{T}_{MC}(\mathcal{C},\delta)$ | $\left(612\mathcal{C}\pi_{\max}\pi_{\min}^{-2}\log(\frac{2s}{\delta})\right)^2$ |
| $\underline{T}_{cl,1}(\rho,\tau)$ | $\frac{(1-\rho)^2}{4n^{1.5}\sqrt{s\tau}\bar{\sigma}^4}$ |
| $\underline{T}_{N}(\mathcal{C},\delta,\rho,\tau)$ | $\max\{\underline{T}_{MC}(\mathcal{C},\frac{\delta}{2}),\underline{T}_{cl,1}(\rho,\tau)\}$ |
| $\underline{T}_{id}(\mathcal{C},\delta,T,\rho,\tau)$ | $\frac{\tau\sqrt{sC\log(T)}}{\pi_{\min}(1-\rho)}\left((\sqrt{2\log(nT)}+\sqrt{2\log(2/\delta)})^2+C_{\mathbf{z}}^2\|B_{1:s}\|^2\right)\log\left(\frac{36s(n+p)C\log(T)}{\delta}\right)(n+p)$ |
| $\underline{T}_{id,N}(L,\delta,T,\rho,\tau)$ | $\max\left\{\underline{T}_{id}(\frac{L}{\log(T)},\frac{\delta}{2L},T,\rho,\tau),\underline{T}_{N}(\frac{L}{\log(T)},\frac{\delta}{2L},\rho,\tau)\right\}$ |
| $\underline{T}_{rgt,^-}(\delta,T)$ | $\mathcal{O}(\frac{\bar{s}(n+p)}{\sqrt{\pi_{\min}}}\bar{}^4_{\mathbf{A},\mathbf{B},\mathbf{T}}\log(\frac{1}{\delta})\log^4(T))$ |
| | $\mathcal{O}(\frac{\bar{s}(n+p)}{\pi_{\min}}\bar{}^2_{\mathbf{A},\mathbf{B},\mathbf{T}}\log(\frac{1}{\delta})\log^2(T))$ (when $B_{1:s}$ is known) |
| $\underline{T}_{\mathbf{x}_0}(\delta)$ | $\frac{1}{\gamma\log(1/\rho)}\max\{\frac{2}{\log(\gamma)},\log(\frac{\pi^2\sqrt{ns\tau}}{3\delta})\}$ |
| $\underline{T}_{rgt}(\delta,T)$ | $\max\left\{\underline{T}_{\mathbf{x}_0}(\delta),\underline{T}_{rgt,^-}(\delta,T),\underline{T}_{MC,1}(\delta),\underline{T}_{id,N}(\underline{L},\delta,T,\bar{\rho},\bar{\tau})\right\}$ |

with $ij$-th $n^2\times n^2$ block given by $[\tilde{L}^{(0)}]_{ij}=[T]_{ji}L_j^{(0)}\otimes L_j^{(0)}$. $\tau(\cdot)$ is as in Definition A.1 and $\rho(\cdot)$ denotes the spectral radius. For the infinite-horizon MJS-LQR($A_{1:s}$, $B_{1:s}$, $T$, $Q_{1:s}$, $R_{1:s}$) problem, we let $P_{1:s}^\star$ denote the solution to cDARE given by (5.2) and $K_{1:s}^\star$ denotes the optimal controller which can be computed via (5.1) with $P_{1:s}^\star$. Similarly, we define $L_{1:s}^\star$ and $\tilde{L}^\star$ to be the corresponding closed-loop state matrix and augmented closed-loop state matrix respectively and $\rho^\star:=\rho(\tilde{L}^\star)$. $\pi_{\max}$ and $\pi_{\min}$ are the largest and smallest elements in the stationary distribution of the ergodic Markov matrix $T$. For the definition of $\underline{T}_{rgt,^-}(\delta,T)$, notation $\bar{}^{}_{\mathbf{A},\mathbf{B},\mathbf{T}}$ is defined in Table 3. As a slight abuse of notation, $T$ in $\underline{T}_{rgt,^-}(\delta,T)$ (as well as Table 4) and $\mathcal{C}$ are merely arguments to be replaced with specific quantities depending on the context.

Table 3 lists the notations related to infinite-horizon MJS perturbation results closely following the notations in [18]. It provides several sensitivity parameters, e.g., how the optimal controller $K_{1:s}^\star$ varies with perturbations in the MJS parameters $A_{1:s}$, $B_{1:s}$, and $T$ and how the MJS-LQR cost $J$ varies with the controller $K_{1:s}$. It also provides certain upper bounds on the variations in $A_{1:s}$, $B_{1:s}$, $T$, and $K_{1:s}$ such that the perturbation theory holds. In this table, $R_{1:s}^{-1}:=\{R_i^{-1}\}_{i=1}^s$ and recall $\|\cdot\|_+:=\|\cdot\|+1$.

## A.1 MJS Covariance Dynamics Under MSS

Consider MJS($A_{1:s}$, $B_{1:s}$, $T$) with process noise $w_t\sim\mathcal{N}(0,\Sigma_{\mathbf{w}})$ and input $u_t=K_{\omega(t)}x_t+z_t$ under a stabilizing controller $K_{1:s}$ and excitation for exploration $z_t\sim\mathcal{N}(0,\Sigma_{\mathbf{z}})$. Let $L_i:=A_i+B_iK_i$ be the closed-loop state matrix. Let $\tilde{L}\in\mathbb{R}^{sn^2\times sn^2}$ be the augmented closed-loop state matrix with $ij$-th $n^2\times n^2$ block given by $[\tilde{L}]_{ij}=[T]_{ji}L_j\otimes L_j$. Let $\tau_{\tilde{L}}>0$ and $\rho_{\tilde{L}}\in[0,1)$ be two constants such that $\|\tilde{L}^k\|\le\tau_{\tilde{L}}\rho_{\tilde{L}}^k$. By definitions of $\tau(\tilde{L})$

Table 3: Notations — MJS-LQR Perturbation

| | |
|---|---|
| $\xi$ | $\min\{\|\mathsf{B}_{1:s}\|_+^{-2}\|\mathsf{R}_{1:s}^{-1}\|_+^{-1}\|\mathsf{L}_{1:s}^{\star}\|_+^{-2}, \underline{\sigma}(\mathsf{P}_{1:s}^{\star})\}$ |
| $\xi'$ | $\|\mathsf{A}_{1:s}\|_+^2\|\mathsf{B}_{1:s}\|_+^4\|\mathsf{P}_{1:s}^{\star}\|_+^3\|\mathsf{R}_{1:s}^{-1}\|_+^2$ |
| $\Gamma_{\star}$ | $\max\{\|\mathsf{A}_{1:s}\|_+, \|\mathsf{B}_{1:s}\|_+, \|\mathsf{P}_{1:s}^{\star}\|_+, \|\mathsf{K}_{1:s}^{\star}\|_+\}$ |
| $C_{\mathbf{A},\mathbf{B},\mathbf{T}}^{\mathbf{K}}$ | $28\sqrt{ns}\tau(\tilde{\mathsf{L}}^{\star})(1-\rho^{\star})^{-1}\left(\underline{\sigma}(\mathsf{R}_{1:s})^{-1}+\Gamma_{\star}^3\underline{\sigma}(\mathsf{R}_{1:s})^{-2}\right)\Gamma_{\star}^3\xi'$ |
| $C_{\mathbf{K}}^{J}$ | $2s^{1.5}\sqrt{n}\min\{n,p\}(\|\mathsf{R}_{1:s}\|+\Gamma_{\star}^3)\frac{\tau(\tilde{\mathsf{L}}^{\star})}{1-\rho^{\star}}$ |
| $\bar{\ }_{\mathbf{K}}$ | $\min\left\{\|\mathsf{K}_{1:s}^{\star}\|, \frac{\sqrt{1-\rho^{\star}}}{2\sqrt{s}\tau(\tilde{\mathsf{L}}^{\star})(1+2\|\mathsf{L}_{1:s}^{\star}\|)\|\mathbf{B}_{1:s}\|}\right\}$ |
| $\bar{\ }_{\mathbf{A},\mathbf{B},\mathbf{T}}^{LQR}$ | $\frac{(1-\rho^{\star})\min\{\Gamma_{\star},\underline{\sigma}(\mathbf{R}_{1:s})^2,\kappa\}}{28\sqrt{ns}\tau(\tilde{\mathsf{L}}^{\star})\Gamma_{\star}^3(\underline{\sigma}(\mathbf{R}_{1:s})+\Gamma_{\star}^3)}\xi'^{-1}$ |
| $\bar{\ }_{\mathbf{A},\mathbf{B},\mathbf{T}}$ | $\min\left\{\frac{\xi(1-\rho^{\star})^2}{204ns\tau(\tilde{\mathsf{L}}^{\star})^2\xi'}, \|\mathsf{B}_{1:s}\|, \underline{\sigma}(\mathsf{Q}_{1:s}), \bar{\ }_{\mathbf{A},\mathbf{B},\mathbf{T}}^{LQR}\right\}$ |

and $\rho(\tilde{\mathsf{L}})$, one can choose them for $\tau_{\tilde{\mathsf{L}}}$ and $\rho_{\tilde{\mathsf{L}}}$ respectively. Let $\Sigma_i(t) := \mathsf{E}[\mathsf{x}_t\mathsf{x}_t^{\top}\mathbf{1}_{\{\omega(t)=i\}}]$, $\Sigma(t) := \mathsf{E}[\mathsf{x}_t\mathsf{x}_t^{\top}]$,

$$\mathsf{s}_t := \begin{bmatrix}\mathrm{vec}(\Sigma_1(t))\\ \vdots \\ \mathrm{vec}(\Sigma_s(t))\end{bmatrix},\ \tilde{\mathsf{B}}_t := \begin{bmatrix}\sum_{j=1}^s \ _{t-1}(j)\mathsf{T}_{j1}(\mathsf{B}_j\otimes\mathsf{B}_j)\\ \vdots \\ \sum_{j=1}^s \ _{t-1}(j)\mathsf{T}_{js}(\mathsf{B}_j\otimes\mathsf{B}_j)\end{bmatrix},\ \text{and}\ \ \tilde{\Pi}_t := \ _t\otimes \mathsf{I}_{n^2}. \tag{A.4}$$

The following lemma shows how $\mathsf{s}_t$ depends on $\mathsf{s}_0$, $\Sigma_{\mathbf{z}}$, and $\Sigma_{\mathbf{w}}$, which will be used to upper bound $\mathsf{E}[\|\mathsf{x}_t\|^2]$ in Lemma A.4.

**Lemma A.3** *The vectorized covariance $\mathsf{s}_t$ has the following dynamics,*

$$\mathsf{s}_t = \tilde{\mathsf{L}}^t\mathsf{s}_0 + (\tilde{\mathsf{B}}_t + \tilde{\mathsf{L}}\tilde{\mathsf{B}}_{t-1} + \cdots + \tilde{\mathsf{L}}^{t-1}\tilde{\mathsf{B}}_1)\mathrm{vec}(\Sigma_{\mathbf{z}}) + (\tilde{\Pi}_t + \tilde{\mathsf{L}}\tilde{\Pi}_{t-1} + \cdots + \tilde{\mathsf{L}}^{t-1}\tilde{\Pi}_1)\mathrm{vec}(\Sigma_{\mathbf{w}}).$$

**Proof** To begin, we evaluate $\Sigma_i(t)$, from the equivalent MJS dynamics $\mathsf{x}_{t+1} = \mathsf{L}_{\omega(t)}\mathsf{x}_t + \mathsf{B}_{\omega(t)}\mathsf{z}_t + \mathsf{w}_t$, as follows,

$$\begin{aligned}\mathsf{E}[\mathsf{x}_{t+1}\mathsf{x}_{t+1}^{\top}\mathbf{1}_{\{\omega(t+1)=i\}}] &= \sum_{j=1}^s \mathsf{E}[\mathsf{L}_j\mathsf{x}_t\mathsf{x}_t^{\top}\mathsf{L}_j\mathbf{1}_{\{\omega(t+1)=i,\omega(t)=j\}}]\\ &\quad + \sum_{j=1}^s \mathsf{E}[\mathsf{B}_j\mathsf{z}_t\mathsf{z}_t^{\top}\mathsf{B}_j^{\top}\mathbf{1}_{\{\omega(t+1)=i,\omega(t)=j\}}] + \mathsf{E}[\mathsf{w}_t\mathsf{w}_t^{\top}\mathbf{1}_{\{\omega(t+1)=i\}}].\end{aligned} \tag{A.5}$$

Since $\mathsf{w}_t \sim \mathcal{N}(0, \Sigma_{\mathbf{w}})$ and $\mathsf{z}_t \sim \mathcal{N}(0, \Sigma_{\mathbf{z}})$, we get

$$\Sigma_i(t+1) = \sum_{j=1}^s \mathsf{T}_{ji}\mathsf{L}_j\Sigma_j(t)\mathsf{L}_j^{\top} + \sum_{j=1}^s \ _t(j)\mathsf{T}_{ji}\mathsf{B}_j\Sigma_{\mathbf{z}}\mathsf{B}_j^{\top} + \ _{t+1}(i)\Sigma_{\mathbf{w}}. \tag{A.6}$$

Vectorizing both sides of the above equation, we have

$$\begin{aligned}\mathrm{vec}(\Sigma_i(t+1)) &= \sum_{j=1}^s \mathsf{T}_{ji}(\mathsf{L}_j\otimes\mathsf{L}_j)\mathrm{vec}(\Sigma_j(t))\\ &\quad + \sum_{j=1}^s \ _t(j)\mathsf{T}_{ji}(\mathsf{B}_j\otimes\mathsf{B}_j)\mathrm{vec}(\Sigma_{\mathbf{z}}) + \ _{t+1}(i)\mathrm{vec}(\Sigma_{\mathbf{w}}).\end{aligned}$$

Stacking this for every $i \in [s]$, we obtain

$$\begin{bmatrix}\mathrm{vec}(\Sigma_1(t+1))\\ \vdots \\ \mathrm{vec}(\Sigma_s(t+1))\end{bmatrix} = \tilde{\mathsf{L}}\begin{bmatrix}\mathrm{vec}(\Sigma_1(t))\\ \vdots \\ \mathrm{vec}(\Sigma_s(t))\end{bmatrix} + \tilde{\mathsf{B}}_{t+1}\mathrm{vec}(\Sigma_{\mathbf{z}}) + \tilde{\Pi}_{t+1}\mathrm{vec}(\Sigma_{\mathbf{w}}). \tag{A.7}$$

Propagating this dynamics from $t$ to 0 gives the desired result. ■

We next provide a key lemma that upper bounds $\mathsf{E}[\|\mathsf{x}_t\|^2]$ and $\|\Sigma(t)\|_F$, which are later used extensively in system identification analysis.

**Lemma A.4** *For $\mathsf{E}[\|\mathsf{x}_t\|^2]$ and $\|\Sigma(t)\|_F$, under MSS given in Definition 3.1, we have*

$$\mathsf{E}[\|\mathsf{x}_t\|^2] \le \sqrt{ns}\cdot\tau_{\tilde{\mathbf{L}}}\rho_{\tilde{\mathbf{L}}}^t\cdot\mathsf{E}[\|\mathsf{x}_0\|^2] + n\sqrt{s}(\|\mathsf{B}_{1:s}\|^2\|\Sigma_{\mathbf{z}}\| + \|\Sigma_{\mathbf{w}}\|)\frac{\tau_{\tilde{\mathbf{L}}}}{1-\rho_{\tilde{\mathbf{L}}}}, \tag{A.8a}$$

$$\|\Sigma(t)\|_F \le \sqrt{s}\cdot\tau_{\tilde{\mathbf{L}}}\rho_{\tilde{\mathbf{L}}}^t\cdot\mathsf{E}[\|\mathsf{x}_0\|^2] + \sqrt{ns}(\|\mathsf{B}_{1:s}\|^2\|\Sigma_{\mathbf{z}}\| + \|\Sigma_{\mathbf{w}}\|)\frac{\tau_{\tilde{\mathbf{L}}}}{1-\rho_{\tilde{\mathbf{L}}}}. \tag{A.8b}$$

**Proof** First we derive an upper bound for $\mathsf{E}[\|\mathsf{x}_t\|^2]$. The upper bound for $\|\Sigma(t)\|_F$ follows similarly. For state $\mathsf{x}_t$, we have

$$\begin{aligned}
\mathsf{E}[\|\mathsf{x}_t\|^2] &= \sum_{i=1}^{s}\mathsf{E}[\|\mathsf{x}_t\|^2\mathbf{1}_{\{\omega(t)=i\}}] = \sum_{i=1}^{s}\mathsf{tr}\left(\mathsf{E}[\mathsf{x}_t\mathsf{x}_t^\top\mathbf{1}_{\{\omega(t)=i\}}]\right) = \sum_{i=1}^{s}\mathsf{tr}(\Sigma_i(t)) \\
&= \sum_{i=1}^{s}\sum_{j=1}^{n}\lambda_j(\Sigma_i(t)) \le \sqrt{ns}\sqrt{\sum_{i=1}^{s}\sum_{j=1}^{n}\lambda_j^2(\Sigma_i(t))} \\
&\le \sqrt{ns}\sqrt{\sum_{i=1}^{s}\|\Sigma_i(t)\|_F^2}.
\end{aligned}$$

Then, by definition of $\mathsf{s}_t$ in (A.4), we have

$$\mathsf{E}[\|\mathsf{x}_t\|^2] \le \sqrt{ns}\|\mathsf{s}_t\|. \tag{A.9}$$

Now, applying the dynamics of $\mathsf{s}_t$ from Lemma A.3, we have

$$\begin{aligned}
\mathsf{E}[\|\mathsf{x}_t\|^2] &\le \sqrt{ns}\left(\|\tilde{\mathbf{L}}^t\|\|\mathsf{s}_0\| + \sum_{t'=1}^{t}\|\tilde{\mathbf{L}}^{t-t'}\|\|\tilde{\mathbf{B}}_{t'}\mathsf{vec}(\Sigma_{\mathbf{z}})\| + \sum_{t'=1}^{t}\|\tilde{\mathbf{L}}^{t-t'}\|\|\tilde{\mathbf{\Pi}}_{t'}\mathsf{vec}(\Sigma_{\mathbf{w}})\|\right) \\
&\le \sqrt{ns}\tau_{\tilde{\mathbf{L}}}\left(\rho_{\tilde{\mathbf{L}}}^t\|\mathsf{s}_0\| + \sum_{t'=1}^{t}\rho_{\tilde{\mathbf{L}}}^{t-t'}\|\tilde{\mathbf{B}}_{t'}\mathsf{vec}(\Sigma_{\mathbf{z}})\| + \sum_{t'=1}^{t}\rho_{\tilde{\mathbf{L}}}^{t-t'}\|\tilde{\mathbf{\Pi}}_{t'}\mathsf{vec}(\Sigma_{\mathbf{w}})\|\right),
\end{aligned} \tag{A.10}$$

where the second line follows from $\|\tilde{\mathbf{L}}^t\| \le \tau_{\tilde{\mathbf{L}}}\rho_{\tilde{\mathbf{L}}}^t$.

Now, we evaluate $\|\mathsf{s}_0\|$, $\|\tilde{\mathbf{B}}_{t'}\mathsf{vec}(\Sigma_{\mathbf{z}})\|$, and $\|\tilde{\mathbf{\Pi}}_{t'}\mathsf{vec}(\Sigma_{\mathbf{w}})\|$ separately. For the first term, we have

$$\|\mathsf{s}_0\| = \sqrt{\sum_{i=1}^{s}\|\Sigma_i(0)\|_F^2} = \sqrt{\sum_{i=1}^{s}{}_i(0)^2\|\mathsf{E}[\mathsf{x}_0\mathsf{x}_0^\top]\|_F^2} \le \|\mathsf{E}[\mathsf{x}_0\mathsf{x}_0^\top]\|_F \le \mathsf{E}[\|\mathsf{x}_0\|^2]. \tag{A.11}$$

Let $[\tilde{\mathbf{B}}_{t'}]_i$ denote the $i$th block of $\tilde{\mathbf{B}}_{t'}$, i.e., $[\tilde{\mathbf{B}}_{t'}]_i = \sum_{j=1}^{s}{}_{t-1}(j)\mathsf{T}_{ji}(\mathsf{B}_j\otimes\mathsf{B}_j)$, then

$$\begin{aligned}
\|\tilde{\mathbf{B}}_{t'}\mathsf{vec}(\Sigma_{\mathbf{z}})\| = \sqrt{\sum_{i=1}^{s}\|[\tilde{\mathbf{B}}_{t'}]_i\mathsf{vec}(\Sigma_{\mathbf{z}})\|^2} &\le \sum_{i=1}^{s}\|[\tilde{\mathbf{B}}_{t'}]_i\mathsf{vec}(\Sigma_{\mathbf{z}})\| \\
&= \sum_{i=1}^{s}\|\sum_{j=1}^{s}{}_{t'-1}(j)\mathsf{T}_{ji}(\mathsf{B}_j\otimes\mathsf{B}_j)\mathsf{vec}(\Sigma_{\mathbf{z}})\| \\
&= \sum_{i=1}^{s}\|\sum_{j=1}^{s}{}_{t'-1}(j)\mathsf{T}_{ji}(\mathsf{B}_j\Sigma_{\mathbf{z}}\mathsf{B}_j^\top)\|_F \\
&\le \|\mathsf{B}_{1:s}\|^2\|\Sigma_{\mathbf{z}}\|\cdot\sum_{i=1}^{s}\|\sum_{j=1}^{s}{}_{t'-1}(j)\mathsf{T}_{ji}\mathsf{I}_n\|_F \\
&= \|\mathsf{B}_{1:s}\|^2\|\Sigma_{\mathbf{z}}\|\cdot\sum_{i=1}^{s}\|{}_{t'}(i)\mathsf{I}_n\|_F \\
&\le \sqrt{n}\|\mathsf{B}_{1:s}\|^2\|\Sigma_{\mathbf{z}}\|.
\end{aligned} \tag{A.12}$$

Lastly, we have

$$\|\tilde{\Pi}_{t'}\mathrm{vec}(\Sigma_{\mathbf{w}})\| = \sqrt{\sum_{i=1}^{s} \| \quad _{t'}(i)\mathrm{vec}(\Sigma_{\mathbf{w}})\|^2} \le \|\mathrm{vec}(\Sigma_{\mathbf{w}})\| = \|\Sigma_{\mathbf{w}}\|_F = \sqrt{n}\|\Sigma_{\mathbf{w}}\|. \tag{A.13}$$

Plugging (A.11)–(A.13) into (A.10), we obtain

$$\begin{aligned} \mathsf{E}[\|\mathsf{x}_t\|^2] &\le \sqrt{ns}\tau_{\tilde{\mathbf{L}}}\left(\rho_{\tilde{\mathbf{L}}}^t \mathsf{E}[\|\mathsf{x}_0\|^2] + \sqrt{n}\|\mathsf{B}_{1:s}\|^2\|\Sigma_{\mathbf{z}}\| \sum_{t'=1}^{t} \rho_{\tilde{\mathbf{L}}}^{t-t'} + \sqrt{n}\|\Sigma_{\mathbf{w}}\| \sum_{t'=1}^{t} \rho_{\tilde{\mathbf{L}}}^{t-t'}\right), \\ &\le \sqrt{ns}\cdot\tau_{\tilde{\mathbf{L}}}\rho_{\tilde{\mathbf{L}}}^t \cdot \mathsf{E}[\|\mathsf{x}_0\|^2] + n\sqrt{s}(\|\mathsf{B}_{1:s}\|^2\|\Sigma_{\mathbf{z}}\| + \|\Sigma_{\mathbf{w}}\|)\frac{\tau_{\tilde{\mathbf{L}}}}{1-\rho_{\tilde{\mathbf{L}}}}, \end{aligned} \tag{A.14}$$

which gives the bound for $\mathsf{E}[\|\mathsf{x}_t\|^2]$ in (A.8a).

To obtain the bound for $\|\Sigma(t)\|_F$ in (A.8b), note that $\|\Sigma(t)\|_F = \|\sum_{i=1}^{s}\Sigma_i(t)\|_F \le \sqrt{s}\sqrt{\sum_{i=1}^{s}\|\Sigma_i(t)\|_F^2} \le \sqrt{s}\|\mathsf{s}_t\|$. We then follow a similar line of reasoning as above to get the statement of the lemma. This completes the proof. ■

## A.2 Supporting Lemmas

In this section, we provide a list of lemmas that will be useful for the subsequent proofs.

**Lemma A.5** *Suppose $\mathsf{z} \sim \mathcal{N}(0, \Sigma_{\mathbf{z}})$ with $\Sigma_{\mathbf{z}} \in \mathsf{R}^{p\times p}$. For any $t \ge (3+2\sqrt{2})p$, we have*

$$\mathsf{P}(\|\mathsf{z}\|^2 \ge 3\|\Sigma_{\mathbf{z}}\|t) \le e^{-t}.$$

**Proof** From [31, Proposition 1], we have for any $t > 0$,

$$\mathsf{P}(\|\mathsf{z}\|^2 \ge \mathsf{tr}(\Sigma_{\mathbf{z}}) + 2\sqrt{\mathsf{tr}(\Sigma_{\mathbf{z}}^2)t} + 2\|\Sigma_{\mathbf{z}}\|t) \le e^{-t},$$

which implies

$$\mathsf{P}(\|\mathsf{z}\|^2 \ge p\|\Sigma_{\mathbf{z}}\| + 2\sqrt{p}\|\Sigma_{\mathbf{z}}\|\sqrt{t} + 2\|\Sigma_{\mathbf{z}}\|t) \le e^{-t}.$$

We can see that when $t \ge (3+2\sqrt{2})p$, we have $p + 2\sqrt{p}\sqrt{t} \le t$, which implies $p\|\Sigma_{\mathbf{z}}\| + 2\sqrt{p}\|\Sigma_{\mathbf{z}}\|\sqrt{t} \le \|\Sigma_{\mathbf{z}}\|t$. Therefore, we have

$$\mathsf{P}(\|\mathsf{z}\|^2 \ge 3\|\Sigma_{\mathbf{z}}\|t) \le e^{-t}.$$

■

**Lemma A.6** *Let $\mathsf{x}_t$ be the MJS state and define the noise-removed state $\tilde{\mathsf{x}}_t = \mathsf{x}_t - \mathsf{w}_{t-1}$ which is independent of $\mathsf{w}_{t-1}$. Let $\tilde{\mathsf{x}}_t$ be zero mean with $\mathsf{E}[\|\tilde{\mathsf{x}}_t\|] \le B$ and $\mathsf{w}_t$ has i.i.d. entries with variance $\sigma_{\mathbf{w}}^2$ bounded in absolute value by $c_{\mathbf{w}}\sigma_{\mathbf{w}}$ for some $c_{\mathbf{w}} > 0$. Consider the conditional random vector*

$$\mathsf{y}_t \sim \{\mathsf{x}_t \mid \|\mathsf{x}_t\| \le 3B\}.$$

*If $c_{\mathbf{w}}\sigma_{\mathbf{w}}\sqrt{n} \le B$, then $\mathrm{Cov}[\mathsf{y}_t\mathsf{y}_t^\square] \ge \sigma_{\mathbf{w}}^2 I_n/2$.*

**Proof** Observe that $\|\mathsf{w}_t\| \le c_{\mathbf{w}}\sigma_{\mathbf{w}}\sqrt{n} \le B$. Define the events

$$E_1 = \{\|\mathsf{x}_t\| \le 3B\},\ E_2 = \{\|\tilde{\mathsf{x}}_t\| \le 2B\}.$$

Clearly $E_2 \subset E_1$ as $\|\mathsf{w}_t\| \le B$. Now, observe that

$$\begin{aligned} \mathrm{Cov}[\mathsf{y}_t\mathsf{y}_t^\square] &= \mathrm{Cov}[\mathsf{y}_t\mathsf{y}_t^\square \mid E_2]\,\mathsf{P}(E_2 \mid E_1) \\ &\ge \mathrm{Cov}[\mathsf{y}_t\mathsf{y}_t^\square \mid E_2]\,\mathsf{P}(E_2). \end{aligned}$$

Note that $\mathsf{P}(E_2) \geq 0.5$ from Markov bound as $\mathsf{E}[\|\tilde{\mathbf{x}}_t\|] \leq B$. Additionally, on the event $E_2$, $\tilde{\mathbf{x}}_t$ and $\mathbf{w}_{t-1}$ are independent. Thus, we further have

$$\begin{aligned}
\mathrm{Cov}[\mathbf{y}_t \mathbf{y}_t^\top] &\geq \mathrm{Cov}[\mathbf{y}_t \mathbf{y}_t^\top \mid E_2]\,\mathsf{P}(E_2) \\
&\geq \mathrm{Cov}[\mathbf{w}_{t-1} \mathbf{w}_{t-1}^\top \mid E_2]\,\mathsf{P}(E_2) \\
&\geq 0.5 \cdot \mathrm{Cov}[\mathbf{w}_{t-1} \mathbf{w}_{t-1}^\top] \\
&\geq \sigma_{\mathbf{w}}^2 I_n/2.
\end{aligned}$$

■

**Lemma A.7** *Let $\mathbf{z} \sim \mathcal{N}(0, \sigma_{\mathbf{z}}^2 I_p)$. Consider the conditional random vector $\mathbf{y} \sim \{\mathbf{z} \mid \|\mathbf{z}\| \leq c\sigma_{\mathbf{z}}\sqrt{p}\}$, where $c \geq 6$ is a fixed constant. Then $\mathrm{Cov}[\mathbf{y}\mathbf{y}^\top] \geq \sigma_{\mathbf{z}}^2 I_p/2$.*

**Proof** This proof gives a lower bound on the covariance of truncated Gaussian vector $\mathbf{z} \mid \|\mathbf{z}\| \leq c\sigma_{\mathbf{z}}\sqrt{p}$. Note that, $\mathbf{z}' = \mathbf{z}/\sigma_{\mathbf{z}}$ is $\mathcal{N}(0, I_p)$. Set variable $X = \|\mathbf{z}'\|^2$. We have the following Lipschitz Gaussian tail bound (we use Lipschitzness of the $_2$ norm and use minor calculus and relaxations)

$$\mathsf{P}(\|\mathbf{z}'\|^2 \geq 4tp) \leq \begin{cases} 1 & \text{if} \quad t \leq 1 \\ e^{-tp/2} & \text{if} \quad t \geq 1. \end{cases}$$

This implies the following tail bound for $X$

$$Q(t) = \mathsf{P}(X \geq t) \leq \begin{cases} 1 & \text{if} \quad t \leq 4p \\ e^{-t/8} & \text{if} \quad t \geq 4p. \end{cases}$$

Fix $\kappa \geq 4$. Using integration-by-parts, this implies that

$$\begin{aligned}
\mathsf{E}[X \mid X \geq \kappa p]\,\mathsf{P}(X \geq \kappa p) &= -\int_{\kappa p}^{\infty} x dQ(x) = -[xQ(x)]_{\kappa n}^{\infty} + \int_{\kappa p}^{\infty} Q(x) dx, \\
&\leq (\kappa p + 8) e^{-\kappa p/8}. \qquad \text{(A.15)}
\end{aligned}$$

The final line as a function of $\kappa p$ is decreasing when $\kappa p \geq 36$. Specifically it is upper bounded by $1/2$ when $\kappa \geq 36$ (as $p \geq 1$). Now define the event

$$E_z = \{\|\mathbf{z}\| \leq \sigma_{\mathbf{z}}\sqrt{\kappa}\sqrt{p}\}.$$

For $\sqrt{\kappa} \geq 6$, $\sqrt{\kappa}$ will map to the $c$ in the statement of the lemma. Observe that this is also the event $X \leq \kappa p$. Following (A.15), this implies

$$\begin{aligned}
\mathsf{E}[\|\mathbf{z}\|^2 \mid E_z^c]\,\mathsf{P}(E_z^c) &\leq \mathsf{E}[\sigma_{\mathbf{z}}^2 X \mid E_z^c]\,\mathsf{P}({E_z}^c) \\
&\leq \sigma_{\mathbf{z}}^2\,\mathsf{E}[X \mid E_z^c]\,\mathsf{P}(E_z^c) \\
&\leq \sigma_{\mathbf{z}}^2/2.
\end{aligned}$$

This also yields the covariance bound of the tail event

$$\mathsf{E}[\mathbf{z}\mathbf{z}^\top \mid E_z^c]\,\mathsf{P}(E_z^c) \leq \mathsf{E}[\|\mathbf{z}\|^2 I_p \mid E_z^c]\,\mathsf{P}(E_z^c) \leq \sigma_{\mathbf{z}}^2 I_p/2.$$

Finally, from the conditional decomposition, observe that

$$\mathsf{E}[\mathbf{z}\mathbf{z}^\top] = \mathsf{E}[\mathbf{z}\mathbf{z}^\top \mid E_z^c]\,\mathsf{P}(E_z^c) + \mathsf{E}[\mathbf{z}\mathbf{z}^\top \mid E_z]\,\mathsf{P}(E_z) \implies \mathsf{E}[\mathbf{z}\mathbf{z}^\top \mid E_z]\,\mathsf{P}(E_z) \geq \sigma_{\mathbf{z}}^2 I_p/2.$$

To conclude, observe that $\mathsf{E}[\mathbf{z}\mathbf{z}^\top \mid E_z] = \mathsf{E}[\mathbf{y}\mathbf{y}^\top]$, where $\mathbf{y}$ is the conditional vector defined by truncating $\mathbf{z}$. Thus, we found

$$\mathsf{E}[\mathbf{y}\mathbf{y}^\top]\,\mathsf{P}(E_z) \geq \sigma_{\mathbf{z}}^2 I_p/2 \implies \mathsf{E}[\mathbf{y}\mathbf{y}^\top] \geq \sigma_{\mathbf{z}}^2 I_p/2.$$

■

**Theorem A.8** *[65, Theorem 5.41 (Isotropic)] Let $\mathbf{X}$ be an $N \times d$ matrix whose rows $\mathbf{x}_i \in \mathbb{R}^d$ are independent isotropic. Let $m$ be such that $\|\mathbf{x}_i\| \le \sqrt{m}$ almost surely for all $i \in [N]$. Then, for every $t \ge 0$, with probability $1 - 2d \cdot e^{-ct^2}$, we have*

$$\sqrt{N} - t\sqrt{m} \le \underline{\sigma}(\mathbf{X}) \le \|\mathbf{X}\| \le \sqrt{N} + t\sqrt{m}.$$

**Corollary A.9 (Non-isotropic)** *Let $\mathbf{X}$ be an $N \times d$ matrix whose rows $\mathbf{x}_i \in \mathbb{R}^d$ are independent with covariance $\mathbf{\Sigma}_i$. Suppose each covariance obeys*

$$\sigma_{\min}^2 \le \underline{\sigma}(\mathbf{\Sigma}_i) \le \|\mathbf{\Sigma}_i\| \le \sigma_{\max}^2.$$

*Let $m$ be such that $\|\mathbf{x}_i\| \le \sqrt{m}$ almost surely for all $i \in [N]$. Then, for every $t \ge 0$, with probability $1 - 2d \cdot e^{-ct^2}$, we have*

$$\sigma_{\min}\sqrt{N} - t\sqrt{m} \le \underline{\sigma}(\mathbf{X}) \le \|\mathbf{X}\| \le \sigma_{\max}\sqrt{N} + t\frac{\sigma_{\max}}{\sigma_{\min}}\sqrt{m}.$$

**Proof** Let $\mathbf{x}_i' = \mathbf{\Sigma}_i^{-1/2}\mathbf{x}_i$. Observe that $\mathbf{x}_i'$ are independent isotropic. Define the matrix $\mathbf{X}'$ with rows $\mathbf{x}'$. Note that $\|\mathbf{x}_i'\| \le \|\mathbf{x}_i\|/\sigma_{\min} \le \sigma_{\min}^{-1}\sqrt{m}$. Thus, applying Theorem A.8 on $\mathbf{X}'$, for every $t \ge 0$, with probability $1 - 2d \cdot e^{-ct^2}$, we have

$$\sqrt{N} - t\sigma_{\min}^{-1}\sqrt{m} \le \underline{\sigma}(\mathbf{X}') \le \|\mathbf{X}'\| \le \sqrt{N} + t\sigma_{\min}^{-1}\sqrt{m}. \tag{A.16}$$

Next, observing that $\mathbf{X}^\top\mathbf{X} = \sum_{i=1}^N \mathbf{x}_i\mathbf{x}_i^\top = \sum_{i=1}^N \sqrt{\mathbf{\Sigma}_i}\mathbf{x}_i'\mathbf{x}_i'^\top\sqrt{\mathbf{\Sigma}_i}$, we find that

$$\begin{aligned}\sigma_{\min}^2\mathbf{X}'^\top\mathbf{X}' = \sigma_{\min}^2\sum_{i=1}^N \mathbf{x}_i'\mathbf{x}_i'^\top \le \mathbf{X}^\top\mathbf{X} &= \sum_{i=1}^N \sqrt{\mathbf{\Sigma}_i}\mathbf{x}_i'\mathbf{x}_i'^\top\sqrt{\mathbf{\Sigma}_i} \\ &\le \sigma_{\max}^2\sum_{i=1}^N \mathbf{x}_i'\mathbf{x}_i'^\top = \sigma_{\max}^2\mathbf{X}'^T\mathbf{X}',\end{aligned}$$

which implies that

$$\sigma_{\min}\underline{\sigma}(\mathbf{X}') \le \underline{\sigma}(\mathbf{X}) \le \|\mathbf{X}\| \le \sigma_{\max}\|\mathbf{X}'\|.$$

Plugging this into (A.16) completes the proof. ∎

# B Sys ID Analysis

We first list in Table 4 a few shorthand notations to be used in this appendix. They are mainly used in the fictional sub-trajectories analysis in Appendices B.2 and B.4. Notations on the inside the parentheses are arguments to be replaced with context-depending variables.

## B.1 Estimating T

The following theorem adapted from [67, Lemma 7] provides the sample complexity result for estimating Markov matrix $\mathbf{T}$, which is a corresponds to the sample complexity on $\|\hat{\mathbf{T}} - \mathbf{T}\|$ in Theorem 4.1.

**Theorem B.1** *Suppose we have an ergodic Markov chain $\mathbf{T} \in \mathbb{R}^{s\times s}$ with mixing time $t_{MC}$ and stationary distribution $_\infty \in \mathbb{R}^s$. Let $\pi_{\min} := \min_{i\in[s]} {}_\infty(i)$ and $\pi_{\max} := \max_{i\in[s]} {}_\infty(i)$. Given a state sequence $\omega(0), \omega(1), \ldots, \omega(T)$ of the Markov chain, define the empirical estimator $\hat{\mathbf{T}}$ of the Markov matrix as follows,*

$$[\hat{\mathbf{T}}]_{ij} = \frac{\sum_{t=1}^{T-1} \mathbf{1}_{\{\omega(t)=i,\omega(t+1)=j\}}}{\sum_{t=1}^{T-1} \mathbf{1}_{\{\omega(t)=i\}}},$$

Table 4: Notations — Sampling Periods

| | |
|---|---|
| $\underline{c}_{\mathbf{w}}(T,\delta)$ | $\sqrt{2\log(nT)}+\sqrt{2\log(2/\delta)}$ |
| $\beta_+(\rho,\tau,c_{\mathbf{w}})$ | $\sqrt{\frac{2\sqrt{\bar{s}}(c_{\mathbf{w}}^2+C_{\mathbf{z}}^2\|\mathbf{B}_{1:s}\|^2)\tau}{1-\rho}}$ |
| $\beta_+'(\rho,\tau,c_{\mathbf{w}},\mathsf{K}_{1:s})$ | $c_{\mathbf{w}}+\beta_+(\rho,\tau,c_{\mathbf{w}})(\|\mathsf{A}_{1:s}\|+\|\mathsf{B}_{1:s}\|\|\mathsf{K}_{1:s}\|)+C_{\mathbf{z}}\sqrt{p/n}\|\mathsf{B}_{1:s}\|$ |
| $\underline{C}_{sub,\mathbf{x}}(\bar{x}_0,\delta,T,\rho,\tau)$ | $\frac{2}{\log(\rho^{-1})}+\frac{2}{\log(\rho^{-1})\log(T)}\log\left(\frac{24n\sqrt{\bar{s}}\tau\max\{\bar{x}_0^2,\bar{\sigma}^2\}}{\delta(1-\rho)}\right)\Big\}$ |
| $\underline{C}_{sub,\bar{\mathbf{x}}}(\delta,T,\rho,\tau)$ | $\frac{1}{\log(\rho^{-1})}+\frac{1}{\log(\rho^{-1})\log(T)}\log(\frac{72\sqrt{n\bar{s}}s^{1.5}\tau}{\delta})$ |
| $\underline{C}_{sub,N}(\bar{x}_0,\delta,T,\rho,\tau)$ | $\max\left\{C_{MC},\underline{C}_{sub,\mathbf{x}}(\bar{x}_0,\frac{\delta}{2},T,\rho,\tau),\underline{C}_{sub,\bar{\mathbf{x}}}(\delta,T,\rho,\tau)\right\}$ |
| $\underline{L}_{id,t_0}(\bar{x}_0,\rho,\tau,c_{\mathbf{w}})$ | $\frac{\log\left((1-\rho)\bar{x}_0^2/(c_{\mathbf{w}}^2\bar{\sigma}_{\mathbf{w}}^2+\bar{\sigma}_{\mathbf{z}}^2\|\mathbf{B}_{1:s}\|)\right)}{1-\rho}$ |
| $\underline{L}_{id,cov}(\rho,\tau,c_{\mathbf{w}},\mathsf{K}_{1:s})$ | $1+\frac{2\log\left(8c^2\beta_+(\rho,\tau,c_{\mathbf{w}})\beta_+'(\rho,\tau,c_{\mathbf{w}},\mathbf{K}_{1:s})n\sqrt{n\bar{s}}\tau\right)}{1-\rho}$ |
| $\underline{L}_{id,tr1}(\delta,T,\rho,\tau,c_{\mathbf{w}},\mathsf{K}_{1:s})$ | $1+\frac{2\log\left(2\sqrt{n\bar{s}}\tau T\beta_+'(\rho,\tau,c_{\mathbf{w}},\mathbf{K}_{1:s})/(\beta_+(\rho,\tau,c_{\mathbf{w}})\delta)\right)}{1-\rho}$ |
| $\underline{L}_{id,tr2}(\delta,T,\rho,\tau,c_{\mathbf{w}},\mathsf{K}_{1:s})$ | $1+\frac{2}{(1-\rho)}\log\left(\frac{192c^2\tau\beta_+'(\rho,\tau,c_{\mathbf{w}},\mathbf{K}_{1:s})(1+\beta_+(\rho,\tau,c_{\mathbf{w}}))n\sqrt{\bar{s}(n+p)}T}{\pi_{\min}\delta}\right)$ |
| $\underline{L}_{id,tr3}(\delta,T,\rho,\tau,c_{\mathbf{w}},\mathsf{K}_{1:s})$ | $1+\frac{2}{(1-\rho)}\log\left(\frac{c_{\mathbf{w}}\sigma_{\mathbf{w}}\beta_+'(\rho,\tau,c_{\mathbf{w}},\mathbf{K}_{1:s})\tau n\sqrt{n\bar{s}}T^2}{\delta(1+\beta_+(\rho,\tau,c_{\mathbf{w}}))\sqrt{(n+p)}\left(C\sigma_{\mathbf{w}}\sqrt{n+p}+C_0\sqrt{\log(2s/\delta)}\right)}\right)$ |
| $\underline{L}_{id}(\bar{x}_0,\delta,T,\rho,\tau,c_{\mathbf{w}},\mathsf{K}_{1:s},L)$ | $\max\left\{\underline{L}_{id,t_0}(\bar{x}_0,\rho,\tau,c_{\mathbf{w}}),\underline{L}_{id,cov}(\rho,\tau,c_{\mathbf{w}},\mathsf{K}_{1:s}),\right.$<br>$\underline{L}_{id,tr1}(\frac{\delta}{36L},T,\rho,\tau,c_{\mathbf{w}},\mathsf{K}_{1:s}),\underline{L}_{id,tr2}(\frac{\delta}{36L},T,\rho,\tau,c_{\mathbf{w}},\mathsf{K}_{1:s}),$<br>$\left.\underline{L}_{id,tr3}(\frac{\delta}{36L},T,\rho,\tau,c_{\mathbf{w}},\mathsf{K}_{1:s}),\underline{C}_{sub,N}\left(\bar{x}_0,\frac{\delta}{2L},T,\rho,\tau\right)\log(T)\right\}$ |

*Assume for some $\delta>0$, $T\geq \underline{T}_{MC,1}(C_{MC},\frac{\delta}{4}):=\left(68C_{MC}\pi_{\max}\pi_{\min}^{-2}\log(\frac{4s}{\delta})\right)^2$, where $C_{MC}$ is defined in Table 2. Then, we have with probability at least $1-\delta$,*

$$\|\hat{\mathsf{T}}-\mathsf{T}\|\leq 4\pi_{\min}^{-1}\|\mathsf{T}\|\sqrt{\frac{17\pi_{\max}C_{MC}\log(T)}{T}\log\left(\frac{4sC_{MC}\log(T)}{\delta}\right)}. \tag{B.1}$$

**Proof** We first consider the estimators computed using a sub-trajectory of $\omega(0),\omega(1),\ldots,\omega(T)$, then combine them together to show the error bound for $\hat{\mathsf{T}}$ in the claim. For $C_{MC}$ defined in Table 2, let $L=C_{MC}\log(T)$. Then, for $l=0,1,\ldots,L-1$, define $\hat{\mathsf{T}}^{(l)}\in\mathsf{R}^{s\times s}$ such that $[\hat{\mathsf{T}}^{(l)}]_{ij}=\frac{\sum_{k=1}^{\lfloor T/L\rfloor}\mathbf{1}_{\{\omega(kL+l)=i,\omega(kL+1+l)=j\}}}{\sum_{k=1}^{\lfloor T/L\rfloor}\mathbf{1}_{\{\omega(kL+l)=i\}}}$. In other words, $\hat{\mathsf{T}}^{(l)}$ is the estimator computed using data with subsampling period $L$. Following the proof of [67, Lemma 7], we know for any $\epsilon<\pi_{\min}/2$, suppose $L\geq 6t_{MC}\log(\epsilon^{-1})$.

$$\mathsf{P}\left(\|\hat{\mathsf{T}}^{(l)}-\mathsf{T}\|\leq 4\pi_{\min}^{-1}\|\mathsf{T}\|\epsilon\right)\geq 1-4s\exp\left(-\frac{T\epsilon^2}{17\pi_{\max}L}\right). \tag{B.2}$$

By setting $\delta=4s\exp\left(-\frac{T\epsilon^2}{17\pi_{\max}L}\right)$, one can also interpret the above result as: for all $\delta>0$, suppose

$$L\geq 3t_{MC}\log\left(\frac{T}{17\pi_{\max}L\log(\frac{4s}{\delta})}\right), \tag{B.3}$$

then when

$$T\geq 68L\pi_{\max}\pi_{\min}^{-2}\log(\frac{4s}{\delta}), \tag{B.4}$$

we have with probability at least $1-\delta$

$$\|\hat{\mathsf{T}}^{(l)}-\mathsf{T}\|\leq 4\pi_{\min}^{-1}\|\mathsf{T}\|\sqrt{\frac{17\pi_{\max}C_{MC}\log(T)\log(\frac{4s}{\delta})}{T}}. \tag{B.5}$$

One can verify (B.3) holds by plugging in $L = C_{MC}\log(T)$ and using definition $C_{MC} := t_{MC}\cdot\max\{3, 3 - 3\log(\pi_{\max}\log(s))\}$; (B.4) holds under the premise condition $T \geq \underline{T}_{MC,1}(C_{MC}, \frac{\delta}{4}) := \left(68C_{MC}\pi_{\max}\pi_{\min}^{-2}\log(\frac{4s}{\delta})\right)^2$.

Note that by definition, $\hat{\mathsf{T}}$ can be viewed as a convex combination of $\hat{\mathsf{T}}^{(l)}$ for all $l = 0, 1, \ldots, L$, thus by triangle inequality and union bound, we have with probability $1 - L\delta$,

$$\|\hat{\mathsf{T}} - \mathsf{T}\| \leq 4\pi_{\min}^{-1}\|\mathsf{T}\|\sqrt{\frac{17\pi_{\max}C_{MC}\log(T)\log(\frac{4s}{\delta})}{T}}. \tag{B.6}$$

Finally, by replacing $L\delta$ with $\delta$, we could show (B.1) and conclude the proof. ∎

## B.2 Estimation of $\mathsf{A}_{1:s}$ and $\mathsf{B}_{1:s}$ from a Single Trajectory (Main SYSID Analysis)

In this section, we estimate the MJS dynamics $\mathsf{A}_{1:s}$ and $\mathsf{B}_{1:s}$ from finite samples obtained from a single trajectory of (3.1). To estimate a coarse model of the unknown system dynamics, we use the method of linear least squares. By running experiments in which the system starts at state $\mathsf{x}_0$ and the dynamics evolve with a given input, we can record the resulting state, excitation and mode observations. Let $\mathsf{K}_{1:s}$ stabilizes the system (3.1) in the mean square sense according to Def. 3.1. Then, choosing the input to be $\mathsf{u}_t = \mathsf{K}_{\omega(t)}\mathsf{x}_t + \mathsf{z}_t$, the state updates as follows,

$$\mathsf{x}_{t+1} = (\mathsf{A}_{\omega(t)} + \mathsf{B}_{\omega(t)}\mathsf{K}_{\omega(t)})\mathsf{x}_t + \mathsf{B}_{\omega(t)}\mathsf{z}_t + \mathsf{w}_t = \mathsf{L}_{\omega(t)}\mathsf{x}_t + \mathsf{B}_{\omega(t)}\mathsf{z}_t + \mathsf{w}_t, \tag{B.7}$$

where $\{\mathsf{z}_t\}_{t=0}^{\infty} \overset{\text{i.i.d.}}{\sim} \mathcal{N}(0, \sigma_{\mathsf{z}}^2 I_p)$ is the i.i.d. excitation for exploration and we let $\mathsf{L}_{\omega(t)} := \mathsf{A}_{\omega(t)} + \mathsf{B}_{\omega(t)}\mathsf{K}_{\omega(t)}$. Observe that the closed-loop state update (B.7) can be expanded as follows,

$$\mathsf{x}_t = \begin{cases} \mathsf{x}_0 & \text{if} \quad t = 0, \\ \mathsf{L}_{\omega(0)}\mathsf{x}_0 + \mathsf{B}_{\omega(0)}\mathsf{z}_0 + \mathsf{w}_0 & \text{if} \quad t = 1, \\ \prod_{j=0}^{t-1}\mathsf{L}_{\omega(j)}\mathsf{x}_0 + \sum_{t'=0}^{t-2}\prod_{j=t'+1}^{t-1}\mathsf{L}_{\omega(j)}\mathsf{B}_{\omega(t')}\mathsf{z}_{t'} + \mathsf{B}_{\omega(t-1)}\mathsf{z}_{t-1} & \\ \qquad + \sum_{t'=0}^{t-2}\prod_{j=t'+1}^{t-1}\mathsf{L}_{\omega(j)}\mathsf{w}_{t'} + \mathsf{w}_{t-1} & \text{if} \quad t \geq 2. \end{cases} \tag{B.8}$$

### B.2.1 Estimation from Bounded States

To estimate the unknown system dynamics, we run the system for $T$ time-steps and collect the samples $(\omega(t), \mathsf{z}_t, \mathsf{x}_t, \mathsf{x}_{t+1})_{t=0}^{T-1}$. Then, we run Alg. 1 to get the estimates $(\hat{\mathsf{A}}_i, \hat{\mathsf{B}}_i)$ for all $i \in [s]$. Our learning method is described in Alg. 1. For the ease of analysis, we first derive the estimation error bounds with the following assumption on the noise.

**Assumption 3 (subGaussian noise)** *Let $\{\mathsf{w}_t\}_{t=0}^{T-1} \overset{i.i.d.}{\sim} \mathcal{D}_w$. There exists $\sigma_{\mathsf{w}} > 0$ and $c_{\mathsf{w}} \geq 1$ such that, each entry of $\mathsf{w}_t$ is i.i.d. zero-mean subGaussian with variance $\sigma_{\mathsf{w}}^2$ and we have $\|\mathsf{w}_t\|_\infty \leq c_{\mathsf{w}}\sigma_{\mathsf{w}}$.*

Later on, we will relax this assumption to get the estimation error bounds with the Gaussian noise. To proceed, we first show that the Euclidean norm of the states $\mathsf{x}_t$ in (B.7) can be upper bounded in expectation. The following result, which is a corollary of Lemma A.4, accomplishes this.

**Corollary B.2 (Bounded states)** *Let $\mathsf{x}_t$ be the state at time $t$ of the MJS (B.7), with initial state $\mathsf{x}_0 \sim \mathcal{D}_x$ such that $\mathsf{E}[\mathsf{x}_0] = 0$, $\mathsf{E}[\|\mathsf{x}_0\|^2] \leq \beta_0^2 n$ for some $\beta_0 > 0$. Suppose Assumption 1 on the system and the Markov chain and Assumption 3 on the process noise hold. Suppose $\{\mathsf{z}_t\}_{t=0}^{\infty} \overset{i.i.d.}{\sim} \mathcal{N}(0, \sigma_{\mathsf{z}}^2 I_p)$. Let $C_{\mathsf{z}} := \frac{\sigma_{\mathsf{z}}}{\sigma_{\mathsf{w}}}$ be a constant, $\mathsf{h}_t := [\frac{1}{\sigma_{\mathsf{w}}}\mathsf{x}_t^\top \ \frac{1}{\sigma_{\mathsf{z}}}\mathsf{z}_t^\top]^\top$ be the concatenated state and define*

$$t_0 := \frac{\log\left((1-\rho_{\tilde{\mathsf{L}}})\beta_0^2 n/(c_{\mathsf{w}}^2\sigma_{\mathsf{w}}^2 + \sigma_{\mathsf{z}}^2\|\mathsf{B}_{1:s}\|)\right)}{1-\rho_{\tilde{\mathsf{L}}}}, \tag{B.9}$$

$$\beta_+^2 := \frac{2\sqrt{s}(c_{\mathsf{w}}^2 + C_{\mathsf{z}}^2\|\mathsf{B}_{1:s}\|^2)\tau_{\tilde{\mathsf{L}}}}{1-\rho_{\tilde{\mathsf{L}}}}. \tag{B.10}$$

*Then, for all $t \geq t_0$, we have*

$$\mathsf{E}[\|\mathsf{x}_t\|^2] \leq \sigma_{\mathbf{w}}^2 \beta_+^2 n \quad \textit{and} \quad \mathsf{E}[\|\mathsf{h}_t\|^2] \leq (1+\beta_+^2)(n+p). \tag{B.11}$$

**Proof** Recall from Lemma A.4 that the states $\mathsf{x}_t$ can be bounded in expectation as follows,

$$\mathsf{E}[\|\mathsf{x}_t\|^2] \leq \tau_{\tilde{\mathbf{L}}} \sqrt{s} \left( \rho_{\tilde{\mathbf{L}}}^t \beta_0^2 \sqrt{n} + \frac{c_{\mathbf{w}}^2 \sigma_{\mathbf{w}}^2 + \sigma_{\mathbf{z}}^2 \|\mathsf{B}_{1:s}\|^2}{1-\rho_{\tilde{\mathbf{L}}}} \right) n \leq \frac{2\sigma_{\mathbf{w}}^2 \sqrt{s}(c_{\mathbf{w}}^2 + C_{\mathbf{z}}^2 \|\mathsf{B}_{1:s}\|^2)\tau_{\tilde{\mathbf{L}}} n}{1-\rho_{\tilde{\mathbf{L}}}}, \tag{B.12}$$

where we get the last inequality by choosing the timestep $t$ to satisfy the following lower bound,

$$\rho_{\tilde{\mathbf{L}}}^t \leq \frac{c_{\mathbf{w}}^2 \sigma_{\mathbf{w}}^2 + \sigma_{\mathbf{z}}^2 \|\mathsf{B}_{1:s}\|^2}{(1-\rho_{\tilde{\mathbf{L}}})\beta_0^2 \sqrt{n}} \impliedby t \geq t_0 := \frac{\log\left((1-\rho_{\tilde{\mathbf{L}}})\beta_0^2 n / (c_{\mathbf{w}}^2 \sigma_{\mathbf{w}}^2 + \sigma_{\mathbf{z}}^2 \|\mathsf{B}_{1:s}\|)\right)}{1-\rho_{\tilde{\mathbf{L}}}}. \tag{B.13}$$

This gives the advertised upper bound on $\mathsf{E}[\|\mathsf{x}_t\|^2]$ for $t \geq t_0$. Using Jensen's inequality, this further implies

$$\mathsf{E}[\|\mathsf{x}_t\|] \leq \sigma_{\mathbf{w}} \sqrt{\frac{2s^{1/2}(c_{\mathbf{w}}^2 + C_{\mathbf{z}}^2 \|\mathsf{B}_{1:s}\|^2)\tau_{\tilde{\mathbf{L}}} n}{1-\rho_{\tilde{\mathbf{L}}}}} \quad \text{for} \quad t \geq t_0. \tag{B.14}$$

Next, using standard results on the distribution of squared Euclidean norm of a Gaussian vector, we have $\mathsf{E}[\|\mathsf{z}_t\|^2] = \sigma_{\mathbf{z}}^2 p$ for all $t \geq 0$. Combining this with (B.12) , we get the following upper bound on the expected squared norm of $\mathsf{h}_t := [\frac{1}{\sigma_{\mathbf{w}}} \mathsf{x}_t^\top \ \frac{1}{\sigma_{\mathbf{z}}} \mathsf{z}_t^\top]^\top$, that is, for all $t \geq t_0$, we have

$$\begin{aligned} \mathsf{E}[\|\mathsf{h}_t\|^2] = \frac{1}{\sigma_{\mathbf{w}}^2} \mathsf{E}[\|\mathsf{x}_t\|^2] + \frac{1}{\sigma_{\mathbf{z}}^2} \mathsf{E}[\|\mathsf{z}_t\|^2] &\leq \frac{2\sqrt{s}(c_{\mathbf{w}}^2 + C_{\mathbf{z}}^2 \|\mathsf{B}_{1:s}\|^2)\tau_{\tilde{\mathbf{L}}} n}{1-\rho_{\tilde{\mathbf{L}}}} + p, \\ &\leq \left(1 + \frac{2\sqrt{s}(c_{\mathbf{w}}^2 + C_{\mathbf{z}}^2 \|\mathsf{B}_{1:s}\|^2)\tau_{\tilde{\mathbf{L}}}}{1-\rho_{\tilde{\mathbf{L}}}}\right)(n+p). \end{aligned} \tag{B.15}$$

This gives the advertised upper bound on $\mathsf{E}[\|\mathsf{h}_t\|^2]$ for $t \geq t_0$. Using Jensen's inequality, this further implies

$$\mathsf{E}[\|\mathsf{h}_t\|] \leq \sqrt{\left(1 + \frac{2s^{1/2}(c_{\mathbf{w}}^2 + C_{\mathbf{z}}^2 \|\mathsf{B}_{1:s}\|^2)\tau_{\tilde{\mathbf{L}}}}{1-\rho_{\tilde{\mathbf{L}}}}\right)(n+p)} \quad \text{for} \quad t \geq t_0. \tag{B.16}$$

This completes the proof. ∎

Suppose the MJS in (B.7) is run for $T$ timesteps and we have access to the trajectory $(\mathsf{x}_t, \mathsf{z}_t, \omega(t))_{t=0}^T$. Our proof strategy is based on the observation that a single MJS trajectory can be split into multiple weakly dependent sub-trajectories, defined as follows.

**Definition B.3 (Sub-trajectories of bounded states)** *Let sampling period $L \geq 1$ be an integer. Let $\tau_k = \tau + kL$ be the sub-sampling indices, where $0 \leq \tau \leq L-1$ is a fixed o set and $k = 1, 2, \ldots, \lfloor \frac{T-L}{L} \rfloor$. We sub-sample the trajectory $(\mathsf{x}_t, \mathsf{z}_t, \omega(t))_{t=0}^T$ at time indices $\tau_k \in S_i^{(\tau)}$, where*

$$S_i^{(\tau)} := \{\tau_k \mid \omega(\tau_k) = i, \ \|\mathsf{x}_{\tau_k}\| \leq c\sigma_{\mathbf{w}} \beta_+ \sqrt{n}, \ \|\mathsf{z}_{\tau_k}\| \leq c\sigma_{\mathbf{z}} \sqrt{p}\}, \tag{B.17}$$

*to obtain the $\tau_{th}$ sub-trajectory $\{(\mathsf{x}_{\tau_k+1}, \mathsf{x}_{\tau_k}, \mathsf{z}_{\tau_k}, \omega(\tau_k))\}_{\tau_k \in S_i^{(\tau)}}$.*

Note that, $S_i = \bigcup_{\tau=0}^{L-1} S_i^{(\tau)}$, where $S_i$ is as defined in Alg. 1. This shows that a single trajectory with bounded states and excitations $\{(\mathsf{x}_{t+1}, \mathsf{x}_t, \mathsf{z}_t, \omega(t))\}_{t \in S_i}$ can be split into $L$ weakly dependent sub-trajectories $\{(\mathsf{x}_{\tau_k+1}, \mathsf{x}_{\tau_k}, \mathsf{z}_{\tau_k}, \omega(\tau_k))\}_{\tau_k \in S_i^{(\tau)}}$ for $0 \leq \tau \leq L-1$. To proceed, we first show that the covariance of the bounded states has the following properties.

**Lemma B.4 (Covariance of bounded states)** *Consider the same setup of Corollary B.2. Let $t_0$ and $\beta_+$ be as in (B.9) and (B.10) respectively. Let $c \geq 6$ be a fixed constant. For all $t \geq t_0$, we have*

$$(\sigma_{\mathbf{w}}^2/2) I_n \preceq \Sigma[\mathsf{x}_t \mid \|\mathsf{x}_t\| \leq c\sigma_{\mathbf{w}}\beta_+\sqrt{n},\ \ \|\mathsf{z}_t\| \leq c\sigma_{\mathbf{z}}\sqrt{p}] \preceq c^2\sigma_{\mathbf{w}}^2\beta_+^2 n I_n, \tag{B.18}$$

$$(1/2) I_{n+p} \preceq \Sigma[\mathsf{h}_t \mid \|\mathsf{x}_t\| \leq c\sigma_{\mathbf{w}}\beta_+\sqrt{n},\ \ \|\mathsf{z}_t\| \leq c\sigma_{\mathbf{z}}\sqrt{p}] \preceq c^2(1+\beta_+^2)(n+p) I_{n+p}. \tag{B.19}$$

**Proof** Let $\mathsf{x}_t$ be the state at time $t$ of the MJS given by (B.7), with initial state $\mathsf{x}_0 \sim \mathcal{D}_x$ such that $\mathsf{E}[\mathsf{x}_0] = 0$ and define the noise-removed state $\tilde{\mathsf{x}}_t = \mathsf{x}_t - \mathsf{w}_{t-1}$ which is independent of $\mathsf{w}_{t-1}$. Observe that $\mathsf{E}[\tilde{\mathsf{x}}_t] = 0$ because both $\mathsf{x}_t$ and $\mathsf{w}_{t-1}$ are zero-mean. Next, from Corollary B.2, we know that, for all $t \geq t_0$, we have $\mathsf{E}[\|\mathsf{x}_t\|] \leq \sigma_{\mathbf{w}}\beta_+\sqrt{n}$. Combining this with $\mathsf{E}[\|\mathsf{w}_{t-1}\|] \leq c_{\mathbf{w}}\sigma_{\mathbf{w}}\sqrt{n}$, we have

$$\mathsf{E}[\|\tilde{\mathsf{x}}_t\|] \leq \mathsf{E}[\|\mathsf{x}_t\|] + \mathsf{E}[\|\mathsf{w}_{t-1}\|] \leq 2\sigma_{\mathbf{w}}\beta_+\sqrt{n}. \tag{B.20}$$

To proceed, consider the conditional random variable $\mathsf{y}_t \sim \{\mathsf{x}_t \mid \|\mathsf{x}_t\| \leq c\sigma_{\mathbf{w}}\beta_+\sqrt{n}\}$. To lower bound the covariance matrix $\Sigma[\mathsf{y}_t]$, observe that $\|\mathsf{w}_{t-1}\| \leq c_{\mathbf{w}}\sigma_{\mathbf{w}}\sqrt{n} \leq \sigma_{\mathbf{w}}\beta_+\sqrt{n}$ and $\mathsf{E}[\|\tilde{\mathsf{x}}_t\|] \leq 2\sigma_{\mathbf{w}}\beta_+\sqrt{n}$, where $\beta_+$ is given by (B.10). Therefore, using Lemma A.6 from Section A.2 with $B = 2\sigma_{\mathbf{w}}\beta_+\sqrt{n}$, we can lower bound $\Sigma[\mathsf{y}_t]$ as follows,

$$\Sigma[\mathsf{y}_t] = \Sigma[\mathsf{x}_t \mid \|\mathsf{x}_t\| \leq c\sigma_{\mathbf{w}}\beta_+\sqrt{n}] \succeq (\sigma_{\mathbf{w}}^2/2) I_n, \tag{B.21}$$

where we use $c \geq 6$ to get the last inequality. Next, we upper bound the covariance matrix $\Sigma[\mathsf{y}_t]$ as follows,

$$\|\Sigma[\mathsf{y}_t]\| = \|\mathsf{E}[\mathsf{y}_t\mathsf{y}_t^\top]\| \leq \mathsf{E}[\|\mathsf{y}_t\|^2] = \mathsf{E}[\|\mathsf{x}_t\|^2 \mid \|\mathsf{x}_t\| \leq c\sigma_{\mathbf{w}}\beta_+\sqrt{n}] \leq c^2\sigma_{\mathbf{w}}^2\beta_+^2 n, \tag{B.22}$$

Combining (B.21) and (B.22), we get the first statement of the lemma. Next, using a similar argument with Lemma A.7, we can show that, when $c \geq 6$, we also have

$$(\sigma_{\mathbf{z}}^2/2) I_p \preceq \Sigma[\mathsf{z}_t \mid \|\mathsf{z}_t\| \leq c\sigma_{\mathbf{z}}\sqrt{p}] \preceq (2\sigma_{\mathbf{z}}^2 p) I_p. \tag{B.23}$$

Finally, combing the derived bounds on $\Sigma[\mathsf{x}_t \mid \|\mathsf{x}_t\| \leq c\sigma_{\mathbf{w}}\beta_+\sqrt{n}]$ and $\Sigma[\mathsf{z}_t \mid \|\mathsf{z}_t\| \leq c\sigma_{\mathbf{z}}\sqrt{p}]$, we get the second statement of the lemma,

$$(1/2) I_{n+p} \preceq \Sigma[\mathsf{h}_t \mid \|\mathsf{x}_t\| \leq c\sigma_{\mathbf{w}}\beta_+\sqrt{n}, \|\mathsf{z}_t\| \leq c\sigma_{\mathbf{z}}\sqrt{p}] \preceq c^2(1+\beta_+^2)(n+p) I_{n+p}. \tag{B.24}$$

This completes the proof. ∎

To proceed, let $\mathsf{h}_t := [\frac{1}{\sigma_{\mathbf{w}}}\mathsf{x}_t^\top \ \frac{1}{\sigma_{\mathbf{z}}}\mathsf{z}_t^\top]^\top$ be the concatenated state and $\Theta_i^\top := [\sigma_{\mathbf{w}} L_i \ \ \sigma_{\mathbf{z}} B_i]$ for all $i \in [s]$. Then the output of each sample in $\{(\mathsf{x}_{t+1}, \mathsf{x}_t, \mathsf{z}_t, \omega(t))\}_{t \in S_i}$ can be related to the inputs as follows,

$$\mathsf{x}_{t+1} = \Theta_{\omega(t)}^\top \mathsf{h}_t + \mathsf{w}_t \quad \text{for all} \quad t \in S_i. \tag{B.25}$$

Next, to carry out finite sample identification of $\Theta_i^\top$ using the method of linear least squares, we define the following concatenated matrices,

$$\mathsf{Y}_i = \begin{bmatrix} \mathsf{x}_{t_1+1}^\top \\ \mathsf{x}_{t_2+1}^\top \\ \vdots \\ \mathsf{x}_{t_{|S_i|}+1}^\top \end{bmatrix}, \quad \mathsf{H}_i = \begin{bmatrix} \mathsf{h}_{t_1}^\top \\ \mathsf{h}_{t_2}^\top \\ \vdots \\ \mathsf{h}_{t_{|S_i|}}^\top \end{bmatrix}, \quad \mathsf{W}_i = \begin{bmatrix} \mathsf{w}_{t_1}^\top \\ \mathsf{w}_{t_2}^\top \\ \vdots \\ \mathsf{w}_{t_{|S_i|}}^\top \end{bmatrix}, \tag{B.26}$$

that is, $\mathsf{Y}_i$ has $\{\mathsf{x}_{t+1}^\top\}_{t \in S_i}$ on its rows, $\mathsf{H}_i$ has $\{\mathsf{h}_t^\top\}_{t \in S_i}$ on its rows and $\mathsf{W}_i$ has $\{\mathsf{w}_t^\top\}_{t \in S_i}$ on its rows. Similarly, we also construct $\mathsf{Y}_i^{(\tau)}$, $\mathsf{H}_i^{(\tau)}$ and $\mathsf{W}_i^{(\tau)}$ by (row-wise) stacking $\{\mathsf{x}_{\tau_k+1}^\top\}_{\tau_k \in S_i^{(\tau)}}$, $\{\mathsf{h}_{\tau_k}^\top\}_{\tau_k \in S_i^{(\tau)}}$ and $\{\mathsf{w}_{\tau_k}^\top\}_{\tau_k \in S_i^{(\tau)}}$ respectively. Then, we have $\mathsf{Y}_i = \mathsf{H}_i\Theta_i^{\top\top} + \mathsf{W}_i$. Our goal in this paper is to solve the following least squares problems,

$$\hat{\Theta}_i^\top = \arg\min_{\Theta_i} \frac{1}{2|S_i|}\|\mathsf{Y}_i - \mathsf{H}_i\Theta_i^\top\|_F^2. \tag{B.27}$$

The least squares estimator of $\Theta_i^{\square\square}$ is $\hat{\Theta}_i^{\square} = \mathsf{H}_i^{\dagger}\mathsf{Y}_i = \left(\mathsf{H}_i^{\square}\mathsf{H}_i\right)^{-1}\mathsf{H}_i^{\square}\mathsf{Y}_i$ and its estimation error is given by

$$\begin{aligned}\|\hat{\Theta}_i - \Theta_i^{\square}\| = \|\left(\mathsf{H}_i^{\square}\mathsf{H}_i\right)^{-1}\mathsf{H}_i^{\square}\mathsf{W}_i\| \le \|\left(\mathsf{H}_i^{\square}\mathsf{H}_i\right)^{-1}\|\|\mathsf{H}_i^{\square}\mathsf{W}_i\| &= \frac{\|\mathsf{H}_i^{\square}\mathsf{W}_i\|}{\lambda_{\min}\left(\mathsf{H}_i^{\square}\mathsf{H}_i\right)}, \\ &\overset{(a)}{=} \frac{\|\sum_{\tau=0}^{L-1}\mathsf{H}_i^{(\tau)\square}\mathsf{W}_i^{(\tau)}\|}{\lambda_{\min}\left(\sum_{\tau=0}^{L-1}\mathsf{H}_i^{(\tau)\square}\mathsf{H}_i^{(\tau)}\right)}, \\ &\overset{(b)}{=} \frac{\sum_{\tau=0}^{L-1}\|\mathsf{H}_i^{(\tau)\square}\mathsf{W}_i^{(\tau)}\|}{\sum_{\tau=0}^{L-1}\lambda_{\min}\left(\mathsf{H}_i^{(\tau)\square}\mathsf{H}_i^{(\tau)}\right)},\end{aligned} \tag{B.28}$$

where we obtain (a) from the fact that $\mathcal{S}_i = \bigcup_{\tau=0}^{L-1}\mathcal{S}_i^{(\tau)}$ and (b) follows from using triangular inequality and Weyl's inequality for Hermitian matrices. If we upper bound the terms $\|\mathsf{H}_i^{(\tau)\square}\mathsf{W}_i^{(\tau)}\|$ and lower bound the terms $\lambda_{\min}\left(\mathsf{H}_i^{(\tau)\square}\mathsf{H}_i^{(\tau)}\right)$, for all $0 \le \tau \le L-1$, we can use (B.28) to upper bound the estimation error $\|\hat{\Theta}_i - \Theta_i^{\square}\|$. However, because $\mathsf{H}_i^{(\tau)}$ has non-i.i.d. rows, it is not straightforward to bound the terms $\|\mathsf{H}_i^{(\tau)\square}\mathsf{W}_i^{(\tau)}\|$ and $\lambda_{\min}\left(\mathsf{H}_i^{(\tau)\square}\mathsf{H}_i^{(\tau)}\right)$ directly. To resolve this issue, we rely on the notion of stability and use perturbation based techniques to indirectly bound these terms in the following sub-sections.

### B.2.2 Estimation from Truncated States

**Definition B.5 (Truncated state vector [51])** *Consider the state equation (B.7). Given $t \ge L > 0$, the $L$-truncation of $\mathsf{x}_t$ is denoted by $\mathsf{x}_{t,L}$ and is obtained by driving the system with excitation $\mathsf{z}_\tau'$ and additive noise $\mathsf{w}_\tau'$ until time $t$, where*

$$\mathsf{v}_\tau' = \begin{cases} 0 & \text{if} \quad \tau < t-L \\ \mathsf{v}_\tau & \text{else} \end{cases}. \tag{B.29}$$

*In words, the $L$-truncated state vector $\mathsf{x}_{t,L}$ is obtained by unrolling $\mathsf{x}_t$ until time $t-L$ and setting $\mathsf{x}_{t-L}$ to 0.*

Using a truncation argument, we can obtain independent samples from a single trajectory which will be used to capture the effect of learning from a single trajectory. With high probability over the mode observation, truncated states can be made very close to the original states with sufficiently large truncation length. From (B.8), we have

$$\mathsf{x}_t - \mathsf{x}_{t,L} = \prod_{j=t-L}^{t-1}\mathsf{L}_{\omega(j)}\mathsf{x}_{t-L}. \tag{B.30}$$

As a corollary of Lemma A.4, observe that for a closed loop autonomous system $\mathsf{x}_{t+1} = \mathsf{L}_{\omega(t)}\mathsf{x}_t$, mean-square stability implies that, for any initial conditions $\mathsf{x}_0$ and $\omega_0$, we have $\mathsf{E}[\|\mathsf{x}_t\|^2] \le \sqrt{ns}\tau_{\tilde{\mathsf{L}}}\rho_{\tilde{\mathsf{L}}}^t\|\mathsf{x}_0\|^2$. Combining this argument with (B.30), we have

$$\begin{aligned}&\mathsf{E}[\|\mathsf{x}_t - \mathsf{x}_{t,L}\|^2] = \mathsf{E}[\|\prod_{j=t-L}^{t-1}\mathsf{L}_{\omega(j)}^{+}\mathsf{x}_{t-L}\|^2] \le \sqrt{ns}\tau_{\tilde{\mathsf{L}}}\rho_{\tilde{\mathsf{L}}}^L\|\mathsf{x}_{t-L}\|^2, \\ \implies &\mathsf{E}[\|\mathsf{x}_t - \mathsf{x}_{t,L}\|] \le \sqrt{(ns)^{1/2}\tau_{\tilde{\mathsf{L}}}\rho_{\tilde{\mathsf{L}}}^L}\|\mathsf{x}_{t-L}\| \le \sqrt{ns}\tau_{\tilde{\mathsf{L}}}\rho_{\tilde{\mathsf{L}}}^{L/2}\|\mathsf{x}_{t-L}\|,\end{aligned} \tag{B.31}$$

where the expectation is over the Markov modes $\{\omega(j)\}_{j=t-L}^{t-1}$ and we get the last relation by using Jensen's inequality. Moreover, if we also have $\|\mathsf{x}_{t-L}\| \le c\sigma_{\mathsf{w}}\beta_+\sqrt{n}$, then we can make $\mathsf{E}[\|\mathsf{x}_t - \mathsf{x}_{t,L}\|]$ arbitrarily small by picking a sufficiently large truncation length $L \ge 1$,

$$\mathsf{E}[\|\mathsf{x}_t - \mathsf{x}_{t,L}\| \mid \|\mathsf{x}_{t-L}\| \le c\sigma_{\mathsf{w}}\beta_+\sqrt{n}] \le \tau_{\tilde{\mathsf{L}}}\rho_{\tilde{\mathsf{L}}}^L\|\mathsf{x}_{t-L}\| \le c\sigma_{\mathsf{w}}\beta_+ n\sqrt{s}\tau_{\tilde{\mathsf{L}}}\rho_{\tilde{\mathsf{L}}}^{L/2}. \tag{B.32}$$

To proceed, we carry out the truncation of the sub-trajectories introduced in Def. B.3 to get the truncated sub-trajectories defined as follows.

**Definition B.6 (Truncated sub-trajectories)** *Let sampling period $L \geq 1$ be an integer. Let $\tau_k = \tau + kL$ be the sub-sampling indices, where $0 \leq \tau \leq L-1$ is a fixed o set and $k = 1, 2, \ldots, \lfloor \frac{T-L}{L} \rfloor$. Let $S_i^{(\tau)}$ be as in Def. B.3. For each $\tau_k \in S_i^{(\tau)}$, let $\tau_k^{'} \in S_i^{(\tau)}$ denotes the largest time index smaller than $\tau_k$. Given the $\tau_{th}$ sub-trajectory $\{(\mathsf{x}_{\tau_k+1}, \mathsf{x}_{\tau_k}, \mathsf{z}_{\tau_k}, \omega(\tau_k))\}_{\tau_k \in S_i^{(\tau)}}$ from Def. B.3, we truncate each state $\mathsf{x}_{\tau_k}$ by $\tau_k - \tau_k^{'} - 1$ to get the $\tau_{th}$ truncated sub-trajectory $\{(\bar{\mathsf{x}}_{\tau_k+1}, \bar{\mathsf{x}}_{\tau_k}, \mathsf{z}_{\tau_k}, \omega(\tau_k))\}_{\tau_k \in S_i^{(\tau)}}$, where*

$$\bar{\mathsf{x}}_{\tau_k} := \mathsf{x}_{\tau_k, \tau_k - \tau_k' - 1} \quad \textit{and} \quad \bar{\mathsf{x}}_{\tau_k+1} := \mathsf{L}_{\omega(\tau_k)} \bar{\mathsf{x}}_{\tau_k} + \mathsf{B}_{\omega(\tau_k)} \mathsf{z}_{\tau_k} + \mathsf{w}_{\tau_k}. \tag{B.33}$$

*If $\tau_k$ is the smallest time index in $S_i^{(\tau)}$, we set $\tau_k^{'} = 0$.*

Note that $\tau_k - \tau_k^{'} \geq L$ by definition. Hence, the truncation lengths used to obtain $\{\bar{\mathsf{x}}_{\tau_k}\}_{\tau_k \in S_i^{(\tau)}}$ are always larger than $L-1$. Next, we show that when $L \geq 1$ is sufficiently large enough, then the truncated states $\{\bar{\mathsf{x}}_{\tau_k}\}_{\tau_k \in S_i^{(\tau)}}$ as well as the Euclidean distance between the truncated and non-truncated states can be bounded with high probability over the modes.

**Lemma B.7 (Bounded states (truncated))** *Consider the same setup of Corollary B.2. Let $\{\mathsf{x}_{\tau_k}\}_{\tau_k \in S_i^{(\tau)}}$ be the bounded states and $\{\bar{\mathsf{x}}_{\tau_k}\}_{\tau_k \in S_i^{(\tau)}}$ be the truncated states from Def. B.3 and B.6 respectively. Let*

$$\beta_+^{'} := c_{\mathbf{w}} + \beta_+ \|\mathsf{L}_{1:s}\| + C_{\mathbf{z}} \sqrt{p/n} \|\mathsf{B}_{1:s}\|, \tag{B.34}$$

$$L_{tr1}(\rho_{\tilde{\mathbf{L}}}, \delta) := 1 + \frac{2 \log\left(2\sqrt{ns}\tau_{\tilde{\mathbf{L}}} T \beta_+^{'} / (\beta_+ \delta)\right)}{1 - \rho_{\tilde{\mathbf{L}}}}, \tag{B.35}$$

$$\textit{and} \quad L \geq \max\{t_0, L_{tr1}(\rho_{\tilde{\mathbf{L}}}, \delta)\}. \tag{B.36}$$

*Then, with probability at least $1-\delta$ over the modes, for all $\tau_k \in S_i^{(\tau)}$ and all $i \in [s]$, we have*

$$\|\mathsf{x}_{\tau_k} - \bar{\mathsf{x}}_{\tau_k}\| \leq (1/2) c \sigma_{\mathbf{w}} \beta_+ \sqrt{n} \quad \textit{and} \quad \|\bar{\mathsf{x}}_{\tau_k}\| \leq (3/2) c \sigma_{\mathbf{w}} \beta_+ \sqrt{n}. \tag{B.37}$$

**Proof** To begin, using Assumption 1 and (B.31), the impact of truncation can be bounded in expectation over the modes as follows,

$$\begin{aligned} \mathsf{E}[\|\mathsf{x}_{\tau_k} - \bar{\mathsf{x}}_{\tau_k}\|] = \mathsf{E}[\|\mathsf{x}_{\tau+kL} - \mathsf{x}_{\tau+kL, (k-k')L-1}\|] &\leq \sqrt{ns}\tau_{\tilde{\mathbf{L}}} \rho_{\tilde{\mathbf{L}}}^{((k-k')L-1)/2} \|\mathsf{x}_{\tau+k'L+1}\|, \\ &\leq \sqrt{ns}\tau_{\tilde{\mathbf{L}}} \rho_{\tilde{\mathbf{L}}}^{(L-1)/2} \|\mathsf{x}_{\tau_k'+1}\|, \end{aligned} \tag{B.38}$$

where we get the last inequality from the fact that $k - k^{'} \geq 1$, and the expectation is over the Markov modes at timesteps $\tau + k^{'}L + 1, \tau + k^{'}L + 2, \ldots \tau + kL - 1$. To proceed, observe that, for all $\tau_k \in S_i^{(\tau)}$, we have

$$\begin{aligned} \|\mathsf{x}_{\tau_k+1}\| = \|\mathsf{L}_{\omega(\tau_k)} \mathsf{x}_{\tau_k} + \mathsf{B}_{\omega(\tau_k)} \mathsf{z}_{\tau_k} + \mathsf{w}_{\tau_k}\| &\leq \max_{i \in [s]} \|\mathsf{L}_i\| \|\mathsf{x}_{\tau_k}\| + \max_{i \in [s]} \|\mathsf{B}_i\| \|\mathsf{z}_{\tau_k}\| + \|\mathsf{w}_{\tau_k}\|, \\ &\leq c\sigma_{\mathbf{w}} \beta_+ \sqrt{n} \|\mathsf{L}_{1:s}\| + c\sigma_{\mathbf{z}} \sqrt{p} \|\mathsf{B}_{1:s}\| + c_{\mathbf{w}} \sigma_{\mathbf{w}} \sqrt{n}, \\ &\leq c\sigma_{\mathbf{w}} (c_{\mathbf{w}} + \beta_+ \|\mathsf{L}_{1:s}\| + C_{\mathbf{z}} \sqrt{p/n} \|\mathsf{B}_{1:s}\|) \sqrt{n}, \\ &= c\sigma_{\mathbf{w}} \beta_+^{'} \sqrt{n}, \end{aligned} \tag{B.39}$$

where we set $\beta_+^{'} := c_{\mathbf{w}} + \beta_+ \|\mathsf{L}_{1:s}\| + C_{\mathbf{z}} \sqrt{p/n} \|\mathsf{B}_{1:s}\|$ and $C_{\mathbf{z}} := \sigma_{\mathbf{z}} / \sigma_{\mathbf{w}}$. Combining (B.39) with (B.38), for all $\tau_k \in S_i^{(\tau)}$ and all $i \in [s]$, we have

$$\mathsf{E}[\|\mathsf{x}_{\tau_k} - \bar{\mathsf{x}}_{\tau_k}\|] \leq c\sigma_{\mathbf{w}} \beta_+^{'} n \sqrt{s} \tau_{\tilde{\mathbf{L}}} \rho_{\tilde{\mathbf{L}}}^{(L-1)/2}, \tag{B.40}$$

$$\implies \mathsf{P}\left(\|\mathsf{x}_{\tau_k} - \bar{\mathsf{x}}_{\tau_k}\| \leq \frac{c\sigma_{\mathbf{w}} \beta_+^{'} n \sqrt{s} \tau_{\tilde{\mathbf{L}}} \rho_{\tilde{\mathbf{L}}}^{(L-1)/2} T}{\delta}\right) \geq 1 - \delta, \tag{B.41}$$

where we get the high probability bound by using Markov inequality and union bounding over all bounded states. This further implies that, with probability at least $1-\delta$ over the modes, for all $\tau_k \in S_i^{(\tau)}$ and all $i \in [s]$, we have

$$\|\bar{\mathsf{x}}_{\tau_k}\| \le \|\mathsf{x}_{\tau_k}\| + \|\mathsf{x}_{\tau_k} - \bar{\mathsf{x}}_{\tau_k}\| \le c\sigma_{\mathsf{w}}\beta_+\sqrt{n} + \frac{c\sigma_{\mathsf{w}}\beta_+' n\sqrt{s}\tau_{\tilde{L}}\rho_{\tilde{L}}^{(L-1)/2}T}{\delta} \le (3/2)c\sigma_{\mathsf{w}}\beta_+\sqrt{n}, \tag{B.42}$$

where we get the last inequality by choosing $L \ge 1$ via

$$\begin{aligned}\frac{c\sigma_{\mathsf{w}}\beta_+' n\sqrt{s}\tau_{\tilde{L}}\rho_{\tilde{L}}^{(L-1)/2}T}{\delta} \le c\sigma_{\mathsf{w}}\beta_+\sqrt{n}/2 &\iff \rho_{\tilde{L}}^{(L-1)/2} \le \frac{\delta\beta_+}{2\sqrt{ns}\tau_{\tilde{L}}T\beta_+'},\\ &\impliedby L \ge 1 + \frac{2\log\left(2\sqrt{ns}\tau_{\tilde{L}}T\beta_+'/(\beta_+\delta)\right)}{1-\rho_{\tilde{L}}}.\end{aligned} \tag{B.43}$$

This also implies that, with probability at least $1-\delta$ over the modes, for all $\tau_k \in S_i^{(\tau)}$ and all $i \in [s]$, we have $\|\mathsf{x}_{\tau_k} - \bar{\mathsf{x}}_{\tau_k}\| \le (1/2)c\sigma_{\mathsf{w}}\beta_+\sqrt{n}$. This completes the proof. ■

By construction, conditioned on the modes, $\bar{\mathsf{x}}_{\tau_k} = \mathsf{x}_{\tau_k,\tau_k-\tau_k'-1}$ only depends on the excitation and noise $\{\mathsf{z}_t, \mathsf{w}_t\}_{t=\tau+k'L+1}^{\tau+kL-1}$. Note that the dependence ranges $[\tau+k'L+1, \tau+kL-1]$ are disjoint intervals for each $(k,k')$ pairs. Hence, $\{\bar{\mathsf{x}}_{\tau_k}\}_{\tau_k \in S_i^{(\tau)}}$ should all be independent of each other. However, this is not the case because $\{\bar{\mathsf{x}}_{\tau_k}\}_{\tau_k \in S_i^{(\tau)}}$ are obtained by truncating only bounded states $\{\mathsf{x}_{\tau_k}\}_{\tau_k \in S_i^{(\tau)}}$. Therefore, we will look for a subset of independent truncated states within $\{\bar{\mathsf{x}}_{\tau_k}\}_{\tau_k \in S_i^{(\tau)}}$, as follows.

**Definition B.8 (Subset of bounded states)** *Let sampling period $L \ge 1$ be an integer. Let $\tau_k = \tau + kL$ be the sub-sampling indices, where $0 \le \tau \le L-1$ is a fixed o set and $k = 1, 2, \ldots, \lfloor \frac{T-L}{L} \rfloor$. We sub-sample the trajectory $(\mathsf{x}_t, \mathsf{z}_t, \omega(t))_{t=0}^T$ at time indices $\tau_k \in \bar{S}_i^{(\tau)}$, where*

$$\bar{S}_i^{(\tau)} := \{\tau_k \mid \omega(\tau_k) = i,\ \|\mathsf{x}_{\tau_k}\| \le c\sigma_{\mathsf{w}}\beta_+\sqrt{n},\ \|\bar{\mathsf{x}}_{\tau_k}\| \le (1/2)c\sigma_{\mathsf{w}}\beta_+\sqrt{n},\ \|\mathsf{z}_{\tau_k}\| \le c\sigma_{\mathsf{z}}\sqrt{p}\}, \tag{B.44}$$

*to obtain a subset of the $\tau_{th}$ sub-trajectory, denoted by $\{(\mathsf{x}_{\tau_k+1}, \mathsf{x}_{\tau_k}, \mathsf{z}_{\tau_k}, \omega(\tau_k))\}_{\tau_k \in \bar{S}_i^{(\tau)}}$.*

Next, we show that, conditioned on the modes, the samples in $\{(\bar{\mathsf{x}}_{\tau_k+1}, \bar{\mathsf{x}}_{\tau_k}, \mathsf{z}_{\tau_k}, \omega(\tau_k))\}_{\tau_k \in \bar{S}_i^{(\tau)}}$ are independent.

**Lemma B.9 (Conditional independence)** *Consider the MJS (B.7). Suppose $\{\mathsf{z}_t\}_{t=0}^{\infty} \overset{i.i.d.}{\sim} \mathcal{N}(0, \sigma_{\mathsf{z}}^2 I_p)$ and $\{\mathsf{w}_t\}_{t=0}^{\infty} \overset{i.i.d.}{\sim} \mathcal{D}_w$ satisfies Assumption 3. Suppose the sampling period $L \ge 1$ satisfies (B.36). Let $S_i^{(\tau)}$ and $\bar{S}_i^{(\tau)}$ be as in Def. B.3 and B.8 respectively. Then, with probability at least $1-\delta$ over the mode, we have,* (a) *$\{\bar{\mathsf{x}}_{\tau_k}\}_{\tau_k \in \bar{S}_i^{(\tau)}}$ is a subset of $\{\bar{\mathsf{x}}_{\tau_k}\}_{\tau_k \in S_i^{(\tau)}}$,* (b) *conditioned on the modes, $\{\bar{\mathsf{x}}_{\tau_k}\}_{\tau_k \in \bar{S}_i^{(\tau)}}$ are all independent,* (c) *conditioned on the modes, $\{\bar{\mathsf{x}}_{\tau_k}\}_{\tau_k \in \bar{S}_i^{(\tau)}}$, $\{\mathsf{z}_{\tau_k}\}_{\tau_k \in \bar{S}_i^{(\tau)}}$ and $\{\mathsf{w}_{\tau_k}\}_{\tau_k \in \bar{S}_i^{(\tau)}}$ are all independent of each other.*

**Proof** The first statement is a direct implication of Def. B.3 and B.8. To prove the second statement, consider $\{\mathsf{x}_{\tau_k}\}_{\tau_k \in S_i^{(\tau)}}$ which contains states bounded by $c\sigma_{\mathsf{w}}\beta_+\sqrt{n}$. From (B.8), observe that, each state can be decomposed into $\mathsf{x}_{\tau_k} = \bar{\mathsf{x}}_{\tau_k} + \tilde{\mathsf{x}}_{\tau_k}$, where $\bar{\mathsf{x}}_{\tau_k}$ is the truncated state and $\tilde{\mathsf{x}}_{\tau_k}$ captures the impact of the past states at time index $\tau + k'L + 1$. When the sampling period $L$ satisfies (B.36), then from Lemma B.7, with probability at least $1-\delta$ over the modes, for all $\tau_k \in S_i^{(\tau)}$ and all $i \in [s]$, we have $\|\tilde{\mathsf{x}}_{\tau_k}\| \le (1/2)c\sigma_{\mathsf{w}}\beta_+\sqrt{n}$. Furthermore, from Def. B.8, for all $\tau_k \in \bar{S}_i^{(\tau)}$ and all $i \in [s]$, we have $\|\bar{\mathsf{x}}_{\tau_k}\| \le (1/2)c\sigma_{\mathsf{w}}\beta_+\sqrt{n}$. Combining these results, with probability at least $1-\delta$ over the modes, for all $\tau_k \in \bar{S}_i^{(\tau)}$ and all $i \in [s]$, we have,

$$\|\mathsf{x}_{\tau_k}\| \le \|\bar{\mathsf{x}}_{\tau_k}\| + \|\tilde{\mathsf{x}}_{\tau_k}\| \le c\sigma_{\mathsf{w}}\beta_+\sqrt{n}. \tag{B.45}$$

Secondly, by construction, conditioned on the modes, $\bar{\mathsf{x}}_{\tau_k} = \mathsf{x}_{\tau_k,\tau_k-\tau_k'-1}$ only depends on the excitation and noise $\{\mathsf{z}_t, \mathsf{w}_t\}_{t=\tau+k'L+1}^{\tau+kL-1}$. Note that the dependence ranges $[\tau+k'L+1, \tau+kL-1]$ are disjoint intervals for

each $(k, k')$ pairs. Hence, conditioned on the modes, the samples in the set $\{\bar{\mathsf{x}}_{\tau_k}\}_{\tau_k \in \bar{S}_i^{(\tau)}}$ are all independent of each other.

To show the independence of $\{\bar{\mathsf{x}}_{\tau_k}\}_{\tau_k \in \bar{S}_i^{(\tau)}}$ and $\{\mathsf{z}_{\tau_k}\}_{\tau_k \in \bar{S}_i^{(\tau)}}$, observe that $\mathsf{z}_{\tau_k} = \mathsf{z}_{\tau+kL}$ have timestamps $\tau + kL$; which is not covered by $[\tau + k'L + 1, \tau + kL - 1]$ – the dependence ranges of $\{\bar{\mathsf{x}}_{\tau_k}\}_{\tau_k \in \bar{S}_i^{(\tau)}}$. Identical argument shows the independence of $\{\bar{\mathsf{x}}_{\tau_k}\}_{\tau_k \in \bar{S}_i^{(\tau)}}$ and $\{\mathsf{w}_{\tau_k}\}_{\tau_k \in \bar{S}_i^{(\tau)}}$. Lastly, $\{\mathsf{z}_{\tau_k}\}_{\tau_k \in \bar{S}_i^{(\tau)}}$ and $\{\mathsf{w}_{\tau_k}\}_{\tau_k \in \bar{S}_i^{(\tau)}}$ are independent of each other by definition. Hence, $\{\bar{\mathsf{x}}_{\tau_k}\}_{\tau_k \in \bar{S}_i^{(\tau)}}$, $\{\mathsf{z}_{\tau_k}\}_{\tau_k \in \bar{S}_i^{(\tau)}}$ and $\{\mathsf{w}_{\tau_k}\}_{\tau_k \in \bar{S}_i^{(\tau)}}$ are all independent of each other. This completes the proof. ■

Next, we state a lemma similar to Lemma B.4 to show that the truncated states have nice covariance properties.

**Lemma B.10 (Covariance of truncated states $\{\bar{\mathsf{x}}_{\tau_k}\}_{\tau_k \in \bar{S}_i^{(\tau)}}$)** *Consider the setup of Corollary B.2. Let $t_0$, $\beta_+$ and $\beta_+'$ be as in (B.9), (B.10) and (B.34) respectively. Let $c \geq 6$ be a fixed constant and $\{\bar{\mathsf{x}}_{\tau_k}\}_{\tau_k \in \bar{S}_i^{(\tau)}}$ be as in Lemma B.9. Define*

$$L_{cov}(\rho_{\tilde{L}}) := 1 + \frac{2\log\left(8c^2\beta_+\beta_+'n\sqrt{ns}\tau_{\tilde{L}}\right)}{1-\rho_{\tilde{L}}}, \tag{B.46}$$

*and suppose, the sampling period $L \geq 1$ obeys,*

$$L \geq \max\{t_0, L_{cov}(\rho_{\tilde{L}})\}. \tag{B.47}$$

*Then, for all $\tau_k \in \bar{S}_i^{(\tau)}$, we have*

$$(\sigma_{\mathsf{w}}^2/4)\, I_n \preceq \Sigma[\bar{\mathsf{x}}_{\tau_k}] \preceq 2c^2\sigma_{\mathsf{w}}^2\beta_+^2 n\, I_n. \tag{B.48}$$

$$(1/4)\, I_{n+p} \preceq \Sigma[\bar{\mathsf{h}}_{\tau_k}] \preceq 2c^2(1+\beta_+^2)(n+p)\, I_{n+p}. \tag{B.49}$$

**Proof** To begin, for all $\tau_k \in \bar{S}_i^{(\tau)}$, we upper bound the difference between the covariance of truncated and non-truncated states as follows,

$$\begin{aligned}
\|\mathsf{E}[\bar{\mathsf{x}}_{\tau_k}\bar{\mathsf{x}}_{\tau_k}^{\square} - \mathsf{x}_{\tau_k}\mathsf{x}_{\tau_k}^{\square}]\| &= \|\mathsf{E}[\bar{\mathsf{x}}_{\tau_k}\bar{\mathsf{x}}_{\tau_k}^{\square} - \mathsf{x}_{\tau_k}\bar{\mathsf{x}}_{\tau_k}^{\square} + \mathsf{x}_{\tau_k}\bar{\mathsf{x}}_{\tau_k}^{\square} - \mathsf{x}_{\tau_k}\mathsf{x}_{\tau_k}^{\square}]\|, \\
&\leq \mathsf{E}[\|\bar{\mathsf{x}}_{\tau_k}\|\|\mathsf{x}_{\tau_k} - \bar{\mathsf{x}}_{\tau_k}\|] + \mathsf{E}[\|\mathsf{x}_{\tau_k}\|\|\mathsf{x}_{\tau_k} - \bar{\mathsf{x}}_{\tau_k}\|], \\
&\leq (1/2)c\sigma_{\mathsf{w}}\beta_+\sqrt{n}\,\mathsf{E}[\|\mathsf{x}_{\tau_k} - \bar{\mathsf{x}}_{\tau_k}\|] + c\sigma_{\mathsf{w}}\beta_+\sqrt{n}\,\mathsf{E}[\|\mathsf{x}_{\tau_k} - \bar{\mathsf{x}}_{\tau_k}\|], \\
&\leq 2c^2\sigma_{\mathsf{w}}^2\beta_+\beta_+'n\sqrt{ns}\tau_{\tilde{L}}\rho_{\tilde{L}}^{(L-1)/2},
\end{aligned} \tag{B.50}$$

where we used Def. B.8 to obtain the second last inequality and (B.40) to obtain the last inequality. Combining this with Lemma B.4, for all $\tau_k \in \bar{S}_i^{(\tau)}$, assuming $c \geq 6$, we have

$$\begin{aligned}
\lambda_{\min}(\Sigma[\bar{\mathsf{x}}_{\tau_k}]) &\geq \lambda_{\min}(\Sigma[\mathsf{x}_{\tau_k}]) - \|\mathsf{E}[\bar{\mathsf{x}}_{\tau_k}\bar{\mathsf{x}}_{\tau_k}^{\square} - \mathsf{x}_{\tau_k}\mathsf{x}_{\tau_k}^{\square}]\|, \\
&\geq \sigma_{\mathsf{w}}^2/2 - 2c^2\sigma_{\mathsf{w}}^2\beta_+\beta_+'n\sqrt{ns}\tau_{\tilde{L}}\rho_{\tilde{L}}^{(L-1)/2} \geq \sigma_{\mathsf{w}}^2/4,
\end{aligned} \tag{B.51}$$

where we get the last inequality by choosing $L \geq 1$ via

$$\begin{aligned}
\sigma_{\mathsf{w}}^2/4 \geq 2c^2\sigma_{\mathsf{w}}^2\beta_+\beta_+'n\sqrt{ns}\tau_{\tilde{L}}\rho_{\tilde{L}}^{(L-1)/2} &\iff \rho_{\tilde{L}}^{(L-1)/2} \leq \frac{1}{8c^2\beta_+\beta_+'n\sqrt{ns}\tau_{\tilde{L}}}, \\
&\impliedby L \geq 1 + \frac{2\log\left(8c^2\beta_+\beta_+'n\sqrt{ns}\tau_{\tilde{L}}\right)}{1-\rho_{\tilde{L}}}.
\end{aligned} \tag{B.52}$$

This also implies that we have the following upper bound on the covariance spectral norm, that is, for all $\tau_k \in \bar{S}_i^{(\tau)}$, assuming $c \geq 6$, we have

$$\|\Sigma[\bar{\mathsf{x}}_{\tau_k}]\| \leq \|\Sigma[\mathsf{x}_{\tau_k}]\| + \|\mathsf{E}[\bar{\mathsf{x}}_{\tau_k}\bar{\mathsf{x}}_{\tau_k}^{\square} - \mathsf{x}_{\tau_k}\mathsf{x}_{\tau_k}^{\square}]\| \leq c^2\sigma_{\mathsf{w}}^2\beta_+^2 n + \sigma_{\mathsf{w}}^2/4 \leq 2c^2\sigma_{\mathsf{w}}^2\beta_+^2 n. \tag{B.53}$$

Combining (B.51) and (B.53), we get the first statement of the lemma, which is combined with (B.23) to obtain the second statement of the lemma. This completes the proof. ■

To proceed, consider the $\tau_{th}$ truncated sub-trajectory $\{(\bar{\mathbf{x}}_{\tau_k+1}, \bar{\mathbf{x}}_{\tau_k}, \mathbf{z}_{\tau_k}, \omega(\tau_k))\}_{\tau_k \in S_i^{(\tau)}}$ given by Def. B.6. Let $\bar{\mathbf{h}}_{\tau_k} := [\frac{1}{\sigma_{\mathbf{w}}}\bar{\mathbf{x}}_{\tau_k}^{\top} \ \frac{1}{\sigma_{\mathbf{z}}}\mathbf{z}_{\tau_k}^{\top}]^{\top}$. Similar to (B.26), we construct $\bar{\mathbf{Y}}_i^{(\tau)}$, $\bar{\mathbf{H}}_i^{(\tau)}$ and $\mathbf{W}_i^{(\tau)}$ by (row-wise) stacking $\{\bar{\mathbf{x}}_{\tau_k+1}^{\top}\}_{\tau_k \in S_i^{(\tau)}}$, $\{\bar{\mathbf{h}}_{\tau_k}^{\top}\}_{\tau_k \in S_i^{(\tau)}}$ and $\{\mathbf{w}_{\tau_k}^{\top}\}_{\tau_k \in S_i^{(\tau)}}$ respectively. As an intermediate step, we lower bound $\lambda_{\min}\big(\bar{\mathbf{H}}_i^{(\tau)\top}\bar{\mathbf{H}}_i^{(\tau)}\big)$ and upper bound $\|\bar{\mathbf{H}}_i^{(\tau)\top}\mathbf{W}_i^{(\tau)}\|$. This will, in turn, allow us to lower and upper bound the non-truncated terms $\lambda_{\min}\big(\mathbf{H}_i^{(\tau)\top}\mathbf{H}_i^{(\tau)}\big)$ and $\|\mathbf{H}_i^{(\tau)\top}\mathbf{W}_i^{(\tau)}\|$ respectively via,

$$\lambda_{\min}\big(\mathbf{H}_i^{(\tau)\top}\mathbf{H}_i^{(\tau)}\big) \geq \lambda_{\min}\big(\bar{\mathbf{H}}_i^{(\tau)\top}\bar{\mathbf{H}}_i^{(\tau)}\big) - \|\mathbf{H}_i^{(\tau)\top}\mathbf{H}_i^{(\tau)} - \bar{\mathbf{H}}_i^{(\tau)\top}\bar{\mathbf{H}}_i^{(\tau)}\|, \tag{B.54}$$

$$\|\mathbf{H}_i^{(\tau)\top}\mathbf{W}_i^{(\tau)}\| \leq \|\bar{\mathbf{H}}_i^{(\tau)\top}\mathbf{W}_i^{(\tau)}\| + \|\mathbf{H}_i^{(\tau)\top}\mathbf{W}_i^{(\tau)} - \bar{\mathbf{H}}_i^{(\tau)\top}\mathbf{W}_i^{(\tau)}\|. \tag{B.55}$$

For this purpose, our next lemma lower bounds the eigenvalues of the matrix $\bar{\mathbf{H}}_i^{(\tau)\top}\bar{\mathbf{H}}_i^{(\tau)}$ and upper bounds the error term $\|\bar{\mathbf{H}}_i^{(\tau)\top}\mathbf{W}_i^{(\tau)}\|$.

**Theorem B.11 (Bounding $\lambda_{\min}(\bar{\mathbf{H}}_i^{(\tau)\top}\bar{\mathbf{H}}_i^{(\tau)})$ and $\|\bar{\mathbf{H}}_i^{(\tau)\top}\mathbf{W}_i^{(\tau)}\|$)** *Consider the setup of Corollary B.2. Let $t_0$, $\beta_+$, $\beta_+'$, $L_{tr1}(\rho_{\tilde{\mathbf{L}}}, \delta)$ and $L_{cov}(\rho_{\tilde{\mathbf{L}}})$ be as in (B.9), (B.10), (B.34), (B.35) and (B.46), respectively. Let $C, C_0 > 0$ and $c \geq 6$ be fixed constants. Let $\bar{\mathbf{H}}_i^{(\tau)}$, $\tilde{\mathbf{H}}_i^{(\tau)}$ and $\mathbf{W}_i^{(\tau)}$ be constructed by (row-wise) stacking $\{\bar{\mathbf{h}}_{\tau_k}^{\top}\}_{\tau_k \in S_i^{(\tau)}}$, $\{\bar{\mathbf{h}}_{\tau_k}^{\top}\}_{\tau_k \in \bar{S}_i^{(\tau)}}$ and $\{\mathbf{w}_{\tau_k}^{\top}\}_{\tau_k \in S_i^{(\tau)}}$ respectively. Suppose the sampling period $L \geq 1$ and the number of independent samples $|\bar{S}_i^{(\tau)}|$ satisfy the following lower bounds,*

$$L \geq \max\{t_0, L_{tr1}(\rho_{\tilde{\mathbf{L}}}, \delta), L_{cov}(\rho_{\tilde{\mathbf{L}}})\}, \tag{B.56}$$

$$|\bar{S}_i^{(\tau)}| \geq 16c^2(1+\beta_+^2)\log\Big(\frac{2s(n+p)}{\delta}\Big)(n+p). \tag{B.57}$$

*Then, with probability at least $1-3\delta$, for all $i \in [s]$, we have*

$$\lambda_{\min}(\bar{\mathbf{H}}_i^{(\tau)\top}\bar{\mathbf{H}}_i^{(\tau)}) \geq \lambda_{\min}(\tilde{\mathbf{H}}_i^{(\tau)\top}\tilde{\mathbf{H}}_i^{(\tau)}) \geq \frac{|\bar{S}_i^{(\tau)}|}{16}, \tag{B.58}$$

$$\|\bar{\mathbf{H}}_i^{(\tau)\top}\mathbf{W}_i^{(\tau)}\| \leq 2c(1+\beta_+)\sqrt{|S_i^{(\tau)}|(n+p)}\Big(C\sigma_{\mathbf{w}}\sqrt{n+p} + C_0\sqrt{\log\Big(\frac{2s}{\delta}\Big)}\Big). \tag{B.59}$$

**Proof** To begin, recall that not all the rows of $\bar{\mathbf{H}}_i^{(\tau)}$ are independent. Therefore, to lower bound $\lambda_{\min}(\bar{\mathbf{H}}_i^{(\tau)\top}\bar{\mathbf{H}}_i^{(\tau)})$, we first consider the matrix $\tilde{\mathbf{H}}_i^{(\tau)}$ which is constructed by (row-wise) stacking $\{\bar{\mathbf{h}}_{\tau_k}^{\top}\}_{\tau_k \in \bar{S}_i^{(\tau)}}$. Observe that, conditioned on the modes, the matrix $\tilde{\mathbf{H}}_i^{(\tau)}$, which is a sub-matrix of $\bar{\mathbf{H}}_i^{(\tau)}$, has independent rows from Lemma B.9.

• **Lower bounding $\underline{\sigma}(\tilde{\mathbf{H}}_i^{(\tau)})$**: Using Lemma B.9, we observe that, conditioned on the modes, the rows of $\tilde{\mathbf{H}}_i^{(\tau)}$ are all independent. Secondly, by definition, each row of $\tilde{\mathbf{H}}_i^{(\tau)}$ can be deterministically bounded as follows: for all $\tau_k \in \bar{S}_i^{(\tau)}$, we have

$$\|\bar{\mathbf{h}}_{\tau_k}\|^2 \leq \frac{1}{\sigma_{\mathbf{w}}^2}\|\bar{\mathbf{x}}_{\tau_k}\|^2 + \frac{1}{\sigma_{\mathbf{z}}^2}\|\mathbf{z}_{\tau_k}\|^2 \leq (1/4)c^2\beta_+^2 n + c^2 p \leq c^2(1+\beta_+^2)(n+p). \tag{B.60}$$

Thirdly, from Lemma B.10, when $c \geq 6$ and $L \geq \max\{t_0, L_{cov}(\rho_{\tilde{\mathbf{L}}})\}$, then for all $\tau_k \in \bar{S}_i^{(\tau)}$, we have

$$(1/4)\mathbf{I}_{n+p} \preceq \boldsymbol{\Sigma}[\bar{\mathbf{h}}_{\tau_k}] \preceq 2c^2(1+\beta_+^2)(n+p)\mathbf{I}_{n+p}. \tag{B.61}$$

Therefore, we can use Corollary A.9 to lower bound $\underline{\sigma}(\tilde{\mathbf{H}}_i^{(\tau)})$. Specifically, using Corollary A.9 with $\sigma_{\min} = 1/2$ and $m = c^2(1+\beta_+^2)(n+p)$, with probability at least $1-\delta$, for all $i \in [s]$, we have

$$\underline{\sigma}(\tilde{\mathbf{H}}_i^{(\tau)}) \geq \frac{\sqrt{|\bar{S}_i^{(\tau)}|}}{2} - c\sqrt{(1+\beta_+^2)(n+p)\log\Big(\frac{2s(n+p)}{\delta}\Big)} \geq \frac{\sqrt{|\bar{S}_i^{(\tau)}|}}{4}, \tag{B.62}$$

as long as $|\bar{S}_i^{(\tau)}|$ satisfies the following lower bound,

$$\frac{\sqrt{|\bar{S}_i^{(\tau)}|}}{4} \geq c\sqrt{(1+\beta_+^2)(n+p)\log\big(\frac{2s(n+p)}{\delta}\big)}$$
$$\Longleftarrow |\bar{S}_i^{(\tau)}| \geq 16c^2(1+\beta_+^2)(n+p)\log\big(\frac{2s(n+p)}{\delta}\big). \tag{B.63}$$

• **Lower bounding** $\lambda_{\min}(\bar{\mathsf{H}}_i^{(\tau)\square}\bar{\mathsf{H}}_i^{(\tau)})$: Using Lemma B.9, we have, $\{\bar{\mathsf{h}}_{\tau_k}\}_{\tau_k\in\bar{S}_i^{(\tau)}}$ is a subset of $\{\bar{\mathsf{h}}_{\tau_k}\}_{\tau_k\in S_i^{(\tau)}}$. As a result, (B.62) also implies that, with probability at least $1-\delta$, for all $i\in[s]$, we have

$$\underline{\sigma}(\bar{\mathsf{H}}_i^{(\tau)}) \geq \underline{\sigma}(\tilde{\mathsf{H}}_i^{(\tau)}) \geq \frac{\sqrt{|\bar{S}_i^{(\tau)}|}}{4} \implies \lambda_{\min}(\bar{\mathsf{H}}_i^{(\tau)\square}\bar{\mathsf{H}}_i^{(\tau)}) \geq \frac{|\bar{S}_i^{(\tau)}|}{16}, \tag{B.64}$$

as long as $|\bar{S}_i^{(\tau)}|$ satisfies the lower bound in (B.63).

• **Upper bounding** $\|\bar{\mathsf{H}}_i^{(\tau)\square}\mathsf{W}_i^{(\tau)}\|$: Using Lemma B.7, when $L\geq\max\{t_0, L_{tr1}(\rho_{\tilde{L}},\delta)\}$, with probability at least $1-\delta$ over the modes, for all $\tau_k\in S_i^{(\tau)}$ and all $i\in[s]$, we have, $\|\bar{\mathsf{h}}_{\tau_k}\|^2\leq c^2(1+(9/4)\beta_+^2)(n+p)$. This implies that, with probability at least $1-\delta$ over the modes, for all $i\in[s]$, we have

$$\|\bar{\mathsf{H}}_i^{(\tau)}\| \leq \|\bar{\mathsf{H}}_i^{(\tau)}\|_F \leq c(1+2\beta_+)\sqrt{|S_i^{(\tau)}|(n+p)} \leq 2c(1+\beta_+)\sqrt{|S_i^{(\tau)}|(n+p)}.$$

To proceed, let $\bar{\mathsf{H}}_i^{(\tau)}$ have singular value decomposition $\mathsf{U}\Sigma\mathsf{V}^\square$ with $\|\Sigma\|\leq 2c(1+\beta_+)\sqrt{|S_i^{(\tau)}|(n+p)}$. Since $\mathsf{W}_i^{(\tau)}$ has i.i.d. $\sigma_{\mathbf{w}}$-subGaussian entries (Assumption 3), $\mathsf{U}^\square\mathsf{W}_i^{(\tau)}\in\mathsf{R}^{(n+p)\times n}$ has i.i.d. $\sigma_{\mathbf{w}}$-subGaussian columns. As a result, applying Theorem 5.39 of [65], with probability at least $1-\delta$, for all $i\in[s]$, we have

$$\|\mathsf{U}^\square\mathsf{W}_i^{(\tau)}\| \leq C\sigma_{\mathbf{w}}\sqrt{n+p} + C_0\sqrt{\log\big(\frac{2s}{\delta}\big)}. \tag{B.65}$$

This implies that, with probability at least $1-2\delta$, for all $i\in[s]$, we have

$$\|\bar{\mathsf{H}}_i^{(\tau)\square}\mathsf{W}_i^{(\tau)}\| \leq \|\Sigma\|\|\mathsf{U}^\square\mathsf{W}_i^{(\tau)}\| \leq 2c(1+\beta_+)\sqrt{|S_i^{(\tau)}|(n+p)}\Big(C\sigma_{\mathbf{w}}\sqrt{n+p} + C_0\sqrt{\log\big(\frac{2s}{\delta}\big)}\Big). \tag{B.66}$$

This completes the proof. ■

### B.2.3 Estimation from Non-truncated States

Coming back to the original problem of estimating the unknown dynamics from dependent samples, observe that the estimation error (B.28) in the case of single trajectory can be upper bounded as follows,

$$\begin{aligned}\|\hat{\Theta}_i - \Theta_i^\square\| &\leq \frac{\sum_{\tau=0}^{L-1}\|\mathsf{H}_i^{(\tau)\square}\mathsf{W}_i^{(\tau)}\|}{\sum_{\tau=0}^{L-1}\lambda_{\min}(\mathsf{H}_i^{(\tau)\square}\mathsf{H}_i^{(\tau)})},\\ &\leq \frac{\sum_{\tau=0}^{L-1}\big(\|\bar{\mathsf{H}}_i^{(\tau)\square}\mathsf{W}_i^{(\tau)}\| + \|\mathsf{H}_i^{(\tau)\square}\mathsf{W}_i^{(\tau)} - \bar{\mathsf{H}}_i^{(\tau)\square}\mathsf{W}_i^{(\tau)}\|\big)}{\sum_{\tau=0}^{L-1}\big(\lambda_{\min}(\bar{\mathsf{H}}_i^{(\tau)\square}\bar{\mathsf{H}}_i^{(\tau)}) - \|\mathsf{H}_i^{(\tau)\square}\mathsf{H}_i^{(\tau)} - \bar{\mathsf{H}}_i^{(\tau)\square}\bar{\mathsf{H}}_i^{(\tau)}\|\big)}.\end{aligned} \tag{B.67}$$

Therefore, to upper bound the estimation error $\|\hat{\Theta}_i - \Theta_i^\square\|$ with dependent samples, we need to upper bound the impact of truncation, captured by $\|\mathsf{H}_i^{(\tau)\square}\mathsf{W}_i^{(\tau)} - \bar{\mathsf{H}}_i^{(\tau)\square}\mathsf{W}_i^{(\tau)}\|$ and $\|\mathsf{H}_i^{(\tau)\square}\mathsf{H}_i^{(\tau)} - \bar{\mathsf{H}}_i^{(\tau)\square}\bar{\mathsf{H}}_i^{(\tau)}\|$ for all $i\in[s]$ and all $0\leq\tau\leq L-1$. This is done in the following theorem.

**Theorem B.12 (Small impact of truncation)** *Consider the same setup of Corollary B.2. Let $t_0$, $\beta_+$, $\beta_+^{'}$ and $L_{tr1}(\rho_{\tilde{\mathbf{L}}}, \delta)$ be as in (B.9), (B.10), (B.34) and (B.35), respectively. Suppose the sampling period $L$ obeys $L \geq \max\{t_0, L_{tr1}(\rho_{\tilde{\mathbf{L}}}, \delta)\}$. Let $\bar{\mathbf{H}}_i^{(\tau)}$ and $\mathbf{W}_i^{(\tau)}$ be constructed by (row-wise) stacking $\{\bar{\mathbf{h}}_{\tau_k}^{\square}\}_{\tau_k \square \mathcal{S}_i^{(\tau)}}$ and $\{\mathbf{w}_{\tau_k}^{\square}\}_{\tau_k \square \mathcal{S}_i^{(\tau)}}$ respectively. Then, with probability at least $1-\delta$ over the modes, for all $i \in [S]$, we have*

$$\|\mathbf{H}_i^{(\tau)\square}\mathbf{H}_i^{(\tau)} - \bar{\mathbf{H}}_i^{(\tau)\square}\bar{\mathbf{H}}_i^{(\tau)}\| \leq \frac{3c^2\beta_+^{'}(1+\beta_+)\tau_{\tilde{\mathbf{L}}}\rho_{\tilde{\mathbf{L}}}^{(L-1)/2}n\sqrt{s(n+p)}|\mathcal{S}_i^{(\tau)}|T}{\delta}, \tag{B.68}$$

$$\|\mathbf{H}_i^{(\tau)\square}\mathbf{W}_i^{(\tau)} - \bar{\mathbf{H}}_i^{(\tau)\square}\mathbf{W}_i^{(\tau)}\| \leq \frac{cc_{\mathbf{w}}\sigma_{\mathbf{w}}\beta_+^{'}\tau_{\tilde{\mathbf{L}}}\rho_{\tilde{\mathbf{L}}}^{(L-1)/2}n\sqrt{ns}|\mathcal{S}_i^{(\tau)}|T}{\delta}. \tag{B.69}$$

**Proof** We begin by simplifying the the term $\|\mathbf{H}_i^{(\tau)\square}\mathbf{H}_i^{(\tau)} - \bar{\mathbf{H}}_i^{(\tau)\square}\bar{\mathbf{H}}_i^{(\tau)}\|$ as follows,

$$\begin{aligned}
\|\mathbf{H}_i^{(\tau)\square}\mathbf{H}_i^{(\tau)} - \bar{\mathbf{H}}_i^{(\tau)\square}\bar{\mathbf{H}}_i^{(\tau)}\| &= \|\sum_{\tau_k \square \mathcal{S}_i^{(\tau)}} \left(\mathbf{h}_{\tau_k}\mathbf{h}_{\tau_k}^{\square} - \bar{\mathbf{h}}_{\tau_k}\bar{\mathbf{h}}_{\tau_k}^{\square}\right)\|, \\
&\leq |\mathcal{S}_i^{(\tau)}| \max_{\tau_k \square \mathcal{S}_i^{(\tau)}} \|\mathbf{h}_{\tau_k}\mathbf{h}_{\tau_k}^{\square} - \bar{\mathbf{h}}_{\tau_k}\bar{\mathbf{h}}_{\tau_k}^{\square}\|, \\
&= |\mathcal{S}_i^{(\tau)}| \max_{\tau_k \square \mathcal{S}_i^{(\tau)}} \|\mathbf{h}_{\tau_k}\mathbf{h}_{\tau_k}^{\square} - \mathbf{h}_{\tau_k}\bar{\mathbf{h}}_{\tau_k}^{\square} + \mathbf{h}_{\tau_k}\bar{\mathbf{h}}_{\tau_k}^{\square} - \bar{\mathbf{h}}_{\tau_k}\bar{\mathbf{h}}_{\tau_k}^{\square}\|, \\
&\leq |\mathcal{S}_i^{(\tau)}| \max_{\tau_k \square \mathcal{S}_i^{(\tau)}} \left(\|\mathbf{h}_{\tau_k}\|\|\mathbf{h}_{\tau_k} - \bar{\mathbf{h}}_{\tau_k}\| + \|\bar{\mathbf{h}}_{\tau_k}\|\|\mathbf{h}_{\tau_k} - \bar{\mathbf{h}}_{\tau_k}\|\right).
\end{aligned} \tag{B.70}$$

We will upper bound each of these terms separately and combine them together in (B.70) to get the desired upper bound. First of all, observe that, for all $i \in [S]$, each row of $\mathbf{H}_i^{(\tau)}$ is deterministically bounded as follows,

$$\|\mathbf{h}_{\tau_k}\|^2 \leq \frac{1}{\sigma_{\mathbf{w}}^2}\|\mathbf{x}_{\tau_k}\|^2 + \frac{1}{\sigma_{\mathbf{z}}^2}\|\mathbf{z}_{\tau_k}\|^2 \leq c^2\beta_+^2 n + c^2 p \leq c^2(1+\beta_+^2)(n+p). \tag{B.71}$$

Similarly, from Lemma B.7, when $L \geq \max\{t_0, L_{tr1}(\rho_{\tilde{\mathbf{L}}}, \delta)$, we observe that, with probability at least $1-\delta$ over the modes, for all $i \in [S]$, each row of $\bar{\mathbf{H}}_i^{(\tau)}$ can be bounded as follows,

$$\|\bar{\mathbf{h}}_{\tau_k}\|^2 \leq \frac{1}{\sigma_{\mathbf{w}}^2}\|\bar{\mathbf{x}}_{\tau_k}\|^2 + \frac{1}{\sigma_{\mathbf{z}}^2}\|\mathbf{z}_{\tau_k}\|^2 \leq c^2(9/4)\beta_+^2 n + c^2 p \leq 4c^2(1+\beta_+^2)(n+p). \tag{B.72}$$

To proceed, recall from (B.41) that, with probability at least $1-\delta$ over the modes, for all $\tau_k \in \mathcal{S}_i^{(\tau)}$ and all $i \in [S]$, we have

$$\|\mathbf{h}_{\tau_k} - \bar{\mathbf{h}}_{\tau_k}\| = \left\|\begin{bmatrix}\frac{1}{\sigma_{\mathbf{w}}}\mathbf{x}_{\tau_k} \\ \frac{1}{\sigma_{\mathbf{z}}}\mathbf{z}_{\tau_k}\end{bmatrix} - \begin{bmatrix}\frac{1}{\sigma_{\mathbf{w}}}\bar{\mathbf{x}}_{\tau_k} \\ \frac{1}{\sigma_{\mathbf{z}}}\mathbf{z}_{\tau_k}\end{bmatrix}\right\|_2 = \frac{1}{\sigma_{\mathbf{w}}}\|\mathbf{x}_{\tau_k} - \bar{\mathbf{x}}_{\tau_k}\| \leq \frac{c\beta_+^{'}n\sqrt{s}\tau_{\tilde{\mathbf{L}}}\rho_{\tilde{\mathbf{L}}}^{(L-1)/2}T}{\delta}. \tag{B.73}$$

Combining (B.71), (B.72) and (B.73) into (B.70), with probability at least $1-\delta$ over the modes, for all $i \in [S]$, we have

$$\|\mathbf{H}_i^{(\tau)\square}\mathbf{H}_i^{(\tau)} - \bar{\mathbf{H}}_i^{(\tau)\square}\bar{\mathbf{H}}_i^{(\tau)}\| \leq \frac{3c^2\beta_+^{'}(1+\beta_+)\tau_{\tilde{\mathbf{L}}}\rho_{\tilde{\mathbf{L}}}^{(L-1)/2}n\sqrt{s(n+p)}|\mathcal{S}_i^{(\tau)}|T}{\delta}. \tag{B.74}$$

Using a similar line of reasoning, with probability at least $1-\delta$ over the modes, for all $i \in [S]$, we also have

$$\begin{aligned}
\|\mathbf{H}_i^{(\tau)\square}\mathbf{W}_i^{(\tau)} - \bar{\mathbf{H}}_i^{(\tau)\square}\mathbf{W}_i^{(\tau)}\| &= \|\sum_{\tau_k \square \mathcal{S}_i^{(\tau)}} \left(\mathbf{h}_{\tau_k}\mathbf{w}_{\tau_k}^{\square} - \bar{\mathbf{h}}_{\tau_k}\mathbf{w}_{\tau_k}^{\square}\|\right), \\
&\leq |\mathcal{S}_i^{(\tau)}| \max_{\tau_k \square \mathcal{S}_i^{(\tau)}} \|\mathbf{h}_{\tau_k} - \bar{\mathbf{h}}_{\tau_k}\|\|\mathbf{w}_{(j_k)}^{\square}\|, \\
&\leq \frac{cc_{\mathbf{w}}\sigma_{\mathbf{w}}\beta_+^{'}\tau_{\tilde{\mathbf{L}}}\rho_{\tilde{\mathbf{L}}}^{(L-1)/2}n\sqrt{ns}|\mathcal{S}_i^{(\tau)}|T}{\delta}.
\end{aligned} \tag{B.75}$$

This completes the proof. ■

Combining Theorems B.11 and B.12, we obtain our result on the estimation of MJS in (B.7) from finite samples obtained from a single trajectory.

**Theorem B.13 (Learning with bounded noise)** *Consider the same setup of Corollary B.2. Let $t_0$, $\beta_+$, $\beta_+'$, $L_{tr1}(\rho_{\tilde{\mathbf{L}}}, \delta)$ and $L_{cov}(\rho_{\tilde{\mathbf{L}}})$ be as in (B.9), (B.10), (B.34), (B.35) and (B.46), respectively. Let $S_i^{(\tau)}$ and $\bar{S}_i^{(\tau)}$ be as in Def. B.3 and B.8 respectively and assume $|\bar{S}_i^{(\tau)}| \geq \frac{\pi_{\min} T}{2L}$, for all $i \in [s]$, with probability at least $1-\delta$. Suppose $\|\mathsf{K}_{1:s}\| \leq C_K$ for some constant $C_K > 0$. Let $C, C_0 > 0$, and $c \geq 6$ be fixed constants. Define*

$$L_{tr_2}(\rho_{\tilde{\mathbf{L}}}, \delta) := 1 + \frac{2}{(1-\rho_{\tilde{\mathbf{L}}})} \log\Big(\frac{192c^2\tau_{\tilde{\mathbf{L}}}\beta_+'(1+\beta_+)n\sqrt{s(n+p)}T}{\pi_{\min}\delta}\Big), \tag{B.76}$$

$$L_{tr_3}(\rho_{\tilde{\mathbf{L}}}, \delta) := 1 + \frac{2}{(1-\rho_{\tilde{\mathbf{L}}})} \log\Big(\frac{c_{\mathbf{w}}\sigma_{\mathbf{w}}\beta_+'\tau_{\tilde{\mathbf{L}}} n\sqrt{ns}T^2}{\delta(1+\beta_+)\sqrt{(n+p)}\big(C\sigma_{\mathbf{w}}\sqrt{n+p} + C_0\sqrt{\log(2s/\delta)}\big)}\Big). \tag{B.77}$$

*Suppose the sampling period $L$ and the trajectory length $T$ satisfy*

$$L \geq \max\{t_0, L_{cov}(\rho_{\tilde{\mathbf{L}}}), L_{tr1}(\rho_{\tilde{\mathbf{L}}}, \frac{\delta}{18L}), L_{tr2}(\rho_{\tilde{\mathbf{L}}}, \frac{\delta}{18L}), L_{tr3}(\rho_{\tilde{\mathbf{L}}}, \frac{\delta}{18L})\} \tag{B.78}$$

$$T \geq \frac{32L}{\pi_{\min}} c^2(1+\beta_+^2)\log\Big(\frac{36sL(n+p)}{\delta}\Big)(n+p). \tag{B.79}$$

*Then, solving the least-squares problem (B.27), with probability at least $1-\delta/2$, for all $i \in [s]$, we have*

$$\|\hat{\mathsf{A}}_i - \mathsf{A}_i\| \leq \frac{192c(1+\beta_+)}{\pi_{\min}} \frac{(\sigma_{\mathbf{z}} + C_K\sigma_{\mathbf{w}})}{\sigma_{\mathbf{z}}} \sqrt{\frac{L(n+p)}{T}} \Bigg(C\sqrt{n+p} + (C_0/\sigma_{\mathbf{w}})\sqrt{\log\Big(\frac{36sL}{\delta}\Big)}\Bigg),$$

$$\|\hat{\mathsf{B}}_i - \mathsf{B}_i\| \leq \frac{192c(1+\beta_+)}{\pi_{\min}} \frac{\sigma_{\mathbf{w}}}{\sigma_{\mathbf{z}}} \sqrt{\frac{L(n+p)}{T}} \Bigg(C\sqrt{n+p} + (C_0/\sigma_{\mathbf{w}})\sqrt{\log\Big(\frac{36sL}{\delta}\Big)}\Bigg).$$

**Proof** To begin, using Theorem B.11 along-with the assumption made in the statement of the theorem regarding $|\bar{S}_i^{(\tau)}|$, with probability at least $1-4\delta$, for all $i \in [s]$, we have

$$\lambda_{\min}(\bar{\mathsf{H}}_i^{(\tau)\square}\bar{\mathsf{H}}_i^{(\tau)}) \geq \frac{|\bar{S}_i^{(\tau)}|}{16} \geq \frac{\pi_{\min} T}{32L}, \tag{B.80}$$

as long as the trajectory length $T$ satisfies the following lower bound,

$$T \geq \frac{32L}{\pi_{\min}} c^2(1+\beta_+^2)\log\Big(\frac{2s(n+p)}{\delta}\Big)(n+p). \tag{B.81}$$

Combining this with Theorem B.12, with probability at least $1-5\delta$, for all $i \in [s]$, we have

$$\begin{aligned}
\lambda_{\min}(\mathsf{H}_i^{(\tau)\square}\mathsf{H}_i^{(\tau)}) &\geq \lambda_{\min}(\bar{\mathsf{H}}_i^{(\tau)\square}\bar{\mathsf{H}}_i^{(\tau)}) - \|\mathsf{H}_i^{(\tau)\square}\mathsf{H}_i^{(\tau)} - \bar{\mathsf{H}}_i^{(\tau)\square}\bar{\mathsf{H}}_i^{(\tau)}\|, \\
&\geq \frac{\pi_{\min} T}{32L} - \frac{3c^2\beta_+'(1+\beta_+)\tau_{\tilde{\mathbf{L}}}\rho_{\tilde{\mathbf{L}}}^{(L-1)/2} n\sqrt{s(n+p)}T^2}{\delta L}, \\
&\geq \frac{\pi_{\min} T}{64L},
\end{aligned} \tag{B.82}$$

where we get the last inequality by choosing $L \geq 1$ via,

$$\begin{aligned}
&\frac{\pi_{\min} T}{64L} \geq \frac{3c^2\beta_+'(1+\beta_+)\tau_{\tilde{\mathbf{L}}}\rho_{\tilde{\mathbf{L}}}^{(L-1)/2} n\sqrt{s(n+p)}T^2}{\delta L}, \\
\iff\ &\rho_{\tilde{\mathbf{L}}}^{(L-1)/2} \leq \frac{\pi_{\min}\delta}{192c^2\tau_{\tilde{\mathbf{L}}}\beta_+'(1+\beta_+)n\sqrt{s(n+p)}T}, \\
\impliedby\ &L \geq 1 + \frac{2\log\big(192c^2\tau_{\tilde{\mathbf{L}}}\beta_+'(1+\beta_+)n\sqrt{s(n+p)}T/(\pi_{\min}\delta)\big)}{(1-\rho_{\tilde{\mathbf{L}}})}.
\end{aligned} \tag{B.83}$$

Similarly, combing Theorems B.11 and B.12, with probability at least $1-4\delta$, for all $i \in [s]$, we also have

$$\begin{aligned}
\|\mathsf{H}_i^{(\tau)\square}\mathsf{W}_i^{(\tau)}\| &\le \|\bar{\mathsf{H}}_i^{(\tau)\square}\mathsf{W}_i^{(\tau)}\| + \|\mathsf{H}_i^{(\tau)\square}\mathsf{W}_i^{(\tau)} - \bar{\mathsf{H}}_i^{(\tau)\square}\mathsf{W}_i^{(\tau)}\|, \\
&\le 2c(1+\beta_+)\sqrt{\frac{T(n+p)}{L}}\left(C\sigma_{\mathbf{w}}\sqrt{n+p} + C_0\sqrt{\log\left(\frac{2s}{\delta}\right)}\right) + \frac{cc_{\mathbf{w}}\sigma_{\mathbf{w}}\beta_+'\tau_{\tilde{\mathbf{L}}}\rho_{\tilde{\mathbf{L}}}^{(L-1)/2}n\sqrt{ns}T^2}{\delta L}, \\
&\le 3c(1+\beta_+)\sqrt{\frac{T(n+p)}{L}}\left(C\sigma_{\mathbf{w}}\sqrt{n+p} + C_0\sqrt{\log\left(\frac{2s}{\delta}\right)}\right), \qquad \text{(B.84)}
\end{aligned}$$

where we get the last inequality by choosing $L \ge 1$ via,

$$\begin{aligned}
\frac{cc_{\mathbf{w}}\sigma_{\mathbf{w}}\beta_+'\tau_{\tilde{\mathbf{L}}}\rho_{\tilde{\mathbf{L}}}^{(L-1)/2}n\sqrt{ns}T^2}{\delta L} &\le c(1+\beta_+)\sqrt{\frac{T(n+p)}{L}}\left(C\sigma_{\mathbf{w}}\sqrt{n+p} + C_0\sqrt{\log\left(\frac{2s}{\delta}\right)}\right) \\
\iff \rho_{\tilde{\mathbf{L}}}^{(L-1)/2} &\le \frac{\delta(1+\beta_+)\sqrt{TL(n+p)}\left(C\sigma_{\mathbf{w}}\sqrt{n+p} + C_0\sqrt{\log(2s/\delta)}\right)}{c_{\mathbf{w}}\sigma_{\mathbf{w}}\beta_+'\tau_{\tilde{\mathbf{L}}}n\sqrt{ns}T^2}, \\
\impliedby \qquad L &\ge 1 + \frac{2}{(1-\rho_{\tilde{\mathbf{L}}})}\log\left(\frac{c_{\mathbf{w}}\sigma_{\mathbf{w}}\beta_+'\tau_{\tilde{\mathbf{L}}}n\sqrt{ns}T^2}{\delta(1+\beta_+)\sqrt{(n+p)}\left(C\sigma_{\mathbf{w}}\sqrt{n+p} + C_0\sqrt{\log(2s/\delta)}\right)}\right). \qquad \text{(B.85)}
\end{aligned}$$

Finally combining (B.82) and (B.84) into (B.28) and union bounding over all $0 \le \tau \le L-1$, with probability at least $1-9L\delta$, for all $i \in [s]$, we have

$$\begin{aligned}
\|\hat{\Theta}_i - \Theta_i^{\square}\| &\le \frac{\sum_{\tau=0}^{L-1}\|\mathsf{H}_i^{(\tau)\square}\mathsf{W}_i^{(\tau)}\|}{\sum_{\tau=0}^{L-1}\lambda_{\min}\left(\mathsf{H}_i^{(\tau)\square}\mathsf{H}_i^{(\tau)}\right)}, \\
&\le \frac{192c(1+\beta_+)}{\pi_{\min}}\sqrt{\frac{L(n+p)}{T}}\left(C\sigma_{\mathbf{w}}\sqrt{n+p} + C_0\sqrt{\log\left(\frac{2s}{\delta}\right)}\right). \qquad \text{(B.86)}
\end{aligned}$$

To proceed, using standard result from linear algebra that the spectral norm of a sub-matrix is upper bounded by the norm of the original matrix, with probability at least $1-9L\delta$, for all $i \in [s]$, we have

$$\begin{aligned}
\|\hat{\mathsf{L}}_i - \mathsf{L}_i\| &\le \frac{192c(1+\beta_+)}{\pi_{\min}}\sqrt{\frac{L(n+p)}{T}}\left(C\sqrt{n+p} + (C_0/\sigma_{\mathbf{w}})\sqrt{\log\left(\frac{2s}{\delta}\right)}\right), \\
\|\hat{\mathsf{B}}_i - \mathsf{B}_i\| &\le \frac{192c(1+\beta_+)}{\pi_{\min}}\frac{\sigma_{\mathbf{w}}}{\sigma_{\mathbf{z}}}\sqrt{\frac{L(n+p)}{T}}\left(C\sqrt{n+p} + (C_0/\sigma_{\mathbf{w}})\sqrt{\log\left(\frac{2s}{\delta}\right)}\right), \\
\implies \|\hat{\mathsf{A}}_i - \mathsf{A}_i\| &\le \|\hat{\mathsf{L}}_i - \mathsf{L}_i\| + \|\mathsf{K}_i\|\|\hat{\mathsf{B}}_i - \mathsf{B}_i\|, \\
&\le \frac{192c(1+\beta_+)}{\pi_{\min}}\frac{(\sigma_{\mathbf{z}} + C_K\sigma_{\mathbf{w}})}{\sigma_{\mathbf{z}}}\sqrt{\frac{L(n+p)}{T}}\left(C\sqrt{n+p} + (C_0/\sigma_{\mathbf{w}})\sqrt{\log\left(\frac{2s}{\delta}\right)}\right). \qquad \text{(B.87)}
\end{aligned}$$

Finally replacing $\delta$ with $\delta/(18L)$ we get the statement of the theorem. This completes the proof. ■

Next, we use the following lemma to relax the Assumption 3 on the noise.

**Lemma B.14 (From Bounded to Unbounded Noise)** *Let* $\mathsf{g} \overset{i.i.d.}{\sim} \mathcal{N}(0, \sigma_{\mathbf{w}}^2 I_{nT})$ *and* $\mathsf{a}$ *be two independent vectors. Let* $\mathsf{g}'$ *be the truncated Gaussian distribution* $\mathsf{g}' \sim \{\mathsf{g} \mid \|\mathsf{g}\|_\infty \le c_{\mathbf{w}}\sigma_{\mathbf{w}}\}$. *Let* $S_{\mathbf{g},\mathbf{a}}$ *be the indicator function of an event defined on vectors* $\mathsf{g}, \mathsf{a}$ *s.t.*

$$\mathsf{E}[S_{\mathbf{g}',\mathbf{a}}] \ge 1 - \delta/2.$$

*That is, the event holds, on the bounded variable* $\mathsf{g}'$*, with probability at least* $1-\delta/2$*. Then, if the bound above holds for* $c_{\mathbf{w}} > C_\delta := \sqrt{2\log(nT)} + \sqrt{2\log(2/\delta)}$*, we also have that*

$$\mathsf{E}[S_{\mathbf{g},\mathbf{a}}] \ge 1 - \delta.$$

*That is, the probability that event holds on the unbounded variable* $\mathsf{g}$ *is at least* $1-\delta$*.*

**Proof** Let E be the event $\{\mathbf{g} \mid \|\mathbf{g}\|_\infty \le c_{\mathbf{w}}\sigma_{\mathbf{w}}\}$. If $c_{\mathbf{w}} \ge \sqrt{2\log(nT)} + \sqrt{2\log(2/\delta)}$, using Gaussian tail bound and the fact that $\mathsf{E}[\|\mathbf{g}\|_\infty] \le \sigma_{\mathbf{w}}\sqrt{2\log(nT)}$, observe that $\mathsf{P}(E) \ge 1 - e^{-\frac{(c_{\mathbf{w}}\sigma_{\mathbf{w}} - \mathbb{E}[\|\mathbf{g}\|_\infty])^2}{2}} \ge 1 - \delta/2$. Therefore, we have

$$\mathsf{E}[S_{\mathbf{g},\mathbf{a}}] \ge \mathsf{E}[S_{\mathbf{g},\mathbf{a}}|E]\,\mathsf{P}(E) = \mathsf{E}[S_{\mathbf{g}',\mathbf{a}}]\,\mathsf{P}(E) \ge (1-\delta/2)^2 \ge 1-\delta.$$

This completes the proof. ∎

Combining Theorem B.13 and Lemma B.14, we get the following result on learning the MJS dynamics when the process noise is Gaussian.

**Corollary B.15 (Learning with un-bounded noise)** *Consider the same setup of Theorem B.13 except that Assumption 3 is replaced with* $\{\mathbf{w}_t\}_{t=0}^{\infty} \overset{i.i.d.}{\sim} \mathcal{N}(0, \sigma_{\mathbf{w}}^2 I_n)$ *and the threshold for bounding the noise satisfies,*

$$c_{\mathbf{w}} \ge \sqrt{2\log(nT)} + \sqrt{2\log(2/\delta)}. \tag{B.88}$$

*Suppose* $\|\mathsf{B}_{1:s}\| \le C_B$ *for some* $C_B > 0$ *and the trajectory length* $T$ *satisfies,*

$$T \gtrsim \frac{\tau_{\tilde{L}}\sqrt{s}L}{\pi_{\min}(1-\rho_{\tilde{L}})}\Big((\sqrt{2\log(nT)} + \sqrt{2\log(2/\delta)})^2 + C_{\mathbf{z}}^2 C_B^2\Big)\log\Big(\frac{36sL(n+p)}{\delta}\Big)(n+p). \tag{B.89}$$

*Then, solving the least-squares problem* (B.27)*, with probability at least* $1-\delta$*, for all* $i \in [s]$*, we have*

$$\begin{aligned}\|\hat{\mathsf{A}}_i - \mathsf{A}_i\| &\lesssim \frac{(\sigma_{\mathbf{z}} + C_K\sigma_{\mathbf{w}})}{\sigma_{\mathbf{z}}}\frac{\tau_{\tilde{L}}\sqrt{s}\big(\sqrt{2\log(nT)} + \sqrt{2\log(2/\delta)} + C_{\mathbf{z}}C_B\|\big)}{\pi_{\min}(1-\rho_{\tilde{L}})}\sqrt{\frac{L(n+p)}{T}} \\ &\quad\times\left(C\sqrt{n+p} + (C_0/\sigma_{\mathbf{w}})\sqrt{\log\Big(\frac{36sL}{\delta}\Big)}\right), \qquad\text{(B.90)}\\ \|\hat{\mathsf{B}}_i - \mathsf{B}_i\| &\lesssim \frac{\sigma_{\mathbf{w}}}{\sigma_{\mathbf{z}}}\frac{\tau_{\tilde{L}}\sqrt{s}\big(\sqrt{2\log(nT)} + \sqrt{2\log(2/\delta)} + C_{\mathbf{z}}C_B\|\big)}{\pi_{\min}(1-\rho_{\tilde{L}})}\sqrt{\frac{L(n+p)}{T}} \\ &\quad\times\left(C\sqrt{n+p} + (C_0/\sigma_{\mathbf{w}})\sqrt{\log\Big(\frac{36sL}{\delta}\Big)}\right). \qquad\text{(B.91)}\end{aligned}$$

At this point, we are only left with verifying the assumption that, for all $i \in [s]$, with probability at least $1-\delta$, we have $|\bar{S}_i^{(\tau)}| \ge \frac{\pi_{\min}T}{2L}$ for some choice of $L$ and $T$. In the following, we will state a lemma to show that the above assumption indeed holds for certain choice of $L$ and $T$. The detailed analysis for obtaining a lower bound on $|\bar{S}_i^{(\tau)}|$ is given in Sec. B.4. Specifically, the following result can be obtained by applying union bound to Lemma B.21 over $\tau = 0, 1, \ldots, L-1$.

**Lemma B.16** *Let* $\bar{S}_i^{(\tau)}$ *be as in Def. B.8 and consider the setup of Alg. 1. Assume* $c_{\mathbf{x}} \ge \underline{c}_{\mathbf{x}}(\rho_{\tilde{L}}, \tau_{\tilde{L}})$, $c_{\mathbf{z}} \ge \underline{c}_{\mathbf{z}}$, $L \ge \underline{C}_{sub,N}\left(\beta_0\sqrt{n}, \frac{\delta}{L}, T, \rho_{\tilde{L}}, \tau_{\tilde{L}}\right)\log(T)$*, and* $T \ge \underline{T}_N\left(\frac{L}{\log(T)}, \frac{\delta}{L}, \rho_{\tilde{L}}, \tau_{\tilde{L}}\right)$*, where* $\underline{c}_{\mathbf{x}}(\rho,\tau)$*,* $\underline{c}_{\mathbf{z}}$*, and* $\underline{T}_N(\mathcal{C},\delta,\rho,\tau)$ *are defined in Table 2, and* $C_{sub,N}(\bar{x}_0,\delta,T,\rho,\tau)$ *is defined in Table 4. Then with probability at least* $1-\delta$*, for all* $i \in [s]$ *and all* $\tau = 0, 1, \ldots, L-1$ *we have*

$$|\bar{S}_i^{(\tau)}| \ge \frac{\pi_{\min}T}{2L}. \tag{B.92}$$

### B.2.4 Finalizing the SYSID: Proof of Theorem 4.1

To finalize, we combine Corollary B.15 and Lemma B.16 to get our main result on learning the unknown MJS dynamics. The following theorem is a more refined and precise version of our main system identification result in Theorem 4.1.

**Theorem B.17 (Main result)** *Consider the MJS (B.7), with initial state $\mathsf{x}_0 \sim \mathcal{D}_x$ such that $\mathsf{E}[\mathsf{x}_0] = 0$, $\mathsf{E}[\|\mathsf{x}_0\|^2] \le \beta_0^2 n$ for some $\beta_0 > 0$. Suppose Assumption 1 on the system and the Markov chain holds. Suppose $\{\mathsf{z}_t\}_{t=0}^{\infty} \overset{i.i.d.}{\sim} \mathcal{N}(0, \sigma_{\mathsf{z}}^2 I_p)$, $\{\mathsf{w}_t\}_{t=0}^{\infty} \overset{i.i.d.}{\sim} \mathcal{N}(0, \sigma_{\mathsf{w}}^2 I_n)$ and the threshold for bounding the noise satisfies,*

$$c_{\mathsf{w}} \ge \sqrt{2\log(nT)} + \sqrt{2\log(2/\delta)}. \tag{B.93}$$

*Suppose $\|\mathsf{B}_{1:s}\| \le C_B$ and $\|\mathsf{K}_{1:s}\| \le C_K$ for some $C_B, C_K > 0$. Let $t_0$, $\beta_+$, $\beta_+'$, $L_{tr1}(\rho_{\tilde{\mathsf{L}}}, \delta)$, $L_{cov}(\rho_{\tilde{\mathsf{L}}})$, $L_{tr2}(\rho_{\tilde{\mathsf{L}}}, \delta)$ and $L_{tr3}(\rho_{\tilde{\mathsf{L}}}, \delta)$ be as in (B.9), (B.10), (B.34), (B.35), (B.46), (B.76) and (B.77), respectively. Suppose $c_{\mathsf{x}} \ge \underline{c}_{\mathsf{x}}(\rho_{\tilde{\mathsf{L}}}, \tau_{\tilde{\mathsf{L}}})$, $c_{\mathsf{z}} \ge \underline{c}_{\mathsf{z}}$, where $\underline{c}_{\mathsf{x}}(\rho, \tau)$ and $\underline{c}_{\mathsf{z}}$ are defined in Table 2. Let $C, C_0 > 0$, and $c \ge 6$ be fixed constants. Suppose the trajectory length $T$ satisfies*

$$T \gtrsim \max\left\{ \begin{array}{c} \frac{\tau_{\tilde{\mathsf{L}}}\sqrt{s}L}{\pi_{\min}(1-\rho_{\tilde{\mathsf{L}}})}\left(\left(\sqrt{2\log(nT)} + \sqrt{2\log(2/\delta)}\right)^2 + C_{\mathsf{z}}^2 C_B^2\right)\log\left(\frac{36sL(n+p)}{\delta}\right)(n+p), \\ \underline{T}_N\left(\frac{L}{\log(T)}, \frac{\delta}{L}, \rho_{\tilde{\mathsf{L}}}, \tau_{\tilde{\mathsf{L}}}\right) \end{array} \right\} \tag{B.94}$$

$$\text{where } L \ge \max\left\{ \begin{array}{c} t_0, L_{cov}(\rho_{\tilde{\mathsf{L}}}), L_{tr1}(\rho_{\tilde{\mathsf{L}}}, \frac{\delta}{18L}), L_{tr2}(\rho_{\tilde{\mathsf{L}}}, \frac{\delta}{18L}), L_{tr3}(\rho_{\tilde{\mathsf{L}}}, \frac{\delta}{18L}), \\ \underline{C}_{sub,N}\left(\beta_0\sqrt{n}, \frac{\delta}{L}, T, \rho_{\tilde{\mathsf{L}}}, \tau_{\tilde{\mathsf{L}}}\right)\log(T) \end{array} \right\}, \tag{B.95}$$

*where $\underline{T}_N(\mathcal{C}, \delta, \rho, \tau)$ and $C_{sub,N}(\bar{x}_0, \delta, T, \rho, \tau)$ are defined in Table 2 and 4 respectively. Then, solving the least-squares problem (B.27), with probability at least $1-\delta$, for all $i \in [s]$, we have*

$$\|\hat{\mathsf{A}}_i - \mathsf{A}_i\| \lesssim \frac{(\sigma_{\mathsf{z}} + C_K\sigma_{\mathsf{w}})}{\sigma_{\mathsf{z}}} \frac{\tau_{\tilde{\mathsf{L}}}\sqrt{s}\left(\sqrt{2\log(nT)} + \sqrt{2\log(2/\delta)} + C_{\mathsf{z}}C_B\|\right)}{\pi_{\min}(1-\rho_{\tilde{\mathsf{L}}})}\sqrt{\frac{L(n+p)}{T}} \tag{B.96}$$
$$\times\left(C\sqrt{n+p} + (C_0/\sigma_{\mathsf{w}})\sqrt{\log\left(\frac{36sL}{\delta}\right)}\right),$$

$$\|\hat{\mathsf{B}}_i - \mathsf{B}_i\| \lesssim \frac{\sigma_{\mathsf{w}}}{\sigma_{\mathsf{z}}} \frac{\tau_{\tilde{\mathsf{L}}}\sqrt{s}\left(\sqrt{2\log(nT)} + \sqrt{2\log(2/\delta)} + C_{\mathsf{z}}C_B\|\right)}{\pi_{\min}(1-\rho_{\tilde{\mathsf{L}}})}\sqrt{\frac{L(n+p)}{T}}$$
$$\times\left(C\sqrt{n+p} + (C_0/\sigma_{\mathsf{w}})\sqrt{\log\left(\frac{36sL}{\delta}\right)}\right). \tag{B.97}$$

**Remark B.3** *Note that in Theorem B.17, with the shorthand notations defined in Tables 2 and 4, the premise conditions (B.93), (B.94), and (B.95) can also be interpreted as the following.*

$$c_{\mathsf{w}} = \underline{c}_{\mathsf{w}}(T, \delta) \tag{B.98}$$
$$T \ge \underline{T}_{id,N}(L, \delta, T, \rho_{\tilde{\mathsf{L}}}, \tau_{\tilde{\mathsf{L}}}) \tag{B.99}$$
$$L \ge \underline{L}_{id}\left(\beta_0\sqrt{n}, \delta, T, \rho_{\tilde{\mathsf{L}}}, \tau_{\tilde{\mathsf{L}}}, \underline{c}_{\mathsf{w}}(T, \delta), \mathsf{K}_{1:s}, L\right). \tag{B.100}$$

*From the definition of $\underline{L}_{id}$, one can see there exists $\underline{L} = \tilde{\mathcal{O}}(\log(T))$ such that (B.100) holds by choosing $L = \underline{L}$. Define shorthand notation $\underline{T}_{id,N,L}(\delta, T, \rho, \tau) := \underline{T}_{id,N}(\underline{L}, \delta, T, \rho_{\tilde{\mathsf{L}}}, \tau_{\tilde{\mathsf{L}}})$, then the premise conditions (B.93), (B.94), and (B.95) can be implied by by the single condition $T \ge \underline{T}_{id,N,L}(\delta, T, \rho, \tau)$, under which the main results in Theorem B.17 still hold.*

## Discussion

• **Sample complexity**: Here, a few remarks are in place. First, the result appears to be convoluted however most of the dependencies are logarithmic (specifically dependency on the failure probability $\delta$ and $\log(T)$ terms). Besides these, the dominant term (when estimating $\mathsf{A}$) reduces to

$$\frac{(\sigma_{\mathsf{z}} + C_K\sigma_{\mathsf{w}})}{\sigma_{\mathsf{z}}} \frac{\tau_{\tilde{\mathsf{L}}}\sqrt{s}}{\pi_{\min}(1-\rho_{\tilde{\mathsf{L}}})} \frac{(n+p)}{\sqrt{T}}.$$

which is identical to our statement in Theorem 4.1. Note that the overall sample complexity grows as $T \gtrsim s(n+p)^2/\pi_{\min}^2$. We remark that, this quadratic growth is somewhat undesirable. A degrees-of-freedom

counting argument would lead to an ideal dependency of $T \gtrsim s(n+p)/\pi_{\min}$. The reason is that, each vector state equation we fit has $n$ scalar equations. The total degrees of freedom for each dynamics pair $(A_i, B_i)$ is $n \times (n+p)$. Additionally, for the least-frequent mode, in steady-state, we should observe $\pi_{\min} T$ equations. Putting these together, we would minimally need $n \times \pi_{\min} T \geq n \times (n+p)$, which means we need $T \geq s(n+p)/\pi_{\min}$ samples to estimate $s$ dynamic pairs $(A_{1:s}, B_{1:s})$. Our analysis indicates that this sub-optimality (at least the quadratic growth in $n$) can be addressed to achieve optimal dependence by establishing a stronger control on the state covariance (e.g. refining (B.18)) as well as a better control on the degree of independence across sampled states (this issue arises during the proof of Theorem B.11).

- **To what extent subsampling is necessary?** We recall that our argument is based on mixing-time arguments which are well-studied in the literature. In Alg. 1, we sub-sample the trajectory for bounded samples and use them to estimate the unknown MJS dynamics. Unfortunately, such a sub-sampling seems unavoidable as long as we don't have a good tail control on the distribution of the state vectors. Specifically, as long as the feature vectors (in our case state vectors) are allowed to be heavy-tailed, existing – to the best of our knowledge – minimum singular value concentration guarantees for the empirical covariance apply under the assumption of boundedness [65]. More recently, self-normalized martingale arguments are employed to address temporal dependencies [55, 59]. We remark that, using martingale-based arguments, it is possible to mitigate the spectral radius dependency by shaving a factor of $1/(1-\rho_{\tilde{L}})$ (e.g. martingale based arguments have milder $\rho_{\tilde{L}}$ dependence [55, 59]).

## B.4 Lower Bounding $|\bar{S}_i^{(\tau)}|$

To begin, we define sub-sampling period $L = C_{sub} \log(T)$, sub-sampling indices $\tau_k = \tau + kL$ for $k = 1, 2, \ldots, \lfloor T/L \rfloor$, and the time index set

$$S_i^{(\tau)} = \left\{ \tau_k \;\middle|\; \omega(\tau_k) = i, \quad \|x_{\tau_k}\| \leq c_x \sqrt{\|\Sigma_w\| \log(T)}, \quad \|z_{\tau_k}\| \leq c_z \sqrt{\|\Sigma_z\|} \right\}$$

by bounding $\|x_t\|$ and $\|z_t\|$, which is used to estimate $A_{1:s}$ and $B_{1:s}$ through least squares (Here we generalize isotropic noise $w_t \sim \mathcal{N}(0, \sigma_w^2 I_n)$ and $z_t \sim \mathcal{N}(0, \sigma_z^2 I_p)$ to $\mathcal{N}(0, \Sigma_w)$ and $\mathcal{N}(0, \Sigma_z)$, respectively.). A fundamental question is: Is $|S_i^{(\tau)}|$ big enough such that there will be enough data available when applying least squares? We provide answer to this question in this section. Lemma B.18 acts as a building block for the later result; Lemma B.19 provides the lower bound on $|S_i^{(\tau)}|$; Corollary B.20 gives a more interpretable lower bound on $|S_i^{(\tau)}|$ when $c_x$ and $c_z$ are large enough; and finally, Lemma B.21 shows how many samples in $S_i^{(\tau)}$ are "weakly" independent, which is the quantity that essentially determines the sample complexity of estimating $A_{1:s}$ and $B_{1:s}$.

For clarity, we reiterate some definitions and define a few new ones here. We are given an MJS$(A_{1:s}, B_{1:s}, T)$ with process noise $w_t \sim \mathcal{N}(0, \Sigma_w)$ and ergodic Markov matrix $T$. With some stabilizing controller $K_{1:s}$, the input is given by $u_t = K_{\omega(t)} x_t + z_t$ where $z_t \sim \mathcal{N}(0, \Sigma_z)$. Let $\bar{\sigma}^2 := \|B_{1:s}\|^2 \|\Sigma_z\| + \|\Sigma_w\|$. Let $L_i := A_i + B_i K_i$. Let $\tilde{L} \in \mathbb{R}^{sn^2 \times sn^2}$ denote the augmented closed-loop state matrix with $ij$-th $n^2 \times n^2$ block given by $[\tilde{L}]_{ij} := [T]_{ji} L_j \otimes L_j$. Let $\rho_{\tilde{L}} \in [0, 1)$ and $\tau_{\tilde{L}} > 0$ be two constants such that $\|\tilde{L}^k\| \leq \tau_{\tilde{L}} \rho_{\tilde{L}}^k$. By definition, one available choice for $\tau_{\tilde{L}}$ and $\rho_{\tilde{L}}$ are $\tau_{\tilde{L}}$ and $\rho(\tilde{L})$, respectively. Let $t_{\mathrm{MC}}(\cdot)$ and $t_{\mathrm{MC}}$ denote the mixing time of $T$ as in Definition A.2. Let $_\infty$ denote the stationary distribution of $T$, $\pi_{\min} = \min_i \; _\infty(i)$, and $\pi_{\max} = \max_i \; _\infty(i)$. Assume the initial state $x_0$ satisfies $\mathbb{E}[\|x_0\|^2] \leq \bar{x}_0^2$ for some $\bar{x}_0 \geq 0$. Lastly, without loss of generality, we consider the sub-trajectory with zero shift, that is, $\tau = 0$, which is identical to any $0 \leq \tau \leq L - 1$.

**Lemma B.18** *Suppose the Markov chain trajectory $\{\omega(0), \omega(1), \ldots\}$ and a sequence of events $\{A_0, A_1, \ldots\}$ are both adapted to filtration $\{\mathcal{F}_0, \mathcal{F}_1, \ldots\}$, i.e. $\omega(t)$ and $1_{\{A_t\}}$ are both $\mathcal{F}_t$-measurable. We assume $\mathbb{E}[1_{\{\omega(t)=j\}} \mid \mathcal{F}_{t-r}] = \mathbb{P}(\omega(t) = j \mid \omega(t-r))$ for all $j \in [s]$, $t$, and $r < t$. For all $i \in [s]$, let*

$$N_i = \sum_{k=1}^{\lfloor T/L \rfloor} 1_{\{\omega(kL)=i\}} 1_{\{A_{kL}\}} \tag{B.101}$$

*and suppose*

$$\mathsf{E}[\mathbf{1}_{\{A_t\}} \mid \mathcal{F}_{t-L}] \geq 1 - p_t \tag{B.102}$$

*for some* $p_t \in [0,1)$ *and* $L = C_{sub}\log(T)$. *Assume* $C_{sub} \geq C_{MC}$, *and for some* $\delta > 0$, $T \geq \underline{T}_{MC,1}(C_{sub}, \delta)$, *where* $C_{MC}$ *and* $\underline{T}_{MC,1}(\mathcal{C}, \delta)$ *are defined in Tables 2 and 4, respectively. Then we have*

$$\mathsf{P}\left(\bigcap_{i=1}^{s}\left\{N_i \geq \frac{T\ \ _\infty(i)}{C_{sub}\log(T)}\left(1 - \frac{1}{\ _\infty(i)}\sqrt{\log(\frac{s}{\delta})\frac{17C_{sub}\pi_{\max}\log(T)}{T}}\right) - \sum_{k=1}^{\lfloor T/L\rfloor} p_{kL}\right\}\right) \geq 1-\delta. \tag{B.103}$$

**Proof** For some $\ < \pi_{\min}/2$, we temporarily let $L \geq 6t_{MC}\log(\ ^{-1})$. From the proof for [17, Lemma 13 (47)], we know this guarantees $L \geq t_{MC}(\ /2)$. By definition of $t_{MC}(\cdot)$, we know $\max_i \|([\mathsf{T}^L]_{i,\cdot})^{\top} - \ _\infty\|_1 \leq \ \leq \pi_{\min}/2$, and since $([\mathsf{T}^L]_{i,\cdot})\mathbf{1} = \ ^{\top}_{\infty}\mathbf{1} = 1$, we further have

$$\max_i \|([\mathsf{T}^L]_{i,\cdot})^{\top} - \ _\infty\|_\infty \leq \frac{}{2} \leq \frac{\pi_{\min}}{4}. \tag{B.104}$$

For simplicity, we assume $\lfloor T/L\rfloor = T/L =: \tilde{T}$. To ease the notation, we let $\tilde{\omega}(k) := \omega(kL)$, $\tilde{A}_k := A_{kL}$, and $\tilde{\mathcal{F}}_k := \mathcal{F}_{kL}$. Then, one can see $\tilde{\omega}(k)$ and $\tilde{A}_k$ are both $\tilde{\mathcal{F}}_k$-measurable. Define $\ _k, \Delta_k \in \mathsf{R}^s$ such that

$$\ _j(i) := \mathbf{1}_{\{\tilde{\omega}(j)=i\}}\mathbf{1}_{\{\tilde{A}_j\}} - \mathsf{E}[\mathbf{1}_{\{\tilde{\omega}(j)=i\}}\mathbf{1}_{\{\tilde{A}_j\}} \mid \tilde{\mathcal{F}}_{j-1}], \tag{B.105}$$

$$\Delta_k(i) := \sum_{j=1}^{k}\ _j(i). \tag{B.106}$$

Note that for all $i \in [s]$, $\{\Delta_k(i), \tilde{\mathcal{F}}_k\}$ forms a martingale as

$$\begin{aligned}
\mathsf{E}[\Delta_{k+1}(i) \mid \tilde{\mathcal{F}}_k] &= \mathsf{E}[\sum_{j=1}^{k+1}\ _j(i) \mid \tilde{\mathcal{F}}_k] \\
&= \sum_{j=1}^{k}\ _j(i) + \mathsf{E}[\mathbf{1}_{\{\tilde{\omega}(k+1)=i\}}\mathbf{1}_{\{\tilde{A}_{k+1}\}} - \mathsf{E}[\mathbf{1}_{\{\tilde{\omega}(k+1)=i\}}\mathbf{1}_{\{\tilde{A}_{k+1}\}} \mid \tilde{\mathcal{F}}_k] \mid \tilde{\mathcal{F}}_k] \\
&= \sum_{j=1}^{k}\ _j(i) = \Delta_k(i),
\end{aligned} \tag{B.107}$$

thus $\ _k(i) = \Delta_k(i) - \Delta_{k-1}(i)$ can be viewed as the martingale difference sequence. Since $\mathsf{E}[\ _k(i) \mid \tilde{\mathcal{F}}_{k-1}] = 0$, we have $\mathsf{E}[\ _k(i)^2 \mid \tilde{\mathcal{F}}_{k-1}] = \mathrm{Var}(\ _k(i) \mid \tilde{\mathcal{F}}_{k-1}) = \mathrm{Var}(\mathbf{1}_{\{\tilde{\omega}(j)=i\}}\mathbf{1}_{\{\tilde{A}_j\}} \mid \tilde{\mathcal{F}}_{k-1}) \leq \mathsf{E}[\mathbf{1}^2_{\{\tilde{\omega}(k)=i\}}\mathbf{1}^2_{\{\tilde{A}_k\}} \mid \tilde{\mathcal{F}}_{k-1}] \leq \mathsf{E}[\mathbf{1}_{\{\tilde{\omega}(k)=i\}} \mid \tilde{\mathcal{F}}_{k-1}] = \mathsf{P}(\tilde{\omega}(k) = i \mid \tilde{\omega}(k-1)) = [\mathsf{T}^L]_{\omega((k-1)L),i}$. By the choice of $L$, using (B.104), we know $[\mathsf{T}^L]_{\omega((k-1)L),i} \leq \ _\infty(i) + \max_j \|([\mathsf{T}^L]_{j,\cdot})^{\top} - \ _\infty\|_\infty \leq 2\pi_{\max}$. Thus,

$$\sum_{k=1}^{\tilde{T}}\mathsf{E}[\ _k(i)^2 \mid \tilde{\mathcal{F}}_{k-1}] \leq 2\pi_{\max}\tilde{T}. \tag{B.108}$$

With this, and the fact that $|\ _k(i)| < 1$, we have

$$\begin{aligned}
\mathsf{P}\left(N_i - \sum_{k=1}^{\tilde{T}}\mathsf{E}[\mathbf{1}_{\{\tilde{\omega}(k)=i\}}\mathbf{1}_{\{\tilde{A}_k\}} \mid \tilde{\mathcal{F}}_{k-1}] \geq \tilde{T}\frac{}{2}\right) &\overset{\text{(i)}}{=} \mathsf{P}(\Delta_{\tilde{T}}(i) \geq \tilde{T}\ /2) \\
&\overset{\text{(ii)}}{\leq} \exp(-\frac{\tilde{T}\ ^2/8}{2\pi_{\max} + \ /6}) \\
&\overset{\text{(iii)}}{\leq} \exp(-\frac{T\ ^2}{17\pi_{\max}L}),
\end{aligned} \tag{B.109}$$

where (i) follows from the definition of $N_i$ and $\Delta_{\tilde{T}}(i)$; (ii) follows from Freedman's inequality [26, Theorem 1.6], and (iii) follows since $\epsilon \le \pi_{\min}/2$. Note that

$$\begin{aligned}
&\left|\sum_{k=1}^{\tilde{T}} \mathsf{E}[\mathbf{1}_{\{\tilde{\omega}(k)=i\}}\mathbf{1}_{\{\tilde{A}_k\}} \mid \tilde{\mathcal{F}}_{k-1}] - \tilde{T}\ \epsilon_\infty(i)\right| \\
\le &\left|\sum_{k=1}^{\tilde{T}} \mathsf{E}[\mathbf{1}_{\{\tilde{\omega}(k)=i\}} \mid \tilde{\mathcal{F}}_{k-1}] - \tilde{T}\ \epsilon_\infty(i)\right| + \left|\sum_{k=1}^{\tilde{T}} \mathsf{E}[\mathbf{1}_{\{\tilde{\omega}(k)=i\}} \mid \tilde{\mathcal{F}}_{k-1}] - \mathsf{E}[\mathbf{1}_{\{\tilde{\omega}(k)=i\}}\mathbf{1}_{\{\tilde{A}_k\}} \mid \tilde{\mathcal{F}}_{k-1}]\right| \\
\le &\tilde{T} \max_j |[\mathsf{T}^L]_{j,i} - \epsilon_\infty(i)| + \left|\sum_{k=1}^{\tilde{T}} \mathsf{E}[\mathbf{1}_{\{\tilde{A}_k^c\}} \mid \tilde{\mathcal{F}}_{k-1}]\right| \\
\le &\tilde{T}\frac{\epsilon}{2} + \sum_{k=1}^{\tilde{T}} p_{kL}.
\end{aligned} \tag{B.110}$$

Then, combining this with (B.109) and applying union bound, we have with probability at least $1 - s\exp(-\frac{T\epsilon^2}{17\pi_{\max}L})$,

$$\bigcap_{i=1}^{s}\left\{N_i \ge \frac{T}{L}\epsilon_\infty(i) - \frac{T}{L}\epsilon - \sum_{k=1}^{T/L} p_{kL}\right\} \tag{B.111}$$

when $\epsilon < \pi_{\min}/2$ and $L \ge 6t_{\mathsf{MC}}\log(\epsilon^{-1})$. Then, similar to the proof of Lemma B.1, we know if we pick $L = C_{sub}\log(T)$ with $C_{sub} \ge t_{\mathsf{MC}}\cdot\max\{3, 3-3\log(\pi_{\max}\log(s))\}$, and for some $\delta > 0$, we pick the trajectory length $T \ge \left(68C_{sub}\pi_{\max}\pi_{\min}^{-2}\log(\frac{s}{\delta})\right)^2$, with probability at least $1-\delta$, we have

$$\bigcap_{i=1}^{s}\left\{N_i \ge \frac{T\ \epsilon_\infty(i)}{C_{sub}\log(T)}\left(1 - \frac{1}{\epsilon_\infty(i)}\sqrt{\log(\frac{s}{\delta})\frac{17C_{sub}\pi_{\max}\log(T)}{T}}\right) - \sum_{k=1}^{T/L} p_{kL}\right\}. \tag{B.112}$$

■

**Lemma B.19** *For some $\delta > 0$, we assume $c_{\mathbf{z}} \ge (\sqrt{3}+\sqrt{6})\sqrt{p}$, $C_{sub} \ge \max\{C_{MC}, \underline{C}_{sub,\mathbf{x}}(\bar{x}_0, \delta, T, \rho_{\tilde{\mathbf{L}}}, \tau_{\tilde{\mathbf{L}}})\}$ and $T \ge \max\{\underline{T}_{MC,1}(C_{sub}, \frac{\delta}{2}), \underline{T}_{cl,1}(\rho_{\tilde{\mathbf{L}}}, \tau_{\tilde{\mathbf{L}}})\}$, where $C_{MC}$, $\underline{T}_{MC,1}(\mathcal{C}, \delta)$ and $\underline{T}_{cl,1}(\rho, \tau)$ are defined in Table 2, and $\underline{C}_{sub,\mathbf{x}}(\bar{x}_0, \delta, T, \rho, \tau)$ is defined in Table 4. Then, with probability at least $1-\delta$, the following intersected events occur*

$$\begin{aligned}
\bigcap_{i=1}^{s}\Bigg\{|S_i^{(\tau)}| \ge \frac{T\ \epsilon_\infty(i)}{C_{sub}\log(T)}\Bigg(1 - \frac{1}{\epsilon_\infty(i)}\sqrt{\log(\frac{2s}{\delta})\frac{17C_{sub}\pi_{\max}\log(T)}{T}} \\
- \frac{2n\sqrt{s}\tau_{\tilde{\mathbf{L}}}\bar{\sigma}^2}{\epsilon_\infty(i)c_{\mathbf{x}}^2\|\Sigma_{\mathbf{w}}\|\log(T)(1-\rho_{\tilde{\mathbf{L}}})} - \frac{1}{\epsilon_\infty(i)}e^{-\frac{c_{\mathbf{z}}^2}{3}}\Bigg)\Bigg\}.
\end{aligned} \tag{B.113}$$

**Proof** For simplicity, we assume $\lfloor T/L \rfloor = T/L$. We let $\mathcal{F}_t$ denote the sigma algebra generated by $\{\{\omega(r)\}_{r=0}^t, \mathsf{w}_{0:t}, \mathsf{z}_{0:t}, \mathsf{x}_0\}$, and let $A_t = \{\|\mathsf{x}_t\| \le c_{\mathbf{x}}\sqrt{\|\Sigma_{\mathbf{w}}\|\log(T)}, \|\mathsf{z}_t\| \le c_{\mathbf{z}}\sqrt{\|\Sigma_{\mathbf{z}}\|}\}$, then by Lemma B.18, when $C_{sub} \ge C_{MC} = t_{MC}\cdot\max\{3, 3-3\log(\pi_{\max}\log(s))\}$ and $T \ge \underline{T}_{MC,1}(C_{sub}, \frac{\delta}{2}) = \left(68C_{sub}\pi_{\max}\pi_{\min}^{-2}\log(\frac{2s}{\delta})\right)^2$, with probability at least $1-\frac{\delta}{2}$, we have

$$\begin{aligned}
\bigcap_{i=1}^{s}\Bigg\{|S_i^{(\tau)}| \ge \frac{T\ \epsilon_\infty(i)}{C_{sub}\log(T)}\cdot\left(1 - \frac{1}{\epsilon_\infty(i)}\sqrt{\log(\frac{2s}{\delta})\frac{17C_{sub}\pi_{\max}\log(T)}{T}}\right) \\
\underbrace{- \sum_{k=1}^{T/L} \mathsf{P}\left(A_t^c \mid \mathsf{x}_{(k-1)L+1}, \omega((k-1)L)\right)}_{=P}\Bigg\}.
\end{aligned} \tag{B.114}$$

For term $P$, we have

$$\begin{aligned} P &= \sum_{k=1}^{T/L} \mathsf{P}\left( \|\mathsf{x}_{kL}\| \geq c_{\mathbf{x}} \sqrt{\|\Sigma_{\mathbf{w}}\| \log(T)} \bigcup \|\mathsf{z}_{kL}\| \geq c_{\mathbf{z}} \sqrt{\|\Sigma_{\mathbf{z}}\|} \mid \mathsf{x}_{(k-1)L+1}, \omega((k-1)L) \right) \\ &\leq \underbrace{\sum_{k=1}^{T/L} \mathsf{P}\left( \|\mathsf{z}_{kL}\| \geq c_{\mathbf{z}} \sqrt{\|\Sigma_{\mathbf{z}}\|} \mid \mathsf{x}_{(k-1)L+1}, \omega((k-1)L) \right)}_{=: P_1} \\ &\quad + \underbrace{\sum_{k=1}^{T/L} \mathsf{P}\left( \|\mathsf{x}_{kL}\| \geq c_{\mathbf{x}} \sqrt{\|\Sigma_{\mathbf{w}}\| \log(T)} \mid \mathsf{x}_{(k-1)L+1}, \omega((k-1)L) \right)}_{=: P_2}. \end{aligned} \tag{B.115}$$

For term $P_1$, we know from Lemma A.5 that when $c_{\mathbf{z}} \geq (\sqrt{3}+\sqrt{6})\sqrt{p}$, we have $P_1 = \sum_{k=1}^{T/L} \mathsf{P}\left( \|\mathsf{z}_{kL}\| \geq c_{\mathbf{z}} \sqrt{\|\Sigma_{\mathbf{z}}\|} \right) \leq \frac{T}{L} e^{-\frac{c_{\mathbf{z}}^2}{3}}$. Now we consider term $P_2$. From Lemma A.4, we know

$$\mathsf{E}[\|\mathsf{x}_{kL}\|^2 \mid \mathsf{x}_{(k-1)L+1}, \omega((k-1)L)] \leq \sqrt{ns}\tau_{\tilde{\mathbf{L}}} \rho_{\tilde{\mathbf{L}}}^{L-1} \|\mathsf{x}_{(k-1)L+1}\|^2 + \frac{n\sqrt{s}\tau_{\tilde{\mathbf{L}}}\bar{\sigma}^2}{1-\rho_{\tilde{\mathbf{L}}}}, \tag{B.116}$$

thus by Markov inequality, we have

$$\begin{aligned} P_2 &\leq \sum_{k=1}^{T/L} \frac{1}{c_{\mathbf{x}}^2 \|\Sigma_{\mathbf{w}}\| \log(T)} \left( \sqrt{ns}\tau_{\tilde{\mathbf{L}}} \rho_{\tilde{\mathbf{L}}}^{L-1} \|\mathsf{x}_{(k-1)L+1}\|^2 + \frac{n\sqrt{s}\tau_{\tilde{\mathbf{L}}}\bar{\sigma}^2}{1-\rho_{\tilde{\mathbf{L}}}} \right) \\ &\leq \frac{1}{c_{\mathbf{x}}^2 \|\Sigma_{\mathbf{w}}\| \log(T)} \left( \frac{T}{L} \frac{n\sqrt{s}\tau_{\tilde{\mathbf{L}}}\bar{\sigma}^2}{1-\rho_{\tilde{\mathbf{L}}}} + \sqrt{ns}\tau_{\tilde{\mathbf{L}}} \rho_{\tilde{\mathbf{L}}}^{L-1} \sum_{k=1}^{T/L} \|\mathsf{x}_{(k-1)L+1}\|^2 \right). \end{aligned} \tag{B.117}$$

Now, we seek to upper bound $\rho_{\tilde{\mathbf{L}}}^{L-1} \sum_{k=1}^{T/L} \|\mathsf{x}_{(k-1)L+1}\|^2$ with high probability. Note that the assumption $C_{sub} \geq \underline{C}_{sub,\mathbf{x}}(\bar{x}_0, \delta, T, \rho_{\tilde{\mathbf{L}}}, \tau_{\tilde{\mathbf{L}}})$ implies the following

$$L = C_{sub} \log(T) \geq \frac{1}{\log(\rho_{\tilde{\mathbf{L}}}^{-1})} \max\left\{ \log(2), 2\log\left(\frac{8\sqrt{ns}\tau_{\tilde{\mathbf{L}}}\bar{x}_0^2}{\delta}\right), 2\log\left(4T\frac{n\sqrt{s}\tau_{\tilde{\mathbf{L}}}\bar{\sigma}^2}{\delta(1-\rho_{\tilde{\mathbf{L}}})}\right) + 2, \right\}. \tag{B.118}$$

Then, we have

$$\begin{aligned} \mathsf{P}\left( \rho_{\tilde{\mathbf{L}}}^{L-1} \sum_{k=1}^{T/L} \|\mathsf{x}_{(k-1)L+1}\|^2 \leq \frac{T}{L} \frac{(1-\rho_{\tilde{\mathbf{L}}})}{4Tn\sqrt{s}\tau_{\tilde{\mathbf{L}}}\bar{\sigma}^2} \right) &\overset{\text{(i)}}{\geq} \mathsf{P}\left( \rho_{\tilde{\mathbf{L}}}^{L-1} \sum_{k=1}^{T/L} \|\mathsf{x}_{(k-1)L+1}\|^2 \leq \frac{T}{L} \rho_{\tilde{\mathbf{L}}}^{\frac{L}{2}-1} \right) \\ &\geq \mathsf{P}\left( \bigcap_{k=1}^{T/L} \left\{ \|\mathsf{x}_{(k-1)L+1}\|^2 \leq \rho_{\tilde{\mathbf{L}}}^{-\frac{L}{2}} \right\} \right) \\ &\geq 1 - \sum_{k=1}^{T/L} \mathsf{P}\left( \|\mathsf{x}_{(k-1)L+1}\|^2 \geq \rho_{\tilde{\mathbf{L}}}^{-\frac{L}{2}} \right) \\ &\overset{\text{(ii)}}{\geq} 1 - \sum_{k=1}^{T/L} \rho_{\tilde{\mathbf{L}}}^{\frac{L}{2}} \left( \sqrt{ns}\tau_{\tilde{\mathbf{L}}} \rho_{\tilde{\mathbf{L}}}^{(k-1)L+1} \bar{x}_0^2 + \frac{n\sqrt{s}\tau_{\tilde{\mathbf{L}}}\bar{\sigma}^2}{1-\rho_{\tilde{\mathbf{L}}}} \right) \\ &\geq 1 - \rho_{\tilde{\mathbf{L}}}^{\frac{L}{2}+1} \frac{\sqrt{ns}\tau_{\tilde{\mathbf{L}}}\bar{x}_0^2}{1-\rho_{\tilde{\mathbf{L}}}^{L}} - \frac{T}{L} \rho_{\tilde{\mathbf{L}}}^{\frac{L}{2}} \frac{n\sqrt{s}\tau_{\tilde{\mathbf{L}}}\bar{\sigma}^2}{1-\rho_{\tilde{\mathbf{L}}}} \\ &\overset{\text{(iii)}}{\geq} 1 - 2\rho_{\tilde{\mathbf{L}}}^{\frac{L}{2}} \sqrt{ns}\tau_{\tilde{\mathbf{L}}}\bar{x}_0^2 - \frac{\delta}{4L} \\ &\overset{\text{(iv)}}{\geq} 1 - \frac{\delta}{4} - \frac{\delta}{4} = 1 - \frac{\delta}{2}, \end{aligned} \tag{B.119}$$

where (i) follows from (B.118) which gives $\rho_{\tilde{\mathbf{L}}}^{\frac{L}{2}-1} \le \frac{(1-\sqrt{\rho_{\tilde{\mathbf{L}}}})}{4Tn\sqrt{s}\tau_{\tilde{\mathbf{L}}}\bar{\sigma}^2}$; (ii) follows from Lemma A.4 and Markov inequality; (iii) follows from (B.118) which gives $\rho_{\tilde{\mathbf{L}}}^{L} \le \frac{1}{2}$ and $\rho_{\tilde{\mathbf{L}}}^{\frac{L}{2}} \le \frac{\delta(1-\sqrt{\rho_{\tilde{\mathbf{L}}}})}{4Tn\sqrt{s}\tau_{\tilde{\mathbf{L}}}\bar{\sigma}^2}$ and (iv) follows from (B.118) which gives $\rho_{\tilde{\mathbf{L}}}^{\frac{L}{2}} \le \frac{\delta}{8}\frac{1}{n\sqrt{s}\tau_{\tilde{\mathbf{L}}}\bar{x}_0^2}$. Therefore, we have with probability at least $1-\frac{\delta}{2}$

$$P_2 \le \frac{1}{c_{\mathbf{x}}^2\|\Sigma_{\mathbf{w}}\|\log(T)}\left(\frac{T}{L}\frac{n\sqrt{s}\tau_{\tilde{\mathbf{L}}}\bar{\sigma}^2}{1-\rho_{\tilde{\mathbf{L}}}} + \frac{1-\rho_{\tilde{\mathbf{L}}}}{4L\sqrt{n}\bar{\sigma}^2}\right), \tag{B.120}$$

and thus,

$$\begin{aligned} P \le P_1 + P_2 &\le \frac{1}{c_{\mathbf{x}}^2\|\Sigma_{\mathbf{w}}\|\log(T)}\left(\frac{T}{L}\frac{n\sqrt{s}\tau_{\tilde{\mathbf{L}}}\bar{\sigma}^2}{1-\rho_{\tilde{\mathbf{L}}}} + \frac{1-\rho_{\tilde{\mathbf{L}}}}{4L\sqrt{n}\bar{\sigma}^2}\right) + \frac{T}{L}e^{-\frac{c_{\mathbf{z}}^2}{3}} \\ &\le \frac{1}{c_{\mathbf{x}}^2\|\Sigma_{\mathbf{w}}\|\log(T)}\left(\frac{T}{L}\frac{2n\sqrt{s}\tau_{\tilde{\mathbf{L}}}\bar{\sigma}^2}{1-\rho_{\tilde{\mathbf{L}}}}\right) + \frac{T}{L}e^{-\frac{c_{\mathbf{z}}^2}{3}}, \end{aligned} \tag{B.121}$$

where the second inequality follows from $T \ge \underline{T}_{cl,1}(\rho_{\tilde{\mathbf{L}}}, \tau_{\tilde{\mathbf{L}}})$. Plugging this into (B.114), we have with probability at least $1-\delta$,

$$\begin{aligned} \bigcap_{i=1}^{s}\Bigg\{|S_i^{(\tau)}| \ge \frac{T}{C_{sub}\log(T)}\cdot\Bigg(1 - \frac{1}{\mathcal{C}_\infty(i)}\sqrt{\log(\frac{2s}{\delta})\frac{17C_{sub}\pi_{\max}\log(T)}{T}} \\ - \frac{2n\sqrt{s}\tau_{\tilde{\mathbf{L}}}\bar{\sigma}^2}{\mathcal{C}_\infty(i)c_{\mathbf{x}}^2\|\Sigma_{\mathbf{w}}\|\log(T)(1-\rho_{\tilde{\mathbf{L}}})} - \frac{1}{\mathcal{C}_\infty(i)}e^{-\frac{c_{\mathbf{z}}^2}{3}}.\Bigg)\Bigg\}, \end{aligned} \tag{B.122}$$

which concludes the proof. ∎

Note that when $T$, $c_{\mathbf{x}}$, and $c_{\mathbf{z}}$ are sufficiently large enough, we could obtain a more interpretable version of Lemma B.19 which is presented as follows.

**Corollary B.20** *For some $\delta > 0$, assume $c_{\mathbf{x}} \ge \frac{1}{3}\underline{c}_{\mathbf{x}}(\rho_{\tilde{\mathbf{L}}}, \tau_{\tilde{\mathbf{L}}})$, $c_{\mathbf{z}} \ge \underline{c}_{\mathbf{z}}$, $C_{sub} \ge \max\{C_{MC}, \underline{C}_{sub,\mathbf{x}}(\bar{x}_0, \delta, T, \rho_{\tilde{\mathbf{L}}}, \tau_{\tilde{\mathbf{L}}})\}$, and $T \ge \underline{T}_N(C_{sub}, 2\delta, \rho_{\tilde{\mathbf{L}}}, \tau_{\tilde{\mathbf{L}}}) := \max\{\underline{T}_{MC}(C_{sub}, \delta), \underline{T}_{cl,1}(\rho_{\tilde{\mathbf{L}}}, \tau_{\tilde{\mathbf{L}}})\}$, where $\underline{c}_{\mathbf{x}}(\rho, \tau)$, $\underline{c}_{\mathbf{z}}$, $C_{MC}$, $\underline{T}_{MC}(\mathcal{C}, \delta)$, $\underline{T}_{cl,1}(\rho, \tau)$ are defined in Table 2, and $\underline{C}_{sub,\mathbf{x}}(\bar{x}_0, \delta, T, \rho, \tau)$ is defined in Table 4. Then, with probability at least $1-\delta/2$, the following intersected events occur*

$$\bigcap_{i=1}^{s}\left\{|S_i^{(\tau)}| \ge \frac{T\pi_{\min}}{2C_{sub}\log(T)}\right\}. \tag{B.123}$$

Now we provide a result on how many data in $S_i^{(\tau)}$ are "weakly" independent, which is the quantity that essentially determines the sample complexity of estimating $\mathsf{A}_{1:s}$ and $\mathsf{B}_{1:s}$ in Algorithm 1. We first define a few notations. Let $\tau_{i,1}, \ldots, \tau_{i,|S_i^{(\tau)}|}$ denote the elements in $S_i^{(\tau)}$, and let $\tau_{i,0} = 0$. Define $\bar{\mathbf{x}}_{\tau_{i,k}}$ such that

$$\bar{\mathbf{x}}_{\tau_{i,k}} = \sum_{j=1}^{\tau_{i,k-1}}\left(\prod_{k=1}^{j-1}\mathsf{L}_{\omega(t-k)}\right)\left(\mathsf{B}_{\omega(t-j)}\mathsf{z}_{t-j} + \mathsf{w}_{t-j}\right) + \mathsf{B}_{\omega(\tau_{i,k}-1)}\mathsf{z}_{\tau_{i,k}-1} + \mathsf{w}_{\tau_{i,k}-1}. \tag{B.124}$$

One can view $\bar{\mathbf{x}}_{\tau_{i,k}}$ as follows: set $\mathbf{x}_{\tau_{i,k-1}} = 0$, then propagate the dynamics to time $\tau_{i,k}$ following the same noise and mode switching sequences, $\mathsf{w}_{\tau_{i,k-1}:\tau_{i,k}-1}$, $\mathsf{z}_{\tau_{i,k-1}:\tau_{i,k}-1}$, $\{\omega(t')\}_{t'=\tau_{i,k-1}}^{\tau_{i,k}-1}$. Or, one can also view $\bar{\mathbf{x}}_{\tau_{i,k}}$ as the contribution of noise $\mathbf{x}_t$ and $\mathsf{z}_t$ that propagate $\mathbf{x}_{\tau_{i,k-1}}$ to $\mathbf{x}_{\tau_{i,k}}$. And it is easy to see that

$$\mathbf{x}_{\tau_{i,k}} - \bar{\mathbf{x}}_{\tau_{i,k}} = \left(\prod_{k=1}^{\tau_{i,k-1}}\mathsf{L}_{\omega(t-k)}\right)\mathbf{x}_{\tau_{i,k-1}}. \tag{B.125}$$

Define $\bar{S}_i^{(\tau)} \subseteq S_i^{(\tau)}$ such that

$$\bar{S}_i^{(\tau)} := \left\{ \tau_k \middle| \omega(\tau_k)=i,\ \|\mathsf{x}_{\tau_k}\| \le c_{\mathbf{x}}\sqrt{\|\Sigma_{\mathbf{w}}\|\log(T)},\ \|\mathsf{z}_{\tau_k}\| \le c_{\mathbf{z}}\sqrt{\Sigma_{\mathbf{z}}},\ \|\bar{\mathsf{x}}_{\tau_k}\| \le \frac{c_{\mathbf{x}}\sqrt{\|\Sigma_{\mathbf{w}}\|\log(T)}}{2} \right\}.$$

The next lemma provides a lower bound on $|\bar{S}_i^{(\tau)}|$.

**Lemma B.21** *Assume $c_{\mathbf{x}} \ge \underline{c}_{\mathbf{x}}(\rho_{\tilde{\mathbf{L}}}, \tau_{\tilde{\mathbf{L}}})$, $c_{\mathbf{z}} \ge \underline{c}_{\mathbf{z}}$, $C_{sub} \ge \underline{C}_{sub,N}(\bar{x}_0, \delta, T, \rho_{\tilde{\mathbf{L}}}, \tau_{\tilde{\mathbf{L}}}) := \max\{C_{MC}, \underline{C}_{sub,\mathbf{x}}(\bar{x}_0, \frac{\delta}{2}, T, \rho_{\tilde{\mathbf{L}}}, \tau_{\tilde{\mathbf{L}}}), \underline{C}_{sub,\bar{\mathbf{x}}}(\delta, T, \rho_{\tilde{\mathbf{L}}}, \tau_{\tilde{\mathbf{L}}})\}$, and $T \ge \underline{T}_N(C_{sub}, \delta, \rho_{\tilde{\mathbf{L}}}, \tau_{\tilde{\mathbf{L}}})$, where $\underline{c}_{\mathbf{x}}(\rho, \tau)$, $\underline{c}_{\mathbf{z}}$, $\underline{C}_{sub,\mathbf{x}}(\bar{x}_0, \delta, T, \rho, \tau)$, and $\underline{C}_{sub,\bar{\mathbf{x}}}(\delta, T, \rho, \tau)$ are defined in Table 4, and $C_{MS}$ and $\underline{T}_N(\mathcal{C}, \delta, \rho, \tau)$ are defined in Table 2. Then with probability at least $1-\delta$, the following intersected events occur*

$$\bigcap_{i=1}^{s} \left\{ |\bar{S}_i^{(\tau)}| \ge \frac{T\pi_{\min}}{2C_{sub}\log(T)} \right\}. \tag{B.126}$$

**Proof** We define sets $R_i^{(\tau)} \subseteq S_i^{(\tau)}$ and $\bar{R}_i^{(\tau)} \subseteq \bar{S}_i^{(\tau)}$ such that

$$R_i^{(\tau)} := \left\{ \tau_k \middle| \omega(\tau_k)=i,\ \|\mathsf{x}_{\tau_k}\| \le \frac{c_{\mathbf{x}}\sqrt{\|\Sigma_{\mathbf{w}}\|\log(T)}}{3},\ \|\mathsf{z}_{\tau_k}\| \le c_{\mathbf{z}}\sqrt{\Sigma_{\mathbf{z}}} \right\}.$$
$$\bar{R}_i^{(\tau)} := \left\{ \tau_k \middle| \omega(\tau_k)=i,\ \|\mathsf{x}_{\tau_k}\| \le \frac{c_{\mathbf{x}}\sqrt{\|\Sigma_{\mathbf{w}}\|\log(T)}}{3},\ \|\mathsf{z}_{\tau_k}\| \le c_{\mathbf{z}}\sqrt{\Sigma_{\mathbf{z}}},\ \|\bar{\mathsf{x}}_{\tau_k}\| \le \frac{c_{\mathbf{x}}\sqrt{\|\Sigma_{\mathbf{w}}\|\log(T)}}{2} \right\}.$$

Note that $\bar{R}_i^{(\tau)} \subseteq R_i^{(\tau)}$. We will first (i) lower bound $|R_i^{(\tau)}|$ and (ii) show $|\bar{R}_i^{(\tau)}| = |R_i^{(\tau)}|$, then we could lower bound $|\bar{S}_i^{(\tau)}|$ since $|\bar{S}_i^{(\tau)}| \ge |\bar{R}_i^{(\tau)}|$ and conclude the proof.

Using Corollary B.20, we see under given assumptions, with probability at least $1-\frac{\delta}{2}$,

$$\bigcap_{i=1}^{s} \left\{ |R_i^{(\tau)}| \ge \frac{T\pi_{\min}}{2C_{sub}\log(T)} \right\}. \tag{B.127}$$

Let $\zeta_{i,1}, \ldots, \zeta_{i,\lfloor R_i^{(\tau)} \rfloor}$ denote the elements in $R_i^{(\tau)}$. It is easy to see $\{\zeta_{i,1}, \ldots, \zeta_{i,\lfloor R_i^{(\tau)} \rfloor}\} \subseteq \{\tau_{i,1}, \ldots, \tau_{i,\lfloor S_i^{(\tau)} \rfloor}\}$. Consider an arbitrary $\zeta_{i,j} \in R_i^{(\tau)}$ and $\tau_{i,j'} \in S_i^{(\tau)}$ denote the counterpart of $\zeta_{i,j}$ such that $\tau_{i,j'} = \zeta_{i,j}$. By definition of $R_i^{(\tau)}$, we have

$$\|\mathsf{x}_{\tau_{i,j'}}\| \le \frac{c_{\mathbf{x}}\sqrt{\|\Sigma_{\mathbf{w}}\|\log(T)}}{3}. \tag{B.128}$$

From (B.125), together with Lemma A.4, we have

$$\begin{aligned} \mathsf{E}[\|\mathsf{x}_{\tau_{i,j'}} - \bar{\mathsf{x}}_{\tau_{i,j'}}\|^2] &\le \sqrt{ns}\tau_{\tilde{\mathbf{L}}}\rho_{\tilde{\mathbf{L}}}^{\tau_{i,j'}-\tau_{i,j'-1}}\mathsf{E}[\|\mathsf{x}_{\tau_{i,j'-1}}\|^2] \\ &\le \sqrt{ns}\tau_{\tilde{\mathbf{L}}}\rho_{\tilde{\mathbf{L}}}^{L}(c_{\mathbf{x}}^2\|\Sigma_{\mathbf{w}}\|\log(T)), \end{aligned} \tag{B.129}$$

where the second inequality follows from $\tau_{i,j'} - \tau_{i,j'-1} \ge L$ and $\|\mathsf{x}_{\tau_{i,j'-1}}\| \le c_{\mathbf{x}}\sqrt{\|\Sigma_{\mathbf{w}}\|\log(T)}$ by definition of $S_i^{(\tau)}$. Then, by Markov inequality, we have

$$\mathsf{P}\left( \|\mathsf{x}_{\tau_{i,j'}} - \bar{\mathsf{x}}_{\tau_{i,j'}}\| \le \frac{c_{\mathbf{x}}\sqrt{\|\Sigma_{\mathbf{w}}\|\log(T)}}{6} \right) \ge 1 - 36\sqrt{ns}\tau_{\tilde{\mathbf{L}}}\rho_{\tilde{\mathbf{L}}}^{L}. \tag{B.130}$$

Then, using union bound, we have

$$\begin{aligned}
&\mathsf{P}\left(\bigcap_{i\in[s]}\bigcap_{j'}\left\{\|\mathsf{x}_{\tau_{i,j'}}-\bar{\mathsf{x}}_{\tau_{i,j'}}\|\le\frac{c_{\mathbf{x}}\sqrt{\|\Sigma_{\mathbf{w}}\|\log(T)}}{6}\right\}\right)\\
\ge&1-36\sqrt{n}s^{1.5}|R_i^{(\tau)}|\tau_{\tilde{\mathsf{L}}}\rho_{\tilde{\mathsf{L}}}^{L}\\
\ge&1-36\sqrt{n}s^{1.5}T\tau_{\tilde{\mathsf{L}}}\rho_{\tilde{\mathsf{L}}}^{L}\\
\ge&1-\frac{\delta}{2},
\end{aligned}\tag{B.131}$$

where the last line follows from $L=C_{sub}\log(T)$ and $C_{sub}\ge\underline{C}_{sub,\bar{\mathbf{x}}}(\delta,T,\rho_{\tilde{\mathsf{L}}},\tau_{\tilde{\mathsf{L}}})$ in the assumption. Note that $\|\bar{\mathsf{x}}_{\zeta_{i,j}}\|=\|\bar{\mathsf{x}}_{\tau_{i,j'}}\|\le\|\mathsf{x}_{\tau_{i,j'}}\|+\|\mathsf{x}_{\tau_{i,j'}}-\bar{\mathsf{x}}_{\tau_{i,j'}}\|$. This together with (B.128) and (B.131) gives, with probability at least $1-\frac{\delta}{2}$,

$$\bigcap_{i\in[s]}\bigcap_{j\in[|R_i^{(\tau)}|]}\left\{\|\bar{\mathsf{x}}_{\zeta_{i,j}}\|\le\frac{c_{\mathbf{x}}\sqrt{\|\Sigma_{\mathbf{w}}\|\log(T)}}{2}\right\}.\tag{B.132}$$

This implies for any $i$, for any $\tau_k\in R_i^{(\tau)}$, we have $\tau_k\in\bar{R}_i^{(\tau)}$, i.e. $R_i^{(\tau)}\subseteq\bar{R}_i^{(\tau)}$. Thus, we have $R_i^{(\tau)}=\bar{R}_i^{(\tau)}$ and $|R_i^{(\tau)}|=|\bar{R}_i^{(\tau)}|$. Combining this with (B.127), we have with probability at least $1-\delta$,

$$\bigcap_{i=1}^{s}\left\{|\bar{R}_i^{(\tau)}|\ge\frac{T\pi_{\min}}{2C_{sub}\log(T)}\right\}.\tag{B.133}$$

Finally, we could conclude the proof by noticing $|\bar{S}_i^{(\tau)}|\ge|\bar{R}_i^{(\tau)}|$. ∎

# C MJS Regret Analysis

Consider MJS-LQR($\mathsf{A}_{1:s},\mathsf{B}_{1:s},\mathsf{T},\mathsf{Q}_{1:s},\mathsf{R}_{1:s}$) with dynamics noise $\mathsf{w}_t\sim\mathcal{N}(0,\Sigma_{\mathbf{w}})$, some arbitrary initial state $\mathsf{x}_0$ and stabilizing controller $\mathsf{K}_{1:s}$. The input is $\mathsf{u}_t=\mathsf{K}_{\omega(t)}\mathsf{x}_t+\mathsf{z}_t$ where exploration noise $\mathsf{z}_t\sim\mathcal{N}(0,\Sigma_{\mathbf{z}})$. Let $\mathsf{L}_i:=\mathsf{A}_i+\mathsf{B}_i\mathsf{K}_i$. Let $\tilde{\mathsf{L}}\in\mathsf{R}^{sn^2\times sn^2}$ denote the augmented closed-loop state matrix with $ij$-th $n^2\times n^2$ block given by $[\tilde{\mathsf{L}}]_{ij}:=[\mathsf{T}]_{ji}\mathsf{L}_j\otimes\mathsf{L}_j$. Let $\tau_{\tilde{\mathsf{L}}}>0$ and $\rho_{\tilde{\mathsf{L}}}\in[0,1)$ be two constants such that $\|\tilde{\mathsf{L}}^k\|\le\tau_{\tilde{\mathsf{L}}}\rho_{\tilde{\mathsf{L}}}^k$. By definition, one available choice for $\tau_{\tilde{\mathsf{L}}}$ and $\rho_{\tilde{\mathsf{L}}}$ are $\tau(\tilde{\mathsf{L}})$ and $\rho(\tilde{\mathsf{L}})$.

We define the following cumulative cost conditioned on the initial state $\mathsf{x}_0$, initial mode $\omega(0)$, and controller $\mathsf{K}_{1:s}$.

$$J_T(\mathsf{x}_0,\omega(0),\{\mathsf{K}_{1:s},\Sigma_{\mathbf{z}}\}):=\sum_{t=1}^{T}\mathsf{E}[\mathsf{x}_t^\top\mathsf{Q}_{\omega(t)}\mathsf{x}_t+\mathsf{u}_t^\top\mathsf{R}_{\omega(t)}\mathsf{u}_t\mid\mathsf{x}_0,\omega(0),\mathsf{K}_{1:s}].\tag{C.1}$$

The definition of this cumulative cost coincides with the cost $\sum_{t=1}^{T_i}c_{T_0+\cdots+T_{i-1}+t}$ in the definition of $\mathrm{Regret}_i$ in (5.4) with $\mathsf{x}_0,\omega(0),\mathsf{K}_{1:s}$ setting to $\mathsf{x}_0^{(i)},\omega^{(i)}(0),\mathsf{K}_{1:s}^{(i)}$ since $\mathrm{Regret}_i$ depends on randomness in $\mathcal{F}_{i-1}$ only through $\mathsf{x}_0^{(i)},\omega^{(i)}(0),\mathsf{K}_{1:s}^{(i)}$. In the remainder of this appendix, for simplicity, we will drop the conditions $\mathsf{x}_0,\omega(0),\mathsf{K}_{1:s}$ in the expectation and simply write $\mathsf{E}[\,\cdot\mid\mathsf{x}_0,\omega(0),\mathsf{K}_{1:s}]$ as $\mathsf{E}[\cdot]$. So, for any measurable function $f$, $\mathsf{E}[f(\mathsf{x}_0,\omega(0),\mathsf{K}_{1:s})]=f(\mathsf{x}_0,\omega(0),\mathsf{K}_{1:s})$. Note that even though the results in this appendix are derived for conditional expectation $\mathsf{E}[\,\cdot\mid\mathsf{x}_0,\omega(0),\mathsf{K}_{1:s}]$, most of them also hold for the total expectation $\mathsf{E}[\cdot]$.

For the infinite-horizon case, we define the following infinite-horizon average cost without exploration noise $\mathsf{z}_t$ and starting from $\mathsf{x}_0=0$.

$$J(0,\omega(0),\{\mathsf{K}_{1:s}\}):=\limsup_{T\to\infty}\frac{1}{T}J_T(0,\omega(0),\{\mathsf{K}_{1:s},0\}).\tag{C.2}$$

Let $P^*_{1:s}$ denote the solution to cDARE($A_{1:s}, B_{1:s}, T, Q_{1:s}, R_{1:s}$) defined in (5.2). Let $K^*_{1:s}$ denote the resulting infinite-horizon optimal controller computed using $P^*_{1:s}$ and following (5.1). Note that the infinite-horizon optimal average cost $J^*$ in (3.4) is achieved if the optimal controller $K^*_{1:s}$ is used, i.e.

$$J^* = J(0, \omega(0), \{K^*_{1:s}\}). \tag{C.3}$$

Note that if the underlying Markov chain $T$ is ergodic, for any initial state $x_0$ and mode $\omega(0)$, $J^* = J(x_0, \omega(0), \{K^*_{1:s}\})$. Let $L^*_i = A_i + B_i K^*_i$ for all $i \in [s]$ denote the closed-loop state matrix when the optimal controller $K^*_{1:s}$ is used. Define the augmented state matrix $\tilde{L}^*$ such that its $ij$-th block is given by $[\tilde{L}^*]_{ij} := [T]_{ji} L^*_j \otimes L^*_j$. From [14], we know $K^*_{1:s}$ is stabilizes the MJS, thus $\rho^* := \rho(\tilde{L}^*) < 1$.

Since $\text{Regret}_i$ defined in (5.4) can be written as

$$\text{Regret}_i = J_T(x_0^{(i)}, \omega^{(i)}(0), \{K^{(i)}_{1:s}, \sigma^2_{z,i} I_\rho\}) - TJ^*, \tag{C.4}$$

to evaluate $\text{Regret}(T)$, it suffices to evaluate $J_T(x_0, \omega(0), \{K_{1:s}, \Sigma_z\}) - TJ^*$ for generic $x_0$, $\omega(0)$, $K_{1:s}$, and $\Sigma_z$. The outline of this Appendix C is as follows.

- In Appendix C.1, we restate perturbation results [18] on $J(0, \omega(0), \{K_{1:s}\}) - J^*$.
- In Appendix C.2, we evaluate $J_T(x_0, \omega(0), \{K_{1:s}, \Sigma_z\}) - TJ(0, \omega(0), \{K_{1:s}\})$. Then, applying the results in Appendix C.1, we can bound the single epoch regret $J_T(x_0, \omega(0), \{K_{1:s}, \Sigma_z\}) - TJ^*$.
- In Appendix C.3, we stitch regrets for all epochs together, and combine them with identification results in Appendix B to bound $\text{Regret}(T)$.

## C.1 MJS-LQR Perturbation Results

We first present a lemma on the perturbation of augmented closed-loop state matrix if we use a controller $K_{1:s}$ that is close to the optimal $K^*_{1:s}$.

**Lemma C.1 (Lemma 9 in [18] )** *For an arbitrary controller $K_{1:s}$, let $L_i = A_i + B_i K_i$ for all $i \in [s]$, and let $\tilde{L}$ be the augmented state matrix such that its $ij$-th block is given by $[\tilde{L}]_{ij} := [T]_{ji} L_j \otimes L_j$. Assume $\|K_{1:s} - K^*_{1:s}\| \le \bar{\epsilon}_K$, where $\bar{\epsilon}_K$ is defined in Table 3. Then, we have*

$$\|\tilde{L}^k\| \le \tau(\tilde{L}^*)\left(\frac{1+\rho^*}{2}\right)^k, \quad \forall k \in \mathbb{N}, \tag{C.5}$$

$$\rho(\tilde{L}) \le \frac{1+\rho^*}{2}. \tag{C.6}$$

*Thus controller $K_{1:s}$ is stabilizing.*

The following perturbation results show how much the infinite-horizon average cost deviates depending on the deviations from the optimal controller and how much the optimal controller deviates depending on the model accuracy for the MJS-LQR problem.

**Lemma C.2 (Perturbation of Infinite-horizon MJS-LQR, Corollary 11 and Theorem 6 in [18])** *Infinite-horizon MJS-LQR($A_{1:s}, B_{1:s}, T, Q_{1:s}, R_{1:s}$) problems have the following perturbation results. Note that notations $\bar{\epsilon}_K$, $\bar{\epsilon}_{A,B,T}$, and $C^K_{A,B,T}$ are defined in Table 3.*

1. *Suppose we have an arbitrary controller $K_{1:s}$ such that $\|K_{1:s} - K^*_{1:s}\| \le \bar{\epsilon}_K$. Then, we have*

$$J(0, \omega(0), \{K_{1:s}\}) - J^* \le C^J_K \|\Sigma_w\| \|K_{1:s} - K^*_{1:s}\|^2. \tag{C.7}$$

2. *Suppose there is an arbitrary MJS($\hat{A}_{1:s}, \hat{B}_{1:s}, \hat{T}$) with $\epsilon_{A,B} := \max\{\|\hat{A}_{1:s} - A_{1:s}\|, \|\hat{B}_{1:s} - B_{1:s}\|\} \le \bar{\epsilon}_{A,B,T}$, and $\epsilon_T := \|\hat{T} - T\|_\infty \le \bar{\epsilon}_{A,B,T}$. Then, there exists an optimal controller $K_{1:s}$ to the infinite-horizon MJS-LQR($\hat{A}_{1:s}, \hat{B}_{1:s}, \hat{T}, Q_{1:s}, R_{1:s}$) and it can be computed using (5.1) and (5.2), and we have*

$$\|K_{1:s} - K^*_{1:s}\| \le C^K_{A,B,T}(\epsilon_{A,B} + \epsilon_T). \tag{C.8}$$

*By definition of $\bar{\epsilon}_{A,B,T}$, we see $\|K_{1:s} - K^*_{1:s}\| \le \bar{\epsilon}_K$, thus Lemma C.1 is applicable.*

## C.2 Single Epoch Regret Analysis

Recall the definitions of $\tilde{\mathsf{B}}_t$ and $\tilde{\Pi}_t$ in (A.4) of Appendix A.1. Furthermore, we define

$$\tilde{\Pi}_\infty = \quad_\infty \otimes \mathsf{I}_{n^2}, \quad \tilde{\mathsf{R}}_t = \sum_{i=1}^{s} \quad_t(i)\mathsf{R}_i. \tag{C.9}$$

For a set of matrices $\mathsf{V}_{1:s}$, define the following reshaping mapping

$$\mathcal{H}(\begin{bmatrix}\mathsf{V}_1\\ \vdots \\ \mathsf{V}_s\end{bmatrix}) = \begin{bmatrix}\mathrm{vec}(\mathsf{V}_1)\\ \vdots \\ \mathrm{vec}(\mathsf{V}_s)\end{bmatrix}, \tag{C.10}$$

and let $\mathcal{H}^{-1}$ denote the inverse mapping of $\mathcal{H}$. Let

$$\mathsf{M}_i := \mathsf{Q}_i + \mathsf{K}_i^\top \mathsf{R}_i \mathsf{K}_i, \qquad \mathsf{M} := [\mathsf{M}_1, \ldots, \mathsf{M}_s]. \tag{C.11}$$

We define

$$\begin{aligned} N_{0,t} &= \mathrm{tr}(\mathsf{M}\mathcal{H}^{-1}(\tilde{\mathsf{L}}^t \begin{bmatrix}\mathrm{vec}(\Sigma_1(0))\\ \vdots \\ \mathrm{vec}(\Sigma_s(0))\end{bmatrix})),\\ N_{\mathbf{z},1,t} &= \mathrm{tr}(\mathsf{M}\mathcal{H}^{-1}((\tilde{\mathsf{B}}_t + \tilde{\mathsf{L}}\tilde{\mathsf{B}}_{t-1} + \cdots + \tilde{\mathsf{L}}^{t-1}\tilde{\mathsf{B}}_1)\mathrm{vec}(\Sigma_{\mathbf{z}}))),\\ N_{\mathbf{w},t} &= \mathrm{tr}(\mathsf{M}\mathcal{H}^{-1}((\tilde{\Pi}_t + \tilde{\mathsf{L}}\tilde{\Pi}_{t-1} + \cdots + \tilde{\mathsf{L}}^{t-1}\tilde{\Pi}_1)\mathrm{vec}(\Sigma_{\mathbf{w}}))),\\ N_{\mathbf{z},2,t} &= \mathrm{tr}(\tilde{\mathsf{R}}_t \Sigma_z), \end{aligned} \tag{C.12}$$

and

$$S_{0,T} = \sum_{t=1}^{T} N_{0,t}, \quad S_{\mathbf{z},1,T} = \sum_{t=1}^{T} N_{\mathbf{z},1,t}, \quad S_{\mathbf{w},T} = \sum_{t=1}^{T} N_{\mathbf{w},t}, \quad S_{\mathbf{z},2,T} = \sum_{t=1}^{T} N_{\mathbf{z},2,t}. \tag{C.13}$$

First, we provide the exact expression for the cumulative cost. It will be used later to analyze the regret.

**Lemma C.3 (Cumulative Cost Expression)** *For the cost $J_T(\mathsf{x}_0, \omega(0), \{\mathsf{K}_{1:s}, \Sigma_{\mathbf{z}}\})$ defined in (C.1), we have*

$$J_T(\mathsf{x}_0, \omega(0), \{\mathsf{K}_{1:s}, \Sigma_{\mathbf{z}}\}) = S_{0,T} + S_{\mathbf{z},1,T} + S_{\mathbf{z},2,T} + S_{\mathbf{w},T}. \tag{C.14}$$

**Proof** For the expected cost at time $t$, we have

$$\begin{aligned} \mathbb{E}[\mathsf{x}_t^\top \mathsf{Q}_{\omega(t)} \mathsf{x}_t + \mathsf{u}_t^\top \mathsf{R}_{\omega(t)} \mathsf{u}_t] &= \sum_{i=1}^{s} \mathrm{tr}\left(\mathbb{E}[\mathsf{Q}_{\omega(t)} \mathsf{x}_t \mathsf{x}_t^\top \mathbf{1}_{\{\omega(t)=i\}}] + \mathbb{E}[\mathsf{R}_{\omega(t)} \mathsf{u}_t \mathsf{u}_t^\top \mathbf{1}_{\{\omega(t)=i\}}]\right)\\ &= \sum_{i=1}^{s} \mathrm{tr}\left((\mathsf{Q}_i + \mathsf{K}_i^\top \mathsf{R}_i \mathsf{K}_i)\Sigma_i(t) + \quad_t(i)\mathsf{R}_i \Sigma_{\mathbf{z}}\right)\\ &= \sum_{i=1}^{s} \mathrm{tr}\left(\mathsf{M}_i \Sigma_i(t)\right) + N_{\mathbf{z},2,t}, \end{aligned} \tag{C.15}$$

where the second equality follows since $\mathsf{u}_t = \mathsf{K}_{\omega(t)}\mathsf{x}_t + \mathsf{z}_t$. Now plugging in the dynamics of $\Sigma_i(t)$ in Lemma A.3, we can conclude the proof. ∎

Next, before proceeding, we provide several properties regarding the operator $\mathrm{tr}(\mathsf{M}\mathcal{H}(\cdot))$ that shows up in (C.12) and (C.13), which will be used later to evaluate $J_T(\mathsf{x}_0, \omega(0), \{\mathsf{K}_{1:s}, \Sigma_{\mathbf{z}}\}) - TJ(0, \omega(0), \{\mathsf{K}_{1:s}\})$.

**Lemma C.4 (Properties of Cost Building Bricks)** *For any $t, t' \in \mathbb{N}$, we have*

*(L1)* $\mathrm{tr}(\mathsf{M}\mathcal{H}^{-1}(\tilde{\mathsf{L}}^t \mathsf{v})) \le \sqrt{ns}\,\|\mathsf{M}_{1:s}\|\,\|\tilde{\mathsf{L}}^t\|\,\|\mathsf{v}\|$, *where* $\mathsf{v} := [\mathrm{vec}(\mathsf{V}_1)^\top, \ldots, \mathrm{vec}(\mathsf{V}_s)^\top]^\top$ *for some* $\mathsf{V}_{1:s}$ *such that* $\mathsf{V}_i \succeq 0$ *for all* $i \in [s]$;

(L2) $\mathrm{tr}(\mathbf{M}\mathcal{H}^{-1}(\tilde{\mathbf{L}}^t\tilde{\mathbf{B}}_{t'}\mathrm{vec}(\boldsymbol{\Sigma}_{\mathbf{z}}))) \le n\sqrt{s}\|\mathbf{M}_{1:s}\|\|\tilde{\mathbf{L}}^t\|\|\mathbf{B}_{1:s}\|^2\|\boldsymbol{\Sigma}_{\mathbf{z}}\|$;

(L3) $\mathrm{tr}(\mathbf{M}\mathcal{H}^{-1}(\tilde{\mathbf{L}}^t\tilde{\boldsymbol{\Pi}}_{t'}\mathrm{vec}(\boldsymbol{\Sigma}_{\mathbf{w}}))) \le n\sqrt{s}\|\mathbf{M}_{1:s}\|\|\tilde{\mathbf{L}}^t\|\|\boldsymbol{\Sigma}_{\mathbf{w}}\|$;

(L4) $|\mathrm{tr}(\mathbf{M}\mathcal{H}^{-1}(\tilde{\mathbf{L}}^t(\tilde{\boldsymbol{\Pi}}_{t'}-\tilde{\boldsymbol{\Pi}}_{\infty})\mathrm{vec}(\boldsymbol{\Sigma}_{\mathbf{w}})))| \le \tau_{MC}n\sqrt{s}\|\mathbf{M}_{1:s}\|\|\tilde{\mathbf{L}}^t\|\|\boldsymbol{\Sigma}_{\mathbf{w}}\|\rho_{MC}^{t'}$, *where* $\tau_{MC}$ *are* $\rho_{MC}$ *are given in Definition A.2, and* $\tilde{\boldsymbol{\Pi}}_{\infty}$ *is given in* (C.9)

**Proof** Let $[\cdot]_i$ denote the $i$th sub-block of an $s\times 1$ block matrix. Let $\mathrm{vec}^{-1}$ denote the inverse mapping of $\mathrm{vec}$, i.e., $\mathrm{vec}^{-1}([\mathbf{v}_1^\top,\cdots,\mathbf{v}_r^\top]^\top) = [\mathbf{v}_1,\cdots,\mathbf{v}_r]$ for a set of vectors $\{\mathbf{v}_i\}_{i=1}^r$. It can be easily seen that for any set of matrices $\mathbf{A},\mathbf{B},\mathbf{C}$ and $\mathbf{X}$, we have $\mathbf{AXB}=\mathbf{C}$ if and only if $(\mathbf{B}^\top\otimes\mathbf{A})\mathrm{vec}(\mathbf{X}) = \mathrm{vec}(\mathbf{C})$. This together with the definitions of $\tilde{\mathbf{B}}_t$, $\tilde{\boldsymbol{\Pi}}_t$ in (A.4), $\tilde{\boldsymbol{\Pi}}_\infty$, $\tilde{\mathbf{R}}_t$ in (C.9), and $\mathcal{H}()$ in (C.10) yields the following preliminary results

$$[\mathcal{H}^{-1}(\tilde{\mathbf{L}}^t\mathbf{v})]_i \ge 0, \tag{C.16a}$$

$$\mathrm{vec}^{-1}([\tilde{\mathbf{B}}_{t'}\mathrm{vec}(\boldsymbol{\Sigma}_{\mathbf{z}})]_i) \ge 0, \tag{C.16b}$$

$$\mathrm{vec}^{-1}([\tilde{\boldsymbol{\Pi}}_{t'}\mathrm{vec}(\boldsymbol{\Sigma}_{\mathbf{w}})]_i) \ge 0, \tag{C.16c}$$

$$\mathrm{vec}^{-1}([|\tilde{\boldsymbol{\Pi}}_{t'}-\tilde{\boldsymbol{\Pi}}_{\infty}|\mathrm{vec}(\boldsymbol{\Sigma}_{\mathbf{w}})]_i) \ge 0, \tag{C.16d}$$

$$|\mathrm{tr}(\mathbf{M}\mathcal{H}^{-1}(\tilde{\mathbf{L}}^t(\tilde{\boldsymbol{\Pi}}_{t'}-\tilde{\boldsymbol{\Pi}}_{\infty})\mathrm{vec}(\boldsymbol{\Sigma}_{\mathbf{w}})))| \le \mathrm{tr}(\mathbf{M}\mathcal{H}^{-1}(\tilde{\mathbf{L}}^t|\tilde{\boldsymbol{\Pi}}_{t'}-\tilde{\boldsymbol{\Pi}}_{\infty}|\mathrm{vec}(\boldsymbol{\Sigma}_{\mathbf{w}}))), \tag{C.16e}$$

where $|\cdot|$ here denotes the element-wise absolute value of a matrix. Now, let us consider (L1). We observe that

$$\begin{aligned}\mathrm{tr}(\mathbf{M}\mathcal{H}^{-1}(\tilde{\mathbf{L}}^t\mathbf{v})) = \mathrm{tr}(\sum_{i=1}^{s}\mathbf{M}_i[\mathcal{H}^{-1}(\tilde{\mathbf{L}}^t\mathbf{v})]_i) &\le \|\mathbf{M}_{1:s}\|\cdot\mathrm{tr}(\sum_{i=1}^{s}[\mathcal{H}^{-1}(\tilde{\mathbf{L}}^t\mathbf{v})]_i)\\ &\le \sqrt{n}\|\mathbf{M}_{1:s}\|\|\sum_{i=1}^{s}[\mathcal{H}^{-1}(\tilde{\mathbf{L}}^t\mathbf{v})]_i\|_F,\end{aligned} \tag{C.17}$$

where the first inequality uses (C.16a) and the definition that $\|\mathbf{M}_{1:s}\| = \max_{i\in[s]}\|\mathbf{M}_i\|$; and the last inequality follows from Cauchy-Schwarz inequality and the fact that $[\mathcal{H}^{-1}(\tilde{\mathbf{L}}^t\mathbf{v})]_i \in \mathbb{R}^{n\times n}$. Now, for the last term on the R.H.S. of (C.17), we have

$$\begin{aligned}\|\sum_{i=1}^{s}[\mathcal{H}^{-1}(\tilde{\mathbf{L}}^t\mathbf{v})]_i\|_F \le \sum_{i=1}^{s}\|[\mathcal{H}^{-1}(\tilde{\mathbf{L}}^t\mathbf{v})]_i\|_F &\le \sqrt{s}\sqrt{\sum_{i=1}^{s}\|[\mathcal{H}^{-1}(\tilde{\mathbf{L}}^t\mathbf{v})]_i\|_F^2}\\ &= \sqrt{s}\|\mathcal{H}^{-1}(\tilde{\mathbf{L}}^t\mathbf{v})\|_F\\ &= \sqrt{s}\|\tilde{\mathbf{L}}^t\mathbf{v}\|\\ &\le \sqrt{s}\|\tilde{\mathbf{L}}^t\|\|\mathbf{v}\|,\end{aligned} \tag{C.18}$$

where the second equality holds since $\mathcal{H}^{-1}$ is a reshaping operator, and $\tilde{\mathbf{L}}^t\mathbf{v}$ is a vector. Substituting (C.18) into (C.17) gives (L1).

To show (L2), we combine (C.16b) with (L1) to get $\mathrm{tr}(\mathbf{M}\mathcal{H}^{-1}(\tilde{\mathbf{L}}^t\tilde{\mathbf{B}}_{t'}\mathrm{vec}(\boldsymbol{\Sigma}_{\mathbf{z}}))) \le \sqrt{ns}\|\mathbf{M}_{1:s}\|\|\tilde{\mathbf{L}}^t\|\|\tilde{\mathbf{B}}_{t'}\mathrm{vec}(\boldsymbol{\Sigma}_{\mathbf{z}})\|$. Then, using the upper bound for $\|\tilde{\mathbf{B}}_{t'}\mathrm{vec}(\boldsymbol{\Sigma}_{\mathbf{z}})\|$ derived in (A.12) completes the proof of (L2).

To establish (L3), we combine (C.16c) with (L1) to obtain

$$\mathrm{tr}(\mathbf{M}\mathcal{H}^{-1}(\tilde{\mathbf{L}}^t\tilde{\boldsymbol{\Pi}}_{t'}\mathrm{vec}(\boldsymbol{\Sigma}_{\mathbf{w}}))) \le \sqrt{ns}\|\mathbf{M}_{1:s}\|\|\tilde{\mathbf{L}}^t\|\|\tilde{\boldsymbol{\Pi}}_{t'}\mathrm{vec}(\boldsymbol{\Sigma}_{\mathbf{w}})\|. \tag{C.19}$$

Then, using the upper bound for $\|\tilde{\boldsymbol{\Pi}}_{t'}\mathrm{vec}(\boldsymbol{\Sigma}_{\mathbf{w}})\|$ derived in (A.13) gives (L2).

Finally, let us consider (L4). It follows from (C.16e) and (C.16d) in conjunction with (L1) that

$$|\mathrm{tr}(\mathbf{M}\mathcal{H}^{-1}(\tilde{\mathbf{L}}^t|\tilde{\boldsymbol{\Pi}}_{t'}-\tilde{\boldsymbol{\Pi}}_{\infty}|\mathrm{vec}(\boldsymbol{\Sigma}_{\mathbf{w}})))| \le \sqrt{ns}\|\mathbf{M}_{1:s}\|\|\tilde{\mathbf{L}}^t\|\||\tilde{\boldsymbol{\Pi}}_{t'}-\tilde{\boldsymbol{\Pi}}_{\infty}|\mathrm{vec}(\boldsymbol{\Sigma}_{\mathbf{w}})\|. \tag{C.20}$$

Now, using (C.9), we obtain

$$\begin{aligned}\|\|\tilde{\Pi}_{t'} - \tilde{\Pi}_\infty|\mathrm{vec}(\Sigma_{\mathbf{w}})\| &= \sqrt{\sum_{i=1}^{s} \|\|[\tilde{\Pi}_{t'}]_i - [\tilde{\Pi}_\infty]_i|\mathrm{vec}(\Sigma_{\mathbf{w}})\|^2} \\ &= \sqrt{\sum_{i=1}^{s} \|\| \quad_{t'}(i) - \quad_\infty(i)|\mathrm{vec}(\Sigma_{\mathbf{w}})\|^2} \\ &= \| \quad_{t'} - \quad_\infty \| \|\mathrm{vec}(\Sigma_{\mathbf{w}})\| \\ &\le \| \quad_{t'} - \quad_\infty \|_1 \|\Sigma_{\mathbf{w}}\|_F \\ &\le \tau_{MC}\sqrt{n}\|\Sigma_{\mathbf{w}}\|\rho_{MC}^{t'},\end{aligned}$$

where the last line follows from Definition A.2. Substituting the above inequality in (C.20) completes the proof of (L4). ■

The following lemma provides a bound for the difference $J_T(x_0, \omega(0), \{K_{1:s}, \Sigma_z\}) - TJ(0, \omega(0), \{K_{1:s}\})$ using an arbitrary stabilizing controller $K_{1:s}$. Based on this result, we will provide in Proposition C.6 a uniform upper bound for this difference when using any controllers $K_{1:s}$ that are close to $K_{1:s}^\star$.

**Lemma C.5** *For an arbitrary stabilizing controller $K_{1:s}$, we have*

$$\begin{aligned}&J_T(x_0, \omega(0), \{K_{1:s}, \Sigma_z\}) - TJ(0, \omega(0), \{K_{1:s}\}) \\ \le&\sqrt{ns}\|M_{1:s}\| \cdot \|x_0\|^2 + \frac{n\sqrt{s}\tau_{\tilde{L}}}{1-\rho_{\tilde{L}}}\|M_{1:s}\|\|B_{1:s}\|^2\|\Sigma_z\|T \\ &+ n\|R_{1:s}\|\|\Sigma_z\|T + n\sqrt{s}\tau_{MC}\tau_{\tilde{L}}\|M_{1:s}\|\|\Sigma_w\|\frac{\rho_{MC}}{\rho_{MC}-\rho_{\tilde{L}}}\left(\frac{\rho_{MC}}{1-\rho_{MC}} - \frac{\rho_{\tilde{L}}}{1-\rho_{\tilde{L}}}\right),\end{aligned} \tag{C.21}$$

*where $\tau_{MC}$ and $\rho_{MC}$ are given in Definition A.2, $\tau_{\tilde{L}}$ and $\rho_{\tilde{L}}$ are constants defined in the beginning of Appendix C, and $M = [M_1, \ldots, M_s]$ with $M_i = Q_i + K_i^\star R_i K_i$.*

**Proof** From Lemma C.3, we know

$$\begin{aligned}J_T(x_0, \omega(0), \{K_{1:s}, \Sigma_z\}) &= S_{0,T} + S_{z,1,T} + S_{z,2,T} + S_{w,T}, \\ J(0, \omega(0), \{K_{1:s}\}) &= \limsup_{T\to\infty} \frac{1}{T}(S_{0,T} + S_{w,T}) =: S_0 + S_w.\end{aligned}$$

where $S_0 := \limsup_{T\to\infty} \frac{1}{T} S_{0,T}$ and $S_w := \limsup_{T\to\infty} \frac{1}{T} S_{w,T}$. Next, we will evaluate each term on the RHSs separately.

For $S_{0,T}$, letting $s_0 = \begin{bmatrix}\mathrm{vec}(\Sigma_1(0)) \\ \vdots \\ \mathrm{vec}(\Sigma_s(0))\end{bmatrix}$, we have

$$\begin{aligned}S_{0,T} = \sum_{t=1}^{T} \mathrm{tr}(M\mathcal{H}^{-1}(\tilde{L}^t s_0)) &\le \sqrt{ns}\|M_{1:s}\|\|\tilde{L}^t\|\|s_0\| \\ &\le \sqrt{ns}\|M_{1:s}\| \cdot \mathsf{E}[\|x_0\|^2] \\ &= \sqrt{ns}\|M_{1:s}\| \cdot \|x_0\|^2,\end{aligned}$$

where the second line follows from Item (L1) in Lemma C.4; the third line follows from (A.11) in Lemma A.4. And from the discussion at the beginning of Appendix C, we can get rid of $\mathsf{E}[\cdot]$. Then it is easy to see $S_0 = 0$, as long as $\|x_0\|^2$ is bounded.

For $S_{\mathbf{z},1,T}$, we have

$$\begin{aligned} S_{\mathbf{z},1,T} &= \sum_{t=1}^{T}\sum_{t'=0}^{t-1} \mathrm{tr}(\mathsf{M}\mathcal{H}^{-1}(\tilde{\mathsf{L}}^{t'}\tilde{\mathsf{B}}_{t-t'}\mathrm{vec}(\Sigma_{\mathbf{z}}))) \\ &\le n\sqrt{s}\|\mathsf{M}_{1:s}\|\|\mathsf{B}_{1:s}\|^2\|\Sigma_{\mathbf{z}}\|(\sum_{t=1}^{T}\sum_{t'=0}^{t-1}\|\tilde{\mathsf{L}}^{t'}\|) \\ &\le \frac{n\sqrt{s}\tau_{\tilde{\mathsf{L}}}}{1-\rho_{\tilde{\mathsf{L}}}}\|\mathsf{M}_{1:s}\|\|\mathsf{B}_{1:s}\|^2\|\Sigma_{\mathbf{z}}\|T, \end{aligned} \tag{C.22}$$

where the first inequality follows from Item (L2) in Lemma C.4, and the second inequality follows from the fact $\|\tilde{\mathsf{L}}^{t'}\| \le \tau_{\tilde{\mathsf{L}}}\rho_{\tilde{\mathsf{L}}}^{t'}$

For $S_{\mathbf{z},2,T}$, we have

$$S_{\mathbf{z},2,T} = \sum_{t=1}^{T}\mathrm{tr}(\sum_{i=1}^{s} {}_t(i)\mathsf{R}_i\Sigma_{\mathbf{z}}) \le n\|\mathsf{R}_{1:s}\|\|\Sigma_{\mathbf{z}}\|T. \tag{C.23}$$

For $S_{\mathbf{w},T}$, we have

$$S_{\mathbf{w},T} = \sum_{t=1}^{T}\sum_{t'=0}^{t-1}\mathrm{tr}(\mathsf{M}\mathcal{H}^{-1}(\tilde{\mathsf{L}}^{t'}\tilde{\Pi}_{t-t'}\mathrm{vec}(\Sigma_{\mathbf{w}}))). \tag{C.24}$$

To evaluate it, we first define the following terms:

$$S_{\mathbf{w},T}^{(\infty)} := \sum_{t=1}^{T}\sum_{t'=0}^{t-1}\mathrm{tr}(\mathsf{M}\mathcal{H}^{-1}(\tilde{\mathsf{L}}^{t'}\tilde{\Pi}_{\infty}\mathrm{vec}(\Sigma_{\mathbf{w}}))), \tag{C.25}$$

$$S_{\mathbf{w}}^{(\infty)} := \limsup_{T\to\infty}\frac{1}{T}S_{\mathbf{w},T}^{(\infty)}, \tag{C.26}$$

where $\tilde{\Pi}_{\infty}$ is defined in (C.9). Note that $S_{\mathbf{w},T}^{(\infty)}$ and $S_{\mathbf{w}}^{(\infty)}$ are the counterparts of $S_{\mathbf{w},T}$ and $S_{\mathbf{w}}$ except that the initial mode distribution ${}_0$ is the stationary distribution ${}_{\infty}$.

Then, we have

$$\begin{aligned} |S_{\mathbf{w},T} - S_{\mathbf{w},T}^{(\infty)}| &= \left|\sum_{t=1}^{T}\sum_{t'=0}^{t-1}\mathrm{tr}(\mathsf{M}\mathcal{H}^{-1}(\tilde{\mathsf{L}}^{t'}(\tilde{\Pi}_{t-t'}-\tilde{\Pi}_{\infty})\mathrm{vec}(\Sigma_{\mathbf{w}})))\right| \\ &\le \tau_{MC}n\sqrt{s}\|\mathsf{M}_{1:s}\|\|\Sigma_{\mathbf{w}}\|(\sum_{t=1}^{T}\sum_{t'=0}^{t-1}\|\tilde{\mathsf{L}}^{t'}\|\rho_{MC}^{t-t'}) \\ &\le \tau_{MC}n\sqrt{s}\|\mathsf{M}_{1:s}\|\|\Sigma_{\mathbf{w}}\|(\sum_{t=1}^{\infty}\sum_{t'=0}^{t-1}\tau_{\tilde{\mathsf{L}}}\rho_{\tilde{\mathsf{L}}}^{t'}\rho_{MC}^{t-t'}) \\ &\le n\sqrt{s}\tau_{MC}\tau_{\tilde{\mathsf{L}}}\|\mathsf{M}_{1:s}\|\|\Sigma_{\mathbf{w}}\|\frac{\rho_{MC}}{\rho_{MC}-\rho_{\tilde{\mathsf{L}}}}(\frac{\rho_{MC}}{1-\rho_{MC}}-\frac{\rho_{\tilde{\mathsf{L}}}}{1-\rho_{\tilde{\mathsf{L}}}}) \end{aligned} \tag{C.27}$$

where the first inequality follows from Item (L4) in Lemma C.4. Thus,

$$S_{\mathbf{w}} = \limsup_{T\to\infty}\frac{1}{T}S_{\mathbf{w},T} = \limsup_{T\to\infty}\frac{1}{T}(S_{\mathbf{w},T}-S_{\mathbf{w},T}^{(\infty)}) + \limsup_{T\to\infty}\frac{1}{T}S_{\mathbf{w},T}^{(\infty)} = S_{\mathbf{w}}^{(\infty)}. \tag{C.28}$$

Since $\sum_{t=1}^{T}\sum_{t'=0}^{t-1}\tilde{\mathsf{L}}^{t'} = (\mathsf{I}-\tilde{\mathsf{L}})^{-1}T - (\mathsf{I}-\tilde{\mathsf{L}})^{-2}\tilde{\mathsf{L}}(\mathsf{I}-\tilde{\mathsf{L}}^{T})$ and $\sum_{t'=0}^{\infty}\tilde{\mathsf{L}}^{t'} = (\mathsf{I}-\tilde{\mathsf{L}})^{-1}$ we have

$$\begin{aligned} S_{\mathbf{w}} = S_{\mathbf{w}}^{(\infty)} &= \mathrm{tr}(\mathsf{M}\mathcal{H}^{-1}(\limsup_{T\to\infty}\frac{1}{T}\sum_{t=1}^{T}\sum_{t'=0}^{t-1}\tilde{\mathsf{L}}^{t'}\tilde{\Pi}_{\infty}\mathrm{vec}(\Sigma_{\mathbf{w}}))) \\ &= \mathrm{tr}(\mathsf{M}\mathcal{H}^{-1}((\mathsf{I}-\tilde{\mathsf{L}})^{-1}\tilde{\Pi}_{\infty}\mathrm{vec}(\Sigma_{\mathbf{w}}))) \\ &= \sum_{t'=0}^{\infty}\mathrm{tr}(\mathsf{M}\mathcal{H}^{-1}(\tilde{\mathsf{L}}^{t'}\tilde{\Pi}_{\infty}\mathrm{vec}(\Sigma_{\mathbf{w}}))). \end{aligned} \tag{C.29}$$

Thus,

$$\begin{aligned} TS_{\mathbf{w}} = TS_{\mathbf{w}}^{(\infty)} &= \sum_{t=1}^{T}\sum_{t'=0}^{\infty} \mathrm{tr}(M\mathcal{H}^{-1}(\tilde{L}^{t'}\tilde{\Pi}_{\infty}\mathrm{vec}(\Sigma_{\mathbf{w}}))) \\ &\geq \sum_{t=1}^{T}\sum_{t'=0}^{t-1} \mathrm{tr}(M\mathcal{H}^{-1}(\tilde{L}^{t'}\tilde{\Pi}_{\infty}\mathrm{vec}(\Sigma_{\mathbf{w}}))) \\ &= S_{\mathbf{w},T}^{(\infty)} \end{aligned} \tag{C.30}$$

where the inequality holds since each trace summand is non-negative. Therefore,

$$\begin{aligned} S_{\mathbf{w},T} &\leq S_{\mathbf{w},T}^{(\infty)} + |S_{\mathbf{w},T} - S_{\mathbf{w},T}^{(\infty)}| \\ &\overset{(C.28)}{\leq} TS_{\mathbf{w}} + |S_{\mathbf{w},T} - S_{\mathbf{w},T}^{(\infty)}| \\ &\overset{(C.27)}{\leq} TS_{\mathbf{w}} + n\sqrt{s}\tau_{MC}\tau_{\tilde{L}}\|M_{1:s}\|\|\Sigma_{\mathbf{w}}\|\frac{\rho_{MC}}{\rho_{MC}-\rho_{\tilde{L}}}\left(\frac{\rho_{MC}}{1-\rho_{MC}} - \frac{\rho_{\tilde{L}}}{1-\rho_{\tilde{L}}}\right). \end{aligned} \tag{C.31}$$

Finally, combining all the results we have so far, we have

$$\begin{aligned} &J_T(x_0, \omega(0), \{K_{1:s}, \Sigma_{\mathbf{z}}\}) - TJ(0, \omega(0), \{K_{1:s}\}) \\ =&S_{0,T} + S_{\mathbf{z},1,T} + S_{\mathbf{z},2,T} + S_{\mathbf{w},T} - T(S_0 + S_{\mathbf{w}}) \\ \leq&\sqrt{ns}\|M_{1:s}\| \cdot \|x_0\|^2 \\ &+ \frac{n\sqrt{s}\tau_{\tilde{L}}}{1-\rho_{\tilde{L}}}\|M_{1:s}\|\|B_{1:s}\|^2\|\Sigma_{\mathbf{z}}\|T \\ &+ n\|R_{1:s}\|\|\Sigma_{\mathbf{z}}\|T \\ &+ n\sqrt{s}\tau_{MC}\tau_{\tilde{L}}\|M_{1:s}\|\|\Sigma_{\mathbf{w}}\|\frac{\rho_{MC}}{\rho_{MC}-\rho_{\tilde{L}}}\left(\frac{\rho_{MC}}{1-\rho_{MC}} - \frac{\rho_{\tilde{L}}}{1-\rho_{\tilde{L}}}\right) \end{aligned} \tag{C.32}$$

which concludes the proof. ■

We now provide a uniform upper bound on the regret $J_T(x_0, \omega(0), \{K_{1:s}, \Sigma_{\mathbf{z}}\}) - TJ^{\square}$ for any stabilizing controller $K_{1:s}$ that is close enough to the optimal controller $K_{1:s}^{\square}$.

**Proposition C.6** *For every $K_{1:s}$ such that $\|K_{1:s} - K_{1:s}^{\square}\| \leq \bar{\ }_{\mathbf{K}}$, we have*

$$\begin{aligned} J_T(x_0, \omega(0), \{K_{1:s}, \Sigma_{\mathbf{z}}\}) - TJ^{\square} \leq& C_{\mathbf{K}}^{J}\|K_{1:s} - K_{1:s}^{\square}\|^2\|\Sigma_{\mathbf{w}}\|T \\ &+\sqrt{ns}M\|x_0\|^2 \\ &+n\sqrt{s}\frac{2\tau(\tilde{L}^{\square})\|B_{1:s}\|^2 M}{1-\rho^{\square}}\|\Sigma_{\mathbf{z}}\|T \\ &+n\|R_{1:s}\|\|\Sigma_{\mathbf{z}}\|T \\ &+n\sqrt{s}\frac{2\tau(\tilde{L}^{\square})\tau_{MC}M\rho_{MC}}{2\rho_{MC}-1-\rho^{\square}}\left(\frac{\rho_{MC}}{1-\rho_{MC}} - \frac{1+\rho^{\square}}{1-\rho^{\square}}\right)\|\Sigma_{\mathbf{w}}\|, \end{aligned} \tag{C.33}$$

*where $M := \|Q_{1:s}\| + 4\|R_{1:s}\|\|K_{1:s}^{\square}\|^2$, and $\bar{\ }_{\mathbf{K}}$ and $C_{\mathbf{K}}^{J}$ are defined in Table 3.*

**Proof** When $\|K_{1:s} - K_{1:s}^{\square}\| \leq \bar{\ }_{\mathbf{K}}$, from Lemma C.1, we know $\|\tilde{L}^k\| \leq \tau(\tilde{L}^{\square})(\frac{1+\rho^{\star}}{2})^k$, thus we could set $\tau_{\tilde{L}}$ and $\rho_{\tilde{L}}$ to be $\tau(\tilde{L}^{\square})$ and $\frac{1+\rho^{\star}}{2}$. By definition, we know $\bar{\ }_{\mathbf{K}} \leq \|K_{1:s}^{\square}\|$, thus $\|M_{1:s}\| \leq \|Q_{1:s}\| + \|R_{1:s}\|\|K_{1:s}\|^2 \leq$

$\|Q_{1:s}\| + \|R_{1:s}\|(\|K^\star_{1:s}\| + \bar{\kappa})^2 \le \|Q_{1:s}\| + 4\|R_{1:s}\|\|K^\star_{1:s}\|^2 = M$. Then applying Lemma C.5, we have

$$\begin{aligned}
&J_T(x_0, \omega(0), \{K_{1:s}, \Sigma_z\}) - TJ(0, \omega(0), \{K_{1:s}\}) \\
\le& \sqrt{ns} M \|x_0\|^2 \\
&+ n\sqrt{s}\frac{2\tau(\tilde{L}^\star)\|B_{1:s}\|^2 M}{1-\rho^\star}\|\Sigma_z\| T \\
&+ n\|R_{1:s}\|\|\Sigma_z\| T \\
&+ n\sqrt{s}\frac{2\tau(\tilde{L}^\star)\tau_{MC} M \rho_{MC}}{2\rho_{MC} - 1 - \rho^\star}\left(\frac{\rho_{MC}}{1-\rho_{MC}} - \frac{1+\rho^\star}{1-\rho^\star}\right)\|\Sigma_w\|
\end{aligned} \tag{C.34}$$

Now note that when $\|K_{1:s} - K^\star_{1:s}\| \le \bar{\kappa}$, we have $J(0, \omega(0), \{K_{1:s}\}) - J^\star \le C^J_K \|\Sigma_w\|\|K_{1:s} - K^\star_{1:s}\|^2$ using Lemma C.2. Combining this with (C.34), we could conclude the proof. ∎

## C.3 Stitching Every Epoch

In this section, we stitch the upper bounds on $\text{Regret}_i$ for every epoch $i$ and build a bound on the overall regret $\text{Regret}(T)$.

We define the estimation error after epoch $i$ as $\varepsilon^{(i)}_{A,B} = \max\{\|A^{(i)}_{1:s} - A_{1:s}\|, \|B^{(i)}_{1:s} - B_{1:s}\|\}$, $\varepsilon^{(i)}_T = \|T^{(i)} - T\|_\infty$. Furthermore, we also define $\varepsilon^{(i)}_K := \|K^{(i)}_{1:s} - K^\star_{1:s}\|$ where $K^\star_{1:s}$ is the optimal controller for the infinite-horizon MJS-LQR$(A_{1:s}, B_{1:s}, T, Q_{1:s}, R_{1:s})$. We define the following events for every epoch $i$.

$$\begin{aligned}
\mathcal{A}_i &= \left\{\text{Regret}_i \le \mathcal{O}\left(sp\left(\varepsilon^{(i-1)}_{A,B} + \varepsilon^{(i-1)}_T\right)^2 \sigma^2_w T_i + \sqrt{ns}\|x^{(i)}_0\|^2 + n\sqrt{s}\sigma^2_{z,i} T_i + c_A\right)\right\} \\
\mathcal{B}_i &= \left\{\varepsilon^{(i)}_{A,B} \le \bar{\varepsilon}_{A,B,T},\ \varepsilon^{(i)}_T \le \bar{\varepsilon}_{A,B,T},\ \varepsilon^{(i+1)}_K \le \bar{\varepsilon}_K\right\} \\
\mathcal{C}_i &= \left\{\varepsilon^{(i)}_{A,B} \le \mathcal{O}\left(\log(\frac{1}{\delta_{id,i}})\frac{\sigma_{z,i} + \sigma_w}{\sigma_{z,i}\pi_{\min}}\frac{\sqrt{s}(n+p)\log(T_i)}{\sqrt{T_i}}\right),\right. \\
&\qquad \left.\varepsilon^{(i)}_T \le \mathcal{O}\left(\log(\frac{1}{\delta_{id,i}})\frac{1}{\pi_{\min}}\sqrt{\frac{\log(T_i)}{T_i}}\right)\right\} \\
\mathcal{D}_i &= \left\{\|x^{(i+1)}_0\|^2 = \|x^{(i)}_{T_i}\|^2 \le \frac{\bar{x}^2_0}{\delta_{x_0,i}}\right\}.
\end{aligned} \tag{C.35}$$

where $c_A, \bar{x}_0$ are constants, $\bar{\varepsilon}_{A,B,T}$ and $\bar{\varepsilon}_K$ are defined in Table 3, and $\delta_{id,i}$ and $\delta_{x_0,i}$ within $[0,1]$ denotes the failure probability for event $\mathcal{C}_i$ and $\mathcal{D}_i$. Note that $\mathcal{O}(\cdot)$ hides terms that are invariant to epochs such as $\rho^\star, \|A_{1:s}\|, \|B_{1:s}\|$, etc.

Event $\mathcal{A}_i$ describes how epoch $i$ regret depends on initial state $\|x^{(i)}_0\|^2$, exploration noise variance $\sigma^2_{z,i}$, and the accuracy of the estimated MJS dynamics $A^{(i-1)}_{1:s}, B^{(i-1)}_{1:s}, \hat{T}$ after epoch $i-1$, which is used to computed epoch $i$ controller $K^{(i)}_{1:s}$. Event $\mathcal{B}_i$ indicates whether the estimated dynamics and resulting controllers are good enough. $\mathcal{C}_i$ describes the dynamics estimation error after epoch $i$, and when epoch $T_i$ is chosen appropriately, $\mathcal{B}_i$ can be implied. Lastly, event $\mathcal{D}_i$ bounds the initial state of each epoch, as the initial state plays a vital role in regret upper bound $\mathcal{A}_i$. We see events $\mathcal{A}_{i+1}, \mathcal{B}_i, \mathcal{C}_i, \mathcal{D}_i$ are $\mathcal{F}_i$-measurable, i.e. these events can be determined using random variables $x_0, w_t, z_t, \omega(t)$ up to epoch $i$. Let $\mathcal{E}_i := \mathcal{A}_{i+1} \cap \mathcal{B}_i \cap \mathcal{C}_i \cap \mathcal{D}_i$. Note that even though $\mathcal{A}_{i+1}$ is for the conditional expected regret of the epoch $i+1$ with randomness coming from $x^{(i+1)}_0 = x^{(i)}_{T_i}, \omega^{(i+1)}(0) = \omega^{(i)}(T_i)$, and controller $K^{(i+1)}_{1:s}$ computed from $A^{(i)}_{1:s}, B^{(i)}_{1:s}, T^{(i)}$, thus $\mathcal{A}_{i+1}$ is $\mathcal{F}_i$-measurable.

Then, we have the following results regarding the conditional probabilities of these events. First, Proposition C.7 says given the event $\mathcal{B}_{i-1}$ (a good controller is applied during epoch $i$) and event $\mathcal{D}_i$ (the initial state of epoch $i$, $x^{(i)}_0$ is bounded), then $\mathcal{D}_i$ could occur, i.e. $x^{(i)}_{T_i}$, the final state of epoch $i$, a.k.a. $x^{(i+1)}_0$ the initial state of epoch $i+1$, is also bounded.

**Proposition C.7** *Suppose* $\frac{\sqrt{n\bar{s}}\bar{\tau}\bar{\rho}^{T_i}}{\delta_{\mathbf{x}_0,i-1}} < 1$ *and* $\bar{x}_0^2 \geq \frac{n\sqrt{\bar{s}}(\|\mathbf{B}_{1:s}\|^2+1)\sigma_{\mathbf{w}}^2\bar{\tau}}{(1-\bar{\rho})(1-\sqrt{n\bar{s}}\bar{\tau}\bar{\rho}^{T_i}/\delta_{\mathbf{x}_0,i-1})}$ for $i \geq 1$. *Then,*

$$\mathsf{P}(\mathcal{D}_i \mid \cap_{j=0}^{i-1}\mathcal{E}_j) = \mathsf{P}(\mathcal{D}_i \mid \mathcal{B}_{i-1}\mathcal{D}_{i-1}) > 1-\delta_{\mathbf{x}_0,i},$$

*and* $\mathsf{P}(\mathcal{D}_0) \geq 1-\delta_{\mathbf{x}_0,0}$.

**Proof** For epoch $i = 1, 2, \ldots$, given event $\mathcal{B}_{i-1}$, we know $\overset{(i+1)}{\mathbf{K}} \leq \bar{\mathbf{K}}$. Let $\tilde{\mathbf{L}}^{(i)}$ denote the augmented closed-loop state matrix. By Lemma C.1, we know $\|(\tilde{\mathbf{L}}^{(i)})^k\| \leq \tau(\tilde{\mathbf{L}}^{(i)})(\frac{1+\rho^\star}{2})^k$. Thus, if we pick $\bar{\tau} := \max\{\tau(\tilde{\mathbf{L}}^{(0)}), \tau(\tilde{\mathbf{L}}^{(1)})\}, \bar{\rho} := \max\{\rho(\tilde{\mathbf{L}}^{(0)}), \frac{1+\rho^\star}{2}\}$, this can be generalized to $i = 0$ case, i.e. for every epoch $i = 0, 1, 2, \ldots$, we have $\|(\tilde{\mathbf{L}}^{(i)})^k\| \leq \bar{\tau}\bar{\rho}^k$.

For $i = 1, 2, \ldots$, event $\mathcal{D}_{i-1}$ implies $\|\mathbf{x}_0^{(i)}\|^2 \leq \frac{\bar{x}_0^2}{\delta_{\mathbf{x}_0,i-1}}$. Then, according to Lemma A.4, we know

$$\begin{aligned}
\mathsf{E}[\|\mathbf{x}_{T_i}^{(i)}\|^2 \mid \mathcal{B}_{i-1}, \mathcal{D}_{i-1}] &\leq \sqrt{n\bar{s}} \cdot \bar{\tau}\bar{\rho}^{T_i}\frac{\bar{x}_0^2}{\delta_{\mathbf{x}_0,i-1}} + n\sqrt{\bar{s}}(\|\mathbf{B}_{1:s}\|^2\frac{\sigma_{\mathbf{w}}^2}{\sqrt{T_i}} + \sigma_{\mathbf{w}}^2)\frac{\bar{\tau}}{1-\bar{\rho}}. \\
&\leq \frac{\sqrt{n\bar{s}} \cdot \bar{\tau}\bar{\rho}^{T_i}}{\delta_{\mathbf{x}_0,i-1}}\bar{x}_0^2 + (1 - \frac{\sqrt{n\bar{s}}\bar{\tau}\bar{\rho}^{T_i}}{\delta_{\mathbf{x}_0,i-1}})\bar{x}_0^2 \qquad \text{(C.36)} \\
&\leq \bar{x}_0^2,
\end{aligned}$$

where the second line follows from the assumptions in the proposition statement. Using Markov inequality, we have

$$\mathsf{P}(\|\mathbf{x}_{T_i}^{(i)}\|^2 \leq \frac{\bar{x}_0^2}{\delta_{\mathbf{x}_0,i}} \mid \mathcal{B}_{i-1}, \mathcal{D}_{i-1}) \geq 1-\delta_{\mathbf{x}_0,i},$$

which implies $\mathsf{P}(\mathcal{D}_i \mid \mathcal{B}_{i-1}, \mathcal{D}_{i-1}) \geq 1-\delta_{\mathbf{x}_0,i}$. For $i = 0$, similarly, we have $\mathsf{E}[\|\mathbf{x}_{T_0}^{(0)}\|^2] \leq n\sqrt{\bar{s}}(\|\mathbf{B}_{1:s}\|^2\frac{\sigma_{\mathbf{w}}^2}{\sqrt{T_i}} + \sigma_{\mathbf{w}}^2)\frac{\bar{\tau}}{1-\bar{\rho}} \leq \bar{x}_0^2$, thus $\mathsf{P}(\mathcal{D}_0) \geq 1-\delta_{\mathbf{x}_0,i}$.

Finally, note that given a good stabilizing controller (event $\mathcal{B}_{i-1}$) and a bounded initial state (event $\mathcal{D}_{i-1}$) for epoch $i$, the final state of epoch $i$ only depends on randomness in epoch $i$, thus $\mathsf{P}(\mathcal{D}_i \mid \cap_{j=0}^{i-1}\mathcal{E}_j) = \mathsf{P}(\mathcal{D}_i \mid \mathcal{B}_{i-1}\mathcal{D}_{i-1})$. ■

Proposition C.8 describes that given the event $\mathcal{C}_i$ (the estimated MJS dynamics after epoch $i$ has estimation errors decays with $T_i$), when epoch $i$ has length $T_i$ large enough, then the event $\mathcal{B}_i$ (the estimated dynamics and controllers computed with it will be good enough) occurs.

**Proposition C.8** *Suppose every epoch* $i$ *has length* $T_i \geq \underline{T}_{rgt,\bar{}}(\delta_{id,i}, T_i)$. *Then,*

$$\mathsf{P}(\mathcal{B}_i \mid \mathcal{C}_i, \cap_{j=0}^{i-1}\mathcal{E}_j) = \mathsf{P}(\mathcal{B}_i \mid \mathcal{C}_i) = 1$$

**Proof** When $\mathcal{C}_i$ occurs, since $\sigma_{\mathbf{z},i}^2 = \frac{\sigma_{\mathbf{w},i}^2}{\sqrt{T_i}}$, we have

$$\overset{(i)}{\mathbf{A},\mathbf{B}} \leq \mathcal{O}(\log(\frac{1}{\delta_{id,i}})\frac{\sqrt{\bar{s}}(n+p)}{\pi_{\min}}\frac{\log(T_i)}{T_i^{0.25}}), \quad \overset{(i)}{\mathbf{T}} \leq \mathcal{O}(\log(\frac{1}{\delta_{id,i}})\frac{1}{\pi_{\min}}\sqrt{\frac{\log(T_i)}{T_i}}).$$

We know when $T_i \geq \mathcal{O}(\frac{\sqrt{\bar{s}}(n+p)}{\pi_{\min}}\bar{}^4_{\mathbf{A},\mathbf{B},\mathbf{T}}\log(\frac{1}{\delta_{id,i}})\log^4(T_i)) =: \underline{T}_{rgt,\bar{}}(\delta_{id,i}, T_i)$, we have $\overset{(i)}{\mathbf{A},\mathbf{B}} \leq \bar{}_{\mathbf{A},\mathbf{B},\mathbf{T}}$, $\overset{(i)}{\mathbf{T}} \leq \bar{}_{\mathbf{A},\mathbf{B},\mathbf{T}}$. Then according to Lemma C.2, we have $\overset{(i+1)}{\mathbf{K}} \leq \bar{}_{\mathbf{K}}$. Thus $\mathsf{P}(\mathcal{B}_i \mid \mathcal{C}_i) = 1$. Finally, note that given the estimation error sample complexity in $\mathcal{C}_i$ for epoch $i$, events happen before epoch $i$ does not influence $\mathcal{B}_i$, so $\mathsf{P}(\mathcal{B}_i \mid \mathcal{C}_i, \cap_{j=0}^{i-1}\mathcal{E}_j) = \mathsf{P}(\mathcal{B}_i \mid \mathcal{C}_i)$=1. ■

Next, Proposition C.9 says given the $\mathcal{B}_{i-1}$ (a good controller is used in epoch $i$), then the event $\mathcal{C}_i$ could occur, i.e. dynamics learned using the trajectory of epoch $i$, will be accurate enough.

**Proposition C.9** *For $c_{\mathbf{x}} \geq \underline{c}_{\mathbf{x}}(\bar{\rho}, \bar{\tau}), c_{\mathbf{z}} \geq \underline{c}_{\mathbf{z}}$, $T_i \geq \max\{\underline{T}_{MC,1}(\frac{\delta_{id,i}}{8}), \underline{T}_{id,N}(\frac{\delta_{id,i}}{2}, \bar{\rho}, \bar{\tau})\}$, we have for $i = 1, 2, \ldots,$*

$$\mathsf{P}(\mathcal{C}_i \mid \cap_{j=0}^{i-1}\mathcal{E}_j) = \mathsf{P}(\mathcal{C}_i \mid \mathcal{B}_{i-1}) \geq 1 - \delta_{id,i}. \tag{C.37}$$

*And $\mathsf{P}(\mathcal{C}_0) \geq 1 - \delta_{id,0}$.*

**Proof** By Lemma B.1, we know for every epoch $i = 0, 1, \ldots,$ when $T_i \geq \underline{T}_{MC,1}(\frac{\delta_{id,i}}{8})$, we have with probability at least $1 - \frac{\delta_{id,i}}{2}$, $\varepsilon_{\mathbf{T}}^{(i)} \leq \mathcal{O}\left(\log(\frac{1}{\delta_{id,i}})\frac{1}{\pi_{\min}}\sqrt{\frac{\log(T_i)}{T_i}}\right)$.

For epoch $i = 1, 2, \ldots,$ given event $\mathcal{B}_{i-1}$, we know $\varepsilon_{\mathbf{K}}^{(i)} \leq \bar{\varepsilon}_{\mathbf{K}}$. Let $\tilde{\mathsf{L}}^{(i)}$ denote the augmented closed-loop state matrix. By Lemma C.1, we know $\|(\tilde{\mathsf{L}}^{(i)})^k\| \leq \tau(\tilde{\mathsf{L}}^{\star})(\frac{1+\rho^{\star}}{2})^k$. Thus, if we pick $\bar{\tau} := \max\{\tau(\tilde{\mathsf{L}}^{(0)}), \tau(\tilde{\mathsf{L}}^{\star})\}, \bar{\rho} := \max\{\rho(\tilde{\mathsf{L}}^{(0)}), \frac{1+\rho^{\star}}{2}\}$, this can be generalized to $i = 0$ case, i.e. for every epoch $i = 0, 1, 2, \ldots,$ we have $\|(\tilde{\mathsf{L}}^{(i)})^k\| \leq \bar{\tau}\bar{\rho}^k$.

Suppose $c_{\mathbf{x}} \geq \underline{c}_{\mathbf{x}}(\bar{\rho}, \bar{\tau}), c_{\mathbf{z}} \geq \underline{c}_{\mathbf{z}}$, and $T_i \geq \underline{T}_{id,N}(\frac{\delta_{id,i}}{2}, \bar{\rho}, \bar{\tau})$ hold for $i = 0, 1, \ldots$. Then, from Theorem B.17, we know for every $i = 0, 1, \ldots,$ with probability at least $1 - \frac{\delta_{id,i}}{2}$, $\varepsilon_{\mathbf{A},\mathbf{B}}^{(i)} \leq \mathcal{O}\left(\log(\frac{1}{\delta_{id,i}})\frac{\sigma_{\mathbf{z},i}+\sigma_{\mathbf{w}}}{\sigma_{\mathbf{z},i}\pi_{\min}}\frac{\sqrt{s}(n+p)\log(T_i)}{T_i}\right)$.

Applying union bound to $\varepsilon_{\mathbf{T}}^{(i)}$ and $\varepsilon_{\mathbf{A},\mathbf{B}}^{(i)}$, we could show $\mathsf{P}(\mathcal{C}_0) \geq 1 - \delta_{id,i}$ and $\mathsf{P}(\mathcal{C}_i \mid \mathcal{B}_{i-1}, \mathcal{D}_{i-1}) \geq 1 - \delta_{id,i}$. Finally, note that given a good stabilizing controller (event $\mathcal{B}_{i-1}$) and bounded initial state (event $\mathcal{D}_{i-1}$) for epoch $i$, the estimation error sample complexity (event $\mathcal{C}_i$) does not depend on events happen before epoch $i$, so $\mathsf{P}(\mathcal{C}_i \mid \cap_{j=0}^{i-1}\mathcal{E}_j)) = \mathsf{P}(\mathcal{C}_i \mid \mathcal{B}_{i-1}, \mathcal{D}_{i-1})$. ∎

Finally, Proposition C.10 simply describes how the regret of epoch $i$ depends on the accuracy of the estimated dynamics after epoch $i-1$.

**Proposition C.10** *For $\mathcal{A}_i$– $\mathcal{C}_i$ given in (C.35), we have*

$$\mathsf{P}(\mathcal{A}_i \mid \mathcal{B}_{i-1}, \mathcal{C}_{i-1}, \mathcal{D}_{i-1}, \cap_{j=0}^{i-2}\mathcal{E}_j) = \mathsf{P}(\mathcal{A}_i \mid \mathcal{B}_{i-1}) = 1.$$

**Proof** From Proposition C.6, we know that for every epoch $i = 1, 2, \ldots,$ given $\|\mathsf{K}_{1:s}^{(i)} - \mathsf{K}_{1:s}^{\star}\| \leq \bar{\varepsilon}_{\mathbf{K}}$ in $\mathcal{B}_{i-1}$, we have with probability 1

$$\begin{aligned}
\text{Regret}_i \leq & C_{\mathbf{K}}^{J}\|\mathsf{K}_{1:s}^{(i)} - \mathsf{K}_{1:s}^{\star}\|^2\sigma_{\mathbf{w}}^2 T_i \\
& + \sqrt{ns}M\|\mathsf{x}_0^{(i)}\|^2 \\
& + n\sqrt{s}\frac{2\tau(\tilde{\mathsf{L}}^{\star})\|\mathsf{B}_{1:s}\|^2 M}{1-\rho^{\star}}\sigma_{\mathbf{z},i}^2 T_i \\
& + n\|\mathsf{R}_{1:s}\|\sigma_{\mathbf{z},i}^2 T_i \\
& + n\sqrt{s}\frac{2\tau(\tilde{\mathsf{L}}^{\star})\tau_{MC}M\rho_{MC}}{2\rho_{MC}-1-\rho^{\star}}\left(\frac{\rho_{MC}}{1-\rho_{MC}} - \frac{1+\rho^{\star}}{1-\rho^{\star}}\right)\sigma_{\mathbf{w}}^2.
\end{aligned} \tag{C.38}$$

Let $c_{\mathsf{A}}$ denote the last term in (C.38), which is a constant over epochs. Note that from $\varepsilon_{\mathbf{A},\mathbf{B}}^{(i-1)} \leq \bar{\varepsilon}_{\mathbf{A},\mathbf{B},\mathbf{T}}$, $\varepsilon_{\mathbf{T}}^{(i-1)} \leq \bar{\varepsilon}_{\mathbf{A},\mathbf{B},\mathbf{T}}$ in event $\mathcal{B}_{i-1}$, we know $\|\mathsf{K}_{1:s}^{(i)} - \mathsf{K}_{1:s}^{\star}\| \leq C_{\mathbf{A},\mathbf{B},\mathbf{T}}^{\mathbf{K}}(\varepsilon_{\mathbf{A},\mathbf{B}}^{(i-1)} + \varepsilon_{\mathbf{T}}^{(i-1)})$ by Lemma C.2. Plugging this into (C.38), we have

$$\text{Regret}_i \leq \mathcal{O}\left(s\cdot p\left(\varepsilon_{\mathbf{A},\mathbf{B}}^{(i-1)} + \varepsilon_{\mathbf{T}}^{(i-1)}\right)^2\sigma_{\mathbf{w}}^2 T_i + \sqrt{ns}\|\mathsf{x}_0^{(i)}\|^2 + n\sqrt{s}\sigma_{\mathbf{z},i}^2 T_i + c_{\mathsf{A}}\right) \tag{C.39}$$

where term $s\cdot p$ comes from term $s\min\{n, p\}$ in the definition of $C_{\mathbf{K}}^{J}$ in Appendix C.1. This shows $\mathsf{P}(\mathcal{A}_i \mid \mathcal{B}_{i-1}) = 1$. Finally, note that given a good controller (event $\mathcal{B}_{i-1}$) for epoch $i$, the regret for epoch $i$ can be upper bounded (event $\mathcal{A}_i$) without dependence on other events, thus $\mathsf{P}(\mathcal{A}_i \mid \mathcal{B}_{i-1}, \mathcal{C}_{i-1}, \mathcal{D}_{i-1}, \cap_{j=0}^{i-2}\mathcal{E}_j) = \mathsf{P}(\mathcal{A}_i \mid \mathcal{B}_{i-1})$. ∎

### C.3.1 Proof for Theorem 5.1

**Theorem C.11 (Complete version of Theorem 5.1)** *Assume that the initial state $\mathbf{x}_0 = 0$, and Assumption 1 and 2 hold. Suppose $c_{\mathbf{x}} \geq \underline{c}_{\mathbf{x}}(\bar{\rho}, \bar{\tau})$, $c_{\mathbf{z}} \geq \underline{c}_{\mathbf{z}}$, $T_0 \geq \mathcal{O}(\underline{T}_{rgt}(\delta, T_0))$, and $\bar{x}_0^2 = \frac{n\sqrt{s}(\|\mathbf{B}_{1:s}\|^2+1)\sigma_{\mathbf{w}}^2\bar{\tau}}{(1-\bar{\rho})(1-\sqrt{ns}\bar{\tau}\bar{\rho}^{T_0}\gamma\pi^2/3\delta)}$. Then, with probability at least $1-\delta$, Algorithm 2 achieves*

$$\mathrm{Regret}(T) \leq \mathcal{O}\left(\frac{s^2p(n^2+p^2)\sigma_{\mathbf{w}}^2}{\pi_{\min}^2}\log^2\left(\frac{\log^2(T)}{\delta}\right)\log^2(T)\sqrt{T} + \frac{\sqrt{ns}\log^3(T)}{\delta}\right). \tag{C.40}$$

**Proof** In this proof, we will first show the intersected event $\cap_i \mathcal{E}_i = \cap_i\{\mathcal{A}_{i+1}\cap\mathcal{B}_i\cap\mathcal{C}_i\cap\mathcal{D}_i\}$ implies the desired regret bound, then we evaluate the occurrence probability of $\cap_i\mathcal{E}_i$ using Proposition C.8 to C.10. In the following, we set $\delta_{id,i} = \delta_{\mathbf{x}_0,i} = \frac{3}{\pi^2}\cdot\frac{\delta}{(i+1)^2}$.

With the choices $T_i = \gamma T_{i-1}$, $\sigma_{\mathbf{z},i}^2 = \frac{\sigma_{\mathbf{w}}^2}{\sqrt{T_i}}$, and $\delta_{id,i} = \delta_{\mathbf{x}_0,i} = \frac{3}{\pi^2}\cdot\frac{\delta}{(i+1)^2}$, event $\mathcal{E}_i = \mathcal{A}_{i+1}\cap\mathcal{B}_i\cap\mathcal{C}_i\cap\mathcal{D}_i$ implies the following.

$$\begin{aligned}
&\mathrm{Regret}_{i+1}\\
\leq&\mathcal{O}(1)\log^2\left(\frac{(i+1)^2}{\delta}\right)sp\left(\frac{\sigma_{\mathbf{z},i}+\sigma_{\mathbf{w}}}{\sigma_{\mathbf{z},i}\pi_{\min}}\cdot\frac{\sqrt{s}(n+p)\log(T_i)}{\sqrt{T_i}} + \frac{\sqrt{\log(T_i)}}{\pi_{\min}\sqrt{T_i}}\right)^2\sigma_{\mathbf{w}}^2T_{i+1}\\
&+\mathcal{O}\left(\frac{(i+1)^2}{\delta}\right)\sqrt{ns}\bar{x}_0^2 + \mathcal{O}(n\sqrt{s}\sigma_{\mathbf{z},i+1}^2T_{i+1}) + \mathcal{O}(1)\\
\leq&\mathcal{O}(1)\log^2\left(\frac{(i+1)^2}{\delta}\right)\frac{s^2p(n^2+p^2)\gamma}{\pi_{\min}^2}\frac{(\sigma_{\mathbf{z},i}+\sigma_{\mathbf{w}})^2}{\sigma_{\mathbf{z},i}^2}\sigma_{\mathbf{w}}^2\log^2(T_i)\\
&+\mathcal{O}\left(\frac{(i+1)^2}{\delta}\right)\sqrt{ns}\bar{x}_0^2 + \mathcal{O}(n\sqrt{s}\sigma_{\mathbf{z},i+1}^2T_{i+1})\\
\leq&\mathcal{O}(1)\log^2\left(\frac{(i+1)^2}{\delta}\right)\frac{s^2p(n^2+p^2)\gamma}{\pi_{\min}^2}\left(\frac{\sigma_{\mathbf{w}}^4}{\sigma_{\mathbf{z},i}^2}\log^2(T_i) + \sigma_{\mathbf{z},i+1}^2T_i\right)\\
&+\mathcal{O}\left(\frac{(i+1)^2}{\delta}\right)\sqrt{ns}\bar{x}_0^2\\
\leq&\mathcal{O}(1)\log^2\left(\frac{(i+1)^2}{\delta}\right)\frac{s^2p(n^2+p^2)\gamma}{\pi_{\min}^2}\sigma_{\mathbf{w}}^2\sqrt{T_i}\log^2(T_i) + \mathcal{O}\left(\frac{(i+1)^2}{\delta}\right)\sqrt{ns}\bar{x}_0^2
\end{aligned} \tag{C.41}$$

We have $M := \mathcal{O}(\log_\gamma(\frac{T}{T_0}))$ epochs at time $T$. Using the fact $T_i = \mathcal{O}(T_0\gamma^i)$, event $\cap_{i=0}^{M-1}\mathcal{E}_i$ implies

$$\begin{aligned}
&\mathrm{Regret}(T)\\
=&\mathcal{O}(\sum_{i=1}^{M}\mathrm{Regret}_i)\\
\leq&\mathcal{O}(1)\log^2\left(\frac{\log^2(T)}{\delta}\right)\frac{s^2p(n^2+p^2)\sigma_{\mathbf{w}}^2}{\pi_{\min}^2}\left(\gamma\sum_{i=1}^{M}\sqrt{T_i}\log^2(T_i)\right) + \mathcal{O}\left(\frac{\sqrt{ns}\log^3(T)}{\delta}\right)
\end{aligned} \tag{C.42}$$

For the term $\gamma \sum_{i=1}^{M} \sqrt{T_i} \log^2(T_i)$, we have

$$
\begin{aligned}
&\gamma \sum_{i=1}^{M} \sqrt{T_i} \log^2(T_i) \\
\leq& \mathcal{O}(1)\gamma\sqrt{T_0}\left(\log^2(T_0)\sum_{i=1}^{M}\sqrt{\gamma}^i + \log^2(\gamma)\sum_{i=1}^{M}\sqrt{\gamma}^i i^2\right) \\
\leq& \mathcal{O}(1)\gamma\sqrt{T_0}\log^2(\gamma)\sum_{i=1}^{M}\sqrt{\gamma}^i i^2 \\
\leq& \mathcal{O}(1)\gamma\sqrt{T_0}\log^2(\gamma) M\sqrt{\gamma}^M \left(\frac{\sqrt{\gamma}}{\sqrt{\gamma}-1}\right)^3\left(M - \frac{1}{\sqrt{\gamma}}\right) \\
\leq& \mathcal{O}(1)\gamma\sqrt{T}\log^2(\gamma)\frac{\log(\frac{T}{T_0})}{\log(\gamma)}\left(\frac{\sqrt{\gamma}}{\sqrt{\gamma}-1}\right)^3\left(\frac{\log(\frac{T}{T_0})}{\log(\gamma)} - \frac{1}{\sqrt{\gamma}}\right) \\
\leq& \mathcal{O}(1)\sqrt{T}\log(\frac{T}{T_0})\left(\frac{\sqrt{\gamma}}{\sqrt{\gamma}-1}\right)^3\left(\gamma\log(\frac{T}{T_0}) - \sqrt{\gamma}\log(\gamma)\right) \\
\leq& \mathcal{O}(\log^2(T)\sqrt{T}).
\end{aligned}
\tag{C.43}
$$

Plugging this back into (C.42), we have

$$
\text{Regret}(T) \leq \mathcal{O}\left(\frac{s^2 p(n^2+p^2)\sigma_{\mathbf{w}}^2}{\pi_{\min}^2}\log^2\left(\frac{\log^2(T)}{\delta}\right)\log^2(T)\sqrt{T} + \frac{\sqrt{ns}\log^3(T)}{\delta}\right) \tag{C.44}
$$

which shows the regret bound in (C.40).

Now we are only left to show the occurrence probability of regret bound (C.40) is larger than $1-\delta$. To do this, we will combine Proposition C.7, C.8, C.9, and C.10 over all $i = 0, 1, \ldots, M-1$. Note that for each individual $i$, these propositions hold only when certain prerequisite conditions on hyper-parameters $c_{\mathbf{x}}$, $c_{\mathbf{z}}$, $T_0$, and $\bar{x}_0$ are satisfied. We first show that under the choices $T_i = \gamma T_{i-1}$, $\sigma_{\mathbf{z},i}^2 = \frac{\sigma_{\mathbf{w}}^2}{T_i}$, and $\delta_{id,i} = \delta_{\mathbf{x}_0,i} = \frac{3}{\pi^2}\cdot\frac{\delta}{(i+1)^2}$ these hyper-parameter conditions can be satisfied for all $i = 0, 1, \ldots, M-1$.

- Proposition C.7 requires that for $i = 1, 2, \ldots,$ conditions $\frac{\sqrt{n\bar{s}\bar{\tau}}\bar{\rho}^{T_0\gamma^i} i^2\pi^2}{3\delta} < 1$ and $\bar{x}_0^2 \geq \frac{n\sqrt{\bar{s}}(\|\mathbf{B}_{1:s}\|^2+1)\sigma_{\mathbf{w}}^2\bar{\tau}}{(1-\bar{\rho})(1-\sqrt{n\bar{s}\bar{\tau}}\bar{\rho}^{T_0\gamma^i} i^2\pi^2/3\delta)}$ need to be satisfied. One can check when $T_0 \geq \frac{1}{\gamma\log(1/\bar{\rho})}\max\{\frac{2}{\log(\gamma)}, \log(\frac{\pi^2\sqrt{n\bar{s}\bar{\tau}}}{3\delta})\} =: \underline{T}_{\mathbf{x}_0}(\delta)$, and picking $\bar{x}_0^2 \geq \frac{n\sqrt{\bar{s}}(\|\mathbf{B}_{1:s}\|^2+1)\sigma_{\mathbf{w}}^2\bar{\tau}}{(1-\bar{\rho})(1-\sqrt{n\bar{s}\bar{\tau}}\bar{\rho}^{T_0\gamma}\pi^2/3\delta)}$ would suffice.
- Proposition C.8 requires that for $i = 0, 1, \ldots,$ condition $T_0\gamma^i \geq \underline{T}_{rgt,\gamma}(\frac{3\delta}{\pi^2(i+1)^2}, T_0\gamma^i)$ holds, which can be satisfied when one chooses $T_0 \geq \mathcal{O}(\underline{T}_{rgt,\gamma}(\delta, T_0))$.
- Proposition C.9 requires the following to hold: $c_{\mathbf{x}} \geq \underline{c}_{\mathbf{x}}(\bar{\rho}, \bar{\tau})$, $c_{\mathbf{z}} \geq \underline{c}_{\mathbf{z}}$, and $T_0\gamma^i \geq \max\left\{\underline{T}_{MC,1}(\frac{3\delta}{8\pi^2 i^2}), \underline{T}_{id,N}(\frac{3\delta}{2\pi^2(i+1)^2}, \bar{\rho}, \bar{\tau})\right\}$. The last one can be satisfied when $T_0 \geq \mathcal{O}(\max\{\underline{T}_{MC,1}(\delta), \underline{T}_{id,N}(\delta, \bar{\rho}, \bar{\tau})\})$.
- Proposition C.10 requires no conditions on hyper-parameters.

Therefore, when $c_{\mathbf{x}} \geq \underline{c}_{\mathbf{x}}(\bar{\rho}, \bar{\tau})$, $c_{\mathbf{z}} \geq \underline{c}_{\mathbf{z}}$,

$$
T_0 \geq \mathcal{O}(\max\{\underline{T}_{\mathbf{x}_0}(\delta), \underline{T}_{rgt,\gamma}(\delta, T_0), \underline{T}_{MC,1}(\delta), \underline{T}_{id,N}(\delta, \bar{\rho}, \bar{\tau})\}) =: \mathcal{O}(\underline{T}_{rgt}(\delta, T_0)),
$$

we can apply Propositions C.7, C.8, C.9, and C.10 to every epoch $i = 0, 1, \ldots, M-1$. First note that Propositions C.7 and C.9 give the following

$$
\begin{aligned}
\mathsf{P}(\mathcal{D}_i \mid \cap_{j=0}^{i-1}\mathcal{E}_j) = \mathsf{P}(\mathcal{D}_i \mid \mathcal{B}_{i-1}\mathcal{D}_{i-1}) > 1 - \frac{3\delta}{\pi^2(i+1)^2}, \quad \mathsf{P}(\mathcal{D}_0) \geq 1 - \frac{3\delta}{\pi^2} \\
\mathsf{P}(\mathcal{C}_i \mid \cap_{j=0}^{i-1}\mathcal{E}_j) = \mathsf{P}(\mathcal{C}_i \mid \mathcal{B}_{i-1}) \geq 1 - \frac{3\delta}{\pi^2(i+1)^2}, \quad \mathsf{P}(\mathcal{C}_0) \geq 1 - \frac{3\delta}{\pi^2}.
\end{aligned}
$$

Then combining the probability bounds in Propositions C.7, C.8, C.9, and C.10, we have

$$\begin{aligned}
&\mathsf{P}\left(\text{Regret bounds in (C.40) holds}\right)\\
\geq&\mathsf{P}(\cap_{i=0}^{M-1}\mathcal{E}_i)\\
=&\mathsf{P}(\mathcal{A}_M, \mathcal{B}_{M-1}, \mathcal{C}_{M-1}, \mathcal{D}_{M-1} \mid \cap_{i=0}^{M-2}\mathcal{E}_i)\cdot\mathsf{P}(\cap_{i=0}^{M-2}\mathcal{E}_i)\\
=&\mathsf{P}(\mathcal{C}_{M-1}, \mathcal{D}_{M-1} \mid \cap_{i=0}^{M-2}\mathcal{E}_i)\cdot\mathsf{P}(\cap_{i=0}^{M-2}\mathcal{E}_i)\\
\geq&\left(1-\delta_{id,M-1}-\delta_{\mathbf{x}_0,M-1}\right)\cdot\mathsf{P}(\cap_{i=0}^{M-2}\mathcal{E}_i)\\
\geq&\prod_{i=0}^{M-1}\left(1-\delta_{id,i}-\delta_{\mathbf{x}_0,i}\right)\\
\geq&1-\sum_{i=0}^{M-1}\left(\delta_{id,i}+\delta_{\mathbf{x}_0,i}\right)\\
\geq&1-\delta.
\end{aligned} \tag{C.45}$$

where the last line holds since $\sum_{i=0}^{M-1}\frac{1}{(i+1)^2}\leq\frac{\pi^2}{6}$. ■

## C.4 Regret Under Uniform Stability — Proof for Theorem 5.2

As we discussed in Section 5.2, under MSS, the regret upper bound in Theorem 5.1 (or the complete version Theorem C.11) involves $\frac{1}{\delta}$ dependency on failure probability $\delta$. By checking the proof for Theorem C.11, we can see the only source for $\frac{1}{\delta}$ is event $\mathcal{D}_i$ in (C.35) and the corresponding Proposition C.7, which provides $1-\delta$ probability bound for event $\mathcal{D}_i$ – the initial state $\mathsf{x}_0^{(i+1)}$ of epoch $i+1$, a.k.a. the final state $\mathsf{x}_{T_i}^{(i)}$ of epoch $i$, is bounded by $\|\mathsf{x}_0^{(i+1)}\|^2=\|\mathsf{x}_{T_i}^{(i)}\|^2\leq\mathcal{O}(\frac{1}{\delta})$. In Proposition C.7, we get this bound using Markov inequality $\|\mathsf{x}_{T_i}^{(i)}\|^2\leq\mathsf{E}[\|\mathsf{x}_{T_i}^{(i)}\|^2]/\delta$ and Lemma A.4 which provides an upper bound on the numerator $\mathsf{E}[\|\mathsf{x}_{T_i}^{(i)}\|^2]$ under MSS. From event $\mathcal{A}_i$ in (C.35) we see the regret of epoch $i$ directly depends on its epoch initial state $\|\mathsf{x}_0^{(i)}\|^2$, thus in the final cumulative regret, the cumulative impact of initial states from all epochs, $\sum_i\|\mathsf{x}_0^{(i)}\|^2$ with order $\frac{1}{\delta}$, will show up, as given in (C.42). Therefore, whether $\frac{1}{\delta}$ terms can be relaxed directly hinges on whether one could refine Proposition C.7 to get a tighter dependency on $\delta$.

This refinement, however, is not possible under the MSS assumption only, and we can easily construct a toy example to show that the $\frac{1}{\delta}$ dependency resulting from the Markov inequality cannot be improved. Consider a two-mode, one-dimensional, autonomous MJS:

$$\begin{cases}x_{t+1}=2x_t\\ x_{t+1}=0.5x_t\end{cases}\text{ with Markov matrix }\mathsf{T}=\begin{bmatrix}0.1 & 0.9\\ 0.1 & 0.9\end{bmatrix}$$

with $x_0\sim\mathcal{N}(0,1)$, and $\mathsf{P}(\omega(0)=1)=0.1$. It is easy to check this MJS is MSS by the spectral radius criterion discussed below Definition 3.1. Also note that with probability $0.1^t$, $\omega(0:t-1)=1$ and $x_t=2^tx_0$. Therefore, for any $a>0$,

$$\mathsf{P}(x_t\geq a)=\sum_{\omega(0:t-1)}\mathsf{P}(x_t\geq a\mid\omega(0:t-1))\mathsf{P}(\omega(0:t-1)) \tag{C.46}$$

$$\geq\mathsf{P}(x_t\geq a\mid\omega(0:t-1)=1)\mathsf{P}(\omega(0:t-1)=1)=0.1^t\cdot\mathsf{P}(x_0\geq 2^{-t}a)\,. \tag{C.47}$$

where the inequality in (C.47) is extremely loose since we condition only on the most improbable event. For standard Gaussian $x_0$, $\mathsf{P}(x_0\geq a)\geq\frac{C}{a}\exp(-\frac{a^2}{2})$ for some absolute constant $C$. Thus $\mathsf{P}(x_t\geq a)\geq C\frac{0.2^t}{a}\exp(-\frac{2^{-2t}a^2}{2})$. From this, we see that for any $a>0$, any $t\geq\log(a)/\log(2)$, we have $\mathsf{P}(x_t\geq a)\geq C\frac{0.2^t}{\sqrt{e}a}$. We can observe that though when $t$ grows slower than $\log(a)$, the tail of $x_t$ has exponential decay, the Markov

inequality decay, i.e. $\frac{1}{a}$, will eventually show up when $t$ gets larger. Interpretation from failure probability $\delta$ perspective is the following: letting $\delta = C\frac{\Phi_x 2^t}{\bar{e}a}$, we have $\mathsf{P}(x_t \le C\frac{\Phi_x 2^t}{\bar{e}\delta}) \le 1-\delta$, which means any $\delta$ dependency lighter than $\frac{1}{\delta}$ must have probability less than $1-\delta$. This further implies that in the regret analysis of adaptive control, in order to obtain better probability dependency, the time horizon has to be limited, which greatly impairs its value in practice.

Intuitively, MSS assumption only provides us with stable behavior of $\|\mathsf{x}_t\|^2$ in the expectation (w.r.t. mode switchings) sense, and having only this first-order moment information is of little use compared with the deterministic Lyapunov stability typically used for LTI systems, which allows one to bound $\|\mathsf{x}_t\|^2$ with only $\log(\frac{1}{\delta})$ dependence ( [15, Lemma C.5]). Then, one may wonder naturally: Does there exist a deterministic version of stability for switched systems? Can this stability (if exists) help build similar dependence for switched systems? The answers to both questions are yes and will be discussed in this appendix. In short, if there exists uniform stability for the MJS, we can adapt Proposition C.7 such that $\|\mathsf{x}_0^{(i)}\|^2$ can instead be bounded much more tightly by $\|\mathsf{x}_0^{(i)}\|^2 \le \mathcal{O}(\log(\frac{1}{\delta}))$, thus the $\frac{1}{\delta}$ dependency can improve to $\log(\frac{1}{\delta})$ in the regret bound (5.5) (or (C.40)). The final improved regret bound is presented in Theorem 5.2. In order to show it, we will need to adapt Proposition C.7 together with several related results (Lemma A.4, Lemma C.1, Lemma C.2) to the uniform stability case, and we append suffix "a" in the result label to denote the adapted versions.

To begin with, recall $\mathsf{K}_{1:\mathsf{S}}^\star$ is the optimal controller for the infinite-horizon MJS-LQR($\mathsf{A}_{1:\mathsf{S}}, \mathsf{B}_{1:\mathsf{S}}, \mathsf{T}, \mathsf{Q}_{1:\mathsf{S}}, \mathsf{R}_{1:\mathsf{S}}$) and define the closed-loop state matrix $\mathsf{L}_i^\star = \mathsf{A}_i + \mathsf{B}_i\mathsf{K}_i^\star$ for all $i$. We let $\theta^\star$ denote the joint spectral radius of $\mathsf{L}_{1:\mathsf{S}}^\star$, i.e. $\theta^\star := \lim_{l\to\infty}\max_{\omega_{1:l}\in[\mathsf{S}]^l}\|\mathsf{L}_{\omega_1}^\star\cdots\mathsf{L}_{\omega_l}^\star\|^{\frac{1}{l}}$. We say $\mathsf{L}_{1:\mathsf{S}}^\star$ is uniformly stable if and only if $\theta^\star < 1$. Similar to $\tau$ in Definition A.1, define $\kappa^\star := \sup_{l\in\mathsf{N}}\max_{\omega_{1:l}\in[\mathsf{S}]^l}\|\mathsf{L}_{\omega_1}^\star\cdots\mathsf{L}_{\omega_l}^\star\|/(\theta^\star)^l$. Note that the pair $\{\theta^\star, \kappa^\star\}$ for uniform stability is just the counterpart of $\{\rho^\star, \tau(\tilde{\mathsf{L}}^\star)\}$ for MSS defined in Appendix C. Similar as before, Table 5 lists all the shorthand notations to be used in this appendix for quick reference.

Table 5: Notations — Uniform Stability

| | |
|---|---|
| $\bar{\sigma}^2$ | $\|\mathsf{B}_{1:\mathsf{S}}\|^2\|\Sigma_{\mathsf{z}}\| + \|\Sigma_{\mathsf{w}}\|$ or $\|\mathsf{B}_{1:\mathsf{S}}\|^2\sigma_{\mathsf{z},0}^2 + \sigma_{\mathsf{w}}^2$ |
| $\bar{\theta}$ | $(1+\theta^\star)/2$ |
| $\bar{\kappa}$ | $\kappa^\star$ |
| $\bar{\mathsf{K}}^{us}_{\mathsf{A},\mathsf{B}}$ | $\frac{1-\rho^\star}{2\kappa^\star\|\mathsf{B}_{1:\mathsf{S}}\|}$ |
| $\bar{\mathsf{K}}_{\mathsf{A},\mathsf{B},\mathsf{T}}$ | $\min\{\bar{\mathsf{K}}^{us}_{\mathsf{A},\mathsf{B}}, \bar{\mathsf{K}}_{\mathsf{A},\mathsf{B}}\}$ |
| $\bar{\bar{\mathsf{K}}}_{\mathsf{A},\mathsf{B},\mathsf{T}}$ | $\min\{\bar{\mathsf{K}}_{\mathsf{A},\mathsf{B},\mathsf{T}}, \frac{\bar{\mathsf{K}}}{2C_{\mathsf{A},\mathsf{B},\mathsf{T}}^{\bar{\mathsf{K}}}}\}$ |
| $\bar{x}^{us}$ | $2\bar{\kappa}^2\bar{\sigma}^2(6\max\{\sqrt{n}e^{3n}, \sqrt{p}e^{3p}\} + \frac{5}{(1-\theta)^2})^2$ |
| $\underline{T}^{us}_{\mathsf{x}_0}(\delta)$ | $\max\{\frac{54\bar{\kappa}^4\bar{\sigma}^2}{(1-\theta)\bar{x}^{us}\log(1/\theta)\log(\gamma)}, \frac{1}{\gamma\log(1/\theta)}\log(6\bar{\kappa}^2 + \frac{54n\sqrt{\bar{s}\bar{\kappa}^4\bar{\sigma}^2\log(\pi^2/3\delta)}}{(1-\theta)(1-\bar{\rho})\bar{x}^{us}\delta})\}$ |
| $\underline{T}^{us}_{rgt,\bar{\mathsf{K}}}(\delta, T)$ | $\mathcal{O}(\frac{\bar{s}(n+p)}{\sqrt{\pi_{\min}}}\bar{\bar{\mathsf{K}}}^{-4}_{\mathsf{A},\mathsf{B},\mathsf{T}}\log(\frac{1}{\delta})\log^4(T))$ |
| | $\mathcal{O}(\frac{\bar{s}(n+p)}{\pi_{\min}}\bar{\bar{\mathsf{K}}}^{-2}_{\mathsf{A},\mathsf{B},\mathsf{T}}\log(\frac{1}{\delta})\log^2(T))$ (when $\mathsf{B}_{1:\mathsf{S}}$ is known) |
| $\underline{T}^{us}_{rgt}(\delta, T)$ | $\max\{\underline{T}^{us}_{\mathsf{x}_0}(\delta), \underline{T}^{us}_{rgt,\bar{\mathsf{K}}}(\delta, T), \underline{T}_{MC,1}(\delta), \underline{T}_{id,N}(\underline{L}, \delta, T, \bar{\rho}, \bar{\tau})\}$ |

The following Lemma A.4a bounds the state $\mathsf{x}_t$ under the designed input in this work. Compared with its counterpart Lemma A.4 which is only able to bound $\mathsf{E}[\|\mathsf{x}_t\|^2]$, Lemma A.4a provides high-probability bound for $\|\mathsf{x}_t\|^2$.

**Lemma A.4a** *Consider an MJS($\mathsf{A}_{1:\mathsf{S}}, \mathsf{B}_{1:\mathsf{S}}, \mathsf{T}$) with noise $\mathsf{w}_t \sim \mathcal{N}(0, \Sigma_{\mathsf{w}})$. Consider controller $\mathsf{K}_{1:\mathsf{S}}$, and let $\mathsf{L}_{1:\mathsf{S}}$ denote the closed-loop state matrices with $\mathsf{L}_i = \mathsf{A}_i + \mathsf{B}_i\mathsf{K}_i$. Assume there exist constants $\kappa$ and $\theta \in [0,1)$ such that, for any sequence $\omega_{1:l} \in [\mathsf{S}]^l$ with any length $l$, $\|\mathsf{L}_{\omega_1}\cdots\mathsf{L}_{\omega_l}\| \le \kappa\theta^l$. Let the input be $\mathsf{u}_t = \mathsf{K}_{\omega_t}\mathsf{x}_t + \mathsf{z}_t$*

*with* $z_t \sim \mathcal{N}(0, \Sigma_z)$. *Then, for any* $t \geq e^{6\max\{n,p\}}$, *with probability at least* $1-\delta$, *we have*

$$\|x_t\|^2 \leq 3\kappa^2\theta^{2t}\|x_0\|^2 + \frac{18\kappa^2\bar{\sigma}^2}{(1-\theta)^2}\log(\frac{1}{\delta}) + c \tag{C.48}$$

*where* $\bar{\sigma}^2 := \|\Sigma_w\| + \|B_{1:s}\|^2\|\Sigma_z\|$ *and* $c := 2\kappa^2\bar{\sigma}^2(6\max\{\sqrt{n}e^{3n}, \sqrt{p}e^{3p}\} + \frac{5}{(1-\theta)^2})^2$.

**Proof** From the MJS dynamics (3.1) and plugging in the input $u_t = K_{\omega(t)}x_t + z_t$, we have the following.

$$\begin{aligned} x_t = \left(\prod_{h=0}^{t-1} L_{\omega(h)}\right)x_0 + \sum_{i=0}^{t-2}\left(\prod_{h=i+1}^{t-1} L_{\omega(h)}\right)B_{\omega(i)}z_i + B_{\omega(t-1)}z_{t-1} \\ \sum_{i=0}^{t-2}\left(\prod_{h=i+1}^{t-1} L_{\omega(h)}\right)w_i + w_{t-1}. \end{aligned} \tag{C.49}$$

Then, by triangle inequality and the assumption that $\|L_{\omega_1}\cdots L_{\omega_l}\| \leq \kappa\theta^l$, we have

$$\begin{aligned} \|x_t\| &\leq \kappa\theta^t\|x_0\| + \kappa\|B_{1:s}\|\sum_{i=0}^{t-1}\theta^{t-i-1}\|z_i\| + \kappa\sum_{i=0}^{t-1}\theta^{t-i-1}\|w_i\| \\ &= \kappa\theta^t\|x_0\| + \kappa\|B_{1:s}\|\sum_{i=0}^{t-1}\theta^{i}\|z_{t-i-1}\| + \kappa\sum_{i=0}^{t-1}\theta^{i}\|w_{t-i-1}\|. \end{aligned} \tag{C.50}$$

For each $w_{t-i-1}$, using Lemma A.5 (replacing $e^{-t}$ with $\delta_i$), we have with probability $1-\delta_i$,

$$\|w_{t-i-1}\| \leq \sqrt{3\|\Sigma_w\|}\log^{0.5}(\frac{1}{\min\{\delta_i, \bar{\delta}_n\}}), \tag{C.51}$$

where $\bar{\delta}_n := e^{-(3+2\sqrt{2})n}$, and $n$ is the dimension of vector $w_{t-i-1}$. In the following, for all $i = 0, 1, \ldots, t-1$, we set $\delta_i = \frac{3}{\pi^2}\frac{\delta}{(i+1)^2}$. First note that when $i \geq \bar{i} := \sqrt{\frac{3\delta}{\pi^2\bar{\delta}_n}} - 1$, we have $\min\{\delta_i, \bar{\delta}_n\} = \delta_i$, i.e. $\delta_i \leq \bar{\delta}_n$, and $\min\{\delta_i, \bar{\delta}_n\} = \bar{\delta}_n$ otherwise. Then, applying union bound for all $i$, we know with probability at least $1 - \frac{\delta}{2}$,

$$\begin{aligned} \sum_{i=0}^{t-1}\theta^i\|w_{t-i-1}\| &\leq \sqrt{3\|\Sigma_w\|}\sum_{i=0}^{t-1}\theta^i\log^{0.5}(\frac{1}{\min\{\delta_i, \bar{\delta}_n\}}) \\ &\leq \sqrt{3\|\Sigma_w\|}\left(\sum_{i=0}^{\bar{i}}\theta^i\log^{0.5}(\frac{1}{\delta_i}) + (\bar{i}+1)\log^{0.5}(\frac{1}{\bar{\delta}_n})\right). \end{aligned} \tag{C.52}$$

In the above equation for the term $\sum_{i=0}^{\bar{i}}\theta^i\log^{0.5}(\frac{1}{\delta_i})$, we have $\sum_{i=0}^{\bar{i}}\theta^i\log^{0.5}(\frac{1}{\delta_i}) = \sum_i\theta^i\log^{0.5}(\frac{\pi^2(i+1)^2}{3\delta}) \leq \sum_i\theta^i(\log^{0.5}(\frac{1}{\delta}) + \sqrt{2}\log^{0.5}(\frac{\pi(i+1)}{\sqrt{3}})) \leq \frac{1}{1-\theta}\log^{0.5}(\frac{1}{\delta}) + \sqrt{2}\sum_i\theta^i\frac{\pi(i+1)}{\sqrt{3}} \leq \frac{1}{1-\theta}\log^{0.5}(\frac{1}{\delta}) + \frac{\sqrt{2}\pi}{\sqrt{3}}\frac{1}{(1-\theta)^2}$. And for the term $(\bar{i}+1)\log^{0.5}(\frac{1}{\bar{\delta}_n})$ in (C.52), by the definitions of $\bar{i}$ and $\bar{\delta}_n$, we have $(\bar{i}+1)\log^{0.5}(\frac{1}{\bar{\delta}_n}) \leq \sqrt{2n}e^{3n}$. Plugging these two results back into (C.52), we have, with probability at least $1 - \frac{\delta}{2}$,

$$\sum_{i=0}^{t-1}\theta^i\|w_{t-i-1}\| \leq \frac{\sqrt{3\|\Sigma_w\|}}{1-\theta}\log^{0.5}(\frac{1}{\delta}) + \frac{5\sqrt{\|\Sigma_w\|}}{(1-\theta)^2} + 3\sqrt{n}e^{3n}\sqrt{\|\Sigma_w\|}. \tag{C.53}$$

Similarly, with probability at least $1 - \frac{\delta}{2}$,

$$\sum_{i=0}^{t-1}\theta^i\|z_{t-i-1}\| \leq \frac{\sqrt{3\|\Sigma_z\|}}{1-\theta}\log^{0.5}(\frac{1}{\delta}) + \frac{5\sqrt{\|\Sigma_z\|}}{(1-\theta)^2} + 3\sqrt{p}e^{3p}\sqrt{\|\Sigma_z\|}. \tag{C.54}$$

Plugging (C.53) and (C.54) back into (C.50) and applying union bound, we have, with probability $1-\delta$,

$$\|x_t\| \le \kappa\theta^t\|x_0\| + \frac{\sqrt{3}\kappa(\sqrt{\|\Sigma_w\|}+\|B_{1:s}\|\sqrt{\|\Sigma_z\|})}{(1-\theta)^2}\log^{0.5}(\frac{1}{\delta}) + \kappa(\sqrt{\|\Sigma_w\|}+\|B_{1:s}\|\sqrt{\|\Sigma_z\|})\left(3\max\{\sqrt{n}e^{3n},\sqrt{p}e^{3p}\}+\frac{5}{(1-\theta)^2}\right). \quad \text{(C.55)}$$

Taking squares of both sides and using Cauchy-Schwartz inequality, we have

$$\|x_t\|^2 \le 3\kappa^2\theta^{2t}\|x_0\|^2 + \frac{18\kappa^2\bar{\sigma}^2}{1-\theta}\log(\frac{1}{\delta}) + c \quad \text{(C.56)}$$

where $\bar{\sigma}^2 := \|\Sigma_w\| + \|B_{1:s}\|^2\|\Sigma_z\|$ and $c := 6\kappa^2\bar{\sigma}^2(3\max\{\sqrt{n}e^{3n},\sqrt{p}e^{3p}\}+\frac{5}{(1-\theta)^2})^2$. ■

The following Lemma C.1a describes that given a set of matrices that have joint spectral radius smaller than 1, i.e. uniformly stable, moderate perturbation can preserve the uniform stability. On the other hand, its counterpart, Lemma C.1, considers perturbation results for MSS.

**Lemma C.1a (Joint Spectral Radius Perturbation)** *Assume $\theta^\square < 1$. For an arbitrary controller $K_{1:s}$ and resulting closed-loop state matrices $L_{1:s}$ with $L_i = A_i + B_iK_i$, let $\theta(L_{1:s})$ denote the joint spectral radius of $L_{1:s}$. Assume $\|K_{1:s} - K^\square_{1:s}\| \le \epsilon_K^{us} := \frac{1-\theta^\star}{2\kappa^\star\|B_{1:s}\|}$, then for any sequence $\omega_{1:l} \in [s]^l$ with any length $l$,*

$$\|\prod_{j=1}^{l} L_{\omega_j}\| \le \bar{\kappa}\bar{\theta}^l \quad \text{(C.57)}$$

$$\theta(L_{1:s}) \le \bar{\theta}. \quad \text{(C.58)}$$

*where $\bar{\kappa} = \kappa^\square$ and $\bar{\theta} = \frac{1+\theta^\star}{2}$.*

**Proof** Let $E_i := L_i - L_i^\square$, then we see $\|E_i\| \le \|B_{1:s}\|\epsilon_K^{us}$ and $\prod_{j=1}^{l} L_{\omega_j} = \prod_{j=1}^{l}(L^\square_{\omega_j} + E_{\omega_j})$. In the expansion of $\prod_{j=1}^{l}(L^\square_{\omega_j} + E_{\omega_j})$, for each $h = 0, 1, \ldots, l$, there are $\binom{l}{h}$ terms, each of which is a product where $E$ has degree $h$ and $L^\square$ has degree $l-h$. We let $F_{h,I}$ with $h = 0, 1, \ldots, l$ and $I \in [\binom{l}{h}]$ to index such terms. Note that $\|F_{h,I}\| \le (\kappa^\square)^{h+1}(\theta^\square)^{l-h}(\|B_{1:s}\|\epsilon_K^{us})^h$. Then, we have

$$\begin{aligned}\|\prod_{j=1}^{l} L_{\omega_j}\| &\le \sum_{h=0}^{l}\sum_{I\in[\binom{l}{h}]}\|F_{h,I}\| \\ &\le \sum_{h=0}^{l}\binom{l}{h}(\kappa^\square)^{h+1}(\theta^\square)^{l-h}(\|B_{1:s}\|\epsilon_K^{us})^h \\ &\le \kappa^\square(\kappa^\square\|B_{1:s}\|\epsilon_K^{us}+\theta^\square)^l.\end{aligned} \quad \text{(C.59)}$$

Then (C.57) follows from the fact that $\epsilon_K^{us} \le \frac{1-\theta^\star}{2\kappa^\star\|B_{1:s}\|}$ and $\bar{\theta} := \frac{1+\theta^\star}{2}$. To proceed, noticing that $\theta(L_{1:s}) = \lim_{l\to\infty}\max_{\omega_{1:l}\in[s]^l}\|\prod_{j=1}^{l} L_{\omega_j}\|^{\frac{1}{l}}$ and using the result in (C.57), we can show (C.58). ■

In Lemma C.1a, if $K_{1:s}$ is computed by solving the infinite-horizon MJS-LQR($\hat{A}_{1:s}, \hat{B}_{1:s}, \hat{T}, Q_{1:s}, R_{1:s}$) for some estimated MJS($\hat{A}_{1:s}, \hat{B}_{1:s}, \hat{T}$), the following result provides the required estimation accuracy such that the resulting $K_{1:s}$ is uniformly stabilizing.

**Lemma C.2a** *Under the setup of Lemma C.2, if $\max\{\epsilon_{A,B}, \epsilon_T\} \le \bar{\epsilon}_{A,B,T}$, then we have $\|K_{1:s} - K^\square_{1:s}\| \le \epsilon_K$, and Lemma C.1a is applicable.*

Recall we defined events $\mathcal{A}_i, \mathcal{B}_i, \mathcal{C}_i, \mathcal{D}_i$ in (C.60) to analyze the events happen in each epoch of the regret. To adapt to the uniform stability assumption, we redefine event $\mathcal{B}_i$ and $\mathcal{D}_i$ while keep $\mathcal{A}_i$ and $\mathcal{C}_i$ as before. For easier reference, We list all of them below.

$$
\begin{aligned}
\mathcal{A}_i &= \left\{ \text{Regret}_i \le \mathcal{O}\left( sp\left( \Gamma_{\mathbf{A},\mathbf{B}}^{(i-1)} + \Gamma_{\mathbf{T}}^{(i-1)} \right)^2 \sigma_{\mathbf{w}}^2 T_i + \sqrt{ns}\|\mathsf{x}_0^{(i)}\|^2 + n\sqrt{s}\sigma_{\mathbf{z},i}^2 T_i + c_{\mathsf{A}} \right) \right\} \\
\mathcal{B}_i &= \left\{ \Gamma_{\mathbf{A},\mathbf{B}}^{(i)} \le \bar{\Gamma}_{\mathbf{A},\mathbf{B},\mathbf{T}},\ \Gamma_{\mathbf{T}}^{(i)} \le \bar{\Gamma}_{\mathbf{A},\mathbf{B},\mathbf{T}},\ \Gamma_{\mathbf{K}}^{(i+1)} \le \bar{\Gamma}_{\mathbf{K}} \right\}, \forall i = 0, 1, \dots \\
\mathcal{C}_i &= \left\{ \Gamma_{\mathbf{A},\mathbf{B}}^{(i)} \le \mathcal{O}\left( \log(\frac{1}{\delta_{id,i}}) \frac{\sigma_{\mathbf{z},i} + \sigma_{\mathbf{w}}}{\sigma_{\mathbf{z},i}\pi_{\min}} \frac{\sqrt{s}(n+p)\log(T_i)}{\sqrt{T_i}} \right), \right. \\
&\qquad \left. \Gamma_{\mathbf{T}}^{(i)} \le \mathcal{O}\left( \log(\frac{1}{\delta_{id,i}}) \frac{1}{\pi_{\min}} \sqrt{\frac{\log(T_i)}{T_i}} \right) \right\} \\
\mathcal{D}_i &= \left\{ \|\mathsf{x}_0^{(i+1)}\|^2 = \|\mathsf{x}_{T_i}^{(i)}\|^2 \le \frac{18\bar{\kappa}^2\bar{\sigma}^2}{(1-\bar{\theta})^2} \log(\frac{1}{\delta_{\mathbf{x}_0,i}}) + 2\bar{x}^{us} \right\}, \forall i = 1, 2, \dots, \\
\mathcal{D}_0 &= \left\{ \|\mathsf{x}_0^{(1)}\|^2 = \|\mathsf{x}_{T_0}^{(0)}\|^2 \le \frac{n\sqrt{s}\bar{\tau}\bar{\sigma}^2/(1-\bar{\rho})}{\delta_{\mathbf{x}_0,0}} \right\},
\end{aligned}
\tag{C.60}
$$

where $\bar{x}^{us} := 2\bar{\kappa}^2\bar{\sigma}^2(6\max\{\sqrt{n}e^{3n}, \sqrt{p}e^{3p}\} + \frac{5}{(1-\theta)^2})^2$, $\bar{\Gamma}_{\mathbf{K}} := \min\{\bar{\Gamma}_{\mathbf{K}}^{us}, \bar{\Gamma}_{\mathbf{K}}\}$ , $\bar{\Gamma}_{\mathbf{A},\mathbf{B},\mathbf{T}} := \min\{\bar{\Gamma}_{\mathbf{A},\mathbf{B},\mathbf{T}}, \frac{\bar{\Gamma}_{\mathbf{K}}}{2C_{\mathbf{A},\mathbf{B},\mathbf{T}}^{\mathbf{K}}}\}$ and $\|\mathsf{B}_{1:s}\|^2\sigma_{\mathbf{z},0}^2 + \sigma_{\mathbf{w}}^2$. Event $\mathcal{D}_i$ describes the initial state magnitude of epoch $i+1$. Since Algorithm 2 requires initial MSS stabilizing controller $\mathsf{K}_{1:s}^{(0)}$ for epoch 0, and as in the proof for the following Proposition C.7a, epoch $1, 2, \dots$ have uniformly stabilizing controller, thus we define $\mathcal{D}_0$ and $\mathcal{D}_1, \mathcal{D}_2, \dots$ separately.

**Proposition C.7a** *Assuming* $T_i \ge \frac{1}{2\log(1/\theta)} \log\left(6\bar{\kappa}^2 + \frac{54\bar{\kappa}^4\bar{\sigma}^2}{(1-\theta)\bar{x}^{us}} \log(\frac{1}{\delta_{\mathbf{x}_0,i-1}})\right)$; $T_1 \ge \frac{1}{2\log(1/\theta)} \log\left(\frac{3n\sqrt{s}\bar{\kappa}^2\bar{\tau}\bar{\sigma}^2}{(1-\bar{\rho})\bar{x}^{us}\delta_{\mathbf{x}_0,0}}\right)$, *we have*

$$
\mathsf{P}(\mathcal{D}_i \mid \mathcal{B}_{i-1}, \mathcal{D}_{i-1}) \ge 1 - \delta_{\mathbf{x}_0,i} \tag{C.61}
$$

*and* $\mathsf{P}(\mathcal{D}_0) \ge 1 - \delta_{\mathbf{x}_0,0}$.

**Proof** For the initial epoch 0, i.e. $i = 0$, since we assume in Algorithm 2 that the initial controller $\mathsf{K}_{1:s}^{(0)}$ stabilizes the MJS in the mean-squared sense, similar to the proof for Proposition C.7, we have $\mathsf{E}[\|\mathsf{x}_{T_0}^{(0)}\|^2] \le n\sqrt{s}(\|\mathsf{B}_{1:s}\|^2\sigma_{\mathbf{z},0}^2 + \sigma_{\mathbf{w}}^2)\frac{\bar{\tau}}{1-\bar{\rho}}$. Then by Markov inequality, with probability $1 - \delta_{\mathbf{x}_0,0}$, $\|\mathsf{x}_{T_0}^{(0)}\|^2 \le \frac{n\sqrt{s}\bar{\tau}\bar{\sigma}^2/(1-\bar{\rho})}{\delta_{\mathbf{x}_0,0}}$ where $\bar{\sigma}^2 := \|\mathsf{B}_{1:s}\|^2\sigma_{\mathbf{z},0}^2 + \sigma_{\mathbf{w}}^2$. This shows $\mathsf{P}(\mathcal{D}_0) \ge 1 - \delta_{\mathbf{x}_0,0}$.

For epoch $i = 1, 2, \dots$, given event $\mathcal{B}_{i-1}$, we know $\Gamma_{\mathbf{K}}^{(i)} \le \bar{\Gamma}_{\mathbf{K}} \le \bar{\Gamma}_{\mathbf{K}}^{us}$. Let $\mathsf{L}_{1:s}^{(i)}$ denote the closed-loop state matrices for epoch $i$, then by Lemma C.1a, $\Gamma_{\mathbf{K}}^{(i)} \le \bar{\Gamma}_{\mathbf{K}}^{us}$ implies that for any $l$ and any sequence $\omega_{1:l} \in [s]^l$, $\|\prod_{j=1}^{l} \mathsf{L}_{\omega_j}^{(i)}\| \le \bar{\kappa}\bar{\theta}^l$. Then using the bound on $\|\mathsf{x}_t\|$ in Lemma A.4a, we have, with probability $1 - \delta_{\mathbf{x}_0,i}$,

$$
\|\mathsf{x}_{T_i}^{(i)}\|^2 \le \frac{18\bar{\kappa}^2\bar{\sigma}^2}{(1-\bar{\theta})^2} \log(\frac{1}{\delta_{\mathbf{x}_0,i}}) + 3\bar{\kappa}^2\bar{\theta}^{2T_i}\|\mathsf{x}_0^{(i)}\|^2 + \bar{x}^{us} \tag{C.62}
$$

where $\bar{x}^{us} := 2\bar{\kappa}^2\bar{\sigma}^2(6\max\{\sqrt{n}e^{3n}, \sqrt{p}e^{3p}\} + \frac{5}{(1-\theta)^2})^2$.

- When $i = 1$, given $D_0$, i.e. $\|\mathsf{x}_0^{(1)}\|^2 \le \frac{n\sqrt{s}\bar{\tau}\bar{\sigma}^2/(1-\bar{\rho})}{\delta_{\mathbf{x}_0,0}}$, the above (C.62) gives $\|\mathsf{x}_{T_1}^{(1)}\|^2 \le \frac{18\bar{\kappa}^2\bar{\sigma}^2}{(1-\theta)^2}\log(\frac{1}{\delta_{\mathbf{x}_0,1}}) + 3\bar{\kappa}^2\bar{\theta}^{2T_1}\frac{n\sqrt{s}\bar{\tau}\bar{\sigma}^2/(1-\bar{\rho})}{\delta_{\mathbf{x}_0,0}} + \bar{x}^{us}$. One can check that when $T_1 \ge \frac{1}{2\log(1/\theta)}\log\left(\frac{3n\sqrt{s}\bar{\kappa}^2\bar{\tau}\bar{\sigma}^2}{(1-\bar{\rho})\bar{x}^{us}\delta_{\mathbf{x}_0,0}}\right)$, we have that $3\bar{\kappa}^2\bar{\theta}^{2T_1}\frac{n\sqrt{s}\bar{\tau}\bar{\sigma}^2/(1-\bar{\rho})}{\delta_{\mathbf{x}_0,0}} \le \bar{x}^{us}$, which gives

$$
\|\mathsf{x}_{T_1}^{(1)}\|^2 \le \frac{18\bar{\kappa}^2\bar{\sigma}^2}{(1-\bar{\theta})^2}\log(\frac{1}{\delta_{\mathbf{x}_0,1}}) + 2\bar{x}^{us}. \tag{C.63}
$$

- When $i = 2, 3, \ldots$, given event $\mathcal{D}_{i-1}$, i.e. $\|\mathbf{x}_0^{(i)}\|^2 \le \frac{18\bar{\kappa}^2\bar{\sigma}^2}{(1-\bar{\theta})^2}\log(\frac{1}{\delta_{\mathbf{x}_0,i-1}}) + 2\bar{x}^{us}$, the above (C.62) gives $\|\mathbf{x}_{T_i}^{(i)}\|^2 \le \frac{18\bar{\kappa}^2\bar{\sigma}^2}{(1-\bar{\theta})^2}\log(\frac{1}{\delta_{\mathbf{x}_0,i}}) + 3\bar{\kappa}^2\bar{\theta}^{2T_i}\left(\frac{18\bar{\kappa}^2\bar{\sigma}^2}{(1-\bar{\theta})^2}\log(\frac{1}{\delta_{\mathbf{x}_0,i-1}}) + 2\bar{x}^{us}\right) + \bar{x}^{us}$. Similarly, when the trajectory length during the $i_{th}$ epoch, $T_i \ge \frac{1}{2\log(1/\theta)}\log\left(6\bar{\kappa}^2 + \frac{54\bar{\kappa}^4\bar{\sigma}^2}{(1-\theta)\bar{x}^{us}}\log(\frac{1}{\delta_{\mathbf{x}_0,i-1}})\right)$, we further have

$$\|\mathbf{x}_{T_i}^{(i)}\|^2 \le \frac{18\bar{\kappa}^2\bar{\sigma}^2}{(1-\bar{\theta})^2}\log(\frac{1}{\delta_{\mathbf{x}_0,i}}) + 2\bar{x}^{us}. \tag{C.64}$$

Combining (C.63) and (C.64), we can claim: for epoch $i = 1, 2, \ldots$, when $T_1 \ge \frac{1}{2\log(1/\theta)}\log\left(\frac{3n\sqrt{s\bar{\kappa}^2\bar{\tau}\bar{\sigma}^2}}{(1-\bar{\rho})\bar{x}^{us}\delta_{\mathbf{x}_0,0}}\right)$ and $T_i \ge \frac{1}{2\log(1/\theta)}\log\left(6\bar{\kappa}^2 + \frac{54\bar{\kappa}^4\bar{\sigma}^2}{(1-\theta)\bar{x}^{us}}\log(\frac{1}{\delta_{\mathbf{x}_0,i-1}})\right)$, we have $\mathsf{P}(\mathcal{D}_i \mid \mathcal{B}_{i-1}, \mathcal{D}_{i-1}) \ge 1 - \delta_{\mathbf{x}_0,i}$. ■

The following Proposition C.8a says that if a good controller is used in epoch $i$, then the final state $x_{T_i}^{(i)}$ of epoch $i$ (the initial state of epoch $i+1$) can be bounded.

**Proposition C.8a** *Suppose every epoch $i$ has length $T_i \ge \underline{T}^{us}_{rgt,-}(\delta_{id,i}, T_i)$. Then,*

$$\mathsf{P}(\mathcal{B}_i \mid \mathcal{C}_i, \cap_{j=0}^{i-1}\mathcal{E}_j) = \mathsf{P}(\mathcal{B}_i \mid \mathcal{C}_i) = 1 \tag{C.65}$$

### C.4.1 Proof for Theorem 5.2

**Theorem C.12 (Complete version of Theorem 5.2)** *Assume that the initial state $\mathbf{x}_0 = 0$, Assumptions 1 and 2 hold, and $\mathsf{L}_{1,s}^{\square}$ is* uniformly stable*. Suppose $c_{\mathbf{x}} \ge \underline{c}_{\mathbf{x}}(\bar{\rho}, \bar{\tau}), c_{\mathbf{z}} \ge \underline{c}_{\mathbf{z}}$, $T_0 \ge \mathcal{O}(\underline{T}^{us}_{rgt}(\delta, T_0))$. Then, with probability at least $1-\delta$, Algorithm 2 achieves*

$$\mathrm{Regret}(T) \le \mathcal{O}\left(\frac{s^2p(n^2+p^2)\sigma_{\mathbf{w}}^2}{\pi_{\min}^2}\log\left(\frac{\log^2(T)}{\delta}\right)\log^2(T)\sqrt{T}\right). \tag{C.66}$$

**Proof** The proof is almost the same as the proof for the MSS regret upper bound in Theorem C.11 in Appendix C.3.1, thus we only present the key steps and omit certain details of intermediate steps.

In the following, we set $\delta_{id,i} = \delta_{\mathbf{x}_0,i} = \frac{3}{\pi^2}\cdot\frac{\delta}{(i+1)^2}$. Similar to the counterpart (C.41), event $\mathcal{E}_i = \mathcal{A}_{i+1}\cap\mathcal{B}_i\cap\mathcal{C}_i\cap\mathcal{D}_i$ implies the following: for $i = 1, 2, \ldots$,

$$\begin{aligned}
&\mathrm{Regret}_{i+1}\\
\le&\mathcal{O}(1)\log\left(\frac{(i+1)^2}{\delta}\right)sp\left(\frac{\sigma_{\mathbf{z},i}+\sigma_{\mathbf{w}}}{\sigma_{\mathbf{z},i}\pi_{\min}}\cdot\frac{\sqrt{s}(n+p)\log(T_i)}{\sqrt{T_i}} + \frac{\sqrt{\log(T_i)}}{\pi_{\min}\sqrt{T_i}}\right)^2\sigma_{\mathbf{w}}^2T_{i+1}\\
&+\mathcal{O}(1)\log\left(\frac{i+1}{\delta}\right)\frac{18\sqrt{ns}\bar{\kappa}^2\bar{\sigma}^2}{(1-\bar{\theta})^2} + \mathcal{O}(n\sqrt{s}\sigma_{\mathbf{z},i+1}^2T_{i+1}) + \mathcal{O}(1)\\
\le&\mathcal{O}(1)\log\left(\frac{(i+1)^2}{\delta}\right)\frac{s^2p(n^2+p^2)\gamma}{\pi_{\min}^2}\sigma_{\mathbf{w}}^2\sqrt{T_i}\log^2(T_i) + \mathcal{O}(1)\log\left(\frac{i+1}{\delta}\right)\frac{18\sqrt{ns}\bar{\kappa}^2\bar{\sigma}^2}{(1-\bar{\theta})^2};
\end{aligned} \tag{C.67}$$

and for $i = 0$,

$$\mathrm{Regret}_1 \le \mathcal{O}(1)\log\left(\frac{1}{\delta}\right)\frac{s^2p(n^2+p^2)\gamma}{\pi_{\min}^2}\sigma_{\mathbf{w}}^2\sqrt{T_0}\log^2(T_0) + \mathcal{O}(1)\left(\frac{1}{\delta}\right)\frac{n^{1.5}s\bar{\tau}\bar{\sigma}^2}{1-\bar{\rho}}. \tag{C.68}$$

Note that the difference between (C.67) ($i = 1, 2, \ldots$) and (C.68) ($i = 0$) is due to the difference between the event $\mathcal{D}_i$ for $i = 1, 2, \ldots$ and event $\mathcal{D}_0$. Compared with the MSS counterpart (C.41), we see the $\frac{(i+1)^2}{\delta}$ dependence in (C.41) is now replaced with $\log\left(\frac{i+1}{\delta}\right)$. For all $M := \mathcal{O}(\log_\gamma(\frac{T}{T_0}))$ epochs, similar to the

counterpart (C.42), event $\cap_{i=0}^{M-1}\mathcal{E}_i$ implies

$$\begin{aligned}&\text{Regret}(T)\\ =&\mathcal{O}\Big(\sum_{i=1}^{M}\text{Regret}_i\Big)\\ \le&\mathcal{O}\bigg(\frac{s^2p(n^2+p^2)\sigma_{\mathbf{w}}^2}{\pi_{\min}^2}\log\Big(\frac{\log^2(T)}{\delta}\Big)\sqrt{T}\log^2(T)+\frac{18\sqrt{ns}\bar{\kappa}^2\bar{\sigma}^2}{(1-\bar{\theta})^2}\log\Big(\frac{\log(T)}{\delta}\Big)\log(T)\bigg)\end{aligned}\tag{C.69}$$

which shows the main result (C.66). Note that in the above summation, we have omit $\frac{1}{\delta}$ term in $\text{Regret}_1$ since it does not scale with time and can be dominated by the rest.

Now we are only left to show the occurrence probability of regret bound (C.66) is larger than $1-\delta$. To do this, we will combine Proposition C.7a, C.8a, C.9, and C.10 over all $i=0,1,\dots,M-1$. Note that for each individual $i$, these propositions hold only when certain prerequisite conditions on hyper-parameters $c_{\mathbf{x}}$, $c_{\mathbf{z}}$, and $T_0$ are satisfied. We first show that under the choices $T_i=\gamma T_{i-1}$, $\sigma_{\mathbf{z},i}^2=\frac{\bar{\sigma}_{\mathbf{w}}^2}{T_i}$, and $\delta_{id,i}=\delta_{\mathbf{x}_0,i}=\frac{3}{\pi^2}\cdot\frac{\delta}{(i+1)^2}$ these hyper-parameter conditions can be satisfied for all $i=0,1,\dots,M-1$.

- Proposition C.7a requires the following to hold: $T_0\gamma^i\ge\frac{1}{2\log(1/\theta)}\log\Big(6\bar{\kappa}^2+\frac{54\bar{\kappa}^4\bar{\sigma}^2}{(1-\theta)\bar{x}^{us}}\log(\frac{i^2\pi^2}{3\delta})\Big)$ and $T_0\gamma\ge\frac{1}{2\log(1/\theta)}\log\Big(\frac{\pi^2n\sqrt{s\bar{\kappa}^2\bar{\tau}\bar{\sigma}^2}}{(1-\bar{\rho})\bar{x}^{us}\delta}\Big)$. One can check $T_0\ge\max\{\frac{54\bar{\kappa}^4\bar{\sigma}^2}{(1-\theta)\bar{x}^{us}\log(1/\theta)\log(\gamma)},\frac{1}{\gamma\log(1/\theta)}\log(6\bar{\kappa}^2+\frac{54n\sqrt{s\bar{\kappa}^4\bar{\sigma}^2\log(\pi^2/3\delta)}}{(1-\theta)(1-\bar{\rho})\bar{x}^{us}\delta})\}=:\underline{T}_{\mathbf{x}_0}^{us}(\delta)$ would suffice.
- Proposition C.8a requires that for $i=0,1,\dots,$ condition $T_0\gamma^i\ge\underline{T}_{rgt,\text{-}}^{us}(\frac{3\delta}{\pi^2(i+1)^2},T_0\gamma^i)$ holds, which can be satisfied when one chooses $T_0\ge\mathcal{O}(\underline{T}_{rgt,\text{-}}^{us}(\delta,T_0))$.
- Proposition C.9 requires the following to hold: $c_{\mathbf{x}}\ge\underline{c}_{\mathbf{x}}(\bar{\rho},\bar{\tau}),c_{\mathbf{z}}\ge\underline{c}_{\mathbf{z}}$, and $T_0\gamma^i\ge\max\big\{\underline{T}_{MC,1}(\frac{3\delta}{8\pi^2i^2}),\underline{T}_{id,N}(\frac{3\delta}{2\pi^2(i+1)^2},\bar{\rho},\bar{\tau})\big\}$. The last one can be satisfied when we have $T_0\ge\mathcal{O}(\max\{\underline{T}_{MC,1}(\delta),\underline{T}_{id,N}(\delta,\bar{\rho},\bar{\tau})\})$.
- Proposition C.10 requires no conditions on hyper-parameters.

Therefore, when $c_{\mathbf{x}}\ge\underline{c}_{\mathbf{x}}(\bar{\rho},\bar{\tau}),c_{\mathbf{z}}\ge\underline{c}_{\mathbf{z}}$, $T_0\ge\mathcal{O}(\max\{\underline{T}_{\mathbf{x}_0}^{us}(\delta),\underline{T}_{rgt,\text{-}}^{us}(\delta,T_0),\underline{T}_{MC,1}(\delta),\underline{T}_{id,N}(\delta,\bar{\rho},\bar{\tau})\})=:\mathcal{O}(\underline{T}_{rgt}^{us}(\delta,T_0))$, we can apply Proposition C.7a, C.8a, C.9, and C.10 to every epoch $i=0,1,\dots,M-1$. Similar to (C.45), this gives $\mathsf{P}(\text{Regret bounds in (C.66) holds})\ge\mathsf{P}(\cap_{i=0}^{M-1}\mathcal{E}_i)\ge1-\delta$. ∎

**Remark C.5** *Note that though system identification result Theorem 4.1 might also benefit from the newly added uniform stability in this appendix, but since the dependencies $\sqrt{T}$ and $\log(\frac{1}{\delta})$ in Theorem 4.1 are close the the optimal ones for LTI systems, so we only focus on refining the regret upper bound under uniform stability and leave adapting the entire framework to uniform stability to potential future work.*