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

Yahya Sattar [*1] **Zhe Du** [*2] **Davoud Ataee Tarzanagh** [2] **Necmiye Ozay** [2] **Laura Balzano** [2] **Samet Oymak** [1]

Abstract

Learning how to effectively control unknown dynamical systems is crucial for autonomous systems. This task becomes more challenging 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 $\tilde{\mathcal{O}}(1/\sqrt{T})$. 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 the proposed adaptive control scheme achieves $\tilde{\mathcal{O}}(\sqrt{T})$ regret, which can be improved to $\hat{\mathcal{O}}(\log(T))$ with partial knowledge of the system. Our analysis introduces innovations to handle MJS specific challenges (e.g. Markovian jumps) and provides insights into system theoretic quantities that affect learning accuracy and control performance.

[*]Equal contribution [1]Department of Electrical and Computer Engineering, University of California, Riverside, USA [2]Department of Electrical Engineering and Computer Science, University of Michigan, Ann Arbor, USA. Correspondence to: Yahya Sattar <ysatt001@ucr.edu>, Zhe Du <zhedu@umich.edu>.

Proceedings of the 38^{th} International Conference on Machine Learning, PMLR 139, 2021.

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 (Campi & Kumar, 1998; Abbasi-Yadkori & Szepesvári, 2011; Dean et al., 2019; Mania et al., 2019; Lale et al., 2020a; Abeille & Lazaric, 2020).

A generalization of linear dynamical systems that can capture dynamics that switch between multiple linear systems, called modes, according to an underlying finite Markov chain is Markov jump linear systems (MJSs). 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 allows different control goals under different modes. While the MJS-LQR problem is also well understood when one has perfect knowledge of the system dynamics (Chizeck et al., 1986; Costa et al., 2006), in practice, it is not always possible to know 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 controlling investments in financial markets may be modeled as MJS-LQR problems with unknown system dynamics. Earlier works have aimed at analyzing the asymptotic properties (i.e., stability) of adaptive controllers for unknown MJS both in continuous-time (Caines & Zhang, 1995) and discrete-time (Xue & Guo, 2001) settings, however, despite the practical importance of MJS, non-asymptotic sample complexity results and regret analysis for MJS 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

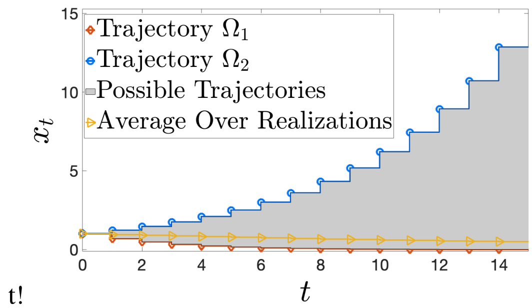

Figure 1. State trajectories for a two-mode MJS $\begin{cases} x_{t+1} = 0.7x_t \\ x_{t+1} = 1.2x_t \end{cases}$ with Markov matrix $\begin{bmatrix} 0.7 & 0.3 \\ 0.4 & 0.6 \end{bmatrix}$ and $\mathbf{x}_0 = 1$. Red and blue curves: mode switching sequence $\Omega_1 = \{1, 1, \ldots\}$, $\Omega_2 = \{2,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

is that, typically, the stability of MJS is understood only in the *mean-square sense*. This is in stark contrast to deterministic stability (e.g., as in LQR), where the system is guaranteed to converge towards an equilibrium point in the absence of noise. On the other hand, the convergence of MJS trajectories towards an equilibrium depends heavily on how the switching between modes occurs.

Figure 1 shows an example (reproduced from (Costa et al., 2006)) 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. (Sarkar et al., 2019)). 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. 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 the 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. Importantly, our guarantees are optimal in the trajectory length T. Specifically, our contributions are:

(I) System identification: For an MJS with s modes, the system dynamics involve Markov chain matrix $\mathbf{T} \in \mathbb{R}^{s \times s}$ and s state-input matrix pairs $(\mathbf{A}_i, \mathbf{B}_i)_{i=1}^{s}$. We provide an algorithm (Alg. 1) to estimate these dynamics with the optimal error rate of $\tilde{\mathcal{O}}(1/\sqrt{T})$ [1]. Specifically, the sample complexity grows as $T \gtrsim \text{poly}(s)(n+p)$ where n and p are the state and input dimensions respectively.

(II) $\tilde{\mathcal{O}}(\sqrt{T})$-regret bound: We employ the system identification guarantees for the MJS-LQR. When system dynamics are unknown, we show that our certainty-equivalent adaptive MJS-LQR algorithm (Alg. 2) achieves a regret of $\tilde{\mathcal{O}}(\sqrt{T})$. Remarkably, this coincides with the optimal regret bound for the standard LQR problem obtained via certainty equivalence (Mania et al., 2019). Furthermore, we show that when the input matrices are known, the regret bound can be significantly improved to $\mathcal{O}(\text{polylog}(T))$, which coincides with the case in (Cassel et al., 2020) for standard LQR.

2. Preliminaries and Problem Setup

We use boldface uppercase (lowercase) letters to denote matrices (vectors). For a matrix $\mathbf{V}$, $\rho(\mathbf{V})$ denotes its spectral radius. The Kronecker product of two matrices $\mathbf{M}$ and $\mathbf{N}$ is denoted as $\mathbf{M} \otimes \mathbf{N}$. $\mathbf{V}_{1:s}$ denotes a set of s matrices $\{\mathbf{V}_i\}_{i=1}^{s}$ of same dimensions. We define $[s] := \{1, 2, \ldots, s\}$. Throughout, $\tilde{\mathcal{O}}(\cdot)$ and $\hat{\mathcal{O}}(\cdot)$ hide polylog$(\frac{1}{\delta})$ and poly$(\frac{1}{\delta})$ terms respectively.

2.1. Markov Jump Linear Systems

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

$$\begin{aligned} \mathbf{x}_{t+1} &= \mathbf{A}_{\omega(t)}\mathbf{x}_t + \mathbf{B}_{\omega(t)}\mathbf{u}_t + \mathbf{w}_t \\ \text{s.t.} \quad & \omega(t) \sim \text{Markov Chain}(\mathbf{T}). \end{aligned} \tag{1}$$

where $\mathbf{x}_t \in \mathbb{R}^n$, $\mathbf{u}_t \in \mathbb{R}^p$ and $\mathbf{w}_t \in \mathbb{R}^n$ are the state, input, and process noise of the MJS at time t. Throughout, we assume $\mathbf{x}_0 \sim \mathcal{D}_x$ and $\{\mathbf{w}_t\}_{t=0}^{\infty} \overset{\text{i.i.d.}}{\sim} \mathcal{N}(0, \sigma_{\mathbf{w}}^2 \mathbf{I}_n)$. There are s modes in total, and the dynamics of mode i is given by the state matrix $\mathbf{A}_i$ and input matrix $\mathbf{B}_i$. The active mode at time t is indexed by $\omega(t) \in [s]$. The MJS mode switching sequence $\{\omega(t)\}_{t=0}^{\infty}$ follows a Markov chain with transition matrix $\mathbf{T} \in \mathbb{R}^{s \times s}$ such that for all $t \geq 0$, the ij-th element of $\mathbf{T}$ denotes the conditional probability $[\mathbf{T}]_{ij} := \mathbb{P}\big(\omega(t+1) = j \mid \omega(t) = i\big)$, $\forall\, i, j \in [s]$. Throughout, we assume that the initial state $\mathbf{x}_0$, Markov chain $\{\omega(t)\}_{t=0}^{\infty}$, and noise $\{\mathbf{w}_t\}_{t=0}^{\infty}$ are mutually independent. We use MJS$(\mathbf{A}_{1:s}, \mathbf{B}_{1:s}, \mathbf{T})$ to refer to an MJS with state equation (1) parameterized by $(\mathbf{A}_{1:s}, \mathbf{B}_{1:s}, \mathbf{T})$.

For mode-dependent state-feedback controller $\mathbf{K}_{1:s}$ that yields the input $\mathbf{u}_t = \mathbf{K}_{\omega(t)}\mathbf{x}_t$, we use $\mathbf{L}_i := \mathbf{A}_i + \mathbf{B}_i\mathbf{K}_i$ to denote the closed-loop state matrix for mode i. We use $\mathbf{x}_{t+1} = \mathbf{L}_{\omega(t)}\mathbf{x}_t$ to denote the noise-free autonomous MJS, either open-loop ($\mathbf{L}_i = \mathbf{A}_i$) or closed-loop ($\mathbf{L}_i = \mathbf{A}_i + \mathbf{B}_i\mathbf{K}_i$). 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 1 (Mean-square stability (Costa et al., 2006))**.**

[1] Here $\tilde{\mathcal{O}}(\cdot)$ hides ploylogarithmic factors in T, $1/\delta$ etc.

We say MJS in (1) *with* $\mathbf{u}_t = 0$ *is mean-square stable (MSS) if there exists* $\mathbf{x}_\infty, \mathbf{\Sigma}_\infty$ *such that for any initial state* $\mathbf{x}_0$ *and mode* $\omega(0)$*, as* $t \to \infty$*, we have*

$$\|\mathsf{E}[\mathbf{x}_t] - \mathbf{x}_\infty\| \to 0, \quad \|\mathsf{E}[\mathbf{x}_t\mathbf{x}_t^\top] - \mathbf{\Sigma}_\infty\| \to 0. \tag{2}$$

In the noise-free case ($\mathbf{w}_t = 0$*), we have* $\mathbf{x}_\infty = 0$*,* $\mathbf{\Sigma}_\infty = 0$*. We say MJS in* (1) *with* $\mathbf{w}_t$=0 *is (mean-square) stabilizable if there exists mode-dependent controller* $\mathbf{K}_{1:s}$ *such that the closed-loop MJS* $\mathbf{x}_{t+1} = (\mathbf{A}_{\omega(t)} + \mathbf{B}_{\omega(t)}\mathbf{K}_{\omega(t)})\mathbf{x}_t$ *is MSS. We call such* $\mathbf{K}_{1:s}$ *a stabilizing controller.*

The (mean-square) stability of a noise-free autonomous MJS is related to the spectral radius of an augmented state matrix $\tilde{\mathbf{L}} \in \mathsf{R}^{sn^2 \times sn^2}$ with ij-th $n^2 \times n^2$ block given by $[\tilde{\mathbf{L}}]_{ij} := [\mathbf{T}]_{ji}\mathbf{L}_j \otimes \mathbf{L}_j$. Specifically, if $\rho(\tilde{\mathbf{L}}) < 1$, a noise-free autonomous MJS can be shown to satisfy MSS (Costa et al., 2006).

Assumption A1. *The MJS in* (1) *is stabilizable, and its underlying Markov chain (*$\mathbf{T}$*) is ergodic.*

Stabilizability allows us to use a mixing argument to obtain weakly dependent sub-trajectories by properly subsampling the original trajectory, and ergodicity guarantees that the Markov chain converges to a unique stationary distribution. Throughout, π_∞ denotes the stationary distribution of $\mathbf{T}$ with $\pi_{\min} := \min_i \pi_\infty(i)$. We further define the mixing time (Levin & Peres, 2017) of $\mathbf{T}$ as $t_{\mathsf{MC}} := \inf\left\{t \in \mathsf{N} : \max_{i \in [s]} \|([\mathbf{T}^t]_{i,:})^\top - \pi_\infty\|_1 \leq 0.5\right\}$, where $[\mathbf{T}^t]_{i,:}$ denotes the i_{th} row of $\mathbf{T}^t$. Note that t_{MC} plays a key role in the mixing time of the overall MJS. In the analysis, $\pi_{\min}$ guarantees one could obtain enough data for each mode, while the mixing time t_{MC} of the MJS determines the fraction of the data that provably helps towards learning the system.

2.2. Problem Formulation

In this paper, 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 and the mode observation, when one has the flexibility to design the input so that the collected data has nice statistical properties. In the MJS setting, one needs to estimate both the state/input matrices $\mathbf{A}_{1:s}, \mathbf{B}_{1:s}$ for every mode as well as the Markov matrix $\mathbf{T}$. In this work, we seek to estimate the MJS dynamics using only a single trajectory $\{\mathbf{x}_t, \mathbf{u}_t, \omega(t)\}_{t=0}^T$ and provide finite sample guarantees. Section 3 presents our system identification results. **(B) Online Linear Quadratic Regulator.** In this paper, we consider the following finite-horizon Markov jump system linear quadratic regulator (MJS-LQR) problem:

$$\begin{aligned} \inf_{\mathbf{u}_{0:T}} \quad & J(\mathbf{u}_{0:T}) := \sum_{t=0}^{T} \mathsf{E}\left[\mathbf{x}_t^\top \mathbf{Q}_{\omega(t)}\mathbf{x}_t + \mathbf{u}_t^\top \mathbf{R}_{\omega(t)}\mathbf{u}_t\right] \\ \text{s.t.} \quad & \mathbf{x}_t, \omega(t) \sim \text{MJS}(\mathbf{A}_{1:s}, \mathbf{B}_{1:s}, \mathbf{T}). \end{aligned} \tag{3}$$

The goal is to design inputs to minimize the expected quadratic cost composed of positive semi-definite matrices $\mathbf{Q}_{1:s}$ and $\mathbf{R}_{1:s}$ under the MJS dynamics. We will use MJS-LQR$(\mathbf{A}_{1:s}, \mathbf{B}_{1:s}, \mathbf{T}, \mathbf{Q}_{1:s}, \mathbf{R}_{1:s})$ to denote MJS-LQR problem (3). We assume the following for cost matrices.

Assumption A2. *For all* $i \in [s]$*, (a)* $\mathbf{R}_i \succ 0$*, (b)* $\mathbf{Q}_i \succ 0$*.*

We assume the state $\mathbf{x}_t$ and mode $\omega(t)$ can be observed at time t. With these observations, instead of a fixed and open-loop input sequence, one can design closed-loop policies that generate real-time input based on current observations, e.g. mode-dependent state-feedback controllers. When the dynamics $\mathbf{A}_{1:s}, \mathbf{B}_{1:s}, \mathbf{T}$ of the MJS are known, one can solve for the optimal controllers recursively via coupled discrete Riccati equations (Costa et al., 2006). In our work, we assume the dynamics are unknown, and only the design parameters $\mathbf{Q}_{1:s}$ and $\mathbf{R}_{1:s}$ 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 infinite-horizon cost $T \cdot J^\star$ where $J^\star$ is the optimal infinite-horizon average cost,

$$J^\star := \limsup_{T \to \infty} \frac{1}{T} \inf_{\mathbf{u}_{0:T}} J(\mathbf{u}_{0:T}), \tag{4}$$

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 (Dean et al., 2018), in MJS-LQR setting, the analysis requires additional consideration of Markov chain mixing, which is addressed in this paper.

3. System Identification for MJS

Our MJS identification procedure is given in Algorithm 1. We assume one has access to a stabilizing controller $\mathbf{K}_{1:s}$, which is a standard assumption in data-driven control (Dean et al., 2018). Note that if the open-loop MJS is already MSS, then one can simply set $\mathbf{K}_{1:s}^{(0)} = 0$. The following theorem gives our main results on learning the dynamics of an unknown MJS from finite samples obtained from a single trajectory.

Algorithm 1 MJS-SYSID

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

Set subsampling period $L = C_{\text{sub}} \log(T)$

Set subsampling indices $\tau_k = kL$ for $k = 1, 2, \dots, \lfloor T/L \rfloor$

Estimate $\mathbf{A}_{1:s}, \mathbf{B}_{1:s}$: **for all** modes $i \in [s]$ **do**: $S_i = \{\tau_k \mid \omega(\tau_k) = i,\ \|\mathbf{x}_{\tau_k}\| \leq c_{\mathbf{x}}\sigma_{\mathbf{w}}\sqrt{\log(T)},\ \|\mathbf{z}_{\tau_k}\| \leq c_{\mathbf{z}}\sigma_{\mathbf{z}}\}$,

$$\hat{\mathbf{\Theta}}_{1,i}, \hat{\mathbf{\Theta}}_{2,i} = \arg\min_{\mathbf{\Theta}_1, \mathbf{\Theta}_2} \sum_{k \in S_i} \|\mathbf{x}_{k+1} - \mathbf{\Theta}_1 \mathbf{x}_k/\sigma_{\mathbf{w}} - \mathbf{\Theta}_2 \mathbf{z}_k/\sigma_{\mathbf{z}}\|^2,$$

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

Estimate T: $$[\hat{\mathbf{T}}]_{ji} = \frac{\sum_{k=1}^{\lfloor T/L \rfloor} \mathbf{1}_{\{\omega(\tau_k)=i, \omega(\tau_k - 1)=j\}}}{\sum_{k=1}^{\lfloor T/L \rfloor} \mathbf{1}_{\{\omega(\tau_k - 1)=j\}}},$$

Output: $\hat{\mathbf{A}}_{1:s}, \hat{\mathbf{B}}_{1:s}, \hat{\mathbf{T}}$.

Theorem 1 (Identification of MJS)**.** *Suppose we run Algorithm 1 with $c_{\mathbf{x}} = \mathcal{O}(\sqrt{n})$ and $c_{\mathbf{z}} = \mathcal{O}(\sqrt{p})$. Let $\rho = \rho(\tilde{\mathbf{L}})$, where $\tilde{\mathbf{L}}$ is the augmented state matrix of the closed-loop MJS. Suppose $\mathbf{w}_t \overset{i.i.d.}{\sim} \mathcal{N}(0, \sigma_{\mathbf{w}}^2 \mathbf{I}_n)$. Suppose the trajectory length $T \geq \tilde{\mathcal{O}}(\log^2(T)(n+p)/\pi_{\min})$ and the sampling factor satisfies $C_{sub} \geq t_{\mathsf{MC}} \cdot \mathcal{O}(1/(1-\rho))$. Then, under Assumption A1, with probability at least $1-\delta$, for all $i \in [s]$, we have*

$$\max\left\{\begin{matrix}\|\hat{\mathbf{A}}_i - \mathbf{A}_i\|, \\ \|\hat{\mathbf{B}}_i - \mathbf{B}_i\|\end{matrix}\right\} \leq \tilde{\mathcal{O}}\left(\frac{\sigma_{\mathbf{z}} + \sigma_{\mathbf{w}}}{\sigma_{\mathbf{z}}\pi_{\min}} \frac{(n+p)\log(T)}{\sqrt{T}}\right),$$

$$\text{and } \|\hat{\mathbf{T}} - \mathbf{T}\|_{\infty} \leq \tilde{\mathcal{O}}\left(\frac{1}{\pi_{\min}}\sqrt{\frac{\log(T)}{T}}\right). \tag{5}$$

Corollary 1. *Consider the setting of Algorithm 1. When $\mathbf{B}_{1:s}$ are known, setting $\sigma_{\mathbf{z}} = 0$ and solving only for the state matrices leads to a stronger upper bound $\|\hat{\mathbf{A}}_i - \mathbf{A}_i\| \leq \tilde{\mathcal{O}}(\frac{(n+p)\log(T)}{\pi_{\min}\sqrt{T}})$.*

Our system identification result achieves optimal statistical error rate of $\tilde{O}(1/\sqrt{T})$. The sample complexity grows quadratically in state dimension n, which can potentially be improved to linear via a more refined control on the state-covariance (see (Simchowitz et al., 2018; Dean et al., 2019) for standard linear systems). It also grows with the inverse of the minimum mode frequency as $\pi_{\min}^{-1}$. Note that, $\pi_{\min}$ dictates the trajectory fraction of the least-frequent mode, thus, $\pi_{\min}^{-1}$ multiplier is not avoidable. In Corollary 1, we show that, when the knowledge of $\mathbf{B}$ is assumed, $\mathbf{A}$ can be estimated regardless of the exploration strength $\sigma_{\mathbf{z}}$. This is because the excitation for the state matrix arises from $\mathbf{w}_t$.

Proof outline for Theorem 1: Our proof strategy for Algorithm 1 addresses the challenges introduced by MJS and mean-square stability. We only emphasize the core technical challenges. In Algorithm 1, we subsample the trajectory. At a high-level, this will help us upper/lower bound the empirical covariance matrix formed by the subsampled state-input pairs $(\mathbf{x}_{\tau_k}, \mathbf{z}_{\tau_k})$ for all $\tau_k \in S_i$. Initial subsampling (with spacing L) aims to reduce the statistical dependence across the input data $(\mathbf{x}_t, \mathbf{z}_t)_{t \geq 0}$ to obtain a weakly-dependent sub-trajectory with indices τ_k. This dependence is due to the mode sequence $\omega(t)$ – unique to the MJS setting – and the system's memory (contribution of the earlier states on the current state). Thus L is primarily a function of the mixing-time of $\mathbf{T}$ and the spectral radius of the MJS system. Unlike related works on sysid and regret analysis (Simchowitz et al., 2018; Dean et al., 2018; Lale et al., 2020a; Oymak & Ozay, 2019; Lale et al., 2020b), mean-square stability does not lead to strong high-probability bounds, as one can only bound $\|\mathbf{x}_t\|$ or $\mathbf{x}_t\mathbf{x}_t^\intercal$ in the expectation sense. The second subsampling restricts our attention to the *bounded* $(\mathbf{x}_{\tau_k}, \mathbf{z}_{\tau_k})$ pairs on mode i. This boundedness enables us to control the covariance matrix despite MSS and potentially heavy-tailed states via non-asymptotic toolset (e.g. Thm 5.44 of (Vershynin, 2010)). However, heavy-tailed empirical covariance lower bounds require independence and our subsampled data are only "approximately independent" (coupled over modes and history). To make matters worse, the fact that we sample bounded states introduces further dependencies. To resolve this, we introduce a novel strategy to construct 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 to deal with actual (non-truncated) states. The fnal igredient is showing that, for each mode $1 \leq i \leq s$, with high probability, this carefully-crafted subset contains enough samples to ensure a well-conditioned covariance (with excitation provided by $\mathbf{z}_t, \mathbf{w}_t$). With this in place, after two rounds of subsampling, least-squares will accurately estimate $\mathbf{A}$ and $\mathbf{B}$ for all modes with rate $1/\sqrt{T}$.

4. Adaptive Control for MJS-LQR

Our adaptive MJS-LQR 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 the newly collected trajectory.

Similar to the discussion in Section 3, we assume at the beginning one has access to a stabilizing controller $\mathbf{K}_{1:s}^{(0)}$. During epoch i, controller $\mathbf{K}_{1:s}^{(i)}$ is used together with additive exploration noise $\mathbf{z}_t$ to boost learning. At the end of epoch i, the trajectory during this epoch is used to obtain a new MJS dynamics estimate $\mathbf{A}_{1:s}^{(i)}, \mathbf{B}_{1:s}^{(i)}, \mathbf{T}^{(i)}$ through Algorithm 1. Then, we set the controller $\mathbf{K}_{1:s}^{(i+1)}$ for epoch $i+1$ to be the optimal controller for the infinite-horizon MJS-LQR($\mathbf{A}_{1:s}^{(i)}, \mathbf{B}_{1:s}^{(i)}, \mathbf{T}^{(i)}, \mathbf{Q}_{1:s}, \mathbf{R}_{1:s}$), which can be solved

Algorithm 2 Adaptive MJS-LQR

Input: Initial epoch length T_0, initial stabilizing controller $\mathbf{K}_{1:s}^{(0)}$, epoch incremental ratio $\gamma > 1$, data bound $c_\mathbf{x}$, $c_\mathbf{z}$, subsampling factor C_{sub}.
for $i = 0, 1, 2, \dots$ **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 MJS for T_i steps with $\mathbf{u}_t^{(i)} = \mathbf{K}_{\omega(t)}^{(i)} \mathbf{x}_t^{(i)} + \mathbf{z}_t^{(i)}$ with $\mathbf{z}_t^{(i)} \overset{\text{i.i.d.}}{\sim} \mathcal{N}(0, \sigma_{\mathbf{z},i}^2 \mathbf{I}_p)$ and record the trajectory $\mathcal{D}_i := \{\mathbf{x}_t^{(i)}, \mathbf{z}_t^{(i)}, \omega^{(i)}(t)\}_{t=0}^{T_i}$.
$\hat{\mathbf{A}}_{1:s}^{(i)}, \hat{\mathbf{B}}_{1:s}^{(i)}, \hat{\mathbf{T}}^{(i)}$
$=$ MJS-SYSID($\mathbf{K}_{1:s}^{(i)}, \sigma_\mathbf{w}^2, \sigma_{\mathbf{z},i}^2, \mathcal{D}_i, c_\mathbf{x}, c_\mathbf{z}, C_{sub}$)
$\mathbf{K}_{1:s}^{(i+1)} \leftarrow$ optimal controller for the infinite-horizon MJS-LQR($\hat{\mathbf{A}}_{1:s}^{(i)}, \hat{\mathbf{B}}_{1:s}^{(i)}, \hat{\mathbf{T}}^{(i)}, \mathbf{Q}_{1:s}, \mathbf{R}_{1:s}$).
end for

efficiently via value iteration or via LMIs (Costa et al., 2006). Note that this control design based on the estimated dynamics is also referred to as certainty equivalent control.

To have theoretically guaranteed performance, i.e. sublinear regret, the key is to have a subtle scheduling of epoch lengths T_i and exploration noise variance $\sigma_{\mathbf{z},i}^2$. We choose T_i to increase exponentially with rate $\gamma > 1$, and set $\sigma_{\mathbf{z},i}^2 = \sigma_\mathbf{w}^2/\sqrt{T_i}$, which collectively guarantee $\mathcal{O}(\log(T)\sqrt{T})$ regret when combined with the system identification result from Theorem 1. Intuitively, this scheduling has interpretations from two folds: (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 exploration noise and deploy (exploit) the controller for a longer time. Note that the scheduling rate γ has similar role to the discount factor in reinforcement learning: smaller γ aims to reduce short-term cost while larger γ aims to reduce long-term cost.

4.1. Regret Analysis

We define filtration $\mathcal{F}_{-1}, \mathcal{F}_0, \mathcal{F}_1, \dots$ such that $\mathcal{F}_{-1} := \sigma(\mathbf{x}_0, \omega(0))$ is the sigma algebra generated by the initial state and mode, and $\mathcal{F}_i := \sigma(\mathbf{x}_0, \omega(0), \{\{\omega^{(j)}(t)\}_{t=1}^{T_j}\}_{j=0}^{i}, \mathbf{w}_0, \{\mathbf{w}_{1:T_j}^{(j)}\}_{j=0}^{i}, \mathbf{z}_0, \{\mathbf{z}_{1:T_j}^{(j)}\}_{j=0}^{i})$ is the sigma algebra generated by the randomness up to epoch i. Note that since the initial state $\mathbf{x}_0^{(i)}$ of epoch i is 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 $\omega(0)^{(i+1)}$. Suppose time step t belongs to epoch i, then we define the following conditional expected cost at time t. $c_t = \mathbb{E}[\mathbf{x}_t^\intercal \mathbf{Q}_{\omega(t)} \mathbf{x}_t + \mathbf{u}_t^\intercal \mathbf{R}_{\omega(t)} \mathbf{u}_t \mid \mathcal{F}_{i-1}]$, and cumulative cost as $J_T = \sum_{t=1}^{T} c_t$. We define the total regret and epoch-i regret as

$$\text{Regret}(T) = J_T - TJ^\star,$$

$$\text{Regret}_i = (\sum_{t=1}^{T_i} c_{T_0 + \dots + T_{i-1} + t}) - T_i J^\star. \tag{6}$$

One can refer to Appendix C.4 for more discussion on the regret definition. With these definitions, we have the following result.

Theorem 2 (Sub-linear regret). *If T_0, C_{sub}, $c_\mathbf{x}$, and $c_\mathbf{z}$ are large enough, then under Assumption A1 and A2, with probability at least $1-\delta$, Algorithm 2 achieves*

$$\text{Regret}(T) \le \hat{\mathcal{O}}(\log(T)) + \tilde{\mathcal{O}}(\log^2(T)\sqrt{T}). \tag{7}$$

From Corollary 1, we know when $\mathbf{B}_{1:s}$ are known, no further exploration noise is needed to learn $\mathbf{A}_{1:s}$ or $\mathbf{T}$, this applies to the adaptive MJS-LQR setting as well. Getting rid of exploration noise improves the regret as follows.

Corollary 2 (Poly-log regret). *When $\mathbf{B}_{1:s}$ are known, it suffices to set $\sigma_{\mathbf{z},i} = 0$ for all i in Algorithm 2. Then, Algorithm 2 achieves* $\text{Regret}(T) \le \hat{\mathcal{O}}(\log(T)) + \tilde{\mathcal{O}}(\log^3(T))$.

Proof outline for Theorem 2: For simplicity, we only show the dominant $\mathcal{O}(\log^2(T)\sqrt{T})$ term. Define the estimation error after epoch i as $\epsilon_{\mathbf{A},\mathbf{B}}^{(i)} := \max_{j \in [s]} \max\{\|\hat{\mathbf{A}}_j^{(i)} - \mathbf{A}_j\|, \|\hat{\mathbf{B}}_j^{(i)} - \mathbf{B}_j\|\}$, $\epsilon_\mathbf{T}^{(i)} := \|\hat{\mathbf{T}}^{(i)} - \mathbf{T}\|_\infty$. Using perturbation result (Du et al., 2021) for infinite-horizon MJS-LQR together with new finite-horizon cost analysis, we can bound epoch-i regret as follows: $\text{Regret}_i \le \mathcal{O}\left(T_i \sigma_{\mathbf{z},i}^2 + T_i \sigma_\mathbf{w}^2 \left(\epsilon_{\mathbf{A},\mathbf{B}}^{(i-1)} + \epsilon_\mathbf{T}^{(i-1)}\right)^2\right)$. Next, plugging in the upper bounds on the estimation errors $\epsilon_{\mathbf{A},\mathbf{B}}^{(i)} \le \mathcal{O}\left(\frac{\sigma_{\mathbf{z},i} + \sigma_\mathbf{w}}{\sigma_{\mathbf{z},i}} \frac{\log(T_i)}{\sqrt{T_i}}\right)$, $\epsilon_\mathbf{T}^{(i)} \le \mathcal{O}\left(\sqrt{\frac{\log(T_i)}{T_i}}\right)$ from Theorem 1, and using the exploration variance $\sigma_{\mathbf{z},i}^2 = \frac{\sigma_\mathbf{w}^2}{\sqrt{T_i}}$, we have $\text{Regret}_i \le \mathcal{O}(\sigma_\mathbf{w}^2 \sqrt{\gamma} \sqrt{T_i} \log^2(T_i))$. Finally, we have: $\text{Regret}(T) = \sum_{i=1}^{O(\log_\gamma(\frac{T}{T_0}))} \text{Regret}_i \le \mathcal{O}\left(\sigma_\mathbf{w}^2 \log(\frac{T}{T_0}) \sqrt{\frac{T}{T_0}} \left(\frac{\sqrt{\gamma}}{\sqrt{\gamma}-1}\right)^3 \left(\sqrt{\gamma} \log(\frac{T}{T_0}) - \log(\sqrt{\gamma})\right)\right) = \mathcal{O}(\log^2(T)\sqrt{T})$.

5. 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 poly-logarithmic factors.

Acknowledgements

Y. Sattar and S. Oymak were supported in part by NSF under grant CNS-1932254. 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. L. Balzano was supported in part by NSF CAREER award CCF-1845076, NSF BIGDATA award IIS1838179, ARO YIP award W911NF1910027, and the Institute for Advanced Study Charles Simonyi Endowment.

References

Abbasi-Yadkori, Y. and Szepesvári, C. Regret bounds for the adaptive control of linear quadratic systems. In *Proc. of COLT*, pp. 1–26. JMLR Workshop and Conference Proceedings, 2011.

Abeille, M. and Lazaric, A. Efficient optimistic exploration in linear-quadratic regulators via lagrangian relaxation. In *ICML*, pp. 23–31. PMLR, 2020.

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

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

Cassel, A., Cohen, A., and Koren, T. Logarithmic regret for learning linear quadratic regulators efficiently. In *International Conference on Machine Learning*, pp. 1328–1337. PMLR, 2020.

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

Cohen, A., Hasidim, A., Koren, T., Lazic, N., Mansour, Y., and Talwar, K. Online linear quadratic control. In *International Conference on Machine Learning*, pp. 1029–1038. PMLR, 2018.

Costa, O. L. V., Fragoso, M. D., and Marques, R. P. *Discrete-time Markov jump linear systems*. Springer, 2006.

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

Dean, S., Mania, H., Matni, N., Recht, B., and Tu, S. On the sample complexity of the linear quadratic regulator. *FOCM*, pp. 1–47, 2019.

Du, Z., Ozay, N., and Balzano, L. Mode clustering for markov jump systems. *arXiv preprint arXiv:1910.02193*, 2019.

Du, Z., Sattar, Y., Tarzanagh, D. A., Balzano, L., Oymak, S., and Ozay, N. Certainty equivalent quadratic control for markov jump systems. *arXiv e-prints*, 2021.

Freedman, D. A. On tail probabilities for martingales. *the Annals of Probability*, pp. 100–118, 1975.

Goel, G. and Hassibi, B. The power of linear controllers in lqr control. *arXiv preprint arXiv:2002.02574*, 2020.

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

Lale, S., Azizzadenesheli, K., Hassibi, B., and Anandkumar, A. Explore more and improve regret in linear quadratic regulators. *arXiv preprint arXiv:2007.12291*, 2020a.

Lale, S., Azizzadenesheli, K., Hassibi, B., and Anandkumar, A. Logarithmic regret bound in partially observable linear dynamical systems. *arXiv preprint arXiv:2003.11227*, 2020b.

Levin, D. A. and Peres, Y. *Markov chains and mixing times*, volume 107. American Mathematical Soc., 2017.

Mania, H., Tu, S., and Recht, B. Certainty equivalence is efficient for linear quadratic control. In *NeurIPS*, 2019.

Oymak, S. Stochastic gradient descent learns state equations with nonlinear activations. In *Conference on Learning Theory*, pp. 2551–2579, 2019.

Oymak, S. and Ozay, N. Non-asymptotic identification of lti systems from a single trajectory. *American Control Conference*, 2019.

Sarkar, T. and Rakhlin, A. Near optimal finite time identification of arbitrary linear dynamical systems. In *ICML*, pp. 5610–5618. PMLR, 2019.

Sarkar, T., Rakhlin, A., and Dahleh, M. Nonparametric system identification of stochastic switched linear systems. In *2019 IEEE 58th Conference on Decision and Control (CDC)*, pp. 3623–3628. IEEE, 2019.

Simchowitz, M., Mania, H., Tu, S., Jordan, M. I., and Recht, B. Learning without mixing: Towards a sharp analysis of linear system identification. *arXiv preprint arXiv:1802.08334*, 2018.

Tu, S., Boczar, R., Packard, A., and Recht, B. Non-asymptotic analysis of robust control from coarse-grained identification. *arXiv preprint arXiv:1707.04791*, 2017.

Vershynin, R. Introduction to the non-asymptotic analysis of random matrices. *arXiv preprint arXiv:1011.3027*, 2010.

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

Zhang, A. and Wang, M. State compression of markov processes via empirical low-rank estimation. *arXiv preprint arXiv:1802.02920*, 2018.

A. Preliminaries

We use boldface uppercase (lowercase) letters to denote matrices (vectors). For a matrix $\mathbf{V}$, $\rho(\mathbf{V})$, $\underline{\sigma}(\mathbf{V})$, $\|\mathbf{V}\|$, $\|\mathbf{V}\|_1$, $\|\mathbf{V}\|_F$ denote its spectral radius, smallest singular value, spectral norm, ℓ_1 norm, and Frobenius norm, respectively. We use $\mathbf{vec}(\mathbf{V})$ to denote the vectorization of a matrix $\mathbf{V}$. We let $\|\mathbf{V}\|_+ := \|\mathbf{V}\| + 1$. The Kronecker product of two matrices $\mathbf{M}$ and $\mathbf{N}$ is denoted as $\mathbf{M} \otimes \mathbf{N}$. $\mathbf{V}_{1:s}$ denotes a set of s matrices $\{\mathbf{V}_i\}_{i=1}^s$ of same dimensions. We define $[s] := \{1, 2, \dots, s\}$, $\underline{\sigma}(\mathbf{V}_{1:s}) := \min_{i \in [s]} \underline{\sigma}(\mathbf{V}_i)$, $\|\mathbf{V}_{1:s}\| := \max_{i \in [s]} \|\mathbf{V}_i\|$, and $\|\mathbf{V}_{1:s}\|_+ := \max_{i \in [s]} \|\mathbf{V}_i\|_+$. $\mathbf{I}_n$ denotes the identity matrix with dimension n, and $\mathbf{1}_n$ denotes the all 1 vector with dimension n. $\mathbf{1}_{\{\}}$ denotes the indicator function. We use $\lesssim$ and $\gtrsim$ for inequalities that hold up to a constant factor.

We define the following to quantify the decay of a square matrix $\mathbf{M}$.

Definition 2. *For a square matrix $\mathbf{M}$ with $\rho(\mathbf{M}) \leq 1$, let*

$$\tau(\mathbf{M}) := \sup_{k \in \mathbb{N}} \{\|(\mathbf{M})^k\| / \rho(\mathbf{M})^k\}. \tag{8}$$

Note that $\tau(\mathbf{M})$ is finite by Gelfand's formula, and it is easy to see that $\tau \geq 1$. This quantity measures the transient response of a non-switching system with state matrix $\mathbf{M}$ and can be upper bounded by its $\mathcal{H}_\infty$ norm (Tu et al., 2017). In this work, we will mainly use this to evaluate the augmented state matrix for an MJS defined in Sec 2.1.

For a Markov chain with transition matrix $\mathbf{T}$, we let $\pi_0 \in \mathbb{R}^s$ denote the initial state distribution and π_t denote the transient state distribution, i.e. $\mathbb{P}(\omega(t) = i) = \pi_t(i)$. Then, it is easy to see $\pi_t = \mathbf{T}^t \pi_0$. By triangle inequality, we have $\|\pi_t - \pi_\infty\|_1 \leq \max_{i \in [s]} \|([\mathbf{T}^t]_{i,:})^\top - \pi_\infty\|_1$. Thus, for the ergodic Markov matrix $\mathbf{T}$, we define the following to quantify the convergence of $\|\pi_t - \pi_\infty\|_1$.

Definition 3. *For an ergodic Markov matrix $\mathbf{T} \in \mathbb{R}^{s \times s}$, let $\tau_{MC} > 0$ and $\rho_{MC} \in [0, 1)$ be two constants (Levin & Peres, 2017, Theorem 4.9) such that*

$$\max_{i \in [s]} \|([\mathbf{T}^t]_{i,:})^\top - \pi_\infty\|_1 \leq \tau_{MC} \rho_{MC}^t. \tag{9}$$

Furthermore, define

$$t_{MC}(\epsilon) := \min \left\{ t \in \mathbb{N} : \max_{i \in [n]} \frac{1}{2} \|([\mathbf{T}^t]_{i,:})^\top - \pi_\infty\|_1 \leq \epsilon \right\}. \tag{10}$$

When parameter ϵ is omitted, it denotes $t_{MC} := t_{MC}(\frac{1}{4})$, i.e. the mixing time defined in Sec 2.1.

Note that $\tau(\mathbf{M})$ and τ_{MC} have similar roles except $\tau(\mathbf{M})$ is usually used to study state matrices while τ_{MC} is for Markov matrices. And when $\mathbf{T}$ is ergodic, $\tau(\mathbf{T} - \mathbf{I})$ and τ_{MC} are equivalent.

A.1. MJS Covariance Dynamics Under MSS

Consider MJS$(\mathbf{A}_{1:s}, \mathbf{B}_{1:s}, \mathbf{T})$ with process noise $\mathbf{w}_t \sim \mathcal{N}(0, \mathbf{\Sigma}_\mathbf{w})$ and input $\mathbf{u}_t = \mathbf{K}_{\omega(t)} \mathbf{x}_t + \mathbf{z}_t$ under a stabilizing controller $\mathbf{K}_{1:s}$ and noise $\mathbf{z}_t \sim \mathcal{N}(0, \mathbf{\Sigma}_\mathbf{z})$. Let $\mathbf{L}_i := \mathbf{A}_i + \mathbf{B}_i \mathbf{K}_i$. Let $\tilde{\mathbf{L}} \in \mathbb{R}^{sn^2 \times sn^2}$ denote the augmented closed-loop state matrix with ij-th $n^2 \times n^2$ block given by $[\tilde{\mathbf{L}}]_{ij} := [\mathbf{T}]_{ji} \mathbf{L}_j \otimes \mathbf{L}_j$. Let $\tau_{\tilde{\mathbf{L}}} > 0, \rho_{\tilde{\mathbf{L}}} \in [0, 1)$ be two constants such that $\|\tilde{\mathbf{L}}^k\| \leq \tau_{\tilde{\mathbf{L}}} \rho_{\tilde{\mathbf{L}}}^k$. By definition, one available choice for $\tau_{\tilde{\mathbf{L}}}$ and $\rho_{\tilde{\mathbf{L}}}$ are $\tau(\tilde{\mathbf{L}})$ and $\rho(\tilde{\mathbf{L}})$. Let $\mathbf{\Sigma}_i(t) := \mathbb{E}[\mathbf{x}_t \mathbf{x}_t^\top \mathbf{1}_{\{\omega(t) = i\}}]$, $\mathbf{\Sigma}(t) := \mathbb{E}[\mathbf{x}_t \mathbf{x}_t^\top]$, and

$$\mathbf{s}_t := \begin{bmatrix} \mathbf{vec}(\mathbf{\Sigma}_1(t)) \\ \vdots \\ \mathbf{vec}(\mathbf{\Sigma}_s(t)) \end{bmatrix}, \; \tilde{\mathbf{B}}_t := \begin{bmatrix} \sum_{j=1}^s \pi_{t-1}(j) \mathbf{T}_{j1} (\mathbf{B}_j \otimes \mathbf{B}_j) \\ \vdots \\ \sum_{j=1}^s \pi_{t-1}(j) \mathbf{T}_{js} (\mathbf{B}_j \otimes \mathbf{B}_j) \end{bmatrix}, \; \tilde{\mathbf{\Pi}}_t := \pi_t \otimes \mathbf{I}_{n^2}. \tag{11}$$

The following lemma shows how $\mathbf{s}_t$ depends on $\mathbf{s}_0$, $\mathbf{\Sigma}_\mathbf{z}$, and $\mathbf{\Sigma}_\mathbf{w}$, which is later used to upper bound $\mathbb{E}[\|\mathbf{x}_t\|^2]$ in Lemma 2.

Lemma 1. *The vectorized covariance $\mathbf{s}_t$ has the following dynamics*

$$\mathbf{s}_t = \tilde{\mathbf{L}}^t \mathbf{s}_0 + (\tilde{\mathbf{B}}_t + \tilde{\mathbf{L}} \tilde{\mathbf{B}}_{t-1} + \cdots + \tilde{\mathbf{L}}^{t-1} \tilde{\mathbf{B}}_1) \mathbf{vec}(\mathbf{\Sigma}_\mathbf{z}) + (\tilde{\mathbf{\Pi}}_t + \tilde{\mathbf{L}} \tilde{\mathbf{\Pi}}_{t-1} + \cdots + \tilde{\mathbf{L}}^{t-1} \tilde{\mathbf{\Pi}}_1) \mathbf{vec}(\mathbf{\Sigma}_\mathbf{w}). \tag{12}$$

Proof. To evaluate $\boldsymbol{\Sigma}_i(t)$, from the equivalent MJS dynamics $\mathbf{x}_{t+1} = \mathbf{L}_{\omega(t)}\mathbf{x}_t + \mathbf{B}_{\omega(t)}\mathbf{z}_t + \mathbf{w}_t$, we have

$$\begin{aligned}\mathsf{E}[\mathbf{x}_{t+1}\mathbf{x}_{t+1}^\intercal \mathbf{1}_{\{\omega(t+1)=i\}}] = &\sum_{j=1}^{s} \mathsf{E}[\mathbf{L}_j \mathbf{x}_t \mathbf{x}_t^\intercal \mathbf{L}_j \mathbf{1}_{\{\omega(t+1)=i, \omega(t)=j\}}] \\ &+ \sum_{j=1}^{s} \mathsf{E}[\mathbf{B}_j \mathbf{z}_t \mathbf{z}_t^\intercal \mathbf{B}_j^\intercal \mathbf{1}_{\{\omega(t+1)=i, \omega(t)=j\}}] + \mathsf{E}[\mathbf{w}_t \mathbf{w}_t^\intercal \mathbf{1}_{\{\omega(t+1)=i\}}].\end{aligned} \tag{13}$$

Since $\mathbf{w}_t \sim \mathcal{N}(0, \boldsymbol{\Sigma}_{\mathbf{w}})$ and $\mathbf{z}_t \sim \mathcal{N}(0, \boldsymbol{\Sigma}_{\mathbf{z}})$, we get

$$\boldsymbol{\Sigma}_i(t+1) = \sum_{j=1}^{s} \mathbf{T}_{ji} \mathbf{L}_j \boldsymbol{\Sigma}_j(t) \mathbf{L}_j^\intercal + \sum_{j=1}^{s} \pi_t(j) \mathbf{T}_{ji} \mathbf{B}_j \boldsymbol{\Sigma}_{\mathbf{z}} \mathbf{B}_j^\intercal + \pi_{t+1}(i) \boldsymbol{\Sigma}_{\mathbf{w}}. \tag{14}$$

Vectorizing both sides, we have

$$\begin{aligned}\mathbf{vec}(\boldsymbol{\Sigma}_i(t+1)) = \sum_{j=1}^{s} \mathbf{T}_{ji} (\mathbf{L}_j \otimes \mathbf{L}_j) \mathbf{vec}(\boldsymbol{\Sigma}_j(t)) + \sum_{j=1}^{s} \pi_t(j) \mathbf{T}_{ji} (\mathbf{B}_j \otimes \mathbf{B}_j) \mathbf{vec}(\boldsymbol{\Sigma}_{\mathbf{z}}) \\ + \pi_{t+1}(i) \mathbf{vec}(\boldsymbol{\Sigma}_{\mathbf{w}}).\end{aligned} \tag{15}$$

Stacking this for every i, we have

$$\begin{bmatrix} \mathbf{vec}(\boldsymbol{\Sigma}_1(t+1)) \\ \vdots \\ \mathbf{vec}(\boldsymbol{\Sigma}_s(t+1)) \end{bmatrix} = \tilde{\mathbf{L}} \begin{bmatrix} \mathbf{vec}(\boldsymbol{\Sigma}_1(t)) \\ \vdots \\ \mathbf{vec}(\boldsymbol{\Sigma}_s(t)) \end{bmatrix} + \tilde{\mathbf{B}}_{t+1} \mathbf{vec}(\boldsymbol{\Sigma}_{\mathbf{z}}) + \tilde{\boldsymbol{\Pi}}_{t+1} \mathbf{vec}(\boldsymbol{\Sigma}_{\mathbf{w}}). \tag{16}$$

Propagate this dynamics from t to 0, we could conclude the proof. □

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

Lemma 2. *For* $\mathsf{E}[\|\mathbf{x}_t\|^2]$ *and* $\|\boldsymbol{\Sigma}(t)\|_F$*, we have*

$$\mathsf{E}[\|\mathbf{x}_t\|^2] \le \sqrt{ns} \cdot \tau_{\tilde{\mathbf{L}}} \rho_{\tilde{\mathbf{L}}}^t \cdot \mathsf{E}[\|\mathbf{x}_0\|^2] + n\sqrt{s}(\|\mathbf{B}_{1:s}\|^2 \|\boldsymbol{\Sigma}_{\mathbf{z}}\| + \|\boldsymbol{\Sigma}_{\mathbf{w}}\|) \frac{\tau_{\tilde{\mathbf{L}}}}{1-\rho_{\tilde{\mathbf{L}}}}, \tag{17}$$

$$\|\boldsymbol{\Sigma}(t)\|_F \le \sqrt{s} \cdot \tau_{\tilde{\mathbf{L}}} \rho_{\tilde{\mathbf{L}}}^t \cdot \mathsf{E}[\|\mathbf{x}_0\|^2] + \sqrt{ns}(\|\mathbf{B}_{1:s}\|^2 \|\boldsymbol{\Sigma}_{\mathbf{z}}\| + \|\boldsymbol{\Sigma}_{\mathbf{w}}\|) \frac{\tau_{\tilde{\mathbf{L}}}}{1-\rho_{\tilde{\mathbf{L}}}}. \tag{18}$$

Proof. First we prove the bound for $\mathsf{E}[\|\mathbf{x}_t\|^2]$, and the bound for $\|\boldsymbol{\Sigma}(t)\|_F$ follows similarly. For state $\mathbf{x}_t$, we have $\mathsf{E}[\|\mathbf{x}_t\|^2] = \sum_{i=1}^{s} \mathsf{E}[\|\mathbf{x}_t\|^2 \mathbf{1}_{\{\omega(t)=i\}}] = \sum_{i=1}^{s} \mathbf{tr}\ (\mathsf{E}[\mathbf{x}_t \mathbf{x}_t^\intercal \mathbf{1}_{\{\omega(t)=i\}}]) = \sum_{i=1}^{s} \mathbf{tr}(\boldsymbol{\Sigma}_i(t)) = \sum_{i=1}^{s} \sum_{j=1}^{n} \lambda_j(\boldsymbol{\Sigma}_i(t)) \le \sqrt{ns}\sqrt{\sum_{i=1}^{s} \sum_{j=1}^{n} \lambda_j^2(\boldsymbol{\Sigma}_i(t))} \le \sqrt{ns}\sqrt{\sum_{i=1}^{s} \|\boldsymbol{\Sigma}_i(t)\|_F^2}$. Then, by definition of $\mathbf{s}_t$, we have

$$\mathsf{E}[\|\mathbf{x}_t\|^2] \le \sqrt{ns}\|\mathbf{s}_t\|. \tag{19}$$

Now, applying the dynamics of $\mathbf{s}_t$ in Lemma 1, we have

$$\begin{aligned}\mathsf{E}[\|\mathbf{x}_t\|^2] &\le \sqrt{ns}\Big(\|\tilde{\mathbf{L}}^t\|\|\mathbf{s}_0\| + \sum_{t'=1}^{t} \|\tilde{\mathbf{L}}^{t-t'}\|\|\tilde{\mathbf{B}}_{t'} \mathbf{vec}(\boldsymbol{\Sigma}_{\mathbf{z}})\| + \sum_{t'=1}^{t} \|\tilde{\mathbf{L}}^{t-t'}\|\|\tilde{\boldsymbol{\Pi}}_{t'} \mathbf{vec}(\boldsymbol{\Sigma}_{\mathbf{w}})\|\Big) \\ &\le \sqrt{ns}\tau_{\tilde{\mathbf{L}}}\Big(\rho_{\tilde{\mathbf{L}}}^t \|\mathbf{s}_0\| + \sum_{t'=1}^{t} \rho_{\tilde{\mathbf{L}}}^{t-t'} \|\tilde{\mathbf{B}}_{t'} \mathbf{vec}(\boldsymbol{\Sigma}_{\mathbf{z}})\| + \sum_{t'=1}^{t} \rho_{\tilde{\mathbf{L}}}^{t-t'} \|\tilde{\boldsymbol{\Pi}}_{t'} \mathbf{vec}(\boldsymbol{\Sigma}_{\mathbf{w}})\|\Big),\end{aligned} \tag{20}$$

where the second line follows from $\|\tilde{\mathbf{L}}^t\| \le \tau_{\tilde{\mathbf{L}}} \rho_{\tilde{\mathbf{L}}}^t$. Now we evaluate $\|\mathbf{s}_0\|$, $\|\tilde{\mathbf{B}}_{t'} \mathbf{vec}(\boldsymbol{\Sigma}_{\mathbf{z}})\|$, and $\|\tilde{\boldsymbol{\Pi}}_{t'} \mathbf{vec}(\boldsymbol{\Sigma}_{\mathbf{w}})\|$ separately.

$$\|\mathbf{s}_0\| = \sqrt{\sum_{i=1}^{s} \|\boldsymbol{\Sigma}_i(0)\|_F^2} = \sqrt{\sum_{i=1}^{s} \pi_i(0)^2 \|\mathsf{E}[\mathbf{x}_0 \mathbf{x}_0^\intercal]\|_F^2} \le \|\mathsf{E}[\mathbf{x}_0 \mathbf{x}_0^\intercal]\|_F \le \mathsf{E}[\|\mathbf{x}_0\|^2]. \tag{21}$$

Let $[\tilde{\mathbf{B}}_{t^0}]_i$ denote the ith block of $\tilde{\mathbf{B}}_{t^0}$, i.e. $[\tilde{\mathbf{B}}_{t^0}]_i = \sum_{j=1}^{s} \quad t \quad 1(j)\mathbf{T}_{ji}(\mathbf{B}_j \otimes \mathbf{B}_j)$, then

$$
\begin{aligned}
\|\tilde{\mathbf{B}}_{t^0}\mathbf{vec}(\boldsymbol{\Sigma}_\mathbf{z})\| = \sqrt{\sum_{i=1}^{s} \|[\tilde{\mathbf{B}}_{t^0}]_i\mathbf{vec}(\boldsymbol{\Sigma}_\mathbf{z})\|^2} &\leq \sum_{i=1}^{s} \|[\tilde{\mathbf{B}}_{t^0}]_i\mathbf{vec}(\boldsymbol{\Sigma}_\mathbf{z})\| \\
&= \sum_{i=1}^{s} \|\sum_{j=1}^{s} \quad t^0 \quad 1(j)\mathbf{T}_{ji}(\mathbf{B}_j \otimes \mathbf{B}_j)\mathbf{vec}(\boldsymbol{\Sigma}_\mathbf{z})\| \\
&= \sum_{i=1}^{s} \|\sum_{j=1}^{s} \quad t^0 \quad 1(j)\mathbf{T}_{ji}(\mathbf{B}_j\boldsymbol{\Sigma}_\mathbf{z}\mathbf{B}_j^{|})\|_F \\
&\leq \|\mathbf{B}_{1:s}\|^2\|\boldsymbol{\Sigma}_\mathbf{z}\| \cdot \sum_{i=1}^{s} \|\sum_{j=1}^{s} \quad t^0 \quad 1(j)\mathbf{T}_{ji}\mathbf{I}_n\|_F \\
&= \|\mathbf{B}_{1:s}\|^2\|\boldsymbol{\Sigma}_\mathbf{z}\| \cdot \sum_{i=1}^{s} \| \quad t^0(i)\mathbf{I}_n\|_F \\
&\leq \sqrt{n}\|\mathbf{B}_{1:s}\|^2\|\boldsymbol{\Sigma}_\mathbf{z}\|.
\end{aligned}
\tag{22}
$$

At last,

$$
\|\tilde{\boldsymbol{\Pi}}_{t^0}\mathbf{vec}(\boldsymbol{\Sigma}_\mathbf{w})\| = \sqrt{\sum_{i=1}^{s} \| \quad t^0(i)\mathbf{vec}(\boldsymbol{\Sigma}_\mathbf{w})\|^2} \leq \|\mathbf{vec}(\boldsymbol{\Sigma}_\mathbf{w})\| = \|\boldsymbol{\Sigma}_\mathbf{w}\|_F = \sqrt{n}\|\boldsymbol{\Sigma}_\mathbf{w}\|, \tag{23}
$$

Plugging (21), (22), and (23) into (20), we have

$$
\begin{aligned}
\mathsf{E}[\|\mathbf{x}_t\|^2] &\leq \sqrt{ns}\tau_{\tilde{\mathbf{L}}}\Big(\rho_{\tilde{\mathbf{L}}}^t \mathsf{E}[\|\mathbf{x}_0\|^2] + \sqrt{n}\|\mathbf{B}_{1:s}\|^2\|\boldsymbol{\Sigma}_\mathbf{z}\|\sum_{t^0=1}^{t}\rho_{\tilde{\mathbf{L}}}^{t \quad t^0} + \sqrt{n}\|\boldsymbol{\Sigma}_\mathbf{w}\|\sum_{t^0=1}^{t}\rho_{\tilde{\mathbf{L}}}^{t \quad t^0}\Big), \\
&\leq \sqrt{ns}\cdot\tau_{\tilde{\mathbf{L}}}\rho_{\tilde{\mathbf{L}}}^t \cdot \mathsf{E}[\|\mathbf{x}_0\|^2] + n\sqrt{s}(\|\mathbf{B}_{1:s}\|^2\|\boldsymbol{\Sigma}_\mathbf{z}\| + \|\boldsymbol{\Sigma}_\mathbf{w}\|)\frac{\tau_{\tilde{\mathbf{L}}}}{1-\rho_{\tilde{\mathbf{L}}}},
\end{aligned}
\tag{24}
$$

which proves the bound for $\mathsf{E}[\|\mathbf{x}_t\|^2]$ in (17). To prove the bound for $\|\boldsymbol{\Sigma}(t)\|_F$ in (18), note that $\|\boldsymbol{\Sigma}(t)\|_F = \|\sum_{i=1}^{s}\boldsymbol{\Sigma}_i(t)\|_F \leq \sqrt{s}\sqrt{\sum_{i=1}^{s}\|\boldsymbol{\Sigma}_i(t)\|_F^2}$ □

A.2. Notations

In this appendix, we define a few notations to ease the exposition in the appendix. Note that for notations under parameterized form, e.g. (δ, ρ, τ), it means that one can choose the parameters freely to get different deterministic quantities.

Table 1 introduces some notation and defines some constants about the shortest trajectory (initial epoch) length such that theoretical performance guarantees can be given. Recall $\mathbf{K}_{1:s}^{(0)}$ is the stabilizing controller for epoch 0 in Algorithm 2. We let $\mathbf{L}_i^{(0)} := \mathbf{A}_i + \mathbf{B}_i\mathbf{K}_i^{(0)}$ for all $i \in [s]$ denote its closed-loop matrix, and $\tilde{\mathbf{L}}^{(0)} \in \mathsf{R}^{sn^2 x sn^2}$ denote the augmented closed-loop state matrix with ij-th $n^2 \times n^2$ block given by $[\tilde{\mathbf{L}}^{(0)}]_{ij} := [\mathbf{T}]_{ji}\mathbf{L}_j^{(0)} \otimes \mathbf{L}_j^{(0)}$. $\tau(\cdot)$ is given in Definition 2, and $\rho(\cdot)$ denotes the spectral radius. For the infinite-horizon MJS-LQR($\mathbf{A}_{1:s}, \mathbf{B}_{1:s}, \mathbf{T}, \mathbf{Q}_{1:s}, \mathbf{R}_{1:s}$) problem, we let $\mathbf{P}_{1:s}^{?}$ denote the solution to the coupled discrete algebraic Riccati equations (Costa et al., 2006), and $\mathbf{K}_{1:s}^{?}$ denote the optimal controller, which can be computed with $\mathbf{P}_{1:s}^{?}$. Similarly, we define $\mathbf{L}_{1:s}^{?}$ and $\tilde{\mathbf{L}}^{?}$ as the corresponding closed-loop state matrix and augmented closed-loop state matrix, and $\rho^{?} := \rho(\tilde{\mathbf{L}}^{?})$. $\pi_{\max}$ and $\pi_{\min}$ are the largest and smallest elements in the stationary distribution of the ergodic Markov matrix $\mathbf{T}$. The constant $\bar{\epsilon}_{\mathbf{A},\mathbf{B},\mathbf{T}}^{5}$ is defined in Table 3.

Table 2 provides constants, and presents the associated notation, regarding the smallest values of the tuning parameters $c_\mathbf{x}$, $c_\mathbf{z}$, and C_{sub} in Algorithm 1 and 2 for which our results are valid. Note that $\|\boldsymbol{\Sigma}_\mathbf{w}\|$ reduces to $\sigma_\mathbf{w}^2$ if $\mathbf{w}_t \sim \mathcal{N}(0, \boldsymbol{\Sigma}_\mathbf{w})$ as in Section 2 of this work. t_{MC} is the mixing time of Markov matrix $\mathbf{T}$ given in Definition 3. $\bar{x}_0$ is a constant used to bound the initial state of each epoch, it is discussed and used in Proposition 4.

Table 3 lists the notations related to infinite-horizon MJS perturbation results closely following the notation in (Du et al., 2021). It provides several sensitivity parameters, e.g. how the optimal controller $\mathbf{K}_{1:s}^{?}$ varies with perturbations in the

Table 1. Notations — Trajectory Length Bounds

$\bar{\sigma}^2$	$\\|\mathbf{B}_{1:s}\\|^2\sigma_{\mathbf{z}}^2 + \sigma_{\mathbf{w}}^2$
(depending on context)	$\\|\mathbf{B}_{1:s}\\|^2\\|\mathbf{\Sigma}_{\mathbf{z}}\\| + \\|\mathbf{\Sigma}_{\mathbf{w}}\\|$ $\\|\mathbf{B}_{1:s}\\|^2\sigma_{\mathbf{z};0}^2 + \sigma_{\mathbf{w}}^2$
$\bar{\tau}$	$\max\{\tau(\tilde{\mathbf{L}}^{(0)}), \tau(\tilde{\mathbf{L}}^?)\}$
$\bar{\rho}$	$\max\{\rho(\tilde{\mathbf{L}}^{(0)}), \frac{1+\ \star}{2}\}$
$\beta_+(\rho, \tau)$	$\sqrt{\frac{2\sqrt{s}(\log(nT)+\ _{\mathbf{z}}^2 = \ _{\mathbf{w}}^2 \\|\mathbf{B}_{1:s}\\|^2)}{1}}$
$\underline{T}_{MC;1}(\delta)$	$\left(68 C_{sub}\pi_{\max}\pi_{\min}^{-2}\log(\frac{s}{\ })\right)^2$
$\underline{T}_{MC}(\delta)$	$\left(612 C_{sub}\pi_{\max}\pi_{\min}^{-2}\log(\frac{2s}{\ })\right)^2$
$\underline{T}_{cl;1}(\rho, \tau)$	$\frac{(1-\ )^2}{4n^{1.5}\sqrt{s}\ ^4}$
$\underline{T}_N(\delta, \rho, \tau)$	$\max\{\underline{T}_{MC}(\frac{\ }{2}), \underline{T}_{cl;1}(\rho, \tau)\}$
$\underline{T}_{id}(\delta, \rho, \tau)$	$\left(4c^2 C_{sub}(4+\beta_+^2(\rho, \tau))\log(\frac{2s(n+p)}{\ })(n+p)\pi_{\min}^{-1}\right)^2$
$\underline{T}_{id;N}(\delta, \rho, \tau)$	$\max\left\{\underline{T}_N(\frac{\ }{6}, \rho, \tau), \underline{T}_{id}(\frac{\ }{6}, \rho, \tau)\right\}$
$\underline{T}_{rgt;\ }(\delta)$	$\tilde{\mathcal{O}}(\bar{\epsilon}_{\mathbf{A};\mathbf{B};\mathbf{T}}^{-5})$
$\underline{T}_{rgt}(\delta)$	$\max\left\{\underline{T}_{id;N}(\delta, \bar{\rho}, \bar{\tau}), \underline{T}_{rgt;\ }(\delta)\right\}$

MJS parameters $\mathbf{A}_{1:s}, \mathbf{B}_{1:s}$, and $\mathbf{T}$ and how the cost J varies with the controller $\mathbf{K}_{1:s}$, together with certain upper bounds on $\mathbf{A}_{1:s}, \mathbf{B}_{1:s}, \mathbf{T}$, and $\mathbf{K}_{1:s}$ such that the perturbation theory holds. In this table, we use $\\|\cdot\\|_+ := \\|\cdot\\| + 1$ and $\mathbf{R}_{1:s}^{-1} := \{\mathbf{R}_i^{-1}\}_{i=1}^s$.

A.3. Supporting Lemmas

We provide a list of lemmas that will be useful for the subsequent proofs.

Lemma 3. *Suppose $\mathbf{z} \sim \mathcal{N}(0, \mathbf{\Sigma}_{\mathbf{z}})$ with $\mathbf{\Sigma}_{\mathbf{z}} \in \mathbb{R}^{p \times p}$. Suppose $t \geq (3 + 2\sqrt{2})p$, then we have*

$$\mathsf{P}(\|\mathbf{z}\|^2 \geq 3\|\mathbf{\Sigma}_{\mathbf{z}}\|t) \leq e^{-t}. \tag{25}$$

Proof. From (Hsu et al., 2012, Proposition 1), we have for any $t > 0$,

$$\mathsf{P}(\|\mathbf{z}\|^2 \geq \mathbf{tr}(\mathbf{\Sigma}_{\mathbf{z}}) + 2\sqrt{\mathbf{tr}(\mathbf{\Sigma}_{\mathbf{z}}^2)t} + 2\|\mathbf{\Sigma}_{\mathbf{z}}\|t) \leq e^{-t}, \tag{26}$$

which implies

$$\mathsf{P}(\|\mathbf{z}\|^2 \geq p\|\mathbf{\Sigma}_{\mathbf{z}}\| + 2\sqrt{p}\|\mathbf{\Sigma}_{\mathbf{z}}\|\sqrt{t} + 2\|\mathbf{\Sigma}_{\mathbf{z}}\|t) \leq e^{-t}. \tag{27}$$

We can see when $t \geq (3 + 2\sqrt{2})p$, we have $p + 2\sqrt{p}\sqrt{t} \leq t$, which implies $p\|\mathbf{\Sigma}_{\mathbf{z}}\| + 2\sqrt{p}\|\mathbf{\Sigma}_{\mathbf{z}}\|\sqrt{t} \leq \|\mathbf{\Sigma}_{\mathbf{z}}\|t$. Therefore, we have

$$\mathsf{P}(\|\mathbf{z}\|^2 \geq 3\|\mathbf{\Sigma}_{\mathbf{z}}\|t) \leq e^{-t}. \tag{28}$$

□

Lemma 4. *Let $\mathbf{x}_t$ be the state and define the noise-removed state $\tilde{\mathbf{x}}_t = \mathbf{x}_t - \mathbf{w}_{t-1}$ which is independent of $\mathbf{w}_{t-1}$. Let $\tilde{\mathbf{x}}_t$ be zero mean with $\mathsf{E}[\|\tilde{\mathbf{x}}_t\|_{\ell_2}] \leq B$ and $\mathbf{w}_t$ has i.i.d. entries with variance σ_w^2 bounded in absolute value by $c_w\sigma_w$ for some $c_w > 0$. Consider the conditional random vector*

$$\mathbf{y}_t \sim \{\mathbf{x}_t \mid \|\mathbf{x}_t\|_{\ell_2} \leq 3B\}.$$

If $c_w\sigma_w\sqrt{n} \leq B$, then $Cov[\mathbf{y}_t\mathbf{y}_t^T] \succeq \sigma_w^2\mathbf{I}_n/2$.

Proof. Observe that $\|\mathbf{w}_t\|_{\ell_2} \leq c_w\sigma_w\sqrt{n} \leq B$. Define the events

$$E_1 = \{\|\mathbf{x}_t\|_{\ell_2} \leq 3B\},\ E_2 = \{\|\tilde{\mathbf{x}}_t\|_{\ell_2} \leq 2B\}.$$

Table 2. Notations — Tuning Parameter Bounds

$\beta_+^0(\rho,\tau)$	$\sqrt{\log(nT)} + \beta_+(\rho,\tau)s\tau\rho + \sigma_{\mathbf{z}}/\sigma_{\mathbf{w}}\sqrt{p/n}\|\mathbf{B}_{1:s}\|$
$\underline{c}_{\mathbf{x}}(\rho,\tau)$	$3\sqrt{\frac{18n\sqrt{s}\ \ ^2}{\min\|\mathbf{\Sigma}_{\mathbf{w}}\|(1\ \ )}}$ or $3\sqrt{\frac{18n\sqrt{s}\ \ ^2}{\min\ _{\mathbf{w}}^2(1\ \ )}}$
$\underline{c}_{\mathbf{z}}$	$\max\left\{(\sqrt{3}+\sqrt{6})\sqrt{p}, \sqrt{3\log(\frac{6}{\min})}\right\}$
$\underline{C}_{sub,MC}$	$t_{MC}\cdot\max\{3, 3-3\log(\pi_{\max}\log(s))\}$
$\underline{C}_{sub,\mathbf{x}}(\bar{x}_0,\delta,T,\rho,\tau)$	$\frac{2}{\log(\ \ ^1)}+\frac{2}{\log(\ \ ^1)\log(T)}\log\left(\frac{24n\sqrt{s}\ \ \max\{x_0^2,\ ^2\}}{(1\ \ )}\right)\}$
$\underline{C}_{sub,\mathbf{x}}(\delta,T,\rho,\tau)$	$\frac{1}{\log(\ \ ^1)}+\frac{1}{\log(\ \ ^1)\log(T)}\log(\frac{72\sqrt{ns}^{1.5}}{\ })$
$\underline{C}_{sub,N}(\bar{x}_0,\delta,T,\rho,\tau)$	$\max\left\{\underline{C}_{sub,MC}, \underline{C}_{sub,\mathbf{x}}(\bar{x}_0,\frac{}{2},T,\rho,\tau), \underline{C}_{sub,\mathbf{x}}(\delta,T,\rho,\tau)\right\}$
$\underline{C}_{sub,id,t_0}(\bar{x}_0,T,\rho,\tau)$	$\frac{\log\left((1\ \ _{\mathbf{L}})x_0^2/(\log(nT)\ _{\mathbf{w}}^2+\ _{\mathbf{z}}^2\|\mathbf{B}_{1:s}\|)\right)}{(1\ \ )\log(T)}$
$\underline{C}_{sub,id,cov}(T,\rho,\tau)$	$\frac{4}{1\ \ }+\frac{2\log\left(8c^2\ _+(\ ,\ )\ ^0_+(\ ,\ )n\sqrt{ns}\ \right)}{(1\ \ )\log(T)}$
$\underline{C}_{sub,id,tr1}(\delta,T,\rho,\tau)$	$\frac{6}{1\ \ }+\frac{2\log\left(2\sqrt{ns}\ \ ^0_+(\ ,\ )/(\ _+(\ ,\ )\ )\right)}{(1\ \ )\log(T)}$
$\underline{C}_{sub,id,tr2}(\delta,T,\rho,\tau)$	$\frac{6}{1\ \ }+\frac{2\log\left(64c^2C_{sub}\ \ ^0_+(\ ,\ )(1+2\ _+(\ ,\ ))n\sqrt{s(n+p)}/(\ \min\ )\right)}{(1\ \ )\log(T)}$
$\underline{C}_{sub,id,tr3}(\delta,T,\rho,\tau)$	$\frac{8}{1\ \ }+\frac{2\log\left(c_w\ ^0_+(\ ,\ )\ n\sqrt{ns}/(C\ (1+2\ _+(\ ,\ ))(n+p))\right)}{(1\ \ )\log(T)}$
$\underline{C}_{sub,id}(\bar{x}_0,\delta,T,\rho,\tau)$	$\max\{\underline{C}_{sub,id,t_0}(\bar{x}_0,T,\rho,\tau), \underline{C}_{sub,id,cov}(T,\rho,\tau),$ $\underline{C}_{sub,id,tr1}(\frac{}{6},T,\rho,\tau), \underline{C}_{sub,id,tr2}(\frac{}{6},T,\rho,\tau), \underline{C}_{sub,id,tr3}(\frac{}{6},T,\rho,\tau)\}$
$\underline{C}_{sub,id,N}(\bar{x}_0,\delta,T,\rho,\tau)$	$\max\left\{\underline{C}_{sub,N}\left(\bar{x}_0,\frac{}{2},T,\rho,\tau\right), \underline{C}_{sub,id}(\delta,T,\rho,\tau)\right\}$
$\underline{C}_{sub,rgt}(\bar{x}_0,\delta,T)$	$\underline{C}_{sub,id,N}(\bar{x}_0,\delta,T,\bar{\rho},\bar{\tau})$

Clearly $E_2 \subset E_1$ as $\|\mathbf{w}_t\|_{\ell_2} \leq B$. Now, observe that

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

Note that $\mathsf{P}(E_2) \geq 0.5$ from Markov bound as $\mathsf{E}[\|\tilde{\mathbf{x}}_t\|_{\ell_2}] \leq B$. Additionally, on the event E_2, $\tilde{\mathbf{x}}_t$ and $\mathbf{w}_{t-1}$ are independent. Thus, we further have

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

□

Lemma 5. *Let* $\mathbf{z} \sim \mathcal{N}(0,\sigma_z^2\mathbf{I}_p)$. *Consider the conditional random vector* $\mathbf{y} \sim \{\mathbf{z} \mid \|\mathbf{z}\|_{\ell_2} \leq c\sigma_z\sqrt{p}\}$, *where* $c \geq 6$ *is a fixed constant. Then* $Cov[\mathbf{y}\mathbf{y}^T] \succeq \sigma_z^2\mathbf{I}_p/2$.

Proof. This proof shows a lower bound on the truncated Gaussian covariance $\mathbf{z} \mid \|\mathbf{z}\|_{\ell_2} \leq c\sigma_z\sqrt{p}$.

Note $\mathbf{z}' = \mathbf{z}/\sigma_z$ is $\mathcal{N}(0,1)$. Set variable $X = \|\mathbf{z}'\|_{\ell_2}^2$. We have the following Lipschitz Gaussian tail bound (we use Lipschitzness of the ℓ_2 norm and use minor calculus and relaxations)

$$\mathsf{P}(\|\mathbf{z}'\|_{\ell_2}^2 \geq 4tp) \leq \begin{cases}1 & \text{if} \quad t \leq 1\\ e^{-tp/2} & \text{if} \quad t \geq 1.\end{cases}.$$

This implies the following tail bound for X

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

Table 3. Notations — MJS-LQR Perturbation

ξ_1	$\|\mathbf{A}_{1:s}\|_+^2 \|\mathbf{B}_{1:s}\|_+^4 \|\mathbf{P}^?_{1:s}\|_+^3 \|\mathbf{R}_{1:s}^{\ 1}\|_+^2$
ξ_2	$\max\{\|\mathbf{A}_{1:s}\|_+, \|\mathbf{B}_{1:s}\|_+, \|\mathbf{P}^?_{1:s}\|_+, \|\mathbf{K}^?_{1:s}\|_+\}$
ξ_3	$\min\{\|\mathbf{B}_{1:s}\|_+{}^{2} \|\mathbf{R}_{1:s}^{\ 1}\|_+{}^{1} \|\mathbf{L}^?_{1:s}\|_+{}^{2}, \underline{\sigma}(\mathbf{P}^?_{1:s})\}$
$C^{\mathbf{P}}_{\mathbf{A};\mathbf{B};\mathbf{T}}$	$6\sqrt{ns}\tau(\tilde{\mathbf{L}}^?)(1-\rho^?)\ \ ^1\xi_1$
$C^{\mathbf{K}}_{\mathbf{A};\mathbf{B};\mathbf{T}}$	$28\xi_2^3\left(\underline{\sigma}(\mathbf{R}_{1:s})\ \ ^1 + \xi_2^3\underline{\sigma}(\mathbf{R}_{1:s})\ \ ^2\right) C^{\mathbf{P}}_{\mathbf{A};\mathbf{B};\mathbf{T}}$
$C^J_{\mathbf{K}}$	$2s\min\{n,p\}(\|\mathbf{R}_{1:s}\| + \xi_2^3)\frac{(\mathbf{L}^\star)}{1 \quad \star}$
$\bar{\epsilon}^{LQR}_{\mathbf{P}}$	$\min\left\{\xi_2, \frac{(1 \quad \star)\_(\mathbf{R}_{1:s})^2}{180 \sqrt{\ } \bar{s}\ \ ^6_2\ (\mathbf{L}^\star)(\_(\mathbf{R}_{1:s}) + \ ^3_2)}\right\}$
$\bar{\epsilon}_{\mathbf{K}}$	$\min\left\{\|\mathbf{B}_{1:s}\|\ \ ^1, \|\mathbf{K}^?_{1:s}\|, \frac{1 \quad \star}{2\sqrt{\ } \bar{s}\ (\mathbf{L}^\star)(1+2k\mathbf{L}^\star_{1:s}k)k\mathbf{B}_{1:s}k)}\right\}$
$\bar{\epsilon}_{\mathbf{A};\mathbf{B};\mathbf{T}}$	$\min\left\{\frac{\ _3(1 \quad \star)^2}{204ns\ (\mathbf{L}^\star)^2\ \ _1}, \|\mathbf{B}_{1:s}\|, \underline{\sigma}(\mathbf{Q}_{1:s}), \frac{\overset{LQR}{\mathbf{P}}}{2C^{\mathbf{P}}_{\mathbf{A},\mathbf{B},\mathbf{T}}}, \frac{\mathbf{K}}{2C^{\mathbf{K}}_{\mathbf{A},\mathbf{B},\mathbf{T}}}\right\}$

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

$$\mathsf{E}[X \mid X \geq \kappa p]\, \mathsf{P}(X \geq \kappa p) = -\int_p^1 x dQ(x) = -[xQ(x)]_n^1 + \int_p^1 Q(x)dx \tag{29}$$

$$\leq (\kappa p + 8)e^{\quad p=8}. \tag{30}$$

The final line as a function of κp 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

$$E_z = \{\|\mathbf{z}_t\|_{\ell_2} \leq \sqrt{\kappa}\sigma_z\sqrt{p}\}.$$

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 (30), this implies

$$\mathsf{E}[\|\mathbf{z}\|^2_{\ell_2} \mid E_z^c]\, \mathsf{P}(E_z^c) \leq \mathsf{E}[\sigma_z^2 X \mid E_z^c]\, \mathsf{P}({E_z}^c) \tag{31}$$

$$\leq \sigma_z^2\, \mathsf{E}[X \mid E_z^c]\, \mathsf{P}(E_z^c) \tag{32}$$

$$\leq \sigma_z^2/2. \tag{33}$$

This also yields the covariance bound of the tail event

$$\mathsf{E}[\mathbf{z}\mathbf{z}^T \mid E_z^c]\, \mathsf{P}(E_z^c) \preceq \mathsf{E}[\|\mathbf{z}\|^2_{\ell_2}\mathbf{I}_n \mid E_z^c]\, \mathsf{P}(E_z^c) \leq \sigma_z^2\mathbf{I}_p/2.$$

Finally, from conditional decomposition, observe that

$$\sigma_z^2\mathbf{I}_p\, \mathsf{E}[\mathbf{z}\mathbf{z}^T] = \mathsf{E}[\mathbf{z}\mathbf{z}^T \mid E_z^c]\, \mathsf{P}(E_z^c) + \mathsf{E}[\mathbf{z}\mathbf{z}^T \mid E_z]\, \mathsf{P}(E_z) \implies \mathsf{E}[\mathbf{z}\mathbf{z}^T \mid E_z]\, \mathsf{P}(E_z) \succeq \sigma_z^2\mathbf{I}_p/2.$$

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

$$\mathsf{E}[\mathbf{y}\mathbf{y}^T]\, \mathsf{P}(E_z) \succeq \sigma_z^2\mathbf{I}_p/2 \implies \mathsf{E}[\mathbf{y}\mathbf{y}^T] \succeq \sigma_z^2\mathbf{I}_p/2.$$

□

Theorem 3 (Isotropic (Vershynin, 2010))**.** *Let $\mathbf{X}$ be an $N \times d$ matrix whose rows $\mathbf{x}_i \in \mathsf{R}^d$ are independent isotropic. Let m be such that $\|\mathbf{x}_i\|_{\ell_2} \leq \sqrt{m}$ almost surely for all i. Then, for every $t \geq 0$, with probability $1 - 2d \cdot e^{\quad ct^2}$, we have*

$$\sqrt{N} - t\sqrt{m} \leq s_{\min}(\mathbf{X}) \leq s_{\max}(\mathbf{X}) \leq \sqrt{N} + t\sqrt{m}.$$

Corollary 3 (Non-isotropic)**.** *Let $\mathbf{X}$ be an $N \times d$ matrix whose rows $\mathbf{x}_i \in \mathsf{R}^d$ are independent with covariance $\mathbf{\Sigma}_i$. Suppose each covariance obeys*

$$\sigma^2_{\min} \leq s_{\min}(\mathbf{\Sigma}_i) \leq s_{\max}(\mathbf{\Sigma}_i) \leq \sigma^2_{\max}.$$

Let m be such that $\|\mathbf{x}_i\|_{\ell_2} \leq \sqrt{m}$ almost surely for all i. Then, for every $t \geq 0$, with probability $1 - 2d \cdot e^{\quad ct^2}$, we have

$$\sigma_{\min}\sqrt{N} - t\sqrt{m} \leq s_{\min}(\mathbf{X}) \leq s_{\max}(\mathbf{X}) \leq \sigma_{\max}\sqrt{N} + t\frac{\sigma_{\max}}{\sigma_{\min}}\sqrt{m}.$$

Proof. Let $\mathbf{x}_i' = \boldsymbol{\Sigma}_i^{-1/2}\mathbf{x}_i$. Observe that $\mathbf{x}_i'$ are independent isotropic. Define the matrix $\mathbf{X}'$ with rows $\mathbf{x}'$. Note that $\|\mathbf{x}_i'\|_{\ell_2} \leq \|\mathbf{x}_i\|_{\ell_2}/\sigma_{\min} \leq \sigma_{\min}^{-1}\sqrt{m}$. Thus, applying Theorem 3 on $\mathbf{X}'$, for every $t \geq 0$, with probability $1-2d\cdot e^{-ct^2}$, we have

$$\sqrt{N} - t\sigma_{\min}^{-1}\sqrt{m} \leq s_{\min}(\mathbf{X}') \leq s_{\max}(\mathbf{X}') \leq \sqrt{N} + t\sigma_{\min}^{-1}\sqrt{m}. \tag{34}$$

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

$$\sigma_{\min}^2\mathbf{X}'^\top\mathbf{X}' = \sigma_{\min}^2\sum_{i=1}^N \mathbf{x}_i'\mathbf{x}_i'^T \preceq \sum_{i=1}^N \sqrt{\boldsymbol{\Sigma}_i}\mathbf{x}_i'\mathbf{x}_i'^\top\sqrt{\boldsymbol{\Sigma}_i} \preceq \sigma_{\max}^2\sum_{i=1}^N \mathbf{x}_i'\mathbf{x}_i'^\top = \sigma_{\max}^2\mathbf{X}'^T\mathbf{X}'.$$

Thus, we found that

$$\sigma_{\min}s_{\min}(\mathbf{X}') \leq s_{\min}(\mathbf{X}) \leq s_{\max}(\mathbf{X}) \leq \sigma_{\max}s_{\max}(\mathbf{X}').$$

Plugging this into (34), we conclude with the theorem's statement. This completes the proof. ☐

B. Sys ID Analysis

B.1. Estimating $\mathbf{T}$

The following lemma 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 1.

Lemma 6. *Suppose we have an ergodic Markov chain $\mathbf{T} \in \mathbb{R}^{s\times s}$ with mixing time t_{MC} and stationary distribution $\pi_\infty \in \mathbb{R}^s$. Let $\pi_{\max} := \max_{i\in[s]} \pi_\infty(i)$ and $\pi_{\min} := \min_{i\in[s]} \pi_\infty(i)$. 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,*

$$\hat{\mathbf{T}}_{ij} = \frac{\sum_{k=1}^{\lfloor T/L\rfloor}\mathbf{1}_{\{\omega(kL)=i,\omega(kL+1)=j\}}}{\sum_{k=1}^{\lfloor T/L\rfloor}\mathbf{1}_{\{\omega(kL)=i\}}}$$

where $L = C_{sub}\log(T)$. Assume $C_{sub} \geq \underline{C}_{sub,MC}$, and for some $\delta > 0$, $T \geq \underline{T}_{MC,1}(\frac{\delta}{4})$, where $\underline{C}_{sub,MC}$ and $\underline{T}_{MC,1}(\delta)$ are defined in Table 2 and Table 1 respectively. Then, we have with probability at least $1-\delta$,

$$\|\hat{\mathbf{T}} - \mathbf{T}\| \leq 4\pi_{\min}^{-1}\|\mathbf{T}\|\sqrt{\log\left(\frac{4s}{\delta}\right)\frac{17\pi_{\max}C_{sub}\log(T)}{T}}. \tag{35}$$

Proof. Following the proof for (Zhang & Wang, 2018, Lemma 7) and the proof for (Du et al., 2019, Lemma 17), we have for $\epsilon < \pi_{\min}/2$, if we temporarily pick $L \geq 6t_{MC}\log(\epsilon^{-1})$.

$$\mathbb{P}\left(\|\hat{\mathbf{T}} - \mathbf{T}\| \leq 4\pi_{\min}^{-1}\|\mathbf{T}\|\epsilon\right) \geq 1 - 4s\exp\left(-\frac{T\epsilon^2}{17\pi_{\max}L}\right). \tag{36}$$

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 $\delta > 0$, pick

$$L \geq 3t_{MC}\log\left(\frac{T}{17\pi_{\max}L\log(\frac{4s}{\delta})}\right), \tag{37}$$

then when

$$T \geq 68L\pi_{\max}\pi_{\min}^{-2}\log(\frac{4s}{\delta}), \tag{38}$$

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

$$\|\hat{\mathbf{T}} - \mathbf{T}\| \leq 4\pi_{\min}^{-1}\|\mathbf{T}\|\sqrt{\frac{17\pi_{\max}L\log(\frac{4s}{\delta})}{T}} \tag{39}$$

Finally, by picking $L = C_{sub}\log(T)$, the lower bound $C_{sub} \geq t_{MC}\cdot\max\{3, 3-3\log(\pi_{\max}\log(4s))\}$ in the assumption guarantees (37), and the lower bound $T \geq \left(68C_{sub}\pi_{\max}\pi_{\min}^{-2}\log(\frac{4s}{\delta})\right)^2$ in the assumption guarantees (38). Plugging this choice L into (39), we conclude the proof. ☐

B.2. Lower Bounds for $|S_i|$

In Algorithm 1, we define subsampling period $L = C_{sub}\log(T)$, subsampling indices τ_k for $k = 1, 2, \dots, \lfloor T/L \rfloor$, and the time index set $S_i = \{\tau_k \mid \omega(\tau_k) = i, \ \ \|\mathbf{x}_{\tau_k}\| \le c_{\mathbf{x}}\sqrt{\|\boldsymbol{\Sigma}_{\mathbf{w}}\|\log(T)}, \ \ \|\mathbf{z}_{\tau_k}\| \le c_{\mathbf{z}}\sqrt{\|\boldsymbol{\Sigma}_{\mathbf{z}}\|}\}$ by bounding $\|\mathbf{x}_t\|$ and $\|\mathbf{z}_t\|$, which is used to estimate $\mathbf{A}_{1:s}$ and $\mathbf{B}_{1:s}$ through least squares. (Here we generalize isotropic noise $\mathbf{w}_t \sim \mathcal{N}(0, \sigma_{\mathbf{w}}^2)$ and $\mathbf{z}_t \sim \mathcal{N}(0, \sigma_{\mathbf{z}}^2)$ to $\mathcal{N}(0, \boldsymbol{\Sigma}_{\mathbf{w}})$ and $\mathcal{N}(0, \boldsymbol{\Sigma}_{\mathbf{z}})$.) A fundamental question is: is $|S_i|$ big enough such that there will be enough data available when applying least squares? We provide answers this question in this section. Lemma 7 acts as a building block for later result; Lemma 8 provides the lower bound on $|S_i|$; Corollary 4 gives a more interpretable lower bound on $|S_i|$ when $c_{\mathbf{x}}$ and $c_{\mathbf{z}}$ are large enough; and finally, Lemma 9 shows how many data in S_i are "weakly" independent, which is the quantity that essentially determines the sample complexity of estimating $\mathbf{A}_{1:s}$ and $\mathbf{B}_{1:s}$ in Algorithm 1.

For clarity, we reiterate some definitions and define a few new ones here. Given an MJS($\mathbf{A}_{1:s}, \mathbf{B}_{1:s}, \mathbf{T}$) with process noise $\mathbf{w}_t \sim \mathcal{N}(0, \boldsymbol{\Sigma}_{\mathbf{w}})$ and ergodic Markov matrix $\mathbf{T}$. With some stabilizing controller $\mathbf{K}_{1:s}$, the input is given by $\mathbf{u}_t = \mathbf{K}_{\omega(t)}\mathbf{x}_t + \mathbf{z}_t$ where $\mathbf{z}_t \sim \mathcal{N}(0, \boldsymbol{\Sigma}_{\mathbf{z}})$. Let $\bar{\sigma}^2 := \|\mathbf{B}_{1:s}\|^2\|\boldsymbol{\Sigma}_{\mathbf{z}}\| + \|\boldsymbol{\Sigma}_{\mathbf{w}}\|$. Let $\mathbf{L}_i := \mathbf{A}_i + \mathbf{B}_i\mathbf{K}_i$. Let $\tilde{\mathbf{L}} \in \mathbb{R}^{sn^2 \times sn^2}$ denote the augmented closed-loop state matrix with ij-th $n^2 \times n^2$ block given by $[\tilde{\mathbf{L}}]_{ij} := [\mathbf{T}]_{ji}\mathbf{L}_j \otimes \mathbf{L}_j$. Let $\tau_{\tilde{\mathbf{L}}} > 0, \rho_{\tilde{\mathbf{L}}} \in [0,1)$ be two constants such that $\|\tilde{\mathbf{L}}^k\| \le \tau_{\tilde{\mathbf{L}}}\rho_{\tilde{\mathbf{L}}}^k$. By definition, one available choice for $\tau_{\tilde{\mathbf{L}}}$ and $\rho_{\tilde{\mathbf{L}}}$ are $\tau_{\tilde{\mathbf{L}}}$ and $\rho(\tilde{\mathbf{L}})$. Let $t_{MC}(\cdot)$ and t_{MC} denote the mixing time of $\mathbf{T}$ as in Definition 3. Let $\boldsymbol{\pi}_\infty$ denote the stationary distribution of $\mathbf{T}$, and $\pi_{\min} = \min_i \boldsymbol{\pi}_\infty(i)$, $\pi_{\max} = \max_i \boldsymbol{\pi}_\infty(i)$. Assume the initial state $\mathbf{x}_0$ is bounded, i.e. $\|\mathbf{x}_0\| \le \bar{x}_0$ for some $\bar{x}_0 > 0$.

Lemma 7. *Suppose the Markov chain trajectory $\{\omega(0), \omega(1), \dots\}$ and a sequence of events $\{A_0, A_1, \dots\}$ are both adapted to filtration $\{\mathcal{F}_0, \mathcal{F}_1, \dots\}$, i.e. $\omega(t)$ and $\mathbf{1}_{\{A_t\}}$ are both $\mathcal{F}_t$-measurable. We assume $\mathsf{E}[\mathbf{1}_{\{\omega(t)=j\}} \mid \mathcal{F}_{t-r}] = \mathsf{P}(\omega(t) = j \mid \omega(t-r))$ for all $j \in [s], t$, and $r < t$. For all $i \in [s]$, let*

$$N_i = \sum_{k=1}^{\lfloor T/L \rfloor} \mathbf{1}_{\{\omega(kL)=i\}}\mathbf{1}_{\{A_{kL}\}} \tag{40}$$

and suppose

$$\mathsf{E}[\mathbf{1}_{\{A_t\}} \mid \mathcal{F}_{t-L}] \ge 1 - p_t \tag{41}$$

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

$$\mathsf{P}\left(\bigcap_{i=1}^{s}\left\{N_i \ge \frac{T\boldsymbol{\pi}_\infty(i)}{C_{sub}\log(T)}\left(1 - \frac{1}{\boldsymbol{\pi}_\infty(i)}\sqrt{\log(\frac{s}{\delta})\frac{17C_{sub}\pi_{\max}\log(T)}{T}}\right) - \sum_{k=1}^{\lfloor T/L \rfloor} p_{kL}\right\}\right) \ge 1 - \delta. \tag{42}$$

Proof. For some $\epsilon < \pi_{\min}/2$, we temporarily let $L \ge 6t_{MC}\log(\epsilon^{-1})$. From the proof for (Du et al., 2019, Lemma 17), we know this guarantees $L \ge t_{MC}(\epsilon/2)$. By definition of $t_{MC}(\cdot)$, we know $\max_i \|([\mathbf{T}^L]_{i,:})^\intercal - \boldsymbol{\pi}_\infty\|_1 \le \epsilon \le \pi_{\min}/2$, and since $([\mathbf{T}^L]_{i,:})\mathbf{1} = \boldsymbol{\pi}_\infty^\intercal\mathbf{1} = 1$, we further have

$$\max_i \|([\mathbf{T}^L]_{i,:})^\intercal - \boldsymbol{\pi}_\infty\|_\infty \le \frac{\epsilon}{2} \le \frac{\pi_{\min}}{4}. \tag{43}$$

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

$$\boldsymbol{\delta}_j(i) := \mathbf{1}_{\{\tilde{\omega}(j)=i\}}\mathbf{1}_{\{\tilde{A}_j\}} - \mathsf{E}[\mathbf{1}_{\{\tilde{\omega}(j)=i\}}\mathbf{1}_{\{\tilde{A}_j\}} \mid \tilde{\mathcal{F}}_{j-1}] \tag{44}$$

$$\boldsymbol{\Delta}_k(i) := \sum_{j=1}^{k} \boldsymbol{\delta}_j(i). \tag{45}$$

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

$$\begin{aligned}\mathsf{E}[\mathbf{\Delta}_{k+1}(i) \mid \tilde{\mathcal{F}}_k] &= \mathsf{E}[\sum_{j=1}^{k+1} \delta_j(i) \mid \tilde{\mathcal{F}}_k] \\ &= \sum_{j=1}^{k} \delta_j(i) + \mathsf{E}[\mathbf{1}_{\{\tilde\omega(k+1)=i\}}\mathbf{1}_{\{\mathcal{A}_{k+1}\}} - \mathsf{E}[\mathbf{1}_{\{\tilde\omega(k+1)=i\}}\mathbf{1}_{\{\mathcal{A}_{k+1}\}} \mid \tilde{\mathcal{F}}_k] \mid \tilde{\mathcal{F}}_k] \\ &= \sum_{j=1}^{k} \delta_j(i) = \mathbf{\Delta}_k(i),\end{aligned} \tag{46}$$

thus $\delta_k(i) = \mathbf{\Delta}_k(i) - \mathbf{\Delta}_{k-1}(i)$ can be viewed as the martingale difference sequence. Since $\mathsf{E}[\delta_k(i) \mid \tilde{\mathcal{F}}_{k-1}] = 0$, we have $\mathsf{E}[\delta_k(i)^2 \mid \tilde{\mathcal{F}}_{k-1}] = \mathrm{Var}(\delta_k(i) \mid \tilde{\mathcal{F}}_{k-1}) = \mathrm{Var}(\mathbf{1}_{\{\tilde\omega(j)=i\}}\mathbf{1}_{\{\mathcal{A}_j\}} \mid \tilde{\mathcal{F}}_{k-1}) \le \mathsf{E}[\mathbf{1}^2_{\{\tilde\omega(k)=i\}}\mathbf{1}^2_{\{\mathcal{A}_k\}} \mid \tilde{\mathcal{F}}_{k-1}] \le \mathsf{E}[\mathbf{1}_{\{\tilde\omega(k)=i\}} \mid \tilde{\mathcal{F}}_{k-1}] = \mathsf{P}(\tilde\omega(k) = i \mid \tilde\omega(k-1)) = [\mathbf{T}^L]_{\omega((k-1)L),i}$. By the choice of L, using (43), we know $[\mathbf{T}^L]_{\omega((k-1)L),i} \le \pi_\infty(i) + \max_j \|([\mathbf{T}^L]_{j,:})^\intercal - \pi_\infty\|_1 \le 2\pi_{\max}$. Thus,

$$\sum_{k=1}^{\tilde T} \mathsf{E}[\delta_k(i)^2 \mid \tilde{\mathcal{F}}_{k-1}] \le 2\pi_{\max}\tilde{T}. \tag{47}$$

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

$$\begin{aligned}&\mathsf{P}\left(N_i - \sum_{k=1}^{\tilde T} \mathsf{E}[\mathbf{1}_{\{\tilde\omega(k)=i\}}\mathbf{1}_{\{\mathcal{A}_k\}} \mid \tilde{\mathcal{F}}_{k-1}] \ge \tilde{T}\frac{\epsilon}{2}\right) \\ &\overset{\text{(i)}}{=}\mathsf{P}(\mathbf{\Delta}_{\tilde T}(i) \ge \tilde{T}\epsilon/2) \\ &\overset{\text{(ii)}}{\le} \exp(-\frac{\tilde{T}\epsilon^2/8}{2\pi_{\max} + \epsilon/6}) \\ &\overset{\text{(iii)}}{\le} \exp(-\frac{T\epsilon^2}{17\pi_{\max}L})\end{aligned} \tag{48}$$

where (i) follows from the definition of N_i and $\mathbf{\Delta}_{\tilde T}(i)$; (ii) follows from Freedman's inequality (Freedman, 1975, Theorem 1.6), and (iii) follows from $\epsilon \le \pi_{\min}/2$. Note that

$$\begin{aligned}&\left|\sum_{k=1}^{\tilde T} \mathsf{E}[\mathbf{1}_{\{\tilde\omega(k)=i\}}\mathbf{1}_{\{\mathcal{A}_k\}} \mid \tilde{\mathcal{F}}_{k-1}] - \tilde{T}\pi_\infty(i)\right| \\ \le&\left|\sum_{k=1}^{\tilde T} \mathsf{E}[\mathbf{1}_{\{\tilde\omega(k)=i\}} \mid \tilde{\mathcal{F}}_{k-1}] - \tilde{T}\pi_\infty(i)\right| + \left|\sum_{k=1}^{\tilde T} \mathsf{E}[\mathbf{1}_{\{\tilde\omega(k)=i\}} \mid \tilde{\mathcal{F}}_{k-1}] - \mathsf{E}[\mathbf{1}_{\{\tilde\omega(k)=i\}}\mathbf{1}_{\{\mathcal{A}_k\}} \mid \tilde{\mathcal{F}}_{k-1}]\right| \\ \le&\tilde{T}\max_j |[\mathbf{T}^L]_{j,i} - \pi_\infty(i)| + \left|\sum_{k=1}^{\tilde T} \mathsf{E}[\mathbf{1}_{\{\mathcal{A}_k^c\}} \mid \tilde{\mathcal{F}}_{k-1}]\right| \\ \le&\tilde{T}\frac{\epsilon}{2} + \sum_{k=1}^{\tilde T} p_{kL}.\end{aligned} \tag{49}$$

Then, combining this with (48) and applying union bound, we have with probability at least $1 - s\exp(-\frac{T\epsilon^2}{17\pi_{\max}L})$,

$$\bigcap_{i=1}^{s}\left\{N_i \ge \frac{T}{L}\pi_\infty(i) - \frac{T}{L}\epsilon - \sum_{k=1}^{T/L} p_{kL}\right\} \tag{50}$$

when $\epsilon < \pi_{\max}/2$ and $L \geq 6t_{MC}\log(\epsilon\ \ ^{1})$.

Then, similar to the proof of Lemma 6, we know if we pick $L = C_{sub}\log(T)$ with

$$C_{sub} \geq t_{MC} \cdot \max\{3, 3 - 3\log(\pi_{\max}\log(s))\}. \tag{51}$$

and for $\delta > 0$, with

$$T \geq \left(68C_{sub}\pi_{\max}\pi_{\min}^{2}\log(\frac{s}{\delta})\right)^2 \tag{52}$$

Then, with probability at least $1 - \delta$,

$$\bigcap_{i=1}^{s}\left\{N_i \geq \frac{T\ \ _1(i)}{C_{sub}\log(T)}\left(1 - \frac{1}{\ \ _1(i)}\sqrt{\log(\frac{s}{\delta})\frac{17C_{sub}\pi_{\max}\log(T)}{T}}\right) - \sum_{k=1}^{T=L} p_{kL}\right\}. \tag{53}$$

□

Lemma 8. *Assume* $c_{\mathbf{z}} \geq (\sqrt{3}+\sqrt{6})\sqrt{p}$. *For some* $\delta > 0$, *we assume* $C_{sub} \geq \max\{\underline{C}_{sub;MC}, \underline{C}_{sub;\mathbf{x}}(\bar{x}_0, \delta, T, \rho_{\mathbf{L}}, \tau_{\mathbf{L}})\}$ *and* $T \geq \max\{\underline{T}_{MC;1}(\frac{}{2}), \underline{T}_{cl;1}(\rho_{\mathbf{L}}, \tau_{\mathbf{L}})\}$ *where* $\underline{C}_{sub;MC}$ *and* $\underline{C}_{sub;\mathbf{x}}(\bar{x}_0, \delta, T, \rho, \tau)$ *are defined in Table 2, and* $\underline{T}_{MC;1}(\delta)$ *and* $\underline{T}_{cl;1}(\rho, \tau)$ *are defined in Table 1. Then, with probability at least* $1 - \delta$, *the following intersected events occur*

$$\bigcap_{i=1}^{s}\left\{|S_i| \geq \frac{T\ \ _1(i)}{C_{sub}\log(T)}\left(1 - \frac{1}{\ \ _1(i)}\sqrt{\log(\frac{2s}{\delta})\frac{17C_{sub}\pi_{\max}\log(T)}{T}}\right.\right.$$
$$\left.\left. - \frac{2n\sqrt{s}\tau_{\mathbf{L}}\bar{\sigma}^2}{\ \ _1(i)c_{\mathbf{x}}^2\|\mathbf{\Sigma_w}\|\log(T)(1-\rho_{\mathbf{L}})} - \frac{1}{\ \ _1(i)}e^{\ \ \frac{c_{\mathbf{z}}^2}{3}}\right\} \tag{54}$$

Proof. For simplicity, we assume $\lfloor T/L \rfloor = T/L$. We let $\mathcal{F}_t$ denote the sigma algebra generated by $\{\{\omega(r)\}_{r=0}^{t}, \mathbf{w}_{0:t}, \mathbf{z}_{0:t}, \mathbf{x}_0\}$, and let $A_t = \{\|\mathbf{x}_t\| \leq c_{\mathbf{x}}\sqrt{\|\mathbf{\Sigma_w}\|\log(T)}, \|\mathbf{z}_t\| \leq c_{\mathbf{z}}\sqrt{\|\mathbf{\Sigma_z}\|}\}$, then by Lemma 7, we know when (i) $C_{sub} \geq \underline{C}_{sub;MC} = t_{MC} \cdot \max\{3, 3 - 3\log(\pi_{\max}\log(s))\}$, and (ii) $T \geq \underline{T}_{MC;1}(\frac{}{2}) = \left(68C_{sub}\pi_{\max}\pi_{\min}^{2}\log(\frac{2s}{})\right)^2$, we have with probability at least $1 - \frac{}{2}$,

$$\bigcap_{i=1}^{s}\left\{|S_i| \geq \frac{T\ \ _1(i)}{C_{sub}\log(T)} \cdot \left(1 - \frac{1}{\ \ _1(i)}\sqrt{\log(\frac{2s}{\delta})\frac{17C_{sub}\pi_{\max}\log(T)}{T}}\right)\right.$$
$$\left. - \underbrace{\sum_{k=1}^{T=L}\mathsf{P}\left(A_t^c \mid \mathbf{x}_{(k\ \ 1)L+1}, \omega((k-1)L)\right)}_{=:P}\right\}. \tag{55}$$

For term P, we have

$$\begin{aligned} P &= \sum_{k=1}^{T=L}\mathsf{P}\left(\|\mathbf{x}_{kL}\| \geq c_{\mathbf{x}}\sqrt{\|\mathbf{\Sigma_w}\|\log(T)} \bigcup \|\mathbf{z}_{kL}\| \geq c_{\mathbf{z}}\sqrt{\|\mathbf{\Sigma_z}\|} \mid \mathbf{x}_{(k\ \ 1)L+1}, \omega((k-1)L)\right) \\ &\leq \underbrace{\sum_{k=1}^{T=L}\mathsf{P}\left(\|\mathbf{z}_{kL}\| \geq c_{\mathbf{z}}\sqrt{\|\mathbf{\Sigma_z}\|} \mid \mathbf{x}_{(k\ \ 1)L+1}, \omega((k-1)L)\right)}_{=:P_1} \\ &\quad + \underbrace{\sum_{k=1}^{T=L}\mathsf{P}\left(\|\mathbf{x}_{kL}\| \geq c_{\mathbf{x}}\sqrt{\|\mathbf{\Sigma_w}\|\log(T)} \mid \mathbf{x}_{(k\ \ 1)L+1}, \omega((k-1)L)\right)}_{=:P_2}. \end{aligned} \tag{56}$$

For term P_1, we know from Lemma 3 that when $c_{\mathbf{z}} \geq (\sqrt{3}+\sqrt{6})\sqrt{p}$, we have

$$P_1 = \sum_{k=1}^{T=L}\mathsf{P}\left(\|\mathbf{z}_{kL}\| \geq c_{\mathbf{z}}\sqrt{\|\mathbf{\Sigma_z}\|}\right) \leq \frac{T}{L}e^{\ \ \frac{c_{\mathbf{z}}^2}{3}}. \tag{57}$$

Now we consider term P_2. From Lemma 2, we know

$$\mathsf{E}[\|\mathbf{x}_{kL}\|^2 \mid \mathbf{x}_{(k-1)L+1}, \omega((k-1)L)] \leq \sqrt{ns}\tau_{\mathbf{L}}\rho_{\mathbf{L}}^{L-1}\|\mathbf{x}_{(k-1)L+1}\|^2 + \frac{n\sqrt{s}\tau_{\mathbf{L}}\bar{\sigma}^2}{1-\rho_{\mathbf{L}}}, \tag{58}$$

thus by Markov inequality, we have

$$\begin{aligned} P_2 &\leq \sum_{k=1}^{T/L} \frac{1}{c_{\mathbf{x}}^2\|\mathbf{\Sigma_w}\|\log(T)} \left(\sqrt{ns}\tau_{\mathbf{L}}\rho_{\mathbf{L}}^{L-1}\|\mathbf{x}_{(k-1)L+1}\|^2 + \frac{n\sqrt{s}\tau_{\mathbf{L}}\bar{\sigma}^2}{1-\rho_{\mathbf{L}}}\right) \\ &\leq \frac{1}{c_{\mathbf{x}}^2\|\mathbf{\Sigma_w}\|\log(T)} \left(\frac{T}{L}\frac{n\sqrt{s}\tau_{\mathbf{L}}\bar{\sigma}^2}{1-\rho_{\mathbf{L}}} + \sqrt{ns}\tau_{\mathbf{L}}\rho_{\mathbf{L}}^{L-1}\sum_{k=1}^{T/L}\|\mathbf{x}_{(k-1)L+1}\|^2\right). \end{aligned} \tag{59}$$

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

$$L{=}C_{sub}\log(T){\geq}\frac{1}{\log(\rho_{\mathbf{L}}^{-1})}\max\left\{\log(2), 2\log(\frac{8\sqrt{ns}\tau_{\mathbf{L}}\bar{x}_0^2}{\delta}), 2\log(4T\frac{n\sqrt{s}\tau_{\mathbf{L}}\bar{\sigma}^2}{\delta(1-\rho_{\mathbf{L}})})+2,\right\}. \tag{60}$$

Then, we have

$$\begin{aligned} &\mathsf{P}\left(\rho_{\mathbf{L}}^{L-1}\sum_{k=1}^{T/L}\|\mathbf{x}_{(k-1)L+1}\|^2 \leq \frac{T}{L}\frac{(1-\rho_{\mathbf{L}})}{4Tn\sqrt{s}\tau_{\mathbf{L}}\bar{\sigma}^2}\right) \\ \overset{\text{(i)}}{\geq}&\mathsf{P}\left(\rho_{\mathbf{L}}^{L-1}\sum_{k=1}^{T/L}\|\mathbf{x}_{(k-1)L+1}\|^2 \leq \frac{T}{L}\rho_{\mathbf{L}}^{\frac{L}{2}-1}\right) \\ \geq&\mathsf{P}\left(\bigcap_{k=1}^{T/L}\left\{\|\mathbf{x}_{(k-1)L+1}\|^2 \leq \rho_{\mathbf{L}}^{-\frac{L}{2}}\right\}\right) \\ \geq&1-\sum_{k=1}^{T/L}\mathsf{P}\left(\|\mathbf{x}_{(k-1)L+1}\|^2 \geq \rho_{\mathbf{L}}^{-\frac{L}{2}}\right) \\ \overset{\text{(ii)}}{\geq}&1-\sum_{k=1}^{T/L}\rho_{\mathbf{L}}^{\frac{L}{2}}\left(\sqrt{ns}\tau_{\mathbf{L}}\rho_{\mathbf{L}}^{(k-1)L+1}\bar{x}_0^2 + \frac{n\sqrt{s}\tau_{\mathbf{L}}\bar{\sigma}^2}{1-\rho_{\mathbf{L}}}\right) \\ \geq&1-\rho_{\mathbf{L}}^{\frac{L}{2}+1}\frac{\sqrt{ns}\tau_{\mathbf{L}}\bar{x}_0^2}{1-\rho_{\mathbf{L}}^L} - \frac{T}{L}\rho_{\mathbf{L}}^{\frac{L}{2}}\frac{n\sqrt{s}\tau_{\mathbf{L}}\bar{\sigma}^2}{1-\rho_{\mathbf{L}}} \\ \overset{\text{(iii)}}{\geq}&1-2\rho_{\mathbf{L}}^{\frac{L}{2}}\sqrt{ns}\tau_{\mathbf{L}}\bar{x}_0^2 - \frac{\delta}{4L} \\ \overset{\text{(iv)}}{\geq}&1-\frac{\delta}{4}-\frac{\delta}{4} = 1-\frac{\delta}{2} \end{aligned} \tag{61}$$

where (i) follows from (60) which gives $\rho_{\mathbf{L}}^{\frac{L}{2}-1} \leq \frac{(1-\rho_{\mathbf{L}})}{4Tn\sqrt{s}\tau_{\mathbf{L}}\bar{\sigma}^2}$; (ii) follows from Lemma 2 and Markov inequality; (iii) follows from (60) which gives $\rho_{\mathbf{L}}^L \leq \frac{1}{2}$ and $\rho_{\mathbf{L}}^{\frac{L}{2}} \leq \frac{(1-\rho_{\mathbf{L}})}{4Tn\sqrt{s}\tau_{\mathbf{L}}\bar{\sigma}^2}$ and (iv) follows from (60) which gives $\rho_{\mathbf{L}}^{\frac{L}{2}} \leq \frac{\delta}{8\sqrt{ns}\tau_{\mathbf{L}}\bar{x}_0^2}$. Therefore, we have with probability at least $1-\frac{\delta}{2}$

$$P_2 \leq \frac{1}{c_{\mathbf{x}}^2\|\mathbf{\Sigma_w}\|\log(T)}\left(\frac{T}{L}\frac{n\sqrt{s}\tau_{\mathbf{L}}\bar{\sigma}^2}{1-\rho_{\mathbf{L}}} + \frac{1-\rho_{\mathbf{L}}}{4L\sqrt{n}\bar{\sigma}^2}\right), \tag{62}$$

and thus,

$$\begin{aligned} P = P_1 + P_2 &\leq \frac{1}{c_{\mathbf{x}}^2 \|\boldsymbol{\Sigma}_{\mathbf{w}}\| \log(T)} \left(\frac{T}{L} \frac{n\sqrt{s}\tau_{\mathbf{L}}\bar{\sigma}^2}{1-\rho_{\mathbf{L}}} + \frac{1-\rho_{\mathbf{L}}}{4L\sqrt{n}\bar{\sigma}^2} \right) + \frac{T}{L} e^{-\frac{c_{\mathbf{z}}^2}{3}} \\ &\leq \frac{1}{c_{\mathbf{x}}^2 \|\boldsymbol{\Sigma}_{\mathbf{w}}\| \log(T)} \left(\frac{T}{L} \frac{2n\sqrt{s}\tau_{\mathbf{L}}\bar{\sigma}^2}{1-\rho_{\mathbf{L}}} \right) + \frac{T}{L} e^{-\frac{c_{\mathbf{z}}^2}{3}}, \end{aligned} \tag{63}$$

where the second inequality follows from $T \geq \underline{T}_{cl,1}(\rho_{\mathbf{L}}, \tau_{\mathbf{L}})$. Plugging this into (55), we have with probability at least $1-\delta$,

$$\bigcap_{i=1}^{s} \left\{ |S_i| \geq \frac{T\pi_1(i)}{C_{sub}\log(T)} \cdot \left(1 - \frac{1}{\pi_1(i)} \sqrt{\log(\frac{2s}{\delta}) \frac{17 C_{sub} \pi_{\max} \log(T)}{T}} - \frac{2n\sqrt{s}\tau_{\mathbf{L}}\bar{\sigma}^2}{\pi_1(i) c_{\mathbf{x}}^2 \|\boldsymbol{\Sigma}_{\mathbf{w}}\| \log(T)(1-\rho_{\mathbf{L}})} - \frac{1}{\pi_1(i)} e^{-\frac{c_{\mathbf{z}}^2}{3}}. \right) \right\}, \tag{64}$$

which concludes the proof. □

Note that when T, $c_{\mathbf{x}}$, and $c_{\mathbf{z}}$ are large enough, we could obtain a more interpretable version of Lemma 8.

Corollary 4. *Assume $c_{\mathbf{x}} \geq \frac{1}{3}\underline{c}_{\mathbf{x}}(\rho_{\mathbf{L}}, \tau_{\mathbf{L}})$, $c_{\mathbf{z}} \geq \underline{c}_{\mathbf{z}}$, $T \geq \max\{\underline{T}_{MC}(\delta), \underline{T}_{cl,1}(\rho_{\mathbf{L}}, \tau_{\mathbf{L}})\}$, and $C_{sub} \geq \max\{\underline{C}_{sub,MC}, \underline{C}_{sub,\mathbf{x}}(\bar{x}_0, \delta, T, \rho_{\mathbf{L}}, \tau_{\mathbf{L}})\}$, where $\underline{c}_{\mathbf{x}}(\rho, \tau)$, $\underline{c}_{\mathbf{z}}$, $\underline{C}_{sub,MC}$, $\underline{C}_{sub,\mathbf{x}}(\bar{x}_0, \delta, T, \rho, \tau)$ are defined in Table 2, and $\underline{T}_{MC}(\delta)$ and $\underline{T}_{cl,1}(\rho, \tau)$ are defined in Table 1. Then, with probability at least $1-\delta$, the following intersected events occur*

$$\bigcap_{i=1}^{s} \left\{ |S_i| \geq \frac{T\pi_{\min}}{2C_{sub}\log(T)} \right\}. \tag{65}$$

Remark 1. *One may notice that in Lemma 8 (Corollary 4), the lower bounds on subsampling factor C_{sub} and trajectory length T depend on each other through $\underline{C}_{sub,\mathbf{x}}(\bar{x}_0, \delta, T, \rho_{\mathbf{L}}, \tau_{\mathbf{L}})$ and $\underline{T}_{MC,1}(\delta)$ ($\underline{T}_{MC}(\delta)$), which makes it difficult for parameter choice and performance validation. Note that $\underline{C}_{sub,\mathbf{x}}(\bar{x}_0, \delta, T, \rho_{\mathbf{L}}, \tau_{\mathbf{L}})$ only depends on T through the term $\frac{1}{\log(T)}$, thus an additional absolute lower bound on T, e.g. the trivial choice $T \geq 2$, would eliminate $\frac{1}{\log(T)}$ and the dependency. By doing this, though the lower bounds on T still depend on C_{sub}, the lower bounds on C_{sub} will no longer depend on T. We keep $\frac{1}{\log(T)}$ mainly for the purpose of regret analysis in Appendix C.3.*

Now we provide a result on how many data in S_i are "weakly" independent, which is the quantity that essentially determines the sample complexity of estimating $\mathbf{A}_{1:s}$ and $\mathbf{B}_{1:s}$ in Algorithm 1, with more details to come in Appendix B.3. We first define a few notations. Let $\tau_{i,1}, \ldots, \tau_{i,|S_i|}$ denote the elements in S_i, and let $\tau_{i,0} = 0$. Define $\bar{\mathbf{x}}_{\tau_{i,k}}$ such that

$$\bar{\mathbf{x}}_{\tau_{i,k}} = \sum_{j=1}^{\tau_{i,k}-1} \left(\prod_{k=1}^{j-1} \mathbf{L}_{\omega(t-k)} \right) \left(\mathbf{B}_{\omega(t-j)} \mathbf{z}_{t-j} + \mathbf{w}_{t-j} \right) + \mathbf{B}_{\omega(\tau_{i,k}-1)} \mathbf{z}_{\tau_{i,k}-1} + \mathbf{w}_{\tau_{i,k}-1}. \tag{66}$$

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, $\mathbf{w}_{\tau_{i,k-1}:\tau_{i,k}-1}$, $\mathbf{z}_{\tau_{i,k-1}:\tau_{i,k}-1}$, $\{\omega(t')\}_{t'=\tau_{i,k-1}}^{\tau_{i,k}-1}$. Or, one can also view $\bar{\mathbf{x}}_{\tau_{i,k}}$ as the contribution of noise $\mathbf{x}_t$ and $\mathbf{z}_t$ that propagate $\mathbf{x}_{\tau_{i,k-1}}$ to $\mathbf{x}_{\tau_{i,k}}$. And it is easy to see that

$$\mathbf{x}_{\tau_{i,k}} - \bar{\mathbf{x}}_{\tau_{i,k}} = \left(\prod_{k=1}^{\tau_{i,k}-1} \mathbf{L}_{\omega(t-k)} \right) \mathbf{x}_{\tau_{i,k-1}}. \tag{67}$$

Define $\bar{S}_i \subseteq S_i$ such that

$$\bar{S}_i := \left\{ \tau_k \middle| \omega(\tau_k) = i, \|\mathbf{x}_{\tau_k}\| \leq c_{\mathbf{x}} \sqrt{\|\boldsymbol{\Sigma}_{\mathbf{w}}\| \log(T)}, \|\mathbf{z}_{\tau_k}\| \leq c_{\mathbf{z}} \sqrt{\boldsymbol{\Sigma}_{\mathbf{z}}}, \|\bar{\mathbf{x}}_{\tau_k}\| \leq \frac{c_{\mathbf{x}} \sqrt{\|\boldsymbol{\Sigma}_{\mathbf{w}}\| \log(T)}}{2} \right\}. \tag{68}$$

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

Lemma 9. *Assume $c_{\mathbf{x}} \geq \underline{c}_{\mathbf{x}}(\rho_{\mathbf{L}}, \tau_{\mathbf{L}})$, $c_{\mathbf{z}} \geq \underline{c}_{\mathbf{z}}$, $C_{sub} \geq \underline{C}_{sub;N}\,(\bar{x}_0, \delta, T, \rho_{\mathbf{L}}, \tau_{\mathbf{L}})$, and $T \geq \underline{T}_N\,(\delta, \rho_{\mathbf{L}}, \tau_{\mathbf{L}})$, where $\underline{c}_{\mathbf{x}}(\rho, \tau)$, $\underline{c}_{\mathbf{z}}$, and $C_{sub;N}\,(\bar{x}_0, \delta, T, \rho, \tau)$ are defined in Table 2, and $\underline{T}_N\,(\delta, \rho, \tau)$ is defined in Table 1. Then with probability at least $1-\delta$, the following intersected events occur*

$$\bigcap_{i=1}^{s} \left\{ |\bar{S}_i| \geq \frac{T\pi_{\min}}{2C_{sub}\log(T)} \right\}. \tag{69}$$

Proof. We define sets $R_i \subseteq S_i$ and $\bar{R}_i \subseteq \bar{S}_i$ such that

$$R_i := \left\{ \tau_k \middle| \omega(\tau_k) = i, \|\mathbf{x}_{\ \ k}\| \leq \frac{c_{\mathbf{x}}\sqrt{\|\mathbf{\Sigma_w}\|\log(T)}}{3}, \|\mathbf{z}_{\ \ k}\| \leq c_{\mathbf{z}}\sqrt{\mathbf{\Sigma_z}} \right\}. \tag{70}$$

$$\bar{R}_i := \left\{ \tau_k \middle| \omega(\tau_k) = i, \|\mathbf{x}_{\ \ k}\| \leq \frac{c_{\mathbf{x}}\sqrt{\|\mathbf{\Sigma_w}\|\log(T)}}{3}, \|\mathbf{z}_{\ \ k}\| \leq c_{\mathbf{z}}\sqrt{\mathbf{\Sigma_z}}, \|\bar{\mathbf{x}}_{\ \ k}\| \leq \frac{c_{\mathbf{x}}\sqrt{\|\mathbf{\Sigma_w}\|\log(T)}}{2} \right\}. \tag{71}$$

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

Using Corollary 4, we see under given assumptions, with probability at least $1 - {}_{2}$,

$$\bigcap_{i=1}^{s} \left\{ |R_i| \geq \frac{T\pi_{\min}}{2C_{sub}\log(T)} \right\}. \tag{72}$$

Let $\zeta_{i;1}, \dots, \zeta_{i;jR_ij}$ denote the elements in R_i. It is easy to see $\{\zeta_{i;1}, \dots, \zeta_{i;jR_ij}\} \subseteq \{\tau_{i;1}, \dots, \tau_{i;jS_ij}\}$. Consider an arbitrary $\zeta_{i;j} \in R_i$ and $\tau_{i;j^0} \in S_i$ denote the counterpart of $\zeta_{i;j}$ such that $\tau_{i;j^0} = \zeta_{i;j}$. By definition of R_i, we have

$$\|\mathbf{x}_{\ \ i,j^0}\| \leq \frac{c_{\mathbf{x}}\sqrt{\|\mathbf{\Sigma_w}\|\log(T)}}{3} \tag{73}$$

From (67), together with Lemma 2, we have

$$\begin{aligned} \mathsf{E}[\|\mathbf{x}_{\ \ i,j^0} - \bar{\mathbf{x}}_{\ \ i,j^0}\|^2] &\leq \sqrt{ns}\tau_{\mathbf{L}}\rho_{\mathbf{L}}^{\ \ i,j^0 \ \ \ i,j^0 \ \ 1}\mathsf{E}[\|\mathbf{x}_{\ \ i,j^0 \ \ 1}\|^2] \\ &\leq \sqrt{ns}\tau_{\mathbf{L}}\rho_{\mathbf{L}}^{L}(c_{\mathbf{x}}^2\|\mathbf{\Sigma_w}\|\log(T)), \end{aligned} \tag{74}$$

where the second inequality follows from $\tau_{i;j^0} - \tau_{i;j^0 \ \ 1} \geq L$ and $\|\mathbf{x}_{\ \ i,j^0 \ \ 1}\| \leq c_{\mathbf{x}}\sqrt{\|\mathbf{\Sigma_w}\|\log(T)}$ by definition of S_i. Then, by Markov inequality, we have

$$\mathsf{P}\left(\|\mathbf{x}_{\ \ i,j^0} - \bar{\mathbf{x}}_{\ \ i,j^0}\| \leq \frac{c_{\mathbf{x}}\sqrt{\|\mathbf{\Sigma_w}\|\log(T)}}{6} \right) \geq 1 - 36\sqrt{ns}\tau_{\mathbf{L}}\rho_{\mathbf{L}}^{L}. \tag{75}$$

Then, using union bound, we have

$$\begin{aligned} &\mathsf{P}\left(\bigcap_{i2[s]}\bigcap_{j^0} \left\{ \|\mathbf{x}_{\ \ i,j^0} - \bar{\mathbf{x}}_{\ \ i,j^0}\| \leq \frac{c_{\mathbf{x}}\sqrt{\|\mathbf{\Sigma_w}\|\log(T)}}{6} \right\} \right) \\ &\geq 1 - 36\sqrt{ns}^{1:5}|R_i|\tau_{\mathbf{L}}\rho_{\mathbf{L}}^{L} \\ &\geq 1 - 36\sqrt{ns}^{1:5}T\tau_{\mathbf{L}}\rho_{\mathbf{L}}^{L} \\ &\geq 1 - \frac{\delta}{2}. \end{aligned} \tag{76}$$

where the last line follows from $L = C_{sub}\log(T)$ and $C_{sub} \geq \underline{C}_{sub;\mathbf{x}}\,(\delta, T, \rho_{\mathbf{L}}, \tau_{\mathbf{L}})$ in the assumption. Note that $\|\bar{\mathbf{x}}_{\ \ i,j}\| = \|\bar{\mathbf{x}}_{\ \ i,j^0}\| \leq \|\mathbf{x}_{\ \ i,j^0}\| + \|\mathbf{x}_{\ \ i,j^0} - \bar{\mathbf{x}}_{\ \ i,j^0}\|$. This together with (73) and (76) gives, with probability at least $1 - {}_{2}$,

$$\bigcap_{i2[s]}\ \bigcap_{j2[jR_ij]} \left\{ \|\bar{\mathbf{x}}_{\ \ i,j}\| \leq \frac{c_{\mathbf{x}}\sqrt{\|\mathbf{\Sigma_w}\|\log(T)}}{2} \right\}. \tag{77}$$

This implies for any i, for any $\tau_k \in R_i$, we have $\tau_k \in \bar{R}_i$, i.e. $R_i \subseteq \bar{R}_i$. Thus, we have $R_i = \bar{\mathsf{R}}_i$ and $|R_i| = |\bar{R}_i|$. Combining with (72), we have with probability at least $1-\delta$,

$$\bigcap_{i=1}^{s} \left\{ |\bar{R}_i| \geq \frac{T\pi_{\min}}{2C_{sub}\log(T)} \right\}. \tag{78}$$

Finally, we could conclude the proof by noticing $|\bar{S}_i| \geq |\bar{R}_i|$. □

B.3. Estimation of $\mathbf{A}_{1:s}$ and $\mathbf{B}_{1:s}$ from a Single Trajectory (Main SYSID Analysis)

In this section we estimate the MJS dynamics $\mathbf{A}_{1:s}$ and $\mathbf{B}_{1:s}$ from finite samples obtained from a single trajectory of (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 $\mathbf{x}_0 \sim \mathcal{D}_x$ and the dynamics evolve with a given input, we can record the resulting state, excitation and mode observations. Let $\mathbf{K}_{1:s}$ stabilizes the system (1) in the mean square sense according to Definition 1. Then, choosing the input to be $\mathbf{u}_t = \mathbf{K}_{\omega(t)}\mathbf{x}_t + \mathbf{z}_t$, the state updates as follows,

$$\mathbf{x}_{t+1} = (\mathbf{A}_{\omega(t)} + \mathbf{B}_{\omega(t)}\mathbf{K}_{\omega(t)})\mathbf{x}_t + \mathbf{B}_{\omega(t)}\mathbf{z}_t + \mathbf{w}_t = \mathbf{L}_{\omega(t)}\mathbf{x}_t + \mathbf{B}_{\omega(t)}\mathbf{z}_t + \mathbf{w}_t, \tag{79}$$

where $\{\mathbf{z}_t\}_{t=0}^{\infty} \overset{\text{i.i.d.}}{\sim} \mathcal{N}(0, \sigma_z^2 \mathbf{I}_p)$ is the i.i.d. excitation for exploration and we let $\mathbf{L}_{\omega(t)} := \mathbf{A}_{\omega(t)} + \mathbf{B}_{\omega(t)}\mathbf{K}_{\omega(t)}$. Observe that the closed-loop state update (79) can be expanded as follows,

$$\mathbf{x}_t = \begin{cases} \mathbf{x}_0 & \text{if} \quad t=0, \\ \mathbf{L}_{\omega(0)}\mathbf{x}_0 + \mathbf{B}_{\omega(0)}\mathbf{z}_0 + \mathbf{w}_0 & \text{if} \quad t=1, \\ \prod_{j=0}^{t-1}\mathbf{L}_{\omega(j)}\mathbf{x}_0 + \sum_{\tau=0}^{t-2}\prod_{j=\tau+1}^{t-1}\mathbf{L}_{\omega(j)}\mathbf{B}_{\omega(\tau)}\mathbf{z}_\tau + \mathbf{B}_{\omega(t-1)}\mathbf{z}_{t-1} & \\ \qquad + \sum_{\tau=0}^{t-2}\prod_{j=\tau+1}^{t-1}\mathbf{L}_{\omega(j)}\mathbf{w}_\tau + \mathbf{w}_{t-1} & \text{if} \quad t \geq 2, \end{cases} \tag{80}$$

B.3.1. Estimation from Bounded States

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

Assumption A3 (subGaussian noise). *Let $\{\mathbf{w}_t\}_{t=0}^{T-1} \overset{i.i.d.}{\sim} \mathcal{D}_w$. There exists $\sigma_w > 0$ and $c_w \geq 1$ such that, each entry of $\mathbf{w}_t$ is i.i.d. zero-mean subGaussian with variance σ_w^2 and we have $\|\mathbf{w}_t\|_\infty \leq c_w\sigma_w$.*

Later on we will relax this assumption to get the estimation error bounds with Gaussian noise. To proceed, we first show that the Euclidean norm of the states $\mathbf{x}_t$ in (79) can be upper bounded in expectation. The following result, which is a corollary of Lemma 2, accomplishes this.

Corollary 5 (Bounded states). *Let $\mathbf{x}_t$ be the state at time t of the MJS given by (79), with initial state $\mathbf{x}_0 \sim \mathcal{D}_x$ such that $\mathsf{E}[\mathbf{x}_0] = 0$, $\mathsf{E}[\|\mathbf{x}_0\|_2^2] \leq \beta_0^2 n$. Suppose Assumption A1 on the system and the Markov chain and Assumption A3 on the process noise hold. Suppose $\{\mathbf{z}_t\}_{t=0}^{\infty} \overset{i.i.d.}{\sim} \mathcal{N}(0, \sigma_z^2\mathbf{I}_p)$. Let $c_z := \frac{\sigma_z}{\sigma_w}$ be a constant and let $\mathbf{h}_t := [\frac{1}{\sigma_w}\mathbf{x}_t^\intercal \ \frac{1}{\sigma_z}\mathbf{z}_t^\intercal]^\intercal$ and define*

$$t_0 := \frac{\log\left((1-\rho_{\mathbf{L}})\beta_0^2 n/(c_w^2\sigma_w^2 + \sigma_z^2\|\mathbf{B}_{1:s}\|)\right)}{1-\rho_{\mathbf{L}}}, \tag{81}$$

$$\beta_+^2 := \frac{2\sqrt{s}(c_w^2 + c_z^2\|\mathbf{B}_{1:s}\|^2)\tau_{\mathbf{L}}}{1-\rho_{\mathbf{L}}}, \tag{82}$$

Then, for all $t_0 \leq t \leq T-1$, we have

$$\mathsf{E}[\|\mathbf{x}_t\|_2^2] \leq \sigma_w^2\beta_+^2 n \quad \textit{and} \quad \mathsf{E}[\|\mathbf{h}_t\|_2^2] \leq (1+\beta_+^2)(n+p). \tag{83}$$

Proof. Recall from Lemma 2 that the states $\mathbf{x}_t$ is bounded in expectation as

$$\mathsf{E}[\|\mathbf{x}_t\|_2^2] \leq \tau_{\mathbf{L}}\sqrt{s}\left(\rho_{\mathbf{L}}^t\beta_0^2\sqrt{n} + \frac{c_w^2\sigma_w^2 + \sigma_z^2\|\mathbf{B}_{1:s}\|^2}{1-\rho_{\mathbf{L}}}\right)n,$$

$$\leq \frac{2\sigma_w^2\sqrt{s}(c_w^2 + c_z^2\|\mathbf{B}_{1:s}\|^2)\tau_{\mathbf{L}} n}{1-\rho_{\mathbf{L}}}, \tag{84}$$

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

$$\rho_{\mathbf{L}}^t \leq \frac{c_w^2\sigma_w^2 + \sigma_z^2\|\mathbf{B}_{1:s}\|^2}{(1-\rho_{\mathbf{L}})\beta_0^2\sqrt{n}} \impliedby t \geq t_0 := \frac{\log\big((1-\rho_{\mathbf{L}})\beta_0^2 n/(c_w^2\sigma_w^2 + \sigma_z^2\|\mathbf{B}_{1:s}\|)\big)}{1-\rho_{\mathbf{L}}}. \tag{85}$$

This gives the advertised upper bound on $\mathsf{E}[\|\mathbf{x}_t\|_{\ell_2}^2]$ for $t \geq t_0$. Using Jensen's inequality, this further implies

$$\mathsf{E}[\|\mathbf{x}_t\|_{\ell_2}] \leq \sigma_w\sqrt{\frac{2\sqrt{s}(c_w^2 + c_z^2\|\mathbf{B}_{1:s}\|^2)\tau_{\mathbf{L}} n}{1-\rho_{\mathbf{L}}}} \quad \text{for} \quad t \geq t_0. \tag{86}$$

Next, using standard results on the distribution of squared Euclidean norm of a Gaussian vector, we have $\mathsf{E}[\|\mathbf{z}_t\|_{\ell_2}^2] = \sigma_z^2 p$ for all $t \geq 0$. Combining this with (84) , we get the following upper bound on the squared norm of the concatenated state $\mathbf{h}_t := [\frac{1}{\sigma_w}\mathbf{x}_t^\top \ \frac{1}{\sigma_z}\mathbf{z}_t^\top]^\top$, that is, for $t \geq t_0$, we have

$$\begin{aligned}\mathsf{E}[\|\mathbf{h}_t\|_{\ell_2}^2] = \frac{1}{\sigma_w^2}\mathsf{E}[\|\mathbf{x}_t\|_{\ell_2}^2] + \frac{1}{\sigma_z^2}\mathsf{E}[\|\mathbf{z}_t\|_{\ell_2}^2] &\leq \frac{2\sqrt{s}(c_w^2 + c_z^2\|\mathbf{B}_{1:s}\|^2)\tau_{\mathbf{L}} n}{1-\rho_{\mathbf{L}}} + p, \\ &\leq \left(1 + \frac{2\sqrt{s}(c_w^2 + c_z^2\|\mathbf{B}_{1:s}\|^2)\tau_{\mathbf{L}}}{1-\rho_{\mathbf{L}}}\right)(n+p). \end{aligned} \tag{87}$$

This gives the advertised upper bound on $\mathsf{E}[\|\mathbf{h}_t\|_{\ell_2}^2]$ for $t \geq t_0$. Using Jensen's inequality, this further implies

$$\mathsf{E}[\|\mathbf{h}_t\|_{\ell_2}] \leq \sqrt{\left(1 + \frac{2\sqrt{s}(c_w^2 + c_z^2\|\mathbf{B}_{1:s}\|^2)\tau_{\mathbf{L}}}{1-\rho_{\mathbf{L}}}\right)(n+p)} \quad \text{for} \quad t \geq t_0. \tag{88}$$

This completes the proof. □

Suppose the MJS in (79) is run for T timesteps and we have access to the trajectory $(\mathbf{z}_t, \mathbf{x}_t, \mathbf{x}_{t+1})_{t=0}^{T-1}$. We sub-sample this trajectory to obtain weakly dependent sub-trajectories as follows.

Definition 4 (Sub-trajectories of bounded states)**.** *Let sampling period $L \geq 1$ be an integer. Set the sub-trajectory length $N = \lfloor\frac{T-L}{L}\rfloor$. We sub-sample the trajectory $(\mathbf{z}_t, \mathbf{x}_t, \mathbf{x}_{t+1})_{t=0}^{T-1}$ at points $\tau + L, \tau + 2L, \ldots, \tau + NL$ to get the τ_{th} sub-trajectory $(\mathbf{z}_{(j)}, \mathbf{x}_{(j)}, \mathbf{x}_{(j)+1})_{j=1}^N$, defined as*

$$(\mathbf{z}_{(j)}, \mathbf{x}_{(j)}, \mathbf{x}_{(j)+1}) := (\mathbf{z}_{\tau+jL}, \mathbf{x}_{\tau+jL}, \mathbf{x}_{\tau+jL+1}) \quad \textit{for} \quad j = 1, \ldots, N, \tag{89}$$

where $0 \leq \tau \leq L-1$ is a fixed offset. Given the τ_{th} sub-trajectory $(\mathbf{z}_{(j)}, \mathbf{x}_{(j)}, \mathbf{x}_{(j)+1})_{j=1}^N$, for each $i \in [s]$, we further subsample at $\|\mathbf{x}_{(j)}\|_{\ell_2} \leq c\sigma_w\beta_+\sqrt{n}$, $\|\mathbf{z}_{(j)}\|_{\ell_2} \leq c\sigma_z\sqrt{p}$ and $\omega(\tau + jL) = i$ to obtain $(\mathbf{z}_{(j_k)}, \mathbf{x}_{(j_k)}, \mathbf{x}_{(j_k)+1})_{k=1}^{N_i}$, at timesteps $\tau + j_1 L, \tau + j_2 L, \ldots, \tau + j_{N_i} L$, where j_k can take values between 1 to N and $j_k < j_\ell$ when $k < \ell$.

For notational convenience, we also denote the noise and the mode observation at time $\tau + j_k L$ as $\mathbf{w}_{(j_k)}$ and $\omega_{(j_k)}$ respectively. From Def. 4 it is clear that, $(\mathbf{z}_{(j_k)}, \mathbf{x}_{(j_k)}, \mathbf{x}_{(j_k)+1})_{k=1}^{N_i}$ is a sub-trajectory of $(\mathbf{z}_{(j)}, \mathbf{x}_{(j)}, \mathbf{x}_{(j)+1})_{j=1}^N$ and has the following covariance properties.

Lemma 10 (Covariance of sampled states)**.** *Consider the setup of Lemma 5. Let t_0 and β_+ be as in* (81) *and* (82) *respectively. Let $c \geq 6$ be a fixed constant and let $\mathbf{h}_t := [\frac{1}{\sigma_w}\mathbf{x}_t^\top \ \frac{1}{\sigma_z}\mathbf{z}_t^\top]^\top$. We have*

$$(\sigma_w^2/2)\mathbf{I}_n \preceq \boldsymbol{\Sigma}[\mathbf{x}_t \mid \|\mathbf{x}_t\|_{\ell_2} \leq c\sigma_w\beta_+\sqrt{n}] \preceq 2\sigma_w^2\beta_+^2 n\mathbf{I}_n. \tag{90}$$

$$(1/2)\mathbf{I}_{n+p} \preceq \boldsymbol{\Sigma}[\mathbf{h}_t \mid \|\mathbf{x}_t\|_{\ell_2} \leq c\sigma_w\beta_+\sqrt{n}, \|\mathbf{z}_t\|_{\ell_2} \leq c\sigma_z\sqrt{p}] \preceq 2(1+\beta_+^2)(n+p)\mathbf{I}_{n+p}. \tag{91}$$

Proof. Let $\mathbf{x}_t$ be the state at time t of the MJS given by (79), with initial state $\mathbf{x}_0 \sim \mathcal{D}_x$ such that $\mathsf{E}[\mathbf{x}_0] = 0$ and define the noise-removed state $\tilde{\mathbf{x}}_t = \mathbf{x}_t - \mathbf{w}_{t-1}$ which is independent of $\mathbf{w}_{t-1}$. Observe that $\mathsf{E}[\tilde{\mathbf{x}}_t] = 0$ because both $\mathbf{x}_t$ and $\mathbf{w}_{t-1}$ are zero-mean. Next, from Corollary 5, we know that $\mathbb{E}[\|\mathbf{x}_t\|_{\ell_2}] \leq \sigma_w \beta_+ \sqrt{n}$. Combining this with $\mathbb{E}[\|\mathbf{w}_{t-1}\|_{\ell_2}] \leq c_w \sigma_w \sqrt{n}$, we have

$$\mathbb{E}[\|\tilde{\mathbf{x}}_t\|_{\ell_2}] \leq \mathbb{E}[\|\mathbf{x}_t\|_{\ell_2}] + \mathbb{E}[\|\mathbf{w}_{t-1}\|_{\ell_2}] \leq 2\sigma_w \beta_+ \sqrt{n} \tag{92}$$

To proceed, consider the conditional random variable

$$\mathbf{y}_t \sim \{\mathbf{x}_t \mid \|\mathbf{x}_t\|_{\ell_2} \leq c\sigma_w \beta_+ \sqrt{n}\} \tag{93}$$

Then using law of total probability, we have

$$\begin{aligned} \|\mathbf{\Sigma}[\mathbf{y}_t]\| = \|\mathbb{E}[\mathbf{y}_t \mathbf{y}_t^\top]\| \leq \mathbb{E}[\|\mathbf{y}_t\|_{\ell_2}^2] &= \mathbb{E}[\|\mathbf{x}_t\|_{\ell_2}^2 \mid \|\mathbf{x}_t\|_{\ell_2} \leq c\sigma_w \beta_+ \sqrt{n}], \\ &\leq \frac{\mathbb{E}[\|\mathbf{x}_t\|_{\ell_2}^2]}{\mathbb{P}(\|\mathbf{x}_t\|_{\ell_2} \leq c\sigma_w \beta_+ \sqrt{n})}, \\ &\leq 2\sigma_w^2 \beta_+^2 n, \end{aligned} \tag{94}$$

where we get the final upper bound from the Markov inequality $\mathbb{P}(\|\mathbf{x}_t\|_{\ell_2} \leq c\sigma_w \beta_+ \sqrt{n}) \geq 1 - 1/c$ and the assumption that $c \geq 6$. To lower bound the covariance matrix $\mathbf{\Sigma}[\mathbf{y}_t]$, observe that $\|\mathbf{w}_{t-1}\|_{\ell_2} \leq c_w \sigma_w \sqrt{n} \leq \sigma_w \beta_+ \sqrt{n}$, where β_+ is given by (82). Therefore, we can use Lemma 4 from Section A.3 to lower bound $\mathbf{\Sigma}[\mathbf{y}_t]$ as follows,

$$\mathbf{\Sigma}[\mathbf{y}_t] = \mathbf{\Sigma}[\mathbf{x}_t \mid \|\mathbf{x}_t\|_{\ell_2} \leq c\sigma_w \beta_+ \sqrt{n}] \succeq (\sigma_w^2/2)\mathbf{I}_n. \tag{95}$$

Combining (94) and (95) we get the statement of the lemma for $\mathbf{\Sigma}[\mathbf{x}_t \mid \|\mathbf{x}_t\|_{\ell_2} \leq c\sigma_w \beta_+ \sqrt{n}]$. Using a similar argument with Lemma 5, we can show that, when $c \geq 6$, we have

$$(\sigma_z^2/2)\mathbf{I}_p \preceq \mathbf{\Sigma}[\mathbf{z}_t \mid \|\mathbf{z}_t\|_{\ell_2} \leq c\sigma_z \sqrt{p}] \preceq (2\sigma_z^2 p)\mathbf{I}_p. \tag{96}$$

Finally, combing the derived bounds on $\mathbf{\Sigma}[\mathbf{x}_t \mid \|\mathbf{x}_t\|_{\ell_2} \leq c\sigma_w \beta_+ \sqrt{n}]$ and $\mathbf{\Sigma}[\mathbf{z}_t \mid \|\mathbf{z}_t\|_{\ell_2} \leq c\sigma_z \sqrt{p}]$, we have

$$(1/2)\mathbf{I}_{n+p} \preceq \mathbf{\Sigma}[\mathbf{h}_t \mid \|\mathbf{x}_t\|_{\ell_2} \leq c\sigma_w \beta_+ \sqrt{n}, \|\mathbf{z}_t\|_{\ell_2} \leq c\sigma_z \sqrt{p}] \preceq 2(1 + \beta_+^2)(n+p)\mathbf{I}_{n+p}. \tag{97}$$

This completes the proof. □

To proceed, let $\mathbf{h}_{(j_k)} := [\frac{1}{\sigma_w}\mathbf{x}_{(j_k)}^\top \ \frac{1}{\sigma_z}\mathbf{z}_{(j_k)}^\top]^\top$ and $\mathbf{\Theta}_i^\star := [\sigma_w \mathbf{L}_i \ \ \sigma_z \mathbf{B}_i]$ for $i \in [s]$. Then the output of each sample $(\mathbf{z}_{(j_k)}, \mathbf{x}_{(j_k)}, \mathbf{x}_{(j_k)+1})$ is related to the inputs as follows,

$$\mathbf{x}_{(j_k)+1} = \mathbf{\Theta}_{\omega(j_k)}^\star \mathbf{h}_{(j_k)} + \mathbf{w}_{(j_k)} \quad \text{for} \quad k = 1, \dots, N_i \ \text{ and } \ j_k \in [1, N]. \tag{98}$$

To carry out finite sample identification of $\mathbf{\Theta}_i^\star$ using the method of linear least squares, we define the following concatenated matrices,

$$\mathbf{Y}_i = \begin{bmatrix} \mathbf{x}_{(j_1)+1}^\top \\ \mathbf{x}_{(j_2)+1}^\top \\ \vdots \\ \mathbf{x}_{(j_{N_i})+1}^\top \end{bmatrix}, \quad \mathbf{H}_i = \begin{bmatrix} \mathbf{h}_{(j_1)}^\top \\ \mathbf{h}_{(j_2)}^\top \\ \vdots \\ \mathbf{h}_{(j_{N_i})}^\top \end{bmatrix}, \quad \mathbf{W}_i = \begin{bmatrix} \mathbf{w}_{(j_1)}^\top \\ \mathbf{w}_{(j_2)}^\top \\ \vdots \\ \mathbf{w}_{(j_{N_i})}^\top \end{bmatrix}. \tag{99}$$

Then, we have $\mathbf{Y}_i = \mathbf{H}_i \mathbf{\Theta}_i^{\star\top} + \mathbf{W}_i$. Our goal in this paper is to solve the following least squares problems,

$$\hat{\mathbf{\Theta}}_i^\top = \arg\min_{\mathbf{\Theta}_i} \frac{1}{2N_i}\|\mathbf{Y}_i - \mathbf{H}_i \mathbf{\Theta}_i^\top\|_F^2. \tag{100}$$

The least squares estimator of $\mathbf{\Theta}_i^{\star\top}$ is $\hat{\mathbf{\Theta}}_i^\top = \mathbf{H}_i^\dagger \mathbf{Y}_i = (\mathbf{H}_i^\top \mathbf{H}_i)^{-1}\mathbf{H}_i^\top \mathbf{Y}_i$ and the estimation error is given by

$$\|\hat{\mathbf{\Theta}}_i - \mathbf{\Theta}_i^\star\| = \|(\mathbf{H}_i^\top \mathbf{H}_i)^{-1}\mathbf{H}_i^\top \mathbf{W}_i\| \leq \|(\mathbf{H}_i^\top \mathbf{H}_i)^{-1}\|\|\mathbf{H}_i^\top \mathbf{W}_i\| = \frac{\|\mathbf{H}_i^\top \mathbf{W}_i\|}{\lambda_{\min}(\mathbf{H}_i^\top \mathbf{H}_i)}. \tag{101}$$

If we can get an upper bound on $\|\mathbf{H}_i^\top \mathbf{W}_i\|$ and a lower bound on $\lambda_{\min}(\mathbf{H}_i^\top \mathbf{H}_i)$, we can use (101) to give an upper bound on the estimation error $\|\hat{\mathbf{\Theta}}_i - \mathbf{\Theta}_i^\star\|$ as well. However, because $\mathbf{H}_i$ has non-i.i.d. rows, we cannot directly get such bounds. To resolve this issue, we rely on the notion of stability and use perturbation based techniques to indirectly obtain these bounds.

B.3.2. Estimation from Truncated States

Definition 5 (Truncated state vector(Oymak, 2019)). *Consider the state equation* (79)*. Given $t \geq L > 0$, the L-truncation of $\mathbf{x}_t$ is denoted by $\mathbf{x}_{t;L}$ and is obtained by driving the system with excitation $\mathbf{z}'$ and additive noise $\mathbf{w}'$ until time t, where*

$$\mathbf{v}' = \begin{cases} 0 & \text{if} \quad \tau < t - L \\ \mathbf{v} & \text{else} \end{cases}. \tag{102}$$

In words, the L-truncated state vector $\mathbf{x}_{t;L}$ is obtained by unrolling $\mathbf{x}_t$ until time $t - L$ and setting $\mathbf{x}_{t-L}$ 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 (80), we have

$$\mathbf{x}_t - \mathbf{x}_{t;L} = \prod_{j=t-L}^{t-1} \mathbf{L}_{\omega(j)} \mathbf{x}_{t-L}. \tag{103}$$

As a corollary of Lemma 2, observe that for a closed loop autonomous system $\mathbf{x}_{t+1} = \mathbf{L}_{\omega(t)} \mathbf{x}_t$, mean-square stability implies that, for any initial conditions $\mathbf{x}_0$ and ω_0, we have $\mathbb{E}[\|\mathbf{x}_t\|_{\ell_2}^2] \leq \sqrt{ns}\tau_{\mathbf{L}}\rho_{\mathbf{L}}^t \|\mathbf{x}_0\|_{\ell_2}^2$. Combining this argument with (103), we have

$$\begin{aligned} \mathbb{E}[\|\mathbf{x}_t - \mathbf{x}_{t;L}\|_{\ell_2}^2] = \mathbb{E}[\|\prod_{j=t-L}^{t-1} \mathbf{L}^+_{\omega(j)} \mathbf{x}_{t-L}\|_{\ell_2}^2] \leq \sqrt{ns}\tau_{\mathbf{L}}\rho_{\mathbf{L}}^L \|\mathbf{x}_{t-L}\|_{\ell_2}^2, \\ \implies \mathbb{E}[\|\mathbf{x}_t - \mathbf{x}_{t;L}\|_{\ell_2}] \leq \sqrt{(ns)^{1/2}\tau_{\mathbf{L}}\rho_{\mathbf{L}}^L} \|\mathbf{x}_{t-L}\|_{\ell_2} \leq \sqrt{ns}\tau_{\mathbf{L}}\rho_{\mathbf{L}}^{L/2} \|\mathbf{x}_{t-L}\|_{\ell_2}, \end{aligned} \tag{104}$$

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 $\|\mathbf{x}_{t-L}\|_{\ell_2} \leq c\sigma_w\beta_+\sqrt{n}$, then we can make $\mathbb{E}[\|\mathbf{x}_t - \mathbf{x}_{t;L}\|_{\ell_2}]$ arbitrarily small by picking a sufficiently large truncation length $L \geq 1$,

$$\mathbb{E}[\|\mathbf{x}_t - \mathbf{x}_{t;L}\|_{\ell_2} \mid \|\mathbf{x}_{t-L}\|_{\ell_2} \leq c\sigma_w\beta_+\sqrt{n}] \leq \tau_{\mathbf{L}}\rho_{\mathbf{L}}^L \|\mathbf{x}_{t-L}\|_{\ell_2} \leq c\sigma_w\beta_+ n\sqrt{s}\tau_{\mathbf{L}}\rho_{\mathbf{L}}^{L/2}.$$

To proceed, we carry out the truncation of the sub-trajectories introduced in Def. 4 to get the truncated sub-trajectories defined as follows.

Definition 6 (Truncated sub-trajectories). *Consider τ_{th} sub-trajectory $(\mathbf{z}_{(j_k)}, \mathbf{x}_{(j_k)}, \mathbf{x}_{(j_k)+1})_{k=1}^{N_i}$ from Def. 4. We truncate the states by $(j_k - j_{k-1})L - 1$, to get the τ_{th} truncated sub-trajectory $(\mathbf{z}_{(j_k)}, \bar{\mathbf{x}}_{(j_k)}, \bar{\mathbf{x}}_{(j_k)+1})_{k=1}^{N_i}$, defined as*

$$(\mathbf{z}_{(j_k)}, \bar{\mathbf{x}}_{(j_k)}, \bar{\mathbf{x}}_{(j_k)+1}) := (\mathbf{z}_{\tau+j_kL}, \mathbf{x}_{\tau+j_kL;(j_k-j_{k-1})L-1}, \mathbf{x}_{\tau+j_kL+1;(j_k-j_{k-1})L}), \tag{105}$$

for $k = 1, \dots, N_i$, where j_k can take values between 1 to N and we set $j_0 = 0$.

Next we show that if the sampling period $L \geq 1$ is sufficiently large enough, then the truncated states $\{\bar{\mathbf{x}}_{(j_k)}\}_{k=1}^{N_i}$ as well as the difference between the truncated and non-truncated states are bounded with high probability over the modes.

Lemma 11 (Bounded states (truncated)). *Consider the setup of Lemma 5. Let $\{\mathbf{x}_{(j_k)}\}_{k=1}^{N_i}$ and $\{\bar{\mathbf{x}}_{(j_k)}\}_{k=1}^{N_i}$ be the (truncated) states from Def. 4 and 6 respectively. Let t_0 and β_+ be as in* (81) *and* (82) *respectively. Let $c_z := \frac{\sigma_z}{\sigma_w}$ be a constant and*

$$\beta'_+ := c_w + \beta_+\|\mathbf{L}_{1:s}\| + c_z\sqrt{p/n}\|\mathbf{B}_{1:s}\|, \tag{106}$$

$$L_{tr1}(\rho_{\mathbf{L}}, \delta) := 1 + \frac{2\log\left(2\sqrt{ns}\tau_{\mathbf{L}}T\beta'_+/(\beta_+\delta)\right)}{1 - \rho_{\mathbf{L}}}, \tag{107}$$

$$\text{and} \quad L \geq \max\{t_0, L_{tr1}(\rho_{\mathbf{L}}, \delta)\} \tag{108}$$

Then, with probability at least $1 - \delta$ over the modes, we have

$$\|\mathbf{x}_{(j_k)} - \bar{\mathbf{x}}_{(j_k)}\|_{\ell_2} \leq (1/2)c\sigma_w\beta_+\sqrt{n} \quad \text{for} \quad k = 1, \dots, N_i \;\; \text{and} \;\; i \in [s], \tag{109}$$

$$\text{and} \quad \|\bar{\mathbf{x}}_{(j_k)}\|_{\ell_2} \leq (3/2)c\sigma_w\beta_+\sqrt{n} \quad \text{for} \quad k = 1, \dots, N_i \;\; \text{and} \;\; i \in [s]. \tag{110}$$

Proof. To begin, using Assumption A1 and (104), the impact of truncation is bounded in expectation over the modes as follows,

$$\begin{aligned}
\mathbb{E}[\|\mathbf{x}_{(j_k)} - \bar{\mathbf{x}}_{(j_k)}\|_{\ell_2}] &= \mathbb{E}[\|\mathbf{x}_{\tau+j_kL} - \mathbf{x}_{\tau+j_kL,(j_k-j_{k-1})L-1}\|_{\ell_2}], \\
&\leq \sqrt{ns}\tau_{\mathbf{L}}\rho_{\mathbf{L}}^{((j_k-j_{k-1})L-1)/2}\|\mathbf{x}_{\tau+j_{k-1}L+1}\|_{\ell_2}, \\
&\leq \sqrt{ns}\tau_{\mathbf{L}}\rho_{\mathbf{L}}^{(L-1)/2}\|\mathbf{x}_{\tau+j_{k-1}L+1}\|_{\ell_2}, \qquad (111)
\end{aligned}$$

where the expectation is over the Markov modes at timesteps $\tau + j_{k-1}L + 1, \tau + j_{k-1}L + 2, \dots \tau + j_kL - 1$ and we used the fact that $L \leq (j_k - j_{k-1})L$ from Def. 6. Furthermore, observe that $(\mathbf{z}_{(j_k)}, \mathbf{x}_{(j_k)})_{k=1}^{N_i}$ are the bounded samples. Therefore, we have

$$\begin{aligned}
\|\mathbf{x}_{\tau+j_kL+1}\|_{\ell_2} &= \|\mathbf{L}_{\omega(\tau+j_kL)}\mathbf{x}_{\tau+j_kL} + \mathbf{B}_{\omega(\tau+j_kL)}\mathbf{z}_{\tau+j_kL} + \mathbf{w}_{\tau+j_kL}\|_{\ell_2}, \\
&\leq \max_{i\in[s]}\|\mathbf{L}_i\|\|\mathbf{x}_{\tau+j_kL}\|_{\ell_2} + \max_{i\in[s]}\|\mathbf{B}_i\|\|\mathbf{z}_{\tau+j_kL}\|_{\ell_2} + \|\mathbf{w}_{\tau+j_kL}\|_{\ell_2}, \\
&\leq c\sigma_w\beta_+\sqrt{n}\|\mathbf{L}_{1:s}\| + c\sigma_z\sqrt{p}\|\mathbf{B}_{1:s}\| + c_w\sigma_w\sqrt{n}, \\
&\leq c\sigma_w(c_w + \beta_+\|\mathbf{L}_{1:s}\| + c_z\sqrt{p/n}\|\mathbf{B}_{1:s}\|)\sqrt{n}, \\
&= c\sigma_w\beta'_+\sqrt{n}, \qquad (112)
\end{aligned}$$

where we set $\beta'_+ := c_w + \beta_+\|\mathbf{L}_{1:s}\| + c_z\sqrt{p/n}\|\mathbf{B}_{1:s}\|$. Combining (112) with (111), for all $k = 1, 2\dots, N_i$ and $i \in [s]$, we have

$$\mathbb{E}[\|\mathbf{x}_{(j_k)} - \bar{\mathbf{x}}_{(j_k)}\|_{\ell_2}] \leq c\sigma_w\beta'_+ n\sqrt{s}\tau_{\mathbf{L}}\rho_{\mathbf{L}}^{(L-1)/2}, \qquad (113)$$

$$\implies \mathbb{P}(\|\mathbf{x}_{(j_k)} - \bar{\mathbf{x}}_{(j_k)}\|_{\ell_2} \leq c\sigma_w\beta'_+ Tn\sqrt{s}\tau_{\mathbf{L}}\rho_{\mathbf{L}}^{(L-1)/2}/\delta) \geq 1 - \delta, \qquad (114)$$

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, we have

$$\begin{aligned}
\|\bar{\mathbf{x}}_{(j_k)}\|_{\ell_2} \leq \|\mathbf{x}_{(j_k)}\|_{\ell_2} + \|\mathbf{x}_{(j_k)} - \bar{\mathbf{x}}_{(j_k)}\|_{\ell_2} &\leq c\sigma_w\beta_+\sqrt{n} + c\sigma_w\beta'_+ Tn\sqrt{s}\tau_{\mathbf{L}}\rho_{\mathbf{L}}^{(L-1)/2}/\delta, \\
&\leq (3/2)c\sigma_w\beta_+\sqrt{n}. \qquad (115)
\end{aligned}$$

where we get the last inequality by choosing L via

$$\begin{aligned}
c\sigma_w\beta'_+ Tn\sqrt{s}\tau_{\mathbf{L}}\rho_{\mathbf{L}}^{(L-1)/2}/\delta \leq c\sigma_w\beta_+\sqrt{n}/2 \iff \rho_{\mathbf{L}}^{(L-1)/2} &\leq \frac{\delta\beta_+}{2\sqrt{ns}\tau_{\mathbf{L}}T\beta'_+}, \\
\impliedby L &\geq 1 + 2\frac{\log\left(2\sqrt{ns}\tau_{\mathbf{L}}T\beta'_+/(\beta_+\delta)\right)}{1-\rho_{\mathbf{L}}}. \qquad (116)
\end{aligned}$$

This also implies that, with probability at least $1-\delta$ over the modes, we have $\|\mathbf{x}_{(j_k)} - \bar{\mathbf{x}}_{(j_k)}\|_{\ell_2} \leq (1/2)c\sigma_w\beta_+\sqrt{n}$. This completes the proof. ☐

The following lemma states that the τ_{th} truncated sub-trajectory $(\mathbf{z}_{(j_k)}, \bar{\mathbf{x}}_{(j_k)}, \bar{\mathbf{x}}_{(j_k)+1})_{k=1}^{N_i}$ has independent samples.

Lemma 12 (Conditional independence)**.** *Suppose $\{\mathbf{z}_t\}_{t=0}^{\infty} \overset{i.i.d.}{\sim} \mathcal{N}(0, \sigma_z^2\mathbf{I}_p)$ and $\{\mathbf{w}_t\}_{t=0}^{\infty} \overset{i.i.d.}{\sim} \mathcal{D}_w$ satisfies Assumption A3. Let $\mathcal{X} := \{\mathbf{x}_{(j_k)}\}_{k=1}^{N_i}$ be the set of all states obtained by sampling the states bounded by $c\sigma_w\beta_+\sqrt{n}$ and $\mathcal{X}' := \{\mathbf{x}_{(l_k)}\}_{k=1}^{N'_i}$ be the set of all states such that $\bar{\mathbf{x}}_{(l_k)}$ are upper bounded by $(1/2)c\sigma_w\beta_+\sqrt{n}$. Let the sampling period L obeys* (108)*. Then, with probability at least $1-\delta$ over the modes, for all $i \in [s]$, the set $\mathcal{X}'$ is a subset of $\mathcal{X}$. Secondly, conditioned on the modes, the truncated states $\{\bar{\mathbf{x}}_{(l_k)}\}_{k=1}^{N'_i}$ obtained by truncating $\mathcal{X}'$ are all independent of each other. Moreover, $\{\bar{\mathbf{x}}_{(l_k)}\}_{k=1}^{N'_i}, \{\mathbf{z}_{(l_k)}\}_{k=1}^{N'_i}, \{\mathbf{w}_{(l_k)}\}_{k=1}^{N'_i}$ are all independent of each other.*

Proof. By construction $\bar{\mathbf{x}}_{(j_k)}$ only depends on the vectors $\{\mathbf{z}_t, \mathbf{w}_t\}_{t=\tau+j_{k-1}L+1}^{\tau+j_kL-1}$. Note that the dependence ranges $[\tau + j_{k-1}L + 1, \tau + j_kL - 1]$ are disjoint intervals for each $k's$. Hence, $\{\bar{\mathbf{x}}_{(j_k)}\}_{k=1}^{N_i}$ should all be independent of each other.

However, this is not the case because $\{\bar{\mathbf{x}}_{(j_k)}\}_{k=1}^{N_i}$ are obtained by truncating only bounded states $\{\mathbf{x}_{(j_k)}\}_{k=1}^{N_i}$. Therefore, we will look for a subset of independent truncated states within $\{\mathbf{x}_{(j_k)}\}_{k=1}^{N_i}$ as follows.

Let $\mathcal{X} := \{\mathbf{x}_{(j_k)}\}_{k=1}^{N_i}$ be the set of all states obtained by sampling the states bounded by $c\sigma_w\beta_+\sqrt{n}$. Each state can be written as $\mathbf{x}_{(j_k)} = \bar{\mathbf{x}}_{(j_k)} + \tilde{\mathbf{x}}_{(j_k)}$ where $\tilde{\mathbf{x}}_{(j_k)}$ captures the impact of the past states at timestep $\tau + j_{k-1}L + 1$. When the sampling period L obeys (108), then from Lemma 11, with probability at least $1-\delta$ over the modes, all $\tilde{\mathbf{x}}_{(j_k)}$ are upper bounded by $(1/2)c\sigma_w\beta_+\sqrt{n}$. Let $\mathcal{X}' := \{\mathbf{x}_{(l_k)}\}_{k=1}^{N_i'}$ be the set of all states such that $\bar{\mathbf{x}}_{(l_k)}$ are upper bounded by $(1/2)c\sigma_w\beta_+\sqrt{n}$. Then, using Lemma 11, with probability at least $1-\delta$ over the modes, for all $k = 1,\dots,N_i'$ and all $i \in [s]$, we have,

$$\|\mathbf{x}_{(l_k)}\|_{\ell_2} \le \|\bar{\mathbf{x}}_{(l_k)}\|_{\ell_2} + \|\tilde{\mathbf{x}}_{(l_k)}\|_{\ell_2} \le c\sigma_w\beta_+\sqrt{n}. \tag{117}$$

As a result, with probability at least $1-\delta$ over the modes, the set $\mathcal{X}'$ is a subset of $\mathcal{X}$. Secondly, conditioned on the modes, truncated states $\{\bar{\mathbf{x}}_{(l_k)}\}_{k=1}^{N_i'}$ obtained by truncating $\mathcal{X}'$ are all independent of each other.

To show the independence of $\{\bar{\mathbf{x}}_{(l_k)}\}_{k=1}^{N_i'}$ and $\{\mathbf{z}_{(l_k)}\}_{k=1}^{N_i'}$, observe that the inputs $\mathbf{z}_{(l_k)} = \mathbf{z}_{\tau+l_kL}$ have timestamps $\tau + l_kL$; which is not covered by $[\tau + l_{k-1}L+1, \tau + l_kL - 1]$ - the dependence ranges of $\{\bar{\mathbf{x}}_{(l_k)}\}_{k=1}^{N_i'}$. Identical argument shows the independence of $\{\bar{\mathbf{x}}_{(l_k)}\}_{k=1}^{N_i'}$ and $\{\mathbf{w}_{(l_k)}\}_{k=1}^{N_i'}$. Lastly, $\{\mathbf{z}_{(l_k)}\}_{k=1}^{N_i'}$ and $\{\mathbf{w}_{(l_k)}\}_{k=1}^{N_i'}$ are independent of each other by definition. Hence, $\{\bar{\mathbf{x}}_{(l_k)}\}_{k=1}^{N_i'}, \{\mathbf{z}_{(l_k)}\}_{k=1}^{N_i'}, \{\mathbf{w}_{(l_k)}\}_{k=1}^{N_i'}$ are all independent of each other. This completes the proof. □

Next, we state a lemma similar to Lemma 10 to show that the truncated states have nice covariance properties.

Lemma 13 (Covariance of truncated states ($\mathcal{X}'$))**.** *Consider the setup of Lemma 5. Let t_0, β_+ and β_+' be as in* (81)*,* (82) *and* (106) *respectively. Let $c \ge 6$ be a fixed constant and let $\mathcal{X}' := \{\mathbf{x}_{(l_k)}\}_{k=1}^{N_i'}$ be the set of all states such that $\bar{\mathbf{x}}_{(l_k)}$ are upper bounded by $(1/2)c\sigma_w\beta_+\sqrt{n}$ (see Lemma 12) and $\bar{\mathbf{h}}_{(l_k)} := [\frac{1}{\sigma_w}\bar{\mathbf{x}}_{(l_k)}^\top \ \frac{1}{\sigma_z}\mathbf{z}_{(l_k)}^\top]^\top$. Let*

$$L_{cov}(\rho_{\mathbf{L}}) := 1 + \frac{2\log\left(8c^2\beta_+\beta_+' n\sqrt{ns}\tau_{\mathbf{L}}\right)}{1-\rho_{\mathbf{L}}} \tag{118}$$

Suppose, the sampling period L obeys,

$$L \ge \max\{t_0, L_{cov}(\rho_{\mathbf{L}})\}. \tag{119}$$

Then, we have

$$(\sigma_w^2/4)\mathbf{I}_n \preceq \mathbf{\Sigma}[\bar{\mathbf{x}}_{(l_k)}] \preceq 4\sigma_w^2\beta_+^2 n\mathbf{I}_n. \tag{120}$$

$$(1/4)\mathbf{I}_{n+p} \preceq \mathbf{\Sigma}[\bar{\mathbf{h}}_{(l_k)}] \preceq 4(1+\beta_+^2)(n+p)\mathbf{I}_{n+p}. \tag{121}$$

Proof. For ease of notation, we will drop the subscript to write $\mathbf{x}_{(l_k)}, \bar{\mathbf{x}}_{(l_k)}, \mathbf{h}_{(l_k)}$ and $\bar{\mathbf{h}}_{(l_k)}$ as $\mathbf{x}, \bar{\mathbf{x}}, \mathbf{h}$ and $\bar{\mathbf{h}}$ respectively. To begin, we upper bound the difference between the covariance of truncated and non-truncated states as follows,

$$\begin{aligned}
\|\mathbb{E}[\bar{\mathbf{x}}\bar{\mathbf{x}}^\top - \mathbf{x}\mathbf{x}^\top]\| &= \|\mathbb{E}[\bar{\mathbf{x}}\bar{\mathbf{x}}^\top - \mathbf{x}\bar{\mathbf{x}}^\top + \mathbf{x}\bar{\mathbf{x}}^\top - \mathbf{x}\mathbf{x}^\top]\|, \\
&\le \mathbb{E}[\|\bar{\mathbf{x}}\|_{\ell_2}\|\mathbf{x}-\bar{\mathbf{x}}\|_{\ell_2}] + \mathbb{E}[\|\mathbf{x}\|_{\ell_2}\|\mathbf{x}-\bar{\mathbf{x}}\|_{\ell_2}], \\
&\le (1/2)c\sigma_w\beta_+\sqrt{n}\,\mathbb{E}[\|\mathbf{x}-\bar{\mathbf{x}}\|_{\ell_2}] + c\sigma_w\beta_+\sqrt{n}\,\mathbb{E}[\|\mathbf{x}-\bar{\mathbf{x}}\|_{\ell_2}], \\
&\le 2c^2\sigma_w^2\beta_+\beta_+' n\sqrt{ns}\tau_{\mathbf{L}}\rho_{\mathbf{L}}^{(L-1)/2}.
\end{aligned} \tag{122}$$

Combining this with Lemma 10, when $c \ge 6$, we have,

$$\begin{aligned}
\lambda_{\min}(\mathbf{\Sigma}[\bar{\mathbf{x}}]) &\ge \lambda_{\min}(\mathbf{\Sigma}[\mathbf{x}]) - \|\mathbb{E}[\bar{\mathbf{x}}\bar{\mathbf{x}}^\top - \mathbf{x}\mathbf{x}^\top]\|, \\
&\ge \sigma_w^2/2 - 2c^2\sigma_w^2\beta_+\beta_+' n\sqrt{ns}\tau_{\mathbf{L}}\rho_{\mathbf{L}}^{(L-1)/2} \ge \sigma_w^2/4,
\end{aligned} \tag{123}$$

where we get the last inequality by choosing L via,

$$\sigma_w^2/4 \ge 2c^2\sigma_w^2\beta_+\beta_+' n\sqrt{ns}\tau_{\mathbf{L}}\rho_{\mathbf{L}}^{(L-1)/2} \iff \rho_{\mathbf{L}}^{(L-1)/2} \le \frac{1}{8c^2\beta_+\beta_+' n\sqrt{ns}\tau_{\mathbf{L}}},$$

$$\Longleftarrow L \geq 1 + 2\frac{\log\left(8c^2\beta_+\beta_+^0 n\sqrt{ns}\tau_{\mathbf{L}}\right)}{1-\rho_{\mathbf{L}}}. \tag{124}$$

This also implies, we have the following upper bound,

$$\|\mathbf{\Sigma}[\bar{\mathbf{x}}]\| \leq \|\mathbf{\Sigma}[\mathbf{x}]\| + \|\operatorname{E}[\bar{\mathbf{x}}\bar{\mathbf{x}}^\top - \mathbf{x}\mathbf{x}^\top]\| \leq 2\sigma_w^2\beta_+^2 n + \sigma_w^2/4 \leq 4\sigma_w^2\beta_+^2 n. \tag{125}$$

Combing the upper and lower bounds, we get the first statement of the lemma as follows,

$$(\sigma_w^2/4)\mathbf{I}_n \preceq \mathbf{\Sigma}[\bar{\mathbf{x}}_{(l_k)}] \preceq 4\sigma_w^2\beta_+^2 n\mathbf{I}_n. \tag{126}$$

Combining this with (96), we obtain the second statement of the lemma as follows,

$$(1/4)\mathbf{I}_{n+p} \preceq \mathbf{\Sigma}[\bar{\mathbf{h}}_{(l_k)}] \preceq 4(1+\beta_+^2)(n+p)\mathbf{I}_{n+p}. \tag{127}$$

This completes the proof. □

As an intermediate step, consider the estimation of the unknown dynamics from the τ_{th} truncated sub-trajectory $(\mathbf{z}_{(j_k)}, \bar{\mathbf{x}}_{(j_k)}, \bar{\mathbf{x}}_{(j_k)+1})_{k=1}^{N_i}$ given in Def. 6. Let $\bar{\mathbf{h}}_{(j_k)} := [\frac{1}{\sigma_w}\bar{\mathbf{x}}_{(j_k)}^\top \; \frac{1}{\sigma_z}\mathbf{z}_{(j_k)}^\top]^\top$ and $\mathbf{\Theta}_i^\star := [\sigma_w\mathbf{L}_i \;\; \sigma_z\mathbf{B}_i]$ for $i \in [s]$. Then the output of each sample $(\mathbf{z}_{(j_k)}, \bar{\mathbf{x}}_{(j_k)}, \bar{\mathbf{x}}_{(j_k)+1})$ is related to the inputs as follows,

$$\bar{\mathbf{x}}_{(j_k)+1} = \mathbf{\Theta}_{\omega(j_k)}^\star\bar{\mathbf{h}}_{(j_k)} + \mathbf{w}_{(j_k)} \quad \text{for} \quad k = 1,\dots,N_i \;\; \text{and} \;\; j_k \in [1, N]. \tag{128}$$

To carry out finite sample identification of $\mathbf{\Theta}_i^\star$ using the method of linear least squares, we define the following concatenated matrices,

$$\bar{\mathbf{Y}}_i = \begin{bmatrix} \bar{\mathbf{x}}_{(j_1)+1}^\top \\ \bar{\mathbf{x}}_{(j_2)+1}^\top \\ \vdots \\ \bar{\mathbf{x}}_{(j_{N_i})+1}^\top \end{bmatrix}, \quad \bar{\mathbf{H}}_i = \begin{bmatrix} \bar{\mathbf{h}}_{(j_1)}^\top \\ \bar{\mathbf{h}}_{(j_2)}^\top \\ \vdots \\ \bar{\mathbf{h}}_{(j_{N_i})}^\top \end{bmatrix}, \quad \mathbf{W}_i = \begin{bmatrix} \mathbf{w}_{(j_1)}^\top \\ \mathbf{w}_{(j_2)}^\top \\ \vdots \\ \mathbf{w}_{(j_{N_i})}^\top \end{bmatrix}. \tag{129}$$

Then, we have $\bar{\mathbf{Y}}_i = \bar{\mathbf{H}}_i\mathbf{\Theta}_i^{\star\top} + \mathbf{W}_i$. Our goal in this section is to solve the following least squares problem,

$$\hat{\mathbf{\Theta}}_i^\top = \arg\min_{\mathbf{\Theta}_i} \frac{1}{2N_i}\|\bar{\mathbf{Y}}_i - \bar{\mathbf{H}}_i\mathbf{\Theta}_i^\top\|_F^2. \tag{130}$$

The least squares estimator of $\mathbf{\Theta}_i^{\star\top}$ is $\hat{\mathbf{\Theta}}_i^\top = \bar{\mathbf{H}}_i^\dagger\bar{\mathbf{Y}}_i = \left(\bar{\mathbf{H}}_i^\top\bar{\mathbf{H}}_i\right)^{-1}\bar{\mathbf{H}}_i^\top\bar{\mathbf{Y}}_i$ and the estimation error is given by

$$\|\hat{\mathbf{\Theta}}_i - \mathbf{\Theta}_i^\star\| = \|\left(\bar{\mathbf{H}}_i^\top\bar{\mathbf{H}}_i\right)^{-1}\bar{\mathbf{H}}_i^\top\mathbf{W}_i\| \leq \|\left(\bar{\mathbf{H}}_i^\top\bar{\mathbf{H}}_i\right)^{-1}\|\|\bar{\mathbf{H}}_i^\top\mathbf{W}_i\| = \frac{\|\bar{\mathbf{H}}_i^\top\mathbf{W}_i\|}{\lambda_{\min}\left(\bar{\mathbf{H}}_i^\top\bar{\mathbf{H}}_i\right)}. \tag{131}$$

To upper bound the estimation error, we need to lower bound $\lambda_{\min}\left(\bar{\mathbf{H}}_i^\top\bar{\mathbf{H}}_i\right)$ and upper bound $\|\bar{\mathbf{H}}_i^\top\mathbf{W}_i\|$. The following lemma lower bounds the eigenvalues of the empirical covariance matrix $\bar{\mathbf{H}}_i^\top\bar{\mathbf{H}}_i$ and upper bounds the error term $\|\bar{\mathbf{H}}_i^\top\mathbf{W}_i\|$.

Theorem 4 (Bounding $\lambda_{\min}(\bar{\mathbf{H}}_i^\top\bar{\mathbf{H}}_i)$ and $\|\bar{\mathbf{H}}_i^\top\mathbf{W}_i\|$). *Consider the same setup of Lemma 5. Let t_0, β_+, β_+^0, L_{tr1} and L_{cov} be as in* (81), (82), (106), (107) *and* (118) *respectively. Let $C_{sub} > 0$ be a system related constant. Let $C, C_0 > 0$ and $c \geq 6$ be fixed constants. Suppose the sampling period L and the trajectory length T satisfy*

$$L \geq \max\{t_0, L_{tr1}(\rho_{\mathbf{L}}, \delta), L_{cov}(\rho_{\mathbf{L}})\}, \tag{132}$$

$$T \geq 4c^2 C_{sub}(4+\beta_+^2)\log(2s(n+p)/\delta)\log(T)(n+p)/\pi_{\min}.. \tag{133}$$

Let $\mathcal{X} := \{\mathbf{x}_{(j_k)}\}_{k=1}^{N_i}$ and $\mathcal{X}^0 := \{\mathbf{x}_{(l_k)}\}_{k=1}^{N_i^0}$ be as in Lemma 12 and the matrices $\bar{\mathbf{H}}_i$ and $\mathbf{W}_i$ are constructed as in (129). *Suppose, $N_i^0 \geq \pi_{\min}T/(C_{sub}\log(T))$ with probability at least $1-\delta$ for all $i \in [s]$. Then, with probability at least $1-4\delta$, for all $i \in [s]$, we have*

$$\lambda_{\min}(\bar{\mathbf{H}}_i^\top\bar{\mathbf{H}}_i) \geq \frac{\pi_{\min}T}{16C_{sub}\log(T)},$$

$$\textit{and} \quad \|\bar{\mathbf{H}}_i^T\mathbf{W}_i\| \leq c(1+2\beta_+)\sqrt{\frac{T(n+p)}{\log(T)}}\left(C\sigma_w\sqrt{n+p} + C_0\sqrt{\log\left(\frac{2s}{\delta}\right)}\right).$$

Proof. To begin, recall from Lemma 12 that, $\mathcal{X}^0$ is a subset of $\mathcal{X}$ with probability at least $1-\delta$ over the modes. Secondly, conditioned on the modes, the truncated states $\{\bar{\mathbf{x}}_{l_k}\}_{k=1}^{N_i^0}$ obtained by truncating $\mathcal{X}^0$ are all independent of each other. Now, let $\bar{\mathbf{h}}_{(l_k)} := [\frac{1}{\sigma_w}\bar{\mathbf{x}}_{(l_k)}^\top \ \frac{1}{\sigma_z}\mathbf{z}_{(l_k)}^\top]^\top$ and $\bar{\mathbf{H}}_i^0 \in \mathbb{R}^{N_i^0 \times (n+p)}$ has $\bar{\mathbf{h}}_{(l_k)}^\top$'s in its rows. Then, for every mode observation the rows of $\bar{\mathbf{H}}_i^0$ are independent of each other. Moreover, conditioned on the modes, each row of $\bar{\mathbf{H}}_i^0$ is deterministically bounded as follows,

$$\|\bar{\mathbf{h}}_{(l_k)}\|_{\ell_2}^2 \leq \frac{1}{\sigma_w^2}\|\bar{\mathbf{x}}_{(l_k)}\|_{\ell_2}^2 + \frac{1}{\sigma_z^2}\|\mathbf{z}_{(j_k)}\|_{\ell_2}^2 \leq (1/4)c^2\beta_+^2 n + c^2 p, \\ \leq c^2(1+(1/4)\beta_+^2)(n+p). \tag{134}$$

To proceed, from Lemma 13, when $c \geq 6$ and $L \geq \max\{t_0, L_{cov}\}$, then all the rows of $\bar{\mathbf{H}}_i^0$ satisfy the following covariance properties,

$$(1/4)\mathbf{I}_{n+p} \preceq \mathbf{\Sigma}[\bar{\mathbf{h}}_{(l_k)}] \preceq 4(1+\beta_+^2)(n+p)\mathbf{I}_{n+p}. \tag{135}$$

We are now ready to use a corollary of Theorem 5.41 of (Vershynin, 2010) to lower bound the singular values of $\bar{\mathbf{H}}_i^0$. Specifically, using Corollary 3 with $\sigma_{\min} = 1/2$ and $m = c^2(1+(1/4)\beta_+^2)(n+p)$, with probability at least $1-\delta$, for all $i \in [s]$, we have

$$s_{\min}(\bar{\mathbf{H}}_i^0) \geq \frac{\sqrt{N_i^0}}{2} - c\sqrt{(1+(1/4)\beta_+^2)(n+p)\log(2s(n+p)/\delta)} \geq \frac{\sqrt{N_i^0}}{4}, \tag{136}$$

as long as N_i^0 satisfies the following lower bound,

$$\frac{\sqrt{N_i^0}}{4} \geq c\sqrt{(1+(1/4)\beta_+^2)(n+p)\log(2s(n+p)/\delta)} \\ \Longleftarrow \quad N_i^0 \geq 4c^2(4+\beta_+^2)(n+p)\log(2s(n+p)/\delta). \tag{137}$$

• **Lower bounding $\lambda_{\min}(\bar{\mathbf{H}}_i^\top \bar{\mathbf{H}}_i)$:** Recall from Lemma 12 that, with probability at least $1-\delta$ over the modes, the rows of $\bar{\mathbf{H}}_i^0$ are a subset of the rows of $\bar{\mathbf{H}}_i$. As a result, (136) also implies that, with probability at least $1-3\delta$, for all $i \in [s]$ we have

$$s_{\min}(\bar{\mathbf{H}}_i) \geq s_{\min}(\bar{\mathbf{H}}_i^0) \geq \frac{\sqrt{N_i^0}}{4} \geq \sqrt{\frac{\pi_{\min}T}{16C_{sub}\log(T)}}, \tag{138}$$

$$\Longrightarrow \lambda_{\min}(\bar{\mathbf{H}}_i^\top \bar{\mathbf{H}}_i) \geq \frac{\pi_{\min}T}{16C_{sub}\log(T)}. \tag{139}$$

as long as $T \geq 4c^2C_{sub}(4+\beta_+^2)\log(2s(n+p)/\delta)\log(T)(n+p)/\pi_{\min}$, where we used the assumption made in the statement of the lemma that, $N_i^0 \geq \pi_{\min}T/(C_{sub}\log(T))$ with probability at least $1-\delta$.

• **Upper bounding $\|\bar{\mathbf{H}}_i^\top \mathbf{W}_i\|$:** To upper bound the singular values of $\bar{\mathbf{H}}_i$, from Lemma 11, it is straightforward to show that each row of $\bar{\mathbf{H}}_i$ is bounded with probability at least $1-\delta$ over the modes as $\|\bar{\mathbf{h}}_{(j_k)}\|_{\ell_2}^2 \leq c^2(1+(9/4)\beta_+^2)(n+p)$. This implies, with probability at least $1-\delta$ over the modes, for all $i \in [s]$, we have

$$s_{\max}(\bar{\mathbf{H}}_i) \leq \|\bar{\mathbf{H}}_i\|_F \leq c(1+2\beta_+)\sqrt{N_i(n+p)} \leq c(1+2\beta_+)\sqrt{T(n+p)/\log(T)}. \tag{140}$$

To proceed, let $\bar{\mathbf{H}}_i$ have singular value decomposition $\mathbf{U}\mathbf{\Sigma}\mathbf{V}^T$ with $\|\mathbf{\Sigma}\| \leq c(1+2\beta_+)\sqrt{T(n+p)/\log(T)}$. Since $\mathbf{W}_i$ has i.i.d. σ_w-subGaussian entries (A3), $\mathbf{U}^T\mathbf{W}_i \in \mathbb{R}^{n \times p}$ has i.i.d. σ_w-subGaussian columns. As a result, applying Theorem 5.39 of (Vershynin, 2010), with probbaility at least $1-\delta$, for all $i \in [s]$, we have

$$\|\mathbf{U}^T\mathbf{W}_i\| \leq C\sigma_w\sqrt{n+p} + C_0\sqrt{\log(2s/\delta)}. \tag{141}$$

This implies, with same probability we also have,

$$\|\bar{\mathbf{H}}_i^T\mathbf{W}_i\| \leq \|\mathbf{\Sigma}\|\|\mathbf{U}^T\mathbf{W}_i\| \leq c(1+2\beta_+)\sqrt{\frac{T(n+p)}{\log(T)}}\left(C\sigma_w\sqrt{n+p} + C_0\sqrt{\log\left(\frac{2s}{\delta}\right)}\right). \tag{142}$$

This completes the proof. $\square$

B.3.3. Estimation from Non-truncated States

Going back to the original problem of estimating the unknown dynamics from dependent samples, recall that the estimation error (101) in the case of single trajectory is upper bounded as follows.

$$\|\hat{\boldsymbol{\Theta}}_i - \boldsymbol{\Theta}_i^\star\| \leq \frac{\|\mathbf{H}_i^\top \mathbf{W}_i\|}{\lambda_{\min}(\mathbf{H}_i^\top \mathbf{H}_i)} \leq \frac{\|\bar{\mathbf{H}}_i^\top \mathbf{W}_i\| + \|\mathbf{H}_i^\top \mathbf{W}_i - \bar{\mathbf{H}}_i^\top \mathbf{W}_i\|}{\lambda_{\min}(\bar{\mathbf{H}}_i^\top \bar{\mathbf{H}}_i) - \|\mathbf{H}_i^\top \mathbf{H}_i - \bar{\mathbf{H}}_i^\top \bar{\mathbf{H}}_i\|}. \tag{143}$$

This implies that, in order to upper bound the estimation error $\|\hat{\boldsymbol{\Theta}}_i - \boldsymbol{\Theta}_i^\star\|$ with dependent samples, we need to upper bound the impact of truncation, captured by $\|\mathbf{H}_i^\top \mathbf{H}_i - \bar{\mathbf{H}}_i^\top \bar{\mathbf{H}}_i\|$ and $\|\mathbf{H}_i^\top \mathbf{W}_i - \bar{\mathbf{H}}_i^\top \mathbf{W}_i\|$.

Theorem 5 (Small impact of truncation)**.** *Consider the same setup of Lemma 5. Let t_0, β_+, β_+^0 and L_{tr1} be as in* (81)*,* (82)*,* (106) *and* (107) *respectively. Suppose the sampling period L obeys $L \geq \max\{t_0, L_{tr1}\}$. Let $\mathbf{H}_i$, $\bar{\mathbf{H}}_i$ and $\bar{\mathbf{W}}_i$ be the matrices given by* (99) *and* (129) *respectively. Then, with probability at least $1-\delta$ over the modes, for all $i \in [s]$, we have*

$$\frac{1}{N_i}\|\mathbf{H}_i^\top \mathbf{W}_i - \bar{\mathbf{H}}_i^\top \mathbf{W}_i\| \leq c c_w \sigma_w \beta_+^0 \tau_{\bar{L}} \rho_{\bar{L}}^{(L-1)/2} n\sqrt{ns}T/\delta, \tag{144}$$

$$\textit{and} \quad \frac{1}{N_i}\|\mathbf{H}_i^\top \mathbf{H}_i - \bar{\mathbf{H}}_i^\top \bar{\mathbf{H}}_i\| \leq 2c^2 \beta_+^0 (1+2\beta_+) \tau_{\bar{L}} \rho_{\bar{L}}^{(L-1)/2} n\sqrt{s(n+p)}T/\delta. \tag{145}$$

Proof. To begin, consider the difference

$$\begin{aligned}
\frac{1}{N_i}\|\mathbf{H}_i^\top \mathbf{H}_i - \bar{\mathbf{H}}_i^\top \bar{\mathbf{H}}_i\| &= \frac{1}{N_i}\|\sum_{k=1}^{N_i}(\mathbf{h}_{(j_k)}\mathbf{h}_{(j_k)}^\top - \bar{\mathbf{h}}_{(j_k)}\bar{\mathbf{h}}_{(j_k)}^\top)\|, \\
&\leq \max_{1\leq k\leq N_i} \|\mathbf{h}_{(j_k)}\mathbf{h}_{(j_k)}^\top - \bar{\mathbf{h}}_{(j_k)}\bar{\mathbf{h}}_{(j_k)}^\top\|, \\
&= \max_{1\leq k\leq N_i} \|\mathbf{h}_{(j_k)}\mathbf{h}_{(j_k)}^\top - \mathbf{h}_{(j_k)}\bar{\mathbf{h}}_{(j_k)}^\top + \mathbf{h}_{(j_k)}\bar{\mathbf{h}}_{(j_k)}^\top - \bar{\mathbf{h}}_{(j_k)}\bar{\mathbf{h}}_{(j_k)}^\top\|, \\
&\leq \max_{1\leq k\leq N_i} \left(\|\mathbf{h}_{(j_k)}(\mathbf{h}_{(j_k)} - \bar{\mathbf{h}}_{(j_k)})^\top\| + \|(\mathbf{h}_{(j_k)} - \bar{\mathbf{h}}_{(j_k)})\bar{\mathbf{h}}_{(j_k)}^\top\|\right), \\
&\leq \max_{1\leq k\leq N_i} \left(\|\mathbf{h}_{(j_k)}\|_{\ell_2}\|\mathbf{h}_{(j_k)} - \bar{\mathbf{h}}_{(j_k)}\|_{\ell_2} + \|\mathbf{h}_{(j_k)} - \bar{\mathbf{h}}_{(j_k)}\|_{\ell_2}\|\bar{\mathbf{h}}_{(j_k)}\|_{\ell_2}\right).
\end{aligned} \tag{146}$$

We will upper bound each of these terms separately and combine them together in (146) to get the desired upper bound. First of all, observe that each row of $\mathbf{H}_i$ is deterministically bounded as follows.

$$\|\mathbf{h}_{(j_k)}\|_{\ell_2}^2 \leq \frac{1}{\sigma_w^2}\|\mathbf{x}_{(j_k)}\|_{\ell_2}^2 + \frac{1}{\sigma_z^2}\|\mathbf{z}_{(j_k)}\|_{\ell_2}^2 \leq c^2\beta_+^2 n + c^2 p \leq c^2(1+\beta_+^2)(n+p). \tag{147}$$

Similarly, using Lemma 11 each row of $\bar{\mathbf{H}}_i$ is bounded with probability at least $1-\delta$ over the modes as follows.

$$\|\bar{\mathbf{h}}_{(j_k)}\|_{\ell_2}^2 \leq c^2(1+(9/4)\beta_+^2)(n+p). \tag{148}$$

To proceed, recall from (114) that, with probability at least $1-\delta$ over the modes, for all $k = 1,\dots,N_i$ and $i \in [s]$, we have

$$\begin{aligned}
\|\mathbf{h}_{(j_k)} - \bar{\mathbf{h}}_{(j_k)}\|_{\ell_2} = \left\| \begin{bmatrix} \frac{1}{\sigma_w}\mathbf{x}_{(j_k)} \\ \frac{1}{\sigma_z}\mathbf{z}_{(j_k)} \end{bmatrix} - \begin{bmatrix} \frac{1}{\sigma_w}\bar{\mathbf{x}}_{(j_k)} \\ \frac{1}{\sigma_z}\mathbf{z}_{(j_k)} \end{bmatrix} \right\|_{\ell_2} &= \frac{1}{\sigma_w}\|\mathbf{x}_{(j_k)} - \bar{\mathbf{x}}_{(j_k)}\|_{\ell_2}, \\
&\leq c\sigma_w \beta_+^0 n\sqrt{s}\tau_{\bar{L}}\rho_{\bar{L}}^{(L-1)/2}T/\delta.
\end{aligned} \tag{149}$$

Combining (147), (148) and (149) into (146), with probability at least $1-\delta$ over the modes, for all $i \in [s]$, we have

$$\frac{1}{N_i}\|\mathbf{H}_i^\top \mathbf{H}_i - \bar{\mathbf{H}}_i^\top \bar{\mathbf{H}}_i\| \leq 2c^2\beta_+^0(1+2\beta_+)\tau_{\bar{L}}\rho_{\bar{L}}^{(L-1)/2} n\sqrt{s(n+p)}T/\delta. \tag{150}$$

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

$$\frac{1}{N_i}\|\mathbf{H}_i^\top \mathbf{W}_i - \bar{\mathbf{H}}_i^\top \mathbf{W}_i\| = \frac{1}{N_i}\|\sum_{k=1}^{N_i}\mathbf{h}_{(j_k)}\mathbf{w}_{(j_k)}^\top - \sum_{k=1}^{N_i}\bar{\mathbf{h}}_{(j_k)}\mathbf{w}_{(j_k)}^\top\|,$$

$$
\begin{aligned}
&\leq \frac{1}{N_i} \sum_{k=1}^{N_i} \|(\mathbf{h}_{(j_k)} - \bar{\mathbf{h}}_{(j_k)})\mathbf{w}_{(j_k)}^\top\|, \\
&\leq \max_{1 \leq k \leq N_i} \|\mathbf{h}_{(j_k)} - \bar{\mathbf{h}}_{(j_k)}\|_{\ell_2} \|\mathbf{w}_{(j_k)}\|_{\ell_2}, \\
&\leq c c_w \sigma_w \beta_+^0 \tau_{\mathbf{L}} \rho_{\mathbf{L}}^{(L-1)/2} n\sqrt{ns}T/\delta. && (151)
\end{aligned}
$$

This completes the proof. □

Combining Lemma 4 with Theorem 5, we get our main result on the estimation of MJS in (79) from finite samples obtained from a single trajectory.

Theorem 6 (Learning with bounded noise)**.** *Consider the least squares problem* (100)*. Suppose Assumption A1 on the MJS and the Markov chain and A3 on the process noise hold. Suppose* $\{\mathbf{z}_t\}_{t=0}^{\infty} \overset{i.i.d.}{\sim} \mathcal{N}(0, \sigma_z^2 \mathbf{I}_p)$*. Suppose* $\mathbf{x}_0 \sim \mathcal{D}_x$ *such that* $\mathbb{E}[\mathbf{x}_0] = 0$*,* $\mathbb{E}[\|\mathbf{x}_0\|_{\ell_2}^2] \leq \beta_0^2 n$*. Let* t_0*,* β_+*,* β_+^0*,* L_{tr1} *and* L_{cov} *be as in* (81), (82), (106), (107) *and* (118) *respectively. Let* $C_{sub} > 0$ *be a system related constant. Let* $C, C_0 > 0$ *and* $c \geq 6$ *be fixed constants. Let*

$$
L_{tr_2}(\rho_{\mathbf{L}}, \delta) := 1 + \frac{2\log\left(64c^2 C_{sub}\tau_{\mathbf{L}}\beta_+^0(1+2\beta_+)n\sqrt{s(n+p)}T/(\pi_{\min}\delta)\right)}{1-\rho_{\mathbf{L}}}, \quad (152)
$$

$$
L_{tr_3}(\rho_{\mathbf{L}}, \delta) := 1 + \frac{2\log\left(c_w\beta_+^0\tau_{\mathbf{L}} nT\sqrt{Tns}/(C\delta(1+2\beta_+)(n+p)\sqrt{\log(T)})\right)}{1-\rho_{\mathbf{L}}}. \quad (153)
$$

Let $\Gamma > 0$ *be a constant such that* $\|\mathbf{K}\| \leq \Gamma$*. Suppose the sampling period* L *and the trajectory length* T *satisfy*

$$
L \geq \max\{t_0, L_{cov}(\rho_{\mathbf{L}}), L_{tr1}(\rho_{\mathbf{L}}, \delta), L_{tr2}(\rho_{\mathbf{L}}, \delta), L_{tr3}(\rho_{\mathbf{L}}, \delta)\} \quad (154)
$$

$$
T \geq 4c^2 C_{sub}(4+\beta_+^2)\log(2s(n+p)/\delta)\log(T)(n+p)/\pi_{\min}. \quad (155)
$$

Let $\mathcal{X} := \{\mathbf{x}_{(j_k)}\}_{k=1}^{N_i}$ *and* $\mathcal{X}' := \{\mathbf{x}_{(l_k)}\}_{k=1}^{N_i'}$ *be as in Lemma 12. Suppose* $N_i' \geq \pi_{\min}T/(C_{sub}\log(T))$ *with probability at least* $1-\delta$ *for all* $i \in [s]$*. Then, with probability at least* $1-5\delta$*, for all* $i \in [s]$*, we have*

$$
\|\hat{\mathbf{A}}_i - \mathbf{A}_i\| \leq \frac{64cC_{sub}(\sigma_z+\sigma_w)\Gamma(1+2\beta_+)\sqrt{\log(T)}}{\sigma_z\pi_{\min}}\left(\frac{C(n+p)}{\sqrt{T}} + \frac{C_0\sqrt{(n+p)\log(2s/\delta)}}{\sigma_w\sqrt{T}}\right),
$$

$$
\|\hat{\mathbf{B}}_i - \mathbf{B}_i\| \leq \frac{64cC_{sub}\sigma_w(1+2\beta_+)\sqrt{\log(T)}}{\sigma_z\pi_{\min}}\left(\frac{C(n+p)}{\sqrt{T}} + \frac{C_0\sqrt{(n+p)\log(2s/\delta)}}{\sigma_w\sqrt{T}}\right).
$$

Proof. To begin, combing the results of Theorems 4 and 5, when L and T obey (132) and (133) respectively, then with probability at least $1-5\delta$, for all $i \in [s]$, we have

$$
\begin{aligned}
\lambda_{\min}(\mathbf{H}_i^\top\mathbf{H}_i) &\geq \lambda_{\min}(\bar{\mathbf{H}}_i^\top\bar{\mathbf{H}}_i) - \|\mathbf{H}_i^\top\mathbf{H}_i - \bar{\mathbf{H}}_i^\top\bar{\mathbf{H}}_i\|, \\
&\geq \frac{\pi_{\min}T}{16C_{sub}\log(T)} - 2c^2\beta_+^0(1+2\beta_+)\tau_{\mathbf{L}}\rho_{\mathbf{L}}^{(L-1)/2} n\sqrt{s(n+p)}T^2/(\delta\log(T)), \\
&\geq \frac{\pi_{\min}T}{32C_{sub}\log(T)}, && (156)
\end{aligned}
$$

where we get the last inequality by choosing the sampling period L as follows,

$$
\begin{aligned}
&\frac{\pi_{\min}}{32C_{sub}\log(T)} \geq 2c^2\tau_{\mathbf{L}}\beta_+^0 T(1+2\beta_+)\rho_{\mathbf{L}}^{(L-1)/2} n\sqrt{s(n+p)}/(\delta\log(T)), \\
\iff\ &\rho_{\mathbf{L}}^{(L-1)/2} \leq \frac{\pi_{\min}\delta}{64c^2C_{sub}\tau_{\mathbf{L}}\beta_+^0(1+2\beta_+)n\sqrt{s(n+p)}T}, \\
&\qquad \Longleftarrow\ L \geq 1 + 2\frac{\log\left(64c^2C_{sub}\tau_{\mathbf{L}}\beta_+^0(1+2\beta_+)n\sqrt{s(n+p)}T/(\pi_{\min}\delta)\right)}{(1-\rho_{\mathbf{L}})}. && (157)
\end{aligned}
$$

Similarly, combing the results of Theorems 4 and 5, when L and T obey (132) and (133) respectively, then with probability at least $1-5\delta$, for all $i \in [s]$, we also have

$$
\begin{aligned}
\|\mathbf{H}_i^{!}\mathbf{W}_i\| &\leq \|\bar{\mathbf{H}}_i^{!}\mathbf{W}_i\| + \|\mathbf{H}_i^{!}\mathbf{W}_i - \bar{\mathbf{H}}_i^{!}\mathbf{W}_i\|, \\
&\leq c(1+2\beta_+)\sqrt{\frac{T(n+p)}{\log(T)}}\Big(C\sigma_w\sqrt{n+p} + C_0\sqrt{\log\big(\frac{2s}{\delta}\big)}\Big) \\
&\qquad + cc_w\sigma_w\beta_+^0\tau_{\mathbf{L}}\rho_{\mathbf{L}}^{(L-1)/2} n\sqrt{ns}T^2/(\delta\log(T)), \\
&\leq 2c(1+2\beta_+)\sqrt{\frac{T(n+p)}{\log(T)}}\Big(C\sigma_w\sqrt{n+p} + C_0\sqrt{\log\big(\frac{2s}{\delta}\big)}\Big)
\end{aligned} \tag{158}
$$

where we get the last inequality by choosing the sampling period L as follows,

$$
\begin{aligned}
&\frac{cc_w\sigma_w\beta_+^0\tau_{\mathbf{L}}\rho_{\mathbf{L}}^{(L-1)/2} n\sqrt{ns}T^2}{\delta\log(T)} \leq c(1+2\beta_+)\sqrt{\frac{T(n+p)}{\log(T)}}\Big(C\sigma_w\sqrt{n+p} + C_0\sqrt{\log\big(\frac{2s}{\delta}\big)}\Big), \\
&\iff \rho_{\mathbf{L}}^{(L-1)/2} \leq \frac{\delta(1+2\beta_+)\sqrt{\log(T)(n+p)}}{c_w\sigma_w\beta_+^0\tau_{\mathbf{L}} n\sqrt{ns}T\sqrt{T}}\Big(C\sigma_w\sqrt{n+p} + C_0\sqrt{\log\big(\frac{2s}{\delta}\big)}\Big), \\
&\Longleftarrow L \geq 1 + 2\frac{\log\big(c_w\beta_+^0\tau_{\mathbf{L}} nT\sqrt{Tns}/(C\delta(1+2\beta_+)(n+p)\sqrt{\log(T)})\big)}{(1-\rho_{\mathbf{L}})}.
\end{aligned} \tag{159}
$$

Finally combining (156) and (158), with probability at least $1-5\delta$, for all $i \in [s]$, we have

$$
\|\hat{\mathbf{\Theta}}_i - \mathbf{\Theta}_i^\star\| \leq \frac{64cC_{sub}(1+2\beta_+)\sqrt{\log(T)}}{\pi_{\min}}\Big(\frac{C\sigma_w(n+p)}{\sqrt{T}} + \frac{C_0\sqrt{(n+p)\log\big(2s/\delta\big)}}{\sqrt{T}}\Big). \tag{160}
$$

Lastly, using the standard result from Linear algebra that the norm of a sub-matrix is less than the norm of the original matrix, with probability at least $1-5\delta$, for all $i \in [s]$, we have

$$
\begin{aligned}
\|\hat{\mathbf{L}}_i - \mathbf{L}_i\| &\leq \frac{64cC_{sub}(1+2\beta_+)\sqrt{\log(T)}}{\pi_{\min}}\Big(\frac{C(n+p)}{\sqrt{T}} + \frac{C_0\sqrt{(n+p)\log\big(2s/\delta\big)}}{\sigma_w\sqrt{T}}\Big), \\
\|\hat{\mathbf{B}}_i - \mathbf{B}_i\| &\leq \frac{64cC_{sub}\sigma_w(1+2\beta_+)\sqrt{\log(T)}}{\sigma_z\pi_{\min}}\Big(\frac{C(n+p)}{\sqrt{T}} + \frac{C_0\sqrt{(n+p)\log\big(2s/\delta\big)}}{\sigma_w\sqrt{T}}\Big).
\end{aligned}
$$

This further implies that, with probability at least $1-5\delta$, for all $i \in [s]$, we have

$$
\begin{aligned}
\|\hat{\mathbf{A}}_i - \mathbf{A}_i\| &\leq \|\hat{\mathbf{L}}_i - \mathbf{L}_i\| + \|\mathbf{K}\|\|\hat{\mathbf{B}}_i - \mathbf{B}_i\|, \\
&\leq \frac{64cC_{sub}(\sigma_z+\sigma_w)\Gamma(1+2\beta_+)\sqrt{\log(T)}}{\sigma_z\pi_{\min}}\Big(\frac{C(n+p)}{\sqrt{T}} + \frac{C_0\sqrt{(n+p)\log\big(2s/\delta\big)}}{\sigma_w\sqrt{T}}\Big).
\end{aligned}
$$

This completes the proof. □

Next, we use the following lemma to relax the Assumption A3 on the noise.

Lemma 14 (From Bounded to Unbounded Noise)**.** *Let* $\mathbf{g} \overset{i.i.d.}{\sim} \mathcal{N}(0, \sigma_w^2\mathbf{I}_{nT})$ *and* $\mathbf{a}$ *be two independent vectors. Let* $\mathbf{g}^0$ *be the truncated Gaussian distribution* $\mathbf{g}^0 \sim \{\mathbf{g} \mid \|\mathbf{g}\|_{\ell_\infty} \leq c_w\sigma_w\}$. *Let* $S_{\mathbf{g},\mathbf{a}}$ *be the indicator function of an event defined on vectors* $\mathbf{g}, \mathbf{a}$ *s.t.*

$$
\mathsf{E}[S_{\mathbf{g}^0,\mathbf{a}}] \geq 1 - \delta/2.
$$

That is, the event holds with probability at least $1-\delta/2$. *Then, if the bound above holds for* $c_w > C := \sqrt{2\log(nT)} + \sqrt{2\log(2/\delta)}$, *we also have that*

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

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

Proof. Let E be the event $\{\mathbf{g} \mid \|\mathbf{g}\|_{\ell_1} \leq c_w\sigma_w\}$. If $c_w \geq \sqrt{2\log(nT)} + \sqrt{2\log(2/\delta)}$, using Gaussian tail bound and the fact that $\mathsf{E}[\|\mathbf{g}\|_{\ell_1}] \leq \sigma_w\sqrt{2\log(nT)}$, observe that $\mathsf{P}(E) \geq 1 - e^{-\frac{(c_w\sigma_w - \mathbb{E}[\|\mathbf{g}\|_{\ell_1}])^2}{2}} \geq 1 - \delta/2$. Therefore, we have

$$\mathsf{E}[S_{\mathbf{g};\mathbf{a}}] \geq \mathsf{E}[S_{\mathbf{g};\mathbf{a}}|E]\,\mathsf{P}(E) = \mathsf{E}[S_{\mathbf{g}^0;\mathbf{a}}]\,\mathsf{P}(E) \geq (1-\delta/2)^2 \geq 1-\delta.$$

This completes the proof. □

Combining Theorem 6 and Lemma 14, we get the following corollary on learning the MJS dynamics.

Corollary 6 (Learning with un-bounded noise)**.** *Consider the least squares problem* (100)*. Suppose Assumption A1 on the MJS and the Markov chain hold. Suppose* $\{\mathbf{z}_t\}_{t=0}^{1} \overset{i.i.d.}{\sim} \mathcal{N}(0,\sigma_z^2\mathbf{I}_p)$ *and* $\{\mathbf{w}_t\}_{t=0}^{1} \overset{i.i.d.}{\sim} \mathcal{N}(0,\sigma_w^2\mathbf{I}_n)$*. Suppose* $\mathbf{x}_0 \sim \mathcal{D}_x$ *such that* $\mathsf{E}[\mathbf{x}_0] = 0$*,* $\mathsf{E}[\|\mathbf{x}_0\|_{\ell_2}^2] \leq \beta_0^2 n$*. Let* $c_w := \sqrt{2\log(nT)} + \sqrt{2\log(12/\delta)}$ *and* t_0*,* β_+*,* β_+^0*,* L_{tr1}*,* L_{cov}*,* L_{tr2} *and* L_{tr3} *be as in* (81), (82), (106), (107), (118), (152) *and* (153) *respectively. Let* $C_{sub} > 0$ *be a system related constant. Let* $C, C_0 > 0$ *and* $c \geq 6$ *be fixed constants. Suppose the sampling period* L *and the trajectory length* T *satisfy*

$$L \geq \max\{t_0, L_{cov}(\rho_{\mathbf{L}}), L_{tr1}(\rho_{\mathbf{L}},\delta/6), L_{tr2}(\rho_{\mathbf{L}},\delta/6), L_{tr3}(\rho_{\mathbf{L}},\delta/6)\} \tag{161}$$

$$T \geq 4c^2C_{sub}(4+\beta_+^2)\log(12s(n+p)/\delta)\log(T)(n+p)/\pi_{\min}. \tag{162}$$

Let $\mathcal{X} := \{\mathbf{x}_{(j_k)}\}_{k=1}^{N_i}$ *and* $\mathcal{X}^0 := \{\mathbf{x}_{(l_k)}\}_{k=1}^{N_i^0}$ *be as in Lemma 12. Suppose* $N_i^0 \geq \pi_{\min}T/(C_{sub}\log(T))$ *with probability at least* $1-\delta/6$ *for all* $i \in [s]$*. Then, with probability at least* $1-\delta$*, for all* $i \in [s]$*, we have*

$$\|\hat{\mathbf{A}}_i - \mathbf{A}_i\| \leq \frac{C_{sub}(\sigma_z+\sigma_w)\sqrt{s}\big(\log(nT) + \sqrt{\log(T)\log(12/\delta)}\big)}{\sigma_z\pi_{\min}}\left(\frac{C(n+p)}{\sqrt{T}} + \frac{C_0\sqrt{(n+p)\log\left(12s/\delta\right)}}{\sigma_w\sqrt{T}}\right),$$

$$\|\hat{\mathbf{B}}_i - \mathbf{B}_i\| \leq \frac{C_{sub}\sigma_w\sqrt{s}\big(\log(nT) + \sqrt{\log(T)\log(12/\delta)}\big)}{\sigma_z\pi_{\min}}\left(\frac{C(n+p)}{\sqrt{T}} + \frac{C_0\sqrt{(n+p)\log\left(12s/\delta\right)}}{\sigma_w\sqrt{T}}\right).$$

B.3.4. Finalizing the SYSID: Proof of Theorem 1

Lastly, we combine Lemma 9 and Corollary 6 to get our final result on learning the MJS dynamics. The following theorem is a more refined and precise version of our main SYSID result Theorem 1.

Theorem 7 (Main result)**.** *Suppose we run Algorithm 1 with* $c_{\mathbf{x}} \geq 6\underline{c}_{\mathbf{x}}(\rho_{\mathbf{L}},\tau_{\mathbf{L}})$*,* $c_{\mathbf{z}} \geq 6\underline{c}_{\mathbf{z}}$*, where* $\underline{c}_{\mathbf{x}}(\rho,\tau)$ *and* $\underline{c}_{\mathbf{z}}$ *are defined in Table 2. Suppose* $\mathbf{x}_0 \sim \mathcal{D}_x$ *such that* $\mathsf{E}[\mathbf{x}_0] = 0$*,* $\mathsf{E}[\|\mathbf{x}_0\|_{\ell_2}^2] \leq \beta_0^2 n$ *and* $\|\mathbf{x}_0\| \leq \bar{x}_0$*. Suppose Assumption A1 on the MJS and the Markov chain hold. Suppose* $\{\mathbf{z}_t\}_{t=0}^{1} \overset{i.i.d.}{\sim} \mathcal{N}(0,\sigma_z^2\mathbf{I}_p)$ *and* $\{\mathbf{w}_t\}_{t=0}^{1} \overset{i.i.d.}{\sim} \mathcal{N}(0,\sigma_w^2\mathbf{I}_n)$*. Let* $c_w := \sqrt{2\log(nT)} + \sqrt{2\log(12/\delta)}$ *and* t_0*,* β_+*,* β_+^0*,* L_{tr1}*,* L_{cov}*,* L_{tr2} *and* L_{tr3} *be as in* (81), (82), (106), (107), (118), (152) *and* (153) *respectively. Let* $C, C_0 > 0$ *and* $c \geq 6$ *be fixed constants. Let*

$$C_{sys}(\rho_{\mathbf{L}},\delta) := \max\{t_0, L_{cov}(\rho_{\mathbf{L}}), L_{tr1}(\rho_{\mathbf{L}},\delta), L_{tr2}(\rho_{\mathbf{L}},\delta), L_{tr3}(\rho_{\mathbf{L}},\delta)\}/\log(T). \tag{163}$$

Choose $C_{sub} \geq \max\{C_{sys}(\rho_{\mathbf{L}},\delta/6), \underline{C}_{sub;N}(\bar{x}_0,\delta/6,T,\rho_{\mathbf{L}},\tau_{\mathbf{L}})\}$*, where* $C_{sub;N}(\bar{x}_0,\delta,T,\rho,\tau)$ *is defined in Table 2. Let*

$$T_{sys}(\delta) := 4c^2C_{sub}(4+\beta_+^2)\log(2s(n+p)/\delta)\log(T)(n+p)/\pi_{\min}. \tag{164}$$

Choose the trajectory length $T \geq \max\{T_{sys}(\delta/6), \underline{T}_N(\delta/6,\rho_{\mathbf{L}},\tau_{\mathbf{L}})\}$*, where* $\underline{T}_N(\delta,\rho,\tau)$ *is defined in Table 1. Then, with probability at least* $1-\delta$*, for all* $i \in [s]$*, we have*

$$\|\hat{\mathbf{A}}_i - \mathbf{A}_i\| \leq \frac{C_{sub}(\sigma_z+\sigma_w)\sqrt{s}\big(\log(nT) + \sqrt{\log(T)\log(12/\delta)}\big)}{\sigma_z\pi_{\min}}\left(\frac{C(n+p)}{\sqrt{T}} + \frac{C_0\sqrt{(n+p)\log\left(12s/\delta\right)}}{\sigma_w\sqrt{T}}\right),$$

$$\|\hat{\mathbf{B}}_i - \mathbf{B}_i\| \leq \frac{C_{sub}\sigma_w\sqrt{s}\big(\log(nT) + \sqrt{\log(T)\log(12/\delta)}\big)}{\sigma_z\pi_{\min}}\left(\frac{C(n+p)}{\sqrt{T}} + \frac{C_0\sqrt{(n+p)\log\left(12s/\delta\right)}}{\sigma_w\sqrt{T}}\right).$$

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 δ and $\log(T)$ terms). Besides these, the dominant term (when estimating $\mathbf{A}$ reduces to)

$$(1+\frac{\sigma_w}{\sigma_z})\frac{\sqrt{s}}{\pi_{\min}}\frac{n+p}{\sqrt{T}}.$$

which is identical to our statement in Theorem 7 except the additional factor of $\sqrt{s}$. In Theorem 7, we opted to put s under $\tilde{\mathcal{O}}$ notation to emphasize $\pi_{\min}$ dependence. Note that, $\pi_{\min}^{-1} \geq s$ by definition. The overall sample complexity grows as T & $s(n+p)^2/\pi_{\min}^2$. We also remark that, this quadratic growth is somewhat undesirable. A degrees-of-freedom counting argument would lead to an ideal dependency of T & $n+p/\pi_{\min}$. The reason is that, each vector state equation we fit has n scalar equations. The total degrees of freedom for each dynamics pair $(\mathbf{A}_s, \mathbf{B}_s)$ 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 $T \geq (n+p)/\pi_{\min}$. Our analysis indicates that this suboptimality (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 (90)) as well as a better control on the degree of independence across sampled states (this issue arises during the proof of Theorem 4).

• **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 fact, more recently, self-normalized martingale arguments are employed to address temporal dependencies (Sarkar & Rakhlin, 2019; Simchowitz et al., 2018). Our argument uses mixing-time however it necessitates a fairly-sophisticated double-subsampling argument and we would like to discuss to what extent this can be avoided. The second subsampling grabs the bounded states and uses them to regress the system dynamics. Unfortunately, this stage 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 (Vershynin, 2010). However, the first subsampling can potentially be avoided in two ways. The first option is adapting self-normalized martingale arguments to our settings. This would essentially help argue that, even if subsequent states are not independent, a new state carries sufficient excitation compared to the earlier one. Here, a potential challenge is keeping track of martingale filtrations and ensuring excitability as we still sample the bounded states within the trajectory. The second option is a stitching argument (e.g. (Oymak, 2019)). In this argument, say we know that, regression using subsampling $(x_L, x_{2L}, \dots, x_{NL})$ with $N = T/L$ samples works with high probability (i.e. returns a decent estimate of $\mathbf{A}, \mathbf{B}$ after the secondary bounded-state subsampling). Note that, the identical argument would also show that any shifted subsampling $(x_{\tau+L}, x_{\tau+2L}, \dots, x_{\tau+NL})$ would also work for any $0 \leq \tau \leq L-1$. Thus, we have L regression problems all of which return a good estimate of $\mathbf{A}, \mathbf{B}$. These regression problems can then be sticthed to show that, solving the combined regression over not-subsampled data still works (using a union bound over all problems) (see (Oymak, 2019) as an example). We left these variations as future refinements. Note that, these don't substantially affect the sample complexity or regret bound as the subsampling period is logarithmic. However, it is possible to mitigate the spectral radius dependency by shaving a factor of $1/(1-\rho_{\tilde{\mathbf{L}}})$ (e.g. martingale based arguments have milder $\rho_{\tilde{\mathbf{L}}}$ dependence (Simchowitz et al., 2018; Sarkar & Rakhlin, 2019)).

C. MJS Regret Analysis

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

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

$$J_T(\mathbf{x}_0, \omega(0), \{\mathbf{K}_{1:s}, \mathbf{\Sigma_z}\}) := \sum_{t=1}^{T} \mathbb{E}[\mathbf{x}_t^\top \mathbf{Q}_{\omega(t)}\mathbf{x}_t + \mathbf{u}_t^\top \mathbf{R}_{\omega(t)}\mathbf{u}_t \mid \mathbf{x}_0, \omega(0), \mathbf{K}_{1:s}] \tag{165}$$

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 Regret$_i$ in (6) with $\mathbf{x}_0, \omega(0), \mathbf{K}_{1:s}$ setting to $\mathbf{x}_0^{(i)}, \omega^{(i)}(0), \mathbf{K}_{1:s}^{(i)}$ since Regret$_i$ depends on randomness in $\mathcal{F}_{i-1}$ only through $\mathbf{x}_0^{(i)}, \omega^{(i)}(0), \mathbf{K}_{1:s}^{(i)}$. In the remainder of this appendix, for simplicity, we will drop the conditions $\mathbf{x}_0, \omega(0), \mathbf{K}_{1:s}$ in the expectation and simply write $\mathsf{E}[\,\cdot \mid \mathbf{x}_0, \omega(0), \mathbf{K}_{1:s}]$ as $\mathsf{E}[\cdot]$. So, for any measurable function f, $\mathsf{E}[f(\mathbf{x}_0, \omega(0), \mathbf{K}_{1:s})] = f(\mathbf{x}_0, \omega(0), \mathbf{K}_{1:s})$. Note that even though the results in this appendix are derived for conditional expectation $\mathsf{E}[\,\cdot \mid \mathbf{x}_0, \omega(0), \mathbf{K}_{1:s}]$, 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 $\mathbf{z}_t$ and starting from $\mathbf{x}_0 = 0$.

$$J(0, \omega(0), \{\mathbf{K}_{1:s}\}) := \limsup_{T \to \infty} \frac{1}{T} J_T(0, \omega(0), \{\mathbf{K}_{1:s}, 0\}) \tag{166}$$

Let $\mathbf{P}_{1:s}^\star$ denote the solution to the coupled discrete algebraic Riccati equations (Costa et al., 2006). Let $\mathbf{K}_{1:s}^\star$ denote the resulting infinite-horizon optimal controller computed using $\mathbf{P}_{1:s}^\star$. Note that the infinite-horizon optimal average cost $J^\star$ in (4) is achieved if the optimal controller $\mathbf{K}_{1:s}^\star$ is used, i.e.

$$J^\star = J(0, \omega(0), \{\mathbf{K}_{1:s}^\star\}). \tag{167}$$

Note that if the underlying Markov chain $\mathbf{T}$ is ergodic, for any initial state $\mathbf{x}_0$ and mode $\omega(0)$, $J^\star = J(\mathbf{x}_0, \omega(0), \{\mathbf{K}_{1:s}^\star\})$. Let $\mathbf{L}_i^\star = \mathbf{A}_i + \mathbf{B}_i \mathbf{K}_i^\star$ for all $i \in [s]$ denote the closed-loop state matrix when the optimal controller $\mathbf{K}_{1:s}^\star$ is used. Define the augmented state matrix $\tilde{\mathbf{L}}^\star$ such that its ij-th block is given by $[\tilde{\mathbf{L}}^\star]_{ij} := [\mathbf{T}]_{ji} \mathbf{L}_j^\star \otimes \mathbf{L}_j^\star$. From (Costa et al., 2006), we know $\mathbf{K}_{1:s}^\star$ is stabilizes the MJS, thus $\rho^\star := \rho(\tilde{\mathbf{L}}^\star) < 1$.

Since Regret$_i$ defined in (6) can be written as

$$\text{Regret}_i = J_T(\mathbf{x}_0^{(i)}, \omega^{(i)}(0), \{\mathbf{K}_{1:s}^{(i)}, \sigma_{\mathbf{z},i}^2 \mathbf{I}_p\}) - T J^\star, \tag{168}$$

to evaluate Regret(T), it suffices to evaluate $J_T(\mathbf{x}_0, \omega(0), \{\mathbf{K}_{1:s}, \mathbf{\Sigma}_\mathbf{z}\}) - TJ^\star$ for generic $\mathbf{x}_0$, $\omega(0)$, $\mathbf{K}_{1:s}$, and $\mathbf{\Sigma}_\mathbf{z}$. The outline of this Appendix C is as follows.

- In Appendix C.1, we restate perturbation results (Du et al., 2021) on $J(0, \omega(0), \{\mathbf{K}_{1:s}\}) - J^\star$.
- In Appendix C.2, we evaluate $J_T(\mathbf{x}_0, \omega(0), \{\mathbf{K}_{1:s}, \mathbf{\Sigma}_\mathbf{z}\}) - TJ(0, \omega(0), \{\mathbf{K}_{1:s}\})$. Then, applying the results in Appendix C.1, we can bound the single epoch regret $J_T(\mathbf{x}_0, \omega(0), \{\mathbf{K}_{1:s}, \mathbf{\Sigma}_\mathbf{z}\}) - TJ^\star$.
- In Appendix C.3, we stitch regrets for all epochs together, and combine them with identification results in Appendix B to bound 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 $\mathbf{K}_{1:s}$ that is close to the optimal $\mathbf{K}_{1:s}^\star$.

Lemma 15 (Spectral Radius Perturbation of Augmented State Matrix)**.** *For an arbitrary controller $\mathbf{K}_{1:s}$, let $\mathbf{L}_i = \mathbf{A}_i + \mathbf{B}_i \mathbf{K}_i$ for all $i \in [s]$, and let $\tilde{\mathbf{L}}$ be the augmented state matrix such that its ij-th block is given by $[\tilde{\mathbf{L}}]_{ij} := [\mathbf{T}]_{ji} \mathbf{L}_j \otimes \mathbf{L}_j$. Assume $\|\mathbf{K}_{1:s} - \mathbf{K}_{1:s}^\star\| \leq \bar{\epsilon}_\mathbf{K}$, where $\bar{\epsilon}_\mathbf{K}$ is defined in Table 3. Then, we have*

$$\|\tilde{\mathbf{L}}^k\| \leq \tau(\tilde{\mathbf{L}}^\star)(\frac{1+\rho^\star}{2})^k, \forall k \in \mathbb{N} \tag{169}$$

$$\rho(\tilde{\mathbf{L}}) \leq \frac{1+\rho^\star}{2} \tag{170}$$

And controller $\mathbf{K}_{1:s}$ is stabilizing.

Proof. Let $\mathbf{\Delta}_j = \mathbf{K}_j - \mathbf{K}_j^\star$, then from the assumption we know $\|\mathbf{\Delta}_j\| \leq \bar{\epsilon}_\mathbf{K}$. Consider the ij-th block difference between $\tilde{\mathbf{L}}^\star$ and $\tilde{\mathbf{L}}$, we have

$$[\tilde{\mathbf{L}}]_{ij} - [\tilde{\mathbf{L}}]_{ij}^\star = \mathbf{T}_{ji} \left((\mathbf{B}_j \mathbf{\Delta}_j) \otimes (\mathbf{B}_j \mathbf{\Delta}_j) + (\mathbf{B}_j \mathbf{\Delta}_j) \otimes \mathbf{L}^\star + \mathbf{L}^\star \otimes (\mathbf{B}_j \mathbf{\Delta}_j) \right) \tag{171}$$

which gives

$$\begin{aligned}\|[\tilde{\mathbf{L}}]_{ij} - [\tilde{\mathbf{L}}]^{\star}_{ij}\| \leq& \mathbf{T}_{ji}\left(\|\mathbf{B}_{1:s}\|^2\bar{\epsilon}_{\mathbf{K}}^2 + 2\|\mathbf{B}_{1:s}\|\|\mathbf{L}_{1:s}\|\bar{\epsilon}_{\mathbf{K}}\right)\\ \leq& \mathbf{T}_{ji}(1 + 2\|\mathbf{L}^{\star}_{1:s}\|)\|\mathbf{B}_{1:s}\|\bar{\epsilon}_{\mathbf{K}}\\ \leq& \mathbf{T}_{ji}\frac{1-\rho^{\star}}{2\sqrt{s}\tau(\tilde{\mathbf{L}}^{\star})}\end{aligned} \tag{172}$$

where the last two inequalities follow from the two upper bounds $\bar{\epsilon}_{\mathbf{K}} \leq \|\mathbf{B}_{1:s}\|^{-1}$ and $\bar{\epsilon}_{\mathbf{K}} \leq \frac{1-\rho^{\star}}{2\sqrt{s}\tau(\tilde{\mathbf{L}}^{\star})(1+2\|\mathbf{L}^{\star}_{1:s}\|)\|\mathbf{B}_{1:s}\|}$ in the definition of $\bar{\epsilon}_{\mathbf{K}}$. Then we have

$$\|\tilde{\mathbf{L}} - \tilde{\mathbf{L}}^{\star}\| \leq \sqrt{\sum_{i=1}^{s}\sum_{j=1}^{s}\|[\tilde{\mathbf{L}}]_{ij} - [\tilde{\mathbf{L}}]^{\star}_{ij}\|^2} \leq \frac{1-\rho^{\star}}{2\tau(\tilde{\mathbf{L}}^{\star})}. \tag{173}$$

Applying (Mania et al., 2019, Lemma 5) to this bound, we have

$$\|\tilde{\mathbf{L}}^k\| \leq \tau(\tilde{\mathbf{L}}^{\star})(\tau(\tilde{\mathbf{L}}^{\star})\|\tilde{\mathbf{L}} - \tilde{\mathbf{L}}^{\star}\| + \rho^{\star}) \leq \tau(\tilde{\mathbf{L}}^{\star})(\frac{1+\rho^{\star}}{2})^k \tag{174}$$

which shows (169). By Gelfand's formula, we have

$$\rho(\tilde{\mathbf{L}}) \leq \lim_{k\to\infty}(\tau(\tilde{\mathbf{L}}^{\star})(\frac{1+\rho^{\star}}{2})^k)^{\frac{1}{k}} = \frac{1+\rho^{\star}}{2} \tag{175}$$

hence (170) holds. Since $\rho^{\star} < 1$, we know $\rho(\tilde{\mathbf{L}}) < 1$, thus $\mathbf{K}_{1:s}$ 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 16 (Perturbation of Infinite-horizon MJS-LQR (Du et al., 2021))**.** *Infinite-horizon MJS-LQR($\mathbf{A}_{1:s}, \mathbf{B}_{1:s}, \mathbf{T}, \mathbf{Q}_{1:s}, \mathbf{R}_{1:s}$) problems have the following perturbation results. Note that notations $\bar{\epsilon}_{\mathbf{K}}, \bar{\epsilon}_{\mathbf{A},\mathbf{B},\mathbf{T}}$, and $C^{\mathbf{K}}_{\mathbf{A},\mathbf{B},\mathbf{T}}$ are defined in Table 3.*

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

$$J(0, \omega(0), \{\mathbf{K}_{1:s}\}) - J^{\star} \leq C^{J}_{\mathbf{K}}\|\mathbf{\Sigma}_{\mathbf{w}}\|\|\mathbf{K}_{1:s} - \mathbf{K}^{\star}_{1:s}\|^2. \tag{176}$$

2. *Suppose there is an arbitrary MJS($\hat{\mathbf{A}}_{1:s}, \hat{\mathbf{B}}_{1:s}, \hat{\mathbf{T}}$) with $\epsilon_{\mathbf{A},\mathbf{B}} := \max\{\|\hat{\mathbf{A}}_{1:s} - \mathbf{A}_{1:s}\|, \|\hat{\mathbf{B}}_{1:s} - \mathbf{B}_{1:s}\|\} \leq \bar{\epsilon}_{\mathbf{A},\mathbf{B},\mathbf{T}}$, and $\epsilon_{\mathbf{T}} := \|\hat{\mathbf{T}} - \mathbf{T}\|_{\infty} \leq \bar{\epsilon}_{\mathbf{A},\mathbf{B},\mathbf{T}}$. Then, the optimal controller $\mathbf{K}_{1:s}$ to the infinite-horizon MJS-LQR($\hat{\mathbf{A}}_{1:s}, \hat{\mathbf{B}}_{1:s}, \hat{\mathbf{T}}, \mathbf{Q}_{1:s}, \mathbf{R}_{1:s}$) exists and can be computed using coupled discrete algebraic Riccati equations, and*

$$\|\mathbf{K}_{1:s} - \mathbf{K}^{\star}_{1:s}\| \leq C^{\mathbf{K}}_{\mathbf{A},\mathbf{B},\mathbf{T}}(\epsilon_{\mathbf{A},\mathbf{B}} + \epsilon_{\mathbf{T}}). \tag{177}$$

   By definitions of $\bar{\epsilon}_{\mathbf{A},\mathbf{B},\mathbf{T}}$, we see $\|\mathbf{K}_{1:s} - \mathbf{K}^{\star}_{1:s}\| \leq \bar{\epsilon}_{\mathbf{K}}$, thus (169) *and* (170) *also hold.*

C.2. Single Epoch Regret Analysis

Recall the definitions of $\tilde{\mathbf{B}}_t$ and $\tilde{\mathbf{\Pi}}_t$ in (11) of Appendix A.1. Furthermore, we define

$$\tilde{\mathbf{\Pi}}_{\infty} = \pi_{\infty} \otimes \mathbf{I}_{n^2}, \quad \tilde{\mathbf{R}}_t = \sum_{i=1}^{s}\pi_t(i)\mathbf{R}_i. \tag{178}$$

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

$$\mathcal{H}(\begin{bmatrix}\mathbf{V}_1\\ \vdots\\ \mathbf{V}_s\end{bmatrix}) = \begin{bmatrix}\mathbf{vec}(\mathbf{V}_1)\\ \vdots\\ \mathbf{vec}(\mathbf{V}_s)\end{bmatrix}, \tag{179}$$

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

Let

$$\mathbf{M}_i := \mathbf{Q}_i + \mathbf{K}_i^\top \mathbf{R}_i \mathbf{K}_i \qquad \mathbf{M} := [\mathbf{M}_1, \dots, \mathbf{M}_s]. \tag{180}$$

We define

$$\begin{aligned} N_{0,t} &= \mathbf{tr}(\mathbf{M}\mathcal{H}^{-1}(\tilde{\mathbf{L}}^t \begin{bmatrix} \mathbf{vec}(\mathbf{\Sigma}_1(0)) \\ \vdots \\ \mathbf{vec}(\mathbf{\Sigma}_s(0)) \end{bmatrix})) \\ N_{\mathbf{z},1,t} &= \mathbf{tr}(\mathbf{M}\mathcal{H}^{-1}((\tilde{\mathbf{B}}_t + \tilde{\mathbf{L}}\tilde{\mathbf{B}}_{t-1} + \cdots + \tilde{\mathbf{L}}^{t-1}\tilde{\mathbf{B}}_1)\mathbf{vec}(\mathbf{\Sigma}_\mathbf{z}))) \\ N_{\mathbf{w},t} &= \mathbf{tr}(\mathbf{M}\mathcal{H}^{-1}((\tilde{\mathbf{\Pi}}_t + \tilde{\mathbf{L}}\tilde{\mathbf{\Pi}}_{t-1} + \cdots + \tilde{\mathbf{L}}^{t-1}\tilde{\mathbf{\Pi}}_1)\mathbf{vec}(\mathbf{\Sigma}_\mathbf{w}))) \\ N_{\mathbf{z},2,t} &= \mathbf{tr}(\tilde{\mathbf{R}}_t \mathbf{\Sigma}_z), \end{aligned} \tag{181}$$

and

$$S_{0,T} = \sum_{t=1}^T N_{0,t}, \quad S_{\mathbf{z},1,T} = \sum_{t=1}^T N_{\mathbf{z},1,t}, \quad S_{\mathbf{w},T} = \sum_{t=1}^T N_{\mathbf{w},t}, \quad S_{\mathbf{z},2,T} = \sum_{t=1}^T N_{\mathbf{z},2,t}. \tag{182}$$

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

Lemma 17 (Cumulative Cost Expression)**.** *For the cost $J_T(\mathbf{x}_0, \omega(0), \{\mathbf{K}_{1:s}, \mathbf{\Sigma}_\mathbf{z}\})$ defined in* (165)*, we have*

$$J_T(\mathbf{x}_0, \omega(0), \{\mathbf{K}_{1:s}, \mathbf{\Sigma}_\mathbf{z}\}) = S_{0,T} + S_{\mathbf{z},1,T} + S_{\mathbf{z},2,T} + S_{\mathbf{w},T}. \tag{183}$$

Proof. For the expected cost at time t, we have

$$\begin{aligned} \mathsf{E}[\mathbf{x}_t^\top \mathbf{Q}_{\omega(t)}\mathbf{x}_t + \mathbf{u}_t^\top \mathbf{R}_{\omega(t)}\mathbf{u}_t] &= \sum_{i=1}^s \mathbf{tr}\left(\mathsf{E}[\mathbf{Q}_{\omega(t)}\mathbf{x}_t\mathbf{x}_t^\top \mathbf{1}_{\{\omega(t)=i\}}] + \mathsf{E}[\mathbf{R}_{\omega(t)}\mathbf{u}_t\mathbf{u}_t^\top \mathbf{1}_{\{\omega(t)=i\}}]\right) \\ &= \sum_{i=1}^s \mathbf{tr}\left((\mathbf{Q}_i + \mathbf{K}_i^\top \mathbf{R}_i\mathbf{K}_i)\mathbf{\Sigma}_i(t) + \pi_t(i)\mathbf{R}_i\mathbf{\Sigma}_\mathbf{z}\right) \\ &= \sum_{i=1}^s \mathbf{tr}\left(\mathbf{M}_i\mathbf{\Sigma}_i(t)\right) + N_{\mathbf{z},2,t}, \end{aligned} \tag{184}$$

where the second equality follows since $\mathbf{u}_t = \mathbf{K}_{\omega(t)}\mathbf{x}_t + \mathbf{z}_t$. Now plugging in the dynamics of $\mathbf{\Sigma}_i(t)$ in Lemma 1, we can conclude the proof. □

Next, before proceeding, we provide several properties regarding the operator $\mathbf{tr}(\mathbf{M}\mathcal{H}(\cdot))$ that shows up in (181) and (182), which will be used later to evaluate $J_T(\mathbf{x}_0, \omega(0), \{\mathbf{K}_{1:s}, \mathbf{\Sigma}_\mathbf{z}\}) - TJ(0, \omega(0), \{\mathbf{K}_{1:s}\})$.

Lemma 18 (Properties of Cost Building Bricks)**.** *For any $t, t' \in \mathbb{N}$, we have*

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

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

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

(L4) $|\mathbf{tr}(\mathbf{M}\mathcal{H}^{-1}(\tilde{\mathbf{L}}^t(\tilde{\mathbf{\Pi}}_{t'} - \tilde{\mathbf{\Pi}}_\infty)\mathbf{vec}(\mathbf{\Sigma}_\mathbf{w})))| \leq \tau_{MC} n\sqrt{s}\|\mathbf{M}_{1:s}\|\vert\vert\vert\tilde{\mathbf{L}}^t\vert\vert\vert\|\mathbf{\Sigma}_\mathbf{w}\|\rho_{MC}^{t'}$, *where τ_{MC} are ρ_{MC} are given in Definition 3, and $\tilde{\mathbf{\Pi}}_\infty$ is given in* (178)

Proof. Let $[\cdot]_i$ denote the ith sub-block of an $s \times 1$ block matrix. Let $\mathbf{vec}^{-1}$ denote the inverse mapping of $\mathbf{vec}$, i.e., $\mathbf{vec}^{-1}([\mathbf{v}_1^\top, \cdots, \mathbf{v}_r^\top]^\top) = [\mathbf{v}_1, \cdots, \mathbf{v}_r]$ 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 $(\mathbf{B}^\top \otimes \mathbf{A})\mathbf{vec}(\mathbf{X}) = \mathbf{vec}(\mathbf{C})$. This together with the definitions of $\tilde{\mathbf{B}}_t$, $\tilde{\mathbf{\Pi}}_t$ in (11), $\tilde{\mathbf{\Pi}}_\infty$, $\tilde{\mathbf{R}}_t$ in (178), and $\mathcal{H}()$ in (179) yields the following preliminary results

$$[\mathcal{H}^{-1}(\tilde{\mathbf{L}}^t \mathbf{v})]_i \succeq 0, \tag{185a}$$

$$\mathbf{vec}^{-1}([\tilde{\mathbf{B}}_{t'} \mathbf{vec}(\mathbf{\Sigma_z})]_i) \succeq 0, \tag{185b}$$

$$\mathbf{vec}^{-1}([\tilde{\mathbf{\Pi}}_{t'} \mathbf{vec}(\mathbf{\Sigma_w})]_i) \succeq 0, \tag{185c}$$

$$\mathbf{vec}^{-1}([|\tilde{\mathbf{\Pi}}_{t'} - \tilde{\mathbf{\Pi}}_\infty| \mathbf{vec}(\mathbf{\Sigma_w})]_i) \succeq 0, \tag{185d}$$

$$|\mathbf{tr}(\mathbf{M}\mathcal{H}^{-1}(\tilde{\mathbf{L}}^t(\tilde{\mathbf{\Pi}}_{t'} - \tilde{\mathbf{\Pi}}_\infty)\mathbf{vec}(\mathbf{\Sigma_w})))| \le \mathbf{tr}(\mathbf{M}\mathcal{H}^{-1}(\tilde{\mathbf{L}}^t|\tilde{\mathbf{\Pi}}_{t'} - \tilde{\mathbf{\Pi}}_\infty|\mathbf{vec}(\mathbf{\Sigma_w}))), \tag{185e}$$

where $|\cdot|$ here denotes the element-wise absolute value of a matrix.

Now, let us consider (L1). We observe that

$$\begin{aligned} \mathbf{tr}(\mathbf{M}\mathcal{H}^{-1}(\tilde{\mathbf{L}}^t\mathbf{v})) = \mathbf{tr}(\sum_{i=1}^{s} \mathbf{M}_i [\mathcal{H}^{-1}(\tilde{\mathbf{L}}^t\mathbf{v})]_i) &\le \|\mathbf{M}_{1:s}\| \cdot \mathbf{tr}(\sum_{i=1}^{s} [\mathcal{H}^{-1}(\tilde{\mathbf{L}}^t\mathbf{v})]_i) \\ &\le \sqrt{n}\|\mathbf{M}_{1:s}\| \|\sum_{i=1}^{s} [\mathcal{H}^{-1}(\tilde{\mathbf{L}}^t\mathbf{v})]_i\|_F, \end{aligned} \tag{186}$$

where the first inequality uses (185a) 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 (186), we have

$$\begin{aligned} \|\sum_{i=1}^{s} [\mathcal{H}^{-1}(\tilde{\mathbf{L}}^t\mathbf{v})]_i\|_F \le \sum_{i=1}^{s} \|[\mathcal{H}^{-1}(\tilde{\mathbf{L}}^t\mathbf{v})]_i\|_F &\le \sqrt{s}\sqrt{\sum_{i=1}^{s} \|[\mathcal{H}^{-1}(\tilde{\mathbf{L}}^t\mathbf{v})]_i\|_F^2} \\ &= \sqrt{s}\|\mathcal{H}^{-1}(\tilde{\mathbf{L}}^t\mathbf{v})\|_F \\ &= \sqrt{s}\|\tilde{\mathbf{L}}^t\mathbf{v}\| \\ &\le \sqrt{s}\|\tilde{\mathbf{L}}^t\|\|\mathbf{v}\|, \end{aligned} \tag{187}$$

where the second equality holds since $\mathcal{H}^{-1}$ is a reshaping operator, and $\tilde{\mathbf{L}}^t\mathbf{v}$ is a vector. Substituting (187) into (186) gives (L1).

To show (L2), we combine (185b) with (L1) to get $\mathbf{tr}(\mathbf{M}\mathcal{H}^{-1}(\tilde{\mathbf{L}}^t\tilde{\mathbf{B}}_{t'}\mathbf{vec}(\mathbf{\Sigma_z}))) \le \sqrt{ns}\|\mathbf{M}_{1:s}\|\|\tilde{\mathbf{L}}^t\|\|\tilde{\mathbf{B}}_{t'}\mathbf{vec}(\mathbf{\Sigma_z})\|$. Then, using the upper bound for $\|\tilde{\mathbf{B}}_{t'}\mathbf{vec}(\mathbf{\Sigma_z})\|$ derived in (22) completes the proof of (L2).

To establish (L3), we combine (185c) with (L1) to obtain

$$\mathbf{tr}(\mathbf{M}\mathcal{H}^{-1}(\tilde{\mathbf{L}}^t\tilde{\mathbf{\Pi}}_{t'}\mathbf{vec}(\mathbf{\Sigma_w}))) \le \sqrt{ns}\|\mathbf{M}_{1:s}\|\|\tilde{\mathbf{L}}^t\|\|\tilde{\mathbf{\Pi}}_{t'}\mathbf{vec}(\mathbf{\Sigma_w})\|. \tag{188}$$

Then, using the upper bound for $\|\tilde{\mathbf{\Pi}}_{t'}\mathbf{vec}(\mathbf{\Sigma_w})\|$ derived in (23) gives (L2).

Finally, let us consider (L4). It follows from (185e) and (185d) in conjunction with (L1) that

$$|\mathbf{tr}(\mathbf{M}\mathcal{H}^{-1}(\tilde{\mathbf{L}}^t|\tilde{\mathbf{\Pi}}_{t'} - \tilde{\mathbf{\Pi}}_\infty|\mathbf{vec}(\mathbf{\Sigma_w})))| \le \sqrt{ns}\|\mathbf{M}_{1:s}\|\|\tilde{\mathbf{L}}^t\|\||\tilde{\mathbf{\Pi}}_{t'} - \tilde{\mathbf{\Pi}}_\infty|\mathbf{vec}(\mathbf{\Sigma_w})\|. \tag{189}$$

Now, using (178), we obtain

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

where the last line follows from Definition 3. Substituting the above inequality in (189) completes the proof of (L4). □

The following lemma provides a bound for the difference $J_T(\mathbf{x}_0, \omega(0), \{\mathbf{K}_{1:s}, \mathbf{\Sigma}_\mathbf{z}\}) - TJ(0, \omega(0), \{\mathbf{K}_{1:s}\})$ using an arbitrary stabilizing controller $\mathbf{K}_{1:s}$. Based on this result, we will provide in Proposition 1 a uniform upper bound for this difference when using any controllers $\mathbf{K}_{1:s}$ that are close to $\mathbf{K}^\star_{1:s}$.

Lemma 19. *For an arbitrary stabilizing controller $\mathbf{K}_{1:s}$, we have*

$$\begin{aligned}
&J_T(\mathbf{x}_0, \omega(0), \{\mathbf{K}_{1:s}, \mathbf{\Sigma}_\mathbf{z}\}) - TJ(0, \omega(0), \{\mathbf{K}_{1:s}\}) \\
\leq &\sqrt{ns}\|\mathbf{M}_{1:s}\| \cdot \|\mathbf{x}_0\|^2 \\
&+ \frac{n\sqrt{s}\tau_{\tilde{\mathbf{L}}}}{1-\rho_{\tilde{\mathbf{L}}}}\|\mathbf{M}_{1:s}\|\|\mathbf{B}_{1:s}\|^2\|\mathbf{\Sigma}_\mathbf{z}\|T \\
&+ n\|\mathbf{R}_{1:s}\|\|\mathbf{\Sigma}_\mathbf{z}\|T \\
&+ n\sqrt{s}\tau_{MC}\tau_{\tilde{\mathbf{L}}}\|\mathbf{M}_{1:s}\|\|\mathbf{\Sigma}_\mathbf{w}\|\frac{\rho_{MC}}{\rho_{MC}-\rho_{\tilde{\mathbf{L}}}}\left(\frac{\rho_{MC}}{1-\rho_{MC}} - \frac{\rho_{\tilde{\mathbf{L}}}}{1-\rho_{\tilde{\mathbf{L}}}}\right)
\end{aligned} \tag{190}$$

where τ_{MC} and ρ_{MC} are given in Definition 3, $\tau_{\tilde{\mathbf{L}}}$ and $\rho_{\tilde{\mathbf{L}}}$ are constants defined in the beginning of Appendix C, and $\mathbf{M} = [\mathbf{M}_1, \dots, \mathbf{M}_s]$ with $\mathbf{M}_i = \mathbf{Q}_i + \mathbf{K}_i^\top \mathbf{R}_i \mathbf{K}_i$.

Proof. From Lemma 17, we know

$$J_T(\mathbf{x}_0, \omega(0), \{\mathbf{K}_{1:s}, \mathbf{\Sigma}_\mathbf{z}\}) = S_{0;T} + S_{\mathbf{z};1;T} + S_{\mathbf{z};2;T} + S_{\mathbf{w};T} \tag{191}$$

$$J(0, \omega(0), \{\mathbf{K}_{1:s}\}) = \limsup_{T\to\infty}\frac{1}{T}(S_{0;T} + S_{\mathbf{w};T}) =: S_0 + S_\mathbf{w}. \tag{192}$$

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

For $S_{0;T}$, letting $\mathbf{s}_0 = \begin{bmatrix}\mathbf{vec}(\mathbf{\Sigma}_1(0)) \\ \vdots \\ \mathbf{vec}(\mathbf{\Sigma}_s(0))\end{bmatrix}$, we have

$$\begin{aligned}
S_{0;T} &= \sum_{t=1}^{T}\mathbf{tr}(\mathbf{M}\mathcal{H}^{-1}(\tilde{\mathbf{L}}^t\mathbf{s}_0)) \\
&\leq \sqrt{ns}\|\mathbf{M}_{1:s}\|\|\tilde{\mathbf{L}}^t\|\|\mathbf{s}_0\| \\
&\leq \sqrt{ns}\|\mathbf{M}_{1:s}\| \cdot \mathsf{E}[\|\mathbf{x}_0\|^2] \\
&= \sqrt{ns}\|\mathbf{M}_{1:s}\| \cdot \|\mathbf{x}_0\|^2
\end{aligned} \tag{193}$$

where the second line follows from Item (L1) in Lemma 18; the third line follows from (21) in Lemma 2. 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 $\|\mathbf{x}_0\|^2$ is bounded.

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

$$\begin{aligned}
S_{\mathbf{z};1;T} &= \sum_{t=1}^{T}\sum_{t'=0}^{t-1}\mathbf{tr}(\mathbf{M}\mathcal{H}^{-1}(\tilde{\mathbf{L}}^{t'}\tilde{\mathbf{B}}_{t-t'}\mathbf{vec}(\mathbf{\Sigma}_\mathbf{z}))) \\
&\leq n\sqrt{s}\|\mathbf{M}_{1:s}\|\|\mathbf{B}_{1:s}\|^2\|\mathbf{\Sigma}_\mathbf{z}\|\left(\sum_{t=1}^{T}\sum_{t'=0}^{t-1}\|\tilde{\mathbf{L}}^{t'}\|\right) \\
&\leq \frac{n\sqrt{s}\tau_{\tilde{\mathbf{L}}}}{1-\rho_{\tilde{\mathbf{L}}}}\|\mathbf{M}_{1:s}\|\|\mathbf{B}_{1:s}\|^2\|\mathbf{\Sigma}_\mathbf{z}\|T
\end{aligned} \tag{194}$$

where the first inequality follows from Item (L2) in Lemma 18, and the second inequality follows from the fact $\|\tilde{\mathbf{L}}^{t'}\| \leq \tau_{\tilde{\mathbf{L}}}\rho_{\tilde{\mathbf{L}}}^{t'}$

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

$$S_{\mathbf{z},2,T} = \sum_{t=1}^{T} \mathbf{tr}(\sum_{i=1}^{s} \pi_t(i)\mathbf{R}_i \mathbf{\Sigma}_{\mathbf{z}}) \leq n\|\mathbf{R}_{1:s}\|\|\mathbf{\Sigma}_{\mathbf{z}}\|T \tag{195}$$

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

$$S_{\mathbf{w},T} = \sum_{t=1}^{T}\sum_{t'=0}^{t-1} \mathbf{tr}(\mathrm{M}\mathcal{H}^{-1}(\tilde{\mathbf{L}}^{t'}\tilde{\mathbf{\Pi}}_{t-t'}\mathbf{vec}(\mathbf{\Sigma}_{\mathbf{w}}))). \tag{196}$$

To evaluate it, we first define the following terms:

$$S_{\mathbf{w},T}^{(\infty)} := \sum_{t=1}^{T}\sum_{t'=0}^{t-1} \mathbf{tr}(\mathrm{M}\mathcal{H}^{-1}(\tilde{\mathbf{L}}^{t'}\tilde{\mathbf{\Pi}}_{\infty}\mathbf{vec}(\mathbf{\Sigma}_{\mathbf{w}}))), \tag{197}$$

$$S_{\mathbf{w}}^{(\infty)} := \limsup_{T\to\infty} \frac{1}{T} S_{\mathbf{w},T}^{(\infty)}, \tag{198}$$

where $\tilde{\mathbf{\Pi}}_{\infty}$ is defined in (178). 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 π_0 is the stationary distribution π_{∞}

Then, we have

$$\begin{aligned} |S_{\mathbf{w},T} - S_{\mathbf{w},T}^{(\infty)}| &= \left| \sum_{t=1}^{T}\sum_{t'=0}^{t-1} \mathbf{tr}(\mathrm{M}\mathcal{H}^{-1}(\tilde{\mathbf{L}}^{t'}(\tilde{\mathbf{\Pi}}_{t-t'} - \tilde{\mathbf{\Pi}}_{\infty})\mathbf{vec}(\mathbf{\Sigma}_{\mathbf{w}}))) \right| \\ &\leq \tau_{MC} n\sqrt{s}\|\mathbf{M}_{1:s}\|\|\mathbf{\Sigma}_{\mathbf{w}}\|(\sum_{t=1}^{T}\sum_{t'=0}^{t-1} \|\tilde{\mathbf{L}}^{t'}\|\rho_{MC}^{t-t'}) \\ &\leq \tau_{MC} n\sqrt{s}\|\mathbf{M}_{1:s}\|\|\mathbf{\Sigma}_{\mathbf{w}}\|(\sum_{t=1}^{\infty}\sum_{t'=0}^{t-1} \tau_{\tilde{\mathbf{L}}}\rho_{\tilde{\mathbf{L}}}^{t'}\rho_{MC}^{t-t'}) \\ &\leq n\sqrt{s}\tau_{MC}\tau_{\tilde{\mathbf{L}}}\|\mathbf{M}_{1:s}\|\|\mathbf{\Sigma}_{\mathbf{w}}\| \frac{\rho_{MC}}{\rho_{MC} - \rho_{\tilde{\mathbf{L}}}}(\frac{\rho_{MC}}{1-\rho_{MC}} - \frac{\rho_{\tilde{\mathbf{L}}}}{1-\rho_{\tilde{\mathbf{L}}}}) \end{aligned} \tag{199}$$

where the first inequality follows from Item (L4) in Lemma 18. Thus,

$$S_{\mathbf{w}} = \limsup_{T\to\infty}\frac{1}{T}S_{\mathbf{w},T} = \limsup_{T\to\infty}\frac{1}{T}(S_{\mathbf{w},T} - S_{\mathbf{w},T}^{(\infty)}) + \limsup_{T\to\infty}\frac{1}{T}S_{\mathbf{w},T}^{(\infty)} = S_{\mathbf{w}}^{(\infty)}. \tag{200}$$

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

$$\begin{aligned} S_{\mathbf{w}} = S_{\mathbf{w}}^{(\infty)} &= \mathbf{tr}(\mathrm{M}\mathcal{H}^{-1}(\limsup_{T\to\infty}\frac{1}{T}\sum_{t=1}^{T}\sum_{t'=0}^{t-1}\tilde{\mathbf{L}}^{t'}\tilde{\mathbf{\Pi}}_{\infty}\mathbf{vec}(\mathbf{\Sigma}_{\mathbf{w}}))) \\ &= \mathbf{tr}(\mathrm{M}\mathcal{H}^{-1}((\mathbf{I}-\tilde{\mathbf{L}})^{-1}\tilde{\mathbf{\Pi}}_{\infty}\mathbf{vec}(\mathbf{\Sigma}_{\mathbf{w}}))) \\ &= \sum_{t'=0}^{\infty}\mathbf{tr}(\mathrm{M}\mathcal{H}^{-1}(\tilde{\mathbf{L}}^{t'}\tilde{\mathbf{\Pi}}_{\infty}\mathbf{vec}(\mathbf{\Sigma}_{\mathbf{w}}))). \end{aligned} \tag{201}$$

Thus,

$$\begin{aligned} TS_{\mathbf{w}} = TS_{\mathbf{w}}^{(\infty)} &= \sum_{t=1}^{T}\sum_{t'=0}^{\infty}\mathbf{tr}(\mathrm{M}\mathcal{H}^{-1}(\tilde{\mathbf{L}}^{t'}\tilde{\mathbf{\Pi}}_{\infty}\mathbf{vec}(\mathbf{\Sigma}_{\mathbf{w}}))) \\ &\geq \sum_{t=1}^{T}\sum_{t'=0}^{t-1}\mathbf{tr}(\mathrm{M}\mathcal{H}^{-1}(\tilde{\mathbf{L}}^{t'}\tilde{\mathbf{\Pi}}_{\infty}\mathbf{vec}(\mathbf{\Sigma}_{\mathbf{w}}))) \\ &= S_{\mathbf{w},T}^{(\infty)} \end{aligned} \tag{202}$$

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

$$
\begin{aligned}
S_{\mathbf{w};T} &\leq S^{(1)}_{\mathbf{w};T} + |S_{\mathbf{w};T} - S^{(1)}_{\mathbf{w};T}| \\
&\overset{(200)}{\leq} T S_{\mathbf{w}} + |S_{\mathbf{w};T} - S^{(1)}_{\mathbf{w};T}| \\
&\overset{(199)}{\leq} T S_{\mathbf{w}} + n\sqrt{s}\tau_{MC}\tau_{\tilde{\mathbf{L}}}\|\mathbf{M}_{1:s}\|\|\boldsymbol{\Sigma}_{\mathbf{w}}\|\frac{\rho_{MC}}{\rho_{MC}-\rho_{\tilde{\mathbf{L}}}}\left(\frac{\rho_{MC}}{1-\rho_{MC}} - \frac{\rho_{\tilde{\mathbf{L}}}}{1-\rho_{\tilde{\mathbf{L}}}}\right).
\end{aligned}
\tag{203}
$$

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

$$
\begin{aligned}
&J_T(\mathbf{x}_0, \omega(0), \{\mathbf{K}_{1:s}, \boldsymbol{\Sigma}_{\mathbf{z}}\}) - TJ(0, \omega(0), \{\mathbf{K}_{1:s}\}) \\
=& S_{0;T} + S_{\mathbf{z};1;T} + S_{\mathbf{z};2;T} + S_{\mathbf{w};T} - T(S_0 + S_{\mathbf{w}}) \\
\leq& \sqrt{ns}\|\mathbf{M}_{1:s}\| \cdot \|\mathbf{x}_0\|^2 \\
&+ \frac{n\sqrt{s}\tau_{\tilde{\mathbf{L}}}}{1-\rho_{\tilde{\mathbf{L}}}}\|\mathbf{M}_{1:s}\|\|\mathbf{B}_{1:s}\|^2\|\boldsymbol{\Sigma}_{\mathbf{z}}\|T \\
&+ n\|\mathbf{R}_{1:s}\|\|\boldsymbol{\Sigma}_{\mathbf{z}}\|T \\
&+ n\sqrt{s}\tau_{MC}\tau_{\tilde{\mathbf{L}}}\|\mathbf{M}_{1:s}\|\|\boldsymbol{\Sigma}_{\mathbf{w}}\|\frac{\rho_{MC}}{\rho_{MC}-\rho_{\tilde{\mathbf{L}}}}\left(\frac{\rho_{MC}}{1-\rho_{MC}} - \frac{\rho_{\tilde{\mathbf{L}}}}{1-\rho_{\tilde{\mathbf{L}}}}\right)
\end{aligned}
\tag{204}
$$

which concludes the proof. □

We now provide a uniform upper bound on the regret $J_T(\mathbf{x}_0, \omega(0), \{\mathbf{K}_{1:s}, \boldsymbol{\Sigma}_{\mathbf{z}}\}) - TJ^?$ for any stabilizing controller $\mathbf{K}_{1:s}$ that is close enough to the optimal controller $\mathbf{K}^?_{1:s}$.

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

$$
\begin{aligned}
J_T(\mathbf{x}_0, \omega(0), \{\mathbf{K}_{1:s}, \boldsymbol{\Sigma}_{\mathbf{z}}\}) - TJ^? \leq& C^J_{\mathbf{K}}\|\mathbf{K}_{1:s} - \mathbf{K}^?_{1:s}\|^2\|\boldsymbol{\Sigma}_{\mathbf{w}}\|T \\
&+ \sqrt{ns}M\|\mathbf{x}_0\|^2 \\
&+ n\sqrt{s}\frac{2\tau(\tilde{\mathbf{L}}^?)\|\mathbf{B}_{1:s}\|^2 M}{1-\rho^?}\|\boldsymbol{\Sigma}_{\mathbf{z}}\|T \\
&+ n\|\mathbf{R}_{1:s}\|\|\boldsymbol{\Sigma}_{\mathbf{z}}\|T \\
&+ n\sqrt{s}\frac{2\tau(\tilde{\mathbf{L}}^?)\tau_{MC}M\rho_{MC}}{2\rho_{MC}-1-\rho^?}\left(\frac{\rho_{MC}}{1-\rho_{MC}} - \frac{1+\rho}{1-\rho}\right)\|\boldsymbol{\Sigma}_{\mathbf{w}}\|,
\end{aligned}
\tag{205}
$$

where $M := \|\mathbf{Q}_{1:s}\| + 4\|\mathbf{R}_{1:s}\|\|\mathbf{K}^?_{1:s}\|^2$, and $\bar{\epsilon}_{\mathbf{K}}$ and $C^J_{\mathbf{K}}$ are defined in Table 3.

Proof. When $\|\mathbf{K}_{1:s} - \mathbf{K}^?_{1:s}\| \leq \bar{\epsilon}_{\mathbf{K}}$, from Lemma 15, we know $\|\tilde{\mathbf{L}}^k\| \leq \tau(\tilde{\mathbf{L}}^?)(\frac{1+~^*}{2})^k$, thus we could set $\tau_{\tilde{\mathbf{L}}}$ and $\rho_{\tilde{\mathbf{L}}}$ to be $\tau(\tilde{\mathbf{L}}^?)$ and $\frac{1+~^*}{2}$. By definition, we know $\bar{\epsilon}_{\mathbf{K}} \leq \|\mathbf{K}^?_{1:s}\|$, thus $\|\mathbf{M}_{1:s}\| \leq \|\mathbf{Q}_{1:s}\| + \|\mathbf{R}_{1:s}\|\|\mathbf{K}_{1:s}\|^2 \leq \|\mathbf{Q}_{1:s}\| + \|\mathbf{R}_{1:s}\|(\|\mathbf{K}^?_{1:s}\| + \bar{\epsilon}_{\mathbf{K}})^2 \leq \|\mathbf{Q}_{1:s}\| + 4\|\mathbf{R}_{1:s}\|\|\mathbf{K}^?_{1:s}\|^2 = M$. Then applying Lemma 19, we have

$$
\begin{aligned}
&J_T(\mathbf{x}_0, \omega(0), \{\mathbf{K}_{1:s}, \boldsymbol{\Sigma}_{\mathbf{z}}\}) - TJ(0, \omega(0), \{\mathbf{K}_{1:s}\}) \\
\leq& \sqrt{ns}M\|\mathbf{x}_0\|^2 \\
&+ n\sqrt{s}\frac{2\tau(\tilde{\mathbf{L}}^?)\|\mathbf{B}_{1:s}\|^2 M}{1-\rho^?}\|\boldsymbol{\Sigma}_{\mathbf{z}}\|T \\
&+ n\|\mathbf{R}_{1:s}\|\|\boldsymbol{\Sigma}_{\mathbf{z}}\|T \\
&+ n\sqrt{s}\frac{2\tau(\tilde{\mathbf{L}}^?)\tau_{MC}M\rho_{MC}}{2\rho_{MC}-1-\rho^?}\left(\frac{\rho_{MC}}{1-\rho_{MC}} - \frac{1+\rho}{1-\rho}\right)\|\boldsymbol{\Sigma}_{\mathbf{w}}\|
\end{aligned}
\tag{206}
$$

Now note that when $\|\mathbf{K}_{1:s} - \mathbf{K}^?_{1:s}\| \leq \bar{\epsilon}_{\mathbf{K}}$, we have $J(0, \omega(0), \{\mathbf{K}_{1:s}\}) - J^? \leq C^J_{\mathbf{K}}\|\boldsymbol{\Sigma}_{\mathbf{w}}\|\|\mathbf{K}_{1:s} - \mathbf{K}^?_{1:s}\|^2$ using Lemma 16. Combining this with (206), we could conclude the proof. □

C.3. Stitching Every Epoch — Proof for Theorem 2

In this section, we stitch the upper bounds on 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 $\epsilon_{\mathbf{A},\mathbf{B}}^{(i)} = \max\{\|\mathbf{A}_{1:s}^{(i)} - \mathbf{A}_{1:s}\|, \|\mathbf{B}_{1:s}^{(i)} - \mathbf{B}_{1:s}\|\}, \epsilon_{\mathbf{T}}^{(i)} = \|\mathbf{T}^{(i)} - \mathbf{T}\|_{\infty}$, and define $\epsilon_{\mathbf{K}}^{(i)} := \|\mathbf{K}_{1:s}^{(i)} - \mathbf{K}_{1:s}^{\star}\|$ where $\mathbf{K}_{1:s}^{\star}$ is the optimal controller for the infinite-horizon MJS-LQR$(\mathbf{A}_{1:s}, \mathbf{B}_{1:s}, \mathbf{T}, \mathbf{Q}_{1:s}, \mathbf{R}_{1:s})$. We define the following events for every epoch i.

$$
\begin{aligned}
\mathcal{A}_i &= \left\{ \text{Regret}_i \leq \mathcal{O}\left(c_A + \|\mathbf{x}_0^{(i)}\|^2 + T_i \sigma_{\mathbf{z},i}^2 + T_i \sigma_{\mathbf{w}}^2 \left(\epsilon_{\mathbf{A},\mathbf{B}}^{(i-1)} + \epsilon_{\mathbf{T}}^{(i-1)} \right)^2 \right) \right\} \\
\mathcal{B}_i &= \left\{ \epsilon_{\mathbf{A},\mathbf{B}}^{(i)} \leq \bar{\epsilon}_{\mathbf{A},\mathbf{B},\mathbf{T}}, \epsilon_{\mathbf{T}}^{(i)} \leq \bar{\epsilon}_{\mathbf{A},\mathbf{B},\mathbf{T}}, \epsilon_{\mathbf{K}}^{(i+1)} \leq \bar{\epsilon}_{\mathbf{K}} \right\} \\
\mathcal{C}_i &= \left\{ \epsilon_{\mathbf{A},\mathbf{B}}^{(i)} \leq \tilde{\mathcal{O}}\left(\frac{\sigma_{\mathbf{z},i} + \sigma_{\mathbf{w}}}{\sigma_{\mathbf{z},i} \pi_{\min}} \frac{(n+p)\log(T_i)}{\sqrt{T_i}} \right), \epsilon_{\mathbf{T}}^{(i)} \leq \tilde{\mathcal{O}}\left(\frac{1}{\pi_{\min}} \sqrt{\frac{\log(T_i)}{T_i}} \right) \right\} \\
\mathcal{D}_i &= \left\{ \|\mathbf{x}_0^{(i+1)}\|^2 = \|\mathbf{x}_{T_i}^{(i)}\|^2 \leq \hat{\mathcal{O}}(\bar{x}_0^2) \right\}.
\end{aligned}
\tag{207}
$$

where $c_A, \bar{x}_0$ are constants, $\bar{\epsilon}_{\mathbf{A},\mathbf{B},\mathbf{T}}$ and $\bar{\epsilon}_{\mathbf{K}}$ are defined in Table 3, and $\hat{\mathcal{O}}$ and $\tilde{\mathcal{O}}(\cdot)$ hide $\text{poly}(\frac{1}{\delta})$ and $\text{polylog}(\frac{1}{\delta})$ respectively. Note that $O, \hat{\mathcal{O}}$, and $\tilde{\mathcal{O}}$ may hide terms that are invariant to epochs such as $\rho^{\star}$, $\|\mathbf{A}_{1:s}\|$, $\|\mathbf{B}_{1:s}\|$, etc.

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 $\mathbf{x}_0, \mathbf{w}_t, \mathbf{z}_t, \omega(t)$ 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 $(i+1)$th epoch with randomness coming from $\mathbf{x}_0^{(i+1)} = \mathbf{x}_{T_i}^{(i)}, \omega^{(i+1)}(0) = \omega^{(i)}(T_i)$, and controller $\mathbf{K}_{1:s}^{(i+1)}$ computed from $\mathbf{A}_{1:s}^{(i)}, \mathbf{B}_{1:s}^{(i)}, \mathbf{T}^{(i)}$, thus $\mathcal{A}_{i+1}$ is $\mathcal{F}_i$-measurable. Let $\mathcal{E}_i := \mathcal{A}_{i+1} \cap \mathcal{B}_i \cap \mathcal{C}_i \cap \mathcal{D}_i$.

Then, we have the following results regarding the conditional probabilities of these events.

Proposition 2. *Suppose for every epoch i, its epoch length $T_i \geq \underline{T}_{rgt,\epsilon}(\delta)$. Then,*

$$
\mathsf{P}(\mathcal{B}_i \mid \mathcal{C}_i, \cap_{j=0}^{i-1} \mathcal{E}_j) = \mathsf{P}(\mathcal{B}_i \mid \mathcal{C}_i) = 1 \tag{208}
$$

Proof. When $\mathcal{C}_i$ occurs, we have $\epsilon_{\mathbf{A},\mathbf{B}}^{(i)} \leq \tilde{\mathcal{O}}(\frac{1}{T_i^{0.25}}), \epsilon_{\mathbf{T}}^{(i)} \leq \tilde{\mathcal{O}}(\frac{1}{T_i^{0.25}})$ as $\sigma_{\mathbf{z},i}^2 = \sigma_{\mathbf{w}}^2/\sqrt{T_i}$. We know when $T_i \geq \tilde{\mathcal{O}}(\bar{\epsilon}_{\mathbf{A},\mathbf{B},\mathbf{T}}^{-5})$, we have $\epsilon_{\mathbf{A},\mathbf{B}}^{(i)} \leq \bar{\epsilon}_{\mathbf{A},\mathbf{B},\mathbf{T}}, \epsilon_{\mathbf{T}}^{(i)} \leq \bar{\epsilon}_{\mathbf{A},\mathbf{B},\mathbf{T}}$. Then according to Lemma 16, we have $\epsilon_{\mathbf{K}}^{(i+1)} \leq \bar{\epsilon}_{\mathbf{K}}$. 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. ☐

Proposition 3. *For $c_{\mathbf{x}} \geq \underline{c}_{\mathbf{x}}(\bar{\rho}, \bar{\tau}), c_{\mathbf{z}} \geq \underline{c}_{\mathbf{z}}$, $C_{sub} \geq \underline{C}_{sub,id,N}\left(\bar{x}_0, \frac{\delta}{2}, T_i, \bar{\rho}, \bar{\tau}\right)$, $T_i \geq \underline{T}_{id,N}(\frac{\delta}{2}, \bar{\rho}, \bar{\tau})$, we have*

$$
\mathsf{P}(\mathcal{C}_i \mid \cap_{j=0}^{i-1} \mathcal{E}_j)) = \mathsf{P}(\mathcal{C}_i \mid \mathcal{B}_{i-1}, \mathcal{D}_{i-1}) \geq 1 - \delta. \tag{209}
$$

and $\mathsf{P}(\mathcal{C}_0) \geq 1 - \delta$.

Proof. By Lemma 6, we know for every epoch $i = 0, 1, \ldots$, when $C_{sub} \geq \underline{C}_{sub,MC}$, $T_i \geq \underline{T}_{MC,1}(\frac{\delta}{8})$, we have with probability at least $1 - \frac{\delta}{2}$, $\epsilon_{\mathbf{T}}^{(i)} \leq \tilde{\mathcal{O}}\left(\frac{1}{\pi_{\min}} \sqrt{\frac{\log(T_i)}{T_i}} \right)$.

Note that for all i, given event $\mathcal{D}_{i-1}$, we know $\|\mathbf{x}_0^{(i)}\| \leq \hat{\mathcal{O}}(\bar{x}_0)$. And for epoch 0, we have $\|\mathbf{x}_0^{(0)}\| = 0$. For epoch $i = 1, 2, \ldots$, given event $\mathcal{B}_{i-1}$, we know $\epsilon_{\mathbf{K}}^{(i+1)} \leq \bar{\epsilon}_{\mathbf{K}}$. Let $\tilde{\mathbf{L}}^{(i)}$ denote the augmented closed-loop state matrix. By Lemma 15, we know $\|(\tilde{\mathbf{L}}^{(i)})^k\| \leq \tau(\tilde{\mathbf{L}}^{\star})(\frac{1+\rho^{\star}}{2})^k$. Thus, if we pick $\bar{\tau} := \max\{\tau(\tilde{\mathbf{L}}^{(0)}), \tau(\tilde{\mathbf{L}}^{\star})\}, \bar{\rho} := \max\{\rho(\tilde{\mathbf{L}}^{(0)}), \frac{1+\rho^{\star}}{2}\}$, we know for every epoch $i = 0, 1, 2, \ldots$, we have $\|(\tilde{\mathbf{L}}^{(i)})^k\| \leq \bar{\tau}\bar{\rho}^k$. Then, from Theorem 7, we know for every i: when $c_{\mathbf{x}} \geq \underline{c}_{\mathbf{x}}(\bar{\rho}, \bar{\tau}), c_{\mathbf{z}} \geq \underline{c}_{\mathbf{z}}$, $T_i \geq \underline{T}_{id,N}(\frac{\delta}{2}, \bar{\rho}, \bar{\tau})$, $C_{sub} \geq \underline{C}_{sub,id,N}\left(\bar{x}_0, \frac{\delta}{2}, T_i, \bar{\rho}, \bar{\tau}\right)$, with probability at least $1 - \delta$, $\epsilon_{\mathbf{A},\mathbf{B}}^{(i)} \leq \tilde{\mathcal{O}}\left(\frac{\sigma_{\mathbf{z},i} + \sigma_{\mathbf{w}}}{\sigma_{\mathbf{z},i} \pi_{\min}} \frac{(n+p)\log(T_i)}{\sqrt{T_i}} \right)$.

By union bound, we could show $\mathsf{P}(\mathcal{C}_0) \geq 1 - \delta$ and $\mathsf{P}(\mathcal{C}_i \mid \mathcal{B}_{i-1}, \mathcal{D}_{i-1}) \geq 1 - \delta$. 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 $\mathsf{P}(\mathcal{C}_i \mid \cap_{j=0}^{i-1} \mathcal{E}_j)) = \mathsf{P}(\mathcal{C}_i \mid \mathcal{B}_{i-1}, \mathcal{D}_{i-1})$. ☐

Proposition 4. *Suppose* $\bar{x}_0^2 \geq \frac{n\sqrt{s}(\|\mathbf{B}_{1:s}\|^2+1)\sigma_{\mathbf{w}}^2}{(1-\bar{\rho})(1-\sqrt{ns}\bar{\tau}\bar{\rho}^{T_0})}$ *Then, for every epoch* i,

$$\mathsf{P}(\mathcal{D}_i \mid \cap_{j=0}^{i-1}\mathcal{E}_j) = \mathsf{P}(\mathcal{D}_i \mid \mathcal{B}_{i-1}\mathcal{D}_{i-1}) > 1-\delta \tag{210}$$

and $\mathsf{P}(\mathcal{D}_0) \geq 1-\delta$.

Proof. For epoch $i = 1, 2, \ldots$, given $\epsilon_{\mathbf{K}}^{(i)} \leq \bar{\epsilon}_{\mathbf{K}}$ in event $\mathcal{B}_{i-1}$, we know by Lemma 15 that controller $\mathbf{K}_{1:s}^{(i)}$ gives augmented state matrix $\tilde{\mathbf{L}}^{(i)}$ such that $\|(\tilde{\mathbf{L}}^{(i)})^k\| \leq \tau(\tilde{\mathbf{L}}^\star)(\frac{1+\rho^\star}{2})^k$. By definitions of $\bar{\tau}$ and $\bar{\rho}$, we further have $\|(\tilde{\mathbf{L}}^{(i)})^k\| \leq \bar{\tau}\bar{\rho}^k$.

Event $\mathcal{D}_{i-1}$ implies $\|\mathbf{x}_0^{(i)}\|^2 \leq \hat{\mathcal{O}}(\bar{x}_0^2)$. Then, according to Lemma 2, we know

$$\begin{aligned}\mathsf{E}[\|\mathbf{x}_{T_i}^{(i)}\|^2 \mid \mathcal{B}_{i-1}, \mathcal{D}_{i-1}] &\leq \sqrt{ns}\cdot\bar{\tau}\bar{\rho}^{T_i}\hat{\mathcal{O}}(\bar{x}_0^2) + n\sqrt{s}(\|\mathbf{B}_{1:s}\|^2\frac{\sigma_{\mathbf{w}}^2}{\sqrt{T_i}} + \sigma_{\mathbf{w}}^2)\frac{\bar{\tau}}{1-\bar{\rho}}. \\ &\leq \sqrt{ns}\cdot\bar{\tau}\bar{\rho}^{T_i}\hat{\mathcal{O}}(\bar{x}_0^2) + (1-\sqrt{ns}\bar{\tau}\bar{\rho}^{T_0})\bar{x}_0^2 \\ &\leq \hat{\mathcal{O}}(\bar{x}_0^2)\end{aligned} \tag{211}$$

where the second line follows from the assumption on $\bar{x}_0^2$ in the proposition statement. Using Markov inequality, we have

$$\mathsf{P}(\|\mathbf{x}_{T_i}^{(i)}\|^2 \leq \frac{\hat{\mathcal{O}}(\bar{x}_0^2)}{\delta} \mid \mathcal{B}_{i-1}, \mathcal{D}_{i-1}) \geq 1-\delta. \tag{212}$$

which implies $\mathsf{P}(\mathcal{D}_i \mid \mathcal{B}_{i-1}, \mathcal{D}_{i-1}) \geq 1-\delta$. Similarly, we could show $\mathsf{P}(\mathcal{D}_0) \geq 1-\delta$.

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 5. *We have*

$$\mathsf{P}(\mathcal{A}_i \mid \mathcal{B}_{i-1}, \mathcal{C}_{i-1}, \mathcal{D}_{i-1}, \cap_{j=0}^{i-2}\mathcal{E}_j) = \mathsf{P}(\mathcal{A}_i \mid \mathcal{B}_{i-1}) = 1. \tag{213}$$

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

$$\begin{aligned}\text{Regret}_i \leq & C_{\mathbf{K}}^J\|\mathbf{K}_{1:s}^{(i)} - \mathbf{K}_{1:s}^\star\|^2\sigma_{\mathbf{w}}^2 T_i \\ &+\sqrt{ns}M\|\mathbf{x}_0^{(i)}\|^2 \\ &+n\sqrt{s}\frac{2\tau(\tilde{\mathbf{L}}^\star)\|\mathbf{B}_{1:s}\|^2 M}{1-\rho^\star}\sigma_{\mathbf{z};i}^2 T_i \\ &+n\|\mathbf{R}_{1:s}\|\sigma_{\mathbf{z};i}^2 T_i \\ &+n\sqrt{s}\frac{2\tau(\tilde{\mathbf{L}}^\star)\tau_{MC}M\rho_{MC}}{2\rho_{MC}-1-\rho^\star}(\frac{\rho_{MC}}{1-\rho_{MC}} - \frac{1+\rho}{1-\rho})\sigma_{\mathbf{w}}^2,\end{aligned} \tag{214}$$

Let c_A denote the last term in (214), which is a constant over epochs. Note that from $\epsilon_{\mathbf{A};\mathbf{B}}^{(i-1)} \leq \bar{\epsilon}_{\mathbf{A};\mathbf{B};\mathbf{T}}, \epsilon_{\mathbf{T}}^{(i-1)} \leq \bar{\epsilon}_{\mathbf{A};\mathbf{B};\mathbf{T}}$ in event $\mathcal{B}_{i-1}$, we know $\|\mathbf{K}_{1:s}^{(i)} - \mathbf{K}_{1:s}^\star\| \leq C_{\mathbf{A};\mathbf{B};\mathbf{T}}^{\mathbf{K}}(\epsilon_{\mathbf{A};\mathbf{B}}^{(i-1)} + \epsilon_{\mathbf{T}}^{(i-1)})$ by Lemma 16. Plugging this into (214), we have

$$\text{Regret}_i \leq O\left(s\cdot p\left(\epsilon_{\mathbf{A};\mathbf{B}}^{(i-1)} + \epsilon_{\mathbf{T}}^{(i-1)}\right)^2\sigma_{\mathbf{w}}^2 T_i + \sqrt{ns}\|\mathbf{x}_0^{(i)}\|^2 + n\sqrt{s}\sigma_{\mathbf{z};i}^2 T_i + c_A\right) \tag{215}$$

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 2

Theorem 8 (Complete version of Theorem 2). *If $c_{\mathbf{x}} \geq \underline{c}_{\mathbf{x}}(\bar{\rho}, \bar{\tau})$, $c_{\mathbf{z}} \geq \underline{c}_{\mathbf{z}}$, $C_{sub} \geq \mathcal{O}(\underline{C}_{sub;rgt}(\bar{x}_0, \delta, 2))$, $T_0 \geq \mathcal{O}(\underline{T}_{rgt}(\delta))$, then under Assumption A1 and A2, with probability at least $1-\delta$, Algorithm 2 achieves*

$$\text{Regret}(T) \leq \hat{\mathcal{O}}(\log(T)) + \tilde{\mathcal{O}}\left(\frac{s(n^2p+p^3)\sigma_{\mathbf{w}}^2}{\pi_{\min}^2}\sqrt{T}\log^2(T)\right). \tag{216}$$

Proof. In this proof, we will first show the intersected event $\cap_i \mathcal{E}_i$ implies the desired regret bound, then we evaluate the occurrence probability of $\cap_i \mathcal{E}_i$, and we left the discussion of requirements on tuning parameters $c_{\mathbf{x}}$, $c_{\mathbf{z}}$, C_{sub}, and T_0 in the end.

Using the definitions of events $\mathcal{A}_{i+1}$, $\mathcal{B}_i$, $\mathcal{C}_i$, $\mathcal{D}_i$, $T_i = \gamma T_{i-1}$, $\sigma_{\mathbf{z};i}^2 = \frac{\sigma_{\mathbf{w}}^2}{T_i}$, and (215) , we know event $\mathcal{E}_{i+1}$ implies the following.

$$\begin{aligned}
\text{Regret}_i &\leq \mathcal{O}(c_A) + \hat{\mathcal{O}}(\bar{x}_0^2) + \mathcal{O}(n\sqrt{s}\sigma_{\mathbf{w}}^2\sqrt{T_i}) \\
&\quad + \tilde{\mathcal{O}}(1)s\cdot pT_i\sigma_{\mathbf{w}}^2\left(\frac{\sigma_{\mathbf{z};i-1}+\sigma_{\mathbf{w}}}{\sigma_{\mathbf{z};i-1}\pi_{\min}}\cdot\frac{(n+p)\log(T_{i-1})}{\sqrt{T_{i-1}}} + \frac{\log(T_{i-1})}{\sqrt{T_{i-1}}}\right)^2 \\
&\leq \hat{\mathcal{O}}(1) + \mathcal{O}(n\sqrt{s}\sigma_{\mathbf{w}}^2\sqrt{T_i}) + \tilde{\mathcal{O}}(1)s\cdot p\left(\frac{n+p}{\pi_{\min}}\right)^2\sigma_{\mathbf{w}}^2 T_i\frac{\sigma_{\mathbf{w}}^2}{\sigma_{\mathbf{z};i-1}^2}\frac{\log^2(T_{i-1})}{T_{i-1}} \\
&\leq \hat{\mathcal{O}}(1) + \tilde{\mathcal{O}}(1)\frac{s(n^2p+p^3)\sigma_{\mathbf{w}}^2}{\pi_{\min}^2}\sqrt{\gamma}\sqrt{T_i}\log^2(T_i)
\end{aligned}\tag{217}$$

We have $M := \mathcal{O}(\log(\frac{T}{T_0}))$ 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

$$\begin{aligned}
&\text{Regret}(T) = \mathcal{O}(\sum_{i=0}^{M-1}\text{Regret}_i) \\
\leq&\hat{\mathcal{O}}(\log(T)) + \tilde{\mathcal{O}}(1)\frac{s(n^2p+p^3)\sigma_{\mathbf{w}}^2}{\pi_{\min}^2}\sqrt{\gamma}\sum_{i=0}^{M-1}\sqrt{T_i}\log^2(T_i) \\
\leq&\hat{\mathcal{O}}(\log(T)) + \tilde{\mathcal{O}}(1)\frac{s(n^2p+p^3)\sigma_{\mathbf{w}}^2}{\pi_{\min}^2}\sqrt{T_0}\sqrt{\gamma}\log^2(\gamma)\sum_{i=0}^{M-1}\sqrt{\gamma}^i\cdot i^2 \\
\leq&\hat{\mathcal{O}}(\log(T)) + \tilde{\mathcal{O}}(1)\frac{s(n^2p+p^3)\sigma_{\mathbf{w}}^2}{\pi_{\min}^2}\sqrt{T_0}\sqrt{\gamma}\log^2(\gamma)M\sqrt{\gamma}^M\left(\frac{\sqrt{\gamma}}{\sqrt{\gamma}-1}\right)^3\left(M-\frac{1}{\sqrt{\gamma}}\right) \\
\leq&\hat{\mathcal{O}}(\log(T)) + \tilde{\mathcal{O}}(1)\frac{s(n^2p+p^3)\sigma_{\mathbf{w}}^2}{\pi_{\min}^2}\sqrt{T}\log(\frac{T}{T_0})\left(\frac{\sqrt{\gamma}}{\sqrt{\gamma}-1}\right)^3\left(\sqrt{\gamma}\log(\frac{T}{T_0})-\log(\gamma)\right) \\
\leq&\hat{\mathcal{O}}(\log(T)) + \tilde{\mathcal{O}}\left(\frac{s(n^2p+p^3)\sigma_{\mathbf{w}}^2}{\pi_{\min}^2}\sqrt{T}\log^2(T)\right),
\end{aligned}\tag{218}$$

which shows the regret bound in (216).

Then, we evaluate the occurrence probability of this event. For the probability bounds in Proposition 3, 4, we define $\delta_{sysid;i}$ and $\delta_{\mathbf{x}_0;i}$ such that

$$\mathsf{P}(\mathcal{D}_i \mid \cap_{j=0}^{i-1}\mathcal{E}_j) = \mathsf{P}(\mathcal{D}_i \mid \mathcal{B}_{i-1}\mathcal{D}_{i-1}) > 1-\delta_{\mathbf{x}_0;i}, \quad \mathsf{P}(\mathcal{D}_0) \geq 1-\delta_{\mathbf{x}_0;i} \tag{219}$$

$$\mathsf{P}(\mathcal{C}_i \mid \cap_{j=0}^{i-1}\mathcal{E}_j)) = \mathsf{P}(\mathcal{C}_i \mid \mathcal{B}_{i-1}, \mathcal{D}_{i-1}) \geq 1-\delta_{sysid;i}, \quad \mathsf{P}(\mathcal{C}_0) \geq 1-\delta_{sysid;0}. \tag{220}$$

Then, applying the probability bounds in Proposition 2, 3, 4, and 5, we have

$$
\begin{aligned}
&\mathsf{P}\,(\text{Regret bounds in (216) holds}) \\
\geq&\mathsf{P}(\cap_{i=0}^{M-1}\mathcal{E}_i) \\
=&\mathsf{P}(\mathcal{A}_M, \mathcal{B}_{M-1}, \mathcal{C}_{M-1}, \mathcal{D}_{M-1} \mid \cap_{i=0}^{M-2}\mathcal{E}_i) \cdot \mathsf{P}(\cap_{i=0}^{M-2}\mathcal{E}_i) \\
=&\mathsf{P}(\mathcal{B}_{M-1}, \mathcal{C}_{M-1}, \mathcal{D}_{M-1} \mid \cap_{i=0}^{M-2}\mathcal{E}_i) \cdot \mathsf{P}(\cap_{i=0}^{M-2}\mathcal{E}_i) \\
\geq& (1-\delta_{sysid;M-1}-\delta_{\mathbf{x}_0;M-1}) \cdot \mathsf{P}(\cap_{i=0}^{M-2}\mathcal{E}_i) \\
\geq& \prod_{i=0}^{M-1}(1-\delta_{sysid;i}-\delta_{\mathbf{x}_0;i}) \\
\geq&1-\sum_{i=0}^{M-1}(\delta_{sysid;i}+\delta_{\mathbf{x}_0;i}) \qquad (\delta_{\mathbf{x}_0;i}, \delta_{sysid;i} := \frac{3}{\pi^2}\cdot\frac{\delta}{(i+1)^2}) \\
\geq& 1-\delta.
\end{aligned}
\tag{221}
$$

Finally, we discuss the requirements on tuning parameters $c_{\mathbf{x}}$, $c_{\mathbf{z}}$, C_{sub}, and T_0 such that Proposition 2, 3, 4, and 5 hold for $\delta_{\mathbf{x}_0;i}, \delta_{sysid;i} = \frac{3}{\pi^2}\cdot\frac{\delta}{(i+1)^2}$ for all $i \in [M]$. For data bounds $c_{\mathbf{x}}$ and $c_{\mathbf{z}}$ used in Proposition 3, choosing $c_{\mathbf{x}} \geq \underline{c}_{\mathbf{x}}(\bar{\rho}, \bar{\tau}), c_{\mathbf{z}} \geq \underline{c}_{\mathbf{z}}$ would suffice. Proposition 2 and 3 require $T_i \geq \underline{T}_{rgt}(\frac{\delta_{sysid,i}}{2}) = \mathcal{O}(\text{polylog}(\frac{i+1}{\delta}))$ for all i. This, can be satisfied when picking $T_0 \geq \mathcal{O}(\underline{T}_{rgt}(\delta)) = \mathcal{O}(\text{polylog}(\frac{1}{\delta}))$ since $T_i = \mathcal{O}(T_0\gamma^i)$. Proposition 3 requires $C_{sub} \geq \underline{C}_{sub;rgt}(\bar{x}_0, \frac{\delta_{sysid,i}}{2}, T_i) = \mathcal{O}(1+\frac{\text{polylog}(\frac{i+1}{\delta})}{\log(T_i)}) = \mathcal{O}(1+\frac{\text{polylog}(\frac{i+1}{\delta})}{\log(T_0\gamma^i)})$ for each epoch i. Since this quantity decreases with i and T_0, the requirements can be satisfied by picking $C_{sub} \geq \mathcal{O}(\underline{C}_{sub;rgt}(\bar{x}_0, \delta, 2))$. This concludes the proof. □

Remark 2. *From the proof, one can see if the exploration noise variance is $\sigma_{\mathbf{z};i}^2 = \frac{\sigma_{\mathbf{w}}^2\sqrt{\log(T_i)}}{\sqrt{T_i}}$, then one can slightly improve the regret bound from $\mathcal{O}(\sqrt{T}\log^2(T))$ to $\mathcal{O}(\sqrt{T}\log(T))$.*

C.4. Discussions on Regret Definitions

Bounding random regret with high probability — Note that the definition of Regret_i in (6) for epoch i is a conditional expectation. This expectation is with respect to all the randomness, i.e. $\mathbf{w}_t, \mathbf{z}_t, \omega(t)$, in epoch i, yet still preserves randomness from earlier epochs 0 to $i-1$. Specifically, this earlier randomness affects current Regret_i through the initial state $\mathbf{x}_0^{(i)}$, mode $\omega(0)^{(i)}$, and controller $\mathbf{K}_{1:s}^{(i)}$. This type of expected regret is widely used to evaluate the online control performance (Goel & Hassibi, 2020; Cohen et al., 2018). Further progress is made in (Dean et al., 2018; Lale et al., 2020b), where one is able to bound the regret as a completely random variable with high probability. On the basis of our existing work, to show results for the completely random regrets, one needs to have a tail inequality for the cumulative cost, i.e. $\mathsf{P}\left(|(\sum_t \mathbf{x}_t^\intercal \mathbf{Q}_{\omega(t)}\mathbf{x}+\mathbf{u}_t^\intercal \mathbf{R}_{\omega(t)}\mathbf{u}_t) - \mathsf{E}[\sum_t \mathbf{x}_t^\intercal \mathbf{Q}_{\omega(t)}\mathbf{x}+\mathbf{u}_t^\intercal \mathbf{R}_{\omega(t)}\mathbf{u}_t]| > t\right) \leq e^{-t}$, and then combining this with Theorem 2 would suffice. This tail inequality can be obtained easily for non-switched stable systems (Dean et al., 2018, Lemma D.2). This is because the cumulative cost can be expressed as a quadratic form of all the process noise $\mathbf{w}_t$, which turns out to be sub-exponential. And since its covariance is constructed by the powers of the state matrix of the system, with stability, one can fairly easily show that the covariance is bounded and also obtain proper sub-exponential tail inequalities. However, for MJS, this covariance will depend on the Gramian matrix constructed with closed-loop state matrix $\mathbf{L}_{1:s}$ for all possible switching sequences, i.e. $\mathbf{L}_{i_t}\mathbf{L}_{i_{t-1}}\cdots\mathbf{L}_{i_1}\mathbf{L}_{i_1}^\intercal\cdots\mathbf{L}_{i_{t-1}}^T\mathbf{L}_{i_t}^T$ with $(i_1, i_2, \dots, i_t) \in [s]^t$. With MSS, one can only bound this Gramian matrix in expectation with respect to the switching sequence, but does not preclude the existence of a possible sequence such that this Gramian matrix can be explosive. This, in turn prevents one from using a high probability tail bound. This is also why we restrict our attention to the expected regret requiring only stability in the expected sense, which can be guaranteed with MSS.

We realize with some stronger notions of stability, one may obtain high probability tail bound for MJS, which can be used to bound the completely random regret with high probability. For example, one may assume $\|\mathbf{L}_i\| < 1$ for all $i \in [s]$, this guarantees the state $\mathbf{x}_t$ will decay in a deterministic way, and the MJS behaves no different from a non-switched stable system. A slightly weaker stability that can help is that the joint spectral radius of $\mathbf{L}_{1:s}$ decays exponentially. Similar assumptions are used in (Sarkar et al., 2019) for identification of stochastic systems. Finally, an even weaker and fairly realistic condition is a probabilistic assumption on joint spectral radiuses, rather than enforcing exponential decay on all joint spectral radiuses (i.e. for all switching sequences). Specifically, following our proof strategy, it appears that, if with high probability over the mode switches, the joint spectral radius of the realized trajectory is small, then, our analysis applies and high probability regret guarantees can be achieved. Here, the basic idea is that, intuitively, the fraction of explosive trajectories are exponentially small in the trajectory length.

D. Additional Numerical Experiments and Details

In this section, we provide additional experiments to investigate the efficiency of the proposed algorithms on synthetic datasets. Throughout, we show results from a synthetic experiment where entries of the true system matrices $(\mathbf{A}_{1:s}, \mathbf{B}_{1:s})$ were generated randomly from a standard normal distribution. We scaled each $\mathbf{A}_i$ to have spectral radius equal to 0.3 to obtain a mean square stable MJS. For the cost matrices $(\mathbf{Q}_{1:s}, \mathbf{R}_{1:s})$, we set $\mathbf{Q}_i = \underline{\mathbf{Q}}_i \underline{\mathbf{Q}}_i^\top$, and $\mathbf{R}_i = \underline{\mathbf{R}}_i \underline{\mathbf{R}}_i$ where $\underline{\mathbf{Q}}_i$ and $\underline{\mathbf{R}}_i$ were generated using `randn` function. The Markov matrix $\mathbf{T}$ was sampled from a Dirichlet Process $\texttt{Dir}((s-1)\cdot\mathbf{I}_s+1)$, where $\mathbf{I}_s$ denotes the identity matrix. We assume that we had equal probability of starting in any initial mode.

Since our main contribution is estimating $\mathbf{A}_{1:s}$ and $\mathbf{B}_{1:s}$ of the MJS, we omit the plots for estimating $\mathbf{T}$. Let $\hat{\Psi}_i = [\hat{\mathbf{A}}_i, \hat{\mathbf{B}}_i]$ and $\Psi_i = [\mathbf{A}_i, \mathbf{B}_i]$. Let $\|\hat{\Psi} - \Psi\|/\|\Psi\|$ denotes the maximum of $\|\hat{\Psi}_i - \Psi_i\|/\|\Psi_i\|$ over the modes. We use $\|\hat{\Psi} - \Psi\|/\|\Psi\|$ to investigate the convergence behaviour of MJS-SYSID. We also use the regret function defined in (6) to evaluate the performance of Adaptive MJS-LQR. In all the aforementioned algorithms, the depicted results are averaged over 50 independent replications.

All algorithms have been implemented in the MATLAB R2020a environment and run on a Mac machine equipped with a 1.8 GHz Intel Core i5 processor and 8 GB 1600 MHz DDR3 of memory.

D.1. Performance of MJS-SYSID

Next, we investigate the performance of MJS-SYSID method and compare different initialization of 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) $\sigma_{\mathbf{w}} = 0.1, \sigma_{\mathbf{z}} \in \{0.01, 0.05, 0.1, 0.2\}$ and (ii) $\sigma_{\mathbf{z}} = 0.1, \sigma_{\mathbf{w}} \in \{0.01, 0.05, 0.1, 0.2\}$. The number of states, inputs, and modes are set to $n = 10$, $p = n - 2$, and $s = 10$, respectively.

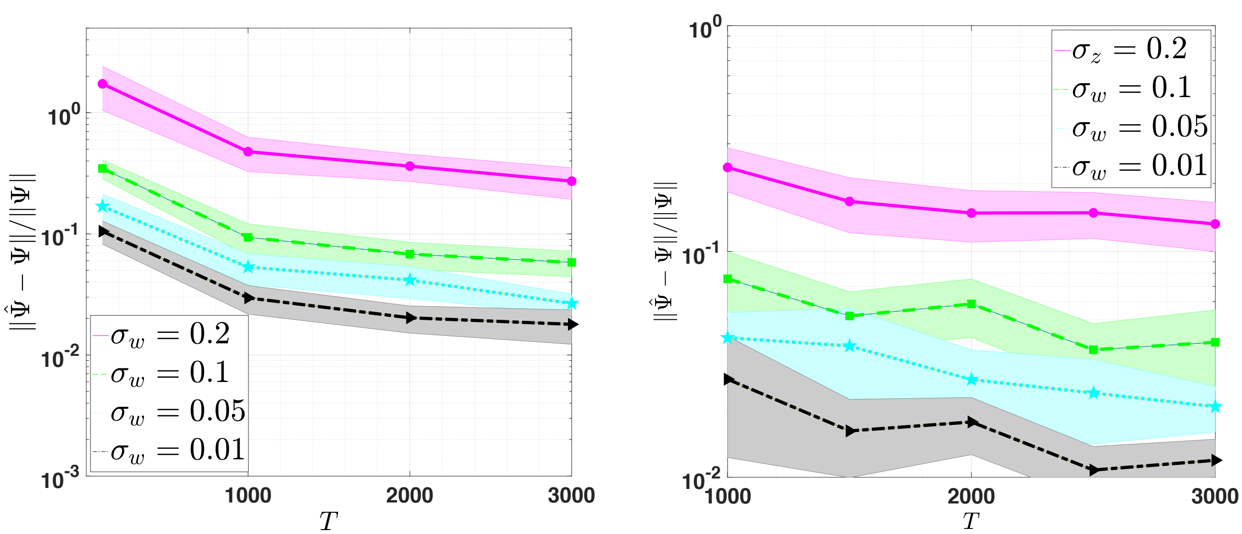

Figure 2. Performance profiles of MJS-SYSID with varying (left) σ_w and (right) σ_z.

Figure 2 demonstrates how the relative estimation error $\|\hat{\Psi} - \Psi\|/\|\Psi\|$ changes as T increases. Each curve on the plot represents a fixed σ_{w} and σ_{z}. These empirical results are all consistent with the theoretical bound of MJS-SYSID given in (5). In particular, the estimation errors degrade with increasing both σ_{w} and σ_{z}.

Now, we fix $\sigma_w = \sigma_z = 0.1$ and investigate the performance of the MJS-SYSID with varying number of states, inputs, and modes. Figure 3 shows how the estimation error $\|\hat{\Psi} - \Psi\|/\|\Psi\|$ changes with (left) $s = 10$, $n \in \{5, 10, 15, 20\}$, $p = n - 2$

and (right) $n = 10$, $p = n - 2$, $s \in \{10, 15, 20, 25\}$. As we can see, the MJS-SYSID has better performance with small n, p and s which is consistent with (5).

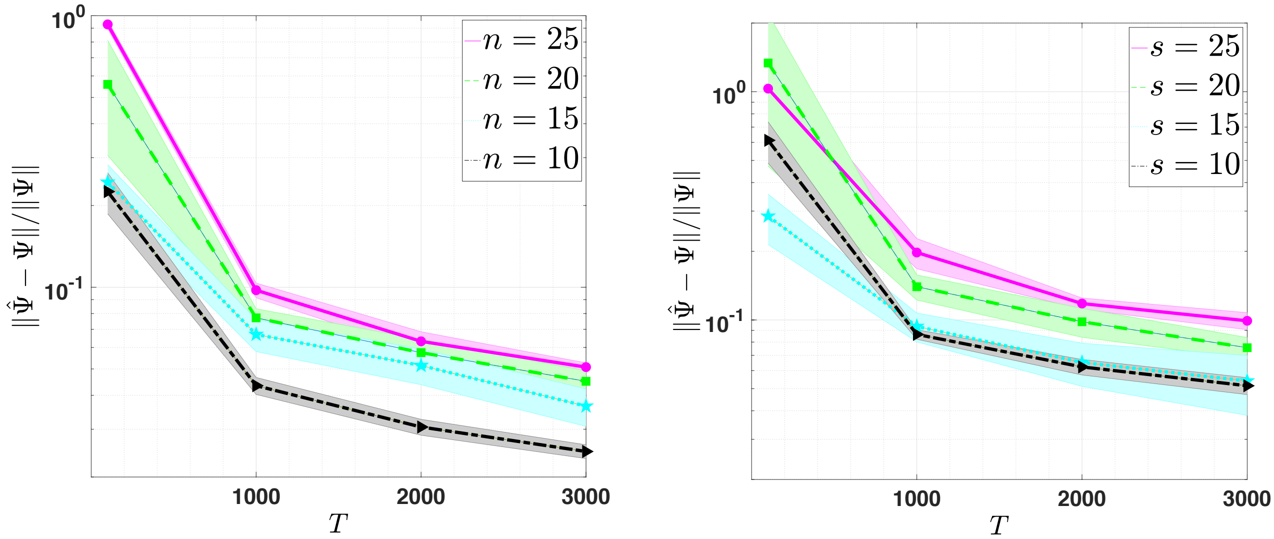

Figure 3. Performance profiles of MJS-SYSID with varying (left) the number of states n and (right) number of modes s.

Our final experiments investigate the system identification bounds in the practically relevant setting where state matrices $\mathbf{A}_{1:s}$ are unknown but input matrices $\mathbf{B}_{1:s}$ are known. In particular, we set $\sigma_z = 0$ an investigate the influence of the number of states n, number of modes s, and σ_w on estimation performance. The relative estimation errors with respect to these parameters are illustrated in Figure 4. One can easily seen that the relative estimation errors improve with partial knowledge. Further, similar to Figures 2 and 3, the estimation error $\|\hat{\Psi} - \Psi\|/\|\Psi\|$ indeed increases with n, s, and σ_w which is consistent with Corollary 1.

D.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 $T_0 = 125$, incremental ratio $\gamma = 2$ and select five epochs to run Algorithm 2. The initial stabilizing controller $\hat{\mathbf{K}}_{1:s}^{(0)}$ is generated randomly from a standard normal distribution. During epoch i, controller $\hat{\mathbf{K}}_{1:s}^{(i)}$ is given by

$$\hat{\mathbf{K}}_j = -\big(\mathbf{R}_j + \hat{\mathbf{B}}_j^\top \varphi_j(\mathbf{P}_{1:s})\hat{\mathbf{B}}_j\big)^{-1}\hat{\mathbf{B}}_j^\top \varphi_j(\mathbf{P}_{1:s})\hat{\mathbf{A}}_j, \tag{222}$$

where $(\hat{\mathbf{A}}_{1:s}, \hat{\mathbf{B}}_{1:s})$ are the output of Algorithm 1 and $\varphi_j(\mathbf{P}_{1:s})$ are approximate solutions of the inexact cDARE given by

$$\mathbf{P}_j = \hat{\mathbf{A}}_j^\top \varphi_j(\mathbf{P}_{1:s})\hat{\mathbf{A}}_j + \mathbf{Q}_j - \hat{\mathbf{A}}_j^\top \varphi_j(\mathbf{P}_{1:s})\hat{\mathbf{B}}_j\big(\mathbf{R}_j + \hat{\mathbf{B}}_j^\top \varphi_j(\mathbf{P}_{1:s})\hat{\mathbf{B}}_j\big)^{-1}\hat{\mathbf{B}}_j^\top \varphi_j(\mathbf{P}_{1:s})\hat{\mathbf{A}}_j \tag{223}$$

for all $j \in [s]$. To solve Equation (223), we set the stopping condition $\|\hat{\mathbf{P}}_{1:s} - \mathbf{P}_{1:s}\| \leq 10^{-6}$ and the maximum number of iteration to 10000.

Figure 5 demonstrates how regret bounds vary with (top left) $\sigma_{\mathbf{w}} \in \{0.01, 0.05, 0.1, 0.2\}$, $n = 10$, $p = n - 2$, $s = 10$, (top right) $\sigma_{\mathbf{w}} = 0.01$, $n = 10$, $p = n - 2$, and $s \in \{5, 10, 15, 20\}$, and (bottom) $\sigma_{\mathbf{w}} = 0.01$, $n \in \{5, 10, 15, 20\}$, $p = n - 2$, $s = 10$. We see that the regret degrades as σ_w, n, and s increase. We also see that when the exploration noise $\sigma_z = \sigma_w$ is large (T is small), the regret become worse quickly as n and s go bigger. These results are consistent with the theoretical bounds in Theorem 2.

Next, we investigate the regret bounds in the practically relevant setting where state matrices $\mathbf{A}_{1:s}$ are unknown but input matrices $\mathbf{B}_{1:s}$ are known. In particular, we determine $\hat{\mathbf{K}}_{1:s}$ and $\mathbf{P}_{1:s}$ by solving (222) and (223), respectively, where $\hat{\mathbf{B}}_{1:s}$ are replaced with their known values $\mathbf{B}_{1:s}$. The numerical results are shown in Figure 6. This figure illustrates that the regret bound indeed increases with n, s, and σ_w. However, the bound is smaller than the unknown setting where one needs to estimate $\hat{\mathbf{B}}$ via Algorithm 1.

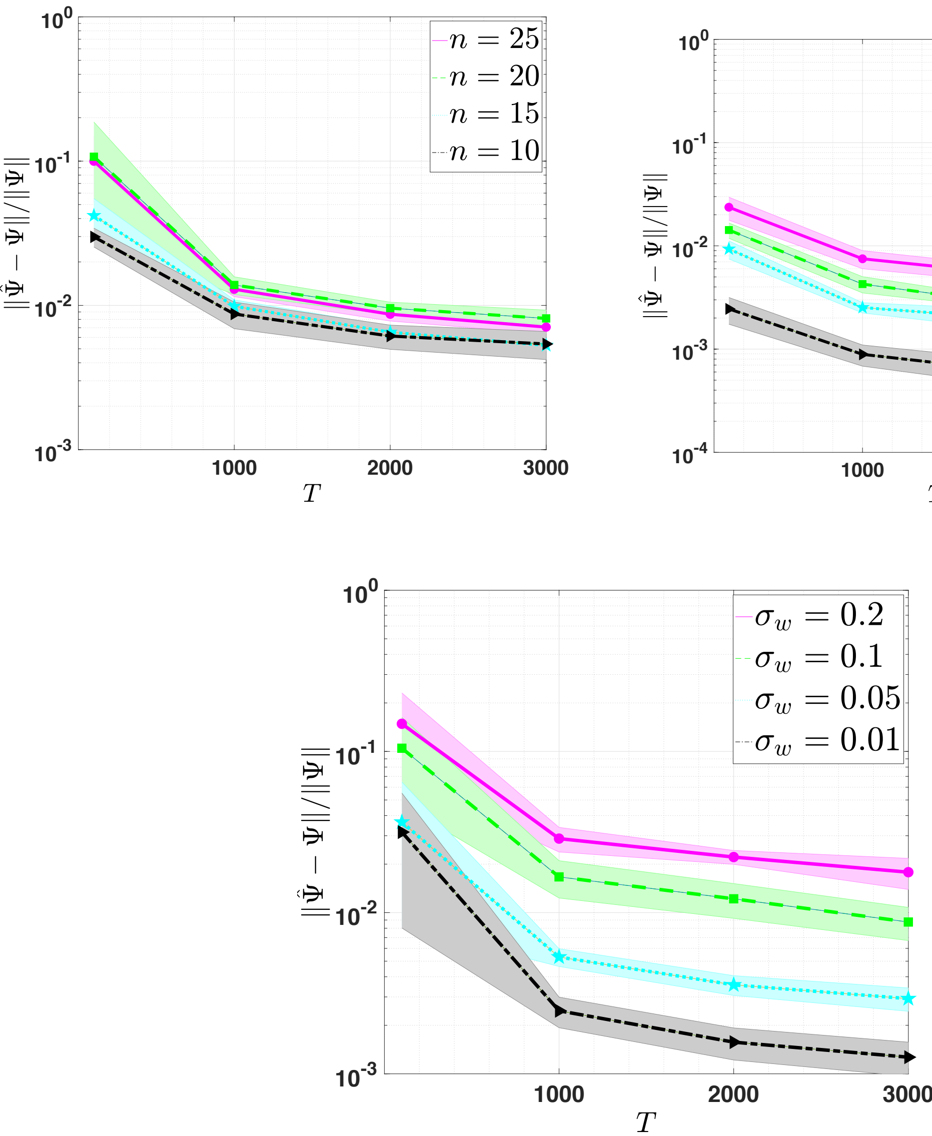

Figure 4. Performance profiles of MJS-SYSID with known $\mathbf{B}$ and varying (top left) the number of states n, (top right) number of modes s, and (bottom) σ_w.

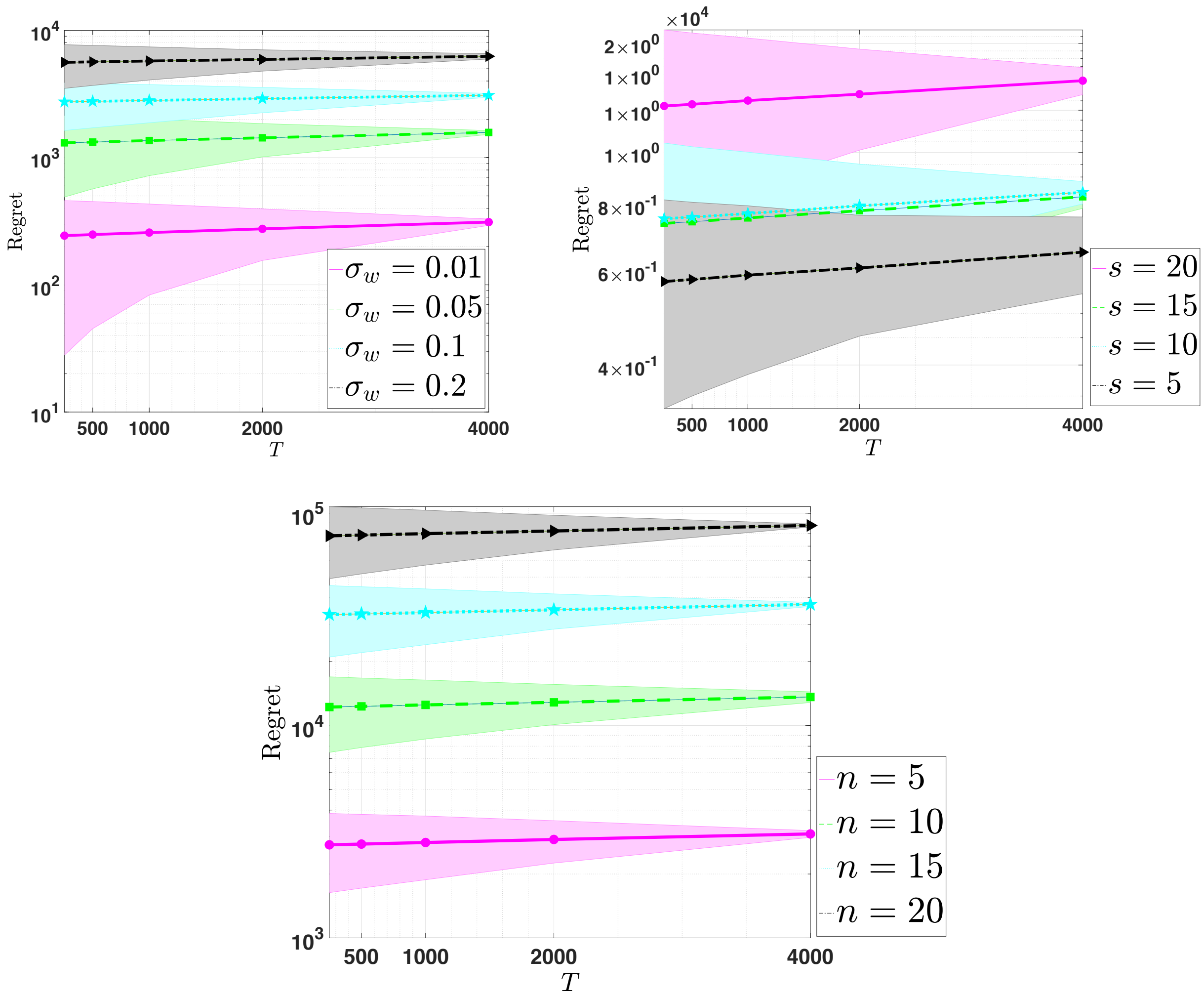

Figure 5. Performance profiles of Adaptive MJS-LQR with varying (top left) process noise σ_w, (top right) number of modes s, and (bottom) number of states n.

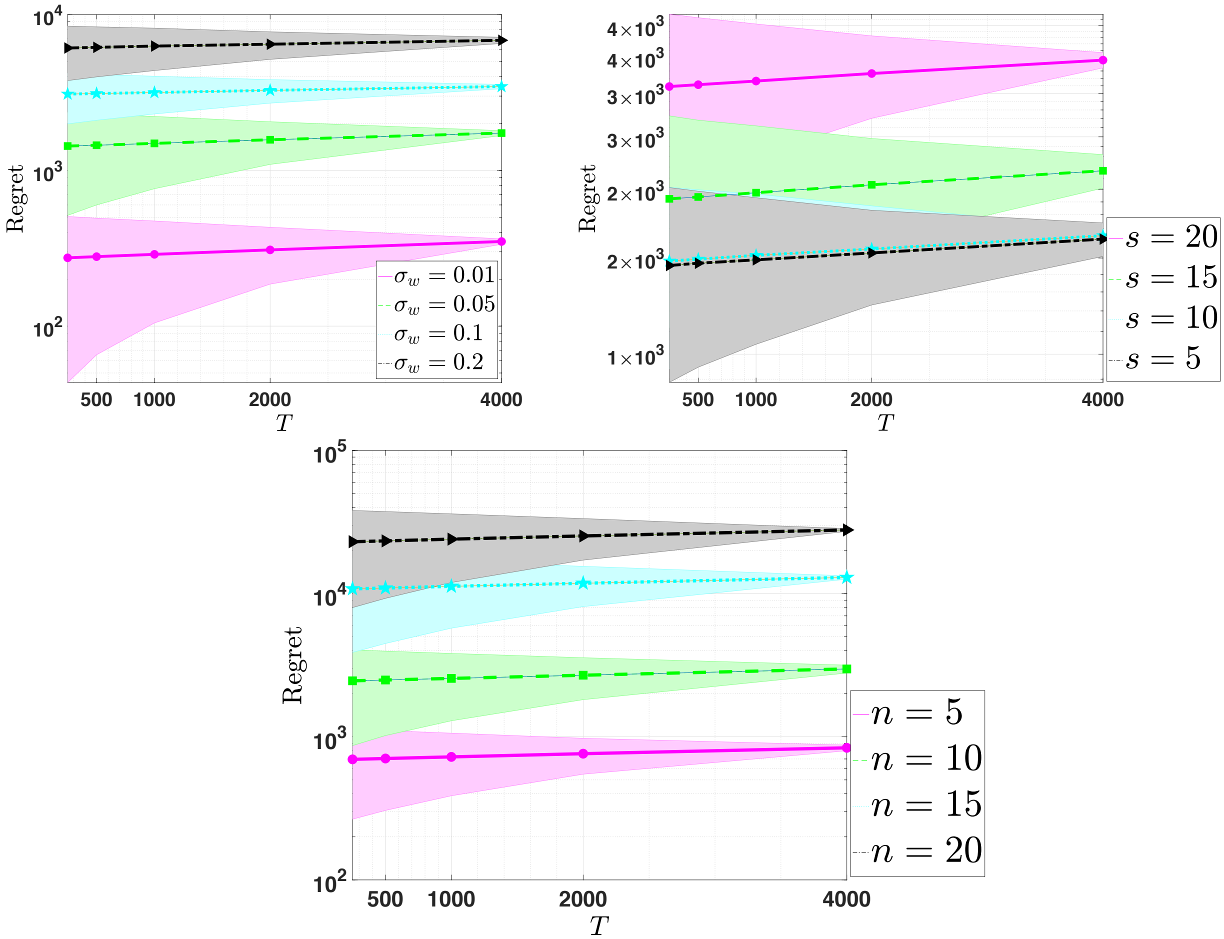

Figure 6. Performance profiles of Adaptive MJS-LQR with known $\mathbf{B}$ and varying (top left) process noise σ_w, (top right) number of modes s, and (bottom) number of states n.