



HAL
open science

Learning stability of partially observed switched linear systems

Zheming Wang, Raphaël Jungers, Mihály Petreczky, Bo Chen, Li Yu

► **To cite this version:**

Zheming Wang, Raphaël Jungers, Mihály Petreczky, Bo Chen, Li Yu. Learning stability of partially observed switched linear systems. 2023. hal-04304091

HAL Id: hal-04304091

<https://hal.science/hal-04304091v1>

Preprint submitted on 24 Nov 2023

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

Learning stability of partially observed switched linear systems

Zheming Wang ^{*1}, Raphaël M. Jungers ^{†2,3}, Mihály Petreczky ^{‡4}, Bo Chen ^{§1}, and Li Yu ^{¶1}

¹The Department of Automation, Zhejiang University of Technology, Hangzhou 310023, China

²The ICTEAM Institute, UCLouvain, Louvain-la-Neuve,1348, Belgium

³The Department of Computer Sciences, University of Oxford, OX1 3QD Oxford, United Kingdom

⁴The Centre de Recherche en Informatique, Signal et Automatique de Lille, UMR CNRS 9189, CNRS; Ecole Centrale Lille, Université de Lille, Villeneuve dAscq 59651, France

Abstract

This paper deals with learning stability of partially observed switched linear systems under arbitrary switching. Such systems are widely used to describe cyber-physical systems which arise by combining physical systems with digital components. In many real-world applications, the internal states cannot be observed directly. It is thus more realistic to conduct system analysis using the outputs of the system. Stability is one of the most frequent requirement for safety and robustness of cyber-physical systems. Existing methods for analyzing stability of switched linear systems often require the knowledge of the parameters and/or all the states of the underlying system. In this paper, we propose an algorithm for deciding stability of switched linear systems under arbitrary switching based purely on observed output data. The proposed algorithm essentially relies on an output-based Lyapunov stability framework and returns an estimate of the joint spectral radius (JSR). We also prove a probably approximately correct error bound on the quality of the estimate of the JSR from the perspective of statistical learning theory.

Keywords— Stability, switched systems, scenario approach, observability

1 Introduction

Verification of safety and robustness of AI systems has gained significant attention in recent years [1, 2, 3]. This is particularly important in the context of application of machine learning algorithms

^{*}wangzheming@zjut.edu.cn

[†]raphael.jungers@uclouvain.be. Raphaël M. Jungers is a FNRS honorary Research Associate. This project has received funding from the European Research Council (ERC) under the European Union’s Horizon 2020 research and innovation programme under grant agreement No 864017 - L2C. Raphaël M. Jungers is also supported by the Innoviris Foundation and the FNRS (Chist-Era Druid-net). He is currently on sabbatical leave at Oxford University, Department of Computer Science, Oxford, UK.

[‡]mihaly.petreczky@centralelille.fr

[§]bchen@zjut.edu.cn

[¶]lyu@zjut.edu.cn

to *cyber-physical systems* (CPS), which tend to be safety-critical (e.g., autonomous vehicles, etc.), see [4, 5]. On the other hand, stability is one of the most basic requirements for safety and robustness of CPS [4, 5, 6]. Due to the interactions between the cyber and physical components, CPSs can be modelled by so called hybrid systems [5, 6, 7], of which switched linear systems (SLS) represent a subclass [8, 9]. It is thus of interest to investigate verification of stability of SLSs.

The topic of stability is a classical one in dynamical systems theory and in control theory [10, 11], and has been studied for SLSs [7, 8, 9]. In the above-cited works, stability analysis techniques typically require a dynamical model of the system. However, obtaining an accurate model of real-world engineering systems by first principles can be challenging in particular for CPSs with hybrid behaviors. From the point of view of conventional control theory, system identification where dynamical models are learned from data could be a reasonable intermediate step for system analysis. In fact, system identification is a well-studied subject with a huge literature, see, e.g., [12, 13] and the references therein. Early work mainly focuses on linear systems and tends to provide asymptotic guarantees see [12] for an overview of the classical literature (including extensions to nonlinear systems). For hybrid systems, there also exists a rich literature on learning SLS models, see [14, 15, 16, 17, 18] and the references therein. Many of existing techniques assume the knowledge of the switching signal [19, 20]. When the switching signal is unavailable, auto-regressive models are often used to approximate SLSs, which can be computationally expensive for an accurate approximation [16]. In fact, it is proved in [21] that learning SLS models is NP-hard without the knowledge of the switching signal. In this paper, we address stability of SLS state-space representations with unobserved switching. Our goal is to decide stability of a SLS directly based only on the observed data generated by it. More precisely, we consider discrete-time SLSs with no inputs under arbitrary switching, and we are interested in deciding if the system is stable, i.e., if the joint spectral radius (JSR) [22] of the system matrices is smaller than 1. Our motivation is to bypass the identification phase, in order to avoid both computational burden and potential modelling errors.

We assume that we can observe random samples of the output of the system. Moreover, we assume that each sampled output trajectory is generated from a certain initial state and switching signal. By output we mean a function of the hidden internal state. We then formulate an optimization problem on the observed outputs, such that the solution of this optimization problem gives an upper bound on the JSR of the underlying system with a certain precision and with a certain probability. That is, the optimization problem gives an upper bound on the JSR with high probability. From the perspective of statistical learning theory [23, 24], our result provides a *probably approximately correct* (PAC) bound, which brings important insights into the relation between the size of the sample and the precision of the solution.

The papers which are the closest to the present one are [25, 26, 27, 28] where the celebrated scenario approach is applied in order to infer stability of the system from a sampled set of observations. They show that, even though it is not obvious that the scenario approach can be used in this context, the geometric properties of SLSs allow to retrieve firm, probabilistic, stability guarantees under a set of mild assumptions. We refer the reader to [29, 30, 31, 32, 33] for good introductions to the scenario approach. In contrast to [25, 26, 27, 28] where fully observed SLSs are considered, this paper considers partially observed SLSs, i.e., we no longer assume that the whole internal state can be observed. The extension from the fully observed case to the partially observed case allows us to consider a wider range of practical systems where the internal state is not directly accessible. However, the challenge is that we need to construct a Lyapunov function from the output trajectory instead of the state trajectory. More specifically, for this extension, we need to develop an output-based Lyapunov stability analysis technique which serves as a basis for the proposed data-driven approach under observability conditions. Similar to [25], we show that the data-based solution converges to the model-based solution as the sample size increases. In addition, we explicitly derive a convergence rate which provides important insights into the relation between the sample size and the precision of the solution. With the obtained convergence rate, we also show that the output-based Lyapunov function is PAC learnable in the sense of Valiant's definition in [34] using the proposed algorithm.

Our work is also closely related with the research on learning safety certificates, such as Lyapunov functions and contraction metrics, which guarantee stability, see [35, 36, 37, 38, 39, 40, 41, 42, 43, 44]. The latter references require the knowledge of the underlying system, except

[37, 36, 42, 44]. Moreover, all these works require the knowledge of the full state and hybrid behaviors are not taken into consideration. Hence, these techniques are not suitable for the stability analysis of partially observed switched systems.

The contribution of this paper is threefold. First, we propose a stability learning approach for SLSs with only partial observation under observability assumptions. Second, we provide a PAC bound on the JSR with an explicit convergence rate, which allows to derive the *sample complexity* of the proposed learning approach. Third, we conduct a comparison with hybrid system identification techniques [14, 15] using numerical examples. More precisely, by the numerical experiments in Section 5, we show that hybrid system identification techniques [14, 15] not only are computationally expensive but also generate modelling errors which leads to false stability inference.

The rest of the paper is organized as follows. In Section 2, we present the formal problem formulation and some preliminary results on stability of SLSs. Section 3 is the main part of the paper, where we present the proposed output-based Lyapunov framework from the stability learning problem to probabilistic guarantees. In Section 4, we discuss some practical issues about the proposed analysis. In Section 5, we provide numerical simulations and comparison with hybrid system identification techniques.

Notation. We denote by \mathbb{R}^+ and \mathbb{Z}^+ the set of all non-negative real numbers and the set of all non-negative integers respectively. For a square matrix Q , $Q \succ (\succeq) 0$ means Q is positive definite (semi-definite). For a symmetric $Q \succ 0$, let $\kappa(Q) := \lambda_{\max}(Q)/\lambda_{\min}(Q)$. Consider the set $\mathcal{M} := \{1, 2, \dots, M\}$ for some given integer $M \in \mathbb{Z}^+$, \mathcal{M}^k denotes the k -Cartesian product of \mathcal{M} for any $k \in \mathbb{Z}^k$. Let $\sigma = (\sigma_0, \sigma_1, \dots, \sigma_{k-1})$ be an element of \mathcal{M}^k . For any $a, b \in \mathbb{Z}^+$ with $b \geq a$, we denote the segment $(\sigma_a, \dots, \sigma_b)$ by $\sigma_{a:b}$. For consistence, let $\sigma_{a:b} = \emptyset$ when $a > b$. For any $p \geq 1$, the p -norm of a vector/matrix x is $\|x\|_p$ ($\|x\|$ is the 2-norm by default), and let $\|x\|_F$ denote the Frobenius norm.

2 Problem statement and preliminaries

Switched linear systems Below we will define the notion of switched linear systems and recall some basic properties of such systems. These properties will allow us to relate stability of the system with the observed behavior.

A discrete-time switched linear system (SLS) is a dynamical system with output of the form

$$x(t+1) = A_{\sigma(t)}x(t), \quad y(t) = C_{\sigma(t)}x(t), \quad t \in \mathbb{Z}^+ \quad (1)$$

where $x(t) \in \mathbb{R}^n$ is the state vector, $y(t) \in \mathbb{R}^p$ is the output and $\sigma : \mathbb{Z}^+ \rightarrow \mathcal{M} := \{1, 2, \dots, M\}$ is a time-dependent switching signal that indicates the current active mode of the system among M possible modes in $\{A_1, A_2, \dots, A_M\}$. We will use the tuple $\Sigma = (n, \{(A_i, C_i) : i \in \mathcal{M}\})$ to denote the switched linear system above.

Intuitively, a SLS is just a collection of linear dynamical systems defined on the same state-space. During the evolution of the SLS, one switches from one linear system to another according to the switching signal. For more details on switched systems see [45, 9].

Informally, we would like to decide stability of (1) based on a finite number of observed output data points. In order to state the problem formally, first we will define below what we mean by stability. Then, we will explain our assumption on the data collection mechanism.

Stability Let us recall some basic stability results on discrete-time SLSs. We begin with the formal definition of asymptotic stability below.

Definition 1 (Asymptotic stability) *The discrete-time SLS (1) is asymptotically stable if for any initial state $x(0) \in \mathbb{R}^n$ and switching signal $\sigma : \mathbb{Z}^+ \rightarrow \mathcal{M}$, $\lim_{t \rightarrow +\infty} x(t) = 0$.*

To characterize this asymptotic stability property, we recall the concept of the joint spectral radius (JSR) [22] of a SLS Σ :

$$\rho(\Sigma) := \lim_{k \rightarrow \infty} \max_{\sigma \in \mathcal{M}^k} \|A_{\sigma}\|^{1/k}. \quad (2)$$

It is well known that the SLS Σ is asymptotically stable under arbitrary switching if and only if $\rho(\Sigma) < 1$, see, e.g., [22]. In general, computing the exact value of the JSR $\rho(\Sigma)$ is a difficult problem. Hence, in practice, we approximate $\rho(\Sigma)$ by computing lower and upper bounds within a Lyapunov framework, where the template of the Lyapunov function $V : \mathbb{R}^n \times \mathcal{M}^k \rightarrow \mathbb{R}$ is specified for some $k \in \mathbb{Z}^+$. More precisely, given a Lyapunov function V above, such that for any sequence $(i_1, \dots, i_k) \in \mathcal{M}^k$, we solve the following Lyapunov inequality for some choice of integers $T > 0$:

$$\begin{aligned} V(A_{i_1} \cdots A_{i_{T-k}} x, (i_{T-k+1} \cdots i_T)) &\leq \\ \gamma^{2(T-k)} V(x, i_1, \dots, i_k), \quad \forall (i_1, \dots, i_T) \in \mathcal{M}^k, \forall x \in \mathbb{R}^n \end{aligned} \quad (3)$$

where $\gamma \in \mathbb{R}^+$ and V is parameterized by some variables. We refer the reader to [22] for some popular Lyapunov templates. In this paper, we consider quadratic Lyapunov functions in the form of $V(x, i_1, \dots, i_k) = x^T \mathcal{P}_{(i_1, \dots, i_k)} x$ for some positive definite matrix $\mathcal{P}_{(i_1, \dots, i_k)} \succ 0$.

Data collection We randomly generate multiple trajectories of (1) where the initial state x_0 is uniform and i.i.d. in the unit sphere \mathbb{S}_{n-1} and the switching signal σ_t is uniform and i.i.d. in \mathcal{M} for any $t \in \mathbb{Z}^+$. Suppose we generate $N \in \mathbb{Z}^+$ trajectories of length $T \in \mathbb{Z}^+$, the sample is denoted as

$$\omega_N := \{(x_0^i, \boldsymbol{\sigma}^i) : i = 1, 2, \dots, N\}. \quad (4)$$

where $\boldsymbol{\sigma}^i = (\sigma_0^i, \sigma_1^i, \dots, \sigma_{T-1}^i) \in \mathcal{M}^T$. For each sampling pair, we measure the output trajectory data

$$y_t^i = C_{\sigma_t^i} x_t^i, x_{t+1}^i = A_{\sigma_t^i} x_t^i, 0 \leq t \leq T-1, \forall i \quad (5)$$

where the subscript t denotes the time instant and the superscript i denotes each trajectory. The whole observed data set is denoted by

$$\mathcal{D}_{obs} := \{y_t^i : 0 \leq t \leq T-1, i = 1, 2, \dots, N\}. \quad (6)$$

Formal problem formulation Suppose System (1) is not known, but the observations \mathcal{D}_{obs} are available and they correspond to the random sample ω_N in (4). Find an estimate $\gamma^*(\omega_N)$ of the JSR $\rho(\Sigma)$. Note that we do not require the information on the switching signal.

3 Learning stability from output data

In this section, we present a data-driven stability analysis approach for SLSs with partial observation. We start with the procedure to estimate the JSR of a SLS in the form of (1). We then present the probabilistic guarantees which describe the quality of that estimate.

3.1 Estimating JSR from data

We propose to estimate the JSR and the corresponding Lyapunov function by solving an optimization problem based on observed data. In order to state the optimization problem, for each trajectory, given the observed data set \mathcal{D}_{obs} and some $k \in \mathbb{Z}^+$, let us define the following time-series data

$$v_k^i = \begin{pmatrix} y_0^i \\ y_1^i \\ \vdots \\ y_{k-1}^i \end{pmatrix}, z_k^i = \begin{pmatrix} y_{T-k}^i \\ y_{T-k+1}^i \\ \vdots \\ y_{T-1}^i \end{pmatrix}, \forall i. \quad (7)$$

For notational convenience, let the time-series data set be denoted by, $\forall k \leq T-1$,

$$\mathcal{D}_k := \{(v_k^i, z_k^i) : i = 1, 2, \dots, N\}. \quad (8)$$

Given the data set \mathcal{D}_k for a sufficiently large $k \in \mathbb{Z}^+$, we estimate the JSR by solving the following scenario (or sampled-based) program

$$\min_{\gamma \geq 0, P} (\gamma, \|P\|_F) \quad (9a)$$

$$\text{s.t. } z^\top P z \leq \gamma^{2(T-k)} v^\top P v, \forall (v, z) \in \mathcal{D}_k, \quad (9b)$$

$$I \preceq P \preceq \bar{\lambda} I. \quad (9c)$$

where the minimization is implemented in the *lexicographic order*¹ of the components of the objective function. Let the solution be denoted by $(\gamma_k^*(\omega_N), P_k^*(\omega_N))$ and suppose that it is unique. *The quantity $\gamma_k^*(\omega_N)$ will be our estimate of the JSR of the unknown system (1).* If $\gamma_k^*(\omega_N)$ is sufficiently smaller than 1, then we can conclude that the underlying system is stable, at least with high probability.

Computational complexity By fixing γ , the constraint (9b) is linear, so that the optimization problem (9) can be solved efficiently using SDP solvers [46] and bisection on γ . The computational complexity of solving it is polynomial in the number of data points (note that we do not need to know the number of modes). In particular, solving (9) is expected to be much less computationally expensive than identifying the underlying SLS, which essentially has a complexity growing exponentially with the number of modes. This is supported by numerical results, see Table 1 in Section 5.

Intuition Intuitively, the solution of (9) results in a Lyapunov function $V(x, (i_1, \dots, i_k))$ which can be rewritten as a quadratic function of the outputs and which satisfies (3) with $\gamma \leq \gamma_k^*(\omega_N)$. More precisely, given any $k \in \mathbb{Z}^+$, consider a switching sequence $\sigma = (\sigma_0, \sigma_1, \dots, \sigma_{T-1}) \in \mathcal{M}^T$. For any initial state x of (1) and any integer $k, \ell \in \mathbb{Z}^+$ with $k + \ell \leq T$, define

$$Y_{\sigma, x, \ell, k} = \begin{pmatrix} y(\ell) \\ y(\ell + 1) \\ \vdots \\ y(\ell + k - 1) \end{pmatrix} \quad (10)$$

where $(y(0), \dots, y(T-1))$ is the output generated by (1) for the switching signal $\sigma(i) = \sigma_i$, $i = 0, \dots, T-1$ and initial state $x(0) = x$. We consider quadratic Lyapunov functions of the output trajectory which can be expressed as

$$V(x, \sigma) = Y_{\sigma, x, 0, k}^\top P Y_{\sigma, x, 0, k} \quad (11)$$

where $P \in \mathbb{R}^{kp \times kp}$ is a positive definite matrix.

It then follows that if V satisfies (3), then the constant γ is an upper bound on the JSR of (1) when k and T satisfy certain conditions, as we will see in Theorem 1. In turn, V satisfies (3), if it satisfies the following inequality: for every $\sigma \in \mathcal{M}^T$

$$Y_{\sigma, x, T-k, k}^\top P Y_{\sigma, x, T-k, k} \leq \gamma^{2(T-k)} Y_{\sigma, x, 0, k}^\top P Y_{\sigma, x, 0, k} \quad (12)$$

for every initial state x . Since any output sequence $(y(0), y(1), \dots, y(T-1))$ arises as $Y_{\sigma, x, 0, T}$ for some initial state x and switching $\sigma \in \mathcal{M}^T$, it follows that (12) is equivalent to requiring that

$$\begin{aligned} & \begin{pmatrix} y(T-k) \\ y(T-k+1) \\ \vdots \\ y(T-1) \end{pmatrix}^\top P \begin{pmatrix} y(T-k) \\ y(T-k+1) \\ \vdots \\ y(T-1) \end{pmatrix} \\ & \leq \gamma^{2(T-k)} \begin{pmatrix} y(0) \\ y(1) \\ \vdots \\ y(k-1) \end{pmatrix}^\top P \begin{pmatrix} y(0) \\ y(1) \\ \vdots \\ y(k-1) \end{pmatrix} \end{aligned} \quad (13)$$

¹The first component comes first: $(\gamma_1, \|P_1\|_F) < (\gamma_2, \|P_2\|_F)$ if $\gamma_1 < \gamma_2$ or else $\gamma_1 = \gamma_2$ and $\|P_1\|_F < \|P_2\|_F$.

holds for any output sequence $(y(0), y(1), \dots, y(T-1))$ and generated by (1). Hence, in order to find a bound of the JSR, we have to solve the optimization problem for some sufficiently large $T \in \mathbb{Z}^+$

$$\begin{aligned} (\gamma_k^o, P_k^o) &:= \arg \min_{\gamma \geq 0, P \succeq I} (\gamma, \|P\|_F) \\ \text{s.t. } P, \gamma &\text{ satisfy (13) for all output trajectories} \\ &(y(0), y(1), \dots, y(T-1)) \end{aligned} \quad (14)$$

The scenario program (9) can be in fact considered as a sampled version of the robust optimization problem (14).

3.2 Pathwise observability

Since our goal is to learn stability from the output trajectory, first we need to recall some technical concepts on observability of switched systems. By incorporating these concepts into Lyapunov stability analysis, we present new results that are needed for developing formal probabilistic stability guarantees in the sequel.

Definition 2 A switching sequence $\sigma \in \mathcal{M}^k$ is said to be observable for (1), if the following implication holds:

$$Y_{\sigma, x, 0, k} = 0 \implies x = 0,$$

where $Y_{\sigma, x, 0, k}$ is as in (10). The smallest $k \in \mathbb{Z}^+$ such that there exists an observable switching sequence $\sigma \in \mathcal{M}^k$ is called the observability index of Σ , denoted by $h(\Sigma)$.

Intuitively, a switching signal is observable, then the state of the system can be reconstructed from the observed outputs. When a SLS Σ has no observable switching signal, we let $h(\Sigma) = \infty$. We also recall a stronger observability condition from [47], called *pathwise observable*.

Definition 3 The SLS (1) is said pathwise observable if there exists $k \in \mathbb{Z}^+$ such that every switching signal $\sigma \in \mathcal{M}^k$ is observable. We refer to the smallest such integer k as the pathwise observability index, denoted by $\mathcal{H}(\Sigma)$.

Intuitively, *pathwise observability* means that the state of the system can be reconstructed from the observed outputs, for any choice of the switching signal. When an SLS Σ is not *pathwise observable*, we say that $\mathcal{H}(\Sigma) = \infty$. Let us also point out that, the condition of *pathwise observable* is decidable as shown in [47, 48] by providing explicit upper bounds on $\mathcal{H}(\Sigma)$.

For any $a, b \in \mathbb{Z}^+$, let us also define $A_{\sigma_{a:b}} := A_{\sigma_b} \cdots A_{\sigma_a}$. By convention, when $a > b$, let $A_{\sigma_{a:b}} = I$. Following [49], for any $\sigma = (\sigma_0, \sigma_1, \dots, \sigma_{k-1}) \in \mathcal{M}^k$ of length $k \in \mathbb{Z}^+$, we define the path-dependent observability matrix

$$\mathcal{O}_\Sigma(\sigma) := \begin{pmatrix} C_{\sigma_0} \\ C_{\sigma_1} A_{\sigma_0} \\ \vdots \\ C_{\sigma_{k-1}} A_{\sigma_{k-2}} \cdots A_{\sigma_1} A_{\sigma_0} \end{pmatrix}. \quad (15)$$

Remark 1 Using the definition of path-dependent observability matrices, a switching sequence $\sigma \in \mathcal{M}^k$ for some $k \in \mathbb{Z}^+$ is said to be observable if $\text{rank}(\mathcal{O}_\Sigma(\sigma)) = n$. Following this, we say that a SLS $\Sigma = (n, \{(A_i, C_i) : i \in \mathcal{M}\})$ is pathwise observable if there exists $k \in \mathbb{Z}^+$ such that $\text{rank}(\mathcal{O}_\Sigma(\sigma)) = n$ for any $\sigma \in \mathcal{M}^k$.

Note that the output trajectory can be rewritten as $Y_{\sigma, x, 0, k} = \mathcal{O}_\Sigma(\sigma)x$, the quadratic Lyapunov function in (11) can be rewritten as $V(x, \sigma) = x^\top \mathcal{O}_\Sigma(\sigma)^\top P \mathcal{O}_\Sigma(\sigma)x$. For notational convenience, we then define the following matrices based on the switching sequence

$$\mathcal{P}_\Sigma^\sigma(P, \ell, k) = \mathcal{O}_\Sigma(\sigma_{\ell:k+\ell-1})^\top P \mathcal{O}_\Sigma(\sigma_{\ell:k+\ell-1}). \quad (16)$$

With the definitions above, we obtain a Lyapunov stability result for SLSs.

Theorem 1 Consider an SLS, denoted by $\Sigma = (n, \{(A_i, C_i) : i \in \mathcal{M}\})$, suppose that Σ admits at least one observable switching signal, and denote $h(\Sigma)$ the observability index as in Definition 2. Assume that there exist $k \in \mathbb{Z}^+$, $\ell \in \mathbb{Z}^+$, $\gamma \geq 0$ and $P \succ 0$ such that $k \geq h(\Sigma)$, and for any $\sigma \in \mathcal{M}^{k+\ell}$,

$$A_{\sigma_{0:\ell-1}}^\top \mathcal{P}_\Sigma^\sigma(P, \ell, k) A_{\sigma_{0:\ell-1}} \preceq \gamma^{2\ell} \mathcal{P}_\Sigma^\sigma(P, 0, k), \quad (17)$$

where $\mathcal{P}_\Sigma^\sigma(P, 0, k)$ and $\mathcal{P}_\Sigma^\sigma(P, \ell, k)$ are given as in (16). Then, $\rho(\Sigma) \leq \gamma$.

Proof: For any $q \in \mathbb{Z}^+$ and $\sigma \in \mathcal{M}^{q\ell+k}$, from (17), it holds that

$$\begin{aligned} & A_{\sigma_{(q-1)\ell:q\ell-1}}^\top \mathcal{P}_\Sigma^\sigma(P, q\ell, k) A_{\sigma_{(q-1)\ell:q\ell-1}} \\ & \preceq \gamma^{2\ell} \mathcal{P}_\Sigma^\sigma(P, (q-1)\ell, k). \end{aligned}$$

This implies that

$$\begin{aligned} & A_{\sigma_{0:q\ell-1}}^\top \mathcal{P}_\Sigma^\sigma(P, q\ell, k) A_{\sigma_{0:q\ell-1}} \\ & \preceq \gamma^{2\ell} A_{\sigma_{0:(q-1)\ell-1}}^\top \mathcal{P}_\Sigma^\sigma(P, (q-1)\ell, k) A_{\sigma_{0:(q-1)\ell-1}} \\ & \quad \vdots \\ & \preceq \gamma^{2q\ell} \mathcal{P}_\Sigma^\sigma(P, 0, k), \quad \forall \sigma \in \mathcal{M}^{q\ell+k}, q \in \mathbb{Z}^+. \end{aligned} \quad (18)$$

Since $k \geq h(\Sigma)$, there exists $\bar{\sigma} \in \mathcal{M}^k$ such that $\text{rank}(\mathcal{O}_\Sigma(\bar{\sigma})) = n$ from Definition 2. Thus, there exist $\bar{c} \geq \underline{c} > 0$ such that $\mathcal{O}_\Sigma(\bar{\sigma})^\top \mathcal{O}_\Sigma(\bar{\sigma}) \succeq \underline{c}I$ and $\mathcal{O}_\Sigma(\sigma)^\top \mathcal{O}_\Sigma(\sigma) \preceq \bar{c}I$ for any $\sigma \in \mathcal{M}^k$. With this, we have that, for any $\sigma \in \mathcal{M}^{q\ell+k}$ with $\sigma_{q\ell:k+q\ell-1} = \bar{\sigma}$,

$$\begin{aligned} A_{\sigma_{0:q\ell-1}}^\top \mathcal{P}_\Sigma^\sigma(P, q\ell, k) A_{\sigma_{0:q\ell-1}} & \succeq \lambda_{\min}(P) \underline{c} A_{\sigma_{0:q\ell-1}}^\top A_{\sigma_{0:q\ell-1}}, \\ \mathcal{P}_\Sigma^\sigma(P, 0, k) & \preceq \bar{c} \lambda_{\max}(P) I. \end{aligned}$$

The two inequalities above, together with (18), imply that

$$\lambda_{\min}(P) \underline{c} A_{\sigma_{0:q\ell-1}}^\top A_{\sigma_{0:q\ell-1}} \preceq \gamma^{2q\ell} \bar{c} \lambda_{\max}(P) I$$

which means that

$$\|A_\sigma\| \leq \gamma^{q\ell} \sqrt{\kappa(P) \frac{\bar{c}}{\underline{c}}}, \quad \forall \sigma \in \mathcal{M}^{q\ell} \quad (19)$$

Hence,

$$\rho(\Sigma) = \lim_{q \rightarrow \infty} \max_{\sigma \in \mathcal{M}^{q\ell}} \|A_\sigma\|^{\frac{1}{q\ell}} \leq \gamma \lim_{q \rightarrow \infty} \left(\kappa(P) \frac{\bar{c}}{\underline{c}} \right)^{\frac{1}{2q\ell}} = \gamma.$$

□

The stability result in Theorem 1 serves as a basis for the rest of the paper. Observe that the constraint (17) is not amenable as it is for a data-driven approach, as it explicitly uses the system matrices. In order to leverage it in a data-driven framework, we now provide an interpretation of this stability condition from the perspective of robust optimization. Suppose the overall length of the output trajectory is T . For any $k \in \mathbb{Z}^+$, we formulate the following robust optimization problem

$$(\gamma_k^o, P_k^o) := \arg \min_{\gamma \geq 0, P \succeq I} (\gamma, \|P\|_F) \quad (20a)$$

$$\begin{aligned} \text{s.t. } & x^\top A_{\sigma_{0:\ell-1}}^\top \mathcal{P}_\Sigma^\sigma(P, T-k, k) A_{\sigma_{0:\ell-1}} x \\ & \leq \gamma^{2(T-k)} x^\top \mathcal{P}_\Sigma^\sigma(P, 0, k) x, \forall (x, \sigma) \in \mathbb{S}_{n-1} \times \mathcal{M}^T \end{aligned} \quad (20b)$$

where $\mathcal{P}_\Sigma^\sigma(P, 0, k)$ and $\mathcal{P}_\Sigma^\sigma(P, \ell, k)$ are given as in (16). By homogeneity, (20b) is equivalent to (17). From this robust optimization formulation, the stability condition in Theorem 1 means that for any initial state $x(0)$ the aggregated quadratic Lyapunov function in (11) is decreasing at a rate of γ . The problem (20) can be also seen as a reformulation of (14)

If all the dynamics matrices are nonsingular, Theorem 1 in fact implies the *pathwise observability* property, as stated in the following corollary.

Corollary 1 Suppose the conditions in Theorem 1 hold. Suppose the matrices $\{A_i\}_{i \in \mathcal{M}}$, $\ell \geq k$, and $\gamma > 0$, the stability condition (17) also implies that $\text{rank}(\mathcal{O}_\Sigma(\boldsymbol{\sigma})) = n$ for any $\boldsymbol{\sigma} \in \mathcal{M}^k$.

Proof: Since $\ell \geq k$, there is no overlapping between $\boldsymbol{\sigma}_{0:k-1}$ and $\boldsymbol{\sigma}_{l:k+l-1}$. Again, let $\tilde{\boldsymbol{\sigma}} \in \mathcal{M}^k$ be such that $\text{rank}(\mathcal{O}_\Sigma(\tilde{\boldsymbol{\sigma}})) = n$. With the fact that A_i is invertible for any $i \in \mathcal{M}$ (as Σ is reversible), the condition (17) implies that, for any $\boldsymbol{\sigma} \in \mathcal{M}^{k+\ell}$ with $\boldsymbol{\sigma}_{l:k+l-1} = \tilde{\boldsymbol{\sigma}}$,

$$\mathcal{P}_\Sigma^\sigma(P, 0, k) \succeq \gamma^{-2\ell} A_{\boldsymbol{\sigma}_{0:\ell-1}}^\top \mathcal{P}_\Sigma^\sigma(P, \ell, k) A_{\boldsymbol{\sigma}_{0:\ell-1}} \succ 0. \quad (21)$$

This means that $\mathcal{O}_\Sigma(\boldsymbol{\sigma})$ is full-column rank for any $\boldsymbol{\sigma} \in \mathcal{M}^k$. \square

Remark 2 From Corollary 1, we can see that pathwise observability is not a conservative condition in our output-based Lyapunov framework in which the switching signal is not available.

3.3 Almost Lyapunov stability

Before we prove our main result, we introduce the concept of almost Lyapunov stability which allows to describe the stability with unseen regions in the state space. More precisely, we aim to derive formal guarantees on the JSR from an almost stability condition in which the aggregated Lyapunov functions based on the matrices as defined in (16) are non-increasing everywhere except on a small subset whose measure is bounded by $\varepsilon \in (0, 1)$.

To formally state the stability result with an almost Lyapunov function, the following definition is also needed. Given any $\varepsilon \in (0, 1)$, let

$$\delta(\varepsilon) := \begin{cases} \sqrt{1 - \mathcal{I}^{-1}(2\varepsilon; \frac{n-1}{2}, \frac{1}{2})} & \varepsilon \in [0, \frac{1}{2}) \\ 0 & \varepsilon \geq \frac{1}{2} \end{cases} \quad (22)$$

where $\mathcal{I}(x; a, b)$ the regularized incomplete beta function defined as

$$\mathcal{I}(x; a, b) := \frac{\int_0^x t^{a-1} (1-t)^{b-1} dt}{\int_0^1 t^{a-1} (1-t)^{b-1} dt}. \quad (23)$$

A geometric interpretation of the function $\delta(\cdot)$ is given in Figure 1.

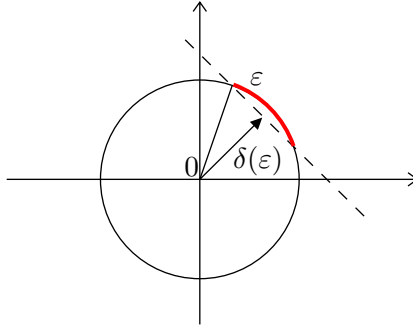


Figure 1: Illustration of $\delta(\varepsilon)$: ε is the uniform (probability) measure of the spherical cap in red and $\delta(\varepsilon)$ is the distance to the base of the spherical cap.

With this definition, we derive the following stability result under the Lyapunov condition (17) with a violating subset.

Theorem 2 Consider a pathwise observable SLS Σ as in (1). Suppose there exist $k \in \mathbb{Z}^+$, $\ell \in \mathbb{Z}^+$, $\gamma \geq 0$, $P \succ 0$ and a subset $S \subseteq \mathbb{S}_{n-1}$ such that, $k \geq \mathcal{H}(\Sigma)$, and for any $\boldsymbol{\sigma} \in \mathcal{M}^{k+\ell}$,

$$\begin{aligned} x^\top A_{\boldsymbol{\sigma}_{0:\ell-1}}^\top \mathcal{P}_\Sigma^\sigma(P, \ell, k) A_{\boldsymbol{\sigma}_{0:\ell-1}} x \leq \gamma^{2\ell} x^\top \mathcal{P}_\Sigma^\sigma(P, 0, k) x, \\ \forall x \in \mathbb{S}_{n-1} \setminus S. \end{aligned} \quad (24)$$

where $\mathcal{P}_\Sigma^\sigma(P, 0, k)$ and $\mathcal{P}_\Sigma^\sigma(P, \ell, k)$ are given as in (16). Then,

$$\rho(\Sigma) \leq \frac{\gamma}{\sqrt[\ell]{\delta\left(\frac{\mu(S)\chi_\Sigma(P, k)}{2}\right)}} \quad (25)$$

where $\mu(\cdot)$ is the uniform probability measure on the unit sphere \mathbb{S}_{n-1} , $\delta(\cdot)$ is defined in (22) and

$$\chi_\Sigma(P, k) := \max_{\sigma \in \mathcal{M}^k} \sqrt{\frac{\det(\mathcal{P}_\Sigma^\sigma(P, 0, k))}{\lambda_{\min}(\mathcal{P}_\Sigma^\sigma(P, 0, k))^n}}. \quad (26)$$

The proof of Theorem 2 is given in the appendix. From the definition of $\chi_\Sigma(P, k)$ in (26), *pathwise observability* is crucial for the stability result in Theorem 2 in the sense that it guarantees the boundedness of $\chi_\Sigma(P, k)$. From Theorem 2, more explicit stability guarantees can be also derived, as shown in the following corollary.

Corollary 2 *Given the same conditions as in Theorem 2, it holds that $\chi_\Sigma(P, k) \leq \sqrt{(c_k \kappa(P))^{n-1}}$ and*

$$\rho(\Sigma) \leq \frac{\gamma}{\sqrt[\ell]{\delta\left(\frac{\mu(S)\sqrt{(c_k \kappa(P))^{n-1}}}{2}\right)}}, \quad (27)$$

where

$$\kappa(P) = \frac{\lambda_{\max}(P)}{\lambda_{\min}(P)} \text{ and } c_k := \max_{\sigma \in \mathcal{M}^k} \kappa(\mathcal{O}_\Sigma(\sigma)^\top \mathcal{O}_\Sigma(\sigma))$$

Moreover, when $I \preceq P \preceq \bar{\lambda}I$ for some $\bar{\lambda} \geq 1$,

$$\rho(\Sigma) \leq \frac{\gamma}{\sqrt[\ell]{\delta\left(\frac{\mu(S)\sqrt{(c_k \bar{\lambda})^{n-1}}}{2}\right)}}. \quad (28)$$

Proof: By definition, since Σ is pathwise observable with $\mathcal{H}(\Sigma) \leq k$, there exists $r > 0$ such that $rI \preceq \mathcal{O}_\Sigma(\sigma)^\top \mathcal{O}_\Sigma(\sigma) \preceq rc_k I$ for any $\sigma \in \mathcal{M}^k$, which leads to the following inequalities:

$$\begin{aligned} \lambda_{\max}(\mathcal{P}_\Sigma^\sigma(P, 0, k)) &\leq \lambda_{\max}(\mathcal{O}_\Sigma(\sigma)^\top \mathcal{O}_\Sigma(\sigma)) \lambda_{\max}(P) \\ &\leq rc_k \lambda_{\max}(P), \\ \lambda_{\min}(\mathcal{P}_\Sigma^\sigma(P, 0, k)) &\geq \lambda_{\min}(\mathcal{O}_\Sigma(\sigma)^\top \mathcal{O}_\Sigma(\sigma)) \lambda_{\min}(P) \\ &\geq r \lambda_{\min}(P). \end{aligned}$$

Using these inequalities, we arrive at

$$\begin{aligned} \sqrt{\frac{\det(\mathcal{P}_\Sigma^\sigma(P, 0, k))}{\lambda_{\min}(\mathcal{P}_\Sigma^\sigma(P, 0, k))^n}} &\leq \sqrt{\frac{\lambda_{\max}(\mathcal{P}_\Sigma^\sigma(P, 0, k))^{n-1}}{\lambda_{\min}(\mathcal{P}_\Sigma^\sigma(P, 0, k))^{n-1}}} \\ &\leq \sqrt{(c_k \kappa(P))^{n-1}}, \forall \sigma \in \mathcal{M}^k. \end{aligned}$$

Hence, it holds that $\chi_\Sigma(P, k) \leq \sqrt{(c_k \kappa(P))^{n-1}}$. Putting this inequality into (25), we obtain (27). Note that the function $\delta(\cdot)$ is decreasing. When $\kappa(P) \leq \bar{\lambda}$, (27) implies (28). \square

With the results above, we are able to extend the technique in [25] which only considers the fully observed case to the partially observed case.

3.4 Main result: Probabilistic stability certificates

In the rest of this section, we formally present our main result, that is, stability certificates based on the solution of the scenario program (9) in a probabilistic sense.

Our derivation relies on the scenario approach (also known as scenario optimization) [29, 30, 32]. In order to use [32], we need to fulfill a non-degeneracy assumption, formalized below.

Definition 4 For the SLS (1) and $k, T \in \mathbb{Z}^+$, a switching sequence $\sigma \in \mathcal{M}^T$ is degenerate if there exist $P \succ 0$ and $\gamma \geq 0$ such that for any initial state x ,

$$Y_{\sigma, x, T-k, k}^\top P Y_{\sigma, x, T-k, k} = \gamma^{2(T-k)} Y_{\sigma, x, 0, k}^\top P Y_{\sigma, x, 0, k}$$

where $Y_{\sigma, x, 0, k}, Y_{\sigma, x, T-k, k}$ are as in (10).

With the discussions and definitions above, we are now ready to present the main result of this section.

Theorem 3 (Main result) Consider the SLS Σ as given in (1), where the initial state is i.i.d with the uniform distribution over the unit sphere \mathbb{S}_{n-1} and the switching signal is uniform and i.i.d. in \mathcal{M} . Given $N, T \in \mathbb{Z}^+$, the sample set ω_N as defined in (4), and $k \leq T - 1$, let $(\gamma_k^*(\omega_N), P_k^*(\omega_N))$ be the unique solution of the scenario program (9). Assume that Σ is path-wise observable with $\mathcal{H}(\Sigma) \leq k$ and that none of the switching sequences in \mathcal{M}^T is degenerate in the sense of Definition 4. For any $\epsilon \in (0, 1)$, with probability no smaller than $1 - \phi(\epsilon; d, N)$,

$$\rho(\Sigma) \leq \frac{\gamma^*(\omega_N)}{\sqrt[T-k]{\delta\left(\frac{\epsilon M^T \chi_\Sigma(P^*(\omega_N), k)}{2}\right)}} \quad (29)$$

where $d = \frac{kp(kp+1)}{2}$, and $\chi_\Sigma(\cdot, k)$ is given in (26), $\delta(\cdot)$ is defined as in (22), and

$$\phi(\epsilon; d, N) := 1 - \mathcal{I}(\epsilon; d, N - d + 1). \quad (30)$$

The proof of Theorem 3 is presented in the appendix. Let us emphasize that the assumption that none of the sequences is degenerate in Theorem 3 is not conservative in practice. In fact, this assumption is a particular case of Assumption 2 in [32], see Section 3.4 in [32] for discussions on relaxing such an assumption. In the fully observed case, this assumption actually means that none of the matrices $\{A_i\}_{i=1}^M$ is similar to a Barabanov matrix, which is diagonalizable with all the eigenvalues having the same modulus, see [26]. Let us also point out that the function $\chi_\Sigma(\cdot, k)$ as defined in (26) implicitly depends on the parameters of the underlying system Σ . In Subsection 4.3 we present an algorithm for estimating $\chi_\Sigma(\cdot, k)$.

The error bound of Theorem 3 implies that the true JSR is smaller than the sample-based solution $\gamma^*(\omega_N)$ multiplied with the correction factor

$$\frac{1}{\sqrt[T-k]{\delta\left(\frac{\epsilon M^T \chi_\Sigma(P^*(\omega_N), k)}{2}\right)}}$$

with a high probability $1 - \phi(\epsilon; d, N)$. The probability $1 - \phi(\epsilon; d, N)$ converges to 1 as $N \rightarrow \infty$. That is, the more data points we have, the more certain we are that (29) holds. The factor depends on the accuracy level ϵ , and it tends to 1 as $\epsilon \rightarrow 0$. We will explicitly discuss the convergence rate in Section 4.1.

The correction factor above is *a posteriori*, in the sense that its numerical value depends on the outcome of the scenario program (9). For this reason it may be difficult to evaluate the quality of this factor. Below we present a modified error bound where the factor with which $\gamma^*(\omega_N)$ is multiplied does not depend on data. This then allows to bound that factor using prior knowledge on the set of possible models. This is done by exploiting the constraint $I \preceq P \preceq \lambda I$ in (9).

Corollary 3 *Suppose that the conditions in Theorem 3 hold. There exists a constant $c \geq 1$, for any $\beta \in (0, 1)$, with probability no smaller than $1 - \beta$,*

$$\rho(\Sigma) \leq \frac{\gamma_k^*(\omega_N)}{\tau^{-k} \sqrt{\delta \left(\frac{\epsilon M^T \sqrt{(c\bar{\lambda})^{n-1}}}{2} \right)}} \quad (31)$$

with $\epsilon = \mathcal{I}^{-1}(1 - \beta; d, N - d + 1)$ and $d = \frac{kp(kp+1)}{2}$.

Proof: This is a direct consequence of Theorem 3 and Corollary 2 with $c = c_k$. \square

The constant c in (31) depends on the underlying system as shown in Corollary 2. Corollary 3 enables us to know *a priori* how much data is needed to reach certain accuracy and confidence level. That is, given a fixed confidence level β and a fixed ϵ , we are able to compute N such that (31) holds with probability $1 - \beta$. Note that this requires the knowledge of the constant c , which depends on the matrices of the underlying system. Suppose that the underlying unknown system comes from a certain family of SLSs, we can choose c to be an upper bound on the constant $\max_{\sigma \in \mathcal{M}^k} \kappa(\mathcal{O}_\Sigma(\sigma)^\top \mathcal{O}_\Sigma(\sigma))$ of each Σ where Σ varies through all possible SLSs. The constant c can be hence considered as a counterpart of VC dimensions in PAC-style error bounds [23], and it captures the complexity of the model class. Alternatively, c could also be estimated from data with high probability, as we will see in Theorem 4 in Section 4.

Remark 3 *In fact, our analysis in Section 3.4 can be extended to the case when the initial state is not uniformly distributed, but its distribution \mathbb{P} satisfies the following regularity condition: There exists an increasing function $\nu : [0, 1] \rightarrow [0, 1]$ such that, for any $S \subseteq \mathbb{S}_{n-1}$, $\mathbb{P}\{x : x/\|x\| \in S\} \geq \nu(\mu(S))$, where $\mu(\cdot)$ is the uniform probability measure on the unit sphere \mathbb{S}_{n-1} . This condition is satisfied for example for Gaussian distributions. Under this condition, the inequality in (29) becomes*

$$\rho(\Sigma) \leq \frac{\gamma^*(\omega_N)}{\tau^{-k} \sqrt{\delta \left(\frac{\nu^{-1}(\epsilon M^T) \chi_\Sigma(P^*(\omega_N), k)}{2} \right)}}. \quad (32)$$

Remark 4 *The proposed approach can be also extended to disturbed systems with bounded additive disturbances in the form of*

$$x(t+1) = A_{\sigma(t)}x(t) + w(t), \quad y(t) = C_{\sigma(t)}x(t) + v(t) \quad (33)$$

where $w(t)$ and $v(t)$ are bounded. Similar probabilistic guarantees can be derived with *a-priori* information on the bounds of the disturbances.

4 Technical issues

In this section, we discuss some technical issues of the proposed data-driven stability analysis approach.

4.1 Sample complexity

We first provide some insights into the relation between the size of the sample and the precision of the solution of the scenario program (9) from the perspective of PAC learning [24]. Let us formally define the *sample complexity* of the scenario program (9) below.

Definition 5 (Sample complexity) *Given $\beta \in (0, 1)$, $\varepsilon \geq 0$ and any integer $k \geq \mathcal{H}(\Sigma)$, the sample complexity of the scenario program (9), denoted by $N_s(\varepsilon, \beta)$, is the minimal $N \in \mathbb{Z}^+$ such that*

$$\mathbb{P}^N \{ \omega_N : \gamma_k^o \leq (1 + \varepsilon) \gamma_k^*(\omega_N) \} \geq 1 - \beta, \quad (34)$$

where γ_k^o is given in (14).

In the following proposition, we present some results on the *sample complexity* for the proposed data-driven approach.

Proposition 1 *Suppose the conditions in Theorem 3 hold. For any $\beta \in (0, 1)$ and $\varepsilon \geq 0$, let the sample complexity $N_s(\varepsilon, \beta)$ be defined as in Definition 5. Then, the following results hold:*

(i) *There exists a constant $c \geq 1$ such that*

$$N_s(\varepsilon, \beta) \leq \mathcal{N}_\phi\left(\frac{\mathcal{I}\left(1 - \frac{1}{(1+\varepsilon)^{2(T-k)}}, \frac{n-1}{2}, \frac{1}{2}\right)}{M^T \chi(Q) \sqrt{(c\lambda)^{n-1}}}, \beta\right) \quad (35)$$

$$= O\left(\frac{1}{\beta} \left(\frac{1}{\varepsilon}\right)^{\frac{n}{2}}\right). \quad (36)$$

where $\mathcal{N}_\phi(\varepsilon, \beta)$ denotes the minimal $N \in \mathbb{Z}^+$ satisfying $\phi(\varepsilon; d, N) \leq \beta$ for any $\varepsilon \in (0, 1)$.

(ii) *When the system Σ is not asymptotically stable under arbitrary switching, i.e., $\rho(\Sigma) > 1$, for any $N \geq N_s(\varepsilon, \beta)$,*

$$\mathbb{P}^N \{\omega_N : \gamma^*(\omega_N) < \frac{1}{1+\varepsilon}\} \leq \beta. \quad (37)$$

The proof is given in the appendix. From Proposition 1, the *sample complexity* is polynomial in $1/\varepsilon$ and $1/\beta$, which implies PAC learnability in the sense of Valiant's definition in [34]. In addition, Property (ii) of Proposition 1 provides a measure of the risk of stability learning using the proposed data-driven approach.

Remark 5 *Using tighter explicit bounds of $\phi(\varepsilon; d, N)$ in [32, 50], it can be even shown that $N_s(\varepsilon, \beta) = O\left(\ln\left(\frac{1}{\beta}\right)\left(\frac{1}{\varepsilon}\right)^{\frac{n}{2}}\right)$.*

Alternatively, we can also define sample complexity based on the absolute error, i.e., the minimal $N \in \mathbb{Z}^+$ such that

$$\mathbb{P}^N \{\omega_N : \gamma^o \leq \gamma^*(\omega_N) + \varepsilon\} \geq 1 - \beta. \quad (38)$$

The analysis above can be adapted to this alternative definition. To avoid repetition, we do not intend to provide the details.

4.2 Pathwise observability index estimation

The results in the previous section all rely on the condition that $k \geq \mathcal{H}(\Sigma)$. However, the exact value of $\mathcal{H}(\Sigma)$ is often unavailable. Though explicit bounds on $\mathcal{H}(\Sigma)$ exist in [47], they are only useful for small values of n and M .

We now show an efficient procedure to estimate $\mathcal{H}(\Sigma)$. Let us define, $\forall k \geq 1$,

$$\xi_k^o := \inf_{\xi \geq 0} \xi \quad (39a)$$

$$\begin{aligned} \text{s.t. } & Y_{\sigma, x, k, k}^\top Y_{\sigma, x, k, k} \leq \xi^{2k} Y_{\sigma, x, 0, k}^\top Y_{\sigma, x, 0, k}, \\ & \forall (x, \sigma) \in \mathbb{S}_{n-1} \times \mathcal{M}^{2k} \end{aligned} \quad (39b)$$

where $Y_{\sigma, x, k, k}$ and $Y_{\sigma, x, 0, k}$ are given in (10). The following proposition shows an important property of the sequence $\{\xi_k^o\}_{k \geq 1}$.

Proposition 2 *Suppose the SLS Σ is pathwise observable. Let $\{\xi_k^o\}_{k \geq 1}$ be defined as in (39). Then, $\xi_k^o < \infty$ if and only if $k \geq \mathcal{H}(\Sigma)$.*

Given a sample set ω_N with a sufficiently long horizon T and the observed trajectories $\{y_t^i : 0 \leq t \leq T-1, 1 \leq i \leq N\}$, empirical estimates of $\{\xi_k^o : 1 \leq k \leq \lfloor \frac{T}{2} \rfloor\}$ are given as

$$\xi_k(\omega_N) := \max_{1 \leq i \leq N} \frac{\|\tilde{v}_k^i\|}{\|v_k^i\|}, \quad (40)$$

where v_k^i is given in (7) and

$$\tilde{v}_k^i = ((y_k^i)^\top \quad (y_{k+1}^i)^\top \quad \cdots \quad (y_{2k-1}^i)^\top)^\top. \quad (41)$$

With these estimates, we then consider the minimal k such that $\xi_k(\omega_N)$ is less than some user-defined threshold. This will be illustrated by a numerical example in the next section.

4.3 Explicit bounds

The probabilistic bound in Theorem 3 implicitly depends on the system matrices which are not available. In this section, we show that explicit bounds can also be derived by estimating $\chi_\Sigma(\cdot, k)$.

Given any $k \geq \mathcal{H}(\Sigma)$ and a sample set ω_N with the corresponding measurements $\{y_t^i : t = 0, 1, \dots, T-1, i = 1, 2, \dots, N\}$, let

$$\bar{\zeta}_k(\omega_N) := \max_{1 \leq i \leq N} \|v_k^i\|, \quad \underline{\zeta}_k(\omega_N) := \min_{1 \leq i \leq N} \|v_k^i\|, \quad (42)$$

where v_k^i is defined in (7). For any $\varepsilon \in (0, 1)$ and $\varepsilon' \in (0, 1)$, let us define

$$\psi_{\varepsilon, \varepsilon'}(\omega_N) := \frac{1}{\delta\left(\frac{\varepsilon M^k}{2}\right) \frac{\bar{\zeta}_k(\omega_N)}{\underline{\zeta}_k(\omega_N)} - \sqrt{2 - 2\delta\left(\frac{\varepsilon' M^k}{2}\right)}}. \quad (43)$$

This allows to estimate $\chi_\Sigma(\cdot, k)$ in Theorem 3, which leads to an explicit probabilistic upper bound on $\rho(\Sigma)$, as stated in the following theorem.

Theorem 4 *Consider the same conditions as in Theorem 3. For any $\bar{\varepsilon}, \varepsilon, \varepsilon' \in (0, 1)$, with probability no smaller than $1 - \phi(\bar{\varepsilon}; d, N) - (1 - \varepsilon)^N - (1 - \varepsilon')^N$,*

$$\rho(\Sigma) \leq \frac{\gamma^*(\omega_N)}{T^{-k} \sqrt{\delta\left(\frac{\bar{\varepsilon} M^T (\psi_{\varepsilon, \varepsilon'}(\omega_N))^{n-1} \sqrt{(P^*(\omega_N))^{n-1}}}{2}\right)}} \quad (44)$$

where $\psi_{\varepsilon, \varepsilon'}(\omega_N)$ is given in (43).

Remark 6 (Checking stability) *We are now ready to continue our discussion on checking stability. As it was mentioned in Section 3.1, the main idea is to solve (9) and then to check if $\gamma_k^*(\omega_N)$ is sufficiently smaller than 1. If we want to make this idea reliable, we can proceed as follows. One option is to choose the confidence level β and check if the right-hand side of (31) is smaller than*

1, i.e., if $\gamma_k^(\omega_N) \leq T^{-k} \sqrt{\delta\left(\frac{\varepsilon M^T \sqrt{(c\bar{\lambda})^{n-1}}}{2}\right)}$. We can then conclude that the underlying system is*

stable, with probability $1 - \beta$ over the data. The drawback of this approach that we need to estimate c , for which we need some a-priori knowledge on the underlying system. Another option is to use Theorem 4: we choose $\varepsilon, \varepsilon'$ such that $\beta = \phi(\bar{\varepsilon}; d, N) - (1 - \varepsilon)^N - (1 - \varepsilon')^N$ and then we check if

$\gamma_k^(\omega_N) \leq T^{-k} \sqrt{\delta\left(\frac{\bar{\varepsilon} M^T (\psi_{\varepsilon, \varepsilon'}(\omega_N))^{n-1} \sqrt{(P^*(\omega_N))^{n-1}}}{2}\right)}$. If the latter is the case, then the underlying*

system is stable with probability $1 - \beta$ over the sampled data. Note that $P^(\omega_N)$ is computed at the same time as $\gamma_k^*(\omega_N)$, when (9) is solved.*

5 Numerical examples

In this section, we illustrate the performance of the proposed approach on numerical examples. To show the advantages of our approach, we make a comparison with the identification-based approach in which we identify the model using hybrid system identification techniques and analyze the stability based on the model. Our codes are available via <https://github.com/zhemingwang/DataDrivenStabilityAnalysis>, and all experiments are run on a MacBook Pro with an Intel Core i7 and 16 GB of RAM.

5.1 Comparison with hybrid system identification

To analyze properties of the system, a natural idea is to first identify a dynamical model which then allows us to use well-documented model-based techniques. For hybrid systems, there also exist many identification techniques, see, e.g., [14, 15]. In the comparison, we consider switched auto-regressive models, which are quite popular in hybrid system identification, see, e.g., [16]. For stability analysis, we convert the switched auto-regressive model into its state-space form using the transformation described in [51] and compute the JSR of the augmented model using the JSR toolbox [52]. Note that the equivalence between switched affine models and switched ARX models has been proved in [51] under the pathwise observability condition. In particular, let us emphasize that such a conversion preserves internal stability under the pathwise observability condition. The key step in the identification procedure is clustering where we group or label the output trajectories. For each group, we then estimate the parameters of its auto-regressive model by solving a least-square problem. A few clustering algorithms that we use in the simulation are the following: Generalized Principal Component Analysis (GPCA) [53], Sparse Subspace Clustering (SSC) [54], Sparse Subspace Clustering by Orthogonal Matching Pursuit (SSC-OMP), [55], Elastic net Subspace Clustering (EnSC) [56], and Piecewise Affine Regression and Classification (PARC) [57].

The numerical example is a three-dimensional switched linear system with 3 modes and 2 outputs. We then generate an output data set of length $T = 5$ and set $k = 3$ in (7). The order of the auto-regressive models is set to be k and the maximal number of groups in the clustering algorithms is set to be M^{k-1} . The results are given in Figure 2, which shows the JSR estimation of our approach and the identification-based approach for different clustering algorithms as the number of data points increases. From this figure, we can see that the modeling error in hybrid system identification can lead to a false stability inference. Note that more data does not necessarily result in more accurate system identification from an algorithmic point of view, especially when there is a clustering step in the whole procedure. This is supported by Figure 2 as the JSR estimation can become worse as N increases. In contrast, our approach is guaranteed to provide better JSR estimation as N increases. Another practical issue of the identification-based approach is that in the clustering algorithms there are some parameters that need to be tuned carefully. We also make a comparison with the two standard clustering algorithms GPCA and SSC in terms of computational time in Table 1.

N		500	1000	1500	2000
Our approach		22.08	50.53	61.18	62.48
GPCA	T_{id}	4.92	40.46	137.90	256.27
	T_{JSR}	15.42	32.93	14.22	5.19
SSC	T_{id}	894.20	2550.08	3554.40	5015.54
	T_{JSR}	67.90	124.18	67.10	120.97

Table 1: Computational time (second) for different algorithms: T_{id} denotes the computational time for the identification step and T_{JSR} denotes the computational time for computing the JSR using the model.

5.2 A six-dimensional example

We also apply the proposed data-driven approach to a higher-dimensional example. We consider a switched linear system with $n = 6$, $M = 3$, $p = 2$ where the entries of the dynamics and output matrices are chosen randomly from the uniform distribution over $[-1, 1]$. We then sample a finite set of initial states following the Gaussian distribution $\mathcal{N}(0, I)$ and project them onto the unit sphere. In this way, we obtain a set of initial states that are uniformly distributed on the unit sphere. With this data set, we formulate Problem (9) with $k = 3$. This is a valid option as it can be verified that the system is *pathwise observable* with the *pathwise observability index* $\mathcal{H}(\Sigma) = 2$.

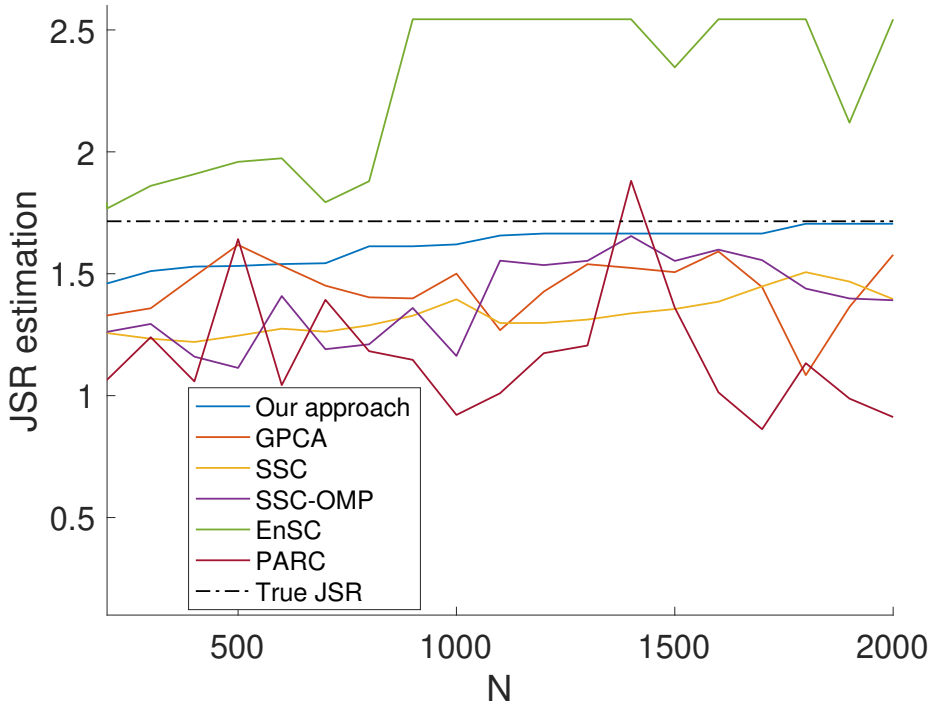


Figure 2: Comparison with the identification-based approach using different hybrid system identification algorithms.

When the information of $\mathcal{H}(\Sigma)$ is not available, we use the procedure in Section 4.2 to estimate $\mathcal{H}(\Sigma)$. Finally, we are ready to solve Problem (9). The simulation results are given in Figure 4 for different values of horizon length T . As expected, $\gamma^*(\omega_N)$ gradually approaches the true JSR (which can be obtained from the JSR toolbox [52]) as N and T increase.

6 Conclusions

In this paper, we have studied the problem of output-based stability learning of a special family of hybrid systems, namely switched linear systems, using a time-series output data set. We leverage a recently introduced approach for state-based data-driven stability analysis. However, the fact that only a (possibly low-dimensional) output is observed incurs additional theoretical challenges that we tackle here. With the proposed output-based Lyapunov framework, we are able to derive probabilistic stability certificates for partially observed switched linear systems. We have also presented additional procedures which allow to use explicit upper bounds on the JSR of switched linear systems. To validate our approach, we have made a comparison with the identification-based approach in which a model is identified using existing clustering algorithms in the literature. Numerical results suggest that it is beneficial to use our direct analysis approach as modeling errors can lead to false stability inferences.

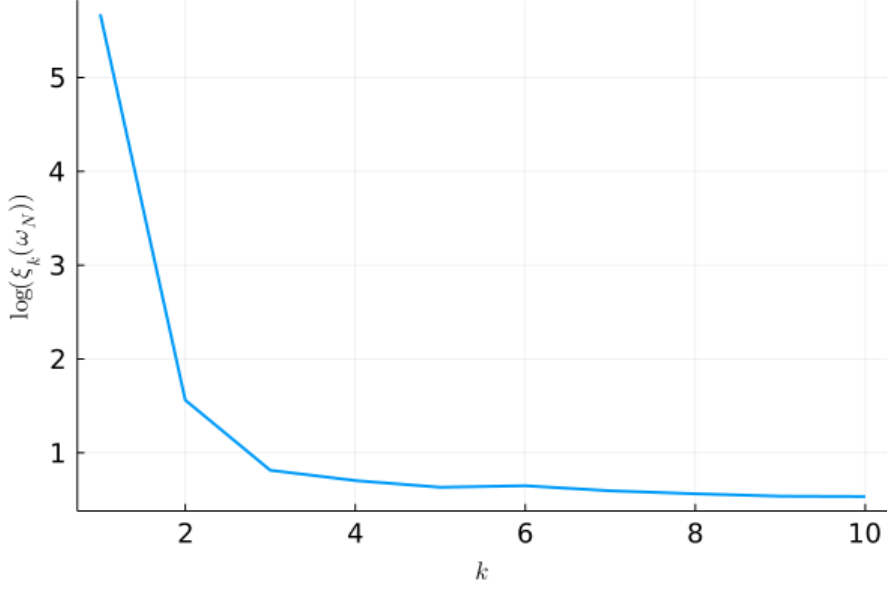


Figure 3: Estimates of ξ_k^o for different values of k .

Appendix

A1 Proof of Theorem 2

Since Σ is pathwise observable, $\mathcal{P}_\Sigma^\sigma(P, 0, k) \succ 0$ and $\mathcal{P}_\Sigma^\sigma(P, \ell, k) \succ 0$ for any $\sigma \in \mathcal{M}^{k+\ell}$. We pick an arbitrary $\sigma \in \mathcal{M}^{k+\ell}$ and consider the Cholesky decompositions of $\mathcal{P}_\Sigma^\sigma(P, 0, k)$ and $\mathcal{P}_\Sigma^\sigma(P, \ell, k)$:

$$U^\top U = \mathcal{P}_\Sigma^\sigma(P, 0, k), \tilde{U}^\top \tilde{U} = \mathcal{P}_\Sigma^\sigma(P, \ell, k),$$

where $U \in \mathbb{R}^{n \times n}$ and $\tilde{U} \in \mathbb{R}^{n \times n}$ are upper triangular matrices. Hence, by the change of coordinates, (24) becomes

$$\begin{aligned} & (\tilde{U} A_{\sigma_{0:\ell-1}} U^{-1} x)^\top \tilde{U} A_{\sigma_{0:\ell-1}} U^{-1} x \\ & \leq \gamma^{2\ell} x^\top x, \quad \forall x \in U(\mathbb{S}_{n-1} \setminus S). \end{aligned} \quad (45)$$

The homogeneity of the dynamics implies that

$$\begin{aligned} & (\tilde{U} A_{\sigma_{0:\ell-1}} U^{-1} x)^\top \tilde{U} A_{\sigma_{0:\ell-1}} U^{-1} x \\ & \leq \gamma^{2\ell} x^\top x, \quad \forall x \in \Pi_{\mathbb{S}_{n-1}}(U(\mathbb{S}_{n-1} \setminus S)), \end{aligned} \quad (46)$$

where $\Pi_{\mathbb{S}_{n-1}}(\cdot)$ denote the projection onto \mathbb{S}_{n-1} . Equivalently, we can write (46) as

$$\begin{aligned} & (\tilde{U} A_{\sigma_{0:\ell-1}} U^{-1} x)^\top \tilde{U} A_{\sigma_{0:\ell-1}} U^{-1} x \\ & \leq \gamma^{2\ell} x^\top x, \quad \forall x \in \mathbb{S}_{n-1} \setminus \Pi_{\mathbb{S}_{n-1}}(US). \end{aligned} \quad (47)$$

Following the same arguments in Theorem 15 in [25], we obtain that

$$\tilde{U} A_{\sigma_{0:\ell-1}} U^{-1} \mathbb{S}_{n-1} \subset \frac{\gamma^\ell}{\delta\left(\frac{\mu(S)\chi(\mathcal{P}_\Sigma^\sigma(P, 0, k))}{2}\right)} \mathbb{B}_n. \quad (48)$$

An illustration of the function $\delta(\cdot)$ is given in Figure 1. Hence, it holds that

$$\begin{aligned} & \left(\tilde{U} A_{\sigma_{0:\ell-1}} U^{-1}\right)^\top \tilde{U} A_{\sigma_{0:\ell-1}} U^{-1} \\ & \preceq \frac{\gamma^{2\ell}}{\left(\delta\left(\frac{\mu(S)\chi(\mathcal{P}_\Sigma^\sigma(P, 0, k))}{2}\right)\right)^2} I, \end{aligned} \quad (49)$$

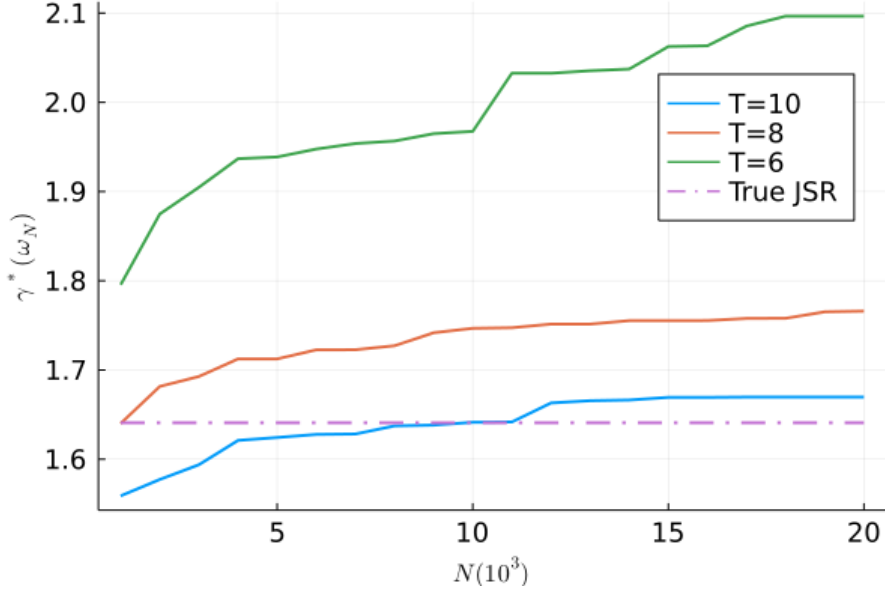


Figure 4: Data-based convergence rates for different horizon lengths.

which implies that

$$\begin{aligned}
& A_{\sigma_{0:\ell-1}}^\top \mathcal{P}_\Sigma^\sigma(P, \ell, k) A_{\sigma_{0:\ell-1}} \\
& \preceq \frac{\gamma^{2\ell}}{\left(\delta\left(\frac{\mu(S)\chi(\mathcal{P}_\Sigma^\sigma(P, 0, k))}{2}\right)\right)^2} \mathcal{P}_\Sigma^\sigma(P, 0, k). \tag{50}
\end{aligned}$$

As σ is chosen arbitrarily, the inequality above holds for any $\sigma \in \mathcal{M}^{k+\ell}$. Note that the function $\delta(\cdot)$ is non-increasing. Therefore, with the definition of $\chi_\Sigma(P, k)$ as in (26), we arrive at

$$\begin{aligned}
& A_{\sigma_{0:\ell-1}}^\top \mathcal{P}_\Sigma^\sigma(P, \ell, k) A_{\sigma_{0:\ell-1}} \\
& \preceq \frac{\gamma^{2\ell}}{\left(\delta\left(\frac{\mu(S)\chi_\Sigma(P, k)}{2}\right)\right)^2} \mathcal{P}_\Sigma^\sigma(P, 0, k), \forall \sigma \in \mathcal{M}^{k+\ell}. \tag{51}
\end{aligned}$$

Finally, from Theorem 1, we arrive at (25).

A2 Proof Theorem 3

To simplify the notation, we drop the subscript k in $(\gamma_k^*(\omega_N), P_k^*(\omega_N))$ in the proof. We use the scenario approach to derive a probabilistic upper bound on the JSR via geometric analysis, following [25, 26]. From Theorem 6 in [26], the chance-constrained theorem in [31, 32] is applicable and we obtain that for any $\epsilon \in (0, 1)$,

$$\mathbb{P}^N\{\omega_N : \mathbb{P}\{V(\omega_N)\} > \epsilon\} \leq \phi(\epsilon; d, N),$$

where

$$\begin{aligned}
V(\omega_N) := & \{(x, \sigma) \in \mathbb{S}_{n-1} \times \mathcal{M}^T : \\
& x^\top A_{\sigma_{0:\ell-1}}^\top \mathcal{P}_\Sigma^\sigma(P^*(\omega_N), T-k, k) A_{\sigma_{0:\ell-1}} x \\
& > (\gamma^*(\omega_N))^{2(T-k)} x^\top \mathcal{P}_\Sigma^\sigma(P^*(\omega_N), 0, k) x\}.
\end{aligned}$$

For any $\sigma \in \mathcal{M}^T$, let $V^\sigma(\omega_N) := \{x \in \mathbb{S}_{n-1} : (x, \sigma) \in V(\omega_N)\}$. With this definition, we claim that, for any $\sigma \in \mathcal{M}^T$ $x \in \mathbb{S}_{n-1} \setminus \bigcup_{\sigma \in \mathcal{M}^T} V^\sigma(\omega_N)$

$$\begin{aligned} & x^\top A_{\sigma_{0:\ell-1}}^\top \mathcal{P}_\Sigma^\sigma(P^*(\omega_N), T-k, k) A_{\sigma_{0:\ell-1}} x \\ & \leq (\gamma^*(\omega_N))^{2(T-k)} x^\top \mathcal{P}_\Sigma^\sigma(P^*(\omega_N), 0, k) x \end{aligned}$$

Note that in the worst case, the sets $\{V^\sigma(\omega_N)\}$ are disjoint. When $\mathbb{P}\{V(\omega_N)\} \leq \epsilon$, $\mu(\bigcup_{\sigma \in \mathcal{M}^T} V^\sigma(\omega_N)) \leq \epsilon M^T$, where $\mu(\cdot)$ denotes the (probability) uniform measure on \mathbb{S}_{n-1} . From Theorem 2, we conclude that, with probability no smaller than $1 - \phi(\epsilon; d, N)$, (29) holds.

A3 Proof of Proposition 1

Proof: (i) Given $\beta \in (0, 1)$ and $\epsilon > 0$, from Corollary 3, we look for a N such that $\eta(\beta, N) \leq 1 + \epsilon$, which is equivalent to

$$\phi\left(\frac{\mathcal{I}\left(1 - \frac{1}{(1+\epsilon)^{2(T-k)}}, \frac{n-1}{2}, \frac{1}{2}\right)}{K_2}, d, N\right) \leq \beta \quad (52)$$

where $K_2 = M^T \chi(Q) \sqrt{(c\lambda)^{n-1}}$. Hence, (35) holds. We then show that $\mathcal{N}(\epsilon, \beta) = O\left(\frac{1}{\beta} \frac{1}{\epsilon^{\frac{n}{2}}}\right)$ as $\epsilon \rightarrow 0$. To do so, we first show that $\phi(\epsilon; d, N)$ can be explicitly bounded from above by $\frac{d}{\epsilon(N+1)}$ for any $\epsilon \in (0, 1)$ and any $N \in \mathbb{Z}^+$. This can be verified using the following manipulations:

$$\begin{aligned} \frac{\phi(\epsilon; d, N)}{\frac{d}{\epsilon(N+1)}} &= \frac{\sum_{i=0}^{d-1} \binom{N}{i} \epsilon^i (1-\epsilon)^{N-i}}{\frac{d}{\epsilon(N+1)}} \\ &= \sum_{i=0}^{d-1} \frac{i+1}{d} \binom{N+1}{i+1} \epsilon^{i+1} (1-\epsilon)^{N-i} \\ &\leq \sum_{i=1}^d \binom{N+1}{i} \epsilon^i (1-\epsilon)^{N+1-i} \\ &\leq \phi(\epsilon; d+1, N+1) \leq 1. \end{aligned} \quad (53)$$

This means that

$$\mathcal{N}_\phi(\epsilon, \beta) \leq \frac{d}{\beta\epsilon}. \quad (54)$$

We then show that $\mathcal{I}(x; \frac{n-1}{2}, \frac{1}{2}) \geq K_1 x^{\frac{n}{2}}$, where $\mathcal{I}(x; a, b)$ is defined as in (23) and $K_1 = \frac{\frac{4}{n}}{\int_0^1 t^{\frac{n-3}{2}} (1-t)^{-\frac{1}{2}} dt}$, as follows

$$\begin{aligned} \mathcal{I}\left(x; \frac{n-1}{2}, \frac{1}{2}\right) &= \frac{\int_0^x t^{\frac{n-3}{2}} (1-t)^{-\frac{1}{2}} dt}{\int_0^1 t^{\frac{n-3}{2}} (1-t)^{-\frac{1}{2}} dt} \\ &= \frac{\int_0^x t^{\frac{n-2}{2}} \frac{1}{(t(1-t))^{\frac{1}{2}}} dt}{\int_0^1 t^{\frac{n-3}{2}} (1-t)^{-\frac{1}{2}} dt} \\ &\geq \frac{2 \int_0^x t^{\frac{n-2}{2}} dt}{\int_0^1 t^{\frac{n-3}{2}} (1-t)^{-\frac{1}{2}} dt} = K_1 x^{\frac{n}{2}}, \end{aligned} \quad (55)$$

where the inequality is due to the fact that $t(1-t) \leq \frac{1}{4}$. Combing (54) and (55) yields

$$\begin{aligned}
& \mathcal{N}_\phi\left(\frac{\mathcal{I}\left(1 - \frac{1}{(1+\varepsilon)^{2(T-k)}}; \frac{n-1}{2}, \frac{1}{2}\right)}{M^T \chi(Q) \sqrt{(c\lambda)^{n-1}}}, \beta\right) \\
& \leq \frac{K_2 d}{\beta \mathcal{I}\left(1 - \frac{1}{(1+\varepsilon)^{2(T-k)}}; \frac{n-1}{2}, \frac{1}{2}\right)} \\
& \leq \frac{K_2 d}{\beta K_1 \left(1 - \frac{1}{(1+\varepsilon)^{2(T-k)}}\right)^{\frac{n}{2}}} \\
& = \frac{K_2 d}{K_1 \beta} \frac{\left(1 + \frac{1}{\varepsilon}\right)^{n(T-k)}}{\left(\left(1 + \frac{1}{\varepsilon}\right)^{2(T-k)} - \left(\frac{1}{\varepsilon}\right)^{2(T-k)}\right)^{\frac{n}{2}}}.
\end{aligned}$$

By some manipulations, we get that $N_s(\varepsilon, \beta) \leq \frac{K_2 d}{K_1 \beta} \left(\frac{1}{\varepsilon}\right)^{\frac{n}{2}}$ as $\varepsilon \rightarrow 0$, which can be equivalently expressed as (36). This completes the proof.

(ii) From the definition of sample complexity in Definition 5, it holds that $\mathbb{P}^N\{\omega_N : \gamma^o > (1 + \varepsilon)\gamma^*(\omega_N)\} \leq \beta$ for any $N \geq N_s(\varepsilon, \beta)$. The fact that $\rho(\Sigma) > 1$ implies that $\gamma^o > 1$. With this, it can be verified that $\{\omega_N : \gamma^*(\omega_N) < \frac{1}{1+\varepsilon}\} \subseteq \{\omega_N : \gamma^o > (1 + \varepsilon)\gamma^*(\omega_N)\}$, which implies (37).

A4 Proof of Proposition 2

When $k \geq \mathcal{H}(\Sigma)$, $\text{rank}(\mathcal{O}_\Sigma(\boldsymbol{\sigma})) = n$ for any $\boldsymbol{\sigma} \in \mathcal{M}^k$. Hence, ξ_k^o is bounded. (Necessity) The proof goes by contradiction. Suppose there exists $k < \mathcal{H}(\Sigma)$ such that $\xi_k^o < \infty$. From the definition of $\mathcal{H}(\Sigma)$, there exists at least one sequence $\tilde{\boldsymbol{\sigma}} \in \mathcal{M}^k$ such that $\text{rank}(\mathcal{O}_\Sigma(\tilde{\boldsymbol{\sigma}})) < n$. The fact that $\xi_k^o < \infty$ implies that $\mathcal{O}_\Sigma(\tilde{\boldsymbol{\sigma}})x = 0$ implies $\mathcal{O}_\Sigma(\boldsymbol{\sigma})A_{\tilde{\boldsymbol{\sigma}}}x = 0$ for any $\boldsymbol{\sigma} \in \mathcal{M}^k$. In particular, $\mathcal{O}_\Sigma(\tilde{\boldsymbol{\sigma}})x = 0$ implies $\mathcal{O}_\Sigma(\tilde{\boldsymbol{\sigma}})A_{\tilde{\boldsymbol{\sigma}}}x = 0$. Then, there exists a matrix $G \in \mathbb{R}^{pk \times pk}$ such that $\mathcal{O}_\Sigma(\tilde{\boldsymbol{\sigma}})A_{\tilde{\boldsymbol{\sigma}}} = G\mathcal{O}_\Sigma(\tilde{\boldsymbol{\sigma}})$, i.e., each row of $\mathcal{O}_\Sigma(\tilde{\boldsymbol{\sigma}})A_{\tilde{\boldsymbol{\sigma}}}$ can be written as a linear combination of the rows in $\mathcal{O}_\Sigma(\tilde{\boldsymbol{\sigma}})$. As a result, the periodic sequence $\tilde{\boldsymbol{\sigma}}\tilde{\boldsymbol{\sigma}}\cdots$ is unobservable, which contradicts the fact that Σ is pathwise observable.

A5 Proof of Theorem 4

To prove Theorem 4, we first state the following bound on the singular value of the observability matrices.

Lemma 1 *Consider the same conditions as in Theorem 3. Then, for any $\varepsilon \in (0, 1)$ and $\varepsilon' \in (0, 1)$, with probability no smaller than $1 - (1 - \varepsilon)^N - (1 - \varepsilon')^N$,*

$$\frac{\max_{\boldsymbol{\sigma} \in \mathcal{M}^k} \sigma_{\max}(\mathcal{O}_\Sigma(\boldsymbol{\sigma}))}{\min_{\boldsymbol{\sigma} \in \mathcal{M}^k} \sigma_{\min}(\mathcal{O}_\Sigma(\boldsymbol{\sigma}))} \leq \psi(\omega_N) \quad (56)$$

Proof: Since only the first k steps of the trajectory are relevant, we only consider the sample set ω_N of length k . The proof consists of three steps:

Step 1: We consider the robust optimization problem below:

$$\min_{\zeta \geq 0} \zeta : \|\mathcal{O}_\Sigma(\boldsymbol{\sigma})x\| \leq \zeta, \forall (x, \boldsymbol{\sigma}) \in \mathbb{S}_{n-1} \times \mathcal{M}^k. \quad (57)$$

The optimum is exactly $\max_{\boldsymbol{\sigma} \in \mathcal{M}^k} \sigma_{\max}(\mathcal{O}_\Sigma(\boldsymbol{\sigma}))$. We want to show that the solution $\bar{\zeta}_k(\omega_N)$ in (42) provides a probabilistic upper bound for Problem (57). To do so, we show the following chance-constrained result, given an $\varepsilon \in (0, 1)$,

$$\mathbb{P}^N\{\omega_N : \Omega(\bar{\zeta}_k(\omega_N)) > \varepsilon\} = (1 - \varepsilon)^N \quad (58)$$

where $\Omega(\zeta) := \mathbb{P}\{(x, \boldsymbol{\sigma}) \in \mathbb{S}_{n-1} \times \mathcal{M}^k : \|\mathcal{O}_\Sigma(\boldsymbol{\sigma})x\| > \zeta\}$. The function $\Omega : [0, \max_{\boldsymbol{\sigma} \in \mathcal{M}^k} \sigma_{\max}(\mathcal{O}_\Sigma(\boldsymbol{\sigma}))] \rightarrow [0, 1]$ is decreasing. Hence, there exists a unique ζ such that $\Omega(\zeta) = \varepsilon$, denoted by ζ_ε . With this,

the set $\{\omega_N : \Omega(\bar{\zeta}_k(\omega_N)) > \varepsilon\}$ can be equivalently expressed as $\{\omega_N : \omega_N \cap \{(x, \sigma) : \|\mathcal{O}_\Sigma(\sigma)x\| \geq \zeta_\varepsilon\} = \emptyset\}$ (whose measure is $(1 - \varepsilon)^N$), which leads to (58). Let us define the projected violating subset $\tilde{\mathbb{S}}$ as follows:

$$\tilde{\mathbb{S}} := \{x \in \mathbb{S}_{n-1} : \exists \sigma \in \mathcal{M}^k, \|\mathcal{O}_\Sigma(\sigma)x\| > \bar{\zeta}_k(\omega_N)\}.$$

For any $\sigma \in \mathcal{M}^k$, we also define:

$$\tilde{\mathbb{S}}_\sigma := \{x \in \mathbb{S}_{n-1} : \|\mathcal{O}_\Sigma(\sigma)x\| > \bar{\zeta}_k(\omega_N)\} \quad (59)$$

By definition, $\tilde{\mathbb{S}} = \cup_{\sigma \in \mathcal{M}^k} \tilde{\mathbb{S}}_\sigma$, which implies that

$$\mathbb{P}_x\{\tilde{\mathbb{S}}\} \leq \sum_{\sigma \in \mathcal{M}^k} \mathbb{P}_x\{\tilde{\mathbb{S}}_\sigma\} \quad (60)$$

where \mathbb{P}_x denotes the uniform probability measure on \mathbb{S}_{n-1} and the equality holds when the sets $\{\tilde{\mathbb{S}}_\sigma\}_{\sigma \in \mathcal{M}^k}$ are disjoint. With the inequality above, we get that

$$\begin{aligned} \Omega(\bar{\zeta}_k(\omega_N)) &= \sum_{\sigma \in \mathcal{M}^k} \mathbb{P}_x\{\tilde{\mathbb{S}}_\sigma\} \mathbb{P}_\sigma\{\sigma\} = \frac{1}{M^k} \sum_{\sigma \in \mathcal{M}^k} \mathbb{P}_x\{\tilde{\mathbb{S}}_\sigma\} \\ &\geq \frac{1}{M^k} \mathbb{P}_x\{\tilde{\mathbb{S}}\} \end{aligned} \quad (61)$$

where \mathbb{P}_σ denote the uniform distribution on \mathcal{M}^k . This means that $\Omega(\bar{\zeta}_k(\omega_N)) \leq \varepsilon$ implies $\mathbb{P}_x\{\tilde{\mathbb{S}}\} \leq M^k \varepsilon$. Hence,

$$\mathbb{P}^N\{\omega_N : \mathbb{P}_x\{\tilde{\mathbb{S}}\} \leq \varepsilon M^k\} = (1 - \varepsilon)^N \quad (62)$$

from (58). Finally, following the same lines as the proof of [25, Theorem 15], we conclude that, with probability no smaller than $1 - (1 - \varepsilon)^N$,

$$\max_{\sigma \in \mathcal{M}^k} \sigma_{\max}(\mathcal{O}_\Sigma(\sigma)) \leq \frac{\bar{\zeta}_k(\omega_N)}{\delta(\frac{\varepsilon M^k}{2})}. \quad (63)$$

Step 2: Similarly, we define the following robust optimization problem:

$$\max_{\zeta \geq 0} \zeta : \|\mathcal{O}_\Sigma(\sigma)x\| \geq \zeta, \forall (x, \sigma) \in \mathbb{S}_{n-1} \times \mathcal{M}^k. \quad (64)$$

As the constraint above is not convex in x , we cannot repeat the same reasoning in (i). Nevertheless, it still holds that the optimum of (64) is $\min_{\sigma \in \mathcal{M}^k} \sigma_{\min}(\mathcal{O}_\Sigma(\sigma))$, which can be attained, i.e., there exists (x^*, σ^*) such that

$$\|\mathcal{O}_\Sigma(\sigma^*)x^*\| = \min_{\sigma \in \mathcal{M}^k} \sigma_{\min}(\mathcal{O}_\Sigma(\sigma)).$$

For any $\varepsilon \in (0, 1)$, we define the set $\bar{\mathbb{S}} := \{x \in \mathbb{S}_{n-1} : |x^\top x^*| \geq \delta(\frac{\varepsilon M^k}{2})\}$ with $\mathbb{P}_x\{\bar{\mathbb{S}}\} = \varepsilon M^k$. The probability that $\omega_N \cap \bar{\mathbb{S}} \times \{\sigma^*\} \neq \emptyset$ is $1 - (1 - \varepsilon)^N$. In this case, there exists $(\bar{x}, \bar{\sigma}) \in \omega_N$ such that $\sigma = \sigma^*$ and $|\bar{x}^\top x^*| \geq \delta(\frac{\varepsilon M^k}{2})$, which implies that $\|\bar{x} - x^*\| \leq \sqrt{2 - 2\delta(\frac{\varepsilon M^k}{2})}$ or $\|\bar{x} + x^*\| \leq \sqrt{2 - 2\delta(\frac{\varepsilon M^k}{2})}$. From the definition in (42), $\|\mathcal{O}_\Sigma(\bar{\sigma})\bar{x}\| \geq \zeta_k(\omega_N)$. We then consider the case that $\|\bar{x} - x^*\| \leq \sqrt{2 - 2\delta(\frac{\varepsilon M^k}{2})}$ (the analysis is exactly the same for the other case). With these relations, it holds that

$$\begin{aligned} \|\mathcal{O}_\Sigma(\sigma^*)x^*\| &= \|\mathcal{O}_\Sigma(\sigma^*)(\bar{x} + x^* - \bar{x})\| \\ &\geq \|\mathcal{O}_\Sigma(\sigma^*)\bar{x}\| - \|\mathcal{O}_\Sigma(\sigma^*)\| \|x^* - \bar{x}\| \\ &\geq \zeta_k(\omega_N) - \max_{\sigma \in \mathcal{M}^k} \|\mathcal{O}_\Sigma(\sigma)\| \sqrt{2 - 2\delta(\frac{\varepsilon M^k}{2})}. \end{aligned}$$

Based on this inequality, we conclude that, with probability no smaller than $1 - (1 - \varepsilon)^N$,

$$\begin{aligned} \min_{\boldsymbol{\sigma} \in \mathcal{M}^k} \sigma_{\min}(\mathcal{O}_{\Sigma}(\boldsymbol{\sigma})) &\geq \zeta_k(\omega_N) \\ &- \max_{\boldsymbol{\sigma} \in \mathcal{M}^k} \|\mathcal{O}_{\Sigma}(\boldsymbol{\sigma})\| \sqrt{2 - 2\delta\left(\frac{\varepsilon M^k}{2}\right)}. \end{aligned} \quad (65)$$

Step 3: We now combine the results from Steps 1 & 2. From Step 1, for any $\varepsilon \in (0, 1)$, the probability that (63) does not hold is less than $(1 - \varepsilon)^N$. From Step 2, for any $\varepsilon' \in (0, 1)$ the probability that (65) with ε' does not hold is also less than $(1 - \varepsilon')^N$. Hence, the probability that (63) or (65) does not hold becomes $(1 - \varepsilon)^N + (1 - \varepsilon')^N$, which means that the probability that both of (63) and (65) hold is no smaller than $1 - (1 - \varepsilon)^N - (1 - \varepsilon')^N$. Thus, (56) holds with probability no smaller than $1 - (1 - \varepsilon)^N - (1 - \varepsilon')^N$. \square

From the arguments in the proof of Corollary 2, it holds that

$$\chi_{\Sigma}(P, k) \leq \sqrt{(c_k^2 \kappa(P))^{n-1}}$$

where

$$c_k = \frac{\max_{\boldsymbol{\sigma} \in \mathcal{M}^k} \sigma_{\max}(\mathcal{O}_{\Sigma}(\boldsymbol{\sigma}))}{\min_{\boldsymbol{\sigma} \in \mathcal{M}^k} \sigma_{\min}(\mathcal{O}_{\Sigma}(\boldsymbol{\sigma}))}.$$

This, together with Theorem 3, and Lemma 1, leads to the result in Theorem 4.

References

- [1] T. G. J. Rudner and H. Toner. Key concepts in AI safety: Robustness and adversarial examples. Technical report, Center for Security and Emerging Technology, 2021.
- [2] T. G. Dietterich. Robust artificial intelligence and robust human organizations. *Frontiers of Computer Science*, 13(1):1–3, 2019.
- [3] R. Hamon, H. Junklewitz, and I. Sanchez. Robustness and explainability of artificial intelligence. *Publications Office of the European Union*, 2020.
- [4] R. Alur. *Principles of Cyber-Physical Systems*. The MIT Press. MIT Press, 2015.
- [5] E. A. Lee. The past, present and future of cyber-physical systems: A focus on models. *Sensors*, 15(3):4837–4869, 2015.
- [6] Y. Yuan, X. Tang, W. Zhou, W. Pan, X. Li, H.-T. Zhang, H. Ding, and J. Goncalves. Data driven discovery of cyber physical systems. *Nat. Commun.*, 10(4894), 2019.
- [7] R. Goebel, R. G. Sanfelice, and A. R. Teel. *Hybrid Dynamical Systems: Modeling, Stability, and Robustness*. Princeton University Press, New Jersey, NJ, 2012.
- [8] D. Liberzon. *Switching in systems and control*. Springer Science & Business Media, 2003.
- [9] Z. Sun and S. S. Ge. *Switched linear systems : control and design*. Springer, London, 2005.
- [10] E.D. Sontag. *Mathematical Control Theory*. Springer-Verlag, 1990.
- [11] M. W. Hirsch, S. Smale, and R. L. Devaney. *Differential equations, dynamical systems, and an introduction to chaos*. Academic press, 2012.
- [12] L. Ljung. System identification. In *Signal analysis and prediction*, pages 163–173. Springer, 1998.
- [13] J. Schoukens and L. Ljung. Nonlinear system identification: A user-oriented road map. *IEEE Control Systems Magazine*, 39(6):28–99, 2019.
- [14] S. Paoletti, A. Juloski, G. Ferrari-Trecate, and R. Vidal. Identification of hybrid systems: A tutorial. *European Journal of Control*, 13(2-3):242 – 260, 2007.

- [15] F. Lauer and G. Bloch. *Hybrid system identification: Theory and algorithms for learning switching models*, volume 478. Springer, 2018.
- [16] R. Vidal. Recursive identification of switched ARX systems. *Automatica*, 44(9):2274 – 2287, 2008.
- [17] Laurent Bako. Analysis of the least sum-of-minimums estimator for switched systems. *IEEE Transactions on Automatic Control*, 66(8):3733–3740, 2021.
- [18] L. Massucci, F. Lauer, and M. Gilson. A statistical learning perspective on switched linear system identification. *Automatica*, 145:110532, 2022.
- [19] P. B. Cox, R. Tóth, and M. Petreczky. Towards efficient maximum likelihood estimation of lpv-ss models. *Automatica*, 97:392–403, 2018.
- [20] J.-W. van Wingerden and M. Verhaegen. Subspace identification of Bilinear and LPV systems for open- and closed-loop data. *Automatica*, 45(2):372–381, 2009.
- [21] F. Lauer. On the complexity of switching linear regression. *Automatica*, 74:80–83, 2016.
- [22] R. M. Jungers. *The joint spectral radius: theory and applications*, volume 385. Springer Science & Business Media, 2009.
- [23] V. Vapnik. *The nature of statistical learning theory*. Springer science & business media, 1999.
- [24] S. Shalev-Shwartz and S. Ben-David. *Understanding machine learning: From theory to algorithms*. Cambridge university press, 2014.
- [25] J. Kenanian, A. Balkan, R. M. Jungers, and P. Tabuada. Data driven stability analysis of black-box switched linear systems. *Automatica*, 109:108533, 2019.
- [26] G. O. Berger, R. M. Jungers, and Z. Wang. Chance-constrained quasi-convex optimization with application to data-driven switched systems control. In *Learning for Dynamics and Control*, pages 571–583. PMLR, 2021.
- [27] A. Rubbens, Z. Wang, and R. M. Jungers. Data-driven stability analysis of switched linear systems with sum of squares guarantees. In *The 7th IFAC Conference on Analysis and Design of Hybrid Systems*. IFAC, 2021.
- [28] Z. Wang and R. M. Jungers. A data-driven method for computing polyhedral invariant sets of black-box switched linear systems. *IEEE Control Systems Letters*, 5(5):1843 – 1848, 2021.
- [29] G. Calafiore and M. C. Campi. Uncertain convex programs: randomized solutions and confidence levels. *Mathematical Programming*, 102(1):25–46, 2005.
- [30] G. Calafiore and M. C. Campi. The scenario approach to robust control design. *IEEE Transactions on Automatic Control*, 51(5):742–753, 2006.
- [31] M. C. Campi and S. Garatti. The exact feasibility of randomized solutions of uncertain convex programs. *SIAM Journal on Optimization*, 19(3):1211–1230, 2008.
- [32] G. C. Calafiore. Random convex programs. *SIAM Journal on Optimization*, 20(6):3427–3464, 2010.
- [33] K. Margellos, P. Goulart, and J. Lygeros. On the road between robust optimization and the scenario approach for chance constrained optimization problems. *IEEE Transactions on Automatic Control*, 59(8):2258–2263, 2014.
- [34] L. G. Valiant. A theory of the learnable. *Communications of the ACM*, 27(11):1134–1142, 1984.
- [35] S. Chen, M. Fazlyab, M. Morari, G. J. Pappas, and V. M. Preciado. Learning lyapunov functions for piecewise affine systems with neural network controllers. *arXiv preprint arXiv:2008.06546*, 2020.
- [36] N. Boffi, S. Tu, N. Matni, J.-J. Slotine, and V. Sindhvani. Learning stability certificates from data. In *Conference on Robot Learning*, pages 1341–1350. PMLR, 2021.

- [37] A. Taylor, A. Singletary, Y. Yue, and A. Ames. Learning for safety-critical control with control barrier functions. In *Learning for Dynamics and Control*, pages 708–717. PMLR, 2020.
- [38] H. Ravanbakhsh and S. Sankaranarayanan. Learning control lyapunov functions from counterexamples and demonstrations. *Autonomous Robots*, 43(2):275–307, 2019.
- [39] Y.-C. Chang, N. Roohi, and S. Gao. Neural lyapunov control. *Advances in neural information processing systems*, 32, 2019.
- [40] S. M. Richards, F. Berkenkamp, and A. Krause. The lyapunov neural network: Adaptive stability certification for safe learning of dynamical systems. In *Conference on Robot Learning*, pages 466–476. PMLR, 2018.
- [41] H. Tsukamoto, S.-J. Chung, and J.-J. E. Slotine. Contraction theory for nonlinear stability analysis and learning-based control: A tutorial overview. *Annual Reviews in Control*, 52:135–169, 2021.
- [42] Y.-C. Chang and S. Gao. Stabilizing neural control using self-learned almost lyapunov critics. In *2021 IEEE International Conference on Robotics and Automation (ICRA)*, pages 1803–1809. IEEE, 2021.
- [43] C. Dawson, S. Gao, and C. Fan. Learning certified control using contraction metric: A survey of neural lyapunov, barrier, and contraction methods. *arXiv preprint arXiv:2202.11762*, 2022.
- [44] P. Giesl, B. Hamzi, M. Rasmussen, and K. Webster. Approximation of lyapunov functions from noisy data. *Journal of Computational Dynamics*, 7(1), 2020.
- [45] D. Bertsekas, A. Nedić, and A. E. Ozdaglar. *Convex analysis and optimization*. Athena Scientific, 2003.
- [46] S. Boyd and L. Vandenberghe. *Convex Optimization*. Cambridge University Press, 2004.
- [47] M. Babaali and M. Egerstedt. Pathwise observability and controllability are decidable. In *42nd IEEE International Conference on Decision and Control (IEEE Cat. No. 03CH37475)*, volume 6, pages 5771–5776. IEEE, 2003.
- [48] R. M. Jungers, A. Kundu, and W.P.M.H. Heemels. Observability and controllability analysis of linear systems subject to data losses. *IEEE Transactions on Automatic Control*, 63(10):3361–3376, 2017.
- [49] M. Babaali and M. Egerstedt. Observability of switched linear systems. In *International Workshop on Hybrid Systems: Computation and Control*, pages 48–63. Springer, 2004.
- [50] T. Alamo, R. Tempo, A. Luque, and D. R. Ramirez. Randomized methods for design of uncertain systems: Sample complexity and sequential algorithms. *Automatica*, 52:160–172, 2015.
- [51] S. Weiland, A. Lj. Juloski, and B. Vet. On the equivalence of switched affine models and switched arx models. In *Proceedings of the 45th IEEE Conference on Decision and Control*, pages 2614–2618. IEEE, 2006.
- [52] G. Vankeerberghen, J. Hendrickx, and R. M. Jungers. Jsr: A toolbox to compute the joint spectral radius. In *Proceedings of the 17th international conference on Hybrid systems: computation and control*, pages 151–156, 2014.
- [53] R. Vidal, Y. Ma, and S. Sastry. Generalized principal component analysis (GPCA). *IEEE transactions on pattern analysis and machine intelligence*, 27(12):1945–1959, 2005.
- [54] E. Elhamifar and R. Vidal. Sparse subspace clustering: Algorithm, theory, and applications. *IEEE transactions on pattern analysis and machine intelligence*, 35(11):2765–2781, 2013.
- [55] C. You, D. Robinson, and R. Vidal. Scalable sparse subspace clustering by orthogonal matching pursuit. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 3918–3927, 2016.

- [56] C. You, C.-G. Li, D. P. Robinson, and R. Vidal. Oracle based active set algorithm for scalable elastic net subspace clustering. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 3928–3937, 2016.
- [57] A. Bemporad. A piecewise linear regression and classification algorithm with application to learning and model predictive control of hybrid systems. *IEEE Transactions on Automatic Control*, 2022.