Learning stability guarantees for constrained switching linear systems from noisy observations*

Adrien Banse[†]   Zheming Wang   Raphaël M. Jungers[‡§]

February 10, 2023

Abstract

We present a data-driven framework based on Lyapunov theory to provide stability guarantees for a family of hybrid systems. In particular, we are interested in the asymptotic stability of switching linear systems whose switching sequence is constrained by labeled graphs, namely constrained switching linear systems. In order to do so, we provide chance-constrained bounds on stability guarantees, that can be obtained from a finite number of noisy observations.

We first present a method providing stability guarantees from sampled trajectories in the hybrid state-space of the system. We then study the harder situation where one only observes the continuous part of the hybrid states. We show that in this case, one may still obtain formal chance-constrained stability guarantees. For this latter result we provide a new upper bound of general interest, also for model-based stability analysis.

Keywords: Data-driven stability analysis; Switched systems; Graph theory; Chance-constrained optimization.

1 Introduction

Recently, industry, society and technology have increasingly relied on complex systems. For example, *Cyber-Physical Systems* (or *CPS* for short) are dynamical systems in which there are constant interactions between physical and digital, computational devices [1]. Such interactions lead to hybrid dynamics, where discrete and continuous dynamical behaviours are intertwined, placing the analysis of *hybrid systems* at the center of a great research effort [2, 3].

Switching systems are dynamical systems in which a *switching rule* plays a non-trivial role [4–6]. These systems consist of an important family of hybrid systems, that often arise in CPS applications [2]. In particular, switching *linear* systems have been extensively studied recently. Indeed, many model-based techniques have been developed in the past years to analyze their stability [6–9]. Some of these tools have been extended to more general models, in which the switching rule is constrained by a labeled graph. These systems, on which we focus in this paper, are called *constrained switching linear systems* (or *CSLS* for short) and have an extensive range of applications, such as Networked Control Systems [10, 11], or coordination of

multi-agents systems [12]. In this context, the *constrained joint spectral radius* (*CJSR* for short), as well as new families of Lyapunov functions have been introduced to take into account these constraints [13–15].

In many practical applications, the engineer may not have access to the model of the dynamical system. This could occur for multiple reasons, such as the intrinsic complexity of the model, or for security purposes. In such cases, one cannot rely on model-based methods such as those cited above, but rather has to analyze the underlying system in a data-driven fashion. Moreover, the central tool to study stability of switching system, the joint spectral radius, is known to be NP-hard to compute, and without any algebraic solution [16]. For these reasons, novel data-driven techniques have been recently developed for switching linear systems, but where the switching rule is supposed to be arbitrary. Some of them provide chance-constrained bounds on the joint spectral radius [17–19], using *scenario optimization* as their main tool [20]. Other methods extend classical tools from machine learning such as k-means clustering [21]. Recently, [22] also provides formal guarantees on switching linear system identification based on statistical learning theory. In this paper, we present a method based on the scenario optimization to compute an approximate upper bound on the CJSR. Our approach aims at providing a data-driven method for analyzing an unknown system, but could also be used as a randomized analysis for (model-based) computation of the CJSR of a known system.

Contributions In this paper, we provide a data-driven method to assess the quality of approximation of the CJSR of a CSLS based on a finite set of observations of trajectories, thereby generalizing [17–19] to the constrained case and therefore taking into account hybrid state-spaces, which consists in a supplementary technical challenge. We investigate several different types of problems, depending on the information available on the trajectories. Preliminary versions of our results are presented in [23,24], sometimes without proofs, or using a different technique. On top of providing complete proofs our contribution is threefold compared with the latter references. First, we present a unified framework for the different types of bounds, using chance-constrained results on quasi-linear programs [18]. Second, we extend our methods to multiple steps and noisy observations, whereas 1-step observations are assumed in [23], and noise-free observations are assumed in [23,24]. In particular, our method allows to measure the impact of the noise on the approximation of the CJSR, which represents an additional motivation for our data-driven approach compared to the references cited above. Moreover, our results do not depend on the distribution of the noise, which is an asset for practical applications. Finally, we demonstrate our methods on a new numerical example, corresponding to a real-life application.

Outline The rest of this paper is organized as follows. In Section 2, we first define model-based tools for studying stability of CSLS, and then we present chance-constrained optimization results. In Section 3, we provide a probabilistic upper bound on the value of the CJSR, assuming access to a finite number of hybrid states. In Section 4, we show that we are still able to provide a probabilistic upper bound on the CJSR assuming that we only have access to less information, that is only continuous states. Finally, results are illustrated in Section 5 on a numerical example corresponding to a real application, namely Network Control Systems.

Notations In this work, $\mathbb{R}$ is the set of real numbers, and $\mathbb{Z}$ the set of integers. In particular, $\mathbb{Z}_{>0}$ denotes the strictly positive integers. $\mathbb{R}_{>0}$ and $\mathbb{R}_{\geq 0}$ respectively denote the set of positive and non-negative real numbers. For the sake of readability, for any $m \in \mathbb{Z}_{>0}$ we define $[m] = \{1, \ldots, m\}$. The set of real vectors of dimension n is noted $\mathbb{R}^n$. The set of real matrices of dimension $n \times n$ is noted $\mathbb{R}^{n \times n}$, and the set of symmetric positive definite matrices is denoted by $\mathcal{S}^n$. For any real matrix $A \in \mathbb{R}^{n \times n}$, $\lambda^{\max}(A)$ and $\lambda^{\min}(A)$ are respectively the maximal and minimal eigenvalues of A, and $\det(A)$ is its determinant. In terms of norms, $\|\cdot\|$ denotes the Euclidian norm, and $\|\cdot\|_F$ the Frobenius norm. The set $\mathbb{S}^n$ is the unit sphere of dimension n, that is $\{x \in \mathbb{R}^n \,|\, \|x\| = 1\}$, $\mathbb{B}^n$ is the unit ball of dimension n, that is $\{x \in \mathbb{R}^n \,|\, \|x\| \leq 1\}$, and, for any $W \in \mathbb{R}_{>0}$, $B^n[W]$ is the ball of radius W, that is $\{x \in \mathbb{R}^n \,|\, \|x\| \leq W\}$. Finally, for $\Omega \subset \mathbb{R}^n$ and $A \in \mathbb{R}^{n \times n}$, the image of Ω by applying A is noted $A\Omega = \{Ax \,|\, x \in \Omega\}$, the projection of Ω on $\mathbb{S}^n$ is noted $\Pi_{\mathbb{S}^n}(\Omega)$ and $\text{conv}(\Omega)$ stands for the convex hull of Ω.

2 Preliminaries

In this section, we recall some concepts needed to present our methods. We first provide a formal definition of CSLS. Then, we recall some principles of data-driven analysis of these systems from [13,15]. In particular, we define the notion of asymptotic stability, the property that we investigate for such dynamical systems, as well as a Lyapunov framework to study it.

First, we formally define the labeled graph.

Definition 1 (Labeled graph, [25, Definition 2.2.1 and 3.1.1]). A *graph* G is composed of a finite set of *nodes* $\mathcal{U}(G)$ and a finite set of *edges* $\mathcal{E}(G)$. Each edge $e \in \mathcal{E}(G)$ *starts* at $s(e) \in \mathcal{U}(G)$, and *terminates* at $t(e) \in \mathcal{U}(G)$. For a given $m \in \mathbb{Z}_{>0}$, a *labeled graph* $\mathcal{G}$ is a pair (G, σ) where G is a graph and $\sigma : \mathcal{E}(G) \rightarrow [m]$ is the *labeling* and assigns to each edge $e \in \mathcal{E}(G)$ a label $\sigma(e)$ in the finite set $[m]$.

Note that for a given labeled graph $\mathcal{G} = (G, \sigma)$, $\mathcal{U}(G)$ and $\mathcal{E}(G)$ are respectively noted $\mathcal{U}$ and $\mathcal{E}$ when it is clear from the context. Starting from a labeled graph $\mathcal{G} = (G, \sigma)$, and for $l \in \mathbb{Z}_{>0}$, the *language of $\mathcal{G}$ restricted to length l*, noted $\mathcal{L}_{\mathcal{G},l}$ is the set of all possible sequences of labels of length l in the labeled graph $\mathcal{G}$. It is defined as

$$\mathcal{L}_{\mathcal{G},l} = \{(\sigma(e_0), \dots, \sigma(e_{l-1})) \,|\, \forall e_0 \in \mathcal{E}, i \in [l-2] : t(e_i) = s(e_{i+1})\}. \tag{1}$$

The *language* of the labeled graph $\mathcal{G}$, noted $\mathcal{L}_{\mathcal{G}}$, is the set of all possible sequences in the labeled graph $\mathcal{G}$, that is $\mathcal{L}_{\mathcal{G}} = \bigcup_{l \in \mathbb{Z}_{>0}}^{\infty} \mathcal{L}_{\mathcal{G},l}$. An illustration of a labeled graph, is given in Figure 1.
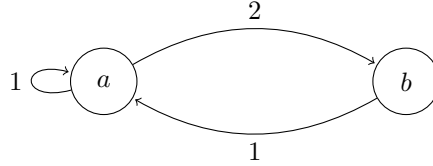

Figure 1: Illustration of a labeled graph $\mathcal{G} = (G, \sigma)$, where $\mathcal{U}(G) = \{a, b\}$, and where $\mathcal{E}(G) = \{e_1, e_2, e_3\}$ such that $s(e_1) = a$, $t(e_1) = a$, $\sigma(e_1) = 1$, $s(e_2) = a$, $t(e_2) = b$, $\sigma(e_2) = 2$, $s(e_3) = b$, $t(e_3) = a$, $\sigma(e_3) = 1$. The language $\mathcal{L}_{\mathcal{G}}$ contains all the words, excepted those that have two consecutive occurences of "2" in them.

In this work, we are interested in a particular subtype of discrete-time hybrid systems, namely constrained switching systems. A hybrid system is a dynamical system whose dynamics is characterized by both continuous and discrete behaviors [1,3]. Starting from $m \in \mathbb{Z}_{>0}$, consider a labeled graph $\mathcal{G} = (G, \sigma)$ where we recall that the labeling is defined as $\sigma : \mathcal{E}(G) \rightarrow [m]$. Consider also a finite set of matrices $\mathcal{A} = \{A_1, \dots, A_m\}$. A *constrained switching linear system* (or *CSLS* for short) is defined by states of the form (e, x) taking values in the hybrid *state-space* $\mathcal{E}(G) \times \mathbb{R}^n$. The dynamics of a CSLS is given by the following hybrid relations:

$$\begin{aligned} e(k+1) &\in \{e \in \mathcal{E}(G) \,|\, s(e) = t(e(k))\} && \text{(discrete behavior)}, \\ x(k+1) &= A_{\sigma(e(k+1))} x(k) && \text{(continuous behavior)}, \end{aligned} \tag{2}$$

with $e(0) = e_0 \in \mathcal{E}$ and $x(0) = x_0 \in \mathbb{R}^n$.

A CSLS is thus entirely characterized by a labeled graph $\mathcal{G} = (G, \sigma)$, and a finite set of matrices $\mathcal{A}$. It is therefore naturally noted $S(\mathcal{G}, \mathcal{A})$.

Remark 1. An *arbitrary switching linear system* (*ASLS* for short), i.e. a switching system whose switching rule is unconstrained, can be seen as a special case of a CSLS. An ASLS defined by a set of matrices $\mathcal{A} = \{A_1, \dots, A_m\}$ is a CSLS $S(\mathcal{F}_m, \mathcal{A})$ where $\mathcal{F}_m = (F_m, \sigma)$ with F_m a directed graph only composed of one node and m self-loops $e_1, \dots, e_m$, and σ such that, for $i = 1, \dots, m$, $\sigma(e_i) = i$. An illustration of $\mathcal{F}_m$ is provided in Figure 2.
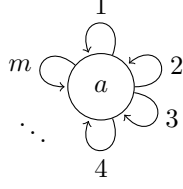
3

Figure 2: Illustration of the labeled graph $\mathcal{F}_m = (F_m, \sigma)$. The language $\mathcal{L}_{\mathcal{F}_m}$ contains all words composed of $1, 2, \ldots, m$.

In this work, we are interested in the asymptotic stability of such hybrid systems. A hybrid system is *asymptotically stable* (or stable, for simplicity) if, for any $(e_0, x_0) \in \mathcal{E} \times \mathbb{R}$ such that $e(0) = e_0$ and $x(0) = x_0$,

$$\lim_{k \to \infty} \|x(k)\| = 0. \tag{3}$$

To study stability, the *constrained joint spectral radius* (*CJSR* for short), has been introduced in [13] to generalize the notion of *joint spectral radius* [6] to the constrained case.

Definition 2 (Constrained joint spectral radius, [13])**.** Given a labeled graph $\mathcal{G}$ and a set of matrices $\mathcal{A}$, the *constrained joint spectral radius* of the CSLS $S(\mathcal{G}, \mathcal{A})$, noted $\rho(\mathcal{G}, \mathcal{A})$, is defined as

$$\rho(\mathcal{G}, \mathcal{A}) = \lim_{k \to \infty} \max_{(\sigma(0),\ldots,\sigma(t-1)) \in \mathcal{L}_{\mathcal{G}, k}} \left\| A_{\sigma(k-1)} \ldots A_{\sigma(0)} \right\|^{1/k}. \tag{4}$$

As the following proposition shows, the CJSR characterizes the stability of a CSLS.

Proposition 1 ([13, Corollary 2.8])**.** *Given a labeled graph $\mathcal{G}$ and a set of matrices $\mathcal{A}$, the CSLS $S(\mathcal{G}, \mathcal{A})$ is stable if and only if $\rho(\mathcal{G}, \mathcal{A}) < 1$.*

Using a slightly different formalism than in [15][1], for some $\gamma \in \mathbb{R}_{>0}$, we now introduce the notion of Lyapunov function for CSLS.

Definition 3 (Lyapunov function)**.** Given a labeled graph $\mathcal{G} = (G, \sigma)$ and a finite set of matrices $\mathcal{A} = \{A_1, \ldots, A_m\}$, a *Lyapunov function* for a CSLS $S(\mathcal{G}, \mathcal{A})$ is a function $V : \mathcal{U} \times \mathbb{R}^n \to \mathbb{R}_{\geq 0}$ such that, for all $u \in \mathcal{U}$, the function $V(u, \cdot)$ is a *norm*[2], and for all $x \in \mathbb{R}^n$, $e \in \mathcal{E}$, there exists $\gamma \in \mathbb{R}_{>0}$ such that

$$V\left(t(e), A_{\sigma(e)}x\right) \leq \gamma V(s(e), x). \tag{5}$$

The quantity $\gamma(V)$ associated to the Lyapunov function V is defined as

$$\gamma(V) = \min_{\gamma}\{\gamma \,|\, \forall e \in \mathcal{E}, x \in \mathbb{R}^n : V\left(t(e), A_{\sigma(e)}x\right) \leq \gamma V(s(e), x)\}. \tag{6}$$

It has been proven in [15, Proposition 2.2] that the CJSR of a given CSLS is also defined as follows:

$$\rho(\mathcal{G}, \mathcal{A}) = \inf_{V}\{\gamma(V) \,|\, V \text{ is a Lyapunov function for } S(\mathcal{G}, \mathcal{A})\}. \tag{7}$$

In this work, we restrict ourselves to the family of quadratic Lyapunov functions.

Definition 4 (Quadratic Lyapunov function)**.** Given a labeled graph $\mathcal{G} = (G, \sigma)$ and a finite set of matrices $\mathcal{A} = \{A_1, \ldots, A_m\}$, a *quadratic Lyapunov function* is a Lyapunov function for the CSLS such that, for all $u \in \mathcal{U}$, $x \in \mathbb{R}^n$, $V(x, u) = \|x\|_{P_u} = \sqrt{x^T P_u x}$ for some $P_u \in \mathcal{S}^n$.

---

[1]Definition 3 is equivalent to the definition of *multinorms* in [15]. We prefer the more general term *Lyapunov function*, that we define in the hybrid framework in Definition 3, and that we generalize by a slight abuse of language to values of γ that may be greater or equal to 1.

[2]In the sense that it is *subadditive*, *absolutely homogeneous* and *positive definite*.

4

Let us now define the following quantity:

$$\gamma^*(\mathcal{G}, \mathcal{A}) = \inf_V \{\gamma(V) \, | \, V \text{ is a quadratic Lyapunov function for } S(\mathcal{G}, \mathcal{A})\}. \tag{8}$$

Since the family of quadratic Lyapunov functions is a subset of the family of all Lyapunov functions, the latter quantity is an upper bound on the CJSR [15, Theorem 3.1]:

$$\rho(\mathcal{G}, \mathcal{A}) \leq \gamma^*(\mathcal{G}, \mathcal{A}). \tag{9}$$

We now introduce the notion of *l-lifting* of a CSLS. This notion and its link to the CJSR have been introduced in [15, Section 3.1].

Definition 5 (*l*-product lifting, [15]). Given a labeled graph $\mathcal{G} = (G, \sigma)$, a finite set of matrices $\mathcal{A} = \{A_1, \ldots, A_m\}$, and a length $l \in \mathbb{N}_{>0}$, the *l-product lift* of $S(\mathcal{G}, \mathcal{A})$ is a CSLS $S(\mathcal{G}^{(l)}, \mathcal{A}^{(l)})$, with

- $\mathcal{G}^{(l)} = (G^{(l)}, \sigma^{(l)})$, where $\mathcal{U}(G^{(l)}) = \mathcal{U}(G)$, and $e' \in \mathcal{E}(G^{(l)})$ if there exists a path $(e_0, \ldots, e_{l-1})$ in G with $s(e') = e_0$ and $t(e') = e_{l-1}$, in which case

$$\sigma^{(l)}(e') = (\sigma(e_0), \ldots, \sigma(e_{l-1})) \in \mathcal{L}_{\mathcal{G}, l}. \tag{10}$$

- $\mathcal{A}_{\mathcal{G}}^{(l)} = \{A_{\sigma(e_{l-1})} \ldots A_{\mathcal{L}(e_0)} \, | \, (\sigma(e_0), \ldots, \sigma(e_{l-1}) \in \mathcal{L}_{\mathcal{G}, l}\}$. For any $e' \in \mathcal{E}(G^{(l)})$ corresponding to the initial path $(e_0, \ldots, e_{l-1})$, the matrix $A_{\sigma^{(l)}(e')} = A_{\sigma(e_{l-1})} \ldots A_{\sigma(e_0)}$.

The *l*-lifting $S(\mathcal{G}^{(l)}, \mathcal{A}_{\mathcal{G}}^{(l)})$ of any CSLS $S(\mathcal{G}, \mathcal{A})$ must be interpreted in the following way: for some $l, N \in \mathbb{N}_{>0}$, if $(x_0, e_0), (x_1, e_1), (x_2, e_2), \ldots, (x_{lN}, e_{lN})$ is a trajectory of $S(\mathcal{G}, \mathcal{A})$, then $(x_0, e_0), (x_l, e_l), (x_{2l}, e_{2l}), \ldots, (x_{lN}, e_{lN})$ is a trajectory of $S(\mathcal{G}^{(l)}, \mathcal{A}_{\mathcal{G}}^{(l)})$. An illustration of the labeled graph $\mathcal{G}^{(l)}$ is given in Figure 3.
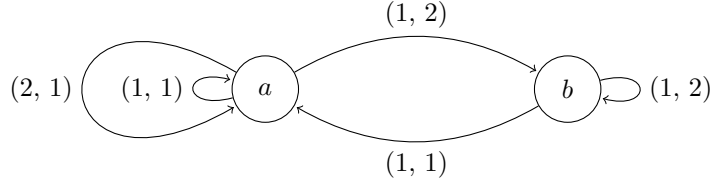

Figure 3: Illustration of $\mathcal{G}^{(2)}$, the 2-lifting labeled graph of $\mathcal{G}$, illustrated in Figure 1.

Using this lift for $l > 1$ allows to refine the upper bound on the CJSR obtained with quadratic Lyapunov functions, compared to the classical case $l = 1$. Indeed, it has been proven in [15, Theorem 3.2] that

$$\rho(\mathcal{G}, \mathcal{A}) \leq \gamma^*(\mathcal{G}^{(l)}, \mathcal{A}_{\mathcal{G}}^{(l)})^{\frac{1}{l}} \leq \rho(\mathcal{G}, \mathcal{A})^{\frac{1}{nl}}, \tag{11}$$

where n is the dimension of the system.

In the following sections, we will show how to construct a quadratic Lyapunov function from a finite amount of data. Using the notion of *l*-lifting, we will present a general method assuming *l*-steps observations. This will then give us an approximate upper bound on the value of the CJSR of the unknown CSLS, extending geometrical properties used in [17].

3 Learning stability guarantees from observations of hybrid states

In this section, we consider a CSLS $S(\mathcal{G}, \mathcal{A})$, where the labeled graph $\mathcal{G}$ and the set of matrices $\mathcal{A}$ are supposed to be unknown. Similarly to [17, 18] in the unconstrained case, we will approximate the CJSR of this system from a finite number of observations of the hybrid state-space of the CSLS. Finally, this will allow for deriving probabilistic stability guarantees for the unknown system.

5

3.1 Problem formulation

For some unknown labeled graph $\mathcal{G} = (G, \sigma)$ and set of matrices $\mathcal{A} = \{A_1, \ldots, A_m\}$, consider the CSLS $S(\mathcal{G}, \mathcal{A})$, for which one can only sample a finite set of N observations of length l. Consider its l-product lift $S(\mathcal{G}^{(l)}, \mathcal{A}_{\mathcal{G}}^{(l)})$, where $\mathcal{G}^{(l)} = (G^{(l)}, \sigma^{(l)})$ according to Definition 5. A *sample of length l* is defined as the triplet (x, e, w), where

- $x \in \mathbb{S}^n$ is an initial continuous state: $x(0) = x$ in $(2)^3$. It is sampled according to the uniform probability measure μ^x on $\mathbb{S}^n$.

- $e \in \mathcal{E}(G^{(l)})$ is an edge of the l-product lift. It is assumed to be sampled according to the uniform probability measure μ_l^e on $\mathcal{E}(G^{(l)})$.

- $w \in B^n[W]$ is a noise on the continuous state-space, where $W > 0$ is a known upper bound. It is sampled according to some probability measure μ^w on $B^n[W]$.

The *sample set* is therefore defined as

$$\omega_N = \{(x_i, e_i, w_i)\}_{i=1,\ldots,N} \subset \Delta, \tag{12}$$

where $\Delta = \mathbb{S}^n \times \mathcal{E}(G^{(l)}) \times B^n[W]$. Each element $(x_i, e_i, w_i) \in \omega_N$ is sampled independently of the others and according to the product probability measure $\pi_l = \mu^x \times \mu_l^e \times \mu^w$.

Remark 2. In the context of this section, we define an *observation* as a hybrid state $(x, u) \in \mathbb{R}^n \times \mathcal{U}(G)$, that is a couple of continuous state and node. The link between the sample (x_i, e_i, w_i) and the pair of observations $(x_i, u_i), (y_i, v_i)$, is that $y_i = A_{\sigma^{(l)}(e_i)} x_i + w_i$, where $s(e_i) = u_i$, and $t(e_i) = v_i$. It is important to highlight that, even though the experiments are defined by the samples (x_i, e_i, w_i), one only observes the pairs of hybrid states $(x_i, u_i), (y_i, v_i)$ as described above.

Our main problem is stated as follows.

Problem 1. Consider an unknown CSLS $S(\mathcal{G}, \mathcal{A})$, where $\mathcal{G} = (G, \sigma)$. For a fixed length $l \in \mathbb{Z}_{>0}$, assume that one is given N pairs of observations, as described in Remark 2. For some user-defined confidence level of at least $\beta \in (0, 1)$, and, knowing $|\mathcal{U}(G)|$, the number of nodes, as well as $|\mathcal{E}(G^{(l)})|$, the number of edges of the l-lifting, provide an upper bound $\rho(\mathcal{G}, \mathcal{A}) \leq \bar{\rho}$.

Remark 3. Although the uniform setting was chosen in the context of this work, we attract the attention of the reader on the fact that more general settings can be considered. For example, one can assume that the distribution on the edges is unknown, and that only an upper bound on the number of nodes and edges is known [24]. One can also assume that labels are sampled instead of edges [23]. However, because the latter introduce conservativism, we choose not to present the last-mentioned frameworks. This remark holds for the results of Section 4 as well.

3.2 Results

In this section, we propose a solution for Problem 1. In order to do so, following the lifting result (11), we will approximate a quadratic Lyapunov function for the unknown l-product lift $S\left(\mathcal{G}^{(l)}, \mathcal{A}^{(l)}\right)$. Let us thus introduce the following quadratic *sampled program*[4]:

$$\mathcal{P}(\omega_N): \min_{\substack{\lambda \in \mathbb{R}_{\geq 0} \\ \forall u \in \mathcal{U}: P_u \in \mathbb{R}^{n \times n}}} \left(\lambda, c\left(P_1, \ldots, P_{|\mathcal{U}|}\right)\right) \tag{13a}$$

$$\text{s.t. } \forall u \in \mathcal{U}: P_u \in \mathcal{X} = \{P \in \mathbb{R}^{n \times n} \mid P \succeq I, \|P\|_F \leq C\}, \tag{13b}$$

$$\forall (x, e, w) \in \omega_N: \left\|A_{\sigma^{(l)}(e)} x + w\right\|_{P_{t(e)}} \leq \lambda^l \|x\|_{P_{s(e)}}, \tag{13c}$$

---

[3]We can restrict ourselves to the unit sphere without loss of generality thanks to *homogeneity* of the dynamical system: for all $\mu > 0$ and $A \in \mathcal{A}$, $A(\mu x) = \mu A x$.

[4]In equation (13), $(\lambda, c(P_1, \ldots, P_{\mathcal{U}}))$ means that we optimize the objective in the *lexicographical order*. That is, $(\lambda_1, c_1) > (\lambda_2, c_2)$ if $\lambda_1 > \lambda_2$, or $\lambda_1 = \lambda_2$ and $c_1 > c_2$, and $(\lambda_1, c_1) = (\lambda_2, c_2)$ if $\lambda_1 = \lambda_2$ and $c_1 = c_2$.

where $c : \mathcal{X}^{|\mathcal{U}|} \to \mathbb{R} : (P_1, \ldots, P_{|\mathcal{U}|}) \mapsto \max(\|P_1\|_F, \ldots, \|P_{|\mathcal{U}|}\|_F)$, and C is a large predefined value (say, $C = 10^6$).

We denote the optimal solutions of the sampled program (13) by $\lambda^*(\omega_N)$ and $\{P_u^*(\omega_N)\}_{u \in \mathcal{U}}$. This program is a data-driven version of the approximation of the CJSR with a quadratic Lyapunov function, as described in (8), and using l-product liftings, as shown in Equation (11). In addition to generalizing data-driven programs expressed in [17,18] to the constrained case, our formulation takes into account noise, as one can see in constraints (13c).

We now define the concept of *non-degenerate sample set* for a sampled program.

Definition 6 (Non-degeneracy, [26, Definition 2.7]). The sample set $\omega_N = \{\delta_1, \ldots, \delta_N\} \subset \Delta$ is *non-degenerate* if there exists a unique set $J \subset [N]$ for which $\beta = \{\delta_j\}_{j \in J}$ is such that $(\lambda^*(\beta), c^*(\beta)) = (\lambda^*(\omega_N), c^*(\omega_N))$.

Given the following assumption, we show that all samples sets considered in this work are non-degenerate.

Assumption 1. All sets $S \subseteq B^n[W]$ such that $\mu^w(S) > 0$ have non-zero Lebesgue measure.

Proposition 2. *Let ω_N be a sample set as described in* (12). *Let Assumption 1 hold, then ω_N is non-degenerate with probability 1.*

A proof of Proposition 2 is provided in A. For now, we consider that Assumption 1 holds, which is a reasonable assumption on the sampling, and therefore that the sample set is always almost-surely non-degenerate. Now, using [18, Theorem 6], we provide a chance-constrained result for the CJSR approximation problem in the presence of noise.

Proposition 3. *Consider an unknown CSLS $S(\mathcal{G}, \mathcal{A})$. For any given number of samples $N \in \mathbb{Z}_{>0}$ such that $N \geq d = |\mathcal{U}|n(n+1)/2$, and for $l \in \mathbb{N}_{>0}$, let $\omega_N \subset \Delta$ be the sample set as described in* (12). *Given $\lambda^*(\omega_N)$ and $\{P_u^*(\omega_N)\}_{u \in \mathcal{U}}$ the solutions of the sampled program* (13), *let*

$$V(\omega_N) = \left\{ (x, e, w) \in \Delta \mid \left\| A_{\sigma^{(l)}(e)} x \right\|_{P_{t(e)}^*(\omega_N)} > \overline{\lambda}_e(\omega_N) \|x\|_{P_{s(e)}^*(\omega_N)} \right\}, \tag{14}$$

where

$$\overline{\lambda}_e(\omega_N) = \lambda^*(\omega_N)^l + \sqrt{\frac{\lambda^{\max}\left(P_{t(e)}^*(\omega_N)\right)}{\lambda^{\min}\left(P_{s(e)}^*(\omega_N)\right)}} W. \tag{15}$$

Then, for any confidence level $\beta \in (0, 1)$, the following holds:

$$\pi_l^N \left(\{ \omega_N \subset \Delta \mid \pi_l \left(V(\omega_N) \right) \leq \varepsilon(\beta, N) \} \right) \geq 1 - \beta, \tag{16}$$

with

$$\varepsilon(\beta, N) = \Phi^{-1}(1 - \beta; d - 1, N). \tag{17}$$

where $\Phi(\cdot; d - 1, N)$ is the regularized incomplete beta function[5]

Proof. For the sake of readability, let $\lambda = \lambda^*(\omega_N)$, and, for $e \in \mathcal{E}$, let $A_e = A_{\sigma^{(l)}(e)}$, $\overline{\lambda}_e = \overline{\lambda}_e(\omega_N)$, $P_{t(e)} = P_{t(e)}^*(\omega_N)$ and $P_{s(e)} = P_{s(e)}^*(\omega_N)$.

(i) One can see that program (13) is the *sampled form* of a *quasi-linear program*, as defined in [18]. Indeed, c is strongly convex, the admissible set for the variables is $\mathcal{X}^{|\mathcal{U}|}$, which is compact, and ω_N is a finite subset of Δ. Now, the set

$$\tilde{V}(\omega_N) = \left\{ (x, e, w) \in \Delta \mid \|A_e x + w\|_{P_{t(e)}} > \lambda^l \|x\|_{P_{s(e)}} \right\} \tag{18}$$

---

[5]See e.g. [27, Equation (6.6.2)].

is the set of violated constraints in the quasi-linear program. Following [18, Theorem 6], since $N \geq d$, and that ω_N is almost surely non-degenerate following Proposition 2, then

$$\pi_l^N \left(\left\{ \omega_N \subset \Delta \mid \pi_l \left(\tilde{V}(\omega_N) \right) \leq \varepsilon(\beta, N) \right\} \right) \geq 1 - \beta \tag{19}$$

holds.

(ii) Second, we prove that $V(\omega_N) \subseteq \tilde{V}(\omega_N)$. Let $(x, e, w) \in V(\omega_N)$, then

$$\overline{\lambda}_e \|x\|_{P_{s(e)}} < \|A_e x\|_{P_{t(e)}}. \tag{20}$$

By subadditivity of the norm, it holds that

$$\overline{\lambda}_e \|x\|_{P_{s(e)}} < \|A_e x + w - w\|_{P_{t(e)}} \tag{21}$$

$$\implies \overline{\lambda}_e \|x\|_{P_{s(e)}} < \|A_e x + w\|_{P_{t(e)}} + \|w\|_{P_{t(e)}}. \tag{22}$$

Now, it is a well known fact that, for any vector $v \in \mathbb{R}^n$ and matrix $P \in \mathcal{S}^n$, $\sqrt{\lambda^{\min}(P)}\|v\| \leq \|v\|_P \leq \sqrt{\lambda^{\max}(P)}\|v\|$. Therefore,

$$
\begin{aligned}
\lambda^l \|x\|_{P_{s(e)}} + W & \sqrt{\frac{\lambda^{\max}\left(P_{t(e)}\right)}{\lambda^{\min}\left(P_{s(e)}\right)}} \sqrt{\lambda^{\min}\left(P_{s(e)}\right)} \\
& < \|A_e x + w\|_{P_{t(e)}} + \sqrt{\lambda^{\max}\left(P_{t(e)}\right)} \|w\| \\
& < \|A_e x + w\|_{P_{t(e)}} + \sqrt{\lambda^{\max}\left(P_{t(e)}\right)} W,
\end{aligned}
\tag{23}
$$

which implies that

$$\lambda^l \|x\|_{P_{s(e)}} < \|A_e x + w\|_{P_{t(e)}}. \tag{24}$$

Now, ii implies that $\pi_l(V(\omega_N)) \leq \pi_l(\tilde{V}(\omega_N))$. Therefore, for any $\varepsilon \in (0, 1)$, $\pi_l(\tilde{V}(\omega_N)) \leq \varepsilon$ implies that $\pi_l(V(\omega_N)) \leq \varepsilon$, and, following i, the proof is completed. $\qquad \square$

From there, using geometrical properties of ellipsoids, one can use the measure of violated constraints of the program (13) to derive guarantees on the model-based optimal objective value, that is $\rho(\mathcal{G}, \mathcal{A})$.

Theorem 1. *Consider an unknown CSLS $S(\mathcal{G}, \mathcal{A})$. For any given number of observations $N \in \mathbb{Z}_{>0}$ such that $N \geq d = |\mathcal{U}|n(n+1)/2$, and for $l \in \mathbb{N}_{>0}$, let $\omega_N \subset \Delta$ be the sampled set as described in (12). Given $\lambda^*(\omega_N)$ and $\{P_u^*(\omega_N)\}_{u \in \mathcal{U}}$ the solutions of the sampled program (13), for any confidence level $\beta \in (0,1)$, the following holds:*

$$\pi_l^N \left(\{ \omega_N \subset \Delta \mid \rho(\mathcal{G}, \mathcal{A}) \leq \overline{\rho}(\omega_N) \} \right) \geq 1 - \beta, \tag{25}$$

where

$$\overline{\rho}(\omega_N) = \max_{e \in \{e' \mid (x, e', w) \in \omega_N\}} \left(\frac{\overline{\lambda}_e(\omega_N)}{\delta \left(\varepsilon(\beta, N) |\mathcal{E}(G^{(l)})| \kappa(P_{s(e)}^*(\omega_N))/2 \right)} \right)^{1/l}, \tag{26}$$

with $\varepsilon(\beta, N)$ and $\overline{\lambda}_e(\omega_N)$ are respectively defined in (17) and (15),

$$\kappa(P) = \sqrt{\det(P)/\lambda^{\min}(P)^n}, \tag{27}$$

and, for any $x \in (0, 1)$, $\delta(x)$ is defined as

$$\delta(x) = \sqrt{1 - \Phi^{-1}(2x; (n-1)/2, 1/2)}. \tag{28}$$

Proof. This theorem is a generalization of [17, Corollary 11, Theorem 15] to the constrained case, and with noisy data. For the sake of readibility, for any $e \in \mathcal{E}$, we write $A_e = A_{\sigma^{(l)}(e)}$, $P_{s(e)} = P^*_{s(e)}(\omega_N)$, $P_{t(e)} = P^*_{t(e)}(\omega_N)$ and $\overline{\lambda}_e(\omega_N) = \overline{\lambda}_e$.

(i) First, we show that, with probability at least $1 - \beta$, there exists a set $\Omega \subset \mathbb{S}^n$ such that

$$\forall e \in \mathcal{E}, \forall x \in \mathbb{S}^n \setminus \Omega : x^T A_e P_{t(e)} A_e x \le \overline{\lambda}_e^2 x^T P_{s(e)} x, \tag{29}$$

and that satisfies

$$\mu^x(\Omega) \le \varepsilon(\beta, N)|\mathcal{E}|. \tag{30}$$

Condition (29) implies that

$$\Omega = \{x \in \mathbb{S}^n \,|\, \exists e \in \mathcal{E} : \|A_e x\|_{P_{t(e)}} > \overline{\lambda}_e \|x\|_{P_{s(e)}} \}. \tag{31}$$

Therefore, since μ_l^e is uniform on $\mathcal{E}$, $\mu^x(\Omega) \le \pi_l(V(\omega_N))|\mathcal{E}|$. Finally, following Proposition 3, (30) holds with probability at least $1 - \beta$.

(ii) Now, for all $e \in \mathcal{E}$, we perform a change of variable. For any $e \in \mathcal{E}$, consider the *Cholesky* factorization of $P_{s(e)}$ and $P_{t(e)}$, that is $P_{s(e)} = L^T_{s(e)} L_{s(e)}$ and $P_{t(e)} = L^T_{t(e)} L_{t(e)}$. Now, for all $e \in \mathcal{E}$, let $B_e = L_{t(e)} A_e L^{-1}_{s(e)}$. Equation (29) becomes

$$\forall e \in \mathcal{E}, \forall x \in L_{s(e)}(\mathbb{S}^n \setminus \Omega) : x^T B_e^T B_e x \le \overline{\lambda}_e^2 x^T x$$
$$\iff \forall e \in \mathcal{E}, \forall x \in \Pi_{\mathbb{S}^n}(L_{s(e)}(\mathbb{S}^n \setminus \Omega)) : x^T B_e^T B_e x \le \overline{\lambda}_e^2 x^T x \tag{32}$$
$$\iff \forall e \in \mathcal{E}, \forall x \in \mathbb{S}^n \setminus \Pi_{\mathbb{S}^n}(L_{s(e)}(\Omega)) : x^T B_e^T B_e x \le \overline{\lambda}_e^2 x^T x,$$

where the second equivalence holds by homogeneity of the system. For the sake of readibility, we write $\Omega'_e = \Pi_{\mathbb{S}^n}(L_{s(e)}(\Omega))$.

(iii) We now link $\mu^x(\Omega)$ and $\mu^x(\Omega'_e)$. Using a similar reasoning as in [17, Theorem 15], for all $e \in \mathcal{E}$, the following holds:

$$\mu^x(\Omega'_e) \le \sqrt{\frac{\det(P_{s(e)})}{\lambda^{\min}(P_{s(e)})^n}} \mu^x(\Omega) = \kappa(P_{s(e)})\mu^x(\Omega). \tag{33}$$

This yields

$$\forall e \in \mathcal{E}, \forall x \in \mathbb{S}^n \setminus \Omega'_e : x^T B_e^T B_e x \le \overline{\lambda}_e^2 x^T x, \tag{34}$$

where, for all $e \in \mathcal{E}$, $\mu^x(\Omega'_e) \le \kappa(P_{s(e)})\mu^x(\Omega)$.

(iv) We now look for the largest sphere included in $\mathrm{conv}(\mathbb{S}^n \setminus \Omega'_e)$. The radius of this sphere is denoted by α. First, (34) is equivalent to

$$\forall e \in \mathcal{E} : B_e(\mathbb{S}^n \setminus \Omega'_e) \subset \overline{\lambda}_e \mathbb{B}^n. \tag{35}$$

Now, [17, Property 2] implies that

$$\forall e \in \mathcal{E} : B_e(\mathrm{conv}(\mathbb{S}^n \setminus \Omega'_e)) \subset \mathrm{conv}(B_e(\mathbb{S}^n \setminus \Omega'_e)) \subset \overline{\lambda}_e \mathbb{B}^n. \tag{36}$$

Now, from the definition of α, for all $e \in \mathcal{E}$, we have that $\alpha \mathbb{S}^n \subset \mathrm{conv}(\mathbb{S}^n \setminus \Omega'_e)$, which implies that $B_e(\alpha \mathbb{S}^n) \subset B_e(\mathrm{conv}(\mathbb{S}^n \setminus \Omega'_e)) \subset \overline{\lambda}_e \mathbb{B}^n$. This yields

$$\forall e \in \mathcal{E} : B_e(\mathbb{S}^n) \subset (\overline{\lambda}_e/\alpha)\mathbb{B}^n. \tag{37}$$

Finally, following [17, Proposition 13] and step iii,

$$\alpha \ge \delta(\mu^x(\Omega'_e)/2) \ge \delta(\kappa(P_{s(e)})\mu^x(\Omega)/2). \tag{38}$$
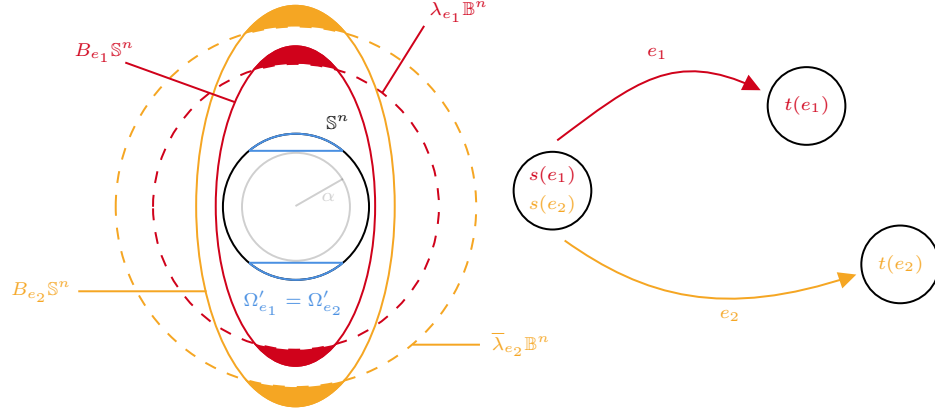
9

Figure 4: Illustration of the proof of Theorem 1. Each node of $\mathcal{G}^{(l)}$ is the starting node of some edge. Suppose a node $s \in \mathcal{U}$ is the starting node of two edges e_1 and e_2. For each edge, there is only one corresponding set $\Omega' = \Omega'_{e_1} = \Omega'_{e_2}$ such that, for $e \in \{e_1, e_2\}$, $B_e(\mathbb{S}^n \setminus \Omega') \subset \overline{\lambda}_e \mathbb{B}^n$. The quantity α is the radius of the largest sphere inscribed in $\operatorname{conv}(\mathbb{S}^n \setminus \Omega')$.

An illustration of Equation (35), as well as the quantity α can be found in Figure 4.

We now summarize. We first define

$$\overline{\rho}_e = \frac{\overline{\lambda}_e}{\delta(\varepsilon(\beta, N)|\mathcal{E}(G^{(l)})|\kappa(P^*_{s(e)}(\omega_N))/2)}. \tag{39}$$

Following steps i, ii, iii, iv, the following holds with probability at least $1 - \beta$:

$$\forall e \in \mathcal{E} : B_e(\mathbb{S}^n) \subset \overline{\rho}_e \mathbb{B}^n. \tag{40}$$

From the definition of B_e, (40) is equivalent to

$$\forall e \in \mathcal{E} : A_e\left(L_{s(e)}^{-1}(\mathbb{S}^n)\right) \subset \overline{\rho}_e\left(L_{t(e)}^{-1}(\mathbb{B}^n)\right), \tag{41}$$

which implies that

$$\forall e \in \mathcal{E} : A_e\left(L_{s(e)}^{-1}(\mathbb{B}^n)\right) \subset \overline{\rho}_e\left(L_{t(e)}^{-1}(\mathbb{B}^n)\right). \tag{42}$$

Since $\overline{\rho}_e \leq \overline{\rho}(\omega_N)^l$ for all $e \in \mathcal{E}$ following step i, the next condition holds with probability at least $1 - \beta$:

$$\forall e \in \mathcal{E} : A_e\left(L_{s(e)}^{-1}(\mathbb{B}^n)\right) \subset \overline{\rho}(\omega_N)^l\left(L_{t(e)}^{-1}(\mathbb{B}^n)\right). \tag{43}$$

Furthermore, since $L_{s(e)}^{-1}(\mathbb{B}^n)$ and $L_{t(e)}^{-1}(\mathbb{B}^n)$ are respectively the one-level set of the quadratic norms $\|\cdot\|_{P_{s(e)}}$ and $\|\cdot\|_{P_{s(e)}}$, condition (43) is equivalent to

$$\forall e \in \mathcal{E}, \forall x \in \mathbb{S}^n : \|A_e x\|_{P_{t(e)}} \leq \overline{\rho}(\omega_N)^l \|x\|_{P_{s(e)}}, \tag{44}$$

which yields that $\gamma^*(\mathcal{G}^{(l)}, \mathcal{A}_{\mathcal{G}}^{(l)}) \leq \overline{\rho}(\omega_N)^l$ following definition (8). Therefore, it holds that

$$\rho(\mathcal{G}, \mathcal{A}) \leq \gamma^*(\mathcal{G}^{(l)}, \mathcal{A}_{\mathcal{G}}^{(l)})^{1/l} \leq \overline{\rho}(\omega_N) \tag{45}$$

with probability at least $1 - \beta$, and the proof is completed. $\qquad\qquad\square$

Remark 4. In parallel to [17, Remark 16], in the proof of Theorem 1, in (33) for all edge $e \in \mathcal{E}$, one can compute a lower bound on $\mu^x(\mathbb{S}^n \setminus \Omega_e')$, the measure of the complement of Ω_e', instead of finding an upper bound on $\mu^x(\Omega_e')$. In this case,

$$\mu^x(\mathbb{S}^n \setminus \Omega_e') \geq \sqrt{\frac{\det(P_{s(e)})}{\lambda^{\max}(P_{s(e)})^n}} \mu^x(\mathbb{S}^n \setminus \Omega) = \sqrt{\frac{\det(P_{s(e)})}{\lambda^{\max}(P_{s(e)})^n}}(1 - \mu^x(\Omega)). \tag{46}$$

As a consequence, one can replace $\varepsilon(\beta, N)|\mathcal{E}(G^{(l)})|\sqrt{\det(P_{s(e)}^*(\omega_N))/\lambda^{\min}(P_{s(e)}^*(\omega_N))^n}$ in the expression of the bound (26) by

$$1 - \left((1 - \varepsilon(\beta, N)|\mathcal{E}(G^{(l)})|)\sqrt{\frac{\det(P_{s(e)}^*(\omega_N))}{\lambda^{\max}(P_{s(e)}^*(\omega_N))^n}} \right), \tag{47}$$

therefore providing an alternative bound. The user can then take the tightest bound as the best probabilistic upper bound on the CJSR.

Theorem 1 provides a probabilistic stability guarantee. For a finite sample set ω_N, and for some confidence level $\beta \in (0, 1)$, if one computes $\overline{\rho}(\omega_N) < 1$, then, following Proposition 1, one has a certificate of stability for the observed CSLS with probability of at least $1 - \beta$. Moreover, it can be proven that, if one assumes no noise on the observations, for l and N going to infinity, the bound (26) is arbitrarily close to the true CJSR. As this result is a consequence of the well known model-based result (11), and follows a similar reasoning as in the proof of [17, Theorem 17], we do not provide it in this paper.

4 Learning stability guarantees by observing continuous states only

In some situations, it can be difficult to observe the discrete part of the dynamics. In this section, we show that one may still be able to derive a probabilistic upper bounds for the underlying CSLS under reasonable assumptions. For this purpose, we first present a model-based result on the CJSR that allows to study the stability of a CSLS via the analysis of an auxiliary arbitrary switching linear system. Then, we formulate the problem of finding a probabilistic upper bound on the CJSR from continuous states. Finally, we provide a solution based on chance-constrained optimization.

4.1 Reducing constrained systems to arbitrary systems

In this subsection we provide a general result, of application in the wider context of model-based analysis of CSLS, which will be useful for our purpose.

Following Remark 1, the quantity $\rho(\mathcal{F}_m, \mathcal{A}_{\mathcal{G}}^{(l)})$ is the (unconstrained) joint spectral radius of the ASLS defined on the (constrained) set of matrices $\mathcal{A}_{\mathcal{G}}^{(l)}$. Proposition 4 states that an upper bound on the initial CJSR can be derived from this quantity. In the next subsection, we will use the continuous data to which we have access to approximate it, and thereby obtain probabilistic stability guarantees.

Proposition 4. *Consider a CSLS $S(\mathcal{G}, \mathcal{A})$. For any length $l \in \mathbb{N}_{>0}$, the following holds:*

$$\rho(\mathcal{G}, \mathcal{A}) \leq \rho(\mathcal{F}_m, \mathcal{A}_{\mathcal{G}}^{(l)})^{1/l}. \tag{48}$$

Moreover, asymptotic equality holds, that is

$$\rho(\mathcal{G}, \mathcal{A}) = \lim_{l \to \infty} \rho(\mathcal{F}_m, \mathcal{A}_{\mathcal{G}}^{(l)})^{1/l}. \tag{49}$$

Proof. First we prove inequality (48). Since for all $t > 0$ and $l > 0$, $\mathcal{A}_{\mathcal{G}}^{(lt)} \subseteq \left(\mathcal{A}_{\mathcal{G}}^{(l)}\right)^t$, Definition 2 and Remark 1 yields

$$
\begin{aligned}
\rho(\mathcal{F}_m, \mathcal{A}_{\mathcal{G}}^{(l)}) &= \lim_{t \to \infty} \max\left\{\|A\|^{1/t} : A \in \left(\mathcal{A}_{\mathcal{G}}^{(l)}\right)^t\right\} \\
&\geq \lim_{t \to \infty} \max\left\{\|A\|^{1/t} : A \in \mathcal{A}_{\mathcal{G}}^{(lt)}\right\}.
\end{aligned}
\tag{50}
$$

Letting $k = lt$ and, again, using Definition 2, the inequality becomes

$$
\begin{aligned}
\rho(\mathcal{F}_m, \mathcal{A}_{\mathcal{G}}^{(l)}) &\geq \lim_{t \to \infty} \max\left\{\|A\|^{l/k} : A \in \mathcal{A}_{\mathcal{G}}^{(k)}\right\} \\
&= \left(\lim_{k \to \infty} \max\left\{\|A\|^{1/k} : A \in \mathcal{A}_{\mathcal{G}}^{(k)}\right\}\right)^l \\
&= \rho(\mathcal{G}, \mathcal{A})^l,
\end{aligned}
\tag{51}
$$

which is the desired result. Now we prove asymptotic equality (49). First, given the description of the CJSR in (7), for any $l > 0$,

$$
\begin{aligned}
\rho(\mathcal{F}_m, \mathcal{A}_{\mathcal{G}}^{(l)}) &\leq \max\left\{\|A\| : A \in \mathcal{A}_{\mathcal{G}}^{(l)}\right\} \\
\iff \rho(\mathcal{F}_m, \mathcal{A}_{\mathcal{G}}^{(l)})^{1/l} &\leq \max\left\{\|A\|^{1/l} : A \in \mathcal{A}_{\mathcal{G}}^{(l)}\right\}.
\end{aligned}
\tag{52}
$$

Taking the limit in both sides, it gives $\lim_{l \to \infty} \rho(\mathcal{F}_m, \mathcal{A}_{\mathcal{G}}^{(l)})^{1/l} \leq \rho(\mathcal{G}, \mathcal{A})$. Combining with (51), it gives the desired result. $\square$

4.2   Problem formulation

In this section, we formulate the problem of finding probabilistic stability guarantees only from continuous information. For some unknown labeled graph $\mathcal{G} = (G, \sigma)$ and set of matrices $\mathcal{A} = \{A_1, \ldots, A_m\}$, consider the CSLS $S(\mathcal{G}, \mathcal{A})$, for which one can sample data containing only information about the continuous states.

More formally, consider the l-product lift of the observed system, that is $S(\mathcal{G}^{(l)}, \mathcal{A}_{\mathcal{G}}^{(l)})$, where $\mathcal{G} = (G^{(l)}, \sigma^{(l)})$ according to Definition 5. One *sample of length l* is defined as the triplet (x, σ, w), where

- $x \in \mathbb{S}^n$ is an initial continuous state: $x(0) = x$ in (2), sampled according to the uniform measure μ^x as described in Section 3.1.

- σ is such that $\sigma = \sigma^{(l)}(e)$, for some unknown edge $e \in \mathcal{E}(G^{(l)})$. σ is sampled according to the probability measure μ_l^σ, which is assumed to be uniform on $\mathcal{L}_{G,l}$.

- $w \in B^n[W]$ is a noise on the continuous state-space, where $W > 0$ is a known upper bound, sampled according to the probability measure μ^w as described in Section 3.1.

In the continuous framework, with $\Delta_{\text{cont}} = \mathbb{S}^n \times [m] \times B^n[W]$, the sample set is therefore defined as

$$
\omega_{N,\text{cont}} = \{(x_i, \sigma_i, w_i)\}_{i=1,\ldots,N} \subset \Delta_{\text{cont}},
\tag{53}
$$

where each element (x_i, σ_i, w_i) is sampled according to the product probability measure $\pi_{l,\text{cont}} = \mu^x \times \mu_l^\sigma \times \mu^w$.

Remark 5. In parallel to Remark 2, but in the framework of this section, we define an *observation* as a continuous state $x \in \mathbb{R}$. The link between the sample (x_i, σ_i, w_i) and the pair of observations x_i, y_i, is that $y_i = A_{\sigma_i} x_i + w_i$.

Our second problem is stated as follows.

Problem 2. Consider an unknown CSLS $S(\mathcal{G}, \mathcal{A})$. For a fixed length $l \in \mathbb{Z}_{>0}$ and number $N \in \mathbb{Z}_{>0}$ of samples, consider the sample set $\omega_{N,\text{cont}}$ as described in (53). For some user-defined confidence level of at least $\beta \in (0, 1)$, and knowing $\mathcal{U}(G)$ and $|\mathcal{L}_{\mathcal{G},l}|$, the number of words of length l generated by $\mathcal{G}$, provide an upper bound $\rho(\mathcal{G}, \mathcal{A}) \leq \bar{\rho}$.

4.3 Results

To answer to Problem 2, we will now approximate a quadratic Lyapunov function for the unknown arbitrary system $S(\mathcal{F}_m, \mathcal{A}_{\mathcal{G}}^{(l)})$ following Proposition 4. In order to do so, let us introduce the following *sampled program*:

$$\mathcal{P}(\omega_{N,\text{cont}}) : \min_{\substack{\lambda \in \mathbb{R}_{\geq 0} \\ P \in \mathbb{R}^{n \times n}}} (\lambda, c(P)) \tag{54a}$$

$$\text{s.t. } P \in \mathcal{X} = \{P \in \mathbb{R}^{n \times n} \,|\, P \succeq I, \|P\|_F \leq C\}, \tag{54b}$$

$$\forall (x, \sigma, w) \in \omega_{N,\text{cont}} : \|A_\sigma x + w\|_P \leq \lambda^l \|x\|_P, \tag{54c}$$

where $c : \mathcal{X} \to \mathbb{R} : P \mapsto \|P\|_F$, and C is sufficiently large. We denote the optimal solutions of the sampled program (13) by $\lambda^*(\omega_{N,\text{cont}})$ and $P^*(\omega_{N,\text{cont}})$.

We first present the following chance-constrained result.

Proposition 5. *Consider an unknown CSLS $S(\mathcal{G}, \mathcal{A})$. For any given number of samples $N \in \mathbb{Z}_{>0}$ such that $N \geq d = n(n+1)/2$, and for $l \in \mathbb{N}_{>0}$, let $\omega_{N,\text{cont}} \subset \Delta_{cont}$ be the sample set as described in (53). Given $\lambda^*(\omega_{N,cont})$ and $P^*(\omega_{N,cont})$ the solutions of the sampled program (54), let*

$$V(\omega_{N,cont}) = \left\{ (x, \sigma, w) \in \Delta_{cont} \,|\, \|A_\sigma x\|_{P^*(\omega_{N,cont})} > \overline{\lambda}(\omega_{N,cont}) \|x\|_{P^*(\omega_{N,cont})} \right\}, \tag{55}$$

where

$$\overline{\lambda}(\omega_{N,cont}) = \lambda^*(\omega_{N,cont})^l + \sqrt{\frac{\lambda^{\max}(P^*(\omega_{N,cont}))}{\lambda^{\min}(P^*(\omega_{N,cont})))}} W. \tag{56}$$

Then, for any confidence level $\beta \in (0,1)$, the following holds:

$$\pi_{l,cont}^N \left(\{\omega_{N,cont} \subset \Delta_{cont} \,|\, \pi_{l,cont} \left(V(\omega_{N,cont}) \right) \leq \varepsilon(\beta, N) \} \right) \geq 1 - \beta, \tag{57}$$

where $\varepsilon(\beta, N)$ is defined in (17).

Proof. For the sake of readability, we note $\lambda = \lambda^*(\omega_{N,\text{cont}})$, $P = P^*(\omega_{N,\text{cont}})$ and $\overline{\lambda} = \overline{\lambda}(\omega_{N,\text{cont}})$. In the same fashion as in the proof of Proposition 3, we first prove that

$$\pi_{l,\text{cont}}^N (\{\omega_{N,\text{cont}} \subset \Delta_{\text{cont}} \,|\, \pi_{l,\text{cont}} \left(\tilde{V}(\omega_{N,\text{cont}}) \right) \leq \varepsilon(\beta, N) \}) \geq 1 - \beta, \tag{58}$$

with

$$\tilde{V}(\omega_{N,\text{cont}}) = \{(x, \sigma, w) \in \Delta_{\text{cont}} \,|\, \|A_\sigma x + w\|_P > \lambda \|x\|_P \}. \tag{59}$$

Again, (58) holds by [18, Theorem 6], since (54) is the sampled form of a quasi-linear program, as defined in [18], and that $\omega_{N,\text{cont}}$ is almost surely non-degenerate following Proposition 2. It remains to prove that $V(\omega_{N,\text{cont}}) \subseteq \tilde{V}(\omega_{N,\text{cont}})$, that is, for any $(x, \sigma, w) \in \Delta_{\text{cont}}$,

$$\|A_\sigma x\|_P > \overline{\lambda} \|A\|_P \implies \|A_\sigma x + w\| > \lambda^l \|A\|_P. \tag{60}$$

The latter holds by applying the exact same argument as in ii in the proof of Proposition 3. $\qquad\square$

Now we present the main theorem of this section by leveraging the model-based result obtained in Proposition 4.

Theorem 2. *Consider an unknown CSLS $S(\mathcal{G}, \mathcal{A})$. For any given number of samples $N \in \mathbb{Z}_{>0}$ such that $N \geq d = n(n+1)/2$, and for $l \in \mathbb{N}_{>0}$, let $\omega_{N,cont} \subset \Delta_{cont}$ be the sample set as described in (53). Given $\lambda^*(\omega_{N,cont})$ and $P^*(\omega_{N,cont})$ the solutions of the sampled program (54), for any confidence level $\beta \in (0,1)$, the following holds:*

$$\pi_{l,cont}^N (\{\omega_{N,cont} \subset \Delta_{cont} \,|\, \rho(\mathcal{G}, \mathcal{A}) \leq \overline{\rho}(\omega_{N,cont}) \}) \geq 1 - \beta, \tag{61}$$

where

$$\overline{\rho}(\omega_{N,cont}) = \left(\frac{\overline{\lambda}(\omega_{N,cont})}{\delta\left(\varepsilon(\beta,N)|\mathcal{L}_{\mathcal{G},l}|\kappa(P^*(\omega_{N,cont}))/2\right)} \right)^{1/l}, \tag{62}$$

with $\varepsilon(\beta,N)$, $\kappa(P^*(\omega_{N,cont}))$, $\delta(\cdot)$ and $\overline{\lambda}(\omega_{N,cont})$ respectively defined in (17), (27), (28) and (56).

Proof. We follow a similar reasoning as in the proof of Theorem 1 where any labeled graph $\mathcal{G}$ is replaced by the labeled graph $\mathcal{F}_m$ as described in Remark 1. We will go through the main steps of the proof of Theorem 1. Again, for the sake of readability, we note $\lambda = \lambda^*(\omega_{N,cont})$, $P = P^*(\omega_{N,cont})$ and $\overline{\lambda} = \overline{\lambda}(\omega_{N,cont})$.

(i) We first show that, with a probability of at least $1 - \beta$, there exists a set $\Omega \subset \mathbb{S}^n$ such that

$$\forall \sigma \in \mathcal{L}_{\mathcal{G},l}, \forall x \in \mathbb{S}^n \setminus \Omega : \|A_\sigma x\|_P \le \overline{\lambda}\|x\|_P, \tag{63}$$

and that satisfies

$$\mu^x(\Omega) \le \varepsilon(\beta,N)|\mathcal{L}_{\mathcal{G},l}|. \tag{64}$$

Again, condition (63) implies that

$$\Omega = \{x \in \mathbb{S}^n \,|\, \exists e \in \mathcal{E} : \|A_e x\|_{P_{t(e)}} > \overline{\lambda}_e \|x\|_{P_{s(e)}}\}. \tag{65}$$

Therefore, since μ_l^σ is uniform on $\mathcal{L}_{\mathcal{G},l}$, $\mu^x(\Omega) \le \pi_{l,cont}(V(\omega_{N,cont}))|\mathcal{L}_{\mathcal{G},l}|$. Finally, following Proposition 5, (64) holds with probability at least $1 - \beta$.

(ii) We now perform only one change of variable. Let $P = L^T L$ be the Cholesky decomposition of P. For all $\sigma \in \mathcal{L}_{\mathcal{G},l}$, let $B_\sigma = L A_\sigma L^T$. For the same reasons as in the proof of Theorem 1, Equation (63) becomes

$$\forall \sigma \in \mathcal{L}_{\mathcal{G},l}, \forall x \in \mathbb{S}^n \setminus \Omega' : x^T B_\sigma^T B_\sigma x \le \overline{\lambda}^2 x^T x, \tag{66}$$

with $\Omega' = \Pi_{\mathbb{S}^n}(L(\Omega))$.

(iii) We now link $\mu^x(\Omega)$ and $\mu^x(\Omega')$. Again, for the same reasons as in the proof of Theorem 1, $\mu^x(\Omega') \le \kappa(P)\mu^x(\Omega)$ holds.

(iv) We now look at the largest sphere included in $\text{conv}(\mathbb{S}^n \setminus \Omega')$. The radius of this sphere is noted α. Again, a similar reasoning as in the proof of Theorem 1 yields $\alpha \ge \delta(\kappa(P)\mu^x(\Omega)/2)$, which yields

$$\forall \sigma \in \mathcal{L}_{\mathcal{G},l} : B_\sigma(\mathbb{S}^n) \subset (\overline{\lambda}/\alpha)\mathbb{B}^n \subset \left(\overline{\rho}(\omega_{N,cont})^l\right)\mathbb{B}^n. \tag{67}$$

Summarizing, (67) is equivalent to

$$\forall \sigma \in \mathcal{L}_{\mathcal{G},l}, \forall x \in \mathbb{S}^n : \|A_\sigma x\|_{P^*(\omega_{N,cont})} \le \overline{\rho}(\omega_{N,cont})^l \|x\|_{P^*(\omega_{N,cont})}, \tag{68}$$

which holds with a probability at least $1 - \beta$. The condition (68) is equivalent to $\gamma^*(\mathcal{F}_m, \mathcal{A}_{\mathcal{G}}^{(l)}) \le \overline{\rho}(\omega_{N,cont})^l$ following definition (8). Therefore, following Proposition 4, one has $\rho(\mathcal{G}, \mathcal{A}) \le \gamma^*(\mathcal{F}_m, \mathcal{A}_{\mathcal{G}}^{(l)})^{1/l} \le \overline{\rho}(\omega_{N,cont})$ with probability at least $1 - \beta$, and the proof is completed. $\square$

Remark 6. In parallel to Remark 4, and for the same reason, one can replace

$$\varepsilon(\beta,N)|\mathcal{L}_{\mathcal{G},l}|\sqrt{\frac{\det(P^*(\omega_N))}{\lambda^{\min}(P^*(\omega_N))^n}} \tag{69}$$

in the expression of the bound (62) by

$$1 - \left((1 - \varepsilon(\beta,N)|\mathcal{L}_{\mathcal{G},l}|)\sqrt{\frac{\det(P^*(\omega_N))}{\lambda^{\max}(P^*(\omega_N))^n}}\right), \tag{70}$$

therefore providing an alternative bound.

Theorem 2 provides a probabilistic stability guarantee in the case where one has access only to continuous states. Indeed in comparaison to Theorem 1, one does not need any information about the starting and terminating nodes for every observation. For a finite sample set $\omega_{N,\text{cont}}$, and for some confidence level $\beta \in (0,1)$, then if one computes $\bar{\rho}(\omega_{N,\text{cont}}) < 1$, then, following Proposition 1, one has a certificate of stability for the observed CSLS with probability of at least $1 - \beta$. Again, it can be proven that, if one assumes no noise on the observations, for l and N going to infinity, the bound (62) is arbitrarily close to the true CJSR, thanks to the asymptotic relation (49). Again we do not explicitly prove this result in this paper.

5   Application to Networked Control Systems

In this section, we illustrate our method on a concrete example, namely Networked Control Systems (or NCS for short)[6]. Networked Control Systems are dynamical systems whose plant is physically separated from its controller, so that the feedback has to go through a network. The failures experienced by the network can typically be modelled by a switching rule. For example, consider a linear dynamical system, whose controller is a linear feedback. Considering that packets can be lost in the network, the dynamics of such system is given by

$$x(k+1) = \begin{cases} (A+BK)x(k) & \text{if the packet is not lost} \\ Ax(k) & \text{otherwise.} \end{cases} \tag{71}$$

An illustration of such dynamical system is given in Figure 5. Now, consider that one has more knowledge
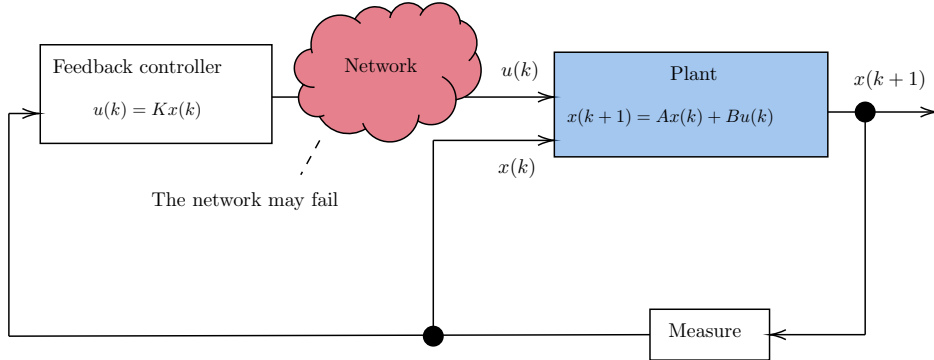

Figure 5: Illustration of a NCS suffering from packet losses.

about the system. For example, consider that packets are never lost more than twice in a row[7]. In this case, this system can be modelled as a CSLS $S(\mathcal{G}, \mathcal{A})$, where $\mathcal{A} = \{A_1, A_2\} = \{A + BK, A\}$ and where $\mathcal{G}$ is as depicted in Figure 6.

More especially, we consider a two-dimensional NCS with a one-dimensional control signal, whose matrices are given as follows:

$$A = \begin{pmatrix} 0.45 & 1.08 \\ 0.36 & 0.09 \end{pmatrix}, B = \begin{pmatrix} 0 \\ 1 \end{pmatrix}, K = \begin{pmatrix} -0.42 & -0.36 \end{pmatrix}. \tag{72}$$

Using model-based methods given in [15], we are able to compute $\rho(\mathcal{G}, \mathcal{A}) \approx 0.70697$.

First, we show the evolution of $\bar{\rho}(\omega_N)$, the bound (26) in case of noisy observations. In order to do so, for an increasing number of observations of length 1, we compute 20 realizations of $\bar{\rho}(\omega_N)$, such that the noises

---

[6]For an introduction, see [28] and references therein.
[7]Although this constrained is quite simple for the sake of the example, the labeled graph framework allows for way more complex constraints.
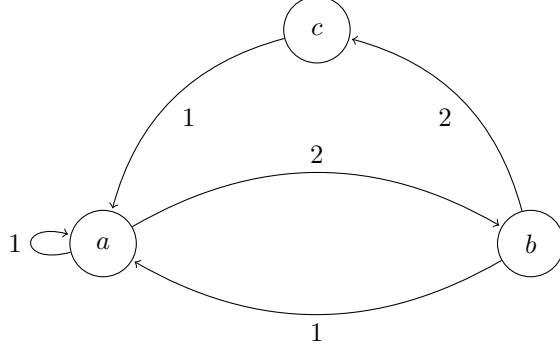
Figure 6: Labeled graph $\mathcal{G} = (G, \sigma)$ such that its language $\mathcal{L}_G$ contains any words except those that have more than two 2 in a row. By using the CSLS $S(\mathcal{G}, \mathcal{A})$ to model Networked Control Systems, it means that the packets cannot be lost more than twice in a row.

w_i are independently and uniformly sampled in $B^n[W]$, for different upper bounds $W \in \{0, 0.01, 0.1\}$[8], and for a confidence level of $\beta = 5\%$. The average is reported in Figure 7. We can observe that, in average, the
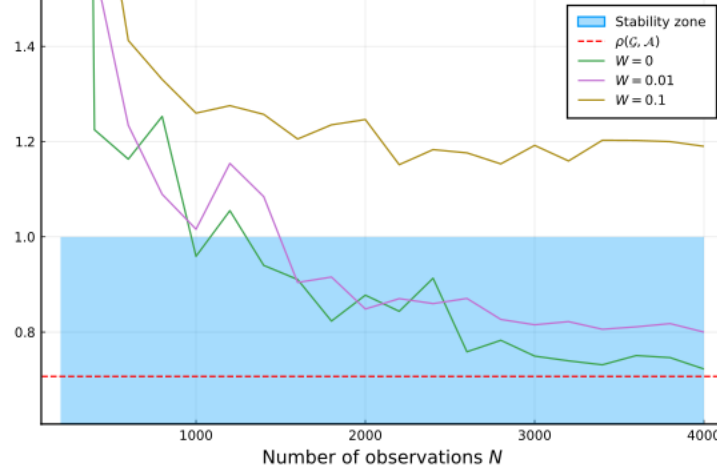

Figure 7: Evolution of the hybrid states bound $\overline{\rho}(\omega_N)$ given in Theorem 1, for different upper bounds on the norm of the noise W. One can observe that a smaller noise allows for a better approximation of the CJSR.

upper bound is still valid, even for noisy observations. In addition, we can see that, for small noises such as $W = 0.01$, our bound is still able to provide stability guarantees with a confidence level of 5%. However, with $W = 0.1$, our bound is unable to provide any guarantee after 4000 observations.

Second, we illustrate that our method is valid for observations of length greater than 1. The evolution of the bound presented in Theorem 1 is presented in Figure 8 for $l \in \{1, 2\}$, and considering no noise. Again, for $\beta = 5\%$, 20 realizations were computed for each considered number of observations, and the average is reported in Figure 8. We also obtain stability guarantees for $l = 2$, although we need slightly more samples.

Finally, we show that we can apply our method even if we do not have access to hybrid states, by applying the method described in Section 4 for a confidence level of $\beta = 5\%$. The evolution of the average of 20 realizations of $\overline{\rho}(\omega_{N,\mathrm{cont}})$, as presented in Theorem 2, is given in Figure 9, for $l \in \{1, 2, 3\}$, where we suppose that the observations are noiseless. The evolution of the bound from hybrid observations is

---

[8]Note that the case $W = 0$ corresponds to a slightly different setting that the one presented in Section 3. However, all results hold in the noise-free setting if the sample set is non-degenerate with probability 1, which one can verifiy in practice.
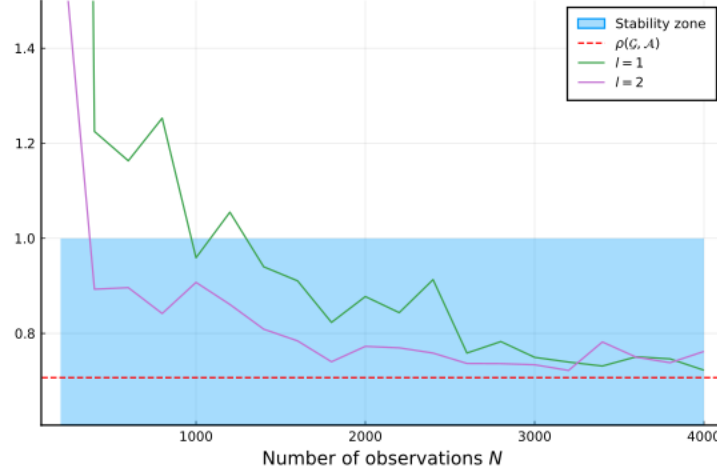
16

Figure 8: Evolution of the hybrid states bound $\overline{\rho}(\omega_N)$ given in Theorem 1, for different lengths of observations l. Even though the bounds evolute differently, they are both providing stability guarantees.

also given as a reference. One can see that the bound computed only from continuous states seems to
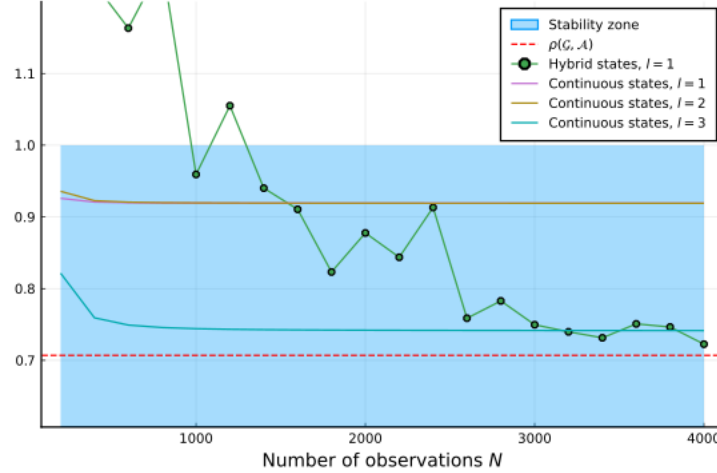

Figure 9: Illustration of the continuous states method. The lines without a marker represent the evolution of $\overline{\rho}(\omega_{N,\text{cont}})$, the bound obtained only from continuous states presented in Theorem 2, for different lengths $l \in \{1, 2, 3\}$. As a reference, the line with a marker is the evolution of the bound obtained from hybrid states, for 1-step observations.

converge to more conservative values that the hybrid method. This is due to the fact that we are actually approximating $\rho(\mathcal{F}_m, \mathcal{A}_{\mathcal{G}}^{(l)})^{1/l}$, which is greater or equal than $\rho(\mathcal{G}, \mathcal{A})$ following Proposition 4. However, in this case, probabilistic stability guarantees were still found for each considered value of l.

6   Conclusions

In this paper, we extended the scope of data-driven stability analysis of hybrid systems. In particular, we considered a more general model of switching linear systems, namely constrained switching linear systems.

Based on white-box stability analysis tools and chance-constrained optimization results, we proposed a method that approximates the CJSR of an observed constrained system in two different frameworks. In the first one, we consider that one has access to hybrid states, that is couples of continuous states and edges in the labeled graph constraining the system. Then, we showed that, even if one does not have access to the full state-space of the hybrid system, one can still provide stability guarantees by considering the joint spectral radius of an auxiliary system. Finally, we showcased our method on a real application of CSLS, namely Networked Control Systems.

For further work, we would like to extend our method to even broader families of dynamical systems, for example stochastic switching systems such as Markovian Jump Linear Systems. Second, other prior knowledge on the noise noise could be investigated, such as Gaussian noise for instance. Finally, we plan to relax some assumptions present in this work such as, for example, the amount of knowledge about the labeled graph required to compute the bounds on the CJSR.

A Proof of Proposition 2

First, we will need the following lemma.

Lemma 1 ([29]). *For $x \in \mathbb{R}^n$, let $p(x)$ be a non-zero polynomial. Then, the zero-set $\{x \in \mathbb{R}^n \,|\, p(x) = 0\}$ has zero Lebesgue measure.*

We note the samples $\omega_N = \{\delta_1, \ldots, \delta_N\}$. Moreover, we say that a subset $\beta \subset \omega_N$ is an *essential set* for ω_N if $(\lambda^*(\beta), c^*(\beta)) = (\lambda^*(\omega_N), c^*(\omega_N))$. Since the objective of the optimization problem (13) is strongly convex, there exists unique λ and $\{P_u\}_{u \in \mathcal{U}}$ such that, for all essential sets β, the following holds for all $(x, e, w) \in \beta$:

$$(A_{\sigma^{(l)}(e)}x + w)^T P_{t(e)}(A_{\sigma^{(l)}(e)}x + w) = \lambda^l x^T P_{s(e)}x. \tag{73}$$

Denote I the set of indices i in ω_N such that (73) holds for $\delta_i = (x, e, w)$. Now, let us look at the probability to sample a degenerate sample set[9]. Since the optimization problem always has at least one tight constraint, if the sample set is degenerate, it must have at least two tight constraints. Then, by symmetry, this probability is

$$\sum_{i=2}^{N} \binom{N}{i} \pi_l^N(\{\omega_N \,|\, \omega_N \text{ is degenerate and } I = [i]\})$$

$$\leq \sum_{i=2}^{N} i \binom{N}{i} \pi_l^N(\{\omega_N \,|\, I = [i] \text{ and } \{\delta_j\}_{j \in [i-1]} \text{ is an essential set for } \omega_N\}). \tag{74}$$

Indeed, if the sample set ω_N is degenerate and $I = [i]$, one of the constraints can be removed from I without changing the solution. Let

$$p_i = \pi_l^N(\{\omega_N \,|\, I = [i] \text{ and } \{\delta_j\}_{j \in [i-1]} \text{ is an essential set for } \omega_N\}), \tag{75}$$

we will show by contradiction that $p_i = 0$ for $i = 2, \ldots, N$. Suppose that $p_i > 0$, then we have

$$\pi_l(\{(x, e, w) \in \Delta \,|\, (A_{\sigma^{(l)}(e)}x + w)^T P_{t(e)}(A_{\sigma^{(l)}(e)}x + w) = \lambda^l x^T P_{s(e)}x\}) > 0. \tag{76}$$

Now, for any $e \in \mathcal{E}$, we define the polynomial

$$p_e(x, w) = \begin{pmatrix} x^T & w^T \end{pmatrix} \begin{pmatrix} A_{\sigma^{(l)}(e)}^T P_{t(e)} A_{\sigma^{(l)}(e)} - \lambda^l P_{s(e)} & A_{\sigma^{(l)}(e)}^T P_{t(e)} \\ P_{t(e)} A_{\sigma^{(l)}(e)} & P_{t(e)} \end{pmatrix} \begin{pmatrix} x \\ w \end{pmatrix}, \tag{77}$$

and let $\mu = \mu^x \times \mu^w$. Since $\mathcal{E}$ is finite, (76) implies that there exists $e \in \mathcal{E}$ such that $\mu(\{(x, w) \in \mathbb{S}^n \times B^n[W] \,|\, p_e(x, w) = 0\}) > 0$. Now, note that if (x, w) is a root of the polynomial p_e, so is $(\alpha x, \alpha w)$, for

---

[9]The set of *degenerate* sample sets refers to the complement of the set of non-degenerate sample sets, as defined in Definition 6.

$\alpha > 0$. Therefore, for all $\alpha > 0$, $\mu(\{(x,w) \in \alpha\mathbb{S}^n \times \alpha B^n[W] \,|\, p(x,w) = 0\}) > 0$. This implies that the set $\{(x,w) \in \mathbb{R}^n \times \mathbb{R}^n \,|\, p(x,w) = 0\}$ has non-zero Lebesgue measure, since so is its projection in the x-coordinate (thanks to the multiplication by α), and its projection on the w-coordinate (by Assumption 1). Therefore, by Lemma 1, the polynomial $p_e(x,w)$ is identically zero, which implies $P_{t(e)} = 0$ following Equation 77. However, since $P_{t(e)}$ is a solution of the program (13), $P_{t(e)} \succeq I$ following constraint (13b), which yields a contradiction. Therefore, for any $i = 1, \ldots, N$, $p_i = 0$, and the proof is completed following inequality (74). $\qquad\square$

References

[1] R. Alur, *Principles of Cyber-Physical Systems*. MIT Press, 2015.

[2] P. Tabuada, *Verification and Control of Hybrid Systems: A Symbolic Approach*. 06 2009.

[3] R. Goebel, R. G. Sanfelice, and A. R. Teel, *Hybrid Dynamical Systems: Modeling, Stability, and Robustness*. Princeton University Press, 03 2012.

[4] D. Liberzon and A. Morse, "Basic problems in stability and design of switched systems," *IEEE Control Systems Magazine*, vol. 19, no. 5, pp. 59–70, 1999.

[5] D. Liberzon, *Switching in Systems and Control*. 2003.

[6] R. Jungers, *The Joint Spectral Radius: Theory and Applications*. Springer Science & Business Media. Google-Books-ID: XcNkhSKpJlEC.

[7] G.-C. Rota and W. G. Strang, "A note on the joint spectral radius," in *Proc. Netherlands Academy*, vol. 22, p. 379–381, 1960.

[8] P. A. Parrilo and A. Jadbabaie, "Approximation of the joint spectral radius using sum of squares," *Linear Algebra and its Applications*, vol. 428, pp. 2385–2402, 2007.

[9] A. A. Ahmadi, R. M. Jungers, P. A. Parrilo, and M. Roozbehani, "Joint spectral radius and path-complete graph lyapunov functions," *SIAM Journal on Control and Optimization*, vol. 52, no. 1, pp. 687–717, 2014.

[10] M. C. F. Donkers, W. P. M. H. Heemels, N. van de Wouw, and L. Hetel, "Stability analysis of networked control systems using a switched linear systems approach," *IEEE Transactions on Automatic Control*, vol. 56, no. 9, pp. 2101–2115, 2011.

[11] R. Jungers, W. M. Heemels, and A. Kundu, "Observability and controllability analysis of linear systems subject to data losses," *IEEE Transactions on Automatic Control*, vol. PP, 09 2016.

[12] A. Jadbabaie, J. Lin, and A. Morse, "Coordination of groups of mobile autonomous agents using nearest neighbor rules," *IEEE Transactions on Automatic Control*, vol. 48, no. 6, pp. 988–1001, 2003.

[13] X. Dai, "A gel'fand-type spectral radius formula and stability of linear constrained switching systems," *Linear Algebra and its Applications*, vol. 436, no. 5, pp. 1099–1113, 2012.

[14] M. Philippe and R. M. Jungers, "Converse lyapunov theorems for discrete-time linear switching systems with regular switching sequences," 2014.

[15] M. Philippe, R. Essick, G. E. Dullerud, and R. M. Jungers, "Stability of discrete-time switching systems with constrained switching sequences," *Automatica*, vol. 72, pp. 242–250, 2016.

[16] V. Blondel and J. Tsitsiklis, "A survey of computational complexity results in systems and control," *Automatica*, vol. 36, p. 1249–1274, 09 2000.

[17] J. Kenanian, A. Balkan, R. M. Jungers, and P. Tabuada, "Data driven stability analysis of black-box switched linear systems," *Automatica*, vol. 109, p. 108533, Nov. 2019.

[18] G. O. Berger, R. M. Jungers, and Z. Wang, "Chance-constrained quasi-convex optimization with application to data-driven switched systems control," in *Proceedings of the 3rd Conference on Learning for Dynamics and Control*, pp. 571–583, PMLR. ISSN: 2640-3498.

[19] A. Rubbens, Z. Wang, and R. M. Jungers, "Data-driven stability analysis of switched linear systems with sum of squares guarantees," 2021.

[20] M. C. Campi and S. Garatti, *Introduction to the Scenario Approach*. MOS-SIAM Series on Optimization, 11 2018.

[21] F. Lauer, "Estimating the probability of success of a simple algorithm for switched linear regression," *Nonlinear Analysis Hybrid Systems*, vol. 8, pp. 31–47, 05 2013.

[22] L. Massucci, F. Lauer, and M. Gilson, "A statistical learning perspective on switched linear system identification," *Automatica*, vol. 145, p. 110532, 2022.

[23] A. Banse, Z. W. Raphaël, and M. Jungers, "Black-box stability analysis of hybrid systems with sample-based multiple lyapunov functions," in *2022 IEEE 61st Conference on Decision and Control (CDC)*, pp. 7284–7289, 2022.

[24] A. Banse, Z. Wang, and R. M. Jungers, "Learning stability guarantees for data-driven constrained switching linear systems," 2022.

[25] D. Lind and B. Marcus, *An Introduction to Symbolic Dynamics and Coding*. Cambridge: Cambridge University Press, 1995.

[26] G. C. Calafiore, "Random convex programs," *SIAM Journal on Optimization*, vol. 20, no. 6, pp. 3427–3464, 2010.

[27] M. Crowder, "Nist handbook of mathematical functions edited by frank w. j. olver, daniel w. lozier, ronald f. boisvert, charles w. clark," *International Statistical Review*, vol. 79, no. 1, pp. 131–132, 2011.

[28] J. P. Hespanha, P. Naghshtabrizi, and Y. Xu, "A survey of recent results in networked control systems," *Proceedings of the IEEE*, vol. 95, no. 1, pp. 138–162, 2007.

[29] R. Caron, "The Zero Set of a Polynomial," tech. rep., University of Windsor, Department of Mathematics and Statistics, 2005.