Data Driven Stability Analysis of Black-box Switched Linear Systems

Joris Kenanian [a], Ayca balkan [a], Raphael M. Jungers [b], Paulo Tabuada [a]

[a] *Department of Electrical Engineering at University of California, Los Angeles (UCLA)*

[b] *UC Louvain - ICTEAM Institute, Louvain-la-Neuve, Belgium*

Abstract

Can we conclude the stability of an unknown dynamical system from the knowledge of a finite number of snapshots of trajectories? We tackle this black-box problem for switched linear systems. We show that, for any given random set of observations, one can give probabilistic stability guarantees. The probabilistic nature of these guarantees implies a trade-off between their quality and the desired level of confidence. We provide an explicit way of computing the best stability-like guarantee, as a function of both the number of observations and the required level of confidence. Our proof techniques rely on geometrical analysis, chance-constrained optimization, and stability analysis tools for switched systems, including the joint spectral radius.

Key words: Learning algorithms; Data-driven verification; Stability; Switched systems; Optimization; Lyapunov methods

1 Introduction

Most of the existing work on stability of dynamical systems is model-based, i.e., it requires the knowledge of a model for the considered system. Although natural in many contexts, a model may not always be available. Cyber-physical systems are an illustration of such difficulty: they consist of a large number of components of different nature (modeled by differential equations, difference equations, hybrid automata, lookup tables, etc.) engaged in complex interactions with each other. Closed-form models for these complex and heterogeneous systems are equally complex or even not available, and therefore one cannot use model-based techniques in these situations. The emphasis that industry places on simulation of such systems is then not surprising, since it is always possible to simulate them despite their complexity. This raises the question of whether one can provide formal guarantees about certain properties of these complex systems, based solely on information obtained via their simulations. We focus here on one of the most important of such properties in the context of control theory: stability.

More formally, we consider a time-varying discrete-time dynamical system of the form:

$$x_{k+1} = f(k, x_k), \tag{1}$$

where $x_k \in X$ is the state of the system and $k \in \mathbb{N}$ is the time index. For the rest of the paper, we use the term *black-box* to refer to systems where we do not have access to the model, i.e., to f, yet we can indirectly learn information about f by observing finite trajectories of length l (in the particular case of $l = 1$, these trajectories become pairs of points (x_k, x_{k+1}) as defined in (1)). We start with the following question to serve as a stepping stone: For some $l \in \mathbb{N}_{>0}$, given N finite trajectories of length l, $(x_{i,0}, x_{i,1}, \ldots, x_{i,l})$, $1 \leq i \leq N$, belonging to the behavior of system (1), (i.e., $x_{i,k+1} = f(k, x_{i,k})$ for any $0 \leq k \leq l-1$ and any $1 \leq i \leq N$), what can we say about the stability of System (1)?

A potential approach to this problem is to first identify the dynamics, i.e., the function f, and then apply existing techniques from the model-based stability analysis literature. If System (1) is linear, its identification and

⋆ A. Balkan, J. Kenanian, and P. Tabuada had their work supported by NSF Grant 1645824 and NSF Expeditions in Computer Augmented Program Engineering. R. Jungers is a Fulbright Fellow and a F.N.R.S. Fellow currently visiting the Dept. of Electrical Engineering at UCLA. His work is supported by the Walloon Region and the Innoviris Foundation.

Email addresses: jkenanian@engineering.ucla.edu (Joris Kenanian), abalkan@ucla.edu (Ayca balkan), raphael.jungers@uclouvain.be (Raphael M. Jungers), tabuada@ucla.edu (Paulo Tabuada).

stability analysis have been extensively studied. If f is not a linear function and in particular if the system is a switched system, there are two main reasons behind our quest to directly work on system behaviors and bypass the identification phase:

- Identification can potentially introduce approximation errors, and can have a high computational complexity. Again, this is the case for switched systems, for which the identification problem is NP-hard (Lauer, 2016);
- Even when the function f is known, in general, stability analysis is a very difficult problem (Blondel and Tsitsiklis, 1999).

A fortiori, the combination of these two steps in an efficient and robust way seems far from obvious.

In recent years, an increasing number of researchers started addressing various verification and design problems in control of black-box systems (Balkan et al., 2016; Blanchini et al., 2017; Duggirala et al., 2013; Huang and Mitra, 2014; Kozarev et al., 2016). The initial idea behind this paper was influenced by the recent efforts in Balkan et al. (2016); Kapinski et al. (2014); Topcu et al. (2008), and Bobiti and Lazar (2015) on using simulation trajectories to find Lyapunov functions for systems with known dynamics. In these works, the main idea is that if one can construct a Lyapunov function candidate decreasing along several finite trajectories starting from different initial conditions, it should hopefully decrease along every other trajectory. Then, once a Lyapunov function candidate is constructed, this intuition is put to test by verifying the candidate function either via off-the-shelf tools as in Topcu et al. (2008) and Kapinski et al. (2014), or via sampling-based techniques as in Bobiti and Lazar (2015). This also relates to almost-Lyapunov functions introduced in Liu et al. (2016), which presents a relaxed notion of stability proved via Lyapunov functions decreasing everywhere except on a small set. These approaches cannot be directly applied to black-box systems, where we do not have access to the dynamics -as in our framework. However, they are based on the following idea that we address in this paper: By observing that a candidate Lyapunov function decreases on a large number of observations, we empirically build a certain confidence that this function is a bona-fide Lyapunov function. *Can we translate this empirical observation on a finite set of points into a confidence that this Lyapunov function decreases in the whole state space?*

Note that, even in the case of a linear system, the connection between these two beliefs is nontrivial. One can easily construct an example where a candidate Lyapunov function decreases everywhere on its levels sets, except for an arbitrarily small subset, yet, almost all trajectories diverge to infinity (see Example 1 in Kenanian et al. (2018)). In this work, we take a first step into more complex systems than the linear case by considering the class of switched linear systems. In addition to the phenomenon exhibited in the above example, switched linear systems seem a priori challenging for black-box stability analysis, as both the identification and "white-box" stability analysis are hard for these systems. Deciding stability of a switched linear system amounts to decide whether its *Joint Spectral Radius* is smaller than 1, which is extremely hard even in the white-box setting (see, e.g., Jungers (2009), Chapter 2, for various complexity results).

We present an algorithm to bound the JSR of an unknown switched linear system from a finite number N of observations of trajectories. This algorithm partially relies on tools from the random convex optimization literature (also known as chance-constrained optimization, see (Calafiore, 2010; Campi and Garatti, 2008; Nemirovski and Shapiro, 2006)), and provides an upper bound on the JSR with a user-defined confidence level. As N increases, this bound gets tighter. Moreover, with a closed form expression, we characterize what is the exact trade-off between the tightness of this bound and the number of samples. In order to understand the quality of our upper bound, the algorithm also provides a deterministic lower bound. Finally, we provide a guarantee of asymptotic convergence between the upper and the lower bound, for large N. The reader can find a developed application of our work to networked control systems in Kenanian et al. (2018).

The organization of the paper is as follows: In Section 2, we introduce the problem studied and provide the necessary background in stability of switched linear systems. Then, based on finite observations for a given switched linear system, we give in Section 3 a deterministic lower bound for the JSR, before presenting in Section 4 the main contribution of this paper, which is the probabilistic upper bound. We illustrate the performance of the presented techniques with some experiments in Section 5, and we propose future extensions of this work in Section 6.

2 Preliminaries

2.1 Notation

We consider the usual finite normed vector space $(\mathbb{R}^n, \ell_2)$, $n \in \mathbb{N}_{>0}$, with ℓ_2 the classical Euclidean norm. We denote by $\|x\|$ the ℓ_2-norm of $x \in \mathbb{R}^n$. We also denote the set of linear functions from $\mathbb{R}^n$ to $\mathbb{R}^n$ by $\mathcal{L}(\mathbb{R}^n)$, and the set of real symmetric matrices of size n by $\mathcal{S}^n$. In particular, the set of positive definite matrices is denoted by $\mathcal{S}^n_{++}$. We write $P \succ 0$ to state that P is positive definite, and $P \succeq 0$ to state that P is positive semi-definite. Given a set $X \subset \mathbb{R}^n$, we denote by $\wp(X)$ its powerset (i.e., the set of all its subsets), and by $X^{\mathbb{N}}$ the set of all possible sequences $(x_n)_{n\in\mathbb{N}}$, $x_n \in X$. For

any $r \in \mathbb{R}_{>0}$, we write $rX := \{rx : x \in X\}$ to denote the scaling of ratio r of X. We denote by $\mathbb{B}$ (respectively $\mathbb{S}$) the ball (respectively sphere) of unit radius centered at the origin. We denote the ellipsoid described by the matrix $P \in \mathcal{S}^n_{++}$ as E_P, i.e., $E_P := \{x \in \mathbb{R}^n : x^T P x = 1\}$. Finally, we denote the spherical projector on $\mathbb{S}$ by $\Pi_{\mathbb{S}}(x) := x/\|x\|$.

In this paper, we only consider simple uniform probability distributions, and we believe that all the concepts can be easily intuitively understood. However, for the sake of rigor, we now develop the proper measure-theoretic setting on which our results build. We consider the classical measure space $(\mathbb{R}^n, \mathcal{B}_{\mathbb{R}^n}, \lambda)$, where $\mathcal{B}_{\mathbb{R}^n}$ is the Borel σ-algebra generated by the topology of $\mathbb{R}^n$ induced by ℓ_2, and where λ denotes the Lebesgue measure. We denote by $\mathcal{B}_{\mathbb{S}}$ the Borel σ-algebra of $\mathbb{S}$, generated by the topology on $\mathbb{S}$ induced from its ambient space $\mathbb{R}^n$. The space $(\mathbb{S}, \mathcal{B}_{\mathbb{S}})$ is provided with the classical, unsigned and finite uniform spherical measure, denoted by σ^{n-1} (n is the dimension of the space where $\mathbb{S}$ is embedded), and derived from the Lebesgue measure λ as follows. For any $\mathcal{A} \subset \mathbb{S}$, the *sector* defined by $\mathcal{A}$, denoted by $\mathbb{S}^{\mathcal{A}}$, is the subset $\{t\mathcal{A},\ t \in [0,1]\} \in \mathcal{B}_{R^n}$. Then, $\forall\ \mathcal{A} \in \mathcal{B}_{\mathbb{S}}$, $\sigma^{n-1}(\mathcal{A}) = \frac{\lambda(\mathbb{S}^{\mathcal{A}})}{\lambda(\mathbb{B})}$. In other words, the spherical measure of a subset of the sphere is related to the Lebesgue measure of the sector of the unit ball it induces. Notice that $\sigma^{n-1}(\mathbb{S}) = 1$.

For $m \in \mathbb{N}_{>0}$, we denote by M the set $M = \{1, 2, \ldots, m\}$. The set M is provided with the classical σ-algebra associated to finite sets: $\Sigma_M = \wp(M)$. We provide (M, Σ_M) with the uniform measure μ_M. For any $l \in \mathbb{N}_{>0}$, we denote by M^l the l-Cartesian product of M, i.e., $M^l = \{(i_1, \ldots, i_l) | i_j \in M, 1 \geq j \geq l\}$. We define Σ_{M^l} as the product $\bigotimes^l \Sigma_M$ (which is here equal to $\wp(M)^l$), and we provide (M^l, Σ_{M^l}) with the uniform product measure $\mu_{M^l} = \otimes^l \mu_M$. We can now define $Z_l = \mathbb{S} \times M^l$ as the Cartesian product of $\mathbb{S}$ and M^l. We provide the set Z_l with the product σ-algebra $\mathcal{B}_{\mathbb{S}} \bigotimes (\Sigma_{M^l})$ generated by $\mathcal{B}_{\mathbb{S}}$ and Σ_{M^l}: $\Sigma = \sigma(\pi_{\mathbb{S}}^{-1}(\mathcal{B}_{\mathbb{S}}), \pi_{M^l}^{-1}(\Sigma_{M^l}))$, where $\pi_{\mathbb{S}} : Z_l \to \mathbb{S}$ and $\pi_{M^l} : Z_l \to M^l$ are the standard projections. On (Z_l, Σ), we define the product measure $\mu_l = \sigma^{n-1} \otimes \mu_{M^l}$. Note that, μ_l is the uniform probability measure on Z_l. We will also need two classical functions to compute our probabilistic upper bound, which are known as the *incomplete beta function* and the *regularized incomplete beta function*

Definition 1 (Olver et al. (2010), 6.6.1)
The incomplete beta function, denoted by B, is given by

$$B : \begin{cases} \mathbb{R}_{>0} \times \mathbb{R}_{>0} \times \mathbb{R}_{>0} \to \mathbb{R}_{\geq 0} \\ (x, a, b) \mapsto B(x, a, b) = \displaystyle\int_0^x t^{a-1}(1-t)^{b-1} dt. \end{cases}$$

Definition 2 (Olver et al. (2010), 6.6.2)
The regularized incomplete beta function, denoted by I, is given by

$$I : \begin{cases} \mathbb{R}_{>0} \times \mathbb{R}_{>0} \times \mathbb{R}_{>0} \to \mathbb{R}_{\geq 0} \\ (x, a, b) \mapsto I(x, a, b) = \dfrac{B(x, a, b)}{B(1, a, b)}. \end{cases}$$

For given values of parameters $a > 0$ and $b > 0$, the inverse of the regularized incomplete beta function with parameters a, b, denoted by $I^{-1}(y, a, b)$, is the function whose output is $x > 0$ such that $I(x, a, b) = y$ (Majumder and Bhattacharjee, 1973).

2.2 Stability of Switched Linear Systems

A *switched linear system*, defined by a set of modes (matrices) $\mathcal{M} = \{A_i, i \in M\}$, is a time-varying discrete-time dynamical system of the form (1), with $f(k, x_k) = A_{\tau(k)} x_k$, that is:

$$x_{k+1} = A_{\tau(k)} x_k, \tag{2}$$

for any $k \in \mathbb{N}$. Here, the signal $\tau \in M^{\mathbb{N}}$ is called the *switching sequence*, and can take arbitrary values in M. Note that such systems are homogeneous, i.e., for any $\gamma > 0$, $f(k, \gamma x_k) = \gamma f(k, x_k)$. In this paper, we assume to not have access to $\mathcal{M}$ nor to the switching sequence. The only information available is (an upper bound on) m, the cardinality of $\mathcal{M}$.

We are interested in the *uniform asymptotic stability* of the system, that is, we want to guarantee the following property:

$$\forall \tau \in M^{\mathbb{N}}, \forall x_0 \in \mathbb{R}^n,\ \|x_k\| \xrightarrow[k \to \infty]{} 0.$$

The joint spectral radius of a set of matrices $\mathcal{M}$ characterizes the stability of the underlying switched linear system (2) defined by $\mathcal{M}$ (Jungers, 2009). This quantity is an extension to switched linear systems of the classical spectral radius for linear systems. It is the maximum asymptotic growth rate of the norm of the state under the dynamics (2), over all possible initial conditions and sequences of matrices of $\mathcal{M}$.

Definition 3 (from Jungers (2009)) *Given a finite set of matrices $\mathcal{M} \subset \mathbb{R}^{n \times n}$, its* joint spectral radius *(JSR) is given by*

$$\rho(\mathcal{M}) = \lim_{k \to \infty} \max_{i_1, \ldots, i_k} \left\{ ||A_{i_1} \ldots A_{i_k}||^{1/k} : A_{i_j} \in \mathcal{M} \right\}.$$

Property 1 (Jungers (2009), Cor. 1.1) *Given a finite set of matrices $\mathcal{M}$, the corresponding switched dynamical system is stable if and only if $\rho(\mathcal{M}) < 1$.*

Definition 4 *Consider a finite set of matrices $\mathcal{M} \subset \mathbb{R}^{n\times n}$. A* common quadratic form (CQF) *for a system (2) with set of matrices $\mathcal{M}$, is a positive definite matrix $P \in \mathcal{S}^n_{++}$ such that for some $\gamma \geq 0$,*

$$\forall A \in \mathcal{M}, A^T P A \preceq \gamma^2 P. \tag{3}$$

CQFs are useful because they can be computed, when they exist, with semidefinite programming (see (Boyd and Vandenberghe, 2004)), and they constitute a stability guarantee (when $\gamma < 1$, they are Lyapunov functions) for switched systems as we formalize next.

Theorem 5 (Jungers (2009), P. 2.8 & T. 2.11) *Consider a finite set of matrices $\mathcal{M}$.*

- *If there exist $\gamma \geq 0$ and $P \succ 0$ such that Equation (3) holds, then $\rho(\mathcal{M}) \leq \gamma$.*
- *If $\rho(\mathcal{M}) < \frac{\gamma}{\sqrt{n}}$, there exists a CQF, P, such that $\forall A \in \mathcal{M}$, $A^T P A \preceq \gamma^2 P$.*

For any $\gamma < 1$, this theorem provides both a Lyapunov and a *converse Lyapunov result*: if there exists a CQF, then our system is stable; if there is, on the contrary, no such stability guarantee, one may conclude a lower bound on the JSR. We obtain then an approximation algorithm for the JSR. It turns out that one can still refine this technique, in order to improve the error factor $1/\sqrt{n}$, and asymptotically get rid of it. This is a well-known technique for the "white-box" computation of the JSR, which we summarize in the following corollary.

Corollary 6 *Fix $\gamma \geq 0$. For any finite set of matrices such that $\rho(\mathcal{M}) < \frac{\gamma}{\sqrt[2l]{n}}$ with $\gamma \geq 0$, there exists a CQF for $\mathcal{M}^l := \{\Pi_{j=1}^l A_{i_j} : A_{i_j} \in \mathcal{M}\}$, that is, a $P \succ 0$ such that:*

$$\forall\ \mathbf{A} \in \mathcal{M}^l,\ \mathbf{A}^T P \mathbf{A} \preceq \gamma^{2l} P. \tag{4}$$

PROOF. It is easy to see from the definition of the JSR that $\rho(\mathcal{M}^l) = \rho(\mathcal{M})^l$. Thus, applying Theorem 5 to the finite set $\mathcal{M}^l$, one directly obtains the corollary. □

Note that, the smaller γ is in Theorem 5, the tighter is the upper bound we get on $\rho(\mathcal{M})$. In order to properly analyze our setting, where the matrices are unknown, let us reformulate (4) in another form. For any $l \in \mathbb{N}_{>0}$, we can consider the optimal solution γ^* of the following optimization problem:

$$\begin{aligned} \min_{\gamma,P} \quad & \gamma \\ \text{s.t.} \quad & (\mathbf{A}x)^T P \mathbf{A}x \leq \gamma^{2l} x^T P x, \mathbf{A} \in \mathcal{M}^l, \forall x \in \mathbb{S} \\ & P \succ 0. \end{aligned} \tag{5}$$

Notice that we restrict the set of constraints by restricting x to $\mathbb{S}$, due to the homogeneity of the system. Homogeneity indeed implies that it is sufficient to show the decrease of a CQF on an arbitrary set enclosing the origin.

The above equation will provide a clear algebraic formalization of our black-box problem: our goal amounts to find a solution to a convex problem with an infinite number of constraints, while only sampling a finite number of them.

2.3 Problem Formulation

Let us now formally present the problem addressed in this paper. We recall that we only observe N finite trajectories of length $l \in \mathbb{N}_{>0}$, i.e., N sequences of states $(x_k, x_{k+1}, \ldots, x_{k+l})$ where x_{k+i+1} and x_{k+i} are related by (2). Note that such sequences of states depend both on the initial state x_k and the switching sequence $\tau(k)$ which is assumed to be unknown. In other words, we do not observe the mode or the matrices used to produce the trajectories. We do not have access to the process through which the system picks the modes. The user's knowledge is limited to the number of modes (or an upper bound on this number) and the dimension of the system. We assume that these trajectories are generated from a finite number of initial conditions $x_{i,0} \in \mathbb{S}$, $1 \leq i \leq N$ enumerating the observations, and that a random sequence of l matrices is applied to each of these points. We randomly draw the initial conditions from $\mathbb{S}$, observe them and the l subsequent state values produced by the system. Sampling the initial conditions from $\mathbb{S}$ is without loss of generality, since any trajectory in $\mathbb{R}^n$ can be rescaled so that $x_{i,0} \in \mathbb{S}$, by homogeneity of the system. To a given observed trajectory $(x_k, x_{k+1}, \ldots, x_{k+l})$, we can associate the corresponding probability event $(x_k, j_1, \ldots, j_l)$ which is another $(l+1)$-tuple. Formally, with the fixed probability space (Z_l, Σ, μ_l), we consider the random variables $X : Z_l \mapsto \mathbb{S}$ and $\theta_i : Z_l \mapsto M$, for $1 \leq i \leq l$. We can note that the random variable X, from which the initial conditions x_k are drawn, has uniform distribution (on $(\mathbb{S}, \mathcal{B}_{\mathbb{S}}, \sigma^{n-1})$). The random variables θ_i, from which the indices j_i of the modes applied at the i^{th} step are drawn, are independent and also have uniform distribution (on the space (M, Σ_M, μ_M)). Thus, to a given random finite set of N trajectories of length l, we can associate an underlying uniform sample of N such $(l+1)$-tuples in Z_l, denoted by

$$\omega_N := \{(x_{i,0}, j_{i,1}, \ldots, j_{i,l}), 1 \leq i \leq N\} \subset Z_l. \tag{6}$$

In other words, a set of N available observations $\{(x_{i,0}, x_{i,1}, \ldots, x_{i,l}), 1 \leq i \leq N\}$ can be rewritten, for all $1 \leq i \leq N$ and $1 \leq k \leq l$, as $x_{i,k} = A_{j_{i,k}} \ldots A_{j_{i,1}} x_{i,0}$, with the $j_{i,k}$ being unobserved variables that take their values in M.

Remark 7 *Let us motivate our assumption on the uniform drawing of the matrices. We assumed that we only have access to random observations of the state of the system, and ignore the process that generates them. In particular, we ignore the process that selects the modes at each time step, and model it as a random process. We*

suppose that with nonzero probability, each mode is active: the problem would indeed not be solvable otherwise, since the system would be unidentifiable with probability 1 *and would prevent to ever observe some of its behaviors. We take this distribution to be uniform since we cannot say that some modes are more likely a priori. Our results extend to the case where the distribution is not uniform as long as we have a nonzero lower bound on the probability of each mode.*

In this work, we aim at understanding what type of guarantees one can obtain on the stability of System (2) (that is, on the JSR of $\mathcal{M}$) from the sample (6). More precisely, we answer the following problem:

Problem 1 *Consider a finite set of matrices, $\mathcal{M}$, describing a switched system* (2)*, and suppose that one has a set of N observations $(x_{i,0}, x_{i,1}, ..., x_{i,l}), i = 1, ..., N$ corresponding to an event ω_N, sampled from Z_l with the uniform measure μ_l.*

- *For a given confidence level $\beta \in [0, 1)$, provide an upper bound on $\rho(\mathcal{M})$, that is, a number $\overline{\rho(\omega_N)}$ such that*

$$\mu_l^N \left(\{\omega_N : \ \rho(\mathcal{M}) \leq \overline{\rho(\omega_N)}\} \right) \geq \beta.$$

- *For the same given level of confidence β, provide a lower bound $\underline{\rho(\omega_N)}$ on $\rho(\mathcal{M})$.*

Remark 8 *We will see in Section 3 that we can even derive a deterministic lower bound for a given (sufficiently high) number of observations.*

We will see in Theorem 17 that for any level of confidence β, it is always possible to provide an upper bound for Problem 1 which tends to the JSR when the number of sampled points increases. In particular, for any (large enough) number of samples, it is always possible to provide such an upper bound that is finite. We can then consider the statistical test with null hypothesis 'H_0 : $\rho(\mathcal{M}) < 1$' and alternate hypothesis '$H_A : \rho(\mathcal{M}) \geq 1$'. This is a Hypothesis test for the question 'Is the JSR strictly lower than 1?', which is equivalent to the question 'Is the system stable?' by Property 1. This test has an a priori fixed probability $1 - \beta$ of false positive (case where $\overline{\rho(\omega_N)} < 1$ and $\rho(\mathcal{M}) > 1$). Theorem 17 will also show that the probability of false negative (case where $\overline{\rho(\omega_N)} > 1$ and $\rho(\mathcal{M}) < 1$) tends to zero when the number of samples N increases.

The key insight is to leverage the fact that conditions for the existence of a CQF for (2) can be obtained by considering a finite number of trajectories in $\mathbb{R}^n$ of the form $(x_k, x_{k+1}, \ldots, x_{k+l})$. Developing that insight leads us to the following algorithm, that is the main result of our work and that solves Problem 1:

Algorithm 1 (Probabilistic upper bound)
Input: *observations produced by a uniform random sample $\omega_N \subset Z_l$ of size $N \geq \frac{n(n+1)}{2} + 1$;*

Input: *β desired level of certainty;*

Compute: *a candidate for the upper bound, $\gamma^*(\omega_N)$, solution of the convex optimization problem* (7)*;*
(observe that (7) *does not require the explicit knowledge of the matrices $\mathbf{A_j}$)*

Compute: *$\varepsilon(\beta, \omega_N)$ the proportion of points where our inference on the upper bound may be invalid;*

Compute: *$\delta(\varepsilon) \leq 1$ a correcting factor; ($\delta \xrightarrow[N\to\infty]{} 1$)*

Output: $\frac{\gamma^*(\omega_N)}{\sqrt[2l]{n}} \leq \rho \leq \frac{\gamma^*(\omega_N)}{\sqrt[l]{\delta(\varepsilon)}}$;
(the right-hand side inequality is valid with probability at least β).

3 A Deterministic Lower Bound

In Section 2.2, we presented an optimization problem, (5), that provides a stability guarantee. Nevertheless, this problem has infinitely many constraints and observing a finite number of trajectories only gives us access to a restriction of it (with finitely many constraints). We consider then the following optimization problem:

$$\begin{aligned} \min_{\gamma, P} \quad & \gamma \\ \text{s.t.} \quad & (\mathbf{A_j}x)^T P \mathbf{A_j} x \leq \gamma^{2l} x^T P x, \, \forall (x, \mathbf{j}) \in \omega_N \\ & P \succ 0, \ \gamma \geq 0. \end{aligned} \tag{7}$$

with optimal solution $\gamma^*(\omega_N)$, and where $\mathbf{A_j} := A_{j_l} A_{j_{l-1}} \ldots A_{j_1}$ and $\mathbf{j} := \{j_1, \ldots, j_l\}$. Note that, (7) can be efficiently solved by semidefinite programming and bisection on the variable γ (see (Boyd and Vandenberghe, 2004)). Note also that solving this program can be done in practice only through the knowledge of the observations: even though the A_{j_i} are not known, the program only requires the knowledge of $\mathbf{A_j}x$, which is known through the observations. In this section, we provide a theorem for a deterministic lower bound based on the observations given by ω_N, whose accuracy depends on the "horizon" l.

Theorem 9 *For an arbitrary $l \in \mathbb{N}_{>0}$ and a given uniform sample $\omega_N \subset Z_l$, by denoting $\gamma^*(\omega_N)$ the optimal solution of the optimization problem* (7)*, we have*

$$\rho(\mathcal{M}) \geq \frac{\gamma^*(\omega_N)}{\sqrt[2l]{n}}.$$

PROOF. Proof. Let $\nu > 0$. By definition of $\gamma^*(\omega_N)$, there exists no matrix $P \in \mathcal{S}^n_{++}$ such that, $\forall x \in \mathbb{S}$, $\forall \mathbf{A_j} \in \mathcal{M}^l$,

$$(\mathbf{A_j}x)^T P \mathbf{A_j} x \leq (\gamma^*(\omega_N) - \nu)^{2l} x^T P x.$$

Taking the contrapositive of Corollary 6, this implies that $\rho(\mathcal{M}) \geq \frac{(\gamma^*(\omega_N)-\nu)^l}{\sqrt[2l]{n}}$. Since this is valid for any $\nu > 0$, we finally obtain the claim.

4 A Probabilistic Stability-like Guarantee

4.1 A Partial Upper Bound

In this section, we show how to compute an upper bound on ρ, with a user-defined confidence $\beta \in [0, 1)$. We do this by constructing an l-step CQF which is valid with probability at least β. The existence of an l-step CQF implies $\rho \leq \gamma^*$ due to Theorem 5. As we will see below, the quality of our bound will depend on geometrical properties of the CQF found; more precisely, the smaller the condition number of the corresponding matrix P, the better will be our bound. In practice, one can minimize the condition number of the solution P in a second step, after computing γ^* from (7). However, for the sake of rigor and clarity of our proofs, we introduce a slightly different optimization problem. We consider for the rest of the discussion the following optimization problem, that we denote by $\mathrm{Opt}(\omega_N)$:

$$\begin{aligned} \min_{P} \quad & \lambda_{\max}(P) \\ \text{s.t.} \quad & (\mathbf{A_j}x)^T P \mathbf{A_j} x \leq ((1+\eta)\gamma^*(\omega_N))^{2l} x^T P x, \\ & \qquad\qquad \forall\ (x, \mathbf{j}) \in \omega_N \\ & P \succeq I, \end{aligned} \tag{8}$$

with $\eta > 0$, and where $\gamma^*(\omega_N)$ is the optimal solution to the optimization problem (7). Let us analyze the relationship between $\mathrm{Opt}(\omega_N)$ and the optimization problem (7). Firstly, thanks to the homogeneity of system (2), feasibility of the optimization problem does not change when replacing the constraint $P \succ 0$ in the initial problem with the constraint $P \succeq I$. Secondly, as discussed above, the objective function $\lambda_{\max}(P)$ (which is convex) can be added once γ^* is computed, in order to minimize the condition number. Lastly, we introduced a "regularization parameter", $\eta > 0$, which ensures strict feasibility of $\mathrm{Opt}(\omega_N)$. As the reader will see, we will derive results valid for arbitrarily small values of η. This will then not hamper the practical accuracy of our technique, while allowing us to derive a theoretical asymptotic guarantee (i.e., for a large number of observations). We denote the optimal solution of $\mathrm{Opt}(\omega_N)$ by $P(\omega_N)$, and drop the explicit dependence of P on ω_N when it is clear from the context.

The intriguing question of whether the optimal solution of this sampled problem is a feasible solution to (5) has been widely studied in the literature (Calafiore, 2010). Under certain technical assumptions, one can bound the proportion of the constraints of the original problem (5) that are violated by the optimal solution of $\mathrm{Opt}(\omega_N)$, with some probability which is a function of the sample size N. In the following theorem, we adapt a classical result from the random convex optimization literature to our problem.

Theorem 10 (adapted from T.3.3 [1] , (Calafiore, 2010)) *Consider the optimization problem $Opt(\omega_N)$ given in (8), where ω_N is a uniform random sample drawn from the set Z_l. Let $d = \frac{n(n+1)}{2}$ be the dimension of the decision variable P of $Opt(\omega_N)$ and $N \geq d + 1$. Then, for all $\varepsilon \in (0, 1]$ the following holds:*

$$\mu_l^N\big\{\omega_N \in Z_l^N : \mu_l\left(V(\omega_N)\right) \leq \varepsilon\big\} \geq \beta(\varepsilon, N), \tag{9}$$

where $\beta(\varepsilon, N) = 1 - \sum_{j=0}^{d} \binom{N}{j} \varepsilon^j (1-\varepsilon)^{N-j}$, and $V(\omega_N)$ is the set $\{(x, \mathbf{j}) \in Z_l : (\mathbf{A_j}x)^T P(\omega_N) \mathbf{A_j} x > (\gamma^*_{\omega_N})^{2l} x^T P(\omega_N) x\}$, i.e., it is the subset of Z_l for which the considered γ^*-contractivity is violated by the optimal solution of $Opt(\omega_N)$.

The quantity ε can also be seen as a function of β and N: $\varepsilon(\beta, N) = 1 - I(1-\beta, N-d, d+1)$ (see the proof of Theorem 3.1 in (Calafiore, 2010)).

Corollary 11 *Consider a set of matrices $\mathcal{M}$ and denote γ^* the optimal solution of (7) and $P \succ 0$ the optimal solution of $Opt(\omega_N)$. Then, with the notation of Theorem 10, for any $\varepsilon > 0$, the set Ω of observations $\omega_N \in Z_l^N$ for which we have*

$$(\mathbf{A_j}x)^T P \mathbf{A_j} x \leq (\gamma^*)^{2l} x^T P x, \forall x \in \mathbb{S} \setminus \tilde{\mathbb{S}}, \forall \mathbf{j} \in M^l \tag{10}$$

for some $\tilde{\mathbb{S}} \subset \mathbb{S}$ with measure $\sigma^{n-1}(\tilde{\mathbb{S}}) \leq \varepsilon m^l$, verifies $\mu_l^N(\Omega) \geq \beta(\varepsilon, N)$.

This result allows us to make abstraction of the probabilistic setting: by accepting a confidence level of β smaller than one, we may assume that all the points except a small set satisfy the Lyapunov equation (5). The case where we have the equality $\sigma(\tilde{\mathbb{S}}) = \varepsilon m^l$ corresponds to the case where every point $x \in \tilde{\mathbb{S}}$ violates (10) for exactly one value of $\mathbf{j}$.

[1] Theorem 3.3 in (Calafiore, 2010) requires $\mathrm{Opt}(\omega_N)$ to satisfy the following technical assumptions:

(1) When the problem $\mathrm{Opt}(\omega_N)$ admits an optimal solution, this solution is unique.

(2) Problem $\mathrm{Opt}(\omega_N)$ is nondegenerate with probability 1.

Here, the first assumption can be enforced if required by adding a tie-breaking rule to $\mathrm{Opt}(\omega_N)$ as explained in Appendix A in (Calafiore and Campi, 2006), while the second assumption can be lifted, as explained in PART 2b in (Campi and Garatti, 2008), thanks to the introduction of "constraint heating".

PROOF. Proof. From Theorem 10, we can assume, with a level of confidence β, that $\mu(V(\omega_N)) \leq \varepsilon$, where $V(\omega_N) = \{(x, \mathbf{j}) \in Z_l : (\mathbf{A_j}x)^T P(\omega_N)\mathbf{A_j}x > (\gamma^*_{\omega_N})^{2l} x^T P(\omega_N)x\}$. Since $\tilde{\mathbb{S}} = \{x \in \mathbb{S} : \exists \mathbf{j} \in M^l : (\mathbf{A_j}x)^T P(\omega_N)\mathbf{A_j}x > (\gamma^*_{\omega_N})^{2l} x^T P(\omega_N)x\}$, we have that

$$\sigma^{n-1}(\tilde{\mathbb{S}}) \leq \mu(V(\omega_N))m^l \leq \varepsilon m^l.$$

The above results allow us to conclude, from a finite number of observations, that with probability β (where β goes to 1 as N goes to infinity), the required property is actually satisfied for the complete sphere $\mathbb{S}$, except on a small set of measure at most $\tilde{\varepsilon} = \varepsilon m^l$. This means that, the ellipsoid E_P computed by Opt(ω_N) is "almost invariant" except on a set of measure bounded by $\tilde{\varepsilon}$. This is represented in Fig. 1. for the case $n = 2$, where the red points of E_P are points that might violate the invariance constraint. Here, the set of red points has measure at most $\tilde{\varepsilon}$.

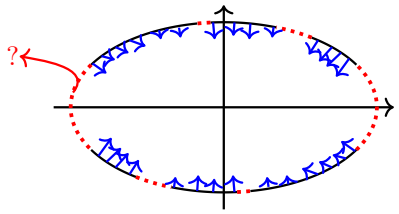

Fig. 1. Representation of the "partial invariance property" obtained by application of the results in Theorem 10. A priori, we know nothing about the images of the (dotted) red points under (2). Our goal is to convert this partial invariance property into a global stability property.

Thus, we are left with the following question:

Problem 2 *What can we conclude on the JSR if the invariance property is satisfied by all points, except a set of measure $\tilde{\varepsilon}$?*

In the course of Theorem 15, we will be able to derive an upper bound by solving the geometric problem of computing the largest scaling of E_P included in the convex hull of the subset of points of E_P that satisfy the invariance property (10). Indeed, this smaller ellipsoid will satisfy a (relaxed) invariance property *for all its points*, thanks to the following key property of switched linear systems.

Property 2 *The dynamics given in* (2) *is convexity-preserving, meaning that for any set of points $X \subset \mathbb{R}^n$,*

$$\{Ax : A \in \mathcal{M}, x \in \textit{convhull } X\} \subset \textit{convhull } \{Ax : A \in \mathcal{M}, x \in X\}.$$

Of course, for a fixed measure $\tilde{\varepsilon}$, this largest ellipsoid will depend on the distribution of points of E_P that violate the constraint. In order to obtain a guarantee on our upper bound, we will look for the smallest such ellipsoid obtained over all possible sets $\tilde{\mathbb{S}}$ of measure $\tilde{\varepsilon}$.

We start by solving this problem in the particular case where $E_P = \mathbb{S}$. In this case, we benefit from the following tool, allowing to explicitly analyse the worst-case distribution.

Definition 12 *We define the* spherical cap *on $\mathbb{S}$ for a given hyperplane $c^T x = k$, as $\mathcal{C}_{c,k} := \{x \in \mathbb{S} : c^T x > k\}$.*

We now define the following function which quantifies the largest-inscribed-sphere problem, for a given subset $X \subset \mathbb{S}$:

$$\Delta : \begin{cases} \wp(\mathbb{S}) \to [0, 1] \\ X \mapsto \sup\{r : r\mathbb{B} \subset \text{convhull } (\mathbb{S} \setminus X)\}. \end{cases} \tag{11}$$

The following proposition tells us that, when the measure of the set X is fixed, Δ is minimized when X is a spherical cap, i.e., the minimal radius δ of the largest sphere $\delta\mathbb{S}$ included in $\mathbb{S} \setminus X$ will be reached when X is a spherical cap.

Proposition 13 *Let $\tilde{\varepsilon} \in [0, 1]$ and $\mathcal{X}_{\tilde{\varepsilon}} = \{X \subset \mathbb{S} : \sigma^{n-1}(X) \leq \tilde{\varepsilon}\}$. Then, the function $\Delta(X)$ attains its minimum over $\mathcal{X}_{\tilde{\varepsilon}}$ for some X which is a spherical cap. We denote by $\delta(\tilde{\varepsilon})$ this minimal value, which takes the following expression:*

$$\delta(\tilde{\varepsilon}) = \sqrt{1 - I^{-1}(2\tilde{\varepsilon}; \frac{n-1}{2}, \frac{1}{2})}.$$

A proof of Proposition 13 is given in (Kenanian et al., 2018). By homogeneity of Program (8), we have $x \in \tilde{\mathbb{S}} \iff -x \in \tilde{\mathbb{S}}$, which implies that the minimal δ will in fact occur when the set of violating points is the union of two symmetric spherical caps, each of measure $\frac{\tilde{\varepsilon}}{2}$.

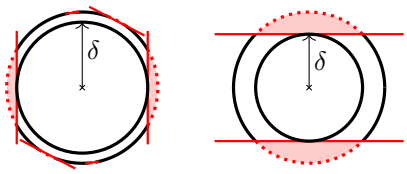

Fig. 2. On the left, a general case of the situation where the ellipse in Fig. 1. is a sphere. On the right, case giving minimal δ. The set of points violating the invariance constraint (in red) is the union of two spherical caps.

Remark 14 *When $\varepsilon \geq \frac{1}{m^l}$, we have $\tilde{\varepsilon} \geq 1$ and $\delta(\tilde{\varepsilon}) = 0$: the only upper bound we can give for the JSR is then $+\infty$.*

4.2 A global upper bound

We now introduce our main theorem, Theorem 15, which provides a solution to Problem 2. In order to use our

solution of previous section, developed for the case $E_P = \mathbb{S}$, we will have to relate $E_{P(\omega_N)}$ to $\mathbb{S}$. We apply thus a change of coordinates bringing E_P to $\mathbb{S}$. Since $P \in \mathcal{S}^n_{++}$, it can be written in its Cholesky form

$$P = L^T L, \tag{12}$$

where L is an upper triangular matrix. Remark that L maps the elements of E_P to $\mathbb{S}$. Since the JSR is not changed by similarity transformations (Jungers, 2009, Proposition 1.3), we can pursue our calculations with the matrices obtained after the change of coordinates.

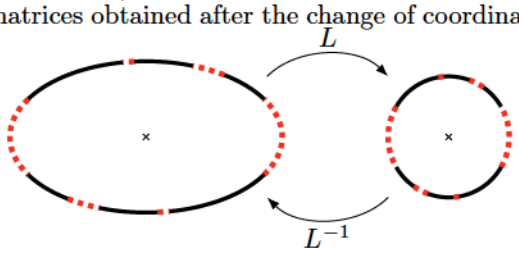

Fig. 3. Change of coordinates to bring our problem back to the case of the unit sphere.

Theorem 15 *Let $\gamma^* \in \mathbb{R}_{>0}$. Consider a set of matrices $\mathcal{M}$, and a matrix $P \succ 0$ optimal solution of Opt(ω_N), satisfying Equation* (10) *for some $\tilde{\mathbb{S}} \subset \mathbb{S}$ where $\sigma^{n-1}(\tilde{\mathbb{S}}) \leq \tilde{\varepsilon}$. Then, we have*

$$\rho(\mathcal{M}) \leq \frac{\gamma^*}{\sqrt[l]{\delta\left(\frac{\tilde{\varepsilon}\kappa(P)}{2}\right)}} \tag{13}$$

with $\kappa(P) = \sqrt{\frac{\det(P)}{\lambda_{\min}(P)^n}}$ and where $\delta(\cdot)$ is given by Proposition 13.

PROOF. Proof. *i)* Since we have seen in the previous section a technique to solve the spherical case, we first bring our problem to the spherical case. To do so, we perform the change of coordinates defined as in (12) by $L \in \mathcal{L}(\mathbb{R}^n)$ which maps the ellipsoid E_P to the sphere $\mathbb{S}$. By defining $\bar{A}_{j_i} = LA_{j_i}L^{-1}$, and $\bar{\mathbf{A}}_{\mathbf{j}} = \bar{A}_{j_l}\bar{A}_{j_{l-1}}\dots\bar{A}_{j_1}$, Equation (10) becomes:

$$(\bar{\mathbf{A}}_{\mathbf{j}}x)^T\bar{\mathbf{A}}_{\mathbf{j}}x \leq (\gamma^*)^{2l}x^Tx,\ \forall x \in L(\mathbb{S}\setminus\tilde{\mathbb{S}}), \forall \mathbf{j} \in M^l. \tag{14}$$

By using the homogeneity of the dynamics, we have:

$$\begin{aligned}&(\bar{\mathbf{A}}_{\mathbf{j}}x)^T\bar{\mathbf{A}}_{\mathbf{j}}x \leq (\gamma^*)^{2l}x^Tx,\ \forall x \in L(\mathbb{S}\setminus\tilde{\mathbb{S}})\\ &\quad\implies (\bar{\mathbf{A}}_{\mathbf{j}}x)^T\bar{\mathbf{A}}_{\mathbf{j}}x \leq (\gamma^*)^{2l}x^Tx,\ \forall x \in \Pi_{\mathbb{S}}\left(L(\mathbb{S}\setminus\tilde{\mathbb{S}})\right),\end{aligned}$$

and therefore we can rewrite (14) as:

$$(\bar{\mathbf{A}}_{\mathbf{j}}x)^T\bar{\mathbf{A}}_{\mathbf{j}}x \leq (\gamma^*)^{2l}x^Tx, \forall x \in \mathbb{S}\setminus\Pi_{\mathbb{S}}(L(\tilde{\mathbb{S}})), \forall \mathbf{j} \in M^l. \tag{15}$$

ii) We now show how to relate $\sigma^{n-1}(\Pi_{\mathbb{S}}(L(\tilde{\mathbb{S}})))$ to $\sigma^{n-1}(\tilde{\mathbb{S}})$, the measure of the violating set in the initial coordinates. Consider $\mathbb{S}^{\tilde{\mathbb{S}}}$, the sector of $\mathbb{B}$ defined by $\tilde{\mathbb{S}}$. We denote $C := L(\tilde{\mathbb{S}})$ and $C' := \Pi_{\mathbb{S}}(L(\tilde{\mathbb{S}}))$. Thus, we have that $\mathbb{S}^{C'} \subset \frac{1}{\lambda_{\min(L)}}\mathbb{S}^C$. This leads to [2]:

$$\sigma^{n-1}(C') = \lambda\left(\mathbb{S}^{C'}\right) \leq \lambda\left(\frac{1}{\lambda_{\min(L)}}L(\mathbb{S}^{\tilde{\mathbb{S}}})\right).$$

Then, the following holds:

$$\begin{aligned}\sigma^{n-1}(C') &\leq \frac{\lambda\left(L(\mathbb{S}^{\tilde{\mathbb{S}}})\right)}{\lambda_{\min}(L)^n}\\ &= \frac{|\det(L)|}{\lambda_{\min}(L)^n}\lambda\left(\mathbb{S}^{\tilde{\mathbb{S}}}\right) &(16)\\ &= \sqrt{\frac{\det(P)}{\lambda_{\min}(P)^n}}\sigma^{n-1}(\tilde{\mathbb{S}}) &(17)\end{aligned}$$

where (16) follows from the fact that $\lambda(Q(X)) = |\det(Q)|\lambda(X)$, for any set $X \subset \mathbb{R}^n$ and $Q \in \mathcal{L}(\mathbb{R}^n)$ (see e.g. (Rudin, 1987)). Hence, we have

$$(\bar{\mathbf{A}}_{\mathbf{j}}x)^T\bar{\mathbf{A}}_{\mathbf{j}}x \leq (\gamma^*)^{2l}x^Tx, \forall x \in \mathbb{S}\setminus\mathbb{S}', \forall \mathbf{j} \in M^l, \tag{18}$$

with $\mathbb{S}' = \Pi_{\mathbb{S}}(L(\tilde{\mathbb{S}}))$ and $\sigma^{n-1}(\mathbb{S}') \leq \sqrt{\frac{\det(P)}{\lambda_{\min}(P)^n}} = \kappa(P)\tilde{\varepsilon}$.

iii) For any such given set $\mathbb{S}'$, we look for the largest sphere included in convhull $(\mathbb{S}\setminus\mathbb{S}')$. By homogeneity of the system, this sphere is centered at the origin, and we denote by α its radius. By (18), finite trajectories of length l initialized on $\mathbb{S}\setminus\mathbb{S}'$ will be in $(\gamma^*)^l\mathbb{B}$:

$$\bar{\mathbf{A}}_{\mathbf{j}}(\mathbb{S}\setminus\mathbb{S}') \subset (\gamma^*)^l\mathbb{B},\ \forall \mathbf{j} \in M^l.$$

Now, combining with Property 2, we have: $\bar{\mathbf{A}}_{\mathbf{j}}(\text{convhull }(\mathbb{S}\setminus\mathbb{S}')) \subset \text{convhull }(\bar{\mathbf{A}}_{\mathbf{j}}(\mathbb{S}\setminus\mathbb{S}')) \subset (\gamma^*)^l\mathbb{B}, \forall \mathbf{j} \in M^l$. Since $\alpha\mathbb{S} \subset \text{convhull }(\mathbb{S}\setminus\mathbb{S}')$, then $\forall \mathbf{j} \in M^l$, $\bar{\mathbf{A}}_{\mathbf{j}}(\alpha\mathbb{S}) = \alpha\bar{\mathbf{A}}_{\mathbf{j}}(\mathbb{S}) \subset \text{convhull }\left(\bar{\mathbf{A}}_{\mathbf{j}}(\mathbb{S}\setminus\mathbb{S}')\right) \subset (\gamma^*)^l\mathbb{B}$, which implies that

$$\bar{\mathbf{A}}_{\mathbf{j}}(\mathbb{S}) \subset \frac{(\gamma^*)^l}{\alpha}\mathbb{B}. \tag{19}$$

iv) Summarizing, since we know that the set $\tilde{\mathbb{S}}$ is symmetric w.r.t. the origin, by Proposition 13, we have that $\alpha \geq \delta(\frac{\tilde{\varepsilon}\kappa(P)}{2})$. Finally, by homogeneity of our

[2] Recall that $\lambda(S)$ is the Lebesgue measure of S, and the spherical measure of any set $C \subset \mathbb{S}$ is given by $\sigma^{n-1}(C) = \lambda\left(\mathbb{S}^C\right)$.

system and the fact that the JSR is invariant under similarity transformations (Jungers, 2009, Proposition 3.1), Equation (19) implies $\rho(\mathcal{M}^l) \leq \frac{(\gamma^*)^l}{\delta(\frac{\tilde{\varepsilon}\kappa(P)}{2})}$, hence $\rho(\mathcal{M}) \leq \frac{\gamma^*}{\sqrt[l]{\delta\left(\frac{\tilde{\varepsilon}\kappa(P)}{2}\right)}}$. $\square$

Remark 16 *There is no conservatism in multiplying ε by m^l, as in the worst case this can happen: if $\varepsilon = 1/m^l$, Theorem 10 does not rule out the pathological case where not a single point satisfies Equation (5) for all $\mathbf{A} \in \mathcal{M}^l$, and thus δ must be equal to 0 since all points might be violating the constraint. However, the multiplication by $\kappa(P)$ is conservative if P has different eigenvalues (this bound is then exactly reached only at a single point on the ellipsoid). We can then, instead of deriving an upper bound on the size of the set of points that violate the constraint, look at a lower bound on the size of the set of points that satisfy the constraint. Taking the complement of this latter set gives another upper bound on the size of the set of violating points. By a similar reasoning as the one conducted above, this second upper bound will be equal to $1 - (1 - \varepsilon m^l)\sqrt{\frac{\det(P)}{\lambda_{\max}(P)^n}}$.*
This provides an alternative upper bound, which can be used if the initial upper bound (13) is infinite, or weaker.

4.3 Main Theorem

We are now ready to prove our main theorem by putting together all the above pieces.

Theorem 17 *Consider an n-dimensional switched linear system as in (2) and a uniform random sampling $\omega_N \subset Z_l$, where $N \geq \frac{n(n+1)}{2} + 1$. For any $\eta > 0$, let $\gamma^*(\omega_N)$ be the optimal solution to (8). Then, for any given $\beta \in [0, 1)$,*

$$\mu_l^N \left\{ \omega_N \in Z_l^N : \rho \leq \frac{\gamma^*(\omega_N)(1+\eta)}{\sqrt[l]{\delta(\beta, \omega_N)}} \right\} \geq \beta,$$

where $\delta(\beta, \omega_N) = \sqrt{1 - I^{-1}(\varepsilon(\beta, N)m^l\kappa(P), \frac{n-1}{2}, \frac{1}{2})}$, with $P \succ 0$ the optimal solution of Opt(ω_N). Moreover, $\lim_{N\to\infty} \delta(\beta, \omega_N) = 1$ with probability 1.

PROOF. Proof. Let us consider $\gamma^*(\omega_N)$ and P as in Equation (8). Then, by taking $\varepsilon := \varepsilon(\beta, N)$ such that $\beta(\varepsilon, N) = \beta$ in Corollary 11, we have with probability at least β

$$(\mathbf{A}_\mathbf{j} x)^T P \mathbf{A}_\mathbf{j} x \leq ((\gamma^*(1+\eta))^{2l}\, x^T P x, \forall x \in \mathbb{S}\backslash\tilde{\mathbb{S}}, \forall \mathbf{j} \in M^l$$

with $\tilde{\mathbb{S}}$ the projection of V on $\mathbb{S}$, and $\sigma^{n-1}(\tilde{\mathbb{S}}) \leq \varepsilon m^l$. Then by Theorem 15, we can compute $\delta(\beta, \omega_N) = \delta(\varepsilon'(\beta, N))$, where

$$\varepsilon'(\beta, N) = \frac{1}{2}\varepsilon(\beta, N)m^l\kappa(P) \tag{20}$$

such that with probability at least β we have:

$$\rho \leq \frac{\gamma^*(\omega_N)(1+\eta)}{\sqrt[l]{\delta(\beta, \omega_N)}},$$

which completes the proof of the first part of the theorem.

Let us prove now that $\lim_{N\to\infty} \delta(\beta, \omega_N) = 1$ with probability 1. We recall that

$$\delta(\beta, \omega_N) = \delta\left(\varepsilon(\beta, \omega_N)m^l\kappa(P(\omega_N))\right).$$

We start by showing that $\kappa\left(P(\omega_N)\right)$ is uniformly bounded in N. The optimization problem Opt(ω_N) given in (8), with (ω_N) replaced by (Z_l) and $(1+\eta)$ replaced by $(1 + \frac{\eta}{2})$ is strictly feasible for any positive parameter η. It then admits a finite optimal value K for some solution $P_{\eta/2}$. Note that, $\lim_{N\to\infty} \gamma^*(\omega_N) = \gamma^*(Z_l)$ with probability 1. Thus, for large enough N, $\gamma^*(\omega_N)(1+\eta) > \gamma^*(Z_l)(1+\frac{\eta}{2})$. This also means that, for large enough N, Opt(ω_N) admits $P_{\eta/2}$ as a feasible solution and thus the optimal value of Opt(ω_N) is upper-bounded by K. In other words, for N large enough, $\lambda_{\max}(P(\omega_N)) \leq K$. Moreover, since $\lambda_{\min}(P(\omega_N)) \geq 1$ (by $P \succeq I$), we also have:

$$\kappa\left(P(\omega_N)\right) = \sqrt{\frac{\det(P(\omega_N))}{\lambda_{\min}(P(\omega_N))^n}} \leq \sqrt{K^n}. \tag{21}$$

We next show that $\lim_{N\to\infty} \varepsilon(\beta, N) = 0$ for any fixed $\beta \in [0, 1)$. Recall that $\varepsilon(\beta, N)$ is implicitely defined by

$$\begin{aligned} 1 - \beta &= \sum_{j=0}^{d} \binom{N}{j} \varepsilon^j (1-\varepsilon)^{N-j} \\ &\leq (d+1)N^d(1-\varepsilon)^{N-d}. \end{aligned} \tag{22}$$

We prove $\lim_{N\to\infty} \varepsilon(\beta, N) = 0$ by contradiction. Assume that $\lim_{N\to\infty} \varepsilon(\beta, N) \neq 0$. This means that, there exists some $c > 0$ such that $\varepsilon(\beta, N) > c$ infinitely often. Then, consider the subsequence N_k such that $\forall k$, $\varepsilon(\beta, N_k) > c$. Then, by (22) we have for any $k \in \mathbb{N}$:

$$1 - \beta \leq (d+1)N_k^d(1-\varepsilon)^{N_k-d} \leq (d+1)N_k^d(1-c)^{N_k-d}.$$

Note that $\lim_{k\to+\infty}(d+1)N_k^d(1-c)^{N_k-d} = 0$, which implies that there exists a k' such that:

$$(d+1)N_{k'}^d(1-c)^{N_{k'}-d} < 1 - \beta,$$

which is a contradiction. Therefore, we must have $\lim_{N\to\infty} \varepsilon(\beta, N) = 0$. Putting this together with (21), we get: $\lim_{N\to\infty} m^l \kappa(P(\omega_N))\varepsilon(\beta, \omega_N) = 0$. By the continuity of the function δ this also implies: $\lim_{N\to\infty} \delta\left(\varepsilon(\beta, \omega_N) m^l \kappa(P(\omega_N))\right) = 1$. □

5 Experimental Results

Theorem 9 and Theorem 17 give us a straightforward algorithm which is summarized at the end of Section 2. In its first part, we look for γ^* by bisection on an interval $[0, U]$ (for the value of U, take, e.g., the maximum value of $||x_{k+l}||$ among the observations made). For a fixed desired accuracy α on that bisection, we solve a feasibility problem (of polynomial complexity in the number of constraints) at most $\lceil \log_2(U/\alpha) \rceil$ times. In our experiments we took $\alpha = 10^{-3}$. Once the result of the bisection is obtained, we solve $\mathrm{Opt}(\omega_N)$. In practice, the parameter η in $\mathrm{Opt}(\omega_N)$ can be put to zero, as it is included in α. Finally, we get δ by using the expression given in Theorem 15. All these computations are also of polynomial complexity.

We illustrate our technique on a 4-dimensional switched system with 6 modes. We fix the confidence level, $\beta = 0.95$, and compute the lower and upper bounds on the JSR for $N := 20 + 200k$, $k \in \{0, \ldots, 29\}$, according to Theorem 9 and Theorem 17, respectively. We take the average performance of our algorithm over 10 different runs. Fig. 4 shows the evolution, as N increases, of the upper and lower bounds for various values of trajectories length l. To further demonstrate the practical performance of our technique, we also provide the true value of the JSR approximated by the JSR Toolbox (Vankeerberghen et al., 2014) for this system, which turns out to be 0.918 ± 0.001. We observed that the performance of the upper bound is much better for trajectories of length 1, while for the lower bound, we benefit by considering trajectories of higher length. While it is expected that longer trajectories improve the accuracy, the decreasing performance for the upper bound comes from the fact that many more points are needed for larger trajectories, because the probability space to be sampled is larger. In our example, our first upper bound smaller than 1 (that is, being a stability guarantee) was obtained for $N = 5820$.

In Fig. 5, we compare the upper bound we obtained with the upper bound given by the (white box) JSR Toolbox, for different values of n and m. Note that the speed of convergence of all the quantities considered decreases when the dimension of the system increases. We nevertheless observe convergence of the upper bound to $\rho(\mathcal{M})$, and convergence of the lower bound to $\frac{\rho(\mathcal{M})}{\sqrt[l]{n}}$. The gap between these two limits is $\frac{\rho}{\sqrt[l]{n}}$ as predicted by Theorem 17. This gap could be improved by considering a more general class of common Lyapunov functions, such

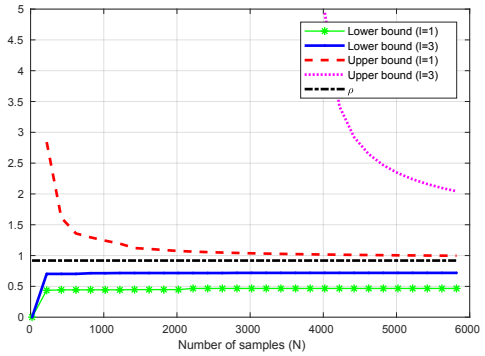

Fig. 4. Evolution of the upper and lower bounds, for various lengths of trajectories, with the number of samples.

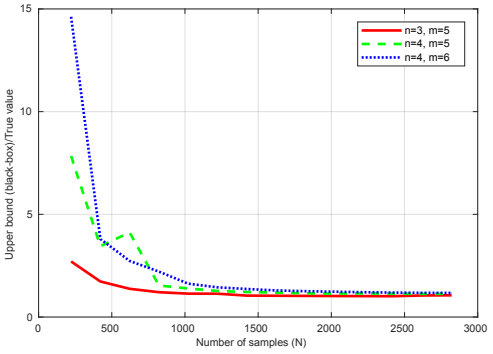

Fig. 5. Convergence of our upper bound when the number of samples increases, for several values of n and m and $l = 1$. The values plotted are the ratios between our upper bound and the true value computed by the JSR Toolbox.

as those that can be described by sum-of-squares polynomials (Papachristodoulou and Prajna, 2002). We leave this for future work.

To illustrate the accuracy of our confidence level beta, we randomly generate 10,000 test cases with systems of dimension between 2 and 7, number of modes between 2 and 6, and size of samples N between 30 and 1000. We take $\beta = 0.95$ and we check if the upper bound computed by our technique is greater than the true value computed by the JSR Toolbox for the system. We get 9921 positive tests, out of 10,000, which gives us a correctness rate of 0.9921 for the upper bound computed. Note that, this probability is significantly above the provided β. This is expected, since our techniques are based on worst-case analysis and thus fairly conservative.

Finally, Fig. 6 shows the evolution of the function δ with the number of samples, for different values of n, at m and l fixed.

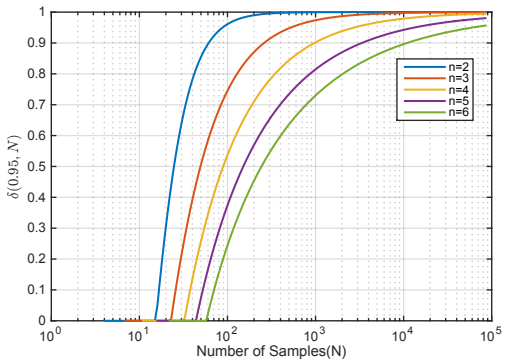

Fig. 6. Average behavior of δ as a function of N for different values of n, with fixed $m = 4$ and $l = 1$.

6 Conclusions

In this paper, we investigated the question of how one can conclude stability of a dynamical system when a model is not available and, instead, we only can observe the evolution of the state of the system. Our goal was to understand how the observation of well-behaved trajectories *intrinsically* implies stability of a system.

As expected, it is not surprising that we need some standing assumptions on the system, in order to allow for any sort of nontrivial stability certificate solely from a finite number of observations.

The novelty of our contribution is twofold: First, we used as standing assumption that the unknown system can be modeled by a switched linear system. This assumption covers a wide range of systems of interest, and to our knowledge no such "black-box" result has been available so far on switched systems. Second, we applied powerful techniques from chance-constrained optimization. Their application was far from obvious, though, and relied on geometric properties of linear switched systems.

We leveraged the concept of 'l-step CQF', and showed that it allows to reach arbitrary precision for our black-box technique. In the switched systems literature, there are other well-known techniques for refining this precision for the white-box problem. For instance, one can replace the LMIs in Theorem 5 by Sum-Of-Squares (SOS) constraints; see (Parrilo and Jadbabaie, 2008) or (Jungers, 2009, Theorem 2.16). Although l-step CQFs seem better suited for our purpose, we leave for further work a more systematic analysis of the behaviour of the different refining techniques.

Notice that, our algorithm can also be used in the white-box framework and becomes then a randomized algorithm to evaluate the JSR of a known system.

In our view, the stability-like guarantees obtained are powerful, in view of the hardness of the general problem. In the future, we plan to investigate how to generalize our results to more complex models of realistic systems.

References

Balkan, A., Tabuada, P., Deshmukh, J.V., Jin, X., Kapinski, J., 2016. Underminer: A Framework for Automatically Identifying Non-converging Behaviors in Black Box System Models, in: Proceedings of the Thirteenth ACM International Conference on Embedded Software (EMSOFT), pp. 1–10. doi:10.1145/2968478.2968487.

Blanchini, F., Fenu, G., Giordano, G., Pellegrino, F.A., 2017. Model-Free Plant Tuning. IEEE Transactions on Automatic Control 62, 2623–2634. doi:10.1109/TAC.2016.2616025.

Blondel, V.D., Tsitsiklis, J.N., 1999. Complexity of Stability and Controllability of Elementary Hybrid Systems. Automatica 35, 479–489. doi:10.1016/S0005-1098(98)00175-7.

Bobiti, R., Lazar, M., 2015. A Delta-sampling Verification Theorem for Discrete-time, Possibly Discontinuous Systems, in: Proceedings of the 18th International Conference on Hybrid Systems: Computation and Control, ACM. pp. 140–148. doi:10.1145/2728606.2728631.

Boyd, S., Vandenberghe, L., 2004. Convex Optimization. Cambridge University Press. doi:10.1017/CBO9780511804441.

Calafiore, G., 2010. Random Convex Programs. SIAM Journal on Optimization 20, 3427–3464. doi:10.1137/090773490.

Calafiore, G., Campi, M.C., 2006. The Scenario Approach to Robust Control Design. IEEE Transactions on Automatic Control 51, 742–753. doi:10.1109/TAC.2006.875041.

Campi, M.C., Garatti, S., 2008. The Exact Feasibility of Randomized Solutions of Uncertain Convex Programs. SIAM Journal on Optimization 19, 1211–1230. doi:10.1137/07069821X.

Duggirala, P.S., Mitra, S., Viswanathan, M., 2013. Verification of Annotated Models from Executions, in: Proceedings of the Eleventh ACM International Conference on Embedded Software, pp. 26:1–26:10. doi:10.1109/EMSOFT.2013.6658604.

Huang, Z., Mitra, S., 2014. Proofs from Simulations and Modular Annotations, in: Proceedings of the 17th International Conference on Hybrid Systems: Computation and Control, ACM. pp. 183–192. doi:10.1145/2562059.2562126.

Jungers, R., 2009. The Joint Spectral Radius: Theory and Applications, in: Lecture Notes in Control and Information Sciences. Springer-Verlag. volume 385. doi:10.1007/978-3-540-95980-9.

Kapinski, J., Deshmukh, J.V., Sankaranarayanan, S., Arechiga, N., 2014. Simulation-guided Lyapunov Analysis for Hybrid Dynamical Systems, in: Proceedings of the 17th International Conference on Hybrid

Systems: Computation and Control, ACM. pp. 133–142. doi:10.1145/2562059.2562139.

Kenanian, J., Balkan, A., Jungers, R.M., Tabuada, P., 2018. Data Driven Stability Analysis of Black-box Switched Linear Systems. arXiv preprint arXiv:1803.03687 .

Kozarev, A., Quindlen, J., How, J., Topcu, U., 2016. Case Studies in Data-Driven Verification of Dynamical Systems, in: Proceedings of the 19th International Conference on Hybrid Systems: Computation and Control, ACM. pp. 81–86. doi:10.1145/2883817.2883846.

Lauer, F., 2016. On the Complexity of Switching Linear Regression. Automatica 74, 80–83. doi:10.1016/j.automatica.2016.07.027.

Liu, S., Liberzon, D., Zharnitsky, V., 2016. On Almost Lyapunov Functions for Non-Vanishing Vector Fields, in: Proceedings of the 55th IEEE Conference on Decision and Control (CDC), pp. 5557–5562. doi:10.1109/CDC.2016.7799123.

Majumder, K.L., Bhattacharjee, G.P., 1973. Algorithm AS 64: Inverse of the Incomplete Beta Function Ratio. Journal of the Royal Statistical Society. Series C (Applied Statistics) 22, 411–414. doi:10.2307/2346798.

Nemirovski, A., Shapiro, A., 2006. Convex Approximations of Chance Constrained Programs. SIAM Journal on Optimization 17, 969–996. doi:10.1137/050622328.

Olver, F.W., Lozier, D.W., Boisvert, R.F., Clark, C.W., 2010. NIST Handbook of Mathematical Functions. Cambridge university press. doi:10.1111/j.1751-5823.2011.00134_18.x.

Papachristodoulou, A., Prajna, S., 2002. On the Construction of Lyapunov Functions Using the Sum of Squares Decomposition, in: Proceedings of 41st IEEE Conference on Decision and Control (CDC), pp. 3482–3487. doi:10.1109/CDC.2002.1184414.

Parrilo, P.A., Jadbabaie, A., 2008. Approximation of the Joint Spectral Radius Using Sum of Squares. Linear Algebra and its Applications 428, 2385–2402. doi:10.1016/j.laa.2007.12.027.

Rudin, W., 1987. Real and Complex Analysis. 3rd ed., McGraw-Hill Book Co.

Topcu, U., Packard, A., Seiler, P., 2008. Local Stability Analysis Using Simulations and Sum-of-Squares Programming. Automatica 44, 2669–2675. doi:10.1016/j.automatica.2008.03.010.

Vankeerberghen, G., Hendrickx, J., Jungers, R., 2014. Jsr: A Toolbox to Compute the Joint Spectral Radius, in: Proceedings of the 17th International Conference on Hybrid Systems: Computation and Control, ACM. pp. 151–156. doi:10.1145/2562059.2562124.