Data Driven Stability Analysis of Black-box Switched Linear Systems with Probabilistic Guarantees

***Abstract*— We explore the general problem of deciding stability of a "black-box system", that is, a system whose model equations are not known. The only information at our possession is a set of observations, that is, couples of vectors of the type $(x(k), x(k+1))$. We adopt a probabilistic approach, and focus on switched systems, which are a widely used model for many complex systems, and are well known to be hard to analyze, even in a non-'black-box setting'.**

We show that, for a given (randomly generated) set of observations, one can give a stability guarantee on the system, for some level of confidence, with a trade-off between the quality of the guarantee and the level of confidence. We provide an explicit way of computing the best stability guarantee, as a function of the number of observations, and the required level of confidence. Our results rely on a geometrical analysis, combining chance-constrained optimization theory with stability analysis tools for switching systems.

I. INTRODUCTION

Today's complex cyber-physical systems are characterized by the interaction of a large number of heterogeneous components. Consequently, the models used to analyze these systems are equally complex and consist of heterogeneous sub-models relying on different assumptions and based on principles from different scientific disciplines. It is not uncommon to encounter a patchwork of differential equations, difference equations, hybrid automata, lookup tables, custom switching logic, low-level legacy code, etc. To further compound the difficulty in analyzing these systems, different components of a complex engineered system are typically designed by different suppliers. Although a high-level specification for these components may be known, detailed models are not available for intellectual property reasons. We are thus faced with a tremendous gap between the existing analysis techniques that rely on closed-form models and the models available in industry. It is, therefore, not surprising the emphasis that industry places on simulation since despite the complexity of models, it is always possible to simulate them. This raises the question of whether we can provide formal guarantees about certain properties of these complex systems based solely on the information obtained via their simulations. In this paper, we focus on one of the most important of such properties in the context of control theory: stability.

More formally, we consider a dynamical system as in:

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

where, $x_k \in X$ is the state and $k \in \mathbb{N}$ is the time index. In this paper, we focus on switched systems, but we are interested in more general forms for the function f as well. We start with the following question to serve as a stepping stone: Given N pairs, (x_1, y_1), (x_2, y_2), $\ldots$, (x_N, y_N) belonging to the behavior of the System (1), (i.e., $y_i = f(k, x_i)$) for some k), what can we say about the stability of the system (1)? For the rest of the paper, we use the term *black-box* to refer to models where we do not have access to its dynamics, yet we can observe f by observing couple of points (x_k, x_{k+1}) as defined in (1).

A potential approach to this problem could be firstly identifying the dynamics, i.e., the function f, and then applying existing techniques from the model-based stability analysis literature. However, unless f is a linear function, there are two main reasons behind our quest to directly work on input-output pairs and bypassing the identification phase: (1) Even when the function f is known, in general, the stability analysis is a very difficult problem [**?**], [**?**]. (2) Identification can also potentially add approximation errors, and can be algorithmically hard. Again, this is the case for switched systems [9]. A fortiori, the combination of these two steps in an efficient and robust way seem far from obvious.

The initial idea behind this paper was influenced by the recent efforts in [13], [8] and [1] in using simulation traces to find Lyapunov functions for systems with known dynamics. Will put Liberzon and Sayan Mitra here. 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 also 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 [13] and [8], or via sampling based techniques as in [1]. Note that, since we do not have access to the dynamics, the second step cannot be directly applied to black-box systems. However, these sampling based ideas raise the following question 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 such candidate Lyapunov function is a bona-fide Lyapunov function. *Can we translate this confidence into a confidence that this Lyapunov function decreases at many points in the state space?*

Note that, even in the case of a 2D linear system the connection between these two beliefs is nontrivial. In fact, one can easily construct an example with one stable and one unstable eigenvalue for which even though almost all trajectories diverge to the infinity, it is possible to construct a Lyapunov function candidate whose level sets are contracting everywhere except a small set. Moreover, the size of this

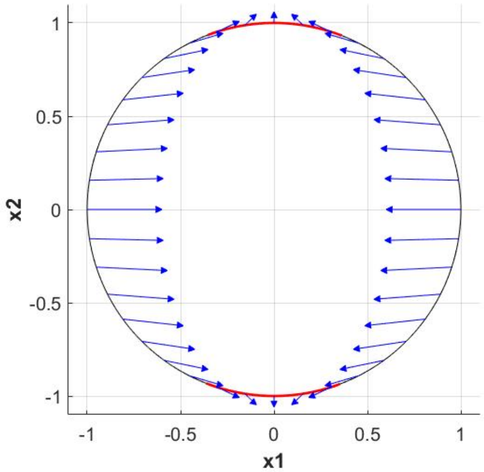

Fig. 1. A simple dynamics and the level set of an 'almost Lyapunov function'. Even though the function decreases at almost all points in the state space, all trajectories are unstable.

"violating set" can be arbitrarily small based on the magnitude of the unstable eigenvalue (see Fig I). The system

$$x^+ = \begin{bmatrix} -2 & 0 \\ 0 & 0.3 \end{bmatrix} x$$

admits a Lyapunov function candidate on the unit circle except the two red areas.

In this paper, we take the a step to close this gap by focusing on switched linear systems. The identification and deciding the stability of arbitrary switched linear systems is NP-hard [7]. Aside from their theoretical value, switched systems are a popular model for many complex systems, as for instance dynamics with (known or unknown) varying parameters. These parameters can model internal properties of the dynamical system such as uncertainties, look-up tables, values in a discrete register as well as exogenous inputs provided by a controller in a closed-loop control system [11], [5].

The stability of switched systems as in Equation (3) is closely related to the *joint spectral radius* (JSR) of the matrices appearing in (3). Deciding stability amounts to deciding whether the JSR is less than one [7]. In this paper, we present an algorithm to bound the JSR of a switched linear system from a finite number N of observations. This algorithm partly relies on tools from the random convex optimization literature (also known as chance-constrained optimization, see [3] [12], [4]), 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.

The organization of the paper is as follows: TO BE FILLED.

II. Preliminaries

A. 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 the set of linear functions in $\mathbb{R}^n$ 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, which are matrices $P \in \mathcal{S}^n$ such that $\forall\ x \in \mathbb{R}^n \setminus \{0\}$, $x^T P x > 0$, is denoted by $\mathcal{S}^n_{++}$. We write $P \succ 0$ to state that P is positive definite. Given a set $X \subset \mathbb{R}^n$, and $r \in \mathbb{R}_{>0}$ we write $rX := \{x \in X : rx\}$ to denote the scaling of this set. We denote by B (respectively 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., $E_P := \{x \in \mathbb{R}^n : x^T P x = 1\}$, and we denote by $\tilde{E}_P$ the volume in $\mathbb{R}^n$ defined by E_P: $\tilde{E}_P = \{x \in \mathbb{R}^n : x^T P x \leq 1\}$. We denote the spherical projector on S by Π_{S}.

For an ellipsoid centered at the origin, and for any of its subsets $\mathcal{A}$, the *sector* defined by $\mathcal{A}$ is the subset

$$\{t\mathcal{A},\ t \in [0,1]\} \subset\ \mathbb{R}^n.$$

A sector induced by $\mathcal{A} \subset E_P$ will be denoted by $E_P^{\mathcal{A}}$. In the particular case of the unit sphere, we instead write $S^{\mathcal{A}}$.

We consider in this work the classical unsigned and finite uniform spherical measure on S, denoted by σ^{n-1}. It is associated to $\mathcal{B}_{\mathrm{S}}$, the spherical Borelian σ-algebra, and is derived from the Lebesgue measure λ. We have $\mathcal{B}_{\mathrm{S}}$ defined by $\mathcal{A} \in \mathcal{B}_{\mathrm{S}}$ if and only if $S^{\mathcal{A}} \in \mathcal{B}_{\mathbb{R}^n}$. The spherical measure σ^{n-1} is defined by

$$\forall\ \mathcal{A} \in \mathcal{B}_{\mathrm{S}},\ \sigma(\mathcal{A}) = \frac{\lambda(S^{\mathcal{A}})}{\lambda(B)}.$$

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 $\sigma^{n-1}(\mathrm{S}) = 1$. Since $P \in \mathcal{S}^n_{++}$, it can be written in its Choleski form $P = UDU^{-1}$, where D is the diagonal matrix of its eigenvalues and U is an orthogonal matrix [2]. Let

$$L := UD^{1/2}U^{-1}. \tag{2}$$

Note that, L^{-1} maps the elements of S to E_P. Then, we define the measure on the ellipsoid σ_P on the σ-algebra $\mathcal{B}_{E_P} := L^{-1}\mathcal{B}_{\mathrm{S}}$, where $\forall\, \mathcal{A} \in \mathcal{B}_{E_P}$, $\sigma_P(\mathcal{A}) = \sigma^{n-1}(L\mathcal{A})$.

For $m \in \mathbb{N}_{>0}$, we denote by M the set $M = \{1, 2, \ldots, m\}$. Set M is provided with the classical σ-algebra associated to the finite sets: $\Sigma_M = \wp(M)$, where $\wp(M)$ is the set of subsets of M. We consider the uniform measure μ_M on (M, Σ_M).

We define $Z = \mathrm{S} \times M$ as the Cartesian product of the unit sphere and M. We denote the product σ-algebra $\mathcal{B}_{\mathrm{S}} \bigotimes \Sigma_M$ generated by $\mathcal{B}_{\mathrm{S}}$ and Σ_M: $\Sigma = \sigma(\pi_{\mathrm{S}}^{-1}(\mathcal{B}_{\mathrm{S}}), \pi_M^{-1}(\Sigma_M))$. On this set, we define the product measure $\mu = \sigma^{n-1} \otimes \mu_M$. Note that, μ is a uniform measure on Z and $\mu(Z) = 1$.

B. Stability of Linear Switched Systems

A *switched linear system* related to a set of modes $\mathcal{M} = \{A_i, i \in M\}$ is of the form:

$$x_{k+1} = A_{\tau(k)} x_k, \tag{3}$$

with switching sequence $\tau : \mathbb{N} \to M$. There are two important properties of linear switched systems that we exploit in this paper.

Property 2.1: Let $\xi(x, k, \tau)$ denote the state of the system (3) at time k starting from the initial condition x and with switching sequence τ. The dynamical system (3) is homogeneous:

$$\xi(\gamma x, k, \tau) = \gamma \xi(x, k, \tau).$$

Property 2.2: The dynamics given in (3) is convexity-preserving, meaning that for any set of points $X \subset \mathbb{R}^n$ we have:

$$f(\text{convhull }(X)) \subset \text{convhull }(f(X)).$$

The joint spectral radius of the set of matrices $\mathcal{M}$ is defined as follows:

Definition [6] Given a finite set of matrices $\mathcal{M} \subset \mathbb{R}^{n\times n}$, its *joint spectral radius* (JSR) is given by

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

Property 2.3 (Corollary 1.1, [7]): Given a finite set of matrices $\mathcal{M}$, the corresponding switched dynamical system is stable if and only if $\rho(\mathcal{M}) < 1$.

Property 2.4 (Proposition 1.3, [7]): Given a finite of matrices $\mathcal{M}$, and any invertible matrix T,

$$\rho(\mathcal{M}) = \rho(T\mathcal{M}T^{-1}),$$

i.e., JSR is invariant under similarity transformations (and is a fortiori a homogeneous function: $\forall \gamma > 0$, $\rho\left(\frac{\mathcal{M}}{\gamma}\right) = \frac{\rho(\mathcal{M})}{\gamma}$.

III. A Deterministic Lower Bound for JSR

CQLFs are useful because they can be computed (if they exist) with Semidefinite Programming (see [2]), and they constitute a stability guarantee for the switched system as the following theorem formalizes:

Theorem 3.1: [6, Prop. 2.8] Consider a bounded set of matrices $\mathcal{M}$. If there exists a $\gamma \geq 0$ and $P \succ 0$ such that

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

then $\rho(\mathcal{M}) \leq \gamma$.

Even though an upper bound is more difficult to obtain in the black box model where only a finite number of observations are available, in this section we leverage Theorem 3.1 in order to derive an easy lower bound in our setting.

Theorem 3.2: [6, Theorem 2.11] For any bounded set of matrices such that $\rho(\mathcal{M}) < \frac{1}{\sqrt{n}}$, there exists a Common Quadratic Lyapunov Function (CQLF) for $\mathcal{M}$, that is, a $P \succ 0$ such that:

$$\forall A \in \mathcal{M},\ A^T P A \preceq P.$$

The following theorem shows that the existence of a CQLF for (3) can be checked by considering N pairs $(x_i, j_i) \in \mathbb{R}^n \times M$, where $i \in \{1, \dots N\}$. Recall that in our setting, we assume that we observe pairs of the form (x_t, x_{t+1}), but we do not observe the mode applied to the system during this time step.

Theorem 3.3: For a given homogenous sampling:

$$\omega_N := \{(x_1, j_1), (x_2, j_2), \dots, (x_N, j_N)\} \subset \mathbb{R}^n \times M,$$

and let $S_{\omega_N} = \{(x_1, y_1), ..., (x_N, y_N)\}$ be the corresponding observations available, satisfying

$$y_i = A_{j_i} x_i.$$

Defining $\gamma^*(\omega_N)$ be the optimal solution of the following optimization problem:

$$\begin{array}{ll} \min & \gamma \\ \text{s.t.} & (y_i)^T P (y_i) \leq \gamma^2 x_i^T P x_i,\ \forall\, i : 1 \leq i \leq N. \\ & P \succ 0 \end{array} \tag{4}$$

Then, we have:

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

Note that, (4) can be efficiently solved by semidefinite programming and bisection on gamma [2].

Proof: By definition of γ^*, for any $\epsilon > 0$, there exists no $P \succ 0$ such that (4) is satisfied. By using Remark 2.4 this means that, there exists no CQLF for the scaled set of matrices $\frac{\mathcal{M}}{(\gamma^*(\omega_N) - \epsilon)}$. Then, by Theorem 3.2 we get:

$$\frac{\rho(\mathcal{M})}{\gamma^*(\omega_N)} \geq \frac{1}{\sqrt{n}}.$$

■

IV. A Probabilistic Stability Guarantee

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 a CQLF which is valid with probability at least β. Note that, the existence of a CQLF implies $\rho \leq 1$ due to Theorem 3.1 and due to Property 2.1, it is enough to show that the CQLF is decreasing on a set enclosing the origin, e.g. the unit sphere S. Therefore, to obtain an upper bound on ρ, we consider the following optimization problem:

$$\begin{array}{ll} \min_{\gamma, P} & \gamma \\ \text{s.t.} & (Ax)^T P (Ax) \leq \gamma^2 x^T P x,\ \forall A \in \mathcal{M},\ \forall\, x \in \mathrm{S}, \\ & P \succ 0. \end{array} \tag{5}$$

Note that, for all $A \in \mathcal{M}$ the optimal P and the optimal γ for the Problem (5) satisfies: $\frac{A}{\gamma}^T P \frac{A}{\gamma} \preceq P$. Therefore, $\rho\left(\frac{\mathcal{M}}{\gamma}\right) \leq 1$, which leads to the upper bound $\rho\left(\mathcal{M}\right) \leq \gamma$. However, solving the optimization problem (5) is impossible for us since it involves infinitely many constraints. We now analyze the relationship between the solutions of the optimization problem (5) and the following optimization problem with finitely many constraints:

$$\begin{array}{ll} \min_P & \lambda_{\max}(P) \\ \text{s.t.} & (A_j x)^T P (A_j x) \leq ((1+\eta)\gamma^*)^2 x^T P x,\ \forall (x, j) \in \omega_N, \\ & P \succeq I. \end{array} \tag{6}$$

Recall that ω_N is an N-uniform random sampling of the set Z, $\eta > 0$, and γ^* is the optimal solution to the optimization problem (**??**). For the rest of the discussion, we refer to the optimization problem (6) by $\mathrm{Opt}(\omega_N)$. We denote its optimal solution by $P(\omega_N)$ and $\gamma^*(\omega_N)$. We drop the explicit dependence of P on ω_N when it is clear from the context. There are a few points that are worth noting about (6). Firstly, due to Property 2.1, we are can replace the constraint $P \succ 0$ with the constraint $P \succeq I$. Moreover, for reasons that will become clear later in the discussion, we chose the objective function as $\lambda_{\max}(P)$, instead of solving a feasibility problem in P. Lastly, the additional η factor is introduced to ensure strict feasibility of (6), which will be helpful in the following discussion.

The curious question whether the optimal solution of the sampled problem $\mathrm{Opt}(\omega_N)$ is a feasible solution to (5) has been widely studied in the literature [3]. It turns out that under certain technical assumptions, one can bound the amount of constraints of the original problem (5) that are violated by the optimal solution of (6), with some with some probability which is a function of the sample size N.

Definition (from [3]) Consider the optimization problem $\mathrm{Opt}(\omega_N)$ where the set of constraints ω_N is randomly sampled according to a measure μ. We define the *constraint violation probability* as:

$$\mathcal{V}^*(\omega_N) = \mathbb{P}\{z \in Z : f(P(\omega_N), z) > 0\}. \tag{7}$$

Note that, since we have $\mathbb{P}(\mathcal{A}) = \frac{\mu(\mathcal{A})}{\mu(Z)}$, we can rewrite (7) as:

$$\mathcal{V}^*(\omega_N) = \frac{\mu(V(\omega_N))}{\mu(Z)},$$

where $V(\omega_N) := \{z \in Z : f(P(\omega_N), z) > 0\}$, i.e., the set of points for which at least one constraint is violated for the sampling ω_N.

Theorem 4.1: Let $d := \frac{n(n+1)}{2} + 1$ and $N \geq d + 1$. Consider the optimization problem $\mathrm{Opt}(\omega_N)$ given in (6), where ω_N is a uniform random sampling of the set Z. For all $\epsilon \in (0, 1)$ the following holds:

$$\mathbb{P}^N\{\mu(V(\omega_N)) \leq \epsilon\} \geq 1 - \sum_{j=0}^{d} \binom{N}{j} \epsilon^j (1-\epsilon)^{N-j}. \tag{8}$$

Before proceeding to the proof, we note that this theorem is an immediate application of Theorem 3.3 in [3], which comes with two technical assumptions on $\mathrm{Opt}(\omega_N)$we do not explicitly state here. The reason is that, even if these assumptions do not hold for the optimization problem 9, it is possible to obtain a slightly modified problem for which they hold and solve that problem instead. We refer the interested reader to [3] for a more detailed discussion of such modification techniques.

Proof: Note that the optimization problem $\mathrm{Opt}(\omega_N)$ can be written as:

$$\begin{array}{ll} \min_{P,t} & t \\ \text{s.t.} & f_{\gamma^*}(P, z) \leq 0, \; \forall \; z \in Z \end{array} \tag{9}$$

where $g_{\gamma^*}(P, z) = \max(g_1(P, z), g_2(P), g_3(P))$, and

$$\begin{aligned} g_1(P, z) &:= (A_j z)^T P (A_j z) - {\gamma^*}^2 z^T P z \\ g_2(P) &:= \lambda_{\max}(-P) + 1. \\ g_3(P) &:= \lambda_{\max}(P) - t. \end{aligned}$$

■

The objective function of (9) is linear while each constraint is convex in P for all $z \in Z$. Also note that, the set of decision variables are in $\mathbb{R}^{\frac{n(n+1)}{2}+1}$. Then, we can invoke Theorem 3.3 in [3] with the optimization problem (9) to conclude the statement of the theorem.

Theorem 4.1 states that the optimal solution of the sampled problem $\mathrm{Opt}(\omega_N)$ violates note more than an ϵ fraction of the constraints in the original optimization problem (5) with probability β, where β goes to 1 as N goes to infinity.

Theorem 4.2: Let $\gamma \in \mathbb{R}_{>0}$. Consider a set of matrices $A \in \mathcal{M}$, and a matrix P satisfying:

$$(A_j x)^T P (A_j x) \leq \gamma^2 x^T P x, \; \forall \, (x, j) \in Z \setminus V, \tag{10}$$

for some $V \subset Z$ where $\mu(V) \leq \epsilon$. Then, defining L as in (2) and $\bar{A}_j = L^{-1} A_j L$, one also has:

$$(\bar{A}_j x)^T (\bar{A}_j x) \leq \gamma^2 x^T x, \; \forall \, x \in \mathrm{S} \setminus \mathrm{S}', \forall \, j \in M,$$

for some $\mathrm{S}' \subset \mathrm{S}$ such that:

$$\sigma(\mathrm{S}') \leq m \epsilon \kappa(P),$$

where $\kappa(P) = \sqrt{\frac{\lambda_{\max}(P)}{\det(P)^n}}$.

Proof: Note that $V \subset \Sigma$. Let $V_{\mathrm{S}} = \pi_{\mathrm{S}}(V)$ and $V_M = \pi_M(V)$. We notice that Σ_M is the disjoint union of its 2^m elements $\{\mathcal{M}_i, i \in \{1, 2, \ldots 2^m\}\}$. Then V can be written as the disjoint union $V = \sqcup_{1 \leq i \leq 2^m} (\mathcal{S}_i, \mathcal{M}_i)$ where $\mathcal{S}_i \in \Sigma(\mathrm{S})$. We notice that $V_{\mathrm{S}} = \sqcup_{1 \leq i \leq 2^m} \mathcal{S}_i$, and

$$\sigma^{n-1}(V_{\mathrm{S}}) = \sum_{1 \leq i \leq 2^m} \sigma^{n-1}(\mathcal{S}_i).$$

We have

$$\begin{aligned} \mu(V) &= \mu(\sqcup_{1 \leq i \leq 2^m} (\mathcal{S}_i, \mathcal{M}_i)) \\ &= \sum_{1 \leq i \leq 2^m} \mu(\mathcal{S}_i, \mathcal{M}_i) \\ &= \sum_{1 \leq i \leq 2^m} \sigma^{n-1} \otimes \mu_M(\mathcal{S}_i, \mathcal{M}_i) \\ &= \sum_{1 \leq i \leq 2^m} \sigma^{n-1}(\mathcal{S}_i) \mu_M(\mathcal{M}_i). \end{aligned}$$

Note that we have

$$\min_{j \in M} \mu_M(\{j\}) = \frac{1}{m}.$$

Then since $\forall \; i$, $\mu_M(\mathcal{M}_i) \geq \frac{1}{m}$, we get:

$$\sigma^{n-1}(V_{\mathrm{S}}) \leq \frac{\mu(V)}{\frac{1}{m}} \leq m\epsilon. \tag{11}$$

This means that

$$(A_j x)^T P (A_j x) \leq \gamma^2 x^T P x, \; \forall \, x \in \mathrm{S} \setminus V_S, \forall \, m \in M, \tag{12}$$

where $\sigma^{n-1}(V_S) \leq m\epsilon$.

We then perform the change of coordinates defined by $L^{-1} \in \mathcal{L}(\mathbb{R}^n)$ which maps S to P defined as in (2). We can then rewrite (12) in this new coordinates system as in:

$$(\bar{A}_j x)^T(\bar{A}_j x) \leq \gamma^2 x^T x,\ \forall\, x \in E_P \setminus L^{-1}(V_S), \forall\, m \in M. \tag{13}$$

Due to the the homogeneity Property 2.1, this implies

$$(\bar{A}_j x)^T(\bar{A}_j x) \leq \gamma^2 x^T x,\ \forall\, x \in \mathrm{S}\backslash\Pi_S(L^{-1}(V_S)), \forall\, m \in M. \tag{14}$$

We now show how to relate $\sigma^{n-1}(V_S)$ to $\sigma^{n-1}(\Pi_S(L^{-1}(V_S)))$. Consider S^{V_S} the sector of B defined by V_S. We denote $C := L^{-1}(S^{V_\mathrm{S}})$ and $V' := \Pi_S(L^{-1}(V_S))$. We have $\Pi_\mathrm{S}(C) = V'$ and $S^{V'} \subset \mathcal{H}_{1/\lambda_{\min}(L^{-1})}(C)$. This leads to:

$$\sigma^{n-1}(V') = \lambda(S^{V'}) \leq \lambda(\mathcal{H}_{1/\lambda_{\min}(L^{-1})}(C)),$$

which means the following holds:

$$\begin{aligned}
\sigma^{n-1}(V') &\leq \frac{1}{\lambda_{\min}(L^{-1})^n}\lambda(L^{-1}(S^{V_\mathrm{S}})) \\
&= \frac{|\det(L^{-1})|}{\lambda_{\min}(L^{-1})^n}\sigma(S^{V_S}), \\
&= \sqrt{\frac{\lambda_{\max}(P)^n}{\det(P)}}\sigma^{n-1}(V_S) \qquad (15)
\end{aligned}$$

where the first equality follows from Remark **??**. Putting together (15), (12), , we get the statement of the theorem where $S' = \Pi_S(L^{-1}(V_S))$. ∎

Lemma 4.3: Let $\epsilon \in (0,1)$. Then, we can compute $\alpha(\epsilon)$ satisfying:

$$\alpha(\epsilon) = \sup\{r : r\mathrm{B} \subset \text{convhull } (\mathrm{S} \setminus X_\epsilon), \tag{16}$$

for all $X_\epsilon \in \{X \subset \mathrm{S} : \sigma^{n-1}(X) \leq \epsilon\}$.

Proof: See Appendix. ∎

Lemma 4.4: Let $\epsilon \in (0,1)$ and $\gamma \in \mathbb{R}_{>0}$. Consider the set of matrices and $A \in \mathcal{M}$ satisfying:

$$(A_j x)^T(A_j x) \leq \gamma x^T x, \quad \forall\, x \in \mathrm{S} \setminus \mathrm{S}', \forall\, j \in, M, \tag{17}$$

where $\mathrm{S}' \subset \mathrm{S}$ and $\sigma^{n-1}(\mathrm{S}') \leq \epsilon$, then we have:

$$\rho(\mathcal{M}) \leq \frac{\gamma}{\alpha(\epsilon)}$$

where $\alpha(\epsilon)$ is defined as in (16).

Proof: Note that, (17) implies that:

$$A_j(\mathrm{S} \setminus \mathrm{S}') \subset \gamma\mathrm{B}.$$

Using Property 2.2 this also implies:

$$A_j\text{convhull } (\mathrm{S} \setminus \mathrm{S}') \subset \text{convhull } (A_j(\mathrm{S} \setminus \mathrm{S}')) \subset \gamma^*\mathrm{B}.$$

By Lemma 4.3 we have:

$$A_j(\alpha(\epsilon)\mathrm{B}) \subset A_j(\text{convhull } (\mathrm{S} \setminus \mathrm{S}')) \subset \mathrm{B}, \quad \forall\, j \in M,$$

where $\alpha(\epsilon)$ is defined as in (26). Therefore, we get:

$$\alpha(\epsilon)A_j(\mathrm{B}) \subset \gamma\mathrm{B}.$$

which implies that $\rho(\mathcal{M}) \leq \frac{\gamma}{\alpha(\epsilon)}$. ∎

Theorem 4.5: Consider an n-dimensional switching system as in (3). For any given $\beta \in (0,1]$, $\eta > 0$ and a uniform random sampling $\omega_N \subset Z$, with $N \geq \frac{n(n+1)}{2} + 1$, and let $\gamma^*(\omega_N)$ be the optimal solution to (6). Then, we can compute $\delta(\beta, \omega_N)$, such that with probability at least β we have:

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

where $\lim_{N\to\infty} \delta(\beta, \omega_N) = 1$.

Proof: Note that, by definition of $\gamma^*(\omega_N)$ we have:

$$(A_j x)^T P(A_j x) \leq (\gamma^*(1+\eta))^2 x^T P x, \quad \forall\, (x,j) \in \omega_N$$

for some $P \succ 0$. Note that the inequality (8) in Theorem 4.1 can be also written as:

$$\mathbb{P}^N\{\mu(V(\omega_N)) \leq \epsilon\} \geq 1 - I(1-\epsilon; N-d, d+1), \tag{18}$$

where $I(\ell; a, b)$ is the regularized in complete beta function. Then, for all $\epsilon \in (0,1)$ satisfying:

$$\epsilon \leq 1 - I^{-1}(1-\beta; N-d, d+1), \tag{19}$$

we have $\mathbb{P}^N\{\mu(V(\omega_N)) \leq \epsilon\} \geq \beta$. Then, by Theorem 4.1 for all ϵ satisfying (19), with probability at least β the following holds:

$$(A_j x)^T P(A_j x) \leq (\gamma^*(1+\eta))^2 x^T P x, \quad \forall (x,j) \in Z \setminus V.$$

By Theorem 4.2, this implies that with probability at least β the following also holds:

$$(\bar{A}_j x)^T(\bar{A}_j x) \leq \gamma^2 x^T x,\ \forall\, x \in \mathrm{S} \setminus \mathrm{S}', \forall\, j \in M,$$

for some S' where $\sigma^{n-1}(S') \leq m\epsilon\kappa(P)$. Then, applying Lemma 4.4, we can compute

$$\delta(\beta, \omega_N) = \alpha(m\kappa(P)(1 - I^{-1}(1-\beta; N-d, d+1)))$$

such that with probability at least β we have:

$$\bar{A}_j\mathrm{B} \subset \frac{\gamma^*(\omega_N)(1+\eta)}{\delta(\beta, \omega_N)}\mathrm{B},\ \forall\, j \in M,$$

By Property 2.4, this means that with probability at least β:

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

which completes the proof of the first part of the theorem. We show that $\lim_{N\to\infty} \delta(\beta, \omega_N) = 1$ in the Appendix D. ∎

V. Experimental Results

VI. Conclusions

Appendix

A. Notation and Background

Before proceeding to the main lemmas we use to prove Lemma 4.3, we first introduce the necessary preliminary definitions and related background.

Let d be a distance on $\mathbb{R}^n$. The distance between a set $X \subset \mathbb{R}^n$ and a point $p \in \mathbb{R}^n$ is $d(X,p) := \inf_{x\in X} d(x,p)$. Note that the map $p \mapsto d(X,p)$ is continuous on $\mathbb{R}^n$.

Definition We define the *spherical cap* on S for a given hyperplane $c^T x = k$ as:

$$\mathcal{C}_{c,k} := \{x \in \mathrm{S} : c^T x > k\}.$$

Remark 1.1: Consider the spherical caps $\mathcal{C}_{c,k_1}$ and $\mathcal{C}_{c,k_2}$ such that $k_1 > k_2$, then we have:

$$\sigma^{n-1}(\mathcal{C}_{c,k_1}) < \sigma^{n-1}(\mathcal{C}_{c,k_2}).$$

Definition A *supporting hyperplane* of a set $X \subset \mathbb{R}^n$ is a hyperplane $\{x : c^T x = k\}$ that has the following two properties:

- $X \subset \{x : c^T x \leq k\}$ or $X \subset \{x : c^T x \geq k\}$.
- $X \cap \{x : c^T x = k\} \neq \emptyset$.

Remark 1.2: [2] Consider a convex set $X \subset \mathbb{R}^n$. For every $x \in \partial X$, there exists a supporting hyperplane containing x. Moreover, if X is smooth, then this supporting hyperplane is unique.

Remark 1.3: The distance between the point $x = 0$ and the hyperplane $c^T x = k$ is $\frac{|k|}{\|c\|}$.

We now define the function $\Delta : \wp(\mathrm{S}) \to [0,1]$ as:

$$\Delta(X) := \sup\{r : r\mathrm{B} \subset \text{convhull } (\mathrm{S} \setminus X)\}. \tag{20}$$

Note that, $\Delta(X)$ can be rewritten as:

$$\Delta(X) = d(\partial \text{convhull } (\mathrm{S} \setminus X), 0). \tag{21}$$

Lemma 1.1: Consider the spherical cap $\mathcal{C}_{c,k}$. We have:

$$\Delta(\mathcal{C}_{c,k}) = \min\left(1, \frac{|k|}{\|c\|}\right).$$

Proof: Note that:

$$\text{convhull } (\mathrm{S} \setminus X) = \left\{x \in \mathrm{B} : c^T x \leq k\right\}.$$

Then the following equalities hold:

$$\begin{aligned}
\Delta(X) &= d(\partial \text{convhull } (\mathrm{S} \setminus X), 0) \\
&= \min(d(\partial \mathrm{B}, 0), d(\partial\{x : c^T x \leq k\}, 0)) \\
&= \min(d(\mathrm{S}, 0), d(\{x : c^T x = k\}, 0)) \\
&= \min\left(1, \frac{|k|}{\|c\|}\right).
\end{aligned}$$

■

Corollary 1.2: Consider the spherical caps $\mathcal{C}_{c,k_1}$ and $\mathcal{C}_{c,k_2}$ such that $k_1 \leq k_2$. Then we have:

$$\Delta(\mathcal{C}_{c,k_1}) \leq \Delta(\mathcal{C}_{c,k_2}).$$

B. Preliminary Results

Lemma 1.3: For any set $X \subset \mathrm{S}$, there exist c and k such that $\mathcal{C}_{c,k}$ satisfies:

$$\mathcal{C}_{c,k} \subset X,$$

and

$$\Delta(\mathcal{C}_{c,k}) = \Delta(X). \tag{22}$$

Proof: Let $\tilde{X} :=$ convhull $(S \setminus X)$. Since d is continuous and the set $\partial\tilde{X}$ is compact, there exists a point $x^* \in \partial\tilde{X}$, such that:

$$\Delta(X) = d(\partial X_S, 0) = \min_{x \in \partial\tilde{X}} d(x, 0) = d(x^*, 0).$$

Next, consider the supporting hyperplane of $\tilde{X}$ at x^*, which we denote by $\left\{x : c^T x = k\right\}$. Note that this supporting hyperplane is a supporting hyperplane of the ball $(\Delta(X)\mathrm{B})$ at x^* since we have:

$$\partial(\Delta(X)\mathrm{B}) \subset \partial\tilde{X} \subset \left\{x : c^T x = k\right\}.$$

By Remark 1.2, this implies that $\left\{x : c^T x = k\right\}$ is in fact the unique supporting hyperplane at x^*. Then we have:

$$\Delta(X) = d(x^*, 0) = d(\{x : c^T x = k\}, 0) = \min\left(1, \frac{|k|}{\|c\|}\right).$$

Now, consider the spherical cap $\mathcal{C}_{c,k}$. Then, by Lemma 1.1 we have $\Delta(\mathcal{C}_{c,k}) = \min\left(1, \frac{|k|}{\|c\|}\right)$. Therefore, $\Delta(X) = \Delta(\mathcal{C}_{c,k})$.

We next show $\mathcal{C}_{c,k} \subset X$. We prove this by contradiction. Assume $x \in \mathcal{C}_{c,k}$ and $x \notin X$. Note that, if $x \notin X$, then $x \in \mathrm{S} \setminus X \subset$ convhull $(\mathrm{S} \setminus X)$. Since $x \in \mathcal{C}_{c,k}$, we have $c^T x > k$. But due to the fact that $x \in$ convhull $(\mathrm{S} \setminus X)$, we also have $c^T x \leq k$, which leads to a contradiction. Therefore, $\mathcal{C}_{c,k} \subset X$. ■

Lemma 1.4: Let $\mathcal{X}_\epsilon = \{X \subset \mathrm{S} : \sigma^{n-1}(X) = \epsilon\}$. Then, for any $\epsilon \in (0,1)$, the function $\Delta(X)$ attains its minimum over X_ϵ for some X which is a spherical cap.

Proof: We prove this via contradiction. Assume that there exists no spherical cap in $\mathcal{X}_\epsilon$ such that $\Delta(X)$ attains its minimum. This means there exists an $X^* \in \mathcal{X}_\epsilon$, where X^* is not a spherical cap and $\arg\min_{X \in \mathcal{X}_\epsilon}(\Delta(X)) = X^*$. By Lemma 1.3, we can construct a spherical cap $\mathcal{C}_{c,k}$ such that $\mathcal{C}_{c,k} \subset X^*$ and $\mathcal{C}_{c,k} = \Delta(X^*)$. Note that, we further have $\mathcal{C}_{c,k} \subsetneq X^*$, since X^* is assumed not to be a spherical cap. This means that, there exists a spherical cap $\sigma^{n-1}(\mathcal{C}_{c,k})$ such that $\sigma^{n-1}(\mathcal{C}_{c,k}) < \epsilon$.

Then, the spherical cap $\mathcal{C}_{c,\tilde{k}}$ with $\sigma^{n-1}(\mathcal{C}_{c,\tilde{k}}) = \epsilon$, satisfies $\tilde{k} < k$ by Remark 1.1. This implies

$$\Delta(\mathcal{C}_{c,\tilde{k}}) < \Delta(\mathcal{C}_{c,k}) = \Delta(X^*)$$

by Corollary 1.2. Therefore, $\Delta(\mathcal{C}_{c,\tilde{k}}) < \Delta(X^*)$. This is a contradiction since we initially assumed that $\Delta(X)$ attains its minimum over $\mathcal{X}_\epsilon$ at X^*. ■

C. Proof of Lemma 4.3

Proof: Let the function $\Delta(X)$ be defined as in (20). Then by Lemma 1.4 we know that:

$$\Delta(X_\epsilon) \geq \Delta(\mathcal{C}_{c,k}), \tag{23}$$

for some spherical cap $\mathcal{C}_{c,k} \subset \mathrm{S}$, where $\sigma^{n-1}(\mathcal{C}_{c,k}) = \epsilon$. It is known (see e.g. [10]) that the area of such $\mathcal{C}_{c,k}$, is given by the equation:

$$\sigma^{n-1}(\mathcal{C}_{c,k}) = \frac{\pi^{d/2}}{\Gamma[\frac{d}{2}]} I\left(1 - \Delta(X_\epsilon)^2; \frac{d-1}{2}, \frac{1}{2}\right), \tag{24}$$

where I is the regularized incomplete beta function. Since, $\sigma^{n-1}(X_\epsilon) \leq \epsilon$, we get the following set of equations:

$$\begin{aligned}\frac{\epsilon\Gamma[\frac{d}{2}]}{\pi^{d/2}} &\leq I\left(1-\Delta(X_\epsilon)^2; \frac{d-1}{2}, \frac{1}{2}\right)\\ 1-\Delta(\mathcal{C}_{c,k})^2 &\leq I^{-1}\left(\frac{\epsilon\Gamma(\frac{d}{2})}{\pi^{d/2}}; \frac{d-1}{2}, \frac{1}{2}\right)\\ \Delta(\mathcal{C}_{c,k})^2 &\geq 1-I^{-1}\left(\frac{\epsilon\Gamma(\frac{d}{2})}{\pi^{d/2}}; \frac{d-1}{2}, \frac{1}{2}\right) \quad (25)\end{aligned}$$

The inequalities (25) and (23) imply the inclusion given in (16), where

$$\alpha(\epsilon) = \sqrt{1-I^{-1}\left(\frac{\epsilon\Gamma(\frac{d}{2})}{\pi^{d/2}}; \frac{d-1}{2}, \frac{1}{2}\right)}. \quad (26)$$

■

D. Proof of $\lim_{N\to\infty}\delta(\beta,\omega_N) = 1$ in Theorem 4.5

We start with the following lemma.

Lemma 1.5: Consider the sequential sampling ω_N. Let $d = \frac{n(n+1)}{2}$ and $P(\omega_N)$ be the optimal solution to the optimization problem $\mathrm{Opt}(\omega_N)$ and let $\lambda_{\max}(P(\omega_N))$ be the optimal objective function value for this $P(\omega_N)$. Then, $\lambda_{\max}(P(\omega_N))$ is uniformly bounded in N.

Proof: We first define the following optimization problem:

$$\begin{aligned}\min_P \quad & \lambda_{\max}(P)\\ \text{s.t.} \quad & (A_jx)^TP(A_jx) \leq (1+\eta)\gamma^2x^TPx,\ \forall(x,j)\in\omega_N,\\ & P \succeq I,\end{aligned} \quad (27)$$

where we denote its optimal solution by $\lambda_{\max}(\gamma,\omega_N)$.

Note that, for all $d\in\mathbb{Z}$ such that $0<d\leq N$ we have $\gamma^*(\omega_d)\leq\gamma^*(\omega_N)$. Also note that,

$$\lambda_{\max}(\gamma^*(\omega_N),\omega_N) \leq \lambda_{\max}(\gamma^*(\omega_d),\omega_N).$$

But note that, there exists a $c>0$ such that $\lambda_{\max}(\gamma^*(\omega_d),\omega_N) < c$ since the problem (27) is strictly feasible for any γ such that $\gamma\leq\gamma^*$. This implies: $\lambda_{\max}(\gamma^*(\omega_d),\omega_N)\leq c$, which completes the proof of this lemma. ■

We now prove that $\lim_{N\to\infty}\epsilon(\beta,N)=0$. Note that, we can upper bound $1-\beta$ as follows:

$$1-\beta = \sum_{j=0}^{d}\binom{N}{j}\epsilon^j(1-\epsilon)^{N-j} \leq (d+1)N^d(1-\epsilon)^{N-d}. \quad (28)$$

We prove $\lim_{N\to\infty}\epsilon(\beta,N)=0$ by contradiction. Assume that $\lim_{N\to\infty}\epsilon(\beta,N)\neq 0$. This means that, there exists some $\delta>0$ such that $\epsilon(\beta,N)>\delta$ infinitely often. Then, consider the subsequence N_k such that $\epsilon(\beta,N_k)>\delta$, $\forall\, k$. Then, by (28) we have:

$$1-\beta \leq (d+1)N_k^d(1-\epsilon)^{N_k-d} \leq (d+1)N_k^d(1-\delta)^{N_k-d}\ \forall\, k\in\mathbb{N}.$$

Note that $\lim_{k\to\infty}(d+1)N_k^d(1-\delta)^{N_k-d} = 0$. Therefore, there exists a k' such that, we have

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

which is a contradiction. Therefore, we must have $\lim_{N\to\infty}\epsilon(\beta,N)=0$.

References

[1] Ruxandra Bobiti and Mircea Lazar. A delta-sampling verification theorem for discrete-time, possibly discontinuous systems. In *Proceedings of the 18th International Conference on Hybrid Systems: Computation and Control*, HSCC '15, pages 140–148, New York, NY, USA, 2015. ACM.

[2] Stephen Boyd and Lieven Vandenberghe. *Convex Optimization*. Cambridge University Press, New York, NY, USA, 2004.

[3] Giuseppe Carlo Calafiore. Random convex programs. *SIAM Journal on Optimization*, 20(6):3427–3464, 2010.

[4] M. C. Campi and S. Garatti. The exact feasibility of randomized solutions of uncertain convex programs. *SIAM Journal on Optimization*, 19(3):1211–1230, 2008.

[5] S.S. Ge and Z. Sun. *Switched linear systems: Control and design*. Communications and control engineering series. Springer, 2005.

[6] R. M. Jungers. The joint spectral radius, theory and applications. In *Lecture Notes in Control and Information Sciences*, volume 385. Springer-Verlag, Berlin, 2009.

[7] Raphaël Jungers. *The joint spectral radius*, volume 385 of *Lecture Notes in Control and Information Sciences*. Springer-Verlag, Berlin, 2009. Theory and applications.

[8] James Kapinski, Jyotirmoy V. Deshmukh, Sriram Sankaranarayanan, and Nikos Arechiga. Simulation-guided lyapunov analysis for hybrid dynamical systems. In *Proceedings of the 17th International Conference on Hybrid Systems: Computation and Control*, HSCC '14, pages 133–142, New York, NY, USA, 2014. ACM.

[9] Fabien Lauer. On the complexity of switching linear regression. *arXiv preprint arXiv:1510.06920*, 2015.

[10] S. Li. Concise formulas for the area and volume of a hyperspherical cap. *Asian Journal of Mathematics & Statistics*, 4:66–70, 2011.

[11] D. Liberzon. *Switching in Systems and Control*. Birkhäuser, Boston, MA, 2003.

[12] Arkadi Nemirovski and Alexander Shapiro. Convex approximations of chance constrained programs. *SIAM Journal on Optimization*, 17(4):969–996, 2006.

[13] Ufuk Topcu, Andrew Packard, and Peter Seiler. Local stability analysis using simulations and sum-of-squares programming. *Automatica*, 44(10):2669 – 2675, 2008.