Data Driven Stability Analysis of Black-box Switched Linear Systems

Marcus Tullius Cicero [a], Julius Caesar [b], Publius Maro Vergilius [c]

[a] *Buckingham Palace, Paestum*

[b] *Senate House, Rome*

[c] *The White House, Baiae*

Abstract

We address the problem of deciding stability of a "black-box" dynamical system (i.e., a system whose model is not known) from a set of observations. The only assumption we make on the black-box system is that it can be described by a switched linear system. We show that, for a given (randomly generated) set of observations, one can give a stability guarantee, 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 both the number of observations and the required level of confidence. Our results rely on geometrical analysis and combining chance-constrained optimization theory with stability analysis techniques for switched systems.

Key words: keywords to pick.

1 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. Rajout: If (1) is linear, its identification and stability analysis have been extensively studied. In this work, we take a first step into more complex systems by considering the class of switched linear systems. Although we restrict ourselves to such systems, we believe that the presented results can be extended to more general classes of dynamical systems. We start with the following question to serve as a stepping stone: For some $l \in \mathbb{N}_{>0}$, given N traces of length l, $(x_0^1, x_1^1, \ldots, x_l^1)$, $(x_0^2, x_1^2, \ldots, x_l^2)$, $\ldots$, $(x_0^N, x_1^N, \ldots, x_l^N)$ belonging to the behavior of the system (1), (i.e., $x_{k+1}^i = f(k, x_k^i)$ for any $k \in \{0, \ldots, l-1\}$ and any $i \in \{1, \ldots, N\}$), what can we say about the stability of the system (1)? For the rest of the paper, we use

⋆ This paper was not presented at any IFAC meeting. Corresponding author M. T. Cicero. Tel. +XXXIX-VI-mmmxxi. Fax +XXXIX-VI-mmmxxv.

Email addresses: `cicero@senate.ir` (Marcus Tullius Cicero), `julius@caesar.ir` (Julius Caesar), `vergilius@culture.ir` (Publius Maro Vergilius).

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 traces of length l (in the particular case of $l = 1$, these traces become couples of points (x_k, y_k) as defined in (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. However, unless f is a linear function, there are two main reasons behind our quest to directly work on system behaviors and bypass the identification phase:

- Even when the function f is known, in general, stability analysis is a very difficult problem [3];
- Identification can potentially introduce approximation errors, and can be algorithmically hard as well. Again, this is the case for switched systems [14].

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

In recent years, increasing number of researchers started addressing various verification and design problems in control of black-box systems [1, 2, 10, 11]. In particular, the initial idea behind this paper was influenced by the recent efforts in [13, 21], and [4] on using simulation traces 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 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 [21] and [13], or via sampling-based techniques as in [4]. This also relates to almost-Lyapunov functions introduced in [16], which presents a relaxed notion of stability proved via Lyapunov functions decreasing everywhere except on a small set. Note that, since we do not have access to the dynamics, these approaches cannot be directly applied to black-box systems. However, these ideas raise the following problem 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 most of the 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 where a candidate Lyapunov function decreases everywhere on its levels sets, except for an arbitrarily small subset, yet, almost all trajectories diverge to infinity. For example, the system

$$x^+ = \begin{bmatrix} 0.14 & 0 \\ 0 & 1.35 \end{bmatrix} x,$$

admits a Lyapunov function candidate on the unit circle except on the two red areas shown in Fig. 1. Moreover, the size of this "violating set" can be made arbitrarily small by changing the magnitude of the unstable eigenvalue. Nevertheless, the only trajectories that do not diverge to infinity are those starting on the stable eigenspace that has zero measure.

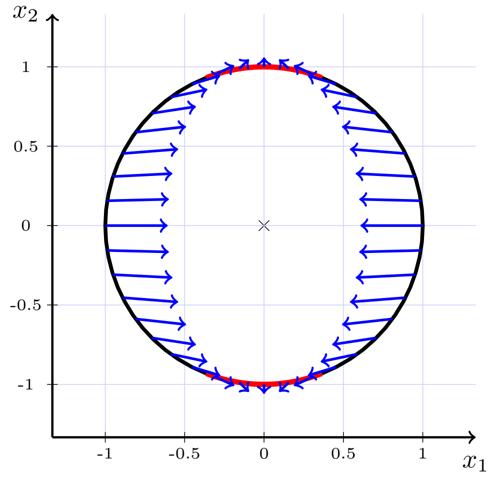


Fig. 1. A simple dynamics and the level set of an "almost Lyapunov function". Even though this function decreases at almost all points in its level set, almost all trajectories diverge to infinity.

In this paper, we take the first steps to infer stability from observations of switched linear systems. In addition to the preceding example, there are other reasons to temper our expectations for proving stability from data. First, identifying an arbitrary switched linear system is NP-hard [12]. Second, the stability of switched linear systems is closely related to a quantity on the matrices modeling the dynamics in each mode, whose computation is itself known to be NP hard: the *joint spectral radius* (JSR). Indeed, deciding stability amounts to deciding whether the JSR is less than 1 [12]. 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 [6, 9, 17]), 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 an asymptotic guarantee on the gap between the upper and the lower bound, for large N.

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 consists in a probabilistic stability guarantee. We illustrate the performance of the presented techniques with some experiments in Section 5, then we conclude in Section 6, while hinting at our related future work.

2 Preliminaries

2.1 Notations

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 $x \in \mathbb{R}^n$. For a distance d on $\mathbb{R}^n$, the distance between a set $X \subset \mathbb{R}^n$ and a point $p \in \mathbb{R}^n$ is given by $d(X, p) := \inf_{x \in X} d(x, p)$. Note that the map $p \mapsto d(X, p)$ is continuous on $\mathbb{R}^n$. Given a set $X \subset \mathbb{R}^n$, ∂X denotes the boundary of set X.

We also 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 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 semidefinite. 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 ratio r of this set. 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 $\Pi_{\mathbb{S}} := x/\|x\|$.

We consider in this work the classical uniform spherical measure on $\mathbb{S}$, denoted by σ^{n-1}, and derived from the Lebesgue measure λ (see the appendix for precise definitions). For $m \in \mathbb{N}_{>0}$, we denote by M the set $M = \{1, 2, \dots, m\}$ and we provide it with the uniform measure μ_M. For any $l \in \mathbb{N}_{>0}$, we denote by M^l the l-Cartesian product of M: $M^l = M \times M \times \cdots \times M$, and similarly we provide it with the uniform measure μ_{M^l}. Finally, we define $Z_l = \mathbb{S} \times M^l$ as the Cartesian product of the unit sphere $\mathbb{S}$ and M^l. On the set Z_l, we define the product measure $\mu_l = \sigma^{n-1} \otimes \mu_{M^l}$. Note that, μ_l is a uniform measure on Z_l and $\mu_l(Z_l) = 1$.

2.2 Stability of Switched Linear Systems

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

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

with $f(k, x_k) = A_{\tau(k)} x_k$ and switching sequence $\tau : \mathbb{N} \to M$.

In this paper, we are interested in the worst-case global stability of this system, that is, we want to guarantee the following property:

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

It is well-known that the joint spectral radius of a set of matrices $\mathcal{M}$ closely relates to the stability of the underlying switched linear systems (2) defined by $\mathcal{M}$. This quantity is an extension to switched linear systems of the classic 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 1 (from [12]) *Given a finite set of matrices $\mathcal{M} \subset \mathbb{R}^{n \times n}$, its* joint spectral radius *(JSR) is given by*

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

Property 2.1 (Corollary 1.1, [12]) *Given a finite set of matrices $\mathcal{M}$, the corresponding switched dynamical system is stable if and only if $\rho(\mathcal{M}) < 1$.*

Property 2.2 (Proposition 1.3, [12]) *Given a finite set of matrices $\mathcal{M}$, and any invertible matrix T,*

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

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

The JSR also relates to a tool classically used in control theory to study stability of systems: Lyapunov functions. We will consider here a family of such functions that is particularly adapted to the case of switched linear systems.

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

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

CQLFs are useful because they can be computed (if they exist) with semidefinite programming (see [5]), and they constitute a stability guarantee for switched systems as we formalize next.

Theorem 3 *[12, Prop. 2.8] Consider a finite set of matrices $\mathcal{M}$. If there exist some $\gamma \geq 0$ and $P \succ 0$ such that*

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

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

It turns out that one can guarantee the accuracy of this Lyapunov technique thanks to the following converse CQLF theorem.

Theorem 4 *[12, Theorem 2.11] For any finite set of matrices such that $\rho(\mathcal{M}) < \frac{1}{\sqrt{n}}$, there exists a CQLF for $\mathcal{M}$, that is, a $P \succ 0$ such that:*

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

Note that, the smaller γ is in 3, the tighter is the upper bound we get on $\rho(\mathcal{M})$. Therefore, we could consider, in particular, the optimal solution γ^* of the following optimization problem:

$$\begin{array}{ll} \min_{\gamma,P} & \gamma \\ \text{s.t.} & (Ax)^T P (Ax) \leq \gamma^2 x^T P x,\ \forall\, A \in \mathcal{M},\ \forall x \in \mathbb{R}^n \\ & P \succ 0. \end{array} \tag{3}$$

Property 2.3 *Let $\xi(x,k,\tau)$ denote the state of the system (2) at time k starting from the initial condition x and with switching sequence τ. The dynamical system (2) is homogeneous: $\xi(\gamma x, k, \tau) = \gamma \xi(x,k,\tau)$.*

Property 2.3 enables us to restrict the set of constraints x to the unit sphere $\mathbb{S}$, instead of considering it as being all $\mathbb{R}^n$. Indeed, the homogeneity implies that it is sufficient to show the decrease of a CQLF on an arbitrary set enclosing the origin, e.g., $\mathbb{S}$. Hence, we consider from now the following optimization problem, with its optimal solution that will also be γ^*:

$$\begin{array}{ll} \min_{\gamma,P} & \gamma \\ \text{s.t.} & (Ax)^T P (Ax) \leq \gamma^2 x^T P x,\ \forall\, A \in \mathcal{M},\ \forall x \in \mathbb{S} \\ & P \succ 0. \end{array} \tag{4}$$

The two theorems above provide us with a *converse Lyapunov result*: if there exists a CQLF, then our system is stable. If, on the contrary, there is no such stability guarantee, one may conclude a lower bound on the JSR. By combining these two results, one obtains an approximation algorithm for the JSR: the upper bound γ^* obtained above is within an error factor of $\frac{1}{\sqrt{n}}$ of the true value. It turns out that one can still refine this technique, in order to improve the error factor, 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 5 *For any finite set of matrices such that $\rho(\mathcal{M}) < \frac{1}{\sqrt[2l]{n}}$, there exists a CQLF for*

$$\mathcal{M}^l := \{A_{i_1}, \ldots, A_{i_l} : A_i \in \mathcal{M}\},$$

that is, a $P \succ 0$ such that:

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

PROOF. It is easy to see from the definition of the JSR that

$$\rho(\mathcal{M}^l) = \rho(\mathcal{M})^l.$$

Thus, applying Theorem 4 to $\mathcal{M}^l$, one directly obtains the corollary.

We can then consider from now, for any $l \in \mathbb{N}_{>0}$, the following optimization problem, whose solution γ^* will be an upper bound on $\rho(\mathcal{M})$:

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

2.3 Problem Formulation

We now formally present the problem we will be considering from now. We recall that our observations are traces of the form $(x_k, x_{k+1}, \ldots, x_{k+l})$ for some arbitrary $l \in \mathbb{N}_{>0}$, and that we do not have access to the mode applied to the system at each time step.

To generate these traces, we assume that we can randomly pick a finite number of initial conditions $x_0^i \in \mathbb{R}^n$, and that a random sequence of l modes is applied to each of these points. Hence, the probability event corresponding to a given observed trace $(x_k, x_{k+1}, \ldots, x_{k+l})$ is another $(l+1)$-tuple $(x_k, j_1, \ldots, j_l)$. More precisely, we assume that we can uniformly sample such $(l+1)$-tuples in $Z_l = \mathbb{S} \times M^l$, giving us a sample denoted by

$$\begin{aligned} \omega_N := &\{(x_0^1, j_{1,1}, \ldots, j_{1,l}), (x_0^2, j_{2,1}, \ldots, j_{2,l}), \ldots, \\ &(x_0^N, j_{N,1}, \ldots, j_{N,l})\} \subset Z_l. \end{aligned}$$

By uniformly sampling, we mean that the points in ω_N are drawn according to the measure μ_l, i.e., the points

x_0^i are drawn from $\mathbb{S}$ according to the classical spherical measure σ^{n-1}, and the modes are drawn from M according to the classical uniform measure μ_M at each time step. The space of sampling for the initial conditions is restricted to $\mathbb{S}$ since, as we recall, by Property 2.3, the system is homogeneous.

Remark 6 *Let us motivate the choice of considering a uniform sampling for the modes. Since we assume that we only have random observations of the state of the system, we do not know the process that generates this state: in particular, we ignore the process that picks the modes at each time step. We model this process with a random distribution. Here, we make the assumption that with nonzero probability, each mode is active. The problem would indeed not make a lot of sense otherwise, since in a such case, with probability* 1*, the system would be unidentifiable and would prevent to ever observe some of its possible behaviors. By default, we take this distribution uniform since we cannot say that some modes are privileged a priori. But we can still take any other distribution satisfying our assumption; if we have a lower bound on the probability of each mode that is strictly positive, our guarantees naturally extend to them.*

From a sample ω_N, we obtain the set of corresponding available observations

$$W_{\omega_N} := \{(x_0^1, x_1^1, \dots, x_l^1), (x_0^2, x_1^2, \dots, x_l^2), \dots, (x_0^N, x_1^N, \dots, x_l^N)\}, \tag{6}$$

which satisfy

$$x_l^i = A_{j_{i,1}} \dots A_{j_{i,l}} x_0^i, \ \forall \, (x_0^i, x_1^i, \dots, x_l^i) \in W_{\omega_N}.$$

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 a finite, uniformly sampled, set of data. More precisely, we answer the following problem:

Problem 7 *Consider a finite set of matrices $\mathcal{M}$, describing a switched system* (2)*, and suppose that one has a set of N observations*

$$\omega_N = \{(x_0^1, j_{1,1}, \dots, j_{1,l}), (x_0^2, j_{2,1}, \dots, j_{2,l}), \dots, (x_0^N, j_{N,1}, \dots, j_{N,l})\},$$

sampled according to the uniform measure μ_l on Z_l.

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

$$\mu_l \left(\{\omega_N : \ \rho(\mathcal{M}) \le \overline{\rho(\omega_N)}\} \right) \ge \beta.$$

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

Remark 8 *We will see in Section 3 that a such level of confidence β is not even required in the case of the lower bound. Indeed, we derive in Theorem 9 a deterministic lower bound for a given (sufficiently high) number of observations.*

The idea from now will be to leverage the fact that conditions for the existence of a CQLF for (2) can be obtained by considering a finite number of traces in $\mathbb{R}^n$ of the form $(x_k, x_{k+1}, \dots, x_l)$. It will lead us to the following algorithm, that is the main result of our paper and that answers Problem 7:

Algorithm 1 ***Input:*** *observations W_{ω_N} corresponding to a uniform random sample $\omega_N \subset Z$ of size $N \ge \frac{n(n+1)}{2} + 1$;*

Input: *β desired level of certainty;*

Compute: *$\gamma^*(\omega_N)$ optimal solution of Opt(ω_N), that we take as candidate for the upper bound;*

Compute: *$\varepsilon(\beta, \omega_N)$ the size of the of points where we might make the wrong inference on the upper bound;*

Compute: *$\delta(\varepsilon)$;*

Output: *$\frac{\gamma^*(\omega_N)}{\sqrt[2l]{n}} \le \rho \le \frac{\gamma^*(\omega_N)}{\sqrt[l]{\delta(\varepsilon)}}$, (upper bound valid with probability at least β and $\delta(\beta, \omega_N) \xrightarrow[N\to\infty]{} 1$).*

3 A Deterministic Lower Bound

In Section 2.2, we gave an optimization problem, (5), that provides a stability guarantee. Nevertheless, solving this problem as stated solely from observation of traces (that gives a finite number of constraints) is not possible since (5) involves infinitely many constraints. We consider then the following optimization problem:

$$\begin{aligned} \min_P \quad & \gamma \\ \text{s.t.} \quad & (A_{j_l} A_{j_{l-1}} \dots A_{j_1} x)^T P (A_{j_l} A_{j_{l-1}} \dots A_{j_1} x) \le \gamma^{2l} x^T P x, \quad \forall (\\ & P \succ 0, \ \gamma \ge 0. \end{aligned} \tag{7}$$

Note that, (7) can be efficiently solved by semidefinite programming and bisection on the variable γ (see [5]). Let us denote from now by $\gamma^*(\omega_N)$ the optimal solution of this problem, which we will use to compute a deterministic lower bound and a probabilistic upper bound on the JSR. In this section, we provide a theorem for a deterministic lower bound based on the observations W_{ω_N}, whose accuracy depends on the "horizon" l.

Theorem 9 *For an arbitrary $l \in \mathbb{N}_{>0}$, and a given uniform sample*

$$\omega_N := \{(x_1, j_{1,1}, \dots, j_{1,l}), (x_2, j_{2,1}, \dots, j_{2,l}), \dots, (x_N, j_{N,1}, \dots, j_{N,l})\} \subset Z_l,$$

by considering $\gamma^*(\omega_N)$ the optimal solution of the optimization problem (7):

$$\begin{array}{ll} min_P & \gamma \\ s.t. & (A_{j_l} A_{j_{l-1}} \dots A_{j_1} x)^T P (A_{j_l} A_{j_{l-1}} \dots A_{j_1} x) \leq \gamma^{2l} x^T P x, \quad \forall (x, j_1, \dots, j_l) \in \omega_N \\ & P \succ 0, \ \gamma \geq 0, \end{array}$$

we have:

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

PROOF. Let $\nu > 0$. By definition of γ^*, there exists no matrix $P \in \mathcal{S}^n_{++}$ such that:

$$(Ax)^T P (Ax) \leq (\gamma^*(\omega_N) - \nu)^{2l} x^T P x, \quad \forall x \in \mathbb{R}^n, \forall A \in \mathcal{M}^l.$$

By Property 2.2, this means that there exists no CQLF for the scaled set of matrices $\frac{\mathcal{M}^l}{(\gamma^*(\omega_N) - \nu)^l}$. Since the above inequality is true for every $\nu \geq 0$, using Theorem 4, and the fact that $\rho(\mathcal{M}^l) = \rho(\mathcal{M})^l$, we conclude:

$$\frac{\rho(\mathcal{M})}{\gamma^*(\omega_N)} \geq \frac{1}{\sqrt[2l]{n}}.$$

4 A Probabilistic Stability 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 a l-step CQLF which is valid with probability at least β. Note that, the existence of a l-step CQLF implies $\rho \leq 1$ due to Theorem 3.

Let us analyze the relationship between the solutions of the theoretical optimization problem (5) and the practical version (7), with finitely many constraints. Even though in practice, one would solve the optimization problem (7) as suggested in the previous section, for the sake of rigor and clarity of our proofs, we introduce a slighly different optimization problem. In this new optimization problem, an objective function is considered, and a "regularization parameter", $\eta > 0$, is added. As the reader will see, we will derive results valid for arbitrarily small values of η, and so this will not hamper the practical accuracy of our technique, while allowing us to derive a theoretical asymptotic guarantee (i.e. for large number of observations).

$$\begin{array}{ll} \min_P & \lambda_{\max}(P) \\ \text{s.t.} & (A_{j_l} A_{j_{l-1}} \dots A_{j_1} x)^T P (A_{j_l} A_{j_{l-1}} \dots A_{j_1} x) \leq ((1+\eta)\gamma^*(\omega_N \\ & \qquad \forall (x, j_1, \dots, j_l) \\ & P \succeq I, \end{array} \tag{8}$$

where $Z_l := \mathbb{S} \times M^l$, $\eta > 0$, and $\gamma^*(\omega_N)$ is the optimal solution to the optimization problem (7). Recall that ω_N is a given uniform random sample on the set Z_l and of size N.

For the rest of the discussion, we refer to the optimization problem (8) by $\text{Opt}(\omega_N)$. We denote its optimal solution by $P(\omega_N)$. 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 (8). Firstly, due to Property 2.3, we 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 (8), which will be helpful in the following discussion.

The curious question whether the optimal solution of the sampled problem $\text{Opt}(\omega_N)$ is a feasible solution to (4) has been widely studied in the literature [6]. It turns out that under certain technical assumptions, one can bound the proportion of the constraints of the original problem (4) that are violated by the optimal solution of (8), with some probability which is a function of the sample size N.

In the following theorem, we adapt a classical result from random convex optimization literature to our problem.

Theorem 10 (adapted from Theorem 3.3 [1], [6]) *Let d be the dimension of $Opt(\omega_N)$ and $N \geq d+1$. Consider the optimization problem $Opt(\omega_N)$ given in (8), where ω_N is a uniform random sample drawn from the set Z_l. Then, for all $\varepsilon \in (0, 1]$ the following holds:*

$$\mu_l^N \{\omega_N \in Z_l^N : \mu_l(V(\omega_N)) \leq \varepsilon\} \geq 1 - \sum_{j=0}^{d} \binom{N}{j} \varepsilon^j (1-\varepsilon)^{N-j}, \tag{9}$$

where μ_l^N denotes the product probability measure on Z_l^N, and $V(\omega_N)$ is defined by

$$V(\omega_N) := \{z = (x, j_1, \dots, j_l) \in Z_l : (A_{j_l} A_{j_{l-1}} \dots A_{j_1} x)^T P(\omega_N) (A_{j_l} A_{j_{l-1}} \dots A_{j_1} x) > (\gamma^*_{\omega_N})^{2l} x^T P(\omega_N) x\},$$

i.e., it is the set of constraints that are violated by the optimal solution of $Opt(\omega_N)$. From now we will denote the left hand-side of (9) by β.

Corollary 11 *Consider a set of matrices $\mathcal{M}$, γ^* optimal solution of* (7) *and matrix $P \succ 0$ optimal solution of $Opt(\omega_N)$. Then P will satisfy:*

$$(A_{j_l}A_{j_{l-1}}\dots A_{j_1}x)^T P(A_{j_l}A_{j_{l-1}}\dots A_{j_1}x) \le (\gamma^*)^{2l}x^T Px, \\ \forall\, x \in \mathbb{S}\setminus\mathbb{S}', \forall\,(j_1,\dots,j_l) \in M^l, \tag{10}$$

for some $\mathbb{S}' \subset \mathbb{S}$ such that $\sigma^{n-1}(\mathbb{S}') \le \varepsilon m^l$.

The proof of Corollary 11 is based on straightforward arguments on measures, and is given in Appendix B. This result allows us to only consider set of violating points on the sphere from now. Note that, this result is conservative: the case where we have the equality $\sigma(\mathbb{S}') = \varepsilon m^l$ corresponds to the case where we have only observed one mode, and have then minimal knowledge on the system for a given ε.

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 εm^l. This means that, the ellipsoid computed by $\mathrm{Opt}(\omega_N)$ is "almost invariant" except on a set of measure bounded by εm^l. This can be representend in the case $n = 2$ by the following plot, where the points of the ellipse in red are points that might violate the contractivity constraint (the set of red points has measure at most εm^l).

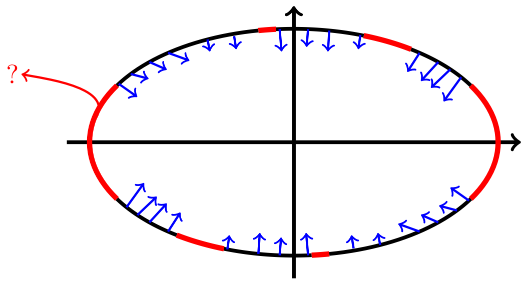


Fig. 2. Representation of the "partial invariance property" obtained by application of the results in 10. A priori, we know nothing about the images of the red points. Our goal is to convert this partial invariance property into a global stability property.

[1] Theorem 3.3 in [6] 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 [7], while the second assumption can be lifted, as explained in PART 2b in [8], thanks to the introduction of a "constraint heating".

Thus, we are left with the following question: "What can we conclude on the JSR if we can assume that the Lyapunov property is satisfied by all points, except a set of measure εm^l?"

4.2 The Spherical Case

We consider first the special and simpler case when the ellipse defined by P is in fact a sphere, and more specifically the unit sphere $\mathbb{S}$. The set of points that can violate the constraint still has measure bounded by εm^l, and is now denoted by $X_{\varepsilon m^l}$. We will see that for this precise case our question becomes a geometric problem, which we describe and solve in this section. Our reasoning uses the following fundamental property of switched linear systems.

Property 4.1 *The dynamics given in* (2) *is convexity-preserving, meaning that for any set of points $X \subset \mathbb{R}^n$ we have:*

$$f(convhull\,(X)) \subset convhull\,(f(X)).$$

This leads us to look for the largest sphere in convhull $(S\setminus X_{\varepsilon m^l})$. By homogeneity of the system, this sphere will be centered at the origin, and we denote by δ its radius. We consider this sphere $\mathbb{S}_\delta$ since, by Property 4.1, we know that the l-traces by system (2) initialized in $\mathbb{S}_\delta$ (hence initialized in convhull $(S \setminus X_{\varepsilon m^l})$), will lie in convhull $(f(S\setminus X_{\varepsilon m^l})) \subset \gamma^l\mathbb{B}$. Then, the rate of growth of $\mathbb{S}_\delta$ after l steps will be upper bounded by $\frac{(\gamma^*)^l}{\delta}$.

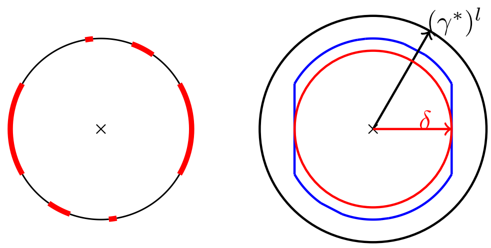


Fig. 3. On the left, our problem in the case of a sphere. On the right, we represent in blue the boundary of the convex hull of the points satisfying the $(\gamma^*)^l$-contractivity constraint on l-traces; in red we represent the largest sphere $\mathbb{S}_\delta$ included in that convex hull ; in black we represent the boundary of the set where elements of all l-traces initialized in that convex hull can lie.

We investigate then the case for which the largest ball included convhull $(S \setminus X_{\varepsilon m^l})$ has the smallest radius δ. This will maximize our upper bound on the growth of rate of the sphere $\mathbb{S}_\delta$ considered, and thus will give us an upper bound on the JSR equal to $\frac{\gamma^*}{\sqrt[l]{\delta}}$. Minimizing δ will then give us the worst case set for a given measure εm^l of points authorized to violate our constraint. This worst case will happen when the set of violating points on $\mathbb{S}$ is a spherical cap (see Proposition 13 below). Let us define this notion.

Definition 12 *We define the* spherical cap *on* $\mathbb{S}$ *for a given hyperplane* $c^T x = k$ *as:*

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

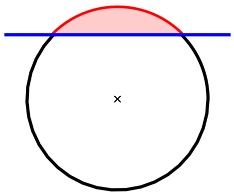


Fig. 4. A spherical cap in the case $n = 2$. The hyperplane $c^T x = k$ defining the spherical cap is in blue, the spherical cap is in red.

We now define the function

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

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

$$\Delta(X) = d(\partial\text{convhull } (\mathbb{S} \setminus X), 0). \tag{12}$$

The following proposition tells us that Δ is minized when X is a spherical cap, i.e., the radius δ of our largest sphere $\mathbb{S}_\delta$ will be reached when $X_{\varepsilon m^l}$ is a spherical cap.

Proposition 13 *Let* $\mathcal{X}_{\varepsilon m^l} = \{X \subset \mathbb{S} : \sigma^{n-1}(X) \leq \varepsilon m^l\}$. *Then, for any* $\varepsilon \in (0, 1/m^l]$, *the function* $\Delta(X)$ *attains its minimum over* $\mathcal{X}_{\varepsilon m^l}$ *for some* X *which is a spherical cap.*

We give a proof of Proposition 13 in Appendix C.

By Property 2.3 (homogeneity of the system), we have $x \in V_{\mathbb{S}} \iff -x \in V_{\mathbb{S}}$, which implies that $X_{\varepsilon m^l}$ will be in fact union of two spherical caps, each of measure $\frac{\varepsilon m^l}{2}$. We denote from now $\varepsilon' := \frac{\varepsilon m^l}{2}$. We illustrate the problem when $X_{\varepsilon m^l}$ is of this form.

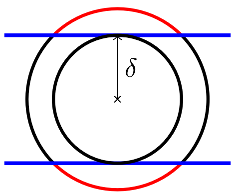


Fig. 5. Illustration of the situation giving the minimal δ, for $n = 2$. The set $V(\omega_N)$ is represented in red, and is the union of two spherical caps, each of measure ε'.

For a given spherical cap of area measure ε', we can compute a closed form expression of the radius of the corresponding largest ball. We denote by α the function that associates to each $\varepsilon' \in (0, 1/2]$ the value of this radius. We have an expression for α given by the following lemma (see the Appendix D for details of its proof).

Lemma 14 *Let* $\varepsilon \in (0, \frac{1}{2}]$ *and* $\alpha : [0, 1] \to \mathbb{R}_{\geq 0}$ *be defined by:*

$$\alpha(\varepsilon) := \inf_{X \in \mathcal{X}_\varepsilon} \sup\{r : r\mathbb{B} \subset \mathit{convhull}\ (\mathbb{S} \setminus X)\}, \tag{13}$$

where $\mathcal{X}_\varepsilon = \{X \subset \mathbb{S} : \sigma^{n-1}(X) \leq \varepsilon\}$. *Then,* $\alpha(\varepsilon)$ *is given by the formula:*

$$\alpha(\varepsilon) = \sqrt{1 - I^{-1}\left(2\varepsilon; \frac{n-1}{2}, \frac{1}{2}\right)}, \tag{14}$$

where I *is the regularized incomplete beta function.*

By exploiting Property 4.1 and Lemma 14 above, we can now show in Lemma 15 how one can compute an upper bound on the JSR when the “almost invariant" set is the unit sphere $\mathbb{S}$.

Proposition 15 *Let* $\varepsilon \in (0, 1/m^l)$ *and* $\gamma^* \in \mathbb{R}_{>0}$. *Consider the set of matrices* $\mathcal{M}$ *and* $A \in \mathcal{M}^l$ *satisfying:*

$$\begin{aligned} &(A_{j_l} A_{j_{l-1}} \dots A_{j_1} x)^T (A_{j_l} A_{j_{l-1}} \dots A_{j_1} x) \leq (\gamma^*)^{2l} x^T x, \\ &\forall\, x \in \mathbb{S} \setminus \mathbb{S}', \forall\, (j_1, \dots, j_l) \in M^l, \end{aligned} \tag{15}$$

where $\mathbb{S}' \subset \mathbb{S}$ *and* $\sigma^{n-1}(\mathbb{S}') \leq \varepsilon m^l$, *then we have:*

$$\rho(\mathcal{M}) \leq \frac{\gamma^*}{\sqrt[l]{\alpha(\frac{\varepsilon m^l}{2})}}$$

where $\alpha(\frac{\varepsilon m^l}{2})$ *is given in* (14).

PROOF. Note that, (15) implies that: $(A_{j_l} A_{j_{l-1}} \dots A_{j_1})(\mathbb{S} \setminus \mathbb{S}') \subset \gamma^l \mathbb{B}$. Using Property 4.1 this also implies:

$$A_{j_l} A_{j_{l-1}} \dots A_{j_1} (\text{convhull } (\mathbb{S} \setminus \mathbb{S}')) \subset \text{convhull } \left(A_{j_l} A_{j_{l-1}} \dots A_{j_1} (\mathbb{S} \setminus\right.$$

Then, by Lemma 14, we have:

$$A_{j_l} A_{j_{l-1}} \dots A_{j_1} (\alpha(\varepsilon)\mathbb{B}) \subset A_{j_l} A_{j_{l-1}} \dots A_{j_1} (\text{convhull } (\mathbb{S} \setminus \mathbb{S}')) \subset \gamma^l \mathbb{B}$$

by definition of $\alpha(\frac{\varepsilon m^l}{2})$ given in (13). Therefore, we get:

$$\alpha\left(\frac{\varepsilon m^l}{2}\right) (A_{j_l} A_{j_{l-1}} \dots A_{j_1} (\mathbb{B})) \subset \gamma^l \mathbb{B},$$

which implies that $\rho(\mathcal{M}^l) \leq \frac{\gamma^l}{\alpha(\frac{\varepsilon m^l}{2})}$ and hence $\rho(\mathcal{M}) \leq \frac{\gamma}{\sqrt[l]{\alpha(\frac{\varepsilon m^l}{2})}}$.

We have then solved the problem when $P(\omega_N) = \mathbb{S}$.

Remark 16 *When $\varepsilon \geq \frac{1}{m^l}$, we have $\delta = 1$ and the upper we can give for the JSR is only $+\infty$.*

4.3 General Case

We now consider the general problem for any $P(\omega_N)$, and transform it into the spherical case presented in the previous section. This enables us to give an answer to this general problem in Theorem 17 below. First, we apply a change of coordinates bringing E_P to $\mathbb{S}$. Since $P \in \mathcal{S}^n_{++}$, it can be written in its Cholesky form

$$P = L^T L, \tag{16}$$

where L is an upper triangular matrix. Note that, L maps the elements of $\mathbb{S}$ to E_P.

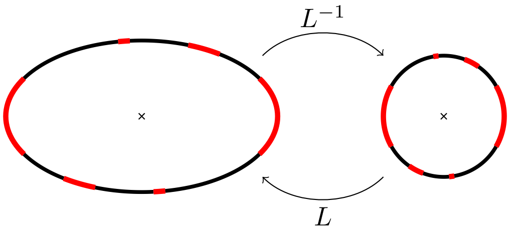


Fig. 6. Change of coordinates to bring our problem back to the case of the unit sphere.

We now know how to compute an upper bound on the JSR when the “almost invariant" ellipsoid is $\mathbb{S}$. Thanks to Property 2.2, if this is not the case, we can simply perform a change of coordinates mapping this ellipsoid to $\mathbb{S}$ and compute the JSR in the new coordinates system instead. To do this, in the next theorem, we bound the measure of violating constraints on $\mathbb{S}$ after the change of coordinates, in terms of the measure of the violated constraints on $\mathbb{S} \times M^l$ in the original coordinates.

Theorem 17 *Let $\gamma^* \in \mathbb{R}_{>0}$. Consider a set of matrices $\mathcal{M}$, and a matrix $P \succ 0$ satisfying* (10)*:*

$$(A_{j_l} A_{j_{l-1}} \dots A_{j_1} x)^T P (A_{j_l} A_{j_{l-1}} \dots A_{j_1} x) \leq (\gamma^*)^{2l} x^T P x,$$
$$\forall x \in \mathbb{S} \setminus \mathbb{S}', \forall\, (j_1, \dots, j_l) \in M^l, \tag{17}$$

for some $\mathbb{S}' \subset \mathbb{S}$ where $\sigma^{n-1}(\mathbb{S}') \leq \varepsilon m^l$. Then, we have

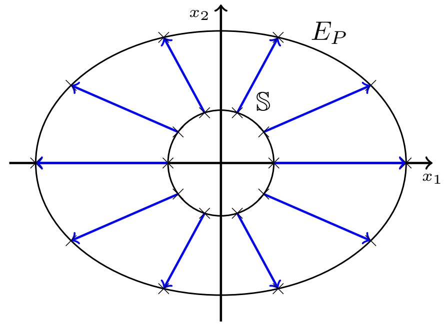
$$\rho(\mathcal{M}) \leq \frac{\gamma^*}{\sqrt[l]{\alpha\left(\frac{\varepsilon m^l \kappa(P)}{2}\right)}}$$

where $\alpha(\frac{\varepsilon m^l \kappa(P)}{2})$ is given in (14)*, and*

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

PROOF.

The assumption (17) is a contraction assumption on points of the ellipsoid E_P. So far, we have represented points of E_P by their corresponding points on $\mathbb{S}$.



We have $y \in E_P \leftrightarrow x \in \mathbb{S}$ with $y = L(x)$, the euclidean norm $y^T y$ becoming $x^T P x$. Hence, (17) can be rewritten as:

$$(A_{j_l} A_{j_{l-1}} \dots A_{j_1} x)^T (A_{j_l} A_{j_{l-1}} \dots A_{j_1} x) \leq (\gamma^*)^{2l} x^T x,$$
$$\forall x \in E_P \setminus L(\mathbb{S}'), \forall\, (j_1, \dots, j_l) \in M^l.$$

Since we have seen in the previous section a technique to solve the spherical case, we perform now the change of coordinates defined as in (16) by $L \in \mathcal{L}(\mathbb{R}^n)$ which maps the ellipsoid E_P to the sphere $\mathbb{S}$. By defining $\bar{A}_{j_i} = L^{-1} A_{j_i} L$, assumption (4.3) becomes:

$$(\bar{A}_{j_l} \bar{A}_{j_{l-1}} \dots \bar{A}_{j_1} x)^T P (\bar{A}_{j_l} \bar{A}_{j_{l-1}} \dots \bar{A}_{j_1} x) \leq (\gamma^*)^{2l} x^T P x,$$
$$\forall x \in L^{-1}(E_P \setminus L(\mathbb{S}')), \forall\, (j_1, \dots, j_l) \in M^l.$$

with

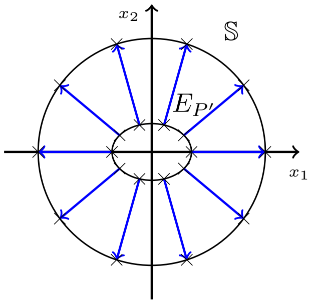
$$L^{-1}(E_P \setminus L(\mathbb{S}')) = \mathbb{S} \setminus \mathbb{S}'.$$

In (4.3), we have brought our original contractivity constraints from points of E_P to points of $\mathbb{S}$. Now, we recall that the points of the ellipsoid E_P in the initial statement were represented through points of $\mathbb{S}$. This sphere becomes after change of coordinates L an ellipsoid $E_{P'}$. Hence, the points of the newly considered level set, $\mathbb{S}$, are now represented through points of $E_{P'}$, and (4.3) can be rewritten as:

$$(\bar{A}_{j_l} \bar{A}_{j_{l-1}} \dots \bar{A}_{j_1} x)^T (\bar{A}_{j_l} \bar{A}_{j_{l-1}} \dots \bar{A}_{j_1} x) \leq (\gamma^*)^{2l} x^T x,$$
$$\forall x \in L^{-1}(\mathbb{S} \setminus \mathbb{S}'),\ \forall\, (j_1, \dots, j_l) \in M^l,$$

with

$$L^{-1}(\mathbb{S} \setminus \mathbb{S}') = E_{P'} \setminus L^{-1}(\mathbb{S}').$$



To reason on the (spherical) measure of violating set in the new coordinates, we project $E_{P'}$ on the unit sphere; $L^{-1}(\mathbb{S}')$ becomes $\Pi_{\mathbb{S}}(L^{-1}(\mathbb{S}'))$.

We now show how to relate $\sigma^{n-1}(\Pi_{\mathbb{S}}(L^{-1}(\mathbb{S}')))$ to $\sigma^{n-1}(\mathbb{S}')$, measure of the violating set in the initial coordinates. Consider $\mathbb{S}^{\mathbb{S}'}$, the sector of $\mathbb{B}$ defined by $\mathbb{S}'$. We denote $C := L^{-1}(\mathbb{S}')$ and $C' := \Pi_{\mathbb{S}}(L^{-1}(\mathbb{S}'))$. We have $\Pi_{\mathbb{S}}(C) = C'$ and $\mathbb{S}^{C'} \subset \mathcal{H}_{1/\lambda_{\min}(L^{-1})}(C)$, where $\mathcal{H}_{1/\lambda_{\min}(L^{-1})}$ is the homothety of ratio $1/\lambda_{\min}(L^{-1})$. This leads to:

$$\sigma^{n-1}(C') = \lambda\left(\mathbb{S}^{C'}\right) \leq \lambda\left(\mathcal{H}_{1/\lambda_{\min}(L^{-1})}(E^{C}_{P'})\right).$$

Then, the following holds:

$$\begin{aligned}
\sigma^{n-1}(C') &\leq \frac{1}{\lambda_{\min}(L^{-1})^n}\lambda(E^{C}_{P'}) \\
&\leq \frac{1}{\lambda_{\min}(L^{-1})^n}\lambda\left(L^{-1}(\mathbb{S}^{\mathbb{S}'})\right) \\
&= \frac{|\det(L^{-1})|}{\lambda_{\min}(L^{-1})^n}\lambda\left(\mathbb{S}^{\mathbb{S}'}\right), && (18) \\
&= \sqrt{\frac{\lambda_{\max}(P)^n}{\det(P)}}\sigma^{n-1}(\mathbb{S}') && (19)
\end{aligned}$$

where (18) 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. [20]).

Finally, thanks to the Property 2.2, we even have:

$$(A_{j_l}A_{j_{l-1}}\dots A_{j_1}x)^T P(A_{j_l}A_{j_{l-1}}\dots A_{j_1}x) \leq (\gamma^*)^{2l}x^TPx, \\ \forall\, x \in \mathbb{S}\Pi_{\mathbb{S}}(L(\mathbb{S}')), \forall\,(j_1,\dots,j_l) \in M^l. \quad (20)$$

with

$$\sigma^{n-1}\left(\Pi_{\mathbb{S}}(L(\mathbb{S}'))\right) \leq \sqrt{\frac{\lambda_{\max}(P)^n}{\det(P)}}\sigma^{n-1}(\mathbb{S}'),$$

which gives us our theorem.

4.4 Main Theorem

We are now ready to prove our main theorem by putting together all the above pieces. For a given level of confidence β, we prove that the upper bound $\gamma^*(\omega_N)$, which is valid solely on finitely many observations, is in fact a true upper bound, at the price of increasing it by the factor $\frac{1}{\sqrt[l]{\delta(\beta,\omega_N)}}$. Moreover, as expected, this factor gets smaller as we increase N and decrease β.

Theorem 18 *Consider an n-dimensional switched linear system as in (2) and a uniform random sampling $\omega_N \subset Z_l$, where $N \geq \frac{n(n+1)}{2} + 1$. Let $\gamma^*(\omega_N)$ be the optimal solution to (8). Then, for any given $\beta \in (0,1)$ and $\eta > 0$, we can compute $\delta(\beta,\omega_N)$, such that with probability at least β we have:*

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

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

PROOF.

By definition of $\gamma^*(\omega_N)$ we have:

$$(A_{j_l}A_{j_{l-1}}\dots A_{j_1}x)^T P(A_{j_l}A_{j_{l-1}}\dots A_{j_1}x) \leq (\gamma^*(1+\eta))^{2l}x^TPx, \\ \forall\,(x,j_1,\dots,j_l) \in \omega_N$$

for some $P \succ 0$.

Then, by rewriting Theorem 10, for

$$\beta = \mu_l^N\left(\{\omega_N \in Z_l^N : \mu_l(V(\omega_N)) \leq \varepsilon\}\right) \geq 1 - I(1-\varepsilon; N-d, d+1), \quad (21)$$

where $I(\ell;a,b)$ is the regularized incomplete Beta function, with probability at least β the following holds:

$$(A_{j_l}A_{j_{l-1}}\dots A_{j_1}x)^T P(A_{j_l}A_{j_{l-1}}\dots A_{j_1}x) \leq ((\gamma^*(1+\eta))^{2l}x^TPx, \\ \forall\,(x,j_1,\dots,j_l) \in Z_l \setminus V.$$

with $\mu_l(V) \leq \varepsilon$, and $\varepsilon(\beta,N) = 1 - I^{-1}(1-\beta; N-d, d+1)$. Thanks to Corollary 11, we even have

$$(A_{j_l}A_{j_{l-1}}\dots A_{j_1}x)^T P(A_{j_l}A_{j_{l-1}}\dots A_{j_1}x) \leq ((\gamma^*(1+\eta))^{2l}x^TPx, \\ \forall\, x \in \mathbb{S} \setminus \mathbb{S}', \forall\,(j_1,\dots,j_l) \in M^l$$

with $\mathbb{S}' = \pi_{\mathbb{S}}(V)$ and $\sigma^{n-1}(\mathbb{S}') \leq \varepsilon m^l$.

By Theorem 17, this implies that with probability at least β the following also holds:

$$(A_{j_l}A_{j_{l-1}}\dots A_{j_1}x)^T(A_{j_l}A_{j_{l-1}}\dots A_{j_1}x) \le (\gamma^*(1+\eta))^{2l}x^Tx,\\ \forall\, x \in \mathbb{S}\setminus\mathbb{S}', \forall\,(j_1,\dots,j_l)\in M^l, \quad (22)$$

for some $\mathbb{S}'$ where $\sigma^{n-1}(\mathbb{S}') \le \varepsilon m^l\kappa(P)$. Then, applying Lemma 15, we can compute

$$\delta(\beta,\omega_N) = \alpha(\varepsilon'(\beta,N)),$$

where

$$\varepsilon'(\beta,N) = \frac{1}{2}\varepsilon(\beta,N)m^l\kappa(P) \quad (23)$$

such that with probability at least β we have:

$$A_{j_l}A_{j_{l-1}}\dots A_{j_1}(\mathbb{B}) \subset \frac{(\gamma^*(\omega_N)(1+\eta))^l}{\delta(\beta,\omega_N)}\mathbb{B},\\ \forall\,(j_1,\dots,j_l)\in M^l,$$

By Property 2.2, this means that with probability at least β:

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

which completes the proof of the first part of the theorem. Note that, the ratio $\frac{1}{2}$ introduced in the expression of ε' is, as we have already mentioned Section 4.2, due to the homogeneity of the system described in Property 2.3. Let us prove now that $\lim_{N\to\infty}\delta(\beta,\omega_N) = 1$ with probability 1.

We recall that, $\delta(\beta,\omega_N) = \alpha\left(\varepsilon(\beta,\omega_N)m^l\kappa(P(\omega_N))\right)$. We start by showing that $\kappa(P(\omega_N))$ is uniformly bounded in N. The optimization problem $\mathrm{Opt}(\omega_N)$ given in (8), with $\gamma^*(\omega_N)$ replaced by $\gamma^*(Z_l)(1+\frac{\eta}{2})$ is strictly feasible, and thus 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, $\mathrm{Opt}(\omega_N)$ admits $P_{\eta/2}$ as a feasible solution and thus the optimal value of $\mathrm{Opt}(\omega_N)$ is bounded by K. In other words, $\lambda_{\max}(P(\omega_N)) \le K$. Moreover, since $\lambda_{\max}(P(\omega_N)) \ge 1$, we also have $\det(P(\omega_N)) \ge 1$, which means that

$$\kappa(P(\omega_N)) = \sqrt{\frac{\lambda_{\max}(P(\omega_N))^n}{\det(P(\omega_N))}} \le \sqrt{K^n}. \quad (24)$$

We next show that for a fixed $\beta\in(0,1)$, $\lim_{N\to\infty}\varepsilon(\beta,N) = 0$. Note that, $\varepsilon(\beta,N)$ is intrinsically defined by the following equation:

$$1-\beta = \sum_{j=0}^{d}\binom{N}{j}\varepsilon^j(1-\varepsilon)^{N-j}.$$

We can then upper bound the term $1-\beta$ as in:

$$1-\beta \le (d+1)N^d(1-\varepsilon)^{N-d}. \quad (25)$$

We prove $\lim_{N\to\infty}\varepsilon(\beta,N) = 0$ by contradiction. Assume that $\lim_{N\to\infty}\varepsilon(\beta,N) \ne 0$. This means that, there exists some $c>0$ such that $\varepsilon(\beta,N) > c$ infinitely often. Then, consider the subsequence N_k such that $\varepsilon(\beta,N_k) > c$, $\forall\,k$. Then, by (25) we have:

$$1-\beta \le (d+1)N_k^d(1-\varepsilon)^{N_k-d} \le (d+1)N_k^d(1-c)^{N_k-d}\,\forall\,k\in\mathbb{N}.$$

Note that $\lim_{k\to+\infty}(d+1)N_k^d(1-c)^{N_k-d} = 0$. Therefore, 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 (24), we get:

$$\lim_{N\to\infty} m^l\kappa(P(\omega_N))\varepsilon(\beta,\omega_N) = 0.$$

By the continuity of the function I^{-1} this also implies: $\lim_{N\to\infty}\alpha\left(\varepsilon(\beta,\omega_N)m^l\kappa(P(\omega_N))\right) = 1$.

5 Experimental Results

5.1 Algorithm

Raphael's other upper bound: There is no conservatism in multiplying ε by m^l, as in the worst case this really happens: if $\varepsilon = 1/m$, it is well possible that one mode is totally forgotten, and that our δ must be equal to zero (because then all points are bad points). However, when multiplying by $\kappa(P)$, we are conservative, because this bound is exactly reached only at a single point on the ellipsoid. So, when we derive an upper bound on the size of the bad points on the sphere after changing the coordinates, we could also derive a "lower bound on the size of the good points". This gives a second upper bound on the size of the bad points by taking the complement. This one can never be larger than one because it is 1−(lower bound on the size of the good points). I think that to obtain that lower bound, we just have to replace lambdamax by lambdamin.

Algorithm: maybe explain we take for the bounds for the bisection on γ, 0 and U, with U the max of the norms y^Ty, and that for some precision η, we run the algorithm at most $\lceil\log_2(U/\eta)\rceil$.

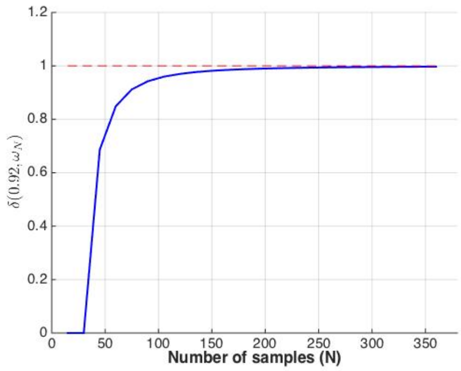


Fig. 7. Evolution of δ with increasing N.

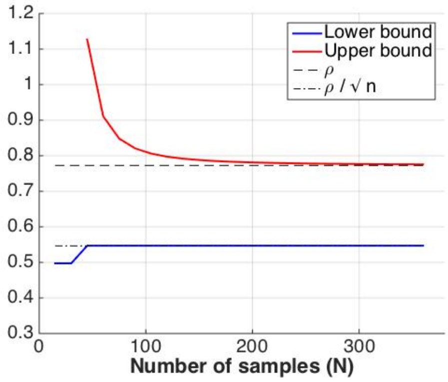


Fig. 8. Evolution of the upper and lower bounds on the JSR with increasing N, for $\beta = 0.92$.

5.2 Experimental results

To do: add Ayca's new plots

We illustrate our technique on a two-dimensional switched system with 4 modes. We fix the confidence level, $\beta = 0.92$, and compute the lower and upper bounds on the JSR for $N := 15 + 15k$, $k \in \{0, \ldots, 23\}$, according to Theorem 9 and Theorem 18, respectively. We illustrate the average performance of our algorithm over 10 different runs in Fig. 7 and Fig. 8. Fig. 7 shows the evolution of $\delta(\beta, N)$ as N increases. We illustrate that δ converges to 1 as expected. In Fig. 8, we plot the upper bound and lower bound for the JSR of the system computed by Theorem 18 and Theorem 9, respectively. To demonstrate the performance of our technique, we also provide the JSR approximated by the JSR toolbox [22], which turns out to be 0.7727. Note that, the plot for the upper bound starts from $N = 45$. This is due to the fact for $N = 15$, and $N = 30$, $\delta(\beta, \omega_N) = 0$, hence it is not possible to compute a nontrivial upper bound for these small values of N. As can be seen, the upper bound approaches to a close vicinity of the real JSR with approximately 200 samples. In addition, the gap between the upper and lower bound converges to a multiplicative factor of $\frac{\rho}{\sqrt{n}}$ as expected.

Note that, if we increase the dimension of the switched system, the convergence of δ to 1 will become much slower. We confirmed this via experiments up to dimension $n = 6$. For example, for dimension $n = 4$, it took $N = 5,000$ to $N = 10,000$ points to reach $\delta = 0.9$. We nevertheless observe convergence of the upper bound to $\rho(\mathcal{M})$, and convergence of the lower bound to $\frac{\rho(\mathcal{M})}{\sqrt{n}}$. The gap between these two limits is $\frac{\rho}{\sqrt{n}}$ and could be improved by considering a more general class of common Lyapunov functions, such as those that can be described by sum-of-squares polynomials [18]. We leave this for future work.

Finally, we randomly generate 10,000 test cases with systems of dimension between 2 and 7, number of modes between 2 and 5, and size of samples N between 30 and 800. We take $\beta = 0.92$ and we check if the upper bound computed by our technique is greater than the actual JSR of the system. We get 9873 positive tests, out of 10,000, which gives us a probability of 0.9873 of the correctness of 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.

5.3 Networked Control System

We now consider a linear time-invariant control system given as $x_{k+1} = Ax_k + Bu_k$, where we do not have access to its dynamics given by the matrices A and B. The control law is of the form $u_k = Kx_k$, where K is also unknown. The open-loop system is unstable with eigenvalues at $\{0.45, 1.1\}$. The controller stabilizes the system by bringing its eigenvalues to $\{0.8, -0.7\}$. The control input is transmitted over a wireless communication channel that is utilized by ℓ users, including the controller. The communication between the users and the recipients is performed based on the IEEE 802.15.4 MAC layer protocol [?], which is used in some of the proposed standards for control over wireless, e.g., WirelessHART [?]. This MAC layer integrates both guaranteed slots and contention based slots. In this example, we consider a beacon-enabled mode of the MAC protocol. In this setup, a centralized control user periodically synchronizes and configures all the users. This period is named Beacon Interval. This interval is divided into two intervals: active and inactive period. The active period is divided into 6 slots. The first 2 slots correspond to the contention access period (CAP), and the next 4 slots correspond to the collusion free period (CFP). In the CAP, the users can only send their message if the channel is "idle" with carrier-sense multiple access with collision avoidance (CSMA/CA). In the CFP however, each user has guaranteed time slots, during which there are no packet losses. In our example, the third and fourth slots are designated for the controller, while the fifth and sixth slots are allocated to the other users. Finally, during the inactive period, all users enter a low-power mode

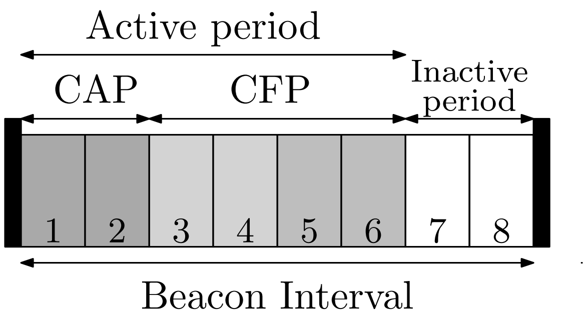


Fig. 9. The time allocation structure of the modified IEEE 802.15.4 MAC layer.

to save energy. We illustrate the overall structure of this communication protocol in Fig. 9. We now want to decide whether the resulting closed-loop network control system is stable by simulating it starting from different initial conditions.

Note that, the closed-loop dynamics of the underlying system when the controller is active is $A_c = A + BK$. Then, we can model the overall networked control system by the linear switched system $x_{k+1} = \bar{A}x_k$, where $\bar{A} \in \mathcal{M}$ and

$$\mathcal{M} = \{A^2A_c^2A^4, A_cAA_c^2A^4, AA_c^3A^4, A_c^4A^4\}. \tag{26}$$

Note that, in (26), each mode corresponds to a different utilization of the CFP by the users. For example, the mode defined by $A_cAA_c^2A^4$ is active when the first slot in the CFP is assigned to the controller and the second slot is assigned to the other users. We assume that all of the users using the channel have an equal probability of being assigned to a time slot during the CFP. Therefore, the probability of each mode in $\mathcal{M}$ being active is $\left\{\frac{1}{(\ell-1)^2}, \frac{1}{\ell(\ell-1)}, \frac{1}{(\ell-1)\ell}, \frac{1}{\ell^2}\right\}$, i.e., the modes given in (26) will not be active with the same probability. Hence, we make use of Remark **??** and update our bounds accordingly. Fig. 10 shows the computed bounds. As can be seen, approximately after 500 samples, the upper bound on the JSR drops below 1, which lets us decide that the given closed-loop networked control system is stable, with probability 0.95.

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 have randomly generated state measurements. Our goal is to understand how the observation of well-behaved trajectories *intrinsically* implies stability of a system. 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.

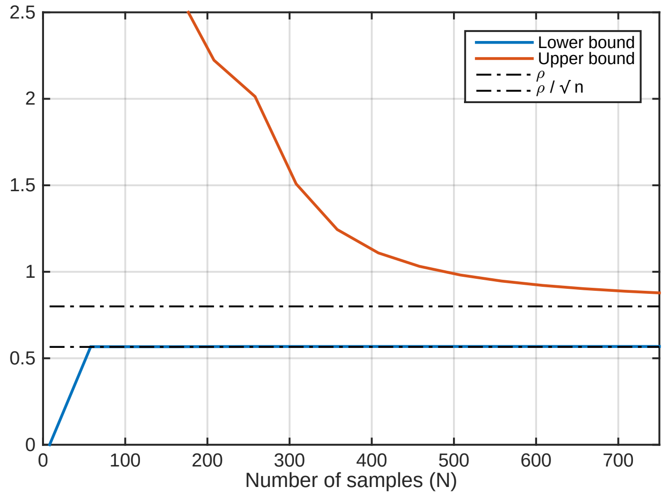


Fig. 10. The evolution of the computed upper and lower bounds on the JSR with respect to the number of simulations collected from the networked control system.

The novelty of our contribution is twofold: First, we use as standing assumption that the unknown system can be described by a switching 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 apply powerful techniques from chance constrained optimization. The application is not obvious, and relies on geometric properties of linear switched systems.

We believe that this guarantee is quite 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 or realistic systems. We are also improving the numerical properties of our technique by incorporating sum-of-squares optimization, and relaxing the sampling assumptions on the observations.

Remark 19 *In the above discussion, we introduce the concept of 'l-step CQLF', and showed that it allows to refine the initial $1/\sqrt(n)$ approximation provided by the CQLF method. In the switching systems literature, there are other techniques for refining this approximation, as for instance replacing the LMIs in Theorem 4 by Sum-Of-Squares (SOS) constraints; see [19] or [12, Theorem 2.16]. It seems that the concept of l-step CQLF is better suited for our purpose, as we briefly discuss below. We leave for further work a more systematic analysis of the behaviours of the different refining techniques.*

Acknowledgements

Acknowledgements, NSF etc.

References

[1] A. Balkan, P. Tabuada, J. V. Deshmukh, X. Jin, and J. Kapinski. Underminer: A framework for automatically identifying non-converging behaviors in black box system models. In *2016 International Conference on Embedded Software (EMSOFT)*, pages 1–10, Oct 2016.

[2] F. Blanchini, G. Fenu, G. Giordano, and F. A. Pellegrino. Model-free plant tuning. *IEEE Transactions on Automatic Control*, PP(99):1–1, 2016.

[3] V. D. Blondel and J. N. Tsitsiklis. Complexity of stability and controllability of elementary hybrid systems. *Automatica*, 35(3):479 – 489, 1999.

[4] R. Bobiti and M. 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.

[5] S. Boyd and L. Vandenberghe. *Convex Optimization*. Cambridge University Press, New York, NY, USA, 2004.

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

[7] G. Calafiore and M. Campi. The scenario approach to robust control design. *IEEE Transactions on Automatic Control*, 51(5):742–753, 2006.

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

[9] 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.

[10] P. S. Duggirala, S. Mitra, and M. Viswanathan. Verification of annotated models from executions. In *Proceedings of the Eleventh ACM International Conference on Embedded Software*, EMSOFT '13, pages 26:1–26:10, Piscataway, NJ, USA, 2013. IEEE Press.

[11] Z. Huang and S. Mitra. Proofs from simulations and modular annotations. In *Proceedings of the 17th International Conference on Hybrid Systems: Computation and Control*, HSCC '14, pages 183–192, New York, NY, USA, 2014. ACM.

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

[13] J. Kapinski, J. V. Deshmukh, and N. Sankaranarayanan, S.and 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.

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

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

[16] S. Liu, D. Liberzon, and V. Zharnitsky. On almost lyapunov functions for non-vanishing vector fields. In *2016 IEEE 55th Conference on Decision and Control (CDC)*, pages 5557–5562, Dec 2016.

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

[18] A. Papachristodoulou and S. Prajna. On the construction of Lyapunov functions using the sum of squares decomposition. In *Proceedings of 41st IEEE Conference on Decision and Control*, pages 3482–3487, Las Vegas, Nevada USA, December 2002.

[19] P. Parrilo and A. Jadbabaie. Approximation of the joint spectral radius using sum of squares. *Linear Algebra and its Applications*, 428(10):2385–2402, 2008.

[20] W. Rudin. *Real and complex analysis*. McGraw-Hill Book Co., New York, third edition, 1987.

[21] U. Topcu, A. Packard, and P. Seiler. Local stability analysis using simulations and sum-of-squares programming. *Automatica*, 44(10):2669 – 2675, 2008.

[22] G. Vankeerberghen, J. Hendrickx, and R. Jungers. Jsr: A toolbox to compute the joint spectral radius. In *Proceedings of the 17th International Conference on Hybrid Systems: Computation and Control*, HSCC '14, pages 151–156, New York, NY, USA, 2014. ACM.

A Precisions on Measures Definitions

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

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

We denote by $E_P^{\mathcal{A}}$ the sector induced by $\mathcal{A} \subset E_P$. In the particular case of the unit sphere, we instead write $\mathbb{S}^{\mathcal{A}}$. We can notice that $E_P^{E_P}$ is the volume in $\mathbb{R}^n$ defined by E_P: $E_P^{E_P} = \{x \in \mathbb{R}^n : x^T P x \leq 1\}$.

The spherical Borelian σ-algebra, denoted by $\mathcal{B}_{\mathbb{S}}$, is defined by

$$\mathcal{A} \in \mathcal{B}_{\mathbb{S}} \iff \mathbb{S}^{\mathcal{A}} \in \mathcal{B}_{\mathbb{R}^n}.$$

We provide $(\mathbb{S}, \mathcal{B}_{\mathbb{S}})$ with the classical, unsigned and finite uniform spherical measure σ^{n-1} defined by

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

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 $\sigma^{n-1}(\mathbb{S}) = 1$.

Since $P \in \mathcal{S}_{++}^n$, we recall that it can be written in its Cholesky form (16). Note that, L^{-1} maps the elements of $\mathbb{S}$ to E_P. Then, we define the measure on the ellipsoid σ_P on the σ-algebra $\mathcal{B}_{E_P} := L^{-1}(\mathcal{B}_{\mathbb{S}})$, where $\forall\, \mathcal{A} \in \mathcal{B}_{E_P}$, $\sigma_P(\mathcal{A}) = \sigma^{n-1}(L\mathcal{A})$.

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 provide (M, Σ_M) with the uniform measure μ_M. Similarly, 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 denote 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 $(Z_l, \mathcal{B}_{\mathbb{S}} \bigotimes(\Sigma_{M^l}))$, we define the product measure $\mu_l = \sigma^{n-1} \otimes \mu_{M^l}$. Note that, μ_l is a uniform measure on Z_l and $\mu_l(Z_l) = 1$.

B Proof of Corollary 11

We prove here Corollary 11 stated in section 4.1.

Let $\mathbb{S}' = \pi_{\mathbb{S}}(V)$. We know that Σ_{M^l} is the disjoint union of its 2^{m^l} elements $\{\mathcal{M}_i^l, i \in \{1, 2, \ldots, 2^{m^l}\}\}$. Then V can be written as the disjoint union $V = \sqcup_{1 \le i \le 2^{m^l}} (\mathcal{S}_i, \mathcal{M}_i^l)$ where $\mathcal{S}_i \in \Sigma_{\mathbb{S}}$. We notice that $\mathbb{S}' = \sqcup_{1 \le i \le 2^{m^l}} \mathcal{S}_i$, and

$$\sigma^{n-1}(\mathbb{S}') = \sum_{1 \le i \le 2^{m^l}} \sigma^{n-1}(\mathcal{S}_i).$$

We have

$$\begin{aligned}\mu_l(V) &= \mu_l \left(\sqcup_{1 \le i \le 2^{m^l}} (\mathcal{S}_i, \mathcal{M}_i^l) \right) = \sum_{1 \le i \le 2^{m^l}} \mu_l \left(\mathcal{S}_i, \mathcal{M}_i^l \right) \\ &= \sum_{1 \le i \le 2^{m^l}} \sigma^{n-1} \otimes \mu_{M^l} \left(\mathcal{S}_i, \mathcal{M}_i^l \right) \\ &= \sum_{1 \le i \le 2^{m^l}} \sigma^{n-1}(\mathcal{S}_i) \mu_{M^l}(\mathcal{M}_i^l).\end{aligned}$$

Note that we have $\min_{(j_1,\ldots,j_l)\in M^l} \mu_{M^l}(\{j_1, \ldots, j_l\}) = \frac{1}{m^l}$. Then since $\forall\, i, \mu_{M^l}(\mathcal{M}_i^l) \ge \min_{(j_1,\ldots,j_l)\in M^l} \mu_{M^l}(\{j_1, \ldots, j_l\}) = \frac{1}{m^l}$, we get:

$$\sigma^{n-1}(sphere') \le \frac{\mu_l(V)}{\frac{1}{m^l}} \le m^l \varepsilon. \tag{B.1}$$

This means that

$$\begin{aligned}(A_{j_l} A_{j_{l-1}} \ldots A_{j_1} x)^T P (A_{j_l} A_{j_{l-1}} \ldots A_{j_1} x) \le \gamma^{2l} x^T P x, \\ \forall\, x \in \mathbb{S} \setminus \mathbb{S}', \forall\, (j_1, \ldots, j_l) \in M^l,\end{aligned} \tag{B.2}$$

where $\sigma^{n-1}(\mathbb{S}') \le m^l \varepsilon$.

C Preliminary Results on Spherical Caps

Before proceeding to the proof of Proposition 14, we first introduce some necessary definitions and related background on spherical caps. We recall that a *spherical cap* on $\mathbb{S}$ for a given hyperplane $c^T x = k$ is defined by $\mathcal{C}_{c,k} := \{x \in \mathbb{S} : c^T x > k\}$.

Remark 20 *Consider the spherical caps $\mathcal{C}_{c,k_1}$ and $\mathcal{C}_{c,k_2}$ such that $k_1 > k_2$, then we have:*

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

Remark 21 *The distance between the point $x = 0$ and the hyperplane $c^T x = k$ is $\frac{|k|}{\|c\|}$.*

We also recall that we defined in section 4.2 the function $\Delta : \wp(\mathbb{S}) \to [0, 1]$ as $\Delta(X) := \sup\{r : r\mathbb{B} \subset \text{convhull } (\mathbb{S} \setminus X)\}$.

Lemma 22 *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:

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

Then the following equalities hold:

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

Corollary 23 *Consider the spherical caps $\mathcal{C}_{c,k_1}$ and $\mathcal{C}_{c,k_2}$ such that $k_1 \le k_2$. Then we have:*

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

Lemma 24 *For any set $X \subset \mathbb{S}$, there exist c and k such that $\mathcal{C}_{c,k}$ satisfies: $\mathcal{C}_{c,k} \subset X$, and $\Delta(\mathcal{C}_{c,k}) = \Delta(X)$.*

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:

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

Next, consider the hyperplane which is tangent to the ball $\Delta(X)\mathbb{B}$ at x^*, which we denote by $\left\{x : c^T x = k\right\}$. 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 defined by this tangent plane i.e., $\mathcal{C}_{c,k}$. Then, by Lemma 22 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 $x \in \mathbb{S} \setminus X \subset \text{convhull}\,(\mathbb{S} \setminus X)$. Since $x \in \mathcal{C}_{c,k}$, we have $c^T x > k$. But due to the fact that $x \in \text{convhull}\,(\mathbb{S} \setminus X)$, we also have $c^T x \leq k$, which leads to a contradiction. Therefore, $\mathcal{C}_{c,k} \subset X$.

We are now able to prove Proposition 13 given in section 4.2, which states that, for any $\varepsilon \in (0,1)$, the function $\Delta(X)$ attains its minimum over $\mathcal{X}_\varepsilon$ for some X which is a spherical cap.

PROOF. We prove this via contradiction. Assume that there exists no spherical cap in $\mathcal{X}_\varepsilon$ such that $\Delta(X)$ attains its minimum. This means there exists an $X^* \in \mathcal{X}_\varepsilon$, where X^* is not a spherical cap and $\arg\min_{X \in \mathcal{X}_\varepsilon}(\Delta(X)) = X^*$. By Lemma 24, 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 $\sigma^{n-1}(\mathcal{C}_{c,k}) < \varepsilon$.

Then, the spherical cap $\mathcal{C}_{c,\tilde{k}}$ with $\sigma^{n-1}(\mathcal{C}_{c,\tilde{k}}) = \varepsilon$, satisfies $\tilde{k} < k$ by Remark 20. This implies

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

by Corollary 23. Therefore, $\Delta(\mathcal{C}_{c,\tilde{k}}) < \Delta(X^*)$. This is a contradiction since we initially assumed that $\Delta(X)$ attains its minimum over $\mathcal{X}_\varepsilon$ at X^*.

D Proof of Lemma 14

We present now a proof of Lemma 14. By Proposition 13 we know that:

$$\alpha(\varepsilon) = \Delta(\mathcal{C}_{c,k}), \tag{D.1}$$

for some spherical cap $\mathcal{C}_{c,k} \subset \mathbb{S}$, where $\sigma^{n-1}(\mathcal{C}_{c,k}) = \varepsilon$. It is known (see e.g. [15]) that the area of such $\mathcal{C}_{c,k}$, is given by the equation:

$$\sigma^{n-1}(\mathcal{C}_{c,k}) = \frac{I\left(1 - \Delta(\mathcal{C}_{c,k})^2; \frac{n-1}{2}, \frac{1}{2}\right)}{2}. \tag{D.2}$$

Since, $\sigma^{n-1}(\mathcal{C}_{c,k}) = \varepsilon$, we get the following set of equations:

$$\begin{aligned}
\varepsilon &= \frac{I\left(1 - \Delta(\mathcal{C}_{c,k})^2; \frac{n-1}{2}, \frac{1}{2}\right)}{2} \\
1 - \Delta(\mathcal{C}_{c,k})^2 &= I^{-1}\left(2\varepsilon; \frac{n-1}{2}, \frac{1}{2}\right) \\
\Delta(\mathcal{C}_{c,k})^2 &= 1 - I^{-1}\left(2\varepsilon; \frac{n-1}{2}, \frac{1}{2}\right).
\end{aligned} \tag{D.3}$$

Then, the inequalities (D.1) and (D.3) imply the inclusion given in (13), which concludes the proof.