Available online at www.sciencedirect.com

ScienceDirect

IFAC Papers Online CONFERENCE PAPER ARCHIVE

IFAC PapersOnLine 58-11 (2024) 31–36

Data-Driven Stability Analysis of Switched Linear Systems Using Adaptive Sampling ⋆

Alexis Vuille * **Guillaume O. Berger** * **Raphaël M. Jungers** *

* *UCLouvain, Louvain-la-Neuve, Belgium (e-mail: firstname.lastname@uclouvain.be).*

Abstract: In the realm of system analysis, data-driven methods have gained a lot of attention in recent years. We introduce a new innovative approach for the data-driven stability analysis of switched linear systems which is adaptive sampling. Our aim is to address limitations of existing approaches, in particular, the fact that these methods suffer from ill-conditioning of the optimal Lyapunov function, which is a direct consequence of the way the data is collected by sampling uniformly the state space. Our adaptive-sampling approach consists in a two-step procedure, in which an optimal sampling distribution is estimated in the first step from data collected with a non-optimal distribution, and then, in the second step, new data points are sampled according to the identified distribution to establish the final probabilistic guarantee for the convergence rate of the system. Numerical experiments show the efficiency of our approach, namely, in terms of the total number of data points needed to guarantee stability of the system with given confidence.

Keywords: data-driven methods, chance-constrained optimization, stability analysis, switched linear systems

1. INTRODUCTION

In recent years, data-driven methods have gained a lot of attention for the study of cyber-physical systems because of the increasing number of applications in which no model of the system is available. At the same time, data has become more and more accessible due to the outbreak of cheap, accurate sensors, user feedback and open-source databases. Finally, statistical learning, the mathematical field of learning from data, has known many great advances in recent years both in theory and practice (Shalev-Shwartz and Ben-David, 2014). All this together opened the door to a new era in control theory where control and system analysis is made from data harvested from observation of the system and comes with formal guarantees of correctness; we refer the reader to (Mitra, 2021, Chapter 11) for an introduction and further references on data-driven verification and control of cyber-physical systems.

In this paper, we consider a prototypical class of cyber-physical systems, known as switched linear systems (Sun and Ge, 2011). Those systems consist in several linear modes among which the system can switch over time. They appear naturally in a wide range of applications (Liberzon, 2003), or as approximations of more complex systems. A crucial question in the study of switched linear systems is their stability analysis (Sun and Ge, 2011). It turns out that the stability analysis of these systems is a very challenging problem in general, even when the model of the system is available (Jungers, 2009). For instance, approximating the rate of convergence of the system (known as the Joint Spectral Radius, or JSR) is known to be NP-hard. Nevertheless, several approximation techniques have been proposed in the last decades leading to good results in the model-based setting (Jungers, 2009; Ahmadi et al., 2014).

Because a model of the system is not always available, several approaches were proposed in recent years for the data-driven analysis of the JSR of switched linear systems (Kenanian et al., 2019; Berger et al., 2021; Rubbens et al., 2021; Wang et al., 2021). These approaches use advanced tools from statistical learning, such as scenario optimization (Calafiore and Campi, 2006), to derive bounds on the JSR and provide probabilistic guarantees on the correctness of the bound.

In this paper, we first discuss two important weaknesses of the previous data-driven approaches (Sec. 3). Namely, (i) that the bounds are often very conservative, and (ii) that they strongly suffer from a bad choice of the distribution used to sample the data. We then propose two approaches to overcome these weaknesses. First, we provide a heuristic formula to replace the bound (Sec. 4), that can be useful when the JSR needs to be approximated from data, but no formal guarantee is needed. Numerical experiments demonstrate the accuracy of the heuristic. Then, to alleviate the dependence on a bad choice of sampling distribution, we propose a two-step approach in which the optimal distribution is first estimated from a batch of data and then new data points are sampled according to this distribution, from which the probabilistic bound on the JSR is derived (Sec. 5). Furthermore, we

⋆ GB is a FNRS Postdoctoral Researcher. RJ is a FNRS honorary Research Associate. This project has received funding from the European Research Council (ERC) under the European Union's Horizon 2020 research and innovation programme under grant agreement No 864017 – L2C.

10.1016/j.ifacol.2024.07.421

discuss the choice of the data sizes to use in each step of the two-step approach (Sec. 6); in particular, we propose a heuristic formula for choosing the data sizes that aims at providing the lowest bound on the JSR, given a total budget of data points. For that, we use the heuristic found in Sec. 4, because choosing the data sizes does not require a formal bound but rather an accurate approximation of the JSR. Finally, we demonstrate on numerical examples the feasibility of our approach and its advantage compared to state-of-the-art approaches in the literature.

Related Works

Data-driven techniques for the verification and control of black-box dynamical systems and cyber-physical systems have received a lot of attention in recent years (Duggirala et al., 2013; Margellos et al., 2014; Coulson et al., 2019; Nejati et al., 2023). Data-driven stability analysis of switched linear systems is studied in Kenanian et al. (2019); Berger et al. (2021); Rubbens et al. (2021); Wang et al. (2021). This work improves on the existing approaches by identifying their main limitations and proposing a way to fix them via adaptive sampling. We further demonstrate in numerical examples the strong benefits of the proposed solution over the previous approaches.

Notation. $\mathbb{R}^{n\times n}_{\succ 0}$ denotes the set of positive definite $n \times n$ matrices. For a natural number N, $[N]$ denotes the set $\{1, \dots, N\}$. We denote the unit sphere in $\mathbb{R}^n$ by $\mathbb{S}^{n-1}$.

Proofs can be found in the extended version of the paper on arXiv.

2. PROBLEM STATEMENT

We consider a discrete-time *switched linear system* with m modes:

$$x(t+1) \in \{Ax(t) : A \in \mathcal{A}\}, \tag{1}$$

wherein $\mathcal{A} = \{A_1, \dots, A_m\} \subseteq \mathbb{R}^{n\times n}$ is a set of m matrices in $\mathbb{R}^{n\times n}$. Since $\mathcal{A}$ characterizes the system in (1), in the following, we will refer to the system simply by $\mathcal{A}$. A *trajectory* of $\mathcal{A}$ is a function $x : \mathbb{N} \to \mathbb{R}^n$ such that for all $t \in \mathbb{N}$, the condition in (1) holds.

We are interested in the stability of $\mathcal{A}$. We remind that $\mathcal{A}$ is *asymptotically stable* if all trajectories of $\mathcal{A}$ converge to the origin. The rate of exponential convergence is called the *Joint Spectral Radius* (JSR) of $\mathcal{A}$.

Definition 1. The *joint spectral radius* of $\mathcal{A}$, denoted by $\rho(\mathcal{A})$, is the infimum of all $r \geq 0$ for which there exists $C \geq 1$ such that every trajectory x of $\mathcal{A}$ satisfies that for all $t \in \mathbb{N}$, $\|x(t)\| \leq Cr^t\|x(0)\|$.

2.1 Quadratic Approximation of the JSR

The JSR is notoriously difficult to approximate, even when the matrices in $\mathcal{A}$ are known (Jungers, 2009). One way to obtain an upper bound on the JSR is by finding a quadratic "Lyapunov" function for the system, and the associated contraction rate of the system.

Definition 2. Given a positive definite matrix $P \in \mathbb{R}^{n\times n}_{\succ 0}$, we define the *contraction rate* of $\mathcal{A}$ with respect to P by

$$\rho(\mathcal{A}, P) = \max_{x\in\mathbb{R}^n\setminus\{0\},\, A\in\mathcal{A}} \sqrt{\frac{(Ax)^\top P(Ax)}{x^\top Px}}.$$

The contraction rate is an upper bound on the JSR (see, e.g., Jungers, 2009, Proposition 2.8):

Theorem 3. For any $P \in \mathbb{R}^{n\times n}_{\succ 0}$, $\rho(\mathcal{A}) \leq \rho(\mathcal{A}, P)$.

Theorem 3 motivates the notion of *best quadratic approximation* of the JSR of $\mathcal{A}$ defined as the smallest contraction rate with respect to a positive definite matrix:

$$\rho_{\text{quad}}(\mathcal{A}) = \inf_{P\in\mathbb{R}^{n\times n}_{\succ 0}} \rho(\mathcal{A}, P).$$

By Theorem 3, ρ_{quad} is an upper bound on ρ.

Note that when $\mathcal{A}$ is known, $\rho_{\text{quad}}(\mathcal{A})$ can be approximated with accuracy $\epsilon > 0$ efficiently with respect to n, $|\mathcal{A}|$ and $\log(\epsilon)$ because the problem can be formulated as a quasi-convex optimization problem (Jungers, 2009, Sec. 2.3.7).

Note that in general the quadratic approximation is not tight i.e., $\rho_{\text{quad}}(\mathcal{A}) < \rho(\mathcal{A})$ and might be too conservative in several situations. Nevertheless it often provides sufficiently accurate upper bounds on the JSR.

In some situations, we want to restrict the set of "Lyapunov" functions to avoid ill-conditioned ones. Therefore, we introduce the class of δ-*conditioned* matrices as the set of symmetric matrices P satisfying $I \preceq P \preceq \delta I$. We then define the *best quadratic approximation of the JSR of* $\mathcal{A}$ *by a* δ*-conditioned matrix* by

$$\rho_{\text{quad}}(\mathcal{A}, \delta) = \min_{P\in\mathbb{R}^{n\times n}_{\succ 0},\, I\preceq P\preceq\delta I} \rho(\mathcal{A}, P).$$

By Theorem 3, $\rho(\mathcal{A}) \leq \rho_{\text{quad}}(\mathcal{A}, \delta)$. Furthermore, it holds that $\lim_{\delta\to\infty} \rho_{\text{quad}}(\mathcal{A}, \delta) = \rho_{\text{quad}}(\mathcal{A}) =: \rho_{\text{quad}}(\mathcal{A}; \infty)$.

In summary, computing the JSR is generally intractable, but we can compute efficiently upper bounds on it by using quadratic "Lyapunov" functions. However, this approach, as many other approaches for approximating the JSR, are model-based because they require the knowledge of $\mathcal{A}$. By contrast, data-driven methods, studied in this paper, allow to approximate the JSR of $\mathcal{A}$, without knowing $\mathcal{A}$, by using observation data (such as a finite set of trajectories) of $\mathcal{A}$. Considering the inherent challenges of approximating the JSR even in the model-based setting, we will focus on the problem of approximating the quadratic approximation of the JSR in a data-driven fashion. In other words, we seek data-driven upper bounds on the best quadratic approximation of the JSR (ρ_{quad}), and ideally these bounds should be as close as possible to ρ_{quad}.

2.2 Data-Driven Analysis and Random Data Collection

In the setting considered in this paper, we collect data by setting the system to an initial state x and observing the state y after one time step. Repeating this process N times yields a data set comprising N one-step trajectories $(x_i, y_i) \in \mathbb{R}^n \times \mathbb{R}^n$ for each $i \in [N]$. As the system is nondeterministic, the matrix in $\mathcal{A}$ responsible for each trajectory cannot be chosen or known. Nevertheless, we assume that the system's switching is a stochastic process, wherein each matrix in $\mathcal{A}$ has a nonzero probability of being applied independently at each time step.

Assumption 4. For each $i \in [N]$, let X_i and Y_i be the random variables describing the initial and terminal states in the i^{th} experiment. We assume that there exists $\alpha \in (0, 1]$ such that for all $A \in \mathcal{A}$ and $i \in [N]$,

$$\mathbb{P}[Y_i = AX_i \mid X_i = x_i,\ (X_j, Y_j) = (x_j, y_j)\ \forall\, j \neq i] \geq \alpha.$$

Regarding the choice of the initial state of each experiment, we assume that x_i can be chosen for each $i \in [N]$. However, because the observation already contains nondeterminism (in the choice of A), we will choose these points randomly according to some distribution that we can design. In particular, we will consider the uniform spherical distribution (possibly after applying a change of basis): [1]

Definition 5. We say that a random variable X has *uniform distribution* on $\mathbb{S}^{n-1}$ if X has the same distribution as $U/\|U\|$ where U is a vector of n independent standard Gaussian random variables [2].

3. PREVIOUS DATA-DRIVEN BOUNDS ON THE JSR

In this section, we present a data-driven bound on the JSR. This bound was proposed in Kenanian et al. (2019), and refined in Berger et al. (2021); other bounds were proposed in Rubbens et al. (2021) and Wang et al. (2021). Note that since the data is random, the data-driven bounds can provide only probabilistic guarantees on the value of the JSR, i.e., one can only guarantee with *some confidence* that the bound computed from the data is a valid upper bound on the JSR. After presenting the bound and the associated guarantees, we discuss some limitations.

3.1 Data-Driven Bound and Probabilistic Guarantees

Given a data set $\mathcal{D} = \{(x_i, y_i)\}_{i=1}^N$ consisting in N one-step trajectories, we formulate the problem of finding a positive definite matrix P with the smallest *data-based contraction rate* defined by

$$\hat{\rho}(\mathcal{D}, P) = \max_{i\in[N]} \sqrt{\frac{y_i^\top P y_i}{x_i^\top P x_i}}.$$

To do that, we solve the following optimization problem:

$$\begin{aligned} \inf_{\gamma\geq 0,\, P\in\mathbb{R}^{n\times n}_{\succ 0}} \quad & \gamma, \\ \text{s.t} \quad & I \preceq P \preceq \delta I, \\ & y_i^\top P y_i \leq \gamma^2 x_i^\top P x_i \quad \forall i \in [N], \end{aligned} \tag{2}$$

wherein $\delta \in [1, \infty]$ is a parameter (if $\delta = \infty$, then there is no upper bound on the eigenvalues of P). If $\delta < \infty$, then the inf is actually a min. The optimal cost of Prob. (2) is denoted by $\gamma_\star(\mathcal{D}, \delta)$, and the optimal solution if it exists is denoted by $P_\star(\mathcal{D}, \delta)$. We assume without loss of generality that if $P_\star(\mathcal{D}, \delta)$ exists, then it is unique; this can be done, e.g., by using a tie-breaking rule (Berger et al., 2021).

The main result regarding Prob. (2) that appears in Berger et al. (2021) is that we can guarantee an upper bound on the JSR with some predetermined confidence level if some very mild and technical assumptions on the matrices in $\mathcal{A}$ holds.

Assumption 6. The matrices in $\mathcal{A}$ do not have the Barabanov property, meaning that for each $A \in \mathcal{A}$, either A is not diagonalizable or A has at least two eigenvalues with different modulus.

[1] Because of the scaling invariance, a data point (x_i, y_i) carries the same information as the data point $(\lambda x_i, \lambda y_i)$ for every $\lambda \neq 0$. This is why it is sufficient to sample on the unit sphere $\mathbb{S}^{n-1}$ in $\mathbb{R}^n$.

[2] I.e., $U = [U_1, \dots, U_n]^\top$ where $U_1, \dots, U_n$ are independent and for each $i \in [n]$, $U_i \sim \mathcal{N}(0, 1)$.

We are now able to present the main result from Berger et al. (2021) regarding Prob. (2):

Theorem 7. Assume that $N \geq d := \frac{n(n+1)}{2}$, that $\{x_i\}_{i=1}^N$ are sampled independently and uniformly on $\mathbb{S}^{n-1}$ and that Assumptions 4 and 6 hold. Let $\beta \in (0, 1]$. Then, with probability at least $1 - \beta$ on the sampling of $\mathcal{D} := \{(x_i, y_i)\}_{i=1}^N$, it holds that [3]

$$\begin{aligned} \gamma_\star(\mathcal{D}, \delta) \leq \rho_{\text{quad}}(\mathcal{A}, \delta) \leq \rho(\mathcal{A}, P_\star(\mathcal{D}, \delta)) \leq \\ \gamma_\star(\mathcal{D}, \delta) \cdot f(\beta, \kappa(P_\star(\mathcal{D}, \delta)), N, d, \alpha, n), \end{aligned} \tag{3}$$

wherein

- $\kappa(P) = \sqrt{\frac{\det(P)}{\lambda_{\min}(P)^n}}$;
- $f(\beta, k, N, d, \alpha, n) = \frac{1}{\sqrt{1 - I^{-1}\left(\frac{k}{\alpha}\Phi^{-1}(\beta; d-1; N); \frac{n-1}{2}; \frac{1}{2}\right)}}$;
- $I^{-1}(y; a; b)$ is the inverse incomplete regularized beta function, i.e., it is the unique $x \in [0, 1]$ such that
$$I(x; a; b) := \frac{\int_0^x t^{a-1}(1-t)^{b-1}\mathrm{d}t}{\int_0^1 t^{a-1}(1-t)^{b-1}\mathrm{d}t} = y;$$
- $\Phi^{-1}(\beta; \zeta; N)$ is the unique $\epsilon \in [0, 1]$ such that
$$\Phi(\epsilon; \zeta; N) := \sum_{i=0}^{\zeta} \binom{N}{i} \epsilon^i (1-\epsilon)^{N-i} = \beta.$$

Before discussing the limitations of Theorem 7 (Sec. 3.2), which motivated this work, we present below two straightforward extensions of Theorem 7.

A-Priori Tightness The inflation factor

$$\bar{f} := f(\beta, \kappa(P_\star(\mathcal{D}, \delta)), N, d, \alpha, n)$$

reflects the tightness of the bound on $\rho_{\text{quad}}(\mathcal{A}, \delta)$ (i.e., how close is the upper bound to the quantity of interest) since

$$\gamma_\star(\mathcal{D}, \delta) \leq \rho_{\text{quad}}(\mathcal{A}, \delta) \leq \gamma_\star(\mathcal{D}, \delta) \cdot \bar{f}.$$

(with confidence $1-\beta$). As we see, the factor $\bar{f}$ depends on the sample set $\mathcal{D}$ via $\kappa(P_\star(\mathcal{D}, \delta))$. Therefore, we cannot fix the tightness a priori, i.e., before knowing $\mathcal{D}$. Nevertheless, we can upper-bound the tightness by noting that

$$I \preceq P \preceq \delta I \quad \Longrightarrow \quad \frac{\det(P)}{\lambda_{\min}(P)^n} \leq \delta^{n-1}.$$

It follows that $\kappa(P_\star(\mathcal{D}, \delta)) \leq \delta^{(n-1)/2}$. Thus, we can obtain an upper bound on $\bar{f}$ that is independent of $\mathcal{D}$:

$$\bar{f} \leq \tilde{f} := f(\beta, \delta^{(n-1)/2}, N, d, \alpha, n). \tag{4}$$

Constrained P Theorem 7 readily extends to the case where P in Prob. (2) is further constrained to be in a linear subspace $U \subseteq \mathbb{R}^{n\times n}$. Indeed, if d_u is the dimension of U, then Theorem 7 applies with d equals d_u. More precisely, (3) just needs be replaced by

$$\begin{aligned} \gamma_\star(\mathcal{D}, \delta) \leq \rho(\mathcal{A}, P_\star(\mathcal{D}, \delta)) \leq \\ \gamma_\star(\mathcal{D}, \delta) \cdot f(\beta, \kappa(P_\star(\mathcal{D}, \delta)), N, d_u, \alpha, n). \end{aligned} \tag{5}$$

For a proof, see Berger et al. (2021).

3.2 Limitations of the Data-Driven Bound

The data-driven bound on $\rho_{\text{quad}}(\mathcal{A}, \delta)$ in Theorem 7 suffers from two important limitations that we explain here.

[3] Note that the first and second inequalities in (3) are always satisfied, while the third one is guaranteed to hold with probability at least $1 - \beta$.

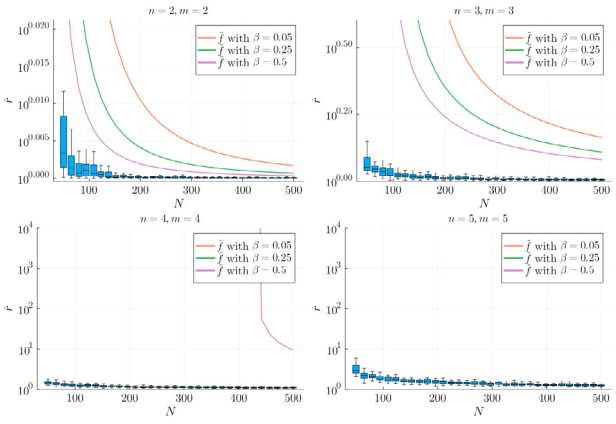

Fig. 1. Boxplots of the ratio $\bar{r}$ defined in (6) for different dimensions n and data sizes $N = |\mathcal{D}|$. For each n, a different randomly generated system with $m := |\mathcal{A}| = n$ matrices was used. For a given n and for each N, the boxplots were obtained by running 32 experiments, with different randomly sampled data sets $\mathcal{D}$. We also represented the bound $\tilde{f} := f(\beta, \delta^{(n-1)/2}, N, d, \alpha, n)$ on the inflation factor $\bar{f}$ for different values of β, and for $\delta = 5$ and $\alpha = 1/m$. For the bottom plots, some curves are not visible because their value is so large ($> 10^4$) that it is outside of the vertical range.

Conservative Upper Bound In numerical experiments, we observe that the upper bound is often conservative, in the sense that $\gamma_\star(\mathcal{D}, \delta) \cdot \tilde{f}$ is often quite large compared to $\rho_{\text{quad}}(\mathcal{A}, \delta)$. This is illustrated in Fig. 1 for randomly generated systems in dimension $n \in \{2, 3, 4, 5\}$. Namely, for different values of N, we sampled N one-step trajectories as described in Theorem 7 and we computed the ratio

$$\bar{r} := \frac{\rho(\mathcal{A}, P_\star(\mathcal{D}, \delta))}{\gamma_\star(\mathcal{D}, \delta)}. \tag{6}$$

We then compare $\bar{r}$ with the a-priori upper bound $\tilde{f}$ on the inflation factor $\bar{f}$, introduced in (4). The ratio of $\tilde{f}$ over $\bar{r}$ measures how conservative the probabilistic upper bound in Theorem 7 is since

$$\frac{\gamma_\star(\mathcal{D}, \delta) \cdot \tilde{f}}{\rho_{\text{quad}}(\mathcal{A}, \delta)} \geq \frac{\gamma_\star(\mathcal{D}, \delta) \cdot \tilde{f}}{\rho(\mathcal{A}, P_\star(\mathcal{D}, \delta))} = \frac{\tilde{f}}{\bar{r}}.$$

In Fig. 1, we observe that this ratio is quite large, especially when n increases and N decreases.

Uniform Sampling Another important limitation of the data-driven bound in Theorem 7 is the factor $\kappa(P_\star(\mathcal{D}, \delta))$. This factor reflects the fact that we use a uniform sampling on the sphere for selecting the initial states x_i for $i \in [N]$, while to be optimal we should use a *uniform sampling on the ellipsoid corresponding to* $P_\star(\mathcal{D}, \delta)$, in a sense that we will precise in Sec. 5. A large value of $\kappa(P_\star(\mathcal{D}, \delta))$ implies that the sampling is far from being uniform on the ellipsoid. Consequently, the factor $f(\beta, \kappa(P_\star(\mathcal{D}, \delta)), N, d, \alpha, n)$ in the upper bound in (3) will blow up (see the extended version on arXiv for an illustration).We are thus facing the following *chicken-and-egg* problem: knowing the ellipsoid requires the data set $\mathcal{D}$ in order to compute $P_\star(\mathcal{D}, \delta)$; and reversely, in order to sample the data set $\mathcal{D}$ uniformly on the ellipsoid, we need the ellipsoid. The main contribution of this work is precisely to solve this problem by considering two data sets $\mathcal{D}_1$ and $\mathcal{D}_2$: the first one is used to compute an ellipsoid, and the second one is obtained by sampling uniformly on the ellipsoid and used to compute the upper bound on $\rho_{\text{quad}}(\mathcal{A}, \delta)$.

4. ESTIMATION OF THE DATA-DRIVEN BOUND'S TIGHTNESS

As noted in Sec. 3.2, the upper bound

$$\tilde{f} := f(\beta, \delta^{(n-1)/2}, N, d, \alpha, n)$$

on the ratio $\bar{r}$ in (6) that measures the ratio of $\rho(\mathcal{A}, P_\star(\mathcal{D}, \delta))$ over $\gamma_\star(\mathcal{D}, \delta)$ is conservative. In this section, we propose a heuristic value $\tilde{f}_{\text{heur}}$ that will be useful to obtain in practice a more accurate estimation of $\bar{r}$, even though it does not come with a probabilistic guarantee. The heuristic provided in this section will be reused in section 6 to build another heuristic based on it. The proposed heuristic is

$$\tilde{f}_{\text{heur}} = f(\beta, \delta^{n/4}, N - d, 1, \alpha, n). \tag{7}$$

We motivate this choice in the extended version available on arXiv.

5. ADAPTIVE SAMPLING: A TWO-STEP APPROACH

The primary goal of adaptive sampling is to address the second limitation of the approach used in Theorem 7 (see Sec. 3.2): namely, that the uniform sampling on the sphere is not optimal and should instead be made uniformly on the ellipsoid associated to $P_\star(\mathcal{D}, \delta)$. However, as noted in Sec. 3.2, we are facing a chicken-and-egg problem because we do not know $P_\star(\mathcal{D}, \delta)$ before sampling $\mathcal{D}$. To resolve this issue, we propose a *two-step approach* in which we first use a data set $\mathcal{D}_1$ to compute $P_\star(\mathcal{D}_1, \delta)$, and then, we build a new data set $\mathcal{D}_2$ by sampling initial states uniformly on the ellipsoid associated to $P_\star(\mathcal{D}_1, \delta)$.

5.1 Two-Step Approach

The two-step approach to compute bounds on $\rho_{\text{quad}}(\mathcal{A}, \delta)$ is described by the following pseudo-code:

Input: A switched linear system $\mathcal{A}$ in the form of a random oracle Σ where each matrix has a probability at least $\alpha > 0$ of being applied. Two positive integers N_1 and N_2. Two conditioning parameters $\delta_1 \geq 1$ and $\delta_2 \geq 1$.

Step 1:

(1) Build a data set $\mathcal{D}_1 = \{(x_i, y_i)\}_{i=1}^{N_1}$ of N_1 one-step trajectories of $\mathcal{A}$ as follows: for each $i \in [N_1]$, sample the initial state x_i uniformly at random on $\mathbb{S}^{n-1}$ and observe the state y_i after feeding to Σ;
(2) Solve Prob. (2) with $\mathcal{D} = \mathcal{D}_1$ to get $P_\star(\mathcal{D}_1, \delta_1)$.

Step 2:

(1) Compute the change of basis $B = P_\star(\mathcal{D}_1, \delta_1)^{-1/2}$.
(2) Build a data set $\mathcal{D}_2 = \{(x_i, y_i)\}_{i=1}^{N_2}$ of N_2 one-step trajectories of $\mathcal{A}$ as follows: for each $i \in [N_2]$, sample a point x_i uniformly at random on $\mathbb{S}^{n-1}$ and define the initial state as $x_i' = Bx_i$, then observe the state y_i' after feeding to Σ, and finally, define $y_i = B^{-1}y_i'$;
(3) Solve Prob. (2) with $\mathcal{D} = \mathcal{D}_2$ to get $\gamma_\star(\mathcal{D}_2, \delta_2)$ and $P_\star(\mathcal{D}_2, \delta_2)$.

(4) Output $\gamma_\star(\mathcal{D}_2,\delta_2)\cdot f(\beta,\kappa(P_\star(\mathcal{D}_2,\delta_2)),N_2,d,\alpha,n)$.

Output: With probability $1-\beta$ on the sampling of $\mathcal{D}_2$, the output is an upper bound on $\rho_{\text{quad}}(\mathcal{A})$.

The change of basis defined in Step 2.1 is motivated by the following observation:

Proposition 8. Let $P_\star(\mathcal{A},\delta)$ provide the smallest contraction rate of $\mathcal{A}$ among all δ-conditioned matrices, that is, $I \preceq P_\star(\mathcal{A},\delta) \preceq \delta I$ and $\rho(\mathcal{A},P_\star(\mathcal{A},\delta)) = \rho_{\text{quad}}(\mathcal{A},\delta)$. Define $B = P_\star(\mathcal{A},\delta)^{-1/2}$ and let $\mathcal{A}' = \{B^{-1}AB : A\in\mathcal{A}\}$. Let $\delta'\geq 1$ and let $P_\star(\mathcal{A}',\delta')$ provide the smallest contraction rate of $\mathcal{A}'$ among all δ'-conditioned matrices. It holds that $\rho(\mathcal{A}',P_\star(\mathcal{A}',\delta')) \leq \rho_{\text{quad}}(\mathcal{A},\delta)$.

Proposition 8 shows that in the appropriate coordinate system, a matrix with low conditioning $\delta\approx 1$ can do as well as a matrix with large conditioning $\delta\gg 1$ in the original coordinate system. Thus, given the appropriate coordinate system, we can set δ close to one, thereby ensuring that $\kappa(P_\star(\mathcal{D},\delta)) \leq \delta^{(n-1)/2}$ is close to one as well.

Unfortunately, we do not know the optimal coordinate system, because $\mathcal{A}$ is unknown so that we cannot compute $P_\star(\mathcal{A},\delta)$. The idea of the two-step approach is to approximate $P_\star(\mathcal{A},\delta)$ in the first step with $P_\star(\mathcal{D}_1,\delta_1)$, and then, use the coordinate system induced by $P_\star(\mathcal{D}_1,\delta_1)$ to sample in the second step. The expectation is that if $P_\star(\mathcal{A},\delta)$ is close to $P_\star(\mathcal{D}_1,\delta_1)$, then $\kappa(P_\star(\mathcal{D}_2,\delta_2))$ will be close to one.

6. TOWARD OPTIMAL DATA SET SIZES

In this section, we address the question: given a maximum total number of samples N (i.e., a *budget*), how to allocate the samples between the first and second step in order to obtain the best probabilistic upper bound? More precisely, how to choose N_1 and N_2, under the constraint that $N_1+N_2=N$, such that

$$\mathbb{E}_{\{(X_i,Y_i)\}_{i=1}^N}\ \gamma_\star(\mathcal{D}_2,\delta_2)\cdot f(\beta,\kappa(P_\star(\mathcal{D}_2,\delta_2)),N_2,d,\alpha,n)$$

is minimal?

Solving exactly the above optimization problem seems out of reach as the output will depend on the underlying *unknown* system. Therefore, in this section, we limit ourselves to first explaining the underlying trade-off for the allocation of the N samples among N_1 or N_2; and second, proposing a heuristic approach for the optimal trade-off.

6.1 Allocation Trade-Off

To minimize the output of the two-step approach, on the one hand, one hopes to have a good approximation of $P_\star(\mathcal{A},\delta_1)$ at the end of step one, because by Proposition 8 this implies that at the end of step two, $\gamma_\star(\mathcal{D}_2,\delta_2)$ can be close to $\rho_{\text{quad}}(\mathcal{A},\delta_1)$ even with $\kappa(P_\star(\mathcal{D}_2,\delta_2))$ close to one. For that, intuitively one needs N_1 as large as possible. But, on the other hand, one needs N_2 to be large enough too, so that the inflation factor $f(\beta,\kappa(P_\star(\mathcal{D}_2,\delta_2)),N_2,d,\alpha,n)$ is close to one. Thus, the optimal choice of N_1 results from a trade-off between taking N_1 large, but not too large so that $N_2=N-N_1$ is still large enough.

Fig. 2 illustrates the trade-off of the two-step approach on randomly generated switched linear systems in dimension 3 and with 3 matrices.

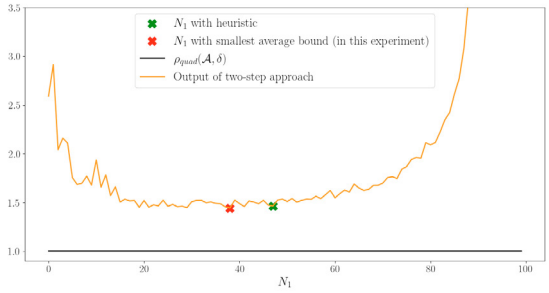

Fig. 2. Output of the two-step approach for different values of N_1, with $N_1+N_2=N:=100$, $n=3$, $|\mathcal{A}|=3$, $\delta_1=\delta_2=5$, $\beta=0.05$ and $\alpha=1/|\mathcal{A}|$. The results are the average output value over 10 randomly generated systems $\mathcal{A}$ for which $\rho_{\text{quad}}(\mathcal{A},\delta_1)=1$.

6.2 Heuristic for Choosing N_1

In this subsection, we propose a heuristic for choosing N_1 that aims at minimizing the output

$$\bar{o} := \gamma_\star(\mathcal{D}_2,\delta_2)\cdot f(\beta,\kappa(P_\star(\mathcal{D}_2,\delta_2)),N_2,d,\alpha,n)$$

of the two-step approach. Let us stress that in the following, we use the term "minimize" in a loose sense, because minimizing a random variable is not well defined, but we mean that we minimize the "magnitude" of the variable with respect to some metric (e.g., the expected value).

The first thing that we note is that [4]

$$\gamma_\star(\mathcal{D}_2,\delta_2) \leq \rho(\mathcal{A},P_\star(\mathcal{D}_1,\delta_1)).$$

Next, we remember from Theorem 7 that

$$\gamma_\star(\mathcal{D}_1,\delta_1) \leq \rho_{\text{quad}}(\mathcal{A},\delta_1).$$

Thus, by combining the above, it holds that

$$\gamma_\star(\mathcal{D}_2,\delta_2) \leq \underbrace{\frac{\rho(\mathcal{A},P_\star(\mathcal{D}_1,\delta_1))}{\gamma_\star(\mathcal{D}_1,\delta_1)}}_{:=\bar{r}_1}\rho_{\text{quad}}(\mathcal{A},\delta_1).$$

In other words, for all systems with same $\rho_{\text{quad}}(\mathcal{A},\delta_1)$, the value of $\gamma_\star(\mathcal{D}_2,\delta_2)$ is essentially bounded by $\bar{r}_1$. In view of this, our first approximation toward the heuristic is to minimize the following quantity which is an upper bound on $\bar{o}$:

$$\bar{o}' := \bar{r}_1\cdot f(\beta,\kappa(P_\star(\mathcal{D}_2,\delta_2)),N_2,d,\alpha,n).$$

Note that minimizing $\bar{o}'$ is still challenging because $\bar{r}_1$ is a random variable whose distribution is hard to analyze due to the many non-linearities involved in the computation of $\gamma_\star(\mathcal{D}_1,\delta_1)$. Luckily, in Sec. 4, we analyzed this quantity and we proposed a heuristic bound for its $(1-\beta)$-quantile: namely, with probability $1-\beta$ on the sampling of $\mathcal{D}_1$,

$$\bar{r}_1 \leq f(\beta,\delta_1^{n/4},N_1-d,1,\alpha,n).$$

In the following, we will use the value $\beta=0.5$ for this bound; in other words, we bound heuristically the median value of $\bar{r}_1$. Hence, our goal is to minimize

$$\bar{o}'' := f(0.5,\delta_1^{n/4},N_1-d,1,\alpha,n) \cdot f(\beta,\kappa(P_\star(\mathcal{D}_2,\delta_2)),N_2,d,\alpha,n).$$

The second factor in the above formula still contains an unknown quantity, namely $\kappa(P_\star(\mathcal{D}_2,\delta_2))$. To remove this

[4] The proof is similar to that of Proposition 8.

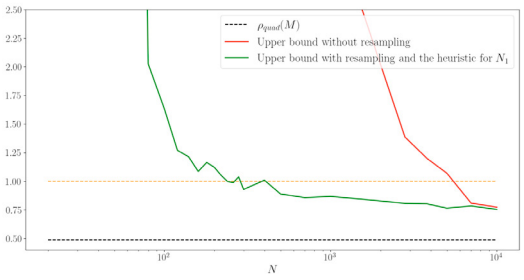

Fig. 3. Data-driven upper-bounds with confidence $1 - \beta = 95\%$ on the JSR of the consensus network with $\alpha = \frac{1}{3}$ in function of N the total number of samples required. We used $\delta_1 = 5$, $\delta_2 = 1$ for the two-step approach (the green curve) and $\delta = 5$ to compute $\gamma_\star(\mathcal{D}, \delta) \cdot \overline{f}$ (the red curve).

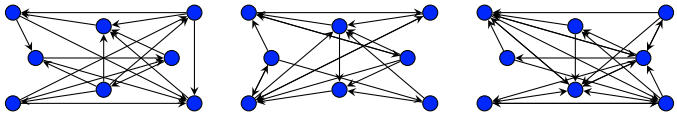

Fig. 4. Three (unknown) interaction networks for the consensus problem in Sec. 7.1.

non-determinism, we bound $\kappa(P_\star(\mathcal{D}_2, \delta_2))$ with its worst-case value, namely $\delta_2^{(n-1)/2}$. This leaves us with the final quantity to minimize with respect to N_1:

$$\hat{o}(N_1) \coloneqq f(0.5, \delta_1^{n/4}, N_1 - d, 1, \alpha, n) \cdot f(\beta, \delta_2^{(n-1)/2}, N - N_1, d, \alpha, n).$$

Summarizing, our heuristic's optimal value of N_1 is computed as $N_1^{\text{opt}} = \arg\min_{N_1 \in [N]} \hat{o}(N_1)$.

7. EXPERIMENTAL RESULTS

7.1 Consensus Network

We consider the problem of consensus in a hidden switched interaction network depicted in Fig. 4. The problem of deciding whether consensus is reached or not can be formulated as a problem of switched linear system stability analysis (Jadbabaie et al., 2003) and since the network is hidden, we do not have a model of the system. We will thus resort to data-driven methods to analyze whether the system reaches consensus or not.

Fig. 3 shows how many data points are needed to certify that consensus is reached with confidence $1 - \beta = 95\,\%$ by using the data-driven methods discussed in this paper. We clearly see the advantage of the adaptive sampling over existing methods without resampling.

REFERENCES

Ahmadi, A.A., Jungers, R.M., Parrilo, P.A., and Roozbehani, M. (2014). Joint spectral radius and path-complete graph Lyapunov functions. *SIAM Journal on Control and Optimization*, 52(1), 687–717. doi: 10.1137/110855272.

Berger, G.O., Jungers, R.M., and Wang, Z. (2021). Chance-constrained quasi-convex optimization with application to data-driven switched systems control. In *Learning for Dynamics and Control 2021*, volume 144 of *Proceedings of Machine Learning Research*, 571–583.

Calafiore, G.C. and Campi, M.C. (2006). The scenario approach to robust control design. *IEEE Transactions on Automatic Control*, 51(5), 742–753. doi: 10.1109/TAC.2006.875041.

Coulson, J., Lygeros, J., and Dörfler, F. (2019). Data-enabled predictive control: In the shallows of the DeePC. In *2019 18th European Control Conference (ECC)*, 307–312. IEEE. doi:10.23919/ECC.2019.8795639.

Duggirala, P.S., Mitra, S., and Viswanathan, M. (2013). Verification of annotated models from executions. In *2013 Proceedings of the International Conference on Embedded Software (EMSOFT)*, 1–10. IEEE. doi: 10.1109/EMSOFT.2013.6658604.

Jadbabaie, A., Lin, J., and Morse, A.S. (2003). Coordination of groups of mobile autonomous agents using nearest neighbor rules. *IEEE Transactions on automatic control*, 48(6), 988–1001. doi:10.1109/TAC.2003.812781.

Jungers, R.M. (2009). *The joint spectral radius: theory and applications*. Springer, Berlin. doi:10.1007/978-3-540-95980-9.

Kenanian, J., Balkan, A., Jungers, R.M., and Tabuada, P. (2019). Data driven stability analysis of black-box switched linear systems. *Automatica*, 109, 108533. doi: 10.1016/j.automatica.2019.108533.

Liberzon, D. (2003). *Switching in systems and control*. Birkhäuser, Boston, MA. doi:10.1007/978-1-4612-0017-8.

Margellos, K., Goulart, P., and Lygeros, J. (2014). On the road between robust optimization and the scenario approach for chance constrained optimization problems. *IEEE Transactions on Automatic Control*, 59(8), 2258–2263. doi:10.1109/TAC.2014.2303232.

Mitra, S. (2021). *Verifying cyber-physical systems: a path to safe autonomy*. MIT Press, Cambridge, MA.

Nejati, A., Lavaei, A., Jagtap, P., Soudjani, S., and Zamani, M. (2023). Formal verification of unknown discrete- and continuous-time systems: a data-driven approach. *IEEE Transactions on Automatic Control*, 68(5), 3011–3024. doi:10.1109/TAC.2023.3255141.

Rubbens, A., Wang, Z., and Jungers, R.M. (2021). Data-driven stability analysis of switched linear systems with sum of squares guarantees. *IFAC-PapersOnLine*, 54(5), 67–72. doi:10.1016/j.ifacol.2021.08.476.

Shalev-Shwartz, S. and Ben-David, S. (2014). *Understanding machine learning: from theory to algorithms*. Cambridge University Press, Cambridge, UK. doi: 10.1017/CBO9781107298019.

Sun, Z. and Ge, S.S. (2011). *Stability theory of switched dynamical systems*. Springer, London. doi:10.1007/978-0-85729-256-8.

Wang, Z., Berger, G.O., and Jungers, R.M. (2021). Data-driven feedback stabilization of switched linear systems with probabilistic stability guarantees. In *2021 60th IEEE Conference on Decision and Control (CDC)*, 4400–4405. IEEE. doi: 10.1109/CDC45484.2021.9683233.