

École polytechnique de Louvain

Data-driven stability analysis of black-box switched linear systems

Author: **Anne RUBBENS**

Supervisor: **Raphaël M. JUNGERS**

Readers: **Julien HENDRICKX, Anthony PAPAVASILIOU, Zheming WANG**

Academic year 2020–2021

Master [120] in Mathematical Engineering

Abstract

Nowadays, we encounter on a daily basis systems of tremendous complexity. For instance, we rely ever more on cyber-physical systems, including self-driving cars, drones or smart assistive technologies. Deeply understanding the properties of these applications is essential to optimize their potential. However, frequently, due to their inherent complexity, there is no available closed-form model of these systems. On the other hand, as the development of new technologies grows massively, so does the amount of information-rich data we have access to. Hence, it is crucial to develop new methods able to analyze complex systems without relying on model-based techniques, but rather on the unprecedented volume of data available.

In this thesis, we propose new data-driven approaches providing probabilistic stability guarantees for black-box linear switched systems, a first class of complex systems whose model is unknown. We construct approximate Lyapunov functions decreasing along a finite number of sampled trajectories by solving a quasi-convex optimization problem, and deduce the stability of the underlying system with a user-defined confidence level. We also explicitly characterize the number of observations required to attain this confidence level.

Recently, techniques relying on common quadratic Lyapunov functions (CQLFs) to provide probabilistic stability guarantees have been proposed. However, they require non trivial assumptions on the system analyzed, and suffer from conservativeness since many stable switched linear systems do in fact not admit a CQLF. Hence, in a first time, we overcome these issues and provide a broader class of probabilistic stability guarantees, by considering Sum of Squares (SOS) Lyapunov functions. Our contribution is twofold: first, we propose a novel approach for CQLFs, allowing to improve the existing guarantees and get rid of all assumptions. Then, we leverage our approach to the SOS setting, providing further improvement for the technique. In a second time, we propose an approach based on adaptive sampling to reduce the sample and computational complexity required to obtain stability guarantees holding with a given confidence level. We learn recursively, from small batches of trajectories, an optimal sampling strategy. This allows the data-driven approach to be tractable for high-dimensional systems, which is the case of the majority of real-world complex systems.

Acknowledgments

First of all, I would like to thank my supervisor, Raphaël Jungers, for his availability throughout the year, his intuition and bright ideas, his contagious interest in the topic and his constructive comments, which motivated me to improve and develop my work further.

I would also like to express my gratitude to Zheming Wang, for his endless patience, his precious advice and ideas and his motivation. Without its unfailing support and insightful comments, I could not have completed this thesis successfully.

I also take this opportunity to thank Anthony Papavasiliou and Julien Hendrickx for agreeing to spend time in reading my thesis.

Finally, I would like to express my gratitude to my friends and family for their presence during this year. In particular, I thank Jeroen for constantly understanding me and cheering me up; Kal for always being there; and Alex for his unconditional love and support.

Contents

List of Notation	ix
List of Figures	x
List of Tables	xviii
I Introduction and background	1
1 Introduction	3
1.1 Motivation	3
1.2 State of the art	5
1.3 Contributions	5
1.4 Structure of the work	7
2 Theoretical background	9
2.1 Switched linear systems and the Joint Spectral Radius	9
2.2 Approximation of the JSR using Lyapunov functions	11
2.2.1 Semidefinite programming	12
2.2.2 Quadratic forms	12
2.2.3 SOS forms	14
2.3 The scenario approach	18
II A broader and more precise class of probabilistic stability guarantees	23
3 Data-driven computation of the JSR	25
3.1 Problem statement	25
3.2 Presentation of an existing technique	28
3.3 Deterministic lower bounds on the JSR	30
3.4 Novel quadratic upper bounds	30

3.4.1	A new approach based on sensitivity analysis	30
3.4.2	New bounds for the quadratic case	33
3.5	Probabilistic SOS-based guarantees	40
4	Numerical experiments	45
4.1	Comparison of the quadratic and SOS-based bounds	46
4.2	Influence of the length of the trace	48
4.3	Influence of the dimension	50
4.4	Comparison of the bounds of Theorems 3.1 and 3.2	51
4.5	Lower bounds on the JSR	52
4.6	Conclusion of the experiments	53
III	Efficient computation using adaptive sampling	55
5	The adaptive sampling approach in the quadratic case	57
5.1	The fast algorithm	58
5.2	An enhanced approach based on adaptive sampling	60
5.2.1	Idea in the white-box setting	60
5.2.2	Two-phase adaptive sampling approach	62
5.2.3	p -phases adaptive sampling approach	65
5.3	Novel theoretical bounds based on adaptive sampling	67
5.4	A variant to the approach	71
5.5	Summary of the approaches based on adaptive sampling	73
6	Numerical experiments and derivation of a heuristic	75
6.1	Online and offline strategies	76
6.2	Comparison of offline and online sampling on Example 2.1	78
6.2.1	Performance of offline sampling	78
6.2.2	Performance of online sampling	80
6.2.3	A computational comparison of the two strategies	82
6.3	Derivation of a heuristic optimizing the adaptive sampling approach	83
6.3.1	For a large number of samples available: online sampling . .	83
6.3.2	For a low number of samples available: offline sampling . . .	86
6.3.3	Summary of the heuristic	87
6.4	Illustration of the performance of the heuristic	88
6.5	Conclusion of the experiments	91
7	Extension of the adaptive sampling approach to the SOS setting	93
7.1	Adaptive sampling in the SOS setting	94
7.2	Numerical illustration of the performance of SOS adaptive sampling	99

IV	Conclusion	103
8	Summary of the contributions and future prospects	105
8.1	Summary of the contributions	106
8.2	Future work and possible further improvements	108
	Bibliography	115
A	Polyhedral stability guarantees	117
A.1	Polyhedral Lyapunov functions	117
A.2	Polyhedral probabilistic upper bounds on the JSR	120
A.2.1	From approximate to actual polyhedral Lyapunov functions	120
A.2.2	Polyhedral upper bounds	121
B	Continuity argument for Lemma 3.3	124
C	Non-degeneracy of Problem (3.14)	126
D	Adaptation of Theorem 3.2 to the adaptive sampling setting	127
E	Derivation of theoretical bounds based on all samples in the adaptive sampling approach	129
F	Comparison of the bounds of Theorems 5.1 and 5.2	134
G	Additional analysis of the online sampling approach	138
H	Derivation of Table 6.2	140
I	Overview of the Julia codes implemented	141

List of Notation

Basics

$\mathbb{N}_0$	Positive integers.
$\mathbb{R}^n$	n -dimensional real space.
$\ x\ $, where $x \in \mathbb{R}^n$	Euclidean norm: $\ x\ := \sqrt{x^T x}$.
$X^{\mathbb{N}}$, where $X \subset \mathbb{R}^n$	All possible sequences $(x_n)_{n \in \mathbb{N}}$, $x_n \in X$.
$\mathbb{B} := \{x \in \mathbb{R}^n : \ x\ < 1\}$	Ball of unit radius centered at the origin in $\mathbb{R}^n$.
$\mathbb{S} := \{x \in \mathbb{R}^n : \ x\ = 1\}$	Sphere of unit radius centered at the origin in $\mathbb{R}^n$.
$\Pi_{\mathbb{S}}(x) := \frac{x}{\ x\ }$	Spherical projector on $\mathbb{S}$.
M	$M := \{1, 2, \dots, m\}$ for $m \in \mathbb{N}_0$.
$M^l := \otimes^l M$, for $l \in \mathbb{N}_0$	l -Cartesian product of M : $M^l = \{(i_1, \dots, i_l) : i_j \in M, j = 1, \dots, l\}$.
Z_l	$Z_l := \mathbb{S} \times M^l$.
$Z_l^N := \otimes^N Z_l$	N -Cartesian product of Z_l .

Measure-theoretic concepts

Our notation concerning these concepts is taken from [LT91].

σ^{n-1}	Uniform spherical measure on $\mathbb{S}$, satisfying $\sigma^{n-1}(\mathbb{S}) = 1$.
μ_M	Uniform measure on M .
$\mu_{M^l} = \otimes^l \mu_M$	Uniform product measure on M^l .
$\mu_l = \sigma^{n-1} \otimes \mu_{M^l}$	Uniform measure on Z_l .
$\mu_l^N = \otimes^N \mu_l$	Uniform product measure on Z_l^N .
$I(x; a, b)$	Regularized incomplete beta function: [MB73] $I : \mathbb{R}_{>0} \times \mathbb{R}_{>0} \times \mathbb{R}_{>0} \longrightarrow \mathbb{R}_{>0}$ $(x, a, b) \longrightarrow I(x; a, b) = \frac{\int_0^x t^{a-1} (1-t)^{b-1} dt}{\int_0^1 t^{a-1} (1-t)^{b-1} dt}.$
$I^{-1}(y; a, b)$	Inversed regularized incomplete beta function: $I^{-1}(y; a, b) = x > 0$ such that $I(x; a, b) = y$.

Matrices

$\mathbb{R}^{n \times n}$

$\mathcal{S}^n$

$\mathcal{S}_{++}^n$

$P \succ 0$, where P is a matrix

$P \succeq 0$, where P is a matrix

$E_P := \{x \in \mathbb{R}^n : x^T P x = 1\}$,

where $P \in \mathcal{S}_{++}^n$

$\|x\|_P := \sqrt{x^T P x}$,

where $x \in \mathbb{R}^n$ and $P \in \mathcal{S}_{++}^n$

$\langle x, y \rangle_P := x^T P y$,

where $x, y \in \mathbb{R}^n$ and $P \in \mathcal{S}_{++}^n$

d_{quad}

$\kappa(P)$, where $P \in \mathbb{R}^{n \times n}$

$\bar{\kappa}(P)$, where $P \in \mathbb{R}^{n \times n}$

I_n

$A \otimes B$, where $A \in \mathbb{R}^{m \times n}$ and $B \in \mathbb{R}^{p \times q}$

$x^{\otimes k}$ where $x \in \mathbb{R}^n$

$n \times n$ matrices.

$n \times n$ real symmetric matrices.

$n \times n$ positive definite matrices.

P positive definite.

P positive semidefinite.

Ellipsoid described by P .

Ellipsoidal norm defined by P .

Ellipsoidal inner product

defined by P .

$d_{quad} := \frac{n(n+1)}{2}$.

$\kappa(P) := \sqrt{\frac{\lambda_{\max}(P)}{\lambda_{\min}(P)}}$

$\bar{\kappa}(P) := \sqrt{\frac{\det(P)}{\lambda_{\min}(P)^n}}$

Identity matrix of size $n \times n$.

Kronecker product in $\mathbb{R}^{np \times mq}$:

$$A \otimes B = \begin{bmatrix} a_{11}B & \cdots & a_{1n}B \\ \vdots & \ddots & \vdots \\ a_{m1}B & \cdots & a_{mn}B \end{bmatrix}$$

k -th Kronecker power of x :

$x^{\otimes k} = x \otimes x^{\otimes(k-1)}$ and $x^{\otimes 1} = x$.

Acronyms

JSR	Joint Spectral Radius
SOS	Sum Of Squares
CQLF	Common Quadratic Lyapunov Function
SDP	Semidefinite Programming
UAS	Uniformly Asymptotically Stable

List of Figures

1.1	Example of a switched system.	4
1.2	Dependencies between the different parts of this thesis.	7
2.1	Illustration of an unstable switched linear system composed of two stable modes. The two left-hand side graphs present the trajectories of the two stable systems associated to each mode, while the right-hand side graph presents an unstable trajectory of the switched linear system.	10
2.2	The two unknown interaction schemes for the consensus problem of 9 nodes.	11
2.3	Representation of the d -lift of $\mathbb{S}$, for $n = 2$ and $d = 2$. This d -lift, given by $\{x \in \mathbb{S}^D : x = y^{[d]} \text{ for } y \in \mathbb{S}\}$, is represented in red on the figure. It only forms a subset of $\mathbb{S}^D$	17
3.1	Relations between the different optimization problems introduced. The left-hand side of the figure is related to the quadratic setting, while the right-hand side presents the extension to the SOS setting. We infer an upper bound on the costs of initial problems, γ^o and γ_{sos} , as a function of the solutions to the respective sampled problems. The sampled problems are relaxations of the initial problems. They only require the knowledge of N trajectories of System (2.1) to be solved. The upper bounds, which approximates the JSR, hold with a certain confidence level β depending on N	27
3.2	Illustration of the dynamics of the unstable system of Example 3.1 on the level set of an approximate Lyapunov function. The trajectories of the system starting on the level set diverge to the infinity, even though the set of points violating the Lyapunov property, represented in red, is arbitrarily small.	28

3.3	Illustration of the existing technique to obtain probabilistic stability guarantees. The largest ellipsoid is $E_{P^*(\omega_N)}$, computed by solving Problem (3.2). The red portions of $E_{P^*(\omega_N)}$ represent the worst-case distribution of the points that potentially violate the Lyapunov property. The inscribed ellipsoid is the largest scaling of $E_{P^*(\omega_N)}$ included in the convex hull of the subset of $E_{P^*(\omega_N)}$ which does not belong to the violating set. All points on this ellipsoid satisfy a relaxed Lyapunov property.	29
3.4	Illustration of a spherical cap of measure ϵ in two dimensions. . . .	32
3.5	Illustration of the argument leading to Proposition 3.1. For each trajectory $(x', \mathbf{j}) \in \omega_s$, where $\gamma^*(\omega_s) = \gamma^o$, there exists a trajectory $(x, \mathbf{j}) \in \omega_N$ such that x belongs to $C(x', \epsilon)$. This holds with probability $\beta(\epsilon, m, N)$	33
4.1	Evolution of several upper bounds on the JSR of the system of Example 4.1 with the number of samples. The bounds of Theorem 3.1 (our new quadratic bound), of Theorem 3.3 (our SOS-based bound, for $d = 2$), of [BJW21] and the true value of the JSR are plotted. From 1500 samples, the SOS-based bound becomes the tightest one.	47
4.2	Evolution of several upper bounds on the JSR of the system of Example 2.1 with the number of samples, for traces of length 2. The bounds of Theorem 3.1, of Theorem 3.3, of [BJW21] and the true value of the JSR are plotted. As expected for systems of large dimension, our bounds outperform the one of [BJW21].	48
4.3	Illustration of the key factor explaining Figure 4.2: the evolution of $\kappa(P^*(\omega_N))$ (appearing in our quadratic bound), $\kappa(P_{sos}^*(\omega_N))$ (appearing in our SOS-based bound) and $\bar{\kappa}(P^*(\omega_N))$ (appearing in the bound of [BJW21]) with the number of samples. $\bar{\kappa}(P^*(\omega_N))$ is significantly larger than the other quantities.	48
4.4	Evolution of the quadratic upper bounds on the JSR of the system of Example 4.2 with the number of samples, for traces of various length. To the contrary of the bound of [BJW21], our bound gets tighter when l increases.	49
4.5	Evolution of the SOS-based upper bound, for $d = 2$, on the JSR of the system of Example 4.2 with the number of samples. The length of the traces ranges between 1 and 5. The bound gets tighter when l increases.	49

4.6	Convergence of the quadratic upper bounds on the JSR of 5-modal systems of various dimensions when the number of samples increases. The ratios between the bounds of Theorem 3.1 (our new quadratic bound) or of [BJW21] and the true value of the JSR are plotted. The ratios increase with n	50
4.7	Convergence of the quadratic upper bounds on the JSR of 3-dimensional systems composed of 5 or 7 modes when the number of samples increases. The ratios between the bounds of Theorem 3.1 or of [BJW21] and the true value of the JSR are plotted. The ratios slightly increase with m	50
4.8	Convergence of the SOS-based upper bounds on the JSR of systems of various dimensions when the number of samples increases. The ratio between the bound of Theorem 3.3 and the true value of the JSR is plotted. The ratio increases with n , while the impact of m is negligible.	51
4.9	Evolution of $\kappa(P_{sos}^*(\omega_N))$ for systems of various dimensions with the number of samples. $\kappa(P_{sos}^*(\omega_N))$ increases with n , while its sensitivity with respect to a variation in m is negligible. The behavior of $\kappa(P_{sos}^*(\omega_N))$ explains Figure 4.8.	51
4.10	Convergence of our quadratic upper bounds on the JSR of systems of various dimensions when the number of samples increases. The ratio between the bounds of Theorem 3.1 or 3.2 and the true value of the JSR is plotted. The bound of Theorem 3.2 is the tightest one when the number of samples used is low.	52
4.11	Evolution of the lower bounds on the JSR of the system of Example 4.2 with the number of samples. The bounds of Lemma 3.1, in the quadratic case, and of Lemma 3.2, in the SOS case for $d = 2$ and $d = 4$, are plotted. The bounds get tighter when d increases.	52
5.1	Change of coordinates L mapping the elements of an ellipsoid E_P to the unit sphere. L is the Cholesky decomposition of P : $P = L^T L$	61

5.2	Relations between the different optimization problems designed for the two-phase adaptive sampling approach, and illustration of the factors impacting the distance $\ P_2^*(\omega_{N_1+N_2}) - I_n\ $. The left-hand side of the figure presents the white-box problems, before and after the change of variables is performed. The right-hand side presents the associated sampled problems. We only have access to the solutions to these problems, approximating the solutions of the initial problems. The closeness of $P_2^*(\omega_{N_1+N_2})$ to the identity depends on $\Delta_{s,1}$ and $\Delta_{s,2}$. $\Delta_{s,1}$ affects the quality of the change of variables by inducing a distance between the solution of $\text{Opt}(L_1)$ and the identity. It decreases when N_1 increases, while $\Delta_{s,2}$ decreases when N_2 increases. A trade-off has thus to be achieved between both quantities.	64
5.3	Summary of the procedure to obtain a probabilistic bound on the JSR using our adaptive sampling approach. The blue box contains the iterations $k = 2, \dots, p - 1$, that is the phases refining the change of variables. After each phase, the change of variables computed is handed to the next phase, as illustrated in red. The external resources provided to the phases, that is the sampling $\omega_{N_1}, \dots, \omega_{N_p}$ are presented in black. In the last phase, one either fix P_p to the identity (right-hand side) or not (left-hand side). This is determined before starting the computations.	73
6.1	Evolution of $\kappa(P_p^*(\omega_{N_1+\dots+N_p}))$ with the ratio of samples used in the first phase, for the system of Example 2.1. In total, 500 trajectories are sampled. $\kappa(P_p^*(\omega_{N_1+\dots+N_p}))$ presents a U-shape and decreases when p increases.	79
6.2	Evolution of the upper bounds of Theorem 5.1 with the ratio of samples used in the first phase, for the system of Example 2.1. For $p \geq 2$, the bounds always outperform the one of Theorem 3.1 (no adaptive sampling).	80
6.3	Evolution of the upper bounds of Theorem 5.2 (for P_p fixed to the identity) with the ratio of samples used in the first phase, for the system of Example 2.1. The bounds always outperform the one of Theorem 3.1.	80
6.4	Evolution of several upper bounds with the ratio of samples used in the first phase, for the system of Example 2.1. The bounds of Theorems 5.1 and 5.2 (two-phase adaptive sampling), of Proposition 5.1 (fast algorithm), of Theorem 3.1 (no adaptive sampling), and the true value of the JSR are plotted.	80

6.5	Evolution of several upper bounds with the total number of samples, for the system of Example 2.1. The bounds of Theorems 5.1 and 5.2, the true value of the JSR of the system and the bound of Theorem 3.1 are plotted. The bound of Theorem 5.2 is the tightest one. The bounds related to adaptive sampling beat the bound of Theorem 3.1 for a sufficiently high number of samples.	81
6.6	Evolution of the number of samples needed for the convergence of $P_1(\omega_{N_1})$ as defined by Procedure Online with the dimension of the system. The blue curve presents the average number of samples required and the blue ribbon is the variance of the experiments. A linear interpolation whose slope is equal to 32 fits the data well. . .	84
6.7	Evolution of the number of phases reached by Algorithm 4 with the total number of samples, for systems of dimension ranging between 3 and 10. In each phase, the number of samples required is determined by Procedure Online. The dots present the average of the observed data, and the curves are interpolations of degree 2.	84
6.8	Evolution of $\kappa(P_k^*(\omega_{N_1+\dots+N_k}))$ and $\kappa(P'^o(\mathbf{L}_{\mathbf{k}-1}))$ with the number of phases, for a 4-dimensional. $\kappa(P_k^*(\omega_{N_1+\dots+N_k}))$ decreases at each phase, and reaches a limit from Phase 4.	85
6.9	Evolution of the upper bounds of Theorem 5.2 (for P_p fixed to the identity) with the ratio of samples used in the first phase, for the system of Example 4.2. In total, 80 trajectories are sampled. The number of phases ranges from 2 to 4. When less than $d_{quad} + 1$ samples are used in the last phase, the tightness of the bounds decreases.	86
6.10	Evolution of several upper bounds with the dimension of the system. The bounds based on adaptive sampling are yielded by the application of Heuristic 1 on 4-modals systems of various dimensions, for a total number of samples fixed to 1500. The bounds of Theorems 5.1 and 5.2 (for P_p set to the identity), of Theorem 3.1 (no adaptive sampling) and the true value of the JSR are plotted. The bound of Theorem 5.2 always outperforms all bounds. When the dimension increases, the number of samples available is not sufficient to allow the bound of Theorem 5.1 to be tighter than the one of Theorem 3.1.	88

6.11	Evolution of several upper bounds with the dimension of the system. The bounds based on adaptive sampling are yielded by the application of Heuristic 1 on 4-modals systems of various dimensions n , for a total number of samples fixed to $100n$. The bounds of Theorems 5.1 and 5.2, of Theorem 3.1 and the true value of the JSR are plotted. The bound of Theorem 5.2 is the tightest. For systems of dimension larger than 25, the number of samples available is not sufficient to allow the bound of Theorem 5.1 to be tighter than the one of Theorem 3.1.	89
6.12	Evolution of several upper bounds on the JSR of a 45-dimensional system with the total number of samples. The bounds based on adaptive sampling are yielded by the application of Heuristic 1. The bounds of Theorems 5.1 and 5.2, of Theorem 3.1 and the true value of the JSR are plotted. The bound of Theorem 5.2 is the tightest. From 5000 samples, all bounds based on adaptive sampling outperform the one of Theorem 3.1.	90
7.1	Relations between the different optimization problems designed for the two-phase adaptive sampling approach in the SOS setting. The factors impacting the distance $\ P_{sos,2}^*(\omega_{N_1+N_2}) - I_D\ $ are emphasized. The left-hand side of the figure presents the white-box problems, before and after the change of variables is performed. The right-hand side presents the associated sampled problems. We only have access to the solutions to these problems, approximating the solutions of the initial problems. The closeness of $P_{sos,2}^*(\omega_{N_1+N_2})$ to the identity depends on $\Delta_{s,1}$, $\Delta_{s,2}$ and Δ_{lift} . The latter appears because the exact optimal change of variables does not exist in general in the SOS setting, to the contrary to the quadratic setting. $\Delta_{s,1}$ and Δ_{lift} affect the quality of the change of variables by inducing a distance between the solution of $\text{Opt}_{sos}(\tilde{L}_1)$ and the identity. $\Delta_{s,1}$ decreases when N_1 increases, while $\Delta_{s,2}$ decreases when N_2 increases.	96
7.2	Evolution of several upper bounds on the JSR of the system of Example 7.1 with the total number of samples used. The SOS-based bounds of Theorems 3.3 and 7.1, the quadratic bounds of Theorems 3.1 and 5.1 and the true value of the JSR are plotted. The bound of Theorem 7.1 outperforms the SOS-based bound obtained without resorting to adaptive sampling from 400 samples. It outperforms all bounds from 3000 samples.	100

8.1	Summary of the different contributions of this thesis. We improved the existing probabilistic bound on the JSR of switched linear systems in several ways. First, we provided a novel class of bounds, based on the computation of approximate SOS Lyapunov functions. Additionally, this allowed us to derive two new quadratic bounds. In a second time, we proposed a technique to reduce the sample and computational complexity required to obtain tight bounds, and hence made the problem tractable for high dimensions. Finally, we combined our two contributions to improve our SOS-based bounds.	106
D.1	Evolution of upper bounds based on adaptive sampling with the dimension of the system. The bounds are yielded by the application of Heuristic 1 on 2-modals systems of various dimension, for a total number of samples of 1500. The bounds of Theorems 5.1 and D.1, as well as the true value of the JSR are plotted. When the dimension is lower than 10, the bound of Theorem D.1 is the tightest.	128
E.1	Illustration of a case, in two dimensions, where $\tilde{\mathbb{S}} \cap \Pi_{\mathbb{S}}(L_1(\tilde{\mathbb{S}})) = \tilde{\mathbb{S}}$. $\tilde{\mathbb{S}}$ is presented in red and $L_1(\tilde{\mathbb{S}})$ is presented in pink. Clearly, the projection of $L_1(\tilde{\mathbb{S}})$ on the sphere covers $\tilde{\mathbb{S}}$	132
G.1	Evolution of $\kappa(P_k^*(\omega_{N_1+\dots+N_k}))$ and $\kappa(P'^o(\mathbf{L}_{k-1}))$ with k , when 1100 samples are used, for the system of Example 2.1. $\kappa(P_k^*(\omega_{N_1+\dots+N_k}))$ decreases at each phase, and reaches a limit from Phase 6. For an equivalent number of samples used, it is much lower than $\kappa(P^*(\omega_N))$.	138
G.2	Evolution of the number of phases reached by Algorithm 4 with the total number of samples, for the system of Example 2.1. In each phase, the number of samples required is determined by Procedure Online.	138
G.3	Evolution of $P_1(\omega_{N_1})$ with the number of samples used in the first phase during Procedure Online, for the system of Example 2.1. We plot $\frac{\ P_1(\omega_{N_1}) - P_1(\omega_{N_1-n})\ }{\ P^o\ }$, which determines the number of samples used in Phase 1, and $\frac{\ P_1(\omega_{N_1}) - P^o\ }{\ P^o\ }$. The latter decreases more slowly, implying that the convergence of $P_1(\omega_{N_1})$ to P^o is not ensured by Procedure Online.	139
G.4	Evolution of $P_2(\omega_{N_1+N_2})$ with the number of samples used in the second phase during Procedure Online, for the system of Example 2.1. We plot $\frac{\ P_2(\omega_{N_1+N_2}) - P_2(\omega_{N_1+(N_2-n)})\ }{\ P'^o(L_1)\ }$ and $\frac{\ P_2(\omega_{N_1+N_2}) - P'^o(L_1)\ }{\ P'^o(L_1)\ }$. Both quantities are smaller than in Phase 1, and $\frac{\ P_2(\omega_{N_1+N_2}) - P'^o(L_1)\ }{\ P'^o(L_1)\ }$ is larger than $\frac{\ P_2(\omega_{N_1+N_2}) - P_2(\omega_{N_1+(N_2-n)})\ }{\ P'^o(L_1)\ }$	139

List of Tables

4.1	Default values of the parameters used in the experiments of Chapter 4.	46
6.1	Default values of the parameters used in Procedure Online and Algorithm 4 for a n -dimensional system.	78
6.2	Sample and computational complexities of the different strategies, required to obtain a bound smaller than 250% of the JSR of the system of Example 2.1. $[0, U]$ is the interval over which we look for γ (or γ_k , for $k = 1, 2, \dots, p$) by bisection with an accuracy α . The expression $\text{poly}(x, y)$ denotes that the computational complexity is polynomial in x and y .	82
6.3	Coefficients of the linear regression $N_k = a * n + b$, where n is the dimension of the system and N_k is the number of samples required to reach convergence in phase k , as determined by Procedure Online.	84
6.4	Maximum number of phases to use, depending on the dimension. This is the optimal number of phases when the number of samples available allows it.	85
I.1	Generation of random systems and trajectories	142
I.2	White-box computations	142
I.3	Functions required by the use of SOS Lyapunov functions	142
I.4	Implementation of the approach of Chapter 3	142
I.5	Implementation of the experiments of Chapter 4	142
I.6	Implementations of the algorithms of Chapter 5, 6 and 7	143
I.7	Implementation of the experiments of Chapters 6 and 7	143

Part I

Introduction and background

Chapter 1

Introduction

1.1 Motivation

Cyber-physical systems are systems composed of networked computers in close and constant interaction with our physical world. Recently, they have been a subject of growing interest [Lee15], [TCRM14]. Indeed, we encounter such systems on a daily basis, in applications ranging from transportation to agriculture and healthcare, which explains the incentive to efficiently design, control and analyze them in spite of their inherent complexity. We can think for instance of self-driving cars; complex robotic systems including smart assistive technologies or drones; or smart grids applications such as the optimal energy management of smart buildings. A large number of cyber-physical systems can be modeled as switched systems. These are systems consisting of several models, describing several laws of dynamics, among which the system can switch at certain points in time [Lib03] [SWM⁺07]. We illustrate the particular example of a light switch in Figure 1.1 . Switched systems whose underlying models are linear form an important family: *switched linear systems*. They can for instance be used to model the design of drug treatment for viral diseases like HIV, mechanical systems such as driving sticks, multi-controller switching schemes, or coordination of autonomous mobile agents [HVCMB11] [JLM03].

A central and challenging questions on switched linear systems is their stability, which is an NP-hard problem [BT99], and many tools have therefore been designed in the past decades for their analysis. Most of these tools are model-based, and build on notions from Lyapunov and invariant set theory [LA09], [BM07], [Lib03]. However, in many practical applications, a closed-form model may not be available directly. Typically, cyber-physical systems, which are composed of many subcomponents of different nature interacting in a complex way with each other, are difficult to model. For instance, consider the optimization of the energy management of

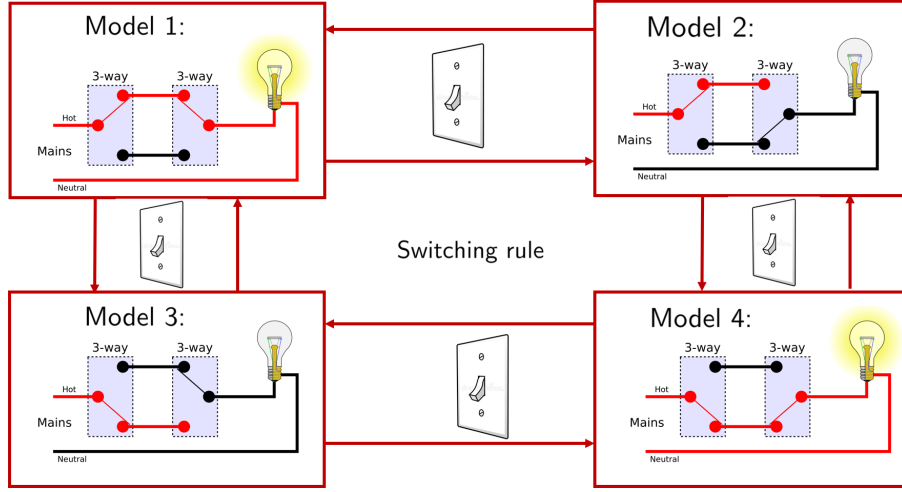


Figure 1.1: Example of a switched system.

smart buildings. The evolution over time of the system is stochastic and a closed-form model for the thermal behavior of these buildings is difficult to obtain [ZHT17]. Hence, stability analysis based on data obtained via simulations or observations of the unknown system solely has received increasing attention lately [KQHT16], [BFGP17], [KBJT19], [SDGDI⁺20], [WJ21], [BJW21]. A first approach consists in recovering the model from the available trajectories, and then using model-based theory for stability analysis of switched linear systems. However, this two-steps procedure presents several drawbacks for switched systems. First, identification is NP-hard, can induce approximation errors, and requires a sufficiently rich set of observations [Lau16]. Second, for various applications, the model may be too complex to be analyzed using classical control techniques. Therefore, in this work, we bypass the explicit identification phase and seek to obtain probabilistic stability guarantees based directly on the simulations, without recovering the underlying model.

The guideline of this thesis is the following question: **After observing a finite number of sampled trajectories of a given length, what formal probabilistic guarantees about the stability of the underlying system can we obtain?**

Of course, there will be a trade-off between their accuracy and the desired level of confidence. We refer to this framework as *data-driven analysis* of *black-box* dynamical systems.

1.2 State of the art

Stability of switched linear systems is characterized by their *Joint Spectral Radius* (JSR), which measures the maximal asymptotic growth rate of a state under the dynamics of the system [Jun09]. Model-based approximations of the JSR have thus been deeply investigated. In particular, one can obtain bounds on the JSR by constructing Lyapunov functions decreasing along all trajectories of the system. This amounts to solving a particular quasi-linear optimization problem, that is a problem of the form $\min_{u \in \mathbb{R}^n, \lambda > 0} \lambda$ s.t. $f_i(u) \leq \lambda$, for $i = 1, \dots, m$ where $f_i(u)$ are linear in u [Epp05] [JAPR17]. For instance, one can construct *Common Quadratic Lyapunov functions* (CQLFs) by solving the following problem:

$$\min_{P \succeq 0, \gamma \geq 0} \gamma \quad \text{s.t.} \quad (Ax)^T P Ax \leq \gamma^2 x^T P x, \quad \forall A \in \mathcal{M}, \forall x \in \mathbb{R}^n,$$

where $\mathcal{M}$ is the set of the matrices associated to each model the system can switch among. Then, a bound on the JSR of the underlying system is provided by the optimal cost.

Recently, this work has been extended to the black-box setting, in which the matrices are unknown and one only has access to a limited number of trajectories (x, Ax) [BJW21] [KBJT19]. The method consists in two steps. First, one constructs approximate CQLFs decreasing along the sampled trajectories, by solving a relaxation of the initial problem:

$$\min_{P \succeq 0, \gamma \geq 0} \gamma \quad \text{s.t.} \quad (Ax)^T P Ax \leq \gamma^2 x^T P x, \quad \forall (x, Ax) \text{ observed.}$$

Then, from the decrease of the approximate CQLF on a finite set of trajectories, it remains to infer a probabilistic upper bound on the JSR of the underlying system, holding with a user-defined confidence. To achieve this step, both [BJW21] and [KBJT19] use tools from scenario optimization theory [CGR18], [Cal10] and geometric properties of switched systems.

The approach developed in [BJW21] and [KBJT19] presents two main weaknesses. First, it relies on the computation of CQLFs, which is conservative since some stable switched systems do not admit a CQLF [LA09]. Second, it invokes classical chance-constrained theorems underlying the scenario approach, which require non-trivial assumptions concerning the set $\mathcal{M}$ to be satisfied. The probabilistic bounds proposed are therefore not valid for all switched linear systems. In this thesis, we answer these two weaknesses and improve the approach in several ways.

1.3 Contributions

In this work, we provide both theoretical results, that is new probabilistic stability guarantees, and practical algorithms to obtain these guarantees. We implement

these algorithms in the programming language Julia [BKSE12] and provide numerical experiments illustrating the performance of our guarantees in comparison with existing techniques. An overview of our code is given in Appendix I. Our work comprises two main contributions.

- First, we derive a **broader and more precise class of probabilistic stability guarantees for black-box switched linear systems** than the ones obtained by relying on CQLFs. To this aim, we consider *Sum of Squares (SOS) Lyapunov functions* of arbitrarily high degree, which yield tighter results and exist in a significantly larger number of cases [PJ08], [Jun09, Section 2.3.7]. We quantify explicitly the trade-off between the number of observations and the tightness of the guarantee.
To get rid of the assumptions required when resorting to scenario optimization theory, and because the geometric tools used in [KBJT19] and [BJW21] do not generalize in a straightforward way to SOS Lyapunov functions of higher degrees, we develop a novel approach, based on *sensitivity analysis* [Ber16] of the quasi-linear optimization problem. Additionally, we thereby also obtain novel upper bounds in the quadratic case, that can be compared with the one obtained in [BJW21], and that hold for all switched linear systems. Then, we extend our approach to the SOS setting. In the white-box setting, approximations of the JSR using SOS Lyapunov functions are tighter [PJ08], and we will observe on several numerical experiments that this generalizes to the black-box setting.
- The main drawback of data-driven stability analysis consists in the huge number of trajectories that have to be sampled to obtain a sufficient confidence level on the derived guarantee. Hence, in a second time, we **reduce the number of samples required to obtain a stability guarantee holding with a given confidence level**. To the contrary of existing techniques [CGC14], we exploit the particular structure of our system and propose an approach based on *adaptive sampling* [DA20], [Set10]. We derive this approach both in the quadratic and in the SOS setting. Our method consists in sampling a small batch of trajectories uniformly, and in learning an optimal change of variables corresponding to an optimal new sampling from the observation of these trajectories. This can be refined by recursively recomputing the change of variables. We propose a heuristic providing the optimal number of trajectories to be sampled in each phase as well as the optimal number of phases to optimize the probabilistic stability guarantees. Then, we compare these optimal guarantees with the one obtained without resorting to adaptive sampling for the same total number samples used. Several numerical experiments illustrate the performance of our approach and how it outperforms the existing bounds.

Publication

The first contribution has been the object of the following publication:

[RWJ21] Anne Rubbens, Zheming Wang, and Raphaël M. Jungers. Data-driven stability analysis of switched linear systems with Sum of Squares guarantees. To be published in *Proceedings of the 7th IFAC Conference on Analysis and Design of Hybrid Systems*. *arXiv preprint: 2101.00112*. 2021.

1.4 Structure of the work

Our work is organized in four parts and eight chapters, including the present introduction and the conclusion, as illustrated in Figure 1.2

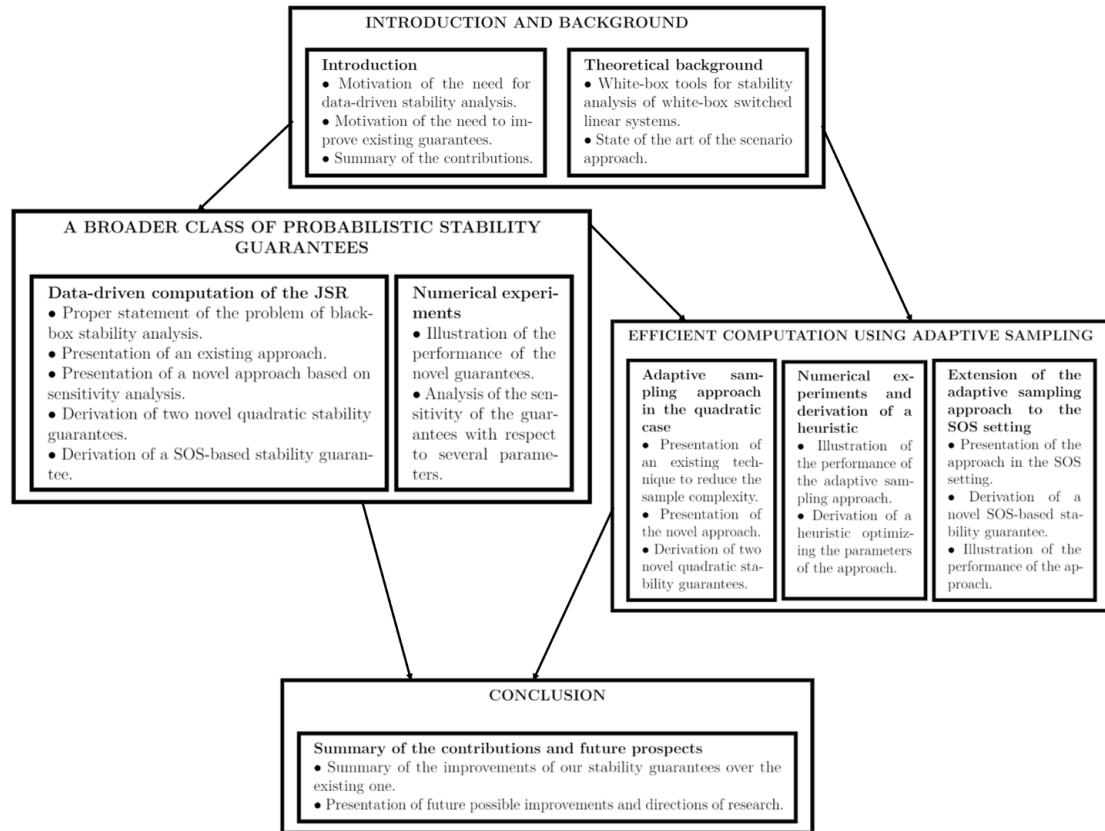


Figure 1.2: Dependencies between the different parts of this thesis.

The first part comprises the introduction and the background. Chapter 2 introduces the main concepts required to fully understand this thesis. It includes the state

of the art of stability of switched linear systems in the white box setting and a presentation of the scenario approach.

The second part contains the formal presentation of our problem and our first contribution. Chapter 3 first presents the problem studied, that is stability analysis of black-box systems, and briefly reviews the existing technique of [KBJT19] and [BJW21]. It also contains the derivation of our new approach relying on sensitivity analysis. It presents our novel probabilistic guarantees in the quadratic case, and the extension of the technique to the SOS setting. We illustrate the performance of our technique in Chapter 4 with numerical experiments, and explore the sensitivity of our bounds with respect to several parameters.

The third part contains our second main contribution, that is the adaptive sampling approach to the problem. In Chapter 5, we formulate our technique in the quadratic setting and derive novel theoretical bounds. In Chapter 6, we then propose a heuristic based on empirical observations, and illustrate the performance of this heuristic with several numerical experiments. Chapter 7 presents the leveraging of the adaptive sampling approach to the SOS setting.

Finally, we conclude our work and present future prospects in Chapter 8.

Chapter 2

Theoretical background

In this chapter, we provide the background required to comprehend our contributions, by presenting a brief review of the involved concepts and proposing references for further interest.

First, we formally present switched linear systems and the main tool used for their stability analysis, that is their *joint spectral radius* (JSR) [Jun09]. Then, we describe two classes of Lyapunov functions: common quadratic Lyapunov functions (CQLFs) and Sum of Squares (SOS) Lyapunov functions. We emphasize their role in the stability analysis of switched systems [AJ15].

Finally, we present the state of the art in scenario-based optimization [CG18] [NY19].

2.1 Switched linear systems and the Joint Spectral Radius

Switched linear systems form a particular family of switched systems, that is systems whose dynamics are governed by several models and a switching sequence selecting the model applying at a given time. They are characterized by a finite set of *modes*: $\mathcal{M} = \{A_i, i \in M\} \subset \mathbb{R}^{n \times n}$ where $M = \{1, 2, \dots, m\}$, which describe the different linear continuous state models; together with a discrete rule deciding the *switching* among them: $\tau \in M^{\mathbb{N}}$. Formally, we consider systems of the form

$$x_{k+1} = A_{\tau(k)}x_k \quad (2.1)$$

where $x_k \in \mathbb{R}^n$ is the state and $k \in \mathbb{N}$ is the time index.

Switched linear systems are *uniformly asymptotically stable* (UAS) if they possess the following property:

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

For linear systems (consisting of one mode only), UAS is characterized by the spectral radius of the matrix A : the system is stable if and only if $\rho(A) = \lim_{k \rightarrow \infty} \|A^k\|^{\frac{1}{k}} = \max\{|\lambda_1(A)|, \dots, |\lambda_n(A)|\} < 1$. Note that the definition is independent of the chosen norm.

This characterization is extended to switched linear systems, by defining the JSR of the set $\mathcal{M}$:

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

which measures the maximum asymptotic growth rate of $\|x_k\|$ under the dynamics (2.1), achieved by taking an arbitrary sequence of modes. The following property relates the JSR and the stability of switched systems:

Property 2.1. [Jun09, Cor. 1.1]

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

The JSR is invariant under similarity transformations and scales with multiplications of the matrices, which allows to conduct stability analysis of several systems equivalently:

Property 2.2. [Jun09, Props. 1.3 and 2.5]

For any bounded set of matrices $\mathcal{M}$, any invertible matrix T , and any $l \in \mathbb{N}_0$, $\rho(\mathcal{M}) = \rho(T\mathcal{M}T^{-1})$ and $\rho(\mathcal{M}^l) = \rho(\mathcal{M})^l$, where $\mathcal{M}^l := \{\prod_{j=1}^l A_{i_j} : A_{i_j} \in \mathcal{M}\}$.

Because switched linear systems do not inherit the properties of their modes, as illustrated on Figure 2.1, and since the number of possible sequences of matrices is exponentially high, computing their JSR is a NP-hard problem.

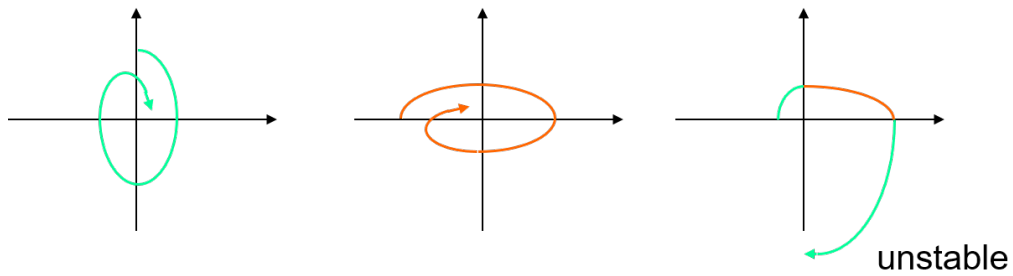


Figure 2.1: Illustration of an unstable switched linear system composed of two stable modes. The two left-hand side graphs present the trajectories of the two stable systems associated to each mode, while the right-hand side graph presents an unstable trajectory of the switched linear system.

We present a running example, adapted from [JLM03], that will serve throughout the rest of this thesis:

Example 2.1. Suppose you can observe the states of n nodes composing a switching and hidden network, that is a network of nodes interacting with each other in several unknown ways, as illustrated on Figure 2.2. At each time step, the states of the nodes are updated by taking into account the influence of the nodes they are currently interacting with, through the rule

$$x_{t+1} = A_{\tau(t)}x_t, \quad \tau(t) \in \{1, \dots, m\},$$

where A_i represent the interaction matrices. These matrices are such that $A_i \mathbf{1} = \mathbf{1}$, where $\mathbf{1}$ is the all-one vector in $\mathbb{R}^n$ for $i = 1, \dots, m$. We seek to verify that the

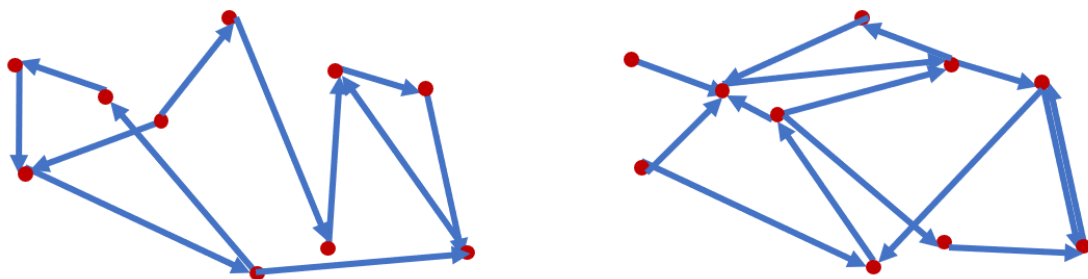


Figure 2.2: The two unknown interaction schemes for the consensus problem of 9 nodes.

nodes eventually reach as consensus, so that $\lim_{t \rightarrow \infty} x(t) = c\mathbf{1}$ for some $c \in \mathbb{R}$. This amounts to verifying whether or not the JSR of the system $\{BA_iB^T, i = 1, \dots, m\}$ is smaller than one, for a fixed matrix $B \in \mathbb{R}^{(n-1) \times n}$ such that $BB^T = I_{n-1}$. In this thesis, we will analyze the consensus of a network of 9 nodes, interacting with each other in 2 different and unknown ways.

2.2 Approximation of the JSR using Lyapunov functions

Over the past decades, many tools have been developed to provide approximations of the JSR of a switched linear system and hence to analyze the stability of the underlying system. Lyapunov functions provide one of the most commonly used tool. They are positive functions $f : \mathbb{R}^n \rightarrow \mathbb{R}_{>0}$ decreasing everywhere in the state space: $f(x_{k+1}) \leq f(x_k)$, $\forall x_k$. Common Lyapunov functions are Lyapunov functions shared

by all modes of the system, that is satisfying $f(A_i x) \leq f(x)$, $\forall A_i \in \mathcal{M}$, $\forall x \in \mathbb{R}^n$. The existence of a common Lyapunov function is a sufficient condition for the system to be UAS, implying that they can serve as stability certificates [Lib03]. Several types of Lyapunov functions exist, providing more or less conservative approximations of the JSR at a more or less expensive computational cost [AJ15].

2.2.1 Semidefinite programming

Computation of Lyapunov functions require the resolution of specific optimization programs. Several of these programs fall under the field of a particular class of convex optimization: *semidefinite programming* (SDP) [VB96]. It consists in the minimization of a linear function under constraints defined by matrix inequalities. SDP programs can be solved efficiently, with polynomial worst-case complexity.

2.2.2 Quadratic forms

CQLFs are the most widely used class of Lyapunov functions. A CQLF is a quadratic Lyapunov function common to all modes of the systems, that is an ellipsoidal norm $\|\cdot\|_P$ with respect to which all modes are contractive: $x^T A^T P A x - x^T P x \leq 0$, $\forall A \in \mathcal{M}$, $\forall x \in \mathbb{R}^n \setminus 0$. Constructing a CQLF hence amounts to solving a SDP program, that we will now define based on *Common quadratic forms* (CQFs).

Definition 2.1. A CQF for System (2.1) with a finite set of matrices $\mathcal{M} \subset \mathbb{R}^{n \times n}$ is a matrix $P \in \mathcal{S}_{++}^n$ such that for some $\gamma \geq 0$

$$(Ax)^T P (Ax) \leq \gamma^2 x^T P x, \quad \forall A \in \mathcal{M}, \quad \forall x \in \mathbb{R}^n \setminus 0. \quad (2.2)$$

When $\gamma < 1$, CQFs become CQLFs and can hence serve as stability certificates for switched linear systems. Conversely, a lower bound on the JSR can be obtained if there exists no such CQLF. This lower bound is refined by considering CQFs for the finite set $\mathcal{M}^l := \{\prod_{j=1}^l A_{i_j} : A_{i_j} \in \mathcal{M}\}$ and using Property 2.2. This also allows to benefit from the consideration of long trajectories. These observations are summarized in the following theorem:

Theorem 2.1

[Jun09, Prop. 2.8, Thm. 2.11]

Consider a finite set of matrices $\mathcal{M} \subset \mathbb{R}^{n \times n}$.

- If there exists $\gamma \geq 0$ and $P \in \mathcal{S}_{++}^n$ such that (2.2) holds, then $\rho(\mathcal{M}) \leq \gamma$.

- If $\rho(\mathcal{M}) < \frac{\gamma}{2\sqrt[n]{n}}$, there exists a CQF $P \in \mathcal{S}_{++}^n$ for $\mathcal{M}^l$ such that

$$(\mathbf{A}x)^T P \mathbf{A}x \leq \gamma^{2l} x^T P x, \quad \forall \mathbf{A} \in \mathcal{M}^l, \forall x \in \mathbb{R}^n \setminus 0$$

The accuracy of the upper bound on the JSR in Theorem 2.1 increases when γ decreases. Hence, we seek for the minimal γ such that (2.2), for the system defined by the set $\mathcal{M}^l$, is satisfied. Therefore, we consider the following quasi-linear optimization problem:

$$\begin{aligned} \min_{P, \gamma \geq 0} \quad & \gamma \\ \text{s.t.} \quad & (\mathbf{A}x)^T P \mathbf{A}x \leq \gamma^{2l} x^T P x, \quad \forall \mathbf{A} \in \mathcal{M}^l, \forall x \in \mathbb{S} \\ & P \succeq I_n, \quad \|P\| \leq C \text{ for a large } C \in \mathbb{R}_{\geq 0}. \end{aligned} \quad (2.3)$$

We denote by (γ^o, P^o) the solution to Problem (2.3), which is a quasi-linear optimization problem. While the optimal cost γ^o is unique, there can be several optimal P^o , that is different feasible P leading to the same optimal cost. When this occurs, we apply a *tie-breaking rule*: P^o denotes the optimal P of smallest condition number, since the accuracy of our upper bound will depend on this number.

Problem (2.3) differs from Program (2.2) in three ways. First, we impose a bound on the norm of P to ensure the compactness of the set of constraints. Since this bound can be arbitrarily high, the solution is not affected. Second, $P \succeq I_n$ replaces the constraint $P \succ 0$, and third, we restrict x to $\mathbb{S}$ instead of $\mathbb{R}^n \setminus 0$. These two latter restrictions do not incur any conservativeness because of the *homogeneity* of System (2.1): for $\lambda > 0$, $A_{\tau(k)} \lambda x_k = \lambda A_{\tau(k)} x_k$. This property implies, for instance, that the decrease of a CQLF on an arbitrary set enclosing the origin is sufficient to serve as a stability certificate. Problem (2.3) can be efficiently solved by bisection on the variable γ . Namely, for a fixed accuracy α , we look for a feasible γ^o by bisection on an interval $[0, U]$, where U is an upper bound on the JSR. In a second step, with γ^o fixed, we apply our tie-breaking rule and obtain P^o . More precisely, once γ^o is obtained, we compute P^o as the solution to:

$$\begin{aligned} \min_{P \succeq I_n} \quad & \lambda_{\max}(P) \\ \text{s.t.} \quad & (\mathbf{A}_j x)^T P \mathbf{A}_j x \leq \gamma^{o2l} x^T P x, \quad \forall \mathbf{A} \in \mathcal{M}^l, \forall x \in \mathbb{S}, \quad \|P\| \leq C. \end{aligned} \quad (2.4)$$

Remark 2.1. When all modes are known, Problem (2.3) can be handled as a semidefinite program, by using matrix inequalities: (γ^o, P^o) is the solution to

$$\min_{P \succeq I_n, \gamma \geq 0} \quad \gamma \quad \text{s.t.} \quad \mathbf{A}^T P \mathbf{A} \preceq \gamma^{2l} P, \quad \forall \mathbf{A} \in \mathcal{M}^l, \quad \|P\| \leq C.$$

It can thus be solved efficiently, by solving at each iteration of the bisection on γ a feasibility SDP problem.

2.2.3 SOS forms

The bound on the JSR obtained by Program (2.3) can, in the white-box setting, be improved by considering SOS polynomials $p(x)$, of degree $2d$. These are polynomials admitting a quadratic decomposition $p(x) = \sum_i p_i(x)^2$, where $p_1, p_2, \dots$ are polynomials of degree d [PJ08]. Indeed, to extend quadratic Lyapunov which are polynomials of degree 2, an idea is to consider polynomials of higher degree, again decreasing with respect to all modes. The following proposition holds:

Proposition 2.1. [Jun09, Thm. 2.13]

Consider a finite set of matrices $\mathcal{M} \subset \mathbb{R}^{n \times n}$. Let $p(x)$ be a strictly positive ($p(x) > 0 \forall x \neq 0$) homogeneous polynomial of degree $2d$, such that for some $\gamma \geq 0$

$$p(Ax) \leq \gamma^2 p(x), \quad \forall A \in \mathcal{M}, \quad \forall x \in \mathbb{R}^n \setminus 0. \quad (2.5)$$

Then, $\rho(\mathcal{M}) \leq \sqrt[d]{\gamma}$.

For $2d > 4$, deciding whether or not a multivariate polynomial of degree $2d$ is strictly positive is NP-hard. However, SOS functions are clearly always positive, and easy to handle. Hence, by releasing the strictly positive assumption on $p(x)$ into the assumption that it is a SOS polynomial, one obtains a tractable and rich class of Lyapunov functions. Before proceeding further, we recall the notion of *symmetric algebra of a vector space*:

Definition 2.2. [Jun09] and [PJ08]

Let $x \in \mathbb{R}^n$. Let D denote the number of different monomials of degree d in n variables, for some $d \in \mathbb{N}_0$:

$$D = \binom{n+d-1}{d}. \quad (2.6)$$

The d -lift of x , denoted by $x^{[d]}$, is a vector in $\mathbb{R}^D$ containing all possible monomials of degree d in n variables, judiciously scaled. $x^{[d]}$ is thus indexed by the D different n -dimensional vectors α such that $\alpha_1 + \dots + \alpha_n = d$ and $\alpha_i \geq 0$ for $i = 1, \dots, n$:

$$x_\alpha^{[d]} = \sqrt{\frac{d!}{\alpha_1! \alpha_2! \dots \alpha_n!}} x_1^{\alpha_1} \dots x_n^{\alpha_n}.$$

The entries of $x^{[d]}$ are scaled so that the lift is norm-preserving up to power d :

$$\|x^{[d]}\| = \|x\|^d. \quad (2.7)$$

The d -lift of the matrix $A \in \mathbb{R}^{n \times n}$ is the unique matrix $A^{[d]} \in \mathbb{R}^{D \times D}$ satisfying

$$A^{[d]} x^{[d]} = (Ax)^{[d]}.$$

The entries of $A^{[d]}$ can be obtained via the following closed-form expression:

$$A_{\alpha\beta}^{[d]} = \frac{\text{per}(A(\alpha, \beta))}{\sqrt{\mu(\alpha)\mu(\beta)}}.$$

In this expression, α, β are the D different d -elements multisets of $\{1, \dots, n\}$ and $\mu(\alpha)$ denotes the product of the factorials of the multiplicities of the entries of α . The permanent of a matrix $B \in \mathbb{R}^{m \times m}$ is denoted by $\text{per}(B)$: $\text{per}(B) = \sum_{\pi \in \Pi_m} \prod_{i=1}^m b_{i, \pi(i)}$, where Π_m is the set of permutations of $\{1, \dots, m\}$. Finally, $A(\alpha, \beta)$ is the $d \times d$ matrix whose entries are given by $a_{\alpha(i), \beta(j)}$.

We illustrate the concept of a d-lift in the following example:

Example 2.2. [PJ08, Ex. 3.1, Ex. 3.2]

Let $n = 2$, and $x = \begin{bmatrix} x_1 \\ x_2 \end{bmatrix}$. Then,

$$x^{[1]} = \begin{bmatrix} x_1 \\ x_2 \end{bmatrix}, \quad x^{[2]} = \begin{bmatrix} x_1^2 \\ \sqrt{2}x_1x_2 \\ x_2^2 \end{bmatrix}, \quad x^{[3]} = \begin{bmatrix} x_1^3 \\ \sqrt{3}x_1^2x_2 \\ \sqrt{3}x_1x_2^2 \\ x_2^3 \end{bmatrix}.$$

The 4 different 3-elements multisets of $\{1, 2\}$ are $\{1, 1, 1\}$, $\{1, 1, 2\}$, $\{1, 2, 2\}$ and $\{2, 2, 2\}$. Hence,

$$A^{[3]} = \begin{bmatrix} a_{11}^3 & \sqrt{3}a_{11}^2a_{12} & \sqrt{3}a_{11}a_{12}^2 & a_{12}^3 \\ \sqrt{3}a_{11}^2a_{21} & a_{11}^2a_{22} + 2a_{11}a_{21}a_{12} & a_{12}^2a_{21} + 2a_{11}a_{12}a_{22} & \sqrt{3}a_{12}^2a_{22} \\ \sqrt{3}a_{11}a_{21}^2 & a_{21}^2a_{12} + 2a_{11}a_{21}a_{22} & a_{22}^2a_{11} + 2a_{22}a_{21}a_{12} & \sqrt{3}a_{12}a_{22}^2 \\ a_{21}^3 & \sqrt{3}a_{21}^2a_{22} & \sqrt{3}a_{21}a_{22}^2 & a_{22}^3 \end{bmatrix}$$

We are now ready to introduce SOS forms to approximate the JSR of $\mathcal{M}$. First, we relate the SOS decomposition to a SDP constraint:

Proposition 2.2. [PJ08, Thm. 2.3]

A homogeneous multivariate polynomial $p(x)$ of degree $2d$ is a SOS polynomial if and only if for some $P \in \mathcal{S}_{++}^D$, where D is given by (2.6),

$$p(x) = (x^{[d]})^T P x^{[d]}.$$

A SOS polynomial thus extends the concept of a quadratic polynomial in a lifted space. We now introduce *SOS forms* similarly as CQFs. Even though we drop the common prefix, it is implicit that the forms are common to all modes.

Definition 2.3. A SOS form for System (2.1) with a finite set of matrices $\mathcal{M} \subset \mathbb{R}^{n \times n}$ is a matrix $P \in \mathcal{S}_{++}^D$ such that for some $\gamma \geq 0$,

$$((Ax)^{[d]})^T P (Ax)^{[d]} \leq \gamma^{2d} (x^{[d]})^T P x^{[d]}, \quad \forall A \in \mathcal{M}, \quad \forall x \in \mathbb{S}, \quad (2.8)$$

where the restriction of x to $\mathbb{S}$ is again due to the homogeneity of System (2.1).

Similarly to CQFs, SOS functions become SOS Lyapunov functions when $\gamma < 1$, and can provide upper and lower bounds on the JSR in any case. These approximations of the JSR using SOS functions are tighter than when resorting to quadratic functions. The improvement holds for all systems but is especially impressive for stable systems that do not admit a CQLF. Considering SOS forms for $\mathcal{M}^l$ improves the accuracy of the lower bounds on the JSR, as summarized in the following theorem:

Theorem 2.2

[Jun09, Thms. 2.13 and 2.16]

Consider a finite set of matrices $\mathcal{M} \subset \mathbb{R}^{n \times n}$.

- If there exists $\gamma \geq 0$ and $P \in \mathcal{S}_{++}^D$ such that (2.8) holds, then $\rho(\mathcal{M}) \leq \gamma$.
- If $\rho(\mathcal{M}) < \frac{\gamma}{2^{1/d}\sqrt{\eta}}$ where $\eta = \min\{D, m\}$, there exists a SOS form $P \in \mathcal{S}_{++}^D$ for $\mathcal{M}^l$ such that

$$((\mathbf{A}x)^{[d]})^T P (\mathbf{A}x)^{[d]} \leq \gamma^{2dl} (x^{[d]})^T P x^{[d]}, \quad \forall \mathbf{A} \in \mathcal{M}^l, \quad \forall x \in \mathbb{S}.$$

We generalize Problem (2.3) as follows:

$$\begin{aligned} & \min_{P, \gamma \geq 0} \gamma \\ & \text{s.t.} \quad (\mathbf{A}x)^{[d]T} P (\mathbf{A}x)^{[d]} \leq \gamma^{2dl} x^{[d]T} P x^{[d]}, \quad \forall \mathbf{A} \in \mathcal{M}^l, \quad \forall x \in \mathbb{S} \\ & \quad \quad P \succeq I_D, \quad \|P\| \leq C \text{ for a large } C \in \mathbb{R}_{\geq 0}. \end{aligned} \quad (2.9)$$

Its solution is denoted by $(\gamma_{sos}^o, P_{sos}^o)$, with the same tie-breaking rule applied to P_{sos}^o when the latter is not unique. When $d = 1$, we recover Problem (2.3).

Remark 2.2. Clearly, when $x \in \mathbb{S}$, $x^{[d]}$ belongs to the unit sphere in $\mathbb{R}^D$, denoted by $\mathbb{S}^D$. However, for $d \geq 2$, the d-lift of $\mathbb{S}$ only consists of a subset of this sphere of higher dimension, as illustrated in Figure 2.3 for $n = 2$ and $d = 2$. Hence, for $d \geq 2$, two differences separate Problems (2.9) and (2.3). First, the dimension of the problem increases from $d_{quad} + 1 = \frac{n(n+1)}{2} + 1$ to $d_{sos} + 1 = \frac{D(D+1)}{2} + 1$, where D is given by (2.6). Second, in Problem (2.9), the constraints hold only for the set $\{x \in \mathbb{S}^D : x = y^{[d]}, y \in \mathbb{S}\}$ instead of the whole sphere $\mathbb{S}^D$.

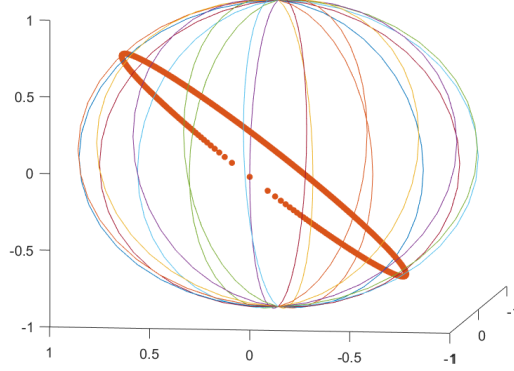


Figure 2.3: Representation of the d -lift of $\mathbb{S}$, for $n = 2$ and $d = 2$. This d -lift, given by $\{x \in \mathbb{S}^D : x = y^{[d]} \text{ for } y \in \mathbb{S}\}$, is represented in red on the figure. It only forms a subset of $\mathbb{S}^D$.

Remark 2.3. Approximations of the JSR can be further improved by using *polyhedral Lyapunov functions* [BM07], defined with respect to *polyhedral sets* of the form $S = \{x \in \mathbb{R}^n : F_i x \leq 1, i = 1, 2, \dots, s\}$, as

$$V(x) = \max_{i=1,2,\dots,s} F_i x.$$

Indeed, they are non-conservative, but can be extremely hard to compute since their complexity is not bounded by the space dimension but by the number of planes describing S . We present a probabilistic bound on the JSR relying on such functions in Appendix A. However, we should keep in mind that the computational complexity of the method will be much higher than the one we propose based on CQLFs and SOS functions.

When we have access to the model, that is to the modes of the system, solving Problems (2.3) and (2.9) is easily handled using matrix inequalities (see Remark 2.1), and provides deterministic stability guarantees [Lib03]. However, in our black-box setting, we do not have access to the modes, and cannot solve Problems (2.3) and (2.9) using matrix inequalities. Moreover, we note that these problems are then defined by an infinite number of constraints, while we only sample a finite number N of trajectories of System (2.1), corresponding to a finite subset of the constraints of Problems (2.3) and (2.9). We hence seek to answer the following question: after observing these N trajectories of length l ; what **probabilistic** upper bound on the JSR of the underlying system, holding with a given confidence level, can we provide?

To answer this question, we first present a method used to solve optimization

problems in which we only have access to a subset of the constraints: the scenario approach.

2.3 The scenario approach

Relying on data to obtain effective control methods, such as Lyapunov functions or invariant sets; and assess efficiently system-theoretic properties despite the complexity of the underlying system and the lack of a closed-form description of the model is a topic of high interest in our data-based world: how to benefit from the potential of the torrents of data we have access to? Hence, scenario-based optimization has been widely studied in the past decade [Cal10], [NY19], [CGR18] [CG18].

Scenario-based optimization is one of the subfields of optimization under uncertainty, that is optimization in which the constraints depend on an uncertainty parameter. Suppose we have a problem of the form

$$\min_{u \in \mathcal{U} \subseteq \mathbb{R}^n} c^T u \text{ s.t. } f(u, \delta) \leq 0, \forall \delta \in \Delta,$$

of solution u^o , where Δ represents some form of uncertainty and $\mathcal{U}$ is a set of deterministic constraints. It is often difficult to derive a solution satisfying all constraints, because we only have access to a subset of them or because the size of Δ causes the problem to be computationally intractable [BEFB94]. In our setting, for instance, $\Delta = (x, \mathbf{A}x) \forall x \in \mathbb{S}, \forall \mathbf{A} \in \mathcal{M}^l$, so that we have a closed-form expression for Δ , but no access to the entirety of it. There exist several approaches to deal with this family of problems [NY19]. For instance, the chance-constrained approach consists in considering δ as a random variable, and in deriving a solution u^* satisfying the constraints up to a given level of probability, that is satisfying $\mu\{f(u^*, \delta) > 0\} \leq \epsilon$, where μ is a probability measure on Δ [SDR09].

The approach suiting our setting best is the scenario approach, which allows to obtain probabilistic guarantees about the unobserved samples in Δ based on a given number of realizations of Δ . In this approach, we consider a finite set $\omega_N \subset \Delta$ composed of N constraints (associated in our setting to the trajectories $(x, \mathbf{A}x)$ we have access to) independently and identically sampled among all possible constraints according to the probability measure μ on Δ . We then solve the sampled optimization problem, denoted by $P(\omega_N)$, defined by these constraints:

$$\min_{u \in \mathcal{U} \subseteq \mathbb{R}^n} c^T u \text{ s.t. } f(u, \delta) \leq 0, \forall \delta \in \omega_N.$$

This problem hence uses available data in a direct way. Since it does not contain any uncertainty and consists of a finite number of constraints, it is efficiently solvable. Its solution, denoted by $u^*(\omega_N)$, is a random variable, which outperforms

the solution of the initial problem: $c^T u^*(\omega_N) \leq c^T u^o$, to the expense of potentially violating some constraints of the initial problem.

The relationship between the initial problem and the sampled problem, depending on the number N of sampled constraints, has been widely studied recently, from two perspectives. First, one can analyse the feasibility of the sampled solution, and bound the measure of the set of initial constraints possibly violated by the sampled solution [Cal10] [CGR18] [CG18]. This perspective is hence related to the chance-constrained approach: given the number N of constraints enforced, one can derive a probability with which the sampled solution is feasible for the chance-constrained problem. Second, one can analyze the objective performance of the sampled solution, that is the distance between its cost and the cost of the initial problem [MSL15].

To analyze the feasibility of the sampled solution, we need the following definition:

Definition 2.4. [Cal10, defn. 2.9]

An *essential set* for $P(\omega_N)$ is a subset $Es \subset \omega_N$, with minimal cardinality, such that $c^T u^*(Es) = c^T u^*(\omega_N)$.

In other terms, an essential set is a set only containing constraints whose removal imply a decrease in the cost of $P(\omega_N)$. One can derive upper bounds on the cardinality of essential sets, for two specific versions of $P(\omega_N)$:

Lem 2.1. [Cal10, Lem. 2.10]

Let $P(\omega_N)$ be a convex problem, such that $f(u, \delta)$ is convex in u for any fixed $\delta \in \Delta$. Then, $|Es| \leq n + 1$ for any essential set Es of $P(\omega_N)$, where n is the dimension of the variable u .

Lem 2.2. [BJW21, Thm. 2]

Let $P(\omega_N)$ be a quasi-linear problem, of the form

$$\min_{u \in \mathbb{R}^n, \lambda \geq 0} \lambda \text{ s.t. } a_\delta^T u \leq \lambda b_\delta^T u, \forall \delta \in \omega_N.$$

Then, $|Es| \leq n$ for any essential set Es of $P(\omega_N)$.

An assumption on $P(\omega_N)$ required to analyze the feasibility of its solution is its non-degeneracy:

Definition 2.5. [Cal10, Lem. 2.11]

$P(\omega_N)$ is said to be non-degenerate if it admits a unique essential set.

We can now present the probabilistic guarantees on the measure of the set of constraints violated by the sampled solution $u^*(\omega_N)$, for non-degenerate problems:

Proposition 2.3. [Cal10, Thm. 3.3]

Let $P(\omega_N)$ be a non-degenerate problem, with a unique solution after, if necessary, the application of a tie-breaking rule. Let $N \geq n$, and let ξ be an upper bound on the cardinality of the essential set of $P(\omega_N)$. Then, for all $\epsilon \in (0, 1)$, the following holds:

$$\mu^N\{\omega_N \in \Delta^N : \mu(V(\omega_N)) \leq \epsilon\} \geq 1 - \sum_{j=0}^{\xi-1} \binom{N}{j} \epsilon^j (1 - \epsilon)^{N-j}$$

where $V(\omega_N) := \{\delta \in \Delta : f(u^*(\omega_N), \delta) > 0\}$ is the set of violated constraints.

Applying Lemmas 2.1 and 2.2 to Proposition 2.3 provides the following corollaries:

Corollary 2.1. Let $P(\omega_N)$ be as in Proposition 2.3, and let it be convex. Then, with the notation of Proposition 2.3, it holds that for all $\epsilon_c \in (0, 1)$,

$$\mu^N\{\omega_N \in \Delta^N : \mu(V(\omega_N)) \leq \epsilon_c\} \geq \beta_c(\epsilon_c, N) = 1 - \sum_{j=0}^n \binom{N}{j} \epsilon_c^j (1 - \epsilon_c)^{N-j}.$$

We can also define $\epsilon_c(\beta_c, N) = 1 - I^{-1}(1 - \beta_c, N - n, n + 1)$.

Corollary 2.2. Let $P(\omega_N)$ be as in Proposition 2.3, and let it be quasi-linear. Then, with the notation of Proposition 2.3, it holds that for all $\epsilon_{ql} \in (0, 1)$,

$$\mu^N\{\omega_N \in \Delta^N : \mu(V(\omega_N)) \leq \epsilon_{ql}\} \geq \beta_{ql}(\epsilon_{ql}, N) = 1 - \sum_{j=0}^{n-1} \binom{N}{j} \epsilon_{ql}^j (1 - \epsilon_{ql})^{N-j}.$$

We can also define $\epsilon_{ql}(\beta_{ql}, N) = 1 - I^{-1}(1 - \beta_{ql}, N - n - 1, n)$.

Proposition 2.3 provides an explicit bound on the number of samples required to obtain, with a high probability, a sampled solution feasible for the chance-constrained problem. However, in general, one cannot straightforwardly translate this knowledge into a confidence interval for the cost of the initial problem as a function of the sampled cost. Hence, analysis of the relationship between the initial and the sampled problems from the performance perspective is far from trivial. There exist problems for which, if a specific set of constraints, even of arbitrary small measure, is not satisfied, the feasible cost decreases greatly, as illustrated in the following example:

Example 2.3. Consider the following problem:

$$\min_{u \in \mathbb{R}} u \text{ s.t. } \begin{cases} u \geq 10^{10}, & \delta \in [0, 10^{-5}] \\ u \geq 0, & \delta \in (10^{-5}, 1] \end{cases}$$

whose optimal solution is given by $u^o = 10^{10}$. By allowing the measure of the set of violated constraints to be of $\epsilon \geq 10^{-5}$, it can happen that the whole set of constraints associated to $\delta \in [0, 10^{-5}]$ gets neglected and that the sampled solution is of 0.

In [MSL15], a probabilistic bound on the optimal cost is given, in the form $\mu^N \{c^T u^o - c^T u^*(\omega_N) \in [0, I(\epsilon)]\} \geq 1 - \beta_c$, where β_c is defined as in Corollary 2.1. However, it only holds for convex problems and it suffers from the curse of dimensionality, so that $I(\epsilon)$ grows exponentially with the dimension of Δ . Hence, other methods have to be derived to obtain a probabilistic bound on the JSR based on the observation of N trajectories, relying on the specific structure of the problem.

We now present a non-exhaustive list of further developments in the scenario approach. We focus on developments that will be of interest in the course of Chapter 5, or attempting to decrease the complexity of the problem.

- Extension of the feasibility analysis to multi-stage optimization [VKR12] [KML16]. In this framework, the decision variable $u = (u_1, \dots, u_m)$ has to satisfy several problems of the form

$$\min_{u_i \in \mathcal{U}_I \subseteq \mathbb{R}^{n_i}} c_i^T u_i \text{ s.t. } f(u_i, u_{i-1}, \delta) \leq 0, \forall \delta \in \Delta_i,$$

for $i = 2, \dots, m$, where $\mathcal{U}_I$ represents a set of deterministic constraints. A bound on the total number of samples to be sampled to obtain a global solution feasible for a large set of constraints is then derived. This total number of samples is divided into m sets of samples used to solve each subproblem. This sequential problem will serve in Chapter 5, and the hereabove mentioned references will be further investigated in this context.

- Attempt to decrease the sample complexity as well as the computational complexity of the approach by solving a two-phase process: the first phase consists in computing $u^*(\omega_N)$ based on N samples, and in the second phase, one only updates part of the decision variable u while maintaining the rest fixed to $u^*(\omega_N)$ [CGC14]. This algorithm is further investigated in Chapter 5.
- Attempts to decrease the computational cost of the approach by relying on distributed algorithms [CSBC14].
- Techniques to decrease the sampled cost while maintaining the same level of probability on the measure of the set of violated constraints, by allowing some sampled constraints to be discarded [Cal10] [CG11]. The choice of the constraints to be discarded to achieve an optimal trade-off between feasibility

and performance is however a NP-hard problem. Moreover, this increases both the computational and sample complexity of the scenario approach.

Part II

A broader and more precise class
of probabilistic stability
guarantees

Chapter 3

Data-driven computation of the JSR

In this chapter, we first formally present the data-driven approach to compute approximations of the JSR, by introducing the black-box setting in the construction of Lyapunov functions described in Chapter 2. We also briefly present the approach proposed in [KBJT19] and [BJW21] to obtain probabilistic upper bounds on the JSR. We then derive deterministic lower bounds on the JSR based on approximate CQLFs and SOS Lyapunov functions.

In a second time, we present our first main contribution by answering the following question:

How to provide probabilistic upper bounds on the JSR of switched linear systems, based on a method which do not rely on chance-constrained theorems and which can be used in the quadratic and in the SOS setting?

We thus propose a novel approach, based on sensitivity analysis of the quasi-linear optimization problems (2.3) and (2.9), which allows to approximate their solution within a given confidence level.

We first derive our approach in the quadratic setting, and obtain two novel upper bounds based on CQLFs. Then, we leverage our technique to the SOS framework and take advantage of the tighter approximation of the JSR allowed by SOS Lyapunov functions.

3.1 Problem statement

In Chapter 2, we proposed a method to obtain deterministic upper bounds on the JSR of a switched linear system whose model is available. The bounds are given by

the optimal costs of the quasi-linear problems (2.3) and (2.9). These problems are defined by an infinite number of constraints since $(\mathbf{A}x)^{[d]T}P(\mathbf{A}x)^{[d]} \leq \gamma^{2dl}x^{[d]T}Px^{[d]}$ must hold $\forall(x, \mathbf{A}) : x \in \mathbb{S}, \mathbf{A} \in \mathcal{M}^l$.

However, in our setting, we only have access to a finite number N of trajectories of length l of System (2.1), as well as to an upper bound on the number of modes of the actual system. The trajectories are denoted by $(x_{i,0}, x_{i,1}, \dots, x_{i,l}) = (x_{i,0}, A_{j_{i,1}}x_{i,0}, \dots, A_{j_{i,l}}x_{i,l-1})$ for $i = 1, \dots, N$. Hence, we can only implement N constraints of Problems (2.3) and (2.9), defined by the N observations $(x_{i,0}, x_{i,1}, \dots, x_{i,l})$. Based on the solutions obtained when solving these relaxations, we seek to infer a bound on the solution of the initial problems, holding with a certain confidence level.

The trajectories are assumed to be generated from N initial states $x_{i,0}$ drawn randomly, uniformly and independently from $\mathbb{S}$ according to σ^{n-1} . Then, N sequences of l modes are drawn randomly, uniformly and independently from M^l according to μ_M and applied to each initial condition. The set of N observations of System (2.1) $\{(x_{i,0}, x_{i,1}, \dots, x_{i,l}), i = 1, 2, \dots, N\}$ is therefore associated to a uniform sample of N $(l+1)$ -tuples in Z_l :

$$\omega_N := \{(x_{i,0}, j_{i,1}, \dots, j_{i,l}), i = 1, 2, \dots, N\} \in Z_l^N. \quad (3.1)$$

We denote $\mathbf{A}_{\mathbf{j}} := A_{j_l}A_{j_{l-1}}\dots A_{j_1}$ and $\mathbf{j} := \{j_1, \dots, j_l\}$.

In the framework of scenario approach, for a given data set ω_N , we define the sampled optimization problems associated to (2.3) and (2.9), that is the relaxed problems defined by the N constraints associated to the N observed trajectories. The sampled optimization problem to be solved in the quadratic case, denoted by $\text{Opt}(\omega_N)$, is:

$$\begin{aligned} \min_{P, \gamma \geq 0} \quad & \gamma \\ \text{s.t.} \quad & (\mathbf{A}_{\mathbf{j}}x)^T P \mathbf{A}_{\mathbf{j}}x \leq \gamma^{2l} x^T P x, \quad \forall(x, \mathbf{j}) \in \omega_N \\ & P \succeq I_n, \quad \|P\| \leq C \text{ for a large } C \in \mathbb{R}_{\geq 0}. \end{aligned} \quad (3.2)$$

Its solution is denoted by $(\gamma^*(\omega_N), P^*(\omega_N))$, where the same tie-breaking rule is applied on $P^*(\omega_N)$ if needed. Observe that Problem (3.2) only requires the knowledge of $\mathbf{A}_{\mathbf{j}}x$, available through the observations. Namely, it does not require the knowledge of $\mathbf{j}$, which is not observed in our setting.

In the SOS case, we consider the following sampled problem, denoted by $\text{Opt}_{\text{sos}}(\omega_N)$ and with solution $(\gamma_{\text{sos}}^*(\omega_N), P_{\text{sos}}^*(\omega_N))$:

$$\begin{aligned} \min_{P, \gamma \geq 0} \quad & \gamma \\ \text{s.t.} \quad & ((\mathbf{A}_{\mathbf{j}}x)^{[d]})^T P (\mathbf{A}_{\mathbf{j}}x)^{[d]} \leq \gamma^{2dl} (x^{[d]})^T P x^{[d]}, \quad \forall(x, \mathbf{j}) \in \omega_N \\ & P \succeq I_D, \quad \|P\| \leq C \text{ for a large } C \in \mathbb{R}_{\geq 0}. \end{aligned} \quad (3.3)$$

Once the solution to Problems (3.2) and (3.3) are obtained, we aim to infer the solutions to Problems (2.3) and (2.9), respectively, within a certain confidence, as illustrated on Figure 3.1.

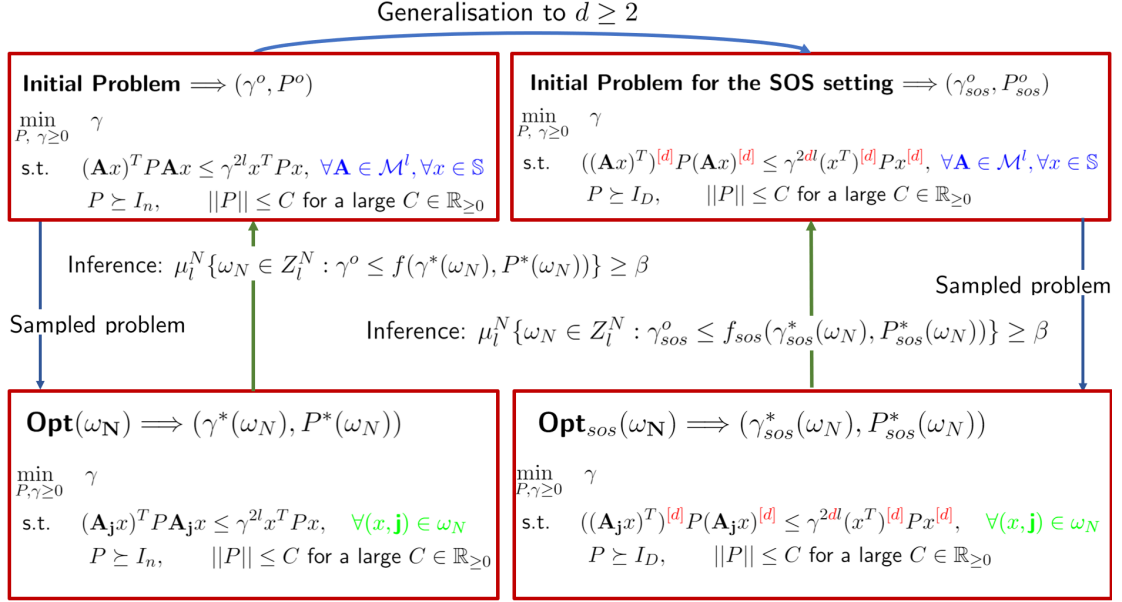


Figure 3.1: Relations between the different optimization problems introduced. The left-hand side of the figure is related to the quadratic setting, while the right-hand side presents the extension to the SOS setting. We infer an upper bound on the costs of initial problems, γ^o and γ_{sos} , as a function of the solutions to the respective sampled problems. The sampled problems are relaxations of the initial problems. They only require the knowledge of N trajectories of System (2.1) to be solved. The upper bounds, which approximates the JSR, hold with a certain confidence level β depending on N .

Observe that solving the sampled problems amounts to computing approximate Lyapunov functions, decreasing on a large number of trajectories instead of everywhere in the state space. Problems (3.2) and (3.3) are both quasi-linear problems, and they are non-degenerate under the assumption that $\mathcal{M}^l$ does not contain any matrix $\mathbf{A}$ such that $\mathbf{A}^T P \mathbf{A} = \gamma^2 P$ for some $P \in \mathcal{S}_{++}^n$ and some $\gamma > 0$ [BJW21, Section 4]. Hence, Corollary 2.2 provides a bound on the measure of the set of trajectories potentially not satisfying the Lyapunov property, holding with a certain confidence level depending on N . However, translating this knowledge into an assumption about the JSR of the system is far from trivial. Indeed, there exist approximate Lyapunov functions decreasing everywhere in the state space except for an arbitrarily small set, while almost all of the trajectories of the analyzed

system diverge to infinity.

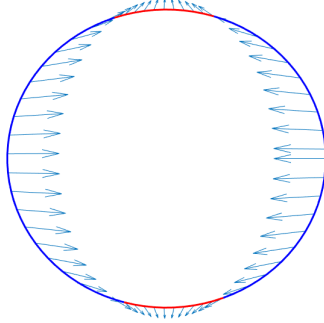


Figure 3.2: Illustration of the dynamics of the unstable system of Example 3.1 on the level set of an approximate Lyapunov function. The trajectories of the system starting on the level set diverge to the infinity, even though the set of points violating the Lyapunov property, represented in red, is arbitrarily small.

Example 3.1. Consider the unstable linear system defined by the matrix

$$A = \begin{bmatrix} 0.1 & 0 \\ 0 & 1.01 \end{bmatrix},$$

and the almost Lyapunov function $V(x) = x_1^2 + x_2^2$, defined by $P = I_2$. It satisfies the Lyapunov property almost everywhere on the sphere: if $x_1^2 + x_2^2 = 1$, then

$$(Ax)^T(Ax) \leq x^T x \iff 0.01x_1^2 + 1.0201x_2^2 \leq x_1^2 + x_2^2 \iff |x_2| \leq 0.99.$$

The set of points violating the Lyapunov property is represented by the red areas on Figure 3.2. However, almost all trajectories diverge to infinity, as illustrated on the same figure.

3.2 Presentation of an existing technique

In [KBJT19] and [BJW21], after the computation of an approximate CQLF, the feasibility property of Corollary 2.2 is translated into a probabilistic bound on the JSR, based on geometric properties of switched linear systems. Indeed, one knows a bound, holding with a certain probability β_{ql} , on the measure of the set of points of the computed ellipsoid $E_{P^*(\omega_N)}$ that do potentially not satisfy the Lyapunov property $(\mathbf{A}_j x)^T P^*(\omega_N) \mathbf{A}_j x \leq \gamma^*(\omega_N)^{2l} x^T P^*(\omega_N) x$, $\forall \mathbf{A}_j \in \mathcal{M}^l$. This bound is obtained by applying Corollary 2.2 to Problem (3.2), and projecting the set of

violating constraints onto the ellipsoid. From this, one can compute the convex hull of the subset of $E_{P^*(\omega_N)}$ whose points satisfy the property with probability β_{ql} . More precisely, since we do not know explicitly which points of $E_{P^*(\omega_N)}$ belong to the set of violating points, we compute this convex hull for the worst distribution of violating points possible. Then, we can compute the largest scaling $\bar{\alpha}$ of $E_{P^*(\omega_N)}$ such that $\bar{\alpha}E_{P^*(\omega_N)}$ is included in this convex hull, as illustrated on Figure 3.3. Because System (2.1) is convexity-preserving, it holds that all trajectories satisfy a relaxed Lyapunov property: $(\mathbf{A}_j x)^T P^*(\omega_N) \mathbf{A}_j x \leq \frac{\gamma^*(\omega_N)^{2l}}{\alpha^2} x^T P^*(\omega_N) x$. Hence, $\rho(\mathcal{M}) \leq \frac{\gamma^*(\omega_N)}{\sqrt[l]{\alpha}}$ holds with probability at least β_{ql} .

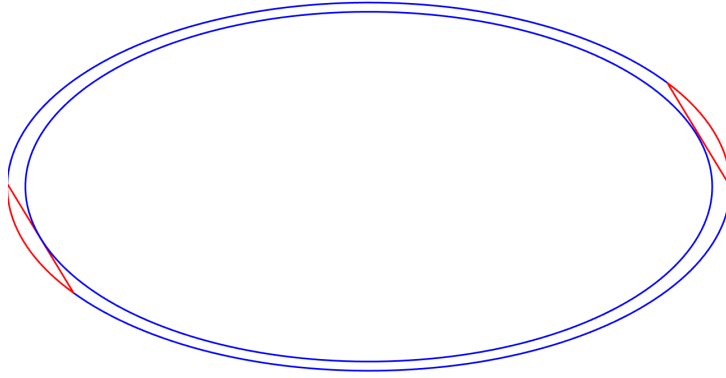


Figure 3.3: Illustration of the existing technique to obtain probabilistic stability guarantees. The largest ellipsoid is $E_{P^*(\omega_N)}$, computed by solving Problem (3.2). The red portions of $E_{P^*(\omega_N)}$ represent the worst-case distribution of the points that potentially violate the Lyapunov property. The inscribed ellipsoid is the largest scaling of $E_{P^*(\omega_N)}$ included in the convex hull of the subset of $E_{P^*(\omega_N)}$ which does not belong to the violating set. All points on this ellipsoid satisfy a relaxed Lyapunov property.

This technique cannot be extended to the SOS setting due to the distribution of the vectors $x^{[d]}$. Indeed, they are not uniformly sampled on the sphere of dimension D when the vectors x are uniformly sampled on $\mathbb{S}$ (see Remark 2.2), while the geometric properties used require a uniform distribution of the sampled points. Moreover, it only holds for a certain class of systems, satisfying the non-degeneracy assumption required to use Corollary 2.2 [BJW21, Section 4].

Therefore, we propose a new method, which can be leveraged to SOS functions and provides bounds valid without any assumption on the system. Before presenting this method, we derive deterministic lower bounds on the JSR that can be obtained by solving the sampled problems (3.2) and (3.3). Combined to the probabilistic upper bounds, these lower bounds allow to infer a confidence interval for the true value of the JSR.

3.3 Deterministic lower bounds on the JSR

The optimal costs of the sampled problems are necessarily smaller than the optimal costs of the initial problems. Hence, by Theorems 2.1 and 2.2, we obtain the following deterministic lower bounds on $\rho(\mathcal{M})$, based respectively on $\gamma^*(\omega_N)$ and $\gamma_{sos}^*(\omega_N)$:

Lem 3.1. [KBJT19, Thm. 3]

Consider a finite set of matrices $\mathcal{M} \subset \mathbb{R}^{n \times n}$. Let ω_N be a uniform random sample drawn from Z_l^N , and $(\gamma^*(\omega_N), P^*(\omega_N))$ be the optimal solution to Problem (3.2). The following holds:

$$\frac{\gamma^*(\omega_N)}{\sqrt[2l]{n}} \leq \rho(\mathcal{M}). \quad (3.4)$$

Lem 3.2. Consider a finite set of matrices $\mathcal{M} \subset \mathbb{R}^{n \times n}$. Let ω_N be a uniform random sample drawn from Z_l^N , and $(\gamma_{sos}^*(\omega_N), P_{sos}^*(\omega_N))$ be the optimal solution to Problem (3.3). The following holds:

$$\frac{\gamma_{sos}^*(\omega_N)}{\sqrt[2dl/\eta]{n}} \leq \rho(\mathcal{M}) \quad (3.5)$$

where $\eta = \min\{m, D\}$.

Note that the lower bound obtained using SOS functions is larger than the one obtained using quadratic functions. Hence, one can expect the gap between the lower and upper bounds on the JSR to be smaller when resorting to SOS functions. We now propose our new probabilistic upper bounds on the JSR, based on the computation of approximate CQLFs or SOS Lyapunov functions.

3.4 Novel quadratic upper bounds

We first derive our method in the quadratic setting, by conducting a sensitivity analysis of Problem (2.3). This provides tools to approximate its solution within a given confidence level.

3.4.1 A new approach based on sensitivity analysis

First, we derive a key result concerning the existence and the cardinality of an essential set for Problem (2.3), that is a finite set of points such that the solutions to

Problems (2.3) and (3.2) defined on this set coincide¹. We denote by $d_{\text{quad}} = \frac{n(n+1)}{2}$ the number of variables of $P \in \mathcal{S}^n$.

Lem 3.3. *Let (γ^o, P^o) be the optimal solution to Problem (2.3). Then, there exists a set $\omega_s \in Z_l^{d_{\text{quad}}+1}$ with $|\omega_s| = d_{\text{quad}} + 1$ such that $\gamma^*(\omega_s) = \gamma^o$, where $\gamma^*(\omega_s)$ is the optimal solution to $\text{Opt}(\omega_s)$.*

Proof. From a standard continuity argument (see Appendix B), it is equivalent to prove that for any $\epsilon > 0$, there exists a set $\omega_s \in Z_l^{d_{\text{quad}}+1}$ with $|\omega_s| = d_{\text{quad}} + 1$ such that $\gamma^*(\omega_s) > \gamma^o - \epsilon$. Problem (2.3) is equivalent to:

$$\min_{P, \gamma \geq 0} \gamma, \quad \text{s.t.} \quad P \in \mathcal{X}_{(x, \mathbf{j}, \gamma)}, \quad \forall (x, \mathbf{j}) \in Z_l \quad (3.6)$$

where

$$\mathcal{X}_{(x, \mathbf{j}, \gamma)} = \{P \in \mathcal{S}^n \mid (\mathbf{A}_{\mathbf{j}}x)^T P \mathbf{A}_{\mathbf{j}}x \leq \gamma^2 x^T P x, P \succeq I_n, \|P\| \leq C\}$$

are sets of dimension d_{quad} . Clearly, all sets $\mathcal{X}_{(x, \mathbf{j}, \gamma)}$ are convex and compact.

Suppose by contradiction that there exists an $\epsilon > 0$ such that, for any $\omega \in Z_l^{d_{\text{quad}}+1}$ with $|\omega| = d_{\text{quad}} + 1$, it holds that $\gamma^*(\omega) \leq \gamma^o - \epsilon$. Let $\gamma' = \gamma^o - \epsilon$. Then, for any $\omega \in Z_l^{d_{\text{quad}}+1}$ with $|\omega| = d_{\text{quad}} + 1$, $\cap_{(x, \mathbf{j}) \in \omega} \mathcal{X}_{(x, \mathbf{j}, \gamma')}$ is nonempty.

However, Helly's theorem [DGK63] states that if every subcollection of cardinality $d + 1$ of a collection of compact convex subsets of $\mathbb{R}^d$ has nonempty intersection, then the whole collection has nonempty intersection.

Hence, $\cap_{(x, \mathbf{j}) \in Z_l} \mathcal{X}_{(x, \mathbf{j}, \gamma')}$ is nonempty. As a result, $\gamma^o \leq \gamma' = \gamma^o - \epsilon$, which leads to a contradiction. $\square$

Based on Lemma 3.3, we will show how to construct, from a set of trajectories ω_N , another set ω'_N such that $\gamma^*(\omega'_N) = \gamma^o$ with a certain probability depending on N and on how close the points of both sets are. Before proceeding further, we introduce the notion of a *spherical cap*:

Definition 3.1. [Li11]

The spherical cap on $\mathbb{S}$ of direction c and measure ϵ , illustrated on Figure 3.4, is defined as $\mathcal{C}(c, \epsilon) := \{x \in \mathbb{S} : c^T x > \|c\| \delta(\epsilon)\}$, where $\delta(\epsilon)$ is given by:

$$\delta(\epsilon) = \sqrt{1 - I^{-1}(2\epsilon; \frac{n-1}{2}, \frac{1}{2})}. \quad (3.7)$$

¹The fundamental difference between Lemma 3.3 and Theorem 2 in [BJW21] lies in the fact that our bound is derived for Problem (2.3) with infinite number of constraints while the bound of [BJW21] is derived for the sampled problem (3.2) with a finite number of constraints.

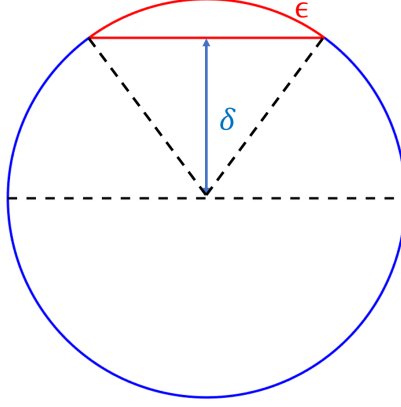


Figure 3.4: Illustration of a spherical cap of measure ϵ in two dimensions.

Proposition 3.1. *Let $\omega_N := \{(x_i, \mathbf{j}_i), i = 1, \dots, N\}$ be a uniform random sample drawn from Z_l^N , with $N \geq d_{\text{quad}} + 1$, and let γ^o be the optimal cost of Problem (2.3). For all $\epsilon \in (0, 1]$, with probability at least*

$$\beta(\epsilon, m, N) := 1 - (d_{\text{quad}} + 1)\left(1 - \frac{\epsilon}{m^l}\right)^N,$$

there exists a set $\omega'_N = \{(x'_i, \mathbf{j}_i), i = 1, \dots, N\}$ such that

- $\gamma^*(\omega'_N) = \gamma^o$, where $\gamma^*(\omega'_N)$ is the optimal solution to $\text{Opt}(\omega'_N)$.
- For $i = 1, \dots, N$: $x'_i \in \mathbb{S}$ and $\|x_i - x'_i\| \leq \sqrt{2 - 2\delta(\epsilon)}$, where $\delta(\cdot)$ is given by (3.7).

Given β , m , N , we can also define $\epsilon(\beta, m, N) = m^l \left(1 - \sqrt[N]{\frac{1-\beta}{d_{\text{quad}}+1}}\right)$.

Proof. By Lemma 3.3, there exists a set ω_s , with $|\omega_s| = d_{\text{quad}} + 1$, such that $\gamma^*(\omega_s) = \gamma^o$. The idea, for each trajectory $(x', \mathbf{j}')$ in ω_s , is to derive the probability that there exist a trajectory $(x, \mathbf{j})$ in ω_N whose initial state belongs to a spherical cap associated to x' and such that $\mathbf{j} = \mathbf{j}'$, as illustrated in Figure 3.5. Given any $(x', \mathbf{j}') \in \omega_s$, let $(x, \mathbf{j})$ be drawn randomly and uniformly from Z_l . Then, the probability that $x \in \mathcal{C}(x', \epsilon)$ and $\mathbf{j} = \mathbf{j}'$ is of $\frac{\epsilon}{m^l}$.

Hence, $\mu_l\{(x, \mathbf{j}) \mid x \notin \mathcal{C}(x', \epsilon) \text{ or } \mathbf{j} \neq \mathbf{j}'\} = 1 - \frac{\epsilon}{m^l}$ for all $(x, \mathbf{j}) \in \omega_N$, and $\mu_l^N\{\omega_N \in Z_l^N \mid x_i \notin \mathcal{C}(x', \epsilon) \text{ or } \mathbf{j}_i \neq \mathbf{j}' \forall (x_i, \mathbf{j}_i) \in \omega_N\} = \left(1 - \frac{\epsilon}{m^l}\right)^N$ by independence of the sampling. These sets, for each $(x', \mathbf{j}') \in \omega_s$, can be disjoint. Hence, the measure of their union, that is the probability that there exists no $(x, \mathbf{j}) \in \omega_N$ such that $x \in \mathcal{C}(x', \epsilon)$ and $\mathbf{j} = \mathbf{j}'$ for a least one $(x', \mathbf{j}') \in \omega_s$, is smaller or equal to $(d_{\text{quad}} + 1)\left(1 - \frac{\epsilon}{m^l}\right)^N$. Therefore, the probability that $\forall (x', \mathbf{j}') \in \omega_s$,

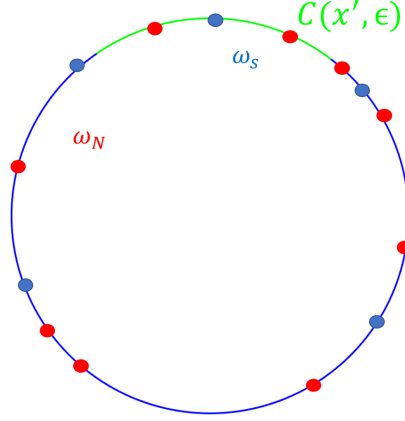


Figure 3.5: Illustration of the argument leading to Proposition 3.1. For each trajectory $(x', \mathbf{j}) \in \omega_s$, where $\gamma^*(\omega_s) = \gamma^o$, there exists a trajectory $(x, \mathbf{j}) \in \omega_N$ such that x belongs to $C(x', \epsilon)$. This holds with probability $\beta(\epsilon, m, N)$.

there exists a point $(x, \mathbf{j}) \in \omega_N$ such that $x \in C(x', \epsilon)$ and $\mathbf{j} = \mathbf{j}'$ is at least of $\beta(\epsilon, m, N) = 1 - (d_{\text{quad}} + 1)(1 - \frac{\epsilon}{m})^N$.

If $x \in C(x', \epsilon)$, then $x'^T x > \|x'\| \delta(\epsilon)$ where $\delta(\epsilon)$ is given by (3.7), and $\|x - x'\|^2 = \|x\|^2 + \|x'\|^2 - 2x^T x' \leq 2 - 2\delta(\epsilon)$. We define the set ω'_N from ω_N as follows: for the $(x, \mathbf{j}) \in \omega_N$ such that $x \in C(x', \epsilon)$ and $\mathbf{j} = \mathbf{j}'$ for some $(x', \mathbf{j}') \in \omega_s$, $(x', \mathbf{j}') \in \omega'_N$. For all other $(x, \mathbf{j}) \in \omega_N$, $(x, \mathbf{j}) \in \omega'_N$.

Hence, with probability larger than $\beta(\epsilon, m, N)$, ω'_N contains ω_s , which implies $\gamma^*(\omega'_N) = \gamma^*(\omega_s) = \gamma^o$. Moreover, by construction, for $i = 1, \dots, N$, $x'_i \in \mathbb{S}$ and $\|x_i - x'_i\| \leq \sqrt{2 - 2\delta(\epsilon)}$. $\square$

3.4.2 New bounds for the quadratic case

Using Proposition 3.1, we provide probabilistic upper bounds on γ^o , based on $(\gamma^*(\omega_N), P^*(\omega_N))$, the only solution we have access to. The idea is to obtain a relationship between $\gamma^*(\omega_N)$ and $\gamma^*(\omega'_N)$, relying on the knowledge that the distance separating the initial states of the sets ω'_N and ω_N is bounded.

To prevent repetitions in the statements of our propositions, we summarize here the notation used in this section:

Assumption 3.1. We consider an n -dimensional switched linear system as in (2.1), whose set of m modes is denoted by $\mathcal{M} \subset \mathbb{R}^{n \times n}$; and a uniform random sampling $\omega_N \in Z_l^N$, where $N \geq d_{\text{quad}} + 1$. We denote by $(\gamma^*(\omega_N), P^*(\omega_N))$ the optimal solution to Problem (3.2) and by (γ^o, P^o) the optimal solution to Problem (2.3).

Proposition 3.2. *Let Assumption 3.1 hold. For any $\beta \in [0, 1)$, let $\epsilon = m^l \left(1 - \sqrt[N]{\frac{1-\beta}{d_{\text{quad}}+1}}\right)$. Then, with probability at least β , the following holds:*

$$\gamma^o \leq \sqrt[l]{\gamma^*(\omega_N)^l + (\gamma^*(\omega_N)^l + B(\mathcal{M}^l))\Delta(\epsilon)\kappa(P^*(\omega_N))}, \quad (3.8)$$

where $\Delta(\epsilon) = \sqrt{2 - 2\delta(\epsilon)}$, $\delta(\cdot)$ is given by (3.7), $B(\mathcal{M}^l) = \max_{\mathbf{A} \in \mathcal{M}^l} \|\mathbf{A}\|$ and $\kappa(P) = \sqrt{\frac{\lambda_{\max}(P)}{\lambda_{\min}(P)}}$.

Proof. Let $\gamma = \gamma^*(\omega_N)$ and $P = P^*(\omega_N)$. By definition,

$$\|\mathbf{A}_{\mathbf{j}}x\|_P \leq \gamma^l \|x\|_P \quad \forall (x, \mathbf{j}) \in \omega_N. \quad (3.9)$$

Consider the Cholesky decomposition $P = L^T L$. Then [GVL13]:

$$\|x\|_P = \sqrt{x^T L^T L x} = \|Lx\| \text{ and } \|L^{-1}x\|_P = \|x\|. \quad (3.10)$$

Since $\|A\|_P = \|LAL^{-1}\|$, $\|L^{-1}\|_P = \|L^{-1}\|$. Hence,

$$\|L\| \|x\| \stackrel{(3.10)}{\leq} \|L\| \|L^{-1}\| \|x\|_P = \kappa(P) \|x\|_P. \quad (3.11)$$

Finally,

$$\|x\|_P = \|(x + \Delta x) - \Delta x\|_P \leq \|x + \Delta x\|_P + \|\Delta x\|_P. \quad (3.12)$$

So, for any Δx such that $\|\Delta x\| \leq \Delta(\epsilon)$ and $x + \Delta x \in \mathbb{S}$, we obtain that $\forall (x, \mathbf{j}) \in \omega_N$

$$\begin{aligned} \|\mathbf{A}_{\mathbf{j}}(x + \Delta x)\|_P &\leq \|\mathbf{A}_{\mathbf{j}}x\|_P + \|\mathbf{A}_{\mathbf{j}}\Delta x\|_P \\ &\stackrel{(3.9), (3.10)}{\leq} \gamma^l \|x\|_P + \|L\mathbf{A}_{\mathbf{j}}\Delta x\| \\ &\stackrel{(3.12), (3.10)}{\leq} \gamma^l \|x + \Delta x\|_P + (\gamma^l + \|\mathbf{A}_{\mathbf{j}}\|) \|L\| \|\Delta x\| \\ &\stackrel{\|x+\Delta x\|=1}{\leq} \gamma^l \|x + \Delta x\|_P + (\gamma^l + \|\mathbf{A}_{\mathbf{j}}\|) \Delta(\epsilon) \|L\| \|x + \Delta x\| \\ &\stackrel{(3.11)}{\leq} \left(\gamma^l + (\gamma^l + B(\mathcal{M}^l))\Delta(\epsilon)\kappa(P) \right) \|x + \Delta x\|_P. \end{aligned}$$

Let $\Delta(\epsilon) = \sqrt{2 - 2\delta(\epsilon)}$, $\beta \in [0, 1)$, and $\epsilon = m^l \left(1 - \sqrt[N]{\frac{1-\beta}{d_{\text{quad}}+1}}\right)$. Then, Proposition 3.1 guarantees the existence of a set $\omega'_N = \{(x_i + \Delta x_i, \mathbf{j}_i), i = 1, \dots, N\}$ such that for $i = 1, \dots, N$, $\Delta x_i \leq \Delta(\epsilon)$ and $x_i + \Delta x_i \in \mathbb{S}$, and such that $\gamma^*(\omega'_N) = \gamma^o$ with probability at least β .

Hence, $\|\mathbf{A}_{\mathbf{j}}x\|_P \leq \left(\gamma^l + (\gamma^l + B(\mathcal{M}^l))\Delta(\epsilon)\kappa(P) \right) \|x\|_P \quad \forall (x, \mathbf{j}) \in \omega'_N$. From the definition of $\gamma^*(\omega'_N)$, $(\gamma^*(\omega'_N))^l \leq \left(\gamma^l + (\gamma^l + B(\mathcal{M}^l))\Delta(\epsilon)\kappa(P) \right)$, which concludes the proof. $\square$

One cannot directly use the probabilistic bound of Proposition 3.2. Indeed, it contains the expression $B(\mathcal{M}^l) = \max_{\mathbf{A} \in \mathcal{M}^l} \|\mathbf{A}\|$. This quantity can be computed as [Jun09, Prop. 2.7]:

$$B(\mathcal{M}^l) = \min_{\lambda \geq 0} \lambda, \text{ s.t. } \|\mathbf{A}x\| \leq \lambda, \forall x \in \mathbb{S}, \forall \mathbf{A} \in \mathcal{M}^l. \quad (3.13)$$

However, we have no direct knowledge of $\mathbf{A} \in \mathcal{M}^l$ and hence cannot solve Problem (3.13). We thus seek to obtain a probabilistic bound on $B(\mathcal{M}^l)$ by solving the sampled problem

$$\min_{\lambda \geq 0} \lambda, \text{ s.t. } \|\mathbf{A}_{\mathbf{j}}x\| \leq \lambda, \forall (x, \mathbf{j}) \in \omega_N, \quad (3.14)$$

which only requires the knowledge of the observations we have access to. Let $\lambda^*(\omega_N)$ denote the solution to Problem (3.14). By combining classical results from chance-constrained optimization and the approach proposed in [KBJT19] and [BJW21], we obtain the following bound:

Proposition 3.3. *Let Assumption 3.1 hold, and let $\lambda^*(\omega_N)$ be the optimal solution to Problem (3.14). Then, for any given $\beta_1 \in [0, 1)$, with probability at least β_1 , we have:*

$$\max_{\mathbf{A} \in \mathcal{M}^l} \|\mathbf{A}\| \leq \frac{\lambda^*(\omega_N)}{\delta(\epsilon_1(\beta_1, m, N))} \quad (3.15)$$

where $\delta(\cdot)$ is given by (3.7) and $\epsilon_1(\beta_1, m, N) = \frac{m^l}{2}(1 - \sqrt[N]{1 - \beta_1})$.

Proof. Problem (3.14) is convex and admits a unique solution. Moreover, there exists a sample $(x, \mathbf{j}) \in \omega_N$ such that $\lambda^*(\omega_N) = \|\mathbf{A}_{\mathbf{j}}x\|$, that is $(x, \mathbf{j}) = \underset{(x, \mathbf{j}) \in \omega_N}{\operatorname{argmax}} \|\mathbf{A}_{\mathbf{j}}x\|$, so that one is a bound on the cardinality of the essential set of Problem (3.14). Let $V(\omega_N)$ be the set $\{(x, \mathbf{j}) \in Z_l : \|\mathbf{A}_{\mathbf{j}}x\| > \lambda^*(\omega_N)\}$ of violated constraints. We seek to bound the probability to sample ω_N such that the measure of this set is larger than ϵ_1 . We consider two cases, depending on whether or not the problem is degenerate.

1. Suppose the problem is non-degenerate with probability one. Then, Corollary 2.1 directly ensures that for all $\epsilon_1 \in (0, 1]$, we have:

$$\mu_l^N \{\omega_N \in Z_l^N : \mu_l(V(\omega_N)) \leq \epsilon_1\} \geq \beta_1(\epsilon_1, N)$$

where $\beta_1(\epsilon_1, N) = 1 - (1 - \epsilon_1)^N$.

2. Suppose the problem is degenerate with probability larger than 0. In this case, there exists at least one matrix $\mathbf{A} \in \mathcal{M}^l$ such that $\mathbf{A}^T \mathbf{A} = \gamma I_n$ for some $\gamma \in \mathbb{R}_{\geq 0}$ (see Appendix C). Let $B(\mathcal{M}^l) = \lambda^o = \|\mathbf{A}^*\|$ be the solution to Problem (3.13). We consider the two following cases:

- $\mathbf{A}^{*T}\mathbf{A}^* = \lambda^o I_n$, implying $\|\mathbf{A}^*x\| = \lambda^o \forall x \in \mathbb{S}$. Then, it holds that if for some $(x, \mathbf{j}) \in \omega_N$, $\mathbf{A}^* = \mathbf{A}_{\mathbf{j}}$, $\lambda^*(\omega_N) = \lambda^o$, so that $\mu_l(V(\omega_N))$ can be positive only if no such $\mathbf{j}$ belongs to ω_N , which happens with probability $(1 - \frac{1}{m})^N$. Hence,

$$\mu_l^N\{\omega_N \in Z_l^N : \mu_l(V(\omega_N)) \geq \epsilon_1\} \leq (1 - \frac{1}{m})^N \leq (1 - \epsilon_1)^N$$

when ϵ_1 is smaller than $\frac{1}{m}$. Then, $\mu_l^N\{\omega_N \in Z_l^N : \mu_l(V(\omega_N)) \leq \epsilon_1\} \geq \beta_1(\epsilon_1, N)$, where $\beta_1(\epsilon_1, N) = 1 - (1 - \epsilon_1)^N$.

- $\mathbf{A}^{*T}\mathbf{A}^* \neq \lambda^o I_n$, so that $\mu_l(V(\omega_N))$ can be positive even though $\mathbf{j}$ such that $\mathbf{A}_{\mathbf{j}} = \mathbf{A}^*$ belongs to ω_N . Let $V'(\lambda) = \mu_l(\{(x, \mathbf{j}) \in Z_l : \|\mathbf{A}_{\mathbf{j}}x\| > \lambda\})$. Then, $1 - V'(\lambda)$ is the cumulative distribution function of the random variable $\|\mathbf{A}_{\mathbf{j}}x\|$. Consequently, $V'(\lambda)$ is a right continuous and non increasing function of λ . Hence, for all $\epsilon_1 \in (0, 1)$, there exists an $\lambda(\epsilon_1) \in \mathbb{R}_{\geq 0}$ satisfying $V'(\lambda(\epsilon_1)) = \epsilon_1$, where $\lambda(\epsilon_1)$ is a decreasing function of ϵ_1 . In addition, for $\lambda^*(\omega_N)$ to be smaller than $\lambda(\epsilon_1)$, it is necessary that ω_N contains no $(x, \mathbf{j}) \in \{(x, \mathbf{j}) \in Z_l : \|\mathbf{A}_{\mathbf{j}}x\| > \lambda(\epsilon_1)\}$, which occurs with probability $(1 - \epsilon_1)^N$. Hence,

$$\begin{aligned} \mu_l^N\{\omega_N \in Z_l^N : \mu_l(V(\omega_N)) \geq \epsilon_1\} &= \mu_l^N\{\omega_N \in Z_l^N : \lambda^*(\omega_N) \leq \lambda(\epsilon_1)\} \\ &\leq (1 - \epsilon_1)^N \end{aligned}$$

and $\mu_l^N\{\omega_N \in Z_l^N : \mu_l(V(\omega_N)) \leq \epsilon_1\} \geq \beta_1(\epsilon_1, N)$ where $\beta_1(\epsilon_1, N) = 1 - (1 - \epsilon_1)^N$.

In each case, we obtain that $\mu_l^N\{\omega_N \in Z_l^N : \mu_l(V(\omega_N)) \leq \epsilon_1\} \geq \beta_1(\epsilon_1, N)$ where $\beta_1(\epsilon_1, N) = 1 - (1 - \epsilon_1)^N$. This result can replace Theorem 4 in [KBJT19]. Then, the proof follows the same lines as the derivation of Theorem 6 in [KBJT19], so that details are omitted here. Note that $\epsilon_1(\beta_1, m, N) = \frac{m^l}{2}(1 - \sqrt[l]{1 - \beta_1})$ is smaller than $\frac{1}{m}$ as soon as $\beta_1 \leq 1 - (1 - \frac{2}{m^l})^N$, which holds for almost all values of β_1 as soon as N is sufficiently high. $\square$

Using sensitivity analysis, we obtain another bound on $B(\mathcal{M}^l)$:

Proposition 3.4. *Let Assumption 3.1 hold and let $\lambda^*(\omega_N) > 0$ be the optimal solution to Problem (3.14). Then, with the same notation as in Proposition 3.3, for any given $\beta_1 \in [0, 1)$, with probability at least β_1 , the following holds:*

$$\max_{\mathbf{A} \in \mathcal{M}^l} \|\mathbf{A}\| \leq \frac{\lambda^*(\omega_N)}{1 - \sqrt{2 - 2\delta(\epsilon_1)}}. \quad (3.16)$$

Proof. Let $\text{Max}(\omega_N)$ denote Problem (3.14). We replace Lemma 3.3 by the following observation concerning Problem (3.13). Consider the trajectory $(x, \mathbf{j})$ such that $\max_{\mathbf{A} \in \mathcal{M}^l} \|\mathbf{A}\| = \|\mathbf{A}_j x\| = \|\mathbf{A}_j(-x)\|$. Then, the two sets $\omega_1 = \{(x, \mathbf{j})\}$ and $\omega'_1 = \{(-x, \mathbf{j})\}$ are such that $\lambda^*(\omega_1) = \lambda^*(\omega'_1) = B(\mathcal{M}^l)$, where $\lambda^*(\omega_1)$ and $\lambda^*(\omega'_1)$ are the optimal solutions to $\text{Max}(\omega_1)$ and $\text{Max}(\omega'_1)$.

Following the lines of Proposition 3.1, we can also prove, for all sets $\omega_N \in Z_i^N$, the existence of a set ω'_N , holding with probability at least $\beta_1(\epsilon_1, m, N) = 1 - (1 - \frac{2\epsilon_1}{m^l})^N$ and satisfying two properties. First, $\lambda^*(\omega'_N) = B(\mathcal{M}^l)$, where $\lambda^*(\omega'_N)$ is the optimal solution to $\text{Max}(\omega'_N)$. Second, for all $(x'_i, \mathbf{j}_i) \in \omega'_N$, $x'_i \in \mathbb{S}$ and there exists a trajectory $(x_i, \mathbf{j}_i) \in \omega_N$ such that $\|x_i - x'_i\| \leq \sqrt{2 - 2\delta(\epsilon_1)}$, where $\delta(\cdot)$ is defined in (3.7). The factor 2 in the expression of $\beta_1(\epsilon_1, m, N)$ follows from the observation that there are two distinct sets whose solution is equal to $B(\mathcal{M}^l)$. We can also

express ϵ_1 as a function of β_1 : $\epsilon_1(\beta_1, m, N) = \frac{m^l(1 - \sqrt[N]{1 - \beta_1})}{2}$.

By definition, $\lambda^*(\omega_N)$ is the minimal λ such that $\|\mathbf{A}_j x\| \leq \lambda \forall (x, \mathbf{j}) \in \omega_N$. Then, $\forall (x, \mathbf{j}) \in \omega_N$, $\|\mathbf{A}_j(x + \Delta x)\| \leq \lambda + \max_{\mathbf{A} \in \mathcal{M}^l} \|\mathbf{A}\| \sqrt{2 - 2\delta(\epsilon_1)}$ for any Δx such that $\|\Delta x\| \leq \sqrt{2 - 2\delta(\epsilon_1)}$. Hence, if $\epsilon_1 = \frac{m^l(1 - \sqrt[N]{1 - \beta_1})}{2}$, we obtain that with probability at least β_1 :

$$\max_{\mathbf{A} \in \mathcal{M}^l} \|\mathbf{A}\| \leq \lambda + \max_{\mathbf{A} \in \mathcal{M}^l} \|\mathbf{A}\| \sqrt{2 - 2\delta(\epsilon)} \iff \max_{\mathbf{A} \in \mathcal{M}^l} \|\mathbf{A}\| \leq \frac{\lambda^*(\omega_N)}{1 - \sqrt{2 - 2\delta(\epsilon_1)}}$$

□

Remark 3.1. The bound of Proposition 3.4 is less tight than the bound of Proposition 3.3 as long as $\delta(\epsilon_1(\beta_1, m, N))$ is larger than 0. However, $\delta(\cdot)$ is defined over $(0, \frac{1}{2})$, and if ϵ_1 is larger than $\frac{1}{2}$, $\delta(\epsilon_1) = 0$. Hence, as soon as $\frac{m^l}{2}(1 - \sqrt[N]{1 - \beta_1}) \geq \frac{1}{2}$, which occurs for a large trace and/or number of modes, or when the number of samples is too small, the bound of Proposition 3.3 becomes infinite while one can keep using the bound of Proposition 3.4.

We now propose our first bound on the JSR, by combining Propositions 3.2 and 3.3.

Theorem 3.1

Let Assumption 3.1 hold and let $\lambda^*(\omega_N)$ be the optimal solution to Problem (3.14). For any $\beta \in [0, 1)$ and $\beta_1 \in [0, 1)$, let $\epsilon = m^l \left(1 - \sqrt[N]{\frac{1 - \beta}{d_{\text{quad}} + 1}}\right)$, and

$\epsilon_1 = \frac{m^l}{2}(1 - \sqrt[l]{1 - \beta_1})$. Then, with probability at least $\beta + \beta_1 - 1$, we have:

$$\rho(\mathcal{M}) \leq \sqrt[l]{\gamma^*(\omega_N)^l + (\gamma^*(\omega_N)^l + A(\epsilon_1))\Delta(\epsilon)\kappa(P^*(\omega_N))}, \quad (3.17)$$

where $\Delta(\epsilon) = \sqrt{2 - 2\delta(\epsilon)}$, $\delta(\cdot)$ is given by (3.7) and $A(\epsilon_1) = \frac{\lambda^*(\omega_N)}{\delta(\epsilon_1)}$.

Proof. By Proposition 3.2, $\mu_l^N\{\omega_N \in Z_l^N \mid (3.8) \text{ holds}\} \geq \beta$, and by Proposition 3.3, $\mu_l^N\{\omega_N \in Z_l^N \mid (3.15) \text{ holds}\} \geq \beta_1$. Hence,

$$\begin{aligned} & \mu_l^N\{\omega_N \in Z_l^N \mid \text{at least one of (3.8) or (3.15) does not hold}\} \\ & \leq \mu_l^N\{\omega_N \in Z_l^N \mid (3.8) \text{ does not hold}\} + \mu_l^N\{\omega_N \in Z_l^N \mid (3.15) \text{ does not hold}\} \\ & \leq (1 - \beta) + (1 - \beta_1). \end{aligned}$$

Therefore, $\mu_l^N\{\omega_N \in Z_l^N \mid (3.8) \text{ and (3.15) hold}\} \geq 1 - (1 - \beta) - (1 - \beta_1)$. If ω_N is such that (3.8) and (3.15) hold, then (3.17) holds. Hence, this relation holds with probability at least $\beta + \beta_1 - 1$, which concludes the proof. $\square$

We can infer another bound on the cost of Problem (2.3) as a function of the solution to Problem (3.2) by resorting to the Cauchy inequality [GVL13]. This bound also holds with probability at least $\beta + \beta_1 - 1$.

Theorem 3.2

Let Assumption 3.1 hold. Then, with the same notation as in Theorem 3.1, with probability at least $\beta + \beta_1 - 1$, the following holds:

$$\rho(\mathcal{M}) \leq \sqrt[2l]{\gamma(\omega_N)^{*2l} + (A(\epsilon_1)^2(2\Delta(\epsilon) + \Delta(\epsilon)^2) + \gamma(\omega_N)^{*2l}(2\Delta(\epsilon) - \Delta(\epsilon)^2))\kappa(P^*(\omega_N))^2}.$$

Proof. We first prove that $\gamma^o \leq$

$$\sqrt[2l]{\gamma(\omega_N)^{*2l} + (B(\mathcal{M}^l)^2(\Delta(\epsilon)^2 + 2\Delta(\epsilon)) + \gamma(\omega_N)^{*2l}(-\Delta(\epsilon)^2 + 2\Delta(\epsilon)))\kappa(P^*(\omega_N))^2},$$

where $B(\mathcal{M}^l) = \max_{\mathbf{A} \in \mathcal{M}^l} \|\mathbf{A}\|$.

Let $\gamma = \gamma^*(\omega_N)$ and $P = P^*(\omega_N)$. By definition,

$$\|\mathbf{A}_{\mathbf{j}}x\|_P^2 \leq \gamma^{2l}\|x\|_P^2 \quad \forall (x, \mathbf{j}) \in \omega_N. \quad (3.18)$$

Let L be the Cholesky decomposition of P . It holds that $\langle x, y \rangle_P = x^T P y = \langle Lx, Ly \rangle$ for $x, y \in \mathbb{R}^n$. Then, for any Δx such that $\|\Delta x\| \leq \Delta(\epsilon)$ and $x + \Delta x \in \mathbb{S}$, we

obtain that $\forall (x, \mathbf{j}) \in \omega_N$

$$\begin{aligned}
& \|\mathbf{A}_{\mathbf{j}}(x + \Delta x)\|_P^2 = \|\mathbf{A}_{\mathbf{j}}x\|_P^2 + \|\mathbf{A}_{\mathbf{j}}\Delta x\|_P^2 + 2\langle \mathbf{A}_{\mathbf{j}}x, \mathbf{A}_{\mathbf{j}}\Delta x \rangle_P \\
& \stackrel{(3.18)}{\leq} \gamma^{2l}\|x\|_P^2 + \|\mathbf{A}_{\mathbf{j}}\Delta x\|_P^2 + 2\langle \mathbf{A}_{\mathbf{j}}x, \mathbf{A}_{\mathbf{j}}\Delta x \rangle_P \\
& = \gamma^{2l}(\|x + \Delta x\|_P^2 + \|\Delta x\|_P^2 - 2\langle x + \Delta x, \Delta x \rangle_P) + \|\mathbf{A}_{\mathbf{j}}\Delta x\|_P^2 + 2\langle \mathbf{A}_{\mathbf{j}}x, \mathbf{A}_{\mathbf{j}}\Delta x \rangle_P \\
& = \gamma^{2l}\|x + \Delta x\|_P^2 - \gamma^{2l}\|\Delta x\|_P^2 - 2\gamma^{2l}\langle Lx, L\Delta x \rangle + \|\mathbf{A}_{\mathbf{j}}\Delta x\|_P^2 + 2\langle L\mathbf{A}_{\mathbf{j}}x, L\mathbf{A}_{\mathbf{j}}\Delta x \rangle \\
& \stackrel{(3.10)}{\leq} \gamma^{2l}\|x + \Delta x\|_P^2 + (\|\mathbf{A}_{\mathbf{j}}\|^2 - \gamma^{2l})\|L\Delta x\|^2 \\
& \quad + 2\gamma^{2l}|\langle Lx, L\Delta x \rangle| + 2\|L\mathbf{A}_{\mathbf{j}}x\|\|L\mathbf{A}_{\mathbf{j}}\Delta x\| \\
& \leq \gamma^{2l}\|x + \Delta x\|_P^2 + (\|\mathbf{A}_{\mathbf{j}}\|^2 - \gamma^{2l})\|\Delta x\|^2\|L\|^2 \\
& \quad + 2(\|\mathbf{A}_{\mathbf{j}}\|^2 + \gamma^{2l})\|x\|\|\Delta x\|\|L\|^2 \\
& \stackrel{(3.11)}{\leq} \left(\gamma^{2l} + \left((B(\mathcal{M}^l)^2 - \gamma^{2l})\Delta(\epsilon)^2 + 2(B(\mathcal{M}^l)^2 + \gamma^{2l})\Delta(\epsilon) \right) \kappa(P)^2 \right) \|x + \Delta x\|_P^2.
\end{aligned}$$

Let $\Delta(\epsilon) = \sqrt{2 - 2\delta(\epsilon)}$, $\beta \in [0, 1)$, and $\epsilon(\beta, m, N) = m^l \left(1 - \sqrt[l]{\frac{1-\beta}{d_{\text{quad}}+1}} \right)$. Then, Proposition 3.1 guarantees the existence of a set $\omega'_N = \{(x_i + \Delta x_i, \mathbf{j}_i), i = 1, \dots, N\}$ such that for $i = 1, \dots, N$, $\Delta x_i \leq \Delta(\epsilon)$ and $x_i + \Delta x_i \in \mathbb{S}$, and such that $\gamma^*(\omega'_N) = \gamma^o$ with probability at least β . Hence, $\forall (x, \mathbf{j}) \in \omega'_N$,

$$\|\mathbf{A}_{\mathbf{j}}x\|_P^2 \leq \left(\gamma^{2l} + \left((B(\mathcal{M}^l)^2 - \gamma^{2l})\Delta(\epsilon)^2 + 2(B(\mathcal{M}^l)^2 + \gamma^{2l})\Delta(\epsilon) \right) \kappa(P)^2 \right) \|x\|_P^2.$$

From the definition of $\gamma^*(\omega'_N)$,

$(\gamma^*(\omega'_N))^{2l} \leq \left(\gamma^{2l} + \left((B(\mathcal{M}^l)^2 - \gamma^{2l})\Delta(\epsilon)^2 + 2(B(\mathcal{M}^l)^2 + \gamma^{2l})\Delta(\epsilon) \right) \kappa(P)^2 \right)$, which concludes the first part of the proof. Then, combining the inequality with Proposition 3.3 as in the proof of Theorem 3.1 concludes the proof. $\square$

Remark 3.2. The bound of Theorem 3.2 is advantaged over the one of Theorem 3.1 in two cases. First, it is less affected by an increase in $\Delta(\epsilon)$, especially when the latter is higher than 1. Indeed, the largest term in the bound of Theorem 3.2 is of order $\sqrt{\Delta(\epsilon)}$ instead of $\Delta(\epsilon)$. $\Delta(\epsilon)$ is larger than 1 when a low number of samples is used or for systems composed of a large number of modes. The bound is also tighter when the difference $B(\mathcal{M}^l)^2 - \gamma^{2l}$ is sufficiently high.

For non-degenerate systems, the bounds of Theorems 3.1 and 3.2 can be compared with the previous bounds from [BJW21], stating that with a confidence level of $\bar{\beta} = \beta_1 + \beta - 1$,

$$\rho(\mathcal{M}) \leq \frac{\gamma^*(\omega_N)}{\sqrt[l]{\delta\left(\frac{\bar{\epsilon}(\bar{\beta}, N)m^l \bar{\kappa}(P^*(\omega_N))}{2}\right)}} \quad (3.19)$$

and

$$\rho(\mathcal{M}) \leq \frac{\gamma^*(\omega_N)}{\sqrt[l]{\delta\left(\frac{1 - (1 - \bar{\epsilon}(\bar{\beta}, N)m^l) \sqrt{\frac{\det(P^*(\omega_N))}{\lambda_{\max}(P^*(\omega_N))^n}}}{2}\right)}}},$$

where $P^*(\omega_N)$, $\gamma^*(\omega_N)$ and $\delta(\cdot)$ are defined as in Theorem 3.1; $\bar{\epsilon}(\bar{\beta}, N) = 1 - I^{-1}(1 - \bar{\beta}, N - d_{\text{quad}} - 1, d_{\text{quad}})$, and $\bar{\kappa}(P) = \sqrt{\frac{\det(P)}{\lambda_{\min}(P)^n}}$.

Remark 3.3. Again, because $\delta(\epsilon) = 0$ when $\epsilon \geq \frac{1}{2}$, our upper bounds present an advantage over the bound in [BJW21]: in the latter, especially when the number of modes m and the length of the traces l increase, a large number of samples N is required to provide a finite upper bound on the JSR, while our bounds are finite in all cases. Indeed, if $\epsilon(\beta, m, N) \geq \frac{1}{2}$ so that $\delta(\epsilon(\beta, m, N)) = 0$, the bounds we provide are nevertheless finite. If $\epsilon_1(\beta_1, m, N) \geq \frac{1}{2}$, we can replace the expression $A(\epsilon_1) = \frac{\lambda^*(\omega_N)}{\delta(\epsilon_1)}$ by the expression provided in Proposition 3.4 (see Remark 3.1) and obtain finite bounds. This is not the case for the bound of [BJW21], which requires for instance $N = 500$ samples to be finite for $m = 5$, $n = 2$ and $l = 3$.

3.5 Probabilistic SOS-based guarantees

We now leverage our new approach to the SOS framework.

The main differences between Problems (2.3) and (2.9) are (i) the increase in the dimension of the optimization problem, from $d_{\text{quad}} + 1$ to $d_{\text{sos}} + 1$, where $d_{\text{sos}} = \frac{D(D+1)}{2}$ and D is given by (2.6); and (ii) the distribution of the vectors $x^{[d]}$, which are not uniformly sampled on the sphere of dimension D when the vectors x are uniformly sampled on $\mathbb{S}$. The latter observation explains why the technique of [BJW21] cannot be leveraged to the SOS setting directly. It however does not affect our sensitivity analysis approach: Lemma 3.3 is straightforwardly extended with the increased dimension taken into account; and we can then apply Proposition 3.1 to Problem (2.9), with only the expression of the confidence level modified:

Proposition 3.5. *Let $\omega_N := \{(x_i, \mathbf{j}_i), i = 1, \dots, N\}$ be a uniform random sample drawn from Z_l , with $N \geq d_{\text{sos}} + 1$, and γ^o be the optimal solution to Problem (2.9). For all $\epsilon_{\text{sos}} \in (0, 1]$, with probability at least*

$$\beta(\epsilon_{\text{sos}}, m, N) = 1 - (d_{\text{sos}} + 1) \left(1 - \frac{\epsilon_{\text{sos}}}{m^l}\right)^N,$$

there exists a set $\omega'_N = \{(x'_i, \mathbf{j}_i), i = 1, \dots, N\}$ such that

- $\gamma_{\text{sos}}^*(\omega'_N) = \gamma^o$, where $\gamma_{\text{sos}}^*(\omega'_N)$ is the optimal solution to $\text{Opt}_{\text{sos}}(\omega'_N)$.
- For $i = 1, \dots, N$: $x'_i \in \mathbb{S}$ and $\|x_i - x'_i\| \leq \sqrt{2 - 2\delta(\epsilon)}$, where $\delta(\cdot)$ is given by (3.7).

Again, to prevent excessively long statements in our next theorems, we present here the notation used in this section:

Assumption 3.2. We consider an n -dimensional switched linear system as in (2.1), whose set of m modes is denoted by $\mathcal{M} \subset \mathbb{R}^{n \times n}$, and approximate SOS Lyapunov functions of degree $2d$. Let $D = \binom{n+d-1}{d}$. We consider a uniform random sampling $\omega_N \in Z_l^N$, where $N \geq d_{\text{sos}} + 1$. We denote by $(\gamma_{\text{sos}}^*(\omega_N), P_{\text{sos}}^*(\omega_N))$ the solution to Problem (3.3), by $(\gamma_{\text{sos}}^o, P_{\text{sos}}^o)$ the optimal solution to Problem (2.9), and by $\lambda^*(\omega_N)$ the optimal solution to Problem (3.14).

We now combine Propositions 3.5 and 3.3 to obtain a SOS-based probabilistic upper bound on $\rho(\mathcal{M})$. Note that we only extend the bound of Theorem 3.1. Indeed, the argument leading to the bound of Theorem 3.2 takes advantage of the relation $\|x + y\|^2 = \|x\|^2 + \|y\|^2 + 2\langle x, y \rangle$. In the SOS setting, we deal with norms of sums containing more than 2 terms, which prevents the bound we could obtain following this argument to be tight.

Theorem 3.3

Let Assumption 3.2 hold. For any $\beta \in [0, 1)$ and $\beta_1 \in [0, 1)$, let $\epsilon_{\text{sos}} = m^l \left(1 - \sqrt[n]{\frac{1-\beta}{d_{\text{sos}}+1}}\right)$ and $\epsilon_1 = \frac{m^l}{2} \left(1 - \sqrt[n]{1-\beta_1}\right)$. Then, with probability at least $\beta + \beta_1 - 1$, we have:

$$\rho(\mathcal{M}) \leq \sqrt[d]{\gamma(\omega_N)^{*dl} + ((\gamma(\omega_N)^{*dl} + A(\epsilon_1)^d)f(d, \epsilon_{\text{sos}}))\kappa(P^*(\omega_N))}$$

where $A(\epsilon_1) = \frac{\lambda^*(\omega_N)}{\delta(\epsilon_1)}$, $f(d, \epsilon_{\text{sos}}) = \sqrt{D}((1 + \Delta(\epsilon_{\text{sos}}))^d - 1 - (1 - \frac{1}{\sqrt{D}})\Delta(\epsilon_{\text{sos}})^d)$, $\Delta(\epsilon_{\text{sos}}) = \sqrt{2 - 2\delta(\epsilon_{\text{sos}})}$ and $\delta(\cdot)$ is given by (3.7).

When $d = 1$, this theorem reduces to Theorem 3.1 as expected. Before proceeding to the proof itself, additional properties concerning d-lifts and Kronecker products are needed.

Proposition 3.6. *There exists a matrix $C_d \in \mathbb{R}^{D \times n^d}$ of norm $\|C_d\| \leq \sqrt{D}$ such that $x^{[d]} = C_d x^{\otimes d}$, where $x^{[d]}$ is the d -lift of x . Moreover, the following relations hold:*

$$\|x^{\otimes d}\| = \|x\|^d \tag{3.20}$$

$$(x + y)^{\otimes d} = (x + y) \otimes \dots \otimes (x + y) = \sum_{k=0}^d K_{xy}(d, k) = x^{\otimes d} + y^{\otimes d} + \sum_{k=1}^{d-1} K_{xy}(d, k) \tag{3.21}$$

where $K_{xy}(d, k)$ is the sum of all possible sequences of d Kronecker products composed of $d - k$ copies of x and k copies of y . For $k = 0, \dots, d$, $K_{xy}(d, k)$ is composed of $\binom{d}{k}$ terms of norm $\|x^{\otimes(d-k)} \otimes y^{\otimes k}\| = \|x\|^{(d-k)} \|y\|^k$.

Proof. The two last relations are well-known [LF72]. In addition, $x^{\otimes d}$ is composed of $\alpha!$ copies of each monomial of degree d x^α . Hence, if the only non-zero entries of α -th row of C_d are $\frac{1}{\sqrt{\alpha!}}$ and lie at columns associated to monomials x^α , $Cx^{\otimes d} = x^d$. Moreover, $\|C_d\|^2 \leq \|C_d\|_F^2 = \sum_\alpha \alpha! \frac{1}{\alpha!} = D$. $\square$

We are now ready to prove Theorem 3.3.

Proof. We recall the binomial theorem [Aig97]: for $n \geq 0$, $x \in \mathbb{R}$,

$$(1+x)^n = \sum_{k=0}^n \binom{n}{k} x^k. \quad (3.22)$$

Let $\gamma = \gamma_{sos}^*(\omega_N)$ and $P = P_{sos}^*(\omega_N)$. By definition,

$$\|(\mathbf{A}_j x)^{[d]}\|_P \leq \gamma^{dl} \|x^{[d]}\|_P, \quad \forall (x, \mathbf{j}) \in \omega_N. \quad (3.23)$$

Let $\psi = \|(x + \Delta x)^{[d]}\|_P$. For any Δx such that $\|\Delta x\| \leq \Delta(\epsilon)$ and $x + \Delta x \in \mathbb{S}$, we obtain that $\forall (x, \mathbf{j}) \in \omega_N$

$$\begin{aligned} & \|(\mathbf{A}_j(x + \Delta x))^{[d]}\|_P \stackrel{Prop. 3.6}{=} \|\mathbf{A}_j^{[d]} C_d (x + \Delta x)^{\otimes d}\|_P \\ & \stackrel{(3.21)}{\leq} \|\mathbf{A}_j^{[d]} C_d x^{\otimes d}\|_P + \|\mathbf{A}_j^{[d]} C_d (\Delta x)^{\otimes d}\|_P + \left\| \sum_{k=1}^{d-1} \mathbf{A}_j^{[d]} C_d K_{x\Delta x}(d, k) \right\|_P \\ & \stackrel{(3.23)}{\leq} \gamma^{dl} \|x^{[d]}\|_P + \|(\mathbf{A}_j \Delta x)^{[d]}\|_P + \sum_{k=1}^{d-1} \|\mathbf{A}_j^{[d]} C_d K_{x\Delta x}(d, k)\|_P \\ & \stackrel{(3.10)}{\leq} \gamma^{dl} \|(x + \Delta x)^{[d]} - \Delta x^{[d]} - \sum_{k=1}^{d-1} C_d K_{x\Delta x}(d, k)\|_P \\ & + \|\mathbf{A}_j \Delta x\|^d \|L\| + \sum_{k=1}^{d-1} \|\mathbf{A}_j^{[d]}\| \|C_d\| \|K_{x\Delta x}(d, k)\| \|L\| \\ & \stackrel{(2.7), Prop. 3.6}{\leq} \gamma^{dl} \psi + (\gamma^{dl} + \|\mathbf{A}_j\|^d) \|\Delta x\|^d \|L\| \\ & + \sum_{k=1}^{d-1} (\gamma^{dl} + \|\mathbf{A}_j\|^d) \sqrt{D} \binom{d}{k} \|x\|^{d-k} \|\Delta x\|^k \|L\| \\ & \stackrel{(3.11), x+\Delta x \in \mathbb{S}}{\leq} \gamma^{dl} \psi + (\gamma^{dl} + \|\mathbf{A}_j\|^d) \Delta(\epsilon)^d \kappa(P) \psi \\ & + \left((\gamma^{dl} + \|\mathbf{A}_j\|^d) \sqrt{D} \sum_{k=1}^{d-1} \binom{d}{k} \Delta(\epsilon)^k \right) \kappa(P) \psi \\ & \stackrel{(3.10)}{\leq} \left(\gamma^{dl} + (\gamma^{dl} + \|\mathbf{A}_j\|^d) \Delta(\epsilon)^d \kappa(P) \right) \psi \\ & + \left((\gamma^{dl} + \|\mathbf{A}_j\|^d) \sqrt{D} \left(\sum_{k=0}^d \binom{d}{k} \Delta(\epsilon)^k - 1 - \Delta(\epsilon)^d \right) \right) \kappa(P) \psi \end{aligned}$$

$$\begin{aligned}
& \stackrel{(3.11), (3.22)}{=} \left(\gamma^{dl} + (\gamma^{dl} + B(\mathcal{M}^l)^d) \Delta(\epsilon)^d \kappa(P) \right) \psi \\
& + \left((\gamma^{dl} + B(\mathcal{M}^l)^d) \sqrt{D} \left((1 + \Delta(\epsilon))^d - 1 - \Delta(\epsilon)^d \right) \right) \kappa(P) \psi \\
& = \left(\gamma^{dl} + (\gamma^{dl} + B(\mathcal{M}^l)^d) \sqrt{D} \left((1 + \Delta(\epsilon))^d - 1 - \left(1 - \frac{1}{\sqrt{D}}\right) \Delta(\epsilon)^d \right) \right) \kappa(P) \psi
\end{aligned}$$

where $B(\mathcal{M}^l) = \max_{\mathbf{A} \in \mathcal{M}^l} \|\mathbf{A}\|$. Let $\beta \in [0, 1)$, $\Delta(\epsilon) = \sqrt{2 - 2\delta(\epsilon_{sos})}$ and $\epsilon_{sos}(\beta, m, N) = m^l \left(1 - \sqrt[N]{\frac{1-\beta}{d_{sos}+1}}\right)$. Proposition 3.5 guarantees the existence of a set $\omega'_N = \{(x_i + \Delta x_i, \mathbf{j}_i), i = 1, \dots, N\}$ such that for $i = 1, \dots, N$, $\Delta x_i \leq \Delta(\epsilon)$ and $x_i + \Delta x_i \in \mathbb{S}$; and such that $\gamma_{sos}^*(\omega'_N) = \gamma_{sos}^o$ with probability at least β . Hence, $\|(\mathbf{A}_{\mathbf{j}} x)^{[d]}\|_P \leq \left(\gamma^{dl} + ((\gamma^{dl} + B(\mathcal{M}^l)^d) \sqrt{D} \left((1 + \Delta(\epsilon))^d - 1 - \left(1 - \frac{1}{\sqrt{D}}\right) \Delta(\epsilon)^d \right) \right) \kappa(P) \|x^{[d]}\|_P$, $\forall (x, \mathbf{j}) \in \omega'_N$. Following the same reasoning as in Theorem 3.1 concludes the proof. $\square$

Similarly to our quadratic bounds, the bound of Theorem 3.3 is always finite.

Chapter 4

Numerical experiments

In this chapter, we conduct several numerical experiments to illustrate the improvement of our quadratic bounds compared to the one of [BJW21], the performance of our SOS-based bound compared to the quadratic bounds, and the accuracy of the lower bounds given in Lemmas 3.1 and 3.2.

We first illustrate the benefits of using approximate SOS Lyapunov functions on a stable system for which no CQLF exists. We then show that for high dimensional systems, our quadratic bounds outperform the one of [BJW21]. In a third time, we evaluate the sensitivity of our bounds with respect to the length of the trace l , the number of modes m of the underlying system, and its dimension n . Finally, we analyze the accuracy of our lower bounds.

For the sake of clarity, we first only plot the bound of Theorem 3.1 in the quadratic case. We then compare it to the bound of Theorem 3.2 in a separate example.

Experimental protocol

Let N denote the number of random trajectories sampled. Then, in each experiment, for every value of m , n , l , d and N analyzed, we solve Problems (3.14), (3.2) in the quadratic case, and (3.3) in the SOS case using N trajectories. For the two last problems, we look for the optimal cost $\gamma^*(\omega_N)$ or $\gamma_{sos}^*(\omega_N)$ by bisection on an interval $[0, U]$, for a reasonable choice of U ($U = \max_{i=1, \dots, N} \|x_{i,l}\|$, for instance) with a fixed accuracy α . We thus solve a feasibility problem at each iteration, that is at most $\lceil \log_2(\frac{U}{\alpha}) \rceil$ times. Once the optimal cost is obtained, we apply the tie-breaking rule on $P^*(\omega_N)$ (or $P_{sos}^*(\omega_N)$) if the latter is not unique. Namely, we obtain a

feasible P of minimal condition number by solving the following problem:

$$\begin{aligned} \min_{P \succeq I_D} \lambda_{\max}(P) \\ \text{s.t. } ((\mathbf{A}_{\mathbf{j}}x)^{[d]})^T P (\mathbf{A}_{\mathbf{j}}x)^{[d]} \leq \gamma^*(\omega_N)^{2d} (x^{[d]})^T P x^{[d]} \quad \forall (x, \mathbf{j}) \in \omega_N, \quad \|P\| \leq C. \end{aligned} \quad (4.1)$$

Then, we compute the bounds of Theorems 3.1, 3.2 and 3.3 as well as the bound of [BJW21]. The average of the bounds is computed over several runs to get rid of the randomness of the experiments. Table 4.1 summarizes the default values of the parameters used in our experiments unless stated otherwise.

Accuracy on the bisection α	10^{-3}
Confidence level β	0.95
Confidence level β_1	0.95
Global confidence level $\beta_1 + \beta - 1$	0.9
Number of runs for each simulation	10
Degree $2d$ of the approximate SOS Lyapunov function	$d = 2$
Number of modes m of the system	$m = 2$
Length l of the traces observed	$l = 1$

Table 4.1: Default values of the parameters used in the experiments of Chapter 4.

4.1 Comparison of the quadratic and SOS-based bounds

In the white-box setting, resorting to SOS Lyapunov functions instead of CQLFs to approximate the JSR of a system provides a significant improvement when the system does not admit a CQLF. Hence, first, to illustrate the improvement of our bounds, especially in the SOS case, compared to the quadratic bound of [BJW21], we consider a stable system for which no CQLF exists:

Example 4.1. ([PJ08, Ex. 2.8])

Consider the 2-modal system of dimension 2 defined by

$$\mathcal{M} = \left\{ \begin{pmatrix} 1 & 0 \\ 1 & 0 \end{pmatrix}, \begin{pmatrix} 0 & 1 \\ 0 & -1 \end{pmatrix} \right\}.$$

The JSR of this set is equal to 1. However, the optimal cost of Problem (2.3) is $\gamma^o = \sqrt{2}$, which is thus the tightest quadratic bound one can obtain on the JSR of the system. On the other hand, by relying on SOS function of degree $2d = 4$, one can obtain an upper bound of $1 + \epsilon$, with ϵ arbitrarily close to 0, so that in the white-box setting, SOS Lyapunov functions outperform CQLFs.

Figure 4.1 presents the evolution of the different upper bounds with the number of samples. We observe that for a large enough number of samples, **the SOS-based bound outperforms the quadratic bound in the black-box setting just like it does in the white-box setting**. We also observe that while the bound of [BJW21] is tighter than our quadratic bound for a sufficiently high number of samples, it is beaten for a small number of samples. As we will notice in Section 3.2, this feature becomes even more advantageous when longer traces are used.

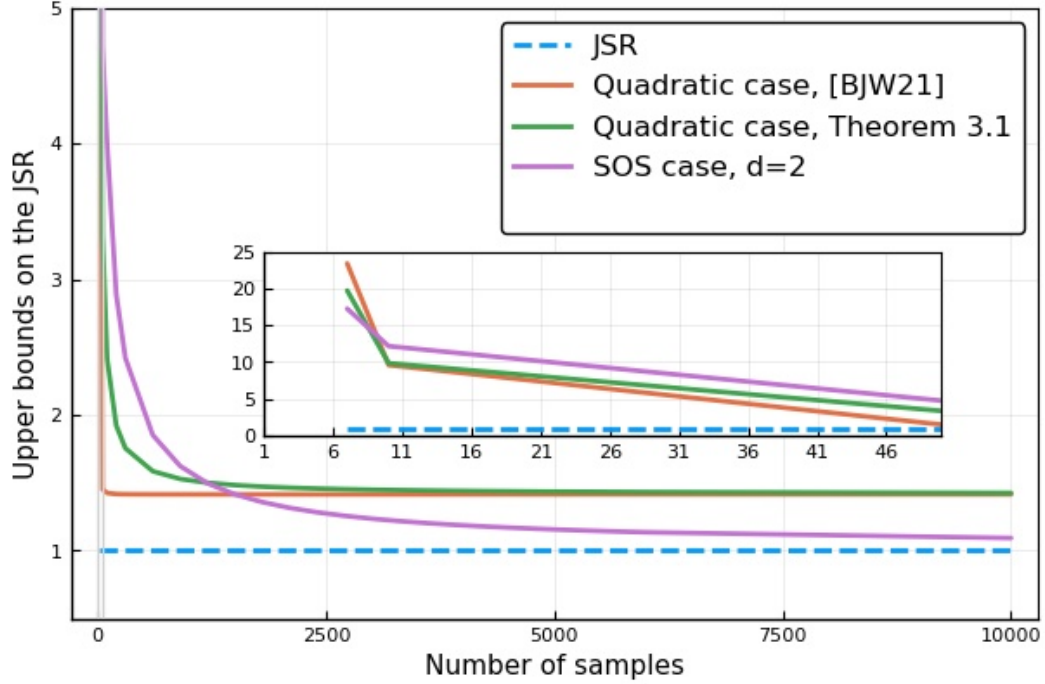


Figure 4.1: Evolution of several upper bounds on the JSR of the system of Example 4.1 with the number of samples. The bounds of Theorem 3.1 (our new quadratic bound), of Theorem 3.3 (our SOS-based bound, for $d = 2$), of [BJW21] and the true value of the JSR are plotted. From 1500 samples, the SOS-based bound becomes the tightest one.

We now illustrate on the 8-dimensional system of Example 2.1 how **our bounds outperform the bound of [BJW21] for high-dimensional systems**. This is due to an asset of our bounds: their robustness with respect to the condition number of $P^*(\omega_N)$. First, the quantity appearing in our bounds is $\kappa(P^*(\omega_N)) = \sqrt{\frac{\lambda_{\max}(P^*(\omega_N))}{\lambda_{\min}(P^*(\omega_N))}}$, which is always smaller than the expression appearing in the bound of [BJW21], that is $\bar{\kappa}(P^*(\omega_N)) = \sqrt{\frac{\det(P^*(\omega_N))}{\lambda_{\min}(P^*(\omega_N))^n}}$. Second, in our bounds, $\kappa(P^*(\omega_N))$

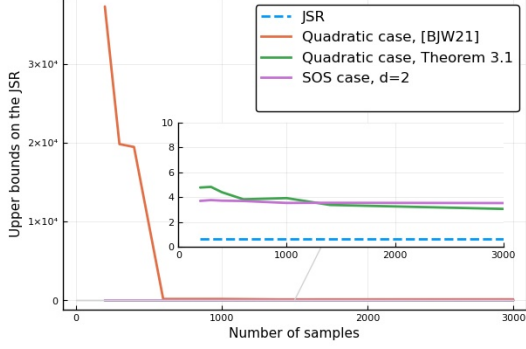


Figure 4.2: Evolution of several upper bounds on the JSR of the system of Example 2.1 with the number of samples, for traces of length 2. The bounds of Theorem 3.1, of Theorem 3.3, of [BJW21] and the true value of the JSR are plotted. As expected for systems of large dimension, our bounds outperform the one of [BJW21].

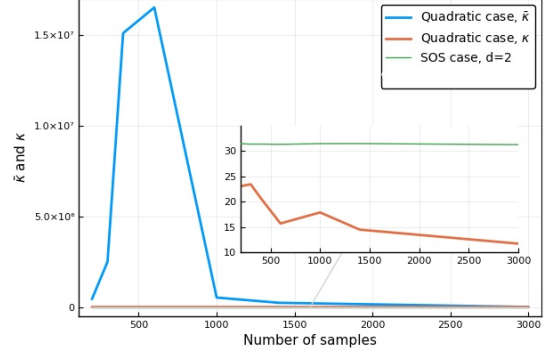


Figure 4.3: Illustration of the key factor explaining Figure 4.2: the evolution of $\kappa(P^*(\omega_N))$ (appearing in our quadratic bound), $\kappa(P_{sos}^*(\omega_N))$ (appearing in our SOS-based bound) and $\bar{\kappa}(P^*(\omega_N))$ (appearing in the bound of [BJW21]) with the number of samples. $\bar{\kappa}(P^*(\omega_N))$ is significantly larger than the other quantities.

is multiplied by $\sqrt{2 - 2\delta(\epsilon)}$, which tends to 0 as N increases. On the contrary, $\bar{\kappa}(P^*(\omega_N))$ appears directly in the expression of $\delta(\cdot)$ in the bound of [BJW21], which causes the bound to become infinite as soon as $\bar{\kappa}(P^*(\omega_N))$ becomes too large. Since both $\kappa(P^*(\omega_N))$ and $\bar{\kappa}(P^*(\omega_N))$ increase rapidly with the dimension of the system, our bounds outperform the one of [BJW21] for high dimensional systems, especially if longer traces are used. We illustrate this property in Figure 4.2, which presents the evolution of the quadratic and SOS-based bounds on the JSR with the number of samples, for traces of length $l = 2$. In this setting, our bounds always outperform the one of [BJW21]. The evolution of $\kappa(P^*(\omega_N))$ and $\bar{\kappa}(P^*(\omega_N))$ with the number of samples, explaining Figure 4.2, is plotted on Figure 4.3.

4.2 Influence of the length of the trace

To illustrate another advantage of our bounds compared to the one of [BJW21], we analyze the impact of the length l of the traces on the upper bounds on the JSR. We consider the system of Example 4.2.

Example 4.2. We consider the following 2-dimensional system composed of 3

modes drawn uniformly, independently and randomly on the sphere:

$$\mathcal{M} = \left\{ \begin{pmatrix} -0.138 & -0.147 \\ 0.446 & 0.874 \end{pmatrix}, \begin{pmatrix} -0.159 & -0.923 \\ -0.288 & -0.297 \end{pmatrix}, \begin{pmatrix} 0.324 & 0.934 \\ 0.461 & -0.297 \end{pmatrix} \right\}.$$

Its JSR is equal to 0.778.

Figures 4.4 and 4.5 present respectively the evolution of the quadratic and SOS-based upper bounds on the JSR with the number of samples, for values of l ranging between 1 and 5. The true value of the JSR approximated by the JSR toolbox [LJP16] is also plotted.

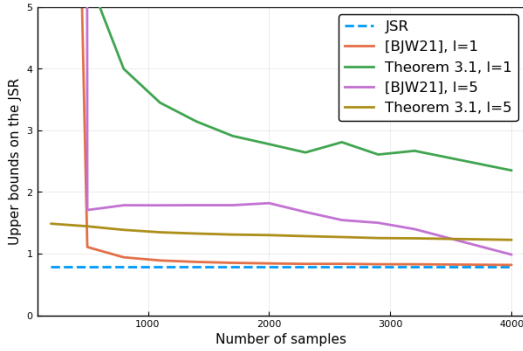


Figure 4.4: Evolution of the quadratic upper bounds on the JSR of the system of Example 4.2 with the number of samples, for traces of various length. To the contrary of the bound of [BJW21], our bound gets tighter when l increases.

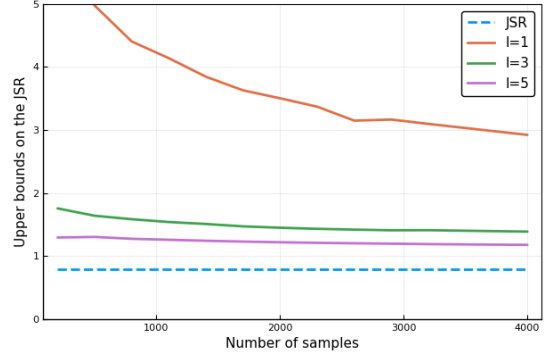


Figure 4.5: Evolution of the SOS-based upper bound, for $d = 2$, on the JSR of the system of Example 4.2 with the number of samples. The length of the traces ranges between 1 and 5. The bound gets tighter when l increases.

We observe that the tightness of the quadratic bound of [BJW21] decreases when the length of the traces increases. However, both in the quadratic and in the SOS setting, our newly proposed bounds do not suffer from this curse of dimensionality: their accuracy increases with the length of the trace. This holds for systems whose modes are drawn on the sphere, and whose JSR is thus relatively small. Indeed, for these systems, the optimal costs of Problems (3.2) and (3.3) are smaller than 5. Then, the bounds of Theorems 3.1 and 3.3 are decreasing in l , especially rapidly when l is small. For systems consisting of modes of larger norm or when longer traces are used, this property does not hold anymore. However, even then, our bounds are less affected by the length of the traces than the bound of [BJW21]. Hence, **our bounds can be used to take advantage of longer trajectories if these are available**. Moreover, we observe as expected that in the quadratic case, our bound outperforms the bound of [BJW21] when a large number of samples

is not available. The number of samples required for the bound of [BJW21] to become the tightest one increases with the length of the traces.

4.3 Influence of the dimension

Our third experiment studies the impact of the number of modes m and of the dimension n of the system on the upper bounds on its JSR. Figures 4.6, 4.7 and 4.8 present the convergence of the quadratic and SOS-based upper bounds when the number of samples increases, for various values of m and n . For each combination of m and n , a random system of m modes of dimension n of unitary Euclidean norm is drawn. The ratios between the probabilistic upper bounds of Theorems 3.1 and 3.3 or of [BJW21] and the true value of the JSR as approximated by the JSR toolbox [LJP16] are plotted.

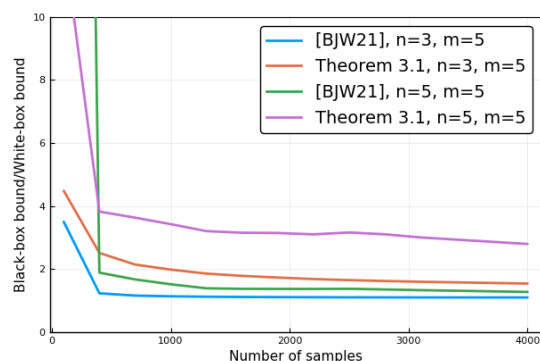


Figure 4.6: Convergence of the quadratic upper bounds on the JSR of 5-modal systems of various dimensions when the number of samples increases. The ratios between the bounds of Theorem 3.1 (our new quadratic bound) or of [BJW21] and the true value of the JSR are plotted. The ratios increase with n .

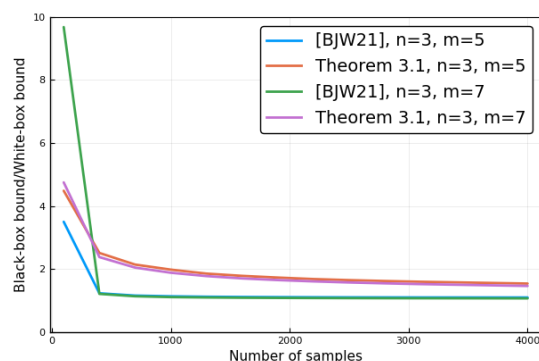


Figure 4.7: Convergence of the quadratic upper bounds on the JSR of 3-dimensional systems composed of 5 or 7 modes when the number of samples increases. The ratios between the bounds of Theorem 3.1 or of [BJW21] and the true value of the JSR are plotted. The ratios slightly increase with m .

In the quadratic case, as expected, our bound outperforms the one of [BJW21] when the number of samples is not sufficiently high, especially when the dimension increases. We observe that in any case, the ratios increase with the number of modes and the dimension of the system. The impact of the dimension is also significantly greater than the impact of the number of modes.

As illustrated on Figure 4.9 in the SOS case, we can explain these observations by the behavior of the condition number of $P_{sos}^*(\omega_N)$ (or $P^*(\omega_N)$). Indeed, it increases

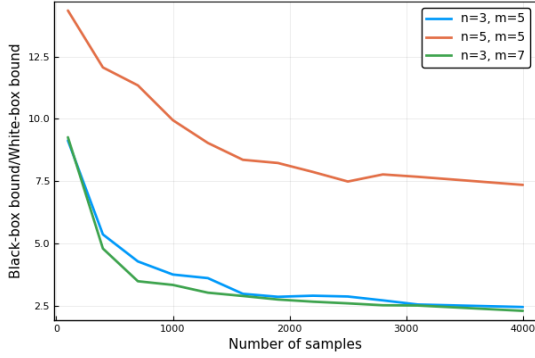


Figure 4.8: Convergence of the SOS-based upper bounds on the JSR of systems of various dimensions when the number of samples increases. The ratio between the bound of Theorem 3.3 and the true value of the JSR is plotted. The ratio increases with n , while the impact of m is negligible.

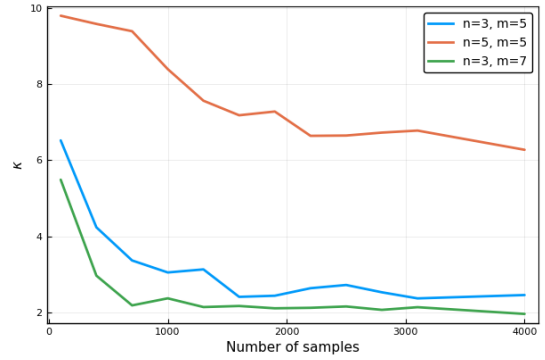


Figure 4.9: Evolution of $\kappa(P_{sos}^*(\omega_N))$ for systems of various dimensions with the number of samples. $\kappa(P_{sos}^*(\omega_N))$ increases with n , while its sensitivity with respect to a variation in m is negligible. The behavior of $\kappa(P_{sos}^*(\omega_N))$ explains Figure 4.8.

with the dimension of the system, while staying more or less equivalent when the number of modes increases. By comparing Figures 4.6, and 4.8, we observe that for larger dimensions, our SOS-based bound is larger than our quadratic bound. This is explained by the increased dimension of the variable P when resorting to approximate SOS Lyapunov functions. Indeed, this rise in the dimension induces an increase in $\kappa(P_{sos}^*(\omega_N))$ compared to $\kappa(P^*(\omega_N))$.

4.4 Comparison of the bounds of Theorems 3.1 and 3.2

Our next experiment consists in the comparison of our two novel quadratic bounds on systems of various dimensions. Figure 4.10 presents the convergence of both bounds on the JSR of n -dimensional systems of unitary norm when the number of samples increases, for various values of n . The ratios between the probabilistic upper bounds of Theorems 3.1 and 3.2 and the true value of the JSR as approximated by the JSR toolbox [LJP16] are plotted.

As expected, **the bound of Theorem 3.2 beats the one of Theorem 3.1 when the number of samples available is low**, a fortiori for low dimensional systems. In this context, it also beats the bound of [BJW21], and can thus be used to obtain the tightest bounds possible when a low number of samples is available.

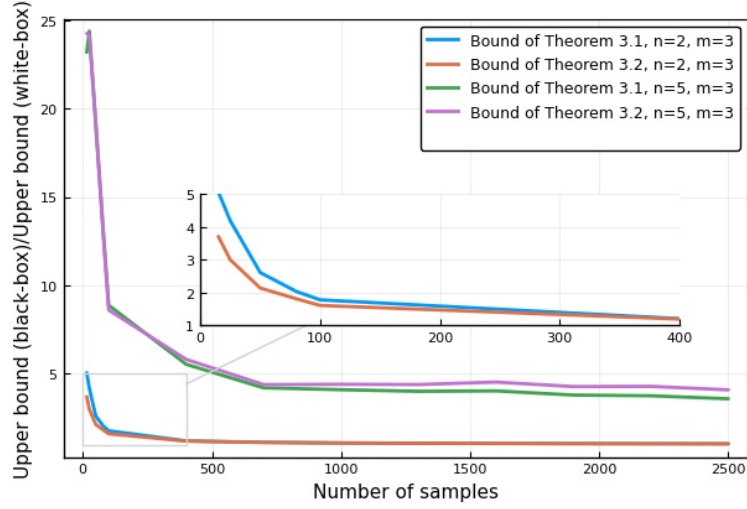


Figure 4.10: Convergence of our quadratic upper bounds on the JSR of systems of various dimensions when the number of samples increases. The ratio between the bounds of Theorem 3.1 or 3.2 and the true value of the JSR is plotted. The bound of Theorem 3.2 is the tightest one when the number of samples used is low.

4.5 Lower bounds on the JSR

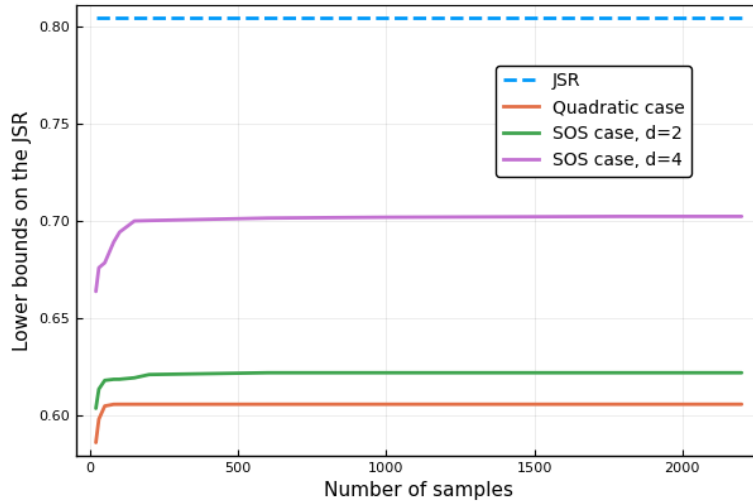


Figure 4.11: Evolution of the lower bounds on the JSR of the system of Example 4.2 with the number of samples. The bounds of Lemma 3.1, in the quadratic case, and of Lemma 3.2, in the SOS case for $d = 2$ and $d = 4$, are plotted. The bounds get tighter when d increases.

Our last experiment illustrates another asset of approximate SOS Lyapunov functions over approximate CQLFs. Indeed, we analyze the behavior of the lower bounds given in Lemmas 3.1 and 3.2 on the system of Example 4.2. More precisely, Figure 4.11 presents the evolution of these bounds, in the quadratic case and in the SOS case with $d = 2$ and $d = 4$, with the number of samples. As expected, the bounds converge to $\frac{\rho(\mathcal{M})}{\sqrt[d]{\min(D, m)}}$. They thus increase with d , and would benefit further from longer traces. Hence, **the gap separating the deterministic lower bound and the probabilistic upper bound on the JSR decreases as d increases.**

4.6 Conclusion of the experiments

In conclusion, we observe that our novel bounds, derived in Theorems 3.1, 3.2 and 3.3, present several improvements in comparison with the bound of [BJW21], both in the quadratic and in the SOS case:

- Our SOS-based bound outperforms the quadratic bounds for systems that do not admit CQLFs, just as it does in the white-box setting.
- In the SOS setting, the gap separating the upper and lower bounds is tighter than in the quadratic setting. The confidence interval for the value of the JSR is thus more accurate.
- Our SOS and quadratic bounds are more robust with respect to the increase in the condition number of $P^*(\omega_N)$ or $P_{sos}^*(\omega_N)$ than the bound of [BJW21]. Moreover, they are always finite.
- Our SOS and quadratic bounds either benefit from the use of longer traces, or are less impacted than the bound of [BJW21].
- Our quadratic bounds, and especially the bound of Theorem 3.2, outperform the bound of [BJW21] when the number of samples available is small. The number of samples required by the bound of [BJW21] to beat our bound increases with the length of the traces and the dimension of the system.

One however has to note that our SOS-based bounds are negatively affected by the dimension of the system. Indeed, due to the rise in the condition number of $P_{sos}^*(\omega_N)$, when d increases, they require a larger number of samples to yield bounds as tight as the quadratic bounds. Moreover, the dimension of the problem to solve increases exponentially with d and the dimension of the system. Hence, since the computational complexity of solving a SDP program depends on both the dimension of the problem and the number of constraints, SOS-based bounds require the resolution of a program of much higher complexity.

Part III

Efficient computation using adaptive sampling

Chapter 5

The adaptive sampling approach in the quadratic case

As illustrated in Chapters 3 and 4, the scenario approach is a successful and handy method to obtain probabilistic bounds on the JSR of switched linear systems. However, this approach presents a major drawback: the sample complexity required to obtain a bound holding with a sufficiently large confidence level increases rapidly with the dimension of the problem and with the desired accuracy of the bound. This induces two potential issues. First, depending on the system analyzed, obtaining a huge number of trajectories might be costly. Second, the computational complexity of the problem to be solved increases with the sample complexity. This is especially problematic for large-scale problems, which might become intractable. In this chapter, we thus provide a novel approach answering the following question:

How to reduce the sample complexity required to obtain a bound on the JSR holding with a given confidence level, while maintaining the accuracy of the solution?

How to reduce the sampled complexity in the scenario approach has been relatively widely investigated recently. Hence, first, we present an existing method designed to this goal, based on a variation of the chance-constrained theory underlying the scenario approach. In a second time, we derive a novel approach, based on adaptive sampling and taking into account the structure of our initial problem. It consists in solving a multi-phase optimization problem, where the total number of samples available is divided into the several phases. We propose two variants of this approach and define different algorithms accordingly. Finally, we derive probabilistic bounds on the JSR based on the solutions computed by these algorithms.

Assumptions and notation

As a reminder, to obtain a quadratic probabilistic upper bound on the JSR, we attempt to approximate the solution (γ^o, P^o) to Problem (2.3):

$$\begin{aligned} \min_{P, \gamma \geq 0} \quad & \gamma \text{ s.t.} \quad (\mathbf{A}x)^T P \mathbf{A}x \leq \gamma^{2l} x^T P x, \quad \forall \mathbf{A} \in \mathcal{M}^l, \forall x \in \mathbb{S} \\ & P \succeq I_n, \quad \|P\| \leq C \text{ for a large } C \in \mathbb{R}_{\geq 0}. \end{aligned}$$

Therefore, we solve an associated sampled quasi-convex optimization problem: Problem (3.2). In this chapter, we will propose a multi-phase algorithm to obtain quadratic upper bounds. Since we will solve Problem (3.2) in the first phase, using N_1 samples, we denote it here by $\text{Opt}_1(\omega_{N_1})$ and its solution by $(\gamma_1^*(\omega_{N_1}), P_1^*(\omega_{N_1}))$:

$$\begin{aligned} \min_{P_1, \gamma_1 \geq 0} \quad & \gamma_1 \text{ s.t.} \quad (\mathbf{A}_{\mathbf{j}}x)^T P_1 \mathbf{A}_{\mathbf{j}}x \leq \gamma_1^{2l} x^T P_1 x, \quad \forall (x, \mathbf{j}) \in \omega_{N_1} \\ & P_1 \succeq I_n, \quad \|P_1\| \leq C \text{ for a large } C \in \mathbb{R}_{\geq 0}. \end{aligned} \quad (5.1)$$

As always, we apply the tie-breaking rule to select the feasible solution P of minimal condition number if needed. We summarize the notation used in the propositions of this chapter in the following assumption:

Assumption 5.1. We consider an n -dimensional switched linear system as in (2.1), whose set of m modes is denoted by $\mathcal{M} \subset \mathbb{R}^{n \times n}$; and a uniform random sampling $\omega_{N_1} \in \mathcal{Z}_l^{N_1}$, where $N_1 \geq d_{\text{quad}} + 1$. We denote by $(\gamma_1^*(\omega_{N_1}), P_1^*(\omega_{N_1}))$ the optimal solution to Problem (5.1) and by (γ^o, P^o) the optimal solution to Problem (2.3).

5.1 The fast algorithm

In [CGC14], a two-phase algorithm, named the *fast algorithm*, is proposed for scenario-based problems similar to Problem (2.3). In the first phase, one solves $\text{Opt}_1(\omega_{N_1})$ with N_1 samples to obtain a first solution $(\gamma_1^*(\omega_{N_1}), P_1^*(\omega_{N_1}))$. In the second phase, one refines the computation of the cost γ while maintaining P fixed to $P_1^*(\omega_{N_1})$, by solving the following one-dimensional problem:

$$\min_{\gamma_f \geq 0} \quad \gamma_f \text{ s.t.} \quad (\mathbf{A}_{\mathbf{j}}x)^T P_1^*(\omega_{N_1}) \mathbf{A}_{\mathbf{j}}x \leq \gamma_f^{2l} x^T P_1^*(\omega_{N_1}) x, \quad \forall (x, \mathbf{j}) \in \omega_{N_1+N_2}, \quad (5.2)$$

where $\omega_{N_1+N_2} = \omega_{N_1} \cup \omega_{N_2}$. We denote the solution to this problem by $\gamma_f^*(\omega_{N_1+N_2})$, and observe that it can easily be computed as:

$$\gamma_f^*(\omega_{N_1+N_2}) = \max \left(\gamma_1^*(\omega_{N_1}), \max_{(x, \mathbf{j}) \in \omega_{N_2}} \sqrt[l]{\frac{\|\mathbf{A}_{\mathbf{j}}x\|_{P_1^*(\omega_{N_1})}}{\|x\|_{P_1^*(\omega_{N_1})}}} \right).$$

Algorithm 1: Fast algorithm adapted to stability analysis of linear switched systems.

Input: Number of samples to use in each phase: N_1 and N_2 .

Result: $(\gamma_2^*(\omega_{N_1}, \omega_{N_2}), P^*(\omega_{N_1}))$

First Phase

- 1 Sampling of a set $\omega_{N_1} \in Z_l^{N_1}$, where $|\omega_{N_1}| = N_1$;
- 2 Computation of $(\gamma_1^*(\omega_{N_1}), P_1^*(\omega_{N_1}))$, solution to $\text{Opt}_1(\omega_{N_1})$;

Second Phase

- 3 Sampling of a set $\omega_{N_2} \in Z_l^{N_2}$ independent from ω_{N_1} , where $|\omega_{N_2}| = N_2$;
- 4 Computation of $\gamma_f^*(\omega_{N_1+N_2})$, solution to Problem (5.2);

We present the global procedure in Algorithm 1, for given numbers of samples N_1 and N_2 to use in each of the two phases.

In comparison to the resolution of Problem (3.2) with $N_1 + N_2$ samples, the total computational complexity is reduced since the problem in the second phase is one-dimensional. We now show that the sample complexity required to obtain a bound of equivalent accuracy holding with a given confidence level is also reduced:

Proposition 5.1. *Let Assumption 5.1 hold. Consider a uniform sampling $\omega_{N_2} \in Z_l^{N_2}$, and let $\gamma_f^*(\omega_{N_1+N_2})$ denote the solution to Problem (5.2). Then, for any $\beta \in [0, 1)$, with probability at least β , it holds that:*

$$\rho(\mathcal{M}) \leq \frac{\gamma_f^*(\omega_{N_1+N_2})}{\sqrt[l]{\delta\left(\frac{\epsilon(\beta, N_1, N_2)m^l \kappa(P_1^*(\omega_{N_1}))}{2}\right)}}$$

where $\delta(\cdot)$ is given in (3.7), $\beta_1(\epsilon, N_1) = \sum_{j=0}^{d_{quad}} \binom{N_1}{j} \epsilon^j (1-\epsilon)^{N_1-j}$, and $\epsilon(\beta, N_1, N_2)$ is the solution to $\beta = 1 - (1 - \epsilon(\beta, N_1, N_2))^{N_2} \beta_1(\epsilon(\beta, N_1, N_2), N_1)$.

Proof. Let $N = N_1 + N_2$, and define $V(\gamma, P) := \{(x, \mathbf{j}) \in Z_l : (\mathbf{A}_{\mathbf{j}}x)^T P \mathbf{A}_{\mathbf{j}}x > \gamma^{2l} x^T P x\}$ as the set of violated constraints for a given pair (γ, P) . [CGC14, Thm. 1] ensures that for all $\epsilon \in (0, 1)$, the following holds:

$$\mu_l^N \{\omega_{N_1+N_2} \in Z_l^N : \mu_l(V(\gamma_f^*(\omega_{N_1+N_2}), P_1^*(\omega_{N_1}))) \leq \epsilon\} \geq 1 - (1 - \epsilon)^{N_2} \beta_1(\epsilon, N_1)$$

where $\beta_1(\epsilon, N_1) = \sum_{j=0}^{d_{quad}} \binom{N_1}{j} \epsilon^j (1 - \epsilon)^{N_1-j}$.

This result can replace Theorem 4 in [KBJT19]. Then, the proof follows the same argument as in the proof of [KBJT19, Thm. 6], so that details are omitted here. $\square$

5.2 An enhanced approach based on adaptive sampling

Attempts to reduce the sample complexity required to obtain bounds holding with a given confidence level such as the fast algorithm present the two following drawbacks:

- They consider the samples to be drawn independently according to the same distribution in each phase. In our setting, this implies that we consider a uniform sampling of the state space in each phase. Similarly, these attempts consider sampled versions of the same initial problem in each phase. In the fast algorithm, for instance, both Problem (5.1) and Problem (5.2) are based on the white-box problem Problem (2.3).
- They hold for all scenario-based optimization programs. Hence, they do not take into account the specific structure of our initial problem, that is Problem (2.3). In particular, they neglect the impact of the condition number of P^o on the accuracy of the probabilistic bounds derived in Theorems 3.1 and 3.2.

Therefore, we propose a new approach overcoming these drawbacks. We consider different samplings in each phase, corresponding to variants of our initial problem, after a specific change of variables has been applied. We base our approach on an analysis of the Problem (2.3), and rely on a key property of the JSR: its invariance under similarity transformations.

We first provide the intuition behind the changes of variables to be performed in the white-box setting. We then derive our approach based on two phases only, before extending our technique to an arbitrary number p of phases.

5.2.1 Idea in the white-box setting

Suppose ω_{N_1} is sampled uniformly. Then, when P^o , the solution to Problem (2.3), possesses a high condition number, so does $P_1^*(\omega_{N_1})$, the solution to Problem (5.1). This prevents the bounds of Theorems 3.1 and 3.2 to be tight. Therefore, we seek for a coordinate system in which P^o is the identity, which allows to get rid of the condition number.

This optimal coordinate system, in the white-box setting, can easily be computed as follows. Let $L \in \mathbb{R}^{n \times n}$ be the matrix defining the change of variables. In the new coordinate system, with $x' = Lx$, the constraints of Problem (2.3) become:

$$\begin{aligned} (\mathbf{A}L^{-1}x')^T P \mathbf{A} L^{-1} x' &\leq \gamma^{2l} (L^{-1}x')^T P L^{-1} x', \quad \forall \mathbf{A} \in \mathcal{M}^l, \forall x' \in LS \\ \iff (\mathbf{A}L^{-1}x')^T P \mathbf{A} L^{-1} x' &\leq \gamma^{2l} (L^{-1}x')^T P L^{-1} x', \quad \forall \mathbf{A} \in \mathcal{M}^l, \forall x' \in \mathbb{S} \end{aligned}$$

by homogeneity of System (2.1). By defining $P' = L^{-T}PL^{-1}$, we obtain the following replacement of Problem (2.3):

$$\begin{aligned} \min_{P', \gamma' \geq 0} \quad & \gamma' \\ \text{s.t.} \quad & (L\mathbf{A}L^{-1}x')^T P' L\mathbf{A}L^{-1}x' \leq \gamma'^{2l} x'^T P' x', \quad \forall \mathbf{A} \in \mathcal{M}^l, \forall x' \in \mathbb{S} \\ & P' \succeq I_n, \quad \|P'\| \leq C \text{ for a large } C \in \mathbb{R}_{\geq 0}. \end{aligned} \quad (5.3)$$

We denote this problem by $\text{Opt}(L)$ and its solution by $(\gamma'^o(L), P'^o(L))$. After performing the change of coordinates, we thus analyze in fact the stability of the set $L\mathcal{M}L^{-1}$. Since by Proposition 2.2, $\rho(\mathcal{M}^l) = \rho(L\mathcal{M}^lL^{-1})$, we can conduct our stability analysis on $L\mathcal{M}^lL^{-1}$ instead of $\mathcal{M}^l$ equivalently: $\gamma'^o = \gamma^o$. Moreover, $P'^o(L) = L^{-T}P^oL^{-1}$ is a feasible solution to $\text{Opt}(L)$.

We thus seek to compute the matrix L such that $L^{-T}P^oL^{-1}$ is the identity. Since $P^o \in \mathcal{S}_{++}^n$, it admits a Cholesky decomposition: $P^o = (L^o)^T L^o$, where L^o maps the elements of E_{P^o} to $\mathbb{S}$, as illustrated on Figure 5.1.

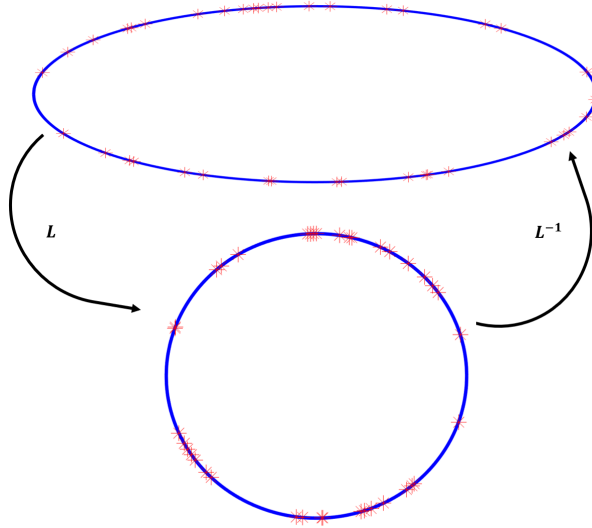


Figure 5.1: Change of coordinates L mapping the elements of an ellipsoid E_P to the unit sphere. L is the Cholesky decomposition of P : $P = L^T L$.

Hence, L^o is exactly the change of coordinates we are looking for. Indeed, in the coordinate system defined by L^o , P^o is mapped to the identity: $P'^o(L^o) = I_n$ is a feasible solution to Problem (5.3).

5.2.2 Two-phase adaptive sampling approach

Having access to a finite number of simulations solely, we cannot compute exactly either P^o or $P^o(L^o)$. Hence, we use adaptive sampling ([DA20], [Set10]) to (i) compute an approximation L_1 of the optimal change of variables L^o , by solving $\text{Opt}_1(\omega_{N_1})$ and (ii) compute a probabilistic bound on the JSR in the new coordinate system, by solving a sampled version of $\text{Opt}(L_1)$. In the second phase, the sampled solution will be an approximation of the identity. This will allow for the bounds of Theorems 3.1 and 3.2 to be tighter for an identical number of samples used. More precisely, the technique consists in the following steps:

1. Uniform sampling of $\omega_{N_1} \in Z_l^{N_1}$.
2. Computation of $(\gamma_1^*(\omega_{N_1}), P_1^*(\omega_{N_1}))$, solution to $\text{Opt}_1(\omega_{N_1})$. From this, we learn an approximation of the optimal change of variables L^o , given by the Cholesky decomposition of $P_1^*(\omega_{N_1})$: $P_1^*(\omega_{N_1}) = L_1^T L_1$. We will approximate the JSR of the system in the coordinate system determined by L_1 .
3. Sampling of ω_{N_2} , composed of N_2 trajectories, uniformly for the system $L_1 \mathcal{M}^l L_1^{-1}$. In our black-box setting, we have no direct access to $L_1 \mathbf{A}_j L_1^{-1}$, for $\mathbf{A}_j \in \mathcal{M}^l$. We therefore use Procedure Generation to sample ω_{N_2} .

Procedure Generation of samples uniformly distributed for $L_1 \mathcal{M}^l L_1^{-1}$	
Input: Number of points to sample: N_2 , change of variables: L_1	
Output: Set of N_2 trajectories sampled uniformly for $L_1 \mathcal{M}^l L_1^{-1}$: ω_{N_2}	
1	for $i = 1, \dots, N_2$ do
2	Sampling of x_i uniformly on $\mathbb{S}$;
3	Computation of $\tilde{x}_i = L_1^{-1} x_i$;
4	Sampling of $\mathbf{j}_i$ uniformly from M^l and generation of $\mathbf{A}_j \tilde{x}_i$;
5	Computation of $y_i = L_1 \mathbf{A}_j L_1^{-1} x_i$;
6	end
7	$\omega_{N_2} = \{(x_i, y_i), i = 1, 2, \dots, N_2\}$;

Note that the initial states of ω_{N_2} are not uniformly sampled on the sphere for the system $\mathcal{M}^l$. They are instead uniformly sampled on the ellipsoid $E_{P_1^*(\omega_{N_1})}$, since $L_1^{-1}(\mathbb{S}) = E_{P_1^*(\omega_{N_1})}$, as illustrated on Figure 5.1.

4. Computation of a bound on the JSR of the system by solving a sampled version of $\text{Opt}(L_1)$. Let the variables of this sampled problem be $\gamma_2 \in \mathbb{R}$ and $P_2 \in \mathbb{R}^{n \times n}$. We would like to benefit from the information contained in both ω_{N_1} and ω_{N_2} . Therefore, we enforce two sets of constraints. First, we

implement a set of constraints associated to ω_{N_2} :

$$(L_1 \mathbf{A}_j L_1^{-1} x)^T P_2 (L_1 \mathbf{A}_j L_1^{-1} x) \leq \gamma_2^{2l} x^T P_2 x, \quad \forall (x, \mathbf{j}) \in \omega_{N_2}.$$

Then, we enforce the constraints of $\text{Opt}_1(\omega_{N_1})$ translated in the new coordinate system:

$$(L_1 \mathbf{A}_j L_1^{-1} x')^T P_2 L_1 \mathbf{A}_j L_1^{-1} x' \leq \gamma_2^{2l} x'^T P_2 x', \quad \forall (x', \mathbf{j}) : x' = L_1 x, (x, \mathbf{j}) \in \omega_{N_1}.$$

To the contrary to the distribution of the trajectories of ω_{N_2} , the distribution of the trajectories of ω_{N_1} is not uniform for the system considered.

Let $\omega_{N_1+N_2} = \omega_{N_1} \cup \omega_{N_2}$. We then solve the following sampled version of $\text{Opt}(L_1)$, denoted by $\text{Opt}_2(\omega_{N_1+N_2})$:

$$\begin{aligned} \min_{P_2, \gamma_2 \geq 0} \quad & \gamma_2 \text{ s.t. } (L_1 \mathbf{A}_j L_1^{-1} x)^T P_2 L_1 \mathbf{A}_j L_1^{-1} x \leq \gamma_2^{2l} x^T P_2 x, \\ & \forall (x, \mathbf{j}) \in \omega_{N_2} \cup \{(x', \mathbf{j}) : x' = L_1 x, (x, \mathbf{j}) \in \omega_{N_1}\} \\ & P_2 \succeq I_n, \quad \|P_2\| \leq C \text{ for a large } C \in \mathbb{R}_{\geq 0}. \end{aligned} \quad (5.4)$$

We denote its solution by $(\gamma_2^*(\omega_{N_1+N_2}), P_2^*(\omega_{N_1+N_2}))$. When N_2 is equal to zero, that is when we only implement the constraints associated to ω_{N_1} , the solution to $\text{Opt}_2(\omega_{N_1+N_2})$ is exactly $(\gamma_1^*(\omega_{N_1}), I_n)$.

We summarize our approach in Algorithm 2.

Algorithm 2: Two-phase adaptive sampling algorithm.	
Input: Number of samples to use in each phase: N_1 and N_2 .	
Result: $(\gamma_2^*(\omega_{N_1+N_2}), P_2^*(\omega_{N_1+N_2}))$	
First Phase	
1	Sampling of a set $\omega_{N_1} \in Z_l^{N_1}$, where $ \omega_{N_1} = N_1$;
2	Computation of $(\gamma_1^*(\omega_{N_1}), P_1^*(\omega_{N_1}))$, solution to $\text{Opt}_1(\omega_{N_1})$;
3	Computation of the change of variables L_1 such that $P_1^*(\omega_{N_1}) = L_1^T L_1$;
Second Phase	
4	Sampling of a set ω_{N_2} uniformly distributed in the coordinate system determined by L_1 , where $ \omega_{N_2} = N_2$, using Procedure Generation ;
5	Computation of $(\gamma_2^*(\omega_{N_1+N_2}), P_2^*(\omega_{N_1+N_2}))$, solution to $\text{Opt}_2(\omega_{N_1+N_2})$;

Remark 5.1. When solving $\text{Opt}_2(\omega_{N_1+N_2})$ as the sampled version of $\text{Opt}(L_1)$, we enforce the constraints associated to the trajectories of both ω_{N_2} and ω_{N_1} . We could also solve a sampled version of $\text{Opt}(L_1)$ where only the constraints associated to ω_2 are enforced. This would decrease the optimal cost of the problem, but $P_2^*(\omega_{N_1+N_2})$ would be further from the identity. Empirically, we determined that the resulting bound on the JSR would hence be less tight.

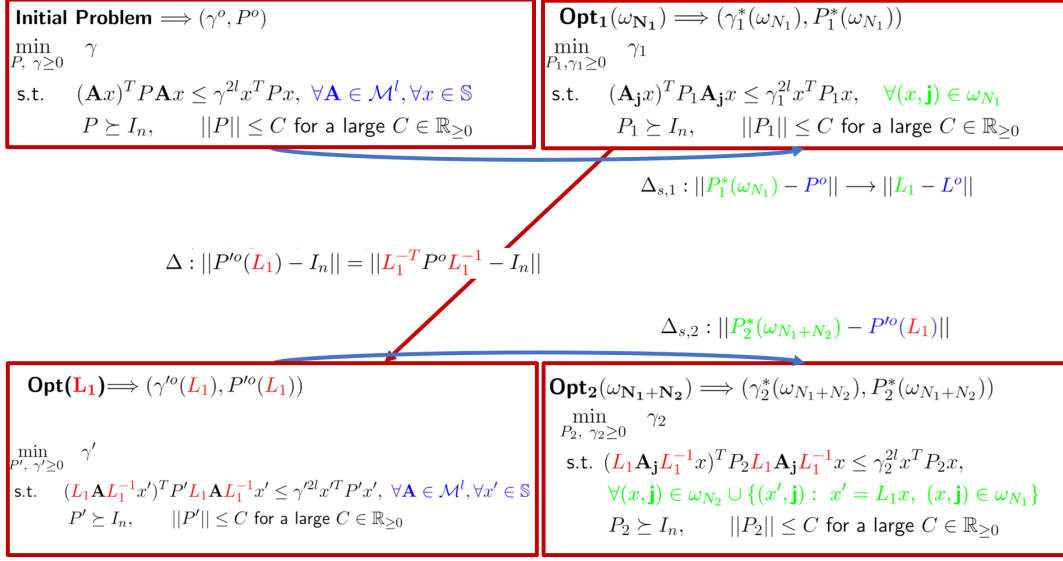


Figure 5.2: Relations between the different optimization problems designed for the two-phase adaptive sampling approach, and illustration of the factors impacting the distance $\|P_2^*(\omega_{N_1+N_2}) - I_n\|$. The left-hand side of the figure presents the white-box problems, before and after the change of variables is performed. The right-hand side presents the associated sampled problems. We only have access to the solutions to these problems, approximating the solutions of the initial problems. The closeness of $P_2^*(\omega_{N_1+N_2})$ to the identity depends on $\Delta_{s,1}$ and $\Delta_{s,2}$. $\Delta_{s,1}$ affects the quality of the change of variables by inducing a distance between the solution of $\text{Opt}(L_1)$ and the identity. It decreases when N_1 increases, while $\Delta_{s,2}$ decreases when N_2 increases. A trade-off has thus to be achieved between both quantities.

The performance of our approach depends on the closeness of $P_2^*(\omega_{N_1+N_2})$ to the identity. This closeness is impacted by two factors, as illustrated on Figure 5.2:

1. The distance $\Delta_{s,1} = \|P_1^*(\omega_{N_1}) - P^o\|$, implying a distance $\|L_1 - L^o\|$ which represents the quality of the change of variables L_1 . This distance is due to the restriction of the infinite number of constraints of Problem (2.3) to N_1 constraints in Problem (5.1). It decreases when the number of samples N_1 used in the first phase increases. It induces a distance $\Delta = \|P'^o(L_1) - I_n\|$ in the solution of Problem (5.3).
2. The distance $\Delta_{s,2} = \|P_2^*(\omega_{N_1+N_2}) - P'^o(L_1)\|$ due to the restriction of the infinite number of constraints of Problem (5.3) to $N_1 + N_2$ constraints in Problem (5.4). It decreases when the number of samples N_2 used in the second phase increases.

Clearly, for a given total number of samples to be used, there will be a trade-off to achieve between these two factors. We propose an approximation for the total distance separating $P_2^*(\omega_{N_1+N_2})$ from the identity, based on $\Delta_{s,1}$ and $\Delta_{s,2}$:

Proposition 5.2. *Let Assumption 5.1 hold. Let $(\gamma_2^*(\omega_{N_1+N_2}), P_2^*(\omega_{N_1+N_2}))$ be the solution to $\text{Opt}_2(\omega_{N_1+N_2})$, and $(\gamma'^o(L_1), P'^o(L_1))$ be the solution to $\text{Opt}(L_1)$, where L_1 is the Cholesky decomposition of $P_1^*(\omega_{N_1})$. Let $\Delta_{s,1} = \|P_1^*(\omega_{N_1}) - P^o\|$ and $\Delta_{s,2} = \|P_2^*(\omega_{N_1+N_2}) - P'^o(L_1)\|$ be the distances separating the solutions of the initial and sampled problems. Then, it holds that:*

$$\|P_2^*(\omega_{N_1+N_2}) - I_n\| \leq \Delta_{s,2} + \Delta_{s,1}. \quad (5.5)$$

Proof. Since the Cholesky decomposition L of a matrix $P \in \mathcal{S}_{++}^n$ is such that $\lambda_{\min}(L) = \sqrt{\lambda_{\min}(P)}$ and since $\|L^{-1}\| = \frac{1}{\lambda_{\min}(L)}$ [GVL13], we have:

$$\begin{aligned} \Delta &= \|P'^o(L_1) - I_n\| = \|L_1^{-T} P^o L_1^{-1} - I_n\| = \|L_1^{-T} (P^o - P_1^*(\omega_{N_1})) L_1^{-1}\| \\ &\leq \|L_1^{-T}\| \|P^o - P_1^*(\omega_{N_1})\| \|L_1^{-1}\| = \frac{\Delta_{s,1}}{\lambda_{\min}(P_1^*(\omega_{N_1}))} = \Delta_{s,1} \end{aligned}$$

where the last equality follows from the constraint $P \succeq I_n$ in $\text{Opt}_1(\omega_{N_1})$ and $\text{Opt}_2(\omega_{N_1+N_2})$, from the homogeneity of these Problems and from the tie-breaking rule applied. Indeed, if $P \in \mathcal{S}_{++}^n$ satisfies $(\mathbf{A}_j x)^T P \mathbf{A}_j x \leq \gamma^{2l} x^T P x$, $\forall (x, \mathbf{j}) \in \omega_N$, then so does αP , for any $\alpha \in \mathbb{R}$. Hence, the feasible $P \succeq I_n \in \mathcal{S}_{++}^n$ whose largest eigenvalue, $\lambda_{\max}(P)$, is minimal will be such that $\lambda_{\min}(P) = 1$. Indeed, then, there exists no $\alpha \in \mathbb{R}$ such that $\lambda_{\min}(\alpha P) = \alpha \lambda_{\min}(P) \geq 1$ and $\lambda_{\max}(\alpha P) = \alpha \lambda_{\max}(P) \leq \lambda_{\max}(P)$. Hence,

$$\begin{aligned} \|P_2^*(\omega_{N_1+N_2}) - I_n\| &= \|P_2^*(\omega_{N_1+N_2}) - P'^o(L_1) + P'^o(L_1) - I_n\| \\ &\leq \|P_2^*(\omega_{N_1+N_2}) - P'^o(L_1)\| + \|P'^o(L_1) - I_n\| \\ &\leq \Delta_{s,2} + \Delta_{s,1}. \end{aligned}$$

□

5.2.3 p -phases adaptive sampling approach

We now refine the idea of adaptive sampling by increasing the number of phases of our approach. After each phase, the change of variables is recomputed until it becomes sufficiently close to the optimal change of variables.

For $k \geq 1$, let $\mathbf{L}_k = L_k \dots L_2 L_1$ and $\mathbf{L}_k^{-1} = L_k^{-1} \dots L_2^{-1} L_1^{-1}$. Let $\mathbf{L}_k = \mathbf{L}_k^{-1} = I_n$ for $k = 0$. Let $\Omega_k = \omega_{N_k} \cup \{(x', \mathbf{j}) : x' = L_{k-1} x, (x, \mathbf{j}) \in \omega_{N_{k-1}}\} \cup \dots \cup \{(x', \mathbf{j}) : x' =$

$\mathbf{L}_{\mathbf{k}-1}x, (x, \mathbf{j}) \in \omega_{N_1}\}$.

We define $\text{Opt}_k(\omega_{N_1+\dots+N_k})$, of solution denoted by $(\gamma_k^*(\omega_{N_1+\dots+N_k}), P_k^*(\omega_{N_1+\dots+N_k}))$, in the following recursive way:

$$\begin{aligned} \min_{P_k, \gamma_k \geq 0} \quad & \gamma_k \text{ s.t. } (\mathbf{L}_{\mathbf{k}-1}\mathbf{A}_{\mathbf{j}}\mathbf{L}_{\mathbf{k}-1}^{-1}x)^T P_k \mathbf{L}_{\mathbf{k}-1}\mathbf{A}_{\mathbf{j}}\mathbf{L}_{\mathbf{k}-1}^{-1}x \leq \gamma_k^{2l} x^T P_k x, \quad \forall (x, \mathbf{j}) \in \Omega_k \\ & P_k \succeq I_n, \quad \|P_k\| \leq C \text{ for a large } C \in \mathbb{R}_{\geq 0}, \end{aligned} \quad (5.6)$$

where L_i , for $i = 1, \dots, k-1$, is the Cholesky decomposition of $P_i^*(\omega_{N_1+\dots+N_i})$, the solution to $\text{Opt}_i(\omega_{N_1+\dots+N_i})$. Note that Problems (5.1) and (5.4) are particular cases of Problem (5.6).

After an initial change of variables L_1 is obtained by solving Problem (5.1) with N_1 trajectories, we solve $\text{Opt}_k(\omega_{N_1+\dots+N_k})$ for $k = 2, \dots, p$. At Phase k , we analyze the stability of the system $\mathbf{L}_{\mathbf{k}-1}\mathcal{M}^l\mathbf{L}_{\mathbf{k}-1}^{-1}$. We sample N_k points uniformly for this system using Procedure Generation, which define a set ω_{N_k} of constraints. We append to this set the $N_1 + \dots + N_{k-1}$ constraints of $\text{Opt}_{k_1}(\omega_{N_1+\dots+N_{k-1}})$, translated in the new coordinate system. We present our approach in Algorithm 3.

Algorithm 3: p -phases adaptive sampling algorithm.	
Input: Number of phases: p , number of samples of each phase: $N_1, \dots, N_p$.	
Result: $(\gamma_p^*(\omega_{N_1+\dots+N_p}), P_p^*(\omega_{N_1+\dots+N_p}))$	
First Phase	
1	Sampling of a set $\omega_{N_1} \in Z_l^{N_1}$, where $ \omega_{N_1} = N_1$;
2	Computation of $(\gamma_1^*(\omega_{N_1}), P_1^*(\omega_{N_1}))$, solution to $\text{Opt}_1(\omega_{N_1})$;
3	Computation of the change of variables L_1 such that $P_1^*(\omega_{N_1}) = L_1^T L_1$;
Second set of phases	
4	for $k = 2$ to p do
5	Sampling of ω_{N_k} uniformly for $\mathbf{L}_{\mathbf{k}-1}\mathcal{M}^l\mathbf{L}_{\mathbf{k}-1}^{-1}$, where $ \omega_{N_k} = N_k$, using Procedure Generation ;
6	Computation of $(\gamma_k^*(\omega_{N_1+\dots+N_k}), P_k^*(\omega_{N_1+\dots+N_k}))$, solution to $\text{Opt}_k(\omega_{N_1+\dots+N_k})$;
7	if $k \neq p$ then
8	Computation of the change of variables L_k such that $P_k^*(\omega_{N_1+\dots+N_k}) = L_k^T L_k$;
9	Computation of $\mathbf{L}_{\mathbf{k}} = L_k \mathbf{L}_{\mathbf{k}-1}$ and $\mathbf{L}_{\mathbf{k}}^{-1} = L_k^{-1} \mathbf{L}_{\mathbf{k}-1}^{-1}$;
10	end
11	end

Assumption 5.2 will be required in our next statements:

Assumption 5.2. We consider $p, N \in \mathbb{N}$, and $N_i \in \mathbb{N} \geq d_{quad} + 1$ for $i = 1, \dots, p$ such that $N_1 + \dots + N_p = N$. For $i = 1, \dots, p$, let ω_{N_i} be a uniform random sampling for System $\mathbf{L}_{i-1} \mathcal{M}^l \mathbf{L}_{i-1}^{-1}$. We denote by $(\gamma_i^*(\omega_{N_1+\dots+N_i}), P_i^*(\omega_{N_1+\dots+N_i}))$ the solution to $\text{Opt}_i(\omega_{N_1+\dots+N_i})$, and by $(\gamma'^o(\mathbf{L}_i), P'^o(\mathbf{L}_i))$ the solution to $\text{Opt}(\mathbf{L}_i)$. L_i is the Cholesky decomposition of $P_i^*(\omega_{N_1+\dots+N_i})$.

The performance of our approach now depends on the closeness of $P_p^*(\omega_{N_1+\dots+N_p})$ to the identity. We thus generalize Proposition 5.2 as follows:

Proposition 5.3. *Let Assumptions 5.1 and 5.2 hold. For $i = 2, \dots, p$, let $\Delta_{s,i} = \|P_i^*(\omega_{N_1+\dots+N_i}) - P'^o(\mathbf{L}_{i-1})\|$ be the distances separating the solutions of the initial and sampled problems. Then, it holds that:*

$$\|P_p^*(\omega_{N_1+\dots+N_p}) - I_n\| \leq \Delta_{s,p} + \Delta_{s,p-1} \quad (5.7)$$

Proof. Since $P'^o(\mathbf{L}_{p-1}) = L_{p-1}^{-T} P'^o(\mathbf{L}_{p-2}) L_{p-1}^{-1}$ is a solution to $\text{Opt}(\mathbf{L}_{p-1})$, we have:

$$\begin{aligned} \|P_p^*(\omega_{N_1+\dots+N_p}) - I_n\| &\leq \|P_p^*(\omega_{N_1+\dots+N_p}) - P'^o(\mathbf{L}_{p-1})\| + \|P'^o(\mathbf{L}_{p-1}) - I_n\| \\ &= \Delta_{s,p} + \|L_{p-1}^{-T} (P'^o(\mathbf{L}_{p-2}) - P_{p-1}^*(\omega_{N_1+\dots+N_{p-1}})) L_{p-1}^{-1}\| \\ &\leq \Delta_{s,p} + \frac{\Delta_{s,p-1}}{\lambda_{\min}(P_{p-1}^*(\omega_{N_1+\dots+N_{p-1}}))} = \Delta_{s,p} + \Delta_{s,p-1} \end{aligned}$$

since $\lambda_{\min}(P_{p-1}^*(\omega_{N_1+\dots+N_{p-1}})) = 1$ by the same argument as in the proof of Proposition 5.2. $\square$

Remark 5.2. $\Delta_{s,k}$ decreases not only when the number of samples N_k used in phase k increases, but also when the condition number of $P_k^*(\omega_{N_1+\dots+N_k})$ decreases. Increasing the number of phase is thus beneficial as long as the decrease in $\kappa(P_k^*(\omega_{N_1+\dots+N_k}))$ compensates the fact that a lower number of samples is used in phase k . For instance, usually, $\Delta_{s,k} \leq \Delta_{s,1}$ despite N_1 being larger than N_k for $k \leq 2$.

5.3 Novel theoretical bounds based on adaptive sampling

To evaluate the performance of our approach, we seek to derive probabilistic bounds on the JSR based on the sets $\omega_{N_1}, \omega_{N_2}, \dots, \omega_{N_p}$ and on the solution $(\gamma_p^*(\omega_{N_1+N_2+\dots+N_p}), P_p^*(\omega_{N_1+\dots+N_p}))$ to $\text{Opt}_p(\omega_{N_1+\dots+N_p})$. Indeed, then, we can compare it with the bounds obtained without resorting to adaptive sampling, given in

Theorems 3.1 and 3.2. For the same total number of trajectories sampled and with the same confidence level, can we improve the bound on the JSR of the system? Obtaining such bounds on the JSR and being able to optimizing them is also crucial to determine the optimal values of $N_1, \dots, N_p$ and p . Two approaches are possible:

- Analysis of $\text{Opt}_p(\omega_{N_1+\dots+N_p})$ on the basis of ω_{N_p} only, that is of the N_p samples whose distribution is uniform in the system considered.
- Analysis of $\text{Opt}_p(\omega_{N_1+\dots+N_p})$ on the basis of all samples used, that is of the totality of the set $\omega_{N_1+\dots+N_p}$.

In both cases, the bounds derived have to be compared with the one that would be obtained by using all the $N_1 + \dots + N_p$ samples in Phase 1. The bounds obtained through the first analysis will be more conservative, but easier to derive. Indeed, the set Ω_p is not identically distributed. In addition, $L_1, \dots, L_{p-1}$ are random variables depending on $\omega_{N_1}, \dots, \omega_{N_{p-1}}$. Therefore, when considering all trajectories sampled, we cannot rely on the results of Chapter 3.

Bound based on the last set of samples only

Basing our analysis of $\text{Opt}_p(\omega_{N_1+\dots+N_p})$ on the Np last samples solely allows us to use the results of Chapter 3. Indeed, then, we consider the matrices $L_1, \dots, L_{p-1}$ and the trajectories of the set $\omega_{N_1+\dots+N_{p-1}}$ to be deterministic. The only random elements remaining are the Np trajectories of ω_{N_p} , sampled uniformly in the system considered, that is $\mathbf{L}_{\mathbf{p}-1}\mathcal{M}^l\mathbf{L}_{\mathbf{p}-1}^{-1}$. Moreover, the JSR of this system is exactly the one of $\mathcal{M}^l$. We propose a bound based on sensitivity analysis, adapted from the one of Theorem 3.1. Doing so requires the introduction of an additional sampled problem, approximating the maximal norm of system $\mathbf{L}_{\mathbf{p}-1}\mathcal{M}^l\mathbf{L}_{\mathbf{p}-1}^{-1}$:

$$\min_{\lambda \geq 0} \lambda, \text{ s.t. } \|\mathbf{L}_{\mathbf{p}-1}\mathbf{A}_{\mathbf{j}}\mathbf{L}_{\mathbf{p}-1}^{-1}x\| \leq \lambda\|x\|, \forall (x, \mathbf{j}) \in \Omega_p \quad (5.8)$$

where L_i , for $i = 1, \dots, p-1$, is the Cholesky decomposition of the solution $P_i^*(\omega_{N_1+\dots+N_i})$ of $\text{Opt}_i(\omega_{N_1+\dots+N_i})$. We denote the solution to Problem (5.8) by $\lambda_p^*(\omega_{N_1+\dots+N_p})$.

Theorem 5.1

Let Assumptions 5.1 and 5.2 hold, and let $\lambda_p^*(\omega_{N_1+\dots+N_p})$ be the optimal solution to Problem (5.8). For any $\beta \in [0, 1)$ and $\beta_1 \in [0, 1)$, let $\epsilon = m^l \left(1 - \sqrt[p]{\frac{1-\beta}{d_{quad}+1}}\right)$ and $\epsilon_1 = \frac{m^l}{2}(1 - \sqrt[p]{1-\beta_1})$. Then, it holds that, with

probability at least $\beta + \beta_1 - 1$:

$$\rho(\mathcal{M}) \leq \sqrt[l]{\gamma_p^*(\omega_{N_1+\dots+N_p})^l + (\gamma_p^*(\omega_{N_1+\dots+N_p})^l + A(\epsilon_1))\Delta(\epsilon)\kappa(P_p^*(\omega_{N_1+\dots+N_p}))},$$

where $\Delta(\epsilon) = \sqrt{2 - 2\delta(\epsilon)}$, $A(\epsilon_1) = \frac{\lambda_p^*(\omega_{N_1+\dots+N_p})}{\delta(\epsilon_1)}$, $\delta(\cdot)$ is given by (3.7) and $\kappa(P) = \sqrt{\frac{\lambda_{\max}(P)}{\lambda_{\min}(P)}}$.

Proof. First, we prove that, with probability at least β , it holds that

$$\begin{aligned} \rho(\mathcal{M}) &= \rho(\mathbf{L}_{\mathbf{p}-1}\mathcal{M}^l\mathbf{L}_{\mathbf{p}-1}^{-1}) \\ &\leq \sqrt[l]{\gamma_p^*(\omega_{N_1+\dots+N_p})^l + (\gamma_p^*(\omega_{N_1+\dots+N_p})^l + B(\mathbf{L}_{\mathbf{p}-1}\mathcal{M}^l\mathbf{L}_{\mathbf{p}-1}^{-1}))\Delta(\epsilon)\kappa(P_p^*(\omega_{N_1+\dots+N_p}))} \end{aligned} \quad (5.9)$$

where $B(\mathbf{L}_{\mathbf{p}-1}\mathcal{M}^l\mathbf{L}_{\mathbf{p}-1}^{-1}) = \max_{\mathbf{A} \in \mathbf{L}_{\mathbf{p}-1}\mathcal{M}^l\mathbf{L}_{\mathbf{p}-1}^{-1}} \|\mathbf{A}\|$.

Clearly, Lemma 3.3, derived for Problem (2.3), also holds for Problem (5.3): it suffices to replace the system $\mathcal{M}^l$ by the system $\mathbf{L}_{\mathbf{p}-1}\mathcal{M}^l\mathbf{L}_{\mathbf{p}-1}^{-1}$. Hence, there exists a set $\omega_s = \{(x_i, \mathbf{j}_i), i = 1, \dots, d_{quad+1}\}$ such that $\gamma'^o(\mathbf{L}_{\mathbf{p}-1}) = \gamma^*(\omega_s)$, where $\gamma^*(\omega_s)$ is the optimal cost of the sampled version of $\text{Opt}(\mathbf{L}_{\mathbf{p}-1})$ implementing only the constraints associated to ω_s .

Moreover, since ω_{N_p} is sampled uniformly in the system considered, Proposition 3.1 also holds for Problem (5.3), only with N replaced by N_p . Then, since the solution $(\gamma_p^*(\omega_{N_1+\dots+N_p}), P_p^*(\omega_{N_1+\dots+N_p}))$ satisfies any subset of constraints of $\text{Opt}_p(\omega_{N_1+\dots+N_p})$, it holds that $\forall (x, \mathbf{j}) \in \omega_{N_p}$, $\|\mathbf{L}_{\mathbf{p}-1}\mathbf{A}_{\mathbf{j}}\mathbf{L}_{\mathbf{p}-1}^{-1}x\|_{P_p^*(\omega_{N_1+\dots+N_p})} \leq \gamma_p^*(\omega_{N_1+\dots+N_p})^l \|x\|_{P_p^*(\omega_{N_1+\dots+N_p})}$. Hence, we can follow the same reasoning as in the proof of Proposition 3.2, applied to system $\mathbf{L}_{\mathbf{p}-1}\mathcal{M}^l\mathbf{L}_{\mathbf{p}-1}^{-1}$ instead of the initial system $\mathcal{M}^l$ and obtain Equation (5.9). Adding the constraints associated to $\omega_{N_1}, \dots, \omega_{N_{p-1}}$ can only increase the probability to sample a trajectory $(\tilde{x}, \mathbf{j})$ whose initial state belongs to a spherical cap associated to x' and such that $\mathbf{j} = \mathbf{j}'$, for each $(x', \mathbf{j}') \in \omega_s$. Hence, the result holds for $\text{Opt}_p(\omega_{N_1+\dots+N_p})$ where all the constraints in Ω_p are enforced.

We now prove that

$$B(\mathbf{L}_{\mathbf{p}-1}\mathcal{M}^l\mathbf{L}_{\mathbf{p}-1}^{-1}) \leq \frac{\lambda_p^*(\omega_{N_1+\dots+N_p})}{\delta(\epsilon_1)} \quad (5.10)$$

holds with probability at least β_1 . Suppose that, in Problem (5.8), we only implement the constraints associated to the uniform sampling ω_{N_p} : $\|\mathbf{L}_{\mathbf{p}-1}\mathbf{A}_{\mathbf{j}}\mathbf{L}_{\mathbf{p}-1}^{-1}x\| \leq \lambda\|x\|$, $\forall (x, \mathbf{j}) \in \omega_{N_p}$, and obtain a solution $\lambda_p^*(\omega_{N_1+\dots+N_p})$. Then, the same argument as in the proof of Proposition 3.3 ensures that:

$$\mu_l^{N_p} \{\omega_{N_p} \in Z_l^{N_p} : \mu_l(V'(\omega_{N_p})) \leq \epsilon_1\} \geq \beta_1(\epsilon_1, N_p)$$

where $V'(\omega_{N_p}) := \{(x, \mathbf{j}) \in Z_l : \|\mathbf{L}_{\mathbf{p}-1} \mathbf{A}_{\mathbf{j}} \mathbf{L}_{\mathbf{p}-1}^{-1} x\| > \lambda_p^*(\omega_{N_1+\dots+N_p})\}$ is the set of violated constraints and $\beta_1(\epsilon_1, N_p) = 1 - (1 - \epsilon_1)^{N_p}$.

Adding the constraints associated to $\omega_{N_1}, \dots, \omega_{N_{p-1}}$ can only increase the cost and decrease the measure of the set of violated constraints, since the trajectories associated to $\omega_{N_1}, \dots, \omega_{N_{p-1}}$ do not, by definition, belong to this set. So, the result holds for the solution of Problem (5.8) considering all constraints: $\mu_l^{N_p}\{\omega_{N_p} \in Z_l^{N_p} : \mu_l(V(\omega_{N_1+\dots+N_p})) \leq \epsilon_1\} \geq \beta_1(\epsilon_1, N_p)$, where $V(\omega_{N_1+\dots+N_p}) := \{(x, \mathbf{j}) \in Z_l : \|\mathbf{L}_{\mathbf{p}-1} \mathbf{A}_{\mathbf{j}} \mathbf{L}_{\mathbf{p}-1}^{-1} x\| \leq \lambda_p^*(\omega_{N_1+\dots+N_p})\}$. Since it holds for all sets $\omega_{N_1}, \dots, \omega_{N_{p-1}}$ independent from ω_p , we also have as a corollary that

$$\mu_l^{N_1+\dots+N_p}\{\omega_{N_1+\dots+N_p} \in Z_l^{N_1+\dots+N_p} : \mu_l(V(\omega_{N_1+\dots+N_p})) \leq \epsilon_1\} \geq \beta_1(\epsilon_1, N_p).$$

This result can replace Theorem 4 in [KBJT19]. Then, following the same lines as in [KBJT19] proves Equation (5.10).

Finally, combining Equations (5.9) and (5.10) in the same way as in the proof of Theorem 3.1 concludes the proof. $\square$

Note that the bound of Proposition 5.1 holds for all $\omega_{N_1}, \omega_{N_2}, \dots, \omega_{N_{p-1}}$. Even though it is based on ω_{N_p} only, it is thus to be compared with the bound of Theorem 3.1 based on $N_1 + \dots + N_p$ samples. In practice, for strategic choices of $N_1, N_2, \dots, N_p$ and p ; the bound of Theorem 5.1 beats the one obtained without resorting to adaptive sampling, despite the fact that the number of trajectories considered is smaller. This is due to the difference $\kappa(P^*(\omega_N)) - \kappa(P_p^*(\omega_{N_1+\dots+N_p}))$.

Remark 5.3. The bound derived in Theorem 5.1 is an adaptation of the one of Theorem 3.1. In Chapter 3, we proposed another quadratic bound on the JSR, that is the one of Theorem 3.2. This bound outperforms the one of Theorem 3.1 when the number of samples is low, or when the dimension of the system analyzed is low (see Remark 3.2). We can thus derive another bound on the JSR based on $(\gamma_p^*(\omega_{N_1+N_2+\dots+N_p}), P_p^*(\omega_{N_1+\dots+N_p}))$, the solution to $\text{Opt}_p(\omega_{N_1+\dots+N_p})$, by adapting the bound of Theorem 3.2. For the sake of clarity, we proceed to its derivation in Appendix D, and provide in the same appendix a comparison of the two novel bounds on a numerical experiment.

Bound based on all samples

Deriving a theoretical bound on the JSR based on the solution of $\text{Opt}_p(\omega_{N_1+\dots+N_p})$ and taking into account the information provided by all samples is far from trivial for two reasons. First, the distribution of Ω_p is not uniform, which prevents us from resorting to the classical chance-constrained theorems such as Proposition 2.3. Second, $L_1, \dots, L_{p-1}$ are themselves random quantities if we consider $\omega_{N_1}, \dots, \omega_{N_{p-1}}$ not to be deterministic. Hence, these quantities cannot appear in the expression

of β , the confidence level on the bound. This constitutes a huge difficulty. For instance, we cannot extend our sensitivity analysis straightforwardly (see Appendix E).

In Appendix E, we propose a bound based on all sets of samples by adapting the technique of [VKR12] and [KML16]. They extend the feasibility theory summarized in Proposition 2.3 to multi-stage optimization. We put it in appendix for several reasons. Indeed, it is always less tight than the bound of Theorem 5.1, it only holds for the case where $p = 2$, and it requires the same assumptions on the system as the bound of [BJW21]. From now on, we thus only consider the bound based on the last set of samples solely.

5.4 A variant to the approach

We now present a variant to our approach, decreasing its computational complexity. It is based on the following observation: if we have used enough phases and enough samples, the solution to $\text{Opt}(\mathbf{L}_{\mathbf{p}-1})$, in the last phase, approximates the identity. If this holds, setting P_p to the identity in $\text{Opt}_p(\omega_{N_1+\dots+N_p})$ does not imply a significant rise in the cost $\gamma_p^*(\omega_{N_1+\dots+N_p})$. Hence, because the dimension of the problem is then reduced to one and because of the complete disappearance of the condition number of $P_p^*(\omega_{N_1+\dots+N_p})$, the bound we can obtain on the JSR will be tighter. In this variant, the problem we solve in the last phase reduces to:

$$\min_{\gamma_p \geq 0} \quad \gamma_p \text{ s.t.} \quad \|\mathbf{L}_{\mathbf{p}-1} \mathbf{A}_{\mathbf{j}} \mathbf{L}_{\mathbf{p}-1}^{-1} x\| \leq \gamma_p^l \|x\|, \quad \forall (x, \mathbf{j}) \in \Omega_p \quad (5.11)$$

where L_i is the Cholesky decomposition of the solution to $\text{Opt}_i(\omega_{N_1+\dots+N_i})$ for $i = 1, \dots, p-1$. We denote this problem by $\text{Opt}_{p,I}(\omega_{N_1+\dots+N_p})$ and its solution by $\gamma_{p,I}^*(\omega_{N_1+\dots+N_p})$. Observe that $\gamma_{p,I}^*(\omega_{N_1+\dots+N_p}) = \sqrt[l]{\lambda_p^*(\omega_{N_1+\dots+N_p})}$, where $\lambda_p^*(\omega_{N_1+\dots+N_p})$ is the solution to Problem (5.8). Note that the set of constraints of this problem associated to $\omega_{N_1}, \dots, \omega_{N_{p-1}}$ is satisfied by $\gamma_p = \gamma_{p-1}^*(\omega_{N_1+\dots+N_{p-1}})$. Hence, $\gamma_{p,I}^*(\omega_{N_1+\dots+N_p})$ can easily be computed as:

$$\gamma_{p,I}^*(\omega_{N_1+\dots+N_p}) = \max \left(\gamma_{p-1}^*(\omega_{N_1+\dots+N_{p-1}}), \max_{(x, \mathbf{j}) \in \omega_{N_2}} \sqrt[l]{\|\mathbf{L}_{\mathbf{p}-1} \mathbf{A}_{\mathbf{j}} \mathbf{L}_{\mathbf{p}-1}^{-1} x\|} \right).$$

For $p = 2$, this approach is similar the fast algorithm, except for the sampling in the second phase. Indeed, we sample the second set of trajectories uniformly on $E_{P_1^*(\omega_{N_1})}$ instead of on the sphere. Similarly to the case where P_p is not set to the identity, we propose a bound on the JSR based on the N_p last points sampled solely:

Theorem 5.2

Let Assumptions 5.1 and 5.2 hold, and let $\gamma_{p,I}^*(\omega_{N_1+\dots+N_p})$ be the optimal solution to $\text{Opt}_{p,I}(\omega_{N_1+\dots+N_p})$. Then, for any $\bar{\beta}_1 \in [0, 1]$, with probability at least $\bar{\beta}_1$, we have:

$$\rho(\mathcal{M}) \leq \frac{\gamma_{p,I}^*(\omega_{N_1+\dots+N_p})}{\sqrt[l]{\delta(\epsilon_1(\bar{\beta}_1, m, N_p))}} \quad (5.12)$$

where $\epsilon_1(\bar{\beta}_1, m, N_p) = \frac{m^l}{2}(1 - \sqrt[l]{1 - \bar{\beta}_1})$ and $\delta(\cdot)$ is given by (3.7).

Proof. Problem (5.11) is a sampled version of the following problem:

$$\min_{\gamma' \geq 0} \quad \gamma' \text{ s.t. } \quad \|\mathbf{L}_{\mathbf{p}-1}\mathbf{A}_{\mathbf{j}}\mathbf{L}_{\mathbf{p}-1}^{-1}x\| \leq \gamma' \|x\|, \quad \forall x \in \mathbb{S}, \quad \forall \mathbf{A}_{\mathbf{j}} \in \mathcal{M}^l$$

which is convex and admits a sample $(x, \mathbf{j})$ as an essential set, namely the sample such that $\gamma_{p,I}^*(\omega_{N_1+\dots+N_p}) = \sqrt[l]{\|\mathbf{L}_{\mathbf{p}-1}\mathbf{A}_{\mathbf{j}}\mathbf{L}_{\mathbf{p}-1}^{-1}x\|}$. Hence, if we only consider the constraints associated to the uniform sampling ω_{N_p} in $\text{Opt}_{p,I}(\omega_{N_1+\dots+N_p})$ and obtain a solution $\gamma_{p,I}^{*l}(\omega_{N_1+\dots+N_p})$, the same argument as in the proof of Proposition 3.3 ensures that for all $\epsilon_1 \in (0, 1]$, we have:

$$\mu_l^{N_p} \{\omega_{N_p} \in Z_l^{N_p} : \mu_l(\tilde{V}'(\omega_{N_p})) \leq \epsilon_1\} \geq \beta_1(\epsilon_1, N_p)$$

where $\beta_1(\epsilon_1, N_p) = 1 - (1 - \epsilon_1)^{N_p}$ and $\tilde{V}'(\omega_{N_p}) := \{(x, \mathbf{j}) \in Z_l : \|\mathbf{L}_{\mathbf{p}-1}\mathbf{A}_{\mathbf{j}}\mathbf{L}_{\mathbf{p}-1}^{-1}x\| > \gamma_{p,I}^{*l}(\omega_{N_1+\dots+N_p})\}$ is the set of violated constraints. By the same argument as in the proof of Proposition 5.1, we obtain as a corollary that $\mu_l^{N_1+\dots+N_p} \{\omega_{N_1+\dots+N_p} \in Z_l^{N_1+\dots+N_p} : \mu_l(\tilde{V}(\omega_{N_1+\dots+N_p})) \leq \bar{\epsilon}_1\} \geq \beta_1(\epsilon_1, N_p)$, where $\tilde{V}(\omega_{N_1+\dots+N_p}) := \{(x, \mathbf{j}) \in Z_l : \|\mathbf{L}_{\mathbf{p}-1}\mathbf{A}_{\mathbf{j}}\mathbf{L}_{\mathbf{p}-1}^{-1}x\| > \gamma_{p,I}^{*l}(\omega_{N_1+\dots+N_p})\}$, which replaces Theorem 4 in [KBJT19]. Then, following the same lines as in [KBJT19] concludes the proof. $\square$

Remark 5.4. For the bound of Proposition 5.2 to be compared with the bound of Proposition 5.1, one must set $\bar{\beta}_1$ to $\beta + \beta_1 - 1$, so that both bounds hold with the same confidence level.

Remark 5.5. For an identical number of samples N_p used in the last phase of the approach and in the context of adaptive sampling, the bound of Theorem 5.2 almost surely outperforms the one the bound of Theorem 5.1. We provide an explanation for this success in Appendix F.

5.5 Summary of the approaches based on adaptive sampling

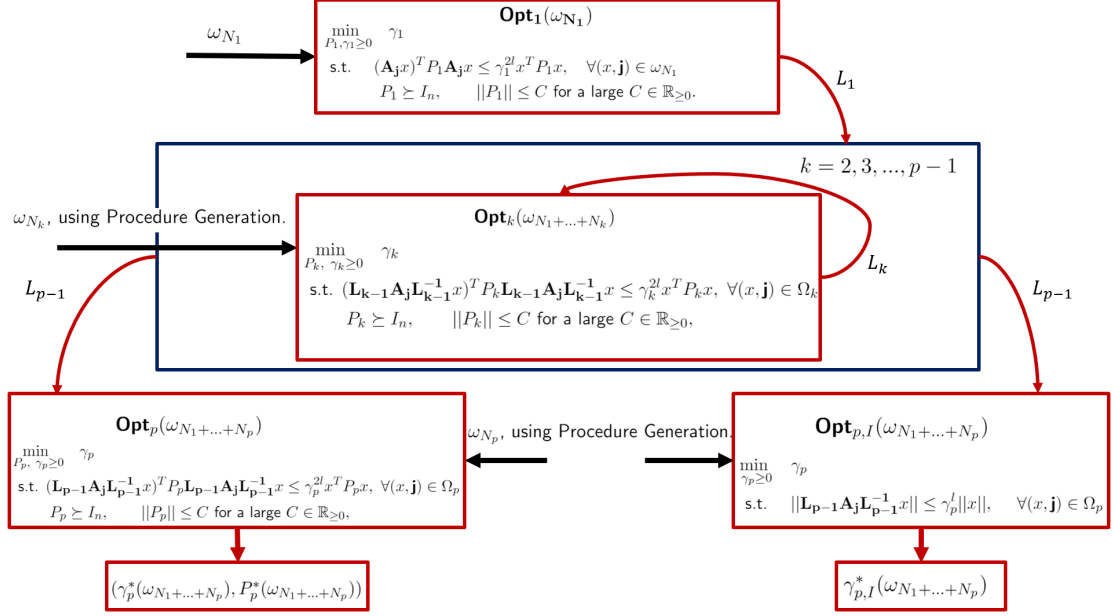


Figure 5.3: Summary of the procedure to obtain a probabilistic bound on the JSR using our adaptive sampling approach. The blue box contains the iterations $k = 2, \dots, p-1$, that is the phases refining the change of variables. After each phase, the change of variables computed is handed to the next phase, as illustrated in red. The external resources provided to the phases, that is the sampling $\omega_{N_1}, \dots, \omega_{N_p}$ are presented in black. In the last phase, one either fix P_p to the identity (right-hand side) or not (left-hand side). This is determined before starting the computations.

Figure 5.3 illustrates the steps required by our adaptive sampling approach to obtain a bound on the JSR. In the last phase, we choose whether or not we fix P_p to the identity. Note that this choice must be taken in advance, and that we cannot simply take the minimum upper bound obtained, since both are probabilistic and only hold with a certain confidence level. Hence, the minimum between the two bounds holds with a lower confidence level. As observed in Remark 5.5, in our adaptive sampling setting, if a sufficiently high number of samples is available, the bound of Theorem 5.2 will almost surely beat the one of Theorem 5.1.

Chapter 6

Numerical experiments and derivation of a heuristic

In this chapter, we seek to optimize our adaptive sampling approach by answering the following question:

Given a total number N of samples available, what is the optimal number of phases p and what is the optimal number of samples $N_1, N_2, \dots, N_p$ to use in each phase?

In Chapter 5, we were only able to provide bounds based on the analysis of the N_p last samples solely. Hence, we cannot directly optimize these bounds to obtain the values of $N_1, \dots, N_p$ and p yielding the tightest bounds on the JSR possible. Therefore, we will determine these optimal values empirically.

First, we propose a strategy, called the *online* strategy, determining the optimal number of samples to use in each phase based on the convergence of $P_k(\omega_{N_1+\dots+N_k})$ to a certain limit. We then compare its performance to what can be obtained by fixing in advance, in an *offline* way, $N_1, \dots, N_p$ and p to judicious values. This allows to empirically confirm that the convergence criterion used by the online strategy yields tight bounds on the JSR. Additionally, it also illustrates how our adaptive sampling approach outperforms the fast algorithm and, for judicious choices of $N_1, \dots, N_p$ and p , the bounds of Chapter 3. In a third time, we conduct a large number of experiments to determine, using online sampling, the optimal values of $N_1, \dots, N_p$ and p to be used in offline strategy.

The information yielded by the experiments allows us to propose a heuristic providing tighter bounds on the JSR than the ones of Chapter 3, holding with the same confidence level and requiring the same sample complexity.

Finally, we illustrate the performance of the heuristic on various systems.

6.1 Online and offline strategies

We present the features of the offline and the online strategies.

- In the offline strategy, we fix the values of $N_1, \dots, N_p$ and p in advance and obtain the solution $(\gamma_p^*(\omega_{N_1+\dots+N_p}), P_p^*(\omega_{N_1+\dots+N_p}))$ by resorting to Algorithm 3. The parameters are thus deterministic.
- In the online strategy, we determine $N_1, \dots, N_p$ and p after the computations started, by resorting to a convergence criterion presented in Procedure Online. In each phase, we first sample N_{init} trajectories uniformly in the coordinate system relevant to the phase, and compute the solution $P_k^*(\omega_{N_1+\dots+N_{init}})$ to $\text{Opt}_k(\omega_{N_1+\dots+N_{init}})$. Then, iteratively, we add a small number n of samples to the total number N of samples used in the phase, and recompute the solution after each increment. The process keeps going until convergence is attained, that is until $\|P_k^*(\omega_{N_1+\dots+N}) - P_k^*(\omega_{N_1+\dots+(N-n)})\|$ is sufficiently small.

Procedure Online determination of N_k	
Input: Initial number of samples, increment n , tolerance tol , phase k .	
Output: Number of samples generated N_k , solution $(\gamma_k^*(\omega_{N_1+\dots+N_k}), P_k^*(\omega_{N_1+\dots+N_k}))$.	
1	Initialization: $P_{old} = I_n, N_k = N_{init}$;
2	Sampling of ω_{N_k} uniformly for System $\mathbf{L}_{k-1}\mathcal{M}^l\mathbf{L}_{k-1}^{-1}$, where $ \omega_{N_k} = N_k$, using Procedure Generation if $k \geq 2$;
3	Computation of $(\gamma_k^*(\omega_{N_1+\dots+N_k}), P_k^*(\omega_{N_1+\dots+N_k}))$, solution to $\text{Opt}_k(\omega_{N_1+\dots+N_k})$;
4	while $\ P_{old} - P_k^*(\omega_{N_1+\dots+N_k})\ > tol$ do
5	Sampling of ω_n uniformly for System $\mathbf{L}_{k-1}\mathcal{M}^l\mathbf{L}_{k-1}^{-1}$, where $ \omega_n = n$, using Procedure Generation if $k \geq 2$;
6	$P_{old} = P_k^*(\omega_{N_1+\dots+N_k}), N_k = N_k + n, \omega_N = \omega_N \cup \omega_n$;
7	Computation of $(\gamma_k^*(\omega_{N_1+\dots+N_k}), P_k^*(\omega_{N_1+\dots+N_k}))$, solution to $\text{Opt}_k(\omega_{N_1+\dots+N_k})$;
8	end

The number of phases p is determined by the total number of samples available or by a parameter $pmax$, as presented in Algorithm 4. Setting a maximum number of phases allows to increase N_p and hence decrease the bound of Theorem 5.1. In the last phase, to decrease the total computational complexity, we directly use all the remaining samples available without resorting to Procedure Online. Note that $N_1, \dots, N_p$ and p are random variables.

The online strategy can only provide useful information when the available number of samples is sufficiently high for Algorithm 4 to reach Phase 3.

Algorithm 4: Online adaptive sampling algorithm.

Input: Total number of samples to be used N , maximum number of phases $pmax$.

Result: $(\gamma_p^*(\omega_{N_1+\dots+N_p}), P_p^*(\omega_{N_1+\dots+N_p}))$

First Phase

- 1 Determination of N_1 and $(\gamma_1^*(\omega_{N_1}), P_1^*(\omega_{N_1}))$, solution to $\text{Opt}_1(\omega_{N_1})$, using Procedure Online;
- 2 Computation of the change of variables L_1 such that $P_1^*(\omega_{N_1}) = L_1^T L_1$;
- 3 $k = 2$;

Second set of phases

- 4 **while** $N_1 + \dots + N_{k-1} \leq N - d_{quad} - 1$ **and** $k \leq pmax$ **do**
- 5 Determination of N_k and $(\gamma_k^*(\omega_{N_1+\dots+N_k}), P_k^*(\omega_{N_1+\dots+N_k}))$, solution to $\text{Opt}_k(\omega_{N_1+\dots+N_k})$, using Procedure Online;
- 6 Computation of the change of variables L_k such that $P_k^*(\omega_{N_1+\dots+N_k}) = L_k^T L_k$;
- 7 Computation of $\mathbf{L}_k = L_k \mathbf{L}_{k-1}$ and $\mathbf{L}_k^{-1} = L_k^{-1} \mathbf{L}_{k-1}^{-1}$;
- 8 $k=k+1$;
- 9 **end**

Last phase

- 10 $p = k$, $N_p = N - (N_1 + \dots + N_{k-1})$;
- 11 Computation of $(\gamma_p^*(\omega_{N_1+\dots+N_p}), P_p^*(\omega_{N_1+\dots+N_p}))$, solution to $\text{Opt}_p(\omega_{N_1+\dots+N_p})$, without using Procedure Online;

Note that the online strategy is computationally expensive, especially for high-dimensional systems. However, it requires no a priori intuition about the optimal values of $N_1, \dots, N_p$ and p . We will now illustrate on numerical experiments that it yields tight bounds on the JSR. This confirms the validity of the convergence criterion to determine the optimal number of samples to use in each phase.

Notation and experimental protocol

We denote by $(\gamma^*(\omega_N), P^*(\omega_N))$ the solution to Problem (3.2), where $N = N_1 + \dots + N_p$, that is the solution obtained without resorting to adaptive sampling. It is with the bound of Theorem 3.1 based on this solution that we compare our adaptive sampling bounds.

The default values of the parameters used in the experiments are the ones of Chapter 4, summarized in Table 4.1. In the offline strategy, by default, we fix a number of samples to be used in the first phase and divide equally the remaining samples between the next phases. Table 6.1 contains the default values of the

parameters used in Procedure Online and in Algorithm 4 when resorting to the online strategy:

Initial number of samples used in Procedure Online in Phase 1	$d_{quad} + 1$
Increment of samples at each iteration of Procedure Online	$n - 2$
Tolerance on the convergence of $P_k(\omega_{N_1+\dots+N_k})$ in Procedure Online	10^{-5}
Maximal number of phases allowed to be reached in Algorithm 4	11

Table 6.1: Default values of the parameters used in Procedure Online and Algorithm 4 for a n -dimensional system.

6.2 Comparison of offline and online sampling on Example 2.1

In Chapter 4, we observed that for high-dimensional systems, due to the rise of $\kappa(P^*(\omega_N)) = \sqrt{\frac{\lambda_{\max}(P^*(\omega_N))}{\lambda_{\min}(P^*(\omega_N))}}$, the number of samples required to obtain relatively tight bounds becomes large. For instance, for the 8-dimensional system of Example 2.1, 4000 samples are required to obtain a bound smaller than 250% of the true value of the JSR. We thus analyze the performance of our approach on this system.

6.2.1 Performance of offline sampling

First, we analyze the benefits of the offline strategy and try to determine empirically, given a total number N of samples to be used, how to divide them between several phases. Clearly, there is a trade-off to achieve between

- (i) the quality of the change of variables measured by $\Delta_{s,p-1} = \|P'^o(\mathbf{L}_{\mathbf{p}-2}) - P_{p-1}^*(\omega_{N_1+\dots+N_p})\|$, increasing with the number of samples $N_1, \dots, N_{p-1}$ used in the first phases and the number of phases (see Remark 5.2).
- (ii) the benefits of increasing the number of samples N_p used in the last phase. First, the distance $\Delta_{s,p} = \|P'^o(\mathbf{L}_{\mathbf{k}-1}) - P_p^*(\omega_{N_1+\dots+N_p})\|$ decreases as N_p increases. Second, N_p appears in the theoretical bounds of Theorems 5.1 and 5.2, and both bounds decrease as N_p increases.

For a total number of samples fixed to 500, we plot the evolution of $\kappa(P_p^*(\omega_{N_1+\dots+N_p}))$ and of the upper bounds on the JSR with the ratio of samples used in the first phase. As expected, we observe on Figure 6.1 that $\kappa(P_p^*(\omega_{N_1+\dots+N_p}))$ presents a U-shape. At first, it decreases significantly when the ratio of samples used in the first phase increases, since the quality of the change of variables then increases.

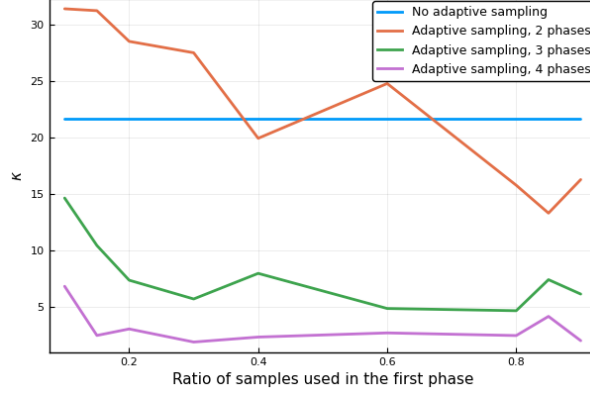


Figure 6.1: Evolution of $\kappa(P_p^*(\omega_{N_1+\dots+N_p}))$ with the ratio of samples used in the first phase, for the system of Example 2.1. In total, 500 trajectories are sampled. $\kappa(P_p^*(\omega_{N_1+\dots+N_p}))$ presents a U-shape and decreases when p increases.

However, when this ratio becomes too high, $\kappa(P_p^*(\omega_{N_1+\dots+N_p}))$ increases again. N_p is indeed too small for $P_p^*(\omega_{N_1+\dots+N_p})$ to be close to $P^{lo}(\mathbf{L}_{\mathbf{p}-1})$, even when the latter is close to the identity. We also observe that $\kappa(P_p^*(\omega_{N_1+\dots+N_p}))$ decreases with the number of phases. When more than 2 phases are used, it is always smaller than $\kappa(P^*(\omega_N))$.

In Figures 6.2, 6.3 and 6.4, we plot the bounds of Theorems 5.1 and 5.2, the true value of the JSR of the system as approximated by the JSR toolbox [LJP16], and the bound given in Proposition 5.1, referring to the fast algorithm. For the sake of clarity, the different bounds are only presented together when two phases are used, on Figure 6.4. Their evolution with the number of phases is presented on separate figures, that is Figures 6.2 and 6.3.

We observe that **all the adaptive sampling bounds decrease with the number of phases, and that they also present a U-shape**. This feature is accentuated since N_p directly appears in the expression of our bounds. Hence, when the ratio of samples used in the first phase increases, even if $\kappa(P_p^*(\omega_{N_1+\dots+N_p}))$ is low, the bounds become less tight. Our new bound of Theorem 5.1 outperforms the one of Theorem 3.1 for a sufficiently high number of phases, and the gap grows for a judicious choice of the ratio of samples used in the first phase. The bounds of Theorem 5.2 always outperform all other bounds as expected given Remark 5.5. Finally, we observe that **the bound resulting from the application of the fast algorithm is significantly larger than our bounds**.

In conclusion, determining judicious values for $N_1, \dots, N_p$ and p is crucial for the bounds yielded by the offline sampling to be as tight as possible. We aim at determining the parameters yielding the minimum values of the U-shape for all types of switched linear systems.

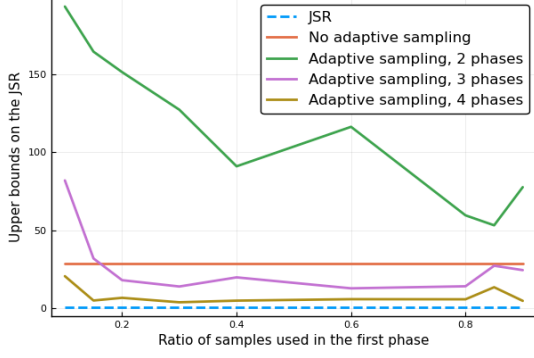


Figure 6.2: Evolution of the upper bounds of Theorem 5.1 with the ratio of samples used in the first phase, for the system of Example 2.1. For $p \geq 2$, the bounds always outperform the one of Theorem 3.1 (no adaptive sampling).

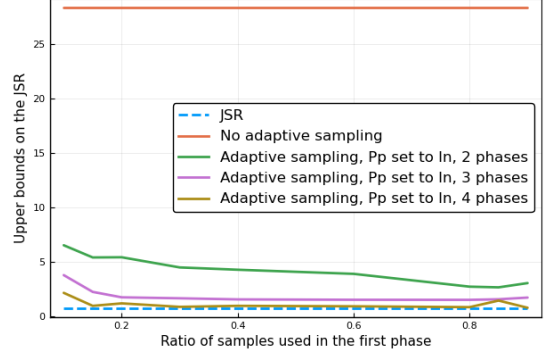


Figure 6.3: Evolution of the upper bounds of Theorem 5.2 (for P_p fixed to the identity) with the ratio of samples used in the first phase, for the system of Example 2.1. The bounds always outperform the one of Theorem 3.1.

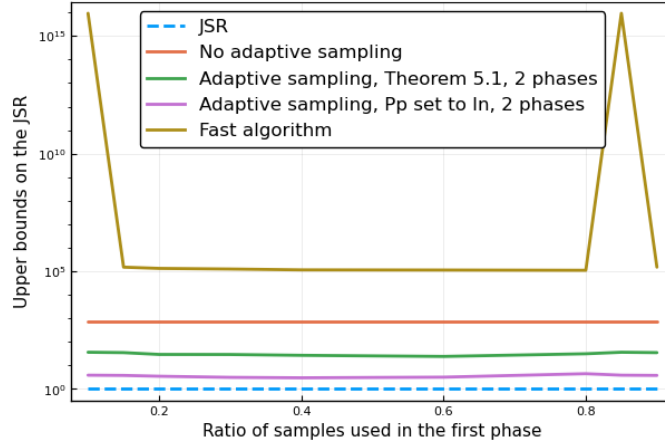


Figure 6.4: Evolution of several upper bounds with the ratio of samples used in the first phase, for the system of Example 2.1. The bounds of Theorems 5.1 and 5.2 (two-phase adaptive sampling), of Proposition 5.1 (fast algorithm), of Theorem 3.1 (no adaptive sampling), and the true value of the JSR are plotted.

6.2.2 Performance of online sampling

We now show that the convergence criterion defined in Procedure Online is valid to ensure the tightness of the bounds of Theorems 5.1 and 5.2. Hence, the number of trajectories to be sampled in each phase as determined by the online strategy can be used as optimal values in the offline strategy.

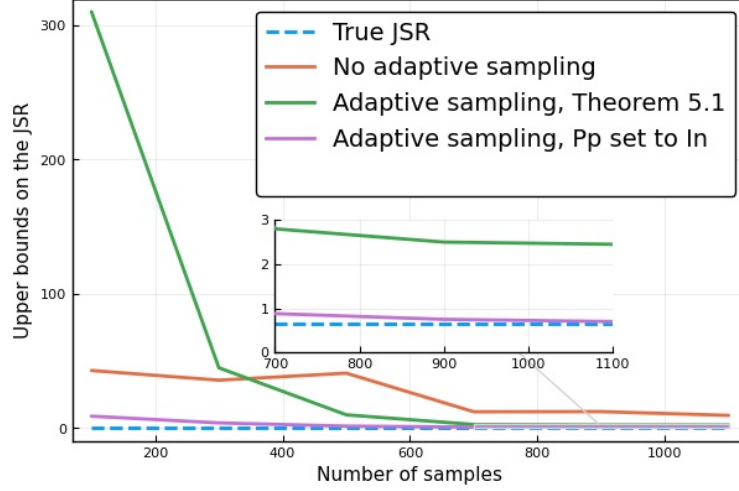


Figure 6.5: Evolution of several upper bounds with the total number of samples, for the system of Example 2.1. The bounds of Theorems 5.1 and 5.2, the true value of the JSR of the system and the bound of Theorem 3.1 are plotted. The bound of Theorem 5.2 is the tightest one. The bounds related to adaptive sampling beat the bound of Theorem 3.1 for a sufficiently high number of samples.

Figure 6.5 presents the evolution of the bounds of Theorems 5.1 and 5.2, depending on whether or not P_p is fixed to the identity, with the number of samples. The true value of the JSR of the system as approximated by the JSR toolbox [LJP16] is also plotted. As expected, the bound of Theorem 5.2 is the tightest one. It only requires 750 samples to become smaller than 1 and hence serve as a stability certificate. We observe that for a small total number of samples, implying that the online adaptive sampling strategy does not reach Phase 3, the bound of Theorem 5.1 does not beat the one obtained without resorting to adaptive sampling. However, when the number of samples increases, it rapidly outperforms the old bound, and provides tight approximations of the JSR. Observe that **for 500 samples used, the order of magnitude of the bounds obtained via online and offline sampling, for a judicious ratio of samples used in the first phase, is equivalent.** Thus, we infer that the convergence criterion we are using to determine how many samples to use in each phase is valid when N is sufficiently high. More insight with respect to the evolution of $P_k(\omega_{N_1+\dots+N_k})$ with the number of phases k and to the number of samples required in each phase is provided in Appendix G.

6.2.3 A computational comparison of the two strategies

We observed that the sample complexity required by both the offline and the online strategies to obtain tight bounds is significantly lower than without resorting to adaptive sampling. However, in the offline strategy, this only holds for a judicious choice of the parameters $N_1, \dots, N_p$ and p of the method. We now answer the following question: if the bounds yielded by the online strategy are always as tight as possible (at least when the number of samples available is high enough), why do we want to rely on the offline strategy? The answer is twofold.

1. Resorting to offline sampling prevents situations where all available samples are used in the first phase. This can occur in online sampling, when the number of samples available is not sufficiently large.
2. In the online sampling strategy, the reduced sample complexity does not imply a reduced computational complexity, which was one of the objectives of resorting to adaptive sampling. Table 6.2 presents the sample and computational complexities required to obtain bounds smaller than 250% of the true value of the JSR by three approaches: no resorting to adaptive sampling, offline sampling or online sampling. When resorting to offline sampling, we use as an input the results obtained via the online strategy, so that the same number of samples is used in each phase of both methods. The derivation of Table 6.2 is presented in Appendix H.

	No adaptive sampling	Offline strategy	Online strategy
Total number of samples required	4000	500	500
Number of phases	1	5	5
Number of samples used in each phase	$N_1 = 4000$	$N_1 = 253, N_2 = 90, N_3 = 24$ $N_4 = 30, N_5 = 103$	
Computational complexity	$\left\lceil \log_2\left(\frac{U}{\alpha}\right) \right\rceil * \text{poly}(4000, d_{quad})$	$\left\lceil \log_2\left(\frac{U}{\alpha}\right) \right\rceil * \text{poly}(1860, d_{quad})$	$\left\lceil \log_2\left(\frac{U}{\alpha}\right) \right\rceil * \text{poly}(14736, d_{quad})$

Table 6.2: Sample and computational complexities of the different strategies, required to obtain a bound smaller than 250% of the JSR of the system of Example 2.1. $[0, U]$ is the interval over which we look for γ (or γ_k , for $k = 1, 2, \dots, p$) by bisection with an accuracy α . The expression $\text{poly}(x, y)$ denotes that the computational complexity is polynomial in x and y .

The computational complexity of the online strategy significantly exceeds the one of the original method, while the contrary is true

for the offline sampling method. Hence, resorting to offline sampling with judiciously chosen values for $N_1, \dots, N_P$ and p allows to take advantage of the adaptive sampling approach on high dimensional systems, for which the computational complexity of the online strategy is too high.

6.3 Derivation of a heuristic optimizing the adaptive sampling approach

We now provide a heuristic combining the assets of online sampling with the low computational complexity of the offline strategy. It determines the optimal number of phases p , as well as the the number of samples to use in each phase, for systems of any dimension and for any total number of samples available. We have asserted that for a sufficiently high number of samples, allowing Algorithm 4 to reach Phase 3, the criterion defined in Procedure Online determining the optimal number of samples to use in each phase yields accurate bounds on the JSR. Hence, we base ourselves on results obtained using the online strategy when a sufficiently high number of samples is available, and on results obtained using offline sampling otherwise.

6.3.1 For a large number of samples available: online sampling

We answer the following questions.

- For systems of any dimension, what is the number of samples required in each phase for the convergence of $P_p^*(\omega_{N_1+\dots+N_p})$ as defined by Procedure Online?
- What is the optimal number of phases to be used? We observe empirically that for some $k \geq 2$, $P_k^*(\omega_{N_1+\dots+N_k})$ reaches a limit close to the identity. From then, at each phase, Procedure Online only requires one or two iterations to converge, and $P_{k+1}^*(\omega_{N_1+\dots+N_{k+1}})$, $P_{k+2}^*(\omega_{N_1+\dots+N_{k+2}})$, ..., do not vary anymore. Clearly, this specific k can be used as the maximum number of phases to be used. Indeed, then, recomputing changes of variables will not decrease $\kappa(P_p^*(\omega_{N_1+\dots+N_p}))$ much, while using more samples in the last phase will increase the accuracy of the bounds of Theorems 5.1 and 5.2.

We consider 2-modal systems of dimension ranging between 3 and 30, sampled randomly and uniformly on the sphere. We first analyze the number of samples required for the convergence of $P_k(\omega_{N_1+\dots+N_k})$, as defined in Procedure Online. Figure 6.6 presents the number of samples required for the convergence of $P_1(\omega_{N_1})$

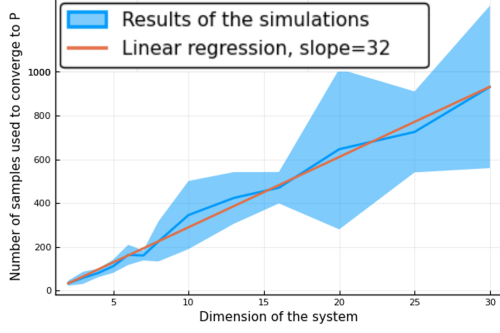


Figure 6.6: Evolution of the number of samples needed for the convergence of $P_1(\omega_{N_1})$ as defined by Procedure Online with the dimension of the system. The blue curve presents the average number of samples required and the blue ribbon is the variance of the experiments. A linear interpolation whose slope is equal to 32 fits the data well.

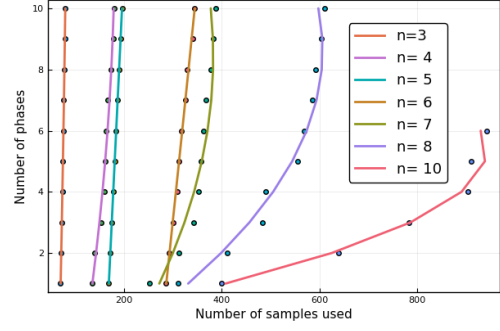


Figure 6.7: Evolution of the number of phases reached by Algorithm 4 with the total number of samples, for systems of dimension ranging between 3 and 10. In each phase, the number of samples required is determined by Procedure Online. The dots present the average of the observed data, and the curves are interpolations of degree 2.

for systems of various dimensions. We plot the average and the variance of the experiments, and we observe that a linear regression fits the data well. Hence, we infer that the number of samples required for the convergence of $P_1(\omega_{N_1})$ is equal to $32 * (n - 1)$, where n is the dimension of the system. We also observe on Figure 6.7 that in the next phases, the number of samples required for the convergence decreases, in a quadratic way for systems of sufficiently high dimension. For systems of lower dimension, from Phase 4, convergence is often reached after one or two iterations, which implies that the optimal number of phases is reached. We approximate the optimal values of $N_2, N_3, \dots$ similarly to N_1 , by linearly interpolating data yielded by our experiments. The coefficients of the linear regression, for each phase, are given in Table 6.3.

Phase	1	2	3	4	5	6	7
a	32	30	20	20	10	10	20
b	-32	-150	-80	-100	-10	-60	-160

Table 6.3: Coefficients of the linear regression $N_k = a * n + b$, where n is the dimension of the system and N_k is the number of samples required to reach convergence in phase k , as determined by Procedure Online.

From Phase 3, the coefficients are obtained based on the observation of systems of di-

mension larger than 6, since for smaller dimensions, convergence of $\kappa(P_k^*(\omega_{N_1+\dots+N_k}))$ to a final limit is then often already reached. From Phase 2, these coefficients yield negative values for systems of small dimension. In this case, we fix the optimal number of samples to be used in each phase to $d_{quad} + 1$, the dimension of the problem.

It now remains to determine the optimal number p of phases to use. Therefore, we analyze the number of phases required for $P_k^*(\omega_{N_1+\dots+N_k})$ and $P'^o(\mathbf{L}_{k-1})$ (unknown in our setting) to converge to a limit close to the identity.

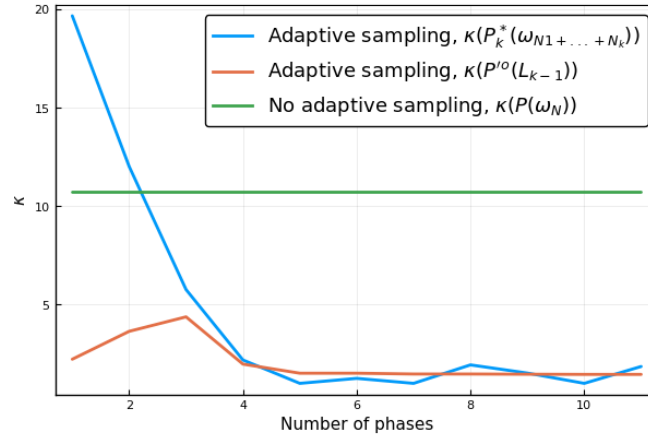


Figure 6.8: Evolution of $\kappa(P_k^*(\omega_{N_1+\dots+N_k}))$ and $\kappa(P'^o(\mathbf{L}_{k-1}))$ with the number of phases, for a 4-dimensional. $\kappa(P_k^*(\omega_{N_1+\dots+N_k}))$ decreases at each phase, and reaches a limit from Phase 4.

For instance, Figure 6.8 plots the evolution of $\kappa(P_k^*(\omega_{N_1+\dots+N_k}))$ and $\kappa(P'^o(\mathbf{L}_{k-1}))$ with the number of phases for a 4-dimensional system. The maximum number of phases to be used for such systems is of 4. By analyzing the evolution of $P_k^*(\omega_{N_1+\dots+N_k})$ and $P'^o(\mathbf{L}_{k-1})$ for systems of various dimensions, we derive the maximum values of p presented in Table 6.4.

Dimension of the system	$1 \leq n \leq 6$	$7 \leq n \leq 10$	$11 \leq n$
Optimal number of phases	4	6	10

Table 6.4: Maximum number of phases to use, depending on the dimension. This is the optimal number of phases when the number of samples available allows it.

Let N_{opt} be the sum of the optimal number of samples to be used in each phase as determined by Table 6.3 for 4, 6 or 10 phases. Note that N_{opt} is much lower than what is needed to obtain tight bounds on the JSR without resorting to adaptive sampling. If the total number of samples available is smaller than N_{opt} , we decrease the number of phases until at least $d_{quad} + 1$ samples can be used in the last phase.

6.3.2 For a low number of samples available: offline sampling

Suppose that the total number of samples available is smaller than $N_1 + N_2 + d_{quad} + 1$, where N_1 and N_2 are the optimal numbers of samples to use in Phases 1 and 2 as determined by Table 6.3. Then, as we have observed in Section 5.7.1, we cannot rely on online sampling to obtain tight bounds on the JSR.

We have observed in the offline sampling strategy that increasing the number of phases yields tighter bounds in the JSR. We show that this holds as long as $d_{quad} + 1$ samples are used in the last phase. We analyze the system of Example 4.2, and fix the total number of samples to 80. Figure 6.9 presents the evolution of the bounds of Theorem 5.2 with the ratio of samples used in the first phase.

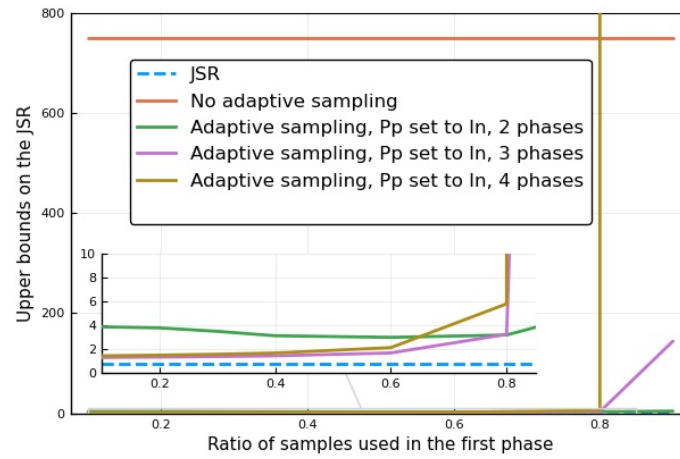


Figure 6.9: Evolution of the upper bounds of Theorem 5.2 (for P_p fixed to the identity) with the ratio of samples used in the first phase, for the system of Example 4.2. In total, 80 trajectories are sampled. The number of phases ranges from 2 to 4. When less than $d_{quad} + 1$ samples are used in the last phase, the tightness of the bounds decreases.

In this setting, when the ratio of samples used in the first phase increases, for a large number of phases, the number of samples to be used in the last phase falls below $d_{quad} + 1$. Then, the bounds obtained increase and do not beat the bound obtained without resorting to adaptive sampling anymore. Hence, if the total number of samples is too low to resort to the online sampling results, we simply use $d_{quad} + 1$ samples in each of the first phases.

6.3.3 Summary of the heuristic

Heuristic 1 presents our method decreasing both the sample and the computational complexity required to obtain tight probabilistic bounds on the JSR of systems of any dimension. It provides the values of the parameters $N_1, \dots, N_p$ and p to be used in Algorithm 4 as a function of the dimension n of the system and of the total number N of available samples.

Heuristic 1: Determination of the optimal values of the parameters of the offline strategy	
Input: Total number of samples to be used N, dimension of the system n, Table 6.3 containing all coefficients a_k and b_k. Result: Input of Algorithm 3: number of phases p, number of samples to be used in each phase: $N_1, N_2, \dots, N_p$.	
<pre> 1 if $n \leq 5$ then 2 $p = 5$; 3 $N_1 = a_1 * n - b_1$; 4 $N_2 = N_3 = N_4 = d_{quad} + 1$; 5 else if $n \leq 10$ then 6 $p = 6$; 7 $N_k = a_k * n + b_k$ for $k = 1, \dots, p - 1$; 8 else 9 $p = 10$; 10 $N_k = a_k * n + b_k$ for $k = 1, \dots, 7$, $N_k = d_{quad+1}$ for $k = 8, 9$; 11 end 12 $N_p = N - \sum_{i=1}^4 N_i$; 13 while $N_p \leq d_{quad+1}$ do 14 $p = p - 1$; 15 $N_p = N - \sum_{i=1}^{p-1} N_i$; 16 end 17 if $p < 5$ then 18 $p = \max_{i=2, \dots, 5} i$ s.t. $N - i * (d_{quad} + 1) > 0$; 19 $N_1 = N - (p - 1) * (d_{quad+1})$, $N_i = d_{quad+1}$ for $i = 2, \dots, p$; 20 end </pre>	

6.4 Illustration of the performance of the heuristic

We illustrate the performance of the heuristic on several numerical experiments. First, we fix the total number of samples available to 1500, and compare the bounds yielded by the application of Heuristic 1 to the one obtained without resorting to adaptive sampling on systems of various dimension. Figure 6.10 presents the evolution of the different bounds and of the true value of the JSR as approximated by the JSR toolbox [LJP16] with the dimension of the systems analyzed.

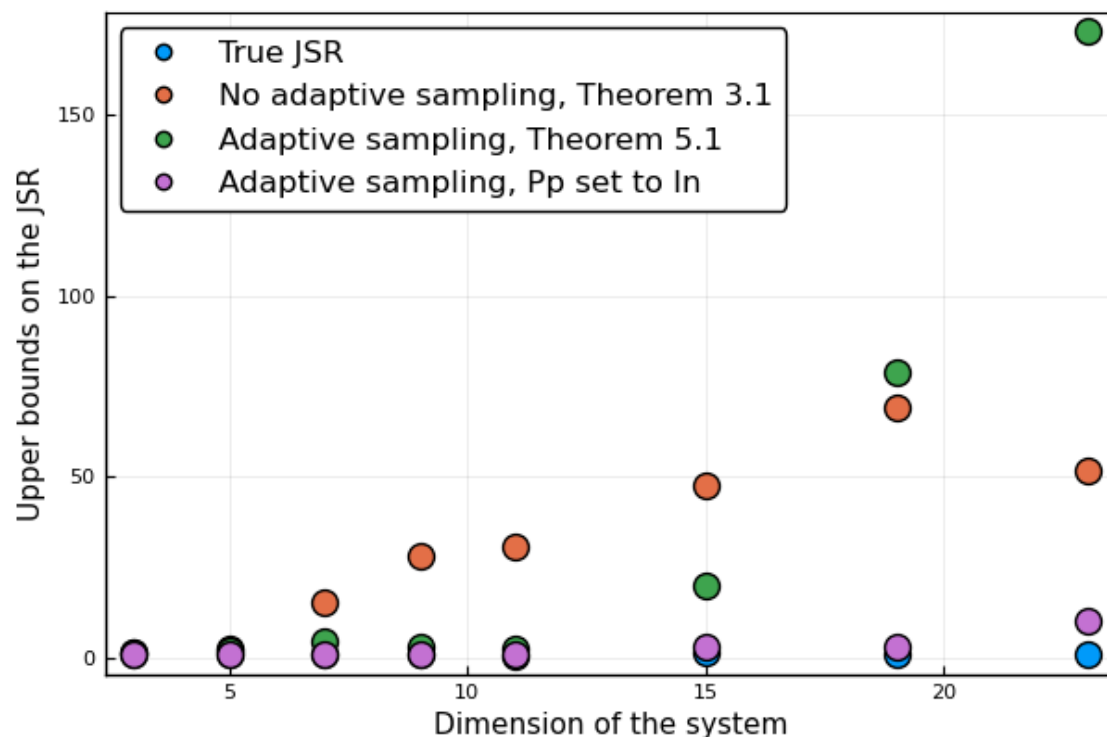


Figure 6.10: Evolution of several upper bounds with the dimension of the system. The bounds based on adaptive sampling are yielded by the application of Heuristic 1 on 4-modals systems of various dimensions, for a total number of samples fixed to 1500. The bounds of Theorems 5.1 and 5.2 (for P_p set to the identity), of Theorem 3.1 (no adaptive sampling) and the true value of the JSR are plotted. The bound of Theorem 5.2 always outperforms all bounds. When the dimension increases, the number of samples available is not sufficient to allow the bound of Theorem 5.1 to be tighter than the one of Theorem 3.1.

We observe that for systems of dimension smaller than 20, the bounds yielded by the heuristic outperform the one of Theorem 3.1. However, for systems of larger

dimensions, the bound of Theorem 5.1 grows rapidly and becomes larger than the one of Theorem 3.1. As expected, the bound obtained when fixing P_p to the identity always beats all other bounds. However, when the dimension increases, it rapidly exceeds the true value of the JSR by more than 80% of its value. Hence, it cannot serve as stability certificate anymore for systems whose JSR is smaller than 1.

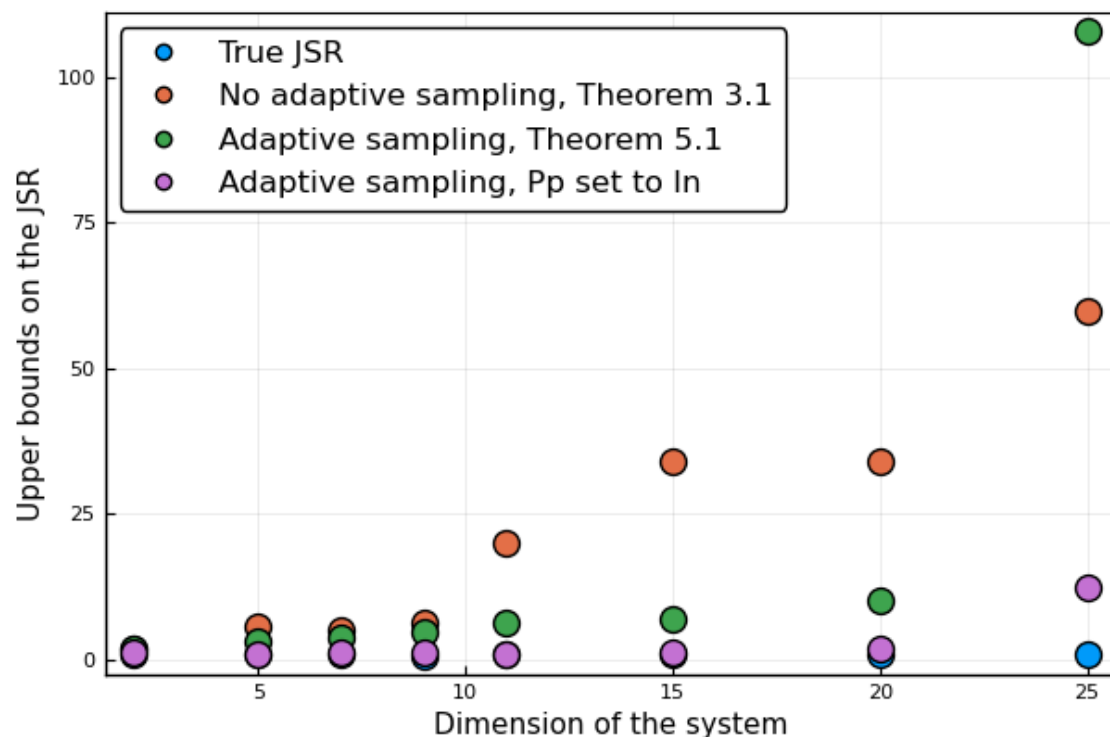


Figure 6.11: Evolution of several upper bounds with the dimension of the system. The bounds based on adaptive sampling are yielded by the application of Heuristic 1 on 4-modals systems of various dimensions n , for a total number of samples fixed to $100n$. The bounds of Theorems 5.1 and 5.2, of Theorem 3.1 and the true value of the JSR are plotted. The bound of Theorem 5.2 is the tightest. For systems of dimension larger than 25, the number of samples available is not sufficient to allow the bound of Theorem 5.1 to be tighter than the one of Theorem 3.1.

Clearly, in higher dimensions, a larger total number of samples is required to obtain tight bounds on the JSR. Hence, we repeat our experiment on Figure 6.11, but fix the total number of samples available to $100n$, where n is the dimension of the system. For low dimensional systems, we observe that the bounds based on adaptive sampling remain tighter than the one of Theorem 3.1, in spite of the decrease in the total number of samples. For high-dimensional systems, the

accuracy of our bounds increases. **The bound of Theorem 5.2 is the tightest. For systems of dimension smaller than 20, it does not exceed the true JSR by more than 40% of its value, which often allows it to serve as a stability certificate.** The bound of Theorem 5.1 beats the one of Theorem 3.1 in a larger number of cases. However, it is still unable to outperform the bound obtained without resorting to adaptive sampling for systems of dimension larger than 25. This does not imply that our adaptive sampling approach is not effective for high-dimensional systems, but only that more samples are required for it to become efficient. To reach tight bounds on the JSR (less than 500% of its value, for instance), the sample complexity of our heuristic is lower than the one required when not resorting to adaptive sampling. We illustrate this property on a 45-dimensional

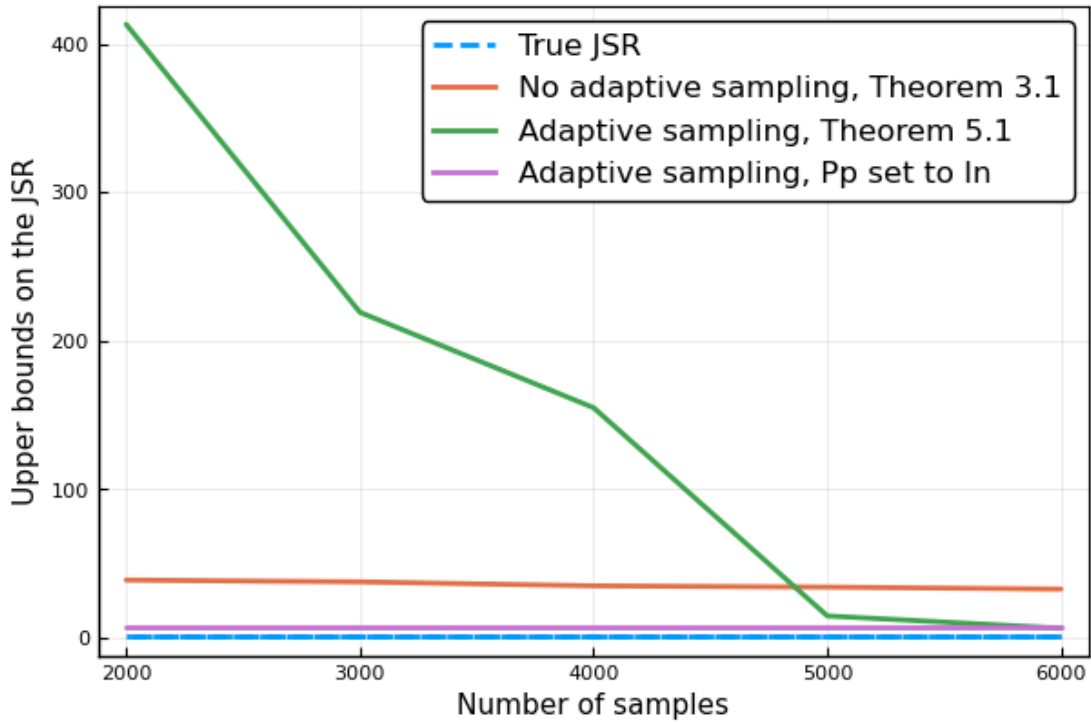


Figure 6.12: Evolution of several upper bounds on the JSR of a 45-dimensional system with the total number of samples. The bounds based on adaptive sampling are yielded by the application of Heuristic 1. The bounds of Theorems 5.1 and 5.2, of Theorem 3.1 and the true value of the JSR are plotted. The bound of Theorem 5.2 is the tightest. From 5000 samples, all bounds based on adaptive sampling outperform the one of Theorem 3.1.

system composed of 4 modes, uniformly and independently sampled on the sphere. Indeed, Figure 6.12 presents the evolution of the different upper bounds on the

JSR with the total number of samples. We observe that **from a sufficiently high number of samples, all bounds yielded by the application of the heuristic outperform the one obtained without resorting to adaptive sampling.**

Finally, we test Heuristic 1 on 1000 systems composed of a number of modes m varying between 2 and 7, and of dimension ranging between 2 and 50. The norm of the modes can be larger than 1. The total number of samples is fixed to $120 * n$ in each experiment. **In 90% of the cases, our adaptive sampling bound of Theorem 5.1 beats the bounds obtained without resorting to adaptive sampling.** Moreover, the bound obtained when fixing P_p to the identity always outperforms all other bounds.

6.5 Conclusion of the experiments

We summarize here the improvements yielded by our adaptive sampling approach in the context of data-driven stability analysis of switched linear systems.

- Even though the bounds of Theorems 5.1 and 5.2 only take into account the last N_p samples and not the total number N of samples, they are able to outperform the bounds obtained without resorting to adaptive sampling, but taking into account the total number of samples used.
- Our approach always yields tighter bounds than the fast algorithm, designed to decrease the sample complexity in the scenario approach.
- Both the online and the offline sampling methods allow to significantly decrease the sample complexity required to obtain tight bounds on the JSR. However, a judicious choice of the parameters used in the offline sampling is needed to obtain tight results. The online strategy provides the tightest bounds on the JSR as soon as the total number of samples is large enough for Algorithm 4 to reach Phase 3.
- As expected in Remark 5.5, the bound of Theorem 5.2, associated to the variant setting P_p to the identity, always beats all other bounds. This presents the auxiliary advantage of reducing even more the computational complexity of the method.
- Heuristic 1, derived based on empirical observations of the online and offline sampling strategies, allows to significantly decrease the sample and the computational complexity required to obtain tight bounds on the JSR holding with a given confidence level. It thus allows the data-driven approach to be

tractable for high-dimensional systems, which was exactly the goal of this thesis.

Chapter 7

Extension of the adaptive sampling approach to the SOS setting

As observed in Chapter 4, resorting to approximate SOS Lyapunov functions to obtain probabilistic bounds on the JSR presents two drawbacks in terms of the computational complexity.

- First, the dimension of the problem increases exponentially with the degree of the function used. This increases the computational complexity, which depends on the scale of the problem to solve.
- Second, because of the increased dimension of the problem solved, $\kappa(P_{sos}^*(\omega_N))$ grows rapidly with the dimension of the system analyzed. Consequently, the sample complexity required to obtain tight bounds on the JSR grows, which also increases the computational complexity of the method.

In this chapter, we overcome this issue by leveraging our adaptive sampling approach to the SOS setting. This allows to decrease the sample complexity required to obtain tight bounds on the JSR, holding with a given confidence level.

We follow the same structure as in Chapter 5. First, we define the optimal change of variables to be computed, and propose an algorithm extending the adaptive sampling approach to the SOS setting. We then provide a SOS-based bound relying on the solutions yielded by this algorithm. Finally, we illustrate the performance of our new SOS-based method on a numerical experiment.

Notation

In the SOS setting, we approximate the solution $(\gamma_{sos}^o, P_{sos}^o)$ to Problem (2.9):

$$\begin{aligned} \min_{P, \gamma \geq 0} \quad & \gamma \quad \text{s.t.} \quad (\mathbf{A}x)^{[d]T} P (\mathbf{A}x)^{[d]} \leq \gamma^{2dl} x^{[d]T} P x^{[d]}, \quad \forall \mathbf{A} \in \mathcal{M}^l, \quad \forall x \in \mathbb{S} \\ & P \succeq I_D, \quad \|P\| \leq C \text{ for a large } C \in \mathbb{R}_{\geq 0}, \end{aligned}$$

by solving an associated sampled problem: Problem (3.3). In this section, similarly to the quadratic setting, we denote it by $\text{Opt}_{sos,1}(\omega_{N_1})$ and its solution by $(\gamma_{sos,1}^*(\omega_{N_1}), P_{sos,1}^*(\omega_{N_1}))$:

$$\begin{aligned} \min_{P_1, \gamma_1 \geq 0} \quad & \gamma_1 \quad \text{s.t.} \quad ((\mathbf{A}_j x)^{[d]})^T P_1 (\mathbf{A}_j x)^{[d]} \leq \gamma_1^{2dl} (x^{[d]})^T P_1 x^{[d]}, \quad \forall (x, \mathbf{j}) \in \omega_{N_1} \\ & P_1 \succeq I_D, \quad \|P_1\| \leq C \text{ for a large } C \in \mathbb{R}_{\geq 0}. \end{aligned} \quad (7.1)$$

7.1 Adaptive sampling in the SOS setting

Similarly to the quadratic case, we seek for a coordinate system in which $P_{sos}^o = I_D$ is the solution to Problem (2.9). We first define it in the white-box setting. Then, we propose our approach in the black-box setting.

Optimal change of variables in the white-box setting

Let $L \in \mathbb{R}^{n \times n}$ be the matrix defining the change of variables. In the new coordinate system, Problem (2.9) becomes:

$$\begin{aligned} \min_{P', \gamma' \geq 0} \quad & \gamma' \\ \text{s.t.} \quad & ((L\mathbf{A}L^{-1}x')^{[d]})^T P' (L\mathbf{A}L^{-1}x')^{[d]} \leq \gamma'^{2dl} (x'^{[d]})^T P' x'^{[d]}, \quad \forall \mathbf{A} \in \mathcal{M}^l, \quad \forall x' \in \mathbb{S} \\ & P' \succeq I_D, \quad \|P'\| \leq C \text{ for a large } C \in \mathbb{R}_{\geq 0} \end{aligned} \quad (7.2)$$

where we have used the fact that $L^{[d]}(\mathbf{A}x)^{[d]} = (L\mathbf{A}x)^{[d]}$ [Bro73]. We denote this problem by $\text{Opt}_{sos}(L)$ and its solution by $(\gamma_{sos}^o(L), P_{sos}^o(L))$. Solving $\text{Opt}_{sos}(L)$ amounts to computing an approximate SOS Lyapunov function for the set $L\mathcal{M}L^{-1}$. Note that $\gamma_{sos}^o(L) = \gamma_{sos}^o$ and that $P_{sos}^o(L) = (L^{[d]})^{-T} P_{sos}^o (L^{[d]})^{-1}$ is a feasible solution.

Let $L^o \in \mathbb{R}^{D \times D}$ be the Cholesky decomposition of P_{sos}^o : $P_{sos}^o = (L^o)^T L^o$. Then, the same reasoning as in the quadratic case yields the following conclusion: the optimal change of variables is given by $L^{[d]} = L^o$. To the contrary to the quadratic case, this change of variables does not exist in general. Indeed, as observed in

Remark 2.2, the d-lift of the set of all matrices in $\mathbf{R}^{n \times n}$ only consists of a subset of the set of all matrices in $\mathbf{R}^{D \times D}$. For instance, when d is even, $l_{11}^{[d]}$ will always be positive. Hence, we approximate the optimal change of variables by the solution of the following least squares problem:

$$\min_{L \in \mathbf{R}^{n \times n}} \|L^{[d]} - L^o\|. \quad (7.3)$$

Even in the white-box setting, we thus only can compute an approximation of the optimal change of variables.

p -phases SOS adaptive sampling

In our black-box setting, we only have access to a finite number of trajectories. Therefore, we do not have access to P_{sos}^o , but only to $P_{sos,1}^*(\omega_{N_1})$, its approximation computed by solving $\text{Opt}_{sos,1}(\omega_{N_1})$ using N_1 uniformly sampled trajectories. Similarly, we cannot compute $P_{sos}^{lo}(L)$ but only its approximation, by solving the sampled problem associated to $\text{Opt}_{sos}(L)$. Hence, our adaptive sampling technique consists of three steps, each inducing a factor increasing the distance separating the final sampled solution from the identity:

- Uniform sampling of $\omega_{N_1} \in Z_l^{N_1}$, and computation of the solution $(\gamma_{sos,1}^*(\omega_{N_1}), P_{sos,1}^*(\omega_{N_1}))$ to $\text{Opt}_{sos,1}(\omega_{N_1})$. This induces a distance $\Delta_{s,1} = \|P_{sos,1}^*(\omega_{N_1}) - P_{sos}^o\|$, representing the quality of the change of variables. It decreases when N_1 increases.
- Determination of the change of variables $\tilde{L}_1$, solution to Problem (7.3). $(\tilde{L}_1)^{[d]} \approx L_1$, where L_1 is the Cholesky decomposition of $P_{sos,1}^*(\omega_{N_1})$. This induces a distance $\Delta_{lift} = \|\tilde{L}_1^{[d]} - L_1\|$ impacting the quality of the change of variables. This distance depends on the dimension of the problem.
- Sampling of ω_{N_2} , consisting of N_2 trajectories, uniformly in the system $\tilde{L}_1 \mathcal{M}^l \tilde{L}_1^{-1}$, using Procedure Generation. We then compute a bound on the JSR by solving a sampled version of $\text{Opt}_{sos}(\tilde{L}_1)$. Similarly to the quadratic case, we choose to benefit from the information contained in both ω_{N_1} and ω_{N_2} . We denote this problem by $\text{Opt}_{sos,2}(\omega_{N_1+N_2})$ and its solution by $(\gamma_{sos,2}^*(\omega_{N_1+N_2}), P_{sos,2}^*(\omega_{N_1+N_2}))$:

$$\begin{aligned} & \min_{P_2, \gamma_2 \geq 0} \quad \gamma_2 \\ & \text{s.t. } ((\tilde{L}_1 \mathbf{A}_j \tilde{L}_1^{-1} x)^{[d]})^T P_2 (\tilde{L}_1 \mathbf{A}_j \tilde{L}_1^{-1} x)^{[d]} \leq \gamma_2^{2dl} (x^{[d]})^T P_2 x^{[d]}, \\ & \quad \forall (x, \mathbf{j}) \in \omega_{N_2} \cup \{(x', \mathbf{j}) : x' = \tilde{L}_1 x, (x, \mathbf{j}) \in \omega_{N_1}\} \\ & \quad P_2 \succeq I_D, \quad \|P_2\| \leq C \text{ for a large } C \in \mathbb{R}_{\geq 0}. \end{aligned} \quad (7.4)$$

This step induces a distance $\Delta_{s,2} = \|P_{sos,2}^*(\omega_{N_1+N_2}) - P_{sos}^{lo}(\tilde{L}_1)\|$ decreasing when N_2 increases.

The performance of our approach depends on the closeness of $P_{sos,2}^*(\omega_{N_1+N_2})$ to the identity. As illustrated on Figure 7.1, this quantity is impacted by one more factor than in the quadratic case: the distance Δ_{lift} .

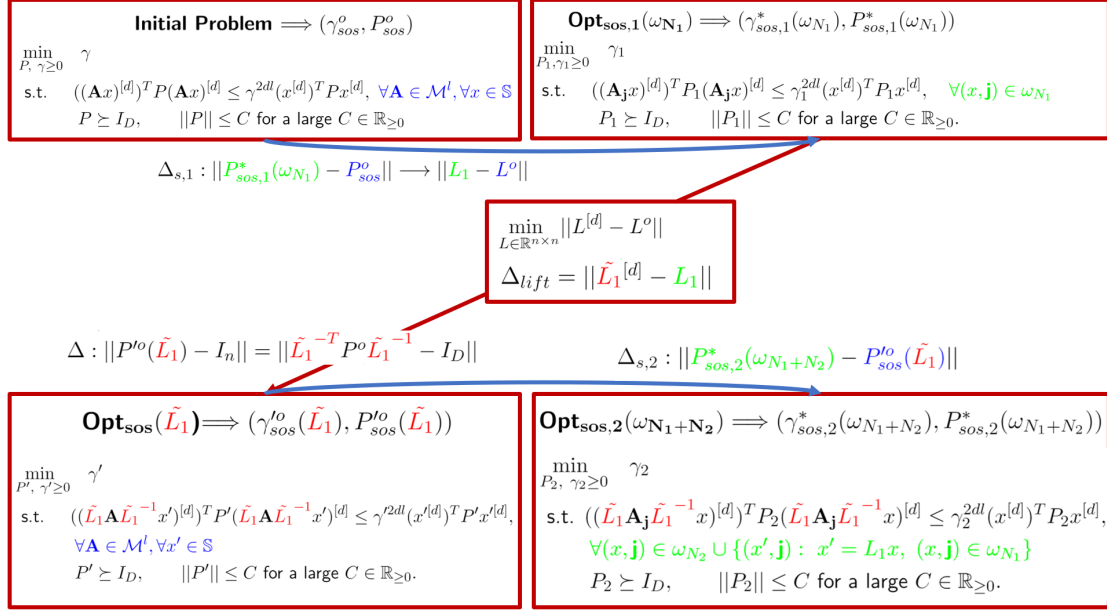


Figure 7.1: Relations between the different optimization problems designed for the two-phase adaptive sampling approach in the SOS setting. The factors impacting the distance $\|P_{sos,2}^*(\omega_{N_1+N_2}) - I_D\|$ are emphasized. The left-hand side of the figure presents the white-box problems, before and after the change of variables is performed. The right-hand side presents the associated sampled problems. We only have access to the solutions to these problems, approximating the solutions of the initial problems. The closeness of $P_{sos,2}^*(\omega_{N_1+N_2})$ to the identity depends on $\Delta_{s,1}$, $\Delta_{s,2}$ and Δ_{lift} . The latter appears because the exact optimal change of variables does not exist in general in the SOS setting, to the contrary to the quadratic setting. $\Delta_{s,1}$ and Δ_{lift} affect the quality of the change of variables by inducing a distance between the solution of $\text{Opt}_{sos}(\tilde{L}_1)$ and the identity. $\Delta_{s,1}$ decreases when N_1 increases, while $\Delta_{s,2}$ decreases when N_2 increases.

Similarly to the quadratic case, the change of variables can be refined by increasing the number of phases. Therefore, for $k \geq 1$, we define $\tilde{\mathbf{L}}_k = \tilde{L}_k \dots \tilde{L}_2 \tilde{L}_1$ and $\tilde{\mathbf{L}}_k^{-1} = \tilde{L}_k^{-1} \dots \tilde{L}_2^{-1} \tilde{L}_1^{-1}$. For $k = 0$, $\tilde{\mathbf{L}}_k = \tilde{\mathbf{L}}_k^{-1} = I_n$. We also define $\tilde{\Omega}_k = \omega_{N_k} \cup \{(x', \mathbf{j}) : x' = \tilde{L}_{k-1} x, (x, \mathbf{j}) \in \omega_{N_{k-1}}\} \cup \dots \cup \{(x', \mathbf{j}) : x' = \tilde{\mathbf{L}}_{k-1} x, (x, \mathbf{j}) \in \omega_{N_1}\}$. Then, we introduce

$\text{Opt}_{\text{sos},k}(\omega_{N_1+\dots+N_k})$, of solution denoted by $(\gamma_{\text{sos},k}^*(\omega_{N_1+\dots+N_k}), P_{\text{sos},k}^*(\omega_{N_1+\dots+N_k}))$, as follows:

$$\begin{aligned}
& \min_{P_k, \gamma_k \geq 0} \quad \gamma_k \\
& \text{s.t.} \quad ((\tilde{\mathbf{L}}_{\mathbf{k}-1} \mathbf{A}_{\mathbf{j}} \tilde{\mathbf{L}}_{\mathbf{k}-1}^{-1} x)^{[d]})^T P_k (\tilde{\mathbf{L}}_{\mathbf{k}-1} \mathbf{A}_{\mathbf{j}} \tilde{\mathbf{L}}_{\mathbf{k}-1}^{-1} x)^{[d]} \leq \gamma_k^{2dl} (x^{[d]})^T P_k x^{[d]}, \quad \forall (x, \mathbf{j}) \in \tilde{\Omega}_k \\
& \quad P_k \succeq I_D, \quad \|P_k\| \leq C \text{ for a large } C \in \mathbb{R}_{\geq 0},
\end{aligned} \tag{7.5}$$

where $\tilde{L}_i$, for $i = 1, \dots, k-1$, is the matrix minimizing $\|\tilde{L}_i^{[d]} - L_i\|$, and L_i is the Cholesky decomposition of $P_{\text{sos},i}^*(\omega_{N_1+\dots+N_i})$, the solution to $\text{Opt}_{\text{sos},i}(\omega_{N_1+\dots+N_i})$. We present our p -phases approach in Algorithm 5.

Algorithm 5: p -phases adaptive sampling algorithm in the SOS setting.	
Input: Number of phases: p , Number of samples of each phase: $N_1, \dots, N_p$.	
Result: $(\gamma_{\text{sos},p}^*(\omega_{N_1+\dots+N_p}), P_{\text{sos},p}^*(\omega_{N_1+\dots+N_p}))$	
First Phase	
1	Sampling of a set $\omega_{N_1} \in Z_l^{N_1}$, where $ \omega_{N_1} = N_1$;
2	Computation of $(\gamma_{\text{sos},1}^*(\omega_{N_1}), P_{\text{sos},1}^*(\omega_{N_1}))$, solution to $\text{Opt}_{\text{sos},1}(\omega_{N_1})$;
3	Computation of L_1 , the Cholesky decomposition of $P_{\text{sos},1}^*(\omega_{N_1})$;
4	Computation of the change of variables $\tilde{L}_1$ minimizing $\ \tilde{L}_1^{[d]} - L_1\ $;
Second set of phases	
5	for $k = 2$ to p do
6	Sampling of ω_{N_k} uniformly for $\tilde{\mathbf{L}}_{\mathbf{k}-1} \mathcal{M}^l \tilde{\mathbf{L}}_{\mathbf{k}-1}^{-1}$, where $ \omega_{N_k} = N_k$, using Procedure Generation ;
7	Computation of $(\gamma_{\text{sos},k}^*(\omega_{N_1+\dots+N_k}), P_{\text{sos},k}^*(\omega_{N_1+\dots+N_k}))$, solution to $\text{Opt}_{\text{sos},k}(\omega_{N_1+\dots+N_k})$;
8	if $k \neq p$ then
9	Computation of L_k , the Cholesky decomposition of $P_{\text{sos},k}^*(\omega_{N_1+\dots+N_k})$;
10	Computation of the change of variables $\tilde{L}_k$ minimizing $\ \tilde{L}_k^{[d]} - L_k\ $;
11	Computation of $\tilde{\mathbf{L}}_{\mathbf{k}} = \tilde{L}_k \tilde{\mathbf{L}}_{\mathbf{k}-1}$ and $\tilde{\mathbf{L}}_{\mathbf{k}}^{-1} = \tilde{L}_k^{-1} \tilde{\mathbf{L}}_{\mathbf{k}-1}^{-1}$;
12	end
13	end

In the quadratic setting, fixing P_p to the identity in $\text{Opt}_p(\omega_{N_1+\dots+N_p})$ allows to decrease the computational complexity of the problem and yields tighter bounds on the JSR for an equivalent sample complexity. However, in the SOS setting, fixing

P_p to the identity in $\text{Opt}_{\text{sos},p}(\omega_{N_1+\dots+N_p})$ reduces the problem to:

$$\begin{aligned} \min_{\gamma_p \geq 0} \quad & \gamma_p \quad \text{s.t.} \quad \|((\tilde{\mathbf{L}}_{\mathbf{p}-1} \mathbf{A}_{\mathbf{j}} \tilde{\mathbf{L}}_{\mathbf{p}-1}^{-1} x)^{[d]})\| \leq \gamma_k^{dl} \|x^{[d]}\|, \quad \forall (x, \mathbf{j}) \in \tilde{\Omega}_k \\ & = \min_{\gamma_p \geq 0} \quad \gamma_p \quad \text{s.t.} \quad \|\tilde{\mathbf{L}}_{\mathbf{p}-1} \mathbf{A}_{\mathbf{j}} \tilde{\mathbf{L}}_{\mathbf{p}-1}^{-1} x\| \leq \gamma_k^l \|x\|, \quad \forall (x, \mathbf{j}) \in \tilde{\Omega}_k \end{aligned} \quad (7.6)$$

where $\tilde{L}_i$, for $i = 1, \dots, p-1$, is the matrix minimizing $\|\tilde{L}_i^{[d]} - L_i\|$, and L_i is the Cholesky decomposition of $P_{\text{sos},i}^*(\omega_{N_1+\dots+N_i})$, the solution to $\text{Opt}_{\text{sos},i}(\omega_{N_1+\dots+N_i})$. The problem thus reduces to a version of $\text{Opt}_{p,I}(\omega_{N_1+\dots+N_p})$ analyzing the system $\tilde{\mathbf{L}}_{\mathbf{p}-1} \mathcal{M}' \tilde{\mathbf{L}}_{\mathbf{p}-1}^{-1}$ instead of $\mathbf{L}_{\mathbf{p}-1} \mathcal{M}' \mathbf{L}_{\mathbf{p}-1}^{-1}$. Because of the distance Δ_{dift} , for a same total number of samples used in the first phases, $\tilde{\mathbf{L}}_{\mathbf{p}-1}$ will be a less accurate approximation of the optimal quadratic change of variables than $\mathbf{L}_{\mathbf{p}-1}$. Hence, this variant in the SOS setting will not be able to beat the quadratic bound. Hence, we only consider the case where P_p is not fixed to the identity.

We introduce the notation used in our next statements in the following assumption:

Assumption 7.1. We consider an n -dimensional switched linear system as in (2.1), whose set of m modes is denoted by $\mathcal{M} \subset \mathbb{R}^{n \times n}$, and approximate SOS Lyapunov functions of degree $2d$. We consider $p, N \in \mathbb{N}$, and $N_i \in \mathbb{N} \geq d_{\text{sos}} + 1$ for $i = 1, \dots, p$ such that $N_1 + \dots + N_p = N$. For $i = 1, \dots, p$, let ω_{N_i} be a uniform random sampling for System $\tilde{\mathbf{L}}_{i-1} \mathcal{M}' \tilde{\mathbf{L}}_{i-1}^{-1}$. We denote by $(\gamma_{\text{sos},i}^*(\omega_{N_1+\dots+N_i}), P_{\text{sos},i}^*(\omega_{N_1+\dots+N_i}))$ the solution to $\text{Opt}_{\text{sos},i}(\omega_{N_1+\dots+N_i})$ for $i = 1, \dots, p-1$, and by $(\gamma_{\text{sos},p}^*, P_{\text{sos},p}^*)$ the solution to $\text{Opt}_{\text{sos},p}(\omega_{N_1+\dots+N_p})$. $\tilde{L}_i$ is the matrix minimizing $\|\tilde{L}_i^{[d]} - L_i\|$, and L_i is Cholesky decomposition of $P_{\text{sos},i}^*(\omega_{N_1+\dots+N_i})$.

Theoretical bound based on the last set of samples only

For the same reasons as in the quadratic setting, we are only able to derive a bound on the JSR of switched linear systems based on the last set of samples solely. Then, since ω_{N_p} is sampled uniformly in the system considered, we can use the results of Chapter 3 and adapt the bound of Theorem 3.3 to our adaptive sampling approach. We obtain the following bound:

Theorem 7.1

Let Assumption 7.1 hold, and let $\lambda_p^*(\omega_{N_1+\dots+N_p})$ be the optimal solution to Problem (5.8). For any $\beta \in [0, 1)$ and $\beta_1 \in [0, 1)$, let $\epsilon_{\text{sos}} = m^l \left(1 - \sqrt[p]{\frac{1-\beta}{d_{\text{sos}}+1}}\right)$ and $\epsilon_1 = \frac{m^l}{2} (1 - \sqrt[p]{1-\beta_1})$. Then, it holds that, with probability at least

$\beta + \beta_1 - 1$:

$$\rho(\mathcal{M}) \leq \sqrt[dl]{\gamma_{sos,p}^{*dl} + ((\gamma_{sos,p}^{*dl} + A(\epsilon_1)^d)f(d, \epsilon_{sos}))\kappa(P_{sos,p}^*)}$$

where $A(\epsilon_1) = \frac{\lambda_p^*(\omega_{N_1+\dots+N_p})}{\delta(\epsilon_1)}$, $\delta(\cdot)$ is given by (3.7), $\kappa(P) = \sqrt{\frac{\lambda_{\max}(P)}{\lambda_{\min}(P)}}$,
 $f(d, \epsilon_{sos}) = \sqrt{D}((1 + \Delta(\epsilon_{sos}))^d - 1 - (1 - \frac{1}{\sqrt{D}})\Delta(\epsilon_{sos})^d)$ and $\Delta(\epsilon_{sos}) = \sqrt{2 - 2\delta(\epsilon_{sos})}$.

Proof. We prove that, with probability at least β , it holds that

$$\rho(\mathcal{M}) \leq \sqrt[dl]{\gamma_{sos,p}^{*dl} + ((\gamma_{sos,p}^{*dl} + B(\tilde{\mathbf{L}}_{\mathbf{p}-1}\mathcal{M}^l\tilde{\mathbf{L}}_{\mathbf{p}-1}^{-1})^d)f(d, \epsilon_{sos}))\kappa(P_{sos,p}^*)} \quad (7.7)$$

where $B(\tilde{\mathbf{L}}_{\mathbf{p}-1}\mathcal{M}^l\tilde{\mathbf{L}}_{\mathbf{p}-1}^{-1}) = \max_{\mathbf{A} \in \tilde{\mathbf{L}}_{\mathbf{p}-1}\mathcal{M}^l\tilde{\mathbf{L}}_{\mathbf{p}-1}^{-1}} \|\mathbf{A}\|$.

By the same reasoning as in the proof of Proposition 5.1, the extended version of Lemma 3.3 as well as Proposition 3.5, derived for Problem (2.9), also hold for Problem (7.2), with N replaced by N_p . Then, since $(\gamma_{sos,p}^*, P_{sos,p}^*)$ satisfy any subset of constraints of $\text{Opt}_{sos,p}(\omega_{N_1+\dots+N_p})$, it holds that $\forall(x, \mathbf{j}) \in \omega_{N_p}$, $\|(\tilde{\mathbf{L}}_{\mathbf{p}-1}\mathbf{A}_{\mathbf{j}}\tilde{\mathbf{L}}_{\mathbf{p}-1}^{-1}x)^{[d]}\|_{P_{sos,p}} \leq \gamma_{sos,p}^{*l}\|x^{[d]}\|_{P_{sos,p}^*}$. Hence, we can follow the same reasoning as in the proof of Theorem 3.3, applied to system $\tilde{\mathbf{L}}_{\mathbf{p}-1}\mathcal{M}^l\tilde{\mathbf{L}}_{\mathbf{p}-1}^{-1}$ instead of the initial system $\mathcal{M}^l$ and obtain Equation (7.7).

By the same reasoning as in the proof of Proposition 5.1, $B(\mathbf{L}_{\mathbf{p}-1}\mathcal{M}^l\mathbf{L}_{\mathbf{p}-1}^{-1}) \leq \frac{\lambda_p^*(\omega_{N_1+\dots+N_p})}{\delta(\epsilon_1)}$ holds with probability at least β_1 . Combining both equations concludes the proof. $\square$

7.2 Numerical illustration of the performance of SOS adaptive sampling

We illustrate the benefits of resorting to adaptive sampling in the SOS setting on a 2-dimensional stable system for which no CQLF exists. Then, in the white-box setting, SOS-based bounds significantly outperform quadratic bounds.

Example 7.1. ([LW95, Ex. 2.1])

Consider the 2-modal system of dimension 2 defined by

$$\mathcal{M} = \left\{ \left(1 - \frac{1}{2000}\right) \alpha^2 \begin{pmatrix} 0 & 0 \\ 1 & 0 \end{pmatrix}, \left(1 - \frac{1}{2000}\right) \alpha^{-1} \begin{pmatrix} \cos \frac{\pi}{4} & \sin \frac{\pi}{4} \\ -\sin \frac{\pi}{4} & \cos \frac{\pi}{4} \end{pmatrix} \right\},$$

where $\alpha = \frac{1}{2} \left(1 + (\cos \frac{\pi}{4})^{-1}\right)$.

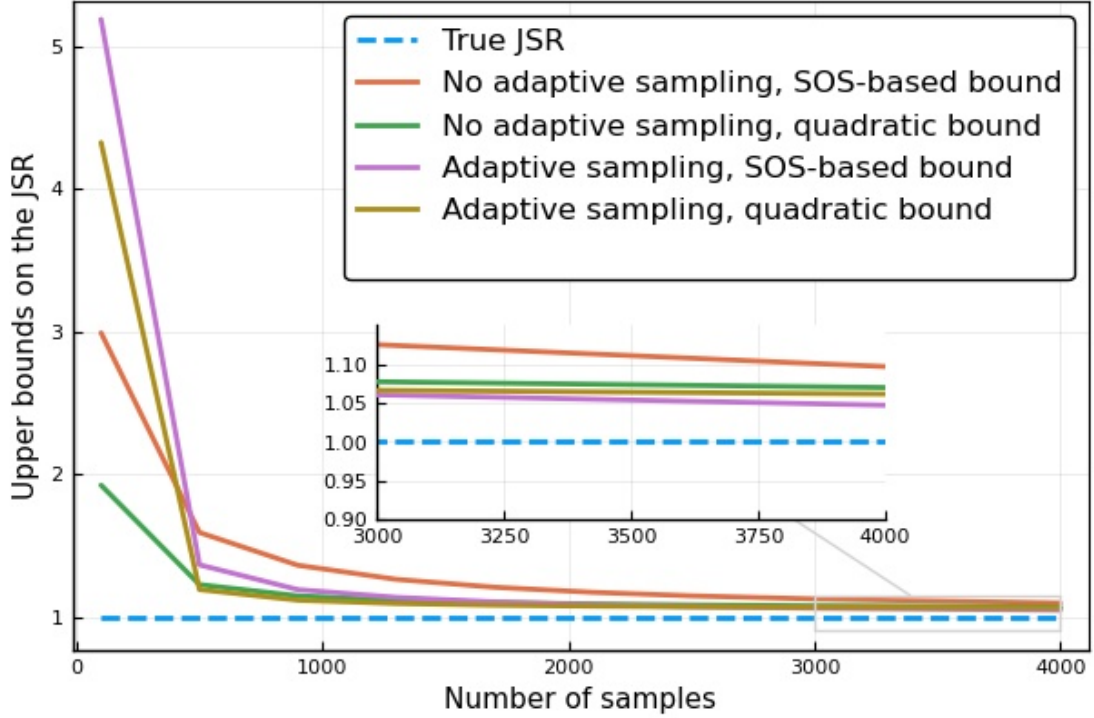


Figure 7.2: Evolution of several upper bounds on the JSR of the system of Example 7.1 with the total number of samples used. The SOS-based bounds of Theorems 3.3 and 7.1, the quadratic bounds of Theorems 3.1 and 5.1 and the true value of the JSR are plotted. The bound of Theorem 7.1 outperforms the SOS-based bound obtained without resorting to adaptive sampling from 400 samples. It outperforms all bounds from 3000 samples.

We rely on Heuristic 1, with n replaced by $D = \binom{d+n-1}{d}$, the new dimension of the problem. Figure 7.2 presents the evolution of the upper bounds on the JSR with the total number of samples. We plot the bounds of Theorems 3.1 and 3.3 (no resorting to adaptive sampling, in the quadratic and in the SOS case), of Theorems 5.1 and 7.1 (adaptive sampling, in the quadratic and in the SOS case), and the true value of the JSR as approximated by the JSR toolbox [LJP16]. We observe that the bound of Theorem 7.1 outperforms all bounds for a sufficiently high number of samples. It beats the SOS-based bound obtained without resorting to adaptive sampling from 400 samples. While the latter requires more than 6000 samples to outperform the one of Theorem 3.1, resorting to adaptive sampling allows to observe the benefits of approximate SOS Lyapunov functions from a much lower number of samples.

This example thus illustrates how combining our two contributions, that is SOS-

based stability guarantees and the adaptive sampling approach, allows to obtain, for a given number of samples and a given confidence level, the tightest bounds on the JSR possible.

Note however than for higher dimensional systems, it often happens that Δ_{lift} is too large for the adaptive sampling bound to beat the bound of Theorem 3.3.

Part IV

Conclusion

Chapter 8

Summary of the contributions and future prospects

In our everyday lives, the range of situations in which we interact with complex and large-scale dynamical systems is ever growing. Who never encountered a drone, was astonished by smart assistive technologies, or heard about the promises of self-driving cars? Deeply understanding the properties of these systems is essential to exploit their potential and ensure the security of their users. However, due to the tremendous complexity of these systems, a closed-form model is not always available, and one cannot directly rely on model-based techniques. Rather, new methods based on data obtained through simulations of the system solely have to be developed.

In this thesis, we handled the problem of stability analysis of black-box switched linear systems, a first class of complex systems whose model is unknown. We met the challenge of providing probabilistic guarantees holding in the whole state space, based on a few observations of the underlying system. Hence, we were able to bridge the gap between data and model-based control. We introduced new data-driven approaches and derived novel probabilistic bounds on the JSR of switched linear systems accordingly. Thereby, we were able to significantly improve an existing bound, relying on the computation of CQLFs [KBJT19] [BJW21]. Moreover, to the contrary of the existing bound, our approach does not require any assumption on the system analyzed.

8.1 Summary of the contributions

We summarize here our main achievements, and illustrate them in Figure 8.1.

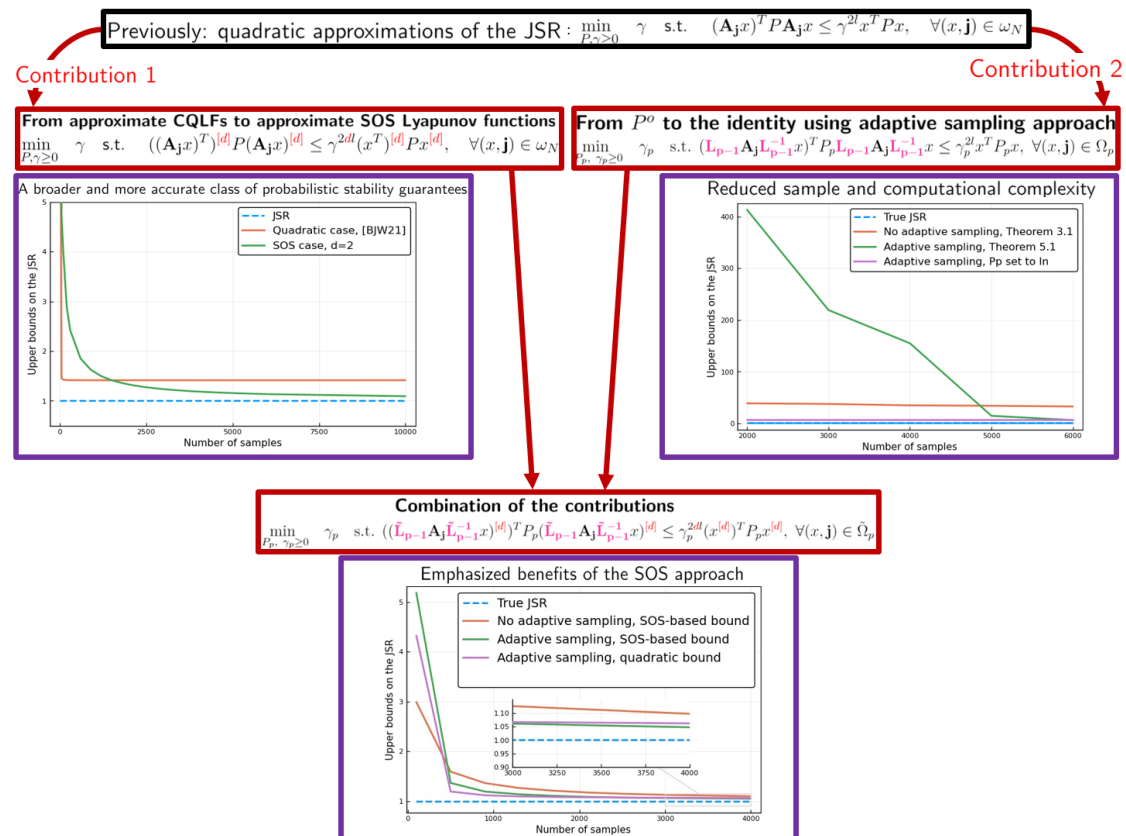


Figure 8.1: Summary of the different contributions of this thesis. We improved the existing probabilistic bound on the JSR of switched linear systems in several ways. First, we provided a novel class of bounds, based on the computation of approximate SOS Lyapunov functions. Additionally, this allowed us to derive two new quadratic bounds. In a second time, we proposed a technique to reduce the sample and computational complexity required to obtain tight bounds, and hence made the problem tractable for high dimensions. Finally, we combined our two contributions to improve our SOS-based bounds.

- First, we proposed in Chapter 3 a novel approach to derive probabilistic bounds on the JSR, relying on the sensitivity analysis of a quasi-linear optimization problem.

This allowed to (i) get rid of the assumptions required when resorting to the classical chance-constrained theory and (ii) propose a novel class of bounds, based on approximate SOS Lyapunov functions, providing more precise guarantees.

This approach yielded the derivation of three novel upper bounds. Firstly, we proposed two novel quadratic bounds, that is the bounds of Theorems 3.1 and 3.2. Then, we derived a SOS-based bound, that is the bound of Theorem 3.3. We illustrated in Chapter 4 the following advantages of these bounds over the one of [BJW21]:

1. They hold for all switched linear systems, to the contrary of the bound of [BJW21] which only holds for non degenerate systems.
 2. They are always finite.
 3. They take advantage of longer traces, while the bound of [BJW21] increases with the length of the trajectories.
 4. They always outperform the bound of [BJW21] for large-scale problems (of dimension larger than 6).
 5. The SOS-based bounds can serve as stability certificate for stable systems that do not admit a CQLF.
 6. The confidence interval for the value of the JSR is smaller in the SOS setting.
- Second, in Chapter 5, we improved our quadratic bounds on the JSR by proposing a novel technique, based on adaptive sampling.

This adaptive sampling approach allowed us to reduce both the sample and the computational complexity required to obtain a bound on the JSR of given accuracy, holding with a certain confidence level.

We derived two quadratic upper bounds based on this approach, that is the bound of Theorem 5.1 and the bound of Theorem 5.2. We then provided a heuristic whose results significantly improve the stability guarantees we can obtain on black-box switched linear systems. Our heuristic presents the following advantages:

1. Again, it does not require any assumption on the system analyzed.

- 2. It outperforms existing techniques to reduce the sample complexity required in the scenario approach.
- 3. It is a tractable method to provide tight bounds on the JSR of systems of any dimension. Indeed, to obtain bounds on the JSR smaller than 250% of its value, the sample and the computational complexity required by the heuristic are significantly smaller (10 to 15 %) than what is required by techniques that do not resort to adaptive sampling.
- Finally, in Chapter 6, we combined our two first contributions, that is our SOS-based bounds and the adaptive sampling approach. We derived an improved SOS-based bound, that is the bound of Theorem 7.1.

This allowed to significantly improve the accuracy of our SOS-based bounds for a given number of samples, and hence to observe the benefits of resorting to approximate SOS Lyapunov functions from a much lower number of samples.

8.2 Future work and possible further improvements

While we already made noticeable advances in the field of data-driven stability analysis of complex dynamical systems, we believe there exist several interesting directions for future research.

- First, we could improve the stability guarantees obtained using the adaptive sampling approach by deriving bounds taking into account not only the last set of samples, but all the trajectories ever sampled. However, as explained in Section 5.3, considering the changes of variables $L_1, \dots, L_{p-1}$ as random variables hugely complexifies the analysis of the problem.
- Another improvement consists in deriving a bound on the distance separating P° , the solution to Problem 2.3, from the sampled solution $P^*(\omega_N)$. This would allow to directly optimize the theorems of Chapters 5 and 6 to obtain the optimal values of the parameters to use in the adaptive sampling approach.
- One could also generalize the data-driven stability analysis of switched linear systems to other Lyapunov functions, for instance to the path-complete Lyapunov framework [PAAJ19].
- To reduce further the computational complexity of the problem, one could rely on distributed algorithms [CSBC14].

- In this thesis, we have considered the stability of switched linear systems under arbitrary switching. In many applications, there exist constraints on the values that the switching sequence can take [Lib03]. One could thus adapt the results derived in our general setting to the case of constrained switching scenario to obtain tighter results in this setting.
- Finally, the data-driven stability analysis can be extended to other classes of complex dynamical systems.

Bibliography

- [Aig97] Martin Aigner. *Combinatorial Theory*. Classics in Mathematics. Springer-Verlag Berlin Heidelberg, 1st edition edition, 1997.
- [AJ15] Amir Ahmadi and Raphaël M. Jungers. Lower bounds on complexity of lyapunov functions for switched linear systems. *IFAC Proceedings Volumes (IFAC-PapersOnline)*, 19, 2015.
- [BEFB94] Stephen Boyd, Laurent El Ghaoui, Eric Feron, and Venkataramanan Balakrishnan. *Linear Matrix Inequalities in System and Control Theory*, volume 15 of *Studies in Applied Mathematics*. SIAM, Philadelphia, PA, June 1994.
- [Ber16] Dimitri P. Bertsekas. *Nonlinear programming*. Athena Scientific, Belmont, 2nd edition, 2016.
- [BFGP17] Franco Blanchini, Gianfranco Fenu, Giulia Giordano, and Felice Andrea Pellegrino. Model-free plant tuning. *IEEE transactions on automatic control*, 62(6):2623–2634, 2017.
- [BJW21] Guillaume O. Berger, Raphaël M. Jungers, and Zheming Wang. Chance-constrained quasi-convex optimization with application to data-driven switched systems control. *arXiv preprint arXiv:2101.01415*, 2021.
- [BKSE12] Jeff Bezanson, Stefan Karpinski, Viral B. Shah, and Alan Edelman. Julia: A fast dynamic language for technical computing. *arXiv preprint arXiv:1209.5145*, 2012.
- [BM07] Franco Blanchini and Stefano Miani. *Set-Theoretic Methods in Control*. Systems and control : foundations and applications. Springer Basel AG, Boston, MA, 2007.
- [Bro73] Roger W. Brockett. Lie algebras and lie groups in control theory. In D. Q. Mayne and R. W. Brockett, editors, *Geometric Methods in System Theory*, pages 43–82, Springer Netherlands, 1973.

- [BT99] Vincent D. Blondel and John N. Tsitsiklis. Complexity of stability and controllability of elementary hybrid systems. *Automatica (Oxford)*, 35(3):479–489, 1999.
- [Cal10] Giuseppe C. Calafiore. Random convex programs. *SIAM journal on optimization*, 20(6):3427–3464, 2010.
- [Cal16] Giuseppe C. Calafiore. On repetitive scenario design. *IEEE Transactions on Automatic Control*, 62, 02 2016.
- [CDT⁺13] Mohammadreza Chamanbaz, Fabrizio Dabbene, Roberto Tempo, V. Venkataramanan, and Qing-Guo Wang. Sequential randomized algorithms for convex optimization in the presence of uncertainty. *IEEE Transactions on Automatic Control*, 61, 04 2013.
- [CG11] Marco C. Campi and Simone Garatti. A sampling-and-discarding approach to chance-constrained optimization: Feasibility and optimality. *Journal of Optimization Theory and Applications*, 148:257–280, 2011.
- [CG18] Marco C. Campi and Simone Garatti. *Introduction to the Scenario Approach*. Society for Industrial and Applied Mathematics, Philadelphia, PA, 2018.
- [CGC14] Algo Carè, Simone Garatti, and Marco C. Campi. Fast—fast algorithm for the scenario technique. *Operations Research*, 62(3):662–671, 2014.
- [CGR18] Marco C. Campi, Simone Garatti, and Federico Alessandro Ramponi. A general scenario theory for nonconvex optimization and decision making. *IEEE transactions on automatic control*, 63(12):4067–4078, 2018.
- [CSBC14] Luca Carlone, Vaibhav Srivastava, Francesco Bullo, and Giuseppe C. Calafiore. Distributed random convex programming via constraints consensus. *SIAM Journal on Control and Optimization*, 52(1):629–662, 2014.
- [DA20] Alex Devonport and Murat Arcak. Data-driven reachable set computation using adaptive gaussian process classification and monte carlo methods. In *2020 American Control Conference (ACC)*, pages 2629–2634. AACC, 2020.
- [DGK63] Ludwig Danzer, Branko Grünbaum, and Victor Klee. Helly’s theorem and its relatives, in convexity. *Proceedings of the Symposium on Pure Mathematics*, 7:101–180, 1963.

- [Epp05] David Eppstein. Quasiconvex programming. *Combinatorial and Computational Geometry*, 52(287-331):3, 2005.
- [GP11] Nicola Guglielmi and Vladimir Protasov. Exact computation of joint spectral characteristics of linear operators. *Foundations of Computational Mathematics*, 13, 2011.
- [GVL13] Gene H. Golub and Charles F. Van Loan. *Matrix Computations*. Johns Hopkins Studies in the Mathematical Sciences. Johns Hopkins University Press, 2013.
- [HVCMB11] Esteban Hernandez-Vargas, Patrizio Colaneri, Richard Middleton, and Franco Blanchini. Discrete-time control for switched positive systems with application to mitigating viral escape. *International Journal of Robust and Nonlinear Control*, 21(10):1093–1111, 2011.
- [JAPR17] Raphaël M. Jungers, Amir Ali Ahmadi, Pablo A. Parrilo, and Mardavij Roozbehani. A characterization of lyapunov inequalities for stability of switched systems. *IEEE Transactions on Automatic Control*, 62(6):3062–3067, 2017.
- [JLM03] Ali Jadbabaie, Jie Lin, and A. Stephen Morse. Coordination of groups of mobile autonomous agents using nearest neighbor rules. *IEEE Transactions on Automatic Control*, 48:988–1001, 2003.
- [Jun09] Raphaël Jungers. *The Joint Spectral Radius: Theory and Applications*, volume 385 of *Lecture Notes in Control and Information Sciences*. Springer Berlin / Heidelberg, 2009.
- [KBJT19] Joris Kenanian, Ayca Balkan, Raphael M. Jungers, and Paulo Tabuada. Data driven stability analysis of black-box switched linear systems. *Automatica (Oxford)*, 109, 2019.
- [KML16] Nikolaos Kariotoglou, Kostas Margellos, and John Lygeros. On the computational complexity and generalization properties of multi-stage and stage-wise coupled scenario programs. *Systems And Control Letters*, 94:63–69, 2016.
- [KQHT16] Alexandar Kozarev, John Quindlen, Jonathan How, and Ufuk Topcu. Case studies in data-driven verification of dynamical systems. In *Proceedings of the 19th International Conference on Hybrid Systems: Computation and Control*, HSCC '16, page 81–86, New York, 2016.

- [LA09] Hai Lin and P.J Antsaklis. Stability and stabilizability of switched linear systems: A survey of recent results. *IEEE transactions on automatic control*, 54(2):308–322, 2009.
- [Lau16] Fabien Lauer. On the complexity of switching linear regression. *Automatica (Oxford)*, 74:80–83, 2016.
- [Lee15] Edward Lee. The past, present and future of cyber-physical systems: A focus on models. *Sensors (Basel, Switzerland)*, 15(3):4837–4869, 2015.
- [LF72] Peter Lancaster and Hanafi K. Farahat. Norms on direct sums and tensor products. *Mathematics of computation*, 26(118):401–414, 1972.
- [Li11] Shengqiao Li. Concise formulas for the area and volume of a hyperspherical cap. *Asian Journal of Mathematics and Statistics*, 4, 2011.
- [Lib03] Daniel Liberzon. *Switching in Systems and Control*. Systems and Control: Foundations and Applications. Birkhauser, 2003.
- [LJP16] Benoît Legat, Raphaël M. Jungers, and Pablo A. Parrilo. Generating unstable trajectories for Switched Systems via Dual Sum-Of-Squares techniques. In *Proceedings of the 19th International Conference on Hybrid Systems: Computation and Control*, HSCC ’16, pages 51–60. ACM, 2016.
- [LT91] Michel Ledoux and Michel Talagrand. *Probability in Banach Spaces: Isoperimetry and Processes*. Classics in Mathematics. Springer Berlin / Heidelberg, 1991.
- [LW95] Jeffrey C. Lagarias and Yang Wang. The finiteness conjecture for the generalized spectral radius of a set of matrices. *Linear Algebra and its Applications*, 214:17–42, 1995.
- [MB73] Kantilal L. Majumder and G. P. Bhattacharjee. Algorithm as 64: Inverse of the incomplete beta function ratio. *Journal of the Royal Statistical Society. Series C (Applied Statistics)*, 22(3):411–414, 1973.
- [MSL15] Peyman Mohajerin Esfahani, Tobias Sutter, and John Lygeros. Performance bounds for the scenario approach and an extension to a class of non-convex programs. *IEEE Transactions on Automatic Control*, 60(1):46–58, 2015.

- [NY19] Chao Ning and Fengqi You. Optimization under uncertainty in the era of big data and deep learning: When machine learning meets mathematical programming. *Computers & Chemical Engineering*, 125:434–448, 2019.
- [PAAJ19] Matthew Philippe, Nikolaos Athanasopoulos, David Angeli, and Raphaël M. Jungers. On path-complete lyapunov functions: Geometry and comparison. *IEEE Transactions on Automatic Control*, 64(5):1947–1957, 2019.
- [PJ08] Pablo A. Parrilo and Ali Jadbabaie. Approximation of the joint spectral radius using sum of squares. *Linear algebra and its applications*, 428:2385–2402, 2008.
- [SDGDI⁺20] Francesco Smarra, Giovanni Domenico Di Girolamo, Vittorio De Iuliis, Achin Jain, Rahul Mangharam, and Alessandro D’Innocenzo. Data-driven switching modeling for mpc using regression trees and random forests. *Nonlinear analysis. Hybrid systems*, 36, 2020.
- [SDR09] Alexander Shapiro, Darinka Dentcheva, and Andrzej Ruszczyński. *Lectures on Stochastic Programming*. Society for Industrial and Applied Mathematics, 2009.
- [Set10] Burr Settles. Active learning literature survey. *University of Wisconsin, Madison*, 2010.
- [SWM⁺07] Robert Shorten, Fabian Wirth, Oliver Mason, Kai Wulff, and Christopher King. Stability criteria for switched and hybrid systems. *SIAM review*, 49(4):545–592, 2007.
- [TCRM14] Paulo Tabuada, Sina Yamac Caliskan, Matthias Rungger, and Rupak Majumdar. Towards robustness for cyber-physical systems. *IEEE transactions on automatic control*, 59(12):3151–3163, 2014.
- [Tes18] Gerald Teschl. *Topics in Real and Functional Analysis*. Graduate Studies in Mathematics. American Mathematical Society, 2018.
- [VB96] Lieven Vandenberghe and Stephen Boyd. Semidefinite programming. *SIAM Rev.*, 38(1):49–95, 1996.
- [VKR12] Phebe Vayanos, Daniel Kuhn, and Berç Rustem. A constraint sampling approach for multi-stage robust optimization. *Automatica*, 48:459–471, 2012.

- [WJ21] Zheming Wang and Raphael M. Jungers. Scenario-based set invariance verification for black-box nonlinear systems. *IEEE control systems letters*, 5(1):193–198, 2021.
- [ZHT17] Datong P. Zhou, Qie Hu, and Claire J. Tomlin. Quantitative comparison of data-driven and physics-based models for commercial building hvac systems. *2017 American Control Conference (ACC)*, pages 2900–2906, 2017.

Appendix A

Polyhedral stability guarantees

A.1 Polyhedral Lyapunov functions

Polyhedral Lyapunov functions form a third important family of Lyapunov functions, that can be used to obtain stability guarantees on switched linear systems [BM07]. They present the advantage of being non-conservative, which means that their existence is both a sufficient and necessary condition for stability. On the other hand, compared to quadratic or SOS Lyapunov functions, they can be extremely complex and hard to compute. Indeed, polyhedral functions are based on *polyhedral sets*, of the form $S = \{x \in \mathbb{R}^n : F_i x \leq 1, i = 1, 2, \dots, s\}$, whose complexity is not bounded by the space dimension but by the number of planes describing the set. This number can be huge even for the simplest systems. We here only consider a subset of polyhedral sets: *C-polytopes*.

Definition A.1. A C-polytope is a polyhedral set that is bounded, convex and includes the origin as an interior point.

For a C-polytope S , let us define the norm

$$\|x\|_S = \min\{\lambda \geq 0 : x \in \lambda S\} = \max_{i=1,2,\dots,s} F_i x.$$

This norm is a Lyapunov function if all modes are contractive with respect to it: $\|Ax\|_S \leq \|x\|_S, \forall A \in \mathcal{M}, \forall x \in \mathbb{R}^n \setminus 0$. In the spirit of CQFs and SOS forms and dropping the common prefix again, we introduce polyhedral forms:

Definition A.2. A polyhedral form for System (2.1) with a finite set of matrices $\mathcal{M} \in \mathbb{R}^{n \times n}$ is a C-polytope $S \subset \mathbb{R}^n$ such that for some $\gamma \geq 0$

$$\|Ax\|_S \leq \gamma \|x\|_S, \forall A \in \mathcal{M}, \forall x \in \mathbb{R}^n \setminus 0. \quad (\text{A.1})$$

When $\gamma < 1$, polyhedral forms can serve as stability certificates. We can also infer a lower bound on the JSR, whose accuracy increases by considering traces of longer length, from the non-existence of such forms. We summarize these observations in the following theorem:

Theorem A.1

[BM07, Thm. 6.27]

Consider a finite set of matrices $\mathcal{M} \in \mathbb{R}^{n \times n}$.

- If there exists $\gamma \geq 0$ and a C-polytope $S \subset \mathbb{R}^n$ such that (A.1) holds, then $\rho(\mathcal{M}) \leq \gamma$.
- If $\rho(\mathcal{M}) < \frac{\gamma}{\sqrt[n]{n}}$, there exists a C-polytope $S \subset \mathbb{R}^n$ corresponding to a polyhedral form for $\mathcal{M}^l$ such that

$$\|\mathbf{A}x\|_S \leq \gamma^l \|x\|_S, \quad \forall \mathbf{A} \in \mathcal{M}^l, \forall x \in \mathbb{R}^n \setminus 0.$$

Similarly to Problems (2.3) and (2.9), we seek for the minimal γ such that (A.1), for the system defined by the set $\mathcal{M}^l$, is satisfied. We therefore consider the following problem, whose solution is denoted by (γ_p^o, S^o) :

$$\min_{S \subset \mathbb{R}^n, \gamma \geq 0} \quad \gamma \text{ s.t.} \quad \|\mathbf{A}x\|_S \leq \gamma^l \|x\|_S, \quad \forall \mathbf{A} \in \mathcal{M}^l, \forall x \in \mathbb{S}, \quad S \text{ C-polytope.} \quad (\text{A.2})$$

The restriction to the sphere is again due to the homogeneity of System (2.1). To the contrary of Problems (2.3) and (2.9), related to quadratic and SOS Lyapunov functions, Problem (A.2) does not belong to the class of SDP programs, and cannot be solved as efficiently. In particular, it requires the computation of *convex hulls* of specific sets, that is the computation of the smallest convex set containing the initial set. We denote by $\text{conv}(S)$ the convex hull of the set $S \subseteq \mathbb{R}^n$. Problem (A.2) can be solved by bisection on γ and resorting to a geometric algorithm [BM07, Section 5.4], [GP11] at each iteration. For a given value of γ , the geometric algorithm requires an initial C-polytope $\mathcal{X}_0$ and recursively computes

$$\mathcal{X}_{k+1} := \text{conv}\left(\mathcal{X}_k \cup \bigcup_{\mathbf{j} \in \mathcal{M}^l} \frac{\mathbf{A}_{\mathbf{j}}}{\gamma} \mathcal{X}_k\right), \quad k = 1, 2, \dots$$

until either, for some $\bar{k}$,

- (i) convergence is attained: $\mathcal{X}_{\bar{k}+1} = \mathcal{X}_{\bar{k}} = \mathcal{X}_{\infty}$. This implies that γ is feasible for Problem (A.2), and we decrease it at the next iteration.
- (ii) $\mathcal{X}_{\bar{k}+1} \subset \text{int}\{\mathcal{X}_0\}$, where $\text{int}\{S\}$ denotes the interior of the set S , that is the

largest open set included in S . This implies that γ is infeasible for Problem (A.2), and we increase it at the next iteration.

It has been proven [BM07, Section 5.4], [GP11] that in either cases, the geometric algorithm terminates in a finite number of steps.

Similarly to the quadratic and SOS setting, we define a sampled problem associated to Problem (A.2):

$$\min_{S \subset \mathbb{R}^n, \gamma \geq 0} \quad \gamma \text{ s.t.} \quad \|\mathbf{A}x\|_S \leq \gamma^l \|x\|_S, \quad \forall (x, \mathbf{j}) \in \omega_N, \quad S \text{ C-polytope.} \quad (\text{A.3})$$

We denote Problem (A.3) by $\text{Opt}_{poly}(\omega_N)$ and its solution by $(\gamma_{poly}^*(\omega_N), S_{poly}^*(\omega_N))$. To extend the white-box algorithm to our sampled setting, we define the following sets:

$$\Omega_k(\omega_N) := \left\{ \frac{\mathbf{A}_{\mathbf{j}}x}{\|x\|_{\tilde{\mathcal{X}}_k(\omega_N)}} : (x, \mathbf{j}) \in \omega_N \right\} \cup \left\{ \frac{-\mathbf{A}_{\mathbf{j}}x}{\| -x \|_{\tilde{\mathcal{X}}_k(\omega_N)}} : (x, \mathbf{j}) \in \omega_N \right\}, \quad (\text{A.4})$$

and propose the following algorithm, adapted from [WJ21]:

Algorithm 6: Data-driven computation of a polyhedral bound on the JSR	
Input: C-polytope X, set $\omega_N \subset Z_l$, accuracy on the bisection $\alpha > 0$, upper bound on the JSR U, tolerance $\epsilon > 0$ Result: $(\gamma_{poly}^*(\omega_N), S_{poly}^*(\omega_N))$	
1	initialization: $ub=U, lb=0$;
2	while $ub - lb > \alpha$ do
3	$\gamma = \frac{ub+lb}{2}, k = 0, \tilde{\mathcal{X}}_0(\omega_N) = X$;
4	while $\text{conv}(\tilde{\mathcal{X}}_k(\omega_N) \cup \Omega_k(\omega_N)) \not\subseteq (1 + \epsilon)\tilde{\mathcal{X}}_k(\omega_N)$ and $\text{conv}(\tilde{\mathcal{X}}_k(\omega_N) \cup \Omega_k(\omega_N)) \not\subseteq (1 + \epsilon)X$ do
5	$\tilde{\mathcal{X}}_{k+1}(\omega_N) = \text{conv}(\tilde{\mathcal{X}}_k(\omega_N) \cup \Omega_k(\omega_N))$;
6	$k = k + 1$;
7	end
8	if $\text{conv}(\tilde{\mathcal{X}}_k(\omega_N) \cup \Omega_k(\omega_N)) \subseteq (1 + \epsilon)\tilde{\mathcal{X}}_k(\omega_N)$ then
9	$ub=\gamma$;
10	else
11	$lb=\gamma$;
12	end
13	end
14	$\gamma_{poly}^*(\omega_N) = ub, S_{poly}^*(\omega_N) = \tilde{\mathcal{X}}_k(\omega_N)$

This algorithm converges to a feasible solution of Problem (A.3) [WJ21, Thm. 1].

The stopping criterion ϵ is included to ensure the finite-time convergence of the algorithm. It now remains to infer a probabilistic bound on the cost of Problem (A.2) from the solution to Problem (A.3). Similarly to the quadratic setting, we can do so by relying on two different approaches: using theorems from chance-constrained optimization, or by relying on sensitivity analysis. We present here both approaches to be able to compare the accuracy of the bounds obtained.

A.2 Polyhedral probabilistic upper bounds on the JSR

A.2.1 From approximate to actual polyhedral Lyapunov functions

Concepts from this section are adapted from [WJ21, Section 4.A].

Let us suppose that we have a C-polytope $S \subset \mathbb{R}^{n-1}$ and a $\gamma > 0$ such that

$$\mu^l(\{x \in \mathbb{S} : \|\mathbf{A}x\|_S \leq \gamma^l \|x\|_S, \forall \mathbf{A} \in \mathcal{M}^l\}) \geq 1 - \epsilon. \quad (\text{A.5})$$

In other terms, S and γ define a ϵ -approximate polyhedral form, which becomes a ϵ -approximate Lyapunov function if $\gamma < 1$. The solution to Problem (A.3), $(\gamma_{poly}^*(\omega_N), S_{poly}^*(\omega_N))$, will satisfy Equation (A.5) for a certain ϵ .

Using geometric tools, we seek to transform this property holding almost everywhere in the state space into a relaxed inequality holding for all points on the sphere. Before stating our proposition, we introduce two geometric concepts:

Definition A.3. Let $S \subseteq \mathbb{R}^n$ be a C-polytope, $\epsilon \in (0, \frac{1}{2})$ and $\mathcal{V}(S)$ be the set of vertices of S . We define:

$$\lambda_{\min}(S, \epsilon) := \min_{u \in \mathcal{V}(S)} \lambda(u, S, \epsilon) \quad (\text{A.6})$$

where $\lambda(u, S, \epsilon) := \max_{\alpha \geq 0} \{\alpha : \alpha u \in \text{conv}(S \cap \bar{C}(u, \epsilon))\}$, $\bar{C}(u, \epsilon) := \{x \in \mathbb{R}^n : u^T x \leq \|x\| \|u\| \delta(\epsilon)\}$, and $\delta(\cdot)$ is given by (3.7).

Definition A.4. Let $S \subseteq \mathbb{R}^n$ be a C-polytope, $\epsilon \in (0, \frac{1}{2})$, $\mathcal{V}(S)$ be the set of vertices of S and $\mathcal{F}(S)$ be the set of facets of S . We define:

$$\underline{\lambda}(S, \epsilon) := \min_{u \in \mathcal{V}(S)} \delta(\epsilon) \frac{d_{\min}(u, S, \epsilon)}{\|u\|} \quad (\text{A.7})$$

where $d_{\min}(u, S, \epsilon) := \min_{f \in \mathcal{F}(S)} d_{\min}^f(u, S, \epsilon)$,

$d_{\min}^f(u, S, \epsilon) := \min_{x \in f} \{\|x\| : u^T x \geq \delta(\epsilon) \|x\| \|u\|\}$ and $\delta(\cdot)$ is given by (3.7).

Remark A.1. We can prove [WJ21, Lem. 1] that $\lambda_{\min}(S, \epsilon) \geq \underline{\lambda}(S, \epsilon)$. Namely, $\underline{\lambda}(S, \epsilon)$ is a relaxation of $\lambda_{\min}(S, \epsilon)$. It is determined by a convex optimization problem and can thus efficiently be computed, to the contrary of $\lambda_{\min}(S, \epsilon)$, which requires the computation of the convex hull of a set composed of nonlinear inequalities.

We now relate the observation (A.5) to a property satisfied everywhere in the state space by S and γ :

Proposition A.1. *Adapted from [WJ21, Prop. 3] Let $\gamma > 0$ and $S \subset \mathbb{R}^n$, a C -polytope, be such that (A.5) is satisfied for a given $\epsilon \in (0, \frac{1}{2})$. Let $\lambda_{\min}(S, \epsilon)$ be given by (A.6) and $\underline{\lambda}(S, \epsilon)$ be given by (A.7). Then, it holds that:*

$$\|\mathbf{A}x\|_S \leq \frac{\gamma^l}{\lambda_{\min}(S, \epsilon)} \|x\|_S \leq \frac{\gamma^l}{\underline{\lambda}(S, \epsilon)} \|x\|_S, \quad \forall x \in \mathbb{S}, \quad \forall \mathbf{A} \in \mathcal{M}^l \quad (\text{A.8})$$

Proof. By [WJ21, Prop. 3], it holds that if $\mu^l(\{x \in \mathbb{S} : \|\mathbf{A}x\|_S \leq \|x\|_S, \forall \mathbf{A} \in \mathcal{M}^l\}) \geq 1 - \epsilon$, then $\forall \mathbf{A} \in \mathcal{M}^l, x \in S$ implies $\mathbf{A}x \in \frac{1}{\lambda_{\min}(S, \epsilon)} S$. Hence, if $\mu^l(\{x \in \mathbb{S} : \|\mathbf{A}x\|_S \leq \gamma^l \|x\|_S, \forall \mathbf{A} \in \mathcal{M}^l\}) \geq 1 - \epsilon \iff \mu^l(\{x \in \mathbb{S} : \|\mathbf{A}x\|_S \leq \|x\|_S, \forall \mathbf{A} \in \frac{\mathcal{M}^l}{\gamma^l}\}) \geq 1 - \epsilon$, then $\forall \mathbf{A} \in \frac{\mathcal{M}^l}{\gamma^l}, x \in S$ implies $\mathbf{A}x \in \frac{1}{\lambda_{\min}(S, \epsilon)} S \iff \forall \mathbf{A} \in \mathcal{M}^l, x \in S$ implies $\mathbf{A}x \in \frac{\gamma^l}{\lambda_{\min}(S, \epsilon)} S$. Since, $\forall x \in \mathbb{S}, \frac{x}{\|x\|_S} \in S$, it holds that, $\forall x \in \mathbb{S}, \forall \mathbf{A} \in \mathcal{M}^l, \mathbf{A}x \in \frac{\gamma^l \|x\|_S}{\lambda_{\min}(S, \epsilon)}$. Therefore, $\|\mathbf{A}x\| \leq \frac{\gamma^l \|x\|_S}{\lambda_{\min}(S, \epsilon)}$ and the proof is concluded. $\square$

Proposition A.1 allows to infer the existence of a polyhedral form holding everywhere (and thus solution to Problem (A.2)) from an approximate polyhedral form, solution of Problem (A.3), valid everywhere except on a subset of measure at most ϵ . It remains to derive the probability with which the first property holds, for the solution of Problem (A.3).

A.2.2 Polyhedral upper bounds

First, based on a sensitivity analysis approach, we obtain the following upper bound on the JSR:

Theorem A.2

(adapted from [WJ21, Thm. 2])

Consider a set of matrices $\mathcal{M}$, and a uniform random sample $\omega_N \in Z_l^N$. Let $(\gamma_{poly}^*(\omega_N), S_{poly}^*(\omega_N))$ be the optimal solution to Problem (A.3). Let $s(\omega_N)$ be the number of samples $(x, \mathbf{j})$ in ω_N such that $\gamma_{poly}^{*l}(\omega_N \setminus \{(x, \mathbf{j})\}) < \gamma_{poly}^{*l}(\omega_N)$

and $\epsilon : \{0, \dots, N\} \rightarrow [0, 1]$ be given by:

$$\epsilon(k) := \begin{cases} 1 & \text{if } k = N \\ 1 - \sqrt[N-k]{\frac{1-\beta}{N \binom{N}{k}}} & \text{if } 0 \leq k < N \end{cases}$$

for any $\beta \in [0, 1)$. Then, with probability at least β , the following holds:

$$\begin{aligned} \rho(\mathcal{M}) &\leq \sqrt[l]{\frac{\gamma_{poly}^*(\omega_N)^l}{\lambda_{\min}(S_{poly}^*(\omega_N), I(4(\delta^2(\epsilon(\beta, N)) - \delta^4(\epsilon(\beta, N)))) ; \frac{n-1}{2}, \frac{1}{2})/2}} \\ &\leq \sqrt[l]{\frac{\gamma_{poly}^*(\omega_N)^l}{\underline{\lambda}(S_{poly}^*(\omega_N), I(4(\delta^2(\epsilon(\beta, N)) - \delta^4(\epsilon(\beta, N)))) ; \frac{n-1}{2}, \frac{1}{2})/2}} \end{aligned} \quad (\text{A.9})$$

where $\delta(\cdot)$ is given by (3.7), $\lambda_{\min}(S, \epsilon)$ is given by (A.6) and $\underline{\lambda}(S, \epsilon)$ is given by (A.7).

The proof follows directly from [WJ21, Thm. 2] and the argument of the proof of Proposition A.1. Based on chance-constrained optimization [CGR18], we obtain another polyhedral probabilistic bound on the JSR of our system:

Theorem A.3

Consider the same notation as in Theorem A.2. For any $\beta \in [0, 1)$, let $\epsilon(\beta, N)$ be such that

$$\beta = 1 - \frac{2m^l \left(1 - \frac{\epsilon(\beta, N)}{m^l}\right)^N}{I\left(\frac{1-\delta(\epsilon(\beta, N))}{2}, \frac{n-1}{2}, \frac{1}{2}\right)}.$$

Then, with probability at least β , the following holds:

$$\rho(\mathcal{M}) \leq \sqrt[l]{\frac{\gamma_{poly}(\omega_N)^{*l}}{\lambda_{\min}(S_{poly}^*(\omega_N), m^l \epsilon(s(\omega_N)))}} \leq \sqrt[l]{\frac{\gamma_{poly}^{*l}(\omega_N)}{\underline{\lambda}(S_{poly}^*(\omega_N), m^l \epsilon(s(\omega_N)))}} \quad (\text{A.10})$$

where $\delta(\cdot)$ is given by (3.7), $\lambda_{\min}(S, \epsilon)$ is given by (A.6) and $\underline{\lambda}(S, \epsilon)$ is given by (A.7).

Proof. A classical chance-constrained theorem [CGR18, Thm. 11] ensures that $\forall \beta \in (0, 1)$,

$$\mu_l^N \{\omega_N \in Z_l^N : \mu^l(V_{poly}(\omega_N) \leq \epsilon(s(\omega_N)))\} \geq \beta$$

where $V_{poly}(\omega_N) = \{(x, \mathbf{j}) \in Z_l : \|\mathbf{A}x\|_{S_{poly}^*(\omega_N)} > \gamma_{poly}^{*l}(\omega_N)\|x\|_{S_{poly}^*(\omega_N)}\}$. Then, by projecting the set of violated constraints onto the sphere, it holds that with probability at least β , $\mu_l\{x \in \mathbb{S} : \|\mathbf{A}x\|_{S_{poly}^*(\omega_N)} \leq \gamma_{poly}^{*l}(\omega_N)\|x\|_{S_{poly}^*(\omega_N)}\} \geq 1 - m^l \epsilon(s(\omega_N))$.

Hence, by Proposition A.1, with probability at least β ,

$$\begin{aligned} \|\mathbf{A}x\|_{S_{poly}^*(\omega_N)} &\leq \frac{\gamma_{poly}^{*l}(\omega_N)}{\lambda_{\min}(S_{poly}^*(\omega_N), m^l \epsilon(s(\omega_N)))} \|x\|_{S_{poly}^*(\omega_N)} \\ &\leq \frac{\gamma_{poly}^{*l}(\omega_N)}{\underline{\lambda}(S_{poly}^*(\omega_N), m^l \epsilon(s(\omega_N)))} \|x\|_{S_{poly}^*(\omega_N)}, \quad \forall x \in \mathbb{S}, \quad \forall \mathbf{A} \in \mathcal{M}^l \end{aligned}$$

which concludes the proof. $\square$

Remark A.2. While presenting the advantage of being non-conservative, the bounds of Propositions A.2 and A.3 are computationally expensive, especially when the dimension increases. Indeed, one has to compute the convex hull of several sets when applying Algorithm 6 to obtain $(\gamma_{poly}^*(\omega_N), S_{poly}^*(\omega_N))$, and the computation of $\underline{\lambda}(S_{poly}^*(\omega_N), m^l \epsilon(s(\omega_N)))$ becomes expensive when the number of vertices of $S_{poly}^*(\omega_N)$ increases, which always happens when the dimension increases.

Appendix B

Continuity argument for Lemma 3.3

We prove the equivalence between the two following statements:

- (S1) There exists a set $\omega \in Z_l^{d_{quad}+1}$ with $|\omega| = d_{quad} + 1$ such that $\gamma^*(\omega) = \gamma^o$.
- (S2) For any $\epsilon > 0$, there exists a set $\omega \in Z_l^{d_{quad}+1}$ with $|\omega| = d_{quad} + 1$ such that $\gamma^*(\omega) > \gamma^o - \epsilon$.

Clearly, if (S1) holds then (S2) holds. Suppose that (S2) holds. Then, we can construct a sequence $\{(\omega_k, \epsilon_k)\}$ where $\omega_k \in Z_l^{d_{quad}+1}$, $|\omega_k| = d_{quad} + 1$ and $\epsilon_{k+1} \leq \epsilon_k$ such that $\epsilon_k \rightarrow 0$ and $\gamma^*(\omega_k) > \gamma^o - \epsilon_k$. Taking the limit on both sides yields $\lim_{k \rightarrow \infty} \gamma^*(\omega_k) \geq \gamma^o$. Hence, since by definition, $\forall \omega_k \in Z_l^{d_{quad}+1}$, $\gamma^*(\omega_k) \leq \gamma^o$, it holds that $\lim_{k \rightarrow \infty} \gamma^*(\omega_k) = \gamma^o$. For (S1) to hold, we have to prove that this limit exists.

By compactness of the set Z_l , $\{\omega_k\}$ (or a subsequence of $\{\omega_k\}$) converges to a limit $\bar{\omega} = \lim_{k \rightarrow \infty} \omega_k$ [Tes18]. It now remains to show that $\gamma^*(\cdot)$ is continuous, which implies that $\gamma^*(\bar{\omega}) = \lim_{k \rightarrow \infty} \gamma^*(\omega_k) = \gamma^o$ so that (S1) holds [Tes18]. For a given $\bar{\epsilon} > 0$,

we define $\bar{\delta} := \frac{\bar{\epsilon}^l}{2\kappa(P) \max_{\mathbf{A} \in \mathcal{M}^l} \|\mathbf{A}\| + \bar{\epsilon}^l}$. From $\bar{\omega} = \{(x_i, \mathbf{j}_i) \in Z_l, i = 1, 2, \dots, d_{quad} + 1\}$, we

define a set $\bar{\omega}' = \{(x'_i, \mathbf{j}_i), i = 1, 2, \dots, d_{quad} + 1\}$ such that $\Delta x_i = \|x_i - x'_i\| \leq \bar{\delta}$ for $i = 1, 2, \dots, d_{quad} + 1$. Then, we prove that $|\gamma^*(\bar{\omega}') - \gamma^*(\bar{\omega})| \leq \bar{\epsilon}$. Indeed,

1. Let us consider $x, \mathbf{j}$ and γ^* such that $\|\Delta x\| \leq \bar{\delta}$ and $\|\mathbf{A}_j x\|_P \leq \gamma^{*l} \|x\|_P$. We then prove that $\|\mathbf{A}_j(x + \Delta x)\|_P \leq (\gamma^{*l} + \bar{\epsilon}^l) \|x + \Delta x\|_P$.

Observe that by definition of Problem (3.2), $\gamma^{*l} \leq \max_{\mathbf{A} \in \mathcal{M}^l} \|\mathbf{A}\|$ which is a feasible solution associated to $P = I_n$, and that $\|x + \Delta x\|_P \geq \sqrt{\lambda_{\min}(P)}(\|x\| - \|\Delta x\|)$.

Hence, by a reasoning similar to the derivation of Proposition 3.2, we have:

$$\begin{aligned}
\|\mathbf{A}_{\mathbf{j}}(x + \Delta x)\|_P &\leq \left(\gamma^{*l} + (\gamma^{*l} + \max_{\mathbf{A} \in \mathcal{M}^l} \|\mathbf{A}\|) \bar{\delta} \frac{\sqrt{\lambda_{\max}(P)}}{\|x + \Delta x\|_P} \right) \|x + \Delta x\|_P \\
&\leq \left(\gamma^{*l} + 2 \max_{\mathbf{A} \in \mathcal{M}^l} \|\mathbf{A}\| \bar{\delta} \frac{\kappa(P)}{1 - \bar{\delta}} \right) \|x + \Delta x\|_P \\
&\leq (\gamma^{*l} + \bar{\epsilon}^l) \|x + \Delta x\|_P
\end{aligned}$$

by definition of $\bar{\delta}$.

2. Conversely, consider x , $\mathbf{j}$ and γ^* such that $\|\Delta x\| \leq \bar{\delta}$ and $\|\mathbf{A}_{\mathbf{j}}(x + \Delta x)\|_P \leq \gamma^{*l} \|x + \Delta x\|_P$. We then prove that $\|\mathbf{A}_{\mathbf{j}}x\|_P \leq (\gamma^{*l} + \bar{\epsilon}^l) \|x\|_P$. Indeed, then,

$$\begin{aligned}
\|\mathbf{A}_{\mathbf{j}}x\|_P &\stackrel{(3.12)}{\leq} \|\mathbf{A}_{\mathbf{j}}(x + \Delta x)\|_P + \|\mathbf{A}_{\mathbf{j}}\Delta x\|_P \leq \gamma^{*l} \|x + \Delta x\|_P + \|\mathbf{A}_{\mathbf{j}}\| \bar{\delta} \|L\| \\
&\leq_{\|x\|=1} \left(\gamma^{*l} + (\gamma^{*l} + \max_{\mathbf{A} \in \mathcal{M}^l} \|\mathbf{A}\|) \bar{\delta} \kappa(P) \right) \|x\|_P \\
&\leq \left(\gamma^{*l} + 2 \max_{\mathbf{A} \in \mathcal{M}^l} \|\mathbf{A}\| \bar{\delta} \frac{\kappa(P)}{1 - \bar{\delta}} \right) \|x\|_P \leq (\gamma^{*l} + \bar{\epsilon}^l) \|x\|_P
\end{aligned}$$

Hence, for a variation of the initial states in ω_k smaller than $\bar{\delta}$, we have proved that $|\gamma^*(\bar{\omega}') - \gamma^*(\bar{\omega})| \leq \bar{\epsilon}$. In addition, from the convergence of $\{\omega_k\}$, for any $0 < \epsilon < 1$, there exists a $l \geq 0$ such that the components in ω_k corresponding to the modes (i.e. $\mathbf{j}_i$) do not change (or have reached the limit) for all $k > l$, which concludes the proof.

Appendix C

Non-degeneracy of Problem (3.14)

Following an argument of [BJW21], we prove that if there exists no matrix $\mathbf{A} \in \mathcal{M}^l$ such that $\mathbf{A}^T \mathbf{A} = \gamma I_n$ for some $\gamma \in \mathbb{R}_{\geq 0}$, then Problem (3.14) is non-degenerate with probability one.

Suppose by contradiction that the probability that $\omega_{N_1} = (x, \mathbf{j}) \in Z_l$ and $\omega_{N'} = (x', \mathbf{j}') \in Z_l$ are two distinct essential sets for Problem (3.14) is non zero. Consequently, the probability that $(x, \mathbf{j})$ is an essential set for Problem (3.4) and that $\|\mathbf{A}_{\mathbf{j}'} x'\| = \lambda^*(\omega_{N_1})$, where $\lambda^*(\omega_{N_1})$ is the solution to Problem (3.14), is non-zero. However, for any non-zero polynomial on $\mathbb{R}^n$ $P(x)$, the set of points $x \in \mathbf{R}^n$ satisfying $P(x) = 0$ has zero Lebesgue measure [Tes18, Problem 2.15]. Hence, if $\|\mathbf{A}_{\mathbf{j}'} x'\| - \lambda^*(\omega_{N_1}) = 0$ holds with a non-zero probability, $\|\mathbf{A}_{\mathbf{j}'} x\| - \lambda^*(\omega_{N_1}) = 0$ for all $x \in \mathbb{S}$, which contradicts the assumption that there exists no matrix $\mathbf{A} \in \mathcal{M}^l$ such that $\mathbf{A}^T \mathbf{A} = \gamma I_n$ for some $\gamma \in \mathbb{R}_{\geq 0}$.

Since ω_{N_1} and ω_{N_2} are arbitrary, the probability that there exist two distinct essential sets for Problem (3.14) is zero: for instance, if $(x, \mathbf{j}) \in \omega_N$ is an essential set for this problem, the probability that $(-x, \mathbf{j})$ also belongs to ω_N is zero.

Appendix D

Adaptation of Theorem 3.2 to the adaptive sampling setting

Theorem 5.1 adapts Theorem 3.1 to provide a bound on the JSR based on the adaptive sampling approach. Similarly, we extend the bound of Theorem 3.2 and propose a second quadratic bound, based on the last N_p samples only.

Theorem D.1

Let Assumptions 5.1 and 5.2 hold. For any $\beta \in [0, 1)$ and $\beta_1 \in [0, 1)$, let $\epsilon = m^l \left(1 - \sqrt[N_p]{\frac{1-\beta}{d_{quad}+1}}\right)$ and $\epsilon_1 = \frac{m^l}{2}(1 - \sqrt[N_p]{1-\beta_1})$. Let $(\gamma_p^*(\Omega_p), P^*(\Omega_p))$ denote the solution to $\text{Opt}_p(\omega_{N_1+\dots+N_p})$. Then, with the same notation as in Theorem 5.1, with probability at least $\beta + \beta_1 - 1$, the following holds:

$$\rho(\mathcal{M}) \leq \sqrt[2l]{\gamma_p^*(\Omega_p)^{2l} + \left(A(\epsilon_1)^2(2\Delta(\epsilon) + \Delta(\epsilon)^2) + \gamma_p^*(\Omega_p)^{2l}(2\Delta(\epsilon) - \Delta(\epsilon)^2)\right) \kappa^2(P_p^*(\Omega_p))}.$$

The proof follows exactly the same lines as the proof of Theorem 5.1.

We now compare the bounds of Theorem 5.1 and 5.12 on a numerical experiment. Figure D.1 plots the evolution of both bounds with the dimension of the system. As expected since the number of samples used in the last phase is low, the bound of Theorem D.1 beats the one of Theorem 5.1 when the dimension of the system is lower than 10. It thus outperforms the bound of Theorem 5.1 in a wider range of situations than the situations in which the bound of Theorem 3.2 outperforms the one of Theorem 3.1.

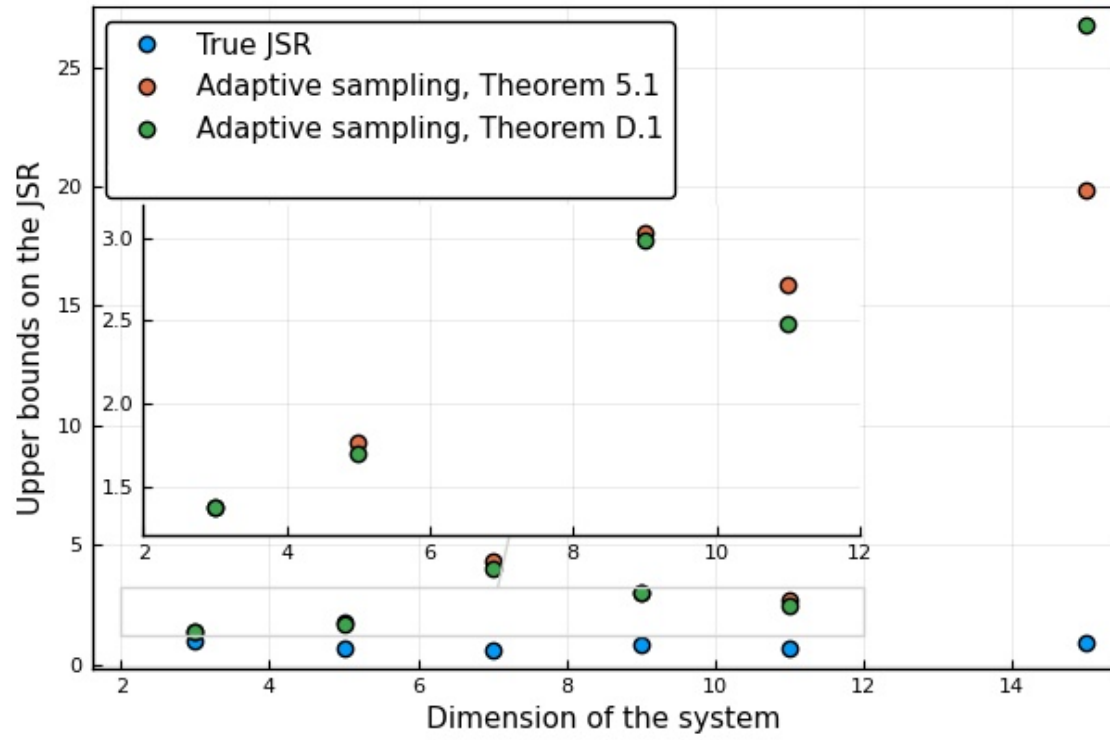


Figure D.1: Evolution of upper bounds based on adaptive sampling with the dimension of the system. The bounds are yielded by the application of Heuristic 1 on 2-modals systems of various dimension, for a total number of samples of 1500. The bounds of Theorems 5.1 and D.1, as well as the true value of the JSR are plotted. When the dimension is lower than 10, the bound of Theorem D.1 is the tightest.

Appendix E

Derivation of theoretical bounds based on all samples in the adaptive sampling approach

We consider the two-phase adaptive sampling approach where P_2 is fixed to the identity in the second phase. We provide a probabilistic bound on the JSR by analysing $\text{Opt}_{2,I}k(\omega_{N_1+N_2})$ based on all points sampled and not only on ω_{N_2} . To this aim, we consider the joint distribution of ω_{N_1} and ω_{N_2} . We base our analysis on the stage-wise coupled scenario approach with re-sampling proposed in [KML16]. In the first phase, we solve $\text{Opt}_1(\omega_{N_1})$ with N_1 samples. In the second phase, we solve $\text{Opt}_2(P_1^*(\omega_{N_1}), \omega_{N_2})$, with solution $\gamma_2^*(P_1^*(\omega_{N_1}), \omega_{N_2})$:

$$\min_{\gamma_2 \geq 0} \quad \gamma_2 \text{ s.t.} \quad \|L_1 \mathbf{A}_j L_1^{-1} x\| \leq \gamma_2^l, \quad \forall (x, \mathbf{j}) \in \omega_{N_2}, \quad (\text{E.1})$$

where $P_1^*(\omega_{N_1}) = L_1^T L_1$. Clearly, for any fixed $P_1^*(\omega_{N_1})$, $\text{Opt}_2(P_1^*(\omega_{N_1}), \omega_{N_2})$ is feasible and its minimizer is unique. Moreover, it is convex with respect to γ_2 . The solution $\gamma_{2,I}^*(\omega_{N_1+N_2})$ to $\text{Opt}_{2,I}k(\omega_{N_1+N_2})$ is given by the maximum between $\gamma_1^*(\omega_{N_1})$ and $\gamma_2^*(P_1^*(\omega_{N_1}), \omega_{N_2})$. We introduce here the notation used this appendix:

Assumption E.1. We consider an n -dimensional switched linear system as in (2.1), whose set of m modes is denoted by $\mathcal{M} \subset \mathbb{R}^{n \times n}$. Let $N_1 \geq d_{quad+1}$, $N_2 \geq 2$ and $N = N_1 + N_2$. We consider two independent samples $\omega_{N_1} \in Z_l^{N_1}$ and $\omega_{N_2} \in Z_l^{N_2}$, whose union is denoted by ω_N . We denote by $(\gamma_1^*(\omega_{N_1}), P_1^*(\omega_{N_1}))$ the solution to $\text{Opt}_1(\omega_{N_1})$ and by $\gamma_2^*(P_1^*(\omega_{N_1}), \omega_{N_2})$ the solution to $\text{Opt}_2(P_1^*(\omega_{N_1}), \omega_{N_2})$. L_1 denotes the Cholesky decomposition of $P_1^*(\omega_{N_1})$.

We will make use of Corollary 2.2. Hence, we require the following assumption on the system analyzed, allowing the problem to be non-degenerate:

Assumption E.2. The system analyzed is defined by a set $\mathcal{M} \subset \mathbb{R}^{n \times n}$ such that $\forall A \in \mathcal{M}$, there exists no $P \in \mathcal{S}_{++}^n$ and $\gamma > 0$ satisfying $A^T P A = \gamma^2 P$.

Then, by combining results from Proposition 2.3, [KML16] and [KBJT19], we obtain the following proposition:

Proposition E.1. *Let Assumptions E.1 and E.2 hold. For any $\beta_1, \beta_2 \in (0, 1)$ such that $\beta = \beta_1 + \beta_2 - 1 \in (0, 1)$, let $\epsilon_1(\beta_1, N_1) = 1 - I^{-1}(1 - \beta_1; N_1 - d_{quad} - 1, d_{quad})$ and $\epsilon_2(\beta_2, N_2) = 1 - \sqrt[N_2]{1 - \beta_2}$. Let $\epsilon(\beta_1, N_1, \beta_2, N_2) = \epsilon_1(\beta_1, N_1) + \epsilon_2(\beta_2, N_2)$. Then, with confidence at least β , it holds that:*

$$\begin{cases} (\mathbf{A}_{\mathbf{j}} x)^T P_1^*(\omega_{N_1}) \mathbf{A}_{\mathbf{j}} x \leq (\gamma_1^*(\omega_{N_1}))^{2l} x^T P_1^*(\omega_{N_1}) x \\ \|L_1 \mathbf{A}_{\mathbf{j}} L_1^{-1} x\| \leq (\gamma_2^*(P_1^*(\omega_{N_1}), \omega_{N_2}))^l \end{cases} \quad \forall x \in \mathbb{S} \setminus \tilde{\mathbb{S}}, \quad \forall \mathbf{j} \in M^l \quad (\text{E.2})$$

where $\tilde{\mathbb{S}} \subset \mathbb{S}$ has measure $\sigma^{n-1}(\tilde{\mathbb{S}}) \leq \epsilon(\beta_1, N_1, \beta_2, N_2) m^l$.

Proof. We define the following sets of violated constraints:

- $V_1(\omega_{N_1}) := \{(x, \mathbf{j}) \in Z_l : (\mathbf{A}_{\mathbf{j}} x)^T P_1^*(\omega_{N_1}) \mathbf{A}_{\mathbf{j}} x > (\gamma_1^*(\omega_{N_1}))^{2l} x^T P_1^*(\omega_{N_1}) x\}$: set of violated constraints in $\text{Opt}_1(\omega_{N_1})$.
- $V_2(P_1^*(\omega_{N_1}), \omega_{N_2}) := \{(x, \mathbf{j}) \in Z_l : \|L_1 \mathbf{A}_{\mathbf{j}} L_1^{-1} x\| > (\gamma_2^*(P_1^*(\omega_{N_1}), \omega_{N_2}))^l\}$: set of violated constraints in $\text{Opt}_2(P_1^*(\omega_{N_1}), \omega_{N_2})$.
- $V(\omega_N) := \{(x, \mathbf{j}) \in Z_l : (\mathbf{A}_{\mathbf{j}} x)^T P_1^*(\omega_{N_1}) \mathbf{A}_{\mathbf{j}} x > (\gamma_1^*(\omega_{N_1}))^{2l} x^T P_1^*(\omega_{N_1}) x \text{ or } \|L_1 \mathbf{A}_{\mathbf{j}} L_1^{-1} x\| > (\gamma_2^*(P_1^*(\omega_{N_1}), \omega_{N_2}))^l\}$: union of violated constraints.

By Corollary 2.2, it holds that

$$\mu_l^{N_1} \{\omega_{N_1} \in Z_l^{N_1} : \mu_l(V_1(\omega_{N_1})) \leq \epsilon_1(\beta_1, N_1)\} \geq \beta_1.$$

By Corollary 2.1, it holds that

$$\mu_l^{N_2} \{\omega_{N_2} \in Z_l^{N_2} : \mu_l(V_2(P_1^*(\omega_{N_1}), \omega_{N_2})) \leq \epsilon_2(\beta_2, N_2)\} \geq \beta_2.$$

Then, we obtain the following joint result from [KML16, Prop. 2]:

$$\mu_l^N \{\omega_N \in Z_l^N : \mu_l(V(\omega_N)) \leq \epsilon\} \geq \beta.$$

Projecting $V(\omega_N)$ on $\mathbb{S}$ in a similar way as in [KBJT19, Cor. 1] concludes the proof. $\square$

Since $P_1^*(\omega_{N_1}) = L_1^T L_1$, we can rewrite Equation (E.2) as follows:

$$\begin{aligned} (L_1 \mathbf{A}_{\mathbf{j}} L_1^{-1} x)^T L_1 \mathbf{A}_{\mathbf{j}} L_1^{-1} x &\leq (\gamma_1^*(\omega_{N_1}))^{2l} x^T x, \quad \forall x \in L_1(\mathbb{S} \setminus \tilde{\mathbb{S}}), \quad \forall \mathbf{j} \in M^l \\ \|L_1 \mathbf{A}_{\mathbf{j}} L_1^{-1} x\| &\leq (\gamma_2^*(P_1^*(\omega_{N_1}), \omega_{N_2}))^l, \quad \forall x \in \mathbb{S} \setminus \tilde{\mathbb{S}}, \quad \forall \mathbf{j} \in M^l. \end{aligned}$$

By homogeneity of the dynamics, the first inequality becomes:

$$\begin{aligned} (L_1 \mathbf{A}_{\mathbf{j}} L_1^{-1} x)^T L_1 \mathbf{A}_{\mathbf{j}} L_1^{-1} x &\leq (\gamma_1^*(\omega_{N_1}))^{2l} x^T x, \quad \forall x \in \mathbb{S} \setminus \Pi_{\mathbb{S}}(L_1(\tilde{\mathbb{S}})), \quad \forall \mathbf{j} \in M^l \\ \iff \|L_1 \mathbf{A}_{\mathbf{j}} L_1^{-1} x\| &\leq (\gamma_1^*(\omega_{N_1}))^l, \quad \forall x \in \mathbb{S} \setminus \Pi_{\mathbb{S}}(L_1(\tilde{\mathbb{S}})), \quad \forall \mathbf{j} \in M^l \end{aligned}$$

Hence, since $\gamma_{2,I}^*(\omega_{N_1+N_2}) = \max(\gamma_1^*(\omega_{N_1}), \gamma_2^*(P_1^*(\omega_{N_1}), \omega_{N_2}))$, we have:

$$\|L_1 \mathbf{A}_{\mathbf{j}} L_1^{-1} x\| \leq (\gamma_{2,I}^*(\omega_{N_1+N_2}))^l, \quad \forall x \in \mathbb{S} \setminus (\tilde{\mathbb{S}} \cap \Pi_{\mathbb{S}}(L_1(\tilde{\mathbb{S}}))), \quad \forall \mathbf{j} \in M^l$$

Let us denote $\sigma^{n-1}(\tilde{\mathbb{S}} \cap \Pi_{\mathbb{S}}(L_1(\tilde{\mathbb{S}}))) = \sigma^{n-1}(\{x \in \tilde{\mathbb{S}} : \exists y \in \tilde{\mathbb{S}} \text{ s.t. } x = \frac{L_1 y}{\|L_1 y\|}\})$ by $\tilde{\epsilon}(\beta_1, N_1, \beta_2, N_2)$. Then, following the same lines as in the derivation of [KBJT19, Thm. 5], we obtain the following upper bound on the JSR:

Theorem E.1

Let Assumptions E.1 and E.2 hold. Then, for any $\beta_1, \beta_2 \in (0, 1)$ such that $\beta = \beta_1 + \beta_2 - 1 \in (0, 1)$, with probability at least β , it holds that:

$$\rho(\mathcal{M}) \leq \frac{\gamma_{p,I}^*(\omega_{N_1+\dots+N_p})}{\sqrt[p]{\delta(\frac{1}{2}\tilde{\epsilon}(\beta_1, N_1, \beta_2, N_2))}} \quad (\text{E.3})$$

where $\tilde{\epsilon}(\beta_1, N_1, \beta_2, N_2) = \sigma^{n-1}(\tilde{\mathbb{S}} \cap \Pi_{\mathbb{S}}(L_1(\tilde{\mathbb{S}})))$,
 $\sigma^{n-1}(\tilde{\mathbb{S}}) = m^l \left(1 - I^{-1}(1 - \beta_1; N_1 - d_{quad} - 1, d_{quad}) + 1 - \sqrt[N_2]{1 - \beta_2}\right)$,
and $\delta(\cdot)$ is given by (3.7).

The bound of Theorem E.1 can outperform the one of Theorem 5.2 only if $\tilde{\epsilon}(\beta_1, N_1, \beta_2, N_2) < \epsilon(\beta_1, N_1, \beta_2, N_2)$. Indeed, otherwise, it is too conservative. Unfortunately, without further assumptions, it can be the case that $\sigma^{n-1}(\tilde{\mathbb{S}} \cap \Pi_{\mathbb{S}}(L_1(\tilde{\mathbb{S}}))) = \sigma^{n-1}(\tilde{\mathbb{S}})$, as illustrated on Figure E.1. However, we believe that if $\epsilon(\beta_1, N_1, \beta_2, N_2)$ is sufficiently high, the measures of these two sets do not coincide. We leave investigation of this track for further work.

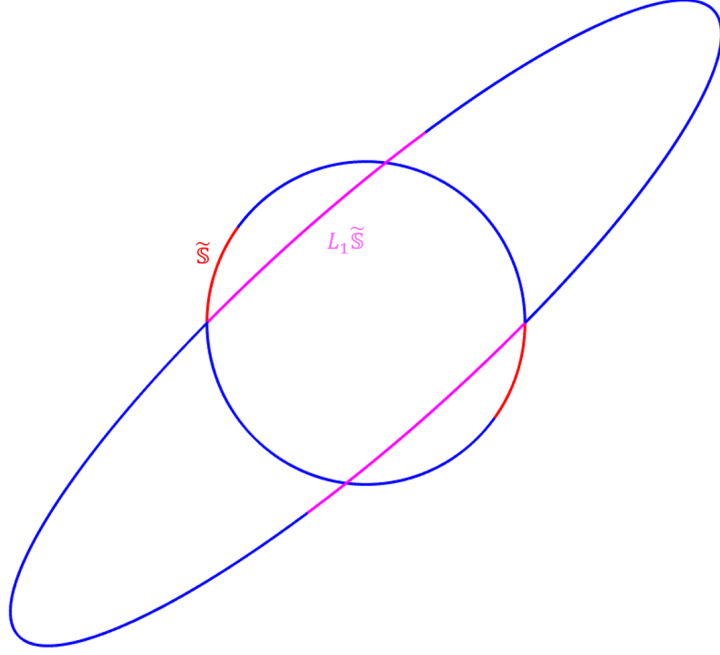


Figure E.1: Illustration of a case, in two dimensions, where $\tilde{\mathbb{S}} \cap \Pi_{\mathbb{S}}(L_1(\tilde{\mathbb{S}})) = \tilde{\mathbb{S}}$. $\tilde{\mathbb{S}}$ is presented in red and $L_1(\tilde{\mathbb{S}})$ is presented in pink. Clearly, the projection of $L_1(\tilde{\mathbb{S}})$ on the sphere covers $\tilde{\mathbb{S}}$.

In a second time, we show why we cannot straightforwardly extend our sensitivity analysis to provide a bound on the JSR based on all trajectories sampled, in the adaptive sampling approach. We consider the case where the number of phases is fixed to 2.

Let $\gamma'^o(L_1)$ be the solution to $\text{Opt}(L_1)$. By Lemma 3.3 and the same argument as in the proof of Theorem 5.1, there exists a set $\omega_s \in Z_l^N$, with $|\omega_s| = d_{quad} + 1$, such that $\gamma^*(\omega_s) = \gamma'^o(L_1)$. $\gamma^*(\omega'_s)$ is the optimal cost of:

$$\begin{aligned} \min_{P, \gamma \geq 0} \quad & \text{s.t. } (\mathbf{L}_{\mathbf{p}-1} \mathbf{A}_{\mathbf{j}} \mathbf{L}_{\mathbf{p}-1}^{-1} x)^T P \mathbf{L}_{\mathbf{p}-1} \mathbf{A}_{\mathbf{j}} \mathbf{L}_{\mathbf{p}-1}^{-1} x \leq \gamma^{2l} x^T P x, \quad \forall (x, \mathbf{j}) \in \omega'_s \\ & P \succeq I_n, \quad \|P\| \leq C \text{ for a large } C \in \mathbb{R}_{\geq 0}. \end{aligned}$$

We now seek to extend Proposition 3.1 to take into account the constraints of $\text{Opt}_2(\omega_{N_1+N_2})$ associated not only to the trajectories in ω_{N_2} , but also to the trajectories in ω_{N_1} . Since N_2 is uniformly sampled in the system considered, $\forall (x', \mathbf{j}) \in \omega_s$, $\mu_l^{N_2} \{\omega_{N_2} \in Z_l^{N_2} \mid x_i \notin \mathcal{C}(x', \epsilon) \text{ or } \mathbf{j}_i \neq \mathbf{j}' \mid \forall (x_i, \mathbf{j}_i) \in \omega_{N_2}\} = (1 - \frac{\epsilon}{m^l})^{N_2}$. This holds by the same argument as in the proof of Proposition 3.1.

If L_1 is fixed, we can also prove that $\forall (x', \mathbf{j}) \in \omega_s$, $\mu_l^{N_1} \{\omega_{N_1} \in Z_l^{N_1} \mid \frac{L_1 x_i}{\|L_1 x_i\|} \notin \mathcal{C}(x', \epsilon) \text{ or } \mathbf{j}_i \neq \mathbf{j}' \mid \forall (x_i, \mathbf{j}_i) \in \omega_{N_1}\} = (1 - \frac{\epsilon}{\sqrt{\det(L_1^T L_1) \lambda_{\min}(L_1^T L_1)^{n_{m^l}}}})^{N_2}$. Indeed, for all

$(x, \mathbf{j}) \in Z_l$, $\frac{L_1 x}{\|L_1 x\|} \in \mathcal{C}(x', \epsilon) \longleftrightarrow \frac{L_1 x}{\|L_1 x\|} \in \mathcal{C}(x', \epsilon) \longleftrightarrow x \in \pi_{\mathbb{S}}(L_1^{-1} \mathcal{C}(x', \epsilon))$, where $\pi_{\mathbb{S}}(A)$ denotes the projection of the set A on the sphere: for any $z \in A$, $\pi_{\mathbb{S}}(z) = \frac{z}{\|z\|}$. Let $\lambda(\cdot)$ be the Lebesgue measure; and for any $A \subset \mathbb{S}$, let $\mathbb{S}^A := \{tA, t \in [0, 1]\}$ and $\sigma^{n-1}(A) = \lambda(\mathbb{S}^A)$. Then, $\frac{1}{\lambda_{\max}(L)} \mathbb{S}^{LA} \subset \mathbb{S}^{\pi_{\mathbb{S}}(LA)}$ for any $A \subset \mathbb{S}$. Hence,

$$\begin{aligned} \sigma^{n-1}(\pi_{\mathbb{S}}(L_1^{-1} \mathcal{C}(x', \epsilon))) &= \lambda(\mathbb{S}^{\pi_{\mathbb{S}}(L_1^{-1} \mathcal{C}(x', \epsilon))}) \geq \lambda\left(\frac{\mathbb{S}^{L_1^{-1} \mathcal{C}(x', \epsilon)}}{\lambda_{\max}(L_1^{-1})}\right) = \frac{\lambda(L_1^{-1}(\mathbb{S}^{\mathcal{C}(x', \epsilon)}))}{\lambda_{\max}(L_1^{-1})^n} \\ &= \frac{|\det(L_1^{-1})| \lambda(\mathbb{S}^{\mathcal{C}(x', \epsilon)})}{\lambda_{\max}(L_1^{-1})^n} = \frac{\sigma^{n-1}(\mathcal{C}(x', \epsilon))}{\sqrt{\det(L_1^T L_1) \lambda_{\min}(L_1^T L_1)^n}} = \frac{\epsilon}{\sqrt{\det(L_1^T L_1) \lambda_{\min}(L_1^T L_1)^n}}. \end{aligned}$$

Then, we are tempted to combine the probabilities and improve Proposition 3.1. However, one crucial mistake has been made: we cannot consider L_1 to be fixed and ω_{N_1} to be random. Hence, the second part of the argument is wrong, and we cannot straightforwardly take into account the information provided by ω_{N_1} .

Appendix F

Comparison of the bounds of Theorems 5.1 and 5.2

We seek to compare the bounds of Theorems 5.2 and 5.1, related respectively to the cases where P_p is fixed to the identity or not. First, we relate the costs of the respective problems:

Lem F.1. *Let ω_{N_p} be a uniform random sampling for system $\mathbf{L}_{\mathbf{p}-1}\mathcal{M}^l\mathbf{L}_{\mathbf{p}-1}^{-1}$. Let $(\gamma_p^*(\omega_{N_1+\dots+N_p}), P_p^*(\omega_{N_1+\dots+N_p}))$ be the optimal solution to $\text{Opt}_p(\omega_{N_1+\dots+N_p})$, and $\gamma_{p,I}^*(\omega_{N_1+\dots+N_p})$ be the optimal solution to $\text{Opt}_{p,I}(\omega_{N_1+\dots+N_p})$. Then, it holds that*

$$\gamma_{p,I}^*(\omega_{N_1+\dots+N_p}) \leq \kappa(P_p^*(\omega_{N_1+\dots+N_p}))\gamma_p^*(\omega_{N_1+\dots+N_p}).$$

Proof. Let $\Omega = \omega_{N_p} \cup \{(x', \mathbf{j}) : x' = L_{p-1}x, (x, \mathbf{j}) \in \omega_{N_{p-1}}\} \cup \dots \cup \{(x', \mathbf{j}) : x' = \mathbf{L}_{\mathbf{p}-1}x, (x, \mathbf{j}) \in \omega_{N_1}\}$. Then, for all $P \in \mathcal{S}_{++}^n$,

$$\begin{aligned} \gamma_{p,I}^*(\omega_{N_1+\dots+N_p}) &= \max_{(x, \mathbf{j}) \in \Omega} \sqrt[l]{\|\mathbf{L}_{\mathbf{p}-1}\mathbf{A}_{\mathbf{j}}\mathbf{L}_{\mathbf{p}-1}^{-1}x\|} \leq \max_{(x, \mathbf{j}) \in \Omega} \sqrt[l]{\frac{\|L^{-1}\| \|\mathbf{L}_{\mathbf{p}-1}\mathbf{A}_{\mathbf{j}}\mathbf{L}_{\mathbf{p}-1}^{-1}x\|_P}{\frac{\|x\|_P}{\|L\|}}} \\ &= \max_{(x, \mathbf{j}) \in \Omega} \sqrt[l]{\kappa(P) \frac{\|\mathbf{L}_{\mathbf{p}-1}\mathbf{A}_{\mathbf{j}}\mathbf{L}_{\mathbf{p}-1}^{-1}x\|_P}{\|x\|_P}} \end{aligned}$$

Taking $P = P_p^*(\omega_{N_1+\dots+N_p})$ concludes the proof. $\square$

Let (γ, P) denote $(\gamma_p^*(\omega_{N_1+\dots+N_p}), P_p^*(\omega_{N_1+\dots+N_p}))$, the optimal solution to $\text{Opt}_p(\omega_{N_1+\dots+N_p})$; and let γ_I denote $\gamma_{p,I}^*(\omega_{N_1+\dots+N_p})$, the solution to $\text{Opt}_{p,I}(\omega_{N_1+\dots+N_p})$. As observed before, $\lambda_p^*(\omega_{N_1+\dots+N_p})$, the solution to Problem (5.8), is given by $\lambda_p^*(\omega_{N_1+\dots+N_p}) = \gamma_I^l$. For given $\beta, \beta_1 \in (0, 1)$; let $\beta_1 = \beta + \beta_1 - 1$,

$\epsilon = m^l \left(1 - \sqrt[p]{\frac{1f-\beta}{d_{quad}+1}}\right)$, $\epsilon_1 = \frac{m^l}{2}(1 - \sqrt[p]{1-\beta_1})$ and $\bar{\epsilon}_1 = \frac{m^l}{2}(1 - \sqrt[p]{1-\bar{\beta}_1})$. We seek to prove that in the context of adaptive sampling, the bound of Theorem 5.1 outperforms the bound of Theorem 5.2. This means that $\forall \beta, \beta_1 \in (0, 1)$,

$$\begin{aligned} \frac{\gamma_I}{\sqrt[l]{\delta(\bar{\epsilon}_1)}} &\leq \sqrt[l]{\gamma^l + (\gamma^l + \frac{\gamma_I^l}{\delta(\epsilon_1)})\sqrt{2-2\delta(\epsilon)\kappa(P)}} \\ \iff \frac{\gamma_I^l}{\delta(\bar{\epsilon}_1)} &\leq \gamma^l + (\gamma^l + \frac{\gamma_I^l}{\delta(\epsilon_1)})\sqrt{2-2\delta(\epsilon)\kappa(P)}, \end{aligned}$$

where $\delta(\cdot)$ is given by (3.7). Note that $\delta(\cdot)$ is a decreasing function. Hence, when $\beta_1 = \beta$,

$$1 - \bar{\beta}_1 \geq 1 - \beta_1 \geq \frac{1 - \beta}{d_{quad} + 1} \implies \bar{\epsilon}_1 \leq \epsilon_1 \leq \epsilon \implies \delta(\bar{\epsilon}_1) \geq \delta(\epsilon_1) \geq \delta(\epsilon). \quad (\text{F.1})$$

In the adaptive sampling context, we consider that $\kappa(P_p^*(\omega_{N_1+\dots+N_p}))$ is relatively low or that N_p , the number of trajectories sampled in the last phase, is itself low. We consider the following cases:

1. $\sqrt{2-2\delta(\epsilon)\kappa(P)} \geq 1$. Then, since $\frac{\gamma_I^l}{\delta(\epsilon_1)}$ is non-negative,

$$\begin{aligned} \frac{\gamma_I^l}{\delta(\epsilon_1)} &\geq \gamma^l \frac{(1 + \sqrt{2-2\delta(\epsilon)\kappa(P)})}{(1 - \sqrt{2-2\delta(\epsilon)\kappa(P)})} \\ \implies \frac{\gamma_I^l}{\delta(\bar{\epsilon}_1)} &\stackrel{(\text{F.1})}{\leq} \frac{\gamma_I^l}{\delta(\epsilon_1)} \leq \gamma^l + (\gamma^l + \frac{\gamma_I^l}{\delta(\epsilon_1)})\sqrt{2-2\delta(\epsilon)\kappa(P)}. \end{aligned}$$

2. $\sqrt{2-2\delta(\epsilon)\kappa(P)} \leq 1$. This implies that $\delta(\epsilon) \geq \frac{1}{2}$. We analyze the function

$$f(\kappa(P), \delta(\epsilon)) = \delta(\epsilon) \frac{1 + \sqrt{2-2\delta(\epsilon)\kappa(P)}}{\kappa(P)^l - \sqrt{2-2\delta(\epsilon)\kappa(P)}^{l+1}} - 1.$$

Indeed, if $f(\kappa(P), \delta(\epsilon))$ is non-negative for all $\kappa(P) \geq 1$, $\delta(\epsilon) \in (0.5, 1)$, then

$$\begin{aligned} \frac{\kappa(P)^l}{\delta(\epsilon)}(1 - \sqrt{2-2\delta(\epsilon)\kappa(P)}) &\leq 1 + \sqrt{2-2\delta(\epsilon)\kappa(P)} \\ \iff \frac{\gamma_I^l}{\delta(\epsilon_1)}(1 - \sqrt{2-2\delta(\epsilon)\kappa(P)}) &\stackrel{\text{Lem. F.1}}{\leq} \frac{\kappa(P)^l \gamma^l}{\delta(\epsilon_1)}(1 - \sqrt{2-2\delta(\epsilon)\kappa(P)}) \\ &\stackrel{(\text{F.1})}{\leq} \frac{\kappa(P)^l \gamma^l}{\delta(\epsilon)}(1 - \sqrt{2-2\delta(\epsilon)\kappa(P)}) \leq \gamma^l + \gamma^l \sqrt{2-2\delta(\epsilon)\kappa(P)} \\ &\stackrel{(\text{F.1})}{\iff} \frac{\gamma_I^l}{\delta(\bar{\epsilon}_1)} \leq \gamma^l + \gamma^l \sqrt{2-2\delta(\epsilon)\kappa(P)} + \frac{\gamma_I^l}{\delta(\epsilon_1)} \sqrt{2-2\delta(\epsilon)\kappa(P)}. \end{aligned}$$

The derivative of $f(\kappa(P), \delta(\epsilon))$ with respect to $\kappa(P)$ is given by:

$$\delta(\epsilon)\kappa(P)^{l-1} \frac{(2 - 2\delta(\epsilon))l\kappa(P)^2 + 2\sqrt{2 - 2\delta(\epsilon)}\kappa(P) - l}{(\kappa(P)^l - \sqrt{2 - 2\delta(\epsilon)}\kappa(P)^{l+1})^2}.$$

Its roots are given by $\kappa(P) = \frac{-1 \pm \sqrt{1+l^2}}{l\sqrt{2-2\delta(\epsilon)}}$, so that $f(\kappa(P), \delta(\epsilon))$ admits a minimum in $\kappa(P)$ when $\kappa(P) = \frac{\sqrt{1+l^2}-1}{l\sqrt{2-2\delta(\epsilon)}}$. Two subcases have to be considered:

- (a) $\frac{\sqrt{1+l^2}-1}{l\sqrt{2-2\delta(\epsilon)}} \leq 1$, implying $\delta(\epsilon) \leq \frac{\sqrt{1+l^2}}{l^2} - \frac{1}{2}$. Note that this only happens when $l = 1$. Then, $f(\kappa(P), \delta(\epsilon))$ is minimal in $\kappa(P)$ when $\kappa(P) = 1$. The derivative of $f(1, \delta(\epsilon))$ with respect to $\delta(\epsilon)$ is given by:

$$\frac{(2\delta(\epsilon) - 1)\sqrt{2 - 2\delta(\epsilon)} - 2\delta(\epsilon)}{(1 - \sqrt{2 - 2\delta(\epsilon)})^2 \sqrt{2 - 2\delta(\epsilon)}}$$

which is always negative for $\delta(\epsilon) \in (0, 1)$. Hence, $f(1, \delta(\epsilon))$ always decreases in $\delta(\epsilon)$, and achieves its minimum for $\delta(\epsilon) = 1$. Since $f(1, 1) = 0$, $f(\kappa(P), \delta(\epsilon))$ is non-negative for all $\kappa(P) \geq 1$, $\delta(\epsilon) \in (0.5, \frac{\sqrt{1+l^2}}{l^2} - \frac{1}{2})$.

- (b) $\frac{\sqrt{1+l^2}-1}{l\sqrt{2-2\delta(\epsilon)}} \geq 1$, implying $\delta(\epsilon) \geq \frac{\sqrt{1+l^2}}{l^2} - \frac{1}{2}$. Then, $f(\kappa(P), \delta(\epsilon))$ is minimal in $\kappa(P)$ when $\kappa(P) = \frac{\sqrt{1+l^2}-1}{l\sqrt{2-2\delta(\epsilon)}}$. We thus analyze

$$f\left(\frac{\sqrt{1+l^2}-1}{l\sqrt{2-2\delta(\epsilon)}}, \delta(\epsilon)\right) = \delta(\epsilon) \left(\frac{l\sqrt{2-2\delta(\epsilon)}}{\sqrt{1+l^2}-1}\right)^l \frac{1 + \frac{\sqrt{1+l^2}-1}{l}}{1 - \frac{\sqrt{1+l^2}-1}{l}} - 1.$$

Clearly, this expression is not always positive: when $\delta(\epsilon) = 1$, for instance, it is equal to -1 . However, the derivative of $f(\frac{\sqrt{1+l^2}-1}{l\sqrt{2-2\delta(\epsilon)}}, \delta(\epsilon))$ with respect to $\delta(\epsilon)$ is given by:

$$\left(\frac{l\sqrt{2-2\delta(\epsilon)}}{\sqrt{1+l^2}-1}\right)^l \frac{\sqrt{2-2\delta(\epsilon)} - \delta(\epsilon) \frac{\sqrt{1+l^2}-1}{l\sqrt{2-2\delta(\epsilon)}} - 2\left(\frac{\sqrt{1+l^2}-1}{l}\right)^2 \sqrt{2-2\delta(\epsilon)}}{\sqrt{2-2\delta(\epsilon)}(1 - \frac{\sqrt{1+l^2}-1}{l})^2}.$$

This expression, depending on l , is either always negative, or is decreasing and admits a unique root in the interval $\delta(\epsilon) \in (0.5, 1]$. Hence, since $f(\frac{\sqrt{1+l^2}-1}{l\sqrt{2-2\delta(\epsilon)}}, 0.5)$ is positive, $f(\frac{\sqrt{1+l^2}-1}{l\sqrt{2-2\delta(\epsilon)}}, \delta(\epsilon))$ remains positive until it reaches a root in the interval $(0.5, 1]$.

Let $\bar{\delta}(\epsilon)$ be the value of this root, and let $\bar{\kappa}(P)$ be equal to $\frac{\sqrt{1+l^2}-1}{l\sqrt{2-2\bar{\delta}(\epsilon)}}$. Then, it holds that $f(\kappa(P), \delta(\epsilon))$ is non-negative for all $\kappa(P) \geq 1$, $\delta(\epsilon) \in (0.5, \bar{\delta}(\epsilon))$ and for all $1 \leq \kappa(P) \leq \bar{\kappa}$, $\delta(\epsilon) \in (0.5, 1]$. At least one of these assumptions will hold in the adaptive sampling context. We illustrate this case for $l = 1$. Then, $\bar{\delta}(\epsilon) = 0.9848$ and $\bar{\kappa}(P) = 2.3776$. Hence, if either $\delta(\epsilon) \leq 0.9848$ or $\kappa(P) \leq 2.3776$ in the last phase, it holds that the bound of Theorem 5.2 always beats the one of Theorem 5.2. Note that the number of samples required to obtain $\delta(\epsilon) \geq 0.9848$ is of 1200 samples in dimension 2 and rapidly increases with the dimension, reaching 10^5 in dimension 4. Such a large number of samples is never used in practice in the last phase of the adaptive sampling approach, since it is sufficient to obtain tight results without resorting to adaptive sampling.

We have thus shown that in the context of adaptive sampling, the bound of Theorem 5.2 almost always beats the one of Theorem 5.2, especially if traces of length 1 are used. Indeed, in this context, the number of samples used in the last phase never exceeds 1200, which is sufficient to ensure this claim.

Appendix G

Additional analysis of the online sampling approach

We consider the system of Example 2.1 and the online strategy with the parameters of Chapter 6. To gain more insight with respect to the behavior of $P'^o(\mathbf{L}_k)$ (unknown in our black-box setting) and $P_k^*(\omega_{N_1+\dots+N_k})$, for $k = 1, \dots, p$, we present on Figure G.1 the evolution of these quantities with the number of phases, when 1100 samples are used. For the sake of comparison, we also plot $\kappa(P^*(\omega_N))$.

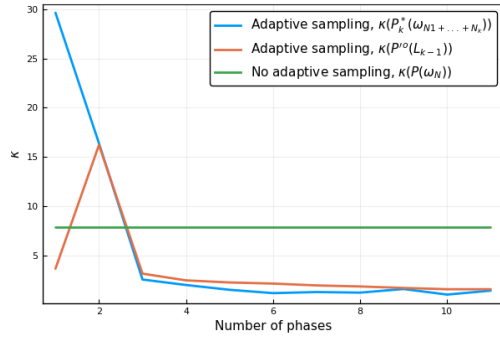


Figure G.1: Evolution of $\kappa(P_k^*(\omega_{N_1+\dots+N_k}))$ and $\kappa(P'^o(\mathbf{L}_{k-1}))$ with k , when 1100 samples are used, for the system of Example 2.1. $\kappa(P_k^*(\omega_{N_1+\dots+N_k}))$ decreases at each phase, and reaches a limit from Phase 6. For an equivalent number of samples used, it is much lower than $\kappa(P^*(\omega_N))$.

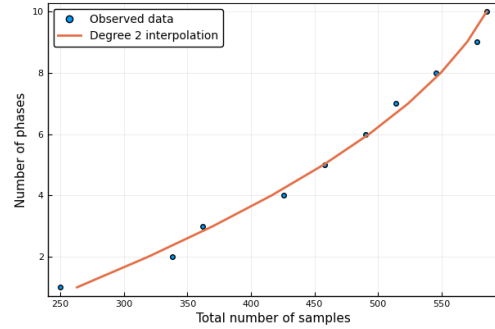


Figure G.2: Evolution of the number of phases reached by Algorithm 4 with the total number of samples, for the system of Example 2.1. In each phase, the number of samples required is determined by Procedure Online.

Figure G.2 presents the number of samples required in each phase. We observe

that after the first change of variables, $\kappa(P^{o}(L_1))$ increases compared to $\kappa(P^o)$. $\kappa(P^{o}(\mathbf{L}_k))$ then decreases with each phase, and becomes smaller than $\kappa(P^o)$ from Phase 3. This rise after the first change of variables means that the latter is not precise enough. This makes sense: for it to be sufficiently precise, a number of samples as large as the one required to obtain a tight cost directly would be needed. However, $\kappa(P_k^*(\omega_{N_1+\dots+N_k}))$ always decreases, and beats $\kappa(P(\omega_N))$ from the third phase, that is after 360 samples (less than the half of the total number of samples) are used. This decrease is due to the enforcement of the constraints associated to all samples in each phase, as predicted in Remark 5.1. This decrease in $\kappa(P_k^*(\omega_{N_1+\dots+N_k}))$ supports the validity of the convergence criterion.

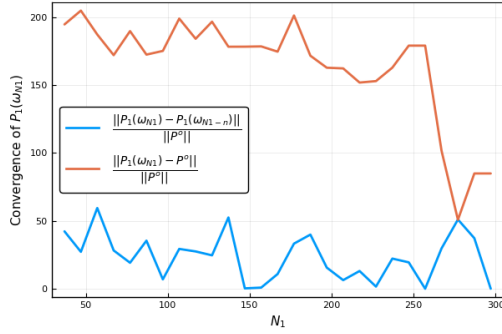


Figure G.3: Evolution of $P_1(\omega_{N_1})$ with the number of samples used in the first phase during Procedure Online, for the system of Example 2.1. We plot $\frac{\|P_1(\omega_{N_1}) - P_1(\omega_{N_1-n})\|}{\|P^o\|}$, which determines the number of samples used in Phase 1, and $\frac{\|P_1(\omega_{N_1}) - P^o\|}{\|P^o\|}$. The latter decreases more slowly, implying that the convergence of $P_1(\omega_{N_1})$ to P^o is not ensured by Procedure Online.

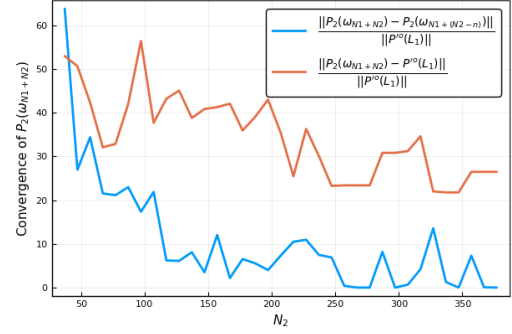


Figure G.4: Evolution of $P_2(\omega_{N_1+N_2})$ with the number of samples used in the second phase during Procedure Online, for the system of Example 2.1. We plot $\frac{\|P_2(\omega_{N_1+N_2}) - P_2(\omega_{N_1+(N_2-n)})\|}{\|P^{o}(L_1)\|}$ and $\frac{\|P_2(\omega_{N_1+N_2}) - P^{o}(L_1)\|}{\|P^{o}(L_1)\|}$. Both quantities are smaller than in Phase 1, and $\frac{\|P_2(\omega_{N_1+N_2}) - P^{o}(L_1)\|}{\|P^{o}(L_1)\|}$ is larger than $\frac{\|P_2(\omega_{N_1+N_2}) - P_2(\omega_{N_1+(N_2-n)})\|}{\|P^{o}(L_1)\|}$.

To explain the rise in $\kappa(P^0(\mathbf{L}_{k-1}))$ after the first change of variables, we present the evolution of $\frac{\|P_k(\omega_{N_1+\dots+N_k}) - P_k(\omega_{N_1+\dots+(N_k-n)})\|}{\|P^{o}(\mathbf{L}_k)\|}$ and $\frac{\|P_k(\omega_{N_1+\dots+N_k}) - P^{o}(\mathbf{L}_k)\|}{\|P^{o}(\mathbf{L}_k)\|}$ during Procedure Online on Figures G.3 and G.4, for $k = 1$ and $k = 2$ respectively. As expected, $\frac{\|P_k(\omega_{N_1+\dots+N_k}) - P^{o}(\mathbf{L}_k)\|}{\|P^{o}(\mathbf{L}_k)\|}$ is larger than $\frac{\|P_k(\omega_{N_1+\dots+N_k}) - P_k(\omega_{N_1+\dots+(N_k-n)})\|}{\|P^{o}(\mathbf{L}_k)\|}$ and decreases more slowly than the latter. Hence, the convergence of $P_k(\omega_{N_1+\dots+N_k})$ as defined in Procedure Online does not ensure the convergence of $P_k(\omega_{N_1+\dots+N_k})$ to $P^{o}(\mathbf{L}_k)$. This explains the rise in $\kappa(P^0(\mathbf{L}_{k-1}))$ after the first change of variables. However, we also observe that this distance $\frac{\|P_k(\omega_{N_1+\dots+N_k}) - P^{o}(\mathbf{L}_k)\|}{\|P^{o}(\mathbf{L}_k)\|}$ decreases at each phase, which explains the performance of our method.

Appendix H

Derivation of Table 6.2

We approximate the computational complexity of the three different approaches as follows. Every problem encountered is a quasi-linear problem solved by looking for γ (or γ_k , for $k = 1, 2, \dots, p$) by bisection on an interval $[0, U]$, with a fixed accuracy α , which requires $\left\lceil \log_2(\frac{U}{\alpha}) \right\rceil$ iterations. At each iteration, a feasibility SDP problem is solved. The complexity of this problem is polynomial in the number of constraints, that is in the number of samples used; and in d_{quad} , the dimension of the problem. Hence, the computational complexity of the initial method is of $\left\lceil \log_2(\frac{U}{\alpha}) \right\rceil * \text{poly}(4000, d_{quad})$. In the offline strategy, for 5 phases used, we solve 5 quasi-linear problems, with respectively N_1 , $N_1 + N_2$, ..., and $N_1 + N_2 + N_3 + N_4 + N_5$ constraints. Hence, the computational complexity is equal to

$$\left\lceil \log_2(\frac{U}{\alpha}) \right\rceil * \text{poly}(5N_1 + 4N_2 + 3N_3 + 2N_4 + N_5, d_{quad}).$$

In the online strategy, at phase k , we solve $\frac{N_k}{n} + 1$ quasi-linear problems, with $N_1 + \dots + N_{k-1}$, $N_1 + \dots + N_{k-1} + n$, ..., and $N_1 + \dots + N_{k-1} + N_k$ constraints, except for the last phase, in which we only solve one quasi-linear problem with $N_1 + N_2 + \dots + N_p$ constraints. Hence, for 5 phases, the computational complexity is equal to

$$\begin{aligned} & \left\lceil \log_2(\frac{U}{\alpha}) \right\rceil * \text{poly}((N_{init}(\frac{N_1 - N_{init}}{n} + 1) + \frac{n}{2}(\frac{N_1 - N_{init}}{n} + 1)\frac{N_1 - N_{init}}{n}) + \dots \\ & + ((N_1 + N_2 + N_3)(\frac{N_4}{n} + 1) + \frac{n}{2}(\frac{N_4}{n} + 1)\frac{N_4}{n}) + \sum_{i=1}^5 N_i, d_{quad}). \end{aligned}$$

Appendix I

Overview of the Julia codes implemented

We briefly present the different functions implemented and the structure linking them. The inputs, outputs, and way to use the functions is heavily commented in the code itself. The code is separated in two parts: the implementation of the functions designed in Part 2, and the implementation of the algorithms of Part 3. They are respectively available as `Rubbens_68821600_2021_Appendix1.ipynb` and `Rubbens_68821600_2021_Appendix2.ipynb`.

To use the functions, several packages have to be installed:

- `Random.jl`: to generate systems and trajectories.
- `LinearAlgebra.jl`: to deal with matrices.
- `JuMP.jl` and `CSDP.jl`: to solve the SDP optimization programs.
- `Ipopt.jl`: to solve the least square approximation of Problem (7.3).
- `IncGammaBeta.jl`: to compute the regularized inverse beta function.
- `HybridSystems.jl` and `SwitchOnSafety.jl`: to compute the true value of the JSR of switched linear systems as approximated by the JSR toolbox [LJP16].
- `Plots.jl` and `PyPlot.jl`: to illustrate the implementations.

Tables G.1 to G.7 present the purpose of the different functions implemented. Each table groups functions related to a specific part of the work.

<i>generatesystem</i>	Generation of a system of m modes in $\mathbb{R}^{n \times n}$, drawn uniformly, independently and randomly on the sphere.
<i>generatetrajectories</i>	Generation of a set of trajectories $\omega_N \in Z_l^N$.

Table I.1: Generation of random systems and trajectories

<i>trueJSR</i>	Computation of the true JSR of a system using the JSR toolbox.
<i>trueSol</i>	Computation of P^o or P_{sos}^o , solution to Problem (2.3) or (2.9).

Table I.2: White-box computations

<i>dlift</i>	Computation of the d -lift of a vector $x \in \mathbb{R}^n$.
<i>dliftmat</i>	Computation of the d -lift of a matrix $L \in \mathbb{R}^{n \times n}$.
<i>inversedlift</i>	Computation of the least square solution to Problem (7.3).

Table I.3: Functions required by the use of SOS Lyapunov functions

<i>proba_bounds</i>	If $d = 1$: Resolution of Problems (3.2) and (3.14), and computation of probabilistic upper bounds on the JSR: bounds of Theorems 3.1 and 3.2, and bound of [WJ21]. If $d = 2$: Resolution of Problems (3.3) and (3.14) and computation of a probabilistic upper bound on the JSR: bound of Theorem 3.3.
---------------------	--

Table I.4: Implementation of the approach of Chapter 3

<i>comparison</i>	Experiments of Sections 4.1 and 4.5.
<i>trace</i>	Experiment of Section 4.2.
<i>convergence</i>	Experiments of Section 4.3 and 4.4.

Table I.5: Implementation of the experiments of Chapter 4

<i>proba_bounds_offline_sampling</i>	Resolution of $\text{Opt}_k(\omega_{N_1+\dots+N_k})$, the k^{th} problem to solve in adaptive sampling.
<i>algo_offline_sampling</i>	Implementation of Algorithm 3, the p -phases offline sampling algorithm.
<i>procedure_online_sampling</i>	Implementation of Procedure Online, determining the number N_k of samples to use in Phase k .
<i>algo_online_sampling</i>	Implementation of Algorithm 4, the p -phases online sampling algorithm.
<i>algo_fast</i>	Implementation of Algorithm 1, the fast algorithm.
<i>heuristic</i>	Implementation of Heuristic 1.

Table I.6: Implementations of the algorithms of Chapter 5, 6 and 7

<i>offline</i>	Experiment of Section 5.7.1.
<i>online</i>	Experiments of Sections 5.7.1 and 6.2.
<i>online_several_dimensions</i>	Experiments of Section 5.7.2.

Table I.7: Implementation of the experiments of Chapters 6 and 7

UNIVERSITÉ CATHOLIQUE DE LOUVAIN

École polytechnique de Louvain

Rue Archimède, 1 bte L6.11.01, 1348 Louvain-la-Neuve, Belgique | www.uclouvain.be/epl