UCL Université catholique de Louvain

École polytechnique de Louvain (EPL)

epl

Sum-Of-Squares programming on Path-complete graphs for switching systems

Dissertation presented by
Benoît LEGAT

for obtaining the Master's degree in
Mathematical Engineering

Supervisor(s)
Raphaël M. JUNGERS

Reader(s)
Matthew PHILIPPE, Laurent JACQUES

Academic year 2015–2016

Acknowledgements

My first thanks would go to my thesis advisor, Prof. Raphaël Jungers. You were always present, in every step of this thesis, providing me essential advice and invaluable help. You put so much effort in improving my writing, I am truly grateful for that.

Secondly, I would like to thank Prof. Pablo Parrilo for his trust in accepting my research stay at MIT, for all our meeting and precious advices who taught me so much, and for our continued collaboration.

Additionally, a special thanks of course to my readers: Laurent Jacques and Matthew Philippe. I hope you will appreciate reading this master's thesis.

I would also like to thank those who helped me in my first research experiences. First Prof. Bernadette Govaerts for our project in Metabolomics in summer 2014, Prof. Jean-Charles Delvenne for our project on PageRank optimization the first semester of 2015, Prof. Michel Goemans for your help on MaxCut at MIT, Prof. Anthony Papavasiliou for your help with the stochastic programming part of the master's thesis. Our interactions in these early days of my research have been crucial.

Of course, help did not only come from professors. I am truly grateful to Matthew Philippe for your priceless help in constrained switched systems, Nikos Athanasopoulos for our compelling discussions on invariant sets at HSCC and Arnaud Cerckel for your decisive help on stochastic programming.

To end this acknowledgements, I would like to thank my family, without which I would not be who I am and Léa Paulus, for her endless love and support.

Contents

List of Notation **vi**

Introduction **x**

1 State of the art **2**
1.1 Polyhedral Computation . . . 3
1.1.1 Convex Analysis . . . 3
1.1.2 Polyhedron and Polytope . . . 5
1.2 Sum of Square Programming . . . 7
1.2.1 Sum of Squares Convexity . . . 7
1.3 Switching Systems . . . 9
1.3.1 CJSR Approximation via SOS . . . 10
1.4 Computation of invariant set for switched systems . . . 13
1.4.1 Polyhedral computation of a minimal invariant superset . . . 13
1.4.2 Polyhedral computation of a maximal invariant subset . . . 14
1.4.3 Nonlinear approximation of invariant sets . . . 15
1.5 The Entropic Cone . . . 20
1.5.1 The entropic cone of 1 variable . . . 21
1.5.2 Kullback-Leibler divergence and convexity . . . 22
1.5.3 The Shannon inequalities . . . 23
1.5.4 The entropic cone of 2 variables . . . 24
1.5.5 The entropic cone of 3 variables . . . 25
1.5.6 The entropic cone of 4 variables . . . 27
1.6 Sculpting the Entropic Cone . . . 29
1.6.1 Generating non-shannon inequalities . . . 29
1.6.2 Formalizing the generation of non-shannon inequalities . . . 32

2 Generating unstable trajectories for switched systems **34**
2.1 Low rank reduction . . . 35
2.2 The p-radius . . . 36
2.3 Approximation guarantees . . . 41
2.4 Finding high-growth sequences . . . 43
2.4.1 Dual SOS program . . . 43
2.4.2 Certificate provided by the dual . . . 44
2.4.3 Randomized algorithm . . . 48
2.4.4 Producing lower bounds . . . 51
2.4.5 Improving the automaton-dependent bounds . . . 52
2.5 Conclusion . . . 53

3 Efficient approximation of controllable sets for switching systems and application to the entropic cone **56**

3.1 Introduction 56
3.2 Computing controllable sets for switched systems 57
3.2.1 Definition of controllable sets 57
3.2.2 Why computing controllable sets is difficult 58
3.2.3 Parallel optimization on controllable sets 58
3.2.4 Nonlinear approximation of controllable sets 60
3.3 The entropic cone as a controllable set 62
3.4 Conclusion 62

Conclusion **64**

References **66**

List of Notation

Basics:

fields, rights	$\mathbb{R}, \mathbb{C}, \mathbb{P}, \mathbb{Q}, \mathbb{Z}$
nonnegative integers	$\mathbb{N}$
standard basis vectors	e_i
$[m]$	$\{1, \dots, m\}$
$A^{-1}\mathcal{S}$ where A is a matrix and $\mathcal{S}$ is a set	$\{ x \mid Ax \in \mathcal{S} \}$

Matrices:

$m \times n$ matrices	$\mathbb{R}^{m \times n}$
transpose of A	$A^\top$
ith column of A	$A_{:,i}$
ith row of A	$A_{i,:}$
matrix of ith to nth column of A	$A_{:,i:}$
matrix of ith to mth row of A	$A_{i:,:}$
p-norm	$\|\cdot\|_p$
$n \times n$ symmetric matrices	$\mathcal{S}^n$
$n \times n$ positive semidefinite matrices	$\mathcal{S}^n_{++}$
$n \times n$ positive definite matrices	$\mathcal{S}^n_{+}$
positive semidefinite	$\succeq$
positive definite	$\succ$

Graphs:

$G(V, E)$	Graph with nodes V and edges E
$E^\top$	$\{ (v, u, \sigma) : (u, v, \sigma) \in E \}$
$G^\top(V, E^\top)$	Graph with reversed edges
E_k	subset of E^k containing all valid paths of length k
$s(i)$ where s is a path	ith node of the path
$s[i]$ where s is a path	ith edge of the path
$s(i:)$ where $s \in E^k$	$(s(i), \dots, s(k))$
$E_k(u, v)$	$\{ s \in E_k^- \mid s(1) = u, s(k+1) = v \}$
$E_k^-(v)$	$\{ s \in E_k^- \mid s(k+1) = v \}$
$E_k^+(v)$	$\{ s \in E_k^+ \mid s(1) = v \}$
$E_k^-[e]$	$\{ s \in E_k^- \mid s[k] = e \}$
$E_k^+[e]$	$\{ s \in E_k^+ \mid s[1] = e \}$

$d^-(v)$	indegree of v
$d^+(v)$	outdegree of v
$\Delta^-(G)$	$\max_{v\in V} d^-(v)$
$\Delta^+(G)$	$\max_{v\in V} d^+(v)$
$d_k^-(v)$	$\lvert E_k^-(v)\rvert$, number of paths of length k ending at v
$d_k^+(v)$	$\lvert E_k^+(v)\rvert$, number of paths of length k starting at v
$\Delta_k^-(G)$	$\max_{v\in V} d_k^-(v)$
$\Delta_k^+(G)$	$\max_{v\in V} d_k^+(v)$

Note that $\Delta_1^-(G) = \Delta^-(G)$, $\Delta_1^+(G) = \Delta^+(G)$ and for any k, $\Delta_k^+(G^\top) = \Delta_k^-(G)$.

Labelled graphs:

$(u, v, \sigma) \in E$	An edge from u to v with label σ
$(\sigma_1, \ldots, \sigma_k)$ is G-admissible	$\sigma_1, \ldots, \sigma_k$ are the respective labels of a path of length k
G_k	subset of $[m]^k$ contaning all G-admissible k-uples
$A_{(u,v,\sigma)}$	A_σ
A_s where $s \in E^k$	$A_{s[k]} \cdots A_{s[1]}$

Convex analysis:

$\mathrm{aff}(\mathcal{S})$	affine hull of $\mathcal{S}$
$\mathrm{conv}(\mathcal{S})$	convex hull of $\mathcal{S}$
$\mathrm{ray}(\mathcal{S})$	smallest cone containing $\mathcal{S}$
$\mathrm{cone}(\mathcal{S})$	convex cone generated by $\mathcal{S}$
$\mathcal{C}^\circ$	polar of a convex set $\mathcal{C}$
$\mathcal{K}^*$	dual of a convex cone $\mathcal{K}$

Optimization:

semidefinite program	SDP

Algebra:

$\alpha \in \mathbb{N}^n$ (for exponents of monomials)	$\lvert\alpha\rvert = \sum \alpha_i$
sum of squares in n variables, degree at most $2d$	$\Sigma_{n,2d}$
if n is clear	Σ_{2d}
sos forms in n variables, degree at most $2d$	$\Sigma_{\mathbf{n,2d}}$
if n is clear	$\Sigma_{\mathbf{2d}}$

List of Figures

1.1 Illustration for Example 1. 5
1.2 Automaton for the running example. The numbers on the edges are their respective labels. 9
1.3 Representation of the solutions to Program 1 with different values of d for the running example. The blue curve represents the boundary of the 1-sublevel set of the optimal solution p_v at each node $v \in V$. The dashed curve is the boundary of the unit circle. 12
1.4 Illustration for the Veronese variety. 17
1.5 The value of $H_b(\alpha)$ for $0 \leq \alpha \leq 1$. 21
1.6 The decomposition of a entropy vector of 2 variables in its atoms. 25
1.7 Lattice representation of entropy vector of $\mathcal{E}_3$. 26
1.8 Extreme rays of $\mathcal{H}_4$ not expressible using the $r^m_{\mathcal{S}}$ notation. 28
1.9 3D visualization of approximations of the Entropic Cone of 4 random variables. . 32
1.10 Outer [DFZ11] and inner [MC13] approximations of $\mathcal{P}_4$. The figures comes from [MC13]. 33

2.1 Simple example of the low rank reduction. 35
2.2 Comparison between Algorithm 1 with $l = 1$ and $l = 2$ 51

3.1 Illustration of the invariant condition for being a controllable set. The matrix A denotes a rotation of $-60°$. We see in this example that the drawn ellispoid is not a controllable set since $(Ax)|_V \notin \mathcal{S}|_V$ for some x. 57
3.2 The polymatroid cone $\mathcal{P}_4$ cut by the feasibility cuts uncovered to find the lower bound -0.15789 of the Ingleton score. 62
3.3 Overview of the Julia packages developped and used. In yellow are low level libraries and solvers, in green are abstraction interface, in blue are modelling interfaces, in red are general purpose packages and in brown are toolboxes. See `http://www.juliaopt.org/` for a comprehensive list of supported solvers. 63

List of Tables

2.1 Quantification of the low rank reduction of Theorem 14. The quantity of the CJSR $(G', \mathcal{M}')$ obtained by the reduction are expressed as a function of the quantity of the original CJSR $(G, \mathcal{M})$. 37

Introduction

Since the 90s, mathematical optimization has played a prominent role in Systems and Control Theory. Through the development of fundamental advances in applied mathematics (e.g., Linear Matrix Inequalities – LMIs, sum of squares, the S-procedure) researchers have designed Control techniques whose theoretical and practical efficiency have revolutionized our ability to cope with complex control systems. In particular, *switched systems* (where the laws of dynamic behaviour change at certain points in time) have been a great source of challenges for control theorists. This simple model encapsulates the complexity of many different applications, ranging from the optimal design of drug treatment for viral diseases like HIV [HVCMB11], to the analysis of consensus systems of autonomous agents [JL+03]. Even though these systems are theoretically extremely hard to analyze, it turns out that *advanced mathematical techniques* allow going beyond these barriers, providing promising methods to handle switched systems. The marriage of deep mathematical results about switched systems, and the powerful machinery of modern optimization techniques has led to innovative results and algorithms, and mathematicians and engineers have today a whole technology to assist them in the analysis and optimal design of switched systems.

A simple class of problems for which efficient methods have been designed is the computation of a (maximal) invariant set for switched systems. This problem is of primary importance in many applications in control, and is known to be very hard (e.g., deciding whether such a nontrivial set exists is Turing-undecidable, and approximating it is NP-hard, see [Jun09]). However, techniques relying on Sum-of-Squares relaxations and other optimization tools have proved useful both for theoretical and practical purposes.

Remarkably, it turns out that several other great challenges in computational engineering, not related a priori with control theory, rely fundamentally on the behaviour of *matrix semigroups*, which are the core mathematical structure underlying (discrete-time) switched systems. In his seminal paper, Claude Shannon studied the entropy of a (single) random source, and showed how this notion is a central tool for optimizing the transmission of the information contained in this source through a noisy channel. A more subtle question arises when there is no longer a single source of information, but several ones, whose probability functions may in general be correlated. The question is: "are there other constraints describing the correlation between several sources?" Are there *Non-Shannon Inequalities*? The answer is yes: in a breakthrough paper in 98, Zhang and Yeung proved that indeed there exist Non-Shannon Inequalities [ZY98]. Since then, our understanding of the problem has merely but slightly improved. While several other Non-Shannon inequalities have been found, no global description of the structure of the *entropic cone* (i.e., the set of values that the entropy can take) is yet available. A solution to this problem could have a dramatic impact on applications, like for instance *Network Coding*. A thorough understanding of the structure of information of multivariate random sources would be of critical help to manage routing the packets of information by maximizing the flow of entropy, constrained by the newly found Information inequalities.

It turns out (even if it is by no means obvious) that all the constraints that have been found on

the entropic cone (the so-called *non-Shannon inequalities*) can actually be expressed as invariance constraints under a family of linear operators. This observation unravels a fundamental link between the dynamics of (maximal) invariant set for switched linear systems. However, there is a technical difference in that in this application, to put it crudely, these constraints are expressed in projected spaces, and not in the full-dimensional space.

In recent years, path-complete graph techniques [AJPR14] provided a new unifying approach to analyse the behaviour of a switched system. However, it requires the generalizations of control techniques for more general switched systems. Moreover, while we have more and more tools to assess the stability of switched systems, no tool allows to efficiently uncover the unstable trajectories of switched systems in case of unstability. The first objective of this master's thesis is to solve these two problems. The first contribution resulted from a work So far, all inequalities in information theory have been obtained through Linear Programming techniques. However, Matúš showed that the entropic cone is not a polyhedron [Mat07a] and Mejia and Montoya showed that it is not even semialgebraic [MM15]. Therefore, any good approximation of the entropic cone would need to have a huge amount of facets, since this cone is high-dimensional and not semialgebraic. since it is high-dimensional and not semialgebraic. In practical applications, one is often looking at the simpler situation of solving an optimization problem involving the entropic cone so we are only looking for an outer approximation that is "close" to the entropic cone "near" the optimum. Moreover, modern optimization tools (LMIs, extended formulations, algebraic liftings, ...) allow to directly handle non-linear techniques. The second objective, based on these observations, is to design efficient algorithmic tools for the optimization on, and approximation of the entropic cone.

This master's thesis contains both theoretical results and implementations with comparison of the computational results of different approaches. It was was started during my research stay at MIT with Prof. Pablo Parrilo in the summer of 2015. The work of my master's thesis resulted in three main contributions in the field of Control and Information Theory.

The first contribution is on the approximation of the joint spectral radius (JSR) of switched systems [LJP16]. This quantity characterizes the maximal asymptotic growth rate of a state of the system. We showed how to exploit the potential low rank structure of a switched system to speed up the approximation of the JSR.

As mentioned above, many algorithms exist for estimating the JSR but not much is known about how find the unstable trajectories. The currently known algorithms select a small sequence with large spectral radius using brute force (or branch-and-bound variants) and repeat this sequence infinitely. We introduce a new approach to this question, using the dual solution of a sum of squares optimization program for JSR approximation. Our algorithm produces an infinite sequence of matrices with an asymptotic growth rate arbitrarily close to the JSR. Unlike the brute force approach, we provide a guarantee on the closeness of the asymptotic growth rate to the JSR. This, in turn, provides new bounds on the quality of the JSR approximation.

The second and third contributions are in the intersection of Information Theory and Control Theory. We introduce the notion of *controllable sets* of controlled switched linear systems and show that the entropic cone is a controllable set for some controlled switched linear system that we define.

Our second contribution is a parallelizable algorithm for the optimization on controllable sets using techniques from Stochastic Programming [LJ16]. We developed a new toolbox `EntropicCone.jl` in Julia [BKSE12] for working with the entropic cone. This algorithm is one of the features implemented in the toolbox. We tested our algorithm to find lower bounds for the Ingleton score. This quantity is usually used to benchmark approximation techniques on the entropy cone. Its exact value is currently unknown, it is only known that it is between

-0.15789 [DFZ11] and -0.09243 [MC13]. Our algorithm obtained the lower bound -0.15789 in under a minute.

Our third contribution is a nonlinear approximation algorithm for controllable sets using sum of squares programming. This algorithm seems to be a promising candidate to overcome the weakness of linear techniques to approximate the high dimensional nonlinear entropic cone. This is ongoing work as we are currently implementing the algorithm in the toolbox `EntropicCone.jl`.

The text is organized in three chapters. The first chapter covers the state of the art on polyhedral computation, sum of squares programming, switched systems and the entropic cone. The second chapter describes the first contribution in detail and in the third chapter, we introduce the notion of controllable set and show how to optimize on and approximate the entropic cone, the second and third contributions.

Contributions

[LJ16] B. Legat and R. M. Jungers. Parallel optimization on the entropic cone. In *Proceedings of the 37rd Symposium on Information Theory in the Benelux*, SITB '16, pages 206–2011, 2016.

[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.

Chapter 1

State of the art

This chapter contains the background material needed for our main contributions presented in Chapter 2 and Chapter 3. We give below a brief description of the topic of each section along with its typical reference.

This master's thesis heavily relies on approximations of convex sets by polyhedral and sum of squares techniques. In Section 1.1, we give broad notions on convexity and polyhedral computations which are respectively typically found in [Roc15] and [Zie95]. In Section 1.2, we give an introduction to sum of squares programming which can typically be found in [BPT12].

In Section 1.3, the notion of switched systems is presented with an emphasis on the *joint spectral radius* (JSR), the typical reference on the matter is [Jun09]. The JSR is the maximal asymptotic growth rate of the state of the system. The notion of invariant set is introduced in Section 1.4. The approximation of quadratic invariant sets is described in [BEGFB94] but we give its generalization for approximation of higher degree.

When the switched system is stable, the invariant sets have the same homothety factor and this factor is equal to the JSR.

The notion of invariant set is generalized and then used to approximate the entropic cone in Chapter 3. In Section 1.5, we introduce the entropic cone and give its basic properties. Then, in Section 1.6, we give the current state of the art in the approximation of the entropic cone. A survey on the entropic cone and its applications can be found in [Yeu08] and a more recent in [Cha11].

1.1 Polyhedral Computation

1.1.1 Convex Analysis

Consider r points $x_1, \ldots, x_r \in \mathbb{R}^n$. A vector $\sum_{i=1}^r \lambda_i x_i$ is

- an *affine combination* if $\sum_{i=1}^r \lambda_i = 1$.
- a *conic combination* if $\lambda_i \geq 0$, $i = 1, \ldots, r$.
- a *convex combination* if it is affine and conic.

The set of λ that are the coefficient of a convex combination is the $r-1$-*dimensional standard simplex* denoted by Δ^{r-1}.

An *affine set* is a set closed to affine combination, a *cone* is a set closed to positive scalar multiplication, a *convex cone* is a set closed to conic combination and a *convex set* is a set closed to convex combination. An arbitrary intersection of affine sets (resp. cone, convex cone, convex set) is an affine set (resp. cone, convex cone, convex set).

The *affine hull* is of a set $\mathcal{S}$ is the smallest affine set containing $\mathcal{S}$,

$$\text{aff}(\mathcal{S}) = \{ \lambda_1 x_1 + \cdots + \lambda_r x_r \mid x_1, \ldots, x_r \in \mathcal{S}, \lambda_1 + \cdots + \lambda_r = 1 \}.$$

The *convex hull* is of a set $\mathcal{S}$ is the smallest convex set containing $\mathcal{S}$,

$$\text{conv}(\mathcal{S}) = \{ \lambda_1 x_1 + \cdots + \lambda_r x_r \mid x_1, \ldots, x_r \in \mathcal{S}, \lambda \in \Delta^{r-1} \}.$$

The smallest cone containing the origin and $\mathcal{S}$ is

$$\text{ray}(\mathcal{S}) = \{ \lambda x \mid x \in \mathcal{S}, \lambda \in \mathbb{R}, \lambda \geq 0 \}$$

The *convex cone generated* by $\mathcal{S}$ is the smallest convex cone containing the origin and $\mathcal{S}$,

$$\text{cone}(\mathcal{S}) = \text{conv}(\text{ray}(\mathcal{S})).$$

Subspaces are affine sets which contain the origin. Each non-empty affine set is the translate of a unique subspace. Its dimension is defined as the dimension of this subspace. The dimension of the empty set $\emptyset$ is -1 by convention.

A set of $m+1$ points $x_0, \ldots, x_m$ is said to be *affinely independent* if $\text{aff}\{x_0, \ldots, x_m\}$ is m-dimensional. In such case, for each point x in the affine hull, the parameter λ of the affine combination is unique and is called the *barycentric coordinate*.

Affine sets of dimension 0, 1, 2, ... are called points, lines, planes, respectively. The $(n-1)$-dimensional affine sets are called hyperplanes. For any hyperplanes, there exists y, α such that the hyperplane is equal to $\{ x \in \mathbb{R}^n \mid \langle x, a \rangle = \alpha \}$. Any affine subsets of $\mathbb{R}^n$ is the intersection of finitely many hyperplanes.

The set $\{ x \in \mathbb{R}^n \mid \langle x, a \rangle \leq \alpha \}$ is called an *halfspace* and is convex.

If $x_0, \ldots, x_m$ are affinely independent, $\text{conv}\{x_0, \ldots, x_m\}$ is an *m-dimensional simplex*. The standard $r-1$-dimensional simplex define above is $\text{conv}\{e_1, \ldots, e_r\}$ where the vectors e_i are the vector of the canonical basis. When $m = 0, 1, 2$, or 3, the simplex is called a point, (closed) line segment, triangle or tetrahedron, respectively.

The dimension of a convex set is defined as the dimension of its affine hull. It is equal to the maximum of the dimension of the simplices contained in it.

A set of $d+2$ or more points is always affinely dependent.

Theorem 1 (Caratheodory)**.** Let R be a collection of points of $\mathbb{R}^n$. If $x \in \text{conv}(R)$, then x is the convex combination of $n+1$ points of R.

An inequality $\langle x, a\rangle \leq \alpha$ is said to be valid on a convex set $\mathcal{C}$ if it holds for any $x \in \mathcal{C}$. A set $\mathcal{F} \subseteq \mathcal{C}$ of a convex set $\mathcal{C}$ is a *face* if and only if there exists a valid linear inequality $\langle x, a\rangle \leq \alpha$ such that $\langle x, a\rangle = \alpha$ for any $x \in \mathcal{F}$. Such a hyperplane is called a *supporting hyperplane* of $\mathcal{F}$. A face of dimension 0, 1 or $\dim(\mathcal{C}) - 1$ is called a vertex, edge or facet respectively. The face of dimension $\dim(\mathcal{C})$ is $\mathcal{C}$ itself and the face of dimension -1 is the emptyset. Faces of dimension between 0 and $\dim(\mathcal{C}) - 1$ are called *proper faces*. The faces of a convex set form a finite partially ordered set (poset) by containment.

The barrier cone of a convex set $\mathcal{C}$ is the set of of y such that for some $\beta \in \mathbb{R}$, $\langle x, y\rangle \leq \beta$ for all $x \in \mathcal{C}$. The polar of a set $\mathcal{S}$ is

$$\mathcal{S}^\circ = \{\, y \in \mathbb{R}^n \mid \langle x, y\rangle \leq 1, \forall x \in \mathcal{S} \,\}.$$

The polar transformation satisfies the following property: For any convex sets $\mathcal{S}, \mathcal{T}$,

$$\mathcal{S} \subseteq \mathcal{T} \Leftrightarrow \mathcal{T}^\circ \subseteq \mathcal{S}^\circ. \tag{1.1}$$

The following three lemma shows how to translates different operations between one convex set and its polar.

Lemma 1 ([Roc15, Corollary 16.3.2])**.** For any convex set $\mathcal{C}$,

$$(A\mathcal{C})^\circ = A^{-\top}(\mathcal{C}^\circ)$$
$$(A^{-1}(\text{cl}\,\mathcal{C}))^\circ = \text{cl}(A^\top(\mathcal{C}^\circ)).$$

Lemma 1 will be quite important for this master's thesis. For this reason, we give an example to give its geometric intuition.

Example 1. We can rewrite $(A\mathcal{C})^\circ = A^{-\top}(\mathcal{C}^\circ)$ as follows:

$$x \in (A\mathcal{C})^\circ \iff A^\top x \in \mathcal{C}^\circ. \tag{1.2}$$

Let $\mathcal{C}$ be a sphere of radius 1 centered at $(0, 0, 2)$ and A be the linear projection on the plane $z = 0$. With these choices of $\mathcal{C}$ and A, the set $A\mathcal{C}$ is the circle of radius 1 of the plane $z = 0$ centered at the origin. This is illustrated by the Figure 1.1. Let x be the normal vector of the line of the plane $z = 0$ shown in Figure 1.1. By (1.2), x is in the polar of the circle, if and only if $A^\top x$, which is the plane perpendicular to the plane $z = 0$ containing the line, is in the polar of the sphere.

Lemma 2 ([Roc15, Corollary 16.4.2])**.** Let $\mathcal{K}_i$ be a non-empty convex cone in $\mathbb{R}^n$ for each $i \in I$. Then

$$(\sum_{i\in\mathcal{K}} \mathcal{K}_i)^\circ = \bigcap_{i\in I} \mathcal{K}_i^\circ$$
$$(\bigcap_{i\in I} \text{cl}\,\mathcal{K}_i)^\circ = \text{cl}(\sum_{i\in I} \mathcal{K}_i^\circ).$$

Lemma 3 ([Roc15, Corollary 16.5.2])**.** Let $\mathcal{C}_i$ be a convex set in $\mathbb{R}^n$ for each $i \in I$. Then

$$(\text{conv}_{i\in I}\, \mathcal{C}_i)^\circ = \bigcap_{i\in I} \mathcal{C}_i^\circ$$
$$(\bigcap_{i\in I} \text{cl}\,\mathcal{C}_i)^\circ = \text{cl}(\text{conv}_{i\in I}\, \mathcal{C}_i^\circ).$$

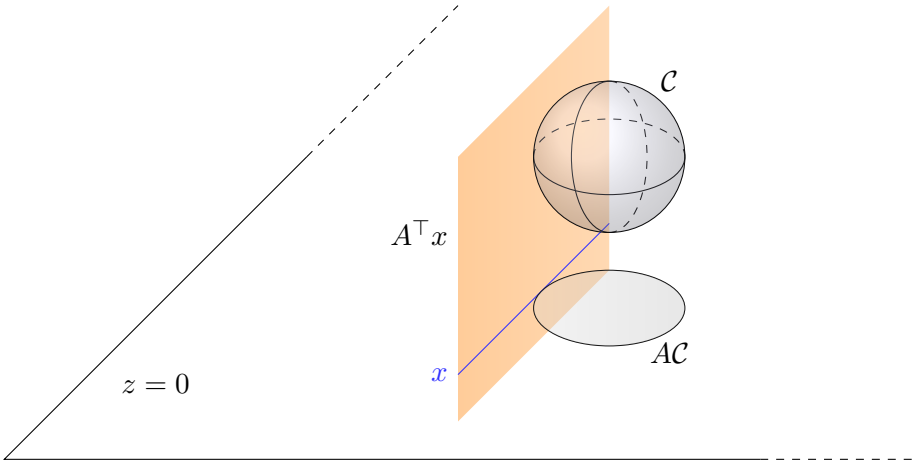


Figure 1.1: Illustration for Example 1.

1.1.2 Polyhedron and Polytope

The intersection of finitely many halfspaces is called a *polyhedral convex set* or a *polyhedron*. The convex hull of finitely many points is called a *polytope*.

Theorem 2 (Weyl-Minkowski)**.** Every polytope is a polyhedron. Every bounded polyhedron is a polytope.

H-representationand V-representation

Theorem 3 (Weyl-Minkowski (continued))**.** For a subset $\mathcal{P}$ of $\mathbb{R}^n$, the following statements are equivalent

- $\mathcal{P}$ has an *H-representation*, that is, there is a matrix $A \in \mathbb{R}^{m\times n}$ and a vector $b \in \mathbb{R}^m$ such that
$$\mathcal{P} = \{\, x \in \mathbb{R}^n \mid Ax \geq b \,\}. \tag{1.3}$$
- $\mathcal{P}$ has a *V-representation*, that is, there is finitely many vectors $v_1, \ldots, v_s \in \mathbb{R}^n$, $r_1, \ldots, r_t \in \mathbb{R}^n$ such that
$$\mathcal{P} = \operatorname{conv}\{v_1, \ldots, v_s\} + \operatorname{cone}\{r_1, \ldots, r_t\}.$$
Or equivalently, there exists matrices $V \in \mathbb{R}^{n\times s}, R \in \mathbb{R}^{n\times t}$ such that
$$\mathcal{P} = \{\, x \in \mathbb{R}^n \mid \exists \lambda \in \Delta^{s-1}, \mu \geq 0, x = V\lambda + R\mu \,\}. \tag{1.4}$$

For a polyhedral cone, the representations can be simplified:

- the H-representation is
$$\mathcal{P} = \{\, x \in \mathbb{R}^n \mid Ax \geq 0 \,\}. \tag{1.5}$$
- and the *R-representation* is
$$\mathcal{P} = \{\, x \in \mathbb{R}^n \mid \exists \lambda \geq 0, x = R\lambda \,\}. \tag{1.6}$$

Every convex set $\mathcal{C} \in \mathbb{R}^n$ can be regarded as the cross-section of a convex cone $\mathcal{K} \in \mathbb{R}^{n+1}$. Indeed, let $\mathcal{K} = \operatorname{cone}\{\, (1, x) \mid x \in \mathcal{C} \,\}$, the set $\mathcal{C}$ is the intersection of $\mathcal{K}$ with the hyperplane $\{\, (\lambda, x) \mid \lambda = 1 \,\}$.

Let A^c, R^c be the matrices A and R of the H-representation and R-representation of $\mathcal{K}$. We have

$$\begin{aligned} A^c &= \begin{bmatrix} -b & A \end{bmatrix} \\ R^c &= \begin{bmatrix} \mathbf{1}^\top & 0 \\ V & R \end{bmatrix} \\ b &= -A^c_{:,1} \\ A &= A^c_{:,2:}. \end{aligned}$$

To extract V and R from R^c we need to scale the columns of R_c by nonnegative scalars so that the first row contains only zeros and ones (we drop the columns containing negative numbers in the first row). Then the columns with one in the first row are put in V and the columns with zeros in the first row are put in R.

The pair (A, R) is called a Double Description (DD)-pair. The DD-pair of the dual cone is $(R^\top, A^\top)$.

$$\mathcal{P}^* = \{\, x \in \mathbb{R}^n \mid R^\top x \geq 0 \,\} \tag{1.7}$$

$$\mathcal{P}^* = \{\, x \in \mathbb{R}^n \mid \exists \lambda \geq 0, x = A^\top \lambda \,\}. \tag{1.8}$$

This shows that going from A to R, which is called the *vertex enumeration* problem, or going from R to A, which is called the *convex hull* problem, are essential the same problems: the *representation conversion* problem.

One way to compute A from R is to eliminate λ from the linear system of (1.6). To compute R from A we can do the same to (1.8). Eliminating variables from a polyhedron is called *polyhedral projection*. That is, we have just provided a polynomial reduction of representation conversion to polyhedral projection of H-representation.

Polyhedral projection of a polyhedron in its V-representation is very easy, it boils down to dropping the eliminated coordinates of vertices and rays. We can see that polyhedral projection of an H-representation can be achieved by two representation conversion and the polyhedral projection of V-representation. This provides a polynomial reduction of H-representation of polyhedral projection to representation conversion.

This shows that representation conversion is essentially "as hard" as polyhedral projection.

A remark on Computational Complexity for large output size

The complexity of algorithms is commonly described using the only input size (in bits) and not the output size. This is not appropriate for representation conversion and polyhedral projection algorithms that need to list the facets (resp. vertices, rays) of a polyhedron. Indeed, the output size itself might not be polynomial in terms of the input size. We say that an algorithm is *output sensitive* if it is in P-TIME in both the input and output size. It was show that none of the known algorithms for the reprensentation conversion are output sensitive [ABS95] but it is still unknown whether there exist an output sensitive algorithm for this problem.

An algorithm with a large output might produce its output incrementally without needing to store its output to continue. Therefore it might run in P-SPACE in its input size even if its output size is exponential in its input size. Such algorithm is said to be *compact* [Fuk96]. This is the case of the reversed search vertex enumeration algorithm implemented in LRS [Avi00].

1.2 Sum of Square Programming

Deciding whether a multivariate polynomial of degree $2d \geq 4$ is nonnegative is known to be NP-hard. However a sufficient condition for a polynomial to be nonnegative is easy to check. We say that a polynomial is a *sum of squares* (SOS) if there exists polynomials $q_1, \ldots, q_M$ such that

$$p(x) = \sum_{k=1}^{M} q_k^2(x).$$

If a polynomial is SOS, then it is obviously nonnegative.

It is well known that if $p(x)$ is an homogeneous polynomial of degree $2d$ then $q_k(x)$ must be an homogeneous polynomial of degree d; this can be shown easily using the Newton polytope of $p(x)$ and [Rez78, Theorem 1]. Let $x^{[d]}$ represent a basis of the homogeneous polynomials of degree d. We can check whether a polynomial is SOS using semidefinite programming thanks to the following theorem.

Theorem 4 ([CLR95, Nes00, Par00, Par03, Sho87])**.** A homogeneous multivariate polynomial $p(x)$ of degree $2d$ is a sum of squares if and only if

$$p(x) = (x^{[d]})^\top Q x^{[d]}$$

where Q is a symmetric positive semidefinite matrix.

Fromt the exact arithmetic viewpoint, the basis $x^{[d]}$ chosen does not affect whether $p(x)$ is SOS or not. A specific choice of basis may however improve the numerical behaviour of the corresponding semidefinite program. Moreover considering a basis satisfying some properties may facilitate the theoretical development. We would like to choose a basis $x^{[d]}$ such that $\|x^{[d]}\|_2 = \|x\|_2^d$ where $\|\cdot\|_2$ is the Euclidean norm.

This leads us to the choice of the *scaled monomial* basis. The elements of this basis are

$$\frac{d!}{\alpha_1!\alpha_2!\cdots\alpha_n!} x_1^{\alpha_1}\cdots x_n^{\alpha_n}.$$

for each n-uples of nonnegative integers α such that $\alpha_1 + \cdots + \alpha_n = d$.

For any matrix $A \in \mathbb{R}^{n\times n}$, the d-lift induces an associated map $A^{[d]} \in \mathbb{R}^{N_d \times N_d}$ which is the unique matrix that satisfies

$$(Ax)^{[d]} = A^{[d]} x^{[d]}. \tag{1.9}$$

We also denote $\mathcal{M}^{[d]} \triangleq \{A_1^{[d]}, \ldots, A_m^{[d]}\}$.

We also have the following property for the standard inner product between vectors

$$\langle x^{[d]}, y^{[d]} \rangle = (\langle x, y \rangle)^d. \tag{1.10}$$

We denote as the set of homogeneous polynomials of degree $2d$ as $\mathbb{R}_{\mathbf{2d}}[x]$, the cone of homogeneous SOS polynomials of degree $2d$ as $\Sigma_{\mathbf{2d}}$ and the dual of $\Sigma_{\mathbf{2d}}$ as $\Sigma_{\mathbf{2d}}^*$.

1.2.1 Sum of Squares Convexity

In Section 1.4 and Section 3.2, we will need to constrain a polynomial to be convex. This can be achieved by constraining its Hessian to be pointwise positive semidefinite. The Hessian is a polynomial matrices, that is, a matrix with polynomial entries. Thus the problem of checking convexity of a polynomial reduces to the problem of checking the pointwise positive

semidefiniteness of a polynomial matrix. Since polynomial nonnegativity is a special case of polynomial matrix positive semidefiniteness, the latter is also NP-hard. Hence this reduction does not provide an efficient algorithm for polynomial convexity.

It turns out that checking the convexity of a polynomial is NP-hard even for quartic polynomials [AOPT13]. Thus checking whether the Hessian is positive semidefinite may actually be the best we can do. We can define the generalisation of sos polynomial for polynomial matrices as follows.

Definition 1 (sos matrices and sos convexity)**.** A symmetric polynomial matrix $P(x) \in \mathbb{R}[x]^{m\times m}$, $x \in \mathbb{R}^n$, is an *sos matrix* if there exists a polynomial matrix $M(x) \in \mathbb{R}[x]^{s\times m}$ for some $s \in \mathbb{N}$, such that $P(x) = M^\top(x)M(x)$.

A polynomial $p(x)$ is *sos-convex* if its Hessian is an sos matrix.

Polynomial matrix semidefiniteness is actually equivalent to polynomial nonnegativity.

Lemma 4. Let $P(x) \in \mathbb{R}[s]^{m\times m}$ be a symmetric polynomial matrix, with $x \in \mathbb{R}^n$. Let $p(x, y) = y^T P(x)y$ be the associated scalar polynomial in $m + n$ variables where $y = [y_1, \ldots, y_m]$.

1. The matrix $P(x)$ is positive semidefinite if and only if $p(x, y)$ is nonnegative.
2. The matrix $P(x)$ is an sos matrix if and only if $p(x, y)$ is a sum of squares (in $\mathbb{R}[x, y]$).

More information on sum of squares matrices can be found in [BPT12, Section 3.3.2] and see [BPT12, Section 3.3.3] or [AP12a, AP13] for sos convexity. Sos convexity also plays a prominent role in certifying the stability of nonlinear switched systems [AJ13b, AJ$^+$13a].

1.3 Switching Systems

A switched linear system is characterized by a finite set of matrices $\mathcal{M} \triangleq \{A_1, A_2, \ldots, A_m\} \subset \mathbb{R}^{n\times n}$ and the iteration

$$x_k = A_{\sigma_k} x_{k-1}, \quad \sigma_k \in [m]. \tag{1.11}$$

The maximal asymptotic growth rate of this iteration is given by the *joint spectral radius* (JSR). The JSR $\rho(\mathcal{M})$ of a finite set of matrices $\mathcal{M}$ is defined as

$$\rho(\mathcal{M}) = \lim_{k\to\infty} \max_{\sigma\in[m]^k} \|A_{\sigma_k} \cdots A_{\sigma_2} A_{\sigma_1}\|^{1/k}.$$

This definition is independent of the norm used.

In some applications the values that σ_k can take may depend on $\sigma_{k-1}, \sigma_{k-2}, \ldots$. These constraints are often conveniently represented using a *finite automaton* and the JSR under such constraints is called constrained joint spectral radius (CJSR) [Dai12]; an example of constrained switched system is given by Example 2 and its automaton is illustrated by Figure 1.2.

The following will serve as a running example.

Example 2 (Running example[1])**.** We borrow the example of [PEDJ15, Section 4]. The set of matrices $\mathcal{M}$ is composed of the following four matrices

$$A_1 = A + B\begin{pmatrix} k_1 & k_2 \end{pmatrix}, \qquad A_2 = A + B\begin{pmatrix} 0 & k_2 \end{pmatrix},$$
$$A_3 = A + B\begin{pmatrix} k_1 & 0 \end{pmatrix}, \qquad A_4 = A.$$

where $k_1 = -0.49$, $k_2 = 0.27$,

$$A = \begin{pmatrix} 0.94 & 0.56 \\ 0.14 & 0.46 \end{pmatrix} \text{ and } B = \begin{pmatrix} 0 \\ 1 \end{pmatrix}.$$

The automaton is represented by Figure 1.2.

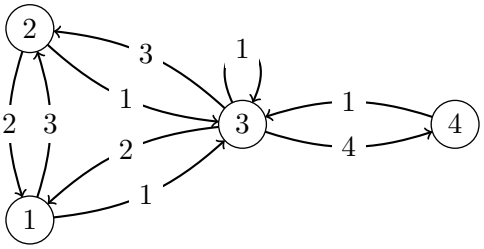


Figure 1.2: Automaton for the running example. The numbers on the edges are their respective labels.

The automaton representing the constraints can be represented by a strongly connected labelled directed graph $G(V, E)$, possibly with parallel edges. The labels are elements of the set $[m]$ and E is a subset of $V \times V \times [m]$. We say that $(u, v, \sigma) \in E$ if there is an edge between node u and node v with label σ.

$$x_k = A_{\sigma_k} x_{k-1}, \quad (\sigma_1, \ldots, \sigma_k) \in G_k. \tag{1.12}$$

[1]The source code and instructions are available at the author's web site to reproduce the numerical results obtained for the running example. The switched system considered in Example 3, Example 10 and Example 12 is simpler than the running example and the results given in those examples can be obtained by hand.

The *arbitrary switching* case (1.11) can be seen as the particular case when the automaton has only one node and m self-loops with labels $1, \ldots, m$.

The definition of the JSR is generalized as follows for constrained systems.
Definition 2 ([Dai12])**.** The *constrained joint spectral radius* (CJSR) of a finite set of matrices $\mathcal{M}$ constrained by an automaton G, denoted as $\rho(G, \mathcal{M})$, is

$$\rho(G, \mathcal{M}) = \lim_{k \to \infty} \hat{\rho}_k(G, \mathcal{M})$$

where

$$\hat{\rho}_k(G, \mathcal{M}) = \max \left\{ \|A_s\|^{1/k} : s \in G_k \right\}.$$

1.3.1 CJSR Approximation via SOS

A key result used in the SOS-based approximation algorithm for the unconstrained JSR [PJ08] is the following:
Proposition 1 ([RS60, Proposition 1])**.** Consider a finite set of matrices $\mathcal{M}$. For any $\epsilon > 0$ there exists a norm $\| \cdot \|$ such that

$$\|A_\sigma x\| \leq (\rho(\mathcal{M}) + \epsilon)\|x\|$$

for all $x \in \mathbb{R}^n$ and $\sigma = 1, \ldots, m$.

Philippe et al. recently gave a generalization of that algorithm, relying on the following extension of Proposition 1 to constrained switched systems.
Proposition 2 ([PEDJ15, Proposition 2.2])**.** Consider a finite set of matrices $\mathcal{M}$ constrained by an automaton $G(V, E)$. For any $\epsilon > 0$, there exists a set of norms $\{ \| \cdot \|_v : v \in V \}$ such that

$$\|A_\sigma x\|_v \leq (\rho(G, \mathcal{M}) + \epsilon)\|x\|_u$$

holds for all edge $(u, v, \sigma) \in E$.

Strictly positive homogeneous polynomials are not necessarily norms because their sublevel sets may not be convex. However we can provide upper bounds for the CJSR using polynomials instead of norms.
Theorem 5. Consider a finite set of matrices $\mathcal{M}$ constrained by an automaton $G(V, E)$. If there exists a strictly positive homogeneous polynomial $p_v(x)$ of degree $2d$ for all $v \in V$ such that

$$p_v(A_\sigma x) \leq \gamma^{2d} p_u(x)$$

holds for all edge $(u, v, \sigma) \in E$. Then $\rho(G, \mathcal{M}) \leq \gamma$.

Proof. Consider a norm $\| \cdot \|$ of $\mathbb{R}^n$ and its corresponding induced matrix norm of $\mathbb{R}^{n \times n}$. For each $v \in V$, we know by compactness of the unit ball in $\mathbb{R}^n$, continuity and strict positivity of $p_v(x)$ that there exist $0 < \alpha_v \leq \beta_v$ such that

$$\alpha_v \|x\|^{2d} \leq p_v(x) \leq \beta_v \|x\|^{2d}$$

for all $x \in \mathbb{R}^n$. Let $\alpha = \min_{v \in V} \alpha_v$ and $\beta = \max_{v \in V} \beta_v$.

For a G-admissible k-uple $(\sigma_1, \sigma_2, \ldots, \sigma_k)$,

$$\|A_{\sigma_k} \cdots A_{\sigma_1}\| = \sup_{x \neq 0} \frac{\|A_{\sigma_k} \cdots A_{\sigma_1} x\|}{\|x\|}.$$

Consider a path such that the ith edge has label σ_i for $i = 1, \dots, k$ and denote the intermediary nodes of that path as $v_0, v_1, \dots, v_k$. For any $x \in \mathbb{R}^n$, we have

$$\begin{aligned}\|A_{\sigma_k} \cdots A_{\sigma_1} x\| &\leq \alpha_{v_k}^{-\frac{1}{2d}} p_{v_k}(A_{\sigma_k} \cdots A_{\sigma_1} x)^{\frac{1}{2d}} \\ &\leq \alpha_{v_k}^{-\frac{1}{2d}} \gamma p_{v_{k-1}}(A_{\sigma_{k-1}} \cdots A_{\sigma_1} x)^{\frac{1}{2d}} \\ &\leq \alpha_{v_k}^{-\frac{1}{2d}} \gamma^k p_{v_0}(x)^{\frac{1}{2d}}\end{aligned}$$

and

$$\|x\| \geq \beta_{v_0}^{-\frac{1}{2d}} p_{v_0}(x)^{\frac{1}{2d}}$$

hence

$$\|A_{\sigma_k} \cdots A_{\sigma_1}\| \leq \Big(\frac{\beta_{v_k}}{\alpha_{v_0}}\Big)^{\frac{1}{2d}} \gamma^k \leq \Big(\frac{\beta}{\alpha}\Big)^{\frac{1}{2d}} \gamma^k.$$

Taking the kth root, the limit $k \to \infty$ and using Definition 2 we obtain the result. □

We relax the positivity condition of Theorem 5 by the more tractable sum of squares (SOS) condition and define $\rho_{\mathrm{SOS},2d}(G, \mathcal{M})$ as the solution of the following SOS program:
Program 1 (Primal)**.**

$$\begin{aligned}\inf_{p_v(x) \in \mathbb{R}_{\mathbf{2d}}[x], \gamma \in \mathbb{R}} \ & \gamma \\ & p_v(x) \text{ is SOS} \quad \forall v \in V && (1.13) \\ & p_v(x) \text{ is strictly positive} \quad \forall v \in V && (1.14) \\ & \gamma^{2d} p_u(x) - p_v(A_\sigma x) \text{ is SOS} \quad \forall (u, v, \sigma) \in E.\end{aligned}$$

Remark 1. In practice we can replace (1.13) and (1.14) by "$p_v(x) - \epsilon\|x\|_2^{2d}$ is SOS" for any $\epsilon > 0$. Note that this constrains $p_v(x)$ to be in the interior of the SOS cone, which is sufficient for $p_v(x)$ to be strictly positive. The bounds given in Section 2.3 are valid if $p_v(x)$ is in the interior of the SOS cone.

By Theorem 5, a feasible solution of Program 1 gives an upper bound for $\rho(G, \mathcal{M})$, and thus, for any positive degree $2d$,

$$\rho(G, \mathcal{M}) \leq \rho_{\mathrm{SOS},2d}(G, \mathcal{M}). \tag{1.15}$$

Example 3. Consider the unconstrained system [AP12b, Example 2.1] with $m = 3$:

$$\mathcal{M} = \{A_1 = e_1 e_2^\top, A_2 = e_2 e_3^\top, A_3 = e_3 e_1^\top\}$$

where e_i denotes the ith canonical basis vector.

For any d, a solution to Program 1 is given by

$$(p(x), \gamma) = (x_1^{2d} + x_2^{2d} + x_3^{2d}, 1).$$

Indeed, for example, with A_1 we have

$$p(e_1 e_2^\top x) = p(x_2, 0, 0) = x_2^{2d} + 0 + 0 \leq x_1^{2d} + x_2^{2d} + x_3^{2d}.$$

Example 4. Let us reconsider our running example; see Example 2. The optimal solution of Program 1 is represented by Figure 1.3 for $2d = 2$, 4, 6, 8, 10 and 12. We can see a big difference between the shape of the sublevel sets for $2d = 2$ and $2d = 4$ while between $2d = 4$ and $2d = 12$, the difference seems to be more subtle.

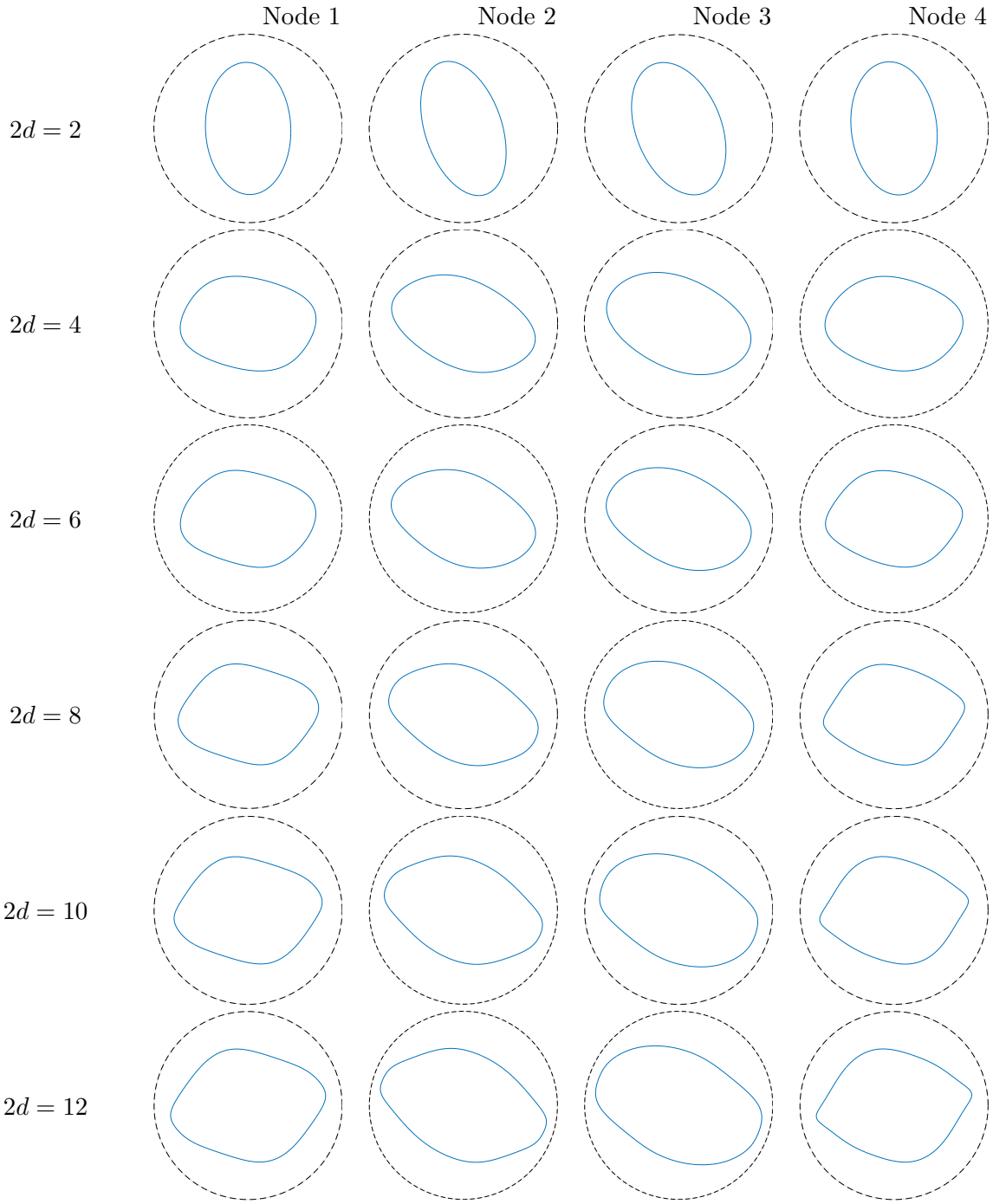


Figure 1.3: Representation of the solutions to Program 1 with different values of d for the running example. The blue curve represents the boundary of the 1-sublevel set of the optimal solution p_v at each node $v \in V$. The dashed curve is the boundary of the unit circle.

1.4 Computation of invariant set for switched systems

An *invariant set* is defined a follows.
Definition 3. A set $\mathcal{S}$ is said to be an *invariant set* of the system (1.11) if

$$\forall x \in \mathcal{S}, \forall \sigma \in [m], A_\sigma x \in \mathcal{S}.$$

which is equivalent to the two following proposition

$$\forall \sigma \in [m], A_\sigma \mathcal{S} \subseteq \mathcal{S} \tag{1.16}$$

$$\forall \sigma \in [m], \mathcal{S} \subseteq A_\sigma^{-1} \mathcal{S}. \tag{1.17}$$

The *minimal invariant (convex) superset* of a set $\mathcal{S}$ is the intersection of all the invariant (convex) supersets of $\mathcal{S}$ and the *maximal invariant (convex) subset* of a set $\mathcal{S}$ is the union (resp. convex hull) of all the invariant (convex) subsets of $\mathcal{S}$.

The minimal (resp. maximal) invariant set is indeed an invariant set since the intersection (resp. union, convex hull) of invariant sets is invariant.

In this section we consider the problem of computing a minimal (resp. maximal) invariant convex set.

1.4.1 Polyhedral computation of a minimal invariant superset

One can compute the minimal invariant superset of $\mathcal{S}_0$ using dynamic programming thanks to the following lemma which is a direct consequence of (1.16).
Lemma 5. If $\mathcal{T} \supseteq \mathcal{S}$ is an invariant set of the system (1.11) then

$$\mathcal{T} \supseteq \mathcal{S} \cup \bigcup_{\sigma \in [m]} A_\sigma \mathcal{S}.$$

The minimal invariant set is therefore obtained by the limit of the sequence defined by the following iteration:

$$\mathcal{S}_{k+1} = \mathcal{S}_k \cup \bigcup_{\sigma \in [m]} A_\sigma \mathcal{S}_k. \tag{1.18}$$

This sequence is increasing using the ordering $\subseteq$ so if it is bounded, it will converge. It turns out that it is bounded if and only if $\rho \leq 1$. Moreover if $\mathcal{S}_0$ is a polytope and $\rho < 1$ then it converges in finitely many iteration.

If one is looking for the minimal *convex* invariant superset of $\mathcal{C}_0$ then the following iteration can be used instead:

$$\mathcal{C}_{k+1} = \text{conv}(\mathcal{C}_k, \text{conv}_{\sigma \in [m]} A_\sigma \mathcal{C}_k). \tag{1.19}$$

Note that if $\mathcal{C}_0$ is polyhedral then so is $\mathcal{C}_k$ for any k. Moreover, if $\mathcal{C}_0$ is a polytope and $\rho < 1$, since it converges in finitely many iteration, $\mathcal{C}_\infty$ will also be a polytope.

From a computational viewpoint, we just need to keep the V-representation of $\mathcal{C}_k$ and for each k we can just

- compute the V-representation of $A_\sigma \mathcal{C}_k$ for each A_σ which amounts to mutliplying each vertex of $\mathcal{C}_k$ by A_σ,
- computing the convex hull, $\text{conv}(\mathcal{C}_k, \text{conv}_\sigma A_\sigma \mathcal{C}_k)$ which amounts to concatenating all the V-representation
- and then, optionally, remove the redundancies in the V-representation.

1.4.2 Polyhedral computation of a maximal invariant subset

One can compute a maximal invariant subset of the set $\mathcal{S}_0$ similarly using the following lemma which is a direct consequence of (1.17).

Lemma 6. If $\mathcal{T} \subseteq \mathcal{S}$ is an invariant set of the switched system (1.11) then

$$\mathcal{T} \subseteq \mathcal{S} \cap \bigcap_{\sigma} A_\sigma^{-1}\mathcal{S}.$$

The maximal invariant set is therefore obtained by the limit of the sequence defined by the following iteration:

$$\mathcal{S}_{k+1} = \mathcal{S}_k \cap \bigcap_{\sigma} A_\sigma^{-1}\mathcal{S}_k. \tag{1.20}$$

If one is looking for the maximal *convex* invariant subset of $\mathcal{C}_0$ then the same iteration (1.20) can be used. If $\mathcal{C}_0$ is convex then so is $\mathcal{C}_k$ for any k since the intersection of convex sets is convex. Also, if $\mathcal{C}_0$ is polyhedral then so is $\mathcal{C}_k$ for any k.

This sequence is decreasing using the ordering $\subseteq$ and bounded by $\emptyset$ so it converges. Therefore we see that the maximal invariant subset of a convex set is convex. It turns out that it converge to an non-empty set if and only if $\rho \leq 1$. Moreover if $\mathcal{C}_0$ is a polytope and $\rho < 1$ then it converges in finitely many iteration and hence $\mathcal{C}_\infty$ is a polytope.

From a computational viewpoint, we just need to keep the H-representation of $\mathcal{C}_k$ and for each k we can just

- compute the H-representation of $A_\sigma^{-1}\mathcal{C}_k$ for each A_σ which amounts to multiplying each facet of $\mathcal{C}_k$ by $A_\sigma^\top$,
- computing the intersection, $\mathcal{C}_k \cap \bigcap_\sigma A_\sigma^{-1}\mathcal{C}_k$ which amounts to concatenating all the H-representation
- and then, optionally, remove the redundancies in the H-representation.

As we can see by comparing (1.19) and (1.20) using Lemma 1 and Lemma 3, computing the minimal invariant *convex* superset of a convex set $\mathcal{C}$ containing the origin in its interior[2] is equivalent to computing the maximal invariant subset of $\mathcal{C}^\circ$ for the system

$$x_{k+1} = A_\sigma^\top x_k.$$

Moreover, the iterates $\mathcal{C}_k$ are polar to each other. The same equivalence applies in the other direction except that we need to take the closure.

This is in fact a consequence of the following general result.

Lemma 7. A convex set $\mathcal{C}$ is an invariant set of the system (1.11) if and only if

$$\forall \sigma \in [m], A_\sigma^\top \mathcal{C}^\circ \subseteq \mathcal{C}^\circ \tag{1.21}$$

or equivalently

$$\forall \sigma \in [m], \mathcal{C}^\circ \subseteq A_\sigma^{-\top}\mathcal{C}^\circ. \tag{1.22}$$

Proof. Using (1.1), we see that (1.16) is equivalent to

$$\forall \sigma \in [m], \mathcal{C}^\circ \subseteq (A_\sigma \mathcal{C})^\circ.$$

Then using Lemma 1, we can rewrite it as

$$\forall \sigma \in [m], \mathcal{C}^\circ \subseteq A_\sigma^{-\top}\mathcal{C}^\circ.$$

□

[2] this technical condition is there only to ensure that the polar of $\mathcal{C}$ exists

1.4.3 Nonlinear approximation of invariant sets

It is well known that convex sets can be approximated by ellipsoid and this approximation can be made arbitrarily precise by using sublevel sets of sum of squares polynomials.
Lemma 8 ([Bar02, Sch13])**.** Given a convex body $\mathcal{C}$, among all ellipsoid containing $\mathcal{C}$, there is a unique one of minimal volume, usually called the *Loewner ellipsoid* of $\mathcal{C}$. Similary, among all ellipsoid contained in $\mathcal{C}$, there is a unique one of maximal volume, known as the *John ellipsoid* of $\mathcal{C}$.

This fact motivates the approximation of the invariant set by quadratic forms, or more generally, by homogeneous polynomial functions. We first show how to do it using ellipsoids and then generalize it to the sublevel sets of homogeneous polynomials of degree $2d$.

The nonlinear computation of invariant set is based in two simple facts. First, the 1-sublevel sets of $f(x)$ is included in the 1-sublevel set of $g(x)$ if and only if $g(x) \leq f(x)$. Second, if $\mathcal{S}$ is the 1-sublevel set of $f(x)$ then $A^{-1}\mathcal{S}$ is the 1-sublevel set of $f(Ax)$:

$$A^{-1}\{\, x \mid f(x) \leq 1 \,\} = \{\, x \mid f(Ax) \leq 1 \,\}. \tag{1.23}$$

Using those two facts and (1.17), we can prove the following lemma.
Lemma 9. The 1-sublevel of a function $f(x)$ is an invariant set of the system (1.11) if and only if

$$\forall \sigma \in [m], f(Ax) \leq f(x).$$

Note that since we will deal with homogeneous functions, we can talk about the invariance of its sublevel sets in general without specifying for which α it is the α-sublevel set.

Quadratic approximation

An ellipsoid (centered at the origin) is a set of the form

$$E = \{\, x \in \mathbb{R}^d \mid \langle Qx, x \rangle \leq 1 \,\} \tag{1.24}$$

where Q is a symmetric positive definite matrix.

The following two lemmas show how to compute the polar and the volume of ellipsoids.
Lemma 10. The polar of the ellipsoid defined by (1.24) is

$$E^\circ = \{\, x \in \mathbb{R}^d \mid \langle Q^{-1}x, x \rangle \leq 1 \,\}.$$

Proof. Let $Q = T^\top T$ be a Cholesky factorization of Q. We can see that TE is the euclidean unit ball B which is self-polar. Therefore, by Lemma 1,

$$E^\circ = (T^{-1}B)^\circ = T^\top B^\circ = T^\top B.$$

□

Lemma 11. The volume of the ellipsoid defined by (1.24) is

$$\operatorname{vol} E = \frac{\operatorname{vol} B}{\sqrt{\det Q}}.$$

where B is the euclidean unit ball.

The following program computes the minimal invariant set of a set $\mathcal{C}_0$. The practical implementation of the constraint (1.26) depends on $\mathcal{C}_0$. If $\mathcal{C}_0$ is an ellipsoid $\{ x \in \mathbb{R}^n \mid \langle Px, x \rangle \leq 1 \}$ then we can replace it with $Q \succeq P$. If $\mathcal{C}_0$ is a polytope, since an ellipsoid is always convex, it suffices to enforce (1.26) for all the vertices of $\mathcal{C}_0$.

Program 2 (Minimal Invariant Superset Quadratic Approximation)**.**

$$\begin{aligned} \underset{Q \in \mathcal{S}^n}{\text{maximize}} \; & \det(Q) \\ & Q \succ 0 && (1.25) \\ & A_\sigma^\top Q A_\sigma \preceq Q \quad \forall \sigma \in [m] \\ & \langle Qx, x \rangle \leq 1 \quad \forall x \in \mathcal{C}_0. && (1.26) \end{aligned}$$

For the maximal invariant subset, we need to impose that the ellipsoid is a subset of $\mathcal{C}_0$. If $\mathcal{C}_0$ is an ellipsoid, it is as easy as for the minimal invariant superset. However, if $\mathcal{C}_0$ is a polytope, contraining every vertex of $\mathcal{C}_0$ to be outside the ellipsoid is not sufficient. Instead, we should contraint every facet of $\mathcal{C}_0$ to be in the polar of the ellipsoid. That is, for every facet x of $\mathcal{C}_0$, we must have $xQ^{-1}x \leq 1$. This constraint is on Q^{-1} but it can be transformed into a constraint on Q using the Schur's complement.

Definition 4. Given a block-partitioned matrix

$$\begin{bmatrix} A & B \\ B^\top & C \end{bmatrix},$$

where A is square and invertible, the *Schur complement* of A is the matrix $C - B^\top A^{-1} B$.

Lemma 12. If $A \succ 0$ then

$$\begin{bmatrix} A & B \\ B^\top & C \end{bmatrix} \succeq 0 \iff C - B^\top A^{-1} B \succeq 0.$$

As we constrain Q to be positive definite with (1.25), by Lemma 12, $\langle Q^{-1}x, x \rangle \leq 1$ if and only if

$$\begin{bmatrix} Q & x \\ x^\top & 1 \end{bmatrix} \succeq 0.$$

Another approach is to do the computation on the polar using Lemma 7. This second approach gives Program 3.

Program 3 (Polar of Maximal Invariant Subset Quadratic Approximation)**.**

$$\begin{aligned} \underset{Q^{-1} \in \mathcal{S}^n}{\text{maximize}} \; & \det(Q^{-1}) \\ & Q^{-1} \succ 0 \\ & A_\sigma Q^{-1} A_\sigma^\top \preceq Q^{-1} \quad \forall \sigma \in [m] \\ & \langle Q^{-1}x, x \rangle \leq 1 \quad \forall x \in \mathcal{C}_0{}^\circ. \end{aligned}$$

Higher-order approximation

In this section we consider the approximation of invariant set using sublevel sets of homogeneous sum of squares polynomials of degree $2d$.

The ideal computation of minimal invariant superset would be as follows.

Program 4 (Minimal Invariant Superset Approximation Sketch)**.**

$$\begin{aligned}\underset{p(x)\in\mathbb{R}_{\mathbf{2d}}[x]}{\text{minimize}}\ & \operatorname{vol}(\{\, x \mid p(x) \le 1 \,\}) \\ & p(x) > 0 && (1.27)\\ & p(A_\sigma x) \le p(x) \quad \forall \sigma \in [m] && (1.28)\\ & p(x) \le 1 \quad \forall x \in \mathcal{C}_0. && (1.29)\end{aligned}$$

None of the four lines of Program 4 are directly implementable. As discussed in Section 1.2, deciding whether a polynomial is positive (or nonnegative) is NP-hard. For this reason, we relax (1.27) and (1.28) respectively to "$p(x)$ is SOS" and "$p(x) - p(A_\sigma x)$ is SOS". The replacement for (1.28) imposes a stronger condition than invariance that we will call *sos invariance*.

Definition 5 (sos invariance)**.** The sublevel set of a polynomial function $p(x)$ is said to be an *sos-invariant set* of the system (1.11) if

$$\forall \sigma \in [m], p(x) - p(A_\sigma x) \text{ is SOS.}$$

We do not know any simple formula for the volume of the sublevel set of $p(x)$. However, as discussed in Section 1.2, if we choose some basis of monomial $x^{[d]}$, there exists some positive semidefinite matrix $Q \in \mathcal{S}^n_{++}$ such that $p(x) = \langle Qx^{[d]}, x^{[d]} \rangle$ so $p(x)$ can be seen as an ellipsoid in a lifted space [PEDJ15].

More precisely, consider the following operator on sets $\mathcal{S}^{[d]} = \{\, x^{[d]} \mid x \in \mathcal{S} \,\}$. Let $E = \{\, x \mid p(x) \le 1 \,\}$. We have the following identity

$$E^{[d]} = \{\, y \mid x^T Q x \le 1 \,\} \cap (\mathbb{R}^n)^{[d]}.$$

That is, $E^{[d]}$ is the intersection of an ellipsoid and the *Veronese variety* $(\mathbb{R}^n)^{[d]}$.

Example 5. Consider the polynomial $p(x) = (x_1^2 + x_2^2)^2 = x_1^4 + 2x_1^2x_2^2 + x_2^4$. We consider the scaled monomial basis introduced in Section 1.2. The Veronese variety is given by

$$(\mathbb{R}^2)^{[2]} = \{\, (x_1^2, x_2^2, x_1x_2) \mid x \in \mathbb{R}^2 \,\} = \{\, y \mid y_3^2 = y_1y_2 \,\}.$$

The set $E^{[d]}$ is the intersection of the ellipsoid $\{\, y \mid y_1^2 + y_2^2 + 2y_3^2 \le 1 \,\}$ with $(\mathbb{R}^2)^{[2]}$. This is illustrated by Figure 1.4.

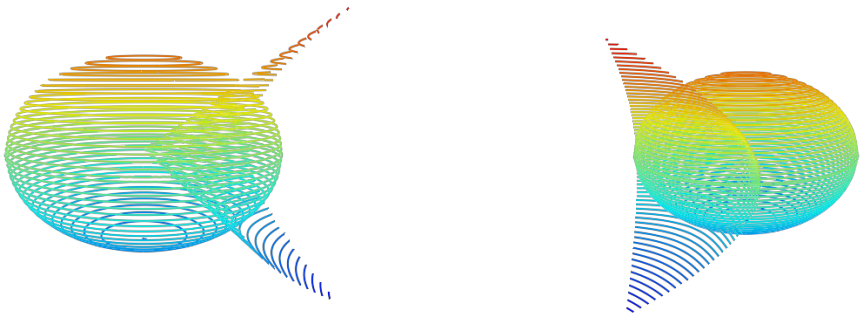


Figure 1.4: Illustration for the Veronese variety.

In view of the preceding discussion, instead of minimizing the volume of $E^{[d]}$, it does not seem unreasonable to minimize the the volume of the ellipsoid $\{\, y \mid \langle Qy, y \rangle \le 1 \,\}$ which is easily computable thanks to Lemma 11. Note however that there is a whole affine space of matrices Q such that $p(x) = (x^{[d]})^T Q x^{[d]}$ so Q can be chosen in the intersection of this affine space with the cone $\mathcal{S}^n_{++}$ of semidefinite matrices. This gives Program 5.

Program 5 (Minimal Invariant Superset Approximation)**.**

$$\begin{aligned}\underset{p(x)\in\mathbb{R}_{\mathbf{2d}}[x]}{\text{maximize}}\ & \det(Q)\\ Q &\succ 0\\ p(x) &= \langle Qx^{[d]}, x^{[d]}\rangle\\ p(x) - p(A_\sigma x) &\text{ is SOS} \quad \forall \sigma \in [m]\\ p(x) &\leq 1 \quad \forall x \in \mathcal{C}_0.\end{aligned}$$

The implementation of (1.29) depends again on $\mathcal{C}_0$. If $\mathcal{C}_0$ is the sublevel set of some polynomial $q(x)$ then we can use the constraint "$q(x) - p(x)$ is SOS". If $\mathcal{C}_0$ is a polytope then we might be tempted, as we did for the quadratic case, to impose that $p(x) \leq 1$ for any extreme point of $\mathcal{C}_0$. However, the sublevel sets of $p(x)$ might not be convex[3]. The fact that $p(x)$ may not convex is also problematic if we are looking for the minimal *convex* invariant superset of $\mathcal{C}_0$. However, we have seen in Section 1.2.1 that we can constraint $p(x)$ to be sos-convex. Therefore, if we are looking for the minimal convex invariant superset, we can add the sos-convexity constraint solution, solving the two above issues at the same time.

We now turn to the problem of computing the maximal invariant subset. As in the quadratic case, the computation is similar when $\mathcal{C}_0$ is the sublevel set of a polynomial. When $\mathcal{C}_0$ is a polytope, as with the quadratic approximation, we have two choices. We can either use Lemma 12 or do the computation on the polar using Lemma 7.

However, for the second option, once we have computed an approximation for the polar as the sublevel set of $p(x)$ found after solving the polar program Program 5 (where the A_σ are replaced by $\mathcal{M}_\sigma^\top$ and $\mathcal{C}_0$ is replaced by $\mathcal{C}_0^\circ$ and the $p(x)$ is constrained to be sos-convex), we still need to compute its polar. But how do we compute the polar of the sublevel set of a polynomial ? Two solution can be considered.

The first solution would be to compute the dual of the sublevel set of $p(x)$ using algebraic techniques. The boundary of $p(x)$, defined by $p(x) = 1$ is an algebraic variety. The dual of this variety is the algebraic boundary of the polar of the 1-sublevel set of $p(x)$. More details on this can be found in [BPT12, Chapter 5], the algorithm for computing the dual variety is [BPT12, Algorithm 5.1]. The drawback of this solution is that it is rather computationaly intensive as it requires the elimination of variables in an algebraic variety.

The second solution would be to take $\langle Q^{-1}x^{[d]}, x^{[d]}\rangle$, that is, take the polar of the ellipsoid of minimum volume in the lifted space. This solution may seem somehow risky. Indeed, if the 1-sublevel set of $p(x)$ is invariant (or even sos-invariant) with the transposed matrices $\mathcal{M}^\top$, does that mean that $\{\, x \mid \langle Q^{-1}x^{[d]}, x^{[d]}\rangle \leq 1 \,\}$ is invariant for $\mathcal{M}$? Lemma 7 does not guarantee that.

We define a stronger notion of invariance.

Definition 6 (lift invariance [AJ16])**.** The sublevel set of a polynomial function $p(x)$ is said to be a *lift-invariant set* of the system (1.11) if there exists a matrix $Q \in \mathcal{S}^n_{++}$ such that $p(x) = \langle Qx^{[d]}, x^{[d]}\rangle$ and

$$\forall \sigma \in [m], Q - A_\sigma^\top Q A_\sigma \succeq 0.$$

Clearly, lift-invariance implies sos-invariance which implies invariance. Moreover, by Lemma 7, if $p(x)$ is lift-invariant for some matrix $Q \in \mathcal{S}^n_{++}$, then $q(x) = \langle Q^{-1}x^{[d]}, x^{[d]}\rangle$ is also lift-invariant and therefore invariant too. Therefore, we can constrain $p(x)$ to be lift-invariant and use polar of the lift. This idea is implemented in Program 6.

[3]In fact, since $p(x)$ is homogeneous, its sublevel sets are convex exactly when it is convex.

Program 6 (Maximal Invariant Subset of a polytope Approximation)**.**

$$\begin{aligned}
\underset{Q^{-1}\in\mathcal{S}^n}{\text{minimize}}\ & \det(Q^{-1}) \\
& Q^{-1} \succ 0 \\
Q^{-1} - A_\sigma^{[d]} Q^{-1} (A_\sigma^{[d]})^\top & \succeq 0 \quad \forall \sigma \in [m] \\
& \langle Q^{-1} x^{[d]}, x^{[d]} \rangle \text{ is sos-convex} \\
& \langle Q^{-1} x^{[d]}, x^{[d]} \rangle \leq 1 \quad \forall \text{ facet } f \text{ of } \mathcal{C}_0.
\end{aligned}$$

One might be concerned about the fact that we constrain the set $\{ x \mid \langle Q^{-1} x^{[d]}, x^{[d]} \rangle \leq 1 \}$ to contain $\mathcal{C}_0^\circ$ but we are taking the dual in the lifted space so how can we be sure that $\{ x \mid \langle Q^{-1} x^{[d]}, x^{[d]} \rangle \leq 1 \}$ is contained in $\mathcal{C}_0$? It turns out that this can be ensured using the following lemma if the basis chosen is the scaled monomial basis.

Lemma 13. Consider two convex sets $\mathcal{C}_1 \in \mathbb{R}^n$ and $\mathcal{C}_2 \in \text{aff}((\mathbb{R}^n)^{[d]})$ where $x^{[d]}$ is the scaled monomial basis. If $\mathcal{C}_1^{[d]} \subseteq \mathcal{C}_2 \cap (\mathbb{R}^n)^{[d]}$ then $\mathcal{C}_2{}^\circ \cap (\mathbb{R}^n)^{[d]} \subseteq (\mathcal{C}_1{}^\circ)^{[d]}$.

Proof. Suppose $y^{[d]} \in \mathcal{C}_2{}^\circ \cap (\mathbb{R}^n)^{[d]}$. We have $\langle y^{[d]}, z \rangle \leq 1$ for any $z \in \mathcal{C}_2$ hence we also have $\langle y^{[d]}, x^{[d]} \rangle \leq 1$ for any $x \in \mathcal{C}_1$. By (1.10), for the scaled monomial basis, $\langle y^{[d]}, x^{[d]} \rangle = (\langle y, x \rangle)^d$ so $\langle y, x \rangle \leq 1$ for any $x \in \mathcal{C}_1$. That is, $y \in \mathcal{C}_1{}^\circ$. □

1.5 The Entropic Cone

In 1948, Shannon published "A Mathematical Theory of Communication" [Sha48]. In this paper, Shannon introduces the entropy of a random variable. Suppose we have a random variable X of alphabet $\mathcal{X}$, he defines the entropy of X as

$$H_b(X) = \sum_{x \in \mathcal{X}} \Pr[X = x] \log_b \frac{1}{\Pr[X = x]}$$

where the basis b is positive. If b is 2 (resp. e), the unit is the bits (resp. nats). Note that $H_b(X) = H_a(X) \log_b(a)$ so the entropies using different basis are equivalent up to a positive constant factor.

This quantity represents the "amount of information contained in X". More precisely, suppose Alice and Bob know the probability distribution of X. They can meet and set up a communication protocol. Then Alice will see realizations x of X and will need to tell Bob the value x using their transmission protocol. The entropy of X is the minimum, over all possible protocols, of the *expected* amount of bits needed to transfer one value.

An optimal protocol is given the entropic coding: for this protocol, each value x is encoded in $\log_2 \frac{1}{\Pr[X=x]}$ bits.

The entropy of several random variables in simply the entropy of their cartesian product:

$$\begin{aligned} H_b(\{X_1, \dots, X_n\}) &= H_b((X_1, \dots, X_n)) \\ &= \sum_{(x_1,\dots,x_n) \in \mathcal{X}_1 \times \cdots \times \mathcal{X}_n} \Pr[(X_1, \dots, X_n) = (x_1, \dots, x_n)] \log_b \frac{1}{\Pr[(X_1, \dots, X_n) = (x_1, \dots, x_n)]}. \end{aligned}$$

By convention, we say that the entropy of an empty set of random variables is 0.

Given a n random variables, we can compute the entropy of any of the 2^n subset of those n variables. The entropic vector of a set of n random variables is a vector h, indexed by the subsets of $[n] = \{1, \dots, n\}$, such that $h_S = H_b(\{ X_i \mid i \in S \})$.

We denote the set of vectors of $\mathbb{R}^{2^n-1}$ that are entropic as:

$$\mathcal{H}_n = \{ h \in \mathbb{R}^{2^n-1} \mid \exists X_1, \dots, X_n, \forall \emptyset \neq S \subseteq [n], h_S = H_b(\{ X_i \mid i \in S \}) \}.$$

We do not include the dimension corresponding to the entropy of the empty set as it is zero. We will see that thanks to this choice, the cone $\mathcal{H}_n$ so that it is solid, i.e. full-dimensional.

The entropic cone has many applications including Network coding [BMR+13, CG07, DFZ05], secret sharing [Bei11, MFP07], guessing games [BCD+13], quantum information [Pip03], conditional independence [MS95], Additive combinatorics [MMT10] and Group Theory [CY02].

The set $\mathcal{E}_n \triangleq \mathbb{R}^{2^n-1}$ of vectors indexed by the nonempty subsets of $[n]$ is called the *entropy space* and its elements are called *entropy vectors*. Entropy vectors can also be viewed as function of $[n] \setminus \emptyset \to \mathbb{R}$. A special class of entropy vectors will be encountered quite often so we will give them a special name.

Definition 7. For a given set $I \subseteq [n]$ and a number m, we define the entropy vector

$$r_I^m = J \mapsto \min(m, |J \cap I|).$$

That is, in the index notation, if $h = r_I^m$ then $h_J = \min(m, |J \cap I|)$.

Note that we do not specify the number of variables n for r as most of the time, it will be clear from the context.

As it is often done in the litterature [Cha11, MC13], we will use shortened notations for set. For instance, the singleton $\{i\}$ will be denoted as i and the set $\{i, j, k\}$ will be denoted as ijk.

1.5.1 The entropic cone of 1 variable

Consider first one binary random variable X with alphabet $\mathcal{X} = \{0, 1\}$ and $\Pr[X = 1] = \alpha$ for some $0 \leq \alpha \leq 1$. We have $H_b(X) = -\alpha \log_b(\alpha) - (1-\alpha)\log_b(1-\alpha) \triangleq H_b(\alpha)$. The function $H_b(\alpha)$ is shown in Figure 1.5. We see that depending on α, $H_b(X)$ can take any value between 0 and $H_b(\frac{1}{2})$ or, more formally, between r_1^0 and $r_1^{H_b(\frac{1}{2})}$. This observation leads to the following two questions.

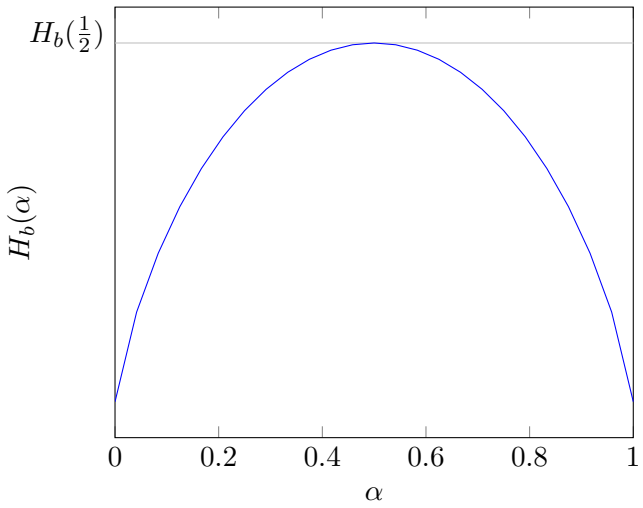


Figure 1.5: The value of $H_b(\alpha)$ for $0 \leq \alpha \leq 1$.

1. Can the entropy take negative values ?
2. Can the entropy take values higher than $H_b(\frac{1}{2})$?

To answer the first question, we generalize the entropy function H_b.

Definition 8. Consider a fixed basis b and two probability mass functions p, q over the *same* alphabet $\mathcal{X}$. The entropy function $H_p(q)$ is defined as

$$H_p(q) = \sum_{x \in \mathcal{X}} p(x) \log \frac{1}{q(x)}.$$

The notation $H_p(p)$ will sometimes be shortened as $H(p)$ or even $H_b(p)$ so specify the base.

Denote the support of p, or a random variable X with probability mass function p, as $\operatorname{supp}(X) = \operatorname{supp}(p) = \{\, x \in \mathcal{X} \mid p(x) > 0 \,\}$. The nonnegativity of $x \mapsto -\log(x)$ for $x \leq 1$ implies the following property of $H_q(p)$.

Proposition 3. For any probability mass functions p, q over the same alphabet,

$$H_p(q) \geq 0$$

with equality if and only if $\operatorname{supp}(p) \cap \operatorname{supp}(q) = \emptyset$ or $|\operatorname{supp}(p)| = 1$.

In particular, if we take p as the joint probability mass function of a set of random variables I, Proposition 3 on $H_p(p)$ gives the following Corollary.

Corollary 1. For any set of random variables X_I,

$$H(X_I) \geq 0$$

with equality if and only if all random variables X_i, $i \in I$, are deterministic.

Corollary 1 answers the first questions, we now turn our attention to the second one. As one can anticipate, using a larger alphabet, we can get a higher entropy. For instance, if $\mathcal{X} = \{00, 01, 10, 11\}$ where the first and second bits of the characters are independent and have the value 1 with probability p_1 and p_2, then the entropy is $H(p_1) + H(p_2)$.

This observation can be generalized as follows.

Lemma 14 (The entropic cone is close to addition)**.** If $g, h \in \mathcal{H}_n$, then $g + h \in \mathcal{H}_n$.

Proof. Let X (resp. Y) be a random variables with entropy vector g (resp. h). We define the variables $Z_i = (X_i, Y_i)$ with X_i and Y_i independent, that is, $\Pr[Z_i = (x, y)] = \Pr[X_i = x] \Pr[Y_i = y]$. The entropy vector of Z is $g + h$. □

In particular, we have the following corollary.

Corollary 2. If $h \in \mathcal{H}_n$, then for any nonnegative integer m, $mh \in \mathcal{H}_n$.

We can now conclude. We have seen that $[0, H_b(\frac{1}{2})] \subseteq \mathcal{H}_1$ which, using Corollary 2, implies that $\mathbb{R}^+ \subseteq \mathcal{H}_1$. From Corollary 1, we have $\mathcal{H}_1 \subseteq \mathbb{R}^+$ hence $\mathcal{H}_1 = \mathbb{R}^+$.

1.5.2 Kullback-Leibler divergence and convexity

Kullback and Leibler introduces the so-called Kullback-Leibler divergence [KL51] between two probability mass functions p, q over the *same* alphabet $\mathcal{X}$:

$$\begin{aligned} D_{\mathrm{KL}}(p\|q) &= H_p(q) - H_p(p) \\ &= \sum_{x \in \mathcal{X}} p(x) \log \frac{p(x)}{q(x)}. \end{aligned}$$

We recall here the Jensen's inequality.

Theorem 6 (Jensen's inequality)**.** If $f \colon \mathbb{R}^n \to \mathbb{R}^n$ is convex on its domain then for any $x_1, \ldots, x_n \in \operatorname{dom} f$ and $\lambda_1, \ldots, \lambda_n \geq 0$ such that $\lambda_1 + \cdots + \lambda_n = 1$,

$$f(\lambda_1 x_1 + \cdots + \lambda_n x_n) \leq \lambda_1 f(x_1) + \cdots \lambda_n f(x_n). \tag{1.30}$$

Moreover if f is strictly convex, the equality occurs if and only if $x_1 = \cdots = x_n$ or one of the $\lambda_i = 1$.

Since $x \mapsto -\log x$ is strictly convex on its domain $\mathbb{R}^+$, the Kullback-Leibler divergence has the following property.

Proposition 4. For any probability mass functions p, q over the same alphabet,

$$D_{\mathrm{KL}}(p\|q) \geq 0$$

with equality if and only if $p = q$.

Proof. By Jensen's inequality on $-\log x$,

$$\begin{aligned} D_{\mathrm{KL}}(p\|q) &= -\sum_{x \in \mathcal{X}} p(x) \log \frac{q(x)}{p(x)} \\ &\geq -\log \sum_{x \in \mathcal{X}} p(x) \frac{q(x)}{p(x)} \\ &= -\log q(x) \\ &= 0. \end{aligned}$$

If $D_{\mathrm{KL}}(p\|q) = 0$, we have either $p(x) = 1$ for some $x \in \mathcal{X}$ or $q(x)/p(x) = q(x')/p(x')$ for any $x, x' \in \mathcal{X}$. In the first case, $-\log q(x) = 0$ so $q(x) = 1$ and $p = q$. In the second case, $q(x)/q(x') = p(x)/p(x')$ for any $x, x' \in \mathcal{X}$ so p and q are proportional. Since $\sum_{x\in\mathcal{X}} p(x) = 1 = \sum_{x\in\mathcal{X}} q(x)$, they are equal. □

1.5.3 The Shannon inequalities

We will see further than $\operatorname{cl}\mathcal{H}_n$ is a convex cone. We can therefore define its dual $(\operatorname{cl}\mathcal{H}_n)^*$, which is[4]

$$\{\, h^* \in \mathcal{E}_n^* \mid \langle h^*, h\rangle \geq 0, \forall h \in \mathcal{H}_n \,\}$$

where the dual entropy space $\mathcal{E}_n^*$ is simply $\mathbb{R}^{2^n-1}$. The dual $(\operatorname{cl}\mathcal{H}_n)^*$ is the set of all linear inequalities satisfied for all entropy vector of $\mathcal{H}_n$.

The set of Shannon inequalities is a subset of the linear inequalities satisfied by all entropic vector. That is, it is a cone $\mathcal{P}_n^* \subseteq (\operatorname{cl}\mathcal{H}_n)^*$. The cone $\mathcal{P}_n^*$ is polyhedral as it is given by all the linear inequalities that are consequence of a finite number of linear inequalities.

These inequalities are the following ones:

normalized

$$h_\emptyset = 0, \tag{1.31}$$

nonnegative For any I,

$$h_I \geq 0. \tag{1.32}$$

nondecreasing If $I \subseteq J$, then

$$h_I \leq h_J. \tag{1.33}$$

submodular For all J, K,

$$h_J + h_K \geq h_{J\cup K} + h_{J\cap K}. \tag{1.34}$$

Or equivalently: If $I \subseteq J$, then for any K,

$$h_{K\cup I} - h_I \geq h_{K\cup J} - h_J. \tag{1.35}$$

We can see that (1.32) is a consequence of (1.31) and (1.33). The inequality (1.32) is a direct consequence of Proposition 3 and the inequality (1.33) is a consequence of Proposition 3 and Proposition 5.

Proposition 5. For any $I \subseteq J$,

$$H_b(X_J) - H_b(X_I) = \sum_{x\in\mathcal{X}_I} \Pr[X_I = x] H_b(X_J|X_I = x)$$

where $X_J|X_I = x$ is the random variable of alphabet $\mathcal{X}_{J\setminus I}$ with probability mass function $y \mapsto \Pr[X_J = y|X_I = x]$.

[4]Note that we do not need the closure in the definition of $(\operatorname{cl}\mathcal{H}_n)^*$

Proof. We have

$$\begin{aligned} H(X_J) - H(X_I) &= \sum_{x \in \mathcal{X}_J} \Pr[X_J = x] \log \frac{\Pr[X_I = x_I]}{\Pr[X_J = x]} \\ &= \sum_{x \in \mathcal{X}_J} \Pr[X_J = x] \log \frac{1}{\Pr[X_J = x | X_I = x_I]} \\ &= \sum_{x \in \mathcal{X}_I} \Pr[X_I = x] \sum_{y \in \mathcal{X}_J} \Pr[X_J = y | X_I = x] \log \frac{1}{\Pr[X_J = y | X_I = x]}. \end{aligned}$$

□

Theorem 7. For any $I \subseteq J$,

$$\langle \Delta_{J|I}, h \rangle \triangleq h_J - h_I \geq 0$$

with equality if and only if $J \setminus I$ is deterministic when I is known.

The submodularity is the consequence of Proposition 4 and the following relation:

$$H_b(J) + H_b(K) - H_b(I) = H_b(p_{J|I} p_{K|I} p_I) \tag{1.36}$$

where $I = J \cap K$.

Theorem 8. If an entropy vector h is entropic and $J \cap K = I$ then

$$\langle \Delta_{J,K|I}, h \rangle \triangleq h_J + h_K - h_{J \cup K} - h_I \geq 0$$

with equality if and only if $p = p_{J|I} p_{K|I} p_I$, that is, $X_{J|I}$ and $X_{K|I}$ are independent.

1.5.4 The entropic cone of 2 variables

The entropic cone of 2 variables lives in $\mathbb{R}^3$. Its superset $\mathcal{P}_3$ has 3 extreme rays (or 2 up to symetry): r_1^1, r_2^1 and r_{12}^1. The first one, r_1^1 is entropic, it corresponds to a deterministic random variable X_2 and a coin flip random variable X_1. The third one is also entropic, it corresponds to a X_1 and X_2 being equal to the same coin flip.

Corollary 2 shows that mr is also entropic for each of our 3 rays r and any nonnegative integer m. In Section 1.5.1 we could also prove that αr_1^1 was entropic for any nonnegative number α.

We can use that fact and symmetry to prove that both the rays generated by r_1^1 and r_2^1 are entropic. To prove that the ray generated by r_{12}^1 is entropic, we can do the same to the variable $X_1 = X_2$.

This technique can be generalized. The idea is to decompose an entropy vector into independent components with positive entropy and then use the fact that each independent component is entropic using the fact that $\mathcal{H}_1 = \mathbb{R}_+$. This may not be entirely clear because we applied this technique to examples that were too trivial For this reason, we now present a more insightful example.

Example 6. Suppose we are given the entropy vector $h = 2r_1^1 + 2.5r_2^1 + 4.2r_{12}^1$. From the fact that $\mathcal{H}_1 = \mathbb{R}_+$, we can define 3 random variables Y_1, Y_2, Y_3 with $H_b(Y_1) = 2$, $H_b(Y_2) = 2.5$ and $H_b(Y_3) = 4.2$. Now, taking $X_1 = (Y_1, Y_3)$ and $X_2 = (Y_2, Y_3)$, the entropy vector of X is h.

The numbers $2, 2.5$ and 4.2 of the previous example are sometimes called the atoms of an entropy vector. If those atoms are nonnegative, we can define independent variables at each atoms using the fact that $\mathcal{H}_1 = \mathbb{R}_+$, we can show that the vector is entropic. Those atoms are commonly represented in a Venn diagram; see Figure 1.6.

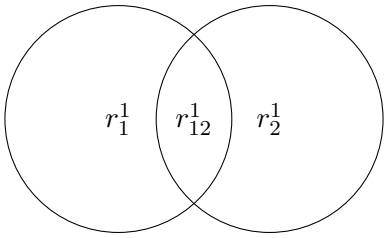


Figure 1.6: The decomposition of a entropy vector of 2 variables in its atoms.

In the general case of entropy vectors of n variables, the $2^n - 1$ vectors r_I^1 provides a base for the entropy space. The matrix of the change of basis is unimodular and can be obtained as follows. Suppose $h = \sum_{I \in [n]\setminus\emptyset} \alpha_I r_I^1$, we have

$$h_J = \sum_{I \subseteq J} \alpha_I$$
$$\alpha_I = \sum_{J \subseteq I} (-1)^{|I \setminus J|} (h_{[n]} - h_{[n]\setminus J}).$$

The first identity is easily obtained using the Venn diagram while the second identity can be obtained using the inclusion exclusion principle.

Definition 9. The *cone of positive atoms* $\mathcal{A}_n$ is defined as

$$\mathcal{A}_n = \{ \sum_{I \subseteq [n]\setminus\emptyset} \alpha_I r_I^1 \mid \alpha \geq 0 \}.$$

As we have seen, using the fact that $\mathcal{H}_1 = \mathbb{R}_+$, we have the following theorem.

Theorem 9. For any natural number n,

$$\mathcal{A}_n \subseteq \mathcal{H}_n.$$

Since $\mathcal{A}_n \subseteq \mathcal{H}_n \subseteq \mathcal{P}_n$ and $\mathcal{A}_2 = \mathcal{P}_2$, we have $\mathcal{A}_2 = \mathcal{H}_2 = \mathcal{P}_2$.

1.5.5 The entropic cone of 3 variables

For 3 variables, the relation $\mathcal{A}_3 = \mathcal{P}_3$ does not hold anymore. The extreme rays of $\mathcal{P}_3$ are the extreme rays of $\mathcal{A}_3$ except for $r_{[3]}^2 \in \mathcal{P}_3$ that is not $\mathcal{A}_3$; see Figure 1.7b for an illustration of $r_{[3]}^2$. This vector is entropic for $b = 2$ as shown by Example 7.

Example 7. Consider the three independent coin tosses X_1, X_2 and $X_3 = X_1 \oplus X_2$ where $\oplus$ is the XOR operator. One can verify that the entropy vector of those random variables is $H_2(X) = r_{[3]}^2$.

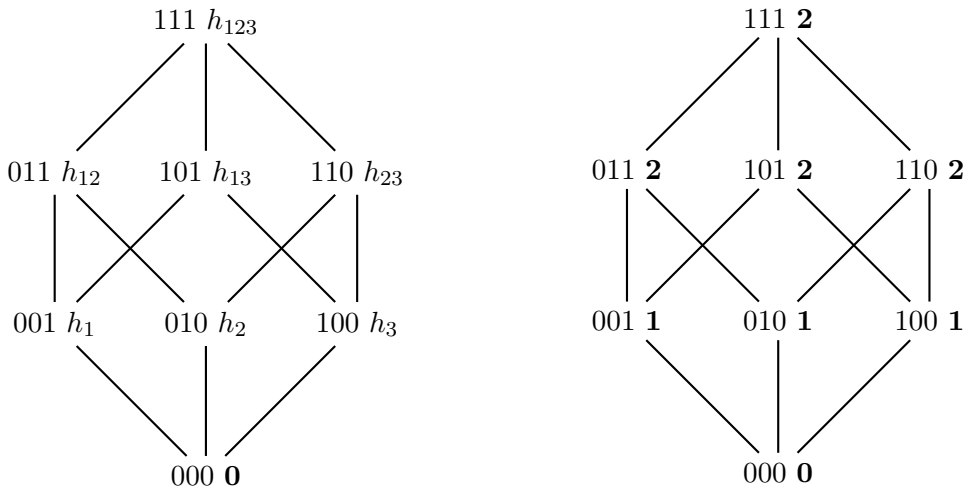
Now, to conclude that "$\mathcal{H}_3 = \mathcal{P}_3$", we would like to show that $\mathcal{H}_n$ is a cone. That is, we would like to generalize Corollary 2 and say that $\mathcal{H}_n$ is closed to multiplication with any nonnegative number. However, this cannot be shown and $\mathcal{H}_n$ is not a cone for $n \geq 3$.

This is shown by Example 8 that is taken from [ZY97].

Example 8 ([ZY97])**.** Consider again $r_{[3]}^2$. We can see that $r_{[3]}^2$ is supported by the following dual entropy vector: $\Delta_{1,2}, \Delta_{2,3}, \Delta_{3,1}, \Delta_{1|23}, \Delta_{2|13}, \Delta_{3|12}$.

Consider a positive multiple $h = \lambda r_{[3]}^2$ for some $\lambda > 0$. It is still supported by the same dual vectors so by Theorem 8, X_1, X_2, X_3 are pairwise independent. Therefore,

$$p_{12} = p_1 p_2$$
$$p_{13} = p_1 p_3$$
$$p_{23} = p_2 p_3.$$



(a) Lattice representation of an entropy vector h.

(b) $r^2_{[3]}$.

Figure 1.7: Lattice representation of entropy vector of $\mathcal{E}_3$.

Moreover, by Theorem 7, there exists functions $x_1(x_2, x_3), x_2(x_1, x_3), x_3(x_1, x_2)$ such that $p_{123} = p_{12}\delta_{x_3(x_1,x_2)} = p_{13}\delta_{x_2(x_1,x_3)} = p_{23}\delta_{x_1(x_2,x_3)}$.

That is, for any x_1, x_2, x_3,

$$p_1(x_1)p_2(x_2)\delta_{x_3(x_1,x_2)}(x_3) = p_1(x_1)p_3(x_3)\delta_{x_2(x_1,x_3)}(x_2) = p_2(x_2)p_3(x_3)\delta_{x_1(x_2,x_3)}.$$

We see with the first inequality that for any x_2, x_3, choosing $x_1 = x_1(x_2, x_3)$, we have

$$p_1(x_1)p_2(x_2) = p_1(x_1)p_3(x_3)$$

hence $p_2(x_2) = p_3(x_3)$. That is, X_1, X_2 and X_3 are uniform and have alphabets of the same length. Therefore, the only entropy vectors of the ray of $r^2_{[3]}$ that are entropic are the vectors

$$h = \log_b(m) r^2_{[3]}.$$

for a positive integer m.

We saw that the full dimensional polyhedron $\mathcal{A}_n$ is included in $\mathcal{H}_n$ but we have just seen that outside this polyhedron, $\mathcal{H}_n$ might not be a cone. At this point one might be wondering how much "porous" $\mathcal{H}_n$ is. It turns out that it is not that porous, there are just barely some "scratch" on the boundary. Morepreciselely, the closure of $\mathcal{H}_n$ is a convex cone (Theorem 10) and the relative interior of its closure is entropic (Theorem 11).

Lemma 15. Consider two alphabets $\mathcal{X}_1, \mathcal{X}_2$ with $\mathcal{X}_1 \cap \mathcal{X}_2 = \emptyset$ and two probability mass functions p_1 and p_2 respectively defined on $\mathcal{X}_1$ and $\mathcal{X}_2$. If $0 \leq \alpha \leq 1$, the probability mass function

$$p(x) = \begin{cases} \alpha p_1(x), & x \in \mathcal{X}_1 \\ (1-\alpha)p_2(x), & x \in \mathcal{X}_2 \end{cases}$$

has an entropy equal to

$$H_b(p) = \alpha H_b(p_1) + (1-\alpha)H_b(p_2) + H_b(\alpha).$$

Corollary 3. If $h, g \in \mathcal{H}_n$ then for any $0 \leq \alpha \leq 1$, $\alpha h + (1-\alpha)g + H_b(\alpha)r^n_{[n]} \in \mathcal{H}_n$.

In particular, taking $g = 0$, and using Corollary 2, we have the following corolarry.

Corollary 4 ([Mat07b, Lemma 4])**.** If $h \in \mathcal{H}_n$ then for any $\alpha, \beta > 0$, $\alpha h + \beta r^n_{[n]} \in \mathcal{H}_n$.

And finally, since we can take β arbitrarilly close to zero in the previous corollary.

Corollary 5 ([Mat07b, Corollary 2])**.** If $h \in \mathcal{H}_n$ then for any nonnegative number α, $\alpha h \in \mathrm{cl}(\mathcal{H}_n)$.

Corollary 5 and Lemma 14 implies the following theorem.

Theorem 10. For integer n, $\mathrm{cl}(\mathcal{H}_n)$ is a convex cone.

As mentioned earlier the only difference between $\mathrm{cl}(\mathcal{H}_n)$ and $\mathcal{H}_n$ is on the boundary of $\mathcal{H}_n$.

Theorem 11. For any integer n,

$$\mathrm{ri}(\mathrm{cl}(\mathcal{H}_n)) \subseteq \mathcal{H}_n.$$

This theorem is the consequence of the fact that $\mathcal{H}_n$ contains the full dimensional cone $\mathcal{A}_n$ as shown by the following lemma that is taken from the proof of [Mat07b, Theorem 1].

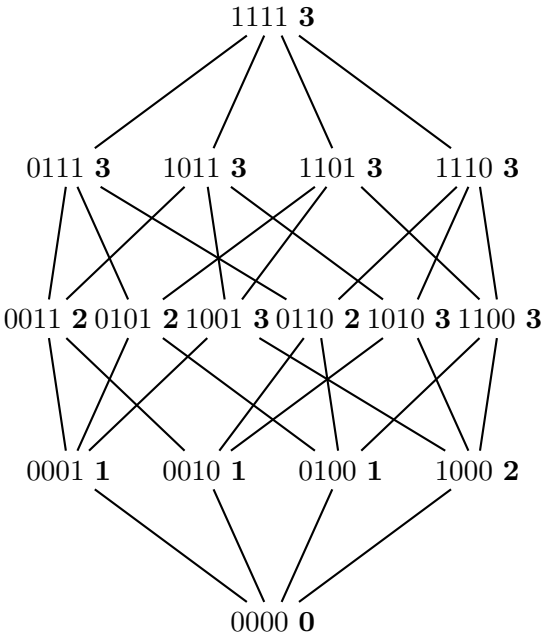
Lemma 16. Consider a set $\mathcal{S}$. If $\mathrm{cl}(\mathcal{S})$ is a convex cone, $\mathcal{S}$ contains a full dimensional cone K and $\mathcal{S}$ is close to addition then

$$\mathrm{ri}(\mathrm{cl}(\mathcal{S})) \subseteq \mathcal{S}.$$

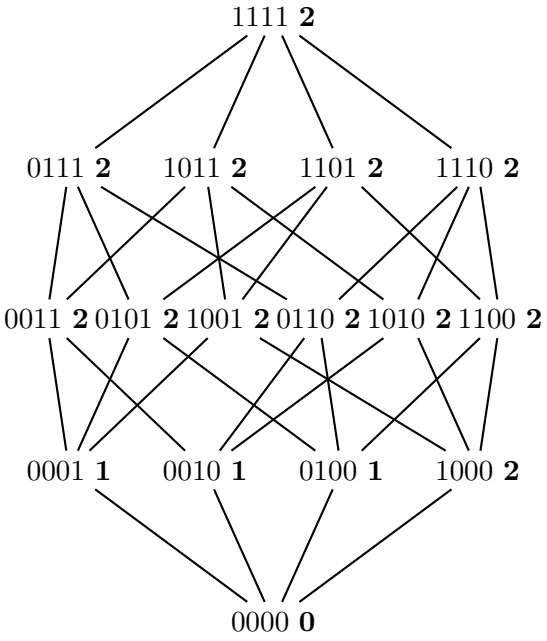
Proof. Suppose $x \in \mathrm{ri}(\mathrm{cl}(\mathcal{S}))$. Since x in in the relative interior, $(x - K) \cap \mathrm{cl}(\mathcal{S})$ should not be empty. Since K is full dimensional, that means that $(x - K) \cap \mathcal{S}$ is also nonempty. Let $y \in (x - K) \cap \mathcal{S}$. Since $\mathcal{S}$ contains K and is closed to addition, $y + K \subseteq \mathcal{S}$ and in particular $x \in \mathcal{S}$. □

1.5.6 The entropic cone of 4 variables

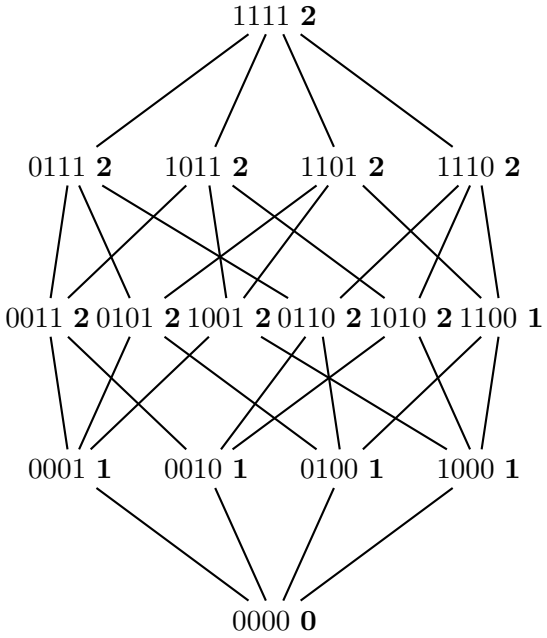
The entropic cone of 4 variables has, in additions to $\mathcal{A}_3$, the extreme rays $r^2_{[3]}, r^2_{[4]}, r^3_{[4]}$ and the three rays given by Figure 1.8 (up to symmetry). The vector $r^2_{[3]}$ is entropic, as seen in Section 1.5.6. The vector $r^3_{[4]}$ is also entropic, it corresponds to X_1, X_2, X_3 being random coin tosses and $X_4 = X_1 \oplus X_2 \oplus X_3$. The vector $r^2_{[4]}$ is entropic too but it is less easy to see.



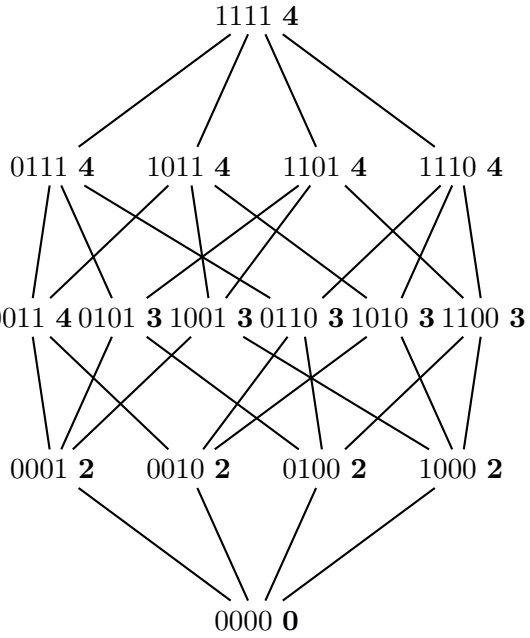
(a) X_1, X_2, X_3 are independent coin tosses and $X_4 = (X_1 \oplus X_3, X_2 \oplus X_3)$.



(b) X_1 and X_2 are independent coin tosses, $X_3 = X_1 \oplus X_2$ and $X_4 = (X_1, X_2)$.



(c) X_1 and X_2 are independent coin tosses and $X_4 = X_3 = X_1 \oplus X_2$.



(d) Non-entropic extreme ray of $\mathcal{P}_4$. Its free expansion is the Vamos matroid [Vam68] (TODO).

Figure 1.8: Extreme rays of $\mathcal{H}_4$ not expressible using the $r^m_{\mathcal{S}}$ notation.

1.6 Sculpting the Entropic Cone

In general, $\mathcal{H}_n \subseteq \mathcal{P}_n$. We have seen that $\mathcal{H}_1 = \mathcal{P}_1, \mathcal{H}_2 = \mathcal{P}_2, \operatorname{cl}\mathcal{H}_3 = \mathcal{P}_3$ but as we will see, $\operatorname{cl}\mathcal{H}_n \subset \mathcal{P}_n$ for $n \geq 4$.

The outer cone $\mathcal{P}_n$ is obtained using Theorem 7 and Theorem 8. As $\mathcal{P}_4$ is a strict superset of $\operatorname{cl}\mathcal{H}_4$, that means that those theorems are not enough. What have we missed ? Do we need a new method to generate new entropic inequalities ? It turns out that we can generate new inequalities using exactly the same method that we used to produce the outer bound $\mathcal{P}_n$.

1.6.1 Generating non-shannon inequalities

The method used to generate $\mathcal{P}_n^*$ is at follows:

1. We consider an entropic vector h and it probability mass function.
2. We apply transformations on the probability mass function.
3. Then use the fact that $x \mapsto -\log(x)$ is nonnegative for $0 \leq x \leq 1$ using Proposition 3 and the fact that $x \mapsto -\log(x)$ is convex using Proposition 4 to derive inequalities between the generated probability mass functions. That is, we enforce $H_p(q) \geq 0$ and $D(p\|q) \geq 0$ between all pairs p, q.
4. Then we look at the implications of those inequalities on the coordinates of the entropy vector h.

We can consider different transformations:

Marginalization For $I \subseteq [n]$, we can transform the probability mass function p to

$$\mathcal{X}_I \to \mathbb{R} : x \mapsto p_I(x_I).$$

Contraction For $I \subseteq J$ and some x_I, we can transform the probability mass function p to

$$\mathcal{X}_{J\setminus I} \to \mathbb{R} : y \mapsto p_{J|I}(y|x_I).$$

Inner-adhesivity For J, K with $I = J \cap K$, we can transform the probability mass function p to

$$\mathsf{ia}_{J,K|I}(p) \colon \mathcal{X}_{J\cup K} \to \mathbb{R} : x \mapsto \frac{p_J(x_J)p_K(x_K)}{p_I(x_I)} = p_{J|I}(x_J|x_I)p_{K|I}(x_K|x_I)p_I(x_I). \tag{1.37}$$

We denote $\mathsf{ia}_{J,K|J\cap K}$ as $\mathsf{ia}_{J,K}$.

Self-adhesivity For $I \subseteq J$, we can transform the probability mass function p to

$$\mathsf{sa}_{J|I}(p) \colon \mathcal{X} \times \mathcal{X}_{J\setminus I} \to \mathbb{R} : x \mapsto \frac{p_{J'}(x_{J'})p_J(x_K)}{p_I(x_I)} = p_{J'|I}(x_{J'}|x_I)p_{J|I}(x_K|x_I)p_I(x_I). \tag{1.38}$$

where $n' = n + |J \setminus I|$, $J' = [n]$ and $K = ([n'] \setminus [n]) \cup I$

One could consider additional types of transformations. For instance, for J, K, L with $I = J \cap K \cap L$, we can transform the probability mass function p to

$$\mathcal{X}_{J\cup K\cup L} \to \mathbb{R} : x \mapsto p_{J|I}(x_J|x_I)p_{K|I}(x_K|x_I)p_{L|I}(x_L|x_I)p_I(x_I).$$

However this transformation can be achieved by doing two inner-adhesivity transformations: one with $(J, K \cup L)$ and then one with $(J \cup K, L)$.

We can see that Theorem 7 was obtained using contraction and then Proposition 3 and Theorem 8 was obtained using inner-adhesivity and then Proposition 4.

If we apply a transformation on a probability mass function that is already a transformation of p, we might not be able to express the resulting probability mass function using p without using summation. For instance, suppose we do an inner-adhesivity transformation using the sets 12 and 13 with a probability mass function p of 3 variables: $q = \mathsf{ia}_{12,13|(}p) = p_{2|1}p_{3|1}p_1$. If we marginalize q on the second variable, it is $q_{13} = p_{13}$ but if we marginalize on the first variable, it is

$$\sum_{x_1 \in \mathcal{X}_1} q(x_1, x_2, x_3) = \sum_{x_1 \in \mathcal{X}_1} p_{2|1}(x_2|x_1)p_{3|1}(x_3|x_1)p_1(x_1)$$

but we cannot simplify it further. This happens when we marginalize on a variable of I after an inner-adhesivity. Suppose q is obtained after applying an inner-adhesivity transformation on p with (J, K) of intersection $I = J \cap K$. Let h (resp. g) be the entropic vector of p (resp. q). Suppose we want to compute g_L in terms of h.

- If $L \cap J = I$ or $L \cap K = I$, then $g_L = h_L$.
- Otherwise, if $I \subseteq L$,

  $$g_L = h_{L\cap J} + h_{L\cap K} - h_I.$$

  Note that the above equation gives $g_L = h_L$ if $L \cap J = I$ or $L \cap K = I$, it was only split in two cases for clarity.
- Otherwise, q_L is obtained from q using marginalization on one of the variables in I and g_L cannot be expressed in terms of h.

Even if some coordinates of g cannot be expressed in terms of g, those coordinates can still be used to derive inequalities on h.

To illustrate this technique, we show in the following example how to prove the Zhang-Yeung inequality.

Example 9. One form of the Zhang-Yeung inequality is as follows:

$$\langle \Delta_{3,4|1} + \Delta_{3,4|2} + \Delta_{1,2} - \Delta_{3,4} + \Delta_{3,4|5} + \Delta_{5,3|4} + \Delta_{4,5|3}, h\rangle \geq 0$$

Let $q = \mathsf{ia}_{5,12|34}p$. Let us abreviate $H(p)$ with p. The Zhang-Yeung inequality is:

$$\begin{aligned}
-p_{134} + \mathsf{ia}_{3,4|1}p - p_{234} + \mathsf{ia}_{3,4|2}p - p_{12} + \mathsf{ia}_{1,2}p - p_{32} \\
+\mathsf{ia}_{3,4}p - p_{345} + \mathsf{ia}_{3,4|5}p - p_{345} + \mathsf{ia}_{5,3|4}p - p_{345} + \mathsf{ia}_{4,5|3}p \geq 0 \\
-p_{134} + p_{13} + p_{14} - p_1 - p_{234} + p_{23} + p_{24} - p_2 - p_{12} + p_1 + p_2 - p_{32} + p_3 + p_4 \\
-p_{345} + p_{35} + p_{45} - p_5 - p_{345} + p_{45} + p_{34} - p_4 - p_{345} + p_{34} + p_{35} - p_3 \geq 0 \\
+p_{13} + p_{35} - p_3 + p_{14} + p_{45} - p_4 \\
+p_{23} + p_{35} - p_3 + p_{24} + p_{45} - p_4 \\
-p_5 - p_{134} - p_{345} + p_{34} - p_{234} - p_{345} + p_{34} - p_{1234} - p_{345} + p_{34} + p_{1234} - p_{12} \geq 0 \\
+p_{13} + p_{35} - p_3 + p_{14} + p_{45} - p_4 \\
+p_{23} + p_{35} - p_3 + p_{24} + p_{45} - p_4 \\
-p_5 - q_{1345} - q_{2345} - q + p_{1234} - p_{12} \geq 0 \\
-q_{135} + p_{13} + p_{35} - p_3 - q_{145} + p_{14} + p_{45} - p_4 \\
-q_{235} + p_{23} + p_{35} - p_3 - q_{245} + p_{24} + p_{45} - p_4 \\
-q_{125} + p^1_{15} + q_{25} - p_5 - q_{1345} + q_{135} + q_{145} - q_{15} \\
-q_{2345} + q_{235} + q_{245} - q_{25} - q + q_{125} + p_{1234} - p_{12} \geq 0 \\
-q_{135} + \mathsf{ia}_{1,5|3}q - q_{145} + \mathsf{ia}_{1,5|4}q - q_{235} + \mathsf{ia}_{2,5|3}q - q_{245} + \mathsf{ia}_{2,5|4}q \\
-q_{125} + \mathsf{ia}_{1,2|5}q - q_{1345} + \mathsf{ia}_{3,4|15}q - q_{2345} + \mathsf{ia}_{3,4|25}q - q + \mathsf{ia}_{34,5|12}q \geq 0 \qquad (1.39) \\
D(q_{135}\|\mathsf{ia}_{1,5|3}q) + D(q_{145}\|\mathsf{ia}_{1,5|4}q) + D(q_{235}\|\mathsf{ia}_{2,5|3}q) + D(q_{245}\|\mathsf{ia}_{2,5|4}q) \\
+D(q_{125}\|\mathsf{ia}_{1,2|5}q) + D(q_{1345}\|\mathsf{ia}_{3,4|15}q) + D(q_{2345}\|\mathsf{ia}_{3,4|25}q) + D(q\|\mathsf{ia}_{34,5|12}q) \geq 0. \qquad (1.40)
\end{aligned}$$

The last inequality is true by Proposition 4 so the Zhang-Yeung inequality is proven.

We used the following probability mass function which are not expressible in terms of p without summation.

$$\begin{aligned}
q_{135} &= \sum_4 p_{5|34}p_{1|34}p_{34} \\
q_{145} &= \sum_3 p_{5|34}p_{1|34}p_{34} \\
q_{235} &= \sum_4 p_{5|34}p_{2|34}p_{34} \\
q_{245} &= \sum_3 p_{5|34}p_{2|34}p_{34} \\
q_{125} &= \sum_{34} p_{5|34}p_{12|34}p_{34} \\
q_{15} &= \sum_{34} p_{5|34}p_{1|34}p_{34} \\
q_{25} &= \sum_{34} p_{5|34}p_{2|34}p_{34}.
\end{aligned}$$

The Zhang-Yeung inequality is a *non-Shannon inequality* as it is not a consequence of the inequalities introduced in Section 1.5.3. It was discovered by Zhang and Yeung in 1998 [ZY98]. The existence of this non-Shannon inequality shows that the inclusion $\operatorname{cl}\mathcal{H}_n \subset \mathcal{P}_n$ is strict for $n \geq 4$.

Matúš and Csirmaz [MC13] recently introduced a way to visualize the projection of a cut of approximations of the Entropic Cone of 4 random variables in 3 dimensions. In this visualization, the polymatroid cone $\mathcal{P}_4$ is the tetrahedron shown in Figure 1.9b.



(a) Polymatroid cone $\mathcal{P}_4$.

(b) The polymatroid cone $\mathcal{P}_4$ cut by the Zhang-Yeung inequality.

Figure 1.9: 3D visualization of approximations of the Entropic Cone of 4 random variables.

1.6.2 Formalizing the generation of non-shannon inequalities

In the previous section, we have seen how to generate new inequalities in terms of transformation on the probability mass function. In practice, instead of looking at all those transformations together and all the possible succession of those transformations, a specific transformation is chosen and we only keep its implication on the entropy space. This may look conservative to only look at it from the entropy space but it makes it more easy to reason about and to compute non-shannon inequalities.

The following theorems translate the invariance over probability mass functions to the entropy space.

The contraction of a polymatroid is defined similarly to the contraction on matroid initially studied by Tutte [Tut58a, Tut58b]; see also [Wel10, Section 4.3] or [Oxl06, Section 3].
Definition 10 (Contraction)**.** The contraction of a polymatroid h onto $J \subset [n]$, denoted by $h \cdot J$ is defined as

$$(h \cdot J)_K = h_{K \cap I} - h_I, \qquad K \subseteq J$$

where $I = [n] \setminus J$. It is also called the contraction of I from h and denoted h/I.

The following theorem is a consequence of Proposition 5 and Theorem 10.
Theorem 12 ([MC13, Lemma 1])**.** The family of cones $\operatorname{cl} \mathcal{H}_n$ is closed to contraction.

We define the following operators

$$\begin{aligned} \mathsf{ia}_{J,K|I}(h) &= \{\, g \in \mathcal{E}_{n'} \mid \forall I \subseteq L \subseteq J \cap K, g_L = h_{L \cap J} + h_{L \cap K} - h_I \,\}, & I = J \cap K, \\ \mathsf{sa}_{J|I}(h) &= \{\, g \in \mathcal{E}_{n'} \mid \forall I \subseteq L \subseteq J' \cap K, g_L = h_{L \cap J'} + h_{L \cap K} - h_I \,\}, & I \subseteq J, \end{aligned}$$

where for inner-adhesivity, $n' = |J \cup K|$ and for self-adhesivity $n' = n + |J \setminus I|$, $J' = [n]$ and $K = ([n'] \setminus [n]) \cup I$.

Again, we denote $\mathsf{ia}_{J,K|J\cap K}$ as $\mathsf{ia}_{J,K}$.

Definition 11. We say that a family of sets $\mathcal{S}_n \subseteq \mathcal{E}_n$ is *inner-adhesive* if for any n, $x \in \mathcal{S}_n$ and $J, K \subseteq [n]$, there exists $y \in \mathcal{S}_{|J \cup K|}$ such that $y \in \mathsf{ia}_{J,K}(x)$ and we say that it is *self-adhesive* if for any n, $x \in \mathcal{S}_n$ and $I \subseteq J \subseteq [n]$, there exists $y \in \mathcal{S}_{n+|J \setminus I|}$ such that $y \in \mathsf{sa}_{J|I}(x)$.

The following theorem is a consequence of equation (1.36), and the transformation (1.37) and (1.38).

Theorem 13. The cone $\mathcal{H}_n$ is inner-adhesive and self-adhesive.

An important numerical quantity, related with the geometric properties of $\mathcal{H}_4$ is the *Ingleton score* defined as

$$\mathbb{I}^* \triangleq \inf_{0 \neq h \in \mathcal{H}_4} \mathbb{I}_{ij}(h)$$

where $\mathbb{I}_{ij}(h) = \langle \square_{ij}, h \rangle / h_{[n]}$ [DFZ11, Definition 3] and $\square_{ij}$ is the Ingleton dual entropy vector [Ing71]. The Ingleton score of an approximation of the entropic cone of four variables is proportional to the "height" of the 3D visualization we introduced. The Ingleton score of $\mathcal{P}_4$ represented in Figure 1.9a is $-\frac{1}{4}$ and the Ingleton score of the approximation represented in Figure 1.9b is $-\frac{1}{6}$.

The current best lower bound on $\mathbb{I}^*$ is equal to -0.15789 [DFZ11]. Upper bounds on $\mathbb{I}^*$ can be obtained from exhibiting four jointly distributed variables for which the entropy vector has low Ingleton score. The current best upper bound on $\mathbb{I}^*$ is equal to -0.09243 [MC13].

The current best lower and upper bound are given respectively by the outer and inner approximation of $\mathcal{P}_4$ illustrated by Figure 1.10.

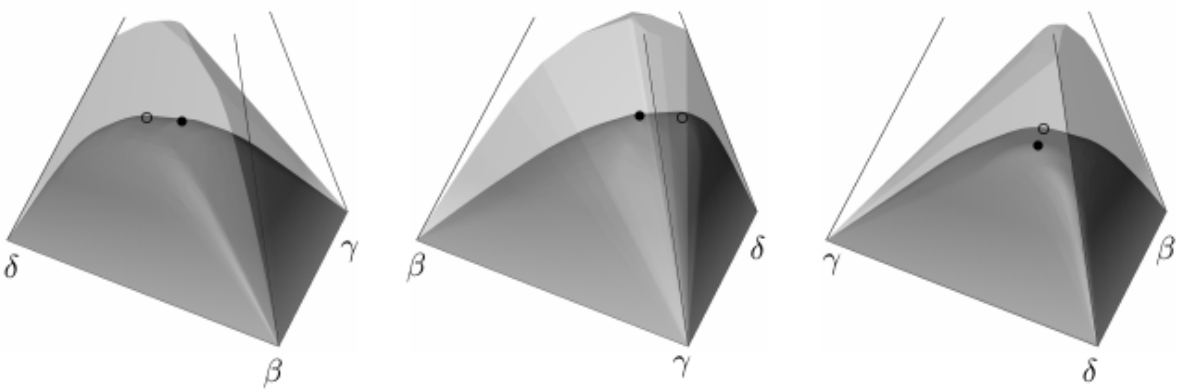


Figure 1.10: Outer [DFZ11] and inner [MC13] approximations of $\mathcal{P}_4$. The figures comes from [MC13].

Chapter 2

Generating unstable trajectories for switched systems

In recent years, the study of the stability of hybrid systems has been the subject of extensive research using methods based on classical ideas from Lyapunov theory and modern mathematical optimization techniques. Even for switched linear systems, arguably the simplest class of hybrid systems, determining stability is undecidable and approximating the maximal asymptotic growth rate is NP-hard [BT00]. Despite these negative results, the vast range of applications has motivated a wealth of algorithms to approximate this quantity.

The JSR was introduced by Rota and Strang [RS60] and has many other applications such as wavelets, the capacity of some particular codes, zero-order stability of ordinary differential equations, congestion control in computer networks, curve design and networked and delayed control systems; see [Jun09] for a survey on the JSR and its applications. Many algorithms exist for estimating the JSR but not much is known on how to generate an infinite sequence of matrices with an asymptotic growth rate close to the JSR. However generating such sequence can be of particular interest, depending on the application, such as exhibiting unstable trajectories for switched linear systems. The currently known algorithms select a small sequence of matrices with high spectral radius using brute force (or branch-and-bound variants) and repeat this sequence infinitely [Gri96, GZ08, JCG14].

Parrilo and Jadbabaie [PJ08] give an efficient approximation algorithm for computing the JSR using sum of squares (SOS) programming. In this chapter, we compute the dual of the corresponding SOS program and give an algorithm that generates an infinite sequence of matrices with an asymptotic growth rate arbitrarily close to the JSR. Unlike the brute force approach, we provide a guarantee on the quality of the asymptotic growth rate. As a by-product of the algorithm, the spectral radius of a finite part of this infinite sequence can be used to give lower bounds on the JSR. We show on numerical examples that the technique works well in practice.

We provide also a new estimate of the accuracy of the SOS-based approximation algorithm for the CJSR which is better than the previously existing one for sufficiently large SOS degree. The existing estimate only depends on the dimension of the matrices while our new one relates the accuracy of the SOS-based approximation algorithm with the combinatorial structure of the automaton representing the constraints.

We start the chapter by showing that the computation of the JSR of low-rank matrices can be reduced to the computation of the CJSR of matrices of dimension equal to the rank of the matrices.

2.1 Low rank reduction

Suppose we want to compute the CJSR of a finite set of matrices $\mathcal{M} \triangleq \{A_1, \dots, A_m\} \subset \mathbb{R}^{n\times n}$ of rank at most r constrained by an automaton $G(V, E)$. For $i = 1, \dots, m$, since the matrix A_i has rank at most r, there exists $X_i, Y_i \in \mathbb{R}^{n\times r}$ such that $A_i = X_i Y_i^T$. Let $A'_{ij} = Y_i^T X_j$ for $i, j = 1, \dots, m$. We have

$$A_{\sigma_k} \cdots A_{\sigma_2} A_{\sigma_1} = X_{\sigma_k} A'_{\sigma_k\sigma_{k-1}} \cdots A'_{\sigma_3\sigma_2} A'_{\sigma_2\sigma_1} Y_{\sigma_1}^T. \tag{2.1}$$

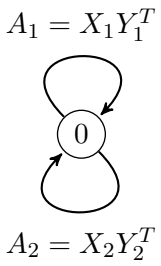
For $r = 1$, the factors $A'_{\sigma_i\sigma_{i+1}}$ in the RHS are scalars which commute with the matrices X_{σ_k} and $Y_{\sigma_1}^T$ so we can isolate the matrix $X_{\sigma_k} Y_{\sigma_1}^T$ whose spectral radius is $A'_{\sigma_1\sigma_k}$. Hence the spectral radius of the RHS is $A'_{\sigma_k\sigma_{k-1}} \cdots A'_{\sigma_2\sigma_1} A'_{\sigma_1\sigma_k}$, this enables Ahmadi and Parrilo [AP12b] to reformulate the computation of the CJSR of rank 1 matrices in terms of the maximum cycle mean problem.

For a general $r \geq 1$, we build instead an automaton G' for which $(\sigma_2\sigma_1, \sigma_4\sigma_3, \dots, \sigma_{2k}\sigma_{2k-1})$ is G'-admissible if $\sigma_{2k} = \sigma_{2k+1}$ for $k = 1, \dots, k-1$ and $(\sigma_1, \sigma_3, \dots, \sigma_{2k-1}, \sigma_{2k})$ is G-admissible. See Figure 2.1 for a simple example. For this automaton G' we show that the CJSR $\rho(G', \mathcal{M}')$ is equal to the original CJSR $\rho(G, \mathcal{M})$.

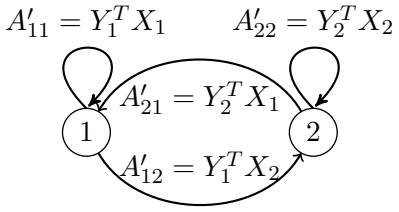
To prove that $\rho(G', \mathcal{M}') = \rho(G, \mathcal{M})$, it could be tempting to use (2.1) and claim that there are constants $\alpha, \beta > 0$ such that

$$\alpha \|A'_{\sigma_k\sigma_{k-1}} \cdots A'_{\sigma_2\sigma_1}\| \leq \|X_{\sigma_k} A'_{\sigma_k\sigma_{k-1}} \cdots A'_{\sigma_2\sigma_1} Y_{\sigma_1}^T\| \leq \beta \|A'_{\sigma_k\sigma_{k-1}} \cdots A'_{\sigma_2\sigma_1}\|.$$

However, the first inequality may not hold if X_{σ_k} or $Y_{\sigma_1}^T$ have a rank strictly smaller than r. For this reason we use twice the right inequality, see the proof of Theorem 14.



(a) Automaton G. We have $V = \{0\}$ and $E = \{(0, 0, 1), (0, 0, 2)\}$.



(b) Automaton G'. We have $V' = \{1, 2\}$ and $E' = \{(1, 1, 11), (1, 2, 12), (2, 1, 21), (2, 2, 22)\}$.

Figure 2.1: Simple example of the low rank reduction.

Theorem 14 (Low Rank Reduction)**.** Consider a finite set of matrices $\mathcal{M} \triangleq \{A_1, \dots, A_m\} \subset \mathbb{R}^{n\times n}$ of rank at most r constrained by an automaton $G(V, E)$.

For a fixed decomposition $A_\sigma = X_\sigma Y_\sigma^T$ for $\sigma = 1, \dots, m$, denote the set of matrices $\mathcal{M}' \triangleq \{A'_{\sigma_1\sigma_2} \mid \sigma_1, \sigma_2 = 1, \dots, m\} \subset \mathbb{R}^{r\times r}$ where $A'_{\sigma_1\sigma_2} = Y_{\sigma_1}^T X_{\sigma_2}$. Define the graph $G'(V', E')$ with $V' \triangleq E$ and

$$E' \triangleq \{\, ((u, v, \sigma_1), (v, w, \sigma_2), \sigma_2\sigma_1) \mid (u, v, \sigma_1), (v, w, \sigma_2) \in E \,\}.$$

Then the two CJSR are the same:

$$\rho(G, \mathcal{M}) = \rho(G', \mathcal{M}').$$

Proof. As the CJSR does not depend on the norm used, we choose a norm $\|\cdot\|$ that is *submultiplicative*, that is $\|AB\| \leq \|A\|\|B\|$ for all matrices A, B. For example, any norm induced by a vector norm is submultiplicative.

Let $\beta = \max_{i=1}^{m} \max\{\|X_i\|, \|Y_i^T\|\}$. If $\beta = 0$, then $\rho(G, \mathcal{M}) = 0 = \rho(G', \mathcal{M}')$. Therefore we may assume that $\beta > 0$. Consider a positive integer k. We first show that $[\hat{\rho}_k(G, \mathcal{M})]^k \leq \beta^2[\hat{\rho}_{k-1}(G', \mathcal{M}')]^{k-1}$. For any G-admissible $(\sigma_1, \ldots, \sigma_k)$, using (2.1) and the submultiplicativity of the norm chosen, we have

$$\begin{aligned}\|A_{\sigma_k} \cdots A_{\sigma_2} A_{\sigma_1}\| &\leq \|X_{\sigma_k}\| \|A'_{\sigma_k\sigma_{k-1}} \cdots A'_{\sigma_3\sigma_2} A'_{\sigma_2\sigma_1}\| \|Y_{\sigma_1}^T\| \\ &\leq \beta^2 \|A'_{\sigma_k\sigma_{k-1}} \cdots A'_{\sigma_3\sigma_2} A'_{\sigma_2\sigma_1}\| \\ &\leq \beta^2 [\hat{\rho}_{k-1}(G', \mathcal{M}')]^{k-1}.\end{aligned}$$

The same way, we now show that $[\hat{\rho}_{k-1}(G', \mathcal{M}')]^{k-1} \leq \beta^2[\hat{\rho}_{k-2}(G, \mathcal{M})]^{k-2}$. For any G'-admissible $(\sigma_2\sigma_1, \ldots, \sigma_k\sigma_{k-1})$,

$$\begin{aligned}\|A'_{\sigma_k\sigma_{k-1}} \cdots A'_{\sigma_3\sigma_2} A'_{\sigma_2\sigma_1}\| &\leq \|Y_k^T\| \|A_{\sigma_{k-1}} \cdots A_{\sigma_2}\| \|X_1\| \\ &\leq \beta^2 [\hat{\rho}_{k-2}(G, \mathcal{M})]^{k-2}.\end{aligned}$$

In summary, we have

$$\hat{\rho}_k(G, \mathcal{M}) \leq \beta^{\frac{2}{k}} [\hat{\rho}_{k-1}(G', \mathcal{M}')]^{\frac{k-1}{k}} \leq \beta^{\frac{4}{k}} [\hat{\rho}_{k-2}(G, \mathcal{M})]^{\frac{k-2}{k}}.$$

Taking the limit of k to ∞ we get $\rho(G, \mathcal{M}) \leq \rho(G', \mathcal{M}') \leq \rho(G, \mathcal{M})$. □

Remark 2. The matrices X_σ, Y_σ of the factorization $A_\sigma = X_\sigma Y_\sigma^T$ are not unique. For any invertible matrix $S \in \mathbb{R}^{r\times r}$, $A_\sigma = (X_\sigma S)(S^{-1}Y_\sigma^T)$ also gives a factorization. However, if $\rho(G', \mathcal{M}')$ is approximated using the sum of squares algorithm of Section 1.3.1, any two factorizations will give the same approximation. The effect of using $X_\sigma S$ and $Y_\sigma S^{-T}$ instead of X_σ and Y_σ will simply be a linear change of variable of the polynomial p_σ; see Section 1.3.1.

If the rank of A_σ is strictly less than r, we have the choice of choosing X_σ and Y_σ with the same rank of A_σ or choosing X_σ with the same rank of A_σ and Y_σ full rank for example. For this choice, the question remains open.

We now discuss the reduction quantitatively. Suppose first that we want to approximate the unconstrained JSR of m matrices of dimension n and rank at most r. Using Theorem 14, we see that the unconstrained JSR is equal to the CJSR of m^2 matrices of rank r constrained by an automaton with m nodes and m^2 edges, the underlying directed graph of the automaton is the complete graph with a loop at every node. This is summarized in Table 2.1a.

Suppose now that we want to approximate the CJSR of m matrices of dimension n and rank at most r constrained by an automaton of $|V|$ nodes and $|E|$ edges. Using Theorem 14, we see that this CJSR is equal to the CJSR of matrices of dimension r. The number of matrices in this CJSR might be lower than m^2 because if (σ_1, σ_2) is not G-admissible, then we do not need $A'_{\sigma_2\sigma_1}$. The size of each important quantity before and after the reduction is given by Table 2.1b. We have just explained why $|\mathcal{M}'| \leq m^2$. We can see that by definition of E', $|E'| \leq |E|^2$ with equality if and only if $|V| = 1$, that is the unconstrained case. Intuitively, the "more" the original CJSR is constrained, the "sparser" G' will be.

2.2 The p-radius

To give the full picture, we include the p-radius in our analysis. The p-radius is a generalization of the joint spectral radius. In this work, we extend its definition to the constrained case.

Quantity	$\mathcal{M}$	$(G', \mathcal{M}')$
dimension	n	r
rank	r	r
matrices	m	m
nodes		m
edges		m^2

(a) Unconstrained case.

Quantity	$(G, \mathcal{M})$	$(G', \mathcal{M}')$
dimension	n	r
rank	r	r
matrices	m	$\leq m^2$
nodes	$\lvert V\rvert$	$\lvert E\rvert$
edges	$\lvert E\rvert$	$\leq \lvert E\rvert^2$

(b) Constrained case.

Table 2.1: Quantification of the low rank reduction of Theorem 14. The quantity of the CJSR $(G', \mathcal{M}')$ obtained by the reduction are expressed as a function of the quantity of the original CJSR $(G, \mathcal{M})$.

Definition 12 (Constrained p-radius)**.** The *constrained p-radius* of a finite set of matrices $\mathcal{M}$ constrained by an automaton $G(V, E)$, denoted as $\rho_p(G, \mathcal{M})$, is

$$\rho_p(G, \mathcal{M}) = \lim_{k\to\infty} \left[|E_k|^{-1} \sum_{v\in V} \hat{\rho}_{p;k;v}(G, \mathcal{M}) \right]^{\frac{1}{pk}}$$

where

$$\hat{\rho}_{p;k;v}(G, \mathcal{M}) = \sum_{s\in E_k^+(v)} \|A_s\|^p.$$

Thus, the CJSR can be defined as the constrained p-radius for $p = \infty$.

Remark 3. Since G is assumed to be strongly connected, we could give the following equivalent definition

$$\hat{\rho}_{p;k}(G, \mathcal{M}) = \lim_{k\to\infty} \left[\max_{v\in V} [d_k^+(v)]^{-1} \hat{\rho}_{p;k;v}(G, \mathcal{M}) \right]^{\frac{1}{pk}} \tag{2.2}$$

or the same definition with "$d_k^-(v)$" instead of "$d_k^+(v)$" and "$s(k) = v$" instead of "$s(1) = v$".

By the equivalence of norms, the definition of the p-radius does not depend on the norm used.

We can show that the p-radius is well defined using the following classical result, known as *Fekete's Lemma.*

Lemma 17 ([Fek23])**.** Let $\{a_n\} : n \geq 1$ be a sequence of real numbers such that

$$a_{m+n} \leq a_m + a_n.$$

Then the limit

$$\lim_{n\to\infty} \frac{a_n}{n}$$

exists and is equal to $\inf\{\frac{a_n}{n}\}$.

Lemma 18. Consider a finite set of matrices $\mathcal{M}$ constrained by an automaton $G(V, E)$ and the sequence $(a_k)_k = \max_{v\in V} \hat{\rho}_{p;k;v}(G, \mathcal{M})$ with a submultiplicative norm. The sequence $\sqrt[k]{a_k}$ converges when $k \to \infty$. Moreover,

$$\lim_{k\to\infty} \sqrt[k]{a_k} = \inf\{\sqrt[k]{a_k}\}.$$

Proof. By submultiplicativity, for any $v \in V$, k and any $k_1, k_2 \geq 0$ such that $k_1 + k_2 = k$,

$$\begin{aligned}
\hat{\rho}_{p;k;v}(G, \mathcal{M}) &= \sum_{u \in V} \sum_{s_1 \in E_{k_1}(v,u), s_2 \in E^+_{k_2}(u)} \|A_{s_2} A_{s_1}\|^p \\
&\leq \sum_{u \in V} \sum_{s_1 \in E_{k_1}(v,u), s_2 \in E^+_{k_2}(u)} \|A_{s_2}\|^p \|A_{s_1}\|^p \\
&= \sum_{u \in V} \hat{\rho}_{p;k_2;u}(G, \mathcal{M}) \sum_{s_1 \in E^-_{k_1}(u), s_1(1) = v} \|A_{s_1}\|^p \\
&\leq a_{k_2} \sum_{u \in V} \sum_{s_1 \in E^-_{k_1}(u), s_1(1) = v} \|A_{s_1}\|^p \\
&\leq \hat{\rho}_{p;k_1;v}(G, \mathcal{M}) a_{k_2}
\end{aligned}$$

hence, in particular, $a_k \leq a_{k_1} a_{k_2}$ and $\log a_k \leq \log a_{k_1} + \log a_{k_2}$. We can conclude by Lemma 17. □

Corollary 6. The following holds

$$\lim_{k \to \infty} \left[\max_{v \in V} \hat{\rho}_{p;k;v}(G, \mathcal{M}) \right]^{\frac{1}{k}} = \lim_{k \to \infty} \left[\sum_{v \in V} \hat{\rho}_{p;k;v}(G, \mathcal{M}) \right]^{\frac{1}{k}}$$

and, in particular, the limit on the right-hand side converges.

Proof. For a finite set of nonnegative numbers, their maximum is always between their average and their sum:

$$\frac{1}{|V|} \sum_{v \in V} \hat{\rho}_{p;k;v}(G, \mathcal{M}) \leq \max_{v \in V} \hat{\rho}_{p;k;v}(G, \mathcal{M}) \leq \sum_{v \in V} \hat{\rho}_{p;k;v}(G, \mathcal{M})$$

or equivalently

$$\max_{v \in V} \hat{\rho}_{p;k;v}(G, \mathcal{M}) \leq \sum_{v \in V} \hat{\rho}_{p;k;v}(G, \mathcal{M}) \leq |V| \max_{v \in V} \hat{\rho}_{p;k;v}(G, \mathcal{M}).$$

By Lemma 18, $\max_{v \in V} \hat{\rho}_{p;k;v}(G, \mathcal{M})$ converges for $k \to \infty$ hence $\sum_{v \in V} \hat{\rho}_{p;k;v}(G, \mathcal{M})$ converges too. Taking the kth root and the limit $k \to \infty$ gives the identity. □

Lemma 19. Consider a finite set of matrices $\mathcal{M}$ constrained by an automaton G. The following relation holds

$$\rho_p(G, \mathcal{M}) = \rho(A(G)) \lim_{k \to \infty} \left[\sum_{s \in E_k} \|A_s\|^p \right]^{\frac{1}{pk}}.$$

Proof. By Corollary 6,

$$\lim_{k \to \infty} \left[\sum_{s \in E_k} \|A_s\|^p \right]^{\frac{1}{pk}}$$

converges. It remains to show that $\lim_{k \to \infty} |E_k|^{-\frac{1}{pk}}$ converges to $\rho(A(G))$.

Consider the matrix norm $\| \cdot \|_\infty$ on $\mathbb{R}^{n \times n}$ induced by the infinity norm on $\mathbb{R}^n$. It is well known that

$$\|A\|_\infty = \max_{1 \leq i \leq n} \sum_{j=1}^{n} |a_{ij}|$$

where a_{ij} is (i,j) entry of A. It is also well known that the (u,v) entry of $A(G)^k$ gives the number of paths of length k starting at node u and ending at node v in G. Hence $\|A(G)^k\|_\infty = \Delta_k^+(G)$. By Gelfand's formula,

$$\rho(A(G)) = \lim_{k\to\infty} \|A(G)^k\|_\infty^{1/k}. \tag{2.3}$$

Since $|E_k|/|V| \le \Delta_k^+(G) \le |E_k|$, and $|V|^{1/k} \to 1$ as $k \to \infty$, we are done. □

In the unconstrained case we know that the p-radius is increasing in p [Zho02]. We can generalize this fact as follows.

Lemma 20. Consider a finite set of matrices $\mathcal{M}$ constrained by an automaton G. For any integers $p \le q$,

$$\rho_p(G,\mathcal{M}) \le \rho_q(G,\mathcal{M}) \le \rho(G,\mathcal{M}) \le \rho(A(G))^{\frac{1}{q}}\rho_q(G,\mathcal{M}) \le \rho(A(G))^{\frac{1}{p}}\rho_p(G,\mathcal{M}).$$

This is a simple consequence of the inequality between ordinary means and the inequality between the p-norms ($p \ge 1$)

$$\|x\|_p = \Big(\sum_{i=1}^n |x_i|^p\Big)^{\frac{1}{p}}.$$

Lemma 21 ([HLP52])**.** For any nonnegative integers $a_1,\dots,a_n$ and positive real numbers $p \le q$,

$$\left(\frac{1}{n}\sum_{i=1}^n a_i^p\right)^{\frac{1}{p}} \le \left(\frac{1}{n}\sum_{i=1}^n a_i^q\right)^{\frac{1}{q}} \le \max\{\, a_i \mid i = 1,\dots,n\,\}.$$

Lemma 22. For any real numbers $1 \le p \le q$,

$$\|x\|_\infty \le \|x\|_q \le \|x\|_p.$$

Since $\|Ax\|^{[d]} = \|A\|^{[d]}\|x\|^{[d]}$, we have the following Lemma that is known in the unconstrained case or for the contrained case with $p = \infty$.

Lemma 23. Consider a finite set of matrices $\mathcal{M}$ constrained by an automaton G, then

$$\rho_p(G,\mathcal{M}) = \rho_1(G,\mathcal{M}^{[p]})^{\frac{1}{p}}$$

and

$$\rho(G,\mathcal{M}) = \rho(G,\mathcal{M}^{[p]})^{\frac{1}{p}}.$$

We say that a cone $\mathcal{K}$ is *proper* if it is closed, solid, convex and pointed. We say that a matrix A *leaves a set* S *invariant* if $AS \subseteq S$ and we say that a set of matrices $\mathcal{M}$ *leaves a proper cone invariant* if there exists a proper cone $\mathcal{K}$ such that each matrix of $\mathcal{M}$ leaves $\mathcal{K}$ invariant.

Lemma 24 ([BN05, Pro97])**.** If a set of m matrices leaves a proper cone $\mathcal{K}$ invariant, then

$$\begin{aligned}\rho_1(\mathcal{M}) &= \frac{1}{m}\lim_{k\to\infty}\left\|\sum_{s\in[m]^k} A_s\right\|^{\frac{1}{k}} \\ &= \frac{1}{m}\rho\left(\sum_{A\in\mathcal{M}} A\right).\end{aligned}$$

We deduce the following corolary of Lemma 23 and Lemma 24.

Corollary 7. If $\mathcal{M}^{[p]}$ leaves a proper cone $\mathcal{K}$ invariant, then

$$\begin{aligned}\rho_p(\mathcal{M}) &= \frac{1}{m}\lim_{k\to\infty}\left\|\sum_{s\in[m]^k} A_s^{[p]}\right\|^{\frac{1}{pk}} \\ &= \frac{1}{m}\rho\left(\sum_{A\in\mathcal{M}} A^{[p]}\right)^{\frac{1}{p}}.\end{aligned}$$

We generalize it to the constrained case using the lifting procedure introduced independantly by Kozyakin [Koz14] and Wang [WRDV14].

Lemma 25. Consider a finite set of matrices $\mathcal{M}$ constrained by an automaton $G(V,E)$. The following identity holds for any $p\in[1,+\infty]$

$$\rho_p(G,\mathcal{M}) = \rho_p(\mathcal{M}')$$

where

$$\mathcal{M}' = \{\,(e_v e_u^\top)\otimes A_\sigma \mid (u,v,\sigma)\in E\,\}.$$

Proof. Consider a vector norm $\|\cdot\|$ of $\mathbb{R}^n$ and the vector norm $\|\cdot\|'$ of $\mathbb{R}^{n|V|}$ such that

$$\|e_1\otimes x_1 + \cdots + e_{|V|}\otimes x_{|V|}\| = \|x_1\| + \cdots + \|x_{|V|}\|.$$

Consider the induced matrix norms $\|\cdot\|$ and $\|\cdot\|'$. It is easy to see that for any nodes $u,v\in V$ and any matrix $B\in\mathbb{R}^{n\times n}$, $\|(e_v e_u^\top)\otimes B\|' = \|B\|$. In particular, given a path $s\in E_k$,

$$\begin{aligned}\|A_s\| &= \|(e_{s(k+1)}e_{s(1)}^\top)\otimes A_s\|' \\ &= \left\|\prod_{i=1}^{k}(e_{s(i+1)}e_i^\top)\otimes A_{s[i]}\right\|'.\end{aligned}$$

□

It is easy to see that if $\mathcal{M}$ leaves the proper cone $\mathcal{K}$ invariant then the set of matrices $\mathcal{M}'$ of Lemma 25 leaves the proper cone $\mathcal{K}^{|V|}$ invariant.

Lemma 26. Consider a finite set of matrices $\mathcal{M}$ constrained by an automaton $G(V,E)$. If $\mathcal{M}$ leaves a proper cone invariant, then

$$\begin{aligned}\rho_1(G,\mathcal{M}) &= \frac{1}{\rho(A(G))}\lim_{k\to\infty}\left\|\sum_{s\in E_k}(e_{s(k+1)}e_{s(1)}^\top)\otimes A_s\right\|^{\frac{1}{k}} \\ &= \frac{1}{\rho(A(G))}\rho\left(\sum_{(u,v,\sigma)\in E}(e_v e_u^\top)\otimes A_\sigma\right).\end{aligned}$$

Theorem 15. Consider a finite set of matrices $\mathcal{M}$ constrained by an automaton G. If $\mathcal{M}^{[p]}$ leaves a proper cone invariant, the following identities hold

$$\begin{aligned}\rho_p(G,\mathcal{M}) &= \frac{1}{\rho(A(G))}\lim_{k\to\infty}\left\|\sum_{s\in E_k}(e_{s(k+1)}e_{s(1)}^\top)\otimes A_s^{[p]}\right\|^{\frac{1}{pk}} \\ &= \frac{1}{\rho(A(G))}\rho\left(\sum_{(u,v,\sigma)\in E}(e_v e_u^\top)\otimes A_\sigma^{[p]}\right)^{\frac{1}{p}}.\end{aligned}$$

Theorem 15 shows that when there is an invariant proper cone, $\rho_p(G, \mathcal{M})$ is as easy to obtain as computing a spectral radius.

It turns out that if p is even then there exists an invariant proper cone.

Lemma 27. Consider a finite set of matrices $\mathcal{M}$ constrained by an automaton G. For any positive integer d, $\mathcal{M}^{[2d]}$ leaves an invariant proper cone. Moreover this cones is the cone of SOS polynomials in the scaled monomial basis.

Proof. Consider an homogeneous SOS polynomial $p(x)$ of degree $2d$ and its coordinates p in the scaled monomial basis. That is, $p(x) = \langle p, x^{[2d]} \rangle$. For any matrix A, we have

$$\langle A^{[2d]} p, x^{[2d]} \rangle = \langle p, (A^{[2d]})^\top x^{[2d]} \rangle = \langle p, (A^\top x)^{[2d]} \rangle = p(A^\top x).$$

Therefore if p is the coordinate vector of an SOS polynomial then $A^{[2d]}p$ is also the coordinate vector of an SOS polynomial. □

2.3 Approximation guarantees

Philippe et al. show, based on [PJ08], the following accuracy guarantees for Program 1.

Theorem 16 ([PEDJ15, Theorem 3.6])**.** Consider a finite set of matrices $\mathcal{M} \subset \mathbb{R}^{n \times n}$ constrained by an automaton G and a positive integer d. The approximation given by Program 1 using homogeneous polynomials of degree $2d$ satisfies:

$$\rho_{\mathrm{SOS},2d}(G, \mathcal{M}) \leq \binom{n+d-1}{d}^{\frac{1}{2d}} \rho(G, \mathcal{M}).$$

This guarantee only depends on the dimension of the matrices and neither on the number of matrices nor on the automaton. In this section, we show the following guarantee that relates the accuracy of Program 1 to the evolution of the number of paths in G for increasing length.

Theorem 17. Consider a finite set of matrices $\mathcal{M}$ constrained by an automaton G and a positive integer d. The approximation given by Program 1 using homogeneous polynomials of degree $2d$ satisfies:

$$\rho_{\mathrm{SOS},2d}(G, \mathcal{M}) \leq \rho(A(G))^{\frac{1}{2d}} \rho_{2d}(G, \mathcal{M})$$

where $A(G)$ is the adjacency matrix of G.

As a corollary of Theorem 17, in the trivial cases such that $\rho(A(G)) = 1$, the approximation is exact. This corresponds to the case where every node of G has indegree and outdegree 1. In that case, the graph forms a cycle of some length k and the CJSR is simply the kth root of the spectral radius of the product of the matrices along this cycle.

The results of Theorem 16, Theorem 17, Lemma 20 and (1.15) are summarized by the following corollary.

Corollary 8. Consider a finite set of matrices $\mathcal{M} \subset \mathbb{R}^{n \times n}$ constrained by an automaton G and a positive integer d, the approximation given by Program 1 using homogeneous polynomials of degree $2d$ satisfies:

$$\min\left\{\binom{n+d-1}{d}, \rho(A(G))\right\}^{-\frac{1}{2d}} \rho_{\mathrm{SOS},2d}(G, \mathcal{M}) \leq \rho(G, \mathcal{M}) \leq \rho_{\mathrm{SOS},2d}(G, \mathcal{M}).$$

where $A(G)$ is the adjacency matrix of the automaton G.

We see that we can have arbitrary accuracy by increasing d.

For the arbitrary switching case, $\rho(A(G)) = m$. Theorem 17 was already known in this particular case [PJ08, Theorem 4.3]. This section can thus be seen as a generalization of [PJ08, Section 4].

Our proof technique relies on the analysis of an iteration in the vector space of polynomials of degree $2d$. When this iteration converges, it converges to a feasible solution of Program 1. By analysing this iteration as affine iterations in this vector space, we derive a sufficient condition for its convergence and thus an upper bound for $\rho_{\mathrm{SOS},2d}(G, \mathcal{M})$.

Consider the iteration

$$\begin{aligned} p_{v,0}(x) &= 0, \\ p_{v,k+1}(x) &= q_v(x) + \frac{1}{\tau} \sum_{(u,v,\sigma)\in E} p_{u,k}(A_\sigma x), \quad v \in V \end{aligned} \tag{2.4}$$

for fixed homogeneous polynomial $q_v(x)$ of degree $2d$ in n variables (not necessarily different) and a constant $\tau > 0$.

When this iteration converges, it converges to a feasible solution of Program 1.

Theorem 18. Consider a constant $\tau > 0$. If there exists homogeneous polynomials $q_v(x)$ in the interior of the SOS cone such that the iteration (2.4) converges then

$$\rho_{\mathrm{SOS},2d}(G, \mathcal{M}) \leq \tau^{\frac{1}{2d}}.$$

Proof. Suppose the iteration converges to the polynomials $p_{v,\infty}(x)$. It is easy to show by induction that $p_{v,k}(x)$ is SOS for all k. It is trivial for $k = 0$ and if it is true for k then it is also true for $k + 1$ by (2.4). Since the SOS cone is closed, $p_{v,\infty}$ is SOS. Now by (2.4), for each $v \in V$,

$$p_{v,\infty}(x) = q_v(x) + \frac{1}{\tau} \sum_{(u,v,\sigma)\in E} p_{u,\infty}(A_\sigma x)$$

so $p_{v,\infty}(x)$ is also in the interior of the SOS cone. For each edge (u, v, σ), by manipulating the above equation, we have

$$\tau p_{v,\infty}(x) - p_{u,\infty}(A_\sigma x) = \tau q_v(x) + \sum_{(u',v',\sigma')\in E} p_{u',\infty}(A_{\sigma'} x)$$

so $\tau p_{v,\infty}(x) - p_{u,\infty}(A_\sigma x)$ is SOS. Therefore $(\{\, p_{v,\infty}(x) : v \in V \,\}, \tau^{\frac{1}{2d}})$ is a feasible solution of Program 1. □

In view of Theorem 18, it is thus natural to analyse under which condition the iteration 2.4 converges. Recall that iteration 2.4 is an affine map on the vector space of homogeneous polynomials of degree $2d$.

Proof of Theorem 17. Iteration 2.4 is an affine map on the vector space of homogeneous polynomials of degree $2d$. It is well known that if the convergence is guaranteed when we only retain the linear part of the affine map then it is also guaranteed for the affine iteration.

Therefore we can analyse instead the following iteration

$$\begin{aligned} p_{v,0}(x) &= q_v(x), \\ p_{v,k+1}(x) &= \frac{1}{\tau} \sum_{(u,v,\sigma)\in E} p_{u,k}(A_\sigma x), \quad v \in V \end{aligned}$$

We can see that

$$p_{v,k+1}(x) = \frac{1}{\tau^k} \sum_{s \in E_k^-(v)} q_{s(1)}(A_s x) \leq \frac{1}{\tau^k} \sum_{s \in E_k^-(v)} q_{s(1)}(A_s x). \tag{2.5}$$

Consider a norm $\|\cdot\|$ of $\mathbb{R}^n$ and its corresponding induced matrix norm of $\mathbb{R}^{n\times n}$. For each $v \in V$, we know by continuity of $q_v(x)$ that there exist $\beta_v > 0$ such that

$$q_v(x) \leq \beta_v \|x\|^{2d}$$

for all $x \in \mathbb{R}^n$. Let $\beta = \max_{v\in V} \beta_v$.

Therefore,

$$\begin{aligned} p_{v,k+1}(x) &\leq \frac{1}{\tau^k} \sum_{s \in E_k^-(v)} \beta_v \|A_s\|^{2d} \|x\|^{2d} \\ &\leq \frac{\beta}{\tau^k} \|x\|^{2d} \sum_{s \in E_k^-(v)} \|A_s\|^{2d}. \end{aligned}$$

By Lemma 19 and Remark 3, if $\tau > \rho(A(G))\rho_{2d}(G, \mathcal{M})$, $\lim_{k\to\infty} p_{v,k}(x) = 0$. We obtain the result by Theorem 18. □

2.4 Finding high-growth sequences

In this section, we introduce Program 7, the dual of Program 1, and use it to give an alternative proof of Theorem 17. Theorem 17, Lemma 20 and (1.15) give

$$\rho(A(G))^{-1}\rho_{\mathrm{SOS},2d}(G, \mathcal{M}) \leq \rho_{2d}(G, \mathcal{M}) \leq \rho(G, \mathcal{M}) \leq \rho_{\mathrm{SOS},2d}(G, \mathcal{M}) \leq \rho(A(G))\rho_{2d}(G, \mathcal{M}).$$

This alternative proof shows that a solution of the dual program provides a certificate for the lower bound $\rho(A(G))^{-1}\rho_{\mathrm{SOS},2d}(G, \mathcal{M})$ to $\rho_{2d}(G, \mathcal{M})$.

It turns out that we can extract high growth sequence from this certificate. We provide Algorithm 1 parametrized by some integer parameter l that produces an high growth sequence using the solution of Program 7. Let $g_{2d,l}$ be the asymptotic growth rate of the sequenc produced. We show the following guarantee in Theorem 20:

$$[\Delta_l^-]^{-1}\rho_{\mathrm{SOS},2d}(G, \mathcal{M}) \leq g_{2d,l} \leq \rho(G, \mathcal{M}) \leq \rho_{\mathrm{SOS},2d}(G, \mathcal{M}) \leq \rho(A(G))\rho_{2d}(G, \mathcal{M})$$

where Δ_l^- is defined in the List of Notation.

2.4.1 Dual SOS program

Consider the feasibility version of Program 1, i.e. γ is fixed and we want to find polynomials such that the constraints hold.

The dual variables are in the dual of the SOS cones, i.e. they are linear functionals over homogeneous polynomials of degree $2d$ that gives a positive value when applied to an SOS polynomials. Barak et al. [BBH+12] introduced the *pseudo-expectation* $\widetilde{\mathbb{E}}$ as a way to interpret those linear functionals. It can be seen as the expectation of a *pseudo-distribution* $\widetilde{\mu}$. That is,

$$\langle \widetilde{\mathbb{E}}, p(x) \rangle = \widetilde{\mathbb{E}}[p(x)] = \int_{\mathbb{S}^{n-1}} p(x)\, \mathrm{d}\widetilde{\mu}. \tag{2.6}$$

We say that $\widetilde{\mu}$ is only a 'pseudo-distribution' (and not a classical distribution) in the sense that it is not necessarily a positive measure, it might be possible that $\widetilde{\mathbb{E}}[p(x)] < 0$ while $p(x)$ is nonnegative. However, if $p(x)$ is an homogeneous SOS polynomial of degree $2d$ then $\widetilde{\mathbb{E}}[p(x)] \geq 0$ since $\widetilde{\mathbb{E}}$ belongs to the dual of the SOS cone.

Note that for any matrix $A \in \mathbb{R}^{n\times n}$, the polynomial $p(Ax)$ depends linearly on $p(x)$. Let $\mathcal{A}$ be the linear map such that $\mathcal{A}(p(x)) = p(Ax)$ for all $p(x) \in \mathbb{R}[x]_{\mathbf{2d}}$. We define the adjoint map $\mathcal{A}^*$, the linear map such that $\langle \mathcal{A}^*\widetilde{\mathbb{E}}, p(x)\rangle = \langle \widetilde{\mathbb{E}}, \mathcal{A}(p(x))\rangle$ for all $\widetilde{\mathbb{E}} \in (\mathbb{R}[x]_{\mathbf{2d}})^*$ and $p(x) \in \mathbb{R}[x]_{\mathbf{2d}}$.

The dual of the feasibility version of the primal is given by

Program 7 (Dual)**.**

$$\sum_{(u,v,\sigma)\in E} \mathcal{A}_\sigma^* \widetilde{\mathbb{E}}_{uv\sigma} \succeq \gamma^{2d} \sum_{(v,w,\sigma)\in E} \widetilde{\mathbb{E}}_{vw\sigma}, \forall v \in V, \tag{2.7}$$

$$\widetilde{\mathbb{E}}_{uv\sigma} \in \Sigma^*_{\mathbf{2d}}, \qquad \forall (u,v,\sigma) \in E,$$

$$\sum_{(u,v,\sigma)\in E} \widetilde{\mathbb{E}}_{uv\sigma}\Big[\sum_{i=1}^n x_i^{2d}\Big] = 1. \tag{2.8}$$

Note that the dual constraint (2.7) is equivalent to:

$$\sum_{(u,v,\sigma)\in E} \widetilde{\mathbb{E}}_{uv\sigma}[p(A_\sigma x)] \geq \gamma^{2d} \sum_{(v,w,\sigma)\in E} \widetilde{\mathbb{E}}_{vw\sigma}[p(x)] \tag{2.9}$$

for all $v \in V$ and for all $p(x) \in \Sigma_{\mathbf{2d}}$.

Example 10. Consider Example 3. For $i = 1, 2, 3$, let $\widetilde{\mathbb{E}}_i$ be the dual solution at the edge corresponding to the matrix A_i. For any d, we can see that the dual solution for $\gamma = 1$ is such that the only monomial x^α such that $\widetilde{\mathbb{E}}_1[x^\alpha]$ (resp. $\widetilde{\mathbb{E}}_2[x^\alpha]$, $\widetilde{\mathbb{E}}_3[x^\alpha]$) is non-zero is x_1^{2d} (resp. x_2^{2d}, x_3^{2d}) and $\widetilde{\mathbb{E}}_1[x_1^{2d}] = \widetilde{\mathbb{E}}_2[x_2^{2d}] = \widetilde{\mathbb{E}}_3[x_3^{2d}] = 1/3$. Note that it means that $\widetilde{\mu}_1 = \delta_{(1,0,0)}/3$, $\widetilde{\mu}_2 = \delta_{(0,1,0)}/3$ and $\widetilde{\mu}_3 = \delta_{(0,0,1)}/3$ where δ_x is the Dirac measure centered on x. Since there also exists a primal solution for $\gamma = 1$ (see Example 3) we have $\rho_{\mathrm{SOS},2d}(G, \mathcal{M}) = 1$ for any d.

Example 11. We continue the running example; see Example 2 and Example 4.

For all d, $\widetilde{\mathbb{E}}_{212} = \widetilde{\mathbb{E}}_{323} = \widetilde{\mathbb{E}}_{344} = \widetilde{\mathbb{E}}_{431} = 0$ hence the node 4 is "unused" by the dual. For $2d = 2, 4, 6, 8$, $\widetilde{\mathbb{E}}_{123} = \widetilde{\mathbb{E}}_{231} = 0$ so the node 2 is "unused" for low degree.

At first, one could think that the dual variables can be used to reduce the systems, e.g. remove nodes or edges. However, as we will see, it would be a mistake to remove the node 2.

For $2d = 10$, $\widetilde{\mathbb{E}}_{123}$ and $\widetilde{\mathbb{E}}_{231}$ are not zero and are of the "same order or magnitude" than $\widetilde{\mathbb{E}}_{131}$ and $\widetilde{\mathbb{E}}_{312}$. Then for $2d = 12$, $\widetilde{\mathbb{E}}_{123}$ and $\widetilde{\mathbb{E}}_{231}$ have "larger magnitude" than $\widetilde{\mathbb{E}}_{131}$ and $\widetilde{\mathbb{E}}_{312}$. This observation will be useful for Example 13.

We can see that while the shape of the primal variables change a lot between $2d = 2$ and $2d = 4$ as mentioned in Example 4, the "important" change for the dual variables happens around $2d = 10$.

It is also interesting to notice that the matrices corresponding to the dual variables have low rank. For example, for $2d = 2$, the measure corresponding to the dual variable $\widetilde{\mathbb{E}}_{131}$ (resp. $\widetilde{\mathbb{E}}_{312}$, $\widetilde{\mathbb{E}}_{331}$) is the Dirac measure $0.324 \times \delta_{(0.917,0.399)}$ (resp. $0.229 \times \delta_{(0.875,0.485)}$, $0.447 \times \delta_{(0.757,-0.653)}$).

2.4.2 Certificate provided by the dual

The approach of Section 2.3, to prove Theorem 17 is to show that it is possible to build a primal solution with $\gamma \leq \rho_{2d}(G, \mathcal{M})\rho(A(G))^{\frac{1}{2d}}$. In this section we give an alternative approach. We show that the existence of a dual solution with γ provides a certificate that $\rho_{2d}(G, \mathcal{M}) \geq \gamma/\rho(A(G))^{\frac{1}{2d}}$.

We would like to define a random variable S_k representing a path of length k chosen "uniformly at random" in the sense that

$$\rho_{2d}(G, \mathcal{M}) = \lim_{k\to\infty} \mathbb{E}[\|A_{S_k}\|^{2d}]^{\frac{1}{2dk}}. \tag{2.10}$$

However, the probability to chose a path uniformly at random from the set of paths of length k is not the same as the probability to be the truncation to length k of a path chosen uniformly at random from the set of paths of length, say $k+1$.

For (2.10) to be satisfied we need the probability that $S_k = s$ to be equal to the probability that s is the truncation to length k of a path chosen uniformly from the path of "infinite length".

To make this precise we define S_k as a Markov Chain. The states are the nodes of G and the transitions are the edges of $G^\top$. The initial state is some fixed state v_0. When in state v, the probability to go to state u using the edge (u, v, σ) of G is equal to

$$\lim_{k\to\infty} \frac{|E_k^-[u, v, \sigma]|}{|E_k^-(v)|},$$

that is, the fraction of path ending at v in G for which the last edge is (u, v, σ).
Remark 4. We walk "backward" in G. We can either see the random sequence generated by this random walk as G-admissible and left infinite or $G^\top$-admissible and right infinite.

We now show how to use the dual variables to provide a lower bound on the $2d$ moment of the random variable $\|A_{S_k}\|^{2d}$, that is, $\rho_{2d}(G, \mathcal{M})^{2d}$. The dual constraint (2.8) implies the following Lemma.
Lemma 28. Given a finite set of matrices $\mathcal{M}$ constrained by an automaton $G(V, E)$, for any edge $(u, v, \sigma) \in E$, the following holds:

$$\sum_{s\in E_k^-(u)} \mathcal{A}_s^* \widetilde{\mathbb{E}}_{s[1]} \succeq \gamma^{2dk} \widetilde{\mathbb{E}}_{uv\sigma}. \tag{2.11}$$

Proof. We prove (2.11) by induction, the case of $k = 0$ being trivial. We can rewrite the left-hand side of (2.11) as

$$\sum_{s\in E_k^-(u)} \mathcal{A}_s^* \widetilde{\mathbb{E}}_{s[1]} = \sum_{s\in E_{k-1}^-} \sum_{(u,s(1),\sigma)\in E} \mathcal{A}_s^* \mathcal{A}_\sigma^* \widetilde{\mathbb{E}}_{us(1)\sigma}. \tag{2.12}$$

By (2.8),

$$\sum_{(u,s(1),\sigma)\in E} \mathcal{A}_\sigma^* \widetilde{\mathbb{E}}_{us(1)\sigma} \succeq \gamma^{2d} \sum_{(s(1),w,\sigma)\in E} \widetilde{\mathbb{E}}_{s(1)w\sigma}.$$

Since the dual variables $\widetilde{\mathbb{E}}_{s(1)w\sigma}$ of the right-hand side are in the dual of the SOS cone, and one of them is $\widetilde{\mathbb{E}}_{s[1]}$, we have

$$\sum_{(u,s(1),\sigma)\in E} \mathcal{A}_\sigma^* \widetilde{\mathbb{E}}_{us(1)\sigma} \succeq \gamma^{2d} \widetilde{\mathbb{E}}_{s[1]}.$$

Left multiplying by $\mathcal{A}_s^*$ and using (2.12) gives

$$\sum_{s\in E_k^-(u)} \mathcal{A}_s^* \widetilde{\mathbb{E}}_{s[1]} \succeq \sum_{s\in E_{k-1}^-(u)} \gamma^{2d} \mathcal{A}_s^* \widetilde{\mathbb{E}}_{s[1]}.$$

□

We define the random variable

$$\widetilde{\mathbb{E}}_k \triangleq \mathbb{E}[\mathcal{A}^*_{S_k}\widetilde{\mathbb{E}}_{S_k[1]}]. \tag{2.13}$$

We would like to use Lemma 28 to provide a lower bound for $\widetilde{\mathbb{E}}_{S_k[1]}$ but there is a technical difficulty. In the left-hand side of (2.11), each term is equally weighted in the sum while in our Markov Chain, the different values of S_k for a fixed k do not have the same probability to occur. Of course, by construction as $k \to \infty$ each possible value of S_k has the same probability to occur.

To make things precise we define a modified random walk parametrized by a length l. For $l \to \infty$ we recover the original Markov Chain. The transition probability from v depends on the index of the edge, if this is the $(ql+r)$th transition for some nonnegative integers q and positive integer $1 \leq r \leq l$ then the transition probability is

$$\frac{|E^-_{l-r+1}[u,v,\sigma]|}{|E^-_{l-r+1}(v)|}.$$

Remark 5. Since the transition probability depends on the index of the edge, this is not a Markov Chain. However we can define an equivalent random walk that is a Markov Chain using $l|V|$ nodes.

The random variable S_k (resp. $\widetilde{\mathbb{E}}_k$) is written $S_{k;l}$ (resp. $\widetilde{\mathbb{E}}_{k;l}$) for the random walk parametrized by l or S_k (resp. $\widetilde{\mathbb{E}}_k$) when there is no ambiguity.

Lemma 29. Consider a finite set of matrices $\mathcal{M}$ constrained by an automaton $G(V,E)$. For any positive integer d and l, if there exists a solution for Program 7 with γ then the following inequality holds for any positive integer k multiple of l:

$$\widetilde{\mathbb{E}}_{k+l} \succeq \frac{\gamma^{2dl}}{\Delta^-_l}\widetilde{\mathbb{E}}_k$$

and for any edge $(v_0, w, \sigma) \in E$:

$$\widetilde{\mathbb{E}}_l \succeq \frac{\gamma^{2dl}}{\Delta^-_l}\widetilde{\mathbb{E}}_{v_0 w\sigma}$$

where Δ^-_l is defined in the List of Notation.

Proof. By Lemma 28,

$$\begin{aligned}
\widetilde{\mathbb{E}}_k &= \sum_{s\in E^-_k(v_0)} \Pr[S_k = s]\mathcal{A}^*_s\widetilde{\mathbb{E}}_{s[1]} \\
&= \sum_{s\in E^-_k(v_0)} \frac{\Pr[S_{k-l} = s(l+1:)]}{d^-_l(s(l+1))}\mathcal{A}^*_s\widetilde{\mathbb{E}}_{s[1]} \\
&\succeq \frac{1}{\Delta^-_l}\sum_{s\in E^-_k(v_0)} \Pr[S_{k-l} = s(l+1:)]\mathcal{A}^*_s\widetilde{\mathbb{E}}_{s[1]} \\
&\succeq \frac{\gamma^{2dl}}{\Delta^-_l}\sum_{s\in E^-_{k-l}(v_0)} \Pr[S_{k-l} = s]\mathcal{A}^*_s\widetilde{\mathbb{E}}_{s[1]} \\
&= \frac{\gamma^{2dl}}{\Delta^-_l}\widetilde{\mathbb{E}}_{k-l}.
\end{aligned}$$

Similarly,

$$\widetilde{\mathbb{E}}_l = \sum_{s\in E^-_l(v_0)} \Pr[S_l = s]\mathcal{A}^*_s\widetilde{\mathbb{E}}_{s[1]}$$

$$= \frac{1}{d_l^-(v_0)} \sum_{s \in E_l^-(v_0)} \mathcal{A}_s^* \widetilde{\mathbb{E}}_{s[1]}$$
$$\succeq \frac{\gamma^{2dl}}{\Delta_l^-} \widetilde{\mathbb{E}}_{v_0 w \sigma}.$$

□

We are looking for a lower bound on the $(2d)$th moment of the asymptotic growth rate of A_{S_k}. The following three lemmas enable us to give one using the lower bound on the growth rate of $\widetilde{\mathbb{E}}_k$ provided by Lemma 29.

Lemma 30. For any matrix $A \in \mathbb{R}^{n \times n}$ and symmetric positive semidefinite matrix Q, the following inequality holds

$$\rho(A^\top Q A) \leq \rho(Q)\rho(A^\top A).$$

Proof. Since Q is symmetric positive semidefinite, it has a Cholesky decomposition $Q = R^\top R$. Therefore we have

$$\rho(A^\top Q A) = \rho(A^\top R^\top R A) = \|RA\|^2 \leq \|R\|^2 \|A\|^2 = \rho(Q)\rho(A^\top A).$$

□

Lemma 31. For any polynomial $p(x) \in \text{int}(\Sigma_{\mathbf{2d}})$, there exists a constant β such that for any matrix A,

$$\beta \|A\|_2^{2d} p(x) - p(Ax) \quad \text{is SOS}$$

where $\|A\|_2 = \rho(A^\top A)^{1/2}$ is the Euclidean norm. Moreover we can choose $\beta = \kappa(Q)$ where $Q \in \mathcal{S}^{\binom{n+d}{d}}$ is the symmetric matrix such that $p(x) = (x^{[d]})^\top Q x^{[d]}$ for all $x \in \mathbb{R}^n$ and $\kappa(Q) = \rho(Q)\rho(Q^{-1})$ is the condition number of Q.

Proof. Consider the matrix Q defined in the statement of the lemma. Note that since $p(x) \in \text{int}(\Sigma_{\mathbf{2d}})$, Q is positive definite. We can see that

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

Moreover, for all $x \in \mathbb{R}^n$,

$$(x^{[d]})^\top Q x^{[d]} \geq \|x^{[d]}\|_2 / \rho(Q^{-1})$$

and by Lemma 30,

$$(x^{[d]})^\top (A^{[d]})^\top Q A^{[d]} x^{[d]} \leq \rho(Q)\rho((A^{[d]})^\top A^{[d]}) \|x^{[d]}\|_2.$$

Using the fact that $\rho(A^{[d]}) = \rho(A)^d$ and $(AB)^{[d]} = A^{[d]} B^{[d]}$ for any matrix A and B, we have

$$\kappa(Q)\rho(A^\top A)^d Q - (A^{[d]})^\top Q A^{[d]} \succeq 0.$$

□

Lemma 32. For any polynomial $p(x) \in \text{int}(\Sigma_{\mathbf{2d}})$, and any pseudo-expectations $\widetilde{\mathbb{E}}_e \in \Sigma_{\mathbf{2d}}^*$ for each $e \in E$, there exists a positive constant τ such that for any matrix A and edge $e \in E$ such that $\widetilde{\mathbb{E}}_e$ is nonzero,

$$\widetilde{\mathbb{E}}_e[p(Ax)] \leq \tau \|A\|_2^{2d}$$

Proof. If all pseudo-expectations are zero, the result is vacuously true. Therefore we can suppose that at least one is nonzero. By Lemma 31, there exists a constant β such that

$$\beta\|A\|_2^{2d}p(x) - p(Ax) \text{ is SOS.}$$

Hence

$$\widetilde{\mathbb{E}}_e[p(Ax)] \leq \beta\|A\|_2^{2d}\widetilde{\mathbb{E}}_e[p(x)].$$

We obtain the result with the constant $\tau = \beta \max_{e\in E} \widetilde{\mathbb{E}}_e[p(x)]$. Since at least one pseudo-expectation is nonzero and $p(x)$ is in the interior of the SOS cone, $\tau > 0$. □

Theorem 19. If there exists dual variables for Program 7 (degree $2d$) with γ, then for any integer $l \geq 1$,

$$\rho_{2d}(G, \mathcal{M}) = \lim_{k\to\infty} \mathbb{E}\big[\|A_{S_k}\|_2^{2d}\big]^{\frac{1}{2dk}} \geq \frac{\gamma}{\|(A(G)^\top)^l\|_\infty^{\frac{1}{2dl}}}.$$

Proof. By (2.8), there exist at least one edge with a nonzero dual variable. Let (v_0, w, σ) be one of them.

Consider the random walk described above with length parameter l and initial state v_0. By Lemma 29, for any k multiple of l,

$$\widetilde{\mathbb{E}}_k \succeq \frac{\gamma^{2dk}}{(\Delta_l^-)^{\frac{k}{l}}}\widetilde{\mathbb{E}}_{v_0 w\sigma} \tag{2.14}$$

Consider a polynomial $p(x) \in \text{int}(\mathbb{R}_{\mathbf{2d}})$. By Lemma 32, there exists a constant τ such that

$$\widetilde{\mathbb{E}}_k[p(x)] = \mathbb{E}[\widetilde{\mathbb{E}}_{S_k[1]}[p(A_{S_k}x)]] \leq \tau\mathbb{E}[\|A_{S_k}\|^{2d}].$$

Therefore, by (2.14), we have

$$\tau\mathbb{E}[\|A_{S_k}\|^{2d}] \geq \frac{\gamma^{2dk}}{(\Delta_l^-)^{\frac{k}{l}}}\widetilde{\mathbb{E}}_{v_0 w\sigma}[p(x)].$$

Since $\widetilde{\mathbb{E}}_{v_0 w\sigma}$ is nonzero, $\widetilde{\mathbb{E}}_{v_0 w\sigma}[p(x)] > 0$. Therefore taking the $(2dk)$th root and the limit $k \to \infty$ we obtain the result. □

Taking the limit $l \to \infty$ and using (2.3), we see that Theorem 17 is a corollary of Theorem 19.

2.4.3 Randomized algorithm

The result of the previous section are theoretical. We see that there exists a sequence that have a high asymptotic growth rate but we do not show how to find it.

In this section we give an algorithm that generates an infinite sequence of matrices such that the asymptotic value of the norm of the product of the matrices is arbitrarily close to the CJSR. Note that by Definition 2, this asymptotic value must be smaller than the CJSR.

Suppose that we are given a polynomial $p_0(x) \in \text{int}(\Sigma_{\mathbf{2d}})$, Since it is in the interior, $\widetilde{\mathbb{E}}_{uv\sigma}[p_0(x)] > 0$ for all $(u, v, \sigma) \in E$ such that $\widetilde{\mathbb{E}}_{uv\sigma}$ is not zero.

The design of the algorithm is to build a $G^\top$-admissible sequence $(v_1, v_0, \sigma_1), (v_2, v_1, \sigma_2), \ldots$ such that

$$\theta_k \triangleq \widetilde{\mathbb{E}}_{v_k v_{k-1}\sigma_k}[p_0(A_{\sigma_1}\cdots A_{\sigma_k}x)]$$

remains "large" for increasing k. Indeed, in view of Lemma 32 this would suggest that the product $A_{\sigma_1} \cdots A_{\sigma_k}$ has a large norm.

The algorithm works similarly to the Markov Chain of the previous section. In the previous section, at each step, we choose a incoming path of length l uniformly at random. In this section, we will try to make smarter choice using the dual variables. However the dual of the SOS cone only provides a partial order between pseudo-expectations. To extend this partial order to a full order, we evaluate the pseudo-expectations at some fixed polynomial $p_0(x) \in \operatorname{int}(\Sigma_{\mathbf{2d}})$.

By (2.8), at least one dual variable is nonzero and we can choose one of such edges, say (v_0, v_{-1}, σ_0). We then append paths of length l one after the other to build a $G^\top$-admissible sequence. To choose which path to add, we use the partial order provided by the dual variables and $p_0(x)$ to give the probability for each edge to be taken. This is described in Algorithm 1.

Algorithm 1 Generates sequences of large asymptotic growth using paths of length l.

Pick an arbitrary polynomial $p_0(x) \in \operatorname{int}(\Sigma_{\mathbf{2d}})$
Pick an edge $(v_0, v_{-1}, \sigma_0) \in E$ such that $\widetilde{\mathbb{E}}_{v_0 v_{-1} \sigma_0}[p_0(x)] > 0$
for $k = 0, l, 2l, \ldots$ **do**
  Compute $a_s \leftarrow \widetilde{\mathbb{E}}_{s[1]}[p_k(A_s x)]$ for each $s \in E_l^-(v_k)$
  Give a probability pr_s for each $s \in E_l^-(v_k)$ using those computed values a_s
  Pick a random s in $E_l^-(v_k)$ according to those probabilities
  Set $(v_{k+l}, \sigma_{k+l}, \ldots, \sigma_{k+1}, v_k) \leftarrow s$
  Set $p_{k+1} \leftarrow p_k(A_s x)$
end for

Under some condition we still have the guarantee given by Lemma 29 and under further conditions, it is satisfied not only expectation but with probability 1. We define the random variable

$$\Theta_k \triangleq \widetilde{\mathbb{E}}_{S_k[1]}[p_0(A_{S_k} x)].$$

Note that $\Theta_k = \widetilde{\mathbb{E}}_k[p_0(x)]$ where $\widetilde{\mathbb{E}}_k$ is the random variable defined (2.13).

We denote the mean of the a_s as

$$\bar{a} = \frac{1}{d_l^-(v_k)} \sum_{s \in E_l^-(v_k)} a_s.$$

Lemma 33. Consider a finite set of matrices $\mathcal{M}$ constrained by an automaton $G(V, E)$. For any positive integer d and l, using Program 7 with any $\gamma < \rho_{\mathrm{SOS},2d}(G, \mathcal{M})$, Algorithm 1 with paths of length l produces a random variable S_k. If the probability vector pr chosen in Algorithm 1 is in the same order than a, that is, for each $s_1, s_2 \in E_l^-(v_k)$, $\mathsf{pr}_{s_1} \leq \mathsf{pr}_{s_2} \Leftrightarrow a_{s_1} \leq a_{s_2}$, then the following inequality is satisfied for all $k \geq 1$ that is a multiple of l:

$$\mathbb{E}[\Theta_{k+l}] \succeq \frac{\gamma^{2dl}}{\Delta_l^-} \mathbb{E}[\Theta_k].$$

Moreover if $a_s < \bar{a} \Rightarrow \mathbf{pr}_s = 0$ then

$$\Theta_{k+l} \succeq \frac{\gamma^{2dl}}{\Delta_l^-} \Theta_k$$

with probability 1.

Proof. By Lemma 28,

$$\mathbb{E}[\bar{a}] \geq \frac{\gamma^{2dl}}{\Delta_l^-} \mathbb{E}[\Theta_k].$$

If pr is in the same order than a, by the Chebyshev's inequality,

$$\begin{aligned}\sum_{s\in E_l^-(v_k)} \mathsf{pr}_s a_s &\geq \frac{1}{d_l^-(v_k)} \left(\sum_{s\in E_l^-(v_k)} a_s\right)\left(\sum_{s\in E_l^-(v_k)} \mathsf{pr}_s\right)\\ &= \bar{a}.\end{aligned}$$

If $\mathbf{pr}_s = 0$ for each s such that $a_s < \bar{a}$, then $a_s \geq \bar{a}$ with probability 1. □

As done in the previous section for $\widetilde{\mathbb{E}}_k$, we now translate the guarantee on Θ_k to a guarantee on A_{S_k}.

Theorem 20. Consider a finite set of matrices $\mathcal{M}$ constrained by an automaton $G(V,E)$. For any positive integer d and l, using Program 7 with any $\gamma < \rho_{\mathrm{SOS},2d}(G,\mathcal{M})$, Algorithm 1 with paths of length l produces a random variable S_k. If pr is in the same order than a, that is, for each $s_1, s_2 \in E_l^-(v_k)$, $\mathsf{pr}_{s_1} \leq \mathsf{pr}_{s_2} \Leftrightarrow a_{s_1} \leq a_{s_2}$, then the following inequality holds:

$$\lim_{k\to\infty} \mathbb{E}[\|A_{S_k}^{2d}\|_2]^{\frac{1}{2dk}} \geq \frac{\gamma}{\|(A(G)^\top)^l\|_\infty^{\frac{1}{2dl}}}$$

Moreover if $a_s < \bar{a} \Rightarrow \mathsf{pr}_s = 0$ where $\bar{a}$ is the mean of a then

$$\lim_{k\to\infty} \|A_{S_k}^{2d}\|_2^{\frac{1}{2dk}} \geq \frac{\gamma}{\|(A(G)^\top)^l\|_\infty^{\frac{1}{2dl}}}$$

with probability 1.

Proof. We give the proof in the case where pr is in the same order than a. The other case is similar. By Lemma 33, for any k multiple of l,

$$\mathbb{E}[\widetilde{\mathbb{E}}_{S_k[1]}[p_0(A_{S_k}x)]] \geq \frac{\gamma^{2dk}}{(\Delta_l^-)^{\frac{k}{l}}} \widetilde{\mathbb{E}}_{v_0 v_{-1}\sigma_0}[p_0(x)] \tag{2.15}$$

By Lemma 32, there exists a constant τ such that

$$\mathbb{E}[\widetilde{\mathbb{E}}_{S_k[1]}[p_0(A_{S_k}x)]] \leq \tau \mathbb{E}[\|A_{S_k}\|^{2d}].$$

Therefore, by (2.15), we have

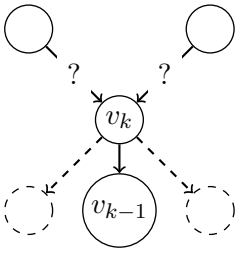
$$\tau \mathbb{E}[\|A_{S_k}\|^{2d}] \geq \frac{\gamma^{2dk}}{(\Delta_l^-)^{\frac{k}{l}}} \widetilde{\mathbb{E}}_{v_0 v_{-1}\sigma_0}[p_0(x)].$$

Since $\widetilde{\mathbb{E}}_{v_0 v_{-1}\sigma_0}$ is nonzero, $\widetilde{\mathbb{E}}_{v_0 v_{-1}\sigma_0}[p_0(x)] > 0$. Therefore taking the $(2dk)$th root and the limit $k \to \infty$ we obtain the result. □

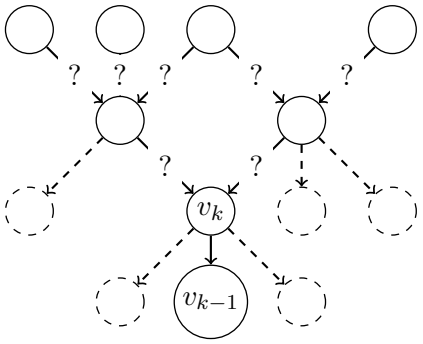
Example 12. Suppose we apply Algorithm 1 with $l = 1$ to Example 10 and the algorithm chooses $p_0(x) = \sum_\alpha c_\alpha x^\alpha$. The start of the sequence produced depends on the order between the coefficients $c_{(2d,0,0)}, c_{(0,2d,0)}, c_{(0,0,2d)}$. If $c_{(2d,0,0)}$ is the largest then the G-admissible left-infinite sequence found is

$$\ldots, 1, 2, 3, 1, 2, 3, 1, 2, 3.$$

The product $A_{\sigma_1}A_{\sigma_2}A_{\sigma_3}\cdots = A_3A_2A_1A_3A_2A_1\cdots$ is periodic and has an asymptotic growth rate $\rho(A_{\sigma_1}A_{\sigma_2}A_{\sigma_3})^{1/3} = 1$. Hence $1 \leq \rho(G,\mathcal{M})$. We saw in Example 10 that $\rho_{\mathrm{SOS},2d}(G,\mathcal{M}) = 1$ for any d. Therefore $\rho(G,\mathcal{M}) = 1$.



(a) One iteration of Algorithm 1 with $l = 1$.



(b) One iteration of Algorithm 1 with $l = 2$.

Figure 2.2: Comparison between Algorithm 1 with $l = 1$ and $l = 2$

2.4.4 Producing lower bounds

By definition of the CJSR, the asymptotic growth rate of the norm of the product of any G-admissible (or $G^\top$-admissible) sequence of matrices gives a lower bound for the CJSR. In particular the sequence produced by Algorithm 1 provides a lower bound for the CJSR.

If there are two integers $\bar{k}, k$ such that the sequence after $\bar{k}$ is periodic of period k, the asymptotic growth rate of the norm is equal to the kth root of the spectral radius of the product of the matrices of one period. This is due to the Gelfand's formula $\rho(A) = \lim_{k\to\infty} \|A^k\|^{1/k}$. From the same identity, we see that the spectral radius of the product of the matrices of one G-admissible cycle gives a lower bound for the CJSR.

More formally, for any natural k, we have

$$\rho_k(G, \mathcal{M}) \leq \rho(G, \mathcal{M})$$

where

$$\rho_k(G, \mathcal{M}) = \max \{ \rho(A_c)^{1/k} : c \in G_k, c \text{ is a cycle} \}.$$

This is the generalization to the constrained case of the first inequality of the three members inequalities; see [Jun09, Section 1.2.2.6].

To find lower bounds for the CJSR, one could generate all the cycles of length smaller than some maximum length and compute the spectral radius for all of them. This brute force approach is not scalable because the number of paths considered grows exponentially with the maximum length. The exponential growth of the brute force approach is the reason why one should choose small l for Algorithm 1.

Gripenberg [Gri96] proposes a branch-and-bound algorithm that prunes the search using an a priori fixed absolute error. Two other branch-and-bound variant exists: the balanced complex polytope algorithm [GZ08] and the invariant conitope algorithm [JCG14].

These algorithms can also be used to produce a G-admissible sequence of matrices of high asymptotic growth rate by reproducing the cycles of high spectral radius infinitely. The advantage of Algorithm 1 is that it provides a guarantee of accuracy given in Theorem 20. Algorithm 1 provides at the same time unstable trajectories and lower bounds of guaranteed accuracy.

We can compute lower bounds using the upper bound provided by Program 1 and Corollary 8 but in practice the trajectories are periodic after some time $\bar{k}$ so we are able to compute much better lower bounds than the pessimistic bound provided by Corollary 8. This is shown by the following example.

Example 13. We first analyse the behaviour of Algorithm 1 for $l = 1$ on our running example; see Example 2, Example 4 and Example 11. We use the recent CSSystem toolbox [CPJ15]. For $2d = 2, 4, 6, 8$, the sequence can only contain the edges $(1, 3, 1)$, $(3, 1, 2)$ and $(3, 3, 1)$ since the other edges have a dual variables of value 0.

To find a cycle with good spectral radius, we generate the sequence for a fixed length and then look through all the cycles of the sequence. The G-admissible cycle found depends on $p_0(x)$, but most of the time, it is

$$(3, 1, 2), (1, 3, 1), (3, 1, 2), (1, 3, 1), (3, 3, 1)^4$$

where the k in exponent means that the edges is taken k times. The 8th root of the corresponding spectral radius is 0.97289.

Using Algorithm 1 with $l = 3$, the produced sequence may include two consecutive edges that have a dual variables of value 0. Even for $2d = 2$, the Algorithm 1 with $l = 3$ finds the G-admissible cycle

$$(3, 1, 2), (1, 3, 1), (3, 1, 2), (1, 2, 3), (2, 3, 1), (3, 3, 1)^3$$

for which the 8th root of the corresponding spectral radius is 0.97482. We can show that it is equal to the CJSR, e.g. using [GZ08]. That is, it is a *spectrum maximizing product* (s.m.p.).

With $2d = 10$ and $2d = 12$, we have seen in Example 11 that $\widetilde{\mathbb{E}}_{123}$ and $\widetilde{\mathbb{E}}_{231}$ are not zero. For $2d = 12$, the s.m.p. is found for any l, initial node and $p_0(x)$. For $2d = 10$, despite the fact that $\widetilde{\mathbb{E}}_{123}$ and $\widetilde{\mathbb{E}}_{231}$ are not zero, the s.m.p. seems to be never found for $l = 1$.

For $2d = 10$ and $l = 2$, it is found from the node 1 and 4[1] with high probability from a random $p_0(x)$ or from $p_0(x)$ equal to the primal solution at this node. From node 3, it is not found using the primal solution of node 3 but it is sometimes found using a random $p_0(x)$. This shows that choosing the primal solution as $p_0(x)$ may not be the best choice. From node 2, it does not seem to be found for any $p_0(x)$.

Instead of increasing l and d, path-complete techniques [AJPR14] can be used to improve the lower bounds provided by Algorithm 1 and the upper bounds provided by Program 1. The idea is to replace the automaton G and the matrices of $\mathcal{M}$ by another automaton G' and other matrices $\mathcal{M}'$ but without changing the set of admissible products of matrices. Typically, the new set $\mathcal{M}'$ contains products of the matrices of the initial set $\mathcal{M}$ and the automaton G' is designed so as to maintain the same set of admissible products.

We tried two examples: The T-product lift [PEDJ15] and M-path-dependent lift [LD06]. Using the 2-product lift, the s.m.p. is found using $2d = 6$ and $l = 1$ and using the 2-path-dependent lift, the s.m.p. is found using $2d = 4$ and $l = 1$. This is concordant with the empirical observation reported in [PEDJ15].

2.4.5 Improving the automaton-dependent bounds

After solving Program 1, we obtain an upper bound on the CJSR and a lower bound thanks to Corollary 8. Running Algorithm 1 provides lower bounds that we can compute if the sequence produced is periodic and this lower bound will always be at least as high than the lower bound produced by Theorem 17.

In this section we show that there is another way to improve the lower bound provided by Theorem 17 using a the dual solution. It will never be higher than the bound provided by

[1] In practice, the dual variables of the edges $(3, 4, 4)$ and $(4, 3, 1)$ are close to zero but not zero. It is possible to start in node 4 even if the sequence will never go back in node 4.

Algorithm 1 but it is a lot cheaper to compute. This lower bound is provided by the following theorem which is a generalization of [PEDJ15, Theorem 3.10].

It is based on the fact that if in an optimal dual solution the dual variable of an edge is 0 then removing the dual variable and the primal constraint of this edge does not affect $\rho_{\mathrm{SOS},2d}(G,\mathcal{M})$. We could prove the following theorem using this fact but we give an altervative proof using Algorithm 1.

Definition 13. Consider a finite set of matrices $\mathcal{M}$ constrained by an automaton G and a positive integer d. The set of edges E_{2d} is the set of edges $e \in E$ such that there exists $\delta > 0$ such that for all $0 < \epsilon < \delta$, there is a solution of Program 7 with $\gamma \geq \rho_{\mathrm{SOS},2d}(G,\mathcal{M}) - \epsilon$ such that $\widetilde{\mathbb{E}}_e = 0$. We define the graph $G_{2d}(V, E_{2d})$ using this set of edges.

Theorem 21. Consider a finite set of matrices $\mathcal{M}$ constrained by an automaton G and a positive integer d. The approximation given by Program 1 using homogeneous polynomials of degree $2d$ satisfies:

$$\rho_{\mathrm{SOS},2d}(G,\mathcal{M}) \leq \rho(G_{2d})^{\frac{1}{2d}} \rho(G,\mathcal{M})$$

where $A(G_{2d})$ is the adjacency matrix of G_{2d}.

Proof. For any edge $e \in E$ such that $\widetilde{\mathbb{E}}_e = 0$, removing this edge does not violate any dual constraint (2.7).

Using Theorem 21 once we have removed all these edges and taking the limit $l \to \infty$ we obtain the result. □

Remark 6. Note that it is not true that $G_{2d_2} \subseteq G_{2d_1}$ when $d_2 \geq d_1$. Indeed, as we have seen with Example 13 the dual variables of edges that were needed to have an s.m.p. are zero for $2d < 10$ so preventing Algorithm 1 to choose those edges even for $l > 1$ prevents it to find the s.m.p. for $2d < 10$.

2.5 Conclusion

We first introduced a technique to exploit the potential low rank structures of switched system in the computation of the CJSR. The downside of this technique is the strong requirement that the matrices must have full rank. We are currently exploring a less restrictive requirement. The rank of a matrix characterize the dimension of the space of vectors that are not mapped to 0 when multiplied by the matrix. This can be generalized by the space of initial state such that all trajectories vanish to the origin. These subspaces are computable in polynomial time using Gurvits iteration [Gur96, PMJ15].

We then introduced a new technique, based on the dual solution of the SOS approximation, to generate a sequence of matrices with asymptotic growth rate close to the CJSR. In Theorem 21, we gave an estimate of its performance. However, in practical examples, the sequence produced achieves a growth rate which is often better than this estimate. This gives rise to the following question: Is it possible to improve the guarantee both for the upper bounds and the lower bounds? There are currently two bounds, Theorem 16 only depending on the dimension of the matrices and Theorem 17 only depending on the automaton. Is there a bound that combines both in a more sophisticated way than Corollary 8?

Our algorithm seems to always produce periodic sequences so in practice we can compute an asymptotic growth rate bound better than what is provided by Corollary 8. This motivates other questions with important algorithmic consequences: Does Algorithm 1 always produce a sequence that is periodic after some $\bar{k}$? Can we give a bound for $\bar{k}$ and the size of the period depending on the dimension of the matrices and on the automaton?

More generally, the techniques developed in this work, based on generating "bad" trajectories for a dynamical system via the dual solution of the natural convex problems used for analysis, naturally extend to many other problems in systems theory. We are currently exploring such possibilities.

Chapter 3

Efficient approximation of controllable sets for switching systems and application to the entropic cone

3.1 Introduction

We mentioned in the introduction that methods for approximation of, and optimization on the entropic cone could have a dramatic impact on Network Coding. However, the entropic cone has a variety of other applications including the capacity region of multi-source network coding [BMR$^+$13], converse theorems for multi-terminal problems of information theory [Yeu12], bounds on the information ratios in secret sharing schemes [Bei11], conditional independence among subvectors of a random vector [Stu06], the guessing numbers of games on directed graphs and entropies of the graphs [Rii06, GR11, BCD$^+$13], additive combinatorics [MMT10], combinatorial inequalities [CY99], Kolmogorov complexity [HRSV97, MMRV02] and determinantal inequalities and group-theoretic inequalities [Cha03, CY02, Cha11].

As we saw in Section 1.4, the notion of invariant sets/reachable sets is very important for switched systems. It can either be computed by polyhedral techniques or by nonlinear techniques using convex optimization.

As we then saw in Section 1.5, the entropic cone can be approximated by a region invariant under some switched system. However, the latter is slightly different to the switched systems studied in Section 1.3 and Section 1.4. In this section, we show that this switched system can be transformed to a *controlled switched system*. In such system, the evolution of the states does not only depends on the dynamic but also on some input signal u_k.

$$x_{k+1} = A_{\sigma_k} x_k + B_{\sigma_k} u_k.$$

For such systems, the notion of invariant set needs to be refined. An appropriate refinement would be the notion of *controllable set*. That is a set S for which $x \in S$ implies that for any transition σ, there exists some input u such that $A_\sigma x + B_\sigma u \in S$.

In this chapter, we provide both a polyhedral and a nonlinear technique that can be used to approximate controllable sets or to solve optimization problems on it. We will then show how to deduce approximation and optimization algorithm for the entropic cone from these.

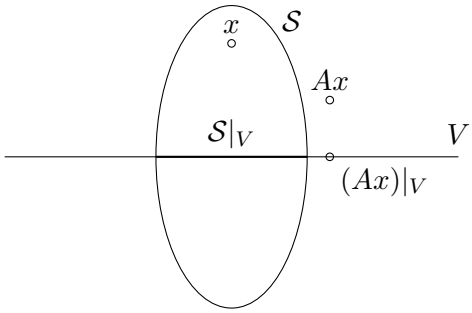


Figure 3.1: Illustration of the invariant condition for being a controllable set. The matrix A denotes a rotation of $-60°$. We see in this example that the drawn ellispoid is not a controllable set since $(Ax)|_V \notin \mathcal{S}|_V$ for some x.

3.2 Computing controllable sets for switched systems

3.2.1 Definition of controllable sets

For a finite set of pairs of matrices $(A, B) \in \mathcal{M} \subseteq \mathbb{R}^{n\times n} \times \mathbb{R}^{n\times d}$, we consider the following switched system with unbouded input:

$$x_{k+1} = A_{\sigma_k} x_k + B_{\sigma_k} u_k. \tag{3.1}$$

where $x_k \in \mathbb{R}^n$ and $u_k \in \mathbb{R}^d$.

We define a controllable set as follows.

Definition 14. A set $\mathcal{S}$ is said to be a *controllable set* of the system (3.1) if

$$\forall x \in \mathcal{S}, \forall (A, B) \in \mathcal{M}, \exists u \in \mathbb{R}^d \text{ such that } Ax + Bu \in \mathcal{S}.$$

Since u_k is unbounded, the only information that matters about B_{σ_k} in the definition of controllable sets is its image. An equivalent condition for a set $\mathcal{S}$ to be controllable is as follows:

$$\forall x \in \mathcal{S}, \forall (A, B) \in \mathcal{M}, (Ax)|_{\mathrm{Im}(B)^\perp} \in \mathcal{S}|_{\mathrm{Im}(B)^\perp}$$

where $\mathcal{S}|_V$ denotes the projection of $\mathcal{S}$ on the subspace V and and $\mathrm{Im}(B)^\perp$ denotes the orthogonal complement of the image of B. This is illustrated by the Figure 3.1.

The projection of $\mathcal{S}$ is a linear transformation on the set $\mathcal{S}$. Consider now, for a finite set of pairs of matrices $(A, E) \in \mathcal{M} \subseteq \mathbb{R}^{n\times n} \times \mathbb{R}^{n\times n}$, the following *algebraic switched system*

$$E_{\sigma_k} x_{k+1} = A_{\sigma_k} x_k. \tag{3.2}$$

For this system, we define a controllable set as follows.

Definition 15. A set $\mathcal{S}$ is said to be a *controllable set* of the system (3.2) if

$$\forall x \in \mathcal{S}, \forall (A, E) \in \mathcal{M}, \exists x \in \mathbb{R}^n \text{ such that } Ax \in E\mathcal{S}. \tag{3.3}$$

Again, for the computation of the controllable set, we can decompose $E = E_1 E_2$ such that E_1 is invertible, $\mathrm{Ker}(E_2) = \mathrm{Ker}(E)$ and $\mathrm{Im}(E_2) = \mathrm{Ker}(E_2)^\perp$. Then an equivalent condition for a set $\mathcal{S}$ to be controllable is as follows:

$$\forall x \in \mathcal{S}, \forall (A, E) \in \mathcal{M}, (E_1^{-1} Ax)|_{\mathrm{Ker}(E)^\perp} \in \mathcal{S}|_{\mathrm{Ker}(E)^\perp}.$$

This shows that we have two different ways to define a controllable set: using a system of the form (3.1) and using a system of the form (3.2).

Remark 7. Systems of the form (3.2) are studied by Trenn [Tre09]. The computation of Lyapunov function, which is similar to the computation of invariant set, is studied in [LT12] but the settings are different because their Lyapunov functions aims at showing the stability *for all* trajectories while in our definition of controllable sets, it is sufficient that one trajectory stays in the controllable set[1].

3.2.2 Why computing controllable sets is difficult

We saw in Section 1.4 how to compute invariant sets of switched systems. We will see in this section that the techniques introduced in Section 1.4 cannot "easily" be generalized for the computation of controllable set of algebraic switched system or switched system with unbounded input.

We consider the following caracterisation of the controllable set

$$\forall x \in \mathcal{S}, \forall (A, V) \in \mathcal{M}, Ax \in \mathcal{S}|_V.$$

The technique of Section 1.4.1 relies on Lemma 5: "For any $A \in \mathcal{M}$, if $x \in \mathcal{S}$ then $Ax \in \mathcal{S}$". So for any $x \in \mathcal{S}$, we can add the point Ax in $\mathcal{S}$, that is what we do in (1.18). For controllable set, we have a different result: For any $A \in \mathcal{M}$, if $x \in \mathcal{S}$ then $\exists y \in \mathcal{S}$ such that $Ax = y|_V$. There is a whole affine space of y such that $y|_V = Ax$ so for given x, we do not know how to modify $\mathcal{S}$ minimally so that $Ax \in \mathcal{S}|_V$.

The technique of Section 1.4.2 relies on Lemma 6: "For any $A \in \mathcal{M}$, $A^{-1}\mathcal{S} \subseteq \mathcal{S}$". The H-representation of $A^{-1}\mathcal{S}$ is easy to compute from the H-representation of $\mathcal{S}$, we simply need to multiply each facet by $A^\top$. For controllable set, the property is different: "For any $A \in \mathcal{M}$, $A^{-1}(\mathcal{S}|_V) \subseteq \mathcal{S}$". So we would first need to compute the projection of $\mathcal{S}$ on V and then multiply each resulting facet by $A^\top$. However, the polyhedral projection of a polyhedra in his H-representation is computationally expensive as discussed in Section 1.1 and we need to work with the H-representation and not the V-representation because we need to compute the intersection of polyhedra and to multiply each facet by $A^\top$.

The technique of Section 1.4.3 relies on the fact that if $\mathcal{S} = \{ x \mid p(x) \leq 1 \}$ then $A^{-1}\mathcal{S} = \{ x \mid p(Ax) \leq 1 \}$. Therefore since, by (1.17), $\mathcal{S}$ must satisfy $\mathcal{S} \subseteq A^{-1}\mathcal{S}$ to be invariant, we enforce the inequality $p(Ax) \leq p(x)$. For controllable sets, the inequality would rather be $q(Ax) \leq p(x)$ where $\mathcal{S}|_V = \{ x \mid q(x) \leq 1 \}$. However, without a simple relation between $q(x)$ and $p(x)$ that would be implementable as constraints in conic programming (e.g. linear, SDP, ...), $q(Ax) \leq p(x)$ does not contraint anything on $p(x)$ since $q(x)$ is free.

3.2.3 Parallel optimization on controllable sets

In this section, we show how to solve optimization problem on controllable sets in an efficient and parallelizable manner using ideas from *Stochastic Programming* [BL11]. In stochastic programming, large scale convex programs are decomposed into smaller convex programs linked together by a markov chain. The convex program at each state u of the markov chain is:

$$\begin{aligned} Q(x, u) = \text{minimize } & c^T y + \mathcal{Q}_u(y) \\ \text{s.t. } & W_u y = h_u - T_u x, \\ & x \geq 0 \end{aligned}$$

[1] In some sense, this is the difference of using an input u and using a perturbation w in (3.1), or using $\exists x \in \mathbb{R}^n$ instead of $\forall x \in \mathbb{R}^n$ in (3.3).

where $\mathcal{Q}_u(y)$ is the sum of $Q(x,v)$ for each state v accessible from u weighted by the probability to go from state u to state v. When the program is infeasible for some x, $Q(x,u)=\infty$. At the initial state of the markov chain, there is no term $-T_u x$ and the solution at this stage is the solution of the original large scale convex program. Note that if v is accessible from different states u, u', the number of variables at u and u' must match for $Q(\cdot,v)$ to be well-defined.

Since we have shown that computing the controllable set of switched system with unbounded input (3.1) or algebraic switched system (3.2) is two equivalent problems, we only consider the latter in this section.

Suppose we are looking for the optimal solution of the following optimization problem:

$$\min_{x\in\mathcal{C}}\langle c,x\rangle$$

where $\mathcal{C}$ is the maximal controllable subset of some convex set $\mathcal{C}_0$.

We consider the following relaxation

$$\min_{x\in\mathcal{C}_T}\langle c,x\rangle \tag{3.4}$$

where $\mathcal{C}_T$ is the maximal subset of $\mathcal{C}_0$ that is controllable up to time T. That is, for any $\sigma_1,\ldots,\sigma_T$, there exist $x_1,\ldots,x_T\in\mathcal{C}_0$ such that $E_{\sigma_k}x_k=A_{\sigma_k}x_{k-1}$ for $k=1,\ldots,T$. We could have equivalently defined $\mathcal{C}_T$ recursively as follows:

$$\mathcal{C}_k=\mathcal{C}_{k-1}\cap\bigcap_{\sigma}A_\sigma^{-1}E_\sigma\mathcal{C}_{k-1}.$$

The problem (3.4) is a convex program in[2] $n(1+m^T)$ variables. Stochastic programming techniques can be used to decompose this huge program. The convex program at the initial state of the markov chain is

$$\begin{aligned}&\text{minimize } \langle c,x\rangle+\mathcal{Q}_0(x)\\ &\text{s.t. } x\in\mathcal{C}_0,\end{aligned}$$

and the convex program at state $(t,\sigma)\in[T]\times[m]$ is

$$\begin{aligned}Q(x,(t,\sigma))=\text{minimize }&\mathcal{Q}_{(t,\sigma)}(y)\\ \text{s.t. }&E_\sigma y\in A_\sigma x\\ &y\in\mathcal{C}_0.\end{aligned}$$

A state (t,σ) is accessible from any state $(t-1,\sigma')$. This means that $\mathcal{Q}_{(t,\sigma)}$ does not depend on σ.

Since there is no objective in states other than the initial state, the probability assigned for each transition does not matter. The only relevant information in $Q(x,(t,\sigma))$ is whether it is infinite or zero.

The convex programs can therefore be rewritten as follows where by convension, an infeasible program as an optimal value if ∞:

$$\begin{aligned}&\text{minimize } \langle c,x\rangle\\ &\quad\text{s.t.}\\ &\qquad x\in\mathcal{C}_0\\ &\qquad x\in\operatorname{dom}(\mathcal{Q}_0(x))\end{aligned}\qquad\qquad\begin{aligned}Q(x,(t,\sigma))=\text{minimize }&0\\ \text{s.t. }&E_\sigma y\in A_\sigma x\\ &y\in\mathcal{C}_0\\ &y\in\operatorname{dom}(\mathcal{Q}_{(t,\sigma)}(y)).\end{aligned}$$

[2] m^T is m to the power T and not the transpose of m.

One can verify that $\mathcal{C}_0 \cap \mathrm{dom}(\mathcal{Q}_t(x)) = \mathcal{C}_{T-t}$.

In stochastic programming, the domain of $Q(\cdot, u)$ is approximated by a polyhedron by starting with the approximation $Q(\cdot, u) = \mathbb{R}^n$ and adding *feasibility cuts*. The feasibility cuts are computed as follows: the convex program at u is solved for some x, if it is infeasible, an unbounded ray of the dual program is computed and used to generate a feasibility cut.

These observations lead to Algorithm 2 for optimizing on controllable sets that is inspired from the *Stochastic Dual Dynamic Programming* algorithm used in stochastic programming. Note that the initial node can have a nonlinear objective function and additionnal nonlinear constraints. This algorithm is easily parallelizable as the convex programs of the different states only need to communicate cuts and optimal solutions.

Algorithm 2 Minimization of $\langle c, x \rangle$ subject to $x \in \mathcal{C}_T \cap \mathbb{F}$ for parameters K, ρ.

for $k = 1, 2, \ldots, K$ **do**
  Pick a set P of ρ random paths of length T starting at the initial state.
  Solve the optimization program

$$\begin{aligned} &\text{minimize } c(x) \\ &\text{s.t. } x \in \mathcal{C}_0 \cap \mathbb{F} \cap \mathrm{dom}(\mathcal{Q}_0), \end{aligned}$$

  where the domain of $\mathcal{Q}_0$ is approximated by the feasibility cuts.
  if the program is infeasible **then**
    return Infeasible
  end if
  for $t = 1, 2, \ldots, T$ **do**
    for all $p \in P$ **do**
      Solve $Q(x, (t, \sigma_{p,i}))$ where x is the value of the optimal solution of the previous state in the path.
      if the program is infeasible **then**
        Add a feasibility cut for $\mathrm{dom}(Q(\cdot, (t, \sigma_{p,i})))$
        Remove p from P
      end if
    end for
  end for
end for

We developed a Julia [BKSE12] package `StochasticDualDynamicProgramming.jl` to solve Stochastic Dual Dynamic Programming problems. Our package use the StructJuMP package [HLP14] for modeling the convex programs in the markov chain structure. For solving the convex programs, it uses MathProgBase [LHDB14] so any solver supported by MathProgBase can be used.

3.2.4 Nonlinear approximation of controllable sets

As done in Section 3.2.3, we consider algebraic switched system (3.2). As seen in Section 3.2.2, the technique of Section 1.4.3 cannot be applied directly to approximate controllable sets.

The issue is that the controllability constraint is

$$\forall \sigma \in [m], A_\sigma \mathcal{S} \subseteq E_\sigma \mathcal{S}. \tag{3.5}$$

while what we need is constraints of the form (1.17) or (1.22) to apply Lemma 9. That is, of the form LHS $\subseteq$ RHS where the LHS and the RHS are either $\mathcal{S}$ or $M^{-1}\mathcal{S}$ for some matrix M since

we can handle it with (1.23). The invariance constraint (3.5) is equivalent to

$$\forall \sigma \in [m], \mathcal{S} \subseteq A_\sigma^{-1} E_\sigma \mathcal{S}$$

but *not* to

$$\forall \sigma \in [m], E_\sigma^{-1} A_\sigma \mathcal{S} \subseteq \mathcal{S}.$$

The main idea to "make these inverse matrices appear" is to generalize Lemma 7 and work on the polar.

Lemma 34. A convex set $\mathcal{C}$ is a controllable set of the system (3.2) if and only if

$$\forall \sigma \in [m], E_\sigma^{-\top} \mathcal{C}^\circ \subseteq A_\sigma^{-\top} \mathcal{C}^\circ. \tag{3.6}$$

Proof. Using (1.1), we see that (3.5) is equivalent to

$$\forall \sigma \in [m], (E_\sigma \mathcal{C})^\circ \subseteq (A_\sigma \mathcal{C})^\circ.$$

Then using Lemma 1, we can rewrite it as

$$\forall \sigma \in [m], E_\sigma^{-\top} \mathcal{C}^\circ \subseteq A_\sigma^{-\top} \mathcal{C}^\circ.$$

□

Taking into account the difference between Lemma 7 and Lemma 34, much of Section 1.4.3 can be reused by doing the computation on the polar.

Definition 16 (sos polar controllability)**.** The sublevel set of a polynomial function $p(x)$ is said to be an *sos-polar-controllable set* of the system (3.2) if

$$\forall \sigma \in [m], p(E_\sigma^\top x) - p(A_\sigma^\top x) \text{ is SOS}.$$

Definition 17 (lift polar controllability)**.** The sublevel set of a polynomial function $p(x)$ is said to be a *lift-polar-controllable set* of the system (3.2) if there exists a matrix $Q \in \mathcal{S}^n_{++}$ such that $p(x) = \langle Qx^{[d]}, x^{[d]} \rangle$ and

$$\forall \sigma \in [m], E_\sigma Q E_\sigma^\top - A_\sigma Q A_\sigma^\top \succeq 0.$$

As discussed in Section 1.4.3, we have two choices: either to compute an sos-lift-controllable set and then to compute the polar using algebraic techniques or to compute a lift-polar-controllable set and then to compute the polar in the lifted space which boils down to a matrix inversion.

The following program can be used to approximate the polar of maximal controllable subsets using sos-polar-invariant sublevel sets of sum of squares polynomials.

Program 8 (Polar of Maximal Controllable Subset Approximation)**.**

$$\begin{aligned} \underset{p(x) \in \mathbb{R}_{\mathbf{2d}}[x]}{\text{maximize}} \; & \det(Q) \\ & Q \succ 0 \\ & p(x) = \langle Qx^{[d]}, x^{[d]} \rangle \\ & p(E_\sigma^\top x) - p(A_\sigma^\top x) \text{ is SOS} \quad \forall \sigma \in [m] \\ & p(x) \leq 1 \quad \forall x \in \mathcal{C}_0{}^\circ. \end{aligned} \tag{3.7}$$

Note that in practice, if $\mathcal{C}_0$ is a polytope, (3.7) can be replaced by $p(x) \leq 1$ for all vertices of $\mathcal{C}_0{}^\circ$/facets of $\mathcal{C}_0$ with the additional constraint that $p(x)$ is sos-convex; see Section 1.4.3 for more details about this.

For the computation of a minimal controllable superset of a polytope $\mathcal{C}_0$, we need to check that every facet of $\mathcal{C}_0$ is in the polar of the sublevel set of $p(x)$. We can see with Lemma 13 that it is sufficient for $x^{[d]}$ to be in the polar of $\{ y \mid \langle Qy, y \rangle \leq 1 \}$ for every facet x of $\mathcal{C}_0$. This can be enforced using the constraint $\langle Q^{-1} x^{[d]}, x^{[d]} \rangle$ for every facet x of $\mathcal{C}_0$ which can be implemented using Lemma 12.

3.3 The entropic cone as a controllable set

We say in Section 1.6 with Theorem 13 that the entropic cone is both inner-adhesive and self-adhesive. Those adhesive invariance are almost the form (3.2). The only difference is that the invariance is between entropic cones of different dimension.

We developed a new toolbox `EntropicCone.jl` in Julia [BKSE12] for working with the entropic cone. In parallel, we created a Julia package `Polyhedra.jl` to separate specific code for the entropic cone and general code for polyhedral computations. The package `Polyhedra.jl` provides an unified interface for polyhedral computation library, much like `MathProgBase.jl` provides an unified interface for optimization solvers. Currently, CDD [Fuk03] and LRS [Avi00] are supported through the `CDDLib.jl` and `LRSLib.jl` packages respectively. The main features of `Polyhedra.jl` are representation conversion, polyhedral projection, intersections, convex/conic hulls and visualization using `GLVisualize.jl`. See Figure 3.3 for a summary of the packages developped and used.

We hooked up our new toolbox `StochasticDualDynamicProgramming.jl` introduced in Section 3.2.3 with `EntropicCone.jl` to make it possible to run Algorithm 2 on the entropic cone. We tested Algorithm 2 to find lower bounds for the Ingleton score and obtained the best known lower bound -0.15789 in under a minute. The cut generated to find this lower bound are represented in Figure 3.2. Comparing Figure 3.2 with the outer approximation of Figure 1.10, we see that, as expected, our approach only generate cuts close to the optimal value.



Figure 3.2: The polymatroid cone $\mathcal{P}_4$ cut by the feasibility cuts uncovered to find the lower bound -0.15789 of the Ingleton score.

We are currently implementing the approximation of the entropic cone using Program 8.

3.4 Conclusion

In this chapter, we have seen both a polyhedral and a nonlinear technique to approximate controllable sets and to solve optimization problems on it. We have shown that the entropic cone

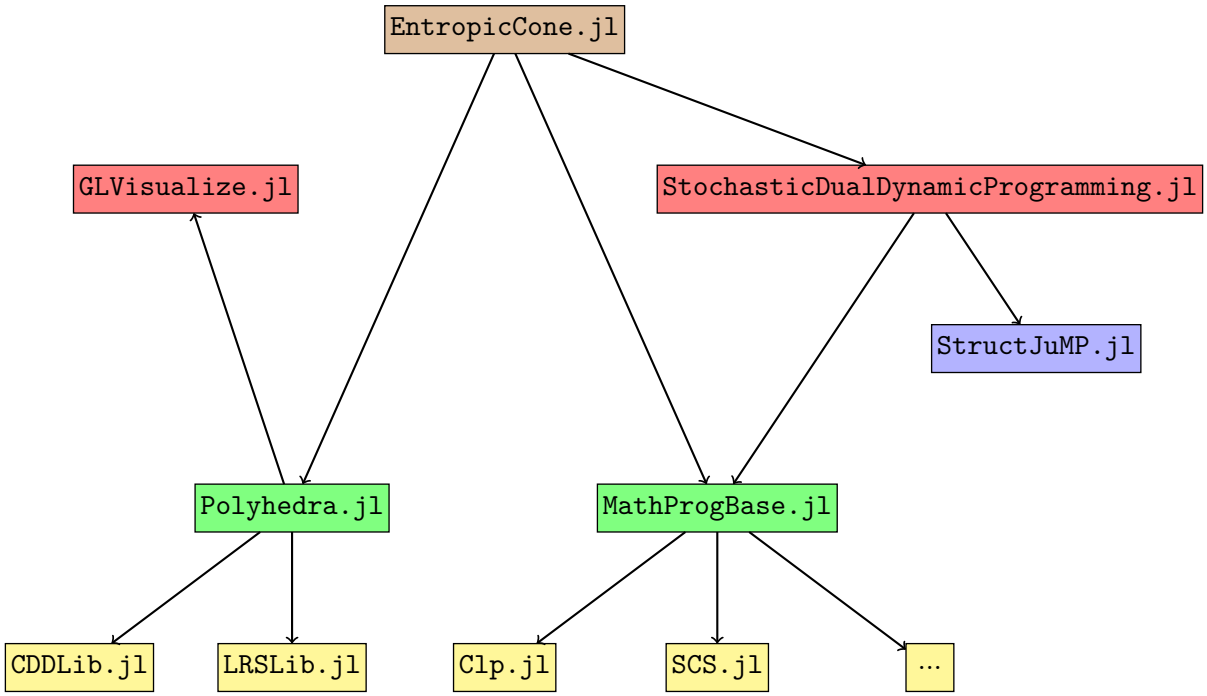


Figure 3.3: Overview of the Julia packages developped and used. In yellow are low level libraries and solvers, in green are abstraction interface, in blue are modelling interfaces, in red are general purpose packages and in brown are toolboxes. See `http://www.juliaopt.org/` for a comprehensive list of supported solvers.

is contained[3] in the maximal controllable subset of the polymatroid cone $\mathcal{P}_n$ for some controlled switched system. Therefore these techniques can directly be applied to the approximaiton of the entropic cone.

The polyhedral approximation technique was already known for the entropic cone but it is impractical because it rely on polyhedral projection in high dimensional space. Nevertheless, we implemented them in the `EntropicCone.jl` toolbox to be able to reproduce the state of the art.

The linear programming approach presented for optimization on controllable set relying on stochastic programming methods seems promising as it only approximate the controllable set close to the optimum. It was implemented in the `EntropicCone.jl` toolbox and presented at SITB 2016 [LJ16]. It was able to find the same lower bound for the ingleton score than the polyhedral approximation technique in under a minute.

As the entropic cone is nonalgebraic and high dimensional, approximating it using polyhedral technique is expected to be expensive. For this reason, information theorists are looking for nonlinear contraints on the entropic cone [CG08]. In this section, we showed how to exploit the approximate the entropic cone using sum of squares and semidefinite programming. We are currently implementing the method in the `EntropicCone.jl` toolbox.

We put a particular emphasis on the reproducibility of our result and the quality and reusability of the code. For this purpose, we reused existing packages as much as possible and made separate packages the code was separate and could be reused.

[3] it is still open whether this inclusion is strict or not

Conclusion

Today, there is increasing interaction between our physical world and the digital devices such as embedded controller and robotics. From this interaction between the continuous dynamics of the physical world and the discrete dynamics of computational elements emerged the modern notion of Cyber-Physical systems. Classical control techniques have not been designed for such systems and we need new tools to handle these complex interactions. In recent years, path-complete graph techniques [AJPR14] provided a new unifying approach to analyse the behaviour of a switched system. While this framework is a crucial step toward to capability to handle these complex systems, many questions still remain to be answered. In Chapter 2, we presented a new result improving our understanding the behavior of modern computational techniques on Path-Complete Graphs. Current technique often seeks to certify the stability of systems but if a system is unstable, one might want to determine the cause of instability. We developed a new tool allowing to exploit the computation on Path-Complete Graphs to uncover the unstable behaviors that make a system unstable.

We are currently working on improving and clarifying the result notably by applying the results to the design of path-complete graphs. The presentation of Chapter 2 already incorporates some enhancement like the relation between the p-radius and the computation of the CJSR. A journal paper including these improvements is in preparation.

For the second contribution, we are currently working on improvements of the implementation to improve the efficiency of the method. As it obtained the best known lower bound for the Ingleton score in under a minute, we should be able to beat it. We noticed that the performance of the methods was rather sensitive to certain parameters like the sampling algorithm for choosing the paths and the way to filter the cuts. A better understanding of the effect of those parameters could significantly enhance our ability to solve optimization on the Entropic Cone which would have a dramatic impact on Network Coding.

We found the third contribution quite recently. Many questions remain to be answered such as: What guarantee on the accuracy of the approximation can we provide ? How can we use path-complete techniques to improve the approximation of invariant sets ? After revolutionizing control theory, The advanced mathematical tool such as Linear Matrix Inequalities, sum of squares and the S-procedure might have the same impact on the approximation of the entropic cone and its many application such as network coding, information theory, cryptography and probability. We do not pretend that our technique will play that role but we hope that it will at least shows that such tools offer a promising alternative to the linear techniques that are currently used.

References

[ABS95] David Avis, David Bremner, and Raimund Seidel. How good are convex hull algorithms? In *Proceedings of the eleventh annual symposium on Computational geometry*, pages 20–28. ACM, 1995.

[AJ+13a] Amir Ali Ahmadi, R Jungers, et al. Sos-convex lyapunov functions with applications to nonlinear switched systems. In *Proceedings of the IEEE Conference on Decision and Control*, 2013.

[AJ13b] Amir Ali Ahmadi and Raphael M Jungers. Switched stability of nonlinear systems via sos-convex lyapunov functions and semidefinite programming. In *Decision and Control (CDC), 2013 IEEE 52nd Annual Conference on*, pages 727–732. IEEE, 2013.

[AJ16] Nikolaos Athanasopoulos and Raphaël M. Jungers. Computing the domain of attraction of switching systems subject to non-convex constraints. In *Proceedings of the 19th International Conference on Hybrid Systems: Computation and Control*, HSCC '16, pages 41–50, New York, NY, USA, 2016. ACM.

[AJPR14] Amir Ali Ahmadi, Raphaël M Jungers, Pablo A Parrilo, and Mardavij Roozbehani. Joint spectral radius and path-complete graph Lyapunov functions. *SIAM Journal on Control and Optimization*, 52(1):687–717, 2014.

[AOPT13] Amir Ali Ahmadi, Alex Olshevsky, Pablo A Parrilo, and John N Tsitsiklis. Np-hardness of deciding convexity of quartic polynomials and related problems. *Mathematical Programming*, 137(1-2):453–476, 2013.

[AP12a] Amir Ali Ahmadi and Pablo A Parrilo. A convex polynomial that is not sos-convex. *Mathematical Programming*, 135(1-2):275–292, 2012.

[AP12b] Amir Ali Ahmadi and Pablo A Parrilo. Joint spectral radius of rank one matrices and the maximum cycle mean problem. In *CDC*, pages 731–733, 2012.

[AP13] Amir Ali Ahmadi and Pablo A Parrilo. A complete characterization of the gap between convexity and sos-convexity. *SIAM Journal on Optimization*, 23(2):811–833, 2013.

[Avi00] David Avis. A revised implementation of the reverse search vertex enumeration algorithm. In *Polytopes—combinatorics and computation*, pages 177–198. Springer, 2000.

[Bar02] Alexander Barvinok. *A course in convexity*, volume 54. American Mathematical Society Providence, 2002.

[BBH+12] Boaz Barak, Fernando GSL Brandao, Aram W Harrow, Jonathan Kelner, David Steurer, and Yuan Zhou. Hypercontractivity, sum-of-squares proofs, and their

applications. In *Proceedings of the forty-fourth annual ACM Symposium on Theory of Computing*, pages 307–326. ACM, 2012.

[BCD+13] Rahil Baber, Damianos Christofides, Anh N Dang, Søren Riis, and Emil R Vaughan. Multiple unicasts, graph guessing games, and non-Shannon inequalities. In *Network Coding (NetCod), 2013 International Symposium on*, pages 1–6. IEEE, 2013.

[BEGFB94] Stephen P Boyd, Laurent El Ghaoui, Eric Feron, and Venkataramanan Balakrishnan. *Linear matrix inequalities in system and control theory*, volume 15. SIAM, 1994.

[Bei11] Amos Beimel. Secret-sharing schemes: a survey. In *Coding and cryptology*, pages 11–46. Springer, 2011.

[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.

[BL11] John R Birge and Francois Louveaux. *Introduction to stochastic programming*. Springer Science & Business Media, 2011.

[BMR+13] Riccardo Bassoli, Hugo Marques, Jose Rodriguez, Kenneth W Shum, and Rahim Tafazolli. Network coding theory: A survey. *Communications Surveys & Tutorials, IEEE*, 15(4):1950–1978, 2013.

[BN05] Vincent D Blondel and Yurii Nesterov. Computationally efficient approximations of the joint spectral radius. *SIAM Journal on Matrix Analysis and Applications*, 27(1):256–272, 2005.

[BPT12] G. Blekherman, P. Parrilo, and R. Thomas. *Semidefinite Optimization and Convex Algebraic Geometry*. Society for Industrial and Applied Mathematics, Philadelphia, PA, 2012.

[BT00] Vincent D Blondel and John N Tsitsiklis. The boundedness of all products of a pair of matrices is undecidable. *Systems & Control Letters*, 41(2):135–140, 2000.

[CG07] Terence Chan and Alex Grant. Entropy vectors and network codes. *arXiv preprint cs/0702063*, 2007.

[CG08] Terence Chan and Alex Grant. Non-linear information inequalities. *Entropy*, 10(4):765–775, 2008.

[Cha03] Terence H Chan. Balanced information inequalities. *Information Theory, IEEE Transactions on*, 49(12):3261–3267, 2003.

[Cha11] Terence Chan. Recent progresses in characterising information inequalities. *Entropy*, 13(2):379–401, 2011.

[CLR95] Man-Duen Choi, Tsit Yuen Lam, and Bruce Reznick. Sums of squares of real polynomials. In *Proceedings of Symposia in Pure mathematics*, volume 58, pages 103–126. American Mathematical Society, 1995.

[CPJ15] Léopold Cambier, Matthew Philippe, and Raphaël Jungers. The CSSystem toolbox. `http://www.mathworks.com/matlabcentral/fileexchange/52723-the-cssystem-toolbox`, August 2015.

[CY99] Ho-Leung Chan and Raymond W Yeung. A combinatorial approach to information inequalities. In *Information Theory and Networking Workshop, 1999*, page 63. IEEE, 1999.

[CY02] Terence H Chan and Raymond W Yeung. On a relation between information inequalities and group theory. *Information Theory, IEEE Transactions on*, 48(7):1992–1995, 2002.

[Dai12] Xiongping Dai. A Gel'fand-type spectral radius formula and stability of linear constrained switching systems. *Linear Algebra and its Applications*, 436(5):1099–1113, 2012.

[DFZ05] Randall Dougherty, Christopher Freiling, and Kenneth Zeger. Insufficiency of linear coding in network information flow. *Information Theory, IEEE Transactions on*, 51(8):2745–2759, 2005.

[DFZ11] Randall Dougherty, Chris Freiling, and Kenneth Zeger. Non-Shannon information inequalities in four random variables. *arXiv preprint arXiv:1104.3602*, 2011.

[Fek23] Michael Fekete. Über die Verteilung der Wurzeln bei gewissen algebraischen Gleichungen mit ganzzahligen Koeffizienten. *Mathematische Zeitschrift*, 17(1):228–249, 1923.

[Fuk96] K Fukuda. Note on new complexity classes enp, ep and cep—an extension of the classes np co-np and p. *ETH Zürich, Institute for Operations Research, Zürich June*, 12:1996, 1996.

[Fuk03] Komei Fukuda. CDD–A C-implementation of the double description method. *Institut fur Operations Research, ETH Z urich, available via anonymous ftp: ifor13. ethz. ch, directory pub/fukuda/cdd*, 2003.

[GR11] Maximilien Gadouleau and Søren Riis. Graph-theoretical constructions for graph entropy and network coding based communications. *Information Theory, IEEE Transactions on*, 57(10):6703–6717, 2011.

[Gri96] Gustaf Gripenberg. Computing the joint spectral radius. *Linear Algebra and its Applications*, 234:43–60, 1996.

[Gur96] Leonid Gurvits. Stability of linear inclusions—part 2. *Linear Algebra Appl.(submitted)*, pages 96–173, 1996.

[GZ08] Nicola Guglielmi and Marino Zennaro. An algorithm for finding extremal polytope norms of matrix families. *Linear Algebra and its Applications*, 428(10):2265–2282, 2008.

[HLP52] Godfrey Harold Hardy, John Edensor Littlewood, and George Pólya. *Inequalities*. Cambridge university press, 1952.

[HLP14] Joey Huchette, Miles Lubin, and Cosmin Petra. Parallel algebraic modeling for stochastic optimization. In *Proceedings of the 1st First Workshop for High Performance Technical Computing in Dynamic Languages*, pages 29–35. IEEE Press, 2014.

[HRSV97] Daniel Hammer, Andrei E Romashchenko, Alexander Shen, and Nikolai K Vereshchagin. Inequalities for shannon entropies and kolmogorov complexities. In *Computational Complexity, 1997. Proceedings., Twelfth Annual IEEE Conference on (Formerly: Structure in Complexity Theory Conference)*, pages 13–23. IEEE, 1997.

[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.

[Ing71] AW Ingleton. Conditions for representability and transversality of matroids. In *Théorie des Matroïdes*, pages 62–66. Springer, 1971.

[JCG14] Raphaël M Jungers, Antonio Cicone, and Nicola Guglielmi. Lifted polytope methods for computing the joint spectral radius. *SIAM Journal on Matrix Analysis and Applications*, 35(2):391–410, 2014.

[JL+03] Ali Jadbabaie, Jie Lin, et al. Coordination of groups of mobile autonomous agents using nearest neighbor rules. *Automatic Control, IEEE Transactions on*, 48(6):988–1001, 2003.

[Jun09] Raphaël Jungers. *The joint spectral radius: theory and applications*, volume 385. Springer Science & Business Media, 2009.

[KL51] Solomon Kullback and Richard A Leibler. On information and sufficiency. *The annals of mathematical statistics*, pages 79–86, 1951.

[Koz14] Victor Kozyakin. The Berger–Wang formula for the Markovian joint spectral radius. *Linear Algebra and its Applications*, 448:315–328, 2014.

[LD06] Ji-Woong Lee and Geir E Dullerud. Uniform stabilization of discrete-time switched and Markovian jump linear systems. *Automatica*, 42(2):205–218, 2006.

[LHDB14] M. Lubin, J. Huchette, I. Dunning, and C. Baldassi. Mathprogbase.jl. `https://github.com/JuliaOpt/MathProgBase.jl`, 2014.

[LT12] Daniel Liberzon and Stephan Trenn. Switched nonlinear differential algebraic equations: Solution theory, Lyapunov functions, and stability. *Automatica*, 48(5):954–963, 2012.

[Mat07a] František Matúš. Infinitely many information inequalities. In *Information Theory, 2007. ISIT 2007. IEEE International Symposium on*, pages 41–44. IEEE, 2007.

[Mat07b] František Matúš. Two constructions on limits of entropy functions. *Information Theory, IEEE Transactions on*, 53(1):320–330, 2007.

[MC13] F. Matúš and L. Csirmaz. Entropy region and convolution. *ArXiv e-prints*, October 2013.

[MFP07] Jaume Martí-Farré and Carles Padró. On secret sharing schemes, matroids and polymatroids. In *Theory of Cryptography*, pages 273–290. Springer, 2007.

[MM15] Carolina Mejia and J Andres Montoya. The almost-entropic regions are not semialgebraic. *arXiv preprint arXiv:1510.02658*, 2015.

[MMRV02] Konstantin Makarychev, Yury Makarychev, Andrei Romashchenko, and Nikolai Vereshchagin. A new class of non-Shannon-type inequalities for entropies. *Communications in Information and Systems*, 2(2):147–166, 2002.

[MMT10] Mokshay Madiman, A Marcus, and Prasad Tetali. Information-theoretic inequalities in additive combinatorics. In *Proc. of the 2010 IEEE Information Theory Workshop*, pages 1–4, 2010.

[MS95] František Matúš and Milan Studený. Conditional independences among four random variables I. *Combinatorics, Probability and Computing*, 4(03):269–278, 1995.

[Nes00] Yurii Nesterov. Squared functional systems and optimization problems. In *High performance optimization*, pages 405–440. Springer, 2000.

[Oxl06] James G Oxley. *Matroid theory*, volume 3. Oxford University Press, USA, 2006.

[Par00] Pablo A Parrilo. *Structured semidefinite programs and semialgebraic geometry methods in robustness and optimization*. PhD thesis, Citeseer, 2000.

[Par03] Pablo A Parrilo. Semidefinite programming relaxations for semialgebraic problems. *Mathematical programming*, 96(2):293–320, 2003.

[PEDJ15] Matthew Philippe, Ray Essick, Geir Dullerud, and Raphaël M Jungers. Stability of discrete-time switching systems with constrained switching sequences. *arXiv preprint arXiv:1503.06984*, 2015.

[Pip03] Nicholas Pippenger. The inequalities of quantum information theory. *Information Theory, IEEE Transactions on*, 49(4):773–789, 2003.

[PJ08] Pablo A Parrilo and Ali Jadbabaie. Approximation of the joint spectral radius using sum of squares. *Linear Algebra and its Applications*, 428(10):2385–2402, 2008.

[PMJ15] Matthew Philippe, Gilles Millerioux, and Raphaël M Jungers. Deciding the boundedness and dead-beat stability of constrained switching systems. *arXiv preprint arXiv:1512.04887*, 2015.

[Pro97] V Yu Protasov. The generalized joint spectral radius. A geometric approach. *Izvestiya: Mathematics*, 61(5):995, 1997.

[Rez78] Bruce Reznick. Extremal PSD forms with few terms. *Duke Math. J.*, 45(2):363–374, 06 1978.

[Rii06] Soren Riis. Information flows, graphs and their guessing numbers. In *Modeling and Optimization in Mobile, Ad Hoc and Wireless Networks, 2006 4th International Symposium on*, pages 1–9. IEEE, 2006.

[Roc15] Ralph Tyrell Rockafellar. *Convex analysis*. Princeton university press, 2015.

[RS60] Gian-Carlo Rota and W Strang. A note on the joint spectral radius. *Proceedings of the Netherlands Academy*, 1960. 22:379–381.

[Sch13] Rolf Schneider. *Convex bodies: the Brunn–Minkowski theory*. Number 151. Cambridge University Press, 2013.

[Sha48] Claude Elwood Shannon. A mathematical theory of communication. *Bell System Technical Journal*, 27:379–423 and 623–656, July and October 1948.

[Sho87] NZ Shor. Class of global minimum bounds of polynomial functions. *Cybernetics and Systems Analysis*, 23(6):731–734, 1987.

[Stu06] Milan Studený. *Probabilistic conditional independence structures*. Springer Science & Business Media, 2006.

[Tre09] Stephan Trenn. *Distributional differential algebraic equations*. Citeseer, 2009.

[Tut58a] William T Tutte. A homotopy theorem for matroids, I. *Transactions of the American Mathematical Society*, 88(1):144–160, 1958.

[Tut58b] WT Tutte. A homotopy theorem for matroids, II. *Transactions of the American Mathematical Society*, 88(1):161–174, 1958.

[Vam68] P Vamos. On the representation of independence structures. *Unpublished manuscript*, 1968.

[Wel10] Dominic James Anthony Welsh. *Matroid theory*. Courier Corporation, 2010.

[WRDV14] Yu Wang, Nima Roohi, Geir E Dullerud, and Mahesh Viswanathan. Stability of linear autonomous systems under regular switching sequences. In *Decision and Control (CDC), 2014 IEEE 53rd Annual Conference on*, pages 5445–5450. IEEE, 2014.

[Yeu08] Raymond W. Yeung. The Science of Information. In *Information Theory and Network Coding*, Information Technology Transmission Processing and Storage, pages 1–4. Springer US, 2008.

[Yeu12] Raymond W Yeung. *A first course in information theory*. Springer Science & Business Media, 2012.

[Zho02] D-X Zhou. The p-norm joint spectral radius and its applications in wavelet analysis. *AMS IP Studies in Advanced Mathematics*, 25:305–326, 2002.

[Zie95] Günter M Ziegler. *Lectures on polytopes*, volume 152. Springer Science & Business Media, 1995.

[ZY97] Zhen Zhang and Raymond W Yeung. A non-Shannon-type conditional inequality of information quantities. *Information Theory, IEEE Transactions on*, 43(6):1982–1986, 1997.

[ZY98] Zhen Zhang and Raymond W Yeung. On characterization of entropy function via information inequalities. *Information Theory, IEEE Transactions on*, 44(4):1440–1452, 1998.

Rue Archimède, 1 bte L6.11.01, 1348 Louvain-la-Neuve **www.uclouvain.be/epl**

