Generating Unstable Trajectories for Switched Systems via Dual Sum-Of-Squares Techniques

Benoît Legat[*]
Université catholique de Louvain
4 Av. G. Lemaître,
1348 Louvain-la-Neuve,
Belgium
benoit.legat@student.uclouvain.be

Raphaël M. Jungers[†]
ICTEAM
Université catholique de Louvain
4 Av. G. Lemaître,
1348 Louvain-la-Neuve,
Belgium
raphael.jungers@uclouvain.be

Pablo A. Parrilo
Laboratory for Information and Decision Systems
Massachusetts Institute of Technology
77 Massachusetts Avenue
Cambridge MA 02139, USA
parrilo@mit.edu

ABSTRACT

The joint spectral radius (JSR) of a set of matrices characterizes the maximal asymptotic growth rate of an infinite product of matrices of the set. This quantity appears in a number of applications including the stability of switched and hybrid systems. Many algorithms exist for estimating the JSR but not much is known about how to generate an infinite sequence of matrices with an optimal asymptotic growth rate. To the best of our knowledge, the currently known algorithms select a small sequence with large spectral radius using brute force (or branch-and-bound variants) and repeats this sequence infinitely.

In this paper 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. The algorithm naturally extends to the case where the allowable switching sequences are determined by a graph or finite automaton. 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. We provide numerical examples illustrating the good performance of the algorithm.

[*]This research took place in the framework of the MIT-UCL Internship Program supported by the International Lhoist Berghmans Innovation Chair.

[†]R. M. J. is a F.R.S.-FNRS Research Associate. His research is supported by the Belgian Interuniversity Attraction Poles, and by the ARC grant 13/18-054 from Communauté française de Belgique.

Permission to make digital or hard copies of all or part of this work for personal or classroom use is granted without fee provided that copies are not made or distributed for profit or commercial advantage and that copies bear this notice and the full citation on the first page. Copyrights for components of this work owned by others than ACM must be honored. Abstracting with credit is permitted. To copy otherwise, or republish, to post on servers or to redistribute to lists, requires prior specific permission and/or a fee. Request permissions from permissions@acm.org.

HSCC'16, April 12–14, 2016, Vienna, Austria.

© 2016 ACM. ISBN 978-1-4503-3955-1/16/04…$15.00

DOI: http://dx.doi.org/10.1145/2883817.2883821

Keywords

Joint spectral radius; Sum of squares programming; Switched Systems; Path-complete Lyapunov functions.

1. INTRODUCTION

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 [5]. Despite these negative results, the vast range of applications has motivated a wealth of algorithms to approximate this quantity.

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

$$x_k = A_{\sigma_k} x_{k-1}, \quad \sigma_k \in \{1, \dots, m\}.$$

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

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

This definition is independent of the norm used.

The JSR was introduced by Rota and Strang [16] 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 [10] 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. To the best of our knowledge, the currently known algorithms select a small sequence of matrices with high spectral radius using brute force (or branch-and-bound variants) and repeats this sequence infinitely [8, 9, 11].

Parrilo and Jadbabaie [14] give an efficient approximation algorithm for computing the JSR using sum of squares (SOS) programming. In this paper 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.

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) [7]; an example of constrained switched system is given by Example 1 and its automaton is illustrated by Figure 1. Philippe et al. generalize the SOS-based approximation algorithm for the CJSR in [15, Section 3]. Our technique is well suited for analysing the more general constrained systems as well.

We provide 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.

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 $\{1, \ldots, m\}$ and E is a subset of $V \times V \times \{1, \ldots, m\}$. We say that $(u, v, \sigma) \in E$ if there is an edge between node u and node v with label σ.

The following will serve as a running example.

Example 1 (Running example[1])**.** We borrow the example of [15, Section 4]. The set of matrices $\mathcal{M}$ is composed of the following four matrices

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

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.

In Section 2, we explain the SOS-based approximation algorithm for the CJSR and give our new estimate for its accuracy. The new bounds explicitly depend on the allowable transitions, through the graph $G(V, E)$.

In Section 3, we compute the dual of the corresponding SOS program and we give our algorithm for generating a sequence of matrices with an asymptotic growth rate close to the CJSR.

[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 2, Example 4 and Example 6 is simpler than the running example and the results given in those examples can be obtained by hand.

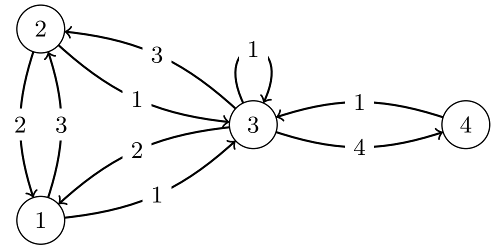


Figure 1: Automaton for the running example. The numbers on the edges are their respective labels.

Notations.

We define the automaton $G^\top(V, E^\top)$ where $E^\top = \{\, (v, u, \sigma) : (u, v, \sigma) \in E \,\}$. We denote as E_k the subset of E^k that represents valid paths of length k. The k-uple $(\sigma_1, \sigma_2, \ldots, \sigma_k)$ is said to be G-*admissible* if $\sigma_1, \ldots, \sigma_k$ are the respective labels of a path of length k. The *arbitrary switching* case, that is, when every tuple is G-admissible, can be seen as the particular case when the automaton has only one node and m self-loops with labels $1, \ldots, m$. We denote the set of all k-uples of $\{1, \ldots, m\}^k$ that are G-admissible as G_k. The sequence $\sigma_1, \sigma_2, \ldots$ is G-admissible (resp. $G^\top$-admissible) if $(\sigma_1, \ldots, \sigma_k)$ (resp. $(\sigma_k, \ldots, \sigma_1)$) is G-admissible for any $k \geq 1$. We denote $A_{\sigma_k} \cdots A_{\sigma_1}$ as A_s where $s = (\sigma_1, \ldots, \sigma_k)$ or s is a path with those respective labels.

We denote 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 $\Sigma_{\mathbf{2d}}^*$.

2. AUTOMATON-DEPENDENT BOUNDS

The definition of the JSR is generalized as follows for constrained systems.

Definition 1 ([7])**.** 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

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

2.1 CJSR Approximation via SOS

A key result used in the SOS-based approximation algorithm for the unconstrained JSR [14] is the following:

Proposition 1 ([16, 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 ([15, 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 1. 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 $i = 1, \ldots, k$ and denote the intermediary nodes of that path as $v_0, v_1, \ldots, v_k$. 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

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

Taking the kth root, the limit $k \to \infty$ and using Definition 1 we obtain the result. □

We relax the positivity condition of Theorem 1 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)**.**

$$\begin{aligned} &\inf_{p_v(x)\in\mathbb{R}_{\mathbf{2d}}[x], \gamma\in\mathbb{R}} \gamma \\ &p_v(x) \text{ is SOS} \quad \forall v \in V && (1) \\ &p_v(x) \text{ is strictly positive} \quad \forall v \in V && (2) \\ \gamma^{2d} p_u(x) - {}&p_v(A_\sigma x) \text{ is SOS} \quad \forall (u, v, \sigma) \in E. \end{aligned}$$

Remark 1. In practice we can replace (1) and (2) 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 2.2 are valid if $p_v(x)$ is in the interior of the SOS cone.

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

$$\rho(G, \mathcal{M}) \leq \rho_{\mathrm{SOS},2d}(G, \mathcal{M}). \tag{3}$$

Example 2. Consider the unconstrained system [2, 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 3. Let us reconsider our running example; see Example 1. The optimal solution of Program 1 is represented by Figure 2 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.

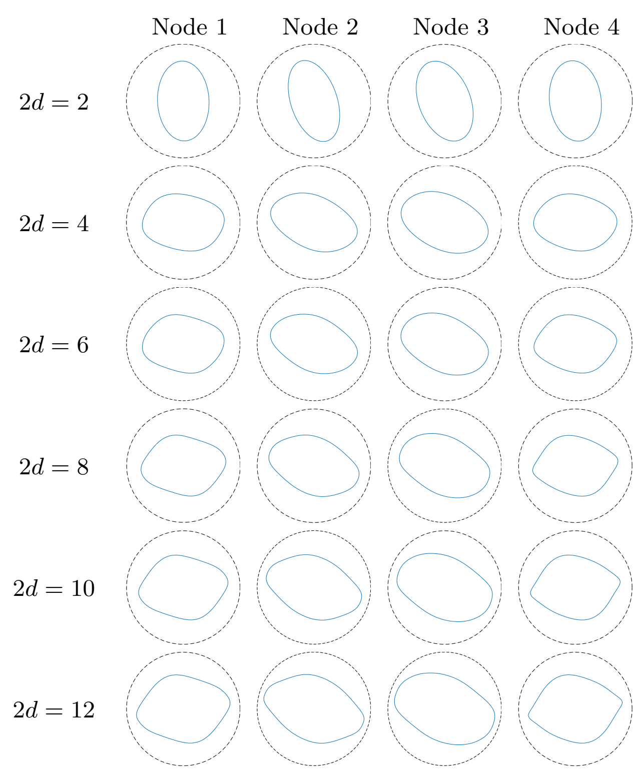


Figure 2: Representation of the solutions to Program 1 with different values of d for the running example. The blue curve represents the boundary of the 1-sublevel set of the optimal solution p_v at each node $v \in V$. The dashed curve is the boundary of the unit circle.

2.2 Approximation guarantees

Philippe et al. show, based on [14], the following accuracy guarantees for Program 1.

Theorem 2 ([15, 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:

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

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 3. 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:

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

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

Remark 2. 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

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

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 $\|A(G)^k\|_\infty$ is the maximum, over all nodes of G, of the number of paths of length k starting at that node. By the Gelfand's formula,

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

Therefore the spectral radius $\rho(A(G))$ represents the "maximal asymptotic growth rate of the number of paths". This is somewhat related to how constrained the automaton is: the more constrained, the better the approximation.

As a corollary of Theorem 3, 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 2, Theorem 3 and (3) are summarized by the following corollary.

Corollary 1. 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:

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

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 3 was already known in this particular case [14, Theorem 4.3]. This section can thus be seen as a generalization of [14, 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 Problem 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

$$\begin{aligned} P_{v,0}(x) &= 0,\\ P_{v,k+1}(x) &= Q_v(x) + \frac{1}{\beta}\sum_{(u,v,\sigma)\in E} P_{u,k}(A_\sigma x), \quad v\in V \end{aligned} \tag{4}$$

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

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

Theorem 4. Consider a constant $\beta > 0$. If there exists homogeneous polynomials $Q_v(x)$ in the interior of the SOS cone such that the iteration (4) converges then

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

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 (4). Since the SOS cone is closed, $P_{v,\infty}$ is SOS. Now by (4), for each $v \in V$,

$$P_{v,\infty}(x) = Q_v(x) + \frac{1}{\beta}\sum_{(u,v,\sigma)\in E} P_{u,\infty}(A_\sigma x)$$

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

$$\beta P_{v,\infty}(x) - P_{u,\infty}(A_\sigma x) = \beta Q_v(x) + \sum_{\substack{(u',v,\sigma')\in E\\ (u',\sigma')\neq(u,\sigma)}} P_{u',\infty}(A_{\sigma'} x)$$

so $\beta P_{v,\infty}(x) - P_{u,\infty}(A_\sigma x)$ is in the interior of the SOS cone. Therefore $(\{\, P_{v,\infty}(x) : v\in V \,\}, \beta^{\frac{1}{2d}})$ is a feasible solution of Program 1. ☐

Remark 3. Before going into the details, we first provide an intuition on how we prove Theorem 3 using the iteration (4). Consider the following iteration (it is not the same as (4) but the sufficient condition of convergence that we develop in Theorem 5 is the same for both iterations)

$$\begin{aligned} P_{v,0}(x) &= Q_v(x),\\ P_{v,k+1}(x) &= \frac{1}{\beta}\sum_{(u,v,\sigma)\in E} P_{u,k}(A_\sigma x), \quad v\in V. \end{aligned}$$

We can see that

$$P_{v,k+1}(x) = \frac{\rho(G,\mathcal{M})^{2dk}}{\beta^k} \sum_{\substack{s\in E_k,\\ s(k+1)=v}} Q_{s(1)}\Big(\frac{A_s}{\rho(G,\mathcal{M})^k}x\Big) \tag{5}$$

where $s(i)$ denotes the ith node of the path s. As $k\to\infty$, $A_s/\rho(G,\mathcal{M})^k$ tends to zero. However, the number of terms in the right-hand side may tend to infinity. It is equal to the number of paths of length k whose last node is v. Therefore $\beta/\rho(G,\mathcal{M})^{2d}$ should be at least the asymptotic value of the kth root of the number of paths of length k. We have seen in Remark 2 that this value is $\rho(A(G))$.

In view of Theorem 4, it is thus natural to analyse under which condition the iteration 4 converges. While we could convert the idea behind Remark 3 into a full proof, we present here a slightly different approach. Recall that iteration 4 is an affine map on the vector space of homogeneous polynomials of degree $2d$. We would like to choose a basis for this vector space to represent the linear operator as a matrix and use a sufficient condition for convergence depending on the spectral radius of the matrix. A simple choice is the monomial basis x^α with α such that $|\alpha| = d$ where x^α denotes $x_1^{\alpha_1}\cdots x_n^{\alpha_n}$ and $|\alpha|$ denotes $\alpha_1+\cdots+\alpha_n$. We choose instead a basis for which each monomial x^α is scaled by the coefficient $\binom{d}{\alpha} = \frac{d!}{\alpha_1!\alpha_2!\cdots\alpha_n!}$ so that the corresponding lift operation for vectors of $\mathbb{R}^n$ that we define below preserves the Euclidean norm (up to power).

Definition 2 (d-lift)**.** Let $N_d = \binom{n+d-1}{d}$. The d-lift of a vector $x \in \mathbb{R}^n$ is the vector $x^{[d]} \in \mathbb{R}^{N_d}$ (indexed by the n-uples $\alpha \in \mathbb{N}^n$ such that $|\alpha| = d$) such that $x^{[d]}_\alpha = \binom{d}{\alpha} x^\alpha$.

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 $N_d \times N_d$ matrix that satisfies $(Ax)^{[d]} = A^{[d]}x^{[d]}$. We also denote $\mathcal{M}^{[d]} \triangleq \{A_1^{[d]},\dots,A_m^{[d]}\}$.

As mentioned above, we have chosen to scale the monomials so that $\|x^{[d]}\|_2 = \|x\|_2^d$ where $\|\cdot\|_2$ is the Euclidean norm. The purpose of this choice is to have the following lemma for the lift operation.

Lemma 1 ([15, Lemma 3.5])**.** Consider a finite set of matrices $\mathcal{M}$ constrained by an automaton G and a positive integer d. The following identity holds:

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

We can now formulate a sufficient condition for the convergence of the iteration 4.

Theorem 5. Let

$$B_{2d} = \sum_{(u,v,\sigma)\in E} (e_u e_v^\top) \otimes A_\sigma^{[2d]}$$

where $\otimes$ is the Kronecker product. If $\beta < \rho(B_{2d})$ then the iteration (4) converges for any homogeneous polynomials $Q_v(x)$ of degree $2d$ in the interior of the SOS cone.

Proof. We can rewrite the polynomials in terms of the basis of scaled monomials. For each $v \in V$, we consider the vector $q_v \in \mathbb{R}^{N_{2d}}$ such that $Q_v(x) = \langle q_v, x^{[2d]}\rangle$. We see that $Q_v(A_\sigma x) = \langle q_v, A_\sigma^{[2d]}x^{[2d]}\rangle = \langle (A_\sigma^{[2d]})^\top q_v, x^{[2d]}\rangle$. The iteration (4) is

$$p_{v,0} = 0,$$
$$p_{v,k+1} = q_j + \frac{1}{\beta}\sum_{u\in V}\Big(\sum_{(u,v,\sigma)\in E} A_\sigma^{[2d]}\Big)^\top p_{u,k}, \quad v \in V$$

which can be rewritten as

$$p_0 = 0, \quad p_{k+1} = q + \frac{1}{\beta} B_{2d}^\top p_k. \tag{6}$$

where the vectors $p_k \in \mathbb{R}^{|V|N_{2d}}$ are (resp. $q \in \mathbb{R}^{|V|N_{2d}}$ is) the concatenation of $p_{1,k},\dots,p_{|V|,k}$ (resp. $q_1,\dots,q_{|V|}$).

The iteration in Equation (6) is an affine iteration, and it is well known that it converges if $\rho(B_{2d}) < \beta$, independently of q. $\square$

Remark 4. If $\rho(B_{2d}) < \beta$, $p_\infty \triangleq \lim_{k\to\infty} p_k$ can be computed by solving the linear system $(I - B_{2d}/\beta)^\top p_\infty = q$.

As a corollary of Theorem 4 and Theorem 5,

$$\rho_{\mathrm{SOS},2d}(G,\mathcal{M}) \le \rho(B_{2d})^{\frac{1}{2d}}. \tag{7}$$

To prove Theorem 3, it remains to find a relation between $\rho(B_{2d})$ and $\rho(G,\mathcal{M})$.

Given a sub-multiplicative norm $\|\cdot\|$ on $\mathbb{R}^{n_1\times n_1}$, consider the sub-multiplicative norm $|||\cdot|||$ on $\mathbb{R}^{n_2 n_1\times n_2 n_1}$ induced by $\|\cdot\|$ defined in [12] as

$$|||M||| \triangleq \max_{1\le i\le N}\sum_{j=1}^{n_2}\|m_{ij}\| \tag{8}$$

for a block $n_2 \times n_2$ matrix $M = (m_{ij})_{i,j=1}^{n_2}$ with $m_{ij} \in \mathbb{R}^{n_1\times n_1}$. We can show that

Lemma 2. Given a norm $\|\cdot\|$ on $\mathbb{R}^{n_1\times n_1}$ and the norm $|||\cdot|||$ defined by (8), for any matrices $A_1,\dots,A_m \in \mathbb{R}^{n_1\times n_1}$ and any nonnegative matrices $B_1,\dots,B_m \in \mathbb{R}^{n_2\times n_2}$, we have

$$\Big|\Big|\Big|\sum_{k=1}^m B_k\otimes A_k\Big|\Big|\Big| \le \Big\|\sum_{k=1}^m B_k\Big\|_\infty \max_{1\le k\le m}\|A_k\|.$$

Proof. Using (8) and the subadditivity of the norm $\|\cdot\|$,

$$\begin{aligned}\Big|\Big|\Big|\sum_{k=1}^m B_k\otimes A_k\Big|\Big|\Big| &= \max_{1\le i\le N}\sum_{j=1}^N\Big\|\sum_{k=1}^m (B_k)_{ij}A_k\Big\| \\ &\le \max_{1\le i\le N}\sum_{j=1}^N\sum_{k=1}^m (B_k)_{ij}\|A_k\| \\ &\le \big(\max_{1\le k\le m}\|A_k\|\big)\max_{1\le i\le N}\sum_{j=1}^N\sum_{k=1}^m (B_k)_{ij} \\ &\le \Big\|\sum_{k=1}^m B_k\Big\|_\infty \max_{1\le k\le m}\|A_k\|.\end{aligned}$$

$\square$

This Lemma allows us to prove the following generalization of [4, Lemma 1].

Lemma 3. For any finite set of matrices $\mathcal{M} = \{A_1,\dots,A_m\}$ of $\mathbb{R}^{n_1\times n_1}$, any nonnegative matrices $B_1,\dots,B_m \in \mathbb{R}^{n_2\times n_2}$ and any automaton $G(V,E)$ such that for each path p that is not G-admissible, $B_p = 0$, we have

$$\rho\Big(\sum_{i=1}^m B_i\otimes A_i\Big) \le \rho\Big(\sum_{i=1}^m B_i\Big)\rho(G,\mathcal{M}).$$

Proof. By Lemma 2,

$$\begin{aligned}\Big\|\Big(\sum_{i=1}^{m} B_i \otimes A_i\Big)^k\Big\| &= \Big\|\sum_{s\in E^k} B_p \otimes A_s\Big\| \\ &= \Big\|\sum_{s\in E_k} B_p \otimes A_s\Big\| \\ &\le \Big\|\sum_{s\in E_k} B_p\Big\|_\infty \max_{s\in E_k} \|A_s\| \\ &= \Big\|\Big(\sum_{i=1}^{m} B_i\Big)^k\Big\|_\infty \max_{s\in E_k} \|A_s\|.\end{aligned}$$

Taking the kth root and the limit of $k \to \infty$ we obtain the result. □

We are now ready to prove Theorem 3.

Proof of Theorem 3. We have

$$\begin{aligned}\rho_{\mathrm{SOS},2d}(G,\mathcal{M}) &\le \rho(B_{2d})^{\frac{1}{2d}} \\ &\le \left[\rho\Big(\sum_{(i,j,\sigma)\in E} e_i e_j^\top\Big)\rho(G,\mathcal{M}^{[2d]})\right]^{\frac{1}{2d}} \\ &= [\rho(A(G))\rho(G,\mathcal{M}^{[2d]})]^{\frac{1}{2d}} \\ &= \rho(A(G))^{\frac{1}{2d}}\rho(G,\mathcal{M}).\end{aligned}$$

where the first inequality is (7), the second is given by Lemma 3, the first equality is given by the fact that

$$\sum_{(i,j,\sigma)\in E} e_i e_j^\top = A(G)$$

and the second equality is given by Lemma 1. □

3. FINDING HIGH-GROWTH SEQUENCES

3.1 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 a SOS polynomials. Barak et al. [3] 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,

$$\langle \widetilde{\mathbb{E}}, p(x)\rangle = \widetilde{\mathbb{E}}[p(x)] = \int_{\mathbb{S}^{n-1}} p(x)\,\mathrm{d}\widetilde{\mu}. \tag{9}$$

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 $\widetilde{\mathbb{E}}[p(x)] \ge 0$ 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 2 (Dual)**.**

$$\sum_{(u,v,\sigma)\in E} \mathcal{A}_\sigma^*\widetilde{\mathbb{E}}_{uv\sigma} \succeq \gamma^{2d} \sum_{(v,w,\sigma)\in E} \widetilde{\mathbb{E}}_{vw\sigma}, \forall v \in V, \tag{10}$$

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

$$\sum_{(u,v,\sigma)\in E} \widetilde{\mathbb{E}}_{uv\sigma}\Big[\sum_{i=1}^{n} x_i^{2d}\Big] = 1. \tag{11}$$

Note that the dual constraint (10) is equivalent to:

$$\sum_{(u,v,\sigma)\in E} \widetilde{\mathbb{E}}_{uv\sigma}[p(A_\sigma x)] \ge \gamma^{2d} \sum_{(v,w,\sigma)\in E} \widetilde{\mathbb{E}}_{vw\sigma}[p(x)] \tag{12}$$

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

Example 4. Consider Example 2. 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 2) we have $\rho_{\mathrm{SOS},2d}(G,\mathcal{M}) = 1$ for any d.

Example 5. We continue the running example; see Example 1 and Example 3.

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 7.

We can see that while the shape of the primal variables change a lot between $2d = 2$ and $2d = 4$ as mentioned in Example 3, 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. $0.229\times\delta_{(0.875,0.485)}$, $0.447\times\delta_{(0.757,-0.653)}$).

3.2 Basic algorithm

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 1, this asymptotic value must be smaller than the CJSR.

Remark 5. The algorithm generates the sequence of matrices in the *reverse order*. We can either see the sequence as G-admissible and left infinite or $G^\top$-admissible and right infinite.

Denote the indegree of a node $v \in V$ as $d^-(v)$ and the maximum indegree of G as $\Delta^-(G) = \max_{v\in V} d^-(v)$. Suppose that we are given a polynomial $p_0(x) \in \operatorname{int}(\Sigma_{\mathbf{2d}})$, i.e. in the interior of the cone of SOS homogeneous polynomials of degree $2d$. 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. By (11), at least one dual variable is nonzero and we can choose one of such edges, say (v_1, v_0, σ_0).

The design of the algorithm is to build a $G^\top$-admissible sequence $(v_1, v_0, \sigma_0), (v_2, v_1, \sigma_1), \dots$ such that

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

remains large for increasing k. Indeed, this would suggest that the product $A_{\sigma_1}\cdots A_{\sigma_k}$ has a large norm. The key idea is that maintaining a large value for θ_k is possible thanks to (12).

Algorithm 1 Producing a sequence of matrices with an asymptotic growth rate close to the CJSR.

Pick an arbitrary polynomial $p_0(x) \in \operatorname{int}(\Sigma_{\mathbf{2d}})$
Pick an edge $(v_1, v_0, \sigma_0) \in E$ such that $\widetilde{\mathbb{E}}_{v_1v_0\sigma_0}[p_0(x)] > 0$
Set $p_1(x) \leftarrow p_0(x)$
for $k = 1, 2, \dots$ **do**
  $(v_{k+1}, v_k, \sigma_k) \leftarrow \arg\max_{(u,v_k,\sigma)\in E} \widetilde{\mathbb{E}}_{uv_k\sigma}[p_k(A_\sigma x)]$
  $p_{k+1} \leftarrow p_k(A_{\sigma_k}x)$
end for

Expanding $p_{k+1}(x)$, we obtain

$$p_{k+1}(x) = p_0(A_{\sigma_1}A_{\sigma_2}\cdots A_{\sigma_k}x).$$

The dual constraint (12) enables us to guarantee the following.

Lemma 4. Consider a finite set of matrices $\mathcal{M}$ constrained by an automaton $G(V, E)$. For any positive integer d, using Program 2 with any $\gamma < \rho_{\mathrm{SOS},2d}(G, \mathcal{M})$, Algorithm 1 produces a $G^\top$-admissible sequence $(v_1, v_0, \sigma_0), (v_2, v_1, \sigma_1), \dots$ that satisfies the following inequality for all $k \geq 1$:

$$\theta_k \geq \frac{\gamma^{2d}}{d^-(v_k)}\theta_{k-1}.$$

Proof. For any integer $k \geq 1$, since $p_0(x)$ is SOS,

$$p_0(A_{\sigma_1}\cdots A_{\sigma_{k-1}}x) \text{ is SOS.}$$

Therefore, by (12),

$$\sum_{(u,v_k,\sigma)\in E} \widetilde{\mathbb{E}}_{uv_k\sigma}[p_0(A_{\sigma_1}\cdots A_{\sigma_{k-1}}A_\sigma x)] \geq \gamma^{2d} \sum_{(v_k,w,\sigma)\in E} \widetilde{\mathbb{E}}_{v_kw\sigma}[p_0(A_{\sigma_1}\cdots A_{\sigma_{k-1}}x)]$$

Since the dual variables $\widetilde{\mathbb{E}}_{v_kw\sigma}$ of the right-hand side are in the dual of the SOS cone, all the terms of the right-hand side are positive. Since one of them is θ_{k-1},

$$\sum_{(u,v_k,\sigma)\in E} \widetilde{\mathbb{E}}_{uv_k\sigma}[p_0(A_{\sigma_1}\cdots A_{\sigma_{k-1}}A_\sigma x)] \geq \gamma^{2d}\theta_{k-1}.$$

Since the left-hand side has $d^-(v_k)$ positive terms and Algorithm 1 picks the term with highest value, the result follows. □

To give a guarantee for the norm of the product of the matrices of the sequence, we introduce the following two Lemmas.

Lemma 5. For any matrix $A \in \mathbb{R}^{n\times n}$ and symmetric positive semidefinite matrix Q, the following inequality holds

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

Lemma 6. For any polynomial $p(x) \in \operatorname{int}(\Sigma_{\mathbf{2d}})$ and any matrix $A \in \mathbb{R}^{n\times n}$, there exists a constant β that does not depend on A such that

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

Moreover we can choose $\beta = \kappa(Q)$ where $Q \in \mathcal{S}^{\binom{n+d}{d}}$ is the symmetric matrix such that $p(x) = (x^{[d]})^\top Qx^{[d]}$ 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 $p(x) \in \operatorname{int}(\Sigma_{\mathbf{2d}})$, Q is positive definite. We can see that

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

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

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

and by Lemma 5,

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

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

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

□

Theorem 6. Consider a finite set of invertible matrices $\mathcal{M}$ constrained by an automaton $G(V, E)$. For any positive integer d, using Program 2 with any $\gamma < \rho_{\mathrm{SOS},2d}(G, \mathcal{M})$, Algorithm 1 produces a $G^\top$-admissible sequence (v_1, v_0, σ_0), $(v_2, v_1, \sigma_1), \dots$ such that

$$\lim_{k\to\infty} \|A_{\sigma_1}\cdots A_{\sigma_k}\|_2^{\frac{1}{k}} \geq \frac{\gamma}{[\Delta^-]^{\frac{1}{2d}}}. \tag{13}$$

Proof. Consider the sequence produced by Algorithm 1. Since the set of edges E is finite, there must be an edge $(\bar{u}, \bar{v}, \bar{\sigma}) \in E$ that appears infinitely many times in the sequence. Let k_1 be the smallest integer such that $(v_{k_1+1}, v_{k_1}, \sigma_{k_1}) = (\bar{u}, \bar{v}, \bar{\sigma})$ and let $q(x) = p_{k_1+1}(x)$. Since the matrices of $\mathcal{M}$ are invertible and $p_0(x) \in \operatorname{int}(\Sigma_{\mathbf{2d}})$, we know that $q(x) \in \operatorname{int}(\Sigma_{\mathbf{2d}})$.

For an arbitrarily large integer K, there exists an integer $k \geq K$ such that $(v_{k_1+k+1}, v_{k_1+k}, \sigma_{k_1+k}) = (\bar{u}, \bar{v}, \bar{\sigma})$. Let $s_K = (\sigma_{k_1+k}, \dots, \sigma_{k_1+1})$. By Lemma 4, we have

$$\widetilde{\mathbb{E}}_{\bar{u}\bar{v}\bar{\sigma}}[q(A_{s_K}x)] \geq \frac{\gamma^{2dk}}{[\Delta^-]^k}\widetilde{\mathbb{E}}_{\bar{u}\bar{v}\bar{\sigma}}[q(x)]. \tag{14}$$

By Lemma 6, there exists a constant β that does not depend on A_{s_K} such that

$$\beta\|A_{s_K}\|_2^{2d}q(x) - q(A_{s_K}x) \text{ is SOS.}$$

Therefore,

$$\widetilde{\mathbb{E}}_{\bar{u}\bar{v}\bar{\sigma}}[\beta\|A_{s_K}\|_2^{2d}q(x)] \geq \widetilde{\mathbb{E}}_{\bar{u}\bar{v}\bar{\sigma}}[q(A_{s_K}x)]$$

and by (14),

$$\beta\|A_{s_K}\|_2^{2d}\widetilde{\mathbb{E}}_{\bar{u}\bar{v}\bar{\sigma}}[q(x)] \geq \frac{\gamma^{2dk}}{[\Delta^-]^k}\widetilde{\mathbb{E}}_{\bar{u}\bar{v}\bar{\sigma}}[q(x)].$$

Since $\widetilde{\mathbb{E}}_{\bar{u}\bar{v}\bar{\sigma}}[q(x)] > 0$,

$$\frac{\gamma^{2dk}}{[\Delta^-]^k} \leq \beta\|A_{s_K}\|_2^{2d}$$

or equivalently

$$\frac{\gamma}{[\beta^{\frac{1}{k}}\Delta^-]^{\frac{1}{2d}}} \leq \|A_{s_K}\|_2^{\frac{1}{k}}.$$

Taking the limit of K to ∞ we obtain the result. $\square$

Example 6. Suppose we apply Algorithm 1 to Example 4 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 4 that $\rho_{\mathrm{SOS},2d}(G, \mathcal{M}) = 1$ for any d. Therefore $\rho(G, \mathcal{M}) = 1$.

3.3 Refinement of the algorithm

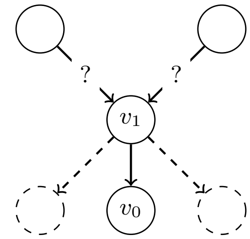
As a consequence of Theorem 6 and Definition 1,

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

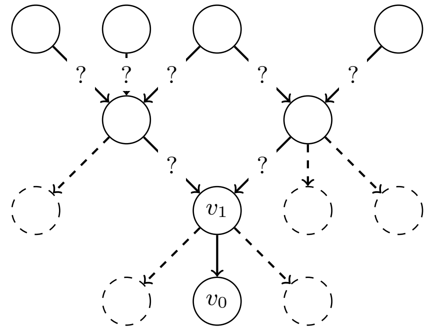
However it is easy to see that $\Delta^- \geq \rho(A(G))$ so Theorem 3 is a stronger statement. This is somewhat surprising, since Algorithm 1 uses more information than the power iteration of Section (2.2) (i.e., the dual solution of the SOS program), so we should be able to improve on it.

Indeed, we give now a generalization of Algorithm 1. The generalized algorithm has an integer parameter l. When $l = 1$, it reduces to Algorithm 1 which has a guarantee of accuracy with Δ^- given by Theorem 6. Increasing l increases the computation time of the generalized algorithm to produce a sequence of the same length but it also improves the guarantee. With $l \to \infty$, it has the same guarantee except that $\rho(A(G))$ replaces Δ^-; see Theorem 7. This section therefore provides an alternative proof of Theorem 3.

Consider the example illustrated by Figure 3. Algorithm 1 considers all the edges ending at v_1 and chooses one of them. This is what is illustrated by Figure 3a. Then it denotes the start of the chosen edge as v_2 and considers all the edges ending at v_2. Instead of doing so, the algorithm could postpone the choice of v_2 and look at all the pairs $(v_3, v_2, \sigma_2), (v_2, v_1, \sigma_1)$ or equivalently, all the paths of length 2 whose last node is v_1 and choose one of them based on the dual variables of the edges (v_3, v_2, σ_2) ignoring the dual variables of the edges (v_2, v_1, σ_1). The intuition behind this is that even if for large enough d the accuracy is arbitrarily good, for small d, we should not follow the dual variables "blindly". Building the sequence with a "larger horizon" is a way to be less affected by the possible imprecision of the dual variables.



(a) One iteration of Algorithm 1.



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

Figure 3: Comparison between Algorithm 1 and Algorithm 2

This idea can be generalized for any length l. To shorten the notation we denote the ith node of a path s as $s(i)$ and the ith label as $s[i]$.

Algorithm 2 Generates sequences of large asymptotic growth using paths of length l.

Pick an arbitrary polynomial $p_0(x) \in \mathrm{int}(\Sigma_{\mathbf{2d}})$
Pick an edge $(v_1, v_0, \sigma_0) \in E$ such that $\widetilde{\mathbb{E}}_{v_1v_0\sigma_0}[p_0(x)] > 0$
Set $p_1(x) \leftarrow p_0(x)$
for $k = 1, 2, \ldots$ **do**
  $(v_{k+l}, \sigma_{k+l-1}, \ldots, \sigma_k, v_k) \leftarrow$
    $\arg\max_{s\in E_l, s(l+1)=v_k} \widetilde{\mathbb{E}}_{s(1)s(2)s[1]}[p_k(A_sx)]$
  $p_{k+1} \leftarrow p_k(A_{\sigma_k}\cdots A_{\sigma_{k+l-1}}x)$
end for

Lemma 4 is generalized into the following Lemma.

Lemma 7. Consider a finite set of matrices $\mathcal{M}$ constrained by an automaton $G(V, E)$. For any positive integer d and l, using Program 2 with any $\gamma < \rho_{\mathrm{SOS},2d}(G, \mathcal{M})$, Algorithm 2 with paths of length l produces a $G^\top$-admissible sequence $(v_1, v_0, \sigma_0), (v_2, v_1, \sigma_1), \ldots$ that satisfies the following inequality for all $k \geq 1$ that is a multiple of l:

$$\theta_k \geq \frac{\gamma^{2dl}}{d_l^-(v_{k-l+1})}\theta_{k-l}.$$

where for a node $v \in V$, $d_l^-(v)$ is the number of paths of length l ending at v.

Proof. For any integer $k \geq 1$ multiple of l, since $p_0(x)$ is SOS, $p_{k-l+1}(x), \ldots, p_k(x)$ are all SOS. Therefore, by (12),

$$\sum_{\substack{s\in E_l,\\ s(l+1)=v_{k-l+1}}} \widetilde{\mathbb{E}}_{s(1)s(2)s[1]}[p_{k-l+1}(A_sx)]$$

$$\geq \gamma^{2d} \sum_{\substack{s\in E_{l-1},\\ s(l)=v_{k-l+1}}} \widetilde{\mathbb{E}}_{s(1)s(2)s[1]}[p_{k-l+1}(A_s x)]$$

$$\vdots$$

$$\geq \gamma^{2dl}\theta_{k-l}$$

Since the first expression has $d_l^-(v_{k-l+1})$ positive terms and Algorithm 2 picks the term with highest value, the result follows. $\square$

Theorem 6 is generalized in the following Theorem.

Theorem 7. Consider a finite set of invertible matrices $\mathcal{M}$ constrained by an automaton $G(V, E)$. For any positive integers d and l and any $\epsilon > 0$, using Program 2 with any $\gamma < \rho_{\mathrm{SOS},2d}(G, \mathcal{M})$, Algorithm 2 with paths of length l produces a $G^\top$-admissible sequence $(v_1, v_0, \sigma_0), (v_2, v_1, \sigma_1), \ldots$ such that

$$\lim_{k\to\infty} \|A_{\sigma_1}\cdots A_{\sigma_k}\|_2^{\frac{1}{k}} \geq \frac{\gamma}{\|[A(G)^\top]^l\|_\infty^{\frac{1}{2dl}}}. \qquad (15)$$

Proof. Consider the sequence produced by Algorithm 2. Since the set of edges E is finite, there must be an edge $(\bar{u}, \bar{v}, \bar{\sigma}) \in E$ that appears infinitely many times in the sequence at multiples of l. Let k_1 be the smallest multiple of l such that $(v_{k_1+1}, v_{k_1}, \sigma_{k_1}) = (\bar{u}, \bar{v}, \bar{\sigma})$ and let $q(x) = p_{k_1+1}(x)$. Since the matrices of $\mathcal{M}$ are invertible and $p_0(x) \in \mathrm{int}(\Sigma_{\mathbf{2d}})$, we know that $q(x) \in \mathrm{int}(\Sigma_{\mathbf{2d}})$.

For an arbitrarily large integer K, there exists an multiple of l $k \geq K$ such that $(v_{k_1+k+1}, v_{k_1+k}, \sigma_{k_1+k}) = (\bar{u}, \bar{v}, \bar{\sigma})$. Let $s_K = (\sigma_{k_1+k}, \ldots, \sigma_{k_1+1})$. By the definition of the matrix norm $\|\cdot\|_\infty$, $\max_{v\in V} d_l^-(v) = \|[A(G)^\top]^l\|_\infty$. Theorefore, by Lemma 7, we have

$$\widetilde{\mathbb{E}}_{\bar{u}\bar{v}\bar{\sigma}}[q(A_{s_K}x)] \geq \frac{\gamma^{2dk}}{\|[A(G)^\top]^l\|_\infty^{k/l}} \widetilde{\mathbb{E}}_{\bar{u}\bar{v}\bar{\sigma}}[q(x)]. \qquad (16)$$

By Lemma 6, there exists a constant β that does not depend on A_{s_K} such that

$$\beta\|A_{s_K}\|_2^{2d} q(x) - q(A_{s_K}x) \text{ is SOS}.$$

Therefore by (16),

$$\beta\|A_{s_K}\|_2^{2d}\widetilde{\mathbb{E}}_{\bar{u}\bar{v}\bar{\sigma}}[q(x)] \geq \frac{\gamma^{2dk}}{\|[A(G)^\top]^l\|_\infty^{k/l}} \widetilde{\mathbb{E}}_{\bar{u}\bar{v}\bar{\sigma}}[q(x)].$$

Hence,

$$\frac{\gamma}{[\beta^{\frac{1}{k}}\|[A(G)^\top]^l\|_\infty^{1/l}]^{\frac{1}{2d}}} \leq \|A_{s_K}\|_2^{\frac{1}{k}}.$$

Taking the limit of K to ∞ we obtain the result. $\square$

Since $\lim_{l\to\infty} \|[A(G)^\top]^l\|_\infty^{\frac{1}{l}} = \rho(A(G)^\top) = \rho(A(G))$, Theorem 3 is a consequence of Theorem 7.

3.4 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 2 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

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

This is the generalization to the constrained case of the first inequality of the three members inequalities; see [10, 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 2.

Gripenberg [8] 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 [9] and the invariant conitope algorithm [11].

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 2 is that it provides a guarantee of accuracy given in Theorem 7. Algorithm 2 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 1 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 1. This is shown by the following example.

Example 7. We first analyse the behaviour of Algorithm 1 on our running example; see Example 1, Example 3 and Example 5. We use the recent CSSystem toolbox [6]. 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 2 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 2 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 [9]. That is, it is a *spectrum maximizing product* (s.m.p.).

With $2d = 10$ and $2d = 12$, we have seen in Example 5 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[2] 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 [1] can be used to improve the lower bounds provided by Algorithm 2 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 [15] and M-path-dependent lift [13]. 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 [15].

4. CONCLUSIONS

We have 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 7, 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? To the best of our knowledge, there are currently two bounds, Theorem 2 only depending on the dimension of the matrices and Theorem 3 only depending on the automaton. Is there a bound that combines both in a more sophisticated way than Corollary 1?

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 1. This motivates other questions with important algorithmic consequences: Does Algorithm 2 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.

[2]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.

5. REFERENCES

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

[2] A. A. Ahmadi and P. A. Parrilo. Joint spectral radius of rank one matrices and the maximum cycle mean problem. In *CDC*, pages 731–733, 2012.

[3] B. Barak, F. G. Brandao, A. W. Harrow, J. Kelner, D. Steurer, and Y. 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.

[4] V. D. Blondel and Y. Nesterov. Computationally efficient approximations of the joint spectral radius. *SIAM Journal on Matrix Analysis and Applications*, 27(1):256–272, 2005.

[5] V. D. Blondel and J. N. Tsitsiklis. The boundedness of all products of a pair of matrices is undecidable. *Systems & Control Letters*, 41(2):135–140, 2000.

[6] L. Cambier, M. Philippe, and R. Jungers. The CSSystem toolbox. http://www.mathworks.com/matlabcentral/fileexchange/52723-the-cssystem-toolbox, August 2015.

[7] X. 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.

[8] G. Gripenberg. Computing the joint spectral radius. *Linear Algebra and its Applications*, 234:43–60, 1996.

[9] N. Guglielmi and M. Zennaro. An algorithm for finding extremal polytope norms of matrix families. *Linear Algebra and its Applications*, 428(10):2265–2282, 2008.

[10] R. Jungers. *The joint spectral radius: theory and applications*, volume 385. Springer Science & Business Media, 2009.

[11] R. M. Jungers, A. Cicone, and N. Guglielmi. Lifted polytope methods for computing the joint spectral radius. *SIAM Journal on Matrix Analysis and Applications*, 35(2):391–410, 2014.

[12] V. Kozyakin. The Berger–Wang formula for the Markovian joint spectral radius. *Linear Algebra and its Applications*, 448:315–328, 2014.

[13] J.-W. Lee and G. E. Dullerud. Uniform stabilization of discrete-time switched and Markovian jump linear systems. *Automatica*, 42(2):205–218, 2006.

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

[15] M. Philippe, R. Essick, G. Dullerud, and R. M. Jungers. Stability of discrete-time switching systems with constrained switching sequences. *arXiv preprint arXiv:1503.06984*, 2015.

[16] G.-C. Rota and W. Strang. A note on the joint spectral radius. *Proceedings of the Netherlands Academy*, 1960. 22:379–381.