Joint spectral characteristics of matrices: a conic programming approach. *

Vladimir Y. Protasov[†], Raphaël M. Jungers, Vincent D. Blondel[‡]

Abstract

We propose a new approach to estimate the joint spectral radius and the joint spectral subradius of an arbitrary set of matrices. We first restrict our attention to matrices that leave a cone invariant. The accuracy of the algorithms, depending on geometric properties of the invariant cone, is estimated. Our algorithms generalize previously known methods which were elaborated for certain particular classes of invariant cones. Then, making use of a lifting procedure, we extend our approach to any set of linear operators, without the common invariant cone assumption. In order to show the good properties of our methods, we consider applications to problems in combinatorics, number theory and discrete mathematics, and improve the state of the art for these problems.

1 Introduction

The joint spectral radius $\hat{\rho}(\mathcal{M})$ of a set of matrices $\mathcal{M}$ is the exponent of the maximal asymptotic growth of a product of matrices taken from this set, when the length of the product grows. The joint spectral subradius $\check{\rho}(\mathcal{M})$ (also called in the literature the lower spectral

*The research of R.J. and V.B. is partially supported by the "Communauté francaise de Belgique - Actions de Recherche Concertées" and by the Belgian Programme on Interuniversity Attraction Poles initiated by the Belgian Federal Science Policy Office. R.J. is a F.R.S.-FNRS Fellow. The research of V.P. is supported by the grants MD-2195.2008.1, by the grant NSh-3233.2008.1 for leading scientific schools and by the RFBR grant No 08-01-00208.

†V.P. is with the Dept. of Mechanics and Mathematics, Moscow State University, Vorobyovy Gory, 119992, Moscow, Russia. E-mail: `v-protassov@yandex.ru`.

‡R.J. and V.B. are with the Division of Applied Mathematics, Université catholique de Louvain, 4 avenue Georges Lemaitre, B-1348 Louvain-la-Neuve, Belgium. E-mail: {`raphael.jungers, Vincent.blondel`}`@uclouvain.be`.

radius) is the minimal growth counterpart. They are formally defined as follows:

$$\begin{aligned}\hat{\rho}(\mathcal{M}) &= \lim_{k\to\infty} \max\left\{\|A_{d_1}\cdots A_{d_k}\|^{1/k} : A_i \in \mathcal{M}\right\}, \qquad (1)\\ \check{\rho}(\mathcal{M}) &= \lim_{k\to\infty} \min\left\{\|A_{d_1}\cdots A_{d_k}\|^{1/k} : A_i \in \mathcal{M}\right\}.\end{aligned}$$

Both these limits exist for all sets of matrices and do not depend on the norm in $\mathbb{R}^n$. In the simplest case, when the set $\mathcal{M}$ consists of one operator A, both the joint spectral quantities become the usual spectral radius $\rho(A)$, which is the largest modulus of eigenvalues of A. This follows from Gelfand's formula $\rho(A) = \lim_{k\to\infty} \|A^k\|^{1/k}$.

The joint spectral radius appeared in [42], the joint spectral subradius in [24]. They have found numerous applications in various areas: in the control of switched systems [3, 28], in subdivision algorithms for approximation and curve design (see [20] for many references), in the study of wavelets and of refinement equations [15,39], in probability theory [36], and in many problems of discrete mathematics, graph theory and combinatorics. See [25] for a survey on these quantities.

The issue of *computing the joint spectral radius* has also been studied in many works. Several negative results exist, showing the difficulty of this problem. We recall them in Theorems 1 and 2 below. Nevertheless, due to its practical importance, many authors have proposed and analyzed different methods of computation or approximation of the joint spectral radius. The first algorithms that appeared were based on exhaustion of matrix products in a special way [15, 22]. For some sets of matrices they can work fast, but no theoretical estimate of the rate of convergence is available. Afterwards several algorithms with theoretically guaranteed rate of convergence were elaborated. Most of them construct a common Lyapunov function in order to approximate the joint spectral radius of the matrices. This function can be constructed with polytopes [23, 33], ellipsoids [2, 10], or "sums of squares" [32]. Also, lifting techniques have been proposed to improve these results [9, 34, 43]. The computational complexity of these methods grows dramatically with the dimension of the matrices. For large dimensions they cannot provide us with good approximation of the joint spectral radius within a reasonable computation time, although they can work fast in many special cases. This is not a surprise because of the negative results detailed below. These results, however, leave us the opportunity, at least theoretical, to come up with algorithms which work efficiently *not for all matrices but for some classes.*

On the other hand, all the algorithms we mentioned only deal with the joint spectral radius $\hat{\rho}$. As for the subradius $\check{\rho}$, to the best of our

knowledge, no algorithm is known, although this notion is also very important in applications. In this paper we consider sets of matrices that have a common invariant cone. These sets of matrices are important for several reasons.
First, special properties of such operators are well known (such as Perron-Frobenius theory; see also [27] for properties related to joint spectral quantities).
Moreover, this case appears very often in applications. For instance, the trackability of sensor networks, the capacity of codes [8, 31], or the analysis of repetition-free languages in combinatorics on words [6, 26] involve nonnegative matrices (for which the positive orthant $\mathbb{R}^n_+ = \{x \in \mathbb{R}^n \mid x_1, \ldots, x_n \geq 0\}$ is a common invariant cone).
Finally as we will see, a simple lifting procedure allows to obtain a set of matrices that leaves a cone invariant, given an arbitrary set of matrices. In view of this, there is no loss of generality to analyze such sets.

For any cone and any set of matrices leaving this cone invariant we introduce the notions of *joint conic radius* and *joint conic subradius.* We show that, first, these notions are both efficiently computable in the framework of conic programming, and, second, they can be used to approximate the joint spectral radius and subradius respectively. This idea was inspired by recent works where the joint conic radius was proposed as an approximation for the joint spectral radius in the special case where the invariant cone is the positive orthant [13, 26].

Outline. In this paper we extend the notion of joint conic radius to an arbitrary cone and introduce its analogue for the subradius (Section 3). Then we prove the main relations between the spectral and conic radii (Theorems 4 and 7), and study their dual analogues (Subsection 3.3). Iterating these relations we come to approximation algorithms for the joint spectral radius and subradius. These results are presented in Section 4, where theoretical estimates of the rate of convergence of the algorithms are also derived. Moreover, applying the lifting procedure we extend our approach to all matrices, possibly without invariant cone. (Note, however, that for the subradius an additional condition will be necessary, so that no nontrivial rate of convergence will be available for arbitrary matrices.) In particular, we shall see that the "best ellipsoidal norm algorithm" (see [10]) can be viewed as a particular case of this general framework. Finally, in section 5 we apply our algorithms to well-known problems in different areas of number theory or combinatorics, namely the asymptotics of the overlap-free language, the density of ones in Pascal's rhombus, and the analysis of Euler's partition function. The dimension of the matrices are from 5 to 20. We shall see that for such sizes, our methods provide a satisfying accuracy. The proof of some technical, or not fundamental results presented in this paper are to be found in an appendix.

2 Notation and auxiliary results

Let us recall some important results on the joint spectral quantities. For any $k \in \mathbb{N}$ we denote by $\mathcal{M}_k$ the set of all products of length k of matrices taken in $\mathcal{M}$:

$$\mathcal{M}_k \quad = \quad \Big\{ A_{d_1} \cdots A_{d_k} \;\Big|\; A_{d_j} \in \mathcal{M},\ j = 1, \cdots, k \Big\}.$$

The two following facts are well-known:

Proposition 1. *With the notations above, for any set of matrices $\mathcal{M}$ and any $k \in \mathbb{N}$,*

$$\hat{\rho}(\mathcal{M}) \quad = \quad [\hat{\rho}(\mathcal{M}_k)]^{1/k}, \qquad \check{\rho}(\mathcal{M}) \quad = \quad [\check{\rho}(\mathcal{M}_k)]^{1/k}.$$

Proposition 2. *With the notations above, for any submultiplicative norm $||\cdot||$ one has*

$$\max\{\rho(A) : A \in \mathcal{M}_k\} \;\le\; \hat{\rho}(\mathcal{M})^k \;\le\; \max\{||A|| : A \in \mathcal{M}_k\},$$

and

$$\check{\rho}(\mathcal{M})^k \;\le\; \min\{\rho(A) : A \in \mathcal{M}_k\} \;\le\; \min\{||A|| : A \in \mathcal{M}_k\}.$$

For the joint spectral radius an equivalent definition exists, which is very convenient in practice.

Proposition 3. *[4] For any bounded set $\mathcal{M}$ such that $\hat{\rho}(\mathcal{M}) \neq 0$, the joint spectral radius can be defined as*

$$\hat{\rho}(\mathcal{M}) = \inf_{||\cdot||} \sup_{A \in \mathcal{M}} \{||A||\}.$$

The joint spectral quantities of the set $\mathcal{M}$, have motivated intense research over the past decades (see [25] for a survey). We have the following negative results:

Theorem 1. *[5,12] The problem of determining, given a set of matrices Σ, if the semigroup generated by Σ is bounded is Turing-undecidable.*
The problem of determining, given a set of matrices Σ, if $\hat{\rho}(\Sigma) \le 1$ is Turing-undecidable.
These two results remain true even if Σ contains only nonnegative rational entries.

Theorem 2. *[11, Theorem 2] Fix any $K > 0$ and $0 < L < 1$. An algorithm providing the value $\tilde{\rho}$ as an approximation of the joint spectral subradius $\check{\rho}$ of a given set is said to be a (K, L)-approximation algorithm if $|\tilde{\rho} - \check{\rho}| < K + L\check{\rho}$.*

- *There exists no (K, L)-approximation algorithm for computing the joint spectral subradius of an arbitrary set Σ.*
- *For the special cases where Σ consists of two integer matrices with binary entries, there exists no polynomial time (K, L)-approximation algorithm for computing the subradius unless $P = NP$.*

Despite these discouraging facts, many researchers have proposed approximation algorithms for the joint spectral radius, because of its growing interest in practical applications. Some recent papers applied semidefinite programming techniques to get approximations of the joint spectral radius. It has been shown independently by several authors [2, 10] that computing the best ellipsoidal norm for a set of matrices can be done efficiently, and it provides an effective upper bound for the joint spectral radius. By "best ellipsoidal norm" we mean the minimal $\gamma > 0$ such that the following SDP program has a solution:

$$\begin{aligned} A_i^T P A_i &\preceq \gamma^2 P \quad \forall A_i \in \mathcal{M} \\ P &\succ 0. \end{aligned} \tag{2}$$

Indeed, it is equivalent to say that

$$\gamma = \max\{||A|| : A \in \mathcal{M}\},$$

where $||\cdot||$ is the matrix norm induced by the vector norm $|x| = \sqrt{x^T P x}$. These norms are precisely the norms whose unit balls are ellipsoids. So, the above equation, together with Proposition 2 yield $\hat{\rho}(\mathcal{M}) \leq \gamma$. The next theorem tells us how tight this bound is:

Theorem 3. *[2, 10] For an arbitrary set of m matrices $\mathcal{M} \subset \mathbb{R}^{n\times n}$, the best ellipsoidal norm approximation $\hat{\rho}^*$ of its joint spectral radius $\hat{\rho}$ satisfies*

$$\max\left\{\frac{1}{\sqrt{n}}, \frac{1}{\sqrt{m}}\right\}\hat{\rho}^* \quad \leq \quad \hat{\rho} \quad \leq \quad \hat{\rho}^*.$$

Moreover, due to Proposition 1 this procedure can be iterated in such a way that one obtains an algorithm providing an arbitrary accuracy within a computation time that can be computed a priori. Indeed, if $\hat{\rho}^*$ is the ellipsoidal norm approximation for $\mathcal{M}_k$, then

$$\frac{1}{n^{1/2k}}\,\hat{\rho}^{*1/k} \quad \leq \quad \hat{\rho} \quad \leq \quad \hat{\rho}^{*1/k}.$$

More recently these techniques were generalized in the framework of Sum-of-Squares computation [32], improving the accuracy of the estimate. Finally, [26] proposes similar semidefinite methods to get approximations of the joint spectral subradius. However, no bound on the accuracy of these methods have been provided.

3 Main results

Let $K \subset \mathbb{R}^n$ be a convex cone. In the following by "cone" we mean a closed, pointed, nondegenerate cone (possessing a nonempty interior) and with the apex at the origin. Any cone defines a corresponding order in $\mathbb{R}^n$: we write $x \geq y$ ($x > y$) for $x - y \in K$ (respectively $x - y \in \mathrm{int}K$). A matrix A possesses an invariant cone K if $AK \subset K$. In this case we say that A is *nonnegative* and write $A \geq 0$. If K is invariant for all matrices of some set $\mathcal{M}$, then it is called an invariant cone of that set.

Definition 1. *For a given compact set $\mathcal{M}$ with an invariant cone K we consider the values*

$$\begin{aligned} \hat{\sigma}_K(\mathcal{M}) &= \inf\Big\{\lambda \geq 0 \quad \Big| \quad \exists\, v > 0 \quad Av \;\leq\; \lambda v \quad \forall A \in \mathcal{M}\Big\}, \\ \check{\sigma}_K(\mathcal{M}) &= \sup\Big\{\lambda \geq 0 \quad \Big| \quad \exists\, v \geq 0, v \neq 0 \quad Av \;\geq\; \lambda v \quad \forall A \in \mathcal{M}\Big\} \end{aligned} \tag{3}$$

and call them the joint conic radius *and the* joint conic subradius *respectively.*

These values depend not only on the set $\mathcal{M}$, but also on the cone $K \subset \mathbb{R}^n$. In the sequel we assume the cone K to be fixed, and use the short notation $\hat{\sigma}(\mathcal{M}), \check{\sigma}(\mathcal{M})$ or just $\hat{\sigma}, \check{\sigma}$, if it is clear which set $\mathcal{M}$ is considered.

These quantities are defined only if the set $\mathcal{M}$ admits an invariant cone. However, if this is not the case a simple procedure allows one to obtain a set of matrices that admits an invariant cone, without changing fundamentally the joint spectral quantities:

Proposition 4. *[9] Let $\mathcal{M}$ be a set of matrices. The semidefinite lifting $\tilde{\mathcal{M}}$ of $\mathcal{M}$:*

$$\tilde{\mathcal{M}} = \{\tilde{A} : \mathbb{R}^{n^2} \to \mathbb{R}^{n^2} X \to A^T X A\} \tag{4}$$

leaves the cone $\mathcal{K}_n$ of symmetric positive semidefinite matrices invariant. Moreover it satisfies $\hat{\rho}(\tilde{\mathcal{M}}) = \hat{\rho}(\mathcal{M})^2, \check{\rho}(\tilde{\mathcal{M}}) = \check{\rho}(\mathcal{M})^2$.

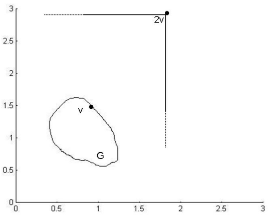

Figure 1: The constant α. For the positive orthant in $\mathbb{R}^2$, $\alpha = 1/2$. That is, for an arbitrary convex set in $\mathbb{R}^2_+$, one is ensured that there exists a point $v \in G$ such that for all $x \in G$, $2v \geq x$.

3.1 The joint spectral radius

In this section we find a relation between the joint spectral radius and the joint conic radius. We start with two simple lemmas. Let us have a compact set $\mathcal{M}$ of matrices in $\mathbb{R}^n$. In the following we call *convex body* a convex compact set with a nonempty interior. To estimate the joint spectral quantities we use the following simple fact.

Lemma 1. *[38] If there exists a convex body $P \subset \mathbb{R}^n$ such that $AP \subset \lambda P$ for all $A \in \mathcal{M}$, then $\hat{\rho}(\mathcal{M}) \leq \lambda$. If there exists a closed set $Q \subset \mathbb{R}^n$ such that $0 \notin Q$ and $AQ \subset \lambda Q$ for all $A \in \mathcal{M}$, then $\check{\rho}(\mathcal{M}) \geq \lambda$.*

As a corollary we obtain

Lemma 2. *Let a set $\mathcal{M}$ possess an invariant cone K. If for some $v \in \mathrm{int}K$ we have $Av \leq \lambda v$ for all $A \in \mathcal{M}$, then $\hat{\rho} \leq \lambda$. If for some $v \in K \setminus \{0\}$ we have $Av \geq \lambda v$ for all $A \in \mathcal{M}$, then $\check{\rho} \geq \lambda$.*

Proof. For the first assertion we apply Lemma 1 for the body $P = (v - K) \cap (-v + K)$. The second one follows from the same Lemma 1 for the set $Q = v + K$. $\square$

In order to find an approximation relation between $\hat{\rho}$ and $\hat{\sigma}$ we introduce the following geometrical characteristic of convex cones:

Definition 2. *For a given cone $K \subset \mathbb{R}^n$ the value $\alpha(K)$ is the largest number such that for any convex compact set $G \subset K$ there exists $v \in G$, for which $\frac{1}{\alpha} v \geq G$.*

In other words, for any $\gamma < \alpha(K)$ and for any convex compact set $G \geq 0$ there is $u \geq G$ such that $\gamma u \in G$. Clearly, α is an affine invariant of a cone.

Theorem 4. *For any set $\mathcal{M}$ of operators with an invariant cone K we have*

$$\alpha\,\hat{\sigma} \quad \le \quad \hat{\rho} \quad \le \quad \hat{\sigma},$$

where $\alpha = \alpha(K), \hat{\sigma} = \hat{\sigma}(\mathcal{M}), \hat{\rho} = \hat{\rho}(\mathcal{M})$.

Proof. The inequality $\hat{\rho} \le \hat{\sigma}$ follows directly from Lemma 2. To prove the inequality $\alpha\hat{\sigma} \le \hat{\rho}$ recall (Proposition 3) that for any $q > \hat{\rho}$ there is a norm in $\mathbb{R}^n$ such that the corresponding matrix norm of each matrix in $\mathcal{M}$ is smaller than q. Denote by G the intersection of the unit ball of that norm with the cone K; then $AG \subset qG$ for any $A \in \mathcal{M}$. On the other hand, for any $\gamma < \alpha$ there is $u \ge G$ such that $\gamma u \in G$. Observe that in this case $\operatorname{int} G \neq \emptyset$, hence $u > 0$. It follows that $A(\gamma u) \in AG \subset qG$, and so $Au \in \frac{q}{\gamma}G \le \frac{q}{\gamma}u$. Thus, $Au \le \frac{q}{\gamma}u$ for all $A \in \mathcal{M}$. Whence $\hat{\sigma} \le \frac{q}{\gamma}$. This holds for arbitrary $q > \hat{\rho}$ and $\gamma < \alpha$, therefore $\hat{\sigma} \le \frac{\hat{\rho}}{\alpha}$, which concludes the proof.

□

Let us add that this theorem had already been proved in the particular case of nonnegative matrices [13].

Theorem 5. *For any cone $K \subset \mathbb{R}^n$ we have* $\alpha(K) \ \ge \ \frac{1}{n}$.

In the proof we use several facts of convex geometry. For a given convex body $G \subset \mathbb{R}^d$ and for any point $z \in \operatorname{int} G$ we consider the Minkowski-Radon constant

$$\tau_z(G) \quad = \quad \inf \Bigl\{ t > 0 \ \Bigm| \ \exists\, x, y \, \in \, \partial G \, , \ z \, = \, tx + (1-t)y \Bigr\}.$$

In other words, $\tau_z(G)$ is the minimal possible ratio $\frac{|z-y|}{|z-x|}$, where x, y are the points of intersection of a line passing through z with the boundary of G. By $\tau(G)$ we denote the value $\tau_z(G)$ for the point $z = \operatorname{gr} G = \frac{1}{\mathrm{Vol} G}\int_G x dx$ (the center of gravity of G), where $\mathrm{Vol} G = \int_G dx$ is the volume of G. The well-known Minkowski-Radon theorem [40] states that $\tau(G) \ge \frac{1}{d}$ for any convex body G. For d-dimensional simplices Δ one has $\tau_z(\Delta) = \frac{1}{d}$ for $z = \operatorname{gr}\Delta$ and $\tau_z(\Delta) < \frac{1}{d}$ for all other points $z \in \Delta$. Another well-known fact will be formulated in the following Lemma. We call a hyperplane H a *plane of support* for the convex set G if $H \cap \partial G \neq \emptyset$ and G lies in one of the closed half-spaces with respect to H.

Lemma 3. *If a convex compact set G lies in a cone $K \subset \mathbb{R}^n$ and contains at least one interior point of K, then there is a hyperplane of support H for G such that*

1) H does not separate the set G from 0*;*

2) the cross-section $H \cap K$ is bounded and has its center of gravity in G.

For the proof it suffices to consider the set $G' = \mathrm{Conv}\{G, 0\}$ and to choose a hyperplane H that cuts from the cone K a convex body that contains G' and has the smallest possible volume. See [1, p. 229] for details.

Proof. of Theorem 5. Let G be an arbitrary convex subset of a cone K. If it does not intersect the interior of K, then it lies on a face of that cone, which is also a cone of a smaller dimension. Invoking the induction argument, we get $\alpha(K) \geq \frac{1}{n-1} > \frac{1}{n}$. Assume now that $G \cap \mathrm{int} K \neq \emptyset$. Applying Lemma 3, we obtain a hyperplane of support H, which does not separate G from the origin. Let us denote $S = K \cap H$ and $v = \mathrm{gr}\, S \in G$. Let $u = nv$. The set $S' = (u - K) \cap H$ is homothetic to S with respect to its center of gravity v with the factor $-(n-1)$, i.e., $S' = v-(n-1)(S-v)$. By the Minkowski-Radon theorem $\tau(S') \geq \frac{1}{n-1}$, therefore $S \subset S'$. Since the hyperplane H separates u from G, it follows that for any $x \in G$ the segment $[x, u]$ intersects H, i.e., it intersects the set S. Thus, for any $x \in G$ the segment $[x, u]$ also intersects S'. This means that $G \subset (u - K)$, and so $u \geq G$. Since $\frac{1}{n}\, u = v \in G$, it now follows that $\alpha(K) \geq \frac{1}{n}$. $\square$

Thus, to estimate the joint spectral radius by the value $\hat{\sigma}(\mathcal{M})$ one needs to compute $\alpha(K)$ for the invariant cone of $\mathcal{M}$. Theorem 5 guarantees that $\alpha \geq \frac{1}{n}$ for any cone in $\mathbb{R}^n$. Therefore, we always have

$$\frac{1}{n}\, \hat{\sigma} \quad \leq \quad \hat{\rho} \quad \leq \quad \hat{\sigma}\,.$$

For some cones we have better bounds. In the following theorem we find precise values of α for three important cases: for n-hedral cones (cones bounded by n hyperplanes passing through the origin) for the cone $\mathcal{K}_n$ of symmetric positive semidefinite $n \times n$-matrices, and for the Lorentzian (or Euclidean, of right spherical) cone.

Theorem 6. *For any n-hedral cone in $\mathbb{R}^n$ we have $\alpha = \frac{1}{n}$; for the Lorentzian cone $\alpha = \frac{1}{2}$; for the cone $\mathcal{K}_n$ of positive semidefinite $n \times n$-matrices we have $\alpha = \frac{1}{n}$.*

The proof is to be found in Appendix A.

3.2 The joint spectral subradius

A straightforward application of Lemma 2 yields the inequality $\check{\sigma} \leq \check{\rho}$. However, in contrast to Theorem 4, there is no inverse estimate. More precisely, there is no positive function of the cone $C(K)$ such that $\check{\rho}(\mathcal{M}) \leq C(K)\check{\sigma}(\mathcal{M})$ for any set of matrices $\mathcal{M}$ that leaves K invariant. We give an example for the case $K = \mathbb{R}^n_+$, and the reader will easily generalize this construction for an arbitrary cone.

Example 1. Let $\mathcal{M} = \{A_1, \ldots, A_n\} \subset \mathbb{R}^n$, where A_j is a matrix whose entries of the jth row are all zeros, and all other entries are ones. Since $A_i A_j = (n-1)A_i$ for all i, j, it follows that $A_{d_1} \cdots A_{d_m} = (n-1)^{m-1} A_{d_1}$ for any product of length m. Hence $\check{\rho} = n - 1$. On the other hand, $\check{\sigma}(\mathcal{M}) = 0$. Indeed, any nonnegative vector $v \neq 0$ has at least one positive coordinate v_j, while $(A_j v)_j = 0$. Hence the inequality $A_j v \geq \lambda v$ implies $\lambda = 0$.

Thus, to obtain an inverse inequality for the subradii we need to impose some extra conditions for the matrices. This situation can actually not be avoided, since Theorem 2 tells us that in general there is no approximation algorithm for the subradius. Example 1 shows that an invariant cone does not suffice. It appears, however, that the existence of a second invariant cone does the job. We start with introducing some more notation. Let us have a cone $K \subset \mathbb{R}^n$. We say that a convex closed cone K' is *embedded in* K if $(K' \setminus \{0\}) \subset \mathrm{int}K$. In this case we call $\{K, K'\}$ an *embedded pair.* Note that the embedded cone K' may be degenerate, i.e., may have an empty interior. An embedded pair $\{K, K'\}$ is called an *invariant pair* for a matrix A if the cones K and K' are both invariant for A. The same is for invariant pairs of a set $\mathcal{M}$.

Definition 3. *For a given embedded pair $\{K, K'\}$ the value $\beta(K, K')$ is the smallest number such that for any line intersecting K and K' by segments $[x, y]$ and $[x', y']$ respectively (with $[x, x'] \subset [x, y']$) one has $1 \leq \frac{|x-y'|}{|x-x'|} \leq \beta$.*

Lemma 4. *Let $\mathcal{M}$ be a set of matrices with an invariant embedded pair (K, K') such that $\check{\rho}(\mathcal{M}) > 0$. For any $p \in [0, \check{\rho}]$ there is a closed convex set $Q \subset K'$ (that may be unbounded) not containing the origin, such that $AQ \subset pQ$ for any $A \in \mathcal{M}$.*

Proof. We suppose without loss of generality that $\check{\rho}(\mathcal{M}) = 1$. If this is not the case we can just scale the matrices by dividing by $\check{\rho}$. Take $v_0 \in K'$. We define the set

$$Q = \mathrm{Conv}\{\lambda A v_0 : A \in \mathcal{M}_k, k \in \mathbb{N}, \lambda \geq 1\}.$$

Since the cone K' is invariant, $Q \subset K'$, and obviously $AQ \subset Q$. Hence, for all $p < 1$, $AQ \subset pQ$. It remains to show that Q does not contain the origin. If this were the case, we could define a series of matrices $A_k \in \mathcal{M}_k : k \in \mathbb{N}$ such that $A_k v_0 \to 0$. However, since $v_0 \in \mathrm{int}K$, this implies that $||A_k|| \to 0$, and so $\check{\rho}(\mathcal{M}) < 1$, which is in contradiction with the assumptions. $\square$

Theorem 7. *For any set $\mathcal{M}$ with an invariant pair $\{K, K'\}$ we have*

$$\breve{\sigma} \quad \leq \quad \breve{\rho} \quad \leq \quad \beta\breve{\sigma},$$

where $\beta = \beta(K, K'), \breve{\sigma} = \breve{\sigma}_K(\mathcal{M}), \breve{\rho} = \breve{\rho}(\mathcal{M})$.

Proof. The inequality $\breve{\sigma} \leq \breve{\rho}$ follows from Lemma 2. To prove that $\breve{\rho} \leq \beta\breve{\sigma}$ we apply Lemma 4 and get a set Q such that $AQ \subset pQ\, \forall A \in \mathcal{M}$.

Draw any hyperplane of support for Q that separates Q from the origin and makes a bounded cross-section of the cone K. Denote this hyperplane by H, let $S = H \cap K, S' = H \cap K'$, and $v \in H \cap Q$. Let us show that $v \leq \beta Q$ (the order is with respect to the cone K). For any ray on H starting at the point v and meeting the boundaries of K' and K at points x' and x respectively one has $\frac{|x-v|}{|x-x'|} \leq \beta$. Hence the set homothetic to S with the factor $\left(1 - \frac{1}{\beta}\right)$ with respect to v contains the set S'. This yields that $S' \subset \left(\frac{1}{\beta}v + K\right)$, therefore $Q \subset \left(\frac{1}{\beta}v + K\right)$, which means $\frac{1}{\beta}v \leq Q$ and hence $v \leq \beta Q$. Since $AQ \subset pQ$, it follows that $\frac{1}{\beta}v \leq \frac{1}{p}AQ$. On the other hand, $v \in Q$, whence $\frac{1}{\beta}v \leq \frac{1}{p}Av$ for any $A \in \mathcal{M}$. This means that $\breve{\sigma} \geq \frac{p}{\beta}$. Taking a limit $p \to \breve{\rho}$, we get $\beta\breve{\sigma} \geq \breve{\rho}$, from which the theorem follows.

□

Remark 1. A simple compactness argument shows that $\beta < \infty$ for any embedded pair. If the cone K is fixed, then the value $\beta(K, K')$ is nondecreasing in the second variable, i.e., if $K'_1 \subset K'_2$, then $\beta(K, K'_1) \leq \beta(K, K'_2)$. The smallest possible value $\beta = 1$ is attained precisely when $\dim K' = 1$, i.e., when K' is a ray. If, on the other hand, a sequence of cones $\{K'_j\}_{j\in\mathbb{N}}$ approaches the boundary of K, that is there are $x_j \in K'_j$ and $x \in \partial K, x \neq 0$ such that $x_j \to x$ as $j \to \infty$, then $\beta(K, K'_j) \to +\infty$.

Now we compute the values $\beta(K, K')$ for several important cases of embedded pairs. For a given point $x = (x_1, \ldots, x_n) \in \mathbb{R}^n_+$ we denote by $x_{\min}$ and $x_{\max}$ its smallest and greatest entries respectively. Similarly, for $X \in \mathcal{K}_n$, $\lambda_{\min}$ and $\lambda_{\max}$ denote the smallest and greatest eigenvalues. Let us recall that $\lambda_{\max} = \max_{|u|=1}(Xu, u)$ and $\lambda_{\min} = \min_{|u|=1}(Xu, u)$. We write K_φ for the Lorentzian cone of angle $\varphi < \frac{\pi}{2}$.

Proposition 5. *If* $\mathbb{R}^n_{+,c} = \left\{x \in \mathbb{R}^n_+ \;\middle|\; x_{\max} \leq c x_{\min}\right\}$, *then*

$$\beta(\mathbb{R}^n_+, \mathbb{R}^n_{+,c}) = c^2;$$

if $\mathcal{K}_{n,c} = \left\{X \in \mathcal{K}_n \;\middle|\; \lambda_{\max} \leq c\lambda_{\min}\right\}$, *then*

$$\beta(\mathcal{K}_n, \mathcal{K}_{n,c}) = c^2;$$

for two coaxial Lorentzian cones K_φ *and* K_ψ $(\varphi > \psi)$ *we have*

$$\beta(K_\varphi, K_\psi) = \Big(\frac{\sin\varphi}{\sin(\varphi-\psi)}\Big)^2.$$

The proof is to be found in appendix B.

Note that if all entries of a matrix A are positive and in each column the ratio between the greatest and the smallest elements does not exceed c, then $A(\mathbb{R}^n_+) \subset \mathbb{R}^n_{+,c}$. This yields, in particular, that A possesses an invariant cone $K' = A(\mathbb{R}^n_+)$ contained in $\mathbb{R}^n_{+,c}$.

Corollary 1. *If all matrices of a set* $\mathcal{M}$ *are positive and in each column of any matrix the ratio between the greatest and the smallest elements does not exceed* c, *then* $\mathcal{M}$ *has an invariant pair, for which* $\beta \le c^2$. *This pair is* $K = \mathbb{R}^n_+, K' = \mathrm{conv}\big\{A(\mathbb{R}^n_+) \bigm| A \in \mathcal{M}\big\}$.

We say that a matrix A with an invariant cone is positive and write $A > 0$ if the cone AK is embedded in K. In this case $\{K, AK\}$ is an invariant pair for A, for which $\beta < \infty$. A set $\mathcal{M}$ is positive if it consists of positive matrices. Since $\mathcal{M}$ is compact, we see that the cone $K' = \mathrm{conv}\big\{AK \bigm| A \in \mathcal{M}\big\}$ is embedded in K, and so all sets of positive matrices admit a constant $\beta < \infty$:

Corollary 2. *If a set of matrices* $\mathcal{M}$ *has an invariant cone* K *such that for all* $A \in \mathcal{M}$, $AK \subset \mathrm{int}K$, *then it has an embedded pair* (K, K') *such that* $\beta(K, K') < \infty$.

3.3 Dual families for improving the accuracy

For a given set $\mathcal{M}$ we denote by $\mathcal{M}^*$ the dual set, which consists of the adjoint matrices. It is well known that the families $\mathcal{M}$ and $\mathcal{M}^*$ have the same joint spectral radius $\hat\rho$ [25], whereas their conic radii $\hat\sigma(\mathcal{M})$ and $\hat\sigma(\mathcal{M}^*)$ are a priori different. Consider the following example:

Example 2. Denote e the column vector whose all entries are equal to one, e_i the column vector whose all entries are equal to zero, except for the ith one which is equal to one, and denote A_i the matrix whose all entries are equal to zero, except the ith row, whose entries are all equal to one:

$$A_i = e_i e^T.$$

The set $\mathcal{M} = \{A_i\}$ has a joint spectral radius equal to one. Indeed, the maximum column-sum is a norm, and is equal to one for all matrices in $\mathcal{M}$. Moreover, the conic radius $\hat\sigma(\mathcal{M})$ is equal to n. In conclusion, the lower bound

$$\hat\sigma(\mathcal{M})/n \quad \le \quad \hat\rho$$

is tight in this case.

However, $\hat{\sigma}(\mathcal{M}^T) = 1 = \hat{\rho}(\mathcal{M})$, and the estimate gives the exact value of the joint spectral radius.

Therefore, combining both values $\hat{\sigma}(\mathcal{M})$ and $\hat{\sigma}(\mathcal{M}^*)$ to estimate the joint spectral radius by Theorem 4, we can obtain better results. That is, one could hope that the following equation holds:

$$(1/f(n)) \min \{\hat{\sigma}(\mathcal{M}), \hat{\sigma}(\mathcal{M}^T)\} \quad \leq \quad \hat{\rho}(\mathcal{M}), \tag{5}$$

where $f(n) < n$. Unfortunately, it appears that the growth $f(n) \approx n$ cannot be avoided, as shown in the next example:

Example 3. Denote e, e_i and $\{A_i\}$ as in Example 2. The set

$$\mathcal{M}' \quad = \quad \left\{ \begin{pmatrix} A_i & 0 \\ 0 & A_j^T \end{pmatrix} \in \mathbb{R}^{2n \times 2n} \quad : \qquad 1 \leq i, j \leq n \right\}$$

has joint spectral radius equal to one. Indeed, the matrices are block diagonal, and so obviously

$$\hat{\rho}(\mathcal{M}') \;=\; \max \{\hat{\rho}(\{A_i\}), \hat{\rho}(\{A_j^T\})\} \;=\; 1.$$

Moreover,

$$\min \{\hat{\sigma}(\mathcal{M}'), \hat{\sigma}(\mathcal{M}'^T)\} \quad = \quad n.$$

The previous example proves the following proposition:

Proposition 6. *The function $f(n)$ giving the accuracy of the joint conic radius in (5) cannot be chosen smaller than $n/2$, where n is the dimension of the matrices, even if one computes the joint conic radius of both the matrices and their transposed.*

The same is for the subradii $\check{\rho}$ and $\check{\sigma}$: For some sets $\mathcal{M}$ taking the transpose helps improving the estimates significantly. As an example, consider once more the set from Example 1 we have $\check{\rho} = n - 1$ and $\check{\sigma}(\mathcal{M}) = 0$. So, the inequality giving the upper bound $\check{\rho} \leq \beta\check{\sigma}(\mathcal{M})$ holds only for $\beta = +\infty$. However, taking the transposes of the matrices, we immediately get $\check{\sigma}(\mathcal{M}^*) = n - 1$, hence in this case we have the best possible situation: the upper bound $\check{\rho} \leq \beta\check{\sigma}(\mathcal{M}^*)$ holds for $\beta = 1$. To see this we observe that the set $\mathcal{M}^*$ has an invariant pair with $K = \mathbb{R}^n_+$ and K' is the one-dimensional cone spanned by the vector of ones e. For this pair $\beta(K, K') = 1$, and by Theorem 7 $\check{\rho} = \check{\sigma}$. However, a construction similar to Example 3 shows that in general one cannot hope to improve the bounds by considering $\max \{\check{\sigma}(\mathcal{M}), \check{\sigma}(\mathcal{M}^*)\}$.

Also, one could ask the same questions for other invariant cones. For instance, by applying the semidefinite lifting (4) to the matrices in $\mathcal{M}$ one gets a set of matrices $\tilde{\mathcal{M}}$ that leaves $\mathcal{K}_n$ invariant. Thus, we could as well apply the lifting to $\mathcal{M}^*$ to get another estimate. Even though the obtained matrices are not the transposed of the initial ones, this estimate will not be better, as shown by the next proposition.

Proposition 7. *Let $\mathcal{M} \subset \mathbb{R}^{n\times n}$ be a set of matrices, and $\tilde{\mathcal{M}} \subset \mathcal{K}_{n(n-1)/2}$ be the semidefinite lifting of $\mathcal{M}$. Then $\hat{\sigma}(\tilde{\mathcal{M}}) = \hat{\sigma}(\tilde{\mathcal{M}^*})$, that is, applying the semidefinite lifting to $\mathcal{M}$ or $\mathcal{M}^*$ does not change the quality of approximation of $\hat{\sigma}$.*

The proof is to be found in Appendix C.

4 Computing the joint spectral quantities

In this section we analyse the following approximation problem: for a given set $\mathcal{M}$ we need to find numbers $\hat{\rho}_*$ and $\check{\rho}_*$ such that $\big|\hat{\rho}_* - \hat{\rho}\big|/\hat{\rho} \le \varepsilon$ and $\big|\check{\rho}_* - \check{\rho}\big|/\check{\rho} \le \varepsilon$. If $\mathcal{M}$ has an invariant cone with a parameter α (or an invariant pair with a parameter β), then $\mathcal{M}_k$ has the same cone (or the same invariant pair). Applying now Theorem 4 and Theorem 7 for the set $\mathcal{M}_k$ we obtain

Corollary 3. *If a set $\mathcal{M}$ has an invariant cone K with parameter α, then for any $k \in \mathbb{N}$*

$$\alpha^{1/k}\big[\hat{\sigma}(\mathcal{M}_k)\big]^{1/k} \quad \le \quad \hat{\rho}(\mathcal{M}) \quad \le \quad \big[\hat{\sigma}(\mathcal{M}_k)\big]^{1/k}. \tag{6}$$

If, in addition, $\mathcal{M}$ has another invariant cone K' embedded in K, then

$$\big[\check{\sigma}(\mathcal{M}_k)\big]^{1/k} \quad \le \quad \check{\rho}(\mathcal{M}) \quad \le \quad \beta^{1/k}\big[\check{\sigma}(\mathcal{M}_k)\big]^{1/k}, \tag{7}$$

where $\beta = \beta(K, K')$.

This result ensures that $\hat{\rho}_* = [\hat{\sigma}(\mathcal{M}_k)]^{1/k}$ gives the desired accuracy ε, whenever $k \ge \frac{\ln \frac{1}{\alpha}}{\varepsilon}$. Moreover, by Theorem 5 $\alpha \ge \frac{1}{n}$ for any cone, and hence for any set with an invariant cone it suffices to take $k \ge \frac{\ln n}{\varepsilon}$. For the joint spectral subradius we take $\check{\rho}_* = [\check{\sigma}(\mathcal{M}_k)]^{1/k}$. This gives the desired approximation for $k \ge \frac{\ln \beta}{-\ln(1-\varepsilon)}$. Note that this value does not exceed $\frac{\ln \beta}{\varepsilon}$. Therefore, to compute the joint spectral subradius it suffices to take $k \ge \frac{\ln \beta}{\varepsilon}$. Let us remember that the parameter β depends on the cones and, in contrast to α, cannot be uniformly estimated for all cones in $\mathbb{R}^n$.

Thus, the problem is reduced to finding the values $\hat{\sigma}(\mathcal{M}_k)$ and $\check{\sigma}(\mathcal{M}_k)$ for a given set of matrices $\mathcal{M}$. We describe the computational issue for finite sets of matrices. Moreover, for the sake of simplicity we restrict ourselves to the case of two matrices $\mathcal{M} = \{A_0, A_1\}$. All the results can easily be generalized to arbitrary finite sets of matrices. We consider separately the joint spectral radius and subradius, and distinct the cases of sets with invariant cones and arbitrary sets.

4.1 Matrices with an invariant cone

We describe practical implementations of our algorithms in two cases: polyhedral cones K (given by systems of homogeneous linear inequalities), and the positive semidefinite cone $\mathcal{K}_n$ in the $\frac{n(n+1)}{2}$ -dimensional space $\mathcal{S}_n$ of symmetric $n \times n$-matrices.

The joint spectral radius. Let a set $\mathcal{M} = \{A_0, A_1\}$ have an invariant cone K. Let this cone be polyhedral, i.e. $K = \big\{x \in \mathbb{R}^n \;\big|\; (a_j, x) \geq 0,\; j = 1, \ldots, m\big\}$. In this case both values $\hat{\sigma}(\mathcal{M}_k)$ and $\check{\sigma}(\mathcal{M}_k)$ can be found by a routine linear programming procedure. For the sake of simplicity we describe it in the case $m = n$, when $a_j = e_j$ (the basis vectors), and K becomes the positive orthant $\mathbb{R}^n_+$. So, both matrices A_0, A_1 have nonnegative entries. Consider the following problem:

$$\begin{array}{lll} \min & r & \\ \text{s.t.} & & \\ x & \geq & e, \\ Ax & \leq & rx, \quad \forall A \in \mathcal{M}_k. \end{array} \tag{8}$$

For each r the feasibility can be checked by solving an LP problem with n variables and $n(2^k+1)+1$ constraints. The minimal $\bar{r}$ can be found by a bisection technique within logarithmic time (with respect to the accuracy), and satisfies $\bar{r} = \hat{\sigma}(\mathcal{M}_k)$. Thus, solving the minimization problem (8) we get the desired approximation of the joint spectral radius.

A very similar algorithm can be used to compute the joint spectral radius for a set of matrices $\mathcal{M} = \{\mathcal{A}_0, \mathcal{A}_1\}$ acting in the space $\mathcal{S}_n$ of symmetric $n \times n$-matrices with the invariant cone $\mathcal{K}_n$. The value $\hat{\sigma}_n(\mathcal{M}_k)$ can be found by solving the following SDP program:

$$\begin{array}{lll} \min & r & \\ \text{s.t.} & & \\ X & \succeq & I, \\ \mathcal{A}X & \preceq & rX, \quad \mathcal{A} \in \mathcal{M}_k. \end{array} \tag{9}$$

For each r the feasibility is checked by solving an SDP problem in the dimension n. The minimal r again can be found by bisection techniques. Since by Theorem 6 $\alpha(\mathcal{K}_n) = \frac{1}{n}$, the number k needed to approximate the joint spectral radius with a given relative precision is the same as for the previous case of nonnegative matrices in $\mathbb{R}^n$.

The joint spectral subradius. We again consider a set of two matrices $\mathcal{M} = \{A_0, A_1\}$ with an invariant polyhedral cone K, and for the sake of simplicity we describe it in the case of the positive orthant $\mathbb{R}^n_+$. To use Theorem 7 we need to impose an extra assumption that there is an embedded invariant cone K'. This will be the case, for example, if the matrices A_0 and A_1 have positive entries (Corollary 1). The value $\check{\sigma}(\mathcal{M}_k)$ is a solution of the following problem:

$$\begin{array}{lcll} \max & & r & \\ \text{s. t.} & & & \\ x & \geq & 0, & \\ Ax & \geq & rx, & A \in \mathcal{M}_k, \\ (x, e) & = & 1. & \end{array} \tag{10}$$

Note that in the definition of $\check{\sigma}$ the vector x does not have to be strictly positive, hence the first constraint is just $x \geq 0$. The reason of the last constraint is to avoid the trivial zero solution. For each r the feasibility is checked by solving an LP problem. The maximal r can be found by bisection.

If a set of linear operators $\mathcal{M}$ acting on $\mathcal{S}_n$ leaves the cone $\mathcal{K}_n$ invariant we can approximate its joint spectral subradius by solving the following program:

$$\begin{array}{lcll} \max & & r & \\ \text{s. t.} & & & \\ X & \succeq & 0, & \\ \mathcal{A}X & \succeq & rX, & \mathcal{A} \in \mathcal{M}_k, \\ \mathrm{tr}X & = & 1. & \end{array} \tag{11}$$

For each r the feasibility is checked by solving an SDP problem in the dimension n. Let us add that in many practical cases our method of approximation of spectral radii works far faster than predicted by the theoretical results (see Section 5).

4.2 Arbitrary matrices

If one does not know of any common invariant cone for the set of matrices $\mathcal{M}$, he could apply Proposition 4 in order to obtain a set that leaves $\mathcal{K}_n$ invariant. Applying the lifting to an arbitrary set $\mathcal{M} = \{A_0, A_1\}$ we come to the special case of algorithm (9) for the cone $\mathcal{K}_n$, when $\mathcal{A}_i X = A_i^T X A_i,\ i = 0, 1$.

The joint spectral radius. Solving the problem

$$
\begin{array}{lll}
\min & & r \\
\text{s.t.} & & \\
X & \succeq & I, \\
A^T X A & \preceq & rX, \quad A \in \mathcal{M}_k\,,
\end{array}
\tag{12}
$$

we find $\bar{r} = \hat{\sigma}(\tilde{\mathcal{M}}_k)$. Since $\hat{\rho}(\tilde{\mathcal{M}}) = \big[\hat{\rho}(\mathcal{M})\big]^2$, invoking Corollary 3 and taking into account that $\alpha(\mathcal{K}_n) = \frac{1}{n}$ (Theorem 6), we obtain $(\bar{r})^{\frac{1}{2k}} n^{-\frac{1}{2k}} \le \hat{\rho}(\tilde{\mathcal{M}}) \le (\bar{r})^{\frac{1}{2k}}$. Hence, the value $(\bar{r})^{\frac{1}{2k}}$ approximates $\hat{\rho}(\mathcal{M})$ with the relative precision $\varepsilon \le \frac{\ln n}{2k}$. This is nothing else but the method of the ellipsoidal norm for computing the joint spectral radius, presented in [2, 10]. In this sense, the ellipsoid method is an important special case of algorithm (9) in the space $\mathcal{S}_n$. We have deduced the main inequality and the estimate of the accuracy using the conic radius, which is a totally different way than in [2, 10].

The joint spectral subradius. In contrast to the case of the joint spectral radius, the lifting has never been implemented in the literature to compute the joint spectral subradius until very recently [26]. Applying Corollary 3 to the set $\tilde{\mathcal{M}}$, we come to the following method. Solving the problem

$$
\begin{array}{lll}
\max & & r \\
\text{s. t.} & & \\
X & \succeq & 0, \\
A^T X A & \succeq & rX, \quad A \in \mathcal{M}_k, \\
\mathrm{tr} X & = & 1,
\end{array}
\tag{13}
$$

we find $\bar{r} = \check{\sigma}(\tilde{\mathcal{M}}_k)$. Since $\check{\rho}(\tilde{\mathcal{M}}) = \big[\check{\rho}(\mathcal{M})\big]^2$, we see that $(\bar{r})^{\frac{1}{2k}} \le \check{\rho}(\mathcal{M}) \le (\bar{r})^{\frac{1}{2k}} \beta^{\frac{1}{2k}}$. Hence, the value $(\bar{r})^{\frac{1}{2k}}$ approximates $\check{\rho}(\mathcal{M})$ with the relative precision $\varepsilon \le 1 - \beta^{-1/2k} \le \frac{\ln \beta}{2k}$.

4.3 Exact computation of the joint spectral quantities in special cases

Theorems 4 and 7 make it possible not only to estimate the joint spectral quantities, but also to find their precise values in some favorable cases. To formulate the next result we need some further notation. Let us recall that by the well-known Perron-Frobenius theorem any linear operator A in $\mathbb{R}^n$ that has an invariant cone $K \subset \mathbb{R}^n$ possesses an eigenvector $v \in K$ such that $Av = \rho(A)\, v$. Any such eigenvector (it may not be unique) is called a *Perron-Frobenius eigenvector* of the operator A.

Proposition 8. *Let a set $\mathcal{M}$ possess an invariant cone K, let also $k \in \mathbb{N}, A \in \mathcal{M}_k$, and $v \in K$ be the Perron-Frobenius eigenvector of the matrix A; then*

a) if $v \in \mathrm{int}K$ and there is $m \in \mathbb{N}$ such that $C(A-B)v \in K$ for all $B \in \mathcal{M}_k, C \in \mathcal{M}_m$, then $\hat{\rho}(\mathcal{M}) = [\rho(A)]^{1/k}$;

b) if there is $m \in \mathbb{N}$ such that $C(B-A)v \in K$ for any $B \in \mathcal{M}_k, C \in \mathcal{M}_m$, then $\check{\rho}(\mathcal{M}) = [\rho(A)]^{1/k}$.

Remark 2. *Note that in the special case $m = 0$ the proposition only requires that $rv \geq Bv\, \forall B \in \mathcal{M}_k$ (resp. $rv \leq Bv$).*

Proof. With possible multiplication by a constant it can be assumed that $\rho(A) = 1$, and hence $Av = v$. Let K_m be the closure of the set $\big\{x \in \mathbb{R}^n \setminus \{0\} \big| Cx \in K \quad \forall C \in \mathcal{M}_m\big\}$. Clearly, K_m is an invariant cone of $\mathcal{M}$ containing K. Define the order in $\mathbb{R}^n$ by the cone K_m and the constants $\hat{\sigma}$ and $\check{\sigma}$ by formulas (3) with respect to this cone. Since $(A-B)v = v - Bv \in K_m$, it follows that $Bv \leq v$ for all $B \in \mathcal{M}_k$, and therefore $\hat{\sigma}(\mathcal{M}_k) \leq 1$. Theorem 4 now implies that $\hat{\rho}(\mathcal{M}_k) \leq 1$. Since $\hat{\rho}(\mathcal{M}) = [\hat{\rho}(\mathcal{M}_k)]^{1/k} \geq [\rho(A)]^{1/k} = 1$, we have $\hat{\rho}(\mathcal{M}) = 1$, from which the assertion (a) follows. Furthermore, if $(B-A)v = Bv - v \geq 0$ for all $B \in \mathcal{M}_k$, then $\check{\sigma}(\mathcal{M}_k) \geq 1$. Hence by Theorem 7 $\check{\rho}(\mathcal{M}_k) \geq 1$. Thus, $1 \;= \rho(A) \geq \min_{B \in \mathcal{M}_k} \rho(B) \geq \check{\rho}(\mathcal{M}_k) \geq 1$. Thus, $\check{\rho}(\mathcal{M}_k) = 1$ and so $\check{\rho}(\mathcal{M}) = 1$, which proves (b). □

Remark 3. This proposition provides sufficient conditions for the joint spectral radius to attain its value at some finite product $A \in \mathcal{M}_k$. For each m, starting with $m = 0$, we verify that $C(A-B)v \in K$ for all $B \in \mathcal{M}_k, C \in \mathcal{M}_m$. If we succeed in finding such m, then the conjecture is approved and $\hat{\rho}(\mathcal{M}) = [\rho(A)]^{1/k}$. For instance, in case of nonnegative matrices, when $K = \mathbb{R}^n_+$, one needs to check that all the vectors $C(A-B)v$ are nonnegative. Note that if v is not the Perron eigenvector of A, the equation $C(A-B)v = C(rI - B)v \in K$ (resp. $C(B-A)v = C(B-rI)v \in K$) still proves that $\hat{\rho} \leq r^{1/k}$, (resp. $\check{\rho} \geq r^{1/k}$) and can then be useful in its own right (see Section 5.2 for an application).

5 Applications

5.1 Overlap-free words

In this section we briefly describe applications where techniques developed in this paper prove useful. We have chosen applications in number theory, because this field has provided many sets of matrices that can be used as test benches nowadays. We start with a recent application: the computation of the asymptotics of overlap-free words. This problem arises in combinatorics on words (for an introduction to combinatorics on words, see [29]), where one is interested in the number u_l of binary overlap-free words of length l. An overlap is a word on the alphabet $\{a, b\}$ of the form $xuxux$, where x is a or b, and u is a word that can be empty. For instance, the word *baabaab* is an overlap. An overlap-free word is a word that does not contain any overlap. Let $r^- = \liminf \frac{\log u_n}{\log n}$ and $r^+ = \limsup \frac{\log u_n}{\log n}$. The following result ([26], see also [7, 14]) allows to express the asymptotics of u_l in terms of the joint spectral quantities.

Theorem 8. *There exist two nonnegative matrices $A_0, A_1 \in \{0, 1, 2\}^{20\times 20}$ such that*

$$r^+ = \log_2 \hat{\rho}(\{A_0, A_1\}),$$
$$r^- = \log_2 \check{\rho}(\{A_0, A_1\}).$$

Thanks to this result, the following accurate estimates appear in [26]:

$$1.2690 \ < \ r^- \ < \ 1.2736 \qquad \text{and} \qquad 1.3322 \ < \ r^+ \ < 1.3326.$$

The inequality $1.2690 < r^-$ was obtained from Theorem 7. Unfortunately, no embedded invariant pair is known for F_0 and F_1, and so it is not possible to obtain an upper bound on r^- with this result. However, it appears that the product $F_1^{10} F_0$ satisfies:

$$r^- \leq \log_2 \left[(\rho(A_1^{10} A_0)^{1/11} \right] = 1.2735...$$

One can verify numerically that this product gives the best possible upper bound among all the matrix products of length less than 14. The upper bound on r^+ can be found by solving the semidefinite program (12) with $k = 14$, while the lower bound is obtained from the simple inequality

$$\hat{\rho} \quad \geq \quad \left[\rho(A_0 A_1)\right]^{1/2} = 2.5179... \tag{14}$$

Remark that the accuracy of this estimate is 0.0003. As we have seen in Section 4.2, in order to ensure such an accuracy, one has to solve the semidefinite program 12 with $k = \ln(n)/(2 \cdot 0.0003) \approx 5000$, which is, of course, enormous. However, Equation (14) shows that the actual cost for obtaining such an accuracy is much lower.

5.2 Pascal's rhombus

Recently, the question of the density of ones in Pascal's rhombus arose in number theory [21]. Pascal's rhombus is a variation of the well-known Pascal's triangle in which each term is equal to the sum of four earlier terms, rather than two. The coefficients in Pascal's rhombus arise from a linear recurrence relation on polynomials: Define $p_0(x) = 1$, $p_1(x) = 1 + x + x^2$, and

$$p_n(x) = (1 + x + x^2)p_{n-1}(x) + x^2 p_{n-2}(x).$$

In [19] the authors show that this leads to a recurrence relation for the value $v(n)$ of the number of odd coefficients in $p_n(x)$. In turn, it is shown that the asymptotic growth of $v(n)$ is in relation with the joint spectral quantities of the following set of matrices:

$$\Sigma = \left\{ \begin{pmatrix} 0 & 1 & 0 & 0 & 0 \\ 1 & 0 & 2 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & 1 \\ 0 & 0 & 0 & 2 & 1 \end{pmatrix}, \begin{pmatrix} 1 & 0 & 2 & 0 & 0 \\ 0 & 0 & 0 & 2 & 1 \\ 1 & 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 \end{pmatrix} \right\}. \tag{15}$$

More precisely,

$$\limsup \frac{\log v_n}{\log n} = \log_2 \hat{\rho}(\{A_0, A_1\}),$$
$$\liminf \frac{\log v_n}{\log n} = \log_2 \check{\rho}(\{A_0, A_1\}).$$

In [19], the authors mention the difficulty to have any kind of estimate for $\check{\rho}(\Sigma)$. It has been conjectured later [18] that $\check{\rho}(\Sigma) = (1 + \sqrt{5})/2 = 1.61803\ldots$ These matrices leave the positive orthant $K = \mathbb{R}^5_+$ invariant, and so, it is possible to apply the program (10) to obtain a lower bound on $\check{\rho}$. It appears, however, that this algorithm does not provide a better lower bound than the trivial value 1. Nevertheless, when applied to the transposed matrices, the algorithm works very well: We obtained the vector $x = (0.196, 0.229, 0.190, 0.190, 0.196)$, which is such that $C(B - (1.618)^{12} I)x \in \mathbb{R}^+$ for any $B \in \mathcal{M}_{12}, C \in \mathcal{M}_6$. This implies (see Remark 3) that $\check{\rho}(\mathcal{M}) \geq 1.618$, which is extremely close to the conjectured value. Note that a good upper bound on $\check{\rho}$ can be obtained as follows: $\check{\rho} \leq \rho(A_0^3 A_1^3)^{(1/6)} = 1.6376$. This is the smallest averaged spectral radius among all products of length less or equal to 18.

Concerning the joint spectral radius of these matrices, the methods developed in this paper are not necessary, as it is easy to prove that $\hat{\rho}(\Sigma) = 2$.

5.3 Euler's binary partition function and generalizations

The binary partition function is a longstanding research topic in number theory. For a given $d \in \mathbb{N} \cup \{\infty\}$ the binary partition function $b_{2,d}(k)$ is defined as the total number of different binary expansions $k = \sum_{j=0}^{\infty} d_j 2^j$, where the "digits" d_j take values from the set $\{0, 1, \dots, d-1\}$. For $d = 2$, obviously, $b(k) \equiv 1$. For $d \geq 3$ the value $b_{2,d}(k)$ grows as $k \to \infty$, and the problem is to find the exponents of this asymptotic growth. For various d this problem was studied by L.Euler [17], K.Mahler [30], N.G. de Bruijn [16], B.Reznick [41], and others. There are certain relations of this problem with the theory of refinement equations and subdivision algorithms [37].

The *generalized partition function* $b_{m,d}(k)$ is defined similarly as the total number of different m-adic expansions $k = \sum_{j=0}^{\infty} d_j m^j$, $d_j \in \{0, 1, \dots, d-1\}$.

Recently, it has been shown [35] that the asymptotic behavior of $b_{m,d}(k)$ as $k \to \infty$ is ruled by the joint spectral quantities of certain sets of matrices $\Sigma_{m,d}$, with binary (0 or 1) entries. More precisely, for all pairs $(m, d) \in \mathbb{N}^2$, there exist constants $C_1, C_2, \lambda_1, \lambda_2$ such that the following holds:

$$C_1\, k^{\lambda_1} \quad \leq \quad b_{m,d}\,(k) \quad \leq \; C_2\, k^{\lambda_2}. \tag{16}$$

Denoting $\hat{\rho}_{m,d}$ and $\check{\rho}_{m,d}$ respectively the joint spectral radius and subradius of $\Sigma_{m,d}$, we have the relations

$$\lambda_1 = \log_m \hat{\rho}_{m,d},$$

$$\lambda_2 = \log_m \check{\rho}_{m,d}.$$

In [35], these joint spectral quantities are deeply analysed for $m = 2$ and for some small values of d. To the best of our knowledge no numerical analysis has been done for other d and for $m \geq 3$. Since the set $\Sigma_{m,d}$ consists of binary matrices it has an invariant cone (the positive orthant), hence we can apply algorithms from Section 4. Take, for example, $m = 3, d = 14$. It follows from [35] that

$$\Sigma_{3,14} = \left\{ \begin{pmatrix} 1&1&1&1&1&0&0\\ 0&1&1&1&1&0&0\\ 0&1&1&1&1&1&0\\ 0&1&1&1&1&1&0\\ 0&0&1&1&1&1&0\\ 0&0&1&1&1&1&1\\ 0&0&1&1&1&1&1 \end{pmatrix}, \begin{pmatrix} 1&1&1&1&1&0&0\\ 1&1&1&1&1&0&0\\ 0&1&1&1&1&0&0\\ 0&1&1&1&1&1&0\\ 0&1&1&1&1&1&0\\ 0&0&1&1&1&1&0\\ 0&0&1&1&1&1&1 \end{pmatrix}, \begin{pmatrix} 1&1&1&1&0&0&0\\ 1&1&1&1&1&0&0\\ 1&1&1&1&1&0&0\\ 0&1&1&1&1&0&0\\ 0&1&1&1&1&1&0\\ 0&1&1&1&1&1&0\\ 0&0&1&1&1&1&0 \end{pmatrix} \right\}.$$

The set $\Sigma_{3,14}$ leaves the cone $\mathbb{R}^n_{+,2} = \{x \in \mathbb{R}^n_+ \mid x_{\max} \le 2x_{\min}\}$, invariant, and thus, combining Proposition 5 with Corollary 3, we have the following bounds on the accuracy of the subradius approximation:

$$\left[\check{\sigma}(\mathcal{M}_k)\right]^{1/k} \quad \le \quad \check{\rho}(\mathcal{M}) \quad \le \quad 4^{1/k}\left[\check{\sigma}(\mathcal{M}_k)\right]^{1/k}. \qquad (17)$$

For $k = 9$ the ratio between the upper and lower bound is $4^{1/9} = 1.1665...$ The algorithm with $k = 9$ provides

$$4.525 \ \le \quad \check{\rho}.$$

Note that $\check{\rho} \le \rho(A_0A_1)^{1/2} = 4.6105$, and so the actual ratio with $k = 9$ is at most $4.6105/4.525 = 1.02$, which is already quite sharp and much better than the predicted 1.1665.

For the joint spectral radius, applying the algorithm with $k = 9$, we find an upper bound equal to 4.8. Note that $\rho \ge \rho(A_1A_2)^{1/2} = 4.72$. Hence, the approximation ratio is actually equal to $4.8/4.72 = 1.02$, which is once again far better than the theoretical ratio $7^{1/9} = 1.24$ provided by Corollary 3. Thus, for $m = 3, d = 14$ we have

$$4.525 \ \le \ \check{\rho}(\Sigma_{3,14}) \ \le \ 4.6105\ ; \qquad 4.72 \ \le \ \hat{\rho}(\Sigma_{3,14}) \ \le \ 4.8$$

Let us consider two other examples of pairs (m, d).

For $m = 3, d = 7$ we have three 3×3-matrices. Our method with $k = 6$ gives:

$$2.4142 \ \le \ \hat{\rho}(\Sigma_{3,7}) \ \le \ 2.416\,.$$

The joint spectral subradius is known to be equal to 2 in this case.

For $m = 4, d = 15$ we have four 5×5-matrices. Our method with $k = 6$ gives:

$$3.7 \ \le \ \check{\rho}(\Sigma_{4,15}) \ \le \ 3.7321\ ; \qquad 3.791287 \ \le \ \hat{\rho}(\Sigma_{4,15}) \ \le \ 3.791288\,.$$

6 Conclusion

In this paper we have pursued several goals:

First, if the joint spectral radius has received much attention in the last decades, and if several algorithms have been proposed that approximate this quantity, to the best of our knowledge we provided here the first approximation algorithm for the joint spectral subradius, which allows moreover for a certified accuracy for certain classes of matrices.

Second, we proposed a framework (conic optimization) that presents many of the known algorithms for the joint spectral radius, and provides simple proofs of their convergence rate. This framework also sheds interesting light on the previously known methods.

Third, we presented these algorithms on several examples, in order to show some empirical facts that seem important: the algorithms performs usually far better than predicted, and some tricks are sometimes determinant to make the difference between computability and impossibility of getting an approximation. A good example of such a trick is the transposition of the matrices: the example in Section 5.2 shows that effect. About this transposition trick, we showed moreover (Section 3.3) that it does not provide a better general accuracy for our algorithms. In practice, our results allow to find accurate estimates very rapidly. As some parameters can be tuned, this allows for "trial and error" approaches, that prove very useful in practice.

We leave some open questions: We have shown on the examples that the accuracy is always far better than the predicted one. Why is it so? Can one prove better convergence rates for some families of matrices? How to find an embedded pair? Is there a weaker condition than the presence of an embedded pair, which seems a bit restrictive?

Acknowledgment

We would like to thank Steven Finch for pointing out the application studied in Section 5.2.

References

[1] Galeev E.M. Alexeev V.M. and Tikhomirov V.M. *Book of problems in optimization.* Nauka, Moscow, Russia, 1984.

[2] T. Ando and M.-H. Shih. Simultaneous contractibility. *SIAM Journal on Matrix Analysis and Applications*, 19(2):487–498, 1998.

[3] N. Barabanov. Lyapunov indicators of discrete inclusions i-iii. *Automation and Remote Control*, 49:152–157, 283–287, 558–565, 1988.

[4] M. A. Berger and Y. Wang. Bounded semigroups of matrices. *Linear Algebra and its Applications*, 166:21–27, 1992.

[5] V. D. Blondel and V. Canterini. Undecidable problems for probabilistic automata of fixed dimension. *Theory of Computing Systems*, 36(3):231–245, 2003.

[6] V. D. Blondel, J. Cassaigne, and R. M. Jungers. On the number of α-power-free words for $2 < \alpha < 7/3$. *Theoretical Computer Science*, 2009. to appear.

[7] V. D. Blondel, J. Cassaigne, and R. M. Jungers. On the number of α-power-free words for $2 < \alpha \leq 7/3$. *Theoretical Computer Science*, 2009. to appear.

[8] V. D. Blondel, R. Jungers, and V. Protasov. On the complexity of computing the capacity of codes that avoid forbidden difference patterns. *IEEE Transactions on Information Theory*, 52(11):5122–5127, 2006.

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

[10] V. D. Blondel, Y. Nesterov, and J. Theys. On the accuracy of the ellipsoid norm approximation of the joint spectral radius. *Linear Algebra and its Applications*, 394(1):91–107, 2005.

[11] V. D. Blondel and J. N. Tsitsiklis. The Lyapunov exponent and joint spectral radius of pairs of matrices are hard - when not impossible - to compute and to approximate. *Mathematics of Control, Signals, and Systems*, 10:31–40, 1997.

[12] 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.

[13] Vincent D. Blondel and Yurii Nesterov. Polynomial-time computation of the joint spectral radius for some sets of nonnegative matrices. 2008. submitted.

[14] J. Cassaigne. Counting overlap-free binary words. *STACS 93, Lecture Notes in Computer Science*, 665:216–225, 1993.

[15] I. Daubechies and J. C. Lagarias. Two-scale difference equations. ii. local regularity, infinite products of matrices and fractals. *SIAM Journal of Mathematical Analysis*, 23:1031–1079, 1992.

[16] N. G. de Bruijn. On mahler's partition problem. *Indagatione Mathematicae*, 10:210–220, 1948.

[17] L. Euler. *Introductio in analysis infinitorum.* Teubner, Leipzig, 1922.

[18] S. Finch. private communication, 2008.

[19] S. Finch, P. Sebah, and Z.-Q. Bai. Odd entries in pascal's trinomial triangle. 2008. arxiv preprint: `http://www.citebase.org/abstract?id=oai:arXiv.org:0802.2654`.

[20] N. Dyn G. A. Derfel and D. Levin. Generalized refinement equations and subdivision processes. *Journal of Approximation Theory*, 80:272–297, 1995.

[21] J. Goldwasser, W. Klostermeyer, M. Mays, and G. Trapp. The density of ones in pascal's rhombus. *Discrete Mathematics*, 204(1-3):231–236, 1999.

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

[23] N. Guglielmi and M. Zennaro. Finding extremal complex polytope norms for families of real matrices. *SIAM Journal of Matrix Analysis and Applications*, 2008.

[24] L. Gurvits. Stability of discrete linear inclusions. *Linear Algebra and its Applications*, 231:47–85, 1995.

[25] R. M. Jungers. The joint spectral radius, theory and applications. In Springer-Verlag, editor, *Lecture Notes in Information and Computer Sciences*. Berlin, 2009. To appear.

[26] R. M. Jungers, V. Protasov, and V. D. Blondel. Overlap-free words and spectra of matrices. *Submitted*, 2007. Preprint: `http://arxiv.org/abs/0709.1794`.

[27] R. M. Jungers, V. Protasov, and V. D. Blondel. Efficient algorithms for deciding the type of growth of products of integer matrices. *Linear Algebra and its Applications*, 428(10):2296–2311, 2008.

[28] V. S. Kozyakin. Structure of extremal trajectories of discrete linear systems and the finiteness conjecture. *Automation and Remote Control*, 68(1):174–209, 2007.

[29] M. Lothaire. *Combinatorics on words*, volume 17 of *Encyclopedia of Mathematics*. Cambridge University Press, 1982.

[30] K. Mahler. J. london math. soc. *Math. Notes*, 15:115–123, 1940.

[31] B. E. Moision, A. Orlitsky, and P. H. Siegel. On codes that avoid specified differences. *IEEE Transactions on Information Theory*, 47:433–442, 2001.

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

[33] V. Y. Protasov. The joint spectral radius and invariant sets of linear operators. *Fundamentalnaya i prikladnaya matematika*, 2(1):205–231, 1996.

[34] V. Y. Protasov. The generalized spectral radius. A geometric approach. *Izvestiya Mathematics*, 61(5):995–1030, 1997.

[35] V. Y. Protasov. Asymptotic behaviour of the partition function. *Sbornik Mathematics*, 191(3-4):381–414, 2000.

[36] V. Y. Protasov. Refinement equations with nonnegative coefficients. *Journal of Fourier Analysis and Applications*, 6:55–77, 2000.

[37] V. Y. Protasov. On the asymptotics of the binary partition function. *Mathematical Notes*, 76(1):144–149, 2004.

[38] V. Y. Protasov. On the regularity of de Rham curves. *Izvestiya Mathematics*, 68(3):567–606, 2004.

[39] V. Y. Protasov. Fractal curves and wavelets. *Izvestiya Mathematics*, 70(5):123–162, 2006.

[40] J. Radon. Über eine Erweiterng des Begriffs der konvexen Funktionen, mit einer Anwendung auf die Theorie der konvexen Körper. *Sitzungsberichte Akademie der Wissenschaften in Wien*, 125:241–258, 1916.

[41] B. Reznick. Some binary partition functions. In B.C. Berndt, H.G. Diamond, H. Halberstam, and A. Hildebrand, editors, *Analytic Number Theory, proceedings of a Conference in Honor of Paul T. Bateman*, pages 451–477, Boston, 1990. Birkhäuser.

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

[43] D.-X. Zhou. The p-norm joint spectral radius for even integers. *Methods and Applications of Analysis*, 5:39–54, 1998.

A Proof of Theorem 6

Proof. Any n-hedral cone is affinely equivalent to the positive orthant $\mathbb{R}^n_+$, so we take $K = \mathbb{R}^n_+$. Theorem 5 yields $\alpha(K) \ge \frac{1}{n}$, it remains to establish the opposite inequality. Let $G = \{x \in \mathbb{R}^n_+ \big| (e, x) \le 1\}$, where $e \in \mathbb{R}^n_+$ is the vector of ones. If $v \ge G$, then each coordinate of the point v is at least 1, hence $(e, v) \ge n$. Since $\alpha v \in G$, it follows that $(\alpha v, e) \le 1$, and hence $\alpha \le \frac{1}{n}$.

For the Lorentzian cone we repeat the proof of Theorem 5 and note that any bounded cross-section S in this case is an ellipsoid, for which $\tau(S) = 1$. Taking now $v = 2z$, where $z = \mathrm{gr}S$, we show in the same way that $v \ge G$ and therefore $\alpha(K) \ge \frac{1}{2}$. It remains to prove that $\alpha(K) \le \frac{1}{2}$. This inequality holds for any cone, not necessarily for Lorentzian one. Indeed, let G be a bounded intersection of K with a hyperplane H. If for some point $v \ge G$ and a number $\gamma > \frac{1}{2}$ we have $z = \gamma v \in G$, then $\frac{|v-z|}{|z|} < 1$. Therefore the set $S' = (v - K) \cap H$ is homothetic to the set $S = K \cap H$ with respect to the point z with a coefficient smaller than 1. Hence there exists a point $x \in S \setminus S'$. Clearly, $x \in G$ and $x \notin (v - K)$, which violates the assumption $v \ge G$. Similarly, for the cone $\mathcal{K}_n$ it will suffice to show that $\tau = \frac{1}{n-1}$ for any of its bounded cross-sections. Let us take such a cross-section S, made by a hyperplane $H = \{X \in \mathcal{K}_n \mid \langle X, B \rangle = n\}$, where $B \in \mathcal{K}_n$ and by definition $\langle X, B \rangle = \mathrm{tr}(XB)$. Since S is bounded, it follows that B is positive definite. Otherwise there is a matrix $V \in \mathcal{K}_n$ such that $\langle V, B \rangle = 0$ (this is seen easily if we diagonalize B in an orthonomal basis), in this case $X + tV \in \mathcal{K}_n \cap H$ and so S is not bounded.
Consider any matrix C, for which $CC^T = B$. Since B is positive definite, it follows that C is nondegenerate. Therefore the map $X \mapsto C^T X C$ is an affine isomorphism of the cone $\mathcal{K}_n$ taking that hyperplane to $H = \{X \in \mathcal{K}_n \mid \langle X, I \rangle = n\}$, where I is the identity matrix in $\mathbb{R}^n$. Indeed,

$$\langle X, B \rangle \;=\; \mathrm{tr}(XB) \;=\; \mathrm{tr}(XCC^T) \;=\; \mathrm{tr}(C^T XC) \;=\; \langle C^T XC\,,\, I \rangle.$$

Therefore,

$$\langle X, B \rangle \;=\; n \qquad \Leftrightarrow \qquad \langle C^T XC\,,\, I \rangle \;=\; n.$$

Thus, all cross-sections of the cone $\mathcal{K}_n$ are affinely equivalent to the set $S = \{X \in \mathcal{K}_n \;\big|\; \mathrm{tr}(X) = n\}$, for which $\mathrm{gr}S = I$. Let $I = tX + (1-t)Y$ for some $X, Y \in \partial\mathcal{K}_n$ and $t \in [0, 1]$. There is an orthogonal basis, in which the matrix X has a diagonal form $X = \mathrm{diag}(x_1, \ldots, x_n)$. Whence in that basis $Y = \mathrm{diag}(y_1, \ldots, y_n)$. Let $x = (x_1, \ldots, x_n), y = (y_1, \ldots, y_n)$ be the corresponding points in $\mathbb{R}^n_+$. Since $\mathrm{tr}(X) = \mathrm{tr}(Y) = n$, we see that x and y are both from the $(n-1)$-dimensional simplex

$\Delta = \{u \in \mathbb{R}^n_+ \mid (e,u) = n\}$ with the center e. If $X, Y \in \partial\mathcal{K}_n$, then $x, y \in \partial\Delta$. Since $e = tx + (1-t)y$ and $\tau(\Delta) = \frac{1}{n-1}$, it follows that $t \geq \frac{1}{n-1}$. Thus, $\tau(S) \geq \frac{1}{n-1}$, from which the theorem follows.

□

B Proof of Proposition 5

Proof. Consider first the case of positive orthant $\mathbb{R}^n_+$. Let a line intersect this cone by a segment $[x, y]$ and the cone $\mathbb{R}^n_{+,c}$ by a segment $[x', y']$. Since the points x, y lie on the boundary of $\mathbb{R}^n_+$, it follows that each of them has at least one zero coordinate. Without loss of generality it can be assumed that $x_1 = 0$. In this case $y_1 \neq 0$, otherwise the segment $[x, y]$ does not intersect the cone $\mathbb{R}^n_{+,c}$. Hence, without loss of generality we assume $y_2 = 0$. Since $[x', y'] \subset [x, y]$, we have $x'_2 > y'_2$. Furthermore, $\frac{x'_2}{x'_1} \leq c$ and $\frac{y'_1}{y'_2} \leq c$, because x' and y' are in $\mathbb{R}^n_{+,c}$. Therefore

$$\frac{|x - y'|}{|x - x'|} = \frac{y'_1}{x'_1} = \frac{y'_1}{y'_2}\frac{y'_2}{x'_1} < \frac{y'_1}{y'_2}\frac{x'_2}{x'_1} \leq c^2,$$

which implies $\beta \leq c^2$.

This upper bound is sharp. Indeed, for the points $x = (0, c, \ldots, c), x' = (1, c, \ldots, c), y' = (c^2, c, \ldots, c)$ we have $\frac{|x-y'|}{|x-x'|} = c^2$. Take now a sequence of points $y^k \in \partial\mathbb{R}^n_+$, for which the direction of the vector $y^k - x$ converges to the direction of $x' - x = (1, 0, \ldots, 0)$ as $k \to \infty$. For the segments $[x, y^k]$ the corresponding ratio tends to c^2 as $k \to \infty$.

The proof for the pair $\mathcal{K}_n, \mathcal{K}_{n,c}$ is similar. Let a line intersect $\mathcal{K}_n$ by a segment $[X, Y]$ and the cone $\mathcal{K}_{n,c}$ by a segment $[X', Y']$. Since the matrices X, Y belong to the boundary of the cone $\mathcal{K}_n$, it follows that there are vectors $a, b \in \mathbb{R}^n, |a| = |b| = 1$, such that $(Xa, a) = (Yb, b) = 0$. Note that $a \neq b$, otherwise $(X'a, a) = 0$, which is impossible, because $X' \in \mathrm{int}\mathcal{K}_n$. We have $(X'b, b) > (Y'b, b)$, and therefore

$$\frac{|X - Y'|}{|X - X'|} = \frac{(Y'a, a)}{(X'a, a)} = \frac{(Y'a, a)}{(Y'b, b)}\frac{(Y'b, b)}{(X'a, a)} < \frac{(Y'a, a)}{(Y'b, b)}\frac{(X'b, b)}{(X'a, a)} \leq c^2,$$

from which we deduce $\beta \leq c^2$. This bound is sharp. To see this we take the matrices $X = \mathrm{diag}(0, c, \ldots, c), X' = \mathrm{diag}(1, c, \ldots, c), Y' = \mathrm{diag}(c^2, c, \ldots, c)$ and applying the same argument as above for the cone $\mathbb{R}^n_+$ prove that $\beta(\mathbb{R}^n_+, \mathbb{R}^n_{+,c}) = c^2$.

The case of the Lorentzian cone is elementary. For the dimension $n = 2$ the inequality $\frac{x-y'}{x-x'} \leq \left(\frac{\sin\varphi}{\sin(\varphi-\psi)}\right)^2$ is a simple consequence of the sine law. In case $n > 2$ we consider the restriction to the two-dimensional plane spanned by the vectors x and y. The cross-sections of the cones K_φ and K_ψ by this plane are also coaxial Lorentzian cones of some angles φ' and ψ', for which, moreover, $\frac{\sin\varphi'}{\sin(\varphi'-\psi')} \leq \frac{\sin\varphi}{\sin(\varphi-\psi)}$. Hence the general case follows from the case $n = 2$.

$\square$

C Proof of Proposition 7

Proof. Recall that $\sqrt{\hat{\sigma}(\tilde{\mathcal{M}})}$ can be interpreted as the solution of this optimization problem

$$\min_{||\cdot||} \quad \gamma \tag{18}$$
$$||A|| \leq \gamma \quad \forall A \in \mathcal{M},$$

where the minimum is taken over all the ellipsoidal norms. Let $|\cdot|$ and $||\cdot||$ be the corresponding vector and matrix norms:

$$|x| = (x^T Sx)^{1/2}, \quad \hat{\sigma}(\tilde{\mathcal{M}}) = \max\{||A|| : A \in \mathcal{M}\},$$

for a positive definite matrix S. The corresponding dual norm is still an ellipsoidal norm, defined by the inverse of the positive definite matrix S :

$$|y|_* = \max_{x^T Sx=1} y^T x = (y^T S^{-1} y)^{1/2}.$$

Now, the induced matrix norm $||\cdot||_*$ satisfies the relation:

$$\max_{A\in\mathcal{M}^*} \{||A||_*\} \leq \max_{A\in\mathcal{M}} \{||A||\}. \tag{19}$$

Indeed,

$$\begin{aligned}
||A^*||_* &= \max_{|y|_*=1} |A^* y|_* && (20)\\
&= \max_{|y|_*=1} \max_{|x|=1} y^T Ax && (21)\\
&\leq \max_{|y|_*=1} \max_{|x|=1} |Ax|.y^T(Ax/|Ax|) && (22)\\
&\leq |Ax|. && (23)
\end{aligned}$$

Since the same reasoning holds by inverting the roles of $|\cdot|$ and $|\cdot|_*$, the result is proved. □