

JOINT SPECTRAL CHARACTERISTICS OF MATRICES: A CONIC PROGRAMMING APPROACH*

VLADIMIR Y. PROTASOV[†], RAPHAËL M. JUNGERS[‡], AND VINCENT D. BLONDEL[‡]

Abstract. We propose a new method to compute the joint spectral radius and the joint spectral subradius of a set of matrices. We first restrict our attention to matrices that leave a cone invariant. The accuracy of our algorithm, depending on geometric properties of the invariant cone, is estimated. We then extend our method to arbitrary sets of matrices by a lifting procedure, and we demonstrate the efficiency of the new algorithm by applying it to several problems in combinatorics, number theory, and discrete mathematics.

Key words. joint spectral radius, conic programming, convex optimization, spectral radius, joint spectral subradius, overlap-free words, Euler’s binary function

AMS subject classifications. 15A18, 15-04, 15A60, 90C22, 93-04

DOI. 10.1137/090759896

1. Introduction. The joint spectral radius $\hat{\rho}(\mathcal{M})$ of a set of $n \times n$ real matrices \mathcal{M} is the exponent of the maximal asymptotic growth of products of matrices from this set when the length of the products grows. The joint spectral subradius $\check{\rho}(\mathcal{M})$ (also called the lower spectral radius) is the minimal growth counterpart:

$$(1.1) \quad \begin{aligned} \hat{\rho}(\mathcal{M}) &= \lim_{k \rightarrow \infty} \max \{ \|A_{d_1} \cdots A_{d_k}\|^{1/k} : A_i \in \mathcal{M} \}, \\ \check{\rho}(\mathcal{M}) &= \lim_{k \rightarrow \infty} \min \{ \|A_{d_1} \cdots A_{d_k}\|^{1/k} : A_i \in \mathcal{M} \}. \end{aligned}$$

Both these limits exist for all sets of matrices and do not depend on the norms used in the definitions. In the simplest case, when the set \mathcal{M} consists of only one matrix A , both these spectral quantities are equal to the spectral radius $\rho(A)$, which is then also equal to the largest magnitude of the eigenvalues of A . This follows from Gelfand’s formula $\rho(A) = \lim_{k \rightarrow \infty} \|A^k\|^{1/k}$.

The joint spectral radius was introduced in [39] and the joint spectral subradius in [23]. These quantities have found numerous applications in various areas: in the control of switched systems [3, 26], in subdivision algorithms for approximation and curve design (see [16] for many references), in the study of wavelets and of refinement equations [14, 36], in probabilistic automata [5], in information theory [7], in probability theory [33], and in many problems of discrete mathematics, graph theory, and combinatorics (see [24] for a survey).

The question of how to compute these quantities efficiently has been the subject of intense research activity in recent years. Let us start with the joint spectral radius $\hat{\rho}$.

*Received by the editors May 26, 2009; accepted for publication (in revised form) by N. Masronardi April 26, 2010; published electronically June 11, 2010.

<http://www.siam.org/journals/simax/31-4/75989.html>

[†]Department of Mechanics and Mathematics, Moscow State University, Vorobyovy Gory, 119992, Moscow, Russia (v-protasov@yandex.ru). This author’s research was supported by grants MD-2195.2008.1, grant NSh-3233.2008.1 for leading scientific schools, and RFBR grant 08-01-00208.

[‡]Division of Applied Mathematics, Université catholique de Louvain, 4 avenue Georges Lemaitre, B-1348 Louvain-la-Neuve, Belgium (raphael.jungers@uclouvain.be, vincent.blondel@uclouvain.be). The research of these authors was partially supported by the “Communauté française de Belgique - Actions de Recherche Concertées” and by the Belgian Programme on Interuniversity Attraction Poles initiated by the Belgian Federal Science Policy Office. The second author is an F.R.S.-FNRS Fellow.

For a given set of matrices \mathcal{M} and for a given $k \geq 0$, we define

$$(1.2) \quad \mathcal{M}_k = \{A_{d_1} \cdots A_{d_k} : A_{d_j} \in \mathcal{M}, j = 1, \dots, k\}.$$

Associated to every submultiplicative norm $\|\cdot\|$ we have the following inequalities for the joint spectral radius:

$$(1.3) \quad \max\{\rho(A)^{1/k} : A \in \mathcal{M}_k\} \leq \hat{\rho}(\mathcal{M}) \leq \max\{\|A\|^{1/k} : A \in \mathcal{M}_k\}.$$

Moreover, the limit of the right-hand side of (1.3) is actually equal to $\hat{\rho}(\mathcal{M})$ as k goes to $+\infty$, and the upper limit of the left-hand side is also equal to $\hat{\rho}(\mathcal{M})$ [4]. These observations provide an elementary method to compute the joint spectral radius to any given accuracy by computing all products of length k until the upper and lower bounds resulting from the inequalities (1.3) are sufficiently close to each other [14, 21]. However, in practice the convergence of the bounds in (1.3) may be extremely slow even for small matrix dimensions. This slow convergence is not surprising because of negative complexity results [11, 40], according to which the computation of the joint spectral radius is a problem that is computationally hard. In particular, unless $P=NP$ there is no algorithm that runs in polynomial time both in n and $1/\varepsilon$ and that approximates $\hat{\rho}$ with a relative accuracy ε .

One possible way to improve the convergence in (1.3) is to choose a special norm in \mathbb{R}^n that depends on the set of matrices. This idea was put to good use in [22, 30] (polyhedral norms), [8, 31, 41] (Kronecker lifting), and [29] (sums of squares). The computational complexity of these methods grows exponentially with the dimension of the matrices.

In this paper we generalize this approach, and we obtain efficient algorithms that perform well, even for large dimensions. The main idea is to select a special set of norms and find the best norm in that set with the help of conic optimization. By “best norm” we mean one that provides the sharpest possible upper bound in (1.3) for a given $k \geq 1$. We will see, for instance, how the best ellipsoidal norm method introduced in [2, 10] appears as a special case of our technique.

An interesting aspect of our method is that it can also be used to approximate the joint spectral subradius $\check{\rho}$. To the best of our knowledge, the method we propose here is the first nontrivial approximation method for the joint spectral subradius. The main difficulty with the subradius is that the inequalities corresponding to (1.3) only provide the following upper bounds for the subradius:

$$(1.4) \quad \check{\rho}(\mathcal{M}) \leq \min\{\rho(A)^{1/k} : A \in \mathcal{M}_k\} \leq \min\{\|A\|^{1/k} : A \in \mathcal{M}_k\}.$$

Both these upper bounds tend to $\check{\rho}$ as $k \rightarrow +\infty$, but no convergent lower bound for $\check{\rho}$ is known; therefore, these bounds cannot be used to derive approximations of guaranteed accuracy. As a contribution of this paper, we use the conic norm to derive a lower bound for the subradius that converges to $\check{\rho}$ as $k \rightarrow +\infty$.

The paper is organized as follows. First, we consider sets of matrices that leave a cone invariant. For these sets we define the notions of joint conic radius and joint conic subradius (section 2). Then we prove the main relations between the spectral and conic radii (Theorems 2.6 and 2.12). In section 3 we iterate these relations, and we derive approximation algorithms for the joint spectral radius and subradius. We then apply a lifting procedure in order to extend our approach to arbitrary sets of matrices, possibly without invariant cone. We also present and analyze in that section general tricks that can improve our algorithms in practice. In section 4 we describe

numerical examples by applying our method to several problems of number theory and combinatorics. We consider problems on the asymptotics of the overlap-free language, the density of ones in Pascal's rhombus, and the analysis of Euler's partition function. The dimensions of the matrices that we consider in these contexts range from 5 to 20.

2. Main results. Let $K \subset \mathbb{R}^n$ be a convex, closed, pointed, and nondegenerate cone with the apex at the origin (for definitions in convex geometry, see [12]). Any such cone defines a partial order in \mathbb{R}^n : we write $x \geq_K y$ ($x >_K y$) for $x - y \in K$ ($x - y \in \text{int}K$). The cone K is an *invariant cone* for the matrix A if $AK \subset K$. In this case we say that A is *nonnegative* and write $A \geq_K 0$. If K is invariant for all matrices of some set \mathcal{M} , then it is said to be an invariant cone for that set.

DEFINITION 2.1. For a given compact set of matrices \mathcal{M} with an invariant cone K , we consider the values

$$(2.1) \quad \begin{aligned} \hat{\sigma}_K(\mathcal{M}) &= \inf \left\{ \lambda \geq 0 \mid \text{there exist } v >_K 0, \quad Av \leq_K \lambda v \quad \text{for all } A \in \mathcal{M} \right\}, \\ \check{\sigma}_K(\mathcal{M}) &= \sup \left\{ \lambda \geq 0 \mid \text{there exist } v \geq_K 0, v \neq 0, \quad Av \geq_K \lambda v \quad \text{for all } A \in \mathcal{M} \right\}, \end{aligned}$$

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 following we assume the cone K to be fixed, and we use the short notation $\hat{\sigma}(\mathcal{M}), \check{\sigma}(\mathcal{M})$, or simply $\hat{\sigma}, \check{\sigma}$, if it is clear what set \mathcal{M} is considered.

These values are well defined, provided the set \mathcal{M} admits an invariant cone. However, if this is not the case, then the following simple procedure can be used. Let us define the *semidefinite lifting* \tilde{A} of the $n \times n$ matrix A by

$$(2.2) \quad \tilde{A} : \mathbb{R}^{n^2} \rightarrow \mathbb{R}^{n^2} : X \rightarrow A^T X A.$$

PROPOSITION 2.2 (see [8, 24]). Let \mathcal{M} be a set of $n \times n$ matrices. The semidefinite lifting $\tilde{\mathcal{M}} = \{\tilde{A} : A \in \mathcal{M}\}$ of \mathcal{M} leaves the cone \mathcal{K}_n of symmetric positive semidefinite matrices invariant. Moreover, the set $\tilde{\mathcal{M}}$ satisfies $\hat{\rho}(\tilde{\mathcal{M}}) = \hat{\rho}(\mathcal{M})^2$ and $\check{\rho}(\tilde{\mathcal{M}}) = \check{\rho}(\mathcal{M})^2$.

As a result of this proposition, sets of matrices can always be transformed into sets of matrices that leave a cone invariant and whose joint spectral radius is squared. But this is, of course, at the cost of squaring the space dimension.

2.1. The joint spectral radius. In this section, we derive 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 $\mathbb{R}^{n \times n}$. In the following we call *convex body* a convex compact set with a nonempty interior.

LEMMA 2.3 (see [35]). If there exists a convex body $P \subset \mathbb{R}^n$ and $\lambda > 0$ 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 the following lemma.

LEMMA 2.4. Let the set \mathcal{M} possess an invariant cone K . If for some $v \in \text{int}K$ we have $Av \leq_K \lambda v$ for all $A \in \mathcal{M}$, then $\hat{\rho} \leq \lambda$. If for some $v \in K \setminus \{0\}$ we have $Av \geq_K \lambda v$ for all $A \in \mathcal{M}$, then $\check{\rho} \geq \lambda$.

Proof. For the first assertion we apply Lemma 2.3 to the body $P = (v - K) \cap (-v + K)$. The second assertion follows from the same lemma for the set $Q = v + K$. \square

In order to find a relation between $\hat{\rho}$ and $\hat{\sigma}$, we introduce the following geometric characteristics of convex cones.

DEFINITION 2.5. For a given cone $K \subset \mathbb{R}^n$, the value $\alpha(K)$, $0 \leq \alpha \leq 1$, is the largest number such that for any convex compact set $G \subset K$, there exists $v \in G$, for which $v \geq_K \alpha G$.

An example of the value of α for the positive orthant in \mathbb{R}^2 is provided in Figure 2.1. Clearly, α is an affine invariant of a cone.

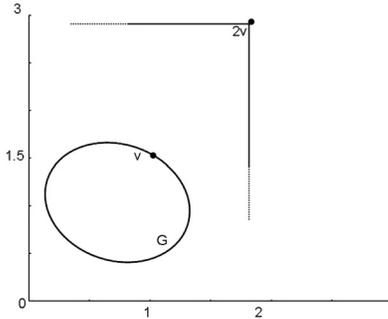


FIG. 2.1. For the positive orthant in \mathbb{R}^2 , we have $\alpha = 1/2$. That is, associated to any compact convex set in \mathbb{R}_+^2 , there is a point $v \in G$ such that $2v \geq_{\mathbb{R}_+^2} x$ for all $x \in G$.

THEOREM 2.6. For any set \mathcal{M} of matrices with an invariant cone K , we have

$$\alpha \hat{\sigma} \leq \hat{\rho} \leq \hat{\sigma},$$

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

Proof. The inequality $\hat{\rho} \leq \hat{\sigma}$ is proved in Lemma 2.4. To prove the inequality $\alpha \hat{\sigma} \leq \hat{\rho}$, we use the well-known fact [24,39] 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 . Fix then such a q and 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 $u \geq_K G$ such that $\gamma u \in G$. Observe that in this case $\text{int}G \neq \emptyset$; hence, $u >_K 0$. It follows that $A(\gamma u) \in AG \subset qG$, and so, $Au \in \frac{q}{\gamma}G \leq_K \frac{q}{\gamma}u$. Thus, $Au \leq_K \frac{q}{\gamma}u$ for all $A \in \mathcal{M}$. Whence $\hat{\sigma} \leq \frac{q}{\gamma}$. This holds for arbitrary $q > \hat{\rho}$ and $\gamma < \alpha$; therefore, $\hat{\sigma} \leq \frac{\hat{\rho}}{\alpha}$, which concludes the proof. \square

Let us add that this theorem is proved in [9] for the particular case of nonnegative matrices. The following theorem provides a universal bound on α .

THEOREM 2.7. For any cone $K \subset \mathbb{R}^n$ we have $\alpha(K) \geq \frac{1}{n}$.

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

$$\tau_z(G) = \inf \left\{ t > 0 \mid \text{there exist } x, y \in \partial G, z = \frac{t}{1+t}x + \frac{1}{1+t}y \right\}.$$

In other words, $\tau_z(G)$ is the minimal possible ratio $\frac{|z-y|}{|z-x|}$, with $|\cdot|$ denoting the Euclidean norm and where x, y are the points of intersection of a line passing through z with the boundary of G . We then denote

$$\tau(G) = \tau_{z'}(G), \quad \text{with} \quad z' = \frac{1}{\text{Vol}G} \int_G x \, dx,$$

that is, with z' the center of gravity of G . The well-known Minkowski–Radon theorem [37] states that $\tau(G) \geq \frac{1}{d}$ for any convex body G . For d -dimensional simplices

Δ , one has $\tau_z(\Delta) = \frac{1}{d}$ for $z = \text{gr}\Delta$ 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 2.8. *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 H does not separate the set G from 0 , and the cross section $H \cap K$ is bounded and has its center of gravity in G .*

For the proof of this lemma it suffices to consider the set $G' = \text{Conv}\{G, 0\}$ 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. We are now in position to prove the above theorem.

Proof of Theorem 2.7. Let G be an arbitrary convex subset of a cone K . If G 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. By induction, we get $\alpha(K) \geq \frac{1}{n-1} > \frac{1}{n}$. Assume now that $G \cap \text{int}K \neq \emptyset$. Applying Lemma 2.8, we obtain a hyperplane of support H , which does not separate G from the origin. Let us denote $S = K \cap H$ and $v = \text{gr}S \in G$. 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, $u \geq_K G$. 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 2.7 guarantees that $\alpha \geq \frac{1}{n}$ for any cone in \mathbb{R}^n . Therefore, we have, by Theorem 2.6,

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

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 Lorentz cone. A *Lorentz cone* of angle $\phi \in]0, \pi/2[$ is the set $K_\phi = \{x \in \mathbb{R}^n : x_1 \tan \phi \geq \sqrt{x_2^2 + \dots + x_n^2}\}$ or a rotation of this set around the origin.

PROPOSITION 2.9. *For any n -hedral cone in \mathbb{R}^n we have $\alpha = \frac{1}{n}$; for a Lorentz cone we have $\alpha = \frac{1}{2}$; for the cone \mathcal{K}_n of positive semidefinite $n \times n$ -matrices, we have $\alpha = \frac{1}{n}$.*

The proof of Proposition 2.9 is in Appendix A.

2.2. The joint spectral subradius. A straightforward application of Lemma 2.4 yields the inequality $\hat{\sigma} \leq \hat{\rho}$. However, in contrast to Theorem 2.6, there is no corresponding upper bound on $\hat{\rho}$. More precisely, there is no positive constant $C(K)$ corresponding to the cone K such that $\hat{\rho}(\mathcal{M}) \leq C(K) \hat{\sigma}(\mathcal{M})$ for any set of matrices \mathcal{M} that leaves K invariant. We show this with an example for the case $K = \mathbb{R}_+^n$.

Example 1. Let $n \geq 2$ and $\mathcal{M} = \{A_1, \dots, A_n\} \subset \mathbb{R}^n$, where A_j is a matrix whose entries of the j th 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, $\hat{\rho} = n-1$. On the other hand, $\hat{\sigma}(\mathcal{M}) = 0$. Indeed, any nonnegative

vector $v \neq 0$ has at least one positive coordinate v_j , while $(A_j v)_j = 0$. Hence, the inequality $A_j v \geq_{\mathbb{R}^n_+} \lambda v$ implies $\lambda = 0$.

Thus, to obtain an upper bound for the subradii, we need to impose some extra conditions on the matrices. This situation can actually not be avoided, since Theorem 2 in [40] shows that there is no algorithm that approximates the subradius for general finite sets of matrices. Example 1 shows that an invariant cone does not suffice to obtain an upper bound. 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 $(K' \setminus \{0\}) \subset \text{int}K$. 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 . This definition extends to sets of matrices.

DEFINITION 2.10. 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 $1 \leq \frac{|x-y'|}{|x-x'|} \leq \beta$.

An illustration of the value $\beta(K, K')$ is provided in Figure 2.2.

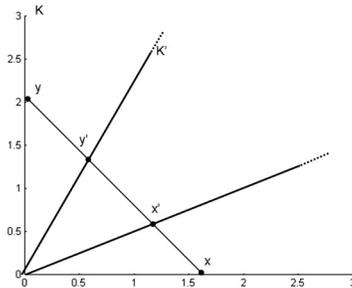


FIG. 2.2. The constant β for an embedded pair (K, K') is equal to the maximum ratio $\frac{|x-y'|}{|x-x'|}$.

LEMMA 2.11. Let \mathcal{M} be a set of matrices such that $\check{\rho}(\mathcal{M}) > 0$, and assume that \mathcal{M} has an invariant embedded pair (K, K') . Then, for any $p : 0 < p \leq \check{\rho}$ 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

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

Since the cone K' is invariant, $Q \subset K'$, and, obviously, $AQ \subset Q$. Hence, for all nonzero $p < 1$, $AQ \subset pQ$. 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 $A_k v_0 \rightarrow 0$. However, since $v_0 \in \text{int}K$, this implies that $\|A_k\| \rightarrow 0$, and so, $\check{\rho}(\mathcal{M}) < 1$, which is in contradiction with the assumptions. \square

THEOREM 2.12. For any set \mathcal{M} with an invariant pair (K, K') , we have

$$\check{\sigma} \leq \check{\rho} \leq \beta \check{\sigma},$$

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

Proof. The inequality $\check{\sigma} \leq \check{\rho}$ follows from Lemma 2.4. To prove that $\check{\rho} \leq \beta \check{\sigma}$, we apply Lemma 2.11 and get a set Q such that $AQ \subset pQ$ for all $A \in \mathcal{M}$ and for all $0 < p \leq \check{\rho}$.

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 $v \leq_K \beta Q$. 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 $(1 - \frac{1}{\beta})$ with respect to v contains the set S' . This yields that $S' \subset (\frac{1}{\beta}v + K)$; therefore, $Q \subset (\frac{1}{\beta}v + K)$, which means $\frac{1}{\beta}v \leq_K Q$, and hence, $v \leq_K \beta Q$. Since $AQ \subset pQ$, it follows that $\frac{1}{\beta}v \leq_K \frac{1}{p}AQ$. On the other hand, $v \in Q$; whence, $\frac{1}{\beta}v \leq_K \frac{1}{p}Av$ for any $A \in \mathcal{M}$. This means that $\tilde{\sigma} \geq \frac{p}{\beta}$. Taking a limit $p \rightarrow \tilde{\rho}$, we get $\beta\tilde{\sigma} \geq \tilde{\rho}$, from which the theorem follows. \square

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 $x_j \rightarrow x$ as $j \rightarrow \infty$), then $\beta(K, K'_j) \rightarrow +\infty$.

Now we compute the values $\beta(K, K')$ for several cases of embedded pairs. For a given point $x = (x_1, \dots, x_n) \in \mathbb{R}_+^n$, we denote by x_{\min} and x_{\max} its smallest and greatest entries, respectively. Similarly, for $X \in \mathcal{K}_n$, λ_{\min} and λ_{\max} denote the smallest and greatest eigenvalues, respectively. 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 Lorentz cone of angle $\varphi < \frac{\pi}{2}$.

PROPOSITION 2.13. *If $\mathbb{R}_{+,c}^n = \{x \in \mathbb{R}_+^n \mid x_{\max} \leq cx_{\min}\}$, then*

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

if $\mathcal{K}_{n,c} = \{X \in \mathcal{K}_n \mid \lambda_{\max} \leq c\lambda_{\min}\}$, then

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

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

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

The proof of Proposition 2.13 is 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 $A(\mathbb{R}_+^n) \subset \mathbb{R}_{+,c}^n$. This yields, in particular, that A possesses an invariant cone $K' = A(\mathbb{R}_+^n)$ contained in $\mathbb{R}_{+,c}^n$.

COROLLARY 2.14. *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 $\beta \leq c^2$. This pair is $K = \mathbb{R}_+^n$, $K' = \text{conv}\{A(\mathbb{R}_+^n) \mid A \in \mathcal{M}\}$.*

If a matrix A with an invariant cone K is such that $AK \subset \text{int}K$, then (K, AK) is an invariant pair for A for which $\beta < \infty$. If a set of such matrices \mathcal{M} is compact, we see that the cone $K' = \text{conv}\{AK \mid A \in \mathcal{M}\}$ is embedded in K , and so, all such sets admit a constant $\beta < \infty$. Indeed, by contradiction, let us suppose that we have a sequence x_i, y_i, x'_i, y'_i , $x_i, y_i \in K$, $x'_i, y'_i \in K'$ such that $\frac{|x_i - y'_i|}{|x_i - x'_i|}$ tends to ∞ . We can suppose (by scaling the vectors) that the vectors x_i belong to the unit ball. Thus,

by compactness of the unit ball, there is a subsequence of x_i that tends to some unit vector x . Now, since $x \notin K'$, $x_i - x'_i$ is bounded away from zero, and thus the points y'_i cannot be bounded. This means, again by compactness, that there is a subsequence of the directions $y_i - x_i$ which tends to the direction of an extremal ray of K' . This direction is thus in the interior of K . But this is impossible because the vector $y_i - x_i$ cannot belong to K , since x_i, y_i belong to its boundary.

COROLLARY 2.15. *If a compact set of matrices \mathcal{M} has an invariant cone K such that $AK \subset \text{int}K$ for all $A \in \mathcal{M}$, then it has an embedded pair (K, K') for which $\beta(K, K') < \infty$. Consequently, there is an algorithm to approximate its joint spectral subradius up to any required accuracy.*

3. Discussion. In this section we first briefly describe how our results can be implemented, and we comment on the approximation they guarantee. We then present some particular cases that can improve in a critical way our algorithms in practice. Subsection 3.2 deals with the transposition of the initial set of matrices. Subsection 3.3 presents some favorable cases where the joint spectral characteristic can be computed exactly; while in subsection 3.4 we show how it is sometimes possible to construct a wider common invariant cone than the classical ones (like \mathbb{R}_+^n or \mathcal{K}_n).

3.1. Implementation. For a given set \mathcal{M} we want to find numbers $\hat{\rho}_*$ and $\check{\rho}_*$ such that $|\hat{\rho}_* - \hat{\rho}|/\hat{\rho} \leq \varepsilon$ and $|\check{\rho}_* - \check{\rho}|/\check{\rho} \leq \varepsilon$. If \mathcal{M} has an invariant cone with parameter α (or an invariant pair with parameter β), then \mathcal{M}_k has the same cone (or the same invariant pair). If the set \mathcal{M} does not have an invariant cone, one can always invoke Proposition 2.2 in order to obtain the corresponding set $\tilde{\mathcal{M}}$ that leaves \mathcal{K}_n invariant. Applying now Theorems 2.6 and 2.12 for the set \mathcal{M}_k , we obtain the following.

COROLLARY 3.1. *If a set \mathcal{M} has an invariant cone K with parameter α , then for any $k \in \mathbb{N}$*

$$(3.1) \quad \alpha^{1/k} [\hat{\sigma}(\mathcal{M}_k)]^{1/k} \leq \hat{\rho}(\mathcal{M}) \leq [\hat{\sigma}(\mathcal{M}_k)]^{1/k}.$$

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

$$(3.2) \quad [\check{\sigma}(\mathcal{M}_k)]^{1/k} \leq \check{\rho}(\mathcal{M}) \leq \beta^{1/k} [\check{\sigma}(\mathcal{M}_k)]^{1/k},$$

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

This result ensures that $\hat{\rho}_* = [\hat{\sigma}(\mathcal{M}_k)]^{1/k}$ gives the desired accuracy ε whenever $k \geq \frac{\ln \frac{1}{\varepsilon}}{\alpha}$. Moreover, by Theorem 2.7, $\alpha \geq \frac{1}{n}$ for any cone, and, hence, for any set with an invariant cone, it suffices to take $k \geq \frac{\ln n}{\varepsilon}$. For the joint spectral subradius, we take $\check{\rho}_* = [\check{\sigma}(\mathcal{M}_k)]^{1/k}$. This gives the desired approximation for $k \geq \frac{\ln \beta}{-\ln(1-\varepsilon)}$. Note that this value does not exceed $\frac{\ln \beta}{\varepsilon}$. Therefore, to compute the joint spectral subradius, it suffices to take $k \geq \frac{\ln \beta}{\varepsilon}$. Let us remember, though, that the parameter β depends on the cones, and, in contrast to α , it cannot be uniformly estimated for all cones in \mathbb{R}^n .

The quantities $\check{\sigma}(\mathcal{M}_k)$ and $\hat{\sigma}(\mathcal{M}_k)$ are easy to compute by applying conic programming methods (together with the bisection method). For instance, let us consider an arbitrary set \mathcal{M} of matrices and lift it using the semidefinite lifting (2.2) to a set $\tilde{\mathcal{M}}$. Now, $\hat{\sigma}(\tilde{\mathcal{M}}_k)$ is given by the following:

$$(3.3) \quad \begin{array}{ll} \min & r \\ X & \geq_{\mathcal{K}_n} I, \\ A^T X A & \leq_{\mathcal{K}_n} rX, \quad A \in \mathcal{M}_k. \end{array}$$

By a simple bisection algorithm, one can find $\bar{r} = \hat{\sigma}(\tilde{\mathcal{M}}_k)$. Since $\hat{\rho}(\tilde{\mathcal{M}}) = [\hat{\rho}(\mathcal{M})]^2$, invoking Corollary 3.1 and taking into account that $\alpha(\mathcal{K}_n) = \frac{1}{n}$ (Proposition 2.9), we obtain $(\bar{r})^{\frac{1}{2k}} n^{-\frac{1}{2k}} \leq \hat{\rho}(\tilde{\mathcal{M}}) \leq (\bar{r})^{\frac{1}{2k}}$. Hence, the value $(\bar{r})^{\frac{1}{2k}}$ approximates $\hat{\rho}(\mathcal{M})$ with the relative precision $\varepsilon \leq \frac{\ln n}{2k}$.

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 our technique with the invariant cone \mathcal{S}_n^+ . We have deduced the main inequality and the estimate of the accuracy using the conic radius, which is a completely different way from that used in [2, 10].

3.2. Transposition. For a given set \mathcal{M} we denote by \mathcal{M}^T the set of transpose matrices. It is well known that the sets \mathcal{M} and \mathcal{M}^T have the same joint spectral radius $\hat{\rho}$ [24], whereas their conic radii $\hat{\sigma}(\mathcal{M})$ and $\hat{\sigma}(\mathcal{M}^T)$ are a priori different.

This is, for instance, the case for the following set of matrices: let $\mathcal{M} = \{A_i\}$, where A_i is the matrix whose entries are all equal to zero, except the i th row whose entries are all equal to one. It is not difficult to see that \mathcal{M} has a joint spectral radius equal to one, but the conic radius $\hat{\sigma}(\mathcal{M})$ is equal to n . In conclusion, the lower bound $\hat{\sigma}(\mathcal{M})/n$ 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 for the transpose set.

Therefore, combining both values $\hat{\sigma}(\mathcal{M})$ and $\hat{\sigma}(\mathcal{M}^T)$ to estimate the joint spectral radius by Theorem 2.6, we can obtain better results. One could hope that the following equation holds:

$$(3.4) \quad (1/f(n)) \min \{\hat{\sigma}(\mathcal{M}), \hat{\sigma}(\mathcal{M}^T)\} \leq \hat{\rho}(\mathcal{M}),$$

where $f(n) < n$. Unfortunately, it appears that the growth $f(n) \approx n$ cannot be avoided as shown by the set

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

where A_i are defined as above. This set has joint spectral radius equal to one, but

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

The previous example proves the following proposition.

PROPOSITION 3.2. *The function f in (3.4) cannot be chosen smaller than $n/2$, where n is the dimension of the matrices.*

The same holds for the subradii $\tilde{\rho}$ and $\tilde{\sigma}$: for some sets \mathcal{M} , taking the transpose of the set helps improve the estimates significantly. However, a construction similar to the one above shows that, in general, one cannot hope to improve the bounds by considering $\max \{\tilde{\sigma}(\mathcal{M}), \tilde{\sigma}(\mathcal{M}^T)\}$.

Also, one could ask the same questions for other invariant cones. For instance, by applying the semidefinite lifting (2.2) to the matrices in \mathcal{M} , one gets another set of matrices $\tilde{\mathcal{M}}$ that leaves \mathcal{K}_n invariant. Thus, we could as well apply the lifting to \mathcal{M}^T to get another estimate. Even though the obtained matrices are not the transpose of the initial matrices, this estimate will not be better, as shown in Appendix C.

3.3. Exact computation of the joint spectral quantities in special cases. Theorems 2.6 and 2.12 make it possible not only to estimate the joint spectral quantities but also to find their precise values in some favorable cases.

PROPOSITION 3.3. *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 : $Av = rv$.*

- (a) If $v \in \text{int}K$ and $rv \geq_K Bv$ for all $B \in \mathcal{M}_k$, then $\hat{\rho}(\mathcal{M}) = r$.
- (b) If $rv \leq_K Bv$ for all $B \in \mathcal{M}_k$, then $\check{\rho}(\mathcal{M}) = r$.

Proof. If $v \in K$ is the Perron-Frobenius eigenvector of the matrix $A \in \mathcal{M}_k$ with eigenvalue r , then (see (1.3) and (1.4))

$$\check{\rho} \leq r^{1/k} \leq \hat{\rho}.$$

Now, the hypothesis in (a) implies that $\hat{\sigma} \leq r^{1/k}$, and the hypothesis in (b) implies that $r^{1/k} \leq \check{\sigma}$. It remains to combine this with Theorems 2.6 and 2.12. \square

This proposition provides sufficient conditions for the joint spectral radius to attain its value at some finite product $A \in \mathcal{M}_k$.

3.4. Constructing wider invariant cones. If a cone K is invariant, but in practice does not deliver a good approximation of the joint spectral quantities, it might be possible to construct a wider cone that gives better approximations.

Let K_m be the closure of the set $\{x \in \mathbb{R}^n \mid Cx \in K \text{ for all } C \in \mathcal{M}_m\}$. Clearly, K_m is an invariant cone of \mathcal{M} containing K . If one defines the order in \mathbb{R}^n by the cone K_m and the constants $\hat{\sigma}$ and $\check{\sigma}$ by formulas (2.1) with respect to this cone, one gets the following proposition, which generalizes Proposition 3.3.

PROPOSITION 3.4. *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 $v \in \text{int}K$ 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 $\check{\rho}(\mathcal{M}) = [\rho(A)]^{1/k}$.*

4. Applications. In this section we briefly describe applications where the techniques developed in this paper prove to be useful. We have chosen two applications in number theory because this field has provided many sets of matrices that can be used as benchmarks. We start with a recent application: the computation of the asymptotics of overlap-free words.

4.1. Overlap-free words. This problem arises in combinatorics on words (for an introduction to combinatorics on words, see [27]), 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

$$r^- = \liminf_{n \rightarrow \infty} \frac{\log u_n}{\log n}, \quad r^+ = \limsup_{n \rightarrow \infty} \frac{\log u_n}{\log n}.$$

The following result ([25], see also [6, 13]) allows us to express the asymptotics of u_l in terms of joint spectral characteristics.

THEOREM 4.1. *There exist two nonnegative matrices $A_0, A_1 \in \{0, 1, 2\}^{20 \times 20}$ such that*

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

Thanks to this result, the following estimates appear in [25]:

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

The inequality $1.2690 < r^-$ was obtained from Theorem 2.12. Unfortunately, no embedded invariant pair is known for A_0 and A_1 , and so, it is not possible to obtain an upper bound on r^- with Theorem 2.12. However, it can be checked by extensive search that the product $A_1^{10}A_0$ satisfies the following:

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

One can verify numerically that this product gives the best possible upper bound among all matrix products of length less than 14. The upper bound on r^+ can be found by solving the semidefinite program (3.3) with $k = 14$, while the lower bound is obtained from the simple inequality

$$(4.1) \quad \hat{\rho} \geq [\rho(A_0A_1)]^{1/2} = 2.5179 \dots$$

Remark that the accuracy of this estimate is 0.0003. As we have seen in section 3, in order to ensure such an accuracy, one has to solve the semidefinite program (3.3) with $k = \ln(n)/(2 \cdot 0.0003) \approx 5000$, which is, of course, enormous. However, (4.1) shows that the actual cost for obtaining such an accuracy is much lower.

4.2. Pascal's rhombus. Recently, the question of the density of ones in Pascal's rhombus arose in number theory [20]. 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

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

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 can be expressed by the joint spectral quantities of the following two matrices:

$$(4.2) \quad \Sigma = \left\{ \begin{pmatrix} 0 & 1 & 0 & 0 & 0 \\ 1 & 0 & 2 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & 1 \\ 0 & 0 & 0 & 2 & 1 \end{pmatrix}, \begin{pmatrix} 1 & 0 & 2 & 0 & 0 \\ 0 & 0 & 0 & 2 & 1 \\ 1 & 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 \end{pmatrix} \right\}.$$

More precisely,

$$\limsup_{n \rightarrow \infty} \frac{\log v_n}{\log n} = \log_2 \hat{\rho}(\{A_0, A_1\}),$$

$$\liminf_{n \rightarrow \infty} \frac{\log v_n}{\log n} = \log_2 \check{\rho}(\{A_0, A_1\}).$$

In [19] the authors mention the difficulty of finding estimates for $\check{\rho}(\{A_0, A_1\})$. It was conjectured later [18] that $\check{\rho}(\{A_0, A_1\}) = (1 + \sqrt{5})/2 = 1.61803 \dots$. These two matrices leave the positive orthant $K = \mathbb{R}_+^5$ invariant, and so, one can try to obtain a lower bound on $\check{\rho}$ with our technique. It appears, however, that this algorithm does not provide a better lower bound than the trivial value 1. Nevertheless, when applied to the transpose 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 Proposition 3.4) 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: $\check{\rho} \leq \rho(A_0^3 A_1^3)^{(1/6)} = 1.6376$. It can be checked by extensive search that 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 $\hat{\rho}(\Sigma) = 2$.

4.3. 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 $k \rightarrow \infty$, and the problem is to find the exponents of this asymptotic growth. For various d this problem was studied by Euler [17], Mahler [28], de Bruijn [15], Reznick [38], and others. There are certain relations of this problem with the theory of refinement equations and subdivision algorithms [34].

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 [32] that the asymptotic behavior of $b_{m,d}(k)$ as $k \rightarrow \infty$ is ruled by the joint spectral quantities of certain sets of matrices $\Sigma_{m,d}$, with binary entries. More precisely, for all pairs $(m, d) \in \mathbb{N}^2$, there exist constants C_1, C_2, λ_1 , and λ_2 such that the following holds:

$$(4.3) \quad C_1 k^{\lambda_1} \leq b_{m,d}(k) \leq C_2 k^{\lambda_2}.$$

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

$$\begin{aligned} \lambda_1 &= \log_m \hat{\rho}_{m,d}, \\ \lambda_2 &= \log_m \check{\rho}_{m,d}. \end{aligned}$$

In [32], these joint spectral quantities are analyzed 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 leaves the positive orthant invariant; hence, we can apply algorithms from section 3. Take, for example, $m = 3, d = 14$. It follows from [32] that

$$\Sigma_{3,14} = \left\{ \begin{aligned} &\left(\begin{array}{cccccccc} 1 & 1 & 1 & 1 & 1 & 0 & 0 \\ 0 & 1 & 1 & 1 & 1 & 0 & 0 \\ 0 & 1 & 1 & 1 & 1 & 1 & 0 \\ 0 & 1 & 1 & 1 & 1 & 1 & 0 \\ 0 & 0 & 1 & 1 & 1 & 1 & 0 \\ 0 & 0 & 1 & 1 & 1 & 1 & 1 \\ 0 & 0 & 1 & 1 & 1 & 1 & 1 \end{array} \right), \left(\begin{array}{cccccccc} 1 & 1 & 1 & 1 & 1 & 0 & 0 \\ 1 & 1 & 1 & 1 & 1 & 0 & 0 \\ 0 & 1 & 1 & 1 & 1 & 0 & 0 \\ 0 & 1 & 1 & 1 & 1 & 1 & 0 \\ 0 & 1 & 1 & 1 & 1 & 1 & 0 \\ 0 & 0 & 1 & 1 & 1 & 1 & 0 \\ 0 & 0 & 1 & 1 & 1 & 1 & 1 \end{array} \right), \\ &\left. \left(\begin{array}{cccccccc} 1 & 1 & 1 & 1 & 0 & 0 & 0 \\ 1 & 1 & 1 & 1 & 1 & 0 & 0 \\ 1 & 1 & 1 & 1 & 1 & 0 & 0 \\ 0 & 1 & 1 & 1 & 1 & 0 & 0 \\ 0 & 1 & 1 & 1 & 1 & 1 & 0 \\ 0 & 1 & 1 & 1 & 1 & 1 & 0 \\ 0 & 0 & 1 & 1 & 1 & 1 & 0 \end{array} \right) \right\}.$$

The set $\Sigma_{3,14}$ leaves the cone $\mathbb{R}_{+,2}^n = \{x \in \mathbb{R}_+^n \mid x_{\max} \leq 2x_{\min}\}$ invariant, and thus, combining Proposition 2.13 with Corollary 3.1, we have the following bounds on the accuracy of the subradius approximation:

$$(4.4) \quad [\check{\sigma}(\mathcal{M}_k)]^{1/k} \leq \check{\rho}(\mathcal{M}) \leq 4^{1/k} [\check{\sigma}(\mathcal{M}_k)]^{1/k}.$$

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

$$4.525 \leq \check{\rho}.$$

Since $\check{\rho} \leq \rho(A_0 A_1)^{1/2} = 4.6105$, the actual ratio with $k = 9$ is at most $4.6105/4.525 = 1.02$, which is 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 $\rho \geq \rho(A_1 A_2)^{1/2} = 4.72$. 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 3.1. Thus, for $m = 3, d = 14$ we have

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

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

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

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

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

5. Conclusion. In this paper we have pursued several goals. First, even though the joint spectral radius has received much attention in the last decades and several algorithms have been proposed to approximate it, to the best of our knowledge, we provide here the first approximation algorithm for the joint spectral subradius.

Second, we propose a general framework (conic optimization) that unifies several of the known algorithms for the joint spectral radius and provides simple proofs of their convergence rate. This framework also allows for new algorithms for computing the joint spectral radius when the matrices share an invariant cone.

Third, we illustrate the effectiveness of our algorithms on several examples. The algorithms perform usually far better than predicted, and some tricks are sometimes useful to obtain good approximations. An example of such a trick is matrix transposition: the example in subsection 4.2 shows that effect. Regarding this transposition trick, we showed moreover (subsection 3.2) that it cannot guarantee to always provide a better accuracy for our algorithms. In practice, our results allow one to find accurate estimates very rapidly. As some parameters can be tuned, this allows for “trial and error” approaches that prove useful in practice.

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

Appendix A. Proof of Proposition 2.9.

Proof. Any n -hedral cone is affinely equivalent to the positive orthant \mathbb{R}_+^n , so we take $K = \mathbb{R}_+^n$. Theorem 2.7 yields $\alpha(K) \geq \frac{1}{n}$; it remains to establish the opposite inequality. Let $G = \{x \in \mathbb{R}_+^n \mid \langle e, x \rangle \leq 1\}$, where $e \in \mathbb{R}_+^n$ is the vector of ones. If $v \geq_{\mathbb{R}_+^n} G$, then each coordinate of the point v is at least 1; hence, $\langle e, v \rangle \geq n$. Since $\alpha v \in G$, it follows that $\langle \alpha v, e \rangle \leq 1$, and hence $\alpha \leq \frac{1}{n}$.

For the Lorentz cone we repeat the proof of Theorem 2.7 and note that any bounded cross section S in this case is an ellipsoid for which $\tau(S) = 1$. Taking now $v = 2z$, where $z = \text{gr}S$, we show in the same way that $v \geq_K G$, and therefore, $\alpha(K) \geq \frac{1}{2}$. It remains to prove that $\alpha(K) \leq \frac{1}{2}$. This inequality holds for any cone, not necessarily for a Lorentz one. Indeed, let G be a bounded intersection of K with a hyperplane H . If for some point $v \geq_K G$ 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 $v \geq_K G$. 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, $\langle X, B \rangle = \text{tr}(XB)$. 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 orthonormal 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,

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

Therefore,

$$\langle X, B \rangle = n \quad \Leftrightarrow \quad \langle C^T X C, I \rangle = n.$$

Thus, all cross sections of the cone \mathcal{K}_n are affinely equivalent to the set $S = \{X \in \mathcal{K}_n \mid \text{tr}(X) = n\}$ for which $\text{gr}S = I$. 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 $X = \text{diag}(x_1, \dots, x_n)$. Whence in that basis $Y = \text{diag}(y_1, \dots, y_n)$. Let $x = (x_1, \dots, x_n), y = (y_1, \dots, y_n)$ be the corresponding points in \mathbb{R}_+^n . Since $\text{tr}(X) = \text{tr}(Y) = n$, we see that x and y are both from the $(n-1)$ -dimensional simplex $\Delta = \{u \in \mathbb{R}_+^n \mid \langle e, u \rangle = n\}$ 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. \square

Appendix B. Proof of Proposition 2.13.

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 $\mathbb{R}_{+,c}^n$ 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 $\mathbb{R}_{+,c}^n$. 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 $\mathbb{R}_{+,c}^n$. Therefore,

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

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

This upper bound is sharp. Indeed, for the points $x = (0, c, \dots, c)$, $x' = (1, c, \dots, c)$, and $y' = (c^2, c, \dots, c)$, 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 $x' - x = (1, 0, \dots, 0)$ as $k \rightarrow \infty$. For the segments $[x, y^k]$ the corresponding ratio tends to c^2 as $k \rightarrow \infty$.

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 $X' \in \text{int}\mathcal{K}_n$. We have $(X'b, b) > (Y'b, b)$, and therefore,

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

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

The case of the Lorentz cone is elementary. For the dimension $n = 2$ the inequality $\frac{|x - y'|}{|x - x'|} \leq \left(\frac{\sin(\varphi + \psi)}{\sin(\varphi - \psi)}\right)^2$ 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 crosssections of the cones K_φ and K_ψ by this plane are also coaxial Lorentz cones of some angles φ' and ψ' , respectively, for which, moreover, $\frac{\sin(\varphi' + \psi')}{\sin(\varphi' - \psi')} \leq \frac{\sin(\varphi + \psi)}{\sin(\varphi - \psi)}$. Hence, the general case follows from the case $n = 2$. \square

Appendix C. Lifting the transpose set does not improve the estimate.

By applying the semidefinite lifting (2.2) 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}^T to get another estimate. Even though the obtained matrices are not the transpose of the initial matrices, this estimate will not be better, as shown by the next proposition.

PROPOSITION C.1. *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 $\hat{\sigma}(\tilde{\mathcal{M}}) = \hat{\sigma}(\tilde{\mathcal{M}}^T)$; that is, applying the semidefinite lifting to \mathcal{M} or \mathcal{M}^T does not change the quality of approximation of $\hat{\sigma}$.*

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

$$(C.1) \quad \min_{\|\cdot\|} \gamma \quad \text{for all } A \in \mathcal{M}, \quad \|A\| \leq \gamma$$

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

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

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 ,

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

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

$$(C.2) \quad \max_{A \in \mathcal{M}^T} \{\|A\|_*\} \leq \max_{A \in \mathcal{M}} \{\|A\|\}.$$

Indeed,

$$(C.3) \quad \|A^T\|_* = \max_{|y|_*=1} |A^T y|_*,$$

$$(C.4) \quad = \max_{|y|_*=1} \max_{|x|=1} y^T A x,$$

$$(C.5) \quad \leq \max_{|y|_*=1} \max_{|x|=1} |A x| \cdot y^T (A x / |A x|),$$

$$(C.6) \quad \leq \|A\|.$$

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

Acknowledgments. We would like to thank Steven Finch for pointing out the application studied in section 4.2. We are also thankful to the reviewers who have provided valuable comments that have helped to improve the quality of the paper.

REFERENCES

- [1] V. M. ALEXEEV, E. M. GALEEV, AND V. M. TIKHOMIROV, *Book of Problems in Optimization*, Nauka, Moscow, Russia, 1984.
- [2] T. ANDO AND M.-H. SHIH, *Simultaneous contractibility*, SIAM J. Matrix Anal. Appl., 19 (1998), pp. 487–498.
- [3] N. BARABANOV, *Lyapunov indicators of discrete inclusions I-III*, Autom. Remote Control, 49 (1988), pp. 152–157, 283–287, 558–565.
- [4] M. A. BERGER AND Y. WANG, *Bounded semigroups of matrices*, Linear Algebra Appl., 166 (1992), pp. 21–27.
- [5] V. D. BLONDEL AND V. CANTERINI, *Undecidable problems for probabilistic automata of fixed dimension*, Theory Comput. Syst., 36 (2003), pp. 231–245.
- [6] V. D. BLONDEL, J. CASSAIGNE, AND R. M. JUNGERS, *On the number of α -power-free words for $2 < \alpha \leq 7/3$* , Theoret. Comput. Sci., 410 (2009), pp. 2823–2833.
- [7] V. D. BLONDEL, R. JUNGERS, AND V. PROTASOV, *On the complexity of computing the capacity of codes that avoid forbidden difference patterns*, IEEE Trans. Inform. Theory, 52 (2006), pp. 5122–5127.
- [8] V. D. BLONDEL AND Y. NESTEROV, *Computationally efficient approximations of the joint spectral radius*, SIAM J. Matrix Anal. Appl., 27 (2005), pp. 256–272.
- [9] V. D. BLONDEL AND Y. NESTEROV, *Polynomial-time computation of the joint spectral radius for some sets of nonnegative matrices*, SIAM J. Matrix Anal. Appl., 31 (2009), pp. 865–876.
- [10] V. D. BLONDEL, Y. NESTEROV, AND J. THEYS, *On the accuracy of the ellipsoid norm approximation of the joint spectral radius*, Linear Algebra Appl., 394 (2005), pp. 91–107.
- [11] V. D. BLONDEL AND J. N. TSITSIKLIS, *The boundedness of all products of a pair of matrices is undecidable*, Systems Control Lett., 41 (2000), pp. 135–140.
- [12] S. BOYD AND L. VANDENBERGHE, *Convex Optimization*, Cambridge University Press, Cambridge, 2004.
- [13] J. CASSAIGNE, *Counting overlap-free binary words*, in STACS 93, Lecture Notes in Comput. Sci. 665, Springer, Berlin, 1993, pp. 216–225.
- [14] I. DAUBECHIES AND J. C. LAGARIAS, *Two-scale difference equations II. Local regularity, infinite products of matrices and fractals*, SIAM J. Math. Anal., 23 (1992), pp. 1031–1079.

- [15] N. G. DE BRUIJN, *On Mahler's partition problem*, Indag. Math. (N. S.), 10 (1948), pp. 210–220.
- [16] G. A. DERFEL, N. DYN, AND D. LEVIN, *Generalized refinement equations and subdivision processes*, J. Approx. Theory, 80 (1995), pp. 272–297.
- [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*, <http://www.citebase.org/abstract?id=oai:arXiv.org:0802.2654> (2008).
- [20] J. GOLDWASSER, W. KLOSTERMEYER, M. MAYS, AND G. TRAPP, *The density of ones in Pascal's rhombus*, Discrete Math., 204 (1999), pp. 231–236.
- [21] G. GRIPENBERG, *Computing the joint spectral radius*, Linear Algebra Appl., 234 (1996), pp. 43–60.
- [22] N. GUGLIELMI AND M. ZENNARO, *Finding extremal complex polytope norms for families of real matrices*, SIAM J. Matrix Anal. Appl., 31 (2009), pp. 602–620.
- [23] L. GURVITS, *Stability of discrete linear inclusions*, Linear Algebra Appl., 231 (1995), pp. 47–85.
- [24] R. M. JUNGERS, *The Joint Spectral Radius: Theory and Applications*, Lecture Notes in Control and Inform. Sci. 385, Springer-Verlag, Berlin, 2009.
- [25] R. M. JUNGERS, V. PROTASOV, AND V. D. BLONDEL, *Overlap-free words and spectra of matrices*, Theoret. Comput. Sci., 410 (2009), pp. 3670–3684.
- [26] V. S. KOZYAKIN, *Structure of extremal trajectories of discrete linear systems and the finiteness conjecture*, Autom. Remote Control, 68 (2007), pp. 174–209.
- [27] M. LOTHAIRE, *Combinatorics on Words*, Encyclopedia of Math. Appl. 17, Cambridge University Press, Cambridge, 1982.
- [28] K. MAHLER, *On a special functional equation*, J. London Math. Soc., 15 (1940), pp. 115–123.
- [29] P. PARRILO AND A. JADBABAIE, *Approximation of the joint spectral radius using sum of squares*, Linear Algebra Appl., 428 (2008), pp. 2385–2402.
- [30] V. Y. PROTASOV, *The joint spectral radius and invariant sets of linear operators*, Fundam. Prikl. Mat., 2 (1996), pp. 205–231.
- [31] V. Y. PROTASOV, *The generalized spectral radius, A geometric approach*, Izv. Math., 61 (1997), pp. 995–1030.
- [32] V. Y. PROTASOV, *Asymptotic behaviour of the partition function*, Sb. Math., 191 (2000), pp. 381–414.
- [33] V. Y. PROTASOV, *Refinement equations with nonnegative coefficients*, J. Fourier Anal. Appl., (2000), pp. 55–77.
- [34] V. Y. PROTASOV, *On the asymptotics of the binary partition function*, Math. Notes, 76 (2004), pp. 144–149.
- [35] V. Y. PROTASOV, *On the regularity of de Rham curves*, Izv. Math., 68 (2004), pp. 567–606.
- [36] V. Y. PROTASOV, *Fractal curves and wavelets*, Izv. Math., 70 (2006), pp. 123–162.
- [37] J. RADON, *Über eine Erweiterung des Begriffs der konvexen Funktionen, mit einer Anwendung auf die Theorie der konvexen Körper*, Sitz.ber. Akad. Wiss. Wien, 125 (1916), pp. 241–258.
- [38] B. REZNICK, *Some binary partition functions*, in *Analytic Number Theory: Proceedings of a Conference in Honor of Paul T. Bateman*, B. C. Berndt, H. G. Diamond, H. Halberstam, and A. Hildebrand, eds., Boston, Birkhäuser, 1990, pp. 451–477.
- [39] G. C. ROTA AND G. STRANG, *A note on the joint spectral radius*, Nederl. Akad. Wetensch. Proc. Ser. A63 Indag. Math., 22 (1960), pp. 379–381.
- [40] J. N. TSITSIKLIS AND V. D. BLONDEL, *The Lyapunov exponent and joint spectral radius of pairs of matrices are hard - when not impossible - to compute and to approximate*, Math. Control, Signals Systems, 10 (1997), pp. 31–40.
- [41] D.-X. ZHOU, *The p-norm joint spectral radius for even integers*, Methods Appl. Anal., 5 (1998), pp. 39–54.